



# Sharding



Sharding & Consistent Hashing

# Agenda

- What is sharding?
- Consistent Hashing
- Horizontal Partitioning vs Sharding
- Example (Code with Postgres)
- Pros & Cons
- Summary

# What is Sharding?



```
SELECT URL  
FROM URL_TABLE  
WHERE URLID = "5FTOJ"
```



URL shortener table with 1 million pages

## URL\_TABLE

id	Url	urlid*
1	<a href="https://www.canva.com/design/DADrSCuKq4I/5sKekxVdctoGGq7Ri9O5GQ/edit">https://www.canva.com/design/DADrSCuKq4I/5sKekxVdctoGGq7Ri9O5GQ/edit</a>	5FTOJ
2	<a href="https://en.wikipedia.org/wiki/Shard_(database_architecture)#Database_architecture">https://en.wikipedia.org/wiki/Shard_(database_architecture)#Database_architecture</a>	CeG0z
..	....	....
..	....	....
1M	<a href="https://www.quora.com/How-does-base64-encoding-work">https://www.quora.com/How-does-base64-encoding-work</a>	J9COp

# What is Sharding?



Which database server is **5FTQJ** in?  
**Server 3!**

```
SELECT URL  
FROM URL TABLE  
WHERE URLID = '5FTQJ'
```



Split 1 million rows  
table into 5 database  
instances.. Same  
schema

# Consistent Hashing



postgres:5432



postgres:5434

Hash("Input 1")  
Hash("Input 2")  
5432  
5433



postgres:5433

# Consistent Hashing



postgres:5432



postgres:5434

Hash("Input3")  
5433



postgres:5433

# Consistent Hashing



postgres:5432



postgres:5434

num("Input2") % 3

1

+5432

=5433



postgres:5433

# Horizontal Partitioning vs Sharding

- HP splits big table into multiple tables in the same database
- 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



# Example Code with Postgres (Url shortener)

- Spin up 3 postgres instances with identical schema
  - 5432, 5433, 5434
- Write to the sharded databases.
- Reads from the sharded databases.

# Pros of Sharding

- Scalability
  - Data
  - Memory
- Security (users can access certain shards)
- Optimal and Smaller index size

# Cons of Sharding

- Complex client (aware of the shard)
- Transactions across shards problem
- Rollbacks
- Schema changes are hard
- Joins
- Has to be something you know in the query

# Summary

- What is sharding?
- Consistent Hashing
- Horizontal Partitioning vs Sharding
- Example (Code with Postgres)
- Pros & Cons