

Big Data: Hadoop Administration and Developer

Session 1: Introduction to Traditional Databases	
➤ Introduction to Database	➤ 3 tier Architecture, Data Models
➤ Entity Relationship Model	➤ ER Diagram
Session 2: SQL (Structured Query Language)	
➤ Create Database, Drop Database	➤ Create Table and Insert Values
➤ Queries, Logical Operators (AND, OR, NOT)	➤ Update, Delete, Like and TOP Clause
Session 3: SQL Continues	
➤ Order by, Group by, Distinct Keyword	➤ SQL Constraints, Using Joins, UNION, NULL values
➤ Using Alias and Truncate, Having Clause	➤ Table Cloning, Subqueries
Session 4: Data Backups	
➤ Backup Entire Database, Backup Single Database	➤ Backup Single Table, Challenges in Traditional Databases, Fragmented Resources
➤ Data Security, Unpredictable cost	➤ Data Retrieval
Session 5: Fundamentals of Linux & VMWare Installation	
➤ Introduction to Linux OS, Types of Linux Version	➤ Their Usage, Linux Installation, Linux Commands
➤ Introduction to VMWare, Installing VMWare	➤ Explaining the Packages
Session 6: Introduction to HADOOP	
➤ Hadoop Architecture, Map Reduce	➤ Hadoop Distributed File System
➤ Environment Setup, Difference between Hadoop1, Hadoop2, Hadoop3	➤ Man in the middle attack
Session 7: HDFS Overview	
➤ HDFS Architecture, Data Node	➤ Importing Data into HDFS, Map Reduce
➤ Map Reduce Job Management, Name Node	➤ HDFS Commands
Session 8: Single Node Cluster Configuration	
➤ Hadoop Prerequisites, Hadoop Installation	➤ Hadoop Configuration
Session 9: Multi Node Cluster Configuration	
➤ Master and Slave Concept with Examples	➤ Hadoop Installation and Configuration with M/S
Session 10: Cluster Maintenance	
➤ Checking HDFS Status, Breaking the Cluster	➤ Adding and Removing Cluster Nodes
➤ Rebalancing the Cluster, Cluster Upgrading	➤ Copying Data Between Cluster
Session 11: Hadoop Administration and Hadoop Ecosystem(Hive)	
➤ Hadoop High Availability, Hadoop Multi Network	➤ Introduction to data Warehouse, Hive Architecture
➤ Installing Hive, Data Management using Hive	➤ Hive Partitioning, Hive Bucketing, Hive Serde, HQL
Session 12: Hadoop Ecosystem (Pig)	
➤ Pig Overview, The need of Apache Pig	➤ Apache Pig Architecture, Downloading and Installing Pig, Pig Latin Basics
➤ Latin Built in functions and data management	➤ Pig UDF, Pig Eval Functions
Session 13: Cluster Monitoring, Troubleshooting & Optimization	
➤ Checking HDFS with fsck, Breaking the cluster	➤ Copying data with distcopy
➤ Rebalancing cluster nodes, Adding and removing cluster nodes	➤ Clusters self-healing features
Session 14: Hadoop Ecosystem (Sqoop and Flume)	
➤ Introduction to Sqoop, Server Installation	➤ Downloading and Installing Packages
➤ Client Installation, Upgrading Server	➤ Sqoop Jobs, Sqoop Incremental Append
➤ Sqoop Advanced, The need of Apache Flume	➤ Data Management using Flume, Installing Flume

Session 15: Restoring Data and Troubleshooting Cluster	
➤ Process Understanding, Data Restoring	➤ Pre-requisites for data restore
➤ Validate Environment Information	➤ Validate Hadoop Cluster Health
➤ Troubleshooting HDFS	➤ Troubleshooting HIVE
Session 16: Introduction to other tools	
➤ Introduction to Spark, Kafka, Presto, Samza	➤ Introduction to Flink, Power BI, Tableau
➤ Connecting with HDFS	➤ Visualization data with power BI Tools
➤ Projects based on Bigdata and Hadoop	