

C2090-102.exam

Number: C2090-102
Passing Score: 800
Time Limit: 120 min



<http://www.gratisexam.com/>

IBM

C2090-102

IBM Big Data Architect

<http://www.gratisexam.com/>

Exam

QUESTION 1

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

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

References:

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

QUESTION 2

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

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

QUESTION 3

Which of the following is true about Cloudant?

- A. It is a byte-value store
- B. It is available as a database-as-a-service (DBaaS)
- C. It runs remotely in the cloud and not on a local system
- D. It offers SQL access to data

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

References:

<http://www-01.ibm.com/software/data/cloudant/>

QUESTION 4

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



<http://www.gratisexam.com/>

- 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

Correct Answer: C

Section: (none)

Explanation

Explanation/Reference:

References:

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

QUESTION 5

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

<http://www.gratisexam.com/>

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

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

References:

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

QUESTION 6

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

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

References:

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

QUESTION 7

A financial company stores all enterprise data for five years in their Teradata data warehouse that costs them expensive storage. The company started an initiative to move historical data to a Hadoop environment and uses federation method to access both sets of data. What would be the IBM Big Data value proposition for this use case?

- A. IBM Logical Data Warehouse and IBM Big SQL
- B. Enterprise Data Warehouse
- C. Pure Data for Analytics
- D. InfoSphere Information Server

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

QUESTION 8

Company A has decided to implement a new data system to support their rapidly growing business. They have an existing 20 TB worth of raw data, with an expected weekly incoming rate of 50 GB of new raw data. The data is mostly text based and unstructured. A typical query can involve pulling in 10 GB of data. Historically, performance has been an issue and currently needs to be addressed. Which of the following would you suggest to support these requirements?



<http://www.gratisexam.com/>

- A. Set up a Hadoop system with commodity HW for scalability
- B. Utilize de-duplication and compression technology
- C. Use a mixture of different disk-types to provide hot/cold storage
- D. Create range partitions for the data

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

QUESTION 9

The analysis layer reads the data digested by the layer massaging and store layer. In some cases, the analysis layer accesses the data directly from the data source. Designing the analysis layer requires careful forethought and planning. Decisions must be made with regard to how to manage the tasks to do which of the following?

- A. Reduce the volume of data
- B. Locate the data sources that can provide data for the required entities

<http://www.gratisexam.com/>

- C. Identify multi-dimension data sources
- D. Identify data import and export tools to adopt and deliver an ETL strategy

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

References:

<http://www.ibm.com/developerworks/library/bd-archpatterns3/>

QUESTION 10

A component of IBM Industry Model forms the basis of the “Logical Data Warehouse Model” that spans across the traditional RDBMS and Hadoop technology. It defines all of the data structures that would be expected to be defined in the Detailed System of Record. What is the name of this component?

- A. Business data model
- B. Atomic warehouse model
- C. Dimensional warehouse models
- D. Metadata management

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

References:

http://www.ibm.com/support/knowledgecenter/SS9NBR_9.1.0/com.ibm.ima.using/comp/bdm/intro.dita

QUESTION 11

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

Correct Answer: D

Section: (none)

Explanation

Explanation/Reference:

References:

https://www.ibm.com/support/knowledgecenter/SSPT3X_3.0.0/com.ibm.swg.im.infosphere.biginsights.bigr.doc/doc/intro.html**QUESTION 12**

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

Correct Answer: D**Section: (none)****Explanation****Explanation/Reference:**

References:

<http://www-03.ibm.com/software/products/en/ibm-streams>**QUESTION 13**

Which data format stores all of the data in a binary format making the files more compact, and will even add in markers to help Map Reduce jobs determine where to break large files for more efficient processing?

<http://www.gratisexam.com/>

- A. Parquet
- B. Avro
- C. ORC
- D. Sequence File

<http://www.gratisexam.com/>

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

References:

<http://www.ibmbigdatahub.com/blog/how-succeed-using-avro-storage-format>

QUESTION 14

Which of the following signifies evidence of congestion in the network?

- A. Packet Discards
- B. Quantity of warning in the user system logs
- C. User complaints
- D. Traffic volumes and rates

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

QUESTION 15

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

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

QUESTION 16

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

Correct Answer: C

Section: (none)

Explanation

Explanation/Reference:

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

QUESTION 17

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

Correct Answer: D

Section: (none)

Explanation

Explanation/Reference:

QUESTION 18

What is the effective meaning of “NoSQL”?

- A. It does not permit the use of SQL
- B. It is not limited to relational database technology
- C. It does not have tables or schemas
- D. It does not permit UPDATE

Correct Answer: C

Section: (none)

Explanation

Explanation/Reference:

References:

<http://www.sitepoint.com/sql-vs-nosql-differences/>

QUESTION 19

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?



<http://www.gratisexam.com/>

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

Correct Answer: C

Section: (none)

Explanation

Explanation/Reference:

QUESTION 20

You need to provision a Hadoop cluster to perform data analysis on customer sales data to predict which products are more popular. Which of the following solutions will let you set up your cluster with the most stability in the platform?

- A. Purchase specific products from multiple Independent Software Vendors (ISV) for your requirements in order to take advantage of vendor-specific features
- B. Develop your own platform of software components to allow for maximum customization
- C. Use a Hybrid of ISV applications to build your customizations on top of that
- D. Leverage the Open Data Platform (ODP) core to provide a stable base against which Big Data solutions providers can qualify solutions

Correct Answer: D

Section: (none)

Explanation

Explanation/Reference:

References:

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

QUESTION 21

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

Correct Answer: D

Section: (none)

Explanation

Explanation/Reference:

References:

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

QUESTION 22

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

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

QUESTION 23

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

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

QUESTION 24

The downside of cloud computing, relative to SLAs, is the difficulty in determining which of the following?

- A. Root cause for service interruptions
- B. Turn-Around-Time (TAT)
- C. Mean Time To Recover (MTTR)
- D. First Call Resolution (FCR)

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

References:

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

QUESTION 25

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?



<http://www.gratisexam.com/>

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

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

QUESTION 26

You need to create an online repository for a company. Data includes pdf, documents, html, images, etc. Total data volume is approximately 1PB. Online repository must be highly available. You are required to propose the least costly solution. Which technology would be preferred for this requirement?

- A. Spark
- B. RDBMS
- C. Hadoop
- D. IBM Infosphere Streams

Correct Answer: C

Section: (none)

Explanation

<http://www.gratisexam.com/>

Explanation/Reference:

QUESTION 27

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

Correct Answer: D

Section: (none)

Explanation

Explanation/Reference:

References:

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

QUESTION 28

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

Correct Answer: D

Section: (none)

Explanation

Explanation/Reference:

References:

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

QUESTION 29

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

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

Correct Answer: C

Section: (none)

Explanation

Explanation/Reference:

QUESTION 30

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

Correct Answer: D

Section: (none)

Explanation

Explanation/Reference:

QUESTION 31

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

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

QUESTION 32

A manufacturing company has decided they need to capture and analyze the log files of their software automation system. Their business users are still trying to define the use cases but would want to start capturing as they have had frequent outages. Given this, which of the following is the best software design recommendation?

- A. ETL tools and a Data Warehouse
- B. Flume and Hadoop
- C. Pure Data for Analytics and Optim
- D. Streams and BigInsights

Correct Answer: D

Section: (none)

Explanation

Explanation/Reference:

QUESTION 33

Company A is searching for a browser-based visualization tool to perform analysis on vast amounts of data in any structure. They want to execute operations such as pivot, slice and dice, among others. Which of the following would meet these requirements?

- A. Streams
- B. BigSheets
- C. Aginity Workbench
- D. Watson Explorer

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

References:

<http://www.dotgroup.co.uk/wp-content/uploads/2014/11/Harness-the-Power-of-Big-Data-The-IBM-Big-Data-Platform.pdf> Page: 132

QUESTION 34

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

A. Component Relationship Diagram



<http://www.gratisexam.com/>

B. Component Interface Diagram

C. Component Interaction Diagram

D. Component Reaction Diagram

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

QUESTION 35

Data in motion is data that is constantly appended with new information. This type of data can be imported into the distributed file system by using which of the following?

A. Scoop

B. Flume

C. Big SQL LOAD

D. Hadoop shell commands

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

References:

http://www.ibm.com/support/knowledgecenter/SSPT3X_2.1.2/com.ibm.swg.im.infosphere.biginsights.import.doc/doc/data_motion.html

QUESTION 36

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

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

QUESTION 37

A media company wants to enhance their subscription services to their end customers by incorporating Twitter, Wikis, Blogs, and Analysts data. The company likes to provide a real-time dashboard view for news articles, press releases, and analysts perspective to end customers based on their subscriptions. What solution blueprint would offer the best fit for this critical business requirement?

- A. Watson Analytics
- B. Big Insights, Spark, and Big SQL
- C. Pure Data for Analytics PDA, and SPSS
- D. InfoSphere Streams, BigInsights, and Watson Explorer

Correct Answer: D

Section: (none)

Explanation

Explanation/Reference:

QUESTION 38

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

Correct Answer: C

Section: (none)

Explanation

Explanation/Reference:

References:

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

QUESTION 39

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

Correct Answer: D

Section: (none)

Explanation

Explanation/Reference:

References:

https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html

QUESTION 40

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

Correct Answer: D

Section: (none)
Explanation

Explanation/Reference:



<http://www.gratisexam.com/>