



Issue 52 • April 2024
datacenterdynamics.com

3D printing a data center

Also inside:

Red Sea terror
Atos' challenges
Khazna's CEO





Vertiv.com

Avoid the Upcharge

Time-shift power usage.

Store energy and use it when needed, avoiding grid constraints and peak demand charges. **Vertiv™ DynaFlex BESS** gives you greater control over your power usage and costs.



Know More About
Dynamic Power Solutions



Contents

April 2024



- 6 **News**
Equinix's CEO steps down, AWS' nuclear DC, and new Nvidia GPUs
- 14 **The Wavehouse: 3D printing a data center**
The story behind the world's first 3D printed data center

- 21 **This Emirati operator who doesn't use diesel** ★
Hassan Alnaqbi's AI-backed business is working for oil exploitation

- 25 **News from the hydrogen frontline**
Yuval Bachar on a 1MW data center

- 29 **The Edge supplement**
From AI to space

- 45 **HAMR - the savior of hard drives**
What happens when hard drive capacities double?

- 49 **Kestler's run**
Building a hyperscale wholesaler for the mega data center era

- 53 **Betting on the US market**
The digital infrastructure holding up online gambling

- 58 **How do you solve a problem like Atos?**
What next for the French IT giant?

- 62 **Running to the storm**
Verizon prepares for natural disasters

- 65 **Cool runnings**
Oil companies like Castrol are looking to data center cooling

- 68 **The Houthis and the Red Sea: A new risk to subsea cables**
Intentional or otherwise, terrorists are now a threat to global connectivity

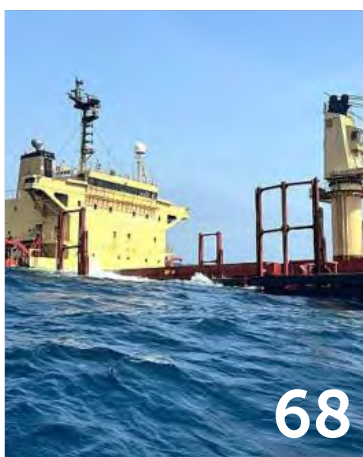
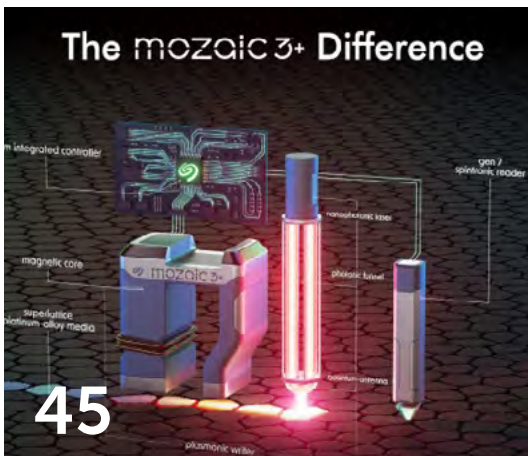
- 74 **The non-profit approach**
Talking to British Heart Foundation's CTO

- 77 **Great X-pectations**
Europe's grand plan for a federal cloud appears to have hit the buffers

- 80 **MWC 2024**
AI comes for mobile

- 82 **Is AI the panacea the chip design industry needs?**
Can AI help the chip crunch?

- 87 **Op-ed: The AI goldilocks zone**





Think Fiber

Think CommScope

The pressure—to deploy faster, manage smarter, improve sustainability—is intensifying. Relax and focus with infrastructure fiber solutions from CommScope. Streamlined and modular, they accelerate deployment and improve change management. High-speed, low-loss performance supports AI, ML, and 800G. Unmatched service means complete peace of mind.

That's grace under pressure.
That's CommScope.

Scan below to learn more or visit
commscope.com/think-fiber



From the Editor

Making a 3D printed data center

Construction is so much work. Wouldn't it be easier if you could just print your next data center? Unfortunately, it's probably not going to be as simple as that.

On-site generation

For the cover, we travel to Germany to understand whether the future of data center construction lies in 3D printing.

We tour The Wavehouse data center, the first of its kind, and talk to the creators about the promises and

3D printing could prove cheaper and faster, while using less metal.

peril of on-site creation - 3D-printed warts and all.

The CEOs

Across the mag, we profile three different CEOs with three very different approaches.

First, there's Khazna's Hassan Alnaqbi, who is developing large data centers in the desert.

Then there's Yuval Bachar, founder of EdgeCloudLink, who hopes to use hydrogen power to fulfil the dream of "zero emission and zero water use" data centers.

Finally, we chat to EdgeCore's Lee Kestler. After a long stint at DuPont Fabros, he's tasked with building a new DuPont for the mega-data center era.

All bets are on-premise

Building a data center can sometimes feel like a bit of a gamble.

But, for this issue, we look at the real risk takers - casinos. Due to regulatory requirements, they have to ensure data centers don't stray too far from the floor.

When the storm comes

Climate change, war, and insurrection.

It's not fun to think about what the future holds, but at least we'll have Internet. We head to a US military base to learn first hand how Verizon trains for the worst.

European failures

Trying to be a tech company in the shadow of the US hyperscalers can be a challenge.

For Atos, once a French IT darling, things have gone from bad to worse.

Elsewhere, the EU's dreams of stitching together a local cloud platform has gone awry. Our new features editor investigates.

Trouble at sea

The Red Sea is in turmoil, and the Internet is one of the casualties.

Houthi attacks appear to have cut submarine cables. What does this mean for the future of Internet stability?

Plus more

Why are oil companies getting into data center cooling? How is the British Heart Foundation using the cloud? What did we learn from MWC?

We've brought on a few new writers to tell more stories about this sector than ever. But don't worry, we're hiring for some more. Stay tuned!

172 hrs

The time it took to 3D print the Wavehouse. But it was spread over 30 days.



Sebastian Moss
Editor-in-Chief

Meet the team

Publisher & Editor-in-Chief
Sebastian Moss
[@SebMoss](#)

Senior Editor
Dan Swinhoe
[@DanSwinhoe](#)

Telecoms Editor
Paul Lipscombe

Compute, Storage & Networking Editor
Charlotte Trueman

Reporter
Georgia Butler

Features Editor
Matthew Gooding

Junior Reporter
Niva Yadav

Head of Partner Content
Claire Fletcher

Partner Content Editor
Chris Merriman
[@ChrisTheDJ](#)

Designer
Eleni Zevgaridou

Head of Sales
Erica Baeta

Conference Director, Global
Rebecca Davison

Content & Project Manager - Live Events
Gabriella Gillett-Perez

Content & Project Manager - Live Events
Matthew Welch

Channel Management Team Lead
Alex Dickins

Channel Manager
Kat Sullivan
Emma Brooks

Director of Marketing Services
Nina Bernard

CEO
Dan Loosemore

Head Office

DatacenterDynamics
22 York Buildings,
John Adam Street,
London, WC2N 6JU

© 2024 Data Centre Dynamics Limited All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, or be stored in any retrieval system of any nature, without prior written permission of Data Centre Dynamics Limited. Applications for written permission should be directed to the editorial team at editorial@datacenterdynamics.com. Any views or opinions expressed do not necessarily represent the views or opinions of Data Centre Dynamics Limited or its affiliates. Disclaimer of liability. Whilst every effort has been made to ensure the quality and accuracy of the information contained in this publication at the time of going to press, Data Centre Dynamics Limited and its affiliates assume no responsibility as to the accuracy or completeness of and, to the extent permitted by law, shall not be liable for any errors or omissions or any loss, damage or expense incurred by reliance on information or any statement contained in this publication. Advertisers are solely responsible for the content of the advertising material which they submit to us and for ensuring that the material complies with applicable laws. Data Centre Dynamics Limited and its affiliates are not responsible for any error, omission or material. Inclusion of any advertisement is not intended to endorse any views expressed, nor products or services offered, nor the organisations sponsoring the advertisement.

Dive even deeper

Follow the story and find out more about DCD products that can further expand your knowledge. Each product is represented with a different icon and color, shown below.



Events



Intelligence



Debates



Training



Awards



CEEDA

News

The biggest data center news stories of the last three months

Equinix CEO Meyers to stand down, Google Cloud's Adaire Fox-Martin to take helm

Equinix has appointed Google Cloud executive Adaire Fox-Martin as its new CEO.

Fox-Martin, who is currently Google Cloud's go-to-market president, will succeed Charles Meyers, who has led the company for six years.

Also head of Google Ireland, Fox-Martin has led Google Cloud's go-to-market teams globally since early last year, encompassing sales, professional services, partner ecosystem, and customer success. She has held senior global positions at Oracle and SAP, where she was a member of the company's executive board with responsibility for sales and service globally, serving more than 440,000 customers.

She has been a member of the Equinix board of Directors since 2020.

"In today's dynamic digital landscape, Equinix has uniquely amassed global reach, highly differentiated ecosystems, strong partner relationships, and an innovative range of product and service offerings - collectively forming a robust and future-proofed platform to address diverse customer challenges," said Fox-Martin.

Myers is moving into a new role as executive chairman of the colocation giant.



Joining Equinix as Americas president in 2010, Meyers has been CEO since 2018.

He took over from Steve Smith, who resigned after "exercising poor judgment with respect to an employee matter."

Meyers said: "I am confident that Adaire's capabilities and experience will be deeply additive to our team and our culture, helping us meet the evolving needs of our customers, fuel our growth, and unlock the extraordinary power of Platform Equinix.

"I am grateful to our board for their support of my desired transition timeline, and excited to assume the executive chairman role and actively support Adaire as she leverages her tremendous global experience to extend and expand our market leadership."

Before Equinix, Meyers held roles at Bell South (later AT&T), IBM, Level3, and VeriSign.

The leadership transition will take place in Q2 2024. Peter Van Camp, currently Equinix executive chairman, will step away from his formal responsibilities as a board member to take the role of special advisor to the board.

bit.ly/MyeMyeCharlieMyeMye

NEWS IN BRIEF

Georgia Senate passes bill suspending tax exemptions

The Georgia Senate has voted to suspend tax breaks for data centers while a special commission reviews grid capacity in the state. The Senate voted 29-22 in favor in March.

Involta announces Brett Lindsey as new CEO

US colocation firm Involta has appointed a new CEO, Brett Lindsey. He replaces Jim Buie, who was at the company for six years. Lindsey left Everstream in 2023, after DigitalBridge bought its parent firm.

Compass co-founder Chris Curtis joins Prologis

Logistics real estate giant Prologis has hired Compass co-founder Chris Curtis to lead a new data center team. The company hired Curtis in a newly created role as global head of data centers.

Kao Data appoints Doug Loewe as CEO

UK data center firm Kao Data has appointed Doug Loewe as its new chief executive officer (CEO). Loewe took the helm in January. Former CEO Lee Myall announced he was stepping down from the role in April 2023; Kao founder David Bloom had been acting as interim CEO.

Network Time Protocol inventor Dr. David L Mills passes away

Dr. David L. Mills, the inventor of the Network Time Protocol (NTP), has passed away aged 85. He created NTP in 1985 to synchronize time across different computer systems and networks, a critical step in the development of the modern Internet.

Exa Infrastructure poaches Aqua Comms CEO

Aqua Comms CEO Jim Fagan has jumped ship and taken the top job at Exa Infrastructure. The two companies announced in February that Fagan was leaving Aqua Comms, effective immediately, and heading across to Exa Infrastructure, where he will take over the CEO role in August.



Microsoft, Google, and AWS drop some data egress fees

Three major US cloud providers have dropped some egress fees ahead of incoming European regulation. Microsoft, Google, and Amazon Web Services (AWS) have each dropped fees for customers looking to exit their respective clouds.

March saw Microsoft announce it will no longer charge data egress fees for customers exiting its Azure cloud, following in the footsteps of rivals AWS and Google Cloud.

The company is the last of the three major cloud providers to make this move ahead of the impending EU Data Act that will come into effect in September 2025.

In a blog post, Microsoft said: "We support customer choice, including the choice to migrate your data away from Azure.

To exit the provider, you must inform Azure of the date you plan to start the transfer and the amount of data you plan to egress. Once logged, you have 60 days to comply from your start date, and must then cancel all subscriptions. After exiting, you can claim credit for your egress charges.

March also saw AWS announce similar plans to remove some of its data transfer fees for customers. Unlike Microsoft and Google, it seems that customers do not have to fully exit AWS as a customer, but similarly has a rigorous approval process and allows only 60 days to complete the migration.

Google was the first of the big three to waive egress fees. In January, the company announced: "Starting today, Google Cloud customers who wish to stop using Google Cloud and migrate their data to another

cloud provider and/or on-premises, can take advantage of free network data transfer to migrate their data out of Google Cloud. This applies to all customers globally."

The free data transfer from Google does come with some caveats. It only applies to customers looking to fully exit Google's Cloud, their application will first be reviewed by the Google Cloud Support team which must approve it. Following approval, customers have only 60 days to fully exit the provider and must then terminate their agreement.

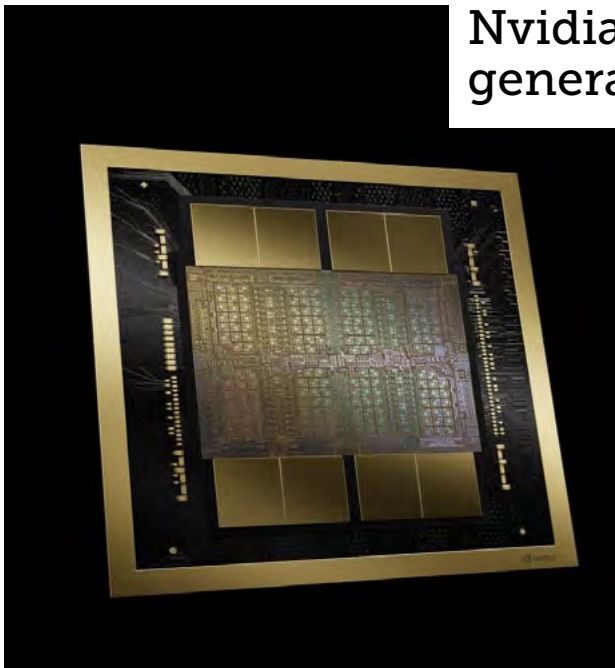
The upcoming EU Data Act requires public and private cloud computing service providers to remove "obstacles to effective switching" between their own and competing cloud services, including commercial, contractual, technical, or organizational hurdles.

The act came into force on January 2024, and provisions of the act will begin to apply 20 months later meaning that businesses must be ready to comply by September 2025

The cloud computing market has been the topic of several discussions and investigations regarding its anti-competitiveness.

The UK's Competition and Markets Authority has launched an investigation into the market after Ofcom identified several features that make it difficult for customers to switch and use multiple cloud suppliers including egress fees, cloud repatriation fees, and discounts.

bit.ly/PassiveEgression



Nvidia announces next GPU generation, Blackwell

Nvidia has unveiled its next GPU family - the Blackwell.

The company claimed that the new GPU architecture would allow customers to build and run real-time generative AI on trillion-parameter large language models at 25x less cost and energy consumption than its predecessor, the Hopper series.

Amazon, Google, Meta, Microsoft, Oracle Cloud, and OpenAI are among the companies that confirmed they will deploy Blackwell GPUs later this year.

Blackwell is named in honor of David Harold Blackwell, a mathematician who specialized in game theory and statistics, and the first Black scholar inducted into the National Academy of Sciences.

Blackwell architecture GPUs are manufactured using a custom-built, two-reticle limit 4NP TSMC process with GPU dies connected by 10TBps chip-to-chip link into a single, unified GPU.

The GPU has 208 billion transistors, an increase on the 80bn in the Hopper series. It is twice the size of the Hopper.

According to Nvidia, the B100, B200, and GB200 operate between 700W and 1,200W, depending on the SKU and type of cooling used.

bit.ly/Blackwellwellwell

Microsoft and Google invest in carbon capture projects



Microsoft and Google are making large investments into carbon capture projects in an effort to reduce their carbon footprints.

Microsoft, which has promised to be carbon negative by 2030, has made a number of investments across several carbon capture methods in recent months.

Microsoft has signed a carbon capture agreement which will see it purchase 95,000 tons of carbon dioxide removal credits from a biochar plant in Mexico.

It has agreed a six-year purchasing deal with The Next 150 and its General Biochar Systems (GBS) business unit, which runs the biochar plant near the city of Guanajuato.

Biochar is produced by burning biomass in a zero-oxygen environment at a temperature of 250°C (482°F) or above. This

thermal process, known as pyrolysis, enables the biochar to store large amounts of carbon. The resulting biochar can then be used as a soil enhancer, and it is thought the biochar retains the carbon for thousands of years.

It is the second biochar agreement Microsoft has signed in recent months. In December 2023, it agreed a deal with Exomad Green, the world's largest producer of biochar, for 32,000 tons of carbon dioxide removal credits to be delivered by June 2024 from its plant in Bolivia. It also previously signed a deal with Carbon Streaming to buy credits from its biochar site in Virginia.

February saw Microsoft two other carbon capture deals. The Redmond-based company paid to have 27,600 tons of carbon dioxide removed from the atmosphere and stored in waste concrete.

Switzerland-based Neustark will capture the CO₂ at source over the next six years, and store it in mineral waste, such as demolition concrete.

Neustark captures CO₂ from biogas plants, liquifies it, and transports it to construction waste recycling sites, where it is injected into demolition concrete, and bound permanently to the surface of the concrete granules by a process of mineralization.

February also saw Microsoft announced it was paying climate finance firm Catona for 350,000 tonnes of carbon removal credits from an agroforestry project in Kenya.

The Lake Victoria Watershed Agroforestry Project in Homa Bay, Kenya will deliver the credits in the next six years. It is financed by Catona and run by Trees for the Future; a 30-year-old global agroforestry organization based in Maryland, USA.

Agroforestry is a land use management system where trees or shrubs are grown around or among crops or pastureland.

The company has also invested in direct air capture projects in the US and Iceland.

Google recently announced it will invest \$35 million into carbon capture projects through Frontier Climate – which makes aggregated purchases of carbon capture credits from a number of companies projects, and technologies.

Google said it had made "recent purchases" through Frontier, but didn't provide details.

bit.ly/GottaCaptureThemAll

Death reported at London substation being built for Microsoft data center



A fatal incident has been reported at an electrical substation being built to supply power to a new Microsoft data center in London.

It is understood that a worker was trapped after an incident involving an aerial work platform at the Park Royal site in Ealing in February, and died of his injuries in hospital the following week.

Few details are available, but the accident took place at the Powergate Project site, where H&MV Engineering is building an electrical substation for Microsoft. The victim was reportedly a man in his late 20s.

A spokesperson for H&MV Engineering said the worker was not a company employee, and it was investigating the incident.

Microsoft has a data center construction site in West Acton (recently visited by Prince William, to discuss concerns about deaths from suicide in construction). The event took place away from the data center site itself, at an electrical substation being built to power the facility.

A Microsoft spokesperson said: "We are deeply saddened by an incident at our construction site in Park Royal, Ealing in which a contractor subsequently lost their life. Our thoughts are with their family, friends, and colleagues at this difficult time. An incident investigation is currently underway, and support being offered to those affected. The health and safety of anyone working on Microsoft sites is a core priority and we are committed to ensuring the highest standards are met in the delivery of our projects."

bit.ly/SubstationDeathLondon

OPERATIONAL EXCELLENCE FOR DATA CENTRES STARTS HERE



THE FUTURE IS WHAT WE MAKE IT

| **Honeywell**

Alphabet spin-out Sidewalk Infrastructure Partners launches data center unit

Alphabet spin-out Sidewalk Infrastructure Partners (SIP) has launched a new data center-focused unit.

"SIP is taking a lead on the private sector solutions by bringing Verrus — a new SIP platform company that builds next-generation data centers — out of stealth after several years of internal research and development," the company announced in March.

SIP said Verrus is "pioneering a new generation of flexible, efficient data centers tailored for both conventional workloads and AI" by matching its infrastructure to the variable uptime requirements and actual workload needs.

As an example, SIP said separate server halls can house availability-sensitive computing like web searches and flexible batch operations like AI training. This workload-aware approach, it said, maximizes efficiency.

SIP said the modular Verrus design can share power assets between halls to improve utilization rather than stranding capacity. Verrus can also deploy grid-scale battery storage rather than diesel generators that can support the grid if needed.

Verrus' orchestrated software system can coordinate uptime preservation while optimizing costs, emissions, and grid support.

Verrus said it utilizes "innovative cooling techniques" to support extreme chip heat dissipation from power-hungry GPUs — likely liquid-cooling, but without providing details.

"In place of redundant, isolated, and unsustainable power assets, we implement optimized electrical and mechanical plant architecture to fluidly

align capacity to loads across the facility. Supply adapts to sync with demand (and vice versa), rather than sitting idle," Verrus' own announcement said. "Our optimization can dynamically calibrate operations to sync with workload variability, grid interactions, and carbon availability."

Verrus' features a number of former Google staff amongst its leadership team. The company is led by CEO Nelson Abramson. Prior to joining Verrus, Nelson served as the global head of Infrastructure for X (formerly Twitter). He also previously worked at Google for 20 years, including on its data center team.

SIP was formed out of Sidewalk Labs in 2019 as an independent entity out of Alphabet.

bit.ly/3TQ1deb



AWS acquires Talen's nuclear data center campus in Pennsylvania

Amazon Web Services (AWS) has acquired Talen Energy's data center campus at a nuclear power station in Pennsylvania.

March saw Talen Energy Corporation announce it has sold its 960MW Cumulus data center campus in Pennsylvania to a 'major cloud service provider' — listed as Amazon in a Talen investor presentation. Amazon is yet to comment on the news.

Talen sold the site and assets of Cumulus Data for \$650 million. The deal comprises \$350 million at close and \$300m escrowed, released upon development milestones anticipated in 2024.

First announced by DCD in July 2021, the 1,200-acre campus draws power from Talen Energy's neighboring 2.5GW nuclear power station in Luzerne County, the Susquehanna Steam Electric Station (SSES). The company broke ground in 2021 and completed the first 48MW, 300,000 square foot (28,870 sqm) hyperscale facility early last year, along with a separate cryptomine facility.

Talen said AWS aims to develop a 960MW data center campus, and has signed a 10-year PPA with the to procure power from the nearby plant.

bit.ly/4asZa6Y

Tract files for 30-building campus in Buckeye, Arizona

Data center park developer Tract is reportedly planning a new campus in Phoenix, Arizona.

Tract is said to be plotting a \$14 billion master-planned data center complex across 1,000 acres in the Buckeye area of Maricopa County.

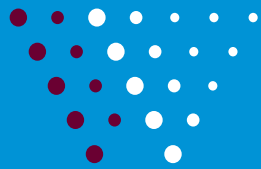
Known as Project Range, the development will span nearly 30 buildings totaling 5.6 million square feet on land north and south of Yuma Road between Jackrabbit Trail and Perryville Road. Buildings will range from 149,000 square feet to 260,000 square feet each.

Work on the project is set to start in 2025 and continue over the next 15 to 18 years.

Tract is led by Grant van Rooyen, president of the van Rooyen Group and founder of US data center firm Cologix. Stonepeak Infrastructure Partners acquired a majority stake in the company in 2017.

Tract is planning other large parks in Reno, Nevada; Eagle Mountain, Utah; and Richmond, Virginia.

bit.ly/3VR7cCm



Winthrop Technologies

Turnkey Delivery of Digital Infrastructure



A dedicated data centre delivery partner,
headquartered in Dublin and delivering
throughout Europe.

winthrop.ie

HPE acquires Juniper Networks for \$14 billion

HPE is to acquire network gear maker Juniper Networks for \$14 billion.

The all-cash transaction was approved by the boards of both companies in January.

The acquisition is expected to double HPE's networking business and "create a new networking leader with a comprehensive portfolio," according to the companies.

The transaction is expected to close in late 2024 or early 2025.

Juniper CEO Rami Rahim will lead the combined HPE networking business.

Founded in 1996 and based in Sunnyvale, California, NYSE-listed Juniper provides networking products, including routers, switches, network management software, network security products, and software-defined networking (SDN) technology.

HPE offers switches via its Aruba Networks unit – acquired in 2015 just prior to Hewlett Packard's split into two companies.

Since it was broken off from HP in 2015, HPE's acquisitions include SGI, Cray, SimpliVity, and Nimble Storage. This is HPE's largest deal since becoming a standalone company by a wide margin and one of the largest in HP's long history.

bit.ly/3xkwwGn

Ardian acquires Verne Global; D9 to wind down after asset sales



Investment firm Ardian has completed the acquisition of European data center firm Verne. Verne's previous owner, D9, is set to wind down after selling off its other assets.

Terms of the sale were not disclosed – though reports suggest £465 million (\$575m).

Ardian said it will support the expansion of Verne with up to \$1.2bn committed investment through equity and debt to "deliver an ambitious growth plan" for Northern Europe.

Ardian said it plans to multiply Verne's existing sold capacity of 29MW for 2023 by "close to four times" in the medium term. This will include expanding existing sites and entering new locations with a focus on Iceland, Finland, Sweden, and Norway, as well as potential opportunities across the rest of Northern Europe.

Triple Point-owned D9 announced it was selling Verne to Ardian in November 2023.

D9 bought Verne for £231 million in 2021 (\$320m at the time). The company operates a 24MW data center campus on a former NATO site near Keflavik, Iceland, offering colocation and high-performance computing services. D9 has since acquired and merged Finnish data center firm Ficolo and London's Volta Data Centres into the Verne brand.

Amid growing debt and a dropping share price, D9 initially said that it would sell a minority stake in Verne, but shareholders pushed for a total sale.

D9 is in the process of winding down and will be selling off its other assets. As well as Verne, D9's investments include subsea cable firm Aqua Comms, UK terrestrial television and radio broadcasting network Arqiva, Irish broadband firm Elio (previously Host Ireland), a stake in the EMIC-1 cable, and the SeaEdge UK1 data center & cable landing station in the UK.

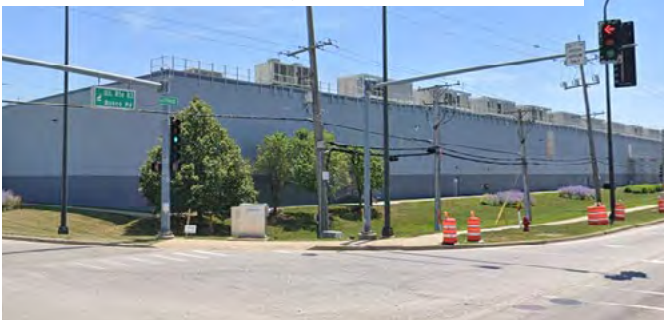
bit.ly/4aITDZI



Peter's factoid

Nvidia CEO Jensen Huang believes that, over the next four to five years, a trillion dollars' worth of data center infrastructure and hardware will be built across the world.

Brookfield completes \$775m deal for bankrupt Cyxtera



Brookfield has completed its \$775m purchase of Cyxtera, merging the bankrupt provider with its Evoque colo brand to create a combined company with more than 50 locations.

The deal, first announced in November, sees Brookfield acquire the majority of Cyxtera's assets. Evoque CEO Spencer Mullee will lead the newly merged organization.

Cyxtera filed for bankruptcy last June with debts of more than a billion dollars and regular quarterly losses.

Combining Evoque and Cyxtera is "a strategic step" that "aligns seamlessly" with Brookfield's vision for data infrastructure investments, according to Udhay Mathialagan, chair of the board at Evoque and CEO for global data centers at Brookfield Infrastructure.

Forty additional data centers are being incorporated into Evoque as part of the deal, most of which are based in the US and Canada.

Cyxtera was formed in 2021 out of CenturyLink's colocation business.

bit.ly/49vmmnA



COMPREHENSIVE ENGINEERING EXCELLENCE

www.anordmardix.com



The Wavehouse: 3D printing a data center



Dan Swinhoe
Senior Editor

The story behind the
world's first 3D printed
data center

Walking down the street of a former US Army Barracks in the quiet German city of Heidelberg, one may come across a strange grey building. Nestled between some pre-WWII offices and a kindergarten, the windowless concrete shell seems oddly out of place.

While the location might seem a little odd on paper, in-the-know *DCD* readers will say a windowless concrete building sounds like your run-of-the-mill data center. But this one is different.

For one, it has no corners. Its rounded external shell undulates and twists along the length of the property. And, instead of traditional blocks or pre-cast walls, this data center's fascias were built up in layers of liquid cement.

Known as the 'Wavehouse Campbell Heidelberg,' this building is the world's first 3D printed data center.

Photos: Dan Swinhoe

3D printing a data center

Heidelberg, located around 50 miles south of Frankfurt on the river Neckar, is a city of around 160,000 people in the state of Baden-Württemberg. It is known for having a reputable university and several high-quality scientific research facilities.

Fittingly, the city is home to Heidelberger Druckmaschinen, a 170-year-old company and one of the world's largest manufacturers of printing presses. The fact the city is now also home to an early landmark of 3D printing, however, is just a chance occurrence made possible by a few local businessmen.

Originally, local managed IT services provider Heidelberg iT and real estate firm Kraus Gruppe planned to develop a traditional right-angled data center on the site.

However, after reading about a 3D-printed residential building built by local materials firm Heidelberg Materials (formerly HeidelbergCement), Hans-Jörg Kraus, managing partner of the Kraus Group, decided to phone his longtime friend Mattias Blatz at Heidelberg iT and ask if he wanted to 3D print the facility instead.

When asked why he wanted to 3D print a building, Kraus smiles at DCD and simply says;

merging with MEG Kommunikationssysteme GmbH, Heidelberg iT can trace its roots back to the launch of its first data center in cooperation with MVV Manet GmbH in 2001. A second facility was launched in 2016.

The company is set to launch a third – traditionally constructed – data center in the coming months that will offer 500kW – the same as the 3D printed facility. But that unremarkable site holds few accolades compared to its 3D-printed compatriot.

The Wavehouse simultaneously has the titles of Europe's largest 3D printed building, the world's first purpose-built 3D printed industrial building, and the world's first 3D printed data center.

The parametric design of the Wavehouse is in stark contrast to another data center recently launched nearby; a 2MW facility from InfraRed Capital-backed Nexspace located less than a mile away and launched in late 2023. That building – boasting smooth green and white panels and right angles at every corner – is the only real competition Heidelberg iT faces in a city not known for its data center developments.

Heidelberg is amongst the warmest regions of Germany – warm enough to grow grapes,

The Wavehouse

Traditionally, data centers are either built using masonry blocks or 'tilt-up construction' – where concrete wall panels are cast on-site and 'tilted' into position using cranes. 3D printing, however, uses a printer to extrude layers of concrete upon one another, building the walls up layer by liquid layer.

The Wavehouse data center is situated to the south of the city center, on the site of the former US Army Barracks. Previously known as Baufeld 5 (literally translating to "Construction site 5") prior to construction and located at Billie-Holiday-Strasse 7 in Baden-Württemberg, the single-story building totals just under 600 sqm (6,600 sq ft); 54 meters long, 11 meters wide, and 9 meters high.

Partners on the project include Peri 3D, which is part of the scaffolding firm Peri Group, as well as architecture firms Mense-Korte and SSV, and Heidelberg Materials.

Peri began work on the facility *last year* using a Cobod BOD2 3D construction printer. The facility was printed over 30 days using around 450 tons of printing concrete from Heidelberg Materials. Printed between late March and mid-August 2023, the facility was

"We had planned a data center here. And then my friend Kraus had the idea with the 3D and I said yes, I want to do it."

*>>Mattias Blatz,
Heidelberg iT*



"I am like a child. I love to make new things. It's beautiful to see."

He continues: "I read in a real estate article that Heidelberg Materials printed a family house. I knew I had to build a data center for Mattias, and thought it would be perfect to print. Then I spoke with people I know from Heidelberg Materials and I said 'I want to print this building.'"

"We had planned a data center here," adds Blatz, owner of Heidelberg iT. "And then my friend [Kraus] had the idea with the 3D and I said yes, I want to do it."

"I was very excited because I've known him so long that I know when he has an idea, I trust him."

Though officially launched in 2007 after

almonds, figs, and palm trees. Blatz says the relatively small local market, temperature, and proximity to Frankfurt mean Heidelberg doesn't need a large data center ecosystem and would struggle to support any hyperscale developments.

"There is only one reason to build a data center in Heidelberg, for the customers," says Blatz. Several healthcare organizations based in the area have sensitive instruments with low latency requirements – he gives MRI equipment as an example – meaning they prefer to keep IT infrastructure local.

"There is a small business economy," he adds. "It would make no sense in the hottest city in Germany to build a data center for other things; it's better to go Iceland, Sweden, Austria, or wherever, but not in Heidelberg."

originally due to take 140 hours, equivalent to printing four square meters of building per hour.

Peri3D tells DCD the total print time was actually around 172 hours over 30 days during this period. Blatz jokes the print took longer due to the amount of press interest the project generated: "Too many news magazines wanted to see how we print! But it was as a pleasure to show all the TV stations; we'd print for an hour or two a day."

At the time of writing, construction on the shell had been completed and Heidelberg iT will now build out the white space and electrical room. The data center will offer 500kW and have capacity for around 100 racks. The fit-out is expected to be complete towards the end of the summer.



The IT firm has a long-term lease on the site, which Blatz says has a planned lifespan of 20 years, though he hopes it could last “much longer” than that. At the time of writing, the company didn’t have an anchor tenant for the facility, but discussions were ongoing.

Heidelberg iT plans to install solar panels on the roof of the facility – around 80kW worth – and will be procuring energy from hydroelectric sources to offset its remaining energy use. Blatz says his company generally has more solar panels at its data centers – such as whole wall deployments – but that it was limited by local planning regulators for this

project. Waste heat for the facility is due to be sent to buildings in the surrounding area once the district heating network is set up.

A learning experience for everyone

Jan van der Velden-Volkmann of SSV Architekten notes his firm hadn’t had any prior 3D printing experience before the Wavehouse, but Mense-Korte had been involved in some previous projects, which was useful for permitting.

“With this technology, we had the chance to

try some more organic forms that are normally not possible,” he says. “That was the aim for us, to test how far can we go with it.”

“We had the chance to make this organic form with the waves on the facade; it’s a very heavy block, but the form seems to be light.”

He notes that because this was a novel design using a novel technology, close collaboration with Peri 3D, Mense-Korte, and Heidelberg Materials during the design process was key.

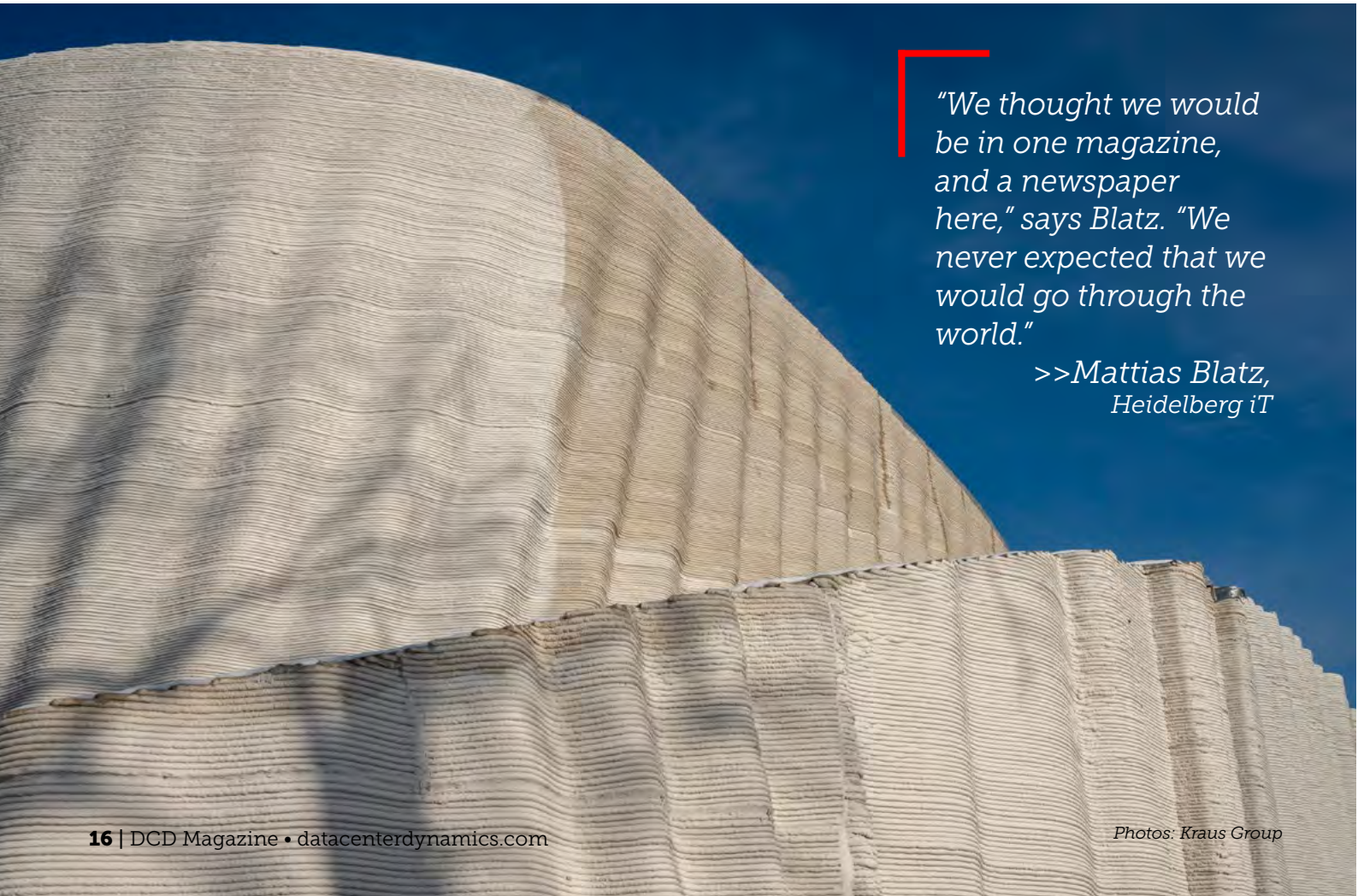
“The important thing is you have to answer various questions very early in the process. Normally we try, and then have a back-and-forth to change things. With this it is very challenging; you have to do a lot in the early stages because it’s hard to change things afterwards.”

Certification for much of the facility has been a challenge. Most of the materials and designs for these kinds of buildings haven’t been proven out. And so the companies had to over-engineer in some places; – the internal hall features multiple protruding support columns that Blatz said the regulators required. The testing required to approve the column-less design would likely have taken several years.

While there were design limitations, van der

“We thought we would be in one magazine, and a newspaper here,” says Blatz. “We never expected that we would go through the world.”

*>>Mattias Blatz,
Heidelberg iT*



Velden-Volkmann says working with the local government overall was “very constructive” during the project. Part of this was because the volume and planned use for the site was largely the same as the previously-planned cuboid data center, but also because officials were keen for the plaudits.

“They were very interested in getting such a building in Heidelberg,” adds van der Velden-Volkmann. “The counselor himself said ‘we want to have this,’ so there was pressure enough [to get it through].”

Heidelberg iT will be targeting certification from TÜVs – a German technical testing & certification organization – to show the facility is secure and resilient, alongside Blue Angel green certification to show environmental chops.

The foibles of printing a data center

The handover ceremony in late January 2024 was attended by around 100 people, including members of the construction, materials science, and national press – a far cry from some data center events with a handful of the same faces over and over. DCD was disappointed that the oversized key handed over during the ceremony was made of styrofoam and not 3D printed; an opportunity missed. But press of all stripes have been regular visitors to the site.

“We thought we would be in one magazine, and a newspaper here,” says Blatz. “We never expected that we would go through the world; that the Chinese might want a picture from us. A Korean team came and wanted to see it. We have had Americans come here. Indian TV as well. We are proud that we have that interest.”

The walls of the Wavehouse comprise two lines of built-up concrete a few inches apart – also known as a double-skin wall – and linked by steel rods. In residential developments the space between the two rows would be filled with insulation; here, it has been filled with more concrete as there’s no need for more heat to be kept in and it improves wall strength.

Due to limitations around the gantry frame that holds the printer, the Wavehouse was printed in three segments. Connecting pieces were then printed and put into place to seal the facility – the seams between segments are noticeable on the inside once pointed out.

During the design phase, Blatz says some of the software the architecture firms were using were struggling to work with some of the shapes and angles being proposed, and required additional programming to accommodate these calculations into their 3D models.

“They figured that out,” he says. “They also spent a lot of time and money in the way to find out how it can work out and how they can calculate if these walls are thick enough.”



Photo: Kraus Group

3D-printed buildings use far less metal than traditional concrete builds and often don't require large amounts of rebar frames. Blatz said this meant there is far less static for the Wavehouse compared to a traditional data center.

Blatz has also made a deliberate design choice around the raised floor learned from a previous facility. A one-meter recess has been cut into the slab for raised floor, but the company has opted not to put any thermal insulation in there (it has been waterproofed). As a result, the colder ambient temperature of the ground reportedly provides 20kW of free cooling to the facility. Only cutting the recess where racks will actually go, rather than right up to the wall, also improves air circulation.

A painting robot from DAW Deutsche Amphibolin-Werke by Robert Murjahn was used for painting the interior – painted Heidelberg iT's shade of green to match Blatz's neon trainers. The outside, however, will remain concrete grey – though living walls are set to be planted at regular intervals along the length of the facility. While the facility won't have perimeter fences – a limitation of its location – the site will feature multiple gates at its entry points. The thick walls also add further security.

On the flip side, Blatz said there were challenges around ensuring the roof of the facility (flat and not 3D printed) was water-tight against the ripples of the 3D printed layers of the building. Likewise, fitting the fireproof doors was also a challenge.

Blatz notes that printing is made more difficult in heavy rain, and that it was best not

to leave too long between printing sessions on segments.

"We learned a lot," Blatz says. "[Some of it was] a lot more work than we thought."

While he didn't provide a number, Mattias said the Wavehouse has cost around double the price of the other, traditionally-built, data center the company is opening – partly because the original plans for a data center on this site were already complete and had to be reworked.

He says, however, they would likely save more money on future builds based on what they've learned during the project.

A one-off data center?

The Wavehouse may remain the only 3D-printed data center for the foreseeable future. While there was one on the cards in the US, the developer opted instead for a warehouse retrofit. Technology constraints still limit the size of any printed building, and regulators are still wary of largely untested designs and materials.

In the US, EdgeCloudLink (ECL) was originally to offer hydrogen-powered off-grid modular data centers, built in 1MW units, using [3D printing construction](#).

ECL previously said it was planning to partner with a local 3D printing construction firm for its initial build, and later buy its own 3D printers. For its first facility, however, the company opted to retrofit an existing warehouse instead of printing a facility, and is looking to focus more on retrofitting existing buildings for now, potentially returning to the 3D printing concept in future.

"I personally believe that the conversion of buildings is probably more powerful than doing the 3D printing," company founder Yuval Bachar recently told DCD (see page 25). "There's so much empty space and it's going to waste right now."

Given the current construction limits around 3D printing in terms of physical size, it seems that printing a hyperscale-sized or high-ceiling multi-story data center is out of reach for now. For a developer willing to test the regulatory limits with local planning boards, an Edge developer could well find utility in this new way of building. But optimism remains.

"The interesting question is what comes afterwards," says SSV's van der Velden-Volkman. "The technology is very new, so you have to go step-by-step to get more knowledge and then to see what comes out next."

While the Wavehouse is the first 3D printed data center, Peri3D is keen that it won't be the last.

"I think that IT infrastructure and respective buildings can absolutely be something that 3D printing could be used for. I think we've done it now," says Lukas Bischofberger, corporate & marketing communications specialist at Peri 3D.

"We don't necessarily think that in 20 or 30 years, every building is coming out of a 3D printer. For some building types, other ways of construction can make more sense. But we still think there's a lot of projects that can really benefit from this."

Both Blatz and Kraus remain keen on the concept. Blatz says he would print another data center in future, while Kraus is already planning his company's next 3D build; a large two-story industrial building twice as tall and three times larger in area than the Wavehouse.

"I will print again," Kraus says. ■



3D PRINTING:

A REVOLUTION IN CONSTRUCTION?

3D printing concrete is said to have originated out of the Rensselaer Polytechnic Institute (RPI) in New York when Joseph Pegna applied additive manufacturing to concrete in 1997.

Then Behrokh Khoshnevis, a professor at the University of Southern California, 3D printed an entire wall in 2004. The first 3D-printed house was built in 2014 in Amsterdam, the Netherlands.

Market reports suggest the global market size of the 3D printing construction industry was \$3.5 billion in 2022, and is estimated to grow to around \$523.3 billion by 2030. To date, however, most printing construction has been in the residential market.

In 2018 Icon and New Story printed a 350 sq ft (32.5 sqm) home in Austin, Texas, in around 48 hours. The company said it was able to build a home for just \$10,000, sparking a wave of interest from developers – though that number was only for the concrete walls rather than the full build-out.

"That really started the fire underneath this whole thing," says Ziyu Xu, founder of California 3D printer provider RIC Technology (previously Advanced Intelligent Construction Technology).

"And then in the last couple of years, buildings have been permitted in North America, and that's a big step. People are starting to buy in and believe in 3D printing."

"This has grown interest from novelists and experts and universities to bold developers. We're past the novelty phase; construction companies are now come in and wanting to try this and figure out how to make this profitable and efficient."

Advocates tout speed and cost as major benefits, while also reducing the number of workers required on-site. Many 3D builds only need two or three people to manage a printer. Xu says 3D printing also ensures constructed buildings are closer to the original designs created by the project architects and site owners.

"I've always looked at things from an architect's standpoint, and each takes a huge amount of communication, effort, and compromise," he says. "And the end result is never what the architect really wants, which is what the owner or developer wants."

"With 3D printing, we are actually taking out a lot of the communication costs and inefficiency out. And the end result will be much more in line with the design that the developers or the owners want."

Most 3D printing construction firms are either startups or subsidiaries of larger construction firms interested in exploring the technology. Founded in 1969, German scaffolding firm Peri began looking at 3D printing in construction and what it could mean for the company. It invested in Cobod in 2018, taking a minority stake in the Danish firm, and officially launched Peri3D in 2022. Today Peri3D sells 3D printers and services, and prints its own buildings.

After being involved in Germany's first 3D printed home – a two-story family house in Bekum – in 2020, Peri has been involved in a multi-story, five-family apartment building in Wallenhausen and a gateway sculpture in Dornstadt, Germany; a 125 sqm office building in Hausleiten, Austria; and a single-family home in Tempe, Arizona.

Today, the number of 3D-printed buildings probably numbers in the low hundreds. Printers have been used on houses in Belgium, South Africa, New Zealand, Russia, France, and more, as well as bridges and wind turbine prototypes. Icon has been used to print a 'vehicle hide structure' – essentially a garage – at Camp Pendleton in California, and says it has printed more than 100 homes. CyBe created a 3D-printed R&D lab for drone research in the UAE.

RIC's Xu notes Accessory Dwelling Units (ADUs – small secondary buildings on residential plots and also known as backyard cottages, granny flats, or laneway houses) are a potential area of growth for 3D printing.

He says the lower cost of entry into the residential market and printer limitations means that's where most of the interest around 3D printing construction is currently focused. The Wavehouse in Germany is the first time a developer has been willing to take on the cost and risk of a commercial industrial building.

On the industrial side in future, Xu notes there's promise around warehouses and cold storage facilities, and he has

been approached by a number of prison development organizations due to the potential wall strength of 3D printed projects.

A new skill to address a labor shortage

The Wavehouse reportedly only required two construction workers while the printer was in operation. Other 3D printing construction projects report similarly small numbers of staff on-site.

This reduction in labor is being heralded as some as a way to help address skills shortages in the construction industry. In the UK alone, a report from last year suggested 83 percent of businesses within the construction industry were feeling the strain from a lack of skilled workers.

"The construction industry severely suffers from labor shortages," says Peri3D's Bischofberger. "It's really hard to get young people into construction and it's really hard to get good, highly qualified people in construction."

"Getting automation into the mix and construction robotics is in our opinion, a good way to not only sort of address the shortage we have but also make the job way more attractive."

According to Peri3D, it takes around six weeks to learn from scratch all the skills needed to manage a 3D-printed build.

"I've been working on 3D printed construction sites a few weeks back, and it's much more laid back and relaxed day-to-day work, simply because you're mainly overseeing the printing process," says Bischofberger. "You're not, for example, doing heavy work putting stone and stone together."

RIC has been working with its US partner Alquist3D to develop a workforce development program with Aims Community College in Greeley, Colorado.

RIC's Xu suggests that while a lot of young people might want to avoid the dirty work of construction, they do want to play with the big robot.

"Training has come a long way. Three years ago, [these printers] were quite difficult to operate and took about a year of training," he says. "But we have been

doing a lot of software development and user interface development, and now we only do a two-week training course for our clients to be able to use them."

Limitations and logistics

While 3D printing buildings can be quicker and easier than a traditional build, it isn't quite as simple as pressing a button and then sitting back until a building is finished.

Most printing technologies require extensive set-up – which in some instances can take days. These frames on which the printing nozzles sit also create the building limitations developers must print within.

Cobod – privately owned by General Electric, with CEMEX, Holcim, and Peri as key shareholders – manufactures the BOD2 construction printers. It is one of the major providers of 3D printing construction machines.

Like rival companies Black Buffalo, 3D Wasp, or Icon, Cobod's technology relies on a gantry system, requiring a scaffolding system, bolted to the slab or on concrete blocks, onto which the printing nozzle is attached.

Peri3D – which uses Cobod printers – says its maximum building dimensions are 13.5 m (44.3 ft) wide, 9 m (29.5 ft) high and as long as desired. Peri says for a machine printing area of 13.5 m (44.3 ft) width, 17 m (55.8 ft) length and 8 m (26.3 ft) height, eight hours would be required to assemble the system. The system can reportedly print at speeds of 1 meter per second (3.3 ft); layer heights can be between 1 and 4 cm (0.4 - 1.5 inch) and layer widths between 3 and 10 cm (1.2 - 3.9 inch).

Different vendors have different systems. Icon's uses a smaller gantry system on wheels – known as Vulcan – for its residential builds.

Other providers – such as Vertico, CyBe or RIC – rely on robotic arm systems. These can be fixed in one place, on a moving track, or attached to gantry systems. Some have their own locomotion, such as RIC or Constructions-3D.

Xu says the gantry systems requiring large frames are "cumbersome" and require cranes, bobcats, and a lot of time and effort to set up and tear down – sometimes up to three days.

"By that time you've already spent \$4,000-\$5,000 on labor and the equipment rentals," he says. "And then you're building a giant metal house before building the actual house and then taking that metal house down after you finish printing, which is not what we really want to achieve by doing this."

RIC originally offered the RIC 0, a static printer relying on a robot arm from Kuka, before moving onto the RIC-1, a printer that could move along fixed rails. It has moved onto the RIC M1, which has seen the company's printing arm attached to a 4x4 wheel system. The RIC-M1 Pro is a combo of wheels with a fixed rail attached. It also offers the RIC-2, which is attached to a scissor lift attached on rails for taller builds. The company says its mobile platforms speed deployment time from days to hours.

Materials

Where 3D printing handheld objects are often done with heated plastic, printed buildings are constructed using cement.

"At the end of the day, it's still your normal day-to-day cement. Is still it's basically Portland cement, plaster, aggregates," says RIC's Xu. "The additives they use are the same as the normal construction industry; the normal accelerators, the normal plasticizers."

While the materials are the same, the proportions of each component are different. The mix needs to be liquid enough to extrude cleanly and quickly, yet viscous enough to stay where it is printed and attach itself to the previous layer. A number of companies are beginning to develop their own mixes specially tailored for 3D printing.

Heidelberg Materials – which supplied the cement for the Wavehouse – says its 3D printing cement 100% recyclable as a mineral building material. It says its printed concrete also contains a binder with around 55 percent CO2 reduction compared to pure Portland cement mineral building material.

"You really have to get a lot of properties right so you don't have cracking," says Peri3D's Bischofberger. "The layers that are extruded on the one hand need to be soft enough or liquid enough that they can be extruded without clogging the pipes. On the other hand, it needs to get stable fast enough so you can put the next layer on top."

"There are height limitations, simply because you can only put the frame to a certain height before the stability gets to a point where you can't ensure the perfect layer quality anymore," says Bischofberger. "That's why right now we limit ourselves to printing heights of around 9 or 9.5 meters. But I know that the company is already developing different ideas and on how to get larger."

Certification and permissions are still a challenge for 3D-printed builds. RIC's Xu notes that currently, many printed

buildings are still currently following international building code standards for CMU (i.e. masonry) builds. But a new building code standard for 3D printing – known as [AC509](#) – is being developed by the International Code Council.

Planning regulators have been erring on the side of caution around the designs and permissioning of these projects, but once an area has its first 3D-printed building go up successfully, regulators are more at ease with the concept and more likely (and quicker) to accept proposals.

"It takes a little bit longer time for the first building in that specific city to go through because the city officials and their engineers want to do their research before they give a permit to a building," says Xu. "Once they're through that learning phase, they have no problem of proceeding with it because they realize it's just any old concrete building."

For now, many builds are over-engineered to ensure they meet current building requirements that weren't designed for 3D prints, which companies say means the technology isn't being used as efficiently or innovatively as it could be.

"There certainly is a lot of potential where developments can go. But as of right now, because we have to start somewhere we need to have a look at the stability, we need to look at how the buildings actually hold up in the long term."

Xu says that currently, the cost of materials is also still a barrier – he suggests printing costs around \$650 per yard at the moment, but the industry will really take off if suppliers hit long terms of bringing that down 50 percent.

"Right now we are one-to-one because the materials are quite expensive," he says. "Once the material costs cut by half, then we're going to actually have about a 30 percent saving compared to a conventional construction."

While 3D-printed concrete is largely similar to traditional concrete, there is still only a small number of 3D-printed buildings more than a few years old. While the industry is confident 3D printed structures will last as long as their traditionally-built cousins, this has yet to be fully proven out on a long-term basis.

"We don't have any long-term studies," says Bischofberger. "So that's a little bit of an area where we just don't have the information on the lifespan yet. But as of right now, nothing points in the direction that the lifespan would be in any way negatively impacted by the fact that 3D printing technology was used in construction." ■

This Emirati operator who doesn't use diesel



Peter Judge
Executive Editor

But Hassan Alnaqbi's AI-backed business is working for oil exploitation

Dubai-based Khazna is the leading provider of wholesale colocation in the United Arab Emirates, putting up buildings for cloud providers and large enterprises. It is growing fast, with 16 data centers between Abu Dhabi and Dubai, and has a strong link to AI, through its owner.

When we spoke to the CEO, Hassan Alnaqbi, Khazna had recently opened two facilities, with another was due to open before this article is published.

Khazna is not much more than a decade old, and started out as a state-directed initiative. It was founded in 2012 by Mubadala Investment Company, a state-owned holding company or National Wealth Fund, which also owned AMD's chipmaking spinoff GlobalFoundries, until it went public via an IPO in 2021.

"The reason Mubadala invested in this particular sector of the business was the lack of any infrastructure availability within the UAE and the entire region," explains Alnaqbi. "The focus from Mubadala back then was to actually create an infrastructure that can support the country into the digital transformation, and also provide a robust solution for specific and niche markets."

Alnaqbi was with Khazna more or less from the start, coming across from du, one

of the UAE's two main telcos (formerly the Emirates Integrated Telecommunications Company) for the period from 2012 to 2014 when the business was being put together.

The Khazna history

"I started the operation of the company," he says. "I launched the business and launched the operation. We had two sites: Abu Dhabi 1 (AUH1) in Masdar City, and Dubai 1 (DXB1) in Maydan. "I remember very well that the growth was very, very slow."

Back then, there was no hyperscale model to speak of, and Mubadala wanted to set up a data center company to support the government and telcos. The company started with 13 customers made up of government entities and telco operators.

The government business was active because the public sector was "very aggressive in the acceleration of their smart government solution, digital transformation, and sectors like cybersecurity," Alnaqbi says.

And that steered Khazna towards a wholesale model. "Our focus was mainly to tackle the requirements of the government, which has critical data it wants to store, and not necessarily to go into the colo mode," he explains. "So Khazna is a wholesale operator that offers the flexibility of enterprises



and government owning their own space securely, within their own remit and not having a shared space, like other colo providers."

Colo was a difficult sale back then, but the telcos were a gateway: "A lot of enterprises had not really adopted any cloud solutions, and were very protective in terms of securing their IT environment," Alnaqbi says. "The telco operators were engaging with small to medium businesses, and large enterprises on the retail side, so they built an ecosystem within Khazna to go and offer shared or colo solutions to the enterprise market."

That was phase one, but things have changed since then - first to the cloud, and then to AI.

"In 2017, we started seeing cloud players coming into the region," Alnaqbi recalls. "The enterprise mindset has changed towards the scalability that the cloud provides and the reduced cost on them from the operational perspective."

That was a "complete mindset shift," he says: "We have shifted our design principles to specifically meet the demands of the cloud players. In 2019, we started landing some of the global players who are leading the cloud domain, on the enterprise side, on the ERP side, and also on the full stack of cloud services."

Then in 2020, Mubadala and du

(which had a minority share) sold Khazna to the Emirati AI technology conglomerate G42.

"G42 took over Khazna prudently, because they are one of the biggest users of data centers," explains Alnaqbi. "That fits with their agenda of leading in AI development."

G42's AI developments include multiple large language models (LLMs), including the first Arabic LLM with more than 30 billion parameters, which is now being adopted by *Microsoft Azure*. It has also funded Cerebras, the developer of giant wafer-scale AI chips

The company's involvement has changed Khazna's focus from "designing specifically for the cloud layer, to looking at how we can embrace the AI infrastructure requirement within the data center space," says Alnaqbi.

In 2021, Khazna expanded through a deal between G42 and the state-owned incumbent telco Etisalat (previously the Emirates Telecommunications Group), now branded as e&.

"Etisalat was our main competition in the UAE, in terms of providing data center solutions, but a lot of telco operators, globally, divested data centers from their portfolio because it's not a core business to them. And that's exactly what happened here," says Alnaqbi.

"We created a joint-venture partnership. G42 is the majority with 60 percent and e& is the minority with 40

percent equity," he says. "We consolidated all their data center space under Khazna and, with that, we became the largest data center operator not only in the UAE but in the entire MENA (Middle East and North Africa) region from a capacity perspective."

That's a big deal: "What we have today is basically more than 300MW that is contracted in terms of capacity contracts," Alnaqbi says. "We have about 175MW active and operational and handed over to the clients, the rest is being handed over as we speak within the next 18 months."

Beyond that, there is fast organic growth: "We have secured about another 250MW which is basically expanding on what we have today within our own ecosystem. That's coming out within the next five years."

And then, there's the projected AI boom. "We have launched our first AI-ready data center, AUH6, which is the first of its kind within the region."

Arabic AI

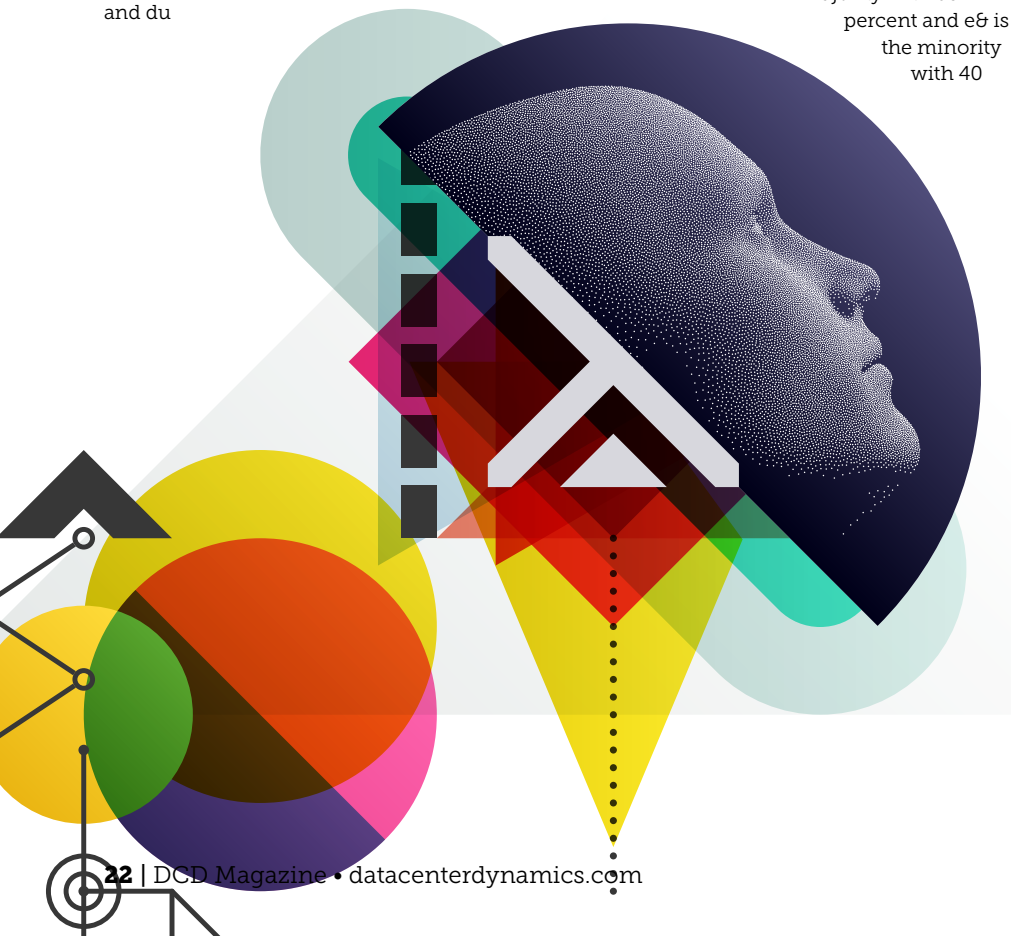
"G42 has taken a different approach to AI," he says. "They realize that to take a lead on this ecosystem, you have to partner with global players. They have announced a partnership with Microsoft, and OpenAI. They are co-developing LLMs that can sit within the Azure platform and can be beneficial to OpenAI."

G42's approach is to focus on "the market we are in," says Alnaqbi. "G42 realized that sitting within an Arabic country, and being part of the country's agenda of developing and taking a leading position on AI, that means focusing on the Arabic languages."

The world has more than 300 million Arabic speakers, and that is a significant opportunity, he says: "There is a huge population that speaks Arabic that may actually require a certain development of LLMs that help certain sectors like banking, financial institutions, investment sectors, and oil and gas, where they can use the Arabic languages to have call centers completely digitalized and AI-driven, with automatic response and voice recognition."

The challenge is that training AI systems needs source data, and there is less of this available in Arabic than in English. "I'm not an AI expert, but I understand it's all about the quality of data," Alnaqbi says. "Apart from having data from the governments, they went to all the Arabic libraries, books, and encyclopedias in Arabic, and scanned books written in Arabic day-by-day and hour-by-hour."

Processing and training on that data allowed G42 to create large datasets and build Jais, an Arabic LLM with over 30 billion parameters: "That's been recognized by Microsoft and I was in Abu Dhabi when Satya [Nadella, Microsoft CEO] mentioned



that we are trying to incorporate the Jais model within the Azure Cloud," says Alnaqbi.

At the same time, Khazna is building infrastructure tailored to the needs of G42 and LLMs like Jais, Alnaqbi says. "We understand that to build a data center for the cloud is a completely different ballgame, than building a data center for AI," he explains.

Beyond the early training facilities, Khazna saw a need for inference, where AI is applied, the CEO says: "The AUH6 development was designed for AI infrastructure from the inference perspective. We're not going to be using this to train models."

AUH6 will deliver a mix of cloud and inference, which Alnaqbi describes as "challenging." The inference system must be able to access data instantly, he says, "but the density of the cabinet is [around 50-60kW], not as high as you need for training. For that, you can need up to 100-150kW per cabinet. That's where the challenge starts!"

And it's a challenge Khazna will be taking on.

G42 has announced the Condor Galaxy cloud AI platform consisting of nine interconnected four exaFLOP AI supercomputers giving a total planned capacity of 36 exaFLOPs. It has delivered the first 54 million core node (CG-1), which includes 64 Cerebras CS-2 systems, and is working to build two more in the US, CG-2 and CG-3.

In the UAE, Alnaqbi reveals: "We are on the verge of finalizing the design to go into the market for building one of our biggest AI training clusters. That will come probably in the next 18 to 20 months. It will be a high-density design built to cater for training AI models and it's going to be one of our biggest clusters."

Alnaqbi has no further details, and *DCD* has asked if the facility is going to be one of G42's Condor Galaxy or a new design. Alnaqbi says the facility has to have high densities. "The issue is how many servers you can put within one cabinet, so the model is trained from one server to another without having to cross 90 meters of distance."

Given those constraints, he says the training facility will be "four or five times what we are doing today in one cluster."

Balancing AI and cloud

Is there a danger that Khazna might shift too quickly to AI, and miss opportunities in cloud and enterprise? "You're absolutely right. I think this is why, within the management and leadership, we wanted to segregate the two tracks so then we don't get into the excitement of the AI track and then forget our own organic business," Alnaqbi says.

The Khazna CEO says this gives a "clear focus on both sides," explaining: "Business as usual for us is taking what we have into the next stage of growth. We focus on deploying cloud and hyperscale data centers for the top players, and we will continue working with them to deploy their cloud - because that's growth by itself. There is huge growth coming up, and we know the region needs a lot of [cloud]."

Meanwhile, "within the company, I'm trying basically to structure AI completely separate from the other businesses so it doesn't get into the excitement and people lose track on the other side.

"It's a challenging thing to do. But luckily, with the leadership, we created a completely different structure for the innovation and deployment of AI. That includes the design team, the project management team delivery, and so on. That's completely segregated."

Within AI, he predicts that energy demands will push alternatives to Nvidia, shifting "from reliance on one particular vendor which has a huge pipeline to other providers."

Obviously, Cerebras is one of those: "Cerebras has an AI infrastructure that allows you to train models faster. They might not yet have the right software, like Nvidia, but they have a really super fast chip that allows a lot of the models to be trained faster."

Commercializing new chips is a challenge, he says: "But I see now that a lot of parties are trying to find different solutions, like Qualcomm, AMD, and others."

Sustainability

Energy demands lead us to talk of sustainability which is, he says, "The other element of AUH6 that we wanted. This is why we selected this location within Masdar City which has been renowned as one of the most sustainable developments not only within UAE but the entire region."

The city is yet another state-funded initiative, being built by Masdar, another Mubadala subsidiary, built with at least \$22 billion of government money. It was announced in 2008 as "the world's most sustainable eco-city" and projected to be built in eight years.

While the project was endorsed by the Worldwide Fund for Nature (WWF) and the US government, others have been skeptical. Professor Geoffrey Heal, of Columbia Business School has called it "a gimmick, a way of attracting publicity and attention."

Eight years on from the projected 2016 completion date, Masdar covers around one-sixth of the original area projected. It has about 10MW of solar power from the Masdar Solar Plant, along with some rooftop solar.

The city uses the grid for the rest of its needs and, as one would expect from a petro-state, that grid is heavily dependent (about 92 percent) on oil and gas, but the country is rapidly developing solar and nuclear power.

The new AUH6, one of several within Masdar, has its own solar plant, built by Emerge (a joint-venture between Masdar Energy and France's EDF), which provides 7MW peak capacity AUH6 is projected to grow to around 32MW.

Building within Masdar means Khazna had to meet environmental standards, and AUH6 meets LEED gold, as well as Pearl 4 under the UAE's own Estidama sustainable building rating system.

The climate in the UAE is not like the rest of the world, he says: "I visited Facebook's data center in Lulea in Sweden - and the challenge there was completely different," Alnaqbi says.

"They cannot pass the freezing air actually coming from outside into the data hall. They have to mix it with the hot air to bring the temperature up and then pass it on to the data center.

"In our case, we don't have the luxury of having cold air so we cannot design our facility with that mindset. We have the opposite issue, which is our hot climate. For most of the year, it is a very harsh and hot environment with lots of sandstone. You need to design your cooling solution to be reliant on what techniques and efficiency you can do within this harsh environment."

With that in mind, AUH6 is designed to use less water: "We are using a new technology which is evaporative cooling, and that doesn't really require a lot of water," Alnaqbi claims.

The technology is adiabatic cooling, which allows some cooling in the winter, and uses a water sprinkler to cool the air, a technique he notes is used in some Arabic houses.

There's some liquid cooling in the racks he says, but not the direct-to-chip cooling used in training architectures, at least not provided by Khazna. "We allow the customer, if they want, to have chip-based cooling, but this will only allow them to go to 150kW in certain cabinets, not for the entire deployment."

The diesel generators on the site use biofuel, he says. "Back in October last year, we announced we are shifting completely away from diesel to biofuel."

Driven by government and clients

That decision is aligned with the government's vision, he says. The UAE has promised that the country's operations will achieve net-zero by 2050, which

is undermined by the nation's plans to continue exploiting oil far beyond that date. Although this will be burnt by others, its total impact on the climate will be far from net-zero.

"The country has invested heavily in nuclear plant," says Alnaqbi. "It may be the only country within the region that has a nuclear plant. We have four reactors that provide about 6GW of clean power, and we have also heavily invested in solar plant. Abu Dhabi and Dubai have invested in more than 10GW of solar plant over the next few years, which will contribute a lot to the grid."

He concedes: "Yes, we are an oil-rich country, but I think we as a company, need to follow the direction of the country."

The company's clients also give it a mandate to focus on the sustainability of its infrastructure, he says: "For a lot of hyperscalers and other customers that we deal with, sustainability becomes one of the criteria for hosting their network," Alnaqbi says.

Hyperscalers have global targets for sustainability, and the conditions in the UAE need explaining, he says. "We work as a partner and we openly discuss our roadmap and their roadmap to find where's the gap."

Alnaqbi regularly meets with the big players' chief sustainability officers "to look at what they're doing in different regions, how they are collaborating with other operators." He says: "We take that seriously and set up sessions with workshops with them. I personally attend a lot of workshops with some of the hyperscaler leadership."

Those sessions determine what is a practical sustainability target: "We come up with a common understanding. This region has its own challenges versus other regions, but they have criteria to follow and they need to tick a box."

After this year, Alnaqbi might not have to go to so many of those workshops. The company doesn't currently have its own sustainability officer, but this year there is a plan "to have a centralized function that will drive sustainability across all functions, not only from DCs, the design and engineering side but also across the entire company."

Whoever fills this role, Alnaqbi is looking forward to introducing them to sustainability, UAE style: "It's going to be really funny to see how they will come to the realization about what they have to do within this environment," he says. "It's completely different here."

Even without a central sustainability leader, Alnaqbi says Khazna's sustainability efforts go beyond emissions: "We make sure our data centers are sustainable not just from the point of view of carbon emissions, but

from the full ecosystem of sustainability, even to the point that we are probably the only operator in the region that has a zero waste policy."

The zero-waste policy was announced in 2023, and certification is on the way, he says.

"Anything you get rid of within the data centers has to be split into materials, and goes to a factory that turns this waste into something - either power or usable material," he says. "You need to make sure that it's zero waste to landfill."

Beyond the UAE

Khazna is also expanding beyond its home territory: "We came to the realization that after we have penetration of data center availability in the UAE, the growth will be very minimal. So we started a few years back looking at markets that are underserved within the region and trying to enter those markets with our knowledge and expertise and the blueprint that we created."

It's a plan to be "the UAE champion," he says, "and transform from being the largest local operator to the largest regional operator for data centers."

Khazna has a project to build a 25MW facility which will be *Egypt's* first hyperscale data center, working with local partner, Benya Group of Cairo.

What's next? "We'd like to start doing something in countries like Saudi Arabia and Kuwait and Qatar and others. We're looking at different regions. Turkey is among those markets."

The region has "huge opportunity for growth in the enterprise market, and a lot of potential for cloud adoption," he says.

And once again government backing comes into play, and some of the business may emerge from 'G2G', or intergovernmental, deals.

"The UAE has a very friendly relationship with the entire region, they guide us to look at a particular region, and that's something we're actually doing. The government is actually looking at technology with a different lens."

UAE President, Mohammed bin Zayed Al Nahyan, who is also the ruler of Abu Dhabi, has set up an Executive Council for technology in general, which includes members of the board of Khazna, as well as of G42.

"So you can see there's a lot of close collaboration - but also there is a vision that the UAE is taking a leading position when it comes to AI, that will also require extensive investment in infrastructure, energy, data centers, and AI infrastructures, but also a lot of investment in human capital."

When UAE ministers sign MOUs with other countries, he says, they have "data centers as one way of collaboration. Our name is always put there to explore collaboration with these governments to help them tackle the lack of data center infrastructure in their market. They present us as the expert in the domain."

Part of that is driven by data sovereignty: "I think the entire region now has rules, but UAE and Bahrain, are actually ahead of the others. They have already created data sovereignty and data localization laws."

"That pushes the cloud players to deploy cloud presence within this region, not offer it from other regions."

When players like Microsoft and Google create "sovereign cloud" products it is up to local providers to deliver facilities to support data residency, as well as security, he says: "We can ensure all the way to the CPU that your cloud is actually completely protected and completely encrypted within the public cloud, so a lot of work has been done on that domain."

He says announcements from AWS and others show that sovereign cloud is emerging in Saudi and Qatar, for example, and other players like China's Tencent ("basically the cloud of the East") is also playing there.

"We've seen a lot of momentum going, and the entire region has invested heavily in connectivity, which is a key component for the success of cloud players within the region. Data centers are the foundation, but on top of that they need low latency."

Above all, energy

Finally, he comes back to energy again, and the unknown demands of AI: "Nobody can tell you what the expectation for AI is," Alnaqbi argues. "If you talk to people who are within that ecosystem, like G42, they expect about 500GW for the next 10 years of AI growth. And that's enormous."

He continues: "I think power becomes the number one constraint for everything that we do in the future. What we have witnessed so far, doesn't equate to more than two percent of the capability of what AI can do. If you imagine increasing this to 20 percent, that will dramatically change."

And he warns: "Energy will be a key factor of success. We have seen recently that there is a scarcity of energy, with some countries and some cities tapping out from the power perspective. AI cannot evolve if there is no data center, and data centers cannot be developed if there is no energy."

"I think that's what we will see AI becoming constrained in some regions and some markets in the future." ■

News from the hydrogen frontline



Peter Judge
Executive Editor

We speak to a CEO running a 1MW facility on fuel cells, with DC power distribution

"This is not a prototype," says Yuval Bachar. "It's a working site. It's production quality."

Back in January 2023, Bachar told us that his brand new company, EdgeCloudLink (ECL) would build modular data centers, running on their own hydrogen-powered microgrids, and make them using 3D printing construction.

One year on, he tells us that it is job done. He is doing the final stages of integration and commissioning a 1MW data center, running on fuel cells, at ECL's headquarters in Mountain View, California.

The facility, built with \$7 million of startup funding from Molex and Hyperwise, takes hydrogen by truck from a local source, uses a proprietary liquid cooling design, and has no diesel generators, using batteries and the utility grid for backup.

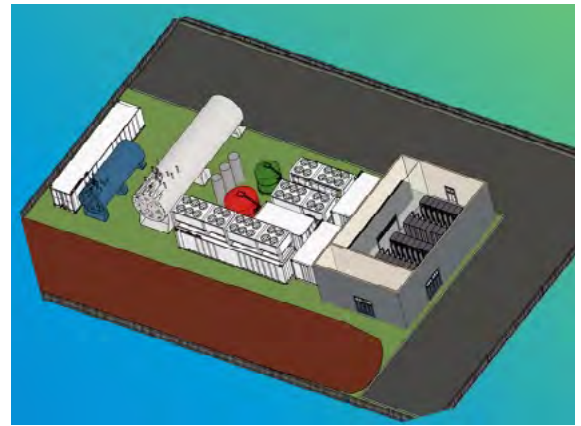
The only part missing so far is the 3D printing: ECL had a vacant warehouse on its plot, so this time around, there was no need to call on the construction printing contractor it had lined up. Bachar assures us that reusing a building is greener than printing a new one and, when 3D printing is needed, it will be the "easy" part of delivering a data center.

"ECL developed a fully sustainable data center which is running on a hydrogen microgrid, not connected to the grid," he tells us. "We can run our data center at 50 to 75kW per rack, and we operate at a PUE [power usage effectiveness] of 1.1, or better."

At \$7 million per MW, the proposal will undercut both cloud and colocation services, he says. With a delivery time of eight months, it can out-deliver them. And, depending on the source of hydrogen, it will also have remarkably low CO2 emissions.

The pitch is a "zero-emission, zero water use community-integrated data center." That means one that contributes to the community - potentially offering heat, spare generating capacity when needed, and clean water as a by-product of its fuel cells.





"We have zero emission and zero water use on our site," he says. "We don't use any water

from the community or energy from the community around us. On the contrary, we give back water to the community. That was our vision a year ago, and we have actually materialized it."

When *DCD* last spoke to Bachar, who has previously held senior data center positions at LinkedIn, Facebook, Cisco, and Microsoft, he was doing the final stage of integration and commissioning at the site, preparing to run 24/7 on hydrogen with a "very strong battery backup system."

The facility fits into a 1,000 sq ft (93 sqm) space in a building, occupying a total of 4,500 sq ft (418 sqm).

Designing for hydrogen

It has a completely redesigned power system, he says: "It's a highly flexible power distribution system, built on top of the hydrogen energy source. We eliminated the diesel generators, we eliminated the UPS systems, and we created a 4-source active-active solution."

In Mountain View, ECL gets its fuel by truck: "We bring a tanker every seven to eight days and it refills our onsite storage."

Bachar would prefer to have a hydrogen pipeline, and thinks these are on the way: "You have a significant amount of pipelines in the United States, and I think there's a tremendous effort right now in Europe to build new hydrogen pipeline," he says. "The best way to actually get the hydrogen is to connect to a pipeline."

ECL's vision of multiple modules on a site will have to wait for the economy provided by a pipeline: "You get a reasonable price and you can create almost unlimited size data centers," Bachar says.

Using tankers places a limit to the capacity of a site, Bachar explains. "How many times do you want to bring in a

refueling tank? If you don't want to do more than two refuels per day then the size of your data center is limited. Around 15 to 16MW it's starting to be a challenge. Unless these are hydrogen-powered trucks, it's actually hurting your sustainability claims."

ECL is working with delivery companies on getting hydrogen-powered trucks, he says "as well as actually trying to focus our energy into places that can get their hydrogen from a pipeline."

A lot of hydrogen reporting concentrates on the spectrum of hydrogen sources available. "Black" or "brown" hydrogen is made from fossil fuels, and is arguably worse than its fossil sources, as energy is lost in the process.

Bachar disagrees: "We will show publicly in the next few weeks that getting hydrogen from any source and any color today to run data centers is superior to any grid in the world from a sustainability perspective. Even if you take a gray hydrogen and deliver it by a truck you're still going to cut the emissions of the data center by 60 to 70 percent."

Green hydrogen is fully sustainable, but other sources of hydrogen will deliver a "dramatic" improvement on power

consumption and carbon footprint, Bachar says. "People don't usually don't look at what is the carbon footprint they get from the grid," he adds. "Even in California, with one of the greenest grids in the world, you can cut your carbon emission by 60 percent if you deliver any kind of hydrogen."

His advice is to optimize a data center to use hydrogen first: "Then you source the hydrogen from the most cost-effective and sustainable source. We're currently getting a blend of green, blue, and gray, and we are striving to get green only," he says.

"By late 2025 or early 2026, we'll be able to supply green hydrogen to every location that we need in the United States - and hopefully in Europe as well.

"Until that time we'll do a blend, targeting 'blue' hydrogen, which has had carbon capture on it when it was produced - so it has a relatively low carbon footprint. In the worst case scenarios, we will blend blue with gray with green."

Data center operators' experiments with hydrogen have included a 3MW generator powered by fuel cells, built by Microsoft in 2022, but most people appear to regard hydrogen as a possibility for the future.

Bachar admits the hydrogen economy is "evolving in different areas of the world at different paces," but says the US is in front thanks to the Biden administration's strategy to create hydrogen hubs across the country, with a goal of getting the price of hydrogen to \$1 per kilogram by 2031.

This is still a long way off, as pump prices in California were at \$36 per kg last year. But even so, Bachar claims it is possible to "create energy with hydrogen prices today at a competitive level to what you can get from the grid."

He says: "We can't compete very well where the electricity price is two cents or four cents per kilowatt hour, but from eight cents to 50 is the area where we operate."

Overall, other factors combine to give a hydrogen facility a lower total cost of ownership (TCO), he says: "We can beat

We have zero emission and zero water use on our site. We don't use any water from the community or energy from the community around us. On the contrary, we give back water to the community.

>>Yuval Bachar



about 60 to 70 percent of the grid prices in the US and in Europe, and within the next two years we'll be able to beat almost 100 percent of the grid prices."

Availability beats price

"That is," Bachar continues, "assuming that the grid is available."

In many places data center builders are facing long waiting lists for connection to the grid at all, but a hydrogen facility can be built more easily: "We can drop this in any place," Bachar says. "Ashburn, Virginia, does not have any power available until 2029. We are probably the only company who can actually build now in Ashburn Virginia."

Quick power makes a big difference: "Because we're not connecting to the grid, we can turn around our data centers in eight months so from the moment the customer comes and says 'I want this size of data center in this location.' When traditional methods can take five years to actually deliver a new site, that is very attractive in a high-demand environment."

ECL is aiming the facility, not at hyperscalers, but at end-user customers who own their IT equipment: "We created the solution be a colocation site or a private data center for the customer," Bachar says. And it can be installed "in any location they want because we are grid agnostic."

Design

The design has a slab floor, so it went into the warehouse space without any major changes.

It has a water-based cooling system: "We're using rear-door heat exchangers to cool the racks. That's how we achieve that level of density with no hot-aisle, cold-aisle isolation," Bachar explains.

ECL stopped short of immersion cooling because Bachar believes it's not ready, it is not approved for powerful AI processors, and is mostly being hyped at the moment: "Most techs in data centers will not accept it," he argues. "If you tell a data center tech

to get a server out of the tub and deliver it dripping into the maintenance area, that's going to be very difficult."

Bachar says the power conditioning system is an integral part of the data center: "It's not sitting in a separate mechanical or electrical room. It's actually right there."

The facility has hydrogen fuel cells, batteries, and cooling systems all put together in a block. "Each block has its own power conditioning system and power distribution system using busways."

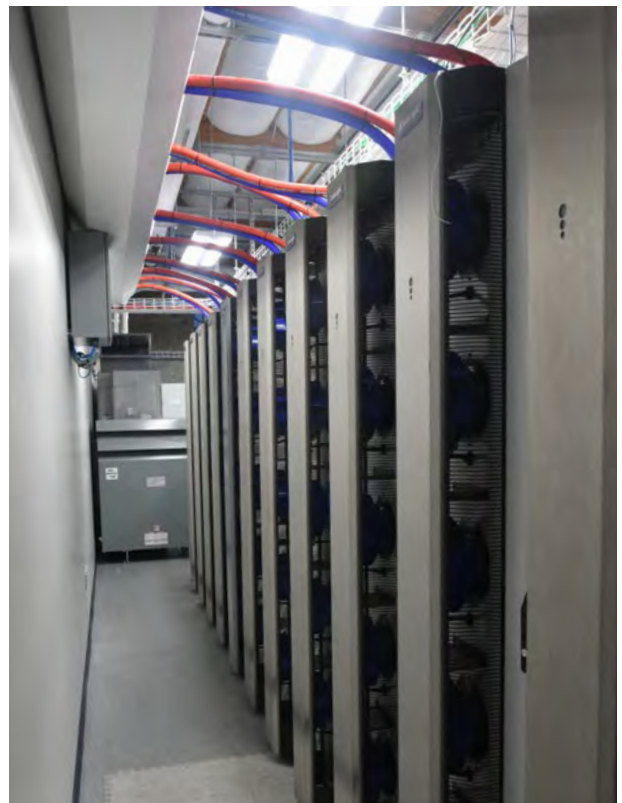
The power system, and the rear-door cooling means ECL can give each rack a different density from 5kW to 75kW. "It gives us the flexibility to change the power per rack and do a mix and match of different powers."

With no diesel generators, the facility has a Tesla Megapack battery that can keep the site online for eight hours. The facility is run by ECL's own management system.

The facility is not doing commercial

Around 15 to 16MW, refueling by truck is a challenge. Unless you're using hydrogen-powered trucks, if you do more than two refuels per day, it's actually hurting your sustainability claims

>>Yuval Bachar



work. It's a showcase for customers: "We'll do testing and bring customers to test their workloads on-site till they feel comfortable with our architecture and technology," Bachar says.

Large potential customers will be let loose on up to 10 racks, the CEO says.

ECL will also have its own separate high-density racks to show the design off: "We are working to put some AI machines there, in a super pod or a supercomputer footprint, to demonstrate the density capabilities - because most customers will not be able to bring in 50 to 75 kW on day one."

Direct DC power

Running high-density racks from a local power source has given ECL the opportunity to rethink how power is moved around the building.

Most data centers use AC bus bars to get electricity from the outside world to the racks, then converted it to DC for the electronics in the racks. Since the fuel cell natively produces DC power, Bachar sees no reason to convert it twice, losing power each time.

"We can actually drive any voltage of DC, or we can drive AC," he says. "For us, DC is better because we save stages of conversion."

Selling capacity in 1MW units, ECL plans to give the customer the choice in future builds: "We can drive a blend of them as well. You can have half DC, half AC if you want."

Setting up DC bus bars could be a challenge if power densities go too high, and the voltage remains comparatively low, he says. "At the amount of power that we use per rack, we're going to have to drive thousands of amps at low voltage into distribution systems," he says.

"You could be close to 10,000-15,000 amps. That requires pretty special copper distribution."

He is using higher voltages, which means a more reasonable level of current, and then stepping the voltage down at the racks. "We're trying to put the voltage high enough to be safe," he says.

"For 1MW, we push at 800V, 750V, or 700V. That's about 2000A, which is reasonable. To increase that five times or six times would be a nightmare."

Helping out

Running high-temperature fuel cells on site, as well as racks of high-powered servers, ECL has a stream of potentially useful by-products: "We are producing a tremendous amount of hot water that we can distribute. We do not need it," he says.

How much gets reused will depend where a future facility is located: "If we are close to the urban area, it makes sense to actually distribute it," Bachar says. "If we're far away from the urban area we can distribute into the building itself for heating."

The hydrogen fuel cells generate water, which is recycled into the data center at present. "The data center does not use water, because we actually have a factory [the fuel cells] which produces water."

But, because the cooling circuit is a closed loop with low loss, there will be a surplus of water. "We have four external loops around the data center. Some of them use some of the water that we actually produce but we definitely have an overproduction of water that we will give, in a cold state or a hot state to our neighbors."

In Mountain View, ECL is planning to offer the water for irrigation, Bachar says. "There's a lot of neighbors over here that require high levels of water, and we're just going to give it to them," he explains.

Ironically, what the fuel cells create is actually too pure for irrigation: "The funny thing is we have to dirty the water because it's too clean. It's de-ionized, distilled water. Other customers like hospitals or research labs might use distilled water. If the community asks us to give them the water, we'll give them the water."

Longer term, it could conceivably be a small revenue stream: "We produce hundreds of gallons per hour, and de-ionized water sells in the US for \$4 per gallon."

But Bachar would rather get the social capital from giving it away: "We see ourselves as a community integrated data center," he says. "We build only one-story buildings, we're trying to be as quiet as possible, we are not taking power from the community, and we are not taking water from the community."

Learning from the permit process

Building in Mountain View was a deliberate tactic, says Bachar, because the city is "very conservative" when it comes to planning permits.

He explains: "We said okay, how long is it going to take us to run for the first time, making all the mistakes and correcting them, and to actually get permits to build a hydrogen-based data center?"

He's pleased with the result: "From start to end, it took us about seven and a half months. And now we know exactly where we made mistakes, and how to accelerate the process. Now we have a very very extensive blueprint that we will take to every new location."

In future, he says, "the cycle to get the permits will shrink dramatically compared to what it took us on the first run."

In future projects, he predicts that the permit process will fold into other parts of the timescale: "Permitting is actually falling into our sourcing cycle. We don't do it in a series once we get a purchase order, we start the permits immediately and we start our sourcing process immediately."

Taking the two together, sourcing and permitting will take about six months, by which time ECL will have all the components it needs. "We will either produce them or purchase them and then we have about two months of integration onsite and and construction."

Re-using buildings

Future construction will be quick, because ECL is building small and modular. Based on ECL's experience in Mountain View, Bachar wants to reuse more vacant buildings such as former offices, which maybe don't have power or are not adequate for traditional data centers.

"A model that we're offering to our customers is conversion," he says. "We can take an existing structure and convert it to a high-end data center with no construction work required at all."

In Mountain View, he says, "we took a warehouse in the back of our headquarters and converted it to one of the most advanced data centers in the world, without doing any construction. All we need to do is run our liquids and power in and out," he says.

"We have a large enough parking lot outside to put our hydrogen generation

and batteries. That's all we need to do the conversion."

He continues: "I personally believe that the conversion of buildings is probably more powerful than doing the 3D printing. There are ghost towns in every city, because people did not come back to the office at the same level as they were before Covid. There's so much empty space and it's going to waste right now."

"We are offering to the real estate partners that manage those sites that we can convert them to a very profitable data center."

Real estate companies are "very receptive" to this idea, Bachar says, because ECL is "creating an advanced data center that will extend the life of that building, and secure the real estate value for years or decades to come."

In future - 3D printing

Where there isn't a vacant building, 3D printing will come into play (for more, see the cover feature), working with environmentally-friendly building materials, to turn out 1MW units.

"Our blocks are very small," he says. "We will be doing 3D printing off-site, which is the most sustainable way of actually doing construction nowadays because it's waste-free. There's no waste coming from the site."

He also promises that ECL has selected a sustainable mix of concrete for the construction printer.

ECL plans to build the blocks one at a time, building them "back-to-back," he says: "We do not scale up, we only scale out and we optimize the way we design the data center to be able to do that in a cost effective way," Bachar says.

"To build a large site of 15 or 25MW, the integration goes in cycles, so we can actually deliver a full site in eight months, more or less."

In the long term, ECL plans to own the 3D construction printers, and make them available for others to use in the community when it has no construction of its own going on, to make housing or community buildings: "There's no reason for us to sit on the printer," says Bachar.

"Our first goal is to be successful as a data center company," he adds. "A secondary goal is to send a message to the industry, that it is possible to build sustainable data centers in 2024. We don't need to wait for 2025, 2030, 2035, or 2050, like we hear from some of the players in this domain."

He explains: "We are at an inflection point in demand for data centers. If we build non-sustainable, traditional data centers because we're in a rush, we will never get another opportunity like this to insert sustainable data centers into the market." ■

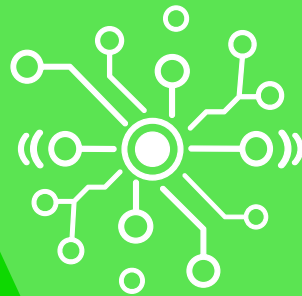


Sponsored by

Life Is On

Schneider
Electric

The Edge Supplement



From AI to space

INSIDE

The evolution of Satellite-connected IoT

> Satellite-based use cases for IoT, with help from the Edge

Banking on Edge AI

> Intel details how the Edge can help AI deployments, even if analysts aren't so sure

Inside Flexnode's plans to solve building headaches

> Modular data Edge centers can provide flexibility

IT professionals get more
CERTAINTY
when they operate at the edge.



APC Smart-UPS Modular Ultra lets you operate in any edge environment.

- Reduces installation cost and saves IT space.
- 50% smaller, 60% lighter with 2.5x more power.
- Simple remote management with EcoStruxure™.
- The most sustainable UPS of its kind.



EcoStruxure IT
modernized DCIM

Smart-UPS™
Modular Ultra

apc.com/edge



Sponsored by

Life Is On

Contents

32. The evolution of Satellite-connected IoT

How can Edge support satellite-based use cases for IoT?

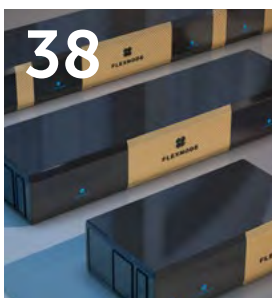
36. How to drive higher productivity on factory floors with industrial edge solutions

38. Banking on Edge AI

Intel details the opportunities at the Edge for AI deployments, and disputes analyst observations that Edge AI has plateaued

41. Inside Flexnode's plans to solve data center build headache

Modular data Edge centers can provide flexibility for the market



Edge meets IoT, AI, and DCs

It's 2024 and conversations around Edge continue to evolve in the data center world.

There's much debate about whether the industry has seen the full benefits of Edge.

A lot is promised though, with Statista projecting the Edge computing market to be worth \$317 billion by 2026. That's just two years away.

If those numbers are to be believed, it shows just how much potential there is around Edge computing, especially around all the use cases poised to proliferate at the Edge. There's a lot of promise around VR, AR, IoT, in particular, while AI continues to get the industry talking.

Satellite-connected IoT

Satellites can serve a range of purposes. They can be used to provide cellular connectivity, detailed weather updates, broadcasting, and for other use cases.

Though they have been around for years, they can provide a platform for IoT use cases to thrive.

IoT goes hand in hand with Edge computing, but the early feeling in the industry is that there's still some way to go before Edge can support satellite companies. Satellite-IoT is seen as the catalyst to drive this, and through the use of Edge, has the potential to support some critical use cases, such as wildfires.

However, how soon these applications are seen in the real

world, may take some time just yet.

Betting on Edge AI

By 2026, Gartner predicts that at least 50 percent of Edge deployments will incorporate machine learning, meaning the need for a platform to support Edge AI deployments has never been greater.

With over 90,000 Edge deployments across 200 million processors sold in the last ten years, Intel saw a gap in the market for an Edge AI platform. In 2022, the company hired Pallavi Mahaja to help build the platform that was finally unveiled at Mobile World Congress earlier this year.

As organizations increasingly look to deploy AI algorithms and models onto local Edge devices to speed up the decision-making and data analysis process, Mahaja discusses the benefits of Edge AI deployments and why she's pleased Intel's new platform is helping the company lead the charge.

Flexnode and its Edge

The difficulty of sourcing land has become a headache for the data center industry, with push back becoming more frequent.

This particular problem has fueled the growth plans of one start-up, Flexnode, a company that deploys Edge data centers, which it says can be deployed in all sorts of environments, while also more efficient for its customers and serving areas beyond the traditional data center hubs.

The distant world of satellite-connected IoT



Georgia Butler
Reporter

Connecting the unconnected

The satellite-connected Internet of Things (IoT) isn't new. According to Ground Control's Alastair MacLeod, some clients have been using a version of the technology for several decades.

Ground Control designs and builds satellite-connected technology including low-power small-form-factor satellite IoT devices and handheld, two-way messaging-enabled trackers. "We have clients that have been with us for almost 30 years, as an example of some of the oldest satellite-IoT use cases we have," says MacLeod, highlighting a water company that needs to remotely monitor its treatment sites.

Those older cases are usually limited to the ultra-isolated locations where satellite connectivity was the only option, and if you go back to that point 30 years ago, that client would have needed somewhere between a 0.9



Alastair MacLeod
Ground Control

meter and 1.6-meter parabolic antenna - expertly installed - to enable their satellite-connected IoT.

Things have changed drastically since then, and IoT itself has experienced significant developments.

The first IoT device was a connected toaster, created in 1990 by John Romkey for a bet. It was linked to the Internet via Transmission Control Protocol/Internet Protocol (TCP/IP) and could be turned on or off remotely. Romkey subsequently added a crane arm to insert the bread into the device, creating an end-to-end toasting system.

Romkey's kitchen experiments came before the term IoT had even been coined, but in the decades since connected devices have become ubiquitous, with advancements such as 5G networks and Edge computing enabling rapid data transfers and localized data processing which reduce machine-decision time. Latterly, AI has added the possibility of learned pattern behavior and, as a result, better responses.

This is fabulous for places with solid connectivity and power infrastructure

- there's no reason powerful Edge computing solutions cannot be deployed if the requisite energy and networking infrastructure are available. But what about those more remote places? What about when you are at the edge of civilization?

Luckily, satellite technology has also been undergoing a parallel revolution.

The first satellites were geostationary (GEO) - placed at an altitude of around 22,300 miles and directly above the equator, and not moving relative to the point on Earth directly below.

By the 1980s, Low Earth Orbit (LEO) satellites began to present a genuine challenge to GEO. These satellites reside at an altitude below 1,200 miles and are constantly on the move, orbiting the Earth in 128 minutes or less with an eccentricity of less than 0.25.

The proliferation of satellites

The number of active satellites in orbit has increased from 533 in 1994, to 8,377 in January 2024 - up a massive 1,472 percent. The bulk of that growth has occurred in the last few years, and is concentrated around LEO craft.

According to Ground Control's Macleod, decreasing launch costs have been a big factor.

"The cost of putting things in space has gone down dramatically and, depending on when you draw your yardstick, it could be by a factor of 10, or if you go back far enough, a factor of 100," he says. "This has meant that the assets put in space don't have to cost billions of dollars or have a 30-year life,

so more entrants have come into the market.

When the so-called "space tax" was higher, the only motivation to use satellite connectivity was if lives, or millions of dollars, were at stake. But as with any technology, as it becomes more affordable, it can reach a larger commercial audience, and in 2022 satellites were dragged into the world of 5G through the most recent update to 3GPP, Release-17.

3GPP Release-17

3GPP is an initiative that provides specifications across cellular telecommunications technologies. It was first founded in 1998, and set out to establish technical specifications and reports for a 3G mobile system - to standardize connectivity.

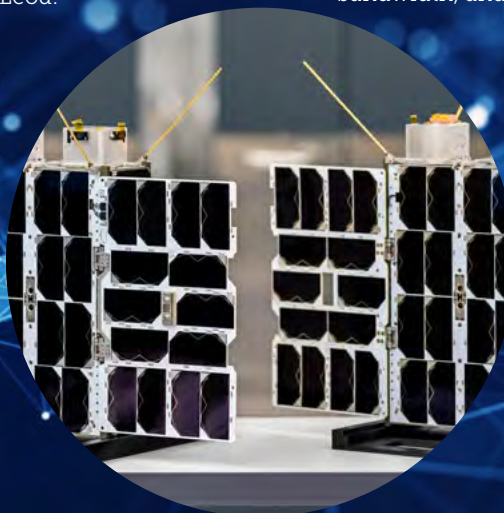
Release-17 was officially frozen, meaning it received no further additions or changes, in June 2022, and brought about a big change in the market.

"3GPP is the starting pistol on the convergence between space-based IoT, or satellite IoT, and terrestrial IoT, on what might come next, and broadly that is Narrowband IoT (NB-IoT)," explains MacLeod.

"[NB-IoT] is using the terrestrial IoT technology and protocol, and putting it in the space segment so that, eventually, you can have a device that switches between the terrestrial and space network."

Release-17 includes a range of specifications for 5G non-terrestrial networking (NTN) for platforms at various orbits, including GEO, MEO (medium earth orbit), LEO, and HAPS (high altitude pseudo satellites). Specific to IoT is the release of 5G IoT-NTN, which provides throughputs ranging from one-100 Kbps, to Release-17 NB-IoT and eMTC (a low power wide area technology that provides a comparatively high data rate service for IoT applications) devices.

The release also includes two frequency bands to support this, both of which are in the 5G FR1 frequency range 255 in the L-band (1626.5-1660.5 MHz UL/1525-1559 MHz DL) with 34+34 MHz FDD bandwidth, and n256 in the S-band



(1980-2010 MHz UL / 2170-2200 MHz DL) with 30+30 MHz FDD bandwidth.

MacLeod notes, however, that to date there is no one offering devices truly compatible with Release-17 and able to switch between space and terrestrial IoT network types seamlessly according to the 3GPP standards.

Andrew Nuttall, CTO and co-founder of Skylo, a GEO satellite connectivity provider, estimates that there are currently around 10 billion devices connected over cellular networks. "There's maybe a few million mobile satellite-connected devices, so if you look at those two ecosystems it's around four orders of magnitude," Nuttall says.



"People often have not just one, but multiple cellular devices connected in parallel. You can buy satellite phones and devices today, but it's completely separate. You are buying separate hardware, separate equipment, and you're paying different people."

The vision is that IoT, and mobile phones, will be designed so that as they cross out of terrestrial connectivity, they can automatically switch to satellite. Devices will no longer be either or, they will be both, offering a much more reliable network as when a device loses contact with the terrestrial network and permanently available alternative can be used.

"Satellite is wonderful from a coverage perspective," says Nuttall. "Anytime you see the sky, you have satellite connectivity. The challenge lies in it being a separate device, and that ecosystem has not really

proliferated or grown at scale."

Getting to that point, MacLeod predicts that we will first see people using 3GPP-type standards over satellite links, but they won't immediately be interoperating.

"Things can change, but in order to make the space segment super efficient, it currently uses a data protocol that's referred to as NIDD - non-IP-based data delivery - which is optimized for trickier links," explains MacLeod.

"But NB-IoT doesn't use it, so the current style of addressing data communication in space isn't mirrored by that on the ground network. Of course, that will change, but none of us knows exactly how long it will take."

The silicon needed to enable devices to switch seamlessly is currently under development by the likes of Qualcomm and others, but isn't yet widely available. Current estimated timelines seem to suggest it will be later this year. Regardless, a number of players are entering the space of space and making their play for a part of it.

Edge computing

There is some disagreement on what this will mean for Edge computing.

IoT and Edge computing, in many use cases, go hand-in-hand. The issue lies in the location of these satellite-connected IoT devices.

"Anytime you see the sky, you have satellite connectivity. The challenge lies in it being a separate device, and that ecosystem has not really proliferated or grown at scale."



For the most part, if you are opting for a satellite connection rather than a terrestrial one, you are in a remote place with very limited connectivity of other kinds. But alongside limited connectivity is often a limited power supply.

"Electrical power can be as scarce or even more scarce than connectivity," says MacLeod.

The amount of power needed to run a satellite terminal in itself can vary significantly. At the higher end there is, for example, the Gen 2 Standard Actuated Starlink dish that uses 50-75 watts.

"We've had customers complain because something we sell uses five watts, and occasionally it might reach nine when it is transmitting," MacLeod says.

These power constraints also apply to Edge computing solutions. If you are struggling to find five watts for satellite data transmission, you are likely also going to be limited in the power you can use for data processing at the location.

But the other side of the coin is that satellite airtime is generally more expensive than terrestrial airtime, meaning that keeping data to a minimum is preferable, and consolidating that data may require Edge computing.

"What you want to do at the Edge is have as few end-points as possible and use as little data as possible," MacLeod explains.

"A classic setup would be sensors all over the place which are then connected to LoRaWAN gateways to do concentration, and then those gateways communicate to the satellite terminal and send it back."

Currently, MacLeod hasn't seen many examples of this in the real world, but Skylo's Nuttall predicts that the use of Edge computing is going to increase as satellite IoT grows. "If anything, this increases the need and desire to do Edge computing," he says.

There are a few different motivations for Edge computing. "One, is you want to reduce the amount of data you backhaul, so you can make decisions locally," says Nuttall.

The other key reason is latency, but if your Edge computing needs are to reduce latency by a few milliseconds, then realistically satellite connectivity isn't the way to go.

"Additionally, this [satellite-connected IoT] is a narrowband service, this is not a broadband service. So if you're generating gigabytes of data then you should not be backhauling that over this network, you should be doing your analysis locally and then just backhauling the results or outcomes."

One scenario where Edge computing and satellite connectivity could coincide was posited by Shahbaz Ali, head of products at Sateliot, which provides satellite IoT services.

"An example could be an emergency forest fire response service in, say, the Amazon rainforest," Ali says. "Edge computing would be necessary to take a decision right there and then based on data that sensors collected to trigger an anti-fire system, and you can't have 100 percent connectivity guaranteed by any terrestrial network."

As for whether that compute would ever happen actually on a satellite itself, or a space-based data center type deployment, Ground Control's MacLeod is skeptical, saying he does not expect to see it "in the foreseeable future."

This is because the cost of getting compute into orbit, in addition to the cost to get the data transmitted to space, makes such a task impractical. "There's too many things driving you in the opposite direction, the common sense direction, to do that," MacLeod says.

GEO vs LEO

Skylo has committed to GEO satellites for its offering. Though data transfers from these craft incur higher latency than their LEO counterparts, Nuttall says the company made this choice because its key purpose is to "solve

the convergence of ecosystems by eliminating behavior changes."

"There are plenty of different ways to connect devices over satellite, but there is a very specific set of solutions that enable that convergence," he explains.

"We want to use standardized frequency bands because we don't want to force people to choose bespoke technology solutions, we want to be able to use the globally harmonized regulated mobile satellite service spectrum assets, and we want to provide continuous and contiguous connectivity. To meet those three criteria, the only applicable satellites today are GEO."

The current latency for those GEO satellites is around 500 milliseconds, which depending on your use case, could matter a lot or not at all.

Skylo has yet to officially launch its satellite, but another company, OQTechnology, began investigating NB-IoT connected to LEO satellites in 2016, with the plan to build a constellation or network specifically to connect to low-power IoT devices.

Currently, OQ has 10 LEO satellites in orbit and has global coverage which it claims is commercially viable. It also purports to be Release-17 compliant.

Connectivity via LEO satellites is variable because, unlike GEO, they are constantly on the move. For OQ, revisit time - the time between which a satellite will return to the same point on the earth twice - sits between four-six hours.

"That's good for some applications," says founder and CEO of OQ, Omar Gaise. "A lot of IoT devices don't have to constantly communicate and they are actually made for 'burst' communication. You can have them just powered by battery, and wake up to send a burst of data, and then go back to sleep."

But as the company adds more satellites to its network, those periods of Internet desert will shrink.

Sateliot, one of the companies that drum up a lot of noise in the IoT-satellite world, is also looking to utilize LEO for this purpose.

It was a major contributor to Release-17, and claims to be the world's first NB IoT LEO satellite operator that is implementing that standard for the market.

Similarly to Skylo, however, Sateliot is not yet commercial. The company had deployed, as of December 31, 2023, five prototype satellites, even though it plans around 256 in the long term. As for when further satellites will rocket into space, the company has a launch date for some time this year, although it has not been shared publicly.

When those satellites have been launched, and the standardized NB IoT connectivity becomes available, mobile network operators will be able to offer seamless connections to satellite and terrestrial networks and Sateliot says it will be able to serve more time-critical applications in an affordable way.

Sateliot's Ali concedes that while things are moving along in the world of satellite IoT, cost is still a constraint.

"What is stopping satellite-IoT from having massive adoption are the existing cost barriers," Ali says. "Satellite connectivity hardware prices sit somewhere between \$80 and \$350, and then the connectivity is between \$10 to \$35 per device per month

"That is very expensive when looking at sectors such as agriculture or livestock where you need to deploy millions of devices. There's no business case for that."

Overall, there is still some skepticism about the technology. OQ's Gaise puts some of this down to the bad name the sector got after a lot of satellite IoT startups appeared, only to disappear. "This gave a bad vibe that the model doesn't work for satellite operators," he says.

MacLeod also dampens some of the hype, dispelling what he describes as some people "throwing a lot of capital up in the sky" along with claims of radical new technologies and a "new dawn."

"Ultimately, the irritating thing about engineering and radio transmission physics, is the laws of physics don't cease to exist. Space is always going to be further away." ●

How to drive higher productivity on factory floors with industrial edge solutions



Data latency is one of the main issues driving manufacturers towards more edge computing. Many industrial applications require rapid response to assure product safety and quality. Computing power and storage are being placed at the industrial edge to lower data transport time and increase availability. Edge computing brings bandwidth-intensive and latency-sensitive applications closer to the user or data source.

Simplifying and standardizing industrial edge computing technology

A global network of technology partners acts as a provider of these industrial edge solutions. Although edge computing solutions tend to be small in size and footprint compared to traditional data centers, their function is often critical. Users who deploy edge systems face the challenges of not having IT resources on-site to support these systems. In addition, edge systems are often made up of multiple parts and pieces from multiple vendors, and configuring, delivering, and installing edge systems requires collaboration.

In such an environment, strong partnerships emerge as a critical success factor. No one company can do it alone. The task of properly managing multiple, local and remote edge systems can also be complex. This is why key OT technology providers like *Schneider Electric* partner with *Dell Technologies*, a global giant in IT server and storage solutions.

With deep networks of IT and OT

Global market changes have forced manufacturers to rapidly adjust their factory systems to be more flexible and agile. Part of the operational reassessment involves developing an ability to accommodate more product variety, shorter resourcing of raw materials, faster production, and rapid delivery. These marketplace demands are driving the increased pace of digital transformation at many plant sites. The trend is often referred to as Industry 4.0, and it involves merging traditional OT or operations technologies with IT *edge computing* systems.

The manufacturing industry is rapidly deploying edge computing across a wide variety of use cases. According to *Grand View Research*, the edge

computing market was valued at \$7.43 billion USD in 2021, and it is expected to expand at a compound annual growth rate (CAGR) of 38.9% from 2022 through 2030.

Why the industrial edge is growing

The *industrial edge* is one of the fastest-growing segments of industrial automation. Examples of devices within industrial edge environments include sensors, programmable logic controllers (PLCs), and local machines with connectivity capabilities. To increase plant profitability, manufacturers must correctly measure asset performance, rapidly identify problem areas, and make crucial, real-time changes.

partners across the globe, both Dell and Schneider Electric provide the programs and resources needed to simplify the configuration, delivery, installation, operation, and maintenance of edge computing systems.

Below are several examples of the Schneider Electric and Dell collaboration in the industrial edge space.

Micro data centers as solutions

Micro data centers are small, preassembled, pretested, and integrated solutions that are configured into an enclosure. The racks, servers, storage devices, power distribution units, uninterruptible power supplies (UPS), and management software can be made ready to operate before arriving at the end-user location. The core elements of the system all work together, so, end-users in edge environments, most of whom are unfamiliar with the workings of IT, have peace of mind knowing these systems can operate on their own, typically without any local support required. Additional security devices, such as racks with card key access and server room security cameras, can also be easily added to the micro data center configuration.

When Schneider Electric's *EcoStruxure™ Micro Data Centers* are matched with Dell Technologies ruggedized Vx-Rail P-Series High Performance Servers, end users gain access to an integrated solution that is both reliable and remotely manageable.

Reference designs for rapid deployment

Reference designs are a series of tested, validated, and documented plans for assembling and laying out specific physical infrastructure components (racks, power, cooling, security) aligned to common customer environments. They can be refined to meet particular needs and provide 90% of what is needed to help expedite the delivery of the edge computing project. Schneider Electric has jointly developed specific *reference designs* for industrial edge environments, including the Dell IT components. Reference design libraries are used to select an integrated solution. Those designs are then refined to accommodate the unique requirements of the particular factory OT edge environment.

Reference designs provide benefits to both end users and partners in several ways:

- **Reduced engineering time** – Pre-tested and validated design reduces trial and error of matching up compatible components.
- **Rapid deployment speed** – Preassembled units mean ready-to-go solutions are available for operation upon delivery without delays from waiting for missing components.
- **Lower risk of human error** – Fewer people are required for systems installation and setup and built-in physical security greatly reduce the possibility of unanticipated errors.



- **Higher sustainability** – Micro data center solutions only require one shipment rather than the shipping of multiple separate parts and pieces. This has a positive impact on reducing *Scope 3 emissions*.

Networks of Managed Service Providers (MSPs) for support

Dell and Schneider Electric channel partners are expanding their business models to offer remote edge computing support services. This helps end-users by providing them with a vast network of edge computing experts for management, technical support, and predictive maintenance services. In addition, these MSPs support end-users by remotely managing edge sites to ensure high system availability, energy savings, and predictable system performance.

Based on the individual end user's business requirements, the power of this partner ecosystem works to manage and simplify the client/vendor interface as much as possible. The success of the partnership ecosystem not only depends on enlisting organizations with different micro data center and edge computing competencies, but also can include partners who are experts within specific geographic regions to accommodate customers with an extended global presence. Enterprise organizations want to experience consistent support for edge IT (computing, network, and storage) and OT (power, cooling, environmental monitoring, and security) assets across all geographies. Combined, these part-

nerships give end users peace of mind knowing that project deployment risks are minimized.

Practical solutions for the industrial edge

Download the reference design to learn more about Schneider Electric and Dell's micro data center, particularly for industrial edge applications. Also, discover what *Schneider Electric's and Dell's alliance* can do for your organization.. ●



Flexnode: Deploying scalable modular data centers at the Edge



Paul Lipscombe
Telco Editor

Following a recent seed funding, Flexnode has its eyes on further deployments, both in the US and internationally

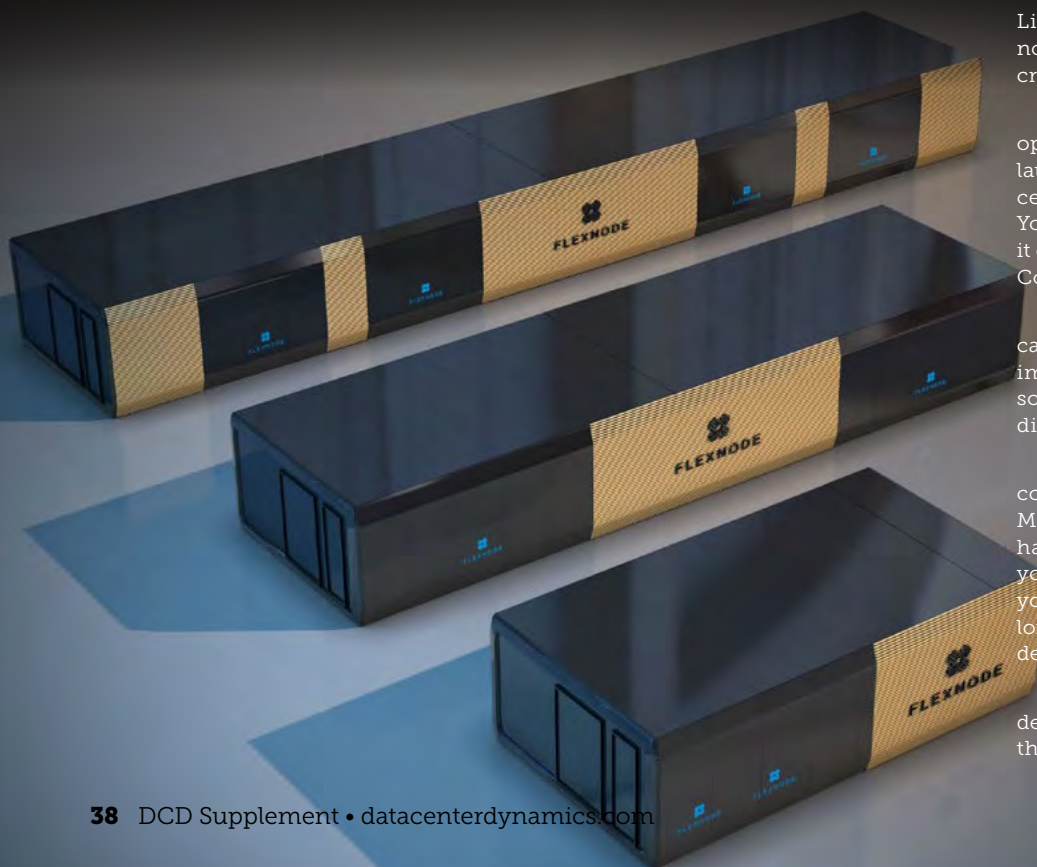
"We have one site deployed so far, and we're going to deploy a couple more this year," says Andrew Lindsey, CEO and co-founder of Flexnode. "A large part of this process was creating a strong foundation,"

Flexnode designs, deploys, and operates Edge data centers. The company launched its first and currently only data center last year in Long Island City, New York, but has ambitions to deploy more so it can serve customers beyond the East Coast.

The site has a 1MW data room (80kW capable racks); liquid-cooled IT (precision immersion and now cold plate); GPU solutions and CPU solutions operational; digital twin running.

Lindsey, who co-founded the company in 2019 along with Robert Mazer, said: "If you look at what has happened in the built environment when you see various solutions come about, you need to have those partners on board long before you start going to market and deploying things,"

He adds that the company wants to deploy one or two more data centers by the end of the year.



Transition to the data center market

Linsley's background is in construction engineering. He previously worked at Alpha Consortium, a company founded by his father, that focuses on mission-critical projects.

It was while working there that he identified what he saw as a challenge in the data center construction industry.

"The interest came about for Flexnode when we discovered that project timelines weren't getting shorter, they were getting longer," he says. "We were seeing this in full view, where land was getting much more difficult to find in these data center hub locations, and if you could find it, it was prohibitively expensive."

Sourcing power for these data centers was becoming a challenge for the industry, says Lindsey, as is the overall supply chain, which he says was getting "blocked up."

"We can't train up people fast enough to be able to deliver the kinds of infrastructure that we're looking at over the next five years," he says. "We're talking about north of a billion square feet in five years, just nutty amounts of infrastructure on any measure. And as the power companies are saying to us today, this is a once in a multi-generational problem."

Flexnode, however, wasn't created to build traditional data centers, but rather a strategy focused around developing a modular data center approach, designed to be flexible and scalable.

The company had identified this approach as the most sustainable long-term strategy for its own data center growth plans. Lindsey says it had also become aware of the perception from the public around data center construction in recent years.

"We have groups now that are far more aware of that 600,000 sq ft facility that's sitting in their backyard and communities are pretty angry about the fact that they bought this new house that had this pretty view, and now they can't see more than 30 ft because there's a 60 ft wall behind it. Five years ago this was just rumblings and now it's in full view."

Take me to the Edge

Requirements around the build-out of data centers have also led to a rethink of how these buildings need to be constructed.



Andrew Lindsey

Building data centers at the Edge is the way forward to combat some of the challenges that the industry faces, Lindsey argues.

"We see Edge as a way to activate infrastructure very quickly, where today, as you know, we have a wildly low vacancy rate," he says.

"That should continue through the next several years, and people still need data, so we see this as a way to scale out where we have now. We're now on track over the next five years to be able to scale out at gigawatt scale."

Edge is already supporting several use cases at present, such as the mining industry in remote areas where power isn't always easy to come by. It's also tipped to underpin autonomous vehicles, smart grids, and medical advancements in the coming years.

Designed to be deployed wherever

Flexnode's data centers are designed to be deployed easily, in various locations, indoor or outdoor, claims Lindsey.

"We wanted to create an integrated platform that allows for our customers to build and deploy data centers, with an experience that's much more akin to building and buying a car," he explains, noting that Flexnode focuses on connecting three key parts.

"The first part was an industrialized building system that is designed for disassembly, configurability, adaptability, and designed to go anywhere."

The second part, he says, is focused on its fully digitized process which helps its customers design and configure their data centers. The final part, he adds, is the ecosystem of partners Flexnode

partners with, spanning engineering and construction.

Flexnode works with several partners to deploy its data centers, which Lindsey says allows the company to build everywhere. The company works with Arup Engineering, SHoP Architects, UK-based structural engineering firm Thornton Tomasetti, Jensen Hughes, JE Dunn Construction, and Boyce.

US businesses have to be prepared for an array of natural disasters, and Lindsey adds that Flexnode has designed its modular data centers to be able to survive in some of the harshest environments, notably category-five hurricanes, tornados, and earthquakes.

He explains that Flexnode integrates a system that can collapse down into components, which can fit in a variety of spaces. The data centers can go from four to 36-rack configurations.

"We have a fully digitized process that starts from the earliest stages of design and configuration," he says. "That digital process continues through the entire lifecycle of a data center."

Indeed, Lindsey says his company creates a digital twin of every data center it builds. "What it is for us is everything from a configurator to a management and controls platform, to a training platform for our service people to a monitoring platform for our customers," he says. "And that together really makes up a digital twin."



Investment

Flexnode raised \$8.85 million in a Series A funding round in March 2023. The money will go towards helping the company scale out its operations, says Lindsey.

The raise came after Flexnode secured a \$3.5m grant from the US Department of Energy as part of the COOLERCHIPS



"I will say that we do have international expansion in our sights, both because of several opportunities that are developing in Europe and in the APAC areas. But also because we're designed for it."

program in May of last year.

Working for ARPA-E – the energy unit of the US government's Advanced Research Projects Agency innovation agency – it will develop a prefabricated, modular Edge data center with a novel manifold microchannel heatsink, a chassis-based hybrid immersion cooling approach, and an additive manufacturing-enabled dry cooling heat exchanger system. This will all be housed in a topology-optimized container.

The investment follows years of hard work from the company's founding team, Lindsey says. "For the first couple of years it was our own money," he says. "We bootstrapped pretty heavily, put a lot of resources into it. And then we started seeking external capital, which is where we've had some raises."

Lindsey tells DCD that the company aims to deploy more across the US later this year, but won't just deploy data centers for the sake of it.

"We're not planning on deploying these data centers like Bird Scooters," he says, referring to the US micromobility company which has scooters operating across many US states.

"It's just not our approach. Because it's a building, we don't see it as a very, very large appliance, and as a result, we respect the communities that we go into."

He says Flexnode seeks to "offer value to the broader communities that we go into," adding: "This isn't a lift, ship, drop, and leave - it's about integrating into those communities as well."

AI and the challenges it presents

The emergence of AI has presented challenges across all industry sectors, and the data center business is no exception.

But as AI promises so much, the

technology will need a lot of compute power to operate. It will require a huge number of high-density servers to accommodate AI workloads, and this means many operators are looking at how their facilities are designed.

Lindsey states that Flexnode is aware of the challenges that AI poses for the industry.

"It's a huge problem," he says about the challenges around AI. "We've had a lot of discussions that are ongoing around groups that are rolling solutions out at a building scale that support AI."

"But these buildings are still facing the same project challenges that every other one has confronted before them. Those five to seven-year project timelines are still real for those other buildings, regardless of how cool their technology is inside their facility."

He notes that it's difficult for older data centers to retrofit their existing buildings because of the demands of AI.

Flexnode has worked with clients to address the growing demand for AI capabilities as part of its phased AI Edge data center approach.

According to Flexnode, the expansion focuses on air-cooled systems, providing immediate relief to the capacity constraints. It says that this initial phase is designed with the future in mind, "allowing for a smooth transition to liquid cooling systems as the demand for AI capabilities grows and the need for more efficient cooling solutions becomes apparent."

International ambitions

Looking further into the future, Lindsey wants to take Flexnode global, specifically in the EMEA and APAC regions.

"I will say that we have international expansion in our sights, both because of

several opportunities that are developing in Europe and in the APAC area, but also because we're designed for it," says Lindsey.

"We have the ability to go into locations that have heat reuse requirements, and sustainability requirements. These are things that we are truly tuned for."

He says this is likely to take two to three years.



As for what's on the horizon for the rest of this year, he says another Series A round is being lined up.

"We are planning to start this within 2024," he adds. "We're planning to build out our constellation of sites where we can support customers to deploy anywhere and everywhere."

Although he admits the transition into the data center world has been challenging at times, Lindsey is keen for Flexnode to leave its stamp on the market.

"Has it been stressful at times? Absolutely. We have gone into the deepest and darkest corners of the built environment," he says.

"But I will also say that it has been unbelievably exciting. We have a real opportunity to revolutionize the built environment, not only in the world of IT, but the truly built world of environment, methodologies, and technologies." ●

Why Intel is making big bets on Edge AI



Charlotte Trueman
Compute, Storage, &
Networking Editor

Intel's new corporate vice president, Pallavi Mahajan, talks about the growing need for Edge AI and why the company's new platform is so necessary

As is the case with all things AI in recent history, Edge AI deployments have not been immune to exponential growth.

As the pendulum has swung from centralized to distributed deployments, AI has driven the majority of growth in Edge computing, with organizations increasingly looking to deploy AI algorithms and models onto local Edge devices, removing the need to constantly rely on cloud infrastructure.

As a result, research from Gartner shows that at least 50 percent of Edge deployments by the year 2026 will incorporate machine learning, a figure that sat at around five percent in the year 2022.



● The Edge Supplement

Edge is not the cloud

Businesses want the Edge to bring in the same agility and flexibility as the cloud, said Pallavi Mahajan, corporate vice president of Intel's network and Edge group software. But, she notes, it's important to differentiate between Edge AI and cloud AI.

"Edge is not the cloud, it is very different from the cloud because it is heterogeneous," she says. "You have different hardware, you have different servers, and you have different operating systems."

Such devices can include anything from sensors and IoT devices to routers, integrated access devices (IAD), and wide area network (WAN) access devices.

One of the benefits of Edge AI is that by storing all your data in an Edge environment rather than a data center, even when large data sets are involved, speeds up the decision-making and data analysis processes, both of which are vital for AI applications that have been designed to provide real-time insights to organizations.

Another benefit borne out of the proliferation of generative AI is that, inferencing can happen in an Edge environment, reducing the time required to send data to a centralized server and receive a response.

Meanwhile, talent shortages, the growing need for efficiency, and the desire to improve time to market through the delivery of new services have all caused businesses to double down on automation.

Alluding to the aforementioned benefits of Edge computing, Mahajan



said there are three things driving its growth right now: businesses looking for new and different ways to automate and innovate, which will in turn improve their profit margins; the growing need for real-time insights, which means data has to stay at the Edge; and new regulations around data privacy, which means companies have to be more mindful about where customer data is being stored.

Add to that the fact that AI has now become a ubiquitous workload, it's no surprise that organizations across all sectors are looking for ways to deploy AI at the Edge.

A large number of organizations deploy smart devices to support their day-to-day business operations, be that MRI machines in hospitals, sensors in factories, or cameras in shops, all of which generate a lot of data that can deliver valuable real-time insights.

GE Healthcare is one Intel customer that uses Edge AI to support the real-time insights generated by its medical devices.

"There are three things driving Edge computing's growth right now: businesses looking for new and different ways to automate and innovate; the growing need for real-time insights; and new regulations around data privacy"

The American healthcare company wanted to use AI in advanced medical imaging to improve patient outcomes, so partnered with Intel to develop a set of AI algorithms that can detect critical findings on a chest X-ray.

Mahajan explains that, in real-time, GE's X-ray machines scan the images that are being taken and automatically detect if there's something wrong with a scan or if there's an anomaly that needs further investigation.

While the patient is still at the hospital, the machine can also advise the physician to take more images, perhaps from different angles, to make sure nothing is being missed.

The AI algorithm is embedded in the imaging device, instead of being on the cloud or a centralized server, meaning any potentially critical conditions can be identified and prioritized almost immediately.

"Experiences are changing," Mahajan says. "How quickly you can consume the data and how quickly you can use the data to get real-time insights, that's what Edge AI is all about."

Intel brings AI to the Edge

Mahajan joined Intel in 2022, having previously held software engineering roles at Juniper Networks and HPE. She explains she was hired specifically to help build Intel's new Edge AI platform.

Unveiled at Mobile World Congress (MWC) in February 2024, the platform is an evolution of the solution codenamed Project Strata that Intel first announced at its Intel Innovation event last year.

"[Intel] has been working at the Edge for many, many years... and we felt there was a need for a platform for the Edge," she explains. Intel says it has over

“According to Gartner, at least 50 percent of Edge deployments by the year 2026 will incorporate machine learning, a figure that sat at around five percent in the year 2022”

90,000 Edge deployments across 200 million processors sold in the last ten years.

Traditionally, businesses looking to deploy automation have had to do so in a very siloed way. In contrast, Mahajan explains that Intel’s new platform will enable customers to have one server that can host multiple solutions simultaneously.

The company has described its Edge AI offering as a “modular and open software platform that enables enterprises to build, deploy, run, manage and scale Edge and AI solutions on standard hardware.”

The new platform has been designed to help customers take advantage of Edge AI opportunities and will include support for heterogeneous components in addition to providing lower total cost of ownership and zero-touch, policy-based management of infrastructure and applications, and AI across a fleet of Edge nodes with a single pane of glass.

The platform consists of three key components: the infrastructure layer and the AI application layer, with the industry solutions layer sitting on top.

Intel provides the software, the infrastructure, and its silicon, and Intel’s customers then deploy their solutions directly on top of it.

“The infrastructure layer enables you to go out and securely onboard all of your devices,” Mahajan says. “It enables you to remotely manage these devices and abstracts the heterogeneity of the hardware that exists at the Edge. Then, on top of it, we have the AI application layer.”



This layer consists of a number of capabilities and tools, including application orchestration, low-code and high-code AI model and application development, and horizontal and industry-specific Edge services such as data thinning and annotation.

The final layer consists of the industry solutions and, to demonstrate the wide range of use cases the platform can support, it has been launched alongside an ecosystem of partners, including Amazon Web Services, Capgemini, Lenovo, L&T Technology Services, Red Hat, SAP, Vericast, Verizon Business, and Wipro.

Mahajan also lists some of the specific solutions Intel’s customers have already deployed on the platform, citing one manufacturer that is automatically detecting welding defects by training its AI tool on photos of good and bad welding jobs.

“What this platform enables you to do is build and deploy these Edge native applications which have AI in them, and then you can go out and manage, operate, and scale all these Edge devices in a very secure manner,” Mahajan says.

At the time of writing, a release date had not been confirmed for Intel’s Edge AI platform. However, during MWC, the company said it would be “later this quarter.”

‘AI Everywhere’

Although Gartner predicted in 2023 that Edge AI had two years before it hit its plateau, Intel is confident this is not the case, and has made the Edge AI platform a central part of its ‘AI Everywhere’ vision.

Alongside its Edge AI platform, Intel also previewed its Granite Rapids-D processor at MWC.

Designed for Edge solutions, it has built-in AI acceleration and will feature the latest generation of Performance-cores (P-cores).

Writing on X, the social media platform previously known as Twitter, in October 2023, Intel’s CEO Pat Gelsinger said: “Our focus at Intel is to bring AI everywhere – making it more accessible to all, and easier to integrate at scale across the continuum of workloads, from client and Edge to the network and cloud.”

As demonstrated by the recent slew of announcements, Intel believes that Edge AI has just reached its peak, with Mahajan stating that all industries go through what she described as “the S Curve of maturity.”

Within this curve, the bottom of the ‘S’ represents those tentative first forays into exploring a new technology, where organizations run pilot programs and proof-of-concepts, while the top of the curve is the point at which the market has fully matured.

“This is where I think we are now,” she says, with the platform representing the level of maturity. She believes Intel was “the first to read the need for [an Edge AI] platform.”

She continues: “This is the feedback that we got back from after the launch at MWC, that everybody was saying ‘Yes, this market needs a platform.’

“I’m sure there will be more platforms to come but I’m glad that Intel has been a leader here.” ●

Realize your company's IT potential

Software to design, monitor, and manage your IT space

62% of IT outages can be attributed to IT infrastructure failure¹. Our Data Center Infrastructure Management (DCIM) 3.0 offer provides device monitoring, health assessments, and more so you can:

- Run simulated impacts to expose vulnerabilities in IT infrastructure and address them immediately
- Reduce physical threats by monitoring IT environmental conditions
- Improve sustainability efforts by tracking PUE, energy, and carbon emissions

#CertaintyInAConnectedWorld

apc.com/edge



EcoStruxure™ IT
modernized DCIM

APC Smart-UPS™
Modular Ultra

¹Uptime Institute Global Data Center Survey, 2018

HAMR - the savior of hard drives



Peter Judge
Executive Editor

What happens when hard drive capacities double?

For years now, solid-state storage makers have been telling us that hard disk drives (HDDs) have a limited life ahead of them, as their performance is eclipsed by Flash memory,

Against that, hard drive makers keep increasing their capacity and delivering lower costs per Tbyte of storage, which have kept HDDs in use for the majority of storage.

"In the data center and the enterprise environment, right now, about 90 percent of all data sits on disk drives," says B S Teh, chief commercial officer at Seagate. "There's always talk about technology that

will take over disk drives, but I think with innovation, it's always proven itself to be the mass capacity storage of choice."

HAMR time

He is talking to DCD because he has an announcement to make: "This is not just any technology announcement. I would categorize this as one of the most impactful and significant technology shifts, that the [hard drive] industry has seen, in its entire history."

That's a big claim, but he's talking about heat-assisted magnetic recording (HAMR), a technology which is set to double the capacity of HDDs, by yet again increasing

the density of the data they can hold.

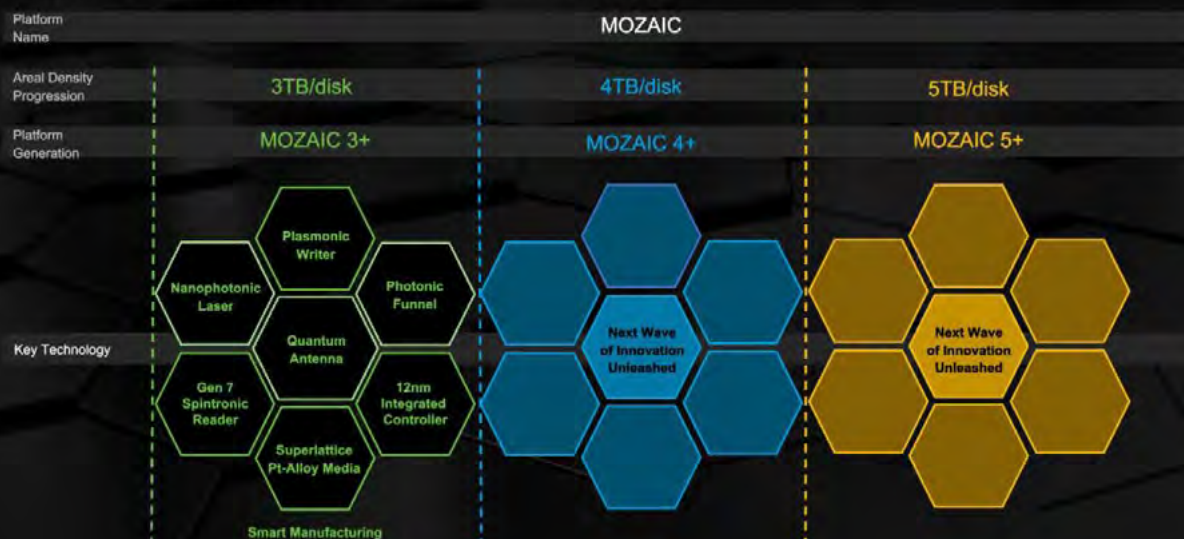
"We are reducing the size of the magnetic particles, and we're also squeezing them a lot tighter together," he says. "The result is a dramatic increase in areal density,"

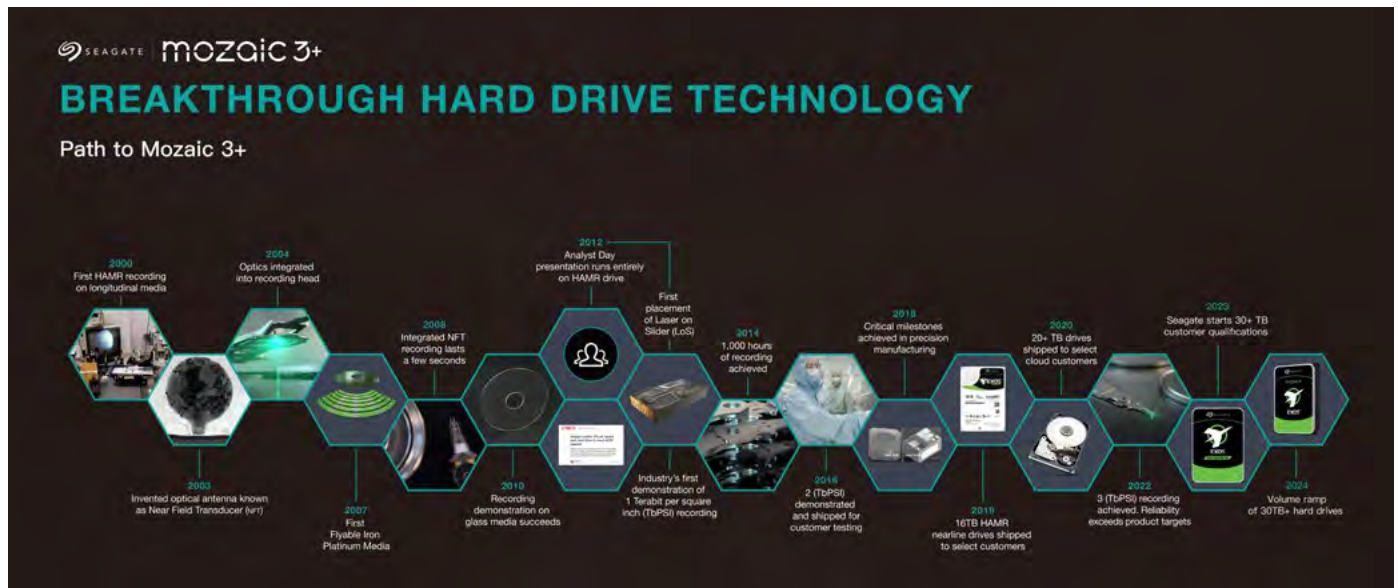
The basics of the technique are easy to explain. HDDs work by writing and rewriting the magnetic polarity of tiny grains in the disk platters as they spin. Density increases as the particles get smaller, but so does the risk that the grains will lose their magnetism, or be remagnetized by nearby reads and writes.

"If you put two magnetic particles too

Tomorrow's advancements are happening now

Mozaic represents both current and future generation technologies driven by Seagate innovations that increase areal density.





closely together, it becomes unstable," says Teh. "That's just physics."

It's possible to magnetize smaller grains, in a platinum alloy medium, but there's a problem. It doesn't work at room temperature. HAMR gets around that by quickly heating up a tiny dot on the medium, so it can be magnetized, and then letting it cool and stabilize.

"To raise the temperature so you can write on it," he explains, "we use a laser beam that will zap the magnetic bit, and heat it up so that we can then read and write."

Stagnant HDDs

It's a significant moment because, frankly, HDDs have been in the doldrums.

Shipments of HDDs have been dropping for the last two years, and Seagate's revenue fell by 50 percent. To be honest, some may be wondering whether we should care about his new change,

Teh thinks HDDs have been hit by a general slowdown in the economy, supply chain problems, and also a period during the Covid pandemic when customers overbought: "The fact is that the last few years was a very tough time for the industry."

Against that backdrop, HAMR, which Seagate is launching under the name Mozaic 3, is not actually that new and surprising. Seagate first planned to deliver it way back in 2009, and that date got pushed to 2015, 2019, and 2021. *DCD* has been hearing about it for years.

Teh explains: "I think there are two challenges. Number one, is the challenge of productizing the technology. There's been more than 20 years of R&D and hundreds of millions of dollars in investment, starting

from the year 2000."

Secondly, he says: "There is the continuous innovation of the existing technology." Thanks to the addition of helium, and other tweaks, "the older technology continued to have legs, so we were able to extend the lifecycle of the technology and push out the platform transition."

Alongside the development of HAMR, there have been other ideas including a stopgap called "shingled" recording and microwave assisted magnetic recording (MAMR) which he sees as "bridging" technologies: "MAMR is not a technology that will allow you to go from A to B, but it, may take you from A to A+."

Boost in demand

Now is the time to announce the change because AI will demand more storage, especially as it moves from a focus on training to inference (application): "Training is very GPU and compute-centric, but when you move to inference, where you're actually using the models, your compute requirement drops dramatically.

"You don't need those expensive GPUs to run every query, and it becomes more and more data-centric, because you're generating and creating more data as you use the application."

Between 2022 and 2026, he believes the global "datasphere" (the world's total storage) will more than double. HAMR will enable that, he says, "With HAMR technology, we're able to double the growth that we've seen in the last 10 years."

In recent times, faced with slack demand, densities have gone up by less than 10 percent a year, he says: "Now, we're talking about doubling that to over

20 percent CAGR of the areal density," he beams. "To put it another way, in about four years, you can see a doubling of the capacity in areal density and consequently in the size of the disk drive that you see."

As well as greater capacities HAMR will also address demands for space and power, he says. HDDs will hold more storage in the same size units, which have essentially the same cost and environmental footprint.

"For a given floor space, you can increase the amount of capacity that you can deploy in a data center," he says, "If you take out all the 16TB drives and replace them with 30TB units, you immediately save 40 percent in Watts per terabyte."

He thinks that the new units will have the same service life, lasting around 10 years: "We've got every reason to believe that it'll be just as reliable or even more reliable than previous generations."

Adding tiny lasers to the read/write head will add an increment to the price, he admits. While Seagate has produced the new heads and media in-house, along with the controller chips, the laser is new: "We never used a laser before, and went with an outside partner to develop that."

The partner produces lasers that are fast and focused enough to operate on tiny grains moving at speed, and small enough to mount on the read/write head. In time Seagate will make its own lasers: "We've felt the need to develop it in-house as well, to complement the external partner that we have."

Price and space

As to the price, he says: "The overall cost per terabyte is still lower than the previous generations, I will say that there is no premium - in fact, you get TCO benefits by

moving into the new platform versus the older platform on a cost per Tbyte basis."

The first qualification units are already shipping and Teh hopes to get major customers buying it in quantity by the end of Q1 2024. The first customers will be large cloud service providers, but eventually lower-end drives will have HAMR.

"We're coming up with variations of the drive," he says. "Today, if you were to buy a 12TB hard disk, you probably need one with at least six platters inside. In the future we're coming out with variants drive with fewer platters, so you could conceivably get 12TB with five disks or less, which actually makes it a lower cost product."

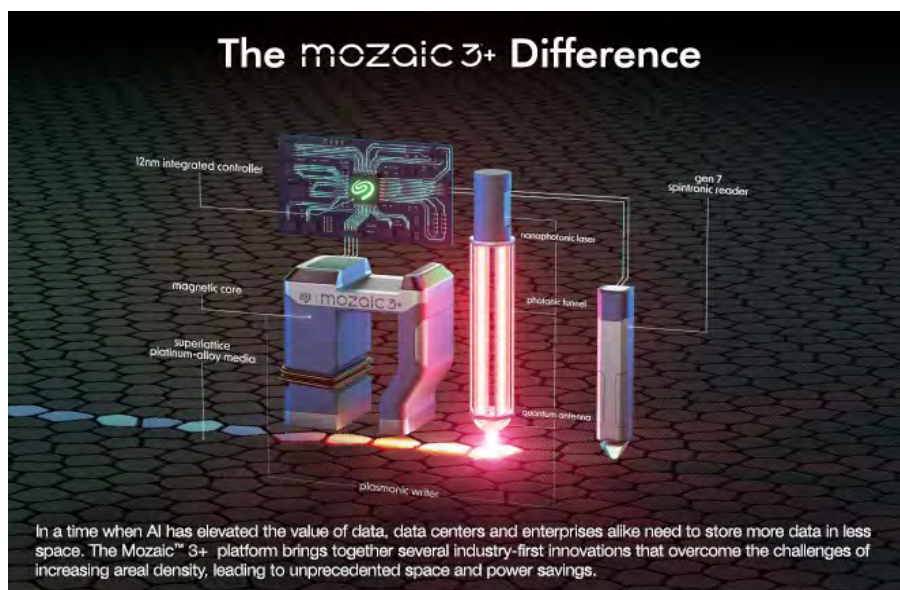
Other vendors will follow with HAMR, he says: "We certainly hope they will, because it's just going to be beneficial for the industry."

For now, though: "We are the only company that has this technology commercially available. We're the first to be able to announce and ship a three-terabyte hard disk product. And the good news is that we've got follow-on products to grow to four terabytes per disk and five terabyte products as well.

"This not just a one-trick pony. It's a new platform, an inflection point that will take us to bigger and better places."

Eye on the competition

Solid-state drives continue to expand - and their presence clearly has had a bigger role



in the declining market for HDDs than Teh likes to admit.

"It's our view is that all the devices that are in mass production today have a role to play," he says, adding in tape to the array of options. "It's all complimentary. We don't see a scenario where one technology takes over the other completely."

HAMR will let HDDs keep their price advantage he says: "Solid State is at least five times more expensive. If you're a cloud service provider, and you are selling gigabytes for dollars, you can't pay five

times more per gigabyte."

Considering the possibility that all HDDs would be replaced by solid-state drives, he says: "They're going to have to invest over \$200 billion in capex just to have enough capacity - and consider the amount of time to actually get the fabs up. No, the industry doesn't have the money to invest \$200 billion.

"Over time, yes, solid state will have a higher percentage of the share overall. But no, it's not going to replace hard drives for a long, long time." ■

HARD DRIVE HISTORY

Hard disks drives (HDDs) have been around for nearly 70 years, and have come a long way since the first units invented by IBM, which were the size of "two refrigerators" and stored less than 4 Mbyte on 52 removable platters.

IBM followed up with the "Winchester" drive in 1973, which set the pattern for the next 50 years, with non-removable media, and ever increasing densities that boosted storage. At the same time the size of the platters shrank from 14in, to 8in, and then on to 5¼ in and 3½ in units which became standard in PCs during the 1980s.

The platters are made of glass, alloy or ceramic material, coated with a magnetic layer 10,000 times thinner than a piece of paper. In that layer are grains of material which can be magnetized, storing data. The platters spin, and the magnetic recording heads "fly" over them,

magnetizing individual grains to record bits.

During this time, storage densities on the platters went from 2000 bits per square inch in 1957, to densities above 1.4 Tbits per square inch - an improvement of 700 million to one. Prices fell, by a factor of about seven billion, starting at more than \$9,000 per Mbyte and falling to less than \$15 per Tbyte.

Despite these improvements, HDDs have been under pressure from solid-state drives (SSDs) using technologies including NAND, which have higher data-transfer rates, higher storage density, and much lower latency and access time.

In response, HDD makers led by Seagate, Western Digital, and Toshiba, have come up with a series of improvements that have kept HDD price

per Tbyte lower than the competition - essentially finding any way to make the size of the grains smaller, while keeping their magnetization stable.

From 2005, hard drives began using perpendicular recording. Magnetic domains pointing downward took less space. Since then, Western Digital introduced the idea of filling the drive with helium, which reduced friction and turbulence, allowing more platters in the drive, and smaller recording tracks, so more data can be packed on each one.

Meanwhile, the number of platters in a drive has increased to 12 in high-end models.

Right now, the next step is HAMR (heat assisted magnetic recording) but beyond that, who knows what other techniques might keep hard drives in play. ■



Strategic data center partner

Expert delivery. Global Scale. Local precision.

Equans Data Centers is a strategic data center partner, delivering new build, fit-out, retrofit, and upgrade covering general contracting for hyperscaler and colocation requirements.

Bringing construction and engineering, global standardisation, and precision localisation, we deliver the highest quality, speed, and scalability, meeting increasing capacity needs across Europe.



Find out more

equansdatacenters.com

Kestler's run



Sebastian Moss
Editor-in-Chief

EdgeCore's CEO on building a hyperscale wholesaler for the mega data center era

EdgeCore CEO Lee Kestler wouldn't want you to think his company is exciting.

"We're not an innovation company, some pretend to be. We are not. We are a real estate developer. We are very focused, we're deliberate and intentional with what we're doing. We have very ruthless prioritization."

EdgeCore was founded in 2018 with a plan to build wholesale data centers for the largest data center companies. "We're looking to acquire land parcels and operate at the intersection of energy and networks," Kestler said.





Lee Kestler
CEO, EdgeCore

"When we started DuPont Fabros, that was the first time you ever saw something more than 30MW in a building. Everybody thought we were nuts. This industry pushes itself to a point where success is always teetering with failure"

The company has styled itself after DuPont Fabros, which was acquired a year before EdgeCore's founding by Digital Realty for \$8 billion - ending a decade-long run as a huge wholesale data center operator.

In late 2022, EdgeCore was itself acquired by Swiss investment firm Partners Group with a plan to double down on that strategy, backed by a \$1.2 billion war chest. The operator has also secured more than \$2bn in green loans to fund its build-out.

Kestler was brought in alongside the Partners acquisition with a clear mandate to build his own DuPont - something he had prior form for, after spending more than eight years at that company.

"When I joined DuPont Fabros, we were one of the early pioneers of wholesale, large-scale, energy-efficient data centers," he recalled. "And that turned out very well for us. The industry has just exploded since."

Between his lengthy run at DuPont, and his past year at EdgeCore, Kestler had two other roles that he believes will deeply inform how he hopes to build the company.

The first was at Vantage, where - as chief commercial officer - he helped the data center firm grow with private equity backing. "It was in a very similar spot to where EdgeCore is now," he said.

The second job may end up proving more useful for the inflection point that the sector now faces. "When I left Vantage [in 2021], I intentionally retired from the data center sector," he said.

Instead, Kestler moved over to the utility space, spending a year at energy developer

NextEra. "I now understand renewables, hydrogen, battery storage, and a lot about EV charging," he said. "I understand what it means to sit with somebody at the utility level, and I can understand exactly what's going on."

This, he hopes, will give his company an edge in finding reliable power at a time when the grid is increasingly constrained and data centers are increasingly power-hungry. "We are going to be 300MVA or bigger. It's not like you need 30MW anymore."

To adapt to the bigger and denser realities of today, EdgeCore has rolled out a new data center design based on 12MW blocks and 36MW floor plates. "They're pretty capable, and are still using air-cooled chillers," he said. "We are not going to use a lot of water. We're not developing evaporative chilled water plants in our designs."

The market's demands for liquid cooling may change, however, so EdgeCore is pre-installing support at the mechanical level so that it is easier to upgrade in the future.

While the company is looking for places to support large-scale developments, it will build at those sites in increments. "I could build a 144MW building across four stories, but I may only take up two floors to start," he said. "It may be seven or eight years before the rest of it gets filled. And that's okay."

Hyperscalers are already talking about infrastructure eight years ahead, he said, making planning for the long run in a period of AI-fueled potentially short-term excess possible.

"We think we understand where the puck is going," Kestler said. "This industry is not short for self promotion on what's happening. I mean, come on, let's get real. Everybody talks about Ohio and Salt Lake in the Pacific Northwest. Atlanta is getting a lot of attention, and Texas."

But "there's not enough money for any of us to go out and buy land at all those locations. So you pick your spot, and then you work to monetize the vision."

Over the next decade, Kestler believes that "a wave of these 300+ MVA type deployments are necessary because of the constraints in the utility market, the needs for future planning of that capacity," he explained.

"And it's proven - when we started DuPont Fabros, that was the first time you ever saw something more than 30MW in a building. Everybody thought we were nuts. This industry pushes itself to a point where success is always teetering with failure."

The market is now much larger than the days of DuPont, which means a much greater opportunity for success. "We will have three to five new markets in the next three years," Kestler said, "with 300MVA plus campuses."

"The market is not going to go away; AI is not a sugar rush. The development cycle for all of these types of projects is 36 to 48 months out."

At the same time, hyperscalers are ramping up their own data center build-outs. "But they need help - and they've said that. So I feel good about it." ■



How to harden critical infrastructure

Use Roxtec cable and pipe seals

[Learn more](#)



Grundfos Data Center Solutions

Keep your cool

Efficient water solutions
for effective data flow

Meet your efficiency and redundancy goals

Smart pumps offering up to IE5 efficiency, redundant solutions that meet up to 2N+1 redundancy – whether you're planning a Tier 1 or Tier 4 data center, you can rely on Grundfos to keep your data center servers cool with a 75-year history of innovation and sustainability at the core of our corporate strategy.

Your benefits:

- High-efficiency cooling solutions saving water and energy
- Redundancy meeting up to Tier 4 requirements
- End-to-end partnership process

GRUNDFOS 

Possibility in every drop

Betting on the US market



Georgia Butler
Reporter

The digital infrastructure holding up online gambling

If you have ever been to a Las Vegas casino, it's likely you will remember the experience vividly. The lack of windows, and clocks. The sound of coins jingling, of slot machines, of cries of delight and angry frustration. If you aren't a gambler, you may have marveled at the sheer lack of seats available for anything other than betting or playing.

Casinos are an experience all of their own, like a candy store where the sugar high comes from the constant flow of money rushing to and from - let's face it, mostly to - the house. The chance, however small, of myriad financial problems being solved by a single spin of the roulette wheel means gambling remains an enticing prospect for many, so it is no wonder that the sector seems endlessly popular, and this popularity is spreading to the online, virtual world.



This happened earlier in Europe than in the US but slowly, state by state, online gambling - also known as iGaming or simply 'gaming' by those in the industry - has opened up on the other side of the Atlantic too.

As with anything based on the Internet, iGaming needs to be hosted somewhere, and that appears to be where significant complexity has arisen.

Continent 8 Technologies is one company providing the digital infrastructure for the gaming industry. It offers colocation and cloud services specifically to the gaming and online gambling markets. "It's a credit to our founder, Michael Tobin," says Justin Cosnett, chief product officer at Continent 8, of the business's move into the US market.



Justin Cosnett
Head of Solution Architecture

"The transactions have to happen inside an actual casino, so we haven't yet managed to get to a position where we could take a data center or even build a data center in a casino property"



"Michael always had the ambition to service the US market once it was regulated. He took the incredibly unusual step of getting a casino license in Atlantic City to service New Jersey and then build a data center in Atlantic City."

From there, Continent 8 has been opening colocation sites as each state legalizes the online gaming industry.

"We do data centers where other people wouldn't, because lots of other data center providers are looking for cheap power and great connectivity. We don't actually need to be in those places, we go where the vertical takes us and locations are needed for regulatory purposes."

This technique of keeping a keen eye on where new locations are opening up is also being followed by competitor Internet Vikings.

Rickard Vikström, founder and CEO of Internet Vikings, told DCD that, similarly, the company is attempting to offer its services in every state applicable.

Moving into the US

Internet Vikings is based in Sweden, and along with Continent 8 began its operations

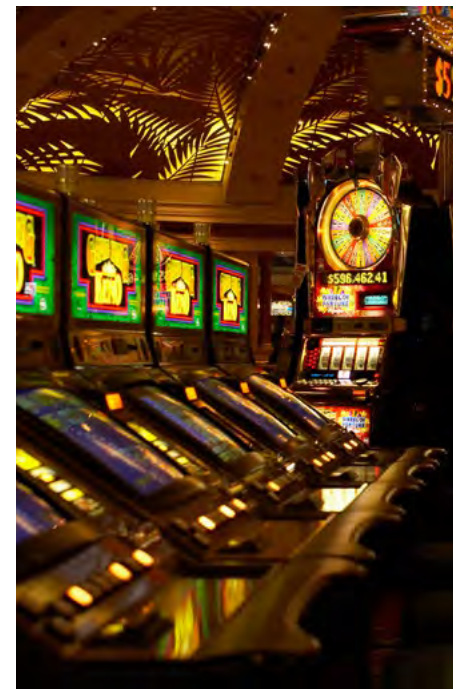
in Europe as the market is, according to Vikström, 10-15 years ahead of the US in terms of regulations.

This is not to say that gambling was not popular in the US, but that a variety of federal and state laws have held the sector back.

In 1961, the Interstate or Federal Wire Act was established, prohibiting "the transmission in interstate or foreign commerce of bets or wagers or information assisting in the placing of bets or wagers on any sporting event or contest" or a communication that enables the recipient to receive money or credit for bets or wagers via a wire communication facility.

"The Wire Act prevents you from placing bets on sporting events across state lines," explains Cosnett. "It's meant that each state has to regulate and have that transaction happening in the state - so even users on their mobile devices using a mobile app have to be in that state."

Beyond that, there was the Professional and Amateur Sports Protection Act (PASPA) of 1992 which effectively outlawed sports betting nationwide excluding a few states, and then the 2007 Unlawful Internet Gambling Enforcement Act which prohibited gambling businesses from "knowingly accepting payments in



connection with the participation of another person in a bet or wager that involves the use of the Internet and that is unlawful under any federal or state law."

In combination, these three laws make the US a hostile environment for online gambling, and sports betting - both major parts of iGaming

However, in 2018 PASPA was overturned, thus freeing up states to begin making their own choices.

This has created an extremely challenging market, with each state having its own regulations, and hosting service providers must get licensed to comply with every set of rules.

"It depends on which state you are in. It's so fragmented because every state has its own rule book so it's impossible to just know 'In this state, you do this or do that'," explains Vikström.

"I've had to send in fingerprints and tax

returns for the last 10 years, I've had interviews with investigators. Everything in my life has been gone through. Then they need to check the servers and our company to make sure we are a good supplier, and also the shareholders behind us."

These regulations slow down the process of business expansion for the likes of Internet Vikings and Continent 8, but they have both now made the process as efficient as possible.

According to Cosnett, Continent 8 was, at one point, opening new locations once every two months over a period of six months as new states added wiggle room for iGaming activities.

"There's no guarantee that every state is going to regulate, but all of our customers that want to come to the US want to be first to market, so we've had to compress our ability to open up a new location to between two and six weeks we can be ready for customers to start installing their equipment," explains Cosnett.

Currently, Continent 8 has been leasing space in other colocation facilities - always at a Tier III quality and with good connectivity and capacity availability - but in the future, the company is looking to acquire and own its premises.

Not only do the likes of Internet Vikings and Continent 8 need to be licensed, but in some states the colocation facility they operate from will need licensing. "That is a bit of a competitive differentiator," argues Cosnett. "We might look like we're just buying or leasing someone else's data center, but we're also bringing that regulatory approval to operate and provide our customers as well as the network and cybersecurity."

But even beyond the complex regulatory landscape, iGaming as a sector has a key set of needs.

Speed and safety

According to Vikström, iGaming is very similar as a sector to e-commerce in terms of its IT needs.

"It's more about security, more about making sure it's always online, rather than a lot of capacity needed because, at the end of the day, it's not that much data," says Vikström.

Cosnett reaffirms this, noting that Continent 8 has been offering services including DDoS (distributed denial of service) protection, WAF (Web App Firewall), and has added a SOC and SIM service and end-point protection.

"It's a rich target for cyber attacks," says Cosnett. "We've seen that in the press in the land-based casino world over the last year or so." While he does not explicitly name any particular incident, one heavily reported on was in September 2023, when a cyber attack



against MGM Resorts brought down the casino and hotel group's slot machines and hotel room key systems in Las Vegas, as well as its website.

Cosnett continues: "It is just as viable a threat to an online business as it is to a land-based casino. Almost more so because all these guys are doing is servicing the bet, so every second that a site is down or unable to service its customers, significant potential revenue is lost. And that's not just for the gaming company, but for the regulator in tax revenue as well."

Beyond security implications, latency is a key consideration.

"Online sports book, in particular, have a competitive advantage by having low latency connectivity and being able to deliver the user the latest possible betting odds and information about a game, particularly with options like cashing out," says Cosnett.

Because of this, Continent 8 drags its multiprotocol label-switching (MPLS) network to all of its locations to try and speed things up. Label switching can be faster than a routing table lookup because switching can take place directly within the switched fabric and avoids CPU and software involvement. Depending on the data center in question and its location, which is sometimes impacted by the state it is based in, this could be even more important.

State-by-state process

The debate in each state as to whether they should legalize iGaming is a complicated one. Currently, iGaming is only fully legal in seven states, though several more allow online sport betting and other limited gambling services.

A major driver in favor of iGaming is, predictably, money. Once something is regulated, it can be taxed, and such a popular sector has the potential to bring in a lot of tax revenue.

"There are few industries in the world that

tell a regulator or a state or government, 'please regulate us, please take our tax money,'" says Vikström.

New York is one market the vendors are studying closely. Late last year it was revealed that it is expecting its budget deficit to reach \$4.3 billion. Senator Joseph Addabbo proposed legislation to authorize iGaming and iLottery in New York, arguing that it could produce \$1bn in tax revenue for the state. Despite this, the state has not included the sector in its 2024 Senate Budget Proposal.

While iGaming as a whole is not allowed in New York, online sports betting is, though only from a few select locations - the casino resorts.

"The transactions have to happen inside an actual casino, so we haven't yet managed to get to a position where we could take a data center or even build a data center in a casino property," explains Cosnett.

Instead, Continent 8 intends to provide casinos that already have their own servers with better connectivity. That could help them connect with iGamers elsewhere in the state, and also potentially the hyperscale cloud providers as many online gaming platforms use cloud services.

According to Vikström, New York currently has four licensed locations. "There are a lot of rules, and it costs a lot of money to set up commercial operations there," he says.

These casinos do not have a secret and vast colocation data center in their basements for this purpose - Vikström suggests it is closer to "a few racks than a big commercial data center." As with anything with limited supply, these casinos are able to charge a premium for hosting there.

"At the end of the day, we [Internet Vikings] are not paying for it, it's the operators. We forward that cost to them, and they then forward it to the player," says Vikström.

DCD reached out to the licensed casinos in the state, but did not receive any comment.

Notably in New York, those casinos that

are licensed for sports betting exclude any on Native American land, in line with the Indian Gaming Regulatory Act. A Native American casino operator that asked not to be named confirmed this limitation to *DCD*, adding that some are in negotiations with the state for a new Compact - an agreement with the state that Indian Gaming sites need to operate - that would enable tribal casinos to take part.

The tribal casinos are, in many states, a motivation to not allow iGaming, according to Vikström.

"The reason states don't want to legalize it is the Native American tribes," he says. "Tribal gaming and casinos have a big influence politically, and online casinos would cause them to lose a lot of money."

"The government would then have to give something back to the Native American facilities."

The history connecting Native Americans with casinos and gambling is complex and rife with colonialism, dating back to the invasion of the US by European countries in the 1500s which led to conflict with tribes and saw many Native Americans displaced as their land was annexed.

Centuries later, Native American reservations were established by treaties and executive orders, returning some of that land that should never have been stolen in the first place.

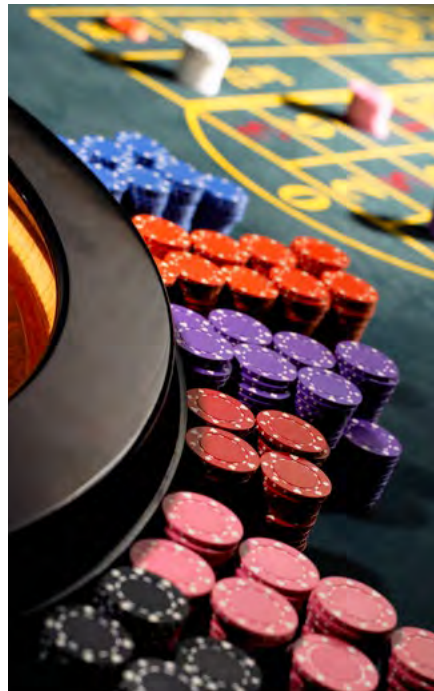
Inequalities prevailed, though, with federal and state government, regulations and grant funding' often leaving the reservations with few and limiting their ability to empower themselves.

In the late 1970s and early 1980s, many Native American tribal governments began establishing casinos and gambling locations which brought in significant revenue to the reservations. In 1988, the Indian Gaming Regulatory Act (IGRA) was established, which is a federal law that dictates the jurisdictional framework governing gaming sites on tribal land.

According to the act, the purpose of it is to protect gaming as a means of generating revenue for the tribes and to boost economic development. But the IGRA states that all Indian gaming must take place on native land, which is problematic for iGaming, which is mobile by nature.

An example of such can be found in the 2018 court case of the State of California vs Iipay Nation of Santa Ysabel. The federally recognized Indian tribe tried to "revitalize its gaming revenue stream" by establishing a served-based bingo game over the Internet. Those servers were located in Iipay's casino on tribal lands which has since closed down.

The casino had an on-site staff member who was supposed to act as a proxy for the patron by placing bets. While the proxy and



"The reason states don't want to legalize it is the Native American tribes - tribal gaming and casinos have a big influence politically, and online casinos would cause them to lose a lot of money"

the server remained within tribal lands, judges ruled this was in contravention of the act as the patron would not necessarily be within tribal land. Thus, the Iipay Nation was not allowed to continue its operations of the online bingo game.

DCD contacted several tribal gaming establishments and gaming regulatory boards for comment, but none were willing to do so.

The ethical dilemma

Another important layer for digital infrastructure providers in the sector is the ethics of gambling itself. The phrase "the house always wins" is well-known, and not inaccurate, because the business model of casinos relies on them winning more money than they lose.

At every level of iGaming, each service provider, be it the gambling platform or the IT hosting company, relies on the customer continuing to place bets that will end in failure.

Cosnett is philosophical about this. "Everyone will have their own personal view," he says. "Continent 8 isn't a company that is full of gamblers. We are a technology provider, rather than a gambling company ourselves. My normal response is that what we've seen is that the best regulation and regulated companies will offer the best consumer protection."

Continent 8 remains in favor of regulation, with Cosnett saying that, if asked by regulators, they will switch customer's operations off.

"Total prohibition has been tried in certain environments, and it's not necessarily successful," he adds. "People respond accordingly, they carry on doing it but in ways that aren't necessarily the most suitable to the providers or the consumer."

This philosophy is further supported by Internet Vikings' Vikström.

"For me, it's easy. It needs to be regulated, licensed, and controlled by someone because it's an industry that can come with a lot of implications, like alcohol, tobacco, or anything like that. People will always do it, so it's much better that it's regulated," he says.

Regulations mean you have to prove things such as source of funds, Vikström says. "If you want to bet with \$100,000, you need to prove that it is legal money and that you actually have an income that can support that kind of gambling."

"In unregulated gaming, they will try and squeeze as much money out of every single person as possible."

Indeed, besides Nevada, which is home to Las Vegas, there does not seem to be a clear link between a state's "friendliness" to gambling and gambling addiction statistics - the results are mostly scattered, but the data itself is also hard to establish. It relies on self-reporting, which many gambling addicts would not do.

The US is unlikely to see a widespread opening up to iGaming. Cosnett puts this down to politics: "At a federal level, trying to get approval has been almost impossible," he says.

"We [Continent 8] view it as part of the entertainment industry but, morally, you can have different views on gambling, and no one is going to win votes by legalizing and or making gambling easier."

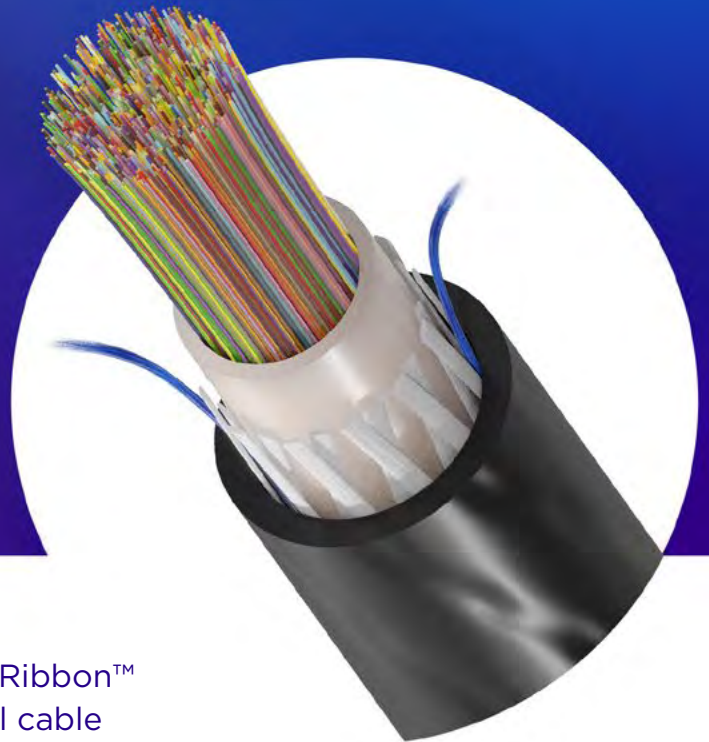
While the political in some states means legalization of iGaming is unlikely to ever happen, there does seem to be a continuous trickle of states changing their policies, with Rhode Island among those joining the legalization list this year.

Be it for better or worse, demand for iGaming is unlikely to go away. ■

NEXT GENERATION THINKING™

MORE FIBER IN LESS SPACE

WITH FREEFORM
RIBBON™
TECHNOLOGY



Sumitomo Electric Lightwave's Freeform Ribbon™ allows for dense fiber packing and a small cable diameter with a non-preferential bend axis thereby increasing density in space-constrained applications.

Sumitomo Electric's patented pliable Freeform Ribbon™ construction is designed to both pack densely in small form factor cables while still being capable to transform quickly, by hand ONLY, to splice-ready form similar to standard ribbon for fast and easy 12ct ribbon splicing (for both in-line and fusion splice-on connector splicing applications).

How do you solve a problem like Atos?



Matthew Gooding
Features Editor

Plagued by debt and bad decisions, is there any way back for the French IT giant?

Given that he built a reputation as France's premier *magicien* when it comes to turning around failing businesses, it is perhaps no surprise that Thierry Breton is not keen to be associated with the problems that have beset one of his former companies, Atos.

Breton, now European Commissioner for the internal market, made his name in the 1990s by successfully waving his magic wand over a trio of struggling state-backed French organizations - national computer maker Groupe Bull, consumer electronics vendor Thomson, and telco France Telecom - and steering them back to profitability.

For a long time, it appeared Breton had also pulled a rabbit out of his hat at Atos. Taking over as CEO in 2009, he helped grow the company into a business of national strategic importance in France, providing digital infrastructure and services to clients across Europe and beyond. Annual revenue at Atos, which was €5.6bn (\$6bn) in 2009, had doubled to €11.5bn (\$12.5bn) by the time Breton departed for Brussels a decade later.

But since 2019, life at Atos has more closely resembled one of the rides at another of Breton's projects, Paris theme park Futuroscope (he was part of the team that originally designed the park in 1986). The company's share price has dipped more sharply than the scariest of rollercoasters amid a series of

strategic blunders and a mounting pile of debt, with its market cap falling from €8.2 billion (\$8.8bn) at the end of 2020 to just over €190 million (\$312m) at the time of writing.

The company's financial results for 2023, released at the end of March did not make for pretty reading, showing a record annual loss of €3.44 billion (\$3.73bn), up from €1bn (\$1.08bn) in 2022.



Distancing himself from Atos's struggles, Breton reportedly told members of France's financial press association at a meeting in March that he has "no responsibility, zero," for Atos's travails, saying they started after his departure. Bloomberg reported that the 69-year-old went even further, with his team issuing a three-page document outlining the wrong turns taken by the company, including the names of the CEOs in charge at the time.

Wherever the finger of blame lays - and for several politicians and shareholders, it is firmly pointed in Breton's direction for laying shaky foundations that have caused the current problems - it is clear that Atos is facing a potentially existential threat. And even if it does survive, one of France's biggest technology companies will never be the same again.

Cracks in the (Tech) Foundations

Atos was founded in 1997 when two French IT businesses, Axime and Sligos, merged. In 2000 it joined forces with Dutch tech vendor Origin to form Atos Origin (the Origin part of the name was eventually dropped in 2011) and began a period of rapid growth driven by a series of acquisitions.

It snapped up the UK and Netherlands consulting arm of KPMG in 2002, and two years later acquired SchlumbergerSema, Schlumberger's IT infrastructure division.

Following Breton's appointment, Atos got even bolder, paying €850 million (\$925m) for Siemens' IT business, and expanding into the US through the purchases of Xerox IT outsourcing (€918 million) and a similar company Syntel, for which it paid €3.1bn (\$3.5bn) in 2018. Along the way, it added one of Breton's former companies, Bull and its supercomputing expertise, for good measure.

The result was a business that, at its 2017 peak, was deriving the majority of its income, some €7 billion (\$7.6bn) annually, managing digital infrastructure. It operates five of its own data centers according to Data Center Platform, three in France and one each in Austria and Germany.

So where did it all go wrong? "With these companies, it can often take a long time for the rot to set in and for it to be visible when the problems were already there," says James Preece, principal analyst at Megabuyte. "Atos is a long-standing business and in my mind just made a few strategic missteps.

"They are probably perceived as a legacy service provider and were slower than others to transition into the world of cloud computing and digital services. On top of that, they've paid a lot for assets that they've struggled to get value from."

Indeed, with such high exposure to managed infrastructure, it is perhaps understandable that Atos was hit harder than most by the advent of cloud computing and the willingness of businesses to embrace IT delivered as a service via the hyperscale public cloud platforms.

Atos did not help itself with an aborted attempt to buy DXC, a company similarly invested in outdated on-premises systems. It saw a \$10bn take-over bid rejected by the US vendor in 2021 and walked away from the deal, but not before its share price had taken a battering as puzzled investors questioned the value of adding another legacy business to its portfolio. Later that year, \$1 billion was wiped off the company's value after accounting errors were discovered.

Breton has pointed to the fact that Atos was one of the first companies to partner with Google Cloud to offer private cloud services as evidence that, under his stewardship, Atos was heading in the right direction. But by summer 2021 the company was in the doldrums and, with revenue flatlining, its directors decided to take radical action, announcing it would slice its business in two, with its legacy managed infrastructure business renamed Tech Foundations, and cloud, cybersecurity and big data divisions rebranded as Eviden. Both companies, it was said at the time, would be publicly traded and operate separately under the Atos umbrella.



Thierry Breton

"Atos needs to leverage what they're good at, which is the HPC business, and try and sell HPC-as-a-service, packaging that for the new AI world,"

>>Beatriz Valle
GlobalData

The business split that never was

The planned split was a sensible idea says Christopher Wiles, senior director analyst at Gartner. "The services they put into Tech Foundations are those focused on the cost, delivered via automation, process efficiencies, and offshore delivery models like data center and workplace," Wiles says. "They have put what I'd call the more value-focused technologies, supporting cloud and cloud migration, big data, data analytics and cybersecurity, into Eviden.

"I can understand why they split it that way, there are a lot of parallels with the IBM and Kyndryl separation though not necessarily their reasons for doing it."

Kyndryl is the ITSP formed when IBM spun off its managed services division in 2021. After struggling to make in-roads initially, its share price has been steadily climbing in recent months on the back of a series of contract wins and partnerships signed with the hyperscale cloud providers.

Any thoughts that Tech Foundations might do similar have been quashed by the fact that, well, the split hasn't happened yet. Despite now operating as separate brands, Eviden and

Tech Foundations remain two sides of the same coin and Atos has spent much of the last three years struggling to come to terms with its debt, which has mounted as income fell. The company has loans of €3.65 billion (\$3.9bn) maturing in 2025, and has been forced to consider drastic action. This included potentially selling part of its prized BDS business unit, a part of Eviden which covers cybersecurity, to Airbus in a deal that would've brought in €1.8 billion (\$2bn). However, the companies announced in March that talks between them had stalled and the deal was off.

Tech Foundations is also up for sale, and last year, a deal was agreed to sell the business in its entirety to EPEI, an investment fund run by Czech billionaire Daniel Kretinskiy. Under the agreement, Atos would have received €100 million (\$108m) and unburdened itself of debt of €1.8bn. Kretinskiy also promised to participate in a €900 million (\$980m) share issue to raise funds for Eviden.

But in November Atos revealed it was seeking to renegotiate its deal with EPEI, apparently to try and squeeze a few more Euros out of Kretinskiy. This did not go well, and in February the two companies said the take-over was off, with the Eviden fundraising canceled too.

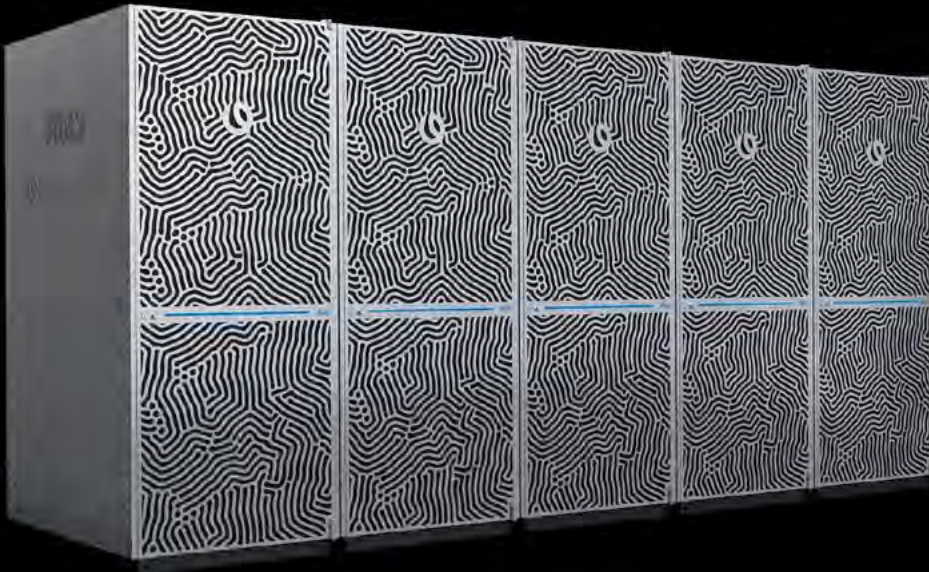
Churn in the boardroom probably hasn't helped. Atos is onto its fourth CEO in three years, with current incumbent Paul Saleh taking on the job in January, replacing Yves Bernaert, who himself had only been in post since October 2023. Bernaert said his departure was driven by a "difference of opinion on the governance to adjust and execute the strategy." Atos chairman Bernard Meunier, one of the prime movers behind the split plan, also left the company last year.

Megabuyte's Preece said dividing the business was "a massive undertaking" for Atos. "If you announce something like that, you need to make sure you can make it happen because you've told the world about it," he says. "If it doesn't go exactly to plan, then you're going to find yourself in a world of pain."

Macroeconomic factors outside the company's control haven't helped, Preece adds. "Market conditions significantly deteriorated after that announcement, through no fault of their own," he explains. "So they tried to do all this logistical work around creating two separate entities and manage an underperforming business, at the same time as the markets were collapsing around them."

What does the future hold for Atos?

With Kretinskiy seemingly out of the picture, Atos has appointed a *mandataire ad hoc*, or



independent broker, to negotiate with its banks in a bid to refinance some of its debt and buy the company some much-needed breathing space.

Speaking to investors after its 2023 results were announced, Saleh said the company is "in discussions with our financial creditors with a view to reaching a refinancing plan by July within the framework of an amicable conciliation procedure."

Digital consultancy OnePoint, which holds 11 percent of Atos shares, has offered to step in with additional funds if necessary according to its CEO David Layani. The Atos board says it has yet to consider or approve such an intervention.

The French government is also closely monitoring the situation, with 10,000 Atos staff based in France. The company also holds a string of public sector contracts; it is a major IT supplier to the French armed forces, and its infrastructure underpins France's nuclear industry through its Bull supercomputing subsidiary.

It is also a technology partner of the International Olympic Committee, meaning it will be responsible for providing much of the digital infrastructure underpinning the 2024 Summer Olympic Games, being held in Paris.

Though Bruno Le Maire, France's minister for the economy and finance, has rejected calls for the state to take over the company, he told *Les Echos* that the government would "use all the means at its disposal to preserve the strategic activities of Atos." It is thought this will mainly involve discouraging foreign buyers from making a move for Eviden or Tech Foundations, though there are unlikely to be many takers for the latter.

For Beatriz Valle, senior analyst at

GlobalData, disposing of Tech Foundations - or making some potentially painful decisions - will be key to Atos forging a profitable future. "The problem with the [Tech Foundations] part of the business is that nobody is bidding against Atos for those legacy outsourcing contracts because they don't make any money," Valle says. If Atos ends up retaining Tech Foundations, she adds, it may be forced to cut a large number of jobs to reduce costs.

Valle believes Atos can flourish providing it trades off its strengths in areas like high-performance computing. It has scored a string of contracts with the EuroHPC joint undertaking, the European Union's supercomputing project, and two machines backed by EuroHPC and based on Eviden's BullSequana XH3000 direct liquid-cooled architecture - the Leonardo supercomputer in Italy and the MareNostrum system at the Barcelona Supercomputing System - featured at positions five and eight respectively in the latest Top500 list of the world's fastest supercomputers.

Eviden has also been awarded a €500 million (\$525m) contract to develop Jupiter, Europe's first exascale machine, which will be based in Germany.

"Atos needs to leverage what they're good at, which is the HPC business, and try and sell HPC-as-a-service, packaging that for the new AI world," Valle says. "They don't have the expertise that a company like IBM has in AI, but they could do something like that through partnerships, and leverage the compliance knowledge they have to offer a 'sovereign cloud' product for European companies."

Gartner's Wiles also believes Atos remains "a strong player in infrastructure services," and says "those capabilities haven't changed." He

"If it doesn't go exactly to plan, then you're going to find yourself in a world of pain"

>>James Preece
Megabuyte

explains: "It's not necessarily whether they've got the capabilities that people are worried about, it's the ramifications of their financial situation."

Wiles says businesses he has spoken to that use Atos products and services are "looking at risk mitigation strategies" in case it goes into debt protection measures equivalent to the Chapter 11 bankruptcy process in the US. "The concern level of our clients [about Atos] has been raised since the turn of the year," Wiles says. "Nobody is acting yet, but people are putting plans in place."

This illustrates the precarious position the company finds itself in, and Megabuyte's Preece says any further missteps could have a potentially devastating impact. "Atos, in its entirety, is burning cash at north of €1bn a year and it has €5bn of debt," he says.

"You can see the writing is on the wall - they are running out of cash, and they need either flexibility from lenders on refinancing or a big transaction to cover a significant chunk of the debt they need to pay off in the next year. They're fighting a lot of fires." ■

Low Carbon Footprint for Your Data Centre: Geniox Tera F

At Systemair, we offer a range of reliable, highly energy-efficient cooling solutions that work at their best all year round with minimal impact on the environment. So you can confidently meet local, national and international energy efficiency standards without compromising the performance of your data centre.



LCA
Life Cycle
Assessment

Running to the storm: How Verizon prepares for natural disasters



Paul Lipscombe
Telecoms Editor

While most run away from storms, network operators have no such option, their networks are critical infrastructure for those in desperate need of help

In August 2005, New Orleans was on the receiving end of one of the worst-ever hurricanes to ever hit the United States.

The Category 5 hurricane killed 1,836 people in the country, across seven states and caused an estimated \$125 billion in damage - the joint most in history.

It's sadly not uncommon for such disasters to hit the US, with hurricanes a regular occurrence each year, though each varies in strength and destruction.

In the event of any natural disaster, be it a hurricane, a tornado, or a wildfire, power will often shut off quickly and mobile networks will go down.

Getting these networks back up and running can be a matter of life and death, and the US carriers recognize this.

In February, DCD was invited by US carrier Verizon to view first-hand how telcos prepare for such disasters.

"We see crisis and change as opportunities, not threats. We run to a crisis, not away," the company often states.

"When natural disasters strike, our team is always ready because we've invested heavily and train relentlessly to be deployed at a moment's notice to assist with rapid recovery efforts of critical communications infrastructure," says Julie Slattery, senior vice president of core engineering & operations, Verizon.

Patriot 24 drill

Verizon paired with the US Military to carry out a training exercise in Hattiesburg, Mississippi. The location is pretty poignant, as it's only a couple of hours away from where Hurricane Katrina caused mass devastation.

Lasting for six days, the training drill saw the Verizon Response Team work alongside

the Air Force, Mississippi National Guard, and Mississippi Emergency Management Agency (MEMA).

Dubbed "Patriot 24," the exercise sought to simulate multiple realistic, high-intensity crisis scenarios, including a Category Four hurricane damaging critical communications at a hospital and a tornado causing a nuclear radiation release with network restoration needed in a hot zone.

The purpose of these drills is for preparation, says Verizon.

"All of these teams do training exercises to sharpen their skills and prepare themselves throughout the year," adds Mark Paff, director for business continuity & event management at Verizon.





"But this is the first opportunity where we had all of them together coordinating on a joint response. We're also doing it with the Air and Army National Guard so to the extent that we can train, learn from each other, and be better prepared for when a situation does occur, that's what makes this training opportunity so special."

Networks go offline

When natural disasters strike, they usually tend to knock out power, causing telcos to lose their network coverage, thus impacting the customers that rely on these networks.

It means that carriers such as Verizon need to be on high alert when the networks go down.

For example, in 2022, when Hurricane Ian hit Florida and South Carolina, Verizon, along with T-Mobile and AT&T, was on hand to restore telecommunication services to those affected.

During that occasion, Verizon deployed its Wireless Emergency Communication Centers (WECC) in areas of Southwestern Florida that were impacted by the storm. The WECC are generator-powered mobile units that have device charging and computer workstations, along with wireless phones, tablets, and other devices.

On this occasion, the telco had to complete site surveys of impaired cell sites, while moving portable generators, and mobile cell towers where needed.

These sorts of responses are the norm during these outages, but to get to this point, a serious amount of preparation is required.

Several teams took part during the drill. This consisted of the carrier's Major Emergency Response Incident Team (MERIT), Dedicated Impact Response Team (DIRT), Emergency Response Team (ERT), and Verizon's Frontline team.

MERIT focuses specifically on responding to hazmat conditions, while DIRT takes care of Verizon's Radio Access Network (RAN) infrastructure. The role of the ERT team is to keep on top of fiber infrastructure, while the Frontline squad works with local and public safety agencies.

The Frontline team is made up of many people who are first responders themselves and is focused on providing solutions to aid first responders at the offset of a disaster. Frontline often provides solutions for critical functions

such as search and rescue.

Verizon's approach to training

During the drill exercise, the various Verizon teams were tasked with a range of drills.

Paff explains that the drills are necessary for the various teams to competently and confidently respond when needed during a crisis.

"We drill all time time," he tells *DCD*. "We design our networks for resiliency. A major storm can sometimes have disruptions for that high resiliency, so we train all the time.

"We also do situational awareness. So my team does 24-7 365 monitoring globally of anything in the world that could disrupt communications, we are a global service provider so we have services in all the major countries in the world.

This means we're constantly monitoring for weather threats, natural disasters, or even man-made events, such as civil unrest, and things of that nature."

When a hurricane looks likely to occur, Paff says that Verizon will usually have advanced warning before it works on its checklists.

"So, we have three sets of checklists that we follow, with the first pre-incident checklist, these are the things that we're doing to prepare all of our response teams in advance of an oncoming threat.

"After this, we have our incident checklist which is after the event occurs, and when we're responding to it, we're making sure that all of the teams are following through with our checklist."

He notes that during this phase, Verizon needs to ensure it has people with the right skillsets in the right areas carrying out repairs. The final checks that Verizon carries out are post-incident, where the carrier ensures that everything is closed down.

Once the incident is finalized, Verizon will then hold after-action reviews and analyze what worked and what maybe didn't during its response efforts, adds Paff.

Support from the military

Also on hand at the drill, were the Air Force and the National Guard. The Military plays a pretty important role during natural disasters.

Although it's the job of carriers such as Verizon to get the network up and running in the wake of a disaster, the Army is also required

to lend its support during these humanitarian efforts.

The military supported the rescue efforts during Hurricane Katrina and will attempt to save lives and deliver supplies.

"So within the first 72 hours, we have our National Guard assets, and they can be on scene usually very quickly if they're in the local area," says Lt Colonel Ashley Nickloes, director of Exercise Patriot.

Nickloes says the National Guard carries out a drill once a year, to prepare troops for a real-life scenario.

It's no secret that military forces will communicate in a different way to the average person, and this could even be a barrier during a rescue mission, however, Nickloes says that they try to keep it as simple as possible.

"Air Force has a different terminology and acronyms that we use to what the army does," she says, noting that even this can lead to confusion.

"One of the things we've tried to do at Patriot is say, 'hey, we can't use acronyms here,' because we're working with our civilian partners. We want to make sure that when we're talking, everybody's talking on the same page, and from the same script so that we can maximize what we're doing, and that nobody feels intimidated. It's about breaking down those barriers."

Driving a response

Typically when a powerful storm hits, or a tornado sweeps through a settlement, cellular coverage is lost. This happens regularly during these disasters and was an important part of the simulation held during Patriot 24.

To support the response efforts, Verizon has a substantial fleet of vehicles on hand across the country.

The telco says it has more than 550 of these portable assets scattered nationwide in anticipation of disasters. The vehicles include fully functional, generator-powered cell sites that can replace or enhance network coverage where necessary.

It's hard to miss the vehicles too, which are heavily branded in red with signage including 'Verizon Response' on them. Verizon says the branding is deliberate, as it aims to bring a source of comfort to those in need of support during disasters.

One of Verizon's advancements in its fleet

in recent years is THOR (Tactical Humanitarian Operations Response), which is essentially a custom Ford F650 that can deploy a 5G mast and satellite dish as it responds to emergencies.

THOR runs on Verizon's ultra-wideband 5G network, which includes mmWave and Edge compute capabilities.

A job for the dog

One particular scenario that played out at Camp Shelby, saw a tornado cause a nuclear radiation release with network restoration needed in a hot zone.

The drill saw Verizon demo its Robotic Emergency Dog, otherwise known more affectionately as RED. But it's not just any dog, RED can provide a secure and reliable private 5G network for communication.

RED can be controlled by a handler, using a tablet-like device for navigation in between areas. Verizon says it has about seven of these robotic 5G canines.

The purpose of RED is for the robotic dog to go into potentially dangerous areas to scope situations. Verizon says it's easily deployable and can be used in remote locations and under harsh conditions.

During this particular drill, RED was tasked with finding various sources of radiation in the buildings.

With the chemical leak in full flow, the engineers in full hazmat were called in to carry out their tasks of clearing the area, along with some who would be splicing fiber cables to replace the damaged comms. To splice a fiber cable is a tough task at the best of times, but MERIT teams are trained to do this while wearing a full hazmat suit and thick gloves.

"This particular drill is satisfying some of the hazmat requirements that our MERIT team is required to take. Our MERIT team consists of about 40 engineers and technicians whose day job is to operate and run all aspects of Verizon's network," says Paff.

He adds that the MERIT team is also trained hazmat technicians, at the "same level that a fire department hazmat team would have to be." These members of the team are required to do 80 hours of hazmat training each year, plus one drill.

"They are trained to go into those very dangerous situations, we always say that we run to a crisis, and respond to bring communication services up which are so critical for our customers, communities, and first responders.

"They'll be in the protective gear is a specially designed suit with breathing apparatus that allows them to be solely fully self-contained.

"They can go into those hazardous situations and do the kinds of communicant

repairs that they need to bring our communication infrastructures back online, including splicing fiber optic cables that are damaged."

Repairing infrastructure

Unsurprisingly, when a hurricane or bad storm hits, cell towers are caught up in the chaos of it all.

If a cell tower is taken down it can be some time before it's safe to erect a replacement site or fix an existing damaged cell structure.

During an event, if this is to occur, Verizon's ERT team is tasked with handling the landline and fiber optic parts of the network. Because fiber optics connect Verizon's cell towers to the network, if they are damaged during a storm event, it's the job of the ERT team to restore them.

However, any fixes to the actual towers themselves are carried out by the DIRT team, explains Paff.

"Our DIRT team is focused on the wireless component of our network, and those are all the people that can restore and fix any damaged towers," he says.

He adds that they'll inspect the tower and base station on the ground to assess any damage to determine if it's from a power perspective or a backhaul fiber optic perspective, allowing Verizon to initiate the next step in its restoration process.

"There's satellite dishes that provide the backhaul antennas that can go up 60 feet, and we have portable towers that we can bring up even higher so they can quickly and temporarily restore communications."

Paff says that small cells can even be put into its drones so that they can fly up and propagate RF signals to a certain area.

To ensure that Verizon can bring temporary communications to people in need, it has a range of solutions, including its COWs (cells on wheels) and TOWs (towers on wheels).

If power isn't easily available, Verizon has other reserves in tow, adds Paff.

"We design our networks with multiple sources of power, with battery reserves at all of our sites, so if it loses power, we have generators in almost all of our major sites.

"We have national fueling agreements with major fuel suppliers to make sure that our generators stay fueled during a crisis event. We have a fleet of portable generators. So if everything fails we will strategically place these mobile generators so that we can quickly get out there and restore power."

Eye in the sky

To get the networks back up, a range of technology is used. Drones are particularly

useful, as they can scale heights that are otherwise impossible.

"Drones are a great asset for getting those eyes on the ground where it may be unsafe for people to go in there," says Paff.

Verizon has previously showcased its usage of drones, doing so during Hurricane Ian, where drones were deployed to provide cell coverage to responders and residents over an approximate five to seven-mile radius.

Drones have also been used to assess wildfires and floods in California and to help spot people trapped in ice in Colorado.

The telco says these drones can fly for up to 1,000 hours and can go as high as 400ft, providing it has the connection.

Significantly, the drones are also 5G-enabled and use Verizon's ultraband spectrum. The drones are operated by FAA (Federal Aviation Authority) certified handlers too.

"Drones are absolutely mission-critical pieces of equipment," states Chris Sanders, senior manager, Verizon Frontline crisis response team. "Our drones are able to provide live real-time streaming, both electro-optical and thermal imagery to our public safety agency via our robust network."

Preparing for the future

The overall purpose of the drills is to prepare for future disasters, but the learnings gained from the events and widespread collaboration are key, says Nickloes.

"The biggest takeaway from these exercises is from meeting partners and building relationships so that we know who we can depend on when we hit the ground. That's really what the exercise is all about."

She adds that technological advancements mean more can be learned from the drills.

While US operators typically compete with each other for market share, this mindset is forgotten during times of disaster, adds Paff.

"We have entered into relationships with other carriers so that if they are damaged and need services they can come to us," he says.

"So, if we've got the capacity and the equipment, we will help them because again, we're all supporting customers and communities and the first responders, so that the whole competitive thing goes by the wayside when you're dealing with a natural disaster or trying to bring communications back up."

With hurricane season fast approaching in the US, along with the threat of other unpredictable weather, the US carriers will likely put their skills into action once again, to provide the critical connectivity that millions rely on. ■

Cool runnings



Matthew Gooding
Features Editor

Oil companies like Castrol are looking to data center cooling as a new income stream

Situated atop Whitchurch Hill in the peaceful Oxfordshire countryside, Bozodown House offers a glimpse of bucolic Edwardian England. Initially constructed in the late 1800s, the manor house was completely rebuilt following a fire in 1907, and its red-brick frontage and imposing gables have remained largely unchanged ever since.

But while the exterior of Bozodown House is a window into the past, those working inside are looking to the future. Since 1976, it has been home to the Pangbourne Technology Centre operated by lubricant specialist Castrol. Best known for making the oil that flows through your car's engine, Castrol, now owned by BP, has big plans for Pangbourne as it looks to develop the next generation of fluids.

Given the automotive industry's rapid transition to electric vehicles (EVs), it is no surprise that a big focus of Castrol's work is creating the fluids required to keep EVs on the road. But it is also turning its expertise to the data center, and earlier this year took delivery of an immersion cooling pod. Delivered by modular data center provider Sonic Edge and featuring immersion cooling tanks from

another vendor, Submer, Castrol intends to use the 100kW 'MegaPod' to demonstrate the effectiveness of cooling fluids it has devised for servers. DCD scaled Whitchurch Hill - no mean feat given its surprisingly steep gradient - to take a look.

From electric vehicles to data centers

Recent years have seen liquid cooling techniques emerge as a realistic alternative to air cooling for data center operators looking to keep their hardware chilled.

Traditional air cooling systems often consume a large amount of energy and are becoming less effective as rack density increases. Added to this, high demand for AI products and services means data centers are being filled with more powerful components, such as GPUs, which generate more heat than their predecessors.

Liquid is better at dispersing heat than air, so liquid cooling techniques are seen as a possible solution for reducing emissions and potentially cutting energy bills. While take-up of these systems has so far been mixed -

Uptime's global survey of the sector last year found that 56 percent of data center operators were still using air cooling for their highest density, 40kW and above, cabinets - interest in the technologies is growing.

Castrol is, perhaps unsurprisingly, keen to capitalize on this demand, according to Nick Barrett, the company's senior engineer for advanced electrification testing. "Castrol sees a huge opportunity to diversify into industrial thermal solutions for data centers - a market that is large and growing fast - and has existing expertise in thermal management," Barrett says.

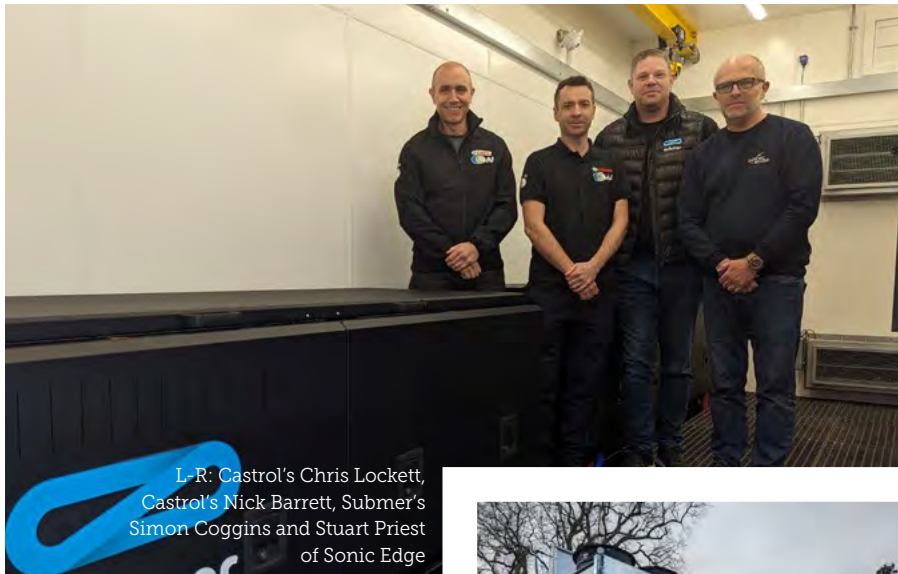
Indeed, thermal management is a key issue facing OEMs developing parts for electric vehicles, and as a result, Castrol is building labs at the Pangbourne Technology Centre where components such as batteries can be put through their paces. The megapod forms part of this array of testing facilities. "We've been working in partnership with OEM car makers for 125 years, and with electrification coming in their needs are changing," says Chris Lockett, Castrol VP for electrification and products innovation. "Part of that involves engineering cooling fluids for electric vehicles, and it makes sense to extend that work into the data center world."

Getting immersed in immersion cooling

The MegaPod, which from the outside resembles a humble shipping container, houses two 50kW Submer Megapod immersion tanks for immersion cooling, a type of liquid cooling which sees IT equipment submerged in liters of dielectric, or non-conducting, fluid.

Simon Coggins, Submer sales director for UK & Ireland, helped deliver the project





L-R: Castrol's Chris Lockett, Castrol's Nick Barrett, Submer's Simon Coggins and Stuart Priest of Sonic Edge

with colleague Franco Caroli, and told *DCD* the tanks are set up for single-phase immersion cooling, where the cooling fluid remains in a liquid state at all times. This differs from two-phase immersion cooling, which involves boiling the liquid in the tanks, causing heat to be dispersed as gas.

Explaining how the system works, Coggins said: "The immersion tank contains a cooling distribution unit, comprising a pump and a plate heat exchanger. The pump circulates the dielectric fluid within the tank, extracting heat from the servers."

The heated dielectric fluid then passes through the plate heat exchanger, where it is cooled by water. Single-phase immersion does not require airtight vessels because "hardly any coolant evaporates," Coggins says, and as such the tanks are known as "open baths."

Though single-phase immersion cooling is not as efficient as its two-phase counterpart, it may prove a more practical solution for the high-density data centers of the future, those running high-performance computing and AI workloads. This is because the cooling equipment required is simpler to install and run than for two-phase, and the fluids themselves often come with a lower price tag.

The long-term viability of two-phase immersion has also been called into question due to the environmental impact of the chemicals used in its cooling fluids. In late 2022 US manufacturer 3M announced it phase out poly-fluoroalkyl substance (PFAS) manufacturing from its product line by 2025, because of increasing concerns over their safety.

PFAS, or so-called "forever chemicals," do not break down when released into the environment, and thus can build up in people or animals over time. 3M's announcement that it is phasing out PFAS covers two compounds, Novec and Fluorinert, which are used in



two-phase immersion. PFAS chemicals have also been classified hazardous by the US Environmental Protection Agency, and are heavily controlled in Europe, making their deployment more tricky.

No such problems impact single-phase immersion, where cooling fluids can be manufactured in a variety of more environmentally friendly ways. Submer has its own cooling fluid made out of a synthetic hydrocarbon, while others on the market are based on hydrocarbon compounds such as vegetable oils, mineral oils, and fluorocarbons. Another oil company, Shell, offers a range of immersion cooling fluid S3 X, which it says is made from natural gas using the company's gas-to-liquids process.

The tanks in the Castrol MegaPod are each filled with one of the company's two data center cooling fluids, known as DC15 and DC20. Castrol is tight-lipped on the make-up of the fluids: "That would be telling," Barrett says when asked what goes into the products. But both are synthetic dielectric liquids designed with slightly different viscosities for different use cases and markets.

Both are capable of housing all types of electrical equipment (*DCD* asks if we can drop our phone in to test this out, but is told it may take a while to clean afterwards), and the Submer pods are designed to cater for the



standard 800mm deep servers you would find in the average rack, as well as larger 900mm models.

A mega opportunity

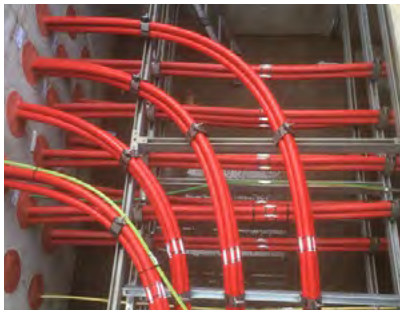
Castrol's Barrett says the MegaPod will give the company the "capability to conduct fluid research and development, IT hardware testing and validation and demonstration of immersion cooling technology."

It is the validation element that is likely to be most beneficial in the short term, with Castrol keen to use the pods to show that its fluids interact effectively with data center equipment. At the time of writing, a test run has been carried out at the pod ahead of the first customers being welcomed onto the site in the coming months.

Moving forward, Castrol hopes to use the MegaPod to demonstrate how data center heat can be reused. The system emits waste heat in liquid form, making it easier to capture and reuse than in other cooling systems.

Submer's Coggins said partnerships like the ones with Castrol are key to encouraging more adoption of immersion cooling. Castrol's fluids were validated for use with Submer equipment last year, and Coggins says one of the main barriers stopping companies from taking the, er, plunge is a lack of available kit. "One of the perceived drawbacks is the lack of an immersion-ready catalog," he says. "We collaborate with component manufacturers to ensure end-to-end solutions, and are seeing a lot more immersion-born SKUs and a growth in support.

"It's an area where demand is growing fast, particularly in the UK. We're seeing a lot of customers becoming more open to the idea of immersion cooling." ■



**EUROPEAN
DATA CENTRE
POWER CABLE
EXPERTS**

**ELAND[®]
CABLES**



**POWERING A
SUSTAINABLE
FUTURE**



**TALK TO US ABOUT
CPD CABLE TRAINING**

OUR ACCREDITATIONS AND COMMITMENTS



www.elandcables.com

The Houthis and the Red Sea: A new risk to subsea cables



Dan Swinhoe
Senior Editor

Intentional or otherwise, terrorists are now a threat to global connectivity

The Red Sea is a major artery of the global Internet. More than a dozen subsea cables pass through it connecting Europe to the Middle East, Africa, and APAC, transporting huge swathes of data traffic.

But this choke point presents a risk. For years the industry has been looking for an alternative way to reach the Indian Ocean from the Mediterranean Sea, but mostly to avoid Egypt and the pricey cost-of-entry requirements to run cables alongside the Suez Canal. However, at the other end of the Red Sea, a more urgent and violent risk is now causing issues for global connectivity.

Rebel groups actively attacking ships in the Red Sea have indirectly damaged a number of cables after a listing vessel dragged its anchor. Others have accused the groups of making direct threats against cables in the area. But how real is the threat?

Yemen, subsea cables, and the Houthis

The Red Sea spans around 2,250km (1,400 miles), running from the Suez Canal in the north to the Bab al-Mandab Strait in the south, before meeting the Arabian Sea and Indian Ocean.

TeleGeography estimates that more than 90

percent of all Europe-Asia capacity is carried by cables through this channel. The firm estimates for the likes of India, Kenya, and the UAE, more than 40 percent of each country's interregional bandwidth is connected to Europe via Red Sea cables.

Also known as the Gate of Grief or the Gate of Tears, the 26km (14-mile) strait runs between Yemen on the Arabian Peninsula and Djibouti and Eritrea in the Horn of Africa, connecting to the Gulf of Aden (and onto the Indian Ocean).

Despite only being the landing point for four cables, around 15 cables currently pass by Yemeni waters. More cables including the Blue-Raman, India-Europe-Xpress, and 2Africa



systems are due online in the coming years, all inevitably passing through the strait.

The Bab al-Mandab is a natural bottleneck between the Middle East and the coast of Africa, meaning any subsea cables connecting Europe to Asia are almost certain to pass close to Yemeni waters (and by extension, Houthi-controlled areas).

On the other side of the strait lies Eritrea, an isolationist country described as the 'North Korea of Africa,' with no subsea cables and little Internet freedom. Bertrand Clesca, partner at subsea consulting firm Pioneer Consulting, tells DCD most cable operators have generally laid cables in Yemeni waters as the country was historically easier to deal with than Eritrea.

The Houthi cable threat – real or imagined?

Like cell towers and data centers, subsea cables are critical infrastructure, providing connectivity to the outside world. This has long been known and exploited during times of conflict; one of the first such moves was made by the British during World War One and saw them cut Germany's undersea telegraph cables. A cable connecting Manila to mainland Asia was cut in 1898, during the Spanish-American War.

But intentional threats of violence against cables are often perceived through the lens of nation-states in deep waters. A 2017 *Policy Exchange report* produced by UK Prime Minister Rishi Sunak – then merely an MP – warned of 'aggressive Russian submarines in the Atlantic' and the need for more international treaties. Few, if any reports, have ever really considered threats to subsea cables from terrorist groups, given their relative obscurity and difficulty in reaching the seafloor.

However, that changed in late 2023. The Middle East Media Research Institute (MEMRI) – a think tank founded by a former Israeli Intelligence officer and a political scientist described as a 'neoconservative and revisionist Zionist' on Wikipedia – said Telegram channels reportedly affiliated with the Houthis had made implied threats against subsea cables in the Red Sea.

"There are maps of international cables connecting all regions of the world through the sea. It seems that Yemen is in a strategic location, as Internet lines that connect entire continents – not only countries – pass near it," one Telegram post said, accompanying a map of cables in the region.

This news was later picked up in mainstream media worldwide after Emily Milliken, SVP and lead analyst at US-based defense intelligence consulting firm Askari Defense & Intelligence, wrote a blog highlighting the MEMRI research, the criticality of the Red Sea in telecoms, and some of the

Who are the Houthis?

Officially known as Ansar Allah, the Houthis are a Shia Islamist political and military organization that emerged in Yemen in the 1990s. The group is largely made up of Zaidi Shia from the Houthi tribe of northern Yemen.

Originally a moderate movement, the group was focused on promoting a 'Zaidi revival' in the country and removing US and Saudi influences in Yemen. Over the years the group's numbers and arsenal have grown quickly – it is widely believed they are being supported by Iran.

After years of growing guerrilla warfare, the group took part in the 2011 Yemen Revolution and refused to concede to a Gulf Cooperation Council deal to quell the uprising. By 2014 the group was in control of parts of the capital city, Sana'a, and officially took control of the government the following year.

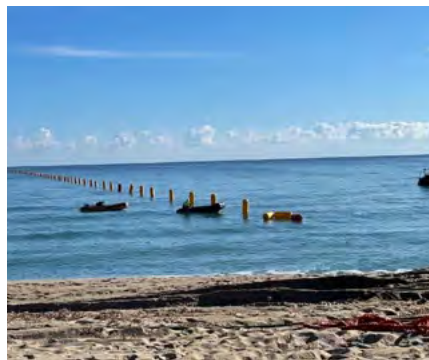
A Saudi-led coalition within Yemen has reclaimed some territory in the name of the UN-backed government. But the conflict has led to the existence of two concurrent Yemeni governments: one in Sana'a under Houthi control, and the

other in Aden, recognized internationally. As well as the former capital, Houthi insurgents currently control large swaths of coastland and all of the former North Yemen except the eastern Marib Governorate.

"This division extends to telecommunications, with each government overseeing different aspects of the sector," says Ahmed Nagi of Crisis Group. "Notably, a significant portion of the telecommunications infrastructure, including that related to the Internet and Red Sea cables, is located within Houthi-controlled areas."

The SeaWeMe-5 and Falcon cables land in Al Hudaydah, under Houthi control.

Falcon also lands in Al Ghaydah to the east, currently under the control of the Saudi-led coalition supporting the UN-backed government. Aden, which lands the Aden-Djibouti and Asia Africa Europe-1 (AAE-1) cables, is under the control of the Southern Transitional Council, a secessionist movement currently allied to the internationally recognized government.



Houthi military capabilities.

"They didn't directly threaten the subsea cables," Milliken tells DCD. "But they alluded to the fact that they were so important in the area. Which, for an organization that has been launching maritime attacks, is obviously a concerning statement."

Government ministries and telecoms firms backed by the UN-recognized government condemned the reported threats to the region's cable infrastructure, while Houthi-backed agencies said the reports were untrue.

The General Corporation for Telecommunications and the Yemeni International Telecommunications Company (TeleYemen), in Aden [i.e. the UN-backed government], said it "strongly condemned" the threats of the Houthis to target international marine cables.

The Ministry of Communications and Information Technology (MTIT) in Sanaa [i.e. the Houthi-controlled ministry] denied there was any danger to subsea cables in the regions. In a comment made through the Internet Society, the Ministry said reports of threats were "fabricated lies" being told to "cover the crimes committed by the Zionist entity in the Gaza Strip."

"The approach of the Government of Yemen, through the MTIT, is to focus on building and developing the telecom and Internet services, and expanding the range of services through the licensed telecom

"They didn't directly threaten the subsea cables. But they alluded to the fact that they were so important in the area. Which, for an organization launching maritime attacks, is a concerning statement"

*>>Emily Milliken
Askari Defense & Intelligence*



institutions and companies," the MTIT added in a separate [statement](#).

Abdul-Malik al-Houthi, leader of the Houthi movement, went so far as publishing a [video](#) saying the group has "no intention" of targeting submarine Internet cables in the region.

While she acknowledges it might not be in their best long-term interest, Milliken says she views the threats to cables as "credible" and is "dismissive" of Houthi claims to the contrary, given previous promises to only target Israeli-related shipping.

"They're trying to show themselves as a bigger threat in the region in order to garner more support and show themselves as an active part of Iran's resistance groups," she says.

But there may be a kernel of truth in Houthi claims to want to keep data flowing in the region, though for purely monetary reasons. The Counter Extremism Project (CEP), a not-for-profit policy organization aiming to 'combat extremist ideologies,' has [said](#) the Houthis have been using its control of large swathes of Yemen's telecoms infrastructure – including operator Yemen Mobile – to monitor and censor its populace while profiting from a growing Internet user base.

The group has reportedly repeatedly raised prices while failing to invest in the country's infrastructure. In 2022, the United Nations [described](#) the telecommunications industry in Yemen as "a major source of revenue" for the Houthis – CEP [estimates](#) potentially tens of millions of dollars.

Intentional damage against digital infrastructure by groups isn't new. Cell towers are often the target of conspiracy theorists, and terrestrial fiber cables have previously been purposefully cut. But this is the first time a terrorist group has been said to be a threat to subsea infrastructure – and motive remains a question if the Houthis really were interested in attacking cables.

Israel is largely dependent on terrestrial cables and subsea cables landing on its Mediterranean coastline, so any cuts in the Red Sea would have minimal impact on the country

"The circumstances that led to that ship being attacked, that could happen today. The threat is continuing, they're still shooting at ships. We're not out of the woods."

>>Doug Madory
Kentik Inc.

the Houthis have said they want to inflict the most damage on.

"They wouldn't really hurt the Western countries that they are targeting; it's not going to really affect Israel or Great Britain, certainly not the United States," says Doug Madory, director of Internet analysis at Internet monitoring firm Kentik Inc. "It's mostly countries either to the south or to the east that are affected by the loss of connectivity including Iran, which is their primary backer."

However, cable damage in the Red Sea could impact Yemen's Saudi neighbors, which could be a boon for the Houthis and Iranians.

"It's essential to consider all possibilities," Ahmed Nagi, senior analyst for Yemen at non-profit think tank Crisis Group, tells *DCD*. "The Houthis might indeed pose a threat to subsea cables, especially if tensions in the maritime lanes escalate further."

One Red Sea cable – the Global Cloud Xchange-owned Falcon system – lands in Iran as well as Saudi and Yemen, so attacking it could impact the Houthis and their only ally. But damaging the Saudi enemy could still appeal.

"The Houthis are not a wholly owned subsidiary of Iran," Milliken says. "It doesn't



necessarily matter if something is in Iran's best interest. They are more than happy to do attacks that don't necessarily benefit Iran and in some cases are detrimental."

Are cables really under threat?

Did the Houthis really make a veiled threat at the cables? It's hard to say. MEMRI no doubt has a pro-Israeli slant that might influence their assessment of the posts. But it's a fact they are attack ships of all stripes in the Red Sea and that is impacting cables indirectly.

The Houthis – officially known as Ansar Allah – have been attacking commercial ships passing by Yemeni water since November 2023. Dozens of ships have been attacked by drones, missiles, and speedboats. The Houthis claim to only be targeting Israel-linked vessels in support of Palestinians in Gaza amid the ongoing invasion by Israel.

US Central Command condemned what it calls the "reckless and indiscriminate attacks" on civilian cargo ships by the Houthis. The US State Department has labeled the Houthis as a 'Specially Designated Global Terrorist group' -- meaning they have committed or pose a significant risk of committing acts of terrorism – ranking them alongside Al-Qaeda, ISIS, Hamas, and Hezbollah.

At the time of writing, the Israel-Gaza conflict continues and Houthis continue to target ships in the Red Sea.

But could the Houthis directly attack subsea cables in the Red Sea if they wanted to? Some military experts have said the Houthis may have the ability, but others still have doubts.

"I can't see any part of the Houthi arsenal



actually being dangerous for the subsea cables," Bruce Jones, senior fellow at the Brookings Institution, told *Foreign Policy*. "The question becomes, do the Iranians have the capability, and would the Iranians take that step?"

"I assess it's a bluff, unless it's an attack on a terminal," former Royal Navy submarine commander Rear Adm John Gower told the *BBC*. "It would need an ally with the capability [of] a submersible plus the ability to locate [the cables]."

"There is nothing I've seen in the Iranian orbat (Order of Battle) that could touch these cables, certainly not their submarines," added former Royal Navy Cdr Tom Sharpe. "Diving is an option but it's deep and busy so I think it would be pushing it."

Wilson Jones, defense analyst at GlobalData, told *AirForce Technology*: "Yemen is near a disproportionate number of international undersea Internet cables. It would be very difficult to stop the Houthis if they made a determined effort to target these cables."

Carolina Pinto, thematic analyst at GlobalData, added that while they likely do not have the technological capabilities to reach cables that are hundreds, if not thousands, of meters underwater, they could "maybe target one or two of the shallowest cables."

While it reaches a maximum depth of 3,040m (9,970ft) in the central Suakin Trough, the Red Sea averages a depth of around 490 m (1,610ft). At its shallowest, however, some points are at depths of as little as 100m (330 ft).

While the Houthis might not have submarines, undersea robots, or the ability to hit the deepest parts of the Red Sea, it's possible to inflict damage on subsea cables without the backing of a major navy.

In March 2013, three divers were *arrested* by the Egyptian Navy off the coast of Alexandria after cutting the SeaMeWe-4 cable by detonating underwater explosives. Internet speeds reportedly fell around 60 percent after the incident. A motive wasn't revealed and it's unclear if they were charged and/or sentenced for the damage.

In 2007, it was reported that police had seized more than 500km of telecom cable taken by fishing vessels to sell for scrap –

including an 11km segment identified as belonging to the SeaMeWe-3 cable.

Askari's Milliken points out that the Houthis are backed by Iran, which has provided a large amount of arms and equipment to build weapons, including maritime-focused equipment. They have been known to use waterborne explosive devices, and the Combating Terrorism Center, an academic institution at the US Military Academy, *suggests* the Houthis have previously undertaken combat diver training on Zuqur and Bawardi islands in the Red Sea.

Regardless of their interest and ability to directly attack cables, their very real and direct attacks on boats are having an impact on future cables set to be laid in the Red Sea. TeleGeography noted in a recent *blog* that a segment of the 2Africa cable set to lie in Yemeni waters is yet to be laid.

Multiple industry sources named the likes of Blue-Raman, Africa-1, and SeaMeWe-6 as other upcoming cables that are likely to be delayed in the Red Sea because of the situation.

Direct threat or friendly fire?

Regardless of whether the Houthis have the motivation or capabilities to directly attack a cable, their attacks on the region's shipping industry have indirectly achieved the same end result.

February 18 saw the *M/V Rubymar* struck by an anti-ship ballistic missile. While one missile missed, the missile that did hit caused the crew to evacuate to a coalition warship and nearby merchant vessel. The drifting and listing *Rubymar*, however, is thought to have caused significant *damage to a number of cables* in the region.

The *Rubymar* crew left the ship anchored but slowly taking on water. It began drifting and it's thought that the ship is to blame for damage caused to three cables in the area. The AAE-1, Seacom/TGN, and Europe India Gateway (EIG) cables were all said to have been impacted around the same time in the area the ship was drifting.

Seacom was initially the only company to publicly confirm cable issues, but Tata later

also acknowledged an impact. Early reports – especially those in Israel – were quick to say the cables had been deliberately damaged by the Houthis, with the *Rubymar* details only coming out later.

DCD's sources say damage to the other cables has also been confirmed – and though the *Rubymar* theory hasn't been completely confirmed, at time of writing it is the industry's best guess consensus as to the cause of the damage.

In a sector where accidental damage to cables by fishing boats and anchors is common, this is likely the first time a cable has been damaged as a by-product of attacks on ships.

The MTIT – controlled by the Houthi administration in Sana'a – denied any involvement in any cable damage.

"MTIT and Government of Yemen reaffirm its obligation to the general position of the Yemen Republic toward the submarine cables [and] keen to keep all telecom submarine cables and its relevant services away from any possible risks," the ministry said.

Yemen's Transportation Ministry, currently under Houthi control, said the "hostilities on Yemen by the British and US naval military units" actually caused the disruption, which "jeopardized the security and safety of international communications and the normal flow of information."

The MTIT said it was "keen" to facilitate the repair of any submarine cables in the region, so long as parties obtain the required licenses and permits from the Sana'a Maritime Affairs Authority.

"The Sana'a MTIT and GoY also brings to notice that the Yemen's decision to ban the passage of Israeli ships does not pertain [to] the other international ships which have been licensed to execute submarine works within the Yemeni territorial waters," the ministry said.

The *Rubymar*, a Belize-flagged, UK-owned bulk carrier, sank on March 2 at approximately 2:15 a.m. (Sana'a time), according to a US CENTCOM update. It was carrying approximately 21,000 metric tons of ammonium phosphate sulfate fertilizer at the time, which present an immediate

environmental risk to the area.

Ahmed Awad bin Mubarak, the foreign minister of Yemen's internationally-recognized government in Aden, said in a post on X: "The sinking of the *Rubymar* is an environmental catastrophe that Yemen and the region have never experienced before. It is a new tragedy for our country and our people. Every day we pay the price for the adventures of the Houthi militia."

No easy fix

The situation may get worse before it gets better. The cable damage has been felt across Africa and Asia, and at time of writing, no repairs have been made.

"It doesn't matter what caused it, right now what matters is how fast can we repair the cables," says Alan Mauldin, research director at TeleGeography. "That's what the focus should be; how challenging is it going to be to get the permits? Can you get insurance? And as a maintenance company, do even you want to send your crew and vessel into an area where there could be active military activity happening?"

The Red Sea still has a number of undamaged cables serving Asia and the east coast of Africa. Normally cable operators have maintenance contracts with cable ship companies to ensure cable breaks are dealt with quickly, but in an active conflict area, the picture is more complicated.

Some industry experts have suggested the current conflict situation may trigger some force majeure clauses in the contracts, but DCD has been unable to confirm this. The Houthi-controlled MTIT may have promised safe passage to cable ship operators with the right permits, but that may be of little reassurance to those aboard when the group has previously kidnapped UN and Red Cross staff.

"I don't know that anybody wants to bet their life on, whether the Houthis correctly or incorrectly ascribe your ship to being a part of their enemy or not," says Kentik's Madory.

Even for cable ship companies willing to put their crews at risk, bureaucracy might be more of a stumbling block. It will be difficult to find insurance that covers sending ships into dangerous waters, and there are likely legal ramifications of doing business dealings with sanctioned organizations linked to terrorist groups to try and secure permits.

For now, there is enough spare capacity on the remaining cables and alternative routes – combined with backup satellite capacity – to mitigate much of the damage.

Intelsat has said its satellites are providing backup connectivity for a number of customers that were sending data via subsea cable before the incident. The operators of SeaMeWe-5 have said demand on its cable has gone up since

the break, while SeaMeWe-4 recently benefited from an upgrade that doubled its capacity from 65Tbps to 122Tbps.

Hong Kong's HGC said that while the cable breaks had "limited" impact in Hong Kong, the company had "devised a comprehensive diversity plan to reroute affected traffic."

"Having a few faults is a normal thing to plan and accommodate for, it happens," says TeleGeography's Mauldin. "But if you have a higher number of breaks on higher capacity cables, there'll be a point where it will start having a major impact on connectivity [in the region]."



Mauldin says we're "not close to that point yet", but a suspected undersea landslide that damaged several cables off the coast of Côte d'Ivoire a few weeks after the *Rubymar* incident has added to the continent's capacity crunch and is unlikely to help matters.

Repairs in the Red Sea are possible, though. In November, Global Cloud Xchange was able to complete scheduled maintenance on the company's Falcon cable in Yemeni waters "in conjunction with the Yemen cable landing party" – likely Houthi-controlled YemenNet.

"This maintenance event had been in planning for the past three months and was notified to all relevant parties and was completed successfully within the agreed maintenance window," the company said.

Pioneer Consulting's Clesca tells DCD it's likely that any repairs to the damaged cables in the Red Sea will be made with the protection of warships; he doesn't think they will remain unfixed for months.

"A few years ago when piracy was a problem, it was a simple measure to have armed guards on board. But here in the Red Sea, if you want to protect yourself, you may need heavier protection," he says.

The International Cable Protection Committee (ICPC) released a statement urging calm to allow the cable to be fixed.

"The ICPC urges governments to coordinate with operators to identify and mitigate causes of damage, expedite permits for repairs, and provide security for vessels engaged in such repairs."

The ICPC does have recommended best practices for governments to protect submarine cables and maintain communications in the event of damage, but these include expected behaviors such as observing treaty obligations including the United Nations Convention on the Law of the Sea, engaging with other states, and treating subsea cable ships as neutral in disputes.

"It's a very unique scenario," says Kentik's Madory. "But the circumstances that led to that ship being attacked, that could happen today. The threat is continuing, they're still shooting at ships. We're not out of the woods."

The long-term danger, however, is that even if there is a ceasefire in Israel, the Houthis will continue to attack shipping lanes in the area. Askari's Milliken says the attacks have been "beneficial" to the group, and even after a ceasefire in Gaza, could continue attacking ships, and even shift to a more profit-focused piracy model.

"I don't think it in their best interest to stop right now," she says. "I think they're getting too much domestic support out of it that they wouldn't stop even if Iran compelled them to. They have more than enough capability stockpiled to continue on alone."

That could, however, lead to a more "wide-scale response" from international forces looking to pacify the area and help business return to usual.

"Should military tensions in the Red Sea heighten, the possibility of the Houthis undertaking such actions increases," says Crisis Group's Nagi. "[But] this could lead to negative consequences for the Houthis, underscoring the complex implications of regional conflicts on global connectivity." ■



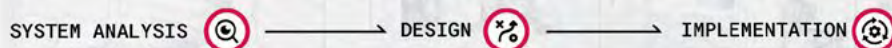
ENGINEERING EXCELLENCE, DELIVERED GLOBALLY.

AXYS Engineering Solutions empowers your facility operations team with the precise tools to stay ahead in your industry.



FROM CONCEPT TO IMPLEMENTATION:

Empowering Your OT Environment with End-to-End Ownership.



Industrial Visualization

- Aveva System Platform
- Factory Talk ME/SE
- GE Cimplicity
- Kepware
- WinCC
- TIA Portal

PLC for Industrial Controls

- | Rockwell | Siemens | Schneider |
|--------------|-------------|--------------|
| -Studio 5000 | -S7 | -Ecostruxure |
| -CCW | -TIA Portal | -Modicon |
| -Logix5000 | | -Quantum |
| -Logix500 | | |

Industrial Network Management:

- MOXA
- Hirschmann
- Cisco

Firewalls and Networks

- Waterfall

Design Software

- AutoCAD
- Solidworks
- EPLAN
- BIM with Revit
- Bluebeam Revu

VFD Controls

- ABB
- Toshiba
- Powerflex
- YASKAWA
- AB/Rockwell

Project Management

- Smartsheets
- Microsoft Project
- Sharepoint
- File Share/OneDrive
- ZOOM/Microsoft Teams
- ProCore

Physical Security/Access Controls

- Risk Management Consulting for Auditing Purposes
- System Design
- Security System Software Integration
- Electronic Locks and Physical Keys
- Motion Sensor and Anti-tailgating for Intrusion Detection
- CCTV Systems / VMS
- Intercom Installations

Our team can support: Greenfield, Retrofits and Expansions

We audit your systems and understand your policy for an efficient and fully compliant implementation.

axys.global



With **65+ years of combined experience**, a team of dedicated specialists will support you and take you to the next level of controlling your Datacenter facility.



The non-profit approach: British Heart Foundation CTO, Alex Duncan

How the third sector organization is using IT to save lives



Georgia Butler
Reporter

We often refer to data centers as the 'digital backbone,' but one could argue that it is in fact the heart, and a key part of the cardiovascular system of modern organizations.

The UK charity 'British Heart Foundation' (BHF) was founded in 1961, and is still beating strong over 60 years later, propped up by its data centers and use of cloud computing.

A non-profit, BHF raises money for research into heart disease, with a mission to eventually eradicate it. But as a large proportion of this comes from BHF stores and shops, the non-profit's IT needs are similar to that of any other retailer.

"We've got 700 shops and stores around the country, and five or six offices. We are a large retailer on our own without even taking into consideration fundraising," says Alex Duncan, CTO at BHF.

In 2023, BHF's retail net profits were £24.9 million (\$31.8m). The charity raised a total of £180.6 million (\$231m) for the year.

Duncan joined BHF in 2022, having previously served in similar roles with companies such as Costa Coffee and the Ambassador Theatre Group.

Fortunately for Duncan, it was not a story of joining an organization with its IT in a mess, "BHF's technology was already in a really good position," says Duncan. "So it's not like there were any burning problems that needed to be fixed."

But a lack of problems does not mean continuing the status quo, and BHF isn't shying away from change.

According to Duncan, the non-profit currently has two data centers that house around half of its operations, the rest is in the cloud. The organization was unable to share the location of said data centers, though conceded that it is a mix of on-premise and colocation.

"Our core systems are currently predominantly on-premise, but we do have a cloud-first strategy so anything new is deployed into the cloud, and we are working on getting all of our systems onto the cloud as quickly as possible," says Duncan.



The organization is currently only working with Microsoft for its cloud computing transformation, though Duncan says that they are not under an exclusivity agreement. Should a use case require it, or innovation be made, BHF would look at other providers as well. But for now, Duncan says "our needs at the moment mean we don't really need to go beyond one provider."

The cloud migration is a multiyear strategy and not one that will be rushed.

"Anytime you do [migrate data to the cloud], it's an upheaval, so we are trying to make sure that we do it at the right time," explains Duncan.

The charity is currently working on its cloud migration roadmap. "We are looking at all of our systems and essentially evaluating what the future is for each of them - how do we make sure that we've got the right foundational layers in terms of the platform data and the integration architecture?"

Once the game plan has been finalized, BHF is expecting the migration process to kick off later this year.

BHF is also looking at how it can use its data to help external organizations to influence research, and potentially policy as well.

The charity assists with the collection of external data looking at the state of cardiovascular research and related issues across the UK.

"We use that to generate reports which help to influence government and policy. Some examples of that are reports on the state of



"We are very conscious of running efficiently given that we are a charity, and of every penny that we spend that doesn't go on research,"

*>>Alex Duncan
British Heart Foundation CTO*

cardiovascular care after Covid, the health of the research community and pipelines of talent, and health inequalities among the general population as well."

While like any other retailer in many senses, as a non-profit organization, there is a different approach to spending.

"We are very conscious of running efficiently given that we are a charity, and of every penny that we spend that doesn't go on research," agrees Duncan.

The charity's accounts are publicly available for anybody to see, and BHF has to report on the percentage spent on research and charitable endeavors, versus on overhead costs. For example, in the year ending March 2023, BHF reports spending £6.9 million (\$8.84m) on Information Technology, up from £4.9m (\$6.28m) the year prior.

In total, BHF had £29.1m (\$37.3m) of IT equipment and software under use in 2023, conversely down by more than £5m (\$6.41m) from the year prior.

This keeps non-profits honest, but it is by no means a reluctance or refusal to invest in technology. Duncan instead argues that it makes BHF spend smarter.

According to the charity's 2023 accounts, around 80 percent of its total income is available for research and medical innovation. Duncan estimates that, on average, BHF funds around £100 million (£128.1m) of research a year, and around £450 million (\$576.55m) of research is happening at any one time.

"We are very conscious that every pound should be spent in the right way. But that doesn't always mean the cheapest way, because that isn't necessarily the best way. It's about being clear on what it is we are trying to do, and the best way of achieving that," she says.

For example, BHF, like many for-profit companies, is exploring the use of artificial intelligence (AI) to improve efficiency.

The organization is always gathering data, similar to that of other retailers. "We use



our data for reporting insights, and we are increasingly looking at how we can use that data to personalize our communications with customers and their journey on the website," says Duncan.

Additionally, BHF uses said data to ensure that it is operating securely, and is as "efficient and cost-effective as possible."

Using AI internally is still very early days for BHF. During a recent talk at the Microsoft Global Non-Profit Leaders Summit 2024, Duncan spoke about how the charity is bringing AI into its logistics, including optimizing stock management, and also employee productivity.

Duncan puts BHF's ability to explore new technologies, in part, down to the nature of a third-sector organization.

"In some ways, it's easier because the benefit of being a charity is that we are making decisions with a much longer timeline and distant horizon, whereas some publicly listed companies or retailers may need to focus on short-term results," says Duncan.

"We can focus on the longevity of investments, and evaluate them from that perspective, which changes your decision-making paradigm."

BHF, despite the relatively recent foray into AI for internal operations, has been funding AI-based research for several years.

Professor Charalambos Antoniades' work is

an example of this.

Antoniades' research has found an AI algorithm that can detect previously invisible changes to the fat around the arteries, enabling doctors to look up to eight years in the future and evaluate if a patient is likely to have a heart attack.

"It's changing the way we practice medicine," says Duncan of this research. "If you are flagged as being at risk, you can have medicine that prevents the heart attack from actually happening."

BHF and Alex Duncan believe so much in AI and data-science-based research, that the non-profit is investing around £20 million (\$25.6m) over the next few years in a data science center.

"We're in a wonderful position as a not-for-profit to be able to bring together whole population data in a way that private organizations can't, and just think about what we can achieve with that when we add AI into the mix," said Duncan at the Microsoft conference.

As for issues such as sustainability, Duncan says that BHF is currently reviewing its ESG strategy. "It's very much top of mind, but we haven't finished it yet."

"Overall, we're in pretty good shape, and in some ways, we are net positive because we contribute to sustainability by the fact that our business model is essentially recycling people's unwanted goods every day. But beyond that, we are working on a strategy to make sure our operations are as sustainable as possible."

Ultimately, the non-profit is investing in its IT in so much as it furthers its mission.

"Where we are different is that we have a wonderful purpose and mission. When our employees come to BHF to work, it's because they believe in what we are doing. It also makes it easier to slip technology in, because we are clear on our purpose," says Duncan.

"Our job is to save lives and save as many as possible through funding research." ■



TruFit™ PDU

Power Distribution Systems

The right products for your mission critical operations

Designed to provide an unmatched combination of power density, safety, and sustainability, the TruFit™ PDU brings highly configurable power distribution to the consumer.

- **From scale to speed to sustainability** – we'll work with you to find the answers.
- **You are the expert** in your business and we're here to support you wherever you are in the world.
- **The right products and services** for mission critical operations.

electrification.us.abb.com



Scan QR code
to learn more



Great X-pectations



Matthew Gooding
Features Editor

Gaia-X, Europe's grand plan for a federal cloud to take on the hyperscalers, appears to have hit the buffers.

NextCloud CEO Frank Karlitschek has been banging the drum for Gaia-X since before it was even called Gaia-X.

"I've been running around for the last 15-20 years speaking to politicians about the problem of monopolies in the cloud sector, and I was in contact with the people in the Ministry of Economics in Germany when they came up with the original Gaia-X idea," he recalls. "I was talking to them about it before that name existed, so I've been involved from the very beginning."

The prospect of Gaia-X was an enticing one: a federated European cloud platform big enough to challenge the market dominance of the US hyperscale providers and meet the data sovereignty needs of businesses.

Organizations from across Europe including NextCloud, an open-source enterprise collaboration platform akin to Slack or Google Workspace, signed up to the Gaia-X association, which was formed to put the grand plan into action. But since its launch in 2019, the project's form has changed somewhat, much to the displeasure of Karlitschek, who says it has been "diluted" and that he no longer attends association meetings.

Karlitschek, who believes the project has been sabotaged from within by the hyperscalers themselves, is not alone in losing faith in Gaia-X. For others, the problems



stem from excessive bureaucracy which has plagued the scheme since its inception. Whatever the cause, it now seems unlikely Gaia-X will deliver its original vision. But could a European federated cloud emerge anyway? And what role will the Gaia-X association play in the continent's rapidly evolving digital infrastructure?

Gaia-X: The story so far

Gaia-X was conceived to solve two problems in the European market for cloud computing infrastructure.

One was the dominance of the market by the hyperscale providers, most of which are based in the US. Globally, three platforms - Amazon's AWS, Microsoft Azure, and Google Cloud - controlled 66 percent of the public cloud market at the end of 2023 according to data from industry analysts Synergy Research Group. In Europe, the trio reigns even more supreme, with Synergy's

most recent publicly available data for the continent, from September 2022, showing a 72 percent combined market share for the three companies.

When one factors in that less than half (42 percent) of European businesses bought cloud services in 2023, per figures from the EU's Eurostat, it seems likely this stranglehold will be strengthened as the overall market grows. Europe-based cloud companies have long claimed that the hyperscalers use their, well, scale to garner an unfair advantage and stifle competition in the market. Trade association CISPE (which, somewhat ironically, counts AWS among its members) has lodged a formal complaint with the European Union against Microsoft over alleged unfair licensing terms for some of its cloud products, which remains unsettled at the time of writing.

As the hyperscaler dominance builds, businesses and public sector organizations have been increasingly concerned with data sovereignty, or where information that is stored in the public cloud geographically resides. The legality of data transfers in and out of Europe to servers and data centers in the US - a practice commonly used by many cloud companies - has been repeatedly challenged in court, with various data-sharing agreements between the EU and the US government being successfully struck down by privacy campaigners because US law is not as stringent as the EU's GDPR

when it comes to data protection.

Faced with the threat of GDPR infringements (and potentially hefty fines), ensuring European data was stored on servers in Europe suddenly became a priority for many organizations, particularly those working in heavily regulated industries, like financial services or healthcare, where much of the information concerned is highly sensitive.

Enter Gaia-X, which Kalitschek says started out as an attempt to build "Airbus, but for the cloud market." He explains: "Airbus was an organization founded in Europe as a counterpart to [US manufacturer] Boeing - a Europe-wide company that builds planes.

"The question was: did we need a Europe-wide company that does cloud? I was very happy when the conversation went in a different direction, because rather than building one company it was decided we should try and create a virtual hyperscaler."

Instead of starting from scratch to try and come up with the next AWS, Gaia-X's architects in the French and German governments resolved to pool the resources of the various smaller companies in the market to create a catalog of products comparable to those on offer from a single hyperscale platform. "You would pick the services you want and combine them together," Kalitschek says. "So you would buy storage from [French business] OVHcloud, some compute from [German cloud company] Ionos and payment services from another vendor and mix and match them together."

Championed by Germany's then minister of economic affairs, Peter Altmaier, and his French counterpart, Bruno Le Maire, the Gaia-X association launched in 2019 with 22 member companies, but really started ramping up its operations two years later. A further 212 companies, including a host of big names from computing and telecoms, joined the association, while Francesco Bonfiglio was appointed CEO.

Bonfiglio, an Italian with 30 years of experience in the IT industry on his CV, left the Gaia-X association last year, and says he is proud of his achievements during his time at the helm. "I had three objectives when I joined the association," he says. "The first was to define what Gaia-X was, which I did in 2021. In 2022, we clarified how Gaia-X solves the problems it intends to solve, and in 2023 we made Gaia-X available to everyone."

Enter the hyperscalers

What Gaia-X has created so far is a series of standards for sharing data in the cloud explains Max Claps, research director for government insights at IDC. "It quickly morphed into into an international initiative that was less about protecting the sovereignty of the European digital infrastructure, and more about creating blueprints that become the gold standard

"Some people wanted a cloud champion, some people wanted a regulator, some people wanted a standards body, and some just wanted restrictions to cut out the Americans."

>>Francesco Bonfiglio

for digital sovereignty, data protection, transparency, accountability and data sharing across the globe," Claps says.

To demonstrate how this works, the Gaia-X initiative launched a series of 'lighthouse' projects in different industry verticals which allow businesses that are certified as Gaia-X compatible to work together through shared data infrastructure. This is in line with the European Union's 2020 data strategy, which aims to create a series of 'data spaces' across the continent where businesses could pool and share information securely to try and promote economic growth.

Last year it also set up two digital clearing houses, automated systems for certifying that a business is Gaia-X compliant and allowing them to access other Gaia-X services. These are operated by Italian cloud provider Aruba and German telco T-Systems.

These digital clearing houses are significant, says Dario Maisto, senior analyst at Forrester, because they take Gaia-X "from a few slides on a Powerpoint to actual services that businesses can consume." He explains: "It's something we can see, that the initiative has achieved for real, and shows that it is working towards an objective."

But he adds, "How many applications have I seen of the clearing house? Not many. No company has asked me, as an analyst, 'How do we access the clearing house?' or 'What is it for?' So it is proving the initiative is working, but whether it meets a market need is a different story."

Talk of data spaces and digital clearing houses is far removed from the virtual hyperscaler plan, but was always part of the initiative's ambition. Bonfiglio explains: "The X in Gaia-X is meant to represent two things - the upper part is for data ecosystems, and the lower part is for the cloud infrastructure. For many reasons, the lower part has been neglected by the association, even though it is the most urgent part we need to address."

Bonfiglio says economic and political

factors are to blame for this. In the case of the former, he explains: "When you understand the power of the data economy, you realize you can do things by sharing data that you cannot do by yourself. There is money to be made or saved by sharing data in a secure, confidential environment, without having to spend anything on technology.

"So the investment in European data spaces is a very positive thing, but this data economy still runs on data platforms that are not under European control."

The political problems stem from the make-up of the Gaia-X association, which spans small and medium-sized businesses right up to the hyperscalers, which were welcomed in amid some controversy. These members, Bonfiglio says, have very different agendas. "There are a lot of internal conflicts of interest between European companies, European countries, and non-European members of the association," he explains. "With an organization of 350+ members, a lot of which are small and medium businesses, not all of them can set the strategy and it's inevitable it will be driven by the bigger companies which can influence governments."

Some European governments have also changed their position on the US cloud providers since the announcement of Gaia-X in 2019, Bonfiglio says. "At that time we had a Trump administration and the signs from the European politicians had to be strong. They had to say they wanted to build a European cloud champion," he says. "But between 2019 and 2021, when I started at Gaia-X, things changed. By then, some people wanted a cloud champion, some people wanted a regulator, some people wanted a standards body, and some just wanted restrictions to cut out the Americans."

Indeed, despite preaching on the merits of digital sovereignty, European governments continue to spend money with the US hyperscalers, with Germany reportedly committing \$3 billion to Oracle Cloud Infrastructure in its 2024 budget.

Bonfiglio is diplomatic when describing this situation, but his frustration is palpable. "The world has changed a little bit and these changes were reflected in the association," he says. "But I'm not a politician, I wanted to make something happen and I knew what was needed. But when it was the moment to change things and make things run faster, things were slowed down by these internal differences. That's when I decided to step out."

NextCloud's Kalitschek is clear that the involvement of the hyperscalers killed off the Gaia-X idea in its original form. "I think it was actively pushed in [its current] direction by the hyperscalers," he says. "What we've got is some specifications and some not-so-useful paperwork. To follow all this kind of paperwork is not possible for mid-size companies."



"It started with a bold vision and ended up being a community where people with good intentions meet and don't achieve much in terms of practical outcomes"

>>Max Claps

Bonfiglio disputes that the hyperscalers are the problem, and says that, even though the decision to admit non-European companies was made before he took up his post, he would have done the same thing. "The statutes of the Gaia-X association were, and are still are, strong enough to prevent anyone from hijacking the project and driving it in the wrong direction," he says. "That the project has become watered down is not the result of the Americans being involved. There were forces on the Gaia-X board, driven by European companies, that were not always going in the same direction."

Of the relationship between Gaia-X and the hyperscalers, he adds: "Cloud - and business in general - are very pragmatic things. If you want to win the battle you have to fight on the battlefield, rather than saying 'I don't want to fight them because they're too strong.' We wanted to find new weapons for the fight [with the hyperscalers] and I think that was the right thing to do."

Looking to the future

The Gaia-X association is continuing its work under a new CEO, Ulrich Ahle, and has set out a 2024 strategy based around four pillars: end-user adoption, market readiness of Gaia-X technology, globalization of the initiative and ecosystem growth.

According to Ahle, the association's mission continues to encompass "both the data and digital infrastructure aspects, recognizing their crucial role in today's digital landscape."

He told DCD: "As data spaces are becoming increasingly important for AI some may perceive a focus on data spaces, [but] it's

imperative to understand that data has become a strategic asset driving sustainability and competitive advantage across various business domains. Moreover, there's a growing need for global collaboration facilitated by shared data."

Ahle continued that Gaia-X is rooted in European values and remains "dedicated to designing a high-performance, competitive, secure, and trustworthy federated data infrastructure."

He says: "As 'data doesn't flow on rainbows,' this infrastructure serves as the backbone for enabling data spaces and fostering seamless collaboration while upholding the principles of sovereignty and security."

"Thus, our approach ensures a balanced emphasis on data and digital infrastructure, aligning with the evolving demands of the digital economy and the values driving Gaia-X's mission."

IDC's Claps believes the current form of Gaia-X is likely to be the extent of its ambitions. "It started with a bold vision and ended up being a community where people with good intentions meet and don't achieve much in terms of practical outcomes," he says.

"That's not to say it isn't useful, it has built a community that has advanced the discussions on this topic, but I don't think it will deliver a product, and nor should it."

Claps argues that the need for a European cloud provider has been reduced by new products on offer from the hyperscalers. All the major cloud vendors now offer some sort of sovereign cloud option that guarantees data is stored on servers in a territory of the user's choice.

"The cloud service providers are always going to deliver products faster than an

institutional initiative like Gaia-X," he says. "I think what's on the market now is sufficient to meet the digital sovereignty requirements of 95 percent of use cases, even in heavily regulated industries. Gaia-X has helped accelerate these regulatory-compliant regional investments from the cloud providers, but at this point, I don't think an additional product is necessary, because it would only confuse things."

Bonfiglio believes the idea of a European federal cloud still has legs, and since leaving Gaia-X has formed his own company, Dynamo, to try and make this a reality.

He explains that it will offer a catalog of cloud products, and any cloud vendor which is signed up to Dynamo will be able to deliver their customers services from this catalog alongside their core offering. The services will need to be compliant with Gaia-X standards to prove their trustworthiness.

As for Gaia-X itself, he says: "I think the common data spaces will be the focus of the association, and it won't evolve the infrastructure technology side of things too much."

"In that area I think it has reached the limit of what it can do, but I am looking forward to continuing to work with the Gaia-X members because what we're doing with Dynamo will be the fulfillment of the promise we made at the very beginning."

As Bonfiglio presses ahead, NextCloud's Karlitschek says his Gaia-X journey is coming to an end. "It's very sad, because I really liked the original idea," he says.

"We're still a member, but I guess I should terminate it at some point to save money. I don't see a lot of benefits there anymore." ■

MWC 2024: An industry turning slowly towards the AI future

Though many of the hot topics at this year's show were familiar, AI is new buzzword in town



Paul Lipscombe
Telecoms Editor

Last year Mobile World Congress was dominated by themes such as sustainability, and the battle between mobile operators and Big Tech over fair share payments for the usage of mobile networks.

This year, however, felt a bit different.

It was the biggest show since the pandemic, with 101,000 people attending the mobile industry's showpiece event in Barcelona, Spain. The King of Spain, Felipe VI was one of those in attendance, along with Pedro Sánchez, the country's Prime Minister.

Unsurprisingly AI was discussed at length, but this should come as no surprise given the hype around the technology in the past 18 months or so and the soaring demand for AI products and services.

This year's show stretched far beyond telecoms, with flying cars, helicopters, and other use cases on display. But all have one thing in common, they'll need reliable networks to underpin them.

AI impossible to ignore

AI was undoubtedly a hot topic on the show floor. Whether people were bigging up the technology, or in some cases skeptical about its potential, it was nigh impossible to avoid.

During one keynote, Vodafone CTO Scott Petty claimed AI is "overhyped at the moment."

Some may argue that he has a point given some of the more outlandish predictions about its potential, but it is clear telcos are taking AI seriously. Members of the Global Telco AI Alliance (GTAA) are planning to establish a joint venture later this year specifically aimed at developing Large Language Models (LLMs).

In a joint statement, Deutsche Telekom, e&Group, Singtel, SoftBank Corp., and SK Telecom revealed the plans after holding the inaugural meeting at the trade show.

Collectively the quintet claim that AI will enable chatbots to become even more human-centric, and say the JV will bring Europe and Asia closer together.

What do we want from AI?

It's tricky to know exactly what AI will bring. For all the promise, there's also a lot of skepticism, as well as fears from some quarters that it render a lot of job roles redundant.

BT has said as much, noting that a *fifth of the 55,000 jobs it plans to cut* by the end of the decade will be replaced by AI. This won't be something that is just confined to the telecoms sector either.

Jürgen Hatheier, chief technology officer for APAC and EMEA at networking solutions vendor Ciena, told *DCD* that there are still a lot of questions to be asked of AI.

"At Ciena, we're currently focusing on the opportunities around AI, and how we can use AI to provide better services, and to beef up our products with those capabilities," he says.

"If you stay around in observation mode for too long, watching AI. It's going to be very hard to catch up. It's not only a technology challenge, I think it's going to be a big commercial modeling challenge, to work out how you can intercept the big money that's expected from AI."

More Open RAN

Away from AI, a topic that was pretty hot once again was Open RAN. More telcos announced partnerships around this, including *Kyivstar and Rakuten Symphony*, and *NTT and AWS*.

Telcos are expected to *invest more than \$30 billion in Open RAN* by the end of the decade.

Analyst firm *Counterpoint Research* expects investments will start to increase year-on-year (YoY) after 2025.

The technology promotes a new breed of telecoms kit that allows providers to mix and match solutions from multiple vendors, rather than being tied to proprietary equipment supplied by a single vendor. It is hoped this will promote competition in the market and offer better value to customers.

Towards the end of last year, AT&T and Ericsson sealed a *lucrative \$14 billion Open-RAN deal*. A big advocate for the technology, AT&T expects 70 percent of its mobile network traffic to flow across Open RAN by 2026.

At this year's show, the duo completed a *cloud RAN call*, using Ericsson's cloud technology over AT&T's commercial 5G network.

"This [Cloud RAN call] was a huge accomplishment for us, and was a good step in the direction of going down the path of this open multi-vendor, programmable RAN network that we that we envisioned," Adam Loddeke, AVP of RAN technology at AT&T, told DCD.

The call was carried out in the South Dallas area. Further calls will take place, but Loddeke warns it will take some time - and lessons will be learned.

"We'll grow the existing trial area in South Dallas into other areas later this year," he says. "The exact volume is still being determined. I would say we're approaching it like a crawl, walk, run scenario.

"We're currently at the crawl stage, where we're learning lessons about how to operationalize."

He noted that Ericsson will be working through its own learnings from the software.

"We also have to do some development within AT&T to support the inclusion of Cloud RAN within our network, so there's IT system development that has to occur, along with processing skills training, because it's new for AT&T, and so you have to scale the workforce as well," Loddeke adds.

Samsung wants to go big

One of the promises of Open RAN is that it will bring on board more hardware and software players, so there were probably a few laughs when some of the big announcements featured the same old names, such as Ericsson or Nokia.

But one new vendor that wants to make some waves in the Open RAN arena is Samsung Networks.

The Korean electronics giant has struck Open RAN partnerships with *Dish Wireless in the US*, *KDDI in Japan*, and more recently *Canadian telco Telus Networks*.

Magnus Ojert, senior vice president, networks business, Samsung Electronics America, told DCD that it sees a real opportunity

to stamp its mark in the Open RAN race.

"With Samsung, when we get into a business, we don't want be the either the fourth or third best player," Ojert says. "We ultimately want to be the number one, we want to be the best, or the top vendor in the world."

The Open RAN approach makes this possible, Ojert says. He believes it has opened the market up for a number of different players, including chip companies and software companies which wouldn't have usually been involved in the network buildout.

5G Advanced in Malaysia

With 5G now commonplace in multiple markets around the world, attention has turned to 5G Advanced (5G-A), or 5.5G as some have labeled it.

5G-A is tipped to drive AI-based capabilities to 5G RAN, core, and operation domains.

Chinese vendor Huawei was talking particularly loud about the technology, a few days after it revealed it had worked with *Malaysian telco Maxis to complete the first 5G-Advanced technology trial* in the country and Southeast Asia.

Huawei boasted that it was able to demonstrate peak speeds of up to 8Gbps.

For context, *Opensignal* revealed in a study last year that South Korea has the fastest average 5G download speeds at 432.5Mbps.

It wasn't just Huawei talking about 5G-A, Ericsson also waded into the discussion. Once again the 5G Advanced conversation centered around Malaysia, with *Ericsson signing an MoU with Malaysia's Digital Nasional Berhad (DNB)* to deploy 5G-A in the country.

Ericsson is the network provider for the DNB, which is Malaysia's first shared 5G network.

"This new '5G Advanced' technology will bring advanced radio security, superior user experience, and enterprise capabilities, as well as empowering sustainability through enhanced energy management," claimed Nasution Mohamed, chief operating officer at DNB.

Spectrum splits opinion

It's not a telecoms conference without a discussion about spectrum, and this year's show was no different.

Thierry Breton, The European Commissioner for Internal Markets, was even drawn into the conversation on spectrum allocation, which is always a divisive issue.

"European businesses and citizens are always at the center of our ambition, but we must balance the focus with the need for a competitive and modern economy," Breton said.

"We should have the same spectrum policy throughout Europe and push for more timely and affordable options."

Breton says the mobile industry needs to do better with the allocation of 6G spectrum.

"In regard to 6G, we cannot afford to encounter delays in the spectrum licensing process with huge disparities in timelines between member states," Breton said. "We cannot tolerate the same outcome as for 5G where the process is still not complete after eight years.

"Investment is needed now to bridge the gap. No less than €200 billion (\$216bn) [is] needed over the next five to six years, and this is only for completing the rollout of 5G," he added.

A separate debate on how much spectrum is required for 6G was also held during MWC. This saw Doug Kirkpatrick, president and CEO at radio-unit maker Eridan, argue that the existing spectrum should be repurposed.

"From a technology perspective, we don't need any more spectrum," Kirkpatrick told delegates. "The challenge is not the spectrum, the challenge is pushing the technologists to push the pieces that you need, we have probably the spectrum in the wrong places.

"We need to think more about what kind of spectrum we want to use for what kind of applications. Moving applications from one spectrum to another is not a zero-sum game; you're going to have a transition period in between. So we will likely need some intermediate spectrum in order to make those movements happen, but when we're done, the answer should be zero."

What next?

The industry is very much in its 5G era now, be that through the 5G standalone deployments, various spectrum launches, and Open RAN architecture underpinning these networks.

Next up is 6G, and although it's not expected to launch commercially before the end of the decade, that doesn't mean it's not being discussed.

"6G is still very raw," says Youssef Latoyef, VP head of strategy and portfolio and cloud services at Ericsson. "The bulk of 6G will be on the radio side.

"From a core perspective, we look at 6G as a bit of an evolution of 5G, especially on the standalone core. So for us, it's about how we continue driving and leveraging the capabilities we introduced in standalone to apply to the 6G networks."

For many, AI will make more of an impact on the telecoms industry when 6G becomes a reality. For now, it's a waiting game, but there's a good chance there will be more 6G talk at next year's show. ■

Using AI to help chip design



Charlotte Trueman
Compute, Storage, &
Networking Editor

Chip companies are currently competing to bring increasingly powerful AI chips to market, but does the technology itself hold the key to helping chip designers make that a reality?

AI is everywhere. It underpins the tools you use in your workplace, it determines the deals offered to you when you go grocery shopping, the customer service representative you reach out to is probably a chatbot, and it is increasingly trying its hand at journalism (although not in this case).

Another area where AI is becoming increasingly pervasive is in the chipmaking industry. Currently, it can take between 18 months to two years to design a chip, and as compute requirements increase, it's a process that is becoming increasingly costly and time-consuming.

The manufacturing process, while less time-consuming, is no less complex and can involve hundreds of steps, making the shift from design to mass production deeply laborious. It's therefore unsurprising that chip companies have started to dip their toe into the AI pool to see if the technology can bring efficiencies to the industry.

When it comes to chip design, Alan Priestley, VP Analyst at Gartner, said that on a very simplistic level, there are a number of things to consider, namely what you want the chip to do, which is reliant on the function of logic blocks; the layout of the chip and the conversion of

those logic blocks to transistors on the silicon surface; and then testing and validating the chip to ensure it does what is intended of it.

Within almost all of these steps, he said that AI tools could be theoretically deployed to speed up the design process, such as automating layout optimization tasks like floor planning and routing, or simulating the behavior of a chip in different scenarios, thus reducing the need for physical prototypes.

Using emerging technologies to assist with chip design is not a new concept Priestly argues, noting that the technology used to develop chips today is incredibly complex by comparison to that of yesteryear.

Priestly, who also has an engineering degree and worked at International Computers and Intel in the past, has some hands-on experience with how things have changed.

"You can't you can't design a current generation chip using the techniques that were out there 30 or 40 years ago."

"If you go back in time, the first chip I ever designed was drawn on paper by hand, and when they made the masks, that was all cut out by hand. So, we've already added computer technology to assist with stuff like the layout design."

Referencing modern-day CPUs, which can contain anything from a few million to billions of transistors, Priestley adds: "You can't design a current-generation chip using the techniques that were out there 30 or 40 years ago; you need sophisticated computer technology to design it today. Adding AI techniques into that design process is just the next step."

In July 2023, speaking at the World Artificial Intelligence Conference (WAIC) in Shanghai, AMD's CEO, Dr. Lisa Su, said that the company had already started to use artificial intelligence for designing chips, adding that she expected that AI-enabled tools would eventually dominate chip design as the complexity of modern processors continues to increase exponentially.

The company's interest in AI-powered chip design was elaborated upon by Andrej Zdravkovic, senior vice president of GPU technologies and engineering software and chief software officer at AMD, in comments to DCD.



"AMD engineering teams are always looking for new ways to use cutting-edge technology in our design process. We have used predictive AI models for many years now and they have proved to be extremely useful in recognizing patterns and helped us improve productivity and reduce development time," he said.

Zdravkovic explained that AMD has been deploying these models to help identify potential issues early in the hardware and software design process, providing the company with additional tools to make informed decisions.

"With the development of large language models and explosion of the generative AI we have started to look at integrating AI into our workflow for the silicon and software design process to more efficiently deliver faster and more innovative designs," he said, adding that AMD is also identifying how AI can help automate and optimize the repetitive tasks including checking and correcting the RTL or software code for best practices, architecture and security standards.

The role of EDA

If you look beneath the hood, in a lot of use cases AI is just being used to automate undeniably dull tasks. In that regard, AI-powered chip design is not a wholly new concept.

Electronic Design Automation (EDA) companies have been around for decades, with the earliest EDA processes being attributed to IBM in the 1950s. However, the continuous scaling of semiconductors is making them increasingly popular with chip manufacturers.

Founded in 1986, Synopsys is one such EDA company that supplies tools and services to semiconductor manufacturing companies. It took its first foray into the world of AI-powered chip design in 2020 when it launched a cloud-based AI software tool called DSO.ai, which uses reinforcement learning to automatically decide how best to place and route blocks of circuitry on a chip.

Previously, the ability to explore the design space without the use of compute was not possible because time limitations meant there were only so many iterations it was humanly possible to run, explains Arvind Narayanan,

executive director of product line management, at Synopsys EDA Group.

However, the company realized that by bringing AI into the process, it would provide its customers with the ability to run millions of combinations within the design space in a reduced time, allowing designers to achieve higher performance, lower power consumption, and smaller chip area with less manual effort.

Following the release of DSO.ai, in March 2023, the company launched Synopsys.ai, an AI-powered design automation suite that deploys generative AI across the entire EDA stack. This suite includes VSO.ai for verification, TSO.ai testing, and ASO.ai for analog design, with the company essentially taking the architecture it used to build its DSO.ai offering and scaling and optimizing it for different workflows and parameters.

Synopsys' full suite of AI-powered tools now forms an end-to-end solution that includes system architecture, design capture, verification, implementation, sign-off, testing, and silicon manufacturing.

The company has since partnered with AMD, Intel, and Nvidia, and counts organizations that design everything from chips



"The challenges currently facing the semiconductor industry are arguably more complex than ever before, meaning that companies of all sizes and from all industries are now turning to EDA organizations and AI tools to help them address some of these issues"

for high-performance computing, to AI, mobile processors, automotive, and electronics among its customer base.

The challenges facing chip designers

Priestley said that EDA companies have traditionally partnered with smaller companies, as bigger organizations already have their own in-house development capabilities.

However, Narayanan said that the challenges currently facing the semiconductor industry are arguably more complex than ever before, meaning that companies of all sizes and from all industries are now turning to EDA organizations and AI tools to help them address some of these issues.

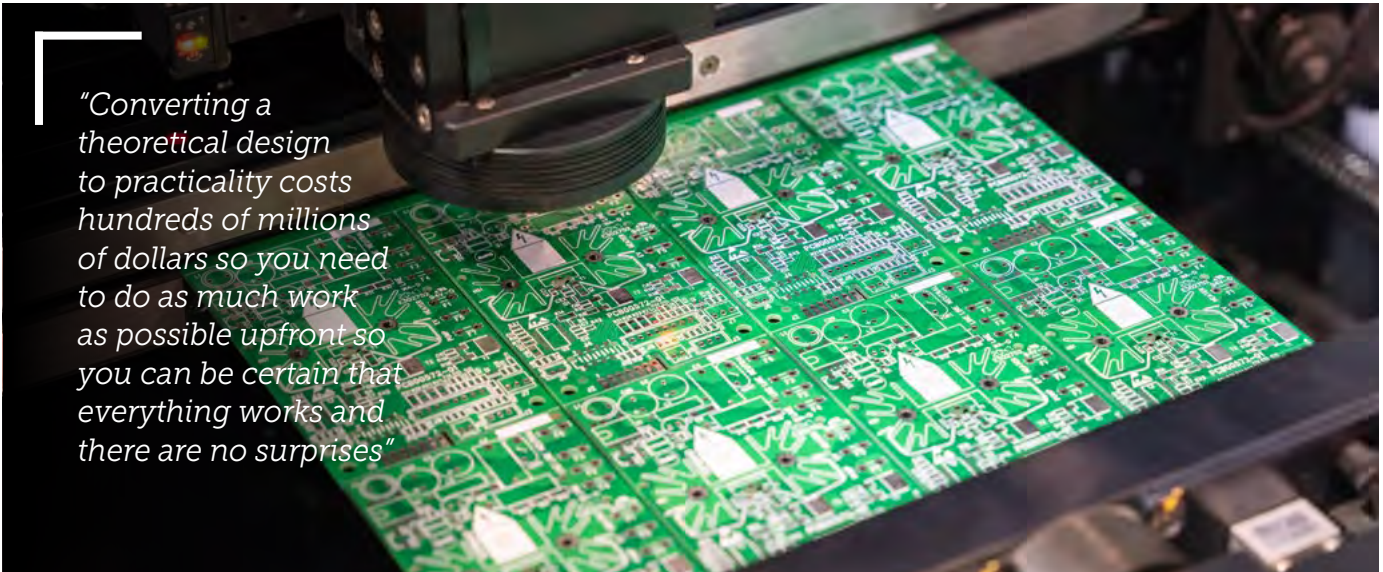
Most conversations relating to the biggest challenge facing the industry will involve a discussion on how much smaller can chips get. Currently, the smallest chips in production are 3nm but the race to mass-produce 2nm is officially underway, with availability expected from 2025.

However, traditional lithography processes – the action of creating minute patterns on a silicon wafer – are reaching their limits because, as transistors have grown smaller, this process has required increasingly complex calculations to work out how to operate at such a small scale.

One company that has tried to tackle this problem is Nvidia, which launched cutLitho, a software library for computational lithography that improves chip design development times, at its GTC conference in March 2023.

At launch, the company said cutLitho would facilitate the development of chips with tinier transistors and wires than is currently achievable, while accelerating time to market and improving the energy efficiency of the data centers that run alongside semiconductor fabs as part of the manufacturing process.

At its 2024 conference, Nvidia said its cutLitho library was now being used by TSMC



"Converting a theoretical design to practicality costs hundreds of millions of dollars so you need to do as much work as possible upfront so you can be certain that everything works and there are no surprises"

and Synopsys for production chip development. The company also claimed that generative AI had allowed cutLitho to provide a 2x boost in performance on top of the accelerated computing it originally offered; which itself was a performance leap of up to 40x beyond current lithography.

Separately, in October 2023, Nvidia announced it had developed ChipNeMo, a large language model (LLM) to help its employees design chips.

Another challenge facing the semiconductor industry that AI could help to address is the talent shortage.

As is the case across most of the technology sector, the semiconductor industry is desperately lacking talent, with a 2022 report from Boston Consulting Group (BCG) titled 'The Growing Challenge of Semiconductor Design Leadership' stating that at the current rate of growth, demand for design workers in the semiconductor industry will exceed supply by nearly 35 percent in 2030.

Consequently, BCG notes that design leaders must leverage "new and future technologies" that are critical for design innovation, including AI.

"Designers can more quickly and effectively meet power, performance, and area targets by leveraging AI-based tools. Reinforcement learning and other AI algorithms can automate less consequential design tasks, freeing engineers to focus on more advanced tasks and decisions," the report reads.

Narayanan said that AI has already made significant inroads when it comes to chip design but he believes there is so much more that can be achieved with the technology, not just in terms of workflow optimization but through the wider benefits it can bring to the industry.

Narayanan references the BCG report

and says that the talent shortage is a very real issue that Synopsys' customers are currently grappling with.

"They see that there is so much potential [for AI] in improving productivity and that's the direction most of the companies are headed," he said.

"[Chip companies] all have the same set of challenges. What they're designing doesn't matter, they all have the same set of challenges. So, when you have technology that can increase the productivity of your existing workforce, it would be foolish not to look at it and adopt it."

"I don't think [AI is] a difficult sell. They're all sold on it already," Narayanan said.

Will the benefits of AI outlive the hype?

As is the case with most emerging technologies, it can be hard to judge if its longevity will outlive its hype.

For now, it appears that the AI bubble won't burst any time soon however, that doesn't mean that all caution should be thrown to the wind just yet.

Gartner's Priestly is more pessimistic than most, saying that we're reaching the peak of the hype cycle and as is often the case, it could very much all come crashing down.

While he acknowledges that AI will undoubtedly accelerate the design process, he warns that the technology is like a black box and therefore the challenge lies within the outcome. He says, that as numerous examples of questionable content created by generative AI have proven, you can't always guarantee that the requests you make will generate the correct response.

"That's the big issue here because chip design is expensive," Priestly said. "Converting a theoretical design to practicality costs hundreds

of millions of dollars so you need to do as much work as possible upfront so you can be certain that everything works and there are no surprises. Especially as the more complex the chip is, the harder it is to test it."

AMD's Zdravkovic echoed Priestly's point that AI will improve the design process by accelerating the speed at which the design process can be completed but warned that human interaction should remain part of the solution, as chip design requires an "in-depth understanding of the complete design space and well-defined interaction and dependency of all system parameters."

Looking to the long-term, however, his comments reflected those Su made in 2023, saying that while recent advances in AI and new capabilities offered by AI are nothing short of phenomenal, it feels like the industry has just started to scratch the surface.

"I firmly believe that over the years we'll move most of the repetitive design work to advanced AI tools, while freeing up our engineers to do creative tasks, and invent new advanced silicon and software architectures," he said.

For Narayanan, the glass isn't half full but rather overflowing. He said that amongst Synopsys' customers at least, where there might once have been some skepticism (as is often the case with any new technology) that seems to have evaporated as customers have seen the value AI can bring to chip manufacturers.

He said that organizations are now realizing that they can do even more without having to grow their workforce, and can accomplish these growing workloads in the same amount of time or even shorter windows of time.

That's where he says Synopsys has seen AI play the most critical role, bridging the gap between what customers need to do and what they can do. ■

Exploring Static UPS systems



Vlad-Gabriel Anghel
DCD>Academy

Part 3

read [part 1](#) & [part 2](#)

So far we have explored in detail what the role of the Uninterruptible Power Supply (UPS) is [in Criticality and redundancy in the data center] and we have analyzed the two main types of uninterruptible power supply systems - static and rotary [in Pick your UPS flavor].

It is time to dig deeper into the different types of static UPS systems. Static UPS systems are the most widespread type of UPS and are responsible for keeping the majority of data centers online in the event of a failure from the grid.

As we have learned previously the design of a data center power delivery system is entirely dependent on the continuity requirements of the IT load the data center hosts. As it is usually the case the IT load will require 24/7 availability, the power delivery system will be built to Tier 3 or Tier 4 standards.

Static UPS systems, compared to their rotary counterparts, do not have any moving parts along their power conversion and distribution systems. The most common types of static UPS systems are as follows:

- **Standby (or off-line) UPS**
- **Line Interactive UPS**
- **Double conversion (or on-line) UPS**

We will cover each of these types of static UPS with a description, a topology diagram, their pros and cons as well as the IEC 62040-3 output performance.

The IEC 62040-3: UPS - method of specifying performance and test requirements is an established standard that we have covered previously in issue #50 and further classifies the UPS in three categories - **Input dependency (AAA)**; **Output Voltage Waveform (BB)**; **Dynamic Output Performance (CC)**.

The standby UPS

The standby UPS (also called off-line) is commonly used to protect loads such as desktop computers. As shown in the diagram below the transfer switch is set to work on

the AC input as the power source (solid line). Should a problem occur with the primary source, the transfer switch changes the power delivery system to switch to the batteries. The name of "standby" comes from the fact that the inverter only starts once the primary power path fails - the inverter effectively "stands by" to carry the load.

A static UPS is characterized by relatively low cost and a small footprint. It further boasts non-sinusoidal output waveforms and high efficiency. The majority of static UPS systems available on the market also provide surge protection and electrical noise filtering.

The IEC standard classifies the input dependency (AAA) as VFD (Voltage and Frequency Dependent) because the standby UPS lacks circuitry to regulate output voltage or frequency when operating in normal mode.

Further, when it comes to output waveform (BB) the IEC standard classifies this as SY because, in normal mode, it passes the sinusoidal AC input voltage through to the IT load. While in energy mode this type of UPS utilizes a square or step wave inverter (a device that converts DC voltage into AC voltage).

Lastly, the dynamic output performance according to the IEC standard is 33. This is because it typically takes around 10 milliseconds for the system to detect a power

failure, move the transfer switch and turn on the inverter.

The complete classification for this type of UPS as per IEC 62040-3 is VFD-SY-33.

The line-interactive UPS

Line interactive UPS systems are best suited for protecting IT equipment present in data centers like servers and associated networking equipment like switches and routers.

The block diagram below shows the inner workings of a line interactive UPS and the main differentiation between this and the standby UPS is that the converter (an inverter or a rectifier capable of processing power in both directions) is constantly connected to the output of the UPS. Should the primary source fail, the transfer switch opens and power flows from the batteries to the IT load.

It is important to note that because the converter is constantly connected to the output of the UPS, switching between power sources yields reduced switching transients as well as increased filtering of electrical noise.

The line interactive UPS system design makes use of a tap-changing AVR (automatic voltage regulator) transformer. By adjusting the transformer taps in line with the input voltage the whole system now has voltage regulation. This is of paramount importance within low

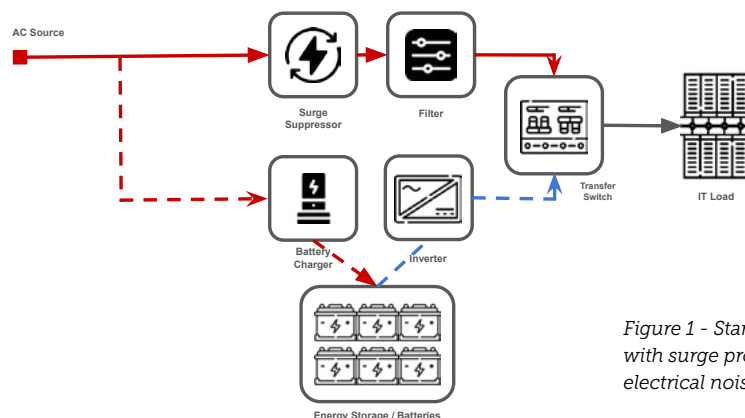


Figure 1 - Standby UPS with surge protection and electrical noise filtration

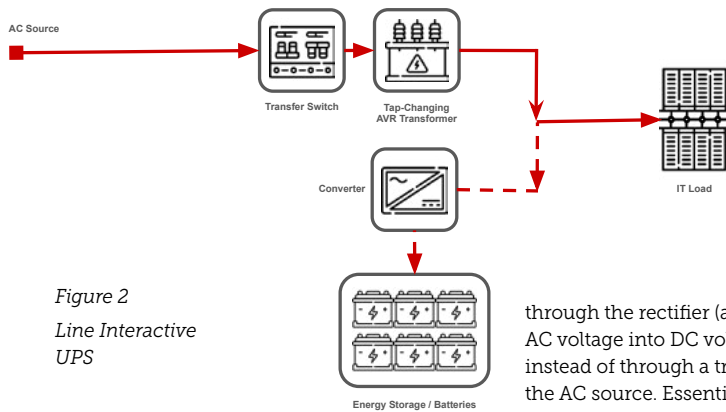


Figure 2
Line Interactive
UPS

voltage conditions, otherwise the UPS would transfer to the battery and in turn shut down the load.

Its small footprint coupled with its high efficiency and high reliability as well as boasting the ability to correct voltage in low or high voltage conditions make this UPS the ideal candidate within the 0.5 to 5kVA power range.

The IEC classification for this type of UPS is **VI-SS-31**. Because of the AVR transformer regulating input voltage the standard classifies the input dependency as Voltage Independent (VI).

Output waveforms are classified as SS because it passes the sinusoidal AC input voltage through to the load when operating in normal mode and utilizes a sine wave inverter in stored energy mode.

Finally, the dynamic output performance is 31 because it usually takes about 10ms to detect a power failure, open the transfer switch and reverse the power flow through the inverter - which handles step loads very well.

The double conversion UPS

Above 10kVA the double conversion UPS dominates the landscape as it is best used to protect large swathes of servers along with networking and storage equipment.

The types of static UPS we have looked at - line interactive and standby - have two modes of operation - normal and energy storage mode. On the contrary, the double conversion UPS will often boast two additional modes of operation - a high-efficiency bypass mode and a high-efficiency PFC (Power Factor Corrected) bypass mode.

Each mode of operation in a double conversion UPS has slightly different power flows, transfer mechanisms and IEC classifications.

The on-line normal mode double conversion UPS is represented in figure 3. The layout is similar to a standby UPS with the only difference being that the power load is fed

through the rectifier (a device that converts AC voltage into DC voltage) and the inverter instead of through a transfer switch fed from the AC source. Essentially the current goes through a double conversion (AC-DC-AC), hence the name.

Should the primary path fail, the rectifier shuts down and the inverter starts drawing power directly from the batteries then feeding the IT load. The primary advantage of this layout is that there is no change in the power path should the primary source fail, no switches open or closed which means the transfer time sits at 0ms. This type of UPS running in on-line normal mode provides almost ideal electrical output performance.

The full IEC standard output performance is classified as **VFI-SS-11**. This is due to the fact that the on-line double conversion UPS operating in normal mode regulates both output voltage and frequency hence it is a voltage and frequency independent system (VFI). Further, because the sinewave inverter powers the load in every mode of operation the output waveform is classified as SS. Finally dynamic output performance is 11 because the inverter continuously supplies power to the load without interruption and can also handle step loads very well.

The double conversion on-line UPS operating in bypass normal mode is illustrated in figure 4. The power path through the load happens through the static bypass transfer switch with the rectifier operating in low power mode to keep batteries charged while the inverter is in standby.

Should the primary power source fail, the transfer switch opens, the inverter kicks in and starts supplying the load directly from the batteries.

Operating this type of UPS in bypass normal mode improves the efficiency of the UPS as well as extending its service life since less component stress is being administered along with a reduced heat output. However, operating in bypass mode means that any power quality problems felt in the primary power path will impact the load directly as the inverter is not in operation.

The full IEC standard output performance for this type of UPS in the bypass normal mode is **VFD-SS-31**.

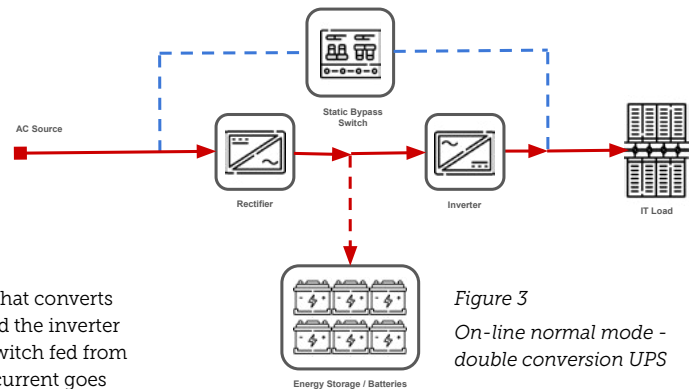


Figure 3
On-line normal mode -
double conversion UPS

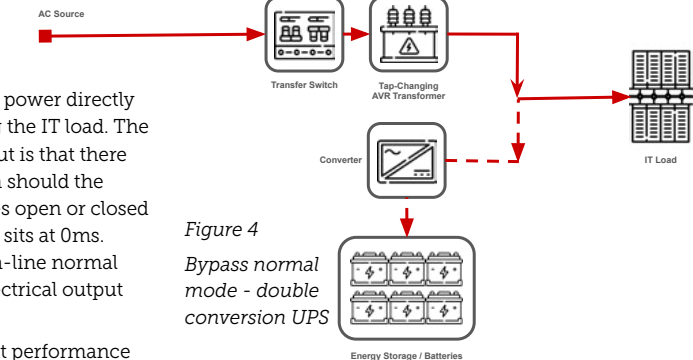


Figure 4
Bypass normal
mode - double
conversion UPS

Because of the use of the transfer switch on the primary power path it does not regulate voltage or frequency, thus being voltage and frequency dependent (VFD). Furthermore, because both power paths end up supplying an AC sine wave, the waveform output is SS. With the transfer time being around 10ms and the inverter handling step loads well, the dynamic output performance is 31.

The PFC bypass mode differs from the usual bypass mode by keeping the inverter online, providing additional required current to ensure that the load receives only sinusoidal current with a high power factor. This alleviates some of the drawbacks of the previous mode of operation and makes the transfer instantaneous. Because of a 0ms transfer time the IEC classification for a double conversion PFC bypass mode UPS system is **VFD-SS-11**.

UPS technology keeps evolving in line with the exponentially increasing demand for clean, uninterrupted and reliable power delivery across a swathe of industries.

Niched approaches like delta conversion, hybrid conversion or ferro-resonant UPS designs further seek to increase efficiency and cater to all proven use cases.

The systems and technologies outlined have long been at the core of enabling today's digital society and will continue to play a functional and foundational role in its evolution as the data center landscape keeps responding to the insatiable demand for data.

■ The good times



The AI goldilocks zone

I have just returned from Nvidia's GTC event in Silicon Valley, a paean to excess and exuberance in the age of artificial intelligence.

For a journalist, my emotions are mixed. On the one hand, as the editor of a publication focused on digital infrastructure, the current sector boom has been nothing but good for us.

We have more to write about than ever before, the ending of industry norms means that data center operators need an independent voice like never before. And, while the editorial team is separate from DCD's wider commercial operations, ballooning company budgets have undoubtedly helped us expand our team to eleven journalists.

It is therefore unsurprising that I have a vested interest in wanting the good times to keep on going. Like any sensible business, we are trying to work out how to take advantage of the current growth opportunity, while being careful to not overexpand if this ends up being a bubble that pops hard.

At the same time, however, another part of me hopes that AI's growth does temper somewhat. As someone that makes a living primarily from writing, and that employs a team of wonderful humans that do the same, it is undeniably unsettling to see the rapid progress of large language models.

I remain convinced that current generative AI approaches have no hope of matching what my team does. Even if hallucinations and accuracy are

solved, the simple act of stringing words together is not what makes us journalists. We work with sources (if you have a tip, get in touch), FOIAs, boots on the ground reporting, and more to add to the Internet's collective knowledge - rather than just rearranging existing datasets.

But the reality is that AI doesn't need to be as good as us to undermine us. The death of link-based search engines would be a death knell for many a publication, while an explosion of AI-generated publications could drown out the human-crafted content, pushing us to the brink.

I find myself wishing for a likely fantasy: A level of AI that keeps growing and financing the next wave of the Internet, but that never quite achieves the intellectual capabilities to displace myself and those around me.

This is a somewhat selfish view, and one that we try to keep out of our reporting - where we aim for the classic neutral news approach that is mostly impossible. But I suspect that I will soon not be alone in this view.

I recently visited the labs of a major hyperscaler where a researcher told me that they were developing robots for lights out data centers. Others are doing the same.

As more jobs in the sector face this moment, where the boom they were a part of fueling starts to render them irrelevant, it may cause many of us to look back at this time as the good ol' days. ■

- Sebastian Moss, Editor-in-Chief



Optimize Your Colocation Operations

Introducing TS IT Colocation Enclosures

- Designed specifically for Colocation installations
- Readily available, in-stock solutions
- Simplified ordering and installation
- Flexible, scalable, and modular
- Comprehensive, standardized designs



**LEARN MORE ABOUT
THE TS IT COLO ENCLOSURE**

Scan the QR code to get started.

