

By **Mr. Gopal Krishna, Sr. BigData Architect**, 15+ Years Of Real Time Exp, 9+ Years exclusive exp on **Hadoop & Spark**. **CCA 175 - Spark and Hadoop Developer Certified Consultant**

Introduction to BIGDATA and HADOOP

- What is Big Data?
- What is Hadoop?
- Relation between Big Data and Hadoop.
- What is the need of going ahead with Hadoop?
- Scenarios to apt Hadoop Technology in REAL TIME Projects

- Challenges with Big Data
 - Storage
 - Processing
- How Hadoop is addressing Big Data Changes

- Comparison with Other Technologies
 - RDBMS
 - Data Warehouse
 - TeraData
- Different Components of Hadoop Echo System
 - Storage Components
 - Processing Components

- Importance of Hadoop Echo System Components
- Other solutions of Big Data
 - Introduction to NO SQL

- Batch Vs Real Time Vs Near Real Time(NRT) Processing
 - Examples of Batch Processing Systems
 - Examples on Real Time Analytics Systems
 - Examples on Near Real Time Systems

SCALA (SCALable LAguage)

- **Introduction to Scala**
 - Why Scala
 - Why Scala is a multi paradigm language?
 - Scala Vs Java
 - Scala Vs Python
 - Interoperability between Scala and Java
 - Scala Data types
 - Scala Packages
 - Scala REPL (Read Evaluate Print Loop)
 -
- **Scala Basics**
 - Variable Declarations
 - Variable Type Inference
 - Interactive Scala – Scala Shell
 - Writing Scala Scripts – Compiling the Scala Programs
 - Defining Functions in Scala

By **Mr. Gopal Krishna, Sr. BigData Architect**, 15+ Years Of Real Time Exp, 9+ Years exclusive exp on **Hadoop & Spark. CCA 175 - Spark and Hadoop Developer Certified Consultant**

- Type casting in Scala
- Different IDEs for Scala

- **Scala Control Structures**
 - If expressions
 - If -Else expressions
 - While Loops
 - Do- While Loops
 - For loop
 - Diff types of for loop
 - For loop with Range
 - For loop with Collection
 - For loop with Filter
 - For loop with Yield.
 - Pattern Matching in Scala
 - Exception Handling in Scala
- How to pass run time arguments in Scala

- **Functional Programing in Scala**
 - What is Functional Programming
 - Difference between Object Oriented and Functional Programing Paradigm
 - Closures in Scala
 - Anonymous Functions in Scala
 - Currying Functions
 - Higher Order Functions
 - Collections in Scala
 - Lists
 - Sets
 - Maps
 - Mutable & Immutable Collections

- **Object Oriented Programing in Scala (Traits & OOPS)**
- Traits Introduction
 - When to use traits in Scala
 - Creating traits basic oops
 - Classes and Objects Basics
 - Pattern Matching in Scala
 - Exception Handling in Scala
 - Constructors in Scala

- **Scala Environment Set Up**
 - Scala set up on Windows
 - ✓ Java Set Up
 - ✓ Scala Set Up
 - Scala set up on Linux
 - ✓ Java Set Up
 - ✓ Scala Set Up

By **Mr. Gopal Krishna, Sr. BigData Architect**, 15+ Years Of Real Time Exp, 9+ Years exclusive exp on **Hadoop & Spark**. **CCA 175 - Spark and Hadoop Developer Certified Consultant**

SPARK [An IN_MEMORY Cluster Computing System]

- **Introduction to Spark**
 - Motivation for Spark
 - Spark Vs Map Reduce Processing
 - Advantages of IN_MEMORY Processing over DISK Based
 - Where to use Spark
 - ROI Comparison of Hadoop Processing over Spark Processing
 - Why Spark Processing is faster than Map Reduce?

- **Architecture of Spark**
 - Comparison between Hadoop & Spark Architectures
 - Spark Master
 - Spark Driver
 - Spark Worker Node
 - Spark Runtime Managers
 - Standalone
 - YARN
 - Apache Mesos
 - How to Start Spark Deamons

- **Spark Basics**
 - Spark Shell Introduction – Standalone Mode
 - Creating Spark Context
 - Creating Spark Conf, Spark Session
 - File Operations in Spark Shell
 - Caching in Spark
 - Real time Examples of Spark
 -

- **Simple Build Tool (SBT)**
 - IDEA IntelliJ IDE Introduction
 - Adding SCALA Plug In to IntelliJ
 - Installing SBT
 - Spark Project creation and building with SBT
 - Running a Spark Project with SBT
 - Verifying Spark Jobs in Spark Web UI
 - Spark-submit - How to deploy Spark Applications with spark-submit command
 - Running Spark project in Clustered Mode

- **Resilient Distributed Dataset (RDD)**
 - What is RDD and why it is important in Spark
 - **RDD Key Features**
 - Immutable
 - Lazily Evaluated
 - Partitioned
 - Cacheable

By **Mr. Gopal Krishna, Sr. BigData Architect, 15+ Years Of Real Time Exp, 9+ Years exclusive exp on Hadoop & Spark. CCA 175 - Spark and Hadoop Developer Certified Consultant**

- How to create a RDD
- Different types of RDDs
- **RDD Operations**
 - Transformations
 - Actions
- Different Transformations in RDD
- Different Actions in RDD
- Loading Data through RDD
- Saving Data
- Key-Value pair RDD
- **Loading and Saving Data - through different File Formats**
 - Text,csv,tsv,Object files
 - As a Hadoop file
- **Key-Value Pair RDD operations**
- Spark Storage Persistence Levels
- Running Spark in a Clustered Mode
- Deploying Application with spark-submit
- Cluster Management
- **Accumulators**
 - Introduction to Accumulators
 - Practical applicability of accumulators
 - Real time examples on Accumulators
- **Broadcast variables**
 - Introduction to Broadcast variables
 - Practical applicability of Broadcast variables
 - Real time examples on Broadcast variables
 -
- **Spark Processing - with different Programming Languages**
 - **Scala**
 - Installing Scala
 - How to use “**spark-shell**”
 - Examples on Spark with Scala
 - **Python**
 - Installing Python
 - How to use “**pyspark**”
 - Examples on Spark with Python
 - **R**
 - Installing R
 - How to use “**SparkR**”
 - Examples on Spark with R Language
- **Spark SQL**
 - Introduction to Spark SQL
 - The SQL Context
 - Hive Vs Spark SQL
 - Spark SQL support for Text Files, Parquet and JSON files

By **Mr. Gopal Krishna, Sr. BigData Architect, 15+ Years Of Real Time Exp, 9+ Years exclusive exp on Hadoop & Spark. CCA 175 - Spark and Hadoop Developer Certified Consultant**

- Data Frames
- Data Sets
- Data frames vs Data sets – Performance Optimization
- Real Time examples of Spark SQL
- **Different File Formats Support in SparkSQL**
 - Text -JSON -CSV -ORC -TSV -Parquet
- **Different Integration with SparkSQL**
 - SparkSQL integration with Hive
 - SparkSQL integration with RDBMS
 - SparkSQL integration with NOSQL(Cassandra)

➤ Spark Streaming

- Introduction to Spark Streaming
- Architecture of Spark Streaming
- RDD vs Discretized Streams(DStreams)
- DStream Operations
- Introduction to SparkStreamingContext(SSC)
- Transformations on DStreams
 - Window Operations
 - Transform Operations
- Spark Streaming Vs Flume
- Introduction to Kafka
- Spark Streaming Integration with Kafka Overview
- Real Time examples of Spark Streaming & Kafka

➤ Spark MLlib

- Introduction to Machine Learning
- Vector Class in MLlib
- Spark MLlib Algorithms introduction
- Classification and Regression Algorithms
- Naïve Bayes Classification Algorithm
- Decision Trees Algorithm Overview

Apache Kafka

- Introduction to Apache Kafka
- Architecture of Kafka
- Real time examples on Kafka usage in enterprise level applications
- Installation of Apache Kafka
- Fail Over Mechanism in Kafka
- Practical Use Cases on Kafka

Apache Storm

- Introduction to Apache Storm
- Flume vs Kafka vs Storm
- Real time examples on Apache Storm usage in enterprise level applications
- Practical Examples on Apache Storm

By **Mr. Gopal Krishna, Sr. BigData Architect, 15+ Years Of Real Time Exp, 9+ Years exclusive exp on Hadoop & Spark. CCA 175 - Spark and Hadoop Developer Certified Consultant**

Offerings from Kelly

- Complete Guidance for **CCA 175 - Spark and Hadoop Certification**
- Providing 2 End to End **Real Time Projects On Spark** with Hadoop Integration
- Good Number of **Proof Of Concepts(POCs)** on Spark
- Proof Of Concepts (POCs) - End to End Execution and demonstration by POC Groups.
- Mock Interviews will be conducted on a one-to-one basis after the course duration.
- Soft Copy Materials for all the Components.
- Detailed Assistance in **RESUME Preparation** with **Real Time Projects** based on your technical back ground.
- Guidance in Resume preparation on a **one - on - one basis**
- All the Real time interview questions will be provided.
- Discussing the new happenings in Spark
- Discussing the Interview Questions in Spark and relevant technologies on a daily basis