

## Big Data by Infosys

Course Objectives	Course Outcomes
<ul style="list-style-type: none"><li>• Big Data Fundamentals and Technology Landscape</li></ul>	<ul style="list-style-type: none"><li>• Implement Big Data Solutions Using Hadoop</li></ul>
<ul style="list-style-type: none"><li>• Master Advanced Hive Techniques</li></ul>	<ul style="list-style-type: none"><li>• Execute Advanced Data Management with Hive</li></ul>
<ul style="list-style-type: none"><li>• Learn Data Ingestion with Sqoop</li></ul>	<ul style="list-style-type: none"><li>• Perform Data Migration Using Sqoop</li></ul>
<ul style="list-style-type: none"><li>• Gain Proficiency in Apache Spark Core</li></ul>	<ul style="list-style-type: none"><li>• Develop and Optimize Spark Core Applications</li></ul>
<ul style="list-style-type: none"><li>• Develop Skills in Spark SQL</li></ul>	<ul style="list-style-type: none"><li>• Build and Query Data Frames with Spark SQL</li></ul>

**Course Duration:** 45 Hours

### PRE-REQUISITES:

Learners who undergo this course would need to understand the following pre-requisites to be able to appreciate and undergo the contents:

- Introduction to Apache Hive
- Scala Programming

### Course Curriculum

#### UNIT I

Why Big Data, Introduction to Big Data technology landscape, Hadoop Demos, More about Big Data

#### UNIT II

Advanced Hive Concepts – Partitions and Buckets, Complex Data types, Hive functions, Hive File formats

#### UNIT III

Data ingestion using Sqoop – Need for Sqoop, Sqoop architecture, Sqoop Data flow model, Data migration using Sqoop

#### UNIT IV

Introduction to Spark, Spark Core – Spark Core architecture, Working with RDDs, RDD operations, Paired RDDs, Spark Core Joins, Spark Persistence

#### UNIT V

Spark SQL – Need for Spark SQL, Overview, Architecture, Workflow, Spark SQL DataFrame and DataSet, Working with different file formats, Spark SQL on Hive

## **INDUSTRY USE CASE:**

Test Projects:

**Use Cases with Infosys Springboard link provided. (Student LMS Access)**

1. [Hive functions - Demo](#)
2. [Hands on try out - Advanced Hive](#)
3. [Hands on try out - Spark Core](#)
4. [DataFrame - Demos](#)
5. [Hands on try out - Spark SQL](#)
6. [Unix Demos](#)
7. [HDFS Demos](#)
8. [MapReduce Demos](#)
9. [Hive Demos](#)