LHC Computing re-costing for 2006-2010 for the CERN T0/T1 center

Introduction

The costing exercise focuses on the implementation of the LHC computing fabric for the CERN T0 + T1 center. This covers the phase 2 of the LCG project (2006-2008) and the following 2 years (2009-2010) of operation.

The following 6 areas are covered :

- **CPU Resources**
- **Disk Storage**
- **Tape Storage**
- **Local Network LAN**
- **Outside Network WAN**
- **System Administration**

more details : http://lcg-computing-fabric.web.cern.ch/LCG-ComputingFabric/lhc_computing_cost_re-calculation.htm

History

- October 1999 — PASTA II
- February 2001 — Report of the steering group of the LHC computing review 'Hoffman' Report
- February 2002 — Task Force 1 Report
- October 2002 — PASTA III
- March 2003 — Re-costing exercise

Ingredients

To estimate the cost and it's evolution over the years several input parameters need to be taken into account :

- **Technology evolution (PASTA)**
- **Today's cost reference points of key components (e.g. 2.4 GHz processor, 120 GB disk , etc.)**
- **The prediction for the future price development (slope) of the components**
- **Estimation of the needed capacity and resources in 2006 and onwards from the experiments**
- **The computing architecture and model**

Requirements

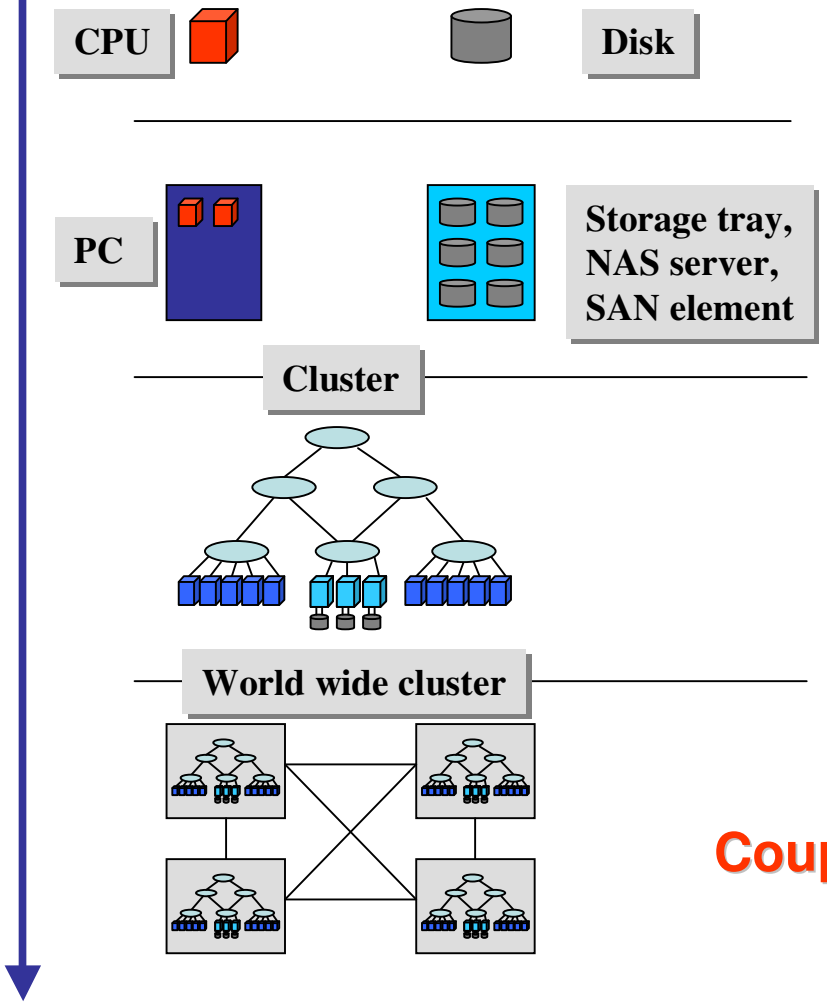
- The requirements are defined by the key experiment parameters like trigger rates and event sizes

- Some changes and refinements due to :

  → Much more experience with data challenges and productions
  → Better estimates and models
  → Different strategies e.g. ATLAS will keep one copy of the raw data at CERN while CMS will export the copy to the Tier 1 centers
  → Optimized model for the 'staging' of the equipment 2006-2008

Architecture (I)

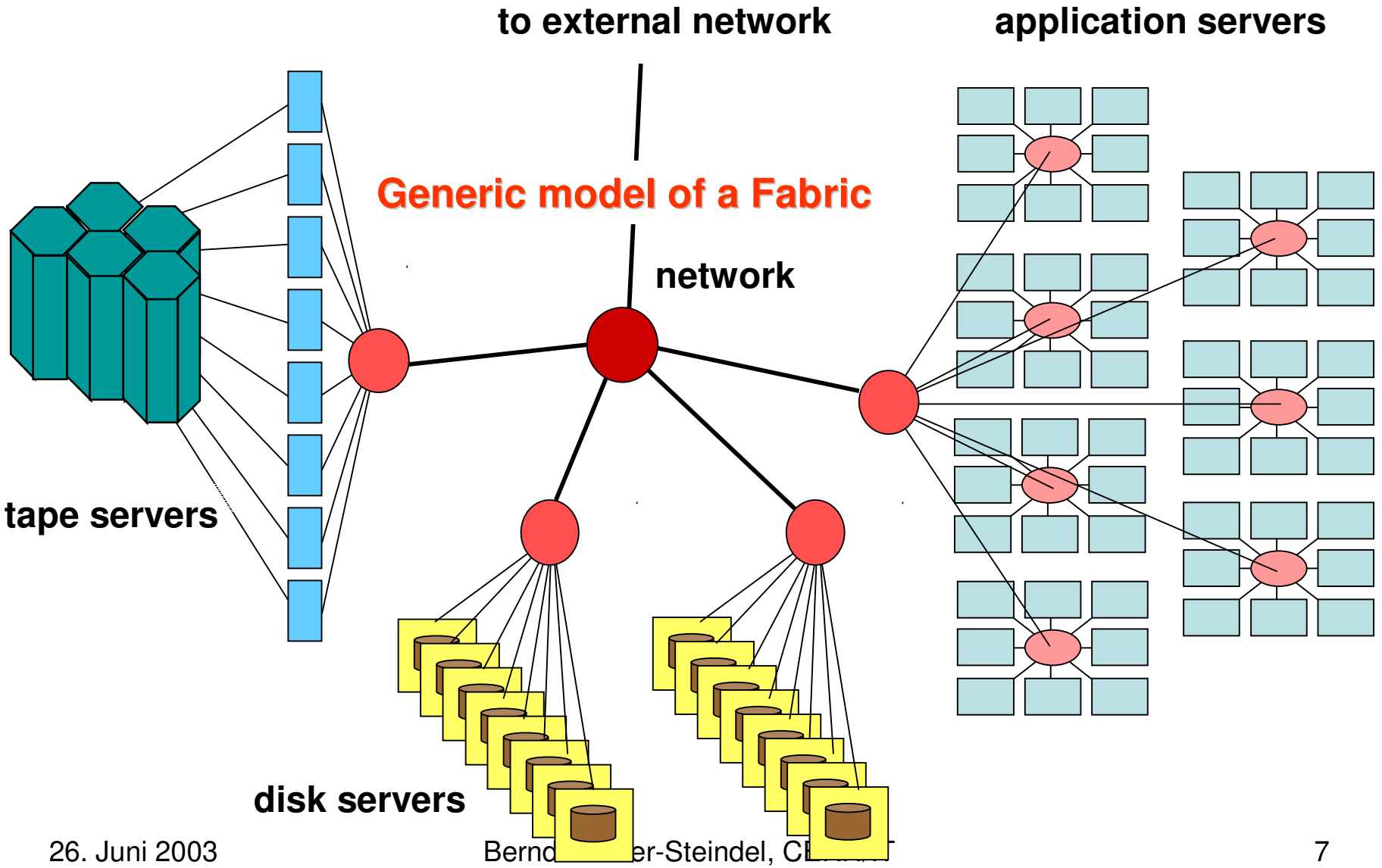
Level of complexity



	Physical and logical coupling	
	Hardware	Software
CPU, Disk	Motherboard, backplane, Bus, integrating devices (memory,Power supply, controller,..)	Operating system, driver
PC, Storage tray, NAS server, SAN element		
Cluster	Network (Ethernet, fibre channel, Myrinet, ….) Hubs, switches, routers	Batch system, load balancing, Control software, Hierarchical Storage Systems
World wide cluster	Wide area network	Grid middleware

Coupling of commodity building blocks

Architecture (II)



CPU Resources (I)

- Still focusing on the INTEL 'deskside' PC
- Have to consider additional costs :
  - → infrastructure (racks, cables, console,etc.)
  - → efficiency (batch system, I/O wait, etc.)
  - → market developments = difference between simple boxes and servers

Market

According to Gartner/Dataquest, the Notebook share in the PC market (Q1 2003) was 33 %. Intel claims that this year the sale of Notebooks will reach 40 million units.

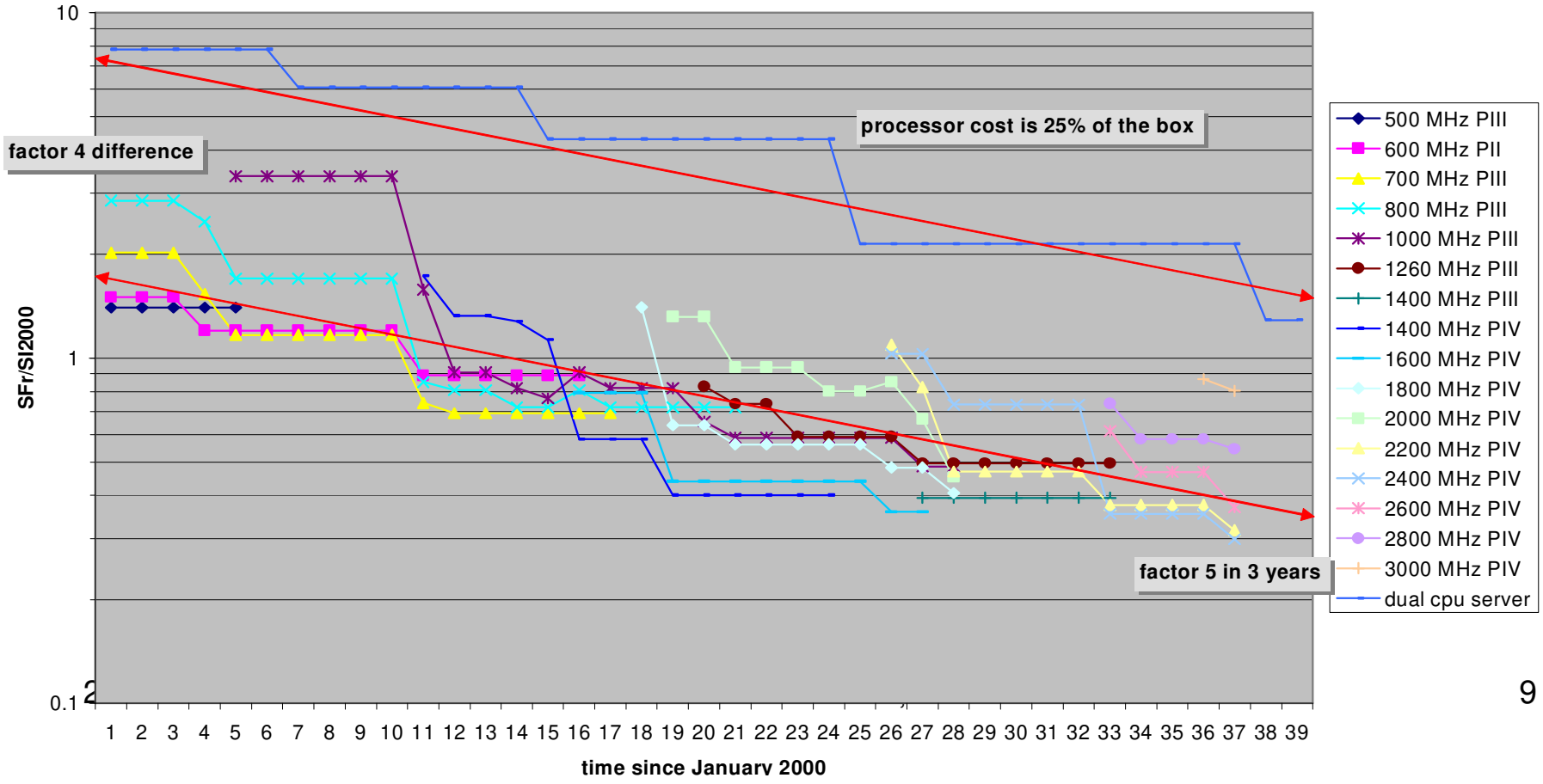
Power

The power consumption per produced SI2000 is still constant. There are plans (INTEL, TeraHertz) to reduce this in the future, but not convincing yet.
→ consequences for the upgrade to 2MW power and cooling in the center

CPU Resources (II)

- Processor reference points and slopes from 'street' prices, the actual purchases in IT during the last 3 years and PASTA III

Processor price/performance evolution



CPU Resources (III)

Year	Resource [million SI2000]	Cost [million CHF]
2005		
2006	3.7	3.9
2007	8.2	3.2
2008	19.1	5.1
2009	25.4	2.0
2010	33.8	2.5

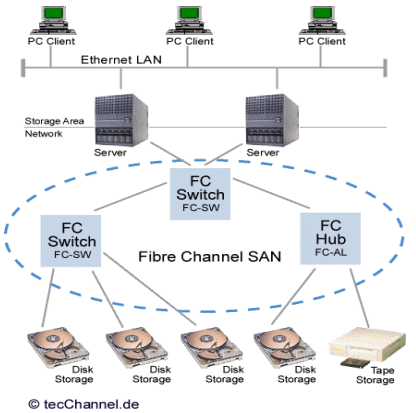
Disk Storage (I)

- Focusing on the commodity mass market
  → IDE, SATA disks

- Extra costs :

  → a few disks are attached to a server, server costs
  → infrastructure (racks, console,etc.)
  → efficiency of space usage (OS/FS overheads,etc.)

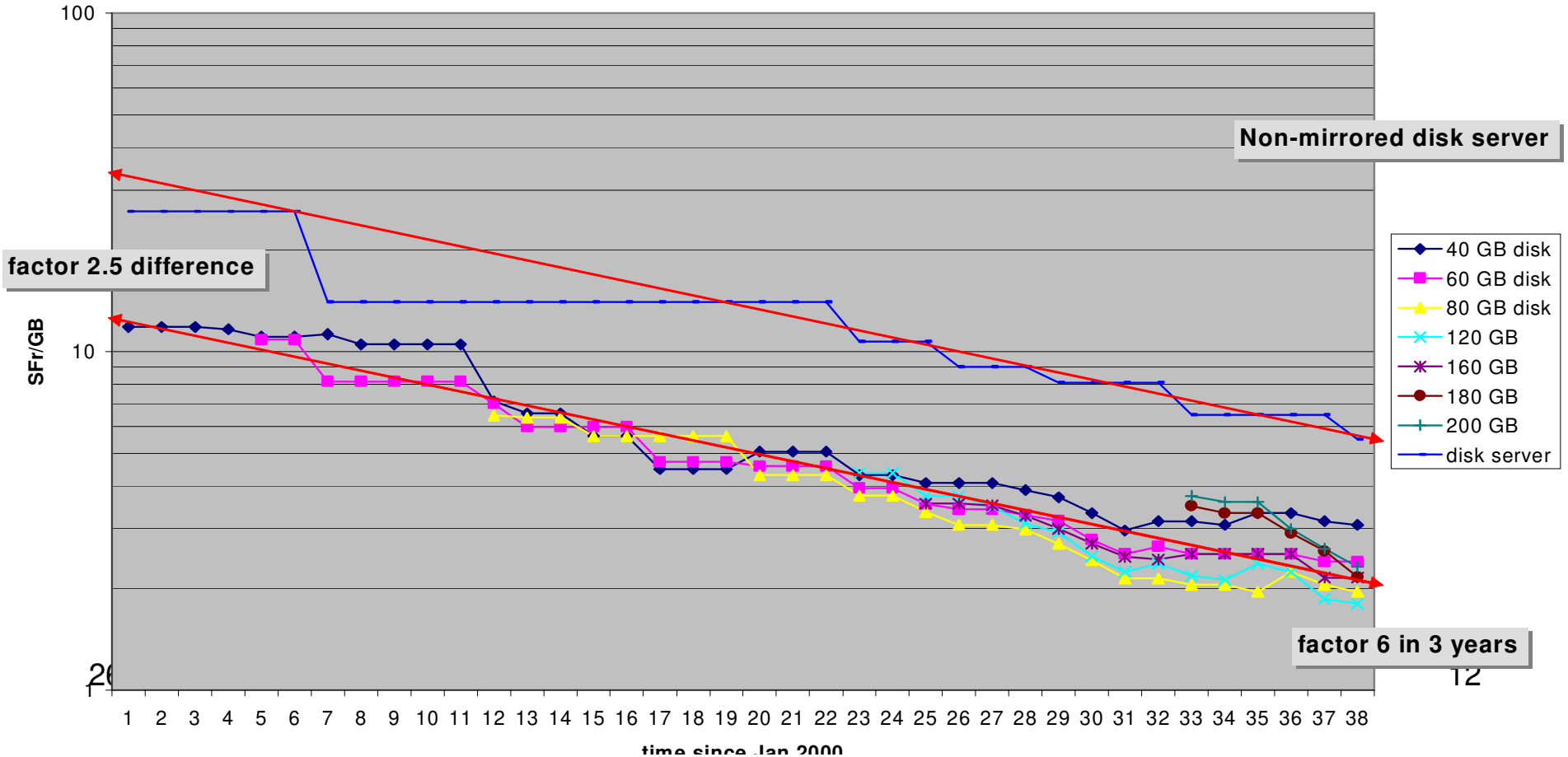
- Have to take into account the need for about 10% high-end storage = more expensive (x4)
  → databases



Disk Storage (II)

- Disk reference points and slopes from 'street' prices ,
  the actual purchases in IT during the last 3 years and PASTA III

Disk price/performance evolution



Disk Storage (III)

Year	Resource [PB]	Cost [million CHF]
2005		
2006	1.0	5.1
2007	2.1	3.4
2008	3.8	3.5
2009	5.0	1.6
2010	6.7	1.4

Tape Storage (I)

- Tape access performance is averaged over the year
  → needs dedicated, guaranteed resources during CDR of heavy-ion period

- Tape storage infrastructure :
  → number of silos for the tapes
  → new building when the number of silos > 16
  → maintenance costs per year

- Technology lifetime is about 5 years
  → replacement of equipment and re-copy of tapes (considerable expenses)
  → timing of the technology change is crucial

Tape Storage (II)

- Tape Storage performance infrastructure (tape drives, tape server, etc.)

Year	Resource [GB/s]	Cost [million CHF]
2005		
2006	1.1	1.2
2007	2.3	3.3
2008	3.9	3.3
2009	4.4	0.7
2010	4.4	0.0

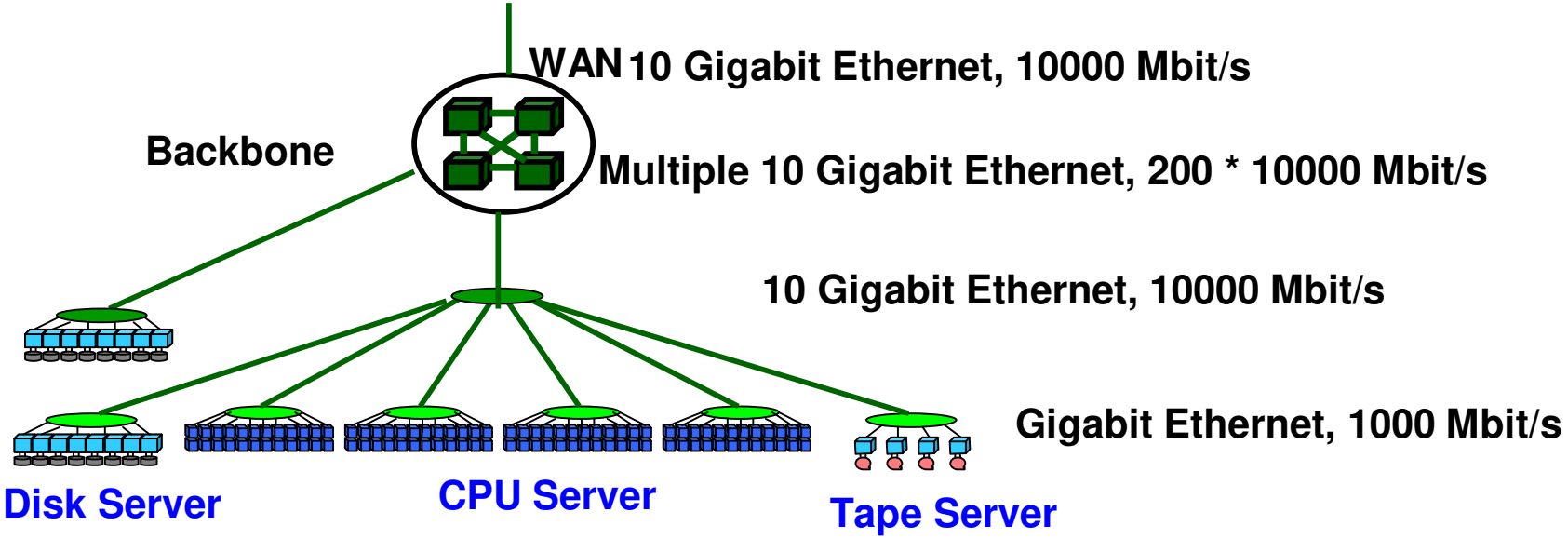
Tape Storage (III)

- Tape storage : tape media, silos, building, replacement

Year	Resource [PB] (total available tape capacity)	Cost [million CHF]
2005		
2006	6.0	5.3
2007	13.8	8.3
2008	25.1	6.5
2009	35.5	6.4
2010	48.4	10.6

Network Infrastructure (I)

- Network architecture based on several Ethernet levels in a hierarchical structure
- Implementation staged between 2005 and 2007 : 20%-50%-100%
- Designed for a 8000 node fabric
- completely new backbone based on 10GE equipment, still some uncertainties in this market for high end routers



Network Infrastructure (II)

Year	Resource [GB/s]	Cost [million CHF]
2005	56	2.2
2006	140	2.2
2007	280	4.3
2008	280	0.9
2009	280	0.9
2010	420	1.6

System Administration (I)

- The previous approached was based on the model of outsourcing the system administration part and was costed at about 1000 SFr per node.
- The new model is based on an in-sourced model, where in the first years 7 administrators and 2 managers are responsible for the sysadmin part, increasing to 12 administrators later
- This new model seems to be more appropriate after the experience from the last years and the fact that the amount of anticipated nodes is 'only' in the range of ~4000.
  This reduces the cost to about 400 SFR per node

System Administration (II)

Year	Resource [FTE]	Cost [million CHF]
2005	7+2	1
2006	7+2	1
2007	7+2	1
2008	12+2	1.5
2009	12+2	1.5
2010	12+2	1.5

Wide Area Network (I)

- Optimistically we should have access to 10 GBit WAN connections already in 2004
- The move to 40 GBit is much more unclear

→ The network providers are still undergoing frequent 'changes' (mergers, chapter 11)
→ Large over-capacity available
→ Todays 40 GBit equipment is very expensive

Wide Area Network (II)

Year	Resource [Gbits/s]	Cost [million CHF]
2005	2.5	1.0
2006	10	2.0
2007	10	2.0
2008	X *10	2.0
2009	40	2.0
2010	40	2.0

Comparison

All units in [million CHF]

Resource	Old 2006-08	New 2006-08	New- Old 2006-08	Old 2009-10	New 2009-10	New - Old 2009-10
CPU+LAN	17.7	19.5	1.8	6.3	6.8	0.5
Disk	6.3	11.9	5.6	2.2	2.9	0.7
Tape	22.5*	27.8	5.3	19.2	17.6	-1.6
WAN	11.4	6.0	- 4.4	6.8	4.0	-2.8
Sysadmin	7.9	3.5	- 4.4	6.6	3.0	-3.6
SUM	65.8	68.7	2.9	41.1	34.3	-6.8
Budget		60.0			34.0	

* **A bug in the original paper is here corrected**

Resource changes

Resource	2006-08	change 2006-08	2009-10	change 2009-10
CPU [MSI2K]	19	12 %	34	19 %
Disk [PB]	3.8	79 %	6.7	58 %
Tape [PB]	25	-1 %	48	-1 %

We are in the process of providing a detailed description and explanation for the resource requirement changes. This will be appended to the note describing the calculations and the used parameters in the re-costing Excel sheets.

Summary

- Better than anticipated price developments for the components
- Not technology changes, but market changes are the most worrying factors
- Exercise needs to be repeated regularly (at least once per year)
- Very fruitful and constructive collaboration between IT and the Experiments