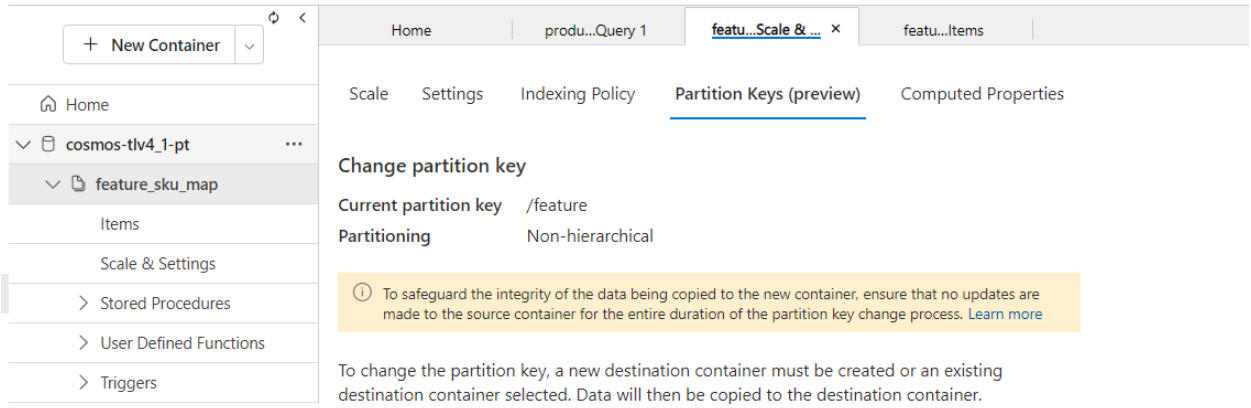
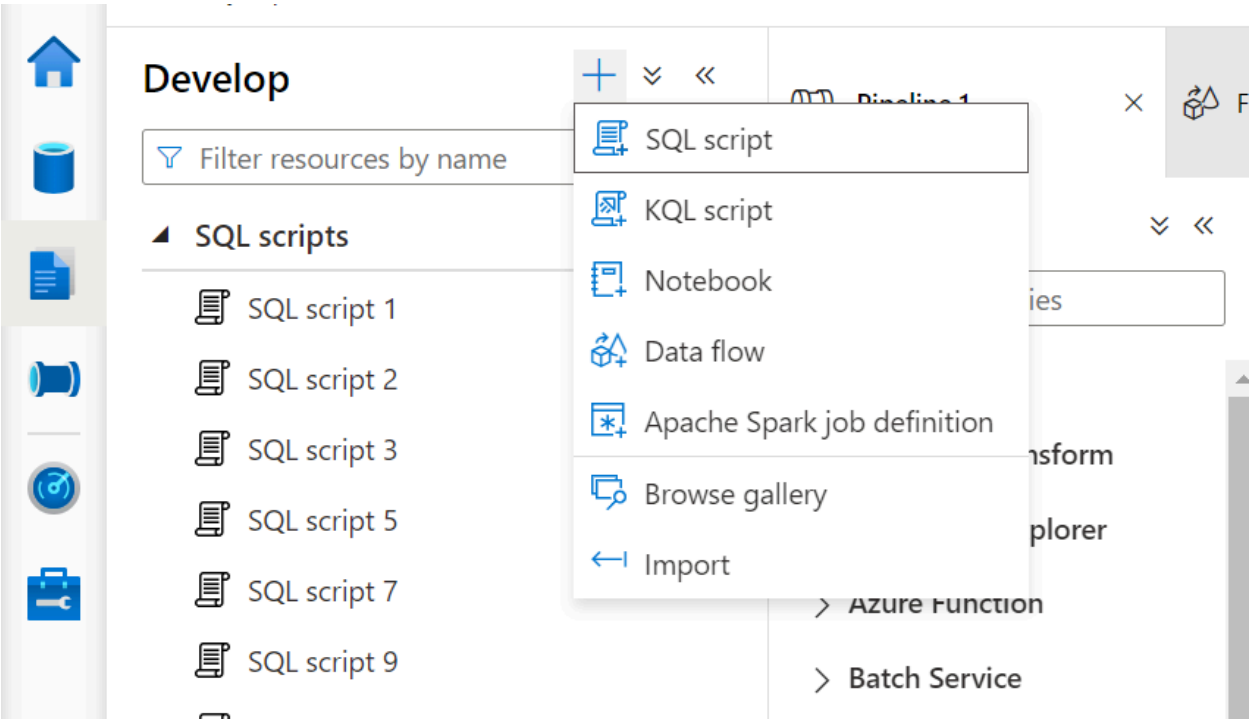
Data Flow pipeline for creating feature sku map container:

1. Create new container for feature_sku_map with feature as partition key.

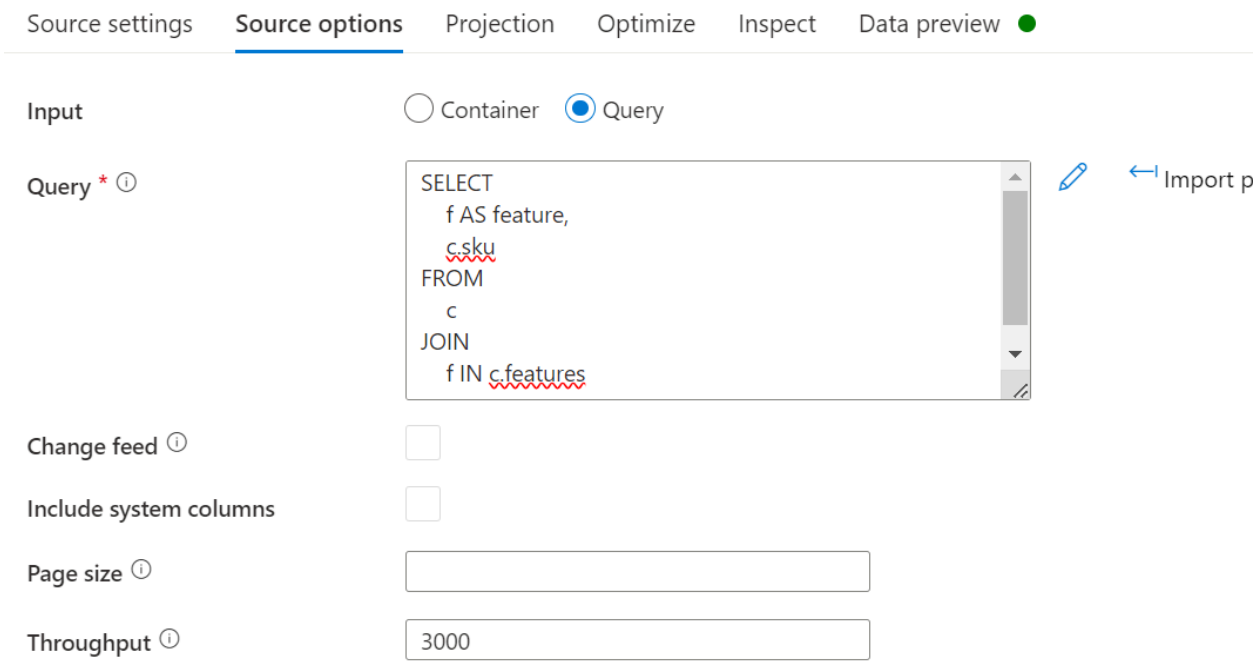


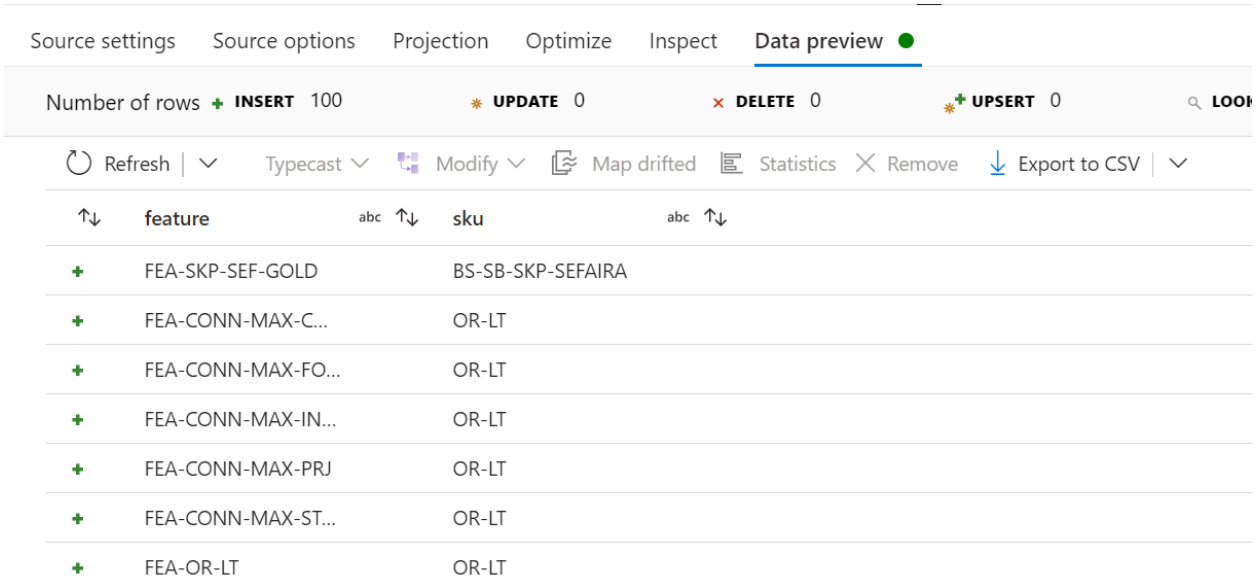
2. Open synapse studio and go to integrates tab.
3. Create new data flow.



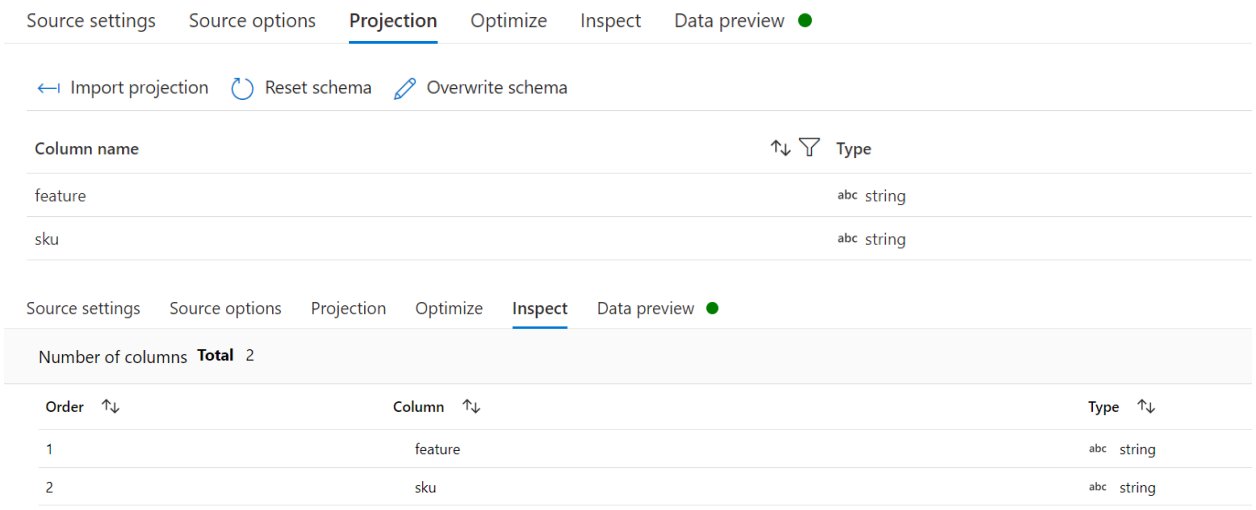
4. Add one source to read details from product_reader container.

5. Use query option to took only feature and sku fields.

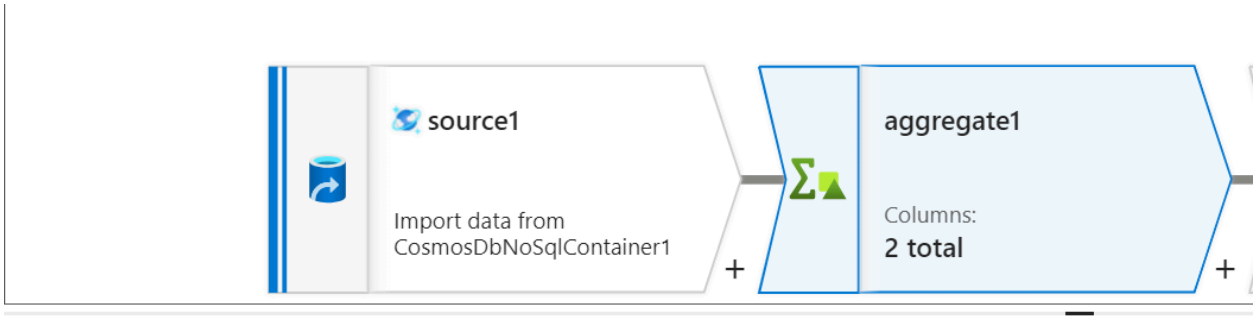


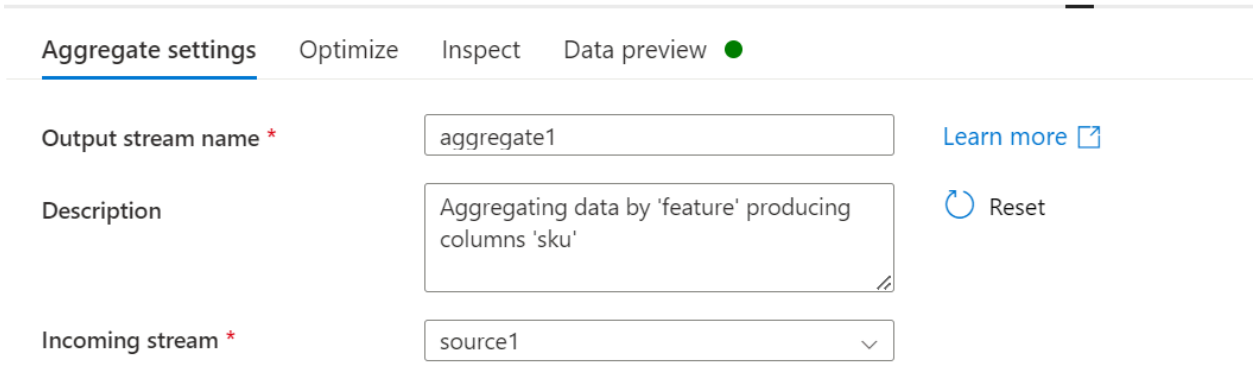


6. It is important to import projection to get columns to pass to next flow.

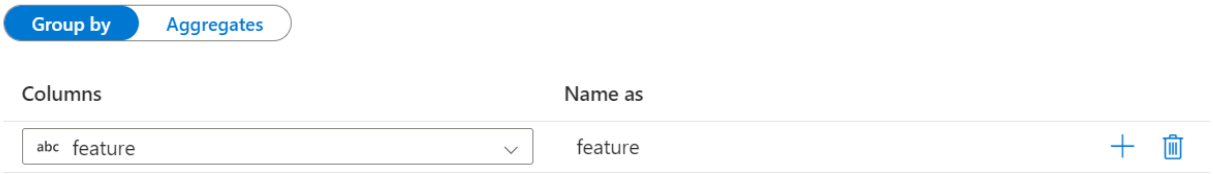


7. Once the data has been ready, next do the transformation. Here, we have to do aggregation.

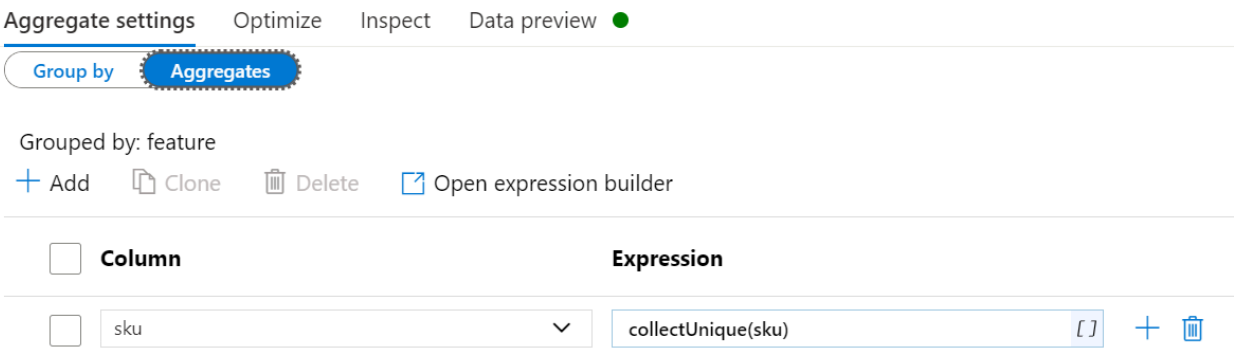


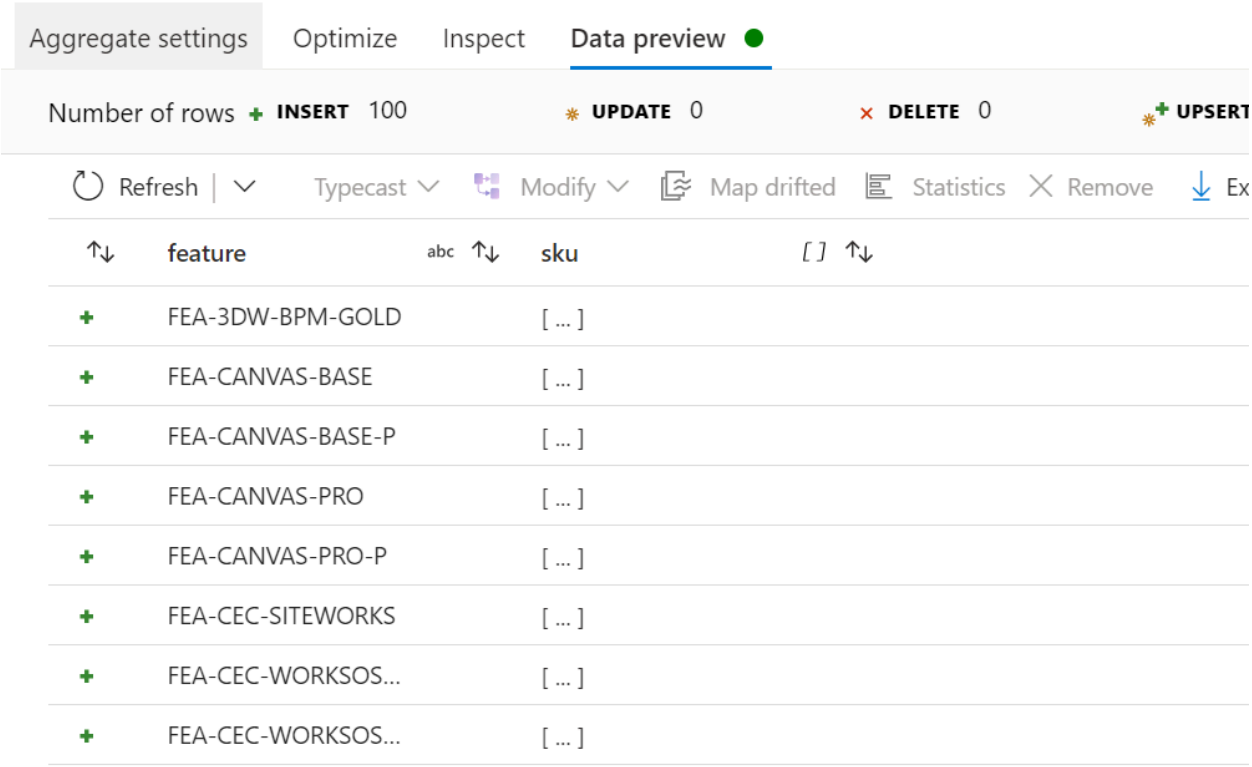


8. Now there is two fields to provide input: Group by and Arregates. For group by, provide “feature”.

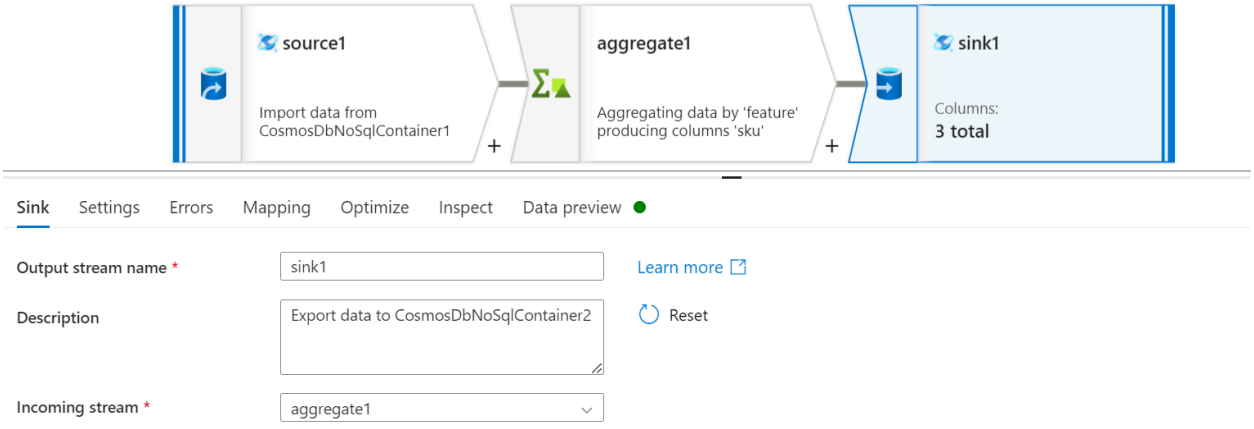


9. For aggregates, give sku column and expression as collectUnique.





10. Now, the final step is to sink the data into result container.



11. For dataset, click new and select Azure Cosmos DB for NoSQL.

New integration 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

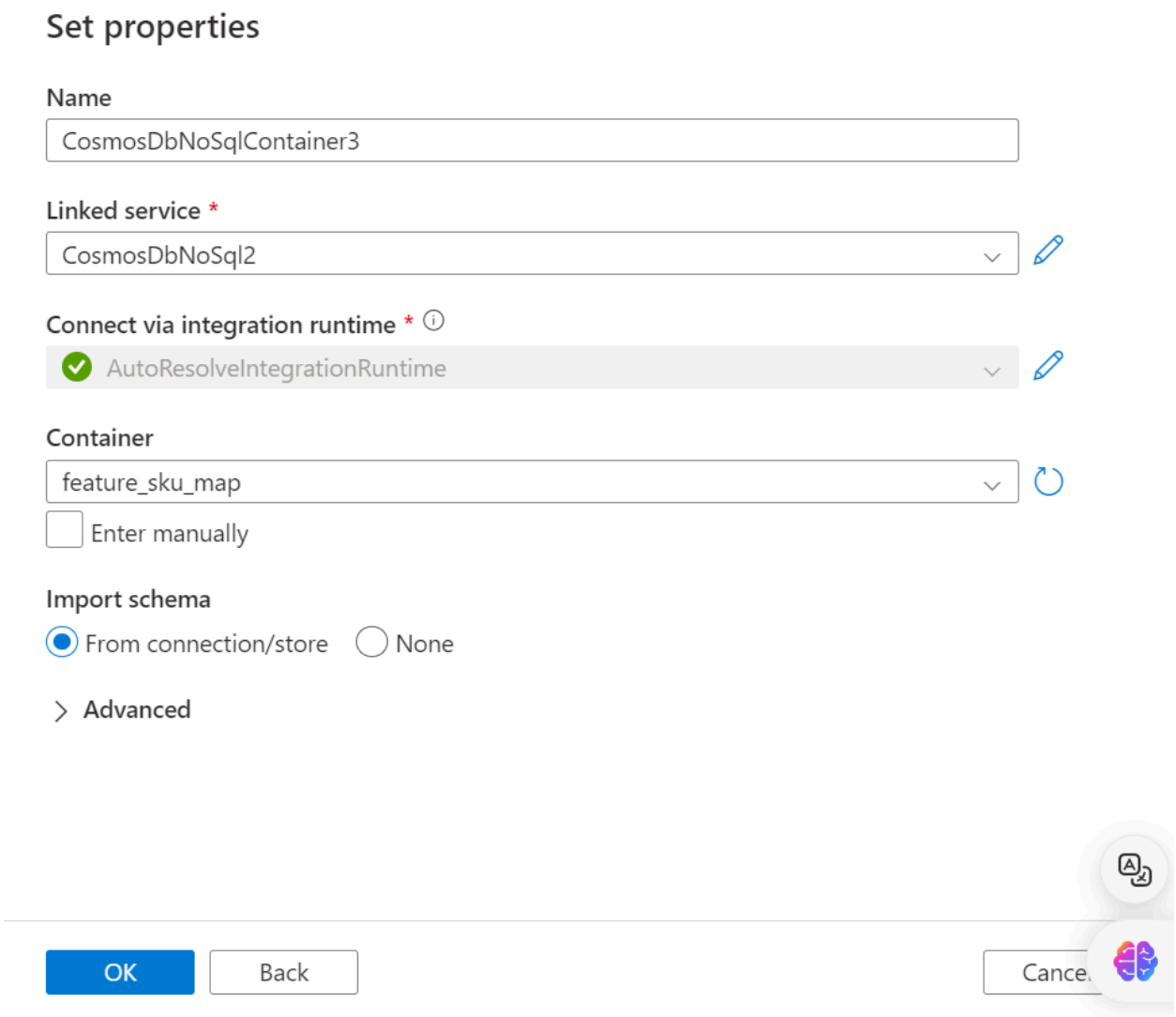
Azure Blob Storage

Azure Cosmos DB for NoSQL

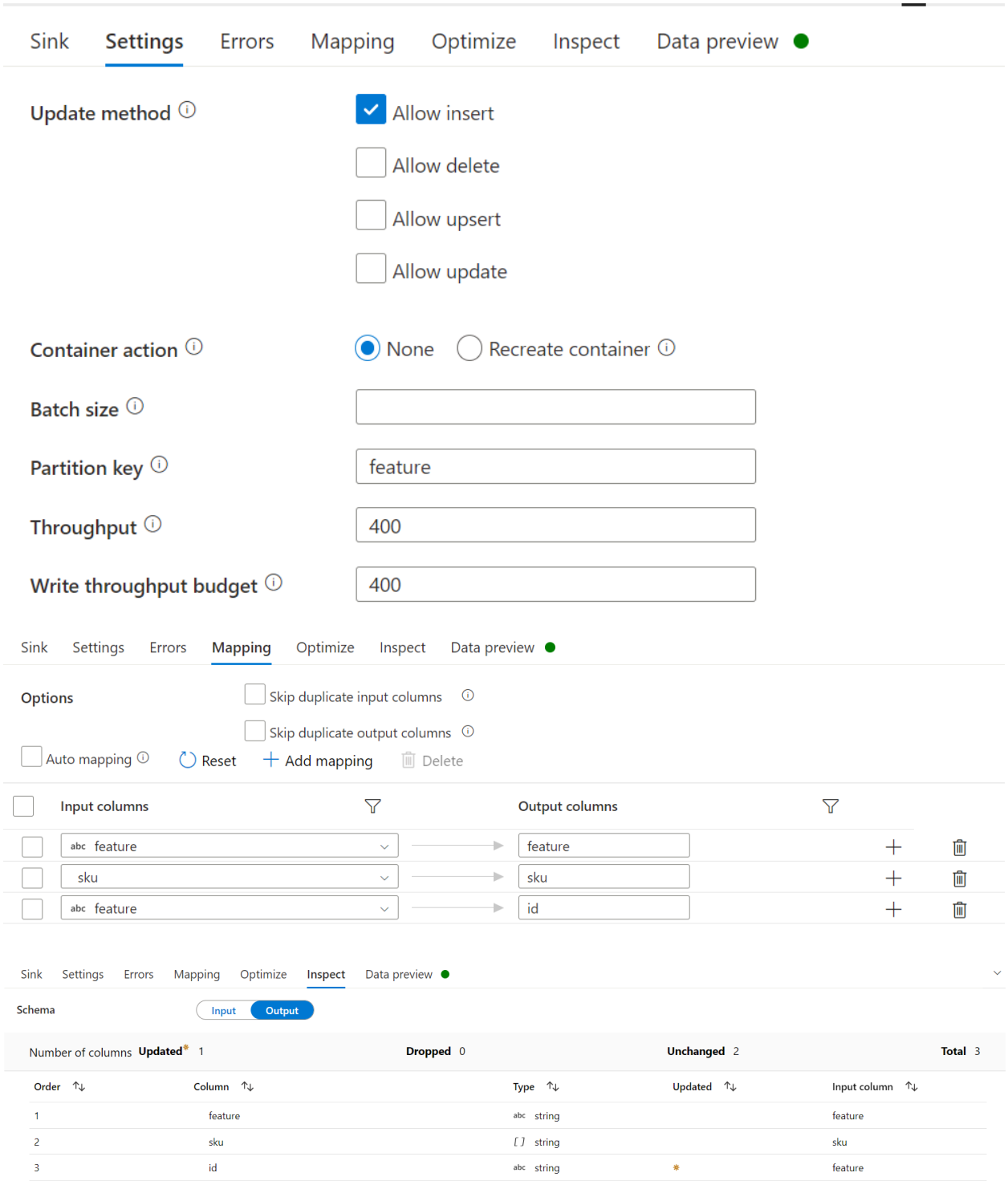
Azure Data Explorer (Kusto)

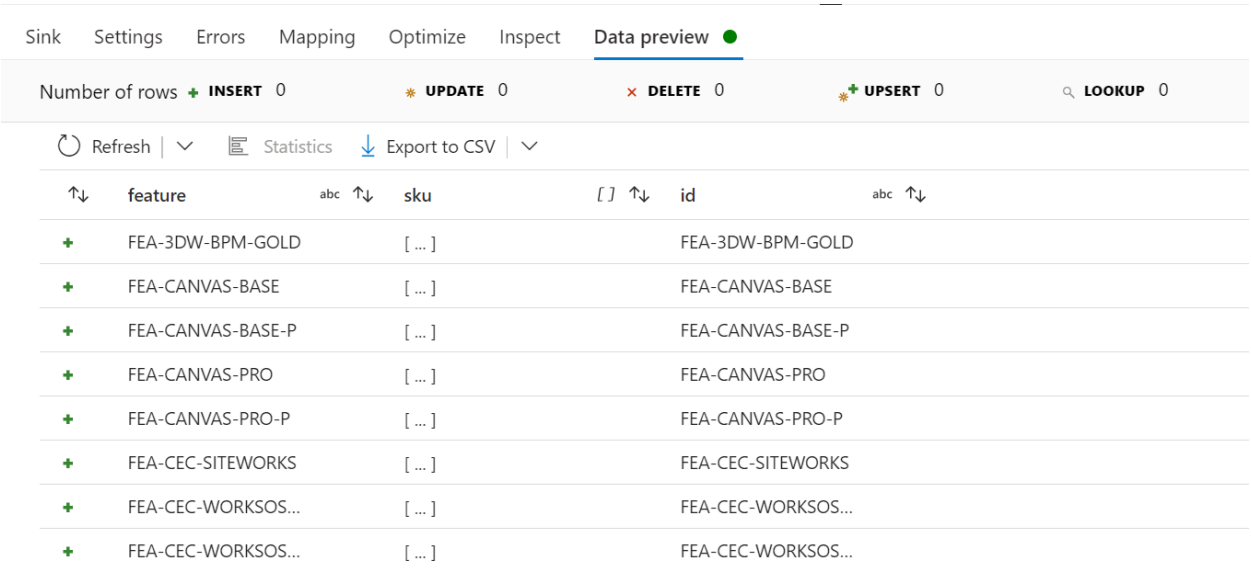
Continue

Cancel

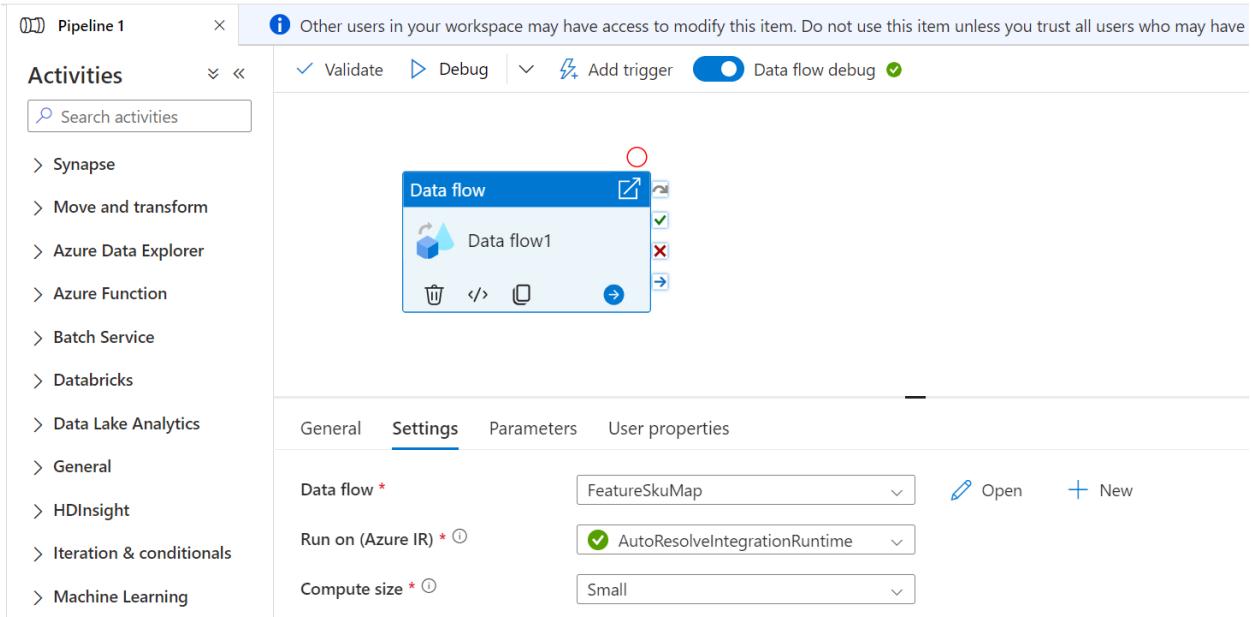


12. Fill the remaining settings and other tabs:


Sink
Settings
Errors
Mapping
Optimize
Inspect
Data preview
Update method
Allow insert
Allow delete
Allow upsert
Allow update
Container action
None
Recreate container
Batch size
Partition key
feature
Throughput
400
Write throughput budget
400
Sink
Settings
Errors
Mapping
Optimize
Inspect
Data preview
Options
Skip duplicate input columns
Skip duplicate output columns
Auto mapping
Reset
Add mapping
Delete
Input columns
Output columns
feature
feature
sku
sku
feature
id
Sink
Settings
Errors
Mapping
Optimize
Inspect
Data preview
Schema
Input
Output
Number of columns
Updated* 1
Dropped 0
Unchanged 2
Total 3
Order
Column
Type
Updated
Input column
1
feature
string
feature
2
sku
string
sku
3
id
string
*
feature



13. Once the data flow has been created successfully, then create pipeline to trigger that data flow.



14. Once setup has been created, then trigger it and check the pipeline execution. Once the execution has finished, we can see the results in “feature_sku_map” container.

Pipeline runs

Triggered | Debug | Rerun | Cancel options | Refresh | Edit columns | List | Gantt

Filter by run ID or name | Chennai, Kolkata, Mu... : **Last 24 hours** | Pipeline name : **Pipeline 1**

Status : **All** | Runs : Latest runs | Triggered by : **All** | Add filter | Copy filters | Export to CSV

Showing 1 - 1 items — Last refreshed 0 minutes ago

Pipeline name	Run start	Run end	Duration	Triggered by	Status
Pipeline 1	10/10/2024, 10:40:22 PM	10/10/2024, 10:44:28 PM	4m 6s	Manual trigger	Succeed