

# PROJECT : Build an End-to-end data pipeline using Azure Data Factory

- **Authentication:** SQL Server Authentication
- **Server name:** nallagah1.database.windows.net

**MAIN LINK :** <https://witty-meadow-04aa4d610.5.azurestaticapps.net/>

## EXPLANATION

### 1. Objective:

The project aims to gather and present a "Student Count by Country" by extracting data from an SQL Server database. The primary objective is to showcase how SQL can be used to group data based on specific categories, in this case, the country of students, and then count the total number of students from each country.

### 2. Tools Used:

- SQL Server Management Studio (SSMS): Used to run SQL queries and manage the database.
- Azure SQL Database: The data is stored and processed in a cloud environment, utilizing Azure services.

### 3. Implementation Steps:

- Step 1: Connect to the Azure SQL Database via SQL Server Management Studio.
- Step 2: Query the student database to group data by country and count the number of students from each country using the SQL query.
- Step 3: Execute the query and review the results.

#### 4. Results:

The query returns a list of countries with the corresponding number of students. For example:

- Bangladesh: 1 student
- Brazil: 9 students
- Canada: 7 students
- China: 13 students
- Germany: 1 student
- A total of 31 rows of results were displayed, showing student counts from different countries.

## COST ANALYSIS

#### 1. Total Projected Monthly Cost:

The forecasted monthly cost for Azure services is **\$10.94**, based on usage trends as of September 2024.

#### 2. Breakdown of Azure Service Costs:

Storage: \$0.17

Azure Data Factory (v2): \$0.17

Bandwidth: < \$0.01

Azure App Service: < \$0.01

Functions: \$0.00

#### 3. Regional Costs:

- **US Central:** \$4.37
- **CA Central:** < \$0.01

## SS 1:- AZURE PORTAL LOGIN

The screenshot shows the Microsoft Azure Portal home page. The browser address bar displays `portal.azure.com/#home`. The page header includes the Microsoft Azure logo, a search bar, and the user's account information: `nallagah@mail.uc.edu` and `UNIVERSITY OF CINCINNATI (MA...`. The main content area is divided into two sections: **Azure services** and **Resources**.

**Azure services** section includes icons for:

- Create a resource
- SQL databases
- Data factories
- Storage accounts
- Reservations
- Quickstart Center
- Virtual machines
- App Services

**Resources** section includes a table of recent resources:

Name	Type	Last Viewed
akash	SQL database	6 days ago
stakash	Resource group	6 days ago
sqlserver111111	SQL server	6 days ago
sqlldb	SQL database	6 days ago
stakash	Storage account	6 days ago
jhsdcjdn	Data factory (V2)	6 days ago

The right sidebar shows the user's profile with the email `nallagah@mail.uc.edu` and links to [View account](#) and [Switch directory](#). A **Sign out** button is also present.

## SS 2:- CREATION OF STORAGE ACCOUNT

The screenshot shows the **Create a storage account** page in the Microsoft Azure Portal. The browser address bar displays `portal.azure.com/#create/Microsoft.StorageAccount`. The page header includes the Microsoft Azure logo, a search bar, and the user's account information: `nallagah@mail.uc.edu` and `UNIVERSITY OF CINCINNATI (MA...`.

The **Create a storage account** page is divided into several sections:

- Basics** (selected tab)
- Advanced**
- Networking**
- Data protection**
- Encryption**
- Tags**
- Review + create**

**Project details** section:

Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription \* `Azure for Students`

Resource group \* `(New) stakash1`

**Instance details** section:

Storage account name \* `stakash1`

Region \* `(US) Central US`

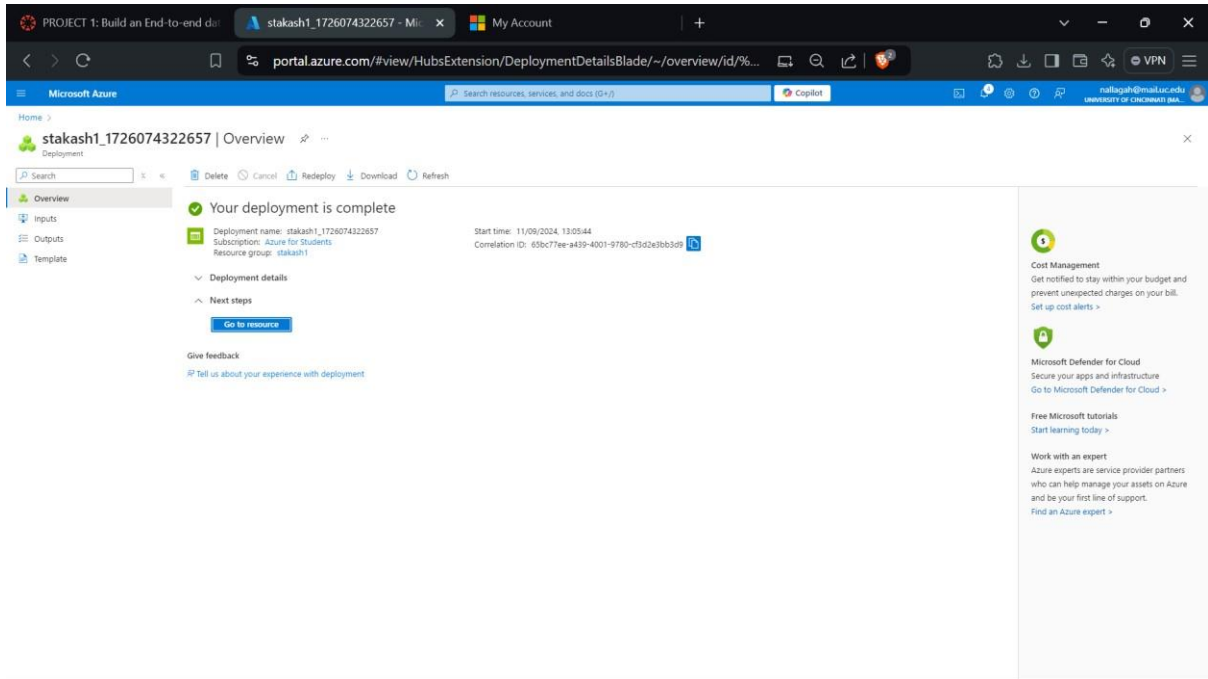
Primary service `Select a primary service`

Performance \* `Standard: Recommended for most scenarios (general-purpose v2 account)`

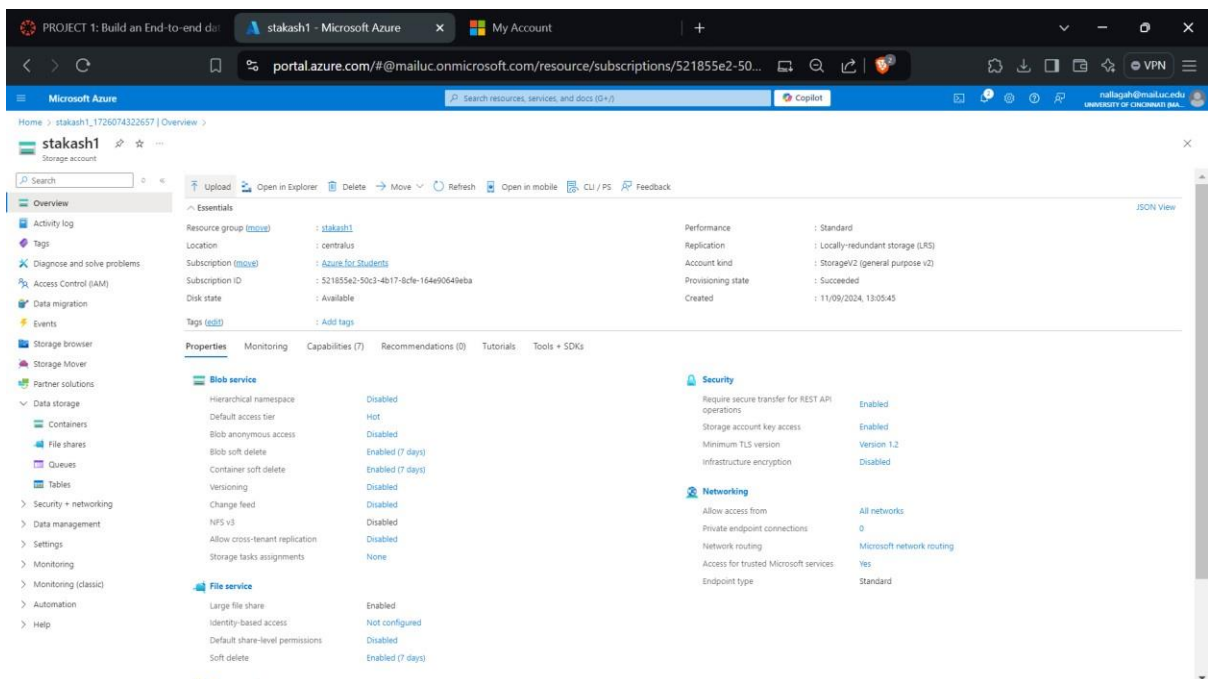
Redundancy \* `Locally-redundant storage (LRS)`

The bottom of the page features a navigation bar with **Previous**, **Next**, and **Review + create** buttons. A **Give feedback** link is also present.

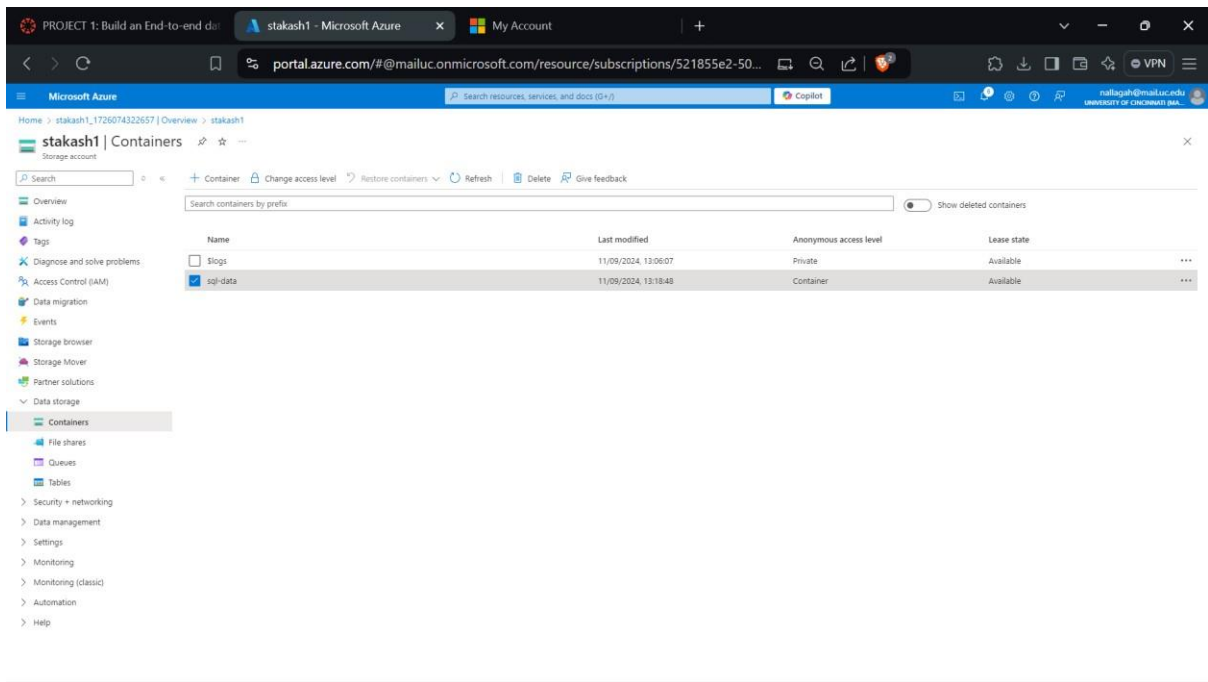
## SS 3:- DEPLOYMENT SCREEN OF STORAGE ACCOUNT



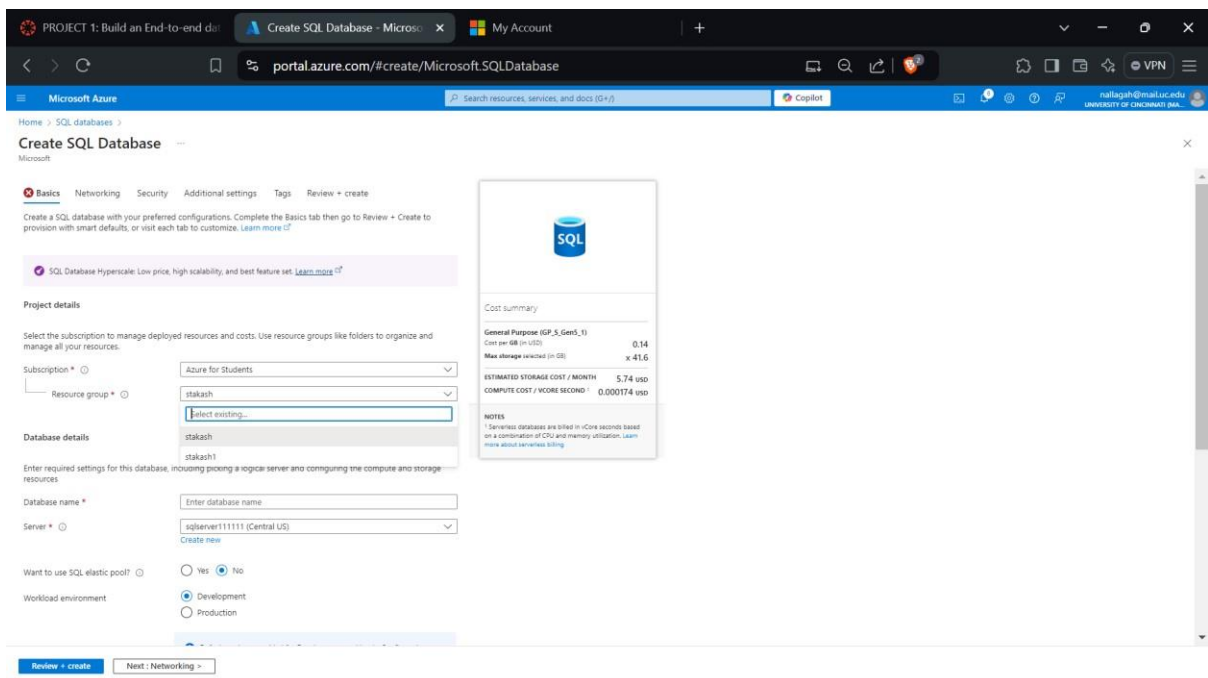
## SS 4:- CONFIGURATION SCREEN OF STORAGE ACCOUNT



## SS 5:- CREATION OF STORAGE CONTAINER



## SS 6:- CREATION OF SQL DATABASE



## SS 7:- DEPLOYMENT SCREEN

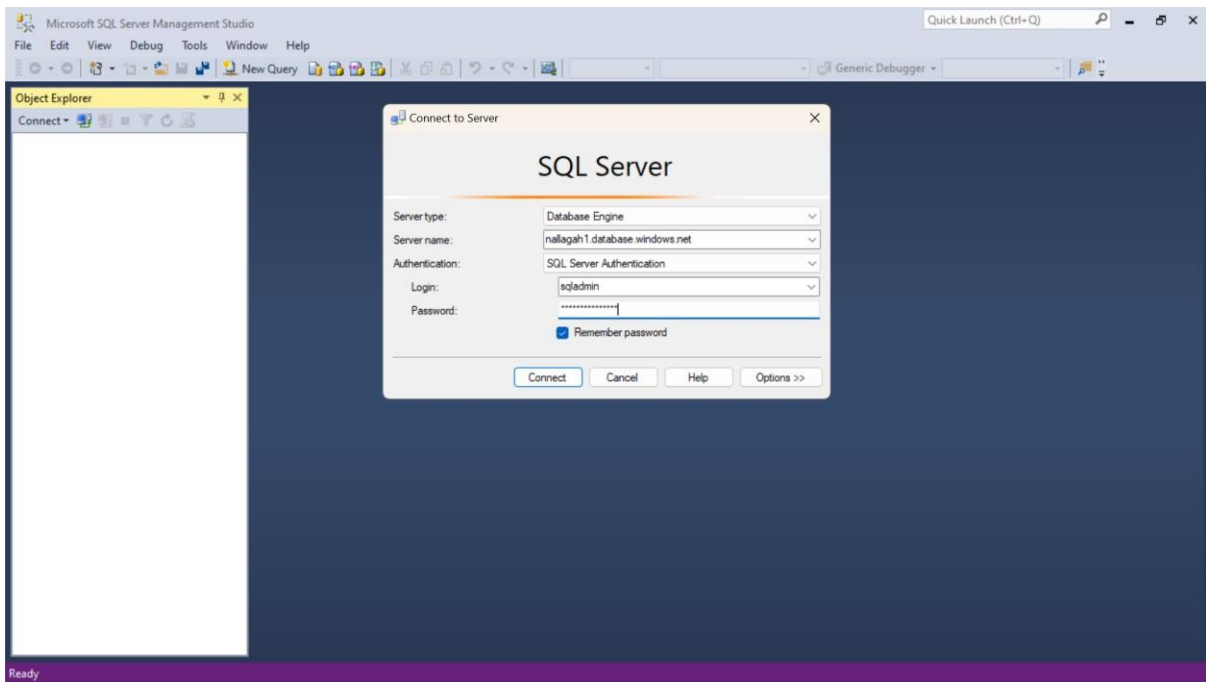
The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes the Microsoft Azure logo, a search bar, and user account information. The main content area displays the deployment details for a resource named 'Microsoft.SQLDatabase.newDatabaseExistingServer\_704c60d557264c6d'. The deployment is marked as 'Complete' with a green checkmark. The deployment details section shows the deployment name, subscription, resource group, start time, and correlation ID. The 'Next steps' section includes a 'Go to resource' button. The right sidebar contains several recommendations: 'Cost management', 'Microsoft Defender for Cloud', 'Free Microsoft tutorials', and 'Work with an expert'.

## SS 8:- SQL SERVER

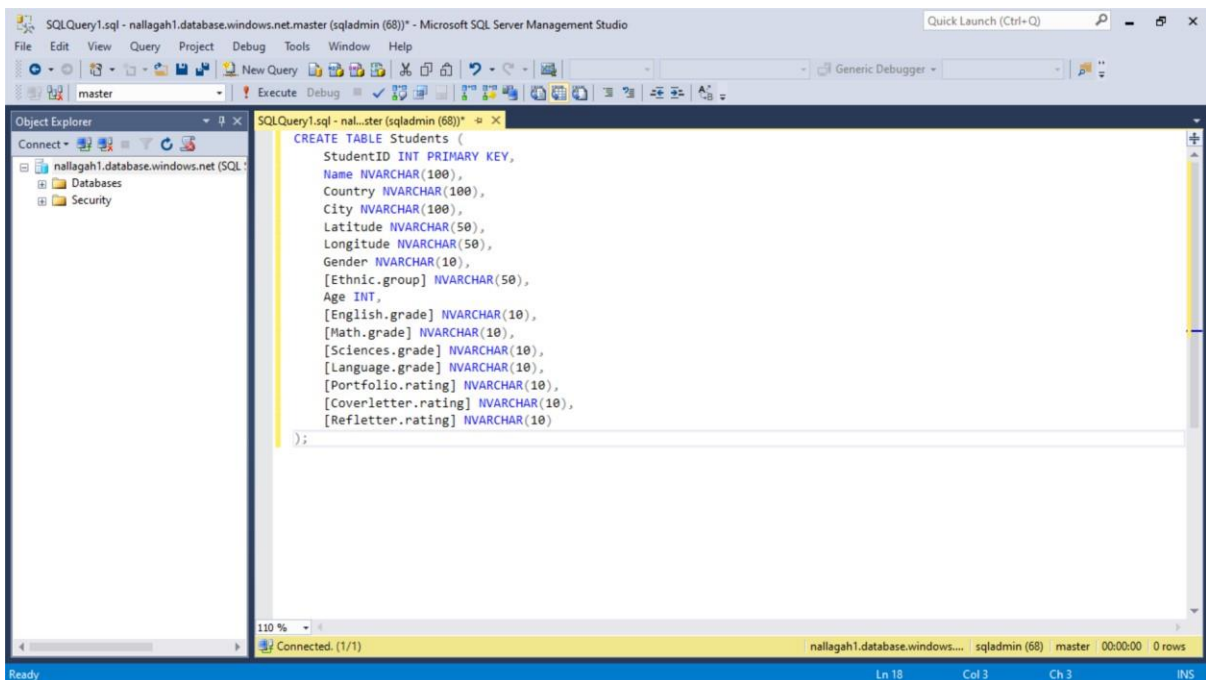
The screenshot shows the Microsoft Azure portal interface for the 'sqlserver11111' resource. The 'Networking' section is selected in the left sidebar. The 'Public network access' section is expanded, showing options to 'Disable' or 'Select networks'. The 'Firewall rules' section is also visible, showing a list of rules and a 'Add a firewall rule' button. A notification banner at the top right indicates 'Successfully updated server firewall rules'.

Rule	Virtual network	Subnet	Address range	Endpoint status	Resource group	Subscription	State
ClientAddress_2024-9-11_15-53-35			208.102.162.194				

## SS 9:- SERVER LOGIN



## SS 10:- CREATION OF TABLE



## SS 11:- AZURE DATA FACTORY STUDIO CR4EATION

The screenshot shows the 'Create Data Factory' wizard in the Azure portal. The browser address bar shows 'portal.azure.com/#create/Microsoft.DataFactory'. The page title is 'Create Data Factory'. The 'Basics' tab is selected, showing options for subscription, resource group, name, region, and version. The 'Subscription' dropdown is set to 'Azure for Students'. The 'Resource group' dropdown is set to 'Create new'. The 'Name' field is empty. The 'Region' dropdown is set to 'East US'. The 'Version' dropdown is set to 'V2'. At the bottom, there are buttons for 'Previous', 'Next', and 'Review + create', along with a 'Give feedback' link.

PROJECT 1: Build an End-to-end data pipeline | Create Data Factory - Microsoft Azure | My Account | ChatGPT

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

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

Dashboard > Data factories > Create Data Factory

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

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

Project details

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

Subscription \* Azure for Students

Resource group \* Create new

Instance details

Name \*

Region \* East US

Version \* V2

Previous Next Review + create

Give feedback

## SS 12:- ADF DASHBOARD

The screenshot shows the Azure Data Factory Studio dashboard. The browser address bar shows 'portal.azure.com/#@mailuc.onmicrosoft.com/resource/subscriptions/521855...'. The page title is 'akash-sql-data Data factory (V2)'. The 'Overview' tab is selected, showing details about the data factory, including its resource group, status, location, subscription, and subscription ID. The 'Launch studio' button is prominent. Below the 'Launch studio' button, there are four tiles: 'Quick Starts', 'Tutorials', 'Template Gallery', and 'Training Modules'. At the bottom, there are three monitoring charts: 'PipelineRuns', 'ActivityRuns', and 'TriggerRuns', each showing a line graph of runs over time.

PROJECT 1: Build an End-to-end data pipeline | akash-sql-data - Microsoft Azure | My Account | ChatGPT

portal.azure.com/#@mailuc.onmicrosoft.com/resource/subscriptions/521855...

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

Dashboard > Microsoft.DataFactory-20240911160253 | Overview >

akash-sql-data Data factory (V2)

Search Delete

Overview | Activity log | Access control (IAM) | Tags | Diagnose and solve problems | Settings | Getting started | Monitoring | Automation | Help

Essentials

Resource group (most) : akash1

Status : Succeeded

Location : Central US

Subscription (most) : Azure for Students

Subscription ID : 52185562-50c3-4b17-8cfe-164e9049eba

Type : Data factory (V2)

Getting started : Quick start

Azure Data Factory Studio

Launch studio

Quick Starts Tutorials Template Gallery Training Modules

Monitoring

PipelineRuns ActivityRuns TriggerRuns



## SS 13:- CREATION OF BLOB STORAGE

The screenshot shows the Microsoft Azure Data Factory portal. The left sidebar contains navigation options: General, Factory settings, Connections, Linked services (selected), Integration runtimes, Microsoft Purview, Source control, Git configuration, ARM template, Author, Triggers, Global parameters, Data flow libraries, Security, Credentials, Customer managed key, Outbound rules, and Managed private endpoints. The main pane is titled 'Linked services' and includes a '+ New' button, a 'Filter by name' input, and an 'Annotations: Any' button. Below this, a table lists the linked services:

Name	Type	Related	Annotations
AzureBlobStorage1	Azure Blob Storage	0	

## SS 14:- CREATION OF AZURE SQL DATA

This screenshot shows the same Microsoft Azure Data Factory portal as the previous one, but with an additional linked service. The table now displays two items:

Name	Type	Related	Annotations
AzureBlobStorage1	Azure Blob Storage	0	
AzureSQLLinkedService	Azure SQL Database	0	

## SS 15:- CREATION (STUDENTCSV) FILE

The screenshot displays the Microsoft Azure Data Factory portal interface. The left sidebar shows the 'Factory Resources' tree with 'studentcsv' selected under 'Datasets'. The main pane shows the 'studentcsv' dataset configuration, which is a 'DelimitedText' dataset. The 'Import schema' button is visible, and a table of columns is shown below it. The right sidebar shows the 'Properties' pane for the dataset.

Column name	Type
StudentID	String
Name	String
Country	String
City	String
Latitude	String
Longitude	String
Gender	String
Ethnic group	String
Age	String
English grade	String
Math grade	String
Sciences grade	String
Languages grade	String

## SS 16:- CREATION OF PIPELINE AND DEBUGGING.

The screenshot displays the Microsoft Azure Data Factory portal interface. The left sidebar shows the 'Factory Resources' tree with 'CopyStudentDataPipeline' selected under 'Pipelines'. The main pane shows the pipeline configuration, which is a 'Copy data' pipeline. The right sidebar shows the 'Properties' pane for the pipeline. Below the pipeline configuration, the 'Pipeline status' is shown as 'Succeeded'. A table of pipeline runs is displayed at the bottom.

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User property
Copy data1	Succeeded	Copy data	9/11/2023 5:09:46 PM	13s	AutoResolveIntegration	

# FINAL RESULT

## SS 17:- QUERYING (STUDENT COUNT BY COUNTRY)

The screenshot displays the Microsoft SQL Server Management Studio interface. The 'Object Explorer' on the left shows the database structure for 'nallagahh'. The 'Query Editor' in the center contains the following SQL query:

```
SELECT Country, COUNT(*) AS StudentCount
FROM Students
GROUP BY Country;
```

The 'Results' pane at the bottom shows the output of the query, which is a table with two columns: 'Country' and 'StudentCount'. The table contains 11 rows of data, with 'China' having the highest student count at 13.

Country	StudentCount
Bangladesh	1
Brazil	9
Canada	7
Chile	1
China	13
Colombia	5
Cuba	1
Dominican Republic	1
Egypt	1
El Salvador	1
Germany	1

The status bar at the bottom indicates that the query was executed successfully, returning 31 rows in 00:00:00 seconds.

## FUNCTION APP

## SS1:- CREATE FUNCTION APP

Microsoft Azure

Home > Function App > Create Function App >

### Create Function App (Consumption)

Basics Storage Networking Monitoring Deployment Tags Review + create

Create a function app, which lets you group functions as a logical unit for easier management, deployment and sharing of resources. Functions lets you execute your code in a serverless environment without having to first create a VM or publish a web application.

**Project Details**

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

Subscription \* Azure for Students

Resource Group \* STORAGEAK

Create new

**Instance Details**

Function App name \* webindexz .azurewebsites.net

Runtime stack \* Node.js

Version \* 20 LTS

Region \* Canada Central

Operating System \* ☐ Linux ☒ Windows

Review + create Previous Next: Storage >

## SS2:- CREATE HTTP TRIGGER TEMPLATE.

Microsoft Azure

Home > webindexz >

### HttpTrigger1 | Code + Test

Code + Test Integration Function Keys Invocations Logs Metrics

Save Discard Refresh Test/Run Get function URL Disable Delete Upload Resource JSON Send us your feedback

webindexz / HttpTrigger1 / index.js

```
9  ...options: {
10  ...  encrypt: true // For Azure SQL
11  ...}
12  ...}
13  ...}
14  const pool = new ConnectionPool(config);
15  try {
16    await pool.connect();
17    const result = await pool.request();
18    .query("SELECT Country, COUNT(*) AS StudentCount FROM Students GROUP BY Country");
19  }
20  context.res = {
21    status: 200,
22    body: result.recordset
23  };
24  catch (err) {
25    context.res = {
26      status: 500,
27      body: err.message
28    };
29  } finally {
30    pool.close();
31  }
32  }
33  }
```

Logs

Connected! You are now viewing logs of Function runs in the current Code + Test panel. To see all the logs for this Function, please go to 'Logs' from the Function menu.

2024-09-21T18:21:13Z [Information] Executing 'Functions.HttpTrigger1' (Reason: 'This function was programmatically called via the host APIs.', Id=191cfcc4-8b8e-4d80-b88d-11405012fefe)

2024-09-21T18:21:13Z [Verbose] Sending invocation id: '191cfcc4-8b8e-4d80-b88d-11405012fefe'

2024-09-21T18:21:13Z [Verbose] Putting invocation id:191cfcc4-8b8e-4d80-b88d-11405012fefe on workerId:b50e7e42-8e9b-4e18-b170-03e2894ec4af

2024-09-21T18:21:13Z [Information] Javascript HTTP trigger function processed a request.

2024-09-21T18:21:13Z [Information] Executed 'Functions.HttpTrigger1' (Succeeded, Id=191cfcc4-8b8e-4d80-b88d-11405012fefe, Duration=78ms)

**Test/Run**

Input Output

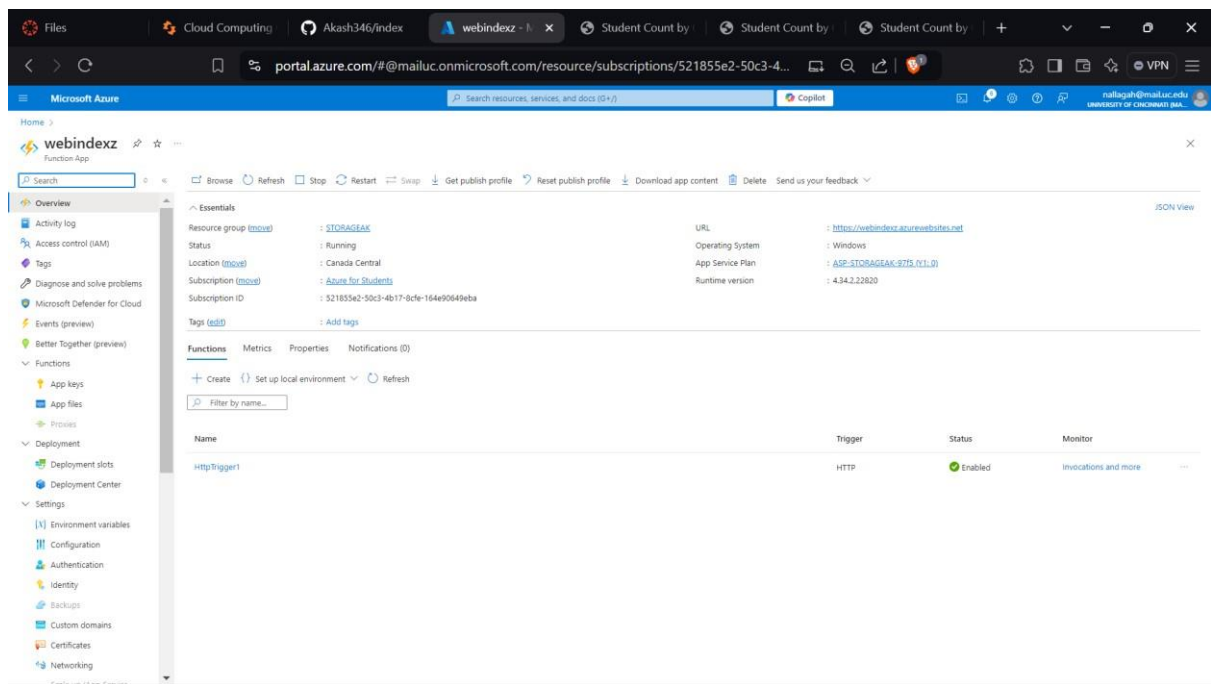
HTTP response code 200 OK

HTTP response content

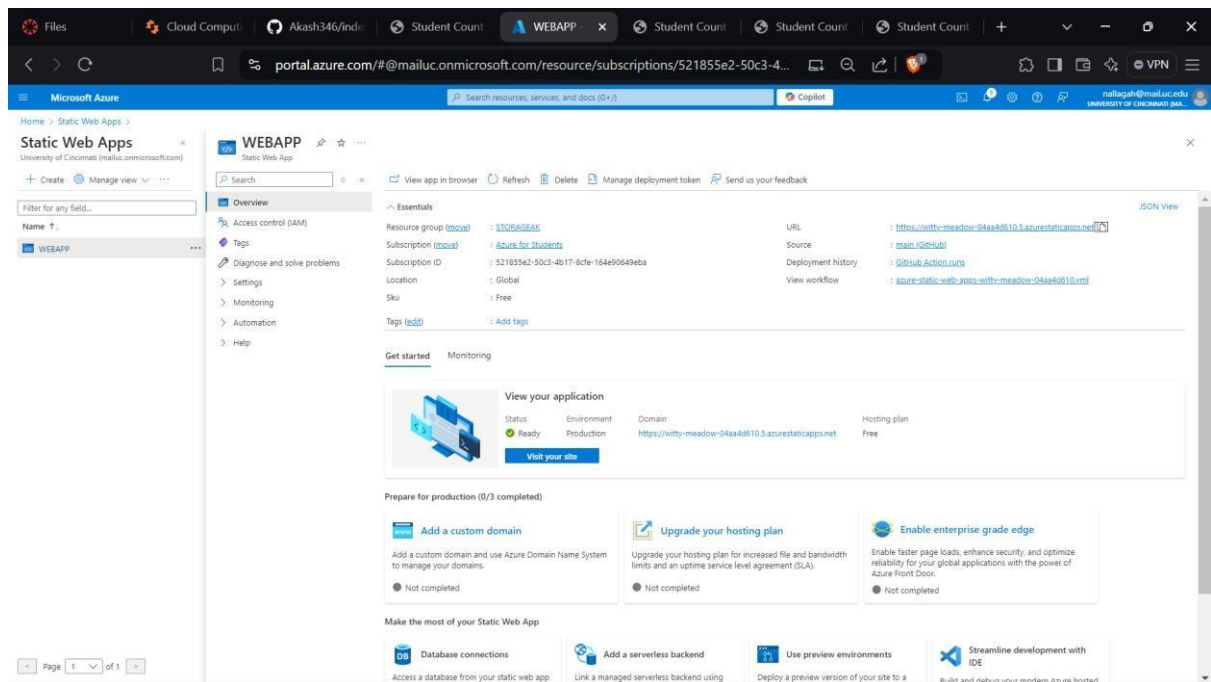
Hello, Azure. This HTTP triggered function executed successfully.

Run Close

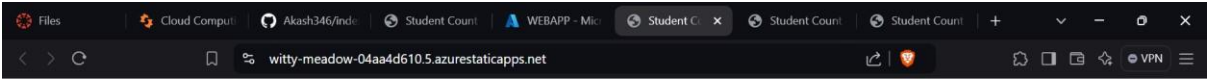
## SS3:- DEPLOYED HTTP TRIGGER TEMPLATE



## SS4:- STATIC WEBAPP



## SS5:- FINAL RESULT(STUDENT COUNT BY COUNTRY IN A LINK)



## Student Count by Country

Country	Student Count
USA	1200
India	800
China	600
Germany	400
Brazil	300

