Exercise: Designing a Data Management Strategy

Scenario:

You are a data architect for a retail company that operates online and offline stores. Your company collects data from various sources, including:

1. **Point-of-Sale Systems (POS)** (transactional data).
2. **Customer Feedback** from surveys and social media.
3. **Web Analytics** from the e-commerce platform.
4. **Supply Chain Data** from logistics and inventory systems.

Your task is to design a data management strategy by answering the following:

1. **Data Storage**
   - Which type of storage (relational database, data lake, or data warehouse) is suitable for each data source? Justify your answer.
2. **Data Integration**
   - Describe how you would integrate data from these sources using ETL/ELT processes. Include a brief flow diagram.

---

1. Data Storage Strategy

Data Source	Storage Type	Description
Point-of-Sale Systems (POS)	Relational Database & Data Warehouse (My SQL, PostgreSql, SQL Server)	ข้อมูลสำหรับงาน POS จำเป็นต้องเก็บในรูปแบบของ Relational Data เพื่อเชื่อมโยงข้อมูลได้ถูกต้อง ทราบที่มาที่ไปของข้อมูล โดยใช้ Database ตัวอย่างเช่น MySQL, PostgreSQL หรือ SQL Server และจะต้องตรวจสอบความสัมพันธ์ระหว่างตาราง และ มีการอัพเดทอยู่เสมอทั้งเรื่องของข้อมูลและสถานะ
Customer Feedback	Data Lake (AWS S3) (Google cloud Storage, Azure Data Lake)	ข้อมูลความเห็นจากลูกค้า มีทั้งแบบ Stauctre Data และ Unstactuer Data(รูป เสียง หรือ วีดีโอ)

Data Source	Storage Type	Description
Web Analytics	Data Lake (AWS S3)	ข้อมูลพฤติกรรมของผู้ใช้ Web Analytics เน้นการวิเคราะห์พฤติกรรมแบบเรียลไทม์ เช่น จำนวนผู้เข้าชม ขณะที่ Data Warehouse ช่วยให้สามารถทำการวิเคราะห์เชิงลึก เช่น ข้อมูลการซื้อขาย หรือการทำธุรกรรมบนเว็บไซต์
Supply Chain Data	Relational Database & Data Warehouse	ข้อมูลสินค้าคงคลังในการดำเนินงานนั้นเหมาะสมกับฐานข้อมูลเชิงสัมพันธ์ (OLTP) ข้อมูลรวมสำหรับการวิเคราะห์แนวโน้มควรย้ายไปยัง Data warehouse เพื่อการนำไปทำรายงานการวิเคราะห์ข้อมูลด้วย BI

2. Data Integration Strategy

Describe how you would integrate data from these sources using ETL/ELT processes. Include a brief flow diagram.

ETL Process:

1. Extract:
    - Data Sources (e.g., MySQL, REST APIs, CSV files)

      Connect and extract data
2. Transform:
    - Data Transformation Layer

      Data cleaning and normalization

      Aggregation and business logic
3. Load:
    - Data Warehouse (e.g., Amazon Redshift, Snowflake)

      Load transformed data
4. Analysis and Reporting:
    - BI Tools (e.g., Tableau, Power BI)

      Generate reports and dashboards

ELT Process:

1. Extract:
    - Data Sources (e.g., MySQL, REST APIs, CSV files)

      Connect and extract data
2. Load:
    - Data Warehouse (e.g., Google BigQuery, Azure Synapse)

      Load raw data
3. Transform:
    - Inside Data Warehouse

      Use SQL or built-in tools to transform data
4. Analysis and Reporting:
    - BI Tools (e.g., Looker, Power BI)

      Generate insights from transformed data

Data Sources

- POS Systems ข้อมูลธุรกรรม, รายการสินค้า, พฤติกรรมการซื้อ
- Web Analytics ข้อมูลผู้ใช้จากเว็บไซต์, การคลิก, Session Duration
- Customer Feedback ข้อมูลจาก Social Media, แบบสอบถามลูกค้า
- Supply Chain Data ข้อมูลสินค้าคงคลัง, การจัดส่ง

Data integration

- POS and Supply Chain

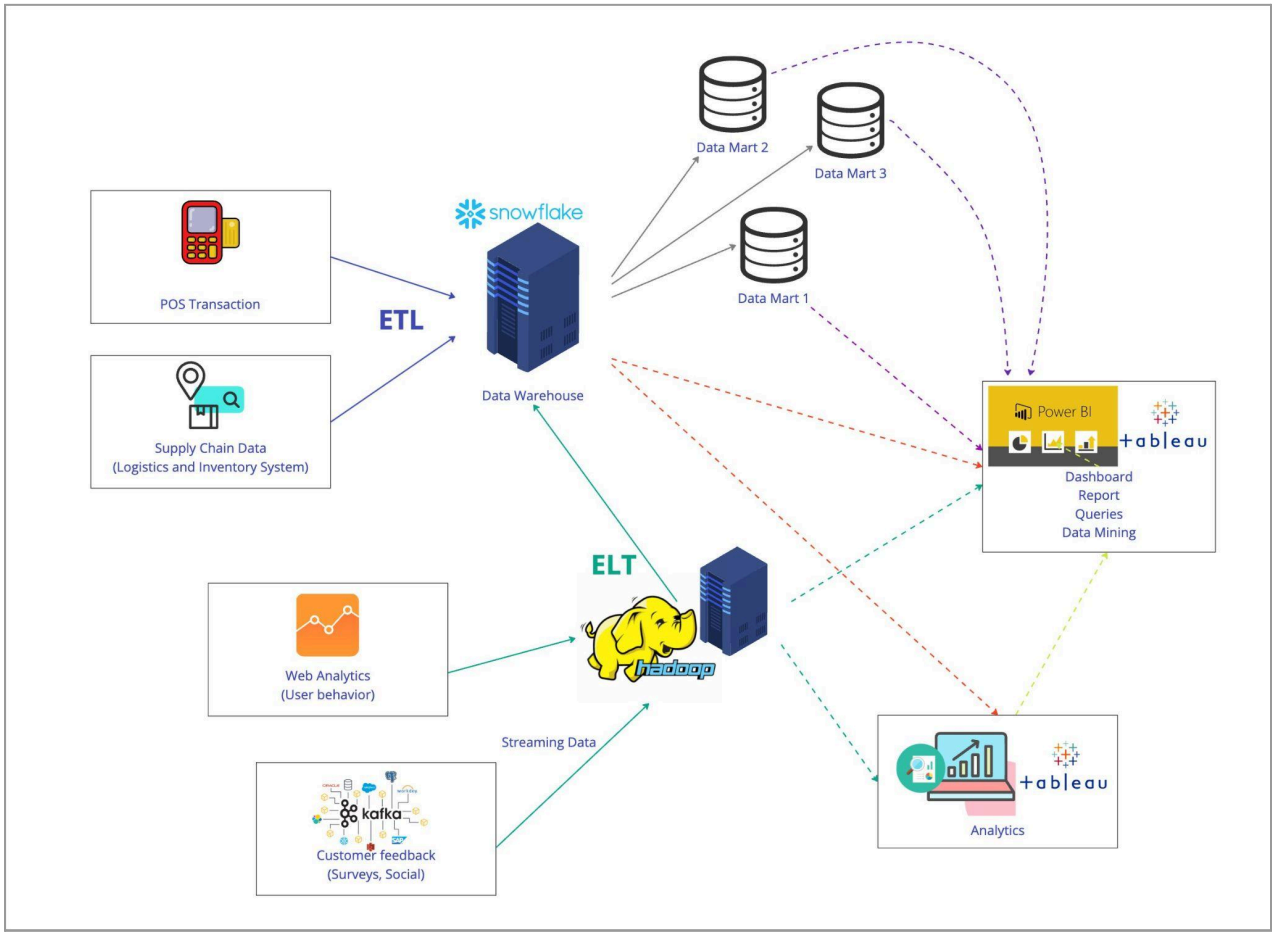
  ETL Tool : Microsoft SSIS, Informatica
- Web Analytics and Customer Feedback

  Streaming Tool : Apache Kafka

Data Storage

- Data warehouse(POS and Supply Chain) AWS Redshift
- Data lake (Web Analytics and Customer Feedback) AWS S3

Data Utilization

- BI Dashboards ⟶ Power BI, Tableau

ETL/ELT processes flow diagram