

PREZZO E PRESTAZIONI DI SNOWPARK: RISULTATI DEI CLIENTI



I clienti registrano valori medi di **prestazioni 4,6 volte più veloci e risparmi sui costi del 35%** con Snowpark rispetto a un ambiente Spark gestito*

PANORAMICA DI SNOWPARK

Snowpark è il set di librerie e ambienti di esecuzione del codice che consente di eseguire Python e altri linguaggi di programmazione accanto ai tuoi dati in Snowflake. Puoi utilizzare Snowpark per creare pipeline di dati, modelli ML, app e altre operazioni di elaborazione dei dati.

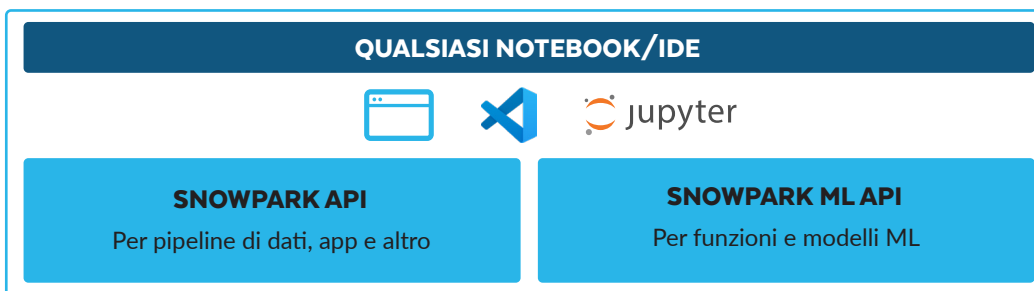
Lato client, Snowpark è costituito da librerie, tra cui Snowpark DataFrame API e API Snowpark ML native con componenti per lo sviluppo e l'esecuzione di modelli.

Lato server, tutto il codice Python e in altri linguaggi di programmazione viene automaticamente trasferito per l'elaborazione agli ambienti di esecuzione del codice del motore Snowflake, tra cui runtime Python, Java e Scala in virtual warehouse o qualsiasi altro runtime eseguito tramite Snowpark Container Services (in public preview). Utilizzando la capacità di calcolo dei warehouse, gli sviluppatori possono servirsi di UDF e stored procedure (sproc) per integrare ed eseguire la logica personalizzata. Snowpark Container Services consente l'utilizzo di runtime personalizzati e l'hosting di container per codice personalizzato all'interno di Snowflake su una varietà di hardware, comprese le GPU.

Migliaia di organizzazioni accelerano lo sviluppo e la performance dei workload di data engineering e AI/ML con Snowpark per Python. Dalla disponibilità generale nel novembre 2022, Snowpark ha continuato a crescere rapidamente, e a gennaio 2024 era utilizzato settimanalmente da più del 40% dei clienti Snowflake. A gennaio 2024, i clienti Snowflake hanno eseguito ogni giorno una media di circa 38 milioni di query in Snowpark (in base ai dati di gennaio 2024).

Il 40% dei clienti Snowflake utilizza Snowpark per eseguire circa 38 milioni di query al giorno

SVILUPPO E DISTRIBUZIONE DI CODICE DA LIBRERIE LATO CLIENT



AMBIENTI DI ESECUZIONE DEL CODICE NEL MOTORE ELASTICO SNOWFLAKE



*METODOLOGIA

DIMENSIONE DEL CAMPIONE: Oltre 50 POC e workload di produzione reali
SETTORI: servizi finanziari, retail e beni di consumo, healthcare e life sciences, settore manifatturiero e industriale, media e entertainment, tecnologia, telecomunicazioni, servizi al consumatore e hospitality, settore pubblico
SEDI: in tutto il mondo
WORKLOAD UTILIZZATI: data engineering, AI/ML

Questo report riepiloga i dati di una serie di casi d'uso di produzione e progetti pilota dei clienti mettendo a confronto la velocità e il costo di Snowpark rispetto ai servizi Spark gestiti nel periodo da novembre 2022 a gennaio 2024. **Tutti i risultati del report sintetizzano i risultati effettivi ottenuti dai clienti con i loro dati reali e non rappresentano data set creati ad hoc come benchmark.**

WORKLOAD VALUTATI:

In base ai risultati dei progetti pilota e dei workload di produzione dei clienti, abbiamo osservato che Snowpark è utilizzato principalmente per i workload di data engineering e AI/ML con Python. Benché Java, Scala e Python siano tutti generalmente disponibili, più del 75% del campione analizzato utilizza Snowpark con Python.

Il 75% del campione analizzato utilizza Snowpark con Python

I data engineer utilizzano Snowpark per trasformazioni dei dati, logica aziendale personalizzata, pipeline che alimentano modelli ML, dashboard di reporting e applicazioni. Queste operazioni includono sia processi ETL, in cui i dati vengono trasformati a partire da un data lake e scritti in Snowflake, sia processi ELT, in cui Snowflake è l'origine e la destinazione dei dati.

I data scientist e gli ingegneri ML utilizzano Snowpark per eseguire feature engineering, addestramento e inferenza di modelli con data set di grandi dimensioni.

RISULTATI:

Dopo aver migrato da Spark a Snowpark, gli utenti registrano notevoli miglioramenti della velocità e dei costi.

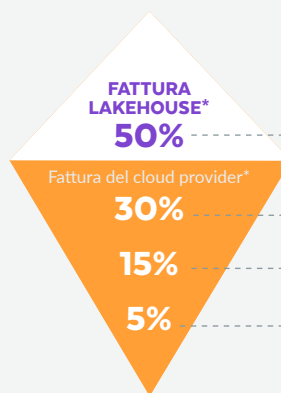
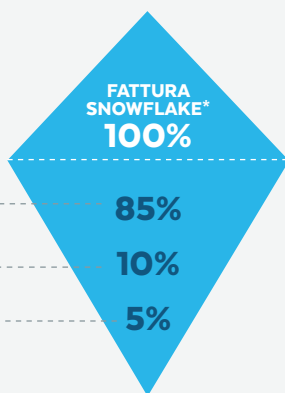
I risparmi sui costi non riflettono tutti i vantaggi offerti da Snowpark in termini di TCO, in quanto omettono i costi operativi indiretti che sono prevalenti per le soluzioni Spark. Ottimizzazioni delle query, sicurezza e governance spesso non sono preconfigurate e pronte all'uso, ma richiedono competenze specializzate per il tuning manuale. Questo può ridurre la produttività e comportare lunghi tempi di configurazione e onboarding, malfunzionamento delle query e aumento del rischio di violazioni a causa di una protezione dei dati inadeguata.

Inoltre, con le soluzioni Spark gestite può essere più difficile monitorare il TCO complessivo, poiché l'infrastruttura cloud e la capacità di calcolo basata su Spark sono fatturate separatamente.

Con Snowpark, i clienti hanno osservato performance medie 4,6 volte più veloci e costi inferiori del 35% rispetto alle soluzioni Spark gestite.

VISIBILITÀ COMPLETA DEI COSTI: ALLA RICERCA DEI COSTI NASCOSTI

Capacità di calcolo Snowflake 85%
Funzionalità opzionali: Snowpipe, viste materializzate, clustering automatico, accelerazione query, task serverless, ecc. 10%
Storage 5%



Costo della capacità di calcolo Spark

Capacità di calcolo cloud (EC2, Azure Virtual Machines, Google Compute)

Cloud storage e costo di recupero (S3, Blob, richieste API)

Trasferimento dei dati nel cloud (rete, connettore Spark, capacità di calcolo, ecc.)

*Le percentuali effettive possono variare a seconda del workload e dell'ambiente del cliente.

SNOWPARK È PIÙ VELOCE DI SPARK: ECCO PERCHÉ

Spark e Snowpark sono entrambi utilizzati per l'elaborazione dei dati su vasta scala, ma i clienti Snowpark ottengono notevoli vantaggi in termini di prezzo e prestazioni rispetto alle soluzioni Spark gestite e open source. Perché? Essenzialmente grazie al motore Snowflake e all'eliminazione dello spostamento dei dati.

Con il motore distribuito di Snowflake, le capacità di archiviazione e di calcolo sono integrate a livello logico, ma separate a livello fisico. Snowflake ha un'architettura dati condivisa multi-cluster che pianifica e ottimizza l'esecuzione dei workload simultanei. Gli sviluppatori SQL sono stati i primi a beneficiare delle numerose ottimizzazioni integrate in questo motore, come il clustering automatico e il micro-partizionamento. Snowpark estende il motore Snowflake oltre SQL, per includere gli sviluppatori che utilizzano Python, Java e Scala.

Un'altra importante fonte di ottimizzazione dei costi e delle prestazioni è l'eliminazione dei motori di elaborazione dati esterni per le pipeline in più linguaggi, che spesso richiedono operazioni di conversione, uscita, acquisizione e caricamento dei dati dalla sorgente. Con Snowpark, non è necessario trasferire i dati tra i vari cluster per i diversi linguaggi e chi sviluppa può elaborare petabyte di dati utilizzando il linguaggio che preferisce su un'unica piattaforma.



L'ESPERIENZA DEI CLIENTI

Snowpark aiuta i clienti in tutti i settori a migliorare le prestazioni e ottimizzare i costi.



TRASFORMAZIONI DEI DATI: OPENSTORE

Migrazione da EMR a Snowpark per le pipeline di dati che alimentano le funzioni interne.

RISULTATI

100%

Raggiunta la parità totale con i precedenti workload PySpark.

80%

Riduzione dell'80% delle ore di manutenzione tecnica.

87%

Openstore ha ridotto dell'87% il runtime delle pipeline E2E.

20X

Scalabilità più facile e affidabilità a lungo termine dopo il passaggio da PySpark a Snowpark, lavorando con un volume di dati oltre 20 volte superiore.

MODELLAZIONE ML: EDF ENERGY

Migrazione da Spark a Snowpark per lo sviluppo e la distribuzione di ML.

RISULTATI



Distribuzione più veloce dei modelli. Riduzione dei tempi di distribuzione da mesi a giorni.



Lancio di prodotti rivolti ai clienti in poche settimane anziché mesi.

3-4X

Quantità dei prodotti aumentata di 3-4 volte.

“Adesso il nostro team lavora molto più velocemente. Prima, potevamo supportare forse sei nuovi data product nel corso di un anno, mentre ora penso che il numero sia triplicato o quadruplicato.”

– DATA SCIENCE LEAD, EDF

LOGICA AZIENDALE PERSONALIZZATA: INTERCONTINENTAL EXCHANGE

Migrazione da EMR a Snowpark per le pipeline di dati per le segnalazioni di vigilanza

RISULTATI

80%

Query ad hoc fino all'80% più veloci.



Creazione più veloce dei report sulla conformità e la sicurezza dei dati.

360°

Visione a 360 gradi di tutte le metriche aziendali e operative.

“Snowpark può davvero contribuire a ottimizzare i costi e parallelamente migliorare le prestazioni.”

– DATA, ANALYTICS & GOVERNANCE HEAD, ICE

Con le librerie e i runtime Snowflake tutti i data scientist e i data engineer che utilizzano Python possono creare con facilità in Snowflake.

Vuoi sapere di più? Visita www.snowflake.com/it/data-cloud/snowpark

INFORMAZIONI SU SNOWFLAKE

Snowflake permette a ogni organizzazione di mobilitare i propri dati grazie al Data Cloud. I clienti utilizzano il Data Cloud Snowflake per unificare i dati contenuti nei silos, esplorare e condividere i dati in totale sicurezza, potenziare le applicazioni basate sui dati, ed eseguire diversi workload di analisi, intelligenza artificiale e machine learning. Ovunque siano i dati o gli utenti, Snowflake offre un'esperienza sui dati unica che si estende a più cloud e aree geografiche. Migliaia di clienti di ogni settore, tra cui 691 della classifica 2023 Forbes Global 2000 (G2K) al 31 gennaio 2024, utilizzano il Data Cloud Snowflake per far crescere le loro aziende.

Scopri di più su snowflake.com/it