

Benvenuti nel cloud computing

Prima di cambiare ufficio almeno un paio di volte, dalla finestra del mio studio al settimo piano ero perfettamente in grado di vedere gli stadi del baseball e del football di Seattle. Questi edifici si trovano uno di fianco all'altro e sono stati ampliati più volte negli anni del boom economico, anche se entrambi i palazzi, costosi e giganteschi, rimangono inutilizzati per la maggior parte del tempo. Un calcolo approssimativo ma nel contempo abbastanza veritiero mi porta a dire che edifici di questo tipo vengono utilizzati appieno per un tempo non superiore all'1% della loro esistenza. Si può altresì affermare che, mediamente, gli stadi di questo tipo rimangono vuoti, e che pertanto risultano inutilizzati centinaia di milioni di dollari di capitale immobiliare. Spesso ricorro all'analogia degli stadi vuoti per spiegare il valore aggiunto del business legato al cloud computing.

Analogamente agli stadi di football, si possono immaginare le condizioni di utilizzo di un grande centro dati aziendale. Un sistema costruito su grande scala è composto da un certo numero di server, costosi e destinati a rapida obsolescenza, che giacciono inutilizzati e in attesa di svolgere elaborazioni sofisticate, di gestire enormi quantità di dati e un numero elevato di visitatori che si collegano al sito web dell'azienda. Da un punto di vista storico una situazione di questo genere è legata alla perenne difficoltà di trovare un equilibrio tra previsioni di traffico e risorse messe a disposizione per gestirlo. Previsioni troppo prudenti portano alla definizione di un data center povero di risorse e rischiano di provocare un "successo disastroso", situazione nella quale un aumento del numero di nuovi utenti comporta il fatto che a ogni utente

In questo capitolo

- **Come evitare un successo disastroso**
- **Saperne di più sul cloud computing**
- **Caratteristiche della nuvola**
- **Modelli di utilizzo del cloud computing**
- **Casi tipici di utilizzo del cloud computing**
- **Riepilogo**

viene offerto un servizio inefficace ed inefficiente; al contrario, previsioni eccessivamente ottimistiche si traducono in costi spropositati e provocano un spreco di risorse che non trova riscontro nell'utilizzo quotidiano del data center.

Questo libro illustra le soluzioni di cloud computing che permettono di risolvere queste difficoltà e di progettare un sistema sempre efficiente e altamente sofisticato da un punto di vista tecnico. Si può tornare per un istante all'analogia degli stadi vuoti: il cloud computing permette di agire come se ogni settimana venisse costruito, utilizzato e poi distrutto uno stadio che abbia dimensioni più funzionali e redditizie. Lo stadio riesce ad avere sempre il numero di posti a sedere, il posteggio, i servizi di ristoro e altre caratteristiche che lo rendono ideale per ogni situazione. In altri termini, uno stadio costruito per 50 persone deve risultare tanto efficiente quanto uno stadio costruito per contenerne fino a 50.000.

È ovvio che una soluzione di questo tipo è impraticabile nel caso di stadi per manifestazioni sportive; al contrario, il cloud computing rende perfettamente ragionevole e praticabile la realizzazione di sistemi personalizzati e operativi in tempo reale. L'infrastruttura di elaborazione dati (server, dispositivi di memorizzazione e banda di trasmissione per il traffico dati) può essere presa e messa a disposizione dalla "nuvola" (*cloud*) costituita dalla rete Internet, sfruttata in base alle necessità dei servizi richiesti e infine rimessa a disposizione della nuvola, il tutto in una manciata di minuti. Questo costituisce la differenza più sostanziale tra le soluzioni di cloud computing e il modello delle infrastrutture precedenti, che risultano tutte statiche e non scalabili. Pagare in base a quello che in effetti serve piuttosto che sopportare costi valutati in funzione di quello che si ritiene possa servire garantisce la definizione di un profilo ideale dei costi dell'applicazione di cloud computing, il che permette in sostanza di ottenere di più utilizzando una quantità inferiore di risorse.

Come evitare un successo disastroso

Si consideri un imprenditore ricco di idee ma limitato nelle risorse a disposizione. L'imprenditore sta pensando a un progetto per un sito innovativo, il cui successo potrà sicuramente oscurare in breve tempo la fama di siti quali Facebook (<http://facebook.com/>) o Twitter (<http://twitter.com/>). Il progetto ha inizio di solito con la stesura di un business plan e con la pianificazione di crescita prevista per i primi sei mesi di vita del sito. Lo studio dei prototipi dell'applicazione da impiegare e l'analisi dei benchmark delle prestazioni porta a concludere che ogni mese sarà necessario acquistare e installare un nuovo server, a condizione che vengano rispettate le previsioni di successo del sito. Il sito stesso non deve mai risultare ai limiti delle proprie capacità, pertanto si dovrà tener conto dei tempi necessari per ordinare, ricevere, installare e configurare ogni nuovo server. Una capacità di riserva adeguata è necessaria per garantire la corretta gestione degli utenti che si dovessero presentare prima dell'installazione completa di un nuovo server; il progetto prevede in questo modo che si debba comunque investire nuove e consistenti somme di denaro, a prescindere dal fatto che nuovi utenti decidano di visitare il sito.

Si supponga ora di aver realizzato il sito, di averlo pubblicato online e di attendere fiduciosi l'arrivo di nuovi utenti. A questo punto si possono verificare tre condizioni, che dipendono dalle previsioni di traffico: queste possono infatti risultare stimate per difetto, per eccesso oppure in modo corretto.

Le previsioni possono essere state pessimistiche, e in questo caso le valutazioni di crescita sono state inferiori alla crescita reale. Invece di un numero limitato di utenti che visitano il sito, il tasso di crescita è stato più elevato del previsto e già i primi utenti hanno esaurito le risorse a disposizione. Il sito è divenuto ben presto sovraccarico, è troppo lento da visitare e i nuovi potenziali utenti se ne vanno del tutto insoddisfatti.

D'altra parte, le previsioni possono essere state troppo ottimistiche. In questo caso la realizzazione del sito ha comportato più risorse rispetto a quelle necessarie. Il progetto ha predisposto un sito per grandi numeri, che però tardano a concretizzarsi. I costi di gestione sono fuori controllo, dato che gli utenti che visitano il sito occupano solo parzialmente le risorse messe a disposizione dai server. Il progetto può perfino fallire a causa dei costi fissi che diventano troppo elevati.

Si può anche ipotizzare che le previsioni di crescita vengano rispettate pienamente, ma anche in questo caso la situazione presenta aspetti critici da valutare con attenzione. Può capitare che un bel giorno ci si accorga che il sito venga richiamato da un link presente sulla pagina iniziale di Digg (<http://digg.com/>), Reddit (<http://reddit.com/>) o Slashdot (<http://slashdot.org/>). Può anche verificarsi il caso in cui un giornalista di una testata famosa decida su due piedi di dare visibilità al sito: un giornalista della CNN può per esempio decidere di riportare l'URL del sito nel banner che scorre in fondo alla pagina del notiziario. Una situazione di questo genere può rappresentare l'avvio del successo che si sta aspettando da tanto tempo.

Sfortunatamente, però, l'infrastruttura rigida del sistema non è quasi mai in grado di supportare esigenze che crescono in modo repentino, pertanto anche in questo caso i nuovi utenti potenziali sono destinati ad abbandonare insoddisfatti il sito. L'eventualità così attesa di ricevere la massima attenzione da parte della Rete si trasforma malauguratamente in un altro "successo disastroso".

È lecito concludere che le previsioni di traffico relative a un sito web rappresentano un compito decisamente improbo. Le probabilità di effettuare valutazioni del tutto scorrette sono elevate, così come sono sempre elevati i costi anche economici che derivano da previsioni errate.

Le tecniche di cloud computing forniscono strumenti utili per affrontare impennate di traffico web, per esempio quelle legate alle situazioni descritte nei paragrafi precedenti. A condizione di pianificare correttamente la progettazione dell'architettura di sistema e di valutare in modo preciso la scalabilità della soluzione, il cloud computing offre la possibilità di affrontare le esigenze di traffico senza compromettere le funzionalità dei server o portare alla bancarotta.

Saperne di più sul cloud computing

È venuto il momento di approfondire il concetto di *cloud computing*. In primo luogo è necessario tenere presente che questo libro si occupa indubbiamente di argomenti tecnici ma vuole evidenziare soprattutto il business che circonda questa nuova tecnologia. Non si tratta solo di sottolineare l'innovazione dovuta alla sua introduzione: il cloud computing è un vero e proprio modello originale di business. La tecnologia è sicuramente interessante in quanto tale e merita tutta l'attenzione che possiamo dedicarle, ma un libro dedicato al cloud computing deve affrontare i nuovi modelli di business, i costi degli investimenti necessari e perfino i diversi aspetti della tecnologia in termini di spese minute e accessorie. Quando ero un giovane appassionato di informatica mi potevo

illudere che questi aspetti “economici” fossero irrilevanti, al punto che trovavo seccante occuparmi di qualcosa che non fosse prettamente tecnologico. Tutto il mio interesse era rivolto alle nuove tecnologie e non volevo parlare di vil denaro! Dopo più di trent’anni di esperienza in prima linea mi sono convinto che un imprenditore con i piedi per terra deve saper combinare gli aspetti tecnici e commerciali se vuole raggiungere obiettivi di successo e duraturi.

Cosa si intende per “nuvola”

In generale quasi tutti gli esperti di informatica sono stati abituati a definire Internet nei termini descritti da un diagramma simile a quello mostrato nella Figura 1.1.

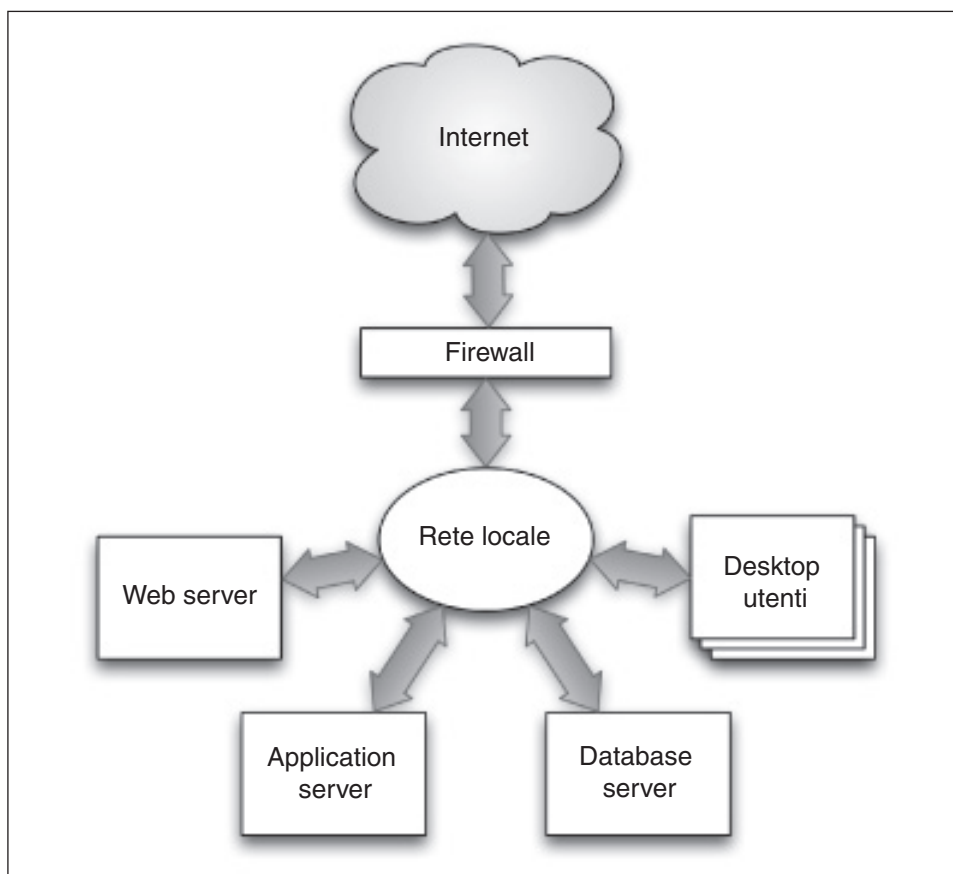


Figura 1.1 Rappresentazione della “nuvola” Internet.

L’immagine della nuvola (*cloud*) è stata da sempre utilizzata per descrivere il concetto di “Internet”, anche se con il tempo il significato di questa immagine è cambiato: se inizialmente la nuvola comprendeva solo le risorse che si trovano su Internet, al momento questa stessa astrazione arriva a includere tutti gli strumenti disponibili per accedere alle risorse.

L'espressione *cloud computing* ha iniziato a diffondersi solo da qualche anno. Per molti si trattava semplicemente di un nuovo termine che descriveva una tecnica web già esistente, piuttosto che una nuova forma di tecnologia; d'altra parte, l'espressione si stava diffondendo tanto rapidamente che alcune applicazioni web già esistenti vennero considerate esempi concreti di applicazione delle tecniche di cloud computing: potenza del marketing! I dettagli delle diverse soluzioni dipendono dalle aziende che forniscono i servizi, ma in generale il cloud computing può essere immaginato come una raccolta coerente e accessibile su larga scala di risorse da condividere in rete per il calcolo e la memorizzazione di dati. Queste risorse vengono allocate utilizzando chiamate (*calls*) relative a servizi web, ovvero un'interfaccia programmabile cui si accede tramite richieste HTTP, e sono messe a disposizione per un tempo che dipende dalla forma di pagamento del servizio, pagamento che viene quantificato in base alle risorse effettivamente impiegate. Da un punto di vista concettuale, un sistema di cloud computing è un ambiente multiutente che opera simultaneamente per un gran numero di utenti. Un sistema di cloud computing deve occuparsi della gestione e del controllo dell'identità degli utenti, è responsabile dell'allocazione delle risorse utilizzate dagli utenti, deve garantire l'accesso esclusivo alle risorse proprietarie di ciascun utente e deve escludere la possibilità che un utente qualsiasi possa interferire con le attività di altri utenti. Il software che caratterizza il cloud computing di un determinato business è da questo punto di vista molto simile al sistema operativo di un personal computer.

Il cloud computing si basa su un certo numero di tecnologie fondamentali, che includono protocolli di rete TCP/IP, connessioni a Internet affidabili, servizi web affini a soluzioni SOAP e REST, apparati hardware, sistemi di virtualizzazione e di pagamento online. Molte di queste tecnologie prevedono soluzioni che risultano perlopiù nascoste, ed è la nuvola che mette a disposizione degli sviluppatori web una configurazione ideale e astratta delle risorse disponibili.

Elaborazione automatica dei dati

Conviene a questo punto analizzare la configurazione che da sempre ha regolato il funzionamento delle risorse IT prima del cloud computing. Le osservazioni riportate nei prossimi paragrafi fanno riferimento indistintamente a risorse che riguardano server, unità di memorizzazione di dati in rete, indirizzi IP, larghezza di banda e dispositivi firewall. Si supponga di lavorare in una grande azienda e di trovarsi nella necessità di rinnovare le risorse IT. È molto probabile che un'operazione di questo genere richieda un processo di revisione che coinvolge un certo numero di persone da formare e di accordi da stipulare. Sarà necessario comunicare con altre persone utilizzando la posta elettronica, creare un sistema di ordini o pagamenti online, oppure più semplicemente si dovrà alzare il telefono e contrattare nuove soluzioni da adottare in termini di requisiti hardware e software. L'interlocutore di riferimento può a sua volta prevedere una sorta di elaborazione "manuale" per approvare la richiesta di nuove risorse IT; deve per esempio ricercare la soluzione hardware più efficace, allocare e configurare le risorse, decidere le forme di cablaggio, quali router e firewall impiegare e così via. Non è insolito che in queste condizioni la soluzione possa richiedere anche 12 o 18 mesi per diventare operativa!

Si consideri per esempio un imprenditore che decide di chiamare il proprio provider ISP (*Internet Service Provider*). L'imprenditore stabilisce con l'ISP una nuova soluzione IT, si

accorda per una forma di pagamento mensile passibile di aumento e ottiene l'hardware richiesto in un tempo che può variare da poche ore a qualche giorno.

Dopo aver avviato una soluzione di questo genere è probabile che l'azienda si trovi impegnata per lungo tempo allo scopo di rientrare dall'investimento. Le grandi aziende inseriscono nel bilancio mensile le spese di gestione dei propri data center e devono assolutamente garantire l'utilizzo pieno delle risorse a disposizione, se possibile per un intervallo di tempo pari al ciclo di vita delle risorse stesse. Da questo punto di vista i servizi ISP sono più flessibili, ma è raro che riescano a effettuare modifiche significative a poche ore di distanza l'una dall'altra.

Il cloud computing introduce in questo scenario la possibilità di escludere il fattore umano dal meccanismo che regola il funzionamento delle risorse IT. Si consideri di nuovo l'imprenditore alle prese con il rinnovamento delle risorse IT della propria azienda, o meglio ancora si consideri il management delle risorse che si occupa di queste operazioni per conto dell'azienda. Il meccanismo di cloud computing si occupa di elaborare una nuova richiesta di servizio web, ovvero una nuova "call". È sempre il meccanismo di cloud computing che si occupa di effettuare i passaggi elencati di seguito, fino a ottenere una soluzione che soddisfi la richiesta iniziale.

1. Accetta la richiesta.
2. Conferma che la richiesta è autorizzata.
3. Convalida la richiesta in funzione dell'account che la effettua.
4. Individua risorse libere e in grado di soddisfare la richiesta.
5. Mette in collegamento le risorse e l'account che effettua la richiesta.
6. Inizializza le risorse da utilizzare.
7. Riporta i dati di identificazione delle risorse che soddisfano la richiesta.

A questo punto l'applicazione di cloud computing ha ottenuto l'accesso esclusivo alle risorse per il tempo necessario. Non appena non ha più bisogno di utilizzare le risorse, è la medesima applicazione che si fa carico di rimetterle a disposizione del cloud computing. Le risorse possono in questo modo essere configurate nuovamente per il loro riutilizzo e vengono segnalate come disponibili; a volte può risultare infatti necessario formattare, cancellare oppure riavviare i dispositivi che costituiscono le risorse da riutilizzare.

Gli sviluppatori IT sono abituati a progettare soluzioni in termini di sistemi orientati agli oggetti, e anche un sistema di cloud computing può essere visto come un oggetto. In sostanza, la definizione teorica di un sistema di cloud computing può essere indicata in PHP come illustrato di seguito (si noti che queste istruzioni non fanno riferimento a un sistema di cloud computing reale; i nomi dei metodi e dei parametri sono stati inclusi solo per illustrare il concetto ideale di cloud computing):

```
class Cloud
{
    public function getDataCenters()
    {
        :
    }

    public function allocateServer($dataCenter, $count)
    {
        :
    }
}
```

```

}

public function releaseServer($server)
{
    :
}

public function allocateDiskStorage($dataCenter, $gb)
{
    :
}

public function releaseDiskStorage($storage)
{
    :
}

:
}

```

Le istruzioni che seguono illustrano l'utilizzo teorico di un sistema di cloud computing. In primo luogo dovete ottenere un elenco di data center disponibili (\$d) e memorizzare un riferimento relativo alla prima voce dell'elenco (\$d1):

```

$c = new Cloud();
$d = $c->getDataCenters();
$d1 = $d[0];

```

Ora potete allocare un server (\$server) del data center e creare uno spazio di memorizzazione dei dati (\$storage):

```

$server = $c->allocateServer($d1, 1);
$storage = $c->allocateDiskStorage($d1, 100);

```

Il punto fondamentale da considerare è che potete scrivere un programma per inizializzare, controllare, analizzare e gestire nel suo complesso l'utilizzo di risorse di cloud computing disponibili su larga scala. La decisione sulle priorità da rispettare e sulla frammentazione delle scelte (che riguardano per esempio l'aumento della capacità del server o l'allocazione della capacità già esistente) era in precedenza affidata a operazioni manuali che avvenivano raramente ed erano condotte direttamente dagli amministratori di sistema; ora le stesse decisioni possono essere affrontate in modo automatico e vengono effettuate a intervalli regolari.

Caratteristiche della nuvola

Dopo aver introdotto i concetti fondamentali del cloud computing si può passare allo studio delle caratteristiche e degli attributi che caratterizzano questo nuovo sistema di gestione delle risorse IT. L'esperienza acquisita grazie alla partecipazione attiva ai forum pubblici di Amazon Web Services fa ritenere che l'illustrazione pratica delle caratteristiche del cloud computing risulti più efficace della semplice definizione del sistema generale, soprattutto quando si vuole spiegare l'essenza stessa del cloud computing e il suo funzionamento teorico.

Caratteristiche generali

Di seguito vengono illustrate alcune caratteristiche generali del cloud computing.

Elasticità

Il cloud computing permette di scalare ampiamente l'utilizzo di risorse in base alle necessità contingenti. Il tempo richiesto per aumentare o diminuire l'utilizzo delle risorse è valutato in secondi o minuti, invece che in settimane o mesi.

Economia di scala

Chi fornisce la soluzione di cloud computing è in grado di sfruttare una vasta economia di scala e propone a prezzi concorrenziali la disponibilità di supporti logistici, sistemi di alimentazione e di raffreddamento, larghezza di banda e hardware. L'ISP fornisce l'infrastruttura del sistema in forma di prodotto, pertanto ha tutto l'interesse a fare in modo che i costi di gestione possano diminuire nel tempo. Il provider è anche in grado di impegnare una parte del proprio personale nelle attività esclusive necessarie per garantire i servizi su scala globale.

Pay-as-you-go

Si tratta di una caratteristica generale che non è legata solo a questioni economiche per un motivo ben preciso: i servizi basati sul cloud computing prevedono che il personale tecnico sia in grado di prendere decisioni che riguardano la gestione delle risorse in modo che queste stesse decisioni abbiano effetto immediato sulla gestione e sui costi complessivi del servizio. Predisporre soluzioni efficienti da un punto di vista economico diventa un obiettivo di tutti.

Caratteristiche economiche

Vediamo ora alcune caratteristiche del cloud computing che interessano gli aspetti economici del sistema.

Nessun investimento iniziale

Dato che il cloud computing prevede una gestione delle risorse in base alle richieste di utilizzo, non è necessario prevedere un investimento iniziale consistente prima che si verifichi la richiesta effettiva delle risorse.

I costi fissi diventano variabili

Invece di prevedere un determinato utilizzo delle risorse per tutta la durata commerciale del contratto (che spesso implica una durata variabile da uno a tre anni), il cloud computing permette di modificare in tempo reale la valutazione dei costi relativi all'utilizzo delle risorse.

Le spese in conto capitale diventano spese di consumo

Le spese in conto capitale avvengono sul lungo termine e si riflettono in un contratto di assistenza pluriennale che riguarda l'utilizzo di una determinata quantità di risorse. Le spese di consumo avvengono, al contrario, in base all'utilizzo effettivo del sistema di cloud computing e possono essere quantificate in tempo reale.

L'allocazione delle risorse è definita in modo puntuale

Il cloud computing permette di quantificare un utilizzo minimo in funzione del tempo impiegato o delle risorse disponibili; è possibile per esempio valutare i costi del servizio

in termini di ore di utilizzo del server oppure quantificare le spese in funzione dei byte memorizzati.

Il business diventa flessibile

Dato che non è più necessario definire un contratto economico a lungo termine che riguarda l'utilizzo delle risorse, si è in grado di rispondere rapidamente a variazioni significative delle transazioni in termini di volume o di tipologia del servizio.

Il business fondamentale del provider

Il provider del cloud computing basa il suo business sulla possibilità di fornire un sistema pubblico di cloud computing e, in quanto tale, ha tutto l'interesse a sviluppare servizi affidabili, adattabili ed efficienti da un punto di vista economico. La soluzione adottata riflette ampiamente le competenze chiave del provider.

I costi sono dotati di proprietà associativa

Il modello flessibile di allocazione delle risorse, proprio del cloud computing, permette di acquisire e mettere in funzione con la stessa facilità 100 server per un'ora oppure un server per 100 ore. Questo offre la possibilità di studiare soluzioni originali anche per quanto riguarda la frammentazione di quei problemi che si verificano inizialmente su larga scala.

Caratteristiche tecniche

Illustriamo ora alcune caratteristiche del cloud computing che riflettono gli aspetti tecnici della soluzione adottata.

Scalabilità in tempi rapidi

È possibile acquistare online nuovo hardware in pochi minuti allo scopo di risolvere richieste inattese di nuove risorse; le richieste possono interessare attività interne al sistema (per esempio un aumento delle operazioni di calcolo) oppure attività esterne (per esempio legate al traffico in un sito web). In alternativa, è possibile rimettere le risorse a disposizione del cloud computing non appena il loro utilizzo non è più necessario.

La scalabilità del sistema sembra infinita

Ogni singolo utente del servizio ha la possibilità di operare come se il sistema offrisse una scalabilità praticamente infinita, anche se questa condizione è vera solo virtualmente. Non è necessario predisporre nuove risorse in anticipo, dato che gli aumenti improvvisi delle richieste costituiscono un problema per il provider del sistema di cloud computing e non più per l'utente del sistema stesso.

Le risorse sono astratte e indifferenziate

Il cloud computing accentua l'interesse nei confronti degli aspetti fondamentali del sistema, ovvero risultati e prestazioni osservabili rispetto alle specifiche tecniche dell'hardware impiegato. È sempre possibile modificare e perfezionare l'hardware che viene utilizzato dal sistema, ma questo compito spetta direttamente al provider dei servizi di cloud computing. Non è più necessario conoscere i dettagli tecnici di una particolare risorsa che può cambiare nel tempo.

La nuvola è composta da componenti singoli

Le risorse IT offerte dal cloud computing sono composte da singoli blocchi di costruzione, ben distinti a livello di prezzo e di consistenza. L'utente dei servizi può in questo

modo stabilire se utilizzare tutti oppure solo alcuni dei servizi messi a disposizione dal sistema di cloud computing.

La sperimentazione di nuove soluzioni è poco costosa

Il cloud computing elimina le barriere economiche che possono ostacolare la sperimentazione di nuove soluzioni. Potete infatti accedere a risorse temporanee per valutare l'efficacia di un'idea originale senza ricorrere a consistenti impegni economici in termini di hardware da impiegare.

Equivoci comuni sul cloud computing

Dopo aver avuto modo in questi ultimi anni di discutere con molte persone, mi sono convinto che esistono ancora molti equivoci attorno al concetto di cloud computing. Alcuni di questi malintesi sono dovuti semplicemente alla difficoltà che molte persone manifestano nei confronti delle innovazioni in generale, altri riflettono invece il fatto che tutte le tecnologie sono in continua evoluzione e che nel tempo stanno comparando servizi e funzionalità sempre più innovativi. Ciò che ora sembra sostanziale è destinato a essere superato dall'offerta di servizi nuovi e più avanzati. Vale comunque la pena tenere presente gli equivoci elencati di seguito, alcuni dei quali derivano dal lavoro svolto presso la University of California di Berkeley.

NOTA

Si veda Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy H. Katz, Andrew Konwinski, Gunho Lee, David A. Patterson, Ariel Rabkin, Ion Stoica e Matei Zaharia, *Above the Clouds: A Berkeley View of Cloud Computing* (Berkeley: University of California, 2009), disponibile all'indirizzo <http://d1smfj0g31qzek.cloudfront.net/abovetheclouds.pdf>.

“Il cloud computing è un capriccio”

La storia anche recente ha mostrato parecchi esempi di tecnologie promettenti e innovative che si sono rapidamente “sgonfiate”, pertanto è più che giustificabile la posizione di chi manifesta scetticismo. È importante essere in grado di reagire in modo tempestivo ed efficace a tutte le modifiche che si possono verificare in un determinato ambiente operativo: questa esigenza rimarrà valida ancora per molto tempo, e le tecnologie di cloud computing costituiscono un esempio concreto di applicazione di questa teoria.

“Il cloud computing richiede una nuova definizione dell'architettura delle applicazioni”

Questa affermazione è piuttosto diffusa. Se da un lato è vero che alcune vecchie applicazioni devono essere ridefinite per sfruttare i vantaggi del cloud computing, è altrettanto vero che applicazioni già in funzione e basate su programmi commerciali oppure open source possono spesso essere trasferite nel cloud computing con interventi minimi o perfino nulli. Non è detto che queste applicazioni possano sfruttare tutte le caratteristiche del cloud computing, ma nella maggior parte dei casi i vantaggi sono comunque significativi.

“Il cloud computing è per natura poco sicuro”

La presenza di dati sensibili per l'azienda in un posto che è “altrove” può diventare una situazione preoccupante e insidiosa per un responsabile IT abituato ad avere il pieno

controllo delle informazioni riservate. I provider di cloud computing sono ben consapevoli di queste preoccupazioni e affrontano il problema con tutta l'attenzione dovuta. In generale preferiscono condividere con gli utenti dei servizi i dettagli relativi alla sicurezza e le policy di protezione dei dati. I responsabili IT possono confidare anche sulla presenza di sistemi avanzati di sicurezza e sul controllo completo del meccanismo di indirizzamento della rete e di supporto della codifica dati, cui si aggiungono apposite certificazioni di sicurezza, per esempio la certificazione SAS 70 (<http://www.sas70.com/>). Nel prossimo capitolo verranno illustrate le tecniche che permettono a sviluppatori, CIO (*Chief Information Officer*) e CTO (*Chief Technical Officer*) di acquisire familiarità con il cloud computing.

“Il cloud computing è un collo di bottiglia”

Cosa succede quando il sistema di cloud computing smette di funzionare? La nuvola ha in sé una ridondanza funzionale e geografica che può garantire la massima disponibilità dei servizi.

“Il cloud computing è un ambiente chiuso”

Il cloud computing permette di eseguire applicazioni già esistenti, che possono pertanto essere attivate e disattivate senza problemi; in altre parole, un sistema di cloud computing offre la possibilità di utilizzare sistemi operativi, middleware e applicazioni web senza o con poche modifiche. D'altra parte, va ricordato che è sempre possibile aggiornare le applicazioni impiegate dal sistema di cloud computing allo scopo di sfruttare pienamente i vantaggi offerti da questa nuova piattaforma web, come verrà ampiamente illustrato nei prossimi capitoli.

“Il cloud computing va bene solo per eseguire codice open source”

Questa affermazione è ormai priva di fondamento. I produttori di sistemi operativi e di software commerciali riconoscono finalmente il cloud computing come un ambiente di riferimento per lo sviluppo di applicazioni web e sono impegnati a garantire che le loro applicazioni web includano licenze compatibili con le soluzioni di cloud computing. I produttori più lungimiranti sono già in grado di offrire licenze software basate su accordi commerciali a tariffa oraria o di tipo pay-as-you-go; per esempio, invece di acquistare la licenza di accesso a un database pagando decine o perfino centinaia di migliaia di dollari, potete avere accesso al medesimo database pagando una tariffa oraria di pochi dollari.

“Le risorse del cloud computing sono troppo costose”

Non è facile confrontare i costi legati all'utilizzo di risorse IT interne e all'impiego di risorse equivalenti basate su un sistema di cloud computing.

NOTA

A questo proposito si può fare riferimento, per esempio, al blog di James Hamilton e in particolare al suo post intitolato *McKinsey Speculates that Cloud Computing May Be More Expensive than Internal IT*, disponibile all'indirizzo <http://perspectives.mvdirona.com/2009/04/21/McKinseySpeculatesThatCloudComputingMayBeMoreExpensiveThanInternalIT.aspx>.

La determinazione precisa dei costi da imputare alle risorse interne richiede un'attenzione contabile che probabilmente è assente nella maggior parte delle aziende di medie e grandi dimensioni. Spesso si trascurano i costi fissi oppure si confrontano erroneamente

i costi delle risorse interne quantificati su base oraria coi costi delle risorse tipiche del cloud computing, scalabili e a costo zero se inutilizzate.

Nei prossimi capitoli verranno illustrate ulteriori ragioni che spiegano perché tutte queste affermazioni sono da considerare equivoci da superare.

Modelli di utilizzo del cloud computing

A questo punto si possono studiare alcuni modelli tipici di utilizzo del cloud computing. Le informazioni che seguono dovrebbero consentire a chiunque di stabilire se la propria applicazione web si può adattare o meno al sistema AWS. I modelli che verranno descritti fanno riferimento a un utilizzo delle risorse valutato in funzione del tempo, anche se occorre tenere presente situazioni diverse una dall'altra. Va infine ricordato che il termine "utilizzo" comprende in genere una combinazione di risorse impiegate dal sistema di cloud computing, ovvero server, unità di memorizzazione e larghezza di banda.

Utilizzo costante nel tempo

Tipico di applicazioni interne all'azienda, per le quali si prevede una scarsa variazione di traffico su base giornaliera oppure su base oraria.

Carico interno di tipo ciclico

Tipico di applicazioni che riguardano elaborazioni batch che si ripetono con regolarità, per esempio legate alla chiusura contabile giornaliera o mensile. In questi casi il carico di lavoro è ampiamente prevedibile, in termini di tempo e di impiego delle risorse.

Carico esterno di tipo ciclico

Questo modello può essere applicato a determinati settori di mercato, per esempio settori che riguardano l'intrattenimento e gli eventi sportivi.

Carico interno con picchi di utilizzo

Modello tipico di ambienti nei quali ricercatori o analisti di mercato possano impiegare il sistema con applicazioni molto impegnative ma per un tempo limitato. Spesso la richiesta di questo genere di servizi non è prevedibile con facilità.

Carico esterno con picchi di utilizzo

Questo modello riguarda siti web che diventano improvvisamente molto popolari, il che si verifica spesso per un tempo limitato.

Tasso costante di crescita nel tempo

In genere questo modello si applica ad applicazioni di un sito web che ha raggiunto la maturità di esercizio. Un aumento di utenti che accedono a queste applicazioni corrisponde di solito a un aumento corrispondente del traffico e delle risorse impiegate.

Casi tipici di utilizzo del cloud computing

In questo paragrafo vengono illustrate alcune situazioni tipiche di implementazione e utilizzo del cloud computing, a partire da casi semplici per arrivare a soluzioni più articolate.

Hosting di semplici siti web statici e di applicazioni web complesse

Una nuvola può includere un semplice sito web statico costituito da pagine HTML statiche, fogli stile CSS e immagini. In effetti, il cloud computing può essere costituito semplicemente dall'hosting di un sito web di questo tipo, cui si può per esempio aggiungere un sistema di distribuzione dei contenuti. Una nuvola favorisce ovviamente l'hosting di siti più complessi, che comprendono per esempio un'elaborazione dati significativa lato server e l'accesso a un database relazionale. I siti che sfruttano il cloud computing a livello di unità di memorizzazione e di elaborazione dati richiedono spesso una disponibilità significativa di risorse per l'elaborazione e la memorizzazione dei dati, oltre che per garantire il livello desiderato di scalabilità.

Supporto dello sviluppo software

Una nuvola è un ambiente ideale per avere a disposizione i requisiti hardware e software richiesti in ogni fase di sviluppo di un'applicazione web. L'utilizzo del cloud computing in fase di sviluppo garantisce di poter sfruttare le risorse necessarie per esaudire tutte le richieste dell'applicazione. Si consideri un team di sviluppatori che sta lavorando su un'applicazione che include un livello web, un livello di applicazioni e un livello di database. Ciascuno di questi livelli deve risiedere su un server fisico diverso dagli altri già in fase di sviluppo. Se si escludono le opportunità offerte da AWS, ognuno degli sviluppatori deve avere a disposizione tre server completi, ciascuno dei quali rimane inutilizzato per la maggior parte del tempo. I costi di sviluppo sono destinati a crescere in modo significativo ogni volta che il progetto coinvolge un nuovo sviluppatore. Il trasferimento dello stesso progetto in una nuvola si traduce nel fatto che ogni sviluppatore ha la possibilità di avviare i server al mattino, di sviluppare e testare le applicazioni web per tutto il giorno e restituire l'utilizzo dei server al sistema di cloud computing alla fine della giornata di lavoro.

Il cloud computing evidenzia la sua importanza innovativa anche in fase di test del software. Anche in questo caso gli sviluppatori possono avviare i server ed eseguire svariati test senza sovraccaricare i rispettivi server di sviluppo. Nel caso di test multipli è possibile utilizzare più server in parallelo, allo scopo di distribuire il carico di lavoro su più risorse. Il cloud computing può essere inoltre utilizzato per supportare il lavoro in un ambiente integrato. In un ambiente di questo tipo le operazioni definite da codice sorgente avviano un processo in più passaggi di ricostruzione dell'applicazione web, di test dell'unità software e di analisi funzionale. Se il codice dell'applicazione web è stato scritto per lavorare in diversi ambienti (per esempio con versioni differenti di Linux) oppure su diverse piattaforme (per esempio Windows e Linux), la nuvola rappresenta una valida soluzione alternativa all'acquisto in proprio di un'infrastruttura completa.

Le risorse di cloud computing permettono di eseguire in ogni fase di sviluppo i test relativi al carico di lavoro e alle prestazioni. Se l'applicazione stessa deve essere eseguita dal sistema di cloud computing, si può essere certi che i test software vengono effettuati in condizioni di carico ottimali, ovvero che si aggiungono risorse quando aumenta il carico di lavoro e che queste vengono rimosse non appena non sono più necessarie.

I test sulle prestazioni di un'applicazione web da distribuire a livello pubblico o aziendale diventano più facili da realizzare quando la nuvola può mettere a disposizione le risorse che permettono di eseguire il test in condizioni di scalabilità che corrispondono al carico

di lavoro previsto per l'applicazione web. Sono molte le aziende che utilizzano le risorse di cloud computing per generare carichi di lavoro equivalenti alla presenza simultanea di centinaia di migliaia di utenti.

Dopo aver distribuito l'applicazione (a prescindere dal fatto che si utilizzi o meno un sistema di cloud computing), la nuvola è comunque in grado di mettere a disposizione le risorse necessarie per eseguire test di compatibilità nel momento in cui si aggiornano i livelli middleware dell'applicazione oppure altri componenti comuni. Un test completo dell'applicazione permette di valutare la plausibilità di aggiornamenti significativi di un sistema di produzione senza rischiare di incorrere in tempi morti.

Programmi di formazione

Una nuvola è in grado di fornire le risorse a breve termine che permettono di supportare svariate tipologie di programmi di formazione.

Si consideri il caso di studenti che stanno imparando come installare, eseguire e controllare un'applicazione web simile a quella descritta nel paragrafo precedente. Gli studenti possono utilizzare i propri computer portatili per accedere a risorse di cloud computing che vengono allocate per la durata della lezione. Al termine delle lezioni le risorse vengono restituite al sistema di cloud computing. Gli studenti hanno pertanto la possibilità di partire da una singola immagine "master" ed evitare in questo modo di perdere il tempo richiesto dall'installazione e configurazione dei pacchetti e delle applicazioni che stanno studiando. I programmi di formazione tradizionali sono costretti a stabilire un limite nel numero di studenti che compongono una classe, a causa dei limiti imposti dall'hardware a disposizione della classe. Al momento alcune aziende del settore stanno proponendo seminari di formazione online basati sull'utilizzo di risorse di cloud computing che mettono a disposizione un nuovo server ogni volta che uno studente si unisce alla classe che sta svolgendo una lezione. Una tecnica simile a questa è già stata utilizzata anche da produttori di software di database e applicazioni web, con risultati molto soddisfacenti.

Demo di applicazioni

È possibile utilizzare le risorse offerte dal cloud computing per l'hosting e la distribuzione di demo e versioni trial di pacchetti di applicazioni web. I programmi demo sono già a disposizione di clienti potenziali quando i programmi stessi sono ancora "caldi", e non è più necessario attendere che venga predisposto un apposito ambiente di prova. I produttori software possono fornire l'accesso a un server ospitato da una nuvola a costi contenuti e con un breve preavviso. I tempi di vendita di un prodotto vengono in questo modo abbreviati in modo significativo e i clienti hanno anche la possibilità di sperimentare direttamente l'utilizzo di risorse basate sul cloud computing. In alcuni casi le demo costruite adottando un sistema di cloud computing possono condurre alla distribuzione di applicazioni web, sempre basate sul cloud computing.

Memorizzazione dei dati

Una nuvola è un ambiente ideale per memorizzare dati pubblici o privati. In questo caso è fondamentale valutare fattori quali la scalabilità, la durata a lungo termine e l'econo-

mia di scala. Il cloud computing permette di memorizzare dati semplici e compatti, che riguardano per esempio il backup di file personali; lo stesso sistema permette altresì di memorizzare dati complessi e di grandi dimensioni, quali possono essere i dati relativi a backup aziendali. Ovviamente il cloud computing è in grado di occuparsi anche della memorizzazione di dati che riguardano situazioni intermedie rispetto a queste.

Spesso la memorizzazione di dati nel cloud computing costituisce solo il primo utilizzo di questo innovativo sistema di servizi web, un primo passo che può portare a considerarne l'impiego in situazioni molto diverse e decisamente più complesse.

Ripristino e continuità del business

Le aziende il cui operato dipende in modo significativo dalle risorse IT devono assolutamente prevedere strategie di intervento nel caso di contrattempi di qualsiasi genere, a prescindere dal fatto che si verifichi una mancanza temporanea o permanente di risorse oppure che risulti impossibile accedere alle stesse. Le strategie di intervento devono tenere presente l'eventualità che si verifichi un incendio, un'inondazione, un terremoto oppure un atto terroristico che possa danneggiare le attività dell'azienda. Molte società prevedono la presenza di un intero data center di riserva; in altri casi i dati aziendali vengono copiati regolarmente in un centro di backup e l'intero pacchetto di dati può essere recuperato in breve tempo. È inutile sottolineare che i costi di realizzazione e di gestione di queste strategie sono indubbiamente molto significativi.

Anche in questo caso il cloud computing offre una soluzione alternativa che garantisce la continuità delle attività aziendali. Invece di investire capitali in hardware che in condizioni normali non viene utilizzato, la rete aziendale nel suo complesso può essere impostata come un gruppo di risorse di cloud computing configurate come un modello e messe a disposizione quando si verifica un problema di funzionamento della rete "normale". Questa soluzione implica un accordo con il provider del sistema di cloud computing al fine di garantire che le risorse necessarie siano disponibili non appena si verifica un problema.

Dopo aver configurato la rete aziendale per garantire la continuità di esercizio, è possibile considerare altri impieghi interessanti del medesimo sistema. Le soluzioni web tradizionali prevedono che la distribuzione di versioni aggiornate di componenti middleware e condivisi richieda numerosi test di compatibilità e sulle prestazioni: questo compito rappresenta spesso una sfida micidiale. Sono molte le aziende che si trovano costrette ad affrontare problemi quasi insormontabili e che non riescono a distribuire l'aggiornamento del codice a causa dell'impossibilità di sviluppare test adeguati prima della distribuzione, e che non vogliono rischiare le conseguenze negative che possono derivare da una distribuzione gestita in modo non corretto.

Si può al contrario valutare la possibilità di avere a disposizione una copia completa (o quantomeno una copia ridotta ma significativa) della rete aziendale e di inglobare le versioni dei componenti da testare, il tutto da verificare a livello di compatibilità e di carico di lavoro utilizzando il cloud computing a un costo decisamente più contenuto.

Elaborazione e rendering multimediale

Molti siti web di successo supportano l'upload di file multimediali, ovvero di musica, immagini fisse o filmati. Dopo aver completato l'upload, i file sono sottoposti a svariate

operazioni, che possono risultare impegnative in termini di elaborazione dati, di gestione dell'I/O o di entrambi. I file vengono esaminati per individuare la presenza di virus e di altre forme di malware. Si controlla la firma digitale dei file musicali (alla ricerca di eventuali violazioni del copyright) e si esegue la loro transcodifica per consentire la riproduzione audio con differenti bit rate. Le immagini vengono ridotte in scala, contrassegnate e verificate per evitare duplicazioni indesiderate, poi riprodotte in formati diversi tra loro. Anche sui video si esegue una transcodifica e una riduzione in scala, cui si aggiunge a volte la frammentazione in brani video di durata inferiore rispetto all'originale. Solo alla fine di queste operazioni gli oggetti multimediali risultano memorizzati definitivamente e sono a disposizione di chi li vuole visualizzare online o scaricare sul proprio computer. Le operazioni di rendering utilizzano una descrizione della scena per generare i frame di un filmato di animazione. Il rendering di ogni frame può richiedere una elaborazione diversa da quella prevista per altri frame. Praticamente esistono pochi dati di input ma si produce una grande quantità di dati di output. L'elaborazione impegna le risorse in modo significativo, dato che è necessario elaborare ogni pixel di ogni frame e si deve tenere conto di effetti dovuti a luci, ombre, colore e movimento. Il cloud computing è un ambiente ideale per eseguire questo genere di elaborazioni dati e di rendering, in particolare considerando la quantità di memoria, il carico di lavoro del processore e la larghezza di banda che è necessario avere a disposizione.

Elaborazione di dati scientifici e finanziari

L'elaborazione di dati a carattere scientifico e finanziario implica spesso la necessità di disporre di data set su vasta scala e può richiedere l'utilizzo di prestazioni significative da parte della CPU. L'analisi di questi dati viene spesso effettuata in base a richieste specifiche, il che comporta a sua volta un impegno eccessivo delle risorse interne, spesso limitate. In effetti, è possibile affermare che molti apparati di calcolo scientifico passano drasticamente da un utilizzo pari allo 0% (non viene eseguita alcuna operazione) a un utilizzo valutato intorno al 100% (tutti i processori a disposizione vengono utilizzati a pieno carico di lavoro). Questa situazione è particolarmente critica nelle università quando l'impiego delle risorse si impenna in prossimità della fine dei semestri e prima dei convegni molto importanti.

L'elaborazione di dati finanziari può riguardare operazioni *ad hoc* (difficili da programmare nel tempo) oppure operazioni che si ripetono con una certa regolarità. Casi tipici che possono essere risolti adottando un sistema di cloud computing sono per esempio l'elaborazione dei pagamenti mensili oppure l'analisi quotidiana della registrazione degli accessi web. Un sito web di grandi dimensioni e molto frequentato può generare ogni ventiquattrore file di log che occupano decine di gigabyte. La quantità di informazioni che può essere estratta da questi file di log giustifica il fatto che la loro analisi costituisce un compito particolarmente delicato. L'accesso puntuale ai dati relativi all'utilizzo del sito permette di ottimizzare la gestione del sito stesso e di valutare più rapidamente gli interventi da effettuare in risposta a modifiche e proiezioni riguardanti l'utilizzo delle risorse web. L'analisi quotidiana dei dati tende a impiegare sempre più tempo e a un certo punto può arrivare a occupare l'intero arco delle ventiquattrore. In questo caso ha senso valutare la possibilità di adottare soluzioni di elaborazione dati in parallelo, che comportano l'utilizzo di più risorse per un tempo inferiore: una situazione ideale per il cloud computing.

Elaborazione in overflow

Le aziende che iniziano a comprendere i vantaggi offerti dal cloud computing si mettono presto alla ricerca di soluzioni ibride, che consentano da un lato di utilizzare le risorse IT già esistenti per svolgere il lavoro di routine e dall'altro trasferiscano il lavoro aggiuntivo nella nuvola. La situazione è analoga all'assunzione di lavoratori temporanei da utilizzare nei periodi di alta stagione.

L'elaborazione in overflow permette alle aziende di adattarsi gradualmente al cloud computing. Con il tempo aumenta il livello di confidenza nei confronti di questa tecnologia innovativa, e la quantità di dati aziendali che si riversa sempre più nella nuvola permette alle aziende di studiare soluzioni diverse e sempre più articolate.

Riepilogo

Esistono molti modi che consentono di utilizzare il cloud computing per l'hosting di applicazioni esistenti, per definire nuove soluzioni e per migliorare l'efficienza e l'efficacia in termini economici di organizzazioni grandi e piccole.

In questo capitolo sono stati introdotti i concetti fondamentali del cloud computing. Un'analogia sportiva ha permesso di valutare il modo in cui il cloud computing permette a singoli utenti e organizzazioni di stabilire un equilibrio ideale tra risorse disponibili e domanda effettiva di servizi web. È stato introdotto il concetto di "disastro da successo" e si è visto come evitare di ritrovarsi in una condizione di questo tipo sfruttando le possibilità offerte dai sistemi AWS. Sono state poi illustrate le caratteristiche del cloud computing e si è visto che questa tecnologia può essere considerata alla stregua di un data center programmabile. Le caratteristiche del cloud computing sono state classificate in tre categorie: generali, economiche e tecniche. Sono stati illustrati gli equivoci che circondano l'idea di cloud computing. Infine, sono stati spiegati i modelli di utilizzo e i casi tipici di impiego del cloud computing.

Nel prossimo capitolo vedremo la soluzione di cloud computing di Amazon Web Services e ci si preparerà a scrivere le prime righe di codice inerenti il cloud computing.