

[DEV-303]: HDF Developer: HBase Essentials

Length : 2 Days
 Delivery Method : Instructor-led (Classroom)

Course Overview

This course is designed for big data analysts who want to use the HBase NoSQL database which runs on top of HDFS to provide real-time read/write access to sparse datasets. Topics include HBase architecture, services, installation and schema design.

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 attend the **HDP Overview: Apache Hadoop Essentials** course.

Formats

Lecture/Discussion	50%
Hands-on Labs	50%

Course Outline

Module 1: A Hadoop Primer and HBase Overview

Lessons

- Distinguish between Hadoop and the Hortonworks Data Platform
- Identify that Hadoop is Comprised of Multiple Apache Projects
- Describe How Hadoop Stores Files and Processes Data
- Describe the Hadoop Distributed File System (HDFS)
- Describe the Reason HBase was Created
- List HBase Features
- List the Components of the HBase Architecture
- Describe an HBase Table as a set of Value Mappings
- Identify HBase as Either a Row or Column Oriented Database
- Describe the Features Available in HBase 1.0
- Describe a High-Level View of the Overall HBase Architectural Design
- Discuss HBase Region Design and Implementation
- List HBase Services
- Use HBase Data Operations
- Describe HBase High (HA) Availability Options

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 HBase Operational Commands
- Outline an HBase Query Operation

Labs

- Running a MapReduce Job
- Using HBase
- Importing Tables from MySQL Into HBase
- Working with Apache ZooKeeper
- Examining HBase Configuration Files

Module 2: HBase Command Line Basics, HBase Installation and Configuration and HBase Schema Design

Lessons

- List the Shell Command Line Categories
- Use General Shell Commands
- Use Data Manipulations Shell Commands
- Use Surgery Tools
- Use Cluster Replication Tools
- Describe General Considerations for Installation and Configuration
- Describe HBase Configuration Requirements
- Describe Apache ZooKeeper Configuration Requirements
- Backup HBase Tables and Metadata
- Identify How to Choose an Appropriate Rowkey
- Describe the HBase Data Model
- Discuss Sample Schema Use Cases
- Describe How to Optimize Block Size
- Describe How to Adjust Cache Size
- Use Bloom Filters
- Discuss General Optimization Methods

LABS

- Using HBase Shell Commands
- Performing a Backup and Using Snapshot
- Exporting with Apache Pig and Importing with ImportTsv
- Setting Block Size and Enabling Bloom Filters
- Demonstration: Using Java Data Access Object

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