IBM

Exam Questions C2090-102

IBM Big Data Architect

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 statements is TRUE?

A. A good use of the BigInsights is as a query-ready archival system for your data warehouse to quickly accessdata
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 processingextension to your existing business intelligence reporting
D. The system ML engine is the preferable option to add unstructured text sentiment analysis to the customerservice 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 3
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 4
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 5
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.biginsight s.analyze.doc/doc/bigsql_performance.html

NEW QUESTION 6
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 7
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

Answer: B

Explanation:
References:
http://www-01.ibm.com/software/data/cloudant/

NEW QUESTION 8
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/fileserve?contentid=239170

NEW QUESTION 9
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?

A. Hadoop
B. Streams
C. Unica
D. Pure Data for Analytics

Answer: C

NEW QUESTION 10
Which of the following statements regarding SPSS is TRUE?

A. SPSS software provides a security framework
B. SPSS analytics are primarily accessed through a scripting language
C. SPSS models can only be applied to data at rest
D. SPSS can directly use BigR syntax such as "bigr.list"

Answer: A

Explanation:
Reference:
http://www.ibmbigdatahub.com/blog/7-really-good-reasons-partner-spss-analytics-
r

NEW QUESTION 10
A media company collects customer behavior data, such as how frequently they tune in, specific viewing habits, and peak usage in real time, in order to improve their services.The company likes to segmentits customers for advertisers by correlating viewing habits with public data, such as voter registration, in order to launch highly targeted campaigns to specific demographics. What technology should their Data Architect consider?

A. InfoSphere Streams, BigInsights, and Pure Data for Analytics PDA
B. BigInsights and Pure Data for Operational Analytics PDOA
C. InfoSphere Streams, Spark, and BigR
D. PureData for Analytics and SPSS

Answer: D

Explanation:
‒Reference:
http://www.ibm.com/software/data/puredata/analytics/nztechnology/analytics.html

NEW QUESTION 13
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

Answer: C

Explanation:
References:
http://www.sitepoint.com/sql-vs-nosql-differences/

NEW QUESTION 16
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 17
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

Answer: A

NEW QUESTION 21
For company B, 85% of their analytics queries only involve about 25% their data; another 10% of the queries will touch 35% of the rest of the data, and only 5% of the queries will touch the remaining 40% of the data. The estimated volume is 50TB growing at 1 TB per year. Which of the following would provide the best value (business benefit) and lowest TCO?

A. Place the entire set of data in a data warehouse with proper partitioning and indexing
B. Place the entire set of data in a hadoop environment – using commodity HW
C. Place the top 25% of data (used by 85% of the query) in a hadoop environment, and the rest in a data warehouse
D. Place the top 25% of data (used by 85% of the query) in a data warehouse, and the rest in a hadoop environment

Answer: C

NEW QUESTION 24
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 inorder 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 solutionsproviders can qualify solutions

Answer: D

Explanation:
References:
http://hortonworks.com/innovation/open-data-platform/

NEW QUESTION 29
Which of the following statements is TRUE regarding cloud applications?

A. Migrating a legacy application to the cloud is a simple solution to drive down cost
B. Architecting and deploying a scalable cloud application requires a private cloud implementation
C. To be truly elastic, a cloud application must be tightly coupled to the physical cloud environment
D. Leveraging a private v
E. public cloud may result in sacrificing some of the core advantages of cloud computing

Answer: B

NEW QUESTION 30
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 32
GPFS is recommended when additional levels of control are required. Which of the following requirements can only be supported by GPFS?

A. There is a requirement where logical isolation and physical isolation need to be supported
B. There is a requirement for separate clusters for analytics and databases
C. There is a need to support placing a MapReduce split on one local disk
D. There is a need to run the NameNode in a separate high availability environment

Answer: A

Explanation:
References: https://www.ibm.com/support/knowledgecenter/SSPT3X_3.0.0/com.ibm.swg.im.infosphere.biginsights.product.doc/doc/over_filesystem_comparison.html

NEW QUESTION 33
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?

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

Answer: B

Explanation:
References:
http://www.ibmbigdatahub.com/blog/how-succeed-using-avro-storage-format

NEW QUESTION 36
You have a need for Storm real time processing and you realize that your Storm processing is detrimental to the timely execution of your MapReduce batch jobs. Which of the following would be your best course of action?

A. Implement a Storm-YARN integration to facilitate the management of elastic workloads
B. Implement the Oozie 2.0 framework optimized for elastic workload management
C. Use a combination of Flume and Oozie 2.0 to enable the Flume built-in elastic flow automation
D. Implement Apache ACE 2.0 for Storm

Answer: C

Explanation:
Reference:
http://www.ibm.com/developerworks/library/bd-ooziehadoop/

NEW QUESTION 41
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 willautomatically loop through each item inthe 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 43
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 44
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

Answer: A

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

NEW QUESTION 46
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 49
Which of the following statements is TRUE about BigSheets?

A. It is a tool that enables line-of-business users to harness the power of Hadoop using a familiar spreadsheet interface
B. It is a browser-based visualization and analysis tool designed to help programmers work with Big Data
C. It is a component of Pure Data for Analytics
D. It is a browser-based visualization and analysis tool included in Streams

Answer: C

Explanation:
Reference:
http://www-01.ibm.com/software/data/puredata/analytics/

NEW QUESTION 51
......

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questons and Answers in PDF Format

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

100% Actual & Verified — Instant Download, Please Click
Order The C2090-102 Practice Test Here