Exam Questions C2090-102

IBM Big Data Architect

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

NEW QUESTION 1
"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 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
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 5
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 6
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 7
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 8
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 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
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 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 14
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 17
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 18
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 22
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 Mainfraime Processing
D. In Memory Computing

Answer: B

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

NEW QUESTION 25
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 29
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 32
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 37
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 42
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 47
In computer science, which of the following represents a set of properties that guarantee that database transactions are processed reliably?

A. Foreign Key
B. Primary Key
C. Buffer Cache
D. ACID

Answer: D

Explanation:
References:
https://en.wikipedia.org/wiki/ACID

NEW QUESTION 51
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 56
You need to provision a Hadoop cluster to perform data analysis on customer sales data to predict which products are more popular. Which of the following solutions will let you set up your cluster with the most stability in the platform?

A. Purchase specific products from multiple Independent Software Vendors (ISV) for your requirements inorder to take advantage of vendor-specific features
B. Develop your own platform of software components to allow for maximum customization
C. Use a Hybrid of ISV applications to build your customizations on top of that
D. Leverage the Open Data Platform (ODP) core to provide a stable base against which Big Data solutionsproviders can qualify solutions

Answer: D

Explanation:
References:
http://hortonworks.com/innovation/open-data-platform/

NEW QUESTION 57
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 62
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 66
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 70
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 71
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 76
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 81
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 83
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 88
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 91
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 95
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 100
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 105
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 108
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 110
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 111
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 116
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 120
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 121
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 122
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 125
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 128
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.i nfosphere. biginsights.bigr.doc/ doc/intro.html

NEW QUESTION 132
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 133
A financial company stores all enterprise data for five years in their Teradata data warehouse that costs them expensive storage. The company started an initiative to move historical data to a Hadoop environment and uses federation method to access both sets of data. What would be the IBM Big Data value proposition for

this use case?

A. IBM Logical Data Warehouse and IBM Big SQL
B. Enterprise Data Warehouse
C. Pure Data for Analytics
D. InfoSphere Information Server

Answer: A

NEW QUESTION 135
......

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