
INTRODUCTION TO APACHE KUDU

Take your knowledge to the next level

////////////////////////////////////
“The professionalism and expansive technical knowledge demonstrated by our instructor were incredible. The quality of the Cloudera training was on par with a university.”

General Dynamics
////////////////////////////////////

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 compare 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