CLOUDERA

Cloudera Training for Apache HBase

Course Overview

Course Type
Instructor-led training course

Level
Intermediate

Duration
3 days

Platform
CDH, HDP, and CDP

Topics Covered

- Apache HBase
- Apache ZooKeeper
- Apache Hive
- Apache Impala
- Apache Spark

About This Training

Our three-day *Cloudera Training for Apache HBase* course empowers participants to use this operational database technology for mission-critical applications capable of storing, locating, and retrieving data well beyond the scale of traditional database management systems.

Students begin by learning about the HBase architecture and typical use cases. Through lecture and more than a dozen hands-on exercises that follow, they build on this foundational knowledge by learning how to design HBase tables and schemas, as well as interact with HBase through both its command-line shell and Java APIs. The course continues beyond the basics to cover performance tuning and cluster management techniques that will help to ensure your services and applications run reliably and efficiently.

What Skills You Will Gain

During this course, you will learn how to:

- Determine whether HBase is appropriate for a given use case
- Employ best practices for schema and row key design
- Create, populate, alter, and remove HBase tables
- Add, locate, retrieve, update, and delete data
- Perform common cluster administration tasks
- Identify and resolve performance bottlenecks

Who Should Take This Course?

This course is appropriate for developers and administrators who intend to use HBase. Prior experience with databases and data modeling is helpful, but not required. Knowledge of Java is assumed. Prior knowledge of the Cloudera platform is not strictly required, but *Cloudera Developer Training for Apache Spark and Hadoop* provides an excellent foundation for this course.

Other Training That Might Interest You

- *Cloudera Developer Training for Apache Spark and Hadoop*
- *Administrator Training: CDP Private Cloud Base*
- *Cloudera Data Analyst Training*
- *Cloudera Operational Database Fundamentals* *(OnDemand library)*

Cloudera Training for Apache HBase

Training Outline (Page 2 of 3)

Introduction to Hadoop and HBase

- Introducing Hadoop
- Core Hadoop Components
- Exercise: Using HDFS
- What is HBase?
- Strengths of HBase
- HBase in Production
- Weaknesses of HBase

HBase Tables

- HBase Concepts
- HBase Table Fundamentals
- Thinking About Table Design
- Exercise: HBase Data Import

HBase Shell

- Creating Tables with the HBase Shell
- Working with Tables
- Exercise: Using the HBase Shell
- Working with Table Data
- Exercise: Data Access in the HBase Shell

HBase Architecture Fundamentals

- HBase Regions
- HBase Cluster Architecture
- HBase and HDFS Data Locality

HBase Schema Design

- General Design Considerations
- Application-Centric Design
- Designing HBase Row Keys
- Other HBase Table Features
- Exercise: Using `MIN_VERSIONS` and Time-To-Live

HBase Schema Design (*continued*)

- General Design Considerations
- Application-Centric Design
- Designing HBase Row Keys

Basic Data Access with the HBase API

- Options to Access HBase Data
- Creating and Deleting HBase Tables
- Retrieving Data with Get
- Retrieving Data with Scan
- Inserting and Updating Data
- Deleting Data
- Exercise: Using the Developer API

More Advanced HBase API Features

- Filtering Scans
- Exercise: HBase Filters
- Client-Side Write Buffer
- Exercise: Using Client-Side Write Buffer
- Best Practices
- HBase Coprocessors
- Exercise: Using Atomic Counters

HBase Write Path

- HBase Write Path
- Exercise: Exploring HBase
- Compaction
- Splits
- Exercise: Flushes and Compactions

HBase Read Path

- How HBase Reads Data
- Block Caches for Reading

[CR:181123c|DS:108|TWW]

Cloudera Training for Apache HBase

Training Outline (Page 3 of 3)

HBase Performance Tuning

- Column Family Considerations
- Schema Design Considerations
- Configuring for Caching
- Memory Considerations
- Dealing with Time Series and Sequential Data
- Pre-Splitting Regions
- Exercise: Detecting Hot Spots

HBase Administration and Cluster Management

- HBase Schema Design
- General Design Considerations
- Application-Centric Design
- Designing HBase Row Keys

HBase Replication and Backup

- HBase Replication
- HBase Backup
- MapReduce and HBase Clusters
- Exercise: Administration

Using Hive and Impala with HBase

- How to use Hive and Impala to Access HBase
- Exercise: Hive and HBase

Appendix A: Accessing Data with Python and Thrift

- Thrift Usage
- Working with Tables
- Getting and Putting Data
- Scanning Data
- Deleting Data
- Counters
- Filters
- Optional Exercise: Using Python and Thrift with HBase

Appendix B: OpenTSDB

Appendix C: `hbase-spark` API

- Introduction
- Architecture and Integration Patterns
- Typing and API Usage
- Future Work
- Optional Exercise: Using `hbase-spark` API

[CR:181123c|DS:108|TWW]