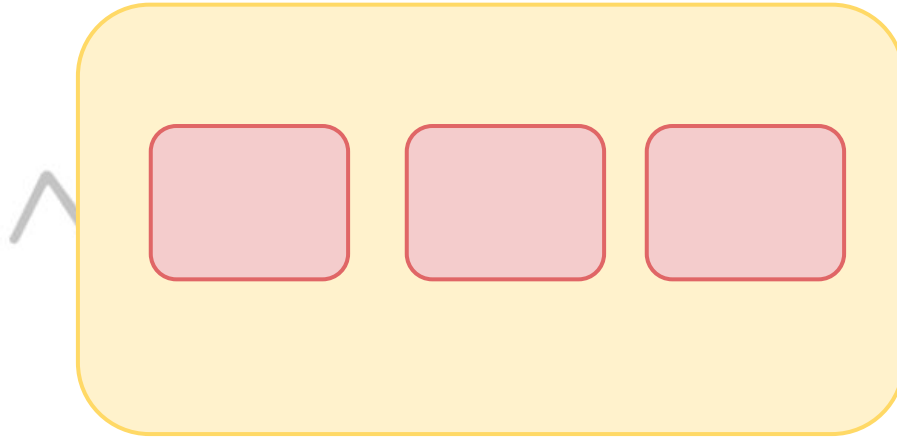


A large, faint watermark of the Analytics Vidhya logo is visible in the background, featuring a line graph icon and the text "Analytics Vidhya".

Bucketing in Hive

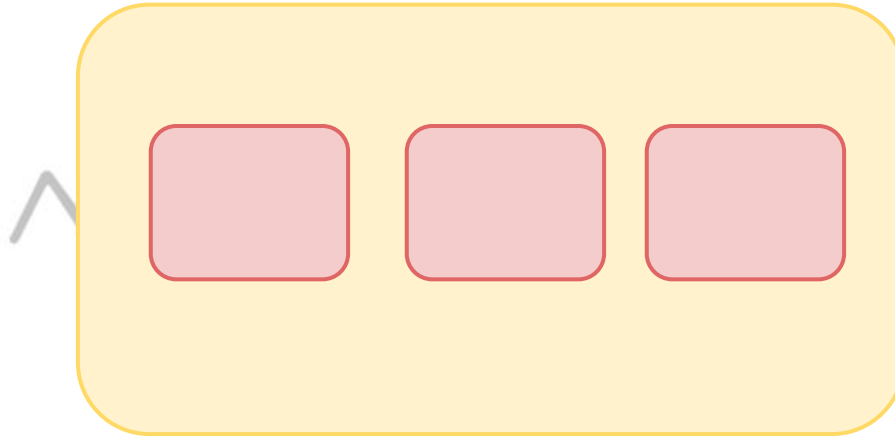
Partitioning in Hive



Partitioning in Hive



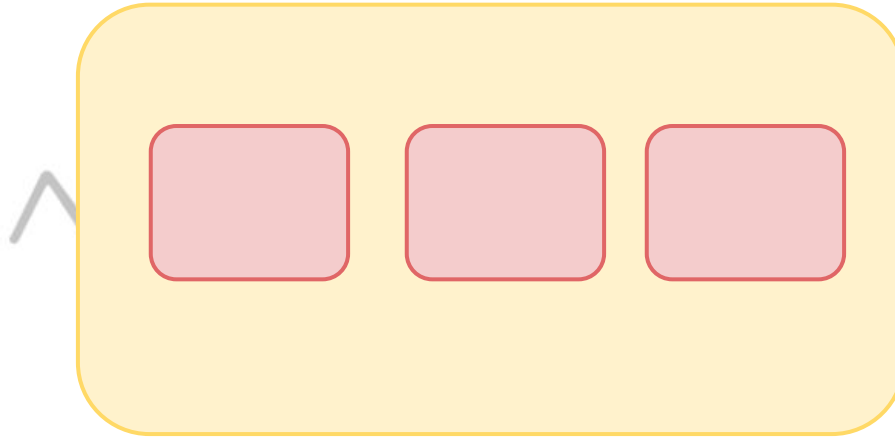
NameNode



Partitioning Problem



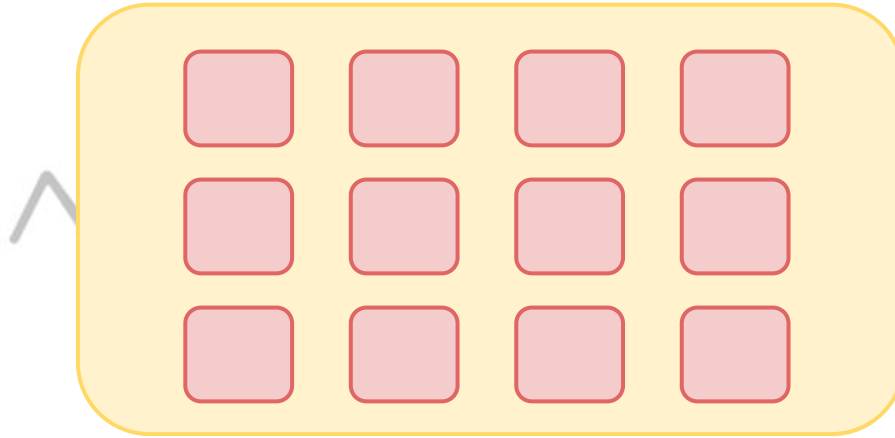
NameNode



Partitioning Problem



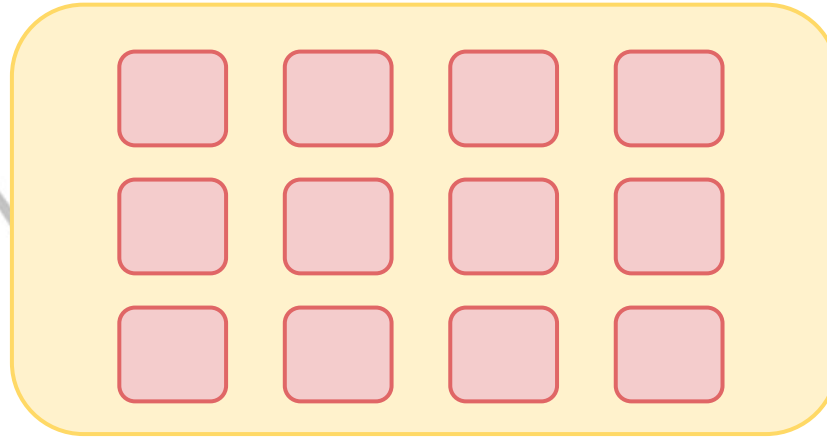
NameNode



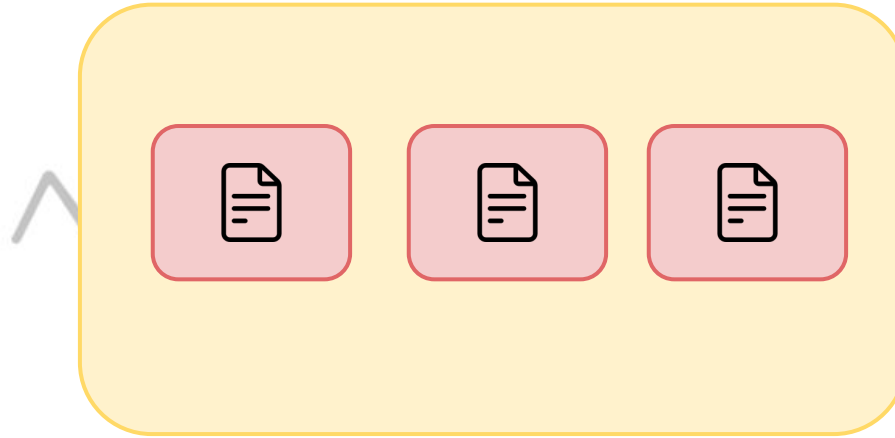
Partitioning Problem



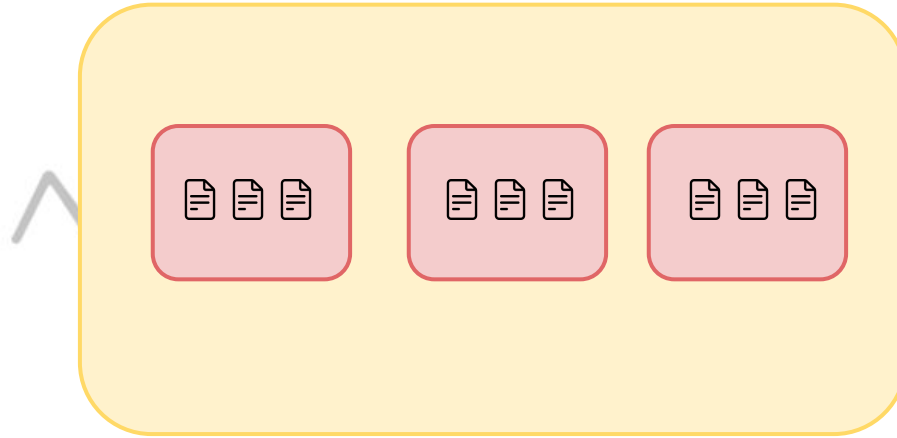
NameNode



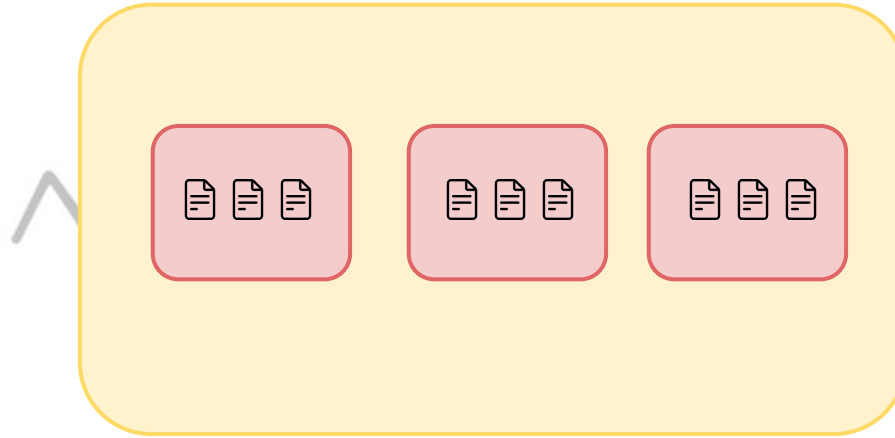
Partitioning Problem



Partitioning Problem



Bucketing in Hive



Bucketing in Hive

- Stores data into different segmented files or Buckets.
- Large number of distinct values on a column.
- Records with same value will be stored in same bucket.

Bucketing in Hive

hash_function (bucketing_column)

mod

number_of_buckets

Analytics
Vidhya

Bucketing in Hive

hash_function (bucketing_column)

mod

number_of_buckets

Patient_ID	Hospital	Department
P103	H002	D001
P334	H001	D001
P103	H003	D002

Bucketing in Hive

hash_function (bucketing_column)

mod

number_of_buckets

Patient_ID	Hospital	Department
P103	H002	D001
P334	H001	D001
P103	H003	D002



Bucketing in Hive

hash_function (bucketing_column)

mod

number_of_buckets

Patient_ID	Hospital	Department
P103	H002	D001
P334	H001	D001
P103	H003	D002



Bucketing in Hive

hash_function (bucketing_column)

mod

number_of_buckets

Patient_ID	Hospital	Department
P103	H002	D001
P334	H001	D001
P103	H003	D002



Bucketing in Hive

hash_function (bucketing_column)

mod

number_of_buckets

Patient_ID	Hospital	Department
P103	H002	D001
P334	H001	D001
P103	H003	D002



Bucketing in Hive

hash_function (bucketing_column)

mod

number_of_buckets

Patient_ID	Hospital	Department
P103	H002	D001
P334	H001	D001
P103	H003	D002



