TDT4225 – Assignment 4

Filip F Egge

November 12, 2015

Algebra with filter

Joining of two tables A and B, $V_A = 400\ B \times 1\ 000\ 000 = 4 \times 10^8 = 400\ MB$, $V_B = 600B \times 10000000 = 6 \times 10^9 = 6000MB$, $V_R = 600B \times 1000000 = 6 \times 10^8 = 600\ MB$. Our workspace is 20 MB, this gives $n = \lceil \frac{V_A}{M} \rceil = \lceil \frac{400\ MB}{20\ MB} \rceil = 20$. $V_{nl}^J = V_A + nV_B + V_R = 400\ MB + 20 \times 6000\ MB + 600\ MB = 121\ GB$

Join with filter.

$$V = 2V_A - M + V_B + V_B \delta_{F_A} \left(\left\lceil \frac{(V_A - M)\delta_{F_{A \cap B}}}{M} \right\rceil \right) + V_R$$

I did not manage to calculate this join with filter.

Parallel algebra

Describe the different partitioning methods used in parallel algebra.

Horizontal fragmentation divides the table into fragment, each fragment corresponds to a group of records stored on a node. Vertical fragmentation has records with key and one or a few attributes for each fragment. The third option is a mix between the two others.

Why is hashing a very good method?

Using hashing is a good way to place records, this requires to indexes and a good hash formula.

Dynamo

Explain the following concepts/techniques used in Dynamo

consistent hashing

Consistent hashing uses a hashing function whose output is represented as a fixed circular space. Each node is assigned a space on this ring, and each item is assigned to a node by hashing the items key. Each node is thus responsible for keeping track of the region between itself and the previous node.

vector clocks

Dynamo uses vector clocks in order to capture causality between different versions of the same object.

sloppy quorum and hinted handoff

Sloppy quorum is a less strict version of the traditional quorum. All read and write operations are performed on the first N healthy nodes from the preference list.

Hinted handoff ensures that read and write operations are not failed due to temporary node or network failures.

merkle trees

Dynamo uses merkle trees to detect inconsistency between replicas. A merkle tree is a hash tree where leaves are hashes of the values of individual keys.

gosip-based membership protocol

Dynamo uses a completely decentralized membership protocol, and updates are spread using gosip or word of mouth.

RamCloud

How does RamClouds ensure ”durability” of data?

RamClouds can use different ways to ensure durability of data, the most common is to ”buffered logging” which uses both disk and memory for backup. A local copy is stored in the primary server DRAM and copies stored in two or more backup servers. On write the primary send a log entry to the backup server who updates its copy.

How does Ousterhout argue that RAMCloud's potential to support ACID transactions is better than for traditional disk-based distributed databases?

Ousterhout argues that RamClouds low latency and fast transactions limits the use for ACID. Since ensuring ACID scales poorly and adds time delays to transactions, a trend in storage systems is to give up some ACID properties to improve scalability.

Facebook TAO

How does Facebook TAO solve the problem that the social graph spans the whole world, and that the data should be close to the user?

Data is divided into logical *shards*, each of which is stored in a logical database. Database servers are responsible for one or more shards. This enables Facebook to place shards at physical servers close to the places the shards relate to.

Google Spanner

How are TimeStamps used in Spanner's transactions?

Timestamps in Google Spanner is used in versioning, where each version of data is timestamped at commit time. Transactional reads and writes use two-phase locking. Transaction can be assigned timestamps when all locks have been acquired, but before any has been released.