**Fastapi:**

I used fast api to extract data because it is best for making micro service and less time take to execute. As compare to flask fast api is fast.

**kafkatopic:**

I used this to send data after one min to pyspark by making kafka topic producer.Kafka is best when you want send data in bulk or realtime .

**pyspark:**

In pyspark you transform all data and merge.In this we can clean data and merge data .Through this we getther all data together .It is best for batch and stream data as compare to pandas.

**PostgressSQL:**

I used postgress data where I dump all data . Data scientist get data from here through api. This is best for getting data fastly.Here All data come

from transformation which is clean.