

# **BIG DATA ANALYSIS & Hadoop**

## **About:-**

Big data is the process of examining large amount of data of variety of types to uncover hidden patterns, unknown correlations and other useful information. The primary goal of big data analytics is to help organizations make better business decision. Big data analytics can be done with software tools commonly used as a part of advanced analytics, disciplines such as predictive analytics and data mining. Hadoop is 100% open or free source, and pioneered a fundamentally new way of storing and processing data. Instead of relying on expensive, proprietary hardware and different systems to store and process data, Hadoop enables distributed parallel processing of huge amounts of data across inexpensive, industry- standard servers that both store and process the data, and can scale without limits. With Hadoop, no data is too big. The objective of this workshop is to understand the challenges in architectures to store, and access the Big data, perform analytics on Big data for data intensive applications. This workshop aims at bringing the students, professionals, academics and research scholars together to share their knowledge for achieving focused development and advancement in the field of BIG DATA ANALYSIS

## **Event Details:-**

DAY 1

### **Fundamentals of Big data**

- What is big data • Application of big data • Tools for big data • How the financial service sector can employ analytics • Role of big data will play in strategic decision making in the financial services industry
- Big Data Cybersecurity Analytics • Scalable and distributed data mining • Large scale data filtering, aggregation and association Big data management • Distributed file systems for big data storage, access, and analytics • Frameworks and tools for big data cybersecurity analytics • Performance modeling, simulation and analysis • Big data applications in cybersecurity • Parallel and distributed algorithms for big data analytics • Big data case studies and applications

### **HADOOP**

- What is Hadoop? • Why Hadoop and its Uses (Demo Part) • Different Components of Hadoop • Introduction to Apache Pig • Map Reduce vs. Apache Pig • SQL vs. Apache Pig • Different Data Types in Pig • Modes of Execution in Pig - Local Mode • Execution Mechanism - Grunt Shell - Script - Embedded • Transformations in Pig • How to Write a Simple Pig script • UDFs in Pig

## DAY 2

### **HDFS (Hadoop Distributed File System)**

• Significance of HDFS in Hadoop • Features of HDFS • 5 daemons of Hadoop - NameNode and its Functionality - DataNode and its Functionality - JobTracker and its Functionality - TaskTrack and its Functionality - Secondary NameNode and its Functionality • Data Storage in HDFS - Introduction about Blocks - Data Replication • Accessing HDFS - CLI (Command Line Interface) and Admin Commands - Java Based approach • Hadoop Archives

### **HIVE**

• HIVE Introduction • HIVE Architecture • HIVE Meta Store • HIVE Integration with Hadoop

### **MAP REDUCE**

• MapReduce Architecture • MapReduce Programming Model - Different Phases of MapReduce Algorithm - Different Data Types in MapReduce - How to Write a Basic MapReduce Program - Joining Datasets In MapReduce Jobs - Map Joins and Reduce Joins - Creating Input and Output Formats in MapReduce Jobs 1. Text Input Format 2. Key Value Input Format 3. Sequence File Input Format - How to Debug MapReduce Jobs in Local and Pseudo Cluster Mode - Unit Testing MR Jobs - Introduction to MapReduce Streaming and Pipelining - Data Localization in MapReduce - Combiner (Mini Reducer and Partitioner) • Distributed Cache • Secondary Sorting Using MapReduce

### **Benefits:-**

- Learn & Interact with renowned Industry Experts.
- Receive an unparalleled education on the art of computer security with personal one on-one attention.
- 2 days Hand on Practical Workshop Big Data and Hadoop.
- Students will get a live Cd free of cost.
- Practical Demonstration application development will be covered during the workshop.
- Three practical projects to be undertaken during the workshop.
- **Certificate of Participation** to all Participants from Aparoksha.

### **Register:-**

Any Individual can register for this workshop.

This is a 2 day workshop (16 hours).

**Date:** 20 - 21 March.

Cost: Rs. 1000 per participant

**Contacts:-**

Anuj Srivastava :- 9044462057

Chaitanya Agrawal :- 8174070273