**ASSET : Policy Advisor**

**Databricks:**

* **Spark:** Use Spark to process and analyze policy data .
* **Delta Lake:** Use Delta Lake to store and manage policy data.
* **MLflow:** Use MLflow to manage machine learning models and deploy them to production.

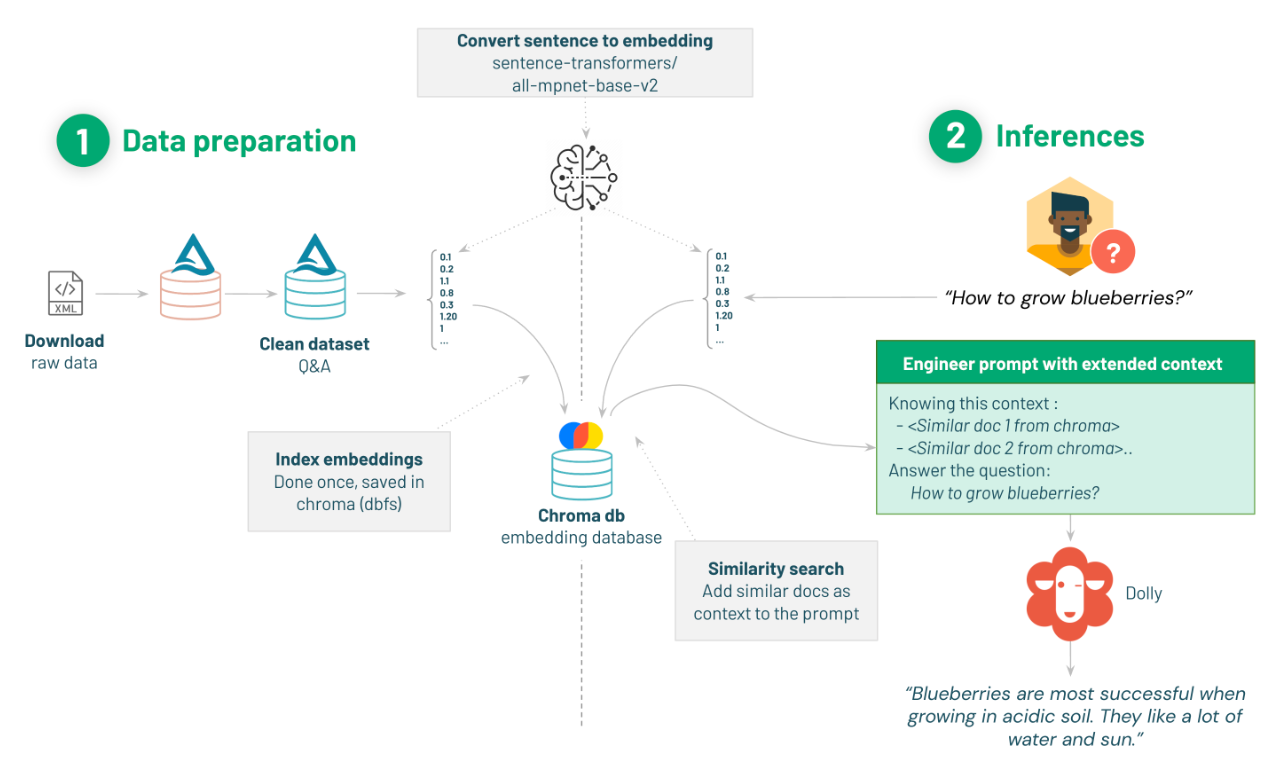
**Natural language processing:**   
  
Use NLP to understand user questions and generate responses (language identification , lexical analysis, lemmatization, entity extraction, intent recognition)

**Large language models (LLMs):**

* **GPT:** GPT is a powerful LLM that can be used to generate text, translate languages, and answer questions in an informative way.
* **Dolly**: is the first open and commercially viable instruction-tuned LLM, created by Databricks. It is designed to efficiently understand and follow instructions provided in natural language, making it an incredibly powerful tool for a wide range of applications.
* **llama2-70B-Chat**: A LLM that is trained on a large corpus of conversational data and can handle various types of dialogues, such as chit-chat, question answering, and task-oriented conversations.

**Implementation:**

1/ Data preparation: ingest and clean our Q&A dataset, transforming them as embeddings in a vector database (exp : chromeDB)

****2/ Q&A inference: leverage an LLM model (exp Dolly ) to answer our query, leveraging our Q&A as extra context for Dolly. This is also known as Prompt Engineering.  
   
These steps are illustrated in the figure down below :