Data Manager: Case Study

Scenario:

An E-commerce company is experiencing rapid growth, offering a wide range of products to its customers. The company has several data sources, such as transactional databases (Oracle, MySQL, SQL Server), NoSQL databases (Cassandra, Elasticsearch), event streaming platforms, CSV files, and external APIs.

The company's vision is to become data-centric, ensuring that everyone in the organization, including non-technical staff, can access and analyze data to inform their daily decisions. The company has a team of data analysts and data scientists who are eager to explore this data to generate insights, KPIs, and customer segmentation. The data engineering team's mission is to provide this data to the organization's data lakehouse regularly, securely, and without incidents.

The marketing team has requested a real-time view of customer interactions on the website. The data will be utilized by an in-house developed Machine Learning solution to predict customer behavior and preferences, assisting the marketing team in making immediate decisions regarding personalized offers and promotions.

The E-commerce company is in a phase of rapid national expansion, with the number of unique transactions currently around 100 million per day and expected to grow exponentially. The data engineering team must design an architecture that not only supports the current scenario but also scales as the company expands.

Requirements:

1. Draw a diagram of the data lakehouse architecture using any cloud service (AWS, Azure, or GCP) that addresses the issues above.
2. Describe how the proposed architecture solves each of the described needs.
3. Your architecture must respect existing regulatory standards (GDPR/LGPD).
4. Your architecture can be multi-cloud.
5. The delivery must be done in the format of a slide presentation in any tool (Microsoft PowerPoint, Google Slides, etc.).

Evaluation Criteria:

1. The ability to design a scalable, robust, and secure data architecture.
2. Understanding of data flow from different sources to the data lakehouse.
3. Knowledge of cloud-based data services and their application.
4. Compliance with data privacy and security regulations.
5. Clear and concise presentation of the proposed architecture and its benefits.