

# Cloud Computing

Diego Casella, Fabio Marcuzzi  
February 24, 2015



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

- 1 Introduzione
- 2 Amazon Elastic Compute Cloud
- 3 Amazon Elastic Compute Cloud - API
- 4 Amazon Elastic MapReduce
- 5 Amazon Elastic MapReduce - API
- 6 Possibili Utilizzi dei servizi AWS
- 7 Conclusioni

In questo quinto seminario, analizzeremo quali strumenti e servizi siano disponibili per potere realizzare una applicazione che sfrutti in pieno le potenzialità offerte dal cloud. Ci concentreremo su di alcuni servizi web più diffusi ed alla luce di essi, quali criticità possono insorgere nel software e quali possibili soluzioni adottare.

Amazon Elastic Compute Cloud, d'ora in poi *Amazon EC2* o *EC2* per brevità, é un servizio che mette a disposizione dell'utente *capacità di calcolo ridimensionabile*, sollevando l'utente dall'onere di investire in hardware performante ed alla sua inevitabile manutenzione.

Le caratteristiche offerte dal servizio *Amazon EC2* sono:

- ambienti di virtual computing, dette *istanze*;
- template pre-configurati di tali istanze, dette *Amazon Machine Images* o *AMI*, contenenti quanto necessario per avere un server operativo;
- storage temporaneo, che serve per immagazzinare dati che poi vengono eliminati quando viene spenta o terminata una associata istanza, detto *instance storage volume*;
- storage persistente che utilizza gli *Amazon Elastic Block Store*, o *Amazon EBS*, organizzato in uno o più *Amazon EBS volumes*;

Le caratteristiche offerte dal servizio *Amazon EC2* sono:

- firewall configurabile;
- indirizzo IP statico per dynamic cloud computing, denominato *Amazon Elastic IP Address*;
- locazioni fisiche multiple, come ad esempio di istanze o volumi *EBS*, dette regioni o *Availability Zones*, che consentono di assegnare istanze a determinate zone geografiche, ad esempio Europa o Stati Uniti, per migliorare la distribuzione del carico di lavoro e la latenza degli applicativi;
- metadata, dette anche *tags*, che possono essere create ed assegnate alle varie risorse *Amazon EC2*;
- reti virtuali logicamente separate dal resto del cloud *AWS*, ma che possono essere connesse alla propria rete, dette *Virtual Private Clouds* o *VPCs*;

## Esempio - AmazonEC2Client

```
AWSCredentials credentials = new AWSCredentials(...);
AmazonEC2Client client = new AmazonEC2Client( credentials );
// allocates an Elastic IP address
AllocateAddressResult r = client.allocateAddress();
// copies an existing AMI CopyImageResult r = client.copyImage(new
CopyImageRequest(...));
// starts an AMI
client.startInstances(new StartInstancesRequest(...));
// stops an AMI
client.stopInstances(new StopInstancesRequest(...));
```

*Amazon Elastic MapReduce*, abbreviato *Amazon EMR*, é un webservice che permette di processare grandi moli di dati. Esso utilizza Apache Hadoop, framework per il calcolo distribuito su server o cluster di servers, il quale si prende carico di assegnare i dati nelle varie istanze di *Amazon EC2*. Viene usato in molteplici ambiti, dal machine learning all'analisi finanziaria.



I vantaggi che *Amazon EMR* fornisce sono:

- si può lanciare un cluster *Amazon EMR* in pochi minuti;
- non serve incaricarsi del setup dei nodi, del/i cluster, di Hadoop, del loro tuning a seconda dei dati processati;
- è scalabile a piacimento, a seconda della dimensione dei dati usati;
- ogni istanza è comunque configurabile singolarmente: il sistemista ha accesso root in ogni singola macchina;
- sicurezza, dato che *Amazon EMR* può essere fatto girare all'interno di una *Amazon VPC*, logicamente separata dall'esterno;

## Esempio - AmazonElasticMapReduceClient

```
client = AmazonElasticMapReduceClient( new AWSCredentials(...) );  
client.addInstanceGroups(new AddInstanceGroupsRequest(...) );  
client.runJobFlow( new RunJobFlowRequest(...) );  
client.terminateJobFlow( new TerminateJobFlowRequest(...) );  
client.listClusters();
```

É ragionevole pensare, alla luce di questi servizi visti fino ad ora, di poter evolvere un applicativo pensato per l'utilizzo nel cloud, in maniera tale da potere scalare in maniera efficiente con la taglia del problema computazionale che gli viene sottoposto. Attraverso l'uso dell'API resa disponibile per i servizi AWS, é ipotizzabile poter evolvere un applicativo nel cloud in maniera tale da poter utilizzare piú potenza di calcolo, quando se ne presenti la necessità, e tutto ciò in maniera del tutto trasparente all'utente finale che non sa, e non é necessario che se ne renda conto, di cosa sta avvenendo in background.

Sebbene questa evoluzione rappresenti un notevole vantaggio per aumentare le performance dell'applicativo in analisi, ci sono ovviamente altri aspetti da tenere in considerazione, come il redesign dell'applicativo per poter sfruttare i nuovi servizi, la gestione delle istanze, dei volumi di storage, ecc..

