TASK 3

Create Incremental Load Pipeline and Automate This on a Daily Basis: Design an ADF pipeline to perform incremental data loads using techniques like watermarking or change tracking. Schedule this pipeline to run daily to ensure only new or modified records are processed, optimizing performance and reducing load.

Step-by-Step Guide:

Step 1: Prerequisites

- Source SQL Server has a column like LastModifiedDate or UpdatedAt or shippedDate in each table.

Step 2: Create a Watermark Table (Metadata Table)

In Azure SQL, create a table to store the last processed timestamp for each table.

```
CREATE TABLE WatermarkControl (
    TableName NVARCHAR(100) PRIMARY KEY,
    LastWatermark DATETIME
);
```

Insert initial values (if needed):

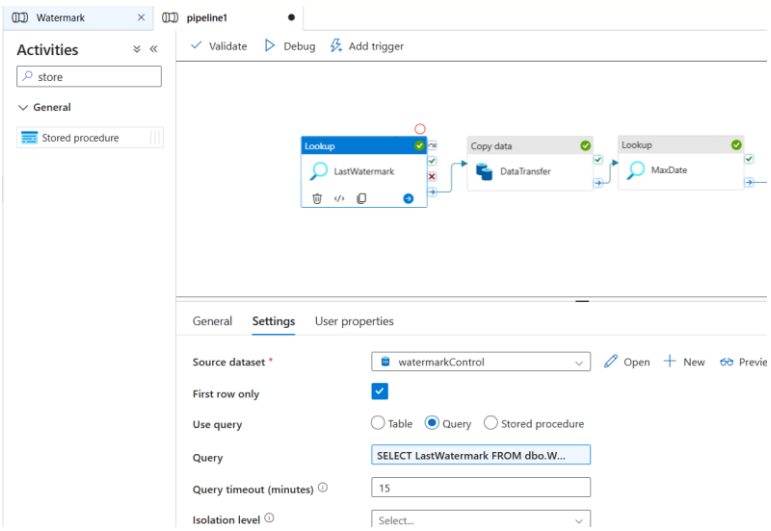
```
INSERT INTO WatermarkControl (TableName, LastWatermark)
VALUES ('sales.orders', '2000-01-01'); -- Initial load date
```

Step 3: Create ADF Pipeline

a. Lookup Activity: Get Last Watermark

- Query the control table:

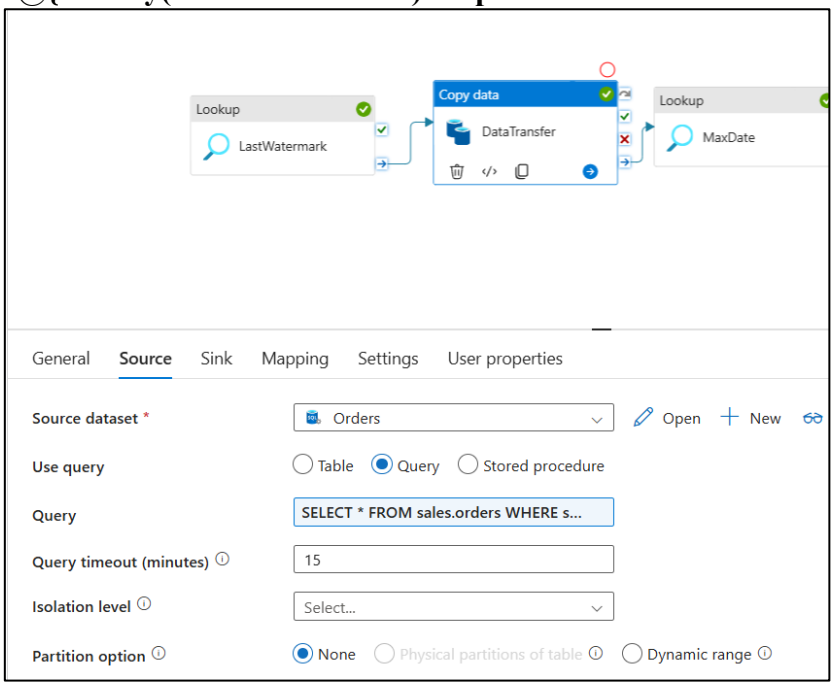
```
SELECT LastWatermark FROM dbo.WatermarkControl;
```



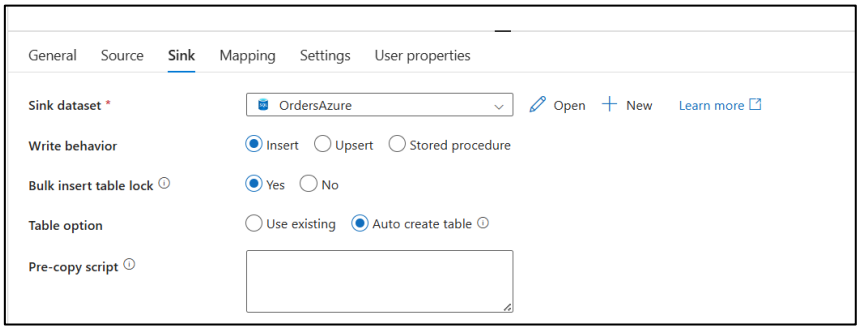
b. Copy Data Activity: Incremental Data Load

- Source Query:

SELECT * FROM sales.orders
WHERE shipped_date >
'@{activity('LastWatermark').output.firstRow.LastWatermark}';



Sink: Azure SQL.



c. Lookup Activity: Retrieve Max Date

- Source Dataset: Azure dataset
- Query:

SELECT MAX(shipped_date) AS MaxDate
FROM sales.orders
WHERE shipped_date >
'@{activity('LastWatermark').output.firstRow.LastWatermark}'

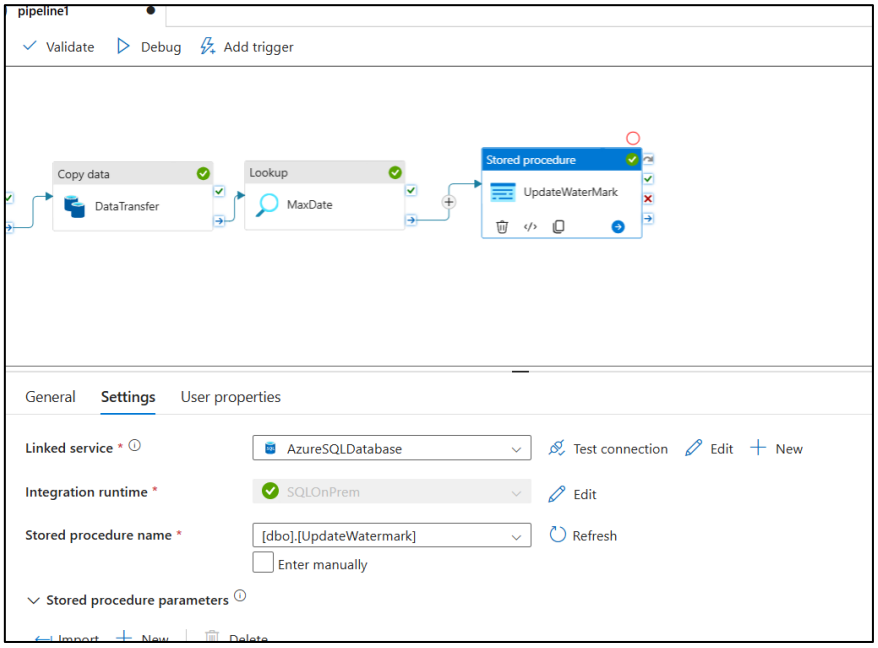
d. Stored Procedure / Script Activity: Update Watermark

- Create a procedure in the destination database in Azure SQL:

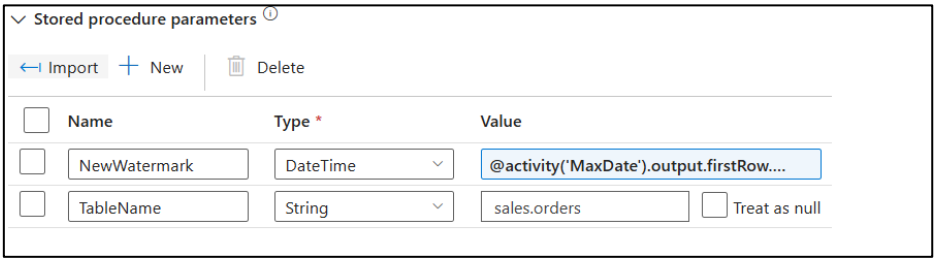
```
CREATE PROCEDURE UpdateWatermark
    @TableName NVARCHAR(100),
    @NewWatermark DATETIME
AS
BEGIN
    SET NOCOUNT ON;

    UPDATE WatermarkControl
    SET LastWatermark = @NewWatermark
    WHERE TableName = @TableName;
END
```

- Select the procedure for updating the watermark.



- Provide parameters for the procedure

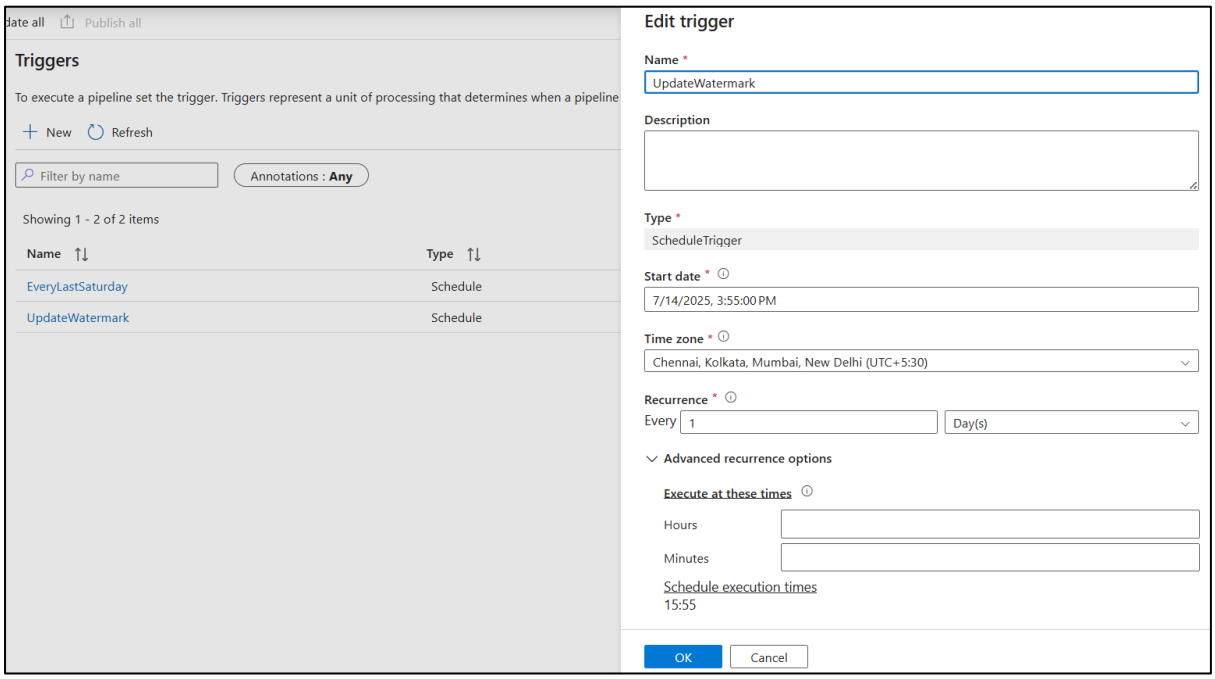


NewWatermark: **@activity('MaxDate').output.firstRow.MaxDate**

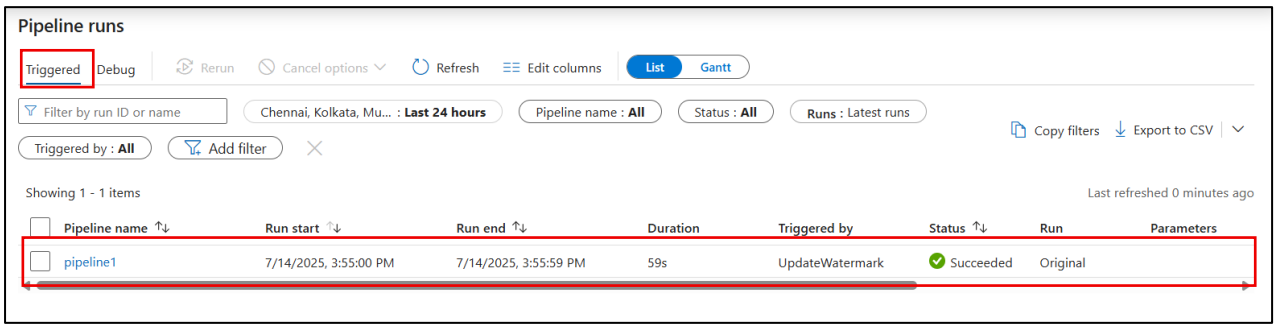
Automation: Schedule Daily Trigger

In ADF:

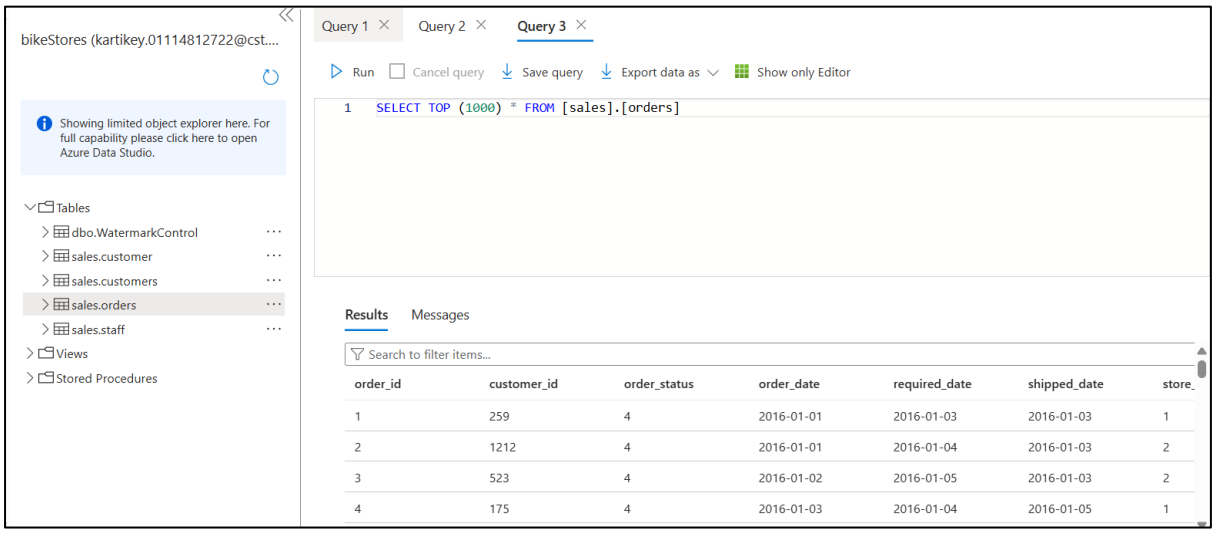
1. Go to your pipeline.
2. Click on **Add Trigger > New/Edit**.
3. Set a **Recurring trigger** with:
    - Frequency: **Day**
    - Time: <time>
    - Start date: Today
4. Link it to your pipeline.



RESULT



ORDERS table updated:



WATERMARK updated for next day: