Exercise Sheet 1: Physical Data Storage (discussed on Friday 09.05.2014)

Exercises will be discussed on **Friday** on week after the respective lecture was given. The handouts are optional and do not have to be handed in. They only serve as optional preparation for the oral exams at the end of the semester.

Exercise 1.1 – Storage Types
a. Explain the main characteristics of offline, nearline and online storage and give an example
b. Why is data not stored in primary storages by default?
c. Why is sequential access so much faster on hard drives than random access?

Exercise 1.2 - Reliability
a. What are common measures to express the reliability of a hard drive?
b. How and why does the reliability differ between hard drives used for desktops or servers?
c. Briefly explain the bathtub curve

Exercise 1.3 - RAID
a. Suppose you have bought **4** hard disks with **500 GB** capacity each and a maximum transfer rate of **80 MB/sec** each. You group these disks either as RAID **0**, RAID **1**, RAID **5**, RAID **6** and RAID **1+0**, respectively. **For each** RAID configuration:

    1. What is the maximum capacity?

    2. What is the maximum transfer rate for reading a **single** block?

    3. What is the maximum transfer rate for writing **two blocks** concurrently? (Assuming that calculating parity does not consume any time. Also, note down if no concurrent write access cannot performed at all.)

b. Assuming a RAID 5 and a RAID 6 array with the **same** idealized MTBF. Why is a RAID 6 still considered more secure than the more efficient RAID 5?

c. You are recovering a lost data fragment in a (4+1) RAID 5. You have D_1=0010 1110, D_2=1100 0101, D_3=1111 0000 with the parity P=1011 1100. What was the data in the lost fragment D_4?

Exercise 1.4 – Network storage
a. Contrast the characteristics of a SAN to those of a NAS
b. You have a small company with 15 employees. Everybody has an own workstation, but they need to store their documents at a central place. Having the choice between a RAID, a NAS, and a SAN, what is the best solution to do this and why?