

Project outputs:

CSV Data Source: Total count: 1004480

<https://www.kaggle.com/datasets/ksabishek/massive-bank-dataset-1-million-rows>

CSV Data stored in BLOB storage:

Blob containers > iubhblobcontainer

Authentication method: Access key (Switch to Microsoft Entra user account)

Add filter

Search blobs by prefix (case-sensitive)

Only show

Showing all 4 items

	Name	Last modified	Access tier	Blob type	Size
<input type="checkbox"/>	db				
<input type="checkbox"/>	dtable				
<input type="checkbox"/>	bankdataset...	4/5/2024, 12:46:44 PM	Hot (Inferred)	Block blob	38.89 MiB
<input type="checkbox"/>	dtablebankdataset.csv	2024, 1:33:02 PM	Hot (Inferred)	Block blob	0

CSV Data stored in PostgreSQL storage:

Query Query History

```
1 SELECT "Id", "Date", "Domain", "Location", "Value", "Transaction_count"
2   FROM iudeschema.bankdata limit 10;
3
4 SELECT count(*)
5   FROM iudeschema.bankdata;
```

Data output Messages Notifications

	Id bigint	Date text	Domain text	Location text	Value text	Transaction_count text
1	164869	01-03-2022	MEDICAL	Konark	872803	646
2	164870	01-03-2022	RETAIL	Bombay	1175425	1533
3	164871	01-03-2022	RESTRAU...	Mon	392852	1673
4	164872	01-03-2022	MEDICAL	Goa	914532	1644
5	164873	01-03-2022	PUBLIC	Doda	601091	1520
6	164874	01-03-2022	INVESTM...	Doda	948989	2064
7	164875	01-03-2022	EDUCATI...	Ajmer	466961	803
8	164876	01-03-2022	EDUCATI...	Ludhiana	875016	1630
9	164877	01-03-2022	EDUCATI...	Ara	335327	1867
10	164878	01-03-2022	MEDICAL	Madurai	627959	933

PostgreSQL storage data count:

The screenshot shows the Databricks SQL interface. At the top, the user is logged in as `iu_iubhstudent@AzureServer`. Below the header, there are icons for file management, a dropdown menu, a filter icon, and a "No limit" dropdown. The main area is divided into two tabs: "Query" (selected) and "Query History". The query editor contains the following SQL code:

```
1 SELECT "Id", "Date", "Domain", "Location",
2     FROM iudeschema.bankdata limit 10;
3
4 SELECT count(*)
5     FROM iudeschema.bankdata;
```

Below the query editor, there are three tabs: "Data output" (selected), "Messages", and "Notifications". The "Data output" tab shows a table with two columns: "count" and "bigint". The first row of data shows a count of 1004480.

	count	bigint
1	1004480	

Parquet data stored on BLOB Azure:

Blob containers > iuhblobcontainer > dtable

Authentication method: Access key ([Switch to Microsoft Entra user account](#))

Search blobs by prefix (case-sensitive)

Showing all 4 items

	Name	Last modified	Access tier	Blob type	Size
	[..]				
	_SUCCESS	4/7/2024, 1:33:02 PM	Hot (Inferred)	Block blob	0
	_committed...	4/7/2024, 1:33:02 PM	Hot (Inferred)	Block blob	122 B
	_started_80...	4/7/2024, 1:33:00 PM	Hot (Inferred)	Block blob	0
<input checked="" type="checkbox"/>	part-00000-...	4/7/2024, 1:33:02 PM	Hot (Inferred)	Block blob	12.76 MiB

Overview

Blob: dtable/part-00000-tid-8016091927049604135-7adf44ff-02b6-4e67-9c62-73a154223644-0-1.c000.snappy.parquet

Save Discard Download Refresh Delete Change tier Acquire lease

Properties

URL	https://iubhblobdb.blob...
LAST MODIFIED	4/7/2024, 1:33:02 PM
CREATION TIME	4/7/2024, 1:33:00 PM
VERSION ID	-
TYPE	Block blob
SIZE	12.76 MiB
ACCESS TIER	Hot (Inferred)
ACCESS TIER LAST MODIFIED	N/A
ARCHIVE STATUS	-
REHYDRATE PRIORITY	-
SERVER ENCRYPTED	true
ETAG	0x8DC56F6C13A583
VERSION LEVEL IMMUTABILITY POLICY	Disabled

