# Cloudera Developer Training for Apache Hadoop

## Duration: 4 Days  Course Code: GK3902

### Overview:

HBase is an open-source, non-relational, distributed database that provides a fault-tolerant, scalable way to store massive quantities of data. In this course, Hadoop developers and administrators will gain the skills needed to install and maintain HBase and develop client code.

### Target Audience:

Developers familiar with Apache Hadoop

### Objectives:

- HBase architecture
- 
- Use the HBase shell to directly manipulate HBase tables
- 
- Design optimal HBase schemas for efficient data storage and recovery
- 
- Connect to HBase using the Java API
- 
- Bulk-load data into HBase using MapReduce
- 
- Administer a HBase cluster
- 
- Resolve performance bottlenecks

### Prerequisites:

- Familiarity with Hadoop's architecture and APIs
- Experience writing basic applications
- Prior programming experience, preferably Java
- Experience with databases and data modeling is helpful, but it is not required

### Testing and Certification

**This course is part of the following programs or tracks:**

- CCSHB: Cloudera Certified Specialist in Apache HBase (CDH4)

## Content:

Data Model

- Tables, Row Keys, and Column Families
- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery
- Data Storage and Bloom Filters
- Modifying Rows
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots
- Administrative Actions Using the Java API
- Accessing Data Using the Java API
- Bulk-Load into HBase
- Required Zookeeper Configurations
- Required Configuration Settings
- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase Operations
- Tuning for Client Operations
- Logging Locations and Troubleshooting Tools
- Flushes and Compactions
- View and Read HFiles
- Detect Hot Spots
- Write an HBase Program
- HBase Input MapReduce
- Bulk Import into HBase
- Install Distributed HBase

- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery
- Data Storage and Bloom Filters
- Modifying Rows
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots
- Administrative Actions Using the Java API
- Accessing Data Using the Java API
- Bulk-Load into HBase
- Required Zookeeper Configurations
- Required Configuration Settings
- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase Operations
- Tuning for Client Operations
- Logging Locations and Troubleshooting Tools
- Flushes and Compactions
- View and Read HFiles
- Detect Hot Spots
- Write an HBase Program
- HBase Input MapReduce
- Bulk Import into HBase
- Install Distributed HBase

HBase API

- Connecting to HBase Using the Java API

- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery
- Data Storage and Bloom Filters
- Modifying Rows
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots
- Administrative Actions Using the Java API
- Accessing Data Using the Java API
- Bulk-Load into HBase
- Required Zookeeper Configurations
- Required Configuration Settings
- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase Operations
- Tuning for Client Operations
- Logging Locations and Troubleshooting Tools
- Flushes and Compactions
- View and Read HFiles
- Detect Hot Spots
- Write an HBase Program
- HBase Input MapReduce
- Bulk Import into HBase
- Install Distributed HBase

- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery
- Data Storage and Bloom Filters
- Modifying Rows
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots
- Administrative Actions Using the Java API
- Accessing Data Using the Java API
- Bulk-Load into HBase
- Required Zookeeper Configurations
- Required Configuration Settings
- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase Operations
- Tuning for Client Operations
- Logging Locations and Troubleshooting Tools
- Flushes and Compactions
- View and Read HFiles
- Detect Hot Spots
- Write an HBase Program
- HBase Input MapReduce
- Bulk Import into HBase
- Install Distributed HBase

- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery
- Data Storage and Bloom Filters
- Modifying Rows
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots
- Administrative Actions Using the Java API
- Accessing Data Using the Java API
- Bulk-Load into HBase
- Required Zookeeper Configurations
- Required Configuration Settings
- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase Operations
- Tuning for Client Operations
- Logging Locations and Troubleshooting Tools
- Flushes and Compactions
- View and Read HFiles
- Detect Hot Spots
- Write an HBase Program
- HBase Input MapReduce
- Bulk Import into HBase
- Install Distributed HBase

Labs

- Use the HBase Shell

HBase Shell

- Creating and Manipulating Data Using the Command-Line Shell

Cluster Architecture

- HMaster, RegionServers, and Zookeeper

- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery
- Data Storage and Bloom Filters
- Modifying Rows
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots
- Administrative Actions Using the Java API
- Accessing Data Using the Java API
- Bulk-Load into HBase
- Required Zookeeper Configurations
- Required Configuration Settings
- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase Operations
- Tuning for Client Operations
- Logging Locations and Troubleshooting Tools
- Flushes and Compactions
- View and Read HFiles
- Detect Hot Spots
- Write an HBase Program
- HBase Input MapReduce
- Bulk Import into HBase
- Install Distributed HBase

- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery
- Data Storage and Bloom Filters
- Modifying Rows
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots
- Administrative Actions Using the Java API
- Accessing Data Using the Java API
- Bulk-Load into HBase
- Required Zookeeper Configurations
- Required Configuration Settings
- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase Operations
- Tuning for Client Operations
- Logging Locations and Troubleshooting Tools
- Flushes and Compactions
- View and Read HFiles
- Detect Hot Spots
- Write an HBase Program
- HBase Input MapReduce
- Bulk Import into HBase
- Install Distributed HBase

- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery
- Data Storage and Bloom Filters
- Modifying Rows
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots
- Administrative Actions Using the Java API
- Accessing Data Using the Java API
- Bulk-Load into HBase
- Required Zookeeper Configurations
- Required Configuration Settings
- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase Operations
- Tuning for Client Operations
- Logging Locations and Troubleshooting Tools
- Flushes and Compactions
- View and Read HFiles
- Detect Hot Spots
- Write an HBase Program
- HBase Input MapReduce

HBase Input MapReduce

- Bulk Import into HBase
- Install Distributed HBase

MapReduce and Bulk Loads

- MapReduce Integration

- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery
- Data Storage and Bloom Filters
- Modifying Rows
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots
- Administrative Actions Using the Java API
- Accessing Data Using the Java API
- Bulk-Load into HBase
- Required Zookeeper Configurations
- Required Configuration Settings
- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase Operations
- Tuning for Client Operations
- Logging Locations and Troubleshooting Tools
- Flushes and Compactions
- View and Read HFiles
- Detect Hot Spots
- Write an HBase Program
- HBase Input MapReduce
- Bulk Import into HBase
- Install Distributed HBase

HBase Configuration

- Standalone and Distributed Run Modes

- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery
- Data Storage and Bloom Filters
- Modifying Rows
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots
- Administrative Actions Using the Java API
- Accessing Data Using the Java API
- Bulk-Load into HBase
- Required Zookeeper Configurations
- Required Configuration Settings
- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase Operations
- Tuning for Client Operations

- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery
- Data Storage and Bloom Filters
- Modifying Rows
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots
- Administrative Actions Using the Java API
- Accessing Data Using the Java API
- Bulk-Load into HBase
- Required Zookeeper Configurations
- Required Configuration Settings
- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase Operations
- Tuning for Client Operations
- Logging Locations and Troubleshooting Tools
- Flushes and Compactions
- View and Read HFiles
- Detect Hot Spots
- Write an HBase Program
- HBase Input MapReduce
- Bulk Import into HBase
- Install Distributed HBase

- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery

- Bulk Import into HBase
- Install Distributed HBase

Storage Architecture

- Client Caching

- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery
- Data Storage and Bloom Filters
- Modifying Rows
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots
- Administrative Actions Using the Java API
- Accessing Data Using the Java API
- Bulk-Load into HBase
- Required Zookeeper Configurations
- Required Configuration Settings
- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase Operations
- Tuning for Client Operations
- Logging Locations and Troubleshooting Tools
- Flushes and Compactions
- View and Read HFiles
- Detect Hot Spots
- Write an HBase Program
- HBase Input MapReduce
- Bulk Import into HBase
- Install Distributed HBase

- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery
- Data Storage and Bloom Filters
- Modifying Rows
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots
- Administrative Actions Using the Java API
- Accessing Data Using the Java API
- Bulk-Load into HBase
- Required Zookeeper Configurations
- Required Configuration Settings
- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase Operations
- Tuning for Client Operations

HBase Administration

- Monitoring HBase Processes

- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery
- Data Storage and Bloom Filters
- Modifying Rows
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots
- Administrative Actions Using the Java API
- Accessing Data Using the Java API
- Bulk-Load into HBase
- Required Zookeeper Configurations
- Required Configuration Settings
- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase

- Logging Locations and Troubleshooting Tools
- Flushes and Compactions
- View and Read HFiles
- Detect Hot Spots
- Write an HBase Program
- HBase Input MapReduce
- Bulk Import into HBase
- Install Distributed HBase

- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery
- Data Storage and Bloom Filters
- Modifying Rows
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots
- Administrative Actions Using the Java API
- Accessing Data Using the Java API
- Bulk-Load into HBase
- Required Zookeeper Configurations
- Required Configuration Settings
- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase Operations
- Tuning for Client Operations
- Logging Locations and Troubleshooting Tools
- Flushes and Compactions
- View and Read HFiles
- Detect Hot Spots
- Write an HBase Program
- HBase Input MapReduce
- Bulk Import into HBase
- Install Distributed HBase

- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery
- Data Storage and Bloom Filters
- Modifying Rows
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots
- Administrative Actions Using the Java API
- Accessing Data Using the Java API
- Bulk-Load into HBase
- Required Zookeeper Configurations
- Required Configuration Settings
- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase Operations
- Tuning for Client Operations
- Logging Locations and Troubleshooting Tools
- Flushes and Compactions
- View and Read HFiles
- Detect Hot Spots
- Write an HBase Program
- HBase Input MapReduce
- Bulk Import into HBase
- Install Distributed HBase

Schema Design

- Creating Column Families
- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery
- Data Storage and Bloom Filters
- Modifying Rows
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots
- Administrative Actions Using the Java API
- Accessing Data Using the Java API
- Bulk-Load into HBase
- Required Zookeeper Configurations
- Required Configuration Settings
- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase Operations
- Tuning for Client Operations
- Logging Locations and Troubleshooting Tools
- Flushes and Compactions
- View and Read HFiles
- Detect Hot Spots
- Write an HBase Program
- HBase Input MapReduce
- Bulk Import into HBase
- Install Distributed HBase

- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery
- Data Storage and Bloom Filters
- Modifying Rows
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots
- Administrative Actions Using the Java API
- Accessing Data Using the Java API
- Bulk-Load into HBase
- Required Zookeeper Configurations
- Required Configuration Settings
- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase Operations
- Tuning for Client Operations
- Logging Locations and Troubleshooting Tools
- Flushes and Compactions
- View and Read HFiles
- Detect Hot Spots
- Write an HBase Program
- HBase Input MapReduce
- Bulk Import into HBase
- Install Distributed HBase

Performance Tuning

- Preventing Network Bandwidth Bottlenecks
- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery
- Data Storage and Bloom Filters
- Modifying Rows
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots
- Administrative Actions Using the Java API
- Accessing Data Using the Java API
- Bulk-Load into HBase
- Required Zookeeper Configurations
- Required Configuration Settings
- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase Operations
- Tuning for Client Operations
- Logging Locations and Troubleshooting Tools
- Flushes and Compactions
- View and Read HFiles
- Detect Hot Spots
- Write an HBase Program
- HBase Input MapReduce
- Bulk Import into HBase
- Install Distributed HBase

- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery
- Data Storage and Bloom Filters
- Modifying Rows
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots
- Administrative Actions Using the Java API
- Accessing Data Using the Java API
- Bulk-Load into HBase
- Required Zookeeper Configurations
- Required Configuration Settings
- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase

- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase Operations
- Tuning for Client Operations
- Logging Locations and Troubleshooting Tools
- Flushes and Compactions
- View and Read HFiles
- Detect Hot Spots
- Write an HBase Program
- HBase Input MapReduce
- Bulk Import into HBase
- Install Distributed HBase

- Operations
- Tuning for Client Operations
- Logging Locations and Troubleshooting Tools
- Flushes and Compactions
- View and Read HFiles
- Detect Hot Spots
- Write an HBase Program
- HBase Input MapReduce
- Bulk Import into HBase
- Install Distributed HBase

- Choosing Column Attributes
- Version and HBase Operations
- Compactions in HBase
- Crash Recovery
- Data Storage and Bloom Filters
- Modifying Rows
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots
- Administrative Actions Using the Java API
- Accessing Data Using the Java API
- Bulk-Load into HBase
- Required Zookeeper Configurations
- Required Configuration Settings
- Performing HBase Backups
- Planning for HBase Capacity
- Java Garbage Collection and HBase Operations
- Tuning for Client Operations
- Logging Locations and Troubleshooting Tools
- Flushes and Compactions
- View and Read HFiles
- Detect Hot Spots
- Write an HBase Program
- HBase Input MapReduce
- Bulk Import into HBase
- Install Distributed HBase

Pre-Split Regions

## Further Information:

For More information, or to book your course, please call us on Head Office 01189 123456 / Northern Office 0113 242 5931

info@globalknowledge.co.uk

www.globalknowledge.co.uk

Global Knowledge, Mulberry Business Park, Fishponds Road, Wokingham Berkshire RG41 2GY UK