Incremental Load → Use Case (2) :

Services required :

1. Azure sql database
2. Azure data factory
3. On prem Mysql

Azure SQL Database creation :

- Creation of the azure sql database is shown in the document below.
- https://docs.google.com/document/d/16iB1EsGKHc6-bcgTPSfqkK6BVf3n8_fpbat42uNOvXc/edit?usp=sharing

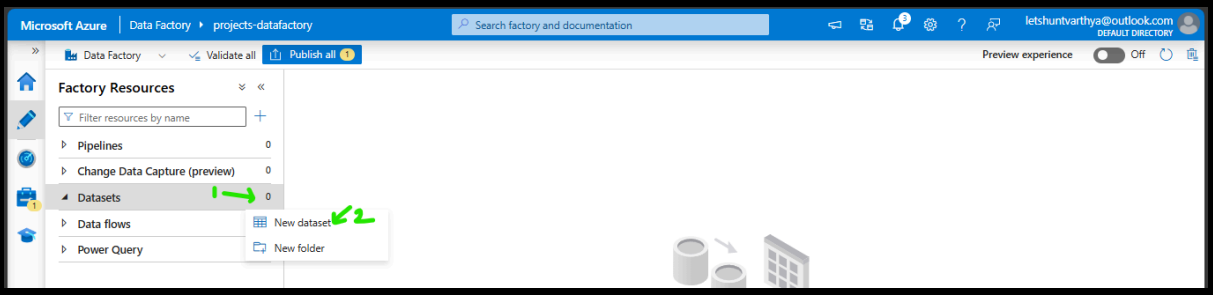
Azure Datafactory Creation :

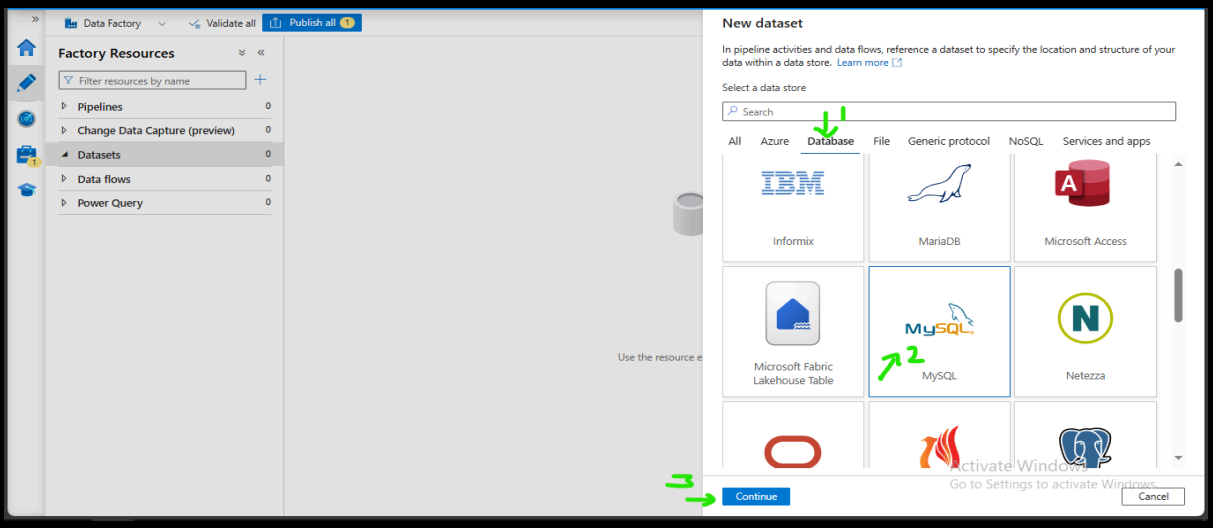
- Creation of the azure data factory is shown in the document below.
- https://docs.google.com/document/d/1IpvA7XumJjbIP0wPWWUlfd_gdvh6jTn_0a12HWv0jcM/edit?usp=sharing

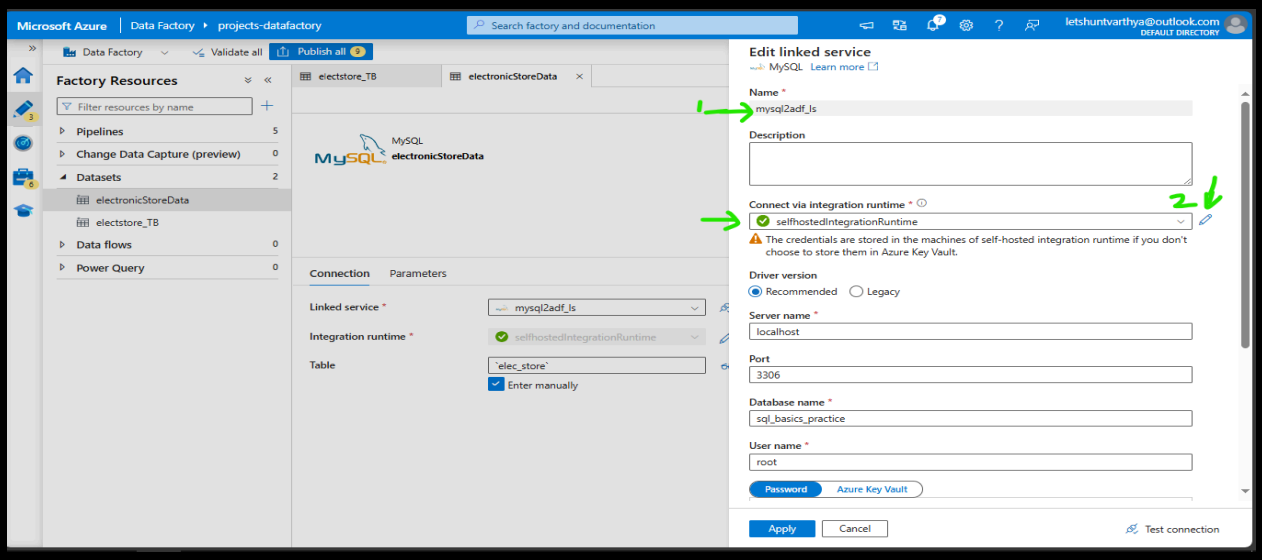
Assuming Elecstore data (initial load(15000)) is already present in the onprem mysql.
Add lastModified column in the elecstore and perform below queries on it.
→ Alter table elc_store add column lastModified date;
→ update elec_store set lastmodifiied = '2024-04-18' where country ='india' #14848
→ update elec_store set lastmodified = '2023-12-31' where country <> 'india'; # 52

Let's bring the initial load to the azure sql database.

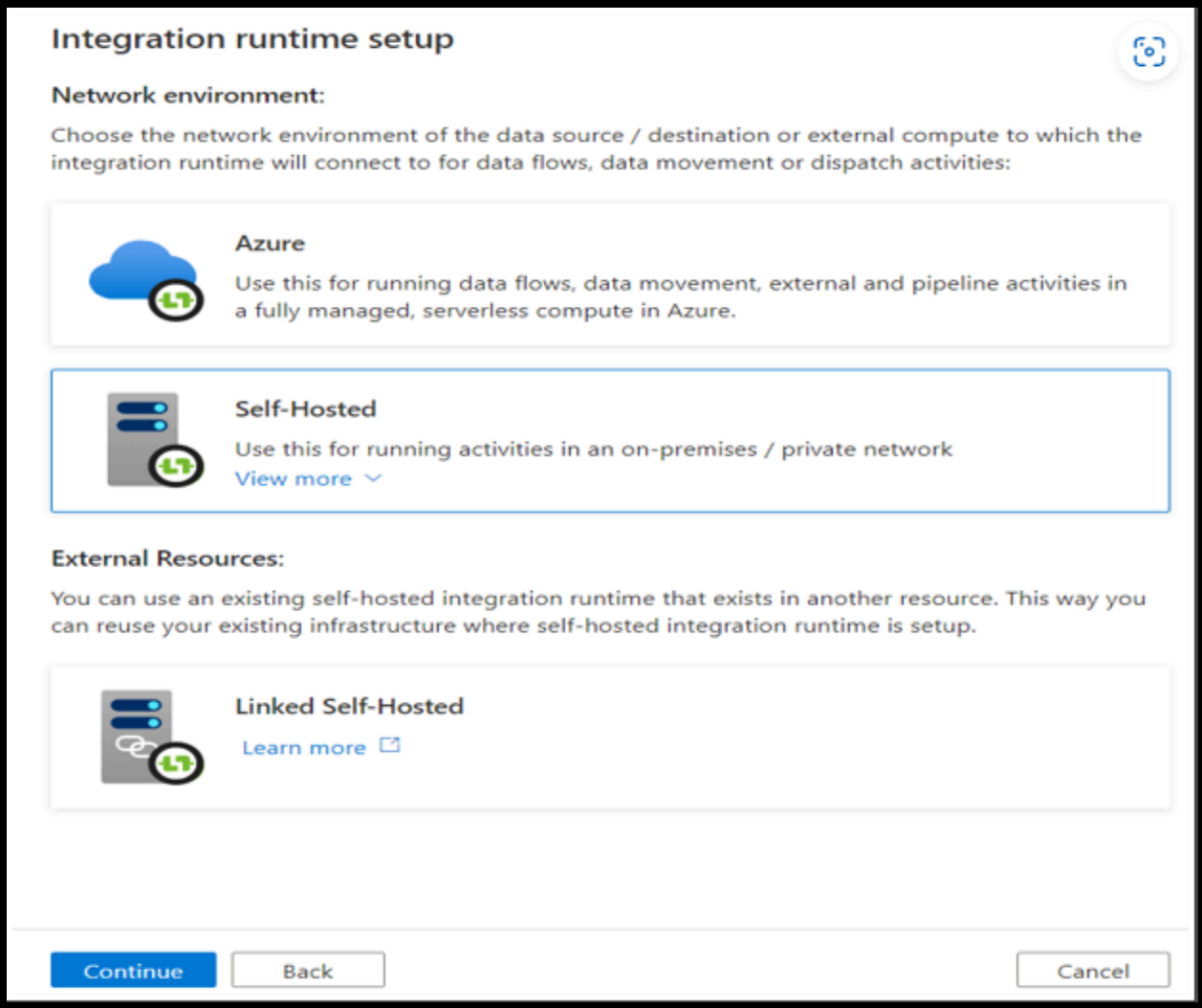
Create linked service and dataset for onprem mysql data.



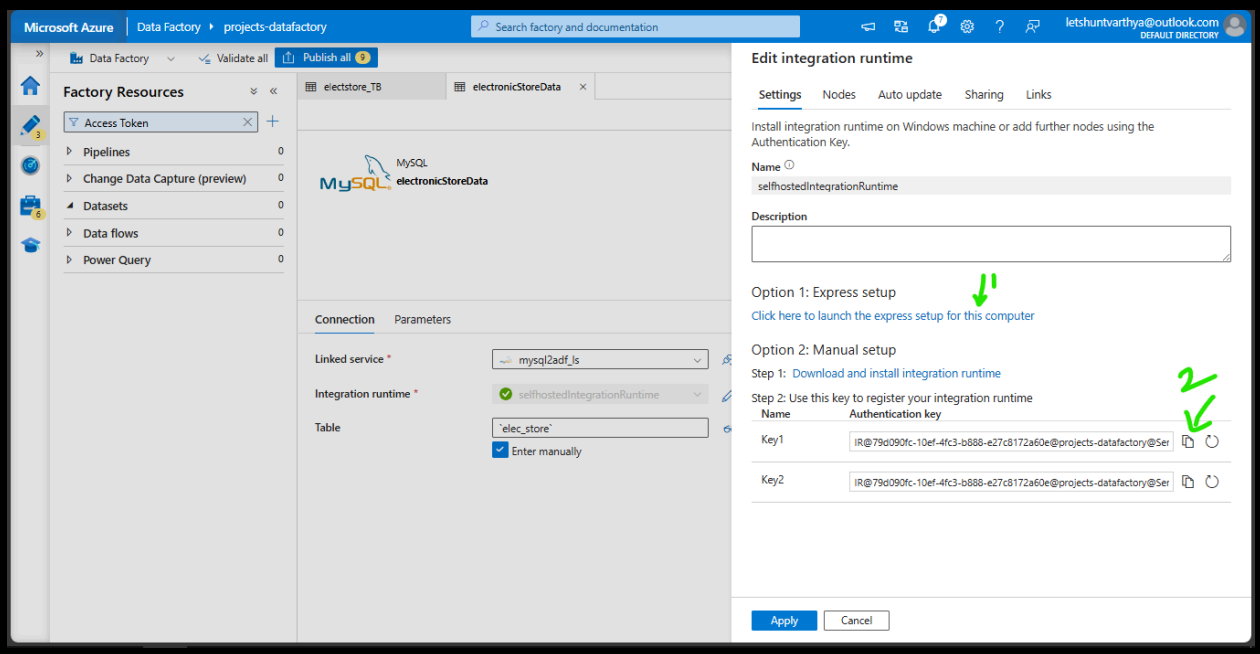


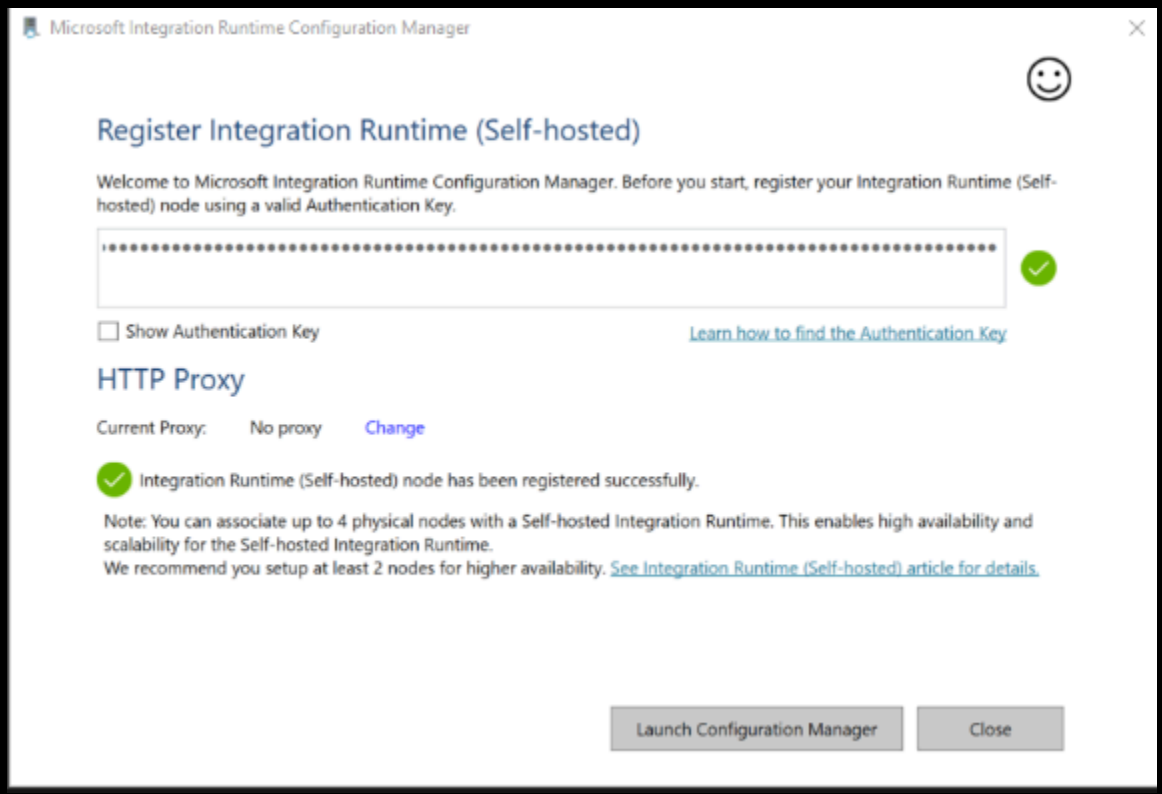


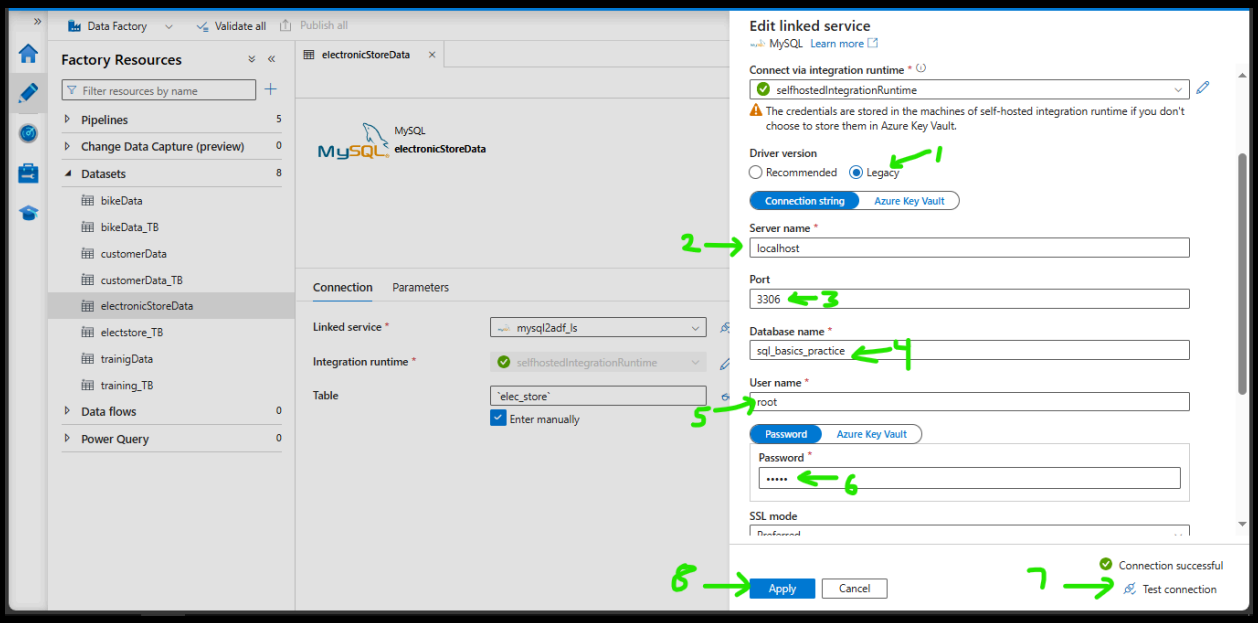
Select the self hostedIR

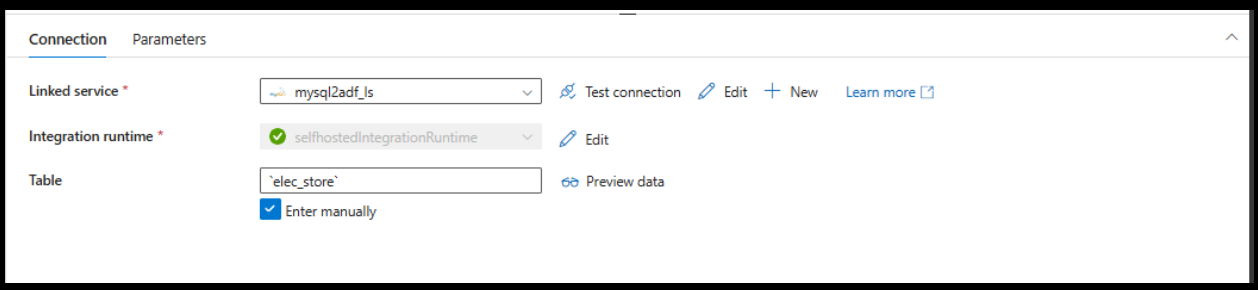


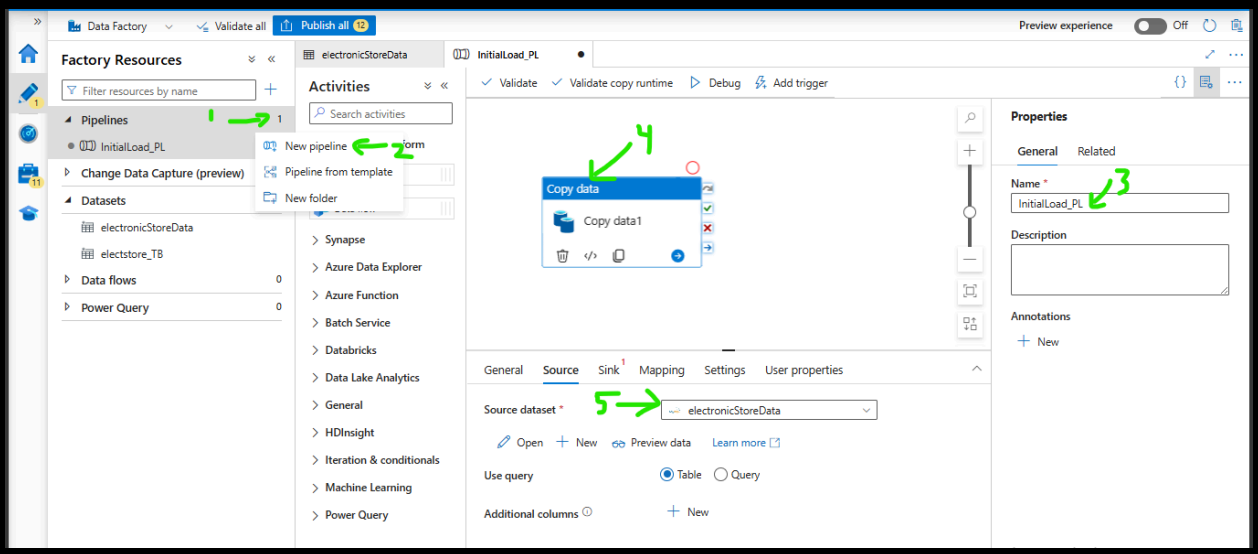
Download selfhostedIR software and paste the key 1 in selfhostedIR

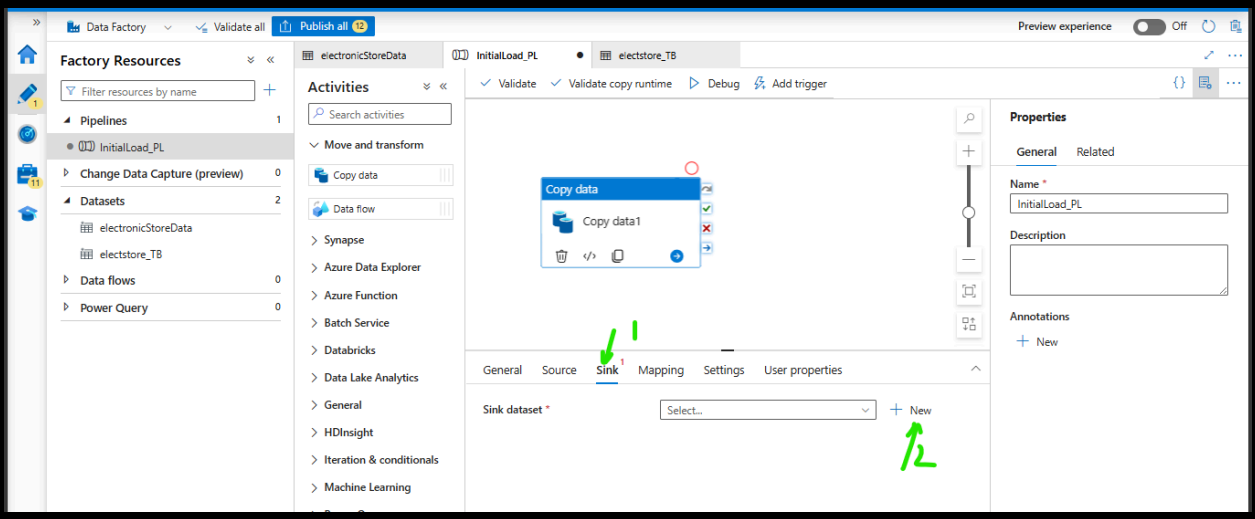

Microsoft Azure
Data Factory ‣ projects-datafactory
Search factory and documentation
letshuntvarthya@outlook.com
DEFAULT DIRECTORY
Data Factory
Validate all
Publish all
Factory Resources
Access Token
Pipelines 0
Change Data Capture (preview) 0
Datasets 0
Data flows 0
Power Query 0
electstore_TB
electronicStoreData
MySQL
electronicStoreData
Connection
Parameters
Linked service * mysql2adf_ls
Integration runtime * selfhostedIntegrationRuntime
Table `elec_store`
Enter manually
Edit integration runtime
Settings
Nodes
Auto update
Sharing
Links
Install integration runtime on Windows machine or add further nodes using the Authentication Key.
Name
selfhostedIntegrationRuntime
Description
Option 1: Express setup
Click here to launch the express setup for this computer
Option 2: Manual setup
Step 1: Download and install integration runtime
Step 2: Use this key to register your integration runtime
Name
Authentication key
Key1
IR@79d090fc-10ef-4fc3-b888-e27c8172a60e@projects-datafactory@Ser
Key2
IR@79d090fc-10ef-4fc3-b888-e27c8172a60e@projects-datafactory@Ser
Apply
Cancel

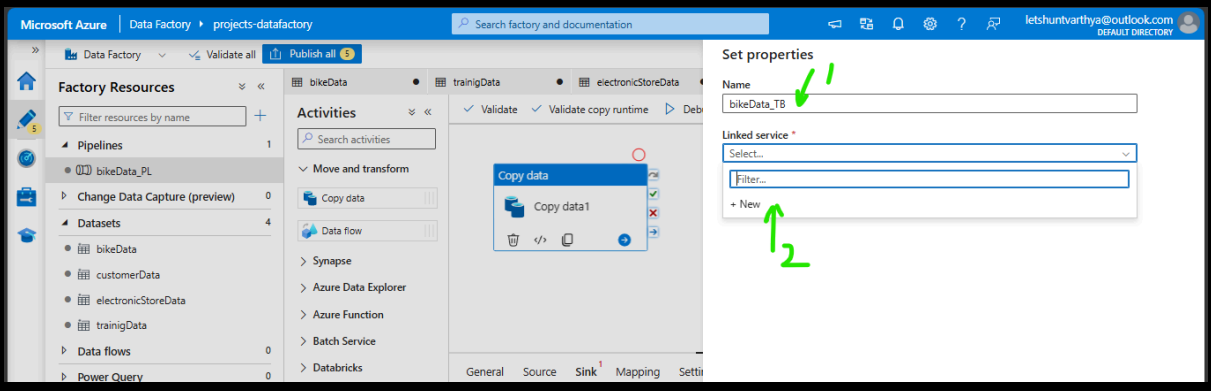

Microsoft Integration Runtime Configuration Manager
Register Integration Runtime (Self-hosted)
Welcome to Microsoft Integration Runtime Configuration Manager. Before you start, register your Integration Runtime (Self-hosted) node using a valid Authentication Key.
Show Authentication Key
Learn how to find the Authentication Key
HTTP Proxy
Current Proxy: No proxy Change
Integration Runtime (Self-hosted) node has been registered successfully.
Note: You can associate up to 4 physical nodes with a Self-hosted Integration Runtime. This enables high availability and scalability for the Self-hosted Integration Runtime.
We recommend you setup at least 2 nodes for higher availability. See Integration Runtime (Self-hosted) article for details.
Launch Configuration Manager
Close

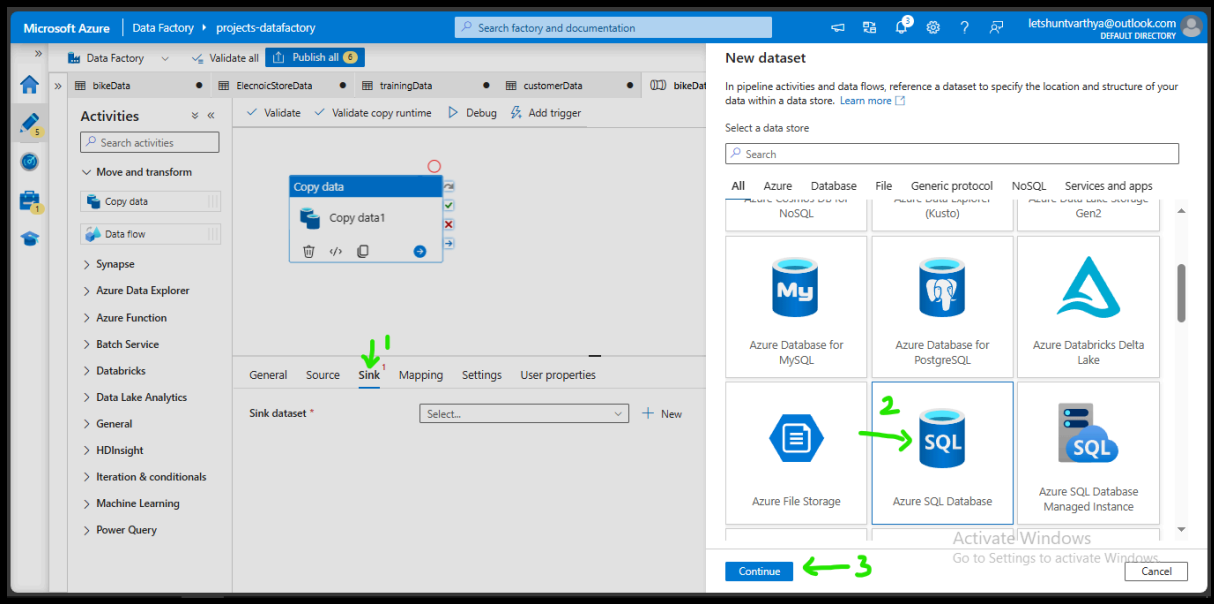

Data Factory
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines 5
Change Data Capture (preview) 0
Datasets 8
bikeData
bikeData_TB
customerData
customerData_TB
electronicStoreData
electstore_TB
trainigData
training_TB
Data flows 0
Power Query 0
electronicStoreData
MySQL
electronicStoreData
Connection
Parameters
Linked service * mysql2adf_ls
Integration runtime * selfhostedIntegrationRuntime
Table `elec_store`
Enter manually
Edit linked service
MySQL Learn more
Connect via integration runtime *
selfhostedIntegrationRuntime
The credentials are stored in the machines of self-hosted integration runtime if you don't choose to store them in Azure Key Vault.
Driver version
Recommended
Legacy
1
Connection string
Azure Key Vault
Server name *
2
localhost
Port
3306
3
Database name *
sql_basics_practice
4
User name *
root
5
Password
Azure Key Vault
Password *
6
SSL mode
8
Apply
Cancel
7
Connection successful
Test connection

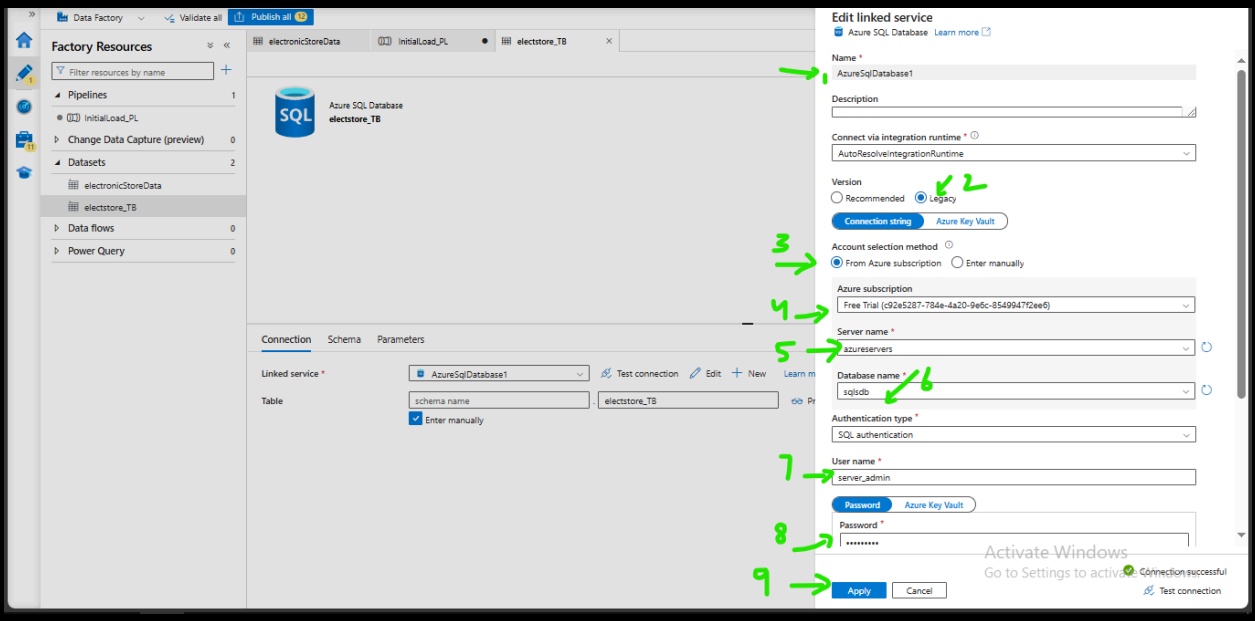

Connection
Parameters
Linked service * mysql2adf_ls
Test connection
Edit
New
Learn more
Integration runtime * selfhostedIntegrationRuntime
Edit
Table `elec_store`
Preview data
Enter manually

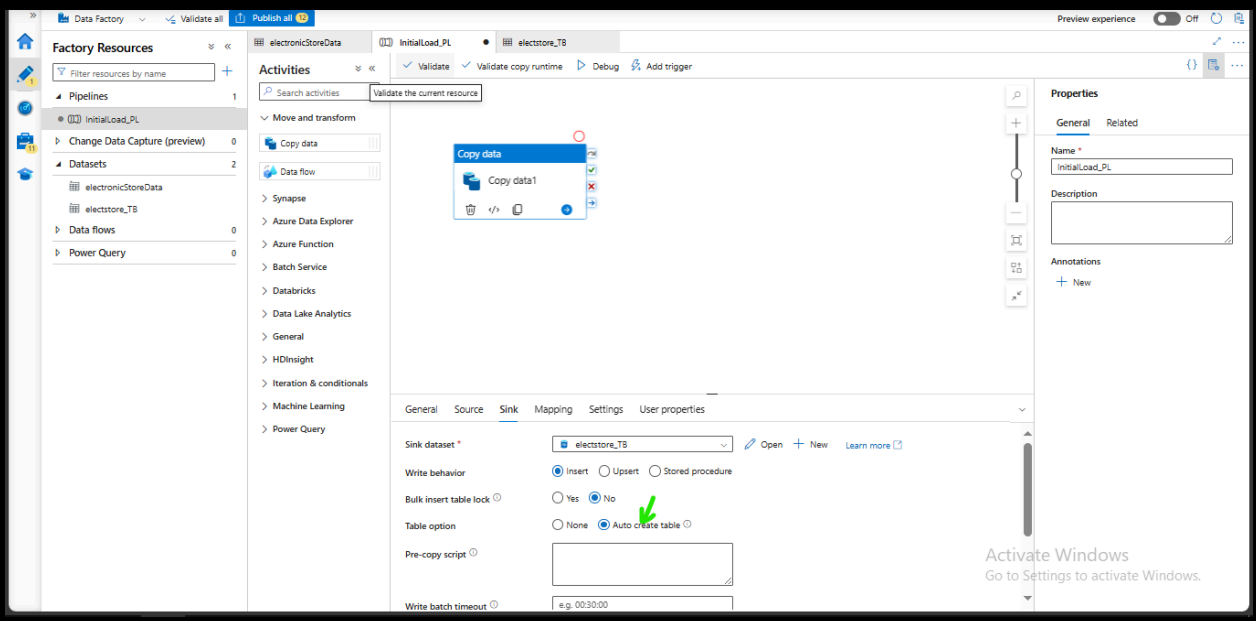

Data Factory
Validate all
Publish all 12
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 1
InitialLoad_PL
Change Data Capture (preview)
Datasets
electronicStoreData
electstore_TB
Data flows 0
Power Query 0
New pipeline
Pipeline from template
New folder
Activities
Search activities
Synapse
Azure Data Explorer
Azure Function
Batch Service
Databricks
Data Lake Analytics
General
HDInsight
Iteration & conditionals
Machine Learning
Power Query
electronicStoreData
InitialLoad_PL
Validate
Validate copy runtime
Debug
Add trigger
Copy data
Copy data1
Properties
General
Related
Name * InitialLoad_PL
Description
Annotations
New
General
Source
Sink 1
Mapping
Settings
User properties
Source dataset * electronicStoreData
Open
New
Preview data
Learn more
Use query
Table
Query
Additional columns
New

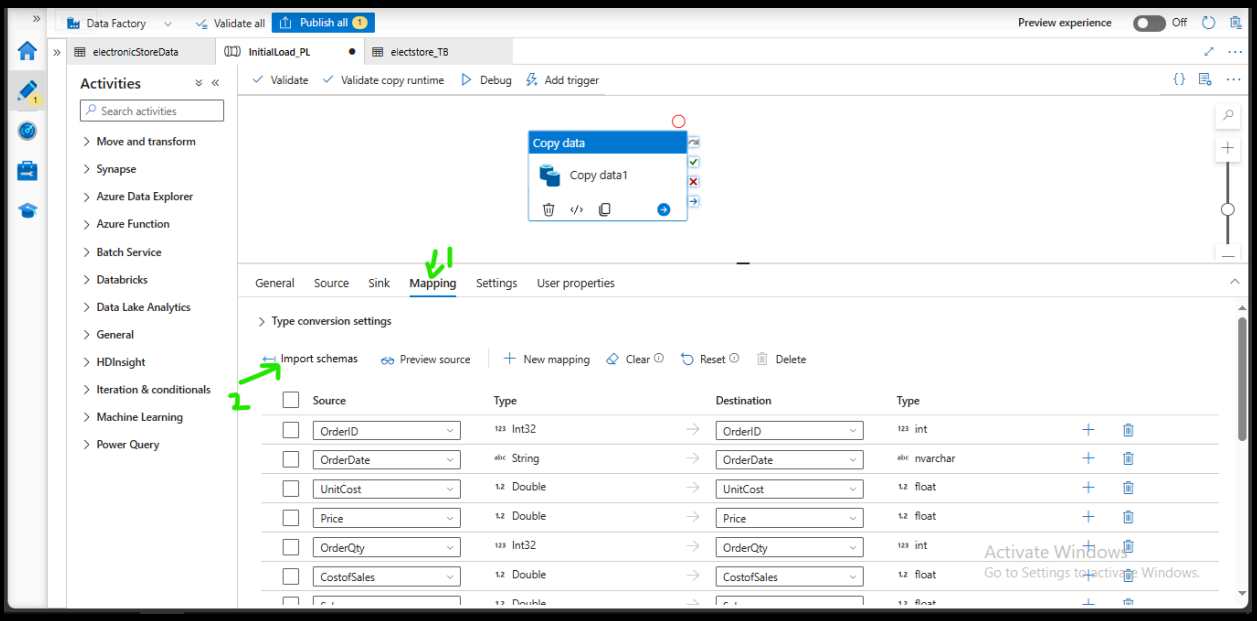

Data Factory
Validate all
Publish all 12
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 1
InitialLoad_PL
Change Data Capture (preview) 0
Datasets 2
electronicStoreData
electstore_TB
Data flows 0
Power Query 0
electronicStoreData
InitialLoad_PL
electstore_TB
Activities
Search activities
Move and transform
Copy data
Data flow
Synapse
Azure Data Explorer
Azure Function
Batch Service
Databricks
Data Lake Analytics
General
HDInsight
Iteration & conditionals
Machine Learning
Validate
Validate copy runtime
Debug
Add trigger
Copy data
Copy data1
Properties
General
Related
Name *
InitialLoad_PL
Description
Annotations
New
General
Source
Sink
Mapping
Settings
User properties
Sink dataset *
Select...
New
1
2


Microsoft Azure
Data Factory
projects-datafactory
Search factory and documentation
letshuntvarthya@outlook.com
DEFAULT DIRECTORY
Data Factory
Validate all
Publish all 5
Factory Resources
Filter resources by name
Pipelines 1
bikeData_PL
Change Data Capture (preview) 0
Datasets 4
bikeData
customerData
electronicStoreData
trainigData
Data flows 0
Power Query 0
bikeData
trainigData
electronicStoreData
Activities
Search activities
Move and transform
Copy data
Data flow
Synapse
Azure Data Explorer
Azure Function
Batch Service
Databricks
Validate
Validate copy runtime
Copy data
Copy data1
General
Source
Sink
Mapping
Set properties
Name
bikeData_TB
Linked service *
Select...
Filter...
+ New
1
2


Microsoft Azure
Data Factory
projects-datafactory
Search factory and documentation
letshuntvarthya@outlook.com
DEFAULT DIRECTORY
Data Factory
Validate all
Publish all 6
bikeData
ElecnoicStoreData
trainingData
customerData
Activities
Search activities
Move and transform
Copy data
Data flow
Synapse
Azure Data Explorer
Azure Function
Batch Service
Databricks
Data Lake Analytics
General
HDInsight
Iteration & conditionals
Machine Learning
Power Query
Validate
Validate copy runtime
Debug
Add trigger
Copy data
Copy data1
General
Source
Sink
Mapping
Settings
User properties
Sink dataset *
Select...
New
New dataset
In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. Learn more
Select a data store
Search
All
Azure
Database
File
Generic protocol
NoSQL
Services and apps
Azure Database for MySQL
Azure Database for PostgreSQL
Azure Databricks Delta Lake
Azure File Storage
Azure SQL Database
Azure SQL Database Managed Instance
Activate Windows
Go to Settings to activate Windows.
Continue
Cancel
1
2
3

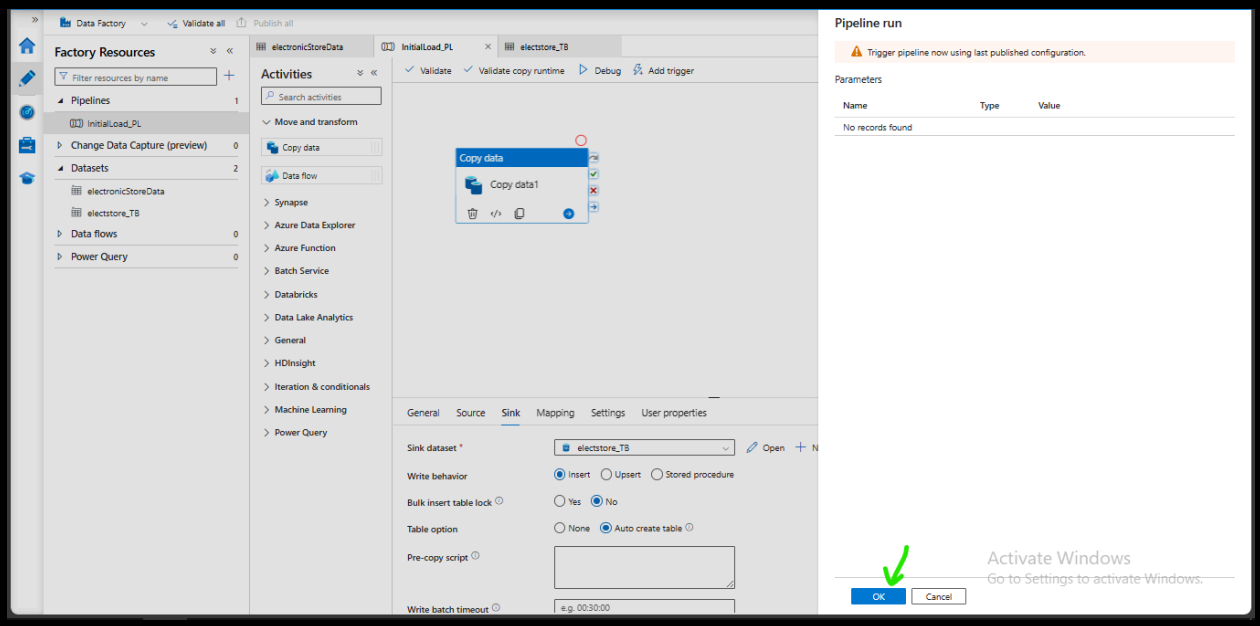

Data Factory
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines
InitialLoad_PL
Change Data Capture (preview)
Datasets
electronicStoreData
electstore_TB
Data flows
Power Query
Azure SQL Database
electstore_TB
Connection
Schema
Parameters
Linked service *
AzureSqlDatabase1
Test connection
Edit
New
Table
schema name
electstore_TB
Enter manually
Edit linked service
Azure SQL Database
Learn more
Name *
AzureSqlDatabase1
Description
Connect via integration runtime *
AutoResolveIntegrationRuntime
Version
Recommended
Legacy
Connection string
Azure Key Vault
Account selection method
From Azure subscription
Enter manually
Azure subscription
Free Trial (c92e5287-784e-4a20-9e6c-8549947f2ee6)
Server name *
azureservers
Database name *
sqlsdb
Authentication type *
SQL authentication
User name *
server_admin
Password
Azure Key Vault
Password *
Connection successful
Test connection
Apply
Cancel
1
2
3
4
5
6
7
8
9

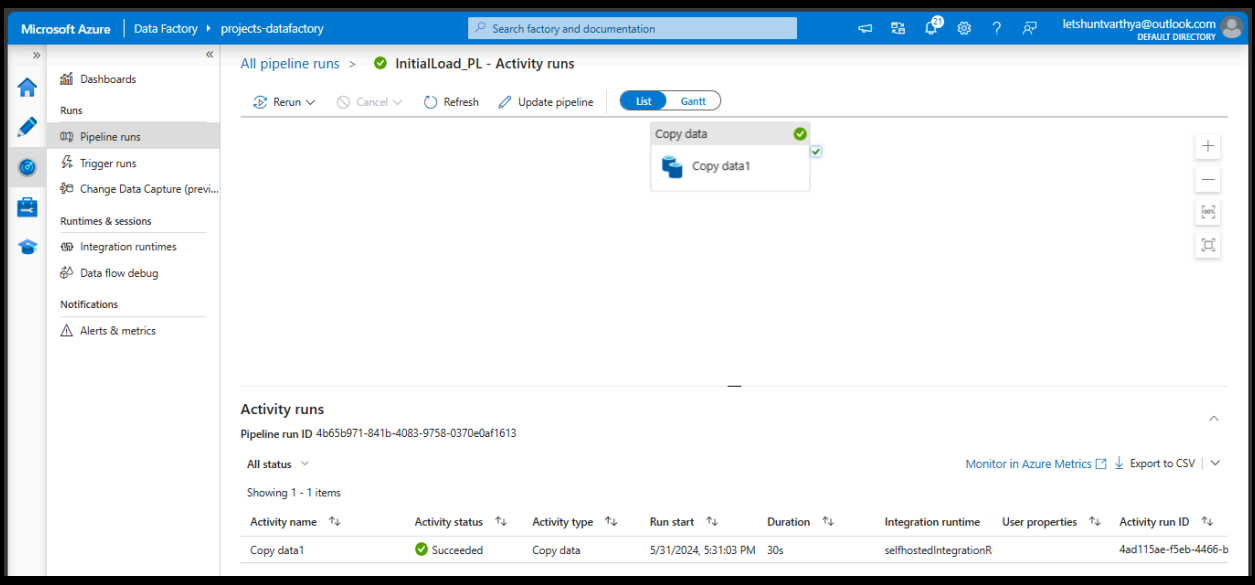

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines
InitialLoad_PL
Change Data Capture (preview)
Datasets
electronicStoreData
electstore_TB
Data flows
Power Query
Activities
Search activities
Validate the current resource
Move and transform
Copy data
Data flow
Synapse
Azure Data Explorer
Azure Function
Batch Service
Databricks
Data Lake Analytics
General
HDInsight
Iteration & conditionals
Machine Learning
Power Query
Validate
Validate copy runtime
Debug
Add trigger
Copy data
Copy data1
Properties
General
Related
Name *
InitialLoad_PL
Description
Annotations
New
General
Source
Sink
Mapping
Settings
User properties
Sink dataset *
electstore_TB
Open
New
Learn more
Write behavior
Insert
Upsert
Stored procedure
Bulk insert table lock
Yes
No
Table option
None
Auto create table
Pre-copy script
Write batch timeout
e.g. 00:30:00

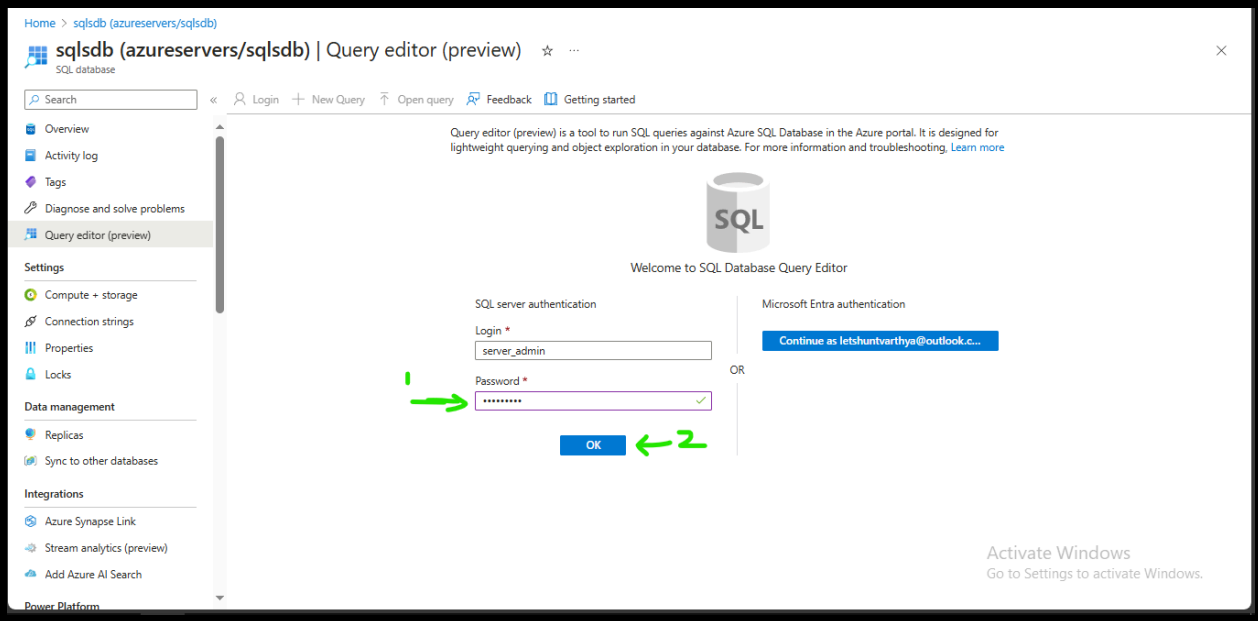


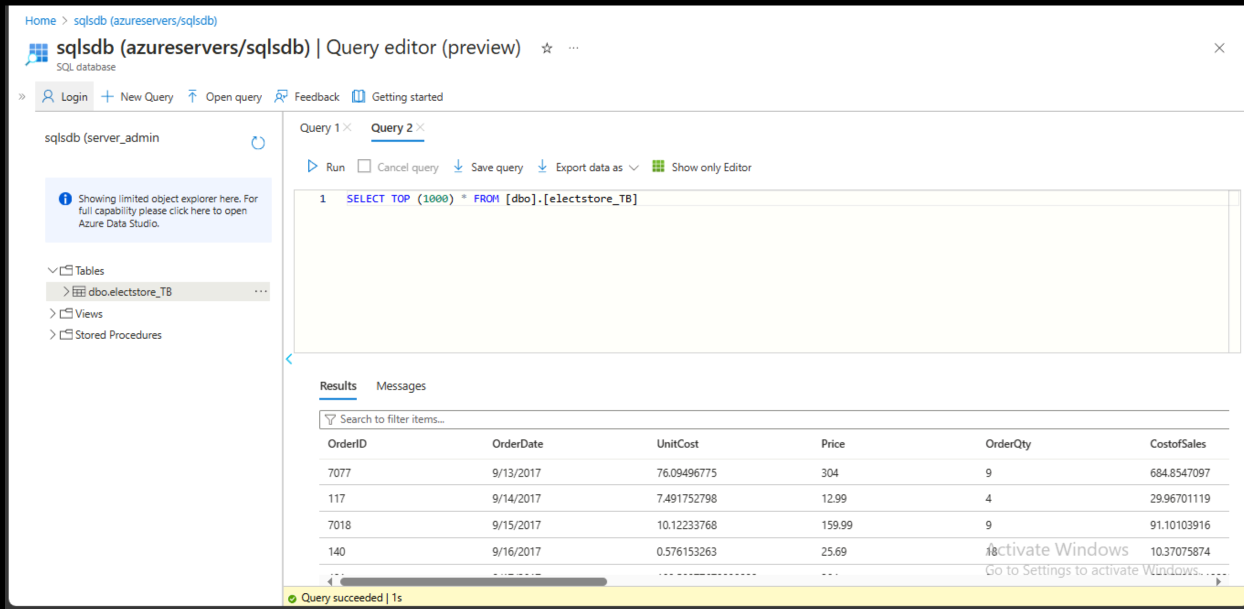
Publish all → publish

Add trigger → trigger now




Microsoft Azure
Data Factory
projects-datafactory
Search factory and documentation
letshuntvarthya@outlook.com
DEFAULT DIRECTORY
Dashboards
Runs
Pipeline runs
Trigger runs
Change Data Capture (previ...
Runtimes & sessions
Integration runtimes
Data flow debug
Notifications
Alerts & metrics
All pipeline runs
InitialLoad_PL - Activity runs
Rerun
Cancel
Refresh
Update pipeline
List
Gantt
Copy data
Copy data1
Activity runs
Pipeline run ID 4b65b971-841b-4083-9758-0370e0af1613
All status
Monitor in Azure Metrics
Export to CSV
Showing 1 - 1 items
Activity name
Activity status
Activity type
Run start
Duration
Integration runtime
User properties
Activity run ID
Copy data1
Succeeded
Copy data
5/31/2024, 5:31:03 PM
30s
selfhostedIntegrationR
4ad115ae-f5eb-4466-b


Home
sqlsdb (azureservers/sqlsdb)
sqlsdb (azureservers/sqlsdb) | Query editor (preview)
SQL database
Search
Login
New Query
Open query
Feedback
Getting started
Overview
Activity log
Tags
Diagnose and solve problems
Query editor (preview)
Settings
Compute + storage
Connection strings
Properties
Locks
Data management
Replicas
Sync to other databases
Integrations
Azure Synapse Link
Stream analytics (preview)
Add Azure AI Search
Power Platform
Query editor (preview) is a tool to run SQL queries against Azure SQL Database in the Azure portal. It is designed for lightweight querying and object exploration in your database. For more information and troubleshooting, Learn more
SQL
Welcome to SQL Database Query Editor
SQL server authentication
Login *
server_admin
Password *
OK
OR
Microsoft Entra authentication
Continue as letshuntvarthya@outlook.c...
1
2
Activate Windows
Go to Settings to activate Windows.



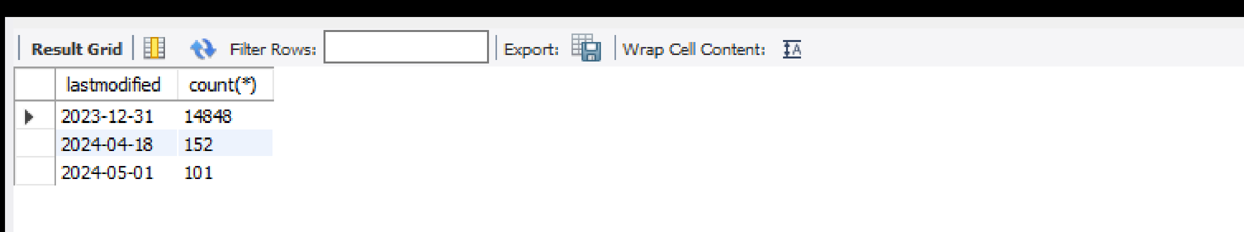
We have brought the initial load to the azure sql database.

Now let's add some data or rows into the initial load in onPrem mysql.

Initial load = 15000

Incremental load = 101

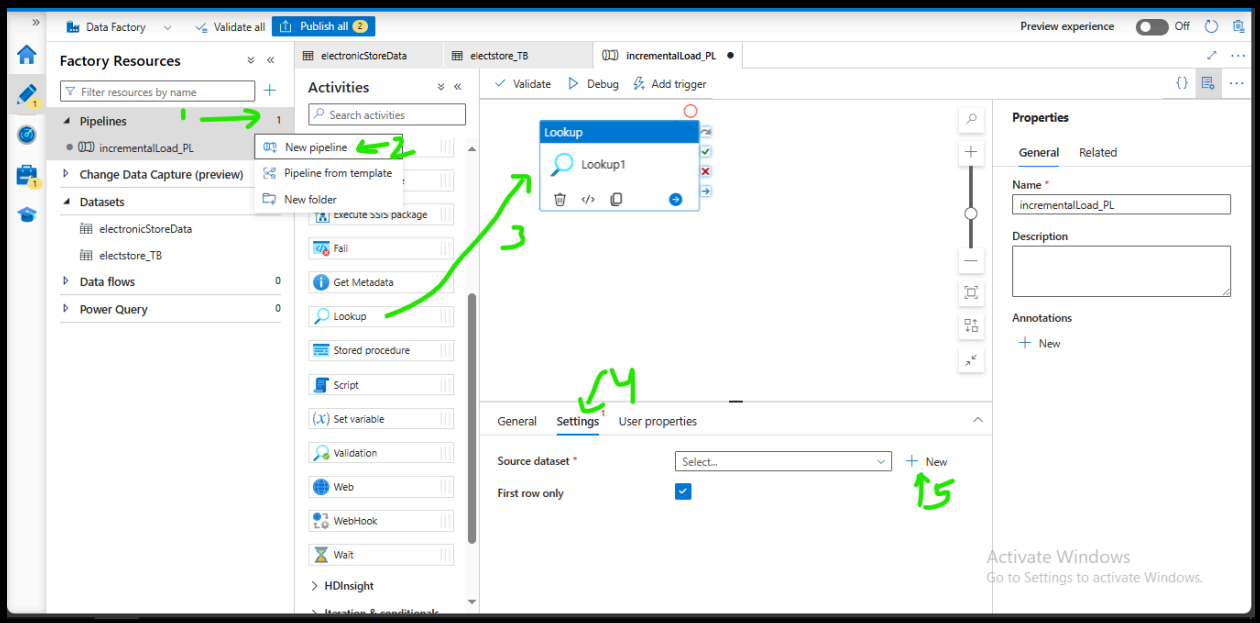
Extra 101 rows where we will add is called as incremental load

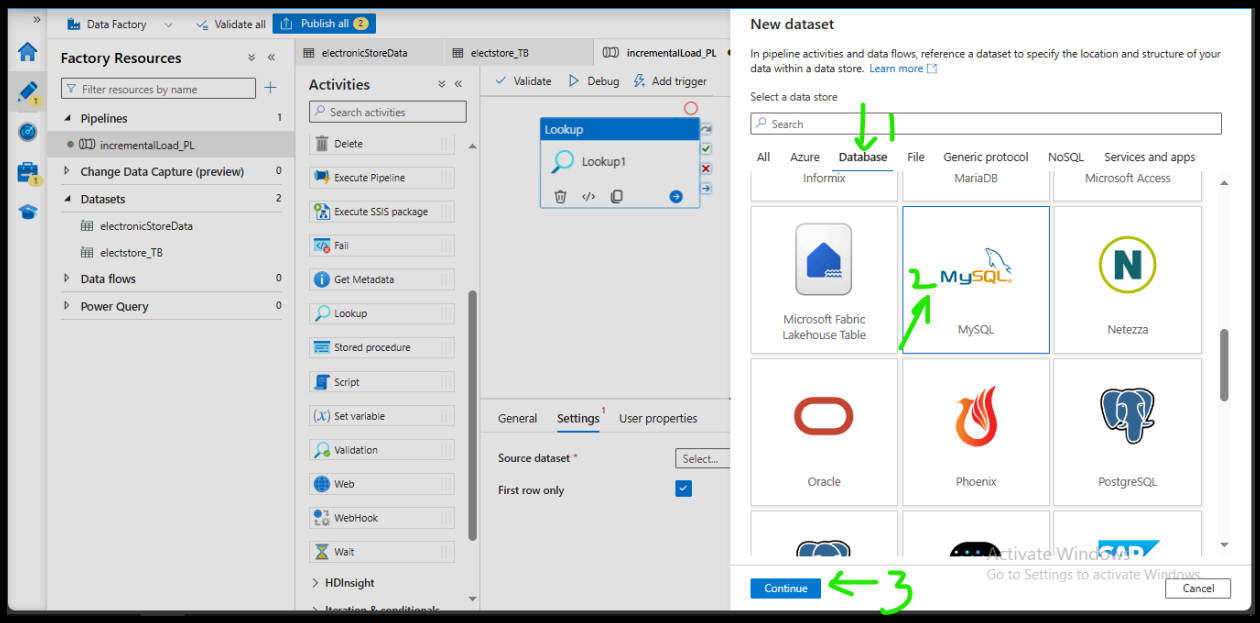


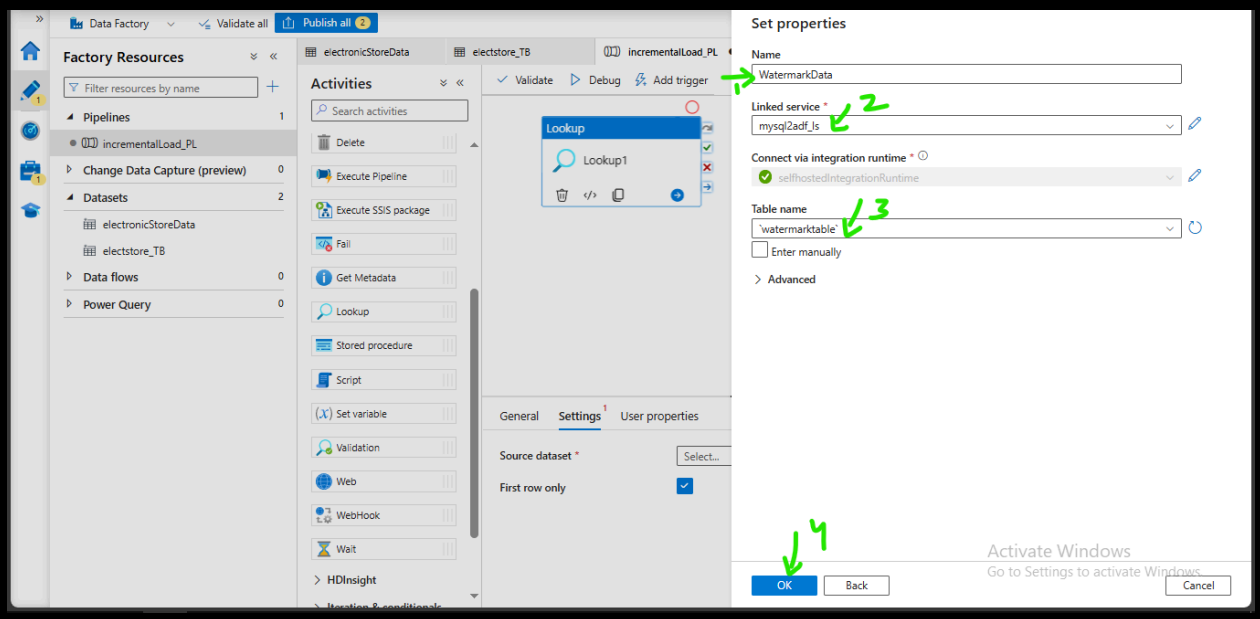
Create a watermark table in mysql which tracks the last updateddate.

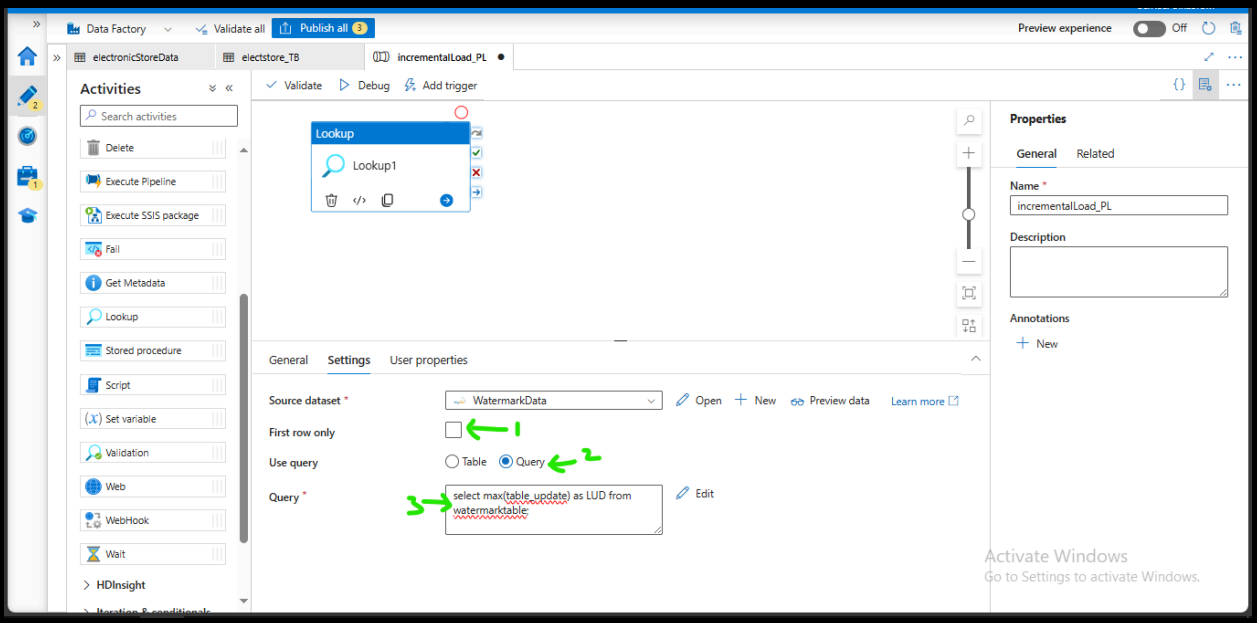
```sql
create table watermarktable (
tablename char(20),
table_update date);

select * from watermarktable;

insert into watermarktable values ('elec_store','2024-12-31');
insert into watermarktable values ('elec_store','2024-04-18');
insert into watermarktable values ('elec_store','2024-05-01');
```

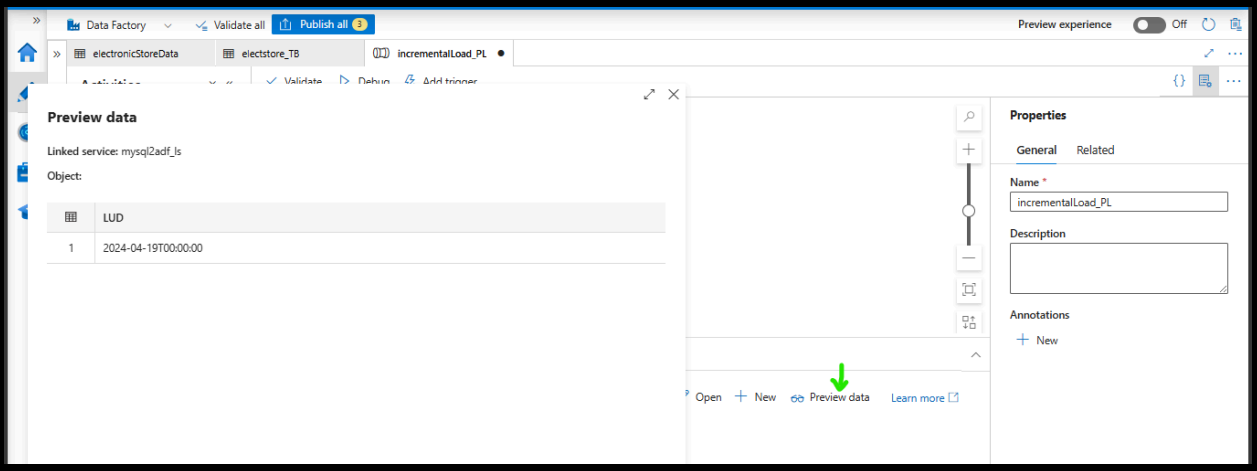
Let's load these 28 records into the Azure SQL Database.



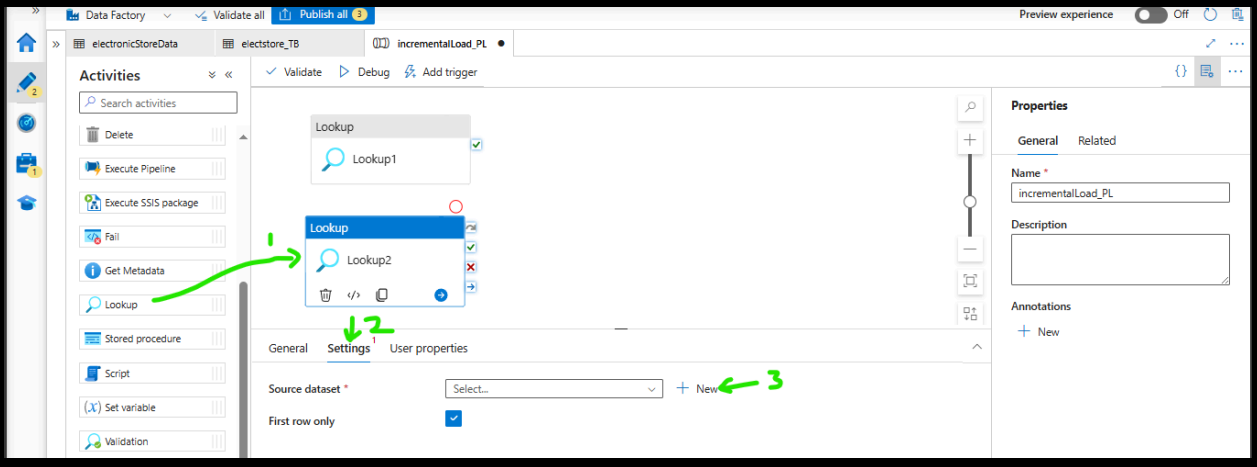

Data Factory
Validate all
Publish all 2
Factory Resources
Filter resources by name
Pipelines 1
incrementalLoad_PL
Change Data Capture (preview) 0
Datasets 2
electronicStoreData
electstore_TB
Data flows 0
Power Query 0
electronicStoreData
electstore_TB
incrementalLoad_PL
Activities
Search activities
Delete
Execute Pipeline
Execute SSIS package
Fail
Get Metadata
Lookup
Stored procedure
Script
Set variable
Validation
Web
WebHook
Wait
HDInsight
Validate
Debug
Add trigger
Lookup
Lookup1
General
Settings 1
User properties
Source dataset *
Select...
First row only
New dataset
In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. Learn more
Select a data store
Search
All
Azure
Database
File
Generic protocol
NoSQL
Services and apps
Informix
MariaDB
Microsoft Access
Microsoft Fabric Lakehouse Table
MySQL
Netezza
Oracle
Phoenix
PostgreSQL
Continue
Cancel
1
2
3

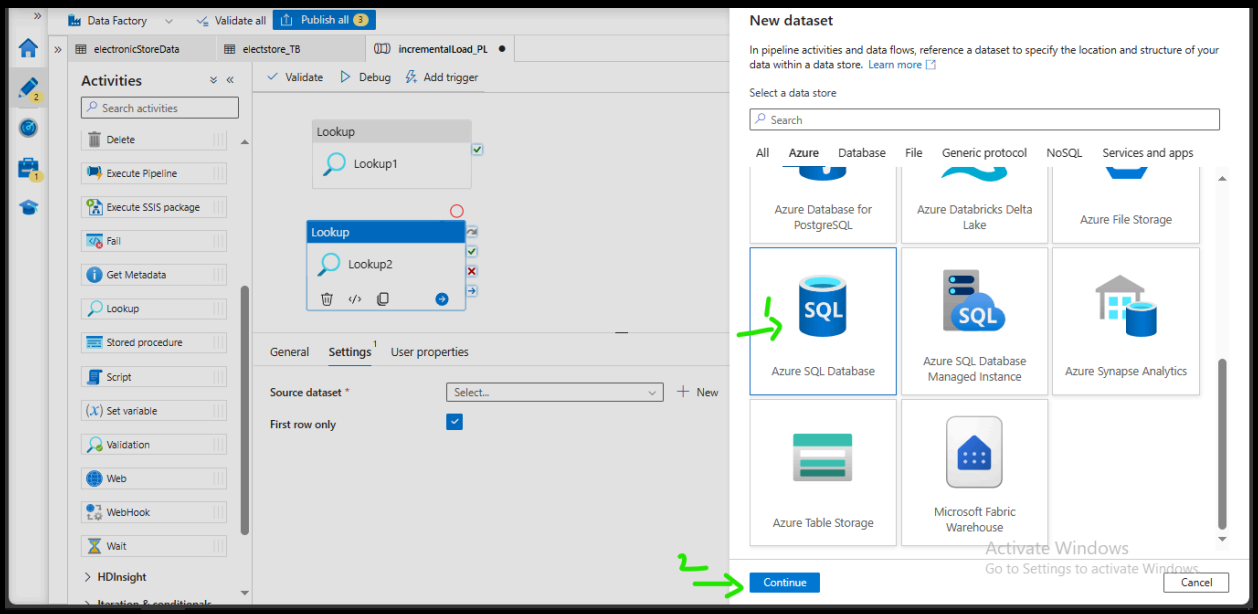

Data Factory
Validate all
Publish all 2
Factory Resources
Filter resources by name
Pipelines 1
incrementalLoad_PL
Change Data Capture (preview) 0
Datasets 2
electronicStoreData
electstore_TB
Data flows 0
Power Query 0
electronicStoreData
electstore_TB
incrementalLoad_PL
Activities
Search activities
Delete
Execute Pipeline
Execute SSIS package
Fail
Get Metadata
Lookup
Stored procedure
Script
Set variable
Validation
Web
WebHook
Wait
HDInsight
Validate
Debug
Add trigger
Lookup
Lookup1
General
Settings 1
User properties
Source dataset *
Select...
First row only
Set properties
Name
WatermarkData
Linked service *
mysql2adf_ls
Connect via integration runtime *
selfhostedIntegrationRuntime
Table name
watermarktable
Enter manually
Advanced
OK
Back
Cancel
1
2
3
4

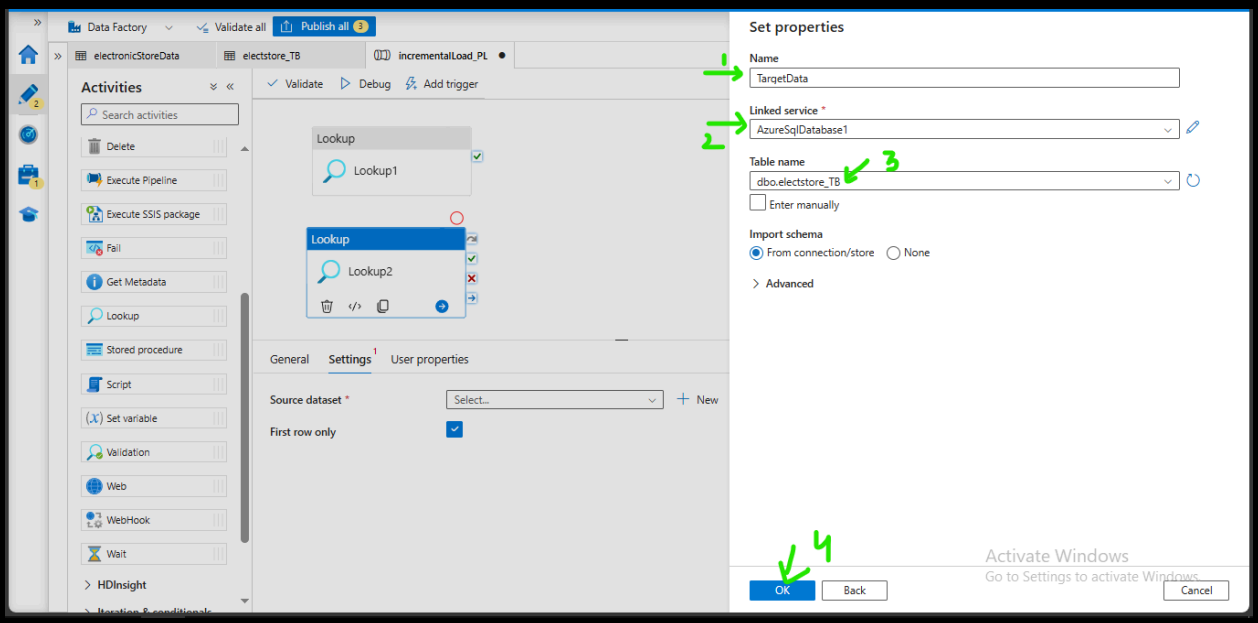


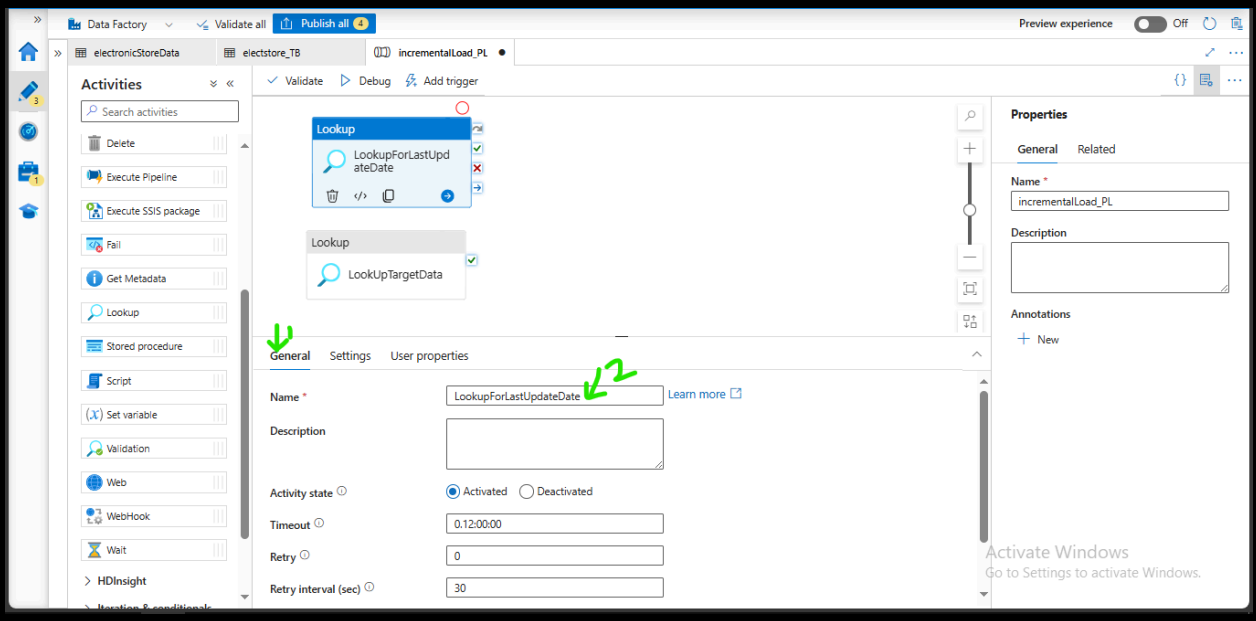


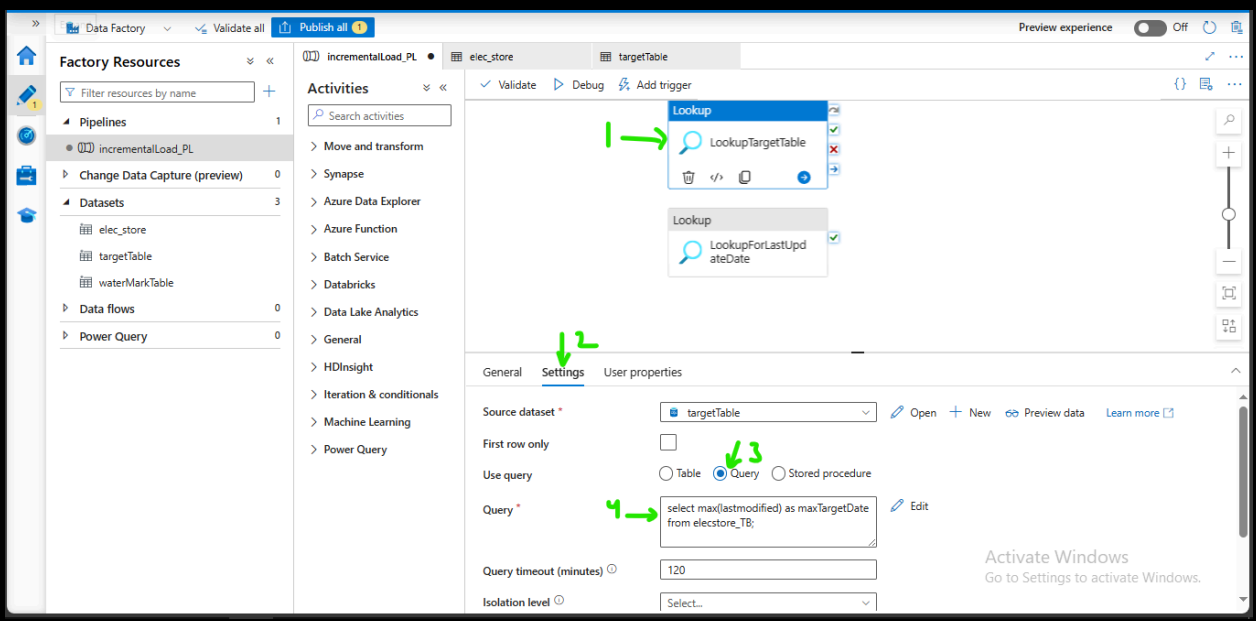
Drag and drop another lookup activity to the field.

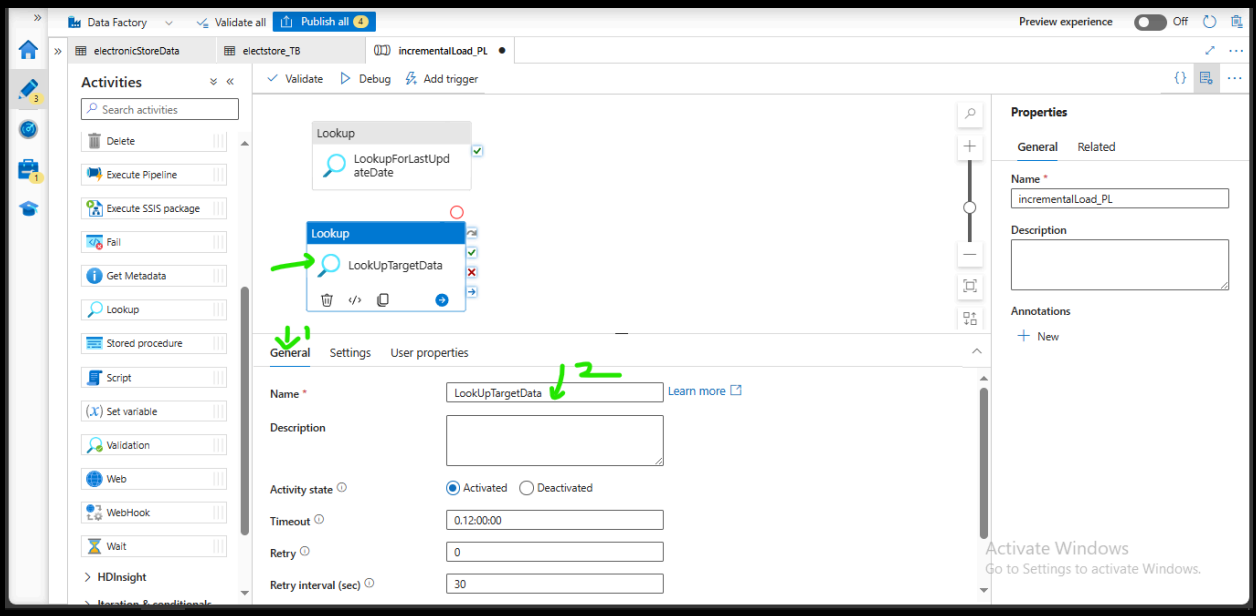


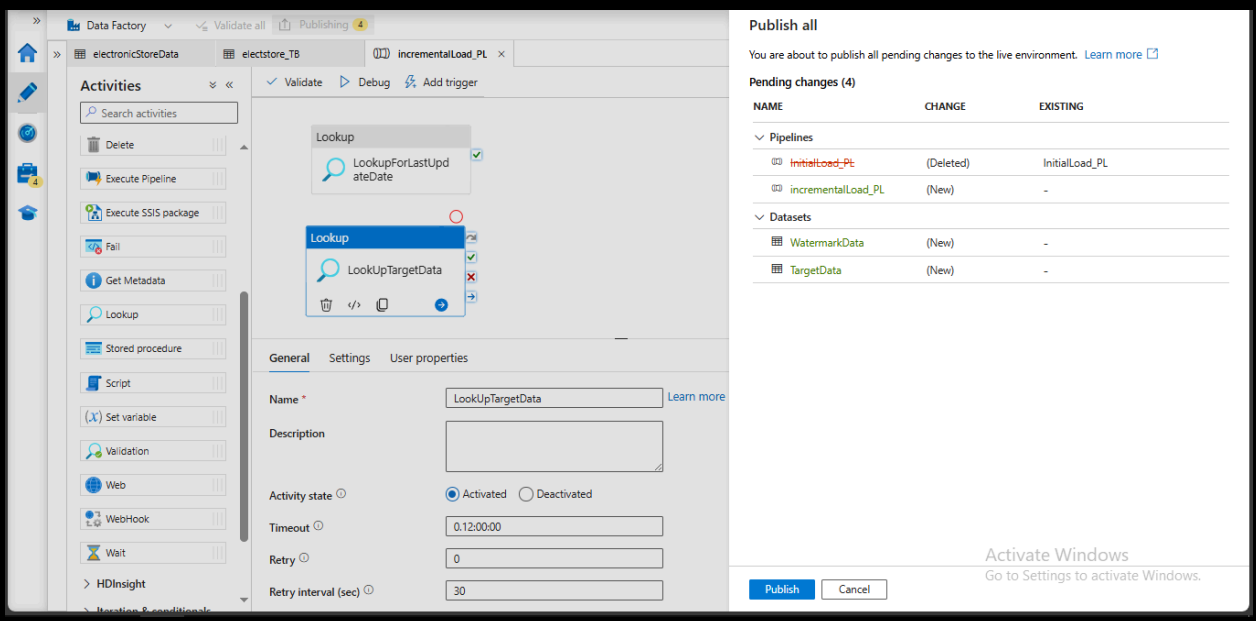

Data Factory
Validate all
Publish all 3
electronicStoreData
electstore_TB
incrementalLoad_PL
Activities
Search activities
Delete
Execute Pipeline
Execute SSIS package
Fail
Get Metadata
Lookup
Stored procedure
Script
Set variable
Validation
Web
WebHook
Wait
HDInsight
Validate
Debug
Add trigger
Lookup
Lookup1
Lookup2
General
Settings
User properties
Source dataset *
Select...
New
First row only
New dataset
In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. Learn more
Select a data store
Search
All
Azure
Database
File
Generic protocol
NoSQL
Services and apps
Azure Database for PostgreSQL
Azure Databricks Delta Lake
Azure File Storage
Azure SQL Database
Azure SQL Database Managed Instance
Azure Synapse Analytics
Azure Table Storage
Microsoft Fabric Warehouse
1
2
Continue
Cancel
Activate Windows
Go to Settings to activate Windows.


Data Factory
Validate all
Publish all 3
electronicStoreData
electstore_TB
incrementalLoad_PL
Activities
Search activities
Delete
Execute Pipeline
Execute SSIS package
Fail
Get Metadata
Lookup
Stored procedure
Script
Set variable
Validation
Web
WebHook
Wait
HDInsight
Validate
Debug
Add trigger
Lookup
Lookup1
Lookup2
General
Settings
User properties
Source dataset *
Select...
New
First row only
Set properties
Name
TarqetData
Linked service *
AzureSqlDatabase1
Table name
dbo.electstore_TB
Enter manually
Import schema
From connection/store
None
Advanced
1
2
3
4
OK
Back
Cancel
Activate Windows
Go to Settings to activate Windows.

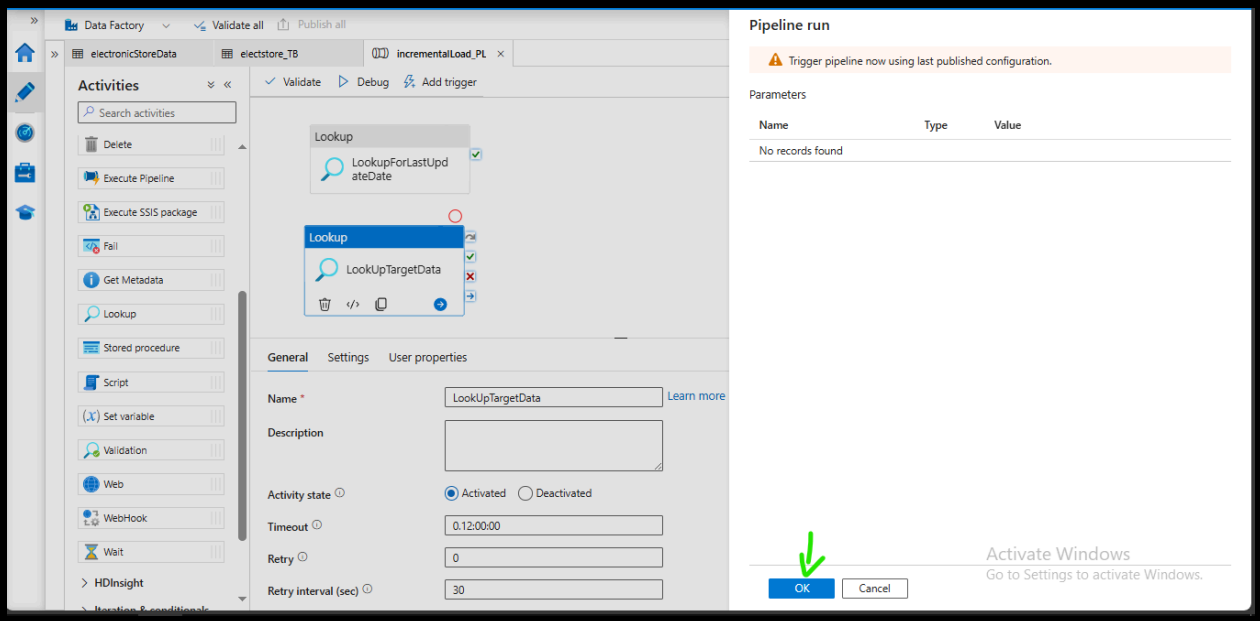

Data Factory
Validate all
Publish all 4
Preview experience
Off
electronicStoreData
electstore_TB
incrementalLoad_PL
Activities
Search activities
Delete
Execute Pipeline
Execute SSIS package
Fail
Get Metadata
Lookup
Stored procedure
Script
Set variable
Validation
Web
WebHook
Wait
HDInsight
Validate
Debug
Add trigger
Lookup
LookupForLastUpdateDate
Lookup
LookUpTargetData
Properties
General
Related
Name *
incrementalLoad_PL
Description
Annotations
New
General
Settings
User properties
Name *
LookupForLastUpdateDate
Learn more
Description
Activity state
Activated
Deactivated
Timeout
0.12:00:00
Retry
0
Retry interval (sec)
30
Activate Windows
Go to Settings to activate Windows.

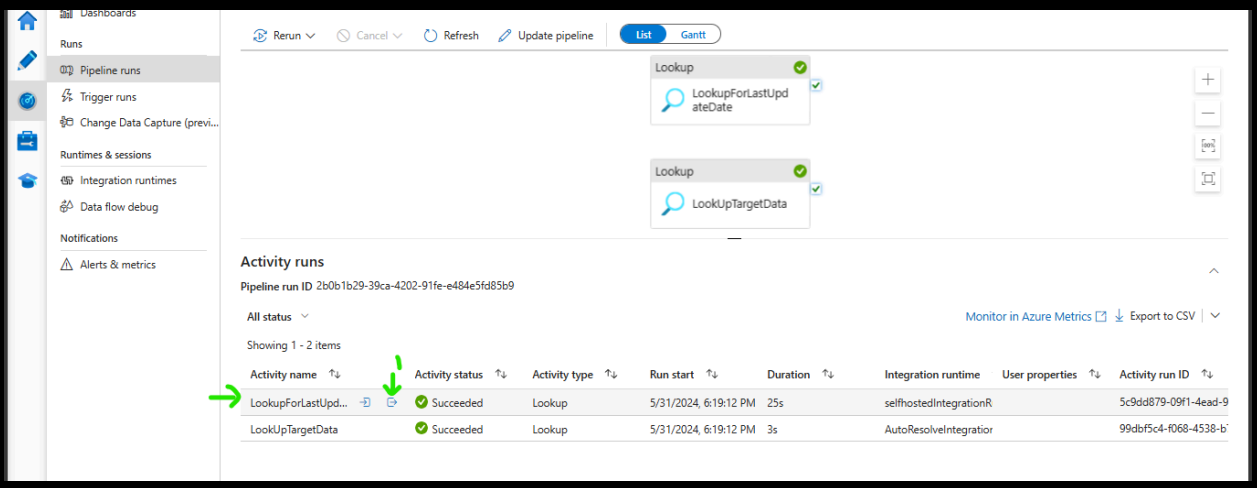

Data Factory
Validate all
Publish all 1
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines
incrementalLoad_PL
Change Data Capture (preview)
Datasets
elec_store
targetTable
waterMarkTable
Data flows
Power Query
incrementalLoad_PL
elec_store
targetTable
Activities
Search activities
Move and transform
Synapse
Azure Data Explorer
Azure Function
Batch Service
Databricks
Data Lake Analytics
General
HDInsight
Iteration & conditionals
Machine Learning
Power Query
Validate
Debug
Add trigger
Lookup
LookupTargetTable
Lookup
LookupForLastUpdateDate
1
2
3
4
General
Settings
User properties
Source dataset *
targetTable
Open
New
Preview data
Learn more
First row only
Use query
Table
Query
Stored procedure
Query *
select max(lastmodified) as maxTargetDate from elecstore_TB;
Edit
Query timeout (minutes)
120
Isolation level
Select...
Activate Windows
Go to Settings to activate Windows.

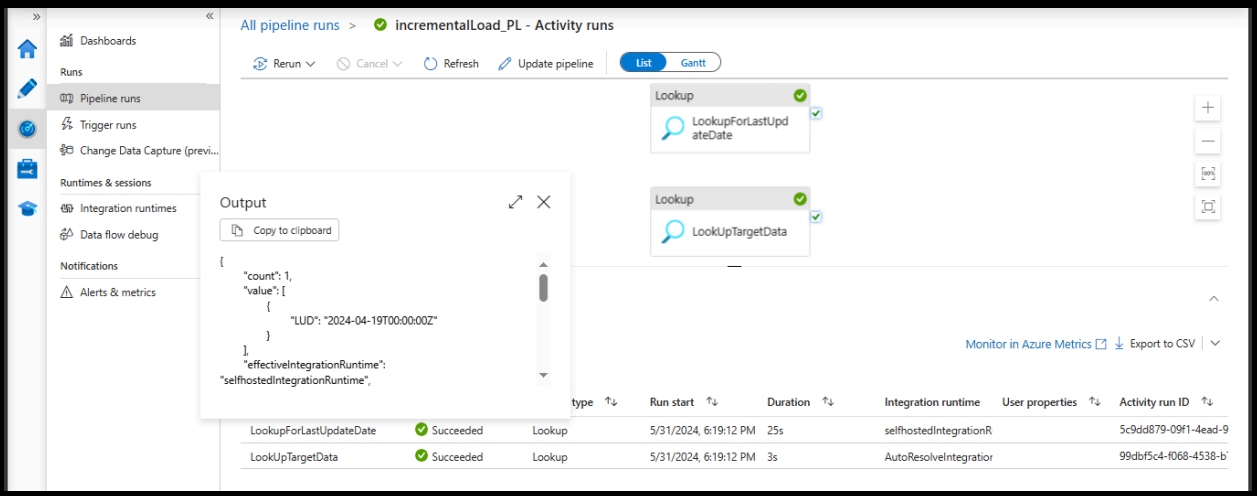


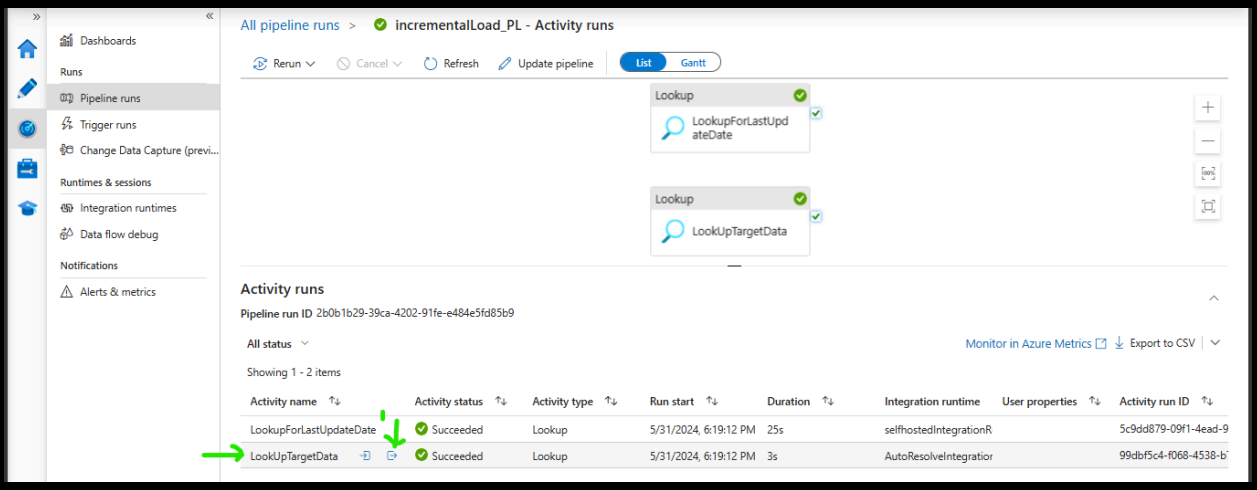


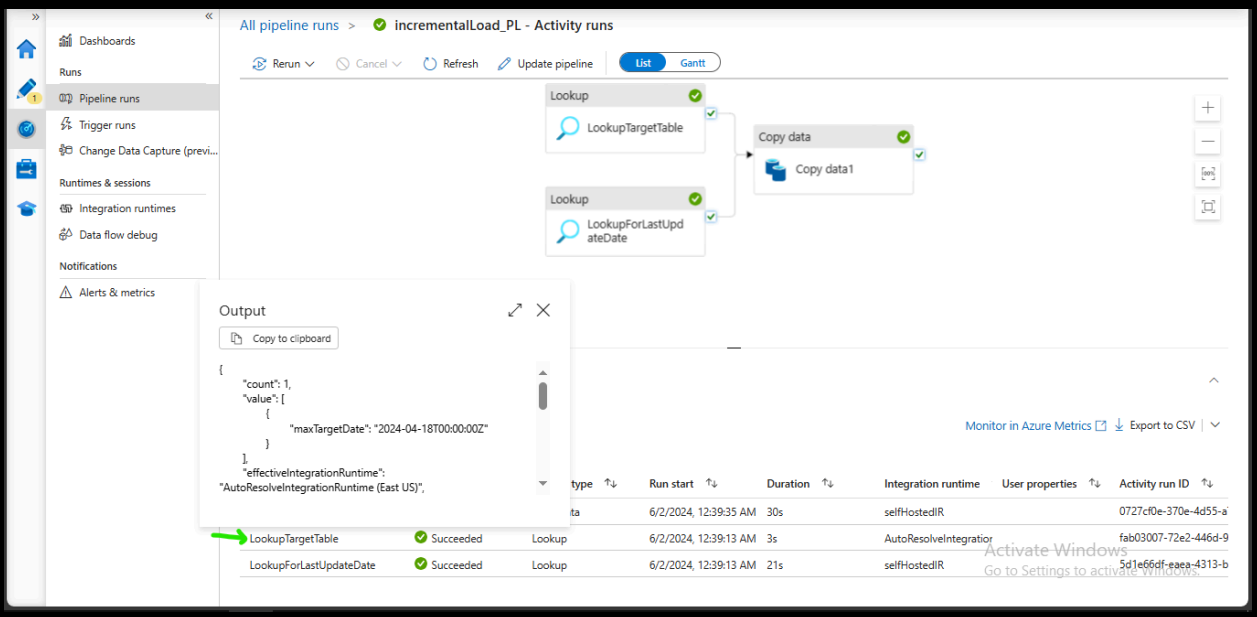
Add Trigger → Trigger Now → ok


Data Factory
Validate all
Publish all
electronicStoreData
electstore_TB
incrementalLoad_PL
Activities
Search activities
Delete
Execute Pipeline
Execute SSIS package
Fail
Get Metadata
Lookup
Stored procedure
Script
Set variable
Validation
Web
WebHook
Wait
HDInsight
Validate
Debug
Add trigger
Lookup
LookupForLastUpdateDate
Lookup
LookUpTargetData
General
Settings
User properties
Name *
LookUpTargetData
Learn more
Description
Activity state
Activated
Deactivated
Timeout
0.12:00:00
Retry
0
Retry interval (sec)
30
Pipeline run
Trigger pipeline now using last published configuration.
Parameters
Name
Type
Value
No records found
OK
Cancel
Activate Windows
Go to Settings to activate Windows.

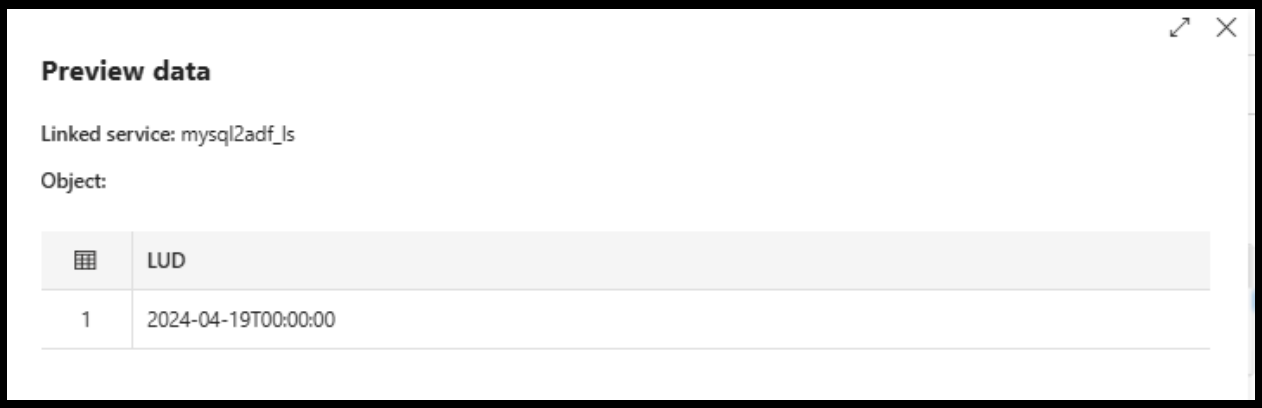

Dashboards
Runs
Pipeline runs
Trigger runs
Change Data Capture (previ...
Runtimes & sessions
Integration runtimes
Data flow debug
Notifications
Alerts & metrics
Rerun
Cancel
Refresh
Update pipeline
List
Gantt
Lookup
LookupForLastUpdateDate
Lookup
LookUpTargetData
Activity runs
Pipeline run ID 2b0b1b29-39ca-4202-91fe-e484e5fd85b9
All status
Monitor in Azure Metrics
Export to CSV
Showing 1 - 2 items
Activity name
Activity status
Activity type
Run start
Duration
Integration runtime
User properties
Activity run ID
LookupForLastUpd...
Succeeded
Lookup
5/31/2024, 6:19:12 PM
25s
selfhostedIntegrationR
5c9dd879-09f1-4ead-9
LookUpTargetData
Succeeded
Lookup
5/31/2024, 6:19:12 PM
3s
AutoResolveIntegratior
99dbf5c4-f068-4538-b


All pipeline runs
incrementalLoad_PL - Activity runs
Dashboards
Runs
Pipeline runs
Trigger runs
Change Data Capture (previ...
Runtimes & sessions
Integration runtimes
Data flow debug
Notifications
Alerts & metrics
Rerun
Cancel
Refresh
Update pipeline
List
Gantt
Lookup
LookupForLastUpdateDate
Lookup
LookUpTargetData
Output
Copy to clipboard
{
"count": 1,
"value": [
{
"LUD": "2024-04-19T00:00:00Z"
}
],
"effectiveIntegrationRuntime":
"selfhostedIntegrationRuntime",
Monitor in Azure Metrics
Export to CSV
Run start
Duration
Integration runtime
User properties
Activity run ID
LookupForLastUpdateDate
Succeeded
Lookup
5/31/2024, 6:19:12 PM
25s
selfhostedIntegrationR
5c9dd879-09f1-4ead-9
LookUpTargetData
Succeeded
Lookup
5/31/2024, 6:19:12 PM
3s
AutoResolveIntegratior
99dbf5c4-f068-4538-b





Below image has the LookupforLastUpdatedDate



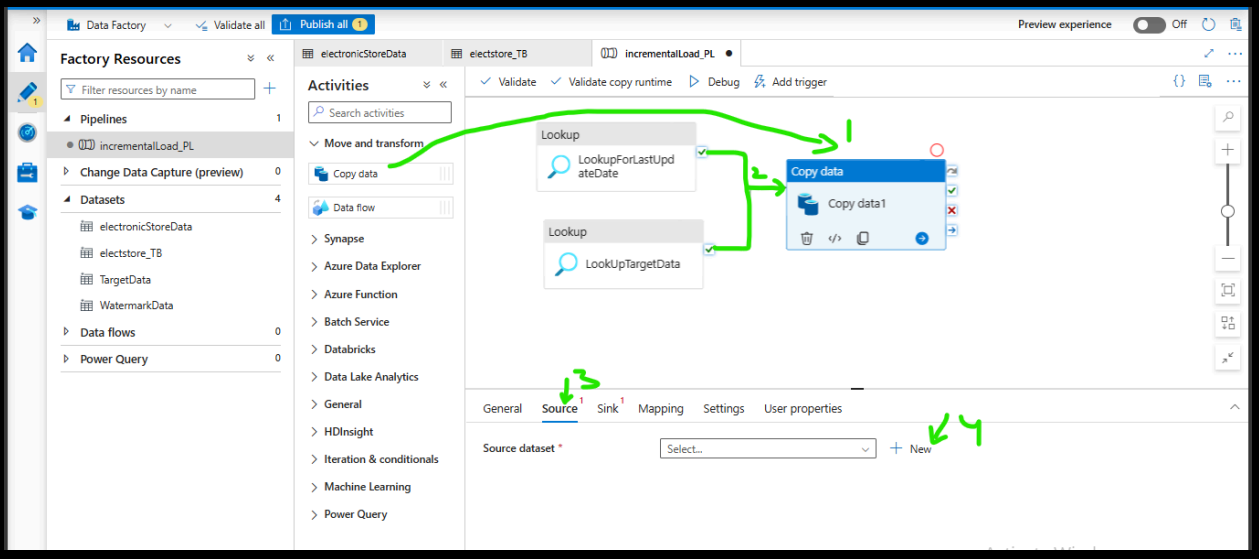
Below image has the LookupforLastUpdatedDate

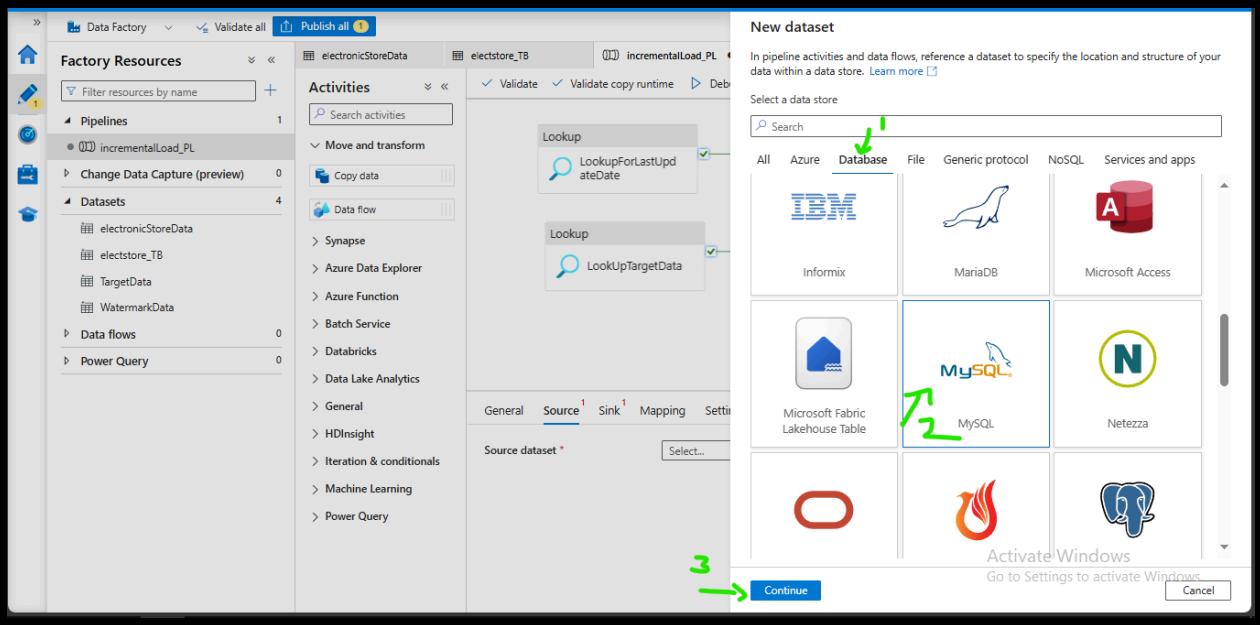
Preview data

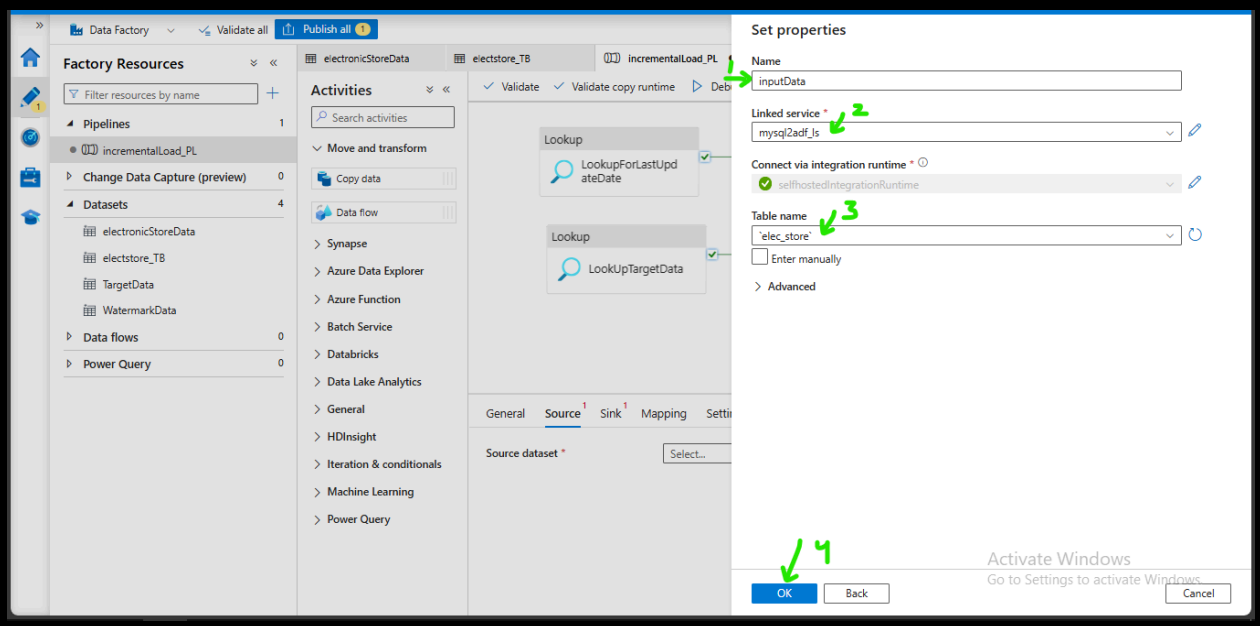
Linked service: AzureSqlDatabase1

Object: dbo.electstore_TB

	OrderID	OrderDate	UnitCost	Price	OrderQty	CostofSales	Sales	Profit
1	7077	9/13/2017	76.09496775	304.0	9	684.8547097	2714.7200000000003	2029.86529000

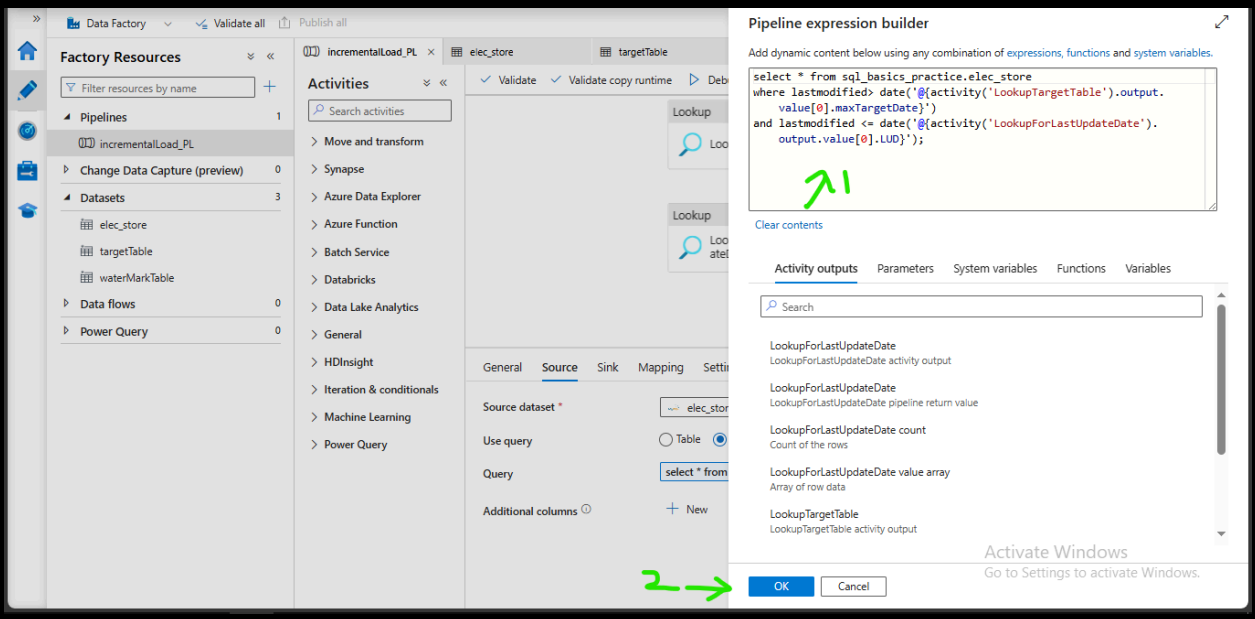


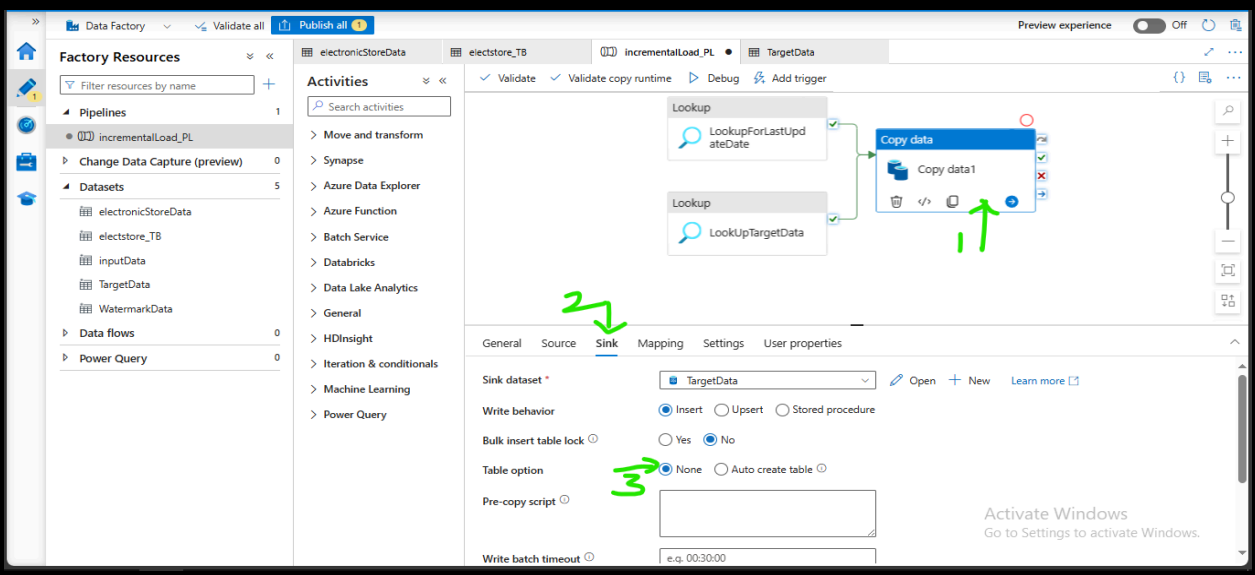




We need to get the data after the lastmodified date '2023-04-19' till 2024-05-01" which is 101 records.

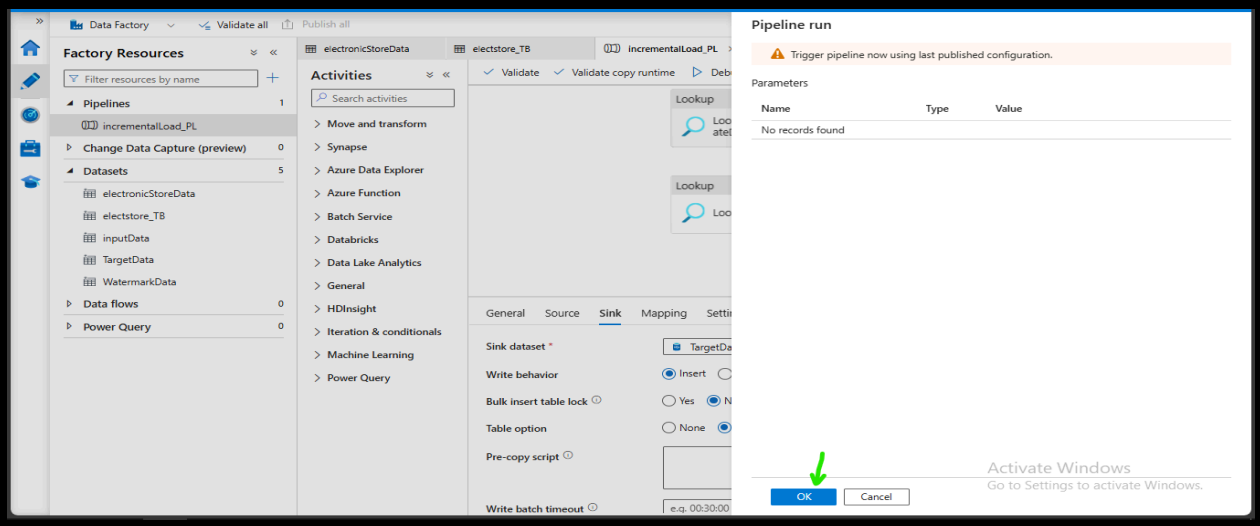
Copy activity → source → Use Query → Query → Add dynamically content

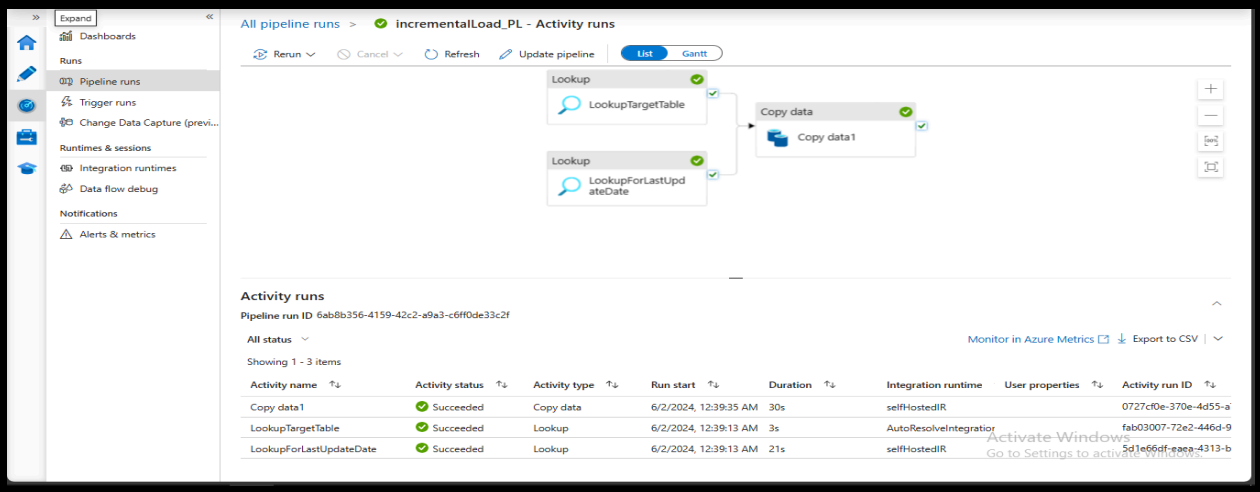




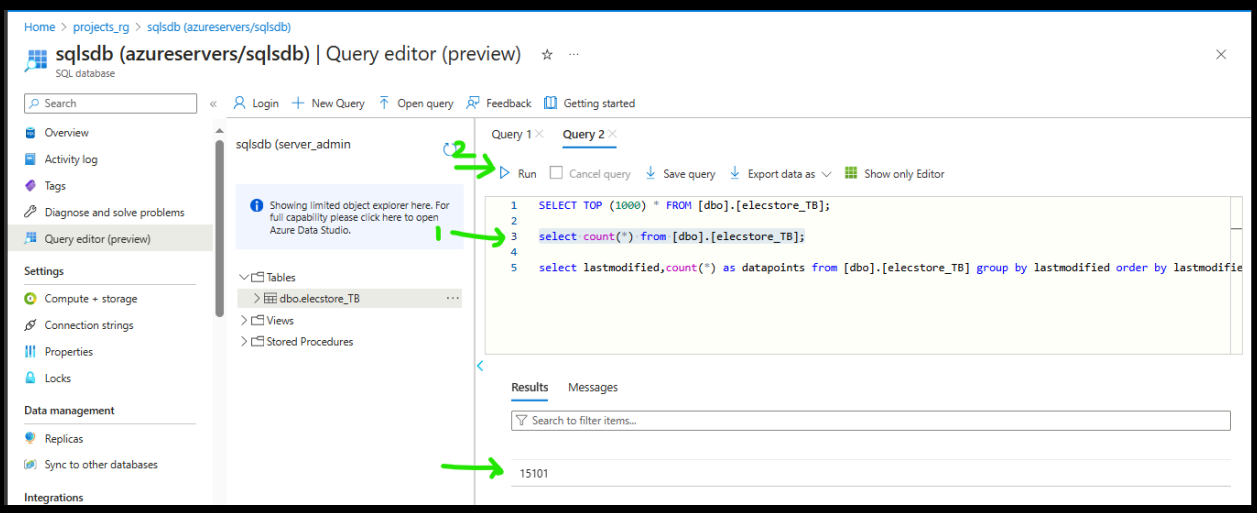
Publish all → publish

Add Trigger → Trigger now

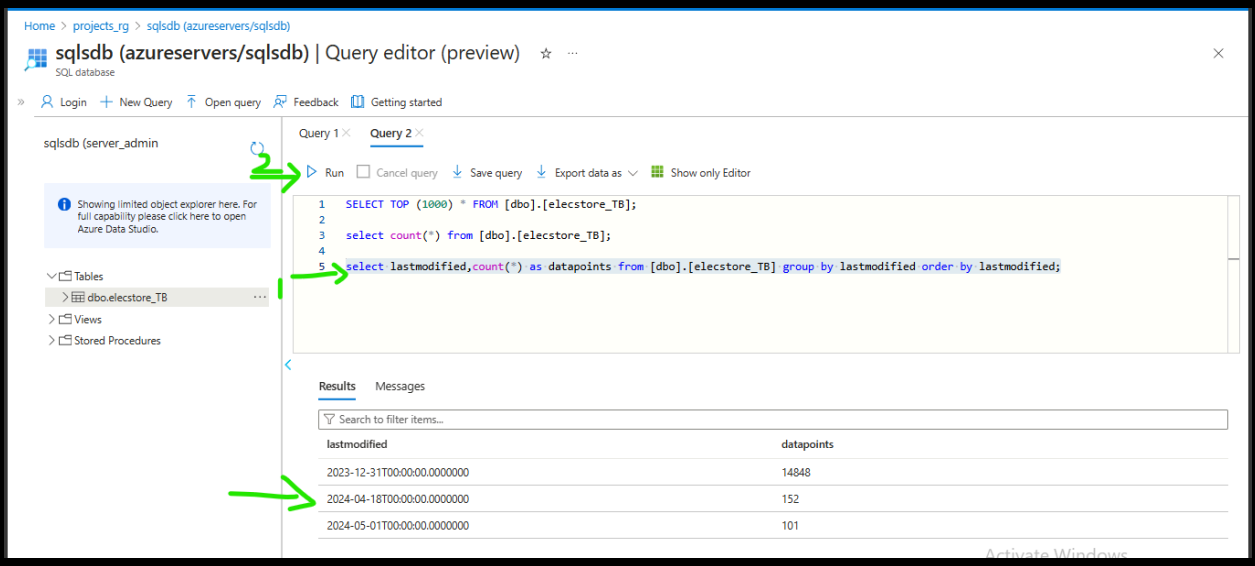




Total 15101 records can be seen in azure sql database.



Groupby lastmodified


Home > projects_rg > sqlsdb (azureservers/sqlsdb)
sqlsdb (azureservers/sqlsdb) | Query editor (preview)
SQL database
Login
New Query
Open query
Feedback
Getting started
sqlsdb (server_admin
Showing limited object explorer here. For full capability please click here to open Azure Data Studio.
Tables
dbo.elecstore_TB
Views
Stored Procedures
Query 1
Query 2
Run
Cancel query
Save query
Export data as
Show only Editor
1 SELECT TOP (1000) * FROM [dbo].[elecstore_TB];
2
3 select count(*) from [dbo].[elecstore_TB];
4
5 select lastmodified,count(*) as datapoints from [dbo].[elecstore_TB] group by lastmodified order by lastmodified;
Results
Messages
Search to filter items...
lastmodified
datapoints
2023-12-31T00:00:00.0000000
14848
2024-04-18T00:00:00.0000000
152
2024-05-01T00:00:00.0000000
101