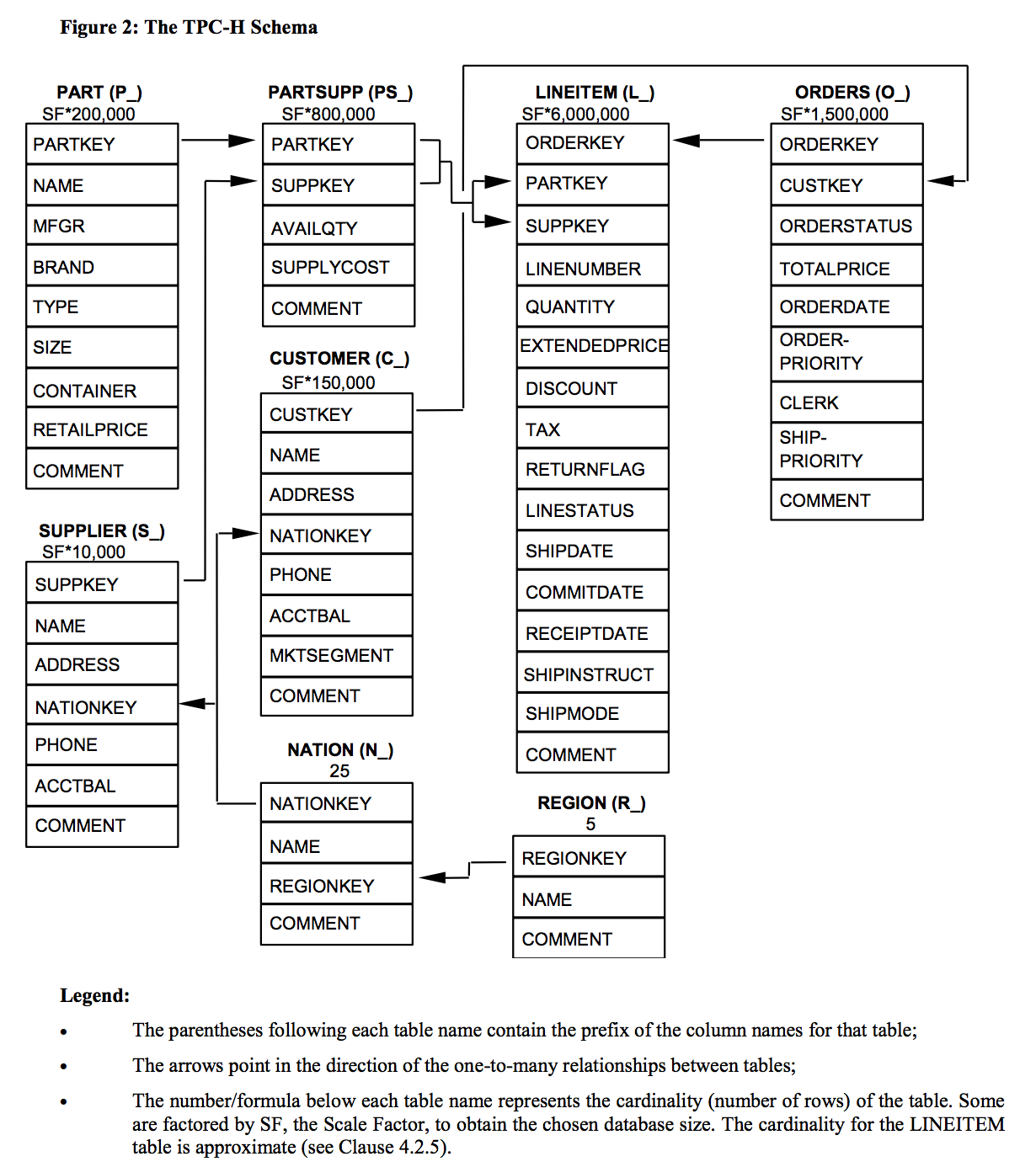
**Retail Case study**

****

| **Data sources:** Flat files , RDB , S3 ,JSON,parquet  Scheduler : Task  Streaming : Streams / Snowpipes  **Visualization :** PowerBI  **Language :** Javascript (Stored procedure) |
| --- |

**Problem Definition:**

The Pricing Summary Report Query provides a summary pricing report for all line items shipped as of a given date. The date is within 60-120 days of the greatest ship date contained in the database.The query lists totals for extended price, discounted extended price, discounted extended price plus tax, average quantity, average extended price, and average discount. These aggregates are grouped by RETURNFLAG and LINESTATUS, and listed in ascending order of RETURNFLAG and LINESTATUS. A count of the number of line items in each group is included



#### Datasets :

<https://github.com/akgeoinsys/retail>

**Tasks :**

1. **Create database retail\_db and schema schema**
2. **Create 8 retail tables based on the diagram and apply cluster key on most used column for querying.**
3. **Create internal stage and load customer ,lineitem ,nation and orders**
4. **Load part,partsupp,region and supplier csv files into AWS s3 bucket snow\_extstage**
5. **Load the data from external stage**
6. **Perform bad records filtering on loading ,file format and compression**
7. **Create snowpipes to continuously load orders data from external stage from s3 bucket to the order staging table ,maintain the change data capture and merge the data to the consumption table**
8. **Share your table to new non-snowflake user**
9. **create clone of the table with time travel before one day and write query to get history data using particular timestamp**
10. **Create stored procedure to insert the data into table after typecasting date format**
11. **Create a visualization for sales performance analysis using powerBI and prepare reports /dashboard for business queries**