

Sno	Topic	Sub-Topic
1	Understanding Big Data and Hadoop	Big Data, Limitations and Solutions of existing Data Analytics Architecture, Hadoop, Hadoop Features, Hadoop Ecosystem, Hadoop 2.x core components, Hadoop Storage: HDFS, Hadoop Processing: MapReduce Framework, Hadoop Different Distributions
2	Hadoop Architecture and HDFS CMDS	Hadoop 2.x Cluster Architecture - Federation and High Availability, A Typical Production Hadoop Cluster, Hadoop Cluster Modes, Common Hadoop Shell Commands, Hadoop 2.x Configuration Files, Single node cluster
3	MapReduce Framework	MapReduce Use Cases, Traditional way Vs MapReduce way, Why MapReduce, Hadoop 2.x MapReduce Architecture, Hadoop 2.x MapReduce Components, YARN MR Application Execution Flow, YARN Workflow, Anatomy of MapReduce Program, Demo on MapReduce. Input Splits, Relation between Input Splits and HDFS Blocks, MapReduce: Combiner & Partitioner
4	Advanced MapReduce	Counters, Distributed Cache, MRunit, Reduce Join, Custom Input Format, Sequence Input Format, Xml file Parsing using MapReduce
5	Pig	About Pig, MapReduce Vs Pig, Pig Use Cases, Programming Structure in Pig, Pig Running Modes, Pig components, Pig Execution, Pig Latin Program, Data Models in Pig, Pig Data Types, Shell and Utility Commands, Pig Latin : Relational Operators, File Loaders, Group Operator, COGROUP Operator, Joins and COGROUP, Union, Diagnostic Operators, Specialized joins in Pig, Built In Functions (Eval Function, Load and Store Functions, Math function, String Function, Date Function, Pig UDF, Piggybank

6	Hive/Impala	Hive Background, Hive Use Case, About Hive, Hive Vs Pig, Hive Architecture and Components, Metastore in Hive, Hive : Thrift Server, Limitations of Hive, Comparison with Traditional Database, Hive Data Types and Data Models, Partitions and Buckets, Hive Tables(Managed Tables and External Tables), Importing Data, Querying Data, Managing Outputs, Hive Script, Hive QL: Joining Tables, Hive Indexes and views Hive query optimizers, User Defined Functions, Hive UDF
7	Advanced Hive with HBase	HBase: Introduction to NoSQL Databases and HBase, HBase v/s RDBMS, HBase Components, HBase Architecture, Run Modes & Configuration, Hive with Hbase examples
8	Advanced Hbase with Phoenix	HBase Data Model, HBase Shell, HBase Client API, Data Loading Techniques, Demos on Bulk Loading, Getting and Inserting Data, Filters in HBase.
9	Oozie, Flume and Sqoop	Sqoop Demo, Imports and exports , Oozie, Oozie Components, Oozie Workflow, Scheduling with Oozie, Flume components , flow and practical example
10	Hadoop Administration	Cluster Management tools and metadata management Ambari, Cloudera Manager and Atlas
11	Hadoop on Cloud Services	Hadoop on Amazon web Services

Distributions : Amazon EMR, Hortonworks and Cloudera

Real Time Project Explanation with Integration of eco-systems

Hadoop on Amazon web Services

Raj

Amazon solution architect

Duration: 2 Weekends

Weekend Batch: Jun-3rd and 4th and Jun-10th and 11th

