

Capitolo 1

Introduzione

Il cloud computing è un modello di distribuzione dei servizi informatici che consente di accedere a risorse e applicazioni tramite Internet, senza doverle mantenere localmente sul proprio computer o server. Attraverso il cloud computing, le risorse come l'archiviazione dei dati, la potenza di calcolo e il software vengono fornite come servizi virtuali da parte di fornitori specializzati noti come provider di cloud (Amazon, Google ecc.). Questi provider gestiscono l'infrastruttura fisica e mettono a disposizione degli utenti le risorse necessarie in base alle loro esigenze. In questo modo, le aziende possono evitare di investire in costosi hardware e software, riducendo i costi operativi e aumentando la flessibilità.

1.1 Cloud Computing

Possiamo distinguere diverse tipologie di servizi cloud in base al tipo di risorse che vengono fornite:

1.1.1 Infrastructure as a Service

L'Infrastructure as a Service (IaaS) è un modello di cloud computing che fornisce risorse informatiche virtualizzate tramite Internet. Con l'IaaS, i provider di cloud mettono a disposizione degli utenti l'infrastruttura fisica necessaria, inclusi server virtuali, storage, reti e altre risorse, consentendo loro di creare e gestire l'ambiente informatico in modo flessibile e scalabile. Attraverso l'IaaS, gli utenti possono evitare di dover investire in hardware e infrastrutture costose, riducendo i costi di gestione e manutenzione. I provider di cloud si occupano dell'acquisizione, dell'installazione e della gestione dell'hardware, nonché della fornitura delle risorse virtuali agli utenti. Questo

modello consente alle aziende di concentrarsi sullo sviluppo delle proprie applicazioni e servizi, piuttosto che preoccuparsi dell'infrastruttura sottostante. La peculiarità sta nel fatto che le risorse vengono istanziate su richiesta o domanda di una piattaforma ha bisogno.

1.1.2 Platform as a Service

Il Platform as a Service (PaaS) offre un ambiente di sviluppo e di esecuzione completo per le applicazioni. I provider di cloud mettono a disposizione degli sviluppatori un insieme di strumenti, framework e servizi che semplificano il processo di sviluppo, test e distribuzione delle applicazioni. PaaS offre un'ampia gamma di strumenti di sviluppo, come linguaggi di programmazione, framework e ambienti di sviluppo integrati (IDE), che semplificano il processo di sviluppo delle applicazioni. Gli sviluppatori possono scrivere il codice, testare e distribuire le applicazioni direttamente nell'ambiente fornito dal PaaS, senza dover configurare manualmente l'infrastruttura. Alcuni esempi di questo modelli li troviamo in Google App Engine, Microsoft Azure App Service e AWS Elastic Beanstalk.

1.1.3 Software as a Service

Il Software as a Service(SaaS) é modello di cloud computing che fornisce agli utenti applicazioni basate su cloud attraverso Internet. Con il SaaS, i provider di cloud ospitano e gestiscono l'infrastruttura e i software applicativi, consentendo agli utenti di accedere e utilizzare le applicazioni tramite Internet, senza dover installare il software sul proprio computer. Consiste nell'utilizzo di programmi installati su un server remoto, cioè fuori del computer fisico o dalla LAN locale, spesso attraverso un server web; l'utente può accedere al programma tramite un browser web, come se fosse un programma installato localmente. I modelli di pagamento del SaaS sono spesso basati sul consumo effettivo delle risorse, consentendo agli utenti di pagare solo per ciò che effettivamente utilizzano. Questo modello di pricing basato su abbonamento o utilizzo può essere vantaggioso per le aziende, in quanto consentono di evitare costi iniziali elevati e di prevedere meglio i costi operativi. Alcuni esempi più comuni di SaaS li troviamo in Google Workspace, Microsoft Office 365, Dropbox.

1.1.4 Function As a Service

Function as a Service(FaaS) é un modello di cloud computing che consente agli sviluppatori di eseguire e gestire le proprie funzioni senza dover gesti-

re l'infrastruttura sottostante. In FaaS, gli sviluppatori suddividono le loro applicazioni in funzioni modulari e indipendenti. Ogni funzione esegue un'attività specifica e può essere attivata in risposta a eventi o richieste specifiche. Ad esempio, una funzione può essere scatenata da un evento di caricamento di un file su un sistema di archiviazione cloud, da una richiesta HTTP o da un timer programmato. Quando una funzione viene attivata, il fornitore di servizi cloud gestisce automaticamente la sua esecuzione, inclusa la gestione delle risorse necessarie. Le funzioni vengono eseguite in ambienti isolati e scalati automaticamente in base alle richieste di carico. Una volta completata l'esecuzione della funzione, le risorse vengono deallocate per massimizzare l'efficienza e minimizzare i costi. FaaS è diventato un'opzione popolare per lo sviluppo di microservizi, serverless application e scenari di elaborazione event-driven, fornendo un modo flessibile ed efficiente per eseguire singole funzioni di codice. Alcuni esempi di FaaS li troviamo in AWS Lambda, Google Cloud Functions, Azure Functions.

Serverless Computing

Serverless computing è un paradigma più ampio del FaaS in cui il fornitore di servizi cloud gestisce completamente l'infrastruttura, mentre FaaS è un sottoinsieme del Serverless computing che si concentra sulla gestione delle funzioni come entità indipendenti. Le funzioni vengono eseguite in modo scalabile e senza la necessità di gestire l'infrastruttura sottostante. [5]

1.2 Obiettivo della tesi

Come già detto nell'ambito del serverless computing, il provider cloud si occuperà di allocare le risorse necessarie per eseguire il codice e di deallocarle una volta terminata l'esecuzione. Questo facilita la distribuzione e la scalabilità del sistema, cercando di automatizzare quest'ultima fase. Il sistema comunque gestirà le risorse attraverso gli elaboratori che al momento saranno a disposizione, attraverso un Allocation Priority Policy (APP) [2]. APP è un linguaggio dichiarativo che permette la configurazione delle politiche di pianificazione per l'esecuzione delle funzioni. Vediamo qui di seguito un esempio del funzionamento:

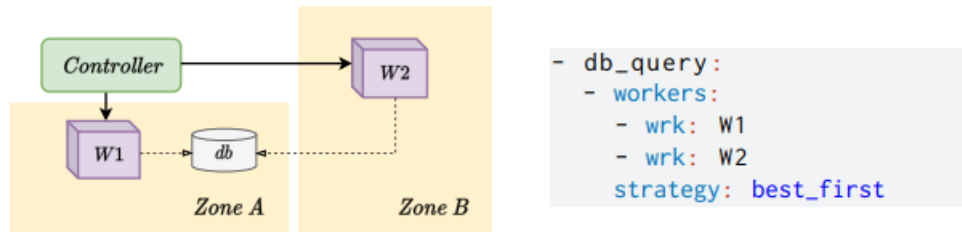


Figura 1.1: Esempio di funzionamento di un sistema serverless [3]

A destra troviamo il codice dichiarativo APP che fornisce le regole per allocare le risorse: abbiamo due **Workers** che sono disponibili, e una strategia che regola come allocare queste risorse nei rispettivi worker; inoltre una **Strategy** che determina come allocare le risorse e secondo quale priorità. Attualmente però la strategy ci permette di allocare secondo poche strategie generiche già consolidate. L'obiettivo diventa ora automatizzare questo processo, adottando procedure automatiche per definire politiche di allocazione delle risorse basate sulle informazioni derivate dall'analisi statica delle funzioni che dovranno essere eseguite.[3] In questa tesi ci concentreremo sulle tecniche che ci permettono, data una funzione scritta in un linguaggio con una grammatica definita nei capitoli successivi, di analizzarne la semantica e da lì estrarre misurazioni quali la complessità (come upper bound di una funzione) al fine di poter ottimizzare lo scheduling dell'esecuzione delle funzioni serverless. Procederemo andando a definire una grammatica per il linguaggio, in seguito illustreremo un interprete per il linguaggio che analizzando un programma in input, ci restituirà le equazioni di costo per ogni funzione. In seguito andremo ad analizzare le equazioni di costo per ogni funzione, attraverso tool(PUBS [1] e CoFloCo[4]) specifici che data un'equazione di costo ci definisce l'upper bound della una funzione.