1/22/2025

608-238-6001 [TEL]

greg@infinityturbine.com [Email]

xai-colossus-cluster-memphis-tn-nvida-h100-by-infinity-turbine

Infinity Turbine LLC

The Xai Colossus MegaCluster in Memphis TN Featuring 100000 Nvida H100

This webpage QR code

Structured Data

<script type= "application/ld+json">
{"@context":"http://schema.org",
"@graph":[
{
"@type" : "Organization",
"@id" : "https://infinityturbine.com/#organization",
"name" : "Infinity Turbine LLC",
"url" : "https://infinityturbine.com",
"sameAs":
["https://www.youtube.com/channel/UCsobpvy0xqc13uvhA71Cv4w",
"https://x.com/InfinityTurbine",
""],
"telephone" : "608-238-6001",
"email" : "greg@infinityturbine.com",
"logo" : "https://infinityturbine.com/logo.png"
},
{
"@type":"WebSite",
"@id":"https://infinityturbine.com",
"url":"https://infinityturbine.com",
"name":"The Xai Colossus MegaCluster in Memphis TN Featuring 100000 Nvida H100",
"description":"Discover the xAI Memphis Megacluster, Elon Musk's AI-driven data center at 3231 Riverport Rd in Memphis, TN. Utilizing 100,000 Nvidia H100 GPUs, this facility, in collaboration with Phoenix Investors, sets a new standard in supercomputing and AI technology, supporting regional economic growth and innovation."
}
,{
"@type":"NewsArticle",
"mainEntityOfPage":{
"@type":"WebPage",
"@id":"https://infinityturbine.com/xai-colossus-cluster-memphis-tn-nvida-h100-by-infinity-turbine.html"},
"headline":"The Xai Colossus MegaCluster in Memphis TN Featuring 100000 Nvida H100",
"image":"https://infinityturbine.com/images/",
"datePublished":"2025-01-22T08:00:00+08:00",
"dateModified":"2025-01-22T09:20:00+08:00",
"author":{
"@type":"Organization",
"name":"Infinity Turbine LLC",
"url":"https://infinityturbine.com"
},
"publisher":{
"@type":"Organization",
"name":"Infinity Turbine LLC",
"logo":{
"@type":"ImageObject",
"url":"https://infinityturbine.com/logo.png"
}}}

]}</script>

Discover the xAI Memphis Megacluster, Elon Musk's AI-driven data center at 3231 Riverport Rd in Memphis, TN. Utilizing 100,000 Nvidia H100 GPUs, this facility, in collaboration with Phoenix Investors, sets a new standard in supercomputing and AI technology, supporting regional economic growth and innovation.

PDF Version of the webpage (first pages)

Xai Memphis SuperCluster

Xai Memphis

xAI, 3231 Riverport Rd, Memphis, TN 38109

The construction and development of the data center involve multiple stakeholders:

xAI: As the owner and operator, xAI oversees the project's direction and integration of AI technologies.
Phoenix Investors: This Milwaukee-based real estate firm owns the facility and collaborates with xAI to repurpose the building for data center operations.
PHOENIX INVESTORS
Local Authorities and Utilities: Entities like the Greater Memphis Chamber and Memphis Light, Gas & Water (MLGW) are involved in facilitating the project's infrastructure needs, including power and water supply.
WREG
The project has progressed rapidly, with the data center reportedly going live in July 2024, utilizing up to 100,000 Nvidia H100 GPUs.
DATACENTER DYNAMICS
However, specific details about the construction companies or contractors responsible for the physical build-out have not been publicly disclosed.

https://phoenixinvestors.com/phoenix-companies/elon-musk-xai-and-a-big-supercomputer-in-memphis-a-look-at-plans-for-the-massive-project/

Elon Musk’s xAI is building an Artificial Intelligence Center in a Phoenix Investors facility in Memphis, Tennessee; Phoenix purchased the 785,000-square-foot facility in December 2023 from Electrolux Group. With this investment, Memphis will be home to the world’s largest supercomputer. The project is expected to open in late 2024 and boost jobs and economic activity in the Memphis area.

https://www.commercialappeal.com/story/money/business/2024/06/05/xai-supercomputer-memphis-electrolux-plant/73987730007/

However, the former Electrolux facility at 3231 Paul Lowry Road is the presumed location for xAI. Milwaukee-based real estate company Phoenix Investors was among the entities thanked during the announcement.

https://wreg.com/news/local/xai-memphis/memphis-chamber-to-announce-major-economic-development/

https://www.electroluxgroup.com/en/electrolux-group-to-divest-memphis-factory-40343/

Elon Musk's artificial intelligence company, xAI, developed the Colossus supercomputer in Memphis, Tennessee, utilizing 100,000 Nvidia H100 GPUs. The project was completed in 122 days, with the system going live in July 2024.
BLOCKONOMI

While specific details about the contractors responsible for the physical build-out and installation of the GPUs and CPUs have not been publicly disclosed, xAI's internal engineering team played a significant role in the rapid deployment of the supercomputer. Nvidia CEO Jensen Huang praised Elon Musk and the xAI team for their "superhuman" effort in building the supercomputer in such a short timeframe.
BUSINESS INSIDER

Additionally, xAI has been actively recruiting for positions related to data center construction and management in Memphis, indicating a combination of internal resources and external contractors may have been utilized.
GREENHOUSE

https://www.electroluxgroup.com/wp-content/uploads/sites/2/2023/09/electrolux-Electrolux-Group-to-divest-Memphis-factory-230921.pdf

1/22/2025

Summary

Elon Musk’s artificial intelligence company, xAI, is developing the xAI Memphis Megacluster, a state-of-the-art data center located at 3231 Riverport Rd in Memphis, Tennessee. This ambitious project, housed in a former Electrolux facility acquired by Phoenix Investors, has rapidly become a major milestone in AI and supercomputing. Launched in July 2024, the center leverages 100,000 Nvidia H100 GPUs to power what is touted as the world's largest supercomputer, designed to meet the demands of advanced AI research and development. This massive infrastructure initiative promises significant job creation and economic growth in the Memphis area, supported by local utilities and infrastructure stakeholders.

1/22/2025

How Tesla Megapack Grid-Scale Batteries are Powering XAI's Colossus Data Center in Memphis, Tennessee

In a significant development in the data center industry, XAI has announced its integration of Tesla's Megapack grid-scale batteries as an energy buffer at the Colossus Data Center in Memphis, Tennessee. This partnership marks a pioneering approach to energy management for data centers, focusing on sustainability, reliability, and efficiency.

Why Data Centers Need Battery Buffers

Data centers like Colossus are the lifeblood of today's digital infrastructure, housing critical information and serving as hubs for cloud computing, artificial intelligence, and machine learning. Powering a data center of this scale requires immense energy resources, not only to operate the servers but also to manage the high cooling demands. With the increasing intensity and frequency of power demands, especially in energy-intensive fields like AI, stability in power delivery becomes critical.

The Tesla Megapack is a grid-scale battery system designed for large-scale applications, capable of storing massive amounts of energy and discharging it to stabilize power grids or offset peak demand. In the case of Colossus, the Tesla Megapack serves as an energy buffer, helping to manage fluctuations in energy demand and acting as a sustainable backup power solution.

The Role of Tesla Megapack at Colossus Data Center

The integration of the Megapack system is a strategic move to ensure that Colossus has a reliable, stable power source without relying solely on the grid or traditional backup generators. Here's how the Tesla Megapack system is enhancing operations:

1. Energy Buffer for Peak Loads:
The Megapack system stores energy during off-peak hours when energy is less expensive or sourced from renewables, such as solar or wind. This stored energy can then be discharged during peak hours to offset power demands, effectively smoothing out power spikes. For a data center, this ensures consistent power for server operations, even during periods of peak usage.

2. Emergency Backup Power:
In the event of a grid failure or other power interruptions, the Megapack provides a buffer of stored energy to ensure that operations continue uninterrupted. Unlike traditional backup generators that run on diesel or natural gas, the Megapack offers a cleaner alternative that supports XAI's sustainability goals.

3. Sustainability Initiatives:
With Tesla's Megapack, XAI is moving towards a more sustainable energy model. The Megapack allows Colossus to operate on renewable energy sources more efficiently, capturing surplus energy from renewable resources and using it later. This approach aligns with broader industry trends of reducing carbon emissions and environmental impact.

4. Reduced Operating Costs:
By integrating the Megapack system, Colossus reduces its dependency on grid energy during high-cost peak periods. The ability to store and discharge energy strategically helps reduce operational expenses, which is especially valuable for data centers facing high cooling and energy demands associated with AI and machine learning applications.

Technical Insights into Tesla Megapack's Role at Colossus

The Colossus data center in Memphis, Tennessee

The Colossus data center in Memphis, Tennessee, operated by xAI, has been described as one of the world's most advanced supercomputers.

Here are some details based on various reports and posts:

Location: The data center is housed in a repurposed manufacturing plant, previously owned by Electrolux, in Memphis, Tennessee.
Purpose: It's designed to power xAI's AI systems, including the "Grok" AI, aiming to accelerate human scientific discovery.
Computing Power: Colossus features up to 100,000 Nvidia H100 GPUs, making it one of the largest supercomputers by GPU count. There's also mention of plans to double its capacity with an additional 100,000 Nvidia Hopper GPUs, although this might be in reference to future expansions or a different metric.
Energy Consumption: The facility consumes up to 150 megawatts of power, which is enough to power approximately 100,000 homes. This has sparked discussions and concerns about its environmental impact.
Cooling and Water Usage: Initially, it was reported that Colossus might need 1 million gallons of water daily for cooling, which raised environmental concerns. However, there's mention of a "Colossus Water Recycle Facility" project aimed at reducing this dependency. This project involves a collaborative effort to treat wastewater for data center cooling, indicating an initiative towards sustainability.
Environmental and Community Impact: The data center's establishment has been controversial due to its location near historically Black neighborhoods with pre-existing environmental health issues, like high asthma rates. There's been criticism regarding its potential to exacerbate local pollution and concerns over its environmental review process.
Networking: It utilizes NVIDIA Spectrum-X Ethernet networking for its high-performance computing needs, showcasing advanced technology in data transfer and connectivity for AI workloads.
Economic Impact: Described as potentially the largest capital investment in Memphis's history, it's expected to bring significant economic activity, although specific job numbers or economic benefits directly linked to the data center haven't been quantified in the provided information.
Secrecy and Controversy: There's been a noted lack of transparency regarding the data center's operations and its environmental impact assessments, leading to community and environmental advocacy groups' concerns.

This data center represents a massive investment in AI infrastructure, aiming to push the boundaries of computational capabilities for artificial intelligence research and application. However, its establishment also highlights significant concerns regarding environmental impact, community health, and corporate transparency in such large-scale tech projects.

1/22/2025

Cooling

At the Colossus data center, which supports xAI's supercomputer cluster, an advanced liquid cooling system is employed. This approach to cooling is pivotal due to the immense computational power housed in the facility, primarily from over 100,000 NVIDIA H100 GPUs. Here's how the cooling is managed:

• Liquid Cooling Technology: The data center uses a direct-to-chip liquid cooling method, where liquid coolants are circulated directly to the components generating the most heat, like GPUs. This method was highlighted as a collaboration between xAI, Supermicro, and NVIDIA, showcasing a significant shift from traditional air cooling to more energy-efficient liquid cooling solutions.

• Rack Configuration: Each rack in the Colossus data center is designed with liquid cooling in mind. Supermicro provided racks where each rack could contain 64 GPUs, with the cooling system integrated into the rack design. This setup allows for easy maintenance and service without having to remove systems from the rack, thanks to quick disconnects for the liquid cooling.

• Efficiency and Scalability: The liquid cooling not only addresses the heat output from such a high number of GPUs but also supports the rapid scalability of the data center. The design facilitates adding more computing power with less concern over thermal management, which is crucial for xAI's plans to expand.

• Energy Considerations: The adoption of liquid cooling aligns with broader trends in data center management towards more sustainable practices. While specific energy consumption figures for Colossus's cooling system weren't detailed, the shift to liquid cooling is generally aimed at reducing the overall power usage effectiveness (PUE), although implementing such systems can initially be more costly and complex.

• Environmental Impact: There's a mention of potential environmental concerns regarding water usage for cooling at such a scale, indicating that while liquid cooling might reduce energy use, the water cycle impact needs careful management.

This cooling strategy at Colossus not only supports the computational demands of AI training but also positions xAI at the forefront of data center technology, focusing on efficiency, scalability, and potentially, environmental sustainability. However, it's worth acknowledging that while liquid cooling reduces energy for cooling, the ecological footprint of water usage and the energy to manage it remains a point of discussion in the broader tech community.

1/22/2025

Heat Dissipated

To calculate the heat dissipated per hour by 100,000 H100 GPUs into the chiller system, we need to consider the power consumption of these GPUs. Here's how we can approach this:

• Power Consumption of H100 GPU: Each H100 GPU consumes up to 700W under peak load.

• Total Power Consumption:
• If all 100,000 GPUs were running at peak load simultaneously:
• $100,000 \text{ GPUs} \times 700 \text{ W/GPU} = 70,000,000 \text{ W}$

• Heat Dissipation: All the power consumed by the GPUs, assuming it's not being used for work (which is an oversimplification but common for heat calculations), turns into heat.

• Conversion to Heat per Hour:
• 1 Watt = 1 Joule per second.
• Therefore, $70,000,000 \text{ W} \times 3600 \text{ seconds/hour} = 252,000,000,000 \text{ Joules/hour}$.

• Heat in More Common Terms:
• 1 Joule = 0.239 calorie.
• $252,000,000,000 \text{ J} \times 0.239 \text{ cal/J} = 60,208,000,000 \text{ calories}$.
• In terms of refrigeration, 1 ton of refrigeration is approximately 12,000 BTU/hour, where 1 BTU ≈ 252 calories.
• $\frac{60,208,000,000 \text{ cal}}{252 \text{ cal/BTU}} = 238,920,238 \text{ BTU/hour}$.
• $\frac{238,920,238 \text{ BTU/hour}}{12,000 \text{ BTU/ton}} \approx 19,910 \text{ tons of refrigeration}$.

Given these calculations, if all 100,000 H100 GPUs were running at full load:

• Heat Dissipation:
• 252,000,000,000 Joules per hour or
• 60,208,000,000 calories per hour or
• Roughly 19,910 tons of refrigeration needed to handle this heat.

• Power Consumption: 70,000,000 W (100,000 GPUs at 700W each)
• Conversion Factor: 1 Watt = 3.412 BTU per hour

Thus:

• Total Heat Dissipation in BTU/hour:
• $70,000,000 \text{ W} \times 3.412 \text{ BTU/W-h} = 238,840,000 \text{ BTU/hour}$

So, if all 100,000 H100 GPUs are running at peak load, they dissipate 238,840,000 BTU per hour into the chiller system.

This calculation assumes continuous maximum load, which might not be the case in real-world scenarios where GPUs might operate below peak capacity or not all units might be active simultaneously. However, for an upper bound estimate, these figures represent the potential heat dissipation into the chiller system per hour.

1/22/2025

Other Cooling Methods

Approaches for advanced methods of cooling beyond traditional liquid cooling for GPUs and other high-performance computing components:

1. Immersion Cooling:
• Single-Phase Immersion: Components are submerged in a thermally conductive, but electrically insulating liquid. Heat is transferred from the components to the liquid, which is then circulated through a heat exchanger to cool it down.
• Two-Phase Immersion: Similar to single-phase but uses a liquid that evaporates as it absorbs heat (boiling point is just above the operating temperature of the components). The vapor then condenses on a cooler surface, releasing heat, and the cycle repeats.

2. Microfluidic Cooling:
• Involves embedding microchannels directly into or very near the silicon of the chip. Coolant is pumped through these microchannels, providing direct cooling to the heat source. This method can achieve very high heat transfer rates due to the close proximity of the coolant to the heat-generating components.

3. Heat Pipes:
• These are passive systems that use phase changes of a working fluid (evaporation and condensation) to efficiently transfer heat from one point to another. While not as advanced as some other methods, heat pipes are highly efficient for certain applications.

4. Thermoelectric Cooling (TEC):
• Uses the Peltier effect, where passing an electric current through junctions of two different metals or semiconductors causes heat to be transferred from one side of the junction to the other. This can be used for spot cooling of specific components.

5. Cryogenic Cooling:
• Involves cooling components to very low temperatures, often below -150°C. This is used in extremely high-performance computing where traditional cooling methods are insufficient. However, it's complex and expensive to maintain.

6. Jet Impingement Cooling:
• High-velocity jets of fluid (usually liquid) are directed perpendicularly onto the surface of the hot component. This method can handle extremely high heat fluxes due to the direct and forceful contact of the coolant with the heat source.

7. Graphene-based Cooling Solutions:
• Research into graphene, which has exceptional thermal conductivity, is ongoing. Prototypes have been developed where graphene is used to conduct heat away from processors more efficiently.

8. Phase Change Cooling:
• Uses materials that change phase (from solid to liquid or gas) at specific temperatures, absorbing a lot of heat during this process. While more common in domestic refrigeration, similar principles can be applied in advanced computing setups.

9. Integrated Photonic Cooling:

Grid Scale Battery Comparison

Based on the information gathered, here's a structured comparison between the Salgenx saltwater battery and Tesla's Mega pack:

Environmental Impact and Sustainability

• Salgenx: Utilizes saltwater, which significantly reduces the reliance on scarce minerals like lithium, cobalt, and nickel. This approach not only lessens the environmental footprint by reducing mining impacts but also leverages the abundance of saltwater, making it inherently more sustainable.

• Tesla Megapack: While Tesla focuses on recycling and sustainable sourcing, it still fundamentally depends on lithium-ion technology, which has environmental concerns related to mining and disposal.

Suggestion: Highlight the sustainability of using saltwater, emphasizing its lower environmental impact and how this could appeal to eco-conscious investors, governments, and consumers.

Safety

• Salgenx: Saltwater batteries are non-flammable, offering a significant safety advantage over lithium-ion batteries, which have been known to catch fire, as seen in some incidents with Tesla's products.

• Tesla Megapack: While robust safety measures are in place, the inherent flammability of lithium-ion batteries remains a concern, though mitigated through design and safety systems.

Suggestion: Market the safety aspect heavily, especially for installations in urban areas or near critical infrastructure, where fire hazards from lithium-ion batteries are a concern.

Cost and Efficiency

• Salgenx: The lower cost of production due to the use of common materials like saltwater could lead to more affordable energy storage solutions. Additionally, the absence of a membrane in its design reduces maintenance and initial costs.

• Tesla Megapack: While Tesla has managed to reduce costs through scale and technology, the raw material costs for lithium-ion batteries still pose a challenge.

Suggestion: Capitalize on the cost-effectiveness for large-scale projects or in regions where capital cost is a deciding factor. Offer detailed cost comparisons in marketing materials.

Scalability and Flexibility

• Salgenx: Offers scalability through increasing electrolyte volume. Its design also allows for integration with other technologies like ORC (Organic Rankine Cycle) for additional functionalities.

• Tesla Megapack: Highly scalable through modular design, but primarily focused on energy storage without additional functions like thermal storage or integration with other energy systems.

150 MW Power to Run Data Center

The xAI facility is demanding a jaw-dropping 150 MW of firm power by the end of 2024. To put that demand in perspective, 150 MW is enough electricity to power 100,000 homes.

1/22/2025