PowerCenter Data Integration 10.1: Big Data Specialist Certification

About the ICS PowerCenter Data Integration Big Data Exam and the Skill Set Inventory

This exam 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 exam. It is an outline of the technical topics and subject areas that are covered in each exam. 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 exam 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 10: Developer, Level 1 (IVA, ILT)
- Informatica Developer Tool 10 for Big Data Developers (IVA, ILT)
- Big Data 10.1 for Developers (IVA, ILT)

Exam Test Domains

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

	% of Test
Big Data Basics	8.70
Data Warehouse Offloading	4.30
Data Ingestion	7.25
Big Data Management Architecture	6.75
Creating Mappings and Polyglot Computing	14
Monitoring and Troubleshooting	6.75
Mapping Challenges and Performance Tuning	10.0
Data Quality	7.75
Complex Files	8.00
NoSQL Databases	5.25
Developer Fundamentals	3.00
Creating Physical Data Objects	3.50
Viewing Data	4.00
Parameters, Parameter Files and Parameter Sets	3.50
Workflows and Applications	7.25

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) Data Integration 10.1 for Big Data.

You are given 90 minutes to complete the exam. Test formats used in this exam are:

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

Exam Policy

- You are eligible for one attempt per exam registration.
- If you do not pass on your first attempt
- You must wait 2 weeks after the exam to take the exam again
- You may take the exam up to three times in one year from the date of your first exam attempt
- You must pay the full exam fee each time you take the exam.

Test Topics

The exam 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 Informatoca server node

When You Are Ready To Test:

To schedule your Informatica Certified Professional (ICP) exam, please follow these steps – you will be presented with the list of current ICP exams to choose from.

1. *Go to the Informatica Certified Professional exam scheduling web site at http://www.webassessor.com/informatica.*
2. *Login or create your new test taker account (top right corner if you do not already have an account).*
    a. *To receive credit for your exam, remember to create your account using your corporate email address.*
3. *Register your exam location, the time and date for your exam of choice.*
4. *Select from the two testing location options (testing center or online), the time that you want and proceed to checkout:*
    a. *You can either take your exam in person at an authorized testing center (Onsite proctored).*
    b. *Or you can take your exam online from your home or office (Online Proctored). To take your exam online, you need an EXTERNAL webcam. A webcam that is built into your laptop/PC is NOT acceptable when passing your exam online. Please review the instructions on your test taker welcome page after you have logged in.*
5. *Proceed to checkout and finalize your exam appointment scheduling (if you have been provided with an exam voucher or coupon, enter it during this step and proceed to checkout).*
6. *Within the next couple of minutes you will receive an email confirming your exam registration details (check your email program's junk/spam folder, if you do not receive the email). You can also verify your exam appointment details at any time via the exam scheduling site.*

Once you have scheduled your exam appointment, you can reschedule it via the exam scheduling site free of charge as per the Informatica Certified Professional program's guidelines: http://www.informatica.com/us/services-and-training/certification/certification-faq/.

informatica University

Worldwide Headquarters, 2100 Seaport Blvd., Redwood City, California 94063, USA
phone: 650.385.5000 fax: 650.385.5500 toll-free in the US: 1.800.653.3871338
www.informatica.com linkedin.com/company/informatica twitter.com/InformaticaCorp

© 2016 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 Corporation 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.