Problem Statement Data Engineering

Business Description

ABC Automobile is a manufacturing company that sells all kinds of automobiles and lately they have been noticing that some areas are performing better than others. As of now, the company does not have any data tools that offer some transparency to the managers making this process quite tedious. After some deep evaluation and feasibility process, the managers at ABC Automobile decided to reach out to Prerana's company to help them build a Data Driven solution which will help them understand sales trend and come up with better future strategies for sales growth.

Challenge

- The complete solution must leverage Azure Resources and include daily/weekly refresh of data should be automated and available on the report.
- Client needs a robust BI solution to support and manage company sales
- Client demands that the managers should be able to access the data solution and be able to share it with other company heads.

Solution

The company has decided to deliver a Sales Dashboard using Azure Cloud Technology. Prerana's company has suggested Azure Synapse for data ingestion (from excel) and transformation and creating the dashboards on Power BI for better insights and self-service capabilities. ABC Automobile would like that the complete process can be automated using Azure Data Services, like extracting the data from SQL DB, ingesting it to ADLS Gen2 using Synapse Pipeline and then transform the data using synapse spark notebooks, creating views on top of the transformed dataset to feed it to Power BI and later auto refresh of Power BI Dashboard is published on a workspace in the Power BI Service so that the managers can share and edit as they wish. The company together with Prerana's company have gathered a list of parameters and Key performance indicators (KPIs) that they would like to be existent in the Power BI Dashboard.

Parameters and Filters

- Automated process of ingesting the data from source (manual/SharePoint).
- Creating tables based on the data model.
- Transforming and Ingesting data to DIM and Fact Tables.
- Daily/Weekly refresh of data.
- Ability to select a date period based on years, months, or days
- Ability to filter by customer, product, and country

Key KPIs:

- Which period had the most sales in a specific year?
- Which country is seeing the most success?
- Which product category should the company continue to invest in?
- What is our Sales trend? Comparison of current year sales with prior year sales Create a measure to compare YOY sales. (Naming – **YOY**)
- What are the Total Sales per Customer? Manager should be able to slice and dice the data by product line, product category and city. (Hint: Use Parameterized field)
- Average for disputed Sales for 2003 and 2005 group by country using DAX.
- Customer's segmentation based on Territory, deal size, total sales in last 2 year.

Additional Material (Security & Alerts)

- The company would like to share the Power BI report to stakeholders however these users can only see data where **'Country = United States'**.

- The company would like to receive a notification via e-mail whenever the **'Total amount in sales for selected period'** KPI exceeds a particular threshold.

Deliverables

1) Automated end to end process in Azure Synapse.
2) A master pipeline to extract the data from SQL DB.
3) ADLS Gen2 containers as a landing and processing zone for the data.
4) Spark Notebooks for curating the raw data.
5) Delta tables.
6) Views on top of curated data that can be fed into Power BI.
7) Power BI Dashboard