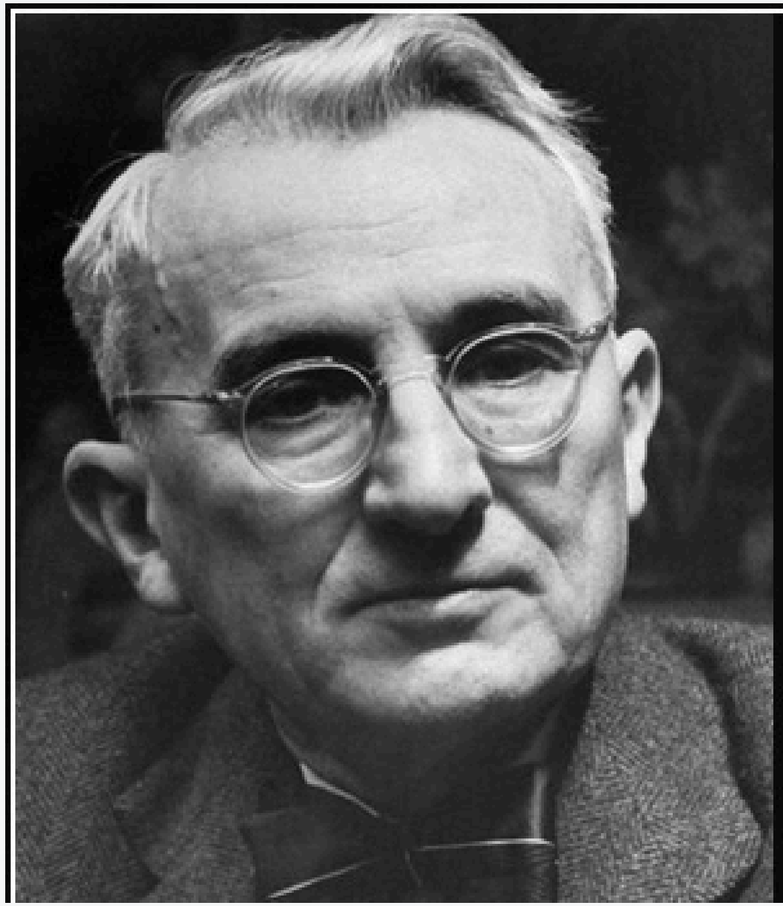
GRADO Y MÁSTER EN INGENIERÍA Y GESTIÓN DE SISTEMAS DE INFORMACIÓN - ÁREA DE ESPECIALIZACIÓN EN INGENIERÍA DE SISTEMAS DE INFORMACIÓN

🤖 TECHNOLOGY CASE

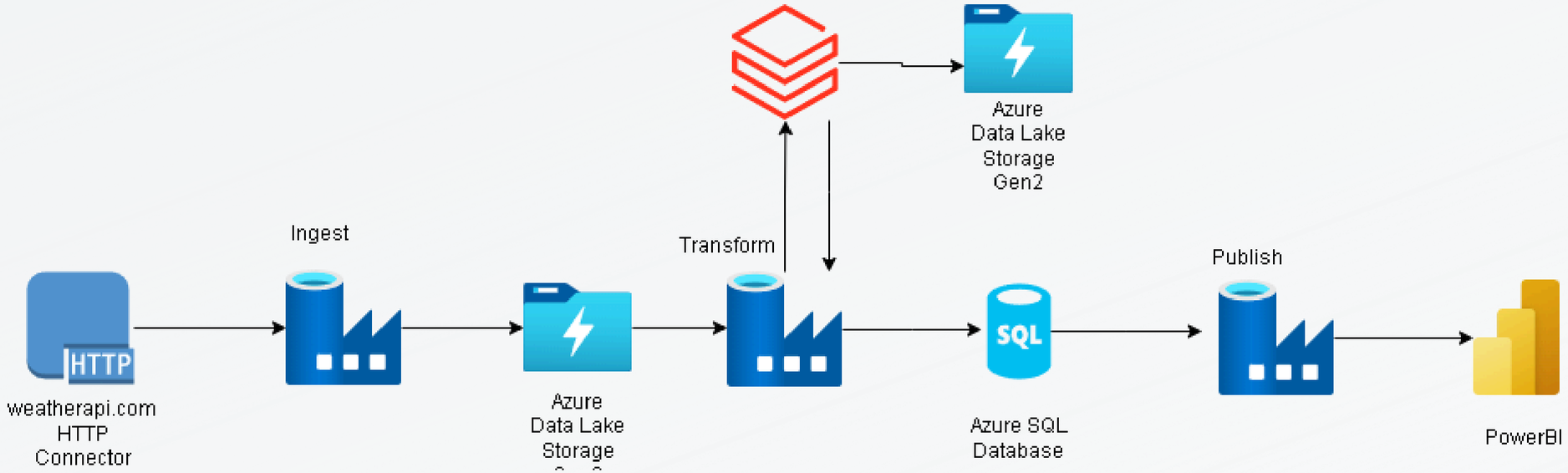
GUSTAVO JOÃO GONÇALVES CALDAS



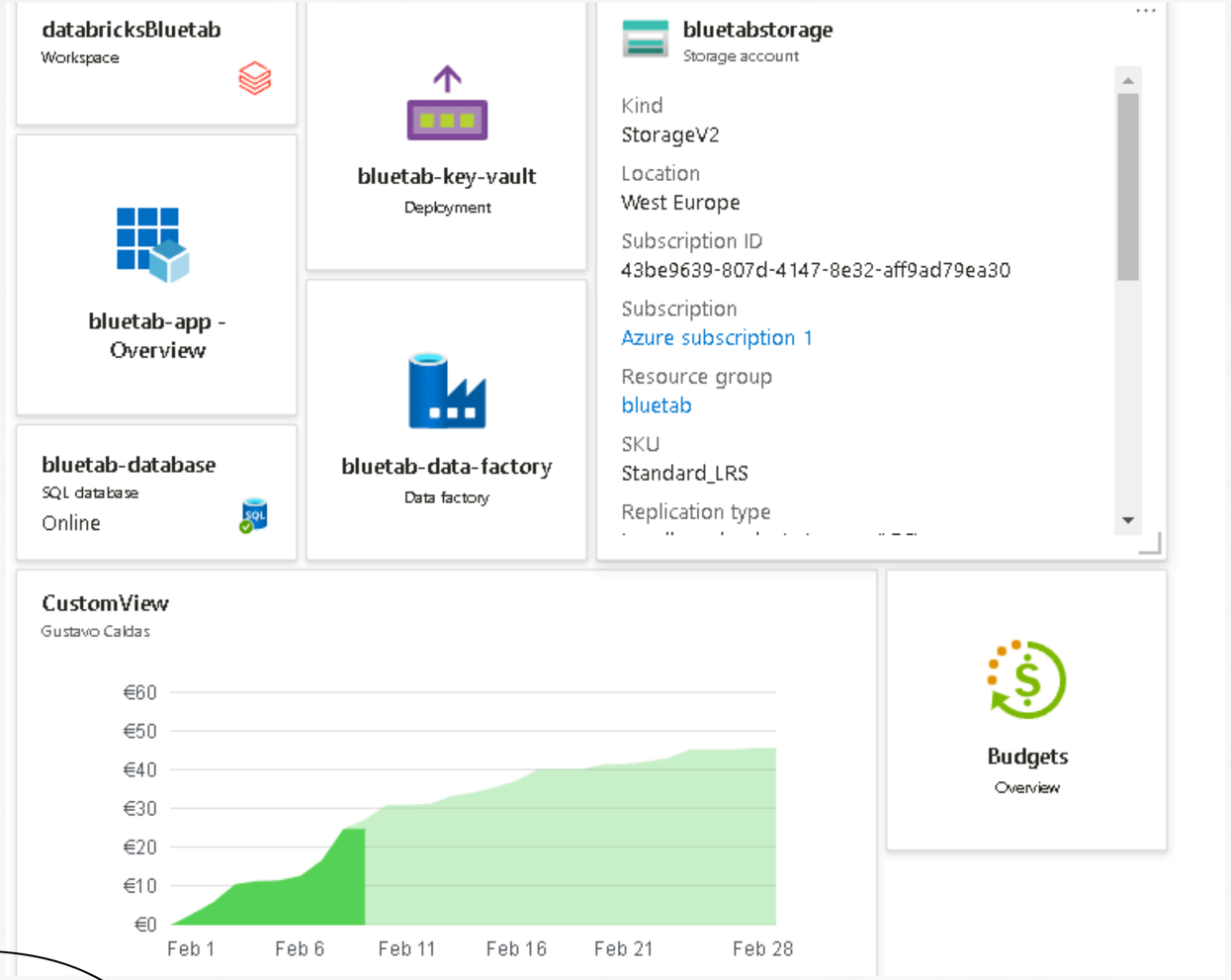
Learning is an active process. We learn by doing.. Only knowledge that is used sticks in your mind.

— Dale Carnegie —

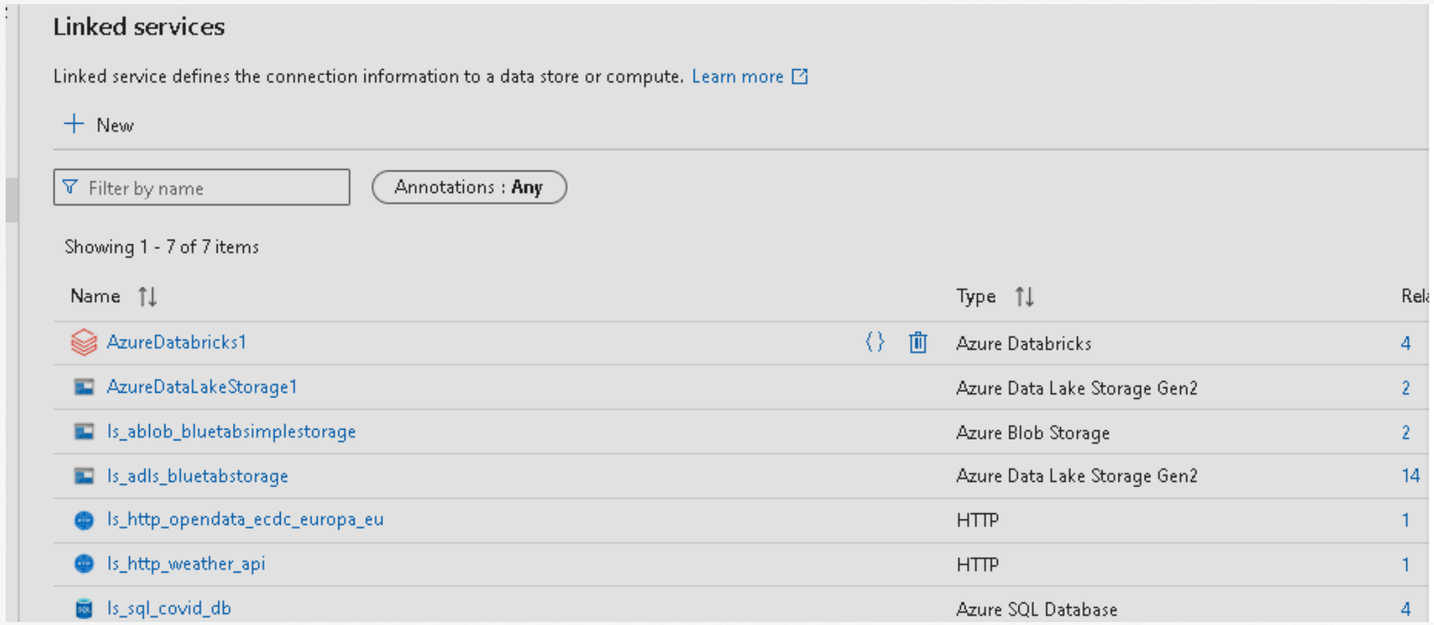
ARCHITECTURE

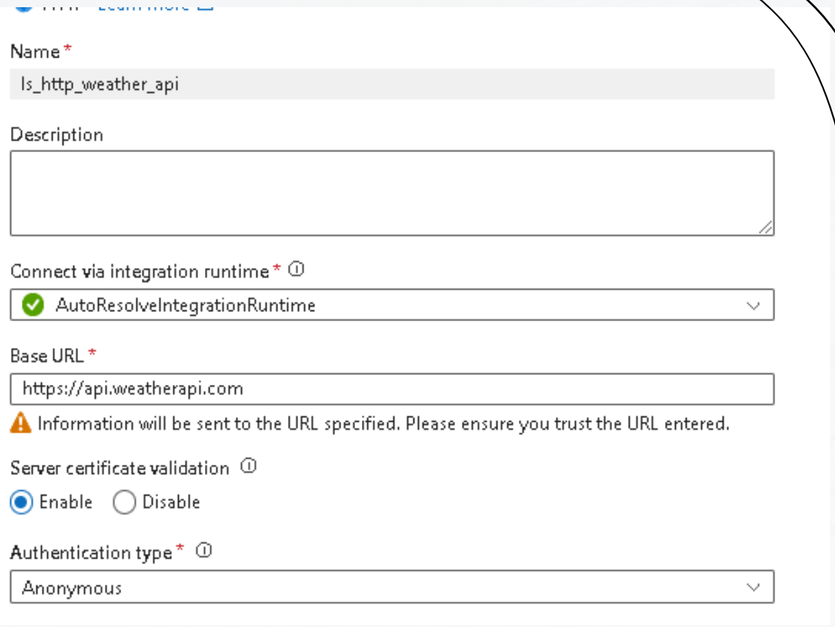


SERVICES


databricksBluetab
Workspace
bluetab-key-vault
Deployment
bluetabstorage
Storage account
Kind
StorageV2
Location
West Europe
Subscription ID
43be9639-807d-4147-8e32-aff9ad79ea30
Subscription
Azure subscription 1
Resource group
bluetab
SKU
Standard_LRS
Replication type
bluetab-app - Overview
bluetab-database
SQL database
Online
bluetab-data-factory
Data factory
CustomView
Gustavo Caldas
€60
€50
€40
€30
€20
€10
€0
Feb 1
Feb 6
Feb 11
Feb 16
Feb 21
Feb 28
Budgets
Overview

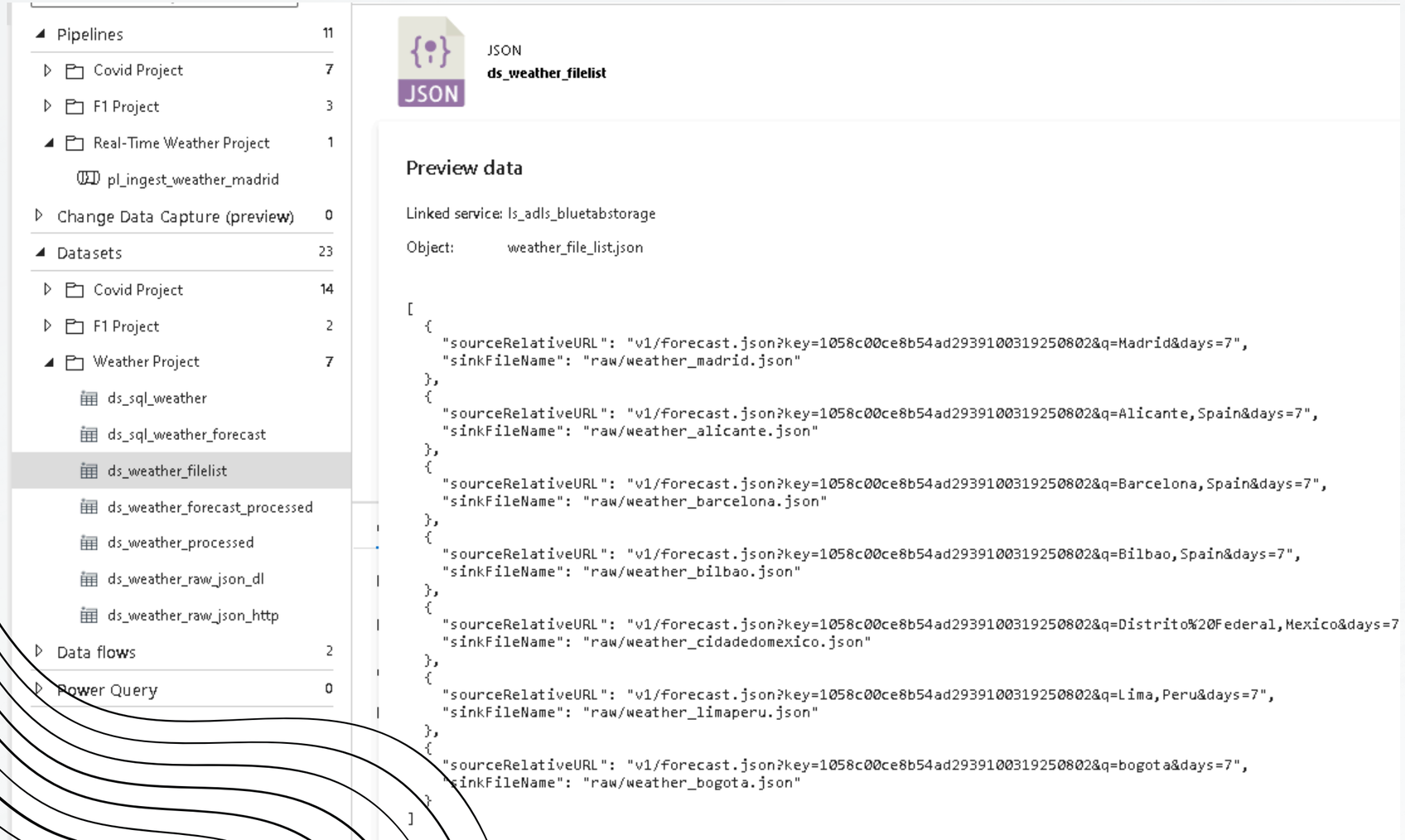
SET-UP
DATA FACTORY LINKED SERVICES



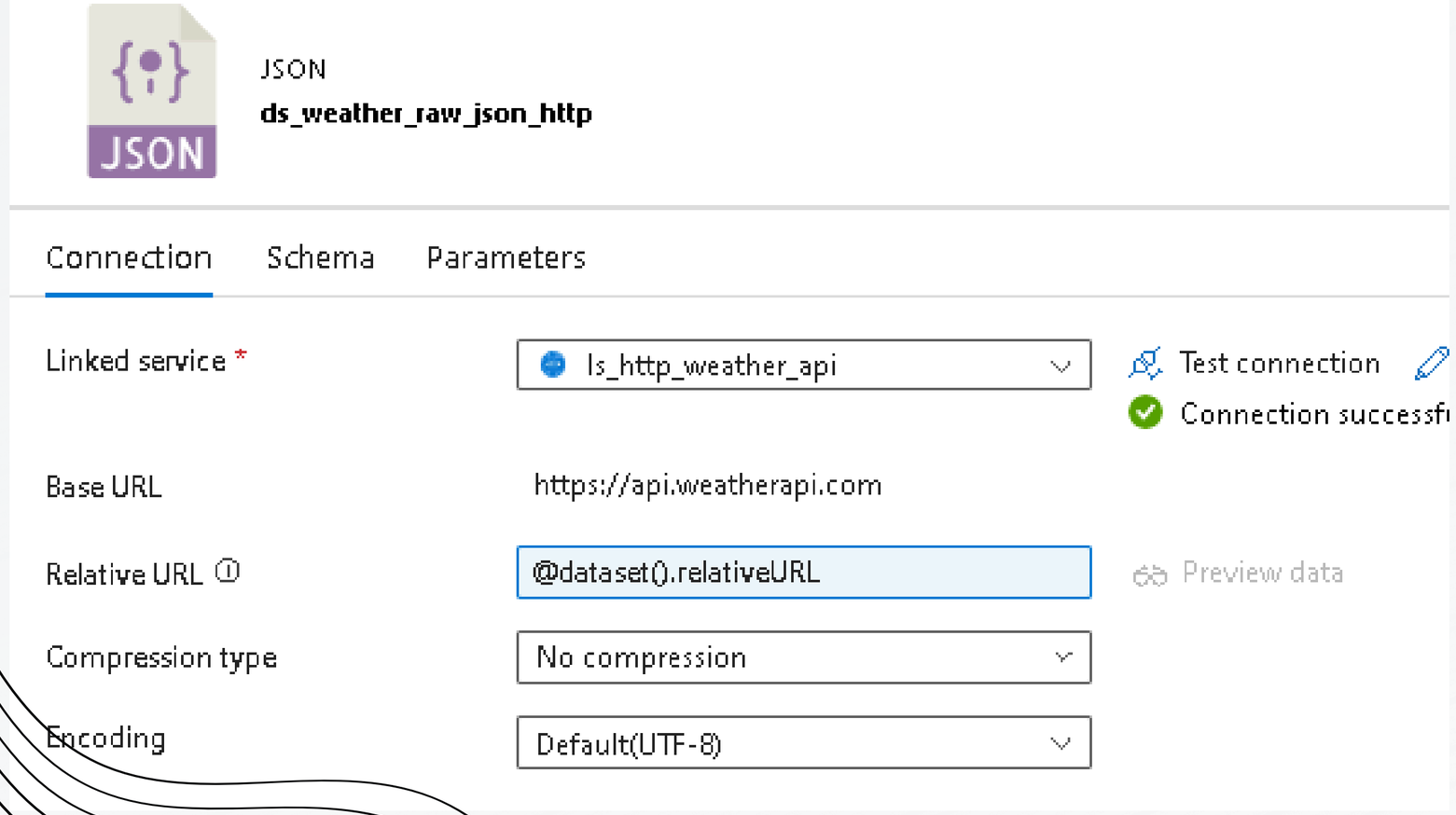


SET-UP

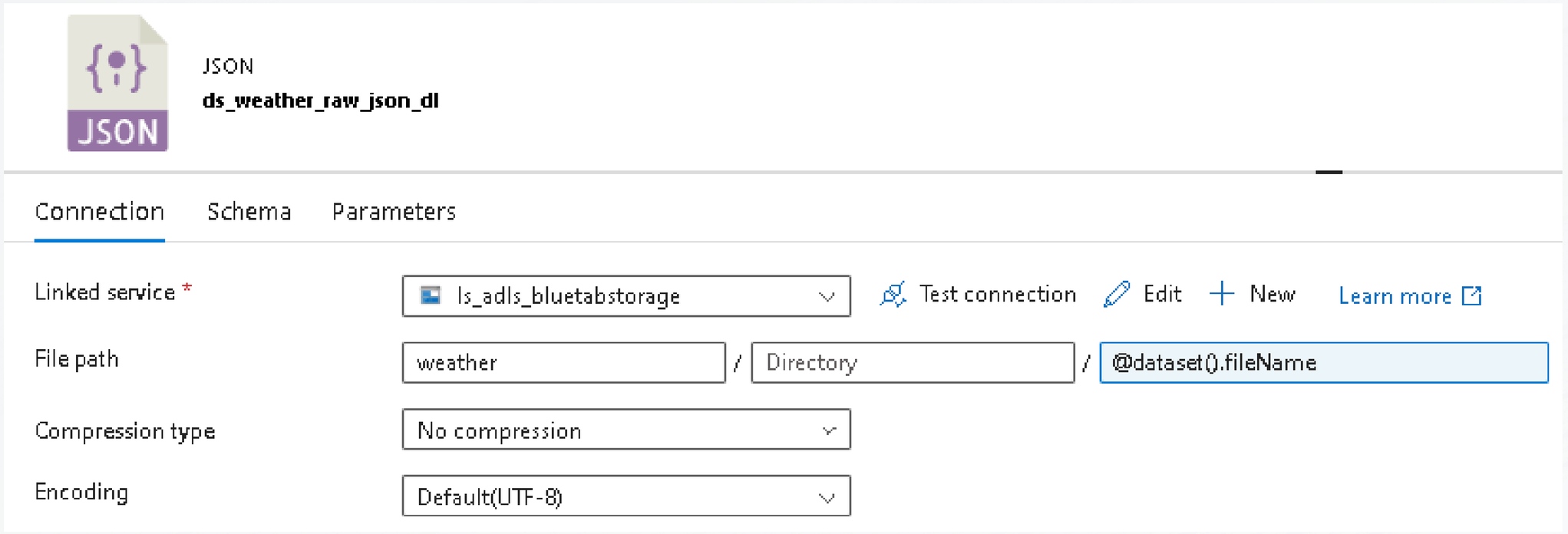
DATA FACTORY FILE LIST



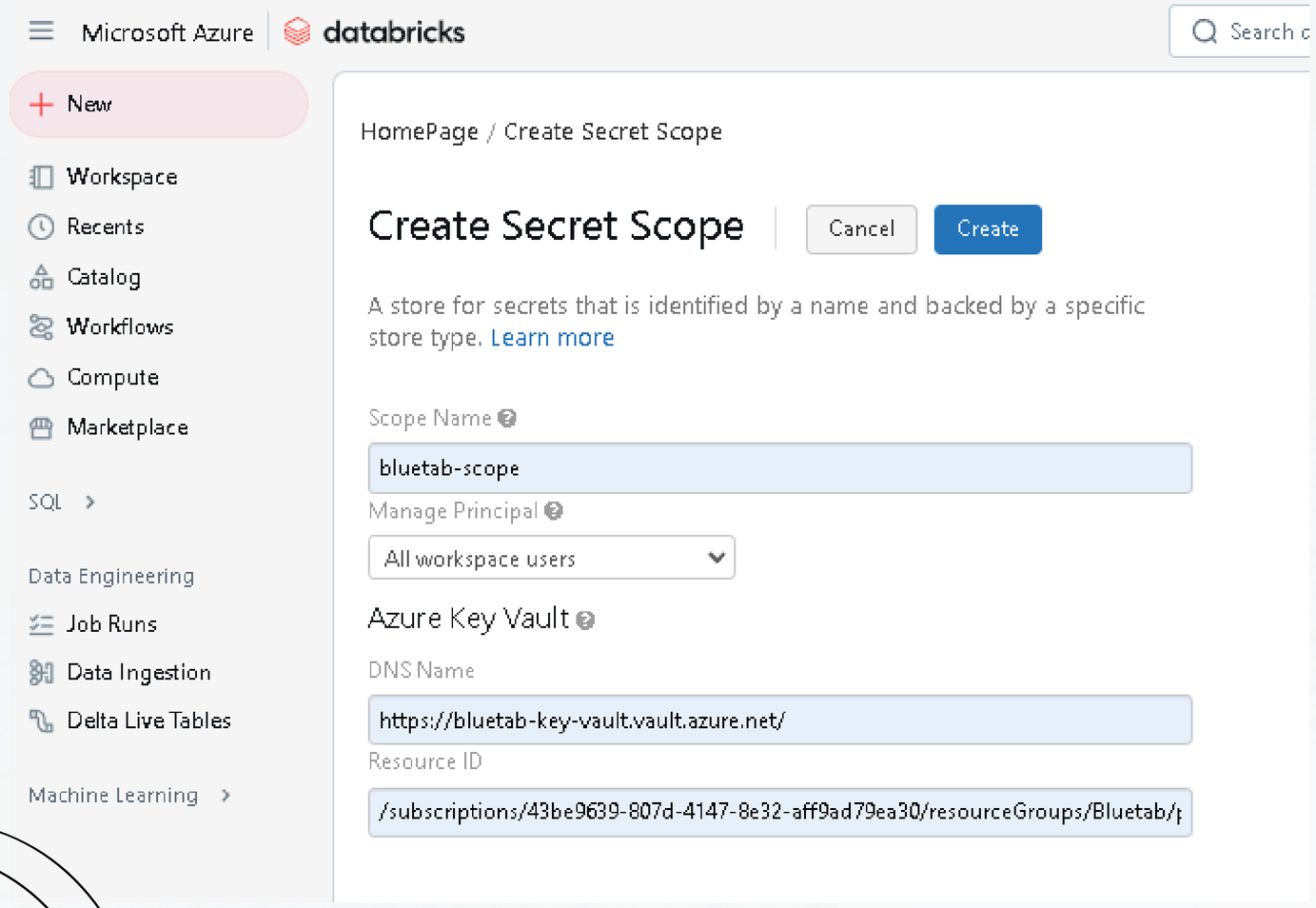
SET-UP
DATA FACTORY DS



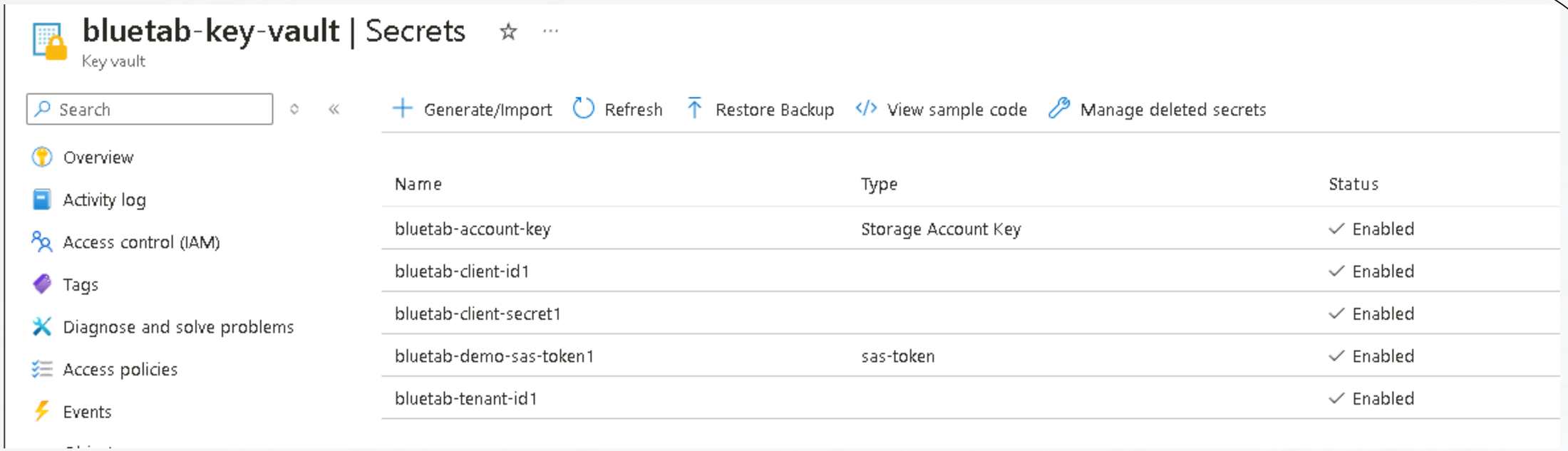
SET-UP
DATA FACTORY DS



SET-UP
DATABRICKS



SET-UP
DATABRICKS



SET-UP
DATABRICKS

Mount Azure Data Lake Containers for the Project

Yesterday (3s) 2

```
def mount_adls(storage_account_name, container_name):

    client_id = dbutils.secrets.get(scope = 'bluetab-scope', key = 'bluetab-client-id1')
    tenant_id = dbutils.secrets.get(scope = 'bluetab-scope', key = 'bluetab-tenant-id1')
    client_secret = dbutils.secrets.get(scope ='bluetab-scope', key ='bluetab-client-secret1')

    configs = {"fs.azure.account.auth.type": "OAuth",
          "fs.azure.account.oauth.provider.type": "org.apache.hadoop.fs.azurebfs.oauth2.ClientCredsTokenProvider",
          "fs.azure.account.oauth2.client.id": client_id,
          "fs.azure.account.oauth2.client.secret": client_secret,
          "fs.azure.account.oauth2.client.endpoint": f"https://login.microsoftonline.com/{tenant_id}/oauth2/token"}

    if any(mount.mountPoint == f"/mnt/{storage_account_name}/{container_name}" for mount in dbutils.fs.mounts()):
        dbutils.fs.unmount(f"/mnt/{storage_account_name}/{container_name}")


    dbutils.fs.mount(
    source = f"abfss://{container_name}@{storage_account_name}.dfs.core.windows.net/",
    mount_point = f"/mnt/{storage_account_name}/{container_name}",
    extra_configs = configs)

    display(dbutils.fs.mounts())
```

Yesterday (23s) 3

```
mount_adls('bluetabstorage', 'weather')
```

Yesterday (1s) 4

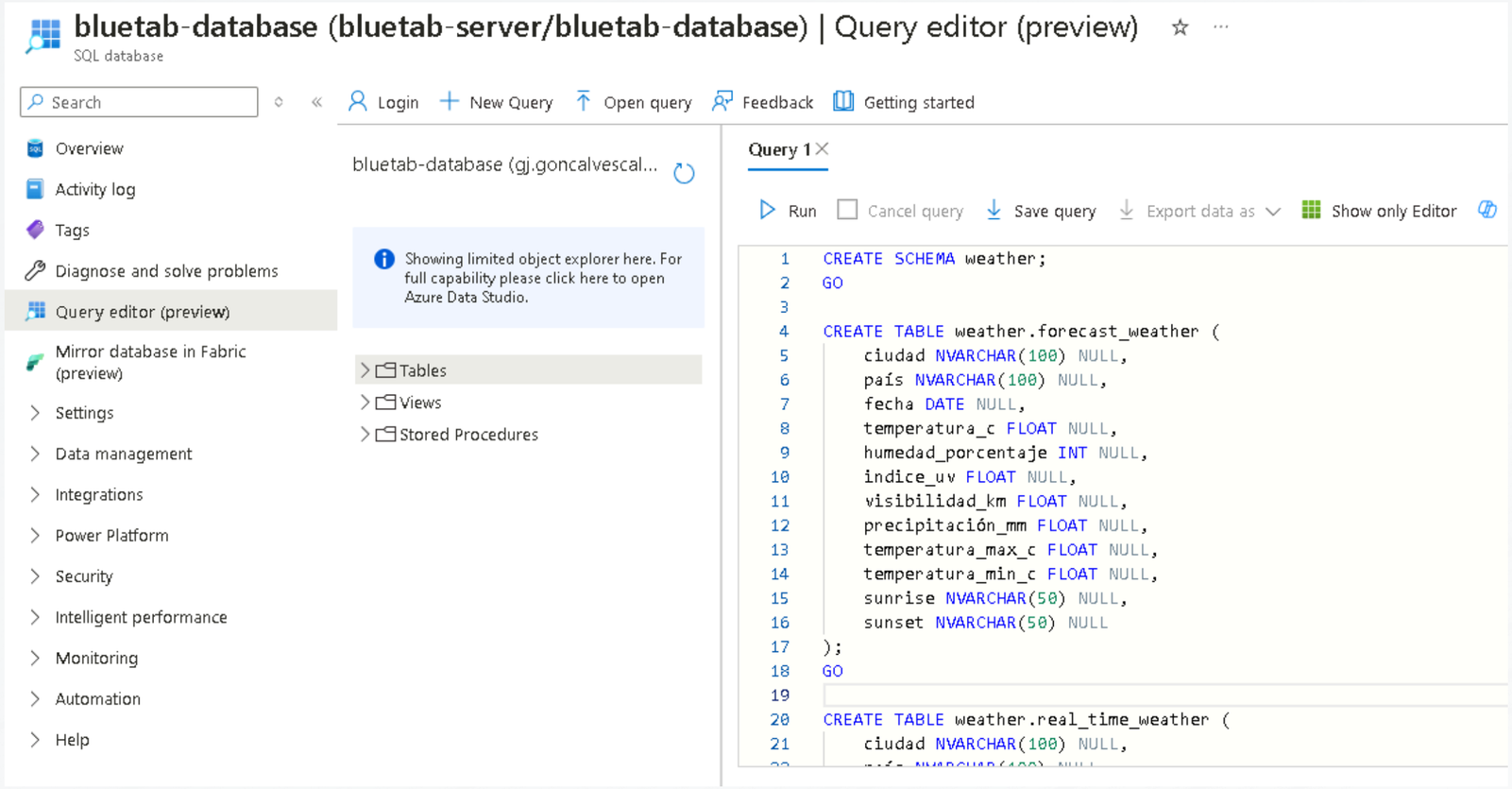
```
display(dbutils.fs.ls("/mnt/bluetabstorage/weather"), truncate=False)
```

(2) Spark Jobs

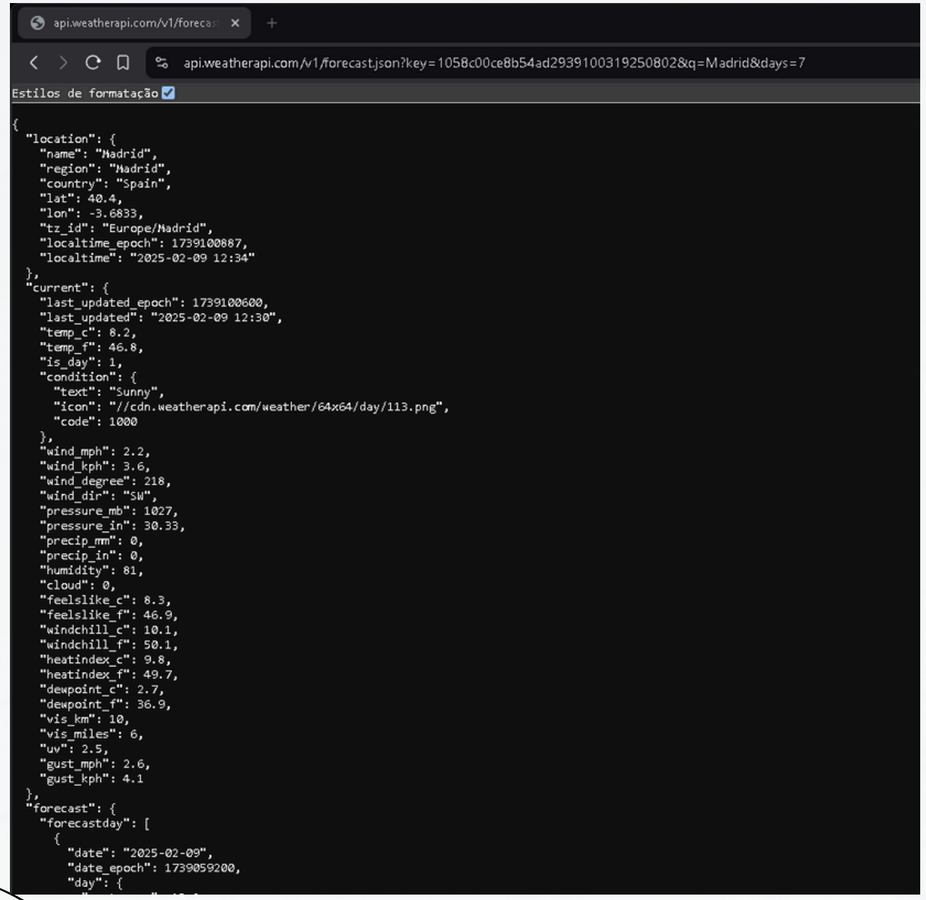
Table +

	path	name	size	modificationTime
1	dbfs:/mnt/bluetabstorage/weather...	raw/	0	17390129210000

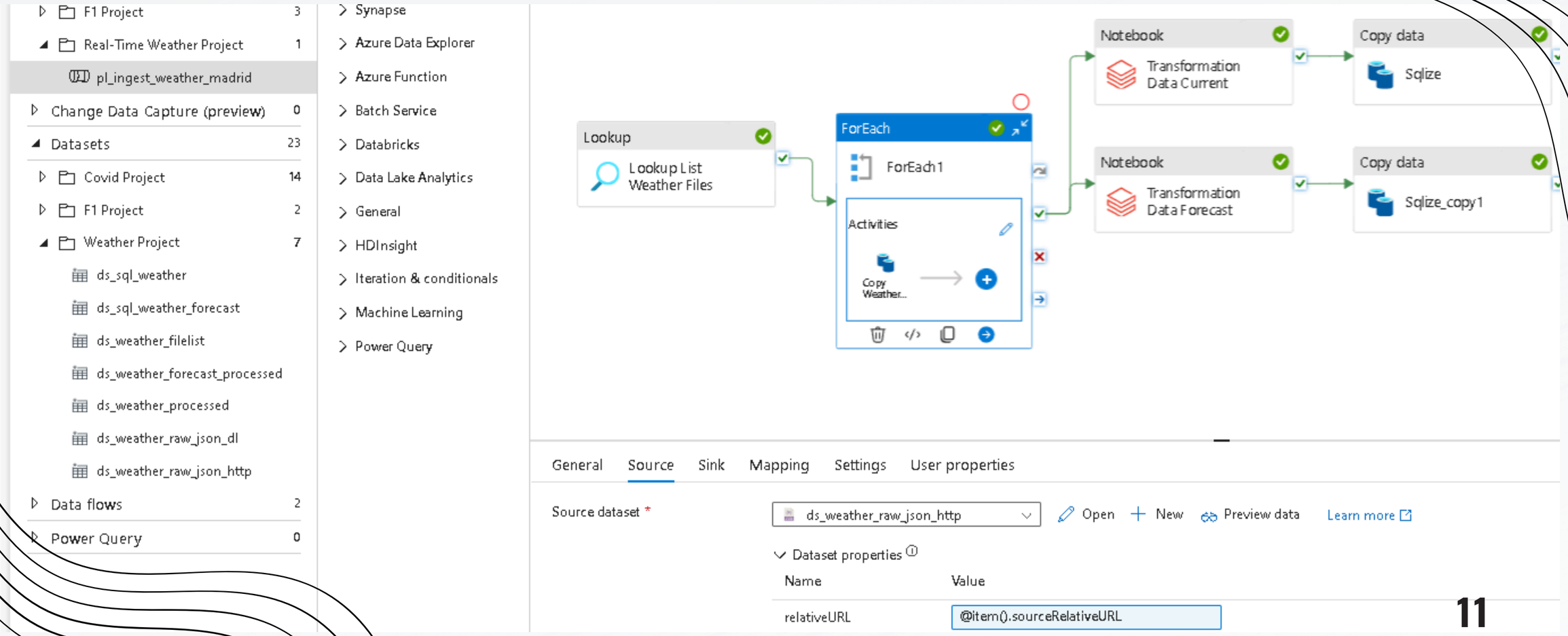
SET-UP
AZURE SQL



API


api.weatherapi.com/v1/forecast.json?key=1058c00ce8b54ad2939100319250802&q=Madrid&days=7
Estilos de formatação
{
"location": {
"name": "Madrid",
"region": "Madrid",
"country": "Spain",
"lat": 40.4,
"lon": -3.6833,
"tz_id": "Europe/Madrid",
"localtime_epoch": 1739100887,
"localtime": "2025-02-09 12:34"
},
"current": {
"last_updated_epoch": 1739100600,
"last_updated": "2025-02-09 12:30",
"temp_c": 8.2,
"temp_f": 46.8,
"is_day": 1,
"condition": {
"text": "Sunny",
"icon": "//cdn.weatherapi.com/weather/64x64/day/113.png",
"code": 1000
},
"wind_mph": 2.2,
"wind_kph": 3.6,
"wind_degree": 218,
"wind_dir": "SW",
"pressure_mb": 1027,
"pressure_in": 30.33,
"precip_mm": 0,
"precip_in": 0,
"humidity": 81,
"cloud": 0,
"feelslike_c": 8.3,
"feelslike_f": 46.9,
"windchill_c": 10.1,
"windchill_f": 50.1,
"heatindex_c": 9.8,
"heatindex_f": 49.7,
"dewpoint_c": 2.7,
"dewpoint_f": 36.9,
"vis_km": 10,
"vis_miles": 6,
"uv": 2.5,
"gust_mph": 2.6,
"gust_kph": 4.1
},
"forecast": {
"forecastday": [
{
"date": "2025-02-09",
"date_epoch": 1739059200,
"day": {

DATAFACTORY
PIPELINE



DATAFACTORY
OUTPUT

15 hours ago (4s) 3 Python

```
df_madrid = extraer_datos_forecast(spark.read.json("dbfs:/mnt/bluetabstorage/weather/raw/weather_madrid.json"))
df_alicante = extraer_datos_forecast(spark.read.json("dbfs:/mnt/bluetabstorage/weather/raw/weather_alicante.json"))
df_barcelona = extraer_datos_forecast(spark.read.json("dbfs:/mnt/bluetabstorage/weather/raw/weather_barcelona.json"))
df_bilbao = extraer_datos_forecast(spark.read.json("dbfs:/mnt/bluetabstorage/weather/raw/weather_bilbao.json"))
df_cidadedomexico = extraer_datos_forecast(spark.read.json("dbfs:/mnt/bluetabstorage/weather/raw/weather_cidadedomexico.json"))
df_limaperu = extraer_datos_forecast(spark.read.json("dbfs:/mnt/bluetabstorage/weather/raw/weather_limaperu.json"))
df_bogota = extraer_datos_forecast(spark.read.json("dbfs:/mnt/bluetabstorage/weather/raw/weather_bogota.json"))
```

(7) Spark Jobs

- df_alicante: pyspark.sql.dataframe.DataFrame = [ciudad: string, país: string ... 10 more fields]
- df_barcelona: pyspark.sql.dataframe.DataFrame = [ciudad: string, país: string ... 10 more fields]
- df_bilbao: pyspark.sql.dataframe.DataFrame = [ciudad: string, país: string ... 10 more fields]
- df_bogota: pyspark.sql.dataframe.DataFrame = [ciudad: string, país: string ... 10 more fields]
- df_cidadedomexico: pyspark.sql.dataframe.DataFrame = [ciudad: string, país: string ... 10 more fields]
- df_limaperu: pyspark.sql.dataframe.DataFrame = [ciudad: string, país: string ... 10 more fields]
- df_madrid: pyspark.sql.dataframe.DataFrame = [ciudad: string, país: string ... 10 more fields]

15 hours ago (1s) 4 Python

```
df_final = df_madrid.unionByName(df_alicante).unionByName(df_barcelona).unionByName(df_bilbao).unionByName(df_cidadedomexico).unionByName
(df_limaperu).unionByName(df_bogota)
df_final = df_final.orderBy("ciudad", "país", "fecha")
display(df_final)
df_final.printSchema()
```

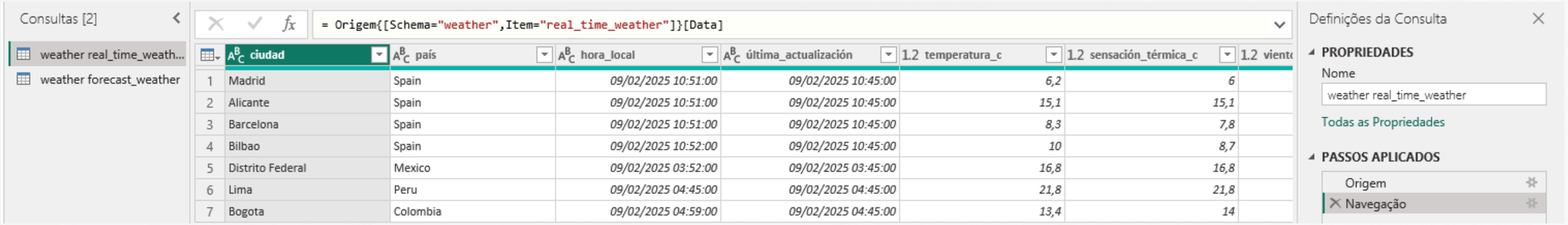
(1) Spark Jobs

- df_final: pyspark.sql.dataframe.DataFrame = [ciudad: string, país: string ... 10 more fields]

Table +

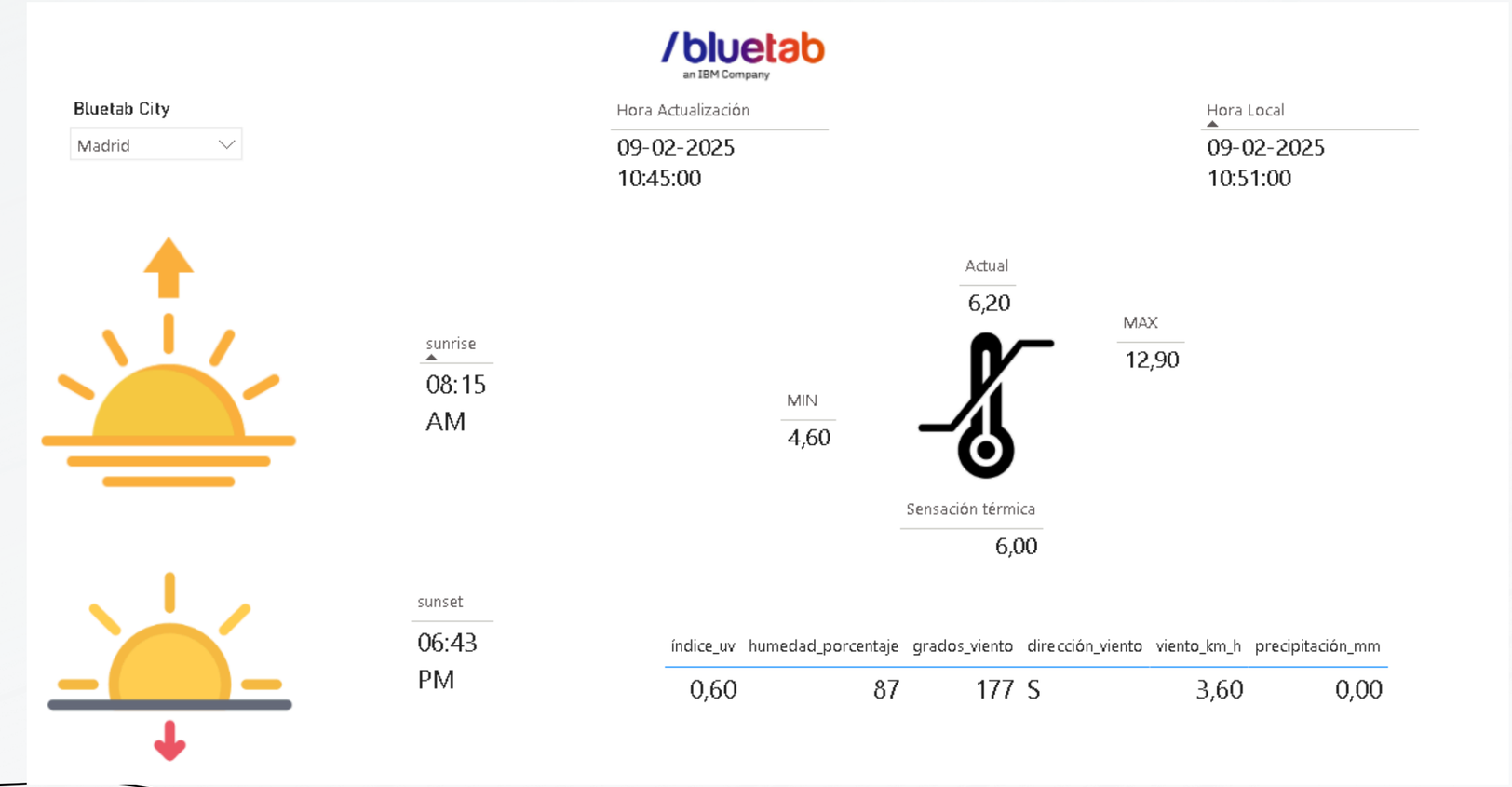
	ciudad	país	fecha	temperatura_c	humedad_porcentaje	índice_uv	visibilidad_km	precipitación_mm	tem
36	Lima	Peru	2025-02-08	22.7	80	3.3	10	0.01	
37	Lima	Peru	2025-02-09	22.6	80	2.9	10	0.02	
38	Lima	Peru	2025-02-10	22.9	75	3.1	10	0.05	
39	Lima	Peru	2025-02-11	23.2	72	3.2	10	0.3	
40	Lima	Peru	2025-02-12	23.3	72	0.1	10	0.01	
41	Lima	Peru	2025-02-13	23.5	77	9	10	0.01	
42	Lima	Peru	2025-02-14	23	81	9	10	0.01	
43	Madrid	Spain	2025-02-08	6.7	73	0.6	10	0	
44	Madrid	Spain	2025-02-09	8.5	68	0.5	10	0	
45	Madrid	Spain	2025-02-10	10.5	67	0.4	10	0	
46	Madrid	Spain	2025-02-11	10.4	66	0.5	10	0.86	
47	Madrid	Spain	2025-02-12	10.4	73	0.3	8.8	11.12	
48	Madrid	Spain	2025-02-13	8.3	71	3	10	0	
49	Madrid	Spain	2025-02-14	10.2	60	3	10	0	

POWERBI
DATA



	ciudad	país	hora_local	última_actualización	temperatura_c	sensación_térmica_c	viento
1	Madrid	Spain	09/02/2025 10:51:00	09/02/2025 10:45:00	6,2	6	
2	Alicante	Spain	09/02/2025 10:51:00	09/02/2025 10:45:00	15,1	15,1	
3	Barcelona	Spain	09/02/2025 10:51:00	09/02/2025 10:45:00	8,3	7,8	
4	Bilbao	Spain	09/02/2025 10:52:00	09/02/2025 10:45:00	10	8,7	
5	Distrito Federal	Mexico	09/02/2025 03:52:00	09/02/2025 03:45:00	16,8	16,8	
6	Lima	Peru	09/02/2025 04:45:00	09/02/2025 04:45:00	21,8	21,8	
7	Bogota	Colombia	09/02/2025 04:59:00	09/02/2025 04:45:00	13,4	14	

POWERBI
DASHBOARD



POWERBI
DASHBOARD

Bluetab City

Madrid

país	ciudad	fecha	sunrise	sunset	temperatura_c	temperatura_max_c	temperatura_min_c	humedad_porcentaje	precipitación_mm	índice_uv	visibilidad_km
Spain	Madrid	09-02-25	08:15 AM	06:43 PM	8,20	12,90	4,60	69	0,00	0,60	10,00
Spain	Madrid	10-02-25	08:14 AM	06:44 PM	9,80	13,60	7,70	68	0,00	0,50	10,00
Spain	Madrid	11-02-25	08:13 AM	06:45 PM	9,80	14,50	7,20	67	0,31	0,50	9,70
Spain	Madrid	12-02-25	08:12 AM	06:46 PM	8,40	9,40	7,90	86	9,81	0,30	7,50
Spain	Madrid	13-02-25	08:10 AM	06:48 PM	9,60	13,30	7,80	70	0,02	0,50	10,00
Spain	Madrid	14-02-25	08:09 AM	06:49 PM	9,40	14,20	5,50	62	0,00	3,00	10,00
Spain	Madrid	15-02-25	08:08 AM	06:50 PM	9,90	14,90	5,70	59	0,00	3,00	10,00

LINKS

WEATHER IN BLUETAB CITIES

LINK

INTERACTIVE POWERBI

LINK

LINKS

DATAFACTORY REPOSITORY

LINK

PROFILE
LINK

GRADO Y MÁSTER EN INGENIERÍA Y GESTIÓN DE SISTEMAS DE INFORMACIÓN - ÁREA DE ESPECIALIZACIÓN EN INGENIERÍA DE SISTEMAS DE INFORMACIÓN

🤖 TECHNOLOGY CASE

GUSTAVO JOÃO GONÇALVES CALDAS