

Exam Questions C2090-102

IBM Big Data Architect

<https://www.2passeasy.com/dumps/C2090-102/>



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

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 access data
- 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 processing extension to your existing business intelligence reporting
- D. The system ML engine is the preferable option to add unstructured text sentiment analysis to the customer service 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

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 4

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 5

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 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/fileservice?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

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 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

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 13

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 15

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 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 23

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 26

By default, Parquet uses which of the following codecs?

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

Answer: A

NEW QUESTION 27

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 28

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 31

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-sqltohadop2/>

NEW QUESTION 35

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 40

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 45

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 the output 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.biginights.analyze.doc/doc/bigsheets_con_workbooks.html

NEW QUESTION 46

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 50

A media company wants to enhance their subscription services to their end customers by incorporating Twitter, Wikis, Blogs, and Analysts data. The company likes to provide a real-time dashboard view for news articles, press releases, and analysts perspective to end customers based on their subscriptions. What solution blueprint would offer the best fit for this critical business requirement?

- A. Watson Analytics

- B. Big Insights, Spark, and Big SQL
- C. Pure Data for Analytics PDA, and SPSS
- D. InfoSphere Streams, BigInsights, and Watson Explorer

Answer: D

NEW QUESTION 52

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 53

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 58

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 60

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 62

The NameNode determines the rack id to which each DataNode belongs via the process outlined in which of the following?

- A. HDFS APIs
- B. Hadoop Rack Awareness
- C. Job Tracker
- D. NameNode High Availability

Answer: D

NEW QUESTION 64

Which of the following statements is TRUE regarding cloud computing solutions?

- A. Cloud security is planned, developed, and layered on top of an application after the application development process is complete
- B. Stateless applications are better candidates for cloud services than applications that maintain state
- C. Cloud solutions rely on scaling up (vertical) scaling v
- D. scale out (horizontal) scaling
- E. Server virtualization is a requirement in a cloud implementation

Answer: D

NEW QUESTION 66

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 67

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 70

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 74

Which of the following is a consideration in sizing an active archive Hadoop infrastructure?

- A. replication factor within Hadoop
- B. Reporting requirements
- C. velocity or rate at which data is being generated
- D. veracity or trustworthiness of the data

Answer: B

Explanation:

Reference:

<http://www.ibm.com/developerworks/library/ba-augment-data-warehouse3/>

NEW QUESTION 78

The YARN High Availability feature adds redundancy in the form of an Active/Standby. Which of the following will pair to remove this otherwise single point of failure?

- A. JobTracker
- B. Data Node
- C. Management Node
- D. Resource Manager

Answer: D

Explanation:

References:

<http://hadoop.apache.org/docs/current/hadoop-yarn/hadoop-yarn-site/ResourceManagerHA.html>

NEW QUESTION 83

Fine grain encryption and security for a Big Data engagement can be enabled and implemented by various IBM, ISV, and open source software products. Which of the following is NOT a security specific software product?

- A. BigInsights

- B. Guardium
- C. OSSEC
- D. Shiro

Answer: A

NEW QUESTION 86

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 89

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 93

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 98

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 99

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 100

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 105

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 110

Which of the following big data components makes all decisions regarding the replication of blocks?

- A. Job Tracker
- B. Edge Node
- C. Name Node
- D. Data Node

Answer: A

NEW QUESTION 114

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 119

Which of the following statements regarding Big R is TRUE?

- A. The Big R API is completely identical to the R API
- B. Big R implements an interpreted computer language
- C. Big R users cannot access Big R capabilities from the standard RStudio client as RStudio is not Hadoopenabled
- D. Big R utilizes the big SQL query engine for processing

Answer: D

Explanation:

_References: https://www.ibm.com/support/knowledgecenter/SSPT3X_3.0.0/com.ibm.swg.im.infosphere.biginsights.bigr.doc/doc/intro.html

NEW QUESTION 123

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 124

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 129

Which one of the following statements about Big SQL is NOT TRUE?

- A. BIG SQL supports querying data but doesn't support updates and deletes
- B. BIG SQL is ANSI 2011 Compliant
- C. BIG SQL is supported to work with Cloudera
- D. BIG SQL supports Fluid Query

Answer: A

Explanation:

Reference:

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

NEW QUESTION 132

.....

THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual C2090-102 Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the C2090-102 Product From:

<https://www.2passeasy.com/dumps/C2090-102/>

Money Back Guarantee

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