

A TECHNICAL HANDBOOK

The Container Data Center

A Practical Handbook for Design, Delivery and Operation of Modular Infrastructure

Design · Power · Cooling · Delivery · Operations · Economics

First Edition

The Container Data Center

A Practical Handbook for Design, Delivery and Operation of Modular Infrastructure

First edition, 2026.

This handbook is a general engineering and commercial reference for containerised, modular and prefabricated data center infrastructure. It is written to be read by the people who specify, buy, build, commission and operate these units.

Important. Nothing in this document constitutes a design, a certification, or engineering advice for any specific installation. Data center infrastructure involves high-energy electrical systems, pressurised mechanical systems, stored chemical energy and life-safety systems. All designs must be prepared, checked and approved by suitably qualified engineers and must comply with the codes, standards and permitting requirements in force at the installation site. Standards cited here are indicative; always confirm the current edition and the version adopted in your jurisdiction. Figures, ranges and durations given in this book are typical industry values intended for planning and comparison, not quotations or performance guarantees.

Prepared for Anomny — AI | IOT | Datacenter

Contents

Contents 3
How to Use This Book 7
1. The Container Data Center Defined 9
1.1 A working definition 9
1.2 What is inside the box 9
1.3 What a container data center is not 10
1.4 A short history 10
1.5 Where containers make sense 11
2. The Business Case for Modular 12
2.1 Four sources of value 12
2.2 The schedule argument, honestly stated 12
2.3 Capital phasing and the cost of being early 13
2.4 Risk transfer and where it stops 13
2.5 A worked comparison 14
3. Form Factors, Configurations and Standards 15
3.1 Physical envelopes 15
3.2 Functional configurations 15
3.3 Density classes 16
3.4 Standards and certification landscape 16
4. Power: From Utility to Rack 19
4.1 The chain 19
4.2 Redundancy topologies 19
4.3 UPS selection 20
4.4 Energy storage 21
4.5 Standby generation 21
4.6 Distribution inside the container 22
5. Cooling: Air, Liquid and Everything Between 23
5.1 Why the container is different 23
5.2 Air cooling architectures 23
5.3 Liquid cooling 23
5.4 Heat rejection 24

5.5 Containment and airflow management 25
5.6 Temperature and humidity setpoints 25
6. The Enclosure: Structure, Sealing and Environment 26
6.1 Structural design 26
6.2 Envelope and thermal performance 26
6.3 Ingress protection and corrosion 26
6.4 Access, ergonomics and serviceability 27
7. Network Architecture and Cabling 29
7.1 Internal topology 29
7.2 The boundary 29
7.3 Cabling practice 29
7.4 Latency and placement 30
8. Fire Protection, Safety and Physical Security 31
8.1 Detection 31
8.2 Suppression 31
8.3 Life safety 32
8.4 Physical security 32
9. Controls, Monitoring and DCIM 34
9.1 What to instrument 34
9.2 Architecture 34
9.3 Alarms that work 35
9.4 Remote operation 35
10. Site Selection and Civil Works 37
10.1 Screening a site 37
10.2 Foundations and the pad 37
10.3 Site services 38
10.4 Environmental and planning 38
11. Factory Build, Testing and Logistics 40
11.1 The build sequence 40
11.2 Factory acceptance testing 40
11.3 Transport 41
11.4 Arrival and installation 42
12. Commissioning 43

12.1 The five levels 43
12.2 Integrated systems testing 43
12.3 Acceptance and handover 44
13. Operations and Maintenance 46
13.1 Operating models 46
13.2 Maintenance schedule 46
13.3 Consumables and spares 47
13.4 Change management 47
14. Efficiency, Metrics and Sustainability 49
14.1 The metrics that matter 49
14.2 Where the overhead actually goes 49
14.3 Design levers, ranked 50
14.4 Water 50
14.5 Sustainability beyond energy 51
15. Reliability, Failure Modes and Incident Response 52
15.1 The failure modes that actually occur 52
15.2 Thermal runaway: the defining container risk 52
15.3 Incident response 53
15.4 Building a realistic availability picture 53
16. Edge Deployments 56
16.1 Why edge, specifically 56
16.2 Design implications 56
16.3 Fleet management 57
17. High-Density AI and HPC Containers 58
17.1 What changed 58
17.2 Liquid cooling in a container 58
17.3 Power for AI 59
17.4 Structural and logistics consequences 59
18. Total Cost of Ownership and Procurement 60
18.1 The full cost picture 60
18.2 Building a defensible model 60
18.3 Selecting a vendor 61
18.4 Contract terms worth fighting for 61

Appendix A. Specification Checklist 64
A.1 Capacity and density 64
A.2 Electrical 64
A.3 Mechanical 64
A.4 Enclosure 65
A.5 Fire, safety and security 65
A.6 Controls, monitoring and network 65
A.7 Commercial and delivery 66
Appendix B. Commissioning Checklist 67
Appendix C. Glossary 68
Appendix D. Standards Index 70

How to Use This Book

This handbook is organised so that it can be read end to end once, then used in pieces for the rest of a project.

- **Part I — Foundations** defines the product, sets out the business case in the terms a finance committee will test, and fixes the three decisions that everything else depends on: envelope, configuration and density.
- **Part II — Engineering the Box** covers the six systems inside the enclosure: power, cooling, structure, network, fire and security, and controls.
- **Part III — Delivering the Project** follows the unit from site selection through factory testing, transport and commissioning.
- **Part IV — Running It** addresses the fifteen years that cost more than the purchase: operations, efficiency and failure.
- **Part V — Applications and Economics** examines edge and AI deployments, and the total cost of ownership.
- **Part VI — Appendices** contains a specification checklist, a commissioning checklist, a glossary and a standards index intended to be copied into live project documents.

Three conventions are used throughout. **Boxed notes** flag a decision or a question to put to a vendor. **Tables** summarise options for comparison and are deliberately sparse, because a table that fits on a page gets used. And figures are given as ranges rather than single values, because the honest answer to most data center questions is "it depends on the site" — the ranges are there to show how much it depends.

> **One request**
>
> Wherever this book asks you to record a measured number — transfer time, minutes to the allowable temperature limit, actual battery autonomy — please record it. Those numbers are the difference between a runbook that is read and a runbook that is filed.

PART I

Foundations

What a container data center is, why organisations buy them, and the shapes they come in.

1. The Container Data Center Defined

A container data center is not a data center that happens to be in a box. It is a data center whose boundary, interfaces and delivery model have been deliberately fixed so that the majority of the engineering happens once, in a factory, rather than repeatedly, on site.

1.1 A working definition

A **container data center** — also called a modular, prefabricated or containerised data center — is a complete, self-contained computing facility built inside a transportable steel enclosure. The enclosure houses IT racks together with the power distribution, cooling, fire protection, monitoring and physical security equipment those racks require. It arrives on site as a finished, factory-tested product, is connected to a small number of external services, and begins operating.

The defining characteristic is not the steel box. It is **the interface**. A conventional data center has a fuzzy boundary: the building, the electrical room, the mechanical plant and the white space are all designed together for one site, by several parties, over many months. A container data center collapses that boundary into a short, explicit list — typically incoming power, a chilled-water or condenser-water connection (or nothing at all, if the unit rejects heat directly to air), fibre, earth, and sometimes potable water and drainage. Everything inside the line is the manufacturer's scope. Everything outside it is the site's scope. That clean split is what makes the model fast, repeatable and comparatively low-risk.

> **The one-sentence version**
>
> A container data center trades site-specific flexibility for factory repeatability, and in doing so converts a construction project into a manufactured product with a delivery date.

1.2 What is inside the box

A representative 40-foot IT container carrying 400 to 600 kW of IT load contains, in rough order of installation:

- **Structure and envelope** — an ISO-dimensioned or custom steel frame, insulated wall and roof panels, a raised or solid floor, personnel doors with panic hardware, and equipment access doors or removable panels.
- **Racks** — usually 6 to 20 enclosures of 42U to 52U, seismically braced to the floor, arranged in a single row with a service aisle, or in two rows with a contained hot or cold aisle.
- **Power chain** — incoming switchgear or a main distribution panel, surge protection, a static transfer switch or automatic transfer switch, UPS modules with batteries, and busway or rack PDU distribution to each cabinet.
- **Cooling** — in-row or perimeter air handlers, rear-door heat exchangers, or direct-to-chip liquid loops with a coolant distribution unit; plus containment, airflow management and, in air-cooled designs, a condenser or dry cooler mounted externally.

- **Fire detection and suppression** — very early smoke detection, conventional detection, and a clean-agent or inert-gas suppression system with its own cylinders and control panel.
- **Controls and monitoring** — a local PLC or building-management controller, environmental sensors, branch-circuit metering, leak detection and a gateway to the customer's monitoring platform.
- **Security** — card or biometric access control, internal and external cameras, door position sensors and tamper detection.
- **Life safety and operator provisions** — emergency lighting, oxygen or gas warning signage, emergency stop, and in many jurisdictions a means of egress that cannot be blocked by suppression discharge.

1.3 What a container data center is not

Three confusions are worth clearing up early, because each of them leads to a different and more expensive mistake.

It is not a shipping container with racks bolted in

A genuine unit is engineered as a pressure-, thermal- and vibration-managed enclosure. Second-hand shipping containers are structurally designed for stacking, not for the concentrated point loads of loaded racks, and their corrugated walls, timber floors and corner castings make sealing, insulating and cable-routing awkward. Most reputable manufacturers now build purpose-made enclosures to ISO external dimensions rather than converting used boxes, precisely because the conversion costs more than it saves once insulation, floor reinforcement and sealing are accounted for.

It is not automatically an "edge" product

Containers are used at the edge, but they are equally used to add 5 MW of capacity to a hyperscale campus in a single procurement round, to house high-density AI clusters that an existing building cannot cool, and to provide swing capacity during a refresh. Edge is a **deployment pattern**, not a synonym for containerised.

It is not intrinsically cheaper per kilowatt

On a pure capital-cost-per-kW basis, a large purpose-built data hall usually wins. The container wins on schedule, on capital phasing, on the ability to deploy where no building exists, and on the avoided cost of over-building capacity that will not be used for three years. Chapter 2 and Chapter 18 develop this properly; the point here is simply that "modular equals cheap" is the wrong mental model and it sets up disappointing business cases.

1.4 A short history

The idea of putting compute in a transportable enclosure is older than the modern cloud. Military and broadcast organisations were moving shelterised electronics in the 1960s and 1970s. The commercial inflection came in the mid-2000s, when large internet operators demonstrated that a container full of servers could be treated as a single unit of capacity — provisioned, monitored and eventually retired as a whole rather than server by server. Commercial vendors followed with productised units, and the

market split into three broad segments that still exist today: micro units of one to four racks for edge and telecom sites; single-container IT modules of roughly 100 to 600 kW; and multi-module campuses where separate power, cooling and IT containers are assembled into what is, functionally, a conventional data center built from parts that were manufactured elsewhere.

Two developments since 2020 have reshaped the segment. The first is **rack density**: AI training and inference hardware has pushed per-rack power from a historical 5 to 10 kW into the 40 to 130 kW range, which makes air cooling marginal and liquid cooling mandatory at the top end. The second is **grid constraint**: in many markets the limiting factor on new capacity is not land or money but the date on which a utility can deliver a connection. Both favour a model where the IT infrastructure is ready before the site is, and where capacity can be added in increments that match whatever power is actually available.

1.5 Where containers make sense

Scenario	Why the container fits	Main watch-out
Grid-constrained expansion	Capacity added in increments that match utility releases; no building permit for a data hall	Utility connection lead time still dominates the schedule
Edge and remote sites	No local construction skill required; unmanned operation; standard unit replicated across many sites	Service logistics, spares and response times at distance
AI / HPC clusters	Purpose-built for liquid cooling and high density that existing buildings cannot support	Heat rejection, water availability and structural loading
Disaster recovery and swing capacity	Deployed quickly, relocated later, financed as equipment rather than property	Long-term siting approvals if "temporary" becomes permanent
Industrial and mining sites	Sealed, filtered and vibration-tolerant enclosure in hostile environments	Corrosion, dust ingress class, and ambient design extremes
Research, defence and mobile operations	Genuinely transportable; can be re-deployed with contents intact	Shock and vibration qualification; road and air transport limits

If none of those apply — if you have a building, a grid connection, in-house construction capability and a stable, predictable growth curve — a conventional fit-out is very often the better answer. Good engineering includes being willing to say so.

2. The Business Case for Modular

The container wins or loses on time, phasing and risk transfer — not on sticker price. This chapter sets out the argument in the terms a finance committee will actually test.

2.1 Four sources of value

Every credible modular business case draws on some combination of four effects. A proposal that rests on only one of them is usually fragile.

1. **Schedule compression.** Factory build runs in parallel with site works. Civil, electrical and utility activities that would be sequential in a construction project overlap with manufacturing, typically taking a project from 18–30 months to 4–9 months from order to live load.
2. **Capital phasing.** Capacity is bought in 200 kW to 1 MW increments as demand materialises, rather than committing to a full data hall on day one. In present-value terms, deferring capital by two or three years is frequently worth more than the per-kW premium the modular unit carries.
3. **Risk transfer.** A single supplier carries design, integration and factory-test responsibility up to a defined interface. Interface failures — the classic cause of commissioning overruns — are discovered in a factory where they cost days, not on a site where they cost months.
4. **Optionality.** The asset is movable and, in many jurisdictions, treated as equipment rather than real property. It can be relocated, sold, re-deployed, or leased. Conventional white space cannot.

2.2 The schedule argument, honestly stated

Modular marketing likes to compare a 12-week factory build against a 24-month construction programme. This is not an honest comparison, because it omits the activities that are on the critical path regardless of build method.

Activity	Conventional build	Container deployment	On critical path for both?
Utility power connection	6–36 months	6–36 months	Yes — usually the true constraint
Permits and planning	3–12 months	1–6 months	Yes, but scope is smaller
Site civil works (pad, ducts, bunds)	4–9 months	6–14 weeks	Yes
Building shell and fit-out	9–18 months	Not applicable	Conventional only
Equipment manufacture	12–30 weeks (separate items)	10–20 weeks (integrated unit)	Yes, but parallel to site

Activity	Conventional build	Container deployment	On critical path for both?
Integration and commissioning	8–16 weeks on site	2–4 weeks on site after FAT	Yes

The realistic claim is this: **modular removes the building from the critical path and shortens the on-site integration tail.** It does not shorten the utility connection, and any business case that assumes it does will miss its date. Where the grid connection is long, the right use of the schedule advantage is to arrive on site fully tested and simply wait — which is a far cheaper way to wait than a half-built building with a contractor on standby.

> **The question to ask the vendor**
>
> "What is on your critical path that is not on mine?" A good vendor answers with specifics: switchgear lead time, chiller lead time, battery cell availability, and the factory slot. A vague answer is a schedule risk you will inherit.

2.3 Capital phasing and the cost of being early

The most under-appreciated cost in data center economics is capacity built and not used. A 5 MW hall commissioned three years before the load arrives carries three years of depreciation, maintenance, insurance, and — critically — standby losses on oversized UPS and cooling plant operating far below their efficient load band.

Modular deployment addresses this directly. Capacity is committed in blocks sized to a demand step. The financial effect is twofold: the capital outlay is deferred, and the installed plant runs closer to its design point, which improves both efficiency and equipment life. Against this, the per-kilowatt price of a modular block is typically **10 to 30 percent above** the marginal cost of adding the same kilowatt inside an already-built hall. Whether phasing beats the premium depends almost entirely on the shape and certainty of the demand curve — steep and certain favours the big hall, shallow or uncertain favours modules.

2.4 Risk transfer and where it stops

The container model transfers integration risk to a manufacturer, which is genuinely valuable. But the transfer stops at the interface, and the interface is where projects still go wrong. Responsibilities that remain squarely with the owner include:

- Utility capacity, connection agreement, protection coordination and metering.
- Ground conditions, foundation design, flood and seismic assessment.
- Heat rejection to the environment, including water supply, quality and discharge consent where evaporative cooling is used.
- Permits, planning consent, noise limits and any environmental impact assessment.

- Network access — fibre route, diversity, and the entry point into the unit.
- Site security, perimeter, access roads and crane or transport access.
- Ongoing operations, unless a managed-service agreement is bought alongside the hardware.

A defensible business case prices all of the above explicitly. The most common cause of modular projects landing 30 to 50 percent above budget is that only the container was priced and the site was assumed to be simple.

2.5 A worked comparison

The figures below are illustrative, not quotations, and are intended to show the **structure** of the comparison rather than any particular market price. Treat them as a template to populate with your own vendor and contractor numbers.

Line item	Conventional 2 MW hall	Four 500 kW containers
Capacity committed at day one	2,000 kW	500 kW, with three options
Time to first live load	22 months	7 months
Capital at day one	Full hall	One module plus site for four
Capital per kW installed	Lower	10–30% higher
Site works scope	Full building, extensive	Pad, ducts, services only
Utility connection	Same for both — usually the binding constraint	Same for both
Stranded capacity if demand halves	High	Low — later modules simply not ordered
Residual value / relocatability	Low, tied to the site	Material — units can be moved or sold
Design freedom (density, topology, future tech)	High	Constrained by the product platform

Two conclusions tend to hold. First, the more uncertain the demand, the stronger the modular case. Second, the longer the grid connection, the less the build method matters to the overall date — which is an argument for spending the waiting period on a factory floor rather than on a construction site.

3. Form Factors, Configurations and Standards

Before specifying anything, fix three decisions: the physical envelope, the functional split between modules, and the density class. Everything downstream follows from them.

3.1 Physical envelopes

Most units follow ISO freight dimensions because the global transport, lifting and handling ecosystem is built around them. Departing from ISO dimensions is possible and sometimes correct, but it turns a routine shipment into a specialist one.

Envelope	External size (approx.)	Typical rack count	Typical IT load	Common use
10 ft / micro	3.0 × 2.4 × 2.6 m	1–3	5–40 kW	Telecom, retail, kerbside edge
20 ft standard	6.1 × 2.4 × 2.6 m	4–8	40–200 kW	Edge, DR, small enterprise
40 ft standard	12.2 × 2.4 × 2.6 m	10–20	150–600 kW	Mainstream IT module
40 ft high cube	12.2 × 2.4 × 2.9 m	10–20	200–800 kW	High density, liquid cooled
45–53 ft / custom	13.7–16.2 m long	18–28	400 kW–1.5 MW	AI clusters, campus modules
Wide-body / skid	Non-ISO, site-assembled	20–40+	1–3 MW	Multi-module campuses

Note the practical ceilings. Road transport in most jurisdictions becomes an abnormal-load exercise beyond roughly 2.9 metres wide or 4.2 metres total travelling height, and beyond 30 to 40 tonnes gross. High-cube units on a standard trailer already approach the height limit under many bridges. These are not abstract limits — they determine the route survey, the permits, the escort requirements and sometimes the entire feasibility of a site.

3.2 Functional configurations

All-in-one

IT, power and cooling in a single enclosure. Simplest to procure, transport and commission; requires only power, fibre and earth. The compromise is space: plant competes with racks for volume, capping density and leaving little room for maintenance access or future change. Best for edge and for loads below roughly 200 kW.

Split — IT module plus plant module

One container holds racks; a second holds UPS, switchgear, batteries and cooling plant. This is the dominant configuration above 300 kW. It raises IT density per container, keeps noisy and heat-producing plant away from the white space, and allows power and cooling to be sized or upgraded independently of IT. It costs more site area, an extra foundation, and interconnecting services that must be designed, protected and commissioned.

Multi-module campus

Several IT modules served by shared power and cooling blocks, often connected by a covered corridor or vestibule so staff can move between modules without going outside. Functionally this is a data center that was manufactured rather than constructed. It brings the scale efficiencies of a conventional facility while retaining phased delivery, but it also reintroduces genuine campus-level engineering: shared distribution, selectivity studies, common controls and a proper master plan.

> **Design rule**
>
> Choose the configuration from the **density and growth plan**, not from the initial load. Retrofitting a split-plant arrangement onto an all-in-one deployment is close to a rebuild. Leaving space and a service route for a future plant module costs almost nothing today.

3.3 Density classes

Density drives the cooling architecture, and the cooling architecture drives nearly everything else — floor loading, water requirements, service access, noise and the resale market for the unit. It is useful to think in four classes:

Class	Per-rack load	Typical cooling approach	Water needed?
Standard	3–10 kW	Perimeter or in-row air, hot-aisle containment	No (DX) or optional
Dense	10–25 kW	Close-coupled in-row air, strict containment	Often chilled water
High	25–60 kW	Rear-door heat exchangers or hybrid air/liquid	Yes
Extreme	60–130+ kW	Direct-to-chip liquid, CDU, sometimes immersion	Yes, with treated loop

A unit designed for the Standard class cannot usually be upgraded into the Extreme class. The floor loading, the pipework routes, the electrical distribution capacity and the heat rejection equipment are all different. If there is any prospect of hosting accelerated computing within the asset's life, specify at least the **structural and hydraulic provisions** for liquid cooling now, even if the initial fit-out is air cooled. Chapter 17 covers this in detail.

3.4 Standards and certification landscape

No single standard governs container data centers end to end. In practice a unit is assessed against a stack of standards belonging to different disciplines, and the supplier's job is to show coherent compliance across all of them.

Domain	Typical references	What it governs
Enclosure and transport	ISO 668, ISO 1496, CSC plate	Dimensions, corner fittings, structural test, transportability
Ingress protection	IEC 60529 (IP), NEMA 250	Dust and water tightness of enclosure and external plant
Electrical safety and installation	IEC 60364, NFPA 70 (NEC), local wiring rules	Installation, protection, earthing, separation
Equipment safety	IEC/UL 62368-1, IEC 61439 (assemblies)	IT and switchgear assembly safety
Thermal environment	ASHRAE TC 9.9 classes A1–A4, ASHRAE 90.4	Allowable inlet conditions and energy performance
Data center design practice	TIA-942, EN 50600 series, Uptime Institute Tier	Topology, redundancy, availability classification
Fire	NFPA 75, NFPA 76, NFPA 2001, ISO 14520, EN 15004	Detection, clean-agent suppression, IT fire protection
EMC	IEC 61000 series, FCC Part 15, EN 55032/35	Emissions and immunity
Seismic and wind	IBC, ASCE 7, Eurocode 8, local codes	Anchorage, bracing, structural qualification
Efficiency and reporting	ISO/IEC 30134 (PUE, WUE, REF and related)	Metric definition and measurement method

Three practical cautions. First, **a certificate on a component is not a certificate on the system** — a listed UPS inside a non-compliant assembly buys you nothing at inspection. Second, classification schemes such as Uptime Institute Tier apply to the whole site, not to the container alone; a 2N container fed by a single utility supply and one generator is not a Tier III facility. Third, the authority having jurisdiction at your site always wins, and its interpretation of whether the unit is a building, a piece of equipment or a temporary structure can change the entire compliance path. Establish that interpretation in writing before manufacturing starts, not after delivery.

PART II

Engineering the Box

Power, cooling, enclosure, network, fire and controls — the six systems that determine whether the unit works.

4. Power: From Utility to Rack

Power is the system most likely to constrain the project and most likely to cause an outage. It is also the one where container design differs least from conventional practice — which makes it a good place to insist on conventional rigour.

4.1 The chain

Trace the path a watt takes and you have the design. From the utility transformer it passes through the site main switchboard, a transfer arrangement that selects between utility and generator, a distribution board inside the container, the UPS, a downstream distribution board, then busway or cabling to the rack PDU, and finally into the server power supply. Every element in that chain has an availability, an efficiency, a maintenance requirement and a failure mode. A design that has not been walked element by element is not a design.

Stage	Typical equipment	Key design decisions
Utility entry	MV transformer, ring main unit, metering	Capacity, fault level, protection coordination, ownership boundary
Standby generation	Diesel or gas genset, fuel store, day tank	Sizing with step load, run time, emissions consent, fuel logistics
Transfer	ATS, STS, or closed-transition switchgear	Transfer time, tolerance of IT load, open vs closed transition
Conditioning	UPS modules, batteries or flywheel	Topology, redundancy, autonomy, battery chemistry
Distribution	Panelboards, busway, branch metering	Selectivity, metering granularity, future capacity
Rack delivery	Rack PDUs, A/B feeds, cord management	Phase balance, outlet types, monitoring level
Earthing and bonding	Earth bar, electrode, equipotential bonding	Enclosure bonding, transport isolation, lightning protection

4.2 Redundancy topologies

Redundancy is a spectrum, not a badge. Four configurations cover almost all container deployments.

- **N** — exactly enough capacity, no spare. Any failure or maintenance action drops load. Acceptable for stateless edge nodes where the application tolerates node loss, and genuinely wrong anywhere else.

- **N+1** — one redundant unit across the set. The workhorse configuration. Tolerates a single module failure or a planned maintenance window without load loss, provided the distribution downstream is also resilient.
- **2N** — two entirely independent systems, each capable of the full load, feeding dual-corded IT through A and B paths. Tolerates the loss of a whole path, including a maintenance shutdown of one side. Roughly doubles the electrical capital cost.
- **2(N+1) or block redundant** — used where a container sits inside a larger campus and shares redundant capacity blocks with neighbouring modules. Efficient at scale, but the sharing arrangement must be designed campus-wide.

The critical and frequently-missed point is that **redundancy is only as good as its weakest link**. A 2N UPS arrangement that converges on a single output board, a single busway, or a single-corded switch in the top of the rack is an N system wearing an expensive costume. Trace both paths on the single-line diagram from utility to server power supply and confirm there is no common element. Where a single element is unavoidable — very often the incoming supply or the earth system — record it explicitly as a known single point of failure rather than letting it hide inside a topology label.

> **Single point of failure audit**
>
> Take the single-line diagram, and for every component ask: "if this fails closed, open, or on fire, what stops?" Mark each answer on the drawing. The components whose answer is "everything" are your real availability figure — not the topology name on the cover sheet.

4.3 UPS selection

Three architectures dominate.

Double-conversion (online)

Rectifier, DC bus, inverter, always in the path. Complete isolation of the load from mains disturbances and the cleanest output. Historically the efficiency penalty was significant; modern units reach roughly 96 to 97 percent in double conversion, narrowing the gap considerably. This is the default for critical loads and the safe answer where the incoming supply is poor.

Line-interactive and eco-mode

The load is fed from conditioned mains, with the inverter on standby, typically reaching 98 to 99 percent efficiency. The trade is transfer time on a mains event and reduced tolerance of unstable supplies. Modern "advanced eco" implementations transfer in well under half a cycle and are widely deployed, but they must be validated against the specific IT load and the site's supply quality — not assumed.

Modular / scalable

Hot-swappable power modules in a common frame. Redundancy is achieved by populating N+1 modules, capacity is added by inserting more, and mean time to repair falls dramatically because a failed module is swapped rather than repaired in place. This suits containers particularly well, because

it lets one frame follow a load that grows from 100 kW to 500 kW over several years. The counterpoint is the shared frame, backplane and static bypass, which remain common-mode elements.

4.4 Energy storage

Technology	Typical autonomy	Footprint / weight	Life	Notes
VRLA	5–15 min	Large and heavy	3–6 yrs	Cheapest capital, highest lifetime cost, thermally sensitive
Lithium-ion (LFP)	5–15 min	50–70% smaller and lighter	10–15 yrs	Now the default; needs BMS, fire strategy and transport compliance
Nickel-zinc / advanced	5–10 min	Compact	8–12 yrs	Good thermal tolerance; narrower supply base
Flywheel	15–60 s	Compact, very heavy point load	15–20 yrs	No chemistry, no thermal runaway; bridges to generator only
Supercapacitor	10–30 s	Compact	10–15 yrs	Ride-through for transfer events, not for outages

In a container, battery choice is unusually consequential. Weight matters because it is concentrated on a small floor area and because the unit must be transported. Volume matters because it competes directly with racks. Thermal tolerance matters because the battery compartment may be exposed to a wider ambient range than a building plant room. And fire strategy matters because lithium chemistries introduce thermal-runaway scenarios that a clean-agent system will not extinguish — a separately-compartmented, ventilated battery space with dedicated detection is the common answer, and in several jurisdictions the mandatory one.

Autonomy sizing deserves a sentence of its own. Battery runtime exists to bridge to the generator, not to ride out an outage. If there is a generator, five to ten minutes is usually ample and anything more is weight and cost. If there is no generator, the honest question is not "how many minutes of battery" but "what is the orderly shutdown sequence, who triggers it, and has it been tested?"

4.5 Standby generation

Generators are usually outside the container scope, but they belong in the design conversation from day one. The recurring errors:

- **Sizing on running load rather than step load.** A UPS with a discharged battery draws heavily on recharge, and the cooling plant restarts simultaneously. The generator must accept the block load without unacceptable frequency and voltage excursion. Run the step-load study.

- **Non-linear load and alternator sizing.** Rectifier loads distort current; alternators are commonly oversized relative to the engine for this reason.
- **Fuel logistics.** Twenty-four hours of autonomy means a defined refuelling contract with a defined response time. Without that contract the tank size is decoration.
- **Emissions and noise consent.** Local limits on stack emissions, test-running hours and night-time noise can constrain siting more than any technical factor.
- **Fuel degradation.** Diesel stored for months requires polishing, biocide treatment and periodic testing. Load-bank testing verifies the whole chain, not just the engine.

4.6 Distribution inside the container

Two approaches compete. **Overhead busway** offers tap-off boxes that can be added or moved live, supports phased rack deployment, keeps the floor clear, and makes future changes cheap; it costs more initially and requires overhead space that competes with cable tray and pipework. **Hard-wired panelboards and cabling** are cheaper and simpler but every change is an electrical works package.

For any unit expected to be reconfigured during its life — which is most of them — busway usually pays for itself on the second change. Whichever is chosen, specify **branch-circuit monitoring at the tap or breaker level**. Without per-circuit measurement you cannot verify phase balance, cannot capacity-plan, cannot bill a tenant, and cannot diagnose a trip. It is inexpensive at build and near-impossible to retrofit neatly.

Finally, phase balance. Single-phase rack loads distributed carelessly across three phases produce neutral currents and derate the whole system. Plan the phase assignment of each rack at design stage and monitor the imbalance continuously; a five-percent imbalance is fine, a twenty-percent imbalance is a problem waiting for a hot day.

5. Cooling: Air, Liquid and Everything Between

Cooling is where container design genuinely diverges from building practice. The volume is small, the heat flux is high, and there is nowhere for a mistake to hide.

5.1 Why the container is different

A conventional data hall has thermal mass and volume. If a cooling unit fails, the air in the room buys minutes. A container has neither. At 400 kW in a 40-foot enclosure, the free air volume may represent **well under a minute** of thermal buffer before inlet temperatures exceed the ASHRAE allowable range. Three consequences follow, and they should shape every cooling decision:

1. Cooling redundancy matters more than in a building, and the changeover must be fast and automatic. Manual intervention is not a strategy.
2. Airflow management is not an efficiency measure, it is an availability measure. Bypass and recirculation that would be merely wasteful in a hall become thermal failures in a box.
3. The control system must be able to shed or throttle load, and the shutdown sequence must be tested. "The servers will just get hot" is not a plan.

5.2 Air cooling architectures

Perimeter cooling

Wall-mounted units pushing air along the enclosure. Simple and cheap; airflow is hard to control over distance and effectiveness falls away above about 8 kW per rack. Acceptable for micro units, rarely the right answer for a 40-foot module.

In-row / close-coupled

Cooling units sit between racks, drawing hot air from the hot aisle and discharging cold air directly into the cold aisle. The short air path makes this efficient and controllable, and it scales comfortably to 20 to 30 kW per rack with good containment. It is the mainstream choice for container IT modules. The cost is rack space — each in-row unit occupies a rack position.

Rear-door heat exchangers

A coil replaces the rack's rear door and removes heat at source; passive versions rely on server fans, active versions add their own. The room becomes, in effect, thermally neutral, and per-rack loads of 40 to 60 kW are achievable. Rear doors require chilled water piped to every cabinet inside the enclosure, which means leak detection, drip trays, careful routing away from electrical distribution, and a service procedure for isolating a single door.

5.3 Liquid cooling

Above roughly 50 kW per rack, air stops being economic and eventually stops being possible. Three liquid approaches are in use.

Approach	How it works	Heat captured to liquid	Main constraints
Direct-to-chip (single phase)	Cold plates on CPU/GPU, coolant loop via CDU	70–85%	Residual air cooling still needed; fluid quality and leak management
Direct-to-chip (two phase)	Phase-change coolant at the cold plate	85–95%	Fluid cost, containment, refrigerant handling rules
Immersion (single phase)	Servers submerged in dielectric fluid	~95–100%	Tank handling, serviceability, fluid volume and weight
Immersion (two phase)	Boiling dielectric with condenser	~100%	Fluid availability and regulatory scrutiny of certain chemistries

For containers, direct-to-chip is the dominant choice today. It uses conventional rack form factors, keeps the servers serviceable by ordinary technicians, and works with the supply chain the industry already has. Its most important design implication is that **it does not remove all the heat**. Memory, drives, network interfaces and power conversion still reject 15 to 30 percent of rack power into the air, so an air-cooling system is still required — smaller, but not absent. Designs that forget this discover it during commissioning, when the enclosure warms up steadily with the liquid loop running perfectly.

Immersion is thermally excellent and increasingly practical, but in a container it raises hard questions about tank weight on the floor, fluid volume during transport, the service workflow when a server must be lifted and drained, and the fire and environmental rules that apply to the fluid. It is a deliberate choice, not a default.

5.4 Heat rejection

Whatever happens inside, the heat must reach the environment. The options carry very different water, power and noise profiles.

Method	Water use	Relative efficiency	Notes
Air-cooled DX / condenser	None	Lowest	Simplest; performance and capacity fall in high ambient
Dry cooler with glycol	None	Moderate	Good in cool climates; large surface area needed
Adiabatic / evaporative assist	Moderate, peaks only	Good	Water quality and treatment; seasonal consumption
Cooling tower + chiller	High	Good at scale	Plant room needed; water treatment and legionella control
Free cooling / economiser	None or low	Best where climate allows	Hours of free cooling are entirely site-dependent

Method	Water use	Relative efficiency	Notes
Heat reuse to district or process	Varies	Best overall system efficiency	Needs a willing, co-located heat consumer

Two site facts govern this choice more than any preference: the **design ambient** (use the local 0.4 percent design condition and check the wet-bulb, not just the dry-bulb) and **water availability and discharge consent**. An adiabatic design specified for a temperate climate and shipped to a hot, arid, water-restricted site will disappoint in exactly the week it matters most.

5.5 Containment and airflow management

In a container, containment is mandatory. Without it, hot exhaust recirculates to server inlets within a metre, and the cooling system fights itself. The essentials are unglamorous and cheap:

- Hot-aisle or cold-aisle containment with sealed ends and a sealed top, including at the junctions with the enclosure walls.
- Blanking panels in every unused rack U. A single open U can recirculate a surprising fraction of a rack's airflow.
- Brush grommets or sealed plates at every cable penetration, floor and rack alike.
- Sealed gaps between racks and between racks and the enclosure structure.
- Airflow-matched fan control so that supply airflow tracks IT demand rather than running flat out.

Measure the result. A simple **return temperature index** and a check that the differential pressure across containment stays slightly positive on the cold side will reveal bypass long before a thermal alarm does.

5.6 Temperature and humidity setpoints

ASHRAE TC 9.9 recommends a supply envelope of roughly 18 to 27 °C for general IT, with allowable classes (A1 through A4) extending considerably wider — class A3 to 40 °C and A4 to 45 °C for equipment rated accordingly. Raising setpoints increases free-cooling hours and reduces energy, but it also reduces thermal buffer and increases server fan power, which at some point cancels the saving. In a container, where buffer is already minimal, running at the very top of the allowable band is a real availability trade, not a free efficiency win.

On humidity: the modern guidance is much more permissive than legacy practice, with a wide allowable band bounded by a dew-point range and an upper relative-humidity limit to manage condensation and contamination risk. Tight humidity control is expensive, and in a well-sealed container with no significant infiltration it is usually unnecessary. What **is** necessary is condensation control wherever chilled surfaces exist — pipework insulation, drip trays and leak detection under every wet component.

6. The Enclosure: Structure, Sealing and Environment

The enclosure is the part of the system customers inspect least and regret most. It determines what the unit tolerates for the next fifteen years.

6.1 Structural design

The enclosure must survive four quite different load cases, and the governing case is often transport rather than service.

- **Static service load** — racks, plant, batteries and people. Fully loaded racks concentrate significant mass on small footprints; floor design is a point-load problem, not a uniformly-distributed-load problem.
- **Transport dynamics** — road, rail and sea impose sustained vibration and shock in all three axes. Sea transport adds racking and torsional loads from vessel motion.
- **Lifting** — corner-casting lifts impose loads entirely different from a forklift or a low-loader transfer, and both must be designed for and marked.
- **Environmental** — wind on a tall, light, high-sail-area object; snow on the roof; seismic acceleration; and in some regions flood buoyancy.

Ask for the calculations. A supplier who can produce a finite-element model, a transport shock-and-vibration assessment and a seismic anchorage design is a different proposition from one who can produce a brochure. Ask specifically whether the unit is qualified to be transported **loaded** with racks installed, and under what conditions — many are not, and the answer changes the entire logistics plan.

6.2 Envelope and thermal performance

The enclosure is a thermal boundary with a poor surface-to-volume ratio and, frequently, direct solar exposure on five surfaces. Design considerations:

- **Insulation** — typically 75 to 150 mm of rigid foam or mineral wool. Mineral wool is preferred where fire performance governs; foam gives better thermal performance per millimetre. Continuous insulation without thermal bridges matters more than raw thickness.
- **Thermal bridging** — steel members that cross the envelope conduct heat efficiently and create condensation points. Thermal breaks at frame junctions are worth specifying explicitly.
- **Solar gain** — a white or high-albedo roof and a ventilated over-roof can reduce peak solar load substantially. In hot climates this is one of the highest-return, lowest-cost decisions available.
- **Vapour control** — a continuous vapour barrier on the warm side, correctly sealed, prevents interstitial condensation inside the wall build-up. Getting this wrong causes corrosion that is invisible until it is structural.

6.3 Ingress protection and corrosion

Ingress class should be specified for the enclosure and for every external component separately. IP54 or IP55 is typical for external plant. In coastal, desert or industrial locations the specification must go further:

Environment	Primary threat	Typical mitigation
Coastal / marine	Chloride-driven corrosion	C5-M coating system, marine-grade fasteners, coated coils, sacrificial anodes
Desert / arid	Fine dust, thermal cycling, UV	IP55+, F7-class filtration, sealed cable entries, UV-stable finishes
Industrial / mining	Corrosive gases, particulates, vibration	Gas-phase filtration, positive pressurisation, corrosion coupon monitoring
Tropical / humid	Condensation, mould, biological growth	Vapour control, conformal-coated boards, continuous dew-point monitoring
Arctic / alpine	Snow load, cold starts, ice on plant	Heated intakes, low-temperature-rated plant, snow-load structure, glycol loops

Corrosion monitoring deserves a mention because it is cheap. ISA-standard corrosion coupons placed in the airstream and read periodically will tell you whether your filtration strategy is working, years before the failures start.

6.4 Access, ergonomics and serviceability

Container interiors are tight, and the aisle that looked adequate on a drawing can be unworkable with a technician, a laptop cart and a server rail extended. Practical minimums to hold the vendor to:

- At least 1,000 mm of clear front aisle; 1,200 mm is far better and allows a server to be withdrawn on its rails while a person passes.
- Rear access to every rack — either a rear aisle or racks on slides. "Pull the rack out to service it" is a legitimate design, but only if the floor, the cabling and the containment support it.
- A door at each end where the length permits, giving two means of escape and a through route for equipment.
- Doors that swing outward, with panic hardware, and that cannot be obstructed by containment, cabling or plant.
- Safe working clearance in front of every electrical panel, per the applicable code — this is a legal minimum, not a design preference, and it is a common cause of inspection failure.
- Lighting adequate for work at the rear of a rack, with emergency lighting on the egress path.
- A documented route by which the largest serviceable component — a UPS module, a CDU pump, a compressor — can actually be removed from the unit.

Walk it before you sign it

At the factory acceptance test, insist on physically doing three things: extend a server on its rails in the aisle; open and close the largest electrical panel door; and remove and replace one UPS module or cooling component. Paper clearances and real clearances differ more often than anyone expects.

7. Network Architecture and Cabling

Network design inside a container is straightforward. Network design at the container boundary is where the outages come from.

7.1 Internal topology

For a single IT module, two patterns cover most cases. **Top-of-rack** switching places one or two access switches in each cabinet, with short copper or DAC links to servers and fibre uplinks to an aggregation pair. It minimises horizontal cable, keeps the containment clear and suits phased rack deployment. **End-of-row** switching concentrates ports in one or two cabinets, improving port utilisation and simplifying management at the cost of a substantial horizontal cabling bundle that consumes containment and airflow space.

In a container, top-of-rack usually wins, for a reason specific to the form factor: bulk cabling is an airflow obstruction in a space with no room to spare. Where the density is high enough that every rack unit is contested, a middle path — shared switching between rack pairs — is a reasonable compromise.

7.2 The boundary

The most consequential network decisions are about what crosses the enclosure wall:

- **Diverse fibre entry.** Two physically separate paths, entering at opposite ends of the unit, following genuinely different routes to genuinely different termination points. Two fibres in the same duct are one fibre with a redundancy label.
- **Route resilience.** Buried ducts at adequate depth, with warning tape, marked routes and as-built records. The commonest cause of container network outages is an excavator, not a switch.
- **Sealed, rated penetrations.** Every entry must maintain ingress protection, fire rating and EMC integrity. Gland plates, not drilled holes.
- **Demarcation.** Define precisely where the carrier's responsibility ends and yours begins, and put a patch panel there. Ambiguous demarcation turns every fault into a jurisdictional argument while the service stays down.
- **Out-of-band management.** A path to the unit's controllers, PDUs, and console servers that does not depend on the production network — ideally a cellular or satellite link. When the production path fails at a remote site, this is the difference between a remote diagnosis and a site visit.

7.3 Cabling practice

- Separate power and data containment, crossing at right angles where they must cross, to limit induced noise.
- Overhead containment is preferred in containers — it keeps the floor clear for airflow, cooling and cleaning, and makes changes visible.

- Respect bend radii, especially for fibre. Cable ties pulled tight on fibre bundles cause attenuation that appears months later as intermittent errors.
- Label both ends of every cable to a documented scheme, and record it in a system that survives staff turnover.
- Seal every penetration with brush grommets or blanking plates. Containment integrity is an airflow and fire-compartment issue simultaneously.
- Leave planned slack and spare ways. Retrofitting containment in an operating container is miserable work in a live electrical environment.

7.4 Latency and placement

For edge deployments the physical location is the design. Light travels through fibre at roughly 200,000 km per second, giving about 1 millisecond of round trip per 100 km of cable route — and cable routes are rarely straight. Add switching, queuing and processing, and a realistic budget is 5 to 15 milliseconds to a site 200 km away over a good network. If an application requires single-digit millisecond response, the container must be close, and no amount of network engineering substitutes for proximity. This is precisely the requirement that justifies edge containers, and it should be stated as a measured budget in the requirements document rather than as an aspiration.

8. Fire Protection, Safety and Physical Security

A small, sealed, occupied space full of electrical energy deserves more fire-safety attention than a large one, not less. Regulators generally agree.

8.1 Detection

Layered detection is standard practice. **Aspirating smoke detection** continuously samples air through a pipe network and detects incipient combustion at very low particle concentrations — typically in the return airstream, where it sees the whole volume. **Point detectors** provide conventional confirmed detection and, in a cross-zoned arrangement, the second signal that authorises suppression release. **Thermal and thermographic monitoring** on busbars, terminations and battery cells catches the developing hot spot before there is any smoke at all, and is the only one of the three that is genuinely preventive.

In high-airflow environments — which every container is — smoke dilutes rapidly and detection sensitivity must be set accordingly. Detection placement is an airflow problem: put detectors where the air actually goes, which means the return path, not the geometric centre of the ceiling.

8.2 Suppression

Agent type	Examples	Advantages	Considerations
Chemical clean agent	FK-5-1-12, HFC-227ea	Fast, low concentration, electrically non-conductive, small cylinder volume	Global-warming and regulatory scrutiny of some agents; decomposition products
Inert gas	IG-55, IG-541, nitrogen, argon	No decomposition products, no environmental impact, occupant-tolerable design concentrations	Large cylinder banks; pressure relief venting is mandatory
Oxygen reduction	Continuous low-oxygen atmosphere	Prevents ignition rather than extinguishing; no discharge event	Continuous energy use; occupational health rules on occupancy
Water mist	High-pressure mist systems	Effective on deep-seated fires; no agent supply limit	Water in an electrical space; careful zoning and drainage needed

Whichever agent is chosen, three container-specific issues must be resolved at design stage. First, **enclosure integrity**: gaseous suppression only works if the agent stays in the space, which requires a door-fan integrity test — not an assumption. Second, **pressure relief**: discharging gas into a small, well-sealed steel box generates a substantial pressure transient that can damage the structure or blow

a door open. Correctly sized relief vents are not optional. Third, **cooling interaction**: the cooling system must shut down and dampers close on discharge, or the agent is extracted within seconds.

> **The lithium question**
>
> Clean-agent systems interrupt combustion; they do not stop thermal runaway in a lithium cell, which supplies its own oxygen and reignites. If the unit contains lithium batteries, the fire strategy must address runaway specifically: a separate fire-rated battery compartment, cell-level monitoring via the BMS, off-gas detection, dedicated ventilation, and an agreed response procedure with the local fire service. Discuss this with your insurer before ordering, not after.

8.3 Life safety

People work inside these units, sometimes alone, sometimes at night, sometimes at remote sites. The non-negotiables:

- Two means of escape wherever the length permits; outward-opening doors with panic hardware that can always be opened from inside without a key or a card.
- Pre-discharge alarm with a time delay, and a manual abort that is reachable from anywhere in the space.
- Emergency lighting and photoluminescent egress marking.
- Emergency power-off where required by code, with the control placed where it cannot be operated accidentally — inadvertent EPO is itself a well-documented cause of outages.
- Oxygen depletion warning and signage where inert gas is used; entry procedures after discharge.
- Lone-worker procedures: check-in, duress alarm, and a documented response if the check-in is missed.
- Arc-flash assessment, labelling and appropriate PPE for every electrical panel, plus lock-out/tag-out provisions on all isolators.

8.4 Physical security

A container is a high-value asset frequently placed in a low-supervision location. Security is layered:

1. **Perimeter** — fencing, anti-ram bollards or blocks, lighting, and clear sight lines. At unmanned sites, this is the layer that buys response time.
2. **Enclosure** — high-security locks, concealed or protected hinges, tamper-evident fasteners, and anti-lift provisions at the corner castings. A container on a low-loader is a theft vector no software control addresses.
3. **Access control** — card plus PIN or biometric, with per-door logging, anti-passback, and time-windowed permissions for contractors.
4. **Surveillance** — internal and external cameras with local recording that survives a network outage, plus verified alarm transmission to a monitoring centre.

5. **Detection** — door and panel position sensors, vibration and tilt sensors on the enclosure, and geolocation for mobile units.
6. **Rack level** — electronic cabinet locks with audit trails where multiple parties or tenants share a unit.

The threat model at a remote site differs from a campus: response time may be measured in tens of minutes or hours, which shifts the emphasis from detection to delay. Design to slow an intruder down long enough for a response to arrive, and assume the network link may be the first thing cut.

9. Controls, Monitoring and DCIM

The control system is what makes an unmanned container viable. It is also, routinely, the least-specified part of the purchase.

9.1 What to instrument

The rule of thumb: instrument anything whose failure you would want to know about before a user does, and anything you will later be asked to prove.

Domain	Minimum instrumentation	Worth adding
Power	Incoming kW/kWh, UPS status and load, battery state	Branch-circuit metering, power quality, harmonic analysis
Cooling	Supply/return air temp, unit status, setpoint, alarms	Per-rack inlet temps, differential pressure, fan power, valve position
Environment	Space temperature and humidity, leak detection	Dew point, particulate count, corrosion coupons, door position
Fire and safety	Panel status, detector state, suppression pressure	Aspirating detector trend data, battery off-gas, thermal imaging
Security	Door position, access events, camera health	Vibration/tilt, geolocation, rack-level lock audit
IT	Rack power, switch and host reachability	Server inlet temp from BMC, workload and utilisation correlation

9.2 Architecture

A sound control architecture for a container has four properties, and each is worth writing into the specification.

1. **Local autonomy.** The unit must protect itself with no network connection. Cooling changeover, alarm escalation, fire response and orderly shutdown all execute locally. A cloud-dependent control system is a cloud-dependent data center.
2. **Open protocols.** Modbus TCP, BACnet/IP, SNMP and Redfish are the common currencies. Proprietary-only integration is a long-term cost and a negotiating weakness at renewal.
3. **Segregated networks.** Building-management and IT networks are separated, with a controlled, documented and monitored gateway between them. Operational technology has a different patch cadence and a different threat model.
4. **Defined data ownership.** Establish who owns the telemetry, where it is stored, for how long, and whether the vendor can access it. This matters for sovereignty, for contract exit, and for security review.

9.3 Alarms that work

The most common failure of a monitoring system is not that it misses events — it is that it reports so many that operators stop reading them. Discipline to impose from day one:

- Every alarm has a defined severity, a defined recipient, and a documented action. If no action exists, it is not an alarm; it is a data point.
- Alarms are classified into **critical** (service at risk now), **major** (redundancy lost, act today) and **minor** (degradation, schedule a fix).
- Deadbands and time delays suppress chatter from parameters that naturally oscillate around a threshold.
- Related alarms are correlated so that one root cause raises one event, not forty.
- Alarm performance is reviewed monthly. Track count per shift, top ten recurring alarms, and mean acknowledgement time; a healthy system produces few alarms and each one gets read.

9.4 Remote operation

For unmanned sites the control system is the operator. Practical requirements:

- Out-of-band access over a path independent of the production network, with strong authentication and full session logging.
- Remote power cycling at the rack outlet level, and serial console access to network equipment and controllers.
- Firmware and configuration management with version control and an automated backup of every controller configuration — recovering a failed controller from an undocumented configuration is a very long day.
- A tested remote-hands procedure: who attends, how they gain access, what they are authorised to do, and how long they take. Test it annually with a real, timed exercise rather than a tabletop.
- Logging retained centrally so that a failed unit's history survives the failure.

> **Design for the worst hour**
>
> Specify the control system against a specific scenario: a summer night, the primary cooling unit has failed, the production network is down, the nearest engineer is two hours away, and the on-call operator has a phone. What does the system do on its own, and what does it tell that operator? Write the answer down. That answer is your specification.

PART III

Delivering the Project

Site, factory, transport and commissioning — turning a purchase order into operating capacity.

10. Site Selection and Civil Works

The container arrives finished. The site does not. Most modular schedule overruns and budget overruns live here.

10.1 Screening a site

Assess candidate sites against seven factors, in roughly this order of severity. The first two eliminate sites; the rest shape cost.

Factor	What to establish	Typical killer
Power	Available capacity, connection date, fault level, voltage	Connection offered three years out, or capacity refused
Connectivity	Carrier presence, diverse routes, lead time to deliver fibre	Single route, or a build cost exceeding the container
Access	Road width, turning radii, bridge heights, axle weight limits	A low bridge or a turn a 40 ft trailer cannot make
Ground	Bearing capacity, water table, contamination, flood zone	Poor ground demanding piled foundations
Consents	Planning status, permitted use, noise limits, emissions	Noise limit that prohibits the generator or dry coolers
Heat rejection	Design ambient, water supply, discharge consent	No water where the design assumed adiabatic cooling
Security	Exposure, response time, perimeter feasibility	No lawful way to fence or light the plot

Run this screen before the design is fixed, not after. The single most expensive pattern in modular projects is a unit specified for an assumed site and then delivered to the site that was actually available.

10.2 Foundations and the pad

Container foundations are usually one of four types, and the choice is a soils question, not a preference.

- **Reinforced concrete slab** — the default. Simple, durable, tolerant, and it doubles as the working platform. Requires curing time in the programme and good ground.
- **Strip or pad footings under corner castings** — less concrete and faster, provided loads genuinely concentrate at the castings and the ground is competent.
- **Piled foundations** — for poor ground, high water table or seismic requirements. Significant cost and schedule; needs geotechnical investigation early.

- **Screw piles or ground screws** — fast, reversible, low-disturbance and increasingly common for temporary or environmentally-sensitive sites. Capacity is soil-dependent and must be proof-tested.

Whatever the type, four details recur in defect lists: **level tolerance** (typically ±3 mm across the bearing points — a container racked by an out-of-level pad will not seal, and doors will bind); **drainage falls** away from the unit with no ponding under it; **service entry ducts** cast in and correctly positioned, because core-drilling a finished slab under a delivered container is a bad day; and **anchorage** designed for wind uplift and seismic loads, with cast-in plates rather than post-fixed anchors wherever possible.

10.3 Site services

The scope that must be delivered to the interface, in the order it is usually needed:

1. **Power** — supply to the incoming terminals at the agreed voltage, capacity and fault level, with metering, protection coordinated to the unit's upstream device, and an earth system meeting the specified resistance.
2. **Generator provision** — base, bund, fuel store, exhaust routing, acoustic treatment and a control interface to the transfer equipment.
3. **Water** — supply and pressure for evaporative or chilled-water systems, treatment provisions, and a permitted discharge route for blowdown.
4. **Fibre** — diverse ducts to the entry points, terminated in a demarcation enclosure, tested and documented.
5. **Drainage** — condensate, blowdown, and surface water, with interceptors where fuel or glycol is present.
6. **Access and hardstanding** — crane standing area with proven bearing capacity, delivery route, turning space and a maintenance vehicle route that is still usable once the unit is installed.
7. **Perimeter and lighting** — fencing, gates, lighting and CCTV poles, installed before the asset arrives rather than after.

> **The interface document**
>
> Produce a single document listing every interface, with its physical location, specification, tolerance, and the party responsible for each side. Have both the vendor and the site contractor sign it. Nearly every modular project failure traces back to an interface that two competent parties each assumed the other owned.

10.4 Environmental and planning

Containers occupy an unhelpful grey area in many planning regimes: not clearly a building, not clearly a vehicle, not clearly temporary. Determine early, in writing, how the local authority classifies the installation, because the answer determines whether you need full planning consent, a building permit, a temporary structures permit, or none of these. Issues that commonly surface:

- **Noise** — dry coolers, chillers and generators produce noise that is measured at the boundary, often with a stricter night-time limit and a penalty for tonal content. Acoustic modelling at design stage is cheaper than attenuation at commissioning.
- **Emissions** — generator stack height, testing hours and permitted run-time can be tightly restricted in urban areas.
- **Visual impact** — screening, colour and siting conditions are common, and can constrain plant placement.
- **Water** — abstraction licences and discharge consents for evaporative systems, plus legionella risk assessment obligations.
- **Flood risk** — assessment and, frequently, a requirement to raise the unit above a defined flood level, which changes the foundation and access design.

11. Factory Build, Testing and Logistics

The factory is where the value is created and where problems are cheapest to fix. Spend time there.

11.1 The build sequence

A typical 40-foot IT module follows a sequence of roughly ten to twenty weeks, with long-lead items — switchgear, chillers, UPS, battery cells — usually setting the start date rather than the assembly time.

Stage	Indicative duration	Customer involvement
Design freeze and approval drawings	2–4 weeks	High — review and sign every drawing
Long-lead procurement	6–20 weeks (parallel)	Confirm selections; approve substitutions in writing
Enclosure fabrication and finishing	3–5 weeks	Low — spot inspection
Mechanical and electrical installation	4–8 weeks	Medium — mid-build inspection visit
Internal fit-out, racks, containment, cabling	2–4 weeks	Medium — verify layout against as-built
Pre-FAT internal testing	1–2 weeks	Low — request test records
Factory acceptance test (FAT)	3–5 days	Very high — attend in person
Preparation for shipment	1 week	Review packing and transport plan

11.2 Factory acceptance testing

The FAT is the single highest-leverage event in the whole project. It is the last moment at which a defect can be fixed by people who built the unit, with parts on the shelf, before transport and before the site is on the critical path. Attend. Send the person who will operate the unit, not only the person who bought it.

A credible FAT covers, at minimum:

- **Dimensional and visual inspection** against the approved drawings, including clearances actually measured rather than assumed.
- **Electrical tests** — insulation resistance, continuity, earth bonding, polarity, protection settings verified against the coordination study, and functional testing of every breaker and interlock.
- **Load-bank testing** at 25, 50, 75 and 100 percent of design load, held long enough at full load to reach thermal steady state — four hours minimum, eight preferred.

- **Thermal verification** — rack inlet temperatures mapped at multiple heights across every rack position under full load, demonstrating the design envelope with margin.
- **Failure simulation** — pull a cooling unit, drop a UPS module, open a feeder, fail a fan, and observe that the system behaves as designed and the alarms say what they should.
- **Battery discharge test** to prove actual autonomy, not catalogue autonomy.
- **Fire system test** — detection response, cross-zone logic, cooling shutdown and damper closure on alarm, and an enclosure integrity (door fan) test.
- **Controls and alarms** — every point on the points list verified end to end, and the alarm actually arriving at the intended recipient.
- **Security systems** — access control, cameras, tamper detection.
- **Documentation review** — as-built drawings, test records, O&M manuals, spares list, certificates.

> **Non-negotiable FAT practice**
>
> Do not accept a FAT performed at partial load, and do not accept a witness certificate for a test you did not see. Full-load thermal steady state is the condition under which the design either works or does not, and it is the one condition a rushed FAT always skips.

11.3 Transport

Transport is an engineering activity, not a shipping formality. The plan must address:

- **Route survey** — physically driven or professionally surveyed, recording bridge heights, road widths, turning radii, gradients, weight-restricted structures and overhead lines.
- **Permits and escorts** — abnormal load notifications, police or escort requirements, and permitted travelling hours. In many jurisdictions these take weeks.
- **Lifting plan** — crane size and position, ground bearing under outriggers, lifting points, spreader beams, rigging and a qualified lift supervisor.
- **Securing** — internal restraint of racks, batteries, plant and any loose items; shock and tilt indicators fitted so that rough handling is evidenced on arrival.
- **Protection** — weather sealing of openings, desiccant, and protection of external plant from road debris.
- **Sea freight** — CSC plating, lashing to vessel requirements, and salt-air protection. Sea transport imposes sustained vibration and racking loads that road transport does not.
- **Insurance** — confirm cover during transit and, specifically, during the lift. The lift is the highest-risk twenty minutes of the entire project.

Whether the unit ships with racks installed is a decision with consequences in both directions. Shipping loaded saves site time and preserves the tested configuration; shipping empty reduces weight, reduces shock exposure to expensive hardware, and sometimes keeps the load within routine transport limits. Take the vendor's written position on what their structural qualification actually permits.

11.4 Arrival and installation

A disciplined arrival sequence prevents most installation disputes:

1. Inspect shock and tilt indicators **before** unloading, and photograph them. This is the evidence that settles any later damage argument.
2. Photograph the unit externally on the trailer, from all sides.
3. Lift and set, checking level at all bearing points and shimming to tolerance before anchoring.
4. Anchor to the foundation per the seismic and wind design.
5. Connect services in sequence — earth first, then power, then mechanical, then data.
6. Perform an internal inspection for transport damage: loose fixings, displaced cables, cracked fittings, disturbed pipework, and rack contents that have shifted.
7. Re-test before energising. Insulation resistance and earth continuity at minimum; transport is hard on terminations.

12. Commissioning

Commissioning proves the installed system does what the design promised. It is the only defensible basis for accepting the asset and the last opportunity to find problems while someone else is still liable for them.

12.1 The five levels

The industry convention divides commissioning into five levels. Container projects compress Levels 1 to 3 into the factory, which is precisely where the schedule advantage comes from — but Levels 4 and 5 must still be done on site, under real site conditions, and skipping them is the commonest false economy in modular delivery.

Level	Name	What happens	Where
L1	Factory witness	Component testing at manufacturers; witness key equipment tests	Component factories
L2	Delivery inspection	Verify equipment against specification on receipt; check for damage	Integrator factory / site
L3	Pre-functional	Static checks, point-to-point, insulation, settings, cleanliness sign-off	Integrator factory
L4	Functional	Each system operated through its full range; alarms and interlocks proven	Site
L5	Integrated systems test	All systems together under load, with failures injected	Site

12.2 Integrated systems testing

Level 5 is the test that matters, because it is the only one that exercises the interactions between systems — and interactions are where real failures live. The unit runs at full load on load banks, with the control system live, and the commissioning team deliberately breaks things.

A representative script:

1. **Utility failure.** Open the incoming supply. Verify UPS pick-up with no load loss, generator start and transfer within the design window, cooling restart sequence, and correct alarm generation.
2. **Generator failure on transfer.** Fail the generator during the transfer. Verify battery autonomy is sufficient for the shutdown sequence and that the sequence actually executes.
3. **Return to utility.** Verify retransfer, generator cool-down, and battery recharge without exceeding supply capacity.

4. **Cooling unit failure.** Disable one unit at full load. Verify changeover, and record the temperature excursion and recovery time against the design limit.
5. **Total cooling failure.** Disable all cooling at full load. Record the rate of temperature rise and the time available before the allowable limit is breached. This number is one of the most operationally useful figures the project will ever produce — write it in the runbook.
6. **UPS module failure and bypass.** Verify redundancy holds and that transfer to bypass and back is bumpless.
7. **Fire alarm sequence.** Verify cooling shutdown, damper closure, alarm transmission and abort function — with the suppression discharge isolated.
8. **Network path failure.** Fail each fibre path in turn and verify the other carries the load, and that out-of-band access remains available.
9. **Control system failure.** Power down the main controller. Verify that plant fails to a safe state and that local control persists.

> **Record the numbers, not just the pass**
>
> For every failure test, record the measured values: transfer time in milliseconds, peak inlet temperature, seconds to reach the allowable limit, recovery time. A tick in a box tells you it worked once. The numbers tell your operators how much time they have when it happens at 3 a.m.

12.3 Acceptance and handover

Acceptance should be conditional on a defined deliverables package, and payment milestones should be tied to it. The package:

- As-built drawings: single-line diagrams, mechanical schematics, layouts, cable schedules, control points list.
- All commissioning records with measured values, signed and dated.
- Settings records: protection settings, control setpoints, alarm thresholds, firmware versions.
- O&M manuals for every system, with maintenance schedules and consumable specifications.
- Spares list, with recommended holdings and lead times for each item.
- Warranty documentation with clear start dates and the conditions that void cover.
- Training delivered to named operators, with records.
- A snagging list with agreed owners and dates, and a defined process for closure.
- Emergency procedures and a runbook that reflects the **tested** behaviour of this unit, not a generic template.

One further recommendation: retain a portion of the contract value — commonly 5 to 10 percent — until the unit has completed a defined period of stable operation, typically thirty days. Infant-mortality failures in electrical and mechanical plant cluster in the first weeks, and a retention is the most reliable mechanism for ensuring the supplier is still attentive when they occur.

PART IV

Running It

Fifteen years of operation cost more than the unit did. This is how that period is managed.

13. Operations and Maintenance

The purchase decision takes months. The operating decision lasts the life of the asset, and it is where most of the money goes.

13.1 Operating models

Model	Description	Suits	Watch-out
In-house	Owner staff operate and maintain, possibly with OEM contracts for specific plant	Campuses, multiple units, existing DC teams	Skills retention; 24/7 cover for a single asset is expensive
OEM service contract	Manufacturer maintains the unit to an agreed schedule and response time	Single units, edge sites, no local team	Response times at distance; scope gaps between OEMs
Managed service	Third party operates the infrastructure end to end against an SLA	Operators wanting capacity, not facilities management	Exit costs; clarity on who owns telemetry and access
Lights-out / remote	No routine attendance; remote monitoring with dispatched remote hands	Distributed edge estates	Everything depends on monitoring quality and dispatch discipline

For a single remote unit, the arithmetic usually favours a service contract or managed service: staffing a 24/7 rota for one container is rarely justifiable. For an estate of ten or more, in-house capability with OEM backup typically becomes cheaper and markedly more responsive. The crossover point is worth calculating explicitly rather than assumed.

13.2 Maintenance schedule

The following is a defensible baseline for a mainstream container data center. Adjust intervals for environment — coastal, dusty or high-vibration sites need shorter cycles on filtration, coatings and terminations.

Interval	Tasks
Continuous (automated)	Environmental and power telemetry, alarm monitoring, leak detection, security events, battery monitoring
Monthly (remote or local)	Alarm log review, capacity and trend review, UPS and battery status check, filter differential pressure, generator remote status, backup verification

Interval	Tasks
Quarterly	Visual inspection inside and out, filter change or clean, generator on-load test, fluid level checks, seal and gasket inspection, cleanliness, camera and access control verification
Semi-annual	Thermographic survey of all electrical terminations, battery impedance testing, cooling coil clean, condenser or dry cooler clean, leak detection functional test, fire panel service
Annual	Full electrical inspection and torque check, protection relay testing, UPS internal service, generator full-load bank test, fire suppression service and integrity test, water system treatment review and legionella assessment, security system audit, disaster recovery exercise
Every 3–5 years	Battery replacement (VRLA) or capacity assessment (lithium), coating inspection and touch-up, structural and anchorage inspection, control firmware lifecycle review, full integrated systems retest

13.3 Consumables and spares

At a remote site, a two-dollar filter that is not on the shelf can cost a thermal event. Hold spares according to consequence and lead time rather than to cost.

- **On site always** — air filters (a full set plus one), fuses of every rating used, common gaskets and seals, leak-detection cable, lamps, and a set of blanking panels.
- **On site if remote** — one spare cooling fan assembly, one UPS power module if modular, a spare controller with a saved configuration, and a spare network transceiver of each type.
- **Contracted with guaranteed response** — compressors, pumps, CDU assemblies, switchgear components, battery strings, generator parts.
- **Documented and reviewed annually** — with lead times recorded and the list revisited when any component goes end-of-life. The spares list is a living document; an obsolete part with a twenty-six-week lead time is a discovery best made during a review, not during a failure.

13.4 Change management

Most data center outages are caused by change, not by component failure. In a container, where systems are tightly coupled and buffers are small, change discipline matters even more than in a building.

1. **Classify.** Standard (pre-approved, low risk, documented), normal (assessed and scheduled), emergency (post-hoc review mandatory).
2. **Assess impact on redundancy.** Any work that removes redundancy is a higher class of change, regardless of how routine the task is. Working on the B path is not low risk simply because the A path is healthy.
3. **Write a back-out plan** with a decision point and a time limit. "We will keep going until it works" is how a two-hour window becomes a twelve-hour outage.

4. **Define the window** and the conditions under which it is abandoned — including weather, staffing and concurrent activity elsewhere.
5. **Update the documentation as part of the change**, not afterwards. Drift between drawings and reality is the root cause behind a large share of serious incidents.
6. **Review.** Every change is reviewed for outcome; every emergency change is reviewed for why it was an emergency.

The concurrent maintenance question

Ask of every planned task: "can this be done with the load live?" If the answer is no, that is a design limitation, and it should be recorded in the risk register — not discovered on the night. Units sold as concurrently maintainable sometimes are not, once the real service procedure for a specific component is examined.

14. Efficiency, Metrics and Sustainability

Efficiency in a container is a design property far more than an operating one. Most of the opportunity is gone once the unit ships — which is exactly why it belongs in the specification.

14.1 The metrics that matter

Metric	Definition	Good range	Caveat
PUE	Total facility energy ÷ IT energy	1.1–1.4 typical; below 1.1 in cold climates	Meaningless without a stated measurement boundary and load level
WUE	Water consumed ÷ IT energy (L/kWh)	0 for dry designs; 0.2–1.8 evaporative	Rises sharply in hot, dry conditions — the worst case, not the average, is what matters
CUE	Carbon emitted ÷ IT energy	Site and grid dependent	Dominated by grid mix; location choice outweighs design choice
ERF	Energy reused ÷ total energy	0 unless heat reuse exists	Needs a co-located heat consumer; rarely available at remote sites
Rack utilisation	Deployed kW ÷ designed kW	Above 70% after ramp	Low utilisation destroys PUE regardless of design quality

Two warnings about PUE in particular. First, it is **load-dependent**: a container designed for 500 kW and running 100 kW will post a poor PUE no matter how good the engineering, because fixed overheads dominate. Always quote PUE with the IT load at which it was measured. Second, it is **boundary-dependent**: whether the measurement includes transformer losses, external heat rejection, lighting and the generator block heater can shift the figure by ten points or more. When comparing vendor claims, insist on the boundary definition and the measurement method per ISO/IEC 30134, and on an annualised figure rather than a design-point one.

14.2 Where the overhead actually goes

In a well-designed air-cooled container operating near design load, the non-IT energy divides roughly as follows. The proportions shift with climate and density, but the ranking rarely does.

Consumer	Share of overhead	Main lever
Heat rejection (compressors, chillers)	45–65%	Free cooling hours, higher setpoints, efficient part-load control
Air movement (fans, internal and external)	15–30%	Containment, variable speed, low-resistance airflow paths

Consumer	Share of overhead	Main lever
UPS and power conversion losses	10–20%	Efficient topology, operating within the efficient load band
Pumps (where water-cooled)	3–8%	Variable flow, higher delta-T across the loop
Lighting, controls, ancillaries	2–5%	Occupancy-controlled LED; small but free to fix

The implication is clear: **cooling design decides efficiency.** Setpoint policy, economiser hours and part-load control are worth an order of magnitude more than any amount of attention to lighting or ancillaries.

14.3 Design levers, ranked

1. **Site the unit where the climate does the work.** Annual free-cooling hours vary by a factor of five or more between sites. This decision dwarfs every other efficiency measure.
2. **Raise the supply temperature** to the top of the ASHRAE recommended range — accepting the reduced thermal buffer consciously and documenting it.
3. **Contain rigorously.** Sealed containment with full blanking is the cheapest measure on this list and among the most effective.
4. **Run the delta-T high.** A larger temperature difference across the cooling loop means less flow, smaller pumps and fans, and more economiser hours.
5. **Use variable speed everywhere** — fans, pumps, compressors — and tune the control so that part load is genuinely efficient rather than merely tolerated.
6. **Size for the load you will have**, using modular plant that operates in its efficient band from day one rather than one large unit loafing at 20 percent.
7. **Capture heat to liquid** where density permits; liquid-cooled heat is both more efficiently removed and far more reusable.

14.4 Water

Evaporative and adiabatic cooling trade water for electricity, and that trade looks very different depending on where you make it. A design consuming 1.5 litres per kWh at a 500 kW unit uses on the order of six thousand cubic metres a year — a rounding error in a temperate industrial area and a serious matter in a water-stressed region. Where water is scarce, contested or expensive, a dry design that accepts a higher PUE is frequently the better overall answer, and is increasingly the one regulators and communities expect.

Where water is used, three obligations follow: treatment to control scale, corrosion and biological growth; a legionella risk assessment with a documented control regime; and a permitted discharge route for blowdown. None of these are optional, and all of them have ongoing cost.

14.5 Sustainability beyond energy

- **Embodied carbon.** Steel, concrete foundations, copper and refrigerant charge all carry embodied emissions. Modular construction generally reduces material waste relative to site construction, and reuse or relocation of a unit is a genuine and under-claimed advantage.
- **Refrigerant selection.** Global-warming potential and phase-down schedules under regional regulation are moving quickly. Specifying a low-GWP refrigerant now avoids a forced and expensive change later in the asset's life.
- **Heat reuse.** The highest-value sustainability measure available, and the hardest to arrange — it requires a heat consumer next door. Liquid-cooled units producing higher-grade heat make this considerably more feasible than air-cooled ones.
- **End of life.** Steel enclosures, copper and aluminium are readily recyclable; batteries and refrigerants require controlled handling. Ask the vendor for a take-back or decommissioning position at the time of purchase, when you still have leverage.

15. Reliability, Failure Modes and Incident Response

Things fail. The measure of an operation is not whether failures occur but how quickly and how calmly they are contained.

15.1 The failure modes that actually occur

Industry incident data has been consistent for years on the ranking below. Note how far down the list genuine component failure appears.

Failure mode	Typical cause	Container-specific aggravation
Power chain failure	UPS, breaker, transfer, upstream supply	Tight coupling; fewer alternative paths than a hall
Cooling failure	Compressor, fan, control, refrigerant or water loss	Very low thermal buffer — minutes or less
Human error during work	Wrong breaker, wrong setting, unplanned isolation	Cramped space; adjacent equipment easily disturbed
Network / fibre cut	Excavation, rodent damage, connector fault	Often a single physical route to a remote site
Control system fault	Firmware, configuration, sensor failure	High automation dependence in unmanned operation
Environmental event	Flood, storm, extreme heat, wildfire smoke	Exposed siting; small envelope; external plant vulnerable
Security incident	Theft, vandalism, unauthorised access	Remote, unmanned, visible and valuable
Fire	Electrical fault, battery event	Confined volume; rapid escalation; slow fire service response

15.2 Thermal runaway: the defining container risk

This is the failure mode that distinguishes container operations from building operations, and it deserves its own procedure. With cooling lost at full load, inlet temperatures in a 40-foot module can move from normal to the ASHRAE allowable limit in a handful of minutes, and to shutdown thresholds shortly after. The response cannot be improvised.

1. **Know the number.** From the Level 5 integrated test, you have a measured rate of temperature rise at full load with cooling failed. That number, and the minutes it gives you, belongs on the first page of the runbook.
2. **Alarm early.** Alarm on rate of change, not only on absolute temperature. A three-degree rise in ninety seconds is actionable long before any threshold is crossed.

3. **Automate the first response.** Load shedding, workload migration or graceful shutdown must be automatic, with thresholds agreed with the application owners in advance. Waiting for a human decision consumes the entire budget of available time.
4. **Define the manual sequence.** Which racks shut down, in what order, and who has authority. Print it; do not rely on a system that may itself be affected.
5. **Provide emergency ventilation** where the environment and security posture permit it — the ability to open the unit and move outside air through it can buy substantial time.
6. **Practise.** A thermal event exercise once a year, with timings recorded.

> **Two numbers every operator should know**
>
> Minutes from cooling failure to the allowable inlet temperature limit at current load. And minutes of battery autonomy at current load with the batteries in their current measured state — not their rated state. Both change as load grows and as batteries age. Recalculate them quarterly.

15.3 Incident response

1. **Detect** — automated alarm with defined severity and recipient; a confirmed, acknowledged alert rather than an email into a shared mailbox.
2. **Assess** — what has failed, what redundancy remains, what is the trajectory, and how long until service is affected.
3. **Stabilise** — protect the load first. Shed non-critical load, migrate workload, or shut down gracefully before an uncontrolled failure does it for you.
4. **Escalate** — a defined ladder with names, numbers and authority levels, including who may authorise a shutdown and who must be told.
5. **Repair** — with parts, access and competent people, under a change record even in an emergency.
6. **Restore** — return to normal configuration deliberately, verifying redundancy is genuinely restored rather than assumed.
7. **Review** — blameless post-incident review within a week, with actions tracked to closure.

For remote sites the assessment step is the hard one, because it happens without eyes on the equipment. This is why instrumentation and out-of-band access are operational necessities rather than luxuries: an operator who can see rack inlet temperatures, per-circuit power and door status can make a correct decision from two hundred kilometres away. An operator with a single "cooling alarm" cannot.

15.4 Building a realistic availability picture

Availability is a property of the whole delivery chain, not of the container. A unit with 2N power and N+1 cooling, fed by a single utility supply with no generator, over a single fibre route, at a site three hours from the nearest engineer, is not a highly available service — regardless of what the internal topology diagram shows.

Model it honestly: list every element in the chain from utility to user, assign each a realistic failure rate and repair time including travel, and compute the result. The exercise almost always shows that **mean time to repair at a remote site dominates everything else**. Which, in turn, means that the highest-value availability investments for a remote container are rarely more redundancy inside the box — they are spares on site, a guaranteed response time, and instrumentation good enough to send the right engineer with the right part the first time.

PART V

Applications and Economics

Edge, AI and the numbers — where containers are being deployed and how to justify them.

16. Edge Deployments

Edge is where the container model is least contested and most operationally demanding. The engineering is easier; the logistics are harder.

16.1 Why edge, specifically

Four drivers push compute out of centralised facilities, and it is worth being precise about which one applies, because they imply different designs.

- **Latency.** Applications that require single-digit millisecond response — industrial control, autonomous systems, augmented reality, some trading and gaming workloads — cannot be served from a distant region. Physics, not network quality, is the constraint.
- **Bandwidth economics.** Video analytics, sensor arrays and industrial telemetry generate data that is expensive to transport and cheap to reduce locally. Processing at source can cut backhaul by one or two orders of magnitude.
- **Data residency and sovereignty.** Regulatory or contractual requirements that data remain within a jurisdiction, a site, or a customer's own premises.
- **Resilience.** Operations that must continue when the wide-area link fails — ports, mines, hospitals, utilities, manufacturing lines.

If none of these apply with force, the edge deployment is probably solving a problem that a region and a CDN already solve more cheaply. Being willing to test that is part of a good design process.

16.2 Design implications

Edge units are smaller, but their requirements are in several respects more demanding than those of a campus module.

Requirement	Why it differs at the edge
Unattended operation	No staff on site, possibly for months; everything must self-protect and self-report
Environmental tolerance	Sites are chosen for location, not for climate; wide ambient range and poor air quality are normal
Standardisation	Tens or hundreds of identical units; any variation multiplies spares, training and documentation cost
Remote management	Full out-of-band access, remote power cycling and console access are functional requirements, not options
Physical security	Publicly accessible or lightly supervised locations with long response times
Serviceability by non-specialists	The person who attends may be a general technician, so procedures must be explicit and components modular

Requirement	Why it differs at the edge
Noise and appearance	Often sited near people; acoustic and visual constraints can be binding
Footprint	Space is frequently the scarcest resource — kerbside, rooftop, compound corner

16.3 Fleet management

One container is a project. Two hundred containers are a fleet, and a fleet is a different discipline. The disciplines that separate a working estate from an expensive one:

1. **Ruthless standardisation.** One design, one firmware baseline, one spares kit, one runbook. Every permitted variant should require justification at a governance forum. Variation is the tax that eats edge economics.
2. **Centralised monitoring with fleet-level analytics.** Comparing units against each other detects drift and incipient failure far earlier than any single unit's thresholds. The unit whose cooling power has risen eight percent relative to its peers has a dirty condenser, and you know it before it alarms.
3. **Staged rollout of change.** Firmware and configuration changes go to a pilot group, then a wave, then the fleet — never everywhere at once. A bad firmware push to two hundred units is an estate-wide outage.
4. **Logistics as a first-class function.** Spares depots positioned by drive time, not by geography; contracted response times; pre-staged kits.
5. **Automated provisioning and configuration management.** Units should be brought online and reconfigured without manual, site-by-site work.
6. **Lifecycle planning.** Units age and reach end-of-life in waves. Plan refresh in cohorts and budget accordingly.

> **The economics of edge**
>
> Edge cost is dominated by the number of sites, not by the capacity at each. Every additional site adds a lease, a connection, a security posture, a maintenance visit and a spares position. The design question is therefore not only "what goes in the box" but "what is the minimum number of boxes that meets the latency requirement" — and that number is usually smaller than the first proposal suggests.

17. High-Density AI and HPC Containers

Accelerated computing has changed the density question from a gradual trend into a step change. It is the strongest current driver of container deployment, and the least forgiving to design.

17.1 What changed

For two decades, average rack density crept from about 3 kW toward 10 kW, and data centers were designed around that pace. GPU-dense racks for AI training arrive at 40 to 130 kW and beyond. That is not an increment; it is an order of magnitude, and it breaks assumptions in four places at once.

Assumption	Traditional	AI / HPC	Consequence
Rack power	5–10 kW	40–130+ kW	Distribution, busway ratings and breaker counts all change
Cooling medium	Air	Liquid, with residual air	Piping, CDUs, fluid management, leak strategy
Rack weight	600–1,200 kg	1,500–2,500+ kg	Floor design, transport limits, lifting and anchorage
Load profile	Steady, diurnal	Step changes of hundreds of kW in seconds	Power quality, UPS transient response, cooling control stability
Network	10–25 GbE, tolerant	400–800 GbE, latency and topology sensitive	Cable volume, optics power, strict length limits

17.2 Liquid cooling in a container

Direct-to-chip liquid cooling is now the mainstream answer, and the container is an unusually good host for it — a purpose-built enclosure can be designed around the piping from the start, which an existing building cannot. The practical design requirements:

- **Coolant distribution unit sizing**, including redundancy. A CDU failure at 100 kW per rack is a very short emergency; N+1 at the CDU level is effectively mandatory.
- **Loop separation.** A clean, treated, closed technology-cooling loop serving the cold plates, heat-exchanged to a facility loop that rejects to the environment. Never plumb servers directly to the facility water.
- **Fluid specification and management.** Chemistry, conductivity, filtration, corrosion inhibition, biological control, and a sampling schedule. Fluid quality is a maintenance activity, not a commissioning one.

- **Leak detection and containment** under every pipe, joint and manifold, with automatic isolation of the affected branch and drip trays with defined drainage paths.
- **Quick-disconnect strategy.** Dry-break couplings at every server, with a documented and rehearsed procedure for removing a node from a live loop.
- **Residual air cooling** sized for the 15 to 30 percent of heat that the liquid loop does not capture. This is the most commonly under-sized element in first-generation liquid designs.
- **Pressure and flow control**, including what happens on a pump failure or an over-pressure event, and how the system proves flow to each cold plate before the hardware draws full power.

17.3 Power for AI

AI training loads are electrically unusual. Synchronised jobs across thousands of accelerators cause the load to step by a large fraction of the total in fractions of a second, repeatedly. Design responses:

- Size the UPS for **transient response**, not just steady-state capacity, and prove it under a real step-load test rather than a smooth ramp.
- Assess the effect of repeated steps on the upstream supply and on the generator; the utility may impose flicker or ramp-rate limits.
- Specify power quality monitoring with sub-cycle capture so that events can be analysed after the fact.
- Consider energy storage sized to smooth steps rather than only to provide autonomy — a different sizing calculation with different chemistry preferences.
- Allow for higher distribution ratings throughout: busway, breakers, cabling and rack PDUs all move up a class or two, and at 130 kW a rack may need several three-phase feeds per side.

17.4 Structural and logistics consequences

A rack at 2,000 kg on a footprint of roughly 0.6 by 1.2 metres is a point load well beyond ordinary raised-floor practice. Container floors for AI modules are typically solid, heavily reinforced structural decks with the racks bolted directly to the frame. The knock-on effects are worth stating plainly: the unit may exceed routine road transport weight limits when loaded, which means shipping empty and populating on site — which in turn means the factory acceptance test cannot include the actual IT hardware, and the site commissioning burden rises. Plan for that sequence explicitly rather than discovering it during transport planning.

Future-proofing without over-building

If AI workloads are even possible within the asset life, specify the **provisions** now: structural capacity for heavy racks, pipe routes and penetrations for liquid, spare electrical capacity and space for a CDU, and heat rejection headroom. Provisions are cheap at design stage and close to impossible to add later. Fitting out for a density you do not yet need is the expensive mistake; failing to leave room for it is the irreversible one.

18. Total Cost of Ownership and Procurement

The purchase price is typically a quarter to a third of what the asset will cost you. This chapter is about the rest, and about buying well.

18.1 The full cost picture

Category	Includes	Notes
Container capital	Enclosure, power, cooling, fire, security, controls, racks	The number vendors quote; usually 25–40% of lifetime cost
Site capital	Foundations, ducts, fencing, lighting, hardstanding, access	Frequently under-estimated by the largest margin
Utility connection	Connection charges, reinforcement, transformer, metering	Can exceed the container cost at constrained sites
Generation and fuel	Genset, bund, tank, acoustic treatment, commissioning	Often a separate package; confirm the interface
Transport and installation	Freight, permits, escorts, crane, rigging, insurance	Route-dependent; get a firm quote, not an allowance
Commissioning	Load banks, commissioning agent, integrated systems test	Do not economise here — see Chapter 12
Energy	IT energy plus overhead, over the asset life	The single largest lifetime line at most sites
Water	Supply, treatment, discharge	Zero for dry designs; material for evaporative
Maintenance	Contracts, spares, consumables, callouts	Typically 3–7% of capital per year
Battery and plant refresh	Battery replacement, compressor or CDU overhaul	Schedule and budget explicitly by year
Connectivity	Circuits, diverse routes, carrier charges	Recurring; diversity roughly doubles it
Decommissioning	Removal, refrigerant and battery disposal, site reinstatement	Offset by residual value; ask for a take-back position

18.2 Building a defensible model

1. Model over the **real asset life** — ten to fifteen years for infrastructure, with an IT refresh cycle of three to five years overlaid.

2. Discount to present value using the organisation's cost of capital. This is where phased modular deployment earns most of its advantage, and omitting it understates the case.
3. Model the **load ramp** honestly. Energy and efficiency both depend on actual load, and a unit at 40 percent utilisation has a very different cost per delivered kilowatt-hour than one at 85 percent.
4. Use a realistic energy price with an escalation assumption, and test the result against a high-price scenario. Energy usually dominates; the model should show how sensitive the answer is to it.
5. Include the cost of downtime, weighted by the modelled availability. If the organisation has never priced an hour of downtime, the TCO exercise is a good occasion to do so.
6. Run sensitivities on the three variables that actually move the answer: energy price, load ramp, and utility connection date.

18.3 Selecting a vendor

Ask for evidence rather than capability statements. The following questions separate integrators from assemblers:

- **Reference sites.** Not a list — visits, and conversations with the operators, preferably about a unit that has been running for three or more years.
- **Engineering evidence.** Structural calculations, thermal modelling, transport shock and vibration assessment, protection coordination study. A supplier who can produce these on request has an engineering function.
- **Test regime.** What exactly does their FAT cover, at what load, for how long, and may we witness all of it?
- **Supply chain.** Whose UPS, whose cooling, whose switchgear? Are these units serviceable by third parties, or locked to the integrator?
- **Service footprint.** Where are engineers based relative to my site, and what is the contracted response time — with evidence that it has been met.
- **Financial stability.** A fifteen-year asset from a supplier with a two-year balance sheet is a risk that belongs in the evaluation.
- **Spares and obsolescence policy.** How long are parts guaranteed, and what happens when a component is discontinued?
- **Documentation sample.** Ask to see a complete O&M package from a real delivered project. It is the most reliable single indicator of delivery quality available before contract.

18.4 Contract terms worth fighting for

- **Performance guarantees** with defined measurement methods: PUE at stated loads and ambients, cooling capacity at design ambient, autonomy at full load — and remedies if they are not met.
- **A liquidated-damages regime** for late delivery, tied to a clearly defined milestone.

- **Payment milestones tied to verification**, including a substantial payment released only on successful FAT witnessed by you.
- **Retention** of 5 to 10 percent until a defined period of stable operation has passed.
- **Defined response times** for warranty callouts, with escalation and remedies.
- **Documentation and source data as a deliverable**, not a courtesy — including controller configurations and protection settings.
- **Change control** with a pre-agreed rate schedule, so that variations are priced by formula rather than by negotiation under time pressure.
- **Intellectual property and access rights** to the control system, so a future service provider can operate the unit without the original vendor's consent.
- **Exit provisions** — what happens on relocation, resale or early termination, and the vendor's obligations in each case.

The decision, in one paragraph

Buy a container data center when speed matters more than cost per kilowatt, when demand is uncertain enough that phasing has real value, when no suitable building exists, or when density exceeds what your existing facilities can cool. Do not buy one because it sounds modern. Price the whole project — site, utility, transport, commissioning, fifteen years of energy and maintenance, and decommissioning — and compare it against the alternative on the same basis. If the modular case is real, it will survive that comparison comfortably.

PART VI

Appendices

Checklists, a glossary and a standards index for use during a live project.

Appendix A. Specification Checklist

Use this as the skeleton of a requirements document or an RFP. Every line should have an answer before a purchase order is raised; a blank is a risk you have not yet priced.

A.1 Capacity and density

- IT load at day one, at year three, and at design maximum (kW).
- Rack count, rack height (U), rack width and depth, and maximum weight per rack.
- Maximum per-rack power, and whether it is uniform or mixed across the row.
- Density class now and anticipated within the asset life (see Chapter 3).
- Provisions required for future liquid cooling: structural, hydraulic, electrical, spatial.

A.2 Electrical

- Incoming supply voltage, frequency, phases, fault level and earthing arrangement.
- Redundancy topology required (N, N+1, 2N, block) and the maintainability standard implied.
- UPS topology, efficiency at the expected load points, and eco-mode policy.
- Battery chemistry, autonomy at full load, expected life and replacement plan.
- Battery compartment separation, ventilation, detection and fire strategy.
- Generator interface: control signals, transfer type, transfer time, step-load acceptance.
- Distribution method (busway or hard-wired) and future change strategy.
- Metering granularity — at minimum, branch-circuit level.
- Rack PDU specification: outlets, phases, monitoring level, A/B configuration.
- Earthing, bonding and lightning protection requirements.
- Arc-flash study, labelling and working-clearance compliance.

A.3 Mechanical

- Cooling architecture and redundancy level.
- Design ambient: dry bulb and wet bulb at the local 0.4% condition, plus the extreme case.
- Supply air temperature setpoint and permitted range; ASHRAE class targeted.
- Containment type and sealing standard.
- Heat rejection method; water availability, quality, treatment and discharge consent.
- Measured time from total cooling failure to the allowable inlet limit at full load.
- Filtration class and the expected change interval for the site environment.
- Leak detection coverage and automatic isolation behaviour.

- Noise limit at the site boundary, day and night, including tonal assessment.

A.4 Enclosure

- External dimensions, gross weight loaded and empty, and centre of gravity.
- Ingress protection class for enclosure and for each item of external plant.
- Coating system appropriate to the corrosion category of the site.
- Insulation type and thickness; thermal bridging and vapour control strategy.
- Structural design: floor point loads, wind, snow, seismic, transport and lifting.
- Whether the unit is qualified to be transported loaded, and under what limits.
- Door arrangement, swing direction, panic hardware and egress provision.
- Aisle and service clearances, verified physically at FAT.

A.5 Fire, safety and security

- Detection: aspirating, point, thermal; zoning and cross-zone logic.
- Suppression agent, design concentration, hold time and cylinder location.
- Pressure relief venting, sized and verified.
- Enclosure integrity (door fan) test as a deliverable.
- Interlocks: cooling shutdown, damper closure, alarm transmission, abort.
- Emergency lighting, egress marking, EPO arrangement and its protection against accidental operation.
- Lone-worker procedure and duress provision.
- Access control type, logging, and anti-passback requirements.
- CCTV coverage, local retention and alarm verification path.
- Tamper, tilt, vibration and geolocation monitoring.

A.6 Controls, monitoring and network

- Points list — every monitored and controlled point, with units and alarm thresholds.
- Protocols supported: Modbus, BACnet, SNMP, Redfish, API.
- Local autonomy behaviour with no network connection.
- Out-of-band management path and its independence from production.
- Data ownership, retention and vendor access rights.
- Fibre entry points, diversity, sealing and demarcation arrangement.
- Internal topology (top-of-rack or end-of-row) and containment capacity.
- Labelling scheme and as-built cable records.

A.7 Commercial and delivery

- Delivery date, milestones and liquidated damages regime.
- FAT scope, duration, load levels and witnessing rights.
- Site commissioning scope, including Level 5 integrated systems testing.
- Documentation deliverables list.
- Spares list with lead times, and recommended site holdings.
- Training scope and named attendees.
- Warranty period, response times and exclusions.
- Retention amount and release conditions.
- Performance guarantees and their measurement method.
- End-of-life, take-back and relocation provisions.

Appendix B. Commissioning Checklist

A condensed field checklist for Levels 4 and 5. Record measured values, not ticks.

#	Test	Record
1	Insulation resistance and earth continuity after transport	MΩ per circuit; earth resistance
2	Protection settings verified against coordination study	Setting per device; discrepancies
3	Phase rotation and voltage balance at all boards	V per phase; % imbalance
4	Load bank at 25 / 50 / 75 / 100% to thermal steady state	kW, hours held, stabilised temperatures
5	Rack inlet temperature map at full load (top/mid/bottom, all racks)	°C per point
6	UPS transfer to bypass and back under load	Transfer time; any load disturbance
7	UPS module failure with load live	Behaviour; alarms raised
8	Battery discharge to end voltage at full load	Measured autonomy in minutes
9	Utility failure: generator start and transfer	Start time; transfer time; frequency/voltage excursion
10	Generator fails during transfer	Shutdown sequence executed; time
11	Retransfer to utility and battery recharge	Time; peak recharge current
12	Single cooling unit failure at full load	Peak inlet temp; recovery time
13	Total cooling failure at full load	°C per minute rise; minutes to allowable limit
14	Leak detection triggered on each zone	Alarm and isolation response
15	Fire detection and interlock sequence (suppression isolated)	Detection time; damper and cooling response
16	Enclosure integrity (door fan) test	Pass level; predicted hold time
17	Each fibre path failed in turn	Failover behaviour; OOB access retained
18	Control system powered down	Plant fail-safe state; local control retained
19	Every alarm point verified end to end to recipient	Points tested; failures
20	Access control, CCTV, tamper and door sensors	Events logged and received

Appendix C. Glossary

Terms used in this book, defined as they are used in the container data center context.

Term	Definition
Adiabatic cooling	Pre-cooling of air by evaporating water before it passes over a heat exchanger, improving capacity at the cost of water consumption.
ASHRAE class (A1–A4)	Thermal environment classes defining recommended and allowable inlet conditions for IT equipment.
ATS	Automatic transfer switch — selects between utility and generator supply.
Busway	Overhead power distribution with movable tap-off boxes, allowing circuits to be added without rewiring.
CDU	Coolant distribution unit — separates and controls the technology cooling loop from the facility loop in liquid-cooled systems.
Clean agent	A gaseous fire-suppression agent that leaves no residue and is safe for energised electronics.
Containment	Physical separation of hot and cold airstreams to prevent recirculation and bypass.
CSC plate	Safety approval plate certifying a container for international transport under the Container Safety Convention.
CUE	Carbon usage effectiveness — carbon emitted per unit of IT energy.
DCIM	Data center infrastructure management — software for monitoring, capacity planning and asset management.
Delta-T	Temperature difference across a cooling loop or IT equipment; a higher delta-T generally means lower flow and better efficiency.
Direct-to-chip	Liquid cooling in which cold plates mounted on processors remove heat directly to a coolant loop.
Door fan test	Pressurisation test verifying that an enclosure is tight enough to hold a suppression agent for its design hold time.
Dry cooler	A finned heat exchanger rejecting heat to ambient air with no water consumption.
Economiser / free cooling	Cooling using ambient conditions with reduced or no mechanical refrigeration.
EPO	Emergency power off — a control that removes power from the space; itself a documented cause of outages.
FAT	Factory acceptance test — verification of the complete unit at the manufacturer before shipment.

Term	Definition
High cube	A container approximately 300 mm taller than the standard ISO height.
Immersion cooling	Submersion of IT equipment in a dielectric fluid that removes heat directly from all components.
IP rating	Ingress protection classification for dust and water, defined in IEC 60529.
IST	Integrated systems test — Level 5 commissioning, exercising all systems together under load with injected failures.
N+1	Redundancy configuration providing one spare unit beyond the capacity required.
OOB	Out-of-band management — access to infrastructure independent of the production network.
PUE	Power usage effectiveness — total facility energy divided by IT energy.
RDHx	Rear-door heat exchanger — a cooling coil replacing a rack's rear door.
Static transfer switch	A solid-state device transferring load between sources within a fraction of a cycle.
Thermal runaway	Self-sustaining exothermic failure in a battery cell; also used for uncontrolled temperature rise after cooling loss.
Tier (Uptime Institute)	A site-level classification of topology and operational sustainability, from Tier I to Tier IV.
VESDA / ASD	Aspirating smoke detection — air-sampling detection capable of identifying incipient combustion.
WUE	Water usage effectiveness — litres of water consumed per unit of IT energy.
2N	Two fully independent systems, each capable of supporting the entire load.

Appendix D. Standards Index

A working index of the standards most often cited in container data center specifications. Always confirm the current edition and the version adopted in your jurisdiction.

Reference	Subject
ISO 668 / ISO 1496	Freight container dimensions, ratings and structural testing
IEC 60529	Degrees of protection provided by enclosures (IP code)
IEC 60364	Low-voltage electrical installations
IEC 61439	Low-voltage switchgear and controlgear assemblies
IEC/UL 62368-1	Audio/video, information and communication technology equipment safety
NFPA 70 (NEC)	National Electrical Code (United States)
NFPA 75	Fire protection of information technology equipment
NFPA 76	Fire protection of telecommunications facilities
NFPA 2001	Clean agent fire extinguishing systems
ISO 14520 / EN 15004	Gaseous fire-extinguishing systems
ASHRAE TC 9.9	Thermal guidelines for data processing environments
ASHRAE 90.4	Energy standard for data centers
TIA-942	Telecommunications infrastructure standard for data centers
EN 50600 series	Information technology — data centre facilities and infrastructures
ISO/IEC 30134 series	Key performance indicators including PUE, REF, ERF, CUE and WUE
ISO/IEC 22237 series	Data centre facilities and infrastructures (international alignment of EN 50600)
ASCE 7 / IBC	Minimum design loads and building code provisions (wind, snow, seismic)
Eurocode 8 (EN 1998)	Design of structures for earthquake resistance
IEC 61000 series	Electromagnetic compatibility
ISA 71.04	Environmental conditions for process control systems — airborne contaminants
Uptime Institute Tier Standard	Topology and operational sustainability classification

End of the handbook. The checklists in Appendices A and B are intended to be copied into project documents and edited. If only one practice from this book is adopted, make it the integrated systems test in Chapter 12 — and the habit of recording measured values rather than ticks.