

➤ **Vendor: Microsoft**

➤ **Exam Code: DP-203**

➤ **Exam Name: Data Engineering on Microsoft Azure**

➤ **New Updated Questions from [Braindump2go](#) (Updated in [August/2021](#))**

### **[Visit Braindump2go and Download Full Version DP-203 Exam Dumps](#)**

#### **QUESTION 139**

You have an Azure Synapse Analytics dedicated SQL pool named Pool1 and a database named DB1. DB1 contains a fact table named Table1.

You need to identify the extent of the data skew in Table1.

What should you do in Synapse Studio?

- A. Connect to the built-in pool and run DBCC PDW\_SHOWSPACEUSED.
- B. Connect to the built-in pool and run DBCC CHECKALLOC.
- C. Connect to Pool1 and query sys.dm\_pdw\_node\_status.
- D. Connect to Pool1 and query sys.dm\_pdw\_nodes\_db\_partition\_stats.

**Answer: A**

#### **Explanation:**

A quick way to check for data skew is to use DBCC PDW\_SHOWSPACEUSED. The following SQL code returns the number of table rows that are stored in each of the 60 distributions. For balanced performance, the rows in your distributed table should be spread evenly across all the distributions.

```
DBCC PDW_SHOWSPACEUSED('dbo.FactInternetSales');
```

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribute>

#### **QUESTION 140**

You are monitoring an Azure Stream Analytics job.

You discover that the Backlogged Input Events metric is increasing slowly and is consistently non-zero.

You need to ensure that the job can handle all the events.

What should you do?

- A. Change the compatibility level of the Stream Analytics job.
- B. Increase the number of streaming units (SUs).
- C. Remove any named consumer groups from the connection and use \$default.
- D. Create an additional output stream for the existing input stream.

**Answer: B**

#### **Explanation:**

Backlogged Input Events: Number of input events that are backlogged. A non-zero value for this metric implies that your job isn't able to keep up with the number of incoming events. If this value is slowly increasing or consistently non-zero, you should scale out your job. You should increase the Streaming Units.

Note: Streaming Units (SUs) represents the computing resources that are allocated to execute a Stream Analytics job. The higher the number of SUs, the more CPU and memory resources are allocated for your job.

Reference:

<https://docs.microsoft.com/bs-cyrl-ba/azure/stream-analytics/stream-analytics-monitoring>

#### **QUESTION 141**

You are designing a star schema for a dataset that contains records of online orders. Each record includes an order

**[DP-203 Exam Dumps](#) [DP-203 Exam Questions](#) [DP-203 PDF Dumps](#) [DP-203 VCE Dumps](#)**

**<https://www.braindump2go.com/dp-203.html>**

date, an order due date, and an order ship date.

You need to ensure that the design provides the fastest query times of the records when querying for arbitrary date ranges and aggregating by fiscal calendar attributes.

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

NOTE: Each correct selection is worth one point.

- A. Create a date dimension table that has a DateTime key.
- B. Use built-in SQL functions to extract date attributes.
- C. Create a date dimension table that has an integer key in the format of YYYYMMDD.
- D. In the fact table, use integer columns for the date fields.
- E. Use DateTime columns for the date fields.

**Answer:** BD

**QUESTION 142**

Hotspot Question

You have two Azure Storage accounts named Storage1 and Storage2. Each account holds one container and has the hierarchical namespace enabled. The system has files that contain data stored in the Apache Parquet format.

You need to copy folders and files from Storage1 to Storage2 by using a Data Factory copy activity. The solution must meet the following requirements:

- No transformations must be performed.
- The original folder structure must be retained.
- Minimize time required to perform the copy activity.

How should you configure the copy activity? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

## Answer Area

Source dataset type:

	▼
Binary	
Parquet	
Delimited text	

Copy activity copy behavior:

	▼
FlattenHierarchy	
MergeFiles	
PreserveHierarchy	

**Answer:**

## Answer Area

Source dataset type:

	▼
Binary	
Parquet	
Delimited text	

Copy activity copy behavior:

	▼
FlattenHierarchy	
MergeFiles	
PreserveHierarchy	

### Explanation:

Box 1: Parquet

For Parquet datasets, the type property of the copy activity source must be set to ParquetSource.

Box 2: PreserveHierarchy

PreserveHierarchy (default): Preserves the file hierarchy in the target folder. The relative path of the source file to the source folder is identical to the relative path of the target file to the target folder.

Incorrect Answers:

FlattenHierarchy: All files from the source folder are in the first level of the target folder. The target files have autogenerated names.

MergeFiles: Merges all files from the source folder to one file. If the file name is specified, the merged file name is the specified name. Otherwise, it's an autogenerated file name.

Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/format-parquet>

<https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-data-lake-storage>

### QUESTION 143

Hotspot Question

You have a SQL pool in Azure Synapse.

You plan to load data from Azure Blob storage to a staging table. Approximately 1 million rows of data will be loaded daily. The table will be truncated before each daily load.

You need to create the staging table. The solution must minimize how long it takes to load the data to the staging table.

How should you configure the table? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

## Answer Area

Distribution: 

	▼
Hash	
Replicated	
Round-robin	

Indexing: 

	▼
Clustered	
Clustered columnstore	
Heap	

Partitioning: 

	▼
Date	
None	

Answer:

## Answer Area

Distribution: 

	▼
Hash	
Replicated	
Round-robin	

Indexing: 

	▼
Clustered	
Clustered columnstore	
Heap	

Partitioning: 

	▼
Date	
None	

### Explanation:

Box 1: Hash

Hash-distributed tables improve query performance on large fact tables. They can have very large numbers of rows and still achieve high performance.

Incorrect Answers:

Round-robin tables are useful for improving loading speed.

**Box 2: Clustered columnstore**

When creating partitions on clustered columnstore tables, it is important to consider how many rows belong to each partition. For optimal compression and performance of clustered columnstore tables, a minimum of 1 million rows per distribution and partition is needed.

**Box 3: Date**

Table partitions enable you to divide your data into smaller groups of data. In most cases, table partitions are created on a date column.

Partition switching can be used to quickly remove or replace a section of a table.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-partition>

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribute>

**QUESTION 144****Hotspot Question**

From a website analytics system, you receive data extracts about user interactions such as downloads, link clicks, form submissions, and video plays.

The data contains the following columns.

Name	Sample value
Date	15 Jan 2021
EventCategory	Videos
EventAction	Play
EventLabel	Contoso Promotional
ChannelGrouping	Social
TotalEvents	150
UniqueEvents	120
SessionWithEvents	99

You need to design a star schema to support analytical queries of the data. The star schema will contain four tables including a date dimension.

To which table should you add each column? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

## Answer Area

EventCategory: 

	▼
DimChannel	
DimDate	
DimEvent	
FactEvents	

ChannelGrouping: 

	▼
DimChannel	
DimDate	
DimEvent	
FactEvents	

TotalEvents: 

	▼
DimChannel	
DimDate	
DimEvent	
FactEvents	

Answer:

## Answer Area

EventCategory: 

	▼
DimChannel	
DimDate	
DimEvent	
FactEvents	

ChannelGrouping: 

	▼
DimChannel	
DimDate	
DimEvent	
FactEvents	

TotalEvents: 

	▼
DimChannel	
DimDate	
DimEvent	
FactEvents	

**Explanation:**

Box 1: DimEvent

Box 2: DimChannel

Box 3: FactEvents

Fact tables store observations or events, and can be sales orders, stock balances, exchange rates, temperatures, etc

Reference:

<https://docs.microsoft.com/en-us/power-bi/guidance/star-schema>

**QUESTION 145**

Hotspot Question

You plan to create a real-time monitoring app that alerts users when a device travels more than 200 meters away from a designated location.

You need to design an Azure Stream Analytics job to process the data for the planned app. The solution must minimize the amount of code developed and the number of technologies used.

What should you include in the Stream Analytics job? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

## Answer Area

Input type:  ▼

Stream
Reference

Function:  ▼

Aggregate
Geospatial
Windowing

Answer:

## Answer Area

Input type:  ▼

Stream
Reference

Function:  ▼

Aggregate
Geospatial
Windowing

**Explanation:**  
Input type: Stream

You can process real-time IoT data streams with Azure Stream Analytics.

Function: Geospatial

With built-in geospatial functions, you can use Azure Stream Analytics to build applications for scenarios such as fleet management, ride sharing, connected cars, and asset tracking.

Note: In a real-world scenario, you could have hundreds of these sensors generating events as a stream. Ideally, a gateway device would run code to push these events to Azure Event Hubs or Azure IoT Hubs.

Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-get-started-with-azure-stream-analytics-to-process-data-from-iot-devices>

<https://docs.microsoft.com/en-us/azure/stream-analytics/geospatial-scenarios>

**QUESTION 146**

Hotspot Question

You are processing streaming data from vehicles that pass through a toll booth.

You need to use Azure Stream Analytics to return the license plate, vehicle make, and hour the last vehicle passed during each 10-minute window.

How should you complete the query? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

**Answer Area**

```

WITH LastInWindow AS
(
    SELECT
        [ ] (Time) AS LastEventTime
        COUNT
        MAX
        MIN
        TOPONE
    FROM
        Input TIMESTAMP BY Time
    GROUP BY
        [ ] (minute, 10)
        HoppingWindow
        SessionWindow
        SlidingWindow
        TumblingWindow
)
SELECT
    Input.License_plate,
    Input.Make,
    Input.Time
FROM
    Input TIMESTAMP BY Time
    INNER JOIN LastInWindow
    ON [ ] (minute, Input, LastInWindow) BETWEEN 0 AND 10
        DATEADD
        DATEDIFF
        DATENAME
        DATEPART
AND Input.Time = LastInWindow.LastEventTime

```

Answer:

**Answer Area**

```

WITH LastInWindow AS
(
    SELECT
        (Time) AS LastEventTime
        COUNT
        MAX
        MIN
        TOPONE
    FROM
        Input TIMESTAMP BY Time
    GROUP BY
        (minute, 10)
        HoppingWindow
        SessionWindow
        SlidingWindow
        TumblingWindow
)
SELECT
    Input.License_plate,
    Input.Make,
    Input.Time
FROM
    Input TIMESTAMP BY Time
    INNER JOIN LastInWindow
    ON (minute, Input, LastInWindow) BETWEEN 0 AND 10
        DATEADD
        DATEDIFF
        DATENAME
        DATEPART
    AND Input.Time = LastInWindow.LastEventTime

```

**Explanation:**

Box 1: MAX

The first step on the query finds the maximum time stamp in 10-minute windows, that is the time stamp of the last event for that window. The second step joins the results of the first query with the original stream to find the event that match the last time stamps in each window.

Query:

```

WITH LastInWindow AS
(
    SELECT
    MAX(Time) AS LastEventTime
    FROM
    Input TIMESTAMP BY Time
    GROUP BY
    TumblingWindow(minute, 10)
)
SELECT
Input.License_plate,
Input.Make,
Input.Time
FROM
Input TIMESTAMP BY Time
INNER JOIN LastInWindow
ON DATEDIFF(minute, Input, LastInWindow) BETWEEN 0 AND 10

```

AND Input.Time = LastInWindow.LastEventTime

Box 2: TumblingWindow

Tumbling windows are a series of fixed-sized, non-overlapping and contiguous time intervals.

Box 3: DATEDIFF

DATEDIFF is a date-specific function that compares and returns the time difference between two DateTime fields, for more information, refer to date functions.

Reference:

<https://docs.microsoft.com/en-us/stream-analytics-query/tumbling-window-azure-stream-analytics>

#### **QUESTION 147**

Hotspot Question

A company plans to use Platform-as-a-Service (PaaS) to create the new data pipeline process. The process must meet the following requirements:

Ingest:

- Access multiple data sources.
- Provide the ability to orchestrate workflow.
- Provide the capability to run SQL Server Integration Services packages.

Store:

- Optimize storage for big data workloads.
- Provide encryption of data at rest.
- Operate with no size limits.

Prepare and Train:

- Provide a fully-managed and interactive workspace for exploration and visualization.
- Provide the ability to program in R, SQL, Python, Scala, and Java.
- Provide seamless user authentication with Azure Active Directory.

Model & Serve:

- Implement native columnar storage.
- Support for the SQL language
- Provide support for structured streaming.

You need to build the data integration pipeline.

Which technologies should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

## Answer Area

### Architecture requirement

### Technology

Ingest

	▼
Logic Apps	
Azure Data Factory	
Azure Automation	

Store

	▼
Azure Data Lake Storage	
Azure Blob storage	
Azure files	

Prepare and Train

	▼
HDInsight Apache Spark cluster	
Azure Databricks	
HDInsight Apache Storm cluster	

Model and Serve

	▼
HDInsight Apache Kafka cluster	
Azure Synapse Analytics	
Azure Data Lake Storage	

Answer:

## Answer Area

### Architecture requirement

### Technology

Ingest

▼
Logic Apps
Azure Data Factory
Azure Automation

Store

▼
Azure Data Lake Storage
Azure Blob storage
Azure files

Prepare and Train

▼
HDInsight Apache Spark cluster
Azure Databricks
HDInsight Apache Storm cluster

Model and Serve

▼
HDInsight Apache Kafka cluster
Azure Synapse Analytics
Azure Data Lake Storage

#### Explanation:

Ingest: Azure Data Factory

Azure Data Factory pipelines can execute SSIS packages.

In Azure, the following services and tools will meet the core requirements for pipeline orchestration, control flow, and data movement: Azure Data Factory, Oozie on HDInsight, and SQL Server Integration Services (SSIS).

Store: Data Lake Storage

Data Lake Storage Gen1 provides unlimited storage.

Note: Data at rest includes information that resides in persistent storage on physical media, in any digital format.

Microsoft Azure offers a variety of data storage solutions to meet different needs, including file, disk, blob, and table storage. Microsoft also provides encryption to protect Azure SQL Database, Azure Cosmos DB, and Azure Data Lake.

Prepare and Train: Azure Databricks

Azure Databricks provides enterprise-grade Azure security, including Azure Active Directory integration. With Azure Databricks, you can set up your Apache Spark environment in minutes, autoscale and collaborate on shared projects in an interactive workspace. Azure Databricks supports Python, Scala, R, Java and SQL, as well as data science frameworks and libraries including TensorFlow, PyTorch and scikit-learn.

Model and Serve: Azure Synapse Analytics

Azure Synapse Analytics/ SQL Data Warehouse stores data into relational tables with columnar storage. Azure SQL Data Warehouse connector now offers efficient and scalable structured streaming write support for SQL Data Warehouse. Access SQL Data Warehouse from Azure Databricks using the SQL Data Warehouse connector.

Note: Note: As of November 2019, Azure SQL Data Warehouse is now Azure Synapse Analytics.

Reference:

[DP-203 Exam Dumps](#) [DP-203 Exam Questions](#) [DP-203 PDF Dumps](#) [DP-203 VCE Dumps](#)

<https://www.braindump2go.com/dp-203.html>

<https://docs.microsoft.com/bs-latn-ba/azure/architecture/data-guide/technology-choices/pipeline-orchestration-data-movement>

<https://docs.microsoft.com/en-us/azure/azure-databricks/what-is-azure-databricks>

**QUESTION 148**

Drag and Drop Question

You have the following table named Employees.

first_name	last_name	hire_date	employee_type
Jane	Doe	2019-08-23	new
Ben	Smith	2017-12-15	Standard

You need to calculate the employee\_type value based on the hire\_date value.

How should you complete the Transact-SQL statement? To answer, drag the appropriate values to the correct targets. Each value 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.

Values	Answer Area
	SELECT
	*,
<input type="text" value="CASE"/>	<input type="text"/>
<input type="text" value="ELSE"/>	WHEN hire_date >= '2019-01-01' THEN 'New'
<input type="text" value="OVER"/>	<input type="text"/> 'Standard'
<input type="text" value="PARTITION BY"/>	END AS employee_type
<input type="text" value="ROW_NUMBER"/>	FROM
	employees

Answer:

Values	Answer Area
	SELECT
	*,
	<input type="text" value="CASE"/>
	WHEN hire_date >= '2019-01-01' THEN 'New'
<input type="text" value="OVER"/>	<input type="text" value="ELSE"/> 'Standard'
<input type="text" value="PARTITION BY"/>	END AS employee_type
<input type="text" value="ROW_NUMBER"/>	FROM
	employees

**Explanation:**

Box 1: CASE

CASE evaluates a list of conditions and returns one of multiple possible result expressions.

[DP-203 Exam Dumps](#)
[DP-203 Exam Questions](#)
[DP-203 PDF Dumps](#)
[DP-203 VCE Dumps](#)

<https://www.braindump2go.com/dp-203.html>

CASE can be used in any statement or clause that allows a valid expression. For example, you can use CASE in statements such as SELECT, UPDATE, DELETE and SET, and in clauses such as select\_list, IN, WHERE, ORDER BY, and HAVING.

Syntax: Simple CASE expression:

CASE input\_expression

WHEN when\_expression THEN result\_expression [ ...n ]

[ ELSE else\_result\_expression ]

END

Box 2: ELSE

Reference:

<https://docs.microsoft.com/en-us/sql/t-sql/language-elements/case-transact-sql>

#### **QUESTION 149**

Drag and Drop Question

You have an Azure Synapse Analytics workspace named WS1.

You have an Azure Data Lake Storage Gen2 container that contains JSON-formatted files in the following format.

```
{
  "id": "66532691-ab20-11ea-8b1d-936b3ec64e54",
  "context": {
    "data": {
      "eventTime": "2020-06-10T13:43:34.553Z",
      "samplingRate": "100.0",
      "isSynthetic": "false"
    },
    "session": {
      "isFirst": "false",
      "id": "38619c14-7a23-4687-8268-95862c5326b1"
    },
    "custom": {
      "dimensions": [
        {
          "customerInfo": {
            "ProfileType": "ExpertUser",
            "RoomName": "",
            "CustomerName": "diamond",
            "UserName": "XXXX@yahoo.com"
          }
        },
        {
          "customerInfo" {
            "ProfileType": "Novice",
            "RoomName": "",
            "CustomerName": "topaz",
            "UserName": "XXXX@outlook.com"
          }
        }
      ]
    }
  }
}
```

You need to use the serverless SQL pool in WS1 to read the files.  
 How should you complete the Transact-SQL statement? To answer, drag the appropriate values to the correct targets.  
 Each value 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.

Values	Answer Area
	select*
	FROM
	<input type="text"/> (
<input type="text" value="opendatasource"/>	BULK 'https://contoso.blob.core.windows.net/contosodw', FORMAT= 'CSV', fieldterminator = '0x0b', fieldquote = '0x0b', rowterminator = '0x0b'
<input type="text" value="openjson"/>	)
<input type="text" value="openquery"/>	with (id varchar(50), contextdateeventTime varchar(50) '\$.context.data.eventTime', contextdatasamplingRate varchar(50) '\$.context.data.samplingRate', contextdataisSynthetic varchar(50) '\$.context.data.isSynthetic', contextsessionisFirst varchar(50) '\$.context.session.isFirst', contextsession varchar(50) '\$.context.session.id', contextcustomdimensions varchar(max) '\$.context.custom.dimensions'
<input type="text" value="openrowset"/>	) as q cross apply <input type="text"/> (contextcustomdimensions)
	with ( ProfileType varchar(50) '\$.customerInfo.ProfileType', RoomName varchar(50) '\$.customerInfo.RoomName', CustomerName varchar(50) '\$.customerInfo.CustomerName', UserName varchar(50) '\$.customerInfo.UserName'
	)

**Answer:**

Values	Answer Area
	select*
	FROM
	<input type="text" value="openrowset"/> (
<input type="text" value="opendatasource"/>	BULK 'https://contoso.blob.core.windows.net/contosodw', FORMAT= 'CSV', fieldterminator = '0x0b', fieldquote = '0x0b', rowterminator = '0x0b'
<input type="text" value="openquery"/>	)
	with (id varchar(50), contextdateeventTime varchar(50) '\$.context.data.eventTime', contextdatasamplingRate varchar(50) '\$.context.data.samplingRate', contextdataisSynthetic varchar(50) '\$.context.data.isSynthetic', contextsessionisFirst varchar(50) '\$.context.session.isFirst', contextsession varchar(50) '\$.context.session.id', contextcustomdimensions varchar(max) '\$.context.custom.dimensions'
	) as q cross apply <input type="text" value="openjson"/> (contextcustomdimensions)
	with ( ProfileType varchar(50) '\$.customerInfo.ProfileType', RoomName varchar(50) '\$.customerInfo.RoomName', CustomerName varchar(50) '\$.customerInfo.CustomerName', UserName varchar(50) '\$.customerInfo.UserName'
	)

**Explanation:**

Box 1: openrowset

The easiest way to see to the content of your CSV file is to provide file URL to OPENROWSET function, specify csv FORMAT.

Example:

SELECT \*

FROM OPENROWSET(

BULK 'csv/population/population.csv',

DATA\_SOURCE = 'SqlOnDemandDemo',

FORMAT = 'CSV', PARSER\_VERSION = '2.0',

FIELDTERMINATOR = ',',

ROWTERMINATOR = '\n'

Box 2: openjson

You can access your JSON files from the Azure File Storage share by using the mapped drive, as shown in the following example:

```
SELECT book.* FROM
OPENROWSET(BULK N't:\books\books.json', SINGLE_CLOB) AS json CROSS APPLY OPENJSON(BulkColumn)
WITH( id nvarchar(100), name nvarchar(100), price float,
pages_i int, author nvarchar(100)) AS book
```

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/query-single-csv-file>

<https://docs.microsoft.com/en-us/sql/relational-databases/json/import-json-documents-into-sql-server>

**QUESTION 150**

Drag and Drop Question

You have an Apache Spark DataFrame named temperatures. A sample of the data is shown in the following table.

Date	Temp
...	...
18-01-2021	3
19-01-2021	4
20-01-2021	2
21-01-2021	2
...	...

You need to produce the following table by using a Spark SQL query.

Year	JAN	FEB	MAR	APR	MAY
2019	2.3	4.1	5.2	7.6	9.2
2020	2.4	4.2	4.9	7.8	9.1
2021	2.6	5.3	3.4	7.9	9.5

How should you complete the query? To answer, drag the appropriate values to the correct targets. Each value 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.

**Values      Answer Area**

```
SELECT * FROM (
  SELECT YEAR(Date) Year, MONTH(Date) Month, Temp
  FROM temperatures
  WHERE date BETWEEN DATE '2019-01-01' AND DATE '2021-08-31'
)
[ ] (
  AVG ( [ ] (Temp AS DECIMAL(4, 1)))
  FOR Month in (
    1 JAN, 2 FEB, 3 MAR, 4 APR, 5 MAY, 6 JUN,
    7 JUL, 8 AUG, 9 SEP, 10 OCT, 11 NOV, 12 DEC
  )
)
ORDER BY Year ASC
```

**Answer:**

**Values      Answer Area**

```
SELECT * FROM (  
    SELECT YEAR(Date) Year, MONTH(Date) Month, Temp  
    FROM temperatures  
    WHERE date BETWEEN DATE '2019-01-01' AND DATE '2021-08-31'  
)  
  
  
  
 PIVOT (  
    AVG (  (Temp AS DECIMAL(4, 1)))  
    FOR Month in (  
        1 JAN, 2 FEB, 3 MAR, 4 APR, 5 MAY, 6 JUN,  
        7 JUL, 8 AUG, 9 SEP, 10 OCT, 11 NOV, 12 DEC  
    )  
)  
ORDER BY Year ASC
```

**Explanation:**

Box 1: PIVOT

PIVOT rotates a table-valued expression by turning the unique values from one column in the expression into multiple columns in the output. And PIVOT runs aggregations where they're required on any remaining column values that are wanted in the final output.

Incorrect Answers:

UNPIVOT carries out the opposite operation to PIVOT by rotating columns of a table-valued expression into column values.

Box 2: CAST

If you want to convert an integer value to a DECIMAL data type in SQL Server use the CAST() function.

Example:

```
SELECT  
CAST(12 AS DECIMAL(7,2) ) AS decimal_value;
```

Here is the result:

```
decimal_value  
12.00
```

Reference:

<https://learnsql.com/cookbook/how-to-convert-an-integer-to-a-decimal-in-sql-server/>

<https://docs.microsoft.com/en-us/sql/t-sql/queries/from-using-pivot-and-unpivot>

**QUESTION 151**

Hotspot Question

The following code segment is used to create an Azure Databricks cluster.

```

{
  "num_workers": null,
  "autoscale": {
    "min_workers": 2,
    "max_workers": 8
  },
  "cluster_name": "MyCluster",
  "spark_version": "latest-stable-scala2.11",
  "spark_conf": {
    "spark.databricks.cluster.profile": "serverless",
    "spark.databricks.repl.allowedLanguages": "sql,python,r"
  },
  "node_type_id": "Standard_DS13_v2",
  "ssh_public_keys": [],
  "custom_tags": {
    "ResourceClass": "Serverless"
  },
  "spark_env_vars": {
    "PYSPARK_PYTHON": "/databricks/python3/bin/python3"
  },
  "autotermination_minutes": 90,
  "enable_elastic_disk": true,
  "init_scripts": []
}

```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.  
 NOTE: Each correct selection is worth one point.

### Answer Area

Statements	Yes	No
The Databricks cluster supports multiple concurrent users.	<input type="radio"/>	<input type="radio"/>
The Databricks cluster minimizes costs when running scheduled jobs that execute notebooks.	<input type="radio"/>	<input type="radio"/>
The Databricks cluster supports the creation of a Delta Lake table.	<input type="radio"/>	<input type="radio"/>

Answer:

## Answer Area

Statements	Yes	No
The Databricks cluster supports multiple concurrent users.	<input checked="" type="radio"/>	<input type="radio"/>
The Databricks cluster minimizes costs when running scheduled jobs that execute notebooks.	<input type="radio"/>	<input checked="" type="radio"/>
The Databricks cluster supports the creation of a Delta Lake table.	<input checked="" type="radio"/>	<input type="radio"/>

### Explanation:

Box 1: Yes

A cluster mode of 'High Concurrency' is selected, unlike all the others which are 'Standard'. This results in a worker type of Standard\_DS13\_v2.

Box 2: No

When you run a job on a new cluster, the job is treated as a data engineering (job) workload subject to the job workload pricing. When you run a job on an existing cluster, the job is treated as a data analytics (all-purpose) workload subject to all-purpose workload pricing.

Box 3: Yes

Delta Lake on Databricks allows you to configure Delta Lake based on your workload patterns.

Reference:

<https://adatis.co.uk/databricks-cluster-sizing/>

<https://docs.microsoft.com/en-us/azure/databricks/jobs>

<https://docs.databricks.com/administration-guide/capacity-planning/cmbp.html>

<https://docs.databricks.com/delta/index.html>

### QUESTION 152

Hotspot Question

You have an enterprise data warehouse in Azure Synapse Analytics that contains a table named FactOnlineSales. The table contains data from the start of 2009 to the end of 2012.

You need to improve the performance of queries against FactOnlineSales by using table partitions. The solution must meet the following requirements:

Create four partitions based on the order date.

Ensure that each partition contains all the orders places during a given calendar year.

How should you complete the T-SQL command? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

## Answer Area

```
CREATE TABLE [dbo].[FactOnlineSales]
([OnlineSalesKey] [int] NOT NULL,
[OrderDateKey] [datetime] NOT NULL,
[StoreKey] [int] NOT NULL,
[ProductKey] [int] NOT NULL,
[CustomerKey] [int] NOT NULL,
[SalesOrderNumber] [varchar](20) NOT NULL,
[SalesQuantity] [int] NOT NULL,
[SalesAmount] [money] NOT NULL,
[UnitPrice] [money] NULL)
WITH (CLUSTERED COLUMNSTORE INDEX)
PARTITION ([OrderDateKey] RANGE  FOR VALUES
```

▼
RIGHT
LEFT

```
)
```

▼
20090101,20121231
20100101,20110101,20120101
20090101,20100101,20110101,20120101

Answer:

## Answer Area

```
CREATE TABLE [dbo].[FactOnlineSales]
([OnlineSalesKey] [int] NOT NULL,
[OrderDateKey] [datetime] NOT NULL,
[StoreKey] [int] NOT NULL,
[ProductKey] [int] NOT NULL,
[CustomerKey] [int] NOT NULL,
[SalesOrderNumber] [varchar](20) NOT NULL,
[SalesQuantity] [int] NOT NULL,
[SalesAmount] [money] NOT NULL,
[UnitPrice] [money] NULL)
WITH (CLUSTERED COLUMNSTORE INDEX)
PARTITION ([OrderDateKey] RANGE 

|       |   |
|-------|---|
|       | ▼ |
| RIGHT |   |
| LEFT  |   |

 FOR VALUES 

|   |                                     |   |
|---|-------------------------------------|---|
| ( |                                     | ) |
|   | 20090101,20121231                   |   |
|   | 20100101,20110101,20120101          |   |
|   | 20090101,20100101,20110101,20120101 |   |

)
```

**Explanation:**

Range Left or Right, both are creating similar partition but there is difference in comparison  
 For example: in this scenario, when you use LEFT and 20100101,20110101,20120101 Partition will be,  
 datecol<=20100101, datecol>20100101 and datecol<=20110101, datecol>20110101 and datecol<=20120101,  
 datecol>20120101

But if you use range RIGHT and 20100101,20110101,20120101  
 Partition will be, datecol<20100101, datecol>=20100101 and datecol<20110101, datecol>=20110101 and  
 datecol<20120101, datecol>=20120101

In this example, Range RIGHT will be suitable for calendar comparison Jan 1st to Dec 31st

Reference:

<https://docs.microsoft.com/en-us/sql/t-sql/statements/create-partition-function-transact-sql?view=sql-server-ver15>

**QUESTION 153**

Hotspot Question

You have an Azure subscription that contains an Azure Data Lake Storage account. The storage account contains a data lake named DataLake1.

You plan to use an Azure data factory to ingest data from a folder in DataLake1, transform the data, and land the data in another folder.

You need to ensure that the data factory can read and write data from any folder in the DataLake1 file system. The solution must meet the following requirements:

- Minimize the risk of unauthorized user access.
- Use the principle of least privilege.
- Minimize maintenance effort.

How should you configure access to the storage account for the data factory? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

**Answer Area**

Use  to authenticate by using

Azure Active Directory (Azure AD)	a managed identity
a shared access signature (SAS)	a stored access policy
a shared key	an Authorization header

Answer:

**Answer Area**

Use  to authenticate by using

Azure Active Directory (Azure AD)	a managed identity
a shared access signature (SAS)	a stored access policy
a shared key	an Authorization header

**Explanation:**

Box 1: Azure Active Directory (Azure AD)

On Azure, managed identities eliminate the need for developers having to manage credentials by providing an identity for the Azure resource in Azure AD and using it to obtain Azure Active Directory (Azure AD) tokens.

Box 2: a managed identity

A data factory can be associated with a managed identity for Azure resources, which represents this specific data factory. You can directly use this managed identity for Data Lake Storage Gen2 authentication, similar to using your own service principal. It allows this designated factory to access and copy data to or from your Data Lake Storage Gen2.

Note: The Azure Data Lake Storage Gen2 connector supports the following authentication types.

Account key authentication

Service principal authentication

Managed identities for Azure resources authentication

Reference:

<https://docs.microsoft.com/en-us/azure/active-directory/managed-identities-azure-resources/overview>

<https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-data-lake-storage>

**QUESTION 154**

Hotspot Question

You are designing an Azure Synapse Analytics dedicated SQL pool.

Groups will have access to sensitive data in the pool as shown in the following table.

Name	Enhanced access
Executives	No access to sensitive data
Analysts	Access to in-region sensitive data
Engineers	Access to all numeric sensitive data

You have policies for the sensitive data. The policies vary by region as shown in the following table.

Region	Data considered sensitive
RegionA	Financial, Personally Identifiable Information (PII)
RegionB	Financial, Personally Identifiable Information (PII), medical
RegionC	Financial, medical

You have a table of patients for each region. The tables contain the following potentially sensitive columns.

Name	Sensitive data	Description
CardOnFile	Financial	Debit/credit card number for charges
Height	Medical	Patient's height in cm
ContactEmail	PII	Email address for secure communications

You are designing dynamic data masking to maintain compliance.

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

### Answer Area

Statements	Yes	No
Analysts in RegionA require dynamic data masking rules for [Patients_RegionA].	<input type="radio"/>	<input type="radio"/>
Engineers in RegionC require a dynamic data masking rule for [Patients_RegionA], [Height]	<input type="radio"/>	<input type="radio"/>
Engineers in RegionB require a dynamic data masking rule for [Patients_RegionB], [Height]	<input type="radio"/>	<input type="radio"/>

Answer:

### Answer Area

Statements	Yes	No
Analysts in RegionA require dynamic data masking rules for [Patients_RegionA].	<input checked="" type="radio"/>	<input type="radio"/>
Engineers in RegionC require a dynamic data masking rule for [Patients_RegionA], [Height]	<input type="radio"/>	<input checked="" type="radio"/>
Engineers in RegionB require a dynamic data masking rule for [Patients_RegionB], [Height]	<input checked="" type="radio"/>	<input type="radio"/>

**Explanation:**

<https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview>

**QUESTION 155**

Drag and Drop Question

You have an Azure Synapse Analytics SQL pool named Pool1 on a logical Microsoft SQL server named Server1.

You need to implement Transparent Data Encryption (TDE) on Pool1 by using a custom key named key1.

Which five actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

**Actions**

- Enable TDE on Pool1.
- Assign a managed identity to Server1.
- Configure key1 as the TDE protector for Server1.
- Add key1 to the Azure key vault.
- Create an Azure key vault and grant the managed identity permissions to the key vault.

**Answer Area**



**Answer:**

**Actions**

**Answer Area**

- Assign a managed identity to Server1.
- Create an Azure key vault and grant the managed identity permissions to the key vault.
- Add key1 to the Azure key vault.
- Configure key1 as the TDE protector for Server1.
- Enable TDE on Pool1.

**Explanation:**

Step 1: Assign a managed identity to Server1

You will need an existing Managed Instance as a prerequisite.

Step 2: Create an Azure key vault and grant the managed identity permissions to the vault Create Resource and setup Azure Key Vault.

Step 3: Add key1 to the Azure key vault

The recommended way is to import an existing key from a .pfx file or get an existing key from the vault. Alternatively, generate a new key directly in Azure Key Vault.

Step 4: Configure key1 as the TDE protector for Server1

Provide TDE Protector key

Step 5: Enable TDE on Pool1

Reference:

<https://docs.microsoft.com/en-us/azure/azure-sql/managed-instance/scripts/transparent-data-encryption-byok-powershell>