CompSci 557

Kayla Goetzke

Phil Sauvey

Chapter 14

2. The 3 V’s of big data are volume, velocity, and variety. Volume is the quantity of the data to be stored. Velocity is the speed at which data is entering the system. Variety is the variations in the structure of the data to be stored.

3.

4. Scaling up happens when you move to faster and larger system. While scaling out is when the workload exceeds the capacity, so you distribute the workload among a number of servers.

5. Stream processing is the process of analyzing the data stream as it enters the system. It decides what data to keep and what to discard before storing it.

6. Feedback loop processing is more focused on the outputs; whereas, stream processing is focused on the inputs.

9. The first key assumption made by Hadoop Distributed File System(HDFS) is that the files will be extremely large. Next, the write-once, read-many model is when a file is created, written and then closed. After it is closed, it cannot be changed. Advancements in HDFS all data to be appended to the end of the file. Another assumption is streaming access. Hadoop is optimized for batch processing of entire files as a continuous stream. The last assumption is fault tolerance. Fault tolerance is when there are multiple copies of the data on multiple devices, so if one fails you can find it on another device.

10.

11.

12.

13. The four basic categories of NoSQL are key-value database, document databases, column-oriented databases, and graph databases.

14.

15.