Storage and Backup

memory - mälu
storage - salvesti(d)
backup - varundus -
subcategory of storage + supporting processes
Hierarchy of Storage

“Classical” view, as seen from the processor:

- primary
- secondary
- tertiary
- off-line
Storage

short term / temporary

long term / permanent
Storage Media

- magnetic storage
  - hard disk drive, floppy disk, tape
- optical storage
  - CD, DVD, HD-DVD, Blue-Ray
- semiconductor storage
  - “Flash” storage, Solid State Drive (SSD)
Hard Disk Drive Interfaces:

- SAS (Serial Attached SCSI)
- SATA (Serial ATA)
- FC (Fibre Channel)
- IDE, ATA (Integrated Device Electronics, Advanced Technology Attachment)
- SCSI (Small Computer System Interface)
HDD: characteristics

- interface, r/w transfer rate, cache
  - transfer rate of the interface ≠ “real” read/write transfer rate
- capacity
- physical dimensions
- hot-swap-ness (kuumvahetatavus)
- rotation speed (rpm)
  - 4200, 5400, 7200, 10k, 15k
- platters, heads, cylinders, sectors, ...
Storage Virtualization: RAID

➔ no virtualization – JBOD
➔ DAS - Direct Attached Storage
➔ RAID (Redundant Array of Inexpensive Disks)
➔ RAID (Redundant Array of Independent Disks)
  ➔ dividing and replicating data among multiple disks
  ➔ the result is presented as one virtual disk
  ➔ implemented in hardware and/or software
Storage Virtualization: RAID

RAID 0 (striped set) - not really “redundant”
Storage Virtualization: RAID

RAID 1 (mirror, peegel)
Storage Virtualization: RAID

combined RAID “levels”:

RAID 0+1 (01)
RAID 1+0 (10)
Activity I
Storage Virtualization: RAID

RAID 5 (*striping with parity*)
Storage Virtualization: RAID

RAID 6 (striping with dual parity)
Storage Virtualization: RAID

- RAID 0 - fast r/w, poor reliability
- RAID 1 - fast r, good reliability, expensive
- RAID 1+0 - fast r/w, good reliability, very expensive
- RAID 5, 6 - fast r, slow w, better economy
  - large caches are used to compensate slow writes
  - write caching needs battery-backed cache
- RAID is not a backup strategy
Storage Virtualization: LVM

- LVM - Logical Volume Management
- volume - an integral chunk of storage
  - physical volume - disk or RAID array
  - logical volume - built from physical volumes by concatenating, striping, mirroring, slicing
- functionality (all on-line): extending-reducing, moving data between physical volumes
- also uses some of the RAID terminology
Storage Virtualization: NAS

➔ NAS (Network Attached Storage)
➔ different level of abstraction compared to DAS
  ➔ still needs a “physical” storage layer beneath
➔ operating on the file system level
  ➔ NFS (Network File System)
  ➔ CIFS (Common Internet File System)
Storage Virtualization: SAN

- SAN - Storage Area Network
  - dedicated storage network
  - many-to-many connections
  - operating on the block level
- block devices over “regular” networks
  - (S)ATA over Ethernet (AoE)
  - iSCSI (Internet SCSI)
Data Backup

data protection process

→ protects against:
  → hardware failures
  → software bugs
  → user errors
  → administrator errors

archiving process

→ ...as sometimes required by law
Data Backup

the main (sole?) purpose of backup is:

enabling a successful restore
Backup Costs

- reliable backup systems are expensive, but to lose your data could cost even more!

- backup system must grow with storage

- files are backed up regardless of the importance of their contents
  - use technology and policies to make storage more effective
Backup Strategy

→ before the backup we must know (from the risk analysis process):
  → what kinds of critical data do we store and process?
  → where are these data sets located?
  → are some of them more critical?
  → how long restore times can we afford?
Backup Strategy

→ the level of the Data Backup service:
  → what data sets to backup (coverage)
  → how often to backup (schedule)
  → how many older backups to retain (history)
  → restore requirements (speed, granularity)
  → archiving requirements
Backup Levels

- full backup (level 0) (*täielik*)
- incremental level 1 (*konsolideeriv*)
- incremental level 2 (*täiendav*)

In the backup schedule (*varundusgraafik*), the levels are combined

- level 0 + level 1
- level 0 + level 2
Backup Schedule

- daily full backups are rare
- most schedules combine full + incremental
  - how often to make full backup?
  - what is the data “change rate”?
- media recycling, archiving
Restore

- restore / recovery
  - file recovery
  - file system recovery
- you can only recover what you have backed up!
- test the recovery
  - test the “full recovery”
Data Backup Security

physical security
- large-scale disasters may affect all of your equipment
- off-site backups are recommended

network security
- sensitive data does not stop being sensitive on the backup tapes
- security measures must be equal or better compared to the system being backed up
Backup Software

- which client platforms are supported?
- do your databases and repositories need special backup interfaces?
- are you being charged by the quantity of data or by the number of clients?
Backup Hardware

→ tape drives
  → (S)DLT - ((Super) Digital Linear Tape) - SDLT-320, SDLT-600, DLT-S4, DLT-V*
  → LTO (Linear Tape-Open), LTO-2, LTO-3, LTO-4...
→ tape library, jukebox - lindirobot
→ disk based backup
  → Disk-to-Disk backup (D2D)
  → Disk-to-Disk-to-Tape backup (D2D2T)
Storage problems

- hard drives tend to be the most unreliable components in the system
- user needs tend to grow faster than storage capacities
- storage capacities tend to grow faster than storage data transfer rates
- storage capacities tend to grow faster than backup system capacities :)