**What You Will Learn:**

Set up the environment in Linux for Hadoop projects using Cloudera Hadoop Distribution CDH 5

Run a MapReduce job

Store data with Apache Hive, and Apache HBase

Index data in HDFS with Apache Solr

Develop a Kafka messaging system

Stream Logs to HDFS with Apache Flume

Transfer data from MySQL database to Hive, HDFS, and HBase with Sqoop

Create a Hive table over Apache Solr

Develop a Mahout User Recommender System

Discover Spark Streaming application development and best practices

Work with the low-level details of discretized streams

Optimize production-grade deployments of Spark Streaming via configuration recipes and instrumentation using Graphite, collectd, and Nagios

Ingest data from disparate sources including MQTT, Flume, Kafka, Twitter, and a custom HTTP receiver

Integrate and couple with HBase, Cassandra, and Redis

Take advantage of design patterns for side-effects and maintaining state across the Spark Streaming micro-batch model

Implement real-time and scalable ETL using data frames, SparkSQL, Hive, and SparkR

Use streaming machine learning, predictive analytics, and recommendations

Mesh batch processing with stream processing via the Lambda architecture

Install Apache Kudu, Impala, and Spark to modernize enterprise data warehouse and business intelligence environments, complete with real-world, easy-to-follow examples, and practical advice

Integrate HBase, Solr, Oracle, SQL Server, MySQL, Flume, Kafka, HDFS, and Amazon S3 with Apache Kudu, Impala, and Spark

Use StreamSets, Talend, Pentaho, and CDAP for real-time and batch data ingestion and processing

Utilize Trifacta, Alteryx, and Datameer for data wrangling and interactive data processing

Turbocharge Spark with Alluxio, a distributed in-memory storage platform

Deploy big data in the cloud using Cloudera Director

Perform real-time data visualization and time series analysis using Zoomdata, Apache Kudu, Impala, and Spark

Understand enterprise big data topics such as big data governance, metadata management, data lineage, impact analysis, and policy enforcement, and how to use Cloudera Navigator to perform common data governance tasks

Implement big data use cases such as big data warehousing, data warehouse optimization, Internet of Things, real-time data ingestion and analytics, complex event processing, and scalable predictive modeling

Study real-world big data case studies from innovative companies, including Navistar, Cerner, British Telecom, Shopzilla, Thomson Reuters, and Mastercard

Leverage Phoenix as an ANSI SQL engine built on top of the highly distributed and scalable NoSQL framework HBase. Learn the basics and best practices that are being adopted in Phoenix to enable a high write and read throughput in a big data space

Handle a petabyte data store by applying familiar SQL techniques

Store, analyze, and manipulate data in a NoSQL Hadoop echo system with HBase

Apply best practices while working with a scalable data store on Hadoop and HBase

Integrate popular frameworks (Apache Spark, Pig, Flume) to simplify big data analysis

Demonstrate real-time use cases and big data modeling techniques

Store big data

Configure big data

Process big data

Schedule processes

Move data among SQL and NoSQL systems

Monitor data

Perform big data analytics

Report on big data processes and projects

Test big data systems

Explains importance of security, auditing and encryption within a Hadoop installation

Describes how the leading players have incorporated these features within their Hadoop distributions and provided extensions

Demonstrates how to set up and use these features to your benefit and make your Hadoop installation secure without impacting performance or ease of use

Kafka, Storm, Spark, Solr, Zookeeper, NiFi, HBase, HDFS, HIVE, YARN, Ranger, Knox, Ambari and Kerberos