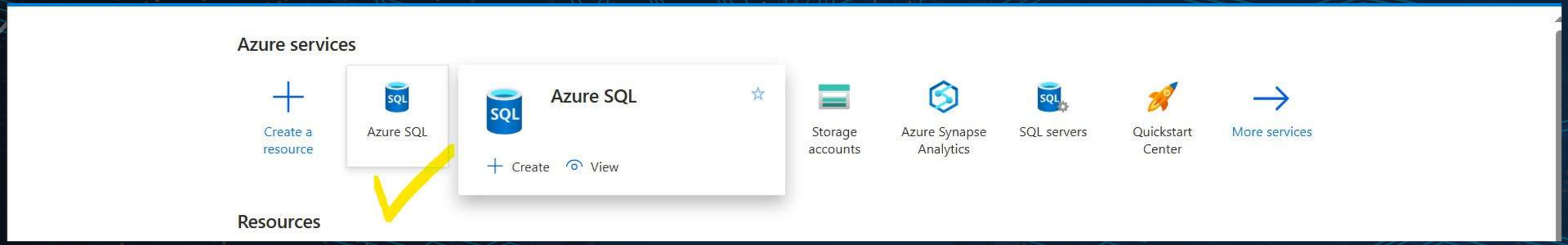
Azure Data Engineer

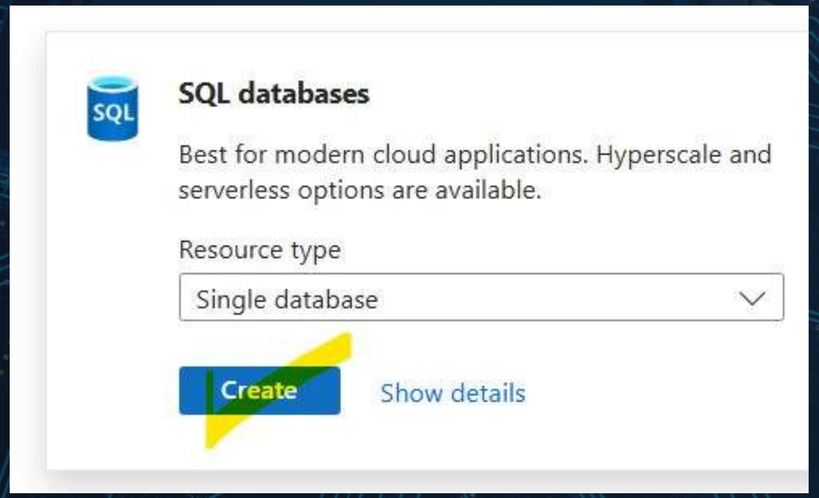
By: Abdulrahman Hussieni

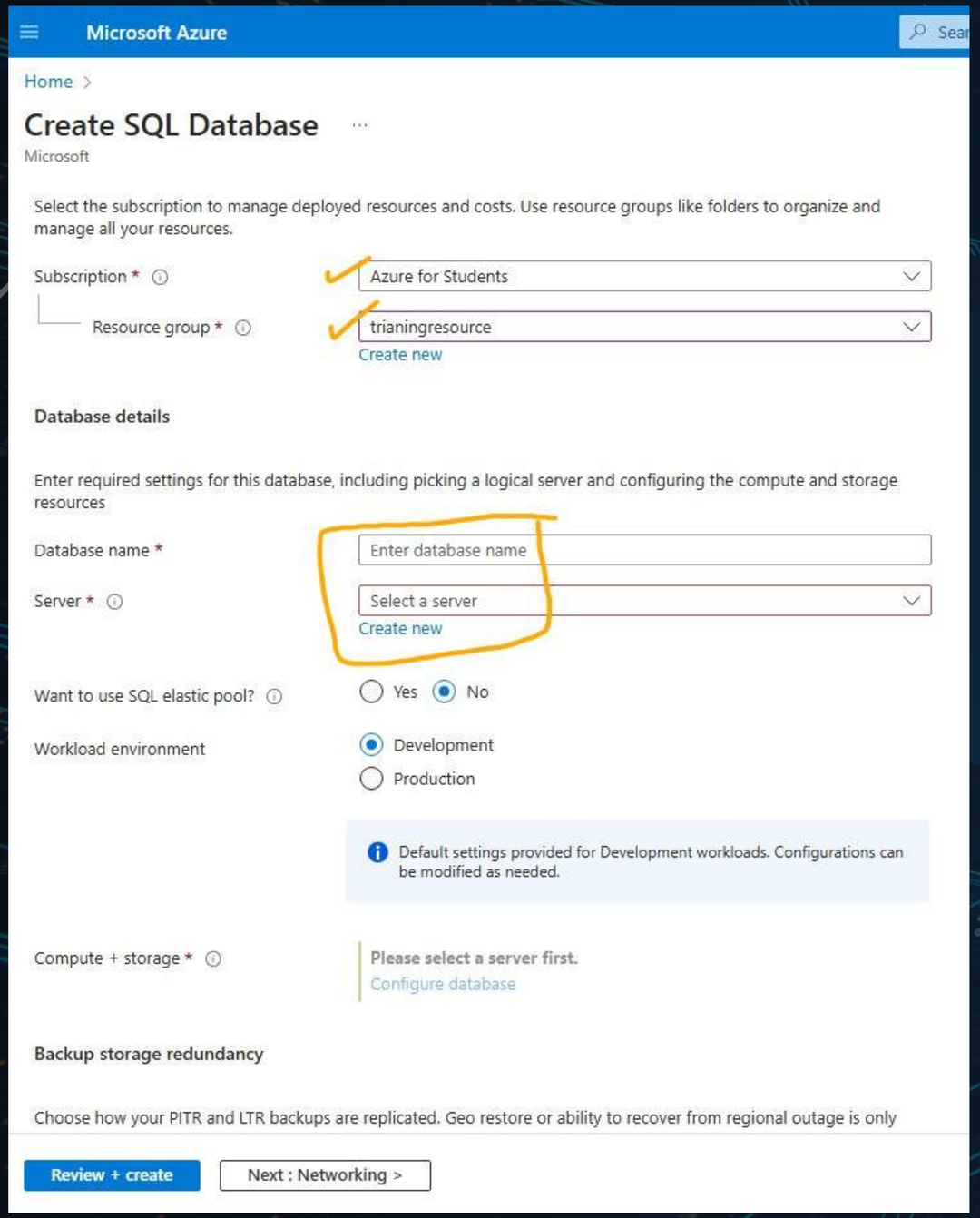
Introduction

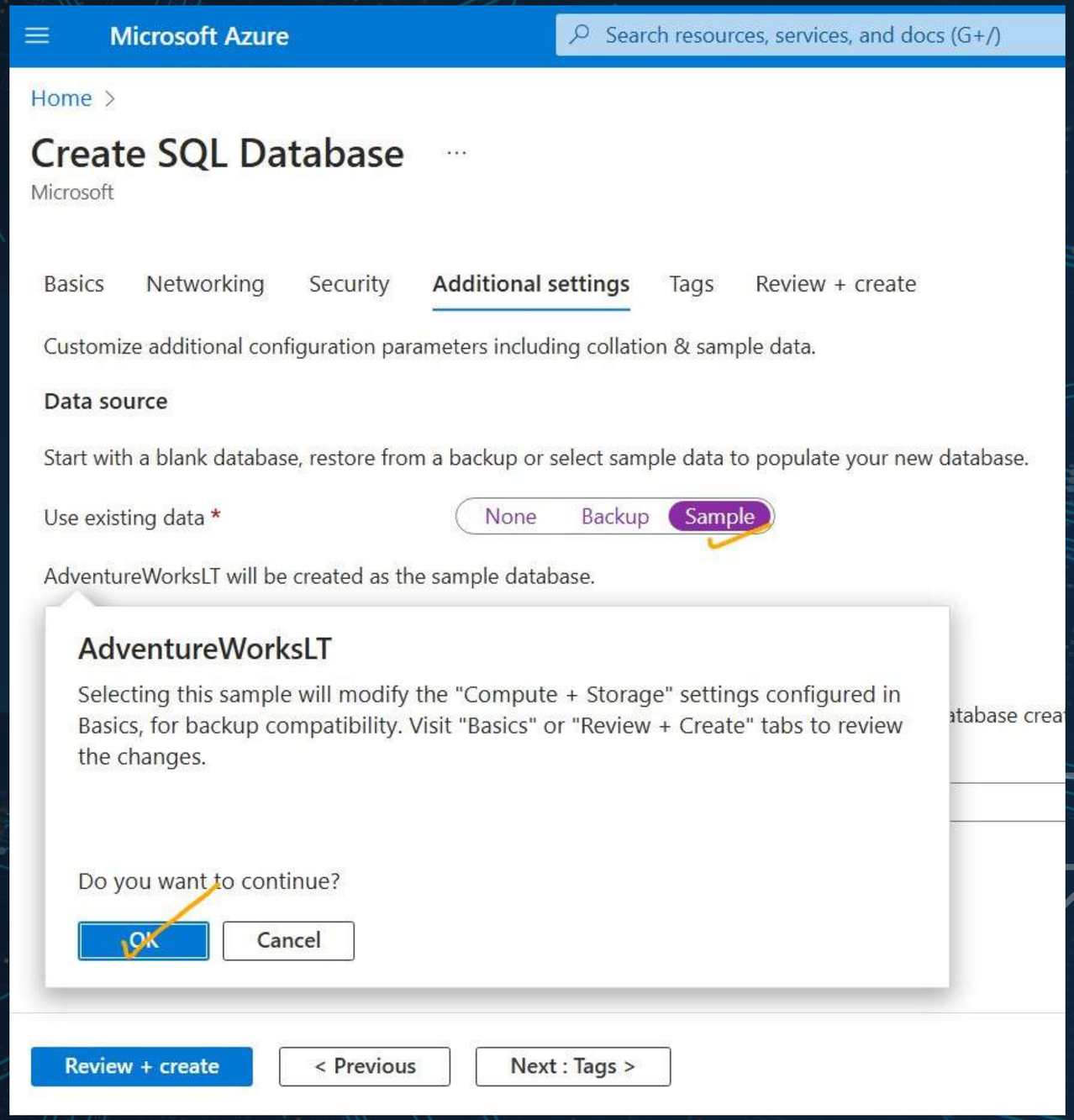
project: we have data avaliable in sql databaseand we want to transfere data in blob storage but with filter condition and data should be avaliable in sorted way

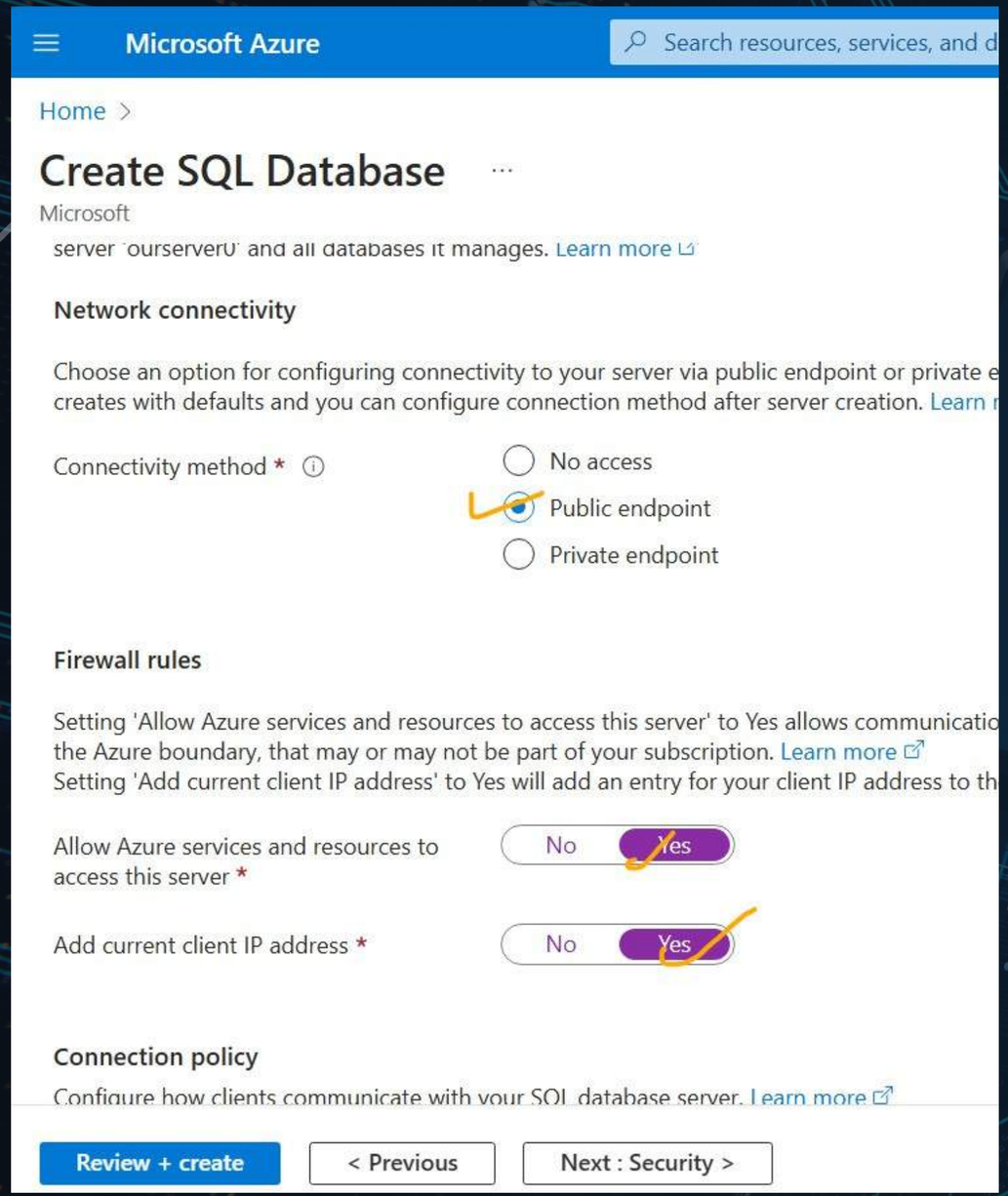
step 1: Azure sql

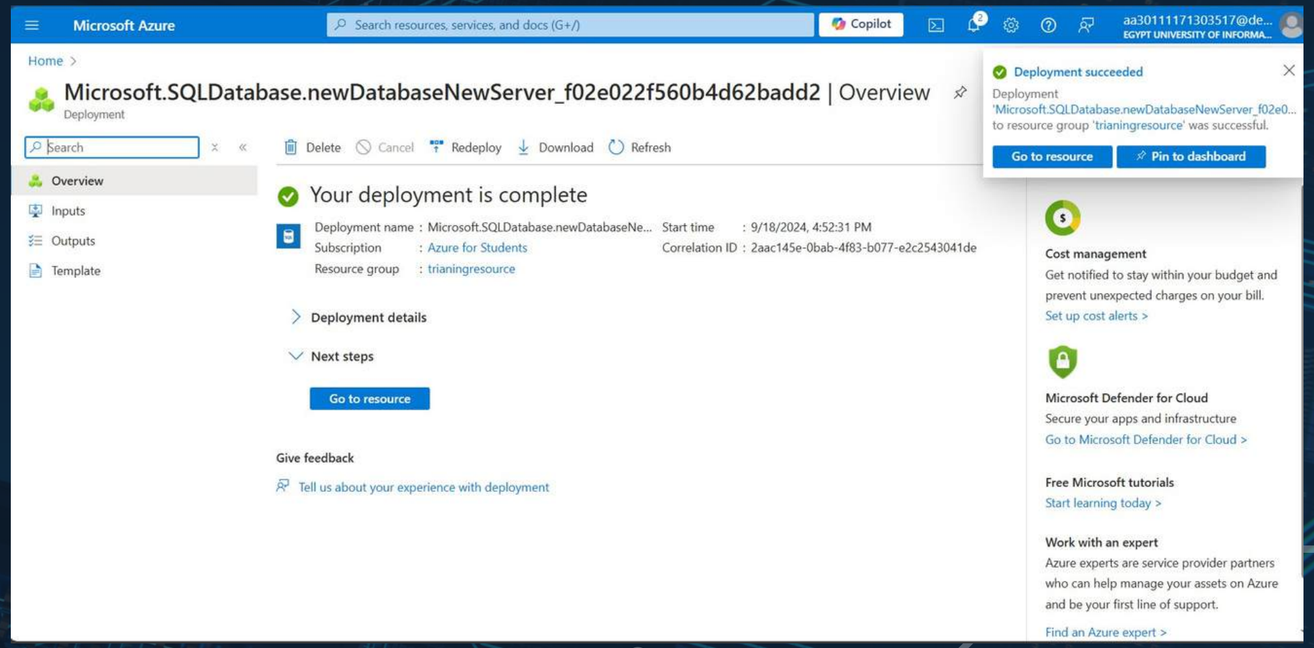


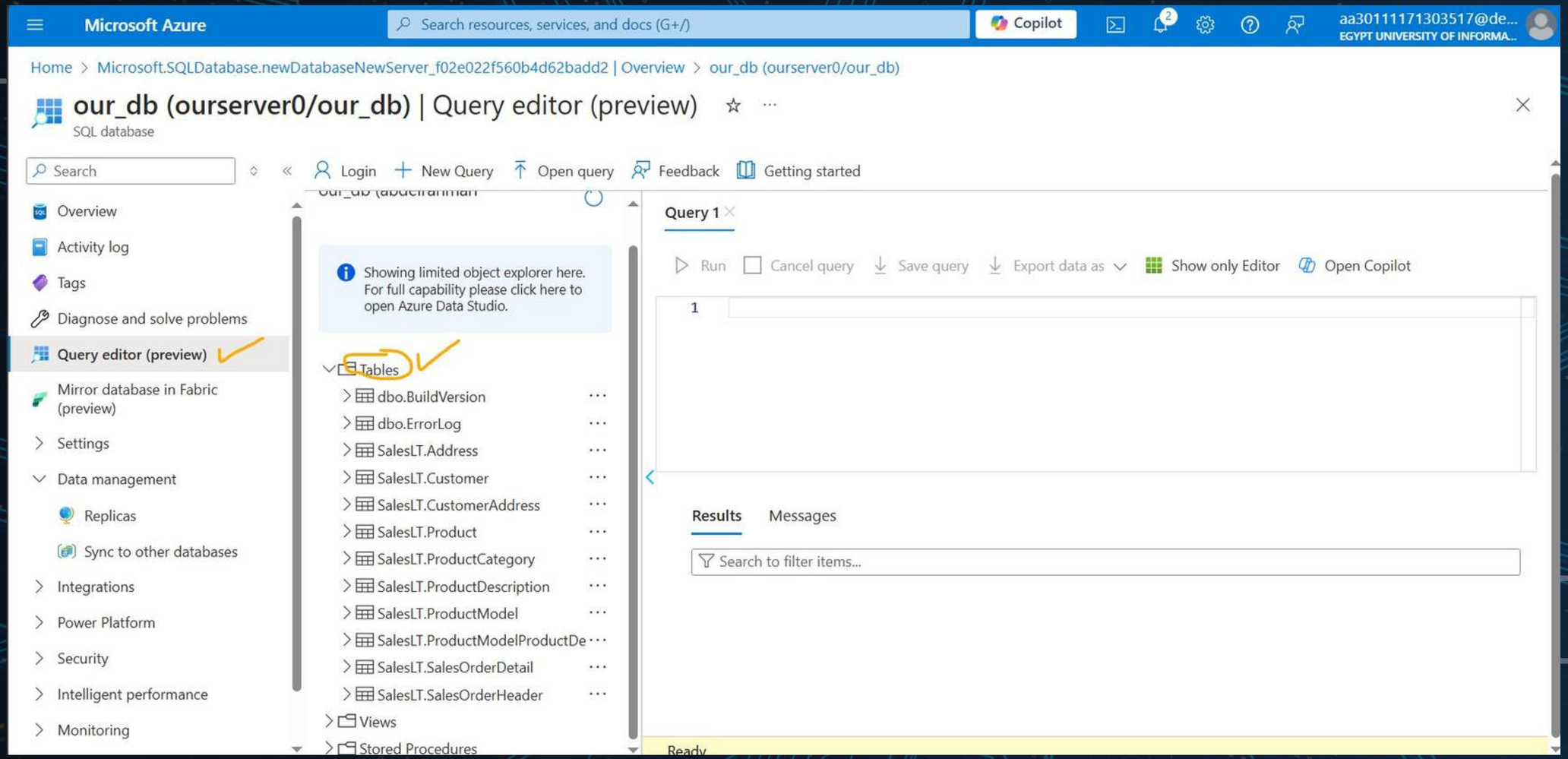



Microsoft Azure
Home
Create SQL Database
Microsoft
Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.
Subscription *
Azure for Students
Resource group *
trianingresource
Create new
Database details
Enter required settings for this database, including picking a logical server and configuring the compute and storage resources
Database name *
Enter database name
Server *
Select a server
Create new
Want to use SQL elastic pool?
Yes
No
Workload environment
Development
Production
Default settings provided for Development workloads. Configurations can be modified as needed.
Compute + storage *
Please select a server first.
Configure database
Backup storage redundancy
Choose how your PITR and LTR backups are replicated. Geo restore or ability to recover from regional outage is only
Review + create
Next : Networking >


Microsoft Azure
Search resources, services, and docs (G+/)
Home
Create SQL Database
Microsoft
Basics
Networking
Security
Additional settings
Tags
Review + create
Customize additional configuration parameters including collation & sample data.
Data source
Start with a blank database, restore from a backup or select sample data to populate your new database.
Use existing data *
None
Backup
Sample
AdventureWorksLT will be created as the sample database.
AdventureWorksLT
Selecting this sample will modify the "Compute + Storage" settings configured in Basics, for backup compatibility. Visit "Basics" or "Review + Create" tabs to review the changes.
Do you want to continue?
OK
Cancel
Review + create
< Previous
Next : Tags >

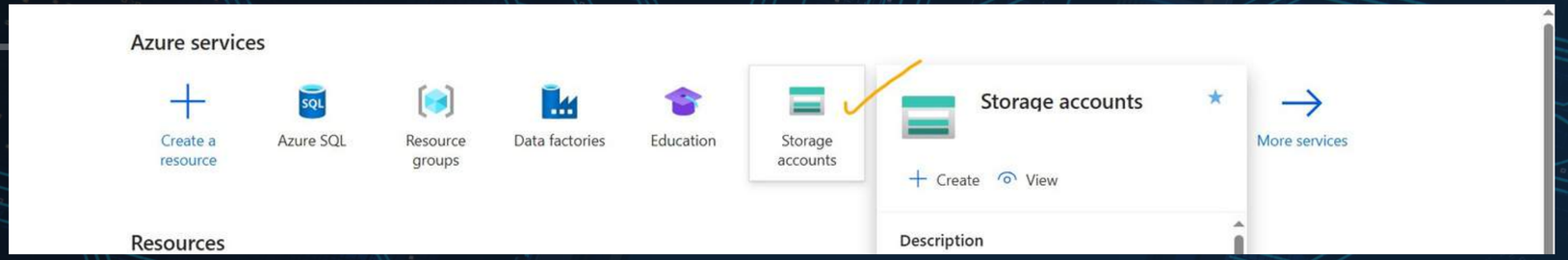

Microsoft Azure
Search resources, services, and d
Home
Create SQL Database
Microsoft
server 'ourserver0' and all databases it manages. Learn more
Network connectivity
Choose an option for configuring connectivity to your server via public endpoint or private e
creates with defaults and you can configure connection method after server creation. Learn
Connectivity method *
No access
Public endpoint
Private endpoint
Firewall rules
Setting 'Allow Azure services and resources to access this server' to Yes allows communicatio
the Azure boundary, that may or may not be part of your subscription. Learn more
Setting 'Add current client IP address' to Yes will add an entry for your client IP address to th
Allow Azure services and resources to access this server *
No
Yes
Add current client IP address *
No
Yes
Connection policy
Configure how clients communicate with your SQL database server. Learn more
Review + create
< Previous
Next : Security >

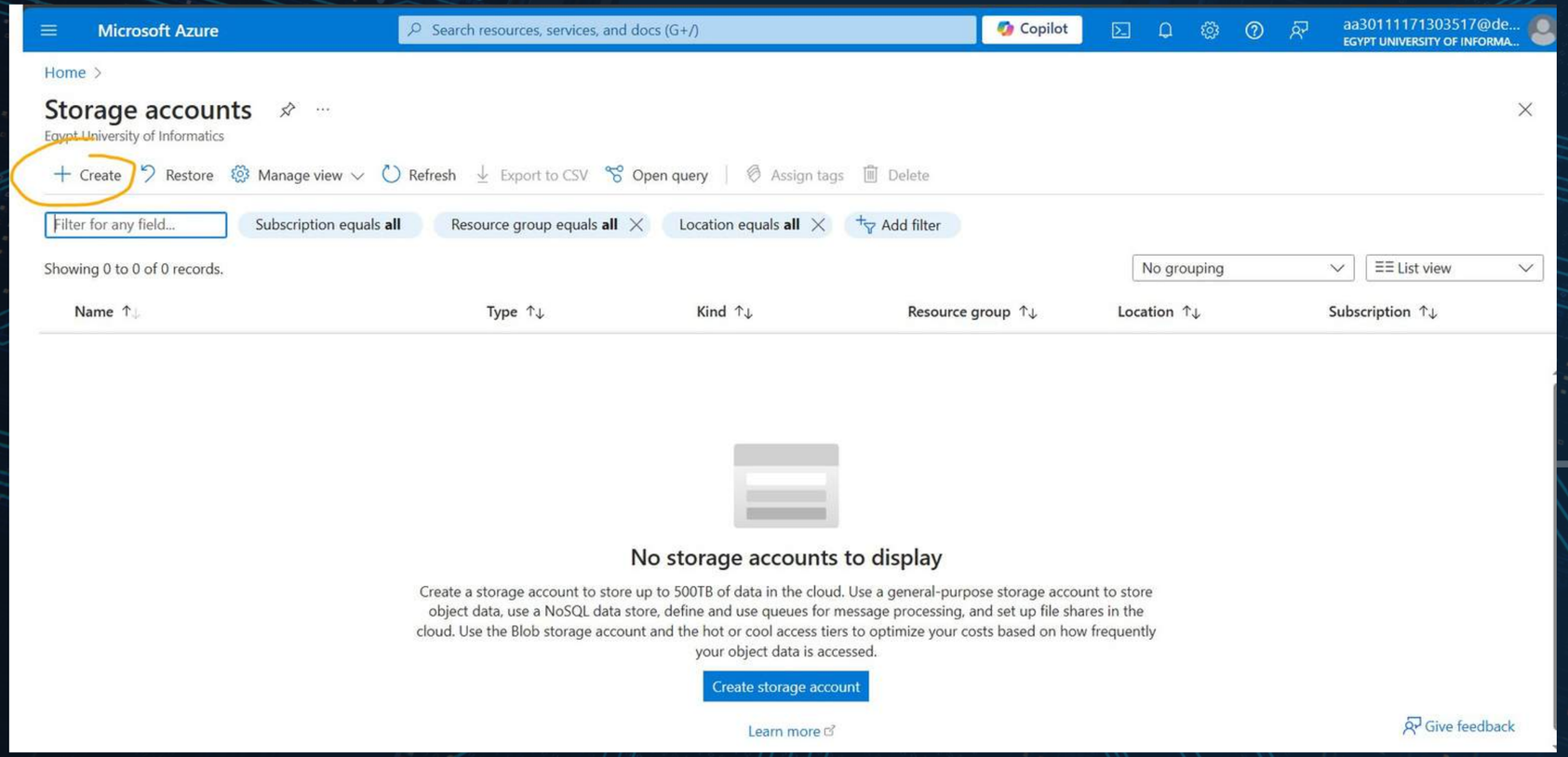

Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
Home
Microsoft.SQLDatabase.newDatabaseNewServer_f02e022f560b4d62badd2 | Overview
Deployment
Search
Overview
Inputs
Outputs
Template
Delete
Cancel
Redeploy
Download
Refresh
Your deployment is complete
Deployment name : Microsoft.SQLDatabase.newDatabaseNe...
Start time : 9/18/2024, 4:52:31 PM
Subscription : Azure for Students
Correlation ID : 2aac145e-0bab-4f83-b077-e2c2543041de
Resource group : trianingresource
Deployment details
Next steps
Go to resource
Give feedback
Tell us about your experience with deployment
Deployment succeeded
Deployment 'Microsoft.SQLDatabase.newDatabaseNewServer_f02e0... to resource group 'trianingresource' was successful.
Go to resource
Pin to dashboard
Cost management
Get notified to stay within your budget and prevent unexpected charges on your bill.
Set up cost alerts >
Microsoft Defender for Cloud
Secure your apps and infrastructure
Go to Microsoft Defender for Cloud >
Free Microsoft tutorials
Start learning today >
Work with an expert
Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.
Find an Azure expert >

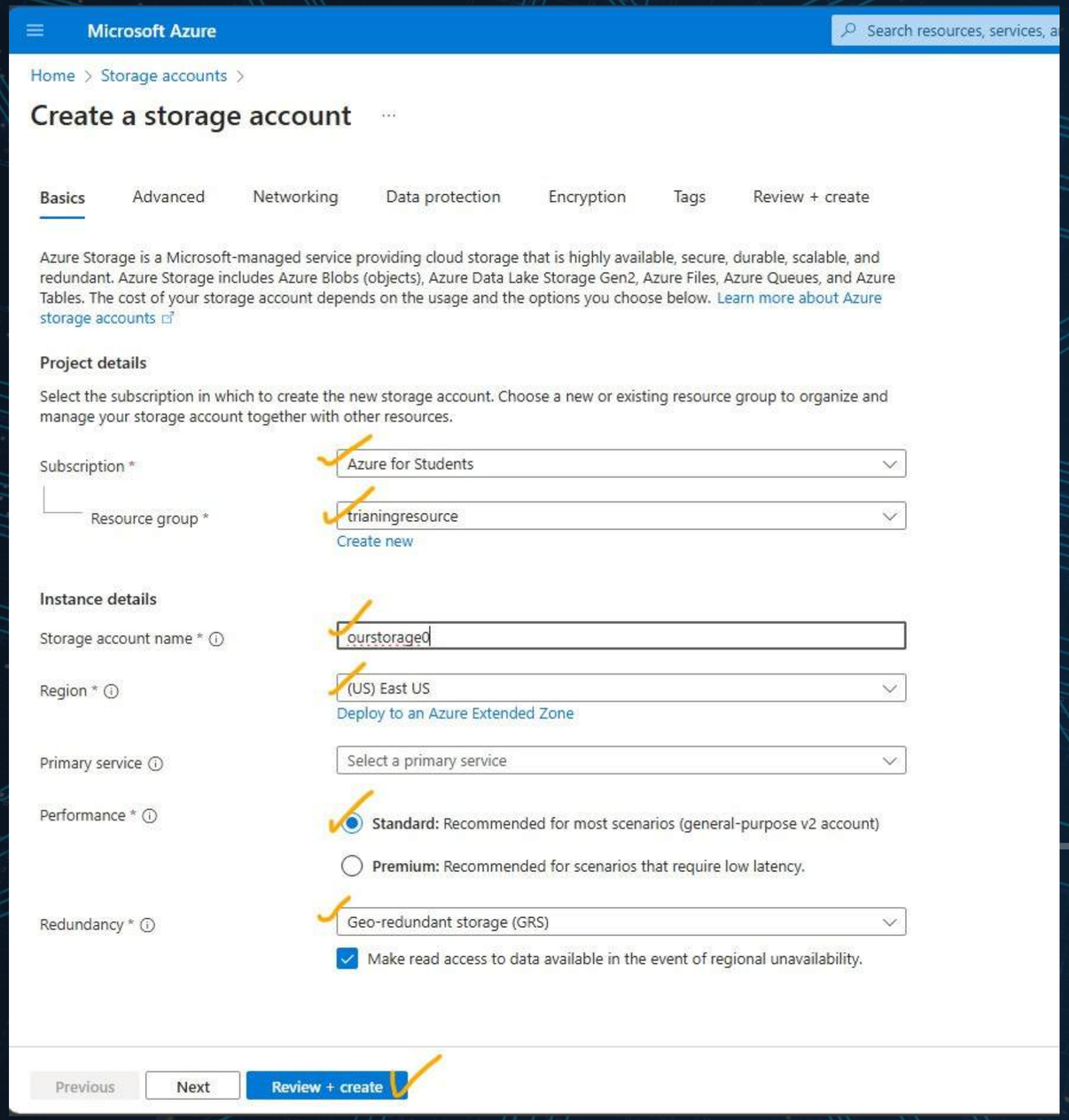

Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
aa30111171303517@de...
EGYPT UNIVERSITY OF INFORMA...
Home > Microsoft.SQLDatabase.newDatabaseNewServer_f02e022f560b4d62badd2 | Overview > our_db (ourserver0/our_db)
our_db (ourserver0/our_db) | Query editor (preview)
SQL database
Search
Login
New Query
Open query
Feedback
Getting started
Overview
Activity log
Tags
Diagnose and solve problems
Query editor (preview)
Mirror database in Fabric (preview)
Settings
Data management
Replicas
Sync to other databases
Integrations
Power Platform
Security
Intelligent performance
Monitoring
Showing limited object explorer here. For full capability please click here to open Azure Data Studio.
Tables
dbo.BuildVersion
dbo.ErrorLog
SalesLT.Address
SalesLT.Customer
SalesLT.CustomerAddress
SalesLT.Product
SalesLT.ProductCategory
SalesLT.ProductDescription
SalesLT.ProductModel
SalesLT.ProductModelProductDe
SalesLT.SalesOrderDetail
SalesLT.SalesOrderHeader
Views
Stored Procedures
Query 1
Run
Cancel query
Save query
Export data as
Show only Editor
Open Copilot
1
Results
Messages
Search to filter items...
Ready

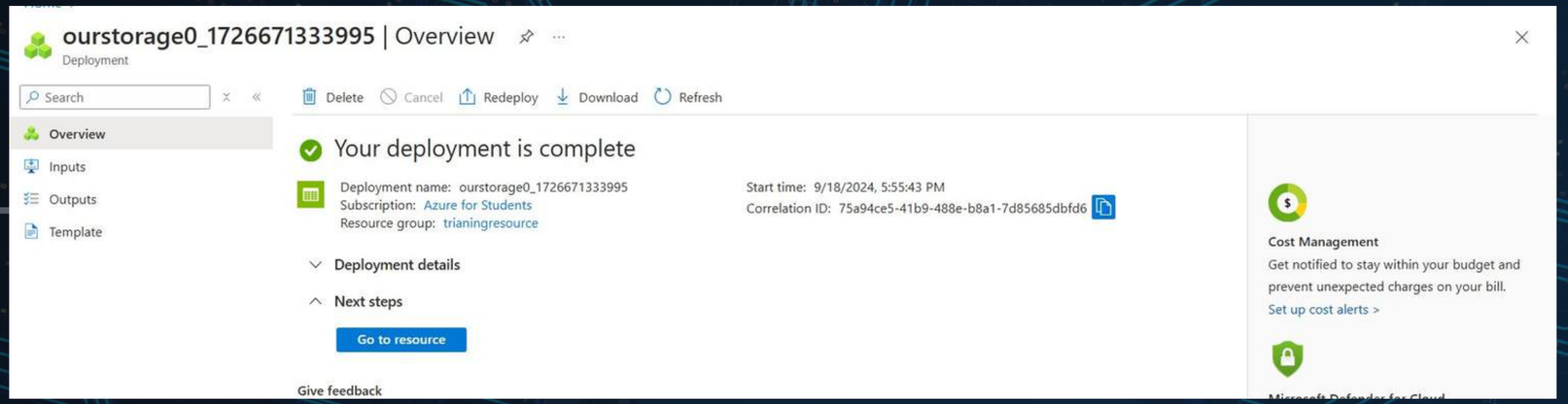
Done

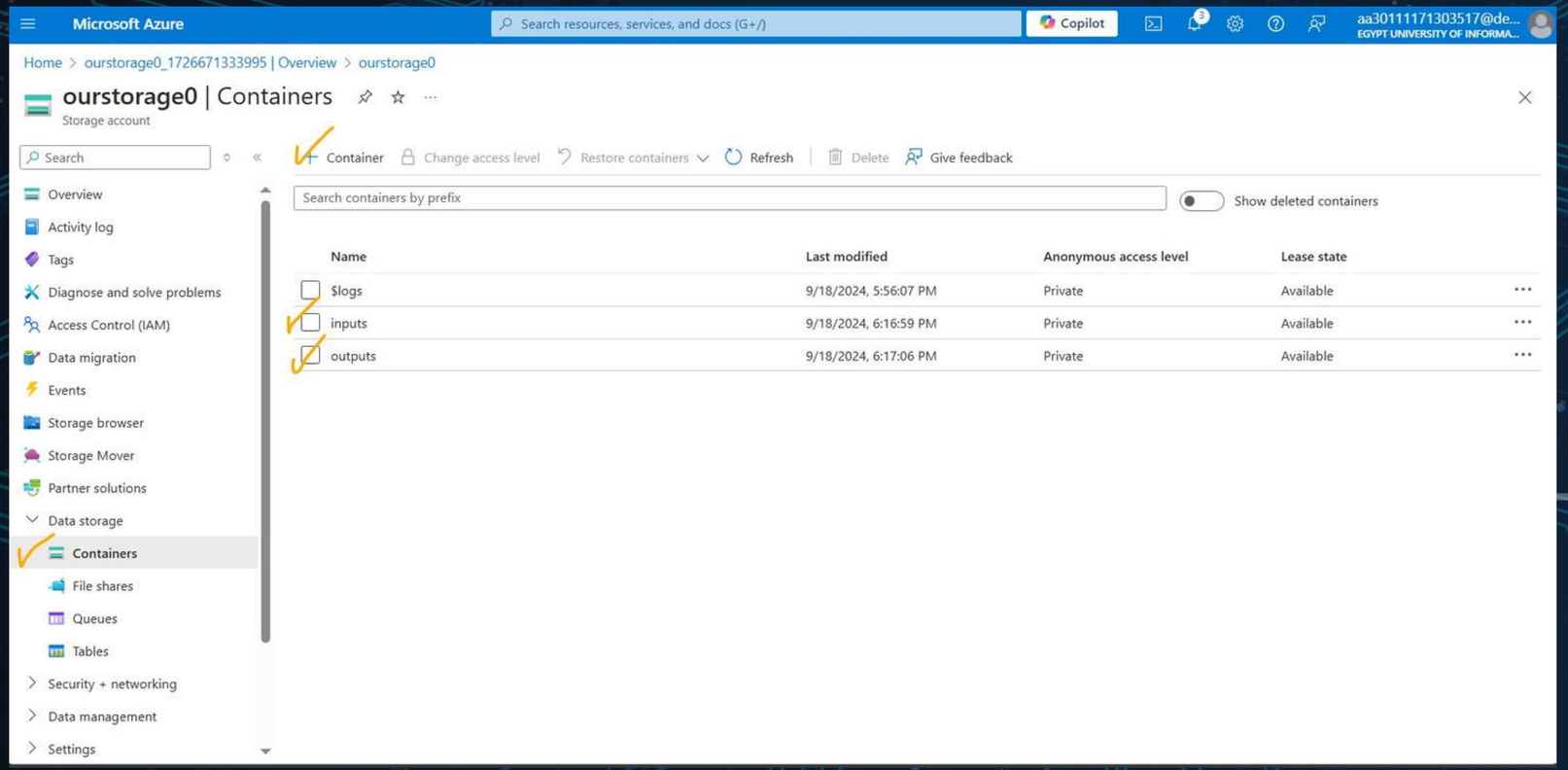
step 2: Azure storage account





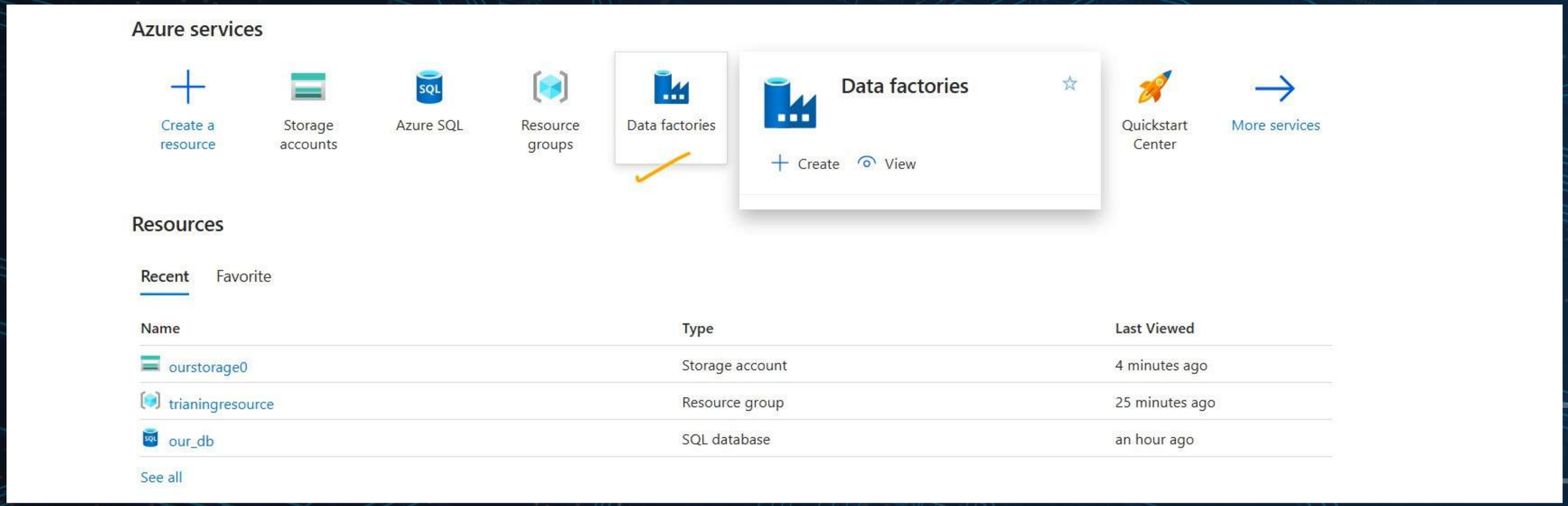

Microsoft Azure
Search resources, services, a
Home > Storage accounts >
Create a storage account
Basics
Advanced
Networking
Data protection
Encryption
Tags
Review + create
Azure Storage is a Microsoft-managed service providing cloud storage that is highly available, secure, durable, scalable, and redundant. Azure Storage includes Azure Blobs (objects), Azure Data Lake Storage Gen2, Azure Files, Azure Queues, and Azure Tables. The cost of your storage account depends on the usage and the options you choose below. Learn more about Azure storage accounts
Project details
Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.
Subscription *
Azure for Students
Resource group *
trianingresource
Create new
Instance details
Storage account name *
ourstorage0
Region *
(US) East US
Deploy to an Azure Extended Zone
Primary service
Select a primary service
Performance *
Standard: Recommended for most scenarios (general-purpose v2 account)
Premium: Recommended for scenarios that require low latency.
Redundancy *
Geo-redundant storage (GRS)
Make read access to data available in the event of regional unavailability.
Previous
Next
Review + create


ourstorage0_1726671333995 | Overview
Deployment
Search
Overview
Inputs
Outputs
Template
Delete
Cancel
Redeploy
Download
Refresh
Your deployment is complete
Deployment name: ourstorage0_1726671333995
Subscription: Azure for Students
Resource group: trianingresource
Start time: 9/18/2024, 5:55:43 PM
Correlation ID: 75a94ce5-41b9-488e-b8a1-7d85685dbfd6
Deployment details
Next steps
Go to resource
Give feedback
Cost Management
Get notified to stay within your budget and prevent unexpected charges on your bill.
Set up cost alerts >


Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
Home > ourstorage0_1726671333995 | Overview > ourstorage0
ourstorage0 | Containers
Storage account
Search
Container
Change access level
Restore containers
Refresh
Delete
Give feedback
Search containers by prefix
Show deleted containers
Overview
Activity log
Tags
Diagnose and solve problems
Access Control (IAM)
Data migration
Events
Storage browser
Storage Mover
Partner solutions
Data storage
Containers
File shares
Queues
Tables
Security + networking
Data management
Settings
Name
Last modified
Anonymous access level
Lease state
$logs
9/18/2024, 5:56:07 PM
Private
Available
inputs
9/18/2024, 6:16:59 PM
Private
Available
outputs
9/18/2024, 6:17:06 PM
Private
Available

Done

step 3: Azure Data Factory



Create Data Factory

Changes on this step may reset later selections you have made. Review all options prior to deployment.

Basics | Git configuration | Networking | Advanced | Tags | Review + create

One-click to create data factory with sample pipeline and datasets. Try it

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * : Azure for Students

Resource group * : trianingresource

Create new

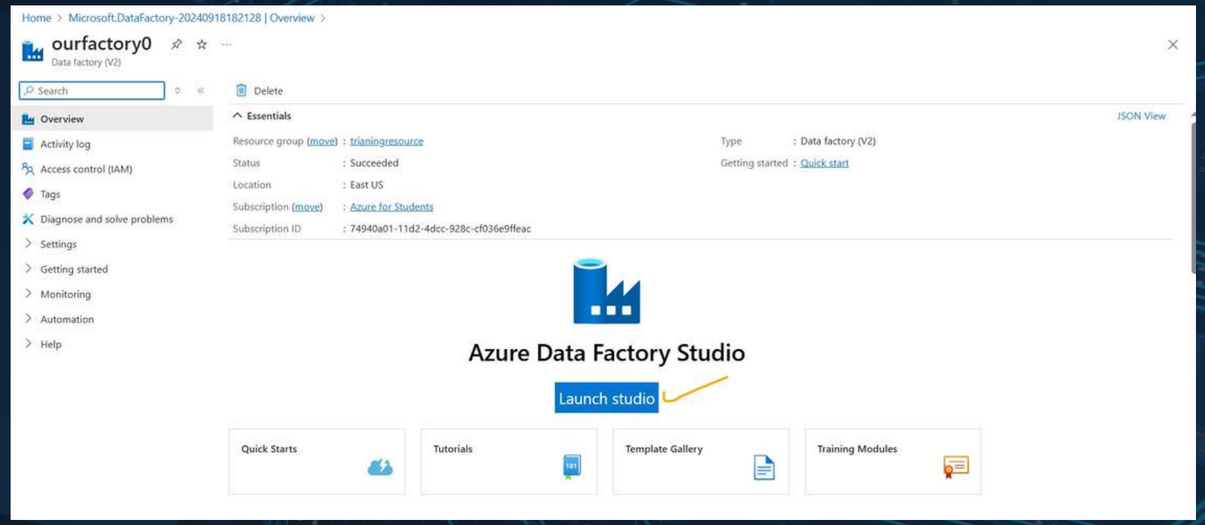
Instance details

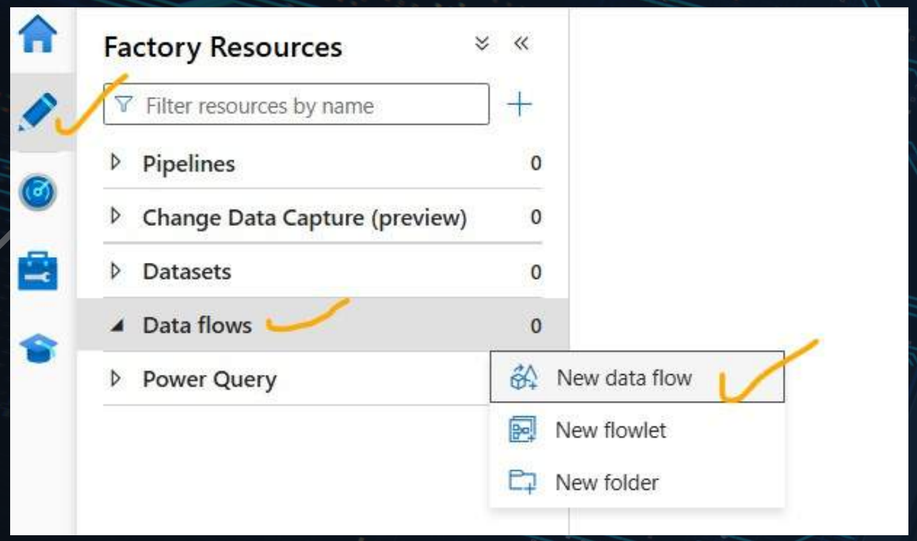
Name * : ourfactory0

Region * : East US

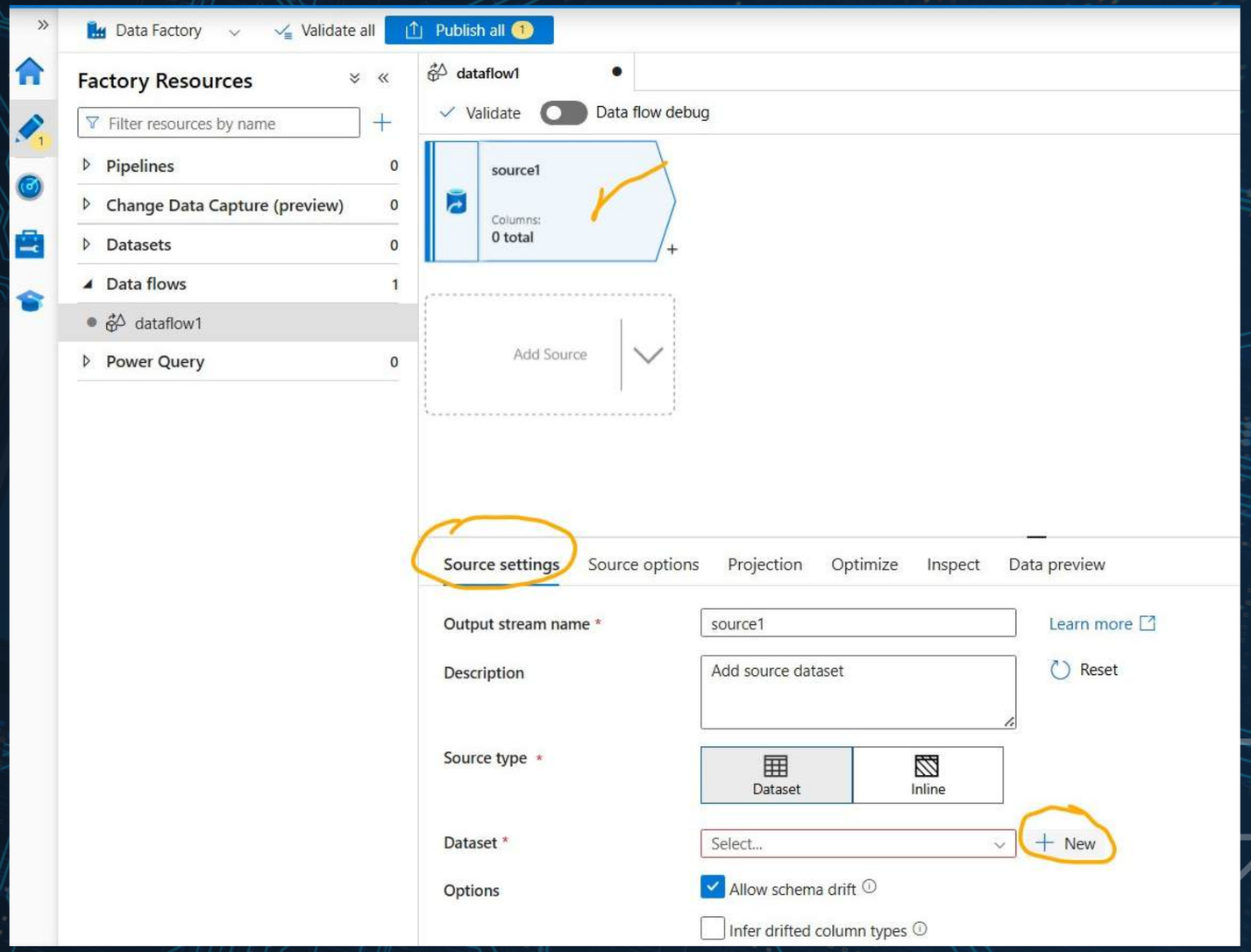
Version * : V2

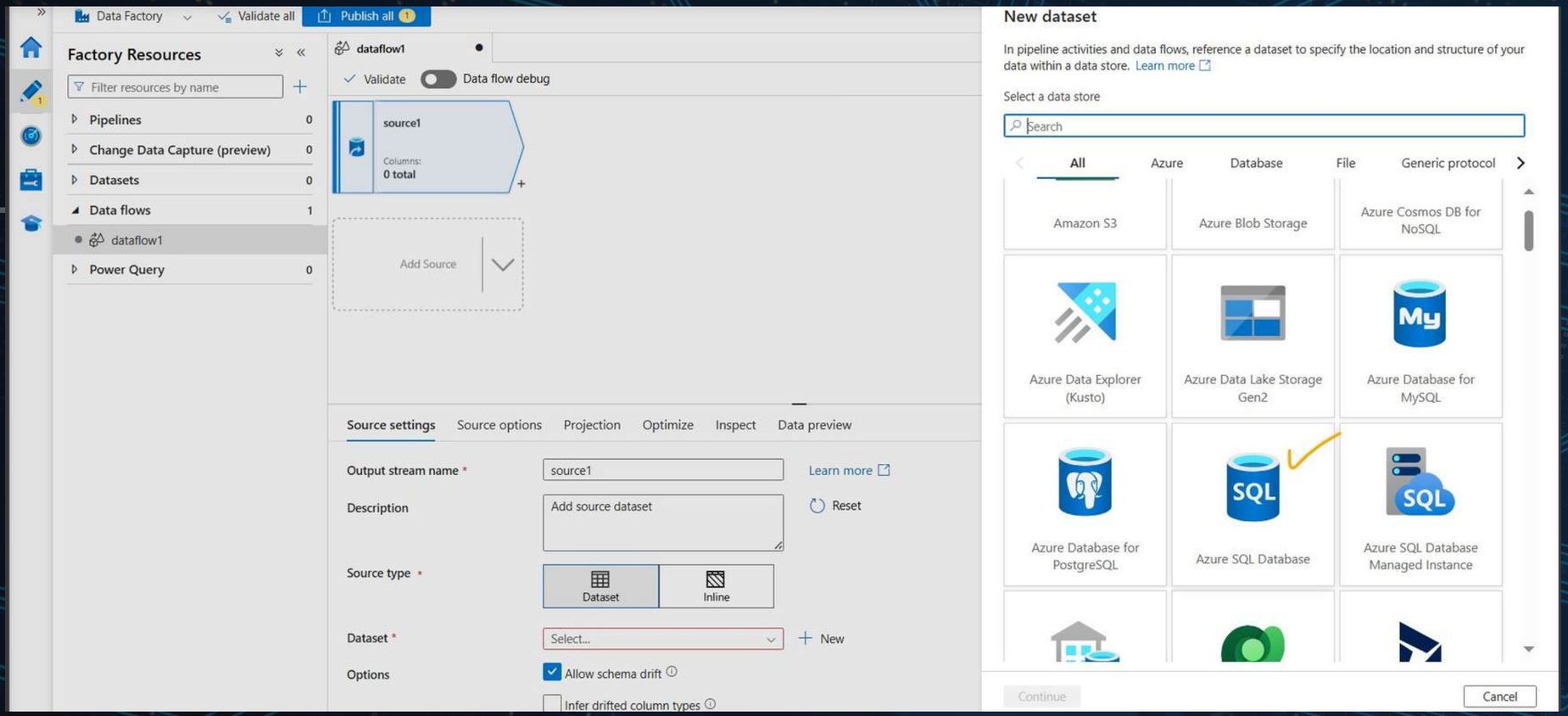
Previous | Next | Review + create



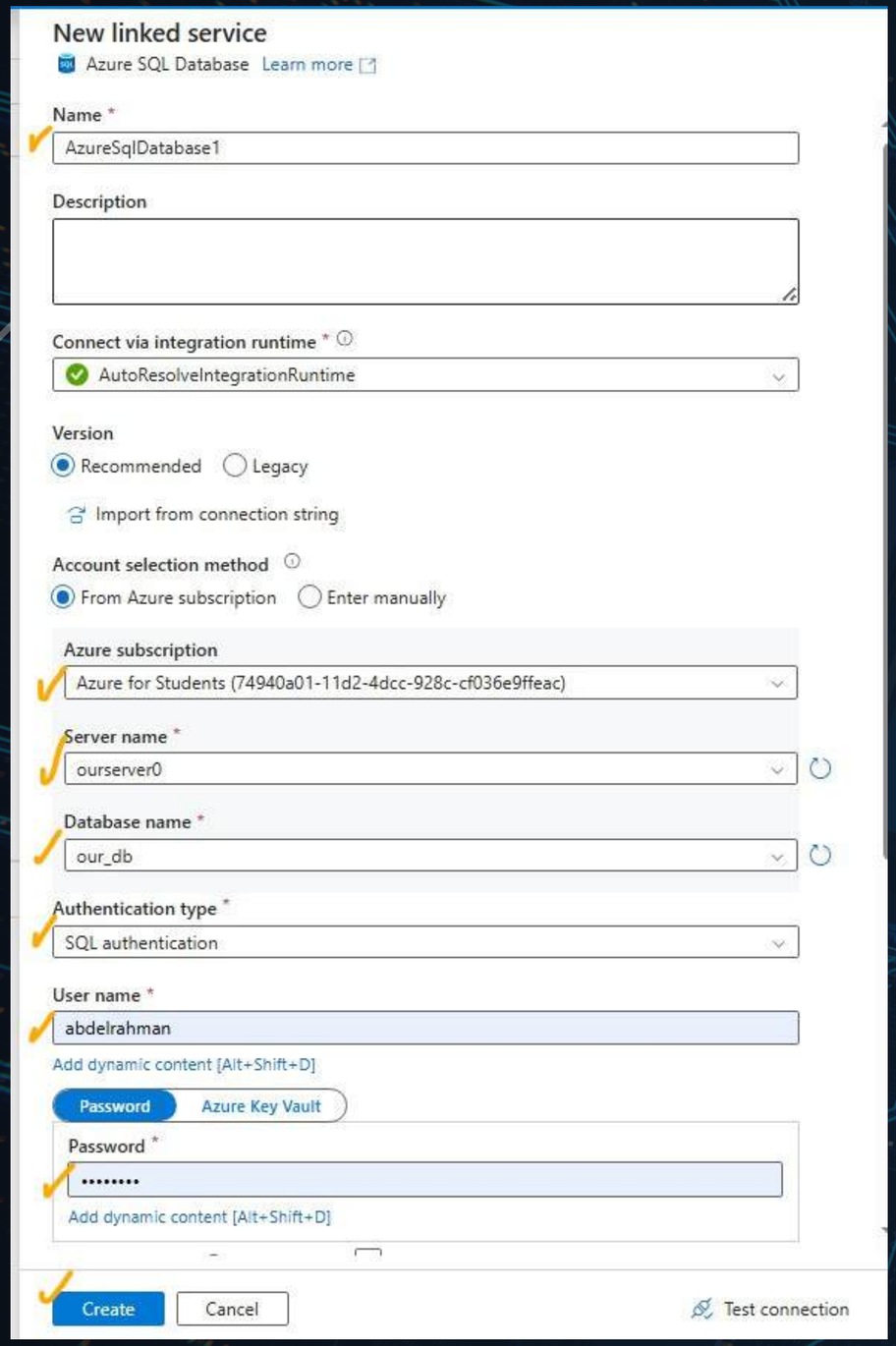

Factory Resources
Filter resources by name
Pipelines 0
Change Data Capture (preview) 0
Datasets 0
Data flows 0
Power Query
New data flow
New flowlet
New folder

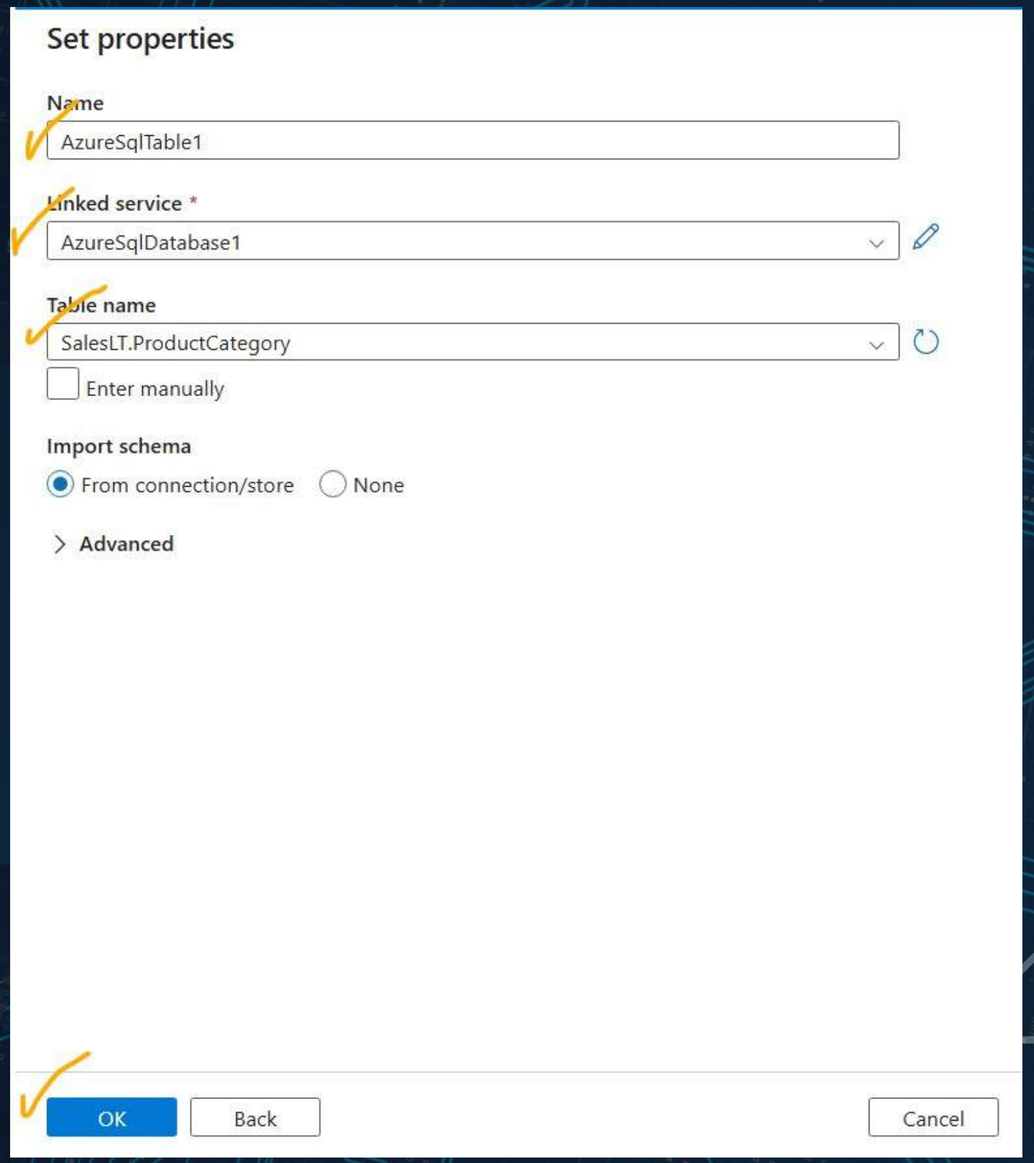
source 1


Data Factory
Validate all
Publish all 1
Factory Resources
Filter resources by name
Pipelines 0
Change Data Capture (preview) 0
Datasets 0
Data flows 1
dataflow1
Power Query 0
dataflow1
Validate
Data flow debug
source1
Columns:
0 total
Add Source
Source settings
Source options
Projection
Optimize
Inspect
Data preview
Output stream name *
source1
Learn more
Description
Add source dataset
Reset
Source type *
Dataset
Inline
Dataset *
Select...
New
Options
Allow schema drift
Infer drifted column types

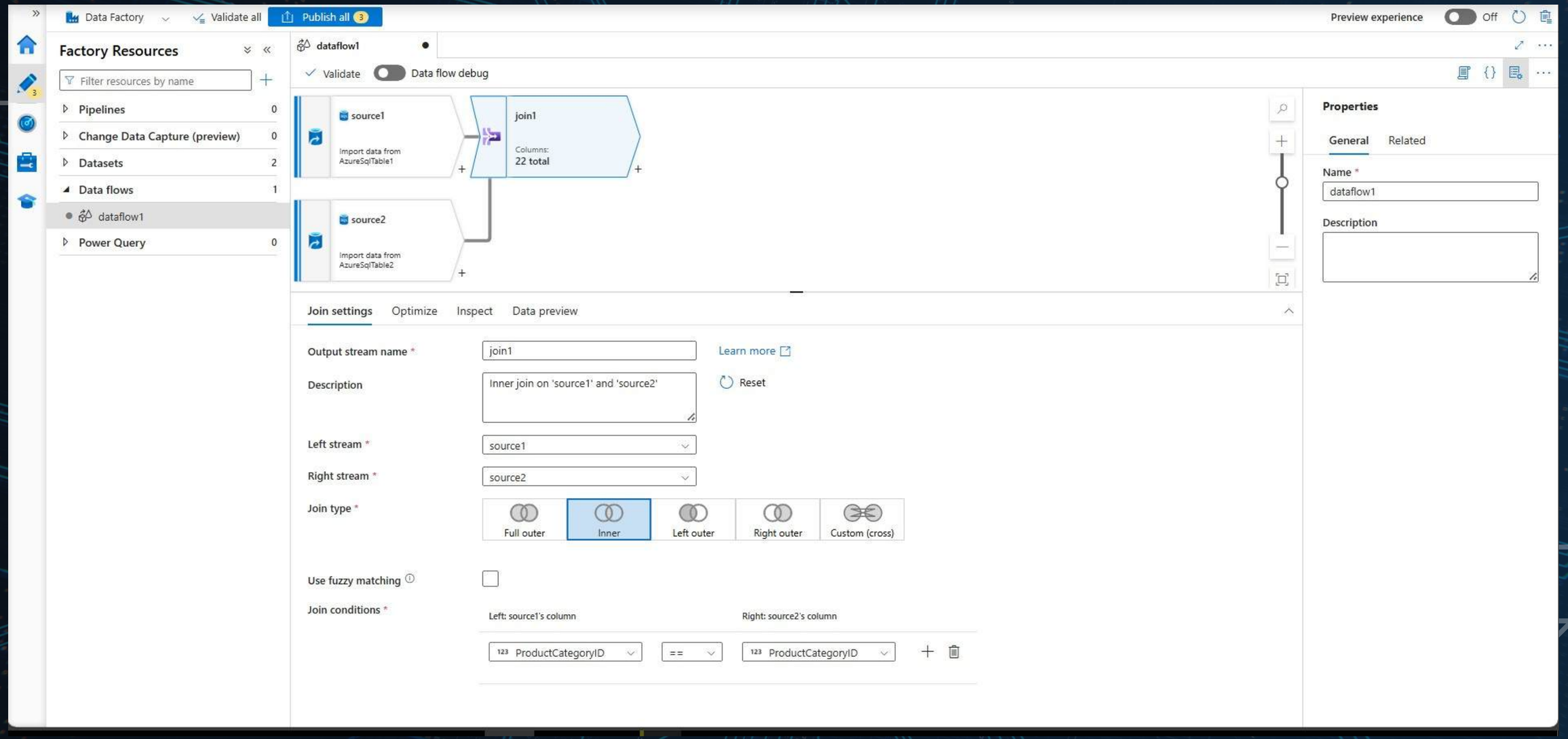

Data Factory
Validate all
Publish all 1
Factory Resources
Filter resources by name
Pipelines 0
Change Data Capture (preview) 0
Datasets 0
Data flows 1
dataflow1
Power Query 0
dataflow1
Validate
Data flow debug
source1
Columns:
0 total
Add Source
Source settings
Source options
Projection
Optimize
Inspect
Data preview
Output stream name *
source1
Learn more
Description
Add source dataset
Reset
Source type *
Dataset
Inline
Dataset *
Select...
New
Options
Allow schema drift
Infer drifted column types
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
Amazon S3
Azure Blob Storage
Azure Cosmos DB for NoSQL
Azure Data Explorer (Kusto)
Azure Data Lake Storage Gen2
Azure Database for MySQL
Azure Database for PostgreSQL
Azure SQL Database
Azure SQL Database Managed Instance
Continue
Cancel

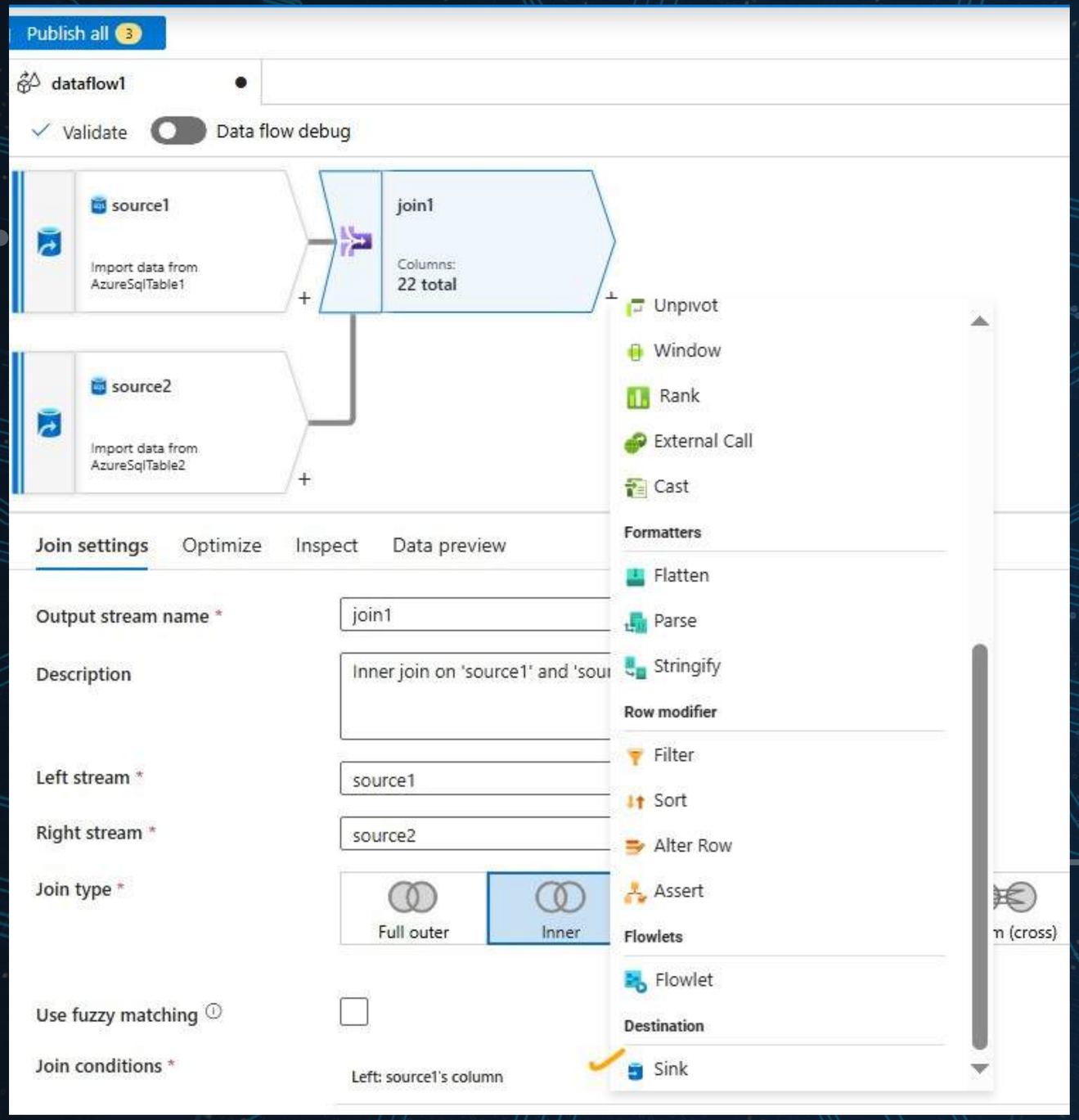
Set properties
Name
AzureSqlTable1
Linked service *
Select...

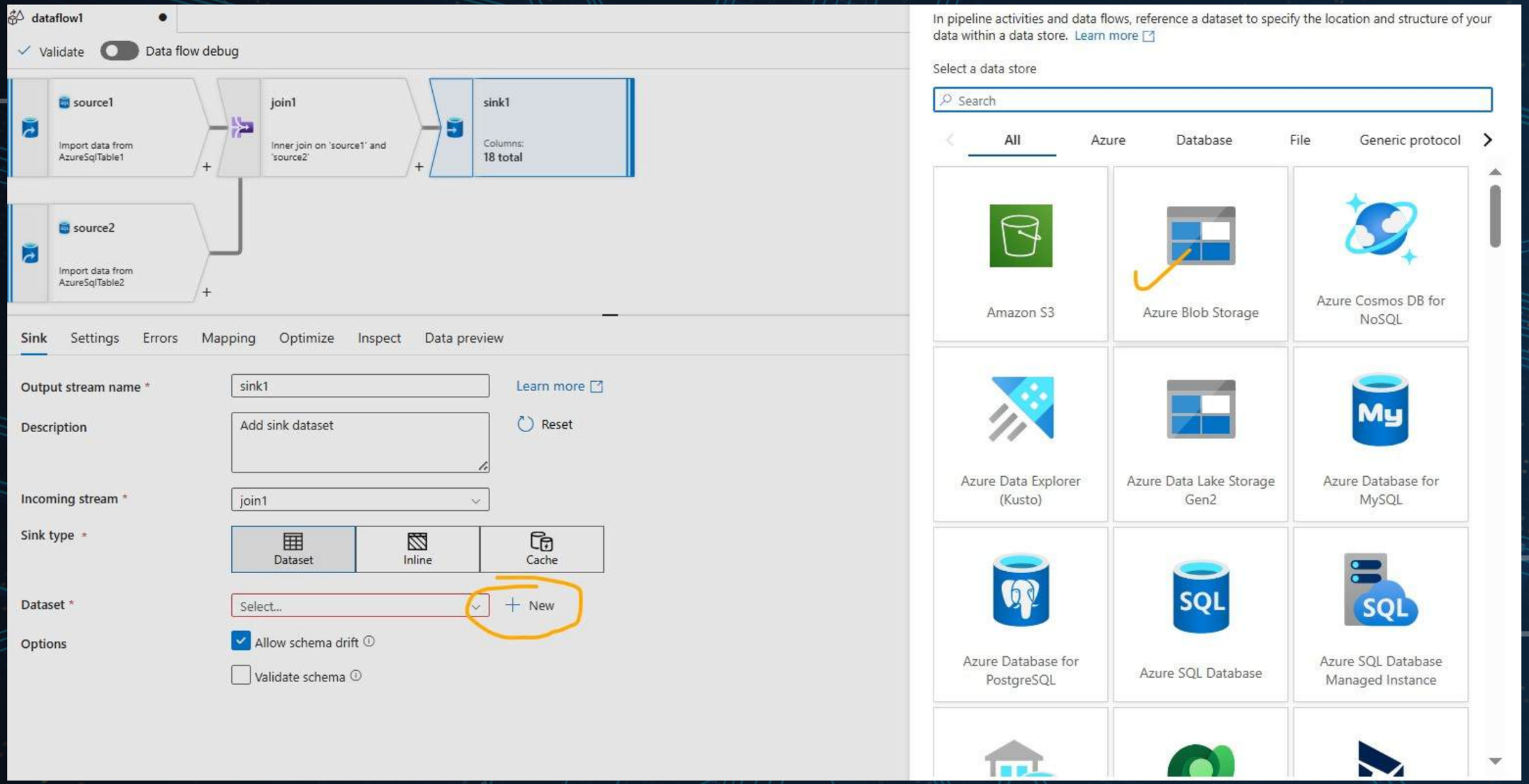


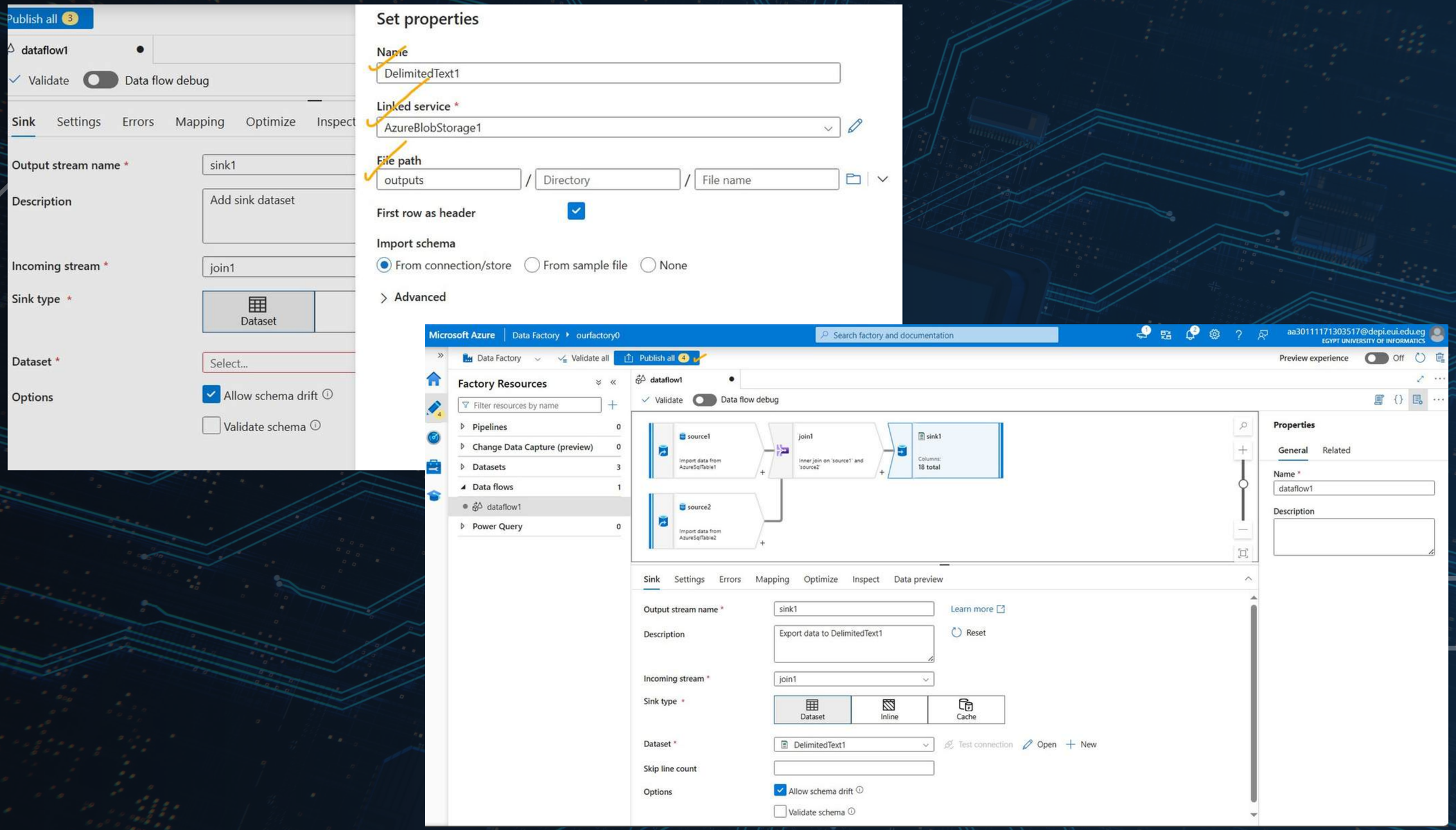


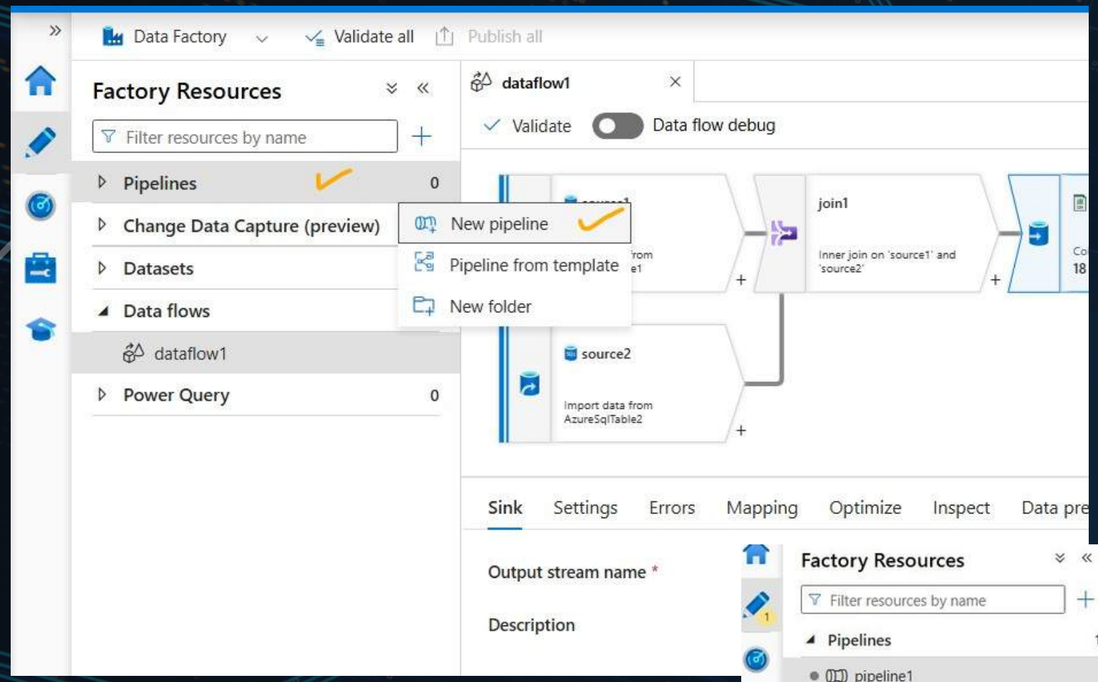
And the second source

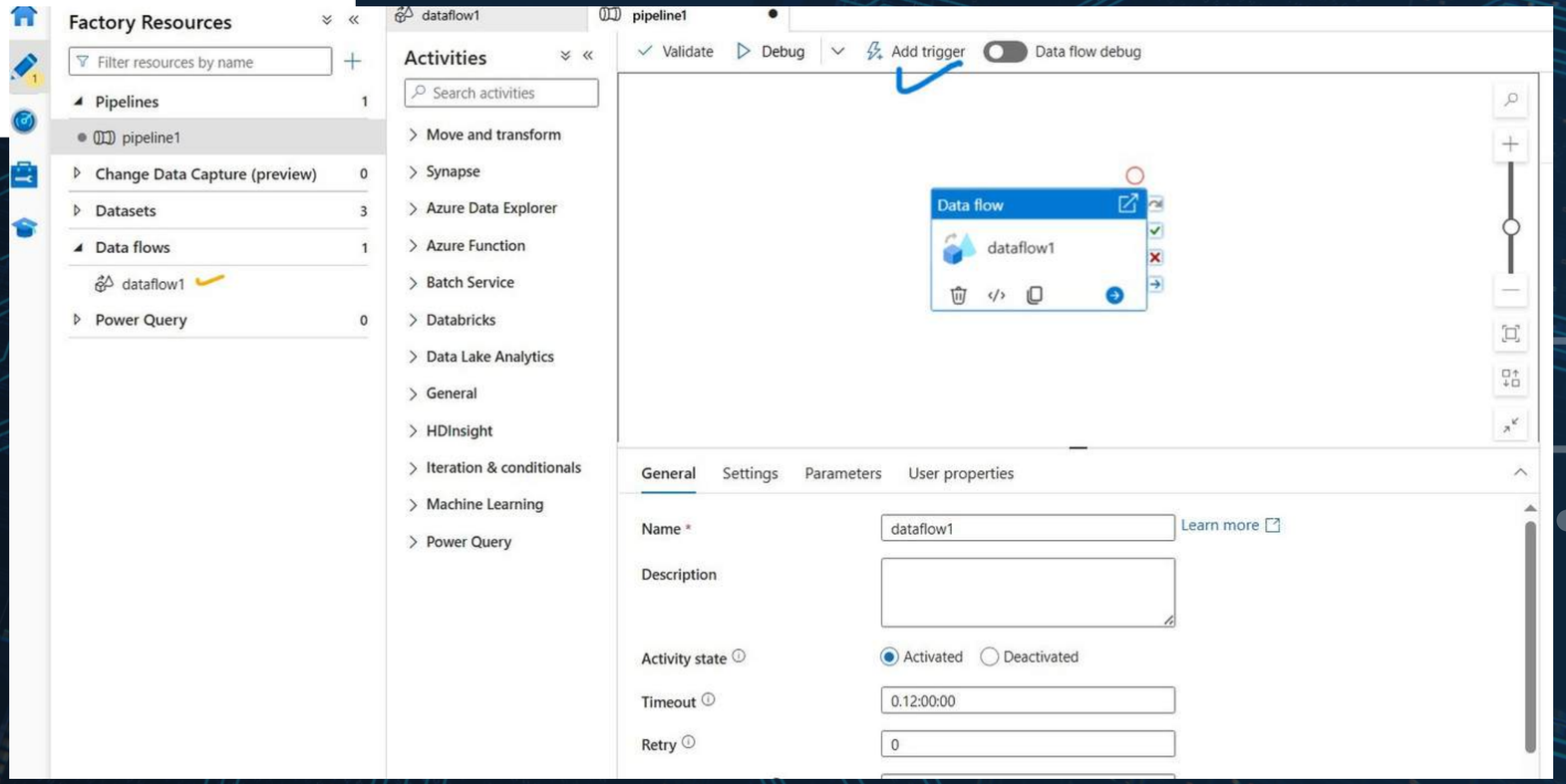

Data Factory
Validate all
Publish all 3
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 0
Change Data Capture (preview) 0
Datasets 2
Data flows 1
dataflow1
Power Query 0
dataflow1
Validate
Data flow debug
source1
Import data from AzureSqlTable1
join1
Columns:
22 total
source2
Import data from AzureSqlTable2
Properties
General
Related
Name *
dataflow1
Description
Join settings
Optimize
Inspect
Data preview
Output stream name *
join1
Learn more
Description
Inner join on 'source1' and 'source2'
Reset
Left stream *
source1
Right stream *
source2
Join type *
Full outer
Inner
Left outer
Right outer
Custom (cross)
Use fuzzy matching
Join conditions *
Left: source1's column
Right: source2's column
ProductCategoryID
==
ProductCategoryID


Publish all 3
dataflow1
Validate
Data flow debug
source1
Import data from AzureSqlTable1
join1
Columns:
22 total
source2
Import data from AzureSqlTable2
Unpivot
Window
Rank
External Call
Cast
Formatters
Flatten
Parse
Stringify
Row modifier
Filter
Sort
Alter Row
Assert
Flowlets
Flowlet
Destination
Sink
Join settings
Optimize
Inspect
Data preview
Output stream name *
join1
Description
Inner join on 'source1' and 'sou
Left stream *
source1
Right stream *
source2
Join type *
Full outer
Inner
n (cross)
Use fuzzy matching
Join conditions *
Left: source1's column

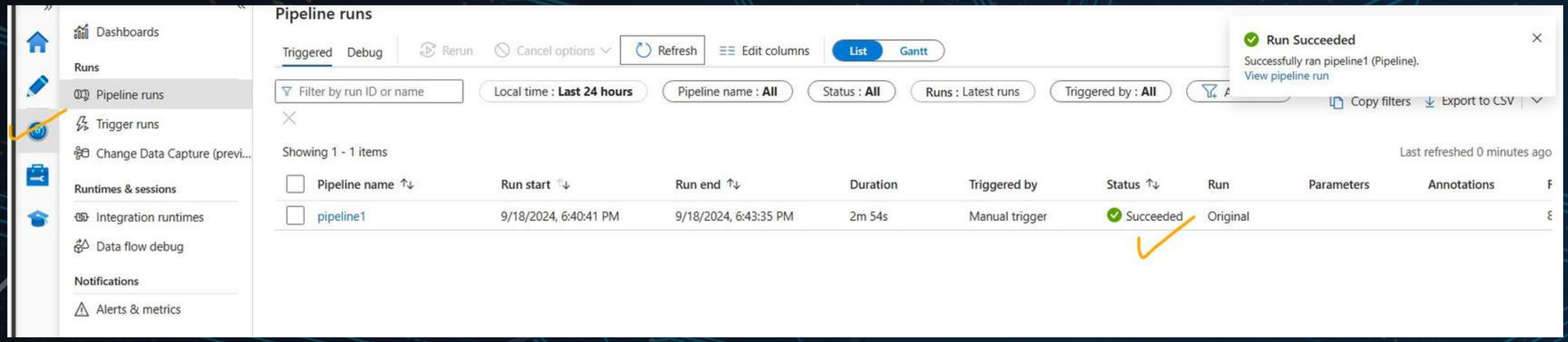

dataflow1
Validate
Data flow debug
source1
Import data from AzureSqlTable1
join1
Inner join on 'source1' and 'source2'
sink1
Columns:
18 total
source2
Import data from AzureSqlTable2
Sink
Settings
Errors
Mapping
Optimize
Inspect
Data preview
Output stream name *
sink1
Learn more
Description
Add sink dataset
Reset
Incoming stream *
join1
Sink type *
Dataset
Inline
Cache
Dataset *
Select...
New
Options
Allow schema drift
Validate schema
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
Amazon S3
Azure Blob Storage
Azure Cosmos DB for NoSQL
Azure Data Explorer (Kusto)
Azure Data Lake Storage Gen2
Azure Database for MySQL
Azure Database for PostgreSQL
Azure SQL Database
Azure SQL Database Managed Instance

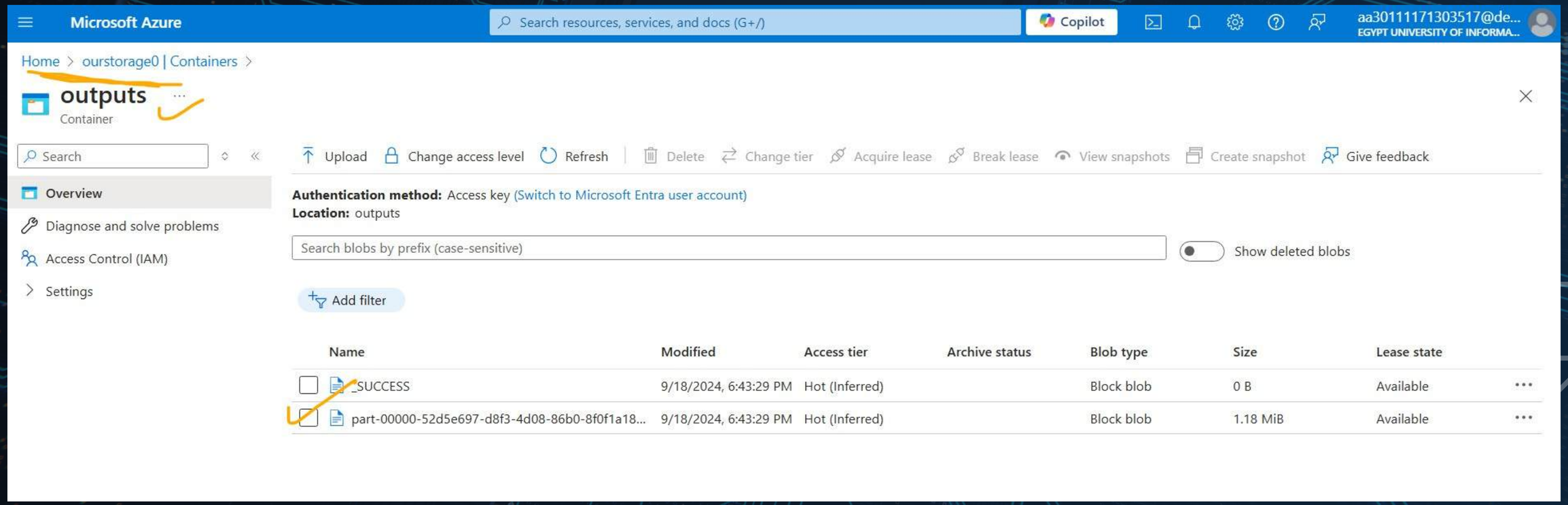

Publish all 3
dataflow1
Validate
Data flow debug
Sink
Settings
Errors
Mapping
Optimize
Inspect
Output stream name *
sink1
Description
Add sink dataset
Incoming stream *
join1
Sink type *
Dataset
Dataset *
Select...
Options
Allow schema drift
Validate schema
Set properties
Name
DelimitedText1
Linked service *
AzureBlobStorage1
File path
outputs
Directory
File name
First row as header
Import schema
From connection/store
From sample file
None
Advanced
Microsoft Azure
Data Factory
ourfactory0
Search factory and documentation
Data Factory
Validate all
Publish all 4
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 0
Change Data Capture (preview) 0
Datasets 3
Data flows 1
dataflow1
Power Query 0
source1
Import data from AzureSqlTable1
join1
Inner join on 'source1' and 'source2'
sink1
Columns: 18 total
source2
Import data from AzureSqlTable2
Properties
General
Related
Name *
dataflow1
Description
Sink
Settings
Errors
Mapping
Optimize
Inspect
Data preview
Output stream name *
sink1
Learn more
Description
Export data to DelimitedText1
Reset
Incoming stream *
join1
Sink type *
Dataset
Inline
Cache
Dataset *
DelimitedText1
Test connection
Open
New
Skip line count
Options
Allow schema drift
Validate schema


Data Factory
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines 0
Change Data Capture (preview)
Datasets
Data flows
dataflow1
Power Query 0
New pipeline
Pipeline from template
New folder
dataflow1
Validate
Data flow debug
join1
Inner join on 'source1' and 'source2'
source2
Import data from AzureSqlTable2
Sink
Settings
Errors
Mapping
Optimize
Inspect
Output stream name *
Description


Factory Resources
Filter resources by name
Pipelines 1
pipeline1
Change Data Capture (preview) 0
Datasets 3
Data flows 1
dataflow1
Power Query 0
dataflow1
pipeline1
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
Data flow debug
Data flow
dataflow1
General
Settings
Parameters
User properties
Name *
dataflow1
Learn more
Description
Activity state
Activated
Deactivated
Timeout
0.12:00:00
Retry
0

output


Pipeline runs
Dashboards
Runs
Pipeline runs
Trigger runs
Change Data Capture (previ...
Runtimes & sessions
Integration runtimes
Data flow debug
Notifications
Alerts & metrics
Triggered
Debug
Rerun
Cancel options
Refresh
Edit columns
List
Gantt
Run Succeeded
Successfully ran pipeline1 (Pipeline).
View pipeline run
Filter by run ID or name
Local time : Last 24 hours
Pipeline name : All
Status : All
Runs : Latest runs
Triggered by : All
Copy filters
Export to CSV
Showing 1 - 1 items
Last refreshed 0 minutes ago
Pipeline name
Run start
Run end
Duration
Triggered by
Status
Run
Parameters
Annotations
pipeline1
9/18/2024, 6:40:41 PM
9/18/2024, 6:43:35 PM
2m 54s
Manual trigger
Succeeded
Original


Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
aa30111171303517@de...
EGYPT UNIVERSITY OF INFORMA...
Home > ourstorage0 | Containers >
outputs
Container
Search
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Upload
Change access level
Refresh
Delete
Change tier
Acquire lease
Break lease
View snapshots
Create snapshot
Give feedback
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: outputs
Search blobs by prefix (case-sensitive)
Show deleted blobs
Add filter
Name
Modified
Access tier
Archive status
Blob type
Size
Lease state
_SUCCESS
9/18/2024, 6:43:29 PM
Hot (Inferred)
Block blob
0 B
Available
part-00000-52d5e697-d8f3-4d08-86b0-8f0f1a18...
9/18/2024, 6:43:29 PM
Hot (Inferred)
Block blob
1.18 MiB
Available

Thank You

FOR YOUR ATTENTION

abdohossieny@gmail.com

01008707953