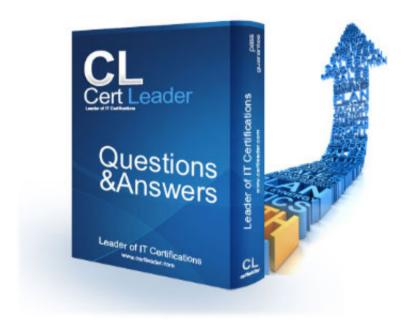# C2090-102 Dumps

IBM Big Data Architect

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

NEW QUESTION 1
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 2
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 3
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 4
A smart meter project is being undertaken by anElectric utility headquartered in San Francisco. They are looking for technology for monitoring thousands of meters, multiple times within an hour, andperforming 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 5
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 6
SPSS and BigInsights integration is enabled by which of the following?

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

Answer: A

Explanation:
References:
http://www.ibm.com/developerworks/library/bd-spss/

NEW QUESTION 7
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 8
It's helpful to look at the characteristics of big data along certain lines – for example, how the data is collected, analyzed and processed. There are many characteristics to consider. Which one of the following is NOT a characteristic that should be considered?

A. Data frequency and size
B. Software
C. Data source
D. Processing methodology

Answer: B

Explanation:
References:
http://www.ibm.com/developerworks/library/bd-archpatterns1/

NEW QUESTION 9
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 10
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/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 14
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 19
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 21
By default, Parquet uses which of the following codecs?

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

Answer: A

NEW QUESTION 22
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 27
The AQL query language is the easiest and most flexible tool to pull structured output from which of the following?

A. Hive data structures
B. Unstructured text
C. Hbase schemas
D. JDBC connected relational data marts

Answer: A

Explanation:
Reference:
http://www.ibm.com/developerworks/library/bd-sqltohadoop2/

NEW QUESTION 28
As you explore the data for a BigSheets workbook, you must run the workbook against the full data set to get the most current results for analysis. Which statement is TRUE regarding running and visualizing data in a workbook?

A. You can create graphs for more than one sheet within the same workbook
B. By default, the first sheet in your workbook is named the Results sheet
C. When you save and run the workbook, the data in a Child Workbook is theoutput for that workbook
D. When you add sheets to workbooks, saving the sheets runs the individual data for the sheet but not for the full workbook

Answer: C

Explanation:
Reference:
https://www- 01.ibm.com/support/knowledgecenter/SSPT3X_4.1.0/com.ibm.swg.im.infosphere.biginsight s.analyze.doc/doc/bigsheets_con_workbooks.html

NEW QUESTION 30
Which of the following statements is TRUE regarding cloud based solutions?

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, onlyin 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

Answer: C

Explanation:
References:
http://www.ibm.com/developerworks/cloud/library/cl-cloudappdevelop/

NEW QUESTION 32
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 34
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 37
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 41
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 43
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 48
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 53
A large application vendor wants to port their existing distributed applications to run on Hadoop. In order to be competitive they need to provide monitoring and keep the size of the monitored applications consistent with the configuration. This implies the ability to deploy a replacement, for example, for any failed components. Which of the following would be a workable solution?

A. Nagios with YARN
B. Slider with YARN
C. Oozie with Lucene
D. OPTIM Performance Manager

Answer: B

NEW QUESTION 58
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_comparis on.html

NEW QUESTION 59
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 61
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. biginsight s.import.doc/doc/data_warehouse_udf_netezza.html

NEW QUESTION 64
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 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.d oc/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 72
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?

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

Answer: B

NEW QUESTION 75
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 77
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 80
Which of the following does NOT consist of collaboration from various organizations to drive innovation and standardization across big data technologies?

A. Cloudera Enterprise
B. Hortonworks Data Platform
C. IBM BigInsights
D. Pivotal Big Data Suite

Answer: C

Explanation:
Reference:
http://www-01.ibm.com/software/data/bigdata/business-partner.html

NEW QUESTION 84
You are asked to provide deep analytics from existing historical data to score new
high volume incoming live data with low latency in real time. Which of the following would you suggest?

A. SPSS Modeler with InfoSphere Streams
B. Info Sphere DataStage with InfoSphere Streams
C. Hive with MapReduce
D. Cassandra with YARN

Answer: B

NEW QUESTION 88
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 90
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 95
......

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