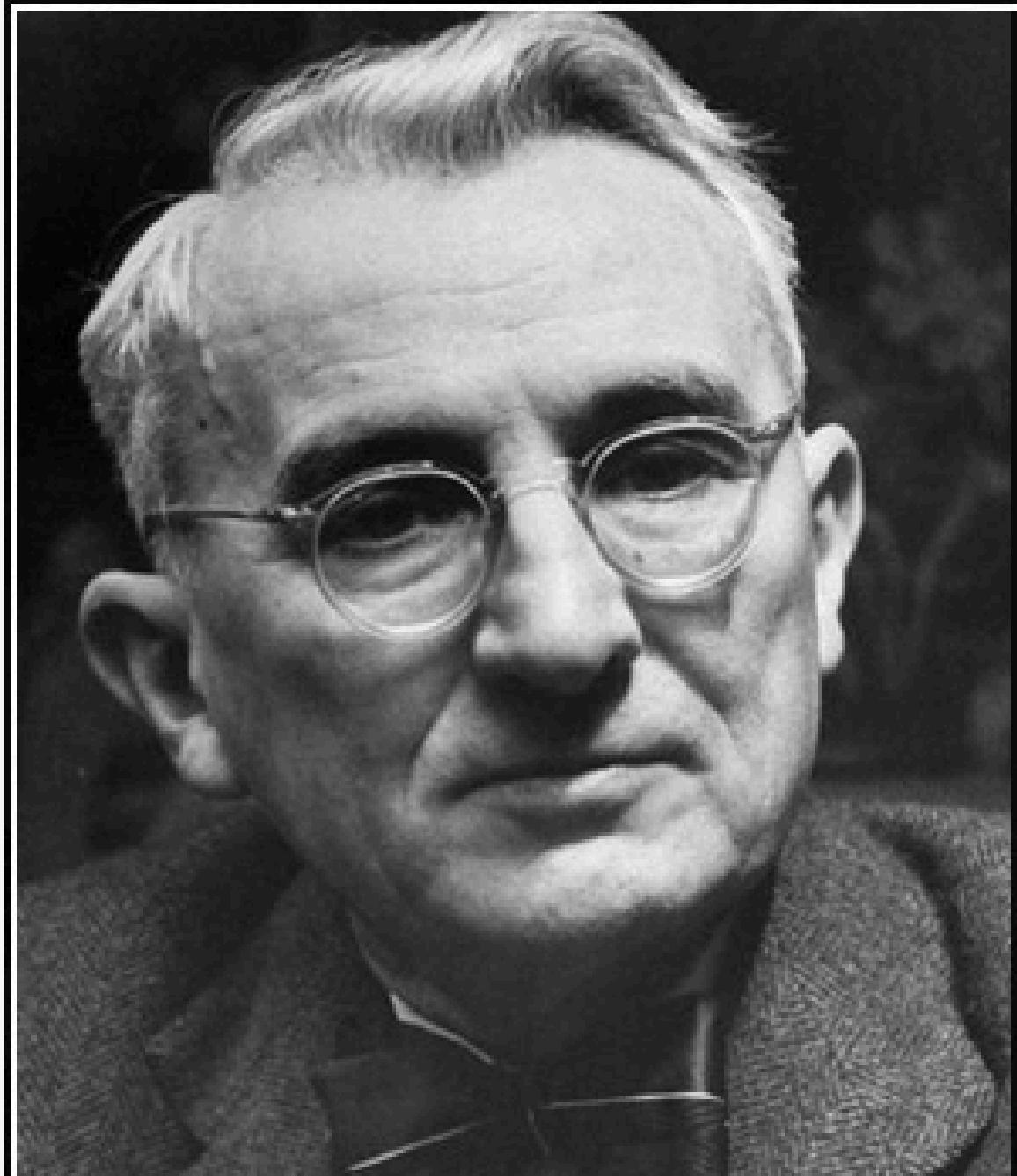


**GRADO Y MÁSTER EN INGENIERÍA Y GESTIÓN DE SISTEMAS DE
INFORMACIÓN - ÁREA DE ESPECIALIZACIÓN EN INGENIERÍA DE
SISTEMAS DE INFORMACIÓN**



TECHNOLOGY CASE

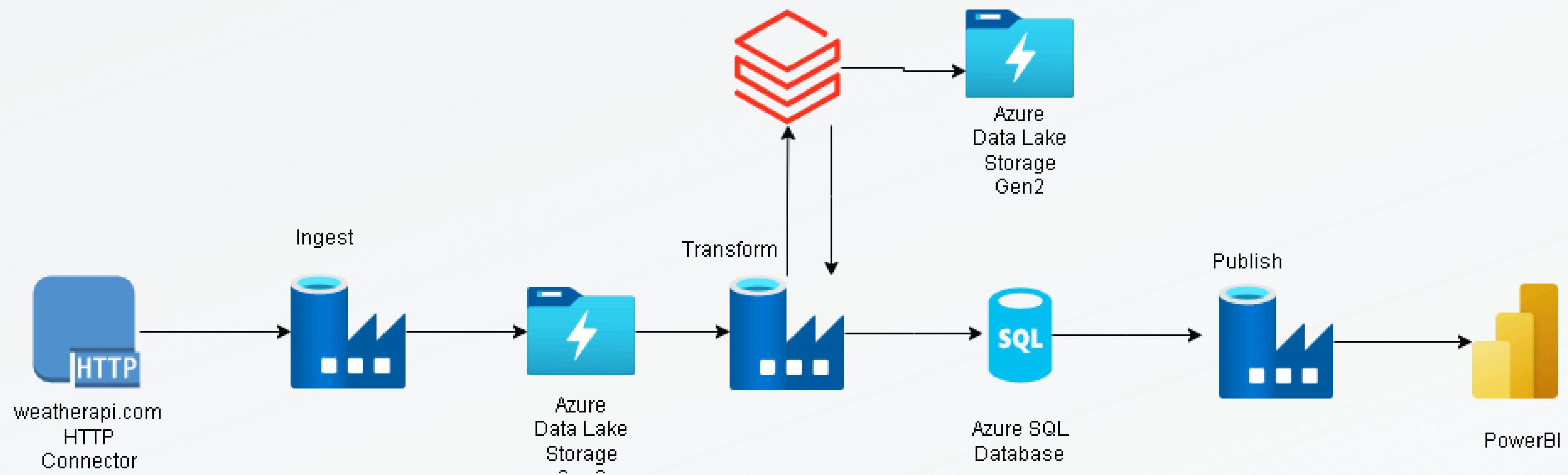
GUSTAVO JOÃO GONÇALVES CALDAS



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

— *Dale Carnegie* —

ARCHITECTURE



SERVICES

The screenshot displays the Microsoft Azure portal interface with several service components visible:

- databricksBluetab** (Workspace): Includes a red cube icon.
- bluetab-app - Overview**: Shows a blue grid icon.
- bluetab-database** (SQL database): Status is Online, indicated by a green checkmark icon.
- bluetab-key-vault** (Deployment): Shows a purple box with an upward arrow icon.
- bluetab-data-factory** (Data factory): Shows a blue cylinder icon.
- bluetabstorage** (Storage account): Includes a teal bar icon. Details shown:
 - Kind: StorageV2
 - Location: West Europe
 - Subscription ID: 43be9639-807d-4147-8e32-aff9ad79ea30
 - Subscription: Azure subscription 1
 - Resource group: bluetab
 - SKU: Standard_LRS
 - Replication type: "Standard"
- CustomView** (Gustavo Caldas): A line chart showing spending over time. The Y-axis ranges from €0 to €60, and the X-axis shows dates from Feb 1 to Feb 28. The spending starts at €0, rises to about €25 by Feb 7, then continues to rise more steadily, reaching approximately €45 by Feb 28.
- Budgets** (Overview): Shows a green circle with a dollar sign icon.

SET-UP DATA FACTORY LINKED SERVICES

Linked services

Linked service defines the connection information to a data store or compute. [Learn more](#)

+ New

Filter by name Annotations : Any

Showing 1 - 7 of 7 items

Name ↑	Type ↑	Rel.
AzureDatabricks1	Azure Databricks	4
AzureDataLakeStorage1	Azure Data Lake Storage Gen2	2
ls_ablob_blueatabsimplestorage	Azure Blob Storage	2
ls_adls_blueatabstorage	Azure Data Lake Storage Gen2	14
ls_http_opendata_ecdc_europa_eu	HTTP	1
ls_http_weather_api	HTTP	1
ls_sql_covid_db	Azure SQL Database	4

Name * ls_http_weather_api

Description

Connect via integration runtime * ⓘ AutoResolveIntegrationRuntime

Base URL * https://api.weatherapi.com

⚠ Information will be sent to the URL specified. Please ensure you trust the URL entered.

Server certificate validation ⓘ Enable Disable

Authentication type * ⓘ Anonymous

SET-UP

DATA FACTORY FILE LIST

The screenshot shows the Azure Data Factory interface. On the left, there's a navigation pane with the following structure:

- Pipelines (11)
 - Covid Project (7)
 - F1 Project (3)
 - Real-Time Weather Project (1)
 - pl_ingest_weather_madrid
 - Change Data Capture (preview) (0)
- Datasets (23)
 - Covid Project (14)
 - F1 Project (2)
 - Weather Project (7)
 - ds_sql_weather
 - ds_sql_weather_forecast
 - ds_weather_filelist**
 - ds_weather_forecast_processed
 - ds_weather_processed
 - ds_weather_raw_json_dl
 - ds_weather_raw_json_http
- Data flows (2)
- Power Query (0)

On the right, the details for the selected dataset 'ds_weather_filelist' are shown. The icon is a JSON file (purple and white). The name is 'ds_weather_filelist'. The preview data section shows the following JSON array:

```
[{"sourceRelativeURL": "v1/Forecast.json?key=1058c00ce8b54ad2939100319250802&q=Madrid&days=7", "sinkFileName": "raw/weather_madrid.json"}, {"sourceRelativeURL": "v1/Forecast.json?key=1058c00ce8b54ad2939100319250802&q=Alicante,Spain&days=7", "sinkFileName": "raw/weather_alicante.json"}, {"sourceRelativeURL": "v1/Forecast.json?key=1058c00ce8b54ad2939100319250802&q=Barcelona,Spain&days=7", "sinkFileName": "raw/weather_barcelona.json"}, {"sourceRelativeURL": "v1/Forecast.json?key=1058c00ce8b54ad2939100319250802&q=Bilbao,Spain&days=7", "sinkFileName": "raw/weather_bilbao.json"}, {"sourceRelativeURL": "v1/Forecast.json?key=1058c00ce8b54ad2939100319250802&q=Distrito%20Federal,Mexico&days=7", "sinkFileName": "raw/weather_cidadedomexico.json"}, {"sourceRelativeURL": "v1/Forecast.json?key=1058c00ce8b54ad2939100319250802&q=Lima,Peru&days=7", "sinkFileName": "raw/weather_limaperu.json"}, {"sourceRelativeURL": "v1/Forecast.json?key=1058c00ce8b54ad2939100319250802&q=bogota&days=7", "sinkFileName": "raw/weather_bogota.json"}]
```

SET-UP DATA FACTORY DS

The screenshot shows the configuration page for a JSON dataset named "ds_weather_raw_json_http". The "Connection" tab is selected. The "Linked service" dropdown is set to "ls_http_weather_api", which has a "Connection successful" status indicated by a green checkmark and a blue "Test connection" button. The "Base URL" is set to "https://api.weatherapi.com". The "Relative URL" field contains the expression "@dataset().relativeURL". The "Compression type" is set to "No compression" and the "Encoding" is set to "Default(UTF-8)".

JSON

ds_weather_raw_json_http

JSON

Connection Schema Parameters

Linked service *

ls_http_weather_api

Test connection

Connection successful

Base URL

https://api.weatherapi.com

Relative URL ⓘ

@dataset().relativeURL

Preview data

Compression type

No compression

Encoding

Default(UTF-8)

SET-UP DATA FACTORY DS

The screenshot shows the configuration page for a JSON dataset named "ds_weather_raw_json_dl". The "Connection" tab is selected. The "Linked service" dropdown is set to "ls_adls_bluetabstorage". The "File path" field contains "weather" / "Directory" / "@dataset().fileName". The "Compression type" is set to "No compression" and the "Encoding" is "Default(UTF-8)".

JSON

ds_weather_raw_json_dl

Connection Schema Parameters

Linked service * ls_adls_bluetabstorage Test connection Edit New Learn more

File path weather / Directory / @dataset().fileName

Compression type No compression

Encoding Default(UTF-8)

SET-UP DATABRICKS

The screenshot shows the Databricks web interface. On the left, there's a sidebar with various navigation options like Workspace, Recents, Catalog, Workflows, Compute, Marketplace, SQL, Data Engineering, Job Runs, Data Ingestion, Delta Live Tables, and Machine Learning. A pink button labeled '+ New' is highlighted. The main area is titled 'Create Secret Scope' with a sub-path 'HomePage / Create Secret Scope'. It includes a 'Scope Name' input field containing 'bluetab-scope', a 'Manage Principal' dropdown set to 'All workspace users', and sections for 'Azure Key Vault' with 'DNS Name' and 'Resource ID' inputs. A 'Cancel' and a blue 'Create' button are at the bottom right.

Microsoft Azure | databricks

+ New

HomePage / Create Secret Scope

Create Secret Scope

Scope Name

bluetab-scope

Manage Principal

All workspace users

Azure Key Vault

DNS Name

`https://bluetab-key-vault.vault.azure.net/`

Resource ID

`/subscriptions/43be9639-807d-4147-8e32-aff9ad79ea30/resourceGroups/Bluetab/t`

SET-UP DATABRICKS

The screenshot shows the Azure Key Vault Secrets page for the 'bluetab-key-vault' resource. The left sidebar includes links for Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Access policies, and Events. The main content area displays a table of secrets with columns for Name, Type, and Status.

Name	Type	Status
bluetab-account-key	Storage Account Key	✓ Enabled
bluetab-client-id1		✓ Enabled
bluetab-client-secret1		✓ Enabled
bluetab-demo-sas-token1	sas-token	✓ Enabled
bluetab-tenant-id1		✓ Enabled

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 (2s) 3
mount_adls('bluetabstorage', 'weather')
```

```
▶  ✓ Yesterday (1s) 4
display(dbutils.fs.ls("/mnt/bluetabstorage/weather"), truncate=False)
```

▶ (2) Spark Jobs

Table +

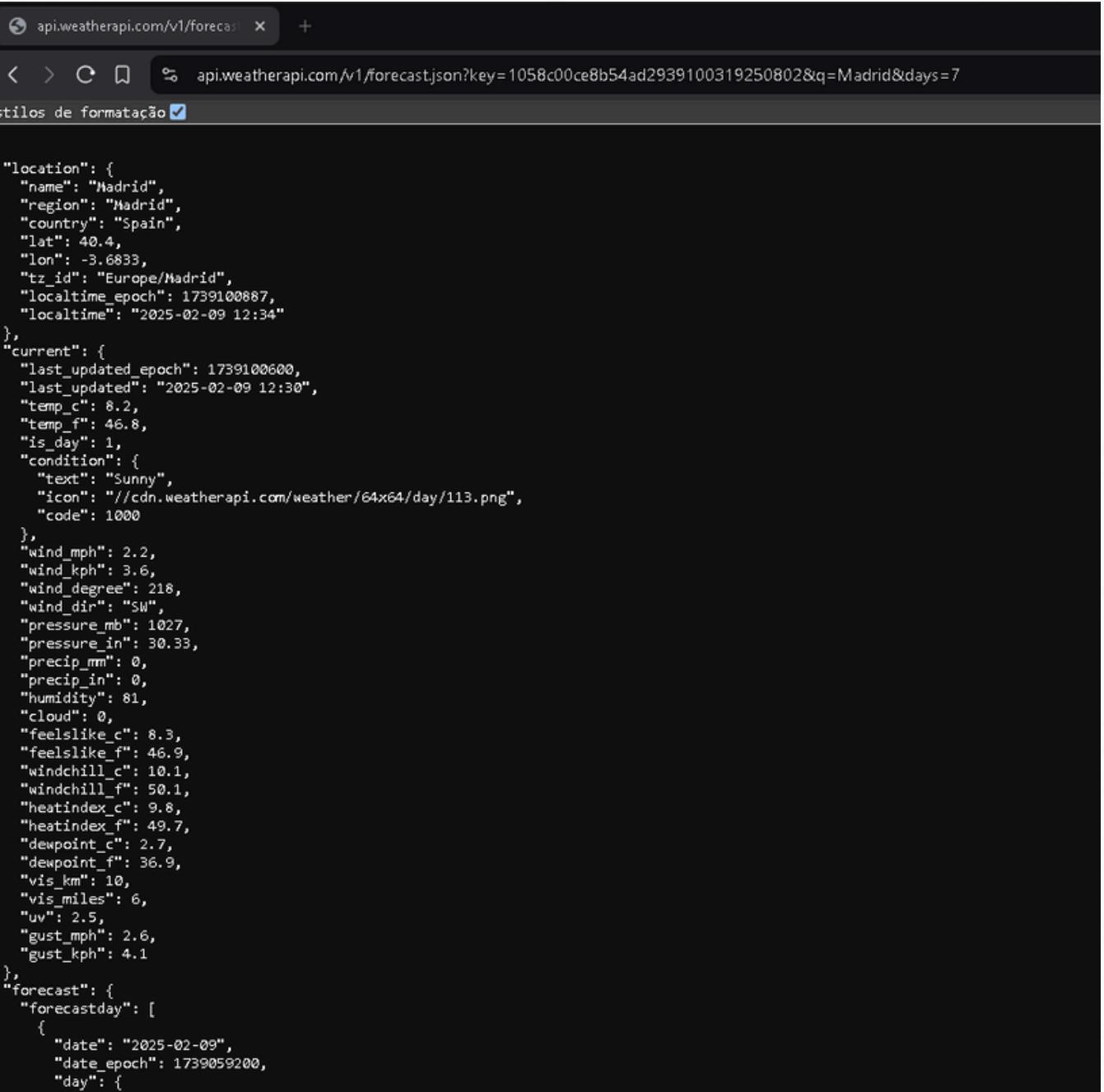
#	path	name	size	modificationTime
1	dbfs:/mnt/bluetabstorage/weather...	raw/	0	1739012921000

SET-UP AZURE SQL

The screenshot shows the Azure SQL Query editor (preview) interface. The left sidebar lists database management options like Overview, Activity log, Tags, and Query editor (preview), which is currently selected. The main area displays a query window titled "Query 1" containing the following T-SQL code:

```
1 CREATE SCHEMA weather;
2 GO
3
4 CREATE TABLE weather.forecast_weather (
5     ciudad NVARCHAR(100) NULL,
6     país NVARCHAR(100) NULL,
7     fecha DATE NULL,
8     temperatura_c FLOAT NULL,
9     humedad_porcentaje INT NULL,
10    indice_uv FLOAT NULL,
11    visibilidad_km FLOAT NULL,
12    precipitación_mm FLOAT NULL,
13    temperatura_max_c FLOAT NULL,
14    temperatura_min_c FLOAT NULL,
15    sunrise NVARCHAR(50) NULL,
16    sunset NVARCHAR(50) NULL
17 );
18 GO
19
20 CREATE TABLE weather.real_time_weather (
21     ciudad NVARCHAR(100) NULL,
22     ...
```

API



A screenshot of a web browser displaying a JSON API response from api.weatherapi.com/v1/forecast.json. The URL in the address bar includes a key, location (Madrid), and days (7). The response is a large block of JSON data.

```
{
  "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": {
          "maxtemp_c": 12.5,
          "maxtemp_f": 54.5,
          "mintemp_c": 7.5,
          "mintemp_f": 45.5,
          "avgtemp_c": 10.0,
          "avgtemp_f": 50.0,
          "avgwind_mph": 3.0,
          "avgwind_kph": 5.0,
          "avghumidity": 82,
          "avghumidity": 82,
          "precip_mm": 0,
          "precip_in": 0,
          "totalsun_hours": 10.0,
          "totalsun_deg": 218.0
        }
      }
    ]
  }
}
```

DATAFACTORY PIPELINE

The screenshot shows the Azure Data Factory pipeline editor interface. On the left, the navigation pane displays the project structure:

- F1 Project (3)
- Real-Time Weather Project (1)
 - pl_ingest_weather_madrid
- Change Data Capture (preview) (0)
- Datasets (23)
 - Covid Project (14)
 - F1 Project (2)
 - Weather Project (7)
 - ds_sql_weather
 - ds_sql_weather_forecast
 - ds_weather_filelist
 - ds_weather_forecast_processed
 - ds_weather_processed
 - ds_weather_raw_json_dl
 - ds_weather_raw_json_http
- Data flows (2)
- Power Query (0)

The main workspace shows a pipeline flow:

```
graph LR; L[Lookup: Weather Files] --> F[ForEach: ForEach1]; F --> N1[Notebook: Transformation Data Current]; F --> N2[Notebook: Transformation Data Forecast]; N1 --> C1[Copy data: Sqlize]; N2 --> C2[Copy data: Sqlize_copy1]
```

The pipeline consists of the following steps:

- A **Lookup** activity named "Weather Files" is connected to an **ForEach** activity.
- The **ForEach** activity is named "ForEach1" and contains an **Activities** section with a single step: "Copy Weather...".
- The **ForEach1** activity branches into two parallel paths, each leading to a **Notebook** activity.
- The first path leads to a **Notebook** named "Transformation Data Current". This notebook is connected to a **Copy data** activity with a sink named "Sqlize".
- The second path leads to a **Notebook** named "Transformation Data Forecast". This notebook is connected to a **Copy data** activity with a sink named "Sqlize_copy1".

Below the pipeline, the "Source" tab of the dataset configuration is visible:

General	Source	Sink	Mapping	Settings	User properties
Source dataset *					
<input type="button" value="Open"/> <input type="button" value="New"/> <input type="button" value="Preview data"/> <input type="button" value="Learn more"/>					
<input type="checkbox"/> Dataset properties					
Name	Value				
relativeURL	@item().sourceRelativeURL				

DATAFACTORY OUTPUT

```
3 15 hours ago (4d)
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]

4 15 hours ago (1d)
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 + Q ▾
Ac ciudad Ac país Ac fecha 1.2 temperatura_°C 1.2 humedad_porcentaje 1.2 índice_uv 1.2 visibilidad_km 1.2 precipitación_mm 1.2 ter
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

Consultas [2]

= Origem{[Schema="weather",Item="real_time_weather"]}[Data]

	A ^B ciudad	A ^B país	A ^B hora_local	A ^B última_actualización	1.2 temperatura_c	1.2 sensación térmica_c	1.2 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	

Definições da Consulta

PROPRIEDADES

Nome
weather real_time_weather

Todas as Propriedades

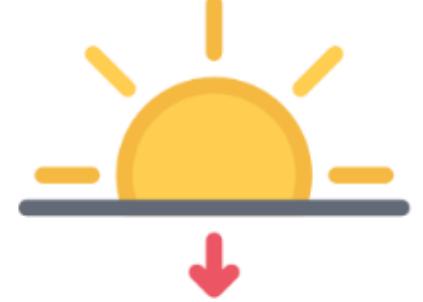
PASSOS APLICADOS

Origem
Navegação

POWERBI DASHBOARD

Bluetab City

Madrid



sunrise
08:15 AM

sunset
06:43 PM

/bluetab
an IBM Company

Hora Actualización
09-02-2025
10:45:00

Hora Local
09-02-2025
10:51:00

Actual
6,20

MIN
4,60

MAX
12,90

Sensación térmica
6,00

índice_uv	humedad_porcentaje	grados_viento	dirección_viento	viento_km_h	precipitación_mm
0,60	87	177	S	3,60	0,00

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

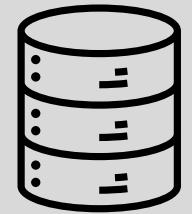
[LINK](#)



INTERACTIVE POWERBI

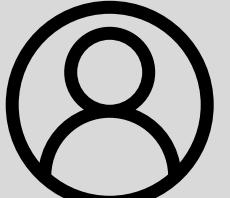
[LINK](#)

LINKS

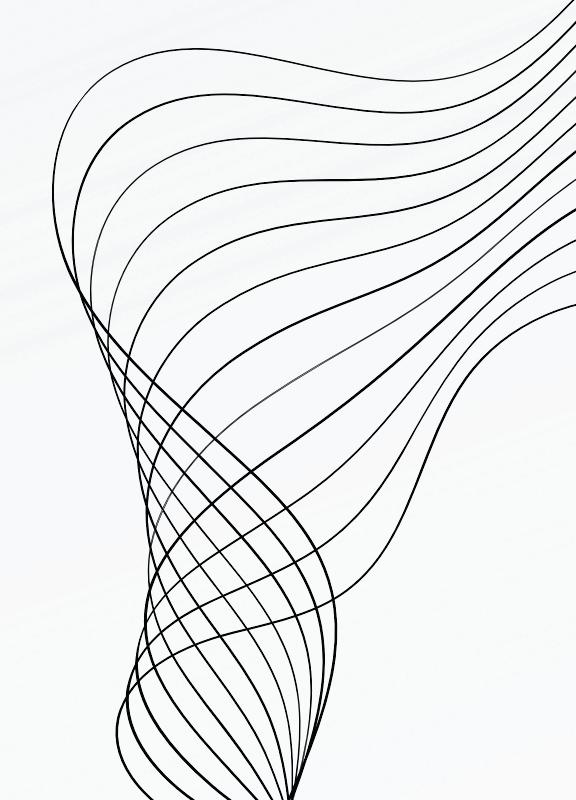


DATAFACTORY REPOSITORY

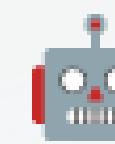
[LINK](#)



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