Clustrex
Software Training

—

BIG DATA

Big Data table of contents:

	Array Inbuilt Function with examples. Real time Array Examples
	String Built in functions with examples. Scenario based Logical string examples.
	Regular Expression and Generic Classes
	Inner Classes Abstract Types Compound Types Self-type Implicit Parameters and Implicit Conversions
Spark	**Spark Introduction** A Quick Comparison of the Five Best Big Data Frameworks. Comparing Scala, java, Python and R in Apache Spark Introduction to Spark. Spark Architecture and spark components. Spark Installation with different OS(EX: Windows,Linux etc.)
	HDFS HDFS installations in Single node cluster HDFS basic commands.
	RDD operations and actions Resilient Distributed dataset Parallelized Collections External Datasets RDD Operations Basics Passing Functions to Spark Working with Key-Value Pairs Actions
	Transformation and RDD Schema Specification

	Introduction to Dataframe and Dataset. Transformations with examples Interoperating with RDDs Inferring the Schema Using Reflection Programmatically Specifying the Schema
	Sample Spark project and spark submit using Eclipse or Intellij Maven Installation and path setting Eclipse integration with spark cluster Types of Deploy modes. Create the Spark simple project. Types of Spark-submits
	Generic Load/Save Functions Manually Specifying Options Run SQL on files directly Save Modes Saving to Persistent Tables Bucketing, Sorting and Partitioning
	Different file formats(Read/Write) ORC Parquet Text CSV AVRO Tab Sequence file Object file
	Aggregation and Sorting Aggregation Sorting Sample project with external files.
Hive	**Introduction to Hive** Describe what Hive is, what it's used for and how it compares to other similar technologies Describe the Hive architecture Describe the main components of Hive

	List interesting ways others are using Hive
	Hive DDL Create databases and tables in Hive, while using a variety of different Data Types Run a variety of different DDL commands Use Partitioning to improve performance of Hive queries Create Managed and External tables in Hive
	Hive DML Load data into Hive Export data out of Hive Run a variety of different Hive QL DML queries
	Hive Operators and Functions Use a variety of Hive Operators in your queries Utilize Hive's Built-in Functions Explain ways to extend Hive functionality
Sqoop	**Introduction to Scoop** Sqoop Installations and Basics Importing Data from RDBMS to HDFS Advance Imports Real Time Use Case Exporting Data from HDFS to RDBMS Parallelism Importing data from RDBMS to Hive Running Sqoop in Oozie
HBase	**HBase** HBase Architecture, Data Flow, and Use cases How to Download & Install Hbase HBase Shell and General Commands Create, Insert, Read Tables in HBase HBase: Limitations, Advantage & Problems HBase Troubleshooting HBase Vs Hive