

Data Ingestion

CREATE EXTERNAL TABLE AS SELECT (CETAS)

CREATE EXTERNAL TABLE AS SELECT (CETAS)

```
CREATE EXTERNAL TABLE {[database_name .] [ schema_name . ] table_name }  
    [(column_name [,...n ] ) ]  
WITH (  
    LOCATION = 'path_to_folder',  
    DATA_SOURCE = external_data_source_name,  
    FILE_FORMAT = external_file_format_name  
)  
AS  
<select_statement>  
[;]
```

Comienza con la sentencia CREATE EXTERNAL TABLE seguida del nombre de la tabla. Puedes especificar el nombre completo de la tabla si quieres que se cree en una base de datos o esquema específico. La clase WITH define la parte de la carpeta donde se van a crear los archivos. Luego tenemos los nombres de la External Data Source y el External File format que se va a escribir. En la parte inferior tenemos la sentencia SELECT.

CREATE EXTERNAL TABLE AS SELECT (CETAS)

```
CREATE EXTERNAL TABLE transformed.taxi_zone  
WITH (  
    LOCATION = 'transformed/taxi_zone',  
    DATA_SOURCE = nyc_taxi_data,  
    FILE_FORMAT = parquet_file_format  
)  
AS  
SELECT *  
FROM raw.taxi_zone ;
```

Veamos un par de ejemplos. En este ejemplo, simplemente estamos seleccionando los datos de nuestra tabla **raw**, que está en el formato CSV, y luego escribimos ese archivo de casos de apoyo y creamos la tabla externa en el archivo de Parquet. Así que lo que hemos hecho es que hemos transformado los datos CSV al formato Parquet para que sea eficiente para las cargas de trabajo analíticas.

En este ejemplo, estamos utilizando la función OPENROWSET para leer los datos de un storage y luego escribir esos datos de nuevo como back file y crear una EXTERNAL TABLE sobre eso. Así que, de nuevo, aquí hemos pasado de un formato CSV a un formato de Parquet para apoyar a las cargas de trabajo analíticas, y también hemos creado una EXTERNAL TABLE en la parte superior.

CREATE EXTERNAL TABLE AS SELECT (CETAS)

```
CREATE EXTERNAL TABLE transformed.taxi_zone
WITH (
    LOCATION = 'transformed/taxi_zone',
    DATA_SOURCE = nyc_taxi_data,
    FILE_FORMAT = parquet_file_format
)
AS
SELECT *
FROM
OPENROWSET(
    BULK 'abfss://nyc-taxi-data@synapsecoursedl.dfs.core.windows.net/raw/taxi_zone.csv',
    FORMAT = 'CSV',
    PARSER_VERSION = '2.0',
    FIRSTROW = 2
)
WITH (
    location_id SMALLINT 1,
    borough VARCHAR(15) 2,
    zone VARCHAR(50) 3,
    service_zone VARCHAR(15) 4
) AS [result];
```

Odemy

Y finalmente, en este ejemplo, estamos creando una EXTERNAL TABLE que contiene el número de zonas de cada municipio (borough). Este es un ejemplo en el que estamos estudiando los resultados agregados al storage. Como puedes ver, sólo necesitas una sentencia SELECT válida y que pueda contener agregaciones gigantes, etc.

CREATE EXTERNAL TABLE AS SELECT (CETAS)

```
CREATE EXTERNAL TABLE transformed.taxi_borough
WITH (
    LOCATION = 'transformed/taxi_borough',
    DATA_SOURCE = nyc_taxi_data,
    FILE_FORMAT = parquet_file_format
)
AS
SELECT borough, COUNT(1) AS number_of_zones
FROM raw.taxi_zone
GROUP BY borough;
```