

# Galaxy on a cluster : user & project management

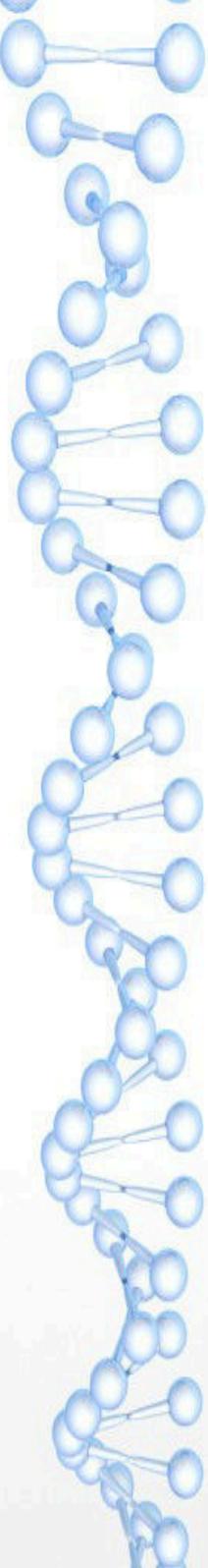
Katerina Michalickova  
Nikolay Vazov

University Center for Information Technologies  
University of Oslo



UiO • University of Oslo





# Prerequisites

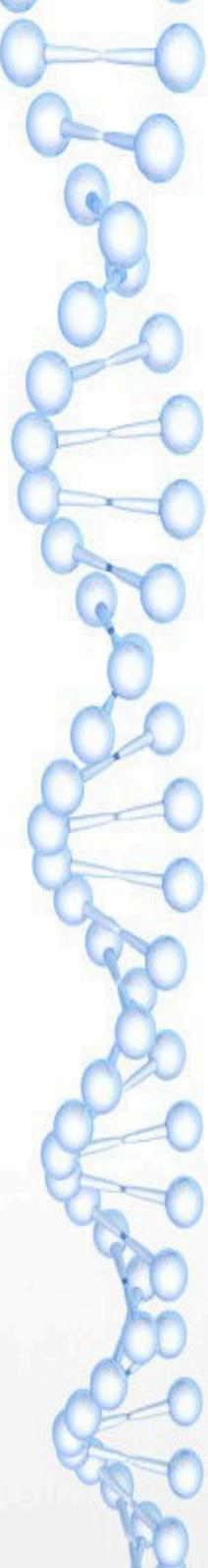
- VirtualBox?

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

- UsersProjects Galaxy VM downloaded?

- <http://folk.uio.no/nikolaiv/GCC2014-Users-Projects.ova>

Password : **UsersProjects2014**



# What are the main cluster features, and requirements?

- Running jobs on a cluster is expensive, requires plenty of resources (processors, disk space, bandwidth, electricity (*yes, the bills are impressive*), and ... daily human expertise)
- Clusters are like banks, they are structured with regard to :
  - **users** : average unix-type users
  - **accounts** (projects) : users belong to accounts
  - **allocations** (granted to each account) which control:
    - *time span validity, e.g 30:12:2014 - 02:02.2015*
    - *use of CPU resources, e.g. 1 000 000 CPU hours*
    - *use of machines : number and type*
    - *job priorities*
    - *etc.*

# Clusters and/vs Galaxy?

or what are our challenges with regard to user DBs

## Cluster

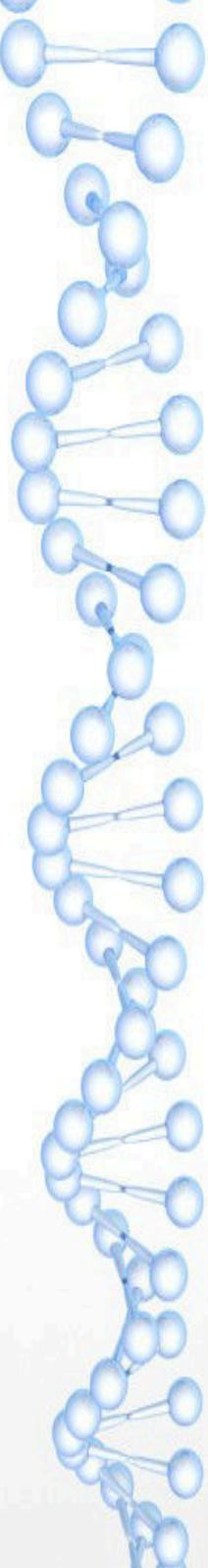
- require complex user DB management
- user DBs are usually dependent on other services responsible for user accounts (LDAP)

## Galaxy

- has a “user-managed” user DB
- does not control user's affiliation at registration

*A very common situation :*





# Clusters and/vs Galaxy?

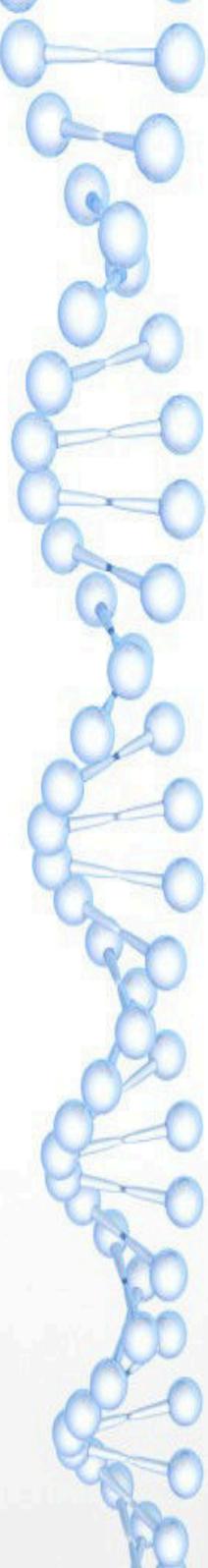
or what are our challenges with regard to resource allocation

## Cluster

- has an integrated accounting system that tracks and manages resource usage
- all cluster users are organized into accounts with respective allocations

## Galaxy

- does not have any accounting system
- the only resource limitations are disk quotas

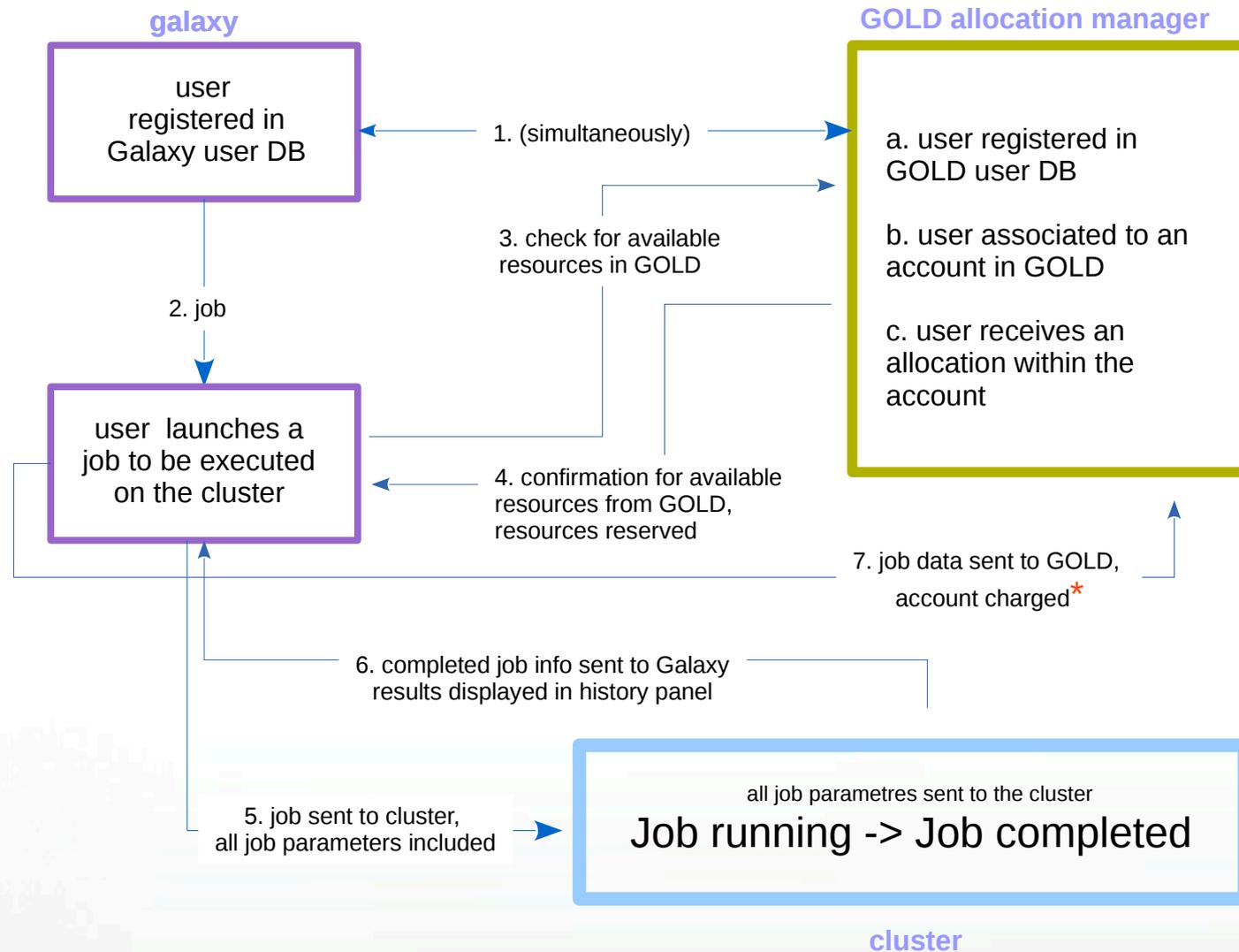


# Clusters not *vs* but *with* Galaxy

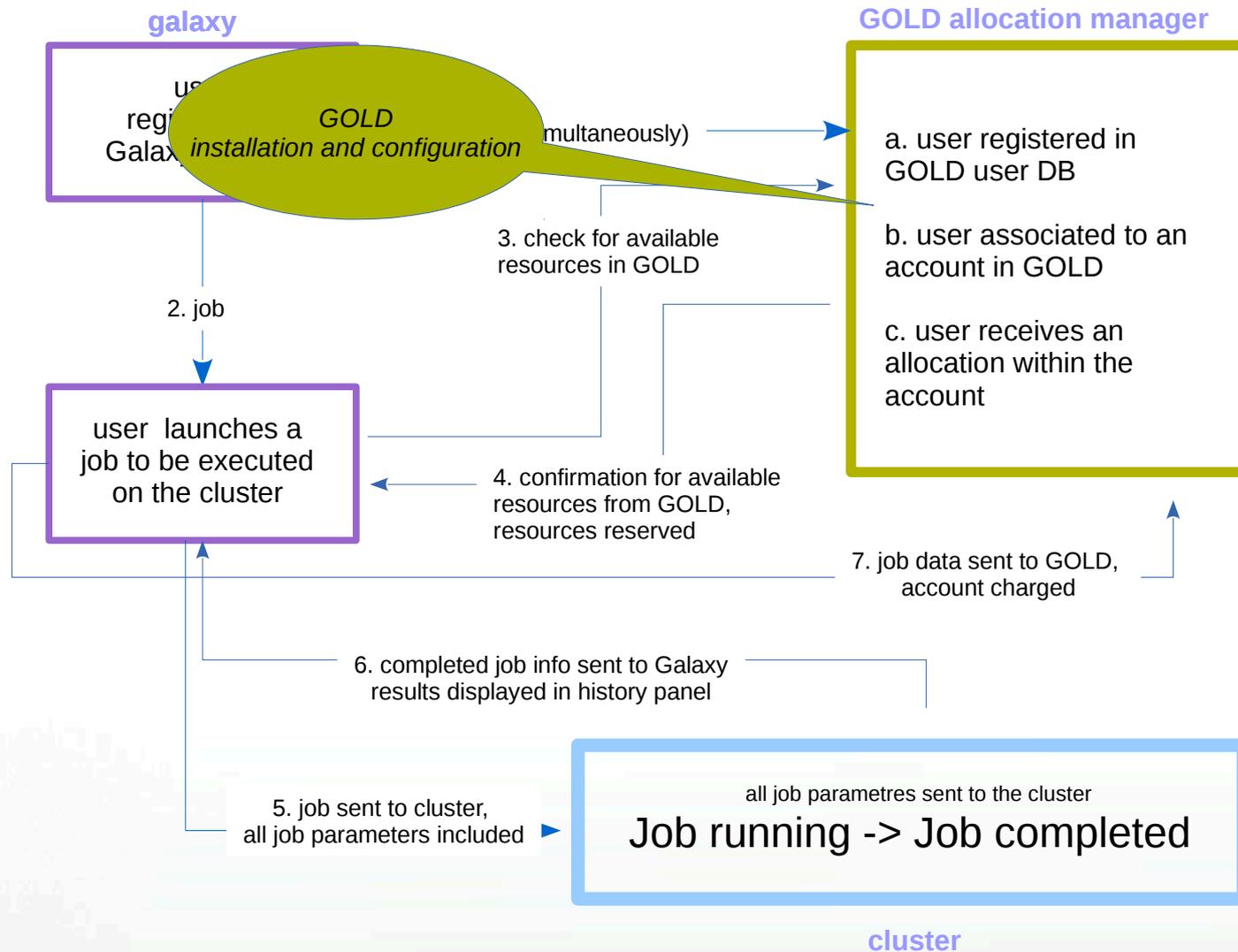
a possible (and working) solution

- all galaxy jobs sent to the cluster belong to one single user : *galaxy*
- information about each Galaxy job is propagated *to* the cluster and *back* to Galaxy when finished; it contains
  - Information about the required/used resources : number of cpu, job duration, memory, etc.
  - Galaxy and Cluster job id
  - Information about the real job owner
- the job data is handled by a separate allocation manager (GOLD) which charges the account of the real Galaxy user

# Clusters and Galaxy : the strategy in (technical) detail

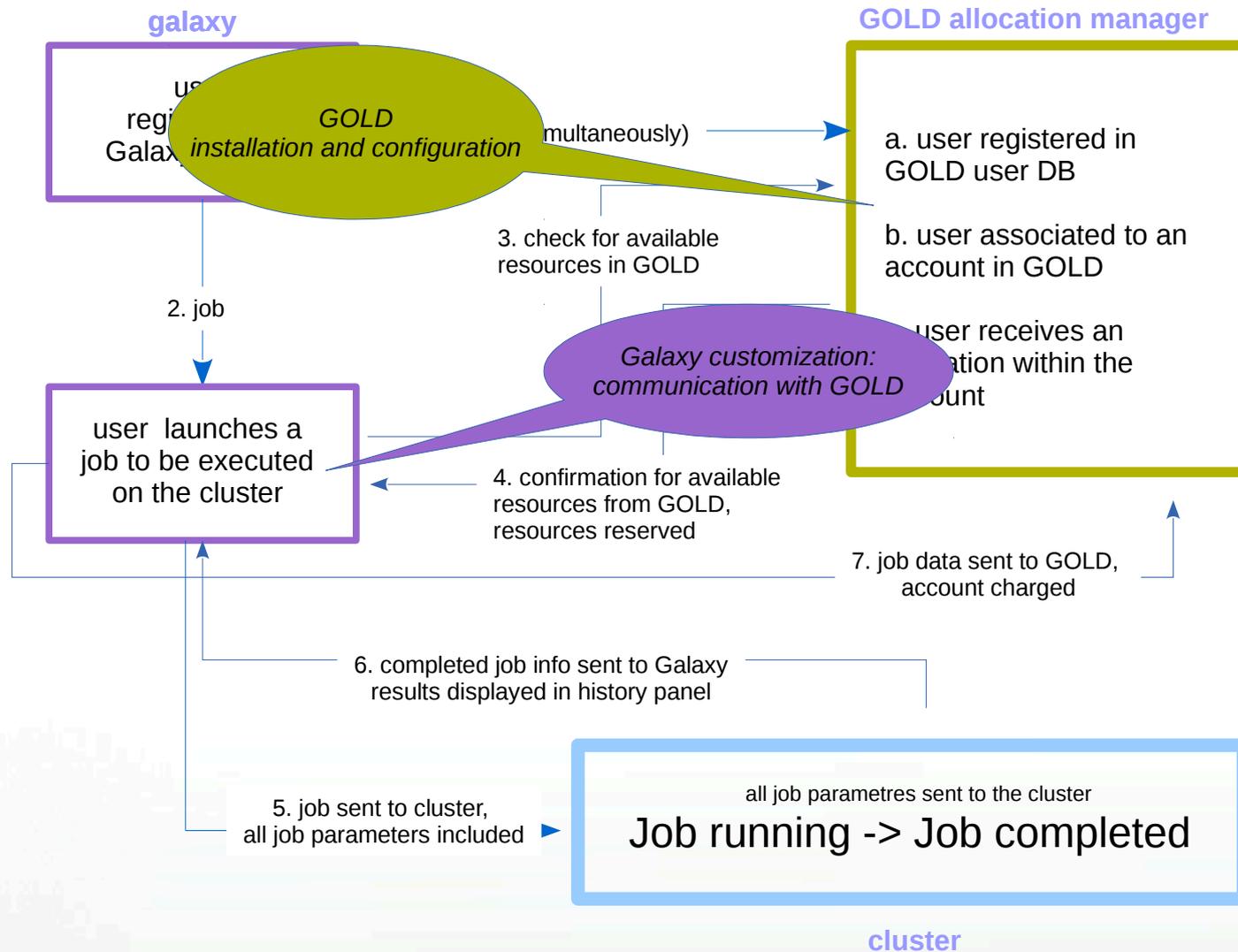


# Clusters + Galaxy : what is this tutorial about

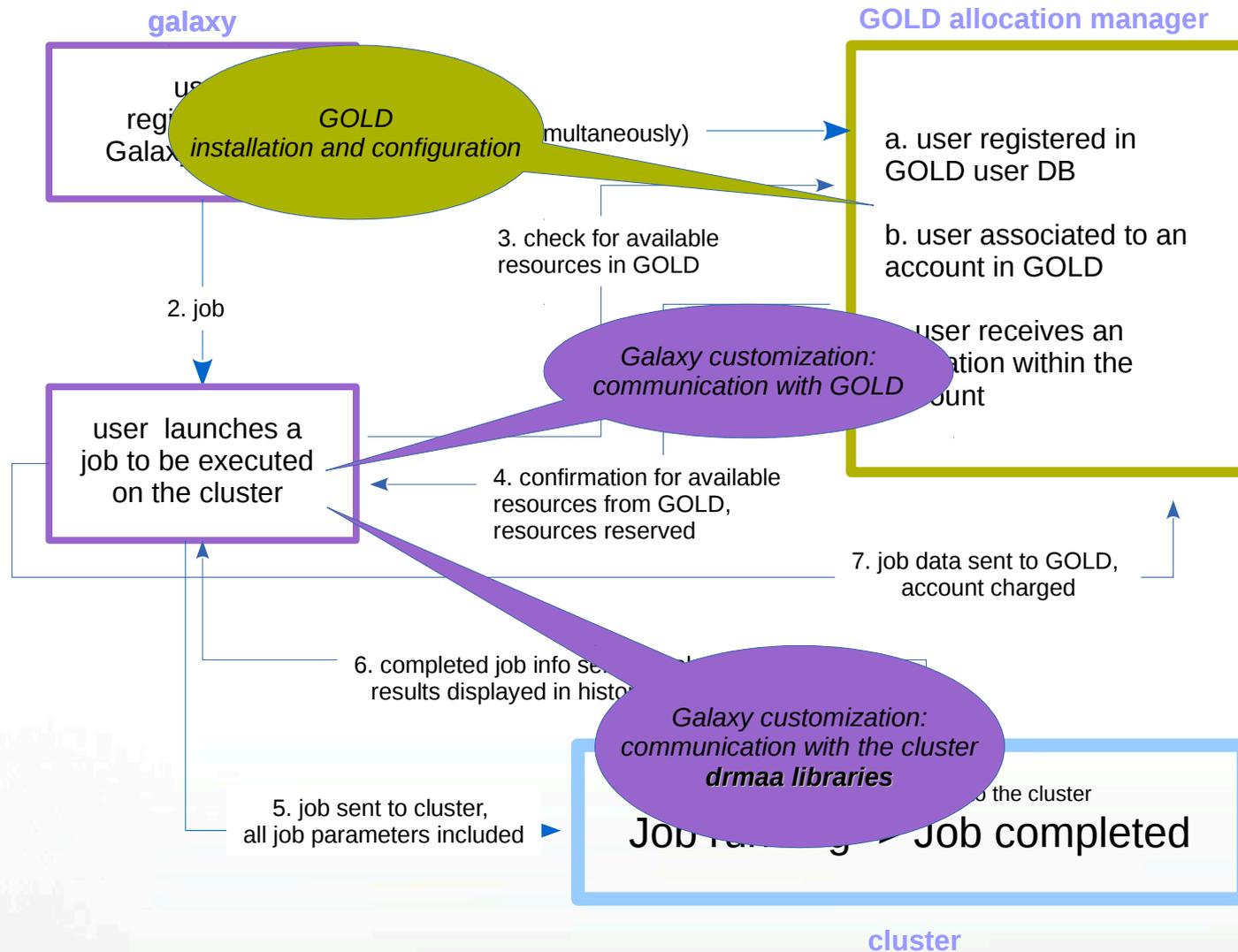


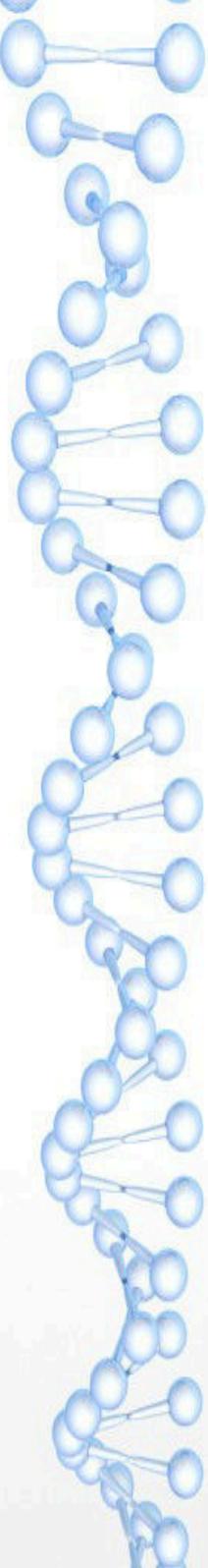
# Clusters + Galaxy :

## what is this tutorial about



# Clusters + Galaxy : what is this tutorial about





# GOLD Allocation Manager

<http://docs.adaptivecomputing.com/gold/>

- open source L(A)PP technology
  - Linux
  - Apache (optional, needed for GUI only)
  - Postgresql (MySQL)
  - Perl
- easy to install (compilation needed)
- terminal and GUI (optional)
  - easy to plug into Galaxy – gold commands are called as subprocesses in python
- excellent manual
- v2.2.0.5 at present, project stopped, the resumed, now 3.x available

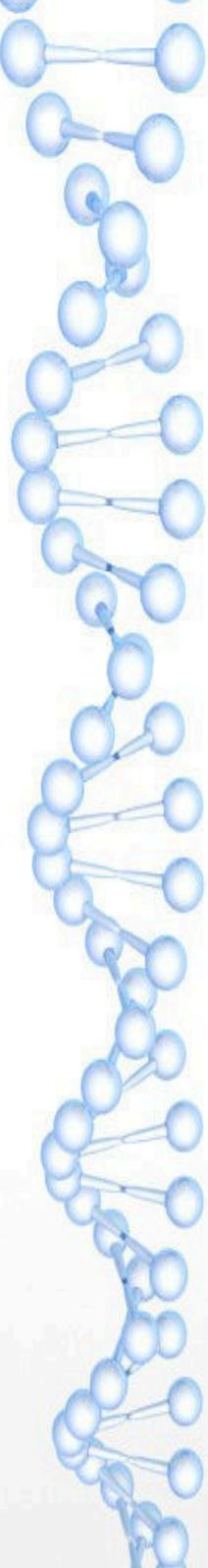
Setup instructions in the file `GOLD-core-installation.txt`

# GOLD – terminal & web

```
[root@galaxy-prod01 ~]# /opt/gold/bin/glsuser | grep added
abegiz3@yahoo.com          True          non-feide user added by magnus.popp@nhm.uio.no
anna.vader@unis.no        True   Anna Vader   anna.vader@unis.no   Manually added non-feide project manager - nikolaiv
laurindodasilva@fas.harvard.edu True          non-feide user added by torbjorn.ekrem@ntnu.no
mgrabowski@email.gwu.edu  True          non-feide user added by t.f.hansen@ibv.uio.no
ovidiu.paun@univie.ac.at  True          non-feide user added by a.k.brysting@ibv.uio.no
ricivito@hotmail.it       True          non-feide user added by geirksa@ifi.uio.no
```

The screenshot shows a web browser window titled "List Users - Mozilla Firefox". The address bar shows the URL "https://gold.galaxy-prod01.hpc.uio.no/index.cgi". The page has a yellow header with the text "List Users". Below the header, there is a search section with "Search:" and "By:" labels. The "Search:" field contains the text "added" and the "By:" dropdown menu is set to "Description". Below the search section is a table with the following columns: Name, Active, CommonName, PhoneNumber, EmailAddress, DefaultProject, Organization, and Description. The table contains seven rows of user data. On the left side of the page, there is a blue sidebar with a "GOLD" logo and a "Manage" menu. The menu items include "Users", "Projects", "Machines", "Accounts", "Allocations", "Reservations", "Quotations", "Jobs", "ChargeRates", "Transactions", and "Organizations". A "Logout" button is located at the bottom of the sidebar.

Name	Active	CommonName	PhoneNumber	EmailAddress	DefaultProject	Organization	Description
abegiz3@yahoo.com	True						non-feide user added by magnus.popp@nhm.uio.no
mgrabowski@email.gwu.edu	True						non-feide user added by t.f.hansen@ibv.uio.no
ovidiu.paun@univie.ac.at	True						non-feide user added by a.k.brysting@ibv.uio.no
anna.vader@unis.no	True	Anna Vader		anna.vader@unis.no			Manually added non-feide project manager - nikolaiv
ricivito@hotmail.it	True						non-feide user added by geirksa@ifi.uio.no
laurindodasilva@fas.harvard.edu	True						non-feide user added by torbjorn.ekrem@ntnu.no



# GOLD integration into Galaxy: *a new feature*

## Users:

- Apply for a project (account) specifying the necessary resources :
  - project duration
  - total of cpu hours
  - preferred applications on the cluster
- After their project approval, the users become Project Managers for their project

## Project managers:

- Manage the projects :
  - delete
  - prolong
  - modify : cpu hours, apps
  - see pending/rejected projects
- Add users to the projects

# Project application form

Lifeportal - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Lifeportal http://docs.ada...ting.com/gold/ List Users

https://lifeportal.uio.no/project\_application Gold allocation

UiO: University of Oslo Analyze Data Workflow Shared Data Visualization Admin ProjectAdmin Reports Apply for a project Help User Using 37.4 KB

## Lifeportal

### Lifeportal project application form

✔ Please fill in the required information

#### Personal information about the project leader

Project responsible, e.g. John Doe. (permitted chars: capital/small letters and blanks)\*:

Job title/position (permitted chars: capital/small letters, digits and blanks):

E-mail address:

Phone no (please, use digits only, e.g. 0047XXX for Norway):

Cell-phone number (please, use digits only, e.g. 0047XXX for Norway)\*:

Institution (Faculty, Department) (permitted chars: capital/small letters and blanks)\*:

Country (permitted chars: capital/small letters and blanks)\*:

#### Project information

Project name (permitted chars: capital/small letters and digits)\*:

CPU hours (specify how many CPU hours you need for the project)(permitted chars: digits)\*:

Preferred applications\* (Click on the box to display the apps) :

Project description (permitted chars: all)\*:

Start date\*:

End date :

I declare that the project **does not** contain sensitive data\*:

I have read the [Lifeportal requirements](#) and accepted them\*:

**Note:** Fields marked \* must be filled out.

Send application form Cancel

# Project selection for job execution

The screenshot shows the Liferay portal interface for job execution. The browser window title is "Liferay - Mozilla Firefox" and the URL is "https://galaxy-test.uio.no/root". The page header includes "UiO University of Oslo" and navigation links like "Analyze Data", "Workflow", "Shared Data", "Visualization", "Admin", "ProjectAdmin", "Reports", "Help", and "User".

The main content area is titled "Abel handshake (version 1.0.0)" and "JOB PARAMETERS". A blue circle highlights the "Projects/Accounts" dropdown menu, which is currently set to "lp9". Below this, there are four configuration sections:

- Number of tasks:** Enter the number of tasks for the current job. (Enter 1, if not sure.) Value: 1
- Number of tasks per node:** Enter the number of tasks per node (max 16). (Leave 1, if not sure.) Value: 1
- Walltime (job duration):** Select how long the job shall be running. Value: DD:00 HH:2 MM:00 SC:00
- Memory per CPU:** Select how much memory you need for the job (the allowed format is digit+Gb/Mb, e.g. 16GB or 3000Mb). Value: 4Gb

An "Execute" button is located below these parameters. Below the configuration area, there is an "About the tool" section stating: "This tool tests the connection to the Liferay computational backend Abel. Abel home page: <http://www.uio.no/hpc>".

The right sidebar shows a "History" list of jobs:

Job ID	Job Name	View	Refresh	Delete
1stTESTafter	23.2 MB			
352:	Abel handshake	👁	🔄	✖
351:	Abel handshake	👁	🔄	✖
350:	Abel handshake	👁	🔄	✖
349:	Abel handshake	👁	🔄	✖
348:	Abel handshake	👁	🔄	✖
347:	1.fasta	👁	🔄	✖
346:	abel-data.txt	👁	🔄	✖
345:	Abel handshake	👁	🔄	✖
344:	Abel handshake	👁	🔄	✖
343:	Dumper.py	👁	🔄	✖
342:	1.fasta	👁	🔄	✖
341:	Abel handshake	👁	🔄	✖
340:	1.fasta	👁	🔄	✖
339:	Abel handshake	👁	🔄	✖
338:	1.fasta	👁	🔄	✖
337:	1.fasta	👁	🔄	✖
336:	lifeportal-leaf-wide1-60x30.jpg	👁	🔄	✖
335:	1.fasta	👁	🔄	✖
334:	Abel handshake	👁	🔄	✖
333:	Dumper.py	👁	🔄	✖

# Project management

Lifeportal - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Lifeportal http://docs.ada...ting.com/gold/ List Users

https://lifeportal.uio.no/project\_admin Gold allocation

UiO University of Oslo Analyze Data Workflow Shared Data Visualization Admin ProjectAdmin Reports Apply for a project Help User Using 37.4 KB

## Lifeportal

Project Administration

- Users
  - Manage users
- Projects
  - Manage projects
  - Show pending projects
  - Show rejected projects

### Lifeportal Project Administration

✔ Lifeportal projects list sorted by project name

#### Full Lifeportal local projects list

**Attention! The Amount of CPU hours below will only be displayed from Start date.**  
The value will be 0.00 before Start and after End

Project owner	Project name	Users	Active	Amount	Description	Start	End
<input type="radio"/> David.Rees@bio.uib.no	lp20	MEMBERS David.Rees@bio.uib.no	True	155.50	Barnacles	2014-02-01	2014-09-30
<input type="radio"/> Trond.Oskars@student.uib.no	lp6	MEMBERS Trond.Oskars@student.uib.no	True	0.00	PhylogenyCephalaspidea	2013-10-21	2013-11-05
<input type="radio"/> Trond.Oskars@um.uib.no	lp34	MEMBERS Trond.Oskars@um.uib.no	True	941.40	AdditionalAnalysesCephas	2014-05-22	2014-09-30
<input type="radio"/> a.k.brysting@ibv.uio.no	lp13	MEMBERS a.k.brysting@ibv.uio.no ovidu.paun@univie.ac.at l.n.olsen@ibv.uio.no m.k.brandrud@ibv.uio.no mariekbr@student.ibv.uio.no mildries@student.matnat.uio.no mildries@student.ibv.uio.no	True	8958.28	Cochlearia	2013-12-03	2014-09-30
<input type="radio"/> a.k.krabberod@ibv.uio.no	lp2	MEMBERS a.k.krabberod@ibv.uio.no jon.brate@ibv.uio.no	True	42281.51	PhyRad	2013-10-17	2014-09-30
<input type="radio"/> a.k.kristoffersen@odont.uio.no	lp27	MEMBERS a.k.kristoffersen@odont.uio.no	True	211.77	IKM16samplicon2014	2014-04-10	2014-09-30
<input type="radio"/> anna.vader@unis.no	lp17	anna.vader@unis.no	True	1000.00	unis - Manually Added	2014-01-30	2014-09-30
<input type="radio"/> arthurah@student.ibv.uio.no	lp38	MEMBERS arthurah@student.ibv.uio.no	True	20000.00	Xcell	2014-06-23	2014-09-30
<input type="radio"/> g.o.s.thomassen@usit.uio.no	lp15	MEMBERS g.o.s.thomassen@usit.uio.no	True	50.00	rektest	2014-01-21	2014-09-30
<input type="radio"/> g.o.s.thomassen@usit.uio.no	lp16	MEMBERS g.o.s.thomassen@usit.uio.no	True	50.00	rkftest21	2014-01-21	2014-09-30
<input type="radio"/> g.o.s.thomassen@usit.uio.no	lp19	MEMBERS g.o.s.thomassen@usit.uio.no	True	10.00	test11	2014-02-10	2014-09-30
<input type="radio"/> g.o.s.thomassen@usit.uio.no	lp5	MEMBERS	False	10.00	fttest01		

# GOLD integration into Galaxy: *code modification*

## Python scripts

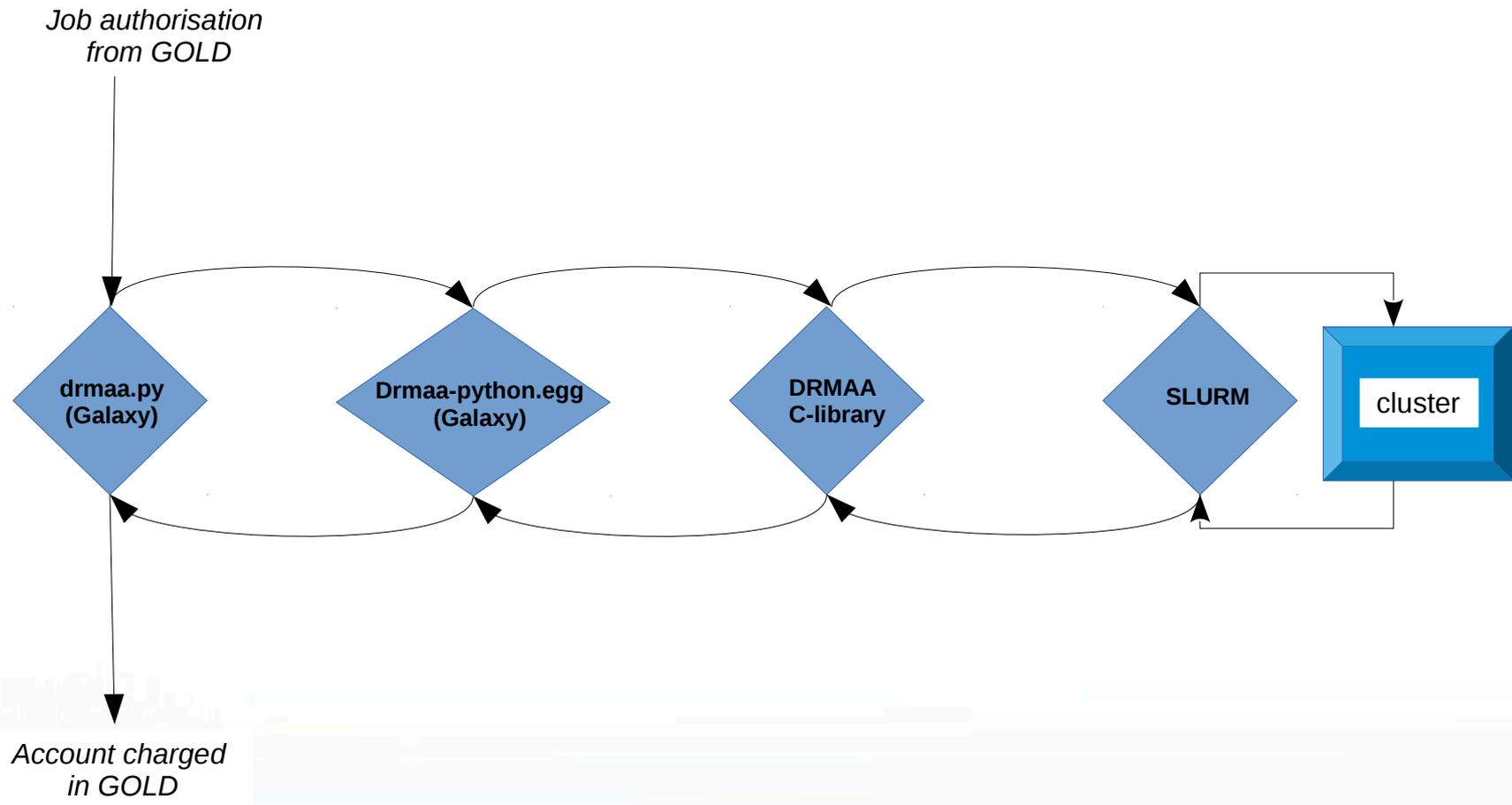
```
[root@galaxy-prod01 BaltimoreVM]# tar tvf gold_pythons.tar
-rwxr-xr-x root/root      65592 2014-06-05 11:47 Accounting.py
-rw-r--r-- root/root      27875 2014-06-11 14:04 config.py
-rwxr-xr-x root/root       7677 2014-06-05 11:52 Dumper.py
-rw-r-xr-- root/root     53724 2014-06-05 11:40 framework_init_.py
-rw-r-xr-- root/root       2019 2014-06-05 11:44 project_application.py
-rwxr-xr-x root/root       2097 2014-06-05 11:48 Project_managers.py
-rw-r-xr-- root/root     85827 2014-06-05 11:40 user.py
-rw-r-xr-- root/root      4821 2014-06-05 11:38 webapps_project_admin.py
-rw-r-xr-- root/root     34920 2014-06-05 11:38 webbase_project_admin.py
-rw-r-xr-- root/root        525 2014-06-05 11:40 web_init_.py
```

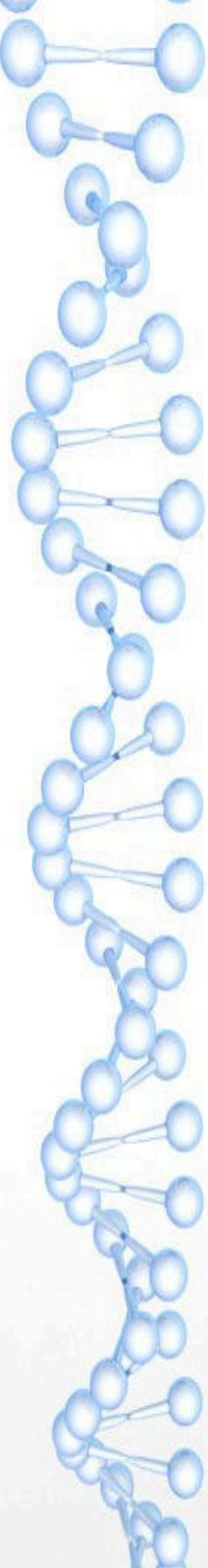
```
[root@galaxy-prod01 BaltimoreVM]# tar tvf gold_makos.tar
-rw-r-xr-- root/root      5792 2014-06-05 11:32 approve_pending_project.mako
-rw-r-xr-- root/root      4386 2014-06-05 11:33 approve_rejected_project.mako
-rw-r-xr-- root/root     13072 2014-06-05 10:57 base_panels.mako
-rw-r-xr-- root/root      3075 2014-06-05 11:31 center_project_admin.mako
-rw-r--r-- root/root      6780 2014-06-05 11:34 center_project_application.mako
-rw-r-xr-- root/root      2166 2014-06-05 11:32 create_new_gold_project.mako
-rw-r-xr-- root/root      5491 2014-06-05 10:56 GOLD_user_register.mako
-rw-r-xr-- root/root      3045 2014-06-05 10:58 index_project_admin.mako
-rw-r-xr-- root/root      1204 2014-06-05 11:33 index_project_application.mako
-rw-r-xr-- root/root      6771 2014-06-05 10:57 login.mako
-rw-r-xr-- root/root      6279 2014-06-05 11:32 manipulate_gold_projects.mako
-rw-r-xr-- root/root     24050 2014-06-05 10:56 project_admin_grid_base.mako
-rw-r-xr-- root/root         49 2014-06-05 10:57 project_admin_grid.mako
-rw-r-xr-- root/root      5189 2014-06-05 10:56 register.mako
-rw-r-xr-- root/root      2923 2014-06-05 11:32 show_pending_projects.mako
-rw-r-xr-- root/root      2982 2014-06-05 11:32 show_rejected_projects.mako
-rw-r-xr-- root/root        254 2014-06-05 11:34 stored_project_application.mako
-rw-r-xr-- root/root      1513 2014-06-24 12:09 confirm_display_gold_project_usage.mako
-rw-r-xr-- root/root        498 2014-06-24 12:09 display_gold_project_usage.mako
```

## Makos

Setup instructions in the file Galaxy-GOLD-code.txt

# Sending Galaxy jobs to the cluster: *chart*

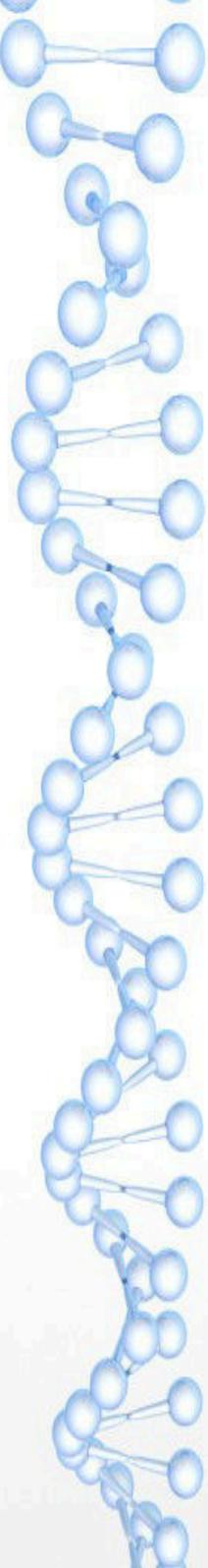




# Sending Galaxy jobs to the cluster

see configuration files on the VM

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



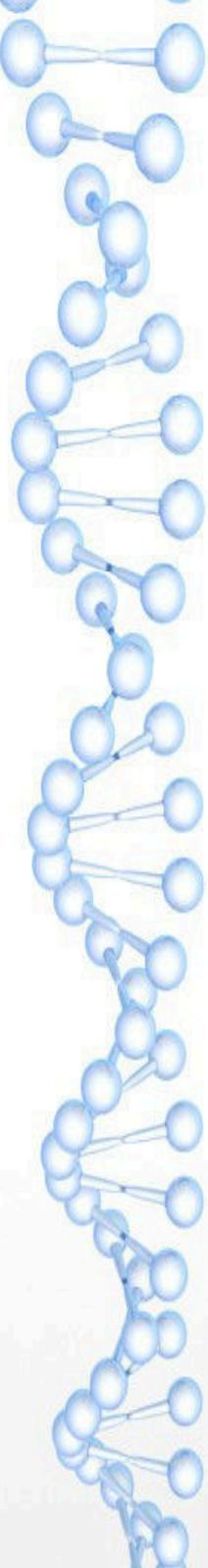
# SLURM / MUNGED client installation

(SLURM manager is already installed on the cluster)

- `apt-get install munge`
- `apt-get install slurm-llnl`
- `apt-get install libslurm-dev`
- `/usr/sbin/create-munge-key`
- `/etc/init.d/munge start`

Setup instructions in the file

`MUNGE-SLURM-DRMAA-installation.txt`

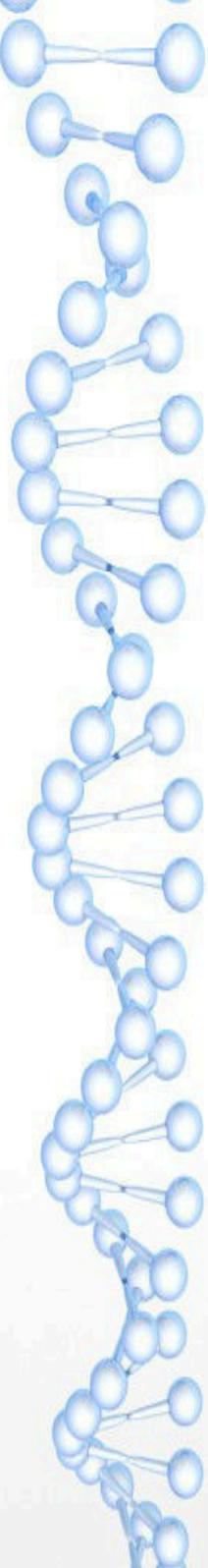


# DRMAA C-library

**DRMAA C-library** is a mediator between *drmaa.py/drmaa.egg* and *SLURM queueing system*

- Copy from `/site/drmaa/slurm-drmaa-1.0.6` on the supplied VM or
- download from <http://apps.man.poznan.pl/trac/slurm-drmaa>

Setup instructions given in the file  
`MUNGE-SLURM-DRMAA-installation.txt`



# Galaxy *code modification* for DRMAA-SLURM

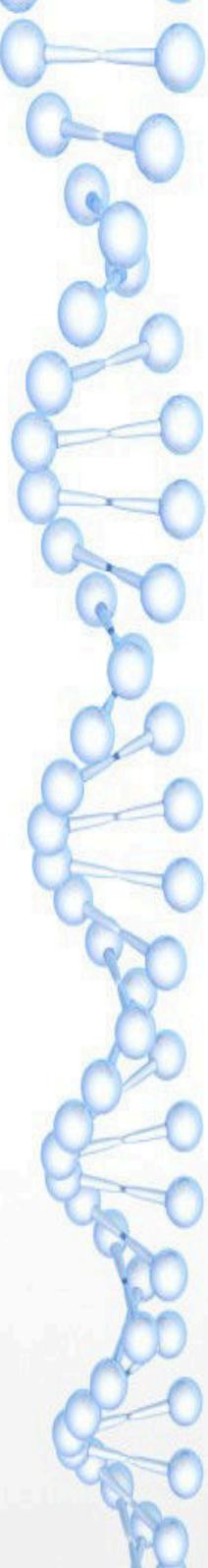
## Python scripts

```
[root@galaxy-prod01 BaltimoreVM]# tar tvf drmaa_pythons.tar
-rw-r-xr-- root/root      20108 2014-06-11 14:20 drmaa.py
-rw-r-xr-- root/root      21953 2014-06-11 14:19 galaxy_jobs_runners__init__.py
-rw-r-xr-- root/root      26474 2014-06-11 14:19 galaxy_tools_actions__init__.py
-rwxr-xr-x root/root     182793 2014-06-11 14:17 galaxy_tools__init__.py
-rw-r-xr-- root/root       7477 2014-06-11 14:20 mapper.py
-rw-r-xr-- root/root      25829 2014-06-11 14:17 tool_runner.py
```

## Makos

```
[root@galaxy-prod01 BaltimoreVM]# tar tvf drmaa_makos.tar
-rw-r-xr-- root/root      18537 2014-06-11 14:21 tool_form.mako
-rw-r-xr-- root/root      16094 2014-06-11 14:21 usit_slurm_switches.mako
```

Setup instructions in the file  
**Galaxy-DRMAA-code.txt**



# What do we send and what do we receive?

## To the cluster

```
>>> DRMAA.PY slurm script #!/bin/sh
#SBATCH -A nn9108k
#SBATCH -t 300:00:00
#SBATCH -D /work/projects/galaxy/data/database_galaxy_prod01/job_working_directory/000/629
#SBATCH -J a.k.krabberod@ibv.uio.no::lp2
#SBATCH -n 1
#SBATCH --mem-per-cpu=4000
#SBATCH --ntasks-per-node=8
#SBATCH --partition=normal

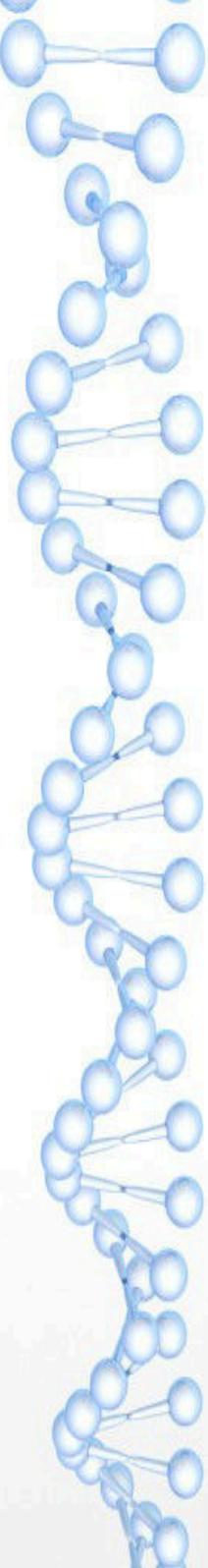
source /etc/profile
source /cluster/bin/jobsetup

cd /work/projects/galaxy/data/database_galaxy_prod01/job_working_directory/000/629
module load blast+; blastn -query "/work/projects/galaxy/data/database_galaxy_prod01/files/000/dataset_922.dat" -db
"/work/databases/bio/ncbi-pre/est_others" -task blastn -evalue 0.001 -out
/work/projects/galaxy/data/database_galaxy_prod01/files/000/dataset_924.dat -outfmt 6 -num_threads 8 -dust yes -strand both
-max_target_seqs 100
echo $? > /work/projects/galaxy/data/database_galaxy_prod01/job_working_directory/000/629/galaxy_629.ec
```

## From the cluster

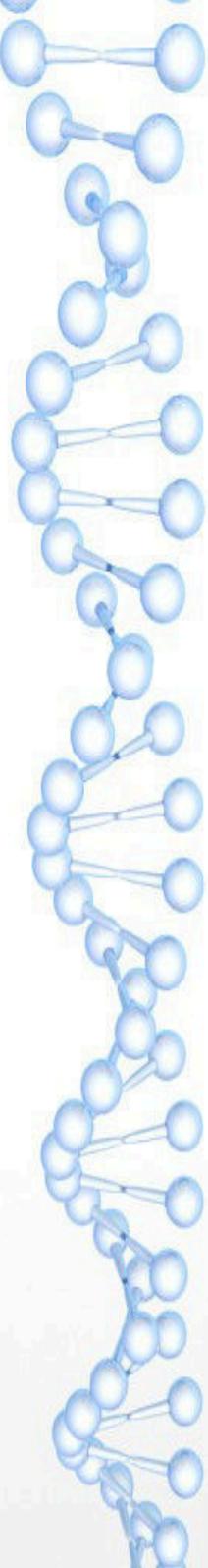
```
{'charge_duration': '451587',
'galaxy_job_id': '629',
'lifeportal_project': 'lp2',
'machine': 'c17-6',
'processes': '8',
'slurm_account': 'nn9108k',
'slurm_job_id': '5200757',
'user': 'a.k.krabberod@ibv.uio.no'}
Successfully charged job 629 for 3612696 credits - MESSAGE FROM GOLD

galaxy.jobs.runners.drmaa DEBUG 2014-02-27 20:27:26,034 Successfully charged job 629 for 3612696 credits
galaxy.jobs.runners.drmaa DEBUG 2014-02-27 20:27:26,034 (629/5200757) state change: job finished normally
```



# Application integration

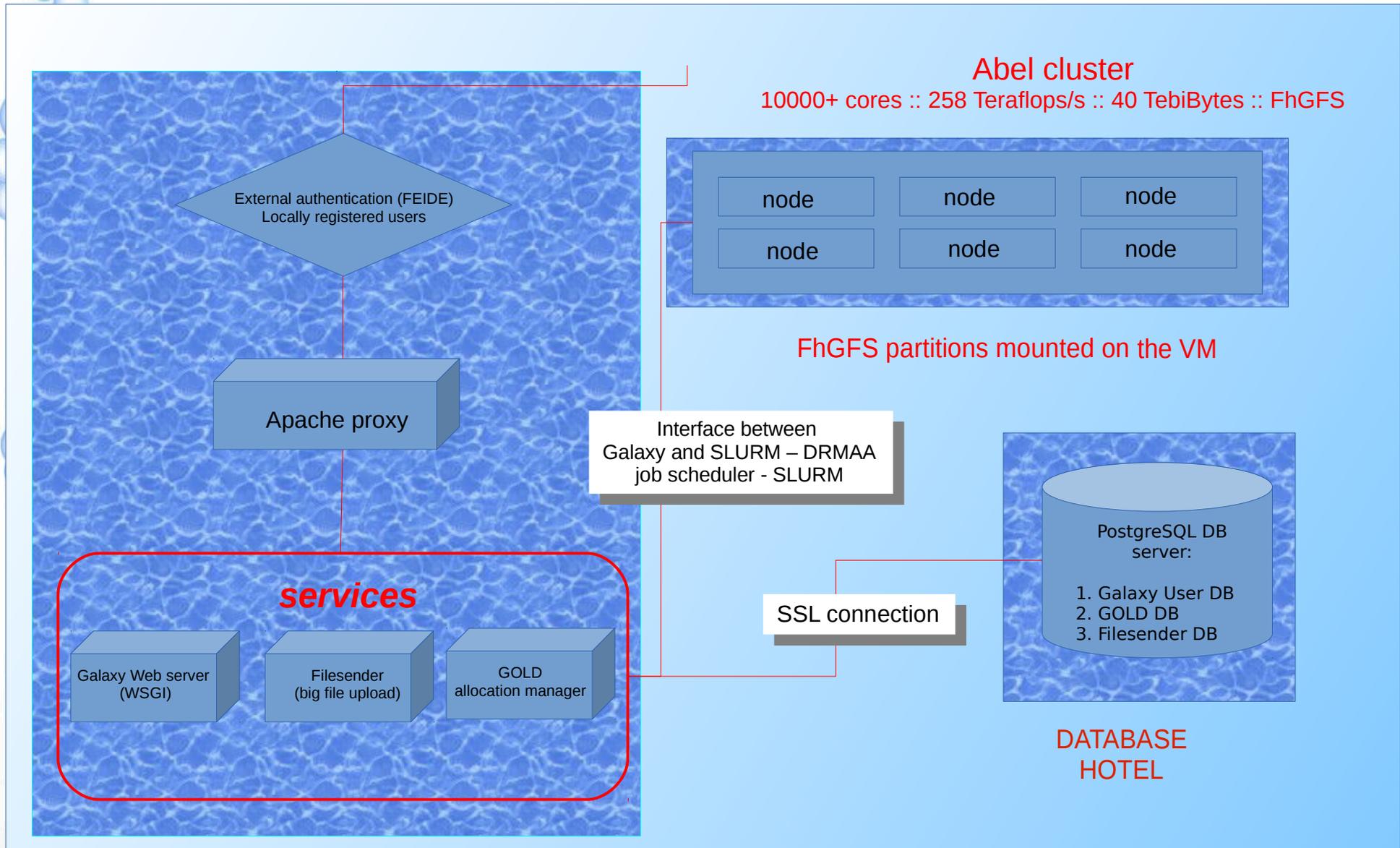
- Applications can be executed locally or on the cluster
- Cluster apps installed on the cluster file system under **module** software environment management
- “module load” statement included in the **xml command block**
- Galaxy internal tools/libraries were exported on the cluster (all /galaxy-dist/lib tree) and dependencies installed

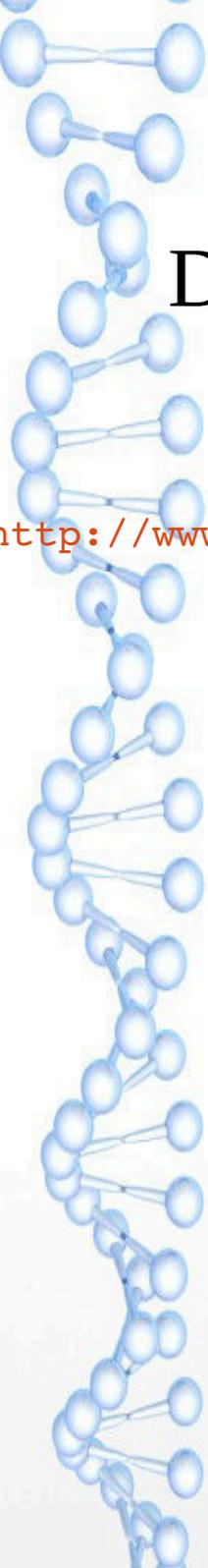


# Work plan

- Go through the files in the VM
  - README-FIRST.txt (in the Desktop>Setup\_Manual )
  - Login / version info file
  - general setup files
  - GOLD setup files
  - Galaxy-GOLD related changes
  - Galaxy-DRMAA related changes
  -
- Run the feature on the VM (Galaxy interface)
  - Start all the services
    - sudo -u daemon munged
    - slurmctld
    - slurmd
    - /opt/gold/sbin/goldd start
    - cd /home/galaxy/galaxy-dist
    - ./run.sh --daemon
    - type "localhost" for the URL (in a browser)
  - DEBUG
    - tail -f /home/galaxy/galaxy-dist/paster.log
  - register, log in, apply for a project, approve a project, GOLD and Galaxy Web
  -
- An example from a production setup (Lifeportal)

# Overview





Detailed setup page for our production server  
can be found here :

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

Write to us at :

[lifeportal-help@usit.uio.no](mailto:lifeportal-help@usit.uio.no)

Link to our production server :

<http://lifeportal.uio.no>