PROTERE™
TECHNOLOGIES

BIG DATA ANALYTICS

Big Data Analytics

Description:

With the advance of IT storage, pcoressing, computation, and sensing technologies, Big Data has become a novel norm of life. Only until recently, computers are able to capture and analysis all sorts of large-scale data from all kinds of fields -- people, behavior, information, devices, sensors, biological signals, finance, vehicles, astronology, neurology, etc. Almost all industries are bracing into the challenge of Big Data and want to dig out valuable information to get insight to solve their challenges.

This course shall provide the fundamental knowledge to equip students being able to handle those challenges. This discipline inherently invoves many fields. Because of its importance and broad impact, new software and hardware tools and algorithms are quickly emerging. A data scientist needs to keep up with this ever changing trends to be able to create a state-of-the-art solution for real-world challenges.

This Big Data Analytics course shall first introduce the overview applications, market trend, and the things to learn. Then, I will introduce the fundamental platforms, such as Hadoop, Spark, and other tools, such as IBM System G for Linked Big Data. Afterwards, the course will introduce several data storage methods and how to upload, distribute, and process them. This shall include HDFS, HBase, KV stores, document database, and graph database. The course will go on to introduce different ways of handling analytics algorithms on different platforms. Then, I will introduce visualization issues and mobile issues on Big Data Analytics. Students will then have fundamental knowledge on Big Data Analytics to handle various real-world challenges.

Afterwards, the course will zoom in to discuss large-scale machine learning methods that are foundations for artificial intelligence and cognitive networks. The course will discuss several methods to optimize the analytics based on different hardware platforms, such as Intel & Power chips, GPU, FPGA, etc. The lectures will conclude with introduction of the future challenges of Big Data, especially on the onging Linked Big Data issues which involves graphs, graphical models, spatio-temporal analysis, cognitive analytics, etc.

Students will choose the topics of their own for a final project. The application domain can be based on the students' own interest. This will be a good opportunity for students to apply what's learned in the class for their needs, either for the future work requirements or for the research problems at hand.

Prerequisites:

The Users, Administrators, Architects, Developers & Support Engineers. It's recommended for anyone in your organization who needs to look at IT data – from help desk staff to systems administrators and developers. This will be a hands-on course. Participants need to know at least one or more programming languages: C, C++, Java, Perl, Python, and/or Javascript to finish homeworks and final project.

Course Outline:

- **Overview of Big Data**
  This includes topics such as history of big data, its elements, career related knowledge, advantages, disadvantages and similar topics.

- **Using Big Data in Businesses**
  This module should focus on the application perspective of Big Data covering topics such as using big data in marketing, analytics, retail, hospitality, consumer good, defense etc.

- **Technologies for Handling Big Data**
  Big Data is primarily characterized by Hadoop. This module cover topics such as Introduction to Hadoop, functioning of Hadoop, Cloud computing (features, advantages, applications) etc.

- **Understanding Hadoop Ecosystem**
  This includes learning about Hadoop and its ecosystem which includes HDFS, MapReduce, YARN, HBase, Hive, Pig, Sqoop, Zookeeper, Flume, Oozie etc.

- **Dig Deep to understand the fundamental of MapReduce and HBase**
  This module should cover the entire framework of MapReduce and uses of mapreduce.

- **Understanding Big Data Technology Foundations**
  This module covers the big data stack i.e. data source layer, ingestion layer, source layer, security layer, visualization layer, visualization approaches etc.

- **Databases and Data Warehouses**
  This module should cover all about databases, polygot persistence and their related introductory knowledge.

- **Using Hadoop to store data**
  This includes an entire module of HDFS, HBase and their respective ways to store and manage data along with their commands.

- **Learn to Process Data using Map Reduce**
  This emphasizes on developing simple mapreduce framework and the concepts applied to it.

- **Testing and Debugging Map Reduce Applications**
  After the applications are developed, the next step is to test and debug it. This modules imparts this knowledge.

- **Learn Hadoop YARN Architechture**
  This module covers the background of YARN, advantages of YARN, working with YARN, backward compatibility with YARN, YARN Commands, log management etc.

- **Exploring Hive**
  This modules introduces you with all the necessary knowledge of Hive.

- **Exploring Pig**
  This modules introduces you with all the necessary knowledge of PIG.

- **Exploring Oozie**
  This modules introduces you with all the necessary knowledge of Oozie.

- **Learn NoSQL Data Management**
  This modules covers all about NoSQL including document databases, relationships, graph databases, schema less databases, CAP Theorem etc.

- **Integrating R and Hadoop and Understanding Hive in Detail**
  This module introduces you to RHadoop, ways to do text mining and related knowledge.

- **Splunk**
  Overview of machine data
  How Splunk works with machine data
  Introduction to Splunk's user interface

- **Start Searching**
  Introduce Splunk and the Search app
  Run basic searches
  Identify the contents of search results
  Control a search job
  Set the time range of a search
  Use the output of a search to refine your search
- **Saving Results and Searches**
  Export search results
  Save and share search results
  Save searches
  Schedule searches
- **Using Fields**
  Understand fields
  Use fields in searches
  Use the fields sidebar
- **Tags and Event Types**
  Understand tags
  Create tags and use tags in a search
  Describe event types and their uses
  Create and use event types in a search
- **Creating Alerts**
  Describe alerts
  Create an alert
  View fired alerts
- **Creating Reports**
  Create reports and charts
  Create dashboards and add reports