

Cloud Data-Center **basato su tecnologie aperte**

Alberto Cammozzo – 29 Ottobre 2014

Studio preliminare sulla realizzabilità di un data-center basato esclusivamente su tecnologie aperte.

Executive Abstract

Le tecnologie aperte (standard aperti e software free/open source) siono estesamente impiegate per la realizzazione (1) dei *data-center* e per (2) l'offerta di servizio *cloud computing*.

(1) Molte delle tecnologie che costituiscono le **infrastrutture** dei *data-center* privati e commerciali sono già open. Vi è uno standard open per l'hardware e diversi standard aperti che riguardano le tecnologie di virtualizzazione e di storage. La maggior parte dei software per gestire la virtualizzazione sono Open Source, così come quelli che ne consentono il monitoraggio e la gestione. I sistemi operativi impiegati sono prevalentemente della famiglia GNU/Linux. Ne consegue che le tecnologie open costituiscono già una buona parte dei data-center disponibili commercialmente o impiegati privatamente.

(2) Per quanto riguarda le offerte di *cloud computing*, sono disponibili soluzioni aperte anche per le diverse modalità nelle quali questo si declina. Da una parte le **piattaforme** integrate che costituiscono le offerte commerciali di Amazon, Google e Microsoft, e che offrono ai programmatori risorse astratte dall'hardware; dall'altra **applicazioni**, ovvero servizi specifici legati a software specializzati, quali quelli di memorizzazione dati e immagini, o la funzionalità collegata a *web applications* basate su browser o smartphone.

In conclusione sembra del tutto possibile tecnicamente realizzare un data-center che offra servizio di cloud computing basato esclusivamente su tecnologie aperte.

Gli aspetti tecnicamente delicati saranno quelli dell'integrazione delle varie tecnologie e della loro scelta in base alle applicazioni prevalenti. Gli aspetti organizzativi e di redditività saranno probabilmente quelli più critici.

Introduzione

Per *data center* si intende il complesso di sistemi informatici e telematici dedicati all'elaborazione dati; come tale è costituito da sistemi sia hardware che software: computer, sistemi di raffreddamento, power management, sicurezza fisica; corredati dal software necessario: sistemi operativi, virtualizzazione, gestione delle memorie di massa, sistemi di telecomunicazione, sicurezza logica, eccetera. Tradizionalmente il data-center risiedeva nel perimetro fisico dell'azienda (il "CED" aziendale). Con il paradigma denominato *cloud computing* si intende invece

l'offerta dei servizi di un data-center a una molteplicità di realtà distribuite ed anche organizzativamente distinte. Convenzionalmente l'offerta di *cloud computing* viene articolate in tre principali modalità: **IaaS** (Infrastructure as a Service), **PaaS** (Platform as a Service) e **SaaS** (Software as a Service).

IaaS – Vengono offerte delle “macchine virtuali” sulle quali l'utente può installare un sistema operativo a scelta e il proprio software, accedendovi remotamente. L'utente non deve occuparsi dell'hardware, che viene virtualizzato, e può cambiarne le caratteristiche a seconda delle esigenze del momento: aggiungere CPU, memoria centrale e di massa, interfacce di rete. La tariffazione tiene conto di queste risorse. Le imprese commerciali che offrono questo tipo di servizio sono moltissime, grandi e piccole.

PaaS – All'utente vengono offerti dei servizi software già articolati ai quali un programmatore può accedere senza doversi preoccupare della sottostante architettura hardware, di sistema operativo e software. Sono di solito disponibili alcuni linguaggi di programmazione e librerie, database, servizi web, eccetera. Esempi di questo tipo sono i servizi commerciali proprietari di Amazon Web Services (AWS), Microsoft Azure o Google App Engine, ma anche quelli basati su tecnologie open come Heroku, Open Shift. AppScale.

SaaS – All'utente vengono offerti dei servizi specifici svolti da un software che gira centralmente attraverso una molteplicità di interfacce (ad esempio Browser, applicazioni client-server, applicazioni su smartphone) e con una diversità di tecnologie (API, Web, o REST). Il programma non viene scaricato sul computer dell'utente (o solo in minima parte) ma gira su risorse gestite da chi lo offre. Il modello di business e di licenza varia dall'uso gratuito con pubblicità al canone fisso o variabile. In questa categoria ricadono la maggior parte delle “app” per smartphone e i *social network*. Ad esempio Dropbox è una applicazione il cui servizio è la memorizzazione di massa remota di file. Molti software usati secondo questo modello sono open source, anche se non girando sul computer dell'utente quest'ultimo non ha la possibilità di intervenire sul codice sorgente.

Data Center Infrastructure

In questa sezione si elencano le tecnologie disponibili per la realizzazione e la gestione di un data-center, considerandone i vari componenti, dall'hardware al sistema operativo, al monitoraggio.

Con lo *Open Compute Project*, Facebook condivide il *design* dell'architettura **hardware** dei suoi data center: dai server alle soluzioni di gestione efficiente dell'energia e delle memorie di massa.

I **software di virtualizzazione** open source disponibili sono numerosi, i più diffusi sono *XEN*, *Linux KVM*, *Virtual Box*. Va notato che *GNU/Linux* fa parte dei **sistemi operativi** che girano negli ambienti virtualizzati. Tra i **sistemi di gestione della virtualizzazione** ci sono soluzioni *open* come *oVirt*. Analoghe tecnologie sono disponibili per la gestione dello **storage**, come *Ceph*.

Numerosi sono sistemi di **monitoraggio** e reportistica che consentono di gestire il corretto funzionamento dei sistemi del data-center e delle macchine virtuali. Tra questi: *openDCIM*, *Nagios*, *CACTI*, *Munin*.

A livello più elevato esistono software dedicati per garantire la robustezza dei dati e dei servizi, offrendo ***fault tolerance*** e replicazione di dati in modo scalabile, (progetti Apache *Mesos* e *Cassandra*), o che consentano il **calcolo distribuito** con la distribuzione del carico di lavoro (*Hadoop*, *Storm*).

Cloud Computing

Per offrire come servizi i sistemi ospitati nel datacenter, secondo la ripartizione già delineata in Infrastrutture (IaaS) , Piattaforme (PaaS) e Software (SaaS), si possono identificare diverse soluzioni che possono essere costruite in modo stratificato: le **applicazioni** software vanno costruite su **piattaforme** cloud che si appoggiano sulle **infrastrutture** che a loro volta rappresentano una astrazione del **data-center** fisico.

Tra le soluzioni **IaaS** open si distinguono *Eucalyptus*, che fornisce ambienti nei quali modulare le risorse in modo compatibile con Amazon AWS, così come fa anche *CloudStack*. *OpenStack*, nato dalla collaborazione tra RackSpace e NASA offre servizi modulari di astrazione dell'hardware, autenticazione ed autorizzazione, connettività ed interfacce utente. *OpenNebula* è focalizzato sulle applicazioni scientifiche ad alta performance. In merito vi sono degli *standard* sviluppati da Open Cloud Computing Interface (OCCI)

Tra gli esempi di sistemi **PaaS** basati su software open source troviamo i già citati *Heroku*, *Open Shift* di RedHat, *AppScale*. Si tratta di servizi che tipicamente vengono offerti gratuitamente sotto una certa soglia di uso e a pagamento oltre tale soglia. E' ragionevole pensare di poter sviluppare piattaforme di sviluppo analoghe con ambienti open come *CloudFoundry*. I componenti necessari quali server database, web, e strumenti collaborativi sono disponibili da tempo come prodotti maturi nell'ecosistema free/open source.

Per lo sviluppo di servizi dedicati destinati alla PA o alle imprese secondo il modello **SaaS** esistono una grande varietà di linguaggi di programmazione open source , tra i quali per lo sviluppo web si distinguono *Ruby on rails* o il più tradizionale *PHP*. I principali sistemi di gestione database sono open: *Postgres SQL*, *MySql* o *MongoDB*. In molti casi esistono delle librerie e ambienti di programmazione per facilitare lo sviluppo di applicazioni cloud (ad es. *Juju*, *jclouds*, *Libcloud*).

Conclusioni

Le tecnologie open e il software free/open source hanno consentito la nascita del cloud computing, pertanto è pensabile costruire un data-center e dei servizi di cloud computing basati su tali presupposti.

Occorre definire le priorità che determineranno maggiormente l'accento da porre: sviluppare un *private cloud* (o meglio un *community cloud*) per le pubbliche amministrazioni o un progetto più marcatamente commerciale? Andrà focalizzato l'aspetto di contenimento dei costi infrastrutturali (sottolineando l'offerta di infrastrutture – IaaS) o l'opportunità di generare applicazioni e servizi per imprese e pubbliche amministrazioni (PaaS e SaaS)? Di conseguenza andranno definiti i modelli di business e di governance.