

# XFRA™

# White Paper

Arch Rao

Chris Lander





# Contents

CHAPTER	PAGE
01 <b>Executive Summary</b>	03
02 <b>Macro Challenge: Compute vs Power</b>	07
2.1 Exploding Compute Demand	
2.2 Grid Capacity & Local Interconnect Limits	
03 <b>What is XFRA?</b>	11
04 <b>SPAN: A Unique Unlock</b>	14
4.1 Unlocking existing electrical headroom	
4.2 Making headroom usable for always-on compute	
4.3 Rapid deployment at scale	
05 <b>Addressing the big questions</b>	17
5.1 Up-time reliability and continuity	
5.2 Sustainability and grid impact	
5.3 Homeowner value and fairness	
5.4 Security and trust	
06 <b>Target Workloads</b>	20
07 <b>XFRA Concept &amp; Architecture</b>	23
7.1 XFRA Node Architecture	
7.2 Fleet Orchestration and Power Events	
7.3 Why This Architecture Matters	
08 <b>Core Value Propositions</b>	26
09 <b>Next Steps</b>	30
10 <b>Risks</b>	32
11 <b>Conclusion &amp; Call To Action</b>	34

01

# Executive Summary

# The world's demand for compute is surging

The world's demand for compute is surging, driven by AI workloads, while the power infrastructure needed to supply it is increasingly constrained.

Expanding centralized data center capacity is slow and capital-intensive: projects can take years to plan, permit, and build, and are often further delayed by generation build-out, distribution limits, and utility interconnection bottlenecks. The result is a widening gap between compute demand and deliverable power.

4,430<sup>TWH</sup>

2025 ACTUAL GENERATION

+25-32%

DEMAND GROWTH BY 2030

80 vs 40<sup>GW/YR</sup>

BUILDOUT GAP

2,600<sup>GW</sup>

INTERCONNECTED QUEUE BACKLOG



XFRA flips this model by converting unused capacity in homes and small businesses into a distributed compute cloud.

On average, residential electrical infrastructure operates at roughly 40% of peak capacity, leaving substantial headroom untapped. The [SPAN smart panel](#) is the unlock, providing real-time, circuit-level visibility and control that allows XFRA to safely and dynamically access that headroom without compromising household needs.

Each XFRA Node pairs a high-performance compute module with a SPAN Panel, and workloads are coordinated through the XFRA Secure Orchestration Layer (XSOL), which securely schedules and routes jobs across the fleet, placing them based on customer latency requirements as well as each Node's real-time energy and hardware capacity. Nodes are designed for resilience with liquid cooling, whole-home battery backup and optional rooftop solar where advantageous.



- \* XFRA is not intended to replace hyperscale data centers, which will remain essential for large-scale AI training and other workloads requiring massive single-instance capacity. Instead, XFRA complements centralized infrastructure by accelerating capacity growth with distributed, dispatchable compute, well suited to AI inference, cloud gaming, content streaming, cloud rendering, and other geographically distributed workloads where rapid deployment, locality, or latency matter. By combining compute with energy storage and controllable load at the edge, XFRA can also reduce peak grid strain and support demand response participation where available.

XFRA is already operating revenue-generating test units and plans to launch a 100-home proof of concept later this year, with a roadmap toward broader U.S. deployment beginning in 2027. XFRA will also expand to larger commercial Nodes and multiple deployment options following initial deployment. Beginning in 2027, XFRA will scale to an annual capacity in excess of 1 gigawatt of AI inference compute, accessing a robust pipeline of new construction starts and commercial Nodes across the country.

02

# Macro Challenge: Compute vs Power

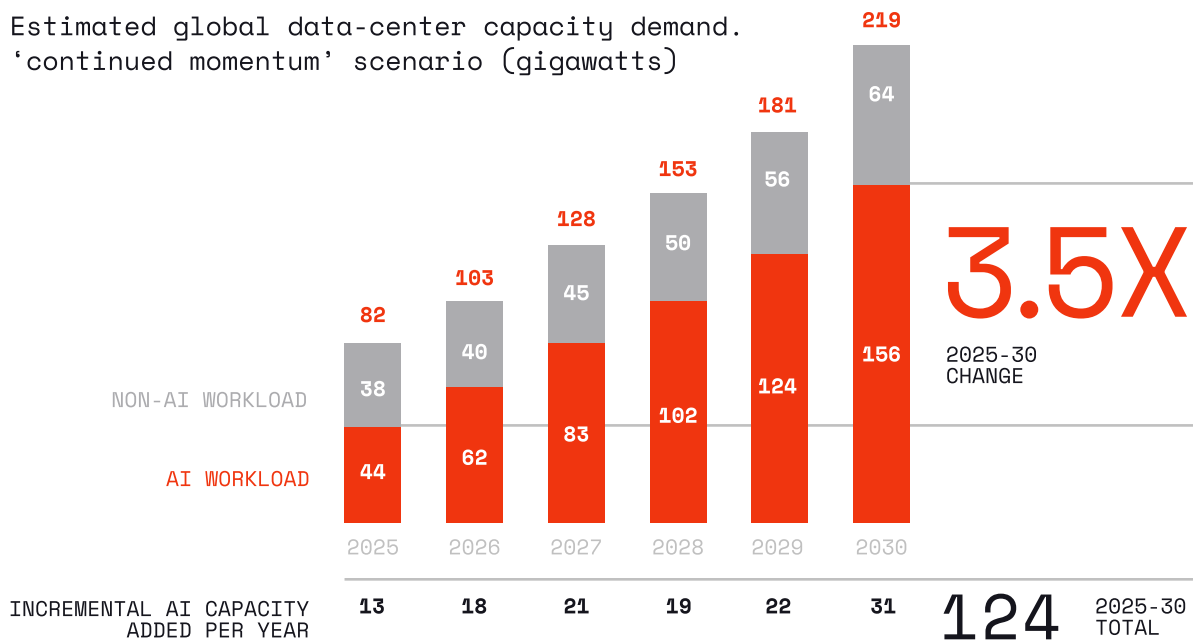
## \* 2.1 Exploding Compute Demand

AI model sizes for training continue to grow rapidly, with many doubling in scale every few months, demanding more compute bandwidth and pushing the limits of even the most advanced data center compute hardware. Other workloads like AI inference including Retrieval-Augmented Generation (RAG) and cloud gaming are experiencing very high rates of demand growth and are consuming GPU compute bandwidth at record rates as services achieve mass adoption.

While AI training continues to demand the largest and most complex data center hardware available, inference models, especially the latest architectures using Mixture of Experts (MoE) and quantization techniques, are becoming more modular and efficient. This trend makes them well-suited for deployment on edge nodes like XFRA.

Both AI and Non-AI workloads will be key drivers of global datacenter capacity demand growth through 2030

Estimated global data-center capacity demand.  
'continued momentum' scenario (gigawatts)



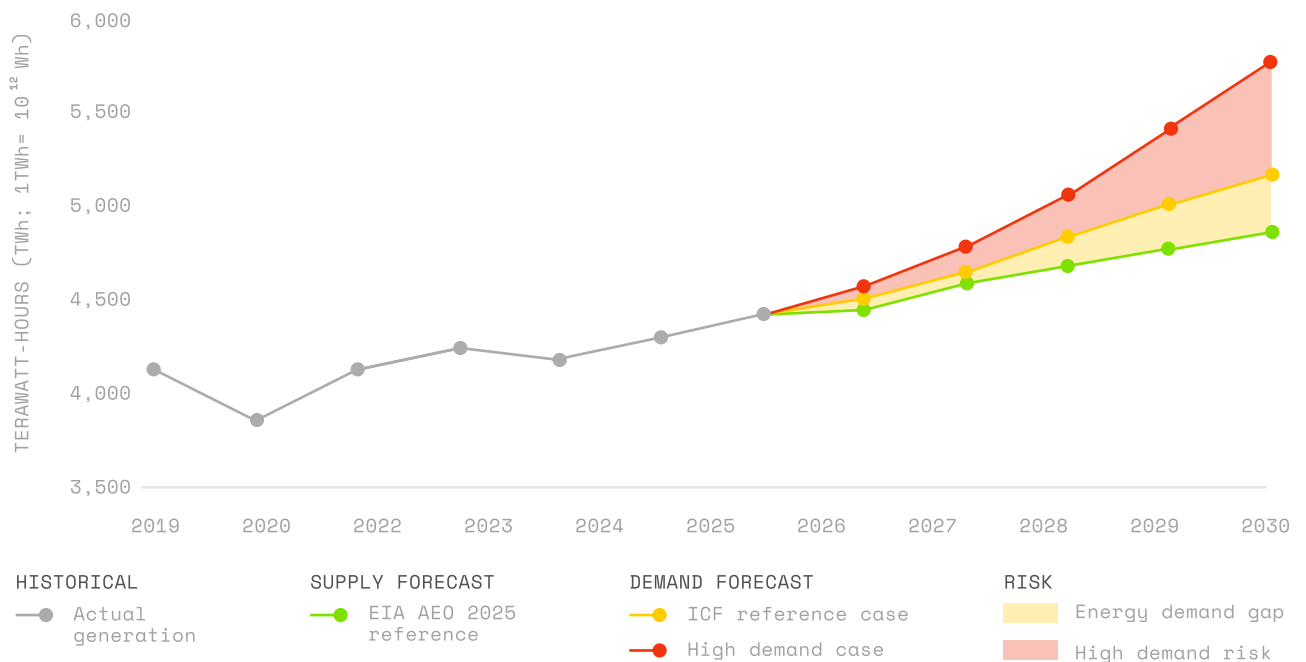
NOTE: Figures may not sum to totals due to rounding

SOURCE: McKinsey Data-Center Demand Model; Gartner Reports; IDC Reports; Nvidia Capital Markets Reports

## \* 2.2 Grid Capacity & Local Interconnect Limits

Despite planned energy generation buildouts, global compute demand is projected to outpace available power generation for decades. In some regions of the world (including many parts of the US), this situation is exacerbated by the retirement and decommissioning of older fossil fuel and nuclear power plants.

### Generation supply vs demand scenarios – U.S. electricity 2019-2030



**SOURCES:**

**Historical:** EIA Monthly Energy Review, Table 7.6 (eia.gov/totalenergy/data/monthly),  
**Supply forecast:** EIA Annual Energy Outlook 2025 Reference Case, Table AB (eia.gov/outlooks/aeo/tables\_ref.php); near-term growth from EIA STEO March 2026 (+1.2% in 2026, +3.1% in 2027),  
**Reference demand:** IF / Pew Charitable Trusts, America's Electricity Future, Sep 2025 (+25% by 2030 from 2023 base). **High demand:** Grid Strategies, National Load Growth Report, Dec 2025 (+32% by 2030; values extrapolated from published growth rate - treat with caution).

✱

Even where total generation is sufficient, distribution bottlenecks choke new compute capacity build-outs. Data centers require large interconnects to the power grid, ranging from tens to hundreds of megawatts, along with transmission upgrades, substations, transformers and sometimes even new local generation. Permitting, equipment lead times and utility interconnection queues are increasingly extending build timelines by years.

The result is a “hyperscaler paradox:” compute demand is scaling quickly, but even if data center build-out can keep up with demand, the power delivery infrastructure to feed those data centers can only scale at utility build-out speed.

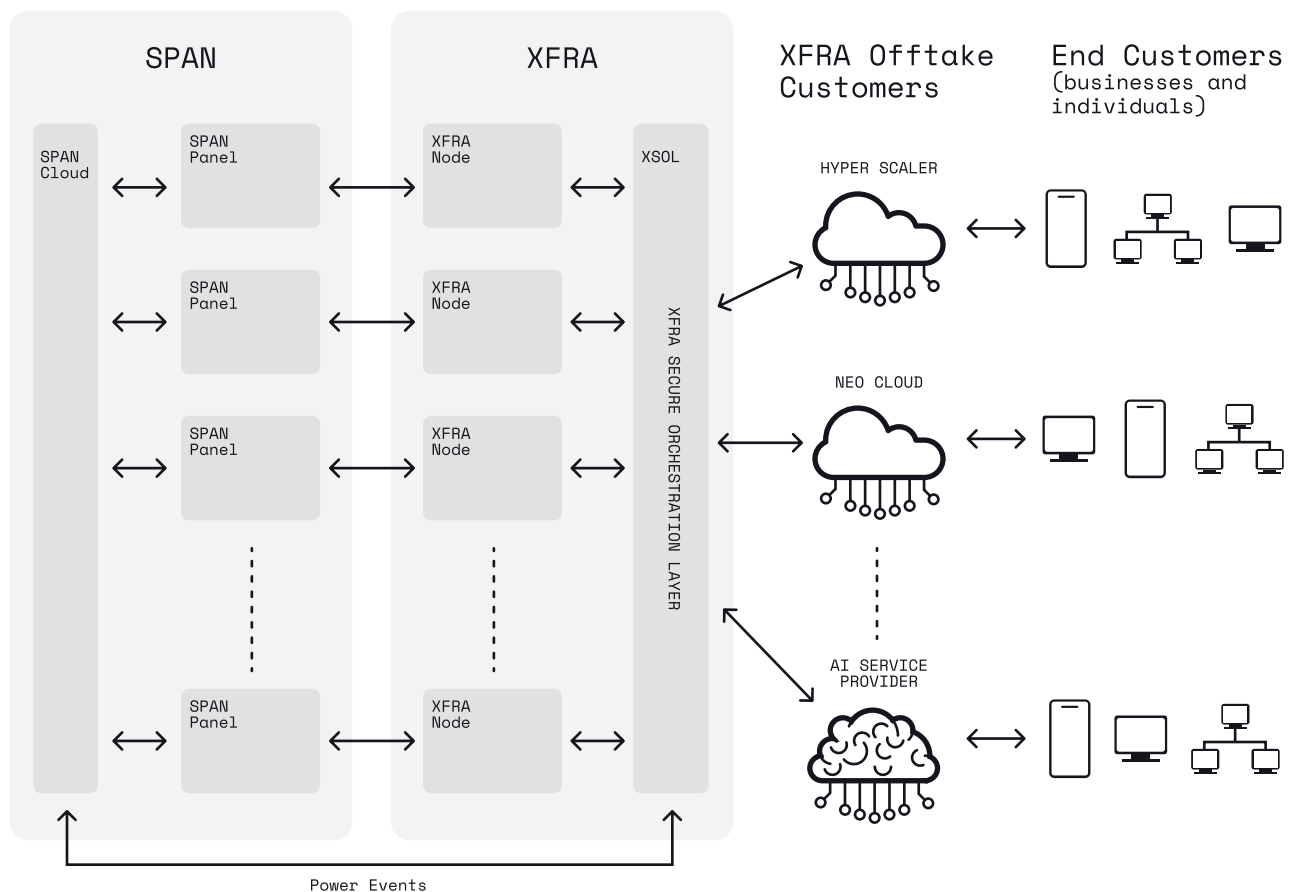
For compute buyers, this constraint shows up as a practical reality: capacity will not always be available where and when it’s needed, and new supply is increasingly gated by power availability.

03

# What is XFRA

\* **XFRA is a distributed compute cloud built from a fleet of XFRA Nodes deployed in homes and small businesses.**

XFRA converts available electrical headroom at the edge into high-demand compute by pairing cutting-edge GPU capacity with real-time energy visibility and control. Workloads are coordinated through the XFRA Secure Orchestration Layer (XSOL), which securely connects compute offtake customers to the fleet and makes many independent Nodes behave like a coherent cloud.



- \* Each XFRA Node integrates XFRA GPU Compute Modules with a high-efficiency cooling system and a local management gateway, and is paired with a SPAN smart panel and whole-home battery. XFRA is designed as a modular system. Nodes are deployed as repeatable building blocks that can be added incrementally across a fleet. Within each Node, capacity scales through modular XFRA GPU Compute Modules, allowing different deployment sizes such as residential and small-business configurations, while using the same core architecture and orchestration layer.

The SPAN smart panel enables XFRA by unlocking unused electrical headroom through real-time, circuit-level visibility and control. This allows XFRA Nodes to be provisioned as always-on loads in standard residential electrical service, while SPAN continuously monitors site conditions and maintains safe operating limits.

The whole-home battery safeguards compute workloads during power outages, buffers short-term demand spikes and responds to utility demand response events. Rooftop solar can also be integrated where advantageous.

At the fleet level, XSOL schedules and routes workloads across Nodes based on customer needs (including latency), GPU and system availability, and real-time energy constraints at each site. XSOL also coordinates graceful workload failover when necessary, ensuring reliable service even when individual Nodes must throttle or pause due to utility grid events.

04

**SPAN's**

**Unique Unlock**

## \* 4.1 Unlocking existing electrical headroom

Residential electrical infrastructure is designed for rare peak conditions, while typical utilization is far lower, leaving substantial capacity underused across the residential grid. In practice, Data from the SPAN installed base shows that standard 200A service homes consistently maintain significant headroom under normal operating conditions, sufficient to support an XFRA Node as a provisioned, always-on load.

The SPAN smart panel is the unlock: it provides real-time, circuit-level visibility and control that makes this headroom measurable and safely usable at scale. SPAN Panels have been certified for performance and safety under UL 3141 certification, the highest safety standard for Power Control Systems (PCS), a category of devices that manage electrical loads to keep power within defined limits. By pairing XFRA Nodes with SPAN smart panels, XFRA can convert underutilized behind-the-meter capacity into high-value compute, leveraging existing infrastructure rather than waiting for new large-scale interconnects and centralized build-outs.

## 4.2 Making headroom usable for always-on compute

Available headroom only becomes useful for compute when it can be operated safely and predictably, like other infrastructure. XFRA achieves this by combining the SPAN circuit-level visibility and control, battery-backed resilience and XSOL fleet orchestration.

XFRA Nodes are provisioned to operate as always-on loads within verified residential capacity. In normal operation, Node power is maintained continuously. If rare residential peaks occur, XFRA preserves Node operation by first drawing on the whole home battery, and in extreme cases by temporarily reducing non-critical flexible loads like EV charging via PowerUp. Node interruption will only occur during defined power events (such as a grid outage, a utility demand response event, or a safety-triggered shutdown), in which case XSOL orchestrates controlled failover to other Nodes in the fleet.

### \* 4.3 Rapid deployment at scale

Because XFRA leverages existing behind-the-meter capacity and the SPAN installed base, capacity can be deployed incrementally, Node by Node and in parallel, without waiting for large centralized site development cycles or new high-capacity interconnects for each deployment. XFRA's modular design supports multiple configurations while maintaining a consistent operating model: standardized Nodes can be added across the fleet, and deployment size can vary by configuration while remaining coordinated through XSOL as a single distributed cloud.

This deployment model enables rapid expansion of geographically distributed GPU capacity while maintaining operational consistency: the same orchestration, security controls, and grid-aware operating policies apply across sites, allowing XFRA to scale quickly without sacrificing reliability or safety.

05

# Addressing the big questions

## **5.1 Up-time reliability and continuity**

XFRA Nodes are designed to operate like infrastructure, not best-effort capacity. Each deployment includes whole-home energy storage that provides a resilient foundation for always-on compute.

When a grid power event occurs (such as an outage, a utility demand response event, or a safety-triggered shutdown), XSOL coordinates controlled failover of workloads to other Nodes in the fleet. XFRA is designed to offer maximum backup for the home during the event by failing over workloads and powering down the Node as quickly as possible, and making the remainder of the battery capacity available for the homeowner.

## **5.2 Sustainability and grid impact**

A core constraint on new compute capacity isn't silicon, it's power delivery. XFRA expands compute capacity by using existing grid infrastructure more effectively, rather than waiting for new, large-scale data center interconnects and upgrades. Residential electrical infrastructure usage is far below peak capacity on average, and XFRA is designed to convert a portion of that underused capacity into productive compute, raising utilization of existing grid assets.

Just as importantly, XFRA can act as a grid “shock absorber.” In periods of grid stress, compute load can be reduced via XSOL so that power is available for other needs, and distributed energy storage can provide local buffering during outages. Where it improves outcomes, deployments can also integrate rooftop solar, reducing net grid draw and increasing local sustainability and resilience.

Finally, the XFRA distributed architecture vastly reduces the number of single points of failure associated with centralized facilities, reducing the end user impact of localized failures.

## **5.2 Sustainability and grid impact (cont)**

For utilities, this model can improve system planning and help maximize the ROI of existing infrastructure by offering a more consistent power demand profile and reducing the need to build significant excess capacity to meet peak-only demands.

## **5.3 Homeowner value and fairness**

XFRA offers a great deal to homeowners. XFRA installs the energy and compute system, including SPAN panel, whole-home battery backup system, along with the XFRA compute Node, at no cost to homeowners. XFRA pays the homeowner a monthly rental to subsidize their energy and high-speed broadband bills such that they are a fraction of what they would normally be. This offers homeowners a sizable discount, and predictability in their monthly billing. When there is a power outage, the Node workloads are quickly transferred to other Nodes in the fleet. The remaining battery is made available to power the home, as a benefit to the homeowner.

## **5.4 Security and trust**

XFRA is designed for enterprise-grade reliability and security in a residential setting. Security spans customer workloads, fleet operations, and physical hardware. XSOL enforces encrypted communications, strong identity and access controls, and workload isolation suitable for multi-tenant environments. Operational monitoring and audit logging support continuous security posture management across the fleet.

Physical security is addressed through secure enclosures, monitoring, and tamper-evident protections. If a device is removed or compromised, XFRA can revoke credentials and disable access to workloads and sensitive data. Together, these measures are designed to protect customers, partners, and homeowners while enabling a distributed fleet to operate safely at scale.

06

# Target Workloads

XFRA is best suited to support workloads that can run on smaller hardware instances, benefit from rapid, incremental capacity expansion and have meaningful latency sensitivity.

This makes XFRA a strong complement to large data centers for serving demand closer to end users while expanding capacity faster than centralized build-outs alone.

XFRA can support many of today's highest-demand workloads, including:

AI inference,  
including Retrieval  
Augmented Generation  
(RAG)

35%

CAGR

Cloud gaming

34%

CAGR

Cloud rendering  
and streaming

8%

CAGR

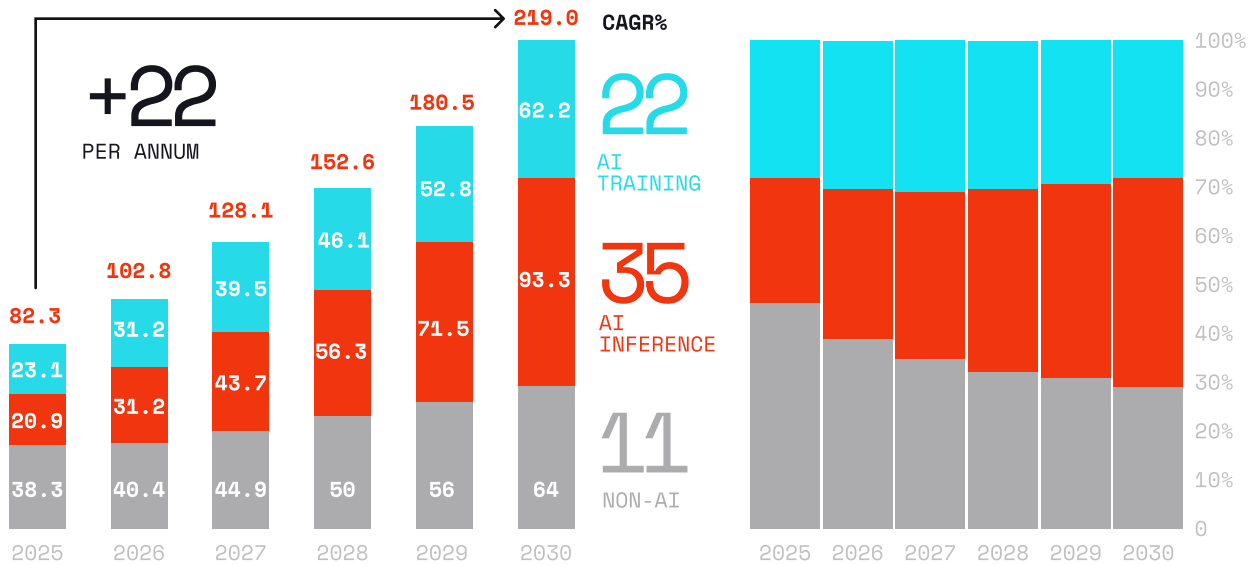
High-performance  
compute and  
simulation

7%

CAGR

For interactive workloads, proximity directly improves user experience and reduces backbone congestion.

Gigawatts



NOTE: Includes all provider types

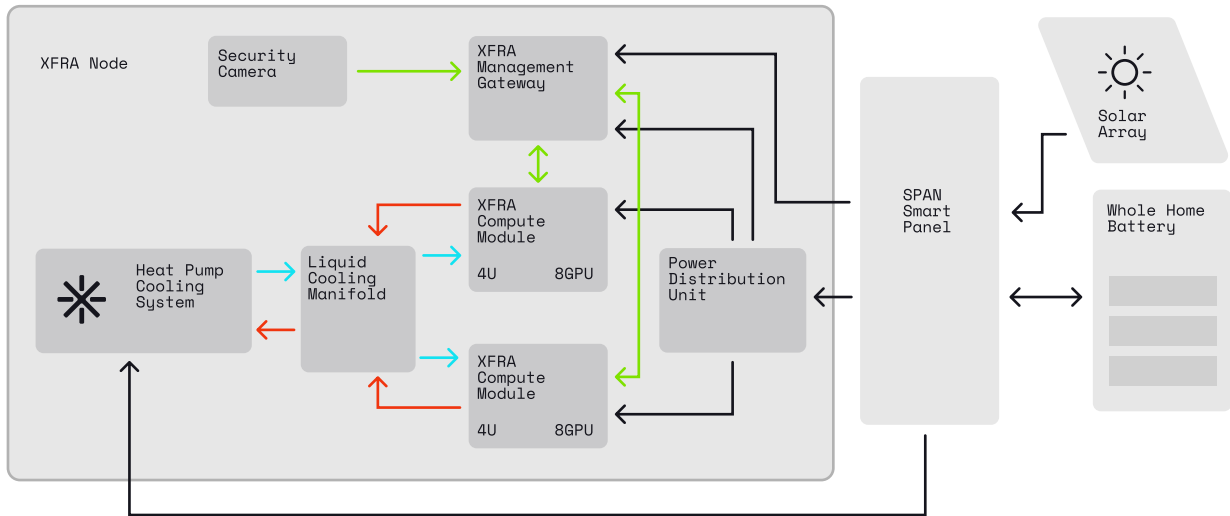
SOURCE: McKinsey Data-Center Demand Model

As AI service architectures become more efficient and more modular, driven by advances in model routing, optimization, and deployment techniques, an increasing share of inference demand can be served by distributed capacity placed closer to users and data sources. XFRA adds geographically distributed, inference-oriented capacity without requiring new large-scale interconnects.

XFRA is not intended to replace hyperscale training infrastructure. While it may support a limited subset of training and fine-tuning workloads, the largest training workloads require single-site cluster scale, specialized networking, and tightly coupled infrastructure best delivered in centralized data centers.

07

# XFRA Concept & Architecture



## 7.1 XFRA Node Architecture

An XFRA Node is built as a repeatable deployment unit, composed of one or more XFRA GPU Compute Modules, a local management gateway and thermal management designed for reliable, homeowner-friendly operation. The Compute Module is purpose-built and equipped with 8 cutting-edge, enterprise-grade GPUs, and ample CPU, memory, high-speed storage and networking bandwidth for optimized operation for the target workloads. The Compute Modules are liquid-cooled for quiet residential operation and to increase hardware and system reliability. Nodes are installed alongside a SPAN smart panel and a whole-home battery Energy Storage System (ESS) and can integrate rooftop solar where advantageous (and homeowner-approved).

This configuration is designed to support always-on compute within verified residential capacity while providing resilience during outages and short-duration disturbances. The ESS provides local buffering and backup power, improving continuity for both the home and the Node during power events.

## **7.2 Fleet Orchestration and Power Events**

At the fleet level, XFRA Nodes connect through a low-latency network fabric and are coordinated by XSOL (the XFRA Secure Orchestration Layer). XSOL securely schedules and routes workloads across the fleet based on customer needs (including latency), Node availability and health, and real-time grid constraints at each site.

During power events such as a grid outage, a utility demand response event, or a safety-triggered shutdown, SPAN surfaces site constraints via the SPAN Cloud, and XSOL coordinates the operational response: workloads are failed-over to other Nodes in the fleet and affected Nodes power down safely. There is also a local safety interlock that can trigger power down from the SPAN smart panel, in the event that the network or SPAN cloud is down and there is a local safety issue. This architecture is designed to deliver predictable, managed behavior across a distributed fleet while keeping homeowner experience and system safety first.

## **7.3 Why This Architecture Matters**

Together, these elements create an infrastructure layer that improves how compute and power are delivered at the edge. Nodes scale incrementally across geographies, reduce reliance on single large sites, and increase resilience by distributing capacity across many independent deployments, while using existing grid infrastructure more effectively.

08

# Core Value Propositions

**XFRA delivers value across the compute and energy ecosystem by turning underutilized behind-the-meter capacity into a managed, geographically distributed compute layer while improving resilience through local storage and secure fleet orchestration.**

### **Compute offtake customers**

XFRA provides rapidly deployable GPU capacity to meet surging demand, delivered as a resilient, secure, and scalable fleet coordinated through XSOL. Because capacity is distributed, workloads can be placed closer to end users, improving performance for latency-sensitive services such as interactive AI inference and cloud gaming.

XFRA also gives compute platforms a path to add incremental capacity without waiting on long-lead site development and utility interconnection timelines, providing flexibility as demand grows.

### **Homeowners (Node hosts)**

Homeowners receive a SPAN smart panel and whole-home energy storage installed at zero cost, improving household resilience during outages and adding a modern energy management foundation. XFRA is designed so household energy use is not compromised, and homeowners benefit from a predictable and improved monthly value proposition that will include bundled services and host compensation.

### **Power utilities and grid operators**

XFRA helps utilities make better use of existing infrastructure by increasing utilization of behind-the-meter capacity and adding distributed storage at the edge. Compute load is curtailed during grid stress events, and storage provides local buffering to smooth short-duration spikes and improve system stability. This can ease peak-driven planning constraints and help improve ROI on grid capital investments by smoothing demand and reducing the need for overbuild in some scenarios.

### **Governments and society**

XFRA supports the growing demand for high-performance compute driven by AI adoption while reducing reliance on slow, capital-intensive centralized build-outs alone. A distributed architecture can improve resilience by reducing dependence on a small number of large sites and by encouraging more geographically balanced infrastructure.



## **SPAN**

XFRA creates a new recurring revenue stream that monetizes the SPAN platform. It strengthens SPAN differentiation in home energy management, expands the value of the installed base, and deepens the SPAN role as the coordination layer between home energy and next-generation compute.

09

# Next Steps

**XFRA is already operating in-house prototypes that are serving paying customers, validating early technical feasibility and initial product-market fit.**

We are now expanding testing to a small number of homes, and will roll out a 100-Node residential proof of concept deployment in early Q3 in partnership with a leading build-to-rent company.

The near-term emphasis is not on scaling prematurely, but on maximizing speed of learning through rapid, system-level iteration, spanning hardware, installation workflows, orchestration software, reliability and customer experience.

Following the proof of concept, we plan to prioritize new-build and build-to-rent deployments as the path to broader U.S. rollout beginning in early 2027, while also developing additional installation options to expand into retrofits (for new and existing SPAN customers) and larger configurations for small-business and commercial customers. Beginning in 2027, XFRA will scale to an annual capacity in excess of 1 gigawatt of AI inference compute, accessing a robust pipeline of new construction starts and commercial Nodes across the country.

10

# Risks

XFRA is a new infrastructure model, and we are approaching it with disciplined risk reduction through staged pilots and measurable milestones.

Key risk areas include:

- **Product-market fit:** Will compute customers commit to purchase capacity at pricing and terms that support a durable business?
- **Technical performance and reliability:** Can we consistently deliver the performance, security, uptime reliability and operational safety expected from a distributed fleet?
- **Deployment and operations:** Can we install, service, and monitor Nodes at scale with a homeowner-friendly experience and sustainable unit economics?
- **Regulatory and utility variability:** How do local rules, programs, and utility processes affect deployment, operating modes, and economics across regions?
- **Supply chain and manufacturing:** Can we secure components and build capacity predictably as demand scales?

We will mitigate these risks methodically through the proof-of-concept phases, iterating with real customers and partners while expanding scope only as reliability, economics, and operational readiness are demonstrated.

11

# Conclusion & Call to Action

## XFRA is a pragmatic response to the growing imbalance between compute demand and power delivery capacity.

By deploying distributed compute Nodes alongside SPAN smart panels, paired with whole-home energy storage and optional solar where advantageous, XFRA converts underutilized behind-the-meter headroom into deployable, managed compute capacity. This approach complements centralized data centers while enabling faster, more geographically distributed expansion that aligns with grid realities.

We are entering an era where compute access will shape competitive advantage across industries and countries. XFRA offers a path to deploy capacity faster, closer to users and with more resilient infrastructure characteristics than a purely centralized model. We believe this is an opportunity for SPAN, and for a broader ecosystem of investors, partners, utilities, and compute platforms, to help build a new layer of distributed infrastructure.

## Want to help build XFRA?

- **Investors:** fund proof-of-concept execution, manufacturing readiness, and scale-out deployment
- **Compute platforms (hyperscalers, neoclouds, AI service providers):** pilot and production offtake, workload requirements and integration partnerships
- **Utilities and energy partners:** collaborate on grid-event operations, demand response participation and regional deployment models
- **Homebuilders, homeowners, and communities:** partner on build-to-rent and new-build deployments and repeatable installation programs to help build a more resilient, sustainable grid and a compute infrastructure layer where hosting communities share in the upside
- **Regulators and policymakers:** support frameworks that enable safe, scalable distributed compute as part of modern grid planning

The demand is real.

The grid constraints are real.

XFRA is designed to bridge the gap.



[xfra.ai](https://xfra.ai)

Last Updated April 2026

©2026 SPAN.IO, INC. ALL RIGHTS RESERVED.