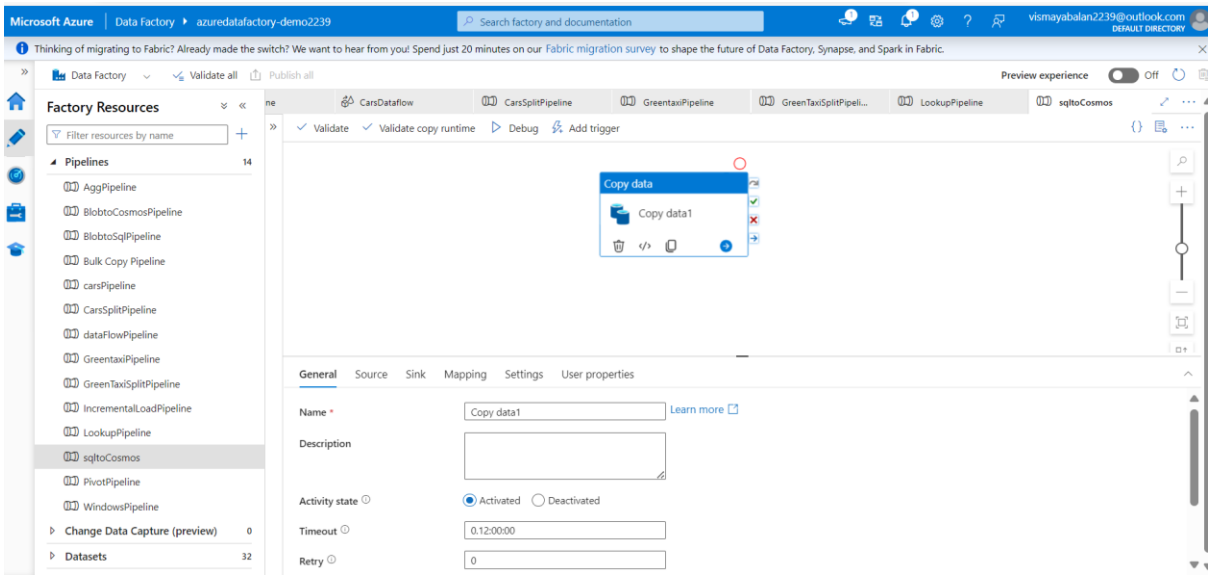
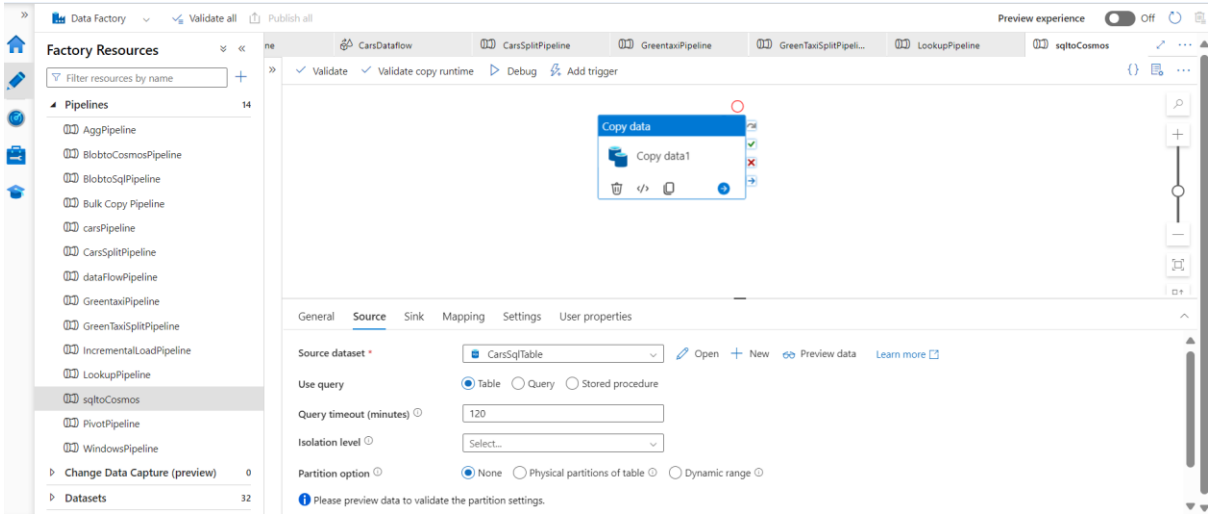
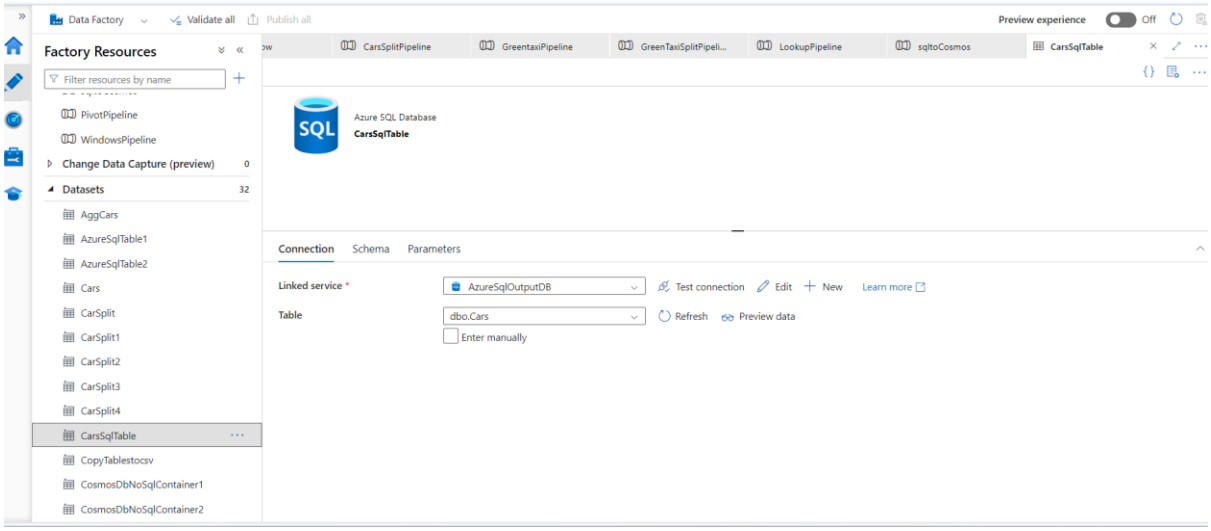
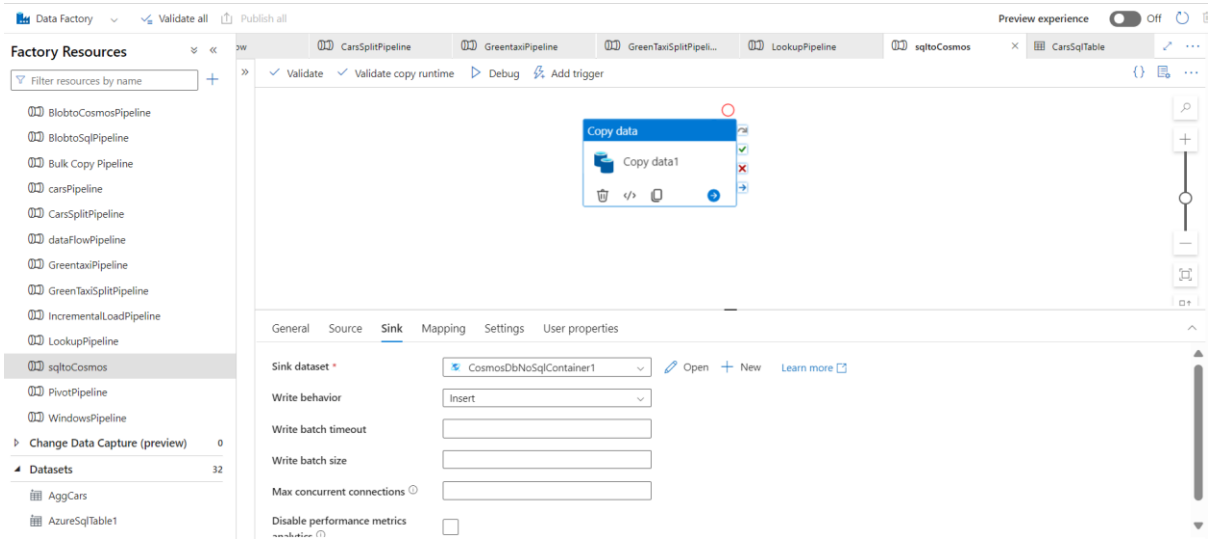
Assignment 10 - Azure Data Factory

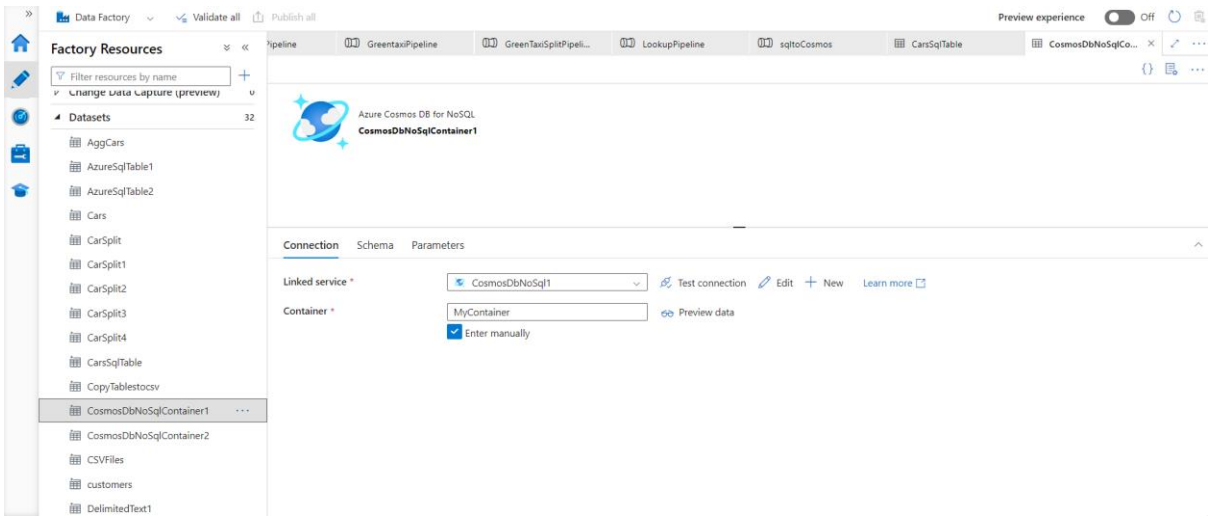
1. Design an ADF pipeline to copy data from an on-premise Azure SQL database to Azure Cosmos DB, ensuring data consistency and performance optimization. Pick correct options of partitioning for better performance.

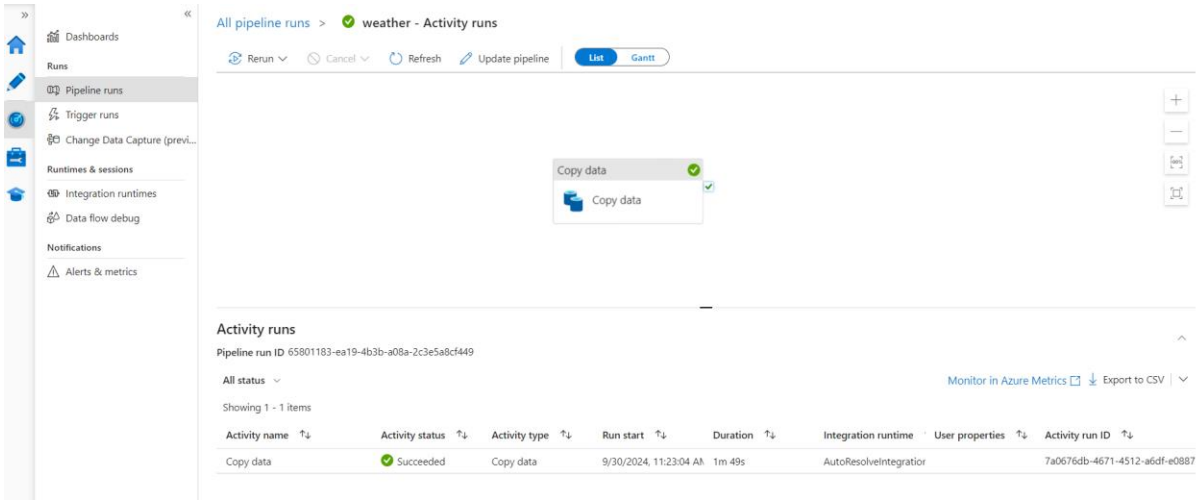




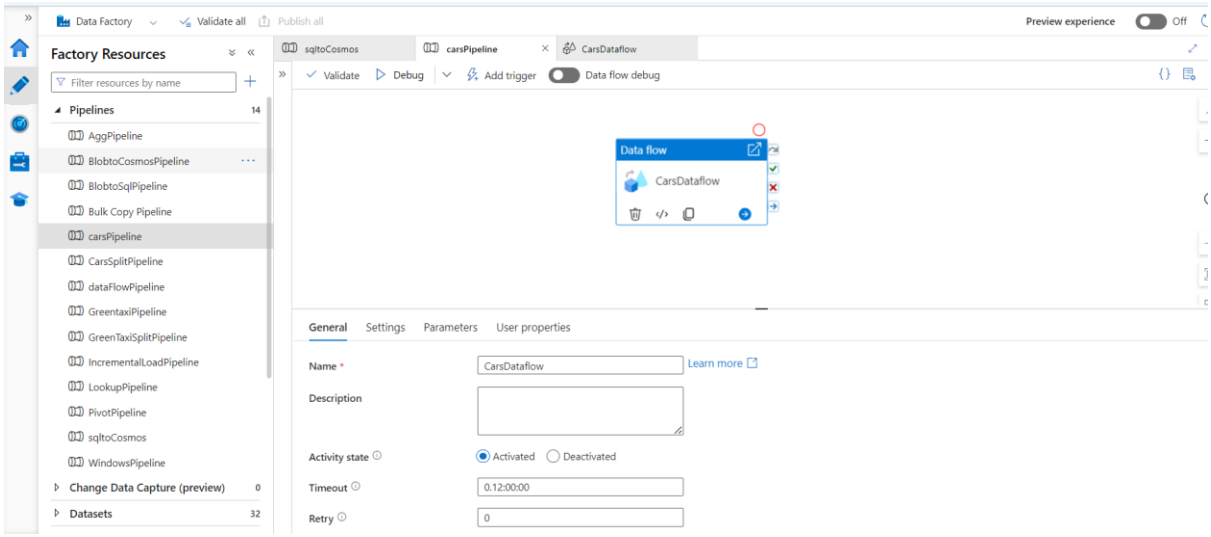

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
PivotPipeline
WindowsPipeline
Change Data Capture (preview) 0
Datasets 32
AggCars
AzureSqlTable1
AzureSqlTable2
Cars
CarSplit
CarSplit1
CarSplit2
CarSplit3
CarSplit4
CarsSqlTable
CopyTablestocsv
CosmosDbNoSqlContainer1
CosmosDbNoSqlContainer2
CarSplitPipeline
GreentaxiPipeline
GreenTaxiSplitPipeli...
LookupPipeline
sqltoCosmos
CarsSqlTable
Azure SQL Database
CarsSqlTable
Connection
Schema
Parameters
Linked service *
AzureSqlOutputDB
Test connection
Edit
New
Learn more
Table
dbo.Cars
Refresh
Preview data
Enter manually

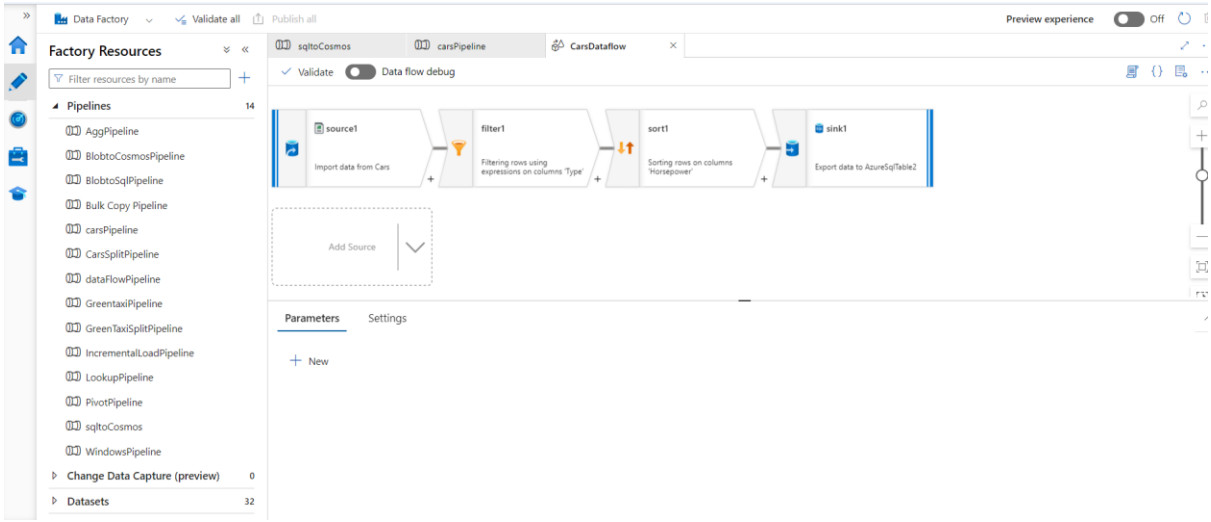

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
BlobtoCosmosPipeline
BlobtoSqlPipeline
Bulk Copy Pipeline
carsPipeline
CarsSplitPipeline
dataFlowPipeline
GreentaxiPipeline
GreenTaxiSplitPipeline
IncrementalLoadPipeline
LookupPipeline
sqltoCosmos
PivotPipeline
WindowsPipeline
Change Data Capture (preview) 0
Datasets 32
AggCars
AzureSqlTable1
CarsSplitPipeline
GreentaxiPipeline
GreenTaxiSplitPipeli...
LookupPipeline
sqltoCosmos
CarsSqlTable
Validate
Validate copy runtime
Debug
Add trigger
Copy data
Copy data1
General
Source
Sink
Mapping
Settings
User properties
Sink dataset *
CosmosDbNoSqlContainer1
Open
New
Learn more
Write behavior
Insert
Write batch timeout
Write batch size
Max concurrent connections
Disable performance metrics

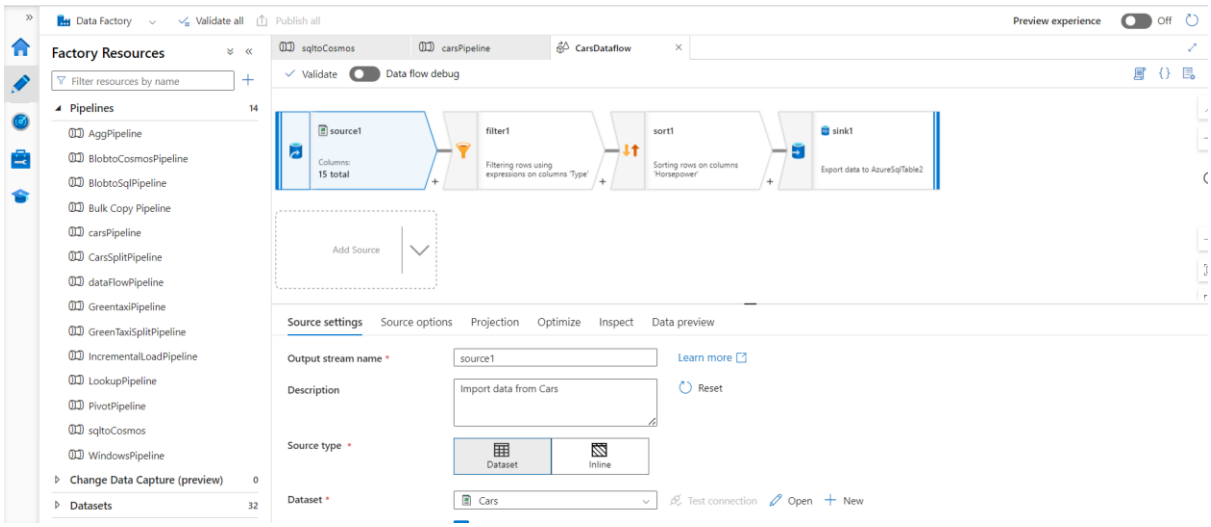

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Change Data Capture (preview)
Datasets 32
AggCars
AzureSqlTable1
AzureSqlTable2
Cars
CarSplit
CarSplit1
CarSplit2
CarSplit3
CarSplit4
CarsSqlTable
CopyTablestocsv
CosmosDbNoSqlContainer1
CosmosDbNoSqlContainer2
CSVFiles
customers
DelimitedText1
GreentaxiPipeline
GreenTaxiSplitPipeli...
LookupPipeline
sqltoCosmos
CarsSqlTable
CosmosDbNoSqlCo...
Azure Cosmos DB for NoSQL
CosmosDbNoSqlContainer1
Connection
Schema
Parameters
Linked service *
CosmosDbNoSql1
Test connection
Edit
New
Learn more
Container *
MyContainer
Preview data
Enter manually

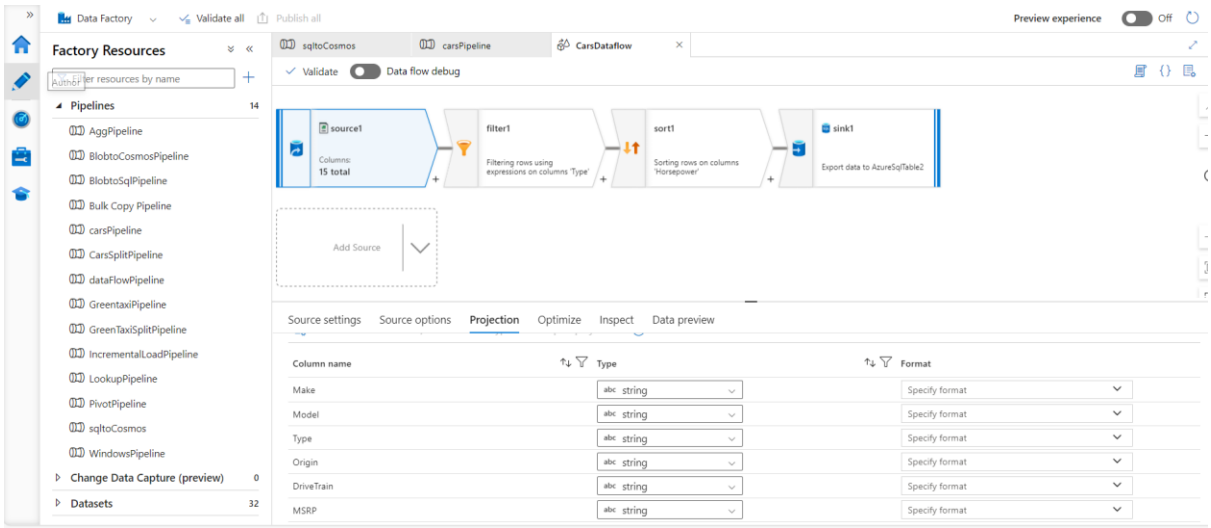


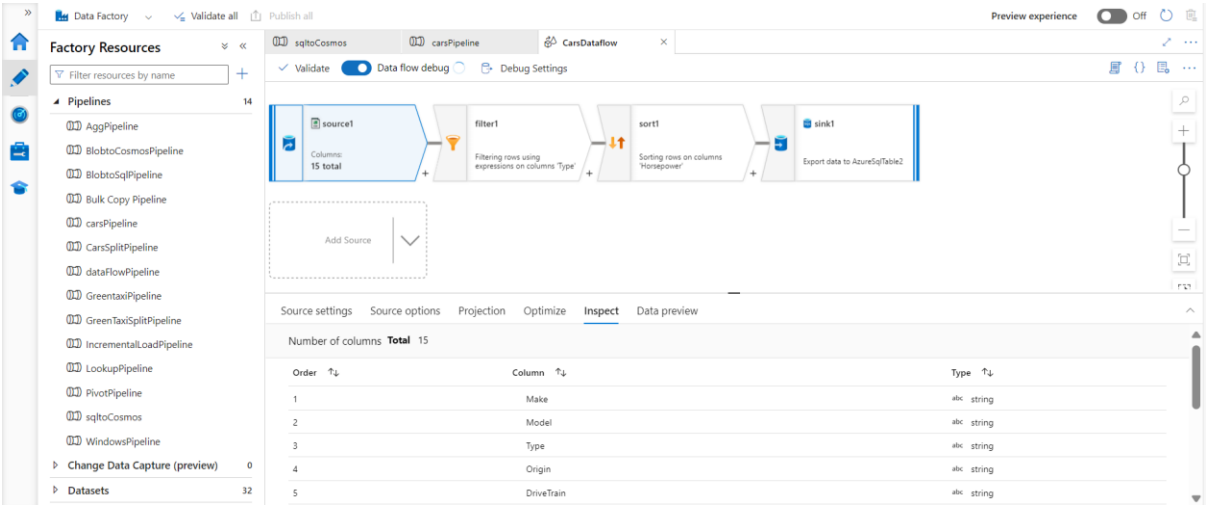
2. Create Pipeline using Azure Data Flow in Azure Data Factory to apply Filter and Sort transformations on datasets.

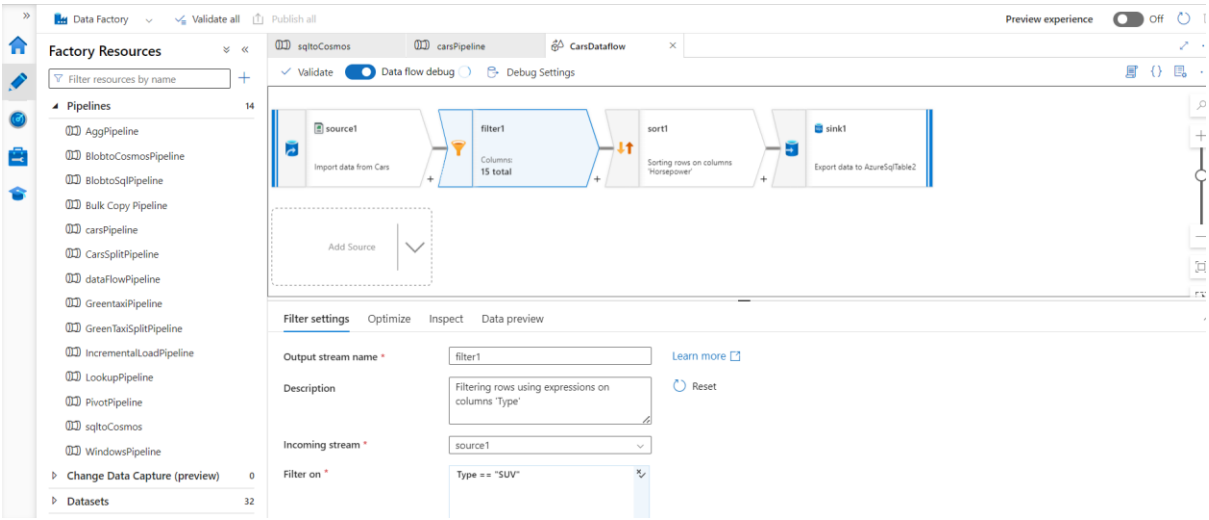


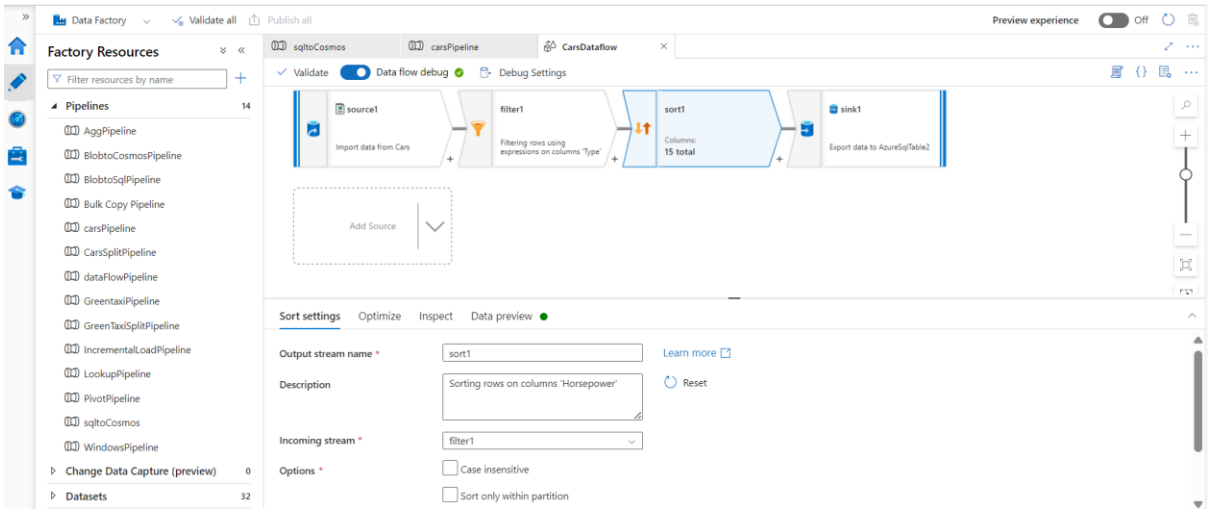

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 14
AggPipeline
BlobtoCosmosPipeline
BlobtoSqlPipeline
Bulk Copy Pipeline
carsPipeline
CarsSplitPipeline
dataFlowPipeline
GreentaxiPipeline
GreenTaxiSplitPipeline
IncrementalLoadPipeline
LookupPipeline
PivotPipeline
sqltoCosmos
WindowsPipeline
Change Data Capture (preview) 0
Datasets 32
sqltoCosmos
carsPipeline
CarsDataflow
Validate
Data flow debug
source1
Import data from Cars
filter1
Filtering rows using expressions on columns 'Type'
sort1
Sorting rows on columns 'Horsepower'
sink1
Export data to AzureSqlTable2
Add Source
Parameters
Settings
New

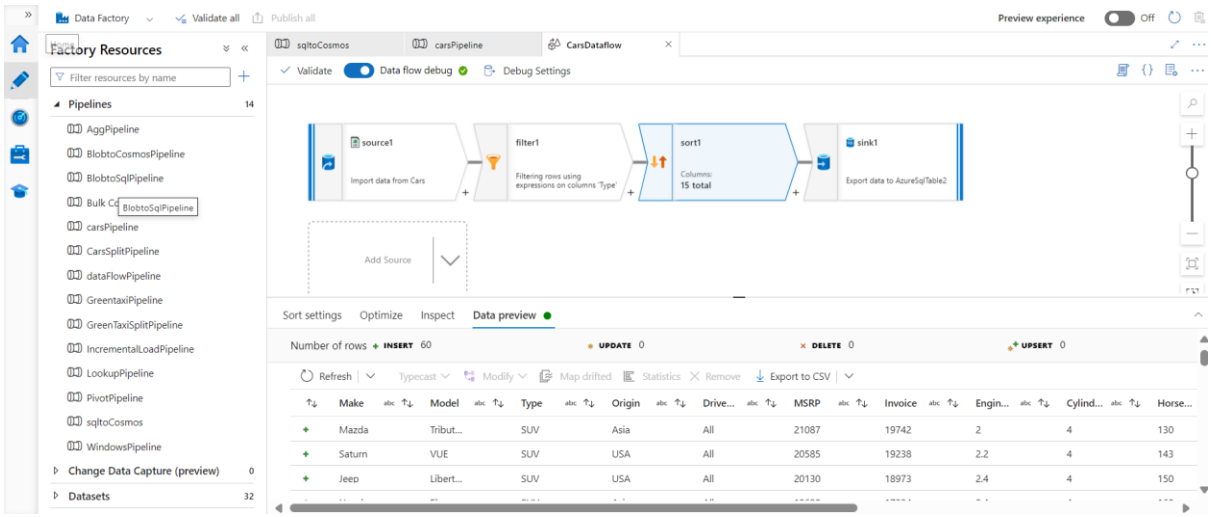

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 14
AggPipeline
BlobtoCosmosPipeline
BlobtoSqlPipeline
Bulk Copy Pipeline
carsPipeline
CarsSplitPipeline
dataFlowPipeline
GreentaxiPipeline
GreenTaxiSplitPipeline
IncrementalLoadPipeline
LookupPipeline
PivotPipeline
sqltoCosmos
WindowsPipeline
Change Data Capture (preview) 0
Datasets 32
sqltoCosmos
carsPipeline
CarsDataflow
Validate
Data flow debug
source1
Columns:
15 total
filter1
Filtering rows using expressions on columns 'Type'
sort1
Sorting rows on columns 'Horsepower'
sink1
Export data to AzureSqlTable2
Add Source
Source settings
Source options
Projection
Optimize
Inspect
Data preview
Output stream name *
source1
Learn more
Description
Import data from Cars
Reset
Source type *
Dataset
Inline
Dataset *
Cars
Test connection
Open
New

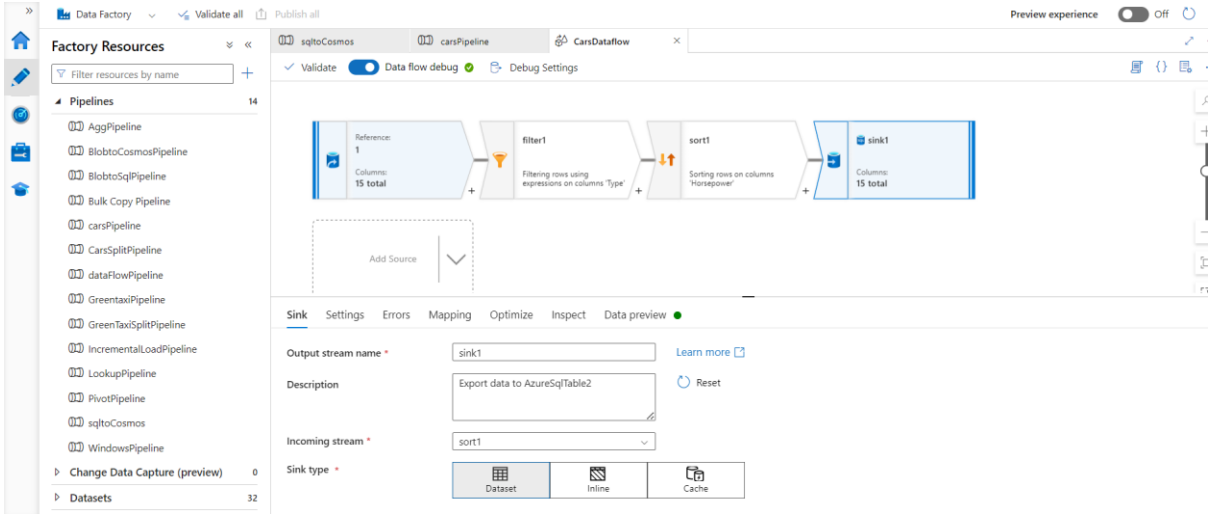

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 14
AggPipeline
BlobtoCosmosPipeline
BlobtoSqlPipeline
Bulk Copy Pipeline
carsPipeline
CarsSplitPipeline
dataFlowPipeline
GreentaxiPipeline
GreenTaxiSplitPipeline
IncrementalLoadPipeline
LookupPipeline
PivotPipeline
sqltoCosmos
WindowsPipeline
Change Data Capture (preview) 0
Datasets 32
sqltoCosmos
carsPipeline
CarsDataflow
Validate
Data flow debug
source1
Columns:
15 total
filter1
Filtering rows using expressions on columns 'Type'
sort1
Sorting rows on columns 'Horsepower'
sink1
Export data to AzureSqlTable2
Add Source
Source settings
Source options
Projection
Optimize
Inspect
Data preview
Column name
Type
Format
Make
Model
Type
Origin
DriveTrain
MSRP
string
Specify format

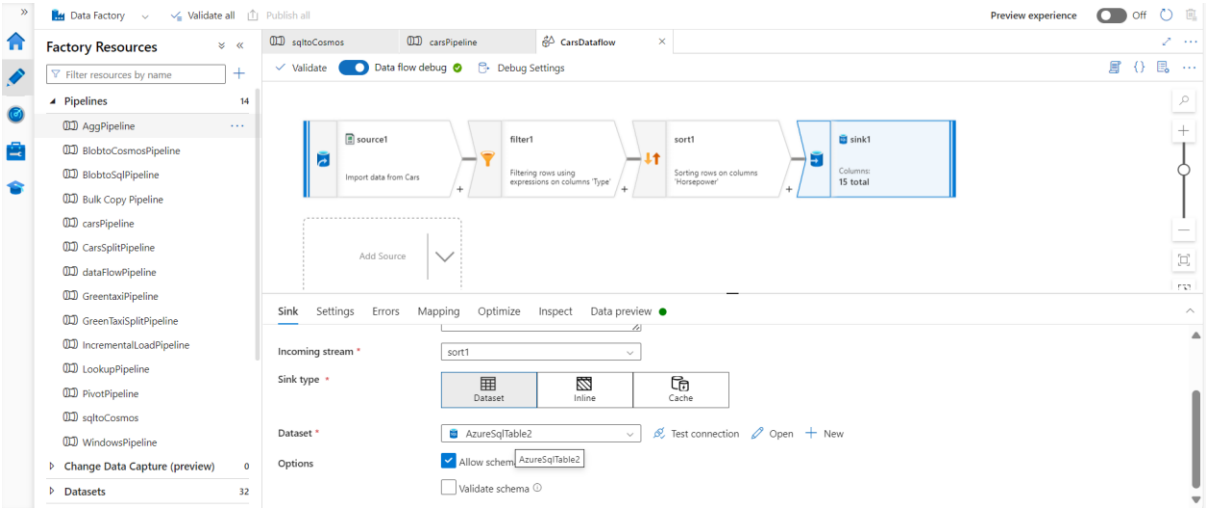

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 14
AggPipeline
BlobtoCosmosPipeline
BlobtoSqlPipeline
Bulk Copy Pipeline
carsPipeline
CarsSplitPipeline
dataFlowPipeline
GreentaxiPipeline
GreenTaxiSplitPipeline
IncrementalLoadPipeline
LookupPipeline
PivotPipeline
sqltoCosmos
WindowsPipeline
Change Data Capture (preview) 0
Datasets 32
sqltoCosmos
carsPipeline
CarsDataflow
Validate
Data flow debug
Debug Settings
source1
Columns:
15 total
filter1
Filtering rows using expressions on columns 'Type'
sort1
Sorting rows on columns 'Horsepower'
sink1
Export data to AzureSqlTable2
Add Source
Source settings
Source options
Projection
Optimize
Inspect
Data preview
Number of columns Total 15
Order
Column
Type
1 Make string
2 Model string
3 Type string
4 Origin string
5 DriveTrain string

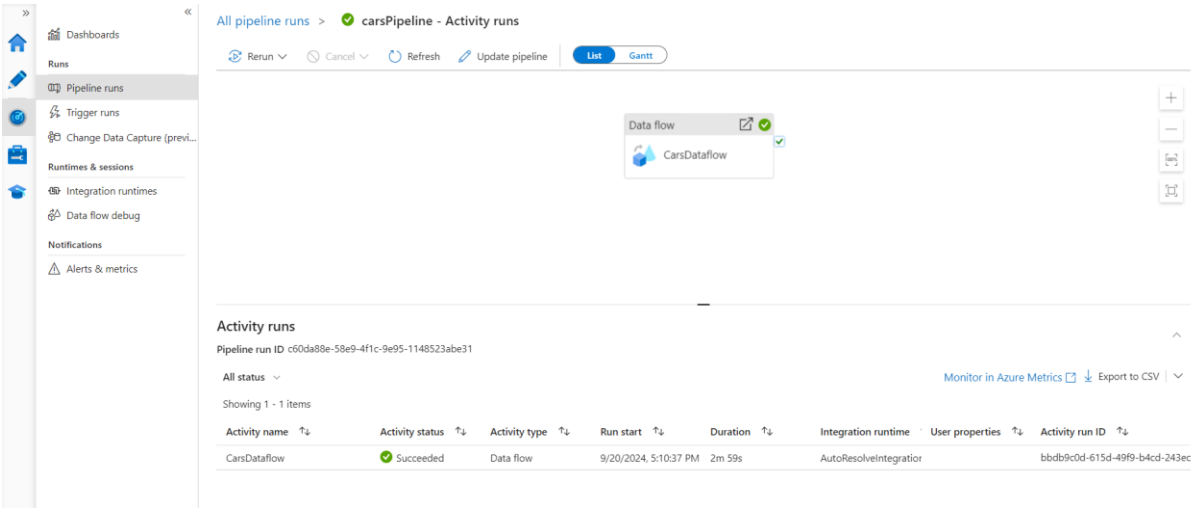

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 14
AggPipeline
BlobtoCosmosPipeline
BlobtoSqlPipeline
Bulk Copy Pipeline
carsPipeline
CarsSplitPipeline
dataFlowPipeline
GreentaxiPipeline
GreenTaxiSplitPipeline
IncrementalLoadPipeline
LookupPipeline
PivotPipeline
sqltoCosmos
WindowsPipeline
Change Data Capture (preview) 0
Datasets 32
sqltoCosmos
carsPipeline
CarsDataflow
Validate
Data flow debug
Debug Settings
source1
Import data from Cars
filter1
Columns:
15 total
sort1
Sorting rows on columns 'Horsepower'
sink1
Export data to AzureSqlTable2
Add Source
Filter settings
Optimize
Inspect
Data preview
Output stream name *
filter1
Learn more
Description
Filtering rows using expressions on columns 'Type'
Reset
Incoming stream *
source1
Filter on *
Type == "SUV"


Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 14
AggPipeline
BlobtoCosmosPipeline
BlobtoSqlPipeline
Bulk Copy Pipeline
carsPipeline
CarsSplitPipeline
dataFlowPipeline
GreentaxiPipeline
GreenTaxiSplitPipeline
IncrementalLoadPipeline
LookupPipeline
PivotPipeline
sqltoCosmos
WindowsPipeline
Change Data Capture (preview) 0
Datasets 32
sqltoCosmos
carsPipeline
CarsDataflow
Validate
Data flow debug
Debug Settings
source1
Import data from Cars
filter1
Filtering rows using expressions on columns 'Type'
sort1
Columns:
15 total
sink1
Export data to AzureSqlTable2
Add Source
Sort settings
Optimize
Inspect
Data preview
Output stream name *
sort1
Learn more
Description
Sorting rows on columns 'Horsepower'
Reset
Incoming stream *
filter1
Options *
Case insensitive
Sort only within partition

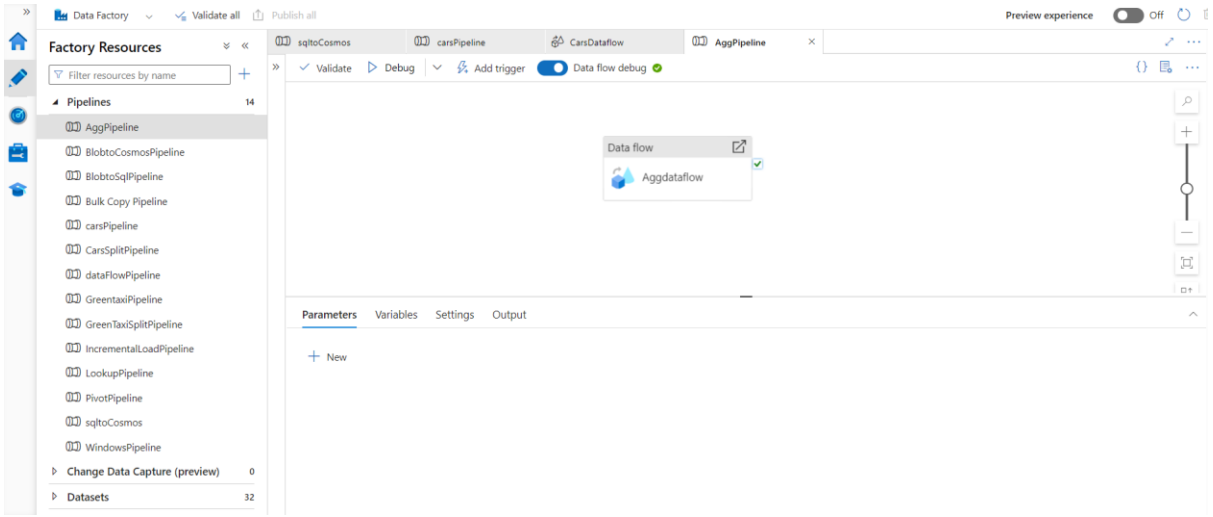

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 14
AggPipeline
BlobtoCosmosPipeline
BlobtoSqlPipeline
carsPipeline
CarsSplitPipeline
dataFlowPipeline
GreentaxiPipeline
GreenTaxiSplitPipeline
IncrementalLoadPipeline
LookupPipeline
PivotPipeline
sqltoCosmos
WindowsPipeline
Change Data Capture (preview) 0
Datasets 32
sqltoCosmos
carsPipeline
CarsDataflow
Validate
Data flow debug
Debug Settings
source1
Import data from Cars
filter1
Filtering rows using expressions on columns 'Type'
sort1
Columns:
15 total
sink1
Export data to AzureSqlTable2
Add Source
Sort settings
Optimize
Inspect
Data preview
Number of rows
INSERT 60
UPDATE 0
DELETE 0
UPSERT 0
Refresh
Typecast
Modify
Map drifted
Statistics
Remove
Export to CSV
Make
Model
Type
Origin
Drive...
MSRP
Invoice
Engin...
Cylind...
Horse...
Mazda
Tribut...
SUV
Asia
All
21087
19742
2
4
130
Saturn
VUE
SUV
USA
All
20585
19238
2.2
4
143
Jeep
Libert...
SUV
USA
All
20130
18973
2.4
4
150

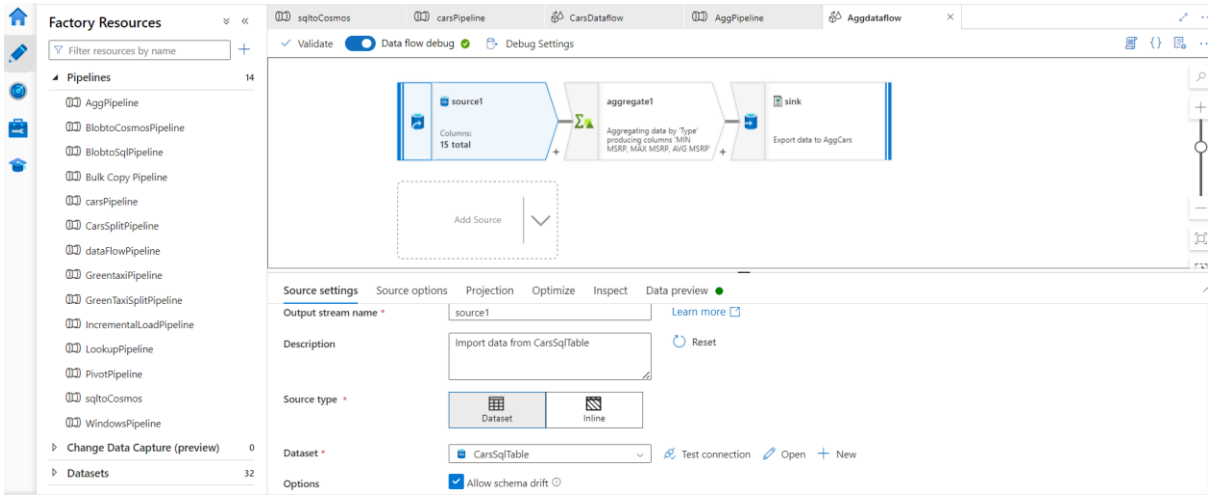

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 14
AggPipeline
BlobtoCosmosPipeline
BlobtoSqlPipeline
Bulk Copy Pipeline
carsPipeline
CarsSplitPipeline
dataFlowPipeline
GreentaxiPipeline
GreenTaxiSplitPipeline
IncrementalLoadPipeline
LookupPipeline
PivotPipeline
sqltoCosmos
WindowsPipeline
Change Data Capture (preview) 0
Datasets 32
sqltoCosmos
carsPipeline
CarsDataflow
Validate
Data flow debug
Debug Settings
Reference:
1
Columns:
15 total
filter1
Filtering rows using expressions on columns 'Type'
sort1
Sorting rows on columns 'Horsepower'
sink1
Columns:
15 total
Add Source
Sink
Settings
Errors
Mapping
Optimize
Inspect
Data preview
Output stream name *
sink1
Learn more
Description
Export data to AzureSqlTable2
Reset
Incoming stream *
sort1
Sink type *
Dataset
Inline
Cache

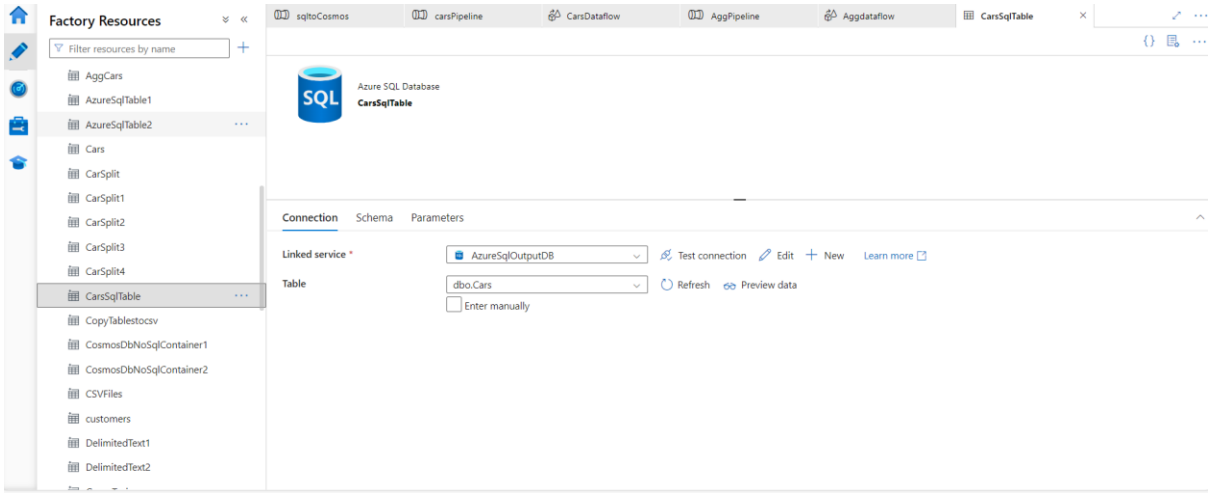


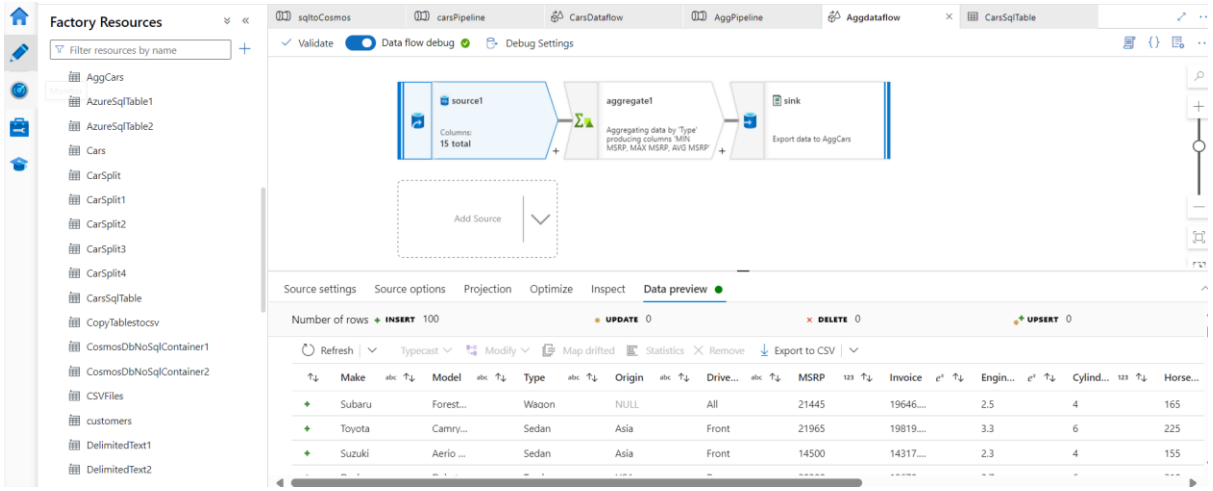


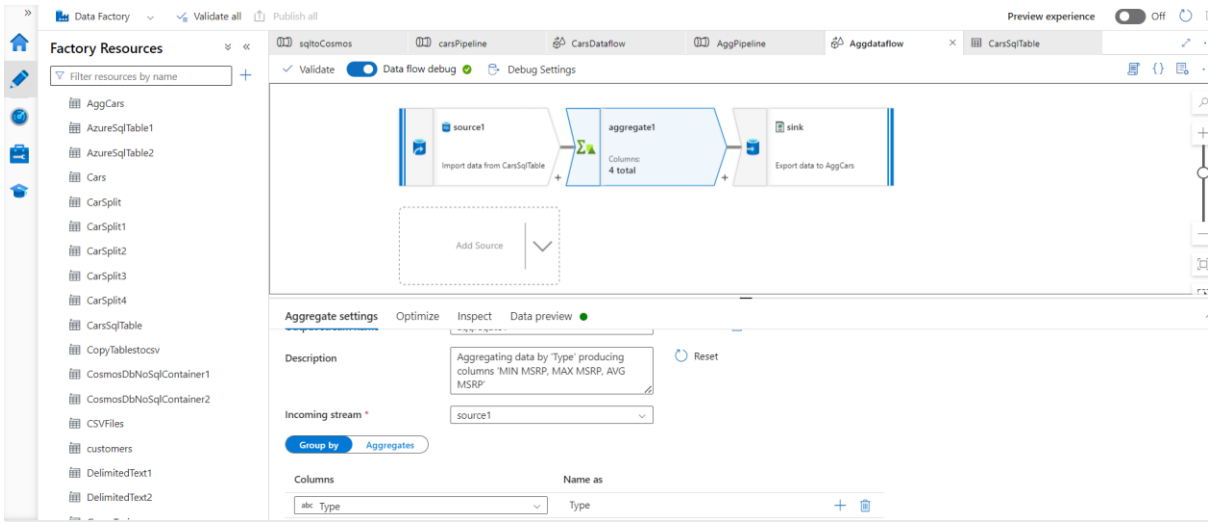
3. Design an ADF pipeline to implement aggregate operations, such as sum, average, max, min and count, within an Azure Data Flow.

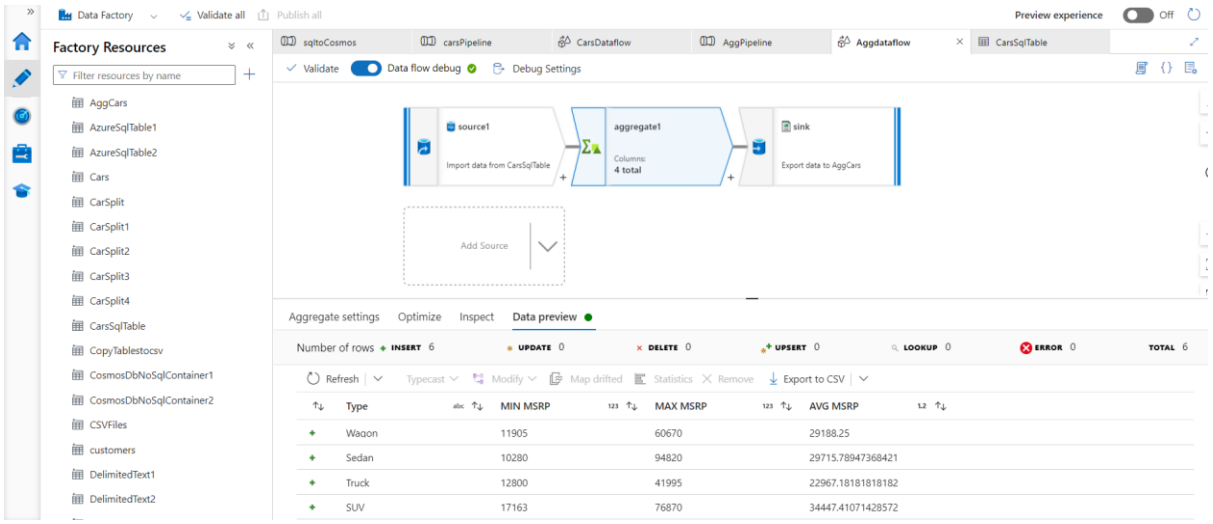

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 14
AggPipeline
BlobtoCosmosPipeline
BlobtoSqlPipeline
Bulk Copy Pipeline
carsPipeline
CarsSplitPipeline
dataFlowPipeline
GreentaxiPipeline
GreenTaxiSplitPipeline
IncrementalLoadPipeline
LookupPipeline
PivotPipeline
sqltoCosmos
WindowsPipeline
Change Data Capture (preview) 0
Datasets 32
sqltoCosmos
carsPipeline
CarsDataflow
AggPipeline
Validate
Debug
Add trigger
Data flow debug
Data flow
Aggdataflow
Parameters
Variables
Settings
Output
New

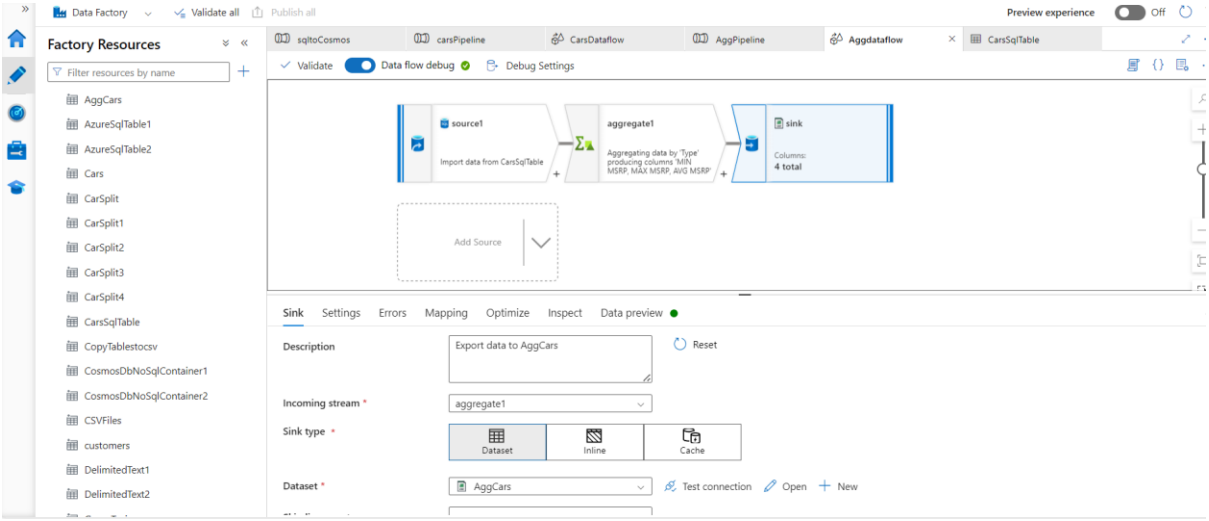

Factory Resources
Filter resources by name
Pipelines 14
AggPipeline
BlobtoCosmosPipeline
BlobtoSqlPipeline
Bulk Copy Pipeline
carsPipeline
CarsSplitPipeline
dataFlowPipeline
GreentaxiPipeline
GreenTaxiSplitPipeline
IncrementalLoadPipeline
LookupPipeline
PivotPipeline
sqltoCosmos
WindowsPipeline
Change Data Capture (preview) 0
Datasets 32
sqltoCosmos
carsPipeline
CarsDataflow
AggPipeline
Aggdataflow
Validate
Data flow debug
Debug Settings
source1
Columns:
15 total
aggregate1
Aggregating data by 'Type'
producing columns 'MIN
MSRP, MAX MSRP, AVG MSRP'
sink
Export data to AggCars
Add Source
Source settings
Source options
Projection
Optimize
Inspect
Data preview
Output stream name *
source1
Learn more
Description
Import data from CarsSqlTable
Reset
Source type *
Dataset
Inline
Dataset *
CarsSqlTable
Test connection
Open
New
Options
Allow schema drift

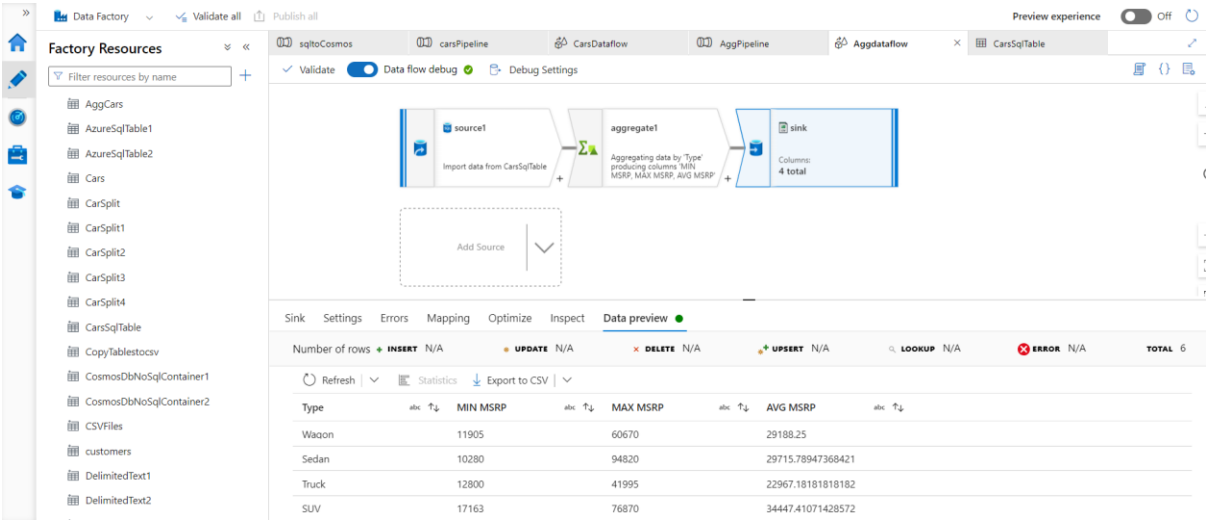

Factory Resources
Filter resources by name
AggCars
AzureSqlTable1
AzureSqlTable2
Cars
CarSplit
CarSplit1
CarSplit2
CarSplit3
CarSplit4
CarsSqlTable
CopyTablestocsv
CosmosDbNoSqlContainer1
CosmosDbNoSqlContainer2
CSVFiles
customers
DelimitedText1
DelimitedText2
sqltoCosmos
carsPipeline
CarsDataflow
AggPipeline
Aggdataflow
CarsSqlTable
Azure SQL Database
CarsSqlTable
Connection
Schema
Parameters
Linked service *
AzureSqlOutputDB
Test connection
Edit
New
Learn more
Table
dbo.Cars
Refresh
Preview data
Enter manually

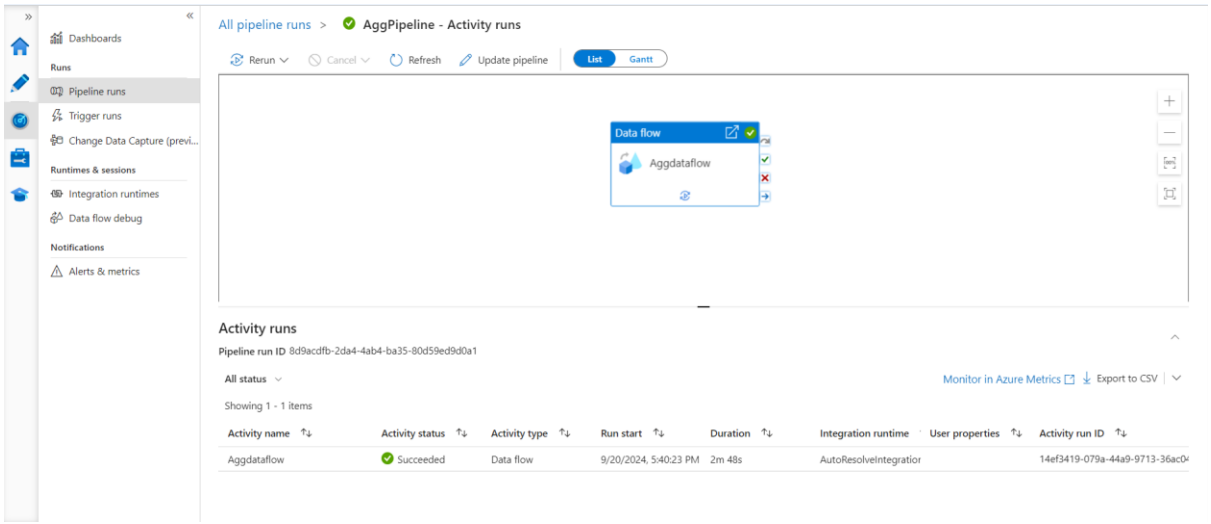

Factory Resources
Filter resources by name
AggCars
AzureSqlTable1
AzureSqlTable2
Cars
CarSplit
CarSplit1
CarSplit2
CarSplit3
CarSplit4
CarsSqlTable
CopyTablestocsv
CosmosDbNoSqlContainer1
CosmosDbNoSqlContainer2
CSVFiles
customers
DelimitedText1
DelimitedText2
sqltoCosmos
carsPipeline
CarsDataflow
AggPipeline
Aggdataflow
CarsSqlTable
Validate
Data flow debug
Debug Settings
source1
Columns:
15 total
aggregate1
Aggregating data by 'Type' producing columns 'MIN MSRP, MAX MSRP, AVG MSRP'
sink
Export data to AggCars
Add Source
Source settings
Source options
Projection
Optimize
Inspect
Data preview
Number of rows
INSERT 100
UPDATE 0
DELETE 0
UPSERT 0
Refresh
Typecast
Modify
Map drifted
Statistics
Remove
Export to CSV
Make
Model
Type
Origin
Drive...
MSRP
Invoice
Engin...
Cylind...
Horse...
Subaru
Forest...
Wagon
NULL
All
21445
19646....
2.5
4
165
Toyota
Camry...
Sedan
Asia
Front
21965
19819....
3.3
6
225
Suzuki
Aerio ...
Sedan
Asia
Front
14500
14317....
2.3
4
155


Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
AggCars
AzureSqlTable1
AzureSqlTable2
Cars
CarSplit
CarSplit1
CarSplit2
CarSplit3
CarSplit4
CarsSqlTable
CopyTablestocsv
CosmosDbNoSqlContainer1
CosmosDbNoSqlContainer2
CSVFiles
customers
DelimitedText1
DelimitedText2
sqltoCosmos
carsPipeline
CarsDataflow
AggPipeline
Aggdataflow
CarsSqlTable
Validate
Data flow debug
Debug Settings
source1
Import data from CarsSqlTable
aggregate1
Columns:
4 total
sink
Export data to AggCars
Add Source
Aggregate settings
Optimize
Inspect
Data preview
Description
Aggregating data by 'Type' producing columns 'MIN MSRP, MAX MSRP, AVG MSRP'
Reset
Incoming stream *
source1
Group by
Aggregates
Columns
Name as
Type
Type

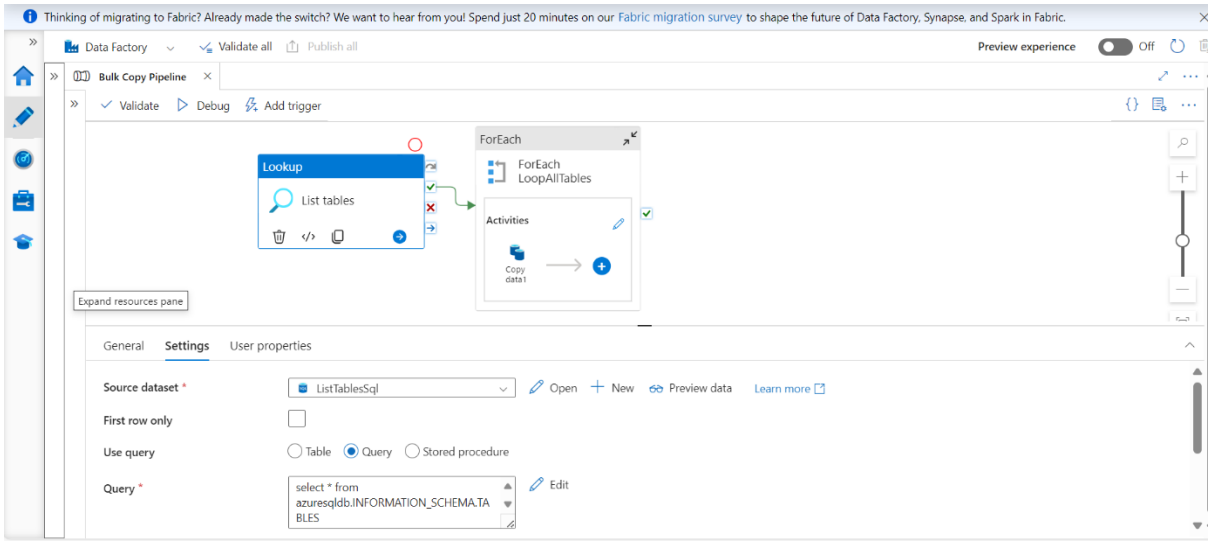

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
AggCars
AzureSqlTable1
AzureSqlTable2
Cars
CarSplit
CarSplit1
CarSplit2
CarSplit3
CarSplit4
CarsSqlTable
CopyTablestocsv
CosmosDbNoSqlContainer1
CosmosDbNoSqlContainer2
CSVFiles
customers
DelimitedText1
DelimitedText2
sqltoCosmos
carsPipeline
CarsDataflow
AggPipeline
Aggdataflow
CarsSqlTable
Validate
Data flow debug
Debug Settings
source1
Import data from CarsSqlTable
aggregate1
Columns:
4 total
sink
Export data to AggCars
Add Source
Aggregate settings
Optimize
Inspect
Data preview
Number of rows
INSERT 6
UPDATE 0
DELETE 0
UPSERT 0
LOOKUP 0
ERROR 0
TOTAL 6
Refresh
Typecast
Modify
Map drifted
Statistics
Remove
Export to CSV
Type
MIN MSRP
MAX MSRP
AVG MSRP
Wagon
11905
60670
29188.25
Sedan
10280
94820
29715.78947368421
Truck
12800
41995
22967.18181818182
SUV
17163
76870
34447.41071428572

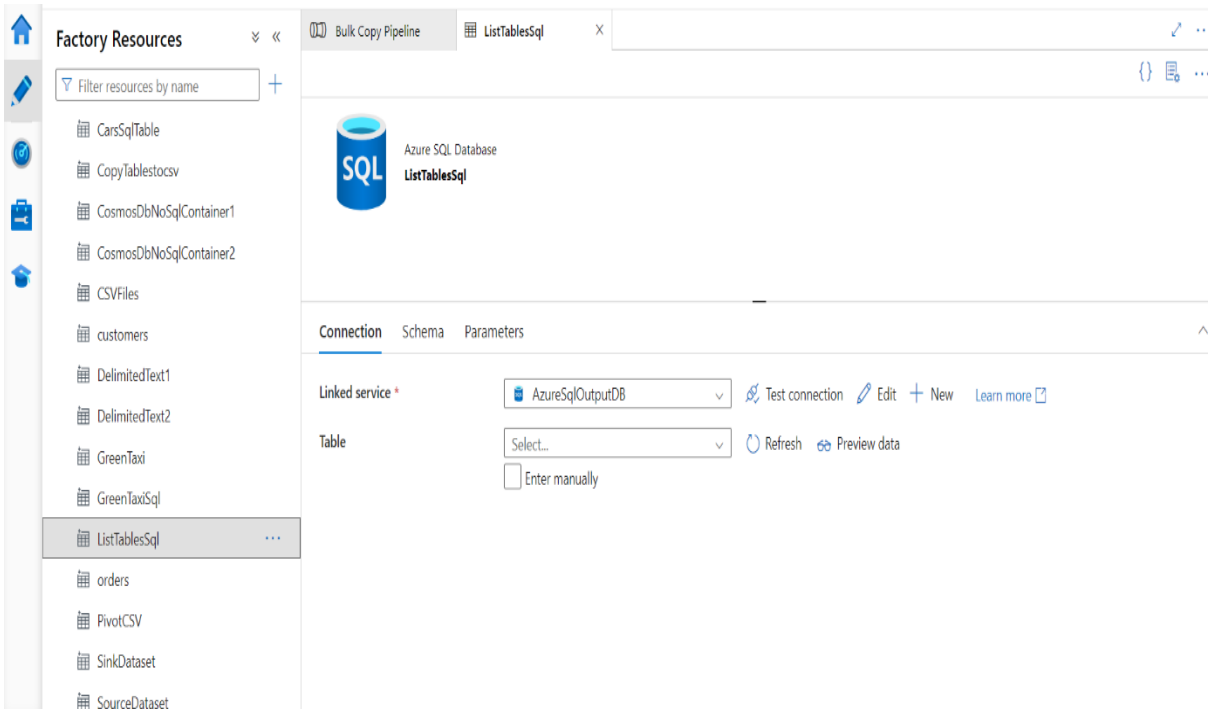

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
AggCars
AzureSqlTable1
AzureSqlTable2
Cars
CarSplit
CarSplit1
CarSplit2
CarSplit3
CarSplit4
CarsSqlTable
CopyTablestocsv
CosmosDbNoSqlContainer1
CosmosDbNoSqlContainer2
CSVFiles
customers
DelimitedText1
DelimitedText2
sqltoCosmos
carsPipeline
CarsDataflow
AggPipeline
Aggdataflow
CarsSqlTable
Validate
Data flow debug
Debug Settings
source1
Import data from CarsSqlTable
aggregate1
Aggregating data by 'Type' producing columns 'MIN MSRP, MAX MSRP, AVG MSRP'
sink
Columns:
4 total
Add Source
Sink
Settings
Errors
Mapping
Optimize
Inspect
Data preview
Description
Export data to AggCars
Reset
Incoming stream *
aggregate1
Sink type *
Dataset
Inline
Cache
Dataset *
AggCars
Test connection
Open
New

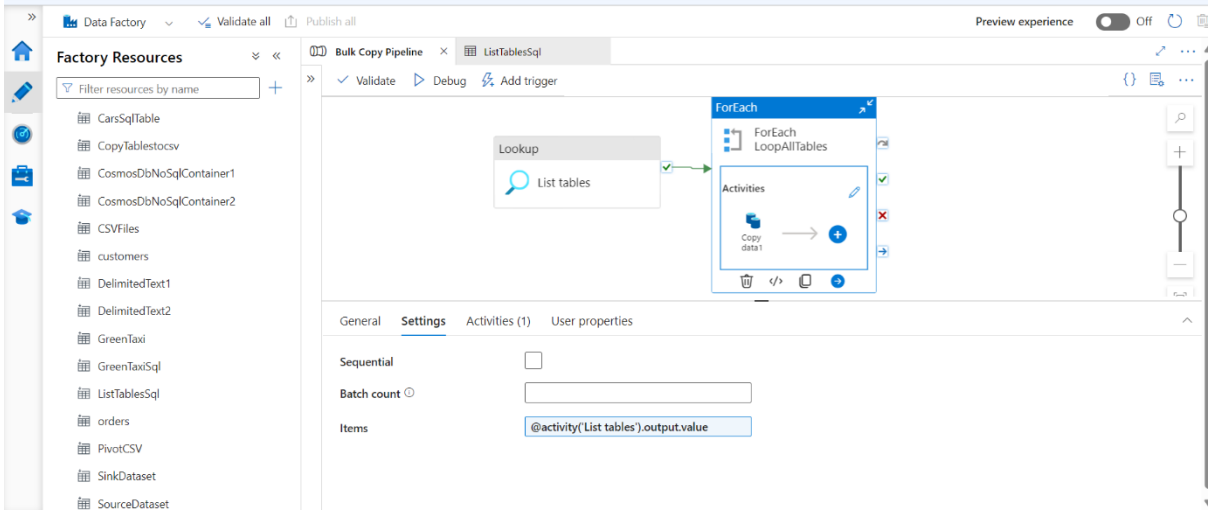

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
AggCars
AzureSqlTable1
AzureSqlTable2
Cars
CarSplit
CarSplit1
CarSplit2
CarSplit3
CarSplit4
CarsSqlTable
CopyTablestocsv
CosmosDbNoSqlContainer1
CosmosDbNoSqlContainer2
CSVFiles
customers
DelimitedText1
DelimitedText2
sqltoCosmos
carsPipeline
CarsDataflow
AggPipeline
Aggdataflow
CarsSqlTable
Validate
Data flow debug
Debug Settings
source1
Import data from CarsSqlTable
aggregate1
Aggregating data by 'Type' producing columns 'MIN MSRP, MAX MSRP, AVG MSRP'
sink
Columns:
4 total
Add Source
Sink
Settings
Errors
Mapping
Optimize
Inspect
Data preview
Number of rows
INSERT N/A
UPDATE N/A
DELETE N/A
UPSERT N/A
LOOKUP N/A
ERROR N/A
TOTAL 6
Refresh
Statistics
Export to CSV
Type
MIN MSRP
MAX MSRP
AVG MSRP
Wagon 11905 60670 29188.25
Sedan 10280 94820 29715.78947368421
Truck 12800 41995 22967.18181818182
SUV 17163 76870 34447.41071428572

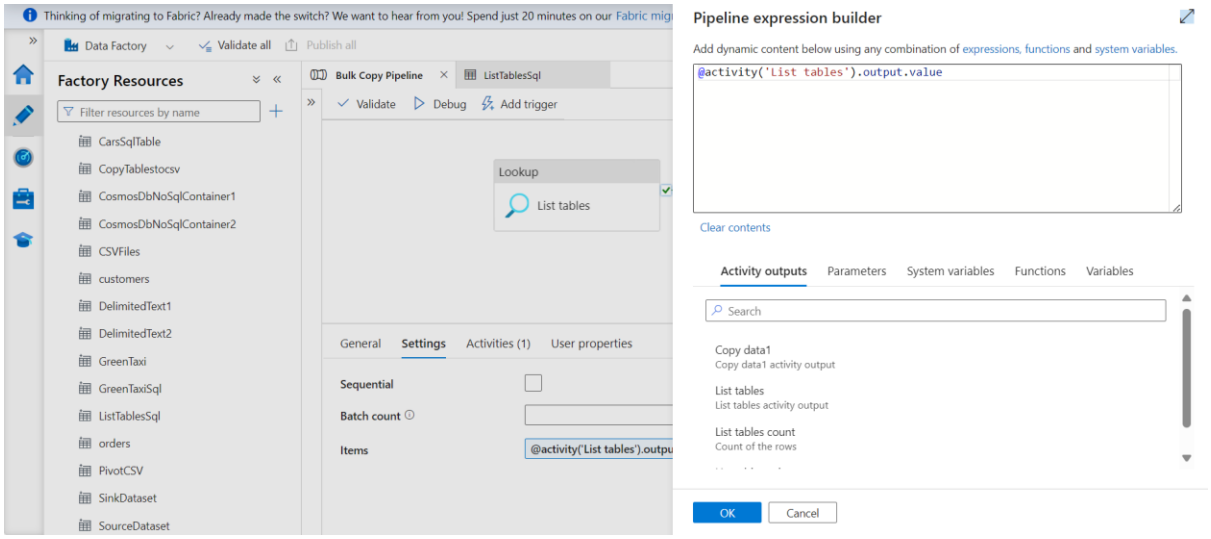


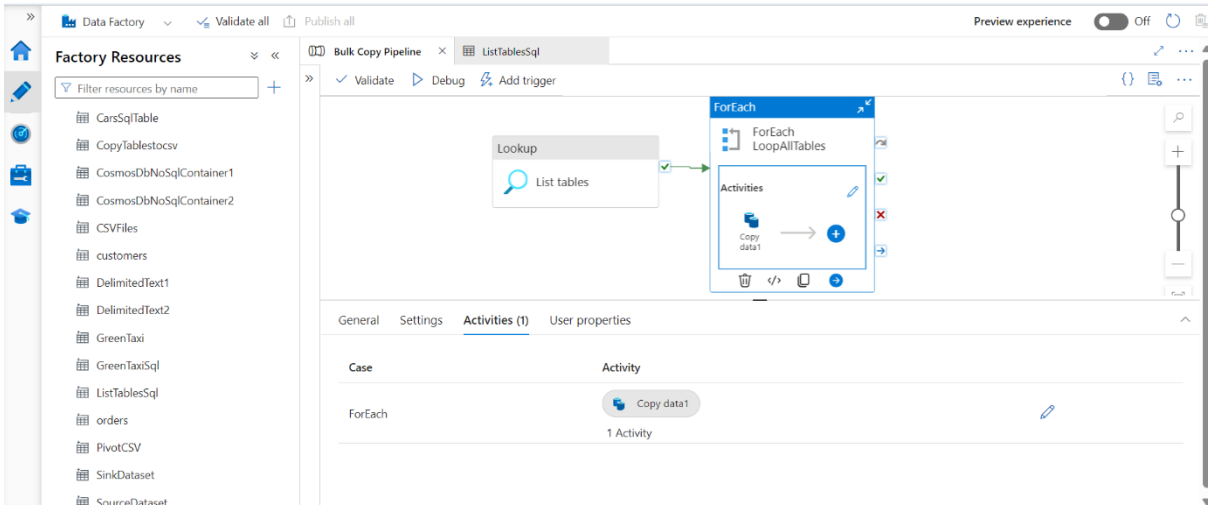
4. Create best approach to bulk copy data from multiple homogenous sources into Azure SQL Database using ADF pipelines. Show usage of Lookup, For Each Loop and Expressions in Azure Data Factory.

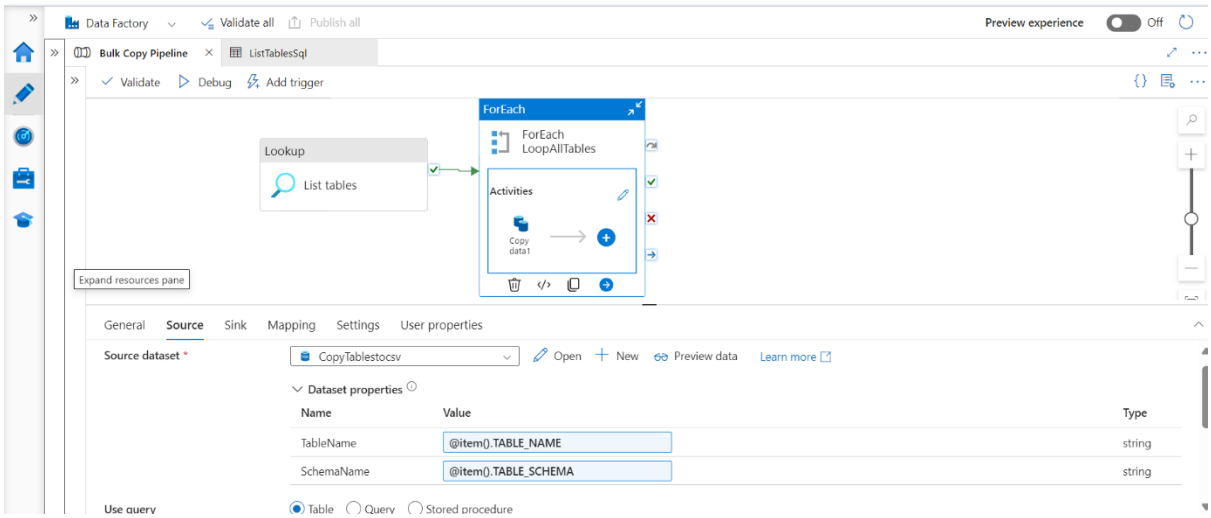


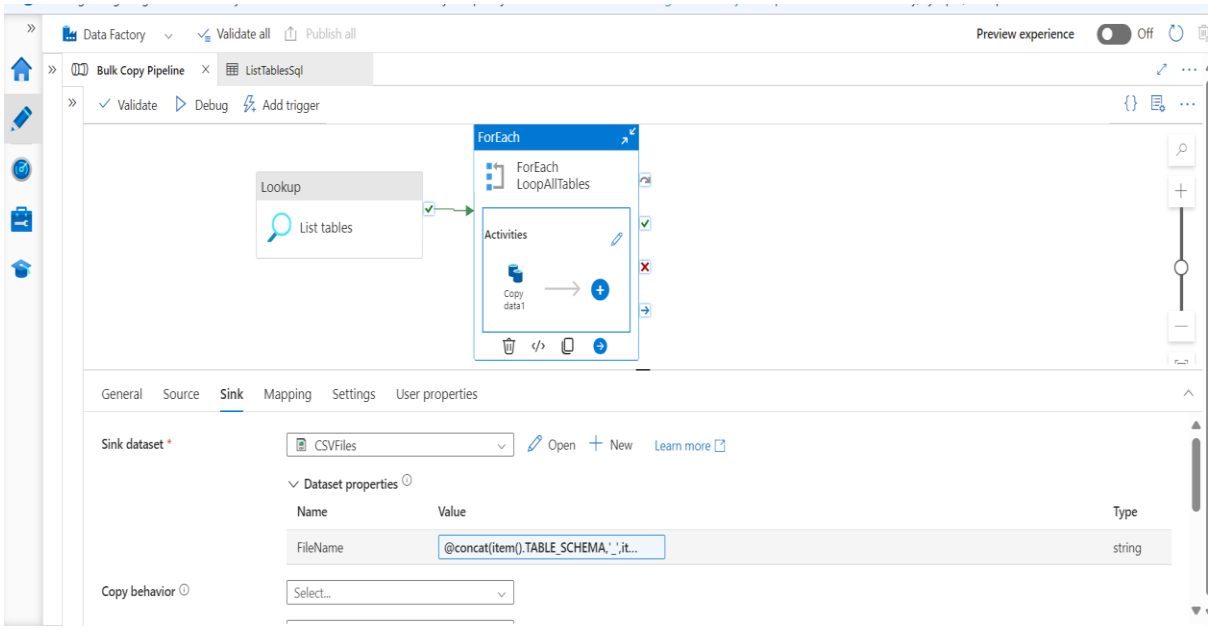

Factory Resources
Filter resources by name
CarsSqlTable
CopyTablestocsv
CosmosDbNoSqlContainer1
CosmosDbNoSqlContainer2
CSVFiles
customers
DelimitedText1
DelimitedText2
GreenTaxi
GreenTaxiSql
ListTablesSql
orders
PivotCSV
SinkDataset
SourceDataset
Bulk Copy Pipeline
ListTablesSql
Azure SQL Database
ListTablesSql
Connection
Schema
Parameters
Linked service *
AzureSqlOutputDB
Test connection
Edit
New
Learn more
Table
Select...
Refresh
Preview data
Enter manually

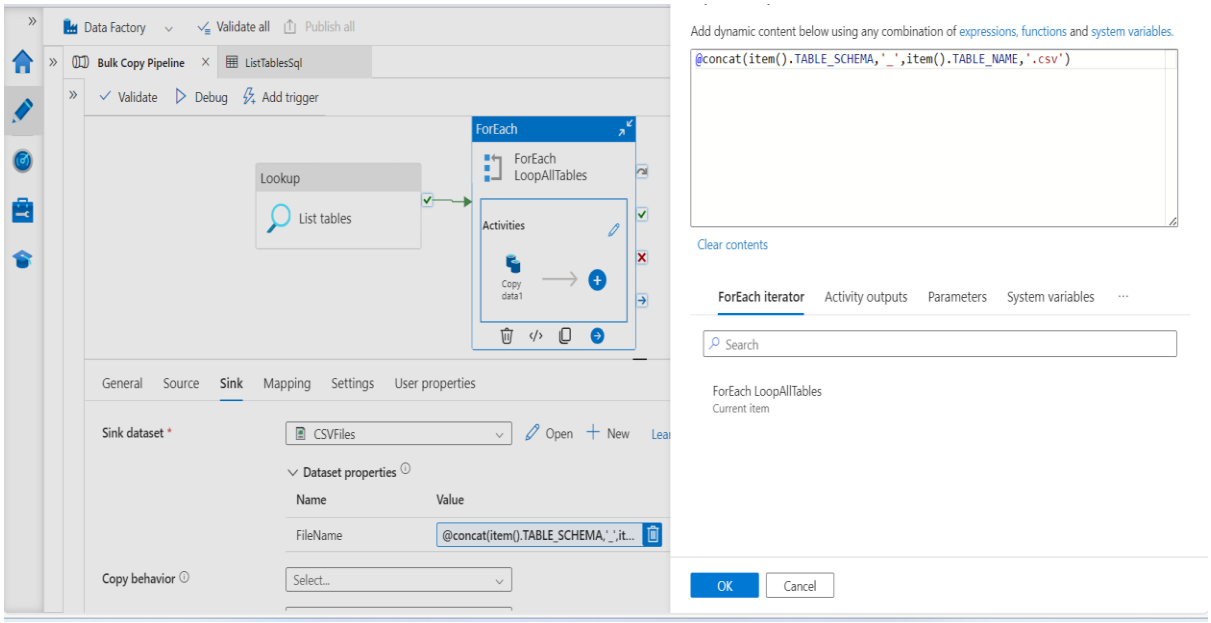

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
CarsSqlTable
CopyTablestocsv
CosmosDbNoSqlContainer1
CosmosDbNoSqlContainer2
CSVFiles
customers
DelimitedText1
DelimitedText2
GreenTaxi
GreenTaxiSql
ListTablesSql
orders
PivotCSV
SinkDataset
SourceDataset
Bulk Copy Pipeline
ListTablesSql
Validate
Debug
Add trigger
Lookup
List tables
ForEach
ForEach LoopAllTables
Activities
Copy data1
General
Settings
Activities (1)
User properties
Sequential
Batch count
Items
@activity('List tables').output.value

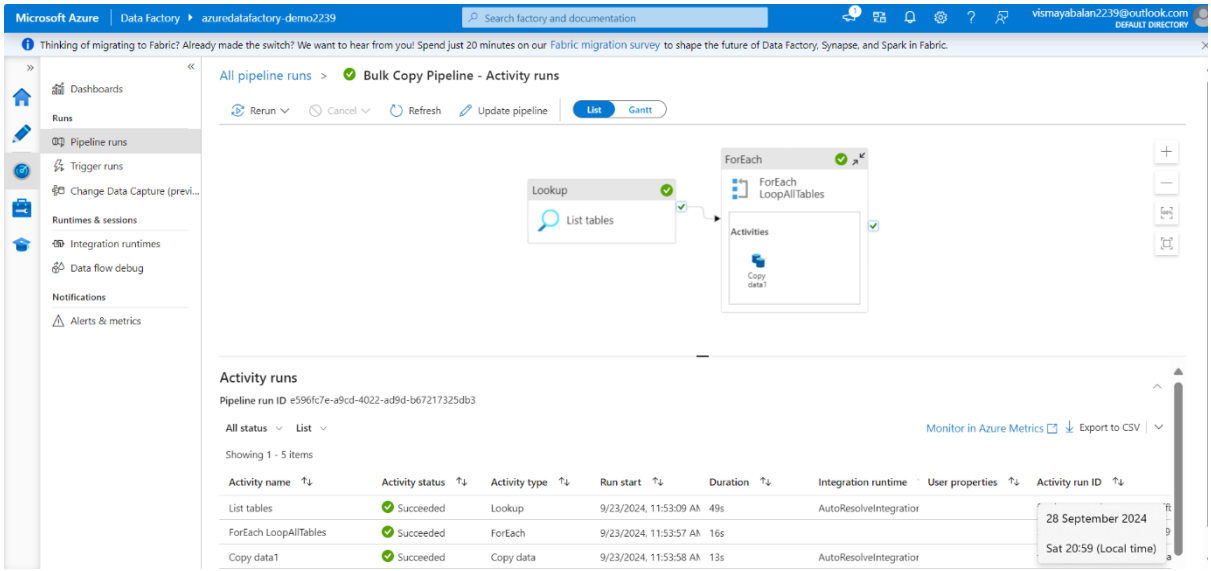

Thinking of migrating to Fabric? Already made the switch? We want to hear from you! Spend just 20 minutes on our Fabric mig
Data Factory
Validate all
Publish all
Factory Resources
Filter resources by name
CarsSqlTable
CopyTablestocsv
CosmosDbNoSqlContainer1
CosmosDbNoSqlContainer2
CSVFiles
customers
DelimitedText1
DelimitedText2
GreenTaxi
GreenTaxiSql
ListTablesSql
orders
PivotCSV
SinkDataset
SourceDataset
Bulk Copy Pipeline
ListTablesSql
Validate
Debug
Add trigger
Lookup
List tables
General
Settings
Activities (1)
User properties
Sequential
Batch count
Items
@activity('List tables').outpu
Pipeline expression builder
Add dynamic content below using any combination of expressions, functions and system variables.
@activity('List tables').output.value
Clear contents
Activity outputs
Parameters
System variables
Functions
Variables
Search
Copy data1
Copy data1 activity output
List tables
List tables activity output
List tables count
Count of the rows
OK
Cancel


Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
CarsSqlTable
CopyTablestocsv
CosmosDbNoSqlContainer1
CosmosDbNoSqlContainer2
CSVFiles
customers
DelimitedText1
DelimitedText2
GreenTaxi
GreenTaxiSql
ListTablesSql
orders
PivotCSV
SinkDataset
Bulk Copy Pipeline
ListTablesSql
Validate
Debug
Add trigger
Lookup
List tables
ForEach
ForEach
LoopAllTables
Activities
Copy data1
General
Settings
Activities (1)
User properties
Case
Activity
ForEach
Copy data1
1 Activity

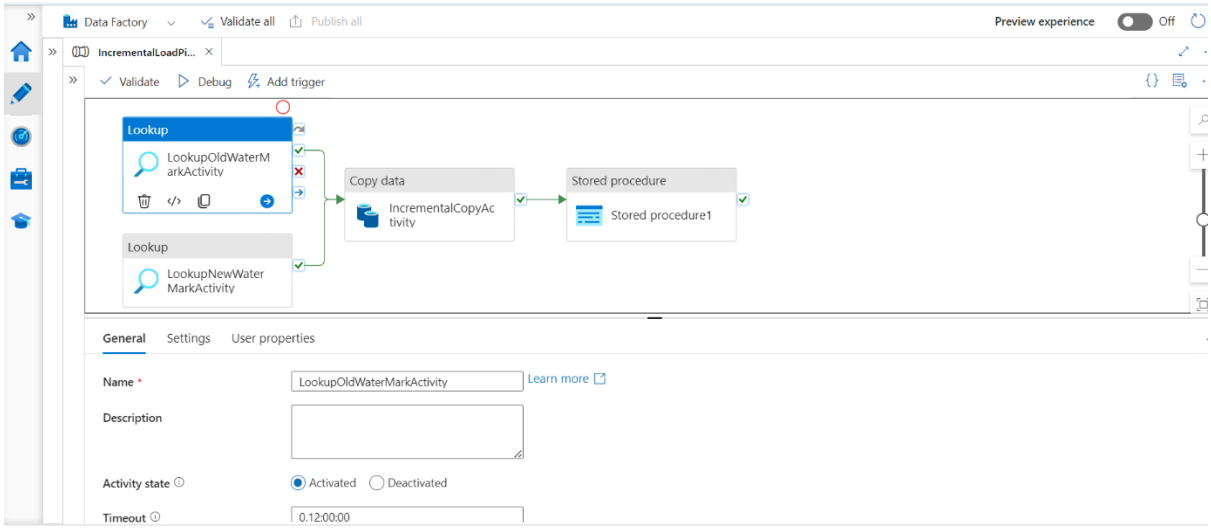

Data Factory
Validate all
Publish all
Preview experience
Off
Bulk Copy Pipeline
ListTablesSql
Validate
Debug
Add trigger
Lookup
List tables
ForEach
ForEach LoopAllTables
Activities
Copy data1
Expand resources pane
General
Source
Sink
Mapping
Settings
User properties
Source dataset *
CopyTablestocsv
Open
New
Preview data
Learn more
Dataset properties
Name
Value
Type
TableName
@item().TABLE_NAME
string
SchemaName
@item().TABLE_SCHEMA
string
Use query
Table
Query
Stored procedure

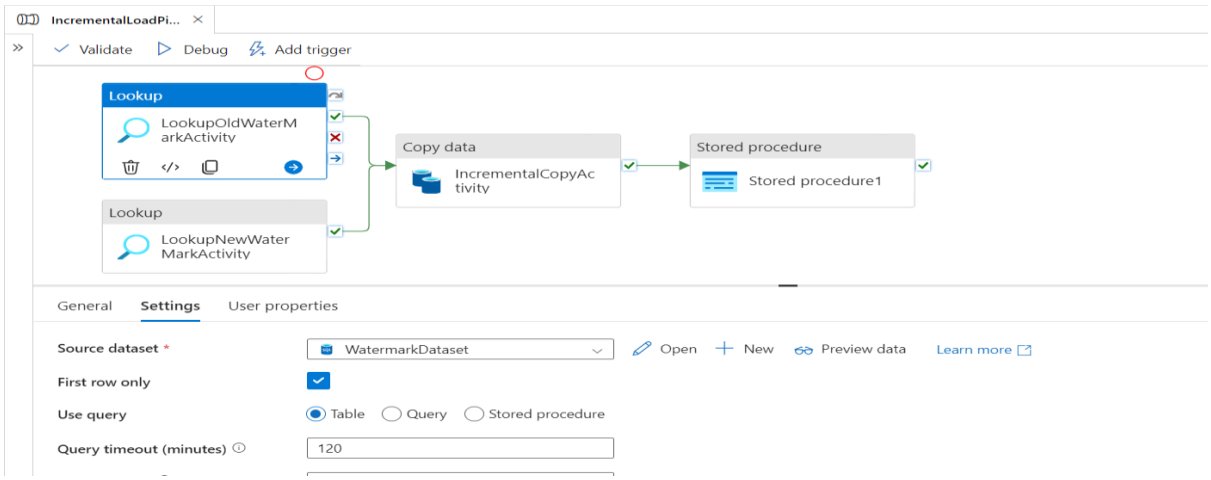

Data Factory
Validate all
Publish all
Preview experience
Off
Bulk Copy Pipeline
ListTablesSql
Validate
Debug
Add trigger
Lookup
List tables
ForEach
ForEach LoopAllTables
Activities
Copy data1
General
Source
Sink
Mapping
Settings
User properties
Sink dataset *
CSVFiles
Open
New
Learn more
Dataset properties
Name
Value
Type
FileName
@concat(item().TABLE_SCHEMA,'_',it...
string
Copy behavior
Select...

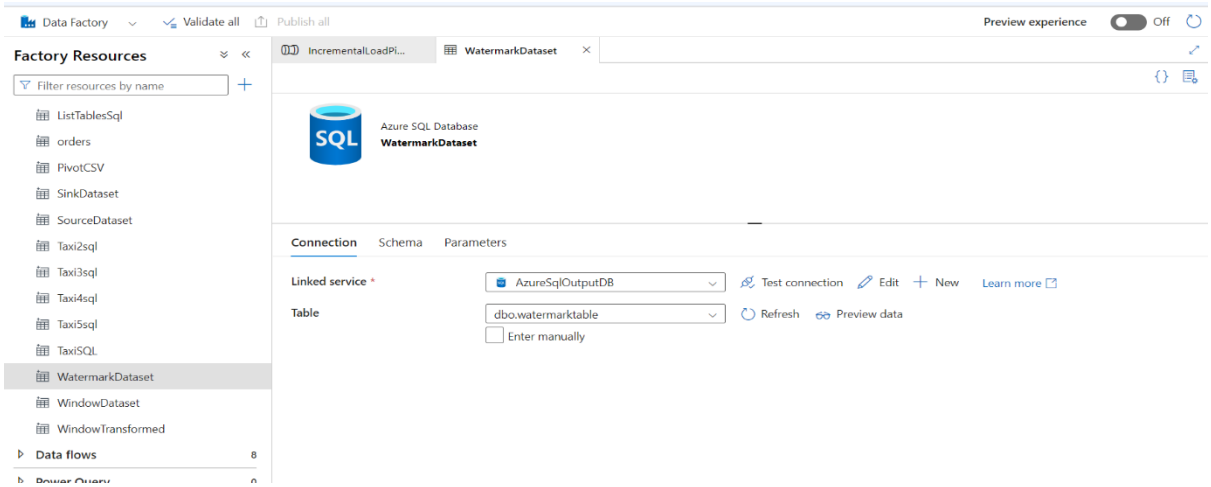


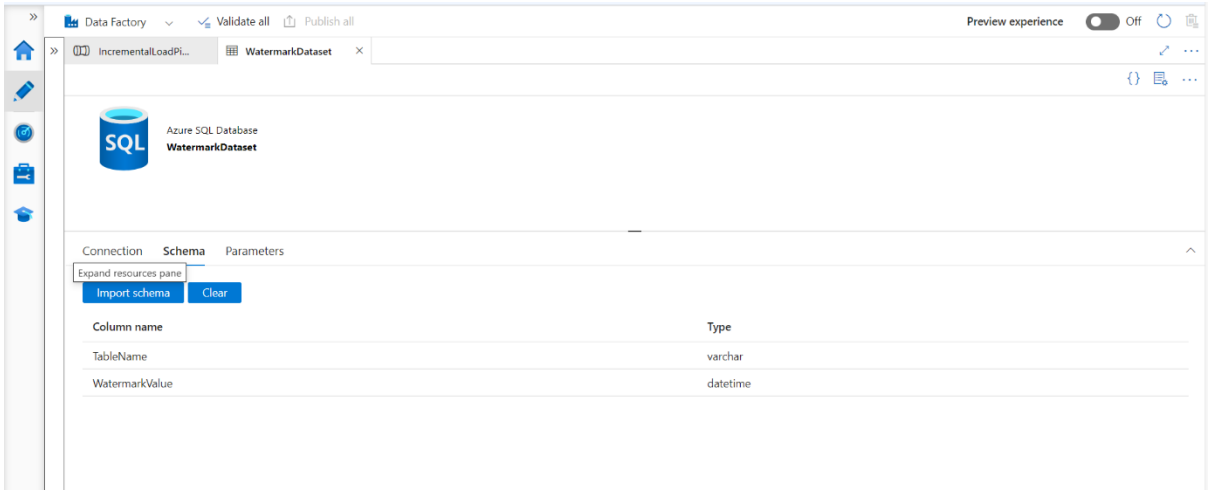


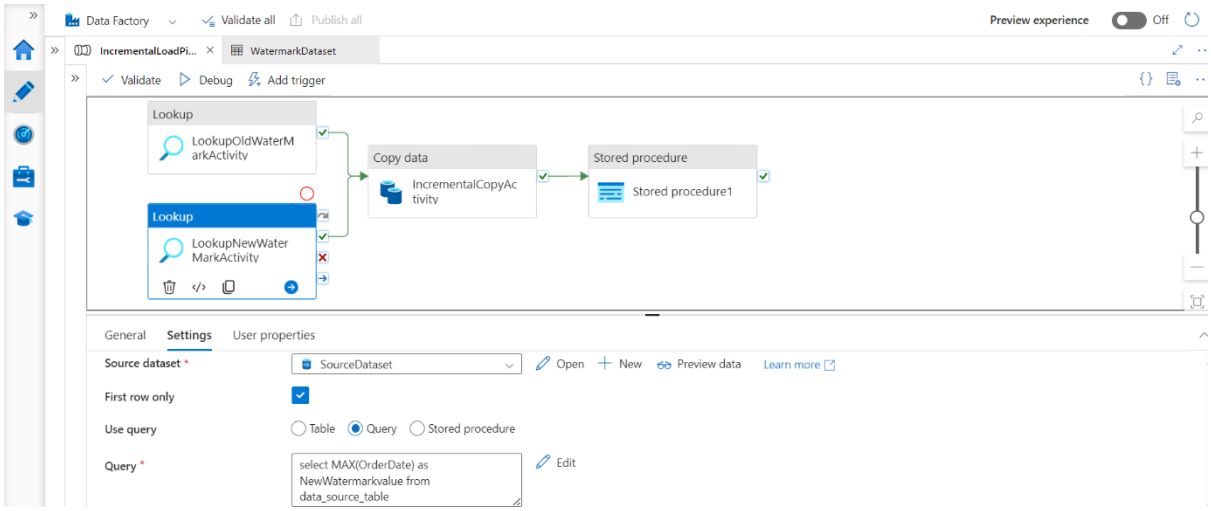
5. Implement incremental load Pipeline in Azure Data Factory for handling datasets, ensuring efficient insert/upsert/updates to the target storage without re-inserting the entire dataset?

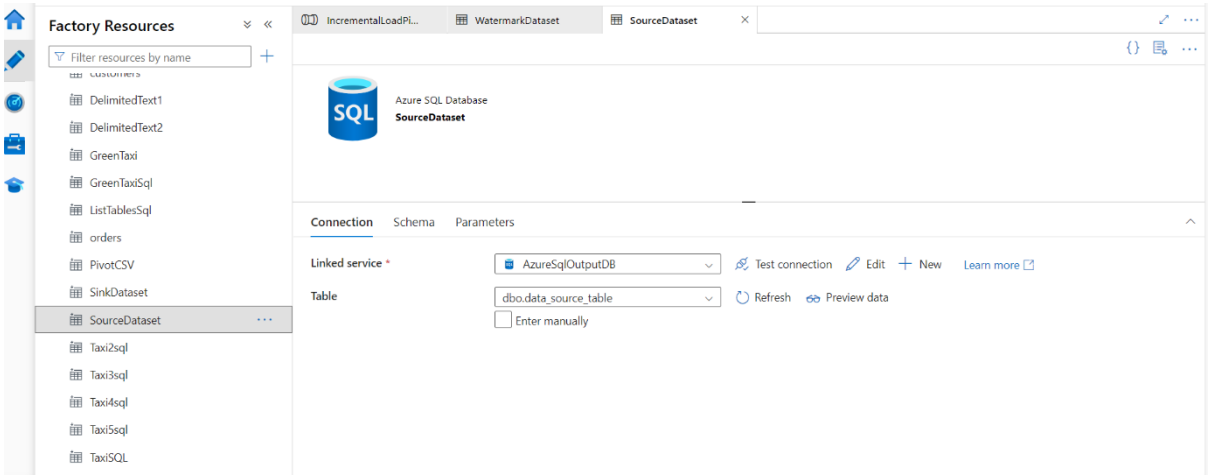

Data Factory
Validate all
Publish all
Preview experience
Off
IncrementalLoadPi...
Validate
Debug
Add trigger
Lookup
LookupOldWaterMarkActivity
Copy data
IncrementalCopyActivity
Stored procedure
Stored procedure1
Lookup
LookupNewWaterMarkActivity
General
Settings
User properties
Name *
LookupOldWaterMarkActivity
Learn more
Description
Activity state
Activated
Deactivated
Timeout
0.12:00:00

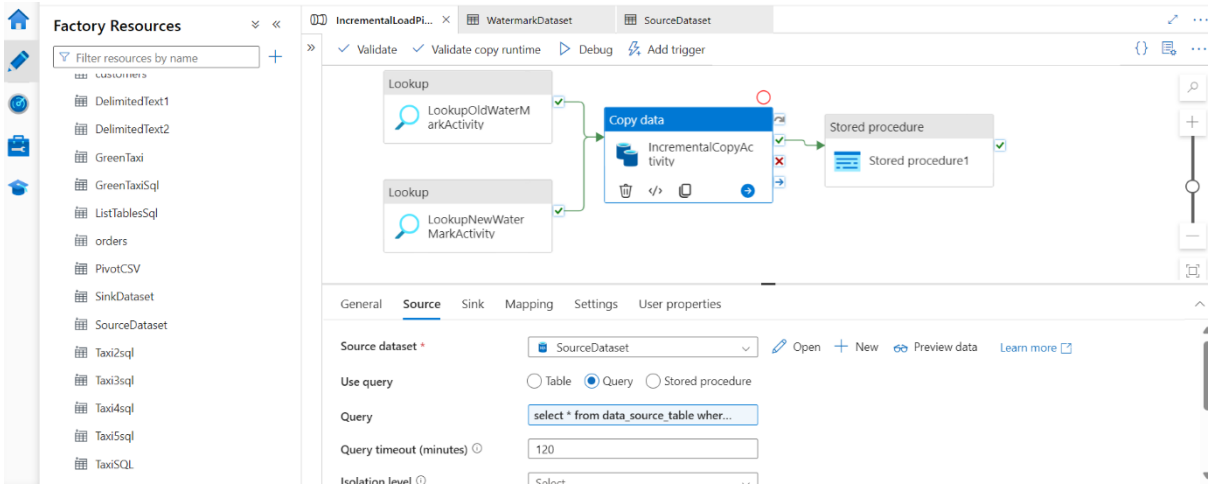

IncrementalLoadPi...
Validate
Debug
Add trigger
Lookup
LookupOldWaterMarkActivity
Copy data
IncrementalCopyActivity
Stored procedure
Stored procedure1
Lookup
LookupNewWaterMarkActivity
General
Settings
User properties
Source dataset *
WatermarkDataset
Open
New
Preview data
Learn more
First row only
Use query
Table
Query
Stored procedure
Query timeout (minutes)
120

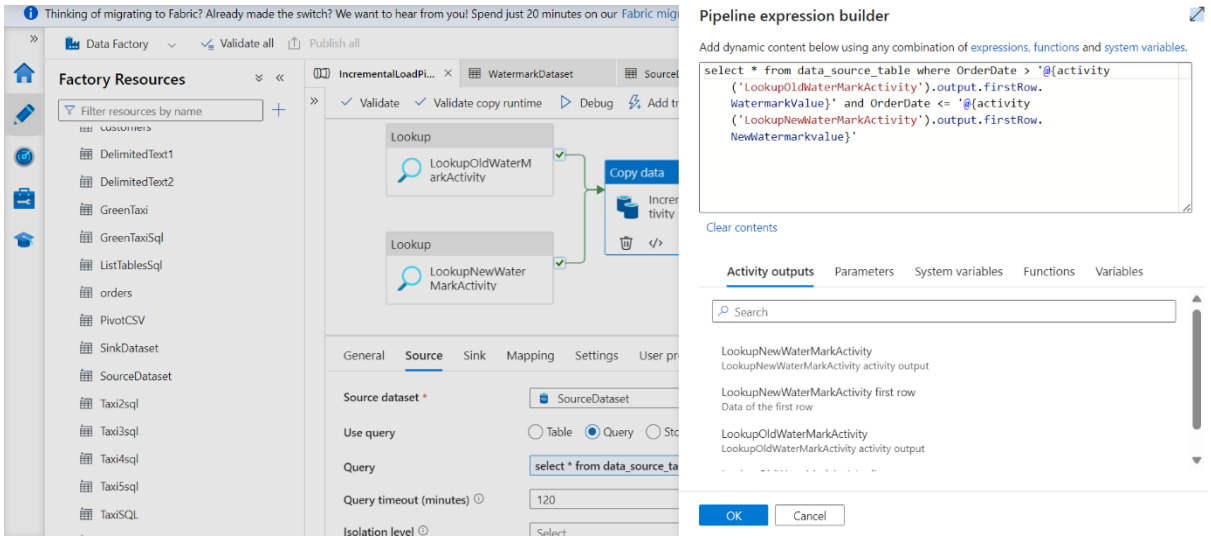

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
ListTablesSql
orders
PivotCSV
SinkDataset
SourceDataset
Taxi2sql
Taxi3sql
Taxi4sql
Taxi5sql
TaxiSQL
WatermarkDataset
WindowDataset
WindowTransformed
Data flows
8
IncrementalLoadPi...
WatermarkDataset
SQL
Azure SQL Database
WatermarkDataset
Connection
Schema
Parameters
Linked service *
AzureSqlOutputDB
Test connection
Edit
New
Learn more
Table
dbo.watermarktable
Refresh
Preview data
Enter manually

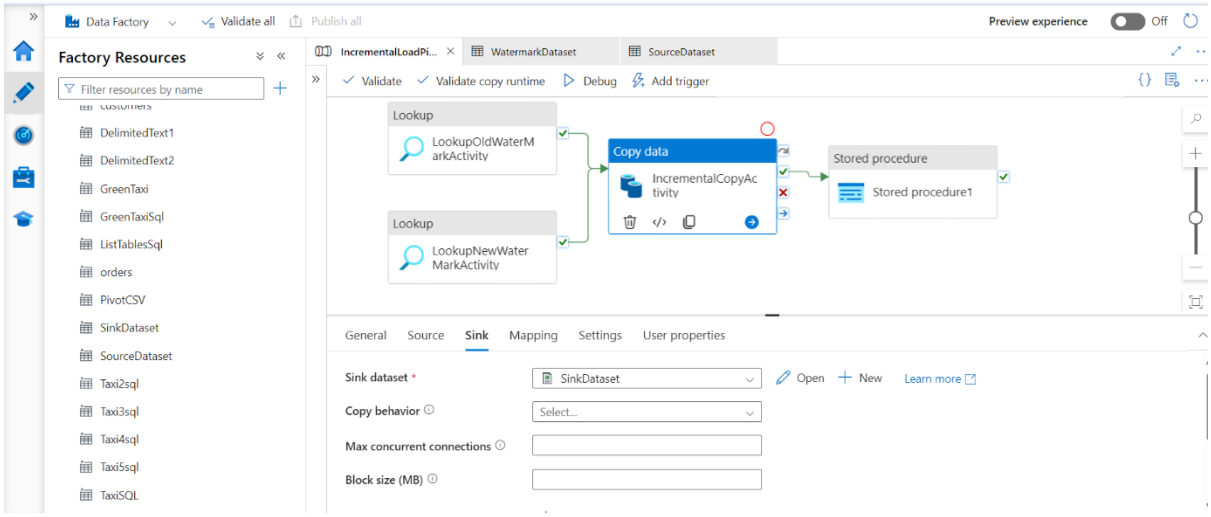

Data Factory
Validate all
Publish all
Preview experience
Off
IncrementalLoadPi...
WatermarkDataset
Azure SQL Database
WatermarkDataset
Connection
Schema
Parameters
Expand resources pane
Import schema
Clear
Column name
Type
TableName
varchar
WatermarkValue
datetime

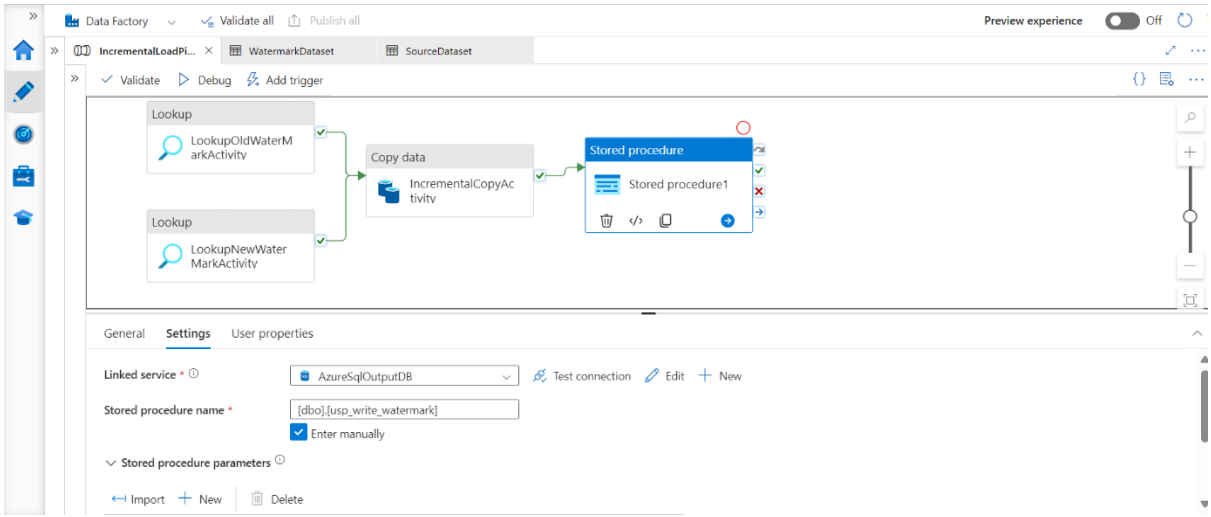

Data Factory
Validate all
Publish all
Preview experience
Off
IncrementalLoadPi...
WatermarkDataset
Validate
Debug
Add trigger
Lookup
LookupOldWaterMarkActivity
Lookup
LookupNewWaterMarkActivity
Copy data
IncrementalCopyActivity
Stored procedure
Stored procedure1
General
Settings
User properties
Source dataset *
SourceDataset
Open
New
Preview data
Learn more
First row only
Use query
Table
Query
Stored procedure
Query *
select MAX(OrderDate) as NewWatermarkvalue from data_source_table
Edit

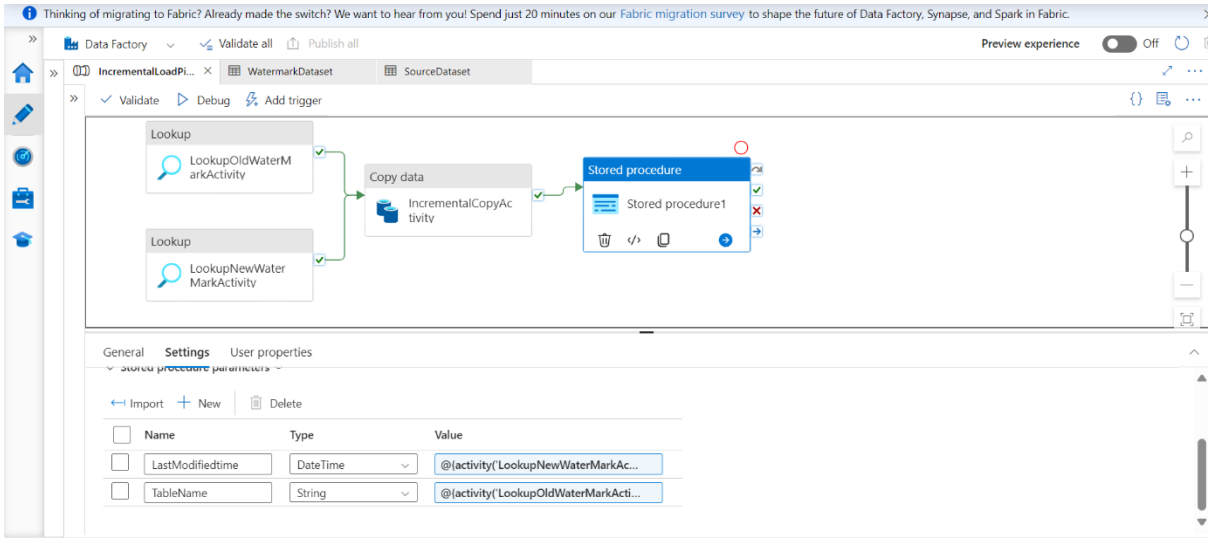

Factory Resources
Filter resources by name
DelimitedText1
DelimitedText2
GreenTaxi
GreenTaxiSql
ListTablesSql
orders
PivotCSV
SinkDataset
SourceDataset
Taxi2sql
Taxi3sql
Taxi4sql
Taxi5sql
TaxiSQL
IncrementalLoadPi...
WatermarkDataset
SourceDataset
Azure SQL Database
SourceDataset
Connection
Schema
Parameters
Linked service *
AzureSqlOutputDB
Test connection
Edit
New
Learn more
Table
dbo.data_source_table
Refresh
Preview data
Enter manually

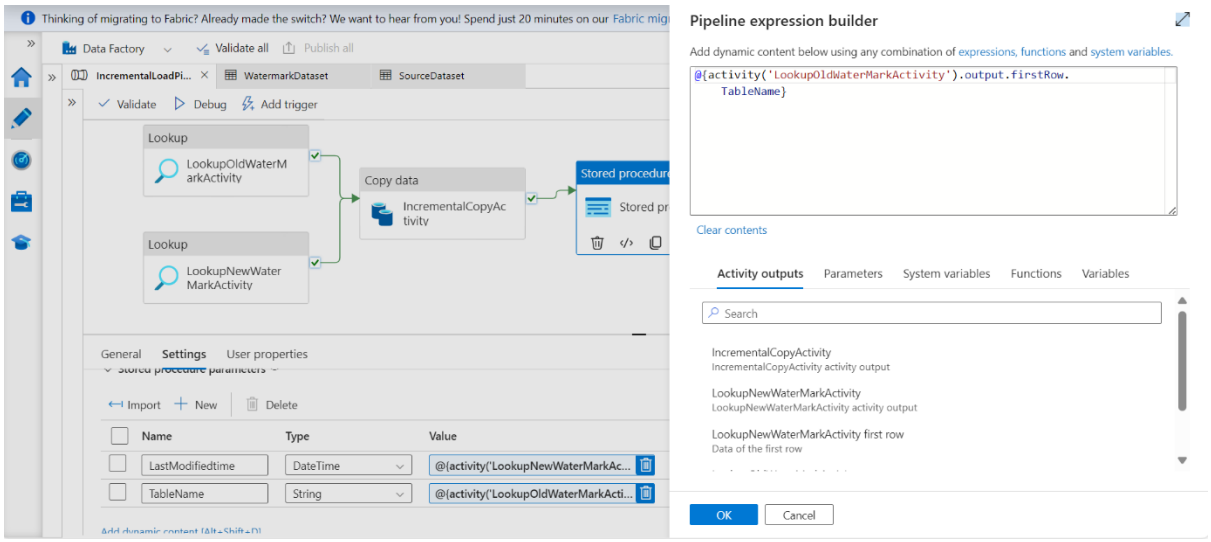

Factory Resources
Filter resources by name
DelimitedText1
DelimitedText2
GreenTaxi
GreenTaxiSql
ListTablesSql
orders
PivotCSV
SinkDataset
SourceDataset
Taxi2sql
Taxi3sql
Taxi4sql
Taxi5sql
TaxiSQL
IncrementalLoadPi...
WatermarkDataset
SourceDataset
Validate
Validate copy runtime
Debug
Add trigger
Lookup
LookupOldWaterMarkActivity
Copy data
IncrementalCopyActivity
Stored procedure
Stored procedure1
Lookup
LookupNewWaterMarkActivity
General
Source
Sink
Mapping
Settings
User properties
Source dataset *
SourceDataset
Open
New
Preview data
Learn more
Use query
Table
Query
Stored procedure
Query
select * from data_source_table wher...
Query timeout (minutes)
120
Isolation level

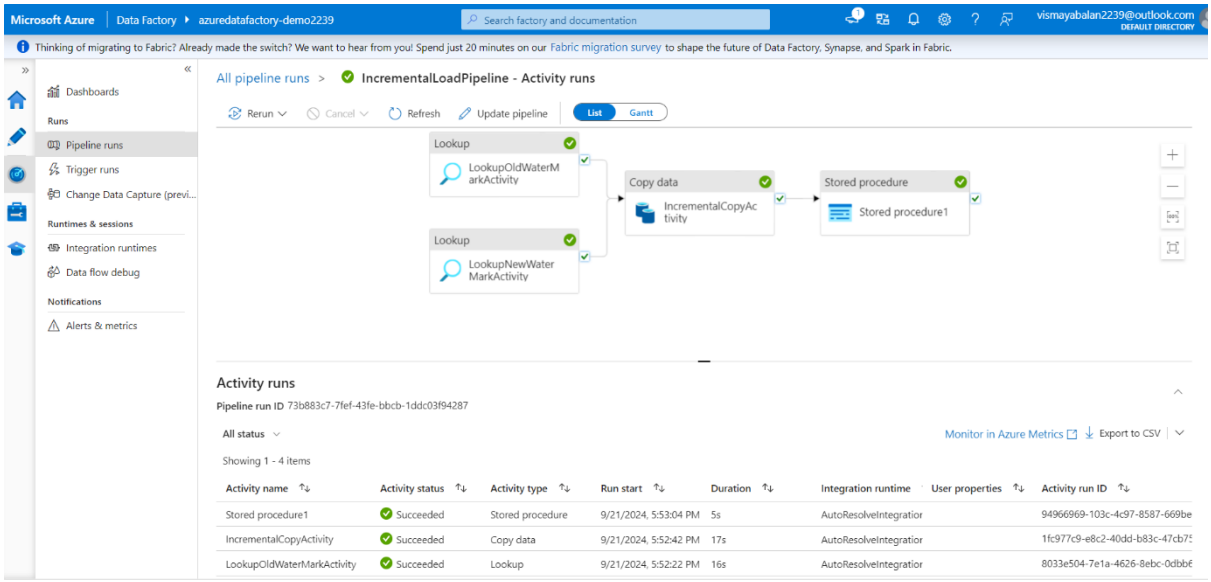

Thinking of migrating to Fabric? Already made the switch? We want to hear from you! Spend just 20 minutes on our Fabric mig
Data Factory
Validate all
Publish all
Factory Resources
Filter resources by name
DelimitedText1
DelimitedText2
GreenTaxi
GreenTaxiSql
ListTablesSql
orders
PivotCSV
SinkDataset
SourceDataset
Taxi2sql
Taxi3sql
Taxi4sql
Taxi5sql
TaxiSQL
Validate
Validate copy runtime
Debug
Lookup
LookupOldWaterMarkActivity
Copy data
Lookup
LookupNewWaterMarkActivity
General
Source
Sink
Mapping
Settings
Source dataset *
SourceDataset
Use query
Table
Query
Query
select * from data_source_ta
Query timeout (minutes)
120
Isolation level
Pipeline expression builder
Add dynamic content below using any combination of expressions, functions and system variables.
select * from data_source_table where OrderDate > '@{activity
('LookupOldWaterMarkActivity').output.firstRow.
WatermarkValue}' and OrderDate <= '@{activity
('LookupNewWaterMarkActivity').output.firstRow.
NewWatermarkvalue}'
Clear contents
Activity outputs
Parameters
System variables
Functions
Variables
Search
LookupNewWaterMarkActivity
LookupNewWaterMarkActivity activity output
LookupNewWaterMarkActivity first row
Data of the first row
LookupOldWaterMarkActivity
LookupOldWaterMarkActivity activity output
OK
Cancel


Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
DelimitedText1
DelimitedText2
GreenTaxi
GreenTaxiSql
ListTablesSql
orders
PivotCSV
SinkDataset
SourceDataset
Taxi2sql
Taxi3sql
Taxi4sql
Taxi5sql
TaxiSQL
IncrementalLoadPi...
WatermarkDataset
SourceDataset
Validate
Validate copy runtime
Debug
Add trigger
Lookup
LookupOldWaterMarkActivity
Lookup
LookupNewWaterMarkActivity
Copy data
IncrementalCopyActivity
Stored procedure
Stored procedure1
General
Source
Sink
Mapping
Settings
User properties
Sink dataset *
SinkDataset
Open
New
Learn more
Copy behavior
Select...
Max concurrent connections
Block size (MB)

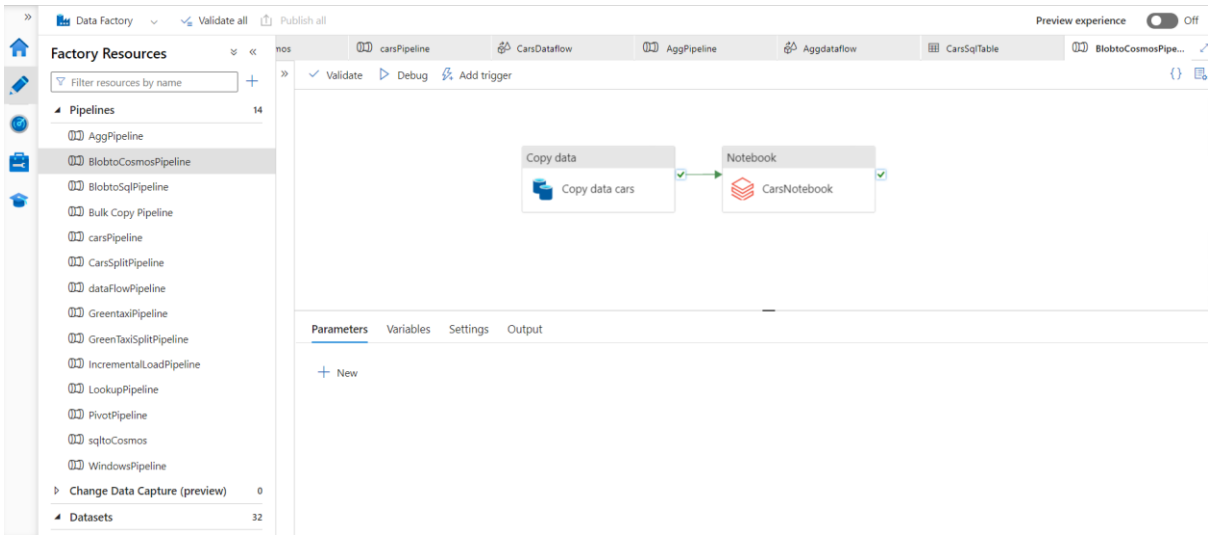

Data Factory
Validate all
Publish all
Preview experience
Off
IncrementalLoadPi...
WatermarkDataset
SourceDataset
Validate
Debug
Add trigger
Lookup
LookupOldWaterMarkActivity
Lookup
LookupNewWaterMarkActivity
Copy data
IncrementalCopyActivity
Stored procedure
Stored procedure1
General
Settings
User properties
Linked service *
AzureSqlOutputDB
Test connection
Edit
New
Stored procedure name *
[dbo].[usp_write_watermark]
Enter manually
Stored procedure parameters
Import
New
Delete

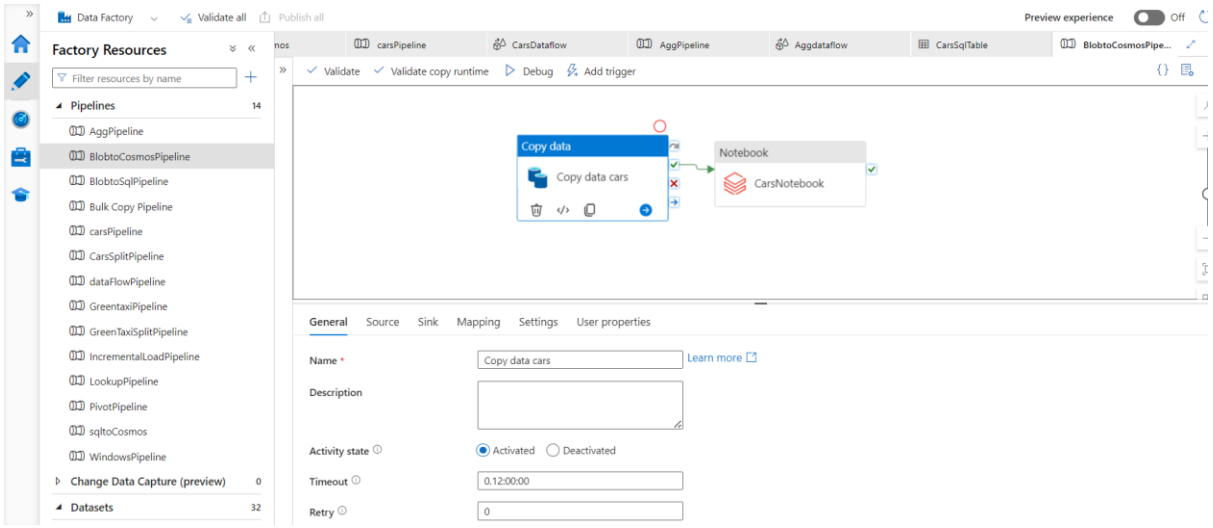

Thinking of migrating to Fabric? Already made the switch? We want to hear from you! Spend just 20 minutes on our Fabric migration survey to shape the future of Data Factory, Synapse, and Spark in Fabric.
Data Factory
Validate all
Publish all
Preview experience
Off
IncrementalLoadPi...
WatermarkDataset
SourceDataset
Validate
Debug
Add trigger
Lookup
LookupOldWaterMarkActivity
Lookup
LookupNewWaterMarkActivity
Copy data
IncrementalCopyActivity
Stored procedure
Stored procedure1
General
Settings
User properties
Import
New
Delete
Name
Type
Value
LastModifiedtime
DateTime
@{activity('LookupNewWaterMarkAc...
TableName
String
@{activity('LookupOldWaterMarkActi...

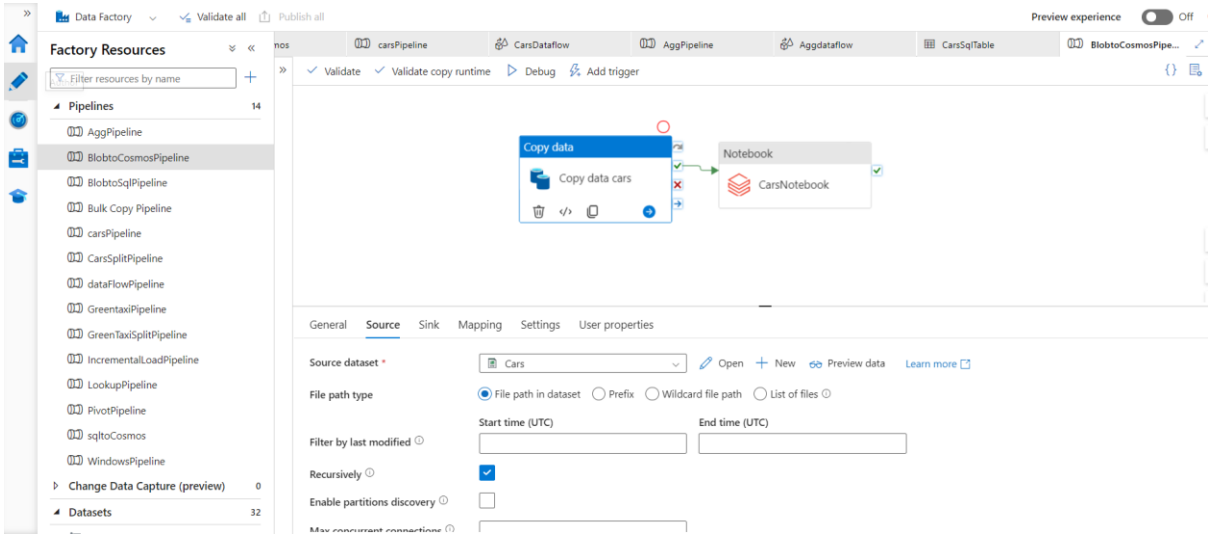

Thinking of migrating to Fabric? Already made the switch? We want to hear from you! Spend just 20 minutes on our Fabric mig
Data Factory
Validate all
Publish all
IncrementalLoadPi...
WatermarkDataset
SourceDataset
Validate
Debug
Add trigger
Lookup
LookupOldWaterMarkActivity
Lookup
LookupNewWaterMarkActivity
Copy data
IncrementalCopyActivity
Stored procedure
General
Settings
User properties
Import
New
Delete
Name
Type
Value
LastModifiedtime
DateTime
@{activity('LookupNewWaterMarkAc...
TableName
String
@{activity('LookupOldWaterMarkActi...
Pipeline expression builder
Add dynamic content below using any combination of expressions, functions and system variables.
@{activity('LookupOldWaterMarkActivity').output.firstRow.
TableName}
Clear contents
Activity outputs
Parameters
System variables
Functions
Variables
Search
IncrementalCopyActivity
IncrementalCopyActivity activity output
LookupNewWaterMarkActivity
LookupNewWaterMarkActivity activity output
LookupNewWaterMarkActivity first row
Data of the first row
OK
Cancel

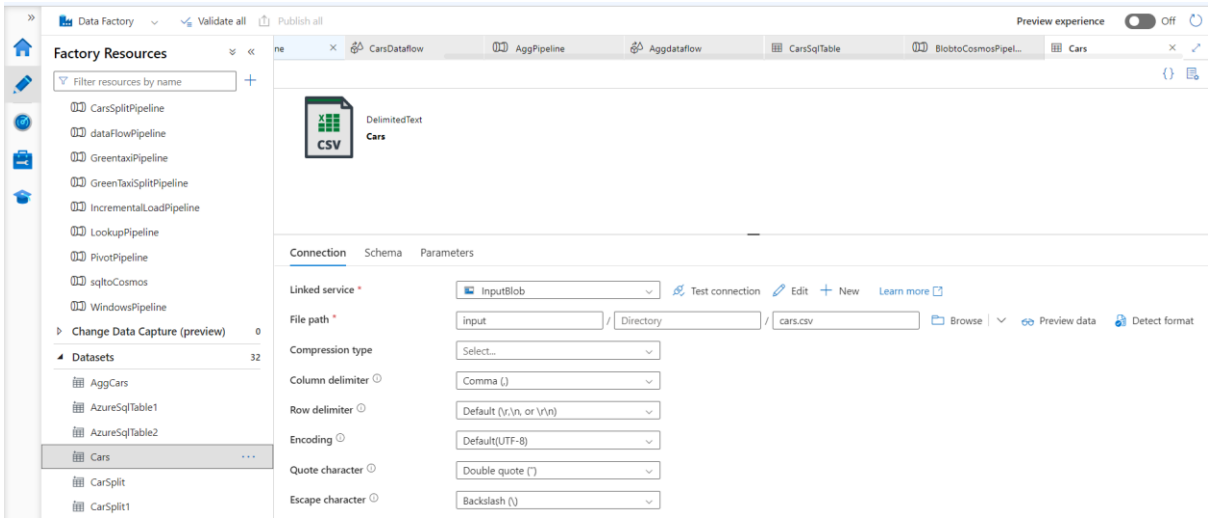


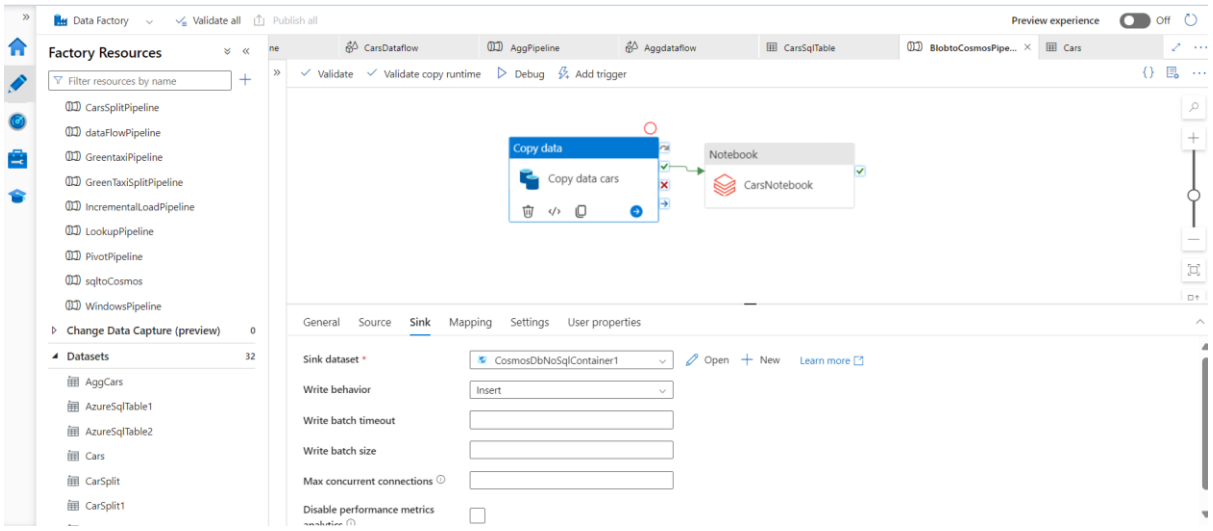
6. What are the key steps to connect Azure Databricks to Cosmos DB for real-time analytics and data transformation using spark and Databricks.

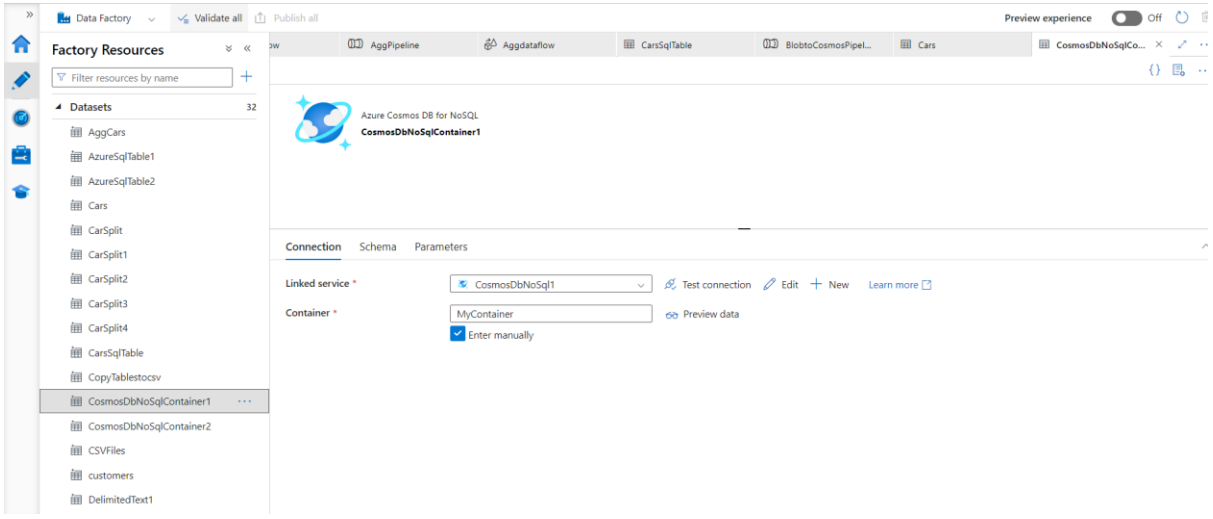


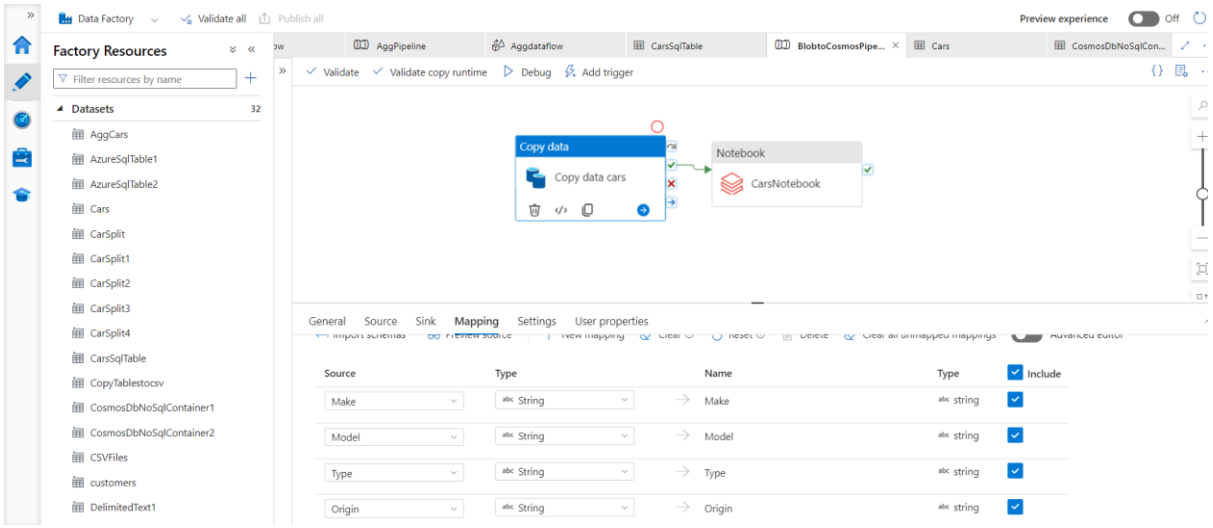

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 14
AggPipeline
BlobtoCosmosPipeline
BlobtoSqlPipeline
Bulk Copy Pipeline
carsPipeline
CarsSplitPipeline
dataFlowPipeline
GreentaxiPipeline
GreenTaxiSplitPipeline
IncrementalLoadPipeline
LookupPipeline
PivotPipeline
sqltoCosmos
WindowsPipeline
Change Data Capture (preview) 0
Datasets 32
carsPipeline
CarsDataflow
AggPipeline
Aggdataflow
CarsSqlTable
BlobtoCosmosPipe...
Validate
Validate copy runtime
Debug
Add trigger
Copy data
Copy data cars
Notebook
CarsNotebook
General
Source
Sink
Mapping
Settings
User properties
Name *
Copy data cars
Learn more
Description
Activity state
Activated
Deactivated
Timeout
0.12:00:00
Retry
0

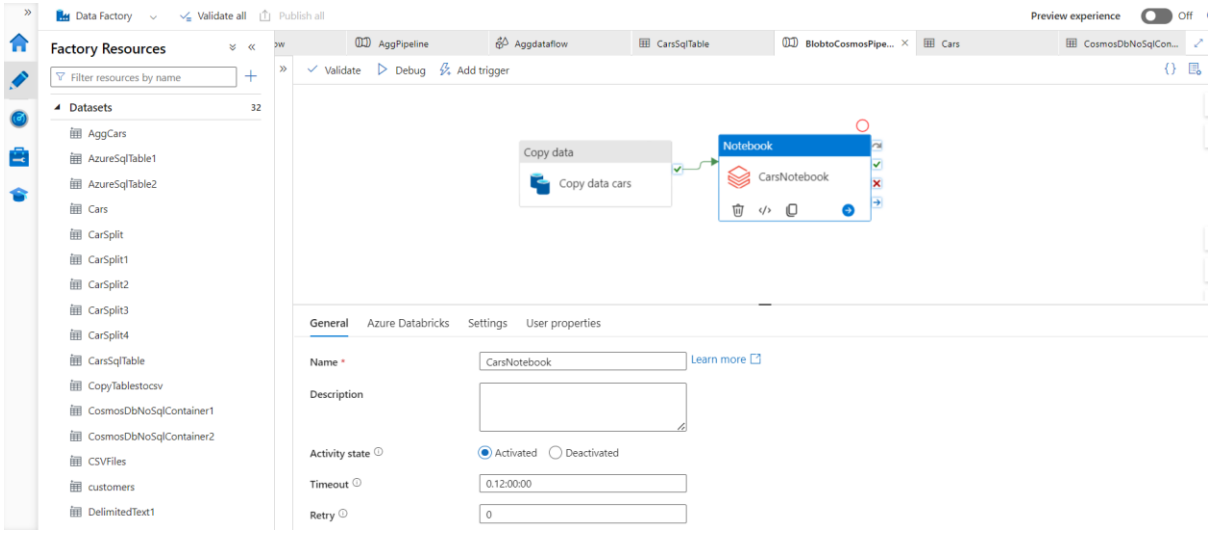

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 14
AggPipeline
BlobtoCosmosPipeline
BlobtoSqlPipeline
Bulk Copy Pipeline
carsPipeline
CarsSplitPipeline
dataFlowPipeline
GreentaxiPipeline
GreenTaxiSplitPipeline
IncrementalLoadPipeline
LookupPipeline
PivotPipeline
sqltoCosmos
WindowsPipeline
Change Data Capture (preview) 0
Datasets 32
Validate
Validate copy runtime
Debug
Add trigger
Copy data
Copy data cars
Notebook
CarsNotebook
General
Source
Sink
Mapping
Settings
User properties
Source dataset *
Cars
Open
New
Preview data
Learn more
File path type
File path in dataset
Prefix
Wildcard file path
List of files
Filter by last modified
Start time (UTC)
End time (UTC)
Recursively
Enable partitions discovery

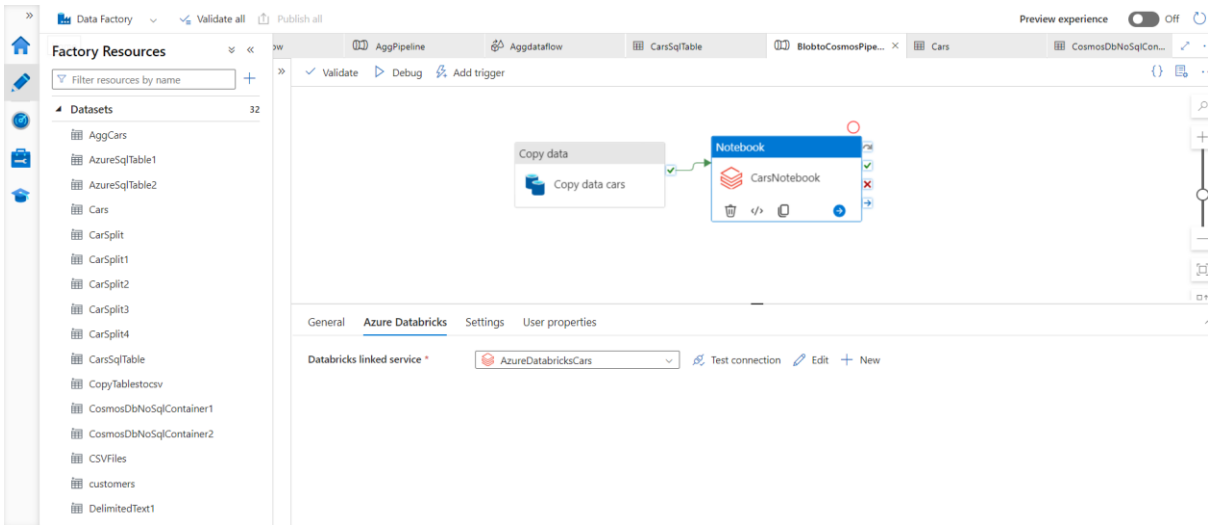

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
CarSplitPipeline
dataFlowPipeline
GreentaxiPipeline
GreenTaxiSplitPipeline
IncrementalLoadPipeline
LookupPipeline
PivotPipeline
sqltoCosmos
WindowsPipeline
Change Data Capture (preview) 0
Datasets 32
AggCars
AzureSqlTable1
AzureSqlTable2
Cars
CarSplit
CarSplit1
CarsDataflow
AggPipeline
Aggdataflow
CarsSqlTable
BlobtoCosmosPipel...
Cars
DelimitedText
Cars
Connection
Schema
Parameters
Linked service *
InputBlob
Test connection
Edit
New
Learn more
File path *
input
Directory
cars.csv
Browse
Preview data
Detect format
Compression type
Select...
Column delimiter
Comma (,)
Row delimiter
Default (\r,\n, or \r\n)
Encoding
Default(UTF-8)
Quote character
Double quote (")
Escape character
Backslash (\)

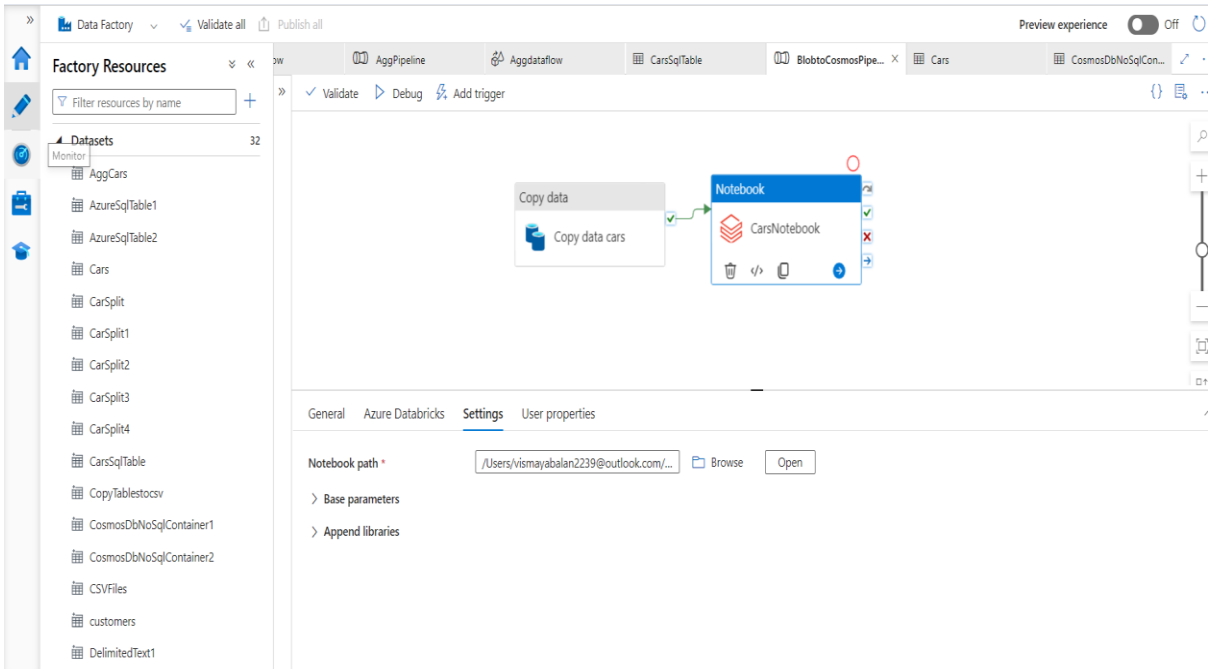

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
CarsSplitPipeline
dataFlowPipeline
GreentaxiPipeline
GreenTaxiSplitPipeline
IncrementalLoadPipeline
LookupPipeline
PivotPipeline
sqltoCosmos
WindowsPipeline
Change Data Capture (preview) 0
Datasets 32
AggCars
AzureSqlTable1
AzureSqlTable2
Cars
CarSplit
CarSplit1
CarsDataflow
AggPipeline
Aggdataflow
CarsSqlTable
BlobtoCosmosPipe...
Cars
Validate
Validate copy runtime
Debug
Add trigger
Copy data
Copy data cars
Notebook
CarsNotebook
General
Source
Sink
Mapping
Settings
User properties
Sink dataset *
CosmosDbNoSqlContainer1
Open
New
Learn more
Write behavior
Insert
Write batch timeout
Write batch size
Max concurrent connections
Disable performance metrics

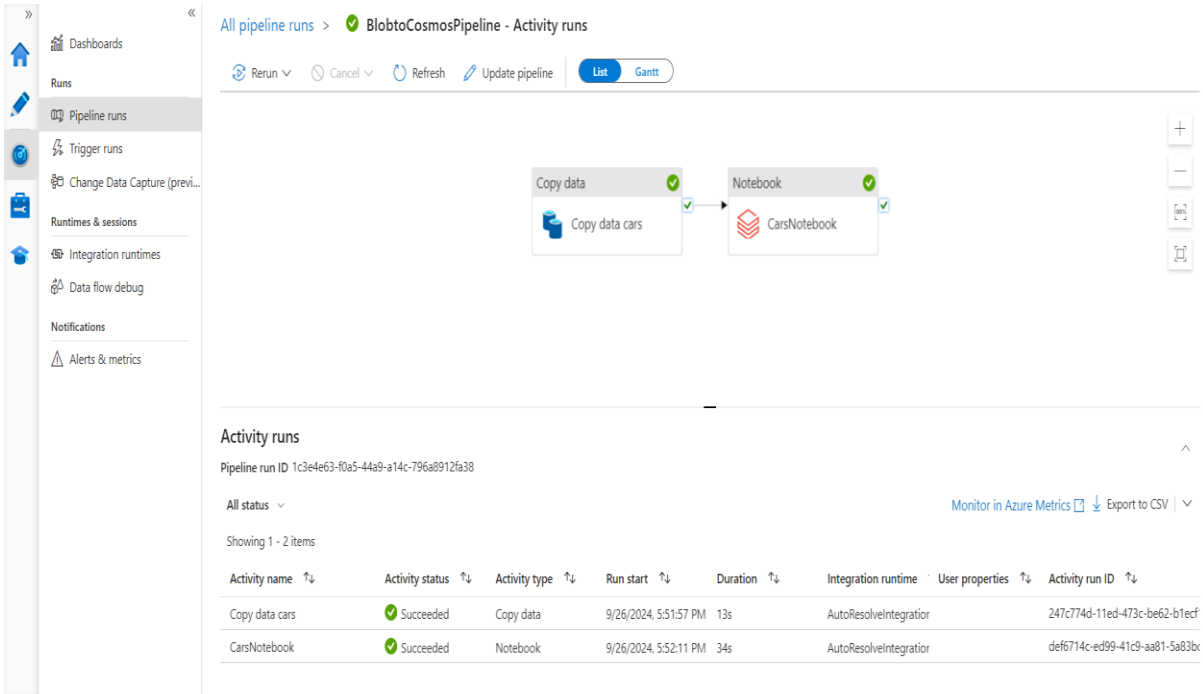

Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Datasets 32
AggCars
AzureSqlTable1
AzureSqlTable2
Cars
CarSplit
CarSplit1
CarSplit2
CarSplit3
CarSplit4
CarsSqlTable
CopyTablestocsv
CosmosDbNoSqlContainer1
CosmosDbNoSqlContainer2
CSVFiles
customers
DelimitedText1
AggPipeline
Aggdataflow
CarsSqlTable
BlobtoCosmosPipel...
Cars
CosmosDbNoSqlCo...
Azure Cosmos DB for NoSQL
CosmosDbNoSqlContainer1
Connection
Schema
Parameters
Linked service *
CosmosDbNoSql1
Test connection
Edit
New
Learn more
Container *
MyContainer
Preview data
Enter manually


Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Datasets 32
AggCars
AzureSqlTable1
AzureSqlTable2
Cars
CarSplit
CarSplit1
CarSplit2
CarSplit3
CarSplit4
CarsSqlTable
CopyTablestocsv
CosmosDbNoSqlContainer1
CosmosDbNoSqlContainer2
CSVFiles
customers
DelimitedText1
AggPipeline
Aggdataflow
CarsSqlTable
BlobtoCosmosPipe...
Cars
CosmosDbNoSqlCon...
Validate
Validate copy runtime
Debug
Add trigger
Copy data
Copy data cars
Notebook
CarsNotebook
General
Source
Sink
Mapping
Settings
User properties
Source
Type
Name
Type
Include
Make
String
Make
string
Model
String
Model
string
Type
String
Type
string
Origin
String
Origin
string


Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Datasets 32
AggCars
AzureSqlTable1
AzureSqlTable2
Cars
CarSplit
CarSplit1
CarSplit2
CarSplit3
CarSplit4
CarsSqlTable
CopyTablestocsv
CosmosDbNoSqlContainer1
CosmosDbNoSqlContainer2
CSVFiles
customers
DelimitedText1
AggPipeline
Aggdataflow
CarsSqlTable
BlobtoCosmosPipe...
Cars
CosmosDbNoSqlCon...
Validate
Debug
Add trigger
Copy data
Copy data cars
Notebook
CarsNotebook
General
Azure Databricks
Settings
User properties
Name *
CarsNotebook
Learn more
Description
Activity state
Activated
Deactivated
Timeout
0.12:00:00
Retry
0


Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Datasets 32
AggCars
AzureSqlTable1
AzureSqlTable2
Cars
CarSplit
CarSplit1
CarSplit2
CarSplit3
CarSplit4
CarsSqlTable
CopyTablestocsv
CosmosDbNoSqlContainer1
CosmosDbNoSqlContainer2
CSVFiles
customers
DelimitedText1
AggPipeline
Aggdataflow
CarsSqlTable
BlobtoCosmosPipe...
Cars
CosmosDbNoSqlCon...
Validate
Debug
Add trigger
Copy data
Copy data cars
Notebook
CarsNotebook
General
Azure Databricks
Settings
User properties
Databricks linked service *
AzureDatabricksCars
Test connection
Edit
New


Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Datasets
32
Monitor
AggCars
AzureSqlTable1
AzureSqlTable2
Cars
CarSplit
CarSplit1
CarSplit2
CarSplit3
CarSplit4
CarsSqlTable
CopyTablestocsv
CosmosDbNoSqlContainer1
CosmosDbNoSqlContainer2
CSVFiles
customers
DelimitedText1
AggPipeline
Aggdataflow
CarsSqlTable
BlobtoCosmosPipe...
Cars
CosmosDbNoSqlCon...
Validate
Debug
Add trigger
Copy data
Copy data cars
Notebook
CarsNotebook
General
Azure Databricks
Settings
User properties
Notebook path *
/Users/vismayabalan2239@outlook.com/...
Browse
Open
Base parameters
Append libraries


Dashboards
Runs
Pipeline runs
Trigger runs
Change Data Capture (previ...
Runtimes & sessions
Integration runtimes
Data flow debug
Notifications
Alerts & metrics
All pipeline runs
BlobtoCosmosPipeline - Activity runs
Rerun
Cancel
Refresh
Update pipeline
List
Gantt
Copy data
Copy data cars
Notebook
CarsNotebook
Activity runs
Pipeline run ID 1c3e4e63-f0a5-44a9-a14c-796a8912fa38
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
Copy data cars
Succeeded
Copy data
9/26/2024, 5:51:57 PM
13s
AutoResolveIntegratior
247c774d-11ed-473c-be62-b1ecf
CarsNotebook
Succeeded
Notebook
9/26/2024, 5:52:11 PM
34s
AutoResolveIntegratior
def6714c-ed99-41c9-aa81-5a83bc