

Yukun Ma
Week #9: Sharding
Page #1

Problem Statement

In this paper, I will show what “sharding” in a database is and how it works. Apart from that, the benefits and challenges the sharding brings to databases will also be evaluated.

References

- [1] Wikipedia. Shard (database architecture) - wikipedia. [https://en.wikipedia.org/wiki/Shard_\(database_architecture\)](https://en.wikipedia.org/wiki/Shard_(database_architecture)). Accessed April 26, 2018.
- [2] Wikipedia. Partition (database) - wikipedia. [https://en.wikipedia.org/wiki/Partition_\(database\)](https://en.wikipedia.org/wiki/Partition_(database)). Accessed April 26, 2018.

Overview of Sharding

Normally, a database shard refers to a horizontal partition of data in a database.[1] And a horizontal partition of a database involves putting different rows in an original table into different tables. Each individual partition of the splitted database is called a shard of the database.

For example, there are a lot of teachers in a university, and the table that is storing the teachers are too large to handle. Then we can use sharding to split the big table into some smaller tables. Specifically, we can split the tables according to different departments that teachers are in. First partition table stores the information of the teachers in Department of Computer Science and Engineering, the second partition table stores the information of the teachers in Department of Software Engineering, and so on.

Critical Thinking

There are many advantages to sharding.

1. After sharding, the number of rows in each table is reduced.
2. Sharding may reduce the cost of each query sometimes for some reason¹.
3. Sharding enables a distribution of the database over multiple machines.

But sharding also brings some challenges.

1. Sharding relies much on the network connections between servers much.
2. Sharding may increase the delay of some query on the database.
3. Sharding adds much complexity on the database.

¹Considering you want to calculate the number of average salary of teachers in some department.

Question

If there are horizontal partitions of data which are also known as shards. Then there should be vertical horizontal of data. If there are vertical partitions, then what are vertical partitions of data in a database? And what is the use of it?

Method

Describe how you are going to answer your own question stated above.

Analysis and Discussion

Vertical partitioning is the process of creating tables with fewer columns.[2] It is kind of like horizontal partitioning. But horizontal partitioning splits the table by rows, while vertical partitioning splits the table by columns.

Table 1: The original table

ID	Name	Birthday
1	Dave	1997-02-24
2	Yanmei	1995-05-23
3	Alice	1996-09-10

Table 2: First partition table

ID	Name
1	Dave
2	Yanmei
3	Alice

Table 3: Second partition table

ID	Birthday
1	1997-02-24
2	1995-05-23
3	1996-09-10

Let's suppose we have a table which has 3 people and their birthdays in it as Table 1 shows. If we split the table by vertical partitioning, the original table may results in two partition tables as Table 2 and Table 3 show.

What is the use of vertical partitioning? When we only need some information in most queries, sometimes it's great to split the table vertically regard to the consideration of efficiency. Also, if there are many rows missing some attributes in the table, it may reduce the size of the database after vertical partitioning. In additional, there is an important technique called database normalization which involves vertical partitioning.