C2090-102

IBM CERTIFIED DATA ARCHITECT - BIG DATA

ANALYTICSEXAM.COM

Exam Summary – Syllabus – Questions

Table of Contents

Introduction to C2090-102 Exam on IBM Certified Data Architect - Big Data

This page is a one-stop solution for any information you may require for IBM Certified Data Architect - Big Data (C2090-102) Certification exam. The IBM C2090-102 Exam Summary, Syllabus Topics and Sample Questions provide the base for the actual IBM Certified Big Data Architect exam preparation, we have designed these resources to help you get ready to take your dream exam.

The IBM Certified Data Architect - Big Data credential is globally recognized for validating IBM Big Data Architect knowledge. With the IBM Certified Big Data Architect Certification credential, you stand out in a crowd and prove that you have the IBM Big Data Architect knowledge to make a difference within your organization. The IBM Certified Data Architect - Big Data Certification (C2090-102) exam will test the candidate's knowledge on following areas.

IBM C2090-102 Certification Details:

Exam Name	IBM Certified Data Architect - Big Data
Exam Code	C2090-102
Exam Duration	90 minutes
Exam Questions	55
Passing Score	60%
Exam Price	$200 (USD)
Training	Test Preparation Resource
Exam Registration	Pearson VUE
Sample Questions	IBM Big Data Architect Certification Sample Question
Practice Exam	**IBM Big Data Architect Certification Practice Exam**

IBM C2090-102 Exam Syllabus:

Objective	Details
Requirements (16%)	- Define the input data structure - Define the outputs - Define the security requirements - Define the requirements for replacing and/or merging with existing business solutions - Define the solution to meet the customer's SLA - Define the network requirements based on the customer's requirements

Objective	Details
Use Cases (46%)	- Determine when a cloud-based solution is more appropriate vs. in-house (and migration plans from one to the other) - Demonstrate why Cloudant would be an applicable technology for a particular use case - Demonstrate why SQL or NoSQL would be an applicable technology for a particular use case - Demonstrate why Open Data Platform would be an applicable technology for a particular use case - Demonstrate why BigInsights would be an applicable technology for a particular use case - Demonstrate why BigSQL would be an applicable technology for a particular use case - Demonstrate why Hadoop would be an applicable technology for a particular use case - Demonstrate why BigR and SPSS would be an applicable technology for a particular use case - Demonstrate why BigSheets would be an applicable technology for a particular use case - Demonstrate why Streams would be an applicable technology for a particular use case - Demonstrate why Netezza would be an applicable technology for a particular use case - Demonstrate why DB2 BLU would be an applicable technology for a particular use case - Demonstrate why GPFS/HPFS would be an applicable technology for a particular use case - Demonstrate why Spark would be an applicable technology for a particular use case - Demonstrate why YARN would be an applicable technology for a particular use case
Applying Technologies (16%)	- Define the necessary technology to ensure horizontal and vertical scalability - Determine data storage requirements based on data volumes - Design a data model and data flow model that will meet the business requirements - Define the appropriate Big Data technology for a given customer requirement (e.g. Hive/HBase or Cloudant) - Define appropriate storage format and compression for given customer requirement
Recoverability (11%)	- Define the potential need for high availability - Define the potential disaster recovery requirements - Define the technical requirements for data retention - Define the technical requirements for data replication - Define the technical requirements for preventing data loss
Infrastructure (11%)	- Define the hardware and software infrastructure requirements - Design the integration of the required hardware and software components - Design the connectors / interfaces / API's between the Big Data solution and the existing systems

C2090-102 Sample Questions:

Q 1: A bank wants to build a system that tracks all ATM and online transactions in realtime.

They want to build a personalized model of their customer's financial activity by incorporating enterprise data as well as social media data. The system must be able to learn and adapt over a period of time.

These personalized models will be used for real time promotions as well as for any fraud or crime detections. Given these requirements, which of the following would recommend?
Options:
A. Netezza

B. Hadoop

C. Spark

D. Cloudand

Q 2: A large Retailer (online and "brick & mortar") processes data for analyzing marketing campaigns for their loyalty club members. The current process takes weeks for processing only 10% of social data.

What is the most costeffective platform for processing and analyzing campaign results from social data on a daily basis using 100% dataset?
Options:
A. High Speed Mainfraime Processing

B. Enterprise Data Warehouse

C. BigInsights Open Data Platform

D. In Memory Computing

Q 3: A media company wants to measure the effectiveness of their advertising campaign. Before they release a movie they prepare and run a campaign for promotion.

Based on the response on Twitter and Facebook they want to decidewhetheror not they should continue a particular campaign. Which of the following should be selected to meet these requirements?
Options:
A. Hadoop

B. Unica

C. Streams

D. Pure Data for Analytics

Q 4: A telecommunication company needs a Big Data solution that could store and analyze multiple years worth of call detail records (CDRs, aprox. 17 billion events per day) containingswitch, billing, and network event data for its millions of subscribers.

Which of the following would you recommend for these requirements?
Options:
A. Pure Data System for Analytics

B. Infosphere DataStage

C. DB2

D. SPSS

Q 5: As you explore the data for a BigSheets workbook, you must run the workbook against the full data set to get the most current results for analysis. Which statement is TRUE regarding running and visualizing data in a workbook?
Options:
A. When you add sheets to workbooks, saving the sheets runs the individual data for the sheet but not for the full workbook

B. By default, the first sheet in your workbook is named the Results sheet

C. When you save and run the workbook, the data in a Child Workbook is the output for that workbook

D. You can create graphs for more than one sheet within the same workbook

Q 6: Faced with a wide area network implementation, you have a need for asynchronous remote updates. Which one of the following would best address this use case?
Options:
A. GPFS File clones can be created from a regular file or a file in a snapshot using the mmclonecommand

B. HDFS NameNode The NameNode keeps an image of the entire file system namespace and file Blockmap in memory. This key metadata item is designed to be compact, such that a NameNode with 4 GB of RAM is plenty to support a huge number of files and directories

C. HDFS Cluster rebalancing is compatible with data rebalancing schemes. A scheme might automatically move data from one DataNode to another if the free space on a DataNode falls below a certain threshold

D. GPFS Active File Management allows data access and modifications even when remote storage cluster is unavailable

Q 7: In designing a new Hadoop system for a customer, the option of using SAN versus DAS was brought up. Which of the following would justify choosing SAN storage?
Options:
A. SAN storage reduces and removes a lot of the HDFS complexity and management issues

B. SAN storage provides better performance than DAS

C. SAN storage supports replication, reducing the need for 3-way replication

D. SAN storage removes the Single Point of Failure for the NameNode

Q 8: The AQL query language is the easiest and most flexible tool to pull structured output from which of the following?
Options:
A. JDBC connected relational data marts

B. Hbase schemas

C. Unstructured text

D. Hive data structures

Q 9: Which of the following is a requirement for data retention and archival?
Options:
A. Hosting location

B. A format and storage repository for archived data

C. Solid-state technology

D. Public cloud

Q 10: Which of the following statements is TRUE regarding cloud computing solutions?
Options:
A. Cloud solutions rely on scaling up (vertical) scaling vs. scale out (horizontal) scaling

B. Server virtualization is a requirement in a cloud implementation

C. Stateless applications are better candidates for cloud services than applications that maintain state

D. Cloud security is planned, developed, and layered on top of an application after the application development process is complete

Answers to C2090-102 Exam Questions:

Question: 1	Answer: A	Question: 2	Answer: C
Question: 3	Answer: B	Question: 4	Answer: A
Question: 5	Answer: C	Question: 6	Answer: A
Question: 7	Answer: C	Question: 8	Answer: D
Question: 9	Answer: B	Question: 10	Answer: B

Note: If you find any typo or data entry error in these sample questions, we request you to update us by commenting on this page or write an email on feedback@analyticsexam.com