Le insidie di AWS CloudTrail: Trappole Nascoste e Best Practices Essenziali

2 Aprile 2025 - 5 min. read

AWS CloudTrail

Intro

Quante volte ci è capitato di dover risalire ad azioni fatte in un ambiente AWS cercando di ricostruire l'accaduto e capire chi ha fatto cosa?

E quante volte abbiamo abbandonato l'investigazione senza cavare un ragno dal buco?

Molti di voi si saranno già visti in questa situazione, soprattutto coloro che lavorano su account AWS condivisi da team numerosi, o, addirittura, su account AWS dedicati a più progetti e su cui lavorano più consulenti esterni.

In alcuni casi, quando la governance di un ambiente AWS non è ben strutturata, bisogna ricorrere a dei metodi alternativi per poter risalire alla root cause di un problema. Per fare questo utilizzeremo uno dei miei servizi preferiti: AWS CloudTrail.

In questo articolo andremo ad analizzare qual è il setup tipico di AWS CloudTrail in un account/organization AWS, quali sono alcuni problemi tipici che potremmo incontrare nell'utilizzo quotidiano, e quali sono gli accorgimenti da prendere per poter facilitare il lavoro e superare i limiti di un setup standard.

Non c'è il Trail

Il primo dei problemi più comuni è l'assenza del Trail stesso.

Capita spesso di dover risalire all'ownership di una particolare risorsa AWS per poter chiedere più informazioni a riguardo. Purtroppo però capita spesso di non avere alcuna

evidenza nei log di AWS CloudTrail perché la risorsa è stata creata tempo prima; il servizio, infatti, tiene di default solo lo storico degli ultimi 90 giorni.

Lesson learned: *all'apertura di un account AWS, impostare sempre un Trail per tenere uno storico a lungo termine di tutte le attività nell'account.*

Setup iniziale

Secondo le best practice di AWS Cloudtrail, dovremmo attivare un Trail e predisporre un Bucket S3 per lo storage a lungo termine.

D'altro canto, secondo le best practice di Amazon S3 dovremmo anche prevedere delle lifecycle policy per variare le storage class gli oggetti memorizzati in base all'utilizzo che ne facciamo.

Per punti bonus possiamo anche creare una tabella Amazon Athena con lo schema di default suggerito da AWS per poter fare query più comodamente.

"Ok, sto facendo la mia query su Athena, ma non vedo risultati oltre una certa data"

Capita di voler investigare su un'azione accaduta in un particolare giorno; andiamo quindi a effettuare una query filtrando per quello specifico arco temporale e.... scopriamo che la query non presenta nessun risultato.

"*Strano...*" – average DevOps

Un occhio poco attento può trarre conclusioni affrettate riguardo all'assenza del dato cercato. Ma se andassimo a controllare i log direttamente nel bucket S3, vedremmo invece che quel dato esiste: i file sono stati messi nel tier Glacier Flexible Retrieval.

Lesson learned: *una query su una tabella Athena di default ignora tutti gli oggetti in S3 Glacier Flexible Retrieval e in S3 Glacier Deep Archive. Athena è in grado di effettuare query su questi file ma è necessario prima abilitare questa opzione per la specifica tabella che si vuole interrogare.*

"Sto cercando chi ha sovrascritto il mio record su Route53 ma non trovo niente".

Spesso cerchiamo delle azioni su particolari servizi di AWS (ad esempio IAM, Route53, ecc.), ma non troviamo niente.

Perchè?

Eppure abbiamo abilitato il Trail come da best practice.

Alcune azioni vengono tracciate solo nella region North Virginia perchè relative a servizi *global;* è il caso ad esempio di modifiche ai record Route53, creazione di IAM Roles, creazione di CloudFront Distributions, ecc...

Lesson learned: *creare il Trail con la feature multi-region abilitata e ricordarsi che i servizi AWS globali sono loggati in North Virginia.*

"Ho un sacco di log e le mie query sono molto lente, come faccio?"

La risposta è il partitioning.

L'alberatura usata da CloudTrail per memorizzare i log su S3 è:

```
s3://<bucket-name>/<optional-prefix>/AWSLogs/<account-id>/CloudTrail/
<region>/<year>/<month>/<day>/<log-file>.json.gz
```

quindi è già ben strutturata, ma questo non basta; per eseguire query in modo efficiente dobbiamo partizionare i dati anche su Athena e abbiamo due opzioni per farlo:

1. Definire le partizioni in modo manuale con `ALTER TABLE ADD PARTITION`

In questo modo, possiamo andare ad effettuare query per ogni partizione creata, ma abbiamo anche un lato negativo: la creazione e la gestione delle partizioni in maniera manuale aggiunge overhead e complessità.

2. Usare la partition projection

Dato che la struttura dei log di AWS CloudTrail è nota a priori, Athena può fare un'inferenza e capire la partizione di destinazione in automatico. Questo ci solleva dall'incarico di dover aggiungere manualmente le nuove partizioni.

Conlusione

Abbiamo visto alcune peculiarità e dettagli poco noti del servizio AWS CloudTrail e dei servizi annessi Amazon S3 e Amazon Athena. Abbiamo imparato che, in alcuni casi, le impostazioni delle lifecycle policy di S3 ci possono intralciare nell'utilizzo quotidiano e

che dobbiamo bilanciare i costi e la data availability. Abbiamo visto che, grazie all'utilizzo delle partizioni, è possibile ottimizzare le query su Athena per avere più performance con poco effort di gestione.

Eravate a conoscenza di questi aspetti?

Vi siete scontrati con altri limiti o - speriamo - con ulteriori best practices?

Fatecelo sapere!

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!

Mehmed Dourmouch

DevOps Engineer. Molto Dev, non così Ops. Mi piace rompere le cose e vedere cosa succede, anche automatizzare tutto. Partecipo spesso a CTF di cybersecurity e nel tempo libero produco cacofonia con la mia chitarra.
