

Real Time Projects on Big Data Using Cloudera (Complete Hands on Practicles)

Hadoop

- ❖ **HDFS**
 - process start & stop commands
 - data transfer commands
 - how to define custom properties
- ❖ **MapReduce**
 - Processing **DATABASE** data
 - Processing **XML** data
 - Processing **PDF** data
 - Processing **JSON** data
 - Processing **IMAGE** data
- ❖ **YARN**
 - Importance of YARN
 - How to do resource scheduling
 - How to observe real time jobs
- ❖ **Real Time Production Cluster**
 - How to define the Real Time Production Cluster Size
 - How to secure the Real Time Production Cluster
 - How to Balance the Real Time Production Cluster
 - How to Transfer the data between Real Time Production Clusters
- ❖ **Performace Improvements**
 - How to improve the performance using Java Code
 - How to improve the performance using proerties
- ❖ **Real Time Usecases on Hadoop**
- ❖ **Interview Questions and Answers**

Hive

- ❖ **Data Model of Hive**
 - Tables
 - Partitions
 - Buckets
- ❖ **Practical Examples on**
 - UDF's
 - SerDe
 - Different Storage options (**AVRO, RC, ORC, PARQUET, TEXT, SEQUENCE**)
 - Different Data Types (**XML, PDF, JSON**)
- ❖ **Hive Integrations**
 - Hive Impala Integrations
 - Hive Hbase Integrations
 - Hive Hcatlog Integrations
 - Hive Phoenix Integrations
 - Hive MongoDB Integrations

- o **Hive Spark Integrations**
- ❖ **Real Time Usecases on Hive**
- ❖ **Interview Questions and Answers**

Pig

- ❖ **Practical Examples on**
 - o UDF's
 - o Different Storage options (**AVRO, ORC, PARQUET, TEXT**)
 - o Different Data Types (**XML, PDF, JSON**)
- ❖ **Pig Integrations**
 - o Pig Hbase Integrations
 - o Pig Hcatlog Integrations
 - o Pig Phoenix Integrations
 - o Pig MongoDB Integrations
 - o **Pig Spark Integrations**
- ❖ **Real Time Usecases on Pig**
- ❖ **Interview Questions and Answers**

HCatalog

- ❖ **HCatalog Integrations**
 - o HCatalog Hive Integrations
 - o HCatalog Pig Integrations
- ❖ **Real Time Usecases on HCatalog**
- ❖ **Interview Questions and Answers**

Impala

- ❖ **Impala vs Hive**
- ❖ **Real Time Usecases on Impala**
- ❖ **Interview Questions and Answers**

HBase

- ❖ **Practical Examples on**
 - o CRUD operations
 - Command Based Solutions
 - Java Based Solutions
 - o Hbase with MapReduce Programs
 - o **Hbase with Spark Programs**
- ❖ **Hbase Integrations**
 - o Hbase Hive Integrations
 - o Hbase Pig Integrations
 - o Hbase Phoenix Integrations
 - o **Hbase Spark Integrations**
- ❖ **Real Time Usecases on Hbase**
- ❖ **Interview Questions and Answers**

Cassandra

- ❖ **Practical Examples on**
 - CRUD operations
 - Command Based Solutions
 - Java Based Solutions
 - Cassandra with MapReduce Programs
 - **Cassandra with Spark Programs**
- ❖ **Cassandra Integrations**
 - **Cassandra Spark Integrations**
- ❖ **Real Time Usecases on Cassandra**
- ❖ **Interview Questions and Answers**

Phoenix

- ❖ **Practical Examples on**
 - CRUD operations
 - Command Based Solutions
 - Java Based Solutions
 - Phoenix with MapReduce Programs
 - **Phoenix with Spark Programs**
- ❖ **Phoenix Integrations**
 - Phoenix Hbase Integrations
 - Phoenix Hive Integrations
 - Phoenix Pig Integrations
 - **Phoenix Spark Integrations**
 - Phoenix Flume Integrations
 - Phoenix Kafka Integrations
- ❖ **Real Time Usecases on Phoenix**
- ❖ **Interview Questions and Answers**

Sqoop

- ❖ **Practical Examples on**
 - Sqoop Import Operations
 - Sqoop Export Operations
- ❖ **Sqoop Integrations**
 - Sqoop HDFS Integrations
 - Sqoop Hive Integrations
 - Sqoop Hbase Integrations
- ❖ **Real Time Usecases on Sqoop**
- ❖ **Interview Questions and Answers**

Flume

- ❖ **Practical Examples on**
 - Sources
 - Channels
 - Sinks
- ❖ **Flume Integrations**

- o Flume Hadoop Integrations
- o Flume Hive Integrations
- o Flume Pig Integrations
- o Flume Hbase Integrations
- o Flume Phoenix Integrations
- o Flume Kafka Integrations
- o **Flume Spark Integrations**
- ❖ **Real Time Usecases on Flume**
- ❖ **Interview Questions and Answers**

Kafka

- ❖ **Practical Examples on**
 - o Producers
 - o Consumers
 - o Brokers
- ❖ **Kafka Integrations**
 - o Kafka Flume Integrations
 - o **Kafka Spark Integrations**
- ❖ **Real Time Usecases on Kafka**
- ❖ **Interview Questions and Answers**

Oozie

- ❖ **Practical Examples on**
 - o Workflows
 - o Schedulers
- ❖ **Oozie Integrations**
 - o Oozie Hdfs Integrations
 - o Oozie MapReduce Integrations
 - o Oozie Hive Integrations
 - o Oozie Pig Integrations
 - o Oozie Hbase Integrations
 - o Oozie Sqoop Integrations
 - o Oozie Flume Integrations
 - o **Oozie Spark Integrations**
- ❖ **Real Time Usecases on Oozie**
- ❖ **Interview Questions and Answers**

Zeppelin

- ❖ **Practical Examples on**
 - o notebooks
- ❖ **Zeppelin Integrations**
 - o Zeppelin Hive Integrations
 - o Zeppelin Phoenix Integrations
 - o Zeppelin RDBMS Integrations
 - o Zeppelin Cassandra Integrations
 - o **Zeppelin Spark Integrations**

- ❖ Real Time Usecases on Zeppelin
- ❖ Interview Questions and Answers

Spark

- ❖ Practical Examples on
 - o Spark Core
 - o Spark SQL
- ❖ Spark Integrations
 - o Spark Hive Integrations
- ❖ Real Time Usecases on Spark
- ❖ Interview Questions and Answers

Day to Day Activity Tools

- ❖ Building Tools
 - o Maven
 - o Sbt
- ❖ Source Repository Tools
 - o SVN
 - o GitHub
- ❖ File Transfer Tools
 - o File Zilla
 - o WinScp
 - o SSH
- ❖ Virtualization Softwares
 - o VMWare
 - o VirtualBox
- ❖ Linux Commands

End to End Projects

- ❖ Complete flow of the Big Data Projects using
 - o Hadoop
 - o Hive, Pig
 - o Hbase, Phoenix, Cassandra
 - o Flume, Kafka
 - o Sqoop
 - o Oozie
 - o Spark
- ❖ Complete flow of the Big Data Usecases using
 - o Hadoop
 - o Hive, Pig
 - o Hbase, Phoenix, Cassandra
 - o Flume, Kafka
 - o Sqoop
 - o Oozie
 - o Spark