Big Data Developer, Specialist Certification

About the ICS Big Data Developer Test and the Skill Set Inventory

This test measures your competency as a Developer utilizing Big Data to include basics such as application services, Hadoop, complex file parsing, the mechanics of data integration using developer tool. Test your knowledge of Big Data Streaming, mappings, Kafka, monitoring, troubleshooting.

The skill set inventory is used to guide your preparation before taking the test. It is an outline of the technical topics and subject areas that are covered in each test. The skill set inventory includes test domain weighting, test objectives and topical content. The topics and concepts are included to clarify the test objectives.

Test takers will be tested on:

- Big Data Integration Course Introduction
- Big Data Basics
- Informatica on Hadoop Architecture
- Data Warehouse Offloading
- Code Migration and Ingestion
- Informatica Polyglot Computing in Hadoop
- Monitoring, Logs, and Troubleshooting
- Hadoop Data Integration Challenges and Performance Tuning
- Complex File Parsing
- NoSQL Databases
- Fundamentals
- Developing Physical Data Objects
- Viewing Data
- Developing Mappings and Transformations
- Working with Dynamic Schema and Dynamic Mappings
- Deploying Applications
- Parameters
- Workflows
- Edge Data Streaming (EDS) Overview
- Big Data Streaming Overview
- Kafka Overview
- Streaming Mappings
- Monitoring Logs and Troubleshooting
- Performance Tuning and Best Practices
- End-to-End Use Case

Training Prerequisites

The skills and knowledge areas measured by this test are focused on product core functionality inside the realm of a standard project implementation. Training materials, supporting documentation and practical experience may become sources of question development.

The suggested training prerequisites for this certification level are the completion of the following Informatica course(s):

- Informatica Developer Tool for Big Data Developers (Instructor Led or onDemand)
- Big Data for Developers (Instructor Led or onDemand)
- Big Data Streaming for Developers (Instructor Led)

Informatica University

Skill Set

Test Domains

The test domains and the extent to which they are represented as an estimated percentage of the test follows:

Title	% of Test
Big Data Integration Course Introduction & Big Data Basics	6%
Informatica on Hadoop Architecture	4%
Data Warehouse Offloading & Code Migration and Ingestion	6%
Informatica Polyglot Computing in Hadoop	3%
Monitoring, Logs, and Troubleshooting	3%
Hadoop Data Integration Challenges and Performance Tuning	3%
Complex File Parsing	4%
NoSQL Databases	3%
Developer Tool Fundamentals	3%
Developing Physical Data Objects	4%
Viewing Data	4%
Developing Mappings and Transformations	4%
Working with Dynamic Schema and Dynamic Mappings	4%
Deploying Applications	4%
Parameters	4%
Workflows	6%
Edge Data Streaming (EDS) Overview	3%
Big Data Streaming Overview	1%
Kafka Overview	3%
Streaming Mappings	16%
Monitoring Logs and Troubleshooting	3%
Performance Tuning and Best Practices	4%
End-to-End Use Case	3%

Question Format

You may select from one or more response offerings to answer a question.

You will score the question correctly if your response accurately completes the statement or answers the question. Incorrect distractors are given as possible correct answers so that those without the required skills and experience may wrongly select that choice.

A passing grade of 70% is needed to achieve recognition as an Informatica Certified Specialist (ICS) in Big Data Developer.

You are given 90 minutes to complete the test. Formats used in this test are:

- Multiple Choice: Select one option that best answers the question or completes the statement
- Multiple Response: Select all that apply to best answer the question or complete the statement
- True/False: After reading the statement or questions select the best answer

Informatica University

Skill Set

Test Policy

- You are eligible for one attempt and re-take, if needed, per test registration.
- If you do not pass on your first attempt
    - Purchase of the test will include one second-attempt if a student does not pass an test.
    - You must wait two weeks after a failed test to take the test again.
    - Any additional retakes are charged the current fee at the time of purchase.
    - Promotions are excluded and cannot be combined.

Informatica University

Skill Set

Test Topics

The test will contain 70 questions comprised of topics that span across the sections listed below. In order to ensure that you are prepared for the test, review the subtopics with each section.

Big Data Basics
- Hadoop concepts
- Hadoop Architecture Components
- The Hadoop Distributed File System (HDFS)
- MapReduce
- "Yet Another Resource Manager" (YARN)

Informatica on Hadoop Architecture
- Explanation of the Big Data world
- Illustrate the Informatica abstraction layer
- Informatica's Polyglot computing engines
- Smart Executor
- Open Source through Innovation
- Connection architecture
- Informatica connections to third party apps on Hadoop

Data Warehouse Offloading
- The requirements of an optimal Data Warehouse
- Data Warehouse Offloading Process

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

Informatica Polyglot Computing in Hadoop
- Hive MR/Tez
- Blaze
- Spark
- The Smart Executor

Monitoring, Logs, and Troubleshooting
- Monitor mappings
- Troubleshooting mappings in Hive

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

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

NoSQL Databases
- NoSQL Databases - an overview
- Informatica HBase support
- Informatica MongoDB support
- Informatica Cassandra support

Fundamentals
- Introduction to the Developer tool
- Brief Overview of Big Data Management Architecture
- Reviewing the Developer interface

Developing Physical Data Objects
- Introduction to the types of physical data objects
- Using relational connections
- Using flat file connections
- Synchronize a flat file data object

Viewing Data
- Introduction to data views
- Troubleshooting configurations
- Reviewing logs
- Monitoring activities

Developing Mappings and Transformations
- Mapping and transformation concepts
- Core transformations
- Developing and validating a mapping
- Using transformations in a mapping

Working with Dynamic Schema and Dynamic Mappings
- Introduction to dynamic mapping concepts
- Developing and running a dynamic mapping
- Reviewing a mapping template

Deploying Applications
- Deploying Applications for Users
- Create, deploy, and run an application

Parameters
- Using Parameters
- Types of Parameters
- Run mappings with parameter file and parameter set

Workflows
- Working with Workflows
- Create a workflow
- Configure a workflow
- Add a conditional sequence flow

Informatica University

Skill Set

Edge Data Streaming (EDS) Overview
- EDS architecture and key features
- EDS Data flow process
- EDS UI
- Create an EDS Service
- Create, deploy and monitor a data flow

Big Data Streaming Overview
- Big Data Streaming overview
- Stream Data Processing with Spark streaming
- BDS component architecture
- BDS key features

Kafka Overview
- Kafka Concepts
- Kafka core APIs
- Kafka models

Streaming Mappings
- Sources in a streaming mapping
- Targets in streaming Mapping
- Lookup sources
- Kafka Data Object Properties

Monitoring Logs and Troubleshooting
- Spark Monitoring
- Viewing Logs
- Troubleshooting

Performance Tuning and Best Practices
- Tune performance of Spark jobs
- Best practices while working with streaming data

End-to-End Use Case
- Use Case
- EDS and BDS – Final Goal

Informatica University

Skill Set

Sample Test Questions

Scorecarding is performed on which of the following?

A. True
B. False

BDM connects to these technologies in a cluster:

A. HDFS
B. Hive
C. SQL
D. Yarn
E. Spark

NoSQL refers to what type of database?

A. non-SQL database
B. non-relational database
C. not-only SQL database
D. All of the Above

Which of the following is TRUE about Edge Data Streaming?

A. You configure the data flow in the Administrator tool.
B. You configure the data flow in the Developer tool.
C. You deploy the data flow configuration to the Data Integration Service.
D. You must choose a Data Integration Service in the Informatica Monitor.

The Developer Tool is a thin client and no installation is necessary.

A. True
D. False

When You Are Ready To Test:

Informatica Specialist Certifications are available Anytime/Anywhere. To become an Informatica Certified Specialist (ICS), please follow these steps.

1. *Go to the Informatica Certification Trainings located **here**.*
2. *Login with your Informatica Passport or create your account.*
3. *Locate the Certification you wish to take, click Certification under the title.*
4. *You will be brought to the Certification Details Page, click Enroll.*
5. *Click Add to Cart and complete your registration/purchase.*
6. *Once you have registered go to My Training and View Your Transcript.*
7. *Now you can simply Launch and take your test Anytime/Anywhere prior to your test's expiry date*
8. *Retake Policy: Current purchases of the test will include one second-attempt if a student does not pass a test. Any additional retakes are charged the current fee at the time of purchase. Promotions are excluded and cannot be combined. You must wait two weeks after a failed test to take the test again.*

Informatica

Worldwide Headquarters, 2100 Seaport Blvd., Redwood City, California 94063, USA
phone: 650.385.5000 fax: 650.385.5500 toll-free in the US: 1.800.653.3871
www.informatica.com linkedin.com/company/informatica twitter.com/Informatica

© 2018 Informatica LLC. All rights reserved. Printed in the U.S.A. Informatica, the Informatica logo, and The Data Integration Company are trademarks or registered trademarks of Informatica LLC in the United States and in jurisdictions throughout the world. All other company and product names may be trade names or trademarks of their respective owners.