

MIGRAZIONE IN CLOUD DEI MOODLE DELL'UNIVERSITÀ DI PADOVA

Angelo Calò, Davide Ferro

Università degli Studi di Padova – Ufficio Digital Learning e Multimedia
{angelo.calò,davide.ferro}@unipd.it

— COMUNICAZIONE—

ARGOMENTO: *Aspetti tecnici – Gestione Moodle di grandi dimensioni*

Abstract

In questo contributo saranno illustrate le motivazioni che hanno portato l'Università degli Studi di Padova alla migrazione in cloud delle proprie istanze Moodle, gli aspetti organizzativi necessari a garantire la continuità del servizio offerto agli oltre 70.000 utenti delle oltre 40 istanze, le considerazioni sulla scelta del fornitore di servizi cloud ed infine le caratteristiche tecniche dell'architettura creata in cloud. Riteniamo che i numeri di utenti e istanze e le dimensioni totali della piattaforma di e-learning dell'Ateneo di Padova possano essere un interessante caso di analisi per capire vantaggi e svantaggi del cloud rispetto alla gestione in casa dell'intera infrastruttura.

Keywords – moodle, cloud, migrazione.

1 LA SITUAZIONE A LUGLIO 2021

Dopo quasi un anno e mezzo di emergenza da pandemia di COVID-19, l'infrastruttura gestita interamente on premise delle numerose istanze Moodle dell'Università di Padova ha cominciato a dare chiari segnali di inadeguatezza nella capacità di sostenere il carico di utenti e richieste contemporanee. L'utilizzo delle risorse, a causa dei lockdown e della trasformazione delle attività didattiche, è moltiplicato a partire da marzo 2020 di un fattore 20 o 30 rispetto al già elevato utilizzo precedente.

La situazione è stata monitorata costantemente fin dall'inizio della pandemia e delle relative restrizioni atte a limitare i contagi, che hanno in particolare segnato prima lo stop della didattica frontale in presenza e poi hanno portato alla gestione "duale" dell'erogazione delle lezioni e degli esami universitari. Fin da subito è stato chiaro che tutti i servizi avrebbero avuto un incremento di richieste e di accessi contemporanei mai visto prima e per il quale l'infrastruttura attuale avrebbe dovuto essere rivista.

L'infrastruttura on premise di gestione di Moodle per l'Università di Padova risulta composta da:

- 49 istanze di Moodle, la maggior parte delle quali dedicate alle strutture didattiche (32 Dipartimenti e 8 Scuole), con l'aggiunta di istanze riservate ai Centri e ai Servizi dell'amministrazione centrale;
- circa 20 macchine virtuali, tra cui 12 application server (web server che erogano le risorse dei Moodle), database, storage server, cache Redis, load balancer, proxy;
- uno storage NFS condiviso da circa 12 TB;
- un sistema di backup per database e storage.

L'architettura pensata in fase di progettazione ha permesso senza particolari problemi di gestire quello che fino a marzo 2020 è stato il tipico volume di richieste e utenti, essendo sicuramente sovradimensionato in termini di capacità di calcolo e risorse disponibili. Si tratta fino all'inizio della pandemia di circa 15'000 utenti giornalieri, di cui 3'000 contemporanei.

A marzo 2020 tutte le attività della didattica si riversano sui servizi online e gli utenti giornalieri dei Moodle passano a 115'000, di cui 27'000 contemporanei, e questi utenti (che in termini assoluti sono incrementati rispettivamente di 7 e 9 volte) risultano molto più attivi, portando un incremento totale delle richieste ai web server dalle 20 alle 30 volte rispetto ai numeri precedenti (da poco più di 2 milioni di richieste giornaliere a mediamente 60 milioni).

2 PRIMI TEST SUL CLOUD

Il cloud non è di per sé la soluzione ad ogni problema, questo era ed è chiaro tuttora. Ma è altrettanto chiaro che permette una elasticità nel considerare e utilizzare le risorse che la tradizionale gestione "in casa" non permette.

Partendo da questa considerazione, l'Ufficio Digital Learning e Multimedia, in considerazione dei brevi tempi a disposizione per risolvere i problemi di sovraccarico dell'infrastruttura in casa, ha sviluppato un piano di riorganizzazione e migrazione dei servizi Moodle. In primo luogo, si è sfruttata l'occasione derivata dalla necessità di gestire gli esami a distanza (unica modalità prevista per la sessione estiva dell'A.A. 2019/2020), utilizzata come banco di prova per testare la fattibilità di un utilizzo intensivo e prevalente di un'infrastruttura in cloud gestita però in maniera autonoma (non quindi come Software as a Service).

In base all'analisi di mercato, delle conoscenze pregresse, della qualità del supporto dedicato e delle attuali convenzioni CRUI, la scelta del provider di servizi cloud è ricaduta su Amazon Web Services (AWS). Non ultimo fattore determinante, la posizione nettamente predominante di AWS nel *Magic Quadrant* 2019 di Gartner [1], in un mercato che vede pochissime alternative.

La progettazione dell'infrastruttura, relativa alla gestione di una sola istanza Moodle dedicata esclusivamente alla gestione degli esami on line, ha permesso di prendere confidenza con i 2 aspetti specifici di una architettura in cloud: l'uso di componenti definite *managed*, ovvero gestite in maniera più immediata e semplice da parte dell'utente come SaaS, in quanto oggetti in gran parte gestiti dal provider cloud in tutte le procedure che normalmente coinvolgono l'amministratore di sistema; l'elasticità con cui considerare le risorse utilizzate nell'architettura.

Il supporto di AWS, nello specifico il team italiano dell'ufficio Education & Research, ha guidato la creazione di questa prima infrastruttura permettendo di ottimizzare le risorse e individuare i componenti e l'architettura generale in grado di gestire correttamente le peculiarità di Moodle.

A seguito della positiva reazione degli utenti (docenti e studenti) e alla stabilità dimostrata anche nei momenti di maggior carico durante le sessioni d'esame, il test di Moodle Esami sul cloud è stato considerato un successo, determinando quindi la decisione di spostare tutte le altre piattaforme della didattica e dei servizi su AWS, avendo ricevuto nel frattempo anche altre positive esperienze da altri atenei occupati nella migrazione sul cloud dei loro Moodle (sia su AWS sia su Azure).

3 MIGRAZIONE IN CLOUD

Dalla fase di test della nuova istanza Moodle Esami su cloud, l'attività relativa alla gestione di Moodle del Settore Digital Learning si è concentrata soprattutto sulla risoluzione dei problemi contingenti (sovraccarico dell'infrastruttura, supporto all'utenza, sviluppo di soluzioni per rendere accessibili da remoto tutte le tecnologie per la didattica) e al contempo alla definizione di una roadmap per la razionalizzazione e lo spostamento in cloud dei Moodle on premise, passo reso sempre più evidente dal continuare della situazione di emergenza e dal crescente numero di accessi alla piattaforma ormai al limite del carico.

Solo con la chiusura delle attività didattiche del secondo semestre dell'A.A. 2020/2021 si sono presentate le circostanze per procedere con il piano di migrazione delle istanze Moodle, con la loro razionalizzazione e riorganizzazione e con la definizione di una architettura dedicata su AWS, in grado di gestire il carico previsto per l'inizio del nuovo anno accademico.

Le attività hanno quindi previsto una fase di progettazione, definizione dell'infrastruttura e selezione dei componenti da utilizzare, con riferimento a quanto delineato nella fase di creazione dell'architettura per Moodle Esami.

Una particolare attenzione è stata data ad alcuni aspetti specifici di questo scenario, che al contrario di Moodle Esami prevedeva una migrazione di istanze preesistenti da on premise al cloud:

- il trasferimento di una grossa mole di dati (circa 12TB) relativi allo storage delle istanze;
- il backup e trasferimento dei database (circa 400GB);
- il collegamento con servizi on premise da non trasferire su AWS, come ad esempio il Single Sign On,

il tutto mantenendo l'usabilità e rendere il processo il più trasparente possibile per gli utenti finali.

3.1 La nuova infrastruttura

Innanzitutto, è stata quindi definita l'architettura della nuova infrastruttura, con l'obiettivo di renderla scalabile, performante e sicura, attraverso la ricerca delle soluzioni e delle componenti AWS maggiormente adatte alle caratteristiche di Moodle. A questo scopo è stata fatta un'analisi approfondita soprattutto sulle componenti:

- DBMS;
- sistema di caching;
- filesystem per lo storage.

Il risultato dell'analisi, relativa a performance, stabilità e facilità di gestione per grossi volumi di carico e un numero elevato di istanze Moodle indipendenti e ognuna con caratteristiche particolari e personalizzazioni ha portato a definire l'architettura semplificata con l'indicazione di massima delle componenti di Figura 1.

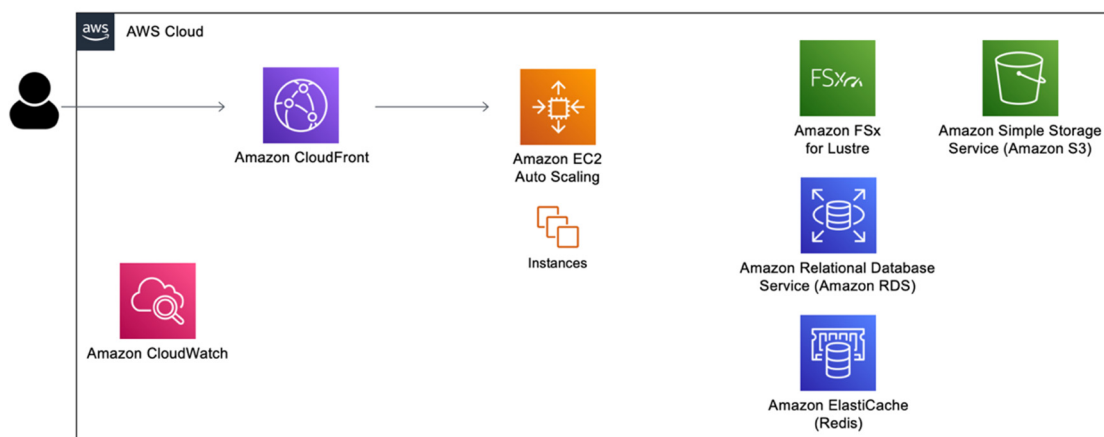


Figura 18 - Architettura semplificata su AWS

Lo schema illustra sinteticamente come le richieste dell'utente dei vari Moodle vengono gestite:

1. i domini dell'Università di Padova dove risiedono le istanze sono serviti tramite una CDN *CloudFront*, che gestisce in maniera automatica ed efficiente le richieste che provengono da tutto il mondo, facendole transitare dal punto geografico più vicino all'utente all'interno della rete proprietaria AWS fino alla "regione" in cui l'infrastruttura risiede (nello specifico in Europa), abbassando la latenza e quindi il ritardo in caricamento delle richieste;
2. la CDN gira le richieste internamente ad un componente (completamente gestito da AWS) *Load Balancer*, che le smista in base al carico all'interno di un gruppo di macchine virtuali (*EC2*) che operano come application server;
3. in base a delle regole personalizzate, in caso di carico elevato negli attuali application server un meccanismo di *Auto Scaling* provvede ad accendere nuove EC2 e quindi rende disponibili al Load Balancer un numero maggiore di macchine virtuali su cui suddividere il carico di richieste. Allo stesso modo, in caso di carico molto basso, l'Auto Scaling provvede a spegnere le macchine virtuali che attualmente sono inutili;
4. a livello applicativo, gli application server eseguono le chiamate interne per il codice PHP e i comandi richiesti a Moodle, e utilizzano connessioni a componenti separate per l'accesso al database (Amazon RDS) e la gestione della cache (Amazon ElastiCache – Redis);
5. per lo storage del codice di Moodle e dei dati è stato scelto, dopo approfonditi test di performance, come file system condiviso Amazon FSx for Lustre [2].

3.2 Altri componenti

A scopo semplificativo nello schema dell'architettura riportato sono stati omessi numerosi dettagli che riguardano l'infrastruttura e il funzionamento di alcuni dei componenti, in particolare:

- l'uso delle *Availability Zone* all'interno della Region, in modo da garantire l'alta affidabilità dei componenti;
- CloudWatch per il monitoraggio e la definizione di allarmi di malfunzionamento o situazioni da verificare;
- S3 per lo storage di dati non live;
- Route53 per la gestione delle *Hosted Zone*;
- Certificate Manager per la gestione trasparente dei certificati SSL/TLS

3.3 Passi della migrazione

La migrazione di un numero così elevato di Moodle, con la relativa mole di dati relativi allo storage e ai database, ha spinto a organizzare lo spostamento in maniera graduale, con un calendario concordato con i vari referenti dei Moodle che si è sviluppata da luglio a dicembre 2021.

Ogni istanza è stata trasferita garantendo il minor tempo di inutilizzabilità possibile, limitando quindi l'uso delle risorse on premise per l'esportazione e il trasferimento dei dati verso AWS, garantendo nel frattempo il pieno utilizzo di tutte le altre istanze non coinvolte dalla migrazione e sfruttando le ore di minor utilizzo delle piattaforme. Questo modo di agire ha inoltre permesso di procedere ad alcune procedure di pulizia e di aggiornamento, garantendo di potersi dedicare ad una o al massimo due istanze per volta.

Un grosso problema, risolto però dalle potenzialità di CloudFront (la CDN di AWS), è stato il fatto che quasi tutti i Moodle risiedono su un unico dominio di terzo livello (elearning.unipd.it) e la gestione simultanea di Moodle su AWS e on premise avrebbe creato problemi di instradamento. Tramite la definizione di *Origins* e *Behaviors* diversi su CloudFront è possibile gestire questa situazione reindirizzando su AWS o on premise in base alle richieste, e quindi in base al Moodle richiesto.

4 VANTAGGI

4.1 I vantaggi attesi

Da tempo è chiaro ormai quali siano i vantaggi generali delle infrastrutture in cloud, lo stesso Piano Triennale (2020-2022) per l'Informatica nella Pubblica Amministrazione [3] promuove il *Cloud Enablement*, secondo i principi di semplificazione, consolidamento e razionalizzazione delle attività IT.

Per questo si attendevano i vantaggi che effettivamente sono stati riscontrati dal momento della migrazione delle prime piattaforme a luglio 2021:

- una gestione molto più rapida dei componenti/servizi, dalla creazione di un database alla modifica di una configurazione di rete che in casa avrebbe richiesto una modifica sostanziale da propagare su vari apparati e macchine virtuali;
- una scalabilità reale, sia verticale (numero di servizi/macchine) che orizzontale (dimensionamento del servizio/macchina). Un esempio su tutti è l'andamento del numero di application server gestiti tramite l'Auto Scaling durante la settimana, in Figura 2 è riportato il grafico che descrive il numero di EC2 durante le giornate dell'ultima settimana di novembre: in base alle politiche definite per gli application server si nota immediatamente la flessibilità e la reattività di un sistema basato sull'effettivo carico e che può scalare in modo automatico per gestire i picchi di richieste. La riga rossa identifica il numero di application server disponibili on premise;
- una *fault tolerance* in grado di ripristinare un problema su un componente in pochi minuti, in maniera completamente trasparente;
- la facilità di creazione di ambienti di test (e anche poi di produzione), grazie anche all'indipendenza dall'hardware, con costi di avvio estremamente limitati;
- un monitoraggio integrato nei servizi AWS;
- maggiore sostenibilità ambientale (AWS dichiara un impatto 3.6 volte minore dei data center tradizionali on premise).

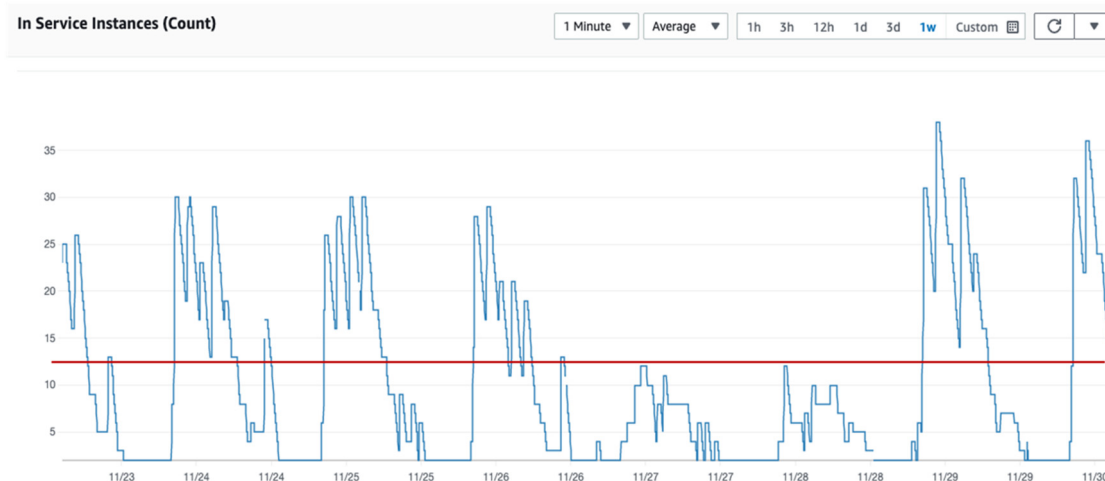


Figura 19 – Andamento del numero di application server tramite Auto Scaling

4.2 I vantaggi inattesi

Oltre a tutto quello che ci aspettava, alcuni risvolti pratici vantaggiosi sono stati in parte una sorpresa. Tra questi i principali sono stati:

- una visione d'insieme dell'infrastruttura con maggiore dettaglio e chiarezza, con la possibilità inoltre di utilizzare l'*Infrastructure as Code* per definire e riproporre intere architetture;
- una segmentazione del networking molto più semplice e sicura;
- una gestione semplificata dell'automatizzazione di alcune procedure tramite API e la possibilità di monitorare con maggiore dettaglio e facilità le risorse.

Infine, la migrazione in sé e l'intero processo di razionalizzazione ha portato in corso d'opera a sviluppare un sistema su architettura ELK per un monitoraggio tramite dashboard ancora più semplici e intuitive delle metriche di utilizzo delle risorse e della suddivisione tra i vari Moodle.

5 CONCLUSIONI

La migrazione in cloud ha senza dubbio risolto un problema contingente, innescato dalla pandemia, ma ha in realtà solo accelerato un processo già in atto di conversione e ammodernamento delle logiche di gestione di un sistema complesso e delicato come l'erogazione di strumenti tecnologici per la didattica universitaria di un Ateneo di grandi dimensioni.

Alla fine di novembre 2021 sono stati registrati questi accessi mensili: più di 2 miliardi di richieste agli application server, 30 TB di traffico totale, 600'000 utenti unici (numeri parziali con la migrazione completata al 90% circa), il tutto con una drastica riduzione dei fermi dovuti a sovraccarico o errori, e la riduzione del tempo di risoluzione dei problemi.

Il progetto di migrazione è sicuramente stato un successo, in grado di mostrare le potenzialità del cloud e della necessaria riorganizzazione delle logiche di progettazione e sviluppo di nuove applicazioni o la conversione delle esistenti in ottica di ottimizzazione e razionalizzazione.

Riferimenti bibliografici

[1] Gartner (2019, Luglio 16), Gartner Magic Quadrant for Cloud Infrastructure as a Service, Worldwide: <https://www.gartner.com/en/documents/3947472>

[2] Amazon FSx for Lustre: <https://aws.amazon.com/fsx/lustre/>

[3] AGID, Agenzia per l'Italia digitale: <https://pianotriennale-ict.italia.it/>