Microsoft Fabric Use Case Demo Steps:

Step 1: Create LakeHouse, upload data to LakeHouse, and create Delta Parquet Table using Fabric Spark

a) Create a new WorkSpace
b) Select data engineering personas from the left bottom of the dashboard
c) Create a new LakeHouse
d) On the WorkSpace newly created, import the spark script, and open it
e) If the newly created LH is not visible, click on remove all LakeHouse and add the new LakeHouse
f) Before proceeding with the next step make sure all the files are uploaded in the files section of LakeHouse and the path is correctly entered in the script as well

Step 2: Create a Star Schema to be used for Reporting

a) Open your Fabric Workspace and create a new data warehouse using the menu on the top left corner
b) Create a new Warehouse or use an existing one. Examples in this demo will use the name CMS_Warehouse
c) Click the + Warehouses button and select the LakeHouse containing the flattened table of CMS data. Once you are done, you can see the LakeHouse table that you created in LakeHouse
d) Create a new SQL view for each of the scripts linked below in this repo. Follow the numeric order in the titles of the scripts:

    01_cms_provider_dim_year.sql

    02_cms_provider_dim_drug.sql

    03_cms_provider_dim_geography.sql

    04_cms_provider_dim_provider.sql

    05_cms_provider_no_null_key.sql

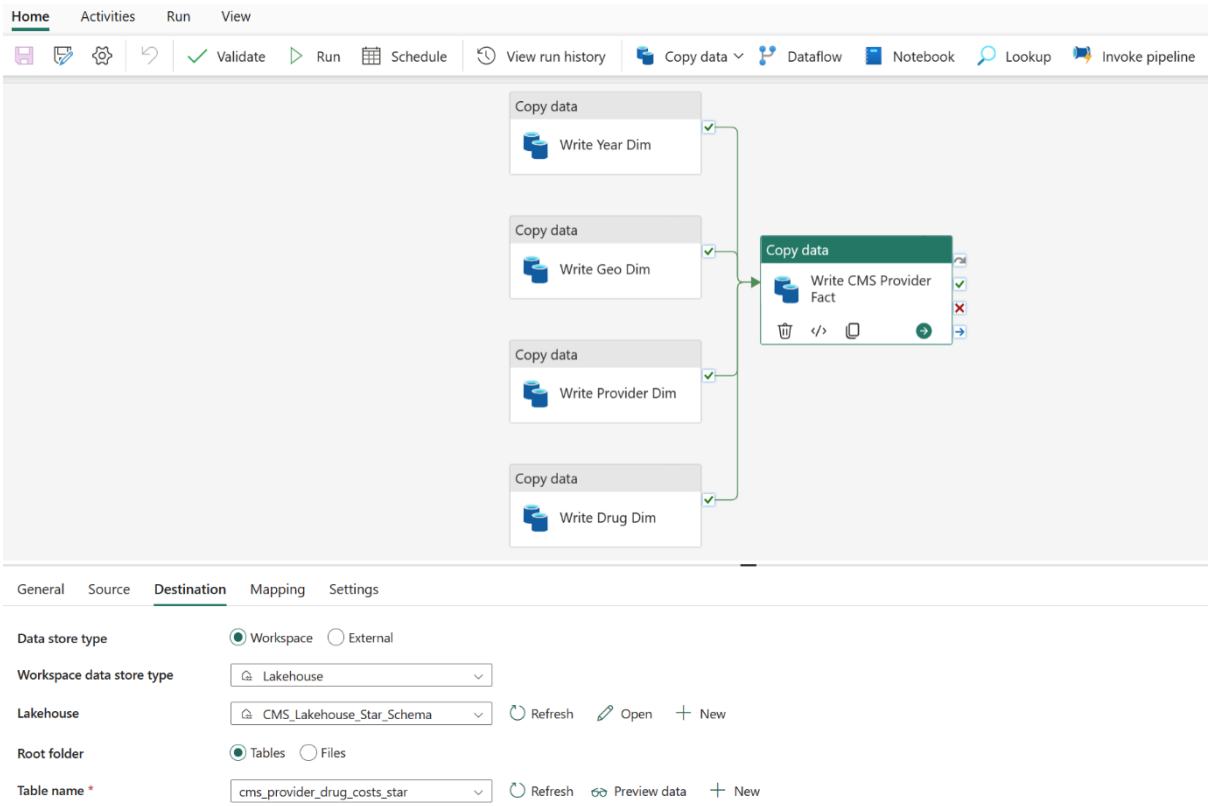
    06_cms_provider_drug_costs_star.sql

Step 3: Create Pipeline and set up Warehouse SQL views to populate LakeHouse tables

a. From the Workspace select +New > Show all > Data pipeline
b. Name the pipeline and select Add > Copy data
c. Rename the Copy data activity to Write Year Dim
d. Change the source to the Workspace's Fabric Warehouse table dbo.cms_dim_year
e. Change the destination to the Workspace's Fabric LakeHouse Tables and name it cms_provider_dim_year
f. Import and validate the schema for the Mapping

g. Select Add > Copy data and begin the process in steps 3-6 above of adding separate activities for all of the SQL views you created in the Warehouse:

Activity name	Warehouse Source table (SQL view)	Lakehouse Destination table (delta parquet)
Write Year Dim	dbo.cms_dim_year	cms_provider_dim_year
Write Geo Dim	dbo.cms_provider_dim_geography	cms_provider_dim_geography
Write Provider Dim	dbo.cms_provider_dim_provider	cms_provider_dim_provider
Write Drug Dim	dbo.cms_provider_dim_drug	cms_provider_dim_drug
Write CMS Provider Fact	dbo.cms_provider_drug_costs_star	cms_provider_drug_costs_star



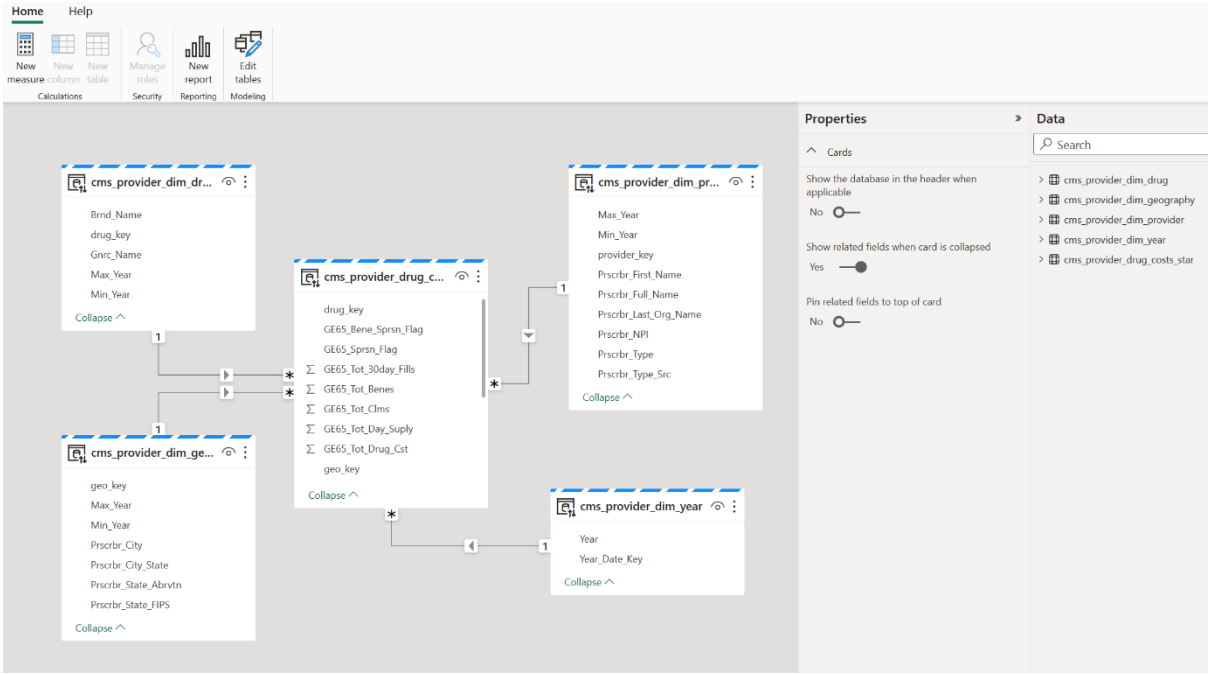
h. On the Pipeline ribbon, click Run and the Pipeline will populate the Fabric LakeHouse with the dimensions and fact table for the CMS data. You do not need to schedule the Pipeline since it is a one-time load.

Step 4: Create the Direct Lake Power BI Star Schema Dataset with DAX expressions and metadata

a) Create powerbi dataset inside the workspace

b) Prepare a new semantic model for the PBI reports to consume
c) consider only the 'costar' fact table and merge with other dim tables
d) rename table cols on the right side in the pbi dataset
e) add your DAX expressions
f) Create relationships between the dimension tables and the fact table

Lakehouse Table Name	Dim Table Primary Key	Fact Table Foreign Key	Cardinality	Cross Filter Direction
cms_provider_dim_year	Year	Year	One to Many	Single
cms_provider_dim_drug	drug_key	drug_key	One to Many	Single
cms_provider_dim_geography	geo_key	geo_key	One to Many	Single
cms_provider_dim_provider	provider_key	provider_key	One to Many	Single



Step 5:

a) Assign user-friendly names to the columns for user-facing values and hide columns that will be built into Calculated Measures (Step 3).
b) All columns on the fact table are hidden since the filter values are now in dimensions.

This is an optional step for renaming the columns so the report column headers will be meaningful. But hiding the fact table cols is necessary as we will show only calculated measures.

Lakehouse Table Name	Lakehouse Table Column Name	New Dataset Column Name	Is hidden
cms_provider_dim_year	Year	Year	No
cms_provider_dim_year	Year_Date_Key	Year_Date_Key	Yes
cms_provider_dim_drug	Brnd_Name	Brand Name	No
cms_provider_dim_drug	Gnrc_Name	Generic Name	No
cms_provider_dim_drug	Max_Year	Max_Year_drug	Yes
cms_provider_dim_drug	Min_Year	Min_Year_drug	Yes
cms_provider_dim_drug	drug_key	drug_key	Yes
cms_provider_dim_geography	Prscrbr_City	Prescriber City	No
cms_provider_dim_geography	Prscrbr_City_State	Prescriber City State	No
cms_provider_dim_geography	Prscrbr_State_Abrvtn	Prescriber State	No
cms_provider_dim_geography	Prscrbr_State_FIPS	Prescriber State FIPS	Yes
cms_provider_dim_geography	Max_Year	Max_Year_geo	Yes
cms_provider_dim_geography	Min_Year	Min_Year_geo	Yes
cms_provider_dim_geography	geo_key	geo_key	Yes
cms_provider_dim_provider	Prscrbr_First_Name	Prescriber First Name	Yes
cms_provider_dim_provider	Prscrbr_Full_Name	Prescriber Full Name	No
cms_provider_dim_provider	Prscrbr_Last_Org_Name	Prescriber Last Name	Yes
cms_provider_dim_provider	Prscrbr_NPI	Prescriber NPI	No
cms_provider_dim_provider	Prscrbr_Type	Prescriber Type	No

Lakehouse Table Name	Lakehouse Table Column Name	New Dataset Column Name	Is hidden
cms_provider_dim_provider	Prscrbr_Type_Src	Prescriber Type Source	Yes
cms_provider_dim_provider	Max_Year	Max_Year_provider	Yes
cms_provider_dim_provider	Min_Year	Min_Year_provider	Yes
cms_provider_dim_provider	provider_key	provider_key	Yes
dbo.cms_provider_drug_costs_star	GE65_Bene_Sprsn_Flag	65 or Older Beneficiaries Suppression Flag	Yes
dbo.cms_provider_drug_costs_star	GE65_Sprsn_Flag	65 or Older Suppression Flag	Yes
dbo.cms_provider_drug_costs_star	GE65_Tot_30day_Fills	65 or Older Total 30 Day Fills	Yes
dbo.cms_provider_drug_costs_star	GE65_Tot_Benes	65 or Older Total Beneficiaries	Yes
dbo.cms_provider_drug_costs_star	GE65_Tot_Clms	65 or Older Total Claims	Yes
dbo.cms_provider_drug_costs_star	GE65_Tot_Day_Suply	65 or Older Total Days Supply	Yes
dbo.cms_provider_drug_costs_star	GE65_Tot_Drug_Cst	65 or Older Total Drug Cost	Yes
dbo.cms_provider_drug_costs_star	Tot_30day_Fills	Tot_30day_Fills	Yes
dbo.cms_provider_drug_costs_star	Tot_Benes	Tot_Benes	Yes
dbo.cms_provider_drug_costs_star	Tot_Clms	Tot_Clms	Yes
dbo.cms_provider_drug_costs_star	Tot_Day_Suply	Tot_Day_Suply	Yes
dbo.cms_provider_drug_costs_star	Tot_Drug_Cst	Tot_Drug_Cst	Yes
dbo.cms_provider_drug_costs_star	Year	Year	Yes
dbo.cms_provider_drug_costs_star	drug_key	drug_key	Yes
dbo.cms_provider_drug_costs_star	geo_key	geo_key	Yes

Lakehouse Table Name	Lakehouse Table Column Name	New Dataset Column Name	Is hidden
dbo.cms_provider_drug_costs_star	provider_key	provider_key	Yes

Step 6: Add the DAX measures as per your reporting requirement

Measure name	DAX Syntax	Format	%format	Thousands separator	Decimal places	Data category
Row Count	`Row Count = COUNTROWS('cms_provider_drug_costs_star')`	Whole Number	No	Yes	0	Uncategorized
Total Claims	`Total Claims = SUM(cms_provider_drug_costs_star[Tot_Clms])`	Whole Number	No	Yes	0	Uncategorized
Total Beneficiaries	`Total Beneficiaries = SUM(cms_provider_drug_costs_star[Tot_Benes])`	Whole Number	No	Yes	0	Uncategorized
Total 30 Day Fills	`Total 30 Day Fills = SUM(cms_provider_drug_costs_star[Tot_30day_Fills])`	Decimal	No	Yes	1	Uncategorized
Total Day Supply	`Total Days Supply = SUM(cms_provider_drug_costs_star[Tot_Day_Suply])`	Whole Number	No	Yes	0	Uncategorized
Total Drug Cost	`Total Drug Cost = SUM(cms_provider_drug_costs_star[Tot_Drug_Cst])`	Currency	No	Yes	0	Uncategorized
Cost per Claim	`Cost per Claim = DIVIDE([Total Drug Cost],[Total Claims])`	Currency	No	Yes	0	Uncategorized
Cost per Day	`Cost per Day = DIVIDE([Total Drug Cost],[Total Days Supply])`	Currency	No	Yes	2	Uncategorized
Days per Claim	`Days per Claim = DIVIDE([Total Days Supply],[Total Claims])`	Decimal	No	Yes	1	Uncategorized

Step 7:

a) From the WorkSpace home dashboard create a new report and call the semantic model
b) Click on the columns with the measures to be viewed as cards in the report-->save

“WS Dashboard-->create new report--> select semantic model--> select the cols and measures for the cards”

The Power BI dataset now exists within Fabric, no caching or refreshing is needed! You can go back to your Workspace and rename the dataset, which shows up as a new artifact in the Fabric Workspace. Now you can adjust some settings for the dataset to potentially enable better performance:

- From the Workspace, click on the ellipse next to the dataset name.
- Choose **Settings**
- Select **Query caching** > **On** to cache reporting results and improve performance for end users
- Turn on the **Large dataset storage format** which may help with the large data volumes in the fact table