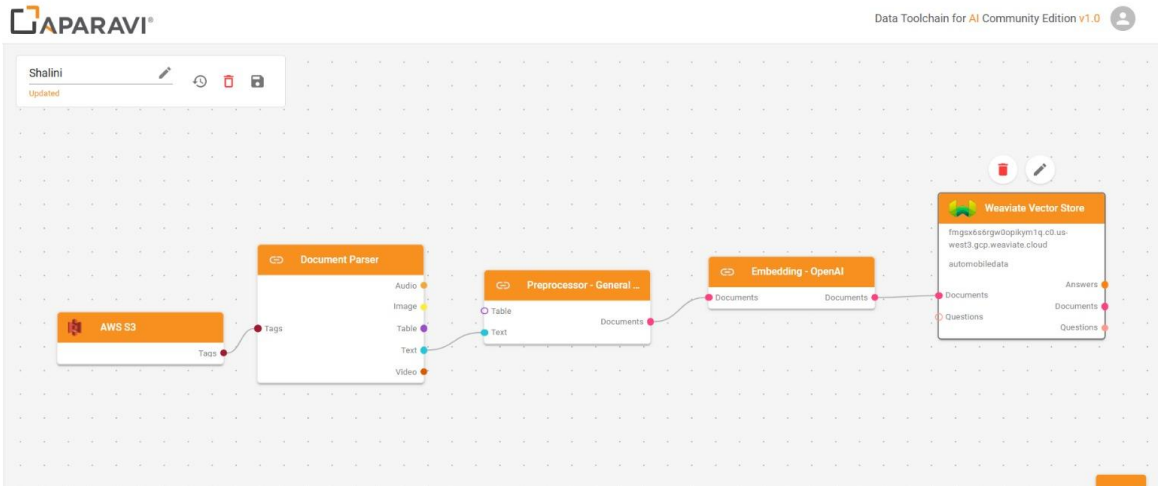
Technical document



1. **AWS S3**

   The software which we were using is called Aparavi. When we were using the Aparavi we had created a new project. Inside the new project we had integrated using the Aws s3 node because we were connecting our data from 3rd party source. This bucket is really important because it is containing the relevant data for the Automobile industry. This connection is ensuring that we can securely access the unstructured data which we need for further processing to gain meaningful insights by creating the chatbot which is our overall goal.

2. **Document Paser**

   Next we have attached the Document parser node which is containing the sored data for the Aws 3 bucket. This parser can then handle the tabular or non-tabular data which is extracting the relevant dataset fields like brand of cars which is crucial for the further analysis.

3. **Preprocessor**

   When we are preparing the data we have used a pre-processor node which is standardising the data or splitting the large content of unstructured data to smaller chunks. This is ensuring that the data is more readable format for the next stage in the process.

4. **Embedding – Open AI**

   We have connected the open AI node to generate the vector embedding which is enabling the semantic searches in Weaviate. This helps to allow the chatbot to extract the data based on the question we have asked and give the answer. This step is really important because we are enabling the features like NPL to understand our chatbot.

5. **Weaviate vector store**

   By sending the embedded dataset to the Weaviate node, we can validate that everything is properly stored in collection and indexed so that it can be retrieved really easy when we are creating the chatbot by answering the small user queries.

6. **Google collab**

   Finally, we had created a google collab notebook to ensure all the necessary coding is done so that we can build the chatbot. This notebook is allowing us to integrate all the necessary components to run efficiently to the end of the pipeline which is resulting in creating a flexible and collaborative working environment.