Problem Set 3

15-440/15-640 Distributed Systems Spring 2016

Assigned: Thursday March 24, 2016

Due: Thursday March 31, 2016 (by the start of class)

Submission procedure:

- Create a .pdf of your answers and upload to Autolab**.**
- If you handwrite your answers, scan into .pdf before uploading. Use a proper scanner, not your smartphone camera. If you answers are illegible, you will receive a zero grade.

Question 1 (14 points)

A. What is the difference between “scale up” and “scale out”?

B. Give two reasons why scale up may not benefit a particular application.

C. Give two reasons why scale out may not benefit a particular application.

Question 2 (24 points)

As a senior engineer at EBooks_R_US, you are asked to improve the scalability of the e-book service so that it can handle at least ***R*** read requests per second (***R*** is specified below). Of course, you want this improvement to be achieved at the least cost.

Currently, there are 15 Wimpy server machines that can each handle 10 reads per second. This gives an ***R*** value of 150 read requests per second. You receive the following quotes for new machines.

Machine Type	Target rate of requests handled per machine (reads/sec)	Price for the new machine (USD)
Beefy-A	20	400
Beefy-B	30	750
Wimpy	10	100

You can scale up by replacing existing machines. Or, you can scale out by adding more machines. An additional expense to be considered is that of system administration. The first 30 machines (of any capacity) can be administered by you personally; no additional personnel are required. After that, every 30 machines or part thereof (of any capacity) require an operator who costs $5400.

For simplicity, you can ignore all other costs such as power, cooling, software licenses, etc. For the following values of R, determine the optimal (i.e., least cost) configuration change (i.e., combination of scale up and scale out). Explain your reasoning in all cases.

A. For R = 300, what is the optimal configuration? What is its cost?
B. For R = 900, what is the optimal configuration? What is its cost?
C. For R = 1000, what is the optimal configuration? What is its cost?

Question 3 (15 points)

Suppose you have access to a compute cluster that can run MapReduce jobs, as well as a supercomputer that can efficiently run MPI jobs. You would like to pick the right tool for the right job. For each of the following, state whether the use case is a better fit for MapReduce, MPI, or neither. Explain your answer in each case. If MapReduce is a better fit, briefly describe what mappers and reducers do. If MPI is a better fit, briefly describe why it wins over MapReduce.

A. Compute the average of all pixels in a 1000x1000 image

B. For each of 10^6 images, compute the average value of all pixels in an image. Each image is of size 1000x1000 pixels.

C. Process an image of size 10^6 by 10^6 pixels, where the processing of each pixel depends on all of the neighboring pixels, and the processing involves several thousand iterations.

Question 4 (12 points)

A cloud storage service allows creation and deletion of blobs (blob = “binary large object”). Once created, a blob cannot be modified. The storage service uses one master node and multiple data nodes. Creation of a new blob occurs at the master node. A unique identifier is first assigned to the blob. The blob is then broken into *chunks* whose size is normally distributed, with a mean of 1 MB and a standard deviation of 100 KB. Chunks are content-addressed. In other words, the SHA-256 hash of a chunk’s content is used as that chunk’s unique identifier. How to reconstruct the blob from its chunks is documented in the *recipe* of that blob, which is stored at the master node.

A. State one advantage and one disadvantage of using content addressing for chunks.
(Hint: think about deletion of blobs.)

B. Consider two alternative designs. The first design maps chunks to nodes based on the low order bits of the chunk identifier. For example, if there are 4096 data notes, the low order 12 bits of the chunk identifier determine the mapping. The second design treats the data nodes as a DHT that is addressed by chunk identifier. Which of these two designs is better? Justify your answer.

Question 5 (15 points)

A client/server model and a DHT are two alternative ways of organizing a large collection of read-only data from many users, such as encrypted backup snapshots of their home PCs. In this context, answer the following questions.

A. State and explain why you might prefer the client/server model.

B. State and explain why you might prefer a DHT.

C. In a real world situation, how will you decide between these alternatives?

Question 6 (20 points)

Which of the following workloads would MapReduce be good for? In each case, provide a couple of sentences justifying your answer.

A. Building a real-time news aggregation service that provides up-to-the-minute news updates. It should be designed to parse feeds from multiple news websites and refresh its index in real-time.

B. Building an interactive data exploration tool which lets users provide their own queries and rapidly see partial results from their queries without waiting for the query to run over an entire dataset.

C. Building a high-frequency trading application. You've been asked to architect a high-frequency stock trading application which needs to consume stock ticker updates and provide quick decisions for trades.

D. Building a database of features from images for face recognition. The input is hundreds of millions of user uploaded and tagged photos, each stored as a file in the local file system. For each image an algorithm must run that extracts features for training face recognition algorithms in a later stage.

E. Crawling through billions of log entries in log files, one per line of input, across thousands of websites and counting the number of unique visitors (unique by IP address) that access those websites. This will feed daily, weekly, and monthly reporting systems that provide website analytics.