

LifePortal – the Galaxy based portal for life sciences at University of Oslo

Nikolay Vazov¹, Katerina Michalickova¹, Georgios Magklaras^{1,2}, Gard Thomassen¹, Hans A. Eide¹

¹University Center for Information Technology (USIT), University of Oslo,

²Biotechnology Center of Oslo & Norwegian Center for Molecular Medicine, University of Oslo

As the demands for simplified and user-centric interfaces to computational resources are increasing, so is the demand for a wider range of applications and tools presented through these interfaces. We selected the Galaxy platform to provide an interface to our high performance computing resources and life sciences software.

The production server release for the LifePortal is set for October 1st, 2013. The LifePortal includes services currently provided by a portal for bioinformatics applications - the Biportal (www.biportal.uio.no).

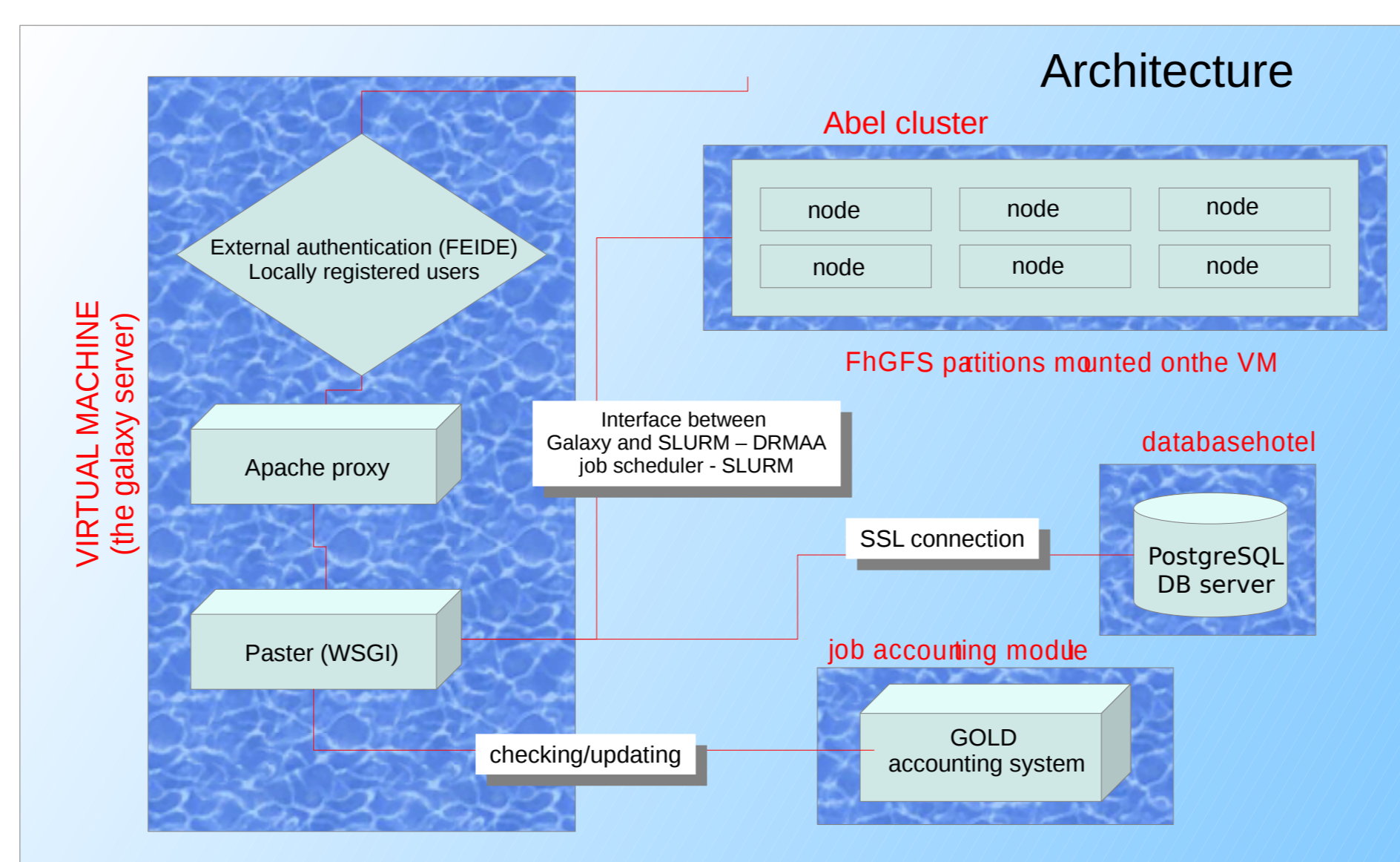
Despite successfully hosting several production Galaxy instances on a single server, we had to introduce modifications to the Galaxy distribution so it can be tailored for our HPC production environment. The adaptations fall into three categories - security, computer cluster job submission and accounting.

The LifePortal will make use of the Norwegian national infrastructure for scientific computing (www.notur.no), specifically the Abel computing cluster at University of Oslo. We are using the Norwegian federated authentication system FEIDE (www.feide.no) to ensure compliance with the terms for usage. We implemented this feature alongside the internal Galaxy user management. Additionally, the Galaxy database has been outsourced to a database hotel using an SSL connection.

The LifePortal Galaxy server submits jobs to the Abel compute cluster using the SLURM batch scheduler system (slurm.schedmd.com). This feature provides a user-friendly interface to our high performance computing resource. Since the computing cluster has fixed user quotas, our Galaxy server has to communicate with an external accounting system (www.clusterresources.com/products/gold-allocation-manager.php).

UiO fully distributed Galaxy setup

- (1) Galaxy DB – exported to a DB hotel (different host)
- (2) Production based Web server solution (Apache proxy server)
- (1) Web framework running on a VM which mounts the cluster partition, data directory on a cluster partition
- (2) Applications run on the cluster and loaded as modules
- (2) Authentication – FEIDE / local
- (3) Porting of applications running in the cluster to Galaxy
- (4) Separate DB for accounting /resource management



1. SSL to the PostgreSQL server (thanks to Nate Coraor)

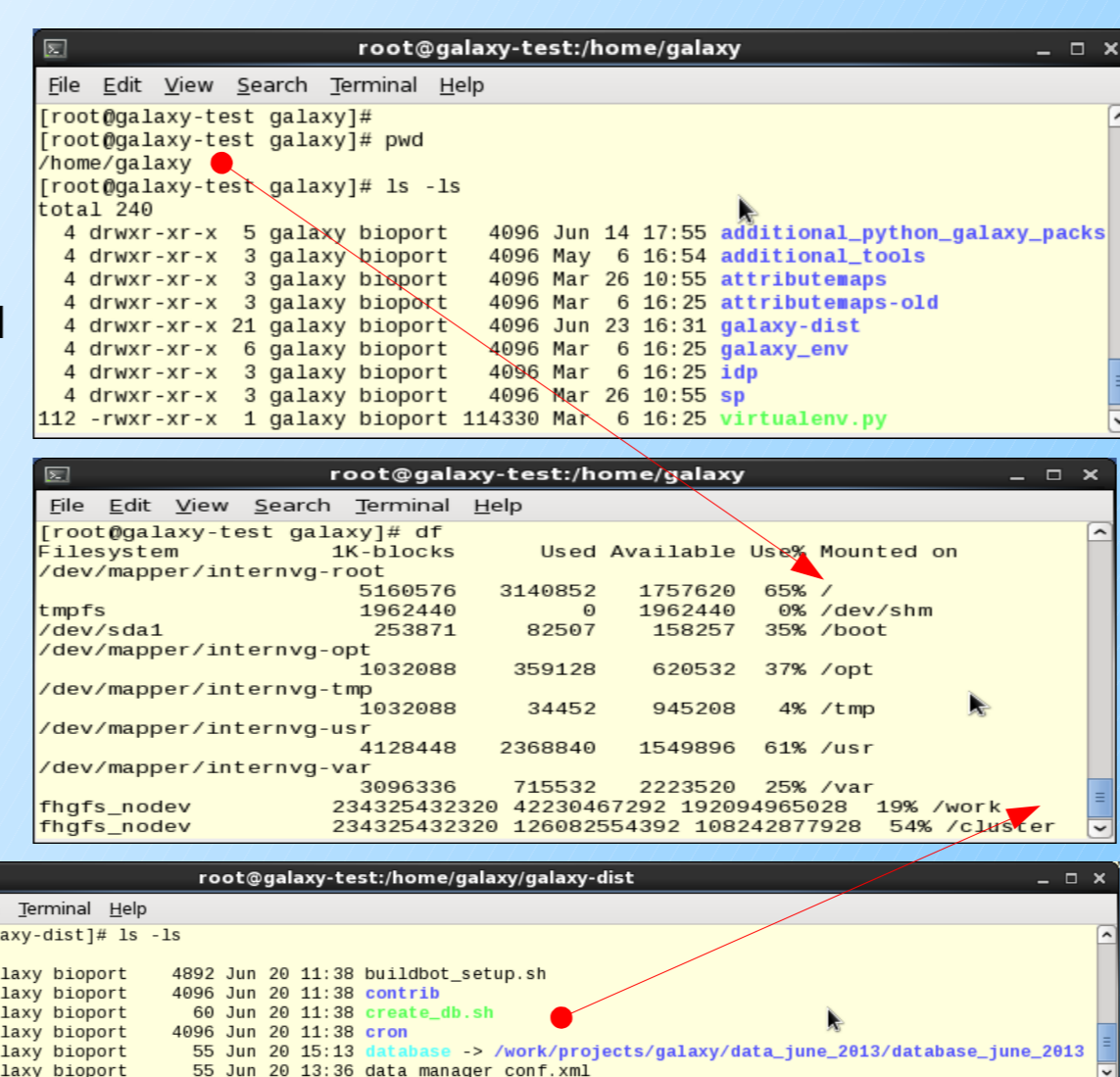
- Install PostgreSQL server (and client) on the Galaxy machine or any other host or order a DB in a DB hotel (UiO)
- Download and recompile *psycpg2-2.0.13 egg*
- In *universe_wsgi.ini* edit the line
`database_connection = postgres://<dbuser>:<password>@<dbhost>:5432/<dbname>?`
`sslmode=require`

2. Apache configuration

- Install Apache with `mod_proxy` and `mod_rewrite` (usually installed automatically)
- Edit the `httpd.conf` file to enable rewrite rules and redirect to `host:port`
- If SSL is needed (recommended), install `mod_ssl`
- Enable SSL
- Edit `universe_wsgi.ini`

3. File architecture

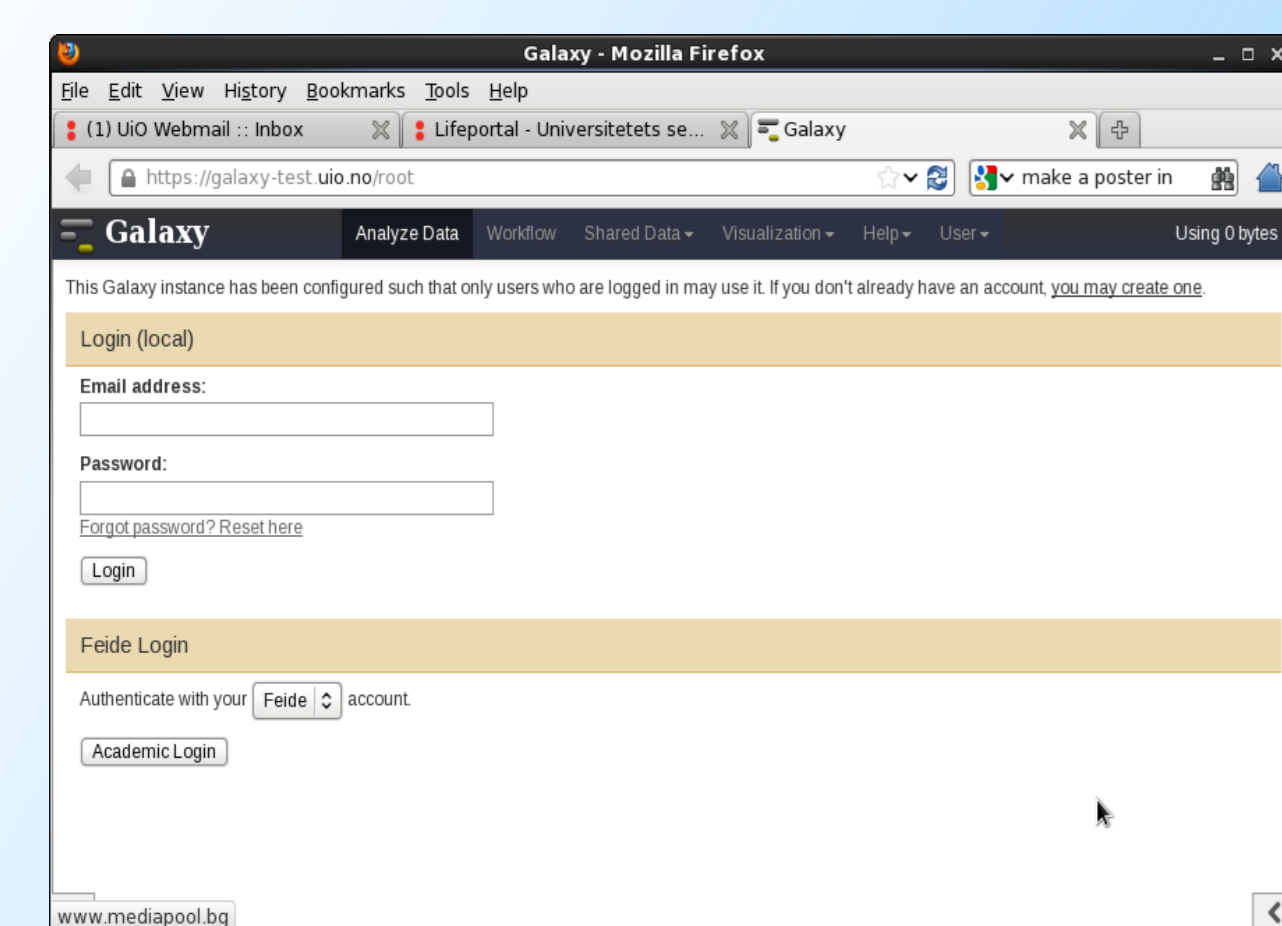
Galaxy on the VM
Mounted fhgfs partitions on the VM
Data /dataset directory



4. Sending jobs to Abel

- `drmaa.py` and `drmaa-python.egg` must be edited to account for all necessary SLURM switches
- DRMAA C-library must be edited and recompiled to accept the modified switches
- SLURM client / munge must be installed and started (running)

5. Local and external (FEIDE) authentication



- Install Pysaml2-0.4.0/ (system-wide is easier)
- Create the keys and metadata files
- Modify all the files in the Galaxy code (about 10) which contain information about the Feide user and connection procedures

6. Porting of applications

- The XML tool definition language supports most of our needs.
- We would welcome if the output “from_work_dir” tag would accept variables and wild cards
- In case of large definition files, it would be beneficial to be able to add graphical elements such as horizontal lines or boxes

7. GOLD – resource management system

- User manipulation
- Resource management
 - jobs
 - projects
 - accounts

