Data Engineering Case Study: Car Insurance Industry

Objective:

Encourage students to apply their knowledge of data engineering concepts to real-world scenarios in the car insurance industry. This activity will help students develop their researching and critical thinking skills.

Instructions:

You will work as a group to research and analyze how data engineering concepts are applied in the car insurance industry. Use the questions and tasks below to guide your research. Be prepared to present your findings and discuss them with the class.

Tasks:

1. **Types of Data**:
   - Identify and describe the different types of data used by car insurance companies (e.g., structured, unstructured, semi-structured data).
   - Provide examples of each type of data specific to the car insurance industry.

2. **Types of Databases**:
   - Research how car insurance companies utilize both relational and non-relational databases.
   - Provide examples of scenarios where each type of database would be most appropriate.

3. **Data Warehouse vs Data Lake vs Lakehouse**:
   - Compare and contrast the use of data warehouses, data lakes, and data lakehouses in the car insurance industry.
   - Find a real-world example of an insurance company (e.g. car insurance generic or Hartford specific) that uses one of these data storage solutions and explain why it suits their needs.

4. **Big Data**:
   - Explain the concept of big data in the context of car insurance.
   - Identify the sources of big data for car insurance companies and describe how this data is collected and processed.

5. **File Formats**:
   - Research the various file formats (XML, JSON, CSV, AVRO, Parquet, ORC) used in the car insurance industry (or Hartford specific).
   - Provide examples of when and why each file format might be used.

6. **Case Study**:

- Investigate how The Hartford leverages data engineering to improve their services, customer experience, and operational efficiency.
- Prepare a summary of your findings, highlighting key points such as data collection methods, data storage solutions, and any innovative data engineering practices they employ.

Deliverables:

- **Presentation**: A 5-minute presentation to share your key insights and findings with the class.