

C2090-102 Dumps

IBM Big Data Architect

<https://www.certleader.com/C2090-102-dumps.html>



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

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 3

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 4

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 5

A smart meter project is being undertaken by an Electric utility headquartered in San Francisco. They are looking for technology for monitoring thousands of meters, multiple times within an hour, and performing corresponding downstream advisory action to influence usage behavior. Which of the following would you recommend to function at the heart of this process?

- A. Hadoop
- B. SPSS
- C. Reporting engine and a portal
- D. Infosphere Streams

Answer: B

NEW QUESTION 6

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 7

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/page569.html

NEW QUESTION 8

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 9

Big data is often defined as the ability to derive new insights from data that has scaled up along three axes known as the three v's. Which of the following is the fourth v? (Hint: It has something to do with the uncertainty.)

- A. volume
- B. variety
- C. velocity
- D. veracity

Answer: C

NEW QUESTION 10

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 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 segment its 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 12

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 13

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 16

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 19

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 23

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

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 29

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 32

For company B, 85% of their analytics queries only involve about 25% of 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 36

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 37

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 39

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

Answer: C

NEW QUESTION 41

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 45

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

Answer: B

Explanation:

References:

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

NEW QUESTION 48

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 52

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 53

.....

Thank You for Trying Our Product

* 100% Pass or Money Back

All our products come with a 90-day Money Back Guarantee.

* One year free update

You can enjoy free update one year. 24x7 online support.

* Trusted by Millions

We currently serve more than 30,000,000 customers.

* Shop Securely

All transactions are protected by VeriSign!

100% Pass Your C2090-102 Exam with Our Prep Materials Via below:

<https://www.certleader.com/C2090-102-dumps.html>