Big Data Hadoop and Spark Developer

simplilearn

Course Introduction

Course Objectives

By the end of this course, you will be able to:

- Learn how to navigate the Hadoop Ecosystem and understand how to optimize its use
- Ingest data using Sqoop, Flume, and Kafka
- Implement partitioning, bucketing, and indexing in Hive
- Work with RDD in Apache Spark
- Process real-time streaming data
- Perform DataFrame operations in Spark using SQL queries
- Implement User-Defined Functions (UDF) and User-Defined Attribute Functions (UDAF) in Spark

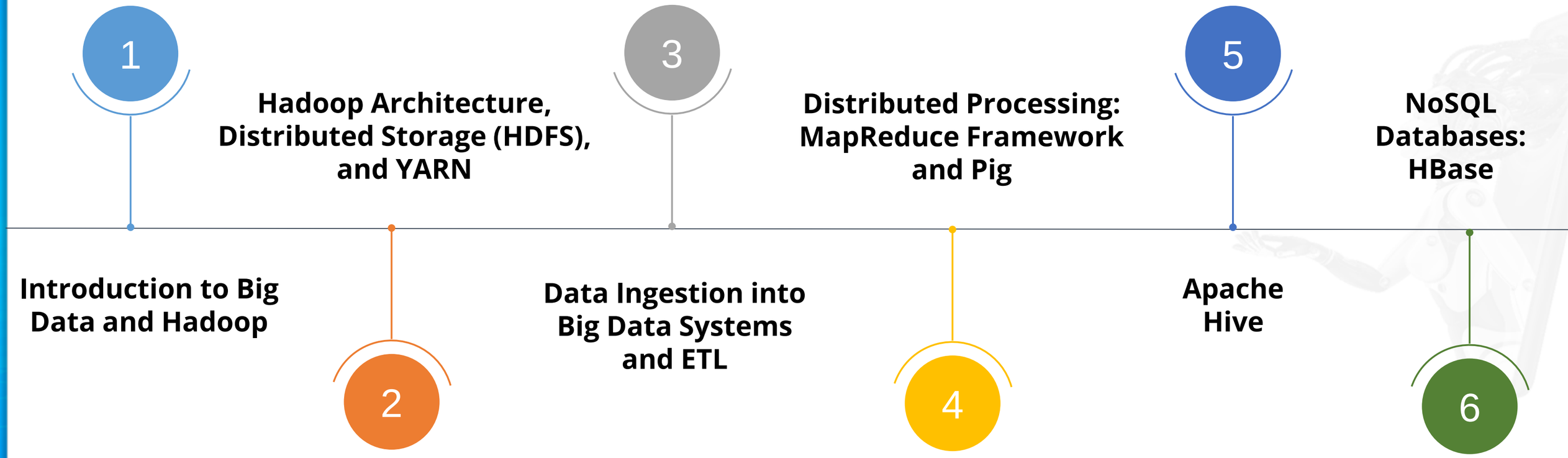
Course Prerequisites

The course requires prior knowledge of the following technologies:

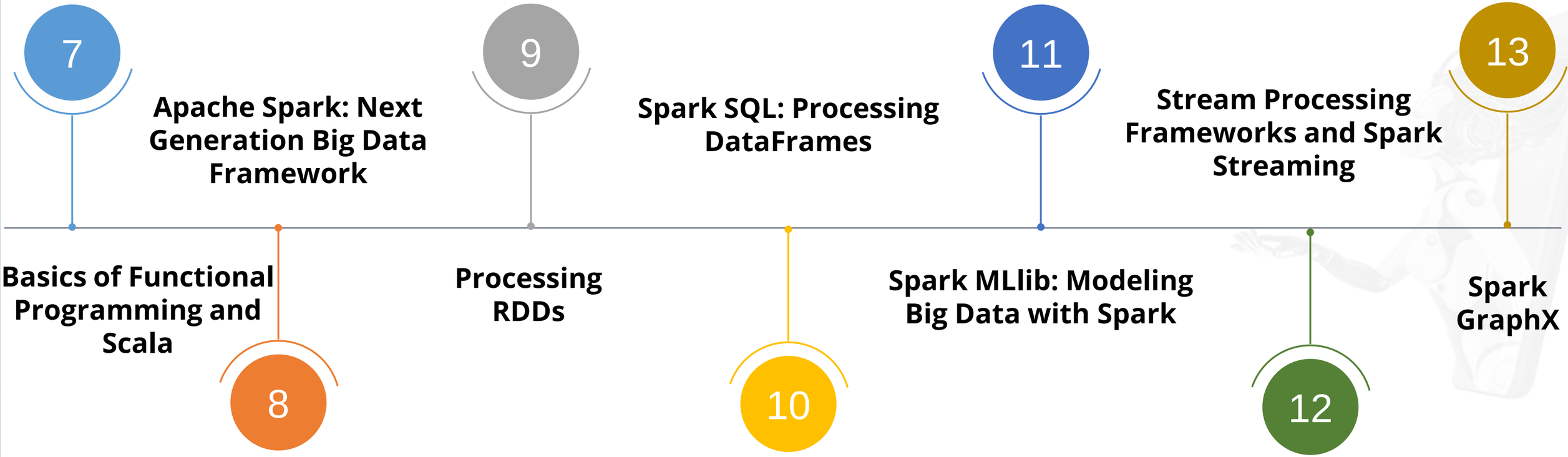
Core Java

SQL

Course Outline



Course Outline



Skills Covered

1. HDFS
1. MapReduce
1. Flume
1. Kafka
1. Hive
1. HBase

Project Highlights

Use Hadoop features to predict patterns and share actionable insights for a car insurance company.

NYSE

Use Hive features for data engineering and analysis of New York stock exchange data.

glassdoor

Perform sentiment analysis on employee review data gathered from Google, Netflix, and Facebook.

amazon

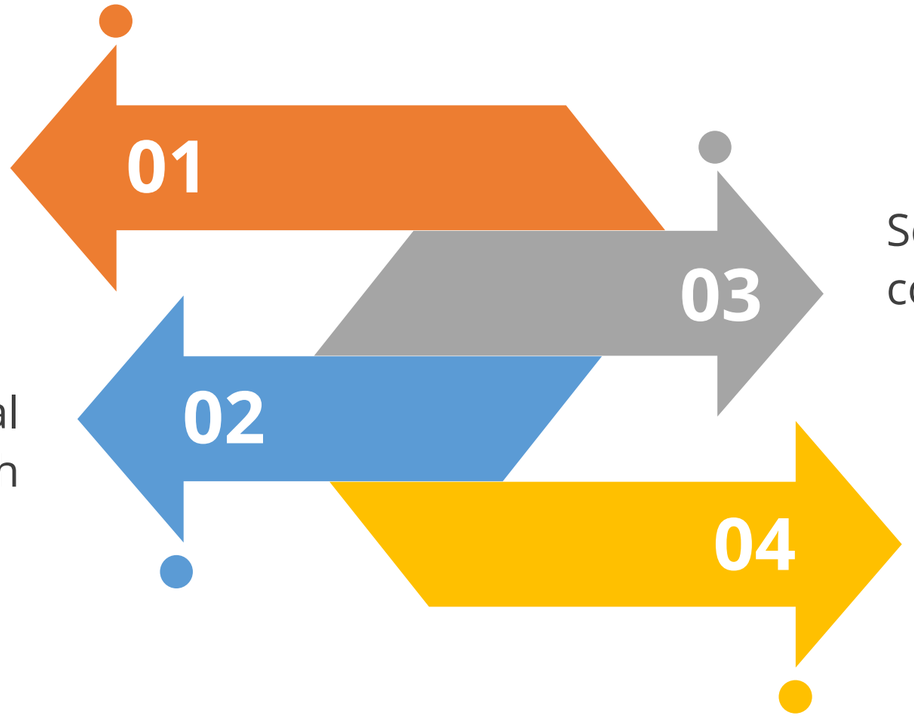
Perform product and customer segmentation to increase the sales of Amazon.

Course Completion Criteria

Complete 85% of the Online Self-Learning (OSL) course

OR

Attend all the Live Virtual Classes (LVC) in a batch



Score a minimum of 75% in course-end assessment

Submit at least one course-end project

DATA AND
ARTIFICIAL INTELLIGENCE