

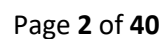
Microsoft Tech Series Lab: Data Lifecycle part 2

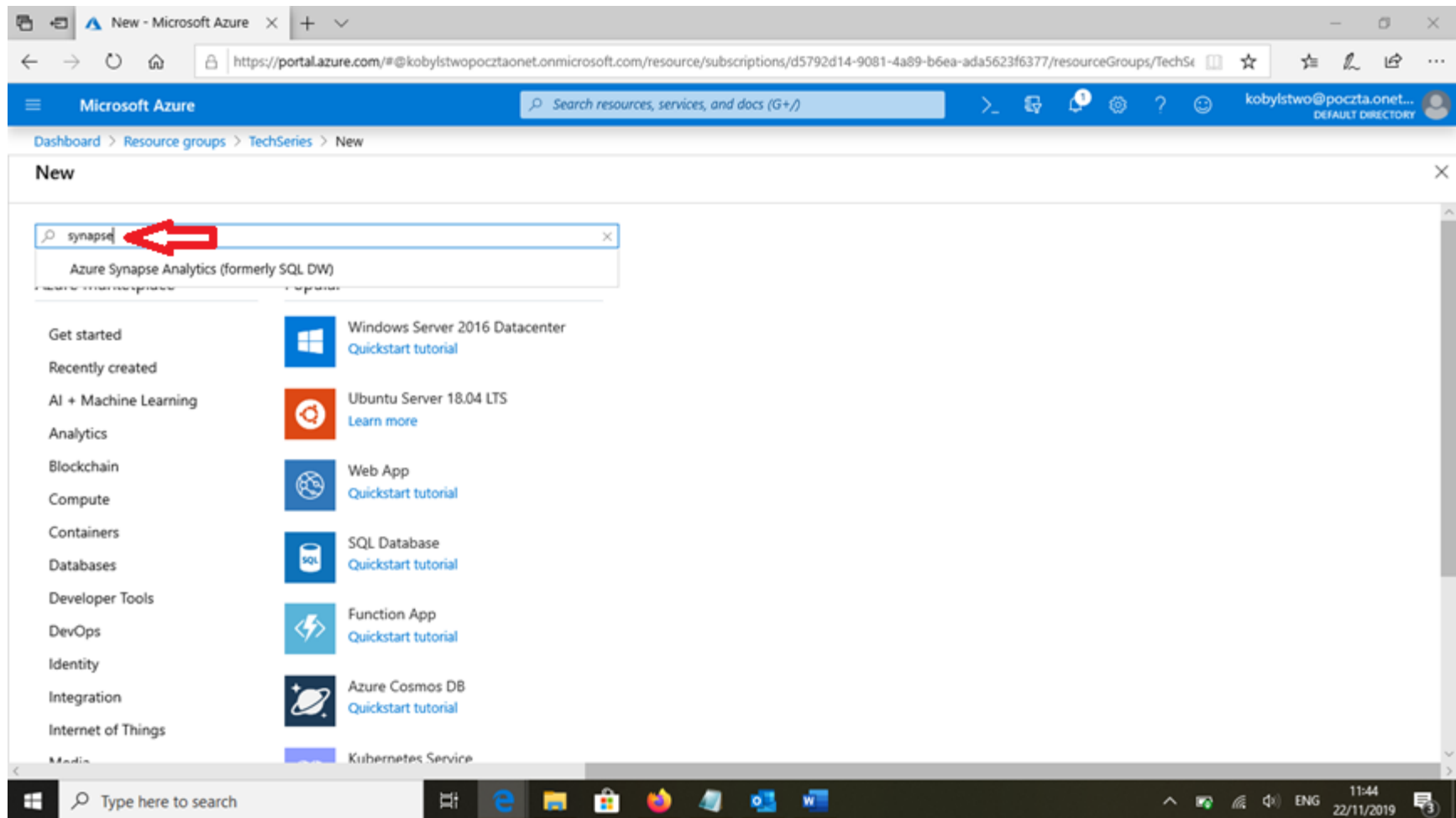
by Hubert K. Kobierzewski from Data Community Poland

The following scenario was covered during Microsoft Tech Series event on Nov 22nd, 2019

(Before you start, follow all steps provided by Damian Widera <https://tiny.pl/tnpbn>)

Use existing virtual SQL Server – this will make our exercise easier.





Marketplace - Microsoft x

https://portal.azure.com/#blade/Microsoft_Azure_Marketplace/MarketplaceOffersBlade/selectedMenuItemId/home/searchQuery/synapse

Microsoft Azure Search resources, services, and docs (G+)

Dashboard > Resource groups > TechSeries > New > Marketplace

Marketplace

My Saved List

Recently created

Service Providers


Categories


- Get Started
- AI + Machine Learning
- Analytics
- Blockchain
- Compute
- Containers
- Databases
- Developer Tools
- DevOps
- Identity

Showing All Results

synapse

Pricing: All Operating System: All Publisher: All






Azure Synapse Analytics
(formerly SQL DW)

Microsoft


Limitless analytics service with unmatched time to insight (Formerly SQL Data Warehouse)



SnapLogic Intelligent Integration Platform

SnapLogic


Unified platform for data and application integration



HushHush SSIS Data Masking Components

HushHush


SQL server 2012 sp2 with SSIS and HushHush data masking components R 1.2.7



Attunity Replicate

Attunity Inc.


Stream real-time data from multiple sources to Azure quickly, easily and securely.



Quest NetVault® Backup

Quest


Cloud Ready Data Protection for Hybrid Data Centers



NetVault® for Office 365

Quest

Cloud Ready Data Protection for Office 365



DataRoad Reflect

DataRoad Technologies LLC

DataRoad Reflect automates movement of data into and out of the cloud

Type here to search

11:44 22/11/2019

The screenshot shows the Azure portal interface for the Azure Synapse Analytics (formerly SQL DW) marketplace page. The browser address bar shows the URL: https://portal.azure.com/#blade/Microsoft_Azure_Marketplace/MarketplaceOffersBlade/selectedMenuItemId/home/searchQuery/synapse. The page header includes the Microsoft Azure logo and a search bar. The breadcrumb navigation shows: Dashboard > Resource groups > TechSeries > New > Marketplace > Azure Synapse Analytics (formerly SQL DW). The main content area features the Azure Synapse Analytics logo and the title "Azure Synapse Analytics (formerly SQL DW)". Below the title is a "Create" button, which is highlighted by a red arrow. The page also includes an "Overview" tab and a "Plans" tab. The "Overview" section describes the service as a limitless analytics service that brings together enterprise data warehousing and Big Data analytics. It states that it gives users the freedom to query data on their terms, using either serverless on-demand or provisioned resources-at scale. The "Useful Links" section includes links to Documentation, Service Overview, and Pricing Details.


Microsoft Azure

Search resources, services, and docs (G+)

Dashboard > Resource groups > TechSeries > New > Marketplace > Azure Synapse Analytics (formerly SQL DW)

Azure Synapse Analytics (formerly SQL DW)

Microsoft

 **Azure Synapse Analytics (formerly SQL DW)** [Save for later](#)

Micro Create

Create

Overview Plans

Azure Synapse Analytics is a limitless analytics service that brings together enterprise data warehousing and Big Data analytics.

It gives you the freedom to query data on your terms, using either serverless on-demand or provisioned resources-at scale. Azure Synapse brings these two worlds together with a unified experience to ingest, prepare, manage, and serve data for immediate BI and machine learning needs.

Simply put, Azure Synapse is Azure SQL Data Warehouse evolved. We have taken the [same industry leading data warehouse](#) to a whole new level of performance and capabilities. Businesses can continue running their existing data warehouse workloads in production today with Azure Synapse and will automatically benefit from the new capabilities which are in preview.

Useful Links

- [Documentation](#)
- [Service Overview](#)
- [Pricing Details](#)

SQL Data Warehouse - 1 x

https://portal.azure.com/#create/Microsoft.SQLDataWarehouse

Microsoft Azure

Search resources, services, and docs (G+)

kobylstwo@poczta.onet...
DEFAULT DIRECTORY

Dashboard > Resource groups > TechSeries > New > Marketplace > Azure Synapse Analytics (formerly SQL DW) > SQL Data Warehouse

SQL Data Warehouse

Microsoft

Welcome to Azure Synapse Analytics (formerly known as Azure SQL Data Warehouse). [Learn more.](#)

Project details

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

Subscription * ⓘ Visual Studio Premium with MSDN

Resource group * ⓘ TechSeries

[Create new](#)

Data warehouse details

Enter required settings for this data warehouse, including picking a logical server and configuring the performance level.

Data warehouse name * techseries ✓

Server * ⓘ hktechseriesserver (West Europe)

[Create new](#)

[Review + create](#) [Next : Additional settings >](#)

Type here to search

11:45
22/11/2019

SQL Data Warehouse - 1 x

https://portal.azure.com/#create/Microsoft.SQLDataWarehouse

Microsoft Azure

Search resources, services, and docs (G+)

kobylistwo@poczta.onet...
DEFAULT DIRECTORY

Dashboard > Resource groups > TechSeries > New > Marketplace > Azure Synapse Analytics (formerly SQL DW) > SQL Data Warehouse

SQL Data Warehouse

Microsoft

Welcome to Azure Synapse Analytics (formerly known as Azure SQL Data Warehouse). [Learn more.](#)

Resource group * ⓘ
TechSeries
[Create new](#)

Data warehouse details

Enter required settings for this data warehouse, including picking a logical server and configuring the performance level.

Data warehouse name *
techseries ✓

Server * ⓘ
hktechseriesserver (West Europe)
[Create new](#)

Performance level * ⓘ
Gen2
DW1000c
[Select performance level](#)

[Review + create](#) [Next : Additional settings >](#)

Type here to search

11:47
22/11/2019

Configure performance

https://portal.azure.com/#create/Microsoft.SQLDataWarehouse

Microsoft Azure

Search resources, services, and docs (G+)

kobylstwo@poczta.onet...
DEFAULT DIRECTORY

Dashboard > Resource groups > TechSeries > New > Marketplace > Azure Synapse Analytics > SQL Data Warehouse

SQL Data Warehouse

Microsoft

Welcome to Azure Synapse Analytics (formerly known as Azure SQL Data Warehouse).

Resource group *
[Create new](#)

Data warehouse details

Enter required settings for this data warehouse, including picking a logical server and

Data warehouse name * 2 (move left and right to change value)

Server *
[Create new](#)

Performance level *

Gen2
DW1000c
[Select performance level](#)

[Review + create](#) [Next : Additional settings >](#)

Configure performance

Configure your performance level that best fits your data warehousing needs.

Gen2

Offers the highest performance and storage scalability options for intensive workloads.

Starting at 1.27 EUR / hour

[Learn more about performance level](#)

Gen1

Offers the lowest compute scale options for less demanding

Not available

Starting at -- / hour

[Learn more about pricing](#)

Scale your system

2.55 EUR / hour
200 cDWU

[Price of your system](#)

[Apply](#)

SQL Data Warehouse - 1 X + -

https://portal.azure.com/#create/Microsoft.SQLDataWarehouse

Microsoft Azure Search resources, services, and docs (G+)

Dashboard > Resource groups > TechSeries > New > Marketplace > Azure Synapse Analytics (formerly SQL DW) > SQL Data Warehouse

SQL Data Warehouse

Microsoft

Welcome to Azure Synapse Analytics (formerly known as Azure SQL Data Warehouse). [Learn more.](#)

Basics * **Additional settings *** Tags Review + create

Customize additional configuration parameters including collation & sample data.

Data source

Start with a blank data warehouse, restore from a backup or select sample data to populate your new database.

Use existing data *

None Backup Sample

SQL data warehouse collation

Data warehouse collation defines the rules that sort and compare data, and cannot be changed after data warehouse creation. The default collation is SQL_Latin1_General_CP1_CI_AS. [Learn more](#)

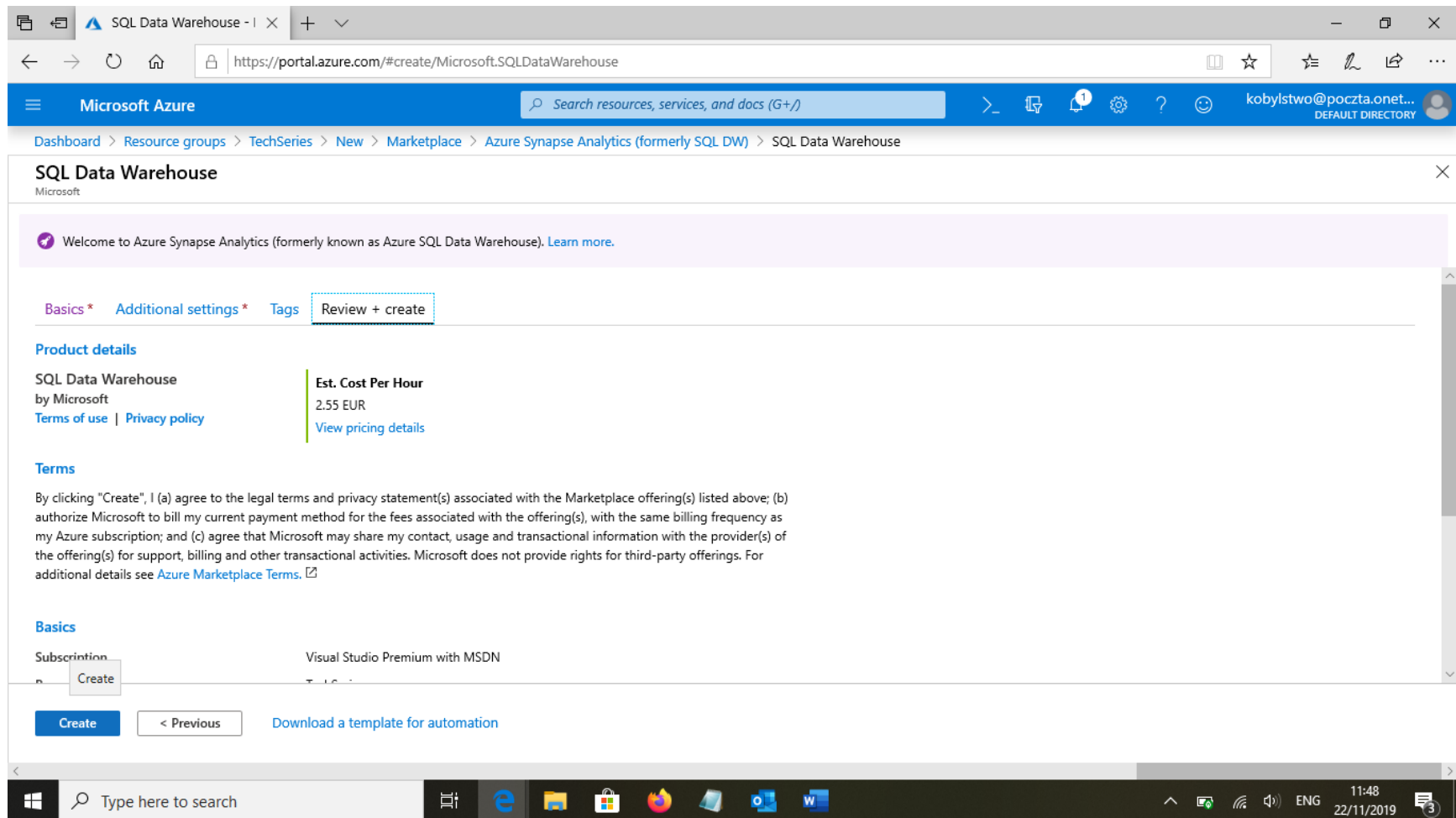
Collation * ⓘ

SQL_Latin1_General_CP1_CI_AS

Review + create < Previous Next : Tags >

Type here to search

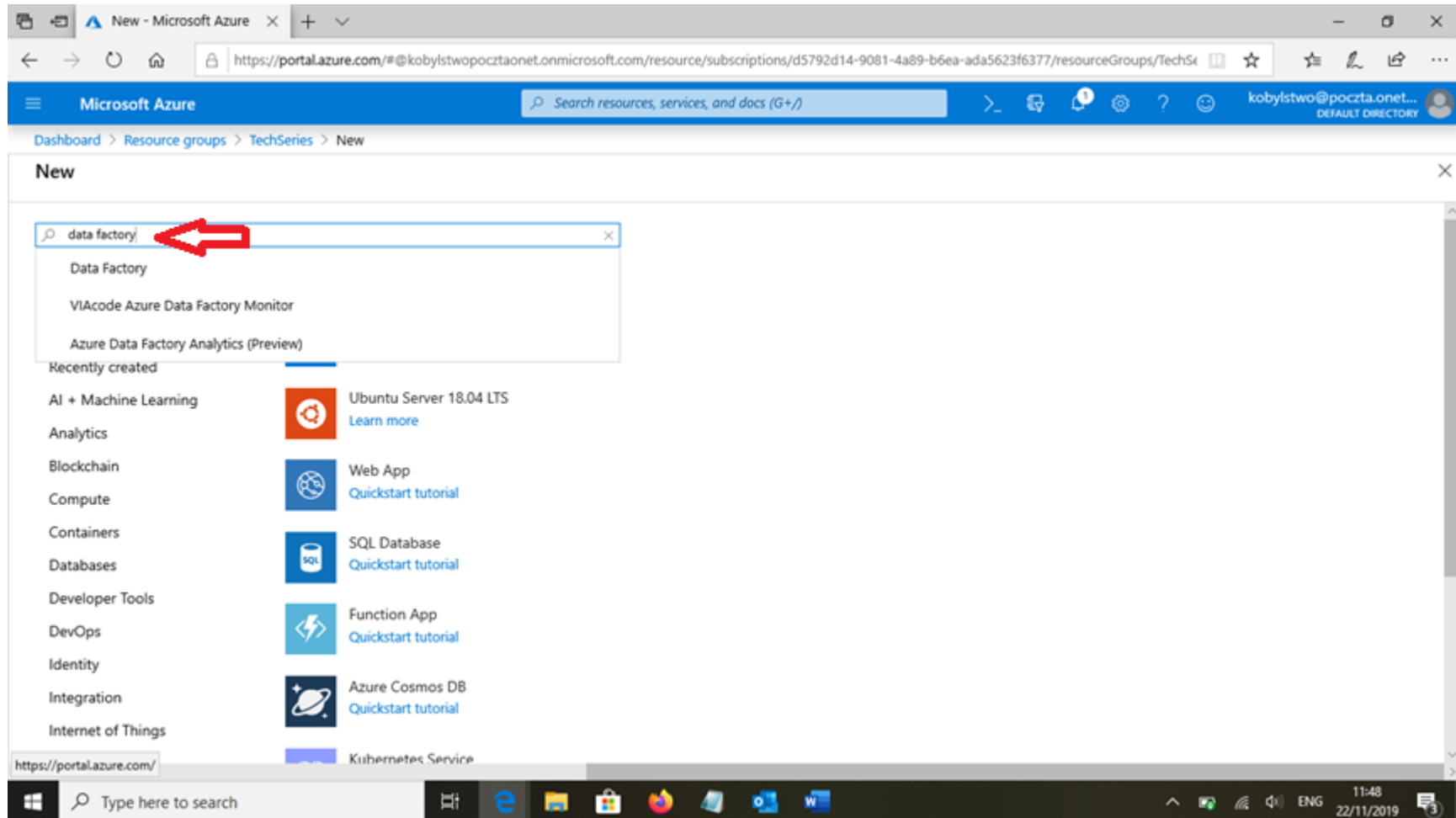
11:47 22/11/2019 ENG



Alternatively you can use the following CLI command:

```
az sql dw create --name techseriesdwh --resource-group TechSeries --server techseriesserver --service-objective DW200c --zone-redundant false
```

Do not wait for creating Azure Synapse Analytics (a.k.a. SQL Data Warehouse) instance, initialize your Data Factory environment. Go to the TechSeries resource group and add a new service.



Don't forget to disable GIT. We will not use it.

The screenshot shows the 'New data factory' page in the Microsoft Azure portal. The browser address bar shows the URL <https://portal.azure.com/#create/Microsoft.DataFactory>. The breadcrumb navigation path is Dashboard > Resource groups > TechSeries > New > Marketplace > Data Factory > New data factory. The form fields are as follows:

- Name ***: techseries (with a green checkmark)
- Version ⓘ**: V2
- Subscription ***: Visual Studio Premium with MSDN
- Resource Group ***: TechSeries (with a 'Create new' link below it)
- Location ⓘ**: West Europe
- Enable GIT ⓘ**: ☐ (highlighted with a red arrow)

A blue 'Create' button is located at the bottom left of the form.

After it has been created you may check the interface by following the url: <https://adf.azure.com/>

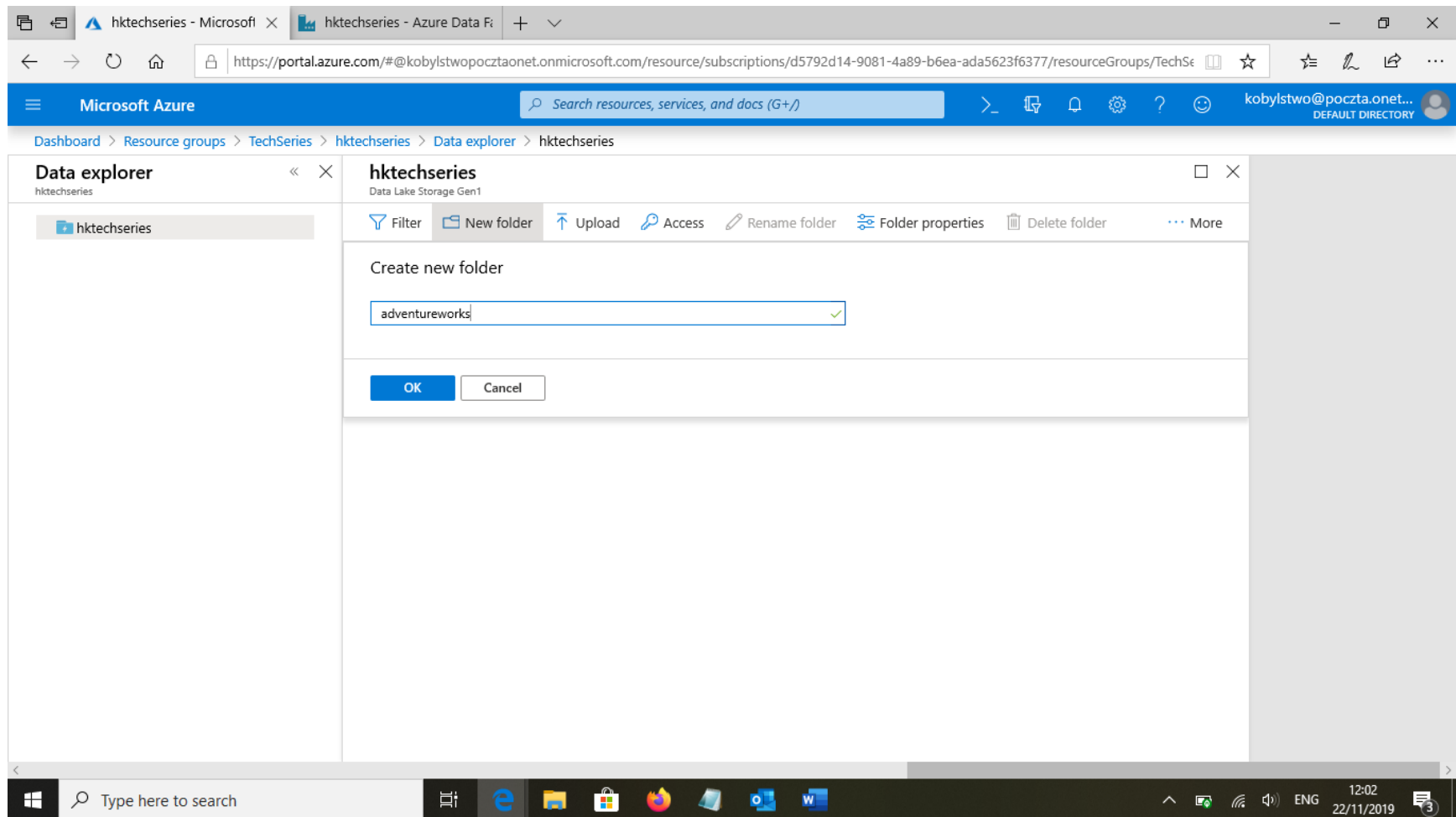
Now, create a folder in your instance of Data Lake. You may use “Data explorer” blade for that.

The screenshot displays the Microsoft Azure portal interface. The browser address bar shows the URL: <https://portal.azure.com/#@kobylistwo@poczta.onmicrosoft.com/resource/subscriptions/d5792d14-9081-4a89-b6ea-ada5623f6377/resourceGroups/TechSeries>. The page title is "hktechseries - Microsoft Azure". The breadcrumb navigation path is "Dashboard > Resource groups > TechSeries > hktechseries".

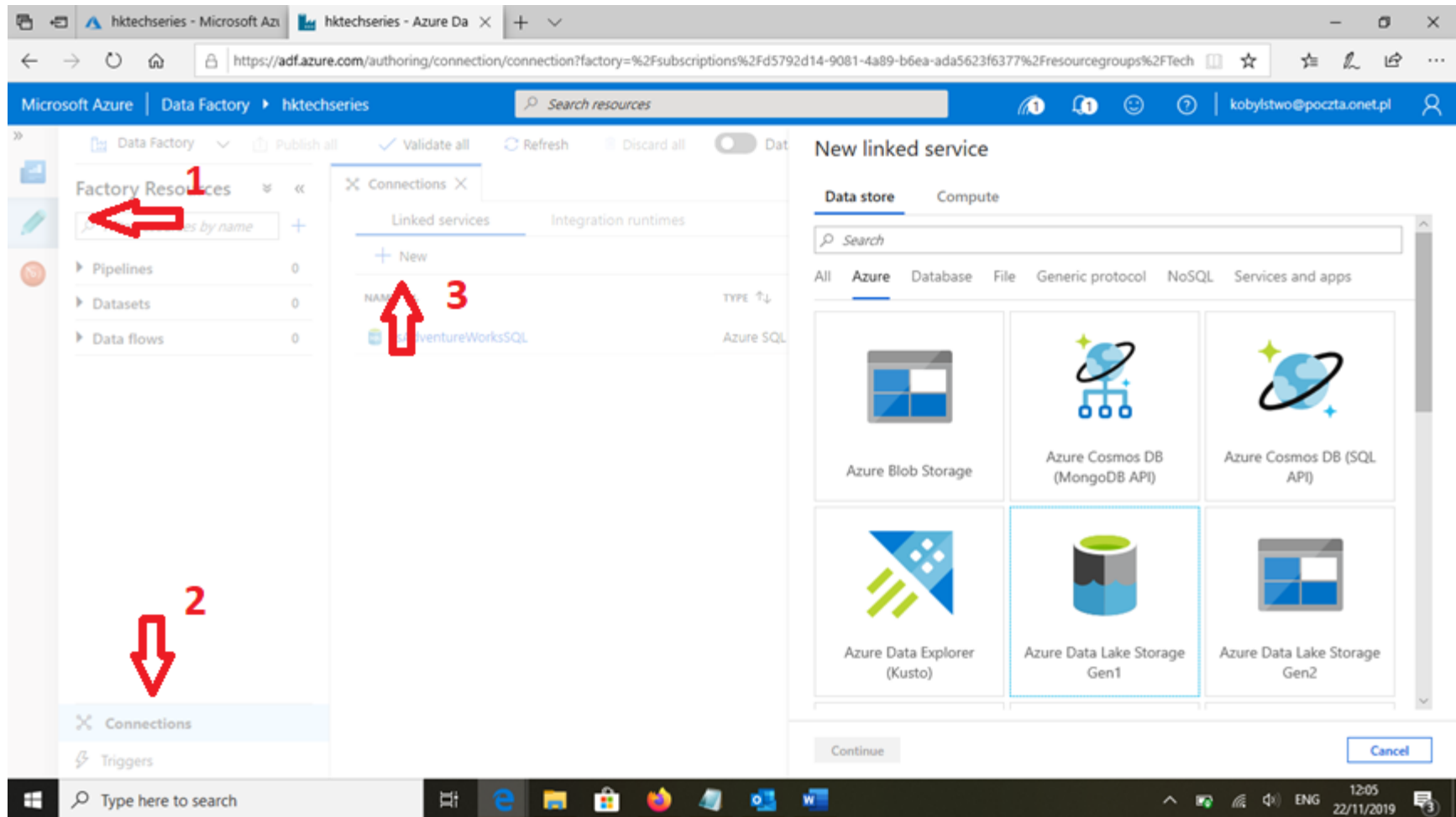
The left-hand navigation pane shows the "Data Lake Storage Gen1" section expanded, with the "Data explorer" option selected. The main content area displays the "Data explorer" blade for the "hktechseries" Data Lake Storage Gen1 instance. The blade includes a search bar, a "Data explorer" tab, and a "Delete" button. A purple banner at the top of the blade states: "Azure Data Lake Storage Gen2 is now GA. Click here to read more about how you can upgrade and speed up your time to insight. →".

The "Essentials" section shows the instance name "hktechseries", the pricing model "Pay-as-You-Go", the start date "November 2019", and the "Estimated cost to date" of "USD 0.00". A link for "Pricing tier info" is provided. Below this, the "Total storage utilization" section is visible, with a filter set to "7 days" and a date range of "11/15/2019-11/22/2019". The utilization chart shows "Total storage utilization", "Data at rest", and "Metadata".

The Windows taskbar at the bottom shows the search bar with the text "Type here to search", the taskbar icons, and the system tray with the date "22/11/2019" and time "12:02".



Go to Data Factory. Create linked services in ADF for Data Lake (name it as IsDataLake) and Azure Synapse Analytics (name it as IsSynapse).



Microsoft Azure | Data Factory | hktechseries

Search resources

Factory Resources

Filter resources by name

Pipelines 0

Datasets 0

Data flows 0

Connections

Triggers

Connections

Linked services

Integration runtimes

+ New

NAME	TYPE
IsAdventureWorksSQL	Azure SQL

New linked service (Azure Data Lake Storage Gen1)

Connect via integration runtime *

AutoResolveIntegrationRuntime

Data Lake Store selection method

☒ From Azure subscription ☐ Enter manually

Azure subscription

Visual Studio Premium with MSDN (d5792d14-9081-4a89-b6ea-ada5623f6377)

Data Lake Store account name *

hktechseries

Tenant *

44fe5703-1610-41b9-bc47-7f2cad8a744d

Authentication type *

Managed Identity

Managed identity application ID: d8970270-205b-4349-ae2d-adaa758424b1
Grant Data Factory service managed identity access to your Azure Data Lake Storage Gen1. [Details](#)

Annotations

+ New

Advanced

Create Back Test connection Cancel

That object is defined by the following JSON code.

```
{
  "name": "lsDataLake",
  "type": "Microsoft.DataFactory/factories/linkedservices",
  "properties": {
    "annotations": [],

```



```
    "type": "AzureDataLakeStore",
    "typeProperties": {
      "dataLakeStoreUri": "https://hktechseries.azuredatalakestore.net/webhdfs/v1",
      "tenant": "44fe5703-1610-41b9-bc47-7f2cad8a744d",
      "subscriptionId": "d5792d14-9081-4a89-b6ea-ada5623f6377",
      "resourceGroupName": "TechSeries"
    }
  }
}
```

Now, a linked service for Synapse.

The screenshot shows the Microsoft Azure Data Factory portal interface. The left sidebar displays 'Factory Resources' with categories: Pipelines (0), Datasets (0), and Data flows (0). The main area is titled 'Connections' and shows a list of linked services: 'IsAdventureWorksSQL' (Azure SQL), 'IsDataLake' (Azure Data Lake), and 'IsSynapse' (Azure Synapse Analytics). A 'New' button is visible above the list. On the right, a modal window titled 'New linked service (Azure Synapse Analytics (formerly SQL...))' is open. It contains the following fields: 'Azure subscription' (Visual Studio Premium with MSDN), 'Server name' (hktechseriesserver), 'Database name' (techseriesdwh), 'Authentication type' (SQL authentication), and 'User name' (damian). There are tabs for 'Password' and 'Azure Key Vault'. The 'Password' tab is active, showing a password field with masked characters and a 'Test connection' button. At the bottom of the modal are 'Create', 'Back', and 'Cancel' buttons. The Windows taskbar at the bottom shows the time as 12:13 on 22/11/2019.

Go to your Azure SQL DB and create following views (use any client you want, Azure Data Studio, SQL Management Studio, etc.). It will create some analytics views on our OLTP database.

CREATE VIEW vSales

```

AS
SELECT oh.[OrderDate]
      ,oh.[SalesOrderNumber]
      ,oh.[CustomerID]
      ,od.ProductID
      ,od.OrderQty
      ,od.UnitPrice
FROM [SalesLT].[SalesOrderHeader] oh
     INNER JOIN [SalesLT].[SalesOrderDetail] od
        ON oh.SalesOrderID = od.SalesOrderID;

CREATE VIEW vProduct
AS
SELECT p.[ProductID]
      ,p.[Name] AS Product
      ,pc.[Name] AS Category
      ,p.[ProductNumber]
      ,p.[Color]
      ,p.[StandardCost]
FROM [SalesLT].[Product] p
LEFT JOIN [SalesLT].[ProductCategory] pc
    ON p.ProductCategoryID = pc.ProductCategoryID;

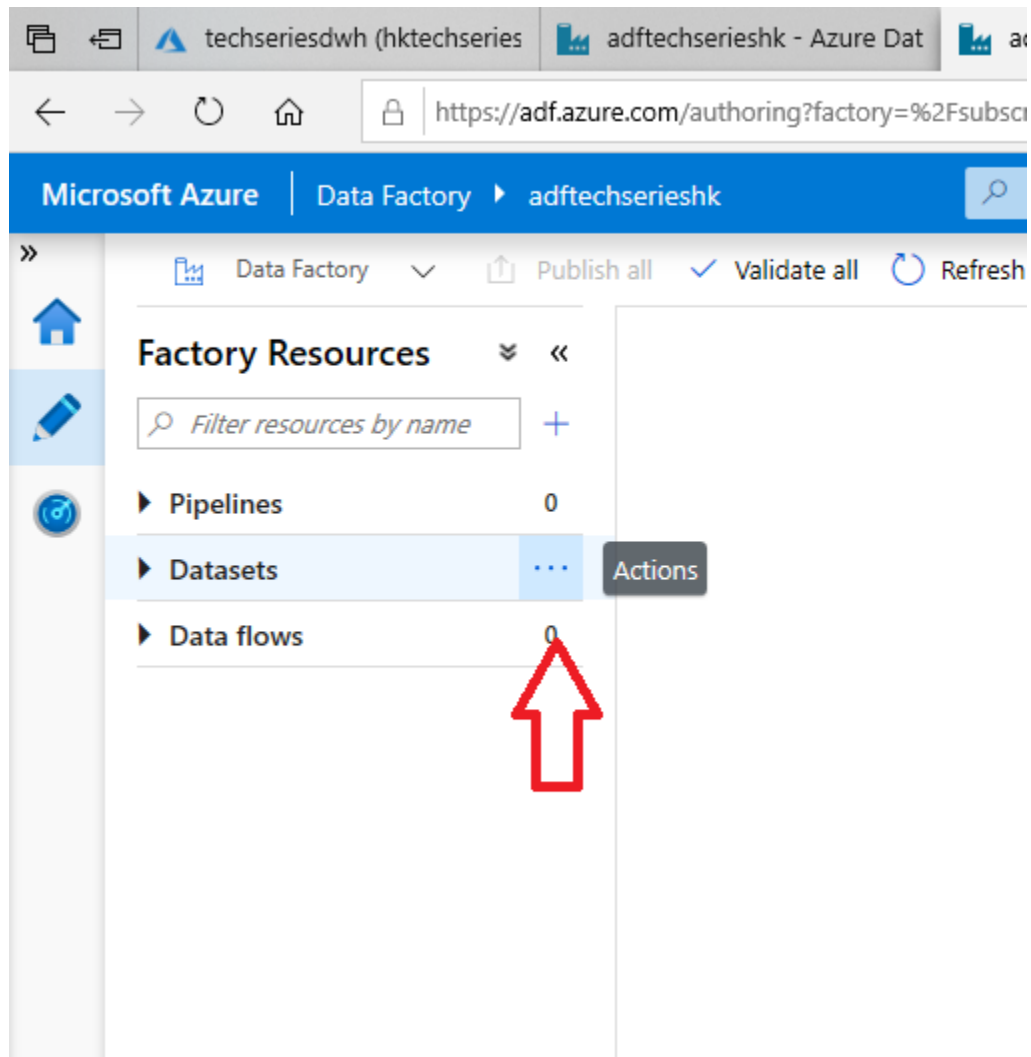
CREATE VIEW vCustomer
AS
SELECT [CustomerID]
      ,[FirstName] + ' ' + [MiddleName] AS Customer
FROM [SalesLT].[Customer];

```

Go to Azure Data Factory, click on Author & Monitor. Alternatively you can go to <https://adf.azure.com/>

Create a connector (called here as Linked Service) to Azure SQL DB. We will be loading data from it. Just follow similar steps as for Data Lake and Synapse Analytics services.

Create source datasets – for Azure SQL DB views created above.



techseriesdwh (hktechseries) adftechserieshk - Azure Dat adftechserieshk - Azure

← → ↻ 🏠 🔒 https://adf.azure.com/authoring?factory=%2Fsubscriptions%2Fd5792d14

Microsoft Azure | Data Factory ▶ adftechserieshk 🔍 Search resources

» Data Factory ▼ Publish all ✓ Validate all ↻ Refresh 🗑 Discard all

🏠

✎

🔗


Factory Resources

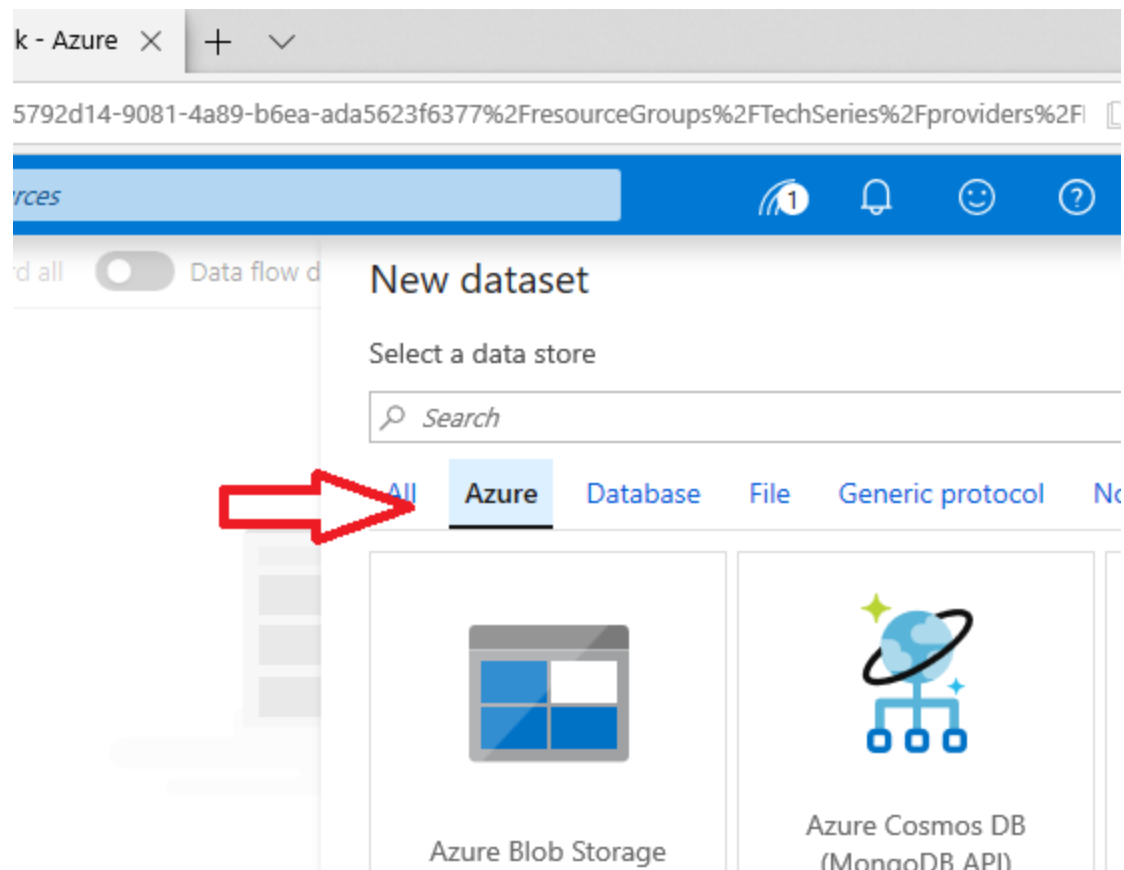
🔍 Filter resources by name +

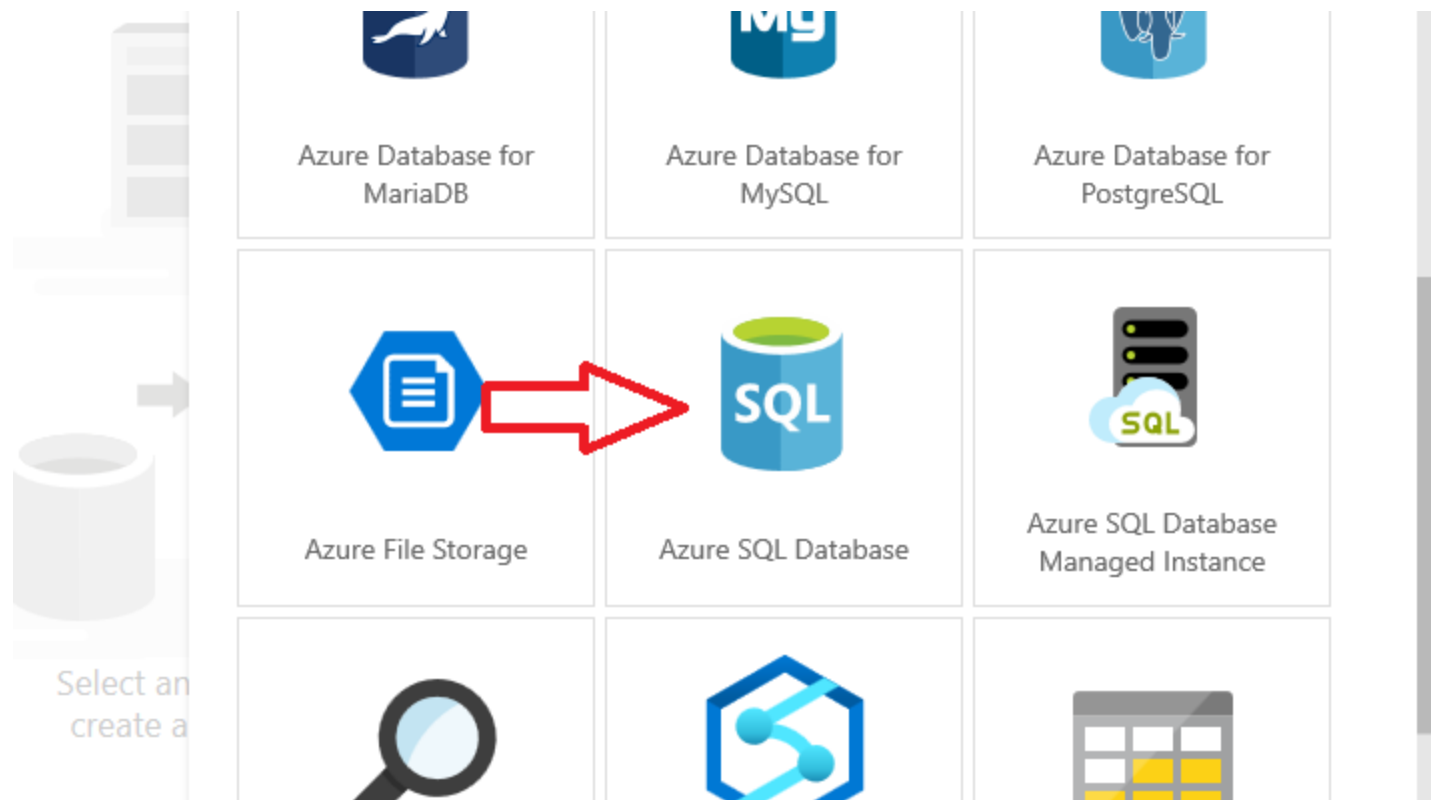
- ▶ Pipelines 0
- ▶ Datasets 0
- ▶ Data flows

📊 New dataset

📁 New folder







techseriesdwh (hktechseries) adftechserieshk - Azure Dat adftechserieshk - Azure

https://adf.azure.com/authoring?factory=%2Fsubscriptions%2Fd5792d14-9081-4a89-b6ea-ada5623f6377%2FresourceGroups%2FTechSeries%2Fproviders%2F...

Microsoft Azure | Data Factory | adftechserieshk Search resources

» Data Factory Publish all Validate all Refresh Discard all Data flow d

Factory Resources

Filter resources by name

Pipelines 0

Datasets 0

Data flows 0

Connections

Triggers

Select an existing connection or create a new one

Set properties

Name
srcSales

Linked service *
IsAdventureWorks

Edit connection

Table name
dbo.vSales

Edit

Import schema
☒ From connection/store ☐ None

OK Back Cancel

You don't have to do anything else with it. Just publish it.

The screenshot shows the Microsoft Azure Data Factory portal interface. At the top, the browser address bar displays the URL: `https://adf.azure.com/authoring/dataset/srcSales?factory=%2Fsubscriptions%2Fd5792d14-9081-4a89-b6ea-ada5623f6377%2FresourceGroups%2FTechSeries9`. The page header includes the Microsoft Azure logo, the text "Data Factory", and the name of the data factory "adftechserieshk". A search bar labeled "Search resources" is also present. Below the header, a navigation bar contains several icons and labels: "Data Factory", "Publish all" (with a yellow notification badge), "Validate all", "Refresh", "Discard all", "Data flow debug" (with a toggle switch), and "ARM template".

The main content area is divided into two sections. On the left, the "Factory Resources" sidebar lists the following items:

- Pipelines: 0
- Datasets: 1
- * srcSales (highlighted)
- Data flows: 0

Below this list are links for "Connections" and "Triggers". On the right, the "srcSales" dataset is selected, showing its configuration. The dataset is identified as an "Azure SQL Database" with the name "srcSales". The configuration is displayed in the "General" tab, which includes fields for "Name" (set to "srcSales"), "Description", and "Annotations" (with a "+ New" button). The "Connection", "Schema", and "Parameters" tabs are also visible.

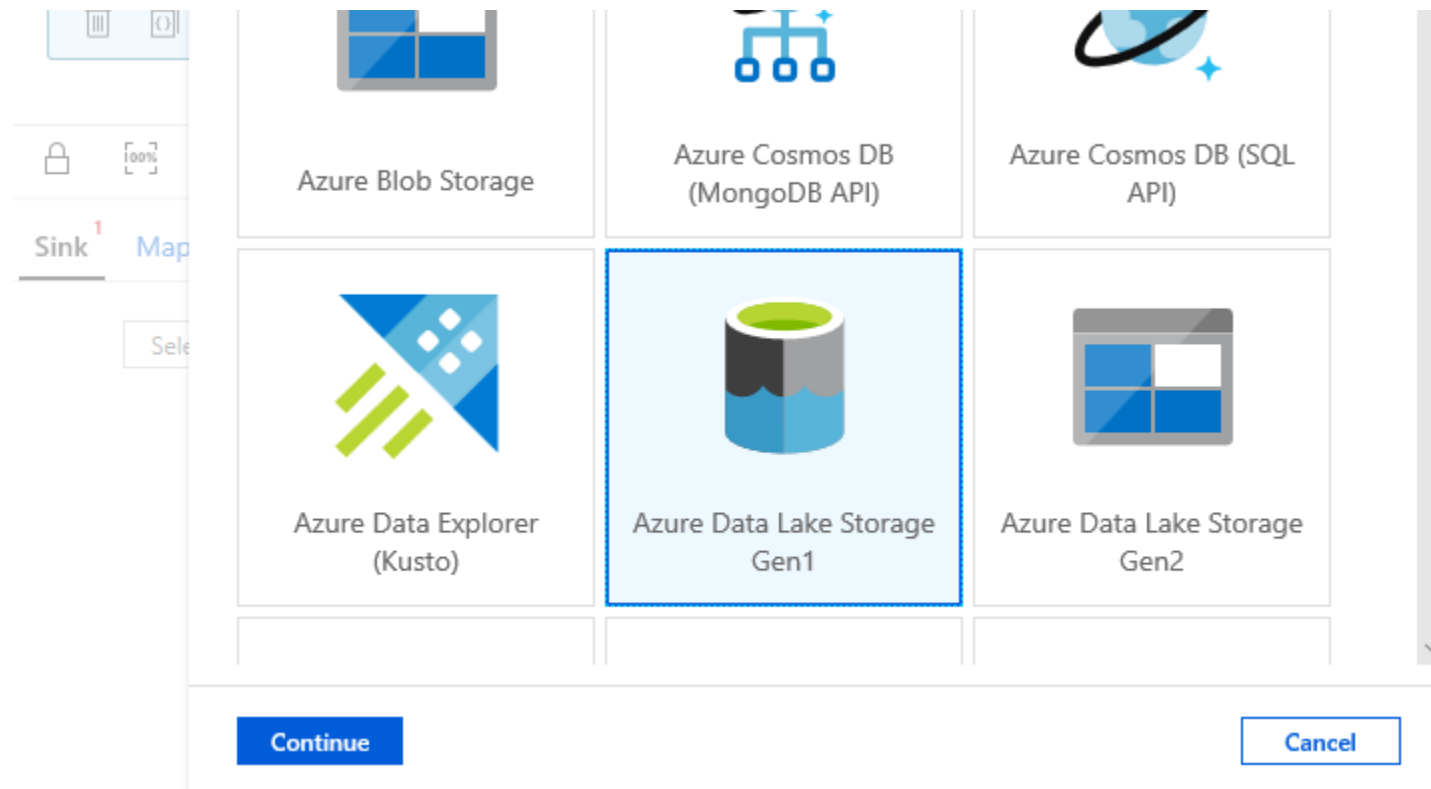
Now, create datasets for the other two views – Customer and Product. Once you have got them published create our first pipeline. Name it properly and drop the Copy data component on the canvas.

The screenshot displays the Microsoft Azure Data Factory portal interface. The top navigation bar shows 'soft Azure | Data Factory > adftechserieshk' with a search bar and notification icons. Below the navigation bar, a toolbar contains buttons for 'Data Factory', 'Publish all' (with a yellow badge '1'), 'Validate all', 'Refresh', 'Discard all', 'Data flow debug' (toggle), and 'ARM template'. The left sidebar, titled 'Factory Resources', lists 'Pipelines' (1), 'Datasets' (3), and 'Data flows' (0). Under 'Datasets', 'srcCustomers', 'srcProduct', and 'srcSales' are listed. The 'Pipelines' section shows '* pipe-copySales' selected. The main canvas area is titled 'pipe-copySales *' and features a toolbar with 'Save as template', 'Validate', 'Debug', and 'Add trigger'. Below this is a search bar for activities and a list of activity categories: 'Move & transform' (containing 'Copy data' and 'Data flow'), 'Azure Data Explorer', 'Azure Function', 'Batch Service', 'Data Lake Analytics', 'Databricks', and 'General'. A red arrow points to the 'Copy data' activity. The right pane shows the 'General' tab for the pipeline configuration, with fields for 'Name' (containing 'pipe-copySales') and 'Description'. A red arrow points to the 'Name' field.

Rename the Copy data component, go to the Source tab and select the Azure SQL DB Sales dataset from the dropdown.

The screenshot displays the Azure Data Factory 'Copy data' component configuration interface. On the left, a sidebar lists various integration runtime options under the 'Move & transform' category, including 'Copy data', 'Data flow', 'Azure Data Explorer', 'Azure Function', 'Batch Service', 'Data Lake Analytics', 'Databricks', 'General', 'HDInsight', 'Iteration & conditionals', and 'Machine Learning'. The 'Copy data' component is highlighted in the main workspace. Below the component, a toolbar contains icons for search, zoom, and other actions. The 'Source' tab is selected, showing the 'Source dataset' dropdown set to 'srcSales'. The 'Use query' section has three radio buttons: 'Table' (selected), 'Query', and 'Stored Procedure'. A red arrow points to the 'Table' radio button. Another red arrow points to the 'General' tab in the top navigation bar. The 'Preview data' button is visible on the right.

Next, go to the Sink tab and create a brand new Data Lake Gen1 dataset for Sales. Use CSV data format.



Select format

Choose the format type of your data



Parquet



DelimitedText



Json



Avro



ORC



Binary

Name it as textSales, select your Data Lake linked service and hit Browse to select folder. You are going to see the access error.

Choose a file or folder





Loading error

Access to <https://hktechseries.azuredatalakestore.net/webhdfs/v1/> is denied. Make sure the ACL and firewall rule is correctly configured in the Azure Data Lake Store account. Service request id: 056572a9-641e-4c1c-b615-fabf1bcac70a Response details: {"RemoteException": {"exception": "AccessControlException", "message": "LISTSTATUS failed with error 0x83090aa2 (Forbidden. ACL verification failed. Either the resource does not exist or the user is not authorized to perform the requested operation.). [056572a9-641e-4c1c-b615-fabf1bcac70a] failed with error 0x83090aa2 (Forbidden. ACL verification failed. Either the resource does not exist or the user is not authorized to perform the requested operation.). [056572a9-641e-4c1c-b615-fabf1bcac70a][2019-11-24T06:43:46.3295417-08:00]", "javaClassName": "org.apache.hadoop.security.AccessControlException"}} The remote

It means one thing – ADF has no rights to access ADLS resources. We are going to fix it the easy way in this tutorial but in your regular environment you should follow a secure path with stronger security bias.

Grant access to DL Gen1 folder. Got to the ADLS resource, use Data explorer and hit Access button in the root folder and tick those three boxes in the “Everyone else” section. Remember – this is not secure.

	00000000-0000-0000-0000-000000000000	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Assigned permissions				
No entries.				
Everyone else				
	Users not covered above will be limited by these permissions	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>



Hit “Save” button after ticking these boxes off.

Do the same thing for our ‘adventureworks” folder and go back to the ADF interface. Select the ‘adventureworks’ folder and press OK button.

Choose a file or folder

↑ 📁 >



📁 adventureworks

📁 adventureworks2

📁 catalog

📁 system

📁 techseries

Enter 'Sales.csv' file name and select 'None' from the Import schema section as we do not have any structure for that file. Do not forget about setting first row as a column header.

Set properties

Name

textSales

Linked service *

lsDataLake

[Edit connection](#)

File path

adventureworks



/ Sales.csv

[Browse](#)



First row as header



Import schema



From connection/store



From sample file



None

► Advanced



Set properties

Name

textSales

Linked service *

IsDataLake

[Edit connection](#)

File path

adventureworks



Sales.csv

[Browse](#)



First row as header

☐

Import schema

☐

From connection/store

☐

From sample file

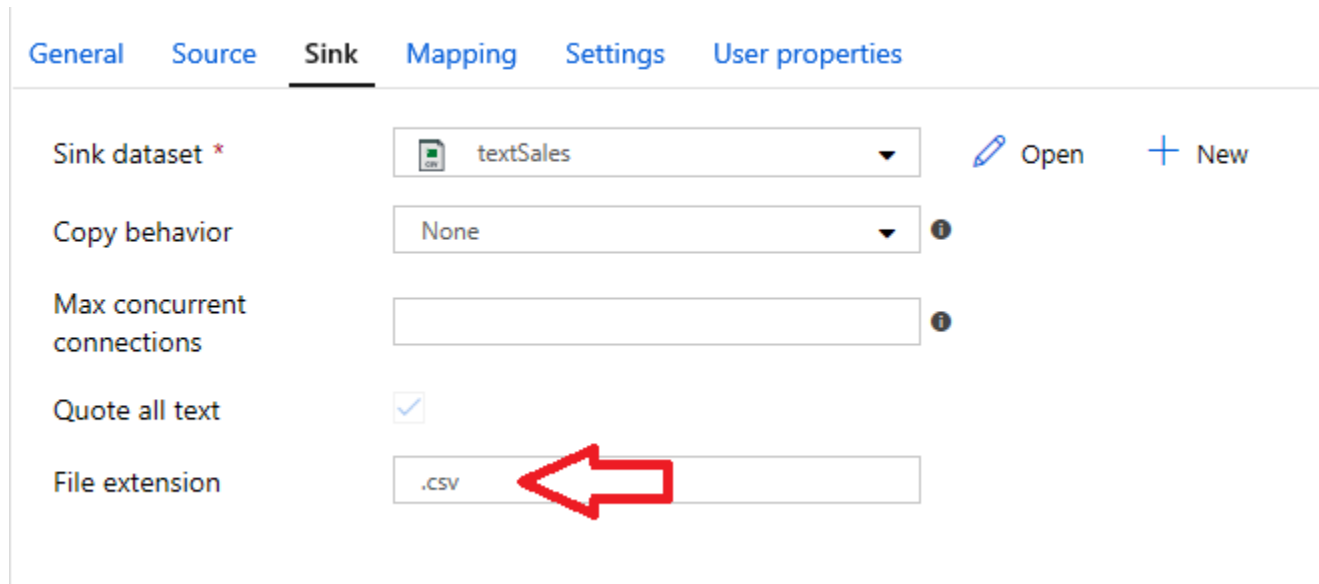
☒

None

► Advanced



After creating your new sink dataset change File extension to '.csv' just to keep it consistent with the previous screen.



The screenshot shows the 'Sink' configuration tab with the following settings:

- Sink dataset ***: textSales (with 'Open' and 'New' buttons)
- Copy behavior**: None (with an information icon)
- Max concurrent connections**: (empty field with an information icon)
- Quote all text**: ☒
- File extension**: .csv (highlighted by a red arrow)

Your pipeline is ready and you may publish it. At the moment you need to add next two Copy data components to your pipeline. You do not have to set any mappings as your destination objects (files on Data Lake Storage instance are not existing as of this moment). Do not forget to publish your finished pipeline. To execute it you may either press the Debug button or Add trigger/Trigger now option. The Debug button enables testing your pipeline before it is published.

Save as template
✓ Validate
▶ Debug
⚙ Add trigger
📄

Copy data

copySales

Trigger now
New/Edit

copyCustomer

Copy data

copyProduct

🗑 📄 📁

🔍 + - 🔒 100% 📐 📏 🔄 📶

Your execution results will appear at the bottom of the screen.

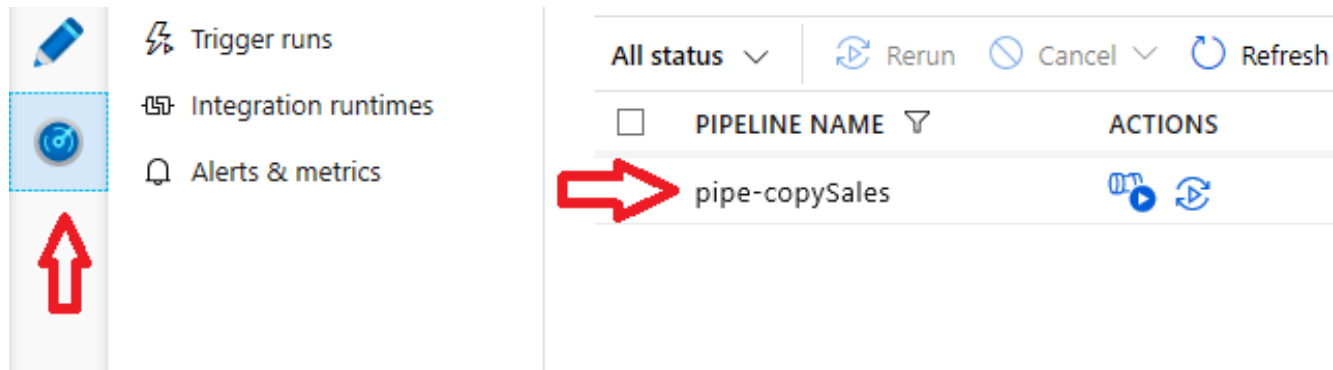
GeneralParametersVariablesOutput



Pipeline run ID: de6017c3-6596-457e-88fc-911de2ce723c [👤] 🔄 ⓘ

NAME	TYPE	RUN START	DURATION	STATUS
copyProduct	Copy	2019-11-24T16:42:04.589	00:00:02	🔄 In Progress
copyCustomer	Copy	2019-11-24T16:42:04.557	00:00:02	🔄 In Progress
copySales	Copy	2019-11-24T16:42:04.542	00:00:02	🔄 In Progress

<

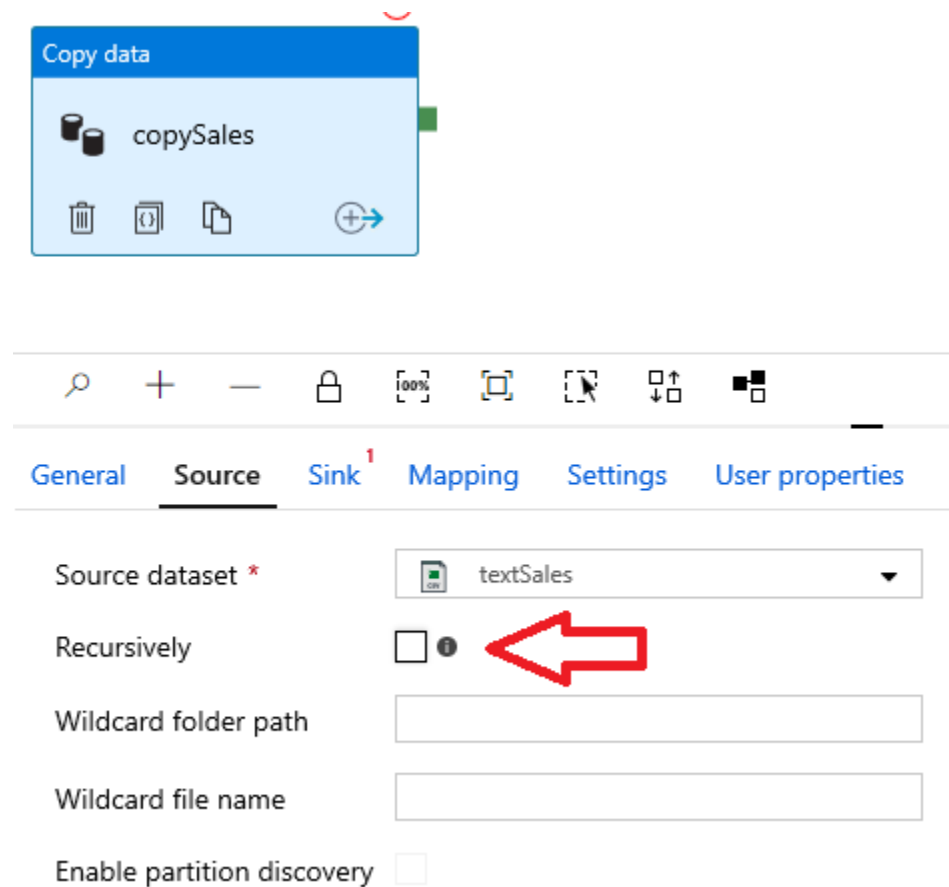
If you have executed the pipeline through a trigger you may observe its progress via the ADF monitoring section.



PIPELINE NAME	ACTIONS
pipe-copySales	 

At this moment you should be ready to create a new pipeline which copies data from Data Lake to Synapse Analytics. Follow similar steps, remember that your source dataset names start with 'text*' and destination ones sit in the Synapse Analytics service. Following screens may be of some help here since your destination tables do not exist yet.

Do not forget to uncheck the 'Recursively' checkbox.



When setting properties of the new dataset pay attention to the following options:

- start its name with 'dst' (just to be consistent with the naming convention)
- check the 'Edit' box in order to enter non-existing table name
- put 'dbo' and 'Sales' in respective text boxes
- select 'None' option in the 'Import schema' section

Set properties

Name 1
dstSales

Linked service *
IsSynapse

[Edit connection](#) 3

Table name 4
dbo . Sales



☒ Edit

Import schema
☐ From connection/store ☒ None 5

► Advanced 2

Once the dataset is created you need to add few Sink options.

General Source **Sink** Mapping Settings User properties

Sink dataset * dstSales  Open  New







Allow PolyBase ☐  

Table option ☐ None ☒ Auto create table  

Pre-copy script

IF OBJECT_ID('[dbo].[Sales]') IS NOT NULL
TRUNCATE TABLE dbo.Sales;

Write batch timeout

Follow similar steps for the other two tables (Customer and Product). Publish and trigger your new pipeline and you should be able to build your first report i.e. in Power BI.