

Big Data for Developers

 now.informatica.com/Big-Data-for-Developers-ILT.html

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. This course is applicable to software version 10.1.

Objectives

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

- Define “Big Data” as it applies to Informatica and ETL/ELT
- Enumerate the primary components of a Hadoop environment
- Describe a process to identify and prioritize migration of expensive Data Warehouse processes to Hadoop
- Migrate PowerCenter mappings to Big Data Management and ingest data into Hadoop
- Use SQOOP and the SQL to Mapping capability to migrate and ingest data into Hadoop
- Describe the Informatica on Hadoop architecture
- Transform data on Hadoop using Informatica polyglot computing
- Enumerate the capabilities of the Informatica engines on Hadoop including Hive MR, Hive Tez, Blaze, and Spark engines
- Identify the optimization methods used by the Informatica Smart Executor
- Utilize Informatica and Hadoop monitoring and troubleshooting
- Parse and transform complex data using the DT transformation and Big Data Parser
- Describe how Informatica parses, reads, and writes NoSQL data collections

Target Audience

- Developer

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

- Hadoop concepts
- Hadoop Architecture Components
- The Hadoop Distributed File System (HDFS)
- MapReduce
- “Yet Another Resource Manager” (YARN) (MapReduce Version 2)

Module 3: Informatica on Hadoop Architecture

- Explanation of the Big Data world
- Explanation of the concept to build once, Deploy Anywhere
- Illustrate the Informatica abstraction layer
- Informatica’s Polyglot computing engines
- Smart Executor
- Open Source through Innovation
- Connection architecture
- Informatica connections to third party applications on Hadoop

Module 4: Data Warehouse Offloading

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

Module 5: Code Migration and Ingestion

- Create and interpret PowerCenter Reuse Reports
- Import PowerCenter Mappings to Developer
- SQOOP
- SQL to Mapping capability for converting SQL code to Informatica mappings

Module 6: Informatica Polyglot Computing in Hadoop

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

Module 7: Monitoring, Logs, and Troubleshooting

- Monitor mappings
- Troubleshooting mappings in Hive

Module 8: Hadoop Data Integration Challenges and Performance Tuning

- Challenges with executing mappings in Hadoop
- Partitioning and Parallel Workflows
- Big Data Management Performance Tuning
- Mapping Level Tuning
- Tips

Module 9: Complex File Parsing

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

Module 10: Accessing NoSQL Databases

- Explain NoSQL Databases
- HBase
- MongoDB
- Cassandra

QUESTIONS?

Instructor Led | Big Data | 3 Days