## Setting up a Galaxy instance as a service

(or ... how can we make simple things more complicated)

Nikolay Vazov

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



#### Astonishment phase

In the beginning there was YOU



#### What about sharing?

Then came your colleagues!



... and said : could I have a look at your results and how did you do that one?



#### You start the real work now





#### And the problems begin piling up ...





#### And the solutions start to flow in ...





### More flexibility ...





## Why did we go for such a solution?

USIT (University Center for Information Technologies), University of Oslo Special group for IT-support for research

#### Domains:

- Bioinformatics (NGS and Phylogeny)
- Geosciences
- Physics
- Chemistry
- Language Engineering (computational linguistics)
- Climate studies
- etc.

#### **Big cluster configuration (ABEL)**:

Number of Cores : 10000+ Number of nodes : 650+ Max Floating point performance : 258 Teraflops/s Total memory : 40 TebiBytes Total local storage : 400 TebiBytes Users : ~1750

#### Galaxy instances

#### **1** Bioinformatics platform lifeportal.uio.no (production):

Users : 535 (end June)

Access to cores : 687

Tools : over 300

1 Geosciences (Geoportal) and 1 Language engineering (Language Analysis Portal) to become in production in early fall

7 test Galaxy instances

Currently 3 galaxy code versions Moving to common code base now under Git repository with branches for each customization



# What do we mean by a distributed (complex) Galaxy instance?

Galaxy running on your machine/dedicated server	VS	Galaxy web-interface and job execution running on different servers (hosts)
SQLite (built-in)	VS	Galaxy DB – a professional relational database (e.g. <i>PostgreSQL</i> ) exported to a DB "hotel"
WSGI server (built-in)	VS	Production-oriented web-server solution (e.g. <i>Apache</i> proxy server)
Applications running on your machine	VS	Applications running on a cluster (loaded as modules)
Datasets on the same machine (storage limitations)	VS	Datasets on a cluster partition (huge amount of data + centralized backup and storage)
No need for authentication	VS	Authentication and authorization - required
No need for accounting	VS	Accounting/resource management - required



#### Galaxy instance : separate galaxy code from data?

Galaxy running on your machine/dedicated server

VS

Galaxy running on a VM/separate physical machine which mounts a cluster partition





#### Professional SQL database





#### And a NoSQL database – Language Analysis Portal



UiO: University of Oslo

### Set up a production web-server





## Remote job execution



#### Data





#### Who is authorized to use the service





## Authentication from within galaxy(1)

File Edit View History Bookmarks Tools	Lifeportal - Mozilla Firefox			_ U X							
Nikolay Aleksandrov Vaz ×	tal × 争										
(  Anttps://lifeportal.uio.no/root	✓ ď	🔍 Abel cluster UiO	→ ☆ 自 🕹	A O <sub>0</sub> ≡							
UiO # University of Oslo	Analyze Data Workflow Shared Data 🗸 Visualization 🗸 Help 🗸 User 🗸			Using 0 bytes							
Academic login	About the Lifeportal										
Authenticate with your Feide 🗢 account.	Lifeportal gives you easy access to the High Performance Computing cluster Abel at the University of Oslo. The Galaxy based Lifeportal has a continuously growing list of services, and among them the most widely used tools from the former Bioportal.										
	Added to Count of Cou										
	Getting access Getting access to the Lifeportal										
	Information for external users										
Email address:	Register for AIR, BIR or CLOTU access										
	Register for SNPexp access										
Password:	Reviewers registration and login										
Login Eorgot password? Reset here	Login Forgot password? Reset here										
Visit the Lifeportal!											
You can explore but you cannot start applications.											
visitor Login											
Partners											
UIO: Universitetet i Oslo Norwegian Bloinformatics Platform											
www.uio.no/english/services/it/research/hpc/lif	eportal/start-using			<							
				1							

# Authentication from within galaxy(2)

HotSpot (00d 23h 56m 🗴 🚦 Lifep	portal × <b>:</b> Lifeporta	al × 🕇		
< 🖨 https://galaxy-test.uio.no/root			🔻 🥙 🔀 🔻 Google	Q ↓ 佘 ☆ 自 三
🛅 Most Visited 👻 📵 Getting Started  🔝	Latest Headlines 🔻			
UiO I University of Oslo	Analyze Data W	/orkflow Shared Data <del>-</del> Visualiza	tion → Help → User →	Using 0 bytes
Lifeportal				
Academic login Authenticate with your eduGain 2 account. AAI@EduHr - Croatian Research and * AAI@EduHr - Croatian Research and Education Federation ACOnet AESIR ARNES	About the Lifeportal Lifeportal gives you easy access to th cluster Abel at the University of Oslo. continuously growing list of services, tools from the former Bioportal. More about Lifeportal Getting access Getting access to the Lifeportal	ne High Performance Computing The Galaxy based Lifeportal has a and among them the most widely used	1	
Academy of Sciences Library Aristotle University of Thessaloniki Art Academy of Latvia - LMA IDP Athena - Research and Innovation	Norwegian Bioinformatics Platform		xy	

>

# Authentication at the web-server level (before accessing galaxy)

Select your identity provider for the GEOPORTAL - Mozilla Firefox
Elle Edit View History Bookmarks Tools Help
Abel cluster UiO
Select your identity provider for the GEOPORTAL
English   Bokmål   Nynorsk   Sámegiella   Dansk   Deutsch   Svenska   Suomeksi   Español   Français   Italiano   Nederlands   Luxembourgish   Czech   Slovenščina   Linturiu kalha   Hoatski   Magyar   Jazyk polski   Português   pracijajno   Titikco   Jazyk   Jazyk
eesti keel   ארארא ארא ארארא ארארא ארארא ארארא ארארא ארארא ארארא ארארא ארא א
Select your identity provider for the GEOPORTAL
Please select the identity provider where you want to authenticate:
Feide - Norwegian educational institutions (test-IdP)
Feide - Norwegian educational institutions (test-idP) NeLS - Norwegian e-Infrastructure for Life Sciences (test-idP)
Apply for access to Geoportal
Copyright © 2007-2010 Feide RnD



#### Resource allocation and accounting

No need for accounting	VS	Separate module (GOLD) for accounting/resource management
------------------------	----	---

Clusters are configurations which share and allocate resources to users

Every job is reported with regard to :

- which project (account) it has been run in
- how many resources it has used (cpu hours, cores)
- which user (project) has been charged for the used resources



#### Resource allocation and accounting - example

9			Liteportal - Mozili	a Firero	х				_ = ×
Eile	<u>E</u> dit <u>V</u> iew Hi <u>s</u> tory <u>B</u> ookmarks <u>T</u> ools <u>H</u> elp								
: Lif	feportal 🛛 🗶 🛛 🎎 http://docs.adating.co	om/gold/ 💥 📋 List Users	💥 📔 😫 Lifeportal		× ( 4	4			
<b>(</b>	Attps://galaxy-test.uio.no/root							🗇 🗸 😂 🔞 🗸 Gold allocation	n 🦊 🐇
UiO	• University of Oslo	Analyze Data Workflo	v Shared Data 🗸 Visualization 🗸	Admin	ProjectAdmin Re	eports	Help <del>+</del>	User <del>-</del>	Using 23.3 MB

#### Lifeportal

Tools	Abel handshake (version 1.0.0)			3	History	0 0				
search tools		1stTESTafter 23.2 MB	Ø 🖻							
Tests Abel handshake	Projects/Accounts	<u>352: Abel handshake</u>	• / ×							
Bioportal Phylogeny Tools	You have logged as user : vazovn@abv.bg. Please	351: Abel handshake	• / ×							
R	lp9 🗘									
<u>Gaussian</u> NCBI BLAST+					349: Abel handshake	• / ×				
<u>Get Data</u>	Number of tasks	Number of tasks per node	Walltime (job duration)	Memory per CPU	348: Abel handshake	• 0 ×				
<u>Text Manipulation</u> <u>Filter and Sort</u>	Enter the number of tasks for the current job. (Enter 1, if not sure.)	Enter the number of tasks per node (max 16). (Leave 1, if not sure.)	Select how long the job shall be running :	Select how much memory you need for the job (the allowed format is digit+Gb/Nb,	<u>347: 1.fasta</u>	• 0 ×				
Join, Subtract and Group	1	1	DD: 00 HH: 2 MM: 00 SC: 00	e.g. 16GB or 3000Mb):	346: abel-data.txt	• 0 %				
Extract Features					345: Abel handshake	• / ×				
Fetch Sequences Operate on Genomic Intervals	Execute				344: Abel handshake	• / X				
Statistics	About the tool This tool tests the connection to	the Lifenortal computational backend Abel			343: Dumper.py	• 0 ×				
Graph/Display Data	Abel home page: http://www.uio.no/hpc	are Eneportar computational backend Aber.			<u>342: 1.fasta</u>	• / ×				
NGS: QC and manipulation					341: Abel handshake	• / ×				
<u>NGS: Mapping</u> NGS: SAM Tools					340: 1.fasta					
Workflows					339: Abel handshake	• 0 %				
<u>All workflows</u>					<u>338: 1.fasta</u>	• / ×				
					337: 1.fasta	• / ×				
					<u>336: lifeportal-</u> leaf-wide1-60x30.jpg	• / %				
					<u>335: 1.fasta</u>	• / ×				
					334: Abel handshake	• 0 ×				
					333: Dumper.py	• / ×				

#### Galaxy on Abel cluster (but could be anywhere else ) physical configuration



# VIRTUAL MACHINE (the galaxy server)

#### Aknowledgements:

Katerina Michalickova for the tool integration Bjørn-Helge Mevik for the SLURM setup Emanuele Lapponi – for the LAP (language analysis portal) setup



## Thank you

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

#### n.a.vazov@usit.uio.no



GCC 2015 – Galaxy as a Service, Norwich, 2015