

SURFnet storage pilot plans

Rogier.Spoor@SURFnet.nl

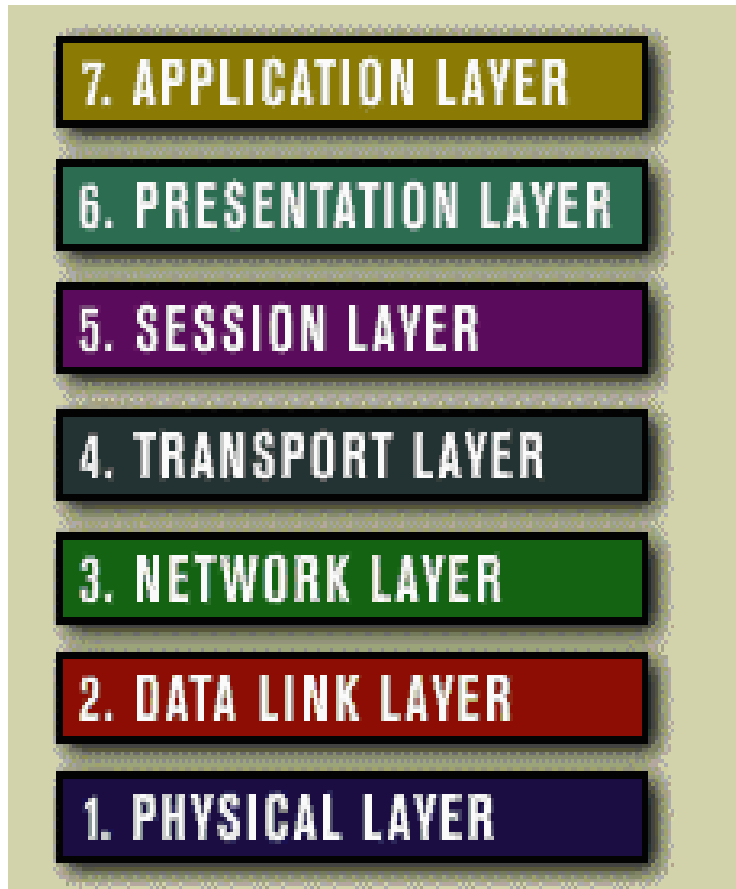
17 september 2009



What do we do ?



OSI-layer



Founded to start a nationwide network



What's our activity focus

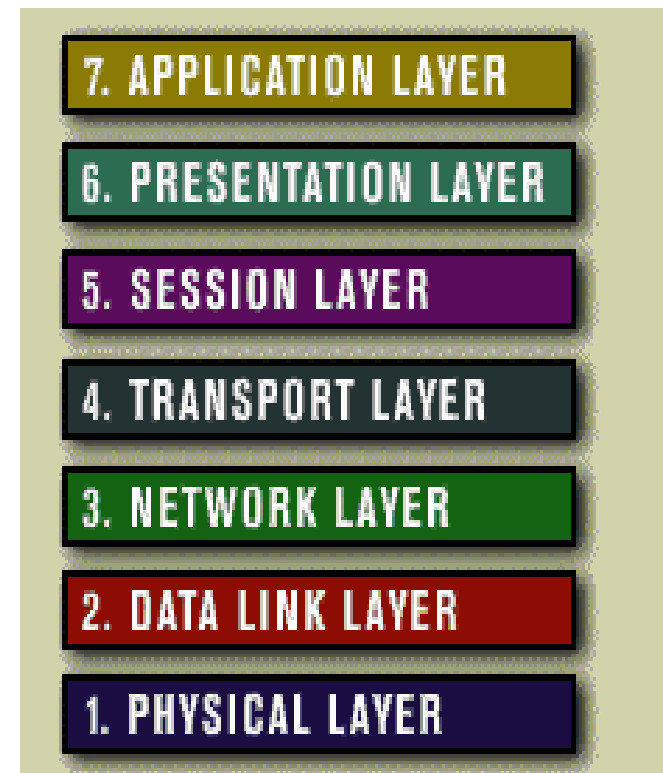
"We make innovation work"

- (online) collaboration
- Thrust
- Expertise centre
- Network
- Authentication and authorization
- Security



Network got commodity

- So when you want to innovate, what's the direction?
- Henry Chesbrough says there're two options:
 - Up in the stack
 - Down in the stack





Down in the stack

- Digg your own fibers (like Aarnet)
 - Build and develop your own network equipment
- ➔ Not logical choices
- Insource the network maintenance
- ➔ is partly feasible



Up in the stack

Options:

- TCP or Session level (stack development)
- Constituent network device management (firewall, router, LAN)
- Application layer

→ Network device management “not done”

→ Only option “the application layer”



What's needed for the application layer?



Foundations are:

- Computing power (cpu)
- Storage
- Network faciliteiten

Architecture:

- scalability
- performance
- reliability (inc. security)



Current application foundation



History based

- Every service own architecture
- Every service own physical hardware

Status

- CPU, Storage en Netwerk not in single architecture
- CPU Virtualisation SURfnet-services since 2008
- New SAN storage in 2009 -> for VM
- Network capacity on demand



Remote storage project

- Small project (in 2007)
- Two phases
 - Remote storage strategies paper

Phase II

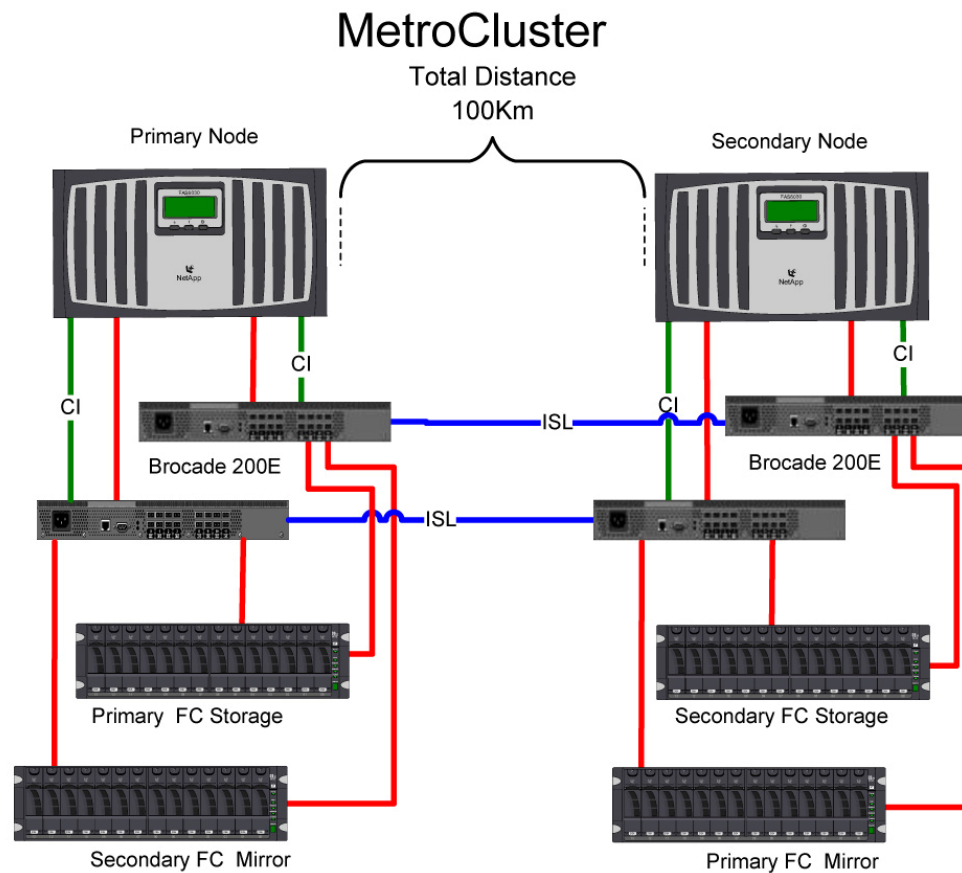
- Practical test
- How to outsource storage?
- Use of lightpaths versus IP
- Long distance (500 km)
- Only ethernet (no FC in backbone)
- Ethernet -> FC converters



How to satisfy storage needs

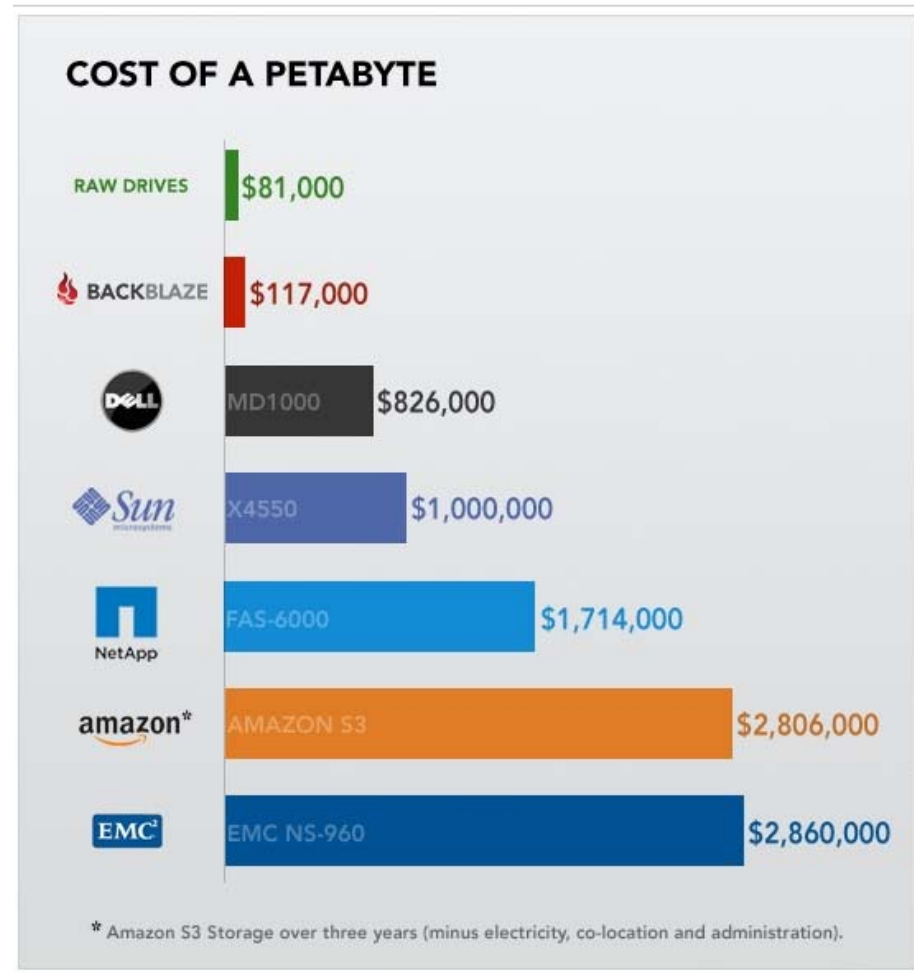


VM environment






Storage costs



<http://blog.backblaze.com/2009/09/01/petabytes-on-a-budget-how-to-build-cheap-cloud-storage/>



Storage type depending needs



- NAS/SAN storage suitable for VM's and DB's
- Cloud storage suitable for archiving data of streaming media
- Clustered storage → research



New storage project

- Combine different storage techniques
- Store VM images, media content
- Offer a single mount point
- Easy scalable (multiple peta)
- Low cost/petabyte

Start project: Q4



Proposed design

- DRBD as "SAN/NAS"
- Mogile/Castor → cloud storage
- Fuse → switching layer



DRBD



- Distributed Replicated Block Device
- Raid 1 mirror across network
- Replication over tcp/ip
- SuperSockets for replication with dolphin
- Sync or asynch replication
- Split brain detection and recovery
- Open source



Castor



- Hardware agnostic
- Massively scalable
- High performance
- Guaranteed data integrity
- Self configuring
- Self managing
- Self healing
- Easy HTTP-based interface

Disadvantage:

- Difficult to manage files inside castor
- Commercial software (\$\$\$)



Mogile



- open source distributed filesystem
- http api
- automatic file replication
- no single point of failure
- Filesystem agnostic
- cheap storage



Glue



Intelligent switching:

- IO to specific file
- File extension (db, vm)
- single mount point
- Namespace virtualization

Proposed solution: fuse implementation

Questions/Suggestions ?