

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

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 3

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 4

Which of the following statements is TRUE?

- A. A good use of the BigInsights is as a query-ready archival system for your data warehouse to quickly access data
- B. BigInsights and Hadoop based systems in general are best used in high concurrency transactional systems
- C. Hadoop map reduce based computing engine is ideal for use as a real-time or near real-time processing extension to your existing business intelligence reporting
- D. The system ML engine is the preferable option to add unstructured text sentiment analysis to the customer service reporting system

Answer: A

Explanation:

References: http://www.ibm.com/support/knowledgecenter/SSPT3X_2.1.2/com.ibm.swg.im.infosphere.biginsights.product.doc/doc/bi_enterprise_architecture.html

NEW QUESTION 5

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 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

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 8

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 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.biginights.analyze.doc/doc/bigsql_performance.html

NEW QUESTION 10

You are building a pricing application combining the output from your internal customer database along with an online feed that contains up to the minute competitor pricing adjustments. Which of the following is the best choice for data level real time output?

- A. Big R
- B. Big SQL
- C. Hadoop native hdfs map reduce based queries
- D. InfoSphere streams

Answer: D

Explanation:

References:

<http://www-03.ibm.com/software/products/en/ibm-streams>

NEW QUESTION 10

SPSS and BigInsights integration is enabled by which of the following?

- A. SPSS Analytic Server
- B. InfoSphere DataStage
- C. SPSS Modeler
- D. SPSS Collaboration & Deployment Services

Answer: A

Explanation:

References:

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

NEW QUESTION 14

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 18

A telecommunication company is experiencing high churn rate. They have more than five million customers. They have more than 400TB of call detail records per year. The majority of their customers are prepaid, so they can change the telecom provider any time. In order to remedy the situation, they want to understand the customer behavior. For that reason they want to build the individual profile of each customer. They also want to enrich the profile by incorporating social media

data. Given this set of requirements, which of the following would you recommend?

- A. Hadoop
- B. RDBMS
- C. Storm
- D. IBM Cloudant

Answer: C

Explanation:

References:

<http://www.ibm.com/developerworks/library/os-twitterstorm/>

NEW QUESTION 21

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_sp ecific/pag e569.html](http://www.ibm.com/developerworks/data/library/cognos/infrastructure/cognos_specific/pag e569.html)

NEW QUESTION 23

Considering a service level requirement (SLR) of less than 3 milliseconds, what task must be performed to meet the SLR?

- A. Collect and analyze SNMP MIB data
- B. Survey a representative amount of end users
- C. Measure switch failure frequency
- D. Inquire when prime network hours occur

Answer: C

NEW QUESTION 25

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 28

A reputable market research firm wants to explore more business opportunities. They have great in house skill in python and machine learning. Their business model is simple, they build the solutions for customers using python and machine learning algorithms and give these solutions to the customer's engineering team for implementation. Given this scenario, which of the following would you recommend?

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

Answer: D

NEW QUESTION 31

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 35

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 36

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 39

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 43

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 47

Which of the following statements is TRUE regarding cloud based solutions?

- A. In a Platform as a Service Cloud deployment, the customer chooses the operating system they want to use
- B. Automated recovery from hardware or network failures is not possible in a public cloud implementation, only in a private clouds
- C. There are benefits to use the cloud even for small-scale applications
- D. Using firewalls to create network boundaries is sufficient for ensuring cloud security

Answer: C

Explanation:

References:

<http://www.ibm.com/developerworks/cloud/library/cl-cloudappdevelop/>

NEW QUESTION 51

Which of the following requirements would NOT be effectively addressed by a NoSQL data store?

- A. Scalability
- B. Reporting
- C. Sparse data
- D. Batch processing

Answer: D

NEW QUESTION 55

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 60

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.doc/bl1adv_clones.htm

NEW QUESTION 61

You are consulting with a company to design a new Big Data System. The CIO insists that all components must be open source. The key requirement is to provide text search on their data. For audit purposes the infrastructure must also be monitored and an alert system in place. To meet these requirements, which of the following would you suggest?

- A. HBase together with OPTIM performance Manager
- B. Text Analytics with OPTIM performance Manager
- C. Nagios with Lucene
- D. Oozie with Lucene

Answer: C

Explanation:

References: <https://www.nagios.org/> https://www.ibm.com/support/knowledgecenter/SSPT3X_2.1.2/com.ibm.swg.im.infosphere.biginsights.product.doc/doc/bi_addtl_technologies.html

NEW QUESTION 64

A large global enterprise customer has a Big Data environment set up on Hadoop. After a year in operation they are now looking to extend access to multiple functions that will need different views into different aspects/portions of the data. As you consider these requirements, which of the following statements is TRUE and also applies to the scenario?

- A. Hadoop does not support multi tenancy but can easily scale to support this by replicating data to new clusters with commodity hardware.
- B. Hadoop can support multi tenancy but only if YARN is used, so if not already used, the customer will need to upgrade to a YARN supported version.
- C. Virtualization methods can be used to support multi-tenancy on Hadoop because one can easily replicate to additional cluster whenever there is a need.
- D. Hadoop can support multi tenancy by using a cluster file system for storage, allowing all nodes to access the data.

Answer: D

Explanation:

Reference:

http://www-01.ibm.com/support/knowledgecenter/STXKQY/411/com.ibm.spectrum.scale.v4r1_1.adv.doc/bl1adv_hadoop.htm?lang=en

NEW QUESTION 69

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

NEW QUESTION 72

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 76

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 79

In BigSheets you can add sheets to workbooks to progressively edit and explore your data. Which of the following is a CORRECT list of types of sheets that BigSheets provides and can contain predefined logic for analyzing data?

- A. Copy, Formula, Update
- B. Filter, Insert, Union
- C. Distinct, Join, Stored Procedure
- D. Complement, Limit, Group

Answer: D

Explanation:

References:

<https://developer.ibm.com/hadoop/blog/2014/08/18/use-bigsheets-analytics/>

NEW QUESTION 81

Which of the following is NOT a valid Service Level Agreement (SLA) metric?

- A. Mean time between failures
- B. Mean time to repair
- C. Identification to responsible party
- D. Identification of failing component

Answer: D

Explanation:

References:

https://en.wikipedia.org/wiki/Service-level_agreement

NEW QUESTION 83

A large application vendor wants to port their existing distributed applications to run on Hadoop. In order to be competitive they need to provide monitoring and keep the size of the monitored applications consistent with the configuration. This implies the ability to deploy a replacement, for example, for any failed components. Which of the following would be a workable solution?

- A. Nagios with YARN
- B. Slider with YARN
- C. Oozie with Lucene
- D. OPTIM Performance Manager

Answer: B

NEW QUESTION 88

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 92

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

NEW QUESTION 93

In order to enable connectivity between Netezza and BigInsights, what must be enabled on the Netezza Performance Server?

- A. Java-based Netezza user-defined functions
- B. ODBC drivers from Netezza
- C. JDBC drivers from Netezza
- D. JAQL Netezza connectors

Answer: A

Explanation:

Reference:

https://www-01.ibm.com/support/knowledgecenter/SSPT3X_4.1.0/com.ibm.swg.im.infosphere.biginsight.s.import.doc/doc/data_warehouse_udf_netezza.html

NEW QUESTION 95

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 96

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 99

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 103

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-llwant-to-know.html>

NEW QUESTION 105

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 109

Which of the following big data components makes all decisions regarding the replication of blocks?

- A. Job Tracker
- B. Edge Node
- C. Name Node
- D. Data Node

Answer: A

NEW QUESTION 113

You have implemented a large Hadoop MapReduce cluster and the applications and users are multiplying. You are now faced with requests for interactive and streaming data applications while you still need to support the original MapReduce batch Jobs. Select the best option for continued support and performance.

- A. Just add several data nodes as Hadoop clusters are designed to scale-up easily
- B. Keep your original cluster configuration, all that is needed is re-optimizing the Oozie- workflow management
- C. Implement Yarn to decouple MapReduce and resource management
- D. Implement Apache Cassandra to automatically optimize multi-tenancy workloads

Answer: D

NEW QUESTION 118

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 121

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.biginsights.bigr.doc/doc/intro.html

NEW QUESTION 122

Which of the following approaches can an organization take to solve the impact on the performance and capacity originated by the variety of data types?

- A. Define a data catalog in a traditional data warehouse
- B. Create different solutions to handle every kind of data
- C. Store a wide range of data formats on the same platform
- D. Define a comprehensive taxonomy and constantly review

Answer: D

Explanation:

Reference:

<http://www.redbooks.ibm.com/redpapers/pdfs/redp5070.pdf> Page: 15

NEW QUESTION 123

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 126

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 127

.....

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