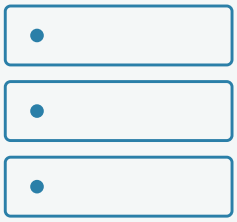
LESSON BOOK



Data Center Systems

Power, heat, cooling, measurement, control and efficiency

Twenty-four lessons · 575 pages

Ricardo Marquez, Ph.D.

ricardo-marquez.com

Data Center Systems

Power, heat, cooling, measurement and control as one physical system

A lesson sequence, written up as a book

Ricardo Marquez

October 5, 2026

How to read this book

This book treats a data center as a physical system rather than as a room of computers. Each chapter expands one lesson of a sequence that follows energy through the facility,

$$\text{electricity} \rightarrow \text{IT equipment} \rightarrow \text{heat} \rightarrow \text{air and water} \rightarrow \text{environment},$$

and then asks what the facility measures of that chain, how it acts on it, and how its efficiency is reported. Chapters 12 to 20 turn to operation under uncertainty: reliability, diagnosis, state estimation, identifiability, capacity as a risk, what-if analysis, the choice of the next measurement, and optimization. Chapters 21 to 24 close the book: the facility as a flexible load on the grid, the choice of model fidelity, the data that are publicly available, and a synthesis. The emphasis throughout is on physical modeling, operation, uncertainty, diagnosis and control, not on a catalogue of equipment.

Three questions recur in every chapter and organize the book: what is *measured*, what is *calculated* from measurements, and what must be *inferred*. A rack inlet temperature is measured; the heat a rack generates is calculated from its electrical power; a recirculation path or a degraded cooling coil is inferred. The numbers used in the examples are illustrative, chosen to make each relation concrete; they are not data from a particular facility.

Where the field uses a shorthand term, a *Jargon* box expands it once and the text then uses the expanded meaning. Every symbol is defined in the notation chapter that follows and is used with the same meaning in every chapter. Each chapter closes with a statement of what to retain, exercises, and a short section connecting the chapter to the TERRA framework, a graph-based way of modeling physical systems as conservation laws on nodes with flows on edges, with probabilistic inference of the states and parameters that the measurements do not give directly.

Contents

How to read this book **ii**

Notation and vocabulary **ix**

1 System architecture **1**
1.1 The facility as an energy-conversion system 1
1.2 The electrical and IT subsystems 4
1.3 The path of heat 5
1.4 Three coupled networks and the control loop 10
1.5 Redundancy and measurement 13
1.6 Diagnosis and the system picture 16
1.7 What to retain 20
1.8 Exercises 20
1.9 Connection to the TERRA framework 21

2 IT loads and computing demand **23**
2.1 Server power 23
2.2 Rack power and rack density 27
2.3 From computing demand to load 29
2.4 Flexibility of the load 31
2.5 Time variation and facility load 34
2.6 Telemetry, diagnosis and state space 37
2.7 What to retain 40
2.8 Exercises 40
2.9 Connection to the TERRA framework 43

3 Electrical power distribution **44**
3.1 Continuity of power 44
3.2 From the utility to the UPS 46
3.3 From the UPS to the server 50
3.4 Redundancy and contingency capacity 53
3.5 Losses and the facility power balance 56
3.6 Topology, state and reliability 59
3.7 What to retain 63
3.8 Exercises 63
3.9 Connection to the TERRA framework 65

4 UPS, batteries and backup power **67**
4.1 Two roles in one transition 67
4.2 Sizing the battery 70
4.3 The generator and the transfer 73

4.4 Load priority and the thermal buffer . . . 76
4.5 Battery health and runtime . . . 78
4.6 Dependence, diagnosis and dual use . . . 81
4.7 What to retain . . . 87
4.8 Exercises . . . 87
4.9 Connection to the TERRA framework . . . 89

5 Heat generation and thermal balance 90
5.1 Conservation of energy . . . 90
5.2 The rack and the air stream . . . 93
5.3 The room and the water loop . . . 95
5.4 Resistance, capacitance and time . . . 98
5.5 Spatial structure and mixing . . . 101
5.6 Closing the balance and inference . . . 105
5.7 What to retain . . . 108
5.8 Exercises . . . 110
5.9 Connection to the TERRA framework . . . 110

6 Airflow and rack cooling 112
6.1 Airflow required by a rack . . . 112
6.2 Separation and mixing . . . 114
6.3 Pressure, resistance and fans . . . 118
6.4 Distribution, containment and hot spots . . . 122
6.5 Measuring and inferring airflow . . . 125
6.6 Network models, CFD and control . . . 127
6.7 What to retain . . . 132
6.8 Exercises . . . 133
6.9 Connection to the TERRA framework . . . 134

7 Cooling equipment 135
7.1 The heat-rejection chain . . . 135
7.2 The coil and the chilled-water loop . . . 137
7.3 Pumps and the chiller . . . 140
7.4 Heat rejection to the atmosphere . . . 144
7.5 The plant as a network with capacities . . . 147
7.6 Measurement, diagnosis and control . . . 151
7.7 What to retain . . . 155
7.8 Exercises . . . 156
7.9 Connection to the TERRA framework . . . 157

8 Thermal dynamics 159
8.1 The dynamic heat balance . . . 159
8.2 The first-order thermal model . . . 162
8.3 Memory, multiple masses and storage . . . 164
8.4 Propagation, delays and diagnosis . . . 168
8.5 Control interaction and dynamic capacity . . . 171
8.6 State space, estimation and identification . . . 174
8.7 What to retain . . . 178
8.8 Exercises . . . 178
8.9 Connection to the TERRA framework . . . 180

9 Sensors and instrumentation **182**
9.1 The measurement equation . . . 182
9.2 What is measured . . . 184
9.3 Time and error . . . 190
9.4 Checks, residuals and observability . . . 194
9.5 Design, placement and estimated quantities . . . 196
9.6 From raw sensors to risk . . . 199
9.7 What to retain . . . 203
9.8 Exercises . . . 203
9.9 Connection to the TERRA framework . . . 205

10 Controls **207**
10.1 The control problem . . . 207
10.2 Feedback controllers . . . 209
10.3 The loops of a cooling plant . . . 215
10.4 Predictive and coordinated control . . . 218
10.5 Control under uncertainty . . . 222
10.6 Diagnosis-aware and supervisory control . . . 225
10.7 What to retain . . . 229
10.8 Exercises . . . 229
10.9 Connection to the TERRA framework . . . 230

11 PUE and energy efficiency **232**
11.1 Power usage effectiveness . . . 232
11.2 What the ratio does not measure . . . 234
11.3 Why PUE varies . . . 237
11.4 Decomposition and local problems . . . 240
11.5 PUE in time . . . 244
11.6 Capacity, evidence and decisions . . . 247
11.7 What to retain . . . 251
11.8 Exercises . . . 252
11.9 Connection to the TERRA framework . . . 253

12 Redundancy and reliability **255**
12.1 Redundancy levels and layers . . . 255
12.2 Dependence and single points of failure . . . 258
12.3 Availability, maintenance and capacity states . . . 262
12.4 Contingency analysis in time . . . 265
12.5 Event trees and fault trees . . . 268
12.6 Operational risk . . . 271
12.7 What to retain . . . 275
12.8 Exercises . . . 275
12.9 Connection to the TERRA framework . . . 276

13 Failure modes and diagnostics **278**
13.1 Symptoms, causes and hypotheses . . . 278
13.2 Air-side failure modes . . . 280
13.3 Water-side and plant failure modes . . . 283
13.4 Sensor, topology and electrical faults . . . 286
13.5 Residual-based diagnosis and ranking . . . 290
13.6 From diagnosis to action . . . 296
13.7 What to retain . . . 299

13.8 Exercises . . . 300
13.9 Connection to the TERRA framework . . . 300

14 State estimation **302**
14.1 Measurements and the physical state . . . 302
14.2 Dynamic state estimation . . . 307
14.3 Static estimation and residuals . . . 311
14.4 Estimated quantities in a data center . . . 314
14.5 Observability and posterior uncertainty . . . 317
14.6 Use and limits of state estimation . . . 319
14.7 What to retain . . . 324
14.8 Exercises . . . 324
14.9 Connection to the TERRA framework . . . 325

15 Uncertainty and identifiability **327**
15.1 Uncertainty and identifiability . . . 327
15.2 Parameter combinations and posterior geometry . . . 331
15.3 Ambiguity in diagnosis . . . 336
15.4 Priors, adequacy and model structure . . . 338
15.5 Capability and the limits of inference . . . 341
15.6 Identifiability relative to the decision . . . 343
15.7 What to retain . . . 347
15.8 Exercises . . . 348
15.9 Connection to the TERRA framework . . . 348

16 Probabilistic diagnosis **350**
16.1 Diagnosis as inference . . . 350
16.2 Bayes' rule on a worked case . . . 351
16.3 Likelihoods from physical models . . . 354
16.4 Identifiability, confidence and scale . . . 358
16.5 Time, normal operation and model error . . . 361
16.6 From posterior to action . . . 364
16.7 What to retain . . . 369
16.8 Exercises . . . 369
16.9 Connection to the TERRA framework . . . 370

17 Capacity and risk **372**
17.1 Capacity as a system property . . . 372
17.2 State-dependent and secure capacity . . . 374
17.3 Local and constrained capacity . . . 376
17.4 Probabilistic capacity . . . 379
17.5 Capacity margin in time . . . 383
17.6 Bottlenecks and capacity planning . . . 387
17.7 What to retain . . . 391
17.8 Exercises . . . 391
17.9 Connection to the TERRA framework . . . 393

18 What-if analysis **395**
18.1 The what-if question . . . 395
18.2 Load scenarios . . . 398
18.3 Failure and weather scenarios . . . 401
18.4 Uncertainty in scenarios . . . 405

18.5 Actions and sensitivity . . . 409
18.6 Model credibility . . . 414
18.7 What to retain . . . 416
18.8 Exercises . . . 417
18.9 Connection to the TERRA framework . . . 418

19 Next-best measurement **420**
19.1 An ambiguous diagnosis . . . 420
19.2 Information gain . . . 423
19.3 Decision value . . . 427
19.4 Active tests and experiment design . . . 429
19.5 Sequential and question-specific measurement . . . 433
19.6 Limits and reporting . . . 436
19.7 What to retain . . . 438
19.8 Exercises . . . 440
19.9 Connection to the TERRA framework . . . 440

20 Data center optimization **442**
20.1 The optimization problem . . . 442
20.2 Constraints and objectives . . . 446
20.3 Workload, sites and storage . . . 448
20.4 Optimization under uncertainty . . . 452
20.5 Dynamics, discrete decisions, contingencies . . . 456
20.6 Optimization in operation . . . 460
20.7 What to retain . . . 462
20.8 Exercises . . . 463
20.9 Connection to the TERRA framework . . . 464

21 Grid interaction and flexible load **466**
21.1 Flexibility of a facility . . . 466
21.2 Electrical and IT resources . . . 470
21.3 Thermal, generation and portfolio . . . 473
21.4 Baselines and forecasts of flexibility . . . 477
21.5 Committing flexibility . . . 479
21.6 Reserve, signals and reporting . . . 483
21.7 What to retain . . . 487
21.8 Exercises . . . 487
21.9 Connection to the TERRA framework . . . 489

22 Digital twins, CFD and reduced models **490**
22.1 Three kinds of model . . . 490
22.2 Computational fluid dynamics . . . 492
22.3 Reduced-order network models . . . 496
22.4 Digital twins and model updating . . . 501
22.5 Discrepancy and calibration . . . 504
22.6 Fidelity matched to the decision . . . 508
22.7 What to retain . . . 512
22.8 Exercises . . . 512
22.9 Connection to the TERRA framework . . . 513

23 Data and public datasets **515**
23.1 Public data and physical inference 515
23.2 Workload, power and thermal datasets 517
23.3 The ideal dataset and data quality 521
23.4 Capability as a function of instrumentation 525
23.5 Ground truth and validation 529
23.6 Question-first use of datasets 531
23.7 What to retain 535
23.8 Exercises 536
23.9 Connection to the TERRA framework 537

24 TERRA for data centers **539**
24.1 Existing tools and the inference gap 539
24.2 Use cases in diagnosis and measurement 543
24.3 Capacity, risk and flexibility 546
24.4 Boundaries of the approach 550
24.5 Outputs for operators 553
24.6 An adoption path 556
24.7 What to retain 560
24.8 Exercises 560
24.9 The book in one page 562

Notation and vocabulary

Every symbol below is used with the same meaning in every chapter.

Symbol	Unit	Meaning
P_{IT}	W	Electrical power consumed by IT equipment (servers, storage, network).
P_{facility}	W	Total electrical power drawn by the facility.
P_{cool}	W	Electrical power consumed by the cooling system.
P_{rack}, P_{server}	W	Electrical power of one rack, of one server.
P_{idle}, P^{max}	W	Power of a server at no load and at full load.
u	–	Utilization of a processor or server, between 0 and 1; also the control input of a controller, by context.
$\dot{Q}$	W	Heat flow rate.
$\dot{Q}_{\mathrm{IT}}$, $\dot{Q}_{\mathrm{rack}}$	W	Heat generated by the IT equipment, by one rack.
$\dot{Q}_{\mathrm{avail}}$, $\dot{Q}_{\mathrm{req}}$	W	Cooling capacity available, cooling required.
$\dot{Q}_{\mathrm{evap}}$, $\dot{Q}_{\mathrm{cond}}$	W	Heat absorbed at a chiller's evaporator, rejected at its condenser.
W_{comp}	W	Compressor power of a chiller.
$\dot{m}$, $\dot{m}_{\mathrm{air}}$, $\dot{m}_w$	kg/s	Mass flow rate; of air; of water.
$\dot{V}$	m^3/s	Volumetric flow rate.
c_p, $c_{p,\mathrm{air}}$, $c_{p,w}$	$\mathrm{J/(kg\,K)}$	Specific heat; of air, about $1.0\,\mathrm{kJ/(kg\,K)}$; of water, about $4.18\,\mathrm{kJ/(kg\,K)}$.
T	K or °C	Temperature.
T_{in}, T_{out}	K	Air temperature entering and leaving a server or rack.
T_{supply}, T_{return}	K	Temperature of supply and return air or water.
T_{hot}	K	Temperature of hot exhaust air.
$T_{\mathrm{CHW,s}}$, $T_{\mathrm{CHW,r}}$	K	Chilled-water supply and return temperature.
T_{amb}, T_{wb}	K	Outdoor dry-bulb and wet-bulb temperature.
T_{set}, T_{max}	K	Setpoint of a controller; upper limit of a temperature.
RH	–	Relative humidity.
Δp	Pa	Pressure difference.
K	varies	Flow-resistance coefficient in $\Delta p = K\dot{V}^2$.
α	–	Recirculation fraction: share of rack inlet air that is hot exhaust.
β	–	Bypass fraction: share of supply air that returns without passing through IT equipment.
R_{th}	K/W	Thermal resistance.
C_{th}, C	J/K	Thermal capacitance.
τ	s	Time constant, $\tau = R_{\mathrm{th}}C_{\mathrm{th}}$.
UA	W/K	Overall conductance of a heat exchanger.
COP	–	Coefficient of performance of a chiller, $\dot{Q}_{\mathrm{evap}}/W_{\mathrm{comp}}$.

Symbol	Unit	Meaning
PUE	–	Power usage effectiveness, $P_{\text{facility}}/P_{\text{IT}}$.
N_{fan}, N_{pump}	rev/min	Rotational speed of a fan, of a pump.
η	–	Efficiency of a converter, pump or power path.
SOC, SOH	–	State of charge and state of health of a battery.
N, $N{+}1$, $2N$	–	Redundancy levels: the number of units required, one spare, two full independent sets.
x, u, d, θ	varies	State, control input, disturbance and parameter vectors of a dynamic model.
y, ϵ, b	varies	Measurement, its random error and its bias.
r	varies	Residual: measured value minus the value a model predicts.
H_i	–	A hypothesis about the cause of an observation.
H_0, H_M	–	The hypotheses of normal operation and of an inadequate model.
$\Pr(H_i \mid y)$	–	Posterior probability of a hypothesis given the measurements.
$p(x \mid y)$	varies	Posterior density of a state (or parameter) given the measurements.
$\hat{x}$, x_{MAP}	varies	An estimate of the state; the maximum a posteriori estimate.
w_k, v_k, K_k	varies	Process noise, measurement noise and Kalman gain at step k.
G	–	Topology: which component serves or feeds which.
$\mathcal{M}_j$	–	A candidate model structure.
$\mathcal{I}$	varies	Fisher information matrix of the parameters.
R_i, A	–	Reliability of a component over an interval; availability.
MTBF, MTTR	h	Mean time between failures; mean time to repair.
C, L, M	W	In Chapters 12 and 17 to 20: capacity, load and capacity margin, $M = C - L$.
C_{50}, C_{95}, C_{99}	W	Capacity exceeded with probability 50, 95 and 99 %.
c	–	A contingency: a specified failure or outage.
ϵ	–	In Chapters 12 to 20 also a risk tolerance, the largest accepted probability of a violation.
L_i, ℓ	varies	Loss or consequence of a hypothesis, scenario or action (ℓ where L is a load).
$\mathcal{R}$	varies	Expected risk: probability times consequence, summed over events.
S, a	–	A scenario; an action.
$\mathcal{H}$	bit	Entropy of a distribution over hypotheses.
IG, VOI	bit, cost	Expected information gain; expected value of information of a measurement.
Φ	–	Standard normal distribution function.
J	varies	Objective of an optimization (in Chapters 15 and 19, a sensitivity matrix).
t_{limit}	s	Time until a temperature reaches its limit at the present rate of rise.
$F^{\downarrow}$, $F^{\uparrow}$	W	Downward and upward flexibility: how far the net grid power can be lowered or raised from its baseline.

Symbol	Unit	Meaning
F_{95}	W	Firm flexibility: the flexibility available with 95 % probability.
$\Delta P(t)$	W	Flexibility trajectory: grid power minus its baseline during and after an event.
$\mathrm{CI}(t)$	kg/Wh	Carbon intensity of grid electricity.
$\mathbf{u}$, p, ρ, μ, k	varies	In Chapter 22: velocity field, pressure, density, viscosity and thermal conductivity of the air.
δ	varies	Model discrepancy: the systematic difference between a model and the system it represents.

Vocabulary

IT equipment Servers, storage and network equipment: the equipment that performs computation and whose power a data center exists to supply.

UPS Uninterruptible power supply: a converter with stored energy that keeps the IT load supplied when the upstream source fails.

PDU Power distribution unit: equipment that distributes and meters power to racks.

CRAH, CRAC Computer room air handler and computer room air conditioner: the units that cool the room air, with a chilled-water coil or with their own refrigeration cycle.

BMS, DCIM Building management system and data center infrastructure management: the supervisory software of the facility plant and of the data-center-specific infrastructure.

Measured, calculated, inferred The three kinds of statement this book keeps apart: a sensor reading, a quantity computed from readings by a known relation, and a conclusion about an unobserved state or cause.

Chapter 1

Data center system architecture: the flow of power, heat, information and control

1.1 The facility as an energy-conversion system

A data center is usually described by what it hosts: servers, storage, network equipment. This book describes it by what it does to energy. Electrical energy enters from the grid, passes through a chain of electrical equipment, is consumed by the IT equipment, and leaves the building as heat carried by air, water and finally the outdoor atmosphere. The question of this chapter is

where energy enters a data center, where it goes, what becomes heat, how that heat is removed, and what keeps the whole chain operating when a component fails.

The answer is a map of the whole facility, drawn at the level of detail that the later chapters fill in one subsystem at a time: the power path in Chapters 3 and 4, the IT load in Chapter 2, heat and air in Chapters 5 and 6, the cooling plant in Chapter 7, its dynamics in Chapter 8, sensors and controls in Chapters 9 and 10, and the efficiency metric in Chapter 11.

Jargon: IT load, UPS, PDU, CRAH, CRAC, chiller, cooling tower, BMS, DCIM, N+1

The *IT load* is the electrical power consumed by the information technology equipment, the servers, storage and network devices that perform the computation; it is the power the facility exists to supply, P_{IT}. A *UPS* (uninterruptible power supply) is a power converter with stored energy, usually batteries, that keeps the IT load supplied without interruption when the upstream source fails, for long enough for a generator to start or for the load to shut down in order. A *PDU* (power distribution unit) divides a feeder into the circuits that supply individual racks, usually with breakers and meters. A *CRAH* (computer room air handler) is a fan unit that cools room air by passing it over a coil of chilled water supplied from a central plant; a *CRAC* (computer room air conditioner) does the same job with its own refrigeration cycle and compressor inside the unit. A *chiller* is a machine that cools a water loop by moving heat out of it with a refrigeration cycle driven by an electric compressor. A *cooling tower* rejects heat from a water loop to the outdoor air, mostly by evaporating a small part of the water. The *BMS* (building management system) is the supervisory control and monitoring system of the mechanical and electrical plant; *DCIM* (data center infrastructure management) is the software that inventories, meters and monitors the data-center-specific infrastructure (racks, PDUs, power and environmental sensors). *N+1* is a redundancy level: if N units are needed to carry the load, $N + 1$ are installed, so that any one can fail or be taken out for maintenance without loss of capacity.

Key idea

A data center converts electrical energy into computation and, from the physical plant's point of view, essentially all of that electrical energy into heat. The facility is a coupled network that delivers power to the IT equipment, transports the resulting heat to the outdoor environment, watches and controls both flows, and duplicates the parts whose failure would stop them.

This section describes the three layers in which the facility is organized and states the energy balance on which the rest of the chapter rests.

1.1.1 Three layers of the facility

The facility can be drawn in three layers. The first is the *energy chain*: grid, electrical infrastructure, IT equipment, heat, cooling system, environment. Every joule that enters the building follows this chain, and its two ends are both outside the building: the energy arrives as electricity from the grid and leaves as heat to the atmosphere. The second layer *watches and controls* the first: sensors measure temperatures, powers, flows, pressures and humidity; controllers in the BMS and the DCIM compare those measurements with targets; and actuators (fans, pumps, valves, chiller staging) change the flows. The third layer gives *resilience*: uninterruptible power supplies and their batteries, standby generators, and redundant equipment of every kind, installed so that the first layer keeps working when a component of it fails.

Figure 1.1 draws the three layers. The middle row of boxes is the energy chain, read from left to right, with solid arrows for the flow of energy: electrical up to the IT equipment, thermal after it. The top row is the resilience layer; its arrows point to the parts of the energy chain each element protects. The bottom row is the sensing and control layer: a dashed path carries measurements from the energy chain to the sensors, the controls act on them, and a second dashed path carries commands from the actuators back to the cooling system. The figure shows each layer at one place for clarity; in a real facility sensors and actuators are distributed along the whole chain, and redundancy appears at almost every stage.

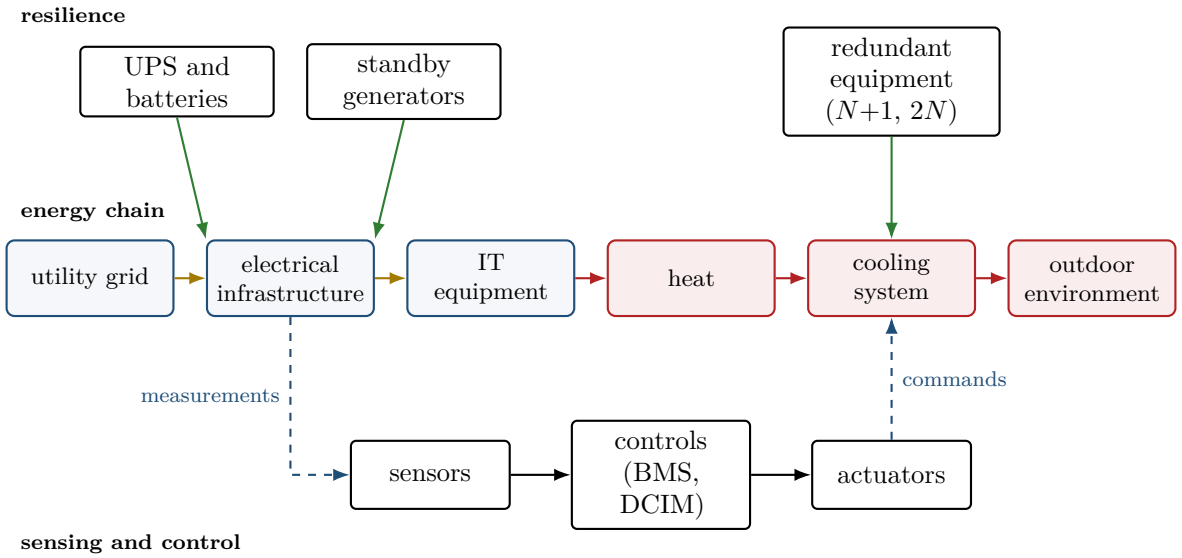


Figure 1.1: The three layers of a data center. Middle: the energy chain, electrical (gold arrows) up to the IT equipment and thermal (red arrows) after it. Top: the resilience layer and the parts of the chain it protects. Bottom: the sensing and control layer, drawn at one place for clarity although sensors and actuators are spread along the whole chain.

1.1.2 Electricity into computation and heat

What a data center does, physically, is convert electrical energy into computation and heat. The computation is the useful product, but it carries almost no energy away with it: the result of a calculation is stored as the states of transistors or sent out as signals whose power is minute compared with the power drawn to produce them. Almost all the electrical energy consumed by the IT equipment therefore ends as heat inside the building. For a server drawing P_{server},

$$\dot{Q}_{\text{server}} \approx P_{\text{server}}, \qquad \dot{Q}_{\text{IT}} \approx P_{\text{IT}}, \tag{1.1}$$

where $\dot{Q}_{\text{server}}$ is the heat flow the server releases and $\dot{Q}_{\text{IT}}$ the heat flow of all the IT equipment. A server at $P_{\text{server}} = 1\,\text{kW}$ releases about 1 kW of heat; the small exceptions, such as the optical power of signals leaving on fiber, the energy of the electrical signals on network cables and the sound of the fans, are fractions of a percent and are ignored. A room whose IT load is $P_{\text{IT}} = 1\,\text{MW}$ therefore produces about 1 MW of heat continuously, and over one day releases $1\,\text{MW} \times 24\,\text{h} = 24\,\text{MWh}$ of heat that must leave the building.

That statement fixes the job of the cooling system. Its job is not to "make the room cold". The room is at a steady temperature only when heat leaves it as fast as it is generated, and the cooling system is the set of equipment that *transports* heat from the IT equipment to the outdoor environment at the rate the IT equipment produces it. A cooling system that removed 0.95 MW from a 1 MW room would not keep it slightly warmer; it would let the room's temperature rise without limit, at a rate set by how much thermal mass the room contains (Section 1.6.4 estimates that rate). Thinking of cooling as transport, rather than as the production of cold, is what makes the rest of the facility readable as a set of conservation laws.

Principle 1.1. *Essentially all the electrical power consumed by IT equipment becomes heat, $\dot{Q}_{\text{IT}} \approx P_{\text{IT}}$. The cooling system does not produce cold; it transports that heat to the outdoor environment, and in steady operation it must carry it at the rate it is produced.*

1.2 The electrical and IT subsystems

Before it becomes heat, the energy passes through two subsystems: the electrical infrastructure that brings it from the grid to the servers, and the IT equipment that consumes it. This section follows the power path, states its two jobs, and describes how IT equipment is grouped from a server up to a data hall.

1.2.1 The power path from grid to server

Electrical power reaches a server through a chain of equipment:

$$\text{utility grid} \rightarrow \text{switchgear} \rightarrow \text{transformer} \rightarrow \text{UPS} \rightarrow \text{PDU} \rightarrow \text{rack} \rightarrow \text{server}.$$

The utility delivers power at medium voltage, typically between about 10 and 35 kV, depending on the country and the size of the site. *Switchgear* is the assembly of breakers, switches and protection relays that connects, disconnects and protects the incoming feeders and the circuits downstream. A *transformer* steps the voltage down to the low voltage used inside the building, typically 400 or 480 V three-phase. The UPS conditions the power and holds it up through interruptions, the PDU divides it among racks, and inside each rack power strips feed the individual servers, whose power supplies convert it to the low direct-current voltages, typically 12 V and below, that processors, memory and drives use. The voltages quoted are typical values; the arrangement of the chain is the same everywhere.

Figure 1.2 draws the chain. The top row of boxes is the path of power, left to right, from the utility to the server. The boxes below it are the elements that keep the path working when part of it fails: a second utility feeder and standby generators connected through transfer equipment at the switchgear, and batteries connected to the UPS. The small red arrows above the transformer, the UPS and the PDU mark the conversion losses of each stage, which leave as heat in the electrical rooms rather than in the data hall. Every stage of the path is a converter or a switch, and each stage is a place where power is measured, where it is lost, and where a failure can interrupt the flow.

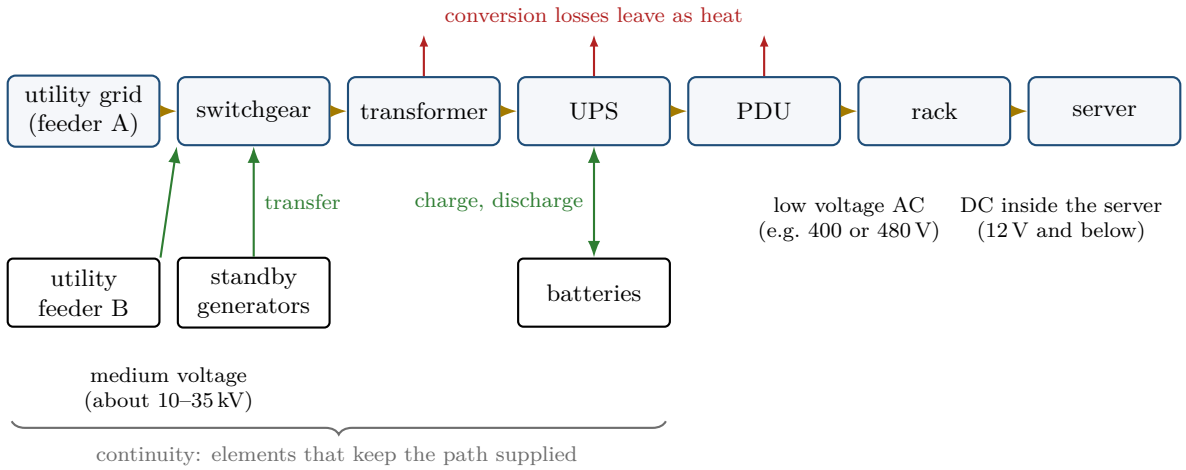


Figure 1.2: The electrical power path from the utility to a server (top row, gold arrows) and the equipment that keeps it supplied when part of it fails (bottom row, green arrows). Red arrows mark the conversion losses of the transformer, the UPS and the PDU. Voltages are typical values.

1.2.2 The two jobs of the electrical system

The electrical system has two jobs. The first is *delivery*: bring power at the right voltage and quality to every server, at whatever rate the servers draw it. The second is *continuity*: keep delivering when something upstream fails, whether that is the utility supply, a transformer, a breaker or a cable. A server does not tolerate even a short interruption: its power supply holds up its output for only a few tens of milliseconds, so an interruption longer than that restarts the server and loses the work in progress.

Continuity is the reason for most of the equipment in Figure 1.2 that is not on the direct path. The UPS and its batteries carry the load through the first seconds and minutes of an outage; batteries are commonly sized for a few minutes at full load. Standby generators, usually diesel, start when the utility fails and take over the load once they have reached speed, typically within ten to fifteen seconds. Redundant utility feeders, breakers that isolate a fault so that the rest of the system keeps running, and transfer equipment that switches the load from one source to another complete the arrangement. Chapter 3 follows the distribution path in detail and Chapter 4 the backup equipment; at the level of this chapter it is enough that the electrical subsystem is both a delivery network and a set of parallel paths kept ready for failure.

1.2.3 From server to data hall

The IT equipment is organized in a hierarchy:

$$\text{server} \rightarrow \text{rack} \rightarrow \text{cluster} \rightarrow \text{data hall}.$$

A *rack* is a standard cabinet, about 0.6 m wide and 1.0 to 1.2 m deep, holding a few to several dozen servers together with their network switches and power strips. A *cluster* is a group of racks that work together on the same workloads, and a *data hall* is the room, or a partitioned zone of a room, that contains many racks together with their cooling units. A rack consumes an electrical power P_{rack} and, by equation (1.1), releases a heat flow

$$\dot{Q}_{\text{rack}} \approx P_{\text{rack}}. \tag{1.2}$$

A rack at $P_{\text{rack}} = 40\,\text{kW}$ releases roughly 40 kW of heat that must be removed from the space it occupies.

The power of a rack matters not only through the total but through its concentration. Conventional enterprise racks have long drawn about 5 to 10 kW each; racks of accelerators for artificial-intelligence workloads draw 40 kW and more, and the densest designs exceed 100 kW. A 40 kW rack with a footprint of about $0.6 \times 1.2 = 0.72\,\text{m}^2$ releases about $40/0.72 \approx 56\,\text{kW}$ per square meter of rack footprint, the heat of a few dozen domestic electric heaters in a single cabinet. High density does not change the energy balance of the facility, but it concentrates the heat spatially, and the air or liquid that carries it away must be delivered to that one place at a correspondingly high rate. Section 1.3.1 quantifies the rate.

1.3 The path of heat

Once the electrical energy has become heat in the IT equipment, it travels through a sequence of media and machines to the outdoor environment. This section follows that path stage by stage: air, the room's airflow arrangement, the cooling unit, the chilled-water loop, the chiller, the condenser loop and the cooling tower.

1.3.1 Air as the first heat-transport medium

In a conventional air-cooled rack, cool air enters at the front of each server, the server's fans pull it across the processors, accelerators, memory and power supplies, and the air leaves at the

rear, warmer by the heat it has picked up. The air is the first medium that carries the heat away from the equipment. The heat it carries follows from the energy balance of the air stream: each kilogram of air that warms by one kelvin absorbs $c_{p,\mathrm{air}}$ joules, so a mass flow $\dot{m}_{\mathrm{air}}$ warming from T_{in} to T_{out} carries

$$\dot{Q} = \dot{m}_{\mathrm{air}}\, c_{p,\mathrm{air}}\, (T_{\mathrm{out}} - T_{\mathrm{in}}), \tag{1.3}$$

where $c_{p,\mathrm{air}} \approx 1.0\,\mathrm{kJ/(kg\,K)}$ is the specific heat of air. For $\dot{m}_{\mathrm{air}} = 2\,\mathrm{kg/s}$ and a temperature rise of $10\,\mathrm{K}$,

$$\dot{Q} = 2\,\mathrm{kg/s} \times 1.0\,\mathrm{kJ/(kg\,K)} \times 10\,\mathrm{K} = 20\,\mathrm{kW}.$$

With air at a density of about $1.2\,\mathrm{kg/m^3}$, $2\,\mathrm{kg/s}$ is about $1.7\,\mathrm{m^3/s}$ of air.

Read the other way, equation (1.3) states the airflow that a given heat requires. The $40\,\mathrm{kW}$ rack of Section 1.2.3, with the same $10\,\mathrm{K}$ rise, needs $\dot{m}_{\mathrm{air}} = 40/(1.0 \times 10) = 4\,\mathrm{kg/s}$, about $3.3\,\mathrm{m^3/s}$, through one cabinet; at $2\,\mathrm{kg/s}$ its air would have to warm by $20\,\mathrm{K}$ instead. The rise is limited by the temperatures the components tolerate and the air can be moved only so fast through a cabinet, which is why high-density racks press against the limits of air as a transport medium. Chapter 5 develops this balance and Chapter 6 the movement of the air.

1.3.2 Hot aisles and cold aisles

If racks were placed in random orientations, the hot exhaust of one rack would blow into the intake of the next. The intake air would be a mixture of cool supply and hot exhaust, the servers would run hotter for the same supply temperature, and the cooling units would receive return air that is cooler than the exhaust and therefore carries less heat per kilogram. This is inefficient, because the cooling system must move more air and supply it colder to achieve the same intake temperature, and potentially dangerous, because a server that breathes another server's exhaust can exceed its temperature limit. To prevent it, racks are arranged in rows with their fronts facing each other across a *cold aisle* and their backs facing each other across a *hot aisle*. The air then follows a loop,

$$\text{cooling unit} \rightarrow \text{cold aisle} \rightarrow \text{servers} \rightarrow \text{hot aisle} \rightarrow \text{cooling unit},$$

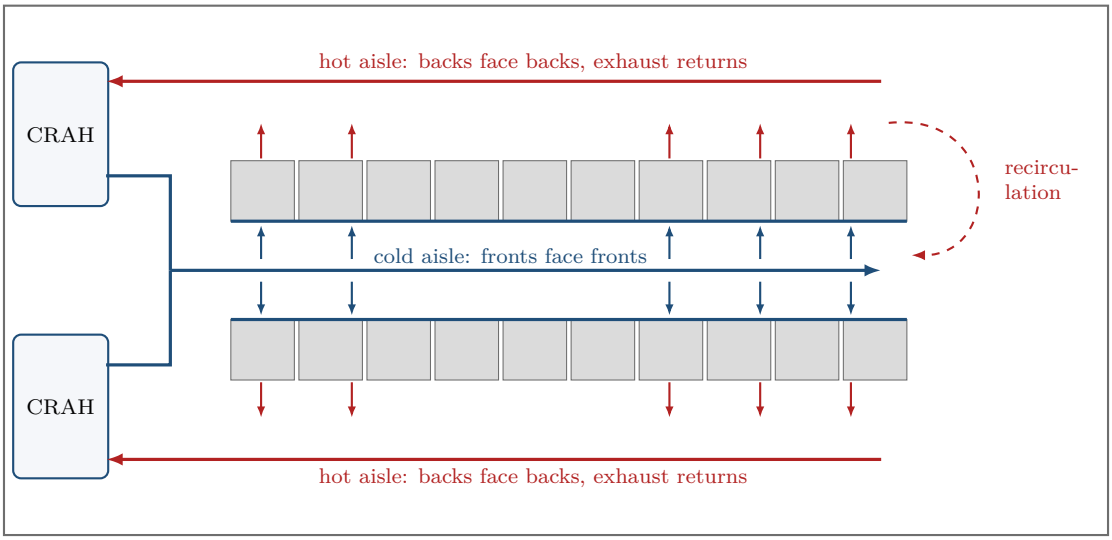
and the arrangement, standard in air-cooled facilities, reduces the mixing of supply and exhaust.

Figure 1.3 shows a plan view of such a room, seen from above. Two rows of racks face each other across the cold aisle in the middle; the hot aisles lie behind them, at the top and the bottom of the plan. Cooling units on the left wall push cool supply air (blue) into the cold aisle, the servers draw it through the racks, and the warm exhaust (red) collects in the hot aisles and returns to the cooling units. The dashed red path at the right end of the upper row is *recirculation*: exhaust that finds its way around the end of the row, or over the top of the racks, back into the cold aisle. The reverse leak, supply air that returns to the cooling units without passing through any server, is *bypass*. Chapter 6 quantifies both through the recirculation fraction α and the bypass fraction β. For an illustration of the cost of mixing: if $\alpha = 0.2$ of the air entering a server is exhaust at $37^{\circ}\mathrm{C}$ and the rest is supply at $22^{\circ}\mathrm{C}$, the server's intake is $0.8 \times 22 + 0.2 \times 37 = 25^{\circ}\mathrm{C}$, three kelvin warmer than the supply.

Air as an intermediary. The arrangement makes explicit a point that is easy to miss: the building's cooling system does not cool the server directly. It cools the air, and the air cools the server,

$$\text{server} \rightarrow \text{air} \rightarrow \text{cooling equipment}.$$

Air is a heat-transport medium between the two, and every property of the air path (how much air each rack receives, how much of it is mixed with exhaust, how much bypasses the servers) sits between the cooling equipment and the temperature the server experiences.



plan view; thick blue edge = rack fronts (air intakes); blue = cool supply air, red = warm exhaust

Figure 1.3: Hot-aisle and cold-aisle arrangement in plan view. Cooling units (left) supply cool air to the cold aisle between two rows of racks whose fronts face each other; servers draw it through and exhaust into the hot aisles behind the rows, from which it returns to the cooling units. The dashed path is recirculation of exhaust around the row end into the cold aisle.

1.3.3 The computer room air handler and air conditioner

The cooling unit at the end of the air loop takes warm return air from the hot aisle and delivers cool supply air to the cold aisle:

$$\text{warm return air} \rightarrow \text{CRAH or CRAC} \rightarrow \text{cool supply air.}$$

Inside a chilled-water CRAH, fans push the return air across a heat exchanger, a coil of finned tubes through which cold water flows. The air gives up heat to the metal of the coil and the metal gives it to the water, so the heat moves from the air to the chilled water and the air leaves the coil cooler. A CRAC achieves the same result with its own refrigeration cycle: the cold surface is the evaporator of a compressor circuit inside the unit, and the heat is rejected through a separate condenser, often outdoors.

With a chilled-water CRAH, the path of the heat has so far been

$$\text{server} \rightarrow \text{air} \rightarrow \text{water.}$$

The CRAH is the first point at which the heat changes medium inside the building. The air side of the unit carries the room's heat at low density and over short distances; the water side carries it in pipes over the long distance to the central plant. The fans of the CRAH also consume electricity, and that electricity becomes heat in the air stream that the coil must remove as well, a first example of the cooling system adding to the heat it handles.

1.3.4 The chilled-water loop

The water that leaves the CRAH coils circulates in a closed loop, the *chilled-water loop*, between the coils and the chillers. In a typical design the coil receives water at a supply temperature $T_{\mathrm{CHW,s}} = 7^{\circ}\mathrm{C}$ and returns it at $T_{\mathrm{CHW,r}} = 12^{\circ}\mathrm{C}$. The heat it carries follows from the same balance as for air, written for water,

$$\dot{Q} = \dot{m}_w \, c_{p,w} \, (T_{\mathrm{return}} - T_{\mathrm{supply}}) = \dot{m}_w \, c_{p,w} \, (T_{\mathrm{CHW,r}} - T_{\mathrm{CHW,s}}), \tag{1.4}$$

with $c_{p,w} \approx 4.18\,\mathrm{kJ/(kg\,K)}$ the specific heat of water. To carry the 1 MW of a 1 MW IT load with a 5 K rise, the loop needs

$$\dot{m}_w = \frac{1000\,\mathrm{kW}}{4.18\,\mathrm{kJ/(kg\,K)} \times 5\,\mathrm{K}} \approx 47.8\,\mathrm{kg/s},$$

about 48 liters per second. The same heat carried by air with a 10 K rise needs $\dot{m}_{\mathrm{air}} = 1000/(1.0 \times 10) = 100\,\mathrm{kg/s}$, about $83\,\mathrm{m^3/s}$. Per unit volume and per kelvin, water carries about $4180\,\mathrm{kJ/(m^3\,K)}$ and air about $1.2\,\mathrm{kJ/(m^3\,K)}$, a ratio of about 3500; that ratio is why heat is moved over long distances by water and why the densest racks bring liquid close to the chips.

At the rack the heat flow is written $\dot{Q} = \dot{m}_{\mathrm{air}} c_{p,\mathrm{air}}\,\Delta T_{\mathrm{air}}$; at the loop it is written $\dot{Q} = \dot{m}_w c_{p,w}\,\Delta T_w$. They are two measurements of one quantity. Aside from losses, the heat added by fans and pumps, small gains through the building envelope, and heat stored while temperatures change, the heat released by the IT equipment, the heat carried by the air, and the heat carried by the chilled water are the same flow:

$$\dot{Q}_{\mathrm{IT}} \approx \dot{m}_{\mathrm{air}} c_{p,\mathrm{air}}\,(T_{\mathrm{out}} - T_{\mathrm{in}}) \approx \dot{m}_w c_{p,w}\,(T_{\mathrm{CHW,r}} - T_{\mathrm{CHW,s}}). \tag{1.5}$$

Equation (1.5) is the beginning of a physics-constrained model of the facility. Each of its three members can be computed from different meters: electrical power, air temperatures and flows, water temperatures and flows. Because they must agree, a disagreement among them is information, either about a meter that is wrong or about a heat path that has not been accounted for.

Principle 1.2. *The heat released by the IT equipment, the heat carried by the air and the heat carried by the chilled water are the same flow, up to losses, added fan and pump heat, and storage. Computed from independent meters, they must agree, and their disagreement is a measurement.*

1.3.5 The chiller

The chilled-water loop delivers warm water to the chiller and receives cold water back:

$$\text{warm chilled water } (12^\circ\mathrm{C}) \rightarrow \text{chiller} \rightarrow \text{cold chilled water } (7^\circ\mathrm{C}).$$

The chiller does not destroy heat; no machine does. It moves heat from the chilled-water loop to a second, warmer water loop, the condenser loop, and since heat does not flow by itself from a colder body to a warmer one, moving it in that direction requires work. The chiller is a vapor-compression refrigeration machine: a refrigerant evaporates at low pressure in a heat exchanger called the *evaporator*, absorbing the heat $\dot{Q}_{\mathrm{evap}}$ from the chilled water; an electric compressor raises the refrigerant's pressure, and with it its condensing temperature, using the power W_{comp}; and the refrigerant condenses at high pressure in a second heat exchanger, the *condenser*, releasing the heat $\dot{Q}_{\mathrm{cond}}$ into the condenser water. Energy conservation for the refrigerant loop in steady operation gives

$$\dot{Q}_{\mathrm{cond}} = \dot{Q}_{\mathrm{evap}} + W_{\mathrm{comp}}, \qquad \mathrm{COP} = \frac{\dot{Q}_{\mathrm{evap}}}{W_{\mathrm{comp}}}, \tag{1.6}$$

where COP, the coefficient of performance, is the heat moved per unit of compressor work. The heat to be rejected outdoors is the IT heat plus the chiller's own work.

With $\dot{Q}_{\mathrm{evap}} = \dot{Q}_{\mathrm{IT}} = 1\,\mathrm{MW}$ and $W_{\mathrm{comp}} = 200\,\mathrm{kW}$, the chiller has $\mathrm{COP} = 1000/200 = 5$, and the condenser must reject

$$\dot{Q}_{\mathrm{cond}} = 1.0\,\mathrm{MW} + 0.2\,\mathrm{MW} = 1.2\,\mathrm{MW},$$

more heat than the servers generated. The compressor's electricity is one of the facility's largest loads after the IT equipment itself, and it enters the building's heat budget twice: once on the electricity bill and once as heat that the rest of the plant must carry outdoors. Chapter 7 describes the chiller and the factors that set its COP.

1.3.6 The cooling tower and the hierarchy of loops

In a water-cooled plant the condenser water carries the 1.2 MW from the chillers to cooling towers outdoors. In an illustrative design it leaves the chiller condensers at 35°C and returns from the towers at 30°C, so by equation (1.4) its flow is $1200/(4.18 \times 5) \approx 57.4\,\mathrm{kg/s}$, larger than the chilled-water flow because it carries the compressor work as well. In the tower the warm water is sprayed over a fill material through which fans draw outdoor air, and a small part of the water evaporates. Evaporation absorbs about 2400 kJ per kilogram of water, so rejecting 1.2 MW mostly by evaporation consumes about $1200/2400 = 0.5\,\mathrm{kg/s}$ of water, roughly $43\,\mathrm{m}^3$ per day, which is replaced as make-up water. Because the tower cools by evaporation, the lowest temperature it can reach is set by the outdoor wet-bulb temperature T_{wb}, not by the dry-bulb temperature T_{amb}. Plants without towers reject the condenser heat through air-cooled condensers or dry coolers instead, at the price of higher condensing temperatures on hot days.

The complete path of the heat in a water-cooled plant is a chain of loops,

$$\text{server} \to \text{air} \to \text{chilled water} \to \text{chiller} \to \text{condenser water} \to \text{cooling tower} \to \text{atmosphere},$$

a hierarchy of heat-transfer loops, each of which picks up heat from the loop before it and hands it to the loop after it, until the last one rejects it outdoors. Each loop has its own fluid, its own flow, its own pumps or fans, and its own temperature range.

Figure 1.4 draws the chain. The boxes are the stages the heat passes through, read along the arrows: left to right along the top row from the chip to the chilled-water loop, then right to left along the bottom row from the chiller to the atmosphere. Under each box is its illustrative temperature level. The brace above the top row marks the stages that carry the IT heat, 1.0 MW; the box at the lower right is the compressor power that joins the flow in the chiller; and the brace below the bottom row marks the stages that carry 1.2 MW. The figure shows two facts at once: the heat flow is the same at every stage until the chiller adds its work, and the temperature falls from the chip to the chilled water and then is raised by the chiller to a level above the outdoor wet-bulb temperature, so that it can flow outdoors.

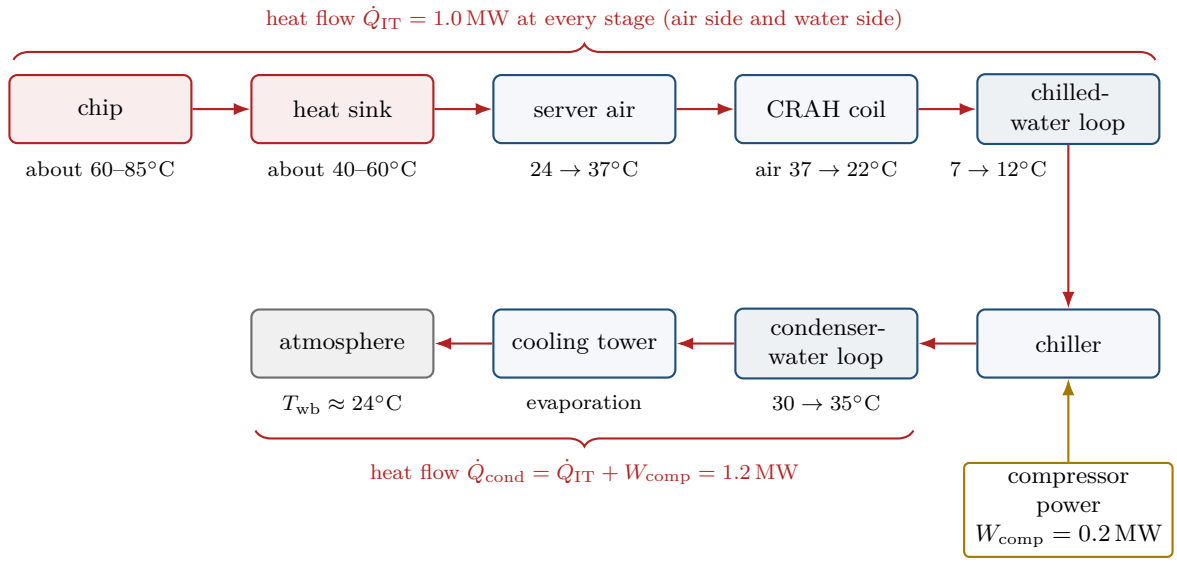


Figure 1.4: The heat-rejection chain of a water-cooled data center, from chip to atmosphere, with illustrative temperature levels under each stage. The IT heat, 1.0 MW, is the same at every stage up to the chiller; the chiller adds its compressor work, 0.2 MW, and the condenser loop, tower and atmosphere carry 1.2 MW (equation (1.6)).

1.3.7 Following one joule

Following a single joule through the facility summarizes the section. One joule of electricity enters a server and is consumed in its transistors, becoming about one joule of heat. That joule moves by conduction from the chip into its heat sink, by convection from the heat sink into the air the server's fans draw through, with the air into the hot aisle and the CRAH, across the CRAH coil into the chilled water, with the chilled water to the chiller's evaporator, through the refrigerant to the condenser, where it is joined by the compressor's work, into the condenser water, and with it to the cooling tower, where it leaves in evaporated water and warmed air to the atmosphere:

chip → heat sink → air → CRAH coil → chilled water → chiller → condenser loop → outside.

Every step of that path except one moves heat from a warmer place to a colder one, which happens by itself provided there is a path and a flow. The one exception is the chiller, which lifts the heat from about 7–12°C to 30–35°C so that it can flow to an atmosphere at a wet-bulb temperature of 24°C. The data center is therefore an energy-transport network: a set of paths along which a conserved quantity, energy, flows from a source to a sink, driven by temperature differences and moved by fans, pumps and one compressor. Each stage has a capacity, the largest flow it can carry at the temperatures available to it, and the network as a whole can carry no more heat than its most constrained stage.

1.4 Three coupled networks and the control loop

The energy chain of the previous sections is one network. A data center contains three, and they are coupled to each other. This section names the three, describes the feedback loop by which control keeps the thermal network within limits, and traces how a change in workload propagates through all of them.

1.4.1 Electrical, thermal-fluid and information networks

The *electrical network* runs from the grid through switchgear, UPS and PDU to the racks. Its quantity is power P, in kilowatts, and its laws are those of circuits: current divides at junctions and power is conserved up to conversion losses. The *thermal-fluid network* runs from the racks through the air to the water and the environment. Its quantities are temperatures T, mass flows $\dot{m}$, heat flows $\dot{Q}$ and pressures p, and its laws are conservation of mass for the air and water, conservation of energy for the heat, and the pressure-flow relations of ducts, pipes, fans and pumps. The *information and control network* runs from the sensors to the controllers to the actuators. Sensors report temperatures, electrical powers, flows, relative humidity RH, pressure differences Δp and equipment states (running, stopped, alarmed); controllers compute commands; actuators manipulate fan speed, pump speed, valve position and chiller staging, the number of chillers running. Its quantities are signals, not energy, and its connections are links between devices rather than pipes or cables carrying power.

Figure 1.5 draws the three networks as stacked bands. The top band is the electrical network, read from left to right; the middle band is the thermal-fluid network, read from right to left, from the rack to the environment; the bottom band is the information and control network. The vertical solid arrow on the right is the main coupling between the top two: the IT power becomes the IT heat. The dashed gold arrows are the second coupling: the electrical network also supplies the cooling equipment. The dashed blue arrows connect the bottom band to the others: measurements flow down into the sensors, commands flow up from the actuators to the fans, pumps, valves and chillers. The three networks share nodes, a rack is a node of all three, and none can be analyzed alone: the plant, the sensors, the controllers and the actuators form one loop,

physical plant → sensors → controllers → actuators → physical plant.

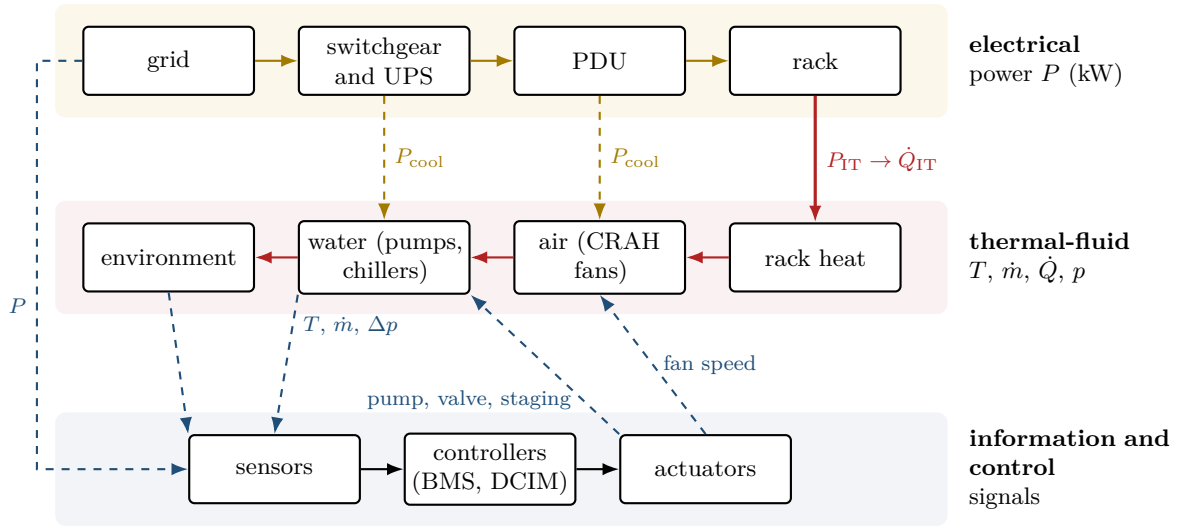


Figure 1.5: The three coupled networks of a data center. The electrical network (top) delivers power; the thermal-fluid network (middle) carries heat from the racks to the environment; the information and control network (bottom) measures both and acts on the cooling equipment. The solid red arrow is the conversion of IT power into heat; the dashed gold arrows are the electrical supply of the cooling equipment.

1.4.2 The control loop

The networks are coupled most visibly through control. Suppose a rack's inlet temperature rises to 29°C, above an illustrative setpoint of 27°C, the upper end of the inlet range commonly recommended for air-cooled IT equipment. The inlet has risen because the cold aisle in front of the rack is receiving less cool air than the servers draw, so part of their intake is recirculated exhaust. A controller in the BMS compares the measurement with the setpoint and raises the speed command u of the CRAH fans. The fans deliver more air, $\dot{m}_{\text{air}}$ rises, the cold aisle fills with supply air and displaces the recirculated exhaust, and the coil's capacity to carry heat, $\dot{Q} = \dot{m}_{\text{air}} c_{p,\text{air}} \Delta T$, rises with the flow. The inlet temperature falls toward the setpoint, the controller sees the smaller error and stops raising the fan speed. The chain

$$T \to \text{sensor} \to \text{controller} \to \text{fan} \to \dot{m}_{\text{air}} \to T$$

is a feedback loop: the output of the process, a temperature, is measured and used to adjust the input that drives it.

Figure 1.6 draws the loop as a block diagram. Signals travel along the arrows. From the left, the setpoint T_{set} enters a comparison (the circle), which subtracts the measured temperature y and passes the error e to the controller. The controller's output is the fan-speed command u; the CRAH fan turns it into an airflow $\dot{m}_{\text{air}}$; the cold aisle and rack, whose heat balance is the process being controlled, turn the airflow into an inlet temperature T_{in}. The disturbances d, drawn entering the process from above, are the inputs nobody commands: the IT load, recirculation paths and the supply-air temperature. The lower path is the feedback: the sensor measures T_{in} and returns the measurement y to the comparison. The loop acts on the error, so it corrects the effect of any disturbance without needing to know which one occurred.

The state form. The loop is one instance of a general model. Writing the facility's physical state as a vector x (temperatures of air volumes, coils, water loops and equipment masses; the

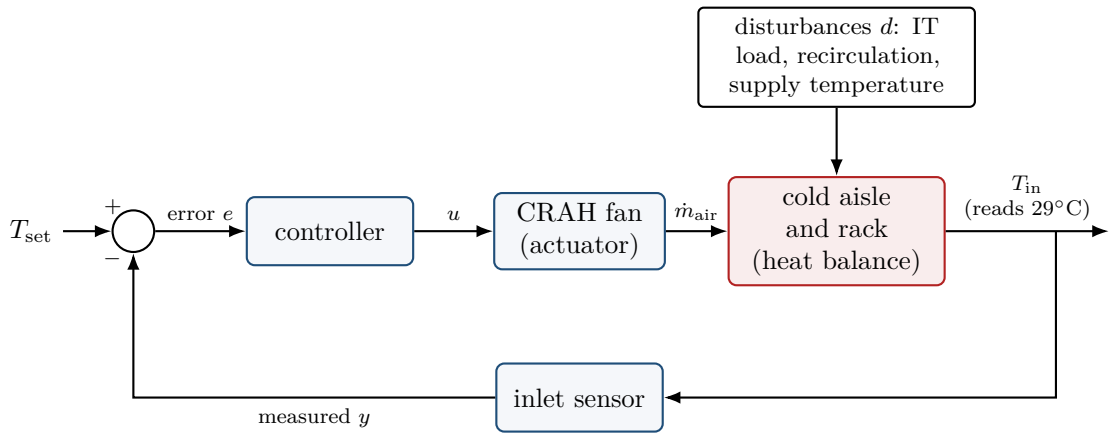


Figure 1.6: The rack-inlet temperature control loop. The controller acts on the error between the setpoint and the measured inlet temperature and commands the CRAH fan speed; the airflow changes the inlet temperature through the heat balance of the cold aisle and rack, which disturbances also drive. Values are illustrative: $T_{\text{set}} = 27°\text{C}$, measured inlet 29°C.

state of charge of the UPS batteries), its controllable inputs as u (fan speeds N_{fan}, pump speeds N_{pump}, valve positions, chiller staging), the inputs nobody controls as d (IT load, outdoor temperatures T_{amb} and T_{wb}), and the physical parameters that are not known exactly as θ (coil conductance UA, recirculation fraction α, thermal resistances, sensor biases), the facility obeys

$$\dot{x} = f(x, u, d, \theta), \qquad y = g(x, \theta) + b + \epsilon, \tag{1.7}$$

where the first equation is the physics, conservation of energy and mass written for each node, and the second says what the sensors report: a function of the state, plus a bias b and a random error ϵ. The controller sees only y. The modeler wants x and θ. The gap between them is the subject of the second half of this book.

1.4.3 Disturbances and the coupling of cooling back into load

A disturbance propagates through all three networks. When the workload rises, the IT power rises, the IT heat rises with it by equation (1.1), rack temperatures begin to rise, the cooling controls respond by raising the air or water flow, and the cooling equipment draws more electricity:

$$\text{workload} \uparrow \to P_{\text{IT}} \uparrow \to \dot{Q}_{\text{IT}} \uparrow \to T_{\text{rack}} \uparrow \to \dot{m}_{\text{air}}, \dot{m}_w \uparrow \to P_{\text{cool}} \uparrow.$$

Because cooling itself consumes electricity, the chain closes on itself:

$$\text{compute demand} \to \text{electrical load} \to \text{heat} \to \text{cooling demand} \to \text{electrical load}.$$

This is a feedback-like coupling between the electrical and thermal networks. It does not run away, because the cooling power is a fraction of the heat it removes, but it amplifies every change in IT load: the facility's electrical demand moves by more than the IT load does.

The facility's electrical demand is the sum of the IT load and everything that supports it,

$$P_{\text{facility}} = P_{\text{IT}} + P_{\text{cool}} + P_{\text{loss}} + P_{\text{other}}, \tag{1.8}$$

where P_{loss} is the conversion loss of the power path (transformers, UPS, PDUs, cables) and P_{other} is lighting and auxiliary loads. As an illustration, a facility with $P_{\text{IT}} = 1.00\,\text{MW}$ and $P_{\text{facility}} = 1.45\,\text{MW}$ might divide its 0.45 MW of overhead into 0.20 MW of chiller compressor power, 0.15 MW of fans, pumps and tower fans, 0.08 MW of electrical losses and 0.02 MW of

lighting and other loads. If the IT load rises by 10 %, to 1.10 MW, and the cooling power scales in proportion to the heat it removes, P_{cool} rises from 0.35 to about 0.385 MW; the conversion losses also grow, so the facility demand rises by more than the 0.10 MW of IT load. Fan power rises faster than in proportion when the airflow must rise, because it grows roughly with the cube of fan speed (Chapter 6), and every watt of fan and pump power ends as heat in the air or water that the plant must remove as well. The ratio $P_{\text{facility}}/P_{\text{IT}}$, here 1.45, is the facility's power usage effectiveness, the subject of Chapter 11.

1.5 Redundancy and measurement

The resilience layer and the measurement layer of Figure 1.1 each change how the facility must be described. This section treats redundancy as parallel paths and states the question it answers as a probability, then lists what a facility measures, what it does not, and how the two are related.

1.5.1 Parallel paths and redundancy levels

Reliability adds parallel paths to every network. The electrical network has UPS A and UPS B, each feeding one of the two power cords of a dual-corded server; the cooling plant has chillers 1, 2 and 3 and often a fourth; the air side has more CRAH units than the load needs. The redundancy level is named by comparison with N, the number of units required to carry the design load. A design at N installs exactly the required units; any failure reduces capacity below the load. A design at $N{+}1$ installs one spare, so any single unit can fail or be taken out for maintenance. A design at $2N$ installs two complete, independent sets, each able to carry the whole load, so an entire set can be lost, as in the A and B electrical paths.

Table 1.1 works one example. The required cooling is 3 MW and each chiller provides 1 MW, so $N = 3$. The table lists, for each design, the number of chillers installed, the installed capacity, and the capacity left after one and after two chillers fail. Only the N design falls short after one failure; $N{+}1$ survives one failure exactly at the required capacity and falls short after a second; $2N$ survives three failures. The same counting appears in contingency analysis of a transmission system, where a network is planned to carry its load after the loss of any one element.

Table 1.1: Redundancy levels for a required cooling of 3 MW provided by 1 MW chillers. Capacities below the requirement are marked as short.

Design	Chillers installed	Installed capacity	Capacity, one unit out	Capacity, two units out
N	3	3 MW	2 MW (short)	1 MW (short)
$N{+}1$	4	4 MW	3 MW	2 MW (short)
$2N$	6	6 MW	5 MW	4 MW

1.5.2 Capacity adequacy as a probability

Table 1.1 answers a deterministic question: whether the installed capacity exceeds the design load after a stated number of failures. The operational question is different. The IT load is uncertain and varies by hour; each chiller's capacity falls on hot days, when its condenser runs warmer; equipment degrades; and at any moment one chiller may already be out for maintenance. The question is then not whether capacity exceeds load, but how likely it is that it does not:

$$\Pr(\dot{Q}_{\text{avail}} < \dot{Q}_{\text{req}}), \tag{1.9}$$

the probability that the available cooling capacity $\dot{Q}_{\text{avail}}$ is below the required cooling $\dot{Q}_{\text{req}}$, given the uncertainty in the IT load, the outdoor temperature, the condition of the equipment and the units unavailable.

An illustrative calculation shows how the probability differs from the count. Let each chiller of the $N{+}1$ plant be available, independently of the others, with probability 0.98. On a mild day each chiller delivers its full 1 MW, and the plant falls short only if fewer than three of the four are available:

$$\Pr(\text{short} \mid \text{mild}) = 1 - \left[0.98^4 + 4 \times 0.98^3 \times 0.02\right] = 1 - (0.9224 + 0.0753) = 0.0023.$$

On a hot day, suppose each chiller's capacity falls to 0.9 MW. Three chillers now give 2.7 MW, less than the 3 MW required, so the plant falls short unless all four are available:

$$\Pr(\text{short} \mid \text{hot}) = 1 - 0.98^4 = 0.0776.$$

If hot conditions occur in 10 % of hours, the probability of a shortfall in a randomly chosen hour is $0.9 \times 0.0023 + 0.1 \times 0.0776 \approx 0.0099$, about four times the mild-day value, and almost all of it comes from hot hours. For comparison, an N plant of three chillers on a mild day falls short with probability $1 - 0.98^3 = 0.0588$. The $N{+}1$ label is the same in both weathers; the risk is not. The numbers are illustrative, and the independence assumption is optimistic, because failures can share a cause (a common electrical supply, a heat wave), a point a later chapter on reliability develops.

Principle 1.3. *Redundancy is a count of parallel units; adequacy is a probability. The operational question is* $\Pr(\dot{Q}_{\text{avail}} < \dot{Q}_{\text{req}})$ *under uncertain load, weather, degradation and unavailable units, and an* $N{+}1$ *design can carry very different risks in different conditions.*

1.5.3 What is measured and what is not

A facility's monitoring system reports many quantities, but far from all of those that matter. Table 1.2 lists an illustrative set of readings a facility might have, followed by quantities calculated from them. The first column names the quantity, the second gives its symbol, the third its value, and the fourth says whether the value is measured or calculated, and from what. The six measured values are the facility power $P_{\text{facility}} = 1.45\,\text{MW}$, the IT power $P_{\text{IT}} = 1.00\,\text{MW}$, a rack inlet temperature of 24°C, the same rack's outlet temperature of 37°C, and chilled-water supply and return temperatures of 7 and 12°C.

The list of what the facility does not know is longer. It typically does not know the airflow through every rack, the amount of heat recirculation in each aisle, the effectiveness of each cooling coil, whether every sensor is unbiased, the true cooling capacity of the plant as it stands today, the thermal resistances between components (chip to heat sink, heat sink to air), or where energy that no submeter records actually goes. These are exactly the quantities that decide whether the facility is operating safely and efficiently, and none of them is read from a single sensor.

1.5.4 Measured, calculated and inferred

Three kinds of statement can be made about the facility, and they are very different claims. A *measured* statement is a sensor reading: $T_{\text{in}} = 24°\text{C}$. It can be wrong, through bias, drift or a misplaced sensor, but it is a direct report. A *calculated* statement applies a known relation to measured values: $\dot{Q} = \dot{m}_{\text{air}} c_{p,\text{air}} (T_{\text{out}} - T_{\text{in}})$, or the 1.54 kg/s airflow of Table 1.2. Its validity depends on the relation and on the assumptions behind it; the airflow calculation assumes that

Table 1.2: An illustrative set of facility readings and the quantities calculated from them. A calculated value is only as good as the relation and the assumption it rests on.

Quantity	Symbol	Value	Kind
Facility power	P_{facility}	1.45 MW	measured (utility meter)
IT power	P_{IT}	1.00 MW	measured (UPS or PDU meters)
Rack inlet air temperature	T_{in}	24°C	measured
Rack outlet air temperature	T_{out}	37°C	measured
Chilled-water supply temperature	$T_{\text{CHW,s}}$	7°C	measured
Chilled-water return temperature	$T_{\text{CHW,r}}$	12°C	measured
Facility overhead	$P_{\text{facility}} - P_{\text{IT}}$	0.45 MW	calculated by subtraction
Power usage effectiveness	PUE	1.45	calculated, $P_{\text{facility}}/P_{\text{IT}}$
Rack air temperature rise	$T_{\text{out}} - T_{\text{in}}$	13 K	calculated by subtraction
Chilled-water temperature rise	$T_{\text{CHW,r}} - T_{\text{CHW,s}}$	5 K	calculated by subtraction
Chilled-water flow	$\dot{m}_w$	47.8 kg/s	calculated from eq. (1.4), assuming all IT heat reaches the water
Rack airflow (rack at 20 kW)	$\dot{m}_{\text{air}}$	1.54 kg/s	calculated from eq. (1.3), assuming all rack heat leaves in the measured air stream

the whole 20 kW leaves the rack in the air stream whose inlet and outlet were measured, and that the two sensors represent the average temperatures of that stream. An *inferred* statement is a conclusion about something no sensor observes, reached by weighing the measurements against models of what could have produced them: "the airflow through this rack is probably inadequate", or

$$\Pr(\text{temperature sensor biased} \mid \text{data}) = 0.65.$$

An inferred statement carries its uncertainty with it, and is revised when new data arrive.

Keeping the three apart is a discipline the rest of the book enforces. The chilled-water flow of 47.8 kg/s in Table 1.2 looks like a fact, but it is a calculation that assumes the water carries all the IT heat; if a flow meter on the loop reads 40 kg/s, the disagreement does not show which of the two is wrong. It shows that one of the meters, or the assumption, is wrong, and choosing among those explanations is an inference.

1.6 Diagnosis as ranked hypotheses and the picture to keep

The coupling of networks, the gaps in measurement and the uncertainty in capacity make a data center a natural object for probabilistic reasoning. This section shows how a single alarm becomes a ranked list of explanations and a choice of the next measurement, then assembles the chapter's picture of the facility, contrasts it with a battery, and states the takeaway.

1.6.1 One symptom, six hypotheses

Suppose an alarm reports that a rack's inlet temperature is high: it reads 29°C where it normally reads about 24°C. Several physical causes produce that symptom, and Table 1.3 lists six. The first column names the hypothesis, the second describes its mechanism, and the third gives the signature it would leave in other measurements, which is what makes the hypotheses distinguishable.

Table 1.3: Six hypotheses for a high rack inlet temperature, with the mechanism of each and the evidence that would distinguish it.

Hypothesis	Mechanism	Signature in other measurements
H_1 higher IT load	rack and neighbors draw more power, so more exhaust and more recirculation	rack PDU power rises; outlet temperature rises
H_2 reduced airflow	less supply air reaches the rack (blocked tiles, a slowed or failed CRAH fan)	low supply pressure or tile airflow at the rack; nearby racks also warm
H_3 air recirculation	exhaust returns to the intake around the row end or over the rack (missing blanking panels, gaps)	supply airflow normal; inlet hottest at the top of the rack or the end of the row
H_4 CRAH degradation	coil fouled or valve faulty, so supply air leaves warmer	CRAH supply-air temperature rises; all racks served by that unit warm
H_5 chilled-water problem	warm or insufficient chilled water at the coils	chilled-water supply temperature rises or loop flow falls
H_6 bad sensor	the inlet sensor is biased or has failed	neighboring sensors and a portable reference read normal

An operator facing the alarm wants the probability of each hypothesis given the data y, $\Pr(H_i \mid y)$. By Bayes' rule it combines how plausible each hypothesis was before the alarm,

$\Pr(H_i)$, with how well each explains what was observed, $\Pr(y \mid H_i)$:

$$\Pr(H_i \mid y) = \frac{\Pr(y \mid H_i)\,\Pr(H_i)}{\sum_j \Pr(y \mid H_j)\,\Pr(H_j)}. \tag{1.10}$$

An illustrative result is the posterior plotted in Figure 1.7: air recirculation 42 %, reduced airflow 31 %, sensor bias 15 %, higher IT load 8 %, CRAH problem 4 %, and a chilled-water problem close to 0 %. The ranking follows from the other readings: the chilled-water supply is on its design value of 7°C, which nearly rules out H_5; the CRAH supply air is normal, which makes H_4 unlikely; the rack's PDU power has risen only slightly, which leaves H_1 a minor share; and the inlet sensors of the neighboring racks are only partly elevated, which keeps a sensor fault in contention. The horizontal axis of the figure lists the six hypotheses in order of probability, and the vertical axis is the posterior probability in percent.

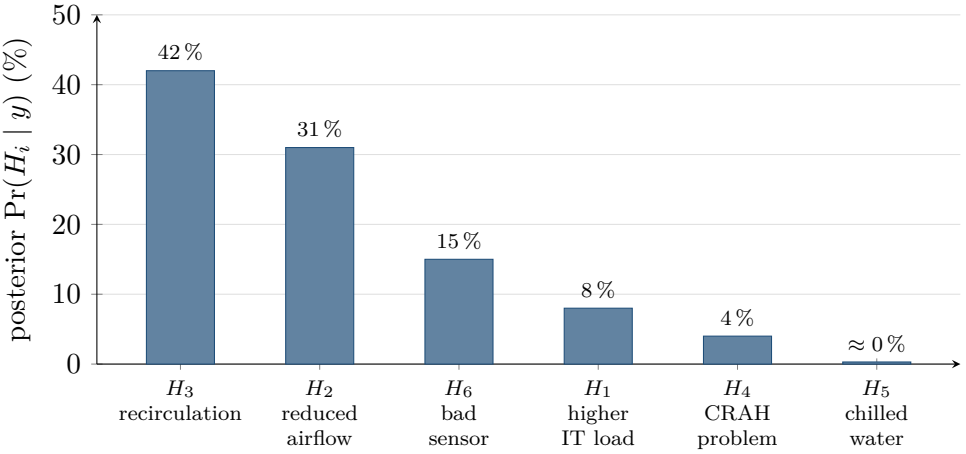


Figure 1.7: Illustrative posterior probabilities of six hypotheses for a high rack inlet temperature, ranked. The two leading hypotheses, recirculation and reduced airflow, together hold 73 % of the probability, and the next measurement is chosen to separate them.

1.6.2 The next measurement

The posterior answers "what is probably wrong"; the operator also needs "what to measure next". The two leading hypotheses differ in one clear signature (Table 1.3): reduced airflow predicts a low supply of air at the rack, recirculation predicts a normal supply with exhaust reaching the intake. A measurement of the supply airflow at the rack, for example the pressure difference between the raised-floor plenum and the cold aisle at the rack's perforated tiles, separates them. Suppose, illustratively, that this reading comes out low with probability 0.85 under H_2, 0.15 under H_3, 0.30 under H_4 (a CRAH fan fault lowers the pressure) and 0.10 under each of the others.

Table 1.4 applies equation (1.10) to both possible outcomes. Its columns are the current posterior, the probability of a low reading under each hypothesis, and the posterior after a low and after a normal reading. A low reading is expected with probability $\sum_j \Pr(\text{low} \mid H_j)\,\Pr(H_j) = 0.36$ and a normal one with probability 0.64. Either outcome reorders the leading pair decisively: a low reading moves reduced airflow from 31 to 73 %; a normal reading moves recirculation from 42 to 56 % and reduced airflow down to 7 %. The ratio of the two leading probabilities, $42/31 \approx 1.35$ before the measurement, becomes 0.24 or 7.7 after it. That is what makes the measurement worth taking: its value lies in how much it is expected to change the ranking. A measurement that both hypotheses predict equally, such as the chilled-water temperatures, would leave the posterior where it is, however precise it was. The sensor-bias

Data Center Systems

These are the opening pages. The complete book runs 575 pages.

The full PDF opens with a token. Write to **rmarquez123@gmail.com**, name the book, and I will send you one.

One token opens every book on the site.

Ricardo Marquez, Ph.D.

ricardo-marquez.com