**Install Python**

* Go to Python official website download Python 3.11.3

<https://www.python.org/downloads/>

* To install required Python Packages for MS SQL parquet file extraction

pip install -r requirements.txt.

**Azure Setup**

* Create a resource group in the azure portal:
* Select **Resource groups**
* Select **Add**
* Enter the values
* Select **Create**.
* Create a Azure Key Vault:
* On the Key Vault section, choose **Create**.
* On the **Create key vault** section select created resource group under **Resource Group**
* Select create.
* Create a Storage Account
* Select **Storage accounts** and select Create**.**
* Register an app with Azure Active Directory:
* In the Azure portal, select **Azure Active Directory** .
* Select **App registrations** and click on **New registration**.
* Enter your application's registration information.
* Click on **Register** to create the application.
* Create a Container
* Click the Storage Account.
* Click on the "Containers" and select **“+Container"** tab to create container.
* You need to assign a name and select the access level then click ok.
* Assign a Key Vault access policy
* Navigate to the key vault and select **Access policies**, then select **Create**
* Select the permissions under **Secret permissions.** Least permissions to read Key Vault secret are ‘List’ and ‘Get’.
* Under the **Principal** selection pane, enter the name of the app in the search field and select the appropriate result.
* select **Create** to save the access policy.

**Parquet Generator Config**

|  |  |
| --- | --- |
| **EDW\_DB\_SERVER\_HOST** | EDW database server name |
| **EDW\_DB\_NAME** | EDW database name |
| **GP\_DB\_SERVER\_HOST** | GP database server name |
| **KEY\_VAULT\_NAME** | Name of the Azure key vault. |
| **EDW\_KEY\_VAULT\_SECRET\_NAME\_DB\_USER\_NAME** | Secret name for EDW database username. |
| **EDW\_KEY\_VAULT\_SECRET\_NAME\_DB\_PASSWORD** | Secret name for EDW database password |
| **GP\_KEY\_VAULT\_SECRET\_NAME\_DB\_ USER\_NAME** | Secret name for GP database username |
| **GP\_KEY\_VAULT\_SECRET\_NAME\_DB\_PASSWORD** | Secret name for GP database password |
| **MAX\_JOB\_COUNT** | Enter the Max job count to make the parallel execution |
| **QUERY\_GET\_ALL\_TABLES\_FOR\_DB** | SQL query to get all tables from GP database |
| **QUERY\_GET\_ALL\_RECORDS\_FROM\_TABLE** | SQL query to fetch all records from the table GP database |
| **AZURE\_STORAGE\_ACCOUNT\_CONNECTION\_STRING** | Connection string for the azure storage account |
| **AZURE\_STORAGE\_ACCOUNT\_CONTAINER\_NAME** | Container name in the azure storage account |

**How to Run Python**

Follow this command to run python file in command prompt.

* Python <File Path\Filename>

Example command:

python gp\_data\_extractor.py

**Screen Blob Storage**

