

Exam Questions C2090-102

IBM Big Data Architect

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



NEW QUESTION 1

The inputs to the Architectural Overview document do NOT include which of the following?

- A. Architectural Goals
- B. Key Concepts
- C. Architectural Overview Diagram
- D. Component Model

Answer: D

NEW QUESTION 2

“The programming model for client developers will hide the complexity of interfacing to legacy systems” is an example of which of the following?

- A. A use case
- B. An architectural decision
- C. A client imperative
- D. An empathy statement

Answer: B

NEW QUESTION 3

Which of the following is the artifact that assists in ensuring that the project is on the right path toward success?

- A. Component Model
- B. Empathy Map
- C. Viability Assessment
- D. Opportunity Plan

Answer: C

NEW QUESTION 4

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?

- A. SAN storage provides better performance than DAS
- B. SAN storage reduces and removes a lot of the HDFS complexity and management issues
- C. SAN storage removes the Single Point of Failure for the NameNode
- D. SAN storage supports replication, reducing the need for 3-way replication

Answer: D

NEW QUESTION 5

In a typical Hadoop HA cluster, two separate machines are configured as which of the following?

- A. Data Nodes
- B. Edge Nodes
- C. Name Nodes
- D. None of the Above

Answer: A

Explanation:

Reference:

<http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-hdfs/HDFSHighAvailabilityWithQJM.html>

NEW QUESTION 6

Service Level Agreements (SLAs) are defined at different levels. Which of the following is NOT a valid level?

- A. Customer-Based SLA
- B. Service-Based SLA
- C. Multilevel SLA
- D. Third-Party SLA

Answer: C

NEW QUESTION 7

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 8

Financial company A needs to ingest, analyze and store 8 billion structured transactions per day (around 2TB to 4TB of data) which translates full table scans, 24/7 access without any downtime, trend analysis, and the search for evidence of fraudulent activity. Which of the following technologies would you recommend to the company to meet these requirements?

- A. IBM Cognos and Unica
- B. Pure Data System for Analytics
- C. Infosphere Streams
- D. The Advanced Text Analytic Toolkit

Answer: B

NEW QUESTION 9

Which one of the following statements is TRUE?

- A. Big SQL uses Hadoop MR framework to process query tasks in parallel
- B. Big SQL executes queries locally on Big SQL server single node on a multi node cluster
- C. Big SQL can process queries in parallel and executes queries locally
- D. Big SQL only works with HDFS

Answer: A

Explanation:

Reference:

https://www-01.ibm.com/support/knowledgecenter/SSPT3X_3.0.0/com.ibm.swg.im.infosphere.biginsights.analyze.doc/doc/bigsql_performance.html

NEW QUESTION 10

An IBM Big Data platform is well suited to deal which of the following kinds of data types?

- A. Structured data in row format only
- B. Semi-structured and unstructured data only
- C. Text data, sensor data, and audio data only
- D. Semi-structured, unstructured, and structured data

Answer: D

Explanation:

Reference:

<https://www-304.ibm.com/industries/publicsector/fileserv?contentid=239170>

NEW QUESTION 10

Data stored in Netezza can be used for model building, scoring, and model refresh. How is SPSS Modeler connectivity to Netezza enabled?

- A. Open Database Connectivity (ODBC) driver provided by Netezza
- B. DataDirect Connect for ODBC drivers
- C. OLE DB drivers provided by Netezza
- D. Java Database Connectivity (JDBC) driver provided by Netezza

Answer: A

Explanation:

Reference: http://www.ibm.com/developerworks/data/library/cognos/infrastructure/cognos_specific/page569.html

NEW QUESTION 14

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 15

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 decide whether or not they should continue a particular campaign. Which of the following should be selected to meet these requirements?

- A. Hadoop
- B. Streams

- C. Unica
- D. Pure Data for Analytics

Answer: C

NEW QUESTION 18

Which of the following is the section of the Component Model that details how the solution integrates?

- A. Component Relationship Diagram
- B. Component Interface Diagram
- C. Component Interaction Diagram
- D. Component Reaction Diagram

Answer: A

NEW QUESTION 22

By default, Parquet uses which of the following codecs?

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

Answer: A

NEW QUESTION 27

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 28

In computer science, which of the following represents a set of properties that guarantee that database transactions are processed reliably?

- A. Foreign Key
- B. Primary Key
- C. Buffer Cache
- D. ACID

Answer: D

Explanation:

References:

<https://en.wikipedia.org/wiki/ACID>

NEW QUESTION 32

If the recovery point objective (RPO) is low, which of the following techniques would be the most appropriate?

- A. Clustering
- B. Database Shadowing
- C. Remote Journaling
- D. Tape Backup

Answer: A

Explanation:

References:

<http://whatis.techtarget.com/definition/recovery-point-objective-RPO>

NEW QUESTION 37

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 38

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?

- A. GPFS Active File Management allows data access and modifications even when remote storage cluster is unavailable
- B. HDFS Cluster rebalancing is compatible with data rebalancing scheme
- C. 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 File clones can be created from a regular file or a file in a snapshot using the mmclone command
- E. HDFS NameNode The NameNode keeps an image of the entire file system namespace and file Blockmap in memory
- F. 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

Answer: C

Explanation:

Reference:

http://www-01.ibm.com/support/knowledgecenter/STXKQY_4.1.1/com.ibm.spectrum.scale.v4_r11.adv.d oc/bl1adv_clones.htm

NEW QUESTION 40

You are going to develop Spark applications for your client. Which of the following cluster managers, that is also part of the Open Data Platform (ODP), would you use to focus on building the business driven application?

- A. Amazon EC2
- B. Apache Mesos
- C. Hadoop Yarn
- D. Standalone deploy mode

Answer: C

NEW QUESTION 42

A bank wants to build a system that tracks all ATM and online transactions in real-time. 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?

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

Answer: D

NEW QUESTION 46

As an Independent Software Vendor (ISV), you are tasked with creating a solution to harness the big data to better understand customer sentiments around your client's product. In order to develop a solution, you want to select the software components that are going to provide the most compatibility with new open source components. In addition to Hadoop and any other software components you may need, which one of the following would you select that is part of the initial release of the Open Data Platform (ODP)?

- A. Spark
- B. Ambari
- C. Hive
- D. HBase

Answer: B

Explanation:

References:

<http://hortonworks.com/innovation/open-data-platform/>

NEW QUESTION 50

Which architecture document is used to help organize projects, manage the complexity of the solution, and ensure that all architecture requirements have been addressed?

- A. Operational Model
- B. Component Model
- C. Connection Model
- D. API Model

Answer: B

NEW QUESTION 54

Which of the following is a consideration in sizing an active archive Hadoop infrastructure?

- A. replication factor within Hadoop
- B. Reporting requirements
- C. velocity or rate at which data is being generated
- D. veracity or trustworthiness of the data

Answer: B

Explanation:

Reference:

<http://www.ibm.com/developerworks/library/ba-augment-data-warehouse3/>

NEW QUESTION 56

The YARN High Availability feature adds redundancy in the form of an Active/Standby. Which of the following will pair to remove this otherwise single point of failure?

- A. JobTracker
- B. Data Node
- C. Management Node
- D. Resource Manager

Answer: D

Explanation:

References:

<http://hadoop.apache.org/docs/current/hadoop-yarn/hadoop-yarn-site/ResourceManagerHA.html>

NEW QUESTION 58

A company has to design a new data system. They will need to support several OLTP applications. Every three days a batch job will run to load specific data into a set of 10 large tables (with historical data) where OLAP analytics will be performed. Performance for both OLTP and OLAP queries is important. Which of the following designs would you suggest to the company?

- A. Use a NoSQL data store such as MongoDB or Cloudant on the cloud to provide needed scalability
- B. Use DB2 Data Partition Feature (DPF), partitioning all tables into different partitions
- C. Use DB2 Data Partition Feature (DPF), only partitioning the 10 tables where Analytics will be run
- D. Use DB2 with BLU Acceleration, use columnar store for the 10 tables where Analytics will be run

Answer: D

NEW QUESTION 63

A quality requirement or constraint that a system (or some part of a system) must satisfy is referred to by which of the following?

- A. Define Objective
- B. Non-functional Requirement
- C. Use Case Definition
- D. Service Level Agreement

Answer: D

NEW QUESTION 64

A telco company is struggling to retain their high value customers. The marketing team has access to the billing data and is using it for customer segmentation. They would like to tap on network and social data to understand customer behavior before crafting any special offer. How best can they achieve this task?

- A. By loading both network and social data in the current Enterprise Data Warehouse, then run analytics
- B. By loading network and social data in BigInsights for exploration then moving resulting data to Enterprise Data Warehouse, and merging with billing data for analytics
- C. By loading network, social, and billing data in BigInsights and using available analytics tools
- D. By creating a dedicated data mart in their current Enterprise Data Warehouse

Answer: B

Explanation:

Reference:

<https://ibmdatawarehousing.wordpress.com/category/ibm-data-warehouse/>

NEW QUESTION 65

Which of the following is a key consideration when determining the potential use of Models in Big Data Deployments?

- A. Stability of schema
- B. Level of Information Processing
- C. Ownership/Governance
- D. All of the above

Answer: B

NEW QUESTION 69

A news organization wants to analyze all the news stories coming in real time and make them available to their users based on the user's interest. Given this requirement, which of the following would you recommend?

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

Answer: A

NEW QUESTION 72

Which of the following does NOT consist of collaboration from various organizations to drive innovation and standardization across big data technologies?

- A. Cloudera Enterprise
- B. Hortonworks Data Platform
- C. IBM BigInsights
- D. Pivotal Big Data Suite

Answer: C

Explanation:

Reference:

<http://www-01.ibm.com/software/data/bigdata/business-partner.html>

NEW QUESTION 74

You are designing storage for a new Hadoop cluster. Which of the following statements is TRUE regarding the usage of SAN or NAS?

- A. SAN or NAS should not be used to set up HDFS
- B. SAN or NAS must be used, if available, to provide backup capabilities
- C. SAN or NAS can be used to support retention policies
- D. SAN or NAS cannot be used if your Hadoop cluster spans 2 sites

Answer: A

Explanation:

References:

<http://www-01.ibm.com/software/data/infosphere/hadoop/hdfs/>

NEW QUESTION 76

Which one of the following statements about Big SQL is NOT TRUE?

- A. BIG SQL supports querying data but doesn't support updates and deletes
- B. BIG SQL is ANSI 2011 Compliant
- C. BIG SQL is supported to work with Cloudera
- D. BIG SQL supports Fluid Query

Answer: A

Explanation:

Reference:

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

NEW QUESTION 79

.....

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