Chapter 6
Warehouse-Scale Computers to Exploit Request-Level and Data-Level Parallelism

Approx. number events in 1st year	Cause	Consequence
1 or 2	Power utility failures	Lose power to whole WSC; doesn't bring down WSC if UPS and generators work (generators work about 99% of time).
4	Cluster upgrades	Planned outage to upgrade infrastructure, many times for evolving networking needs such as recabling, to switch firmware upgrades, and so on. There are about nine planned cluster outages for every unplanned outage.
1000s	Hard-drive failures	2%–10% annual disk failure rate (Pinheiro et al., 2007)
	Slow disks	Still operate, but run 10 × to 20 × more slowly
	Bad memories	One uncorrectable DRAM error per year (Schroeder et al., 2009)
	Misconfigured machines	Configuration led to ~30% of service disruptions (Barroso and HÖlzle, 2009)
	Flaky machines	1% of servers reboot more than once a week (Barroso and HÖlzle, 2009)
5000	Individual server crashes	Machine reboot; typically takes about 5 min (caused by problems in software or hardware).

Figure 6.1 List of outages and anomalies with the approximate frequencies of occurrences in the first year of a new cluster of 2400 servers. We label what Google calls a cluster an *array*; see Figure 6.5. Based on Barroso, L.A., 2010. Warehouse Scale Computing [keynote address]. In: Proceedings of ACM SIGMOD, June 8–10, 2010, Indianapolis, IN.

Month	Number of MapReduce Jobs	Average completion time (s)	Average no. servers per job	Avg. no. cores per server	CPU core years	Input data (PB)	Intermediate data (PB)	Output data (PB)
Sep-16	95,775,891	331	130	2.4	311,691	11,553	4095	6982
Sep-15	115,375,750	231	120	2.7	272,322	8307	3980	5801
Sep-14	55,913,646	412	142	1.9	200,778	5989	2530	3951
Sep-13	28,328,775	469	137	1.4	81,992	2579	1193	1684
Sep-12	15,662,118	480	142	1.8	60,987	2171	818	874
Sep-11	7,961,481	499	147	2.2	40,993	1162	276	333
Sep-10	5,207,069	714	164	1.6	30,262	573	139	37
Sep-09	4,114,919	515	156	3.2	33,582	548	118	99
Sep-07	2,217,000	395	394	1.0	11,081	394	34	14
Mar-06	171,000	874	268	1.6	2002	51	7	3
Aug-04	29,000	634	157	1.9	217	3.2	0.7	0.2

Figure 6.2 Monthly MapReduce usage at Google from 2004 to 2016. Over 12 years the number of MapReduce jobs increased by a factor of 3300. Figure 6.17 on page 461 estimates that running the September 2016 workload on Amazon's cloud computing service EC2 would cost $114 million. Updated from Dean, J., 2009. Designs, lessons and advice from building large distributed systems [keynote address]. In: Proceedings of 3rd ACM SIGOPS International Workshop on Large-Scale Distributed Systems and Middleware, Co-located with the 22nd ACM Symposium on Operating Systems Principles, October 11–14, 2009, Big Sky, Mont.

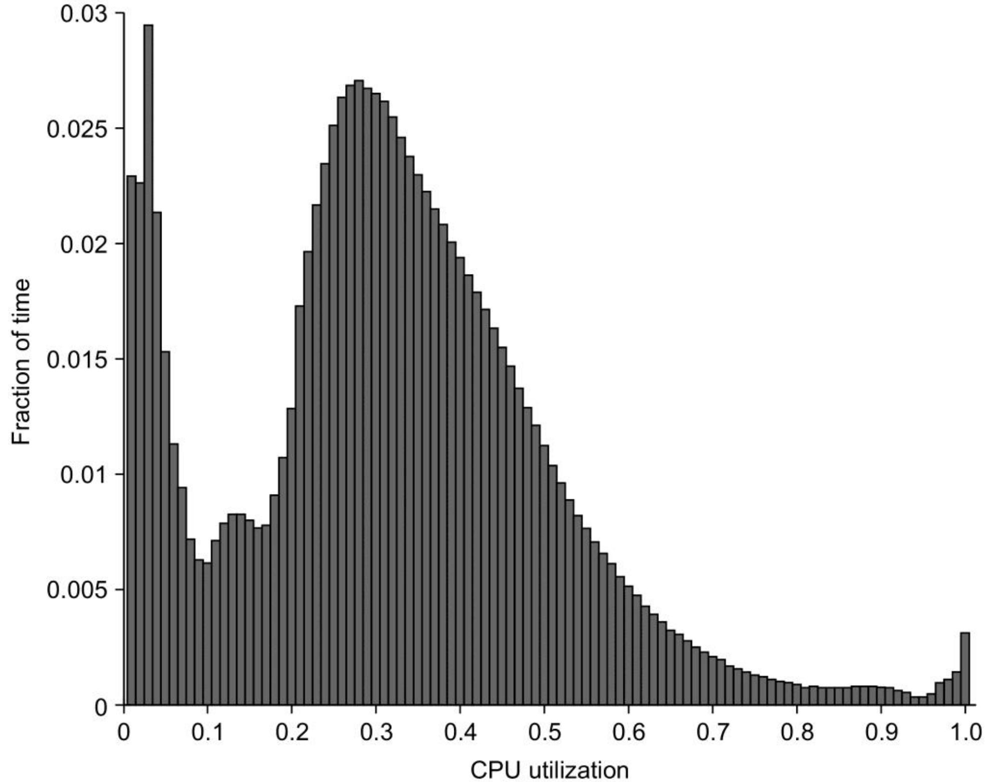

Figure 6.3 Average CPU utilization of more than 5000 servers during a 6-month period at Google. Servers are rarely completely idle or fully utilized, instead operating most of the time at between 10% and 50% of their maximum utilization. The third column from the right in Figure 6.4 calculates percentages plus or minus 5% to come up with the weightings; thus 1.2% for the 90% row means that 1.2% of servers were between 85% and 95% utilized. From Figure 1 in Barroso, L.A., Hölzle, U., 2007. The case for energy-proportional computing. IEEE Comput. 40 (12), 33–37.

Load	Performance	Watts	SPEC weightings	Weighted performance	Weighted watts	Figure 6.3 weightings	Weighted performance	Weighted watts
100%	2,889,020	662	9.09%	262,638	60	0.80%	22,206	5
90%	2,611,130	617	9.09%	237,375	56	1.20%	31,756	8
80%	2,319,900	576	9.09%	210,900	52	1.50%	35,889	9
70%	2,031,260	533	9.09%	184,660	48	2.10%	42,491	11
60%	1,740,980	490	9.09%	158,271	45	5.10%	88,082	25
50%	1,448,810	451	9.09%	131,710	41	11.50%	166,335	52
40%	1,159,760	416	9.09%	105,433	38	19.10%	221,165	79
30%	869,077	382	9.09%	79,007	35	24.60%	213,929	94
20%	581,126	351	9.09%	52,830	32	15.30%	88,769	54
10%	290,762	308	9.09%	26,433	28	8.00%	23,198	25
0%	0	181	9.09%	0	16	10.90%	0	20
Total	15,941,825	4967		1,449,257	452		933,820	380
				ssj_ops/W	3210		ssj_ops/W	2454

Figure 6.4 SPECpower result using the weightings from Figure 6.3 instead of even weightings.

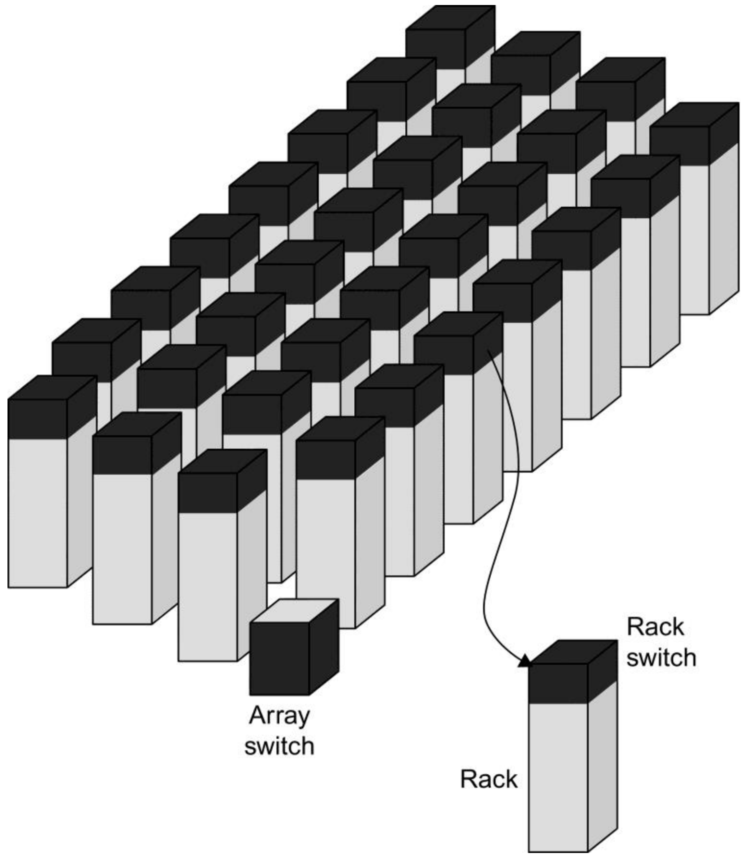

Figure 6.5 Hierarchy of switches in a WSC. Based on Figure 1.1 in Barroso, L.A., Clidaras, J., Hölzle, U., 2013. The datacenter as a computer: an introduction to the design of warehouse-scale machines. Synth. Lect. Comput. Architect. 8 (3), 1–154.

	Local	Rack	Array
DRAM latency (μs)	0.1	300	500
Flash latency (μs)	100	400	600
Disk latency (μs)	10,000	11,000	12,000
DRAM bandwidth (MB/s)	20,000	100	10
Flash bandwidth (MB/s)	1000	100	10
Disk bandwidth (MB/s)	200	100	10
DRAM capacity (GB)	16	1024	31,200
Flash capacity (GB)	128	20,000	600,000
Disk capacity (GB)	2000	160,000	4,800,000

Figure 6.6 Latency, bandwidth, and capacity of the memory hierarchy of a WSC (Barroso et al., 2013). Figure 6.7 plots this same information.

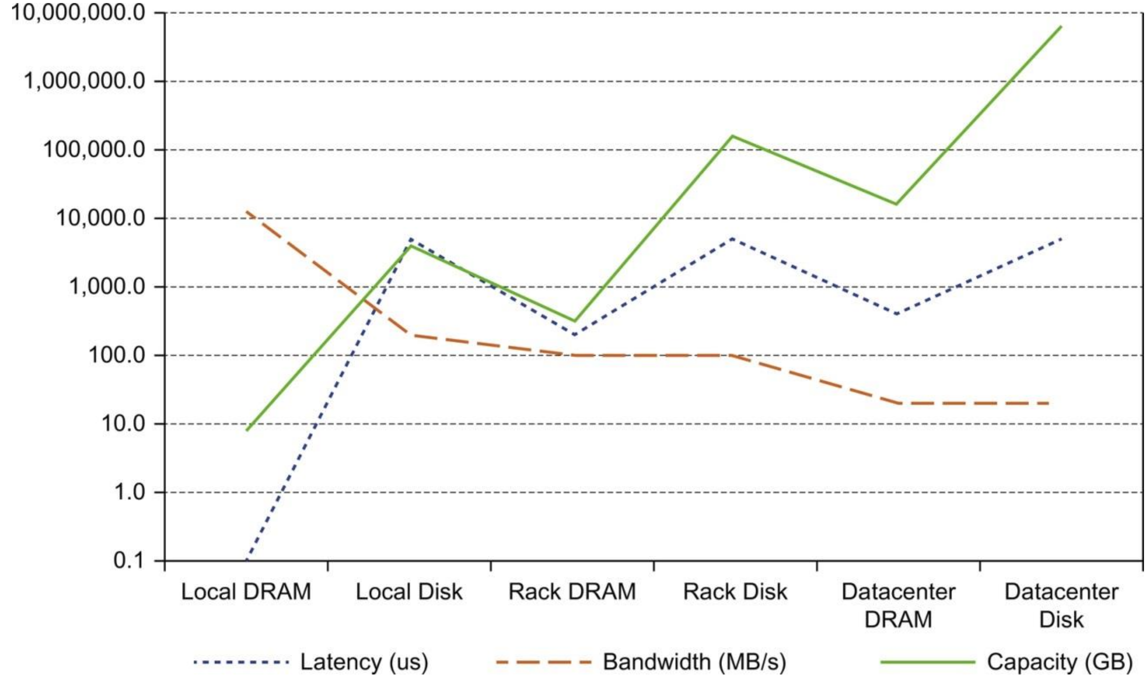

Figure 6.7 Graph of latency, bandwidth, and capacity of the memory hierarchy of a WSC for data in Figure 6.6 (Barroso et al., 2013).

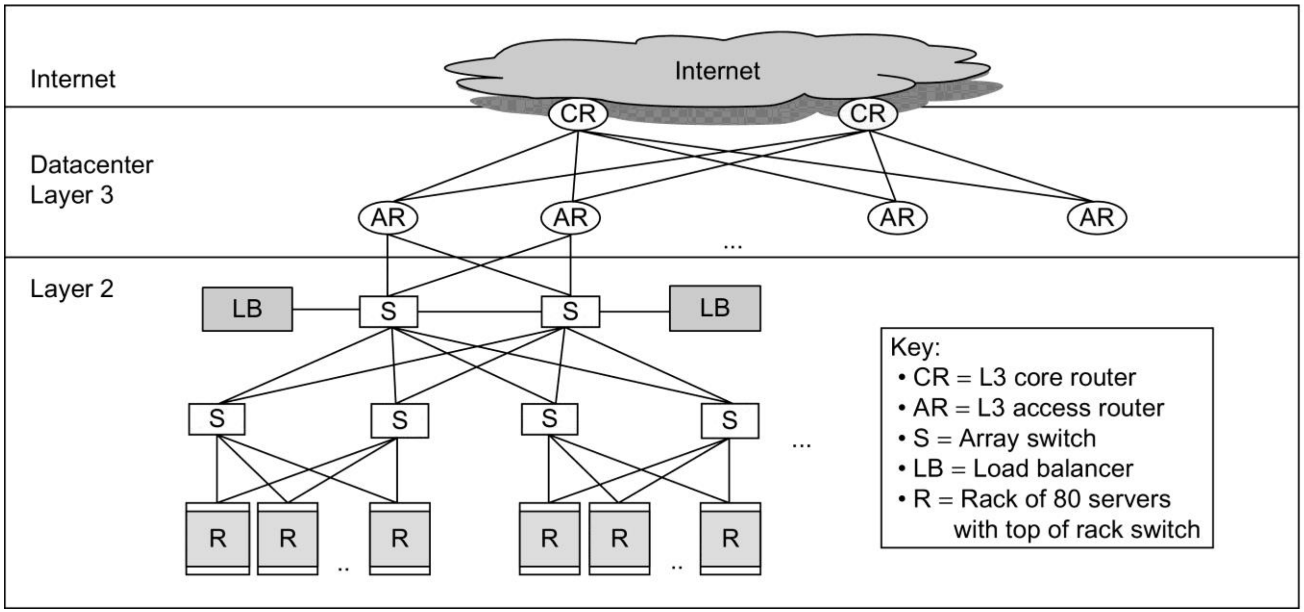

Figure 6.8 A Layer 3 network used to link arrays together and to the Internet (Greenberg et al., 2009). A load balancer monitors how busy a set of servers is and directs traffic to the less loaded ones to try to keep the servers approximately equally utilized. Another option is to use a separate *border router* to connect the Internet to the data center Layer 3 switches. As we will see in Section 6.6, many modern WSCs have abandoned the conventional layered networking stack of traditional switches.

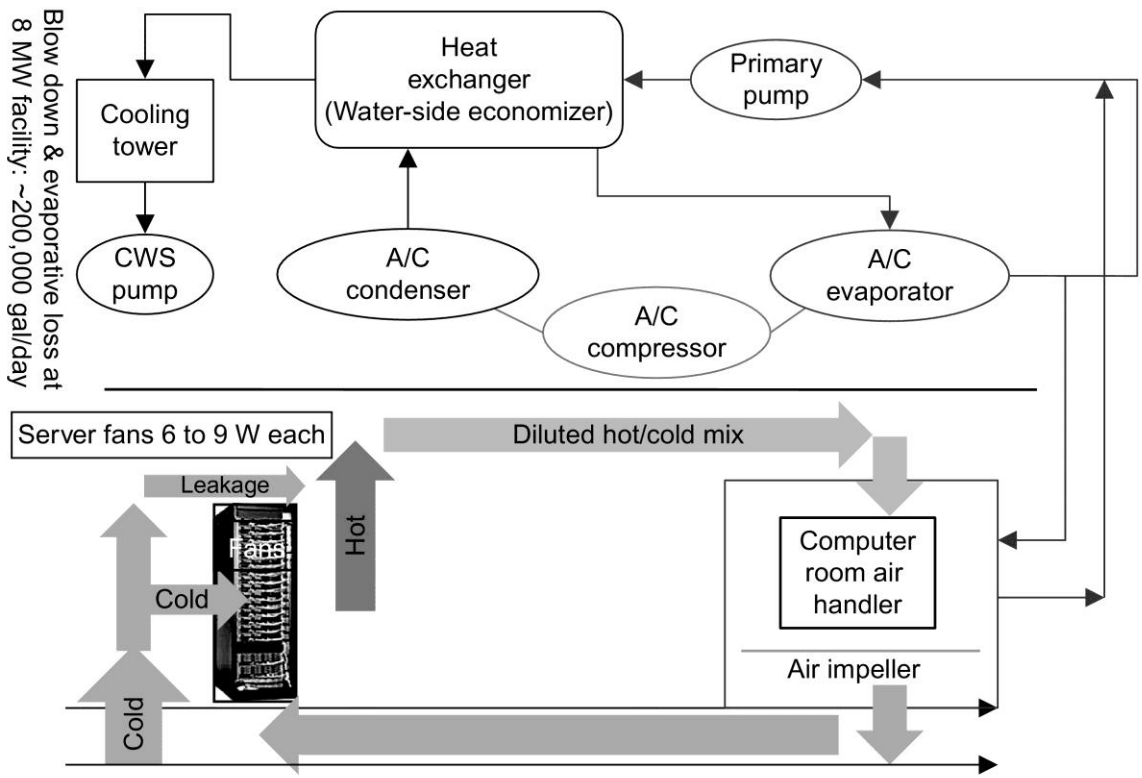

Figure 6.9 Mechanical design for cooling systems. CWS stands for circulating water system. From Hamilton, J., 2010. Cloud computing economies of scale. In: Paper Presented at the AWS Workshop on Genomics and Cloud Computing, June 8, 2010, Seattle, WA. http://mvdirona.com/jrh/TalksAndPapers/JamesHamilton_GenomicsCloud20100608.pdf.

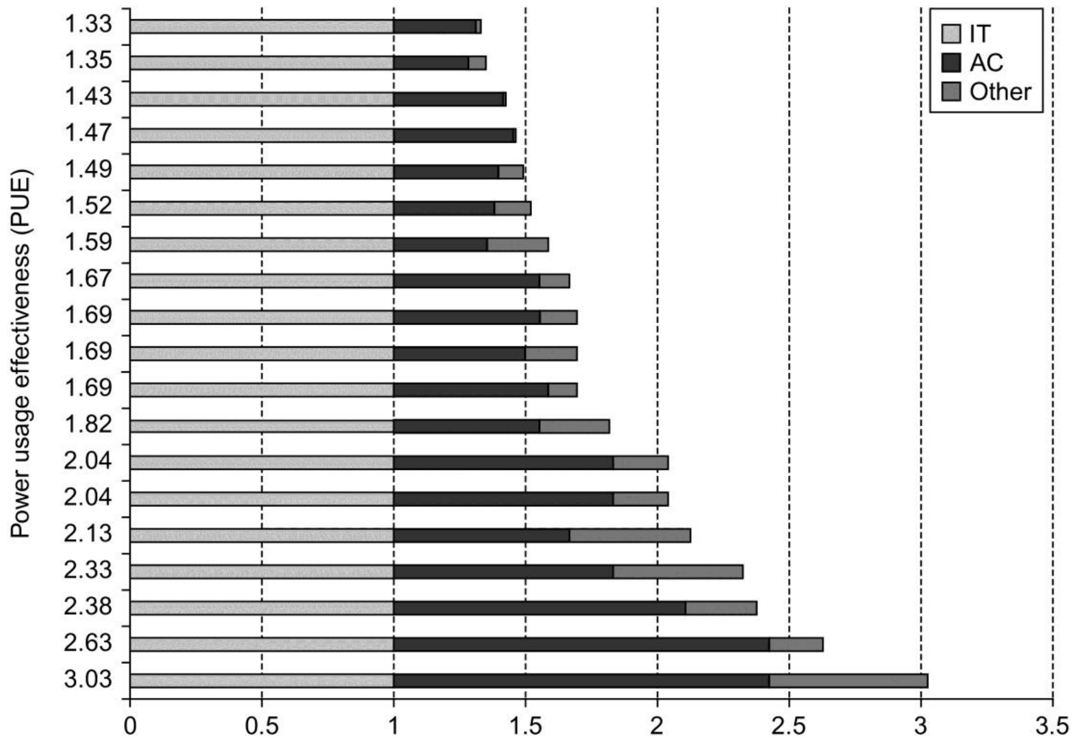

Figure 6.10 Power utilization efficiency of 19 data centers in 2006 (Greenberg et al., 2009). The power for air conditioning (AC) and other uses (such as power distribution) is normalized to the power for the IT equipment in calculating the PUE. Thus, power for IT equipment must be 1.0, and AC varies from about 0.30 to 1.40 times the power of the IT equipment. Power for “other” varies from about 0.05 to 0.60 of the IT equipment.

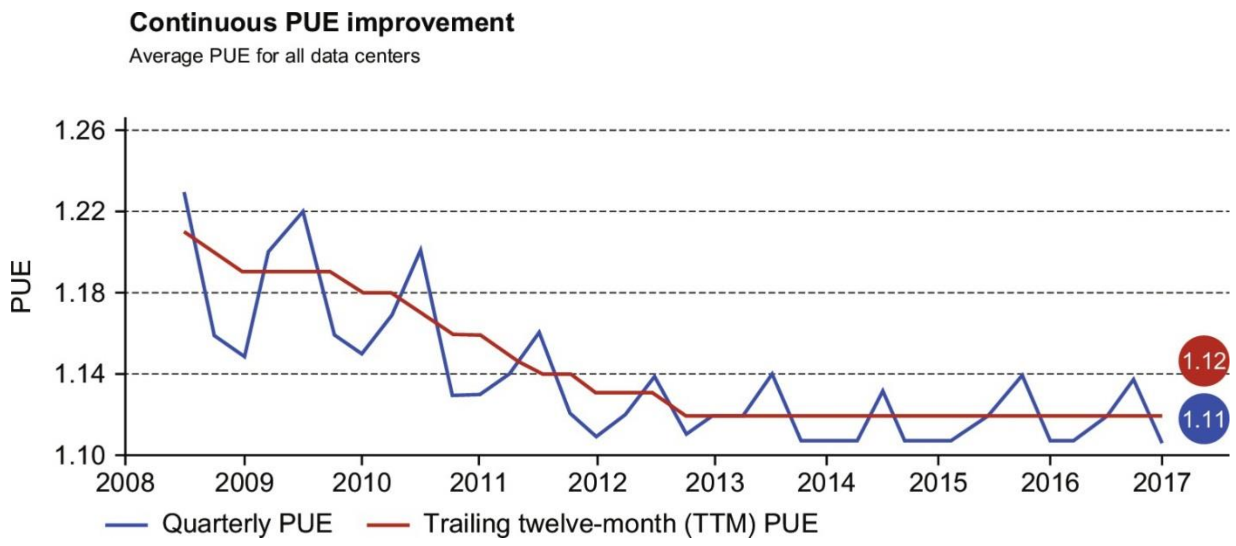

Figure 6.11 Average power utilization efficiency (PUE) of the 15 Google WSCs between 2008 and 2017. The spiking line is the quarterly average PUE, and the straighter line is the trailing 12-month average PUE. For Q4 2016, the averages were 1.11 and 1.12, respectively.

Server delay (ms)	Increased time to next click (ms)	Queries/ user	Any clicks/ user	User satisfaction	Revenue/ user
50	–	–	–	–	–
200	500	–	−0.3%	−0.4%	–
500	1200	–	−1.0%	−0.9%	−1.2%
1000	1900	−0.7%	−1.9%	−1.6%	−2.8%
2000	3100	−1.8%	−4.4%	−3.8%	−4.3%

Figure 6.12 Negative impact of delays at the Bing search server on user behavior (Schurman and Brutlag, 2009).

Size of facility (critical load watts)	8,000,000
Average power usage (%)	80%
Power usage effectiveness	1.45
Cost of power ($/kWh)	$0.07
% Power and cooling infrastructure (% of total facility cost)	82%
CAPEX for facility (not including IT equipment)	**$88,000,000**
Number of servers	45,978
Cost/server	$1450
CAPEX for servers	**$66,700,000**
Number of rack switches	1150
Cost/rack switch	$4800
Number of array switches	22
Cost/array switch	$300,000
Number of layer 3 switches	2
Cost/layer 3 switch	$500,000
Number of border routers	2
Cost/border router	$144,800
CAPEX for networking gear	**$12,810,000**
Total CAPEX for WSC	**$167,510,000**
Server amortization time	3 years
Networking amortization time	4 years
Facilities amortization time	10 years
Annual cost of money	5%

Figure 6.13 Case study for a WSC, rounded to nearest $5000. Internet bandwidth costs vary by application, so they are not included here. The remaining 18% of the CAPEX for the facility includes buying the property and the cost of construction of the building. We added people costs for security and facilities management in Figure 6.14, which were not part of the case study. Note that Hamilton's estimates were done before he joined Amazon, and they are not based on the WSC of a particular company. Based on Hamilton, J., 2010. Cloud computing economies of scale. In: Paper Presented at the AWS Workshop on Genomics and Cloud Computing, June 8, 2010, Seattle, WA. http://mvdirona.com/jrh/TalksAndPapers/JamesHamilton_GenomicsCloud20100608.pdf.

Expense (% total)	Category	Monthly cost	Percent monthly cost
Amortized CAPEX (85%)	Servers	$2,000,000	53%
	Networking equipment	$290,000	8%
	Power and cooling infrastructure	$765,000	20%
	Other infrastructure	$170,000	4%
OPEX (15%)	Monthly power use	$475,000	13%
	Monthly people salaries and benefits	$85,000	2%
	Total OPEX	$3,800,000	100%

Figure 6.14 Monthly OPEX for Figure 6.13, rounded to the nearest $5000. Note that the 3-year amortization of servers means purchasing new servers every 3 years, whereas the facility is amortized for 10 years. Thus, the amortized capital costs for servers are about three times more than for the facility. People costs include three security guard positions continuously for 24 h a day, 365 days a year, at $20 per hour per person, and one facilities person for 24 h a day, 365 days a year, at $30 per hour. Benefits are 30% of salaries. This calculation does not include the cost of network bandwidth to the Internet because it varies by application nor vendor maintenance fees because they vary by equipment and by negotiations.

	Instance	Per hour	Ratio to m4.large	Virtual cores	Compute units	Memory (GiB)	Storage (GB)
General-purpose	t2.nano	$0.006	0.05	1	Variable	0.5	EBS only
	t2.micro	$0.012	0.11	1	Variable	1.0	EBS only
	t2.small	$0.023	0.21	1	Variable	2.0	EBS only
	t2.medium	$0.047	0.4	2	Variable	4.0	EBS only
	t2.large	$0.094	0.9	2	Variable	8.0	EBS only
	t2.xlarge	$0.188	1.7	4	Variable	16.0	EBS only
	t2.2xlarge	$0.376	3.5	8	Variable	32.0	EBS only
	m4.large	$0.108	1.0	2	6.5	8.0	EBS only
	m4.xlarge	$0.215	2.0	4	13	16.0	EBS only
	m4.2xlarge	$0.431	4.0	8	26	32.0	EBS only
	m4.4xlarge	$0.862	8.0	16	54	64.0	EBS only
	m4.10xlarge	$2.155	20.0	40	125	160.0	EBS only
	m4.16xlarge	$3.447	31.9	64	188	256.0	EBS only
	m3.medium	$0.067	0.6	1	3	3.8	1 × 4 SSD
	m3.large	$0.133	1.2	2	6.5	7.5	1 × 32 SSD
	m3.xlarge	$0.266	2.5	4	13	15.0	2 × 40 SSD
	m3.2xlarge	$0.532	4.9	8	26	30.0	2 × 80 SSD
Compute-optimized	c4.large	$0.100	0.9	2	8	3.8	EBS only
	c4.xlarge	$0.199	1.8	4	16	7.5	EBS only
	c4.2xlarge	$0.398	3.7	8	31	15.0	EBS only
	c4.4xlarge	$0.796	7.4	16	62	30.0	EBS only
	c4.8xlarge	$1.591	14.7	36	132	60.0	EBS only
	c3.large	$0.105	1.0	2	7	3.8	2 × 16 SSD
	c3.xlarge	$0.210	1.9	4	14	7.5	2 × 40 SSD
	c3.2xlarge	$0.420	3.9	8	28	15.0	2 × 80 SSD
	c3.4xlarge	$0.840	7.8	16	55	30.0	2 × 160 SSD
	c3.8xlarge	$1.680	15.6	32	108	60.0	2 × 320 SSD

Figure 6.15 Price and characteristics of on-demand general-purpose and compute-optimized EC2 instances in the Virginia region of the United States in February 2017. When AWS started, one EC2 computer unit was equivalent to a 1.0–1.2 GHz AMD Opteron or Intel Xeon of 2006. Variable instances are the newest and cheapest category. They offer the full performance of a high-frequency Intel CPU core if your workload utilizes less than 5% of the core on average over 24 h, such as for serving web pages. AWS also offers Spot Instances at a much lower cost (about 25%). With Spot Instances, customers set the price they are willing to pay and the number of instances they are willing to run, and then AWS runs the bids when the spot price drops below their level. AWS also offers Reserved Instances for cases where customers know they will use most of the instance for a year. They pay a yearly fee per instance and then an hourly rate that is about 30% of column 1 to use the service. If a Reserved Instance is used 100% for a whole year, the average cost per hour including amortization of the annual fee will be about 65% of the rate in the first column. EBS is Elastic Block Storage, which is a raw block-level storage system found elsewhere on the network, rather than in a local disk or local solid stage disk (SSD) within the same server as the VM.

	Instance	Per hour	Ratio to m4.large	Virtual cores	Compute units	Memory (GiB)	Storage (GB)
GPU	p2.xlarge	$0.900	8.3	4	12	61.0	EBS only
	p2.8xlarge	$7.200	66.7	32	94	488.0	EBS only
	p2.16xlarge	$14.400	133.3	64	188	732.0	EBS only
	g2.2xlarge	$0.650	6.0	8	26	15.0	60 SSD
	g2.8xlarge	$2.600	24.1	32	104	60.0	2 × 120 SSD
FPGA	f1.2xlarge	$1.650	15.3	8 (1 FPGA)	26	122.0	1 × 470 SSD
	f1.16xlarge	$13.200	122.2	64 (8 FPGA)	188	976.0	4 × 940 SSD
Memory-optimized	x1.16xlarge	$6.669	61.8	64	175	976.0	1 × 1920 SSD
	x1.32xlarge	$13.338	123.5	128	349	1,952.0	2 × 1920 SSD
	r3.large	$0.166	1.5	2	6.5	15.0	1 × 32 SSD
	r3.xlarge	$0.333	3.1	4	13	30.5	1 × 80 SSD
	r3.2xlarge	$0.665	6.2	8	26	61.0	1 × 160 SSD
	r3.4xlarge	$1.330	12.3	16	52	122.0	1 × 320 SSD
	r3.8xlarge	$2.660	24.6	32	104	244.0	2 × 320 SSD
	r4.large	$0.133	1.2	2	7	15.3	EBS only
	r4.xlarge	$0.266	2.5	4	14	30.5	EBS only
	r4.2xlarge	$0.532	4.9	8	27	61.0	EBS only
	r4.4xlarge	$1.064	9.9	16	53	122.0	EBS only
	r4.8xlarge	$2.128	19.7	32	99	244.0	EBS only
	r4.16xlarge	$4.256	39.4	64	195	488.0	EBS only
Storage-optimized	i2.xlarge	$0.853	7.9	4	14	30.5	1 × 800 SSD
	i2.2xlarge	$1.705	15.8	8	27	61.0	2 × 800 SSD
	i2.4xlarge	$3.410	31.6	16	53	122.0	4 × 800 SSD
	i2.8xlarge	$6.820	63.1	32	104	244.0	8 × 800 SSD
	d2.xlarge	$0.690	6.4	4	14	30.5	3 × 2000 HDD
	d2.2xlarge	$1.380	12.8	8	28	61.0	6 × 2000 HDD
	d2.4xlarge	$2.760	25.6	16	56	122.0	12 × 2000 HDD
	d2.8xlarge	$5.520	51.1	36	116	244.0	24 × 2000 HDD

Figure 6.16 Price and characteristics of on-demand GPUs, FPGAs, memory-optimized, and storage-optimized EC2 instances in the Virginia region of the United States in February 2017.

	Aug-04	Sep-09	Sep-12	Sep-16
Average completion time (h)	0.15	0.14	0.13	0.11
Average number of servers per job	157	156	142	130
Cost per hour of EC2 c4.large instance	$0.100	$0.100	$0.100	$0.100
Average EC2 cost per MapReduce job	$2.76	$2.23	$1.89	$1.20
Monthly number of MapReduce jobs	29,000	4,114,919	15,662,118	95,775,891
Total cost of MapReduce jobs on EC2/EBS	$80,183	$9,183,128	$29,653,610	$114,478,794

Figure 6.17 Estimated cost to run the Google MapReduce workload for select months between 2004 and 2016 (Figure 6.2) using 2017 prices for AWS EC2. Because we are using 2017 prices, these are underestimates of actual AWS costs.

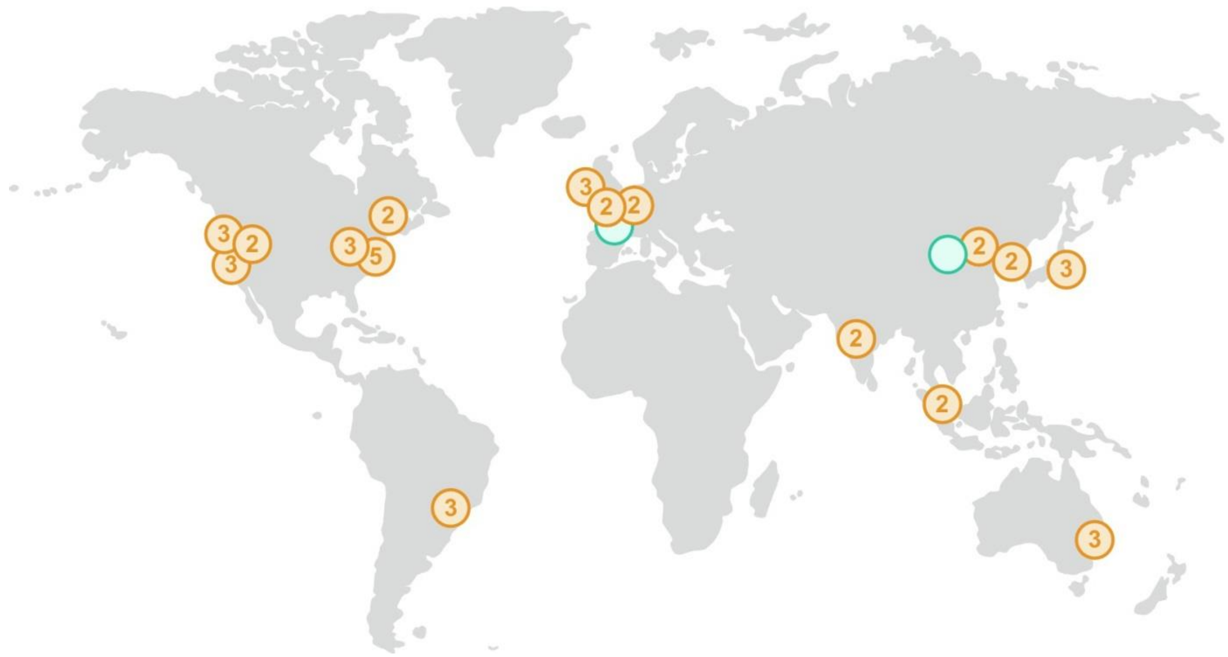

Figure 6.18 In 2017 AWS had 16 sites ("regions"), with two more opening soon. Most sites have two to three *availability zones*, which are located nearby but are unlikely to be affected by the same natural disaster or power outage, if one were to occur. (The number of availability zones are listed inside each circle on the map.) These 16 sites or regions collectively have 42 availability zones. Each availability zone has one or more WSCs. https://aws.amazon.com/about-aws/global-infrastructure/.

Figure 6.19 In 2017 Google had 15 sites. In the Americas: Berkeley County, South Carolina; Council Bluffs, Iowa; Douglas County, Georgia; Jackson County, Alabama; Lenoir, North Carolina; Mayes County, Oklahoma; Montgomery County, Tennessee; Quilicura, Chile; and The Dalles, Oregon. In Asia: Changhua County, Taiwan; Singapore. In Europe: Dublin, Ireland; Eemshaven, Netherlands; Hamina, Finland; St. Ghislain, Belgium. https://www.google.com/about/datacenters/inside/locations/.

Figure 6.20 In 2017 Microsoft had 34 sites, with four more opening soon. https://azure.microsoft.com/en-us/regions/.

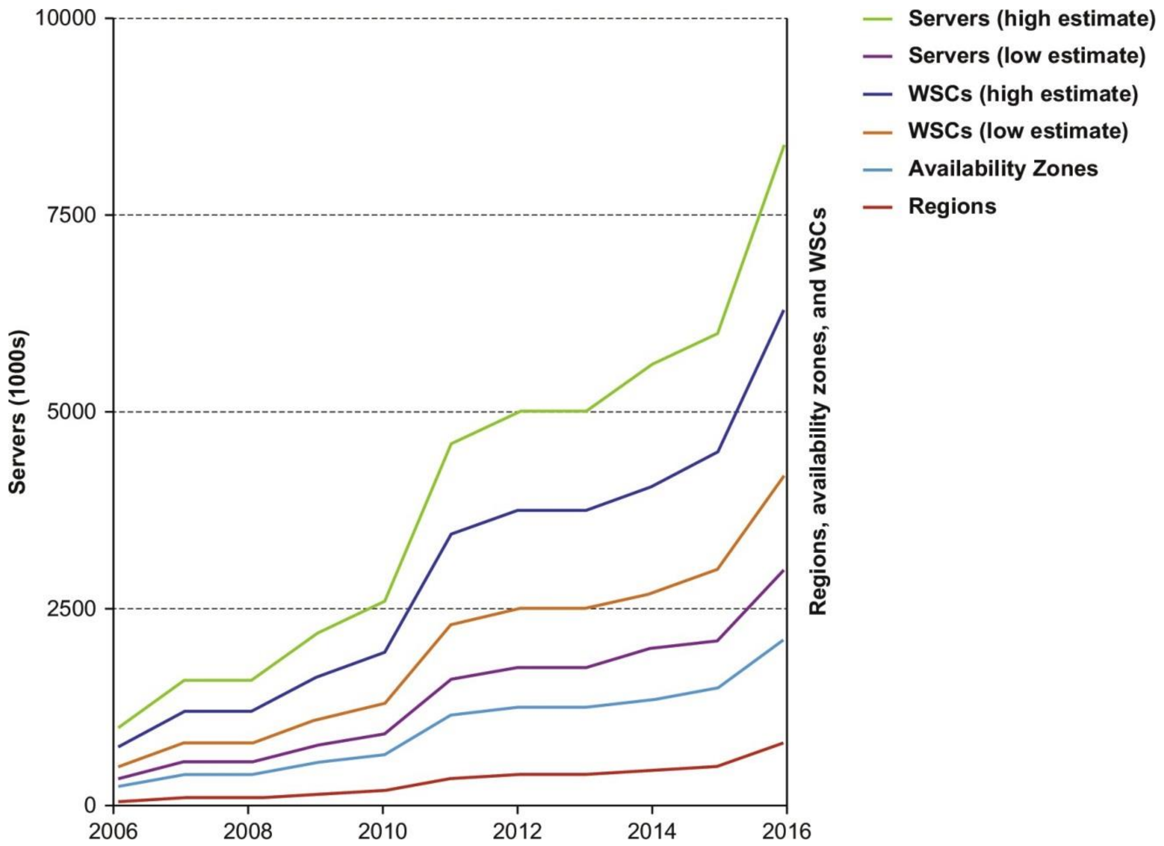

Figure 6.21 Growth of AWS regions and availability zones (right vertical axis) over time. Most regions have two or three availability zones. Each availability zone can have one or more WSCs, with the largest having more than 10 WSCs. Each WSC has at least 50,000 servers, with the biggest having more than 80,000 servers (Hamilton, 2014). Based on two published estimates for the number of AWS servers in 2014 (Clark, 2014; Morgan 2014), we project the number of servers per year (left vertical axis) and WSCs (right vertical access) as a function of the actual number of availability zones.

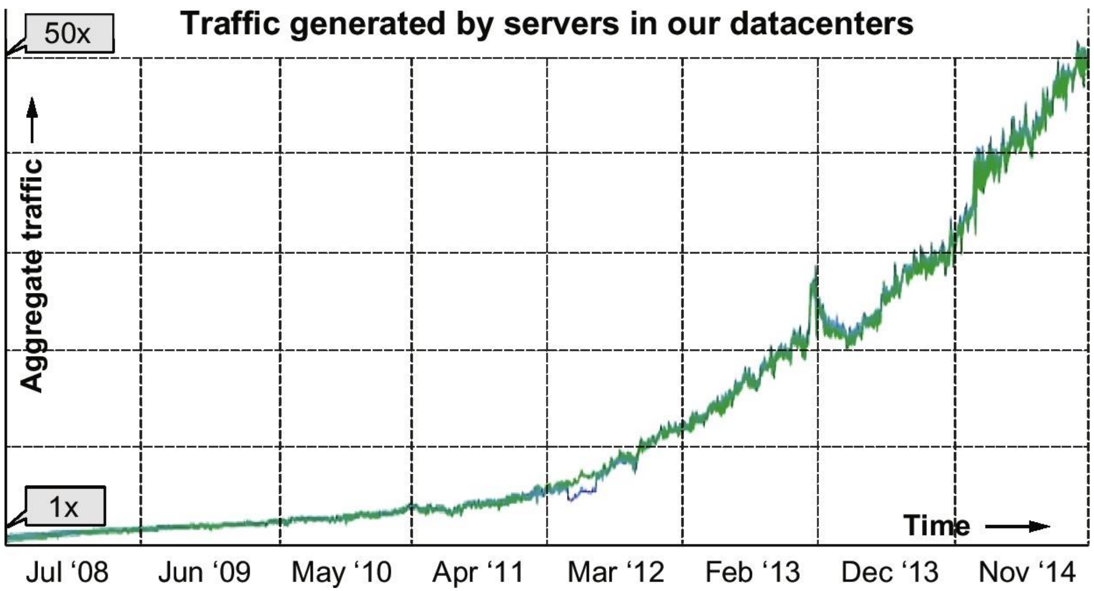

Figure 6.22 Network traffic from all the servers in Google's WSCs over 7 years (Singh et al., 2015).

Data center generation switch	First deployed	Merchant silicon	Top of rack (ToR) switch config	Edge aggregation block	Spine block	Fabric speed	Host speed	Bisection BW
Four-Post CRs	2004	Vendor	48 × 1 Gbps	–	–	10 Gbps	1 Gbps	2 Tbps
Firehose 1.0	2005	8 × 10 Gbps 4 × 10 Gbps (ToR)	2 × 10 Gbps up 24 × 1 Gbps down	2 × 32 × 10 Gbps	32 × 10 Gbps	10 Gbps	1 Gbps	10 Tbps
Firehose 1.1	2006	8 × 10 Gbps	4 × 10 Gbps up 48 × 1 Gbps down	64 × 10 Gbps	32 × 10 Gbps	10 Gbps	1 Gbps	10 Tbps
Watchtower	2008	16 × 10 Gbps	4 × 10 Gbps up 48 × 1 Gbps down	4 × 128 × 10 Gbps	128 × 10 Gbps	10 Gbps	n × 1 Gbps	82 Tbps
Saturn	2009	24 × 10 Gbps	24 × 10 Gbps	4 × 288 × 10 Gbps	288 × 10 Gbps	10 Gbps	n × 10 Gbps	207 Tbps
Jupiter	2012	16 × 40 Gbps	16 × 40 Gbps	8 × 128 × 40 Gbps	128 × 40 Gbps	10/40 Gbps	n × 10 Gbps/ n × 40 Gbps	1300 Tbps

Figure 6.23 Six generations of network switches deployed at Google WSCs (Singh et al., 2015). The Four-Post CRs used commercial 512 port, 1 Gbit/s Ethernet switches, and 48-port, 1 Gbit/s Ethernet Top of Rack (ToR) switches, which allowed 20,000 servers in the array. The goal of Firehose 1.0 was to deliver 1 Gbps of nonblocking bisection bandwidth to each of 10,000 servers, but it ran into problems with the low connectivity of the ToR switch that caused problems when links failed. Firehose 1.1 was the first custom-designed switch with better connectivity in the ToR switch. Watchtower and Saturn followed in the same footsteps, but used new, faster merchant switch chips. Jupiter uses 40 Gbps links and switches to deliver more than 1 Pbit/s of bisection bandwidth. Section 6.7 describes the Jupiter switch and the Edge Aggregation and Spine Blocks of Clos networks in more detail.

Figure 6.24 An on-site substation.

Figure 6.25 This image shows transformers, switch gear, and generators in close proximity to a WSC.

Figure 6.26 Row of servers with the copper bus ducts above that distribute 400 V to the servers. Although hard to see, they are above the shelf on the right side of the photo. It also shows a cold aisle that operators use to service the equipment.

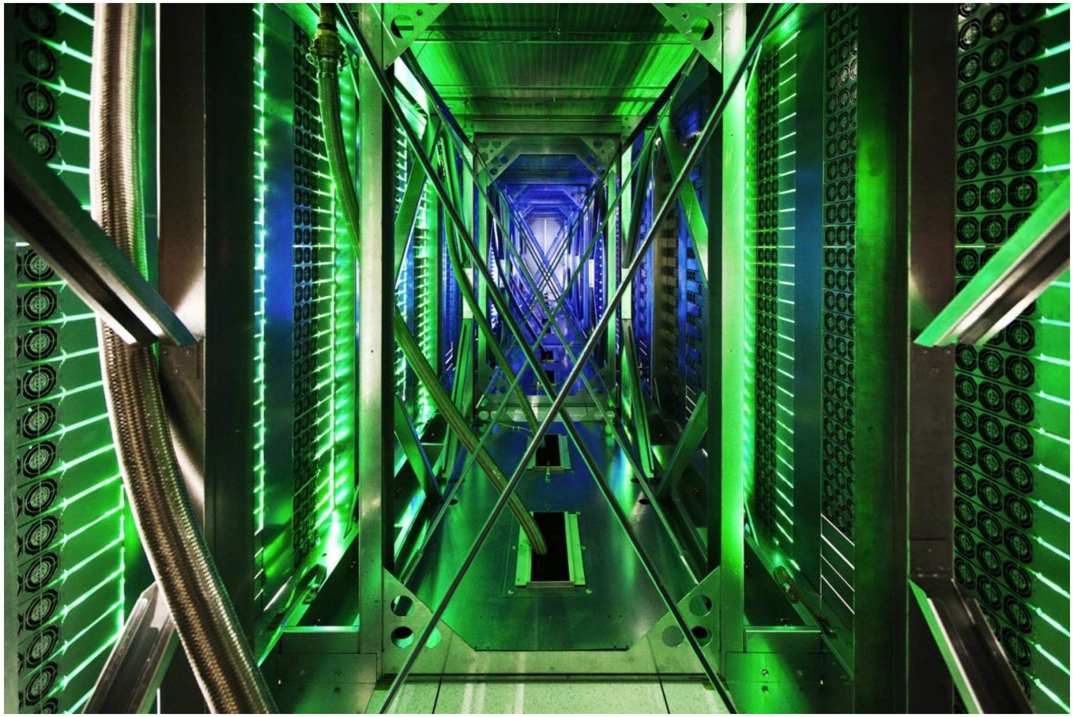

Figure 6.27 Hot aisle in a Google data center, which is clearly not designed to accommodate people.

Figure 6.28 The cool air blows into the room containing the aisles of servers. The hot air goes through large vents into the ceilings where it is cooled before returning to these fans.

Figure 6.29 Steam rising from the cooling towers that transfer heat to the air from the water used to cool equipment.

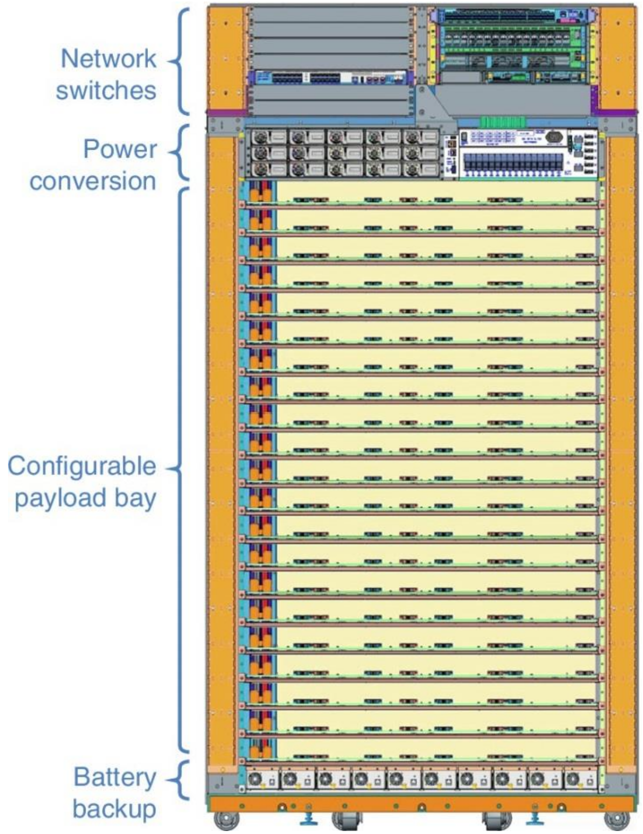

Figure 6.30 A Google rack for its WSC. Its dimensions are about 7 ft high, 4 ft wide, and 2 ft deep (2 m × 1.2 m × 0.5 m). The Top of Rack switches are indeed at the top of this rack. Next comes the power converter that converts from 240 V AC to 48 V DC for the servers in the rack using a bus bar at the back of the rack. Next is the 20 slots (depending on the height of the server) that can be configured for the various types of servers that can be placed in the rack. Up to four servers can be placed per tray. At the bottom of the rack are high-efficiency distributed modular DC uninterruptible power supply (UPS) batteries.

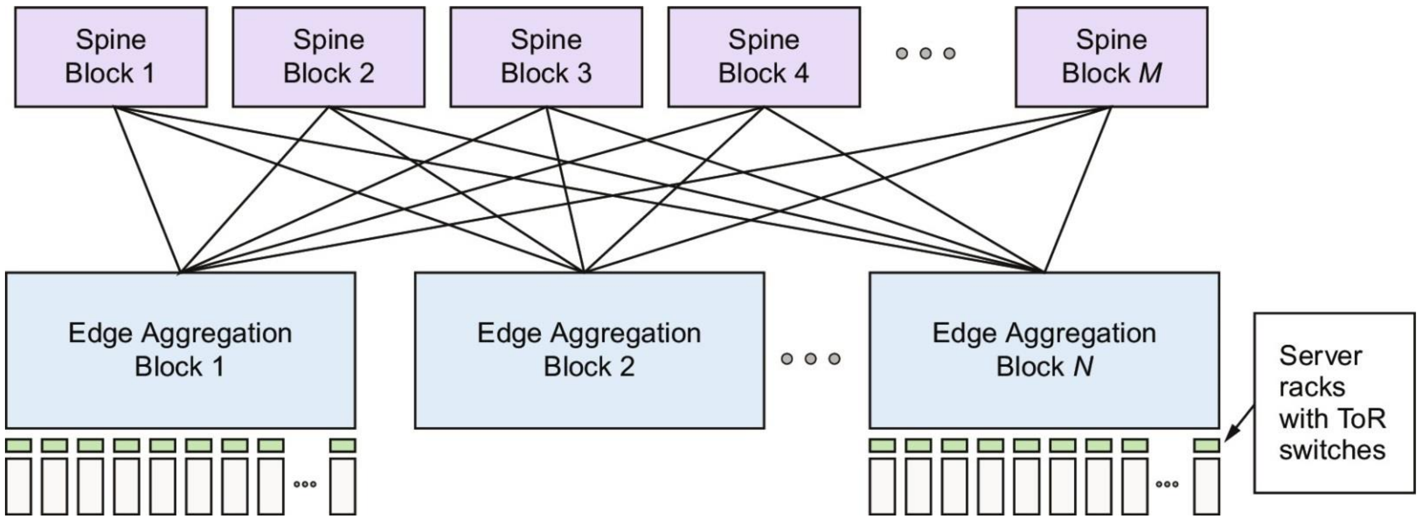

Figure 6.31 A Clos network has three logical stages containing crossbar switches: ingress, middle, and egress. Each input to the ingress stage can go through any of the middle stages to be routed to any output of the egress stage. In this figure, the middle stages are the *M* Spine Blocks, and the ingress and egress stages are in the *N* Edge Activation Blocks. Figure 6.22 shows the changes in the Spine Blocks and the Edge Aggregation Blocks over many generations of Clos networks in Google WSCs.

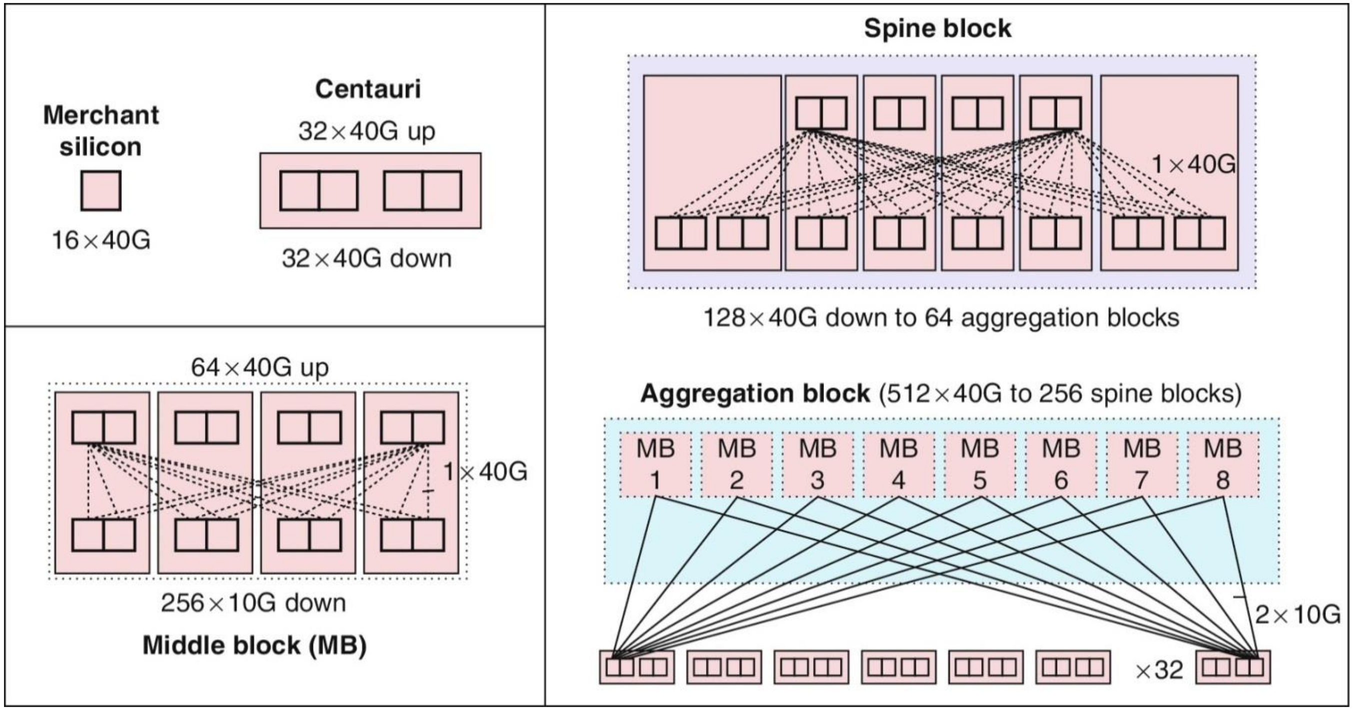

Figure 6.32 Building blocks of the Jupiter Clos network.

Figure 6.33 Middle blocks of the Jupiter switches housed in racks. Four are packed in a rack. A rack can hold two spine blocks.

Figure 6.34 An example server from a Google WSC. The Haswell CPUs (2 sockets × 18 cores × 2 threads = 72 “virtual cores” per machine) have 2.5 MiB last level cache per core or 45 MiB using DDR3-1600. They use the Wellsburg Platform Controller Hub and have a TFP of 150 W.

Mode	Performance	Power
High	100%	100%
Medium	75%	60%
Low	59%	38%

Figure 6.35 Power–performance modes for low-power servers.

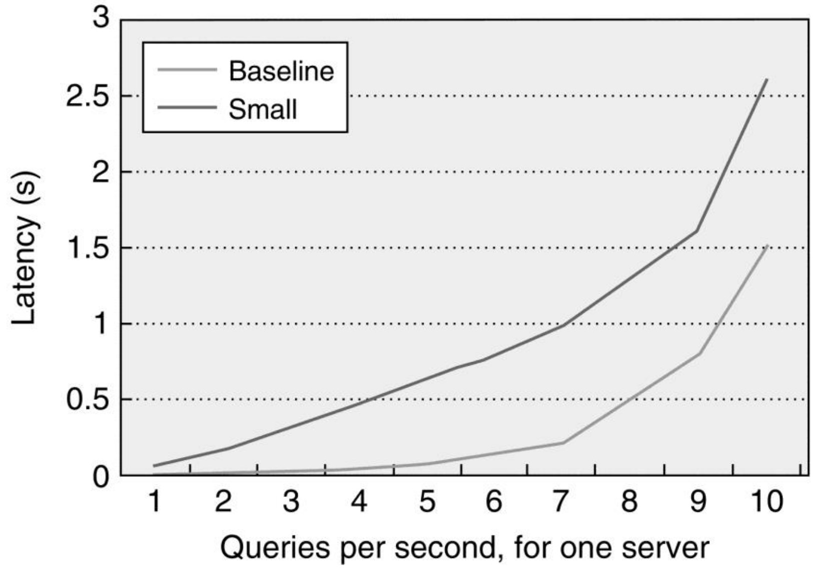

Figure 6.36 Query–response time curve.

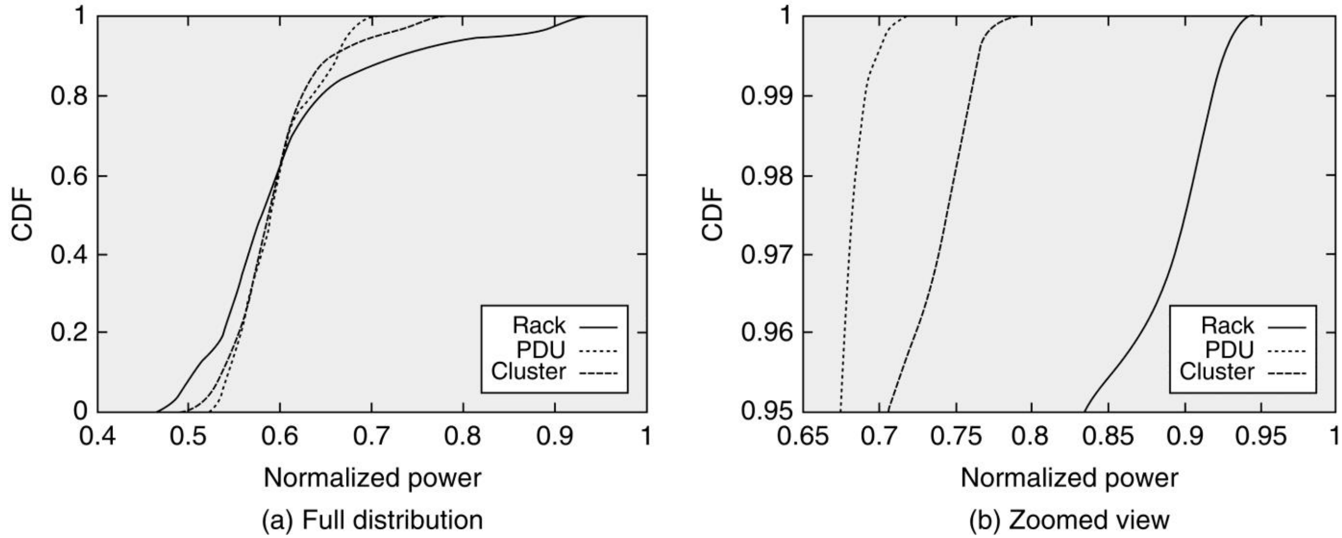

Figure 6.37 Cumulative distribution function (CDF) of a real datacenter.

Activity (%)	0	10	20	30	40	50	60	70	80	90	100
Power, case A (W)	181	308	351	382	416	451	490	533	576	617	662
Power, case B (W)	250	275	325	340	395	405	415	425	440	445	450

Figure 6.38 Power distribution for two servers.

Activity (%)	0	10	20	30	40	50	60	70	80	90	100
No. servers, case A and B	109	80	153	246	191	115	51	21	15	12	8
No. servers, case C	504	6	8	11	26	57	95	123	76	40	54

Figure 6.39 Utilization distributions across cluster, without and with consolidation.

Tier 1	Single path for power and cooling distributions, without redundant components	99.0%
Tier 2	$(N + 1)$ redundancy = two power and cooling distribution paths	99.7%
Tier 3	$(N + 2)$ redundancy = three power and cooling distribution paths for uptime even during maintenance	99.98%
Tier 4	Two active power and cooling distribution paths, with redundant components in each path, to tolerate any single equipment failure without impacting the load	99.995%

Figure 6.40 Overview of data center tier classifications. (Adapted from Pitt Turner IV et al. [2008].).