

Apache Hadoop



What is Apache Hadoop and Why is it Important?

Apache Hadoop is an open-source software framework project written in Java that develops software for distributed storage and processing of large data sets on computer clusters built from commodity hardware. The main focus of the modules in Hadoop is to start from single servers and scale up to thousands of machines, each of which feature local computation and storage. Hadoop handles data by splitting files into large blocks and distributing them across nodes in a cluster. Packaged code for nodes is then transferred by Hadoop to be processed in parallel based on the data that needs to be processed. What makes Hadoop and its features so important is that the data sets are processed faster and more efficiently, making more conventional super computer architecture that rely on a parallel file system less convenient.

Data Analytics and Apache Hadoop

MIS 2502 students are taught how to extract data from a certain source, transform it to a new format, and load it into one database where the data can be viewed. Students are then taught how to properly analyze the resulting data and how to make inferences and conclusions about the data to get a better understanding of the methods taught in the course. These techniques are in a way similar to Apache Hadoop in that the two involve extracting, transforming, and loading data. Hadoop does this on a very wide scale, by taking the large amounts of data and splitting them up so that they can be more easily processed and handled. Almost every assignment in MIS 2502 can be related to this as a lot of them involve processes similar to this.

Apache Hadoop in Practice

Apache Hadoop has been used much more in the IT field as of late. An example of its use is the software company Cloudera using Apache Hadoop for cybersecurity solutions. Cloudera feels that Hadoop-based technologies remove traditional and data constraints to gain visibility and detection capabilities that work well against more modern cyber-attacks. Cloudera has used their experience with Hadoop to educate other industries on the open-source software framework.

Works Cited

"Welcome to Apache™ Hadoop®!" *Welcome to Apache™ Hadoop®!* N.p., n.d. Web. 21 Apr. 2016. <<http://hadoop.apache.org/>>.

"Cloudera CEO and Intel Chief Cloud Security Architect Explain Why Organizations Are Turning to Apache Hadoop for Cybersecurity in Strata + Hadoop World Keynote." *GlobeNewswire News Room*. N.p., n.d. Web. 21 Apr. 2016.
<<https://globenewswire.com/news-release/2016/03/30/824028/0/en/Cloudera-CEO-and-Intel-Chief-Cloud-Security-Architect-Explain-Why-Organizations-Are-Turning-to-Apache-Hadoop-for-Cybersecurity-in-Strata-Hadoop-World-Keynote.html>>.

Created for MIS 2501.