Azure Synapse - Serverless Pool

Dataset

Source link : https://developer.nrel.gov/docs/transportation/alt-fuel-stations-v1/

The data was obtained using an **API** from The **National Renewable Energy Laboratory (NREL) website**. An **API key** was **registered** and used to retrieve the data in **JSON format**. This dataset supports the Alternative Fueling Station Locator on the Alternative Fuels Data Center. The data includes station locations for biodiesel, compressed natural gas, ethanol, electric charging, hydrogen, liquefied natural gas, and propane. However, only **electric charging stations** were **filtered** for analysis.

Challenges

Configuring the REST API linked service to copy JSON data into a file in my ADLS container. Since the JSON data was a very large file, it took a significant amount of time to parse each field and record as needed.

Actions Taken

YouTube tutorials helped me with my linked service, and multiple errors and trials helped me parse the JSON properly.

Data Ingestion

API to JSON file in ADLS

A Copy activity was set up in the pipeline to copy the JSON data into a JSON file in ADLS Gen2.

Copy data

Copy Json

Parameters Variables Settings **Output**

Pipeline run ID: 4de7842b-c6c2-4015-8e8d-ad1d19a143

A REST API linked service was used as the source for the Copy activity, with JSON as the sink to save the data as a JSON file.

Edit linked service

REST Learn more ⬈

Base URL *

https://developer.nrel.gov/api/alt-fuel-stations/v1.json?fuel_type_code=ELEC&api_key=@li

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

Authentication type *

Anonymous ⌄

Server certificate validation ⓘ

◉ Enable   ○ Disable

Auth headers ⓘ

Parameters

+ New | 🗑 Delete

	Name	Type	Default value	
☐	API	String ⌄	W4An9PTBZNUwq1	🗑

> Advanced ⓘ

Pipeline
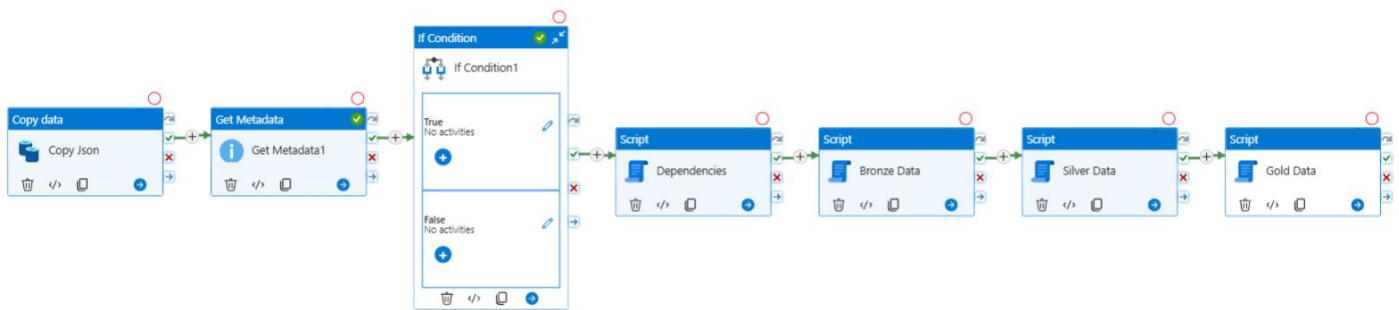
Data Ingestion and Processing Workflow

1. Copy Activity (Data Extraction)

The first step involves extracting the JSON file from the source website using a REST API. The Copy Activity is responsible for transferring this JSON file from the API source to an Azure Data Lake Storage (ADLS) container, which serves as the sink.

2. Metadata Extraction

A Get Metadata Activity is used to retrieve information about the files in the source container. This helps in identifying the available child items (files) for further processing.

3. Conditional Check for File Availability

To ensure the pipeline runs only when there is at least one file available, an If Condition Activity is implemented. If no files are found, the pipeline stops execution; otherwise, the process continues.

4. Automation with Script Activities

To automate the setup and processing, four Script Activities are incorporated:

- Script Activity 1: Creates the database, schema, external table, and external file format necessary for processing and querying the ingested data.

- Script Activities 2, 3, and 4: Handle data transformation as part of the Medallion Architecture:

    o Bronze Layer: Raw data ingestion.

    o Silver Layer: Data cleansing and transformation.

    o Gold Layer: Aggregated and optimized data for reporting and analytics.

Create External Data Source:

```sql
1    CREATE DATABASE big;
2    ALTER DATABASE big COLLATE Latin1_General_100_CI_AI_SC_UTF8;
3    USE big;
4    CREATE SCHEMA Bronze;
5    GO
6    CREATE SCHEMA Silver;
7    GO
8    CREATE SCHEMA Gold;
9    GO
10   CREATE EXTERNAL DATA SOURCE all_stations
11   WITH
12   (
13       LOCATION='abfss://taxi-data@synapsecontainerazbr.dfs.core.windows.net'
14   );
15   GO
```

Create External File Formats

```
16   CREATE EXTERNAL FILE FORMAT csv_format
17   WITH
18   (
19       FORMAT_TYPE = DELIMITEDTEXT,
20       FORMAT_OPTIONS
21       (
22            FIELD_TERMINATOR = ',',
23            FIRST_ROW = 2,
24            STRING_DELIMITER = '"',
25            ENCODING = 'UTF8',
26            USE_TYPE_DEFAULT = FALSE,
27            PARSER_VERSION = '2.0'
28       )
29   );
30   GO
```

Create an external table to represent the raw data (Bronze Layer)

```
32   /*Creating Bronze Table - Raw Data*/
33   IF OBJECT_ID('Bronze.ev_charging_raw') IS NOT NULL
34       DROP EXTERNAL TABLE Bronze.ev_charging_raw;
35   CREATE EXTERNAL TABLE Bronze.ev_charging_raw
36   WITH
37   (
38       LOCATION='Medallion/Bronze',
39       DATA_SOURCE=all_stations,
40       FILE FORMAT=csv format
76   ev_level1_evse_num,
77   ev_level2_evse_num,
78   ev_network,
79   ev_network_web,
80   ev_pricing,
81   ev_renewable_source
82   FROM
83       OPENROWSET(
84           BULK 'Source/*.json',
85           DATA_SOURCE='all_stations',
86           FORMAT = 'CSV',
87           FIELDQUOTE = '0x0b',
88           FIELDTERMINATOR ='0x0b',
89           ROWTERMINATOR = '0x0a',
90           PARSER_VERSION='1.0'
91       )
92   WITH (
93           JsonDoc NVARCHAR(MAX)
94       ) AS ev_json
```

```
 79    CROSS APPLY OPENJSON(JsonDoc)
 80    WITH (
 81        fuel_stations NVARCHAR(MAX) as JSON
 82    )
 83    CROSS APPLY OPENJSON(fuel_stations)
 84    WITH (
 85            access_code VARCHAR(50) '$.access_code',
 86        access_days_time VARCHAR(255) '$.access_days_time',
 87        access_detail_code VARCHAR(200) '$.access_detail_code',
 88        fuel_type_code VARCHAR(50) '$.fuel_type_code',
 89        cards_accepted VARCHAR(100) '$.cards_accepted',
 90        date_last_confirmed VARCHAR(50) '$.date_last_confirmed',
 91        open_date VARCHAR(50) '$.open_date',
 92        groups_with_access_code VARCHAR(100) '$.groups_with_access_code',
 93        id INT '$.id',
 94        maximum_vehicle_class VARCHAR(100) '$.maximum_vehicle_class',
 95        owner_type_code VARCHAR(10) '$.owner_type_code',
 96        station_name VARCHAR(255) '$.station_name',
 97        city VARCHAR(100) '$.city',
 98        [state] VARCHAR(50) '$.state',
 99        street_address VARCHAR(255) '$.street_address',
100        ev_dc_fast_num INT '$.ev_dc_fast_num',
101        ev_level1_evse_num INT '$.ev_level1_evse_num',
102        ev_level2_evse_num INT '$.ev_level2_evse_num',
103        ev_network VARCHAR(255) '$.ev_network',
104        ev_network_web VARCHAR(255) '$.ev_network_web',
105        ev_pricing VARCHAR(255) '$.ev_pricing',
106        ev_renewable_source VARCHAR(200) '$.ev_renewable_source'
107    )
108
```

O/P:

access_code	access_days_ti...	access_detail_c...	fuel_type_code	cards_accepted	date_last_confi...	open_date	groups_with_a...	id	maximum_vehi...	owner_type_co...	station_name
private	(NULL)	(NULL)	CNG	(NULL)	2024-10-11	2010-12-01	Private	17	MD	T	Spire - Montc
private	(NULL)	GOVERNMENT	CNG	(NULL)	2024-02-12	1996-12-15	Private - Gover...	45	LD	LG	Metropolitan
private	(NULL)	(NULL)	CNG	(NULL)	2023-12-13	1997-01-01	Private	64	HD	P	United Parcel
public	24 hours daily	CREDIT_CARD_...	CNG	CREDIT M V Vo...	2024-04-14	1997-01-01	Public - Credit ...	73	MD	T	Arkansas Okla
public	24 hours daily; ...	CREDIT_CARD_...	CNG	A CleanEnergy ...	2024-12-10	1996-11-15	Public - Credit ...	81	MD	SG	Clean Energy
public	24 hours daily; ...	CREDIT_CARD_...	CNG	CleanEnergy D ...	2024-12-10	1996-11-15	Public - Credit ...	84	HD	T	Clean Energy
public	24 hours daily; ...	CREDIT_CARD_...	CNG	CleanEnergy D ...	2024-12-10	2016-07-15	Public - Credit ...	108	HD	T	Clean Energy
public	24 hours daily; ...	CREDIT_CARD_...	CNG	CleanEnergy D ...	2024-01-09	1988-01-15	TEMPORARILY ...	112	MD	T	Canarsie - Na
public	8am-5pm M-F;...	KEY_ALWAYS	CNG	(NULL)	2024-10-11	1998-01-15	Public - Card h...	120	MD	T	Con Edison

Data Cleaning and Transformation: Silver

```
110    /*Creating Silver Table - Transformed Data*/
111  ∨ IF OBJECT_ID('Silver.ev_charging_transformed') IS NOT NULL
112        DROP EXTERNAL TABLE Silver.ev_charging_transformed;
113    CREATE EXTERNAL TABLE Silver.ev_charging_transformed
114    WITH
115  ∨ (
116        LOCATION='Medallion/Silver',
117        DATA_SOURCE=all_stations,
118        FILE_FORMAT=csv_format
119    )
120    AS
```

```
121    SELECT
122    access_code,
123    access_days_time,
124    access_detail_code,
125    fuel_type_code,
126    cards_accepted,
127    CAST(date_last_confirmed AS DATE) AS date_last_confirmed,
128    CAST(open_date AS DATE) AS open_date,
129    groups_with_access_code,
130    id,
131    maximum_vehicle_class,
132    owner_type_code,
133    station_name,
134    city,
135    [state],
136    street_address,
137    CAST(ev_dc_fast_num AS SMALLINT) AS ev_dc_fast_num,
138    CAST(ev_level1_evse_num AS INT) AS ev_level1_evse_num,
139    CAST(ev_level2_evse_num AS INT) AS ev_level2_evse_num,
140    ev_network,
141    ev_network_web,
142    ev_pricing,
143    ev_renewable_source
144    FROM Bronze.ev_charging_raw
```

O/p:

Results Messages

View Table Chart   ↦ Export results ∨

🔍 Search

ards_accepted	date_last_confi...	open_date	groups_with_a...	id	maximum_vehi...	owner_type_co...	station_name	city	state	street_address	ev_dc_fast_num
NULL)	2023-09-14	1999-10-15	Private	1517	(NULL)	LG	LADWP - Trues...	Sun Valley	CA	11797 Truesdal...	2
NULL)	2023-01-10	1995-08-30	Public	1523	LD	P	Los Angeles Co...	Los Angeles	CA	1201 S Figuero...	(NULL)
NULL)	2023-09-14	1999-10-15	Private	1525	LD	LG	LADWP - John ...	Los Angeles	CA	111 N Hope St	12
NULL)	2024-01-09	2018-05-01	Private	1531	(NULL)	LG	LADWP - Hayn...	Long Beach	CA	6801 E 2nd St	1
NULL)	2024-01-09	1999-10-15	Private	1552	(NULL)	LG	LADWP - Harb...	Wilmington	CA	161 N Island Ave	(NULL)
NULL)	2024-01-09	2016-01-01	Private - Gover...	1556	(NULL)	LG	LADWP - Sylma...	Sylmar	CA	13201 Sepulve...	(NULL)
NULL)	2024-01-09	1999-10-15	Private	1572	(NULL)	LG	LADWP - EV Se...	Los Angeles	CA	1630 N Main St	1
NULL)	2024-01-09	2019-04-01	Private	1573	(NULL)	LG	LADWP - Fairfa...	Los Angeles	CA	2311 S Fairfax ...	(NULL)
NULL)	2023-09-14	1999-10-15	Private	1589	(NULL)	LG	LADWP - Palme...	Los Angeles	CA	1212 Palmetto St	(NULL)
NULL)	2024-08-15	1997-07-30	Public	6355	(NULL)	P	Scripps Green ...	La Jolla	CA	10666 N Torrey...	(NULL)

Gold

```sql
151    /*Creating Gold Table - Transformed Data*/
186    END AS owner_type,
187    station_name,
188    city,
189    [state],
190    street_address,
191    ev_dc_fast_num,
192    ev_level1_evse_num,
193    ev_level2_evse_num,
194    ev_network,
195    ev_network_web,
196    ev_pricing,
197    ev_renewable_source,
198    CASE
199        WHEN access_days_time LIKE '%24%' THEN 'Yes'
200        ELSE 'No'
201    END AS allday_open
202    FROM Silver.ev_charging_transformed;
203
171    id,
172    CASE
173        maximum_vehicle_class
174        WHEN 'LD' THEN 'Passengers'
175        WHEN 'MD' THEN 'Medium-duty'
176        WHEN 'HD' THEN 'Heavy-duty'
177    END AS Max_Vehicle_class,
178    CASE
179        owner_type_code
180        WHEN 'FG' THEN 'Federal Government'
181        WHEN 'J' THEN 'Jointly Owned'
182        WHEN 'LG' THEN 'Local Government'
183        WHEN 'P' THEN 'Privately'
184        WHEN 'SG' THEN 'State Government'
185        WHEN 'T' THEN 'Utility'
```

O/P:

renewable...	allday_open
ULL)	No
ULL)	No
ULL)	No
ULL)	No
ULL)	No
ULL)	No
ULL)	No
ULL)	No
ULL)	No

Max_Vehicle_class	owner_type
(NULL)	Local Government
Passengers	Privately
Passengers	Local Government
(NULL)	Local Government
(NULL)	Local Government
(NULL)	Local Government
(NULL)	Local Government
(NULL)	Local Government
(NULL)	Local Government

View

```sql
220    CREATE VIEW Gold.years_functioning
221    AS
222    SELECT
223    access_days_time,
224    cards_accepted,
225    YEAR(date_last_confirmed)-YEAR(open_date) AS active_years,
226    Max_Vehicle_class,
227    owner_type,
228    station_name,
229    city,
230    [state],
231    ev_network,
232    ev_pricing,
233    allday_open
234    FROM Gold.ev_charging_gold
235    WHERE cards_accepted IS NOT NULL;
```

O/P:

Results   Messages

View   Table   Chart   ⊢ Export results ⌄

🔍 Search

| access_days_time	cards_accepted	active_years	Max_Vehicle_cl...	owner_type	station_name	city	state	ev_network	ev_pricing	allday_open
24 hours daily	Proprietor	10	Passengers	Privately	LA County DPR...	Arcadia	CA	POWERFLEX	$0.45 per kWh ...	Yes
(NULL)	Proprietor	9	Passengers	Local Governm...	LA County DP...	Baldwin Park	CA	Non-Networked	(NULL)	No
(NULL)	Proprietor	7	Passengers	Local Governm...	LA County DP...	South Gate	CA	Non-Networked	$0.45 per kWh ...	No
24 hours daily	D	20	Passengers	Local Governm...	Stars Recreatio...	Vacaville	CA	Non-Networked	Free	Yes
Restricted to garage hours; pay l...	A ANDROID_PA...	11	Passengers	Local Governm...	New London Pa...	New London	CT	Non-Networked	Free	No
24 hours daily; guests may reque...	A Cash D M V	12	Passengers	Privately	Super 8 - Ukiah	Ukiah	CA	Non-Networked	$10 per session	Yes
24 hours daily; pay lots	A D M V	13	Passengers	Privately	O'Hare Airport	Chicago	IL	Non-Networked	Free	Yes
24 hours daily; pay lot	A D M V	12	Passengers	State Governm...	Greenville-Spar...	Greer	SC	Non-Networked	Free	Yes
Restricted to garage hours; pay lot	M V	11	Passengers	Local Governm...	Elizabeth Parkin...	Elizabeth	NJ	Non-Networked	Free	No
Garage business hours; pay lot	Cash M V	12	Passengers	Privately	The Green Parki...	Charlotte	NC	Non-Networked	Free	No

Power BI

I used the **workspace SQL endpoint link** from the **Manage** tab in Synapse Studio to connect my Synapse account to Power BI. From there, I selected my database and chose the external tables I wanted to work with. This dashboard is built using the external tables from both the silver and gold schemas:
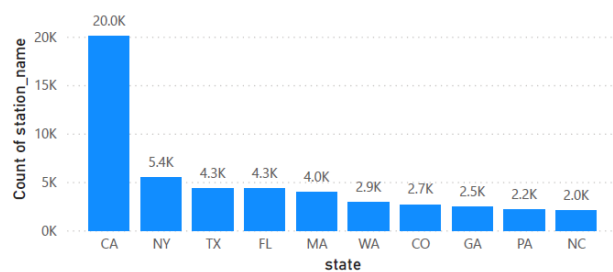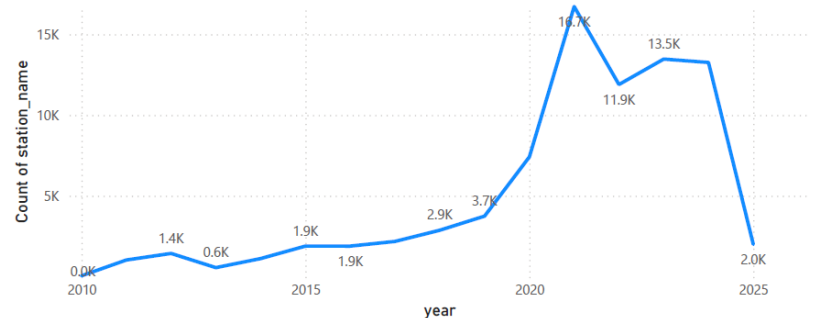
Owner type distribution
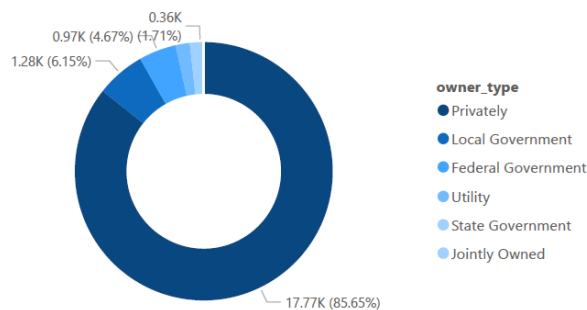
Top 3 EV Networks
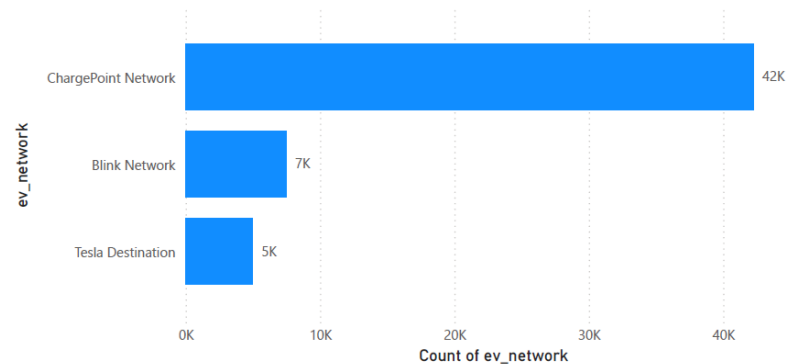
Top 10 States with the Most Charging Stations

Count of Stations opened by Year

Business Recommendations

1. Owner Type Distribution Insight

Most EV charging stations (about 85%) are privately owned. Government-owned and utility-run stations make up a tiny portion. Private companies are leading the way in charging infrastructure, but there's not enough public/government investment.

Recommendation:

- Encourage more public-private partnerships to balance the ownership structure and increase accessibility.

- Provide incentives to local governments and utilities to set up more public EV charging stations, ensuring coverage in rural areas.

2. Top EV Networks Analysis

ChargePoint dominates the market with 42K stations, leaving Blink (7K) and Tesla Destination (5K) way behind. Other networks are barely competing, which could mean fewer choices for EV use.

Recommendation:

Diversify EV network partnerships to reduce dependency on a single provider.

3. Charging Stations Opened by Year

We can notice a massive spike in new station openings around 2020-2021 (peaking at 15K), but things have slowed down since.

Recommendation:

Investigate the reasons for the drop in station openings post-2021 (policy changes, supply chain issues, funding restrictions).

Predicting future station needs using historical growth trends and expected EV adoption rates.