Big Data Management 10.1, Specialist Certification

About the ICS Big Data Management 10.1 Test and the Skill Set Inventory

This test measures your competency in utilizing PowerCenter mappings and workflows at basic and advanced levels in order to perform data integration on Big Data. It will test your ability to integrate PowerCenter with Hadoop clusters and the related Hadoop ecosystem.

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 Basics
- Data Warehouse Offloading
- Data Ingestion
- Big Data Management Architecture
- Creating Mappings and Polyglot Computing
- Monitoring and Troubleshooting
- Mapping Challenges and Performance Tuning
- Data Quality
- Complex Files
- NoSQL Databases
- Developer Fundamentals
- Creating Physical Data Objects
- Viewing Data
- Parameters, Parameter Files and Parameter Sets
- Workflows and Applications

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):

- PowerCenter: Data Integration for Developers (Instructor Led) OR PowerCenter: Developer, Level 1 (onDemand)
- Informatica Developer Tool for Big Data Developers (Instructor Led or onDemand)
- Big Data for Developers (Instructor Led or onDemand)

Test Domains

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

Title	% of Test
Developer Tool for Big Data Developers: Fundamentals	3%
Developer Tool for Big Data Developers: Developing Physical Data Objects	3%
Developer Tool for Big Data Developers: Viewing Data	6%
Developer Tool for Big Data Developers: Developing Mappings and Transformations	9%
Developer Tool for Big Data Developers: Working with Dynamic Schema and Dynamic	4%
Developer Tool for Big Data Developers: Parameters	4%
Developer Tool for Big Data Developers: Workflows	6%
Developer Tool for Big Data Developers: Working with Applications	4%
Big Data Management for Developers: Accessing NoSQL Databases	4%
Big Data Management for Developers: Big Data Basics	9%
Big Data Management for Developers: Big Data Management Architecture	10%
Big Data Management for Developers: Complex File Parsing	6%
Big Data Management for Developers: Data Warehouse Offloading	4%
Big Data Management for Developers: Hadoop Data Integration Challenges and Perfomance	9%
Big Data Management for Developers: Informatica Polyglot Computing in Hadoop	6%
Big Data Management for Developers: Ingestion and Offload	9%
Big Data Management for Developers: Mappings, Monitoring, and Troubleshooting	6%

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 Management 10.1.

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

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.

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 and Architecture
- HDFS
- YARN
- MapReduce

Data Warehouse Offloading
- Challenges with Traditional DW
- Requirements for Offloading
- The Offloading Process

Data Ingestion
- PowerCenter Reuse Reports
- Importing PowerCenter Mappings to Developer
- SQOOP
- SQL to Mapping Feature
- Partitioning and Parallelism

Big Data Management Architecture
- The Informatica Abstraction Layer
- Polyglot Computing
- The Smart Executor
- Open source and innovation
- Connection Architecture

Creating Mappings and Polyglot Computing
- Mapping and Transformation Concepts
- Core Transformations
- Developing and Validating a Mapping
- Configuring and running a mapping in Native and Hadoop environments
- Hive MR/Tez
- Blaze
- Spark
- Native
- The Smart Executor

Monitoring and Troubleshooting
- Configuring and Run Mappings in Native and Hadoop Environments
- Execution Plans
- Monitor Mappings
- Troubleshoot Mappings
- Viewing Mapping Results

Mapping Challenges and Performance Tuning
- Mapping Design Challenges in Hadoop
- Big Data Management Performance Tuning
- Hive Environment Optimization
- Mapping Level Tuning
- DIS Level Tuning
- Cluster Level Tuning

Data Quality
- 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

Complex Files
- The Complex File Reader/Writer
- The Data Processor transformation
- Partitioning
- Parsing and Processing Avro, Parquet, JSON, and XML Files
- Data Processor Transformation Considerations

NoSQL Databases
- CAP Theorem
- HBase
- MongoDB
- Cassandra

Developer Fundamentals
- Introduction to the Developer tool
- The Developer interface

Creating Physical Data Objects
- Types of Physical Data Objects
- Using Relational and Flat File Connections
- Synchronizing Data Objects

Viewing Data
- Viewer Configurations
- Monitoring and Logs

Parameters, Parameter Files and Parameter Sets
- Parameters in Developer
- Parameter Files and Parameter Sets

Workflows and Applications
- Workflows
- Deploying Applications

Sample Test Questions

Which of the following is not a supported Hadoop execution engine?

A.	☒	Spark
B.	☑	Storm
C.	☒	Blaze
D.	☒	Hive

Sequence Generators cannot be used in mappings executed by either the Hive or Spark engine. Workarounds include:

A.	☒	Use the Hive serializer
B.	☒	Use an Expression transformation to increment an input parameter
C.	☒	Use an External Procedure transformation
D.	☑	Use an Expression transformation with the UUID4 function

For best performance, when joining more than two sources

A.	☒	Use concurrent Joiner transformations when possible
B.	☒	Order the sources to join from largest to smallest
C.	☑	Order the sources to join from smallest to largest
D.	☒	Ordering the sources based on size has no impact on performance

Which of the following is not a valid data object in Developer?

A.	☒	Complex File Object
B.	☑	Queue Object
C.	☒	Schema Object
D.	☒	Logical Data Object

Which of the following best describes executing a mapping with the Hive engine?

A.	☑	The DIS converts the mapping metadata into a query language which, when passed to the proper server, is further converted into map/reduce
B.	☒	The DIS converts the mapping metadata into a Scala program which, is then submitted to Yarn for execution on the Hadoop cluster
C.	☒	The DIS converts the mapping metadata into segments and tasklets. An orchestrator process is created to supervise containers on data nodes identified by YARN
D.	☒	The DIS processes the mapping metadata directly on an Informatica server node

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.*

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