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

November 24, 2017

Databases IIB: DBMS-Implementation — Exercise Sheet 7 —

Part a) to c) will be discussed in class, you only have to submit the Homeworks, i.e. Part d) to g). But please think about the questions in a) before the meeting. Send your solution for the homework to the instructor via EMail: `brass@informatik.uni-halle.de` (with "`dbi17`" in the subject line). The official deadline is November 29, 12:00.

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 many random/independent read/writes can a typical disk do per second?
- What amount of data can a disk read/write sequentially per second?
- How much faster can one write blocks of e.g. 4KB sequentially vs. randomly distributed on the disk?
- Why is using RAM in the disk controller for "read ahead" a good idea?
- 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) If you did not attend the last problem session, please try the web administration tool for the Oracle database:

`http://141.48.____.____:1158/em/`

The manual for the Oracle Enterprise Manager is available here:

[https://docs.oracle.com/cd/E11857_01/em.111/e11982/toc.htm]

Chapter 6 is about Database Management. There is also an online help in the web user interface.

c) Under your own account (not as `SYSTEM`), create a table `R` with

- a column `A` of type `NUMERIC(9)` and
- a column `B` of type `VARCHAR(50)`.

Do not declare a key for it, since the following simple PL/SQL procedure inserts many copies of the same row (if you want, you can modify the procedure so that it increments the number in each step):

```
CREATE OR REPLACE PROCEDURE FILL_BLOCK AS
    N NUMBER;
BEGIN
    N := 1;
    WHILE N < 2 LOOP
    INSERT INTO R VALUES(34, 'uvwxy');
        SELECT COUNT(DISTINCT DBMS_ROWID.ROWID_BLOCK_NUMBER(ROWID))
        INTO N FROM R;
    END LOOP;
END;
/
```

You can download a file that contains this procedure definition from the following address:

[http://www.informatik.uni-halle.de/˜brass/dbi17/fill_block.sql]

Execute this file in SQL*Plus.

If you should get the message "procedure created with errors", enter the command "`SHOW ERRORS`". You can execute the procedure with

```
call fill_block();
```

How many rows does the table contain? The number is one more than the rows that fit in one block. If you have still time, you can write an SQL query that selects the block size from the data dictionary (e.g. `USER_SEGMENTS` is accessible as a normal user) and computes the number of bytes that are used for each row. Note that there is some overhead for the block, and the block might not be completely full, so you cannot expect a whole number. We will talk later in the course about computing the storage size of a row.

Homework Exercises

d) Compare SSD drives with magnetic hard disks. Name at least one advantage of the SSD technology and one disadvantage. Please demonstrate the advantage and the disadvantage with some numbers for concrete drives (one magnetic disk and one SSD, you can choose any models you like).

If in the last homework, you looked at the data of a magnetic drive, look now at the data of SSDs, for instance the Samsung 850 Pro with 512 GB for approx. 220€ (some of the tests are for slightly different variants of this SSD):

- [http://www.samsung.com/semiconductor/minisite/ssd/product/consumer/] [850pro.html]
- [https://www.tomshardware.de/] [samsung-sd-850-evo-test,testberichte-241703.html]
- [http://www.pcgameshardware.de/SSD-Hardware-255552/Tests/] [Samsung-SSD-850-Pro-Test-1126978/]
- [http://www.storagereview.com/samsung_850_evo_ssd_2tb_review]
- [https://www.heise.de/preisvergleich/] [samsung-ssd-850-pro-512gb-mz-7ke512bw-a1136104.html]

You find some explanation on the CrystalDisk benchmark here (e.g. "QD32" stands for "queue depth 32", and means that there are 32 independent threads that read or write blocks concurrently):

- [https://wiki.mikejung.biz/CrystalDiskMark]
- [https://www.ssd-ratgeber.de/] [cristaldiskmark-ssd-hdd-benchmark-performance-leistungs-test/]
- [http://www.chip.de/downloads/CrystalDiskMark_32969879.html]

If in the last homework, you already looked at the data of an SSD, then have a look at some magnetic disk, for instance a "Seagate ST4000LM024" (4 TB for approx. 150 €):

- [https://www.seagate.com/de/de/internal-hard-drives/hdd/barracuda/]
- [http://www.storagereview.com/seagate_mobile_hdd_review]
- [https://www.pcwelt.de/produkte/] [Test-Seagate-Desktop_HDD.15-ST4000DM000-Festplatte-7914536.html]
- [https://www.heise.de/preisvergleich/] [seagate-barracuda-compute-4tb-st4000lm024-a1521938.html]

Information about the IOMeter benchmark is available here:

- [http://www.iometer.org/]

- [https://www.itechstorm.com/iometer-tutorial-introduction]
- [http://www.itprotoday.com/management-mobility/]
  [using-iometer-measure-disk-performance]
- [http://www.storagereview.com/articles/200003/20000313OSandBM_5.html]

For Linux, `fio` is often used:

- [https://www.binarylane.com.au/support/solutions/articles/]
  [1000055889-how-to-benchmark-disk-i-o]
- [https://www.thomas-krenn.com/de/wiki/Fio_Grundlagen]
- [https://www.linux.com/learn/inspecting-disk-io-performance-fio]
- [https://www.cyberciti.biz/tips/linux-disk-benchmarking-io.html]
- [https://github.com/axboe/fio]

This is a performance comparison for Oracle using disks vs. an SSD-like device:

- [https://www.sandisk.com/content/dam/sandisk-main/en_us/assets/]
  [resources/enterprise/white-papers/]
  [comparing-performance-of-oracle-disk-array-vs-fusion-iomemory-pcie.pdf]

e) 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.

f) 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.

g) Have a look at the following web page:

- [https://www.backblaze.com/blog/hard-drive-failure-rates-q1-2017/]

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