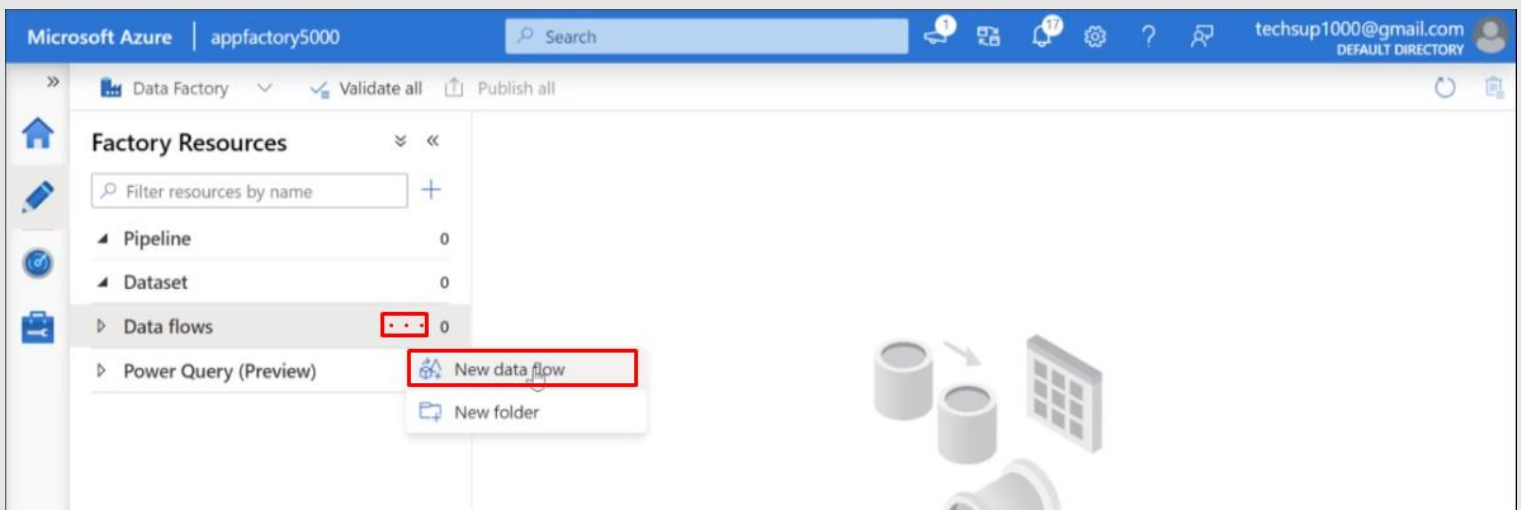


Azure Data Factory Mapping Data Flow para mover datos desde Azure SQL Database a un Dedicated SQL Pool

Veamos cómo crear el **Mapping data flow** para mover los datos desde Azure SQL Database a Azure Synapse para construir nuestra tabla fact de ventas.



Se busca generar el movimiento de datos a la tabla SalesFact que se encuentra en un Dedicated SQL Pool por medio de Mapping Data Flows. Se recreará esta Query utilizando un Data Flow.

```
4 CREATE VIEW [Sales_Fact_View]
5 AS
6 SELECT dt.[ProductID],dt.[SalesOrderID],dt.[OrderQty],dt.[UnitPrice]
7 FROM [Sales].[SalesOrderDetail] dt
8 LEFT JOIN [Sales].[SalesOrderHeader] hd
9 ON dt.[SalesOrderID]=hd.[SalesOrderID]
10
11 SELECT * FROM [Sales_Fact_View]
12
13 -- Then we will create the Sales Fact table from the view
14
15 SELECT [ProductID],[SalesOrderID],[CustomerID],[OrderQty],[UnitPrice]
16 INTO SalesFact
17 FROM Sales_Fact_View
18
19 SELECT * FROM SalesFact
```

Damos click al **Source** y le damos un nombre. Ahora indicamos el Dataset de la fuente.

The screenshot shows the Microsoft Azure Data Factory interface. The top navigation bar includes the Microsoft Azure logo, the workspace name 'appfactory5000', a search bar, and user information 'techsup1000@gmail.com'. The main workspace area displays a data flow named 'SalesFactDataFlow'. The 'Source settings' tab is selected, showing the 'Output stream name' as 'SalesOrderDetailStream' and the 'Source type' as 'Dataset'. The 'Dataset' dropdown is set to 'Select...', and a red box highlights the '+ New' button next to it. The 'Options' section shows 'Allow schema drift' checked, and the 'Sampling' section shows 'Disable' selected. The 'Properties' pane on the right shows the 'General' tab with the 'Name' set to 'SalesFactDataFlow' and a 'Description' field.

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory > SalesFactDataFlow

Validate ☒ Data flow debug ☐

SalesOrderDetailStream
Columns: 0 total

Source settings Source options Projection Optimize

Output stream name * SalesOrderDetailStream

Source type * ☒ Dataset ☐ Inline

Dataset * Select...

Options ☒ Allow schema drift ⓘ
☐ Infer drifted column types ⓘ
☐ Validate schema ⓘ

Sampling * ⓘ ☐ Enable ☒ Disable

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

SQL
Azure SQL Database

Azure SQL Database Managed Instance

Azure Synapse Analytics

SQL
SQL server

Snowflake

Amazon Marketplace Web Service

Continue Cancel

Source dataset: SalesOrderDetailDt

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory > SalesFactDataFlow

Validate ☒ Data flow debug ☐

SalesOrderDetailStream
Columns: 0 total

Source settings Source options Projection Optimize

Output stream name * SalesOrderDetailStream

Source type * ☒ Dataset ☐ Inline

Dataset * Select...

Options ☒ Allow schema drift ⓘ
☐ Infer drifted column types ⓘ
☐ Validate schema ⓘ

Sampling * ⓘ ☐ Enable ☒ Disable

Set properties

Name
SalesOrderDetailDt

Linked service *
AzureSqlDatabase

Table name
Sales.SalesOrderDetail

☐ Edit

Import schema
☒ From connection/store ☐ None

Advanced

OK Back Cancel

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all (2)

SalesFactDataFlow

✓ Validate | Data flow debug

SalesOrderDetailStre...
Columns: 11 total

Source settings | **Source options** | Projection | Optimize | Inspect | Data preview | Description

Input: ☒ Table ☐ Query ☐ Stored procedure

Batch size:

Isolation level: Read uncommitted

Properties
General | Related
Name: SalesFactDataFlow
Description:

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all (2)

SalesFactDataFlow

✓ Validate | Data flow debug

SalesOrderDetailStre...
Columns: 11 total

Source settings | Source options | **Projection** | Optimize | Inspect | Data preview | Description

Import projection | Reset schema

Column name	Type
SalesOrderID	integer
SalesOrderDetailID	integer
CarrierTrackingNumber	string
OrderQty	integer
ProductID	integer
SpecialOfferID	integer

Creamos otra fuente

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all (2)

SalesFactDataFlow

Validate | Data flow debug

SalesOrderDetailStre...

Import data from SalesOrderDetailDt

SalesOrderHeaderStream

Columns: 0 total

Source settings | Source options | Projection | Optimize | Inspect | Data preview | Description

Output stream name * SalesOrderHeaderStream [Learn more](#)

Source type * Dataset Inline

Dataset * Select... [+ New](#)

Options ☒ Allow schema drift

Properties

General Related

Name * SalesFactDataFlow

Description

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all (2)

SalesFactDataFlow

Validate | Data flow debug

SalesOrderDetailStre...

Import data from SalesOrderDetailDt

SalesOrderHeaderStream

Columns: 0 total

Source settings | Source options | Projection | Optimize

Output stream name * SalesOrderHeaderStream

Source type * Dataset Inline

Dataset * Select...

Options ☒ Allow schema drift

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 SQL Database

Azure SQL Database Managed Instance

Azure Synapse Analytics

SQL server

Snowflake

Amazon Marketplace Web Service

Continue

Cancel

Source dataset: SalesOrderHeaderDt

This screenshot shows the 'Set properties' dialog for a dataset named 'SalesOrderHeaderDt' in an Azure Data Factory environment. The dialog is part of a data flow named 'SalesFactDataFlow'. The 'Name' field is set to 'SalesOrderHeaderDt'. The 'Linked service' is 'AzureSqlDatabase'. The 'Table name' is 'Sales.SalesOrderHeader'. The 'Import schema' option is set to 'From connection/store'. The 'Advanced' section is expanded, showing 'Output stream name' as 'SalesOrderHeaderStream', 'Source type' as 'Dataset', and 'Dataset' as 'Select...'. The 'Options' section has 'Allow schema drift' checked. The 'OK' button is highlighted with a mouse cursor.

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

» Data Factory | Validate all | Publish all (2)

» SalesFactDataFlow

✓ Validate Data flow debug

SalesOrderDetailStre...
Import data from SalesOrderDetailDt

SalesOrderHeaderStream
Columns: 0 total

Source settings Source options Projection Optimize

Output stream name * SalesOrderHeaderStream

Source type * Dataset Inline

Dataset * Select...

Options ☒ Allow schema drift ⓘ

Set properties

Name SalesOrderHeaderDt

Linked service * AzureSqlDatabase

Table name Sales.SalesOrderHeader

☐ Edit

Import schema ☒ From connection/store ☐ None

Advanced

OK Back Cancel

This screenshot shows the 'Properties' pane for a data flow named 'SalesFactDataFlow' in an Azure Data Factory environment. The 'Name' field is set to 'SalesFactDataFlow'. The 'Description' field is empty. The 'Source settings' tab is selected, showing 'Output stream name' as 'SalesOrderHeaderStream', 'Source type' as 'Dataset', and 'Dataset' as 'SalesOrderDetailDt'. A context menu is open over the 'SalesOrderHeaderStream' dataset, showing options like 'Join', 'Conditional Split', 'Exists', 'Union', 'Lookup', 'Schema modifier', 'Derived Column', and 'Select'. The 'Join' option is highlighted with a red box.

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

» Data Factory | Validate all | Publish all (3)

» SalesFactDataFlow

✓ Validate Data flow debug

SalesOrderDetailStre...
Columns: 11 total

SalesOrderHeaderStr...
Import data from SalesOrderHeaderDt

Source settings Source options Projection Optimize

Output stream name * SalesOrderHeaderStream

Source type * Dataset Inline

Dataset * SalesOrderDetailDt

Test connection Open New

Properties

General Related

Name * SalesFactDataFlow

Description

Join

Conditional Split

Exists

Union

Lookup

Schema modifier

Derived Column

Select

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all (3)

SalesFactDataFlow

Validate | Data flow debug

SalesOrderDetailStre...
Import data from SalesOrderDetailDt

SalesJoin
Columns: 37 total

SalesOrderHeaderStr...
Import data from SalesOrderHeaderDt

Join settings | Optimize | Inspect | Data preview | Description

Output stream name *
SalesJoin

Left stream *
SalesOrderDetailStream

Right stream *
SalesOrderHeaderStream

Join type *
Full outer | Inner | Left outer | Right outer | Custom (cross)

Properties

General | Related

Name *
SalesFactDataFlow

Description

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all (3)

SalesFactDataFlow

Validate | Data flow debug

Join settings | Optimize | Inspect | Data preview | Previous | Next

Output stream name *
SalesJoin

Left stream *
SalesOrderDetailStream

Right stream *
SalesOrderHeaderStream

Join type *
Full outer | Inner | Left outer | Right outer | Custom (cross)

Conditions *

Left: SalesOrderDetailStream's column
123 SalesOrderID

Right: SalesOrderHeaderStream
123 SalesOrderID

Properties

General | Related

Name *
SalesFactDataFlow

Description

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

>> Data Factory > Validate all > Publish all 3

>> SalesFactDataFlow

✓ Validate Data flow debug

Import data from SalesOrderDetailDt

Import data from SalesOrderHeaderDt

SalesJoin
Columns: 37 total

Rank

Formatters

Flatten

Parse

Row modifier

Filter

Sort

Alter Row

Destination

Sink

Join settings Optimize Inspect Data preview

Output stream name * SalesJoin

Incoming stream * SalesOrderDetailStream

Join stream * SalesOrderHeaderStream

Join type *

Properties

General Related

Name * SalesFactDataFlow

Description

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

>> Data Factory > Validate all > Publish all 3

>> SalesFactDataFlow

✓ Validate Data flow debug

Import data from SalesOrderDetailDt

Import data from SalesOrderHeaderDt

SalesJoin
Left outer join on 'SalesOrderDetailStream' and 'SalesOrderHeaderStream'

sink1
Columns: 34 total

Sink Settings Mapping Optimize Inspect Data preview

Output stream name * SalesFactSink

Incoming stream * SalesJoin

Sink type *

Dataset Inline Cache

Dataset * Select... + New

Options

Allow schema drift

Properties

General Related

Name * SalesFactDataFlow

Description

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all (3)

SalesFactDataFlow

Validate | Data flow debug

SalesOrderDetailStre... | SalesJoin

Import data from SalesOrderDetailDt

Left outer join on 'SalesOrderDetailStream' and 'SalesOrderHeaderStream'

SalesOrderHeaderStr...

Sink | Settings | Mapping | Optimize | Inspect | Data pre

Output stream name * | SalesFactSink

Incoming stream * | SalesJoin

Sink type * | Dataset | Inline

Dataset * | Select...

Options | ☒ Allow schema drift

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 Data Lake Storage Gen2

Azure Database for MySQL

Azure Database for PostgreSQL

Azure SQL Database

Azure SQL Database Managed Instance

Azure Synapse Analytics

Continue | Cancel

Destination dataset: SalesFactSynapse

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all (3)

SalesFactDataFlow

Validate | Data flow debug

SalesOrderDetailStre... | SalesJoin

Import data from SalesOrderDetailDt

Left outer join on 'SalesOrderDetailStream' and 'SalesOrderHeaderStream'

SalesOrderHeaderStr...

Sink | Settings | Mapping | Optimize | Inspect | Data pre

Output stream name * | SalesFactSink

Incoming stream * | SalesJoin

Sink type * | Dataset | Inline

Dataset * | Select...

Options | ☒ Allow schema drift

Set properties

Name | SalesFactSynapse

Linked service * | AzureSynapseAnalytics

☒ Select from existing table ☐ Create new table

Table name | dbo.SalesFact

☐ Edit

Import schema | ☒ From connection/store ☐ None

Advanced

OK | Back | Cancel

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com

Data Factory > Validate all > Publish all 4

SalesFactDataFlow

✓ Validate Data flow debug

Sink Settings **Mapping** Optimize Inspect Data preview Description

Options

- ✓ Skip duplicate input columns
- ✓ Skip duplicate output columns
- ☒ Auto mapping
- Reset Add mapping Delete
- Output format
- 7 mappings: All outputs mapped

Input columns	Output columns
123 ProductID	123 ProductID
123 SalesOrderDetailStream@SalesOrderID	123 SalesOrderID
123 CustomerID	123 CustomerID
123 OrderQty	123 OrderQty
e* UnitPrice	e* UnitPrice
OrderDate	OrderDate
e* TaxAmt	e* TaxAmt

Luego **Validamos y Publicamos**. ¿Pero como ejecutamos esto? Para eso crearemos un Pipeline

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com

Data Factory > Validate all > Publish all

Factory Resources

- Pipeline 0
 - New pipeline
 - Pipeline from template
 - New folder
- Dataset
 - SalesFactSynapse
 - SalesOrderDetailDt
 - SalesOrderHeaderDt
- Data flows 1
 - SalesFactDataFlow
- Power Query (Preview) 0

SalesFactDataFlow

✓ Validate Data flow debug

SalesOrderDetailStre... SalesJoin SalesFactSink

Left outer join on 'SalesOrderDetailStream' and 'SalesOrderHeaderStream'

Columns: 7 total

Import data from SalesOrderHeaderDt

Sink Settings **Mapping** Optimize Inspect Data preview Description

Options

- ✓ Skip duplicate input columns
- ✓ Skip duplicate output columns
- ☒ Auto mapping
- Reset Add mapping Delete
- Output format
- 7 mappings: All

Input columns	Output columns
123 ProductID	123 ProductID

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all 1

SalesFactDataFlow | SalesFactPipeline

Save as template | Validate | Debug | Add trigger | Data flow debug

Activities

Search activities

- Move & transform
 - Copy data
 - Data flow
- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks
- Data Lake Analytics
- General
- HDInsight
- Iteration & conditionals
- Machine Learning
- Power Query

SalesDataFlow

General Settings Parameters User properties

The default timeout for new data flow activities has recently changed from 7 days to 1 day.

Name * SalesDataFlow

Description

Timeout 1.00:00:00

Retry 0

Retry interval 30

Properties

General Related

Name * SalesFactPipeline

Description

Concurrency has moved to the [Settings tab](#).

Annotations

+ New

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all 1

SalesFactDataFlow | SalesFactPipeline

Save as template | Validate | Debug | Add trigger | Data flow debug

Activities

Search activities

- Move & transform
 - Copy data
 - Data flow
- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks
- Data Lake Analytics
- General
- HDInsight
- Iteration & conditionals
- Machine Learning
- Power Query

SalesDataFlow

General **Settings** Parameters User properties

Data flow * SalesFactDataFlow

Run on (Azure IR) * AutoResolveIntegrationRuntime

Compute type * General purpose

Core count * 4 (+ 4 Driver cores)

Logging level * Verbose Basic None

Sink properties

Properties

General Related

Name * SalesFactPipeline

Description

Concurrency has moved to the [Settings tab](#).

Annotations

+ New

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all 1

SalesFactDataFlow | SalesFactPipeline

Activities

- Move & transform
- Copy data
- Data flow
- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks
- Data Lake Analytics
- General
- HDInsight
- Iteration & conditionals
- Machine Learning
- Power Query

Save as template | Validate | Debug | Add trigger | Data flow debug

SalesDataFlow

General | Settings | Parameters | User properties

Run on (Azure IR) * | AutoResolveIntegrationRuntime

Compute type * | General purpose

Core count * | 4 (+ 4 Driver cores)

Logging level * | ☒ Verbose ☐ Basic ☐ None

Sink properties

Staging ☒

Staging linked service | AzureDataLakeStorage | Test connection | Edit | New

Staging storage folder | synapse | / | Directory | Browse

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all

SalesFactDataFlow | SalesFactPipeline

Activities

- Move & transform
- Copy data
- Data flow
- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks

Save as template | Validate | Debug | Add trigger

SalesDataFlow

General | Settings | Parameters | User properties

Run on (Azure IR) * | AutoResolveIntegrationRuntime

Compute type * | General purpose

Core count * | 4 (+ 4 Driver cores)

Publishing completed
Successfully published

Trigger now
Trigger on-demand run of the last published pipeline

New/Edit

Microsoft Azure | appfactory5000

techsup1000@gmail.com
DEFAULT DIRECTORY

Pipeline runs

Triggered Debug Rerun Cancel Refresh Edit columns List Gantt

Search by run ID or name Abu Dhabi, Muscat (... : Last 24 hours Pipeline name : All

Status : All Runs : Latest runs Add filter Show all

Showing 1 - 6 items

	Duration	Triggered by	Status	Error	Run	Parameters
:34 PM	00:04:41	Manual trigger	✓ Succeeded		Original	
8:52 PM	00:00:30	Manual trigger	✓ Succeeded		Original	
7:54 AM	00:00:09	Manual trigger	✓ Succeeded		Original	
:49 AM	00:00:25	Manual trigger	✓ Succeeded		Original	
:36 AM	00:00:28	Manual trigger	✓ Succeeded		Original	
4:47 PM	00:00:23	Manual trigger	✓ Succeeded		Original	

SQLQuery1.sql - appworkspace9000.sql.azuresynapse.net.newpool (sqladminuser (2408)) Executing... - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

newpool

SQLQuery2.sql - ap...sqladminuser (119))* SQLQuery1.sql - app...2408)) Executing... Object Explorer

```
DELETE FROM SalesFact
SELECT * FROM SalesFact
```

100 %

Results Messages

	ProductID	SalesOrderID	CustomerID	OrderQty	UnitPrice	OrderDate	TaxAmt
1	970	58915	29821	1	728.91	2013-10-30 00:00:00.000	2629.3245
2	907	55326	29648	2	63.90	2013-08-30 00:00:00.000	104.0115
3	707	60692	17669	1	34.99	2013-11-22 00:00:00.000	49.996
4	860	50672	29998	3	14.1289	2013-04-30 00:00:00.000	7597.726
5	967	65158	29736	3	1430.442	2014-01-29 00:00:00.000	6500.3898
6	707	47067	29667	1	20.1865	2012-06-30 00:00:00.000	1763.621
7	782	57118	29968	3	1376.994	2013-09-30 00:00:00.000	2925.8926
8	973	72514	24592	1	1700.99	2014-05-09 00:00:00.000	137.1976
9	998	61227	29489	8	323.994	2013-11-30 00:00:00.000	5496.2454
10	922	58890	19602	1	3.99	2013-10-29 00:00:00.000	201.1944
11	708	60745	22411	1	34.99	2013-11-23 00:00:00.000	47.1168
12	712	50682	29891	5	5.1865	2013-04-30 00:00:00.000	3724.0827
13	778	43914	29521	1	2024.994	2011-07-01 00:00:00.000	782.4537
14	716	55822	24718	1	49.99	2013-09-07 00:00:00.000	140.0784
15	954	53477	30087	1	953.628	2013-07-31 00:00:00.000	2666.592

Executing query... appworkspace9000.sql.azuresynapse.net newpool 00:00:00 0 rows