

# DUMPS ARENA

## Designing an Azure Data Solution

Microsoft DP-201

Version Demo

Total Demo Questions: 15

Total Premium Questions: 209

Buy Premium PDF

<https://dumpsarena.co>

[sales@dumpsarena.co](mailto:sales@dumpsarena.co)

[sales@dumpsarena.co](mailto:sales@dumpsarena.co)  
[dumpsarena.co](https://dumpsarena.co)

## Topic Break Down

Topic	No. of Questions
Topic 1, Case Study 1	5
Topic 2, Case Study 2	4
Topic 3, Case Study 3	3
Topic 4, Case Study 4	5
Topic 5, Case Study 5	3
Topic 6, Case Study 6	2
Topic 7, Case Study 7	4
Topic 8, Case Study 8	4
Topic 9, Case Study 9	4
Topic 10, Case Study 10	2
Topic 11, Case Study 11	2
Topic 12, Case Study 12	2
Topic 13, Case Study 13	2
Topic 14, Mixed Questions	167
<b>Total</b>	<b>209</b>

**QUESTION NO: 1 - (DRAG DROP)**

DRAG DROP

You are designing a Spark job that performs batch processing of daily web log traffic.

When you deploy the job in the production environment, it must meet the following requirements:

- Run once a day.
- Display status information on the company intranet as the job runs.

You need to recommend technologies for triggering and monitoring jobs.

Which technologies should you recommend? To answer, drag the appropriate technologies to the correct locations. Each technology may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Select and Place:

**Answer Area**

**Technologies**

- Livy
- Beeline
- Azure Logic App
- Azure API App

**Requirement**

- Triggering of jobs
- Monitoring of jobs

**Technology**

Empty boxes for placing technologies under the requirements.

**ANSWER:**

## Answer Area

## Technologies

Livy

Beeline

Azure Logic App

Azure API App

## Requirement

Triggering of jobs

Monitoring of jobs

## Technology

Livy

Beeline

## Explanation:

Box 1: Livy

You can use Livy to run interactive Spark shells or submit batch jobs to be run on Spark.

Box 2: Beeline

Apache Beeline can be used to run Apache Hive queries on HDInsight. You can use Beeline with Apache Spark.

Note: Beeline is a Hive client that is included on the head nodes of your HDInsight cluster. Beeline uses JDBC to connect to HiveServer2, a service hosted on your HDInsight cluster. You can also use Beeline to access Hive on HDInsight remotely over the internet.

Reference: <https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-livy-rest-interface>  
<https://docs.microsoft.com/en-us/azure/hdinsight/hadoop/apache-hadoop-use-hive-beeline>

## QUESTION NO: 2

You have a Windows-based solution that analyzes scientific data. You are designing a cloudbased solution that performs real-time analysis of the data.

You need to design the logical flow for the solution.

Which two actions should you recommend? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

**A.** Send data from the application to an Azure Stream Analytics job.

**B.** Use an Azure Stream Analytics job on an edge device. Ingress data from an Azure Data Factory instance and build queries that output to Power BI.

- C. Use an Azure Stream Analytics job in the cloud. Ingress data from the Azure Event Hub instance and build queries that output to Power BI.
- D. Use an Azure Stream Analytics job in the cloud. Ingress data from an Azure Event Hub instance and build queries that output to Azure Data Lake Storage.
- E. Send data from the application to Azure Data Lake Storage.
- F. Send data from the application to an Azure Event Hub instance.

**ANSWER: C F**

**Explanation:**

Stream Analytics has first-class integration with Azure data streams as inputs from three kinds of resources:

- Azure Event Hubs
- Azure IoT Hub
- Azure Blob storage

Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-define-inputs>

**QUESTION NO: 3**

You manage a process that performs analysis of daily web traffic logs on an HDInsight cluster.

Each of the 250 web servers generates approximately 10 megabytes (MB) of log data each day. All log data is stored in a single folder in Microsoft Azure Data Lake Storage Gen 2.

You need to improve the performance of the process.

Which two changes should you make? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. Combine the daily log files for all servers into one file
- B. Increase the value of the mapreduce.map.memory parameter
- C. Move the log files into folders so that each day's logs are in their own folder
- D. Increase the number of worker nodes
- E. Increase the value of the hive.tez.container.size parameter

**ANSWER: A C**

**Explanation:**

A: Typically, analytics engines such as HDInsight and Azure Data Lake Analytics has a per-five overhead. If you store your data as many small files, this can negatively affect performance. In general, organize your data into larger sized files for better performance (256MB to 100GB in size). Some engines and applications might have trouble efficiently processing files that are greater than 100GB in size.

C: For Hive workloads, partition pruning of time-series data can help some queries read only a subset of the data which improves performance.

Those pipelines that ingest time-series data, often place their files with a very structured naming for files and folders. Below is a very common example we see for data is structured by date: \DataSet\YYYY\MM\DD\datafile\_YYYY\_MM\_DD.tsv

Notice that the datetime information appears both as folders and in the filename.

Reference: <https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-performance-tuningguidance>

#### QUESTION NO: 4

A company installs IoT devices to monitor its fleet of delivery vehicles. Data from devices is collected from Azure Event Hub.

The data must be transmitted to Power BI for real-time data visualizations.

You need to recommend a solution.

What should you recommend?

- A. Azure HDInsight with Spark Streaming
- B. Apache Spark in Azure Databricks
- C. Azure Stream Analytics
- D. Azure HDInsight with Storm

#### ANSWER: C

##### Explanation:

Step 1: Get your IoT hub ready for data access by adding a consumer group.

Step 2: Create, configure, and run a Stream Analytics job for data transfer from your IoT hub to your Power BI account.

Step 3: Create and publish a Power BI report to visualize the data.

Reference:

<https://docs.microsoft.com/en-us/azure/iot-hub/iot-hub-live-data-visualization-in-power-bi>

#### QUESTION NO: 5

You plan to deploy an Azure SQL Database instance to support an application. You plan to use the DTU-based purchasing model.

Backups of the database must be available for 30 days and point-in-time restoration must be possible.

You need to recommend a backup and recovery policy.

What are two possible ways to achieve the goal? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. Use the Premium tier and the default backup retention policy.
- B. Use the Basic tier and the default backup retention policy.
- C. Use the Standard tier and the default backup retention policy.
- D. Use the Standard tier and configure a long-term backup retention policy.
- E. Use the Premium tier and configure a long-term backup retention policy.

**ANSWER: D E**

**Explanation:**

The default retention period for a database created using the DTU-based purchasing model depends on the service tier:

- Basic service tier is 1 week.
- Standard service tier is 5 weeks. ▪ Premium service tier is 5 weeks.

Incorrect Answers:

B: Basic tier only allows restore points within 7 days.

Reference:

<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-long-term-retention>

**QUESTION NO: 6**

You plan to migrate data to Azure SQL Database.

The database must remain synchronized with updates to Microsoft Azure and SQL Server.

You need to set up the database as a subscriber.

What should you recommend?

- A. Azure Data Factory
- B. SQL Server Data Tools
- C. Data Migration Assistant
- D. SQL Server Agent for SQL Server 2017 or later
- E. SQL Server Management Studio 17.9.1 or later

ANSWER: E

**Explanation:**

To set up the database as a subscriber we need to configure database replication. You can use SQL Server Management Studio to configure replication. Use the latest versions of SQL Server Management Studio in order to be able to use all the features of Azure SQL Database.

Reference:

<https://www.sqlshack.com/sql-server-database-migration-to-azure-sql-database-using-sql-server-transactional-replication/>

**QUESTION NO: 7 - (HOTSPOT)**

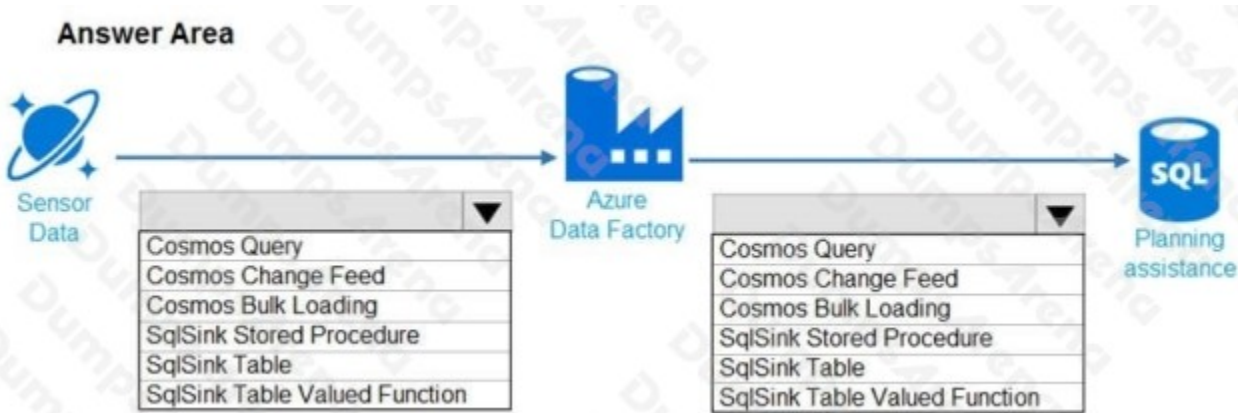
**HOTSPOT**

You need to design the data loading pipeline for Planning Assistance.

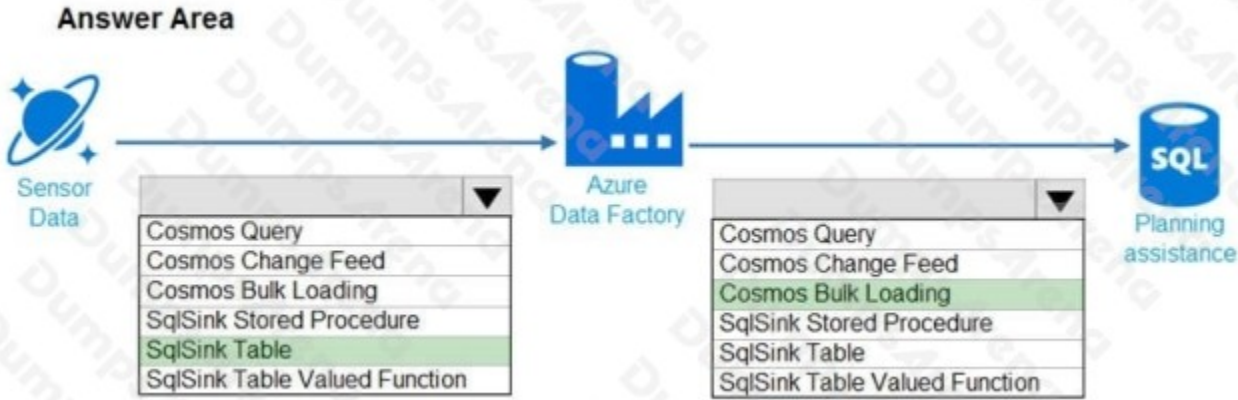
What should you recommend? To answer, drag the appropriate technologies to the correct locations. Each technology may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

**Hot Area:**



ANSWER:



**Explanation:**

Box 1: SqlSink Table

Sensor data must be stored in a Cosmos DB named treydata in a collection named SensorData

Box 2: Cosmos Bulk Loading

Use Copy Activity in Azure Data Factory to copy data from and to Azure Cosmos DB (SQL API).

Scenario: Data from the Sensor Data collection will automatically be loaded into the Planning Assistance database once a week by using Azure Data Factory. You must be able to manually trigger the data load process.

Data used for Planning Assistance must be stored in a sharded Azure SQL Database.

Reference: <https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-cosmos-db>

**QUESTION NO: 8**

You are planning a streaming data solution that will use Azure Databricks. The solution will stream sales transaction data from an online store. The solution has the following specifications:

- The output data will contain items purchased, quantity, line total sales amount, and line total tax amount.
- Line total sales amount and line total tax amount will be aggregated in Databricks.
- Sales transactions will never be updated. Instead, new rows will be added to adjust a sale.

You need to recommend an output mode for the dataset that will be processed by using Structured Streaming. The solution must minimize duplicate data.

What should you recommend?

- A. Append
- B. Complete
- C. Update

**ANSWER: A****Explanation:**

Append Mode: Only new rows appended in the result table since the last trigger are written to external storage. This is applicable only for the queries where existing rows in the Result Table are not expected to change.

Incorrect Answers:

B: Complete Mode: The entire updated result table is written to external storage. It is up to the storage connector to decide how to handle the writing of the entire table.

C: Update Mode: Only the rows that were updated in the result table since the last trigger are written to external storage. This is different from Complete Mode in that Update Mode outputs only the rows that have changed since the last trigger. If the query doesn't contain aggregations, it is equivalent to Append mode.

Reference:

<https://docs.microsoft.com/en-us/azure/databricks/getting-started/spark/streaming>

**QUESTION NO: 9**

You have a line-of-business (LOB) app that reads files from and writes files to Azure Blob storage in an Azure Storage account.

You need to recommend changes to the storage account to meet the following requirements:

Provide the highest possible availability. Minimize potential data loss.

Which three changes should you recommend? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. From the app, query the LastSyncTime of the storage account.
- B. From the storage account, enable soft deletes.
- C. From the storage account, enable read-access geo-redundancy storage (RA-GRS).
- D. From the app, add retry logic to the storage account interactions.
- E. From the storage account, enable a time-based retention policy.

**ANSWER: B C E****Explanation:**

Soft delete protects blob data from being accidentally or erroneously modified or deleted. When soft delete is enabled for a storage account, blobs, blob versions (preview), and snapshots in that storage account may be recovered after they are deleted, within a retention period that you specify.

Geo-redundant storage (with GRS or GZRS) replicates your data to another physical location in the secondary region to protect against regional outages. However, that data is available to be read only if the customer or Microsoft initiates a failover from the primary to secondary region. When you enable read access to the secondary region, your data is available

to be read if the primary region becomes unavailable. For read access to the secondary region, enable readaccess geo-redundant storage (RA-GRS) or read-access geo-zone-redundant storage (RAGZRS).

Reference: <https://docs.microsoft.com/en-us/azure/storage/blobs/soft-delete-overview>

<https://docs.microsoft.com/en-us/azure/storage/common/storage-redundancy#read-access-todata-in-the-secondary-region>

**QUESTION NO: 10**

You need to recommend a security solution to grant anonymous users permission to access the blobs in a specific container only.

What should you include in the recommendation?

- A. access keys for the storage account
- B. shared access signatures (SAS)
- C. Role assignments
- D. the public access level for the container

**ANSWER: D****Explanation:**

You can enable anonymous, public read access to a container and its blobs in Azure Blob storage. By doing so, you can grant read-only access to these resources without sharing your account key, and without requiring a shared access signature (SAS).

Public read access is best for scenarios where you want certain blobs to always be available for anonymous read access.

Reference: <https://docs.microsoft.com/en-us/azure/storage/blobs/storage-manage-access-to-resources>

**QUESTION NO: 11**

You are designing an Azure Databricks cluster that runs user-defined local processes.

You need to recommend a cluster configuration that meets the following requirements:

- Minimize query latency.
- Reduce overall costs without compromising other requirements.
- Maximize the number of users that can run queries on the cluster at the same time.

Which cluster type should you recommend?

- A. Standard with Autoscaling
- B. High Concurrency with Auto Termination
- C. High Concurrency with Autoscaling

D. Standard with Auto Termination

**ANSWER: C**

**Explanation:**

High Concurrency clusters allow multiple users to run queries on the cluster at the same time, while minimizing query latency. Autoscaling clusters can reduce overall costs compared to a statically-sized cluster.

Incorrect Answers:

A, D: Standard clusters are recommended for a single user.

Reference: <https://docs.azuredatabricks.net/user-guide/clusters/create.html> <https://docs.azuredatabricks.net/user-guide/clusters/high-concurrency.html#high-concurrency> <https://docs.azuredatabricks.net/user-guide/clusters/terminate.html> <https://docs.azuredatabricks.net/user-guide/clusters/sizing.html#enable-and-configure-autoscaling>

**QUESTION NO: 12 - (HOTSPOT)**

**HOTSPOT**

You are designing the security for a mission critical Azure SQL database named DB1. DB1 contains several columns that store Personally Identifiable

Information (PII) data

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

- Ensures that DB1 is encrypted at rest
- Ensures that data from the columns containing PII data is encrypted in transit

Which security solution should you recommend for DB1 and the columns? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

**Hot Area:**

### Answer Area

DB1:

	▼
Always Encrypted	
Dynamic data masking	
Row-level security	
Transparent Data Encryption (TDE)	

Columns:

	▼
Always Encrypted	
Dynamic data masking	
Row-level security	
Transparent Data Encryption (TDE)	

ANSWER:

## Answer Area

DB1:

Always Encrypted
Dynamic data masking
Row-level security
Transparent Data Encryption (TDE)

Columns:

Always Encrypted
Dynamic data masking
Row-level security
Transparent Data Encryption (TDE)

### Explanation:

DB1: Transparent Data Encryption

Azure SQL Database currently supports encryption at rest for Microsoft-managed service side and client-side encryption scenarios.

Support for server encryption is currently provided through the SQL feature called Transparent Data Encryption.

Columns: Always encrypted

Always Encrypted is a feature designed to protect sensitive data stored in Azure SQL Database or SQL Server databases. Always Encrypted allows clients to encrypt sensitive data inside client applications and never reveal the encryption keys to the database engine (SQL Database or SQL Server).

Note: Most data breaches involve the theft of critical data such as credit card numbers or personally identifiable information. Databases can be treasure troves of sensitive information. They can contain customers' personal data (like national identification numbers), confidential competitive information, and intellectual property. Lost or stolen data, especially customer data, can result in brand damage, competitive disadvantage, and serious fines--even lawsuits.

Reference: <https://docs.microsoft.com/en-us/azure/security/fundamentals/encryption-atrest> <https://docs.microsoft.com/en-us/azure/security/fundamentals/database-security-overview>

### QUESTION NO: 13

You have a MongoDB database that you plan to migrate to an Azure Cosmos DB account that uses the MongoDB API.

During testing, you discover that the migration takes longer than expected.

You need to recommend a solution that will reduce the amount of time it takes to migrate the data.

What are two possible recommendations to achieve this goal? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. Increase the Request Units (RUs).
- B. Turn off indexing.
- C. Add a write region.
- D. Create unique indexes.
- E. Create compound indexes.

**ANSWER: A B**

**Explanation:**

A: Increase the throughput during the migration by increasing the Request Units (RUs).

For customers that are migrating many collections within a database, it is strongly recommend to configure database-level throughput. You must make this choice when you create the database. The minimum database-level throughput capacity is 400 RU/sec. Each collection sharing database-level throughput requires at least 100 RU/sec.

B: By default, Azure Cosmos DB indexes all your data fields upon ingestion. You can modify the indexing policy in Azure Cosmos DB at any time. In fact, it is often recommended to turn off indexing when migrating data, and then turn it back on when the data is already in Cosmos DB.

Reference: <https://docs.microsoft.com/bs-latn-ba/Azure/cosmos-db/mongodb-pre-migration>

**QUESTION NO: 14**

You need to design a solution to support the storage of datasets. The solution must meet the following requirements:

- Send email alerts when new datasets are added.
- Control access to collections of datasets by using Azure Active Directory groups.
- Support the storage of Microsoft Excel, Comma Separated Values (CSV), and zip files.

What should you include in the solution?

- A. Azure SQL Database
- B. Azure Storage
- C. Azure Cosmos DB
- D. Azure HDInsight

ANSWER: B

**Explanation:**

Reference:

<https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/data-storage> Design data processing solutions

**QUESTION NO: 15 - (DRAG DROP)**

DRAG DROP

You have data on the 75,000 employees of your company. The data contains the properties shown in the following table.

Name	Data populated	Unique values	Distinct values
Employee ID	100%	Each value is unique	75,000
Employee Surname	99%	50% of values are unique	40,000
Employee Given Name	98%	40% of values are unique	20,000
Employee Birth Date	99%	20% of values are unique	200
Employment Start Date	100%	10% of values are unique	100
Current Department	100%	0% of values are unique	25

You need to store the employee data in an Azure Cosmos DB container. Most queries on the data will filter by the Current Department and the Employee Surname properties.

Which partition key and item ID should you use for the container? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

**Select and Place:**

Employee ID    Employee Surname    Current Department

**Answer Area**

Partition key:

Item ID:

ANSWER:



**Explanation:**

Partition key: Current Department

Item ID: Employee ID

Reference:

<https://docs.microsoft.com/en-us/rest/api/storageservices/designing-a-scalable-partitioning-strategy-for-azure-table-storage>