

pNFS

Parallel Network File System

Thijs Stuurman
Thijs.Stuurman@os3.nl

July 2, 2008

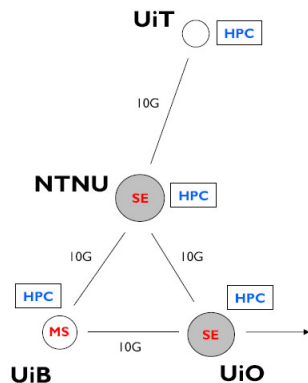


UNIVERSITEIT VAN AMSTERDAM

Masters program System and Network Engineering



- ▶ Introduction
- ▶ Research question
- ▶ Approach
- ▶ Research
 - ▶ pNFS?
 - ▶ How does it work?
 - ▶ pNFS behaviour. . .
 - ▶ pNFS availability & status
 - ▶ Usage scenario
 - ▶ NorStore
- ▶ Conclusion
- ▶ Questions



UNINETT, UNINETT Sigma, UiO (University of Oslo), NTNU (Norwegian University of Science and Technology, UiB (University of Bergen), UiT (University of Tromsø)

Main research question:

" Is pNFS capable of transferring the large amounts of data required in the NorStore context?"

Thesis:

" pNFS will be capable of transferring large amounts of data but is currently still missing good integration with underlying file systems and under heavy development. Therefore it is currently not ready for production use."

Approach:

- ▶ Reading, setting up PoC
- ▶ Analyse
- ▶ Analyse some more...
- ▶ Making sense of it all

Example current solutions:

- ▶ Panasas PanFS
- ▶ IBM GPFS
- ▶ Lustre
- ▶ PVFS2
- ▶ EMC MPFS

Working on pNFS:

- ▶ NetApp
- ▶ Sun Microsystems
- ▶ EMC
- ▶ Panasas
- ▶ CITI/UMICH

pNFS?

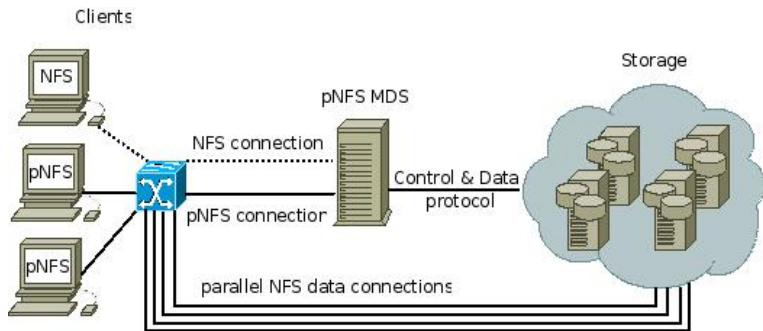


Figure: pNFS network overview

How does it work?

1. Client \Rightarrow MDS OPEN
2. Client \Leftarrow MDS Reply OPEN
3. Client \Rightarrow MDS LAYOUTGET
4. Client \Leftarrow MDS Reply LAYOUTGET
5. Client \Rightarrow DSs Parallel I/O
6. Client \Rightarrow MDS LAYOUTCOMMIT
7. Client \Leftarrow MDS Reply LAYOUTCOMMIT
8. Client \Rightarrow MDS CLOSE
9. Client \Leftarrow MDS Reply CLOSE

- ▶ Layout size & file size
- ▶ Small files
- ▶ Big files

- ▶ File based: Sun, IBM, DESY, NetApp, Red Hat, CITI
- ▶ Object based: Panasas
- ▶ Block/Volume based: EMC

Pnfs != pNFS

- ▶ pNFS file system?

2009...

- ▶ Any decent storage platform!
- ▶ Combined with existing tools as GridFTP

- ▶ pNFS is useful for the NorStore platform and is able to transfer large amounts of data.

Conclusion and future work

Conclusions:

" Is pNFS capable of transferring the large amounts of data required in the NorStore context?"

Thesis:

" pNFS will be capable of transferring large amounts of data but is currently still missing good integration with underlying file systems and under heavy development. Therefore it is currently not ready for production use.

Future work:

- ▶ Wait until pNFS (near)ready!
- ▶ Use a real 10Gbit environment, these things are tricky!
- ▶ Use two different storage back-ends, test interoperability.

Questions?