

C2090-102 Dumps

IBM Big Data Architect

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



NEW QUESTION 1

What are the two levels documented in the Operational Model? (Choose two.)

- A. Logical
- B. Rational
- C. Theoretical
- D. Physical
- E. Middleware

Answer: AC

NEW QUESTION 2

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)

Answer: A

Explanation:

References:

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

NEW QUESTION 3

“The programming model for client developers will hide the complexity of interfacing to legacy systems” is an example of which of the following?

- A. A use case
- B. An architectural decision
- C. A client imperative
- D. An empathy statement

Answer: B

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.biginsights.analyze.doc/doc/bigsql_performance.html

NEW QUESTION 6

Which of the following statements regarding Big R is TRUE?

- A. Unless specified otherwise, Big R automatically assumes all data to be integers
- B. Big R's 'bigr.frame' is equivalent to R's 'data.frames'
- C. When you execute Big R “apply” function, Big R transparently extracts data out of HDFS into the Big R engine
- D. A data analyst using Big R employs Map Reduce programming principles

Answer: A

Explanation:

Reference:

<https://developer.ibm.com/hadoop/docs/biginsights-value-add/big-r/bigr-tutorial/>

NEW QUESTION 7

The analysis layer reads the data digested by the layer messaging 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
- C. Identify multi-dimension data sources
- D. Identify data import and export tools to adopt and deliver an ETL strategy

Answer: B

Explanation:

References:

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

NEW QUESTION 8

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 9

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_specific/pag e569.html

NEW QUESTION 10

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 10

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

- A. Component Relationship Diagram
- B. Component Interface Diagram
- C. Component Interaction Diagram
- D. Component Reaction Diagram

Answer: A

NEW QUESTION 11

A large Retailer (online and “brick & mortar”) processes data for analyzing marketing campaigns for their loyalty club members. The current process takes weeks for processing only 10% of social data. What is the most costeffective platform for processing and analyzing campaign results from social data on a daily basis using 100% dataset?

- A. Enterprise Data Warehouse
- B. BigInsights Open Data Platform
- C. High Speed Mainframe Processing

D. In Memory Computing

Answer: B

Explanation:

References: <http://www.ibm.com/developerworks/data/library/techarticle/dm-1110biginsightsintro/>

NEW QUESTION 12

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 15

By default, Parquet uses which of the following codecs?

- A. SNAPPY
- B. LZO
- C. GZIP
- D. BZIP2

Answer: A

NEW QUESTION 18

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 19

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 21

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 24

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 26

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 28

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 29

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 30

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 33

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?

- 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

Answer: A

NEW QUESTION 35

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 36

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/c/bl1adv_hadoop.htm?lang=en

NEW QUESTION 38

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 39

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 44

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 47

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

Answer: B

Explanation:

References:

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

NEW QUESTION 49

Which architecture document is used to help organize projects, manage the complexity of the solution, and ensure that all architecture requirements have been addressed?

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

Answer: B

NEW QUESTION 52

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 54

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 57

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 59

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.biginsights.import.doc/doc/data_warehouse_udf_netezza.html

NEW QUESTION 64

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 67

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.doc/c_gets_trt_nzsql_help.html

NEW QUESTION 69

Which of the following is a key consideration when determining the potential use of Models in Big Data Deployments?

- A. Stability of schema
- B. Level of Information Processing
- C. Ownership/Governance
- D. All of the above

Answer: B

NEW QUESTION 73

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 78

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 83

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 87

.....

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>