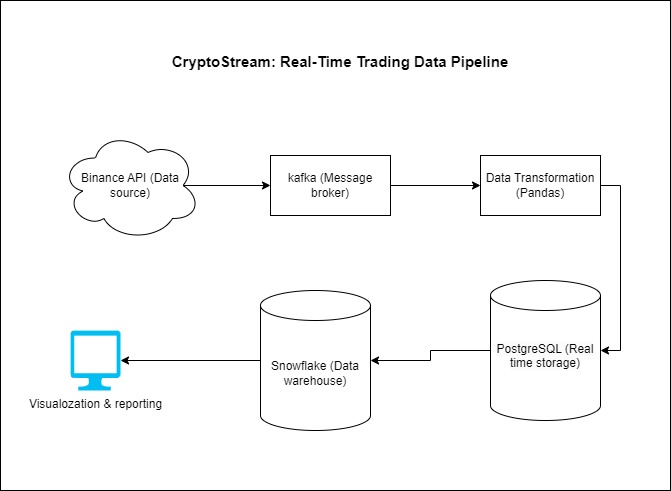
**PROJECT DOCUMENTATION**



**Relevance of Kafka:**

1. **Decoupling the System**: Kafka helps decouple the data flow between the **Binance API** (data source) and the **data transformation/processing** layer. This allows the two systems to operate independently and handle the real-time flow of data in a scalable manner.
2. **Handling Real-time Data Streams**: Since cryptocurrency market data is generated continuously, Kafka can efficiently capture and buffer this high-throughput data stream in real time. Binance provides WebSocket data, and Kafka consumes this data without overwhelming the downstream system (like Pandas for transformation, PostgreSQL for storage, etc.).
3. **Scalability**: As your data load grows, Kafka allows you to scale horizontally by partitioning your topics across multiple brokers, making it highly suitable for handling large volumes of trading data.
4. **Fault Tolerance**: Kafka provides durability and fault tolerance with replication. This ensures that even if part of your system fails, you won't lose any market data.
5. **Buffering for Downstream Systems**: Kafka stores the data stream for a configurable period, allowing downstream systems (like PostgreSQL and Snowflake) to consume the data at their own pace. This avoids bottlenecks.