



# Big Data Technologies

Trainer: Mr. Nilesh Ghule.



# Hive Functions → UDFs

## ① UDF (Scalar fns)

~~~~~  
~~~~~  
~~~~~ → UDF → ~~~~~  
~~~~~  
~~~~~

e.g. UPPER(),  
CONCAT(),  
ABS(), POW(),  
ROUND(), YEAR(),  
...

## ② UDAF (Aggregate fns)

1 ~~~~~  
2 ~~~~~  
1 ~~~~~ → UDAF → 1 ~~~~~  
2 ~~~~~ 2 ~~~~~  
1 ~~~~~

e.g. SUM(), AVG(),  
CORR(), ...

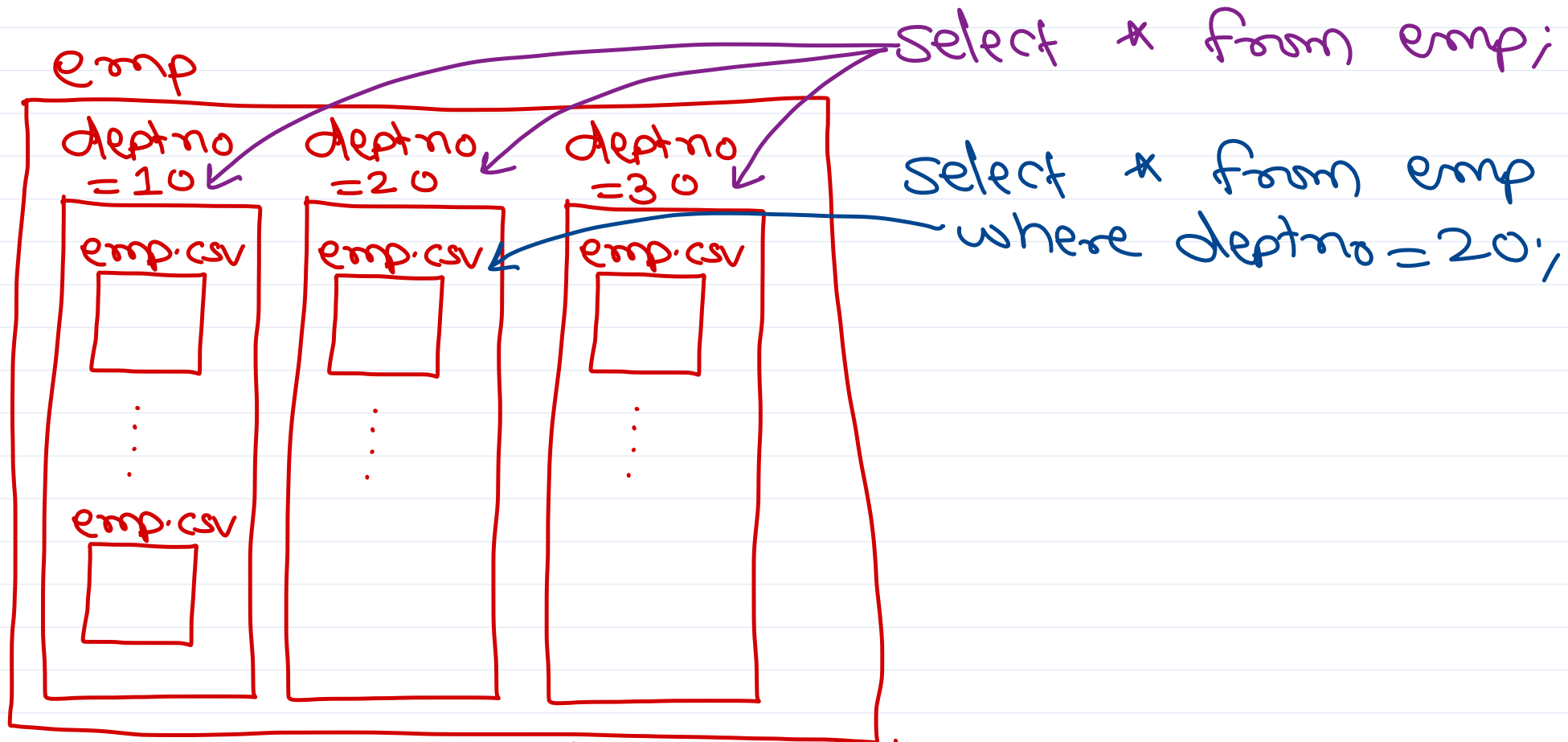
## ③ UDTF (Table fns)

1 ~~~~~  
2 ~~~~~ → UDTF → 1 ~~~~~  
1 ~~~~~  
1 ~~~~~  
2 ~~~~~  
2 ~~~~~

e.g. EXPLODE(), ...



# Partitioning

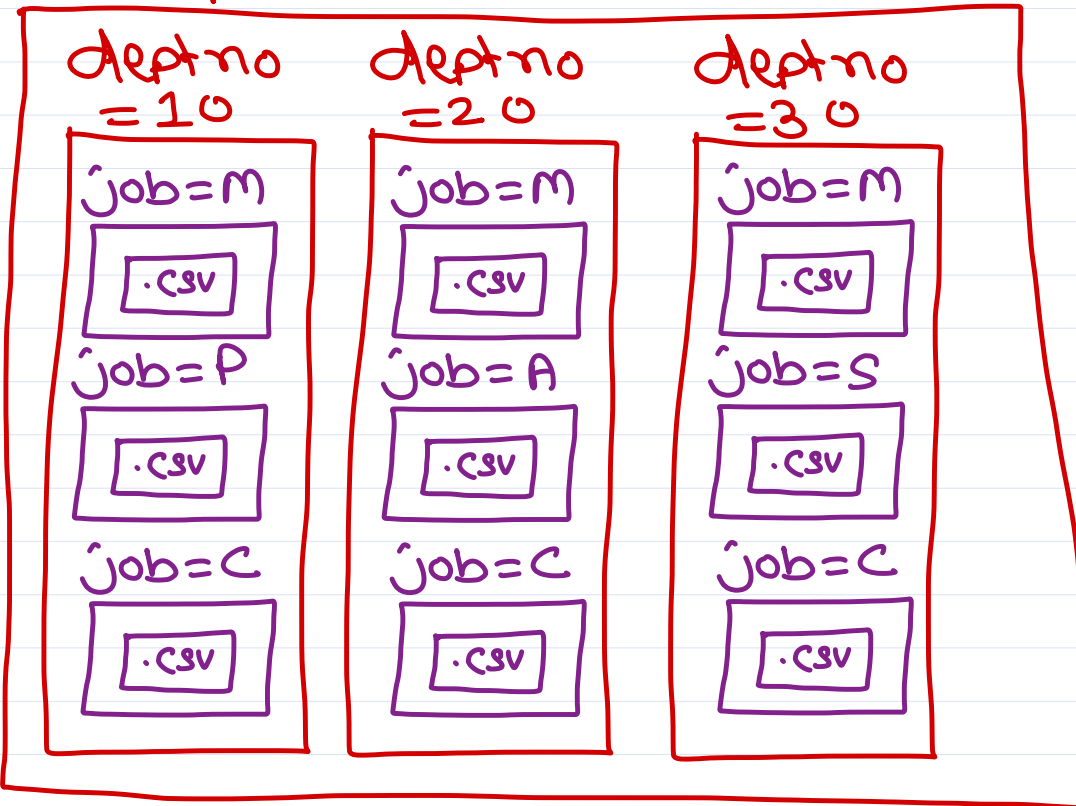


In Hive, when table is partitioned, data is stored in multiple subfolders in HDFS.



# Partitioning

emp

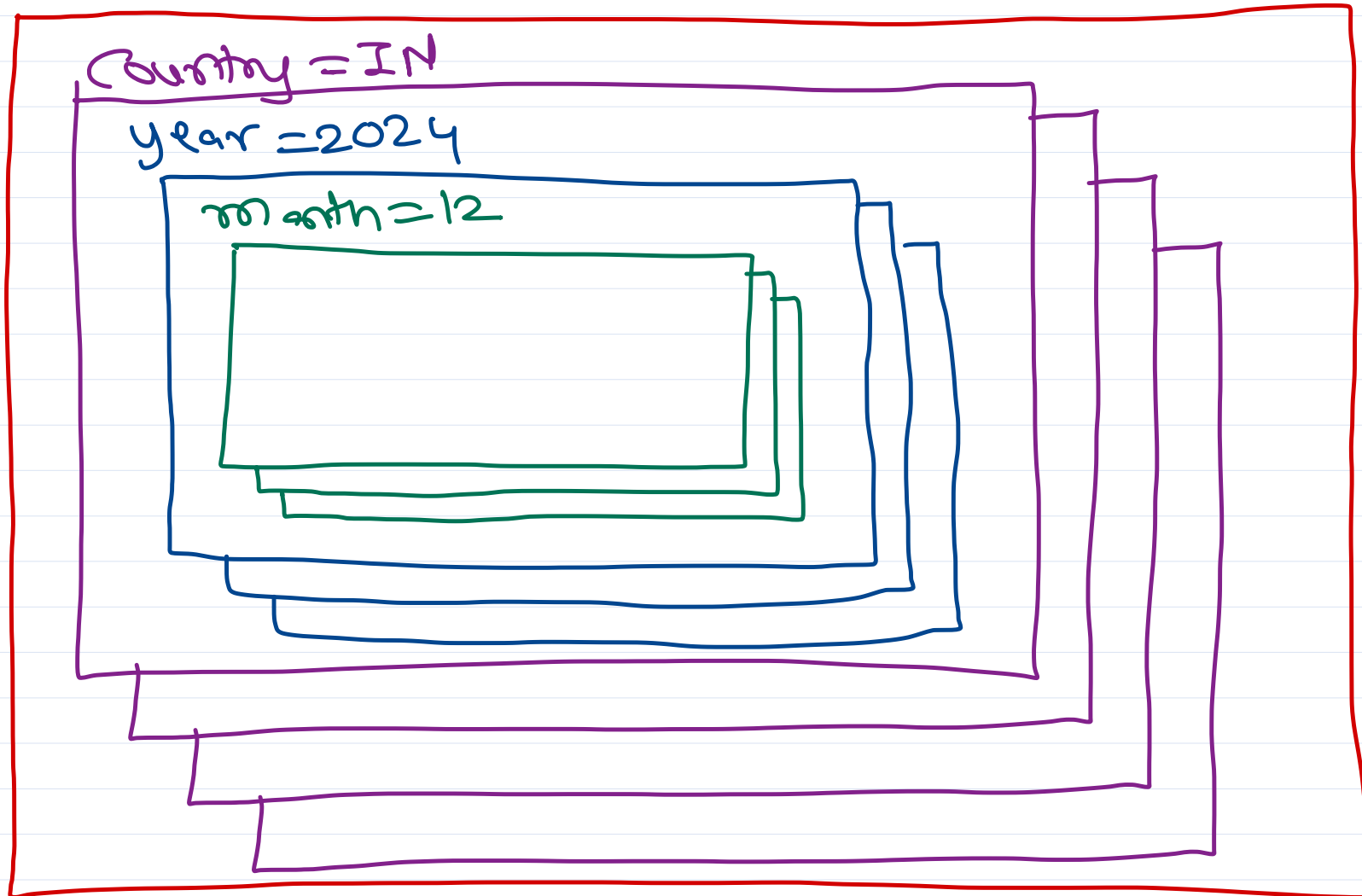


nested partitions



# Partitioning

Order





Thank you!

*Nilesh Ghule <nilesh@sunbeaminfo.com>*

