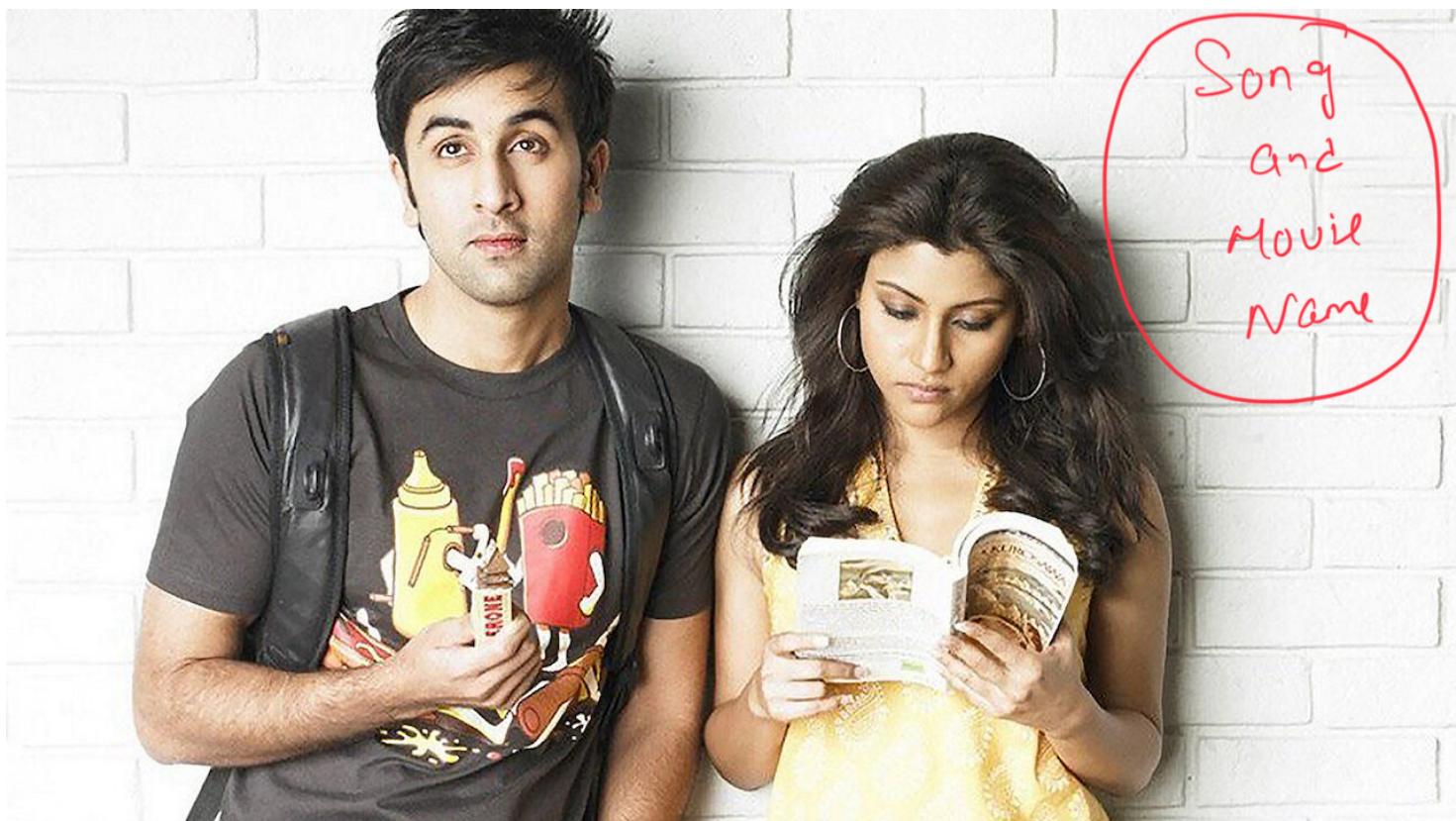


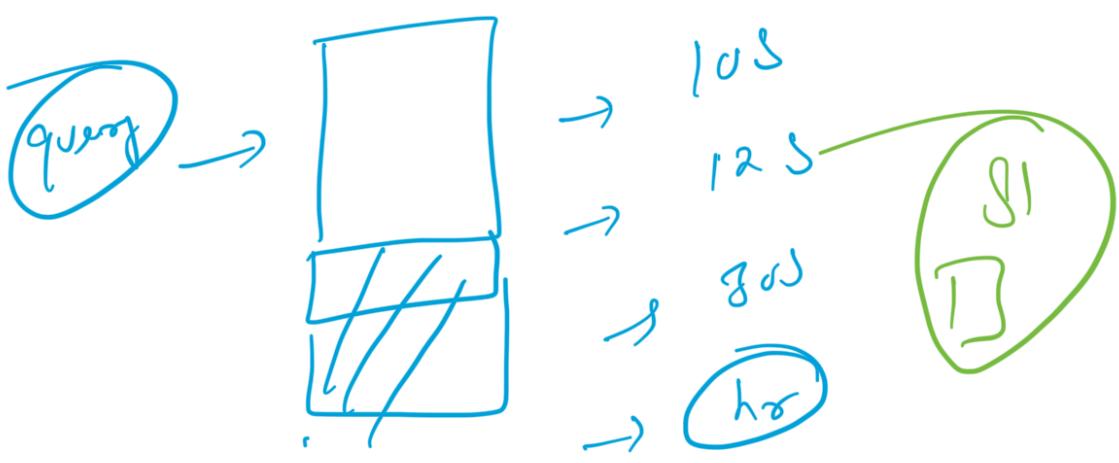
## Agenda

- A. Concepts of Big data
- B. Hadoop address big data challenges
- C. Core components of hadoop



Volume / Velocity

Traditional Data Systems

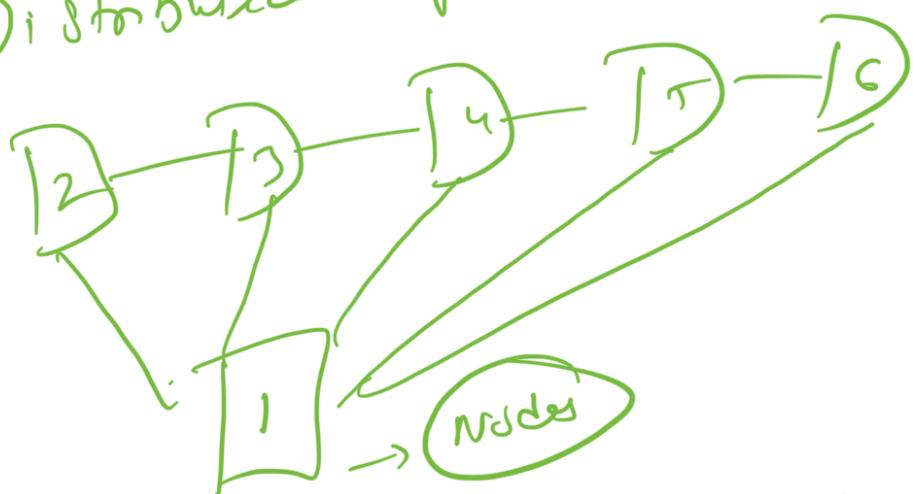


Demerits

- Requires human intervention at various stages
- Query | Information takes a good amount of time to process.
- Lacks systematic usages among strategy, planning, execution & reporting.

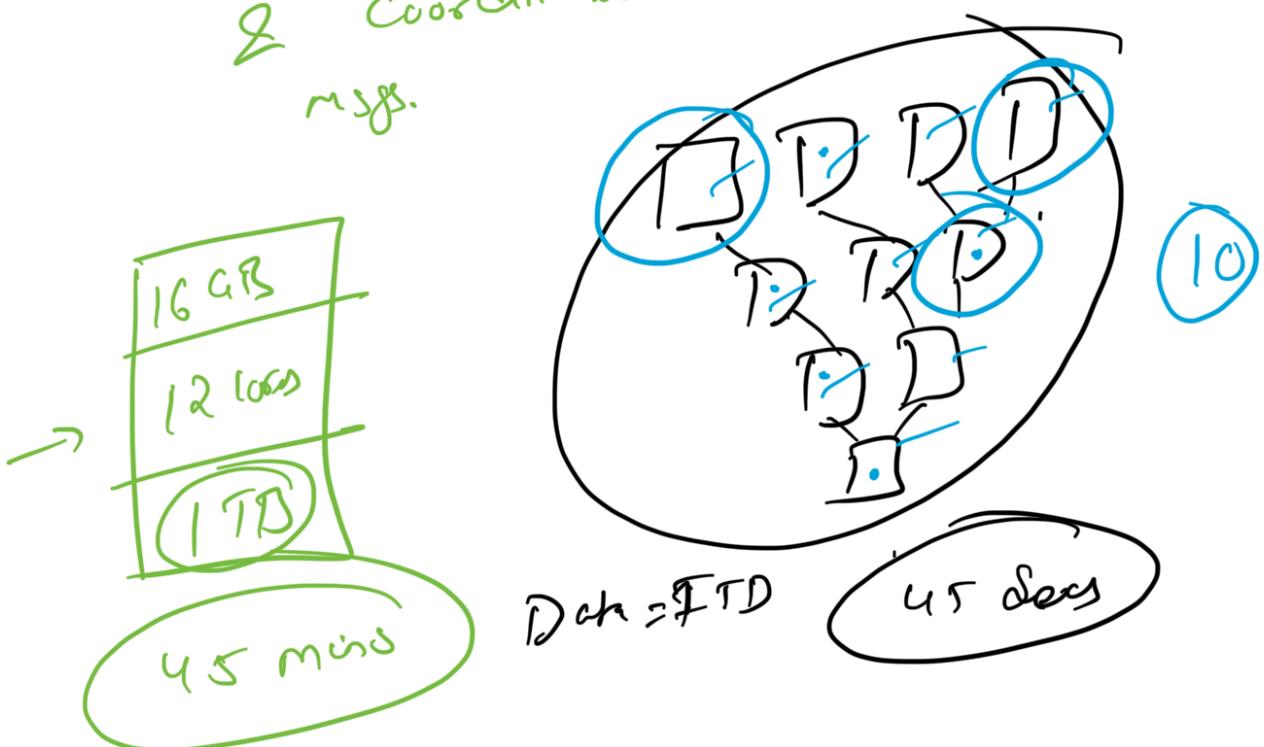
Distributed Systems ?

② Cluster



... in components

\* → A DS is a model in which ~  
located on n/w computers communicate  
& coordinate their actions by passing  
msgs.



- ① System failure :
- ② Limited Bandwidth :
- ③ High Programming Complexity :

Hadoop

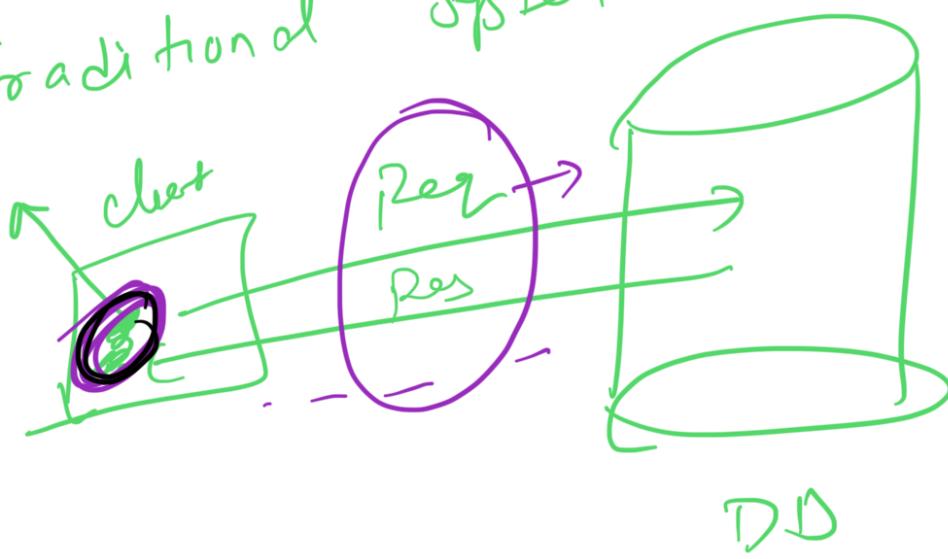
→ Hadoop is a framework that allows  
distributed processing of large datasets  
on a lot of commodity Computers

across clusters  
using Simple programming models.

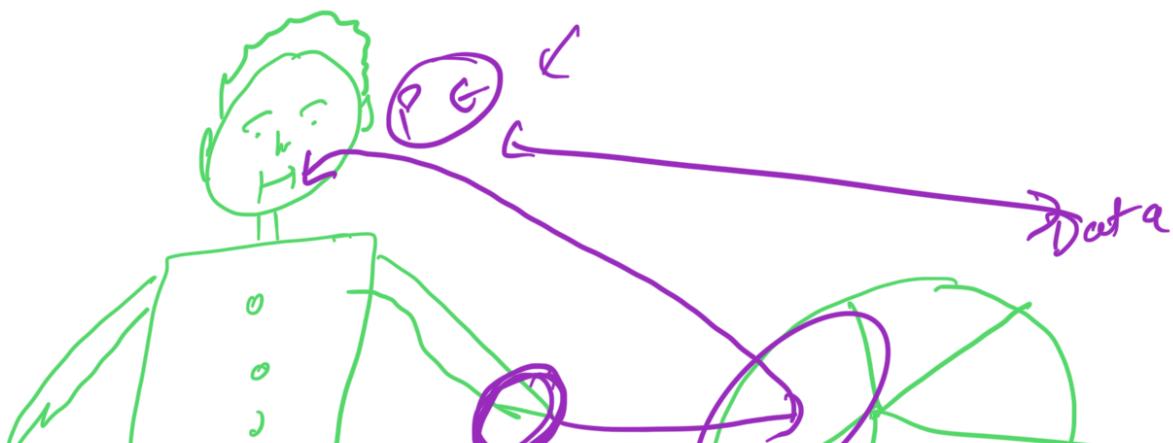
characteristics:-

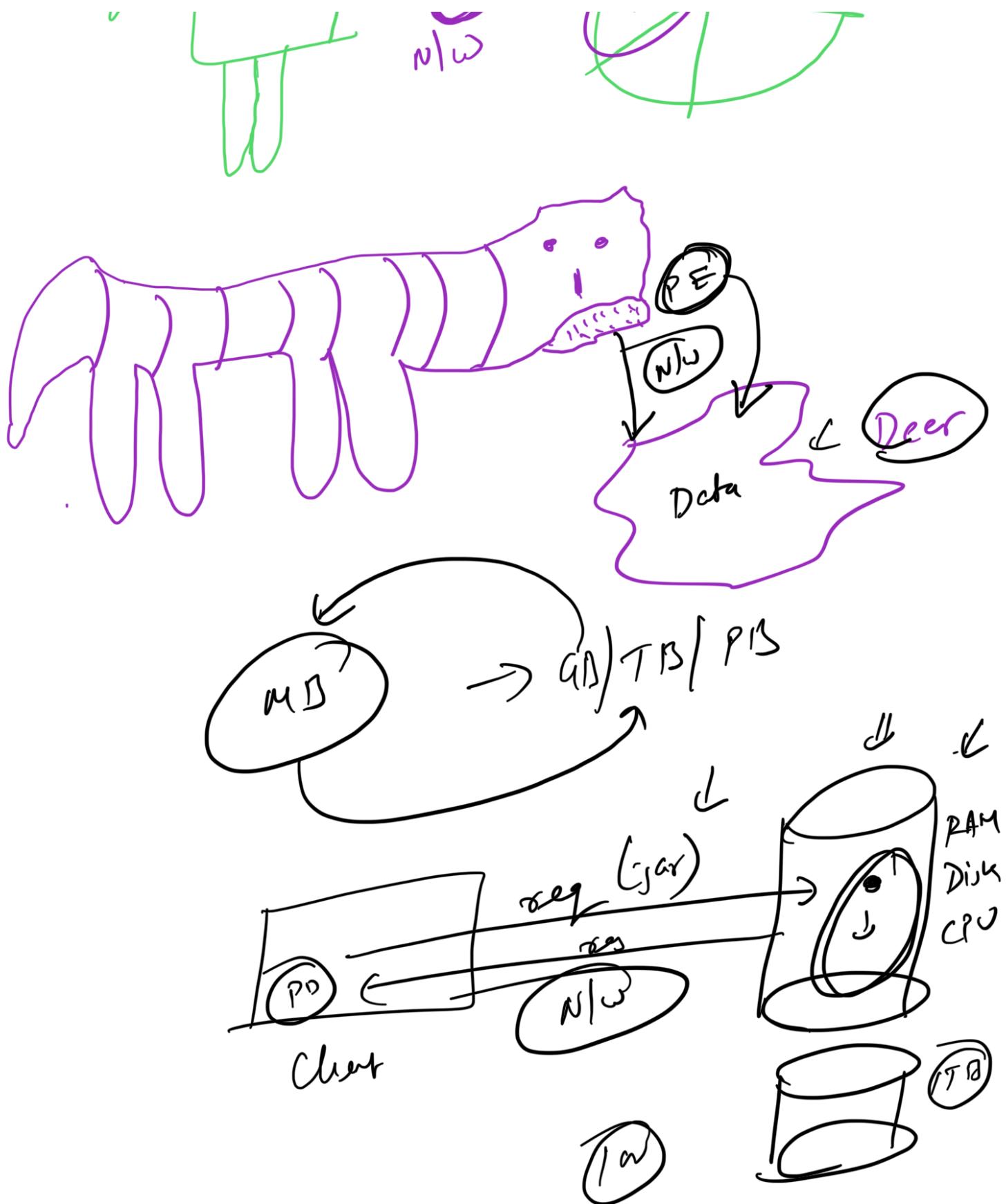
- Scalable ( $rs/n^2$ )
- Reliable (copies)
- Economical
- Flexible

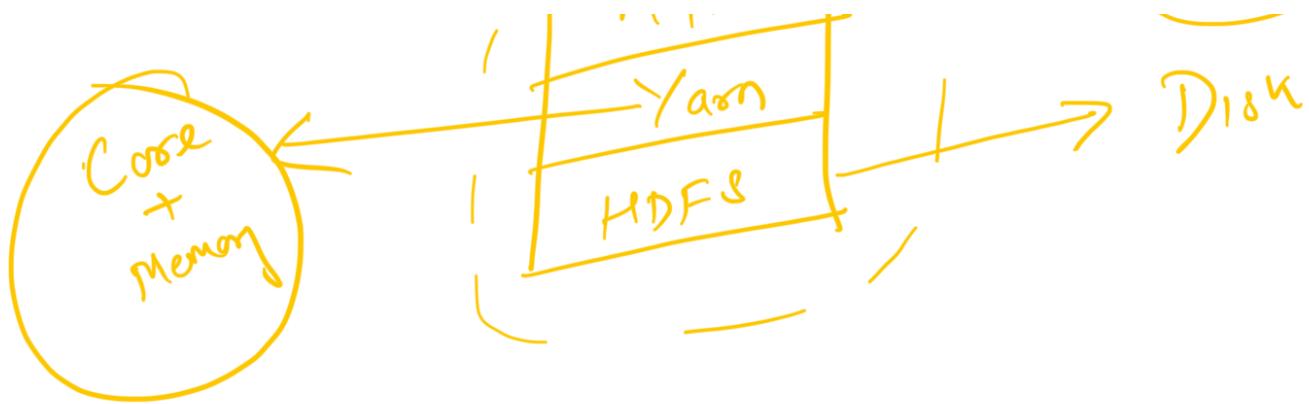
Traditional System



Aditya

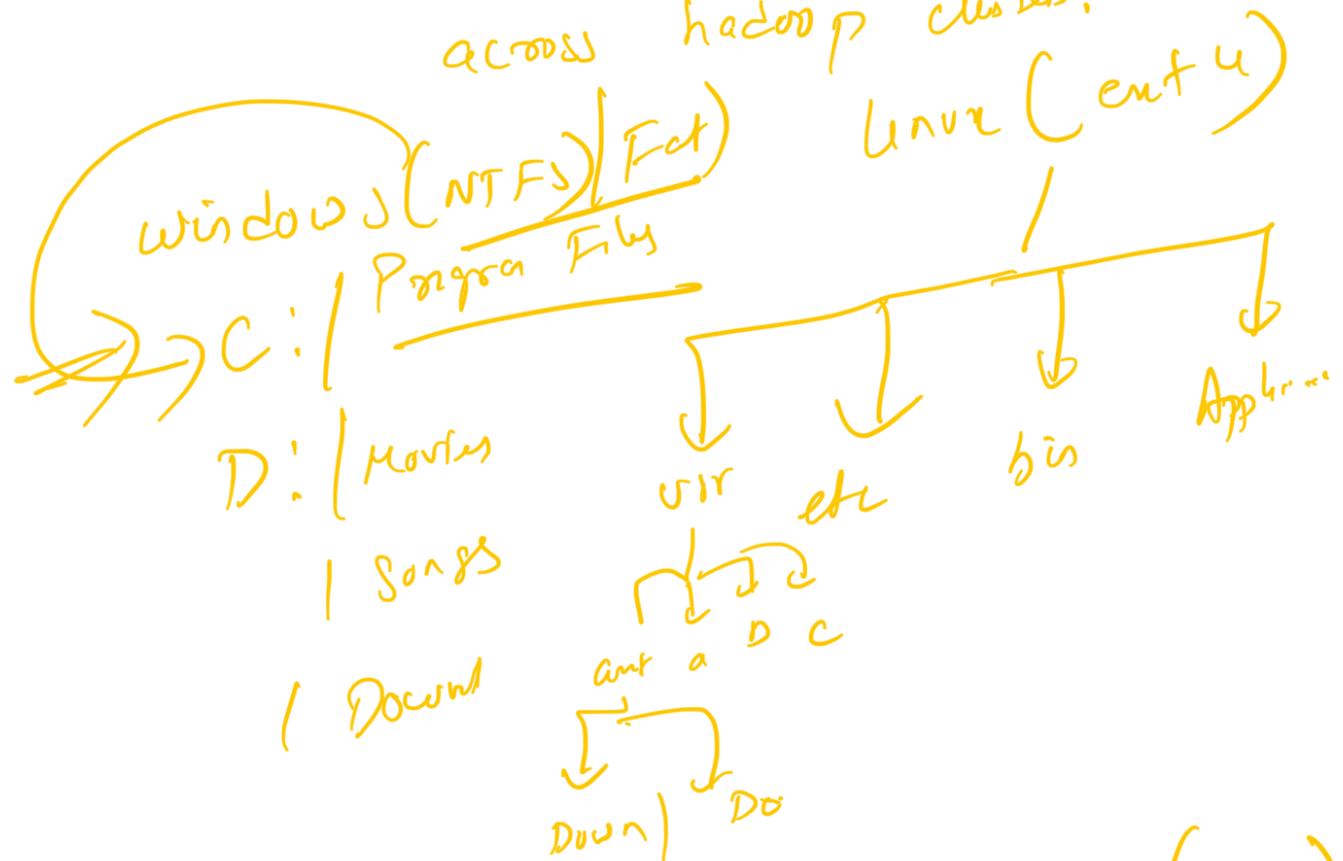






HDFS  
Hadoop Distributed File System (Linux)

↳ that provides access to data across hadoop clusters.



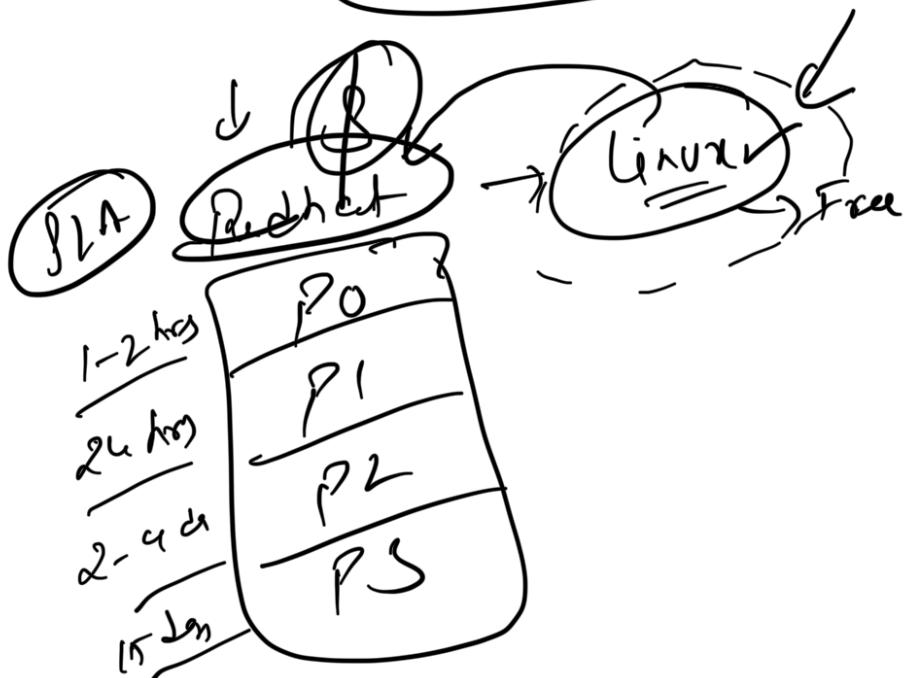
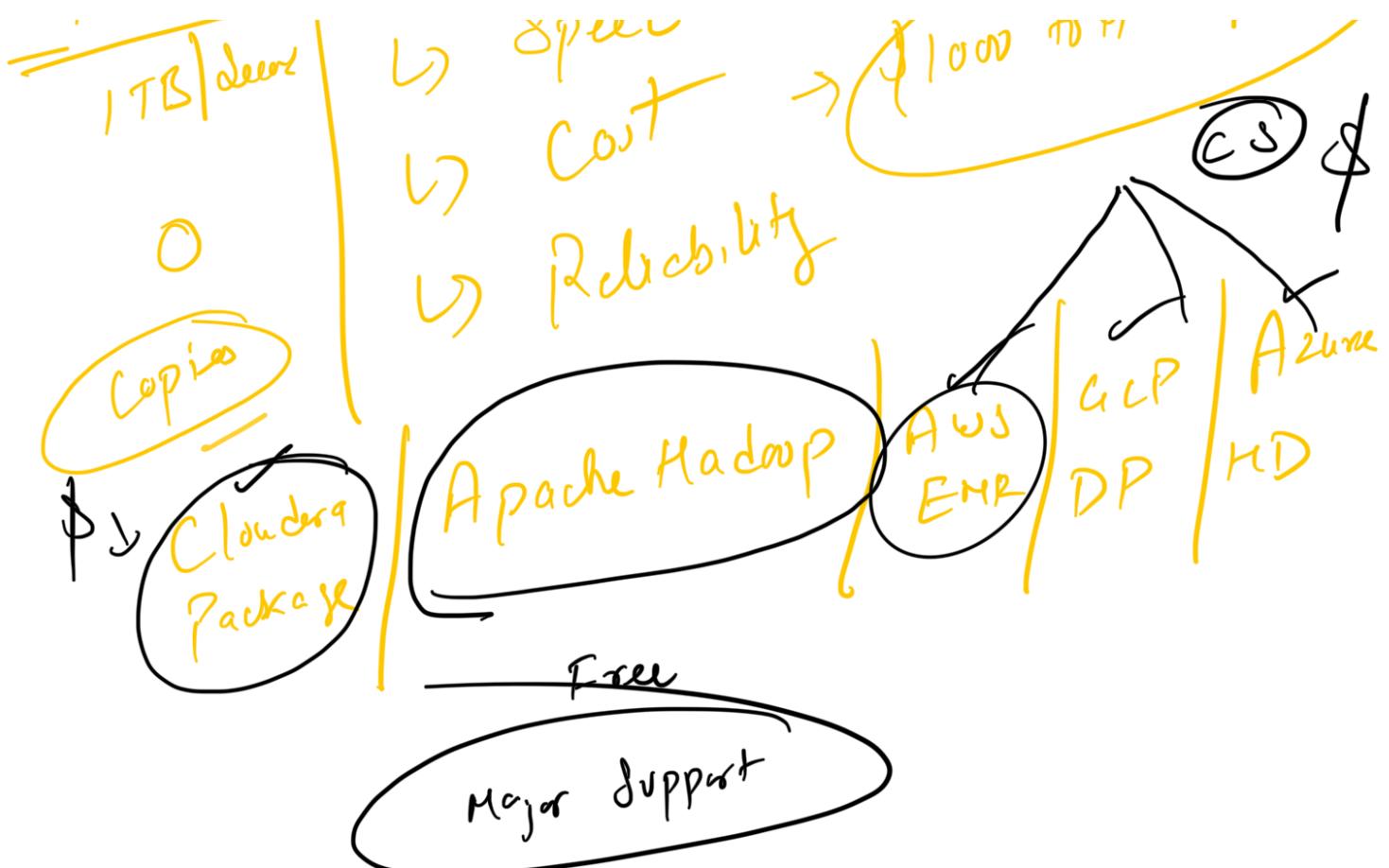
↓ challenges with traditional FS (Windows)

HDFS

↓

○...n.

L. MySQL | TD



## HDFS Architecture

