

CHALLENGE – 2

MOVING DATA/AZURE DATA FACTORY

AI INFRASTRUCTURE AND ARCH.

Name: Samrat Thapa -- 200481763

CREATING DATA FACTORY

AI_infra_CT (samrat-ai-infrastruc...)

Create Data Factory - Microsoft

how to download azcopy using ...

+

portal.azure.com/#create/Microsoft.DataFactory

Imported From IE (2) FASTEST Way to... conversational AI ethics Georgian College... YouTube Inbox (45) - tsamrat... math Data Connectors ... AI for decision maki... machine learning fr... vision system SKINOPATHY AI Infra - AWS Other bookmarks

Microsoft Azure

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

200503303@student.ge... GEORGIAN COLLEGE (GEORGIAN...

Home > Create a resource >

Create Data Factory

Validation Passed

Basics Git configuration Networking Advanced Tags Review + create

TERMS

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s) with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage, and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. See the [Azure Marketplace Terms](#) for additional details.

Basics

Subscription

Azure for Students

Resource group

AI_infrastructure

Name

datafactory-twitter

Region

East US

Version

V2

Networking

Connect via

Public endpoint

Create < Previous Next

Give feedback

Download a template for automation

AI_Infra_C1 (samrat-ai-infrastruct... x

Microsoft.DataFactory-20221112... x

how to download azcopy using... x

+

portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade/~/overview/id/%2Fsubscriptions%2F8c1bcaa6-91ae-435f-9dcb-25a67d95315f%2FresourceGroups%2FAI_Infrastructure%2Fproviders%2FMicrosoft.Resources%2Fdeployment...

Imported From IE (2) FASTEST Way to... conversational AI ethics Georgian College... YouTube Inbox (45) - samrat... math Data Connectors ~ AI for decision maki... machine learning fr... vision system SKINOPATHY AI Infra - AWS Other bookmarks

Microsoft Azure

Search resources, services, and docs (0+)

200503303@student.ge... GEORGIAN COLLEGE (GEORGIAN...

Home >

Microsoft.DataFactory-20221112124255 | Overview

Deployment

Search

Delete Cancel Redeploy Download Refresh

Overview

Inputs

Outputs

Template

✔ Your deployment is complete

Deployment name: Microsoft.DataFactory-20221112124255

Subscription: Azure for Students

Resource group: AI_Infrastructure

Start time: 11/12/2022, 12:44:23 PM

Correlation ID: d30244dd-5bea-4227-9d70-713098f8e6f8

Deployment details

Next steps

Go to resource

Give feedback

Tell us about your experience with deployment

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 >

AI Infra_C1 (samrat-ai-infrastruc...datafactory-twitter - Microsoft A...how to download azcopy using ...


portal.azure.com/#@georgiancollege.onmicrosoft.com/resource/subscriptions/8c1bcaa6-91ae-435f-9dc6-25a67d95315f/resourceGroups/AI_Infrastructure/providers/Microsoft.DataFactory/factories/datafactory-twitter/overview

Imported From IE(2) FASTEST Way to...conversational AIethicsGeorgian College...YouTubeInbox (45) - samrat...mathData Connectors...AI for decision maki...machine learning fr...vision systemSKINOPATHYAI Infra - AWSOther bookmarks

Microsoft AzureSearch resources, services, and docs (0+)

200503303@student.ge...GEORGIAN COLLEGE (GEORGIAN...

Home > Microsoft.DataFactory-20221112124255 | Overview >

datafactory-twitter

Data factory (V2)

Delete

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Settings

Networking

Managed identities

Properties

Locks

Getting started

Quick start

Monitoring

Alerts

Metrics

Diagnostic settings

Logs

Automation

Tasks (preview)

Help

Resource health

New Support Request

Essentials

Resource group (move) : AI_Infrastructure

Status : Succeeded

Location : East US

Subscription (move) : Azure for Students

Subscription ID : 8c1bcaa6-91ae-435f-9dc6-25a67d95315f

Type : Data factory (V2)

Getting started : Quick start

Launch studio

Launch Quick Start

Tutorials

Template Gallery

Training Modules

Monitoring

PipelineRuns

ActivityRuns

TriggerRuns

UPLOADING TWITTER DATA

The screenshot displays the Microsoft Azure portal's Storage browser interface. The left sidebar contains navigation options: Overview, Activity log, Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, Storage browser (selected), Data storage (Containers, File shares, Queues, Tables), Security + networking (Networking, Azure CDN, Access keys, Shared access signature, Encryption, Microsoft Defender for Cloud), and Data management (Redundancy, Data protection). The main pane shows the 'storagesqlsamrat' storage account. Under 'Blob containers', the 'sqldatacontainer' is selected. The 'Add filter' section shows a search for 'Twitter_Data.csv'. The table below lists the blobs in the container.

	Name	Last modified	Access tier	Blob type	Size	Lease state
<input checked="" type="checkbox"/>	Twitter_Data.csv	11/12/2022, 12:46:05 PM	Cool	Block blob	19.93 MiB	Available
<input type="checkbox"/>	winequality-red.csv	11/12/2022, 12:14:27 PM	Cool	Block blob	82.17 KiB	Available

DATA FACTORY PORTAL

Microsoft Azure | Data Factory | datafactory-twitter


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](#)

Data factory


datafactory-twitter

[New](#)



-  **Ingest**
Copy data at scale once or on a schedule.
-  **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



No items to show

Your recently opened resources will show up here.

Discover more

-  [Browse partners \(preview\)](#)
-  [Pipeline templates](#)
-  [SAP pipeline templates](#)

AI_Infra_C1 (samrat-ai-infrastru...datafactory-twitter - Microsoft A...datafactory-twitter - Azure Data

adf.azure.com/en/home/copywizard?factory=%2Fsubscriptions%2F8c1bcaa6-91ae-435f-9dcb-25a67d95315f%2FresourceGroups%2FAI_Infrastructure%2Fproviders%2FMicrosoft.DataFactory%2Ffactories%2Fdatafactory-twitter

Imported From IE(2) FASTEST Way to...conversational AIethicsGeorgian College...YouTubeInbox (45) - samrat...mathData Connectors...AI for decision maki...machine learning fr...vision systemSKINOPATHYAI Infra - AWSOther bookmarks

Microsoft AzureData factorydatafactory-twitter200503303@student.georgianc.on.caGEORGIAN COLLEGE

Copy Data tool

Properties

Source

Dataset

Configuration

Destination

Settings

Review and finish

Source data store

Specify the source data store for the copy task. You can use an existing data store connection or specify a new data store.

Source typeAzure Blob Storage

ConnectionSelect...+ New connection

Previous

Next

New connection

Azure Blob Storage Learn more

Name *AzureBlobStorage1

Description

Connect via integration runtime * AutoResolveIntegrationRuntime

Authentication typeAccount key

Connection string Azure Key Vault

Account selection method

From Azure subscription

Enter manually

Azure subscription * Azure for Students (8c1bcaa6-91ae-435f-9dcb-25a67d95315f)

Storage account name * storagesqsamrat

Additional connection properties

+ New

Test connection

To linked service

To file path

Annotations

+ New

Parameters

Advanced

CreateCancel

Test connection

AI_Infra_C1 (samrat-ai-infrastruc...datafactory-twitter - Microsoft A...datafactory-twitter - Azure Data...+
←→↻adf.azure.com/en/home/copywizard?factory=%2Fsubscriptions%2F8c1bcaa6-91ae-435f-9dcb-25ab7d95315f%2FresourceGroups%2FAI_Infrastructure%2Fproviders%2FMicrosoft.DataFactory%2FFactories%2Fdatafactory-twitter
Imported From IE(2) FASTEST Way to...conversational ALetricsGeorgian College...YouTubeInbox (45) - samrat...mathData Connectors...AI for decision mak...machine learning fr...vision systemSKINOPATHYAI Infra - AWSOther bookmarks
Microsoft Azure | Data Factory | datafactory-twitter200503303@student.georgianc.on.ca
GEORGIAN COLLEGE

Copy Data tool

PropertiesSourceDatasetConfigurationDestinationSettingsReview and finish

Source data store

Specify the source data store for the copy task. You can use an existing data store connection or specify a new data store.

Source typeAzure Blob Storage

Connection *AzureBlobStorage1Edit+New connection

File or folder *
If the identity you use to access the data store only has permission to subdirectory instead of the entire account, specify the path to browse.
sqldatacontainer/Twitter_Data.csvBrowse

Options

Binary copy

Recursively

Enable partition discovery

Max concurrent connections

Filter by last modified

Start time (UTC)

End time (UTC)

< PreviousNext >

Cancel

AI Infra_C1 (samrat-ai-infrastruc...datafactory-twitter - Microsoft A...datafactory-twitter - Azure Data...+
adf.azure.com/en/home/copywizard?factory=%2Fsubscriptions%2F8c1bcaa6-91ae-435f-9dcb-25a67d95315f%2FresourceGroups%2FAI_Infrastructure%2Fproviders%2FMicrosoft.DataFactory%2Ffactories%2Fdatafactory-twitter
Imported From IE(2) FASTEST Way to...conversational AIethicsGeorgian College...YouTubeInbox (45) - samrat...mathData Connectors...AI for decision maki...machine learning fr...vision systemSKINOPATHYAI Infra - AWSOther bookmarks
Microsoft AzureData Factorydatafactory-twitter200503303@student.georgianc.on.caGEORGIAN COLLEGE

Copy Data tool

Properties

Source

Dataset

Configuration

Destination

Settings

Review and finish

File format settings

File format [ⓘ]
DelimitedTextDetect text formatPreview data

Column delimiter
Comma (,)
☐ Edit

Row delimiter
Line feed (\n)
☐ Edit

☒ First row as header [ⓘ]
> Advanced

Compression type
None

Additional columns [ⓘ]
 New

< Previous

Next >

Cancel

AI Infra_C1 (samrat-ai-infrastruc...datafactory-twitter - Microsoft A...datafactory-twitter - Azure Data

adf.azure.com/en/home/copywizard?factory=%2Fsubscriptions%2F8c1bcaa6-91ae-435f-9dcb-25a67d95315f%2FresourceGroups%2FAI_Infrastructure%2Fproviders%2FMicrosoft.DataFactory%2FFactories%2Fdatafactory-twitter

Imported From IE(2) FASTEST Way to...conversational AIethicsGeorgian College...YouTubeInbox (45) - samrat...mathData Connectors...AI for decision maki...machine learning fr...vision systemSKINOPATHYAI Infra. - AWSOther bookmarks

Microsoft AzureData Factorydatafactory-twitter200503303@student.georgianc.on.caGEORGIAN COLLEGE

Copy Data tool

Home

Source

Destination

Dataset

Configuration

Settings

Review and finish

Destination data store

Specify the destination data store for the copy task. You can use an existing data store connection or specify a new data store.

Destination type

All

Connection *

Select...

New connection

Previous

Next

New connection

Connection string

Azure Key Vault

Account selection method ⓘ

From Azure subscription

Enter manually

Azure subscription

Azure for Students (8c1bcaa6-91ae-435f-9dcb-25a67d95315f)

Server name *

samrat-ai-infrastructure

Database name *

master

Authentication type *

SQL authentication

User name *

samratSQL

Password

Azure Key Vault

Password *

Always encrypted ⓘ

Additional connection properties

+ New

Annotations

+ New

Parameters

Advanced ⓘ

Create

Back

Test connection

Cancel

Microsoft Azure | Data Factory > datafactory-twitter

Copy Data tool

Properties

Source

Destination

Dataset

Configuration

Settings

Review and finish

Destination data store

Specify the destination data store for the copy task. You can use an existing data store connection or specify a new data store.

Destination type

All

Connection *

AzureSqlDatabase2

Edit

New connection

Source

Twitter_Data

→

Target

Twitter_Data

(auto-create)

Use existing table

☐ Skip column mapping for all tables

< Previous

Next >

Cancel

Microsoft Azure

Data Factory

datafactory-twitter

200503303@student.georgianc.on.ca

GEORGIAN COLLEGE

Copy Data tool

Properties

Source

Destination

Settings


Review and finish

Review


Deployment

Summary

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

 Azure Blob Storage

→

 Azure SQL Database

Properties

Task name

CopyPipeline_j6m

Task description

Source

Connection name

AzureBlobStorage1

Dataset name

SourceDataset_j6m

Column delimiter

Row delimiter

Escape character

Quote char

First row as header

File name

Twitter_Data.csv

Container

sqldatacontainer

Destination

Connection name

AzureSqlDatabase2

Dataset name

DestinationDataset_j6m

< Previous

Next >

Cancel

Microsoft Azure

Data Factory

datafactory-twitter

200503303@student.georgianc.on.ca

GEORGIAN COLLEGE

Copy Data tool

Properties

Source


Destination


Settings


Review and finish

Review





Deployment

 Azure Blob Storage



 Azure SQL Database

Deployment complete

Deployment step	Status
Validating copy runtime environment	 Succeeded
> Creating datasets	 Succeeded
> Creating pipelines	 Succeeded
> Running pipelines	 Succeeded

Datasets and pipelines have been created. You can now monitor and edit the copy pipelines or click finish to close Copy Data Tool.

Finish

Edit pipeline

Monitor

QUERY EDITOR

The screenshot displays the Microsoft Azure portal interface. At the top, the browser address bar shows the URL: `portal.azure.com/#@georgiancollege.onmicrosoft.com/resource/subscriptions/8c1bcaa6-81ae-435f-9dcb-25a67d95315f/resourceGroups/AI_Infrastructure/providers/Microsoft.Sql/servers/samrat-ai-Infrastructure/databases/AI_Infra_C1`. The page title is "AI_Infra_C1 (samrat-ai-Infrastructure/AI_Infra_C1) | Query editor (preview)".

The interface is divided into several sections:

- Left Navigation Panel:** Contains links to Overview, Activity log, Tags, Diagnose and solve problems, Getting started, Query editor (preview) (selected), Settings, Compute + storage, Connection strings, Properties, Locks, Data management, Replicas, Sync to other databases, Integrations, Stream analytics (preview), Add Azure Search, Power Platform, Power BI, Power Apps, Power Automate, and Security.
- Top Bar:** Includes a search bar, a "Login" button, and links for "New Query", "Open query", and "Feedback".
- Query Editor:** The main area for writing and running queries. It includes a "Run" button, "Cancel query", "Save query", "Export data as", and "Show only Editor" options. The query text area is currently empty, with a line number "1" visible.
- Results Panel:** Located at the bottom, it has tabs for "Results" and "Messages". The "Results" tab is active, showing a search bar with the placeholder text "Search to filter items...".

The status bar at the bottom of the page indicates "Ready".

datafactory-twitter - Microsoft A...datafactory-twitter - Azure Data...AI_Infra_C1 (samrat-ai-infrastruc...+
portal.azure.com/#@georgiancollege.onmicrosoft.com/resource/subscriptions/8c1bcaa6-91ae-435f-9dcb-25a67d95315f/resourceGroups/AI_Infrastructure/providers/Microsoft.Sql/servers/samrat-ai-infrastructure/databases/AI_Infra_...
Imported From IE(2) FASTEST Way to...conversational AIethicsGeorgian College...YouTubeInbox (45) - samrat...mathData Connectors...AI for decision maki...machine learning fr...vision systemSKINOPATHYAI Infra - AWSOther bookmarks
Microsoft AzureSearch resources, services, and docs (0+)
200503303@student.ge...GEORGIAN COLLEGE (GEORGIAN...

Home > AI_Infra_C1 (samrat-ai-infrastructure/AI_Infra_C1)
AI_Infra_C1 (samrat-ai-infrastructure/AI_Infra_C1) | Query editor (preview)
SQL database

Search
LoginNew QueryOpen queryFeedback

Overview
Activity log
Tags
Diagnose and solve problems
Getting started
Query editor (preview)
Settings
Compute + storage
Connection strings
Properties
Locks
Data management
Replicas
Sync to other databases
Integrations
Stream analytics (preview)
Add Azure Search
Power Platform
Power BI
Power Apps
Power Automate
Security

AI_Infra_C1 (samratSQL)
Showing limited object explorer here. For full capability please click here to open Azure Data Studio.
Tables
dbo.Twitter_Data
Views
Stored Procedures

Query 1Query 2

RunCancel querySave queryExport data asShow only Editor

```
1 SELECT TOP (1000) * FROM [dbo].[Twitter_Data]
2 WHERE clean_text LIKE '%high%'
```

ResultsMessages
Search to filter items...

clean_text	category
yes good job highly insensitivearrogant incompetent ploar needs defeated any costnobody knows why mad...	1
higher voting turnout directly proportional bjp victory wonder modi has launched campaigns like dont sit ho...	1
modi came power because five yrs ago the youth india were takenin his promise create jobsyr fast forward fi...	1
malnutrition india and constituencies among worst affected mallikarjun kharge' gulbarga fared worst malnut...	-1
growth story indian economy during modi regime reflects serious concerns gst collections below the estimat...	-1
chowkidar hee chor hain baap chor beta bada chor aur unke guju dost maa tukdo kharidne jurrat bhi kartel...	1
unemployment rate years high male labor force shrunked since 2014 due fear being exposed modi didnt dar...	1
raghuram rajan sent list high profile bank fraudsters modi early months his office and thereafter some remin...	-1

Query succeeded | 31s

DATA QUERY WITH AZURE DATA STUDIO

The screenshot displays the Azure Data Studio interface. On the left, the 'CONNECTIONS' pane shows a tree view of the database structure, with 'dbo.Twitter_Data' selected. The main editor window shows a SQL query: `SELECT TOP (1000) [clean_text], [category] FROM [dbo].[Twitter_Data]`. Below the query, the 'Results' pane shows a table with two columns: 'clean_text' and 'category'. The table contains 19 rows of data. The status bar at the bottom indicates the current position in the query (Ln 1, Col 1) and the number of rows returned (1,000 rows).

SQLQuery_1 - samrat-ai-infrastructure.database.windows.net.AI_Infra_C1 (samratSQL) - Azure Data Studio

Run Cancel Disconnect Change Connection AI_Infra_C1

```
1 SELECT TOP (1000) [clean_text]
2     , [category]
3 FROM [dbo].[Twitter_Data]
```

Results Messages

	clean_text	category
1	when modi promised "minimum ...	-1
2	talk all the nonsense and co...	0
3	what did just say vote for m...	1
4	asking his supporters prefix...	1
5	answer who among these the m...	1
6	kiya tho refresh maarkefir c...	0
7	surat women perform yagna se...	0
8	this comes from cabinet whic...	0
9	with upcoming election india...	1
10	gandhi was gay does modi	1
11	things like demonetisation g...	1
12	hope tuthukudi people would ...	1
13	calm waters wheres the modi ...	1
14	one vote can make all the di...	0
15	one vote can make all the di...	0
16	vote such party and leadersh...	-1
17	vote modi who has not create...	0
18	through our vote ensure govt...	0
19	dont play with the words was...	1

Ln 1, Col 1 Spaces: 4 UTF-8 CRLF SQL 1,000 rows MSSQL 00:00:01 samrat-ai-infrastructure.database.windows.net : AI Infra C1

WITH WHERE CLAUSE

The screenshot displays the Azure Data Studio interface. On the left, the 'CONNECTIONS' pane shows a tree view of the database structure. The 'dbo' schema is expanded, and the 'Twitter_Data' table is selected. The 'Columns' folder is also expanded. The main editor shows a SQL query in the 'SQLQuery_1' window, which is connected to the 'AI_Infra_C1' database. The query is as follows:

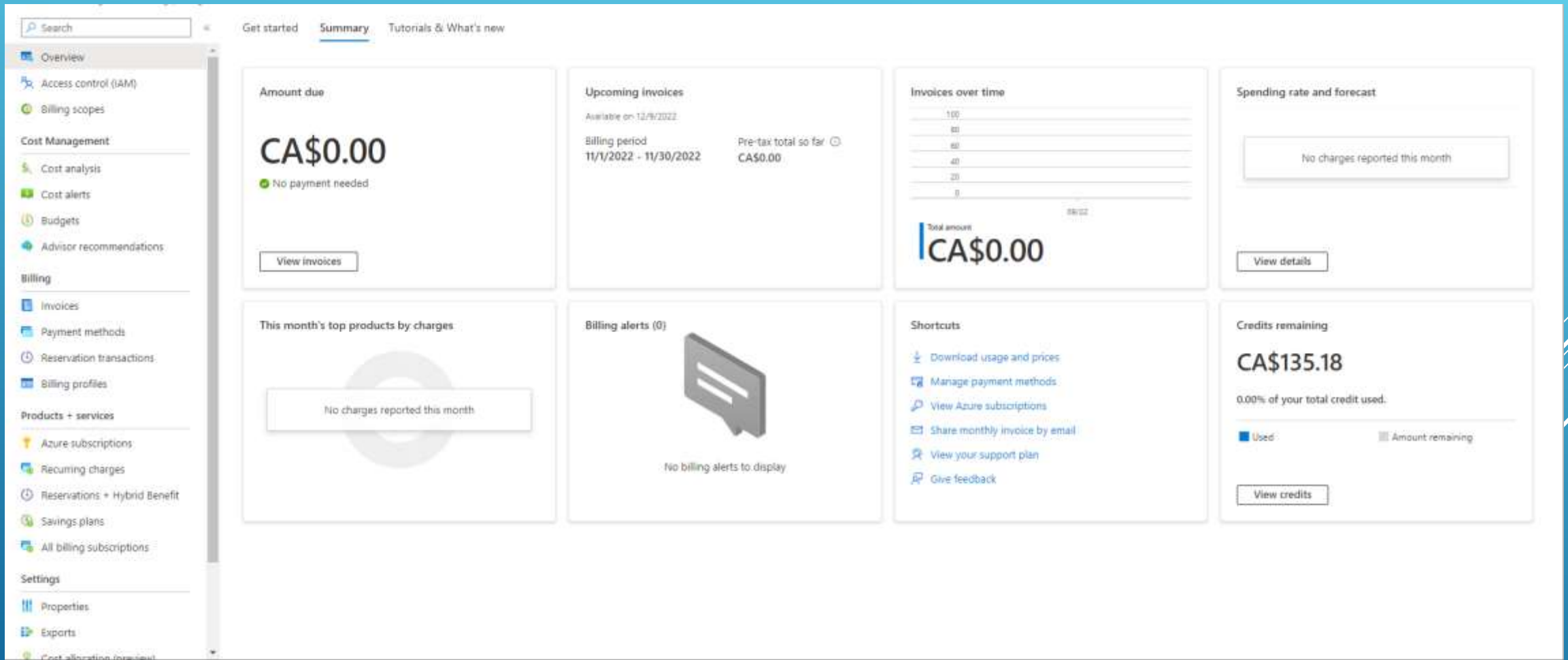
```
1 SELECT TOP (1000) [clean_text]
2     , [category]
3 FROM [dbo].[Twitter_Data]
4 WHERE clean_text LIKE '%high%'
```

The 'Results' pane at the bottom shows the output of the query, displaying 19 rows of data. The columns are 'clean_text' and 'category'.

	clean_text	category
1	yes good job highly insensit...	1
2	higher voting turnout direct...	1
3	modi came power because five...	1
4	malnutrition india and const...	-1
5	growth story indian economy ...	-1
6	chowkidar hee chor hain baap...	1
7	unemployment rate years high...	1
8	raghuran rajan sent list hig...	-1
9	modi while preparing manifes...	1
10	would like play the devils a...	1
11	modi govt controls the price...	1
12	1st all highest court our la...	1
13	why vote modi dynasty vision...	-1
14	today dollar highest ever mo...	1
15	raghuran rajan sent list hig...	-1
16	jawans killed the border cri...	-1
17	the last years the economist...	-1
18	the year high unemployment d...	-1
19	sapna based zirakpur mbl 964...	1

The status bar at the bottom indicates the current position in the query (Ln 4, Col 33) and the execution time (00:00:08).

SUBSCRIPTION CHECK



THANK YOU