Prof. Dr. Stefan Brass
Institut für Informatik
MLU Halle-Wittenberg

December 05, 2019

Databases IIB: DBMS-Implementation — Exercise Sheet 7 —

Please read Part a) and mark questions which you want to discuss in class. You only have to submit Part d) to g). Please upload your solution into the StudIP file folder called "Hausaufgabe_7" in the StudIP entry of the lecture. The deadline is December 10 (the day before the next lecture).

It is permitted to form groups of up two members, but please make sure that both members can fully explain all homeworks submitted by the group. Please upload only one file per group. Name the file such that it contains the names of all group members.

Not all submitted homeworks will be corrected, but all homework exercises will be discussed in class. If you should have questions about your homework, please ask! A precondition for getting credit for this course is that you submit solutions to two thirds of the homeworks. Obviously wrong or very incomplete submissions do not count.

Repetition Questions

a) What would you answer to the following questions in an oral exam?

- What is a typical seek time of a disk? What operation does the seek time measure?
- What is a typical latency time? What does "latency time" mean (in connection with magnetic disks)? On what other performance measure does the latency time depend?
- How long does it take in total to read or write a single block from/to disk? (Of course, this depends on the disk. Name any reasonable value for a standard magnetic hard disk.) What is/are the most important components of the total time?
- How much faster can one write blocks of e.g. 4KB sequentially vs. randomly distributed on the disk?
- Name any interface that is used to attach disks to a computer.
- What is the purpose of RAID systems? Do you know what "RAID" stands for (there are two correct solutions)?
- Explain RAID Level 0 (Striping). What is the performance of a system with four disks compared to the performance of a single disk? What is the big problem of this RAID level?

- Explain RAID Level 1 (Mirroring). What is the performance of a system with two disks compared to the performance of a single disk? What is the disadvantage of this level?
- What is RAID Level 10?
- Explain RAID Level 5 (Striping with parity information distributed over the disks). What happens if a disk is faulty? Discuss the performance of a system with four disks compared to a single disk.
- Why does one not use a very large number of disks (say, 500) with a single parity disk?
- How can you specify in Oracle on which disk(s) a table is stored?
- What is the syntax for specifying a tablespace in the `CREATE TABLE` statement in Oracle?
- How can one avoid fragmentation of the data files on disk? (In the optimal case, each data file consists of consecutive disk blocks.)

In-Class Exercises

b) In this exercise, we use SQL*Plus, the standard command-line interface to an Oracle database. More information can be found in the "SQL*Plus User's Guide and Reference":

[https://docs.oracle.com/en/database/oracle/oracle-database/18/sqpug/]

The "SQL*Plus Quick Reference" is also useful:

[https://docs.oracle.com/cd/B28359_01/server.111/b31190/toc.htm]

Please run the following command in a terminal window to set environment variables:

```
source /home/brass/ora_env
```

Afterwards, call `sqlplus` with your DBA-account in the same window:

```
sqlplus brass_dba@pdb1
```

Then enter:

```
SET AUTOTRACE ON STATISTICS
```

Run a query. The output for "`physical reads`" shows the number of blocks actually read from disk. The output for "`db block gets`" and "`consistent gets`" shows the number of blocks accessed in total (using also blocks that are already cached in memory). The difference between both values is that "`db block gets`" are accesses to the current version of a block, whereas "`consistent gets`" are accesses to the version that contains only committed changes when the current query started. Note that also accesses to the data dictionary (including information needed by the query optimizer) count. If you immediately rerun the same query (using the cached execution plan), the numbers will be much smaller. You find more information about `AUTOTRACE` in Chapter 8 (Part II) of the SQL*Plus Manual.

c) Open three terminal windows (use the one from the previous exercise and two more, the script `ora_env` must be executed in each of them).

- In one window, log into SQL*Plus as a normal user, and start a transaction, e.g. update a row. Make sure that you do not have `AUTOCOMMIT` mode (then SQL*Plus would send a `COMMIT` to the server after any update). E.g. you can enter (inside SQL*Plus):

  ```
  show autocommit
  ```

  If necessary, enter

  ```
  set autocommit off
  ```

- If you never tried what happens when you want to access a row that is locked by somebody else, this would be a good time. Log into SQL*Plus in the second window as the same user, and first write a query that shows the updated row. You will notice that it appears unchanged as long as you do not enter `COMMIT;` in the first window. Then enter an `UPDATE` statement for the same row. It will appear that the system "hangs", i.e. there is no reaction until you enter `COMMIT;` or `ROLLBACK;` in the first window and thereby release the lock. Then the update will proceed as normal. However, please create a situation where one session waits for a lock hold by the other session.

- In the third window, log in as administrator and use the tables `V$TRANSACTION`, `V$SESSION`, `V$TRANSACTION_ENQUEUE`, and `V$LOCK` to find out (1) who is currently logged in, (2) who has an active transaction, and (3) who holds locks that other users wait for.

Homework Exercise 7

d) Suppose you a disk with 12 ms average seek time, 6000 rounds per minute rotation speed, 400 KB per track and an Ultra-320 SCSI interface (320 MByte/s). How long does it take to read a single block of 4 KB on average? It suffices to compute whole ms.

e) You have to configure a RAID system with 4 disks of 2 TB each.

- A requirement is that the failure of a single disk does not lead to a data loss. Under this restriction, you want to maximize the storage capacity of the system. For equal storage capacity, you want to maximize the performance. Which RAID level would you choose?
- What will be the total storage capacity of the system?
- If each disk can read or write 100 blocks per second, how many blocks can read from the entire system per second?
- Give an estimate of how many blocks can be written per second. For simplicity, assume that a read-modify-write cycle takes twice as long as a reading or writing a block (it will actually be less). Also assume that the cache in the RAID controller is so small that it can be neglected.

f) Have a look at the following web page:

- [https://www.backblaze.com/blog/hard-drive-stats-for-2018/]

How many disk drives failed in the year 2018 and what is the total number of drives this company has? Which percentage failed?

g) You can find interesting research about hardware failures in big data centers in the publications of Bianca Schroeder:

[http://www.cs.toronto.edu/ bianca/index.html#publications]

Read at least the abstract of the following article and answer the question what annual replacement rate for hard disks is mentioned there as typical:

[http://www.cs.toronto.edu/˜bianca/papers/fast07.pdf]