

Quando Amazon S3 non basta: i servizi AWS per lo storage gestito

28 Aprile 2023 - 8 min. read

[Amazon FSx for NetApp ONTAP](#)

[Amazon FSx for Windows](#)

[Amazon S3](#)

[Amazon S3 File Gateway](#)

Citando Douglas Adams nella "Guida galattica per autostoppisti", se 42 è la risposta alla domanda fondamentale sulla vita, l'universo e tutto quanto, Amazon S3 lo è per tutti i quesiti riguardanti lo storage su AWS.

O forse no?

Come vedremo, purtroppo, non è una verità assoluta: a volte occorre usare qualcosa di differente. D'altronde si sa, il mondo è bello perché è vario.

Quando si ha a che fare con ambienti ibridi, progetti lift and shift e workload Windows, non è raro trovare cartelle condivise utilizzate da utenti e servizi. Numerose soluzioni software commerciali fanno affidamento a questa tecnologia; in alcuni casi, il refactoring non è una via praticabile.

Come suggerito dal nome, una condivisione di file non è equivalente a un object storage: ha proprietà, comportamenti e scenari di utilizzo diversi. In questo articolo, esamineremo le altre opzioni per migrare e adattare i carichi di lavoro sul Cloud.

Per prima cosa, vediamo di chiarire le differenze chiave fra gli **storage ad oggetti ed i file**.

Lo storage basato su file corrisponde alla nostra idea "tradizionale": a seconda del sistema operativo (Windows, Linux, macOS, i dati vengono memorizzati in una struttura gerarchica e sono identificati da nome (file) e percorso (directory). I metadati (come permessi e proprietà dei file) sono memorizzati separatamente, e il loro design dipende dal filesystem in uso.

Lo storage ad oggetti è diverso: tutti i dati (inclusi metadati e proprietà) sono memorizzati in un namespace piatto. L'accesso è fatto tramite API e facendo riferimento a un identificatore. Una volta scritto l'oggetto, non c'è modo di aggiungere o eliminare porzioni dei dati come in un filesystem tradizionale, ma occorre scriverne una nuova versione per modificare l'oggetto.

Questo approccio rende lo storage ad oggetti più scalabile ed economico, anche se non è purtroppo possibile passare in modo indolore da una tecnologia ad un'altra.

Quali sono quindi le opzioni per un workload Windows che ha bisogno di cartelle condivise? Come diciamo sempre "non c'è una sola risposta possibile".

Amazon mette a disposizione 3 servizi differenti:

- Amazon S3 File Gateway
- Amazon FSx for Windows
- Amazon FSx for NetApp ONTAP

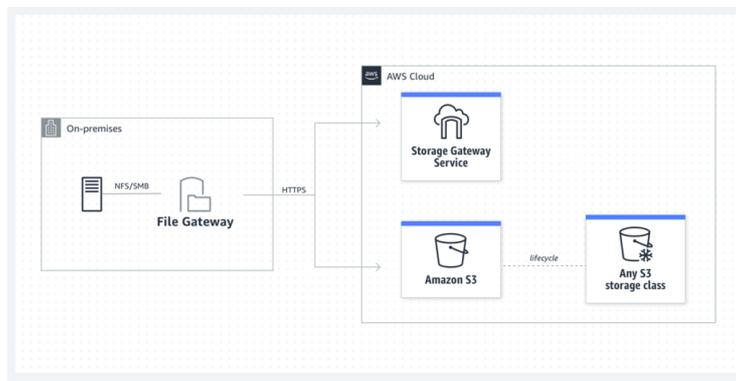
Vediamo ora cosa offrono e il loro caso d'uso.

Amazon S3 File Gateway

Amazon S3 File Gateway combina il mondo dello storage ad oggetti e quello basato file. Potremmo pensare che metta a disposizione uno storage scalabile illimitato con i famosi "Undici Nove" di durabilità. Sfortunatamente, come vedremo, non è tutto oro quel che luccica.

Con Amazon S3 File Gateway, è possibile usare il protocollo SMB per memorizzare i file in S3, sfruttando la sua scalabilità e traducendo automaticamente le chiamate API per la memorizzazione.

Per abbassare i costi di storage e archiviare o eliminare tutti i file è anche possibile usare le **lifecycle policy**. Non ci sono costi di licenza, ed è possibile utilizzare la soluzione on-premise per velocizzare l'accesso ai file mediante l'uso della cache interna.



Questa soluzione ha alcuni svantaggi: modificare file di dati di grandi dimensioni crea nuove versioni dell'oggetto S3 ogni volta, influenzando le prestazioni e i costi.

C'è un limite non modificabile di 50 condivisioni per appliance Amazon S3 File Gateway e, inoltre, Amazon S3 File Gateway usa i metadati degli oggetti memorizzati su Amazon S3 per mappare gli attributi del filesystem. Questo comporta che le File Access Control Entries (ACE), debbano essere memorizzati nei metadati, che hanno però una dimensione limitata a 2KB.

La conseguenza di questo meccanismo è l'errore

```
"ERROR 1344 (0x00000540)"
```

quando si utilizza robocopy per trasferire file che hanno più di 10 Access Control Entries.

(Vedi [Troubleshooting: File Gateway issues](#))

Quando si raggiunge questo limite si hanno 2 possibilità:

1. Riorganizzare i permessi relativi ai file
2. Usare una soluzione AWS differente.

Siccome non esiste nessun servizio AWS che permetta di riorganizzare i permessi (si tratta infatti di una antica arte padroneggiata dai sysadmin :D), descriveremo le altre soluzioni che permettono di superare queste limitazioni.

Amazon FSx for Windows

Amazon FSx for Windows mette a disposizione un fileserver Windows ad alte prestazioni completamente gestito e compatibile con gli strumenti standard di amministrazione Windows. Può essere configurato in modalità Multi-AZ per essere altamente disponibile, anche durante le operazioni di patch (gestite da AWS).

Supporta tutte le funzionalità che una soluzione nativa può offrire, come copie shadow, deduplicazione dei dati e namespace DFS, mentre delega le operazioni di gestione e automatizza la crittografia usando AWS KMS. Usando PowerShell e commandlet sviluppati ad-hoc, è possibile automatizzarne la gestione e migrare in modo semplice le condivisioni esistenti.

Inoltre, è possibile abilitare i backup automatici, usare AWS Backup con diverse pianificazioni e politiche di conservazione dei dati ed avere un appliance di cache on-premise (Amazon Storage Gateway for FSx) per velocizzare i tempi di accesso riducendo il consumo di banda.

Come sempre, c'è un rovescio della medaglia: Amazon FSx per Windows mette a disposizione due diversi livelli di storage: SSD e HDD, senza che sia possibile passare da uno all'altro in modo trasparente. Il tipo di storage è impostabile solo alla creazione di un filesystem Amazon FSx for Windows.

Per ridurre i costi di storage, l'auto-tiering non è un'opzione in questo scenario. Ovviamente c'è un altro servizio per quello!

Amazon FSx for NetApp ONTAP

Amazon FSx for NetApp ONTAP è sempre un servizio gestito che utilizza storage SSD ad alte prestazioni offrendo tutte le funzionalità di un'appliance ONTAP on-premise, come:

- Auto-tiering: i dati acceduti meno frequentemente possono essere spostati automaticamente da storage SSD a HDD
- Accesso multi-protocollo: oltre alle condivisioni Windows SMB, offre volumi NFS e iSCSI).
- È possibile usare gli stessi strumenti di gestione standard
- Snapshot (anche tra regioni differenti) e ripristino di singoli file (file-level restore)

Oltre a tutte queste funzionalità, sul Cloud AWS sono supportate configurazioni Multi-AZ ed è possibile usare IAM per assegnare i permessi alle risorse. Come Amazon FSx for Windows, Amazon applica automaticamente patch e aggiornamenti durante la finestra di manutenzione definita e personalizzabile.

Anche se sembra di aver trovato la soluzione per per tutti gli scenari cloud ibridi ci sono alcuni "ma":

- La capacità minima per un filesystem Amazon FSx for NetApp ONTAP è di 1 Terabyte.

- La curva di apprendimento può essere ripida se non si ha esperienza con questo tipo di storage.
- Occorre considerare attentamente i costi relativi al traffico per il backup ed il trasferimento dei dati.

Trasferimento dei dati

È possibile trasferire i dati usando il buon vecchio RoboCopy o AWS DataSync. Entrambi sono soluzioni affidabili e robuste per la copia dei file. Si possono riutilizzare tutti gli script e le opzioni da riga di comando per robocopy imparate con il sudore durante gli anni, senza modificare o dover imparare nulla da capo.

Scelta del servizio più adatto

Come ripetiamo sempre, nessuna soluzione può soddisfare tutte le esigenze, infatti non abbiamo ancora parlato di **ottimizzazione dei costi**: è ora di aggiungere questa variabile e prendere in considerazione uno scenario di esempio, confrontando costi, prestazioni e alta disponibilità.

Supponiamo di avere 5 Terabyte di dati usati come archivio, di cui solo circa 500 Gigabyte (il 10%) sono acceduti più frequentemente.

Amazon S3 File Gateway

Con Amazon S3 File Gateway, si ha una cache locale on-premise, mentre lo storage è ospitato da S3. Poiché si tratta di una singola appliance, non si ha alta disponibilità, anche se i dati saranno sempre al sicuro e recuperabili perché memorizzati su Amazon S3. Grazie a questo, possiamo usare S3 Intelligent-Tiering

Le prestazioni di accesso ai file saranno buone per i file già presenti in cache, ma, alla richiesta di nuovi file, la latenza aumenterà perché è necessario trasferire i dati. Inoltre saranno necessari almeno 100 Gigabyte di spazio di archiviazione on-prem per la cache.

Usare un appliance locale on-premise non influisce sulla bolletta AWS, ma è bene ricordare che nessun hardware è privo di TCO.

Usando **S3 Intelligent-Tiering** i costi di archiviazione vengono automaticamente ridotti. Supponendo anche che tutti i 5 TB di dati siano acceduti almeno una volta al mese, il costo sarà di circa 70\$ / mese.

Amazon FSx per Windows

Amazon FSx for Windows, come abbiamo detto prima, non permette di fare auto-tiering dei dati; poiché la frequenza di accesso non è conosciuta, abbiamo assunto che tutti i 5 TB di dati siano memorizzati su storage SSD. Usando quindi una configurazione multi-az ed assumendo un fattore di deduplicazione del 50% il costo si può aggirare a circa 687\$/mese.

Scegliendo avere meno prestazioni ed usando quindi storage HDD standard, il costo si aggira a circa 103\$ al mese.

Come Amazon S3 File Gateway, è possibile avere un appliance on-premise per la cache dei file: tutte le considerazioni precedenti sono valide, ma ci saranno dei costi orari aggiuntivi, anche ospitando la cache in un server on-premise.

Amazon FSx per NetApp ONTAP

Assumendo lo stesso fattore di deduplica e con un deploy multi-az, il costo toale è di 581\$ al mese

Di seguito un piccolo specchietto di riassunto:

Monthly cost	Description	Config Summary
687.28 USD	FSx all on SSD with 50% deduplication savings	Desired storage capacity (5 TB), Desired aggregate throughput (0 M...
581.56 USD	Fsx for NetApp ONTAP	Desired aggregate throughput (1 MBps), Desired storage capacity (...)
70.20 USD	Storage Gateway with S3 Intelligent-Tiering cost	S3 INT Average Object Size (16 MB), S3 INT storage (5 TB per month)

I costi delle soluzioni cambiano molto, ma le caratteristiche e le prestazioni sono completamente diverse.

Amazon Storage Gateway è economico, ma senza alta disponibilità, mentre le configurazioni che abbiamo esaminato per Amazon FSx for Windows e Amazon FSx for NetApp ONTAP sono in una configurazione multi-AZ. L'auto-tiering è prezioso e può aiutare a ridurre le bollette, ma con NetApp ONTAP si paga anche per altre funzionalità di valore come snapshot, backup e condivisione multi-protocollo.

Se il workload Windows, una volta migrato, deve essere rifattorizzato usando tecnologie Cloud-native, è possibile usare Amazon S3 File Gateway. Una volta modificata l'applicazione sarà possibile usare quindi S3 senza migrare i dati.

Se invece occorre usare comunque dei servizi di condivisione di file nativi Windows, Amazon FSx for Windows permette di ridurre i costi di gestione.

Amazon for NetApp ONTAP si adatta perfettamente ad uno scenario cloud ibrido, offrendo tutte le funzionalità già usate on-premise.

C'è molta scelta per i servizi di storage e, a seconda delle esigenze, la bolletta mensile può aumentare notevolmente. Come sempre, usare soluzioni Cloud-native (come Amazon S3), aiuta a migliorare la scalabilità, la resilienza e a ridurre i costi, ma occorre considerare le esigenze e trovare la soluzione perfetta.

Migrare i dati può essere anche complicato. Usate già questi servizi? Se avete qualche storia interessante sulle migrazioni, condividetela nei commenti!

About Proud2beCloud

Proud2beCloud è il blog di **beSharp**, APN Premier Consulting Partner italiano esperto nella progettazione, implementazione e gestione di infrastrutture Cloud complesse e servizi AWS avanzati. Prima di essere scrittori, siamo Solutions Architect che, dal 2007, lavorano quotidianamente con i servizi AWS. Siamo innovatori alla costante ricerca della soluzione più all'avanguardia per noi e per i nostri clienti. Su Proud2beCloud condividiamo regolarmente i nostri migliori spunti con chi come noi, per lavoro o per passione, lavora con il Cloud di AWS. Partecipa alla discussione!



Damiano Giorgi

Ex sistemista on-prem, pigro e incline all'automazione di task noiosi. Alla ricerca costante di novità tecnologiche e quindi passato al cloud per trovare nuovi stimoli. L'unico hardware a cui mi dedico ora è quello del mio basso; se non mi trovate in ufficio o in sala prove provate al pub o in qualche aeroporto!
