

JOURNEY SCRAPBOOK

Technical Training - Week 3

04-Sep-2023 to 08-Sep-2023 (Days 5 – 9)

YUVA SAHITH VARMA SANGARAJU

BATCH - 8

Day 5 – Azure Data Storage

- ▶ Storage Accounts and types
- ▶ Benefits of using Azure to store data
- ▶ Automated backup, Data replication, Encryption
- ▶ Azure key vault and Storage tiers
- ▶ Comparing Azure to On-premises storage
- ▶ Storage Account setup and configuration in Azure
- ▶ Containers, Queues, File Shares and Tables
- ▶ Azure Data Lake Storage (ADLS) - Gen II
- ▶ Hierarchical namespace in ADLS
- ▶ Blob Storage vs ADLS
- ▶ Processing Big Data with ADLS

Day 5 continued...

- ▶ Ingestion, Store, Prep & Train, Model & Serve
- ▶ Big Data use cases – Modern DW
- ▶ Advanced and Real time data analytics
- ▶ Creating Storage Account and Container in Blob
- ▶ Uploading file in container
- ▶ Creating Azure Cosmos DB account
- ▶ Performing labs for Azure Storage

Home > sahithstore321_1693803742904 | Overview >

 sahithestore321



Storage account



Search

 Upload  Open in Explorer  Delete  Move  Refresh  Open in mobile  CLI / PS  Feedback

 Overview

Activity log

Tags

Diagnose and solve problems

Access Control (IAM)

➡ Data migration

Events

Storage browser

Gianni M.

Data storage

Containers

File shares

Questions

Security + networking

[Properties](#) [Monitoring](#) [Capabilities \(7\)](#) [Recommendations \(0\)](#) [Tutorials](#) [Tools + SDKs](#)

Blob service

Hierarchical namespace	Disabled
Default access tier	Hot
Blob anonymous access	Disabled
Blob soft delete	Enabled (7 days)
Container soft delete	Enabled (7 days)
Versioning	Disabled

Security

Require secure transfer for REST API operations	Enabled
Storage account key access	Enabled
Minimum TLS version	Version 1.2
Infrastructure encryption	Disabled

sahithstore321 | Shared access signature

X

Storage account

 Search

Give feedback

Tables

Security + networking

Networking

Front Door and CDN

Access keys

Shared access signature

Encryption

Microsoft Defender for Cloud

Data management

Redundancy

Data protection

Object replication

Blob inventory

Static website

Lifecycle management

Allowed services

- Blob
- File
- Queue
- Table

Allowed resource types

- Service
- Container
- Object

Allowed permissions

- Read
- Write
- Delete
- List
- Add
- Create
- Update
- Process
- Immutable storage
- Permanent delete

Blob versioning permissions

- Enables deletion of versions

Allowed blob index permissions

- Read/Write
- Filter

Start and expiry date/time

Start: 09/04/2023

11:07:42 AM

End: 09/04/2023

7:07:42 PM

(UTC+05:30) Chennai, Kolkata, Mumbai, New Delhi

Allowed IP addresses

For example, 168.1.5.65 or 168.1.5.65-168.1.5.70



sahith

Container

...

X

 Search

Upload

Change access level

Refresh

Delete

Change tier

Acquire lease

Break lease

View snapshots

...

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Shared access tokens

Access policy

Properties

Metadata

Authentication method: Access key ([Switch to Azure AD User Account](#))**Location:** sahith

Search blobs by prefix (case-sensitive)

 Show deleted blobs

Add filter

Name	Modified	Access tier	Archive status	Blob type
ER for Employees.png	9/4/2023, 11:28:50 AM	Hot (Inferred)		Block blob

demoshare

SMB File share

...



Search

[Connect](#) [Upload](#) [Refresh](#) [Add directory](#) [Delete share](#) [Change tier](#) [Edit quota](#) [Give feedback](#)

Overview

[Diagnose and solve problems](#)[Access Control \(IAM\)](#)[Browse](#)

Operations

[Snapshots](#)[Backup](#)

ⓘ Enable Backup for file share "demoshare" to protect your data. [Learn more](#)

[JSON View](#)

Essentials

Storage account	:	sahithstore321	Share URL	:	https://sahithstore321.file.core.windows.net/demoshare
Resource group (move)	:	labrg	Redundancy	:	Geo-redundant storage (GRS)
Location	:	East US	Configuration modified	:	9/4/2023, 11:30:13 AM
Primary/Secondary location : Primary: East US, Secondary: West US					
Subscription (move)	:	npunext-1673504988896			
Subscription ID	:	2190116a-14d8-4b0c-8821-f97ab2e43421			

Properties

Capabilities (2)

Tutorials

Size

Maximum capacity	5 TiB
Used capacity	0 B
Tier	Transaction optimized

Feature status

Soft delete	(i)	7 days
Large file shares		Disabled

Performance

1000

Active Directory

Directory service	(i)	Not configured
-------------------	---------------------	--------------------------------

Home > sahithstore321

sahithstore321 | Alerts

Storage account

Search



View as ti

Resource sharing (CORS)

Advisor recommendations

Endpoints

Locks

Monitoring

Insights

Alerts

Metrics

Workbooks

Diagnostic settings

Logs

Monitoring (classic)

Metrics (classic)

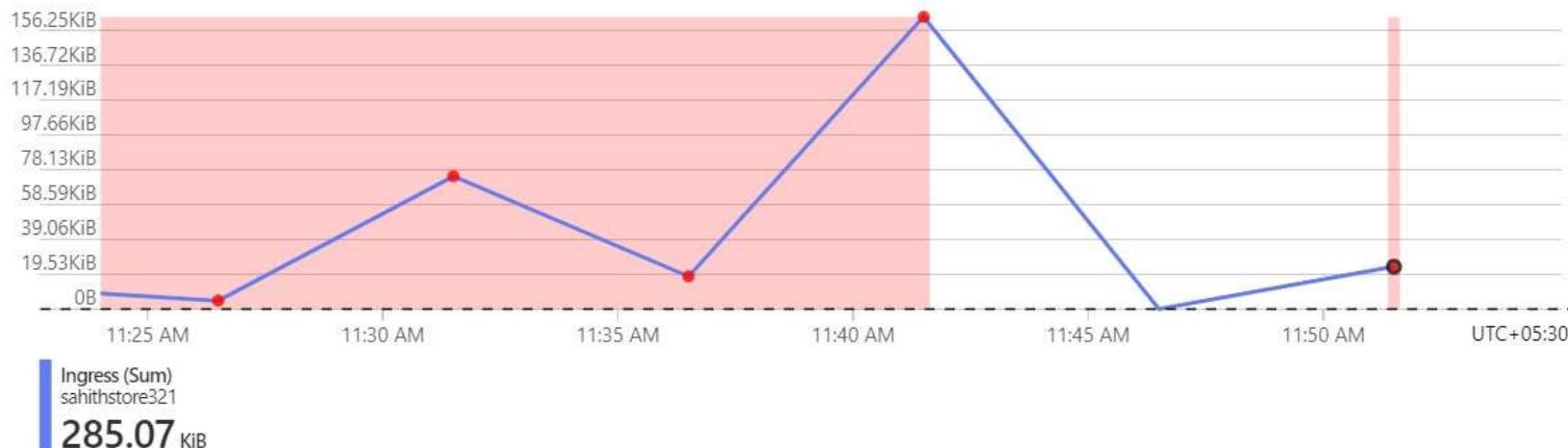
Diagnostic settings (classic)

Usage (classic)

myruleingressandegress

Alert details

Why did this alert fire?



Showing 1 - 1

sahitdhatalake | Quick start ☆ ...

Azure Cosmos DB account

Search

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Cost Management

Quick start

1 Add a container

The "Items" container has been created with a capacity of 400 Request Units per second (RU/s)throughput. To see your container, go to Data Explorer and find the ToDoList database.

2 Download and run your Node.js app

We created a sample Node.js app connected to your "Items" collection. Download and extract the app.

[Download](#)

Run **npm install** and **npm start** and navigate to <http://localhost:3000>.

3 Work with data

...

PowerShell | ? { }

— □ ×

Type "az" to use Azure CLI

Type "help" to learn about Cloud Shell

Storage fileshare subscription 2190116a-14d8-4b0c-8821-f97ab2e43421 is not registered to Microsoft.CloudShell Namespace. Please follow these instructions "<https://aka.ms/RegisterCloudShell>" to register. In future, unregistered subscriptions will have restricted access to CloudShell service.

MOTD: Azure Cloud Shell now includes Predictive IntelliSense! Learn more: <https://aka.ms/CloudShell/IntelliSense>

VERBOSE: Authenticating to Azure ...

VERBOSE: Building your Azure drive ...

PS /home/shellunext> npm install

sahithdatalake | Data Explorer

Azure Cosmos DB account

[Add Azure Cognitive Search](#)[Add Azure Function](#)

Containers

[Browse](#)[Scale](#)[Settings](#)[Document Explorer](#)[Query Explorer](#)[Script Explorer](#)

Monitoring

[Insights](#)[Alerts](#)[Metrics](#)[Logs](#)[Diagnostic settings](#)[Metrics \(Classic\)](#)

NoSQL API

Home

Items - Items

SELECT * FROM c

Edit Filter

id	/cate...
1	personal
2	work

[Load more](#)

```
1  {
2    "id": "1",
3    "category": "personal",
4    "name": "groceries",
5    "description": "Pick an apples and strawberries.",
6    "IsComplete": false,
7    "_rid": "cA1SAPn9D3kBAAAAAAA==",
8    "_self": "dbs/cA1SAA==/colls/cA1SAPn9D3k=/docs/cA1SAPn9D3kBAAAAAAA==/",
9    "_etag": "\"86024cd1-0000-0100-0000-64f5a1160000\"",
10   "_attachments": "attachments/",
11   "_ts": 1693819158
12 }
```

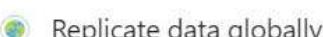


Home > Microsoft.Azure.CosmosDB-20230904142305 | Overview > sahithdatalake

sahithdatalake | Keys



Azure Cosmos DB account



Refresh

URI

https://sahithdatalake.documents.azure.com:443/



Read-write Keys

Read-only Keys

PRIMARY KEY



PowerShell |

— □ ×

Replacing item:

Anderson.1

Querying container:

Items

```
Query returned [{"firstName": "Henriette Thaulow", "gender": "female", "grade": 6, "pets": [{"givenName": "Fluffy"}]}]
```

Deleted item:

Anderson.1

Completed successfully

Press any key to exit

```
PS /home/shellunext> [ ]
```

 **sahithdatalake** | Metrics (Classic) ☆ ...
Azure Cosmos DB account Metrics X[Download as csv](#)[Refresh](#)[Feedback](#)

Monitoring

 Metrics Metrics (Classic)[Overview](#) [Throughput](#) [Storage](#) [Availability](#) [Latency](#) [Consistency](#) [System](#)

Showing data for all containers

Show data for last

1 hour

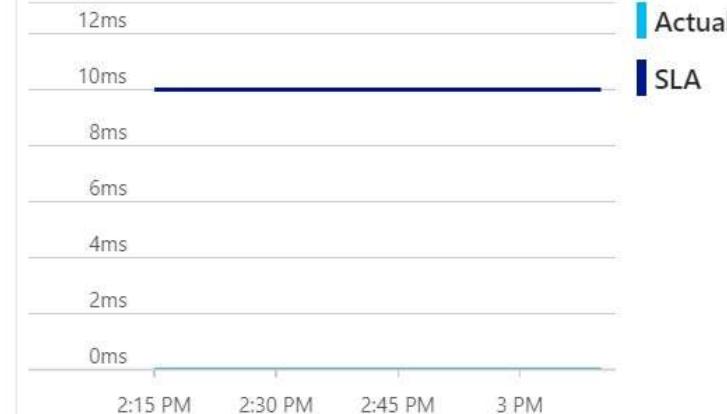


Below charts represent latencies for reads and writes of 1K objects observed by a continuously running test workload with a test Cosmos account in the East US region. These are offered for reference only, and do not represent the actual latencies measured for your application requests.

End to end observed read latency at the 99th percentile ⓘ



End to end observed write latency at the 99th percentile ⓘ



Day 6 – ADLS & ADF

- ▶ Blobs and Containers
- ▶ Anonymous access levels and Shared access tokens
- ▶ Server and Database level firewall
- ▶ Orchestrating Data Movement
- ▶ ETL vs ELT
- ▶ ADF Components and Processes
- ▶ Linked Service, Dataset, Activity and Pipeline
- ▶ Control Flow, Dynamic Parameters and Integration Runtime
- ▶ ADF Security and Triggers
- ▶ ADF Setup and Features

Day 6 continued...

- ▶ Creating a Static website in Storage account using Virtual folders
- ▶ Connecting to SQL VM and SSMS
- ▶ Querying database firewall rules
- ▶ Encryption, keys and secret creation
- ▶ Simple pipeline creation

 **democon** ...
Container Search

Upload

Change access level

Refresh



Delete



Change tier



Acquire lease



Break lease



Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Shared access tokens

Access policy

Properties

Metadata

Authentication method: Access key ([Switch to Azure AD User Account](#))**Location:** democon / virtualFolder

Search blobs by prefix (case-sensitive)

 Show deleted blobs

Add filter

Name	Modified	Access tier	Archive status	Blob type
<input type="checkbox"/> [.]				
<input type="checkbox"/> error.html	9/5/2023, 9:40:59 AM	Cool		Block blob
<input type="checkbox"/> index.html	9/5/2023, 9:40:59 AM	Cool		Block blob

Overview

Blob: Root/Parent/Child/Dir1/error.html

Save Discard Download Refresh Delete Change tier Acquire lease Break lease Give feedback

Properties

URL <https://sahithadls2.blob...>

LAST MODIFIED 9/5/2023, 9:55:02 AM

CREATION TIME 9/5/2023, 9:55:02 AM

VERSION ID -

TYPE Block blob

SIZE 71 B

ACCESS TIER Hot (Inferred)

ACCESS TIER LAST MODIFIED N/A

ARCHIVE STATUS -

REHYDRATE PRIORITY -

SERVER ENCRYPTED true

ETAG 0x8DBADC813158E1F

VERSION-LEVEL IMMUTABILITY POLICY Disabled

Manage ACL

...

container: adlsdemo (storage account: sahithadls2)

Set and manage permissions for:

/Root/Parent/Child/Dir1/error.html

[Learn more about access control lists \(ACLs\)](#)

Access permissions

[+ Add principal](#) [+ Add mask](#)

Security principal	Read	Write	Execute	
Owner: \$superuser	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
Owning group: \$superuser	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Other	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

-  Read and write permissions will only work for a security principal if the security principal also has execute permissions on all parent directories, including the container (root directory).

adlsdemo

...



Container



Upload

+ Add Directory

⟳ Refresh

⟳ Rename

Delete

⇄ Change tier

🔗 Acquire lease

✖ Break lease

↗ Give feedback

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Shared access tokens

Manage ACL

Access policy

Properties

Metadata

Authentication method: Access key ([Switch to Azure AD User Account](#))

Location: adlsdemo / v1 / v2 / v3

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type
<input type="checkbox"/> [..]				
<input type="checkbox"/> error.html	9/5/2023, 10:07:06 AM	Hot (Inferred)		Block blob



SQLQuery3.sql - demoserver0509.database.windows.net.master (sqladmin (68))* - Microsoft SQL Server Management Studio (Administrator)

File Edit View Query Project Tools Window Help

New Query MDX DMX XMLA DAX

master Execute

Object Explorer

sqlvm (SQL Server 16.0.4055.4 - sahit)

- Databases
 - System Databases
 - master
 - Tables
 - Views
 - Synonyms
 - Programmability
 - Service Broker
 - Storage
 - Security
 - model
 - msdb
 - tempdb
 - Database Snapshots
 - Security
 - Server Objects
 - Replication
 - Always On High Availability
 - Management
 - Integration Services Catalogs
 - SQL Server Agent (Agent XPs disabled)
- XEvent Profiler
- demoserver0509.database.windows.net (S)
 - Databases
 - Security
 - Integration Services Catalogs

Quick Launch (Ctrl+Q)

SQLQuery3.sql - de...ter (sqladmin (68))*

SQLQuery2.sql - sql...master (sahith (64))

SQLQuery1.sql - sql...master (sahith (62))

```
select * from sys.database_firewall_rules
select * from sys.firewall_rules
```

Results Messages

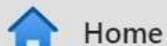
id	name	start_ip_address	end_ip_address	create_date	modify_date
1	3	0.0.0	255.255.255.255	2023-09-05 06:16:20.043	2023-09-05 06:16:20.043

i Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!



Azure Data Factory allows you to configure a Git repository with either Azure DevOps or GitHub. Git is a version control system that allows for easier change tracking and collaboration. [Learn more](#)

[Set up code repository](#)



Home



Author



Monitor



Manage

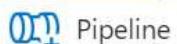


Learning Center

Data factory

sahithadf12

New



Pipeline



Power Query



Data flow



Dataset

scale once or
more



Orchestrate

Code-free data pipelines.



Transform data

Transform your data using
data flows.



Configure SSIS

Manage & run your SSIS
packages in the cloud.

Recent resources

Create a secret

Upload options

Name * ⓘ blobstorage ✓

Secret value * ⓘ ✓

Content type (optional)

Set activation date ⓘ

Set expiration date ⓘ

Enabled

Tags 0 tags

Create

Cancel



Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting

Data Factory

Validate all

Publish all 2

simplecopy12

Activities

Search activities

Move and transform

Copy data

Data flow

Synapse

Azure Data Explorer

Azure Function

Batch Service

Databricks

Data Lake Analytics

General

HDInsight

Iteration & conditionals

Machine Learning

✓ Validate ✓ Validate copy runtime

Copy data

simplecopypipeline



General Source¹ Sink¹ Mapping S...

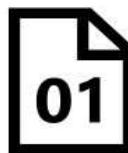
Source dataset *

Select...

Select format



Avro



Binary



DelimitedText



Excel



JSON



ORC



Parquet



XML

Continue

Back

Cancel

Microsoft Azure | Data Factory > sahitac

1 2 3 4 ? 🔎 Shellunext_1693422046249@npunext.onmicrosoft.com UNEXT

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting

Data Factory Validate all Publish all (2)

simplecopy12

Activities Search activities

Move and transform Copy data simplecopypipeline Data flow

Synapse Azure Data Explorer

Copy data simplecopypipeline

Set properties

Name DelimitedText1

Linked service * Select...

First row as header

The screenshot shows the Microsoft Azure Data Factory interface. On the left, there's a sidebar with icons for Home, Pipelines, Datasets, Data Flows, Synapse, and Azure Data Explorer. The main area shows a pipeline named 'simplecopy12'. A 'Copy data' activity is selected, highlighted with a blue border. The 'Set properties' panel is open on the right, showing fields for 'Name' (set to 'DelimitedText1') and 'Linked service' (with a dropdown menu labeled 'Select...'). There are also checkboxes for 'First row as header' and other validation options.

Day 7 – ADF Pipeline and Activities

- ▶ Key vaults and Secrets
- ▶ Ingesting data with copy activity
- ▶ Get metadata, Filter, ForEach loop, Lookup and If-else activities
- ▶ Creation of storage account, SQL database and Data factory
- ▶ ADF Setup and creating pipeline with different activities
- ▶ Creation of SQL VM and running query editor on the VM
- ▶ IR Setup and Pipeline runs

oric, a brand new and exciting

Validate all Publish all

Integration runt

The integration runtime
network environment.

New Refresh

Filter by name

Showing 1 - 1 of 1 ite

Name



Integration runtime setup

Settings Nodes Auto update Sharing Links

Install integration runtime on Windows machine or add further nodes using the Authentication Key.

Name

selfhostedIR321

Option 1: Express setup

[Click here to launch the express setup for this computer](#)

Option 2: Manual setup

Step 1: Download and install integration runtime

Step 2: Use this key to register your integration runtime

Name	Authentication key		
Key1	IR@4f927e28-4b6a-472a-bacc-84a8a8851fff@sahithadf12@ServiceEndpoint		
Key2	IR@4f927e28-4b6a-472a-bacc-84a8a8851fff@sahithadf12@ServiceEndpoint		

Close

Integration runtimes

The integration runtime (IR) is the compute infrastructure to provide the following data integration capabilities across different network environment. [Learn more](#)

+ New  Refresh

 Filter by name

Showing 1 - 2 of 2 items

Name ↑↓	Type ↑↓	Sub-type ↑↓	Status ↑↓	Related ↑↓	Region ↑↓	Version ↑↓
 AutoResolveIntegrationR...	Azure	Public	 Running	0	Auto Resolve	---
 selfhostedIR321	Self-Hosted	---	 Running	0	---	5.32.8597.5

Microsoft Azure | Data Factory | sahithadf12 | Search factory and documentation | Home | Settings | Diagnostics | Update | Help | Shellunext_1693422046249@npunext.onmicrosoft.com UNEXT

Microsoft recently announced the

Data Factory

General

Factory settings

Connections

Linked services

Integration runtimes

Microsoft Purview

Source control

Git configuration

ARM template

Author

Triggers

Global parameters

Data flow libraries

Security

Credentials

Microsoft Integration Runtime Configuration Manager

Home Settings Diagnostics Update Help

Self-hosted node is connecting to the cloud service

Integration Runtime: selfhostedIR321

Node: sqlvm

Stop Service

Data Source Credential ⓘ

Credential store: On-premises

Credential status: -----

Last backup time: N/A

Generate Backup Import Backup

Connecting to the cloud service (Data Factory V2)

Copy Data tool

Properties

Linked service: AzureSqlDatabase132

Object: SalesLT.Customer

Preview Schema

CustomerID	NameStyle	Title	FirstName	MiddleName	LastName	Suffix	CompanyName	SalesPerson
1	false	Mr.	Orlando	N.	Gee		A Bike Store	adventure-works\pamela0
2	false	Mr.	Keith		Harris		Progressive Sports	adventure-works\david8
3	false	Ms.	Donna	F.	Carreras		Advanced Bike Components	adventure-works\jillian0
4	false	Ms.	Janet	M.	Gates		Modular Cycle Systems	adventure-works\jillian0

e ⓘ Dynamic range

< Previous

Next >

Cancel

Overview X | Lumen X | sahitadfl X | Create a X | sahitadfl X | My Acco X | New tab X | The value X | Access d X | +

https://adf.azure.com/en/home/copywizard?factory=%2Fsubscriptions%2F2190116a-14d8-4b0c-8821-f97ab2...

Microsoft Azure | Data Factory > sahitadfl12 | Search factory and documentation

Shellunext_1693422046249@npunext.onmicrosoft.com UNEXT



Copy Data tool



Properties

Source

Destination

Settings

Review and finish

Review

Deployment

Summary

You are running pipeline to copy data from Azure SQL Database to Azure Blob Storage.



Azure SQL Database



Azure Blob Storage

Properties

Edit

Task name CopyPipeline_rn1

Task description

Source

Edit

Connection name AzureSqlData

Dataset name SourceDataset_rn1

Table name SalesLT.Customer

Destination

Edit

Connection name AzureBlobStorage1

Publish all 1

Preview experience

Off

pipeline lookup

Activities

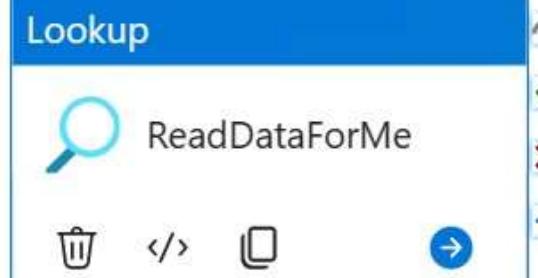
✓ Validate ▶ Debug ⚡ Add trigger



lookup

General

Lookup



General Settings User properties

Source dataset *

SourceDataset_rn1

Open + New Preview data
Learn more

First row only

Properties

General Related

Name *

pipelinelookup

Description

Annotations

+ New

Pipeline name

pipelinelookup

 Copy to clipboard

Output



»

[Copy to clipboard](#)

```
"count": 847,
"value": [
    {
        "CustomerID": 1,
        "NameStyle": false,
        "Title": "Mr.",
        "FirstName": "Orlando",
        "MiddleName": "N.",
        "LastName": "Gee",
        "Suffix": null,
        "CompanyName": "A Bike Store",
        "SalesPerson": "adventure-works\\pamela0",
        "EmailAddress": "orlando0@adventure-works.com",
        "Phone": "245-555-0173",
        "PasswordHash": "L/Rlwzxp4w7RWmEgXX+/A7cXaePEPcp+KwQhl2fJL7w=",
        "PasswordSalt": "1KjXYs4=",
        "rowguid": "3f5ae95e-b87d-4aed-95b4-c3797afcb74f",
        "ModifiedDate": "2005-08-01T00:00:00Z"
    },
    {
        "CustomerID": 2,
        "NameStyle": false,
        "Title": "Mr.",
        "FirstName": "Keith",
        "MiddleName": null,
        "LastName": "Harris",
        "Suffix": null,
        "CompanyName": "Progressive Sports",
        "SalesPerson": "adventure-works\\david8",
        "EmailAddress": "keith.oconnor@progressivesports.com"
    }
]
```

Pipeline runs

Triggered Debug

Rerun

Cancel options

Refresh

Edit columns

List

Gantt

Filter by run ID or name

Monrovia, Reykjavik (... : Last 24 hours)

Pipeline name : All

Status : All

Runs : Latest runs

Triggered by : All

Add filter

X

Copy filters

Export to CSV

Showing 1 - 4 items

Last refreshed 0 minutes ago

<input type="checkbox"/> Pipeline name ↑↓	Run start ↑↓	Run end ↑↓	Duration	Triggered by	Status ↑↓
<input type="checkbox"/> pipelinelookup	9/6/2023, 10:23:37 AM	9/6/2023, 10:23:54 AM	17s	Manual trigger	✓ Success
<input type="checkbox"/> pipelinelookup	9/6/2023, 9:53:25 AM	9/6/2023, 9:53:35 AM	11s	Manual trigger	✓ Success
<input type="checkbox"/> pipelinelookup	9/6/2023, 9:39:28 AM	9/6/2023, 9:39:36 AM	9s	Manual trigger	✓ Success
<input type="checkbox"/> CopyPipeline_rn1	9/6/2023, 9:29:47 AM	9/6/2023, 9:29:56 AM	10s	Manual trigger	✓ Success

Rerun

Cancel

Refresh

Update pipeline

List

Gantt



Activity runs

Pipeline run ID e58a8c6c-80da-4c99-b645-7e8f1b3532b4

All status

[Monitor in Azure Metrics](#) [Export to CSV](#)

Showing 1 - 2 items

Activity name	Activity status	Activity type	Run start	Duration	Log	Inte
ReadInfo	Succeeded	Get Metadata	9/6/2023, 10:23:39 AM	13s		self
ReadDataForMe	Succeeded	Lookup	9/6/2023, 10:23:39 AM	12s		self

Mic

Output

 Copy to clipboard

```
{  
    "structure": [  
        {  
            "physicalName": "CustomerID",  
            "type": "Int32",  
            "logicalType": "Int32",  
            "name": "CustomerID",  
            "physicalType": "int",  
            "precision": 10,  
            "scale": 255,  
            "DotNetType": "System.Int32, mscorelib, Version=4.0.0.0, Culture=neutral, PublicKeyToken=b77a5c561934e089"  
        },  
        {  
            "physicalName": "NameStyle",  
            "type": "Boolean",  
            "logicalType": "Boolean",  
            "name": "NameStyle",  
            "physicalType": "bit",  
            "precision": 255,  
            "scale": 255,  
            "DotNetType": "System.Boolean, mscorelib, Version=4.0.0.0, Culture=neutral, PublicKeyToken=b77a5c561934e089"  
        },  
        {  
            "physicalName": "Title",  
            "type": "String",  
            "logicalType": "String",  
            "name": "Title",  
            "physicalType": "nvarchar",  
            "precision": 255,  
            "scale": 255,  
            "DotNetType": "System.String, mscorelib, Version=4.0.0.0, Culture=neutral, PublicKeyToken=b77a5c561934e089"  
        }  
    ]  
}
```

Day 8 – ADF Data Flows

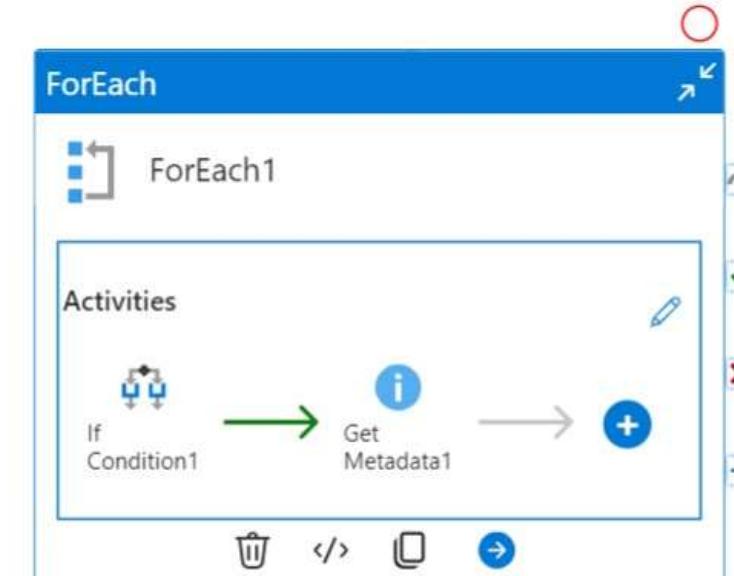
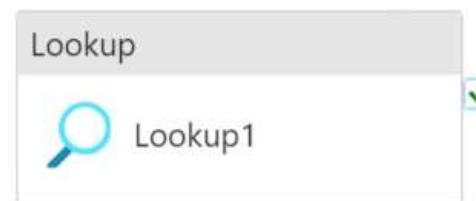
- ▶ Pipeline expression builder to add dynamic content
- ▶ Data flow creation
- ▶ Source and Sink Activities
- ▶ Conditional Split Transformation
- ▶ Derived Column Transformation
- ▶ Union
- ▶ Surrogate Key
- ▶ Mapping and Validate
- ▶ Demonstration of pipeline and data flow



✓ Validate

▷ Debug

⚡ Add trigger

GeneralSettings ¹

Activities (2)

User properties

Name *

ForEach1

Learn more

Description

Output

Copy to clipboard

```
{  
    "lastModified": "2023-09-07T04:39:39Z",  
    "childItems": [  
        {  
            "name": "file1.csv",  
            "type": "File"  
        },  
        {  
            "name": "file2.csv",  
            "type": "File"  
        },  
        {  
            "name": "file3.csv",  
            "type": "File"  
        }  
    ],  
    "effectiveIntegrationRuntime": "AutoResolveIntegrationRuntime (East US)",  
    "executionDuration": 0,  
    "durationInQueue": {  
        "integrationRuntimeQueue": 0  
    },  
    "billingReference": {  
        "activityType": "PipelineActivity",  
        "billableDuration": [  
            {  
                "meterType": "AzureIR",  
                "duration": 0.016666666666666666,  
                "unit": "Hours"  
            }  
        ]  
    }  
}
```



recently announced the public preview of Microsoft Fabric, a brand new and exciting

Data Factory Validate all Publish all (1)

pipeline321

Activities

filter

Iteration & conditionals

Filter

Get Metadata



GetListOfMeta



General

Settings²

User properties

Items *

This property

Add dynamic c

Condition *

This property

Pipeline expression builder

Add dynamic content below using any combination of expressions, functions and system variables.

```
@activity('GetListOfMeta').output
```

Clear contents

Activity outputs

Parameters

System variables

Functions

Variables

Search

GetListOfMeta

GetListOfMeta activity output

GetListOfMeta

GetListOfMeta pipeline return value

GetListOfMeta childItems

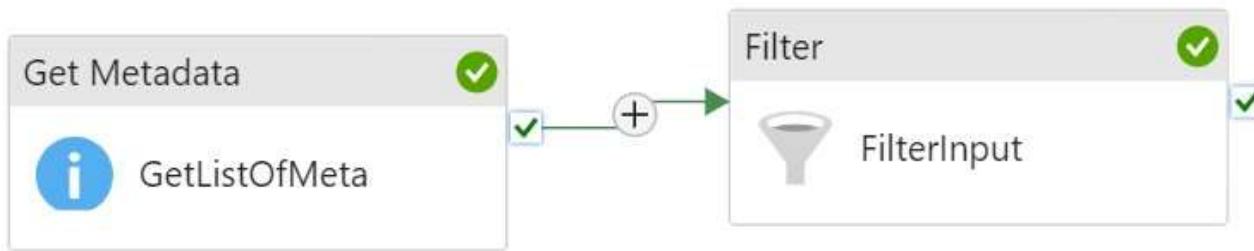
List of subfolders and files in the given folder

GetListOfMeta exists

OK

Cancel

Validate Debug Add trigger



Output

Copy to clipboard

```
{
  "ItemCount": 3,
  "FilteredItemCount": 3,
  "Value": [
    {
      "name": "file1.csv",
      "type": "File"
    }
  ]
}
```



Pipeline status



Succeeded

debug run consumption

Monitor in Azure Metrics



Export to CSV

		Time	Duration	Integration runtime	User properties
FilterInput	Succeeded	9/7/2023, 11:20:59 AM	Less than 1s		
GetListOfMeta	Succeeded	9/7/2023, 11:20:56 AM	2s	AutoResolveIntegration	

Data Factory Validate all Publish all (2)

pipeline321 DelimitedText4

 DelimitedText
DelimitedText4

Connection Schema Parameters

Linked service * AzureBlobStorage1

File path * democon / Directory

Compression type Select...

Column delimiter ⓘ Comma (,)

Row delimiter ⓘ Default (\r,\n, or \r\n)

Encoding ⓘ Default(UTF-8)

Quote character ⓘ Double quote ("")

Pipeline expression builder

Add dynamic content below using any combination of expressions, functions and system variables

@dataset().filename

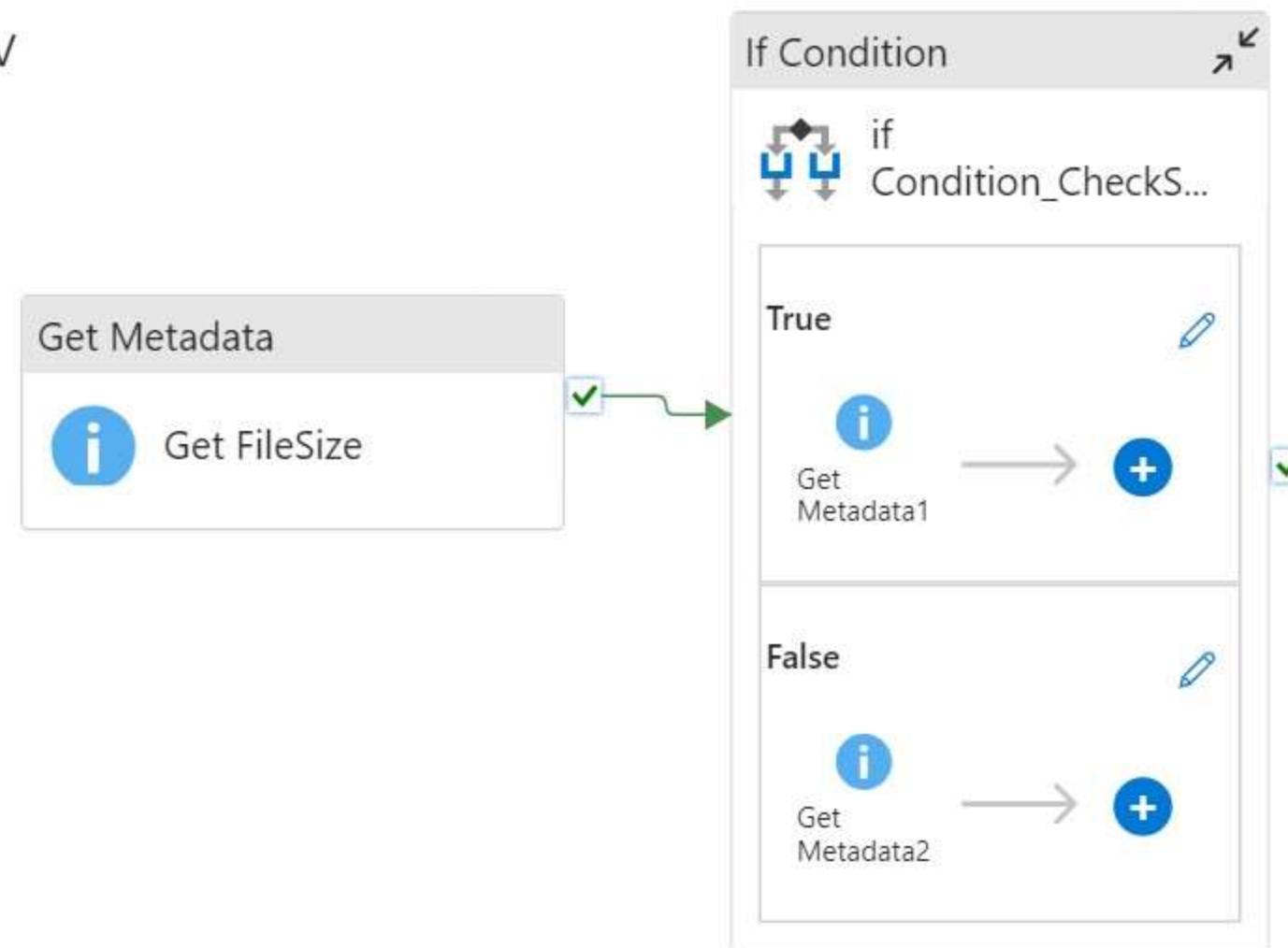
[Clear contents](#)[Parameters](#) [Functions](#)[!\[\]\(2a126c2a36ebe27b46f0d45fbbc8bf84_img.jpg\) Search](#)

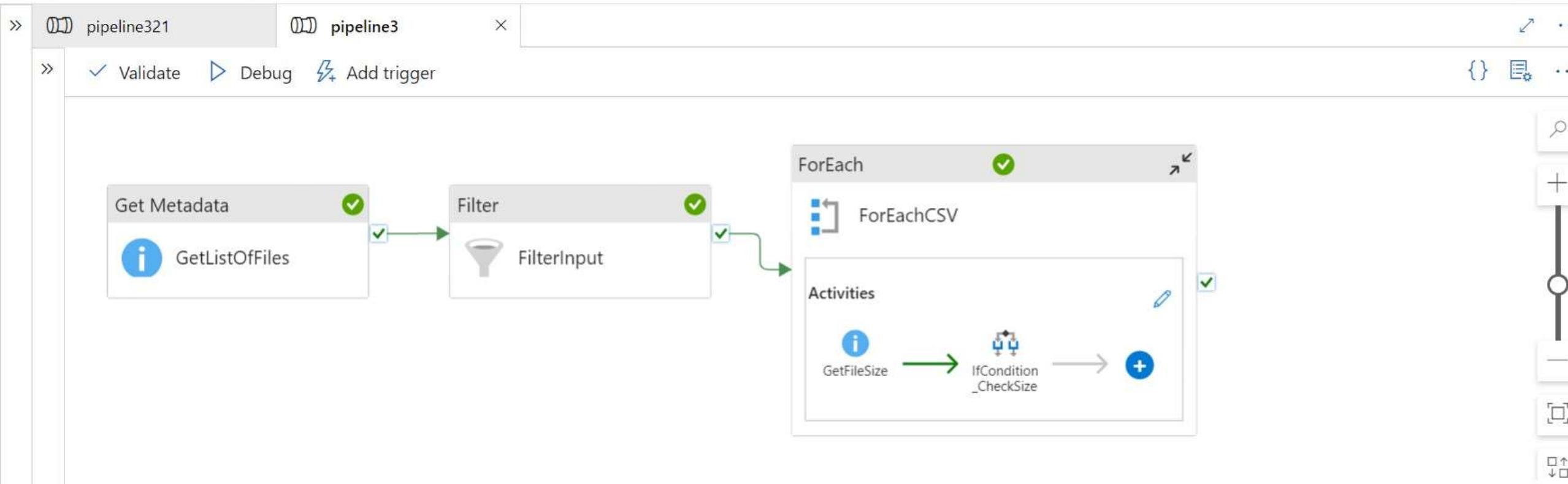
filename

[OK](#)[Cancel](#)

✓ Validate ▶ Debug ⚡ Add trigger

pipeline321 > ForEachCSV





Parameters Variables Settings **Output**

Pipeline run ID: c5ed8c4c-a42f-4b8a-a8b3-f811160cfbb9



Pipeline status **Succeeded**

[View debug run consumption](#)

All status **▼** List **▼**

[Monitor in Azure Metrics](#) [Export to CSV](#) **|** [▼](#)

Showing 1 - 12 of 12 items

Activity name ↑↓	Activity status ↑↓	Run start ↑↓	Duration ↑↓	Integration runtime	User properties ↑↓	Activity run ID ↑↓
CopyFromBlobToSQL	Succeeded	9/7/2023, 2:29:25 PM	10s	AutoResolveIntegration		cac67f04-9e07-4ff0-a686-

pipeline321

pipeline3

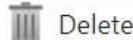
AllCSVFiles

...

Activities



General

✓ Validate ▷ Debug ⚡ Add trigger

{ }

Parameters Variables Settings Output

Pipeline run ID: c5ed8c4c-a42f-4b8a-a8b3-f811160cfbb9



Pipeline status



Succeeded debug run consumption

All status

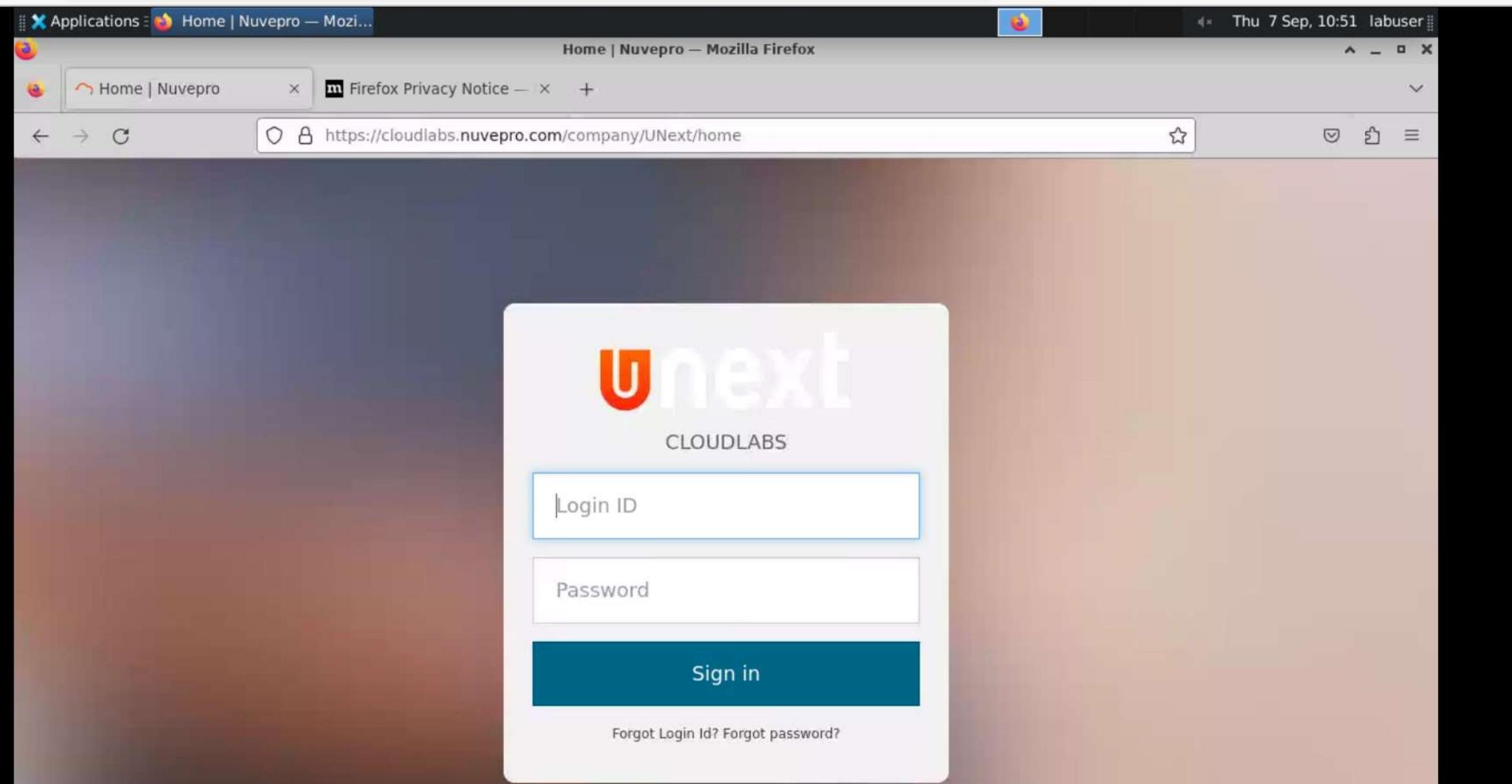
List

Monitor in Azure Metrics

Export to CSV

Showing 1 - 12 of 12 items

Activity name	Activity status	Run start	Duration	Integration runtime	User properties
CopyFromBlobToSQL	✓ Succeeded	9/7/2023, 2:29:25 PM	10s	AutoResolveIntegration	
IfCondition_CheckSize	✓ Succeeded	9/7/2023, 2:29:25 PM	13s		
GetFileSize	✓ Succeeded	9/7/2023, 2:29:22 PM	3s	AutoResolveIntegration	
CopyFromBlobToSQL	✓ Succeeded	9/7/2023, 2:29:11 PM	9s	AutoResolveIntegration	
IfCondition_CheckSize	✓ Succeeded	9/7/2023, 2:29:11 PM	11s		
GetFileSize	✓ Succeeded	9/7/2023, 2:28:58 PM	12s	AutoResolveIntegration	
CopyFromBlobToSQL	✓ Succeeded	9/7/2023, 2:28:46 PM	10s	AutoResolveIntegration	
IfCondition_CheckSize	✓ Succeeded	9/7/2023, 2:28:45 PM	13s		
GetFileSize	✓ Succeeded	9/7/2023, 2:28:41 PM	3s	AutoResolveIntegration	
ForEachCSV	✓ Succeeded	9/7/2023, 2:28:40 PM	1m 1s		



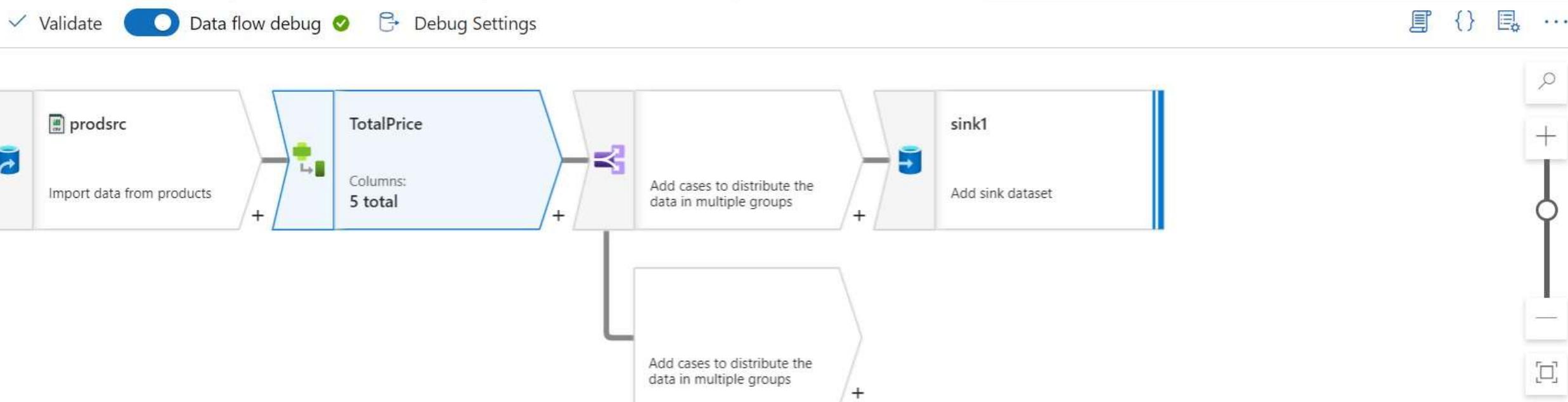
» pipeline321

pipeline3

InsideStats

dataflow1

...



Derived column's settings

Optimize

Inspect

Data preview

Schema

Input

Output

Number of columns **New*** 1**Updated*** 0**Unchanged** 4**Total** 5

Order ↑↓

Column ↑↓

Type ↑↓

Updated ↑↓

Based on ↑↓

1

pid

123 integer

2

pname

abc string

Day 9 – ADF APIs and CDC

- ▶ Implementing a data flow case scenario for Employee information
- ▶ Configuring and running VM to run SQL DB
- ▶ API Integration
- ▶ Using OpenWeatherMapAPI in REST resource
- ▶ Sending data from REST to Blob Storage
- ▶ Extracting modified data – Change Data Capture (CDC)
- ▶ Incremental update and CDC control task
- ▶ Staged Inserts, Updates and Deletes
- ▶ Monitoring the network
- ▶ Metric data, Log data and Alerts
- ▶ Diagnose and solve problems

Day 9 continued...

- ▶ Implementing CDC process by capturing updates to files in SQL DB
- ▶ Azure Monitor case scenario with Alerts
- ▶ Troubleshooting data storage issues
- ▶ Connectivity, Performance and Storage issues

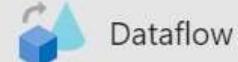
All Publish all 3

pipeline3

InsideStats

✓ Validate ▶ Debug ✓ ⚡ Add trigger

Data flow



Dataflow



General

Settings

Parameters ¹

User p

Data flow parameters ⓘ

Name

Value

No data flow parameters

New trigger

Name *

AutoSchedule

Description

Type *

Schedule

Start date *

9/8/2023, 3:55:28 AM

Time zone *

Chennai, Kolkata, Mumbai, New Delhi (UTC+5:30)

Recurrence *

Every

15

Minute(s)

Specify an end date

Annotations

+ New

Start trigger ⓘ

OK

Cancel

✓ EmpDataflow

Cluster startup time: 1s 424ms Number of transformations: 4 Data flow status: Success

Refresh Auto refresh

On Edit dataflow



Sinks All streams

Sink	Status	Processing time ↑	Highest processing	Rows written ↑	Stages	Lineage
AgeGt23sink	Succeeded	518ms	911ms	4		
AgeLt23sink	Succeeded	1s 735ms	911ms	6		

 dataflowEmp

AzureSqlTable

AgeGt23SqlTable

 pipeline1

 apipipeline

1

四

Copy data

CopyFromAPI



General Source Sink Mapping Settings User properties

Source dataset *  srcRestResource   Open  New  Preview data  Learn more 

Request method ⓘ GET ▾

Request timeout ⓘ 00:01:40

Request interval (ms) ⓘ 10

[Additional headers](#) ⓘ  New

Pagination rules ⓘ + New Delete

Name	Value
RFC5988	True

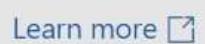
Data Factory Validate all Publish all

apipipeline srcRestResource

 REST
srcRestResource

Connection Parameters

Linked service *

 Test connection  Edit  New 
 Connection successful

Base URL

Relative URL

Edit linked service

 REST [Learn more](#)

Name *

Description

Connect via integration runtime * 

Base URL *

 Information will be sent to the URL specified. Please ensure you trust the URL entered.

Authentication type *

Server Certificate Validation 
 Enable Disable

Auth headers 
 New

 Save  Cancel  Test connection

Factory Resources

Filter resources by name

Pipelines

apipipeline

pipeline1

Change Data Capture (preview)

- > Move and transform
- > Synapse
- > Azure Data Explorer
- > Azure Function
- > Batch Service
- > Databricks
- > Data Lake Analytics
- > General
- > HDInsight
- > Iteration & conditionals
- > Machine Learning
- > Power Query

Datasets

AgeGt23SqlTable

AzureSqlTable1

DelimitedText1

empcsv

srcRestResource

Data flows

dataflowEmp

Power Query

Choose Your Sources

Change Data Capture (CDC) automatically detects data changes at the source and then sends the updated data to the destination.

CDC name *

adfdc1

Source type *

Select...

Continue

Cancel

Preview data

Linked service: EmpBlobStorage

Object:

	lon	lat
1	-0.1257	51.5085

e source and then sends the



Continue

Cancel

srcRestResource apipipeline pipeline1 adfccdc1

Start Set Latency Data flow debug

⚠ This preview feature is licensed to you as part of your Azure subscription. By continuing you agree to the [Preview Terms](#) and [Privacy Statement](#).

i Your insights matter! Participate in our [brief survey](#) about our CDC top-level resource, and help us enhance your experience.

Filter by target table name

AgeGt23SqlDatabase

+ New mapping Delete

Source Table ↑
democon

Target Table ↑
dbo.democon

Columns mapped ↑↓ Auto map

Source Target

Target linked service	Type	Target tables
AgeGt23SqlDatabase	Azure SQL Database	1 edit

employee.csv

Blob

Save Discard Download Refresh | Delete

Overview Versions Snapshots **Edit** Generate SAS

```
1 eid,ename,eage,esalary
2 1,ab,13,10
3 2,cd,15,11
4 3,ef,29,15
5 4,gh,27,90
6 5,ij,20,100
7 6,kl,89,199
8 7,mn,18,33
9 8,op,17,11
10 9,qr,31,17
11 10,st,20,111
12
```

Csv



Preview

[Feedback](#) [Getting started](#)Query 1  Run  Cancel query  Save query  Export data as  Show only Editor

```
1 select * from [dbo].[democon]
```

[Results](#) [Messages](#)

8	op	17	11
9	qr	31	17
10	st	20	11

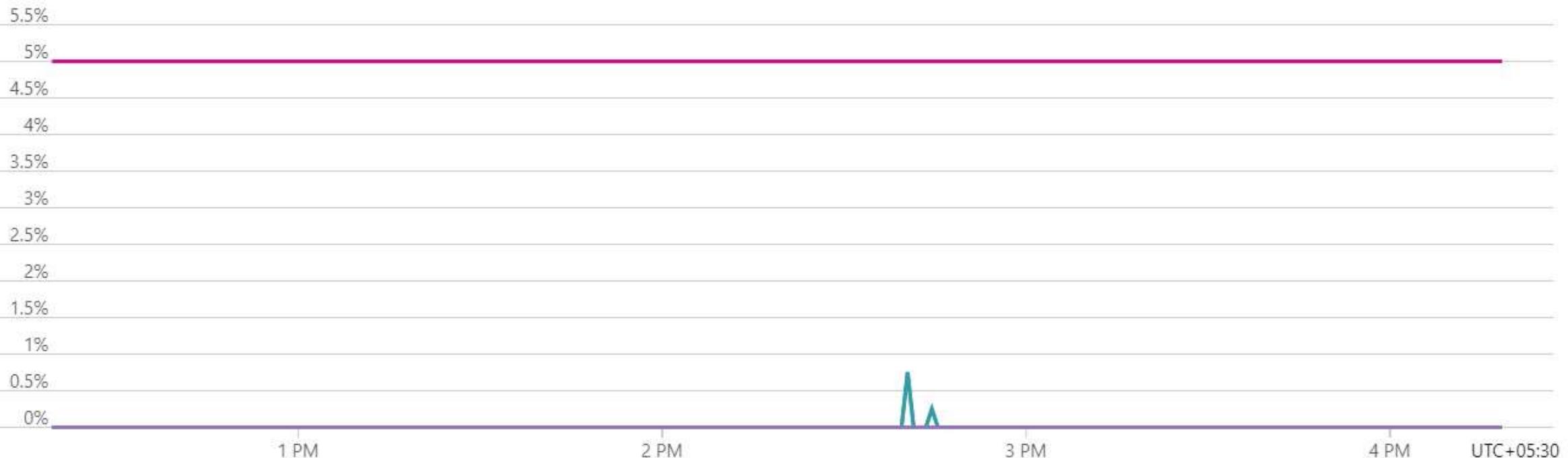
Query succeeded | 0s

[+ New chart](#) [⟳ Refresh](#) [🔗 Share](#) [😊 Feedback](#)

Local Time: Last 4 hours (Automatic - 1 minute)

SQL sahithserver/sahithdb, CPU percentage, Avg [×](#)sahithserver/sahithdb, DTU Limit, Avg [×](#)sahithserver/sahithdb, DTU percentage, Avg [×](#)sahithserver/sahithdb, Data IO percentage, Avg [×](#)

Scope	Metric Namespace	Metric	Aggregation
sahithserver/sahithdb	Sql database standar...	Select metric	Select aggregation

CPU percentage (Avg)
sahithdb

0.0042 %

DTU Limit (Avg)
sahithdb

5

DTU percentage (Avg)
sahithdb

0.0042 %

Data IO percentage (...)
sahithdb

0 %

Create an alert rule

...

Scope Condition Actions Details Tags Review + create

Configure when the alert rule should trigger by selecting a signal and defining its logic.

Signal name * ⓘ

 Used capacity ▼

[See all signals](#)

Alert logic

Threshold ⓘ

Static Dynamic

 Dynamic Thresholds is currently not available for this metric

Aggregation type ⓘ

Average ▼

Operator ⓘ

Greater than ▼

Unit ⓘ

B ▼

Threshold value * ⓘ

30 ✓

[Review + create](#)

[Previous](#)

[Next: Actions >](#)

Monitor | Alerts

Microsoft



Overview

Activity log

Alerts

Metrics

Logs

Change Analysis

Service health

Workbooks

Insights

Applications

[View as timeline \(preview\)](#)[Create](#)[Alert rules](#)[Action groups](#)[Alert processing rules](#)[Columns](#)[Refresh](#)[Search](#)

Subscription : 2190116a-14d8-4b0c-8821-f97ab2e43421

[Add filter](#)[More \(3\)](#)

Total alerts

0

Critical

0

Error

0

Warning

0

Informational

0

Verbose

0

[No grouping](#)[Name ↑↓](#)[Severity ↑↓](#)[Affected resource ↑↓](#)[Alert condition ↑↓](#)[User response ↑↓](#)[Fire time ↑↓](#)



sahithstore1 | Diagnose and solve problems



Storage account



Search



Overview



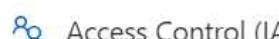
Activity log



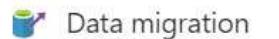
Tags



Diagnose and solve problems



Access Control (IAM)



Data migration



Events



Storage browser



Storage Mover

Data storage



Containers



File shares



Queues



Tables

Security + networking

Common problems

Explore the most common problems for your resource. Select Troubleshoot to run an automated troubleshooter, follow do-it-yourself troubleshooting steps, or explore a wide range of troubleshooting tools.

No health related issues found. [View resource health dashboard](#)

1 change to this resource in the last 72 hours. [View change details](#)

I can't delete my storage account

[Troubleshoot](#)

Move Data to, from, or within Azure Storage

[Troubleshoot](#)

I need help with Import/Export

[Troubleshoot](#)