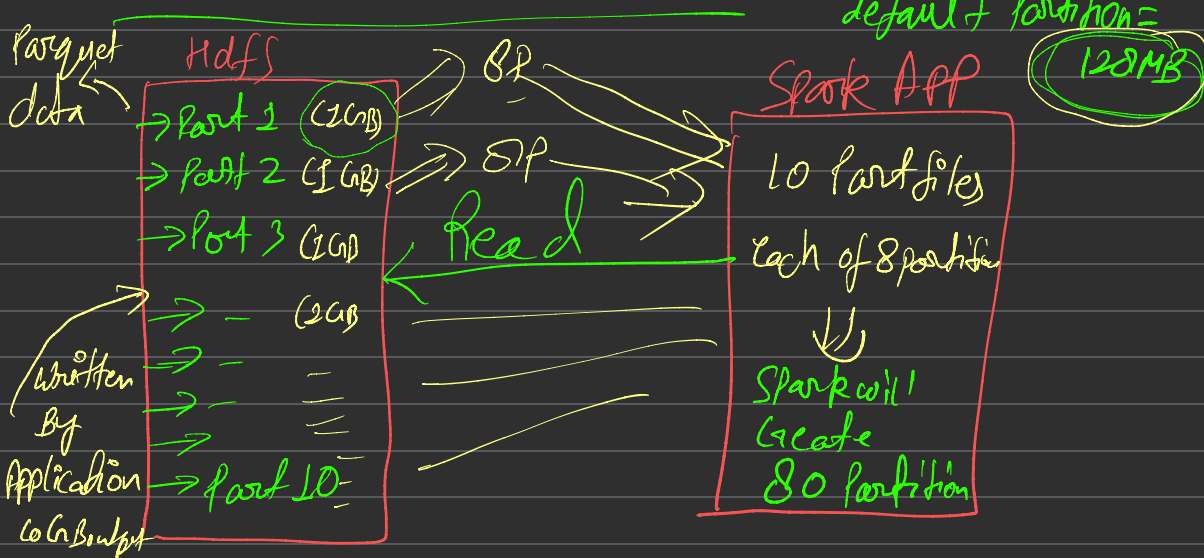
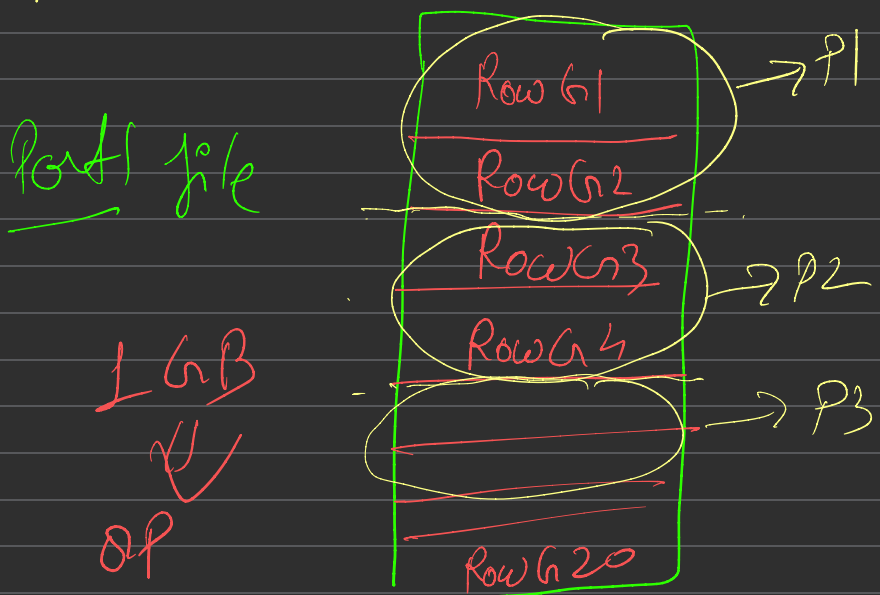
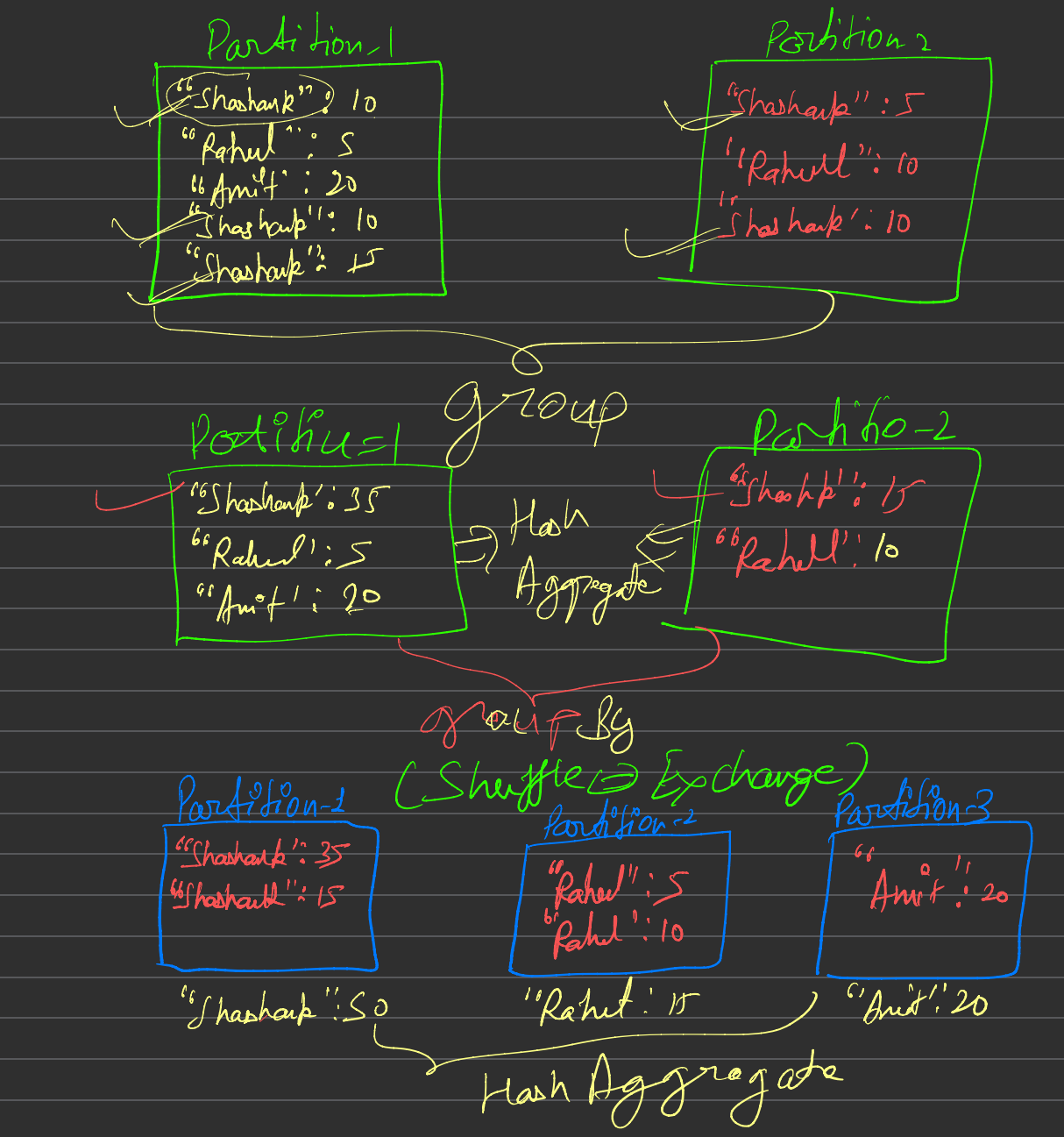
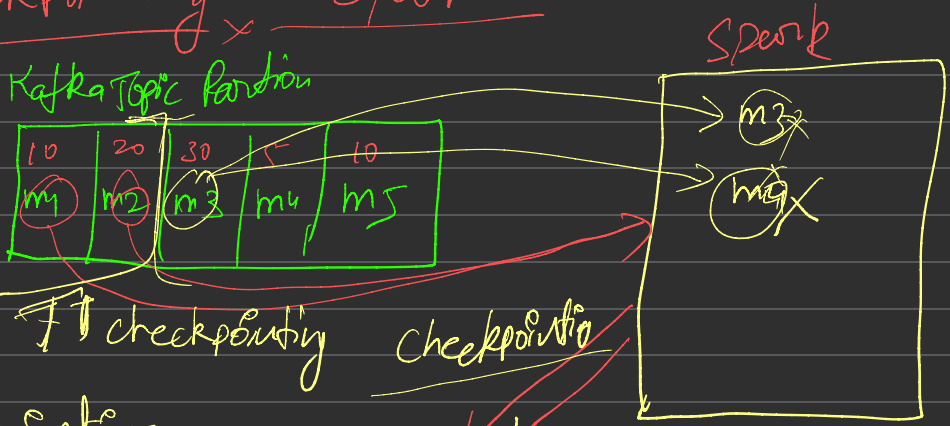
If source data is in parquet then how many partition ?







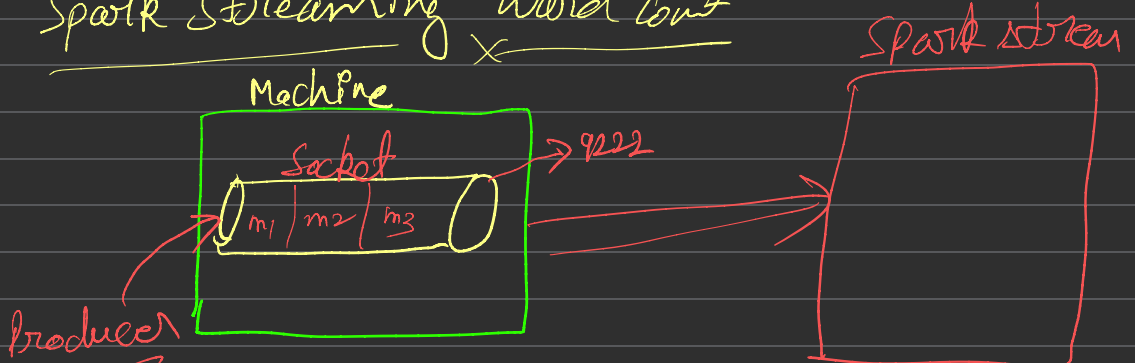
Checkpointing in Spark



Checkpointing
=> /tmp/checkpointing
Checkpointing Interval 10sec

=> Till which offset Storage Should Consumed
=> Source total = 30
File System / Database / RAM

Spark Streaming Word Count



m1 => Hello Shashank
m2 => How are you Shashank
m3 => Shashank are you OK

```
Word = lines.select(explode(split(lines.value," ")).alias("word"))
```