**Guided Reading Questions**

Sato, K., “An Inside Look at Google Big Query,” Google, 2012, accessed 2/3/19, https://cloud.google.com/files/BigQueryTechnicalWP.pdf.

note: skip all the sections that start on pp. 10-11.

1. What is unique or interesting about BigQuery? The technology is built around Google’s proprietary software “Dremel” and is accessible to the public.
2. What are the intended use cases or application areas for BigQuery? BigQuery is intended for one shot use cases to generate information about data in a very quick timeframe. It can be used to get information to even non-programming folks without a significant learning curve.
3. What other technologies have been proposed to address the same or similar use cases? BiqQuery is built on Dremel, but a lot of the buzz is whether or not MapReduce (or the open-source Hadoop) is more effective than BigQuery for handling large amounts of data.
4. What is the relationship between Dremel and BigQuery? Dremel is the private, proprietary software that Google used to build BigQuery. It is the publicly available version of Dremel.
5. What two core technologies are responsible for Dremel’s unprecedented performance? Columnar Storage (which achieves “high compression ratio and scan throughput”) and Tree Architecture(Able to “Dispatch queries and aggregate results across thousands of machines in a few seconds”)
6. What are two advantages of columnar storage? A high compression ratio and scan throughput are both achieved by this format.
7. What kind of typical database operation is column storage not suited for? Columnar is not optimized for update operations.
8. What is the purpose of the tree architecture? Tree architecture enables the quick processing of multiple queries and the aggregation of results in a few seconds.
9. What is the relationship between MapReduce and Hadoop? Hadoop is the open-source version of MapReduce.
10. What are intended use cases of Hadoop? Hadoop has historically been used for log analysis, user activity analysis for social apps, recommendation engines, unstructured data processing, data mining, text mining, and others.
11. What kind of database operations is MapReduce not suitable for? Why? Because MapReduce is optimized for large specific data analytics and batch processing, it runs incredibly slow when trying to do ad-hoc queries or trial and error data analytics. MapReduce is unsuited for these use cases.
12. What advantages does using a cloud-based solution like BigQuery offer to an organization (compared to a traditional hardware/software solution)? Data can be stored and queries can be conducted with absolutely no need to purchase hardware to store anything. A distributed system also enables faster parallel processing capabilities that are not possible on a single sever box.