[June-2018-NewValid Braindump2go 70-775 PDF Dumps and 70-775 VCE Dumps for Passing 70-775 Exam[17-22

 2018 June new Microsoft 70-775 Exam Dumps with PDF and VCE Just Updated Today! Following are some new 70-775 Real Exam Questions:1.|2018 New 70-775 Exam Dumps (PDF & VCE) 38Q&As Download:https://www.braindump2go.com/70-775.html2.|2018 New 70-775 Exam Questions & Answers Download:https://drive.google.com/drive/folders/1uJ5XlxxFUMLKFts7UH_t9bHMkDKxlWDT?usp=sharingQUESTION 17Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.You have an initial dataset that contains the crime data from major cities.You plan to build training models from the training data. You plan to automate the process of adding more data to the training models and to constantly tune the models by using the additional data, including data that is collected in near real-time. The system will be used to analyze event data gathered from many different sources, such as Internet of Things (IoT) devices, live video surveillance, and traffic activities, and to generate predictions of an increased crime risk at a particular time and place.You have an incoming data stream from Twitter and an incoming data stream from Facebook, which are event-based only, rather than time-based. You also have a time interval stream every 10 seconds.The data is in a key/value pair format. The value field represents a number that defines how many times a hashtag occurs within a Facebook post, or how many times a Tweet that contains a specific hashtag is retweeted.You must use the appropriate data storage, stream analytics techniques, and Azure HDInsight cluster types for the various tasks associated to the processing pipeline.You are designing the real-time portion of the input stream processing. The input will be a continuous stream of data and each record will be processed one at a time. The data will come from an Apache Kafka producer.You need to identify which HDInsight cluster to use for the final processing of the input data. This will be used to generate continuous statistics and real-time analytics. The latency to process each record must be less than one millisecond and tasks must be performed in parallel. Which type of cluster should you identify?A.   Apache StormB.   Apache HadoopC.   Apache HBaseD.   Apache SparkAnswer: A Explanation:**https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-storm-overview**QUESTION 18Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.You have an initial dataset that contains the crime data from major cities.You plan to build training models from the training data. You plan to automate the process of adding more data to the training models and to constantly tune the models by using the additional data, including data that is collected in near real-time. The system will be used to analyze event data gathered from many different sources, such as Internet of Things (IoT) devices, live video surveillance, and traffic activities, and to generate predictions of an increased crime risk at a particular time and place.You have an incoming data stream from Twitter and an incoming data stream from Facebook, which are event-based only, rather than time-based. You also have a time interval stream every 10 seconds.The data is in a key/value pair format. The value field represents a number that defines how many times a hashtag occurs within a Facebook post, or how many times a Tweet that contains a specific hashtag is retweeted.You must use the appropriate data storage, stream analytics techniques, and Azure HDInsight cluster types for the various tasks associated to the processing pipeline.You are planning a storage strategy for a large amount of analytic data used for the crime data analytics system. The initial data load involves over 100 billion records, and more than two billion records will be added daily.You already created an Apache Hadoop cluster in HDInsight premium.You need to implement the storage strategy to meet the following requirements:What should you create?A.   a virtual machine (VM) by using the Data Science Virtual Machine template for Windows that has premium storage, a G-series size, and uses Microsoft SQL Server 2016 to store the dataB.   an Azure Data Lake Analytics service by using Azure PowerShellC.   an Azure Data Lake Store account by using the Azure portalD.   an Azure Blob storage account by using the Azure portalAnswer: C Explanation:**https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-get-started-portal**QUESTION 19Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.After you answer a question in this sections, you will NOT be able to return to it. As a result, these questions will not appear in the review screen. You are building a security tracking solution in Apache Kafka to parse security logs. The security logs record an entry each time a user attempts to access an application. Each log entry contains the IP address used to make the attempt and the country from which the attempt originated.You need to receive notifications when an IP address from outside of the United States is used to access the application.Solution: Create two new brokers. Create a file import process to send messages. Run the producer.Does this meet the goal?A.   YesB.   No**Answer: B**QUESTION 20Note: This question is part of a series of questions

that use the same or similar answer choices. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series. Information and details provided in a question apply only to that question.You are implementing a batch processing solution by using Azure HDInsight.You plan to import 300 TB of data.You plan to use one job that has many concurrent tasks to import the data in memory.You need to maximize the amount of concurrent tasks for the job.What should you do?A.   Use a shuffle join in an Apache Hive query that stores the data in a JSON format.B.   Use a broadcast join in an Apache Hive query that stores the data in an ORC format.C.   Increase the number of spark.executor.cores in an Apache Spark job that stores the data in a text format.D.   Increase the number of spark.executor.instances in an Apache Spark job that stores the data in a text format.E.   Decrease the level of parallelism in an Apache Spark job that stores the data in a text format.F.   Use an action in an Apache Oozie workflow that stores the data in a text format.G.   Use an Azure Data Factory linked service that stores the data in Azure Data Lake.H.   Use an Azure Data Factory linked service that stores the data in an Azure DocumentDB database.Answer: CExplanation:**https://blog.cloudera.com/blog/2015/03/how-to-tune-your-apache-spark-jobs-part-2/** QUESTION 21Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.You are planning a big data infrastructure by using an Apache Spark cluster in Azure HDInsight. The cluster has 24 processor cores and 512 GB of memory.The architecture of the infrastructure is shown in the exhibit. (Click the Exhibit button.) The architecture will be used by the following users: Support analysts who run applications that will use REST to submit Spark jobs. Business analysts who use JDBC and ODBC client applications from a real-time view. The business analysts run monitoring queries to access aggregate results for 15 minutes. The results will be referenced by subsequent queries. Data analysts who publish notebooks drawn from batch layer, serving layer, and speed layer queries. All of the notebooks must support native interpreters for data sources that are batch processed. The serving layer queries are written in Apache Hive and must support multiple sessions. Unique GUIDs are used across the data sources, which allow the data analysts to use Spark SQL.The data sources in the batch layer share a common storage container. The following data sources are used: Hive for sales data Apache HBase for operations data HBase for logistics data by using a single region serverYou need to ensure that the support analysts can develop embedded analytics applications by using the least amount of development effort.Which technology should you implement?A.   ZeppelinB.   JupyterC.   Apache AmbariD.   LivyAnswer: DExplanation:
https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-apache-spark-livy-rest-interfaceQUESTION 22You have an Apache Hive table that contains one billion rows.You plan to use queries that will filter the data by using the WHERE clause. The values of the columns will be known only while the data loads into a Hive table.You need to decrease the query runtime.What should you configure?A.   static partitioningB.   bucket samplingC.   parallel executionD.   dynamic partitioningAnswer: CExplanation:**https://www.qubole.com/blog/5-tips-for-efficient-hive-queries/**!!!RECOMMEND!!!1.|2018 New 70-775 Exam Dumps (PDF & VCE) 38Q&As Download:https://www.braindump2go.com/70-775.html2.|2018 New 70-775 Study Guide Video: YouTube Video: YouTube.com/watch?v=l86_bnjB2uU