**Module 3 – Lesson 2 – Demo 1 (M03\_L02\_Demo01)**

# **Objective*:*** The goal of this demo is to demonstrate the ease of copying the by using the CTAS command and to demonstrate how the caching works.

**Setup:**

# **Objective*:*** If SQL pool database is available from a previous step, with the NYCTaxi data, use that, do not create a new database – skip the below steps.

1. Create Synapse SQL Pool.
2. The scripts to create tables for NYC Taxi can be copied from the link below by navigating to the section “**Create tables for the sample data**”.
3. The data can be loaded into the tables created for NYC Taxi by using the below link and navigating to the section “**Load the data into your data warehouse**”.
4. Visual Studio (latest) and SSMS (latest) should be installed and connected to your Synapse SQL Pool DB.

**Note:** Link for creating and loading the data into NYC Taxi

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/load-data-from-azure-blob-storage-using-copy>

You can also create and load data by using the below script files placed with other script files.

**Creating table - NY\_Taxicab\_Data\_Set\_CREATE\_STABLE\_Script.sql**

**Loading Data - NY\_Taxicab\_Data\_Set\_DATA\_LOAD\_Script.sql**

**NOTE: the scripts load a different set of data, which will change the performance and rowcounts, only use the scripts if the UI load is not available.**

**Demo steps:**

1. The below script would drop the dimDate, fctTrip and fctTrip\_RR tables.

IF OBJECT\_ID('dimDate') IS NOT NULL DROP TABLE [dimDate]

IF OBJECT\_ID('fctTrip') IS NOT NULL DROP TABLE [fctTrip]

IF OBJECT\_ID('fctTrip\_RR') IS NOT NULL DROP TABLE [fctTrip\_RR]

1. Run the below scripts to create REPLICATED, HASH DISTRIBUTED and ROUND-ROBIN DISTRIBUTED table using CTAS command, that would copy the table and data from existing table.

--Create tables - REPLICATED / HASH DISTRIBUTED / ROUND-ROBIN DISTRIBUTED

CREATE TABLE [dimDate]

WITH

(DISTRIBUTION = REPLICATE,

CLUSTERED INDEX (DateID) )

AS

SELECT \*

FROM [dbo].[Date]

CREATE TABLE [fctTrip]

WITH

(DISTRIBUTION = HASH(DateID),

CLUSTERED COLUMNSTORE INDEX )

AS

SELECT \*

FROM [dbo].[Trip]

CREATE TABLE [fctTrip\_RR]

WITH

(DISTRIBUTION = ROUND\_ROBIN,

CLUSTERED COLUMNSTORE INDEX )

AS

SELECT \*

FROM [dbo].[Trip]

1. Once the tables have been created you can refresh the explorer and show the new tables, different icons by geometry.
2. Run the below command to Show dimDate status as **not cached**.

SELECT o.[name] AS TableName,

rcs.[state]

FROM sys.pdw\_replicated\_table\_cache\_state rcs

INNER JOIN sys.objects o

ON rcs.object\_id = o.object\_id

WHERE o.[name] = 'dimDate'

The output from the above command reports the state as “**NotReady”** as shown below

A screenshot of a cell phone

Description automatically generated

1. Run the below select command it will **Cache** dimDate.

SELECT top 1 \*

FROM dimDate

1. Re-run the same command from step “4” and now you will **see dimDate status as cached**.

SELECT o.[name] AS TableName,

rcs.[state]

FROM sys.pdw\_replicated\_table\_cache\_state rcs

INNER JOIN sys.objects o

ON rcs.object\_id = o.object\_id

WHERE o.[name] = 'dimDate'

The output from the above command reports the state as “**Ready”** as shown below

A screenshot of a cell phone

Description automatically generated

1. Show table details with DBCC PDW\_SHOWSPACEUSED. RR table is not *exactly* evenly distributed

-- g. Show table details

DBCC PDW\_SHOWSPACEUSED ('fctTrip');

DBCC PDW\_SHOWSPACEUSED ('dbo.fctTrip\_RR');