

# BÀN VỀ HỆ THỐNG DỮ LIỆU & DWH TRONG XU THẾ CLOUD

Trình bày: Trần Xuân Lộc  
Hà Nội- 03/2024



- ☐ Đặc tính của Data Warehouse và mối tương quan trong hệ thống dữ liệu
- ☐ Data Warehouse trong xu thế Cloud
- ☐ Cần chuẩn bị nền tảng/chuyên môn gì?
- ☐ Ví dụ về một yêu cầu tổng hợp số liệu





# ĐẶC TÍNH CỦA DATA WAREHOUSE VÀ MỐI TƯƠNG QUAN TRONG HỆ THỐNG DỮ LIỆU





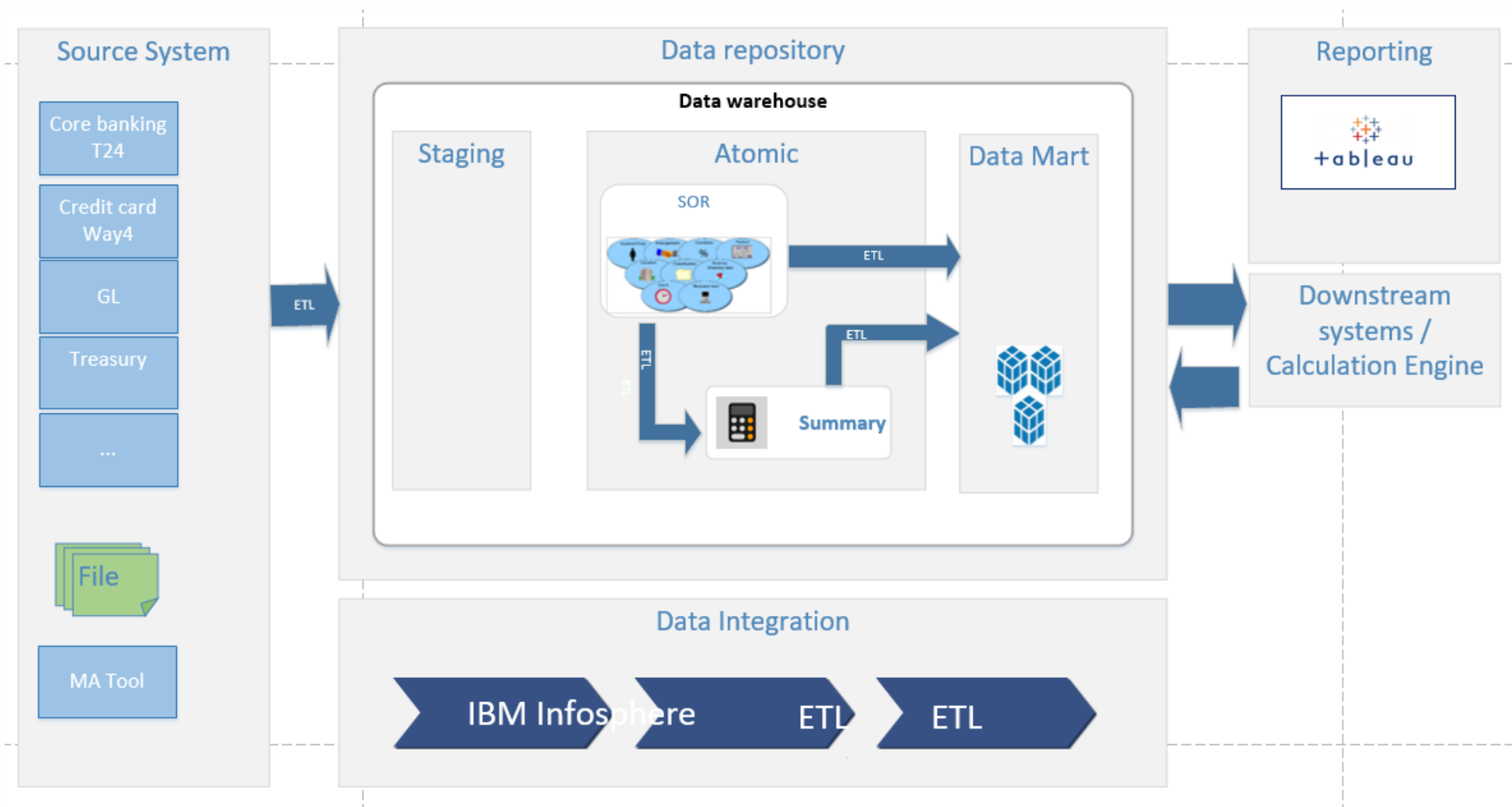
## Các đặc tính của Data Warehouse

- ✓ Subject Oriented: Dữ liệu hướng nghiệp vụ, tổ chức theo chủ đề
- ✓ Integrated: tích hợp dữ liệu từ nhiều nguồn, đóng vai trò là “***Single of Source Data Truth***”
- ✓ Time-Variant: lưu trữ dữ liệu theo lịch sử
- ✓ Non volatile: dữ liệu được lưu trữ, ko thay đổi và ko bị xóa
- ✓ Summarized: dữ liệu tổng hợp và phân loại phục vụ báo cáo /phân tích





# Mô hình kiến trúc cơ bản của hệ thống Data warehouse



## Các trường hợp khai thác số liệu từ DWH

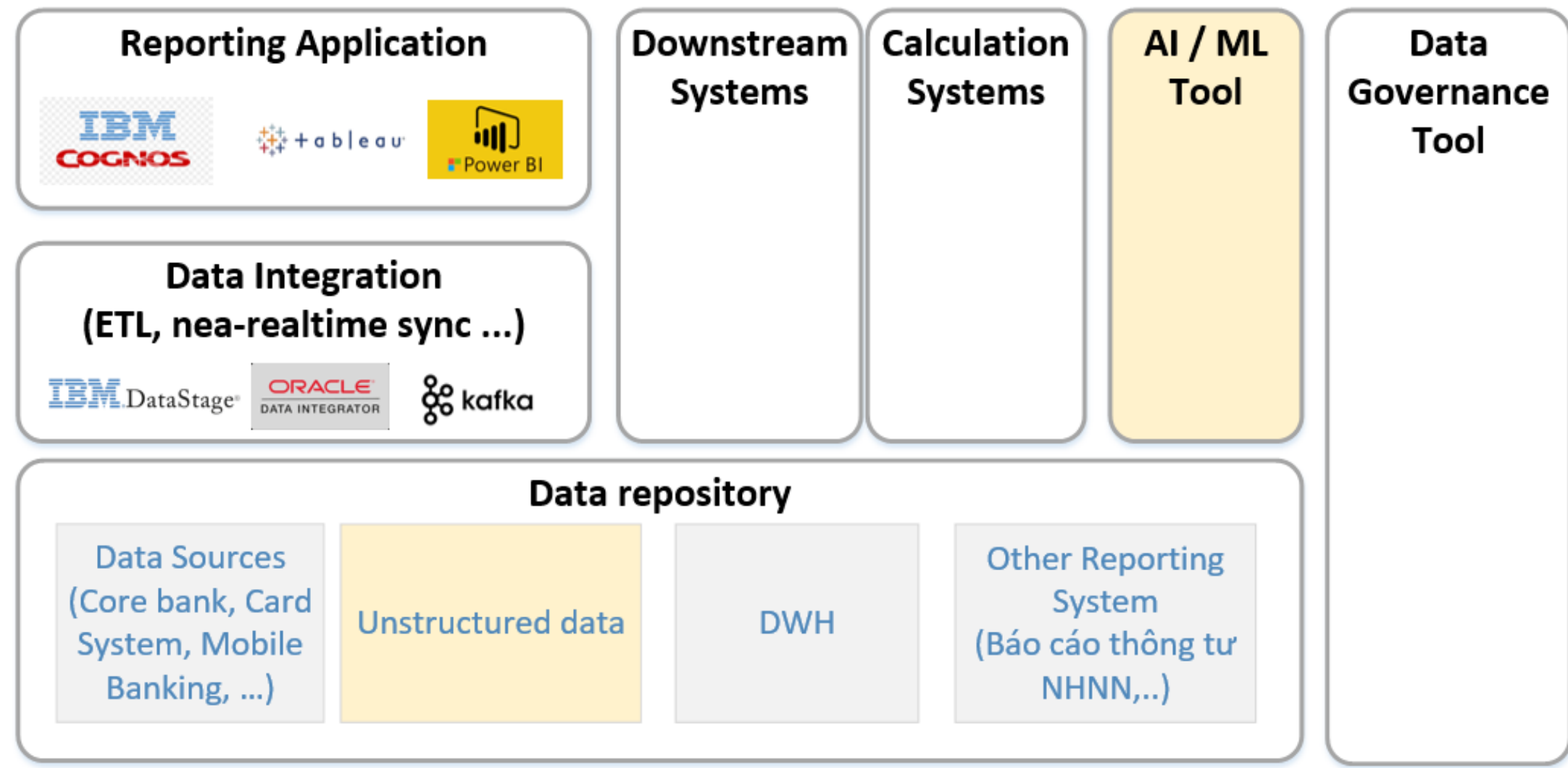


- ☐ Reporting: xây dựng báo cáo tuần thủ, báo cáo hoạt động hàng ngày/định kỳ. Các dạng báo cáo bao gồm:
  - ✓ Dạng sao kê chi tiết
  - ✓ Dạng bảng tổng hợp, so sánh kỳ
  - ✓ Dạng chỉ tiêu
- ☐ Visualization report / Dashboard: xây dựng các báo cáo biểu đồ KPI,...
- ☐ Cung cấp cho các hệ thống Downstream / Calculation Engines/ Hệ thống reporting khác
- ☐ Cung cấp số liệu cho các đơn vị phân tích số liệu(Analysis)





# Một số thành phần trong hệ thống dữ liệu





# DATA WAREHOUSE TRONG XU THẾ CLOUD





## Traditional approaches: Data Warehouse vs Data Lake

High performance  
Governed  
BI



Expensive  
Non-scalable  
Vendor lock-in

### 101: Data Warehouse

*How can we ask enterprisewide questions requiring historical data?*

- A data warehouse gathers structured data from multiple sources into a central repository
- Supports multiple data analytics and business intelligence applications such as enterprise reporting, to answer questions about your business
- Normalized and structured data made it easy to analyze, but was an **expensive choice**

Unstructured data  
AI/ML  
Low-cost storage



Maintenance  
High-complex  
Limited Performance

### 101: Data Lake

*How can we discover what we don't know?*

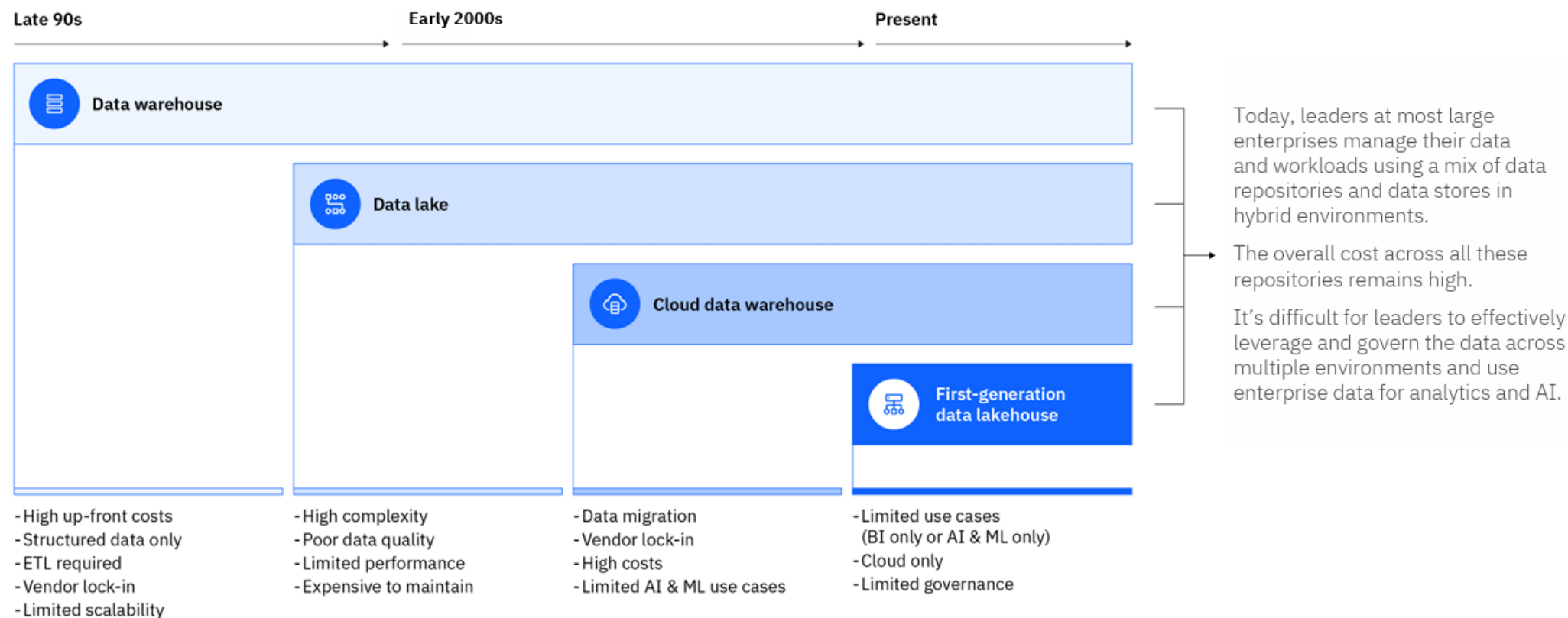
- As volume, velocity and variety of data grew, data lakes were designed for data discovery and data science/machine learning use cases.
- Commonly built on large data platforms such as Hadoop (HDFS)
- Data is stored in RAW and unstructured format = lower cost for large volumes of data
- Highly flexible and scalable
- Difficult to use and govern, and complex to maintain, required data scientist, now many are **data swamps**

Source: IBM





Traditional approaches to addressing these challenges have created more overall complexity and cost, which has led to the emergence of data lakehouse architectures



Source: IBM





# CẦN CHUẨN BỊ NỀN TẢNG KỸ THUẬT / CHUYÊN MÔN GÌ?



## Đối với DWH On-Premise



### ☐ Đối với Dev

- ✓ Kỹ năng SQL
- ✓ Ít nhất một công cụ ETL
- ✓ Ít nhất một công cụ BI

### ☐ Chung

- ✓ Tư duy về mối liên quan giữa các dữ liệu với nhau

### ☐ Đối với Tester / BA

- ✓ Kỹ năng SQL
- ✓ Kỹ năng test dữ liệu
- ✓ Hiểu nghiệp vụ





Vậy mọi thứ có thay đổi hoàn toàn khi  
chuyển lên Cloud???





Bản chất không thay đổi, chỉ thay đổi về  
kiến trúc nền tảng!



# Mọi thứ có thay đổi hoàn toàn khi chuyển lên Cloud?



- ❑ Một số thành phần không thay đổi / kế thừa
  - ✓ Nhu cầu đầu ra của nghiệp vụ
  - ✓ Cơ cấu tổ chức dữ liệu theo góc độ nghiệp vụ
  - ✓ Logic xử lý dữ liệu
- ❑ Một số thành phần thay đổi
  - ✓ Kiến trúc nền tảng như: công cụ ETL, cách thức lưu trữ,...





## ☐ Đối với Dev

- ✓ Kỹ năng SQL
- ✓ Stack công nghệ mới tương ứng
- ✓ Tuy nhiên sau khi đã có Framework làm việc, phần phức tạp nhất vẫn là tổ chức / xử lý dữ liệu.
- ✓ Ít nhất một công cụ BI

## ☐ Chung

- ✓ Tư duy về mối liên quan giữa các dữ liệu với nhau

## ☐ Đối với Tester / BA

- ✓ Kỹ năng SQL
- ✓ Kỹ năng test dữ liệu
- ✓ Hiểu nghiệp vụ

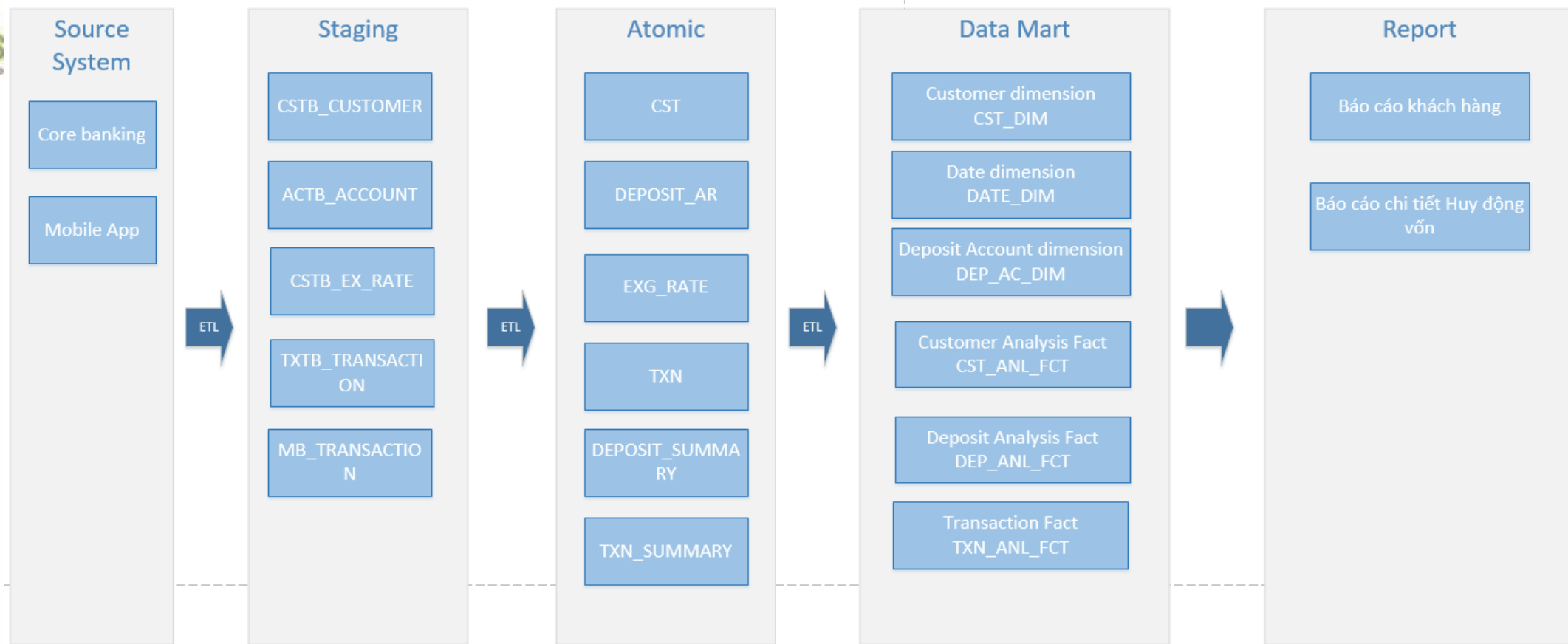






# VÍ DỤ VỀ MỘT YÊU CẦU TỔNG HỢP SỐ LIỆU





- ✓ Quan hệ giữa các bảng như nào?
- ✓ Logic mapping tổng hợp ra các trường trên từng tầng như thế nào?
- ✓ Logic trên báo cáo, quan hệ giữa các chỉ tiêu như nào?
- ✓ Quan hệ giữa các measure/ chỉ tiêu báo cáo như thế nào?



Microsoft Excel  
Worksheet



# Q&A





**XIN TRÂN TRỌNG CẢM ƠN!**

---

