

# Exam Questions C2090-102

IBM Big Data Architect

<https://www.2passeasy.com/dumps/C2090-102/>



#### NEW QUESTION 1

An upstream Oil and Gas Producer needs to optimize the performance of its assets. It needs to calculate Key Performance Indicators for flow rate sensors deployed to monitor the output flow and temperature and pressure for multiple pipelines in an oil field with hundreds of wells. Which of the following would you recommend to meet these requirements?

- A. Datastage ETL jobs should be created
- B. Infosphere Streams should be used
- C. Cognos reports would suffice here
- D. Hadoop based datastorage engine should be used

**Answer:** B

#### NEW QUESTION 2

Which of the following can NOT be performed by the Open Data Platform (ODP)?

- A. Provide development and governance of upstream projects
- B. Contribute to the Apache Software Foundation projects
- C. Produce a set of tools and methods that enable members to create and test differentiated offerings based on the ODP core
- D. Accelerate the delivery of Big Data solutions by providing a well-defined core consisting of all the components of the Apache Software Foundation projects

**Answer:** C

#### NEW QUESTION 3

The Yarn Resource Managers (RM) have an option to embed an ActiveStandbyElector to decide which RM should be the Active. Upon which of the following is it based?

- A. Job Tracker
- B. Zookeeper
- C. Task Tracker
- D. Name Node

**Answer:** C

#### NEW QUESTION 4

As data architect, understanding where patient data resides, how is it related, and how you define security and privacy policies requires knowledge of each of the following?

- A. PII Data
- B. HIPAA Requirements
- C. Privacy by design
- D. Data storage in HDFS

**Answer:** B

#### NEW QUESTION 5

Which one of the following statements about Big SQL is TRUE?

- A. Big SQL doesn't need any secondary indices to access HBase tables
- B. Big SQL processes queries locally either on disk or in memory
- C. Big SQL supports updates in Hive.
- D. Executing Big SQL queries through MapReduce framework would always be a better choice

**Answer:** C

#### Explanation:

References:

<http://www.ibm.com/developerworks/library/bd-bigsq/>

#### NEW QUESTION 6

By default, Parquet uses which of the following codecs?

- A. SNAPPY
- B. LZO
- C. GZIP
- D. BZIP2

**Answer:** A

#### NEW QUESTION 7

You need to set up a distributed storage system for being able to process very large data sets and you want to be able to leverage the Open Data Platform (ODP) Core. Which one of the following would you use?

- A. Apache Spark
- B. IBM GPFS

C. EMC Isilon  
D. HDFS

**Answer:** B

#### NEW QUESTION 8

Data high availability is provided by SAN in traditional architectures by employing a level of RAID. What is the Hadoop equivalent?

A. Data Node  
B. Replication  
C. Quorum Journal Manager  
D. Yarn Resource Manager

**Answer:** B

#### Explanation:

References:

<http://www.computerweekly.com/feature/Big-data-storage-Hadoop-storage-basics>

#### NEW QUESTION 9

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) containing switch, billing, and network event data for its millions of subscribers. Which of the following would you recommend for these requirements?

A. Infosphere DataStage  
B. DB2  
C. Pure Data System for Analytics  
D. SPSS

**Answer:** C

#### NEW QUESTION 10

Which of the following is a requirement for data retention and archival?

A. A format and storage repository for archived data  
B. Public cloud  
C. Hosting location  
D. Solid-state technology

**Answer:** A

#### NEW QUESTION 10

Which language is used to define secondary views in a Cloudant database?

A. Ruby  
B. Java, C++, or Perl  
C. Javascript  
D. Python

**Answer:** C

#### NEW QUESTION 14

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?

A. You can create graphs for more than one sheet within the same 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. When you add sheets to workbooks, saving the sheets runs the individual data for the sheet but not for the full workbook

**Answer:** C

#### Explanation:

Reference:

[https://www-01.ibm.com/support/knowledgecenter/SSPT3X\\_4.1.0/com.ibm.swg.im.infosphere.biginights.analyze.doc/doc/bigsheets\\_con\\_workbooks.html](https://www-01.ibm.com/support/knowledgecenter/SSPT3X_4.1.0/com.ibm.swg.im.infosphere.biginights.analyze.doc/doc/bigsheets_con_workbooks.html)

#### NEW QUESTION 19

The CAP Theorem states that it is not possible for a distributed computer system to guarantee all three of these?

A. Consistency, Accuracy, and Partition tolerance  
B. Concurrency, Availability, and Parallel updates  
C. Concurrency, Accuracy, and Parallel updates  
D. Consistency, Availability, and Partition tolerance

**Answer:** B

#### NEW QUESTION 24

Data Architects assess data according to several common characteristics. One of these characteristics is the type of data. Which of the following are data types that fall into this ontology? (Choose two.)

- A. Master data
- B. Meta data
- C. Boolean
- D. Composite
- E. Integer

**Answer:** AD

#### NEW QUESTION 28

What is the most important aspect of data center disaster recovery?

- A. A complete damage assessment
- B. Control of critical assets
- C. Restoration of business functions
- D. Protection of individual life

**Answer:** D

#### NEW QUESTION 29

Defining your need to enrich existing customer data you realize that you need to process large quantities of Geospatial data and output to your data warehouse in a standard GeoJSON format. Which of the following would provide a business analyst with the desired output?

- A. Big SQL
- B. BigSheets
- C. Hive queries
- D. Text Analytics

**Answer:** C

#### NEW QUESTION 34

What term applies to the data elements in Infosphere Streams?

- A. Tuples
- B. Operators
- C. Sink adapters
- D. Composite operators

**Answer:** B

#### Explanation:

Reference:

[http://www-01.ibm.com/support/knowledgecenter/SSCRJU\\_3.2.1/com.ibm.swg.im.infosphere.streams.glossary.doc/doc/glossary\\_Streams.html?lang=en](http://www-01.ibm.com/support/knowledgecenter/SSCRJU_3.2.1/com.ibm.swg.im.infosphere.streams.glossary.doc/doc/glossary_Streams.html?lang=en)

#### NEW QUESTION 36

A television company is going to launch a new talk show and wants to gauge viewers' interest in various topics. They would like to do social media analysis and find out best responses. They also want to discuss best responses during the next day's show. Given these requirements, which technology you would recommend?

- A. Hadoop
- B. Streams
- C. Content Management
- D. Netezza

**Answer:** C

#### NEW QUESTION 37

On startup, the NameNode enters a special state called Safemode. Replication of data blocks does not occur when the NameNode is in the Safemode state. The NameNode receives Heartbeat and Blockreport messages from the DataNodes. Which of the following contains the list of data blocks that a DataNode is hosting?

- A. Rack Awareness
- B. Snapshot
- C. Replica
- D. Blockreport

**Answer:** D

#### Explanation:

References:

[https://hadoop.apache.org/docs/r1.2.1/hdfs\\_design.html](https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html)

#### NEW QUESTION 41

A major telecommunication company has millions of customers. Most of their customers are prepaid. Being prepaid customers, they can very easily switch to other

vendors. The last four to six months, this company has lost quite a good number of customers to competition. They intend to build a system that can provide them with insight into the customer's social network (e.g. who is the influencer and who is the follower). They also want the ability to monitor the voice and data usage patterns in real time and they want the system to be trained over time to predict possible dissatisfactions. Given this scenario, which one of the following would you recommend?

- A. Hadoop
- B. Spark
- C. Cloudant
- D. Netezza

**Answer:** B

#### NEW QUESTION 44

What is used to capture client requirements for software selection and to evaluate the initial functional "fit" of a vendor's software solution to the business needs of the client?

- A. Operational Model
- B. Requirements Matrix
- C. Viability Assessment
- D. Use Case Model

**Answer:** B

#### NEW QUESTION 47

Which of the following is NOT a valid Big Data platform integration?

- A. Platform plugins
- B. Intraplatform integration
- C. Enterprise integration with other repositories
- D. Network integration

**Answer:** D

#### NEW QUESTION 49

Which of the following statements regarding Big R is TRUE?

- A. Missing data values must be handled by ETL processes prior to analyzing data with Big R
- B. A bigr.frame loads data in memory for optimal performance
- C. A Big R user is responsible for parallelizing the execution of the R functions being used in the R program
- D. Performing a mathematical operation on a Big R vector variable will automatically loop through each item in the vector

**Answer:** D

#### Explanation:

Reference:

<http://www.computerworld.com/article/2497319/business-intelligence-beginner-s-guide-to-r-syntax-quirks-you-ll-want-to-know.html>

#### NEW QUESTION 54

Which of the following statements regarding Big R is TRUE?

- A. The Big R API is completely identical to the R API
- B. Big R implements an interpreted computer language
- C. Big R users cannot access Big R capabilities from the standard RStudio client as RStudio is not Hadoop enabled
- D. Big R utilizes the big SQL query engine for processing

**Answer:** D

#### Explanation:

\_References: [https://www.ibm.com/support/knowledgecenter/SSPT3X\\_3.0.0/com.ibm.swg.im.infosphere.biginights.bigr.doc/doc/intro.html](https://www.ibm.com/support/knowledgecenter/SSPT3X_3.0.0/com.ibm.swg.im.infosphere.biginights.bigr.doc/doc/intro.html)

#### NEW QUESTION 59

Which of the following statements is TRUE regarding Cloud deployment models?

- A. Performance and scalability requirements are a critical factor for deciding between Platform as a Service and Infrastructure as a Service deployment models
- B. In a platform as a Service offering, the customer has root access to the servers
- C. Applications with extremely high transactions volumes are good candidates for Platform as a Service
- D. In an infrastructure as a service deployment, the cloud provider provides security patching, monitoring and fail over capabilities

**Answer:** A

#### NEW QUESTION 61

Company K is designing their Big Data system. In their enterprise, they anticipate every 9 months there will be a big spike of new data on the order of multiple TB. Their company policy also dictates that data older than one year will be archived with a major clean up every 5 years. Cost is also a big issue. Which of the following provides the best design for these requirements?

- A. Estimate the peak volume over a 5 year period and set up a Hadoop system with commodity HW and storage to accommodate that volume

- B. Estimate the peak volume over a 3 year period and set up a Hadoop system with NAS to accommodate the expected volume
- C. Use Cloud elasticity capabilities to handle the peak and valley data volume
- D. Use SAN storage with compression to handle the peak and valley data volume

**Answer:** A

#### NEW QUESTION 62

.....

## THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual C2090-102 Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the C2090-102 Product From:

<https://www.2passeasy.com/dumps/C2090-102/>

## Money Back Guarantee

### C2090-102 Practice Exam Features:

- \* C2090-102 Questions and Answers Updated Frequently
- \* C2090-102 Practice Questions Verified by Expert Senior Certified Staff
- \* C2090-102 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- \* C2090-102 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year