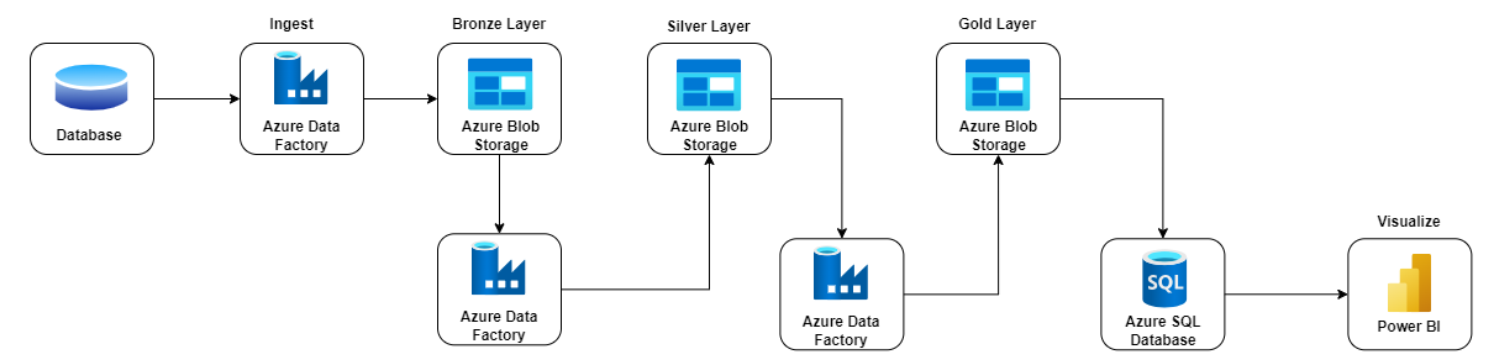
**BÁO CÁO**

## Data flow

Ở dataflow này thì thay vì sử dụng Azure Synapse Analytics tốn nhiều chi phí thì mô phỏng bằng Azure SQL Database.



## Đầu vào, đầu ra của quá trình xử lý là gì?

* Đầu vào của project là một Database ở SQL local.
* Đầu ra của project là một Data Warehouse dưới dạng star schema.

## Tại sao lại không sử dụng SQL local mà phải đưa lên cloud?

### Ưu nhược điểm của SQL local

* Ưu điểm:
* Không cần phải trả phí thuê server hoặc dịch vụ lưu trữ đám mây.
* Có toàn quyền kiểm soát và quản lý cơ sở dữ liệu.
* Hiệu suất cao cho môi trường cục bộ.
* Dễ dàng thiết lập và sử dụng trên máy phát triển cá nhân.
* Nhược điểm:
* Khả năng mở rộng hạn chế.
* Dễ bị mất dữ liệu nếu máy tính bị hỏng hoặc bị mất.
* Yêu cầu người dùng tự quản lý bảo mật và sao lưu dữ liệu.
* Khó truy cập từ các thiết bị ngoài mạng cục bộ.
* Khó khăn khi chia sẻ dữ liệu với người dùng khác hoặc các ứng dụng khác.
* Hạn chế trong việc tích hợp với các dịch vụ đám mây.

### Ưu nhược điểm khi đưa lên cloud

* Ưu điểm:
* Tự động cập nhật và bảo trì.
* Khả năng mở rộng linh hoạt.
* Tối ưu hóa hiệu suất với các công nghệ mới.
* Mã hóa dữ liệu, kiểm soát truy cập, bảo mật cao.
* Sao lưu tự động và phục hồi dữ liệu.
* Dễ dàng kết nối và tích hợp với các dịch vụ khác của Azure.
* Nhược điểm:
* Chi phí cao, theo mức sử dụng.
* Phụ thuộc vào Internet.
* Cần cấu hình và quản lý bảo mật phù hợp.
* Giới hạn về cấu hình phần cứng.
* Nếu doanh nghiệp muốn tối ưu hóa chi phí, tăng cường bảo mật, và đảm bảo hiệu suất hoạt động cao thì sử dụng cơ sở dữ liệu trên đám mây mang lại nhiều lợi ích về khả năng mở rộng, quản lý, bảo mật, và tính sẵn sàng so với cơ sở dữ liệu cục bộ.

## Tại sao lại qua nhiều lớp Azure Blob Storage?

* Từng lớp lưu trữ các loại dữ liệu khác nhau. Ví dụ như lớp **Bronze layer** lưu trữ dữ liệu thô ban đầu, **Silver layer** lưu trữ dữ liệu đã được làm sạch và biến đổi ban đầu, **Gold layer** lưu trữ dữ liệu cuối cùng đã được tinh chỉnh và tối ưu hóa cho mục đích báo cáo và phân tích.
* Giảm thiểu rủi ro mất dữ liệu.
* Dễ dàng theo dõi và kiểm soát chất lượng dữ liệu.

## Sau khi xử lý là OLTP hay OLAP?

Bước Azure SQL Database nên sử dụng Azure Synapse Analytics nhưng tốn nhiều chi phí nên sử dụng Azure SQL Database để mô phỏng một Data Warehouse với mô hình star schema, do đó sau khi qua quá trình xử lý thì nó là một hệ thống OLAP.