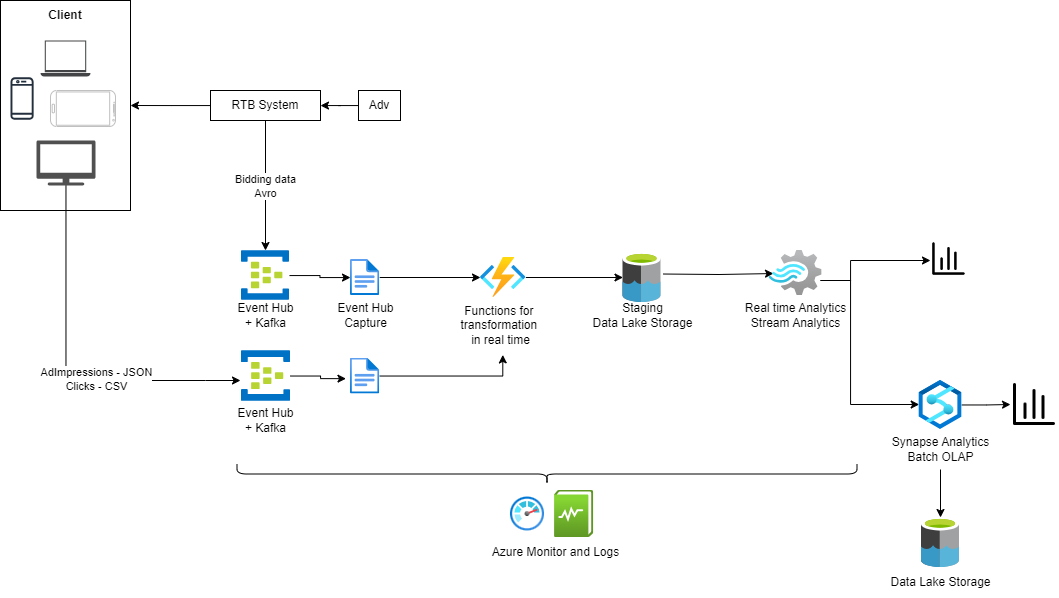
AdvertiseX Assignment

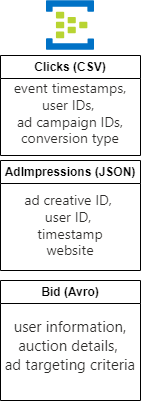
Summary of problem statement:

1. Data Ingestion:
   1. Different sources and types
   2. High volume real-time
2. Data processing
   1. Real-time transformations
   2. Correlate clicks and ad impressions
3. Data Storage
4. Error Handling and Monitoring

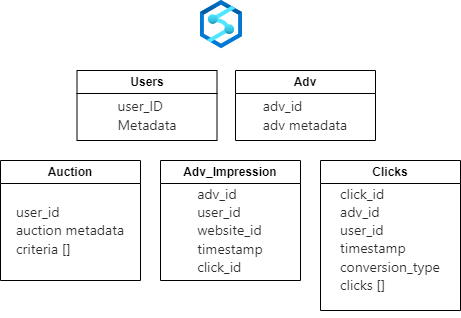
Solution:



1. Data ingestion
   1. Apache Kafka can be used to serialize and deserialize the various types of data sources in real-time. Producers can be developed to read CSV or JSON data from various sources (e.g., files, databases, APIs) and then serialize the data into a format suitable for Kafka messages before publishing them to Kafka topics. Once the data is published to Kafka topics, Kafka consumers can subscribe to these topics to process and consume the messages. Consumers can then deserialize the messages back into their original format (e.g., CSV or JSON) and perform further processing, analysis, or storage as needed.



1. Data Processing
   1. Transformation on real-time windows of data can be performed on a small-scale service like Azure function apps. This will provide serverless compute environment for executing event-driven code. Within the Azure Function, you can implement custom logic to parse, enrich, filter, aggregate, or otherwise manipulate the captured data according to your business
   2. In this case, the incoming data can be filtered for the respective company taken for analysis.
   3. After the staging phase, stream analytics can be used to further refine/aggregate data into gold layer.
   4. Clicks data can be aggregated for each AdImpression. This will help correlate AdImpressions and Clicks using click\_id.
   5. Bidding can be aggregated for each user
2. Data Storage
   1. Data Lake Storage is used for staging purposes to connect real-time end points
   2. Synapse Analytics is used for warehousing purposes and for quick OLAP analysis



1. Error Handling and monitoring
   1. Azure monitor and logging provides an easy way of setting business metrics and analyzing the same for checking system health and performance degradations.
   2. Function Apps can implement logic to detect and address data anomalies, such as missing or malformed data, and implement retry mechanisms for failed data processing tasks to ensure fault tolerance and data integrity.
   3. Track key performance indicators (KPIs) related to business outcomes and objectives, such as ad campaign performance metrics (e.g., CTR, conversion rates, ROI), customer engagement metrics, revenue generated from ad sales, and customer satisfaction scores.