Oracle Big Data Fundamentals Ed 2

Duration: 5 Days

What you will learn

In the Oracle Big Data Fundamentals course, you learn about big data, the technologies used in processing big data and Oracle's solution to handle big data. You also learn to use Oracle Big Data Appliance to process big data, and obtain a hands-on experience in using Oracle Big Data Lite VM. You identify how to acquire the raw data from a variety of sources, and learn to use HDFS and Oracle NoSQL Database to store the data. You learn about data integration options available in Oracle Big Data. These include Oracle Big Data Connectors to move data to and from Oracle Database, Oracle Data Integrator and Oracle GoldenGate for Big Data which provide integration and synchronization capabilities for data unification of relational and Hadoop data, and Oracle Big Data SQL, which enables dynamic, integrated access for all of your data big data, whether it is stored in HDFS, NoSQL, or Oracle Database. Finally, you learn how to analyze your big data using Oracle Big Data SQL, Oracle Advance Analytics, and Oracle Big Data Spatial and Graph.

Learn To:

Define Big Data.

Describe Oracle's Integrated Big Data Solution and its components.

Define Cloudera's distribution of Hadoop and its core components and the Hadoop ecosystem.

Use the Hadoop Distributed File System (HDFS).

Acquire big data using the Command Line Interface, Flume, and Oracle NoSQL Database.

Process big data using MapReduce, YARN, Hive, Oracle XQuery for Hadoop, Solr, and Spark.

Integrate big data and warehouse data using Sqoop, Oracle Big Data Connectors, Copy to Hadoop, Oracle Data Integrator, and Oracle GoldenGate for big data, and Oracle Big Data SQL.

Analyze big data using Oracle Big Data SQL, Oracle Big Data Spatial and Graph, and Oracle Advanced Analytics technologies.

Use and manage Oracle Big Data Appliance.

Identify the key features and benefits of Oracle Big Data Cloud Service.

Identify the key features and benefits of Oracle Big Data Cloud Service - Compute Edition.

Benefits To You

You will benefit from this course as you define the term big data and discuss Oracle's Big Data solution and use cases. You learn about Apache Hadoop and its core components: HDFS, YARN, and MapReduce. You will also learn about some of the major projects in the Hadoop ecosystem. You will learn how to acquire data into HDFS and Oracle NoSQL Database by using CLI, Flume, and Kafka. To process the data stored in HDFS, you run MapReduce and Spark jobs.

You also explore a range of analysis options, including Oracle Advanced Analytics (OAA) (comprised of Oracle Data Mining and Oracle R Enterprise), and Oracle Big Data Spatial and Graph.

You will learn about the Oracle Big Data Appliance, Oracle Big Data Cloud Service, and Oracle Big Data Cloud Service - Compute Edition. You will study case scenarios where Oracle Big Data stands as the perfect solution.

Audience
Application Developers
Database Administrators
Database Developers

Related Training

Required Prerequisites

Database Basics and Administration

Suggested Prerequisites
Exposure to Big Data

Course Objectives
Define Big Data

Describe Oracle's Integrated Big Data Solution and its components

Define Cloudera's distribution of Hadoop and its core components and the Hadoop ecosystem

Use the Hadoop Distributed File System (HDFS)

Acquire big data using the Command Line Interface, Flume, and Oracle NoSQL Database

Process big data using MapReduce, YARN, Hive, Oracle XQuery for Hadoop, Solr, and Spark

Integrate big data and warehouse data using Sqoop, Oracle Big Data Connectors, Copy to Hadoop, Oracle Data Integrator, and Oracle GoldenGate for big data, and Oracle Big Data SQL

Analyze big data using Oracle Big Data SQL, Oracle Big Data Spatial and Graph, and Oracle Advanced Analytics technologies

Use and manage Oracle Big Data Appliance

Identify the key features and benefits of Oracle Big Data Cloud Service

Identify the key features and benefits of Oracle Big Data Cloud Service - Compute Edition

Introduction
Questions About You
Course Objectives
Course Road Map
Oracle Big Data Lite (BDLite) Virtual Machine (VM) Home Page
Starting the Oracle BDLite VM and accessing the Practice Files
Reviewing the Available Big Data Documentation, Tutorials, and Other Resources

Introducing Oracle Big Data Strategy
Characteristics of Big Data
Importance of Big Data
Big Data Opportunities: Some Examples
Big Data Challenges
Big Data implementation examples
Oracle strategy for Big Data: combining Big Data Processing Engines: Hadoop / NoSQL / RDBMS

Using Oracle Big Data Lite Virtual Machine and Movieplex Application
Oracle Big Data Lite VM Used in this Course
Oracle Big Data Lite VM Home Page Sections
Reviewing the Deployment Guide
Downloading and installing Oracle VM VirtualBox and its Extension Pack
Downloading and Running 7-zip Files to create Virtual Box Appliance File
Importing the Appliance File
Staring the Big Data Lite VM and Starting and Stopping Services
Introducing the Oracle Movieplex Case Study

Introduction to the Big Data Ecosystem
Computer Clusters and Distributed Computing
Apache Hadoop
Types of Analysis That Use Hadoop
Types of Data Generated
Apache Hadoop Core Components: HDFS, MapReduce (MR1), and YARN (MR2)
Apache Hadoop Ecosystem
Cloudera's Distribution Including Apache Hadoop (CDH)
CDH Architecture and Components

Introduction to the Hadoop Distributed File System
Hadoop Distributed Filesystem (HDFS) Design Principles, Characteristics, and Key Definitions
Sample Hadoop High Availability (HA) Cluster
HDFS Files and Blocks
Active and Standby Daemons (Services) Functions
DataNodes (DN) Daemons Functions
Writing a File to HDFS: Example
Interacting With Data Stored in HDFS: Hue, Hadoop Client, WebHDFS, and HttpFS

Acquire Data using CLI, Fuse, Flume, and Kafka
Reviewing the Command Line Interface (CLI)
Viewing File System Contents Using the CLI
FS Shell Commands
Loading Data Using the CLI

Overview of FuseDFS
What is Flume?
Kafka topics
Additional Resources

Acquire and Access Data Using Oracle NoSQL Database
What is a NoSQL Database
RDBMS Compared to NoSQL
HDFS Compared to NoSQL
Define Oracle NoSQL Database
Oracle NoSQL models: Key-Value and Table
Acquiring and Accessing Data in a NoSQL DB
Accessing the CLIs (Data, Admin, SQL)
Accessing the KVStore

Introduction to MapReduce and YARN Processing Frameworks
MapReduce Framework Features, Benefits, and Jobs
Parallel Processing with MapReduce
Word Count Examples
Data Locality Optimization in Hadoop
Submitting and Monitoring a MapReduce Job
YARN Architecture, Features, and Daemons
YARN Application Workflow
Hadoop Basic Cluster: MapReduce 1 Versus YARN (MR 2)

Resource Management Using Yarn
Job Scheduling in YARN
First In, First Out (FIFO) Scheduler, Capacity Scheduler, and Fair Scheduler
Cloudera Manager Resource Management Features
Static Service Pools
Working with the Fair Scheduler
Cloudera Manager Dynamic Resource Management: Example
Submitting and Monitoring a MapReduce Job Using YARN
Using the YARN application Command

Overview of Apache Spark
Benefits of Using Spark
Spark Architecture
Spark Application Components: Driver, Master, Cluster Manager, and Executors
Running a Spark Application on YARN (yarn-cluster Mode)
Resilient Distributed Dataset (RDD)
Spark Interactive Shells: spark-shell and pyspark
Word Count Example by Using Interactive Scala
Monitoring Spark Jobs Using YARN's ResourceManager Web UI

Overview of Apache Hive
What is Hive?
Use Case: Storing Clickstream Data
Hadoop Architecture
How is Data Stored in HDFS?
Organizing and Describing Data With Hive
Big Data SQL on Top of Hive Data

Defining Tables Over HDFS
Hive Queries

Overview of Cloudera Impala
Overview of Cloudera Impala
Hadoop: Some Data Access/Processing Options
Cloudera Impala
Cloudera Impala: Key Features
Cloudera Impala: Supported Data Formats
Cloudera Impala: Programming Interfaces
How Impala Fits Into the Hadoop Ecosystem
How Impala Works with Hive

Using Oracle XQuery for Hadoop
XML Review
Oracle XQuery for Hadoop (OXH)
OXH Features
OXH Data Flow
Using OXH: Installation, Functions, Adapters, and Configuration Properties
Running an OXH Query
XQuery Transformation and Basic Filtering
Viewing the Completed Query in YARN's ResourceManager

Overview of Solr
Overview of Solr
Apache Solr (Cloudera Search)
Cloudera Search: Key Capabilities
Cloudera Search: Features
Cloudera Search Tasks
Indexing in Cloudera Search
Types of Indexing
The solrctl Command

Integrating Your Big Data
Unifying Data: A Typical Requirement
Comparing Big Data Processing Engines
Introducing Data Unification Options
When To Use These Options?

Batch Loading Options
Apache Sqoop
Oracle Loader for Hadoop
Oracle Copy to Hadoop

Using Oracle SQL Connector for HDFS
Batch and Dynamic Loading: Oracle SQL Connector for HDFS
OSCH Architecture
Using OSCH
Features
Parallelism and Performance
Performance Tuning
Key Benefits

Loading: Choosing a Connector

Using Oracle Data Integrator and Oracle GoldenGate for Big Data
ETL and Synchronization: Oracle Data Integrator
ODI's Declarative Design
ODI Knowledge Modules (KMs)Simpler Physical Design / Shorter Implementation Time
Using ODI with Big Data Heterogeneous Integration with Hadoop Environments
Using ODI Studio
ODI Studio Components: Overview
ODI Studio: Big Data Knowledge Modules
Oracle GoldenGate for Big Data

Using Oracle Big Data SQL
Barriers to Effective Big Data Adoption
Overcoming Big Data Barriers
Oracle Big Data SQL: The Hybrid Solution
Benefits: Virtualizes data access across Oracle Database, Hadoop and NoSQL stores
Using Oracle Big Data SQL
Query Performance Overview
Deployment Options

Using Oracle Big Data Spatial and Graph
Graph and Spatial Analysis: All About Relationships
What is Oracle Big Data Spatial and Graph (BDSG)?
Strategy (supported platforms, etc)
BDSG: Graph Analysis
Oracle BDSG: Spatial Analysis
Multimedia Analytics Framework
Deployment Options for Oracle BDSG
Additional Resources

Using Oracle Advanced Analytics
Oracle Advanced Analytics (OAA)
OAA: Oracle Data Mining
OAA: Oracle R Enterprise

Oracle Big Data Deployment Options
Introduction to the Oracle Big Data Appliance
Running the Oracle BDA Configuration Generation Utility
Oracle BDA Mammoth Software Deployment Bundle
Using the Oracle BDA mammoth Utility
BDA Hardware and Integrated and Optional Software
Administering and Securing the Oracle BDA
Introduction to the Oracle Big Data Cloud Service
Introduction to the Oracle Big Data Cloud Service – Compute Edition