Big Data Hadoop Course Content

Chapter 1: Introduction to big data-hadoop

- Overview of Hadoop Ecosystem
- Role of Hadoop in Big data– Overview of other Big Data Systems
- Who is using Hadoop
- Hadoop integrations into Exiting Software Products
- Current Scenario in Hadoop Ecosystem
- Installation
- Configuration
- Use Cases ofHadoop (HealthCare, Retail, Telecom)

Chapter 2 : HDFS

- Concepts
- Architecture
- Data Flow (File Read , File Write)
- Fault Tolerance
- Shell Commands
- Data Flow Archives
- Coherency -Data Integrity
- Role of Secondary NameNode

Chapter 3 : Mapreduce

- Theory
- Data Flow (Map – Shuffle – Reduce)
- MapRed vs MapReduce APIs
- Programming [Mapper, Reducer, Combiner, Partitioner]
- Writables
- InputFormat
- Outputformat
- Streaming API using python
- Inherent Failure Handling using Speculative Execution
- Magic of Shuffle Phase
- FileFormats

- Sequence Files

Chapter 4: Hbase

- Introduction to NoSQL
- CAP Theorem
- Classification of NoSQL
- Hbase and RDBMS
- HBASE and HDFS
- Architecture (Read Path, Write Path, Compactions, Splits)
- Installation
- Configuration
- Role of Zookeeper
- HBase Shell Introduction to Filters
- RowKeyDesign -What's New in HBase Hands On

Chapter 5 : Hive

- Architecture
- Installation
- Configuration
- Hive vs RDBMS
- Tables
- DDL
- DML
- UDF
- Partitioning
- Bucketing
- Hive functions
- Date functions
- String functions
- Cast function Meta Store
- Joins
- Real-time HQL will be shared along with database migration project

Chapter 6 : pig

- Architecture
- Installation
- Hive vs Pig
- Pig Latin Syntax
- Data Types
- Functions (Eval, Load/Store, String, DateTime)
- Joins

CREDO SYSTEMZ

- UDFs- Performance
- Troubleshooting
- Commonly Used Functions

Chapter 7 : sqoop

- Architecture , Installation, Commands(Import , Hive-Import, EVal, Hbase Import, Import All tables, Export)
- Connectors to Existing DBs and DW

Chapter 8 : kafka

- Kafka introduction
- Data streaming Introduction
- Producer-consumer-topics
- Brokers
- Partitions
- Unix Streaming via kafka

Chapter 9 : oozie

- Architecture
- Installation
- Workflow
- Coordinator
- Action (Mapreduce, Hive, Pig, Sqoop)
- Introduction to Bundle
- Mail Notifications

CREDO SYSTEMZ

Chapter 10: Hadoop 2.0 and spark

- Limitations in Hadoop
- –HDFS Federation
- High Availability in HDFS
- HDFS Snapshots
- Other Improvements in HDFS2
- Introduction to YARN aka MR2
- Limitations in MR1
- Architecture of YARN
- MapReduce Job Flow in YARN
- Introduction to Stinger Initiative and Tez
- BackWard Compatibility for Hadoop 1.X
- Spark Fundamentals
- RDD- Sample Scala Program- Spark Streaming

Contact Info:

+91 9884412301 | +91 9884312236

Know more about Hadoop

info@credosystemz.com

New # 30, Old # 16A, Third Main Road,
Rajalakshmi Nagar, Velachery, Chennai
(Opp. to MuruganKalyanaMandapam)

BOOK A FREE DEMO

CREDO SYSTEMZ