



- Ingestion

- Engine được lựa chọn cho quá trình data ingestion phải thỏa mãn các đặc điểm sau:
  - ◆ Engine phải hỗ trợ cả hai loại databases MySQL và PostgreSQL
  - ◆ Data được ingest về data lake phải giữ được các thuộc tính và data type của data sources
  - ◆ Khả năng xử lý data theo lô (batch processing)
- Data ingest từ nhiều nguồn nên tốn nhiều tài nguyên lưu trữ vì thế nên data lưu trong DataLake cần được lưu ở dạng được nén như file parquet.

- ETL

- Engine sử dụng trong quá trình ELT phải đủ đơn giản như hỗ trợ sql base để các team có thể dễ dàng tiếp cận và sử dụng.

- Provide data

- Data trong Data Warehouse được đưa vào các data mart dành riêng để cung cấp cho các người dùng bên ngoài như các công ty.
- Các team trong công ty có thể sử dụng data trực tiếp từ Data Warehouse để visualizing.
- Tool dùng để exploring và visualizing phải hỗ trợ cho OLAP