ATTENDANCESHEET

												Student Code	
												Course Enrolled	
BigData Hadoop with MongoDB **Total Duration: 56 Hours** **Each Session: 2 Hours**												Name of the Student	
Session 1 **Big Data & Hadoop Overview** · Evolution of Big Data · What is Big Data? · Five Vs. of Big Data. · Where is Big Data?													1
Session 2 **Big Data & Hadoop Overview (contd.)** · Case Study and Big Data Challenges. · Challenges with traditional database? · Why Hadoop to resolve Big Data challenges? · Introduction to Hadoop? · History of Hadoop. · Overview of various tools(or small Project) of Hadoop.													2
Session 3 **Hadoop Architecture- HDFS** · Hadoop Storage and Processing Architecture · Purpose of design of HDFS. · DFS – Distributed File System. · Components of Hadoop. · HDFS architecture.													3
Session 4 **Hadoop Architecture- HDFS (contd.)** · Hadoop Services. · Block in Hadoop. · Rack Awareness. · HDFS Read Mechanism · HDFS Single Block Write Mechanism · HDFS Multi Blocks Write Mechanism · Hands-On Practice · Assignments													4
Session 5 **MAPREDUCE - Processing Architecture** · What is Hadoop MapReduce? · MapReduce in Nutshell. · Two Advantages of MapReduce. · Hadoop MapReduce approach with an Example. · Hadoop MapReduce | YARN Components · YARN with MapReduce.													5
Session 6 **MAPREDUCE - Processing Architecture (contd.)** · YARN Application Workflow. · MR1 vs MR2 · MapReduce Way: Classes and Packages in MapReduce · Explanation of Complete MapReduce Program · Writing MapReduce Program · Hands-On Practices · Assignments													6
Session 7 **SQOOP - with MYSQL** · What is Sqoop? · Why Sqoop so popular? · How to get it working? · Internal Workflow of Sqoop. · Advantages of Sqoop. · Different Connectors of Sqoop.													7

ATTENDANCESHEET

												Student Code		
												Course Enrolled		
												Name of the Student		
Session 8 **SQOOP - with MYSQL (contd.)** · Sqoop Data Imports · Sqoop Data Exports · Sequence File Format. · Free Form Query. · Incremental Import vs Sqoop Job. · Integration with Hadoop Ecosystem · Hands-On Practices · Assignments														8
Session 9 **PIG** · Entry Of Apache Pig. · What is Apache PIG? · Where should I use PIG? · Where not to use PIG? · Pig vs MapReduce.														9
Session 10 **PIG (contd.)** · Twitter case study on Apache Pig. · Apache Pig Architecture. · Pig Components. · Pig Data Model and Operators. · Running Pig commands and Pig Scripts. · Log Analysis. · Hands-On Practices · Assignments														10
Session 11 **HIVE** · What is Hive? · Where to use Hive? · Why go for Hive when PIG is There? · Hive Architecture. · Hive Components. · Hive Background.														11
Session 12 **HIVE (contd.)** · How Facebook uses Hive? · Limitation of Hive. · Abilities of Hive Query Language. · Differences with Traditional RDBMS. · Hive Types. · Hands-On Practices. · Assignments														12
Session 13 **SPARK** · Big Data Analytics. · Batch VS Real Time Analytics. · What is Apache Spark? · Why Apache Spark? · Spark Ecosystem. · Why Spark when there is a Hadoop?														13
Session 14 **SPARK (contd.)** · Spark vs Hadoop. · Apache Spark Features. · Spark SQL. · RDD · Case Study · Hands On Practices · Assignments														14

ATTENDANCESHEET

												Student Code	
												Course Enrolled	
												Name of the Student	
Session 15 **OOZIE and HUE** · Intro to Hue: · What is Hue? · When to Use it? · **Oozie:** · What is Oozie? · Life without it. · When to use it?													15
Session 16 **OOZIE and HUE (contd.)** · Oozie Workflow · Actions · Control flow · Sqoop Action · Hive Action · HDFS Action · Workflows: Hands-On Examples · Use of Coordinators and Bundles · Hands-On Practices · Assignments													16
Session 17 **FLUME** · What is Flume? · Flume Overview. · Flume Vs Sqoop. · Fume Architecture. · Building Blocks of Flume. · Configuring Flume Agent.													17
Session 18 **FLUME (contd.)** · Configuring Channel and Sink. · Flume Commands. · Real Time streaming – Capturing Twitter Data · Hands-On Practices. · Assignments.													18
Session 19 **KAFKA** · Need of Messaging System? · What is Kafka? · Kafka Features. · Kafka Components. · Kafka Architecture. · Configuring Kafka.													19
Session 20 **KAFKA (contd.)** · Kafka Cluster-Producer, Broker, Consumer · Kafka Messaging · Analyzing Twitter Data using Kafka. · Kafka use case. · Hands-On Practices. · Assignments.													20
Session 21 **HBASE** · Challenges with traditional database? · Rise of NoSQL Database. · What is NoSQL Databases? · Types of NoSQL Database. · What is HBase? · When you have HDFS, why HBase ? · When you have Hive, why HBase ? · CAP Theorem and ACID properties.													21

ATTENDANCESHEET

												Student Code		
												Course Enrolled		
												Name of the Student		
Session 22 **HBASE (contd.)** · Why HBase? · HBase vs RDBMS. · Physical view in HBase. · Logical to Physical view. · HBase Architecture. · HBase Components. · HBase table characteristics · HBase Shell and Commands. · Hands-On Practices. · Assignments.														22
Session 23 **MongoDB** · What is Document Database? · Advantages of using documents. · What is MongoDB? · MongoDB Architecture. · Where to use MongoDB? · Advantages of MongoDB. · Datatype in MongoDB. · MongoDB Data Management.														23
Session 24 **MongoDB (contd.)** · MongoDB document storage system. · MongoDB Collections. · Insert Document in collection. · Update/Save Document · Delete document/collection. · Mongo Import/Export. · Query document. · Aggregation,Map-Reduce,Index in MongoDB. · Text Search & Lookup in MongoDB.														24
Session 25 **Project- Retail Data Analysis using various Tools.** · Load data to MySQL · Sqoop Import Data in AVRO format to HDFS														25
Session 26 **Project- Retail Data Analysis using various Tools.** · Load data into hive · Partitioning and bucketing in Hive														26
Session 27 **Project- Retail Data Analysis using various Tools.** · Load data into pig from Hive using HCatalog · Store results into Hive from Pig using HCatalog														27
Session 28 **Project- Retail Data Analysis using various Tools.** · Load avro data into pig · Load data from Hive into Spark using SparkSQL · Conclusion														28