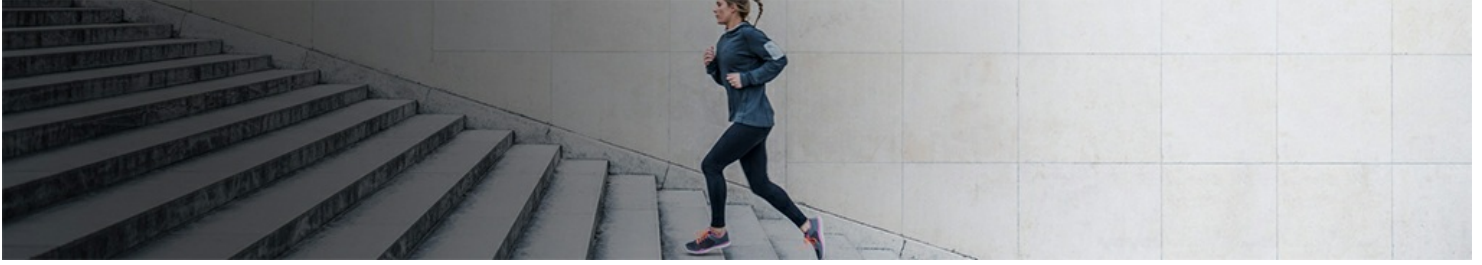


Big Data 10.1 for Developers

now.informatica.com/Big-Data-Management-v101-Developer-ILT.html



Big Data Management v10.1: Developer

[View Course Agenda](#)

Instructor-Led | Big Data | 3 Days

Course Overview

Learn to leverage Informatica Big Data Management for the optimization of data warehousing through the offloading of data processing to Hadoop. Also discover options for the enhancement of a data warehouse through accessing NoSQL databases and complex file parsing.

Objectives

[Enroll Public](#)

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

[Enroll Virtual](#)

- Define “Big Data”
- Identify and prioritize the offloading resource intensive Data Warehouse processes to Hadoop
- Migrate PowerCenter mappings to Big Data Management and ingest data into Hadoop
- Migrate and ingest data into Hadoop using SQOOP and SQL to Mapping
- Describe the Informatica on Hadoop architecture
- Transform data on Hadoop using Informatica polyglot computing
- Differentiate the capabilities of the Informatica engines on Hadoop including Hive MR/Tez, Blaze, and Spark engines
- Leverage the Informatica Smart Executor
- Utilize Informatica and Hadoop monitoring and troubleshooting
- Parse and transform complex data such as JSON, AVRO, and Parquet
- Describe how Informatica parses, reads, and writes NoSQL data collections

Target Audience

- Big Data Management Developers
- Big Data Architects

Prerequisites

- Informatica Developer Tool 10.1 Big Data Management

Agenda

Module 1: Big Data Integration Course Introduction

- Course Agenda
- Accessing the lab environment
- Related Courses

Module 2: Big Data Basics

- What is Big Data?
- Hadoop concepts
- Hadoop Architecture Components
- The Hadoop Distributed File System (HDFS)
- Purposes of a Name Node & Secondary Name Node
- MapReduce
- “Yet Another Resource Manager” (YARN) (MapReduce Version 2)

Module 3: Data Warehouse Offloading

- Challenges with traditional Data Warehousing
- The requirements of optimal Data Warehouse
- The Data Warehouse Offloading Process

Module 4: Ingestion and Offload

- PowerCenter Reuse Reports
- Importin PowerCenter Mappings to Developer
- SQOOP
- SQL to Mapping capability
- Partitioning and parallelism

Module 5: Big Data Management Architecture

- The Big Data world
- Build once, deploy anywhere

Module 7: Mappings, Monitoring, and Troubleshooting

- Configuring and running a mapping in Native and Hadoop environments
- Execution Plans
- Monitor mappings
- Troubleshoot mappings
- Viewing mapping results

Module 8: Hadoop Data Integration Challenges and Performance Tuning

- Describe challenges with executing mappings in Hadoop
- Big Data Management Performance Tuning
- Hive Environment Optimization
 - Mapping Level Tuning
 - DIS Level Tuning
 - Cluster Level Tuning
 - Hadoop environment and cluster tuning
- Tips

Module 9: Data Quality on Hadoop

- The Data Quality process
- Discover insights into your data
- Collaborate and Create Data Improvement Assets
- Modify, Manage, and Monitor Data Quality
- Self Service Data Quality
- Executing Data Quality mappings on Hadoop

Module 10: Complex File Parsing

- The Complex file reader
- The Data Processor transformation
- The Complex file writer
- Performance Considerations: Partitioning

- The Informatica abstraction layer
- Polyglot computing
- The Smart Executor
- Open source and innovation
- Connection architecture
- Connections to third Party applications

Module 6: Informatica Polyglot Computing in Hadoop

- Hive MR/Tez
- Blaze
- Spark
- Native
- The Smart Executor

- Parsing and processing Avro, Parquet, JSON, and XML file
- Data Processor Transformation Considerations

Module 11: Accessing NoSQL Databases

- CAP Theorem
- HBase
- MongoDB
- Cassandra

[Enroll Public](#)

[Enroll Virtual](#)

[Back to Course Overview](#)

QUESTIONS?

Instructor-Led | Big Data | 3 Days