

AI Factory Architecture Evolution Handbook

From GB300 NVL72 to Future AI Factories

A Comprehensive Guide to Building Scalable,
Efficient, and Sustainable AI Infrastructure



ADVANCED
COMPUTE



OPTIMIZED
POWER



EFFICIENT
COOLING



MODULAR
DEPLOYMENT



CAMPUS
SCALE



SUSTAINABLE
ENERGY



RELIABLE & RESILIENT
Built for 24/7 mission
critical operations



SCALABLE & FLEXIBLE
Designed to scale from
rack to gigawatt



COST OPTIMIZED
Lower TCO through
right-sized design



SUSTAINABLE
Energy efficient and
environmentally responsible



GB300 NVL72

Current AI Platform

150kW Rack Power



Vera Rubin NVL72

Next Generation AI Platform

~200–250kW Rack Class



20MW AI Hall

Single AI Hall

Complete Infrastructure Solution



50MW AI Block

2 × 20MW AI Halls

Redundancy and Efficiency



300MW AI Campus

6 × 50MW AI Blocks

Campus Scale Infrastructure



GW AI Factory

Multi-Campus Scale

Full AI Factory Ecosystem



From AI Platforms to AI Factories



BUILD ONCE.



SCALE CONTINUOUSLY.



SUPPORT MULTIPLE GENERATIONS.



INTEGRATE ENERGY AND COMPUTE.



DELIVER LASTING VALUE.



Internal Engineering Reference | Version 2026

1 / 14

Topics: GB300 NVL72 • Vera Rubin • SST • 800V DC Distribution • 20MW Data Hall • AI Factory

AI Factory Architecture Framework

A Framework for Understanding Next-Generation AI Infrastructure



1 COMPUTE EVOLUTION

AI rack platforms continue to evolve with higher density and greater performance

GB300 NVL72
Current Generation



Vera Rubin NVL72
Next Generation



Rubin Ultra
Ultra-Dense Generation



Kyber Platform
Future Generation



Higher Density



Greater Performance



Advanced Cooling



Power Delivery Evolution



2 POWER EVOLUTION

Power architectures evolve to support AI density growth with higher efficiency and scalability

Traditional AC
AC Architecture



Hybrid AC / DC
Transitional Architecture



800V DC
DC Architecture



SST-Enabled
Next-Gen Power Center



Efficiency



Lower Losses



Higher Density



Scalability



3 DEPLOYMENT SCALE

Scalable from rack level to GW-scale AI factories

AI Rack Platform



Rack Level

20MW AI Data Hall



Hall Scale

50MW Power Block



Block Scale

300MW AI Campus



Campus Scale

GW AI Factory



Factory Scale



4 ECOSYSTEM PARTICIPANTS

Built for the entire AI infrastructure ecosystem

INFRASTRUCTURE



Utility

Power delivery and grid modernization



EPC Contractor

Design and build power infrastructure



OEM / Vendor

Provide equipment and technology solutions

OPERATORS & CAPITAL



Hyperscaler

Build and operate massive AI infrastructure



Cloud Provider

Deliver AI compute as a service



Data Center Operator

Operate and manage AI data centers



Infrastructure Investor

Invest in AI infrastructure and platforms

THREE DIMENSIONS
OF AI INFRASTRUCTURE



Compute
Evolution

+



Power
Evolution

+



Deployment
Scale

=

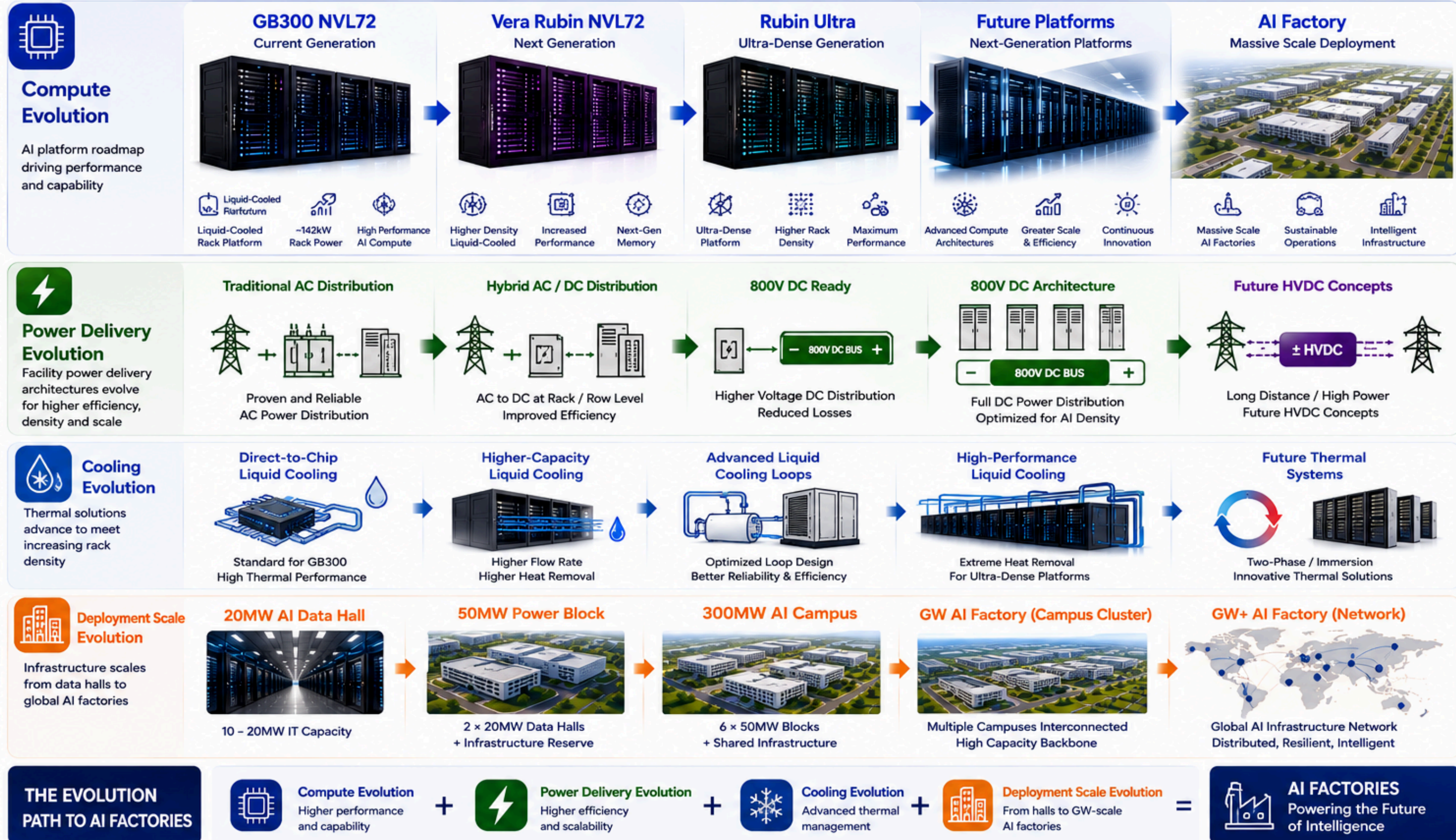


AI Factory
Next-Generation
Infrastructure)



AI Infrastructure Evolution Overview

Compute, Power, Cooling and Deployment Evolution Toward AI Factories



GB300 NVL72 Reference Architecture

Current Generation Platform for AI Factory Deployment



~150kW
Rack Power
Up to ~150kW
Typical Operating Range

Integrated 54V DC
Internal Rack Bus
High Efficiency
DC Power Delivery
to Compute Trays

Liquid Cooling
Direct-to-Chip
Industry Standard
Thermal Solution

Production Ready
Available Today
Proven in Large-Scale
AI Deployments

Current Generation
Reference Platform
AI Factory
Entry Point

GB300 NVL72 REFERENCE DEPLOYMENT ARCHITECTURE

UTILITY / GRID
High Voltage
Incoming Power

FACILITY DISTRIBUTION
Facility Level
Power Distribution

RACK POWER INPUT
AC Power to Rack
Power Shelves

POWER CONVERSION
Inside the Rack
Power Shelves

INTERNAL DC BUS
54V DC Distribution
to Compute Trays

COMPUTE TRAY
GB300 NVL72
Compute Platform

34.5kV MV
Medium Voltage (MV)
• Reliable Power Feed
• Utility or On-site Generation

480V AC
Low Voltage (LV) AC Distribution
• Switchgear / Busway
• Redundant Power Paths
• N+1 or 2N Reliability

480V AC Input to Rack
Three-Phase AC
• Feeds Integrated Power Shelves
• Short, Efficient Power Path

AC → 54V DC
Power Shelf Conversion
• AC to 54V DC Conversion
• High Efficiency (96%+)
• Hot-Swappable Shelf Design

54V DC Bus
Rack Internal DC Distribution
• Low Loss Power Delivery
• Optimized for High Density

Blackwell Ultra GPUs
72x Blackwell Ultra GPUs
• NVLink Switch System
• High-Speed Data Fabric
• Direct-to-Chip Liquid Cooling

GB300 NVL72 RACK OVERVIEW

Top-of-Rack (ToR)
Networking

Compute Trays
(Blackwell Ultra GPUs)
Liquid-Cooled

Power Shelves
(AC to 54V DC)
Hot-Swappable

54V DC Bus
Integrated Busbars
High Current

CDU / Manifold
Coolant Distribution
Leak Detection

Rack Footprint
600mm (W) x 1,200mm (D) x 2,100mm (H)
(Typical)

Rack Power
Up to ~150kW
(Typical Operating Range)

WHY GB300 NVL72 MATTERS

Production Ready Today
GB300 NVL72 is deployed today at scale with a mature supply chain and proven large-scale operations.

Proven Rack Architecture
Standardized rack design with integrated power, networking, and liquid cooling for rapid deployment.

Mature Liquid Cooling Ecosystem
Direct-to-chip liquid cooling with established components, vendors, and operational best practices.

Foundation for AI Factory Deployment
Designed for scale-out clusters and AI factories, supporting tens to hundreds of MW installations.

Compatible with Existing AC Infrastructure
Leverages traditional 480V AC distribution for easier site readiness and deployment flexibility.

TYPICAL RACK SPECIFICATIONS (REFERENCE)

Rack Power (Up to)	~150kW
Power Input	480V AC, 3Ø, 60Hz
DC Bus Voltage (Internal)	54V DC
Power Conversion	AC to 54V DC (Power Shelves)
Cooling	Direct-to-Chip Liquid Cooling
Compute	GB300 NVL72 (72x Blackwell Ultra GPUs)
Rack Height	Standard 42U
Availability	Production Ready

KEY DESIGN CHARACTERISTICS

Compute
72x Blackwell Ultra GPUs
NVLink Full Mesh
High Bandwidth Memory

Power
~150kW Rack Power
480V AC Input
54V DC Internal Bus
High Efficiency

Cooling
Direct-to-Chip Liquid Cooling
Cold Plate Technology
High Thermal Performance

Networking
High-Speed InfiniBand
NVLink Switch System
Low Latency Fabric

Reliability
Redundant Power Paths
Hot-Swappable Modules
High Availability Design

Operations
Production-Ready Platform
Field-Tested at Scale
Ecosystem Support

99 GB300 NVL72 represents the current generation reference architecture for AI Factory deployment, combining proven liquid cooling, rack-scale design, and production-ready infrastructure.”

AI Factory Architecture Handbook | Internal Engineering Reference | Version 2026

4 / 14

Topics: GB300 NVL72 • Vera Rubin • SST • 800V DC • 20MW Data Hall • AI Factory

Vera Rubin NVL72 Platform Evolution

Next-Generation AI Rack Platform – Projected Direction



**~200–250kW
Rack Power Class**
Higher density for
next-gen AI workloads



**Advanced
Liquid Cooling**
Direct-to-chip and
CDU solutions



**Future-Ready
Power Architecture**
Designed to support
evolving power topologies



**Next-Generation
AI Platform**
Higher compute density
and advanced networking



**Expected
2026+
Deployment**
Projected industry
timeline

PROJECTED POWER DELIVERY PATH



UTILITY / GRID
34.5kV MV

Medium Voltage
Power Supply



FACILITY DISTRIBUTION
480V AC

Facility
Power Distribution



**PROJECTED RACK
POWER INPUT**
AC or Hybrid AC/DC

Flexible Rack
Power Input



**PROJECTED IN-RACK
POWER CONVERSION**
Power Electronics
Architecture

High Efficiency
Power Conversion
and Distribution



**PROJECTED
COMPUTE TRAYS**
Next-Generation
Compute Platform

Compute Trays with
High-Density GPUs
and Accelerators

VERA RUBIN NVL72 RACK (REFERENCE OUTLOOK)



* Illustration for reference only



**Projected Power
Architecture**
Modular power shelves
and intelligent power
management



**Advanced Power
Electronics**
High efficiency conversion
and distribution for
future power topologies



**Integrated Cooling
System**
Direct-to-chip liquid
cooling with CDU
distribution manifold

VERA RUBIN NVL72 – PROJECTED DIRECTION

Category	Projected Direction
Rack Power Class	200–250kW+
Cooling	Advanced Liquid Cooling (Direct-to-Chip + CDU)
Network Fabric	Next-Generation NVLink (Higher Bandwidth)
Compute Density	Higher Than GB300 (More GPUs per Rack)
Power Architecture	Under Industry Evolution (AC / Hybrid AC/DC / DC Architectures)
Platform Status	Expected 2026+ (Projected Industry Timeline)



Vera Rubin NVL72 represents the next-generation AI rack platform, delivering higher compute density, advanced networking and evolving power architectures for future AI factories.



Higher Performance
More GPUs, more flops



Power Architecture Evolution
Designed for flexibility
and future readiness



Sustainable and Scalable
Efficient cooling and
infrastructure optimization



AI Infrastructure Power Decision Framework

Selecting the Right Power Architecture for Your AI Factory



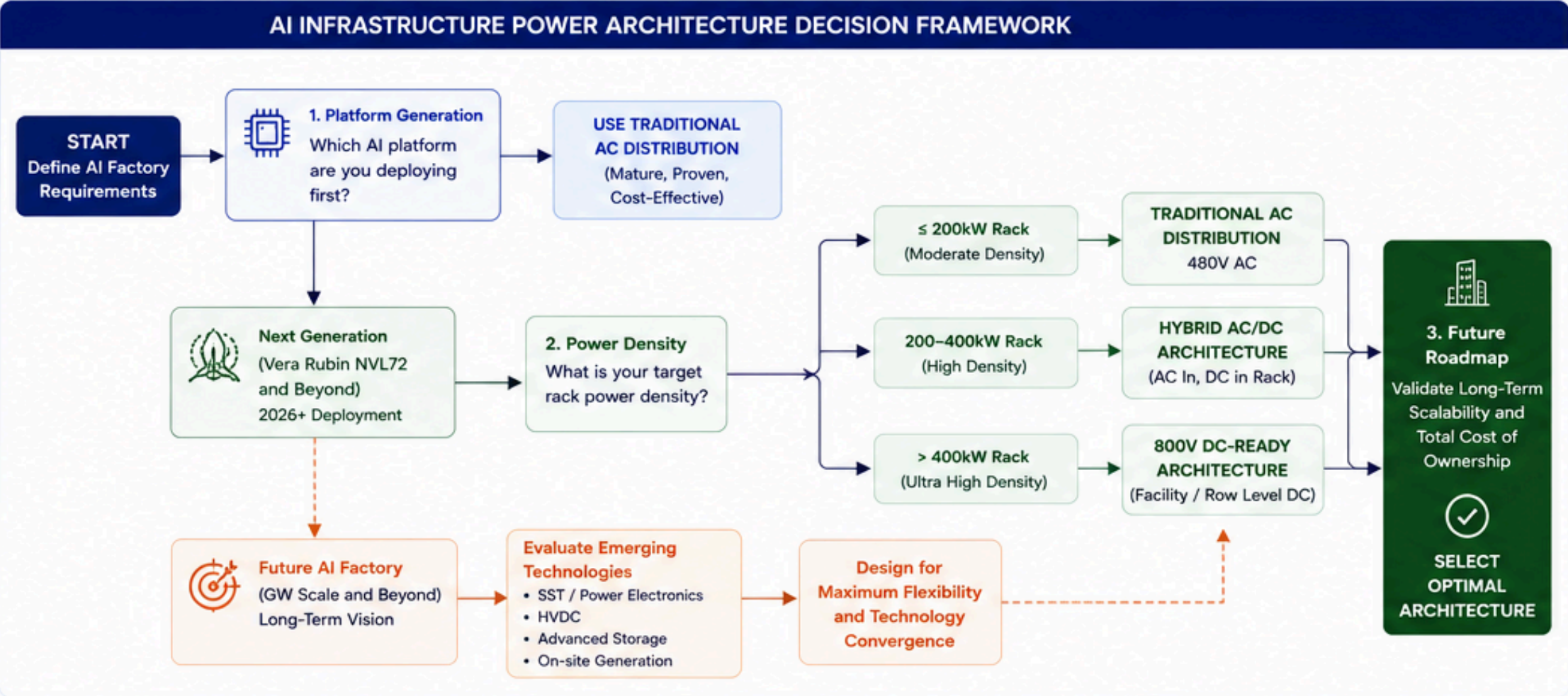
Right Architecture
Match power architecture with platform, density and business objectives

Scalable Investment
Enable gradual evolution while protecting current investments

Operational Reliability
Ensure high availability with proven and future-ready technologies

Energy Efficiency
Optimize total cost of ownership and carbon footprint

Future Flexibility
Design for multiple generations and emerging power technologies



KEY CONSIDERATIONS

Platform Roadmap

- Current: GB300 NVL72 ~150kW
- Next: Rubin NVL72 ~200–250kW
- Future: Higher density platforms

Power Density & Growth

- Current needs vs. future expansion
- Rack power trajectory over 5–10 years

Total Cost of Ownership

- CAPEX: Equipment, Infrastructure
- OPEX: Energy, Cooling, Maintenance

Reliability & Availability

- Power path redundancy
- Technology maturity and track record

Sustainability Goals

- Energy efficiency
- Carbon footprint reduction

Site & Utility Readiness

- Grid capacity and stability
- Local regulations and incentives

DECISION GUIDANCE

- Start with your first platform and business timeline.
- Design for future density and technology evolution.
- Balance performance, cost, risk and flexibility.
- Engage with experienced partners and vendors early.
- Review and evolve your strategy as technology matures.

THERE IS NO ONE-SIZE-FITS-ALL SOLUTION

The optimal power architecture depends on your platform generation, power density targets, deployment timeline, site conditions, and total cost objectives.

MAKE THE RIGHT CHOICE TODAY, BUILD A FUTURE-READY AI FACTORY.

POWER ARCHITECTURE OPTIONS AT A GLANCE				
	OPTION 1 Traditional AC Distribution	OPTION 2 Hybrid AC/DC Architecture	OPTION 3 800V DC-Ready Architecture	OPTION 4 Future Technologies (SST/HVDC)
Rack Input	480V AC	480V AC (AC in), DC in Rack	800V DC (Facility/Row Level)	Varies (High Voltage DC or AC)
Best For	Current Gen (GB300), ≤200kW racks	Next Gen (Rubin), 200–400kW racks	High Density >400kW racks, Future Platforms	GW Scale AI Factory, Long-Term Future
Advantages	- Mature & proven - Lower initial cost - Wide ecosystem	- Higher efficiency - Reduced I²R losses - Good balance of cost & performance	- Highest efficiency potential - Less copper, more power density - Future-ready for scaling	- Ultra-high efficiency - Grid integration benefits - Advanced control & flexibility
Considerations	- Lower efficiency at high density - More copper, more space	- In-rack DC complexity - Requires power shelf technology	- Ecosystem still evolving - Higher initial investment	- Emerging technology - Longer standardization timeline

99 The right power architecture decision today enables reliable operations, optimized TCO, and seamless evolution across multiple generations of AI platforms."

Future Power Conversion Architectures

Evaluating Traditional and Solid-State Approaches for Next-Generation AI Infrastructure



Traditional AC Compatible

Fully compatible with today's proven AC distribution and equipment



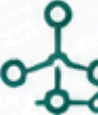
Potentially Higher Efficiency

Power electronics enable higher efficiency and lower losses



Reduced Physical Footprint

Smaller and lighter equipment footprint and simplified layouts

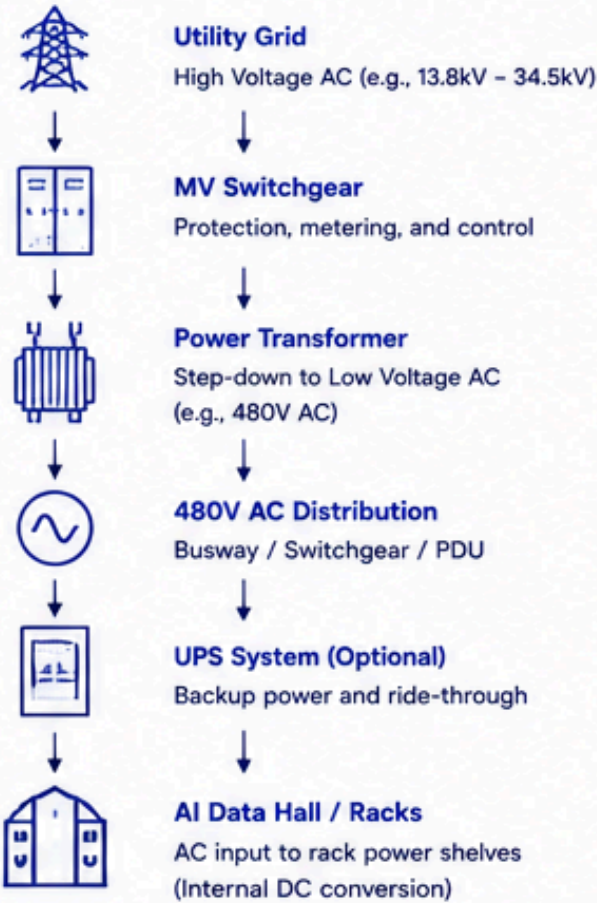


Future Distribution Flexibility

Enables AC/DC hybrid and DC-native distribution for future platforms

TRADITIONAL POWER CENTER ARCHITECTURE

Proven. Mature. Widely Deployed Today.



KEY CHARACTERISTICS

- ✓ **Proven Technology**
Decades of operational history
- ✓ **High Availability**
Mature supply chain and service ecosystem
- ✓ **Lower Initial Cost**
Lower equipment cost today
- ✓ **Widely Supported**
Broad engineer and contractor expertise
- ✓ **Lower Implementation Risk**
Well-understood design, commissioning and operation



Best For: Current generation deployments, shorter timelines, and projects prioritizing lowest risk and proven solutions.

VS.

ARCHITECTURE COMPARISON

Category	Traditional	SST
Technology Maturity	High	Emerging
Market Availability	High	Limited
Conversion Efficiency	Good	Potentially Higher
Footprint	Larger	Smaller
Scalability	Good	Excellent
Power Quality Control	Good	Advanced
Reliability Track Record	Proven	Developing
Initial Cost	Lower Today	Higher Today
Implementation Risk	Lower	Higher
Best Application	Current Deployments	Future AI Factories

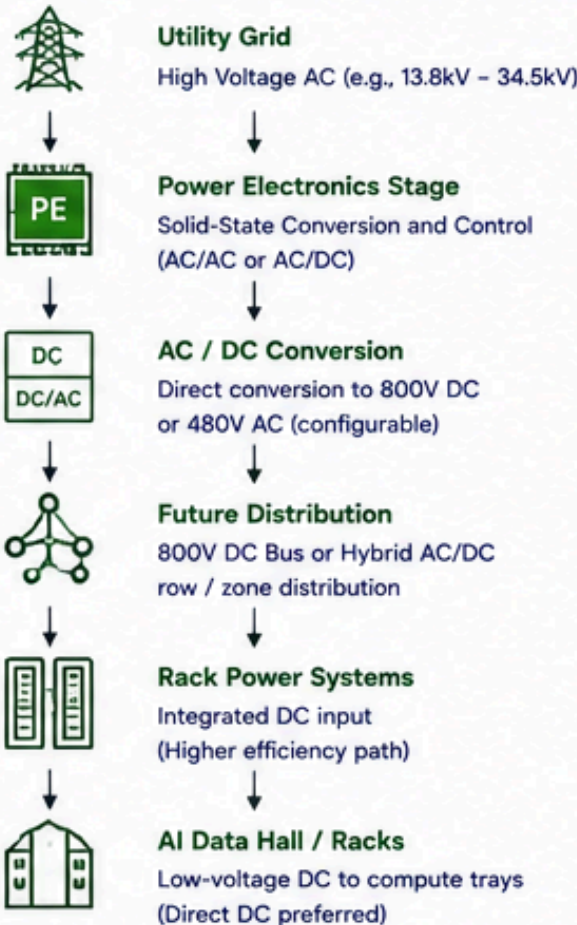


Key Takeaway

SST is an emerging option for future AI factories, not a mandatory requirement for current deployments.

SOLID-STATE POWER CENTER ARCHITECTURE (SST)

Emerging. Power Electronics Based. Future-Oriented.



KEY CHARACTERISTICS

- ✓ **Higher Efficiency Potential**
Fewer conversion stages and lower losses
- ✓ **Smaller Footprint**
Reduced size and weight of power equipment
- ✓ **Power Quality Control**
Advanced control and fast dynamic response
- ✓ **DC Integration Ready**
Enables 800V DC and future DC-native architectures
- ✓ **Future Scalability**
Designed for higher density and future platform evolution



Best For: Future AI factories, ultra-high density deployments, and campuses targeting DC distribution and maximum efficiency.

WHEN TO CONSIDER TRADITIONAL ARCHITECTURE



- Deploying GB300 or current-generation platforms
- Shorter project timelines (12–24 months)
- Leveraging existing utility and infrastructure
- Prioritizing lowest CAPEX and operational risk
- Projects in regions with limited SST availability

WHEN TO CONSIDER SST ARCHITECTURE



- Planning for Rubin Ultra and future ultra-dense platforms
- Long-term campus developments (3+ years horizon)
- Availability of 800V DC infrastructure in the region
- Strong focus on efficiency, footprint and sustainability
- Willingness to adopt emerging technologies with partners

IMPORTANT CONSIDERATIONS



- SST technology and standards are still evolving
- Higher initial cost and longer lead times today
- Requires experienced partners and early engagement
- Utility interconnection and codes may require coordination
- Pilot projects and phased adoption recommended

OUR RECOMMENDATION



Design for flexibility. Keep options open. Adopt architectures that allow migration from traditional to SST/DC in the future without major rework.

99

Both traditional and solid-state power conversion architectures can enable successful AI Factory deployments. The optimal choice depends on project timeline, density targets, utility constraints and technology maturity.



800V DC Distribution Architecture

The Next Step in Power Distribution for AI Factories



INDUSTRY DIRECTION



Higher Efficiency
Lower current, lower losses, higher efficiency



Lower CapEx
Smaller conductors, less copper, reduced cost



Scalable Architecture
Designed for multi-MW and GW-scale AI factories

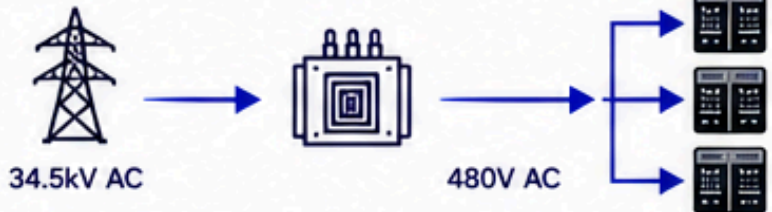


Future Evolution
Foundation for future DC-native infrastructure



Ready Today
Proven components, standards and designs

TODAY'S ARCHITECTURE (AC)



- Industry standard for decades
- Higher current at 480V AC
- Larger conductors, more copper
- Higher I²R losses
- More switchgear and transformation stages

DISTRIBUTION CURRENT COMPARISON

IT LOAD (Per Rack)	TRADITIONAL AC (480V AC)	800V DC ARCHITECTURE (800V DC)	IMPROVEMENT
150 kW	312 A	188 A	↓ 40% Lower Current
250 kW	521 A	313 A	↓ 40% Lower Current
1 MW	2,083 A	1,250 A	↓ 40% Lower Current



Lower Current



Less Copper

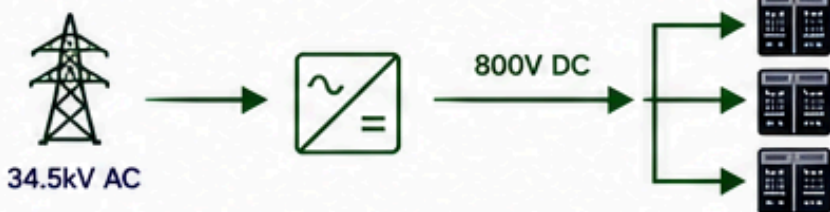


Lower Losses



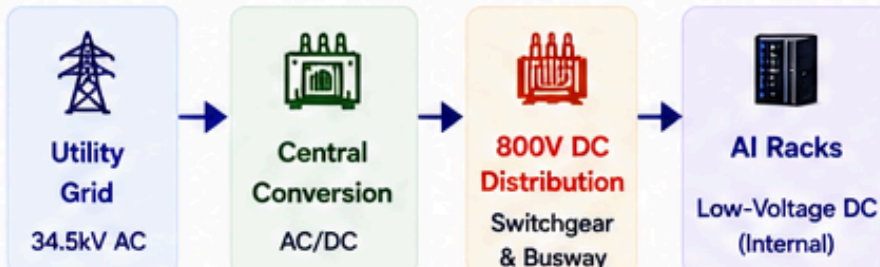
Higher Efficiency

FUTURE-READY ARCHITECTURE (800V DC)



- 800V DC distribution from central conversion
- ~40% lower current vs 480V AC at the same power level
- Smaller conductors, less copper
- Lower I²R losses, higher system efficiency
- Fewer conversion stages, simplified architecture

POWER ARCHITECTURE LAYERS



ADOPTION OUTLOOK



KEY BENEFITS OF 800V DC

Lower Current	~40% lower current vs 480V AC at the same power level
Higher Efficiency	2-4% system efficiency gain at facility level
Reduced Conductor Size	20-40% less copper and smaller cable trays
Lower CapEx	Reduced material, labor and equipment costs
Improved Reliability	Fewer components, less complexity, higher uptime



800V DC is designed to reduce facility-level distribution current. AI racks continue to use internal low-voltage DC distribution (e.g., 50-54V DC bus, depending on platform design).



Leverages proven components and standards to enable safe, reliable and scalable deployments.



BUILD ONCE.



SCALE CONTINUOUSLY.



SUPPORT MULTIPLE GENERATIONS.



INTEGRATE ENERGY AND COMPUTE.



DELIVER LASTING VALUE.



20MW AI Data Hall Reference Design

A Future-Proof Deployment Module for Multi-Generation AI Platforms





20MW
IT Load Capacity
Net IT Power



100–150
AI Racks
Depending on Platform



Direct-to-Chip
Liquid Cooling
High Efficiency



Future-Proof
Deployment Module
Modular & Scalable

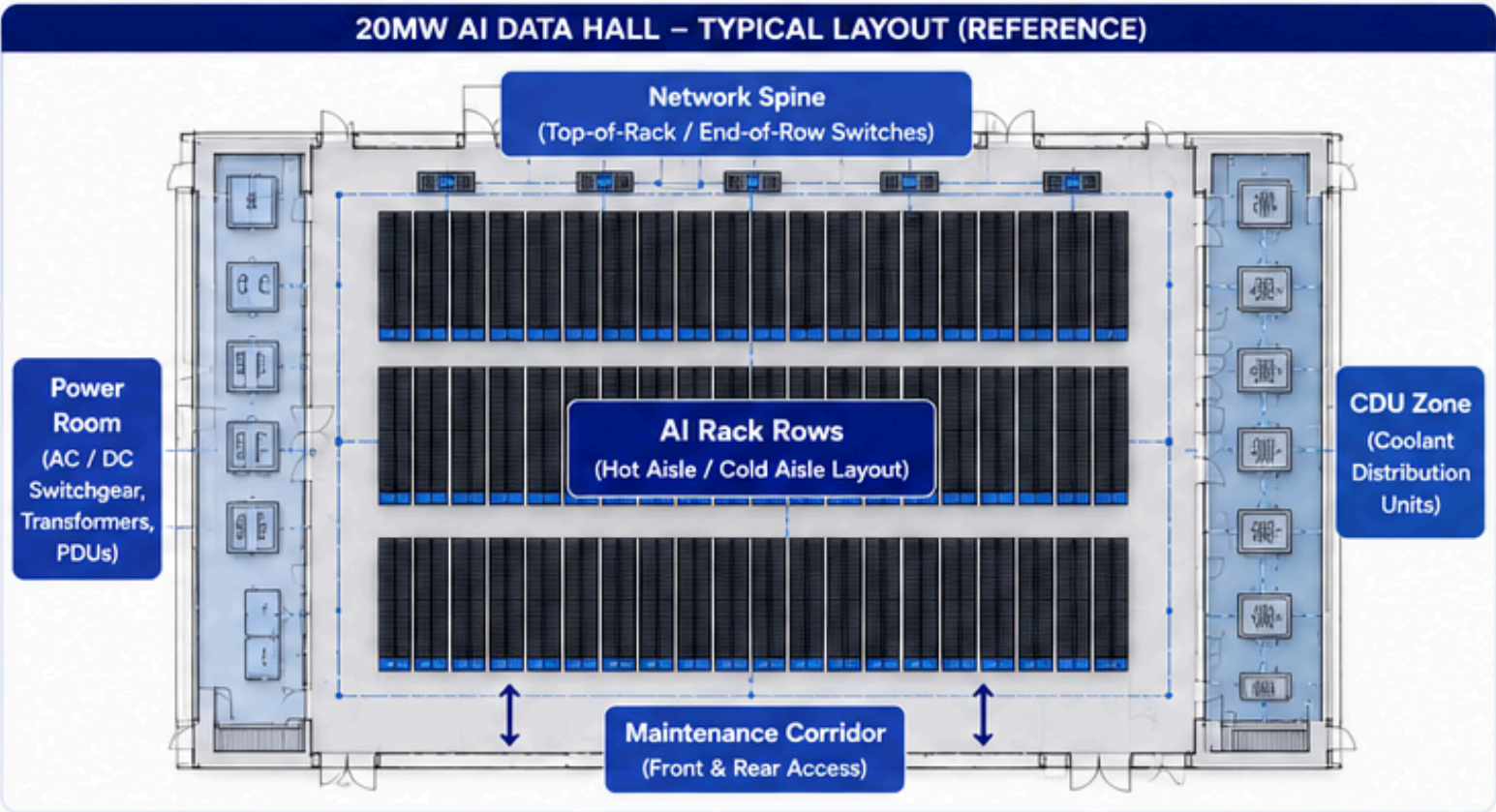
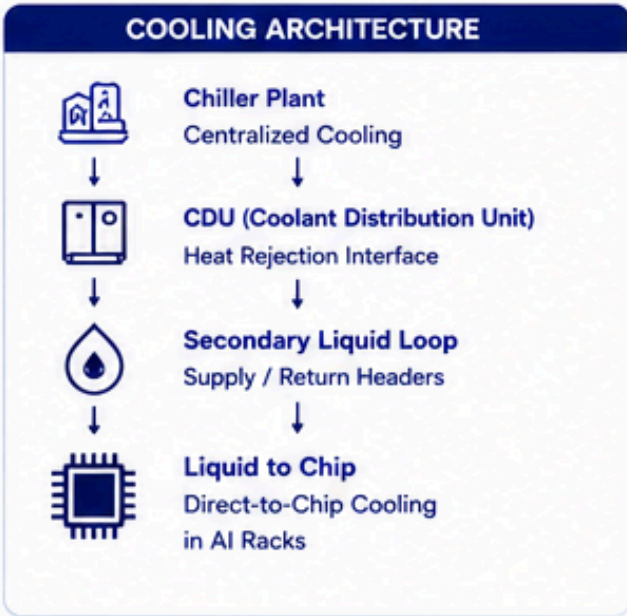
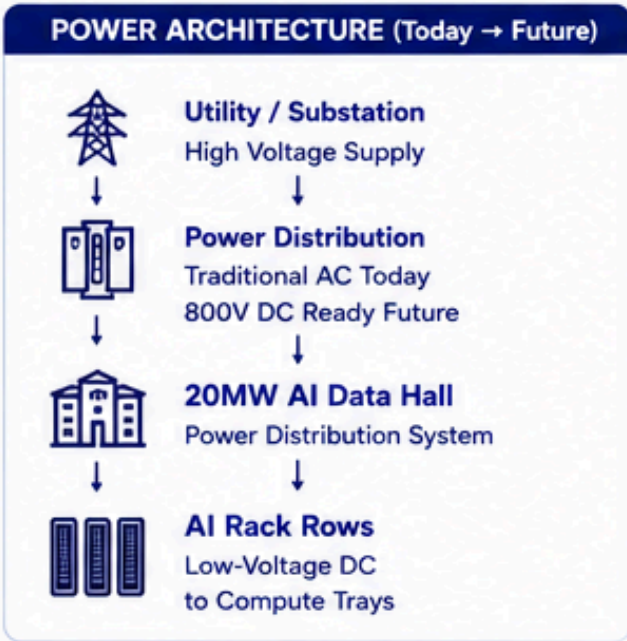
INFRASTRUCTURE EVOLUTION PATH

TODAY
Traditional AC
480V AC Distribution

TRANSITION
Hybrid AC/DC
Flexible Architecture

FUTURE
800V DC READY
800V DC Distribution

Same 20MW Hall Architecture • Smooth Transition • No Major Redesign Required



20MW HALL CAPACITY EXAMPLES (Approximate)

AI Platform	Rack Power Class	Approx. Racks / 20MW Hall	Notes
GB300 NVL72 (Today)	~150kW	~130 Racks	Traditional AC Distribution
Vera Rubin NVL72 (Future)	~200 – 250kW	~80 – 100 Racks	Hybrid AC/DC Ready
Rubin Ultra (Future+)	~400kW Class	~50 Racks	High-Density Cooling
Future Platforms	Higher Density	Fewer Racks	800V DC Ready Infrastructure

FUTURE-READY DESIGN PRINCIPLES

-  **Modular Power Distribution**
Pre-engineered power modules enable easy migration from AC to hybrid or 800V DC.
-  **Flexible Busway & Cable Routing**
Oversized pathways and conduits support higher voltage and current evolution.
-  **Liquid Cooling Backbone**
Standardized CDU placement and piping support future higher-density racks.
-  **Scalable Network Fabric**
Spine-leaf design with expansion capability for future bandwidth demands.
-  **Power Conversion Flexibility**
Rack-level power shelves support different input voltages and future 800V DC feeds.

WHAT STAYS THE SAME vs WHAT EVOLVES

Layer	Stays the Same	Evolves Over Time
 Building Shell / Structure	✓	–
 Hall Layout & Footprint	✓	–
 Cooling Backbone (Piping, CDU Zone)	✓	Upgrade Capacity
 Network Spine (Topology & Pathways)	✓	Upgrade Bandwidth
 Power Distribution Architecture	–	Traditional AC → Hybrid AC/DC → 800V DC Ready
 Rack Platform (Compute Density)	–	GB300 → Rubin → Rubin Ultra → Future





Design Once.
Support Multiple AI Generations.
Adopt Future Power Architectures When Ready.

A properly designed 20MW AI Hall should support multiple generations of AI platforms and evolving power architectures without major facility redesign.

50MW AI Power Block Architecture

The Scalable Building Block for Campus-Level AI Infrastructure





50MW
IT Capacity
Net IT Power



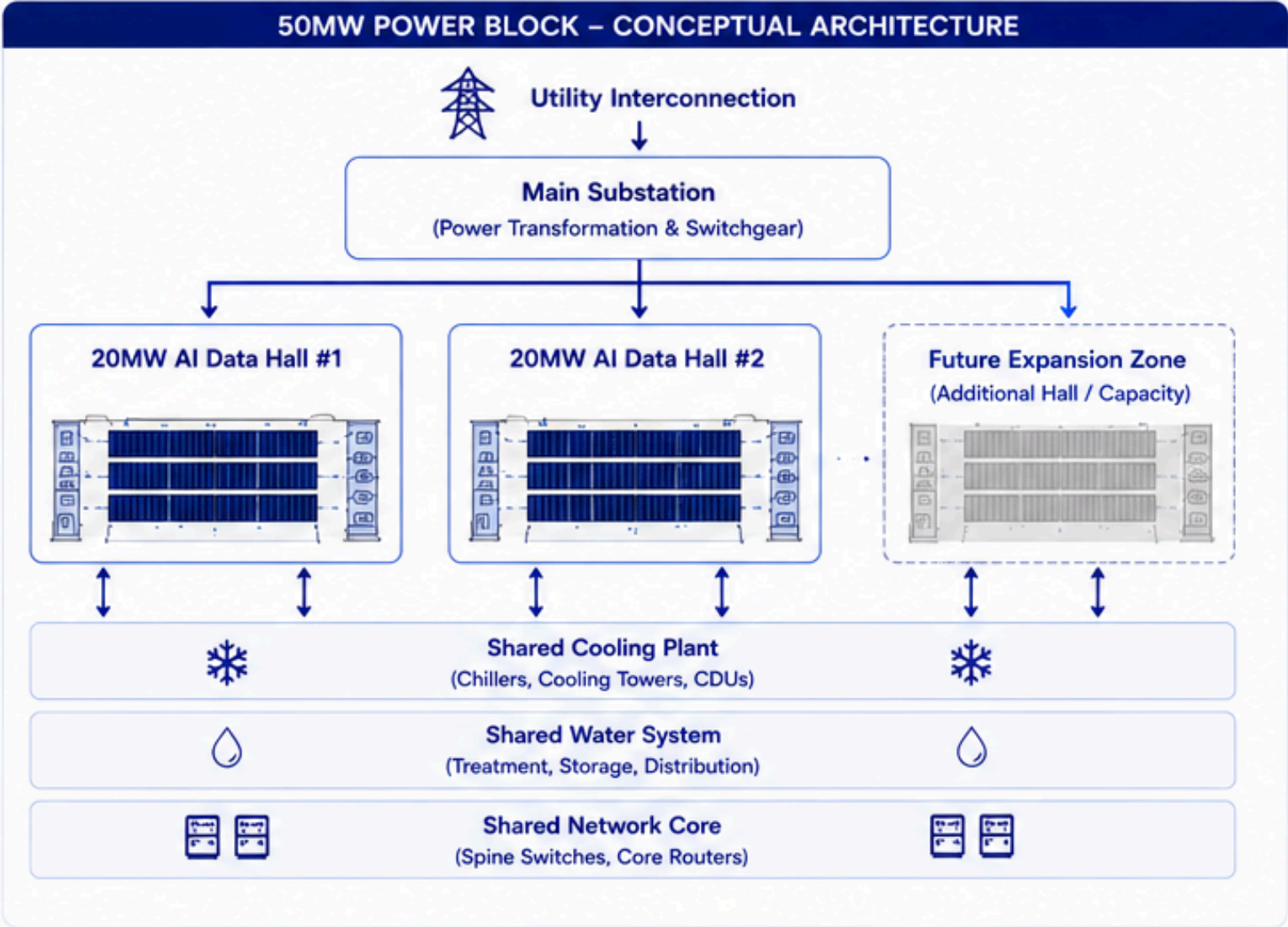
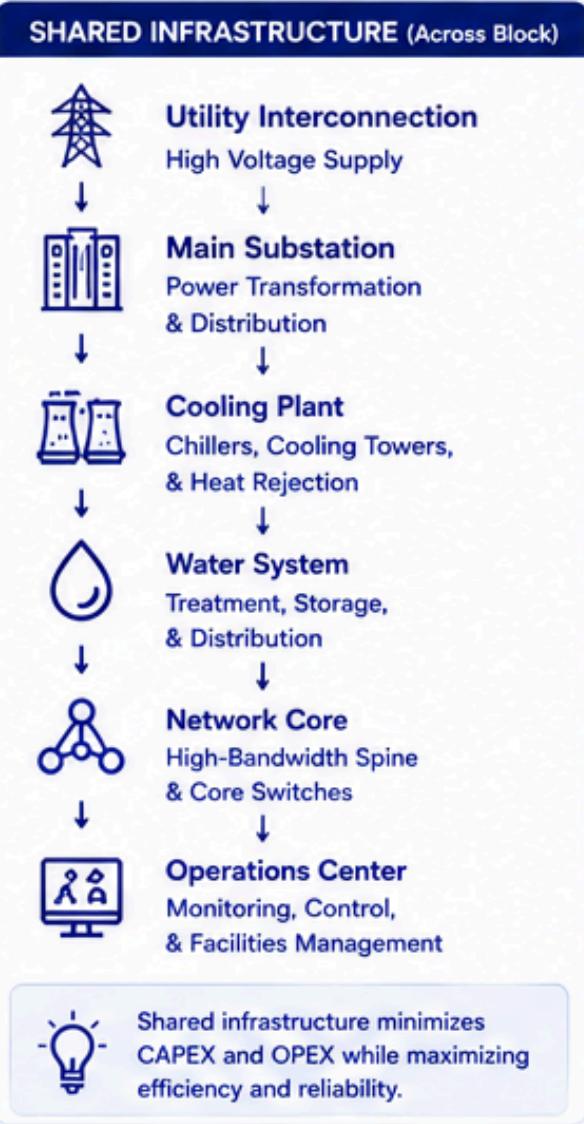
2-3
AI Data Halls
Per Block



Shared Utility Infrastructure
Optimized CAPEX & OPEX



Campus Expansion Ready
Repeatable & Scalable

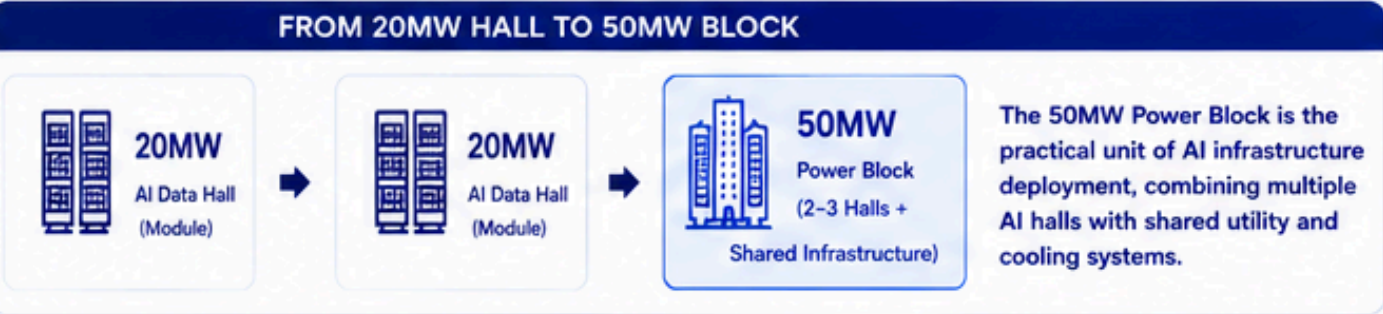


BLOCK CAPACITY & UTILIZATION (N+1 PHILOSOPHY)

50MW Design Capacity (Infrastructure) = **40MW** Initial Deployment (IT Load) + **10MW** Expansion Reserve (Infrastructure Reserve)

- Provides headroom for future IT growth
- Ensures reliability and operational flexibility
- Supports technology evolution without major rework

WHAT IS SHARED?		
Component	Shared Across 50MW Block?	Benefit
Substation (HV/LV)	✓	Lower CAPEX
Cooling Plant	✓	Higher Efficiency
Water System	✓	Lower OPEX
Network Core	✓	Scalability
Security System	✓	Operational Efficiency
AI Data Halls (IT Load)	✗	Scalable Independently



 A properly designed 50MW Power Block provides the right balance of capacity, efficiency, and flexibility—enabling seamless growth from today's deployment to tomorrow's AI factory.

300MW AI Campus Architecture

From Power Blocks to AI Campus – Scale, Integration and Efficiency





300MW Total Capacity
Large-scale AI campus power deployment



6 × 50MW Power Blocks
Modular design for flexibility and growth



Centralized Cooling
High-efficiency thermal management system



Campus Network Core
High-speed, low-latency connectivity backbone



Multi-Generation Support
AC / Hybrid AC-DC / 800V DC ready



UTILITY CONNECTION
Multiple utility feeds for reliability

230kV / 345kV AC



CAMPUS SUBSTATION
High-efficiency power conversion

Step-down to 34.5kV



POWER DISTRIBUTION
Redundant ring main units

34.5kV Distribution



POWER BLOCKS
6 × 50MW blocks for AI workloads

50MW per Block



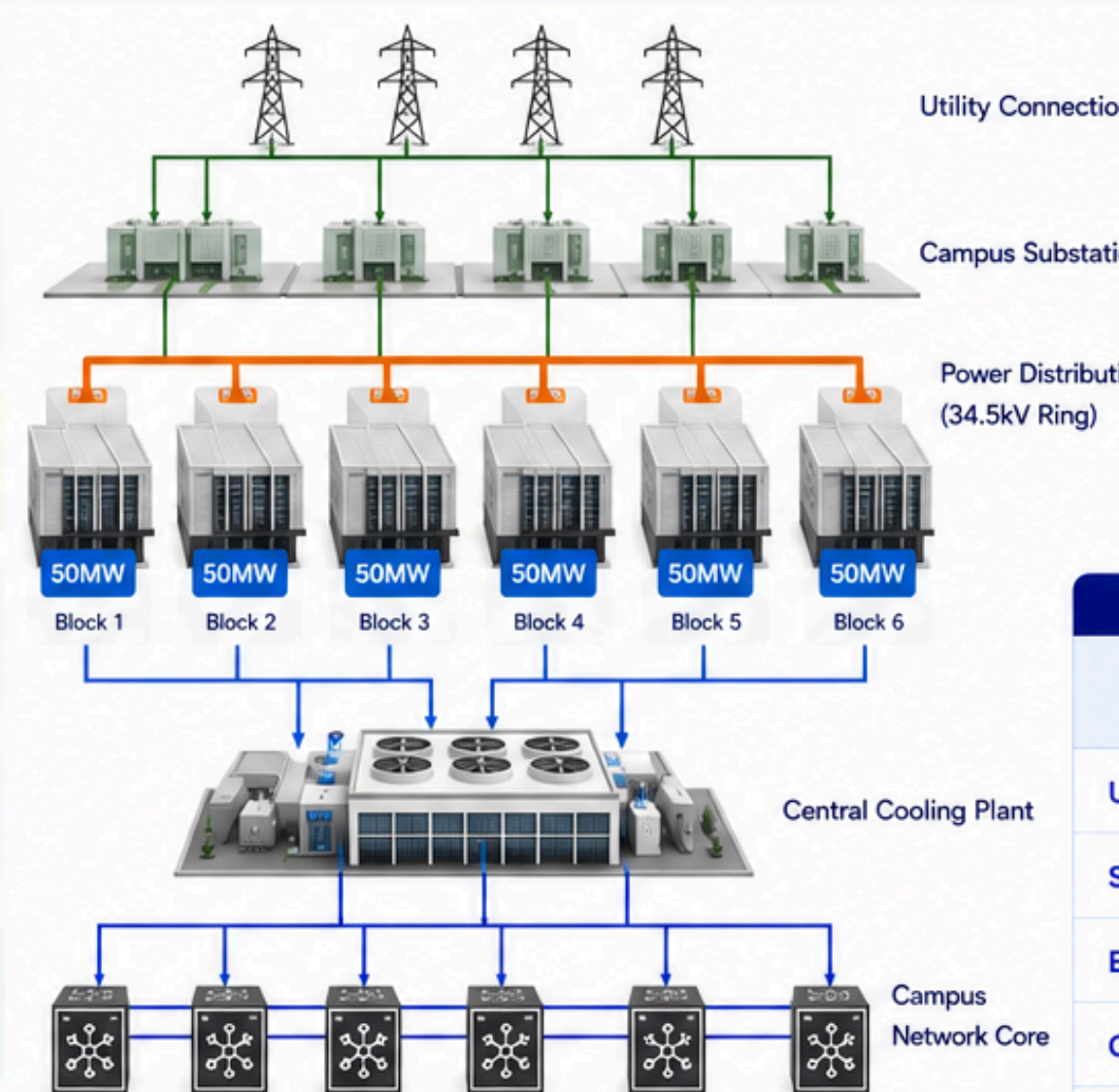
CENTRAL COOLING PLANT
High-efficiency cooling for entire campus

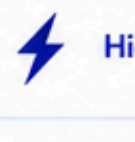
N+1 / N+N Redundancy



CAMPUS NETWORK CORE
High-speed, low-latency network backbone

400G / 800G Ethernet





High-Density Power

300MW total IT capacity for massive AI workloads



Energy Efficient

Optimized power conversion and cooling efficiency



Scalable Architecture

Modular blocks enable seamless future expansion



Resilient Design

Redundant power, cooling and network for high uptime

CAMPUS ECONOMICS – SCALE ADVANTAGE

Metric	Single AI Hall (~20MW)	50MW Power Block	300MW AI Campus
Utility Leverage	Low	Medium	High
Shared Infrastructure	Low	Medium	High
Expansion Flexibility	Medium	High	High
Operational Efficiency	Medium	High	High
Cost Efficiency	Baseline	Improved	Optimized



Reliable Power
Redundant power paths and equipment



Advanced Cooling
Centralized & efficient thermal management



High-Speed Network
Low-latency, high-bandwidth campus connectivity



Operational Excellence
Intelligent monitoring and automation



Lower TCO
Economies of scale drive cost efficiency

Internal Engineering Reference | Version 2026

11 / 14

Topics: GB300 NVL72 • Vera Rubin • SST • 800V DC Distribution • 20MW Data Hall • AI Factory

GW-Scale AI Factory Architecture

Building Next-Generation AI Infrastructure at Unprecedented Scale





GW-Scale Capacity
Gigawatt-scale AI factory for frontier workloads



Multi-Source Energy
Diverse energy portfolio for reliability and resilience



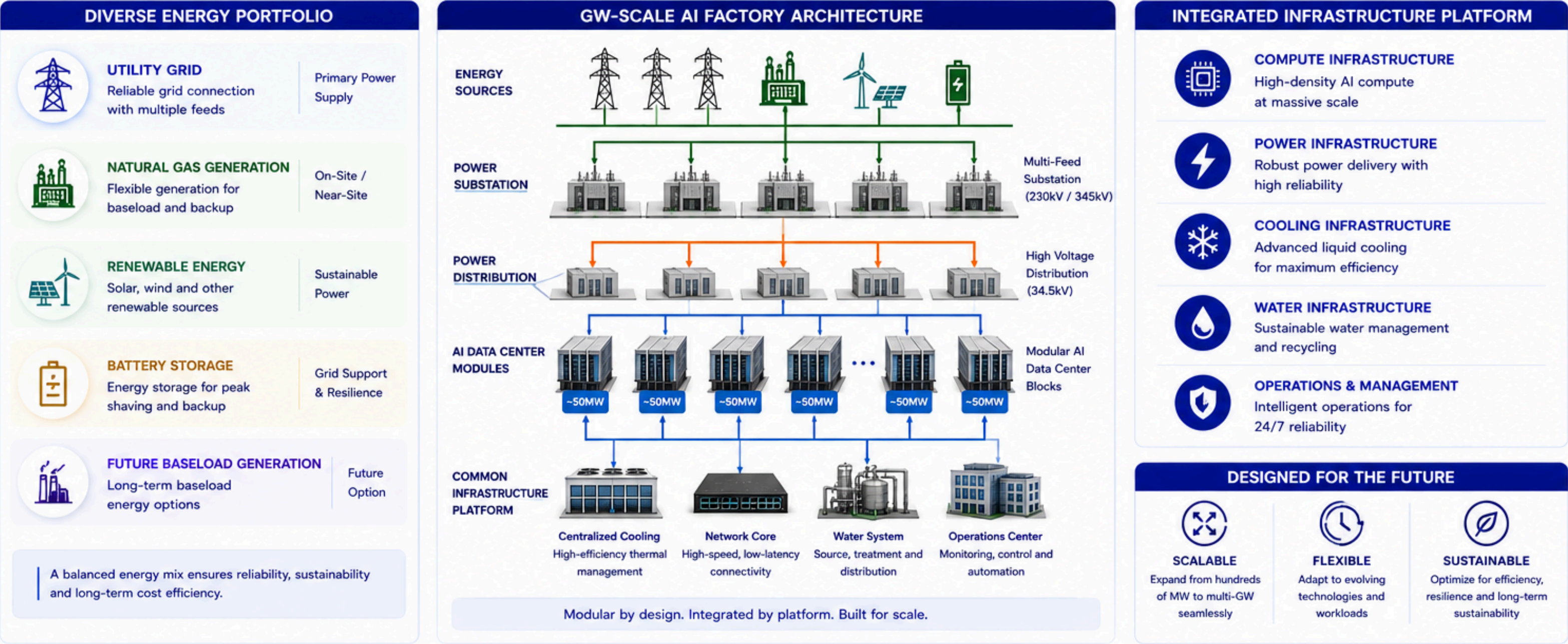
Integrated Infrastructure Platform
Compute, power, cooling and energy convergence



Built for Longevity
Designed for 20+ years of continuous operation



Future-Ready Design
Modular, scalable and technology agnostic



Multi-Generation AI Factory Framework

Designing Infrastructure That Supports Multiple Generations of AI Platforms



Future-Proof Design
Built for continuous technology evolution



Protect Capital Investment
Maximize asset life and ROI



Maximize Utilization
Higher efficiency across generations



Sustainable Platform
Lower TCO and energy intensity



Technology Agnostic
Open architecture for future innovation

WHY MULTI-GENERATION MATTERS



Future-Proof Infrastructure
Designed for continuous change



Protects Capital Investment
Longer asset life with lower replacement cost



Continuous Upgrade Path
Evolve performance without rebuilding the foundation



Technology Agnostic Design
Support current and future AI platforms



Long-Term Asset Utilization
Maximize the value of every infrastructure dollar



WHAT EVOLVES



AI Platform & Rack System
New generations deliver higher performance and capabilities



Power Distribution
Evolving from AC to hybrid AC/DC to 800V DC-ready



Cooling Capacity
Increase capacity and efficiency to support higher densities



Network Throughput
Expanding bandwidth and lower latency



Management & Software
Evolving tools and automation for greater intelligence



These elements evolve as technology advances.

AI PLATFORM EVOLUTION – ONE AI FACTORY, MULTIPLE GENERATIONS

**2026
GB300 NVL72
(Current Generation)**



- Highest Performance Today
- Mature Ecosystem
- Proven at Scale

**2026+
Vera Rubin NVL72
(Next Generation)**



- More Memory
- Higher Bandwidth
- Better Efficiency

**2027+
Rubin Ultra
(Ultra Dense Generation)**



- Higher Density
- More Compute
- Greater Performance

**Future Platforms
(Kyber and Beyond)**



- New Architectures
- Advanced Interconnects
- Continued Innovation



DIFFERENT AI GENERATIONS. SAME AI FACTORY FOUNDATION.
The foundation you build today is ready for tomorrow.

INFRASTRUCTURE LAYERS – BUILT TO EVOLVE

Layer	Purpose	Today (2026)	Future Evolution	Design Approach
Building Shell	Structure & Envelope	Standardized for durability and flexibility	Same	Build once, last for decades
Hall Layout	Space & Deployment	20MW Halls in 50MW Blocks	Same	Consistent for operational efficiency
Cooling Backbone	Thermal Management	High-efficiency cooling infrastructure	Increase capacity (flow, temp, redundancy) Same architecture	Design for higher loads and flexibility
Network Fabric	Connectivity	High-bandwidth spine-leaf fabric	Expand bandwidth (400G → 800G → 1.6T+) Same network topology	Topology remains, capacity grows
Power Distribution	Power Delivery	AC Distribution	Evolve to Hybrid AC/DC → 800V DC-ready	Flexible architecture for future power
Operations Platform	Management & Monitoring	Unified platform	Expand capabilities with AI & Automation	Support multiple generations

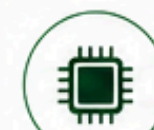
INFRASTRUCTURE LIFECYCLE – DESIGNED FOR LONGEVITY

Asset Layer	Scope	Typical Lifecycle	Notes
AI Platform + Rack System	GPU, Server, Rack, Networking within the rack	5+ Years (Typical 5–7 Years)	Evolve with new generations, not replaced annually
Cooling Systems	CRAH/CRAC, Pumps, CDU, Heat Rejection	10 – 15+ Years	Components upgraded, infrastructure remains
Electrical Systems	MV/LV, Switchgear, UPS, Busway, PDUs	15 – 20+ Years	Designed for expansion and technology change
Campus Infrastructure	Buildings, Roads, Utility Yard, Water, Land	20+ Years	Long-term asset for multi-decade operation



Long-lived infrastructure layers protect capital investment and maximize long-term asset value.

INFRASTRUCTURE REPLACEMENT CYCLE



AI Platform + Rack System
5+ Years
Evolve with each new generation



Support Systems (Power, Cooling, Network)
10+ Years
Upgrade components as needed



Facility Infrastructure
15+ Years
Major systems built for long-term use



Campus Infrastructure
20+ Years
Multi-decade asset foundation



The AI Factory foundation remains. Technology evolves. Value is preserved.

INFRASTRUCTURE COMPATIBILITY PRINCIPLES



Standardization
Common standards enable seamless integration



Modularity
Replace or upgrade components, not the foundation



Scalability
Capacity can be expanded in phases



Interoperability
Open ecosystems and industry standards



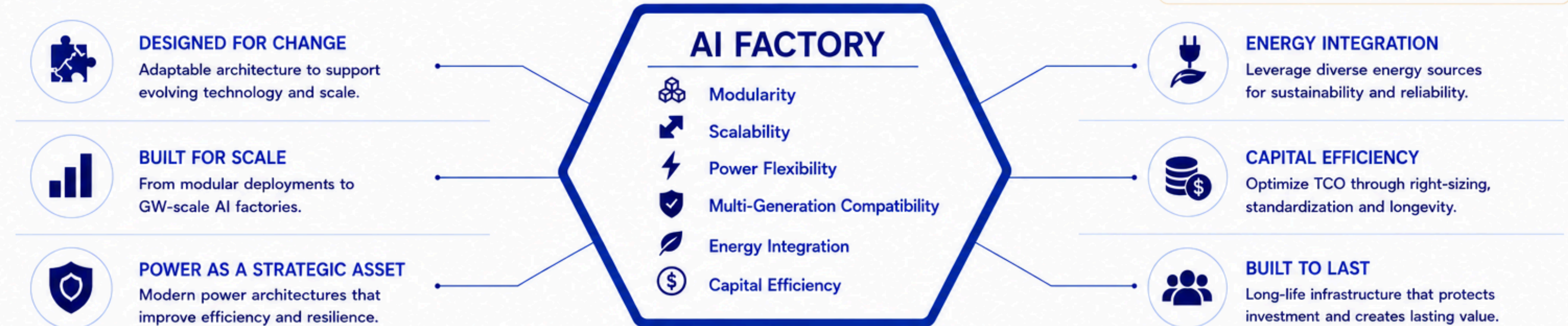
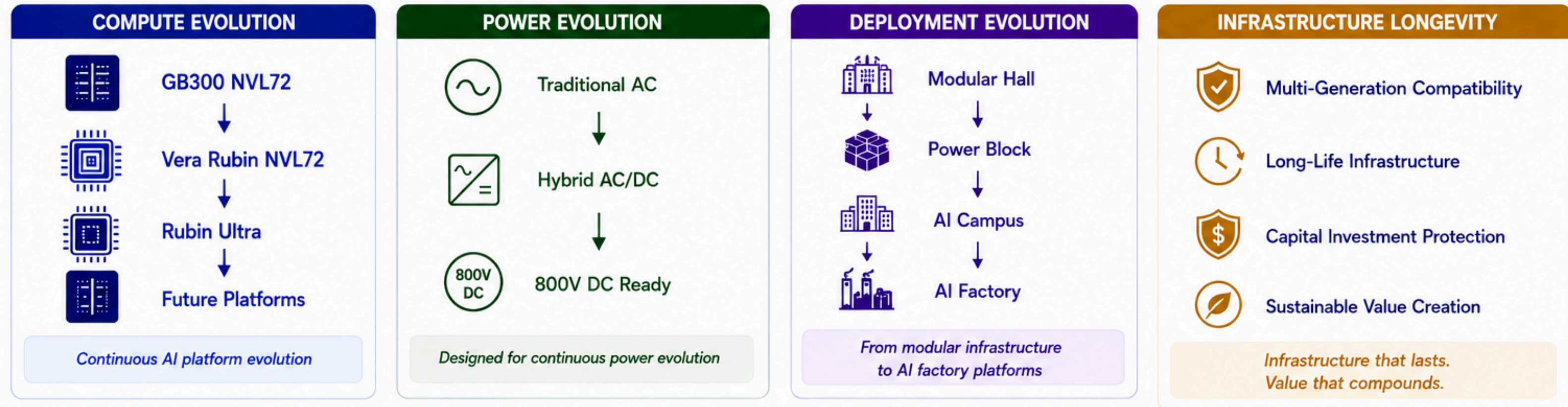
Resilience
Built-in redundancy for long-term reliability



BUILD ONCE. SUPPORT MULTIPLE GENERATIONS. PROTECT LONG-TERM CAPITAL INVESTMENT.

AI Factory Design Principles

Enduring Principles for Building the Next Generation of AI Infrastructure



AI FACTORIES ARE THE FOUNDATIONS OF THE AI ECONOMY.

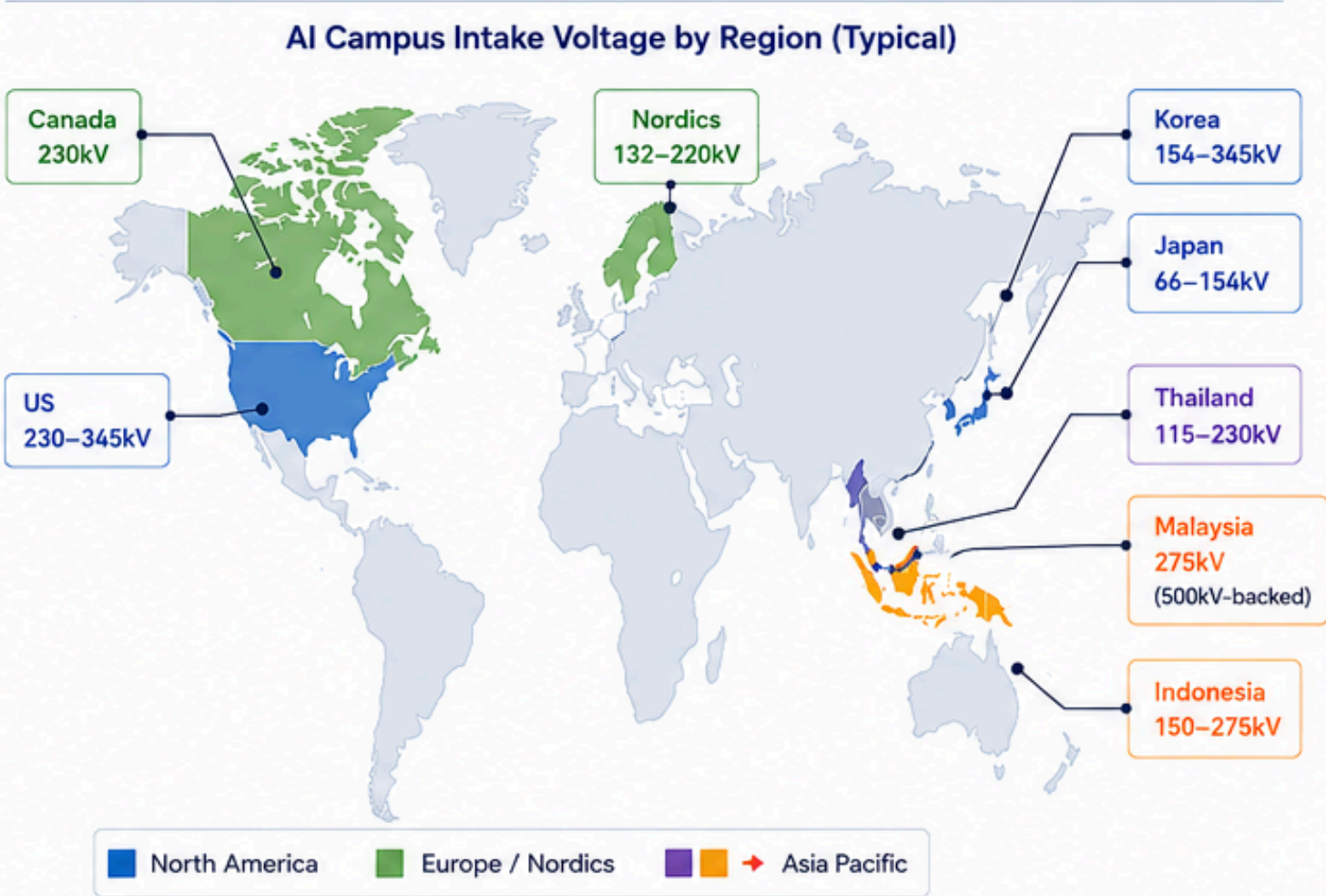
BUILT ON ENDURING PRINCIPLES. DELIVERING LASTING VALUE.



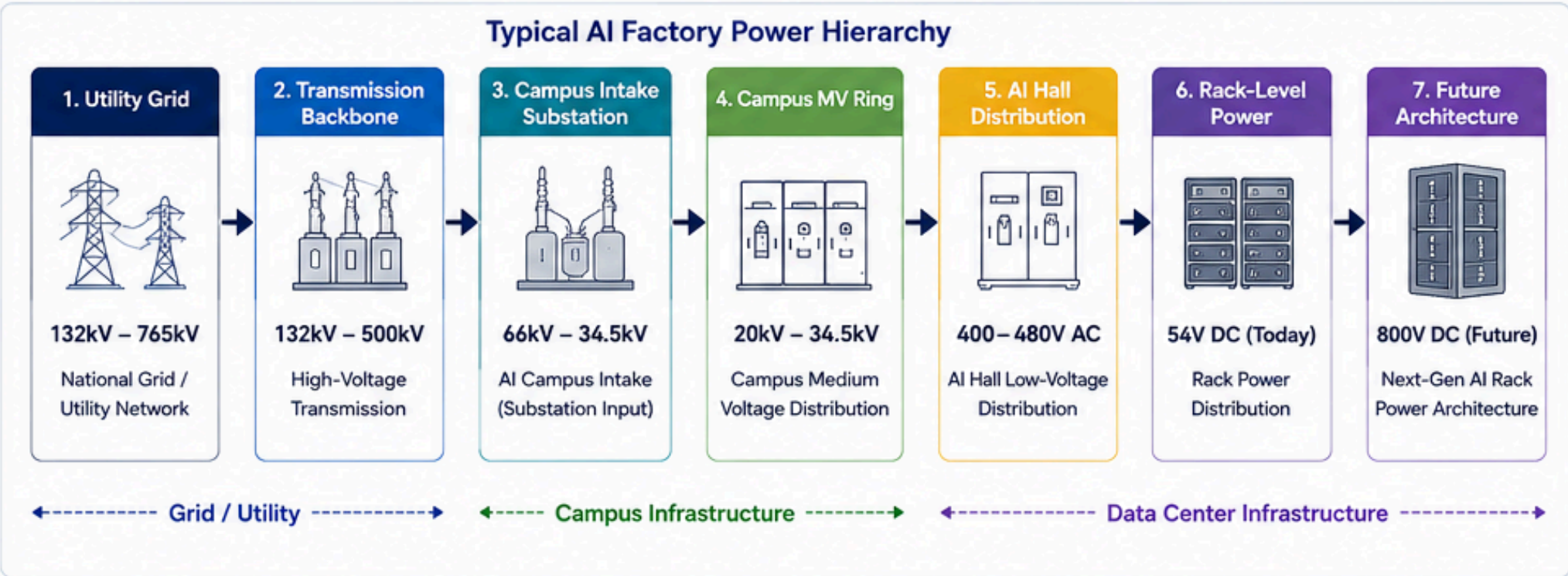
Global Grid-to-Campus Power Architecture Benchmark

Representative High-Voltage Power Architecture for Large-Scale AI Infrastructure Development

Region	Transmission Backbone (Utility Network)	Typical AI Campus Intake (Substation Input)	Campus Distribution (Internal MV Ring)
 Malaysia	500kV / 275kV	275kV (500kV-backed)	33kV
 Indonesia	500kV / 275kV / 150kV	150–275kV	20–33kV
 Thailand	500kV / 230kV / 115kV	115–230kV	22–33kV
 Japan	500kV / 275kV / 154kV	66–154kV	6.6–22kV
 South Korea	765kV / 345kV / 154kV	154–345kV	22.9kV
 Nordic Region (EPG*)	400kV / 220kV / 132kV	132–220kV	11–33kV
 Canada	500kV / 230kV	230kV (500kV-backed)	25–34.5kV
 United States	765kV / 500kV / 345kV / 230kV / 138kV / 115kV	230–345kV	13.8–34.5kV



Note: Voltage levels vary depending on utility, location, and project scale.



Key Observations

- Grid Access Is the Primary Constraint**
The bottleneck for next-generation AI factories is increasingly determined by transmission-level power availability rather than rack-level power density.
- Direct High-Voltage Access Is Emerging**
Large-scale AI campuses (300MW+) are progressively securing direct access to transmission infrastructure, typically ranging from 220kV to 500kV depending on regional utility standards.
- Campus Distribution Remains Regional**
While transmission-level intake voltages are converging globally, campus medium-voltage distribution continues to follow local utility practices, commonly ranging from 20kV to 34.5kV.
- AI Power Architectures Continue to Evolve**
The industry is transitioning from traditional 12V server architectures toward 54V rack-level distribution, with future AI factory designs increasingly evaluating 800V DC architectures to improve efficiency and scalability.

Source: Regional utility transmission standards, hyperscale data center deployment practices, and publicly disclosed AI infrastructure projects across North America and Asia-Pacific.

*EPG: European Power Grid (Nordic Region)

AI Campus Power Standardization Framework

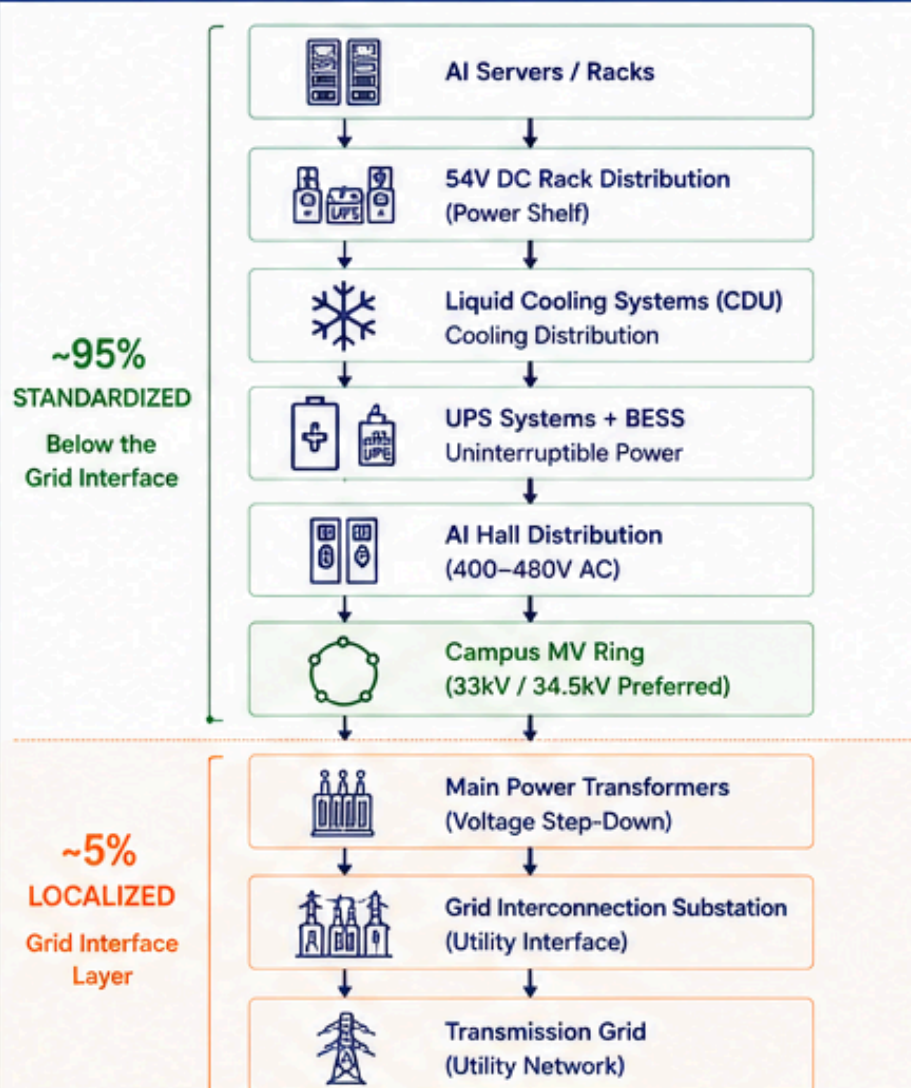
Standardize Below the Grid Interface. Scale Globally.

Maximize global standardization of AI campus power infrastructure while adapting only what is required at the grid interface.

1. STANDARDIZED VS LOCALIZED POWER INFRASTRUCTURE

Infrastructure Layer	Standardized Globally (Minimize Change)	Localized by Country (Minimize Scope)
 AI Servers / Racks	✓	—
 Rack Power Shelf	✓	—
 54V DC Architecture	✓	—
 Liquid Cooling Systems (CDU)	✓	—
 CDU Support Systems	✓	—
 UPS Systems	✓	—
 Battery Energy Storage (BESS)	✓	—
 AI Hall Distribution (400–480V AC)	✓	Minor (voltage range)
 Campus MV Ring (33kV / 34.5kV Preferred)	✓	Minor (voltage level)
 Medium Voltage Switchgear	✓	Minor (breakers, protection)
 Main Power Transformers	—	✓ Yes (ratio, size, configuration)
 Grid Interconnection Substation	—	✓ Yes (design, layout, equipment)
 Utility Protection & Relay Systems	—	✓ Yes (schemes, settings, devices)
 Utility Compliance & Grid Codes	—	✓ Yes (standards, approvals)

2. GLOBAL AI CAMPUS REFERENCE ARCHITECTURE



GLOBAL STANDARDIZATION IMPACT



Standardize ~95% of Infrastructure | Localize ~5% at the Grid Interface

4. KEY INSIGHTS



High Degree of Global Standardization

AI campus infrastructure below the medium-voltage layer can be standardized across countries, enabling rapid replication, lower cost, and supply chain efficiency.



Localization Concentrated at Grid Interface

Differences across regions are primarily at the grid interface: transmission voltage, transformer configuration, protection requirements, and utility compliance.



Platform Repeatability

A standardized architecture enables repeatable engineering design, modular procurement, and scalable deployment across global markets.



Lower Risk, Faster Deployment

Standardization reduces design complexity and project risk, resulting in shorter development cycles and more predictable execution.

3. COUNTRY-LEVEL ADAPTATION MATRIX

Parameter	 Malaysia	 Indonesia	 Thailand	 Japan	 South Korea	 Nordic Region (Nordics)	 Canada	 United States
Typical Utility Intake (Substation Input)	275kV (500kV-backed)	150–275kV	115–230kV	66–154kV	154–345kV	132–220kV	230kV (500kV-backed)	230–345kV
Preferred Platform MV Ring	33kV	33kV	22–33kV	22kV	22.9kV	33kV	34.5kV	34.5kV
Alternative MV Options	22kV	20kV / 33kV	22kV	6.6kV / 22kV	—	20kV / 10kV	25kV	13.8kV / 25kV
Key MV Standards	IEC	IEC	IEC / TIS	JIS	KEC	IEC	CSA	IEEE / NEMA
Standardized Portion (Estimated %)	~90–95%	~90–95%	~90–95%	~85–90%	~90–95%	~90–95%	~90–95%	~90–95%

Note: Voltage levels vary depending on utility, location, and project scale. This framework represents common practices across global AI datacenter and AI factory projects.

5. KEY TAKEAWAY

“

Transmission voltage may vary by country, but the AI campus architecture below the grid interface remains fundamentally the same.



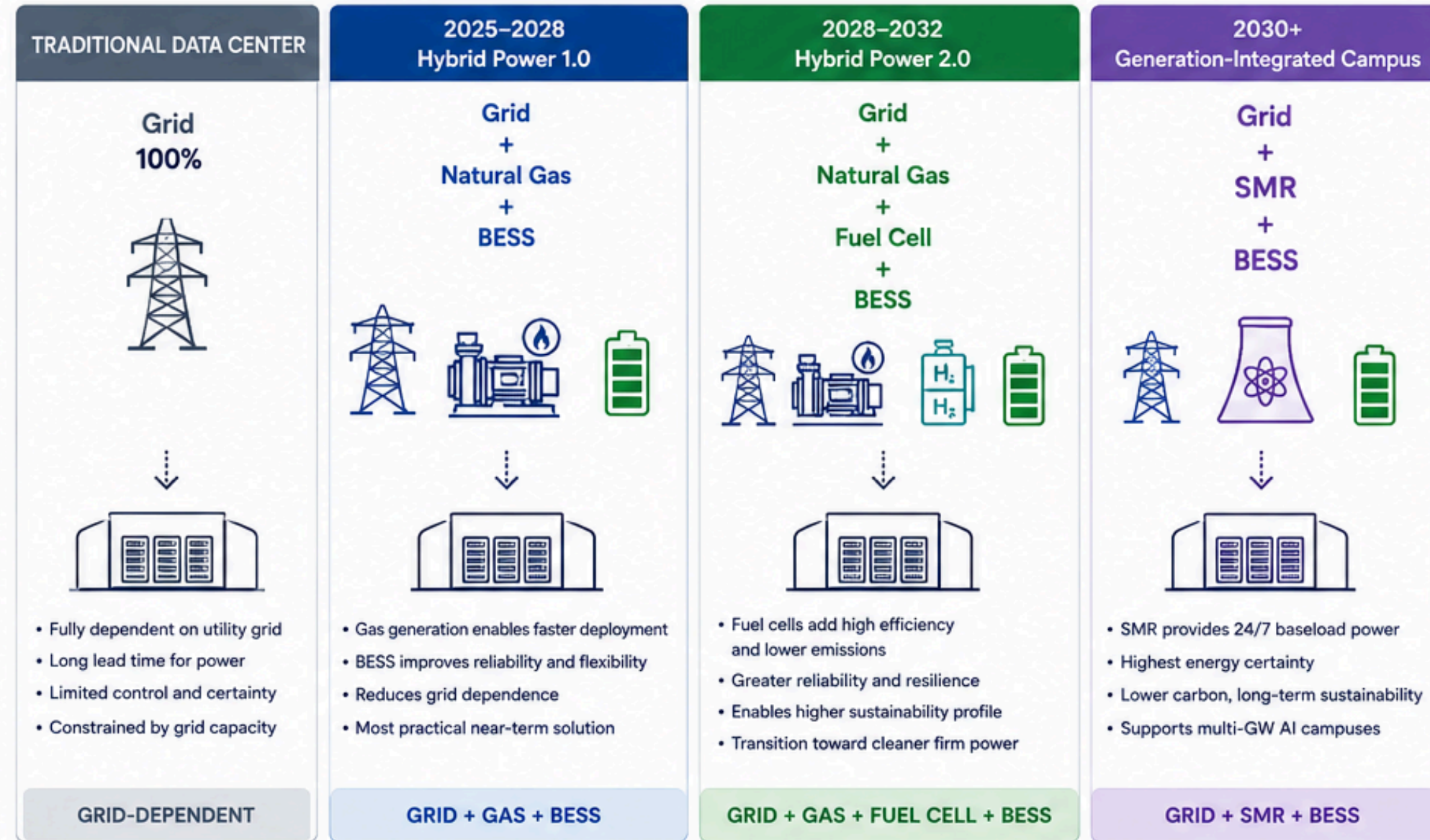
- Standardize ~95% of the infrastructure below the medium-voltage layer.
- Localize only what the utility requires at the grid interface.
- Deliver a global AI infrastructure platform with local adaptability.

AI Campus Power Generation Roadmap

Evolution from Grid-Dependent to Generation-Integrated AI Factories

AI factories require reliable, scalable, and sustainable power. The power strategy is evolving from full grid dependence to generation-integrated campuses, ensuring speed-to-power and energy certainty.

1. AI INFRASTRUCTURE POWER EVOLUTION

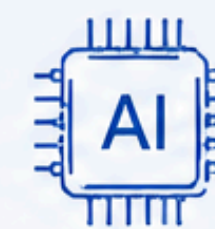


EVOLUTION SUMMARY



2. KEY DRIVERS OF POWER EVOLUTION

AI Workloads Are Becoming Baseload Infrastructure



Traditional Cloud
(Variable Demand)



AI Factory
(24x7 Constant Demand)



AI training and inference require consistent, uninterrupted power.

Grid Availability Is Becoming the Bottleneck



Compute Availability
✗ Not the Constraint

Power Availability
✓ Primary Constraint

Power and energy capacity, not compute, will limit AI scale.

Behind-the-Meter Generation Is Emerging



Grid Power
(Passive)



Hybrid Power
(Shared)



Dedicated Generation
(Controlled)

AI campuses are evolving from power consumers to power system participants.



KEY TAKEAWAY

Future AI campuses will increasingly combine utility power, on-site generation, and energy storage to achieve faster deployment, higher power certainty, and long-term sustainability.



AI factories require 24x7 reliable power at gigawatt scale. The optimal generation mix evolves over time, balancing speed-to-power, cost, scalability, sustainability, and energy security.

1. GENERATION TECHNOLOGY COMPARISON

Technology	Deployment Speed (Time to Power)	Capital Cost (Capex Intensity)	Scalability (GW Scale Capability)	Dispatchability (24x7 Reliability)	AI Suitability (Base Load Support)	Key Advantages
 Utility Grid (Standalone)	★★★★★	★★★★★	★★	★★★★★	★★★	- No on-site Capex - Proven and mature - Subject to grid constraints
 Natural Gas Turbine (CCGT)	★★★★	★★★	★★★★★	★★★★★	★★★★★	- Fast deployment - High power density - Flexible and scalable
 Fuel Cell (PEM / SOFC)	★★★	★★★	★★★★	★★★★★	★★★★	- High efficiency - Low emissions - Modular and scalable
 Solar + BESS	★★	★★	★★	★	★★	- Renewable and clean - Low operating cost - Intermittent by nature
 Wind + BESS	★★	★★	★★	★	★★	- Renewable - Low operating cost - Intermittent by nature
 SMR Nuclear	★	★	★★★★★	★★★★★	★★★★★	24x7 baseload power - Lowest carbon - Highest energy density

★★★★★ Excellent ★★★★ Very Good ★★★ Good ★★ Fair ★ Poor

4. GENERATION TECHNOLOGY SUMMARY INSIGHTS



Natural Gas
Is the Near-Term
Workhorse

Fast, scalable, proven, and capable of supporting hundreds of MW quickly.



Fuel Cells Are the
Clean Transition
Enabler

High efficiency and low emissions make fuel cells ideal for hybrid systems.



SMR Is the
Long-Term
Endgame

Nuclear power provides the most reliable, clean, and scalable baseload for AI factories.



Renewables Play
a Supporting Role

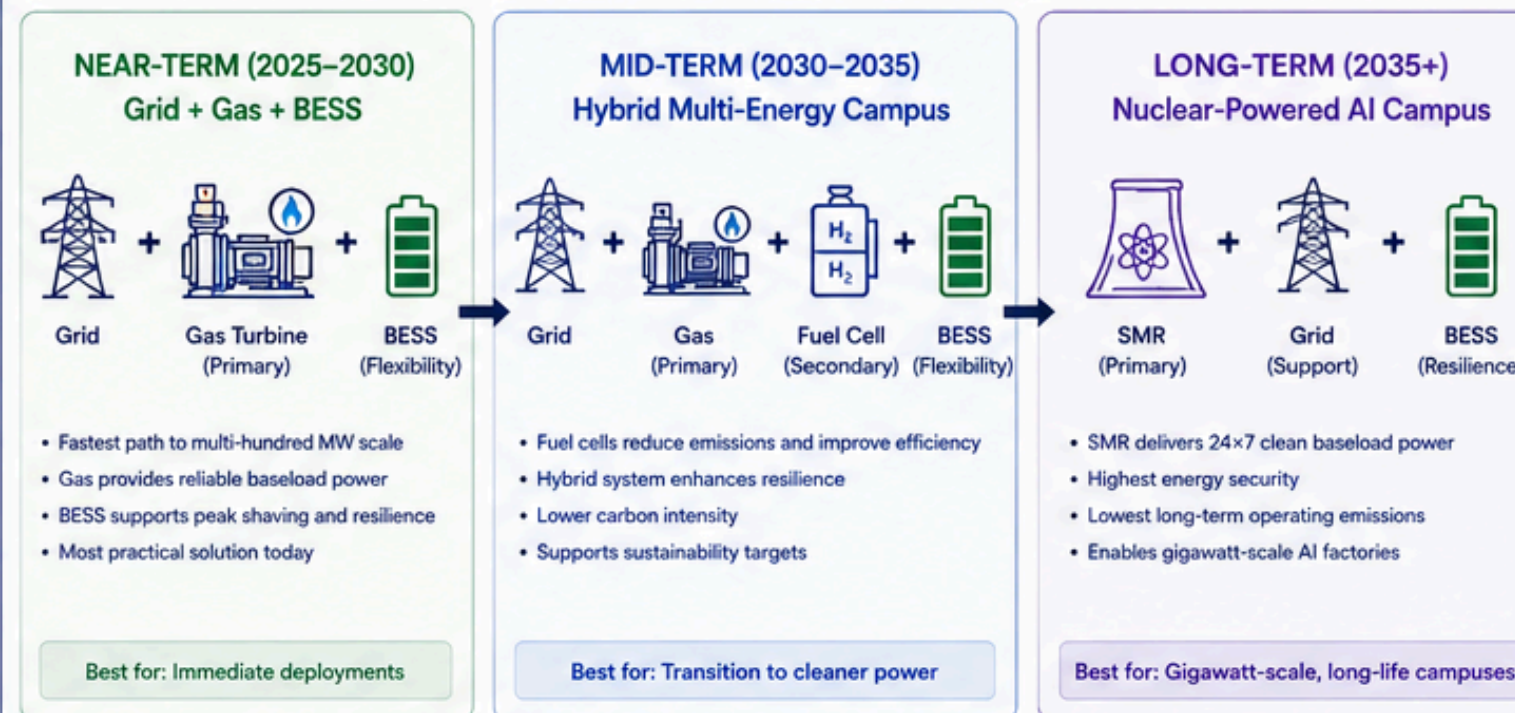
Solar and wind are valuable for sustainability goals but cannot be the primary source for 24x7 AI loads.



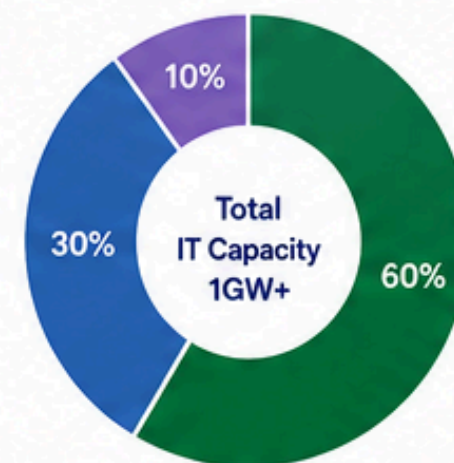
Energy Storage Is
Always Essential

BESS is critical across all stages to ensure flexibility, resilience, and power quality.

2. AI CAMPUS ENERGY ARCHITECTURE EVOLUTION



3. ILLUSTRATIVE 1GW AI FACTORY ENERGY MIX (LONG-TERM VISION)



	SMR Nuclear	600 MW (60%)	24x7 baseload - Lowest carbon - Long-term cost stability
	Utility Grid	300 MW (30%)	- Supports variability - Provides additional capacity - Enhances reliability
	BESS (Battery Energy Storage)	100 MW (10%)	- Peak shaving - Frequency regulation - Backup and resilience

This mix delivers maximum reliability, sustainability, and cost efficiency for future gigawatt-scale AI factories.



**AI FACTORY
EVOLUTION**



Grid-Dependent
(Traditional Data Center)



Grid-Assisted
(Behind-the-Meter Generation)



Generation-Integrated
(Multi-Energy Campus)



Energy Self-Sufficient
(Future AI Factory)

KEY TAKEAWAY

Natural gas is the most practical solution for today's deployments, while SMRs represent the most promising pathway for sustainable, gigawatt-scale AI factories.

APPENDIX E Chinese Power Generation Supply Chain for AI Infrastructure

Domestic Generation Technologies Supporting AI Campus Deployment

China has developed a full-spectrum power generation ecosystem covering gas engines and gas turbines from 1MW to 300MW+, capable of supporting AI campuses from tens of MW to over 1GW.

1. CHINESE SUPPLIER LANDSCAPE

A. MODULAR GAS ENGINE PLATFORM (1–5MW)

Supplier (中文)	Technology	Unit Size	Key Features	Typical Application
 WEICHA 潍柴动力	Gas Engine (天然气发动机) 	1–5MW	- High efficiency - Wide load range - Fast start capability - Modular & scalable	Distributed Generation Data Centers
 JICHA 济柴动力 (CNPC)	Gas Engine (天然气发动机) 	1–3.5MW	- Mature technology - High reliability - Low maintenance	Distributed Generation Industrial Power
 YUCHAI 玉柴机器	Gas Engine (天然气发动机) 	1–5MW	- Compact design - Low fuel consumption - Flexible configuration	Industrial Power Backup Power Microgrid

B. INDUSTRIAL & HEAVY-DUTY GAS TURBINE PLATFORM (5–300MW)

Supplier (中文)	Technology	Unit Size	Key Features	Typical Application
 AECC 中国航发 AECC	Industrial Gas Turbine (工业燃气轮机) 	5–30MW	- Advanced aerodynamics - High efficiency - Low emissions - Quick deployment	AI Campus Industrial Power Cogeneration
 东方电气 DONGFANG ELECTRIC	Heavy-Duty Gas Turbine (重型燃气轮机) 	50–300MW	- F-class technology - High reliability - Proven performance	Utility Power Large Scale Plant Independent Power
 哈尔滨电气 HARBIN ELECTRIC	Heavy-Duty Gas Turbine (重型燃气轮机) 	50–300MW	- F-class technology - High efficiency - Combined cycle ready	Utility Power IPP Projects Cogeneration
 上海电气 SHANGHAI ELECTRIC	Heavy-Duty Gas Turbine (重型燃气轮机) 	50–300MW	- Advanced F-class - High reliability - Integrated solutions	Utility Power Large Scale Plant Industrial Power

2. AI CAMPUS SIZE vs. RECOMMENDED CHINESE SOLUTION

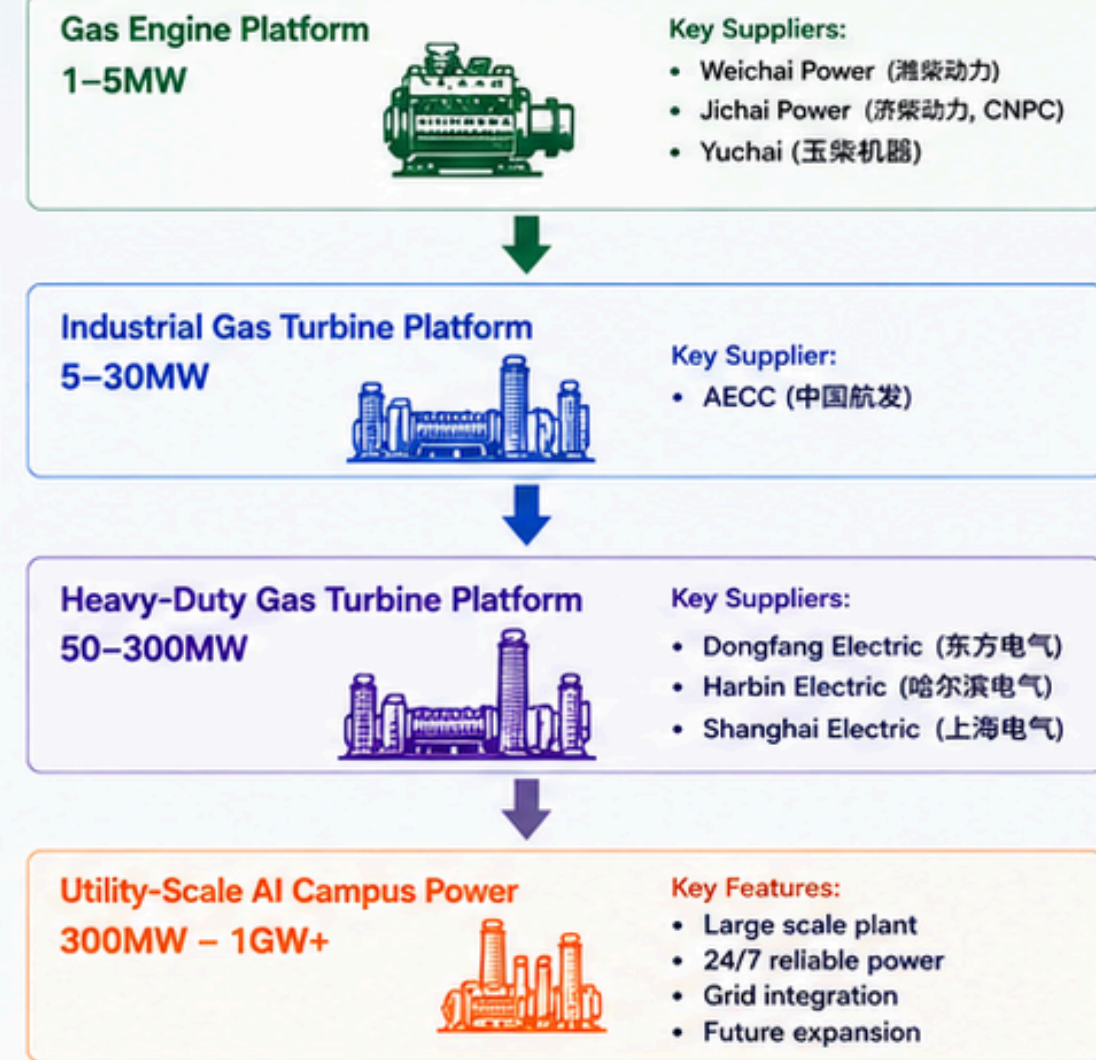
AI Campus Size	Typical Power Demand	Recommended Chinese Solution	Technology Platform
 50MW Campus	30 – 60MW	AECC QD50 Series (5MW Class)	Industrial Gas Turbine
 100MW Campus	60 – 120MW	AECC + Gas Engine Hybrid (Optimized Mix)	Industrial GT + Gas Engine
 200MW Campus	120 – 240MW	Multiple AECC Turbines (QD50 / QD70 Series)	Industrial Gas Turbine
 300MW+ Campus	240 – 500MW	Multiple AECC Turbines or Hybrid with Heavy-Duty GTs	Industrial / Heavy-Duty Gas Turbine
 500MW+ Campus	500MW – 1GW+	Dongfang / Harbin / Shanghai Electric (50–300MW Class)	Heavy-Duty Gas Turbine
 1GW+ Campus (Future AI Factory)	1GW+	Utility-Scale Gas Turbine Plant (F-class / H-class)	Heavy-Duty Gas Turbine

3. INTERNATIONAL DEPLOYMENT READINESS

 LOW CERTIFICATION BARRIER (Generally No Major Re-certification Required)	 Malaysia  Indonesia  Thailand  Vietnam  Philippines  UAE  Saudi Arabia  Africa Typically accepts international standards (ISO / IEC) and project-based approvals.
 MEDIUM CERTIFICATION BARRIER (Local Utility Approval / Additional Standards)	 Japan (日本) - Utility interconnection approval - JIS compliance - Project-specific certification  South Korea (韩国) - Utility interconnection approval - KEC / Local standards compliance - Project-specific certification
 HIGH CERTIFICATION BARRIER (Extensive Certification & Compliance Required)	 United States (美国) - UL / IEEE / ANSI compliance - EPA emissions certification - CARB (if applicable) - Utility interconnection approval  Canada (加拿大) - CSA certification - Utility interconnection approval - Environmental compliance

4. CHINESE AI POWER SUPPLY CHAIN

From Modular to Utility Scale



5. STRATEGIC ASSESSMENT

NEAR-TERM OPPORTUNITY Core Markets for Chinese Suppliers	LONG-TERM CHALLENGE Key Barriers to Enter North America
 Southeast Asia  Middle East  Africa	 United States  Canada
- Lower CapEx Advantage - Shorter Manufacturing Lead Time - Flexible Project Financing - Growing Export Experience - Strong After-Sales Network	- Utility Acceptance - Emissions Compliance (EPA / CARB) - UL / CSA Certification - Financing & Bankability

KEY TAKEAWAY

Chinese power generation suppliers offer a complete portfolio from 1MW gas engines to 300MW+ gas turbines. They are highly competitive across Southeast Asia, the Middle East, and Africa. For North America, certification, utility acceptance, and financing bankability are the primary barriers rather than technology capability.



APPENDIX F

Global Power Generation Supply Chain for AI Infrastructure

Leading International Suppliers Supporting AI Campus Deployment

Global OEMs provide proven gas engine and gas turbine solutions with strong track record in AI campus, data centers, and utility-scale power projects worldwide.

1. GLOBAL SUPPLIER LANDSCAPE

A. Modular Gas Engine Platform (2–20MW)

Supplier	Country	Technology	Unit Size	Key Strengths	AI Campus Suitability
 CATERPILLAR	USA	Gas Engine	2–10MW	- High reliability - Global service network - Fast delivery	★★★★★
 Cummins	USA	Gas Engine	2–10MW	- Fuel flexible - High efficiency - Strong after-sales	★★★★★
 Rolls-Royce mtu	Germany	Gas Engine	2–10MW	- Advanced technology - Low emissions - High power density	★★★★★
 WÄRTSILÄ	Finland	Gas Engine	5–18MW	- High efficiency - Flexible operation - Fast start capability	★★★★★

B. Industrial & Heavy-Duty Gas Turbine Platform (30–500MW)

Supplier	Country	Technology	Unit Size	Key Strengths	AI Campus Suitability
 GE VERNOVA	USA	Heavy-Duty Gas Turbine	30–500MW	- HA-class technology - High efficiency - Global reference fleet	★★★★★
 SIEMENS energy	Germany	Heavy-Duty Gas Turbine	30–500MW	- Advanced H-class turbines - High efficiency - Strong digital capability	★★★★★
 MITSUBISHI POWER	Japan	Heavy-Duty Gas Turbine	30–500MW	- Advanced JAC / H-class - High reliability - Strong project execution	★★★★★
 ANSALDO ENERGIA	Italy	Heavy-Duty Gas Turbine	30–500MW	- Proven technology - Flexible configurations - Strong EPC capability	★★★★☆

2. CURRENT AI CAMPUS DEPLOYMENT TRENDS

Campus Size	Preferred Architecture	Typical Suppliers	Key Characteristics
50–100MW Campus 	5MW Gas Engines * 10–20 Units   	 Cummins mtu	- Modular & scalable - Fast deployment - High redundancy (N+1) - Lower initial CapEx
100–300MW Campus 	5–10MW Gas Engines + BESS + Grid   	 Cummins WÄRTSILÄ	- Phased expansion - BESS for peak shaving - Grid integration - Optimal TCO
300–500MW Campus 	Gas Engines + Heavy-Duty Gas Turbines + Grid   	 WÄRTSILÄ  GE VERNOVA	- Hybrid architecture - Higher efficiency - Improved resilience - Lower \$/MWh
500MW – 1GW+ Campus 	Heavy-Duty Gas Turbines + Grid  	 GE VERNOVA  SIEMENS energy  MITSUBISHI POWER	- Utility-scale solution - High efficiency (60%+) 24/7 baseload capable - Proven at scale
1GW+ Future AI Factory 	SMR + + Grid- +   		Long-term vision for sustainable, 24/7, low-carbon AI factories

3. GLOBAL CERTIFICATION ADVANTAGE

Widely Accepted Standards



Deployment Readiness by Region

Region	Deployment Complexity
United States	Low
Canada	Low
Europe	Low
Australia	Low
Middle East	Low
Southeast Asia	Low
Japan	Medium
South Korea	Medium
China	Low ¹

Note: Based on established certifications and global project experience.
1. Chinese market typically requires local type testing and approvals.

4. STRATEGIC ASSESSMENT

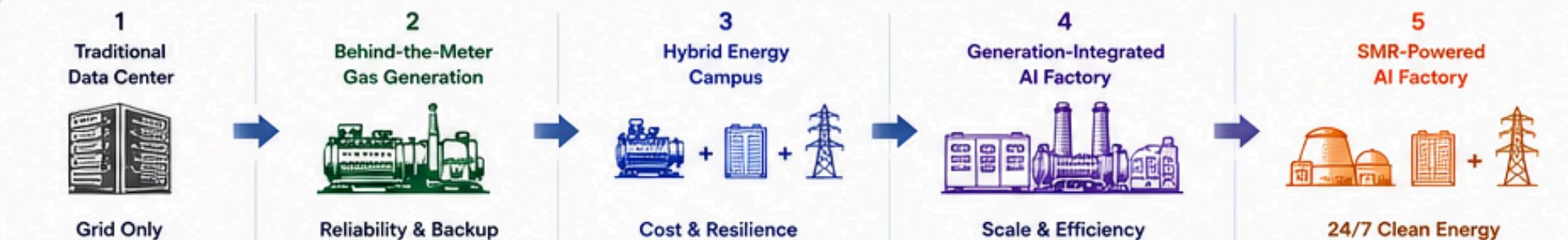
NEAR-TERM AI CAMPUS WINNER
Gas Engine Platform

- ✓ Fast deployment (months)
- ✓ Modular & flexible
- ✓ N+1 redundancy
- ✓ Lower upfront investment
- ✓ Proven technology & supply chain

LONG-TERM AI CAMPUS WINNER
SMR + Grid + BESS

- ✓ 24x7 baseload power
- ✓ Gigawatt-scale capability
- ✓ Lower long-term carbon intensity
- ✓ Energy independence
- ✓ Future-proof infrastructure

5. AI FACTORY ENERGY EVOLUTION



KEY TAKEAWAY

Modular gas engines are the preferred near-term building blocks for AI campuses due to speed, flexibility, and redundancy.
Heavy-duty gas turbines enable large-scale, high-efficiency power for gigawatt campuses and AI factories.
Over the long term, SMRs integrated with grid and BESS will become the foundation of sustainable, gigawatt-scale AI factories.



APPENDIX G

Gas Generation Integration Architecture for AI Campuses

Power Flow, System Architecture & Grid Interaction

This appendix illustrates how on-site gas generation is integrated into an AI campus, how power flows from fuel source to data hall, and how the system interacts with the utility grid.

