**Overview of the approach used**

**How the approach would provide a good data source to the clients**

Kafka clusters contain multiple brokers. The partitions of a particular topic are distributed among several brokers to aid in horizontal scaling. From the user’s (The individual who records audio) point of view, different consumers can access different instances of topic partition. They can therefore upload files in parallel without lagging the system. This would therefore improve the user experience while interacting with the system

From the receiver of the audio point of view, data ETL would be real-time due to the capability built on Kafka. Once a user uploads an audio file equivalent to the text, it is immediately sent to the Kafka cluster, then gets transformed by Spark before being loaded into the data lake. This approach ensures timely access to important data and would provide an efficient platform for Continuous Integration and Continuous Development. The CI/CD would be enabled In such a way that the clients can start developing products with the data available and incrementally fine-tune the system as more data gets continuously loaded into the system.