Jacqueline Henry

April 28th, 2017

Hadoop Technology

A topic that comes up several times in Management Information Systems courses is distributed data technologies, or more specifically, Hadoop. Hadoop is a software from Apache that stores data using the Distributed File System over multiple nodes ("Hadoop and Big Data"). Nodes are computers made for the purpose of storing data. Hadoop has advantages: it is open source (free), it is easy to scale by adding more nodes, is flexible enough that companies can add data as they go without preprocessing, and it is faster the more nodes are added (Phillip). Data storage through Hadoop is not susceptible to hardware failure because if one node fails, the data is automatically transferred to another node.

In MIS2502, processing and storing data was discussed. The class looked at how relational databases work and how SQL can be used to put information in or process that data out. Hadoop can store multiple databases, a data warehouse, or unstructured data. With the massive amounts of storage MapReduce is used. MapReduce maps out the data on all of the nodes, and then reduces those results to formulate an answer to the query(Gualtieri). Companies that have increasing, inconsistent, influxes of data should use Hadoop. Social media is growing, which means companies need to be able to store and analyze large sets of unstructured data. The professional social media website, Linkedin, uses Hadoop for data storage. Hadoop is flexible, making it easy to add more data to the storage. Linkedin needs that feature because every second two new users come aboard ("How LinkedIn uses Hadoop to leverage Big Data Analytics?"). Linkedin uses association mining, talked about in MIS2502, to give connection and job recommendations. The user data on Linkedin is unstructured, therefore Hadoop is an efficient way to store and perform analysis.

Works Cited

"Hadoop & Big Data." *What Is Apache Hadoop?* MapR Technologies, n.d. Web. 28 Apr. 2017.

"How LinkedIn Uses Hadoop to Leverage Big Data Analytics?" *DeZyre*. N.p., 10 Mar. 2016.

Web. 28 Apr. 2017.

Phillips, Nate. "Hadoop vs. Traditional Database: Choose a Big Data Database." *Qubole*.

Qubole, 01 Aug. 2016. Web. 28 Apr. 2017.

What Is Hadoop? Dir. Mike Gualtieri. Perf. Mike Gualtieri. *Youtube*. Youtube, 7 June 2013. Web.

28 Apr. 2017.