

HADOOP ESSENTIALS (15 hours)

<p>Introduction to Big Data (1 hours)</p> <p>Characteristics of Big Data Challenges for Big Data Popular tools used to store, process, analyze & visualize big data Use Cases for Big Data</p>	<p>Hadoop Eco-system & Architecture (2 hours)</p> <p>Hadoop's key characteristics Eco-system & Core Components Where Hadoop fits? Traditional versus Hadoop's data analytics architecture When to use & not use Hadoop? Apache Hadoop & distributions Hadoop Job Trends</p>	<p>HDFS Architecture (3 hours)</p> <p>Introduction to Hadoop Distributed File System HDFS Architecture and Features Files and data blocks Anatomy of a File Read/ Write on HDFS Replication & Rack Awareness</p>
<p>YARN Architecture(3 hours)</p> <p>ARYARN Architecture Classic vs. YARN YARN daemons Speculative Execution HDFS Federation Authentication & High Availability</p>	<p>Hadoop Setup (6 hours)</p> <p>Hadoop Deployment Modes Setting up a Pseudo-distributed cluster Horton works sandbox configuration Linux Terminal Commands Configuration Parameters and Values HDFS File System Operations</p> <p>Hadoop services using Ambari HDFS, MapReduce and YARN Parameters Hadoop WebUI Filesystem & Linux commands</p>	

HADOOP PROGRAMMING & ANALYTICS (40 hours)

<p>MapReduce Basics (3 hours)</p> <p>MapReduce Framework, Architecture & Use Cases Input Splits Hands on with MapReduce Programming Packaging MapReduce Jobs in a JAR</p>	<p>MapReduce Advanced (6 hours)</p> <p>Setting Mapper & Reducer counts Combiners Partitioners & Custom Partitioners Input & Output formats in MapReduce Sequence Files & Compressions Distributed Cache</p>	<p>Hadoop Streaming using Python (4 hours)</p> <p>Hadoop Streaming concepts Hadoop streaming using Python Writing Python scripts for streaming Testing Python scripts Executing YARN jar on Python script</p>
<p>Using Pig (6 hours)</p> <p>Background of Pig Pig architecture Pig Latin basics Pig execution modes Pig processing – loading and transforming data Pig built-in functions Filtering, Grouping, Sorting data Relational Join operators Pig User Defined Functions</p>	<p>Using Hive (6 hours)</p> <p>Background of Hive Hive architecture Warehouse directory & Metastore Hive Query Language Managed & external tables concept Data processing – loading data into tables Using Hive built-in functions Using Joins in Hive Partitioning data using Hive – static & dynamic Bucketing in Hive</p>	<p>Working with HBase (6 hours)</p> <p>HBase Overview HBase Data Model Row Oriented v/s Column Oriented Storage HBase Architecture HBase shell commands Bulk Load data into HBase</p>
<p>Sqoop and Flume – Data Ingestion (6 hours)</p> <p>Setup MySQL RDBMS & Sqoop Sqoop Connectors, Commands Sqoop Options-file Importing data – to HDFS, Hive Exporting data to MySQL Data ingestion using Flume Why Flume? Flume Architecture Ingesting Twitter data into HDFS using Flume</p>	<p>Using Oozie (3 hours)</p> <p>The Oozie DAG architecture Setting up database & Oozie configuration Creating workflows with Oozie Submitting Oozie jobs & monitoring it</p>	

JOB READINESS (20 hours)

Project Work:

Hands on project and use cases from the Retail, Telecom, Healthcare, Airlines, Banking and Finance domains.

Placement Assistance:

Resume Building, Interview Preparation, Mock Interviews with Industry Experts.