CENTRALIZE DATA WAREHOUSE

**Adventure Works Cycles** là một công ty chuyên kinh doanh xe đạp. Hoạt động của công ty bao gồm nhập xe từ các đối tác cung cấp xe đạp khắp thế giới, sau đó phân phối về những cửa hàng của công ty. CSDL của công ty là một OLTP database. Do có mong muốn trích xuất báo cáo vào những khoảng thời gian nhất định mà không gây sức ép đến CSDL, các nhân viên IT đề xuất công ty nên xây dựng một Data Warehouse riêng nhằm phục vụ mục đích báo cáo. Công ty đang có nhu cầu thử nghiệm giải pháp Data Warehouse, vì thế database cần phải có tối thiểu 3 fact tables và **1 dim date**.

Data Warehouse mà IT team hướng đến gồm có các tiêu chuẩn sau:

Landing:

* ELT pipeline phải được run vào non-working hours.
* Data cũ phải được archive mỗi lần ELT.

Staging:

* Staging tables phải chứa thông tin của ELT runs (PipelineRunDate, PipelineRunId, …).
* Staging tables phải thể hiện đúng tình trạng data của source, tại thời điểm ELT.

Dim, Fact:

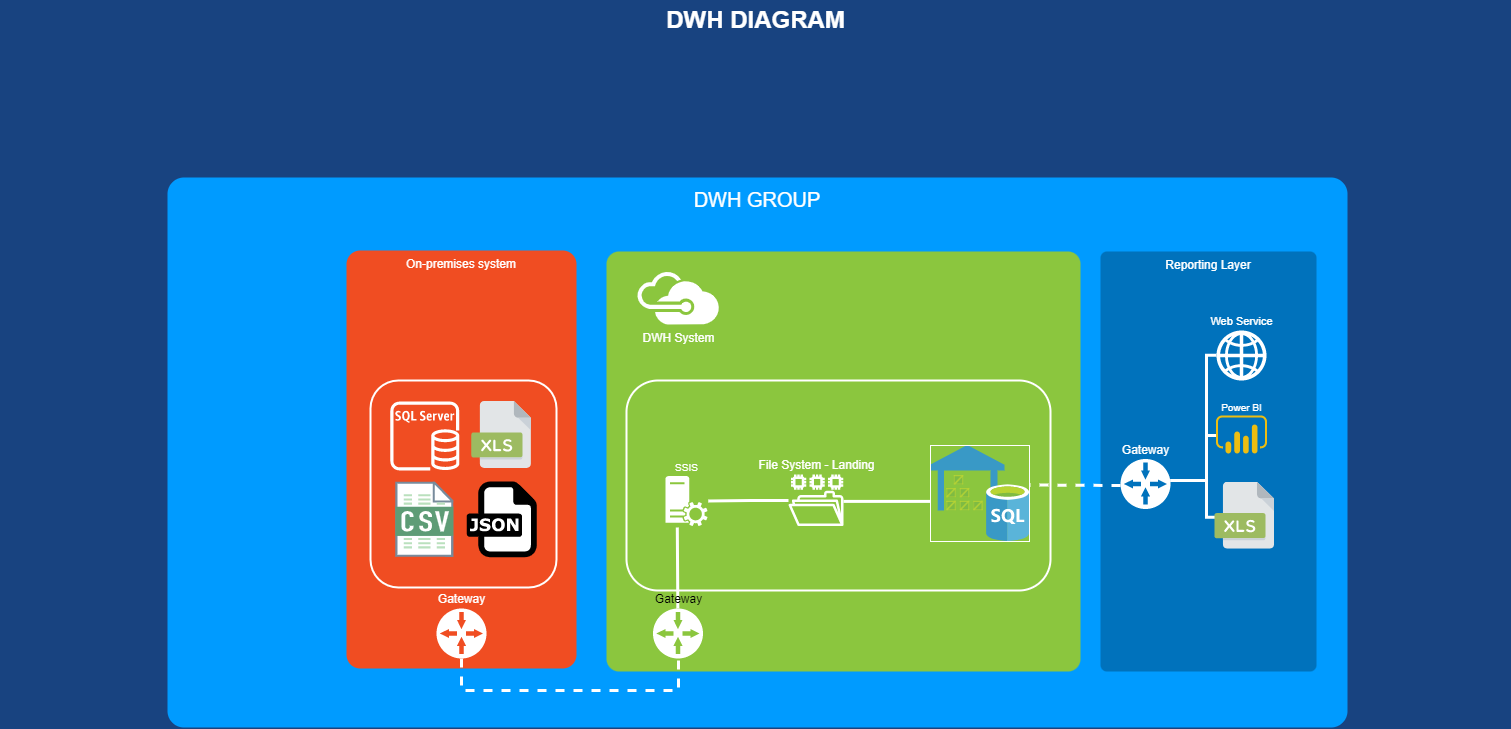
* Database phải ở dạng Star Schema.
* Data của bảng fact phải được incremental load. Data bảng dim có thể cân nhắc giữa full load và incremental load.
* Các bảng dim, fact phải chứa thông tin của ELT runs (PipelineRunDate, …).

Note:

* **ELT Jobs info** & transformation code phải được lưu ở target database / server. SSIS package không được chứa transform code.
* ELT job phải được control thông qua **config table**. Không lưu config trên SSIS package.
* Các ELT layer (landing, staging, dim, fact) phải được đặt trong những package riêng biệt.
* Các object trên SSIS phải đúng **naming convention** ([link](https://sqlkover.com/ssis-naming-conventions-2-0/)).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data Source | Server | Data | Login | Password |
| SQL Database | 192.168.68.27 | BicycleRetailer | Tự đặt | Tự đặt |
| SQL Database | 192.168.68.27 | HRDB | Tự đặt | Tự đặt |
| Csv | Tự bỏ vào 1 nơi nào đó và ko được đụng vào | TransactionHistory | - | - |
| Excel | Tự bỏ vào 1 nơi nào đó và ko được đụng vào | CountryOfBusinessEntity | - | - |
| Json | Tự bỏ vào 1 nơi nào đó và ko được đụng vào | Person | - | - |

High-level Diagram:



Optional Question

+ Tư vấn tất cả các kĩ thuật có thể sử dụng trong tất cả các giai đoạn để pipeline tổng hoàn thành với tốc độ nhanh nhất có thể.

+ User được quyền chỉnh sửa data trong khoảng thời gian lùi 1 năm kể từ ngày đang làm việc, hãy nêu ảnh hưởng của việc chỉnh sửa dữ liệu của những column đang được sử dụng ở dimension lên hệ thống DWH.

+ Business mong muốn được tư vấn một vài report cơ bản xoay quanh việc bán hàng, nhập hàng, tồn kho và sản xuất. Hãy xây dựng, cung cấp và diễn giải ý nghĩa của báo cáo.

**TRÌNH BÀY BÁO CÁO:**

1. **ETL Config**
   1. ETL Control Config
      * Config table
      * Audit table
   2. ETL Pipeline
      * Diagram
      * Pipeline detail
2. **DWH Design**
   1. Landing Layer
      * Tables
      * Stored Procedures
      * Function
      * View
   2. Staging Layer
      * Tables
      * Stored Procedures
      * Function
      * View
   3. Dimension Tables
      * Tables
      * Stored Procedures
      * Function
      * View
   4. Fact Tables
      * Tables
      * Stored Procedures
      * Function
      * View
3. **Optional Question & Answer**