Exam Questions C2090-102

IBM Big Data Architect

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

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
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 accessdata
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 processingextension to your existing business intelligence reporting
D. The system ML engine is the preferable option to add unstructured text sentiment analysis to the customerservice 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 4
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 5
Which of the following can NOT be performed by the Open Data Platform (ODP)?

A. Provide development and governance of upstream projects
B. Contribute to the Apache Software Foundation projects
C. Produce a set of tools and methods that enable members to create and test differentiated offerings based on the ODP core
D. Accelerate the delivery of Big Data solutions by providing a well-defined core consisting of all the components of the Apache Software Foundation projects

Answer: C

NEW QUESTION 6
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 7
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 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
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 15
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 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 23
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 26
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 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 34
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 38
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 40
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 segmentits 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 42
By default, Parquet uses which of the following codecs?

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

Answer: A

NEW QUESTION 47
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 49
Which of the following is a requirement for data retention and archival?

A. A format and storage repository for archived data
B. Public cloud
C. Hosting location
D. Solid-state technology

Answer: A

NEW QUESTION 54
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 58
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 61
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 65
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 69
For company B, 85% of their analytics queries only involve about 25% 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 72
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 74
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 76
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 78
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 81
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 83
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 84
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 88
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 93
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 96
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 99
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 104
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 109
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 114
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 119
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 120
BigInsights is a solution that accomplishes which of the following?

A. Replaces the traditional Data warehouses
B. Can exchange information with the traditional Data warehouses only
C. Includes a connector that enables data exchange between a BigInsights cluster and Netezza appliance in only one way
D. Supports data exchange with a number of sources

Answer: C

Explanation:
Reference:
https://www- 01.ibm.com/support/knowledgecenter/SSPT3X_2.1.1/com.ibm.swg.im.infosphere.biginsight s.install.doc/Install.pdf

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 126
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 128
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 133
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 134
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 136
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 138
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 143
......

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