

Estrarre dati da SAP in modo semplice con Amazon AppFlow

4 Dicembre 2024 - 6 min. read

[Data and Analytics](#)

[Data Lake](#)

In un mondo in cui i dati guidano l'innovazione aziendale, la capacità di raccogliarli, organizzarli e analizzarli in modo efficiente è cruciale. Un data lake è un repository centralizzato che acquisisce e memorizza grandi quantità di dati nella loro forma originale. Grazie alla sua architettura scalabile, un data lake può ospitare tutti i tipi di dati da qualsiasi fonte: strutturati (tabelle di database relazionali, file CSV), semi-strutturati (file XML, pagine web) e non strutturati (immagini, file audio, tweet) - il tutto senza compromettere la fedeltà.

AWS offre un ecosistema completo di servizi che ti consentono di creare e gestire i data lake, rendendoli una base ideale per una vasta gamma di esigenze analitiche. Inoltre, attraverso le integrazioni con servizi di terze parti, è possibile migliorare ulteriormente un data lake, ottenendo una visione complessiva e approfondita dei dati aziendali.

In questo articolo, esploreremo un'integrazione specifica tra Amazon AppFlow e SAP per preparare i dati per una piattaforma di analisi dei dati.

A cosa serve Amazon AppFlow?

Immagina di poter trasferire i dati tra applicazioni SaaS e il tuo ambiente in AWS rapidamente, in modo sicuro e senza complessità.

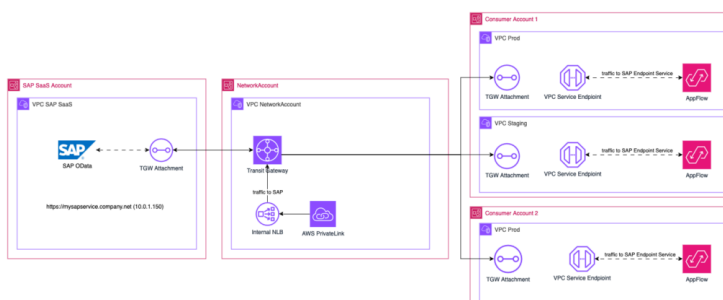
Questa è la magia che offre **Amazon AppFlow**, un servizio che semplifica le integrazioni e ti permette di automatizzare i flussi di dati con pochi click, occupandosi di tutti gli aspetti importanti delle integrazioni dei flussi di dati. Ti solleva dal compito di gestire infrastrutture complesse e di scrivere codice per integrare diverse fonti di dati. Un aspetto cruciale dell'integrazione dei flussi di dati è la privacy del canale di connessione utilizzato.

Fortunatamente, AWS aiuta a risolvere questa problematica con il suo servizio PrivateLink. Amazon AppFlow si integra nativamente con **AWS PrivateLink**, permettendoti di instaurare connessioni a fonti di dati in maniera privata, senza esporre i dati a internet pubblico. Questo garantisce una miglior sicurezza dei dati, soprattutto quando si tratta di informazioni sensibili.

Prerequisiti e AWS Organization

Analizziamo i principali attori nel nostro scenario, come interagiscono tra di loro e come sono organizzati i nostri account AWS:

- **Un account di management**, responsabile della rete centralizzata, che funge da hub centrale per tutti gli altri account all'interno dell'organizzazione. Abbiamo configurato un Transit Gateway in questo account e lo abbiamo condiviso con l'organizzazione. Questo ci consente di collegare il nostro account principale all'account gestito da SAP.
- **Molteplici account figli** all'interno della nostra organizzazione che saranno i consumatori dei dati presenti in SAP Cloud.
- **Un account AWS esterno di proprietà di SAP**. Qui è dove si trovano i nostri dati.



Notate che l'account di proprietà di SAP, così come i nostri account, sono tutti collegati tramite Transit Gateway VPC Attachments. Questa configurazione di rete scalabile garantisce che tutti i nostri account AWS attuali e futuri possano comunicare tra di loro. Un altro aspetto da considerare è la corretta configurazione del routing: seguire la best practice del "least privilege" garantisce che gli account abbiano accesso solo ai dati e ai servizi che sono esplicitamente autorizzati a utilizzare.

Setup di un Endpoint Service

Una volta che i servizi SAP sono stati collegati al nostro account principale, dobbiamo trovare un modo semplice per rendere i dati disponibili ai nostri consumatori. A tal proposito possiamo utilizzare un Endpoint Service supportato da AWS PrivateLink.

PrivateLink ci consente di condividere un servizio ospitato in AWS con altri account di consumatori. Richiede un Network Load Balancer (supporta anche i Gateway Load

Balancer), che riceve le richieste dai consumatori e le inoltra al tuo servizio.

Per configurare una connessione privata su Amazon AppFlow, dobbiamo prima creare e verificare un Endpoint Service che sarà il punto di ingresso ai servizi SAP.

Per creare l'Endpoint Service, è necessario considerare vari aspetti:

- **Un nome DNS.** È necessario definire il nome DNS privato che verrà utilizzato per l'Endpoint Service. Tieni presente che il nome DNS deve essere situato in una Hosted Zone pubblica per poter verificare l'ownership del dominio.
- **Certificato ACM per il dominio.** Considera di richiedere un certificato wildcard per il dominio per prevenire potenziali problemi relativi alle discrepanze dei sottodomini.
- **Network Load Balancer, Target Group e dettagli dell'endpoint SAP.** È necessario creare un Network Load Balancer interno da associare all'Endpoint Service. Hai anche bisogno di alcune informazioni di base sui target per registrarli nel Target Group (ad es. porta, protocollo, indirizzi IP). Nel nostro caso specifico si tratta di un servizio con protocollo TLS sulla porta 44300.
- **Verifica del nome DNS privato.** Ti verrà chiesto di confermare la proprietà del dominio selezionato tramite un record TXT.

Anche se questi passaggi sono più incentrati sui servizi SAP, è possibile applicare gli stessi principi a qualsiasi altro servizio esterno che si desidera integrare con le applicazioni consumer in AWS.

Aspetta alcuni minuti affinché il provisioning e la verifica del dominio abbiano effetto e dovresti essere pronto per andare avanti!

Setup AppFlow Connection

Ora che abbiamo un Endpoint Service, possiamo passare agli account figli di AWS e creare un nuovo Flow usando SAP OData come sorgente dei dati.

Dalla console di Amazon AppFlow, seleziona **SAP OData** come tipo di connettore e inserisci i parametri richiesti per la connessione:

- **Connector type:** SAP OData
- **Application Host URL.** Questo è il nome DNS personalizzato che hai definito durante la configurazione dell'Endpoint (es. <https://mysapservice.example.com>)
- **Path, Port Number e Client Number dell'applicazione.** Possono variare a seconda dei tuoi servizi (es. percorso: `/sap/opu/odata/iwfnd/catalogservice;v=2`, numero di porta: 44300,

numero del client: 100)

- **PrivateLink.** Imposta su “Enabled” e inserisci il Service Name di AWS PrivateLink. Puoi trovarlo nella dashboard del servizio Endpoint VPC. (es. com.amazonaws.vpce.eu-west-1.vpce-svc-12341234333344455)
- **Modalità di autenticazione:** Basic Auth. Nota che l'autenticazione OAuth2 non è fattibile per il nostro caso d'uso poiché richiede l'interazione dell'utente tramite browser e non può essere effettuata tramite PrivateLink.
- **Nome utente e password.** AppFlow creerà automaticamente un segreto su Secrets Manager per le credenziali del tuo servizio, in modo che tu possa successivamente modificarle se fosse necessario.

Una volta che tutti i campi sono compilati, salva e testa la connessione a SAP.

Prossimi Passi

Con l'aiuto di Amazon AppFlow e di un Endpoint Service VPC, abbiamo ottenuto una connessione privata al nostro datastore in SAP e possiamo ora creare un nuovo flusso di dati.

Ad esempio, puoi configurare un Flow usando un bucket S3 come destinazione, e impostarlo per l'esecuzione on-demand oppure in modo ricorrente.

Puoi scegliere tra json o csv come formati di output, o meglio ancora, puoi scegliere il formato **parquet**. Utilizzare il formato parquet consente di avere una memorizzazione dei dati altamente compressa e ottimizzata, il che a sua volta permette di ottimizzare i costi di storage e di interrogazione dei dati con AWS Athena.

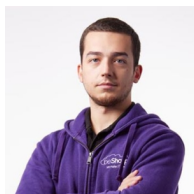
Conclusione

In questo articolo abbiamo trattato solo il connettore SAP OData, ma è solo una delle tante integrazioni che puoi configurare con AppFlow. Abbiamo appena scalfito la superficie di ciò che può essere fatto e delle vaste possibilità che offre un data lake. Da questo punto di partenza puoi espandere il tuo data lake e il sistema di ingestione per accogliere molte altre integrazioni con fonti di dati eterogenee esterne. Questo ti permette di iniziare ad analizzare, trasformare e ottenere preziosi insight dai tuoi dati da un punto di vista centralizzato.

Sentiti libero di esplorare tutte le opzioni di personalizzazione che Amazon AppFlow offre per trovare la configurazione che meglio si adatta al tuo caso d'uso!

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.
