INTRODUCTION TO APACHE KUDU
Take your knowledge to the next level

Cloudera’s Introduction to Apache Kudu training teaches students the basics of Apache Kudu, a data storage system for the Hadoop platform that is optimized for analytical queries. The course covers common Kudu use cases and Kudu architecture. Students will learn how to create, manage, and query Kudu tables, and to develop Spark applications that use Kudu.

Get hands-on experience
Through instructor-led discussion, as well as hands-on exercises, participants will learn topics including:

- A high-level explanation of Kudu
- How does it compares to other relevant storage systems and which use cases would be best implemented with Kudu
- Learn about Kudu’s architecture as well as how to design tables that will store data for optimum performance.
- Learn data management techniques on how to insert, update, or delete records from Kudu tables using Impala, as well as bulk loading methods
- Finally, develop Apache Spark applications with Apache Kudu

What to expect
This material is intended for a broad audience of students involved with either software development or data analysis. This would include software developers, data engineers, DBAs, data scientists, and data analysts.

Students should know SQL. Familiarity with Impala is preferred but not required. Students should also know how to develop Apache Spark applications using either Python or Scala. Basic Linux experience is expected.
Course Contents

1. Introduction

2. Overview and Architecture
   _ What Is Kudu?
   _ Why Use Kudu?
   _ Kudu Use Cases
   _ Architecture Overview
   _ Kudu Tools
   _ Essential Points

3. Apache Kudu Tables
   _ Kudu Tables
   _ Data Storage Options
   _ Designing Schemas
   _ Partitioning Tables for Best Performance
   _ Using Kudu Tools with Tables
   _ Essential Points

4. Using Apache Kudu with Apache Impala
   _ Apache Impala Overview
   _ Creating and Querying Tables
   _ Deleting Tables
   _ Loading and Modifying Data in Kudu Tables
   _ Defining Partitioning Strategy
   _ Essential Points

5. Developing Apache Spark Applications with Apache Kudu
   _ Apache Spark and Apache Kudu
   _ Kudu, Spark SQL, and DataFrames
   _ Managing Kudu Table Data with Scala
   _ Creating Kudu Tables with Scala
   _ Essential Points

6. Conclusion