

Oracle Big Data Fundamentals Ed 1 NEW

Duration: 5 Days

What you will learn

In the Oracle Big Data Fundamentals course, learn to use Oracle's Integrated Big Data Solution to acquire, process, integrate and analyze big data.

Learn To:

Define Big Data.

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

Define the Hadoop Ecosystem and Cloudera's Distribution Including Apache Hadoop (CDH).

Use the Hadoop Distributed File System (HDFS) to store, distribute, and replicate data across the nodes in the Hadoop cluster.

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

Use MapReduce and YARN for distributed processing of the data stored in the Hadoop cluster.

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

Integrate big data and warehouse data using Scoop, Oracle Big Data Connectors, Copy to BDA, Oracle Big Data SQL, Oracle Data Integrator, and Oracle GoldenGate.

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

Use and manage Oracle Big Data Appliance.

Secure your data.

Benefits To You

Increase your Big Data technology portfolio by learning to use a wide range of big data acquisition, processing, integration, and analysis techniques. In addition, you learn about Oracle's engineered systems for Big Data, which provide a variety of data integration and analysis capabilities. Analysis options include Oracle Big Data SQL, Oracle Data Mining, Oracle R Enterprise, and Oracle Big Data Discovery.

Benefit from a hands-on, case-study approach while learning about Oracle's Integrated Big Data Solution.

Audience

Application Developers

Database Administrators

Hadoop Programmer

Hadoop/BigData Cluster Administrator

Related Training

Required Prerequisites

Database Basics (optional)

Exposure to Big Data

Basic knowledge of Hadoop

Course Objectives

Examine MapReduce programs and balance MapReduce jobs

Use the Oracle BigDataLite Virtual Machine

Review Oracle s Big Data Management Architecture and Engineered Systems

Define Big Data

Identify Big Data Use Cases

Define the Hadoop ecosystem and its components

Examine MapReduce programs and balance MapReduce jobs

Use Oracle NoSQL Database

Use Oracle XQuery for Hadoop

Install, use, and administer the Oracle Big Data Appliance

Provide data security and enable resource management

Course Topics

Introduction

Lesson Objectives

Questions About You

Course Objectives

Course Road Map

Practice Environment

Connecting to the Course Environment (Oracle Big Data Lite Virtual Machine) Using VNC

Starting the Oracle Big Data Lite Virtual Version Machine 4.01

Introducing the Movieplex

Big Data and the Oracle Information Management System

Big Data Opportunities and Challenges

Oracle Information Management Architecture

Optimizing/Simplifying Architecture with Engineered Systems

Using Oracle Big Data Lite Virtual Machine

Overview of the Big Data product stack

Access methods

Review the Oracle Big Data Virtual Machine Home page

Deep dive into the Oracle case study

Identify the data structures used

Understand the importance of filtering the data

Identify the Hadoop Command Guide URL, and review the fs and version commands that are used in the practice

Introduction to the Big Data Ecosystem

Lesson Objectives

Computer Clusters

Distributed Computing

The Hadoop Ecosystem

Hadoop Core Components

Choosing a Hadoop Distribution and Version

Types of Analysis That Use Hadoop

Cloudera s Distribution Including Apache Hadoop (CDH) Architecture

Introduction to the Hadoop Distributed File System (HDFS)

Lesson Objectives

Hadoop Distributed Filesystem (HDFS)

Acquire Data using CLI, Fuse-DFS, and Flume

Introducing the CLI

Examining Fuse DFS

Using Flume

Using and Administering Oracle NoSQL Database

Define Oracle NoSQL Database

List Benefits

Load data into the DB

Access NoSQL Data

Plan an Oracle NoSQL Database installation and Node configuration

Configure and Deploy a KVStore

Using the GUI Interface (monitoring the KVStore)

Use the NoSQL Database Table Model (both CLI and Java API)

Introduction to MapReduce

Lesson Objectives

MapReduce

Interacting with MapReduce

MapReduce Daemons (Services) update based on YARN

Interacting With MapReduce

Fault Tolerance

MapReduce Examples

Using YARN to Manage Resources

YARN Overview

YARN: Theme 1

YARN: Theme 2

Job Submission in YARN

YARN Features

MapReduce 2.0: Overview

Overview of Apache Hive and Apache Pig

Apache Hive

Apache Pig

Overview of Cloudera Impala, Solr, and Apache Spark

Examining Cloudera Impala

Integrating Hadoop and Oracle

What is Apache Solr (Cloudera Serach)?

Cloudera Search: Key Capabilities, Features, Tasks, Indexes, and Collections

Introduction to Spark

Resilient Distributed Datasets (RDD) and Directed Acyclic Graph (DAG) Execution Engine

Overview of Scala Language

Using Oracle XQuery for Hadoop

Extensible Markup Language (XML)

XML Elements and Attributes

XML Path (XPath) Language: Node Types and Family Relationships

FLWOR Expressions

Oracle XQuery for Hadoop (OXH) Features and Data Flow

OXH Adapters and Configuration Properties

XQuery Transformation and Basic Filtering

Viewing the Completed OXH Job in YARN

Options for Integrating Your Big Data

Apache Sqoop

Oracle Loader for Hadoop (OLH)

Copy To BDA

Oracle SQL Connector for HDFS (OSCH)

Oracle Data Integrator (ODI) and Oracle GoldenGate (OGG)

Using Oracle Big Data SQL

Context: Exadata and Big Data Appliance

What is Big Data SQL?

Configuring Oracle Big Data SQL

Create Oracle Tables over HDFS data

Leverage the Hive Metastore to Access Data in Hadoop

Apply Oracle Database Security Policies Over Data in Hadoop

Combine HDFS and Oracle data for analysis (SQL Pattern Matching)

Using Oracle Advanced Analytics

Oracle Data Mining (ODM)

Oracle R Enterprise (ORE)

Oracle R Advanced Analytics for Hadoop (ORAAH)

Introducing Oracle Big Data Discovery

Discover Complex Data Using Oracle Big Data Discovery

Oracle R Enterprise (ORE) Performing Complex Event Processing

Decision Making Guidelines

Recommendations

Using the Oracle Big Data Appliance (BDA)

- Identify the Hardware and Software Components of Oracle Big Data Appliance
- The Available Oracle BDA Configurations
- Using the Mammoth Utility
- Using Oracle BDA Configuration Generation Utility
- BDA Configurations: Full Rack, Starter Rack, and In-Rack Expansion
- Critical and Noncritical Nodes in an Oracle BDA CDH Cluster
- Oracle Integrated Lights Out Manager (ILOM)

Managing the Oracle Big Data Appliance

- Mammoth Installation Types and Steps
- Monitoring the Oracle BDA
- Oracle BDA Command-Line Interface
- Monitor BDA with Oracle Enterprise Manager (OEM)
- Hadoop Cluster Monitoring
- Using Cloudera Manager
- Using Cloudera Hue to interact with CDH
- Starting and Stopping Oracle BDA

Balancing MapReduce Jobs

- Define the Perfect Balance Feature of Oracle BDA
- Use Perfect Balance to Balance MapReduce Jobs
- Run Job Analyzer as a Stand-alone Utility or With Perfect Balance
- Identify, Locate, and Read the Generated Reports
- Collect additional Metrics with Job Analyzer
- Configure Perfect Balance
- Use chopping (partitioning of values)
- Troubleshoot Jobs Running with Perfect Balance and Use the Perfect Balance Examples

Securing Your Data on the BDA

- Security Levels
- Authentication, Authorization, Auditing, and Encryption
- BDA Secure Installation: Kerberos, Sentry, Oracle Audit Vault, and Encryption
- Strong Authentication With Kerberos
- Cloudera Navigator
- Configure Perfect Balance