

Big Data: Hadoop Administration

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 Queries	Like and TOP Clause

Session-3: SQL Continues

Order by	Group by
Distinct keyword	SQL Constraints
Using joins	UNION

Session-4: SQL Continues

Union Clause	NULL Values
Using alias and truncate	Having clause
Table Cloning	Subqueries

Session-5: Data Backup

Backup Entire Database	Backup single Database
Backup Single Table	

Session-6: Challenges in Traditional Databases

Fragmented Resources	The Emerging Data Libraries
Database Engine Architecture	Unstructured Data
Data Loss/ Theft	Data Security

Session-7: Challenges continues

Capacity Planning	Backup for backup
Unpredictable cost	Bandwidth Saturation
Data Storage	Data Retrieval

Session-8: Introduction to HADOOP

Hadoop Architecture	MapReduce
Hadoop Distributed File System	Environment Setup
Session-9: HDFS Overview	
HDFS Architecture	Data node
Importing Data into HDFS	MapReduce
MapReduce Job Management	HDFS Commands
Session-10: Single Node Cluster Configuration	
Hadoop Prerequisites	Hadoop Installation & Configuration
Session-11: Multi Node Cluster Configuration	
Hadoop Prerequisites	Hadoop Installation & Configuration
Session-12: Cluster Maintainance	
Checking HDFS Status	Breaking The Cluster
Adding and Removing Cluster Nodes	Rebalancing The Cluster
Copying Data between Cluster	Cluster Upgrading
Session-13: Hadoop Ecosystem (Sqoop)	
Introduction to Sqoop	Downloading & Installing package
Server installation	Client Installation
Upgrading Server	
Session-14: Hadoop Ecosystem (Flume)	
The need of Apache Flume	Downloading & installing Flume
Data management using Flume	
Session-15: Hadoop Ecosystem (Hive)	
Introduction to Data Warehouse	Hive Architecture
Installing Hive	Data management using Hive
Session-16: Hadoop Ecosystem (Pig)	
Pig Overview	The Need of Apache Pig
Apache Pig Architeccture	Downloading & Installing Pig
Pig Latin basics	Latin Built in functions and data management
Session-17: Cluster Monitering, Troubleshooting & Optimisation	
checking HDFS with fsck	Breaking the cluster

copying data with distcopy	Rebalancing cluster Nodes
Adding and removing cluster nodes	clusters self healing feature
Session-18: Data Backup	
Understand the process	Pre requisits for data backup
Backing up hadoop cluster	
Session-19: Restoring Data	
Process understanding	Pre requisits for data restore
Data restoring	
Session-20: Manage Hadoop Log Files	
Understand server Log	Need of Data Visualisation
Apache Zeppelin	Challenges in processing log file
Session-21: Observium	
Introduction to observium	Hardware & Software requirements
OS Installation	Customisation
Session-22: Ganglia	
Overview	Download & Install ganglia
Customisation	
Session-23: Troubleshooting Cluster	
Validate Environment Information	Validate Hadoop Cluster health
Troubleshooting HDFS	Troubleshooting Hive