

Braindump2go Guarantee All Exams 100% Pass

**One Time!** 

## Vendor: Microsoft

# **Exam Code:** DP-200

# **Exam Name:** Implementing an Azure Data Solution

# > New Updated Questions from <u>Braindump2go</u> (Updated in <u>May/2020</u>)

# Visit Braindump2go and Download Full Version DP-200 Exam Dumps

## **QUESTION 186**

Note: 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 section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are designing an Azure SQL Database that will use elastic pools. You plan to store data about customers in a table. Each record uses a value for CustomerID.

You need to recommend a strategy to partition data based on values in CustomerID.

Proposed Solution: Separate data into customer regions by using horizontal partitioning. Does the solution meet the goal?

A. Yes

B. No

## Answer: B

#### **Explanation:**

We should use Horizontal Partitioning through Sharding, not divide through regions.

Note: Horizontal Partitioning - Sharding: Data is partitioned horizontally to distribute rows across a scaled out data tier. With this approach, the schema is identical on all participating databases. This approach is also called "sharding". Sharding can be performed and managed using (1) the elastic database tools libraries or (2) self-sharding. An elastic query is used to query or compile reports across many shards.

References:

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-elastic-query-overview

#### **QUESTION 187**

Note: 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 section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are designing an Azure SQL Database that will use elastic pools. You plan to store data about customers in a table. Each record uses a value for CustomerID.

You need to recommend a strategy to partition data based on values in CustomerID.

Proposed Solution: Separate data into shards by using horizontal partitioning.

Does the solution meet the goal?

A. Yes

B. No

## Answer: A

## Explanation:

Horizontal Partitioning - Sharding: Data is partitioned horizontally to distribute rows across a scaled out data tier. With this approach, the schema is identical on all participating databases. This approach is also called "sharding". Sharding



can be performed and managed using (1) the elastic database tools libraries or (2) self- sharding. An elastic query is used to query or compile reports across many shards.

References:

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-elastic-query-overview

## **QUESTION 188**

You are evaluating data storage solutions to support a new application.

You need to recommend a data storage solution that represents data by using nodes and relationships in graph structures.

Which data storage solution should you recommend?

- A. Blob Storage
- B. Cosmos DB
- C. Data Lake Store
- D. HDInsight

## Answer: B

#### **Explanation:**

For large graphs with lots of entities and relationships, you can perform very complex analyses very quickly. Many graph databases provide a query language that you can use to traverse a network of relationships efficiently. Relevant Azure service: Cosmos DB

References:

https://docs.microsoft.com/en-us/azure/architecture/guide/technology-choices/data-store-overview

#### **QUESTION 189**

You are designing a data processing solution that will implement the lambda architecture pattern. The solution will use Spark running on HDInsight for data processing.

You need to recommend a data storage technology for the solution.

Which two technologies should you recommend? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

- A. Azure Cosmos DB
- B. Azure Service Bus
- C. Azure Storage Queue
- D. Apache Cassandra
- E. Kafka HDInsight

# Answer: AE Explanation:

To implement a lambda architecture on Azure, you can combine the following technologies to accelerate real-time big data analytics:

Azure Cosmos DB, the industry's first globally distributed, multi-model database service.

Apache Spark for Azure HDInsight, a processing framework that runs large-scale data analytics applications Azure Cosmos DB change feed, which streams new data to the batch layer for HDInsight to process The Spark to Azure Cosmos DB Connector

E: You can use Apache Spark to stream data into or out of Apache Kafka on HDInsight using DStreams. References:

https://docs.microsoft.com/en-us/azure/cosmos-db/lambda-architecture

## **QUESTION 190**

A company manufactures automobile parts. The company installs IoT sensors on manufacturing machinery. You must design a solution that analyzes data from the sensors.

You need to recommend a solution that meets the following requirements:

- Data must be analyzed in real-time.
- Data queries must be deployed using continuous integration.
- Data must be visualized by using charts and graphs.
- Data must be available for ETL operations in the future.
- The solution must support high-volume data ingestion.



**One Time!** 

Which three actions should you recommend? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

- A. Use Azure Analysis Services to query the data. Output query results to Power BI.
- B. Configure an Azure Event Hub to capture data to Azure Data Lake Storage.
- C. Develop an Azure Stream Analytics application that queries the data and outputs to Power BI. Use Azure Data Factory to deploy the Azure Stream Analytics application.
- D. Develop an application that sends the IoT data to an Azure Event Hub.
- E. Develop an Azure Stream Analytics application that queries the data and outputs to Power BI.
- Use Azure Pipelines to deploy the Azure Stream Analytics application.
- F. Develop an application that sends the IoT data to an Azure Data Lake Storage container.

#### Answer: BCD

#### **QUESTION 191**

You are designing an Azure Databricks interactive cluster.

You need to ensure that the cluster meets the following requirements:

```
- Enable auto-termination
```

```
- Retain cluster configuration indefinitely after cluster termination.
```

What should you recommend?

- A. Start the cluster after it is terminated.
- B. Pin the cluster
- C. Clone the cluster after it is terminated.
- D. Terminate the cluster manually at process completion.

#### Answer: B

#### Explanation:

To keep an interactive cluster configuration even after it has been terminated for more than 30 days, an administrator can pin a cluster to the cluster list.

References:

https://docs.azuredatabricks.net/user-guide/clusters/terminate.html

#### **QUESTION 192**

You are designing a solution for a company. The solution will use model training for objective classification. You need to design the solution. What should you recommend?

- A. an Azure Cognitive Services application
- B. a Spark Streaming job
- C. interactive Spark queries
- D. Power BI models
- E. a Spark application that uses Spark MLib.

## Answer: E

Explanation:

Spark in SQL Server big data cluster enables AI and machine learning.

You can use Apache Spark MLlib to create a machine learning application to do simple predictive analysis on an open dataset.

MLlib is a core Spark library that provides many utilities useful for machine learning tasks, including utilities that are suitable for:

Classification

Regression

Clustering

Topic modeling

Singular value decomposition (SVD) and principal component analysis (PCA) Hypothesis testing and calculating sample statistics



#### References:

https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-machine-learning-mllib-ipython

#### **QUESTION 193**

## Case study 4 - ADatum Corporation

#### Overview

ADatum Corporation is a retailer that sells products through two sales channels: retail stores and a website. **Existing Environment** 

ADatum has one database server that has Microsoft SQL Server 2016 installed. The server hosts three mission-critical databases named SALESDB, DOCDB, and REPORTINGDB.

SALESDB collects data from the stored and the website.

DOCDB stored documents that connect to the sales data in SALESDB. The documents are stored in two different JSON formats based on the sales channel.

REPORTINGDB stores reporting data and contains server columnstore indexes. A daily process creates reporting data in REPORTINGDB from the data in SALESDB. The process is implemented as a SQL Server Integration Services (SSIS) package that runs a stored procedure from SALESDB.

#### Requirements

#### **Planned Changes**

ADatum plans to move the current data infrastructure to Azure. The new infrastructure has the following requirements:

- Migrate SALESDB and REPORTINGDB to an Azure SQL database.

- Migrate DOCDB to Azure Cosmos DB.

The sales data including the documents in JSON format, must be gathered as it arrives and analyzed online by using Azure Stream Analytics. The analytic process will perform aggregations that must be done continuously, without gaps, and without overlapping.
As they arrive, all the sales documents in JSON format must be transformed into one consistent format.

- Azure Data Factory will replace the SSIS process of copying the data from SALESDB to REPORTINGDB.

#### **Technical Requirements**

The new Azure data infrastructure must meet the following technical requirements:

- Data in SALESDB must encrypted by using Transparent Data Encryption (TDE). The encryption must use your own key.

- SALESDB must be restorable to any given minute within the past three weeks.

- Real-time processing must be monitored to ensure that workloads are sized properly based on actual usage patterns.

- Missing indexes must be created automatically for REPORTINGDB.

- Disk IO, CPU, and memory usage must be monitored for SALESDB.

You need to configure a disaster recovery solution for SALESDB to meet the technical requirements. What should you configure in the backup policy?

- A. weekly long-term retention backups that are retained for three weeks
- B. failover groups
- C. a point-in-time restore
- D. geo-replication

## Answer: C

#### **Explanation:**

Scenario: SALESDB must be restorable to any given minute within the past three weeks.

The Azure SQL Database service protects all databases with an automated backup system. These backups are retained for 7 days for Basic, 35 days for Standard and 35 days for Premium. Point-in-time restore is a self-service capability, allowing customers to restore a Basic, Standard or Premium database from these backups to any point within the retention period.

#### References:

https://azure.microsoft.com/en-us/blog/azure-sql-database-point-in-time-restore/

#### **QUESTION 194**

#### Case study 4 - ADatum Corporation Overview

ADatum Corporation is a retailer that sells products through two sales channels: retail stores and a website.



**One Time!** 

## **Existing Environment**

ADatum has one database server that has Microsoft SQL Server 2016 installed. The server hosts three mission-critical databases named SALESDB, DOCDB, and REPORTINGDB.

SALESDB collects data from the stored and the website.

DOCDB stored documents that connect to the sales data in SALESDB. The documents are stored in two different JSON formats based on the sales channel.

REPORTINGDB stores reporting data and contains server columnstore indexes. A daily process creates reporting data in REPORTINGDB from the data in SALESDB. The process is implemented as a SQL Server Integration Services (SSIS) package that runs a stored procedure from SALESDB.

## Requirements

## Planned Changes

ADatum plans to move the current data infrastructure to Azure. The new infrastructure has the following requirements: - Migrate SALESDB and REPORTINGDB to an Azure SQL database.

- Migrate DOCDB to Azure Cosmos DB.

The sales data including the documents in JSON format, must be gathered as it arrives and analyzed online by using Azure Stream Analytics. The analytic process will perform aggregations that must be done continuously, without gaps, and without overlapping.
As they arrive, all the sales documents in JSON format must be transformed into one consistent format.

- Azure Data Factory will replace the SSIS process of copying the data from SALESDB to REPORTINGDB.

#### **Technical Requirements**

#### The new Azure data infrastructure must meet the following technical requirements:

- Data in SALESDB must encrypted by using Transparent Data Encryption (TDE). The encryption must use your own key.

- SALESDB must be restorable to any given minute within the past three weeks.

- Real-time processing must be monitored to ensure that workloads are sized properly based on actual usage patterns.

- Missing indexes must be created automatically for REPORTINGDB.

- Disk IO, CPU, and memory usage must be monitored for SALESDB.

You need to implement event processing by using Stream Analytics to produce consistent JSON documents.

Which three actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Define an output to Cosmos DB.
- B. Define a query that contains a JavaScript user-defined aggregates (UDA) function.
- C. Define a reference input.
- D. Define a transformation query.
- E. Define an output to Azure Data Lake Storage Gen2.
- F. Define a stream input.

#### Answer: DEF

#### Explanation:

DOCDB stored documents that connect to the sales data in SALESDB. The documents are stored in two different JSON formats based on the sales channel.

The sales data including the documents in JSON format, must be gathered as it arrives and analyzed online by using Azure Stream Analytics. The analytic process will perform aggregations that must be done continuously, without gaps, and without overlapping.

As they arrive, all the sales documents in JSON format must be transformed into one consistent format.

#### **QUESTION 195**

## **Case study 4 - ADatum Corporation**

#### Overview

ADatum Corporation is a retailer that sells products through two sales channels: retail stores and a website.

#### Existing Environment

ADatum has one database server that has Microsoft SQL Server 2016 installed. The server hosts three mission-critical databases named SALESDB, DOCDB, and REPORTINGDB.

SALESDB collects data from the stored and the website.



```
One Time!
```

DOCDB stored documents that connect to the sales data in SALESDB. The documents are stored in two different JSON formats based on the sales channel.

REPORTINGDB stores reporting data and contains server columnstore indexes. A daily process creates reporting data in REPORTINGDB from the data in SALESDB. The process is implemented as a SQL Server Integration Services (SSIS) package that runs a stored procedure from SALESDB.

#### Requirements

#### **Planned Changes**

ADatum plans to move the current data infrastructure to Azure. The new infrastructure has the following requirements:

- Migrate SALESDB and REPORTINGDB to an Azure SQL database.

- Migrate DOCDB to Azure Cosmos DB.

The sales data including the documents in JSON format, must be gathered as it arrives and analyzed online by using Azure Stream Analytics. The analytic process will perform aggregations that must be done continuously, without gaps, and without overlapping.
As they arrive, all the sales documents in JSON format must be transformed into one consistent format.

- Azure Data Factory will replace the SSIS process of copying the data from SALESDB to REPORTINGDB.

#### **Technical Requirements**

The new Azure data infrastructure must meet the following technical requirements:

- Data in SALESDB must encrypted by using Transparent Data Encryption (TDE). The encryption must use your own key.

- SALESDB must be restorable to any given minute within the past three weeks.

- Real-time processing must be monitored to ensure that workloads are sized properly based on actual usage patterns.

- Missing indexes must be created automatically for REPORTINGDB.

- Disk IO, CPU, and memory usage must be monitored for SALESDB.

Which windowing function should you use to perform the streaming aggregation of the sales data?

- A. Tumbling
- B. Hopping
- C. Sliding
- D. Session

## Answer: A

#### **Explanation:**

Scenario: The analytic process will perform aggregations that must be done continuously, without gaps, and without overlapping.

The key differentiators of a Tumbling window are that they repeat, do not overlap, and an event cannot belong to more than one tumbling window.

Incorrect Answers:

B, C: Like hopping windows, events can belong to more than one sliding window.

D: Session windows can have gaps.

References:

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-window-functions

#### **QUESTION 196**

#### Case study 4 - ADatum Corporation Overview

ADatum Corporation is a retailer that sells products through two sales channels: retail stores and a website.

#### **Existing Environment**

ADatum has one database server that has Microsoft SQL Server 2016 installed. The server hosts three mission-critical databases named SALESDB, DOCDB, and REPORTINGDB.

SALESDB collects data from the stored and the website.

DOCDB stored documents that connect to the sales data in SALESDB. The documents are stored in two different JSON formats based on the sales channel.

REPORTINGDB stores reporting data and contains server columnstore indexes. A daily process creates reporting data in REPORTINGDB from the data in SALESDB. The process is implemented as a SQL Server Integration Services (SSIS) package that runs a stored procedure from SALESDB.

## Requirements



Braindump2go Guarantee All Exams 100% Pass

**One Time!** 

#### **Planned Changes**

ADatum plans to move the current data infrastructure to Azure. The new infrastructure has the following requirements:

- Migrate SALESDB and REPORTINGDB to an Azure SQL database.

- Migrate DOCDB to Azure Cosmos DB.

The sales data including the documents in JSON format, must be gathered as it arrives and analyzed online by using Azure Stream Analytics. The analytic process will perform aggregations that must be done continuously, without gaps, and without overlapping.
As they arrive, all the sales documents in JSON format must be transformed into one consistent format.

- Azure Data Factory will replace the SSIS process of copying the data from SALESDB to REPORTINGDB.

#### **Technical Requirements**

The new Azure data infrastructure must meet the following technical requirements:

- Data in SALESDB must encrypted by using Transparent Data Encryption (TDE). The encryption must use your own key.

- SALESDB must be restorable to any given minute within the past three weeks.

- Real-time processing must be monitored to ensure that workloads are sized properly based on actual usage patterns.

- Missing indexes must be created automatically for REPORTINGDB.

- Disk IO, CPU, and memory usage must be monitored for SALESDB.

How should you monitor SALESDB to meet the technical requirements?

A. Query the sys.resource\_stats dynamic management view.

- B. Review the Query Performance Insights for SALESDB.
- C. Query the sys.dm\_os\_wait\_stats dynamic management view.
- D. Review the auditing information of SALESDB.

#### Answer: A

Explanation:

Scenario: Disk IO, CPU, and memory usage must be monitored for SALESDB The sys.resource\_stats returns historical data for CPU, IO, DTU consumption. There's one row every 5 minute for a database in an Azure logical SQL Server if there's a change in the metrics.

Incorrect Answers:

B: Query Performance Insight helps you to quickly identify what your longest running queries are, how they change over time, and what waits are affecting them.

C: sys.dm\_os\_wait\_stats: specific types of wait times during query execution can indicate bottlenecks or stall points within the query. Similarly, high wait times, or wait counts server wide can indicate bottlenecks or hot spots in interaction query interactions within the server instance. For example, lock waits indicate data contention by queries; page IO latch waits indicate slow IO response times; page latch update waits indicate incorrect file layout. References:

https://dataplatformlabs.com/monitoring-azure-sql-database-with-sys-resource\_stats/