IBM

Exam C2090-102

IBM Big Data Architect

Verson: Demo

[Total Questions: 10]

Question No : 1

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

Question No : 2

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

Question No : 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

Question No : 4

Cloudant is defined by which of the following?

A. A JSON document store
B. A key/value store
C. A graph database
D. A column/family store

Answer: C
Reference:

http://www.redbooks.ibm.com/redpapers/pdfs/redp5126.pdf

Question No : 5

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

Question No : 6

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

Question No : 7

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

Question No : 8

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
Reference:

http://www.ibm.com/software/data/puredata/analytics/nztechnology/analytics.html

Question No : 9

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

Question No : 10

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. InfoSphere DataStage with InfoSphere Streams
C. Hive with MapReduce
D. Cassandra with YARN

Answer: B