

Cassandra

1. The Purpose of Cassandra

Understanding What Cassandra is

Exercise: Access the Apache Cassandra website.

Learning What Cassandra is Being Used For.

Exercise: Learn what Cassandra is being used for.

Lab: Open the main course virtual machine

2. Getting Started with the Architecture

Understanding that Cassandra is a Distributed Database.

Exercise: View Cassandra documentation.

Learning What Snitch is for.

Learning What Gossip is for.

Exercise: Check your understanding.

Learning How Data Gets Distributed and Replicated.

Exercise: Access related documentation.

Lab: View the course projects

3. Installing Cassandra

Downloading Cassandra.

Exercise: Locate the tarball.

Ensuring Oracle Java 7 is installed.

Exercise: Check that Oracle Java 7 is installed.

Installing Cassandra.

Exercise: Install Cassandra.

Viewing the Main Configuration File.

Exercise: View the cassandra.yaml file.

Providing Cassandra with Permission to Directories.

Exercise: Provide Cassandra with permission to directories.

Starting Cassandra.

Exercise: Start Cassandra.

Checking Status.

Exercise: Check status.

Accessing the Cassandra system.log File.

Exercise: Access the Cassandra system.log file

4. Communicating with Cassandra

Understanding Ways to Communicate with Cassandra.

Exercise: View CQL commands

Using cqlsh

Exercise: Access cqlsh

Lab: Use cqlsh

5. Creating a Database

Understanding a Cassandra Database

Exercise: View the existing keyspaces

Defining a Keyspace.

Exercise: Define a keyspace

Lab: Create a second database

6. Creating a Table

Exercise: Enter a keyspace

Defining Columns and Data Types

Exercise: Define columns and data types

Defining a Primary Key

Exercise: Define a primary key

Recognizing a Partition Key

Exercise: Recognize a partition key

Specifying a Descending Clustering Order

Exercise: Specify a descending clustering order

Lab: Create a second table

7. Inserting Data

Understanding Ways to Write Data

Using the INSERT INTO command

Exercise: Use the INSERT INTO command

Using the COPY command

Exercise: Use the COPY command

Seeing How Data is Stored in Cassandra

Exercise: See how data is stored in Cassandra

Lab: Insert data

8. Modeling Data

Understanding Data Modeling in Cassandra

Understanding Secondary Indexes

Exercise: Use a WHERE clause

Creating a Secondary Index

Exercise: Create a secondary index

Defining a Composite Partition Key

Exercise: Define a composite partition key.

9. Creating an Application

Understanding Cassandra Drivers

Exercise: View the Cassandra drivers

Exploring the DataStax Java Driver

Exercise: Explore the DataStax Java driver

Setting Up a Development Environment

Exercise: Access a development environment

Creating an Application Page

Exercise: Create an application page

Getting the DataStax Java Driver Classes

Exercise: Get the DataStax Java driver classes

Connecting to a Cassandra Cluster

Exercise: Connect to a Cassandra cluster

Executing a Query

Exercise: Execute a query

Displaying Query Results

Exercise: Display query results

Using an MVC Pattern

Exercise: View an MVC pattern

Lab: Create a second application

10. Updating and Deleting Data

Updating Data

Exercise: Update data

Understanding How Updating Works

Exercise: Get an inside view into how updating works

Deleting Data

Exercise: Delete data

Understanding Tombstones

Exercise: View a tombstone

Using TTLs

Exercise: Create a TTL

Lab: Update and delete data

11. Selecting Hardware

Understanding Hardware Choices

Understanding RAM and CPU

Recommendations

Exercise: Check for RAM

Selecting Storage

Exercise: View available storage on your computer

Deploying in the Cloud

Exercise: View cloud recommendations

12. Adding Nodes to a Cluster

Understanding Cassandra Nodes

Exercise: Set up a second a node

Having a Network Connection

Exercise: Have a network connection

Specifying the IP Address for a Node

Exercise: Specify the IP address for a node

Specifying Seed Nodes

Exercise: Specify seed nodes

Bootstrapping a Node

Exercise: Bootstrap a node

Cleaning Up a Node

Exercise: Clean up a node

Using Cassandra-stress

Exercise: Use Cassandra-stress

Lab: Add a third node

13. Monitoring a Cluster

Understanding Cassandra Monitoring Tools

Using node tool

Exercise: Use node tool

Using JConsole

Exercise: Use JConsole

Learning About OpsCenter

Exercise: Learn about OpsCenter

14. Repairing Nodes

Understanding Repair

Exercise: Modify the replication factor of a key space

Repairing Nodes

Exercise: Repair nodes

Understanding Consistency

Exercise: Specify consistency

Understanding Hinted Handoff

Understanding Read Repair

Exercise: View read repair and hinted handoff settings

Lab: Repair nodes for a key space

15. Removing a Node

Understanding Removing a Node

Decommissioning a Node

Exercise: Decommission a node

Putting a Node Back into Service

Exercise: Put a node back into service

Removing a Dead Node

Exercise: Remove a dead node

Lab: Put a node back into service.

16. Redefining a Cluster for Multiple Data Center.

Redefining for Multiple Data Centers

Changing Snitch Type

Exercise: Change snitch type

Modifying cassandra-rackdc.properties

Exercise: Modify cassandra-rackdc.properties

Changing Replication Strategy

Exercise: Change replication strategy

SPARK

WHY SPARK?

Problems with Traditional Large-Scale Systems

Introducing Spark & Spark Basics

Apache MapReduce vs Apache Spark

Using the Spark Shell

Resilient Distributed Datasets (RDDs)

Functional Programming with Spark

Working with RDDs

RDD OPERATIONS

Key-Value Pair RDDs

Types of RDDs

- ParallelCollectionRDD
- SequenceFileRDD
- DoubleRDD
- PairRDD
- PipedRDD

MapReduce and Pair RDD Operations

Read data from text file

Using HDFS with Spark

SPARK FUNCTIONS

Actions

- Reduce, Collect
- Count, First
- Take, TakeSample
- TakeOrdered
- SaveAsTextFile

Transformations

- Map, Filter
- Flat Map, Map-partitions
- Sample, Union
- Intersection
- Distinct, Cartesian

RDD PARTITIONS AND HDFS DATA LOCALITY

Working with Partitions

Executing Parallel Operations

Caching Overview

Caching and Persistence

Distributed Persistence

Writing Spark Applications

Data Frame

Spark 2.0 data frames

Introduction of Data Frame

Create Data Frame

Data input & output

Gathering data frame information

Using SQL on data frame

Apply filter on data frame

Spark DataFrame Dates And Timestamps

Introduction to Date & timestamps

Working with Dates & timestamps

Spark DataFrame Aggregate Operations

Aggregate and GroupBy concepts

Sorting & Ordering

Spark DataFrame Working With Missing Data

Introduction to Missing Data

Dropping Data

Filling missing data

SPARK APPLICATIONS VS. SPARK SHELL

A Spark Cluster

The Spark Web UI

Parallel Programming with Spark

Creating the SparkContext

Configuring Spark Properties

Building and Running a Spark Application
Logging
Spark, Hadoop, and the Enterprise Data
Center

Spark SQL

Introduction to Spark SQL
Inferring a schema
Applying a schema
Loading and Writing schema
SQL caching and UDF
Spark SQL queries to perform computations

Duration: - 2 Months

Fees: - Rs. 22,000



mapping
minds