

Cloud Computing Resources, Grants, and Datasets

BD2K 2018 Workshop

Aditya Srikanth Kovuri

Sandeep Namburi



Outline

- What is the cloud
- Cloud in genomics and bioinformatics
- Myths of the cloud
- Galaxy Platform
- Cloud computing resources, grants, and datasets
 - Research Computing Grants
 - Cloud as a Home for Collaborative Research

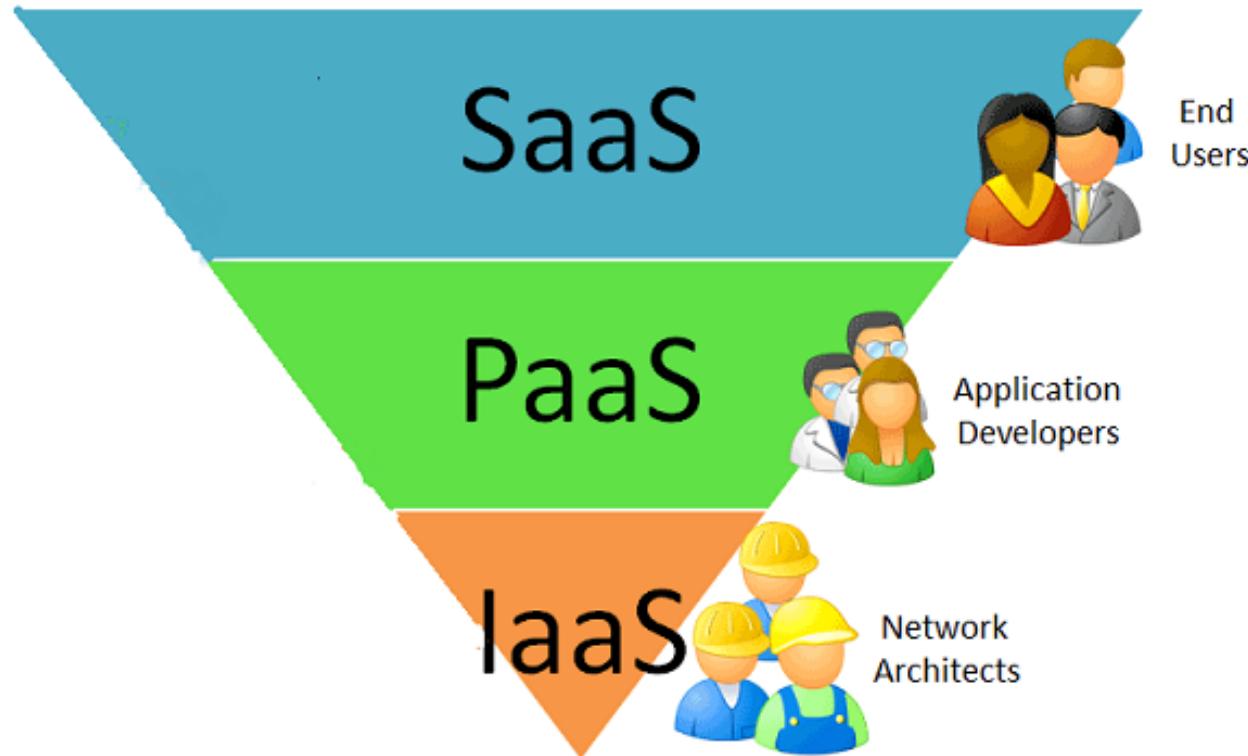


Cloud?

- The National Institute of Standards and Technology describes cloud as “a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.”



Service Models

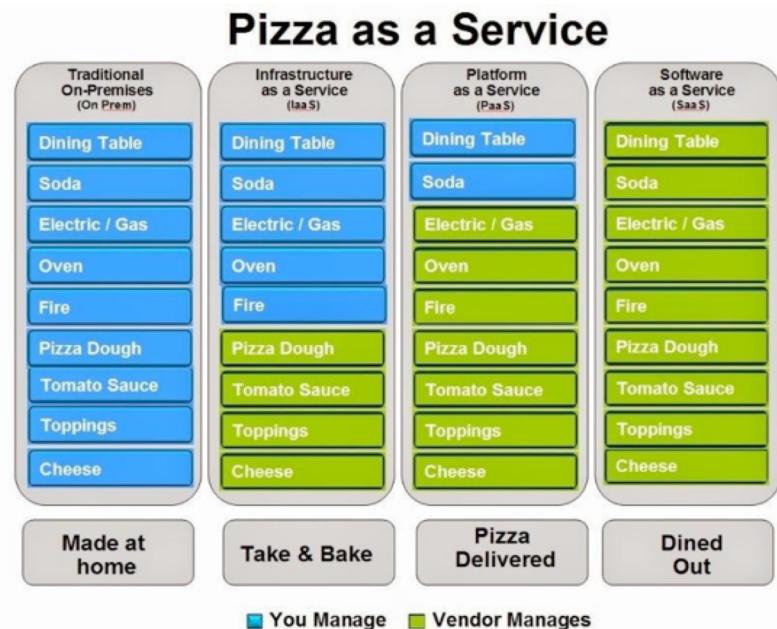
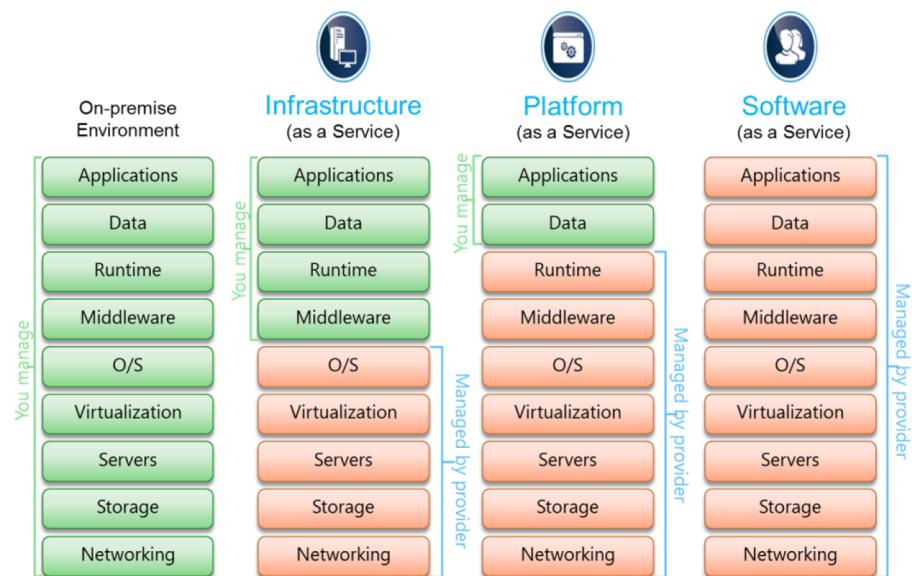


Source: Samisa.org



THE JACKSON LABORATORY

Service Models



Source: <https://www.simple-talk.com/cloud/cloud-development/a-comprehensive-introduction-to-cloud-computing/>



Cloud in Genomics and Bioinformatics

Table 1 Cloud resources in bioinformatics

Resource	Description & availability
Data as a Service (DaaS):	
AWS Public Datasets	Cloud-based archives of GenBank, Ensembl, 1000 Genomes, Model Organism Encyclopedia of DNA Elements, Unigene, Influenza Virus, etc.; http://aws.amazon.com/publicdatasets
Software as a Service (SaaS):	
BGI Cloud (unpublished)	Cloud-based implementations of various genomic analysis applications; http://cloud.genomics.cn
CloudAligner [16]	Fast and full-featured MapReduce-based tool for sequence mapping; http://cloudaligner.sourceforge.net
CloudBLAST [19]	A cloud-based implementation of NCBI BLAST; http://ammatsun.acis.ufl.edu/amwiki/index.php/CloudBLAST_Project
CloudBurst [17]	Highly sensitive short read mapping with MapReduce; http://cloudburst-bio.sourceforge.net
Contrail (unpublished)	Cloud-based <i>de novo</i> assembly of large genomes; http://contrail-bio.sourceforge.net
Crossbow [18]	Read Mapping and SNP calling using cloud computing; http://bowtie-bio.sf.net/crossbow
EasyGenomics (unpublished)	Cloud-based NGS pipelines for whole genome resequencing, exome resequencing, RNA-Seq, small RNA and de novo assembly; http://www.easygenomics.org
eCEO [26]	Cloud-based identification of large-scale epistatic interactions in genome-wide association study (GWAS); http://www.comp.nus.edu.sg/~wangzk/eCEO.html
FX [20]	RNA-Seq analysis tool; http://fx.gmi.ac.kr
Gaea (unpublished)	Cloud-based genome re-sequencing assembly; http://bgiamericas.com/data-analysis/cloud-computing
Hecate (unpublished)	Cloud-based <i>de novo</i> assembly; http://bgiamericas.com/data-analysis/cloud-computing
Jnomics (unpublished)	Cloud-scale sequence analysis suite based on Apache Hadoop; http://sourceforge.net/apps/mediawiki/jnomics
Myrna [21]	Differential gene expression tool for RNA-Seq; http://bowtie-bio.sourceforge.net/myrna
PeakRanger [24]	Cloud-enabled peak caller for ChIP-seq data; http://www.modencode.org/software/ranger
RSD [23]	Reciprocal smallest distance algorithm for ortholog detection using Amazon's Elastic Computing Cloud; http://roundup.hms.harvard.edu
VAT [25]	Variant annotation tool to functionally annotate variants from multiple personal genomes at the transcript level; http://vat.gersteinlab.org
YunBe [22]	Pathway-based or gene set analysis of expression data; http://tinyurl.com/yunbedownload
Platform as a Service (PaaS):	
Eoulsan [27]	Cloud-based platform for high throughput sequencing analyses; http://transcriptome.ens.fr/eoulsan
Galaxy Cloud [28,29]	Cloud-scale Galaxy for large-scale data analysis; http://galaxy.psu.edu
Infrastructure as a Service (IaaS):	
Cloud BioLinux [30]	A publicly accessible virtual machine for high performance bioinformatics computing using cloud platforms; http://cloudbiolinux.org
CloVR [31]	A portable virtual machine for automated sequence analysis using cloud computing; http://clovr.org

Source:Dai et al. Biology direct 2012



THE JACKSON LABORATORY

True or False



- It is Only for Tech Companies

False



- Cloud costs jobs

False



- Big data is not a big deal

False



- It is always cheaper to run in the cloud

False



- Everything can be automated in the cloud, so there's no need for support

False



- Cloud technology is still in its infancy

False



- Data is not as secure in the cloud

False



They are the myths of the cloud!!



© D.Fletcher for CloudTweaks.com

What do I need to maintain my galaxy?





Source:https://www.comicbookmovie.com/guardians_of_the_galaxy/guardians-of-the-galaxy-ranks-as-deadliest-film-ever-a145699



Source:http://izismile.com/2009/08/03/compilation_of_funny_pics_for_the_sysadmin_day_126_pics.html

Why Galaxy?

- Bioinformatics course
 - Computationally focused course?
 - Biology focused course?



Versions Available

- Standalone version
- Docker version
- Public Main Version
- Cloudman Version



Best tools for your gaurdians

- AWS price calculator
- Instances chart
- Usage Alarms



Grants

- How?

