

CS441 HW1: Cloud Datacenter Simulators

I. Background

This project consists of three cloud datacenter simulations, each of which embodies the Platform as a Service (PaaS) model and consists of one datacenter and two brokers. The properties **a** through **e** listed below are largely kept equivalent to more effectively observe changes in results.

- a.** **datacenter1** has an x86 system architecture and uses the Linux operating system and Xen virtual machine monitor. The datacenter is based in the UTC-5:00 time zone. The cost per CPU is \$3.50, the cost per memory is \$0.05, and the cost per storage is 0.1¢.
- b.** **datacenter1** hosts two host machines.
 - i.** **host1** is dual-core and has 2 GB of RAM, 10 GB of bandwidth, and 1 TB of storage.
host1 can process 3 billion instructions per second.
 - ii.** **host2** is quad-core and has 4 GB of RAM, 10 GB of bandwidth, and 1 TB of storage.
host2 can process 4 billion instructions per second.
- c.** **broker1** has three virtual machines, each of which has 1 vCPU, 512 MB of RAM, 1 GB of bandwidth, and an image size of 10 GB. Each VM can process 1 billion instructions per second.
- d.** **broker2** has two virtual machines, each of which has 2 vCPU, 1 GB of RAM, 1 GB of bandwidth, and an image size of 10 GB. Each VM can process 2 billion instructions per second.
- e.** **broker1** and **broker2** are tasked with processing 15 cloudlets each, whose lengths vary between 300 billion and 600 billion instructions. Each cloudlet requires one CPU and has a size of 300 bytes.
- f.** In simulation #1, all virtual machines utilize a time-shared cloudlet scheduler policy, in which resources are shared among cloudlets.
- g.** In simulation #2, all virtual machines utilize a space-shared cloudlet scheduler policy, in which each cloudlet is given access to a VM exclusively until its execution completes.
- h.** In simulation #3, all virtual machines utilize a space-shared cloudlet scheduler policy, and the MapReduce framework is implemented to process cloudlets in parallel.

II. MapReduce

The MapReduce implementation in this project, enabled in simulation #3, is comprised of two phases.

- a.** During the **mapper** phase, cloudlets are clustered into varying numbers, between 2 and 9, of smaller subcloudlets. The lengths, sizes and output sizes of the subcloudlets are determined as a proportionate partition of those of each original cloudlet. Mapped subcloudlets originating from a single cloudlet are assigned to the same virtual machine in order to ensure that data locality is used to guide task allocation.
- b.** During the **reducer** phase, the received cloudlets are aggregated to produce the final output. Received subcloudlets are reassigned to and grouped by their original cloudlet IDs. That is,
 - i.** The maximum of the finish times of the mapped subcloudlets is determined to be the finish time of the aggregated cloudlet,
 - ii.** The minimum of the start times of the mapped subcloudlets is determined to be the start time of the aggregated cloudlet,
 - iii.** The difference between the finish time and the start time is determined to be the CPU time of the aggregated cloudlet, and
 - iv.** The sum of the costs of the mapped subcloudlets is determined to be the cost of the aggregated cloudlet.

III. Results

Simulation #1 Results for broker1									
Cloudlet	Status	Datacenter	VM #	Time (ms)	Start	Finish	CPU	Cost (\$)	
12	SUCCESS	1	3	1,596.14	0.10	1,596.24	50.30%	5,586.49	
2	SUCCESS	1	2	1,639.68	0.10	1,639.78	49.09%	5,738.87	
10	SUCCESS	1	1	1,741.79	0.10	1,741.89	49.82%	6,096.27	
15	SUCCESS	1	3	1,825.72	0.10	1,825.82	50.06%	6,390.03	
1	SUCCESS	1	1	1,846.97	0.10	1,847.07	49.09%	6,464.40	
13	SUCCESS	1	1	1,910.04	0.10	1,910.14	51.02%	6,685.14	
8	SUCCESS	1	2	1,913.78	0.10	1,913.88	50.21%	6,698.23	
7	SUCCESS	1	1	2,035.39	0.10	2,035.49	50.22%	7,123.87	
5	SUCCESS	1	2	2,094.45	0.10	2,094.55	49.46%	7,330.59	
3	SUCCESS	1	3	2,123.11	0.10	2,123.21	50.40%	7,430.90	
4	SUCCESS	1	1	2,158.50	0.10	2,158.60	49.40%	7,554.76	
9	SUCCESS	1	3	2,230.54	0.10	2,230.64	49.96%	7,806.89	
6	SUCCESS	1	3	2,284.74	0.10	2,284.84	50.27%	7,996.58	
11	SUCCESS	1	2	2,306.62	0.10	2,306.72	50.40%	8,073.19	
14	SUCCESS	1	2	2,338.62	0.10	2,338.72	49.20%	8,185.17	

In simulation #1, broker1 received 15 cloudlets in 2,338.62 ms
The total cost for broker1 is: \$105,161.38

Simulation #1 Results for broker2									
Cloudlet	Status	Datacenter	VM #	Time (ms)	Start	Finish	CPU	Cost (\$)	
4	SUCCESS	1	2	533.12	0.10	533.22	51.37%	1,865.93	
6	SUCCESS	1	2	612.22	0.10	612.32	51.77%	2,142.78	
10	SUCCESS	1	2	679.80	0.10	679.90	48.24%	2,379.31	
5	SUCCESS	1	1	712.87	0.10	712.97	49.29%	2,495.05	
15	SUCCESS	1	1	734.10	0.10	734.20	49.97%	2,569.36	
2	SUCCESS	1	2	778.89	0.10	778.99	50.66%	2,726.11	
11	SUCCESS	1	1	799.47	0.10	799.57	50.65%	2,798.14	
12	SUCCESS	1	2	804.59	0.10	804.69	48.63%	2,816.08	
14	SUCCESS	1	2	820.40	0.10	820.50	51.72%	2,871.42	
13	SUCCESS	1	1	825.44	0.10	825.54	49.81%	2,889.04	
8	SUCCESS	1	2	830.00	0.10	830.10	49.10%	2,905.00	
1	SUCCESS	1	1	834.18	0.10	834.28	48.73%	2,919.64	
9	SUCCESS	1	1	855.44	0.10	855.54	52.63%	2,994.05	
7	SUCCESS	1	1	891.51	0.10	891.61	50.45%	3,120.30	
3	SUCCESS	1	1	904.81	0.10	904.91	49.49%	3,166.83	

In simulation #1, broker2 received 15 cloudlets in 904.81 ms
The total cost for broker2 is: \$40,659.05

Simulation #1: Time-shared cloudlet scheduler policy

- Although **broker1** has more virtual machines (3 VMs as opposed to **broker2**'s 2 VMs), the degraded technical specs of those virtual machines result in reduced performance.
 - The total execution time for **broker1** to process 15 cloudlets is 2,338.62 milliseconds.
 - The total execution time for **broker2** to process 15 cloudlets is 904.81 milliseconds.
 - The total cost for **broker1** to process 15 cloudlets is \$105,161.38.
 - The total cost for **broker2** to process 15 cloudlets is \$40,659.05.
 - *Takeaway:* Simply doubling the number of cores to 2 vCPU and the memory to 1 GB resulted in more than a 60% reduction in processing time and cost. In other words, the number of cores and memory (RAM) in a virtual machine have an inversely proportionate relationship with CPU time and cost.
- Although all 15 cloudlets share resources and start at the same time at 0.1 milliseconds, the cloudlets take much longer to complete execution as compared to what is observed in the subsequent simulation.
 - *Takeaway:* Context-switching and saving state information of cloudlets as they are processing and paused to grant other cloudlets CPU access substantially reduces performance and increases cost.

Simulation #2 Results for broker1									
Cloudlet	Status	Datacenter	VM #	Time (ms)	Start	Finish	CPU	Cost (\$)	
3	SUCCESS	1	3	428.90	0.10	429.00	52.12%	1,501.14	
1	SUCCESS	1	1	458.07	0.10	458.17	49.59%	1,603.26	
2	SUCCESS	1	2	538.68	0.10	538.78	50.13%	1,885.39	
6	SUCCESS	1	3	408.70	429.00	837.70	50.65%	1,430.44	
5	SUCCESS	1	2	415.49	538.78	954.27	52.03%	1,454.20	
4	SUCCESS	1	1	533.80	458.17	991.97	48.27%	1,868.29	
9	SUCCESS	1	3	408.44	837.70	1,246.13	47.89%	1,429.53	
8	SUCCESS	1	2	328.34	954.27	1,282.60	50.55%	1,149.18	
7	SUCCESS	1	1	325.80	991.97	1,317.77	51.17%	1,140.30	
12	SUCCESS	1	3	522.69	1,246.13	1,768.83	49.94%	1,829.43	
11	SUCCESS	1	2	502.79	1,282.60	1,785.40	50.24%	1,759.78	
10	SUCCESS	1	1	531.94	1,317.77	1,849.71	50.29%	1,861.78	
15	SUCCESS	1	3	341.12	1,768.83	2,109.95	48.55%	1,193.92	
13	SUCCESS	1	1	409.90	1,849.71	2,259.61	50.69%	1,434.64	
14	SUCCESS	1	2	477.56	1,785.40	2,262.96	49.51%	1,671.47	

In simulation #2, broker1 received 15 cloudlets in 2,262.86 ms

The total cost for broker1 is: \$23,212.76

Simulation #2 Results for broker2									
Cloudlet	Status	Datacenter	VM #	Time (ms)	Start	Finish	CPU	Cost (\$)	
3	SUCCESS	1	1	179.56	0.10	179.66	51.64%	628.47	
2	SUCCESS	1	2	210.94	0.10	211.04	48.95%	738.30	
4	SUCCESS	1	2	245.91	0.10	246.01	47.96%	860.68	
1	SUCCESS	1	1	265.76	0.10	265.86	49.47%	930.15	
5	SUCCESS	1	1	206.08	179.66	385.74	47.75%	721.27	
6	SUCCESS	1	2	232.41	211.04	443.45	51.02%	813.43	
7	SUCCESS	1	1	211.60	265.86	477.46	53.39%	740.60	
8	SUCCESS	1	2	263.73	246.01	509.73	49.02%	923.04	
10	SUCCESS	1	2	170.91	443.45	614.37	52.20%	598.20	
9	SUCCESS	1	1	243.06	385.74	628.80	48.72%	850.71	
11	SUCCESS	1	1	168.87	477.46	646.32	53.71%	591.04	
12	SUCCESS	1	2	190.56	509.73	700.29	48.40%	666.95	
14	SUCCESS	1	2	198.01	614.37	812.38	51.83%	693.04	
13	SUCCESS	1	1	193.55	628.80	822.35	51.39%	677.42	
15	SUCCESS	1	1	220.81	646.32	867.13	54.32%	772.83	

In simulation #2, broker2 received 15 cloudlets in 867.03 ms

The total cost for broker2 is: \$11,206.13

Simulation #2: Space-shared cloudlet scheduler policy

- Although all 15 cloudlets start at varying times, the cloudlets complete execution *much* more rapidly.
 - The total execution time for broker1 to process 15 cloudlets is 2,262.86 milliseconds.
 - The total execution time for broker2 to process 15 cloudlets is 867.03 milliseconds.
 - The total cost for broker1 to process 15 cloudlets is \$23,212.76.
 - The total cost for broker2 to process 15 cloudlets is \$11,206.13.
 - *Takeaway:* Context-switching and saving state information of cloudlets as they are processing and paused to grant other cloudlets CPU access reduces performance and substantially increases cost. Simply shifting from a time-shared cloudlet scheduler policy to one that is space-shared resulted in a 5% improvement in performance, but a 70-80% reduction in cost.

Simulation #3 Results for broker1									
Cloudlet	Status	Datacenter	VM #	Time (ms)	Start	Finish	CPU	Cost (\$)	
1	SUCCESS	1	1	235.61	0.10	235.71	52.15%	824.64	
3	SUCCESS	1	3	278.88	0.10	278.98	64.32%	976.07	
2	SUCCESS	1	2	348.65	0.10	348.75	61.19%	1,220.26	
6	SUCCESS	1	3	295.52	278.98	574.50	54.27%	1,034.33	
5	SUCCESS	1	2	227.66	348.75	576.41	49.46%	796.83	
4	SUCCESS	1	1	373.13	235.71	608.85	51.53%	1,305.97	
9	SUCCESS	1	3	349.03	574.50	923.53	59.36%	1,221.62	
7	SUCCESS	1	1	321.85	608.85	930.70	55.38%	1,126.48	
8	SUCCESS	1	2	357.14	576.41	933.55	55.45%	1,249.98	
11	SUCCESS	1	2	305.80	933.55	1,239.35	53.86%	1,070.32	
10	SUCCESS	1	1	311.51	930.70	1,242.21	56.82%	1,090.29	
12	SUCCESS	1	3	343.12	923.53	1,266.65	56.25%	1,200.91	
14	SUCCESS	1	2	267.57	1,239.35	1,506.93	54.95%	936.51	
15	SUCCESS	1	3	319.32	1,266.65	1,585.97	55.51%	1,117.62	
13	SUCCESS	1	1	411.93	1,242.21	1,654.14	48.13%	1,441.76	

In simulation #3, broker1 received 15 cloudlets (via 77 slaves) in 1,654.04 ms

The total cost for broker1 is: \$16,613.58

Simulation #3 Results for broker2									
Cloudlet	Status	Datacenter	VM #	Time (ms)	Start	Finish	CPU	Cost (\$)	
2	SUCCESS	1	2	67.10	0.10	67.20	64.41%	423.91	
1	SUCCESS	1	1	81.94	0.10	82.04	60.74%	502.87	
4	SUCCESS	1	2	93.75	54.12	147.87	53.89%	569.17	
3	SUCCESS	1	1	114.03	61.83	175.86	60.25%	583.69	
6	SUCCESS	1	2	93.19	136.07	229.26	57.49%	537.58	
5	SUCCESS	1	1	115.67	134.78	250.46	57.40%	539.24	
8	SUCCESS	1	2	89.38	208.27	297.66	56.58%	525.77	
7	SUCCESS	1	1	106.40	214.26	320.66	49.97%	556.44	
10	SUCCESS	1	2	79.30	290.10	369.40	58.48%	456.00	
9	SUCCESS	1	1	85.77	303.04	388.81	57.85%	526.92	
11	SUCCESS	1	1	88.67	385.44	474.11	54.29%	594.70	
12	SUCCESS	1	2	126.51	348.64	475.15	53.85%	631.16	
14	SUCCESS	1	2	112.14	423.22	535.36	58.28%	583.62	
13	SUCCESS	1	1	74.60	470.06	544.66	51.38%	488.66	
15	SUCCESS	1	1	83.77	539.13	622.90	53.02%	448.27	

In simulation #3, broker2 received 15 cloudlets (via 66 slaves) in 662.80 ms

The total cost for broker2 is: \$7,968.01

Simulation #3: Space-shared cloudlet scheduler policy with MapReduce

- MapReduce further improved performance and cost measures.
 - The total execution time for broker1 to process 15 cloudlets is 1,654.04 milliseconds.
 - The total execution time for broker2 to process 15 cloudlets is 662.80 milliseconds.
 - The total cost for broker1 to process 15 cloudlets is \$16,613.58.
 - The total cost for broker2 to process 15 cloudlets is \$7,968.01.
 - Peak CPU utilization increased by up to 10%.
 - *Takeaway:* Dividing data tasks into smaller ones, grouping them via data locality and allowing them to run in parallel results in a 30-40% improvement in performance and cost.