

# Customer Churn Azure Project

---

**Note:** The dataset provided for this assignment was already clean, with no missing values or Personally Identifiable Information (PII). However, for demonstration purposes, a filter condition was implemented to show how missing values can be handled using an Azure Data Factory (ADF) pipeline.

## Project Objective

The goal of this project was to migrate data from an API source to an Azure SQL Database and establish an automated orchestration workflow using Azure Data Factory.

## Technology Stack

- Resource Group
- Storage Accounts
- Azure Data Factory (ADF)
- Azure SQL Database

## ADF Activities Used

- Lookup Activity
- Copy Data Activity
- Data Flow

## Project Workflow

### Step 1: Azure Resources Setup

1. Created a Resource Group in Azure.
2. Added a Storage Account and Azure Data Factory instance under this resource group.
3. Within the Storage Account, created three containers:
  - config – Stores the file name from the API so the pipeline can automatically pick up new files.
  - staging – Contains the raw CSV file loaded from the GitHub repository.
  - transformation – Stores the transformed data in Parquet format for SQL loading.

Screenshots for reference.:

Microsoft Azure

Upgrade

Search resources, services, and docs (G+V)

Copilot

sausifk\_07@outlook.com

EMAIL DIRECTORY (MAGNIFY)

Home >

RG-CCProject

Are there any alerts fired for this resource group?

What are the best practices for managing this resource group?

How do I monitor this resource group?

Resource group

Search

Create

Manage view

Delete resource group

Refresh

Export to CSV

Open query

Assign tags

Move

Delete

Export template

Open in mobile

Overview

Activity log

Access control (IAM)

Tags

Resource visualizer

Events

Settings

Cost Management

Monitoring

Automation

Help

Essentials

Subscriptions

Deployments

Location

Resources

Recommendations

Filter for any field...

Type equals all

Location equals all

Add filter

Showing 1 to 2 of 2 records

Show hidden types

No grouping

List view

Name	Type	Location
ccadstg	Data factory (V2)	Central India
ccadstg	Storage account	Central India

+ Add container   ↑ Upload   ↻ Refresh   🗑 Delete   🔒 Change access level   ↺ Restore containers   🛠 Edit columns

🔍 Search containers by prefix

Showing all 5 items

<input type="checkbox"/>	Name
<input type="checkbox"/>	adbtransform
<input type="checkbox"/>	.config
<input type="checkbox"/>	presentation
<input type="checkbox"/>	.staging
<input type="checkbox"/>	.transformation

Microsoft Azure

Upgrade

Search resources, services, and docs (G+V)

Copilot

sausifk\_07@outlook.com

EMAIL DIRECTORY (MAGNIFY)

Home >

RG-CCProject >

ccadstg | Containers >

staging

Container

Search

+ Add Directory

Upload

Refresh

Delete

Copy

Paste

Rename

Acquire lease

Break lease

Edit columns

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

staging

Authentication method: Access key

Search blobs by prefix (case-sensitive)

Only show active objects

Showing all 1 items

Name	Last modified	Access tier	Blob type	Size	Lease state
customer_churn_data.csv	14/06/2023, 00:32:48	Hot (Inferred)	Block blob	57.29 KiB	Available

Microsoft Azure

Upgrade

Search resources, services, and docs (G+V)

Copilot

sausifk\_07@outlook.com

EMAIL DIRECTORY (MAGNIFY)

Home >

RG-CCProject >

ccadstg | Containers >

staging

Container

Search

+ Add Directory

Upload

Refresh

Delete

Copy

Paste

Rename

Acquire lease

Break lease

Edit columns

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

staging

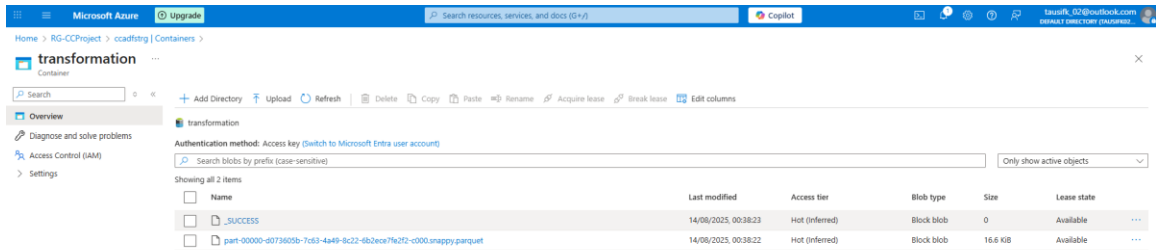
Authentication method: Access key

Search blobs by prefix (case-sensitive)

Only show active objects

Showing all 1 items

Name	Last modified	Access tier	Blob type	Size	Lease state
customer_churn_data.csv	14/06/2023, 00:32:48	Hot (Inferred)	Block blob	57.29 KiB	Available

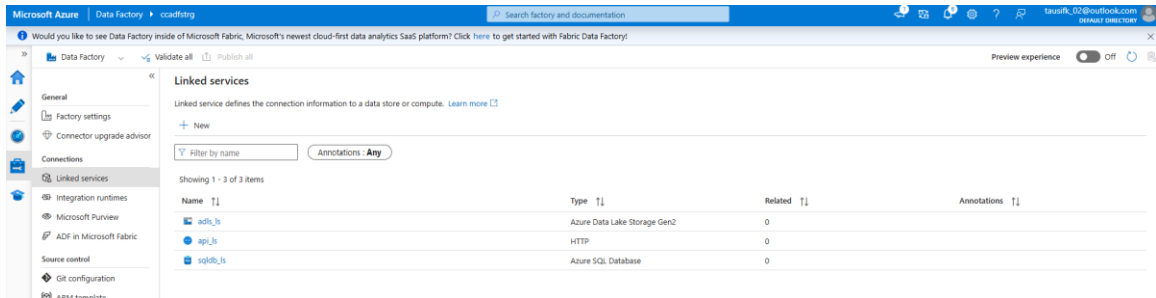


## Step 2: Linked Services Configuration

Configured Linked Services in ADF for connections to:

- Azure Storage Account
- Azure SQL Database

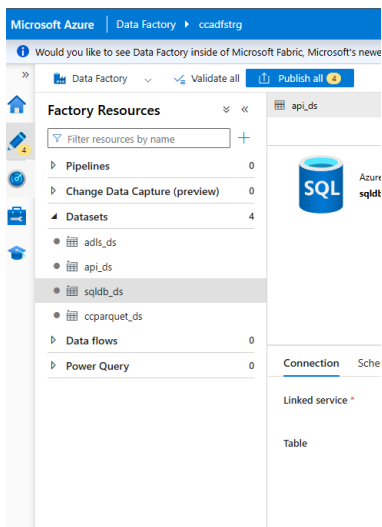
*Screenshot for reference:*



## Step 3: Dataset Creation

Created datasets in ADF for:

- The config file in the Storage Account
- The staging CSV file
- The transformed Parquet file
- The destination SQL Database table



## Step 4: ETL Pipeline Development

Pipeline Workflow:

1. The config file containing the file name is passed through a Lookup Activity to the Copy Data Activity.
2. In **Data Flows**, filtered out rows where "InternetService" = "None".
3. Saved the filtered data as a Parquet file in the transformation folder.
4. Used another **Copy Data Activity** to load the Parquet file into the SQL Database.

The screenshot shows the Microsoft Azure Data Factory interface. The pipeline 'copy from api to g...' is displayed with the following activities:

- Lookup: Getfilename
- Copy data: Copy API to Staging
- Data flow: Transformed data
- Copy data: Transformed to Presentation

The pipeline status is 'Succeeded'. The output table shows the following details:

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties	Activity run ID
Transformed to Presentation	Succeeded	Copy data	8/14/2025, 12:27:07 AM	17s	AutoResolveIntegrationRuntime (Central India)		3c9e938f-5e1a-4223-ba9b-996c05b
Transformed data	Succeeded	Data flow	8/14/2025, 12:24:55 AM	2m 11s	AutoResolveIntegrationRuntime (Central India)		c3135ee-204e-464b-bd89-9f2d99
Copy API to Staging	Succeeded	Copy data	8/14/2025, 12:24:41 AM	14s	AutoResolveIntegrationRuntime (Central India)		66c50d05-897f-4b73-aefb-317002e
Getfilename	Succeeded	Lookup	8/14/2025, 12:24:36 AM	4s	AutoResolveIntegrationRuntime (Central India)		ee2e5e55-0a30-4054-a66c-e0fb84a

The screenshot shows the Microsoft Azure portal interface for the 'staging' container. The container contains one file:

Name	Last modified	Access tier	Blob type	Size	Lease state
customer_shum_data.csv	14/8/2025, 12:24:53 am	Hot (Inferred)	Block blob	57.29 KiB	Available

The screenshot shows the Microsoft Azure portal interface for the 'transformation' container. The container contains two files:

Name	Last modified	Access tier	Blob type	Size	Lease state
_SUCCESS	14/8/2025, 12:26:10 am	Hot (Inferred)	Block blob	0	Available
part-00000-01138846-e089-4353-8d75-cdd9737404c-c000.snappy.parquet	14/8/2025, 12:26:10 am	Hot (Inferred)	Block blob	16.6 KiB	Available

Microsoft Azure | Upgrade | Search resources, services, and docs (Ctrl+K) | Copilot

Home > tk-azure-sql-server > SQL databases > CCpresentationlayer (tk-azure-sql-server/CCpresentationlayer) | Query editor (preview) ☆

Search | Login | New Query | Open query | Feedback | Getting started

CCpresentationlayer (tausifk\_02@outlook.com)

Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

Tables: db.Customer\_churn

Views: Stored Procedures

Query 1: `SELECT TOP (1000) * FROM [dbo].[Customer_churn]`

Run | Cancel query | Save query | Export data as | Show only Editor

Results | Messages

CustomerID	Age	Gender	Tenure	MonthlyCharges	ContractType	InternetService
1	49	Male	4	88.35	Month-to-Month	Fiber Optic
2	43	Male	0	36.67	Month-to-Month	Fiber Optic
3	51	Female	2	63.79	Month-to-Month	Fiber Optic
4	60	Female	6	102.34	One-Year	DSL
6	42	Female	16	119.75	Two-Year	DSL
9	40	Female	53	49.81	Two-Year	Fiber Optic
10	50	Female	10	61.55	Month-to-Month	Fiber Optic
11	40	Female	1	63.53	Month-to-Month	Fiber Optic

Query succeeded | 0s

## Step 5: Orchestration and Automation

- Set a trigger so the pipeline runs automatically at scheduled times. *As per assignment it is to be scheduled for every hour.*
- Screenshots of the setup are included for reference.

Microsoft Azure | Data Factory | ccdatfmg | Search factory and documentation

All pipeline runs > copy from api to gen2 - Activity runs

Rerun | Cancel | Refresh | Update pipeline | List | Gantt

Lookup: GetItemname

Copy data: Copy API to Staging

Data flow: Transformed data

Copy data: TransformedtoPresentation

Activity runs

Pipeline run ID: bf2a3e1e-b233-43f7-9aba-47b579b04529

All status: v

Showing 1 - 4 items

Activity name	Activity status	Activity	Run start	Duration	Integration runtime	User properties	Activity run ID	Log
TransformedtoPresentation	Succeeded	Copy data	8/14/2025, 12:38:28 AM	20s	AutoResolveIntegrationRuntime (Central India)		f9d3c72a-8368-427d-a099-b52b2db4f1bc	
Transformed data	Succeeded	Data flow	8/14/2025, 12:32:50 AM	5m 38s	AutoResolveIntegrationRuntime (Central India)		de2a5a89-73f5-4078-9c31-aaab3cde5e09	
Copy API to Staging	Succeeded	Copy data	8/14/2025, 12:32:37 AM	12s	AutoResolveIntegrationRuntime (Central India)		e6792f19-240b-4b66-ae64-ef2dc05d5ea40	
GetItemname	Succeeded	Lookup	8/14/2025, 12:32:33 AM	4s	AutoResolveIntegrationRuntime (Central India)		6c129561-825a-4fb4-9b60-c191f6e600b5	

Microsoft Azure | Data Factory | ccdatfmg | Search factory and documentation | taussfkh\_02@outlook.com | DEFAULT DIRECTORY

Pipeline runs

Triggered Debug Run Cancel options Refresh Edit columns List Gantt

Filter by run ID or name Chennai, Kolkata, Mu... Last 24 hours Pipeline name: All Status: All Runs: Latest runs Triggered by: All Add filter Copy filters Export to CSV

Showing 1 - 9 items

<input type="checkbox"/>	Pipeline name %s	Run start %s	Run end %s	Duration	Triggered by	Status %s	Run	Parameters	Annotations	Run ID
<input type="checkbox"/>	copy from api to gen2	8/14/2025, 12:32:26 AM	8/14/2025, 12:38:49 AM	6m 23s	Manual trigger	Succeeded	Original			b72a3e6e-b253-43f7-9aba-47b579c654
<input type="checkbox"/>	copy from api to gen2	8/14/2025, 12:19:05 AM	8/14/2025, 12:23:59 AM	4m 54s	Manual trigger	Cancelled	Original			6d8b0e9e-9c29-4c1b-b4c9-b62e388b7
<input type="checkbox"/>	copy from api to gen2	8/14/2025, 12:09:43 AM	8/14/2025, 12:15:02 AM	5m 20s	Manual trigger	Cancelled	Original			9a5b5572-31d4-4bc9-b927-fd8b5a2ec
<input type="checkbox"/>	copy from api to gen2	8/14/2025, 12:07:32 AM	8/14/2025, 12:07:33 AM	2s	Manual trigger	Failed	Original			5d814873-9719-4c7b-aa05-268b78cd1
<input type="checkbox"/>	copy from api to gen2	8/13/2025, 11:40:15 PM	8/13/2025, 11:40:17 PM	2s	Manual trigger	Failed	Original			4dc7cdbc-69d3-4d3a-b75f-638437e36
<input type="checkbox"/>	copy from api to gen2	8/13/2025, 11:21:20 PM	8/13/2025, 11:26:46 PM	5m 26s	Manual trigger	Succeeded	Original			66852149-12f1-4a3c-b87a-ee8738b68
<input type="checkbox"/>	copy from api to gen2	8/13/2025, 11:12:58 PM	8/13/2025, 11:19:10 PM	6m 12s	Manual trigger	Succeeded	Original			a19067cd-9fe1-4bdc-8a7c-ba3e5988e
<input type="checkbox"/>	copy from api to gen2	8/13/2025, 11:10:21 PM	8/13/2025, 11:10:50 PM	29s	Manual trigger	Failed	Original			4229d3a2-7131-47a6-aa21-063ec08a4
<input type="checkbox"/>	copy from api to gen2	8/13/2025, 11:06:11 PM	8/13/2025, 11:06:41 PM	30s	Manual trigger	Failed	Original			0ee747af-5d57-49f1-b0f1-bb8ed3566

Last refreshed 0 minutes ago

## Outcome

This project shows how to:

- Move data from an API to Azure SQL Database automatically.
- Handle files in a dynamic way using a config file.
- Apply simple cleaning steps in ADF.
- Orchestrate the whole process in one workflow.