

State of the Galaxy

2014

5

GDC2010 | CSHL



GCC2011 | Lunteren



GCC2012 | Chicago



GCC2013 | Oslo





First ever Galaxy Hackathon!

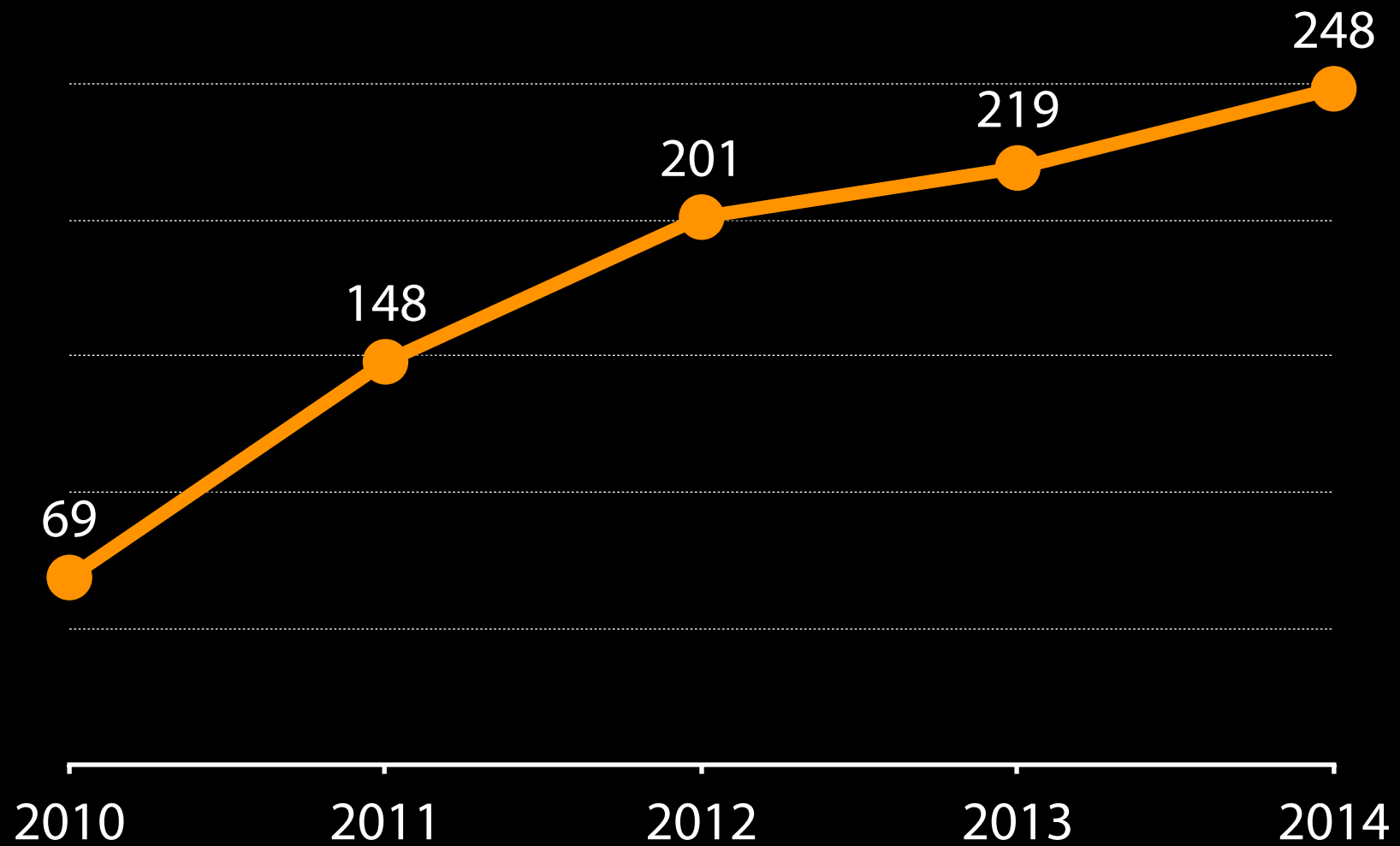
(See talk from Dannon Baker, Brad Chapman, John Chilton, Kyle Ellrott, *et al*)

- 1. State of the Community — GCC participation, code contributions, publications, Galaxy ToolShed et cetera**
- 2. State of Galaxy main — move to TACC (more detail), statistics of Galaxy main usage**
- 3. Some highlights from this last year — biostar, viz, dataset collections, toolshed, data managers, new UI: reference talks/lightning talks to come**
- 4. Where we're going — 10-ish priority areas**
- 5. Acknowledgements**

Community

(gcc, instances, code, toolshed, papers)

GCC participation





DBCLS Galaxy
An Integrated Environment for Databases in Life Science
ライフサイエンスにおけるデータベース統合環境として Galaxy の日本での普及を促進して

DBCLS
Database Center for Life Science

GeneNetwork
University of Tennessee: www.genenetwork.org

Galaxy @ ZURGI

Analyze mapped RNA-Seq data,

SunLab
DNA Data Bank of Japan

REGULATORY GENOMICS GROUP
at Pompeu Fabra University

50+ public instances

Cistrome
A Galaxy Server dedicated to ChIP-seq analysis
WAGENINGUR
For quality of life

The Populus Genome Integrative Explorer
popgenie.org

GENEONTOLOGY
Unifying Biology

uCC
Coláiste na hOllscoile Corcaigh, Éire
University College Cork, Ireland

globus genomics

DTL
DUTCH TECHCENTRE FOR LIFE SCIENCES

LiSIs
Life Sciences Informatics System

FastUniFrac

Ciphers

South Green

GWIS: Online e in minutes...

MGTXA
A Galaxy server for metagenomic taxonomy prediction

Bioinformatics Toolbox
SBML Tools
SBMLsqueezer
SBML2LATEX
BioPAX2GML
Transcription factor annotation
iFPredict
Expression data analysis
RPPA

12 Month Summary

Jun 30 2013 — Jun 30 2014

3738 Commits

Up **+840 (28%)** from previous 12 months

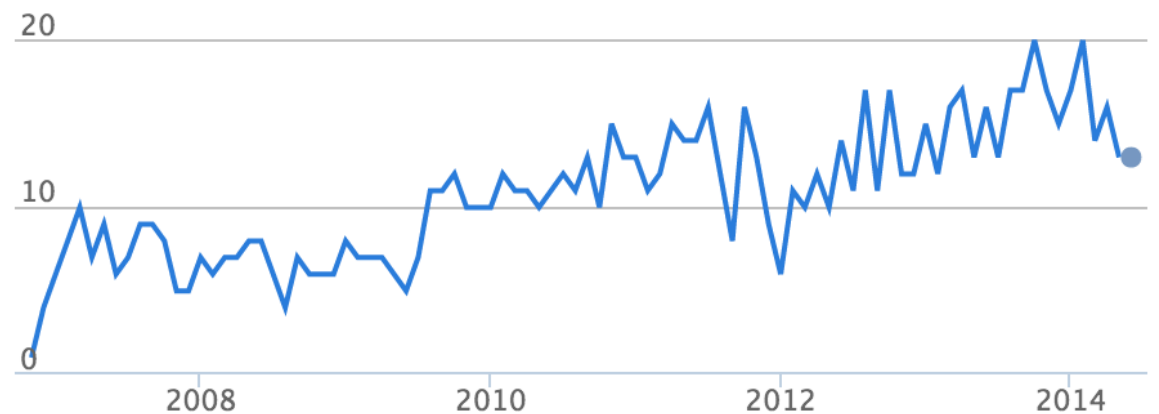
51 Contributors

Up **+6 (13%)** from previous 12 months

Commits per Month



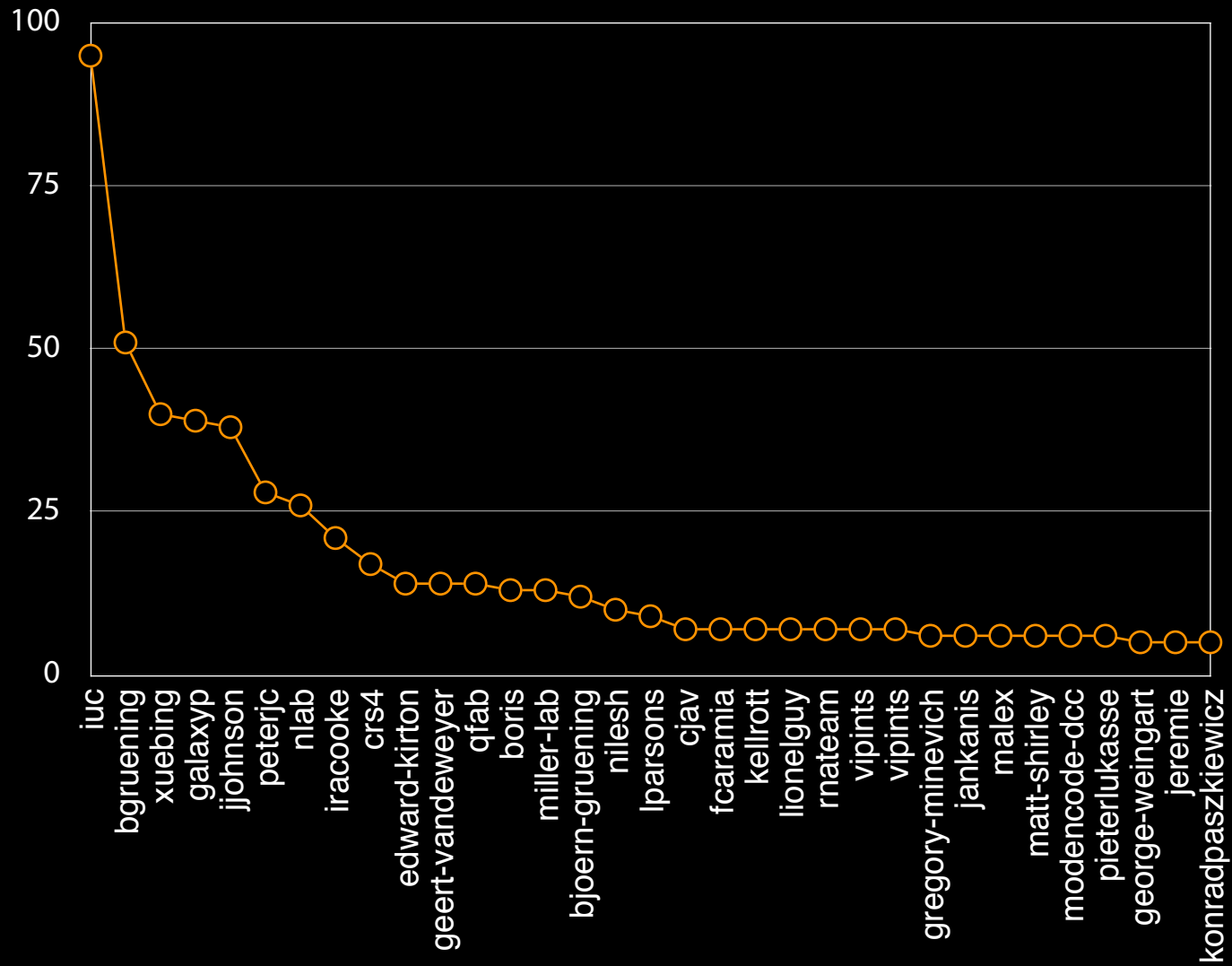
Contributors per Month



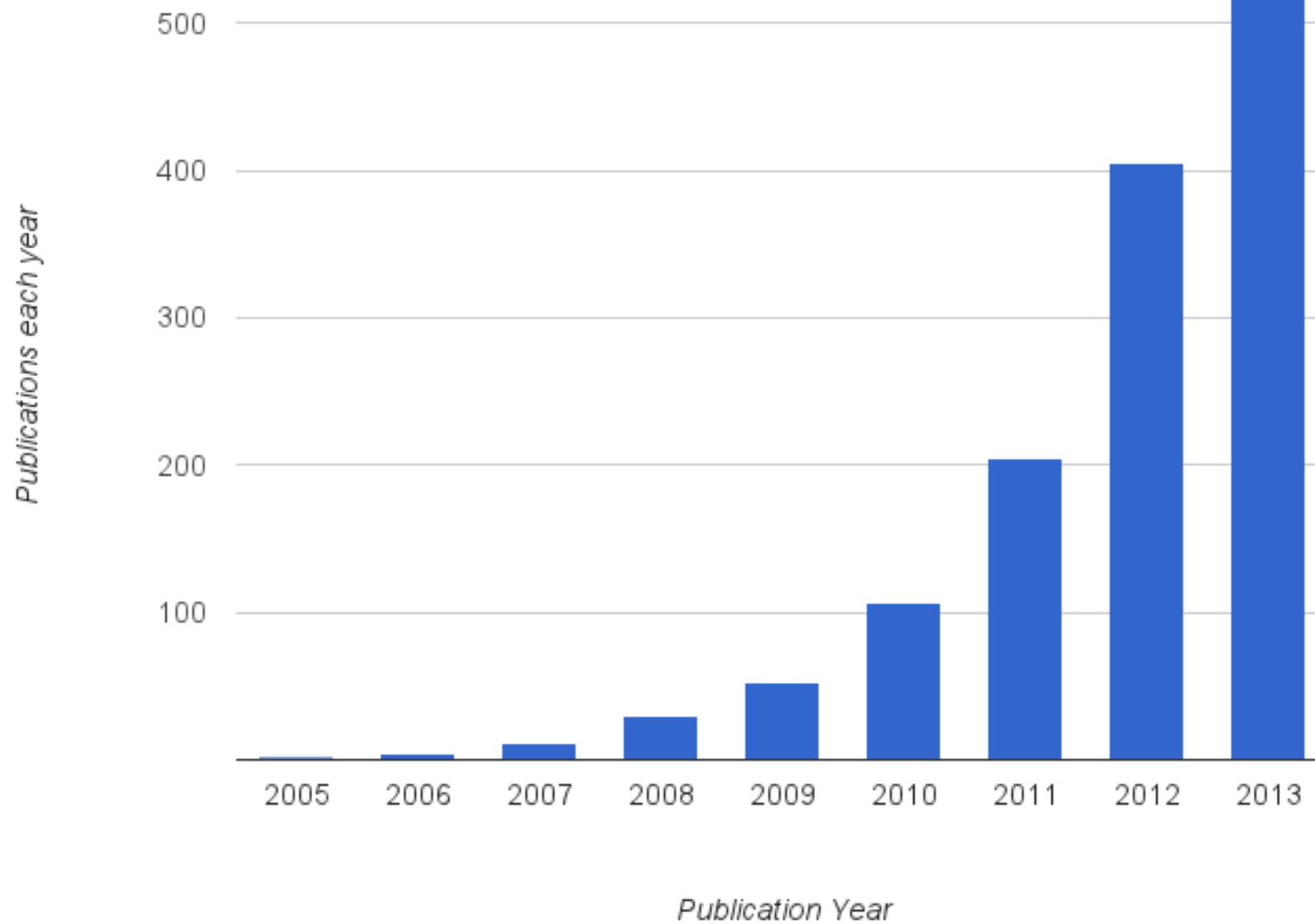
toolshed statistics

- 897 repositories
- 222 unique owners
- 176 Tool dependency package installation recipes
- 2,330 valid tools
- 3,420 valid versions of tools
- 54 exported Galaxy workflows
- 455 custom datatypes
- 62,021 total repository installations

toolshed contributions

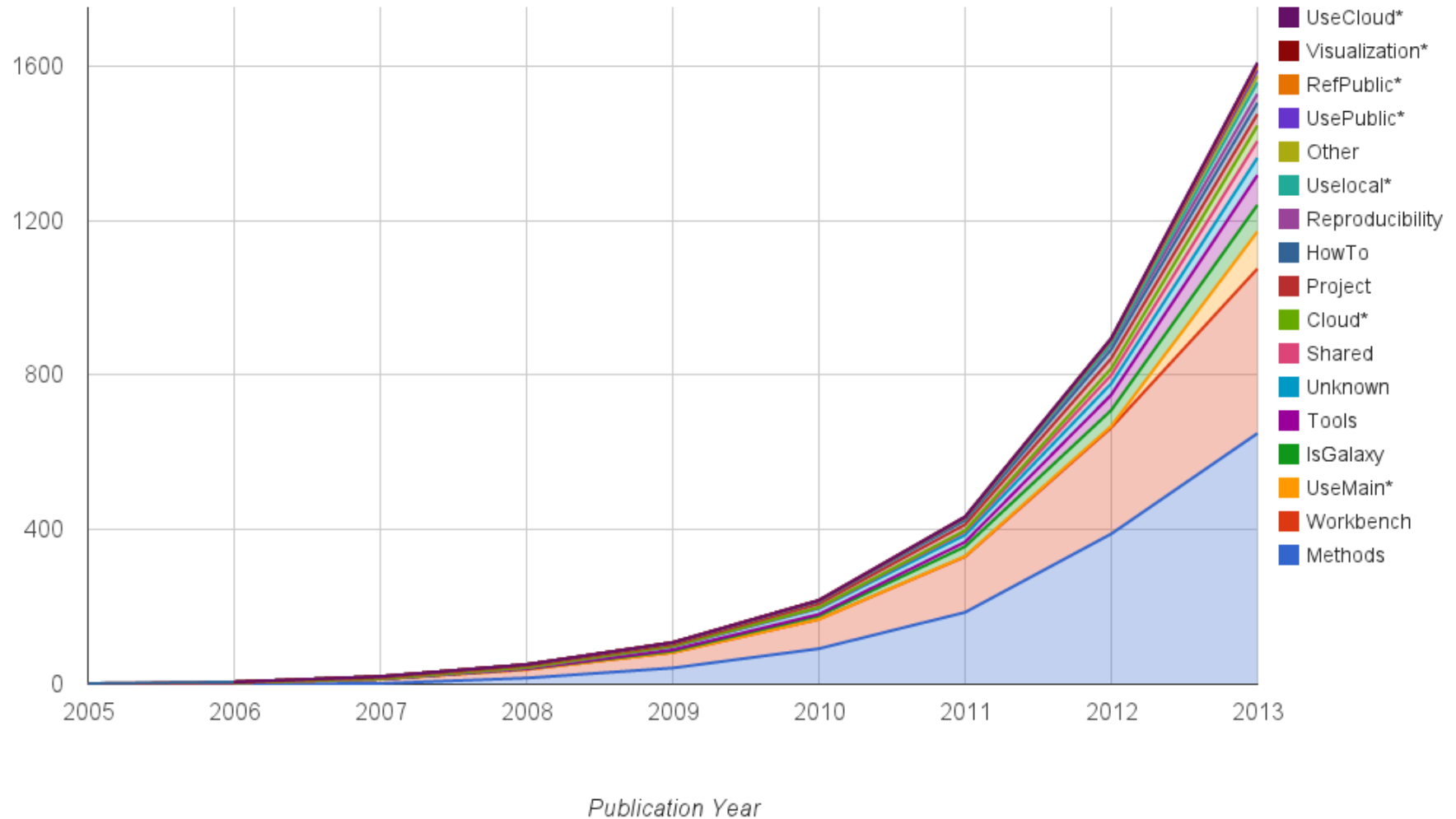


papers



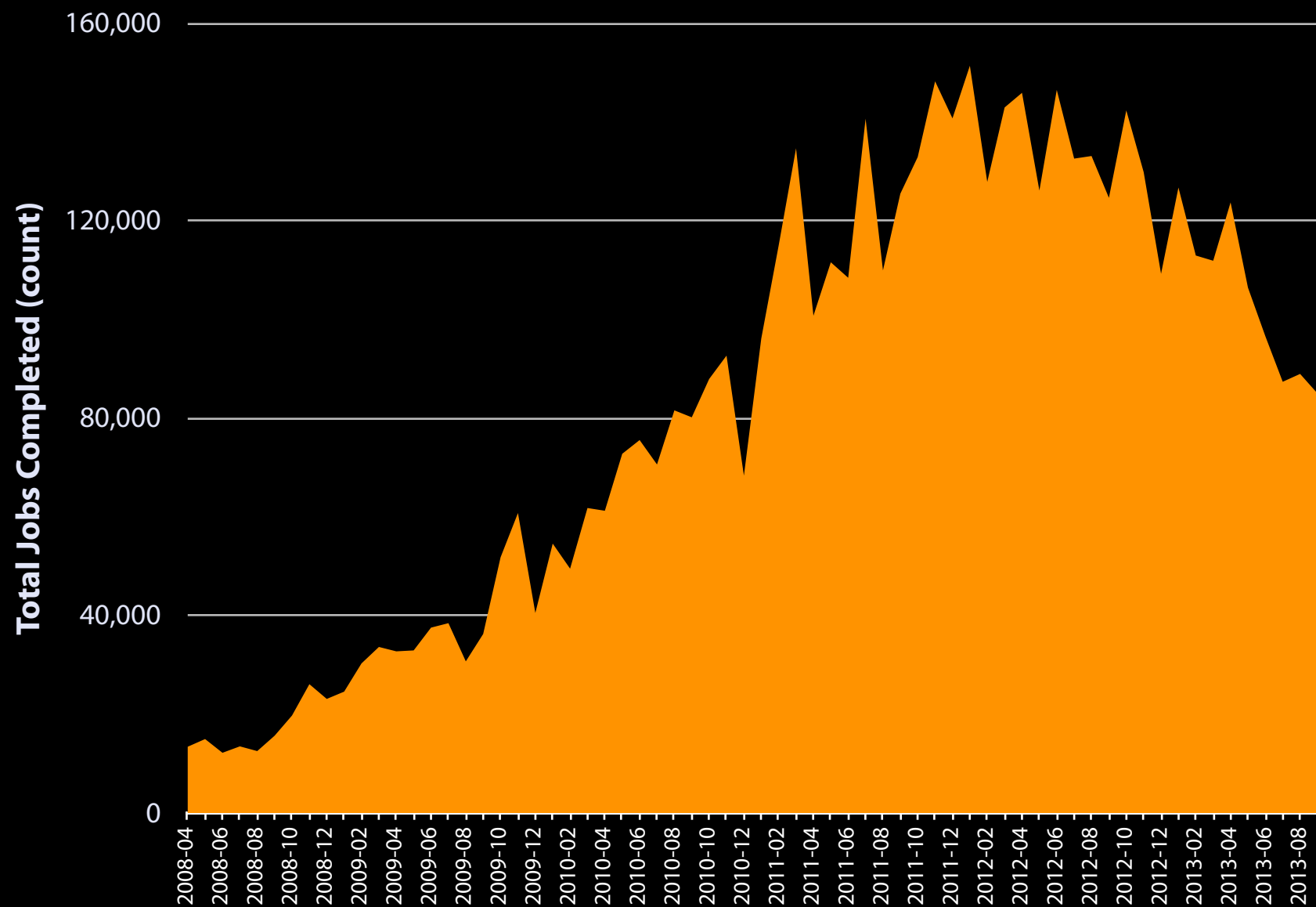
All publications: cumulative by tag

Publications: Cumulative Tags per Year 2005-2013



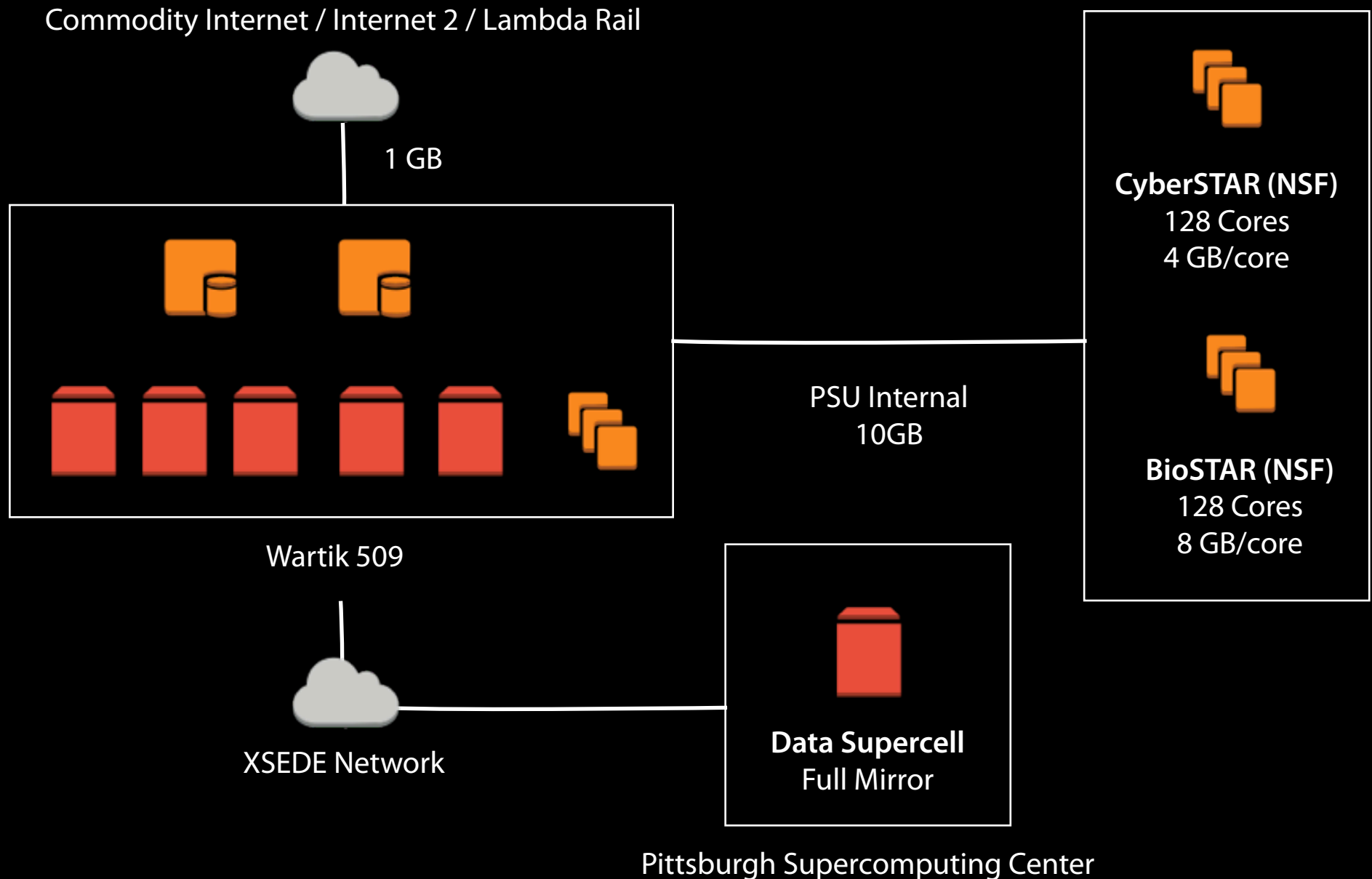
vital signs of galaxy main
(infrastructure, jobs, users)

job numbers crash of 2013





(2013) Galaxy Main Architecture: Extended



Galaxy Main at TACC

All physically co-located at the Texas Advanced Computing Center

Virtual machines



Web Front-end 1



Web Front-end 2



Master Database



Replicate Database



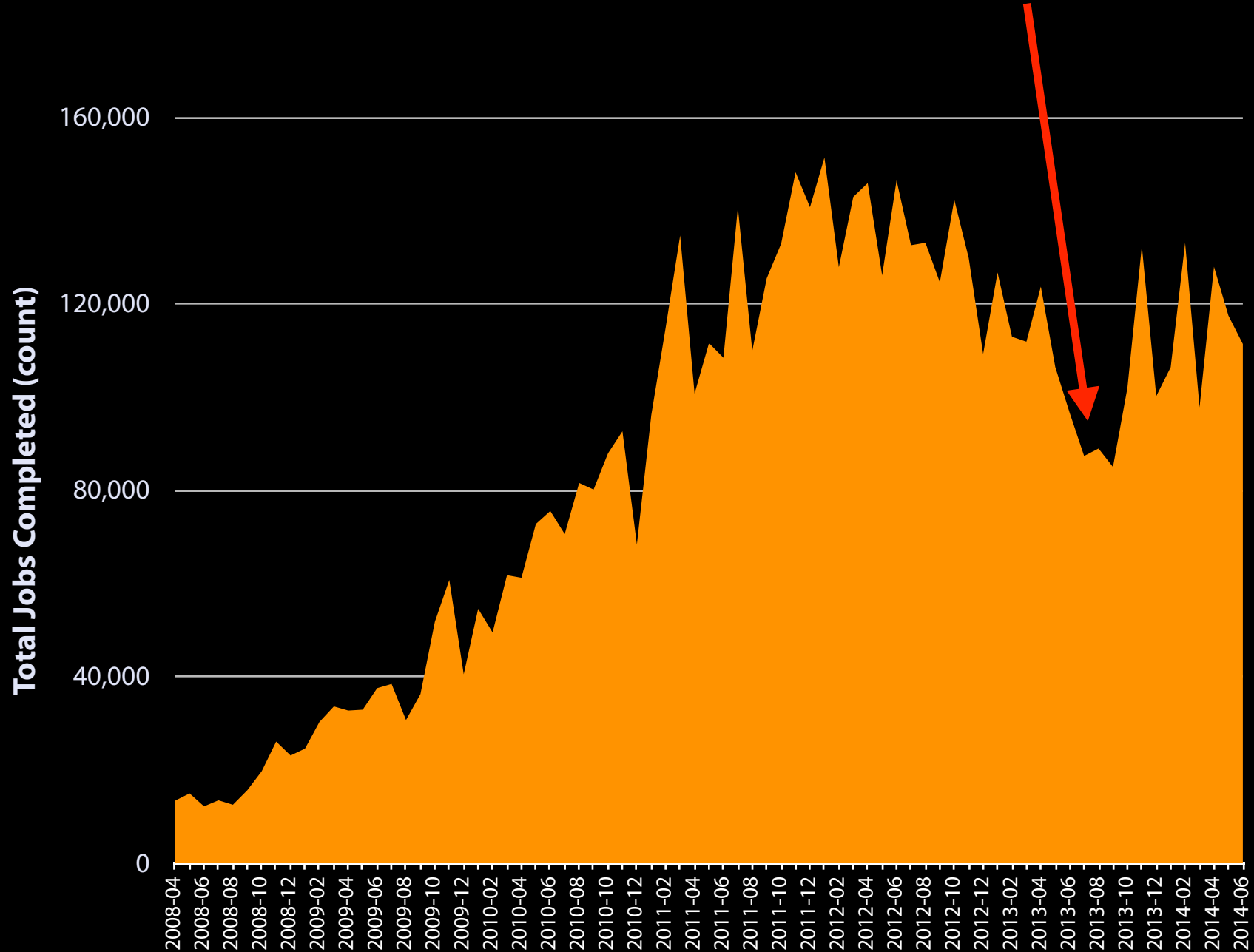
Corral
DDN Storage Appliance



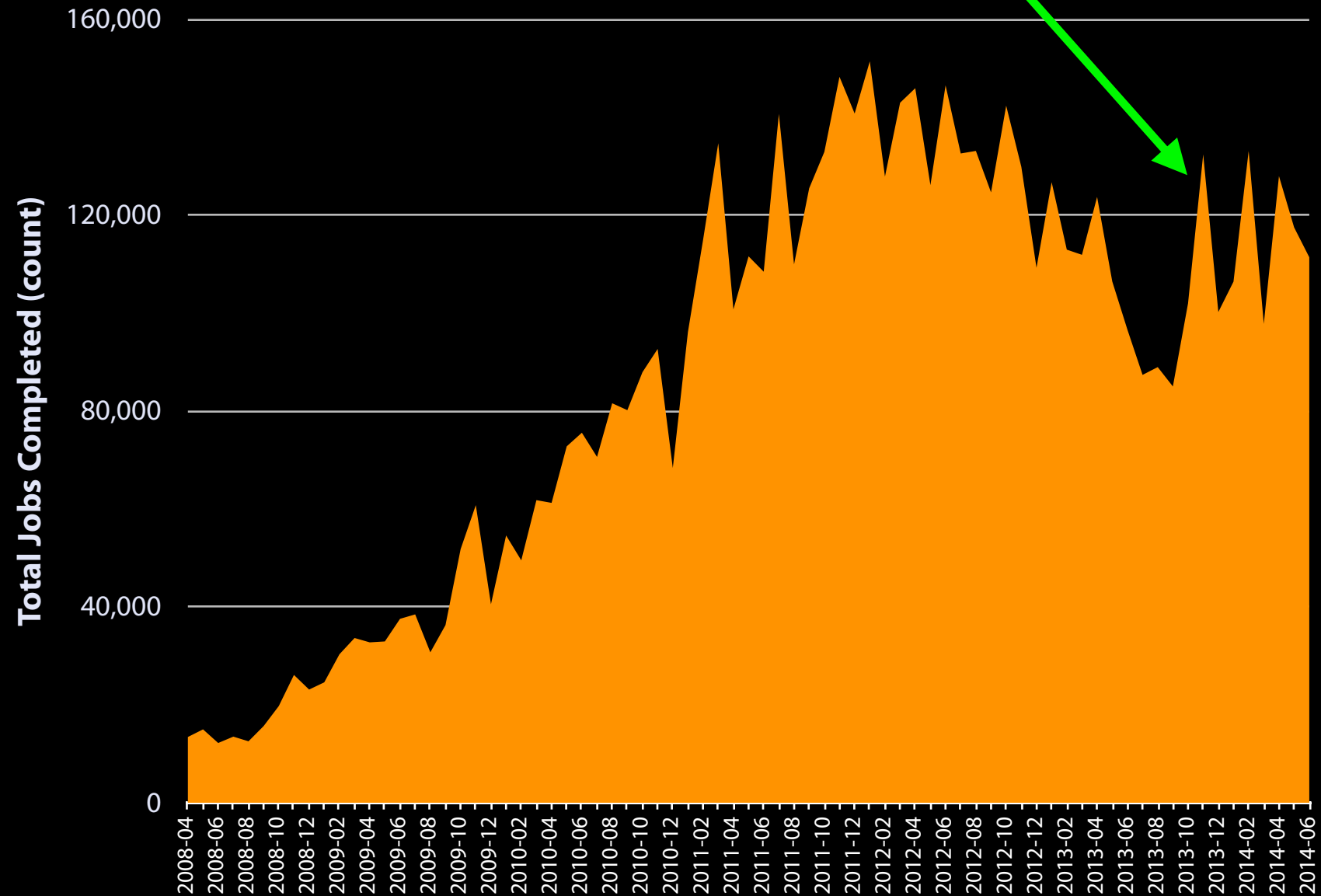
Galaxy Cluster
256 Cores
16 GB/core

Fileservers
926 TB of user data

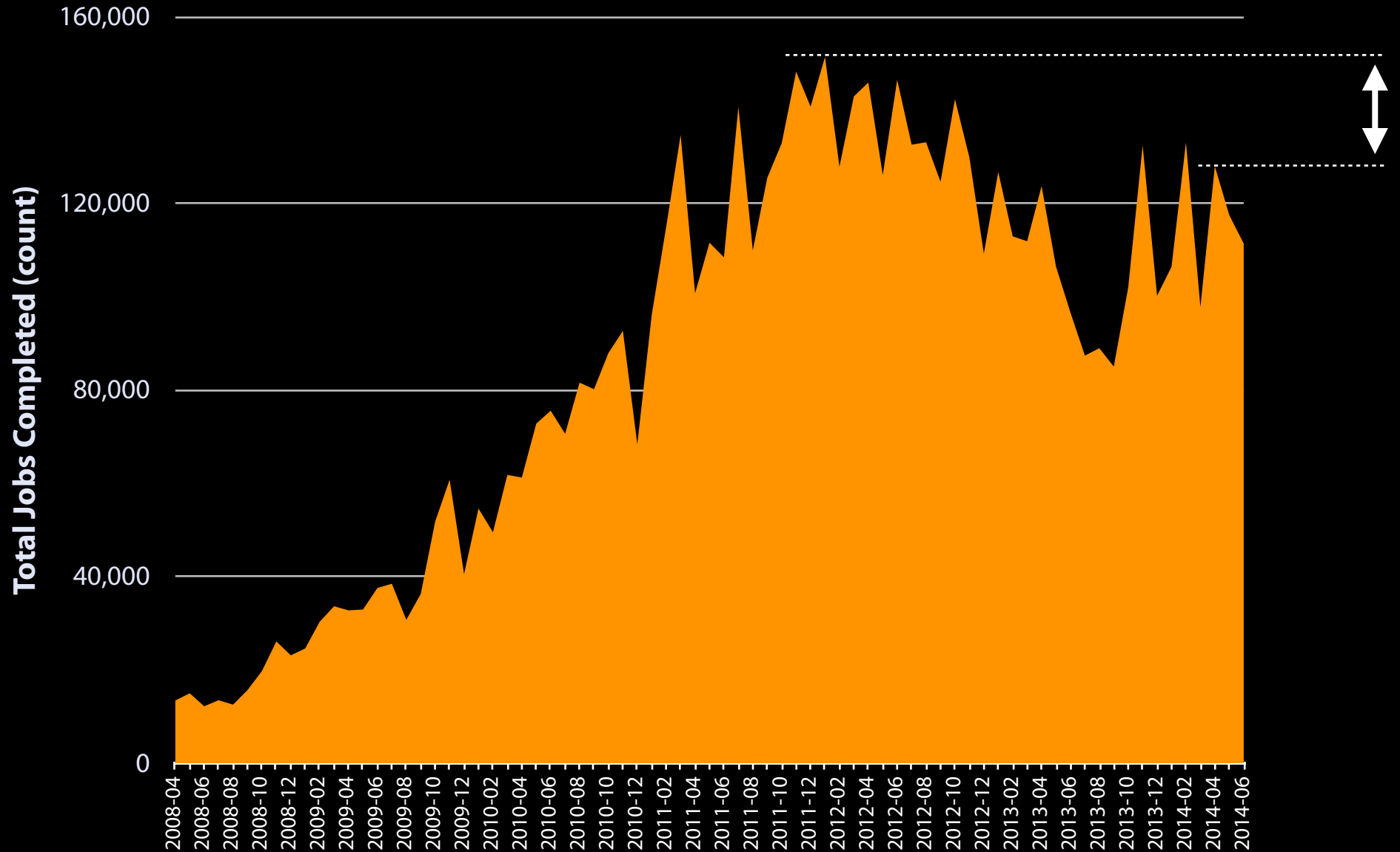
severe resource bottleneck



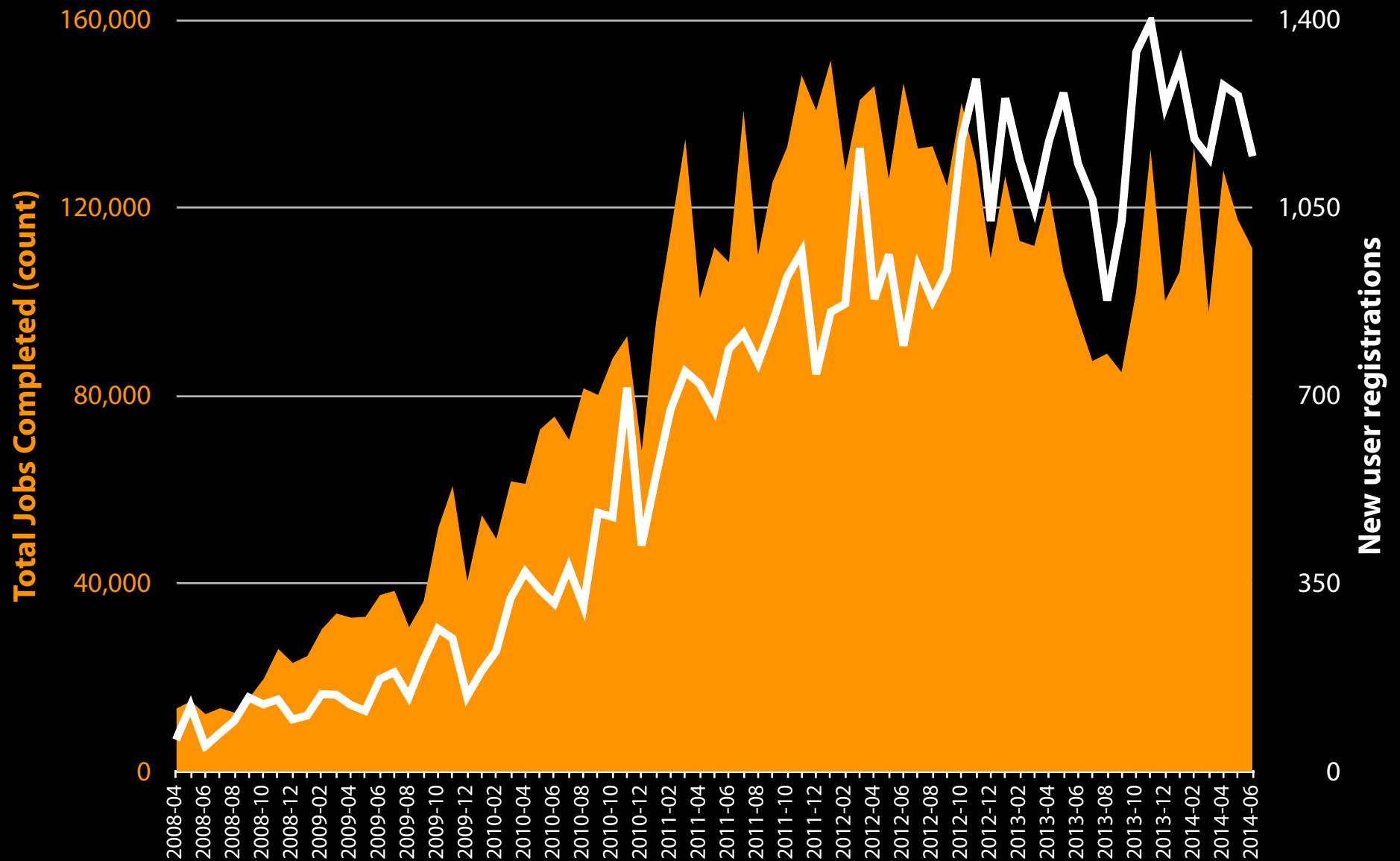
move to TACC



data is bigger = jobs are longer



user dynamics at usegalaxy.org



highlights of the past year
(from Oslo to Baltimore)

Expansion to external resources



Viz framework

127.0.0.1:8080/root/index — Galaxy

Galaxy

Analyze DataWorkflowShared DataVisualizationAdminHelpUser

Tools

search tools

Get Data

Send Data

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

Graph/Display Data

Regional Variation

Multiple regression

Multivariate Analysis

Evolution

Motif Tools

FASTA manipulation

NGS: QC and manipulation

NGS: Mapping

NGS: GATK Tools (beta)

NGS: Simulation

Phenotype Association

Workflows

All workflows

Charts - New Chart

List of created charts:

New ChartStacked areaLast change: 9/3/2014, 14:07

New ChartBar diagramLast change: 9/3/2014, 14:06

New Chart

Y-axis

3984.6

0.0

0.0

10.0

20.0

30.0

40.0

50.0

60.0

70.0

80.0

90.0

99.0

X-axis

○ Stacked

● Stream

○ Expanded

Customize

History

Unnamed history

23.9 KB

1: tabular.txt

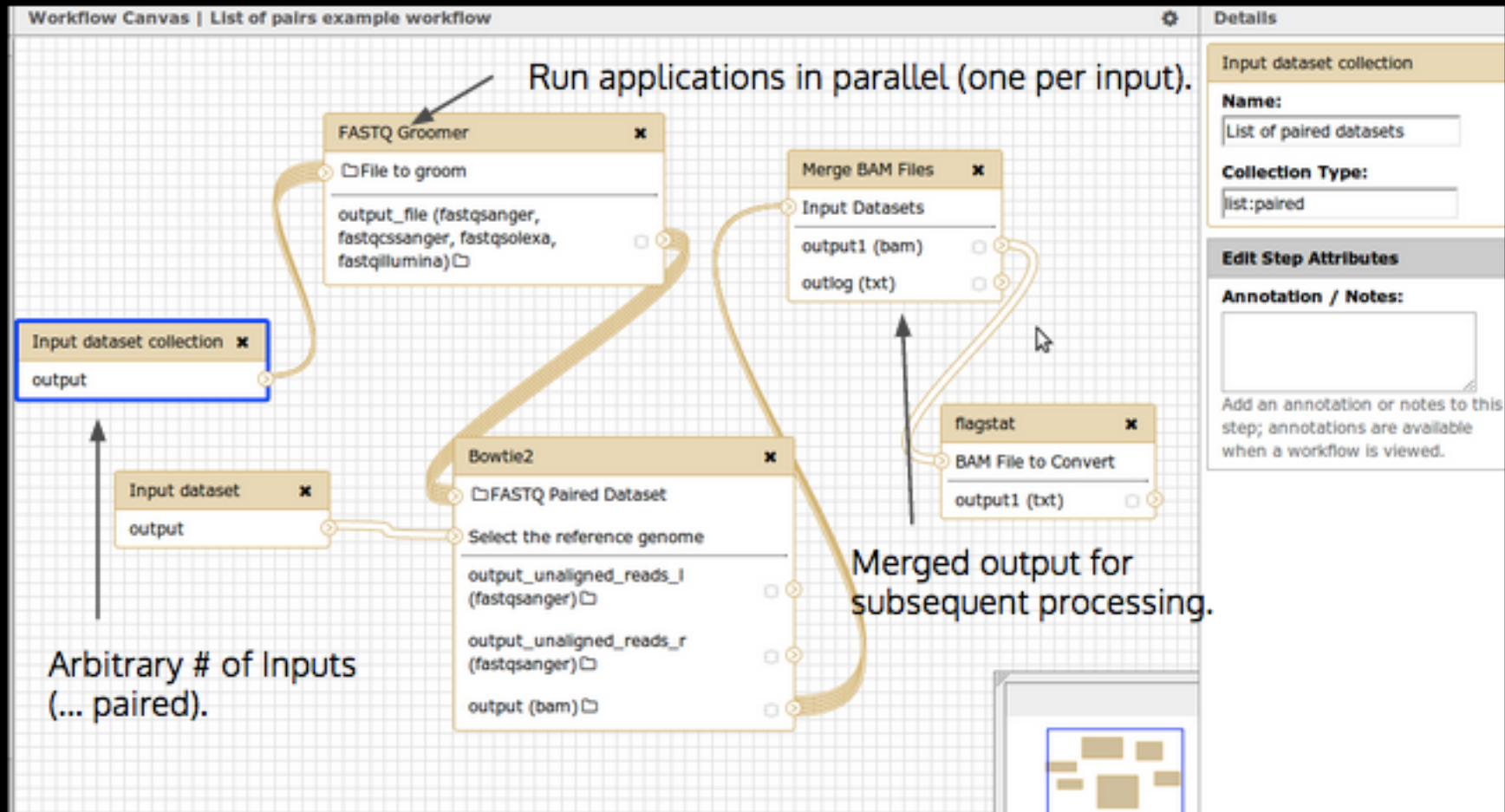
100 regions

format: interval, database: ?

uploaded interval file

1.Chrom	2.Start	3.End	4	5	6
43	14	75	95	85	5
33	100	32	20	17	5
5	60	46	54	34	8
35	73	40	58	21	2
45	3	36	35	5	1
6	84	89	72	30	7

Dataset collections



Galaxy/biostar

Latest Posts

← → ↺ 🏠

https://biostar.usegalaxy.org

☆ 🍌 📄 🔑 ☰

LATEST

OPEN

RNA-SEQ

CHIP-SEQ

SNP

ASSEMBLY

FORUM

PLANET

ALL »

 **Biostars**
GALAXY EXPLAINED

Welcome to Galaxy Biostar! User support for Galaxy!

about • faq • rss

Community

User Login

New Post

Live search: start typing...

or

Q Classic search

Limit to: all time ▾

<prev • 2,950 results • page 1 of 85 • next >

Sort by: update ▾

0 votes

0 answers

4 views

combining indexed paired reads from two lanes of a Hiseq rapidrun

ma-seq

written 8 hours ago by mduffy • 0

0 votes

0 answers

4 views

Forum: Need help with "Tophat2" tool

tophat2 forum

written 12 hours ago by martine.collart • 0

0 votes

1 answer

7 views

WIG file from MACS2

macs2 cistrome

written 16 hours ago by katherinedixon4 • 0 • updated 13 hours ago by Jennifer Hillman Jackson • 15k

0 votes

1 answer

10 views

I want to delete my account

galaxy

written 1 day ago by christine.nolte • 0 • updated 13 hours ago by Jennifer Hillman Jackson • 15k

0 votes

1 answer

20 views

phyloP not recognizing any of the HG19 tags as being HG19

metadata phyloP

written 3 days ago by caffletezbrant • 0 • updated 13 hours ago by Jennifer Hillman Jackson • 15k

1 vote

2 answers

28 views

Problems adding bulid-in genomes for tophat on local galaxy

galaxy tophat

written 4 days ago by bimbam42 • 0

0 votes

1 answer

15 views

rsID from list of chromosome and position

snp

written 2 days ago by sookwah.yee • 0 • updated 1 day ago by Bjoern Gruening • 810

Recent Votes

- C: Problems adding bulid-in genomes for tophat on local galaxy
- C: Condor And Galaxy
- A: Condor And Galaxy
- A: Condor And Galaxy
- C: Split barcodes tool output
- A: Should galaxy default to fastqsanger instead of fastq?
- A: Need help with "Variant Annotator" tool

Recent Locations • All »

- United Kingdom, 10 minutes ago
- United States, 47 minutes ago
- United States, 1 hour ago
- Netherlands, 1 hour ago
- European Union, 1 hour ago

Recent Awards • All »

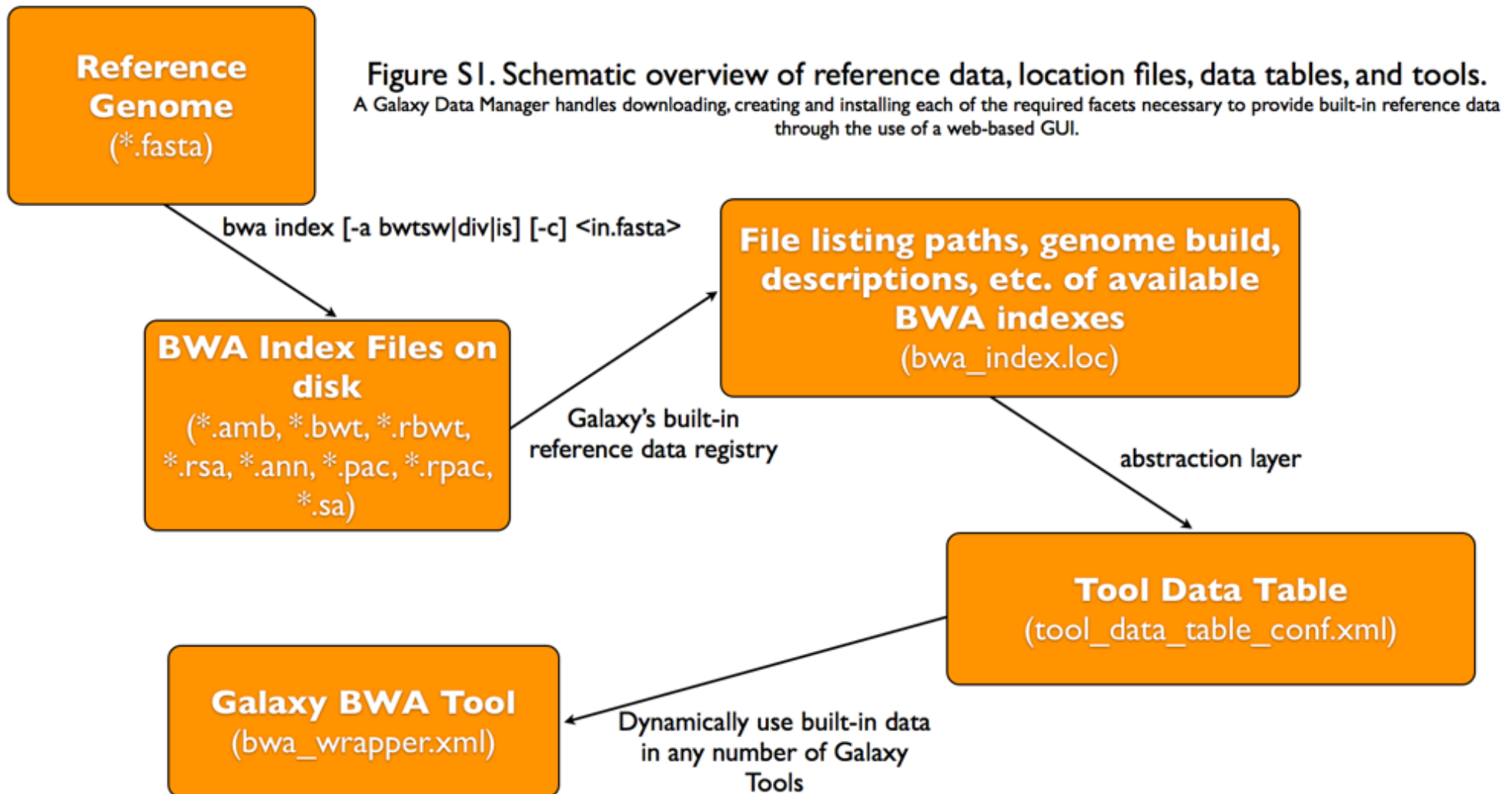
- Supporter 🍌 to Bjoern Gruening • 810
- Scholar 🏆 to Bjoern Gruening • 810
- Teacher 🍌 to fubar • 640
- Autobiographer 📖 to fubar • 640
- Scholar 🏆 to Daniel Blankenberg • 960
- Scholar 🏆 to Bjoern Gruening • 810

Traffic: 15 users visited in the last hour

Toolshed

```
<actions>
  <action type="shell_command">git clone --recursive git://github.com/ekg/freebayes.git</action>
  <action type="shell_command">git checkout 9608597d12e127c847ae03aa03440ab63992fedf</action>
  <action type="shell_command">git submodule update --recursive</action>
  <action type="shell_command">make || ( make clean && sed -i.bak -e 's:LIBS = -lz -lm -L./ -L../vcflib/tabixpp/
  <action type="move_directory_files">
    <source_directory>bin</source_directory>
    <destination_directory>$INSTALL_DIR/bin</destination_directory>
  </action>
```

Data managers



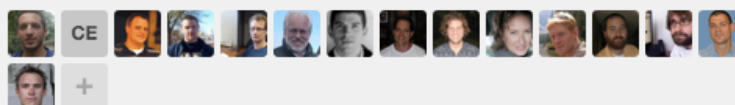
highlights of the future
(from Baltimore to ...)

📁 Galaxy: Development 🔒 Galaxy Project

< [Show sidebar](#)

📁 PRIORITIES 2014 in list [Project in Planning](#) 🔍

Members



Description [Edit](#)

This list contains items in order of priority as identified at March on the Baltimore team meeting.

They should include links to all cards that have been identified as parts of this effort.

☰ User Metrics + Analytics

- 0%
- ☐ Jeremy, James, Martin
 - ☐ [RFC: Implement sophisticated user behavior analysis tool](#)
 - ☐ [Core: Fine-grained logging](#)
 - ☐ [User Metrics and Analytics](#)
- [Add item](#)

☰ UI for large scale analyses

- 0%
- ☐ John, Carl, Anton (user cases)
 - ☐ [Dataset Collections - Models, API, Tools and Workflows.](#)

Add

👤 Members

🏷️ Labels

☑️ Checklist

🕒 Due date

📎 Attachment

Actions

➡️ Move

📄 Copy

👁️ Subscribe ☒

👍 Vote

📁 Archive

[Share and more...](#)

Pull Requests / Patches 🔍

665: Patch for FASTQ paired-end issue

Tools: Bowtie Wrapper Pull Requests from Community

add rna seq metrics and downsample sam to picard tools

Please merge patch to bowtie2 wrapper (adds support for mapping fasta files)

Enhancement: Extract Genomic DNA - GTF parsing

Pull Request #281 - tools/fastq/fastq_paired_end_joiner: added support for recent Illumina headers

[Add a card...](#)

Complete niche-specific tools/workflows

New workflow scheduling engine

User Interface for true large-scale analysis

Visualization/Visual Analyses

ToolShed evolution

Cloud/Virtualization/Distributed Data/
Federation

Scalable Training

thank you!

Organizing Committee

Dave Clements, Johns Hopkins University
Mohammad Heydarian, Johns Hopkins University
Dan MacLean, The Sainsbury Laboratory
Karen Reddy, Johns Hopkins University

Scientific Committee

Jeremy Goecks, George Washington University
Jessica Kissinger, University of Georgia
Karen Reddy, Johns Hopkins University

Hackathon Committee

Dannon Baker, Johns Hopkins University
Brad Chapman, Harvard University
John Chilton, Penn State University
Kyle Ellrott, University of California Santa Cruz (UCSC)

Support

Stacey Hooker (GCC)
Teal Golden and Team (Housing)
Paula Davis (Hackathon)

thank you!

Training Day Infrastructure

Dannon Baker

Dave Bouvier

Registration and Setup Volunteers

Eric Rasche

Carl Eberhard

Teresa Romeo Luperchio

Xianrong (Jose) Wong

Nick Stoler

Martin Čech

BoF organizers

Nikolay Aleksandrov Vazov

Dave Clements

Mo Heydarian

Philip Blood

John Chilton

Nate Coraor

Carrie Ganote

thank you!

Training Day Instructors

Jeremy Goecks (x2)
Aysam Guerler (x2)
Tom Bair
Jennifer Jackson
Nate Coraor
John Chilton
Simon Gladman
Andrew Lonie
Nikolay Vazov (x2)
Katerina Michalickova (x2)
Dannon Baker
Carl Eberhard
Greg Von Kuster

Training Day Instructors

JBjörn Grüning
Peter Cock
Saskia Hiltemann (x2)
Youri Hoogstrate (x2)
Hailiang (Leon) Mei (x2)
Jonas Paulsen
Tonje Lien Gulbrandsen
Morten Johansen
Karen Reddy
JJ Johnson
Dan Blankenberg
Ntino Krampis
Enis Afgan
Ravi Sanka
Brad Chapman

ion torrent



by *life* technologies™



EMC ISILON



globus
genomics

(GIGA)ⁿ
SCIENCE



BIO  MARYLAND

