29 September 2024 10:41 AM

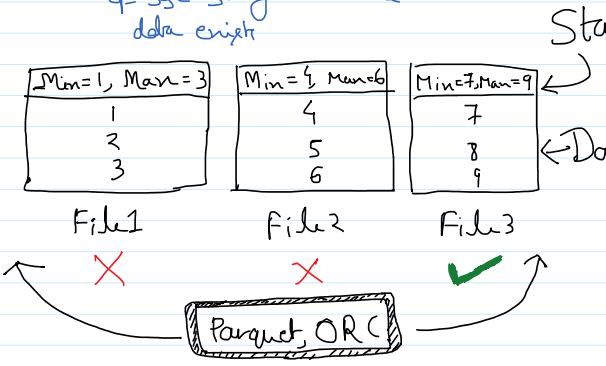
:- Predicate Pushdown :-

SQL Query

Push down specific section of SQL query where the data exists

```
select * from table1 where col1 = 9
```

① Huge Data Transfer via network
② Long Loading Time in memory



:- Z-Order Optimization :-

File1 → id = 5, 8, 9, 20 (min = 5, max = 20)

File2 → id = 10, 15, 30, 70 (min = 10, max = 70)

File3 → id = 2, 6, 30, 45 (min = 2, max = 45)

File4 → id = 10, 12, 18, 19 (min = 10, max = 19)

→ optimize → (Bin-packing)

select * from table where id = 12

File1 → id = 5, 8, 9, 20 (min = 5, max = 20)

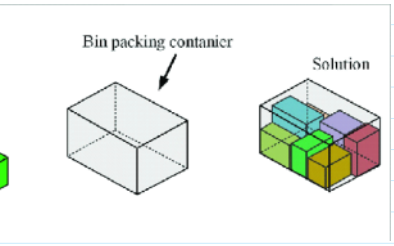
File2 → id = 10, 15, 30, 70 (min = 10, max = 70)

File3 → id = 2, 6, 30, 45 (min = 2, max = 45)

File4 → id = 10, 12, 18, 19 (min = 10, max = 19)

→ optimize → (z-ordering)

select * from table where



File1 + File2

id = 5, 8, 9, 20
10, 15, 30, 70
(min = 5, max = 70)

File3 + File4

= 3, 6, 30, 45
10, 12, 18, 19
min = 3, max = 45)

id = 3, 5, 6, 8, 9, 10, 10, 12
(min = 3, max = 12)

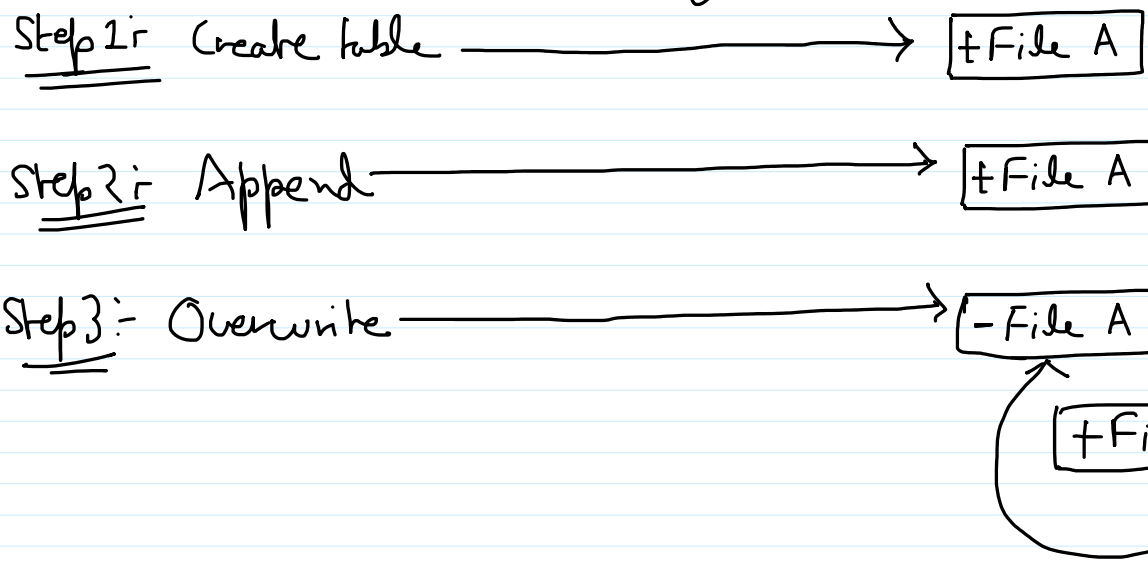
→

id = 15, 19, 19, 20, 30, 30, 45, 70
(min = 15, max = 70)

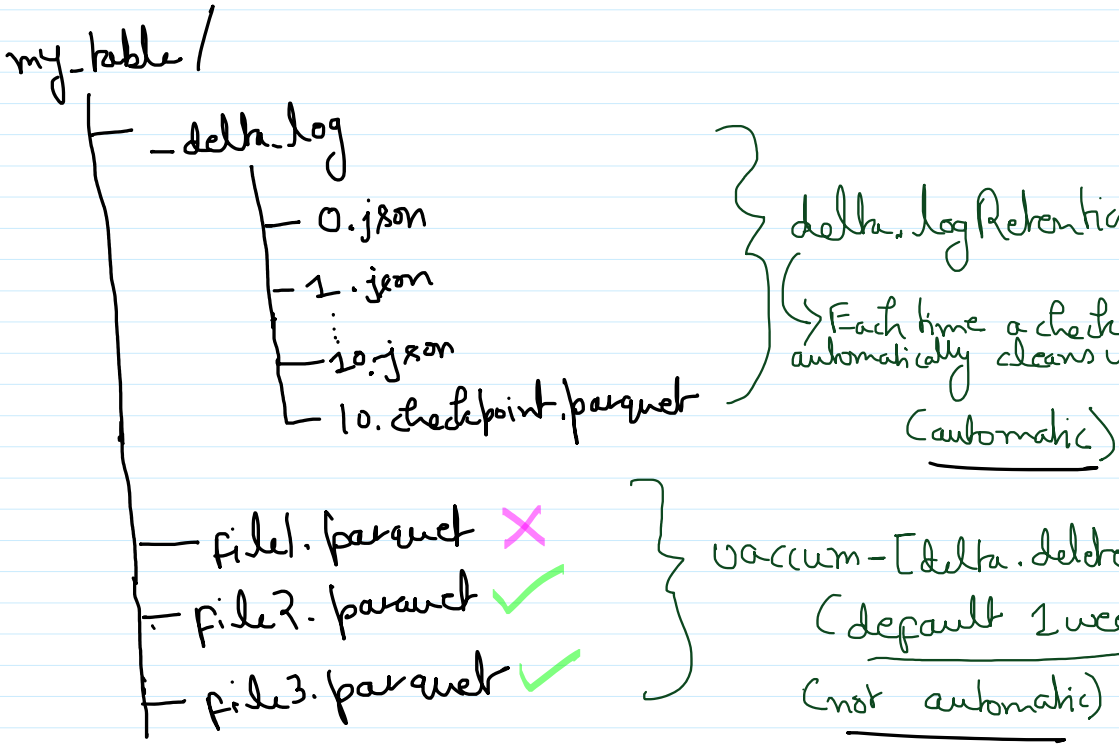
id = 12

So overall, Z-ordering helps your queries run faster because it makes it that data can be skipped.

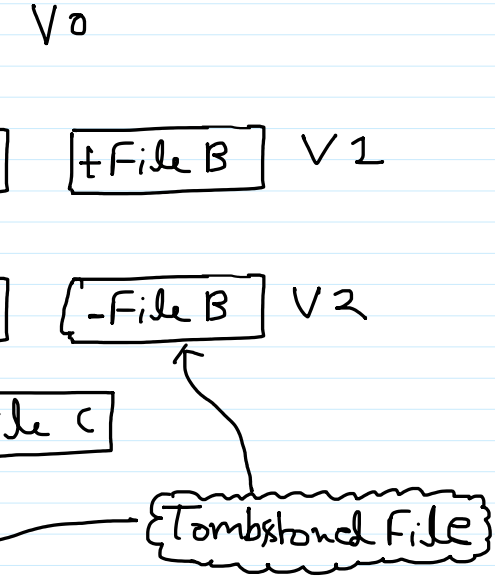
:Vacuuming:



:Deleting log files:



r more likely



nDuration (interval 30 days)

point is written Databricks
up log entries older than
logRetentionDuration.

d.FileRetentionDuration]
eth)

pack time-travel

{delta.deletedFileRetentionDuration} => Don't my

⊙→ If I need to perform time-travel upto

pace time-travel

20 days ?