

19/nov/25: HBASE

- Putting data on cloud is not v stable
- Faida we know, what about nuqsan, in a data engineering perspective
- We cant put whole oracle on cloud as its terabytes
- We need data governance policies
- We need pipelining and data engineering
- Devops methodology
- Data pipelines
- It would be random if not done
- Open source: hbase and casaandara
- Deployments khaas nahi hai

24/nov/2025

- Hive mein partitioning querying sab haath mein hai
- Major compaction
- Minor compaction
- Column family is compact/compressed
- Main cheez is column family
- Can be 3-4
- Too many mean more resources
- Column family kay baad column qualifiers
- Data is stored in a cell
- Get function in hbase gets the fresh version
- If we set it as a specific version phir woh ayega
- Hbase table is sparse we dont care about spaces in ebtween

1/dec/2025

- Follow request leader se ayegi
- Rights ko replicate karna followers mein will be done by leader itself
- Each tree ke andar sab tree is also a node
- So 5 z nodes
- It can have some metadata
- And it can have subtree child directories
- Zookeeper mein tree structure is balanced
- Watches: I cannot keep track of everything, so I put a watch on him
- Data node se name node to yarn (in map task when it completes) so yeh dono iss par watch yarn and watch name node lagadenge taake jese hi hojaye ping hojaye and they don't have to wait for it to be completed
- No waiting now
- Just give it a watch
- Jese hi z node ki entry ayegi woh receive hojayega
- Ab z node mein three types aagayi
- 1) persistent node
- 2) ephemeral node
- 3) sequential node
- Observers are non voting cant participate in voting
- Znode can be watched by clients
- Ordering kabhi kharab nahi hoti
- What is a lock
- Hadoop ecosystem mein kiss application ko lock ki zaroorat hogi?