

Team 1:

Retail

1. Data Lake Task:

- **Task:** Design a Data Lake for a retail company that handles omnichannel customer interactions.
 - **Folder Structure:** Propose a folder structure to store data from different channels (online, in-store, mobile app) and departments (sales, marketing, inventory).
 - **Data Lifecycle Management:** Define policies for data retention, archival, and deletion, considering regulatory requirements and business needs.
 - **Streaming vs. Batch Processing:** Evaluate when to use streaming (e.g., real-time customer behavior analysis) and when to use batch processing (e.g., daily sales reports).

2. Data Warehouse Task:

- **Task:** Develop a Data Warehouse schema for a retail company focused on analyzing customer behavior and sales performance.
 - **Schema Design:** Create a snowflake schema that includes fact tables for sales and promotions, with dimension tables for customers, products, stores, and time.
 - **Data Aggregation:** Define methods for aggregating daily, weekly, and monthly sales data to support executive dashboards.
 - **ETL Process:** Design an ETL workflow that ensures data consistency and accuracy when integrating data from different retail systems (e.g., POS, CRM).
 - 1. **Note:** Just the high level workflow is sufficient.

3. Comparison: Compare Data Warehouse and Database

Team 2:

Banking

1. Data Lake Task:

- **Task:** Implement a Data Lake for a bank to store and analyze transaction data from various channels (ATM, mobile banking, branch transactions).
 - **Folder Structure:** Organize the Data Lake with folders for transactions, customer profiles, fraud detection, and compliance records.
 - **Security and Compliance:** Discuss strategies for securing sensitive financial data and ensuring compliance with regulations like GDPR and PCI DSS. You need to handle Credit Card numbers for fraud detection, Hence you cannot remove the credit card numbers from data completely while ingesting the data and at the same time you need to logically map a credit card with the actual customer without exposing the sensitive card information
 - **Real-time Analytics:** Propose a solution for real-time fraud detection using streaming data from transaction logs.

2. Data Warehouse Task:

- **Task:** Design a Data Warehouse for a bank to manage and analyze risk and credit scoring data.
 - **Schema Design:** Develop a star schema that includes fact tables for credit scores and loan approvals, with dimensions for customers, products, time, and risk categories.
 - **Data Integration:** Plan the integration of data from external credit bureaus and internal transaction systems into the warehouse.
 - **Advanced Analytics:** Suggest the combination of the tables to implement the predictive analytics models within the Data Warehouse to assess customer creditworthiness.
 - **BI Visualisations:** Suggest charts we can create using the data to be created in a BI Tool like PowerBI
- 3. **Comparison:** Compare Data Warehouse and Data Lake

Team3:

Health Care

1. **Data Lake Task:**
 - **Task:** Create a Data Lake for a healthcare provider to collect and store patient data from Appointment systems, medical devices, and genomics research.
 - **Folder Structure:** Design a structure that separates clinical data, imaging data, genomic sequences, and research studies.
 - **Data Governance:** Define governance policies to ensure data quality, privacy, and interoperability across different healthcare data sources.
 - **Data Processing Pipelines:** Propose processing pipelines to handle large volumes of unstructured data (e.g., medical images) and structured data (e.g., patient records).
2. **Data Warehouse Task:**
 - **Task:** Build a Data Warehouse to support hospital management in analyzing patient outcomes and resource utilization.
 - **Schema Design:** Create a star schema with fact tables for patient admissions, treatment outcomes, and resource utilization, and dimension tables for patients, physicians, departments, and time.
 - **Reporting:** Suggest charts we can create using the data to be created in a BI Tool like PowerBI
 - **Data Quality:** Define strategies to ensure data quality in the warehouse, including data cleansing, validation, and consistency checks.
3. **Comparison:** Compare Data Warehouse and Data Mart

Team 4:

Education

1. **Data Lake Task:**

- **Task:** Design a Data Lake for an educational institution to store and analyze data from various sources such as student information systems, online learning platforms, and feedback forms.
 - **Folder Structure:** Propose a folder structure that includes student data, course materials, assessments, and feedback.
 - **Data Security and Privacy:** Discuss methods to secure student data, ensure privacy, and comply with regulations such as FERPA.
 - **Integration and Analytics:** Suggest methods to integrate real-time data from online learning platforms and analyze it to identify students at risk of falling behind.

2. **Data Warehouse Task:**

- **Task:** Develop a Data Warehouse for an educational institution to track and analyze student performance, enrollment trends, and resource usage.
 - **Schema Design:** Design a schema with fact tables for grades, attendance, and resource utilization, and dimension tables for students, courses, instructors, and semesters.
 - **Predictive Analytics:** Suggest charts we can create using the data to be created in a BI Tool like PowerBI
 - **ETL Process:** Outline the ETL process for extracting data from various academic systems and transforming it to support reporting and analysis.

3. **Comparison:** Analyse on how data Lake is integrated with BI Tools (Amazon Quicksight)