



# Cloud Computing

*What it is, how we use it, and why.*

**Charles Jahnke**  
**Systems Administrator**  
**BUMC Research Computing (LinGA)**



**All of that and more...**

# NIST Definition

1. On-demand self-service
2. Broad network access
3. Resource pooling
4. Rapid elasticity
5. Measured service

**NIST**  
National Institute of  
Standards and Technology  
U.S. Department of Commerce

Special Publication 800-145

---

## **The NIST Definition of Cloud Computing**

---

**Recommendations of the National Institute  
of Standards and Technology**

---

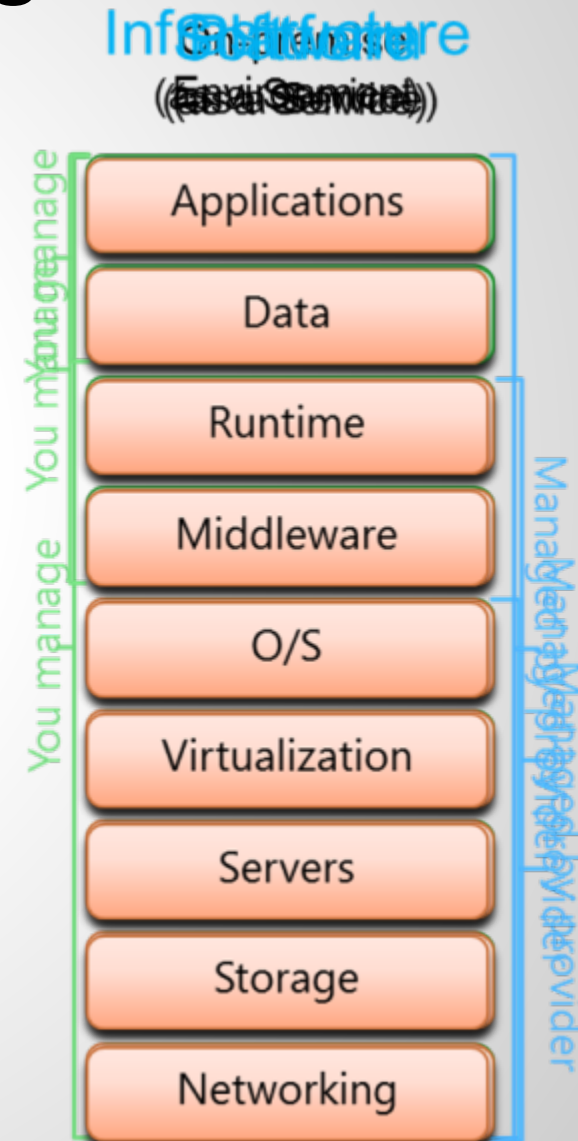
Peter Mell  
Timothy Grance

---

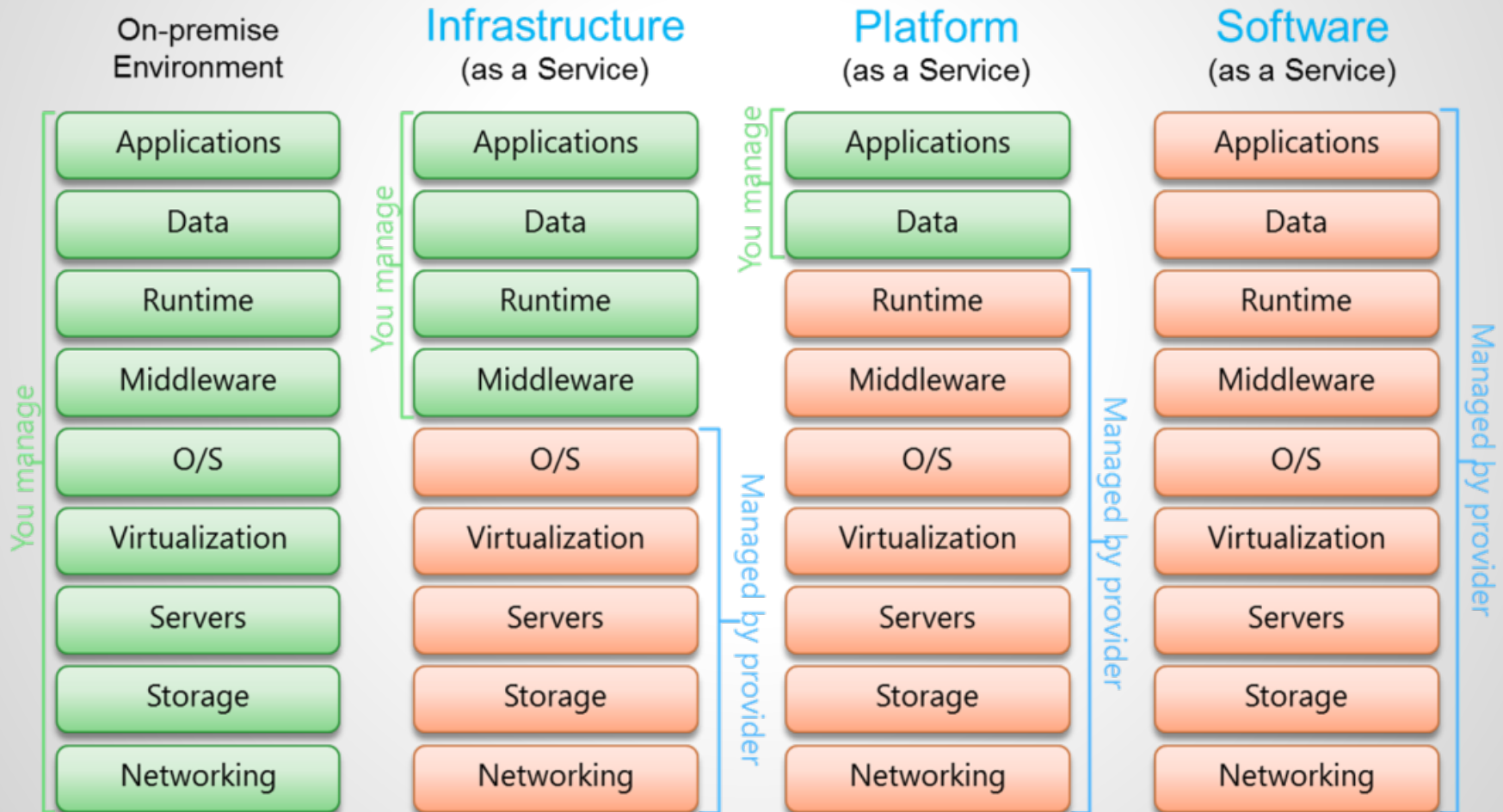
# **Cloud Service Models**

# Flavors and Service Models

- Infrastructure as a Service (IaaS)
  - Infrastructure components
  - Generally: \$/component usage
  - e.g. Amazon Web Services, Rackspace
- Platform as a Service (PaaS)
  - Application backend
  - Generally: \$/data
  - e.g. Google App Engine, MS Azure databases
- Software as a Service (SaaS)
  - Application based
  - Generally: \$/user
  - e.g. Google GMail, ServiceNow



# Flavors and Service Models



# Using the Cloud

*Practical Examples of IaaS, PaaS, and SaaS*

# Infrastructure as a Service (IaaS)

## Infrastructure components

- Compute, storage, network, etc.
- Pay for usage of each component
- Building blocks provide flexibility
- Often lower cost

**Example:**





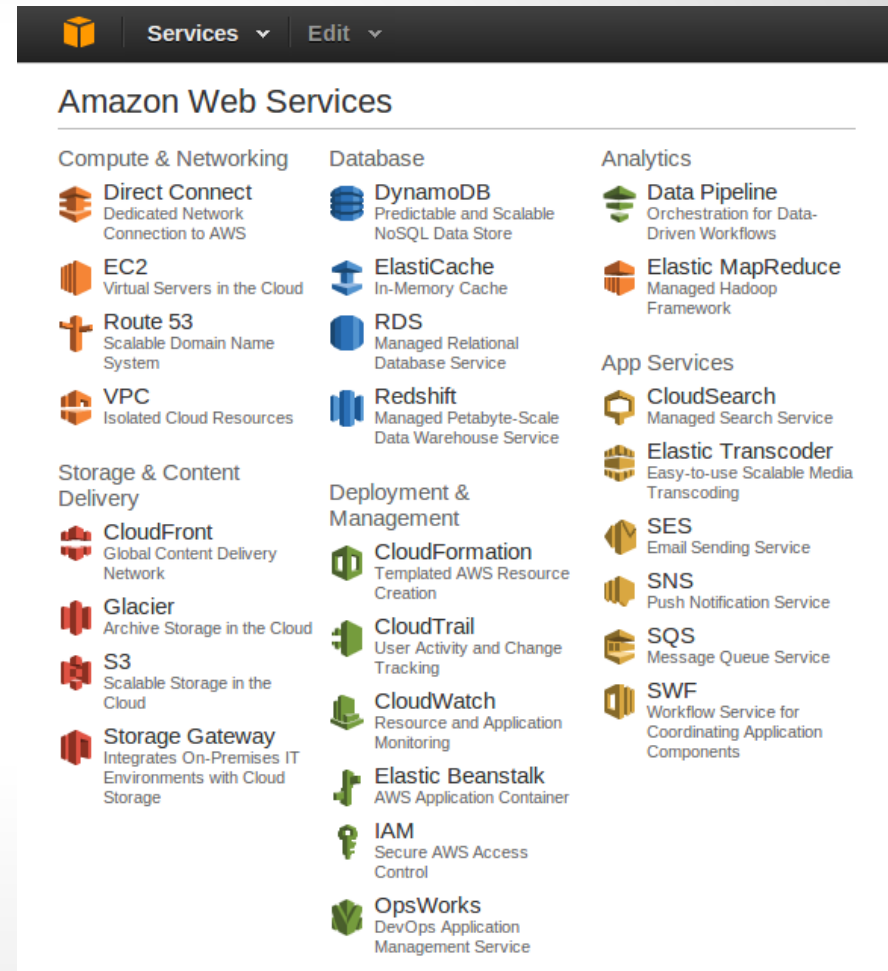
# IaaS - Amazon Web Services

Amazon Provides

- Components

You Provide

- Component Config.
- Middleware
- Runtimes
- Data
- Applications
- **Time and Experience**





Services ▾

Edit ▾

C. Jahnke ▾

Global ▾

Help ▾

## Amazon Web Services

### Compute & Networking



**Direct Connect**  
Dedicated Network Connection to AWS



**EC2**  
Virtual Servers in the Cloud



**Elastic MapReduce**  
Managed Hadoop Framework



**Route 53**  
Scalable Domain Name System



**VPC**  
Isolated Cloud Resources

### Storage & Content Delivery



**CloudFront**  
Global Content Delivery Network



**Glacier**  
Archive Storage in the Cloud



**S3**  
Scalable Storage in the Cloud



**Storage Gateway**  
Integrates On-Premises IT Environments with Cloud Storage

### Database



**DynamoDB**  
Predictable and Scalable NoSQL Data Store



**ElastiCache**  
In-Memory Cache



**RDS**  
Managed Relational Database Service



**Redshift**  
Managed Petabyte-Scale Data Warehouse Service

### Deployment & Management



**CloudFormation**  
Templated AWS Resource Creation



**CloudWatch**  
Resource and Application Monitoring



**Data Pipeline**  
Orchestration for Data-Driven Workflows



**Elastic Beanstalk**  
AWS Application Container



**IAM**  
Secure AWS Access Control



**OpsWorks**  
DevOps Application Management Service

### App Services



**CloudSearch**  
Managed Search Service



**Elastic Transcoder**  
Easy-to-use Scalable Media Transcoding



**SES**  
Email Sending Service



**SNS**  
Push Notification Service



**SQS**  
Message Queue Service



**SWF**  
Workflow Service for Coordinating Application Components

## Additional Resources

### Getting Started

See our documentation to get started and learn more about how to use our services.

### Trusted Advisor

Best practice recommendations to save money, improve fault tolerance, increase performance, and close security gaps.

## Service Health

All services operating normally.

Updated: Nov 08 2013 18:13:00 GMT-0500

[Service Health Dashboard](#)

## Set Start Page

Console Home ▾



**AWS Marketplace**

Find & buy software, launch with 1-Click and pay by the hour.



Services v

Edit v

C. Jahnke v

N. Virginia v

Help v

EC2 Dashboard

Events

Tags

INSTANCES

Instances

Spot Requests

Reserved Instances

IMAGES

AMIs

Bundle Tasks

ELASTIC BLOCK STORE

Volumes

Snapshots

NETWORK &amp; SECURITY

Security Groups

Elastic IPs

Placement Groups

Load Balancers

Key Pairs

Network Interfaces

Launch Instance

Connect

Actions v



Filter: All instances v

All instance types v

Search Instances



No Instances

You do not have any running instances in this region.

First time using EC2? Check out the [Getting Started Guide](#).

Click the Launch Instance button to start your own server.

Launch Instance

Select an instance above





Services ▾

Edit ▾

C. Jahnke ▾

N. Virginia ▾

Help ▾

**1. Choose AMI**

2. Choose Instance Type

3. Configure Instance

4. Add Storage

5. Tag Instance

6. Configure Security Group

7. Review

## Step 1: Choose an Amazon Machine Image (AMI)

[Cancel and Exit](#)

An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. You can select an AMI provided by AWS, our user community, or the AWS Marketplace; or you can select one of your own AMIs.

### Quick Start

My AMIs

AWS Marketplace

Community AMIs

☐ Free tier only ⓘ**Amazon Linux**

Free tier eligible

**Amazon Linux AMI 2013.09** - ami-35792c5c (64-bit) / ami-51792c38 (32-bit)

The Amazon Linux AMI is an EBS-backed, PV-GRUB image. It includes Linux 3.4, AWS tools, and repository access to multiple versions of MySQL, PostgreSQL, Python, Ruby, and Tomcat.

Root device type: ebs    Virtualization type: paravirtual

**Select**☒ 64-bit ☐ 32-bit**Red Hat**

Free tier eligible

**Red Hat Enterprise Linux 6.4** - ami-a25415cb (64-bit) / ami-7e175617 (32-bit)

Red Hat Enterprise Linux version 6.4, EBS-boot.

Root device type: ebs    Virtualization type: paravirtual

**Select**☒ 64-bit ☐ 32-bit**SUSE Linux**

Free tier eligible

**SUSE Linux Enterprise Server 11** - ami-e8084981 (64-bit) / ami-b60948df (32-bit)

SUSE Linux Enterprise Server 11 Service Pack 3 basic install, EBS boot with Amazon EC2 AMI Tools preinstalled; Apache 2.2, MySQL 5.5, PHP 5.3, and Ruby 1.8.7 available

Root device type: ebs    Virtualization type: paravirtual

**Select**☒ 64-bit ☐ 32-bit**Ubuntu Server 12.04.3 LTS** - ami-a73264ce (64-bit) / ami-**Select**



Services ▾

Edit ▾

C. Jahnke ▾

N. Virginia ▾

Help ▾

[1. Choose AMI](#)[2. Choose Instance Type](#)[3. Configure Instance](#)[4. Add Storage](#)[5. Tag Instance](#)[6. Configure Security Group](#)[7. Review](#)

## Step 2: Choose an Instance Type

Amazon EC2 provides a wide selection of instance types optimized to fit different use cases. Instances are virtual servers that can run applications. They have varying combinations of CPU, memory, storage, and networking capacity, and give you the flexibility to choose the appropriate mix of resources for your applications. [Learn more](#) about instance types and how they can meet your computing needs.

Currently selected: t1.micro (up to 2 ECUs, 1 vCPUs, 0.613 GiB memory, EBS only)

All instance types

**Micro instances**

Free tier eligible

General purpose

Memory optimized

Storage optimized

Compute optimized

### Micro instances

Micro instances are a low-cost instance option, providing a small amount of CPU resources. They are suited for lower throughput applications, and websites that require additional compute cycles periodically, but are not appropriate for applications that require sustained CPU performance. Popular uses for micro instances include low traffic websites or blogs, small administrative applications, bastion hosts, and free trials to explore EC2 functionality.

Size	ECUs <small>i</small>	vCPUs <small>i</small>	Memory (GiB)	Instance Storage (GiB) <small>i</small>	EBS-Optimized Available <small>i</small>	Network Performance
t1.micro	up to 2	1	0.613	EBS only	-	Very Low



Micro instances are eligible for the AWS free usage tier. For the first 12 months following your AWS sign-up date, you get up to 750 hours of micro instances each month. When your free usage tier expires or if your usage exceeds the free tier restrictions, you pay standard, pay-as-you-go service rates.

[Learn more](#) about free usage tier eligibility and restrictions

[Cancel](#)[Previous](#)[Review and Launch](#)[Next: Configure Instance Details](#)

# Configure Instance Details

Can get very detailed if desired

- Number of instances
- Networks
- Zones (Geographic location, redundancy)
- Monitoring
- Storage
- Tags
- Security and Security Groups





Services ▾

Edit ▾

C. Jahnke ▾

N. Virginia ▾

Help ▾

[1. Choose AMI](#)[2. Choose Instance Type](#)[3. Configure Instance](#)[4. Add Storage](#)[5. Tag Instance](#)[6. Configure Security Group](#)[7. Review](#)

## Step 2: Choose an Instance Type

Amazon EC2 provides a wide selection of instance types optimized to fit different use cases. Instances are virtual servers that can run applications. They have varying combinations of CPU, memory, storage, and networking capacity, and give you the flexibility to choose the appropriate mix of resources for your applications. [Learn more](#) about instance types and how they can meet your computing needs.

Currently selected: t1.micro (up to 2 ECUs, 1 vCPUs, 0.613 GiB memory, EBS only)

All instance types

**Micro instances**

Free tier eligible

General purpose

Memory optimized

Storage optimized

Compute optimized

### Micro instances

Micro instances are a low-cost instance option, providing a small amount of CPU resources. They are suited for lower throughput applications, and websites that require additional compute cycles periodically, but are not appropriate for applications that require sustained CPU performance. Popular uses for micro instances include low traffic websites or blogs, small administrative applications, bastion hosts, and free trials to explore EC2 functionality.

Size	ECUs <small>i</small>	vCPUs <small>i</small>	Memory (GiB)	Instance Storage (GiB) <small>i</small>	EBS-Optimized Available <small>i</small>	Network Performance
t1.micro	up to 2	1	0.613	EBS only	-	Very Low



Micro instances are eligible for the AWS free usage tier. For the first 12 months following your AWS sign-up date, you get up to 750 hours of micro instances each month. When your free usage tier expires or if your usage exceeds the free tier restrictions, you pay standard, pay-as-you-go service rates.

[Learn more](#) about free usage tier eligibility and restrictions

[Cancel](#)[Previous](#)[Review and Launch](#)[Next: Configure Instance Details](#)



## Launch Status



### Your instance is now launching

The following instance launch has been initiated: [i-2212b25f](#) [View launch log](#)



### Get notified of estimated charges

[Create billing alerts](#) to get an email notification when estimated charges on your AWS bill exceed \$0.0 (in other words, when you have exceeded the free usage tier).

## How to connect to your instance

Your instance is launching, and it may take a few minutes until it is in the **running** state, when it will be ready for you to use. Usage hours on your new instance will start immediately and continue to accrue until you stop or terminate your instance.

Click **View Instances** to monitor your instance's status. Once your instance is in the **running** state, you can **connect** to it from the Instances screen. [Find out](#) how to connect to your instance.

### ▼ Here are some helpful resources to get you started

- [How to connect to your Linux instance](#)
- [Amazon EC2: User Guide](#)
- [Learn about AWS Free Usage Tier](#)
- [Amazon EC2: Discussion Forum](#)

While your instances are launching you can also

[Create status check alarms](#) to be notified when these instances fail status checks. (Additional charges may apply)

[Create and attach additional EBS volumes](#) (Additional charges may apply)

[Manage security groups](#)





Services ▾

Edit ▾

C. Jahnke ▾

N. Virginia ▾

Help ▾

EC2 Dashboard

Events

Tags

INSTANCES

Instances

Spot Requests

Reserved Instances

IMAGES

AMIs

Bundle Tasks

ELASTIC BLOCK STORE

Volumes

Snapshots

NETWORK &amp; SECURITY

Security Groups

Elastic IPs

Placement Groups

Load Balancers

Key Pairs

Network Interfaces

Launch Instance

Connect

Actions ▾



Filter: All instances ▾

All instance types ▾

Search Instances



&lt;&lt; &lt; 1 to 1 of 1 Instances &gt; &gt;&gt;

<input type="checkbox"/>	Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks
<input type="checkbox"/>	GSI Test	i-2212b25f	t1.micro	us-east-1a	running	2/2 check...

Instance: i-2212b25f (GSI Test)

Public DNS: ec2-54-225-33-165.compute-1.amazonaws.com



Description

Status Checks

Monitoring

Tags

Instance ID i-2212b25f

Instance state running

Instance type t1.micro

Private DNS ip-10-178-33-154.ec2.internal

Private IPs 10.178.33.154

Secondary private IPs -

VPC ID -

Public DNS ec2-54-225-33-165.compute-1.amazonaws.com

Public IP 54.225.33.165

Elastic IP -

Availability zone us-east-1a

Security groups launch-wizard-1.  
[view rules](#)Scheduled events No scheduled  
[events](#)

AMI ID ubuntu-precise-12.04-amd64-

File Edit View Search Terminal Help

```
cjahnke@BUMC-PC905908:~/Downloads$ ssh -i GSI.pem ubuntu@ec2-54-225-33-165.compute-1.amazonaws.com
```

Welcome to Ubuntu 12.04.3 LTS (GNU/Linux 3.2.0-54 virtual x86\_64)

\* Documentation: <https://help.ubuntu.com/>

System information as of Mon Nov 18 19:23:18 UTC 2013

System load:	0.0	Processes:	58
Usage of /:	11.1% of 7.87GB	Users logged in:	0
Memory usage:	6%	IP address for eth0:	10.178.33.154
Swap usage:	0%		

Graph this data and manage this system at <https://landscape.canonical.com/>

Get cloud support with Ubuntu Advantage Cloud Guest:  
<http://www.ubuntu.com/business/services/cloud>

Use Juju to deploy your cloud instances and workloads:  
<https://juju.ubuntu.com/#cloud-precise>

0 packages can be updated.  
0 updates are security updates.


The programs included with the Ubuntu system are free software;  
the exact distribution terms for each program are described in the  
individual files in /usr/share/doc/\*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by  
applicable law.

To run a command as administrator (user "root"), use "sudo <command>".  
See "man sudo\_root" for details.





```
ubuntu@ip-10-178-33-154:~$
```

# But that's just one component...





 **Services** ▾ **Edit** ▾

## Amazon Web Services





### Compute & Networking

-  **Direct Connect**  
Dedicated Network Connection to AWS
-  **EC2**  
Virtual Servers in the Cloud
-  **Route 53**  
Scalable Domain Name System
-  **VPC**  
Isolated Cloud Resources







### Storage & Content Delivery

-  **CloudFront**  
Global Content Delivery Network
-  **Glacier**  
Archive Storage in the Cloud
-  **S3**  
Scalable Storage in the Cloud
-  **Storage Gateway**  
Integrates On-Premises IT Environments with Cloud Storage



### Database

-  **DynamoDB**  
Predictable and Scalable NoSQL Data Store
-  **ElastiCache**  
In-Memory Cache
-  **RDS**  
Managed Relational Database Service
-  **Redshift**  
Managed Petabyte-Scale Data Warehouse Service







### Deployment & Management

-  **CloudFormation**  
Templated AWS Resource Creation
-  **CloudTrail**  
User Activity and Change Tracking
-  **CloudWatch**  
Resource and Application Monitoring
-  **Elastic Beanstalk**  
AWS Application Container
-  **IAM**  
Secure AWS Access Control
-  **OpsWorks**  
DevOps Application Management Service

### Analytics

-  **Data Pipeline**  
Orchestration for Data-Driven Workflows
-  **Elastic MapReduce**  
Managed Hadoop Framework

### App Services

-  **CloudSearch**  
Managed Search Service
-  **Elastic Transcoder**  
Easy-to-use Scalable Media Transcoding
-  **SES**  
Email Sending Service
-  **SNS**  
Push Notification Service
-  **SQS**  
Message Queue Service
-  **SWF**  
Workflow Service for Coordinating Application Components

# Platform as a Service (PaaS)

“Platform” or Application Backend

- Often a database or management toolset
- Pay for usage (time) of platform
- Consumer builds on top of platform
- Examples:
  - Google App Engine, Microsoft Azure Databases

**Example:**



**StarCluster**

# PaaS - MIT StarCluster

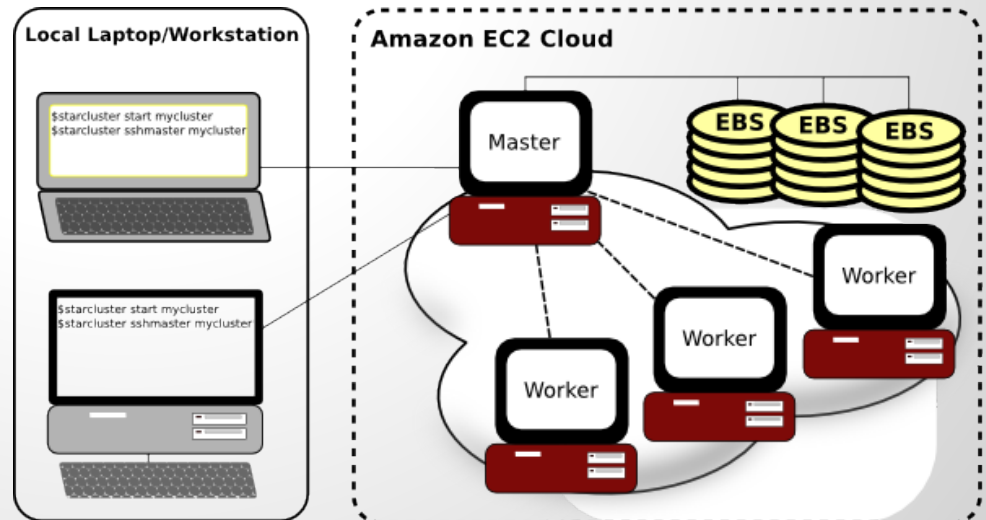
- StarCluster Provides

- Builds on AWS “infrastructure” services
- Cluster configuration
- Middleware and runtime environment
- Tools for management

} Platform

- You provide

- Management
- Applications
- Data



## StarCluster



Home



Biochem



Genetics



Orf



Biogene



Hydro



Molsim



Cluster



Hpc

[star](#) > [cluster](#) > [Home](#)[Home](#)[@MIT](#)[Features](#)[Download](#)[Download AMIs](#)[Documentation](#)[Mailing List](#)[Bugs and Feature Requests](#)[Wiki](#)[Donate](#)

# StarCluster

StarCluster is an open source cluster-computing toolkit for Amazon's Elastic Compute Cloud (EC2) released under the [LGPL license](#).

StarCluster has been designed to automate and simplify the process of building, configuring, and managing clusters of virtual machines on [Amazon's EC2 cloud](#). StarCluster allows anyone to easily create a cluster computing environment in the cloud suited for distributed and parallel computing applications and systems.

## Overview

*Introduction to StarCluster.*

## Features

*Current and planned features*

## Install instructions

*How to install StarCluster.*

## Quick-Start Tutorial

*A Quick-Start tutorial for StarCluster newcomers.*

## Support

*Reporting bugs, issues, questions, feature requests, etc.*

## User Manual

*Complete user manual for StarCluster*

## Guides

*Task-oriented guides*

## Plugins

*Plugin-specific documentation*

## FAQ

*Frequently Asked Questions*

## Contribute

*Taking part in StarCluster's development.*

## Quick-Start Screencast

Getting started with StarCluster:

StarCluster 0.91 Demo

```
(starcluster-demo) demouser@demohost ~$ starcluster start -h | less
(starcluster-demo) demouser@demohost ~$ cpconf
(starcluster-demo) demouser@demohost ~$ starcluster start myfirstcluster
```

# PaaS - Install StarCluster Client

## Manual Install

```
$ wget https://pypi.python.org/packages/source/S/StarCluster/StarCluster-0.94.2.tar.gz  
$ tar xvzf StarCluster-XXX.tar.gz  
$ cd StarCluster-XXX  
$ python distribute_setup.py  
$ python setup.py install
```

## Python Easy Install

```
$ easy_install StarCluster
```

# PaaS - Configure StarCluster

## Write Template

```
$ starcluster
StarCluster - (http://star.mit.edu/cluster) (v. 0.94.2)
Software Tools for Academics and Researchers (STAR)
Please submit bug reports to starcluster@mit.edu

!!! ERROR - config file /home/<user>/.starcluster/config does not exist

Options:
-----
[1] Show the StarCluster config template
[2] Write config template to /home/<user>/.starcluster/config
[q] Quit

Please enter your selection: 2
```

## Configure Cluster Template

```
$ vim /home/<user>/.starcluster/config
```

- Add (minimum) AWS account keys, security keys, cluster names



# PaaS - Configure StarCluster (Cont.)

## Generate AWS Keys

```
$ starcluster createkey mykey -o ~/.ssh/starcluster.rsa

StarCluster - (http://star.mit.edu/cluster) (v. 0.94.2)
Software Tools for Academics and Researchers (STAR)
Please submit bug reports to starcluster@mit.edu

>>> Successfully created keypair: mykey
>>> fingerprint: j6:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:e2
>>> contents:

-----BEGIN RSA PRIVATE KEY-----
<key snipped>
-----END RSA PRIVATE KEY-----

>>> keypair written to /home/<user>/.ssh/starcluster.rsa
```

## Add AWS Keys

```
$ vim /home/<user>/.starcluster/config
```

# PaaS - Start StarCluster

```
$ starcluster start -c smallcluster gsiccluster
```

```
StarCluster - (http://star.mit.edu/cluster) (v. 0.94.2)  
Software Tools for Academics and Researchers (STAR)  
Please submit bug reports to starcluster@mit.edu
```

```
>>> Validating cluster template settings...  
>>> Cluster template settings are valid  
>>> Starting cluster...  
>>> Launching a 2-node cluster...  
>>> Creating security group @sc-gsiccluster...  
Reservation:r-89aa46f1  
>>> Waiting for instances to propagate...  
>>> Waiting for all nodes to be in a 'running' state...  
2/2 ||||| 100%  
>>> Waiting for SSH to come up on all nodes...  
2/2 ||||| 100%  
>>> Waiting for cluster to come up took 1.561 mins  
>>> The master node is ec2-107-22-121-5.compute-1.amazonaws.com  
>>> Configuring cluster...  
>>> Running plugin starcluster.clustersetup.DefaultClusterSetup  
>>> Configuring hostnames...  
2/2 ||||| 100%  
>>> Creating cluster user: sgeadmin (uid: 1001, gid: 1001)  
2/2 ||||| 100%  
>>> Configuring scratch space for user(s): sgeadmin  
2/2 ||||| 100%  
>>> Configuring /etc/hosts on each node  
2/2 ||||| 100%  
>>> Starting NFS server on master
```

. . .



Services v

Edit v

C. Jahnke v

N. Virginia v

Help v

Launch Instance

Connect

Actions v

Filter: Running instances v

All instance types v

Search Instances



1 to 3 of 3 Instances



<input type="checkbox"/>	Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS
<input type="checkbox"/>	GSI Test	i-2212b25f	t1.micro	us-east-1a	running	2/2 check...	None	ec2-54-225-33-11
<input checked="" type="checkbox"/>	master	i-1a0e8561	m1.small	us-east-1c	running	2/2 check...	None	ec2-107-22-121-5
<input type="checkbox"/>	node001	i-180e8563	m1.small	us-east-1c	running	2/2 check...	None	ec2-107-21-76-2

Instance: i-1a0e8561 (master)

Public DNS: ec2-107-22-121-5.compute-1.amazonaws.com

Description

Status Checks

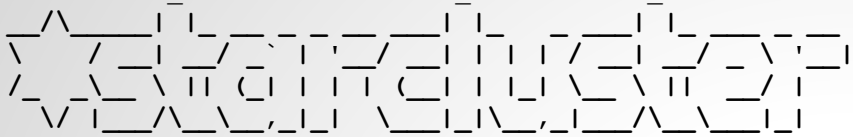
Monitoring

Tags

Instance ID	i-1a0e8561	Public DNS	ec2-107-22-121-5.compute-1.amazonaws.com
Instance state	running	Public IP	107.22.121.5
Instance type	m1.small	Elastic IP	-
Private DNS	ip-10-169-2-55.ec2.internal	Availability zone	us-east-1c
Private IPs	10.169.2.55	Security groups	@sc-gscluster. <a href="#">view rules</a>
Secondary private IPs	-	Scheduled events	<a href="#">No scheduled events</a>
VPC ID	-	AMI ID	starcluster-base-ubuntu-12.04-

## **\$ starcluster sshmaster gsiccluster**

StarCluster - (<http://star.mit.edu/cluster>) (v. 0.94.2)  
Software Tools for Academics and Researchers (STAR)  
Please submit bug reports to [starcluster@mit.edu](mailto:starcluster@mit.edu)



StarCluster Ubuntu 12.04 AMI  
Software Tools for Academics and Researchers (STAR)  
Homepage: <http://star.mit.edu/cluster>  
Documentation: <http://star.mit.edu/cluster/docs/latest>  
Code: <https://github.com/jtriley/StarCluster>  
Mailing list: [starcluster@mit.edu](mailto:starcluster@mit.edu)

### This AMI Contains:

- \* Open Grid Scheduler (OGS - formerly SGE) queuing system
- \* Condor workload management system
- \* OpenMPI compiled with Open Grid Scheduler support
- \* OpenBLAS- Highly optimized Basic Linear Algebra Routines
- \* NumPy/SciPy linked against OpenBlas
- \* IPython 0.13 with parallel support
- \* and more! (use 'dpkg -l' to show all installed packages)

### Open Grid Scheduler/Condor cheat sheet:

- \* `qstat/condor_q` - show status of batch jobs
- \* `qhost/condor_status` - show status of hosts, queues, and jobs
- \* `qsub/condor_submit` - submit batch jobs (e.g. `qsub -cwd ./job.sh`)
- \* `qdel/condor_rm` - delete batch jobs (e.g. `qdel 7`)
- \* `qconf` - configure Open Grid Scheduler system

### Current System Stats:

System load:	0.0	Processes:	81
Usage of /:	27.5% of 9.84GB	Users logged in:	0
Memory usage:	5%	IP address for eth0:	10.169.2.55
Swap usage:	0%		

**root@master:~#**

# Software as a Service (SaaS)

## Application or Software Suite

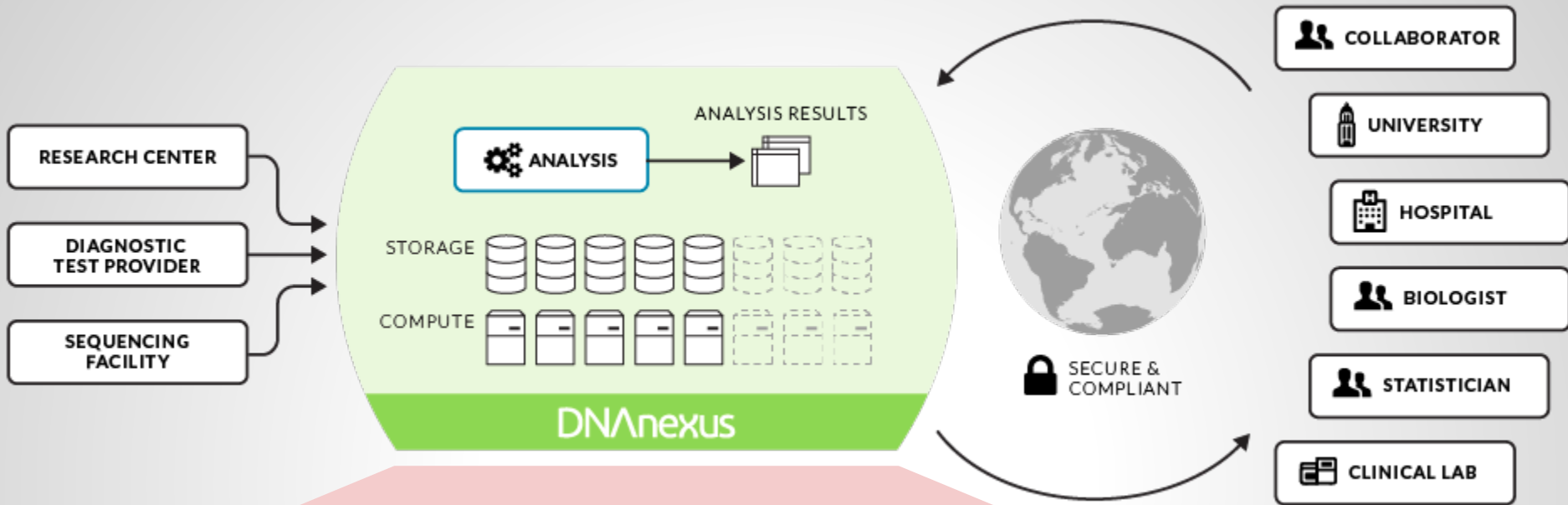
- Graphical Web Interface
- Often pay for each user
- Common Examples:
  - Google GMail
  - ServiceNow

**Example:**



# SaaS - DNA Nexus

- DNA Nexus Provides
  - Compute infrastructure
  - Virtualization, OS and cluster management
  - Standard and Public datasets
    - Study datasets if use their sequencing core.
  - Applications, tools and software
  - Graphical interface
- You provide
  - Study data
  - Application options and order



- Annotation
- Assembly
- Debugging
- Export
- Import
- Mappings Manipulation
- Read Manipulation
- Read Mappings
- Reports
- RNA-Seq
- Statistics
- Variation Calling

Catagories  
(100+ Applications)

# SaaS - DNA Nexus Command Line

## Quick Install

```
$ wget https://wiki.dnanexus.com/images/files/dx-toolkit-v0.76.0-ubuntu-13.10-amd64.tar.gz
$ tar -xzf dx-toolkit*.tar.gz
$ source dx-toolkit/environment
```

## Login to DNA Nexus

```
$ dx login

Acquiring credentials from https://auth.dnanexus.com
Username: <your username>
Password: <your password>

No projects to choose from. You can create one with the command "dx new
project". To pick from projects for which you only have VIEW
permissions, use "dx select --level VIEW" or "dx select --public".
```

## Get to work

```
$ dx select --public

Available public projects:
0) Reference Genomes (VIEW)
1) Demo Data (VIEW)
2) Developer Applets (VIEW)

Pick a numbered choice: 0
Setting current project to: Reference Genomes
```



# SaaS - DNA Nexus Graphic Interface

- GUI Available
- Drag and Drop
- Data Mgmt.
- Application Mgmt.
- Workspace Sharing

The screenshot displays the DNA Nexus Graphic Interface for a workflow titled "HGSC\_Mercury\_v3.2.1\_MendelianExome\_Public". The interface is organized into three main columns: Inputs, App, and Outputs. The workflow consists of several steps, each represented by a card with a visual icon and a text label. The steps are connected by arrows indicating the flow of data. The first step is "BWA - Fix Mate Info" (unconfigured), which takes inputs of "Read1 (b2, gz, bxt, fq)" and "Read2 (b2, gz, bxt, fq)". The second step is "BAM Analyzer Ne..." (configured), which takes the output of the first step and produces outputs like "BAM Map Stats file" and "BAM Analysis Info XML file". The third step is "GATK-Lite Indel R..." (configured), which takes the output of the second step and produces "Realigned sorted mappings". The fourth step is "GATK-Lite Base Q..." (configured), which takes the output of the third step and produces "Recalibrated sorted mappings". The fifth step is "bamUtil BAM Valid..." (configured), which takes the output of the fourth step and produces "Validation report file", "PASS file", and "FAIL file". The sixth step is "Parallel Atlas SNP..." (configured), which takes the output of the fifth step and produces "VCF File", "SNP File", and "Indel File". The interface also shows a "Monitor" tab for viewing job progress and a "Readme" button.

(This will get larger)

HGSC\_Mercury\_v3.2.1\_MendelianExome\_Public

1 app  
unconfigured

12 apps  
configured

Readme

## Run Analysis for HGSC\_Mercury\_v3.2.1\_MendelianExome\_Public

View job progress in your project's Monitor tab. Modifications to an existing workflow will not be saved.

### Inputs

### Application

### Outputs

Workflow Step 1:

- Inputs:**
  - \*.txt \*.fq. \*.fastq\* Read1 (.bz2, .gz, .txt, .fq)
  - \*.txt \*.fq. \*.fastq\* Read2 (.bz2, .gz, .txt, .fq)
  - hg19-BCM-mod\_indexed\_v0.6.2.tar.gz BWA index (.tar.gz)
- Application:** BWA - Fix Mate Info - ... (set inputs)
- Outputs:**
  - Output BAM file (coord sorted)
  - Output bam Index file
  - Mark duplicates metrics file

Workflow Step 2:

- Inputs:**
  - via BWA - Fix Mate info - Mark dupes - Sort - Adjust Quality (Output BAM file (coord sorted))
- Application:** BAM Analyzer New (a... (configured)
- Outputs:**
  - BAM Map Stats file
  - BAM Analysis Info XML file
  - FR\_InsertSizeDist.png
  - BaseQualPerPosition.png

Workflow Step 3:

- Inputs:**
  - via BWA - Fix Mate info - Mark dupes - Sort - Adjust Quality (Output BAM file (coord sorted))
  - hg19-BCM-mod.fa.gz Genome
  - 2 inputs Known indels [array]
- Application:** GATK-Lite Indel Real... (configured)
- Outputs:**
  - \*.bam Realigned sorted mappings

Workflow Step 4:

- Inputs:**
  - via GATK-Lite Indel Realigner (Realigned sorted mappings)
  - hg19-BCM-mod.fa.gz Genome
  - 3 inputs Known variants [array]
- Application:** GATK-Lite Base Quali... (configured)
- Outputs:**
  - \*.bam Recalibrated sorted mappings

Workflow Step 5:

- Inputs:**
  - via GATK-Lite Base Quality Score Recalibrator (Recalibrated sorted mappings)
- Application:** bamUtil BAM Validato... (configured)
- Outputs:**
  - Validation report file
  - PASS file
  - FAIL file

Workflow Step 6:

- Inputs:**
  - via bamUtil BAM Validator v1.0.7 FIX (PASS file)
  - hg19-BCM-mod.fa.gz Reference
- Application:** Parallel Atlas SNP and... (configured)
- Outputs:**
  - VCF File
  - SNP File
  - Indel File
  - SNP Pileup File

# Recap

- Cloud computing provides flexible hosted environment to the public.
- Three Service Models
  - Infrastructure as a Service
  - Platform as a Service
  - Software as a Service
- Easy to get started

# Questions