

[ADM-203]: HDP Operations: HBase Management

Length : 4 Days
 Delivery Method : Instructor-led (Classroom)

Course Overview

This course is designed for administrators who will be installing, configuring and managing HBase clusters. It covers installation with Ambari, configuration, security and troubleshooting HBase implementations. The course includes an end-of-course project in which participants work together to design and implement an HBase schema.

Audience Profile

Architects, software developers, and analysts responsible for implementing non-SQL databases in order to handle sparse datasets commonly found in big data use cases.

Pre-Requisites

Participants must have basic familiarity with data management systems. Familiarity with Hadoop or databases is helpful but not required. Participants new to Hadoop are encouraged to take the **HDP Overview: Apache Hadoop Essentials** course.

Formats

Lecture/Discussion	50%
Hands-on Labs	50%

Course Outline

Module 1: An Apache HBase Overview and Installing HBase

Lessons

- List Data Science Use Cases
- Describe the Characteristics and Operation of HDFS
- Describe the Responsibilities of the NameNode and DataNode
- Describe the Purpose of YARN, including the:
 - ResourceManager
 - NodeManager
 - ApplicationMaster
- Describe the Primary Differences Between Hadoop 1.x and 2.x
- Describe the Function and Purpose of HBase
- List HBase Features and Components
- Describe an HBase Table as a Set of Key-Value Mappings
- Identify HBase as Either a Row-or- Column-Oriented Database
- Describe HBase Operations

AVANTUS TRAINING PTE LTD

80 Jurong East Street 21 #04-04 Devan Nair Institute Singapore 609607
 Main Line: +65 6661 0888 | Fax: +65 6661 0886
 Email: enquiries@AvantusTraining.com
www.AvantusTraining.com

- List the Options for HBase Installation
- List the HBase Minimum System Requirements
- Describe the Process for Installing HBase Using Ambari
- Describe the Process for Confirming a Successful Installation

LABS

- Installing and Configuring HBase with Ambari
- Manually Installing an HBase Cluster

Module 2: Using the HBase Shell and Ingest/ImportTSV

Lessons

- Setting Up a Development Environment
- Using HDFS Commands
- Demonstration: Understanding Map Reduce
- Using Mahout for Machine Learning
- Work with Basic HBase Shell Commands
- List the Categories of Shell Commands Including:
 - General
 - Table Management
 - Data Manipulation
 - Surgery Tools
 - Cluster Replication Tools
 - Security Tools
- Work with Cluster Administration Commands
- Describe the Function and Purpose of the Regionserver
- Identify the Purpose of Key-Value Pairs
- Identify the Purpose of Row Keys
- Identify the Purpose of Column Families
- Describe How to Read and Write Data in HBase
- Describe the Flush Process
- Describe the Compaction Process
- Perform a Bulk Ingest Using ImportTSV
- Describe the Function and Purpose of a CopyTable

LABS

- Using HBase Shell Commands
- Ingesting Data with ImportTSV

Module 3: Managing HA Clusters and Log Files, Backup Recovery and Security

Lessons

- List and Describe Python Programming Concepts
- List the Steps Required to Upgrade HBase
- Configure HBase for High Availability
- View Log Files
- Describe the Function and Purpose of HBase Coprocessors

AVANTUS TRAINING PTE LTD

80 Jurong East Street 21 #04-04 Devan Nair Institute Singapore 609607

Main Line: +65 6661 0888 | Fax: +65 6661 0886

Email: enquiries@AvantusTraining.com

www.AvantusTraining.com

- Describe the Function and Purpose of HBase Filters
- Describe the Process for Using Filters for Scans
- Describe the Process for Protecting HBase Data with Backups
- Describe the Function and Benefits of Snapshots in HBase
- Describe the Process for Performing Snapshots in HBase
- Describe the Process for HBase Replication
- Configure HBase Cluster Replication
- Describe the Purpose of HBase Authentication
- Describe the Purpose and Benefits of HBase Authorization Via ACLs
- Describe the Benefits of Ranger and Knox for HBase Security
- Describe the Process Used to Configure Simple Authentication
- Describe the Secure Bulk Load Process

LABS

- Enabling HBase High Availability
- Viewing Log Files
- Configuring and Enabling Snapshots
- Configuring Cluster Replication
- Enabling Authentication and Authorization in HBase Tables

Module 4: Monitoring HBase, Maintenance, Troubleshooting and Class Project

Lessons

- List Important Metrics to Monitor for an HBase Cluster
- Monitor an HBase Cluster Using Ambari
- Describe the Benefits of OpenTSDB as a Tool for Monitoring
- Describe How to Identify a Region Hot Spot
- Design a Row-Key Schema to Avoid Hot Spotting
- Configure an HBase Table Using Pre-Splitting
- Describe the Region Splitting Process
- Describe the Function of the Load Balancer
- Define Region Sizing
- Describe the Process of Manual Splitting and Merging
- Describe the Process of Resolving Regions Overlap Issues
- Use the Zookeeper Command Line Tool to Check Zookeeper Status and State
- Monitor JVM Garbage Collection Metrics on Region servers
- Resolve Startup Errors for Master server and Region servers
- Tune HBase for Better Performance
- Tune HDFS for Better HBase Performance

LABS

- Diagnosing and Resolving Hot Spotting
- Region Splitting
- Monitoring JVM Garbage Collection
- End of Course Lab Project – Designing an HBase Schema

AVANTUS TRAINING PTE LTD

80 Jurong East Street 21 #04-04 Devan Nair Institute Singapore 609607

Main Line: +65 6661 0888 | Fax: +65 6661 0886

Email: enquiries@AvantusTraining.com

www.AvantusTraining.com