C2090-102 Dumps
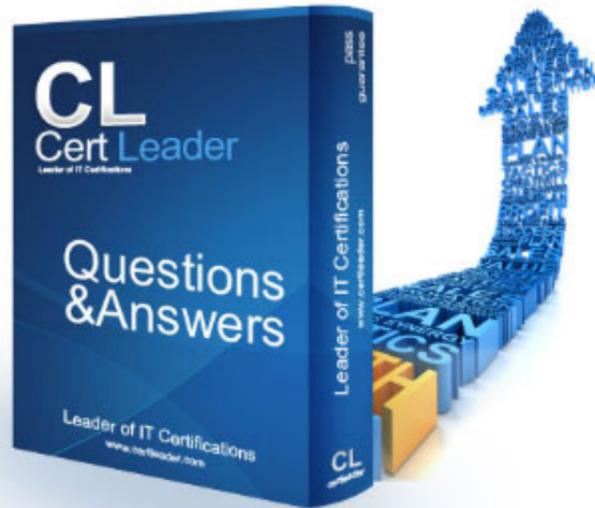
IBM Big Data Architect

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

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
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
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 6
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 7
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 8
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 9
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 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/fileserve?contentid=239170

NEW QUESTION 10
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 11
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

Answer: B

NEW QUESTION 16
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 18
Which of the following is NOT a typical Service Level Agreement (SLA) metric?

A. Abandonment rate
B. Average speed to answer
C. Time service factor
D. Average response time

Answer: D

NEW QUESTION 22
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-bigsql/

NEW QUESTION 25
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 27
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 31
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 35
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 38
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 40
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

Answer: D

NEW QUESTION 44
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 45
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 49
What is the most important aspect of data center disaster recovery?

A. A complete damage assessment
B. Control of critical assets
C. Restoration of business functions
D. Protection of individual life

Answer: D

NEW QUESTION 51
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 54
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 systemnamespace and file Blockmap in memor
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 56
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.ht ml

NEW QUESTION 59
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.do c/bl1adv_hadoop.htm?lang=en

NEW QUESTION 64
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 69
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 74
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 75
Which architecture document is used to help organize projects, manage the complexity of the solution, and ensurethat all architecturerequirements have been addressed?

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

Answer: B

NEW QUESTION 80
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 83
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 84
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 87
Which of the following is NOT an example of unstructured data?

A. HBase table
B. Tweet
C. Netezza table
D. Internet Protocol Detail Record

Answer: C

Explanation:
_References: https://www.ibm.com/support/knowledgecenter/SSULQD_7.2.0/com.ibm.nz.gsg.d oc/c_gets trt_nzsql_help.html

NEW QUESTION 91
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 92
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 96
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 Serviceand 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 andfail over capabilities

Answer: A

NEW QUESTION 97
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 98
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 99
......

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