Cloudera - Introduction to Apache Kudu

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

Who Can Benefit

- 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.

Prerequisites

- 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 Details

Overview and Architecture

- What Is Kudu?
- Why Use Kudu?
- Kudu Use Cases
- Architecture Overview
- Kudu Tools
- Essential Points

Apache Kudu Tables

- Kudu Tables
- Data Storage Options
- Designing Schemas
- Partitioning Tables for Best Performance
- Using Kudu Tools with Tables
Essential Points