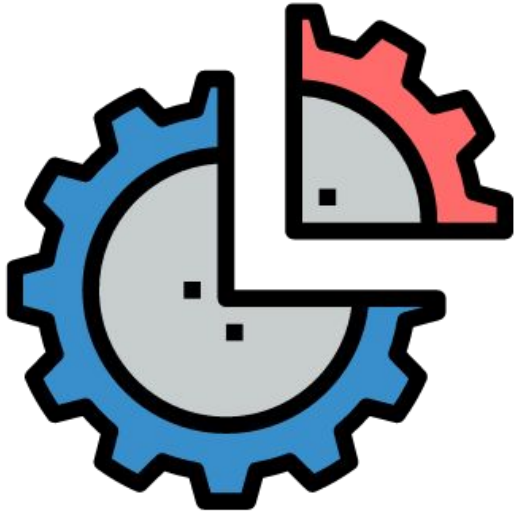

Partitioning

husseinnasser.com

Database Partitioning Explained

Agenda

- What is Partitioning?
- Horizontal Partitioning vs Vertical Partitioning
- Partitioning Types
- Partitioning vs Sharding
- Demo
- Pros & Cons
- Summary

What is Partitioning?

```
SELECT Name
FROM CUSTOMERS_TABLE
WHERE ID = 700,001
```

CUSTOMERS Table

id	NAME
1	Ali
2	Mike
3	Edmond
..	….
...	...
700,001	Kim
700,002	Ali
..	...
1M	James

What is Partitioning?

```
SELECT Name
FROM CUSTOMERS_TABLE
WHERE ID = 700,001
```

Which “partition” is customer 700,001 in?

Partition CUSTOMERS_800K

Split 1 million rows table into 5 tables called partitions.. Same schema

CUSTOMERS_200K

id	NAME
1	Ali
...	..
200,000	Edmond

CUSTOMERS_400K

id	NAME
200,001	James
...	..
400,000	Smith

CUSTOMERS_600K

id	NAME
400,001	Nasser
...	..
600,000	Karen

CUSTOMERS_800K

id	NAME
600,001	Nada
...	..
700,001	Kim
...	...
800,000	Tyler

CUSTOMERS_1M

id	NAME
800,001	Paul
...	..
1,000,000	Rick

Vertical vs Horizontal Partitioning

- Horizontal Partitioning splits rows into partitions
  - Range or list
- Vertical partitioning splits columns partitions
  - Large column (blob) that you can store in a slow access drive in its own tablespace

Partitioning Types

- By Range
  - Dates, ids (e.g. by logdate or customerid from to)
- By List
  - Discrete values (e.g. states CA, AL, etc.) or zip codes
- By Hash
  - Hash functions (consistent hashing)

Horizontal Partitioning vs Sharding

- HP splits big table into multiple tables in the same database, client is agnostic
- Sharding splits big table into multiple tables across multiple database servers
- HP table name changes (or schema)
- Sharding everything is the same but server changes

Demo - Example with Postgres

- Spin up a postgres instance
- create a table and Insert 10 million rows
- Create partitions

Pros of Partitioning

- Improves query performance when accessing a single partition
- Sequential scan vs scattered index scan
- Easy bulk loading (attach partition)
- Archive old data that are barely accessed into cheap storage

Cons of Partitioning

- Updates that move rows from a partition to another (slow or fail sometimes)
- Inefficient queries could accidently scan all partitions resulting in slower performance
- Schema changes can be challenging (DBMS could manage it though)

Summary

- What is Partitioning?
- Horizontal Partitioning vs Vertical Partitioning
- Partitioning Types
- Partitioning vs Sharding
- Pros & Cons