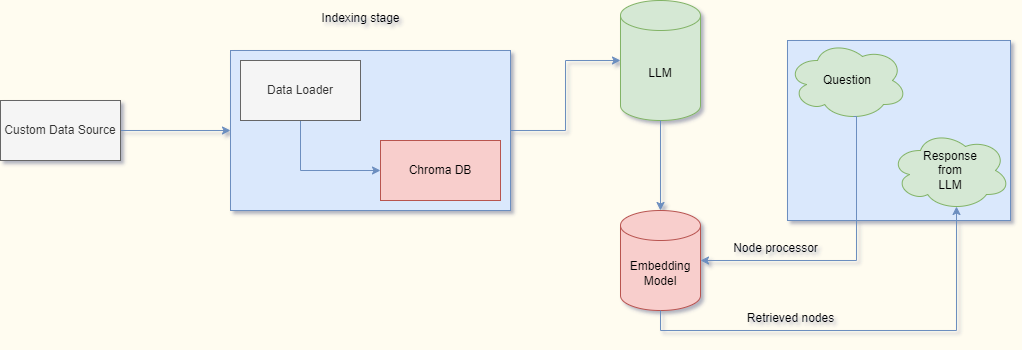
Code Documentation for Chatiiitb

The documentation format is as follows:

* File name
  + Description
  + Function name
    - Documentation
    - Future scope (If possible)
* Folder name
  + Sub-folders
  + Info
* /data
  + **Sub-folders**: There are no sub-folder
  + **Info**: This directory consists of all the files (pdf, markdown, text, etc) to create the vector embeddings
* /local\_chroma\_db
  + **Sub-folders**: There is a folder generated during the creating of vector embeddings.
  + **Info**: This directory consists of the sqlite3 database of the vector and also a folder consisting of binary files for the database
* chat.py
  + **Description:** This is the only executable file in this project. I is responsible for loading data, splitting text into chunks, making API calls to convert text data to vector embeddings and running the chatbot.
  + **Function:** hash\_it
    - **Documentation:** It returns the SHA256 hash for a piece of text to be used as unique identifier for it’s corresponding vector embedding in the vector DB.
    - **Future scope:** As of now, there is no need to make any changes to the function.
  + **Function:** load\_data
    - **Documentation:** It read the data from the files in the `data` directory and passes it through langchain tools to create the vector embeddings and stores it in the `local\_charoma\_db` in a sqlite3 database.
    - **Future scope:** Hyperparameters like chunk\_overlap, chunk\_size, etc can be optimized. Different databases can be used to get better embeddings.

**Chatiiitb Flowchart**

****