UNIVERSITA' DEGLI STUDI DI
NAPOLI FEDERICO II

Facoltà di Ingegneria

Corso di Studi in Ingegneria Informatica

Elaborato finale in **Reti di Calcolatori**

Cloud Computing: la piattaforma Amazon Web Services (AWS)

Anno Accademico 2011/2012

Candidato:

Fabio Riccio

matr. N46000828

Ai miei Nonni,
a Zio Gianni,
a Maria.

Indice

Introduzione

Molto spesso nel corso della storia, le più grandi evoluzioni, tecnologiche e non, passano quasi inosservate, pur rivoluzionando, o meglio evolvendo, le più semplici azioni quotidiane di milioni di persone.

Infatti al giorno d'oggi, leggere la propria posta elettronica dovunque ci si trovi, oppure caricare la propria libreria musicale on-line per ascoltarla successivamente da un altro device, o ancora, pagare il mutuo comodamente da casa, sono azioni semplici e all'ordine del giorno. Eppure, fino a pochissimi anni fa, questi non erano per niente gesti quotidiani.

Quella che in parte è avvenuta e in parte sta ancora avvenendo è un'evoluzione dell'informatica tradizionale.

Secondo l'IDC[1] " è questa la grande portata del cambiamento, che vede il Cloud, insieme ad altri fenomeni, traghettare l'industria ICT verso la cosiddetta "Terza Piattaforma". Sancisce, di fatti, l'ingresso nella "terza era" – come definito da IDC – dopo quella dei mainframe e la seconda fase legata alla diffusione di PC, database, reti, ambienti client-server. Un rinnovamento del settore ICT che promette di espandere radicalmente l'uso delle tecnologie informatiche, portando ad una nuova e intelligente varietà di soluzioni ".

In pratica la tecnologia Cloud Computing permette di utilizzare qualsiasi tipo di documento senza aver bisogno di hard disk e archivi digitali. I software, invece di essere installati direttamente sui propri computer, risiedono in rete, ossia in "Cloud", letteralmente nella "nuvola". I dati, fino ad oggi salvati sui nostro personal computer,

[1] IDC (International Data Corporation) è il primo gruppo mondiale specializzato in ricerche di mercato, servizi di consulenza e organizzazione di eventi nei settori dell'Information Technology, delle telecomunicazioni e della tecnologia consumer.

verranno invece decentrati su diversi server e saranno accessibili, tramite browser e applicazioni, da qualsiasi device dotato di connessione.

I vantaggi saltano subito all'occhio e vanno dalla possibilità di poter accedere ai nostri file, alle nostre foto e ai nostri documenti dovunque ci si trovi e in qualsiasi momento, al poter viaggiare più leggeri senza aver l'obbligo di dover portare con se unità di storage o addirittura gli stessi computer. Non mancano nemmeno gli svantaggi, due su tutti: la necessità di una connessione, che non sempre può essere disponibile, o la temporanea indisponibilità dei server sui quali sono immagazzinati i nostri dati.

L'obiettivo di questo lavoro è definire e analizzare in maniera breve ma completa il Cloud Computing e, nello specifico, la piattaforma Cloud di Amazon: Amazon Web Services (AWS), cercando di illustrare vantaggi e svantaggi di questa soluzione e proponendo un esempio di configurazione di una istanza di macchina virtuale su cloud.

Capitolo 1

Cloud Computing

1.1 Definizioni

Come definire il Cloud Computing, quindi?

Ne esistono varie definizioni, ognuna delle quali analizza e classifica il Cloud Computing secondo vari punti di vista. Descriverò ora la definizione standard di Cloud Computing, alcune definizioni alternative ed una personale definizione.

1.1.1 Definizione del NIST

Non esiste ancora una versione ufficialmente standardizzata per la definizione di Cloud Computing, esiste pero una definizione "standard" data dal NIST[2] ed è la seguente:

"Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics**,** three service models, and four deployment models".

Traducendo letteralmente, il Cloud Computing è un modello che abilita, in modo onnipresente, conveniente e on-demand, l'accesso, tramite rete, ad un pool condiviso di

[2] Il National Institute of Standards and Technology (NIST) è un agenzia del governo degli Stati Uniti d'America che si occupa della gestione delle tecnologie. Fa parte del Dipartimento del Commercio e il suo compito è la promozione dell'economia americana attraverso il lavoro con l'industria per sviluppare standard, tecnologie e metodologie che favoriscano la produzione e il commercio.

risorse di calcolo configurabili (ad esempio, reti, server, storage, applicazioni e servizi) che possono essere rapidamente acquisite e rilasciate, con un minimo sforzo di gestione o di interazione con il fornitore dei servizi**.**

Questo modello è composto da cinque caratteristiche essenziali, tre modelli di servizio, e quattro modelli di distribuzione.

Le cinque caratteristiche essenziali sono:

- **On-demand self-services** – il consumatore deve poter sfruttare autonomamente e direttamente il pool di risorse, come server time o il network storage, cioè senza l'interazione diretta con il fornitore dei servizi.
- **Broad network access** – le risorse disponibili in rete devono essere accessibili attraverso meccanismi standard che promuovano l'uso di client eterogenei, ad esempio smartphone, tablet, computer portatili e semplici postazioni di lavoro.
- **Resource pooling** – le risorse di calcolo del server sono raggruppate per servire più utenti contemporaneamente usando un modello *multi-tenant*[3], con diverse risorse fisiche e virtuali dinamicamente assegnate e riassegnate in base alla domanda dei consumatori. C'è un senso di indipendenza dal luogo, nel senso che il cliente, in genere, non ha alcun controllo o conoscenza sulla posizione esatta delle risorse messe a disposizione, ma può essere in grado di specificare la posizione ad un livello superiore di astrazione (ad esempio paese, stato o data center). Esempi di risorse includono archiviazione, la memoria e la larghezza di banda della rete.
- **Rapid elasticity** – le risorse possono essere elasticamente assegnate e rilasciate, in alcuni casi, automaticamente in modo da scalare rapidamente verso l'esterno e verso l'interno in relazione alla domanda. Per l'utente le risorse disponibili per l'assegnazione spesso sembrano essere illimitate e possono essere assegnate in qualsiasi quantità ed in qualsiasi momento.
- **Measured service** – i sistemi cloud controllano automaticamente e ottimizzano

3

l'uso delle risorse, attraverso il costante monitoraggio dei misuratori di performance appropriati per il tipo di servizio, come possono essere la conservazione, l'elaborazione, la larghezza di banda, e gli account utente attivi. Il monitoraggio viene anche riportato come garanzia di trasparenza sia per il provider che per l'utente del servizio utilizzato.

I tre modelli di servizio sono:

- **Software as a Service (Saas)** – L'utente userà le applicazioni, prodotte dal fornitore del servizio, che eseguono su un'infrastruttura cloud. Queste applicazioni sono accessibili tramite browser o interfacce web. L'utente non gestisce o controlla l'infrastruttura cloud di base, tranne che per alcune limitate impostazioni dell'applicazione specifiche dell'utente. Esempi ne sono GoogleDocs, IBM Lotuslive, Microsoft Online Services.
- **Platform as a Service (Paas)** – E' una evoluzione del modello SaaS nel senso che invece che sfruttare le singole applicazioni, l'utente ha possibilità di utilizzare, a consumo, una piattaforma di sviluppo delle proprie applicazioni e attivarle in cloud pagando per il tempo di utilizzo, la dimensione dell'hardware e le elaborazioni eseguite. L'utente non gestisce o controlla l'infrastruttura cloud di base tra cui la rete, i server, i sistemi operativi o di archiviazione, ma ha il controllo sulle applicazioni distribuite ed eventualmente sulle impostazioni per la configurazione dell'ambiente. Esempi ne sono Force.com Windows Azure.
- **Infrastructure as a Service (Iaas)** – All'utente è data la possibilità di eseguire le applicazioni sfruttando le risorse hardware in cloud, ad esempio unità di elaborazione, storage, reti e altre risorse di calcolo fondamentali. Le risorse non vengono assegnate a prescindere dall'effettivo utilizzo ma vengono assegnate su richiesta al momento in cui una piattaforma ne necessita. L'utente non gestisce o controlla l'infrastruttura cloud di base, ma ha il controllo su sistemi operativi, storage e applicazioni distribuite, ed il controllo limitato per selezionare componenti di rete. Esempi ne sono Amazon.com e Windows Azure.

I quattro modelli di distribuzione sono :

- **Private Cloud** – Detto anche *Internal Cloud* o *Corporate Cloud*, in questo caso l'infrastruttura cloud è un'architettura proprietaria che fornisce, in maniera esclusiva, servizi hosted ad una singola azienda oppure ad una organizzazione, quest'ultima interna all'azienda stessa o esterna, ossia gestita da terzi. Proprietà, gestione e organizzazione dell'infrastruttura sono dell'azienda stessa o date in mano a terzi o ancora ad una combinazione delle due. Si parlerà di on premise e off premise.
- **Community Cloud** – L'infrastruttura cloud è riservata ad una specifica comunità da parte di diverse organizzazioni che condividono problematiche e interessi. Anche in questo caso, proprietà, gestione e organizzazione dell'infrastruttura sono dell'azienda stessa o date in mano a terzi o ancora ad una combinazione delle due e si parlerà di on premise e off premise.
- **Public Cloud** – L'infrastruttura cloud è aperta al grande pubblico, ossia un service provider rende disponibili al pubblico, gratuitamente o secondo un modello *pay-per-use*, servizi come storage e applicazioni. Proprietà, gestione e organizzazione dell'infrastruttura sono di organizzazioni governative, di business o accademiche. Si parlerà di on premise sul provider cloud.
- **Hybrid Cloud** – L'infrastruttura cloud è una composizione di due o più infrastrutture cloud distinte (*private, community o public*) che rimangono soggetti unici, ma sono legati da tecnologia standardizzata o di proprietà che consente la portabilità di dati e applicazioni. Esempi ne sono Amazon Elastic Compute Cloud (EC2), Blue Cloud di IBM, Sun Cloud, Google App Engine e Windows Azure.

1.1.2 Definizioni "alternative"

Il Cloud Computing non sta incontrando i favori tutti e lo si può notare dalle parole non proprio positive di Richard Stallman[4]:

[4] Richard Matthew Stallman (New York, 16 marzo 1953) è un programmatore, un informatico e un attivista statunitense ma soprattutto fondatore del progetto GNU e della Free Software Foundation.

“É stupidità. É peggio della stupidità: è un’ingannevole campagna di marketing. Qualcuno sta dicendo che tutto questo sia inevitabile e ogni volta che si sente qualcuno dirlo, è perchè, molto probabilmente, c’è un insieme di aziende che lottano per renderlo vero" . La centralizzazione e delocalizzazione dei servizi viene vista, da uno dei maggiori esponenti e fautori mondiali del movimento per il software libero, come una minaccia proprio per quest’ultimo, dato che per poter usufruire di alcune applicazioni, non potremo più farlo liberamente e localmente dai nostri dipositivi ma dovremmo interfacciarci direttamente sui server proprietari delle case di sviluppo.

Anche le parole di Steve Wozniak[5] non sono benevole:

“Mi preoccupa tutto quello che riguarda la Nuvola. Penso che sarà orrendo. Penso che ci saranno molti problemi davvero orribili nei prossimi cinque anni. Nel Cloud non possiedi più nulla. Hai già firmato un contratto per cederlo. […] Io voglio percepire di essere proprietario delle mie cose […] Più cose trasferiamo online, meno resterà sotto controllo”.

Quantomeno curiose le parole di Larry Ellison[6]:

“La cosa interessante di cloud computing è che abbiamo ridefinito il cloud computing per includere tutto ciò che già facciamo. Non riesco a pensare a qualcosa che non sia il cloud computing con tutti questi annunci. L’industria del computer è l’unica industria guidata dalla moda più della stessa moda femminile. [...] Faremo annunci di cloud computing. Non combatterò contro questa cosa. Ma non capisco cosa faremo di diverso alla luce del cloud”.

Invece, secondo la Commissione Europea, riunitasi il 27 settembre 2012, “con cloud computing (nuvola informatica) si intende la conservazione, l’elaborazione e l’uso di dati su computer remoti accessibili via Internet. Molti utenti si avvalgono di questi servizi senza nemmeno saperlo. Diversi servizi, come le webmail o i social network, possono essere basati su tecnologie di cloud. Per gli utenti professionali di servizi informatici il cloud computing offre una grande flessibilità in termini di potenza di elaborazione. Se, ad esempio, aumenta la domanda per un determinato servizio, è semplice potenziarne le

[5] Stephen Gary Wozniak (Sunnyvale, 11 agosto 1950) è un informatico statunitense: le sue ricerche e le sue intuizioni portarono alla creazione dell’Apple I, il primo personal computer assemblato.

[6] Lawrence Joseph Ellison (Manhattan, 17 agosto 1944) è un imprenditore e informatico statunitense, co-fondatore e CEO della Oracle Corporation.

capacità in tempi molto più ridotti rispetto ad un'impresa che deve installare un dispositivo nel proprio centro dati".

Insomma come si può notare non esiste una versione unica e in grado di definire in maniera sintetica e precisa il concetto di Cloud Computing, quindi in maniera semplicistica lo definiremo come un qualsiasi servizio che condivide potenza di calcolo e di storage ed in generale una piattaforma di cloud la si può definire come una piattaforma che permette di aggiustare il livello di servizio a quello che serve in un particolare momento, se si ha la capacita di accorpare le risorse in un unico pool da gestire in modo dinamico, spostando i carichi di elaborazione per dare di meno a chi chiede di meno e di più a chi chiede di più.

1.2 Vantaggi e Svantaggi

Come si evince già dai precedenti paragrafi, i vantaggi sono molteplici e per semplicità li distingueremo in economici e tecnici. Iniziamo con i vantaggi economici e il primo di cui si deve tenere conto è sicuramente l'abbassamento del *TCO*, ossia il *Total Cost of Ownership*: aziende piccole, medie e grandi hanno costi ridotti e maggiore flessibilità nel breve e medio termine in quanto non hanno più bisogno di installare, configurare e dismettere sofisticati apparecchi o costose licenze. Inoltre le aziende possono aumentare la loro richiesta di hardware e software necessari, il tutto semplicemente con un cambio contrattuale, ossia on-demand, in base al livello di *QoS* ossia *Quality of Service* desiderato. Per quanto riguarda l'hardware, cè un ritorno al concetto di *thin client*, ossia il computer, ed in generale tutti i device dotati di connessione internet, sono soltanto terminali evoluti dai quali poter accedere ai servizi cloud online. Per quanto riguarda il software vengono invece bypassati gli acquisti delle costose licenze e dei relativi contratti di supporto.

Come secondo vantaggio economico, ne indichiamo uno indiretto: la possibilità di concentrare il focus sul proprio core business. In effetti l'azienda, risparmiando sul personale e le infrastrutture, può concentrarsi unicamente sulla realizzazione dei propri prodotti.

Tra i vantaggi tecnici annoveriamo la maggiore scalabilità e l'accessibilità in mobilità

oltre che l'indipendenza dalle periferiche. Per maggiore scalabilità si intende la capacità di allocare le risorse necessarie a fronte di una maggiore richiesta da parte del cliente.

Per l'accessibilità in mobilità, invece, si intende la possibilità di lavorare e accedere ai servizi da qualsiasi parte del mondo, con una semplice connessione internet, come se si lavorasse dal proprio ufficio. Infine, per l'indipendenza dalle periferiche si intende semplicemente che, essendo tutto gestito online, non si è legati ad alcun tipo di hardware o di configurazione di rete.

Per quanto riguarda gli svantaggi, ci focalizzeremo principalmente su tre problemi: sicurezza, conformità e privacy.

La sicurezza è il problema cruciale, in quanto, non solo non è possibile avere pieno controllo dei dati aziendali, ma la mancanza di uno standard ratificato per i sistemi cloud implica una diversificazione dei sistemi di sicurezza adottati. A questo fattore fondamentale, si aggiunga poi il problema della cancellazione dei dati, che non corrisponde ad una immediata cancellazione fisica degli stessi, e il rischio di rotture hardware e software in cui possono incorrere sia le piccole aziende che le grandi multinazionali.

La conformità invece racchiude al suo interno diverse problematiche come l'integrazione e la migrazione. Non per essere tediosi ma la mancanza di standard definiti a livello internazionale comporta una difficoltosa, se non addirittura impossibile, migrazione da una determinata piattaforma cloud verso altre piattaforme cloud con la relativa perdita di conformità e certificazioni acquisite a livello software e in alcuni casi, a livello hardware, una costosa riconfigurazione delle periferiche, locali e non, con la nuova infrastruttura verso la quale si sta migrando.

Infine la privacy, ossia tutto ciò che riguarda riservatezza e legalità. Per poter usufruire dei servizi cloud, come peraltro per quasi tutti gli altri servizi, vanno accettati i termini d'uso sottoposti dal fornitore che implica che i dati, sensibili e non, che si stanno caricando sui server, verranno trattati secondo le tecniche e le metodologie adottate dal fornitore stesso. Vanno quindi valutate, in particolar modo, le tecniche di sicurezza per quanto riguarda possibili attacchi hacker che, mentre prima colpendo il singolo computer, potevano

prelevare piccole porzioni di dati sensibili, ora invece possono in un sol colpo avere accesso a porzioni ben piu grandi se non addirittura interi settori.

CAPITOLO 2

Amazon Web Services (AWS)

2.1 Definizione e struttura

Per quanto riguarda la definizione di AWS, ci viene incontro Amazon stesso con una definizione propria[7]:

"Nel 2006, Amazon Web Services (AWS) ha iniziato ad offrire servizi di infrastruttura IT alle imprese sotto forma di servizi web - ormai comunemente noto come il cloud computing. Uno dei principali vantaggi del cloud computing è la possibilità di sostituire il capitale speso per le infrastrutture up-front con bassi costi variabili, che scalano con il vostro business. Con il Cloud, le aziende non hanno più bisogno di pianificare e acquistare server e altre infrastrutture IT con settimane o mesi di anticipo. Al contrario, possono avviare istantaneamente centinaia o migliaia di server in pochi minuti e ottenere risultati più velocemente".

Attualmente AWS, che serve con i suoi prodotti centinaia di migliaia di aziende in oltre 190 paesi nel Mondo, ha 9 data center localizzati in 7 diverse regioni del Mondo:

- Singapore (ASIA – PACIFICO) lanciato nel 2010
- Tokyo (ASIA – PACIFICO) lanciato nel 2011
- Sidney (ASIA – PACIFICO) lanciato nel 2012
- Irlanda (EUROPA) lanciato nel 2007

[7] http://aws.amazon.com/what-is-aws/

- Nord Virginia (EST USA) lanciato nel 2006
- Nord California (OVEST USA) lanciato nel 2009
- Oregon (OVEST USA) lanciato nel 2011
- GovCloud (USA) lanciato nel 2011
- Sao Paulo (BRASILE) lanciato nel 2011

Al momento stesso della creazione di un account su AWS, è possibile scegliere una di queste regioni e la scelta può essere definita in base ai costi, a particolari requisiti normativi e aziendali o, ancora, a zone con migliori latenze.

Ogni regione è composta da più *Availabilty Zones* ossia Zone di Disponibilità, che sono una delle caratteristiche di AWS. In pratica, ogni volta che si memorizzano oggetti o dati critici, questi non vengono immediatamente salvati nel data center, bensì attraversano uno step intermedio in cui vengono propagati e salvati in ognuna delle *availability zones* di quella particolare regione. Quando, infine, i nostri dati risultano salvati in ognuna di queste zone, allora vengono definitivamente salvati nella regione scelta. Sono essenzialmente due le ragioni di questo passaggio intermedio:

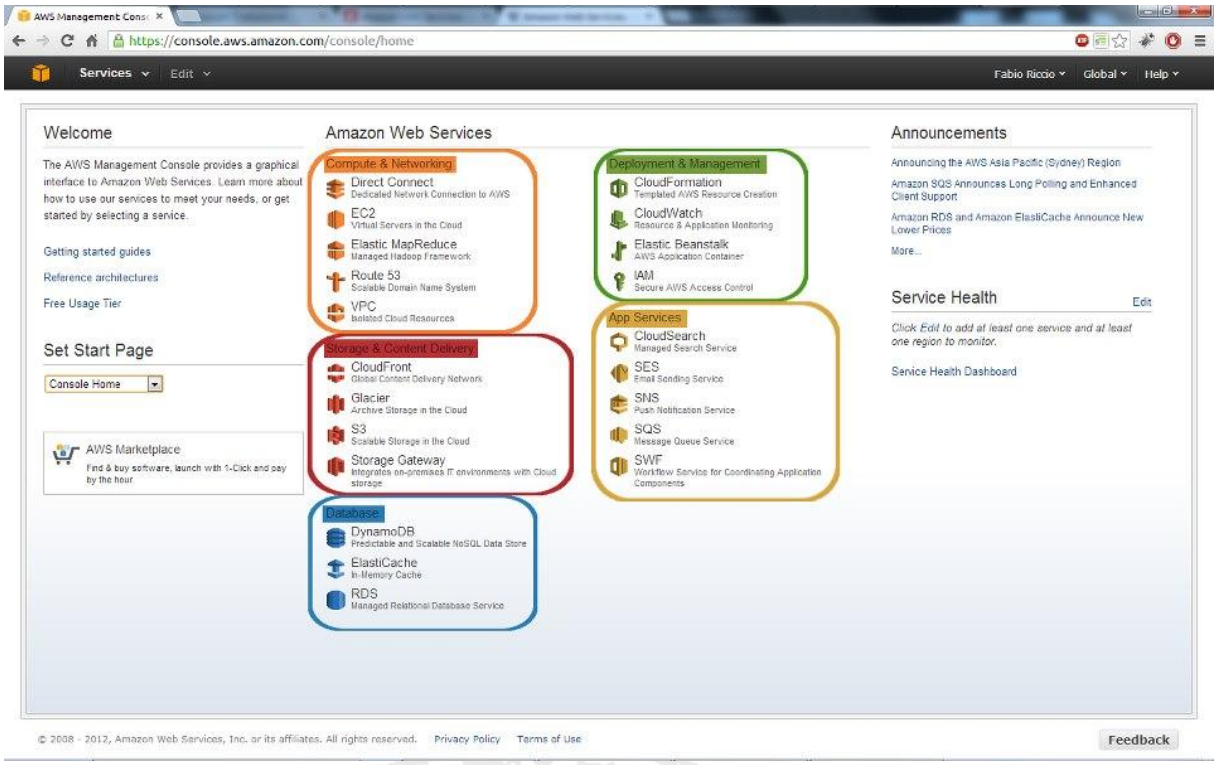
1. una ridondanza, ricercata e voluta, per evitare perdite di dati,
2. salvaguardare le informazioni nel caso una regione di memorizzazione fosse "non disponibile".

Il numero delle Zone di Disponibilità varia per ogni regione, infatti:

- Singapore, Sidney, GovCloud, Sao Paulo – 2 A. Z,
- Tokyo, Irlanda, Nord California, Oregon – 3 A.Z,
- Nord Virginia – 5 A.Z.

2.2 Servizi

I servizi offerti da Amazon Web Services sono classificati in 5 diverse categorie, come si evince dalla AWS Management Console:



- **Compute & Networking**: Direct Connect, EC2, Elastic MapReduce, Route 53, VPC.
- **Deployment & Management**: CloudFormation, CloudWatch, Elastic Beanstalk, IAM.
- **Storage & Content Delivery**: CloudFront, Glacier, S3, Storage Gateway.
- **App Services**: CloudSearch, SES, SNS, SQS, SWF
- **Database**: DynamoDB, ElastiCache, RDS.

Si noti che sono state messe in risalto, evidenziando e cerchiando rispetto alla figura originale, le diverse categorie. Passerò ora in rassegna tutte le categorie, descrivendo i servizi di ognuna.

2.2.1 Compute & Networking

La panoramica comincia con il **Direct Connect**.

Come ci informa la stessa Amazon[8], usando Direct Connect, è possibile stabilire la connettività privata tra AWS e il nostro data center, ufficio o ambiente lavorativo. Questa soluzione permette di ridurre i costi di rete, aumentare la velocità e larghezza di banda e fornire un'esperienza di rete più consistente di una basata su connessioni Internet.

Infatti il Direct Connect consente di stabilire una connessione di rete dedicata tra la propria rete aziendale e una delle locazioni degli AWS Direct Connect. La connessione è basata sullo standard VLAN 802.1Q[9] e può essere ripartita in più interfacce virtuali. Questo consente di utilizzare la stessa connessione per accedere alle risorse pubbliche, come gli oggetti memorizzati in Amazon S3 con un IP dello spazio degli indirizzi pubblici, e le risorse private come le istanze di Amazon EC2 in esecuzione all'interno di un Amazon VPC[10] utilizzando un IP dello spazio degli indirizzi privati, il tutto, quindi, mantenendo una netta separazione tra gli ambienti pubblici e privati. Inoltre, il tutto può essere riconfigurato in qualsiasi momento per soddisfare ogni esigenza.

Per quanto riguarda la **Amazon Elastic Compute Cloud**, meglio nota come **EC2**, faremo una trattazione breve per poi approfondirla nel prossimo capitolo con un esempio di configurazione di un'istanza di macchina virtuale.

Amazon EC2 è un servizio web progettato per facilitare il lavoro degli sviluppatori. Infatti è dotato di un'interfaccia web che permette di richiedere, ottenere e configurare in maniera rapida e diretta la necessaria capacità computazionale, ossia affittare un determinato numero di macchine virtuali sulle quale implementare ed eseguire le proprie applicazioni in fase di sviluppo o testing, il tutto fornendo il controllo completo delle proprie risorse, anche riguardo le latenze legate alle locazioni geografiche. Inoltre, riduce a pochi minuti il tempo necessario per ottenere e avviare queste nuove istanze del server, permettendo di

[8] http://aws.amazon.com/directconnect/

[9] È uno standard IEEE che permette a diverse reti virtuali VLAN di condividere lo stesso collegamento fisico, senza perdere informazioni tra i diversi apparati.

[10] Amazon Virtual Private Cloud (VPC) è un cloud privato all'interno dell'infrastruttura AWS, isolata logicamente dai prodotti di cloud pubblico della stessa infrastruttura.

scalare rapidamente la capacità, sia in up che in down, così come eventuali modifiche dei propri requisiti computazionali, da qui il significato di elasticità. Il tutto pagando solo le capacità effettivamente utilizzate. In sintesi, è un tipo di servizio in cloud la cui caratteristica principale è la possibilità di avviare più istanze di macchine virtuali, da poter adattare dinamicamente alle nostre esigenze.

L'**Amazon Elastic MapReduce**, **EMR**, è un servizio web che consente l'elaborazione di grandi quantità di dati, sull'ordine di grandezza dei petabyte[11], O(PB). Sfrutta il framework open source Hadoop poggiato direttamente sulle infrastrutture di Amazon EC2 e Amazon S3[12]. L'EMR prende il suo nome da due funzioni, seppur in questo caso abbiano scopi diversi, della programmazione funzionale[13]: map() e reduce().

Per spiegare come funziona in generale un modello di MapReduce, useremo le parole di Michael Stonebraker[14]: "Il programma map legge un insieme di record da un file di input, svolge le operazioni di filtraggio e le trasformazioni desiderate, quindi produce una serie di record di output nella forma convenuta (chiave, dati). Mentre il programma map produce questi record, una funzione separata li partiziona in multipli e indipendenti contenitori applicando una funzione alla chiave di ciascun record. Questa funzione è tipicamente hash, sebbene sia sufficiente qualsiasi tipo di funzione deterministica. Una volta che il contenitore è pieno, il suo contenuto viene riversato su disco. Infine il programma map termina producendo una serie di file di output, uno per ciascun contenitore. Dopo essere stati raccolti dal framework MapReduce i record di input vengono raggruppati per chiavi (attraverso operazioni di sorting o hashing) e sottoposti al programma reduce. Come per il programma map, reduce esegue una elaborazione arbitraria attraverso un linguaggio general purpose. Di conseguenza può compiere qualsiasi sorta di operazioni sui record. Per esempio può elaborare alcune funzioni addizionali per altri campi dati del record. Ciascuna istanza reduce può scrivere record a

[11] Petabyte: dall'inglese byte e dal greco Penta, ossia cinque, ad indicare la quinta potenza di mille, ossia 10^15 byte. 1 PB corrisponde, anche se in maniera imprecisa, ad 1 biliardo di byte.
[12] Amazon Simple Storage Service (S3) è il servizio di storage on line di Amazon. Verrà trattato successivamente.
[13] La programmazione funzionale è un paradigma di programmazione. A differenza della programmazione procedurale, basata appunto sull'esecuzione di procedure, la funzionale tratta funzioni, espressioni.
[14] Micheal Ralph Stonebraker (11 ottobre 1943) è uno scienziato informatico specializzato nella ricerca dei database.

un file di output e quest'ultimo rappresenta una parte della risposta soddisfatta da una elaborazione MapReduce"[15]. Il framework Hadoop sui cui è basato l'EMR è un modello ottimizzato e volto alla parallelizzazione dei dati e della loro elaborazione. Consta fondamentalmente di due blocchi principali:

- il file system distribuito Hadoop HDFS, in cui i file sono divisi in blocchi, da ben 64Mb, distribuiti tra diversi nodi del cluster[16]. Anche in questo caso, è ricercata una certa ridondanza dei dati, al fine di sopperire ad eventuali perdite degli stessi.
- Il framework MapReduce, adibito alla trattazione di insiemi di dati di grandi dimensioni, in logica parallela e quindi sfruttando nodi e cluster.

In EMR, l'infrastruttura viene definita elastica perché possibile dimensionare istantaneamente, in negativo e in positivo, la capacità di elaborazione e la banda disponibile a rete e dischi, il tutto con l'aggiunta o la sottrazione di nodi.

Uno dei servizi web fondamentali offerti da AWS è la **Route 53**, ossia il *Domain Name System* (DNS) di Amazon. Un DNS non fa altro che tradurre letteralmente i comuni indirizzi internet che inseriamo e leggiamo nei nostri browser, come www.Amazon.com, con gli IP numerici, come 192.0.1.1, che sono invece usati dai computer per comunicare tra loro. Nel caso specifico, Route 53 affida all'utente la gestione DNS dei propri siti permettendo anche la creazione di eventuali sottodomini associati a diversi web server, che possono essere anche diversi da quelli offerti da Amazon stessa. In pratica serve le richieste utente sia che siano indirizzate a infrastrutture "interne" AWS, come EC2 e S3, sia che siano indirizzate "esternamente" all'universo AWS. Route 53 risponde alle richieste di traduzione DNS con bassa latenza, sfruttando la rete globale di server DNS. La velocità di risposta viene assicurata dall'inoltro della richiesta al server DNS più vicino. La gestione dei record DNS può essere eseguita sia adoperando la Console AWS che le API[17] di facile gestione. Come per altri servizi di AWS, non ci sono contratti vincolanti a lungo termine o requisiti minimi per l'utilizzo di Route 53, si paga solo per:

[15] Che cos'è MapReduce? - http://www.technologytransfer.it/?cis=4;1&rec=80&yy=2009&mm=11
[16] Il cluster è un'unita o un raggruppamento logico di settori contigui di un hard disk.
[17] Application Programming Interface – un insieme di procedure apposite per programmare l'applicativo per cui sono sviluppate.

- la gestione dei domini tramite il servizio,
- il numero di richieste servite.

L'**Amazon Virtual Private Cloud**, **VPC**, permette di creare una sezione privata e isolata di cloud AWS, all'interno della quale posso eseguire servizi AWS su una rete virtuale definita dall'utente stesso. In parole povere, tramite una VPN, *Virtual Private Network*, si possono espandere le capacità e le funzionalità della propria azienda, connettendo l'infrastruttura interna IT aziendale già esistente con le risorse AWS. Anzi, creando una connessione hardware VPN tra il data center aziendale e la VPC, si sfrutta il cloud AWS come un'estensione vera e propria del proprio data center.

In verità sono due le possibilità che conferiscono maggior flessibilità alla connessione VPN:

1. la già citata connessione hardware VPN alla VPC tramite routing statico, ossia tramite la configurazione manuale, dell'amministratore, delle entries delle tabelle di routing,
2. la configurazione della propagazione automatica delle route dalle proprie VPN e dai link al Direct Connect alle tabelle di route della propria VPC, con un grosso guadagno in termini di semplicità avendo rimosso del tutto la necessità del routing statico.

La VPC offre, tra i vari servizi, il controllo completo sulla propria VPN, incluse:

- la selezione del range dello spazio degli indirizzi degli IP,
- la creazione di sottoreti, private e non accessibili dall'esterno,
- la configurazione delle tabelle di route e dei gateway di rete.

2.2.2 Deployment & Management

La survey riguardo il Deployment & Management ha inizio dalla **CloudFormation**.

Si tratta di un servizio che offre un metodo semplice per creare e gestire configurazioni di istanze di EC2, senza doverle ricreare ogni volta e con risparmio di tempo, sulla falsa riga di una Macro[18].

La CloudFormation mette a disposizione dei template di esempio oppure permette di crearne dei propri, liberando sviluppatori e amministratori di sistema da alcune problematiche come l'ordine in cui i servizi devono essere eseguiti oppure i parametri necessari a runtime per eseguire le proprie applicazioni, occupandosene direttamente la CloudFormation. Definendo un certo template per un determinato tipo di istanze non sarà più necessario configurarne lo specifico sistema operativo, gli applicativi da installare e gli eventuali servizi richiesti. E' bene specificare che i template possono essere succesivamente aggiornati e quindi non si tratta di oggetti bloccati. Sviluppo e update dei template possono essere eseguiti direttamente dalla ormai nota Console AWS, oppure sfruttando tool di sviluppo adatti o ancora con applicativi basati sulle apposite API. È un servizio che non comporta costi aggiuntivi: si pagano infatti solo le risorse AWS necessarie per eseguire le applicazioni.

Il **CloudWatch**, come suggerisce lo stesso nome, è un servizio di monitoraggio completo delle risorse AWS. E' uno strumento completo che permette di monitorare sia le istanze di EC2 che quelle di DB di Amazon RDS[19]. È facile intuire come questo prodotto sia orientato principalmente agli amministratori di sistema, il cui scopo principale è valutare appunto le performance del sistema in esecuzione. Questo strumento infatti permette di monitorare i dati e raccogliere le informazioni sullo stato del sistema, valutando in tempo reale la sensibilità dello stesso, ossia la risposta ad eventuali carichi problematici, e di conseguenza consente all'utilizzatore di sviluppare soluzioni atte a superare eventuali criticità e a far eseguire gli applicativi senza interruzioni o rallentamenti. Le metriche di controllo rappresentano certamente uno dei punti di forza di questo strumento, infatti oltre le metriche standard predefinite che ci vengono offerte dal CloudWatch, è possibile,

[18] La Macro può essere definita come un agglomerato di istruzioni, di passi, per automatizzare delle operazioni ripetitive.
[19] Amazon Relational Database Service. Verrà trattato successivamente.

tramite Console AWS, API o tool esterni svilupparne di proprie, personalizzando profondamente la risorsa: ad esempio sarà possibile attivare determinate metriche solo in alcuni periodi temporali o magari solo oltre determinate soglie di carico. Infine, questa particolare risorsa AWS consente di visualizzare grafici dei dati di monitoraggio, impostare allarmi per determinati eventi e analizzare le tendenze per agire in automatico, in base allo stato del proprio ambiente cloud.

Se si vogliono bypassare le problematiche relative alle istanze virtuali, quali possono essere configurazione e avvio della stessa, e dedicarsi unicamente alla propria applicazione, allora **Elastic Beanstalk** è il servizio AWS adatto. Basterà semplicemente caricare l'applicazione in cloud e ogni aspetto dei dati sarà gestito automaticamente, dal load balancing all'auto scaling, pur lasciando all'utente il pieno controllo su tutte le risorse AWS necessarie all'esecuzione. L'Elastic Beanstalk si basa sui servizi Amazon EC2, S3 e SNS[20]. In questo momento supporta i seguenti linguaggi di programmazione:

- .NET con supporto IIS 7.5
- PHP e Python con supporto Apache HTTP Server
- Java con supporto Apache Tomcat.

Anche questo prodotto non comporta costi aggiuntivi, in quanto si paga unicamente per le risorse AWS necessarie a memorizzare ed eseguire l'applicazione.

AWS Identity and Access Management, o in breve **IAM**, è il servizio che si occupa della concessione dell'accesso alle risorse AWS, e relativo utilizzo, ai propri utenti. La vera innovazione risiede nella possibilità di concedere l'accesso alle risorse condivise anche a utenze non gestite con AWS. Inoltre è presente una funzionalità che permette un merge, una fusione, tra le rubriche AWS e quelle aziendali, ossia di sfruttare le credenziali aziendali senza crearne di nuove per l'accesso ai servizi AWS.

[20] Amazon Simple Notification Service. Verrà trattato successivamente.

2.2.3 Storage & Content Delivery

L'**Amazon S3**, anche noto come **Amazon Simple Storage Service** è il servizio di storage online, in cloud, tramite il quale gli utenti archiviano i propri dati, da qualsiasi punto della rete e in qualsiasi momento. Garantisce agli utenti l'accessibilità alla stessa infrastruttura usata da Amazon per gestire la propria rete globale di siti web, caratterizzata da alta scalabilità, velocità e sicurezza. È necessario introdurre due dei componenti fondamentali gestiti dall'S3:

- Buckets,
- Objects.

I Buckets possono essere definiti come una cartella principale, all'interno della quale è possibile memorizzare un numero teoricamente infinito di oggetti, che possono andare dai semplici file di testo alle immagini e video. Seppure il numero di oggetti possa essere molto alto, questi non possono avere una dimensione superiore ai 5 GB. Per ogni account sono consentiti 100 buckets e ogni bucket deve avere un nome univoco. Si noti che parlare di cartella, in realtà, è improprio in quanto il servizio S3 non è un file system gerarchico e quindi non sarebbe applicabile il concetto di cartella, ma viene comunque simulato e gestito come tale. Il bucket va memorizzato in una delle 7 regioni disponibili su Amazon, seguendo i criteri base di bassa latenza e maggiore vicinanza logistica.

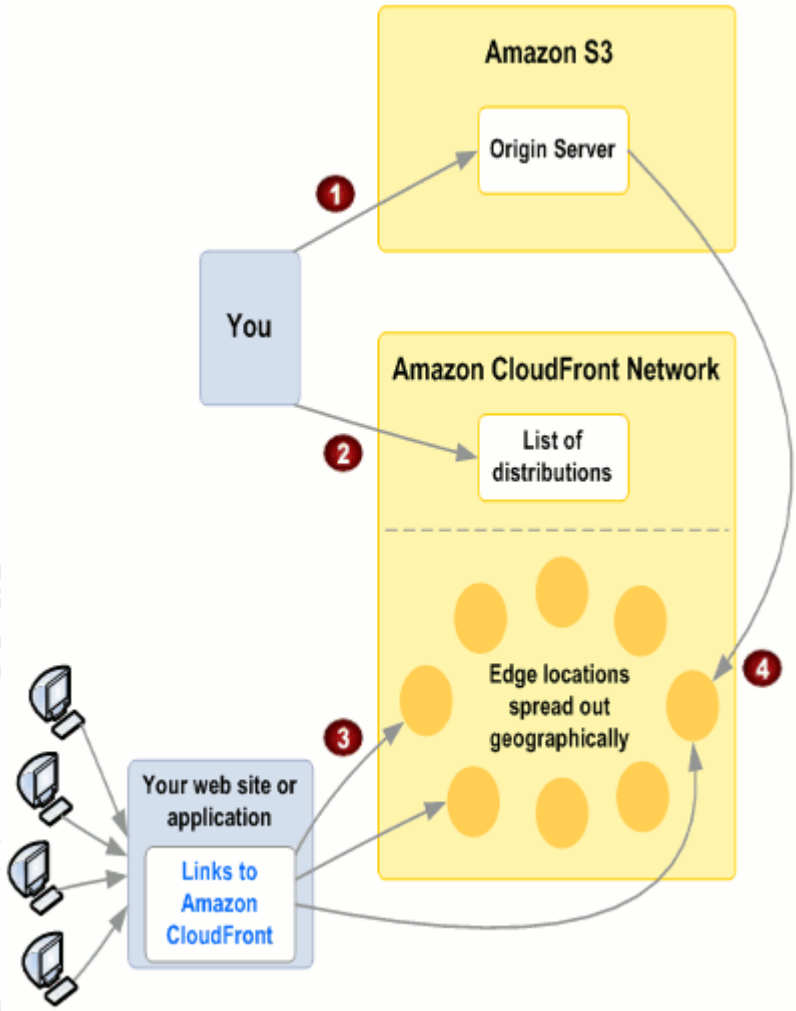
Gli Objects sono le entità minime fondamentali memorizzate in S3. Ogni oggetto è formato da due parti: la prima parte sono i dati veri e propri, in forma binaria, mentre la seconda parte sono i metadati, formati dalla coppia (chiave, valore), che forniscono informazioni sui dati stessi. Avendo specificato che ogni bucket avrà un nome univoco, anche ogni object S3 memorizzato sarà individuabile da un URL univoco del tipo:

*http://NOME_BUCKET.s3.amazonaws.com/NOME_OGGETTO.** .

È facile intuire che un object memorizzato in una regione non sarà mai replicato in un'altra regione, a meno che non sia lo stesso utente a trasferirlo manualmente. L'accesso agli objects è sottoposto ad un severo sistema di controllo, e gli stessi objects sono configurabili come privati o pubblici o "dedicati" ad uno specifico utente di AWS.

Il servizio S3 è dotato di un'opzione particolare, la *Reduced Redundancy Storage* (*RRS*), secondo la quale, accettando una rischio di perdita media annua dello 0,01% degli objects, si ottiene una notevole riduzione dei costi. Questa opzione è stata pensata qualora si voglia (o si debba) memorizzare dati non critici o facilmente recuperabili, in caso di perdita, da altre fonti.

Il **CloudFront** è il servizio dedicato al *content distribution*, ossia la distribuzione di contenuto statico o in streaming, con bassa latenza e alte velocità di trasferimento. È uno dei servizi in maggiore espansione come testimoniano i numerosi webinar[21] che Amazon terrà nei prossimi mesi. La configurazione è rapida e semplice in quanto collegata al servizio S3, infatti basta associare l'object, memorizzato in uno dei propri bucket, al CloudFront ed in automatico questo verrà replicato nelle varie edge locations sparse nel mondo, che verranno sfruttate per far fronte alle richieste geograficamente più vicine.



[21] Seminari e corsi tenuti online.

L'**Amazon Glacier** è un servizio di storage, a costi estremamente bassi, che offre un deposito sicuro e durevole per l'archiviazione dei dati e il loro backup. Al fine di mantenere bassi i costi, Amazon Glacier è ottimizzato per i dati che vengono usati raramente e per i quali i tempi di recupero, anche di diverse ore, sono ammissibili. Con questo servizio, gli utenti possono archiviare grandi o piccole quantità di dati, in modo affidabile, per un minimo di 0,01 dollari per GB/mese, con un notevole risparmio rispetto alle soluzioni on-premise. Il problema principale è che, solitamente, le aziende pagano oltre il dovuto per l'archiviazione dei dati, a causa di due motivi in particolare.

Il primo motivo è che sono costrette a effettuare un costoso pagamento anticipato per la loro soluzione di archiviazione (che non include il costo continuo per le spese operative come l'energia, le strutture, di personale e di manutenzione).

Il secondo motivo è che dal momento che le imprese devono cercare di prevedere di quanta capacità ci sarà bisogno per i propri dati, essi comprensibilmente sovrastimano queste quantità per essere sicuri di avere abbastanza capacità per la ridondanza dei dati e sopperire ad una eventuale crescita inaspettata.

Questo insieme di situazioni genera spreco di denaro e grossi quantitativi di capacità non utilizzata. Amazon Glacier ribalta e risolve queste problematiche con nuove regole: nessun pagamento anticipato e soprattutto costi inferiori dovuti alla rapida scalabilità, verso l'altro o il basso, del sistema che supera il problema della stima della capacità necessaria futura, basandosi sul paradigma "pay-per-use", ossia paghi solo ciò che usi.

Lo **Storage Gateway** è uno strumento rivolto principalmente alle azienda, in quanto si tratta di un software on-premise, quindi da installare in locale, rivolto a fornire una perfetta integrazione tra lo storage fisico di un'azienda e l'infrastruttura AWS di storage. Questo servizio, che supporta i protocolli standard per lo storage aziendale, permette quindi di interfacciarsi con qualsiasi infrastruttura hardware e software esistente in locale al fine di eseguire una copia fisica crittografata dei dati nello storage S3, il tutto con performance a bassa latenza ottenute mantenendo un accesso frequente ai dati on-premise.

Lo Storage Gateway supporta due configurazioni:

1. **Gateway-Cached Volumes**: i dati critici e primari vengono immagazzinati in S3, mentre i dati ad accesso frequente risiederanno solo in locale. In questo modo si risparmiano i costi per la sicurezza e lo storage dei dati importanti, viene minimizzata la necessità di uno scaling[22] per lo storage on-premise e per questi ultimi si riesce ad ottenere anche una latenza molto bassa, fondamentale per i dati ad accesso frequente.
2. **Gatewat-Stored Volumes**: nel caso in cui la bassa latenza sia richiesta per accedere all'intero storage, ossia a tutti i dati, anche i primari saranno memorizzati in locale, mentre in S3 verranno eseguiti, in maniera asincrona, dei backup off-site conservati per un determinato periodo. Questi backup di dati primari possono essere poi recuperati localmente oppure da servizi come EC2, se ad esempio ci fosse un *disaster recovery* in seguito ad un disastro, appunto.

2.2.4 App Services

L'Amazon **CloudSearch**, come suggerisce il nome stesso, è un servizio di ricerca di contenuti all'intero del nostro ambiente cloud. È la stesso servizio su cui è basata la ricerca dei prodotti commerciali su Amazon.com e la sua configurazione è completamente affidata all'utente che può creare dei propri domini di ricerca o può affidarsi ai settaggi di base. Il CloudSearch scala continuamente così da adattarsi ai fattori altamente variabili come possono essere la quantità di dati tra cui cercare o il numero di query in esecuzione, e gli sviluppatori possono modificare i parametri di ricerca e applicare le nuove impostazioni in qualsiasi momento senza dover caricare nuovamente i dati.
Il tutto eseguibile direttamente in cloud svincolando il cliente da appositi software o piattaforme locali.

Il servizio **SES**, **Simple Email Service**, di Amazon è un servizio di mailing divincolato dalla necessità di server dedicati, con conseguente risparmio economico, e rivolto in special modo al settore industriale commerciale permettendo di gestire importanti

[22] Il dimensionamento della capacità.

quantitativi di posta elettronica, come può essere necessario nelle campagne di marketing. In effetti, l'invio di mail su larga scala comporta una serie di problematiche, architetturali e non, di cui tenere conto. Il problema fondamentale da affrontare è quello di massimizzare la ricezione della posta inviata, problema, questo, che trae giovamento dai rigorosi standard degli ISP[23] per il contenuto delle mail. SES agisce in questo senso con dei filtri a monte che garantiscano la conformità agli standard degli ISP. Per aiutare le aziende a migliorare ulteriormente la qualità delle comunicazioni via email con i propri clienti, SES offre un *built-in loop*[24] di feedback, che include le notifiche delle email rinviate e fallite, i tentativi di consegna riusciti con successo e le denunce di spam.

Un altro servizio che sfrutta le email è il **Simple Notifications Service**, **SNS**. Questo servizio, di facile installazione e gestione, permette l'invio di notifiche *push*[25] su diversi protocolli, come SMS, HTTP/HTTPS ed email. Il funzionamento è semplice: si crea un topic, un argomento o un evento, che si voglia sia seguito, e questo funzionerà da punto di accesso per la pubblicazione di messaggi, permettendo ai clienti interessati di sottoscriverne le notifiche. Il creatore del topic ne definisce le varie caratteristiche come il tipo di notifiche che si vuole generare, i protocolli, anche diversi per lo stesso topic, sui quali possono essere trasmesse, la tipologia di clienti che possono sottoscrivere le notifiche e il trigger che le genera. I sottoscrittori del topic possono, a loro volta, anche essere stati aggiunti direttamente dal creatore del topic e hanno possibilità di scegliere il tipo di notifica e il protocollo di invio ossia l'*end-point*, come un numero di telefono, un URL o un indirizzo email. Quando verrà inserito un nuovo messaggio nel topic o si presenta un aggiornamento che ne scateni il trigger, la notifica sarà instradata verso tutti i clienti sottoscrittori e sottoscritti.

Invece il **Simple Queue Service**, **SQS**, si occupa della distribuzione del carico operativo su diverse componenti tramite code di messaggi, facendo comunicare in questo modo

[23] Internet Service Provider sono i fornitori commerciali dei servizi internet.
[24] Si intende un sistema circolare di feedback interno.
[25] È un meccanismo che libera l'utente dalla necessità di controllare periodicamente la presenza di nuove informazioni, aggiornamenti o notizie.

anche i vari servizi di AWS. Gli sviluppatori, infatti, posso semplicemente muovere i dati tra le componenti distribuite dei propri applicativi che svolgono diversi compiti, senza perdita di messaggi o necessità che ogni componente sia sempre disponibile. Il numero di code che si possono creare così come il numero di messaggi è illimitato e possono essere create in 8 delle 9 regioni, in quanto è esclusa la GovCloud[26]. Il corpo del messaggio contiene fino a 64 KB di testo in qualsiasi formato e i messaggi possono essere gestiti in blocchi da 10 messaggi o 64 KB e possono essere letti e inviati simultaneamente, ma rimarranno per 14 giorni al massimo. All'atto della ricezione, il messaggio diventa "*locked*" durante l'elaborazione, impedendo modifiche in contemporanea da parte di altre componenti. Se l'esito dell'elaborazione è negativo, il messaggio torna disponibile, mentre se c'è bisogno di maggior tempo di elaborazione si può andarlo a modificare dinamicamente. Gli sviluppatori possono condividere le proprie code con altri account AWS, anche anonimamente e la condivisione può essere limitata in base a varie opzioni, come l'ora, il giorno e l'indirizzo IP.

L'Amazon **SWF**, ossia **Simple WorkFlow Service**, è il servizio che si occupa del workflow, il flusso di lavoro, per la creazione di applicazioni. Questo servizio infaatti coordina in maniera affidabile tutte le fasi della lavorazione di un'applicazione, indifferentemente dal fatto che si tratti di processi business, applicazioni assicurative oppure di applicativi per l'analisi strutturale dei dati. Con SWF, gli sviluppatori strutturano le fasi di lavorazione in *tasks*, che lavoreranno come applicazioni distribuite, lasciando al servizio l'onere di gestire i problemi di concorrenza e di pianificazione in base alla configurazione settata dall'utente. In pratica, il SWF funge da *hub*[27] di coordinamento per le diverse fasi di sviluppo:

1. il mantenimento dello stato dell'applicazione,
2. monitoraggio delle esecuzioni del workflow e registrazione dei progressi,
3. holding e dispatching dei tasks,
4. verifica che ogni task della applicazione sia stato eseguito.

[26] La GovCloud è la regione riservata alle attività governative statunitensi.
[27] L'hub è un dispositivo multiporta che riceve da una porta delle informazioni per ridistribuirle sulle altre.

2.2.5 Database

Il primo servizio analizzato è il **DynamoDB**, che si occupa della gestione completa di un database *NoSQL*[28] in cloud. Il DynamoDB diffonde il traffico di dati della tabella su un numero di server sufficiente a gestire la capacità richiesta dall'utente e la quantità di dati memorizzati, mantenendo sempre consistenza e alta velocità. Tutti questi dati sono immagazzinati su dischi a stato solido (SSD) e sono automaticamente replicati su 3 availability zones, li dove presenti, di una regione per fornire alta disponibilità e durata nel tempo dei dati. Questo meccanismo di protezione assicura i dati contro i guasti di una singola macchina o di interi impianti. Essendo slegato dalle regole relazionali, il DynamoDB non ha uno schema fisso ed in effetti ogni elemento del database può avere un importante numero di attributi e ospitare diversi tipi di dati, dalle stringhe ai numeri passando per i dati binari. Per rinforzare la sicurezza di questo servizio, sono inoltre usati collaudati metodi di crittografia per l'autenticazione degli utenti, volti a prevenire accessi non autorizzati. In tal senso, l'integrazione con il servizio IAM[29] è una garanzia ulteriore. Un ulteriore esempio di integrazione da evidenziare è quello con l'EMR[30] che permette di eseguire complesse analisi su interi set di *big data*.

L'**ElastiCache** è un servizio relativamente nuovo essendo ancora in fase beta. Si occupa di realizzare dei cluster di cache distribuiti in RAM. Viene utilizzato in particolar modo, in ambito web dinamico, per velocizzare i dati con accesso più frequente servendoli direttamente dalla RAM invece che dai database, che sono più lenti essendo basati su dischi. È possibile quindi realizzare un cluster di cache costituito da un insieme di nodi di cache, aggiungendo o sottraendo i quali, si scala rapidamente il sistema per soddisfare le esigenze del carico di lavoro. Potrebbe capitare che la creazione di alcuni nodi non avvenga con successo, provocando sovraccarichi nel database con conseguenti rallentamenti o crash delle applicazioni che si poggiano su di esso, ma il servizio è

[28] I NoSQL sono una famiglia di database che non segue i principi dei database relazionali, in quanto non usano SQL come linguaggio di gestione e non adottano lo schema tabella-relazione.
[29] Identity and Access Management è il servizio Amazon che si occupa dell'autenticazione e del controllo utente.
[30] Elastic MapReduce è il servizio Amazon che si occupa dell'elaborazione di grandi quantità di dati.

progettato per rilevare e sostituire automaticamente i nodi della cache non riusciti. Sfruttando il CloudWatch[31], inoltre, si può migliorare la visibilità, e quindi lo studio, delle metriche di performance relative ai nodi. Uno dei punti di forza di questo servizio è la completa compatibilità con *Memcached*[32], grazie alla quale è possibile utilizzare, senza problemi, applicativi basati su questo sistema o anche una migraare in maniera completa ad ElastiCache appunto.

Last but not least, è il servizio **Amazon RDS**, ossia **Relational Database Service**. RDS fornisce una gestione semplice ma efficace dei database SQL, attraverso le funzionalità dei motori per database più comuni come MySQL, Oracle o Microsoft SQL Server. Grazie a ciò, è possibile continuare ad utilizzare codice, applicazioni e strumenti, usati per i precedenti database, anche con RDS stesso. Il servizio in questione esegue automaticamente un backup del database dal quale si sta migrando, lasciando all'utente la possibilità di scegliere il periodo di tempo per il quale conservare i backup e stabilire i punti di ripristino e il loro timing. In ogni momento, poi, sarà possibile scalare alcuni parametri fondamentali, come la capacità di archiviazione o le risorse di elaborazione, associati ad un'istanza di database tramite una singola chiamata di API. È anche possibile usare una replica per migliorare affidabilità e disponibilità in caso di sovraccarichi.
Le istanze di database possono essere rifornite con uno storage standard o con archiviazione IOPS[33].

[31] È il servizio Amazon adibito al monitoraggio delle risorse e alla valutazione delle tendenze.
[32] Memcached è un sistema di caching in RAM nato nel 2003 dalla Danga Interactive. A differenza di ElastiCache non ha un sistema di gestione automatizzato e può quindi risultarne più complesso l'utilizzo.
[33] I/O Operations Per Second è una misura delle performance usata per gli hard disk e i drive allo stato solido.

CAPITOLO 3

EC2: Configurazione di un'istanza di macchina virtuale

3.1 Amazon Elastic Compute Cloud (EC2)

Il servizio EC2 è già stato introdotto nel precedente capitolo, ora invece si vogliono approfondire e analizzare alcune sue funzionalità, evidenziandole in maniera tecnica e soprattutto pratica, instanziando una macchina virtuale direttamente dalla Console AWS.

EC2 viene presentato nel 2006 come primo servizio di Cloud Computing su modello *HaaS*, *Hardware as a Service*, ossia un servizio on-demand basato sull'offerta di hardware tramite un pagamento commisurato all'utilizzo e non più vincolato da onerosi contratti di media e lunga durata. Basandosi sulla notevole esperienza mutuata dal settore commerciale, lancia un servizio che di fatto altro non è che un *VPS*, un *Virtual Private Server*, dando così all'utente la sensazione di avere un proprio server, con tanto di RAM, CPU e quant'altro, il tutto tramite la tecnologia di virtualizzazione open-source *Xen*[34]. I vantaggi principali offerti sono quelli che riguardano la velocità e la semplicità della configurazione di un'istanza e la gestione da diverse piattaforme, come possono essere la proprietaria Console AWS o altri IDE che sfruttino le API messe a disposizione da Amazon.

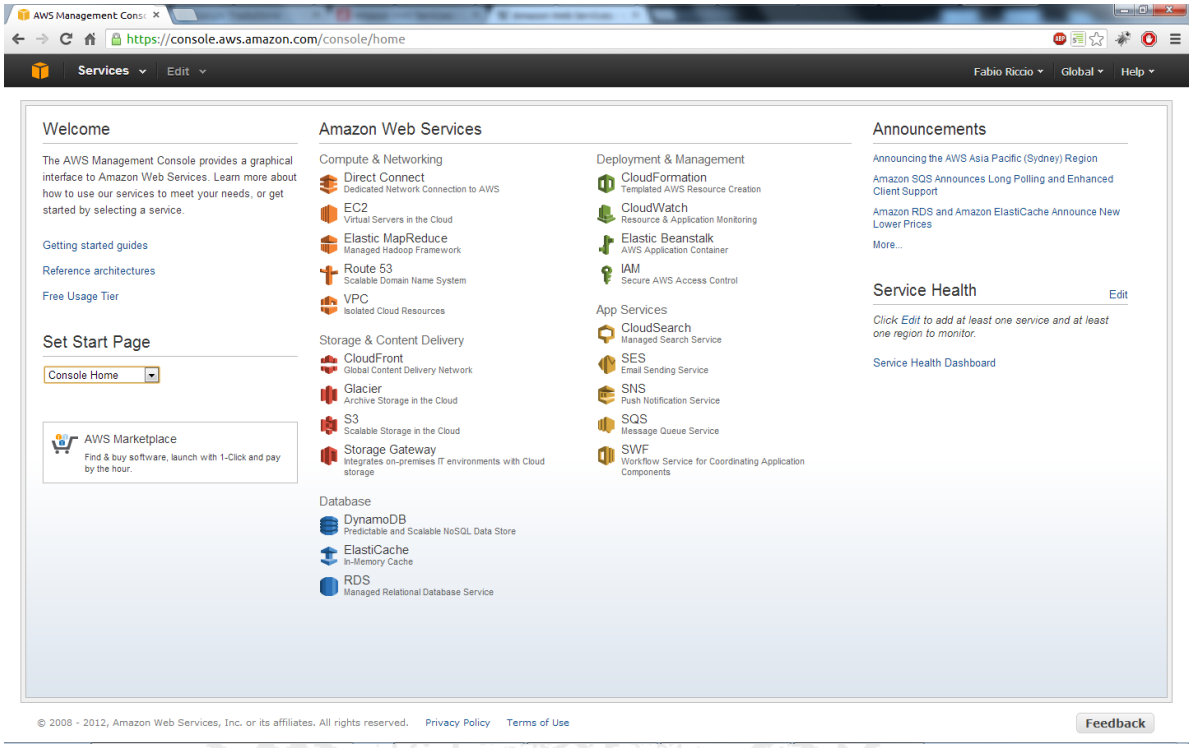
[34] Tecnologia open-source nata e sviluppatasi a Cambridge, che tramite l'installazione di un leggero software chiamato Hypervisor consente ad ogni server fisico l'esecuzione di diversi server virtuali.

3.2 Istanza di una macchina virtuale

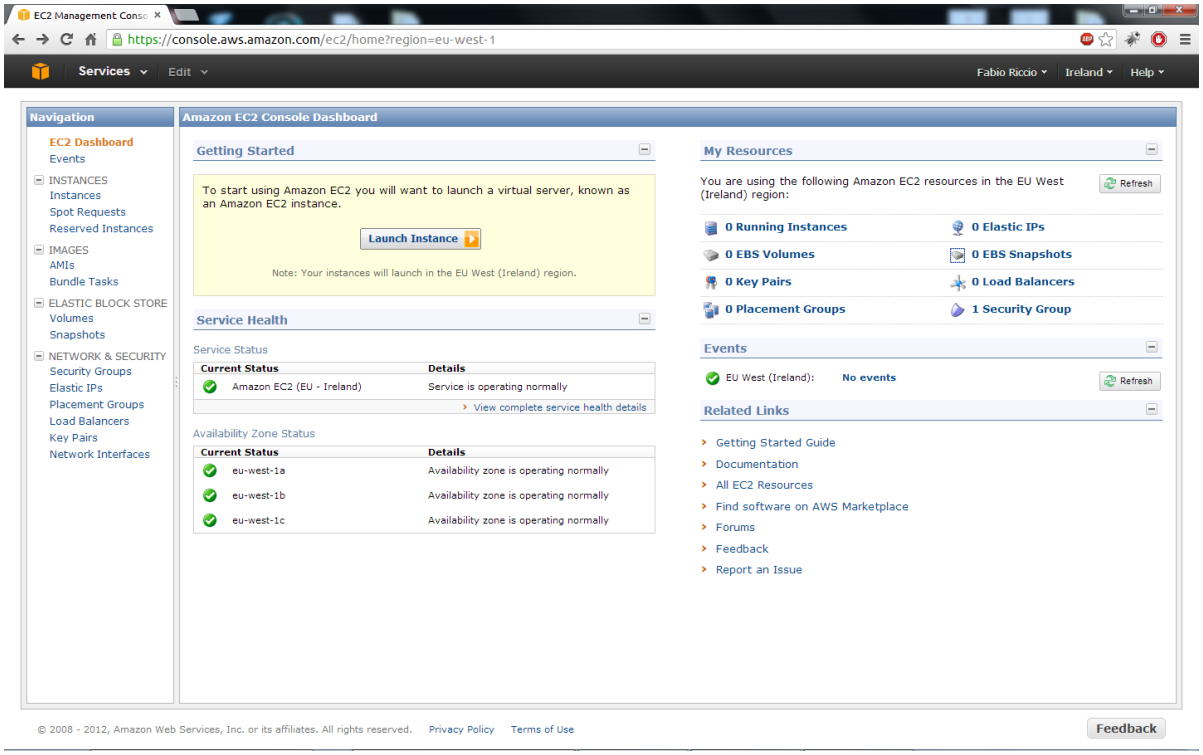
Si procederà ora, step by step, alla creazione di un'istanza di una macchina virtuale, illustrandone i vari passaggi.

Amazon AWS mette a disposizione *Free Usage Tier*, ossia un account gratuito a funzionalità limitate per un anno, per permettere di studiare e valutare il tipo di servizi offerti, il bilanciamento dei carichi, le capacità, le criticità di cui si necessità e tanto altro, per stabilire con quali parametri lavorare in futuro, sia dal punto di vista tecnico che economico. L'accesso al sito richiede una registrazione e l'immissione di una carta di credito. Quest'ultima è necessaria anche in caso di Free Usage Tier, nel caso in cui malauguratamente si dovessero sforare le limitazioni imposte dal servizio.

Una volta loggati, la AWS Console si presenterà così:



Cliccando su EC2, si viene ridiretti su un'altra console ricca di opzioni per la gestione delle istanze:



A parte l'evidente pulsante di Launch Instance, si può notare in alto a destra la regione in cui stiamo operando, con il sommario dei relativi oggetti e servizi attivi. Per ora è presente solo 1 Security Group, settato di default dal sistema. Dal menu a tendina, si può selezionare una regione diversa dalla attuale, in questo caso l'Irlanda. In basso a sinistra invece sono indicate le Availabilty Zones e i loro relativi status, in questo caso tutte le zone funzionano correttamente e non ci sono disservizi. All'estrema sinistra è invece presente un menu che presenta alcune opzioni di configurazione da analizzare a breve.

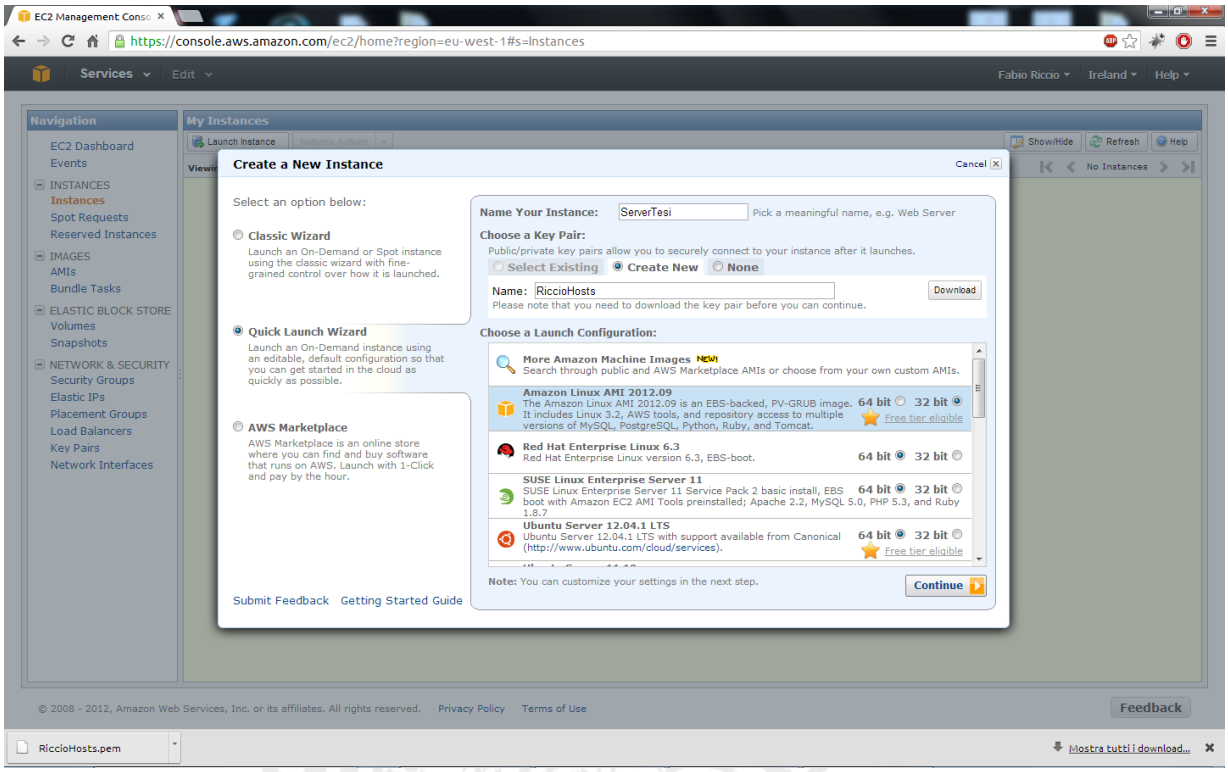
Prima di procedere nel lancio di un'istanza è doveroso spiegare cosa sia un **AMI**, **Amazon Machine Image**.

Si tratta di un'immagine preconfigurata di macchina virtuale con tanto di sistema operativo, selezionabile, nativa di Amazon nel senso che è studiata e adattata proprio per eseguire nell'ambiente virtualizzato di AWS. L'immagine, inoltre, può essere arricchita di

ulteriore software tramite il Marketplace. Si evidenzia quindi che lanciare un'istanza equivale a creare un server virtuale.

Lanciando un'istanza, si presenta un pannello con 3 opzioni per effettuare una creazione classica, una rapida o attingere a software dal nuovo AWS *Marketplace*. In modalità classica si procederà con una configurazione ex novo dell'istanza, con la rapida invece si scelgono istanze già pronte e nel Marketplace invece si possono acquistare AMI di terze parti, gratuite o a pagamento. Nel wizard vanno inoltre specificati il nome dell'istanza e create le *key pair*, una coppia di chiavi per l'accesso all'istanza, da scaricare e conservare altrimenti il wizard stesso impedisce la prosecuzione nel processo di creazione. Si parla di coppia di chiavi in quanto è data una chiave pubblica ed un privata, quella da scaricare appunto, la cui comodità è data dalla possibilità di loggarsi tramite il solo username senza fornrie alcuna password.

Come si nota dall'immagine, si è scelta una AMI Linux a 32 bit.



Dopo una schermata riepilogativa, si viene poi ricondotti ad una pagina in cui si possono gestire degli allarmi che diano avviso su eventuali guasti o insuccessi dell'istanza, creare dei volumi *EBS*, ritornare alle proprie istanze o rilasciare un feedback.

EBS, o **Elastic Block Storage**, è un servizio che fornisce volumi, da 1GB a 1TB, persistenti di storage per le istanze EC2. Si parla di persistenza per evidenziare la differenza con lo storage nativo di un'istanza, che invece al momento dello spegnimento dell'istanza stessa viene perso.

Un volume EBS presenta concettualmente numerose analogie con un normale hard disk, infatti può essere formattato secondo diversi tipi di file system, viene visto dalla macchina virtuale come un nuovo disco locale e ne possono essere collegati diversi in gruppo all'interno della stessa istanza. Ma è bene ricordare che non si tratta effettivamente di hard disk dato che sono soggetti a possibili e rapidi dimensionamenti legati alla scalabilità del sistema o possono essere separati da un'istanza per essere collegati ad un'altra. Ogni volume viene creato in una determinata availabilty zone ed essere collegato solo a istanze presenti nella stessa zona. Per evitare perdite dovute ad un qualsiasi tipo di guasto fisico, ogni volume viene replicato all'interno della zona in cui è stato creato. Inoltre EBS consente di creare degli snapshot di recupero, mantenuti in S3[35], che possono essere utilizzati sia per una recovery in caso di criticità che per creare rapidamente nuovi volumi EBS.

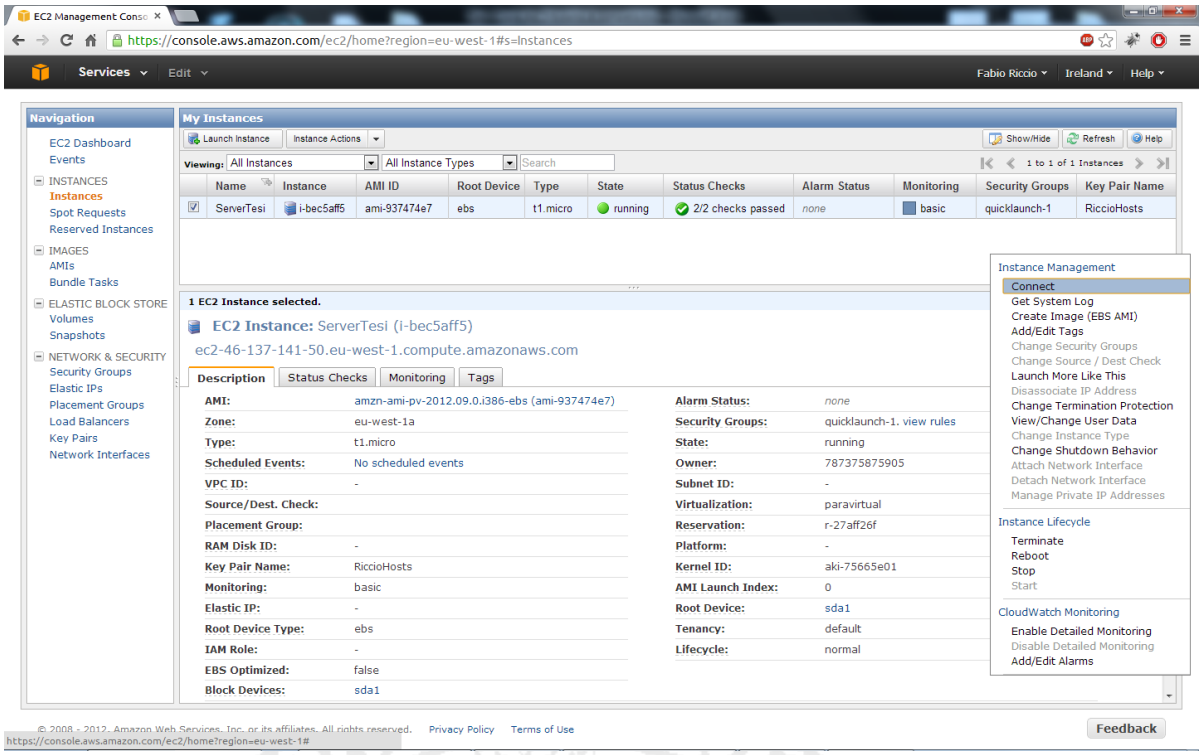
Una volta terminato il processo di creazione, si torna nella dashboard dove ora è visibile l'istanza generata ed è possibile eseguire una serie di operazioni come analisi approfondite di traffico e carico o ancora generare altre istanze.

Sfogliando il menu sulla sinistra si può anche decidere ad esempio di aggiungere volumi EBS, oppure, nel caso ne siano già presenti altri, effettuare o gestire degli snapshot degli stessi.

Possiamo decidere di creare dei nuovi *security group* o cambiare le *policies*[36] dei gruppi

[35] È il servizio di storage online di AWS.
[36] Le policies sono le linee di condotta.

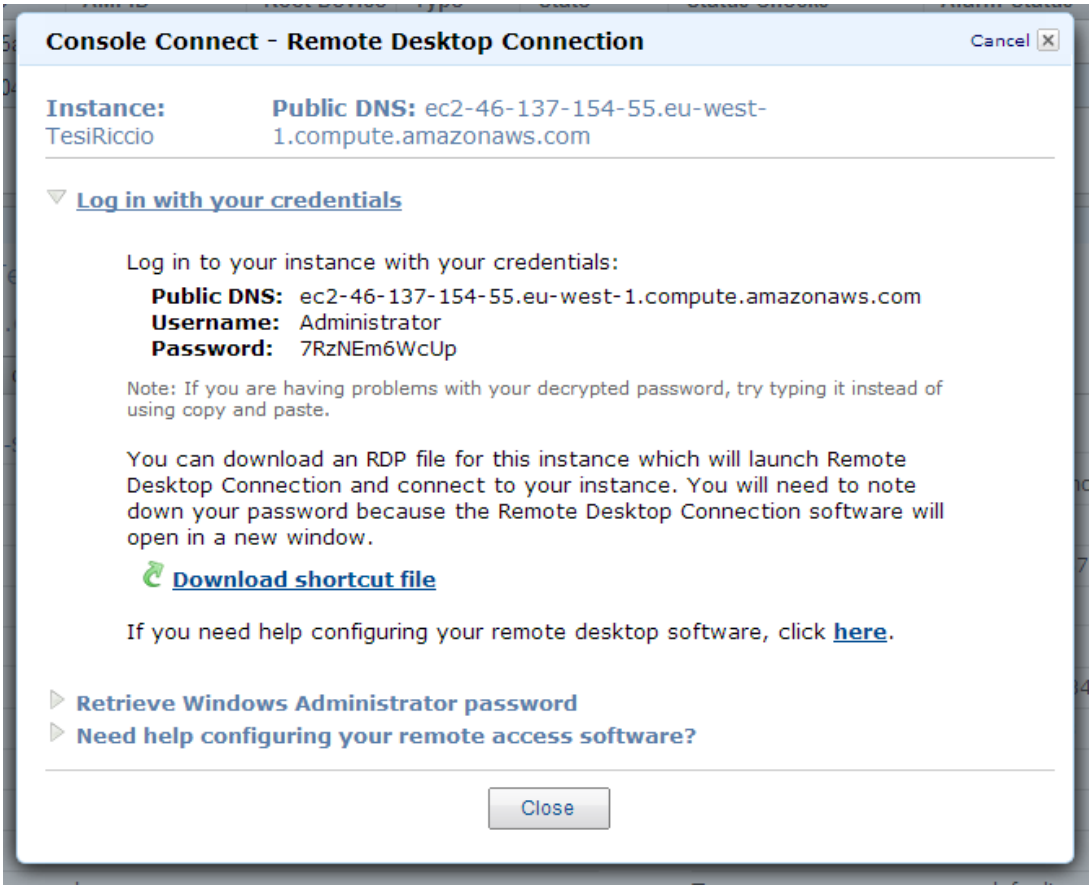
già esistenti, o assegnare e gestire degli IP dinamici o ancora le key pair o i bilanci di carico. Insomma viene offerta al cliente esperto la massima libertà di gestione possibile e a quelli meno esperti, come il sottoscritto, la possibilità in pochi minuti e senza conoscenze eccessivamente approfondite creare un'istanza senza occuparsi di ogni piccola sfaccettatura.



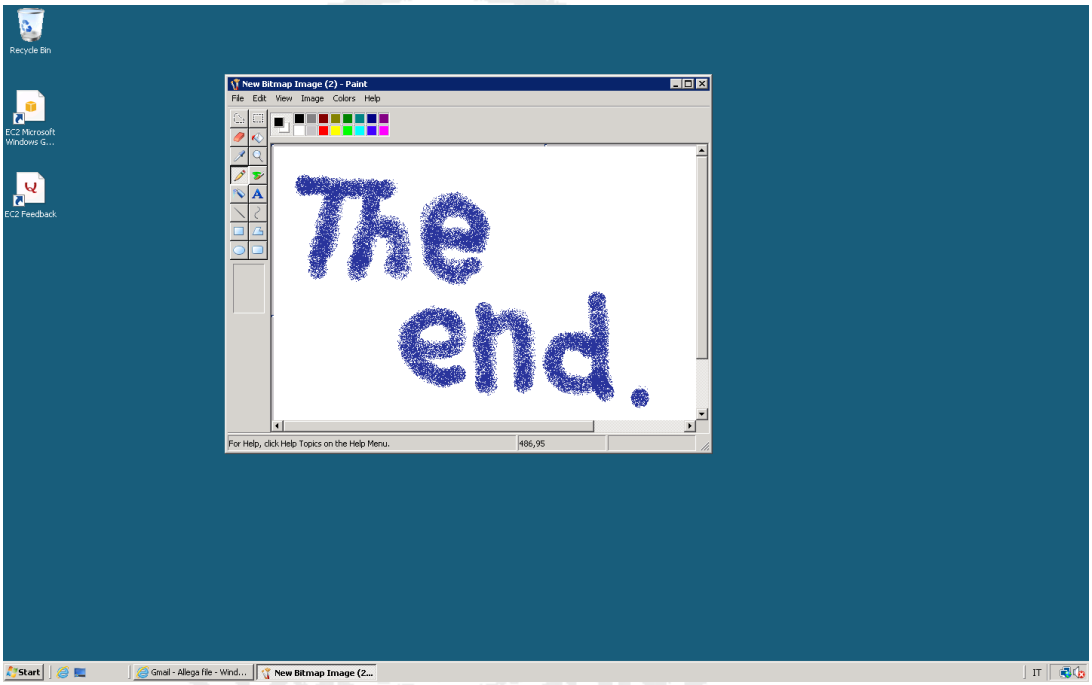
Ripetendo lo stesso procedimento, genero un'istanza di Microsoft Windows Server 2008 a 32 bit, per poter accedere rapidamente da Windows 7 installato sul mio computer.

Si vuole ora verificare il funzionamento di questa istanza e per farlo si deve accedere, cliccando col pulsante destro e selezionando l'opzione connect. Quella che appare è una console di connessione, con le credenziali di accesso alla nostra macchina virtuale e il DNS[37] pubblico. Inoltre offre la possibilità di scaricare sul proprio computer una short cut per un accesso rapido in remoto all'istanza.

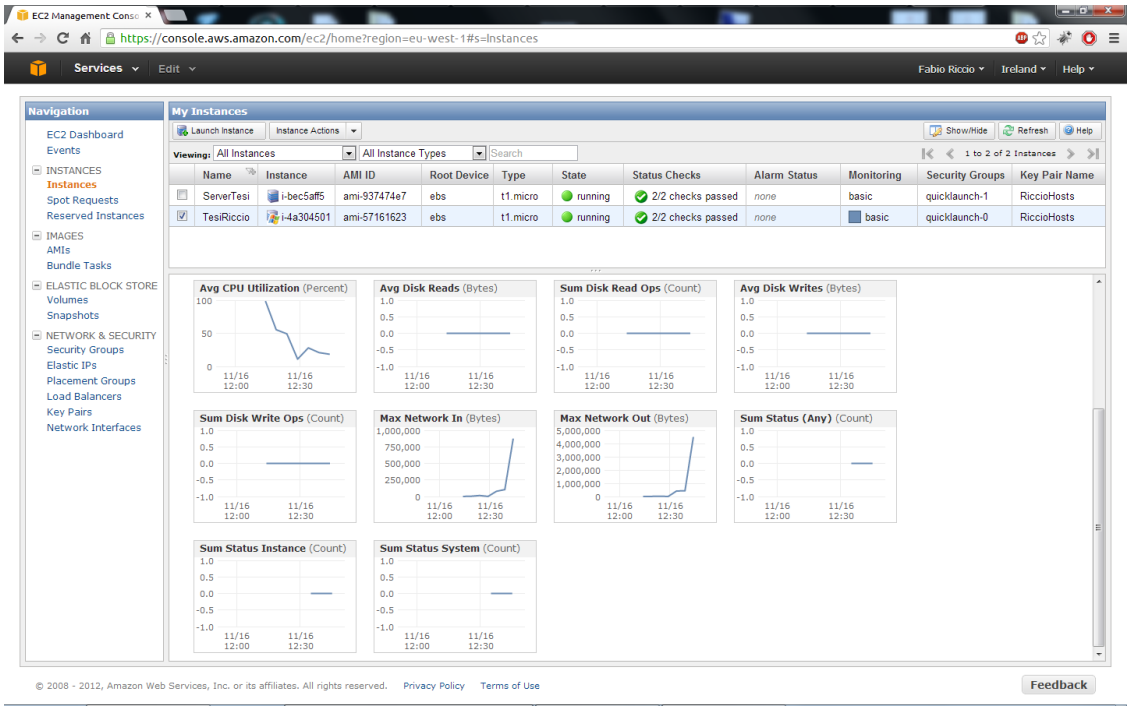
[37] Domain Name System converte gli url internet in indirizzi IP. Può essere utilizzato al posto dell'IP stesso nel caso in cui sia dinamico.



Dopo aver inserito la password, si ha finalmente accesso alla nostra macchina virtuale.



Spegnendo la macchina e tornando alla Console AWS, si può notare come le istanze siano monitorate continuamente e come i servizi di motoring ci supportino anche attraverso l'uso di grafici che forniscono l'andamento di molti parametri, fisici e non, come le statistiche di accesso, la banda utilizzata e l'uso della CPU.



Conclusioni

La piattaforma Amazon Web Services sposa in pieno il concetto dinamico di evoluzione legato al Cloud Computing, non soffermandosi con un approccio solo statico legato ai primi servizi offerti, ma evolvendo essa stessa la sua gamma di soluzioni ai clienti tramite l'offerta di prodotti tecnologicamente nuovi e spesso innovativi, come dimostrano i frequenti webinar e il lancio di servizi recenti, legati ad esempio ai database NoSql.

Certo restano dubbi legati a possibili attacchi hacker ai dati sensibili e ad alcuni crash che, di rado, hanno mandato "down" alcuni famosi siti e applicazioni come Foursquare, Reddit e Blomming, prontamente poi riparati, che dimostrano come ci sia ancora da lavorare per assicurare un servizio, quasi, perfetto.

In conclusione, i punti cardine per il successo di questa piattaforma, messi in evidenza da questo lavoro, sono essenzialmente legati all'abbattimento dei costi, dovuto al concetto di "pay-per-use", e alla facilità di utilizzo, come dimostrato dall'implementazione di istanze macchine virtuali, basate anche su sistemi operativi diversi, in maniera rapida e senza necessità di pesanti configurazioni, che la rendono certamente una delle più importanti realtà di questo settore.

Bibliografia

[1] http://aws.amazon.com

[2] James Marty. 2008, “Programming Amazon Web Services:

S3, EC2, SQS, FPS, and SimpleDB”

[2] http://blog.indigenidigitali.com/amazon-ec2

[3] http://paolobernardi.wordpress.com/2011/10/09/introduzione-ad-hadoop

[4] http://www.cvent.com/events/idc-cloud-symposium-2012/custom-19-8d36c44e7e7a403dbbff471a2f49f9e5.aspx

[5] http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

[6] http://www.cloudtalk.it/multi-tenant-fa-rima-con-cloud-computing-ma-ecco-tutti-i-rischi

[7] http://www.cloudtalk.it/cloudml-una-proposta-di-standard-per-la-descrizione-dei-servizi-cloud

[8] http://www.guardian.co.uk/technology/2008/sep/29/cloud.computing.richard.stallman

[9] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0529:FIN:it:PDF

[10] http://www.mycloudbeans.com/content/cloud-computing-vantaggi-e-svantaggi-clienti?page=0,0

[11] http://www.technologytransfer.it/?cis=4;1&rec=80&yy=2009&mm=11

[12] http://www.hostingtalk.it/articoli/performance/5764/introduzione-memcached-performance-elevate-grazie-ad-una-caching-completa

[12] http://news.cnet.com/8301-13953_3-10052188-80.html

[13] http://www.theregister.co.uk/2012/08/06/wozniak_cloud_will_be_horrendous/