



Center for Translational  
Molecular Medicine

Translating science into better healthcare

# Building a scalable Galaxy cluster for biomedical research in The Netherlands

David van Enckevort

Galaxy Community Conference 2014, Baltimore



# Galaxy adaptation for a clinical setting

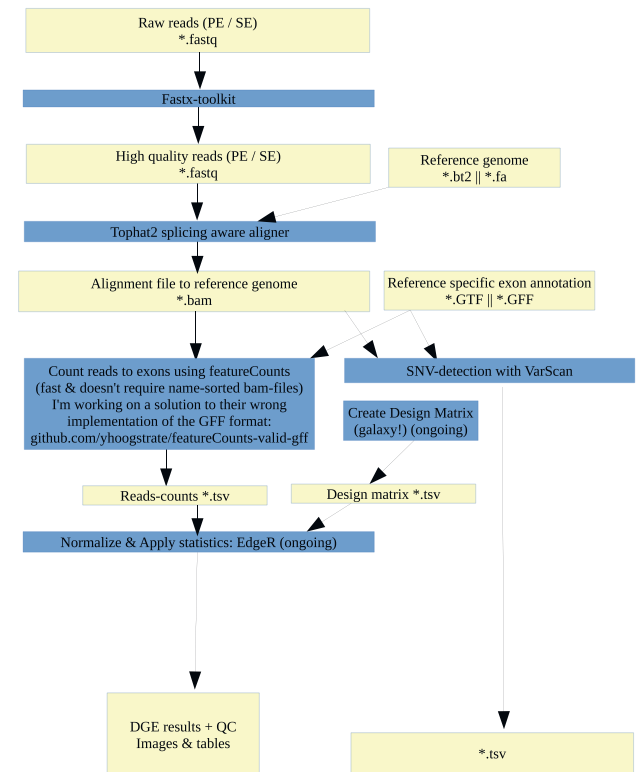
| Requirement                       | Addressed by   |
|-----------------------------------|--|
| Provenance                        | Galaxy   |
| Ease of administration            | Tool Shed / hosting company                                |
| Stable system                     | DTAP instances, Tool Shed, stable releases                 |
| Service levels                    | Hosting company, skilled helpdesk / redundant architecture |
| Certified security of data        | Architecture / hosting company certification               |
| <b>Scalability</b>                | <b>Scalable architecture / CloudMan</b>                    |
| <b>Performance</b>                | <b>HPC architecture</b>                                    |
| <b>Separation between studies</b> | <b>Multi-tenant Cloud architecture / CloudMan</b>          |
| Resource accounting               | Galaxy reporting module, quota & scheduler                 |
| Single Sign On                    | OpenConext (SAML)  |

# Typical Use Cases: NGS, Proteomics

*e.g. RNA-Seq pipeline used for measuring gene expression*

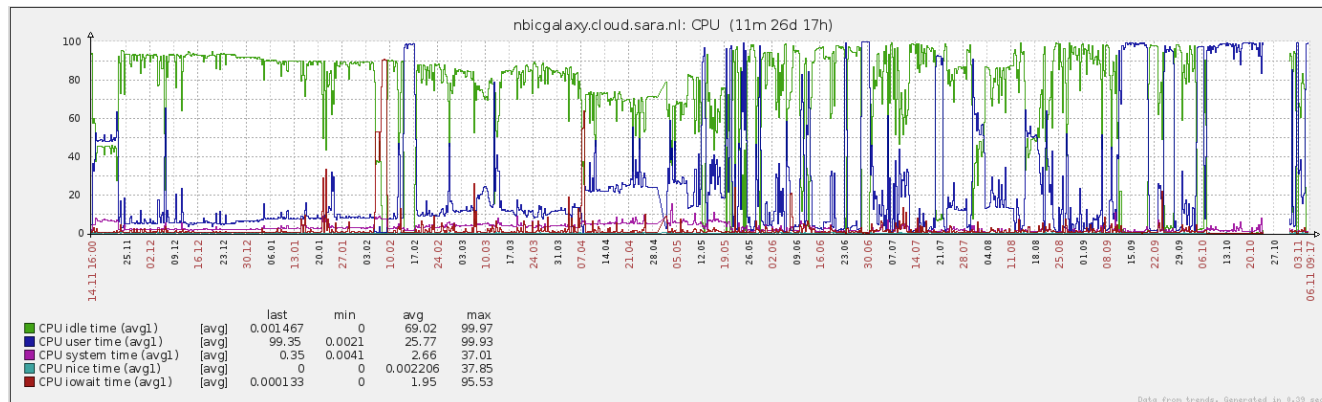
Pipelines are often I/O intensive, because of:

- Input data size typically many gigabytes
- Five-fold increase of the data size in the intermediate files



# Bottlenecks for performance

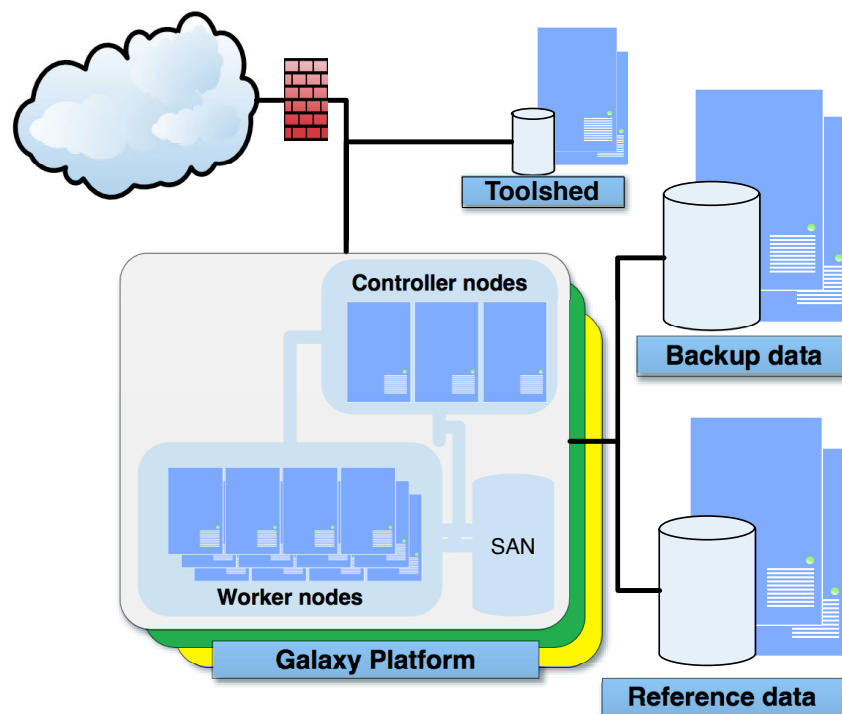
- Based on metrics collected from the NBIC public Galaxy we identified a few key factors:
  1. I/O performance (input/output from storage / hard drive)
  2. Lots of memory
  3. Large storage, allowing for future growth



# Architecture

Key design features:

- **HPC hardware**
- Redundant
- Multi-tenant
- **Multi-tiered storage**
- **Dedicated NFS server**
- Elastically scaling through cloud technology



# CTMM-TraIT partners

netherlands eScience center



University Medical Center  
Utrecht



VU university medical center



PHILIPS



nbic



UMC St Radboud



GENALICE  
TECHNOLOGY FOR PEOPLE & SCIENCE

LEIDS UNIVERSITAIR MEDISCH CENTRUM



SURF SARA



CORDYS  
The Enterprise Cloud Platform



OpenClinica®  
Open Source for Clinical Research



Vancis



Maastricht UMC+  
azM | Maastricht University

