

BID DATA SI SCALABILITATE

Introducere in AWS si Elastic MapReduce

Amazon Web Services (AWS)



Ce este AWS



Infrastructure as a service

Ce este AWS

- O mare ferma de masini disponibile la cerere
- Platesti doar pentru ce folosesti
- Servicii peste ferma de masini
 - ▣ Monitorizare
 - ▣ Auto scalare
 - ▣ Wrapper-e pentru configuratii comune
 - ▣ Imagini preconfigurate pentru masini
 - ▣ etc

Mai exact ce ofera AWS ?

Retelistica si putere de calcul

- EC2 – Masini pentru uz general
- VPC – Retea virtuala (Virtual private cloud)
- Route53 – Serviciu DNS
- Elastic MapReduce – Hadoop

Mai exact ce ofera AWS ?

Stocare si livrare de continut

- Cloud front – Livrare de continut static
- S3 – Stocare **nelimitata** <cheie, valoare>
- Glacier – Stocare ieftina dar cu timp mare de acces

Mai exact ce ofera AWS ?

Baze de date

- DynamoDB – NoSQL Data store
- RDS – Relational database
- ElastiCache – In memory cache

Mai exact ce ofera AWS ?

Gestionare

- ❑ CloudFormation – Infrastructure as code
- ❑ Elastic Beanstalk – Infrastructura pentru aplicatii web
- ❑ CloudWatch - Monitorizare
- ❑ IAM – Gestionare utilizatori

Mai exact ce ofera AWS ?

Aplicatii ca servicii

- ❑ SNS – Sistem de notificari
- ❑ SQS – Cozi multi redundante
- ❑ SES – Servciul de email
- ❑ Elastic transcoder – Convertor video
- ❑ SWF – Serviciu de “Workflow”
- ❑ CloudSearch – Motor de cautare

Argumente PRO Cloud

- Te poti concentra pe aplicatie, nu pe infrastructura
- Putere “nelimitata” de calcul
- Platesti pentru ce folosesti
- etc

Argumente CONTRA Cloud

Ai ales un provider, cu el ramai!

Consola AWS



Services ▾



EC2



VPC



Route 53



S3



SQS



IAM



RDS

Edit ▾

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 **NEW**

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 **NEW**

DevOps Application Management Service

App Services



CloudSearch

Managed Search Service



Elastic Transcoder **NEW**

Easy-to-use Scalable Media Transcoding



SES

Email Sending Service



SNS

Push Notification Service



SQS

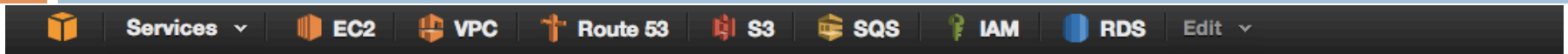
Message Queue Service



SWF

Workflow Service for Coordinating Application Components

Consola AWS: EC2



EC2 Dashboard

Events

Tags

INSTANCES

Instances

Spot Requests

Reserved Instances

IMAGES

AMIs

Bundle Tasks

ELASTIC BLOCK STORE

Volumes

Snapshots

NETWORK & SECURITY

Security Groups

Elastic IPs

Placement Groups

Load Balancers

Key Pairs

Network Interfaces

Resources

You are using the following Amazon EC2 resources in the US West (N. California) region:

0 Running Instances

0 Volumes

0 Key Pairs

Placement Groups not supported in this region

0 Elastic IPs

0 Snapshots

0 Load Balancers

3 Security Groups

... Optimize your resources' cost, performance and security with [AWS Trusted Advisor](#)

Hide

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 US West (N. California) region

Service Health

Service Status:

✓ US West (N. California):
This service is operating normally

Availability Zone Status:

✓ us-west-1a:
Availability zone is operating normally

✓ us-west-1b:
Availability zone is operating normally

[Service Health Dashboard](#)



Scheduled Events



US West (N. California):

No events

API AWS

- Ofera un API http public
- CLI – Command line interface
 - ▣ `aws ec2 describe-instances`
 - ▣ `aws ec2 start-instances --instance-ids i-1348636c`
 - ▣ etc

Elastic MapReduce

- ❑ O distributie de Hadoop
- ❑ Fara HDFS. Foloseste S3 pentru stocare. Datele sunt tot “aproape” de procesare
- ❑ Amazon gestioneaza cluster-ul