SYSTEM ARCHITECTURE AND DESIGN

This web-based SAAS has a 3 tier Client-Server Architecture which segregates the system into two applications, where the client makes requests to the server(through Power BI). In this case, the server is a database(Azure Data Lakehouse Gen2) with application logic represented as stored procedures. The tiers in detail are as follows:

Presentation Tier – Power BI is the user interface where the end user interacts with this desktop application.

Application Tier – This tier is the heart of the design, known as the middle tier where the information is collected in the presentation tier is processed using Python(Spark) and Dax.

Data Tier- Sometimes called the database tier or back-end where the information processed is stored and managed. This solution uses Microsoft Azure for real-time lakehouse processing through Azure SQL database.

Data Lakehouse – Is a new open data management architecture that combines the flexibility, cost-efficiency, and scale of data lakes with the data management and ACID transactions of data warehouses, enabling BI and ML on all data.

Persistent Data Storage- There is not persistent data storage with Azure BLOB in which Azure Data Lake Storage Gen 2 is built on. Azure Blob storage is an object storage solution where each object is stored in a flat name space. Whereas, Azure Data Lake Gen2 allows a hierarchical namespace similar to a HADFS. Azure does have a persistent store for long-term storage of file. In this case, this system uses an Azure Blob storage tier classified as a "hot tier" which is optimal for frequent reading and writing access to stored data. The data will be stored in a parquet file format and then stored in folders. The objects stored in the folders will be JSON files,

excel files, and folders created during the transformation from unstructured to structed data within the landing zones.

Global Control Flow-

- Execution order: Azure Functions is a serverless compute option. It uses an event-driven model, where a piece of code (a function) is invoked by a trigger. In this architecture, when events arrive at Event Hubs, they trigger a function that processes the events and writes the results to storage.
- Time dependency: Is it a near real time system.
- Concurrency: Does your system use multiple threads? NO

- Hardware Requirements

    - What system resources your system depend on?
    - Internet Connection
    - Active accounts for Microsoft 365, Azure, and Power BI.
    - A device that runs Windows 10 Home, Windows 10 Pro, Windows 10 Enterprise, Windows 11 Home, Windows 11 Pro, Windows 11 Enterprise, Windows Server 2016, Windows Server 2019 or Windows Server 2022. ARM devices aren't supported. Minimum hardware- Processor: 1.00 GHz or faster with two or more cores. For unattended mode, four or more cores are needed. Storage: 1 GB,RAM: 2 GB