

CASE STUDY

Liquid Cooled AI Workload at VIRTUS



LIQUID COOLED AI WORKLOAD

VIRTUS Data Centres (VIRTUS) is a trusted global leader in the data centre industry and the UK's leading data centre company, committed to delivering reliable, resilient and responsible digital infrastructure for our customers. As part of ST Telemedia Global Data Centres (STT GDC), VIRTUS Europe is the home of the cloud in London, and at the heart of Europe.

Our Design and Solutions teams work in collaboration with all our customers to create truly tailor-made solutions that benefit from their combined industry leading expertise. This design could be for an entire facility, heightened security or solving complex cooling and power requirements such as those posed by new and developing Al/Machine Learning (ML) workloads.

Overview

Customer A - an existing customer of VIRTUS - approached with a multi-megawatt AI workload with a power density beyond the capability of air-cooling back in 2020.

The project was way beyond the usual capability that a traditional data centre could offer. VIRTUS' facilities are designed with flexibility in mind, enabling alternative technologies to be deployed as required. This allows customers to benefit from agile, flexible solutions that are tailored to their exact requirements.

This mission critical AI deployment had key delivery contract milestones, with very demanding efficiency performance criteria and commissioning requirements.

Challenge

The challenges for both VIRTUS and Customer A, were across multiple levels:

- > Weight of deployment due to the multitude of AI servers
- Extremely high kW density per rack, up to 400kW
- Cooling design capable of nearly instantly responding to fluctuating compute demand
- The need for standard networking IT equipment to be at close proximity to the High Performance Computing (HPC) workload
- Latency sensitive, meaning the workload needed to be able to seamlessly migrate between sites

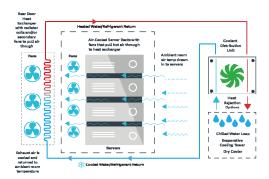
The VIRTUS Solution

Two cooling technologies were implemented to support two styles of IT equipment at close proximity to each other. VIRTUS' standard design was deployed to support the networking IT servers. This therefore afforded Customer A the ability to expand these racks within the overall environment as standard.

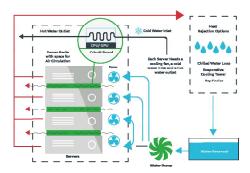
The main HPC deployment required liquid cooling to support the extremely high-powered AI liquid cooled racks. Working with Customer A and the equipment manufacturer, VIRTUS implemented a direct-to-chip liquid cooled solution.



Liquid cooling for racks can be provided via rear door heat exchangers, or direct to chip for very high density racks.



Rear Door Heat Exchanger Cooling



Direct-to-Chip Single Phase

Weight

VIRTUS' designs its data centres to be fully flexible, allowing for not only cooling and power changes, but also structural upgrades. The environment designed and built for Customer A's AI deployment meant that the structural challenges posed by the weight of the deployment were not an issue. The site was also upgraded to support multiple racks with more than three times the weight of standard aircooled racks.

Power

Customer A wanted to minimise the environmental impact of this multi-megawatt system. VIRTUS only sources power from 100% renewable sources, and combined with the overall HPC design, was able to deliver an extremely efficient operational PUE.

Commissioning

Customer A was required to demonstrate that this AI deployment could successfully operate at full capacity, and be able to ramp up and down on demand, in line with varying compute requirements. In air cooled deployments, this is typically achieved by placing heaters in the data halls to simulate the cooling demand of the IT load. However, for this specific deployment, an alternative solution was required to simulate the demand from the liquid cooled racks and demonstrate the supporting mechanical plant could accommodate rapidly fluctuating demand. This was achieved by the development of an industrial boiler system, and the addition of a thermal store within the cooling system to ensure sufficient volume of chilled water was available should the liquid cooled system rapidly demand more capacity. This was combined with a series of operational tests, to successfully prove that

the design would meet these demands, along with a low operational PUE prior to the actual equipment being deployed, without any risk to Customer A's equipment.

Benefit

- > The experience gained by the VIRTUS team to design similar solutions in both operational and future sites that are under construction
- > The project exposed our Operations teams to Al workloads within mixed cooling technologies and dense environments, further demonstrating VIRTUS' agile design, allowing HPC and standard/traditional IT equipment in the same room at very close proximity
- > Extremely low PUE of less than 1.1
- > Supports large Generative AI workloads that are latency sensitive
- In operation since 2021 this early AI deployment provided VIRTUS with valuable knowledge gained during the design, build, commissioning and live operational management
- > Future data centre assets have been and are being designed leveraging this at scale experience, and has afforded VIRTUS a significant head start on a market with growing AI/ML demand

For more information

To learn more about VIRTUS Data Centres, their facilities, migration packages and how they can help your business, visit www.virtusdatacentres.com

