

Capacity Planning Project Sedona

Nate Amsden

November 2010

General Disclaimer

Capacity planning at this stage of the game involves an enormous amount of factors that impact requirements, some of these are:

- Application code efficiency / feature changes (code changes can increase or decrease infrastructure load even when there is no change in traffic)
- Scalability of application (is it effectively able to max out at least one CPU core, reducing the number of instances and thus memory requirements)
- Amount of data stored per user on the environment (whether it is in memory caches or SQL databases), and frequency of access
- Scalability of SQL queries
- Degree of efficiency gains from:
 - “Right sizing” virtual machines to precisely match application requirements
 - Observing the load balancer's ability to improve web server capacity transparently using HTTP Multiplexing
 - Impact of low latency network and storage on application throughput
 - Observing effectiveness of VMware's advanced functionality in increasing capacity utilization including memory de duplication and compression

Add to that the overall dynamic nature of the environment and it makes it very difficult to come up with accurate estimates.

Amazon Capacity (August 2010)

EC2 CPU Units Provisioned	CPU Usage (average)	EC2 CPU Units Usage	Memory Gigabytes Provisioned	Memory Gigabytes Used
2800	7.00%	196	3700	2500

Storage Type	TB of provisioned storage	TB of used storage	Capacity Utilization
EBS Storage	77	4.4	5.70%
Instance Storage	287	6.7	2.00%

Storage I/O requirements are much more difficult to pin down without precise instrumentation. Leveraging my own years of experience running production environments virtualized and otherwise, and knowing how lightly loaded our databases are I estimate our I/O requirements to be roughly 3,500 logical I/Os per second, this is roughly the equivalent of 60 SATA disks in RAID 5, or 18 high speed Fibre channel disks in RAID 5.

Amazon does not give us bandwidth numbers that can be compared to the way a regular Internet provider bills for bandwidth. Amazon bills us per amount of data, and normal providers bill at the rate you transmit or receive data (this can be advantageous as it allows you to burst without paying extra). Amazon charged us for about 40TB of data transfer for August. Rough math puts that at around 120 - 150Mbps.

Initial Capacity Deployment

- Server Capacity
 - 384 x 2.2 Ghz CPU cores **[20% Used initially]**
 - 4096 GB of memory **[61% Used initially]**
 - 56 Gigabit theoretical network throughput/server
 - 2 x 10Gbps Networking for virtual server traffic
 - 2 x 10Gbps Networking for VMware host traffic
 - 2 x 8Gbps Networking for storage traffic
- Storage
 - 27TB of usable storage **[44% Used initially]**
 - 8,000 logical I/Os per second **[43% Used Initially]**
- Network
 - 150Mbps of Internet bandwidth (burst to 1Gbps)
 - Load balancer performance
 - 1 Gbps **[15% Used initially]**
 - Up to 200,000 HTTP requests per second
 - Up to 20,000 SSL transactions per second

Capacity Units

Two options will be presented for each capacity increment:

- **Best case**

- If everything goes as expected this **should** work

- Assumes that application scales to at least 50% CPU capacity from baseline.

- If not, then we become even more memory bound and need more servers.

- **Optimal**

- This is the safer choice without going all out (about 15% additional capacity)

Cumulative Expansion Capacity

Best Case Estimates

Hardware+ Software	1YR Support Costs	1YR Hosting Costs	Capacity Metric	CPU Ghz	Memory Gigabytes	Usable Disk TB	Logical Disk I/O	Physical Disk I/O
\$864,000	\$56,000	\$65,500	Initial Capacity	845	4,096	27	8,000	15,500
\$23,000	\$0	\$21,600	200% traffic	0	0	0	0	0
\$72,000	\$0	\$35,424	300% traffic	0	0	14	4,000	7,700
\$176,000	\$8,700	\$21,600	400% traffic	105	512	14	4,000	7,700
\$180,000	\$16,200	\$21,600	500% traffic	210	1,024	14	4,000	7,700
\$1,315,000	\$80,900	\$165,724	Total – 500%	1,160	5,632	69	20,000	38,600

Optimal Estimates

Hardware+ Software	1YR Support Costs	1YR Hosting Costs	Capacity Metric	CPU Ghz	Memory Gigabytes	Usable Disk TB	Logical Disk I/O	Physical Disk I/O
\$864,000	\$56,000	\$65,500	Initial Capacity	845	4,096	27	8,000	15,500
\$131,000	\$8,100	\$35,424	200% traffic	105	512	0	0	0
\$168,000	\$8,100	\$21,600	300% traffic	105	512	21	6,000	11,600
\$126,000	\$8,100	\$21,600	400% traffic	105	512	14	4,000	7,700
\$222,000	\$16,200	\$21,600	500% traffic	210	1,024	21	6,000	11,600
\$1,511,000	\$96,500	\$165,724	Total – 500%	1,370	6,656	83	24,000	46,400

Peak Capacity above involves two racks, thirteen servers and one storage system

Appendix:

**Technical information behind
capacity utilization and
growth.**

Quick comparison

Based on August 2010 costs, if we were to sustain that level of services and not grow, our bill for three years with 3YR EC2 reserved would be \$3,000,000 (\$5,200,000 with On Demand pricing)

By comparison, three years with at **500% of the traffic** that we have today would run us roughly \$2,100,000, and consume two racks of data center space.

200% traffic growth capacity estimates

Best Case:

Server Utilization:

- No new servers needed
- CPU utilization goes to 40%
- Memory utilization increases to 70%

Storage Capacity:

- Add remaining storage licensing
- Disk space utilization increases to 17 TB (63% of max)
- Disk I/O increases to 7,000 I/Os per second (89% of max)

Network Capacity:

- No new network equipment needed
- Add 150Mbps of Internet Bandwidth

300% traffic growth capacity estimates

Best Case:

Server Utilization:

- No new servers needed
- CPU utilization goes to 60%
- Memory utilization increases to 75%

Storage Utilization:

- Add 4,000 IOPS / 14 TB storage
- Disk space utilization increases to 22 TB (44% of max)
- Disk I/O utilization to 10,500 I/Os per second (88% of max)
- Power for second rack required

Network Utilization:

- No new network equipment needed
- Add 150Mbps of Internet Bandwidth

400% traffic growth capacity estimates

Best Case:

Server Utilization:

- Add 2nd blade chassis
- Add one additional blade
 - CPU usage goes to 71%
 - Memory usage goes to 73%

Storage Capacity:

- Add 4,000 IOPS / 14 TB storage
- Disk space utilization increases to 30 TB (54% of max)
- Disk I/O increases to 15,000 I/Os per second (93% of max)

Network Capacity:

- Network cables, fiber GBICs required
- Add 150Mbps of Internet Bandwidth

500% traffic growth capacity estimates

Best Case:

Server Utilization:

- Add two additional blades
 - CPU usage goes to 72%
 - Memory usage goes to 72%
- Power for second rack required

Storage Capacity:

- Add 12,000 IOPS / 35 TB storage
- Disk space utilization increases to 38 TB (61% of max)
- Disk I/O increases to 18,000 I/Os per second (90% of max)

Network Capacity:

- No additional network equipment required
- Add 150Mbps of Internet Bandwidth

200% traffic growth capacity estimates

Optimal:

Server Utilization:

- Add 2nd blade chassis
- Add one additional blade
 - CPU Utilization goes to 35%
 - Memory usage goes to 62%
- Power for second rack required

Storage Capacity:

- Add remaining storage licensing
- Disk space utilization increases to 17 TB (63% of max)
- Disk I/O increases to 7,000 I/Os per second (87% of max)

Network Capacity:

- Network cables, fiber GBICs required
- Add 150Mbps of Internet Bandwidth

300% traffic growth capacity estimates

Optimal:

Server Utilization:

- Add one additional blade
 - CPU usage goes to 50%
 - Memory usage goes to 60%

Storage Capacity:

- Add 6,000 IOPS / 21 TB storage
- Disk space utilization increases to 22 TB (45% of max)
- Disk I/O utilization to 10,500 I/Os per second (75% of max)

Network Capacity:

- No additional network equipment required
- Add 150Mbps of Internet Bandwidth

400% traffic growth capacity estimates

Optimal:

Server Utilization:

- Add one additional blade
 - CPU usage goes to 64%
 - Memory usage goes to 67%

Storage Capacity:

- Add 4,000 IOPS / 14 TB storage
- Disk space utilization increases to 30 TB (48% of max)
- Disk I/O increases to 15,000 I/Os per second (83% of max)

Network Capacity:

- No additional network equipment required
- Add 150Mbps of Internet Bandwidth

500% traffic growth capacity estimates

Optimal:

Server Utilization:

- Add two additional blades
 - CPU usage goes to 61%
 - Memory usage goes to 61%

Storage Capacity:

- Add 6,000 IOPS / 21 TB storage
- Disk space utilization increases to 38 TB (45% of max)
- Disk I/O increases to 18,000 I/Os per second (75% of max)

Network Capacity:

- No additional network equipment required
- Add 150Mbps of Internet Bandwidth

Initial System Utilization

With current traffic loads the estimated capacity utilization will be:

→ **Server Utilization**

- CPU: 20% or 169 Ghz
- Memory: 61% or 2500 MB
- Network throughput: minimal

→ **Storage Utilization**

- Disk space: Databases: 5 TB
- Disk space: Virtual machines: 7 TB
- Total Disk space requirements: 12 TB
- Disk I/O: 3,500 logical I/Os per second

→ **Network Utilization**

- 150Mbps Internet throughput