

# VL-e Medical

## Data grid facilities in VL-e

Alex de Landgraaf

IBM & Vrije Universiteit, Amsterdam

November 6, 2007

Introduction

Datagrid development / VGFS

Demo VGFS (screencast)

Issues with Datagrid & VGFS

# Who am I, what am I working on

- ▶ Student CS @ VU
- ▶ CEO @ Aperte
- ▶ Intern @ IBM (6 months, 4.5 to go)
- ▶ Goal: Usage of IBM products together with VL-e PoC
- ▶ Assessment of use of Grid infrastructure in a production environment, bridge the gap between research and commercial deployment of nation-wide grids



# Test-case AMC #1: Genome Sequencer FLX

- ▶ Sequences many DNA-pieces at once
- ▶ Processed data not very large (ACTGs, text-files)
- ▶ However algorithms still being improved, wish to store raw image data
- ▶ Raw data: 30GB/day, 8TB/year, stored for X years

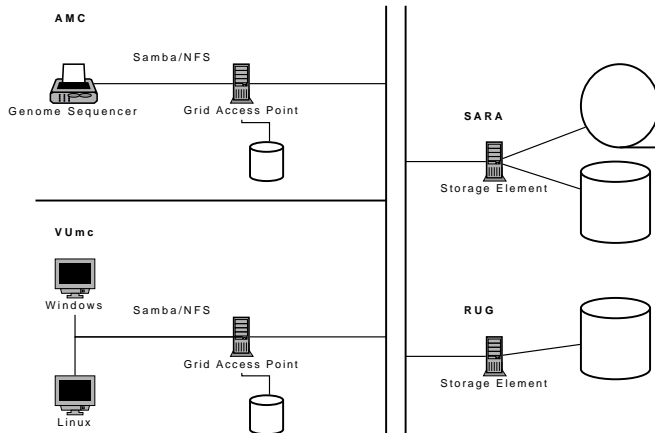


# What I am working on

- ▶ Inventory of IBM products (applicability to case, linking to VL-e)
- ▶ Experimenting with VL-e EGEE/SRM data interface
- ▶ Churning out documents
- ▶ Development of VL-e Grid/Filesystem interface proof-of-concept



# In which direction am I going?



## Technical details VGFS (aka acronym-time)

- ▶ My goal: filesystem-level integration with IBM products,
- ▶ Allow network-level integration (NAS) within organization (NFS/Samba/WebDAV/HTTP/SFTP/etc)
- ▶ Developed VGFS, a Linux FUSE (Filesystem in USErspace) program, runs on modified PoC v2 install
- ▶ Uses combination of LFC/LCG/GFAL EGEE APIs
- ▶ Supports both Buffered I/O and Direct I/O
- ▶ But in very early stage of development

# Time for a demo!

## Short-term issues to be solved

- ▶ Latency and gLite performance (as you have seen)
- ▶ Access control (need to look into gLite ACL support)
- ▶ Stability, POSIX-compliance tests, large file testing
- ▶ Multiple threads (unknown multithreaded-support in certain libraries)
- ▶ Encryption (Hydra or IBM-based product? Anything else?)

## Long-term/deployment issues to be solved

- ▶ Host-based vs. user-based grid certificates
- ▶ Low-maintenance GAP-box (acceptance by administrators, firewall issues, configuration & monitoring, datagrid-access-as-a-service?)
- ▶ Dialog with VL-e resource providers about testing semi-commercial deployment
- ▶ Determine when commercial use of VL-e/BIG GRID is viable and preferable to local storage (cost, trust, privacy, law)
- ▶ Grid-middleware independence (development against SAGA, OGF-defined GAT/JavaGAT successor)

# You may now tell me what I'm doing wrong (Questions?)

