

Range sharding

🧠 Range Sharding কী?

👉 Shard key-এর value range অনুযায়ী data ভাগ করা

Shard-1 → 1 – 1000

Shard-2 → 1001 – 2000

Shard-3 → 2001 – 3000

📌 Shard selection:

if key \in range → that shard

1. Basic Architecture

Client

↓

App / Router

↓

Range Map

↓

Shard-1 | Shard-2 | Shard-3

Range Map বলে দেয়:

1001–2000 → Shard-2

2. Example (Very Clear)

Use case: Orders table

Shard key = order_id

Shard-1 → order_id 1 – 1M

Shard-2 → order_id 1M – 2M

Shard-3 → order_id 2M – 3M

3. Why Use Range Sharding?

✓ Advantages

- **Efficient range queries**
- Easy data locality
- Simple to understand

📌 Interview line:

“Range sharding is ideal for range-based queries.

✗ Disadvantages (IMPORTANT)

- Hot shard problem
- Uneven data distribution
- Manual rebalancing needed

📌 Example:

Latest orders → only last shard

4. Range Query Performance (Big Win)

Query:

orders between 1.5M – 1.6M

✓ Only Shard-2 scanned

✗ Not all shards

5. Re-Sharding (Easier than Hash)

When shard-3 full:

Shard-3 → split into Shard-3 & Shard-4

✓ Only part of data moves

📌 Interview line:

“Range sharding allows easier shard splitting.”

6. Hot Shard Problem (🔥 Interview Favorite)

Problem:

- Recent data heavily accessed
- One shard overloaded

Solutions:

- Add random suffix
- Time + hash hybrid
- Pre-splitting ranges

7. Range vs Hash (DEEP COMPARISON)

Feature	Range Sharding	Hash Sharding
Distribution	Uneven	Even
Range query	✅ Excellent	❌ Poor
Hot shard	High risk	Low
Re-sharding	Easier	Hard
Use case	Time-based data	User-based data

8. Real-World Usage

Databases

- HBase
- Bigtable
- MySQL sharded by date

Systems

- Logs
- Metrics
- Orders by date

9. Hybrid Sharding (Advanced)

👉 Combine **Range + Hash**

Example:

Shard by date range

Inside shard → hash userId

📌 Used by Google, Meta

🏆 FAANG Interview Answer 💎

Say this:

“Range sharding partitions data based on key ranges, making it efficient for range queries and easier to rebalance. However, it risks hot shards and uneven load, so it’s often combined with hashing for better distribution.”

🧠 One-Line Memory Hook

Range sharding = fast range queries, hot shard risk