

Reproduce and share: the key to the new generation scientific portal at UiO based on the Galaxy framework

(Bioportal 2.0 - setup and policies)

Nikolay Vazov, Katerina Michalickova, Georgios Magklaras, Thomas Röblitz, Hans Eide

*University Center for Information Technologies (USIT)
University of Oslo
Norway*



The existing service – the Biportal

The screenshot shows the University of Oslo Biportal website. The browser window title is "University of Oslo Biportal - Mozilla Firefox". The address bar shows "https://www.biportal.uio.no". The website features a navigation menu with links for APPLICATIONS, TUTORIAL, CITATIONS, ABOUT, and DOWNLOADS. Below the navigation is a large banner image with the word "biportal" overlaid. The main content area is divided into several sections:

- LOGIN:** Norwegian university users, Other users, New user? Apply for access.
- STATISTICS:** Members logged in: 26, Current submitted jobs: 25, Jobs submitted last week: 188, Total number of users: 4309, Number of jobs in: 2012: 28039, 2011: 35601.
- APPLICATIONS INSTALLED:** A list of software tools including ADSCREENING, AIR-APPENDER, AIR-IDENTIFIER, AIR-REMOVER, AUTODOCK4, BEAST, BEST, BIR, BLAST, BLASTGRABBER, CLOTU, CLUSTEX, FRISCH, GARLI, GAUSSIAN, HERVEPARSER, LAMARC, MAFFT, MCMCPHASE, MIGRATE, MIRA, MODELTEST, MRBAYES, MRMODELTEST, NEWBLER, OPENBUGS, PAML, PAUP, PHASE, PHRED, PHREDPHRAP, PHYLOBAYES, PHYML, POY, PRANK, PROTTTEST, R, RAXML, SIMWALK2, STRUCTURE, TREEFINDER, and UNPHASED.
- UPCOMING EVENTS:** Univ. of Oslo Blindern campus, Oslo, Oct. 8 - 19 intensive course in High Throughput Sequencing technologies and bioinformatics analysis; University of Oslo, Sept. 17-18 Software Carpentry Boot Camp; UiB (Bergen): October-December 2012 ForBio course Phylogenetic methods; September 17-21 2012, at the Natural History Museum, Oslo, Norway Introduction to Bioinformatics for Biosystematics; Madrid, Spain, 18-22 June 2012 Molecular Phylogenetics course.
- SYSTEM MESSAGES:** [25.09.2012] Bioportal downtime!; [20.08.2012] New computing cluster; [03.07.2012] Short stop of Bioportal web Wednesday at 18:00 CEST; [25.06.2012] Short stop of bioportal web Wednesday at 19:00 CEST; [14.06.2012] Downtime Friday 15. at 12:00 CEST.
- NEWS:** [30.07.2011] RAxML optimised; [27.07.2011] Treefinder upgraded; [12.05.2011] OpenBUGS installed; [09.12.2010] MIRA documentation now updated; [09.12.2010] PRANK documentation now updated.



Bioportal features - jobs

The screenshot shows the University of Oslo Bioportal interface in a Mozilla Firefox browser. The page title is "University of Oslo Bioportal - Mozilla Firefox" and the URL is "https://www.bioportal.uio.no/general/submitjob.php".

The main navigation bar includes: **bioportal** logo, **UNIVERSITY OF OSLO** logo, and menu items: **APPLICATIONS**, **TUTORIAL**, **CITATIONS**, **ABOUT**, **DOWNLOADS**.

Left sidebar menu:

- LOG OUT**
- STATISTICS**
 - Members logged in: 25
 - Current submitted jobs: 27
 - Jobs submitted last week: 187
 - Total number of users: 4310
 - Number of jobs in:
 - 2012: 28039
 - 2011: 35601
- PROJECTS**
- FILE ADMIN**
 - Upload new files to the system
 - Best (selected)
 - Clear all
 - Add file
- SUBMIT JOB**
- JOB ADMIN**
- POST MESSAGE**
 - News
 - Seminar
 - Course
 - Sys. news
- NEW USER ADMIN**
- USER ADMIN**
- RESOURCE ADMIN**
- APP ADMIN**
- BLAST DB ADMIN**
- WHITELIST ADMIN**
- BLACKLIST ADMIN**

Main content area:

- Project:** Choose project: Best
- Input files:** Choose input files from project 'Best':

Filename	Changed	KB	Expires	select
COX1prot.nexus	2011-10-31	16	2012-07-05	<input type="checkbox"/> Edit

Total size: 16 KB
Order files by: File Name
- Application:** Choose application (you can limit the list to a category and/or a computing resource):

App category	Resource	Application
all applications	any resource	BEAST@abel
- Submit job:** Job name:

Load situation on resources:

Titan status: Bioportal queues

Queue	Running (%)	Waiting (%)
bioportal (580)	~25	~15
lowpri (4082)	~5	~15

% usage (Thu Sep 27 11:48:06 2012)

Numbers in parentheses are the number of CPUs available for the queue. Click here to see all queues.



Biportal features - files

The screenshot shows the University of Oslo Biportal website in a Mozilla Firefox browser. The page title is "University of Oslo Biportal - Mozilla Firefox". The address bar shows the URL "https://www.biportal.uio.no/general/files.php". The page features a navigation menu with "APPLICATIONS", "TUTORIAL", "CITATIONS", "ABOUT", and "DOWNLOADS". On the left side, there are several administrative links: "LOG OUT", "STATISTICS" (showing 27 members logged in, 25 current jobs, 187 jobs submitted last week, and 4309 total users), "PROJECTS", "FILE ADMIN" (with an upload button and a dropdown menu set to "Best"), "SUBMIT JOB", "JOB ADMIN", "POST MESSAGE" (with radio buttons for "News", "Seminar", and "Course"), "NEW USER ADMIN", "USER ADMIN", "RESOURCE ADMIN", "APP ADMIN", "BLAST DB ADMIN", "WHITELIST ADMIN", and "BLACKLIST ADMIN". The main content area is titled "Your current files" and contains a table of files with columns for "Filename", "Project", "Changed", "KB", "Expires", and "select". The table lists 15 files, including "COX1prot.nexus", "paster.log", "delete_blastdb", "fasta_simple", "Fall.txt", "all.txt", "peddum2010.txt", "delete_blastdb", "mainparams_vulpes.txt", "mainparams", "Aurora", "All_tabelimit.txt", "All_STRUCTURE.txt", and "vulpes_STRUCTURE.txt". The total size of the files is 4919 KB. Below the table, there is a dropdown menu for "Order Files By" set to "Project" and a section for "Actions" with buttons for "Clear all", "Download", "Delete", "View", "Move to project:", and a dropdown menu set to "Best".

Filename	Project	Changed	KB	Expires	select
COX1prot.nexus	Best	2011-10-31	16	2012-07-05	<input type="checkbox"/> Edit
paster.log	Migrate	2011-10-21	3765	2012-08-04	<input type="checkbox"/> Edit
delete_blastdb	Mira	2011-01-07	1	2011-11-05	<input type="checkbox"/> Edit
fasta_simple	MrBayes	2011-07-27	0	2013-01-19	<input type="checkbox"/> Edit
fasta_simple	Phylogenus	2011-07-25	0	2011-11-25	<input type="checkbox"/> Edit
COX1prot.nexus	Phylogenus	2011-07-24	16	2011-11-25	<input type="checkbox"/> Edit
COX1prot.nexus	Protest	2011-03-09	16	2012-02-11	<input type="checkbox"/> Edit
Fall.txt	Rproj	2010-11-19	2	2011-07-15	<input type="checkbox"/> Edit
all.txt	Rproj	2010-11-19	451	2011-08-23	<input type="checkbox"/> Edit
peddum2010.txt	Rproj	2010-11-19	128	2011-07-15	<input type="checkbox"/> Edit
delete_blastdb	Rproj	2011-01-07	1	2011-07-15	<input type="checkbox"/> Edit
mainparams_vulpes.txt	Structure	2011-11-07	1	2012-03-07	<input type="checkbox"/> Edit
mainparams	Structure	2011-11-07	1	2012-03-07	<input type="checkbox"/> Edit
Aurora	Structure	2011-10-18	3	2012-02-18	<input type="checkbox"/> Edit
All_tabelimit.txt	Structure	2011-11-07	180	2012-03-07	<input type="checkbox"/> Edit
All_STRUCTURE.txt	Structure	2011-11-07	182	2012-03-07	<input type="checkbox"/> Edit
vulpes_STRUCTURE.txt	Structure	2011-11-07	156	2012-03-07	<input type="checkbox"/> Edit
Total size:			4919	KB	



The two keywords: *reproducibility* and *shareability*

- Galaxy allows for file sharing (both input or output files) using complex group/role structure
- Galaxy uses workflows which allow the users to reproduce the entire process (set of jobs) over a number of input files
- Galaxy allows for workflow sharing



What is Galaxy?

- * Presently geared towards genomics – 400 standard programs and further 2100 prepared (the new toolshed)
- * Open-source, developed by core team of ~15 people, and a large community
- * Basically a web-based system for running any kind of programs on any kind of data
- * Runs on laptops, clusters and clouds



Galaxy – origins (http://galaxy.psu.edu)

The screenshot shows a Windows Internet Explorer browser window displaying the Galaxy Project website. The address bar shows the URL <http://galaxy.psu.edu/>. The page content includes the following sections:

- Use Galaxy**: Accompanied by a thumbnail of the Galaxy interface. Below it is the text: [Use the free public server](#).
- Get Galaxy**: Accompanied by a large blue downward-pointing arrow. Below it is the text: [Install locally or in the cloud](#). Below this is a link: [Search all resources](#).
- Learn Galaxy**: Accompanied by a thumbnail of a presentation slide titled "Advanced fastQ manipulation". Below it is the text: [Screencasts, Galaxy 101, ...](#).
- Get Involved**: Accompanied by a thumbnail of a mailing list or wiki page. Below it is the text: [Mailing lists, Tool Shed, wiki](#).

At the bottom of the page, there is a paragraph of text: "The [Galaxy Team](#) is a part of [BX](#) at [Penn State](#), and the [Biology](#) and [Mathematics and Computer Science](#) departments at [Emory University](#). The Galaxy Project is supported in part by [NSF](#), [NHGRI](#), [The Huck Institutes of the Life Sciences](#), [The Institute for CyberScience at Penn State](#), and [Emory University](#)."

The browser window also shows the Windows taskbar at the bottom with the Start button, several application icons, and the system tray displaying "Internet | Protected Mode: On", "100%", "NO", and the date "11:14 12.09.2012".



What does a Galaxy page look like?

The screenshot shows a web browser window titled "Galaxy - Windows Internet Explorer" with the address bar displaying "https://nofish.titan.uio.no/root". The browser interface includes a menu bar with "Convert" and "Select", a "Favorites" bar, and a "Galaxy" toolbar. The main content area is titled "MCMCglmm Univariate Animal Model (version 1.0.1)".

Left Panel (Tools):

- NICK'S TOOLS
- BLAST+
- Running CLC
- Plink
- Nicks MCMCglmm Tools
 - MCMCglmm Univariate Animal Model R version. See full pdf instruction manual.
 - MCMCglmm Univariate Repeated Measurement Animal Model R version. See full pdf instruction manual.
 - MCMCglmm Bivariate Repeated Measurement Animal Model R version. See full pdf instruction manual.
- Nicks Asreml Tools
- Nicks NGS Expression Tools
- Nicks Simulation Tools
- Crimap Tools
- Nicks other R Tools
- rQuant
- palmapper
- EMBOSS
- Blast2Go
- Get Data
- Send Data
- ENCODE Tools
- Lift-Over
- Text Manipulation
- UNIX Tools
- Convert Formats
- FASTA manipulation
- Filter and Sort
- Join, Subtract and Group

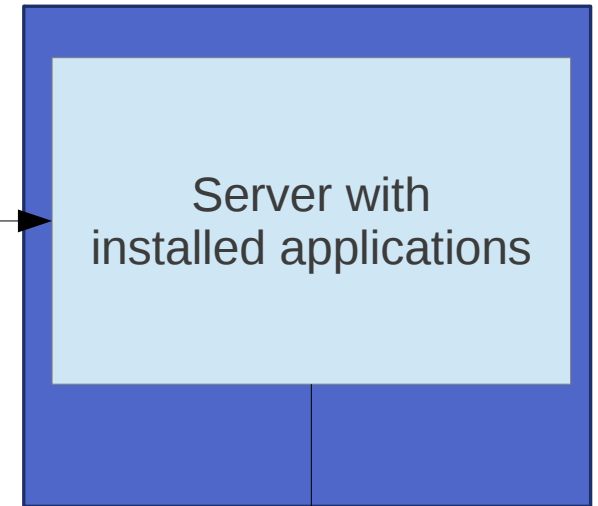
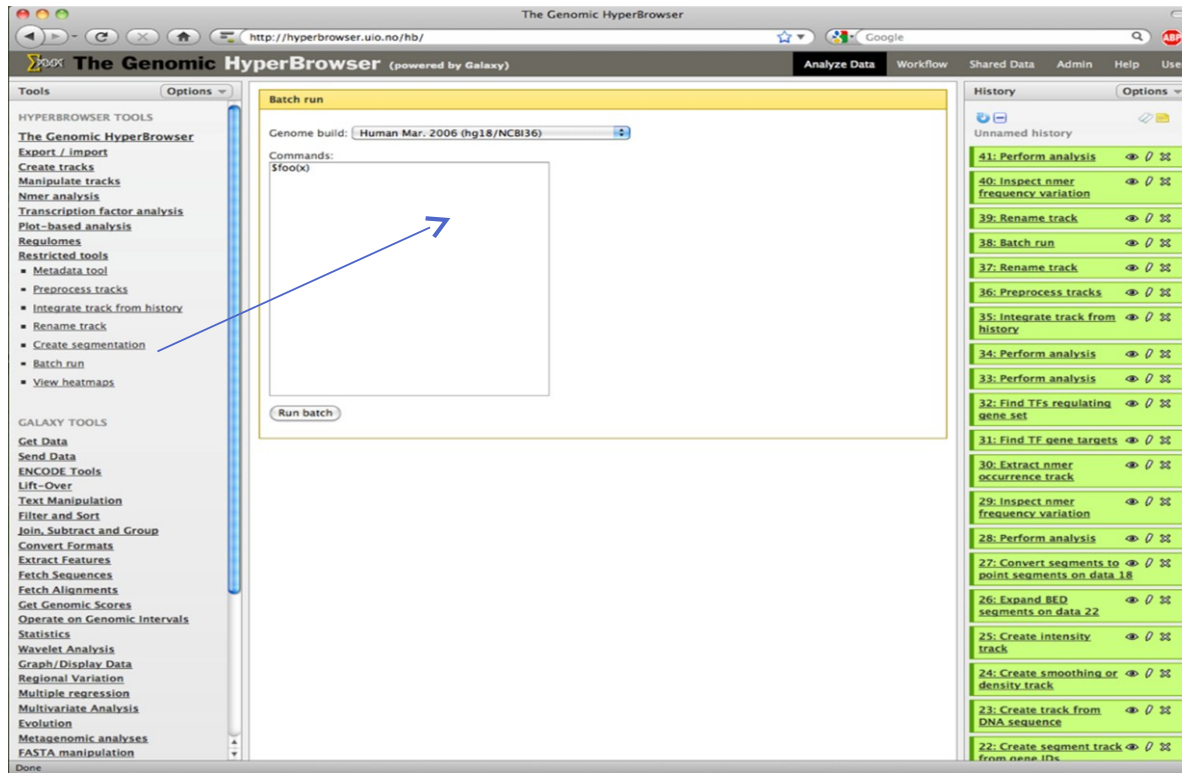
Main Configuration Panel:

- Main data frame WITH column headers.:** 97: problems Nikolay test (missing data=NA. animal MUST be a header)
- y term eg. distance:** distance
- column number containing y term eg. 7:** 7
- fixed terms eg. Sex+origin:** (empty)
- random terms ie. animal:** animal (Assumes only 1 random term = animal)
- Guess of proportion explained by the animal effect:** 0.5 (limit of 0 to 1, with residual must add to 1.0)
- Guess of proportion explained by the residual effect:** 0.5 (limit of 0 to 1, with animal must add to 1.0)
- Degree of belief parameter:** 1.0 (Limit of 1 for univariate [smallest degree of belief that provides a 'proper' prior avoiding numerical problems] to ?)
- Number of run length iterations:** 1000000 (limit of 1 to 10000000)
- Burn in iterations:** 30000 (limit of 1 to 10000000)
- Thinning interval:** 50 (Iteration estimates retained every X iterations, limit of 1 to 10000000)
- Pedigree data frame with NO column headers in order animal sire dam:** 97: problems Nikolay test

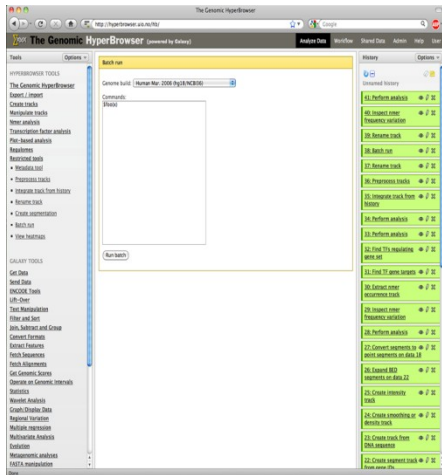
Right Panel (History):

- Clone of 'MCMCglmm AbSwim' shared by 'nicholas.robinson@nofi ma.no' (active items only) 41.0 Kb
- 99: stdout Nikolay test
- 98: stderr Nikolay test
- 97: problems Nikolay test
- 96: renumbered ped file Nikolay test
- 95: response predictions w CI distance
- 94: stdout distance
- 93: stderr distance
- 92: Posterior distribution variance components distance
- 91: Posterior distribution fixed effect distance
- 90: Solutions distance
- 89: PostModePostHeritability distance
- 88: HPDIntervalPostHeritability distance
- 87: HPDInterval distance

A *very* simple view of Galaxy



A simple view of Galaxy



- Scripts generating the web interface to the applications (mako scripts using xml files as input)
- Wrapper scripts launching applications

applications

Built-in web server

Built-in SQLite
• for user manipulation
• for user data manipulation

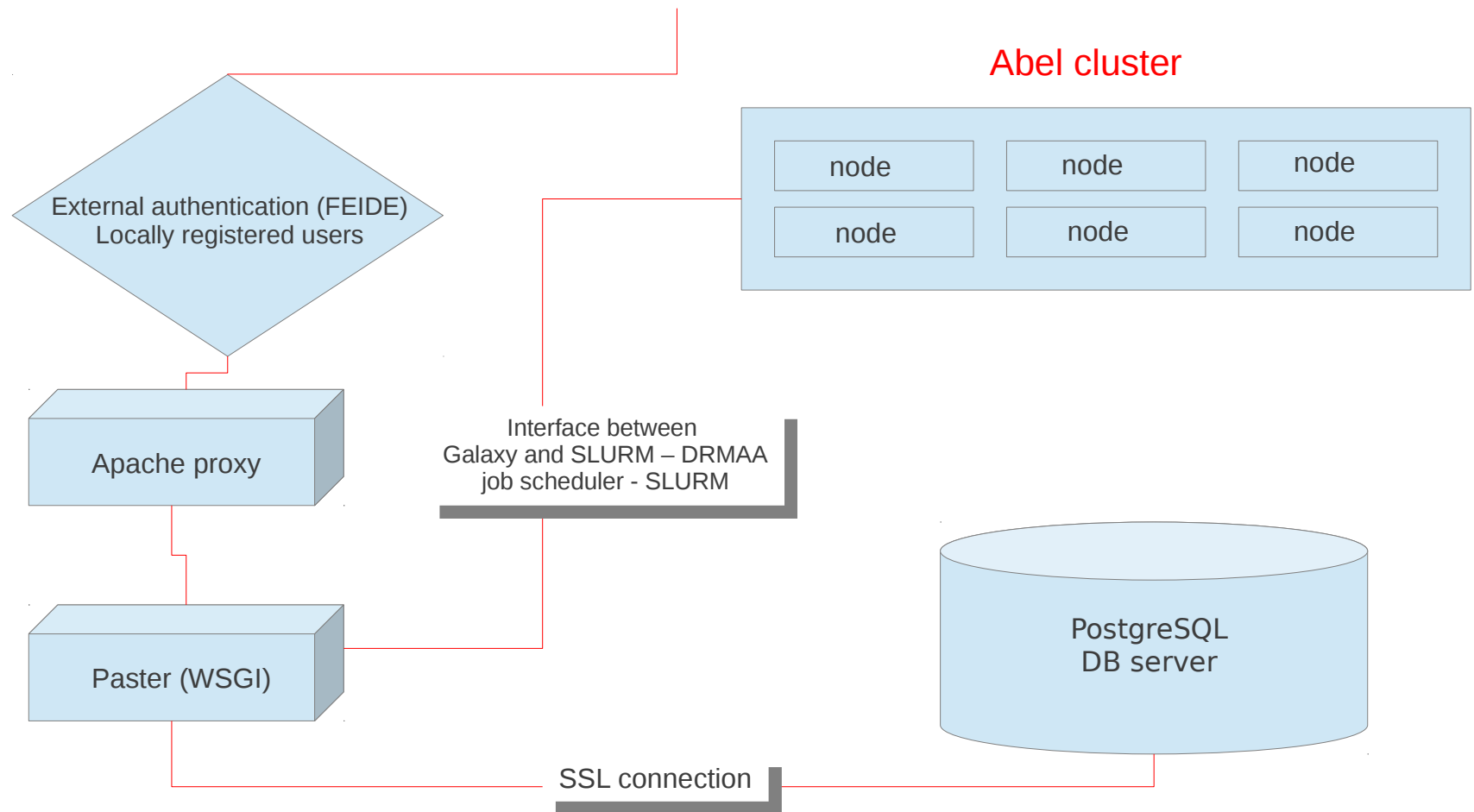


UiO fully distributed Galaxy setup

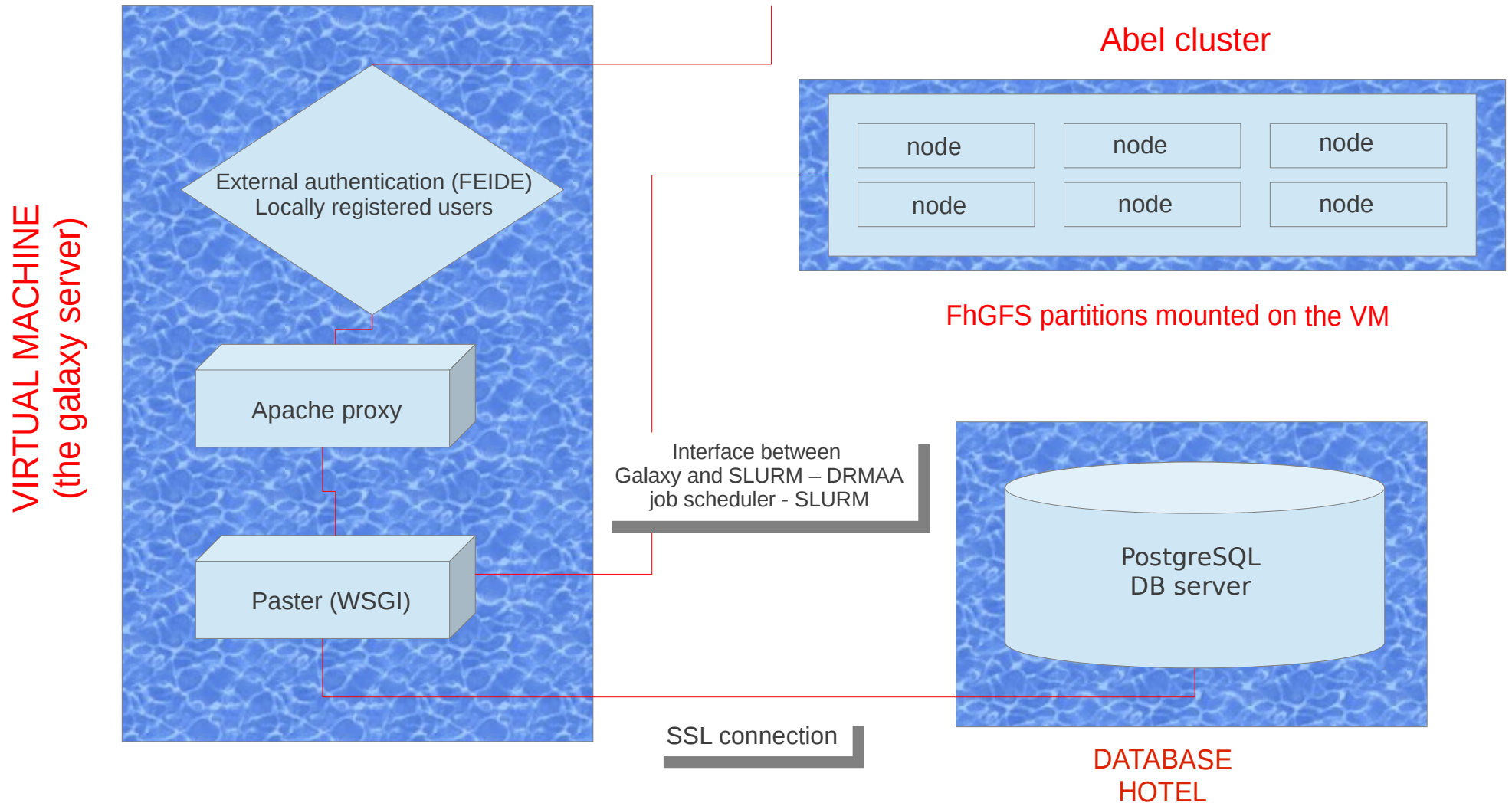
- Galaxy DB – exported to a DB hotel
- Authentication – FEIDE / local
- Production based Web server solution (Apache proxy server)
- Applications run on the cluster and loaded as modules
- Data (dataset) directory on a cluster partition
- Separate DB for accounting /resource management
- Web framework running on a VM which mounts the cluster partition



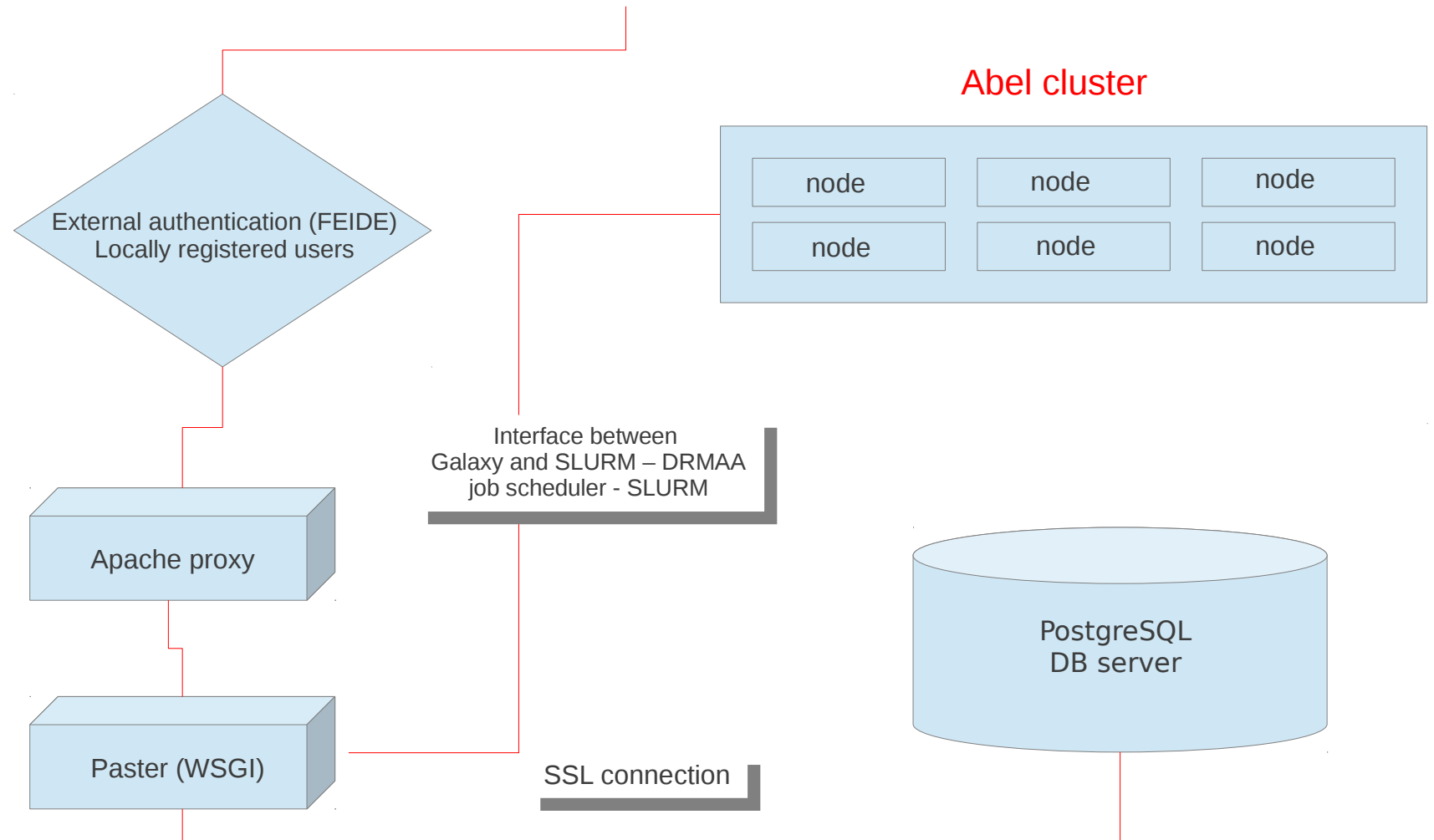
Galaxy in Abel - configuration



Galaxy in Abel – physical configuration



Apache & Paster

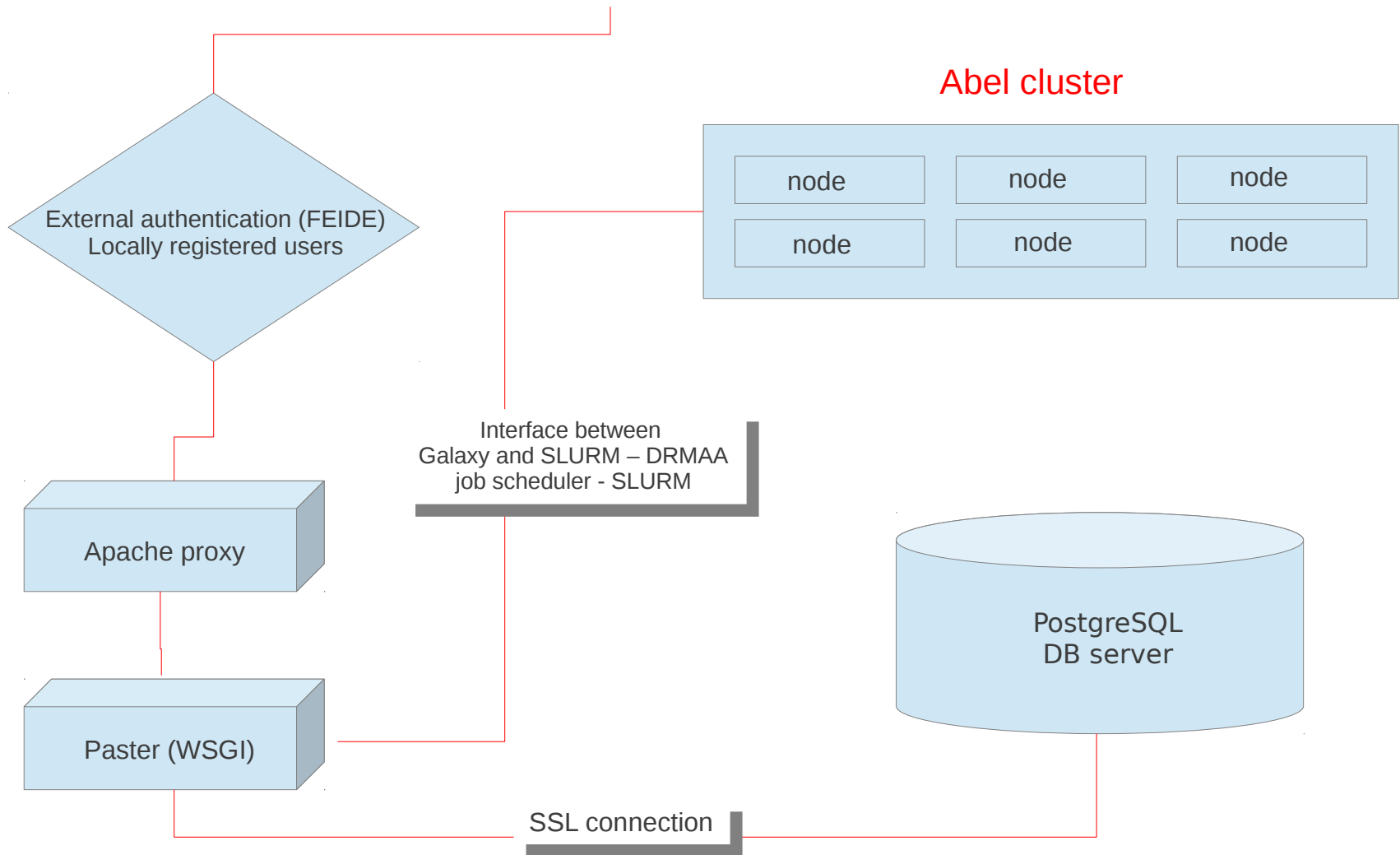


Apache configuration

- 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`
 - `localhost:8080`
- If SSL is needed (recommended), install `mod_ssl`
- Enable SSL (there are a couple of ways to do this, here's one of them)
 - Edit `httpd.conf` to redirect to port 443
 - Generate the keys and place them in the proper location (you may use the same keys for authentication later)
 - Edit `ssl.conf` such that it redirects to `localhost:8080`
- Edit `universe_wsgi.ini` as follows :
 - `host = 127.0.0.1`
 - `port = 8080`



Paster & Database



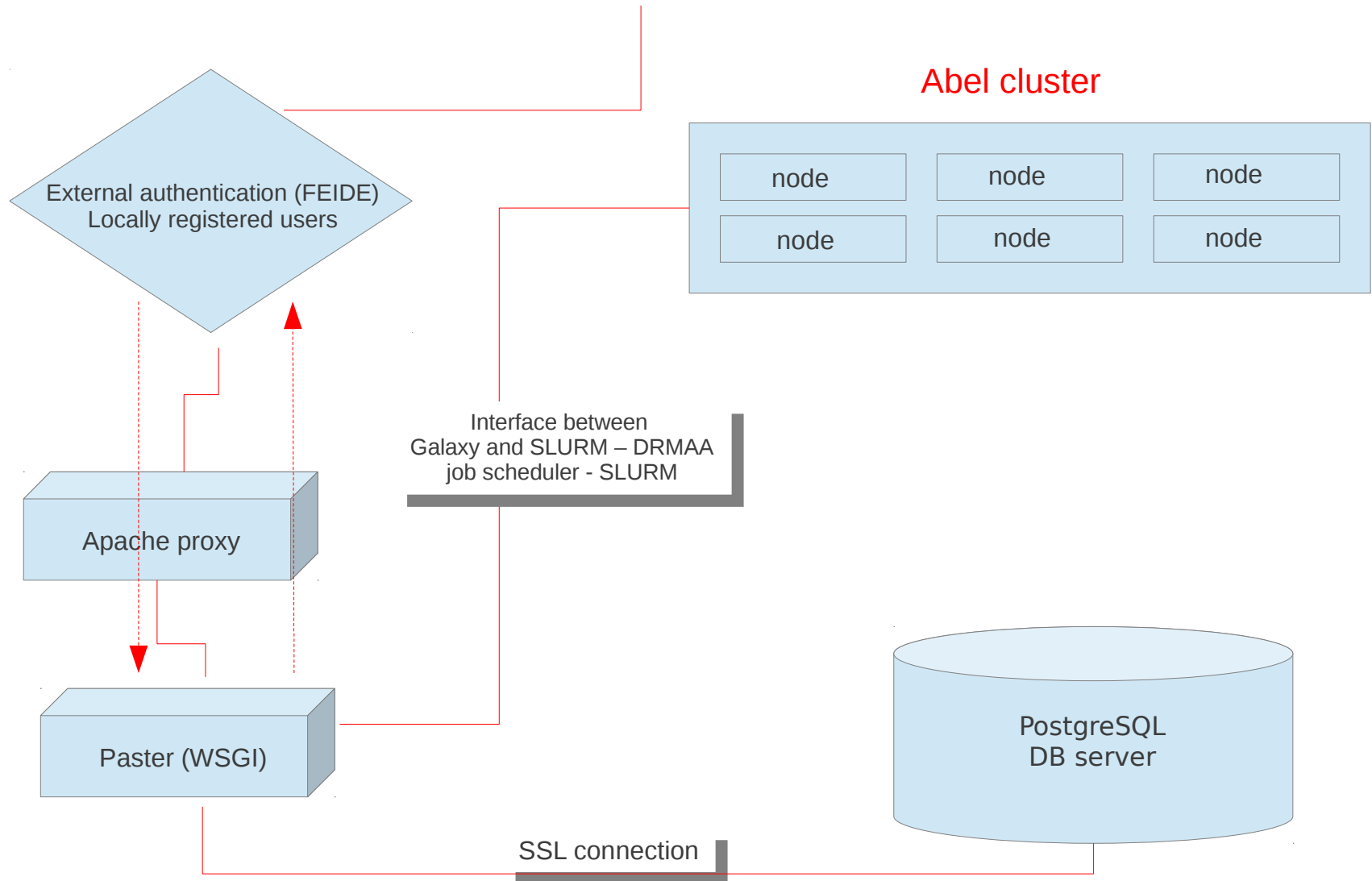
SSL to the PostgreSQL server

(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 *psycopg2-2.0.13 egg*
- In *universe_wsgi.ini* edit the line
database_connection = postgres://<dbuser>:<password>@<dbhost>:5432/<dbname>?sslmode=require



External authentication



Authentication (Feide)

(thanks to Feide team, Roland Hedberg, Paul Boddie)

- Install Pysaml2-0.4.0/ (system-wide is easier)
- Create the keys and metadata files
 - Public/private key for the machine with Galaxy server (virtual machine)
 - Create the metadata files for the SP
 - Generate the XML file containing the data of Service provider (from the metadata file)
 - Get the XML file containing the data of the Identity provider
- Modify all the files in the Galaxy code (about 10) which contain information about the Feide user and connection procedures



Authentication page

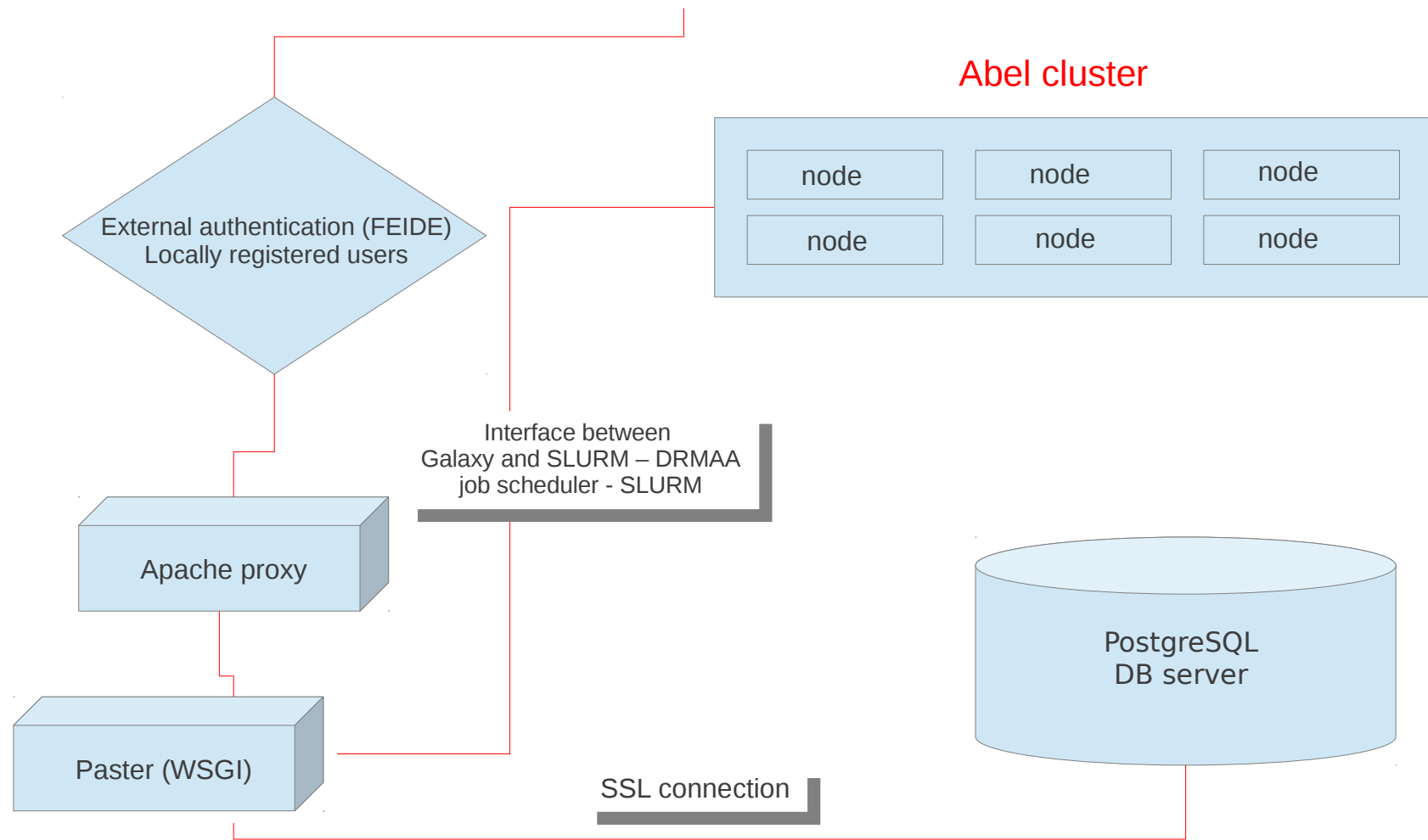
The screenshot shows a web browser window titled "Galaxy - Mozilla Firefox" with the address bar displaying "https://galaxy-test.uio.no/root". The page header includes the "Galaxy" logo and navigation menus for "Analyze Data", "Workflow", "Shared Data", "Visualization", "Help", and "User". A status bar at the top right indicates "Using 0 bytes".

The main content area features a message: "This Galaxy instance has been configured such that only users who are logged in may use it. If you don't already have an account, [you may create one](#)." Below this, there are two login sections:

- Login (local)**: Includes input fields for "Email address:" and "Password:", a link for "Forgot password? Reset here", and a "Login" button.
- Feide Login**: Includes the text "Authenticate with your Feide account." and an "Academic Login" button.

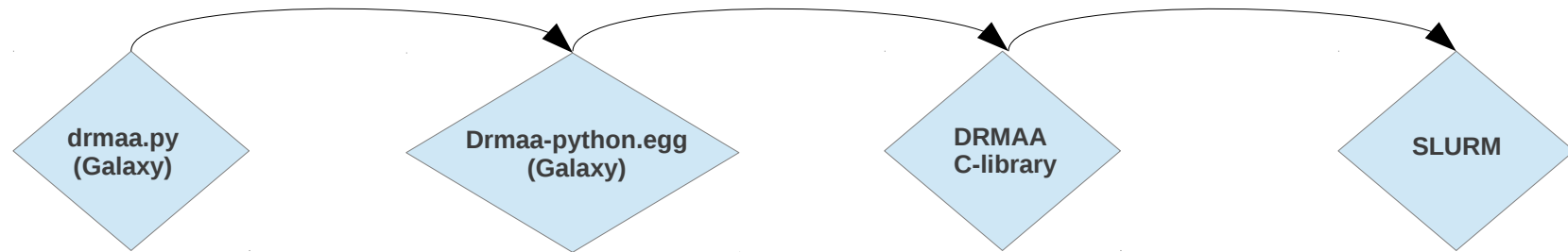


Galaxy & SLURM (Abel's queueing system)



Sending jobs to Abel

(thanks to Bjørn-Helge Mevik)



- 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)



Accounting / Authentication policies

- One system user (`galaxy`) sending jobs to the cluster (not real users)
- All users shall be inserted into the Galaxy user DB
- Galaxy quota management system will be used for space limits
- Customized setup for matching Galaxy users vs Abel projects before running the jobs
- Three types of users
 - A: Visiting user – instant access to limited resources from the portal's pool
 - B: Small project access – locally evaluated access to limited resources from the portal's pool
 - C: Large projects – RFK evaluated projects with access to resources outside the portal's pool



Project (account) selection page

The screenshot shows the Galaxy web interface in a Mozilla Firefox browser window. The browser's address bar displays the URL `https://galaxy-test.uio.no/root`. The Galaxy header includes navigation tabs for 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User', along with the text 'Using 336.5 KB'.

The main content area is titled 'Tell current time (version 1.0.0)'. It features a 'Source file:' section with a dropdown menu showing '3: Tell current time on data 2'. Below this, a message states: 'You have logged as user: *nikolaiv*. Please select the project (account) where you will be running your current job.' A dropdown menu with the text '--select--' is provided for selection, and an 'Execute' button is located below it.

The left sidebar contains a 'Tools' section with a search bar and a list of tool categories, including 'myTools', 'Get Data', 'Send Data', 'ENCODE Tools', 'Lift-Over', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Convert Formats', 'Extract Features', 'Fetch Sequences', 'Fetch Alignments', 'Get Genomic Scores', 'Operate on Genomic Intervals', 'Statistics', 'Wavelet Analysis', 'Graph/Display Data', 'Regional Variation', 'Multiple regression', 'Multivariate Analysis', 'Evolution', 'Motif Tools', 'Multiple Alignments', 'Metagenomic analyses', 'FASTA manipulation', 'NGS: QC and manipulation', 'NGS: Mapping', 'NGS: Indel Analysis', 'NGS: RNA Analysis', 'NGS: SAM Tools', 'NGS: GATK Tools (beta)', 'NGS: Peak Calling', 'NGS: Simulation', 'SNP/WGA: Data: Filters', 'SNP/WGA: QC: LD: Plots', 'SNP/WGA: Statistical Models', 'Phenotype Association', and 'VCF Tools'. At the bottom of the sidebar, there are 'Workflows' and 'Log messages' options.

The right sidebar shows a 'History' section with a refresh icon and a settings icon. It lists 'Unnamed history' (7.1 KB) and three entries: '3: Tell current time on data 2', '2: Tell current time on data 1', and '1: fasta_simple'. Each entry has icons for viewing, deleting, and refreshing.



Thank you

<http://www.usit.uio.no/om/organisasjon/uav/itf/intern-doc/galaxy/galaxy-test.html>

n.a.vazov@usit.uio.no

