# Big Data Training

**Amazon Web Services** 

## Introducción

Cloud computing o computación en nube es un paradigma para ofrecer servicios de computación en Internet.



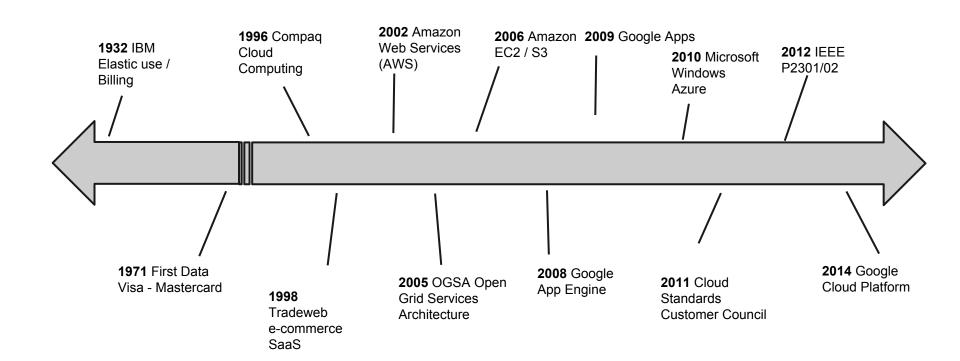
### Introducción

> Ilusión de **recursos** de computación **infinitos** disponibles **bajo demanda**.

> Eliminación de compromisos tempranos de los Cloud Users

> Pagar por el uso de recursos de computación

## **Evolución**



## Estandarización

A Standard Isn't a Document - It's a Process!

- > IEEE P2301 (Cloud Profiles)
- > IEEE P2302 (Intercloud)

# Otros Paradigmas

**Cloud Computing** 

**VS** 

In House



#### **Escalabilidad**

> Crecimiento y decrecimiento de las instancias según la necesidad del sistema. Convertir CapEx a OpEx.

#### Disponibilidad del servicio

> La disponibilidad del servicio es mayor en un ámbito distribuido como también su elasticidad para evitar ataques DoS y DDoS.

#### Costo

> El costo es menor en términos de inversión, mantenimiento, cambio y migración de la tecnología actual.

#### Actualizaciones automáticas

> Las actualizaciones de SW/HW quedan delegadas al proveedor de cloud dependiendo del modelo de servicio.

#### Replicación

> Facilidad de replicar instancias y servicios

> Generación de backups y deploy de servicios para testing.

#### Uso de Energía

> Los datacenters consumen mucha energía. En la nube los nodos son usados de manera más eficiente.

#### **Ataques EDoS (Economic Denial of Sustainability)**

> Ante un ataque DoS/DDoS se genera un escalamiento dinámico del servicio para evitar que el servicio caiga. Esto puede provocar un aumento del costo económico y perjudicar al servicio.

> Solución: Capacity Planning

#### **Vendor Lock-in**

> La mayoría de las API de servicio son cerradas (clientes cautivos!)

> Solución: Standards

#### Confidencialidad de la información y auditabilidad

> La geolocalización distribuida de la información puede generar conflictos con las legislaciones de diferentes países.

> Solución: Cloud Geolocalizadas

Reputación compartida entre Cloud User y Cloud provider

> La mala praxis por parte de un usuario puede afectar la reputación de otro, por ejemplo que su IP sea blacklisteada.

> Solución: políticas en los proveedores de servicio

#### Reproducción de errores en entornos de desarrollo

> La replicación de errores se puede dificultar si no se conoce con detalle la infraestructura sobre la que se trabaja.

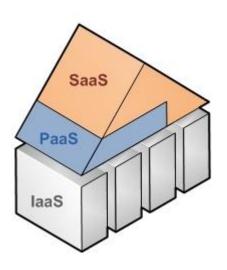
> Solución: soluciones open source

#### Privacidad!

> La información de carácter sensible está expuesta a terceros que pueden copiarla o acceder a ella.

> Solución: proveedores confiables / Criptografía

## Modelos



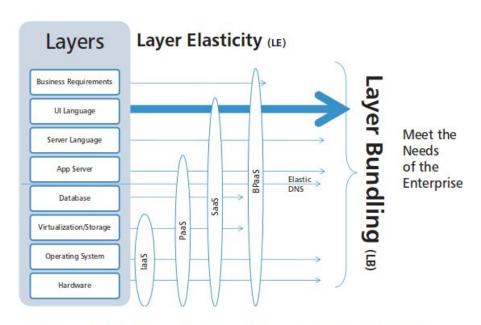


Figure 1 — The cloud is a convergence of layer elasticity and layer bundling.

#### laaS

#### Infrastructure as a Service

> Es un medio de entregar almacenamiento y capacidades de cómputo, virtualización y clustering, entre otros.

(Amazon EC2, Eucalyptus, IBM SBC, GoGrid, OpenNebula, Rackspace, RHEV-M, RimuHosting, Terremark, vSphere, OpenStack, FGCP)

All AWS Services



EC2



Lambda

#### Compute

Storage & Content Delivery

Database

Networking

Administration & Security

Analytics

Application Services

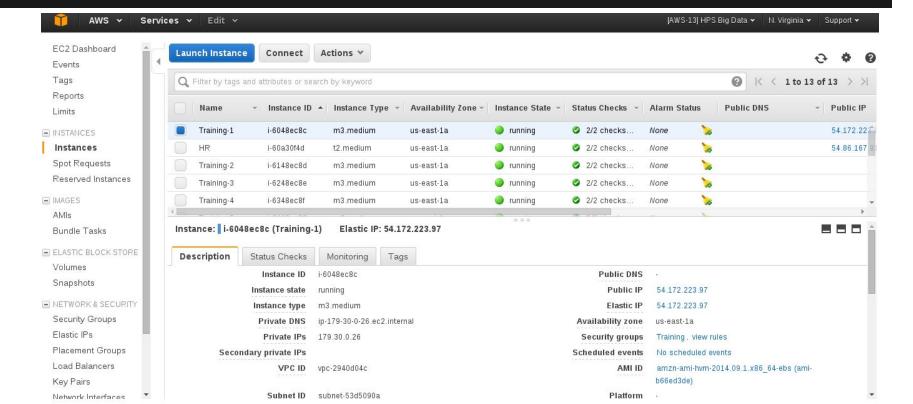
Deployment & Management

Mobile Services

Enterprise Applications

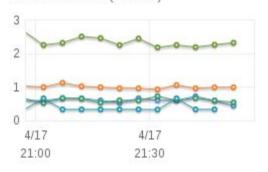
Amazon Elastic Compute Cloud (EC2) provides resizable compute capacity in the cloud

AWS Lambda is a compute service that runs your code in response to events and automatically manages the compute resources for you.

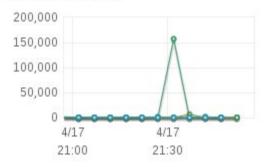


vCPU	ECU	Memoria (GiB)	Almacenamiento de instancias (GB)	Uso de Linux/UNIX
eneración	n actual			
1	3	3.75	1 x 4 SSD	\$0.070 por hora
2	6.5	7.5	1 x 32 SSD	\$0.140 por hora
4	13	15	2 x 40 SSD	\$0.280 por hora
8	26	30	2 x 80 SSD	\$0.560 por hora
	Aeneración 1 2 4	1 3 2 6.5 4 13	1 3 3.75 2 6.5 7.5 4 13 15	1 3 3.75 1 x 4 SSD 2 6.5 7.5 1 x 32 SSD 4 13 15 2 x 40 SSD

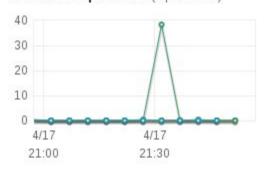
#### CPU Utilization (Percent)



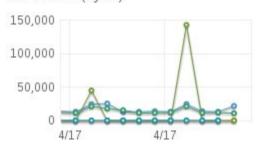
#### Disk Reads (Bytes)



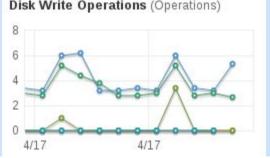
#### Disk Read Operations (Operations)



Disk Writes (Bytes)

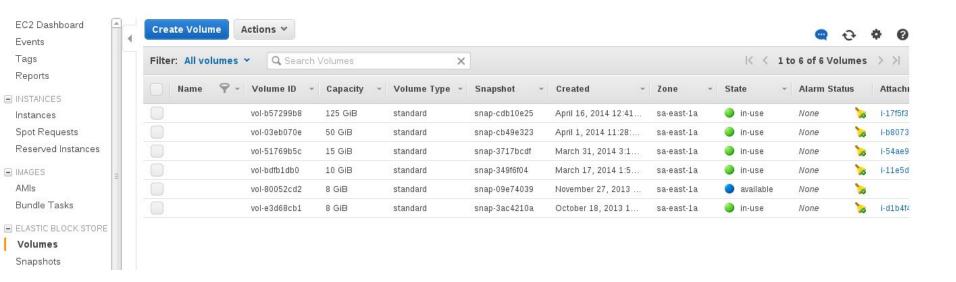


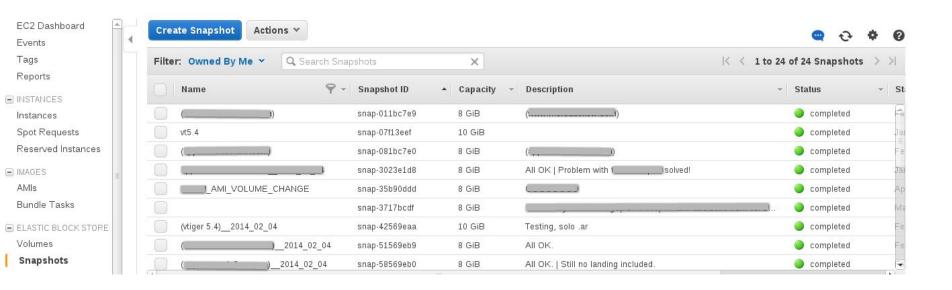
#### Disk Write Operations (Operations)

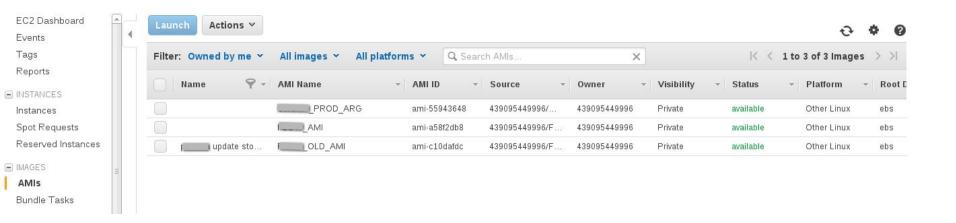


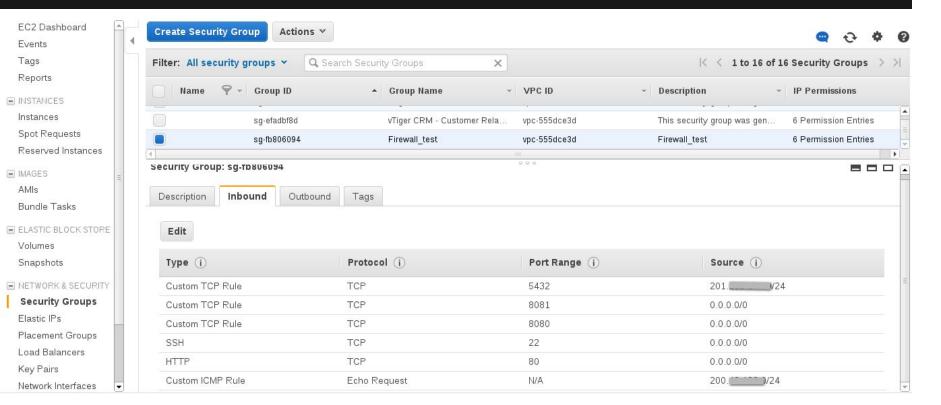
#### Network In (Bytes)

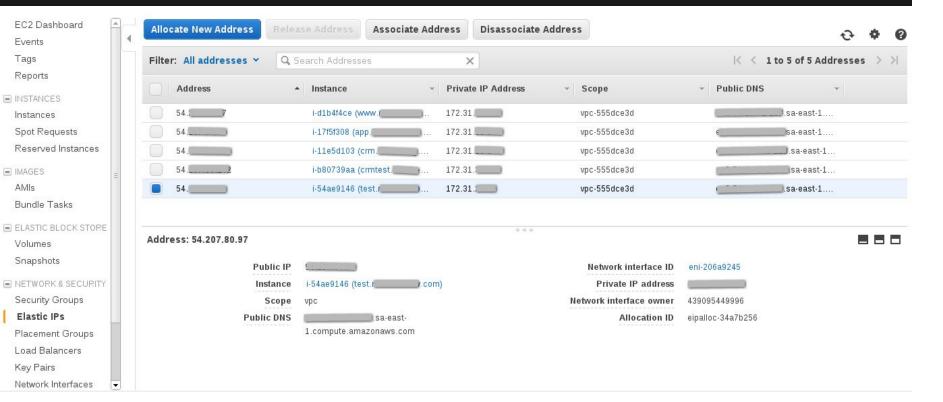














All AWS Services



Glacier

CloudFront

Compute

Storage & Content Delivery >

Amazon Simple Storage Service (S3) can be used to store and retrieve any amount of data

Amazon Glacier is a low-cost storage service that provides secure and durable storage for data archiving and backup.

Amazon CloudFront provides a way to distribute content to end users with low latency and high data transfer speeds.

Database

Networking

Administration & Security

Analytics

Application Services

Deployment & Management

Mobile Services

Enterprise Applications



Storage Gateway

AWS Storage Gateway securely integrates on-premises IT environments with cloud storage for backup and disaster recovery.

All AWS Services

Compute

Storage & Content Delivery

Database

Networking

Administration & Security

Analytics

Application Services

Deployment & Management

Mobile Services

Enterprise Applications



availability.

RDS

Amazon Relational Database Service (RDS) provides familiar SQL databases while automatically managing administrative tasks.

DynamoDB

Amazon DynamoDB is a scalable NoSQL data store that manages

distributed replicas of your data for high



ElastiCache

Amazon ElastiCache improves application performance by allowing you to retrieve information from an inmemory caching system.



Redshift

Amazon Redshift is a fast, fully managed, petabyte-scale data warehouse service that lets you efficiently analyze all your data using your existing business intelligence tools

All AWS Services





🚹 Route 53

Compute

Storage & Content Delivery

Database

Amazon Virtual Private Cloud (VPC) lets you launch AWS resources in a private, isolated cloud. AWS Direct Connect lets you establish a dedicated network connection from your network to AWS. Amazon Route 53 is a scalable and highly available Domain Name System (DNS) and Domain Name Registration service.

Networking

Administration & Security

Analytics

Application Services

Deployment & Management

Mobile Services

Enterprise Applications

All AWS Services



Directory Service



Trusted Advisor

AWS environment and finds

help close security gaps.

AWS Trusted Advisor inspects your

opportunities to save money, improve

system performance and reliability, or



Confia

Compute

Storage & Content Delivery

Database

Networking



Analytics

Application Services

Deployment & Management

Mobile Services

Enterprise Applications

AWS Directory Service provides

managed directories in the cloud.



AWS Identity and Access Management (IAM) lets you securely control access to AWS services and resources



CloudTrail

AWS CloudTrail provides increased visibility into user activity by recording API calls made on your account.



CloudWatch

AWS Config gives you inventory of your

AWS resources, lets you audit resource

configuration history, and notifies you

when resource configurations change.

Amazon CloudWatch provides monitoring for resources and applications.

Amazon Elastic MapReduce lets you

perform big data tasks such as web

indexing, data mining, and log file

All AWS Services



analysis.

u Kinesis



driven workflows

Compute

Storage & Content Delivery

Database

Networking

Administration & Security

#### Analytics

Application Services

Deployment & Management

Mobile Services

Enterprise Applications

W Kine

Amazon Kinesis is a managed service that scales elastically for real-time processing of streaming big data. AWS Data Pipeline is a lightweight orchestration service for periodic, data-

All AWS Services

E SQS

AppStream

SES

Compute

Storage & Content Delivery

Database

Networking

Administration & Security

**Application Services** 

Analytics

Deployment & Management

Mobile Services

Enterprise Applications

Amazon Simple Queue Service (SQS) offers a reliable, highly scalable, hosted queue for storing messages.

Amazon AppStream lets you stream resource intensive applications and games from the cloud to multiple enduser devices

Amazon Simple Email Service (SES) is a bulk and transactional email-sending service.

SWF

Amazon Simple Workflow (SWF) coordinates all of the processing steps within an application.



Elastic Transcoder

Amazon Elastic Transcoder lets you convert your media files in the cloud easily, at low cost, and at scale.



CloudSearch

AWS CloudSearch is a fully managed search service for websites and apps.

### laaS - Amazon EC2

All AWS Services

🥼 Elastic Beanstalk

CloudFormation

CodeDeploy

Compute

Storage & Content Delivery

Database

Networking

Administration & Security

Analytics

Application Services

Deployment & Management >

Mobile Services

Enterprise Applications

AWS Elastic Beanstalk is an application container for deploying and managing applications.

AWS CloudFormation lets you create and update a collection of related AWS resources in a predictable fashion.

AWS CodeDeploy lets you fully automate code deployments.



**OpsWorks** 

AWS OpsWorks is a DevOps platform for managing applications of any scale or complexity on the AWS cloud.

### laaS - Amazon EC2

All AWS Services



Mobile Analytics

SNS

Compute

Storage & Content Delivery

Database

Networking

Administration & Security

Analytics

Application Services

Deployment & Management

Mobile Services

Amazon Cognito is a simple user identity and data synchronization service that helps you securely manage and synchronize app data for your users across their mobile devices.

Amazon Mobile Analytics is a service that lets you easily collect, visualize, and understand app usage data at scale.

Amazon Simple Notification Service (SNS) lets you publish messages to subscribers or other applications.



### laaS - Amazon EC2

All AWS Services

WorkSpaces

Compute

Storage & Content Delivery

Database

Networking

Administration & Security

Analytics

Application Services

Deployment & Management

Mobile Services

**Enterprise Applications** 

Amazon WorkSpaces is a fully managed desktop computing service in the cloud.

Zocalo

Amazon Zocalo is a fully managed, secure enterprise storage and sharing service with strong administrative controls and feedback capabilities that improve user productivity.

#### **PaaS**

#### Platform as a Service

> Es la capa dentro del Cloud Computing que ofrece las herramientas de plataforma. Ej: Lenguajes de programación, pilas de desarrollo (LAMP - MEAN), bases de datos, webservers, etc.

(Heroku - Google App Engine - RedHat OpenShift)



```
The following components will be installed:
     App Engine SDK for Go (Linux, x86 64)
                                                   I 1.9.3 | 31.9 M
    App Engine SDK for Python and PHP
                                                   1.9.3 | 35.9 M
    BigOuery Command Line Tool
                                                        2.0.18 | < 1 M
     BigQuery Command Line Tool (Platform Specific)
                                                   2.0.18 | < 1 M
    Cloud SDK Core Command Line Tools
                                                            1 1
    Cloud SDK Core Libraries (Platform Specific) | 2014.04.08 | < 1 M
     Cloud SDK for Go Developers
                                                            1 1
    Cloud SQL Admin Command Line Interface
                                              | 2014.04.15 | < 1 M
    Cloud Storage Command Line Tool
                                                         3.42 | 1.8 M
    Cloud Storage Command Line Tool (Platform Specific) | 3.42 | < 1 M
     Compute Engine Command Line Tool
                                                      1.15.0 | < 1 M
    Compute Engine Command Line Tool (Platform Specific) | 1.13.0 | < 1 M
```

#### Archivo de configuración app.yaml

```
application: helloworld
version: 1
runtime: go
api_version: go1
handlers:
- url: /.*
script: _go_app
```

Instance Class	Memory Limit	CPU Limit	Cost per Hour per Instance
B1	128 MB	600 Mhz	\$0.05
B2	256 MB	1.2 Ghz	\$0.10
B4	512 MB	2.4 Ghz	\$0.20
B4_1G	1024 MB	2.4 Ghz	\$0.30
B8	1024 MB	4.8 Ghz	\$0.40
F1	128 MB	600 Mhz	\$0.05
F2	256 MB	1.2 Ghz	\$0.10
F4	512 MB	2.4 Ghz	\$0.20
F4_1G	1024 MB	2.4 Ghz	\$0.30



My Account | Help | Sign out

#### My Applications

Application	Title	Storage Scheme	Status
fifth-incentive-553	My First Project 1	High Replication	None Deployed
fine-transit-553	My First Project	High Replication	None Deployed
Create Application			Prev 20 1-2 of 2 Next 20

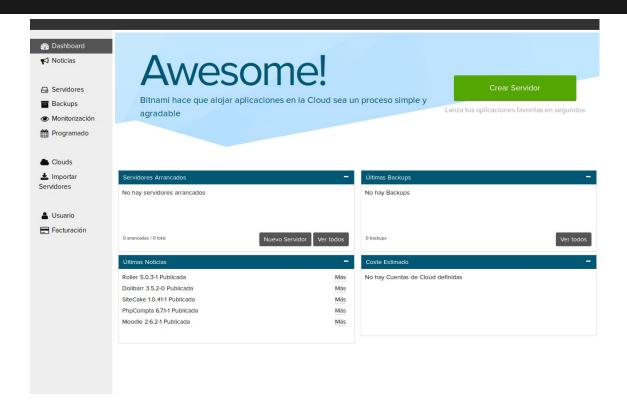


#### SaaS

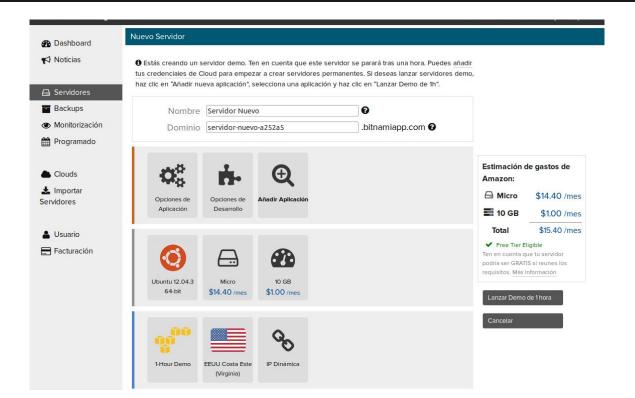
#### Software as a Service

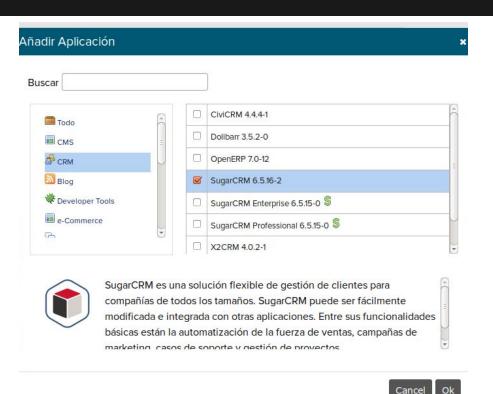
> Es un modelo para ofrecer el software como servicio, dejando al usuario solo la configuracion y utilizacion del mismo.

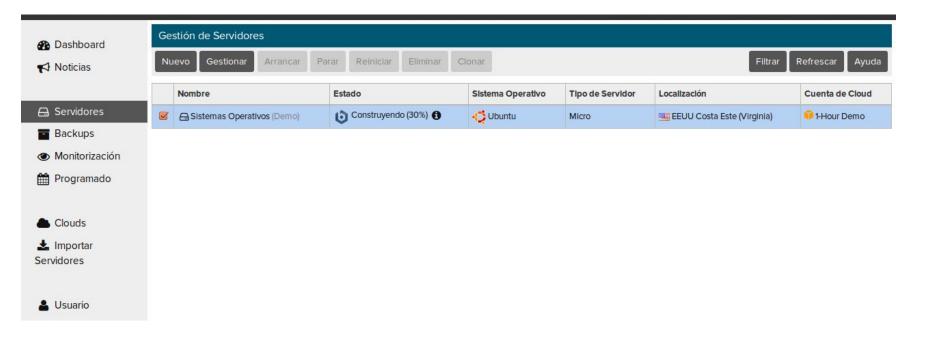
(ERP- CRM -BLOGS -FOROS - NOTICIAS - EMAILS - DOCS)

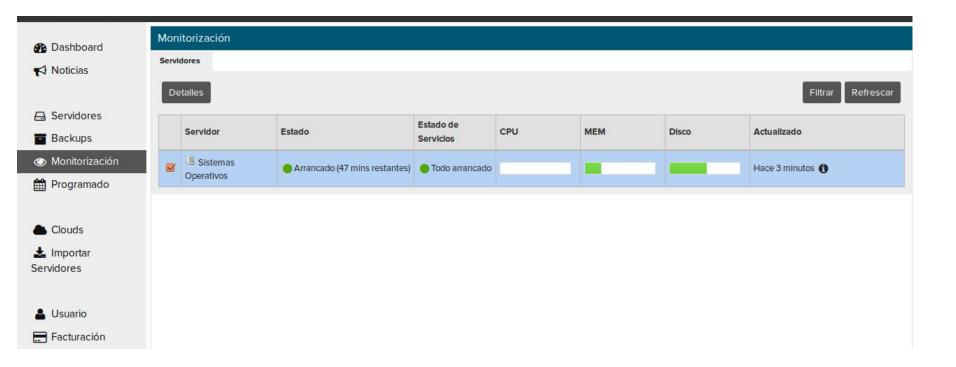


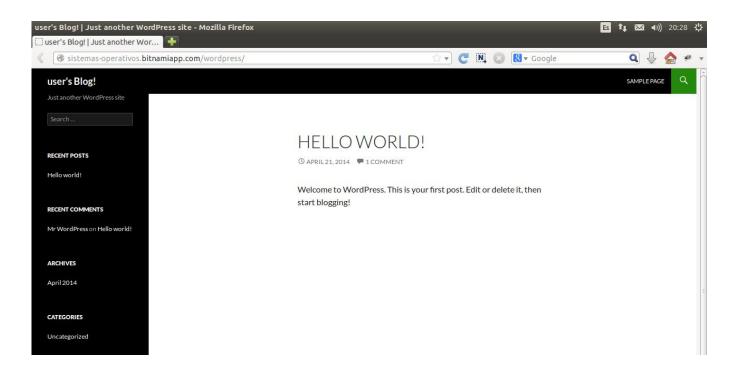
♠ Dashboard	Cuenta de Cloud	Ayuda	
Noticias	Guardar Cerrar Ayuda	Añadiendo una Cuenta de Cloud	
<ul><li>⇒ Servidores</li><li>⇒ Backups</li><li>◆ Monitorización</li></ul>	Configurar Cuenta de Cloud  Nombre Mi Cuenta de AWS  Localización por Defecto EEUU Costa Este (Virginia)	Para poder lanzar servidores con Bitnami Cloud Hosting, debes introducir las credenciales de tu cuenta en Amazon Web Services. Esta información será inmediatamente encriptada en nuestro sistema por seguridad y no será accessible a traves de nuestra interfaz.	
Programado	Credenciales de AWS  Access Key ID		
Clouds ★ Importar Servidores	Secret Access Key  1 Por favor consulta la información sobre cómo registrarte en Amazon EC2 y cómo conseguir tus credenciales	Cuando las credenciales hayan sido añadidas, Bitnami Cloud Hosting generará un nuevo par de claves SSH y	
▲ Usuario  ☐ Facturación	Contraseña de Administración  ① Por favor introduce una contraseña de administración para tu cuenta de Bitnami que será utilizada para encriptar tus credenciales de Cloud y otra información sensible. Se preguntará por esta contraseña antes de realizar algunas operaciones en tu cuenta de Cloud.  Contraseña de Administración  Repetir Contraseña	un 'Security Group'. Si quieres conectarte a los servidores por SSH, utiliza el botón de descarga que aparece al hacer click en la cuenta de cloud para obtener tu clave privada.  Nota: Bitnami Cloud Hosting permite qestionar servidores en varias cuentas	











#### X as a Service

> Dentro de las capas de laaS y SaaS existen múltiples servicios que pueden ofrecerse de forma individual, respetando el paradigma de Cloud Computing.

MBaaS: Mobile Backend as a Service.

> Incluye servicios de push notification, data storage, redes sociales a través de APIs y SDKs (Ej: Apache UserGrid)

#### **DBaaS**

Database as a Service

#### **DRaaS**

Disaster Recovery as a Service

#### **HDaaS**

HelpDesk as a Service

#### **DaaS**

Desktop as a Service

### Conclusiones

> La computación en nube reduce costos, flexibiliza la elasticidad, facilita el deploy y pone énfasis en los aspectos que el consumidor necesita.

> Es un modelo adecuado, sencillo y de bajo costo.

### Links útiles

AWS Status http://status.aws.amazon.com/

AWS Pricing https://aws.amazon.com/ec2/pricing/

AWS Instances http://aws.amazon.com/ec2/instance-types/

RH OpenShift https://www.openshift.com/

Bitnami https://bitnami.com/

IEEE http://cloudcomputing.ieee.org/standards/articles

Open Grid http://forge.gridforum.org/sf/projects/ogsa-wg/

Coursera https://www.coursera.org/course/mobilecloud