Day 5: More Scenario-Based Questions for Azure Data Factory

Welcome to Day 5 of our Azure Data Engineer interview questions and answers series! Continuing with the theme from Day 4, we will dive into more scenario-based questions for Azure Data Factory. These questions will further test your ability to apply Azure Data Factory to real-world data engineering challenges. Let's get started with 10 new scenario-based questions.

1. Scenario: Your company needs to copy data from a REST API endpoint to an Azure SQL Database every hour. How would you set this up in Azure Data Factory?

- **Answer:** To set up this data movement:
    1. Create a linked service for the REST API and Azure SQL Database.
    2. Create datasets for the REST API source and the SQL table destination.
    3. Create a pipeline with a Copy Data activity to move the data from the API to the SQL table.
    4. Schedule the pipeline using a schedule trigger to run every hour.

2. Scenario: You need to perform a lookup operation in Azure Data Factory to fetch a configuration value from an Azure SQL Database table and use it in subsequent activities. Describe how you would do this.

- **Answer:** To perform a lookup operation:
    1. Create a linked service and dataset for the Azure SQL Database table containing the configuration value.
    2. Add a Lookup activity in the pipeline to fetch the configuration value.
    3. Use the output of the Lookup activity in subsequent activities by referencing the lookup result in expressions.

3. Scenario: Your pipeline must process a large number of files stored in an Azure Data Lake Storage Gen2 account. How would you efficiently process these files using Azure Data Factory?

- **Answer:** To efficiently process a large number of files:
    1. Create a linked service for Azure Data Lake Storage Gen2.
    2. Create a dataset with a wildcard path to reference the files.
    3. Use a ForEach activity to iterate over the list of files.
    4. Within the ForEach activity, use a Copy Data activity or a Data Flow activity to process each file.

4. Scenario: You need to transform and load data from a SQL Server database to a Parquet file in Azure Blob Storage. Describe the steps to achieve this using Azure Data Factory.

- **Answer:** To transform and load data:
    1. Create linked services for the SQL Server database and Azure Blob Storage.
    2. Create datasets for the SQL Server table and the Parquet file.
    3. Create a pipeline with a Mapping Data Flow activity.
    4. In the Data Flow, read data from the SQL Server table, apply necessary transformations, and write the output to a Parquet file in Azure Blob Storage.

5. Scenario: You need to send an email notification if a pipeline in Azure Data Factory fails. How would you set this up?

- **Answer:** To send an email notification on pipeline failure:
    1. Set up an Azure Logic App to send email notifications.
    2. In Azure Data Factory, configure the pipeline to call the Logic App using a Web activity on failure.
    3. Pass relevant failure details to the Logic App to include in the email notification.

6. Scenario: You need to implement a data pipeline that reads data from an Azure Event Hub, processes it in real-time, and writes the results to an Azure SQL Database. Explain how you would achieve this.

- **Answer:** To implement real-time data processing:
    1. Set up an Azure Stream Analytics job to read data from the Azure Event Hub.
    2. Configure the Stream Analytics job to process the data and write the results to an Azure SQL Database.
    3. Use Azure Data Factory to orchestrate the process, ensuring that the Stream Analytics job is running and monitoring the output.

7. Scenario: You need to load data from multiple sources (e.g., SQL Server, Oracle, and flat files) into a single data warehouse in Azure Synapse Analytics. Describe your approach using Azure Data Factory.

- **Answer:** To load data from multiple sources:
    1. Create linked services for SQL Server, Oracle, flat files, and Azure Synapse Analytics.
    2. Create datasets for each source and the destination data warehouse.
    3. Create a pipeline with multiple Copy Data activities, each handling a different source.
    4. Use Data Flow activities to transform and merge the data before loading it into the Azure Synapse Analytics data warehouse.

8. Scenario: Your data pipeline must run under specific conditions, such as when a particular file is available in Azure Blob Storage. How would you configure this trigger in Azure Data Factory?

- **Answer:** To configure a trigger based on file availability:
    1. Set up an event-based trigger in Azure Data Factory.
    2. Configure the trigger to monitor the specific Azure Blob Storage location for the arrival of the file.
    3. Define the pipeline to run when the trigger condition is met.

9. Scenario: You need to create a pipeline that performs conditional data processing based on the value of a parameter passed at runtime. Explain how you would implement this in Azure Data Factory.

- **Answer:** To implement conditional data processing:
    1. Create parameters in the pipeline to receive runtime values.

2. Use If Condition activities to evaluate the parameter values.
3. Based on the condition, route the execution to different branches in the pipeline to perform the required data processing.

10. Scenario: You are required to implement a pipeline that processes daily transactional data and updates a fact table in an Azure SQL Data Warehouse, ensuring no duplicate records. Describe your approach.

- **Answer:** To implement this:
    1. Create linked services for the source of the transactional data and Azure SQL Data Warehouse.
    2. Create datasets for the source data and the destination fact table.
    3. Use a Data Flow activity to read the daily transactional data, apply necessary transformations, and deduplicate the records.
    4. Write the transformed and deduplicated data to the fact table, using an Upsert pattern to handle new and existing records.