Details:

1. The data provided is already chunked i.e. you need to only index the contexts of the data. You can index them using any of the different strategies you may like
2. The data is provided in two formats namely csv and jsonl file. Jsolnl file is self explanatory once you load them. The csv file details are provided below:

Squad\_v2\_dataset.csv contains following columns:

“Question”, “answers”, “context\_id”

Context.csv contains following columns:

“Context\_id”, “context”

1. Note: For some questions, the answer is not present in the actual context. However, this will not affect the retrieval metrics calculation as each question is already associated with a context.