

Hadoop \Rightarrow First framework which is capable of doing distributed computation.

\rightarrow HDFS (Hadoop file system)

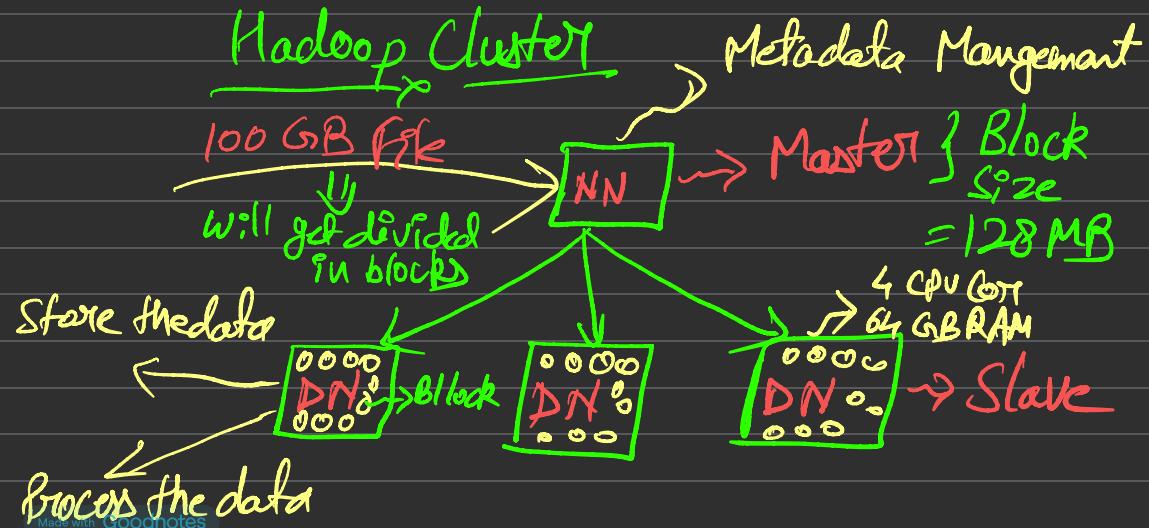
\rightarrow Map-Reduce (Computation)

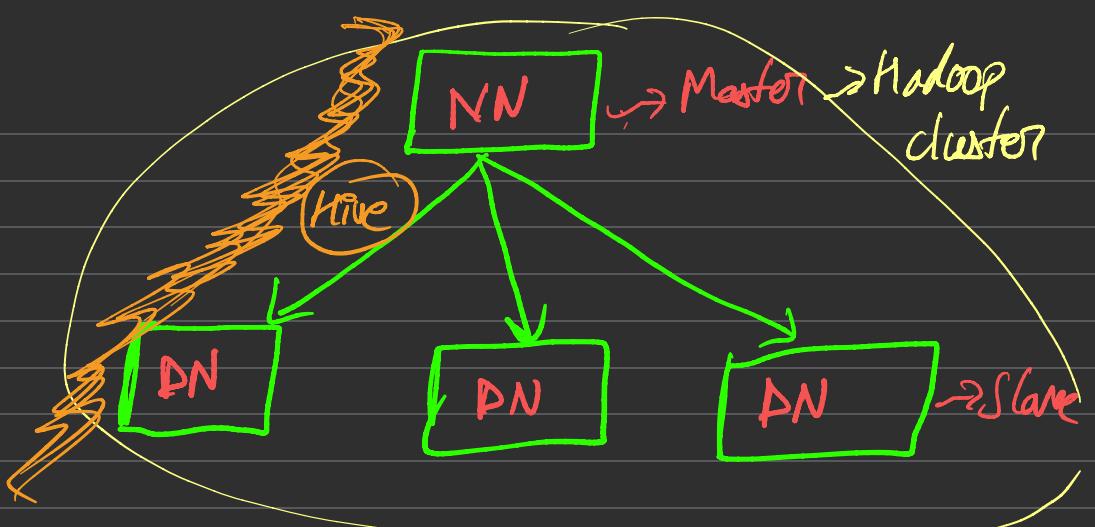
\rightarrow YARN (Resource Manager)



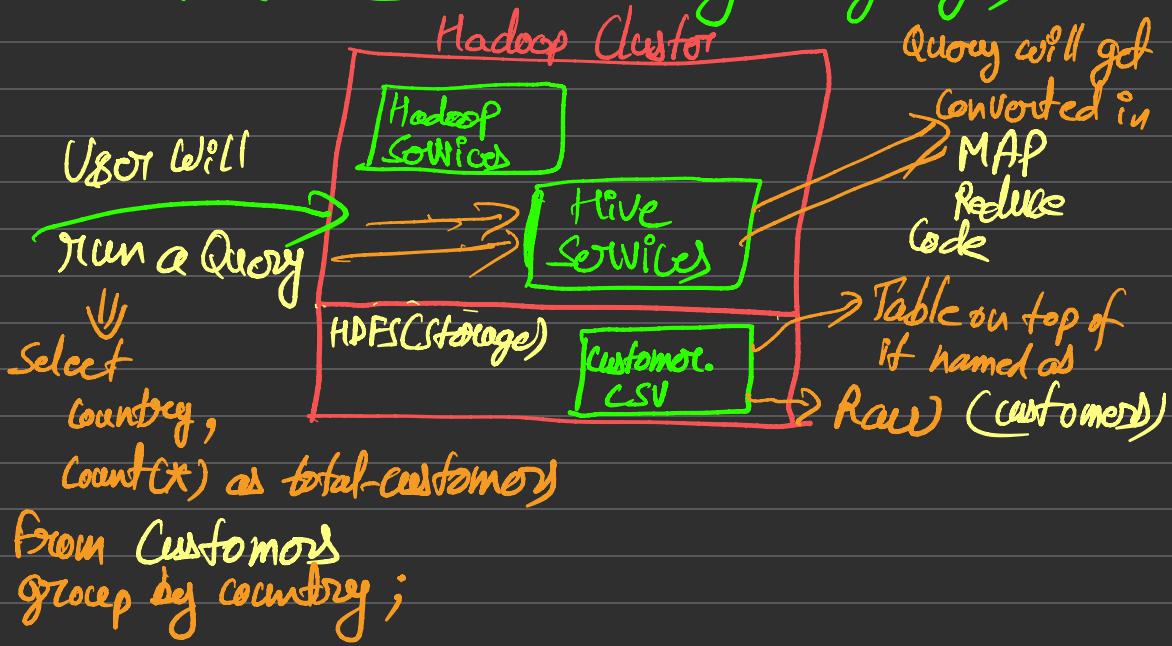
Hive \Rightarrow It is a data warehousing application framework built on top of Hadoop.

$$100\text{GB} = \frac{100 \times 1024 \text{MB}}{120 \text{MB}}$$



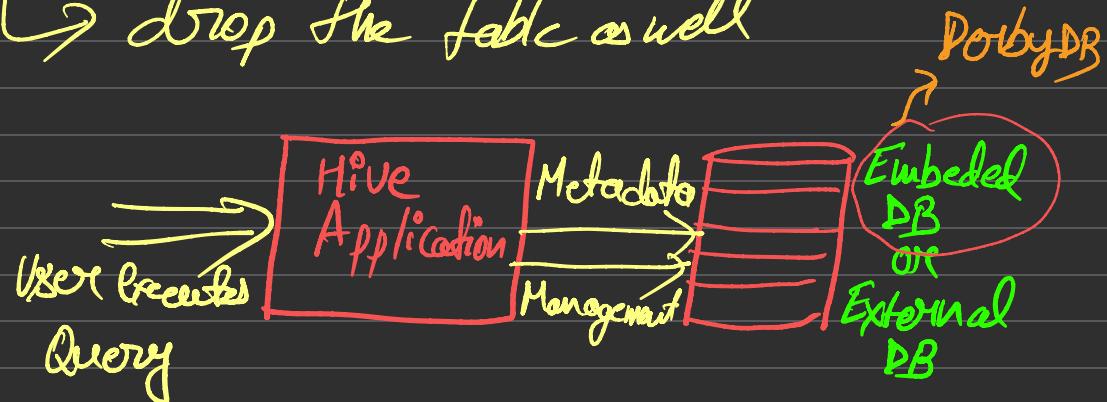


Hive is giving users the flexibility of querying the data with the help of HQL (Hive Query Language).



What we can do with HQL?

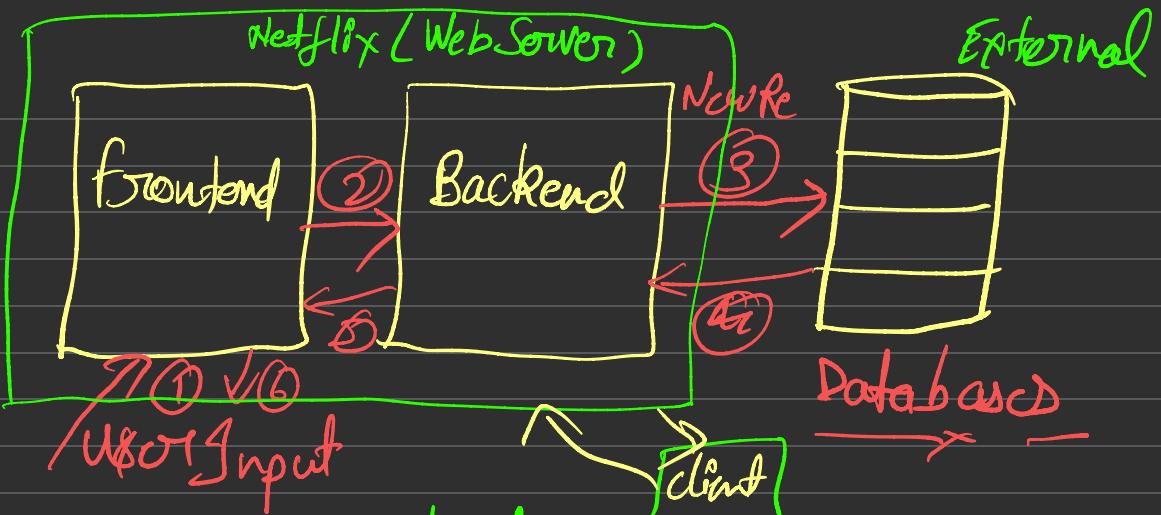
- Create database
- Create a table
- Load the data in table
- Select the data
- drop the table as well



Tightly Coupled



Any transaction
Database
(MySQL)
External DB



Make
a request

