

Galaxy Deployments at Indiana University

Carrie Ganote

National Center for Genome Analysis Support

April 16, 2015



INDIANA UNIVERSITY



INDIANA UNIVERSITY

About IU and NCGAS, and me

A bit about Indiana University:

Very culturally diverse campuses

Known for our Jacob's School of Music and Kelley School of Business



A bit about NCGAS:

National Center for Genome Analysis Support
NSF/NIH/IU funded group offering bioinformatics support/Linux help and lab experience to researchers

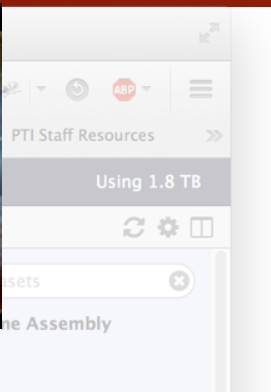

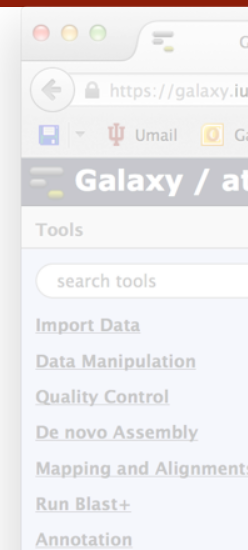
A bit about me:


NCGAS Staff since Oct 2012
First year PhD student, woohoo!
An unhealthy obsession with turkeys





INDIANA UNIVERSITY






NATIONAL CENTER FOR GENOME ANALYSIS SUPPORT


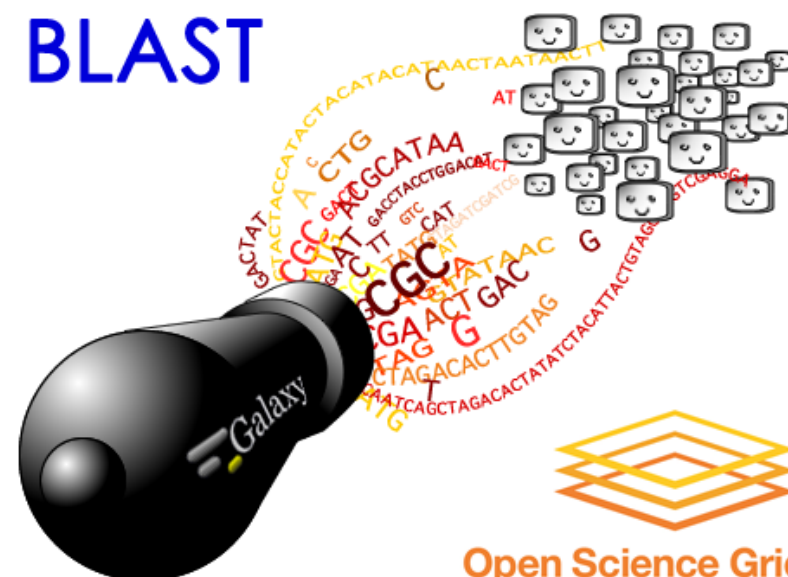
INDIANA UNIVERSITY

Our deployment of Blast is the work of a partner Support and the Open Science Grid. The OSG is:



This instance of the Galaxy is installed and maintained by Nation Award #1062432)

BLAST



Open Science Grid

National Center for Genome Analysis Support: <http://ncgas.org>



INDIANA UNIVERSITY

Galaxy / at IU

https://galaxy.iu.edu


Umail Ganote, Carrie L ... RT at a glance ActivityLog Galaxy Knowledge Base Post to CiteULike Unicornify this pa... IUScholarWorks PTI Staff Resources

Galaxy / at IU Analyze Data Workflows Shared Data Visualization Admin Help User Using 1.8 TB

Tools

search tools

Import Data
Data Manipulation
Quality Control
De novo Assembly
Mapping and Alignments
Run Blast+
Annotation
Statistics
Variants
Clustering/Phylogeny
Visualization
NGS: Mapping
Workflows



**NATIONAL CENTER FOR
GENOME ANALYSIS SUPPORT**
INDIANA UNIVERSITY

✓ Welcome to the Galaxy Instance at Indiana University

Thank you for choosing Galaxy!

Slides from our July Galaxy talks (part of the Bioinformatics Clinic) are available:
[A Short Demo on RNA-Seq](#)
[Moving Large Data onto Galaxy](#)
[Galaxy for Data Provenance](#)

NCGAS is committed to providing support for Indiana University research. Don't hesitate to contact help@ncgas.org if you find that you need a tool that is not supported by our current Galaxy or if you have questions or suggestions. When possible, we will have the requested tool up and running in two working days time, and failing that, will report the status of the request within that time.

 Our deployment of Blast is the work of a partnership between the National Center for Genome Analysis Support and the Open Science Grid. The OSG is supported by grants from the NSF and the DOE.

This instance of the Galaxy is installed and maintained by National Center for Genome Analysis Support [NCGAS](#) (NSF Award #1062432)

History

search datasets

Transcriptome Assembly Workflow
27 shown
30.5 MB

27: Trim Galore on data 1: trimmed reads

26: blastx Trinity on data 1 and data 2: Assembled Transcripts vs nr

25: Trinity on data 1 and data 2: Assembled Transcripts

24: Trinity on data 1 and data 2: log

23: ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/Escherichia_coli_K_12_substr_M_G1655_uid57779/NC_000913.fna

22: FASTQ Quality Trimmer on data 1

21: FASTQ Quality Trimmer on data 1

20: FastQC FASTQ Quality Trimmer on

National Center for Genome Analysis Support: <http://ncgas.org>



INDIANA UNIVERSITY

Fin

Thanks for watching!
Questions and comments:
Email help@ncgas.org



INDIANA UNIVERSITY

IU Galaxy



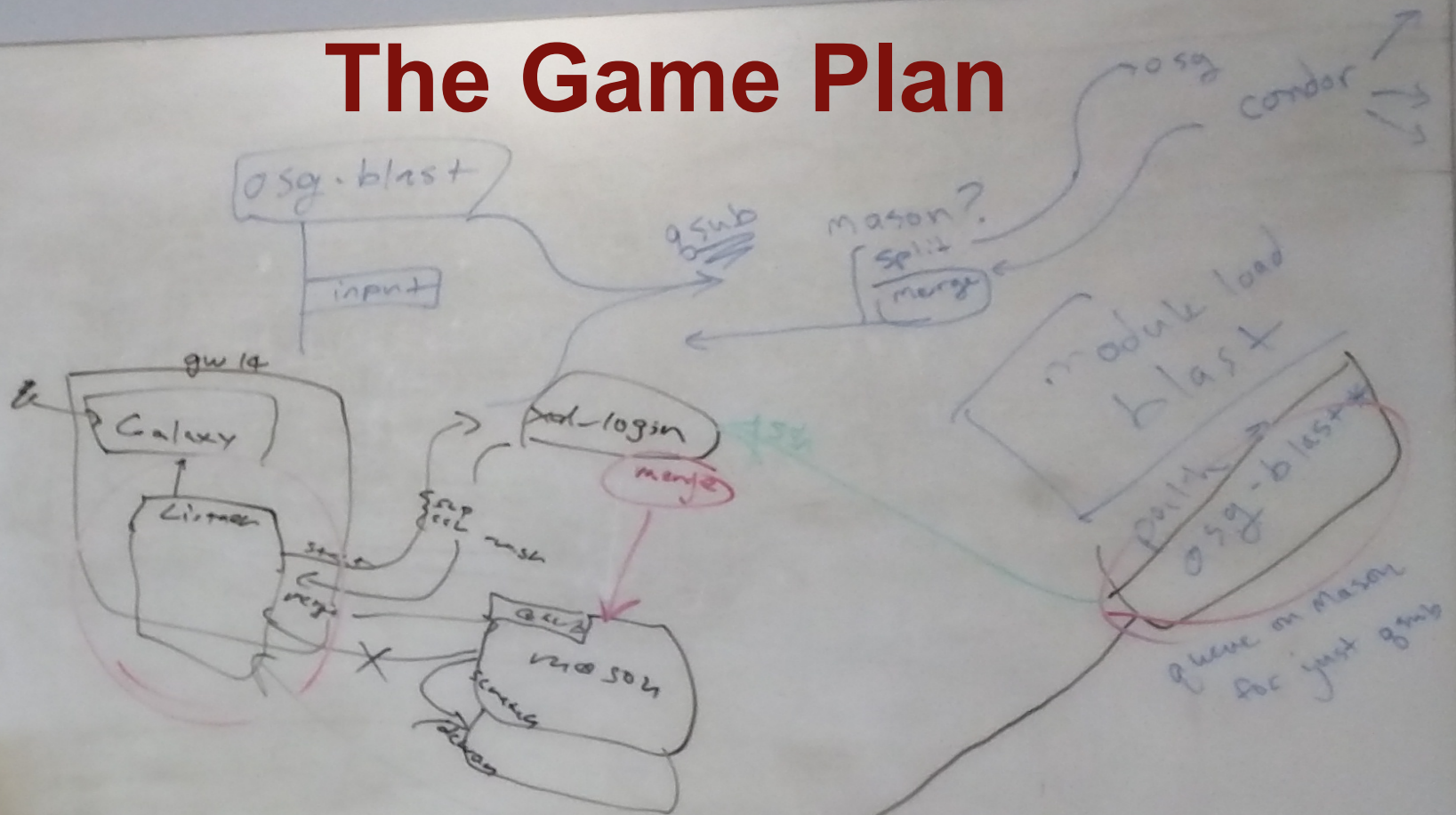
- Running since 2012 on Mason cluster at IU
 - Deployed on Quarry, Big Red2 and soon to be Karst
 - Apache proxy + Kerberos Authentication
 - Lustre Filesystem
 - Most tools, with emphasis on large memory applications
 - University-wide, plus collaborators
 - Still using hg
- Fancy javascript tool menus!**





INDIANA UNIVERSITY

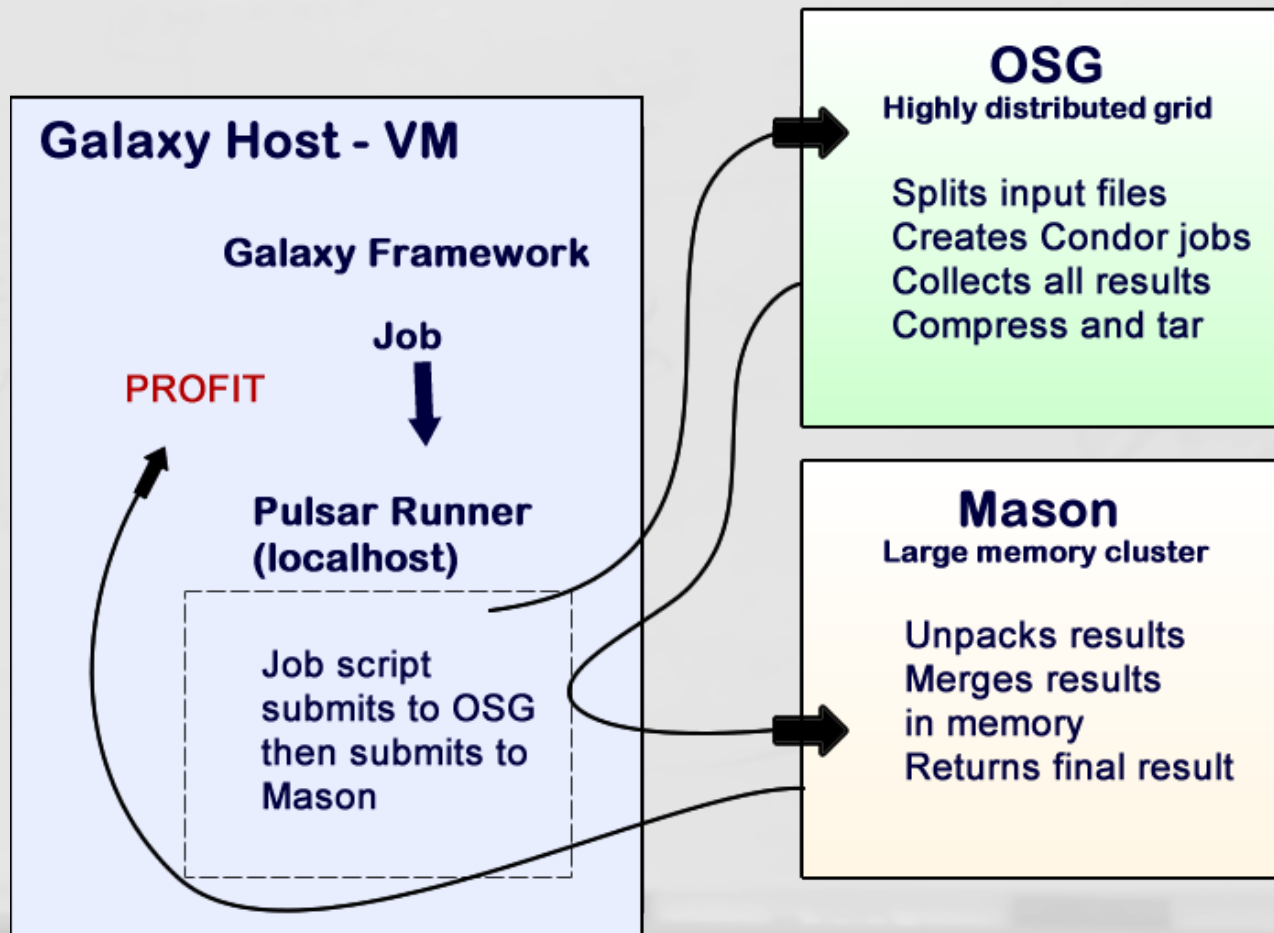
The Game Plan



BLAST is an embarrassingly parallel program – it's our only tool that uses the grid.



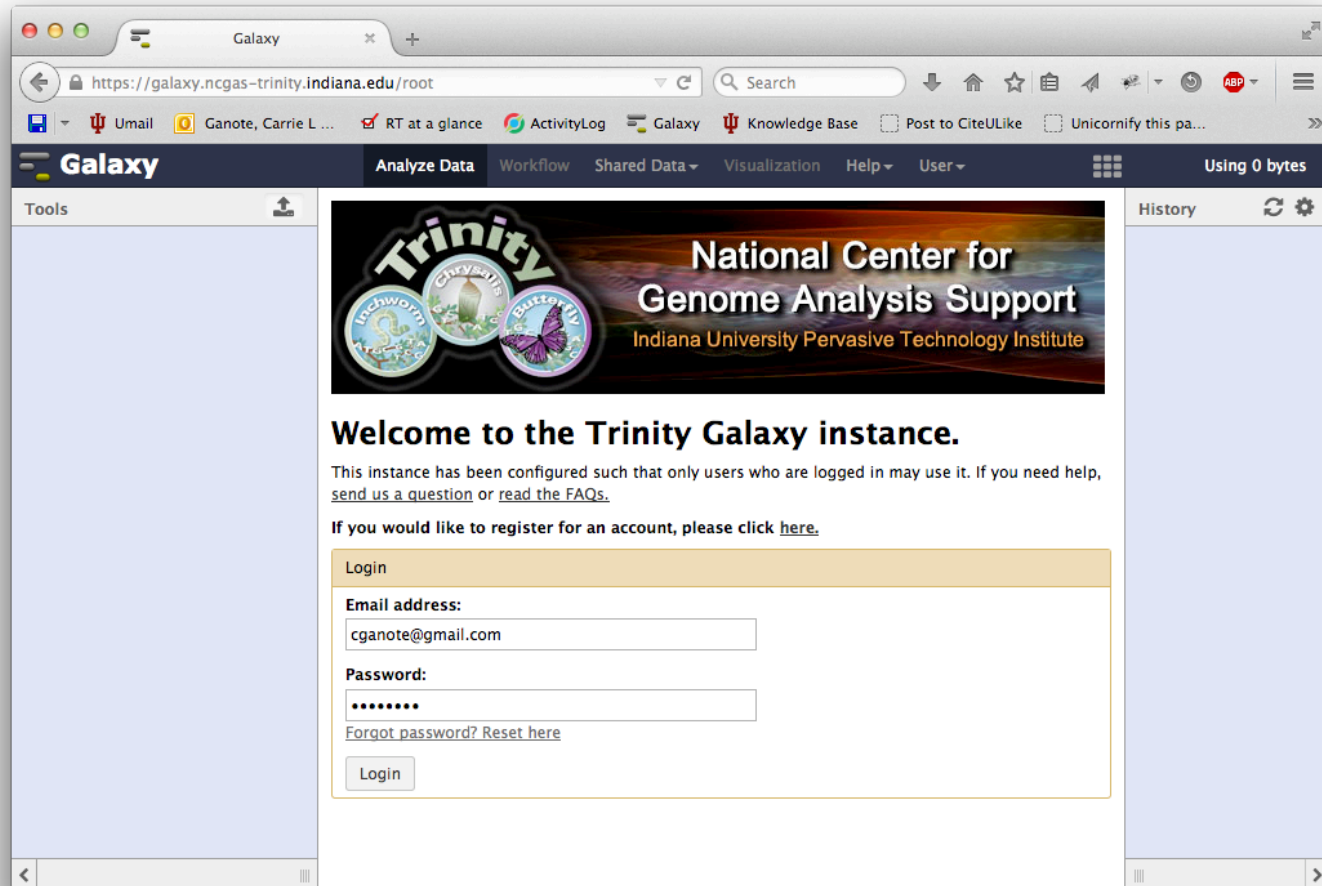
The Game Plan





INDIANA UNIVERSITY

Trinity Galaxy



National Center for Genome Analysis Support: <http://ncgas.org>



INDIANA UNIVERSITY

Trinity Galaxy

Galaxy

https://galaxy.ncgas-trinity.indiana.edu/root

Search

Umail Ganote, Carrie L... RT at a glance ActivityLog Galaxy Knowledge Base Post to CiteULike Unicornify this pa... IUScholarWorks PTI Staff Resources

Galaxy Analyze Data Workflow Shared Data Visualization Help User

Using 0 bytes

Tools

Trinity National Center for Genome Analysis Support
Indiana University Pervasive Technology Institute

Trinity Galaxy Registration Form

Please enter the required information below.

Questions? Need help? Feel free to email the NCGAS team at help@ncgas.org.

Information about You

Full Name:
* must provide value

Institution name:
* must provide value

Institution location, city:
* must provide value
ex, Austin

Institution location, state:
* must provide value
ex, Texas

Institution location, country:
ex, US

Returning?

History

National Center for Genome Analysis Support: <http://ncgas.org>



Trinity Galaxy

- One tool/pipeline emphasis
- Ftp server
- Local users

lib

galaxy

`config.py`

managers

`context.py`

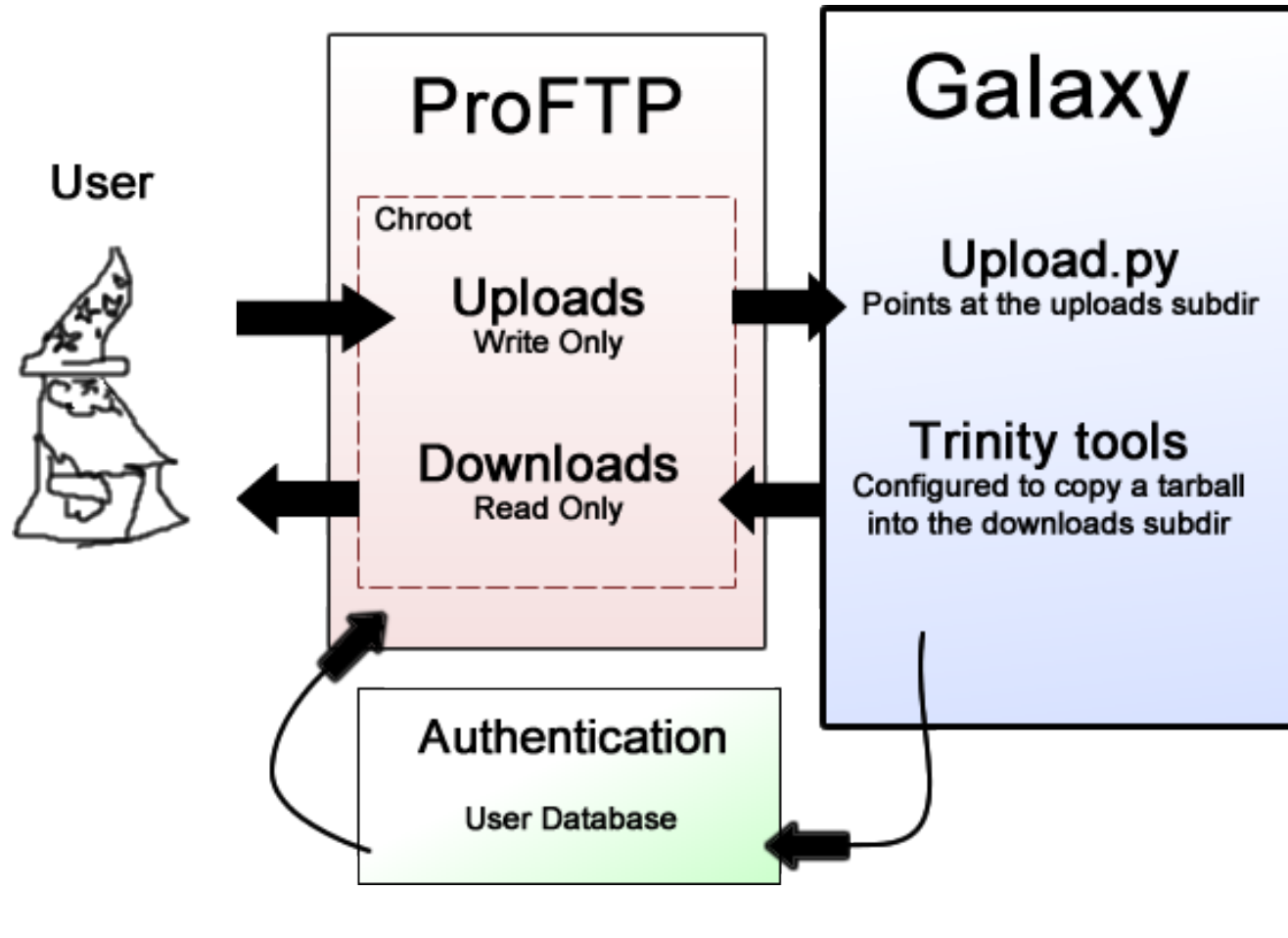
tools

parameters

`grouping.py`

api

`remote_files.py`





Challenges

Job Runners:

- PBS – still breaks with Torque 4x
 - Seems to have trouble interacting with the dynamic runner
- Drmaa – one destination possible?
- Dev servers are always somehow different than production!

Future Goals

- **Trinity needs some more guinea pigs!**
Sign up now!
- IU Galaxy still needs to be moved to Github
- OSG-BLAST is still in beta



Guinea Pig by Peter Griffin



INDIANA UNIVERSITY

Fin (for real)

Thanks for watching!

Questions and comments:

Email help@ncgas.org



PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY



BROAD
INSTITUTE

National Center for Genome Analysis Support: <http://ncgas.org>