BIG DATA –HADOOP

1.Overview of Big Data

- Big Data analytics project lifecycle – brief

- Hadoop Overview

2.Environment setup

- Setting up Single node cluster

- Setting up Multi node cluster

- Hadoop Distributed File System – HDFS

3.HDFS Overview

- HDFS Architecture, Operations

- Data Storage in Hadoop – Replication and Block storage

- Hadoop Daemons and its functionality

- Java based approach (Access HDFS through Java)

- Anatomy of a File Read

- Map Reduce

4.History of Map Reduce

- Map Reduce architecture – Programming paradigm

- Mapper and Reducer concepts

- Hadoop Daemons and its functionality

- Data types in Map Reduce

- Input and output formats of Map Reduce

- Different phases of Map Reduce

- Partitioner and Combiner

- Data Localization in Map Reduce

- Examples of Map Reduce program – Log Analysis

- MR Job Configurations

5.Sqoop

- Overview of Sqoop

- Sqoop Installation and configuration

- Importing data from RDBMS to HDFS

- Conditional based import

- Sqoop – to hive import

- Sqoop – Export to RDBMS

- Sqoop – Meta store

- Sqoop – Jobs configure and execution

6.YARN Framework

- Yarn Architecture

- Advantages of YARN over MR engine V1

- MR Job submission and execution procedure

7.Hive

- Introduction to Hive

- Hive Architecture

- RDBMS vs Hive

- Map Reduce Vs Apache Pig Vs Hive

- Hive shell and HiveQL(HQL)

Crush IT Techno
Add: 84 B, Sheetal Nagar, Above Uchchariya Stationary, Service Road of Apollo Premier (Vijay Nagar Square) Indore (MP) 452001
Contact: 9993049572/8827049572
Web: www.crushittechno.com , Email: crushittechno@gmail.com

Page 2

LinkedIn - https://tinyurl.com/crushittechno   Facebook- https://www.facebook.com/crushittechno
Instagram - https://www.instagram.com/crushittechno   Youtube- https://tinyurl.com/crushittechnoEclasses

- Hive server & Beeline client

- Tables Management (External Vs Managed tables)

- Schemas and data types

- Partitions (Static & Dynamic partitioning)

- Bucketing

- Different file formats – rc, orc, parquet, Json, text formats

- Hive – Built in functions

- Hive – User defined functions (UDF)

- HQL – Query execution plan & performance tuning

8.Introduction to NoSQL

- What is NoSQL Databases?

- Fixed schema(RDBMS) Vs Flexible Schema(NoSQL)

- Json (JavaScript Object Notation) format

- Ideas Behind MongoDB / Cassandra

9.HBase

- HBase Introduction

- HBase architecture

- HBase vs RDBMS (fixed Vs flexible schema)

- Master and Region servers

- HBase commands

10.Introduction to Apache Spark

- Introduction to Spark

- Apache Hadoop Vs Apache Spark

Crush IT Techno
Add: 84 B, Sheetal Nagar, Above Uchchariya Stationary, Service
Road of Apollo Premier (Vijay Nagar Square) Indore (MP)
452001
Contact: 9993049572/8827049572
Web: www.crushittechno.com , Email: crushittechno@gmail.com

Page 3

LinkedIn - https://tinyurl.com/crushittechno   Facebook- https://www.facebook.com/crushittechno
Instagram - https://www.instagram.com/crushittechno   Youtube- https://tinyurl.com/crushittechnoEclasses

- Need for Spark

- Sample Programs to understand the advantages of Spark

11.Other Hadoop Ecosystem Components

- Overview of Apache Pig

- Overview of Flume

- Overview of Oozie

- Overview of Apache Kafka

- HADOOP DISTRIBUTIONS

- Overview & need for Cloudera (CDH) distribution

- Overview of Hortonworks (HDP)