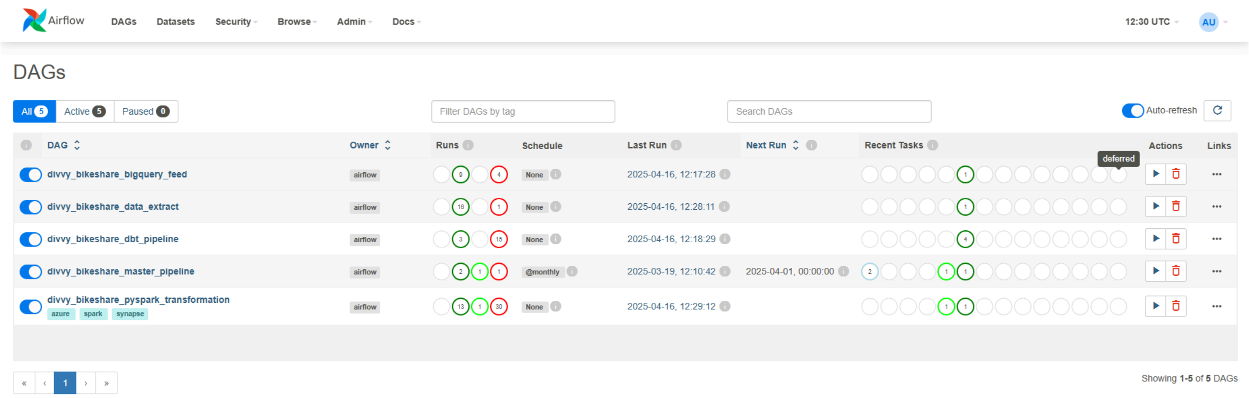
Divvy Bikeshare Analytical Project: Complete pipeline representation

We have five separate pipelines:

- divvy_bikeshare_data_extract
- divvy_bikeshare_pyspark_transformation
- divvy_bikeshare_bigquery_feed
- divvy_bikeshare_dbt_pipeline
- divvy_bikeshare_master_pipeline



DAGs

All 5 | Active 5 | Paused 0

Filter DAGs by tag

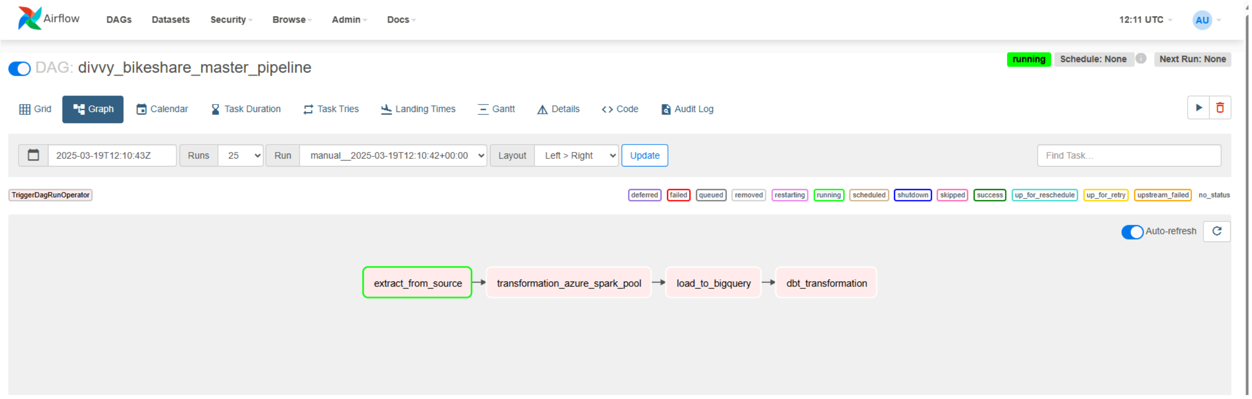
Search DAGs

Auto-refresh

DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks
divvy_bikeshare_bigquery_feed	airflow	9 4	None	2025-04-16, 12:17:28		1
divvy_bikeshare_data_extract	airflow	16 1	None	2025-04-16, 12:28:11		1
divvy_bikeshare_dbt_pipeline	airflow	3 15	None	2025-04-16, 12:18:29		4
divvy_bikeshare_master_pipeline	airflow	2 1 1	@monthly	2025-03-19, 12:10:42	2025-04-01, 00:00:00	2 1 1
divvy_bikeshare_pyspark_transformation azure spark synapse	airflow	13 1 30	None	2025-04-16, 12:29:12		1 1

Showing 1-5 of 5 DAGs

The data pipeline is orchestrated by a master DAG “divvy_bikeshare_master_pipeline” that coordinates the execution of several sub-DAGs in sequence.



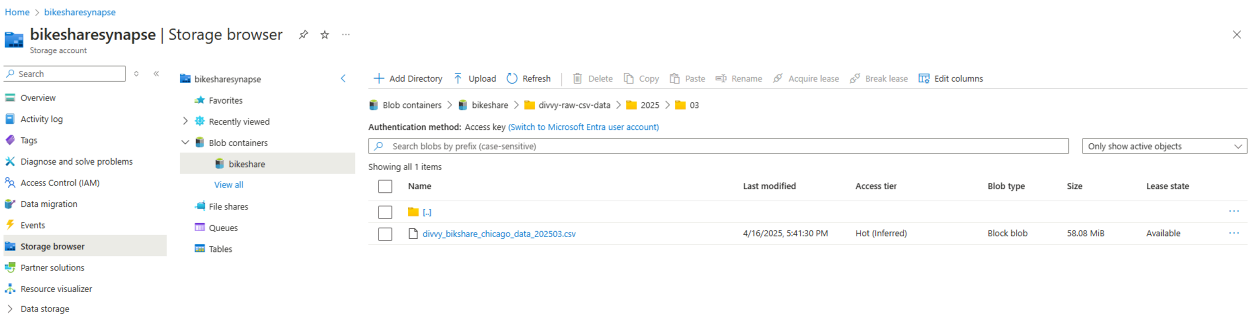
DAG: divvy_bikeshare_master_pipeline

running | Schedule: None | Next Run: None

2025-03-19T12:10:43Z | Runs 25 | Run manual__2025-03-19T12:10:42+00:00 | Layout Left > Right | Update

Find Task...

divvy_bikeshare_data_extract : Downloads monthly Divvy Bikeshare data from S3, extracts ZIP files and standardizes naming and uploads raw CSV data to Azure Blob Storage



Home > bikesharesynapse

bikesharesynapse | Storage browser

Storage account

Blob containers > bikeshare > divvy-raw-csv-data > 2025 > 03

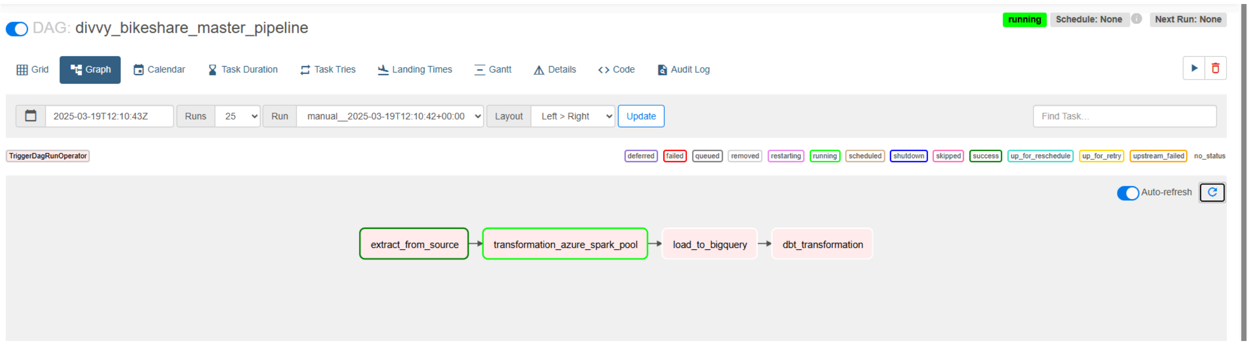
Authentication method: Access key (Switch to Microsoft Entra user account)

Search blobs by prefix (case-sensitive) | Only show active objects

Showing all 1 items

☐	Name	Last modified	Access tier	Blob type	Size	Lease state
☐	[..]					
☐	divvy_bikshare_chicago_data_202503.csv	4/16/2025, 5:41:30 PM	Hot (Inferred)	Block blob	58.08 MiB	Available

divvy_bikeshare_pyspark_transformation is triggered by this and it submits a pyspark job in synapse analytics spark pool.

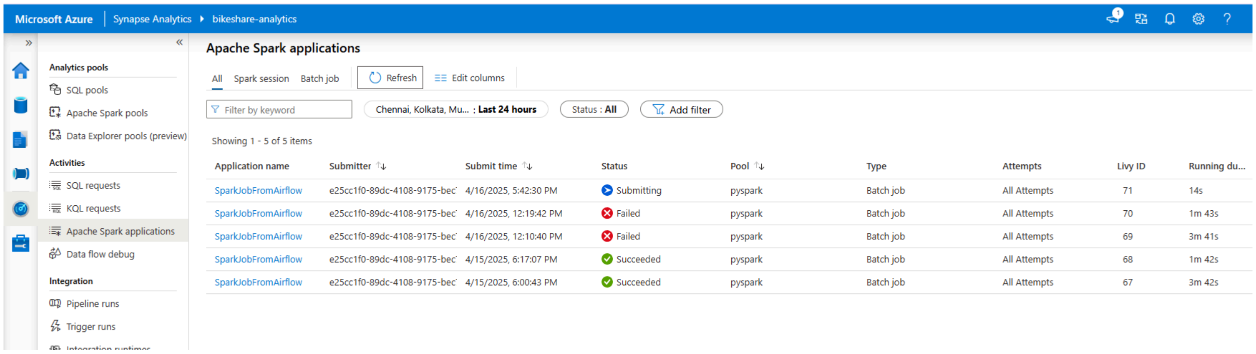


DAG: divvy_bikeshare_master_pipeline

running | Schedule: None | Next Run: None

2025-03-19T12:10:43Z | Runs 25 | Run manual__2025-03-19T12:10:42+00:00 | Layout Left > Right | Update | Find Task...

Executes PySpark jobs for data transformation



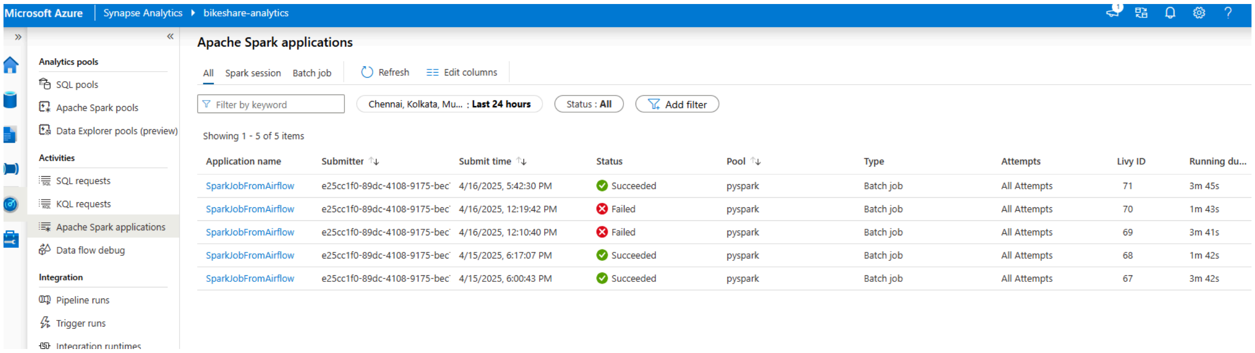
Microsoft Azure | Synapse Analytics ▸ bikeshare-analytics

Apache Spark applications

Chennai, Kolkata, Mu... : **Last 24 hours** | Status : **All**

Showing 1 - 5 of 5 items

Application name	Submitter	Submit time	Status	Pool	Type	Attempts	Livy ID	Running du...
SparkJobFromAirflow	e25cc1f0-89dc-4108-9175-bec	4/16/2025, 5:42:30 PM	Submitting	pyspark	Batch job	All Attempts	71	14s
SparkJobFromAirflow	e25cc1f0-89dc-4108-9175-bec	4/16/2025, 12:19:42 PM	Failed	pyspark	Batch job	All Attempts	70	1m 43s
SparkJobFromAirflow	e25cc1f0-89dc-4108-9175-bec	4/16/2025, 12:10:40 PM	Failed	pyspark	Batch job	All Attempts	69	3m 41s
SparkJobFromAirflow	e25cc1f0-89dc-4108-9175-bec	4/15/2025, 6:17:07 PM	Succeeded	pyspark	Batch job	All Attempts	68	1m 42s
SparkJobFromAirflow	e25cc1f0-89dc-4108-9175-bec	4/15/2025, 6:00:43 PM	Succeeded	pyspark	Batch job	All Attempts	67	3m 42s



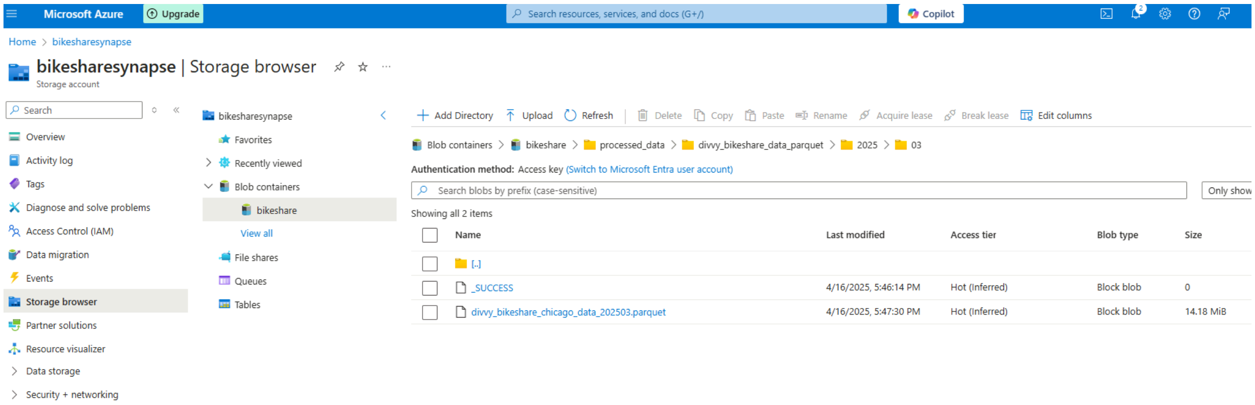
Apache Spark applications

Chennai, Kolkata, Mu... : **Last 24 hours** | Status : **All**

Showing 1 - 5 of 5 items

Application name	Submitter	Submit time	Status	Pool	Type	Attempts	Livy ID	Running du...
SparkJobFromAirflow	e25cc1f0-89dc-4108-9175-bec	4/16/2025, 5:42:30 PM	✓ Succeeded	pyspark	Batch job	All Attempts	71	3m 45s
SparkJobFromAirflow	e25cc1f0-89dc-4108-9175-bec	4/16/2025, 12:19:42 PM	✗ Failed	pyspark	Batch job	All Attempts	70	1m 43s
SparkJobFromAirflow	e25cc1f0-89dc-4108-9175-bec	4/16/2025, 12:10:40 PM	✗ Failed	pyspark	Batch job	All Attempts	69	3m 41s
SparkJobFromAirflow	e25cc1f0-89dc-4108-9175-bec	4/15/2025, 6:17:07 PM	✓ Succeeded	pyspark	Batch job	All Attempts	68	1m 42s
SparkJobFromAirflow	e25cc1f0-89dc-4108-9175-bec	4/15/2025, 6:00:43 PM	✓ Succeeded	pyspark	Batch job	All Attempts	67	3m 42s

After the job is succeeded in synapse analytics, it saves data in Parquet format.



bikesharesynapse | Storage browser

Storage account

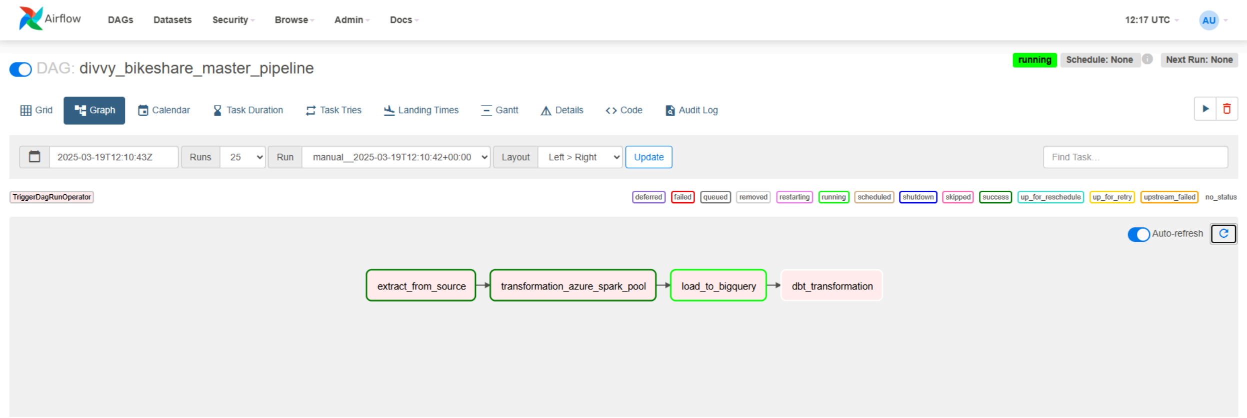
Blob containers > bikeshare > processed_data > divvy_bikeshare_data_parquet > 2025 > 03

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

Showing all 2 items

□	Name	Last modified	Access tier	Blob type	Size
□	[..]				
□	_SUCCESS	4/16/2025, 5:46:14 PM	Hot (Inferred)	Block blob	0
□	divvy_bikeshare_chicago_data_202503.parquet	4/16/2025, 5:47:30 PM	Hot (Inferred)	Block blob	14.18 MiB

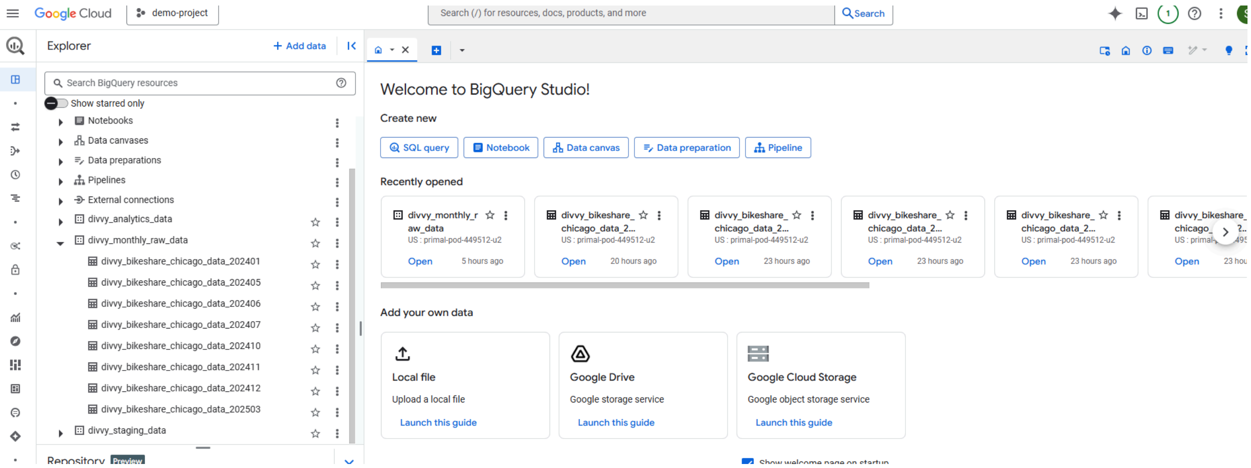
Divvy_bikeshare_bigquery_feed gets triggered it renames processed Parquet files with standardized naming and it loads transformed data into BigQuery



DAG: divvy_bikeshare_master_pipeline

2025-03-19T12:10:43Z | Runs 25 | Run manual__2025-03-19T12:10:42+00:00 | Layout Left > Right

extract_from_source → transformation_azure_spark_pool → load_to_bigquery → dbt_transformation

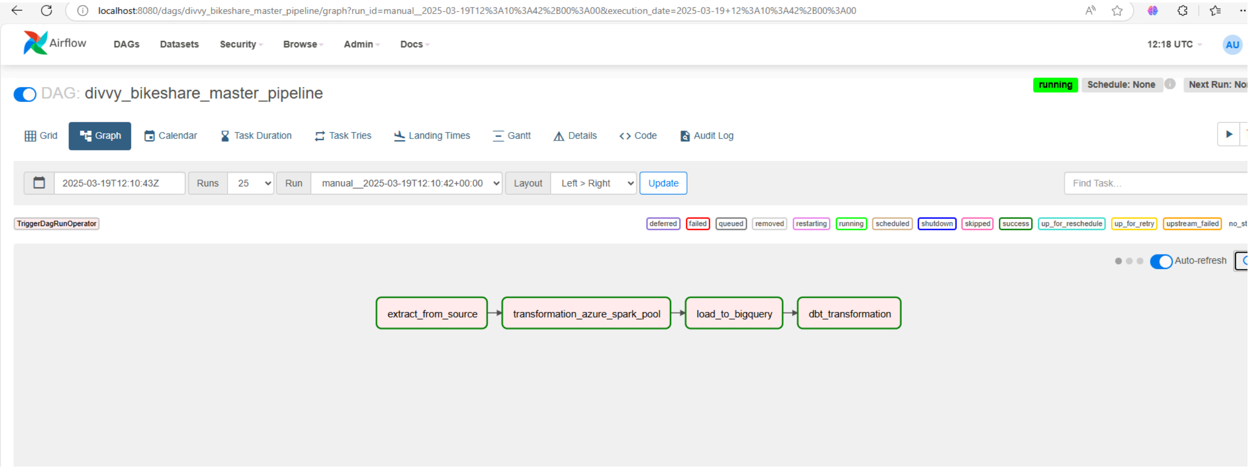


Welcome to BigQuery Studio!

- divvy_analytics_data
- divvy_monthly_raw_data
  - divvy_bikeshare_chicago_data_202401
  - divvy_bikeshare_chicago_data_202405
  - divvy_bikeshare_chicago_data_202406
  - divvy_bikeshare_chicago_data_202407
  - divvy_bikeshare_chicago_data_202410
  - divvy_bikeshare_chicago_data_202411
  - divvy_bikeshare_chicago_data_202412
  - divvy_bikeshare_chicago_data_202503
- divvy_staging_data

divvy_monthly_raw_data	US : primal-pod-449512-u2	5 hours ago
divvy_bikeshare_chicago_data_2...	US : primal-pod-449512-u2	20 hours ago
divvy_bikeshare_chicago_data_2...	US : primal-pod-449512-u2	23 hours ago
divvy_bikeshare_chicago_data_2...	US : primal-pod-449512-u2	23 hours ago
divvy_bikeshare_chicago_data_2...	US : primal-pod-449512-u2	23 hours ago
divvy_bikeshare_chicago_2...	US : primal-...495	23 hou

It triggers the final DAG “divvy_bikeshare_dbt_pipeline”

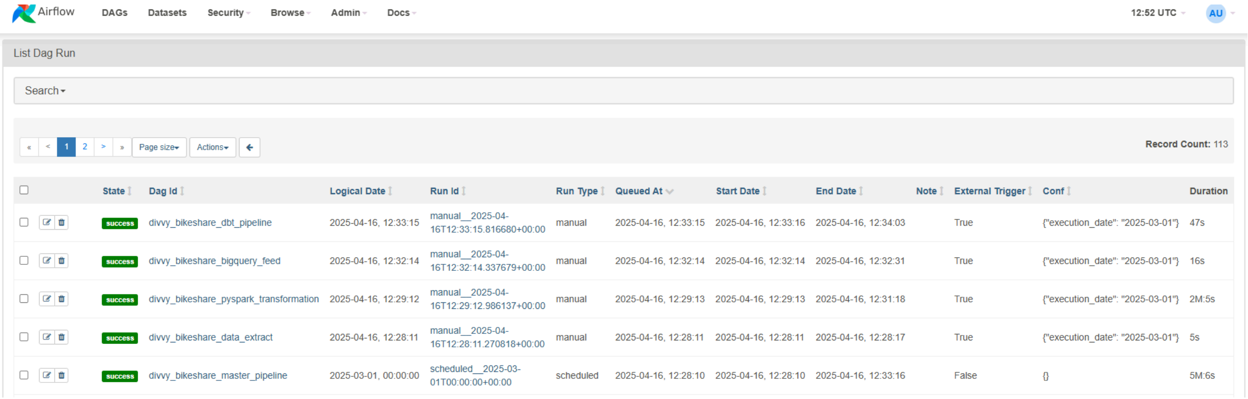


localhost:8080/dags/divvy_bikeshare_master_pipeline/graph?run_id=manual__2025-03-19T12%3A10%3A42%2B00%3A00&execution_date=2025-03-19+12%3A10%3A42%2B00%3A00

DAG: divvy_bikeshare_master_pipeline

running Schedule: None Next Run: No

2025-03-19T12:10:43Z Runs 25 Run manual__2025-03-19T12:10:42+00:00 Layout Left > Right

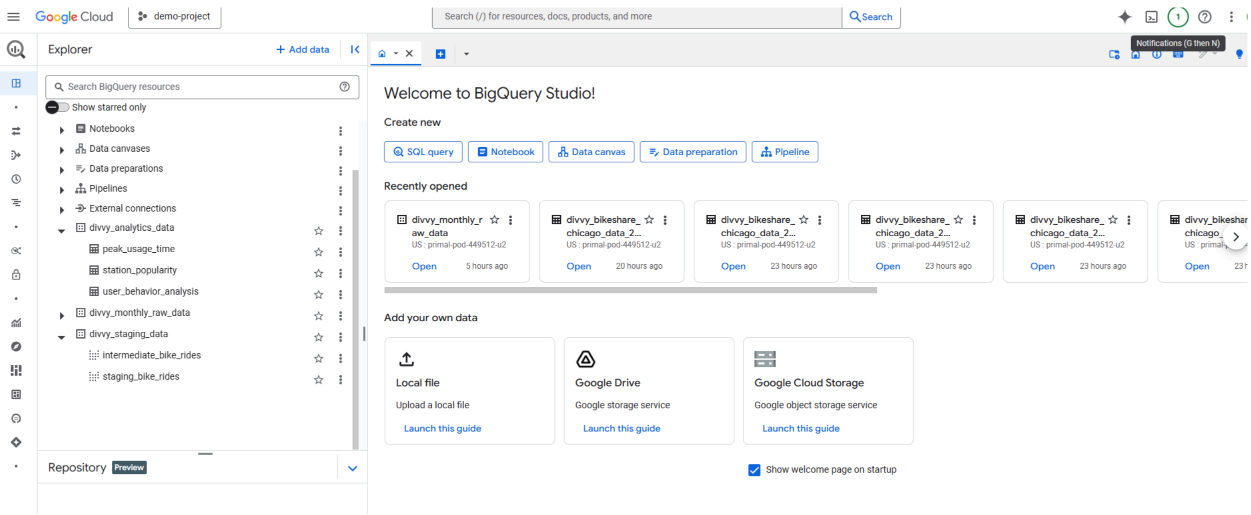


List Dag Run

Record Count: 113

State	Dag Id	Logical Date	Run Id	Run Type	Queued At	Start Date	End Date	Note	External Trigger	Conf	Duration
success	divvy_bikeshare_dbt_pipeline	2025-04-16, 12:33:15	manual__2025-04-16T12:33:15.816680+00:00	manual	2025-04-16, 12:33:15	2025-04-16, 12:33:16	2025-04-16, 12:34:03		True	{"execution_date": "2025-03-01"}	47s
success	divvy_bikeshare_bigquery_feed	2025-04-16, 12:32:14	manual__2025-04-16T12:32:14.337679+00:00	manual	2025-04-16, 12:32:14	2025-04-16, 12:32:14	2025-04-16, 12:32:31		True	{"execution_date": "2025-03-01"}	16s
success	divvy_bikeshare_pyspark_transformation	2025-04-16, 12:29:12	manual__2025-04-16T12:29:12.986137+00:00	manual	2025-04-16, 12:29:13	2025-04-16, 12:29:13	2025-04-16, 12:31:18		True	{"execution_date": "2025-03-01"}	2M:5s
success	divvy_bikeshare_data_extract	2025-04-16, 12:28:11	manual__2025-04-16T12:28:11.270818+00:00	manual	2025-04-16, 12:28:11	2025-04-16, 12:28:11	2025-04-16, 12:28:17		True	{"execution_date": "2025-03-01"}	5s
success	divvy_bikeshare_master_pipeline	2025-03-01, 00:00:00	scheduled__2025-03-01T00:00:00+00:00	scheduled	2025-04-16, 12:28:10	2025-04-16, 12:28:10	2025-04-16, 12:33:16		False	{}	5M:6s

It runs dbt models with dynamic variables and applies business logic transformations and updates in bigquery.



Google Cloud | demo-project | Search (/) for resources, docs, products, and more | Search

Notifications (G then N)

Explorer | + Add data

Search BigQuery resources

Show starred only

- Notebooks
- Data canvases
- Data preparations
- Pipelines
- External connections
- divvy_analytics_data
  - peak_usage_time
  - station_popularity
  - user_behavior_analysis
- divvy_monthly_raw_data
- divvy_staging_data
  - intermediate_bike_rides
  - staging_bike_rides

Repository Preview

Welcome to BigQuery Studio!

Create new

SQL query | Notebook | Data canvas | Data preparation | Pipeline

Recently opened

divvy_monthly_raw_data	US : primal-pod-449512-u2	Open	5 hours ago
divvy_bikeshare_chicago_data_2...	US : primal-pod-449512-u2	Open	20 hours ago
divvy_bikeshare_chicago_data_2...	US : primal-pod-449512-u2	Open	23 hours ago
divvy_bikeshare_chicago_data_2...	US : primal-pod-449512-u2	Open	23 hours ago
divvy_bikeshare_chicago_data_2...	US : primal-pod-449512-u2	Open	23 hours ago
divvy_bikeshar chicago_	US : primal-	Open	23 h

Add your own data

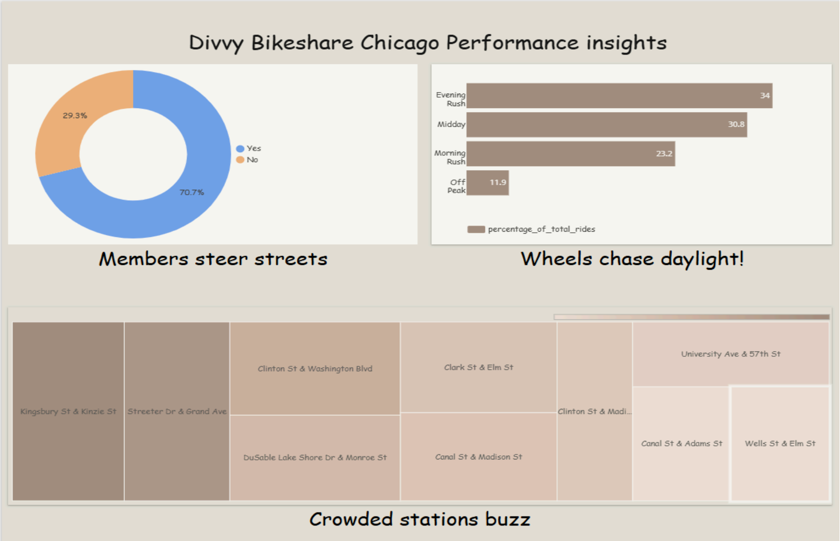
Local file
Upload a local file
Launch this guide

Google Drive
Google storage service
Launch this guide

Google Cloud Storage
Google object storage service
Launch this guide

☑ Show welcome page on startup

It finally updates the dashboard hosted on looker:



Divvy Bikeshare Chicago Performance insights

Members steer streets

Wheels chase daylight!

Crowded stations buzz