Finding the perfect file system for life sciences An impossible mission?

Roberto Fabbretti
Volker Flegel
Vital-IT
Swiss Institute of Bioinformatics

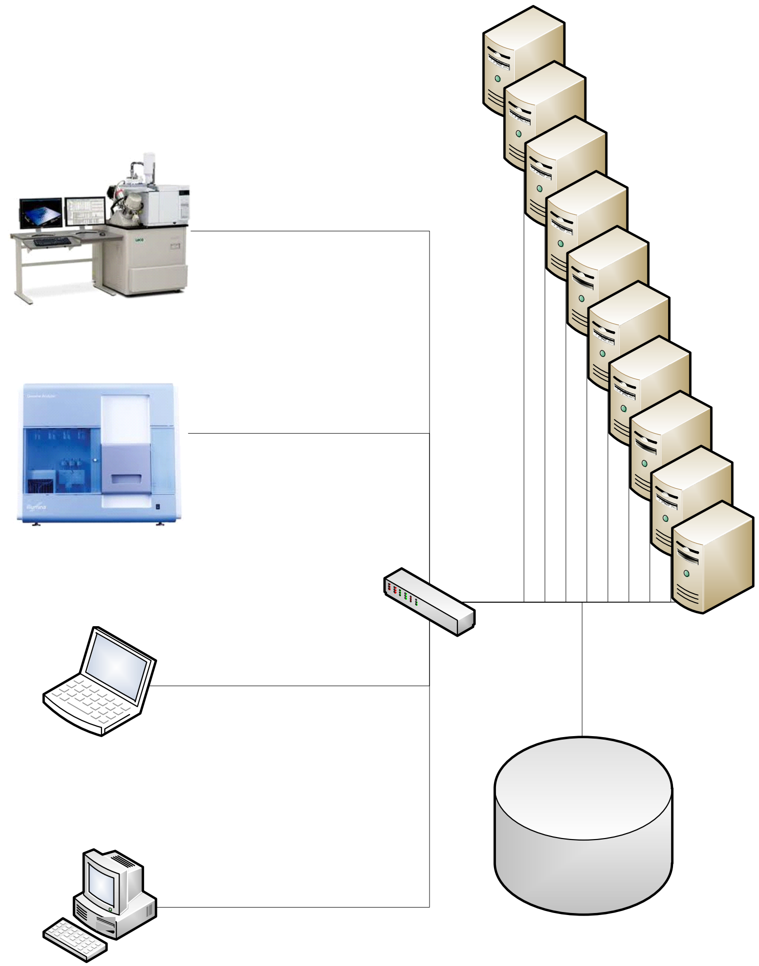
Lausanne, Switzerland

Vital-IT

SIB
Swiss Institute of Bioinformatics

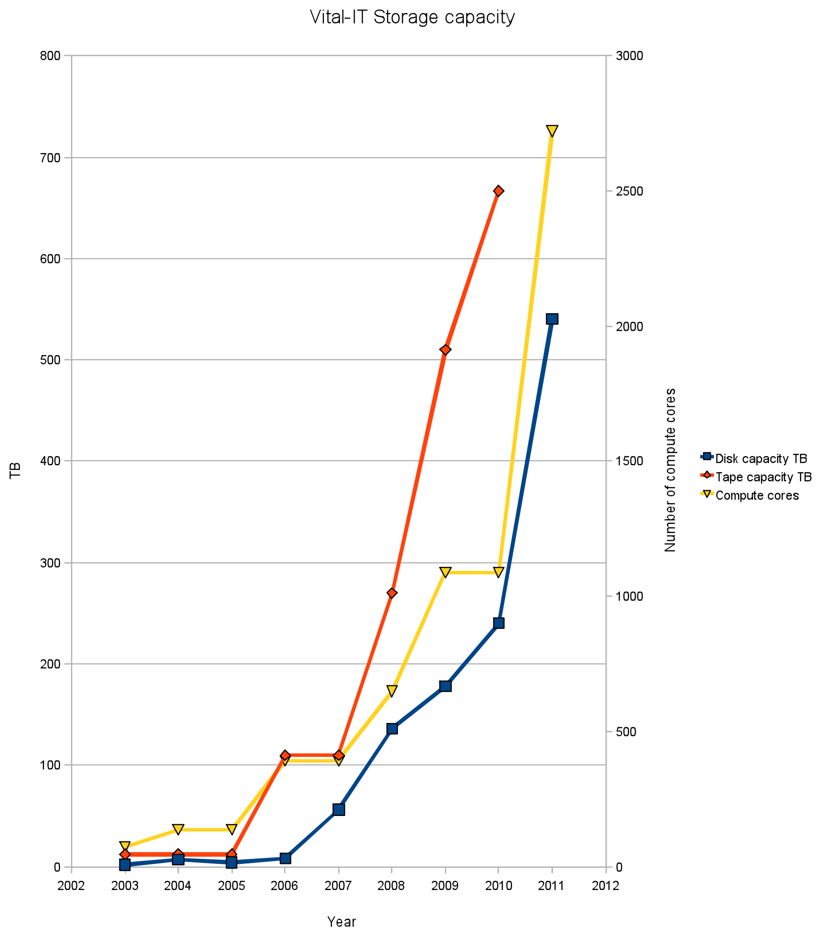
Life sciences are a new Customer of HPC

- Most end-users have little computer literacy
- Groups are small
- Data is mostly produced from Windows workstations
- Some data require a lot of computing to be analyzed
- Some categories of end users are big consumers



A brief history of Vital-IT storage

- 2003 1.7 TB
  - SAN attached storage
  - NFS server
- 2006
  - Lustre 8TB
- 2007
  - Addition of NFS attached storage for dedicated projects
- 2008
  - Implementation of HSM storage

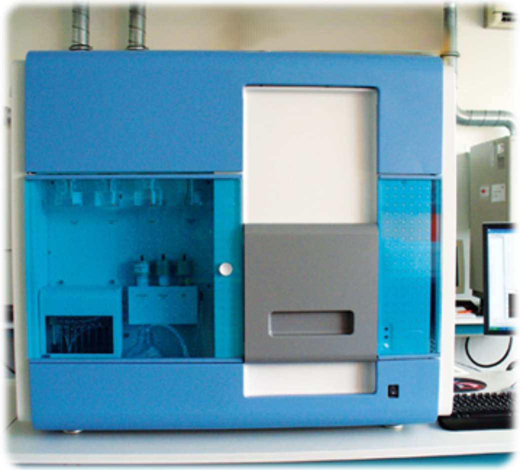


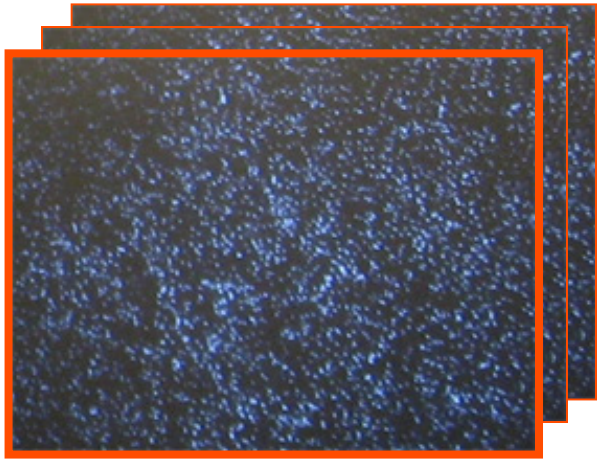
Vital-IT

SIB Swiss Institute of Bioinformatics

Next Generation Sequencing redefine Genomics

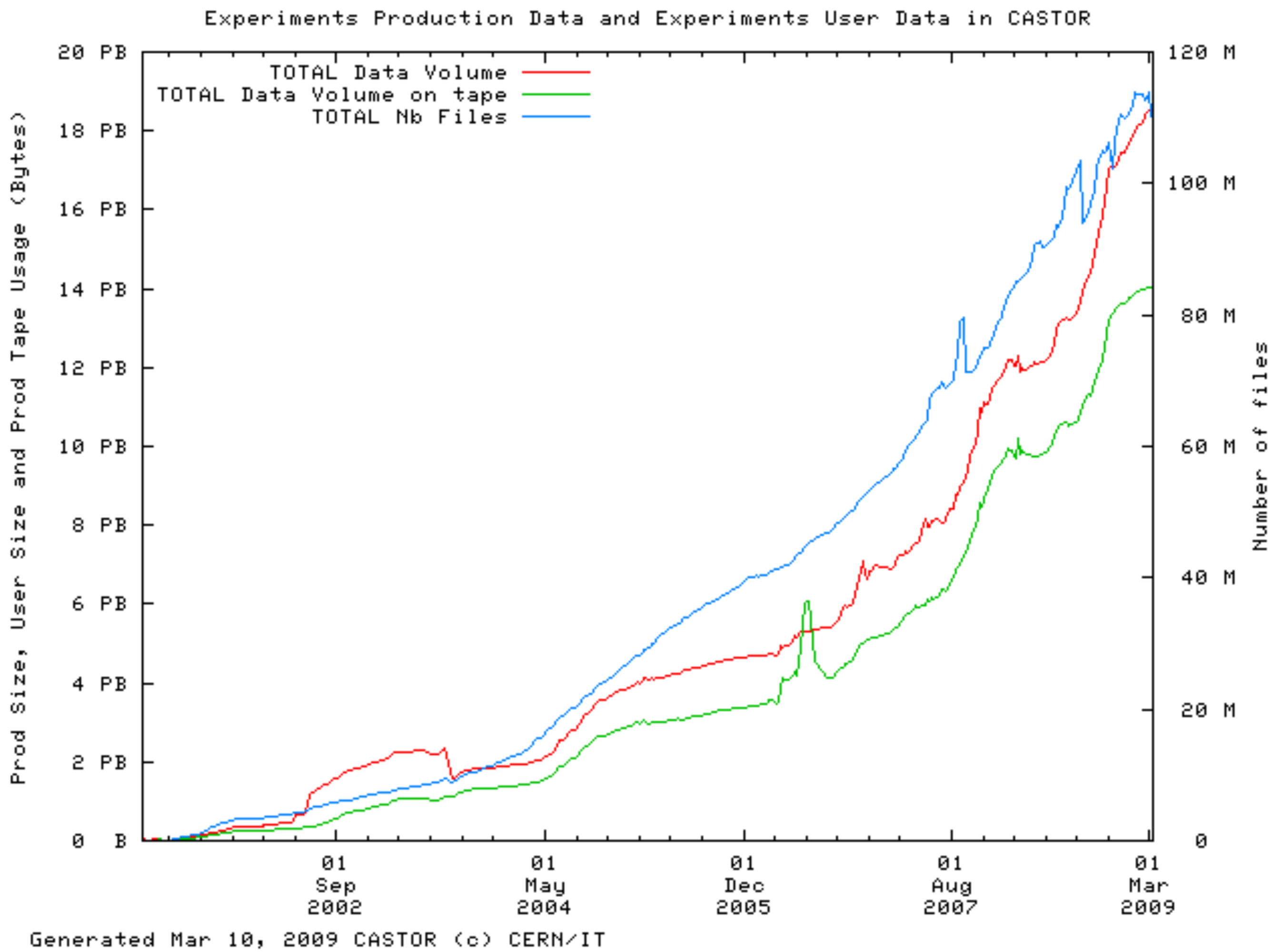
Ultra High Throughput Sequencing (UHTS)
Solexa





Data per run/per week
2007 – **1 Terabytes** of raw and processed data
2008 – **2.5 Terabytes** of raw and processed data
2009 – **7 Terabytes** of raw and processed data

Comparison of the growth with the CERN



Currently on Vital-IT

86 Millions files
for 483 TB

Three machines
operational since 2008

Requirements

Create a storage infrastructure capable of

•scaling to many PB

•hosting hundreds million of files

•providing the lowest cost or people will put their data on usb disks

•eliminating the need of backup

•being accessible from a compute farm

•providing a credible full disaster recovery scenario

What we try to avoid

Courtesy Chris Dagdigian BioTeam

What we have considered

- Lustre
  - Pro
    - Cheap
    - Performance
  - Cons
    - Difficult to set up
    - Poor metadata performance
    - Reliability
    - Manageability
    - Backup
    - Disaster recovery

What we have considered

- Netapp
  - Pro
    - Easy setup
    - Robust
    - Excellent availability
  - Cons
    - Price
    - Performance
    - Standard Ontap OS limited to 16 TB volumes
    - Ontap GX (cluster NAS) not in mainstream
    - Backup
    - Disaster recovery
    - Low storage density

What we have considered

- Isilon
  - Pro
    - Easy setup
    - Robust
    - Excellent availability
  - Cons
    - Price
    - Performance for small configurations
    - Backup
    - Disaster recovery
    - Low storage density

What we have considered

- Panasas
  - Pro
    - Easy manageability
  - Cons
    - Price
    - Poor metadata performance leads to complex configuration
    - Backup
    - Disaster recovery
    - Low storage density

What we have considered

- GPFS+TSM
  - Pro
    - Robust
    - Excellent availability
    - Community
  - Cons
    - Price
    - At the time we considered it TSM could manage only a few million files

Quantum StorNext

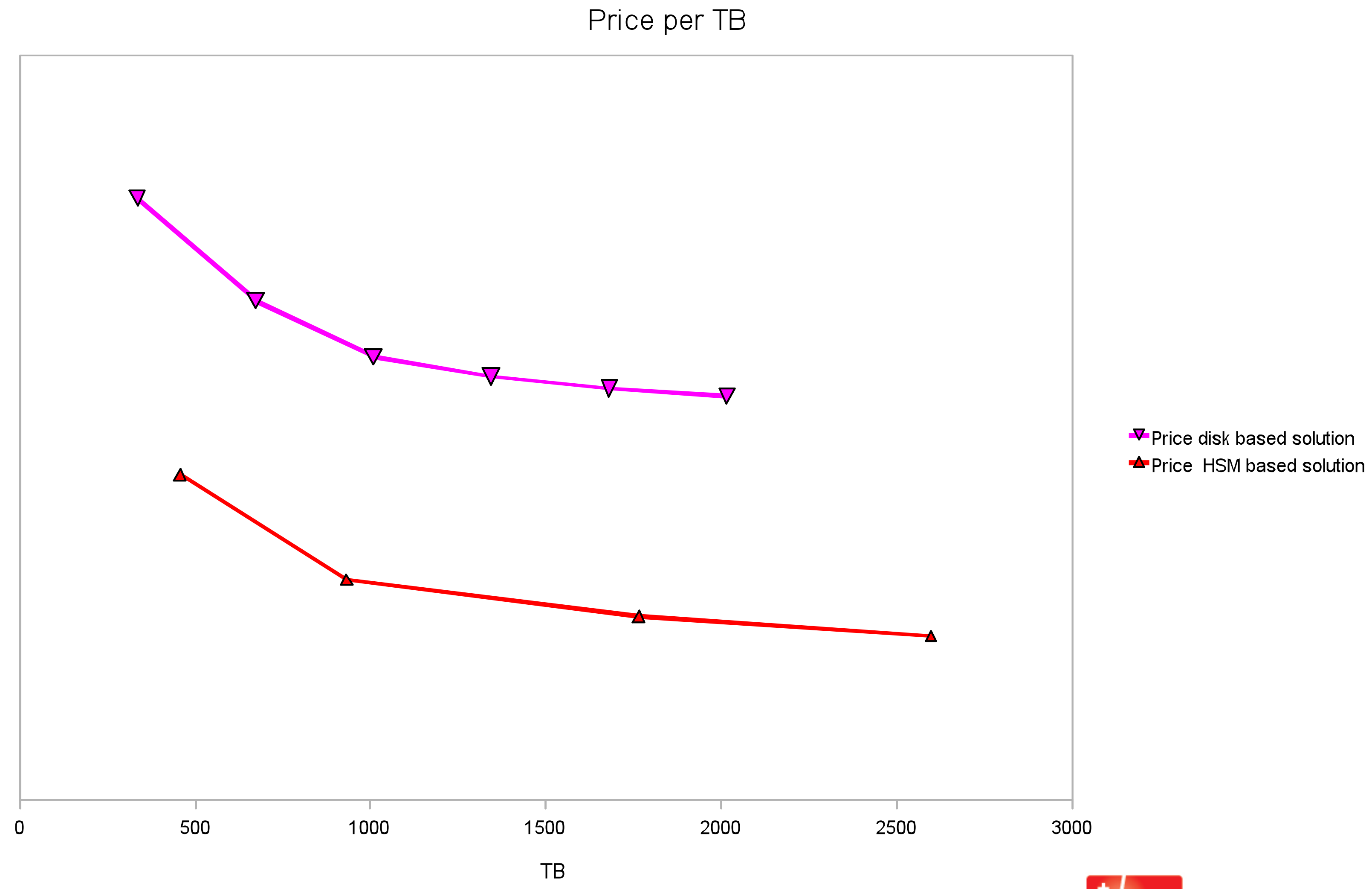
Pros

- Fulfills both requirements HPC file system and HSM
- Offers a tight integration between cluster filesystem and storage manager
- Transport independent (Infiniband, 10 Gb ethernet)
- Hardware vendor independent
- Scalable performance
- Easy resizing of volumes

Cons

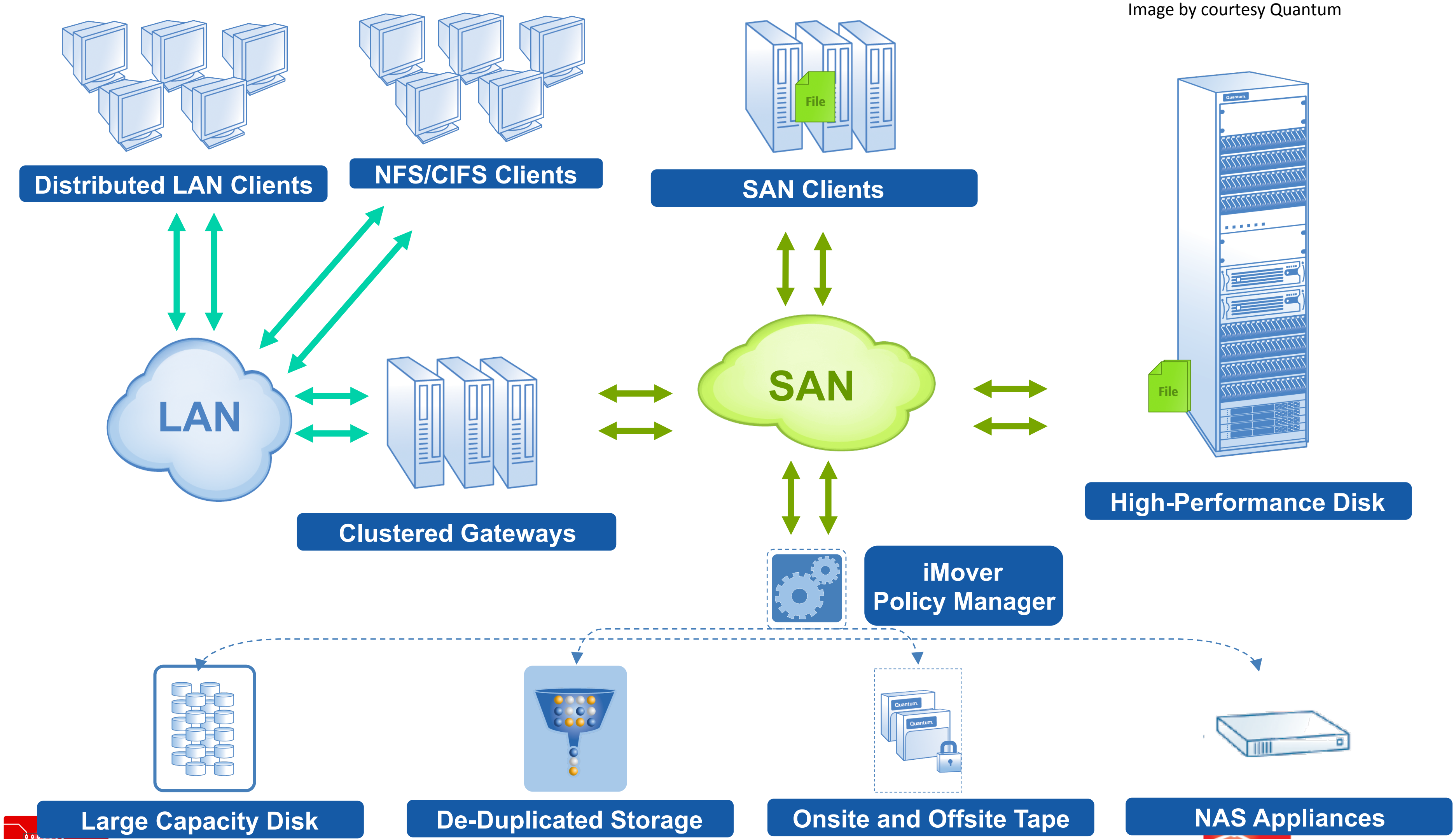
- Price
- Commercial product
- Lots of moving parts

TCO


Price per TB
Price disk based solution
Price HSM based solution
0
500
1000
1500
2000
2500
3000
TB

Quantum StorNext architecture

Image by courtesy Quantum



Quantum StorNext architecture

Quantum StorNext has two components

- Quantum StorNext filesystem
  - Cluster filesystem which can assemble LUNS on a SAN and present them to hosts through a dedicated protocol or NFS/CIFS
  - Directs the I/O to the physical disks through affinities
  - Organizes the migration of data from one physical disk to another while keeping the filesystem view constant
- Quantum StorNext Storage Manager
  - Organizes the copy of data from disk to tape and reciprocally according to defined policies
  - Monitors the fill level of a filesystem
  - If low watermark is reached data are truncated and a pointer is left which is shown to the user as the real file

Storage infrastructure

- HSM front end servers
- 2 metadata controllers
- 3 SAN clients

- Disk cache
- EVA8100 144 TB RAW extensible to 168 TB
- 2 x Transtec SUMO 550 48 TB RAW

- Tape storage
- 660 TB For primary copies
- 95 TB For secondary copies

Vital-IT

SIB Swiss Institute of Bioinformatics

Storage infrastructure

HP SFS cluster file system

- 16 TB storage

Tape libraries

- 830 TB storage

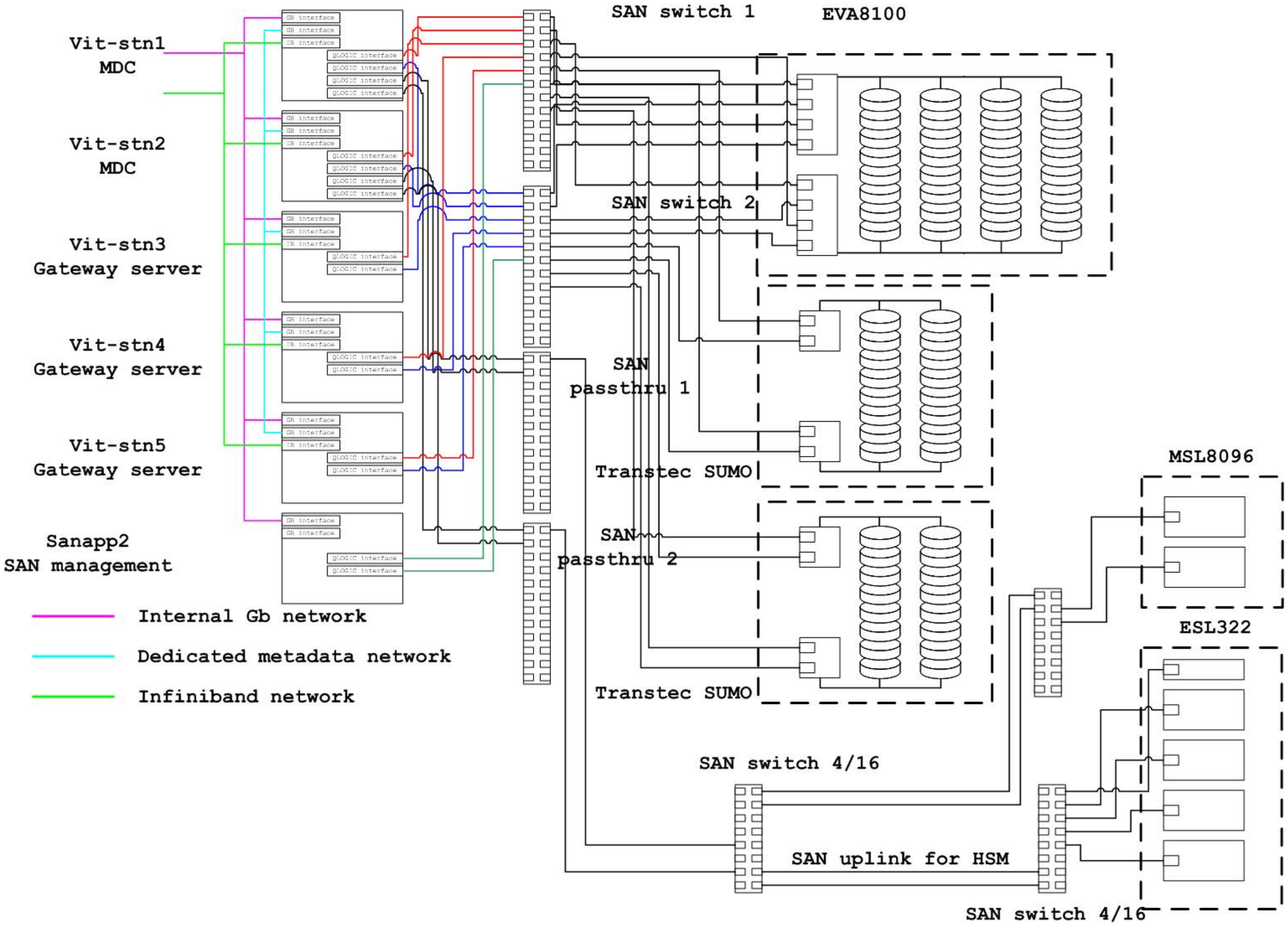
HP EVA 8100

- 74 TB VRAID 5 storage

Transtec Provigo SUMO 550F

- 74 TB RAID 6 storage

Network setup



Host configuration

Two metadata servers

- One active
- One standby

Five gateway servers

- One samba server
- Four NFS servers

CentOS 5

2 network interfaces

- One for the gigabit backbone
- One for the metadata dedicated network

Infiniband interface

Operations

To be successful an HSM has to be used the right way:

All important files must stay on the disk cache.

All useless files have to be truncated in order to stay below low watermark level

Truncation policies are set by directory

Scripting allows to select a certain type of files for truncation

For example

UHTS images are removed from the disk cache after 1 month

Conclusion

- System is operational
- System is robust
- Requires trained operators
- Requires a valid 24/7 support contract
- Allows to make large retrospective analysis of archived data
- Allows to associate biological data from different technology platforms