

1)  
USE DATABASE BIGDATA\_GITHUB;  
CREATE OR REPLACE SCHEMA analytics;

The screenshot shows a SQL IDE interface. The top bar indicates the database is 'BIGDATA\_GITHUB.ANALYTICS'. The query editor contains two lines of SQL: 'USE DATABASE BIGDATA\_GITHUB;' and 'CREATE OR REPLACE SCHEMA analytics;'. Below the editor, the 'Results' tab is active, showing a single row with the status 'Schema ANALYTICS successfully created.' The right sidebar displays 'Query Details' with a progress bar for 'Query duration' and 'Rows'. The 'Query ID' is '01bc279c-0305-2'. A 'Show more' button is visible. At the bottom right, there is an 'Ask Cop' button.

2)  
-- Table 1: Repositories by Commit Volume  
CREATE OR REPLACE TABLE BIGDATA\_GITHUB.analytics.most\_active\_repos (  
    repo\_name STRING,  
    total\_commits INT  
);

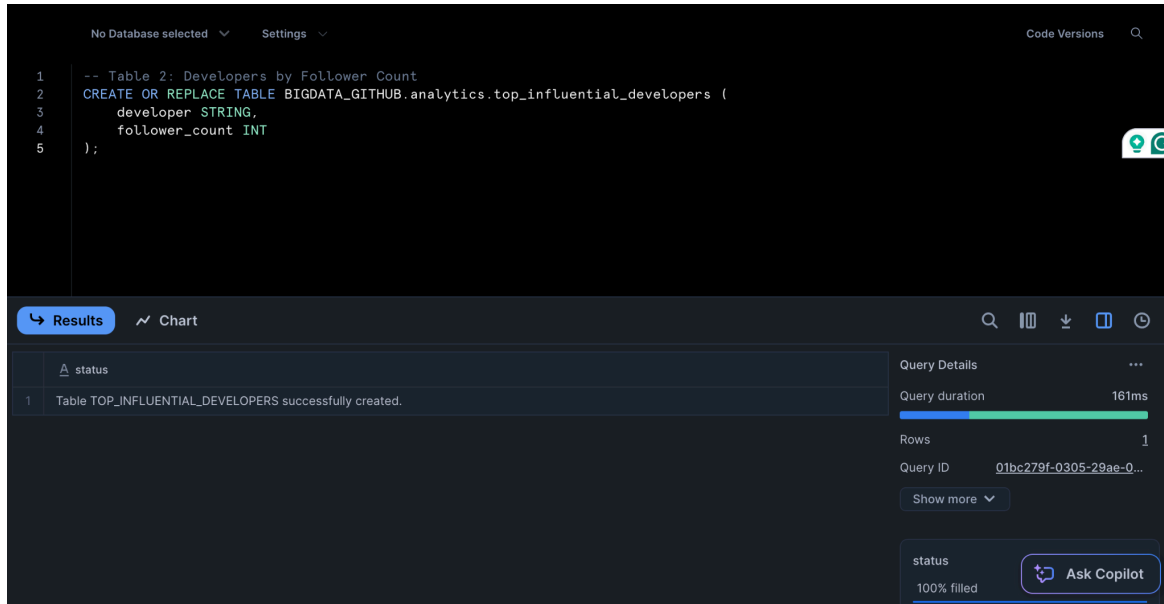
The screenshot shows a SQL IDE interface. The top bar indicates 'No Database selected'. The query editor contains five lines of SQL: a comment '-- Table 1: Repositories by Commit Volume', followed by 'CREATE OR REPLACE TABLE BIGDATA\_GITHUB.analytics.most\_active\_repos (', then 'repo\_name STRING,', 'total\_commits INT', and finally ');'. Below the editor, the 'Results' tab is active, showing a single row with the status 'Table MOST\_ACTIVE\_REPOS successfully created.' The right sidebar displays 'Query Details' with a progress bar for 'Query duration' and 'Rows'. The 'Query ID' is '01bc279e-0305-29ae'. A 'Show more' button is visible. At the bottom right, there is an 'Ask Cop' button.

3)

-- Table 2: Developers by Follower Count

```
CREATE OR REPLACE TABLE BIGDATA_GITHUB.analytics.top_influential_developers (  
    developer STRING,  
    follower_count INT
```

);



4)

```
CREATE OR REPLACE STORAGE INTEGRATION azure_github_int  
    TYPE = EXTERNAL_STAGE  
    STORAGE_PROVIDER = AZURE  
    ENABLED = TRUE  
    AZURE_TENANT_ID = 'e85c5307-76b1-4c48-bc5d-e88373dda261'  
    STORAGE_ALLOWED_LOCATIONS =  
( 'azure://matthewleffler1.blob.core.windows.net/kaggle-datasets' );
```

```
DESC INTEGRATION azure_github_int;
```

```
CREATE OR REPLACE STAGE github_stage  
    URL='azure://matthewleffler1.blob.core.windows.net/kaggle-datasets'
```

```
CREDENTIALS=(AZURE_SAS_TOKEN='sp=rl&st=2025-04-30T06:56:51Z&se=2025-05-14T14:  
56:51Z&spr=https&sv=2024-11-04&sr=c&sig=FHnuulJLF1zV67cmCem1l2EKOjV6REiliVR%2B  
HQ9nXz4%3D')  
    FILE_FORMAT = (TYPE = PARQUET);
```

```
LIST @github_stage/analytics/most_active_repos/;
```

BIGDATA\_GITHUB.ANALYTICS
Settings
Code Versions

```

1 CREATE OR REPLACE STORAGE INTEGRATION azure_github_int
2   TYPE = EXTERNAL_STAGE
3   STORAGE_PROVIDER = AZURE
4   ENABLED = TRUE
5   AZURE_TENANT_ID = 'e85c5307-76b1-4c48-bc5d-e88373dda261'
6   STORAGE_ALLOWED_LOCATIONS = ('azure://matthewleffler1.blob.core.windows.net/kaggle-datasets');
7
8 DESC INTEGRATION azure_github_int;
9
10 CREATE OR REPLACE STAGE github_stage
11   URL='azure://matthewleffler1.blob.core.windows.net/kaggle-datasets'
12   CREDENTIALS=(AZURE_SAS_TOKEN='sp=r&st=2025-04-30T06:56:51Z&se=2025-05-14T14:56:51Z&spr=https&sv=2024-11-04&sr=c&sig=FHnuuIJLF1zV67cmCem1L2EK0jV6REil1VR%2BHQ9nXz4%3D')
13   FILE_FORMAT = (TYPE = PARQUET);
14
15 LIST @github_stage/analytics/most_active_repos/;
16

```

Results
Chart

	name	# size	md5	last_modified
1	azure://matthewleffler1.blob.core.windows.net/kaggle-dataset	0	null	Mon, 5 May 2025 22:30:47 GMT
2	azure://matthewleffler1.blob.core.windows.net/kaggle-dataset	2077	null	Mon, 5 May 2025 22:30:46 GMT

Query Details

Query duration 129ms

Rows 2

Query ID 01bc...

Show more

Ask Copilot

5)

```

COPY INTO BIGDATA_GITHUB.analytics.most_active_repos
FROM @github_stage/analytics/most_active_repos/
FILE_FORMAT = (TYPE = PARQUET)
MATCH_BY_COLUMN_NAME = CASE_INSENSITIVE
PATTERN = '.*.parquet';

```

```

SELECT * FROM BIGDATA_GITHUB.analytics.most_active_repos LIMIT 10;

```

BIGDATA\_GITHUB.ANALYTICS
Settings
Code Versions

```

1 COPY INTO BIGDATA_GITHUB.analytics.most_active_repos
2 FROM @github_stage/analytics/most_active_repos/
3 FILE_FORMAT = (TYPE = PARQUET)
4 MATCH_BY_COLUMN_NAME = CASE_INSENSITIVE
5 PATTERN = '.*.parquet';
6
7 SELECT * FROM BIGDATA_GITHUB.analytics.most_active_repos LIMIT 10;
8

```

Results
Chart

	REPO_NAME	# TOTAL_COMMITS
1	cirosantilli/imagine-all-the-people	60425
2	thiagodecarlo/cocoapods	37637
3	muxi/grpc-pod	36739
4	enkidevs/commit	36373
5	CocoaPods/Specs	32990
6	yangxiangwei/YXWSecond	16046
7	MicrosoftDocs/azure-docs	15478
8	wangxingneng/Specs	14449
9	jlord/patchwork	10482
10	Yann-Fornier/patchwork	10215

Query Details

Query duration 325ms

Rows 10

Query ID 01bc28ed-0305-2812-0...

Show more

REPO\_NAME

100% filled

TOTAL\_COMMITS

10 Queries

Status All

Mon May 5 2025

10:01:32 PM 325ms SELECT \* FROM BIGDATA...

10:01:19 PM 45ms SELECT \* FROM github.a...

10:01:00 PM 1.3s COPY INTO BIGDATA\_GITH...

9:59:58 PM 78ms COPY INTO BIGDATA\_GITH...

9:58:37 PM 128ms COPY INTO BIGDATA\_GITH...

9:58:17 PM 69ms COPY INTO BIGDATA\_GITH...

9:58:05 PM 136ms COPY

5:06:40 PM 37ms CREA

Ask Copilot

6)

```
COPY INTO BIGDATA_GITHUB.analytics.top_influential_developers
FROM @github_stage/analytics/top_influential_developers/
FILE_FORMAT = (TYPE = PARQUET)
MATCH_BY_COLUMN_NAME = CASE_INSENSITIVE
PATTERN = '.*.parquet';
```

```
SELECT * FROM BIGDATA_GITHUB.analytics.top_influential_developers LIMIT 10;
```

The screenshot displays a data warehouse interface with a dark theme. At the top, a SQL query is entered in a text area. Below the query, the 'Results' tab is active, showing a table with 10 rows of data. The table has two columns: 'DEVELOPER' and 'FOLLOWER\_COUNT'. To the right of the table, 'Query Details' are shown, including the query duration (343ms) and the number of rows (10). Further right, a '2 Queries' panel shows a list of executed queries with their timestamps and durations. At the bottom right, there is an 'Ask Copilot' button.

	DEVELOPER	FOLLOWER_COUNT
1	hongyangAndroid	10583
2	singwhatiwanna	6263
3	offensive-security	5300
4	hackedteam	4843
5	evilcos	3267
6	slidenerd	3154
7	yangyangwithgnu	2917
8	botelho	2663
9	litesuits	2643
10	hussien89aa	2624

**Query Details**

- Query duration: 343ms
- Rows: 10
- Query ID: 01bc28f0-0305-1f93-0...

**2 Queries**

- 10:04:33 PM 343ms SELECT \* FROM BIGDATA\_...
- 10:04:17 PM 1.1s COPY INTO BIGDATA\_GITH...

**Ask Copilot**