Cloud computing 101

Fabien Hermenier

Image credit: http://eyepluscamera.files.wordpress.com/
was cloud computing needed?
Then came with affordable PCs

Then we spread out the load for security, performance, manageability

Then we bought tons of servers to support load spikes
Amazon X-mas 2013

426 items sold each second
Where is energy spent?
episode 0
rise of the cloud
Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction.
1. On-demand self-services

Self-provisioning, no human intervention
Postgres Plus Cloud Database Advanced Edition

Database Compatibility for Oracle

24 x 7 premium support included from EnterpriseDB
the Postgres Database Company

LEARN MORE

Featured Products

- Postgres Plus Cloud Database Advanced
  - EnterpriseDB Corporation
  - $0.26 to $20.24/hr for software
- PHP 5.5 - Zend Server Developer Edition
  - Zend Technologies
  - $0.03/hr for software
- SteelCloud
  - SteelCloud STIG Windows 200...
  - $49.00/mo + $0.00 to $0.06/hr for software + Charges for EC2 with Windows

Operating Systems

- Amazon Linux AMI (HVM/64-bit)
  - Amazon Web Services
  - $0.013 to $8.14/hr incl EC2 charges
- CentOS 7 (x86_64) with Updates HVM
  - Centos.org
  - $0.00/hr for software
- Ubuntu
  - Ubuntu Server 14.04 LTS (HVM)
  - Canonical Group L...
  - $0.00/hr for software

Popular Products

- JumpBox
- DBA
- OpenShift
- Jenkins
broad network access

availability over the network
standard mechanisms
3 resource pooling

multi-tenant
virtual or physical resources
on-demand allocation
location independance
reserved instances (yearly based)
on-demand instances (hourly based)
hotspot instances (market based)
Amazon EC2
HotSpot instances

bid over the market price to get the instance
rapid elasticity

fast (de-)allocation of resources
scale to infinity
vertical elasticity

Tiers 1

Tiers 2

Tiers 3
vertical elasticity

Tiers 1

Tiers 2

Tiers 3
horizontal elasticity
horizontal elasticity
5

measured service

metering capabilities
transparent reporting
<table>
<thead>
<tr>
<th>Plan</th>
<th>Size</th>
<th>Hourly Price</th>
<th>Monthly Price</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cloud Sandbox</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labs</td>
<td>M</td>
<td>$0.004</td>
<td></td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>$0.010</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 Core</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 GB RAM</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Steadfast Resources</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>$0.014</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>$0.028</td>
<td></td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>$0.056</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 Core</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 GB RAM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 Core</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4 GB RAM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 Cores</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8 GB RAM</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1 VM/HOST</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>XL 3</td>
<td>$0.110</td>
<td></td>
</tr>
<tr>
<td></td>
<td>XL 4</td>
<td>$0.140</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4 Cores</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16 GB RAM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6 Cores</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>24 GB RAM</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Labs | Get 1 month Free!**

**Launch Now!**

**More Info ▼**
and I will call it cloud computing
If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility... The computer utility could become the basis of a new and important industry.

John McCarthy, 1961
Cluster computing

loosely coupled co-located servers
single tenant
non-interactive workload
rigid jobs

80s
Cloud or not?

- on demand self-services
- broad network access
- resource pooling
- rapid elasticity
- measured service
Cloud or not?

- on demand self-services
- broad network access
- resource pooling
- rapid elasticity
- measured service
grid computing
Ian Foster et al. 2001
Power Grid Analogy
Power grid

- multiple providers
- heterogeneous sources
- multiple clients
- abstract source
- live consumption
- location
- doing * at
- large scale

Computing grid

- virtual organisation
- heterogeneous hw.
- multiple applications
- abstract resources
- batch jobs
- independence
- large scale
Worldwide LHC Computing grid

170 centres to analyse 30 PB / year
Cloud or not?

- on demand self-services
- broad network access
- resource pooling
- rapid elasticity
- measured service
Cloud or not?

- On demand self-services
- Broad network access
- Resource pooling
- Rapid elasticity
- Measured service
Application
Service Provider

remote access to dedicated applications

service oriented

pay as you go
Cloud or not?

- on demand self-services
- broad network access
- resource pooling
- rapid elasticity
- measured service
Cloud or not?

- **on demand self-services**
- **broad network access**
- **resource pooling** *(not real hw resources)*
- **rapid elasticity**
- **measured service**
2002 computers on demand.

Deploy full custom stacks (OS to applications)
to (re)deploy reproducible network experiments

multi-tenant, (limited on purpose) resource pooling,
to (re)deploy reproducible network experiments
to (re)deploy reproducible network experiments
Cloud or not?

- on demand self-services
- broad network access
- resource pooling
- rapid elasticity
- measured service
Cloud or not?

- on demand self-services
- broad network access
- resource pooling
- rapid elasticity
- measured service
Service-oriented Architecture

composable unassociated, loosely coupled units

2001+
exponential grows since 2001
corporate and public services to support its growth
Two pizza rule

If a team can’t be fed by two pizzas then it is too big

- Jeff Bezos (founder/CEO of amazon.com)
800 x tons of API, mini-services devoted to automation, flexibility, on-demand services for public and private use
scalable web services for other websites or client-side applications
SOAP & REST over HTTP
pay as you go
elastic *-oriented services

*data, network or computation
on demand self-services
broad network access
resource pooling
rapid elasticity
measured service
RECAP
I have a dream, it was about Utility Computing

John McCarthy - 1961
web
+ grid computing
+ resources on demand
+ service oriented architectures

cloud computing (2006)
can we talk about cloud computing now?
See?
SaaS
Software as a Service

web access to commercial sw.
“one to many” model
customers don’t handle upgrades
API for integration
### Users.messages

For Users.messages Resource details, see the [resource representation](#) page.

<table>
<thead>
<tr>
<th>Method</th>
<th>HTTP request</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>delete</td>
<td>DELETE /userId/messages/id</td>
<td>Immediately and permanently deletes the specified message. This operation cannot be undone. Prefer <code>messages.trash</code> instead.</td>
</tr>
<tr>
<td>get</td>
<td>GET /userId/messages/id</td>
<td>Gets the specified message.</td>
</tr>
<tr>
<td>insert</td>
<td>POST /userId/messages</td>
<td>Directly inserts a message into only this user's mailbox similar to <code>IMAP APPEND</code>, bypassing most scanning and classification. Does not send a message.</td>
</tr>
<tr>
<td>list</td>
<td>GET /userId/messages</td>
<td>Lists the messages in the user's mailbox.</td>
</tr>
<tr>
<td>modify</td>
<td>POST /userId/messages/idmodify</td>
<td>Modifies the labels on the</td>
</tr>
</tbody>
</table>
# Plans and pricing

GitHub is free to use for public projects. Collaborate on private repositories with any of our paid plans.

Sign up now

## Personal plans

For individuals looking to share their own projects and collaborate with others.

<table>
<thead>
<tr>
<th></th>
<th>Free</th>
<th>Micro</th>
<th>Small</th>
<th>Medium</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$0/month</td>
<td>$7/month</td>
<td>$12/month</td>
<td>$22/month</td>
<td>$50/month</td>
</tr>
<tr>
<td>Collaborators</td>
<td>Unlimited</td>
<td>Unlimited</td>
<td>Unlimited</td>
<td>Unlimited</td>
<td>Unlimited</td>
</tr>
<tr>
<td>Public repositories</td>
<td>Unlimited</td>
<td>Unlimited</td>
<td>Unlimited</td>
<td>Unlimited</td>
<td>Unlimited</td>
</tr>
<tr>
<td>Private repositories</td>
<td>0</td>
<td>5</td>
<td>10</td>
<td>20</td>
<td>50</td>
</tr>
</tbody>
</table>

**Do you offer education discounts?**

Yes, we offer free and discounted plans to students, teachers, and schools. You can find more information and apply on our education site.
PaaS Platform as a Service

- jailed runtime available to host applications
- generic or provider-specific APIs
- no control over the environment

Logos for Elastic Beanstalk, Heroku, and Windows Azure
$ heroku login
...

$ git clone https://github.com/heroku/java-getting-started.git
$ cd java-getting-started

$ heroku create
Creating warm-eyrie-9006... done, stack is cedar-14
http://warm-eyrie-9006.herokuapp.com/ | git@heroku.com:warm-eyrie-9006.git
Git remote heroku added

$ git push heroku master
...
    http://warm-eyrie-9006.herokuapp.com/ deployed to Heroku

$ heroku ps:scale web=1
Scale & estimate your dyno cost

1X
512MB RAM
1x CPU Share
$0.05/dyno-hour

2X
1024MB RAM
2x CPU Share
$0.10/dyno-hour

Performance Dynos
6GB RAM
Superior Quality-of-Service
Low latency, high throughput
$0.80/dyno-hour

Summary

Dynos
$682.50
Databases
$0.00
Support
$0.00
Add-ons
$0.00

$682.50
Estimated monthly cost

Sign up for free

IronMQ from $0/mo
Highly available elastic message queuing service.

Plans

<table>
<thead>
<tr>
<th>Plan</th>
<th>Price</th>
<th>Requests per month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lite</td>
<td>Free</td>
<td>10M</td>
</tr>
<tr>
<td>Starter</td>
<td>$29/mo</td>
<td></td>
</tr>
<tr>
<td>Dev</td>
<td>$129/mo</td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>$499/mo</td>
<td></td>
</tr>
</tbody>
</table>

Requests per month

- Unlimited Queues
- High Availability
- Persistent Messages
- Push Queues

Sign up for free
IaaS
Infrastructure as a Service

deploy arbitrary software stacks that are typically associated with cloud computing. With IaaS, customers can gain:

- **Low-level resources**: Access to the underlying hardware resources such as servers, storage, and networking infrastructure.
- **Complete control**: Full control over the network, storage, and operating system (OS) configuration.

This approach allows for greater customization and flexibility compared to other cloud service models like Platform as a Service (PaaS) or Software as a Service (SaaS).
Resources

You are using the following Amazon EC2 resources in the EU West (Ireland) region:

- 2 Running Instances
- 9 Volumes
- 1 Key Pair
- 0 Placement Groups
- 0 Elastic IPs
- 6 Snapshots
- 2 Load Balancers
- 6 Security Groups

Easily deploy Ruby, PHP, Java, .NET, Python, Node.js & Docker applications with Elastic Beanstalk.

Create Instance

To start using Amazon EC2 you will want to launch a virtual server, known as an Amazon EC2 instance.

Launch Instance

Note: Your instances will launch in the EU West (Ireland) region

Service Health

Service Status:
- EU West (Ireland):
  - This service is operating normally

Availability Zone Status:
- eu-west-1a:
  - Availability zone is operating normally

Scheduled Events

EU West (Ireland):
- No events
### Summary

<table>
<thead>
<tr>
<th>AWS Service Charges</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$0.00</td>
</tr>
</tbody>
</table>

*There are no invoices for the selected month.*

### Details

<table>
<thead>
<tr>
<th>AWS Service Charges</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$0.00</td>
</tr>
</tbody>
</table>

#### Elastic Compute Cloud

<table>
<thead>
<tr>
<th>EU (Ireland) Region</th>
<th>Usage</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon CloudWatch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$0.00 per alarm-month - first 10 alarms</td>
<td>0.373 Alarms</td>
<td>$0.00</td>
</tr>
<tr>
<td>$0.00 per metric-month - first 10 metrics</td>
<td>0.826 Metrics</td>
<td>$0.00</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>$0.00</td>
</tr>
</tbody>
</table>

### Amazon Elastic Compute Cloud running Linux/UNIX

| Region Total                           |          | $0.00    |

| Simple Notification Service            | $0.00    |
| Simple Queue Service                   | $0.00    |
| Simple Storage Service                 | $0.00    |
| CT to be collected                     | $0.00    |
Assume you could start with super reliable servers (MTBF of 30 years) Build computing system with 10 thousand of those Watch one fail per day

"Things will crash. Deal with it!"

Dean Keynote, LADIS 2009
Typical first year for a new google cluster

~0.5 overheating (power down most machines in <5 mins, ~1-2 days to recover)
~1 PDU failure (~500-1000 machines suddenly disappear, ~6 hours to come back)
~1 rack-move (plenty of warning, ~500-1000 machines powered down, ~6 hours)
~1 network rewiring (rolling ~5% of machines down over 2-day span)
~20 rack failures (40-80 machines instantly disappear, 1-6 hours to get back)
~5 racks go wonky (40-80 machines see 50% packetloss)
~8 network maintenances (4 might cause ~30-minute random connectivity losses)
~12 router reloads (takes out DNS and external vips for a couple minutes)
~3 router failures (have to immediately pull traffic for an hour)
~dozens of minor 30-second blips for dns
~1000 individual machine failures
~thousands of hard drive failures
slow disks, bad memory, misconfigured machines, flaky machines, etc.
Long distance links: wild dogs, sharks, dead horses, drunken hunters, etc.
A distributed system is one in which the failure of a computer you didn't even know existed can render your own computer unusable.

Leslie Lamport
Building fault tolerant services

at every level

be pessimistic

deal with failures

deal with inconsistency
October 21th 2016: dynDNS targeted by a DDoS

1.2 Tb/s of DNS lookups

Affected services:

- Airbnb[12]
- Amazon.com[9]
- Ancestry.com[13][14]
- The A.V. Club[15]
- BBC[14]
- The Boston Globe[12]
- Box[16]
- Business Insider[14]
- CNN[14]
- Comcast[17]
- CrunchBase[14]
- DirecTV[14]
- The Elder Scrolls Online[14][18]
- Electronic Arts[17]
- Etsy[12][19]
- FiveThirtyEight[14]
- Fox News[20]
- The Guardian[20]
- GitHub[12][17]
- Grubhub[21]
- HBO[14]
- Heroku[22]
- HostGator[14]
- iHeartRadio[13][23]
- Imgur[24]
- Indiegogo[13]
- Mashable[25]
- National Hockey League[14]
- Netflix[14][20]
- The New York Times[12][17]
- Overstock.com[14]
- PayPal[19]
- Pinterest[17][19]
- Pixlr[14]
- PlayStation Network[17]
- Qualtrics[13]
- Quora[14]
- Reddit[13][17][19]
- Roblox[26]
- Ruby Lane[14]
- RuneScape[13]
- SaneBox[22]
- Seamless[24]
- Second Life[27]
- Shopify[12]
- Slack[24]
- SoundCloud[12][19]
- Squarespace[14]
- Spotify[13][17][19]
- Starbucks[13][23]
- Storify[16]
- Swedish Civil Contingencies Agency[28]
- Swedish Government[28]
- Tumblr[13][17]
- Twilio[13][14]
- Twitter[12][13][17][19]
- Verizon Communications[17]
- Visa[29]
- Vox Media[30]
- Walgreens[14]
- The Wall Street Journal[20]
- Wikia[13]
- Wired[16]
- Wix.com[31]
- WWE Network[32]
- Xbox Live[33]
- Yammer[24]
- Yelp[14]
- Zillow[14]
Jorns-MacBook-Pro:Stack jornjambers$ host -t NS us-east-1.amazonaws.com
us-east-1.amazonaws.com name server ns3.p31.dynect.net.
us-east-1.amazonaws.com name server ns2.p31.dynect.net.
us-east-1.amazonaws.com name server ns4.p31.dynect.net.
us-east-1.amazonaws.com name server ns1.p31.dynect.net.
Jorns-MacBook-Pro:Stack jornjambers$ host -t NS us-east-2.amazonaws.com
us-east-2.amazonaws.com name server pdns1.ultradns.net.
us-east-2.amazonaws.com name server pdns5.ultradns.info.
us-east-2.amazonaws.com name server pdns3.ultradns.org.
us-east-2.amazonaws.com name server u5.amazonaws.com.
us-east-2.amazonaws.com name server ns4.p31.dynect.net.
us-east-2.amazonaws.com name server ns2.p31.dynect.net.
us-east-2.amazonaws.com name server ns1.p31.dynect.net.
us-east-2.amazonaws.com name server u1.amazonaws.com.
us-east-2.amazonaws.com name server u2.amazonaws.com.
ol' school IT

you manage

applications
runtimes
integration/security
database
servers
virtualisation
server HW
storage
network
Paas

you manage

network
storage
server HW
virtualisation
servers
database
integration/security
runtimes

managed by vendor

applications
Open-source PaaS stacks
IaaS

you manage

managed by vendor

- applications
- runtimes
- integration/security
- database
- servers
- virtualisation
- server HW
- storage
- network
Open-source IaaS stacks

- Eucalyptus (2008+)
- OpenNebula (2008+)
- openstack™ (2010+)
- cloudstack (2012+)
vendor lock-in

Cloud Storage

Cloud Datastore

IaaS

PaaS

SaaS
Deployment models
public cloud

general availability to everyone

the “real” cloud
reduced costs
trust issues?
cloud computing vs. fog of war
give me your code & data
Trust in me
I'm aware read my mails

what is my is hacked?
private cloud
self hosted cloud

might reduce TCO
stronger trust
better manageability
multi-clouds

you spread your application
avoid Single Point of Failures
*take the benefits of each cloud
inter-clouds

they outsource your components
agreements between the providers
“cloud of clouds”
fog computing
community cloud

private cloud by and for multiple organizations
RECAP
CLOUD IS ABOUT REDUCING COSTS
CLOUD IS ABOUT SCALABILITY
CLOUD IS ABOUT RESILIENCY
CLOUD IS ABOUT TRUST