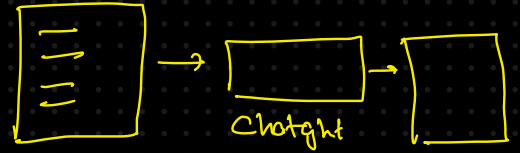8th Feb

1. Many of you are still on hadoop and MR.
2. Please be a lot active on the forums
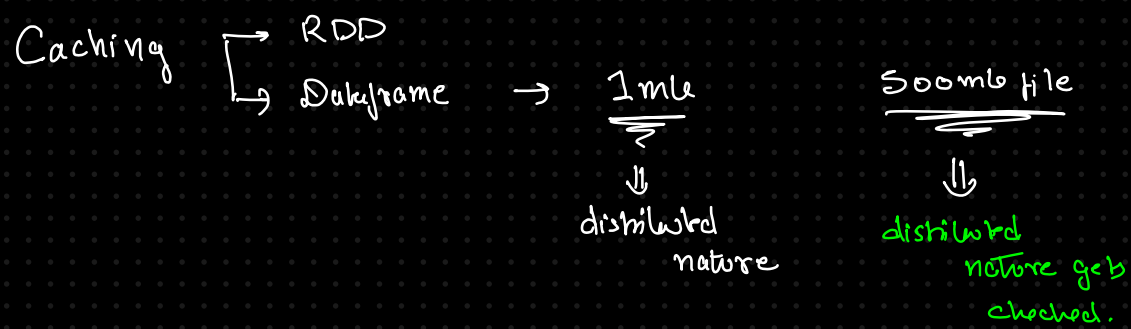3. Please be a lot active on Linked In or socials → LinhedIn

Chatght

February ⇒ Hive + Spark + Project

⇒) Cloud ⇒ Azure ⟶ Data factory

⇒) Airflow, MongoDB, cassandra, Kafka

Caching ⌐→ RDD
         └→ Dataframe   →   1mle
                                ̲̲⇌̲̲
                                ⇓
                              distributd
                                   nature

500mb file
̲̲̲̲=̲̲̲̲
⇓
distributd
    nature geb
      chocked.

Caching Type	Description	When to Use
Eager Caching	Data is cached immediately after the command.	When data is small to medium in size and frequently accessed.
Lazy Caching	Data is cached only when accessed for the first time.	For large datasets with selective access.

Customers - 500mle
̲̲̲̲̲̲̲̲̲̲̲̲̲̲̲̲̲
⇓
500 mb

Cache it lazily