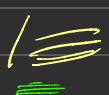




# Partitioning

Warehouse Directory  $\Rightarrow$  /user/hive/warehouse

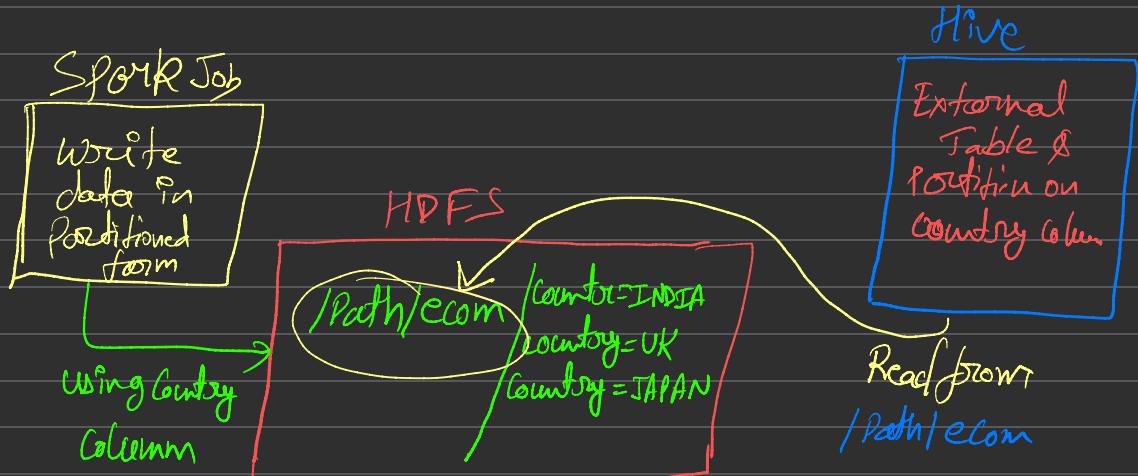
/user/hive/warehouse/hive-dbdbs/ecom/ 

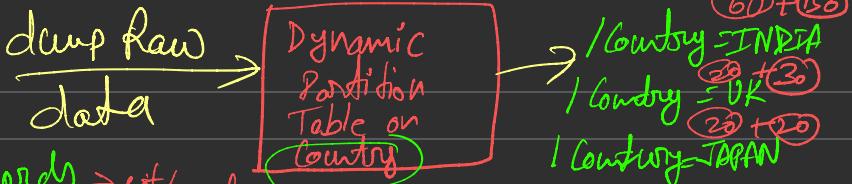
Path

- $\rightarrow$  ① /Path/country=INDIA/state=PUNJAB/city=MOHALI  
 $\rightarrow$  ② /path/country=INDIA/state=PUNJAB/city=AMRITSAR 

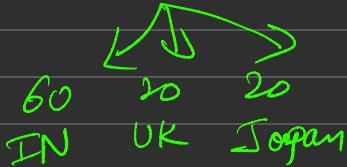
$\Rightarrow$  Internal Table

 External Table

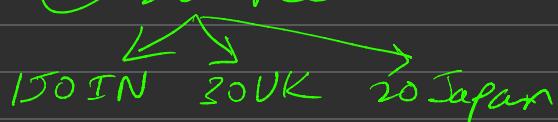




① 100 Records → 1<sup>st</sup> Load



② 200 Records → 2<sup>nd</sup> Load

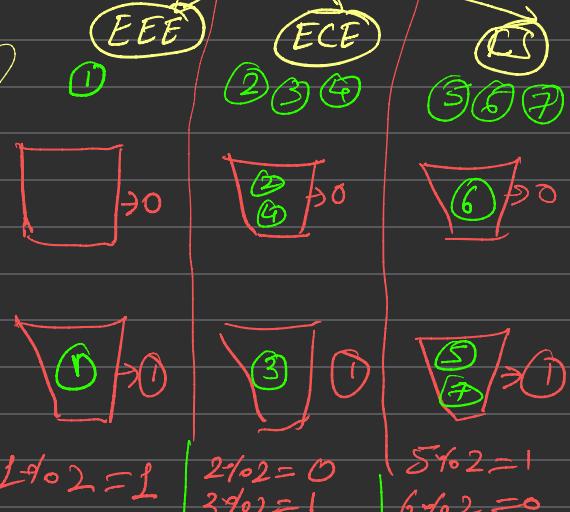


## Bucketing

Col  $stu\_id$  for bucketing (2)

	stu_id	name	dept
1	1	A	EEE
2	1	B	ECE
3	1	C	ECE
4	1	D	ECE
5	1	E	CS
6	1	F	CS
7	1	G	CS

## Partition (dept)



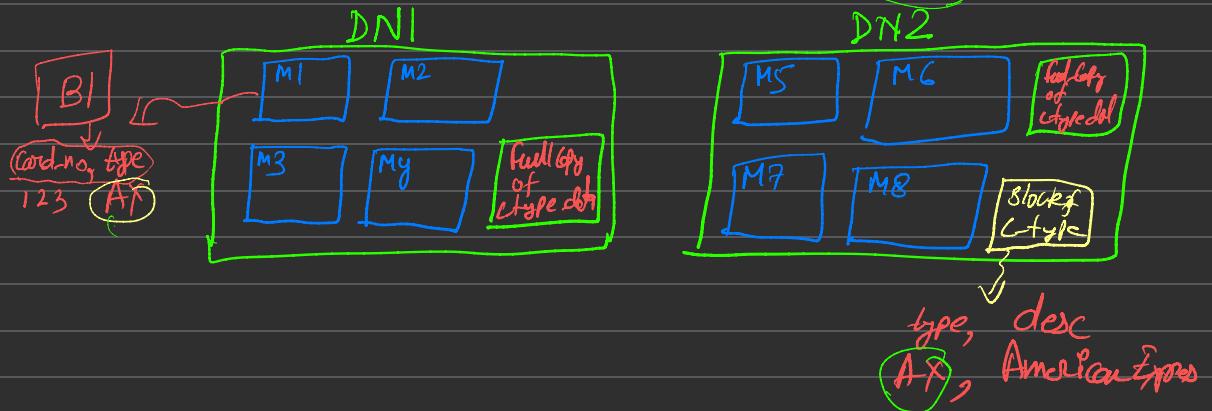
$$6 \times 2 = 0$$

$f(n) \Rightarrow n \text{ is student-id}$   
 $f(1) \Rightarrow 1\%(\text{number of buckets})$

# Map Side Join

(8 Blocks) Credit-card-records (large) 1 GB

(1 Block) Credit-card-type (small) 50 MB



## emp-table

emp\_id, name, dept\_id

100	A	1
200	B	2
300	C	1
400	D	2
500	E	3
600	F	2
700	G	3
800	H	1
900	I	1
1000	J	4

## dept-table

dept\_id, total-empl, name

1	50	CS
2	30	IT
3	20	HR
4	90	Finan.

↑ Buckets

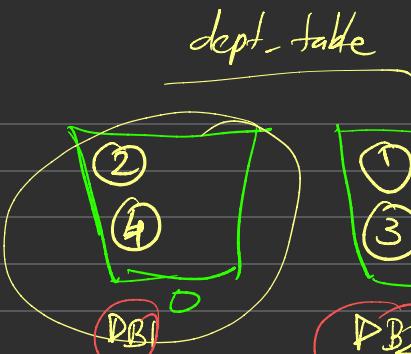
→ 2, 3  
→ 3, 6  
→ 6, 1

2 Buckets

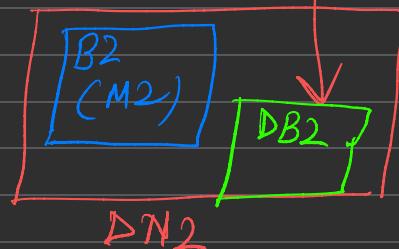
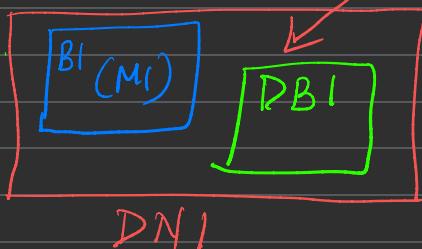
Emp-Tbl

200, 2
400, 2
600, 2
1000, 4
5

100, 1
300, 1
500, 3
700, 3
900, 2



$$1 \neq 2 = 1$$



# Sorted Merge Join

$l1 = [1, 2, 5, 9, 12] \times n$

$l2 = [3, 4, 7, 11, 14, 18, 19] \times m$

$l3 = [1, 2, 3, 4, 5, 7, 9, 11, 12, 14, 18, 19]$

$O(n \log n)$

$l3 = [1, 2, 3, 4, 5, 7, 9, 11, 12, 14, 18, 19]$

$O(n+m) = O(n+m)$





