# Task:

Data is stored in CSV files. The customer wants to have visualized analytical data based on this data set. He would like to store data in a local database and data warehouse in the cloud. The whole workflow should be built to maintain full load and incremental loads.

# Stack:

Python - Vlad

**MS SQL Server**/MariaDB/Postgres for 3rd NF - Pavlo

Talend - Pavlo, Petro

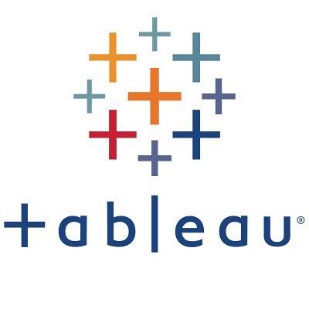
Cloud database (f.e. **AWS** RDS MySQL) for DWH schema - Vlad, MIla

Tableau/**Power BI**/Looker - Olia

## Workflow:

A picture containing text, clipart

Description automatically generatedIcon

Description automatically generated

# **Source:** NorthWind

CSV files will be provided as source files.

Schema for the first DB is provided as a DDL script.

DWH schema should be architecture by the team.

## Схема бази даних

NorthWind компанія займається перевезеннями продуктів. Тестова база MS SQL 2000

Створення бази даних - create\_database\_northwind.sql (необхідно поправити шляхи до фізичних файлів)

Створення  таблиць і заливка даних - northwind\_database.sql (скріпти генерувалися з MS SQL 2012)

## Список таблиць

* Orders  - замовлення на перевезення
* Products - продукти
* Order Details – деталі замовлень
* CustomerCustomerDemo (не використовується)
* CustomerDemographics (не використовується)
* Region – список регіонів
* Territories – список територій
* EmployeeTerritories (не використовується)
* Contacts – список контактів
* Employees – список працівників (хто оформляє замовлення)
* Categories – категорії продуктів
* Customers – замовники (хто замовляє продукти)
* Shippers – перевізники (хто перевозить продукти)
* Suppliers – поставщики продуктів (хто поставляє)

Diagram

Description automatically generated

# Репорти:

1. Отримати середню суму замовлення (загальна сума, вартість перевезення (dbo.orders. freight), кількість операцій) біжуче середнє за три місяці (січень – лютий; січень – лютий – березень; лютий – березень - квітень)) за день (вказаний місяць, діапазон даних) по всіх категоріях продуктів (вибраній категорії, списку категорій) в географічному розділі (звідки і куди перевозимо регіони, країни, штати), в розрізі перевізників, замовників, поставщиків і працівників, які оформляють угоду
2. Вивести найкращі (найгірші) 5 продуктів за кількістю операцій, сумою замовлень (добавити розділ категорій)
3. Вивести найкращі (найгірші) 5 замовників \ поставщиків \ працівників \ перевізників (одного найкращого \ найгіршого)  за кількістю операцій, сумою покупок  (добавити категорії продуктів)
4. Графік замовлень (суми, кількості) за перший тиждень кожного місяця
5. Тижневий звіт за замовленнями (з підсумками за місяць) в розрізі категорій продуктів (період – рік)
6. Медіанне значення замовлень (щомісячне) в розрізі категорій продукту і регіонів (походження і замовлення (звідки і куди перевозять))
7. Рейтинг замовлень по категоріях продуктів (ті, що краще замовляються – зверху)