

Yet another on-demand Galaxy cloud, but only powered by Apache CloudStack Youngki Kim, kt, Korea

Abstract

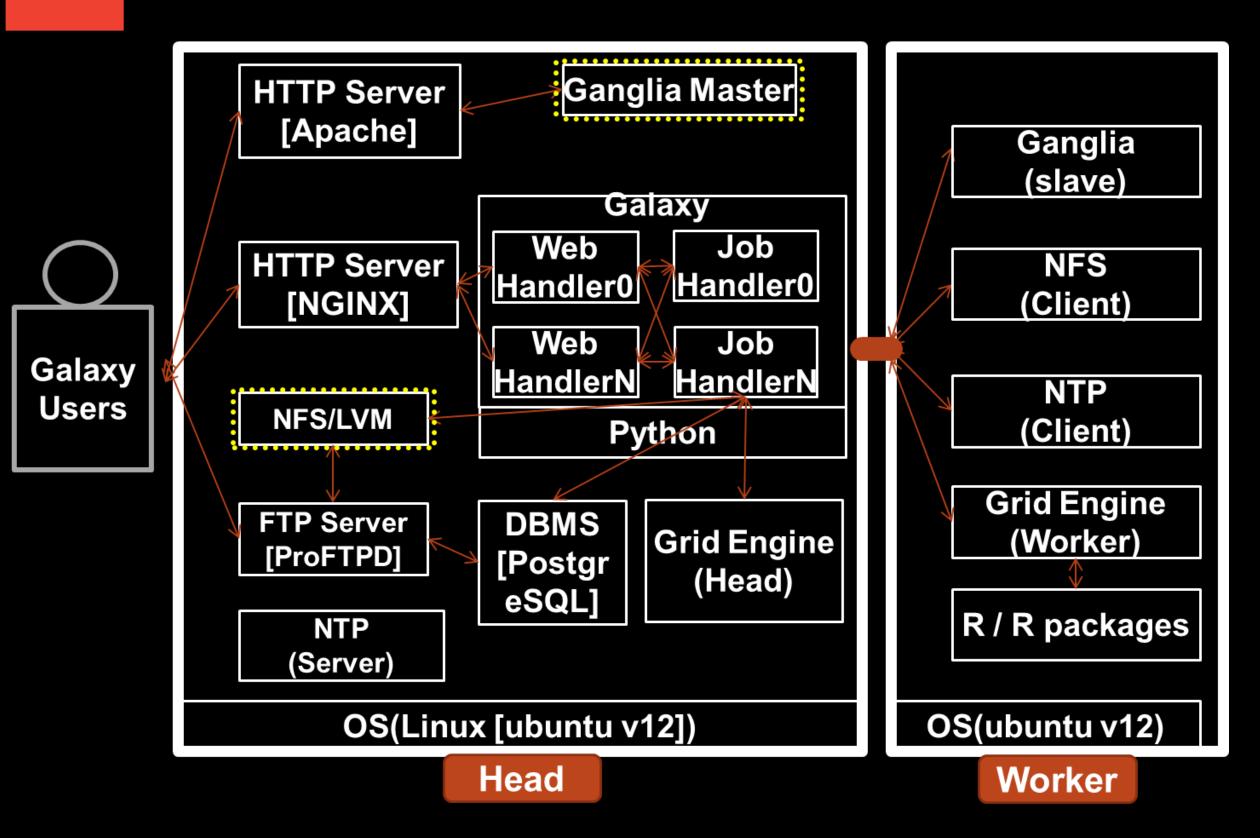
GenomeCloud is an integrated platform for analyzing, interpreting and storing genome data, based on KT's cloud computing infrastructure which uses Apache CloudStack software. GenomeCloud consists of g-Analysis (automated genome analysis pipelines at your fingertips), g-Cluster (easy-of-use and cost-effective genome research infrastructure) and g-Storage (a simple way to store and share genome-specific data).

Because of flexible tool integration architecture and seamless workflow creation functionality, Galaxy was selected to achieve multi purpose goals such as agile pipeline development and bioinformatics education support. To provide on-demand and Apache CloudStack based Galaxy cluster, we have automated virtual machine creation, clustering and various software setup including Galaxy.

Furthermore, seamless integration with GenomeCloud helps researchers not only create and manage Galaxy through a convenient web interface but also fully utilizes genome data in g-Storage. g-Storage is powered by OpenStack Swift and specially designed genome file transfer protocol. Galaxy on the GenomeCloud uses Grid Engine as a Cloud HPC Solutions, Ganglia as a distributed monitoring system and LVM over NFS as a large volume shared storage, all of which are setup automatically upon request.

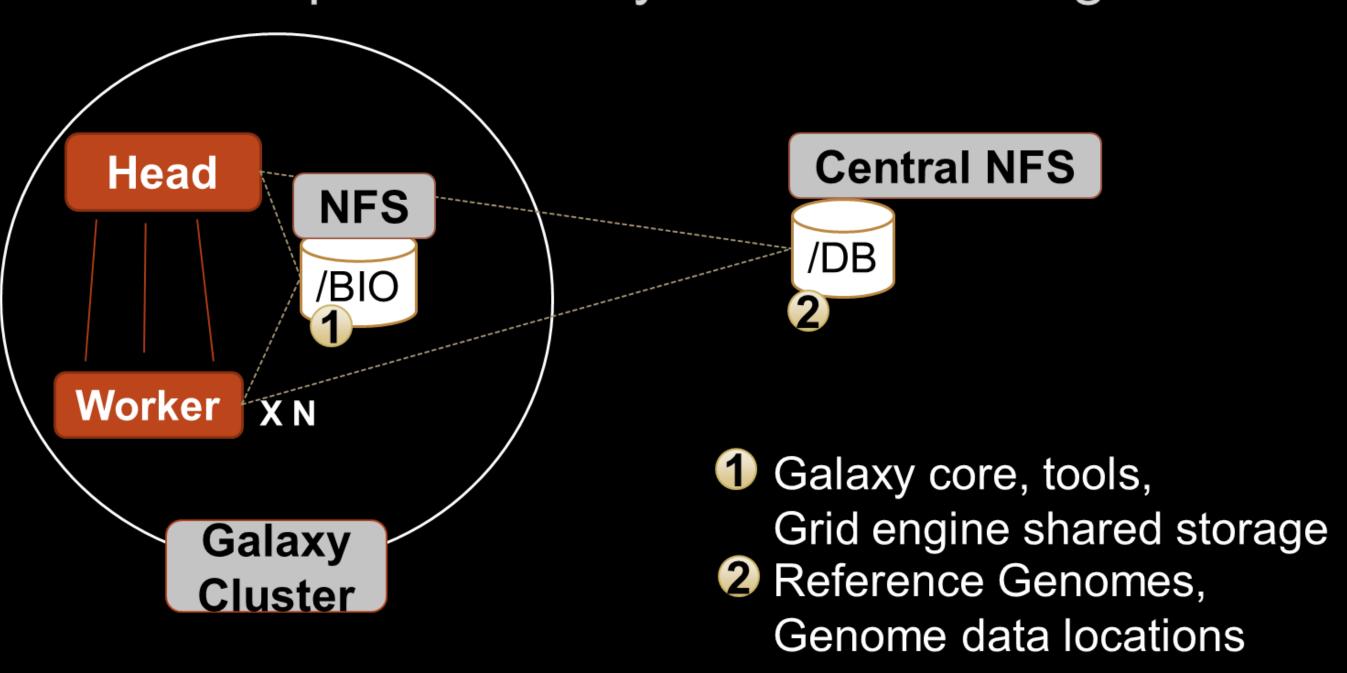
Galaxy on the GenomeCloud

Galaxy software stack



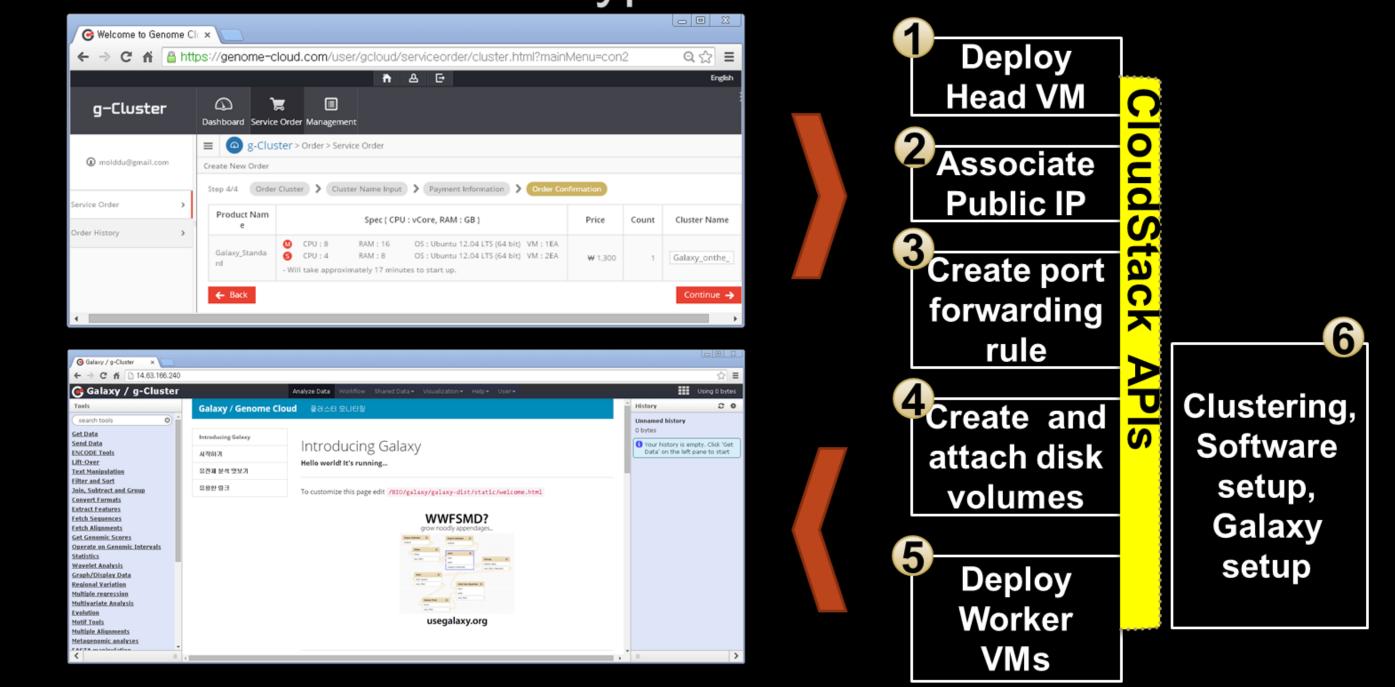
Decoupled multi layer shared storages

Galaxy system architecture



Fully automated cluster creation

Select cluster type → name it → use it



Integration with g-Storage

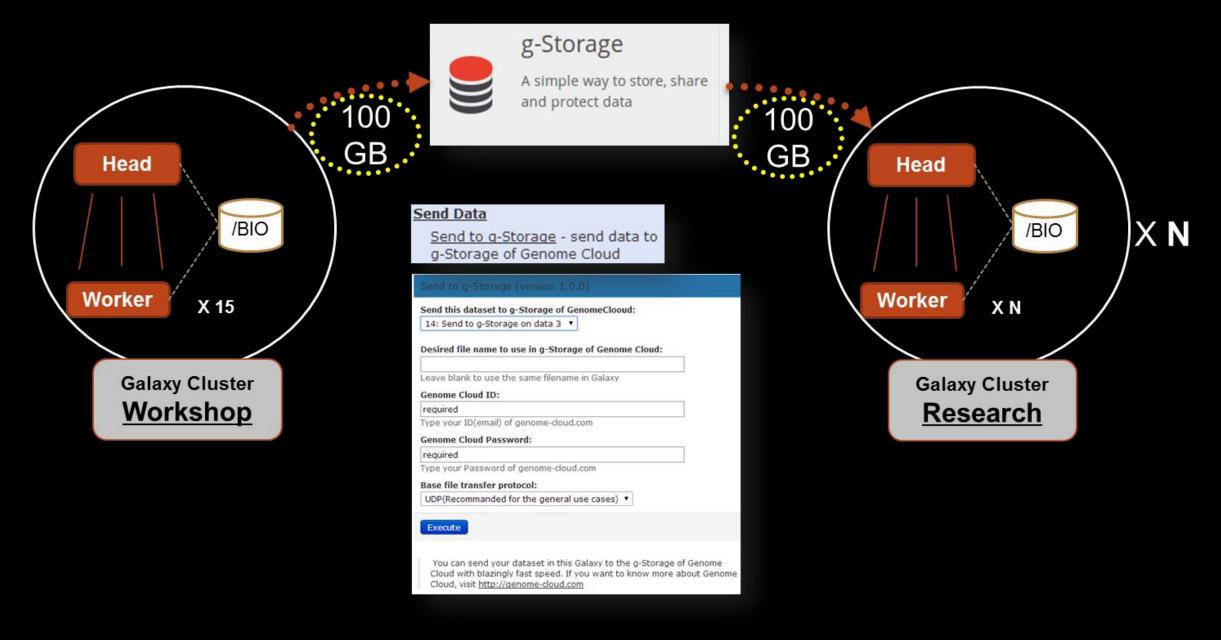
Pre-installed pipelines

Supporting diverse genome references



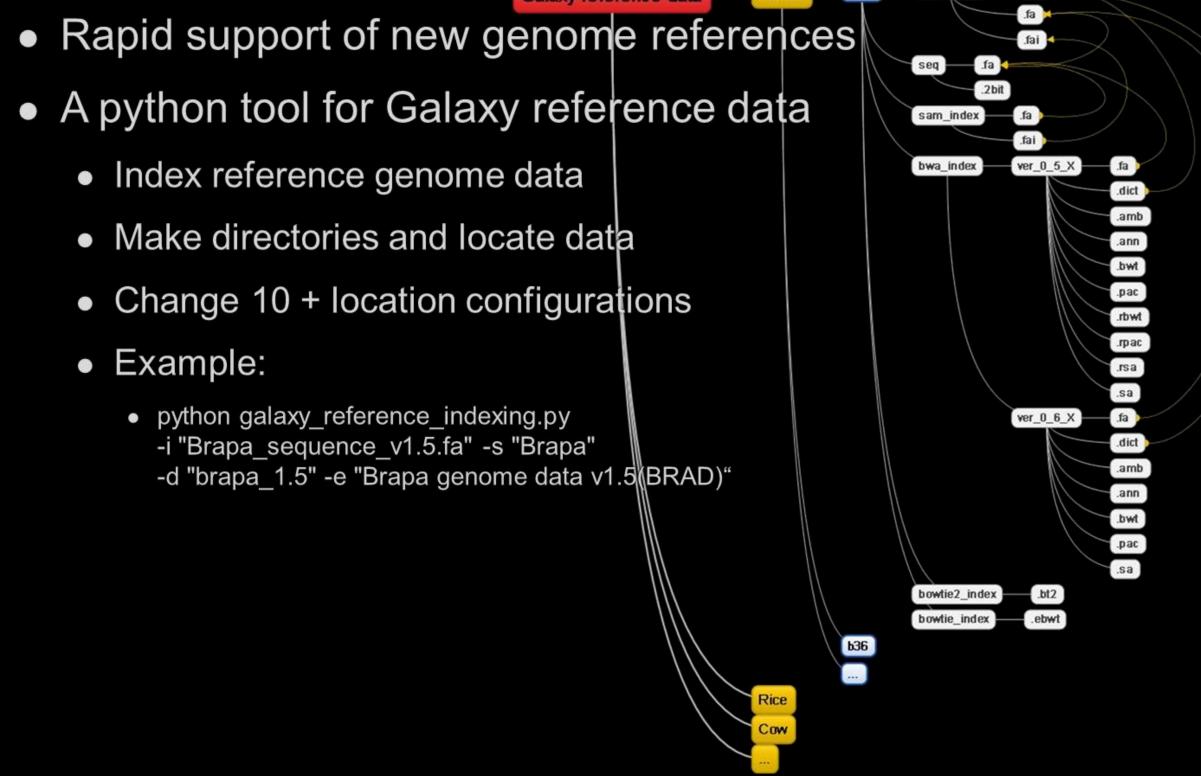
Inter Galaxy cluster data transfer

Develop a galaxy tool for sending large data



Workflows for RNA-Seq(Tuxedo, Ion Proton) analysis

Galaxy/g-Cluster ×				Condition A Cond
← → C ff 14.63.171.237/workflow/editor?id=0	0a248a1f62a0cc04	Q	☆ =	Reads
🕝 Galaxy / g-Cluster	Analyze Data Workfow Shared Data- Vesselzation	- Admin Help- User- Using	ig 466.7 GB	¥¥.
Workflow Canvas KNA-Seq for IonProton Beta			•	TopHat
reads RNA-Seq PASTQ file, reverse reads	Image: Start / Mark Adjustment Summary Image: Start / Mark Adjustment Summary Image: Start / Mark Adjustment Summary SAM/BAM dataset to generate startistics for html_file (html) SAM/BAM dataset to generate startistics for html_file (html) Prist file with file output1 (bam) outing (txt) Here Startistics for html_file (html)	penerate Commer SAM or BAM file of aligned RNA-Seq reads Global model (for use in Trackater) genes_expression (tabular) transcripts_expression (tabular) assemblist_softrms (gf) total_map_mass (bt) Util_map_mass (bt) Util_map_mass (bt)		Mapped reads Cufflinks Assembled transcripts Asse
Junctions (bed) accepted_hts (barn) ummacred_barn (barn) ummacred_barn (barn) ummacred_barn (barn) ummacred_barn (barn) put to the second barn (barn) to utput to utput	Mark Durb. ## real 23 Pest file SAM/BAM dataset to mark duplicates with file output1 (barn) output1 (barn) output1 (barn) outlog (txt) output1 (barn)	Comment SAM or BAM file of aligned RNA-Seq reads Comment Same Reads Comment Same Croup 1 > Replicate 1 > Add file Select backword database (selts) Global model (for use in Trackater) genes_expression (tabular) ranscripts_expression (tabular) assembled_informs (glf) Comment Same Croup 1 > Replicate 1 > Add file Select backword database (selts) Units (for use in Trackater) Compose info (tabular) run_info (tabul) Image: Select backword database (selts) Select backword database (selts) Units (for use in Trackater) Compose info (tabular) splicing_dff (tabular) Image: Select backword database (selts) Units (for use in Trackater) Select backword database (selts) Select backword database (selts) Units (for use in Trackater) Compose info (tabular) splicing_dff (tabular) Image: Select backword database (selts) Units (for use in Trackater) Select backword database (selts) Select backword database (selts) Units (for use in Trackater) Select backword database (selts) Select backword database (selts) Units (for use in Trackater) Select backword database (selts) Select backword database (selts) Units (for use in Trackater) Select backword database (selts) Select backword database (selts)		Cutfmerge Final transcriptome assembly
delettons (bed) junctions (bed) accepted_hts (bam) unmapped_bam (bam)	SAM IN FREED 2 BAN/SAM file FASTQ file output_fastq1 (fastqsanger) rASTQ file output_fastq2 (fastqsanger) output_unalgend_reads_l' (fastqsanger) BAN/SAM file 0 output_fastq1 (fastqsanger) 0 BAN/SAM file 0 output_fastq1 (fastqsanger) 0 BAN/SAM file 0 output_fastq1 (fastqsanger) 0 Dubut_fastq1 (fastqsanger) 0	cds_dff (tabular) cds_dff (tabular) cds_exp_dff (tabular) cds_fpixm_tracking (tabular) cds_count_tracking (tabular) cds_count_tracking (tabular) cds_read_group_tracking (tabular) tss_groups_count_tracking (tabular) tss_groups_read_group_tracking (tabular) tss_groups_read_group_tracking (tabular) tss_groups_read_group_tracking (tabular) tss_groups_read_group_tracking (tabular) tss_groups_count_tracking (tabular) tss_groups_read_group_tracking (tabular) tsoforms_count_tracking (tabular) tsoforms_count_tracking (tabular)	eadby	Mapped reads Cuffdiff Differential expression results
>	output_fastq2 (fastqsanger) PASTQ file PASTQ file output_unaligned_reads_l (fastqsanger) output_unaligned_reads_r (fastqsanger) output (bam)	soforms_count_tracking (tabular) genes_read_group_tracking (tabular) genes_fpkm_tracking (tabular) genes_count_tracking (tabular) genes_count_tracking (tabular) cutifiata (cutifiata)		CummeRbund Expression plots



Use case & Acknowledgement



4 half-day workshops



Dates: 13th, 20th May, 3rd, 10th Jun

Attendees : 50 +

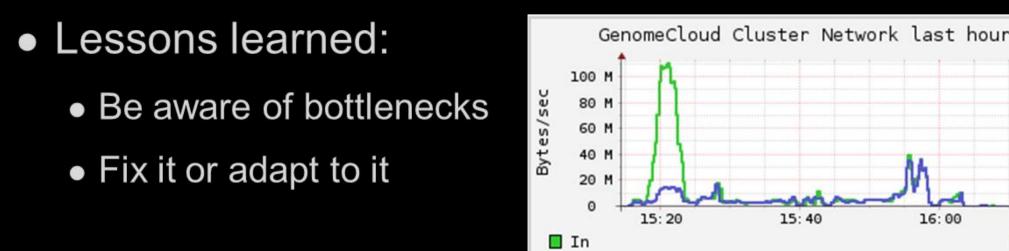
Galaxy: 8 core 16GB X 16 servers

Contents : from fastQC to RNA-Seq Analysis

Off site workshop and home works

of executed jobs: 5,000 +

Feedback: good to continue further research



Galaxy team

• -> We could never start this without you

GenomeCloud team

• -> We could never **finish** this without you

• Daechul choi, Changbum hong, Kwangjoong kim, Wanpyo hong, Hankyu choi, Hosang jeon, Sehyuk yoon, Eunjean jo



Galaxy Community Conference 2014, JHU, USA

