The AI Buildout: Data Centers, Energy,

and the Business of Scale

Contents

Executive Summary 4
Acknowledgments 6
AI Usage Acknowledgment 7
1. Introduction 8
1.1 Two Dimensions of the Challenge 8
1.2 Key Decision Points for Success 8
1.3 Where Projects Fail 9
1.4 Why Efficiency Is Not the Sole Solution 9
1.5 Why Multiple Perspectives Matter 9
2. Part 1: Powering the Buildout 11
2.1 The Real Constraint: Deliverability, Not Generation 11
2.1.1 Categories of Constraints 11
2.1.2 What "Power Available" Actually Means 11
2.1.3 The Timeline Mismatch 12
2.2 Grid-Interactive Design 12
2.2.1 Demand Flexibility 12
2.2.2 What Grid-Interactive Architecture Requires 13
2.3 Solutions that Move Faster than Transmission 13
2.3.1 Behind-the-Meter Battery Storage 13
2.3.2 Phased Power 14
2.3.3 Grid-Side Technologies 14
3. Part 2: Delivering the Buildout 15
3.1 The First 90 Days 15
3.2 Power-Centered Site Selection 15
3.3 Project Economics 15
3.4 Local Engagement 16
3.4.1 Early Engagement 16

3.4.2 What Sustained Presence Delivers 16
3.5 Energy Efficiency 16
3.5.1 Cooling System Selection 16
3.5.3 Power Usage Effectiveness 16
3.5.4 Design Trade-offs 17
4. Recommendations 17
4.1 For Developers and Operators 17
4.2 For Utilities 17
4.3 Local Engagement 18
5. The Path Forward 19

Executive Summary

The physical infrastructure required to power artificial intelligence is a significant real estate, energy, and industry challenge. This whitepaper examines what it takes to build AI-era data center infrastructure, and what constraints stand in the way.

The challenge has two distinct dimensions. The first is powering the buildout: can you get the electricity you need, when you need it? The second is delivering the buildout: can you actually site, permit, and construct the facility? These are related but distinct challenges, with different stakeholders, constraints, and solutions.

The analysis draws on perspectives from across the data center ecosystem: utility executives managing grid constraints, developers delivering capacity, construction professionals navigating schedule risks, regulators balancing competing priorities, and researchers developing new solutions. Key findings include:

On powering the buildout:

- The real bottleneck is power deliverability, not generation. Many regions have energy resources. What they lack is the transmission and distribution infrastructure to move power to specific sites on developer timelines.
- A fundamental timeline mismatch drives the tension. Data center developers typically need power commitments within 6 to 12 months. New distribution sub-stations can take 2 to 3 years. New transmission corridors take 10 to 15 years to permit and build.
- Significant capacity can be unlocked without new infrastructure. Demand flexibility, grid-interactive design, and efficiency improvements represent underutilized tools.

On delivering the buildout:

- Site selection has fundamentally changed. Power availability and timing now matter more than land cost.
- Local resident trust is earned through early, transparent engagement. The projects that succeed treat local officials and local residents as partners, not obstacles.
- Demand will not slow down. Efficiency improvements enable more use, not less.

> *The AI buildout is not a future challenge; it is happening now. The infrastructure decisions made in the next two to three years will shape regional capacity, reliability, and economic competitiveness for decades. No single stakeholder can solve this alone. The path forward requires partnership.*

Acknowledgments

This whitepaper was developed by from proceedings of the Urban Land Institute (ULI) Northwest & University of Washington Foster School of Business Symposium, "The AI Buildout: Data Centers, Energy, and the Business of Scale," held on April 29, 2026 at the University of Washington. It synthesizes lessons learned and practical insights from a single symposium and is not intended as a comprehensive literature review or exhaustive analysis of the data center buildout. The examples and recommendations reflect specific projects and partnerships discussed by panelists and may not apply to all markets, regulatory environments, or individual projects.

The following individuals are acknowledged for their participation in the symposium, not as authors or endorsers of this whitepaper. The synthesis, conclusions, and recommendations are those of the whitepaper's authors and should not be attributed to individual participants or their organizations unless expressly identified:

- Isaac Barrow, Portland General Electric (Moderator)
- Tracy Seslen, University of Washington Foster School of Business (Moderator)
- Bing Liu, Pacific Northwest National Laboratory
- Pouyan Farzanehkari, HDR
- Tim Mirick, Sabey Data Centers
- Leonard Ruff, AIA, Cascade Mission Critical, LLC

Event Organizers

- Tracy Seslen, University of Washington Foster School of Business
- Andrea Newton, Urban Land Institute Northwest

Whitepaper Lead Author

- Anna Brannon, Pacific Northwest National Laboratory

AI Usage Acknowledgment

This paper was developed with the assistance of artificial intelligence tools. AI was used to transcribe and translate panel recordings and to identify and organize recurring themes across participant responses. All structural decisions, interpretive analysis, and conclusions presented in this paper reflect the judgment of the authors.

1. Introduction

Artificial intelligence does not exist only in software. It lives in buildings, in land use decisions, in power systems, in transmission constraints, in construction timelines, and in local trade-offs. The servers that train large language models and run inference workloads require physical space, reliable electricity, and sophisticated cooling, at a scale that is straining existing infrastructure across the country.

In this moment, megawatts matter as much as square footage. The analysis draws on perspectives from across the data center ecosystem: utility executives managing grid constraints, developers delivering capacity, construction professionals navigating schedule risks, regulators balancing competing priorities, and researchers developing new solutions.

1.1 Two Dimensions of the Challenge

Getting a data center built involves solving two related but distinct problems:

- Powering the buildout is the energy challenge. Can enough electricity be secured, at the right voltage and reliability level, on a timeline that matches the project? This involves utilities, grid infrastructure, generation resources, and increasingly creative approaches to flexibility and storage.
- Delivering the buildout is the development challenge. Can a project find a viable site, navigate permitting, manage construction, and maintain local support? This involves developers, construction teams, local officials, policymakers, and other local stakeholders.

A project can fail on either dimension. A site with willing local support and streamlined permitting isn't viable if power can't be delivered for five years. A site with available power isn't viable if permitting stalls or local opposition blocks construction.

1.2 Key Decision Points for Success

Two categories of decisions made at project inception most significantly affect outcomes two to three years later. The sections that follow explain why these decisions matter, what the underlying constraints are, and what approaches are working today.

Design choices:

- On the power side, facilities designed for grid interactivity from day one, with flexible load architectures, telemetry, alignment on guidelines, and microgrid readiness, will have options that retrofitted facilities will not.

- On the delivery side, decisions about site selection, engagement strategy, and procurement timing largely determine whether a project stays on track.

Partnership and engagement timing:

The projects that succeed engage early: utilities before land deals close, communities before permits are filed, procurement before design is finalized. The projects that struggle treat these as sequential steps rather than parallel workstreams.

1.3 Where Projects Fail

Projects fail on both dimensions of the challenge.

- On the power side, engineering assumptions tend to break at the substation and distribution level. Project teams assume substation upgrades will be minor, that distribution circuits can handle additional load, or that cooling profiles will not affect peak demand. When those assumptions prove wrong, the consequences are significant: 6 to 24 months of delay, stranded capital, redesign costs, and in some cases, total site abandonment.
- On the delivery side, projects fail when local engagement starts too late, when permitting timelines are underestimated, or when procurement is not aligned with the construction schedule.

In both cases, a common pattern emerges from speed-to-market pressures. When developers prioritize moving fast, efficiency, local engagement, and early utility coordination often take a back seat. This creates an ironic outcome: the project that cut corners to move faster often ends up slower.

1.4 Why Efficiency Is Not the Sole Solution

Efficiency improvements are essential, but they won't solve the capacity problem. This is Jevons' paradox applied to computing: each generation of more efficient processors enables new applications that consume the gains and then some. More efficient chips don't reduce total electricity demand; they enable new uses that expand it. When computing becomes cheaper, organizations don't return the savings—they do more.

Efficiency matters for managing the intensity of growth. It won't slow the buildout itself.

1.5 Why Multiple Perspectives Matter

The data center challenge looks different depending on where you sit.

- Utilities must balance their obligation to serve new load against their responsibility to existing ratepayers.

- Developers and operators are under pressure to deliver capacity quickly. Their customers have aggressive timelines and will take their projects elsewhere if a region can't deliver.
- Construction professionals see where projects actually fail: in the gap between ambitious schedules and the reality of permitting, procurement, and coordination.
- Engineers and designers translate project specifications into buildable systems. They see where assumptions break down.
- Researchers are developing technologies that could change the equation while facing the challenge of moving innovations from laboratory to deployment.

This whitepaper draws on all of these perspectives.

2. Part 1: Powering the Buildout

2.1 The Real Constraint: Deliverability, Not Generation

A consistent theme across practitioner perspectives: many regions have energy resources, but getting power to specific sites on project timelines is the binding constraint.

This distinction between having power somewhere on the system and being able to deliver it to a specific location is the source of more confusion, frustration, and project failures than any other factor.

2.1.1 Categories of Constraints

Constraints operate at four levels, each with different characteristics and solutions:

- Physical constraints are the most tangible. Can the existing transmission lines carry additional load? Is there capacity at the nearest substation? Can the local distribution infrastructure handle the demand?
- Transmission constraints are particularly acute. The high-voltage lines that move power across long distances were built decades ago, sized for historical demand patterns.
- Power quality constraints become more significant as AI workloads grow. AI training workloads behave differently from traditional data center loads, with rapid ramps that create challenges for grid operators.
- Contractual and legal constraints are less visible yet equally binding. When a utility commits capacity to a customer, that commitment has legal force.

This creates a planning challenge. Utilities must reserve capacity for committed loads, even if those loads haven't materialized yet. To address this, most utilities will implement a contractual load ramp commitment so that the costs for the utility infrastructure improvements are covered in the energy costs over the term of the contract.

2.1.2 What "Power Available" Actually Means

Much of the confusion in data center siting stems from conflating distinct concepts:

- Generation means energy exists somewhere on the system. A region may have substantial generation capacity, but the best generation sources are often far from where the power is needed.
- Transmission is the high-voltage infrastructure that moves power across long distances.
- Distribution is the local infrastructure—substations, circuits, transformers—that delivers power the last few miles to end users.

- Deliverability is the combination of all three: having generation, transmission, and distribution capacity available, at the required voltage and reliability level, on a timeline that matches the project.

Developers sometimes hear that a region "has plenty of power" and assume their project is feasible. The constraint is almost always in the distribution and deliverability, not in regional generation capacity.

2.1.3 The Timeline Mismatch

The mathematics of infrastructure development create a fundamental tension with the pace of data center demand:

- Data center construction, once a site is secured and permits obtained, typically takes 12 to 24 months.
- A new electrical substation can be built in approximately 2 to 3 years, assuming the electrical capacity is available, though transformer procurement is increasingly extending that timeline.
- New generation capacity—wind farms, solar installations, battery storage—can typically be developed in about 3 years.
- New transmission capacity is on a much longer timeline. Permitting and constructing a major transmission corridor takes 10 to 15 years under best-case conditions.

The implication is stark: if a data center project requires new transmission capacity to be feasible, that project cannot proceed on any timeline that matches business reality.

The constraint, paradoxically, is also creating opportunity. As one utility executive observed: "Constraint leads to innovation." The combination of urgent demand and well-funded customers is accelerating technology adoption and partnership models that might otherwise have taken a decade or more to develop.

2.2 Grid-Interactive Design

The conventional framing positions data centers as grid burdens. An alternative framing is emerging: data centers as grid partners that provide flexibility, fast response, and even ancillary services.

2.2.1 Demand Flexibility

Demand flexibility—the ability to reduce or shift load in response to grid conditions—allows data centers to receive faster capacity access in exchange for commitments to reduce load during peak periods.

- Workload flexibility includes shifting non-critical compute jobs to off-peak hours, distributing workloads geographically across data center networks, and applying power capping during peak periods. A recent research partnership between PNNL, NVIDIA, and Emerald AI demonstrated 30 percent peak capacity reduction through workload management techniques.
- Thermal flexibility includes pre-cooling facilities before peak periods, using thermal storage to shift cooling loads, and adjusting temperature setpoints dynamically based on grid conditions.

The gap is not technical potential. The gap is the lack of incentives, limited orchestration tooling, and the need for architectures that safely expose flexible load without risking uptime.

2.2.2 What Grid-Interactive Architecture Requires

Designing for grid interactivity means embedding specific capabilities from day one:

- Telemetry and alignment on guidelines. Real-time communication with grid operators uses consistent protocols. IEEE 2030.5, IEEE 1547, and OpenADR provide frameworks for grid-interactive equipment.
- Orchestration-ready architecture. The ability to modulate load in response to grid signals requires control systems that can receive external inputs and translate them into operational adjustments.
- Microgrid readiness. Facilities designed to operate in island mode during grid outages, or to provide restart support after outages, offer value beyond their own operations.

2.3 Solutions that Move Faster than Transmission

Given that major transmission upgrades take 10 to 15 years, solutions that can be deployed faster are essential for near-term projects.

2.3.1 Behind-the-Meter Battery Storage

Battery energy storage systems are real tools today. In the near term (1 to 3 years), they can meaningfully reduce peak load, support ride-through during brief outages, and enable limited islanding.

Aligned Data Centers, working with battery developer Calibrant, constructed a 30-megawatt, 60-megawatt-hour battery energy storage system adjacent to their data center in PGE's service territory. The battery was funded entirely by the private parties, with zero investment from the utility or ratepayers.

The structure worked because incentives aligned. The data center got accelerated access to capacity. The utility got a dispatchable resource at no capital cost. Ratepayers got reliability benefits.

2.3.2 Phased Power

Phased power is often the fastest practical lever for getting projects moving. This approach allows projects to proceed with 10 to 40 percent of full capacity while waiting for broader infrastructure upgrades.

2.3.3 Grid-Side Technologies

On the utility side of the meter, several technologies extract more capacity from existing transmission and distribution infrastructure:

- Dynamic line ratings use sensors on transmission lines to measure real-time conditions. Traditional ratings assume worst-case conditions; actual capacity on most days is higher.
- Targeted infrastructure modifications, such as coatings that improve heat dissipation or tower height increases on specific segments, can add meaningful capacity faster and cheaper than new construction.
- Advanced capacity modeling uses hour-by-hour Monte Carlo simulation to identify capacity that conservative planning approaches miss. Portland General Electric's partnership with GridCare identified approximately 84 megawatts of additional load capacity on an existing system with no new infrastructure investment.

3. Part 2: Delivering the Buildout

3.1 The First 90 Days

From the construction perspective, project outcomes are largely determined in the first 90 days, long before physical construction begins. Teams that succeed focus early and intensively on three areas:

- Power. Is the utility's timeline for providing service aligned with the project schedule? Have interconnection requirements been clearly defined?
- Permitting. Does the local jurisdiction have relevant codes and experience with data center projects? What's the realistic approval timeline?
- Procurement. Long-lead equipment (transformers, switchgear, generators, cooling systems) must be ordered early. Lead times that were once 6 months may now be 18 months or longer.

The common thread is early alignment. Stress-testing schedules across power, permitting, and procurement—with input from all stakeholders—before committing to timelines distinguishes projects that deliver from those that slip.

3.2 Power-Centered Site Selection

Site selection for data centers has fundamentally changed. Power availability and timing now matter more than land cost. Teams now engage utilities early, often before land deals are finalized.

Three dimensions of power now drive site evaluation:

- Speed: How fast can power be delivered? A site with power available in 12 months is worth far more than one requiring 36 months of utility work.
- Capacity: How much power is available, and what's the pathway to expansion?
- Cost: What does power cost, and what infrastructure investments will the project need to fund?

3.3 Project Economics

Data center economics are driven by power capacity, not building footprint. The industry has largely stopped measuring facilities in square feet; megawatts are the relevant unit.

Land costs have escalated dramatically in areas with available power. Parcels that would have been relatively inexpensive a decade ago now command significant premiums where power access is confirmed.

3.4 Local Engagement

3.4.1 Early Engagement

Data centers consume substantial power, operate continuously, and have characteristics that typical building codes don't fully address. Local residents unfamiliar with data centers often lack frameworks for evaluating proposals.

Experienced developers address this gap directly. When Microsoft built its first data center in Quincy, Washington—then an agricultural area with no data center experience—the company brought local code officials to tour existing facilities before submitting permit applications.

3.4.2 What Sustained Presence Delivers

Quincy, Washington, is a prime case study of how data centers, done right, can be a net local benefit. After more than a decade of data center presence, property taxes for residents have decreased. As a result of this tax revenue, the city has a new fire station, upgraded police facilities, and a renovated high school with resources that exceed what many larger communities provide. A partnership between the data center operator and the City of Quincy resulted in construction of the Quincy Water Reuse Utility, which treats water from the datacenter and recirculates it for re-use.

3.5 Energy Efficiency

3.5.1 Cooling System Selection

Cooling a data center involves two functions: removing heat from the computing equipment and rejecting that heat outside the building. While historically evaporative cooling was commonly used in data centers due to its high energy efficiency, other technologies have evolved that can be just as efficient but utilize little or no water. These options include air-cooled chillers, water-cooled chillers coupled with dry coolers, or hybrid systems.

Efficient cooling systems also bring the cooling medium closer to the chips. While traditional air-based cooling systems coupled with hot and cold aisle containment are still used, the hyperscale data center industry is moving towards close-coupled cooling systems and liquid cooling which delivers cooling directly to chips in a closed-loop system.

3.5.3 Power Usage Effectiveness

Power Usage Effectiveness (PUE) measures total facility power divided by IT equipment power. A PUE of 2.0 means one watt of overhead for every watt delivered to computing equipment.

During the dot-com era, PUE commonly exceeded 2.0 due to reliance on energy-intensive compressor-based cooling. The industry benchmark is now around 1.5, with leading facilities achieving 1.3 or lower. Much of this improvement came from shifting to evaporative cooling, which reduced energy consumption but increased water use. Today, facilities can achieve similar efficiency with air-cooled and closed-loop systems that consume little or no water.

3.5.4 Design Trade-offs

There is no universally optimal cooling approach. The choice depends on local climate, water availability, energy costs, and rack density. A facility in a temperate climate with moderate density may rely entirely on air-based systems. A high-density AI facility in a hot climate may require liquid cooling to racks with hybrid heat rejection. Early engagement with local stakeholders helps identify which constraints matter most for a given site.

4. Recommendations

4.1 For Developers and Operators

Design for grid interactivity from day one. The facilities being built today will operate for decades. Embedding flexibility capabilities into initial designs is far cheaper than retrofitting later.

Engage utilities before finalizing site selection. Power availability, interconnection timelines, and the scope of infrastructure upgrades vary dramatically across locations.

Treat local engagement as a core competency. The developers with the strongest track records invest in relationships before they need them.

Plan for flexibility, not just capacity. The ability to reduce load during peak periods can be the difference between a site that's feasible and one that isn't.

4.2 For Utilities

Embrace advanced modeling and data sharing. Sophisticated analysis can identify capacity that conservative planning approaches miss.

Develop clear, predictable interconnection processes. Uncertainty about timelines, costs, and what infrastructure upgrades a project will trigger is one of the largest barriers developers face.

Explore creative partnership structures. Behind-the-meter storage, efficiency investments, and flexibility commitments can all expand effective capacity without ratepayer-funded capital investment.

Differentiate between serious developers and speculators. Developing mechanisms to identify and prioritize ready-to-proceed projects could improve outcomes for utilities and serious developers alike.

4.3 Local Engagement

Local residents have legitimate interests in data center development: water use, noise, visual impact, rate effects, and local economic benefit. Thoughtful and early engagement can help address these concerns.

Effective engagement starts with transparency about what's being built and why. It includes meaningful commitments on mitigation, clear information about economic impacts, and responsiveness when concerns are raised.

Long-term commitments matter. Load ramp agreements, benefit packages, and sustained presence distinguish serious operators from speculators.

5. The Path Forward

The examples throughout this paper share a common thread. The Quincy benefits came from sustained presence and contractual commitments that gave utilities confidence to invest. The GridCare partnership unlocked 84 megawatts of capacity through sophisticated modeling paired with flexibility commitments. The Aligned/Calibrant battery project accelerated capacity access through a structure where incentives aligned: faster power for the developer, a dispatchable resource for the utility, reliability benefits for ratepayers.

None of these happened by accident. Someone took the risk of going first on transparency, and others responded in kind. These weren't cases where interests magically aligned; they were cases where parties chose to find alignment.

The AI buildout will proceed. The infrastructure decisions made in the next two to three years will shape regional capacity, reliability, and economic trajectory for decades. The question is whether those decisions emerge from partnership—with benefits shared, risks managed, and communities engaged—or from conflict, with each party pursuing narrow interests.