WHITE PAPER 3

From Flexible Load to Grid Resource

A Qualification Framework for Large Flexible Energy Consumers

With data centres as the leading case

Svendsen* · Nyiredy · Milenkovic****

Version: Release candidate 1.0 · October 2026

Contact: arne@thefuturepowergrid.org · robert@thefuturepowergrid.org · mathieu@thefuturepowergrid.org · info@thefuturepowergrid.org

Published by The Future Power Grid. Authors write in an independent capacity unless otherwise stated.

Suggested citation: Svendsen, A. B., Nyiredy, R., & Milenkovic, M. (2026). From Flexible Load to Grid Resource: A Qualification Framework for Large Flexible Energy Consumers. White Paper 3. The Future Power Grid. Draft 1.0 (pre-publication).

"Virtual N-1 asked the grid to accept a new kind of tenant. This paper asks what that tenant must install, how its response can be qualified, and under what conditions an orderly, contracted departure can preserve security for everyone else."

This paper builds on: Svendsen, Nyiredy & Milenkovic (2026). Unlocking Grid Capacity Through Risk-Based Operation. White Paper 1; and From N-1 to Virtual N-1: Extending the Risk-Based Framework for Grid Capacity. White Paper 2.

* *Lead author and corresponding author*

** *Contributing authors*

Table of Contents

Glossary of Acronyms 3
Source Attribution 4
Executive Summary 5
1. Context: From System Side to Load Side 7
2. What Makes Data Centres Different 9
3. A Typology of Large Energy Consumers 11
4. Anatomy of Flexibility — A Taxonomy 12
5. The Qualification Framework 15
6. Reserve and Storage as Active Resources 20
7. Mapping Flexibility to Power System Needs 21
8. Bounding the Potential: An Illustrative System-Level Translation 23
9. Regulatory and Commercial Framework 25
10. Validation Boundaries and Open Questions 27
11. Conclusions and Specific Recommendations 28
References 30
Appendix A — Research Questions for WP3 30
Appendix B — Worked Example: Qualifying a 300 MW AI Campus 32

Glossary of Acronyms

Acronym	Definition
ACER	Agency for the Cooperation of Energy Regulators
aFRR	Automatic Frequency Restoration Reserve
AI	Artificial Intelligence
BESS	Battery Energy Storage System
CENS	Cost of Energy Not Supplied
BYONC	Bring Your Own New Capacity (PJM)
CRU	Commission for Regulation of Utilities (Ireland)
CSAM	Coordinated Security Assessment Methodology (ENTSO-E)
DSA	Dynamic Security Assessment
EENS	Expected Energy Not Supplied
EPRI	Electric Power Research Institute
ERCOT	Electric Reliability Council of Texas
ENTSO-E	European Network of Transmission System Operators for Electricity
FCA	Flexible Connection Agreement
FCR	Frequency Containment Reserve
GET	Grid Enhancing Technologies
GPU	Graphics Processing Unit
HVAC	Heating, Ventilation and Air Conditioning
IEA	International Energy Agency
IRAS	Interim Resource Adequacy Service (PJM)
LOLP	Loss of Load Probability
mFRR	Manual Frequency Restoration Reserve
N-1	Single-contingency operational criterion
N-X	Multi-contingency state (X elements out of service)
NESO	National Energy System Operator (Great Britain)
PJM	PJM Interconnection (US regional transmission organisation)
PRA	Probabilistic Risk Assessment
PUCT	Public Utility Commission of Texas
PUE	Power Usage Effectiveness
SOGL	System Operation Guideline (EU network code)
SPP	Southwest Power Pool (US)
SQSS	Security and Quality of Supply Standard (Great Britain)
TSO	Transmission System Operator
UPS	Uninterruptible Power Supply

Source Attribution

Throughout this paper, a distinction is made between:

- **"IEA estimates"** referring to quantitative values and findings directly reported in IEA (2026a) and IEA (2026b).
- **"Authors' interpretation"** referring to system-level and load-level interpretations derived from these findings.
- **"Operational experience"** referring to results observed in practical implementations.
- **"Proposed framework"** referring to the qualification criteria, tier boundaries, firmness classes and certification arrangements introduced by the authors for consultation and validation. These are not existing technical or regulatory standards.
- **"Industry signal"** referring to public statements by technology and data centre industry leaders. Industry signals are treated as evidence of direction and intent, not as technical validation of the framework or of any capacity figure in this paper.

Unless otherwise stated, percentage-based and quantitative flexibility estimates represent authors' interpretations rather than directly reported IEA results. Capacity and flexibility figures are indicative and require validation through system- and site-specific analysis before application to planning or contractual decisions.

No capacity range proposed in this paper should be interpreted as an IEA estimate unless explicitly identified as such.

This paper is the third in a series. White Paper 1 presented a three-layer framework for PRA, DSA and Grid Enhancing Technologies (indicative +15–30% utilisable capacity); White Paper 2 introduced Virtual N-1 (a further indicative +10–20%). This paper turns to the load side: how a large consumer — with the data centre as the leading case — can be qualified as a dependable resource within that principle.

Executive Summary

White Papers 1 and 2 developed the system-side case for risk-based operation and conditional connection. This paper addresses the unresolved load-side question: **when may a system operator treat the response of a large flexible consumer as a dependable part of system security?**

The answer proposed here is qualification. A flexible capability becomes a grid resource only when its depth, speed and duration are specified; its trigger is compatible with the relevant security mechanism; its response can be verified; and its availability, recovery behaviour and failure modes are explicitly represented in security assessment.

This paper proposes a general qualification architecture for converting flexible load from physical capability into reliability-accounted grid resource, using data centres as the leading case. Figure 4 summarises the proposed qualification chain from physical capability to credited conditional capacity.

The individual building blocks are not new. Demand response, UPS and battery systems, backup generation, remedial action schemes and probabilistic risk assessment already exist and are widely used across the power sector. The novelty is found in their integration into a common qualification chain that transforms flexible load from a passive consumer of electricity into a verifiable, contractable and reliability-accounted grid resource. Qualified flexibility can then be represented within probabilistic risk assessment and used explicitly to support operational security and connection-capacity decisions.

White Paper 1 framed the IEA's estimate of 1,200–1,600 GW of capacity unlockable through operational measures (IEA, 2026a) as a three-layer framework — Decision Layer (PRA/DSA), Physical Layer (GET) and Security Layer (flexibility and markets) — with an indicative +15–30% in utilisable capacity. White Paper 2 added Virtual N-1, a further indicative +10–20%. The IEA has since published a closely related three-layer view of grid optimisation, citing this series among its sources (IEA, 2026b, p. 16); two of the authors were external peer reviewers of that report, so the citation indicates alignment, not validation.

Both earlier papers take the system operator's view. But a conditional-connection contract is only as good as the load's ability to honour it — at the required depth, speed and duration, verifiably, and without harming the process it serves.

The question is pressing. At least 150 GW of data centre projects are in advanced stages of connection queues, and one-fifth of the global build-out is at risk of delay from grid congestion (IEA, 2026a). The European Commission has asked regulators to create frameworks for flexible connection agreements, naming data centres as good candidates (European Commission, 2026), and the compute industry itself now describes data centres that reduce their draw when the grid is constrained (Lex Fridman Podcast, 2026). What is missing is the specification that connects the two sides.

This paper develops that missing half, taking the data centre as the leading case. The argument proceeds in five steps:

- **Data centres are not simply large loads.** Their electrical architecture already contains the building blocks of flexibility — uninterruptible power supplies, batteries and backup generation — supplemented, where the owner also runs the workload, by controllable IT load. The flexibility is latent, not absent.
- **That flexibility can be mapped and quantified.** A load decomposes into critical, shiftable, deferrable and cooling components, and responds through four mechanisms — electrical substitution, computational curtailment, temporal and spatial flexibility — each with a characteristic depth, speed and duration. For most data centres the most dependable response is electrical substitution — online UPS bridging to on-site generation — which reduces grid import without touching the computing service.
- **Qualification, not aspiration, is the missing framework.** To count as a Virtual N-1 resource, a load must demonstrate five criteria — committed depth, speed, duration, trigger compatibility and verifiability — under explicit constraints for availability, recovery and failure behaviour.
- **Qualification makes differentiated firmness contractible.** Once the response is certified, firmness can be offered in classes, allowing earlier connection in exchange for a defined, bounded and compensated obligation to respond.
- **Qualified flexibility can be represented in system risk.** Qualification does not make the response certain; it makes its uncertainty measurable, so that PRA can determine how much conditional capacity keeps risk to

priority load within an agreed threshold. Illustrative translations suggest that qualified data-centre flexibility could make material contribution on data-centre-heavy corridors; Section 8 develops an indicative range of roughly 4-13% of utilisable capacity. This is a share of, not an addition to, the Virtual N-1 potential of White Paper 2.

Read together, the series is not a sequence of claims for "more capacity". It is a single operational logic, of which data centres are the first and clearest application rather than the main story (Figure 1).

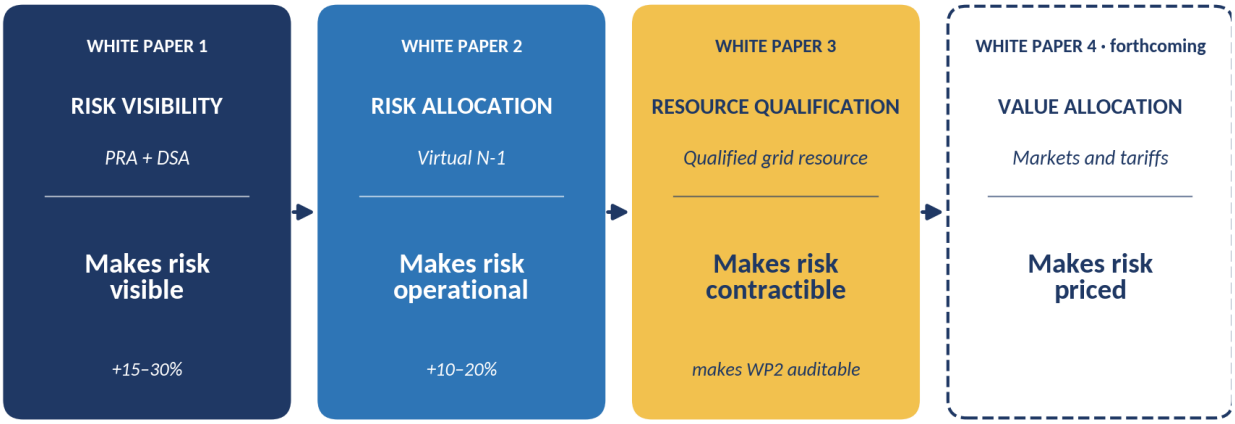


Figure 1 *The logic of the series. WP1 makes risk visible; WP2 makes it operational through differentiated firmness; WP3 makes it contractible by qualifying the load-side response; a forthcoming WP4 addresses how the qualified resource is valued and priced. Data centres are the leading case, not the subject.*

> ***PRA makes risk visible. Virtual N-1 makes risk operational. Qualification makes risk contractible.***

Many of the enabling assets are already installed in modern data centres or increasingly required, and relevant demonstrations exist (Colangelo et al., 2025). What is missing is the qualification, control, telemetry and contractual architecture that lets the grid rely on them — summarised on one page in Figure 4. That work is definitional, regulatory and commercial, and it needs to start now: the data centres that could qualify are already in the queue.

> ***Capability is not a grid resource until it is specified, qualified, triggered and verified. A qualified flexible load can support additional connection without its response being treated as an assumption: the response becomes a defined, testable and reliability-accounted element of system security.***

1. Context: From System Side to Load Side

Grids, not generation, are now the binding constraint on the energy transition (IEA, 2026a). The mismatch is one of time as much as capacity: new grid takes 5 to 15 years to plan, permit and build, while a data centre can be built in 1 to 3 (IEA, 2026a).

White Paper 1 responded with a three-layer framework for risk-based operation. White Paper 2 extended it with Virtual N-1 — Stage 3 of the implementation pathway — a conditional-connection principle under which a defined category of load provides an automatic, pre-specified response to a contingency or a quantified risk threshold, such that the resulting system state satisfies the required security criterion for all priority load (formal definition: Svendsen, Nyiredy & Milenkovic, 2026b, Section 5.2).

That principle rests on an assumption it does not examine: that the conditionally connected load can, in fact, deliver. No system operator would rely on a protection relay whose operating characteristic was merely asserted by its owner; the same discipline must apply to a load that has contracted to be the release valve of the grid.

1.1 The Load-Side Gap

White Paper 2 listed candidate load categories and cited demonstrated results such as TenneT's 9.1 GW of non-firm contracts (Ember, 2026; Ei, 2023). It did not answer the operator's and regulator's next question:

> ***Given a specific large load seeking connection, how do we determine — in advance, and auditably — how much flexibility it can commit, how fast, for how long, under which trigger, and with what confidence that the commitment will hold?***

The same gap is visible in the IEA's latest work. Its report on grid modernisation places flexible and non-firm connections in the top layer of grid optimisation and notes that flexibility can be delivered through pre-agreed operating protocols, but leaves the underlying contractual and regulatory arrangements for further work (IEA, 2026b). That is the territory of this paper: the specification of the load-side response on which such a protocol can rely.

Answering it requires understanding the load's internal electrical architecture, decomposing its demand, and defining the criteria for admission to conditional connection.

1.2 The Policy Moment

Regulators and system operators are already moving towards conditional access for large loads, but without a common definition of what a qualifying load must deliver. Table 1 summarises the most relevant recent signals.

Source	Finding or measure	Relevance to WP3
IEA Electricity 2026	1,200–1,600 GW unlockable, 750–900 GW via non-firm connections; ≥150 GW of data centres in advanced queues.	Scale of the opportunity; non-firm access is the largest single lever
IEA Modernising Grids (2026)	Three layers of grid optimisation; flexibility layer includes non-firm connections and pre-agreed protocols.	WP3 specifies the load side of the flexibility layer
European Commission, COM(2026) 501	Frameworks for flexible connection agreements; data centres good candidates; tripartite agreement and rating scheme.	A qualification standard is the natural technical content of both instruments
UK Independent Review (2026)	Probabilistic operation by 2035; risk metrics and thresholds by 2030; warns that flexibility baselines can be gamed.	Risk-threshold triggers; verification must not rely on gameable baselines
Ireland (CRU, 2025)	New data centres must provide dispatchable generation and/or storage matching the site's maximum import capacity.	On-site assets mandated — but not yet qualified as grid resources

Source	Finding or measure	Relevance to WP3
United States (SPP, PJM, ERCOT)	Fast-track for curtailable large loads (SPP); curtail-first unless new capacity is brought (PJM, pending); disconnection capability ≥75 MW (Texas SB 6).	Conditional access emerging without a common qualification standard

Table 1 *Recent policy and regulatory signals on large-load connection and flexibility (IEA, 2026a; IEA, 2026b; European Commission, 2026; Yu, 2026).*

Across the jurisdictions reviewed here, a common direction is visible: conditional or flexible access for large loads is moving from exception towards an explicit connection option, and several now require the very on-site assets — storage and dispatchable generation — on which flexibility depends. The sources reviewed here do not identify a common framework specifying how such responses should be qualified, verified and counted as a system resource.

1.3 An Industry Signal: Compute Is Converging on the Same Problem

Industry leaders now describe the same capability from the computing side. In a 2026 discussion, NVIDIA CEO Jensen Huang argued that data centres should "gracefully degrade" when the grid asks them to — through backup generation, workload relocation and reduced computing rates — and that utilities could offer several levels of delivery guarantee, so that customers accepting a more conditional one could connect sooner (Lex Fridman Podcast, 2026). On how hard this would be, he answered: "As soon as you could specify, you could engineer it." That places the burden where this paper places it — on the specification.

This is treated strictly as an industry signal (see Source Attribution), not as evidence of power-system reliability. It does not validate the capacity ranges in this series, and NVIDIA has not endorsed Virtual N-1 or this framework (Section 10.2).

1.4 Why the Data Centre Is the Leading Case

Any sufficiently large and controllable load can participate in Virtual N-1; the data centre is the leading case for four reasons. **Scale:** data centres are expected to account for about half of US electricity demand growth to 2030 and dominate demand queues (IEA, 2026a; Ember, 2025); EU capacity is expected to grow from about 12 GW in 2025 to 28 GW by 2030 (European Commission, 2026). **Architecture:** redundant power paths, storage and backup generation — the prerequisites of fast, deep, bounded response — are already present. **Incentive:** the value of faster grid access justifies conditional terms. **Controllability:** UPS and generation can take a large share of the site off the grid within milliseconds to minutes, without touching the computing service.

The framework is written to apply to any large consumer — electrolysers, smelters and other flexible industry share many of the same characteristics — but the data centre is where the case is clearest and most urgent.

1.5 The Series as an Operational Progression

The papers are best read as an operational progression (Figure 1): White Paper 1 makes risk visible, White Paper 2 makes it operational by differentiating firmness, and this paper makes it contractible by qualifying the load-side response. A fourth paper will address how the qualified resource is valued.

The data centre is therefore the leading case, not the subject. The same sequence — visible, operational, contractible — applies to electrolysers, smelters, storage and any other large load that can respond on demand.

2. What Makes Data Centres Different

From a power-system perspective, a data centre is unusual because it is engineered never to lose supply to its critical load. That engineering is what creates flexibility.

2.1 An Architecture Built Around Ride-Through

A large data centre takes grid supply at transmission or sub-transmission voltage (typically 132–220 kV for the largest sites), steps it down through on-site transformers, and distributes it through redundant power paths to the IT load. Between the grid and the servers sit three elements that conventional load does not have:

- **Uninterruptible power supply (UPS).** Batteries (increasingly lithium-ion rather than lead-acid) bridge the gap between a grid disturbance and the response of backup generation, holding critical load for seconds to minutes with no interruption to service. An online UPS responds in effectively 0 ms; a line-interactive UPS in under about 4 ms.
- **Backup generation.** Diesel or gas generators, sized to carry the full critical load, start within seconds. Under Uptime-type certification they are rated for unlimited running hours, and operation on generation is a normal operating mode rather than a degraded one; in practice, duration is limited by fuel logistics and by environmental permits for non-emergency operation.
- **Controllable IT load.** Where the data-centre owner also operates the workload, schedulers and power-capping allow part of the demand to be reduced or shifted on command. In colocation and GPU-cloud (neocloud) facilities the owner of the site and the operators of the workload are different parties with different motivations, and the signal path from software to hardware to infrastructure is long. At high power density, cooling has little thermal inertia and largely follows the IT load.

Each of these exists to protect the facility from the grid. The insight behind Virtual N-1 — and the reason the data centre is the leading case — is that the same equipment can be operated in the other direction: to protect the grid from the facility's demand at moments of system stress.

2.2 From Passive Protection to Active Resource

In the conventional model, UPS, batteries and generators are passive: they defend only the site's own supply. Reframed as a grid resource, they become active. A UPS that can carry critical load for two minutes gives the grid a two-minute window in which the facility can be disconnected with no interruption to computing; backup generation can, under contract, let the site island and relieve the grid.

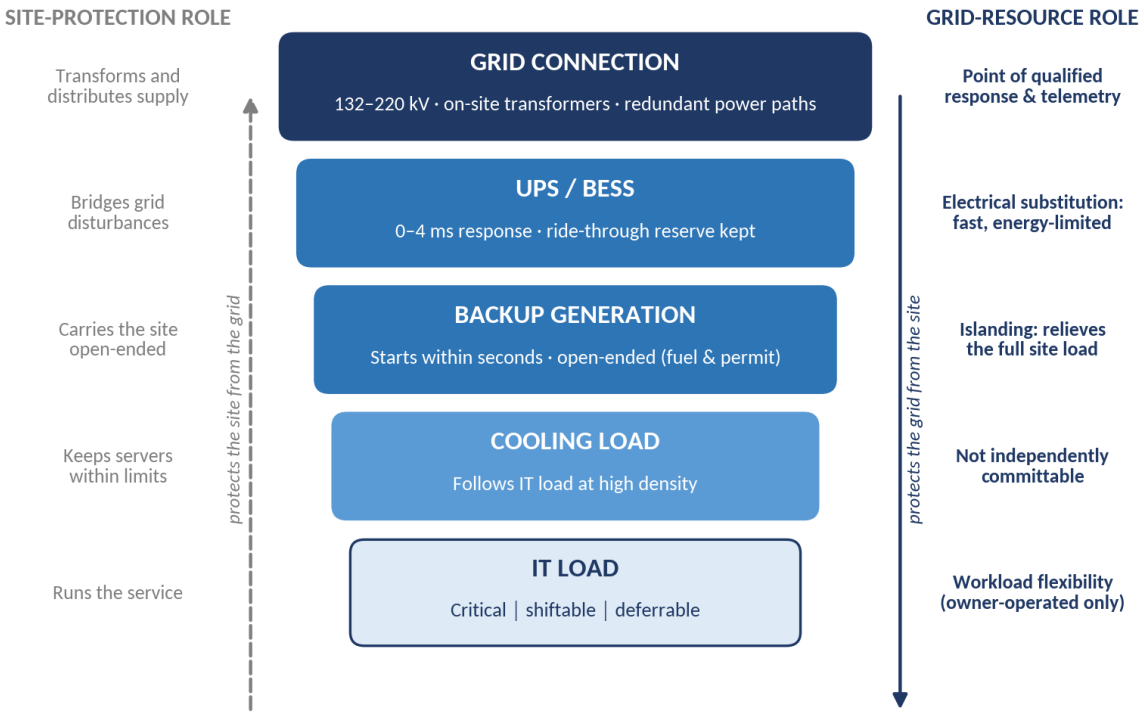


Figure 2 *The data centre as a layered load. The same architecture that makes a data centre a demanding consumer — redundant power, on-site storage, backup generation, controllable cooling — is what qualifies it as a flexibility resource.*

Regulation is already moving in this direction. In Ireland, new data centres must provide dispatchable generation and/or storage matching the site's maximum import capacity (IEA, 2026a). The asset base is therefore being mandated; what is not yet defined is how its response is to be qualified and relied upon.

This reframing is the load-side counterpart to the system-side reframing in White Paper 2. There, the grid learned to treat a class of load as interruptible by design. Here, the load learns to treat its own protective infrastructure as the mechanism that makes interruption harmless.

> ***Key insight: the equipment that makes a data centre a demanding load is the same equipment that qualifies it as a flexible one. Ride-through capability and curtailment capability are two uses of one architecture.***

3. A Typology of Large Energy Consumers

Flexibility potential depends on the type of facility, because size and function determine both the electrical architecture and the commercial logic of conditional connection. The categories below are indicative.

Category	Typical size	Characteristics relevant to flexibility
Enterprise	1–20 MW	Single-business, on-premises. Often highly redundant but individually small; flexibility value lies almost entirely in aggregation.
Colocation	20–100 MW	Multi-tenant. The operator controls the infrastructure — UPS, storage, generation — but not the tenants' IT load, so flexibility is offered on the infrastructure side rather than through workload curtailment.
Hyperscale	100–1,000 MW+	Single-operator, purpose-built. Sophisticated in-house power and workload management, large UPS/BESS and backup generation. The primary near-term Virtual N-1 candidate.
GPU cloud (neocloud)	20–300 MW	Capacity rented to third-party AI users. As in colocation, the site owner controls the infrastructure but not the workload; sub-second power-capping is not acceptable to tenants, so flexibility is offered through UPS and generation.
AI factory	300 MW – 2 GW	Training-focused, GPU-dense, growing fastest. Very large concentrated load; training and batch workloads are often interruptible or shiftable, making deep curtailment credible where the owner runs the workload — whereas real-time inference is not. Failover of heavy AI workloads to another site is limited: the data volumes cannot be moved fast enough, and idle standby AI capacity is costly.

Table 2 *Typology of data centre facilities by size and flexibility characteristics (authors' interpretation).*

Two features matter for what follows. First, the relevant flexibility is concentrated in the larger categories, where one connection point represents hundreds of megawatts. Second, AI training is often interruptible in ways inference is not, which gives AI factories additional workload flexibility where the owner runs the workload (Section 10.1). The most dependable response, however, lies in the power architecture: a colocation or GPU-cloud site with a well-qualified UPS and generation chain may be as creditable as an AI factory.

3.1 Redundancy as Structural Slack

Data centres are built to redundancy standards — the Uptime Institute's Tier classification and N, N+1 and 2N notation (Uptime Institute, 2024) — that provide at least one, and often a full duplicate, set of power and cooling paths beyond what the load requires.

This redundancy keeps critical load supplied through any single failure, but for most of its life it is installed, maintained and unused. Under contract and within bounded windows, the same slack can be made available to the grid; the higher the tier, the greater the latent capacity.

There is an important qualification, and it applies mainly to stored energy. Under Uptime-type certification, running on generation is normal operation and does not reduce the site's maintainability. Battery energy is different: the UPS must always retain enough charge to carry the load until generation has taken over. Only energy above this ride-through reserve can be committed to the grid. The site operator has, in effect, its own N-1 criterion, and qualification must specify how much of the redundant capacity can be committed without compromising it — a direct parallel, at facility scale, to the system-level trade-off that PRA and DSA make explicit in White Paper 1.

> ***Key insight: redundancy is committable flexibility only up to the point where it would compromise the site's own security. Qualification makes that boundary explicit — for the grid and for the data centre operator alike.***

4. Anatomy of Flexibility — A Taxonomy

A data centre's demand is not one indivisible block but a sum of components with sharply different tolerances for interruption. Only part of it is genuinely inflexible; making the decomposition explicit is the empirical basis for qualification (ScienceDirect, 2025).

4.1 Four Components of Load

- **Critical load.** Compute and infrastructure that must not be interrupted — latency-sensitive services, live transactions, control systems. This is the load Virtual N-1 must always protect. It corresponds to priority load in the White Paper 2 definition.
- **Shiftable load.** Workloads that can be moved in time or to another site without loss — batch processing, certain training jobs, replication and backup tasks. Shiftable load is the richest source of deep, planned flexibility.
- **Deferrable load.** Tasks that can be delayed for a bounded period — non-urgent compute, model checkpointing, data movement. Deferrable load provides shallow-to-moderate flexibility on minutes-to-hours timescales.
- **Cooling load.** HVAC and heat rejection, typically 20–40% of total facility demand depending on efficiency (PUE). In older, air-cooled facilities, thermal inertia allows cooling to be modulated for short periods. At the power densities of modern AI facilities thermal inertia is short, servers compensate by raising fan speed, and cooling in practice follows the IT load rather than offering independent flexibility.

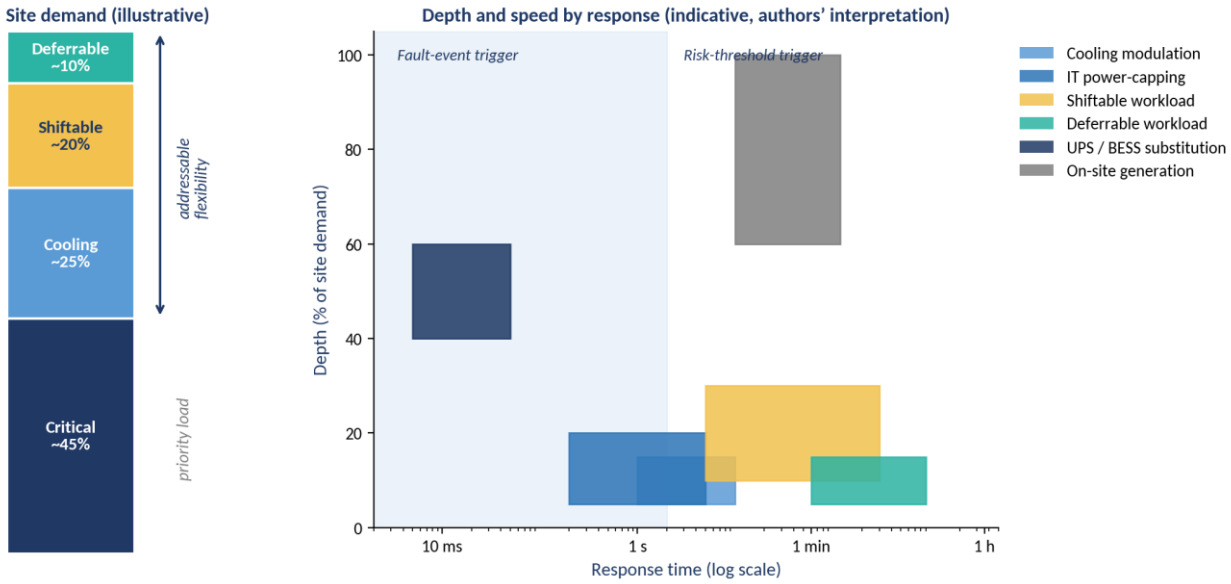


Figure 3 *Decomposition of data centre demand into critical, shiftable, deferrable and cooling components, each with a characteristic depth, speed and duration. Only the non-critical components constitute addressable flexibility. Conceptual illustration; values are indicative.*

4.2 Four Response Mechanisms

Load components describe *what* can respond. Response mechanisms describe *how*. The distinction matters because the same megawatt of shiftable workload can be relieved in different ways, with different consequences for the grid and for the service. Four mechanisms cover the space:

Mechanism	What it does	Grid effect	Binding constraint
Electrical substitution	UPS, BESS or on-site generation replaces grid supply for a bounded period	Grid import falls; compute continues unchanged	Stored energy, fuel, permits; recharge afterwards

Mechanism	What it does	Grid effect	Binding constraint
Computational curtailment	Power capping or reduced computing rate lowers MW directly	Immediate MW reduction at the connection point	Service degradation; requires the site owner to control the workload — not the case in colocation or neocloud
Temporal flexibility	Workload is paused or delayed to a later time	MW reduced now, energy returns later	Job deadlines; rebound when work resumes
Spatial flexibility	Workload is relocated to another data centre or grid area	MW leaves this node, appears at another	Data locality, latency, network capacity, sovereignty; receiving site headroom; limited for heavy AI workloads

Table 3 *Four response mechanisms for data centre flexibility (authors' interpretation, drawing on Lex Fridman Podcast, 2026).*

The first two mechanisms change what the site draws without deferring anything and suit fault-event triggers; the third moves energy in time and always carries a rebound; the fourth moves demand through the data network rather than the electricity network, and its value depends on where the work lands (Section 7.3).

4.3 Three Axes: Depth, Speed and Duration

Three properties govern each response: **depth** (how much load can be shed), **speed** (how quickly) and **duration** (how long it can be held). Depth and speed map onto the two triggers of White Paper 2 — the fault-event trigger needs protection-speed response, the risk-threshold trigger can also use slower ones. Duration determines whether the response lasts until the system operator has re-secured the system by other means.

Response	Typical depth (% of site)	Response time	Sustainable duration	Primary trigger fit	Rebound risk
Cooling modulation	0–5% (follows IT load at high density)	Seconds	Minutes in air-cooled sites; negligible at high density	Not independently committable	Moderate — thermal recovery
IT power-capping	5–20% (owner-operated sites)	Seconds; sub-second not acceptable in colocation/neocloud	Minutes to hours	Risk-threshold	Low to moderate
Shiftable workload	10–30%	Seconds to minutes	Hours	Risk-threshold	High — deferred work returns
Deferrable workload	5–15%	Minutes	Minutes to hours	Risk-threshold	High — deferred work returns
UPS / BESS substitution	Up to critical load	0 ms (online) to ~4 ms (line-interactive)	Minutes — energy above the ride-through reserve	Fault-event	Moderate — battery recharge (can be covered by generation)
On-site generation	Up to full site	Seconds to minutes	Unlimited technically; fuel- and permit-limited	Both triggers	Low — can support recovery

Table 4 *Indicative depth, speed, duration and rebound of data centre flexibility by response (authors' interpretation; site-specific values vary widely).*

The taxonomy yields a clear hierarchy. The backbone is electrical substitution: an online UPS responding within milliseconds and bridging to on-site generation, which can carry a large share of the site for as long as fuel and permits allow, without touching the computing service. Workload flexibility is an addition where the owner also runs the

workload; cooling adds little at high density. Which part of the envelope matters depends on the trigger and on the time needed to re-secure the system.

4.4 Recovery and Rebound

A response does not end when the trigger clears. Deferred compute returns as additional load; a depleted battery recharges; a cooling system recovers its thermal margin. If many sites return simultaneously, the recovery can itself create a new peak on the very corridor that triggered the event. Recovery behaviour is therefore part of the resource, not an afterthought.

Recovery is described by three parameters: the **maximum ramp-up rate** on return, the **rebound energy** (deferred or recharging energy served afterwards) and the **release condition** (automatic, or only on operator release). Ramping generation down gradually while the battery recharges keeps recharge from appearing as a new peak on the grid.

A site therefore has no single quantity called “flexibility”. It has a **response envelope**: MW as a function of time, operating state and trigger, together with a recovery trajectory and a probability that the response will fail when called. Qualification is the process of turning that envelope into a dependable commitment.

Key insight: depth without duration overstates the value of a response. Duration without controlled recovery can move the problem rather than solve it. A dependable resource specifies all three.

5. The Qualification Framework

This is the core contribution of the paper. Virtual N-1 requires that a conditionally connected load behaves as a specified safety mechanism. A qualification framework defines what a load must demonstrate to be admitted to conditional connection, in terms that a system operator can verify and a regulator can audit. It converts the vague notion of a "flexible load" into a certifiable resource.

In Huang's formulation, as soon as you can specify it, you can engineer it (Lex Fridman Podcast, 2026). The qualification framework is that specification, written from the grid side: the properties a system operator must be able to require before a response can be counted in a security decision.

5.1 Definition and Five Qualification Criteria

> ***Qualified Flexible Load — Proposed Definition. A qualified flexible load is a connected demand resource whose committed reduction in grid import is specified by depth, response time and duration; activated through a defined and compatible trigger; verified through testing and telemetry; and represented in system security assessment together with its availability, recovery behaviour and probability of failure on demand.***

A load qualifies as a Virtual N-1 resource to the extent that it can demonstrate five things, in advance and under audit:

- **Committed depth.** A specified minimum curtailment, expressed as a fraction of connected load, that the site will deliver on trigger. The commitment must be firm enough that the system operator can plan around it — the load "pays" for its own N-1 margin through this commitment (White Paper 2, Factor C).
- **Committed speed.** The maximum time from trigger to delivered curtailment, matched to the trigger mechanism the connection relies on. A fault-event trigger requires protection-timescale response; a risk-threshold trigger allows seconds to minutes. The certificate states the requirement precisely — for example, "≤ X ms from validated trigger receipt to Y MW measured at the connection point" — rather than relying on a tier label.
- **Committed duration.** The minimum time for which the committed depth will be sustained once delivered. Duration must at least cover the time the system operator needs to restore a secure state by other means; for energy-limited assets such as UPS and BESS it is the criterion most likely to bind.
- **Trigger compatibility.** Pre-installed, automatic protection or control that responds to the agreed trigger without operator intervention — fault-event, risk-threshold, or both — as specified in White Paper 2. Manual or discretionary response cannot serve as a Virtual N-1 security barrier, since the White Paper 2 definition requires automatic protection. Slower firmness classes (Section 7.2) may, however, admit operator-initiated or centrally dispatched response where the required response time, dispatch chain and reliability are explicitly qualified.
- **Verifiability.** The ability to test, monitor and audit the committed response — through commissioning tests, periodic verification and telemetry — so that the commitment is demonstrable, not merely asserted. This is what allows the load to be counted in reliability accounting.

5.2 Qualification Constraints

The five criteria define what response the load commits to deliver. They do not, by themselves, establish how reliable that response will be available when required. **The relevant question is therefore not only whether a flexible load can respond, but what probability the contracted response will be available and delivered when demanded.** Three constraints describe the conditions under which it can be relied upon, and must be declared alongside it:

- **Availability state.** The committed response may depend on the site's own operating state — battery state of charge above the ride-through reserve, generator availability and fuel stock, workload mix, ambient temperature. Qualification must specify in which states the commitment holds and require the site to signal in real time when it does not.

- **Recovery and rebound.** The maximum ramp-up rate, rebound energy and release condition described in Section 4.4. A qualified resource returns in a controlled way; it does not create a second event.
- **Failure modes.** A protection-based response has a probability of failure on demand, just as a relay does. Dependencies matter as much as failure rates: five response elements may share one control chain — one trigger interface, controller and communications path. Common-mode dependencies within a site, and across sites sharing software, vendors or control platforms, must be identified and represented, since these are the inputs PRA needs (Svendsen, Nyiredy & Milenkovic, 2026a).
- **Cyber and control-path integrity** (part of the failure-mode constraint). Qualification must specify behaviour under loss, corruption or delay of the trigger and telemetry channels, including authentication of the trigger, fail-safe response, local autonomous fallback and configuration-change control. A resource that is part of grid security is also part of its attack surface and should be treated within the framework of the EU Network Code on Cybersecurity for the electricity sector.

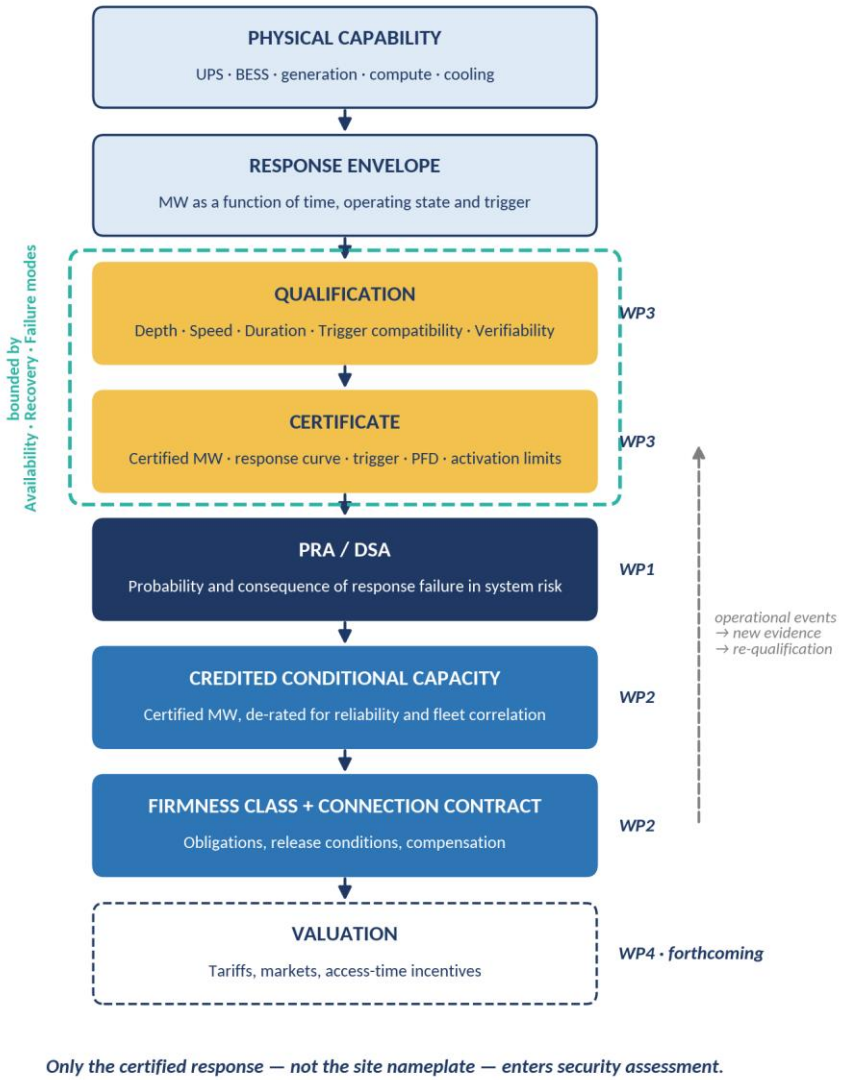


Figure 4 *The argument on one page: from physical capability through certified capability to credited conditional capacity. Qualification turns a response envelope into a certificate, bounded by availability, recovery and failure modes; PRA/DSA determines how much of the certified response can be credited, taking account of reliability and fleet correlation; the contract allocates firmness; valuation follows in WP4. Conceptual illustration.*

5.3 A Qualification Matrix

Combining the criteria with the triggers of White Paper 2 gives a qualification matrix: online UPS substitution and generator hand-over qualify for the fault-event trigger; workload shifting only for the slower risk-threshold trigger. The tiers are illustrative design classes — proposed hypotheses for pilot testing, not normative thresholds.

Tier	Depth committed	Speed	Duration	Trigger	Recovery requirement	Typical enabling assets
Tier 1 — Deep, fast	20–30%	Protection timescale (sub-second)	Minutes, above the ride-through reserve	Fault-event	Operator release; recharge ramp-limited	Online or line-interactive UPS/BESS
Tier 2 — Moderate, fast	Typically 15–25%; potentially much higher where full-site generation hand-over is qualified	Seconds to minutes	Hours; generation unlimited within fuel and permit	Both	Ramp-limited return; generator-supported recharge	On-site generation hand-over
Tier 3 — Moderate, managed	10–20%	Minutes	Hours	Risk-threshold	Coordinated rebound schedule	Shiftable and deferrable workload (owner-operated sites)
Tier 4 — Shallow	5–10%	Minutes to hours	Hours	Risk-threshold	Declared rebound	Deferrable workload, limited cooling headroom

Table 5 *Virtual N-1 qualification matrix for flexible loads (authors' interpretation). Tiers 1–4 retain the curtailment tiers of White Paper 2, Section 8, and restate them as qualification requirements, extended with duration and recovery.*

In practice, a large facility will combine several tiers: part of its load qualified at Tier 1 through UPS substitution, a larger part at Tier 2 through generator hand-over and — where the owner also runs the workload — a further part at Tier 3 through workload shifting. The certificate describes the stack, not a single number. Appendix B works through a complete example for a 300 MW campus.

5.4 The Qualification Certificate (proposed)

The output of qualification is a certificate. It records, for each committed part of the site's response, the values the system operator will rely on:

Field	Certified value
Qualified capacity	MW committed under conditional terms, by tier
Trigger	Fault-event, risk-threshold or both; signal path and interface
Initial response	MW delivered within a stated time from trigger
Sustained response	MW held for a stated minimum duration
Availability conditions	Operating states in which the commitment holds, and how loss of availability is signalled
Recovery limit	Release condition and maximum ramp-up rate (MW/min), including recharge
Telemetry	Measured quantities, resolution and latency
Trigger PFD	Probability of failure on demand of the trigger chain
Response PFD	Probability of failure on demand of each committed response element
Common-mode and dependency structure	Shared trigger paths, communications, software, control platforms, auxiliary supplies, vendors and physical dependencies across committed response elements
Activation limits	Maximum event frequency, cumulative duration and minimum recovery time between activations (generator run-hours, battery cycling, thermal recovery, workload backlog)

Field	Certified value
Resource exclusivity and priority	Capacity reserved for the conditional-connection obligation, and rules for simultaneous participation in other services (e.g. FCR, balancing)
Control-path integrity	Trigger authentication, behaviour on loss or corruption of communications, local fallback, configuration-change control
Verification interval	Frequency and scope of re-testing
Validity and re-qualification	Period of validity, plus change-based triggers: material change to firmware, orchestration software, protection settings, workload composition, BESS or generator configuration, network connection, trigger interface or dependency structure

Table 6 *Proposed content of a flexible-load qualification certificate (proposed framework).*

The certificate, rather than the site's nameplate or theoretical flexibility, is the object used by the system operator in security assessment and reliability accounting. It is not, however, the same as the capacity credited in that assessment. Three quantities should be distinguished: **physical capability** (what the site can do), **certified capability** (what has been specified and verified), and **credited capacity** (what the system operator counts after PRA has taken account of reliability, availability and correlation). A certificate for 100 MW may justify a credit of 85 MW; the difference is determined by the risk model, not by the certificate.

> ***Capability is what the load can do. Certification is what it has demonstrated. Credit is what the system operator can rely on.***

5.5 From Qualification Back to PRA

Qualification closes the loop between the load-side response and the Decision Layer of White Paper 1. The resource is specified; its response is tested; its probability of failure on demand is measured; PRA represents that probability and its consequence in system risk; the system operator determines how much conditional capacity can be admitted within an agreed risk threshold; operational events generate new evidence; and the resource is re-qualified.

> ***Qualification does not make the response certain. It makes its uncertainty measurable. PRA does not require a resource never to fail; it requires the probability and consequence of failure to be represented with sufficient confidence for the operating decision being made.***

Appendix B illustrates why this matters: the same site adds about 2 MWh/yr of expected energy not supplied to priority load when its response is qualified, and about 19 MWh/yr when the same response is assumed but unverified (illustrative values). The assets are identical; what changes is the confidence with which their response can be represented in system risk.

Site qualification is necessary but not sufficient for portfolio credit. If twenty qualified sites share the same GPU platform, orchestration stack, BESS supplier, cloud control plane or telecom provider, the probability that many fail together can be far higher than the individual certificates suggest. System operators must therefore also assess correlated failure across the qualified fleet, and de-rate portfolio credit accordingly.

5.6 From Capability to Grid Resource: The Control Chain

Figure 5 shows the operational sequence: quantified risk defines when a trigger is needed; the trigger activates the contracted response; qualification establishes its characteristics and reliability; the data centre delivers it, grid load falls and compute continues; telemetry and re-verification keep the certificate true.

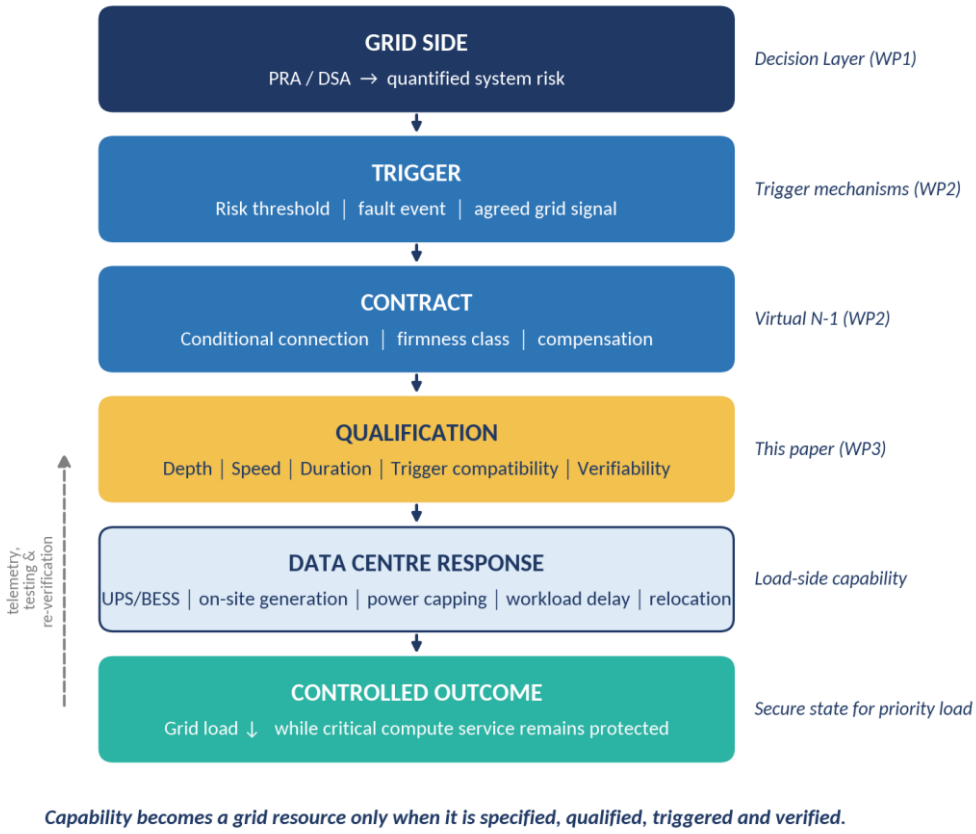


Figure 5 *Operational sequence during a trigger event — from quantified grid risk to a secured post-contingency state: the control chain linking PRA/DSA, trigger, conditional-connection contract, qualification and data centre response. Qualification is the link that makes the chain dependable.*

5.7 What Is New — and What Is Not

None of the individual building blocks is new. Interruptible connections, remedial action and special protection schemes, demand response, UPS and BESS, and probabilistic reliability assessment all exist today. The contribution proposed here is their integration into a common qualification chain: a conditional load response is specified, certified, assigned a probability of failure on demand, represented in PRA, linked to a firmness class and used explicitly to determine admissible connection capacity.

Qualification is equally not a promise of permanent curtailment. It specifies the conditional portion of the load — the fraction that has accepted, and been compensated for, interruptible terms. Critical load remains outside the committed conditional portion and must continue to satisfy the site's specified internal security requirements. Nor is qualification static: a site's certificate changes with its assets, software and workload, and is re-qualified both periodically and on material change.

6. Reserve and Storage as Active Resources

The qualification tiers in Section 5 depend heavily on the site's reserve and storage assets. These assets — originally installed for site protection — determine how deep, how fast and for how long a facility can respond. Understanding them as grid resources rather than passive backup is essential to qualifying the flexibility.

6.1 Backup Generation

Generators — diesel, gas or HVO-fuelled — are sized for the full critical load, start within seconds and, under Uptime-type certification, are rated for unlimited running hours. As a grid resource they let a site island for as long as fuel is secured, turning a minutes-long battery response into an open-ended one; they are the backbone of the most dependable data-centre response. The practical limits — fuel logistics and permits for non-emergency operation — should be addressed explicitly in qualification, preferably with low-carbon fuels.

6.2 Battery Energy Storage (UPS and BESS)

Lithium-ion UPS and dedicated BESS turn the battery from a bridging device into a fast-responding asset — the cleanest form of Virtual N-1 response. Its limit is energy: the UPS must always keep enough charge to carry the site until generation has taken over (its ride-through reserve), and only energy above that reserve can be committed. The remaining capacity can serve frequency response or arbitrage, improving the commercial case (Section 7).

6.3 Hydrogen and Fuel Cells

Fuel cells and hydrogen storage are an emerging alternative to diesel for backup and, potentially, for longer-duration grid support. Their relevance to Virtual N-1 is prospective rather than established, but they extend the same logic: on-site energy that protects the site can also, under contract, protect the grid. As hydrogen infrastructure scales under European and global policy frameworks, fuel-cell-equipped facilities may qualify for deeper and longer commitments than battery-only sites.

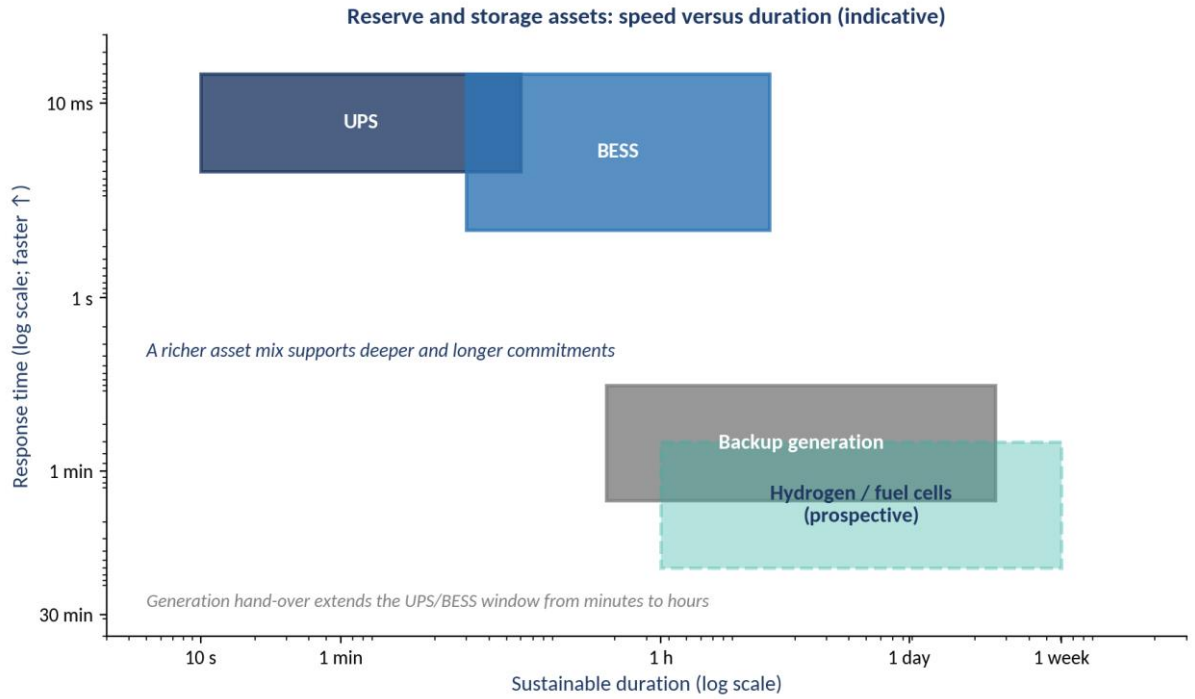


Figure 6 *Reserve and storage assets as a spectrum from fast/short (UPS, BESS) to slow/long (backup generation, hydrogen). Depth and duration of committable flexibility scale with the asset mix. Conceptual illustration; values are indicative.*

7. Mapping Flexibility to Power System Needs

Qualified data centre flexibility is not useful in the abstract; it is useful because it meets specific, defined system needs. The same physical capability — the ability to reduce or shift load quickly and by a known amount — serves several distinct grid services, on different timescales. Mapping the flexibility to these services shows where the value lies and how a single facility can participate in more than one.

System need	Timescale	Data centre capability used	Relation to the series
Frequency containment (FCR)	Seconds	BESS/UPS fast injection or fast load reduction	System service; complements Virtual N-1
Frequency restoration (aFRR/mFRR)	Seconds to minutes	Power-capping, workload shifting, generation	Security Layer (WP1)
Congestion management	Minutes to hours	Shiftable/deferrable load, spatial relocation, islanding	Physical and Security Layers (WP1)
Virtual N-1 conditional connection	Milliseconds to minutes, sustained for the committed duration	Full qualification stack (Section 5)	Core WP2 principle

Table 7 *Mapping of data centre flexibility to power system needs (authors' interpretation).*

Virtual N-1 is the most structurally significant of these services but not the only one. A facility qualified for it can also provide frequency and congestion services in ordinary operation, so the equipment is paid for by everyday market participation and called upon rarely as a safety mechanism — provided the capacity committed to Virtual N-1 is ring-fenced from what is traded elsewhere.

7.1 Precedents in Operation and Regulation

The mapping above is not hypothetical. System operators and regulators are already formalising the relationship between large data centre load and grid security, and operators are already demonstrating the technical response:

- **PJM (US) — bring capacity or be curtailed first.** PJM's proposed Interim Resource Adequacy Service (filed August 2026, pending FERC approval) would curtail new loads of 50 MW or more that do not bring new capacity ahead of pre-emergency load management in capacity shortages. It replaces the withdrawn NCBL concept cited in White Paper 2 (PJM, 2026; Datacenter Knowledge, 2026).
- **SPP (US) — accelerated access for curtailable load.** A 90-day study and approval pathway is available for data centres and other large loads paired with generation or agreeing to a curtailable supply (IEA, 2026a).
- **ERCOT / Texas — mandatory disconnection capability.** Senate Bill 6 (2025) requires new loads of 75 MW or more connecting after 2025 to be capable of disconnection during firm load shed, and lets ERCOT direct backup-generation deployment in emergencies; implementing rules are being finalised in 2026 (Texas Legislature, 2025; PUCT, 2026).
- **Ireland — mandated on-site capacity.** New data centres must provide dispatchable generation and/or storage matching their maximum import capacity (IEA, 2026a) — the asset base required for Tier 1 and 2 qualification, mandated before the qualification itself is defined.
- **Denmark and the Netherlands — flexible connection agreements.** Danish grid operators offer accelerated, lower-cost connection to large loads that accept being disconnected or reduced at any time in case of constraints; in the Netherlands, non-firm contract types are expected to free around 9 GW of hosting capacity (IEA, 2026a; Ember, 2026).
- **Oracle, Phoenix and Nebius, London — demonstrated depth, duration and recovery.** Under EPRI's DCFlex initiative (EPRI, 2024), orchestration software held a 25% power reduction on a 256-GPU cluster for three hours, with a 15-minute ramp-down and no overshoot on return; a later London trial reported ~30% within 30 seconds, sustained up to ten hours (Colangelo et al., 2025; NVIDIA, 2026). This evidences Tier 3 workload flexibility — compute only, at owner-operated sites.

- **Aligned Data Centers, Hillsboro (Oregon) — storage as a connection enabler.** A ~31 MW / 62 MWh battery beside the site lets the utility reduce its draw at peak in exchange for earlier connection (Canary Media, 2025; Latitude Media, 2026).

These precedents show both progress and the gap. The PJM and ERCOT measures are adequacy and emergency mechanisms, activated in system-wide scarcity rather than on a specific contingency, and they define no depth, duration or verification that would count in an N-1 assessment; the demonstrations prove the response but sit outside any certification regime. What is missing is a common qualification framework — the contribution of Section 5.

7.2 Differentiated Firmness Classes

Today the connection offer to a large load is mostly binary: firm after reinforcement, or nothing. The contractual chain reinforces this — end customers demand very high availability, the operator passes it on, and the utility is asked for equivalent firmness even where the service does not need it in every hour (Lex Fridman Podcast, 2026). A hospital, an inference cluster and an interruptible training run all appear in the queue as megawatts, but the consequence of reducing them differs entirely.

Qualification opens a design space between the two extremes: **differentiated firmness classes** for large loads, in which the contracted level of delivery guarantee, the system risk it implies and the certified response capability are linked. Table 8 illustrates one possible structure. It is offered as a design space for regulators and system operators, not as a finished proposal.

Firmness class	Delivery guarantee	Qualification required	Typical load	Commercial logic
Firm (priority)	Standard firm connection under applicable security criteria	None beyond standard connection	Critical infrastructure; latency-sensitive inference	Standard tariff; connection after reinforcement where needed
Conditional — fast	Curtailable on fault-event trigger for committed duration	Tier 1 or 2 certificate	Hyperscale sites with UPS/BESS and generation	Earlier connection; reduced capacity charges
Conditional — managed	Curtailable on risk-threshold trigger	Tier 3 or 4 certificate	Training and batch compute; shiftable workload	Earlier connection; availability payments
Non-firm (energy-only)	Access only when headroom exists	Trigger compatibility and telemetry	Opportunistic compute; flexible industrial load	Lowest cost; fastest access

Table 8 *Illustrative differentiated firmness classes for large loads, linking delivery guarantee to certified response capability (authors' interpretation).*

The value chain is: earlier connection ↔ differentiated firmness ↔ contracted controllability ↔ a different price. The task of White Papers 2 and 3 together is to show under which technical and regulatory conditions that exchange can be made without weakening security for priority load.

7.3 A Forward Look: Compute Follows Power

Spatial flexibility suggests moving the computation to where network and generation capacity are available. For a constrained corridor, relocation is equivalent to deep curtailment; for the wider system it is geographically distributed flexibility.

The idea is introduced here without making it a central claim. It depends on data locality, latency, bandwidth, sovereignty and the receiving site's headroom, and is particularly limited for heavy AI workloads, whose data cannot be moved fast enough. A spatial response should be credited only where the receiving path is itself verified.

8. Bounding the Potential: An Illustrative System-Level Translation

Read this section with its status in mind: the figures below are indicative hypotheses, not modelled results, universal forecasts or transferable capacity figures. They lie within — not on top of — the +10–20% Virtual N-1 estimate of White Paper 2.

How much does qualified data-centre flexibility contribute at system level? Two approaches anchor the estimate from opposite ends: top-down, the data-centre share of White Paper 2's +10–20%; and bottom-up, built from the data-centre load seeking connection and the flexibility it can qualify.

Both are presented here as indicative hypotheses grounded in empirical evidence, not modelled forecasts. The purpose is to show order of magnitude and method, and to be explicit about how the data centre contribution relates to the figure already published in White Paper 2, so that the two are not silently double-counted.

8.1 The Bottom-Up Estimate

The bottom-up estimate follows the method of White Paper 2 (Section 8), restricted to data centre load. The contribution is the product of three factors; the fourth — committed depth — acts as an adequacy condition rather than a multiplier, exactly as in White Paper 2, where the committed curtailment must be sufficient to restore N-1 compliance for priority load.

- **Factor Q — Data centre queue share.** Data centre load seeking connection, as a percentage of current utilisable capacity on data-centre-heavy corridors. In White Paper 2 the total queued load was taken as 20–40% of utilisable capacity; data centres are assumed to represent roughly half to two-thirds of it in such systems.
- **Factor A — Eligible share.** The fraction of that load that is technically and commercially suitable for qualification: a qualifiable electrical architecture with sufficient UPS and on-site generation, trigger and control capability, fuel and permits for non-emergency operation, and commercial willingness. Control of the workload is an addition, not the entry ticket — so colocation and GPU-cloud sites can qualify through electrical substitution. This broader basis could raise Factor A; the ranges below are kept unchanged pending validation.
- **Factor B — Qualified depth and duration (adequacy condition).** The weighted-average committed depth across the eligible fleet, drawn from the tiers in Section 5.3, held for at least the duration the system operator needs to re-secure the system. Where committed depth falls short of the required N-1 margin, less load can be accommodated per MW of headroom; this shortfall is reflected in Factor C.
- **Factor C — System translation efficiency.** How committed curtailment translates into additional connectable load, given the N-1 headroom on the relevant circuits (White Paper 2, Factor C).

Parameter	Low (illustrative)	Central (illustrative)	High (illustrative)
Data centre queue share of utilisable capacity (Factor Q)	15%	20%	25%
Eligible share (Factor A)	35%	50%	60%
Weighted committed depth (Factor B, adequacy condition)	12%	18%	25%
Required N-1 margin (system-dependent)	20%	18%	15%
System translation efficiency (Factor C)	70%	80%	90%
Indicative contribution (Q × A × C)	~4%	~8%	~13%

Table 9 *Illustrative parameter sets for the bottom-up translation (authors' interpretation). Values illustrate method and range, not a forecast.*

The cases are illustrative parameter sets, not points on a probability distribution. The central case — queue share 20%, eligible share 50%, translation efficiency 80% — gives about 8% of utilisable capacity on data-centre-heavy corridors, with roughly 4–13% across the low and high cases (authors' interpretation). Committed depth matches the required N-1 margin in the central case, falls short in the low case and exceeds it in the high case, which is why translation efficiency varies. The product is a screening approximation, not an assumption of independence: qualification rate, availability and concentration may be correlated — for example, hot weather or coincident compute peaks can reduce available response exactly when the system is stressed — and a system study should represent them jointly.

Read against the precedents, these ranges are plausible: TenneT's non-firm contracts unlocked 9.1 GW, about 40% of Dutch peak demand (Ember, 2026), and a Lawrence Berkeley National Laboratory analysis found that 76 GW of additional US load could be integrated if data centres curtailed for only 0.25% of maximum uptime (Knowledge Problem, 2026).

8.2 The Top-Down Anchor

The top-down approach starts from the published figure. White Paper 2 estimated the Virtual N-1 potential at +10–20% of utilisable capacity, drawing on the full range of eligible load categories — data centres, electrolysers, aluminium and other flexible industry, and co-located storage. Data centres are the largest and fastest-growing of these categories in most systems now facing large-load queues (IEA, 2026a; Ember, 2025), but they are not the whole of it.

If data centres represent half to two-thirds of the eligible flexible load in a data-centre-heavy system, their share of the +10–20% Virtual N-1 potential is of the order of +5–13% — a portion of White Paper 2's figure, not an addition to it. It inherits the anchoring of White Paper 2 (v1.2, Section 9.1), which allocates the IEA's 750–900 GW of non-firm potential (≈12–18%) to Virtual N-1 so that nothing is counted twice.

8.3 Reconciling the Two

The two estimates land in a similar range — roughly 4–13% bottom-up and 5–13% top-down. That overlap is a useful consistency check, but not independent validation of the +10–20%, since both rest on the same assumptions about eligible load and on the logic of White Paper 2.

Estimate	Basis	Indicative range	Relation to WP2
Bottom-up	Factors Q × A × C, with depth and duration as adequacy condition	~4–13% on data-centre-heavy corridors	Separate bottom-up construction
Top-down	Data centre share of the WP2 +10–20%	~5–13% of utilisable capacity	Portion of the same figure

Table 10 *Two anchors for the data centre flexibility contribution (authors' interpretation). The estimates are consistent and must not be added — they describe the same underlying capacity.*

These figures describe the same capacity and must not be summed. The data centre contribution lies within the White Paper 2 +10–20%, not on top of it. What this section adds is a transparent method for estimating how much of that figure a given system might expect from its own data centre fleet. Appendix B illustrates the same logic at the level of a single node.

9. Regulatory and Commercial Framework

The qualification framework of Section 5 is only useful if it is embedded in regulation and contract. White Paper 2 set out the system-side regulatory pathway for conditional connection; this section addresses the load-side counterpart: the standards, verification and commercial arrangements needed to certify and reward a qualified flexible load.

9.1 Qualification Standards and Verification

If qualification is to carry regulatory weight, validated criteria will ultimately need to be codified as standards — how depth, speed and duration are measured, how trigger compatibility is certified, and how and how often verification is performed — building on existing conformity assessment for protection systems and generators. The immediate step is pilot testing and pre-standardisation work. Who performs qualification — the system operator or an accredited independent body — remains for each jurisdiction to decide: the framework specifies what must be certified, not who certifies it.

Verification should rest on availability and response, not on estimated baselines, which the UK independent review warns are open to gaming (Yu, 2026). This points to three levels: **commissioning verification** — can the site deliver, tested end to end from trigger to measured MW; **continuous qualification state** — is the capability available now, shown by telemetry; and **event verification** — did the site deliver, assessed against the committed response curve. In short: can, available, did.

9.2 The Commercial Bargain

The load accepts interruptibility and installs the enabling assets; in return it gets faster or cheaper connection and revenue from everyday grid services (Section 7). The contract must specify who bears the cost of a curtailment event, what compensation applies, and how the system operator limits the frequency and duration of curtailment — the questions of White Paper 2, now anchored to a certified resource and a defined firmness class (Section 7.2).

9.3 Treatment in Reliability Accounting

Qualified flexibility is non-firm, and that distinction must be preserved in planning and reliability indices (LOLP, EENS). Qualification makes the non-firm portion quantifiable: a resource with certified depth, speed, duration and probability of failure on demand can be modelled in PRA with explicit, auditable parameters rather than conservative assumptions. As Section 5.5 sets out, the aim is not a response that never fails, but one whose failure probability and consequence are known well enough for the decision being made.

9.4 European and National Pathway

The instruments through which a qualification standard could be introduced are already taking shape. Table 11 sets out the most direct routes.

Body	Required action
European Commission	Use the model tripartite agreement for data centre integration (planned for the second half of 2026) and the EU data centre rating scheme, which already covers flexibility, as vehicles for a common qualification standard, alongside the flexible connection provisions of the European Grids Package (European Commission, 2025; European Commission, 2026).
ACER and ENTSO-E	Extend work on flexible connection agreements and on common methodologies for risk-threshold triggers (White Paper 2) to include a common qualification standard for flexible loads, so that a resource certified in one system is recognised in another, building on CSAM (ACER, 2023; ENTSO-E, 2023).
National regulatory authorities	Develop conditional-connection tariffs and firmness classes that reward certified, tiered flexibility. Define how qualified non-firm load is treated in reliability accounting. In Great Britain, align with the development of risk metrics and thresholds proposed for 2030 (Yu, 2026).

Body	Required action
TSOs and DSOs	Specify trigger interfaces, telemetry and verification requirements; pilot qualification with data centres already in the connection queue.

Table 11 *Illustrative regulatory pathway for load-side qualification (proposed framework).*

10. Validation Boundaries and Open Questions

The estimates and proposals in this paper are indicative and subject to the following boundaries, which should be stated wherever the figures are cited.

10.1 Theoretical versus Realisable Flexibility

Independent analyses caution that the gap between theoretical and realisable data-centre flexibility is real (MIT CEEPR, 2025; Latitude Media, 2026). Four factors matter most. **Inference is not training:** latency-sensitive inference is hard to curtail, so qualification is assessed per facility and per workload, not per nameplate. **Ownership of the workload:** in colocation and GPU-cloud facilities the site owner does not control the workload, so the dependable response is electrical — UPS and generation — whose use depends on fuel and permits (practitioner review of this paper). **Criticality culture:** only verifiable, bounded, automatic commitments are likely to be acceptable to risk-averse operators. **Maturity and timing:** retrofit requires site-specific changes to protection, controls, telemetry and contracts; since a new facility can be built in around 18–24 months (Latitude Media, 2026), near-term potential lies mainly in new builds designed for trigger compatibility.

10.2 What the Industry Signal Does and Does Not Establish

The industry signal discussed in Section 1.3 establishes direction: that leading compute actors see graceful degradation, workload relocation and differentiated delivery guarantees as desirable and technically feasible. It does not establish any of the following, and should not be cited as if it did:

- validation of the 15–30%, 10–20% or 4–13% capacity ranges in this series;
- a universal flexibility depth — the illustrative reduction to around 80% of demand is an example, not a measurement;
- a system study — statements about unused or idle grid capacity are a CEO perspective, not a power-system analysis;
- qualification, telemetry, failure probabilities, reliability accounting or regulatory treatment, none of which were addressed;
- endorsement — NVIDIA has not endorsed Virtual N-1, White Paper 2 or this framework.

This division of labour is a strength rather than a weakness: the industry signal establishes capability and intent; this paper sets out the grid-side qualification and reliability logic that capability must satisfy.

10.3 Methodological Status

The 4–13% data centre contribution is an indicative hypothesis, derived from a structured analysis of empirical evidence and physical system constraints. It is not a modelled result, a universal forecast, or a transferable capacity figure, and it lies within — not on top of — the +10–20% Virtual N-1 estimate of White Paper 2. The qualification criteria and tier boundaries are proposals for consultation and validation, not established standards.

11. Conclusions and Specific Recommendations

This paper has turned the Virtual N-1 principle from a system-side arrangement into a load-side framework. Its argument runs from what makes a data centre different (an architecture built around ride-through), through a taxonomy of the flexibility latent in that architecture and the mechanisms through which it responds, to a qualification framework that certifies the response as a resource, and finally to the firmness classes, system value and regulatory arrangements that make the resource real.

The timing matters. The IEA, the European Commission and the UK independent review all point in the same direction — towards risk-based operation and conditional access for large loads — and the compute industry is independently describing data centres that can reduce their draw on request. What neither side yet has is the common specification between them.

The qualification framework proposed in WP3 should be viewed as technology-neutral. Data centres are the leading example because of their scale and urgency, but the framework is intended to apply equally to other large flexible loads.

> ***The central claim: conditional connection can be treated as a security mechanism only when the load-side response is qualified. Qualification turns flexibility from an assumed capability into a specified, triggered, verified and reliability-accounted commitment. This allows the system operator to evaluate explicitly whether additional conditional connection keeps risk to priority load within the agreed security threshold.***

Seen across the series, the argument is simple: PRA makes risk visible, Virtual N-1 makes it operational, and qualification makes it contractible. The data centre is where that logic becomes concrete first.

11.1 Specific Recommendations

- **For TSOs and system operators:** Begin mapping the flexibility of large loads already in the connection queue using the taxonomy in Section 4. Pilot qualification of data centre flexibility, starting with fault-event triggers and Tier 1/2 resources, which — as White Paper 2 set out for Stage 3 — can be deployed before the full Decision Layer is in place, and design telemetry, duration testing and recovery coordination from the outset.
- **For national regulatory authorities:** Develop qualification standards, firmness classes and conditional-connection tariffs that reward certified, tiered flexibility. Base verification on availability and response curves rather than historical baselines, and define how qualified non-firm load is treated in reliability accounting.
- **For the European Commission, ACER and ENTSO-E:** Use the model tripartite agreement and the EU data centre rating scheme to introduce a common qualification standard for flexible loads, building on the common methodology work on risk-threshold triggers (White Paper 2), so that a resource certified in one system is recognised in another.
- **For data centre and large-load operators:** Assess the qualification tier your facility can reach, and the commercial value of reaching it. Design power, storage and protection infrastructure to be trigger-compatible from the outset, reserve battery capacity for committed duration, and treat flexibility as a revenue-generating capability, not only a connection concession.
- **For policymakers:** Recognise that the loads that could qualify are already in the queue, and that several jurisdictions already mandate the on-site assets on which qualification depends. Defining how their response is qualified is as important as building new grid hardware (IEA, 2026a).
- **For standards bodies (IEC, IEEE, CIGRE, CENELEC):** Codify qualification criteria — depth, speed, duration, trigger compatibility, verification — and the recovery constraint in a vendor-neutral way, analogous to existing standards for grid-connected protection and generation, so that flexible-load certification becomes routine.

11.2 A Proposed First Pilot

Select one constrained transmission corridor and one large load with existing UPS/BESS and controllable demand.

> ***Define the required post-contingency response from the grid side; qualify the site's response against the five criteria; commission-test the full trigger-to-meter chain; estimate the probability of failure on demand; incorporate the qualified response into PRA; and compare the resulting conditional connection offer with the deterministic N-1 baseline.***

Appendix B shows what such a pilot would produce on paper; the pilot would show it in operation.

11.3 A Forward Pointer: Valuation and Markets

This paper answers "what is the resource?"; WP4 will address "what is it worth?" — how a certified, tiered resource should be valued and remunerated through connection tariffs, firmness classes, capacity mechanisms, reserve and balancing markets, and access-time incentives. Nordic discussion of capacity-based tariffs and proposals to trade faster connection for binding flexibility commitments are concrete instances of that question.

11.4 Key Takeaway: The Tenant in the Loft

White Paper 2 ended with a house: a foundation of PRA and DSA, two floors of grid-enhancing technologies and flexibility, and a conditional loft for a new kind of tenant who agrees to vacate on short notice. This paper is about that tenant. The data centre arrives already equipped — batteries to keep its own lights on while it steps out, generators to run independently, workloads it can pause or move. What it has lacked is not capability but terms: a way to prove, in advance and under audit, how fast and how far it will step aside, how long it can stay out and how quietly it will return.

The qualification framework provides those terms. The building code is not replaced by trust in the tenant. The tenant's response becomes part of the engineering case: specified, tested and represented explicitly in the risk calculation.

The building code is not replaced by trust in the tenant: its response is specified, tested and part of the risk calculation.

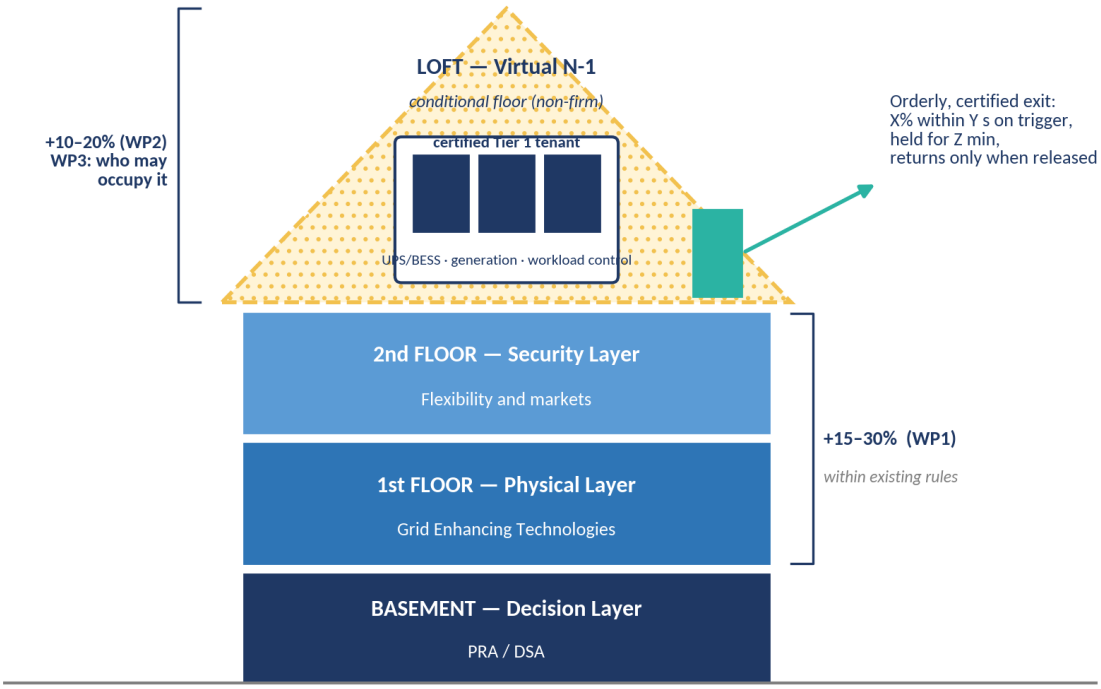


Figure 7 *The tenant in the loft. Extending the White Paper 2 house metaphor: the qualified data centre is the certified occupant of the conditional top floor, equipped with the storage, generation and controllable load that let it vacate the space safely and on demand. Conceptual illustration; each floor must be built before the next.*

References

ACER (2023). Decision on the Coordinated Security Assessment Methodology (CSAM). Agency for the Cooperation of Energy Regulators, Ljubljana.

Canary Media (2025). Oregon utility tries out a faster, cheaper way to power data centers.

Colangelo, P. et al. (2025). Turning AI Data Centers into Grid-Interactive Assets: Results from a Field Demonstration in Phoenix, Arizona. arXiv.

Datacenter Knowledge (2026). PJM's New Deal for Data Centers: Bring Power or Face Cuts.

Ei — Swedish Energy Markets Inspectorate (2023). Conditional grid connections — a literature review. EiR2023:08.

Ember (2025, September). Grids for data centres: ambitious grid planning can win Europe's AI race.

Ember (2026, April). Crossed wires: Grid capacity could block EU energy security.

ENTSO-E (2023). Coordinated Security Assessment Methodology (CSAM). Brussels.

EPRI (2024). DCFlex — Data Center Flexibility Initiative.

European Commission (2025). European Grids Package (COM/2025/1005). Brussels.

European Commission (2026). Strategic Roadmap for Digitalisation and AI in the Energy Sector. Brussels.

IEA (2026a). Electricity 2026. International Energy Agency, Paris.

IEA (2026b). Modernising Grids in the Age of Electricity. International Energy Agency, Paris.

Knowledge Problem (2026). Data Centers, Flexibility, and the Architecture of the Grid.

Latitude Media (2026). A Reality Check on Flexible Data Centers.

Lex Fridman Podcast (2026). Jensen Huang: NVIDIA and the AI Revolution.

MIT CEEPR (2025). Data Center Load Flexibility. Working Paper 2025-14.

NVIDIA (2026). How Emerald AI Makes AI Factories Power-Flexible.

PJM (2026). Proposed Framework to Connect Data Centers Without Compromising Reliability and Affordability.

PUCT — Public Utility Commission of Texas (2026). Rulemakings Implementing SB 6.

ScienceDirect (2025). Data Centres as a Source of Flexibility for Power Systems.

Svendsen, A. B., Nyiredy, R., & Milenkovic, M. (2026a). Unlocking Grid Capacity Through Risk-Based Operation. White Paper 1. The Future Power Grid.

Svendsen, A. B., Nyiredy, R., & Milenkovic, M. (2026b). From N-1 to Virtual N-1: Extending the Risk-Based Framework for Grid Capacity. White Paper 2. The Future Power Grid.

Texas Legislature (2025). Senate Bill 6. 89th Legislature.

Uptime Institute (2024). Tier Classification System for Data Center Infrastructure.

Yu, L. (2026). The Grid We Need Now. UK Government.

Appendix A — Research Questions for WP3

The following research questions structure the work programme underlying this paper and identify where system- and site-specific validation is required. They are intended as an agenda for the collaborative research effort behind White Paper 3.

- **RQ1 — Architecture.** How do the electrical architectures of different data centre types (enterprise, colocation, hyperscale, AI factory) determine the depth, speed, duration and reliability of their available flexibility?
- **RQ2 — Flexibility.** How can total facility demand be decomposed, in practice and with telemetry, into critical, shiftable, deferrable and cooling components, and how stable is that decomposition across operating conditions?
- **RQ3 — Storage and reserve.** How do UPS, BESS, backup generation and emerging hydrogen assets translate into committable Virtual N-1 depth and duration, and how much redundant capacity can be committed without compromising the site's own security?
- **RQ4 — Protection and triggering.** What protection and control architectures are required for fault-event and risk-threshold triggering at a data centre, and how is trigger compatibility certified and verified?
- **RQ5 — Recovery and rebound.** How large are rebound energy and ramp-up rates after a response, and how should recovery be coordinated across multiple sites on the same corridor?
- **RQ6 — Reliability of response.** What is the probability of failure on demand for each response mechanism, and which dependencies — within a site and correlated across a fleet of sites sharing software, vendors or control platforms — must be represented in PRA?
- **RQ7 — Spatial flexibility.** Under which conditions of latency, data locality, sovereignty and receiving-site headroom can workload relocation be credited as a qualified response?
- **RQ8 — Capacity potential.** What is the realistic aggregate contribution of qualified data centre flexibility to Virtual N-1 capacity in a given system, and how sensitive is it to fleet composition, storage penetration, firmness-class design and regulatory framework?
- **RQ9 — Certification stability.** How often must a flexible-load certificate be re-tested or updated as hardware, firmware, workloads, operating policies and site configuration change?

Appendix B — Worked Example: Qualifying a 300 MW AI Campus

This appendix applies the framework to a single, generic site. All parameters are illustrative and chosen for transparency, not taken from any project; the aim is to show the method end to end (authors' interpretation).

B.1 The System and the Request

A 300 MW AI campus requests connection at a 220 kV substation supplied by a corridor of two circuits. Each circuit has a continuous rating of 600 MW and a 15-minute emergency rating of 660 MW. Existing demand in the area peaks at 400 MW, with no local generation. Under deterministic N-1, the firm capacity of the corridor is the rating of one circuit, 600 MW; the firm headroom available for new load is therefore 200 MW. The remaining 100 MW could be offered firm only after corridor reinforcement, assumed here to take 7–10 years — consistent with the 5–15 years the IEA reports for new grid infrastructure (IEA, 2026a).

Parameter	Value	Note
Corridor	2 × 220 kV circuits, 600 MW continuous each	15-minute emergency rating 660 MW
Existing area peak demand	400 MW	Exceeds 300 MW in about half of all hours
Firm N-1 headroom	200 MW	600 MW minus 400 MW
Requested connection	300 MW	AI training and inference campus
Conditional portion	100 MW (33% of site)	Offered under Virtual N-1 (Stage 3)
Circuit trip frequency	0.5 per year	80% transient (successful auto-reclose), 20% permanent
Time to re-secure after a permanent trip	Up to 4 hours	Switching, redispatch or temporary repair

Table B1 *System and request parameters for the worked example (illustrative).*

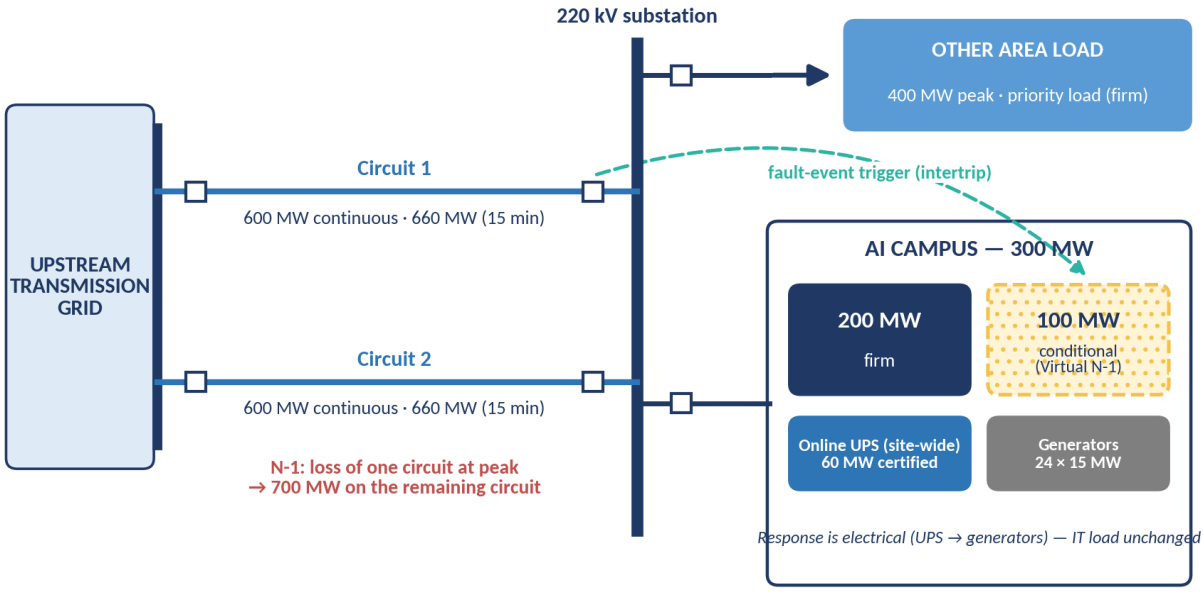


Figure 8 *Simplified system for the worked example: two incoming 220 kV circuits, other area load and the 300 MW AI campus, of which 100 MW is connected on conditional terms and responds to a fault-event trigger. Conceptual illustration; values are indicative.*

B.2 What the Grid Requires

With the campus connected at full import and area demand at peak, the loss of one circuit leaves 700 MW on the remaining circuit. Translating that state into load-side requirements gives four conditions that together define the resource the system operator needs:

- **R1 — Speed:** reduce grid import by at least 40 MW within 1 second, bringing loading below the 660 MW emergency rating.
- **R2 — Depth:** reduce import by at least 100 MW within 15 minutes, bringing loading below the 600 MW continuous rating.
- **R3 — Duration:** hold the 100 MW reduction for up to 4 hours, until the system operator has re-secured the system by other means.
- **R4 — Recovery:** return only when released by the system operator, at no more than 10 MW per minute, and without exceeding the contracted 300 MW import during recovery (including battery recharge).

These requirements follow from thermal limits. A DSA gate completes them: before the conditional connection is offered, DSA must confirm that the post-contingency state — with the response applied as certified, including the 60 MW within milliseconds and the hand-over to generation — remains within transient, voltage and small-signal stability limits. If it does not, the requirements tighten, for example to a faster or deeper initial response. The example therefore uses both parts of the Decision Layer: PRA for risk, DSA for the dynamic feasibility of the post-contingency state.

The trigger is a fault-event trigger — a circuit trip signal delivered by intertrip — so the arrangement can be deployed before a full Decision Layer is in place, as White Paper 2 set out for Stage 3. Curtailment is needed only when area demand exceeds 300 MW; at lower demand, the remaining circuit can carry the full load.

B.3 The Site's Response Stack

The campus is decomposed following Section 4: around 120 MW of critical load (inference, control systems and their cooling), around 100 MW of shiftable training workload, 20 MW of deferrable batch work and 60 MW of cooling. Its protective assets are those of a data centre built to an Uptime-type Tier III standard: a site-wide online UPS/BESS architecture sized for the critical load, of which 60 MW and 30 MWh above the ride-through reserve are certified for grid substitution, and on-site generation sized for the full site load with redundancy — here 24 generator sets of 15 MW, rated for unlimited running hours. Following practitioner input (Section 4), the response is built on electrical substitution rather than on computational curtailment: the IT load itself is not touched.

The reliability values used in this example are illustrative engineering assumptions intended to demonstrate the qualification and PRA methodology. They are not derived from site-specific field data and should not be interpreted as representative industry-wide failure rates.

Response	Mechanism	Depth	Speed	Duration	Tier
UPS/BESS substitution (committed part of site-wide online UPS)	Electrical substitution	60 MW	0–4 ms	≥ 30 min (30 MWh above the ride-through reserve)	Tier 1
Generator hand-over (8 × 15 MW committed of 24)	Electrical substitution	120 MW	≤ 2 min	Unlimited technically; limited by fuel logistics and permit for non-emergency operation	Tier 2
Training workload pause	Temporal flexibility	50 MW	≤ 10 min	≥ 6 h	Not committed (owner-operated sites only)
Cooling	—	Follows IT load	—	—	Not committable at high density

Table B2 *Response stack and qualification of the example site (illustrative).*

The certified capability is 60 MW (20% of site) at Tier 1 and 120 MW (40%) at Tier 2 — above the typical range of Table 5, because generator hand-over substitutes for a large share of demand without affecting computing. R1 is met by 60 MW within milliseconds (margin 20 MW); R2 and R3 by 120 MW of generation from about two minutes (margin 20%), sustained within fuel and permit limits, with the battery bridging the hand-over. Workload pause is not committed: it is available only where the owner runs the workload and would touch the service. Figure 9 shows the response over time.

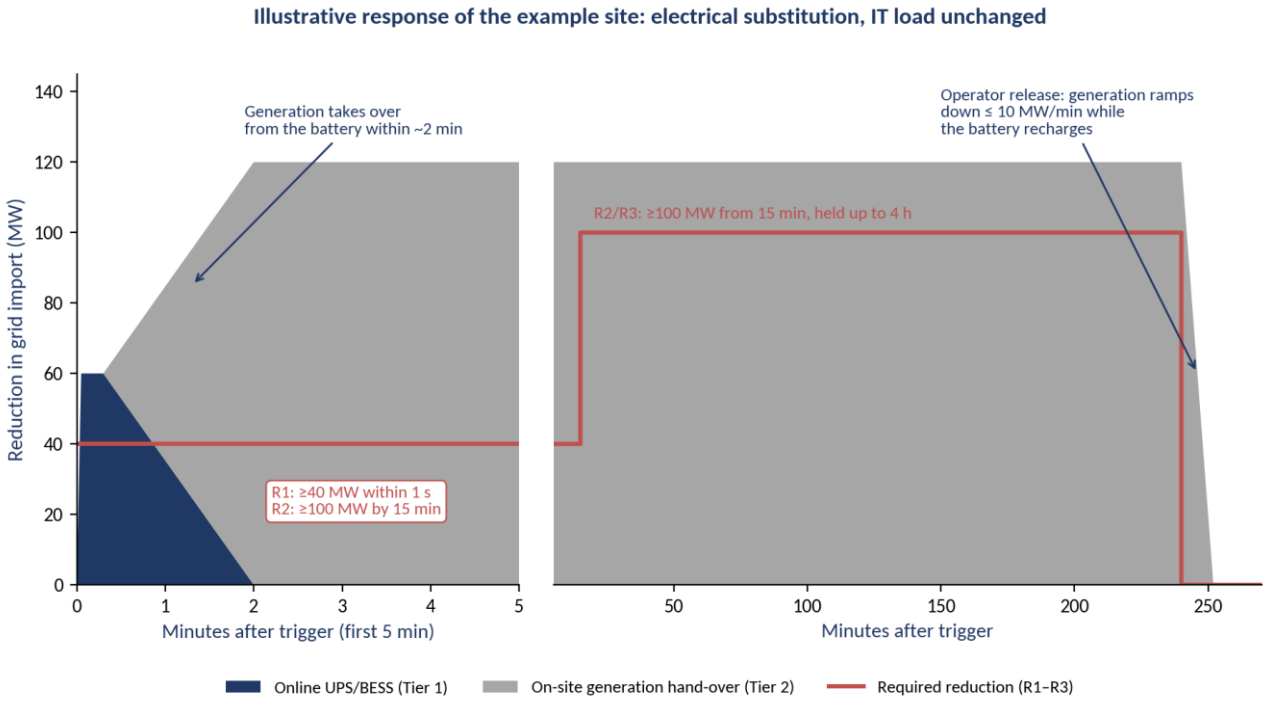


Figure 9 *Illustrative response of the example site to a permanent circuit outage at peak load. The online UPS/BESS responds within milliseconds and bridges the hand-over to on-site generation, which carries the reduction for as long as required; on release, generation ramps down gradually while the battery recharges. Conceptual illustration; values are indicative.*

B.4 Declared Constraints

- **Availability state.** The Tier 1 commitment holds only while the committed 30 MWh is available above the ride-through reserve that the UPS needs for its primary role of bridging to generation. The Tier 2 commitment holds while fuel stock covers at least the declared duration and the permit for non-emergency operation is valid. Loss of either condition is signalled by telemetry.
- **Recovery.** Return is on operator release only, at no more than 10 MW per minute. Generation is ramped down gradually and kept partly in operation while the battery recharges, so that recharge does not raise grid import above the contracted 300 MW.
- **Failure modes.** The dominant failure of Tier 1 is the battery and its trigger interface; the dominant failure of Tier 2 is a common-mode failure of the generator fleet — shared control and synchronisation system, fuel supply or switchgear — rather than individual set failures, which are covered automatically by the 16 uncommitted sets.

B.5 Reliability Check

The Decision Layer can now test whether the arrangement keeps risk to priority load within an agreed threshold, using the declared figures. Of 0.5 circuit trips per year, about half occur when curtailment is actually needed: roughly 0.2 short events per year, covered by the battery for seconds, and 0.05 long events per year — about one in twenty years — requiring generation for up to four hours. If the response fails in either case, the remaining circuit is assumed to trip on overload, interrupting the area's 700 MW of demand for an assumed restoration time of one hour: C = 700 MW × 1 h = 700 MWh. For illustration purposes, an interruption consequence of 700 MW for one hour is assumed. Actual studies should use corridor-specific restoration times and consequence distributions.

The calculation can be reproduced from the stated assumptions. Illustrative probabilities of failure on demand (PFD) are: trigger chain 0.001; UPS/BESS 0.01; common-mode failure of the generator fleet 0.002 (individual set failures are covered by the uncommitted sets). Consequence of response failure C = 700 MW × 1 h = 700 MWh.

Step	Expression	Value
(1) Binding short events	$f_s = 0.5 \times 0.8 \times 0.5$	0.20 per year
(2) Binding long events	$f_l = 0.5 \times 0.2 \times 0.5$	0.05 per year
(3) R1 failure	$P_{R1} \approx P_{trig} + P_{BESS} = 0.001 + 0.01$	≈ 0.011
(4) R3 failure	$P_{R3} \approx P_{gen,cm}$	≈ 0.002
(5) Added EENS, qualified	$\Delta EENS = [f_s \times P_{R1} + f_l \times (P_{R1} + P_{R3})] \times C$	≈ 2.0 MWh/yr
(6) Added EENS, unverified	Same, with $P_{trig} = 0.05$, $P_{BESS} = 0.05$ (ride-through reserve not verified) and $P_{gen,cm} = 0.05$ (islanding and synchronisation untested): $P_{R1} \approx 0.10$, $P_{R3} \approx 0.05$	≈ 19 MWh/yr

Table B3 *Reproducible calculation chain for the reliability check (illustrative; first-order approximation, rare-event terms neglected).*

Three qualifications apply. First, the result illustrates how measured reliability changes the risk calculation; it is not evidence of a specific factor, and in practice all PFD values would come from component data, commissioning tests and observed performance — for UPS and generators, from an extensive existing record. Second, independence is assumed except where a dependency is represented (the generator fleet's common-mode term); a real study needs dependency modelling of shared control, communication, auxiliary supply and fuel. Third, trips and high demand are assumed independent in time; in Nordic systems storms and icing coincide with winter peak, raising the frequency of binding events — exactly the weather-driven correlation PRA captures and deterministic N-1 does not.

Quantity	Qualified and verified	Unverified (assumed)
Trigger chain, probability of failure on demand	0.001	0.05
R1 fails (trigger or battery)	≈ 0.011	≈ 0.10
R3 fails (common-mode failure of generation)	≈ 0.002	≈ 0.05
Added expected energy not supplied to priority load	≈ 2 MWh/yr	≈ 19 MWh/yr
Expected grid energy replaced by on-site generation	≈ 24 MWh/yr	—
Effect on the computing service	None — IT load unchanged	—
Expected generator running time for grid events	≈ 0.2 h/yr (plus routine testing)	—

Table B4 *Indicative reliability check for the example (authors' interpretation; probabilities illustrative).*

The comparison is the point. With a verified trigger chain, ride-through reserve and islanding, the conditional connection adds about 2 MWh/yr of expected energy not supplied to priority load — a figure the Decision Layer can compare against an agreed threshold. With the same assets unverified, the added risk is roughly ten times higher. The physical capability is identical; what differs is whether it has been qualified (Section 9.3).

B.6 Result

Outcome	Firm-only connection	Qualified Virtual N-1 connection
Connected now	200 MW	300 MW (200 MW firm + 100 MW conditional)
Remaining 100 MW	After reinforcement, 7–10 years	Connected immediately, conditional
Firmness class (Table 8)	Firm	Firm for 200 MW; Conditional — fast for 100 MW

Outcome	Firm-only connection	Qualified Virtual N-1 connection
Corridor peak served under N-1 security for priority load	600 MW	700 MW
Added risk to priority load	—	≈ 2 MWh/yr EENS, verified
Cost to the campus	Delay of one third of capacity	≈ 0.2 h/yr of own generation; certification; no effect on computing

Table B5 *Connection outcome for the example site (illustrative).*

The purpose of this example is not to predict absolute EENS values, but to demonstrate how qualification enables explicit representation of response reliability within PRA and how this can be translated into admissible conditional capacity.

At this node the conditional connection adds 100 MW, or about 17% of the corridor's firm N-1 capacity — within the +10–20% range of White Paper 2. This is a node-level result on a deliberately binding corridor and should not be compared directly with the system-level ranges of Section 8, still less added to them. The use of on-site generation for grid support, even at a few hours per year, requires the appropriate environmental permit for non-emergency operation and should be planned with low-carbon fuels where available.

Worked-example result: +17% conditional capacity at this node (100 MW against 600 MW firm N-1 capacity) — consistent with, but not a validation of, the indicative +10–20% system-level range of White Paper 2.

The example turns the framework into a certificate: "Tier 1: 60 MW within 4 ms for ≥ 30 min, above the ride-through reserve; Tier 2: 120 MW within 2 min, sustained within fuel and permit limits; fault-event trigger; release by operator at ≤ 10 MW/min with generator-supported recharge; declared failure probabilities verified annually." That statement — not the nameplate of the site — is what allows the system operator to connect 300 MW today, without touching the computing service.