

XÂY DỰNG GIẢI PHÁP BUSINESS INTELLIGENCE TRÊN NỀN TẢNG ĐÁM MÂY MICROSOFT AZURE KẾT HỢP ELT ĐỘNG



Công ty: TNHH Bảo Hiểm Nhân Thọ Prudential VN

GVHD: Ths. Trương Hoài Phan

Hướng dẫn: Ths. Lê Bá Thiên



THÀNH VIÊN NHÓM



NHẬT NGUYỄN

Nhóm trưởng



ĐẶC SANG

Thành viên



NỘI DUNG CHÍNH

01

GIỚI THIỆU

Tổng quát vấn đề và công ty

02

CƠ SỞ LÝ THUYẾT

Khái niệm, lý thuyết

03

PHÂN TÍCH VẤN ĐỀ

Phân tích và đề xuất mô hình

04

THỰC NGHIỆM

Xây dựng data lake và data warehouse

05

TRỰC QUAN HÓA

Phân tích dữ liệu

06

KẾT LUẬN

Kết quả, hạn chế và phương hướng phát triển



01

GIỚI THIỆU

TỔNG QUAN CÔNG TY



“LẮNG NGHE. THẤU HIỂU. HÀNH ĐỘNG.”

TẦM NHÌN

Là lựa chọn số 1 về bảo hiểm nhân thọ cho mọi gia đình Việt.

MỤC TIÊU

Giúp mọi người đạt được những điều tốt đẹp trong cuộc sống.



VẤN ĐỀ

- Nhập và lưu trữ dữ liệu từ nhiều nguồn khác nhau
- Dữ liệu lớn và được định dạng nhiều loại tệp
- Đảm bảo và bảo mật dữ liệu

MỤC TIÊU

- Thiết lập quy trình ELT động
- Xây dựng giải pháp BI
- Triển khai trên nền tảng Cloud Azure





02

CƠ SỞ LÝ THUYẾT





CƠ SỞ LÝ THUYẾT



Business Intelligence

Kỹ năng, quy trình, công nghệ và ứng dụng để hỗ trợ ra quyết định



Quy trình ELT và ETL

Quy trình trích xuất, chuyển đổi và tải dữ liệu



Hồ dữ liệu và Kho dữ liệu

Nơi lưu trữ dữ liệu



Microsoft Azure

Hệ thống tổng hợp các dịch vụ điện toán đám mây



Microsoft Power Bi

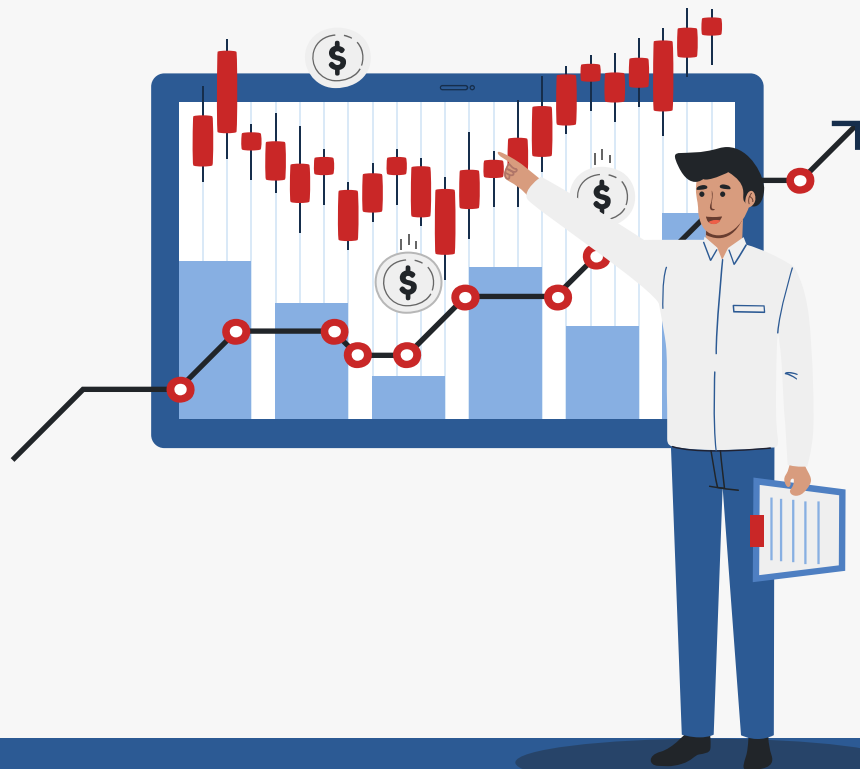
Công cụ trực quan hóa dữ liệu





03

PHÂN TÍCH VẤN ĐỀ





NGUỒN DỮ LIỆU



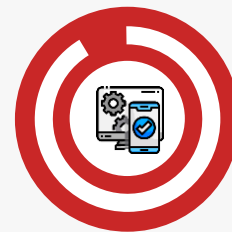
Databases

Ghi nhận hoạt động
bán hàng



Accounting Systems

Ghi nhận thông tin
thanh toán của khách hàng



Web Services

Bình luận khách hàng

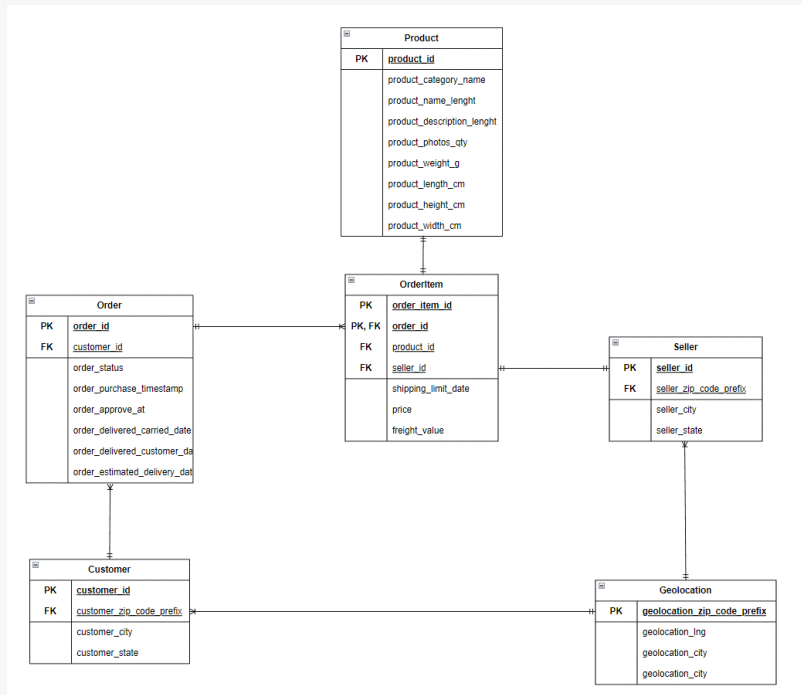


NGUỒN DỮ LIỆU

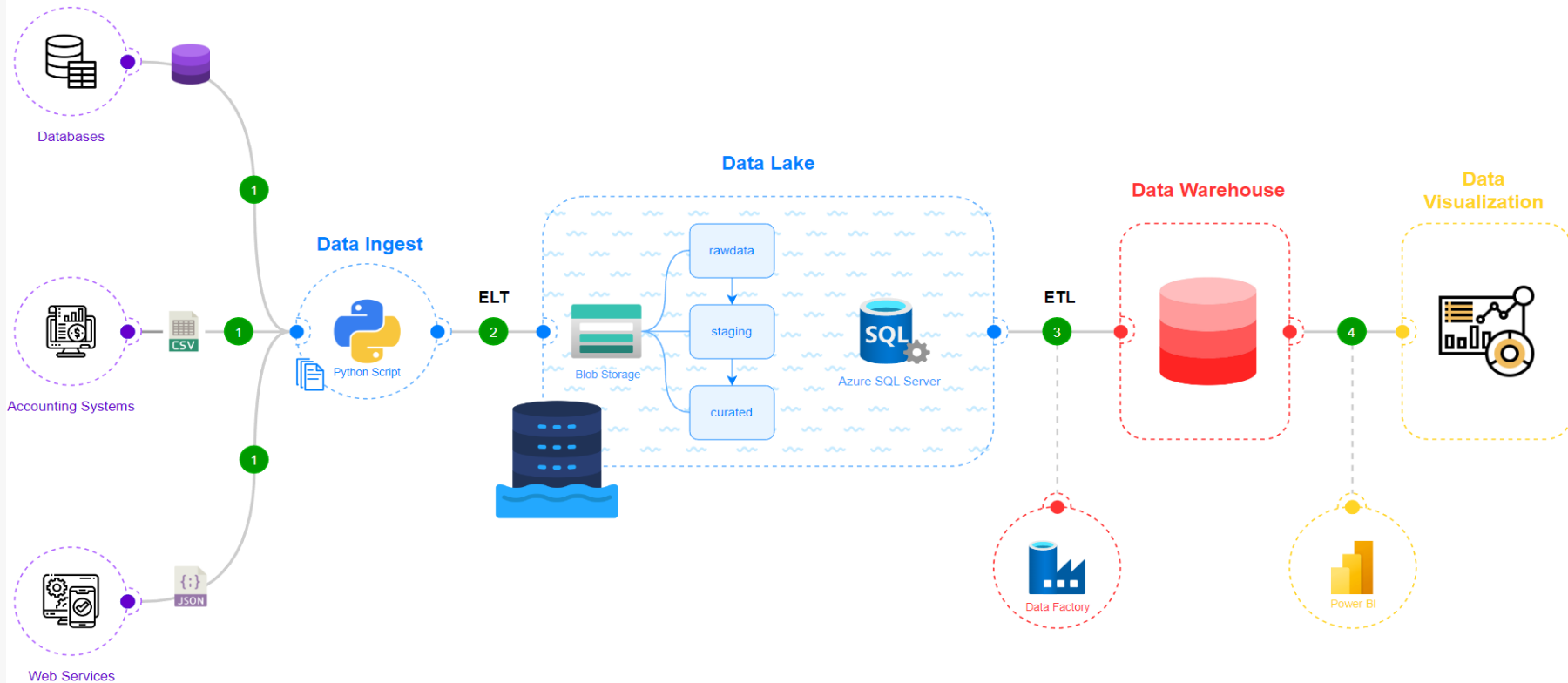


Databases

Mô hình ERD



Data Sources



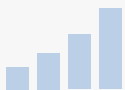
ĐỀ XUẤT MÔ HÌNH



04

THỰC NGHIỆM





VÙNG DỮ LIỆU

RAWDATA

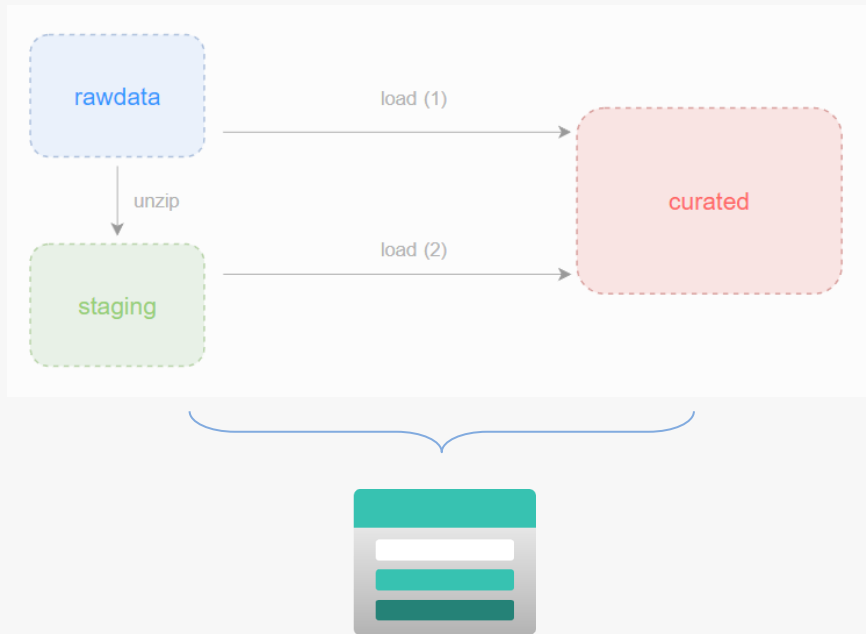
Lưu trữ dữ liệu thô được tập hợp từ nhiều nguồn

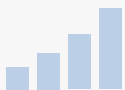
STAGING

Giải nén tất cả các tệp được nén thành tệp zip

CURATED

Tổ chức lưu trữ dữ liệu theo dạng tệp

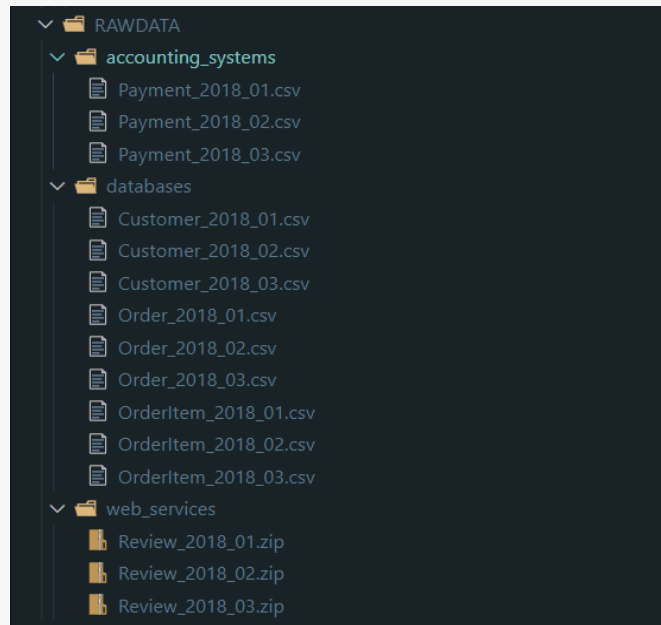


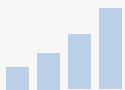


VÙNG DỮ LIỆU

RAWDATA

Lưu trữ dữ liệu thô được tập hợp từ nhiều nguồn



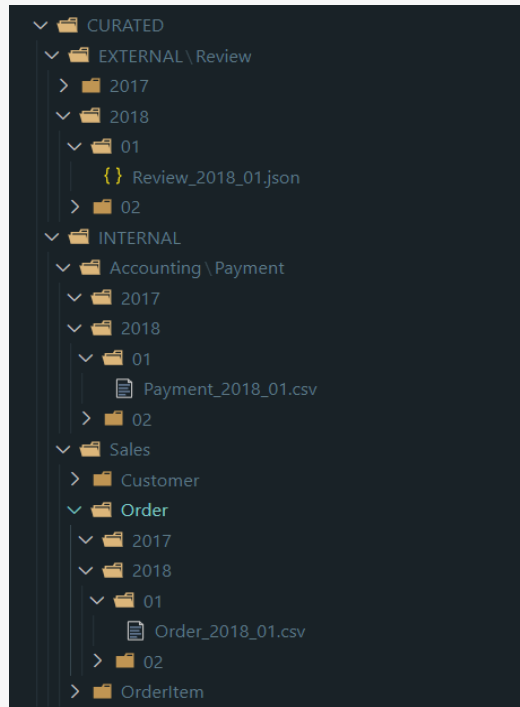


VÙNG DỮ LIỆU

CURATED



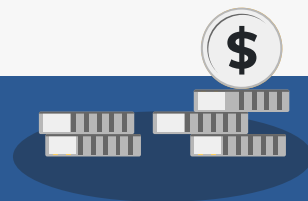
Tổ chức lưu trữ dữ liệu theo
dạng tệp

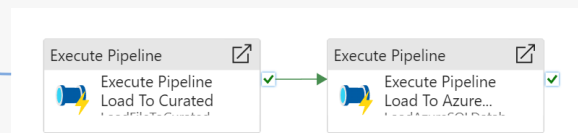
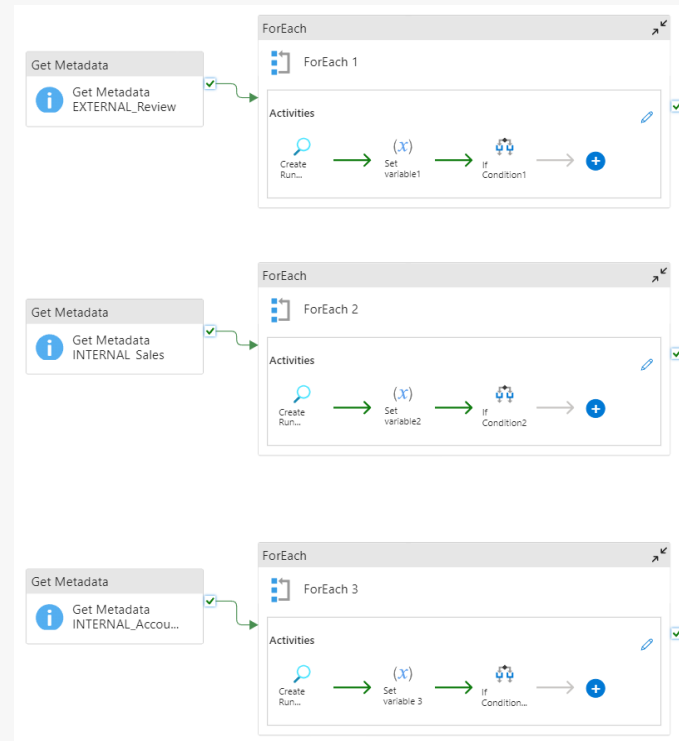
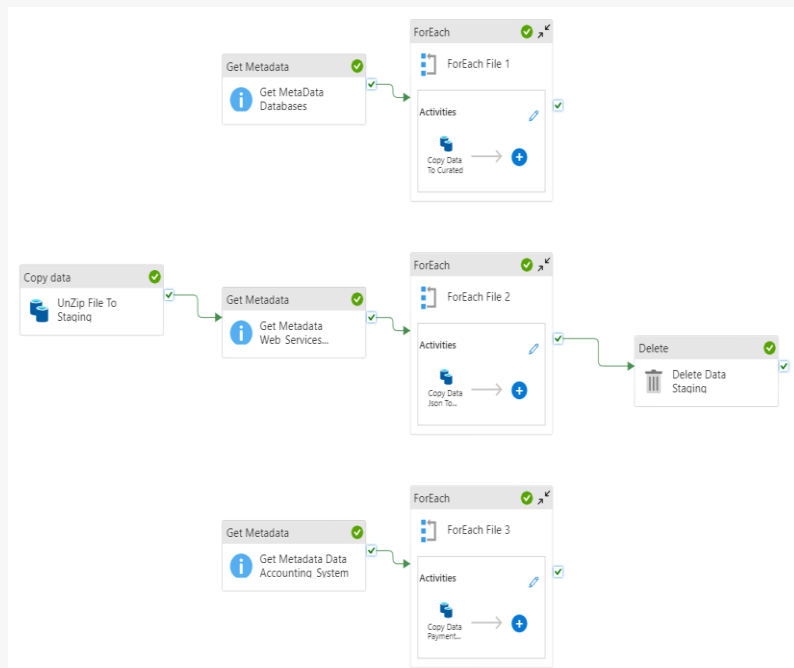




- Lưu trữ dữ liệu cần cho việc phân tích kinh doanh
- Giảm thiểu thời gian truy vấn từ vùng dữ liệu curated
- Ghi nhận quá trình hoạt động của quy trình ELT động

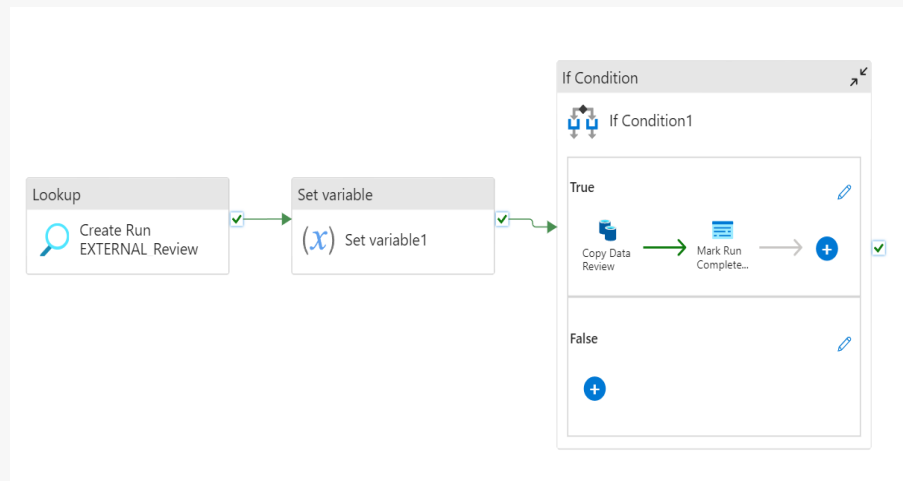
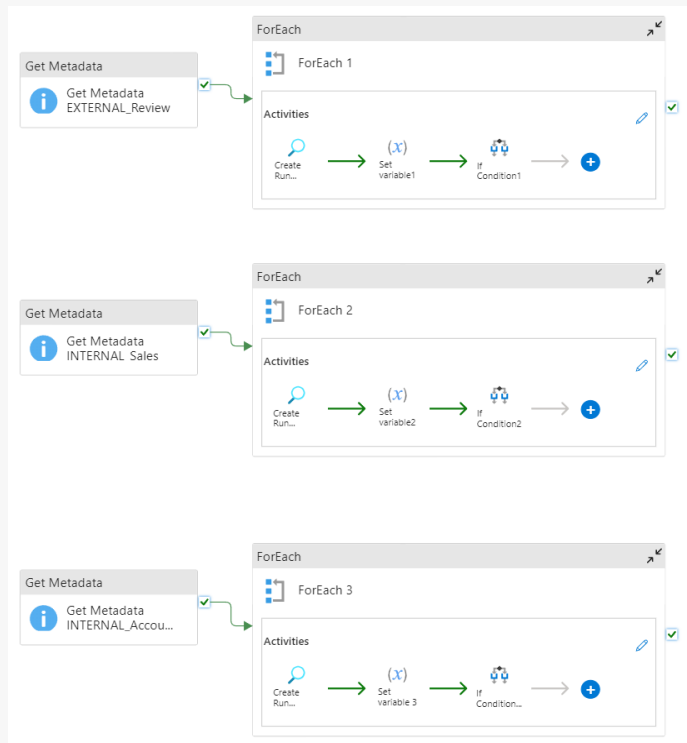
-AZURE SQL SERVER-





QUY TRÌNH ELT ĐỘNG





QUY TRÌNH ELT ĐỘNG





Bus Matrix

	Customer	Seller	Product	Time
Sales performance tracking	X	X	X	X
Orders	X	X	X	X
Customer segmentations	X			X
Seller distribution		X		

Master Data

Customer	Mô tả thông tin về thành phố, tiểu bang, id,... của khách hàng.
Seller	Bao gồm thông tin về cửa hàng, địa chỉ cửa hàng
Product	Thông tin về sản phẩm của công ty như số ký tự trích xuất từ mô tả sản phẩm, chiều dài, trọng lượng, chiều rộng của sản phẩm, ...

Transaction Data

Dữ liệu giao dịch đặt hàng	Mô tả thông tin về các mặt hàng được mua trong mỗi đơn hàng, giá sản phẩm, số lượng, ngày hạn vận chuyển, giá trị cước, tình trạng...
----------------------------	---



XÂY DỰNG KHO DỮ LIỆU



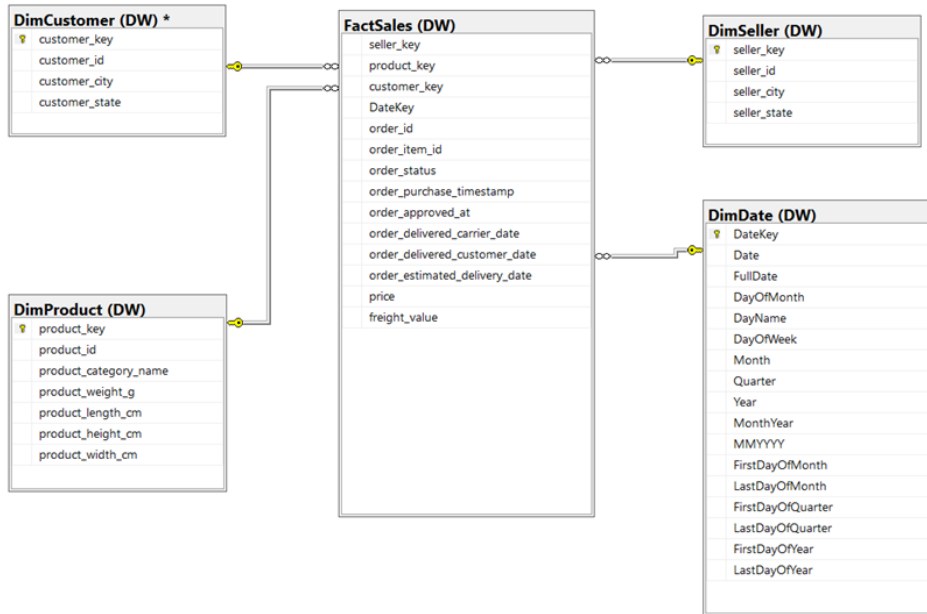
ETL mapping

FactSales							
seller_key	int		DimSeller	seller_key	int		Truy xuất từ [DimSeller].[seller_key]
product_key	int		DimProduct	product_key	int		Truy xuất từ [DimProduct].[product_key]
customer_key	int		DimCustomer	customer_key	int		Truy xuất từ [DimCustomer].[customer_key]
DateKey	int		DimDate	DateKey	int		Truy xuất từ [DimDate].[DateKey]
order_id	varchar(50)		quantad_orders	order_id	varchar(50)		Từ nguồn
order_item_id	varchar(50)		quantad_order_items	order_item_id	varchar(50)		Từ nguồn
order_status	varchar(32)	x	quantad_orders	order_status	varchar(32)	x	Từ nguồn
order_purchase_timestamp	date	x	quantad_orders	order_purchase_timestamp	Datetime	x	Từ nguồn
order_approved_at	date	x	quantad_orders	order_approved_at	Datetime	x	Từ nguồn
order_delivered_carrier_date	date	x	quantad_orders	order_delivered_carrier_date	Datetime	x	Từ nguồn
order_delivered_customer_date	date	x	quantad_orders	order_delivered_customer_date	Datetime	x	Từ nguồn



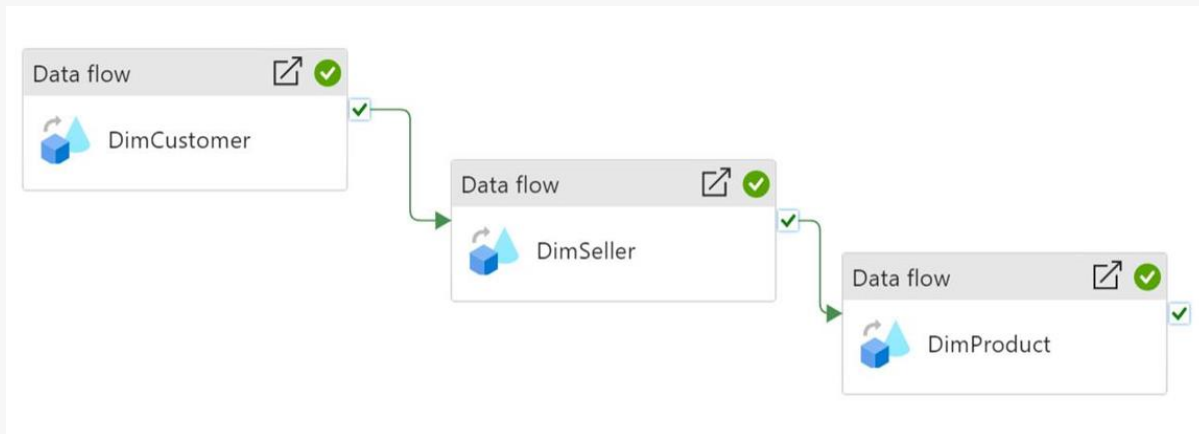


Mô hình nhà kho dữ liệu (Star Schema)





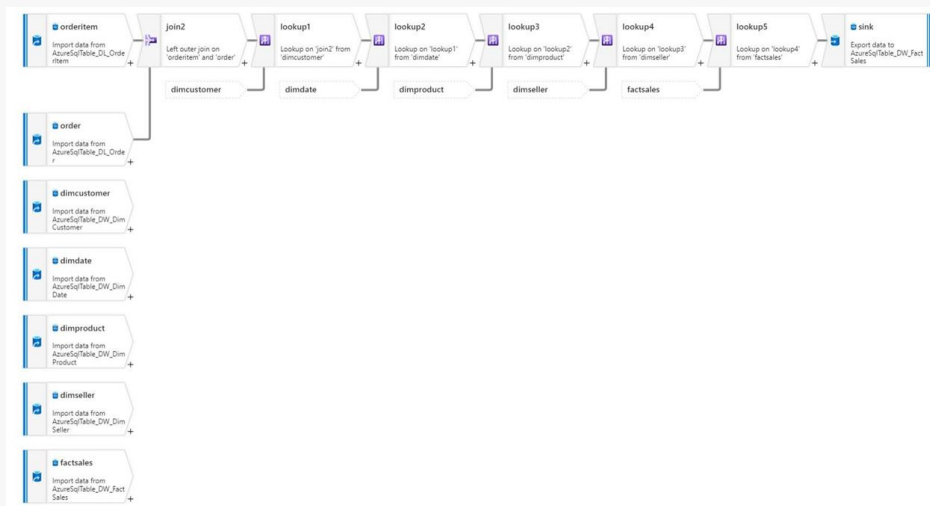
Quy trình ETL



Load Dim



Quy trình ETL



Load Fact

05

PHÂN TÍCH VÀ TRỰC QUAN HÓA



Sales Dashboard

Sale Summary, overview of Product and Customer

Year

2017

Month

All

Revenue

986.87K



No. Orders

44.58K



No. Sold Products

50.86K



No. Sellers

7.43K

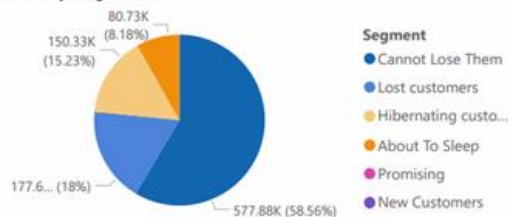


% Delivered

98.03%



Sales by Segments



Segment

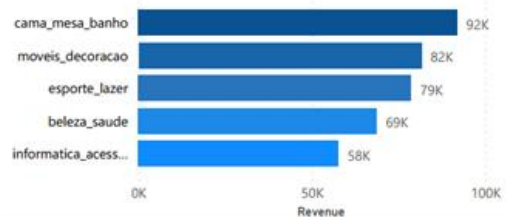
Avg R

Avg F

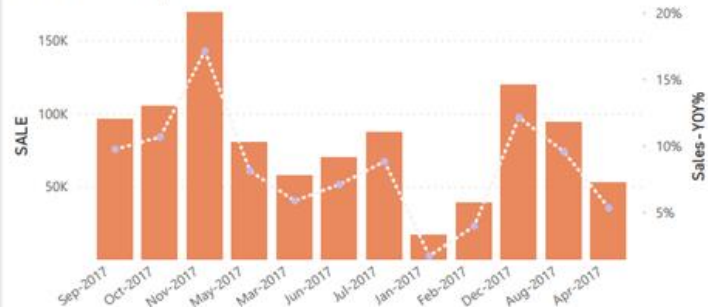
Avg M

About To Sleep	231.95	1.00	17.37
Cannot Lose Them	401.38	1.00	36.27
Hibernating customers	284.51	1.00	13.47

Top 5 Sales By Category



SALES AND SALES-YOY% By month



Revenue allocation by city





06

KẾT LUẬN

KẾT LUẬN

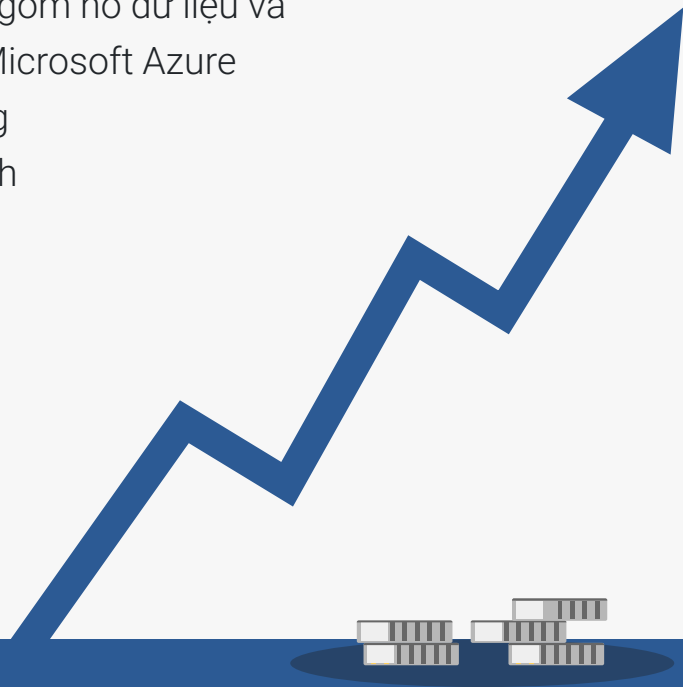
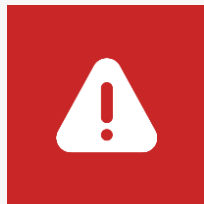


■ Kết quả

- Xây dựng giải pháp BI bao gồm hồ dữ liệu và kho dữ liệu trên nền tảng Microsoft Azure
- Kết hợp quá trình ELT động
- Xây dựng báo cáo phân tích

■ Hạn chế

- Thiếu nguồn dữ liệu chuẩn chỉnh
- Giới hạn chi phí trong quá trình triển khai giải pháp trên nền tảng Cloud Azure
- Đòi hỏi nhiều thời gian thực hiện



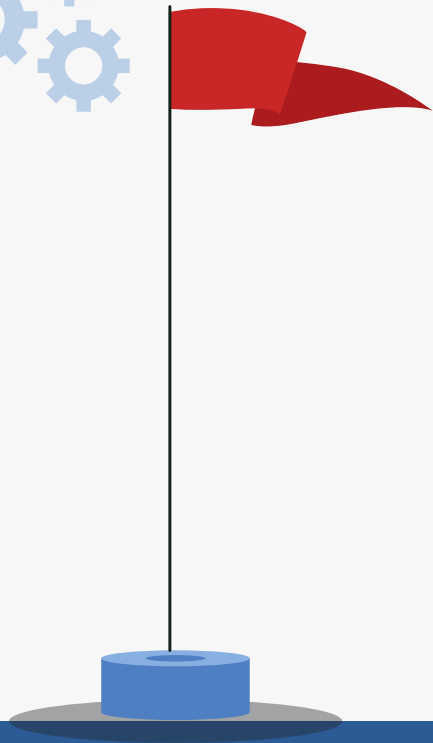
KẾT LUẬN



■ Hướng phát triển

- Xây dựng quy trình tự động hóa
- Xây dựng hệ thống phân tích dự báo
- Xây dựng hệ thống phân tích trải nghiệm khách hàng bằng phương pháp học máy





THANK YOU

