



Big Data Management for Administrators

Instructor-Led | Big Data | 2 Day | 10.2.2

Course Overview

Applicable for users of software versions 10.x. Set up a live BDM environment by performing various administrative tasks such as Hadoop integration, configuration of the Distribution package, security mechanism set up, and performance tuning. Learn to integrate the Informatica domain with the Hadoop eco-system leveraging Hadoop's lightning processing capability to churn huge data sets.

Objectives

After successfully completing this course, students should be able to:

- Create Cluster Configuration Object
- Set up security with Kerberos authentication
- Tune the performance of the system
- Monitor, view, and troubleshoot BDM logs

Target Audience

Administrators

Prerequisites

Big Data Management for Developers ([Instructor Led](#))

Agenda

Module 0: Introduction to Big Data Management Administration Course Overview

Module 1: Introduction to Big Data Management Administration

- BDM and Big Data
- BDM Components
 - Client tools
 - Application Services
 - Databases
- Roles and responsibilities of Informatica Administrator and Big Data Administrator
- BDM engines: Blaze, Spark, and Hive

Module 2: BDM 10.2 Installation and Configuration

- Pre-installation checks
- Create Cluster Configuration objects

Module 3: Big Data Management Security

- Security aspects
- Authentication and authorization
- Kerberos authentication
- Set up security using Apache Ranger and Data Masking
- BDM with Sentry enabled
- Set up Kerberos on Informatica Server

Module 4: BDM Performance Tuning

- Hardware and cluster level tuning
- Tune the following:
 - Blaze Performance
 - Spark Performance
 - Data Integration Service
 - Model Repository Service
 - Sqoop performance

Module 5: Monitoring Mappings and Viewing Hadoop Environment Logs

- View Data Integration Service Generated Logs
- View Logs for Blaze, Spark, and Hive Engines
- Monitor Blaze Engine and view logs
- Monitor Spark Engine and view logs