

MIGRARE AL CLOUD

Come e perché modernizzare la tua piattaforma dati

SOMMARIO

- 3** La voce del mercato
- 4** 1a parte. Il contesto della migrazione: approcci, strategie e requisiti
- 8** 2a parte. Pianificare la migrazione
- 13** 3a parte. Sfruttare i vantaggi della nuova piattaforma dati nel cloud
- 15** Conclusione
- 15** Ulteriori informazioni
- 16** Informazioni su Snowflake

LA VOCE DEL MERCATO

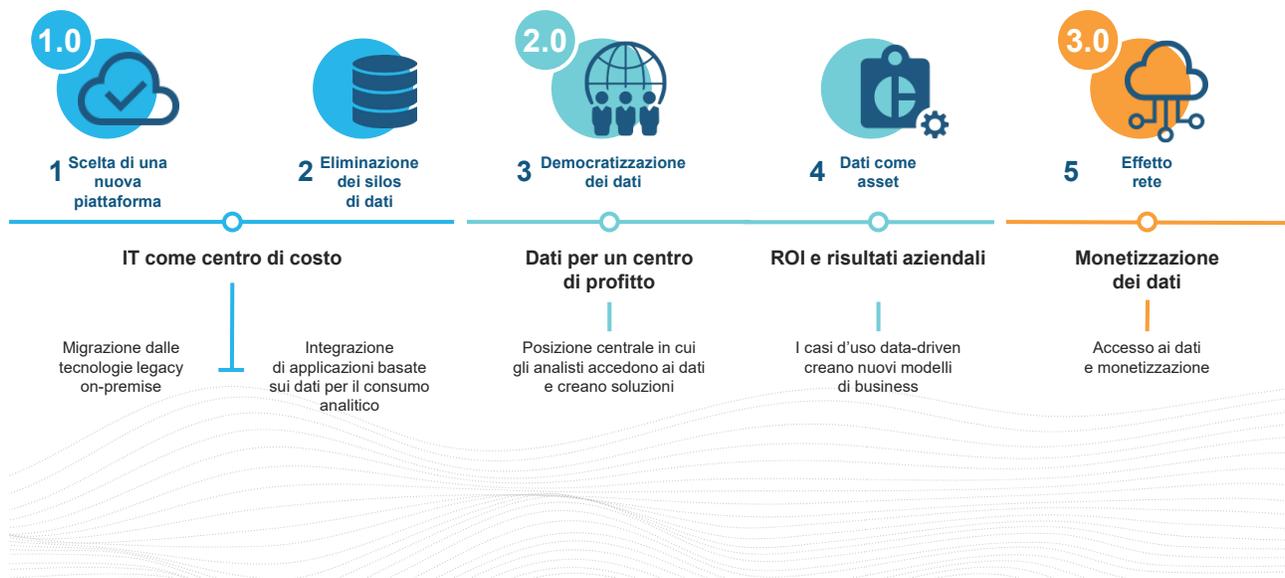
Con la crescita esponenziale dei volumi e della varietà di dati, i sistemi di gestione dei dati legacy non sono più adatti a caricare e integrare i dati in modo semplice e sicuro, consentire un'analisi rapida e democratizzata e rendere gli insight così ottenuti disponibili a tutti gli utenti aziendali che ne hanno bisogno. Inoltre, oggi il semplice accesso ai dati non basta: per cogliere le nuove opportunità di distribuire e monetizzare i dati è necessario modernizzare. Non sorprende che le iniziative di modernizzazione richiedano il trasferimento dei workload sul cloud. In un recente sondaggio condotto da TDWI, solo una piccola percentuale degli intervistati non

utilizzava il cloud e solo il 22% ha dichiarato che i dati risiedevano on-premise. Il 38% degli intervistati utilizzava dati sia nel cloud che on-premise.¹

Molte organizzazioni stanno cogliendo l'opportunità di trasformare e aggiornare il loro approccio ai dati.

Con un'adeguata pianificazione iniziale, la migrazione dei data warehouse e data lake legacy al cloud può generare grandi profitti per la tua organizzazione. Questo ebook propone una roadmap per la migrazione al cloud e l'inizio del percorso di modernizzazione.

Il nuovo percorso di modernizzazione



¹ "TDWI Best Practices Report: Modernizing the Organization to Support Data and Analytics", tdwi.org/research/2022/06/ppm-all-best-practices-report-modernizing-the-organization-support-data-analytics/

PRIMA PARTE

IL CONTESTO DELLA MIGRAZIONE: APPROCCI, STRATEGIE E REQUISITI

Sono molti i motivi per cui le organizzazioni oggi scelgono di adottare il cloud computing. Tuttavia, molte organizzazioni hanno bisogno di un piano, di una possibile visione di quello che sarà il loro futuro nel cloud. Poiché le condizioni di partenza per quanto riguarda capacità analitiche o preparazione per il cloud possono variare, dovrai valutare attentamente la quantità di codice legacy che andrai a trasferire dall'ambiente on-premise all'infrastruttura cloud e la velocità del trasferimento.

I CINQUE SCENARI DI MIGRAZIONE PIÙ COMUNI

Il tipo di migrazione scelto influenzerà in modo significativo la tua strategia di migrazione. Di seguito sono riportati cinque possibili percorsi che molte organizzazioni intraprendono per migrare le proprie soluzioni di dati legacy su una piattaforma moderna:

1. OLTP per il reporting operativo e l'analisi

Questo scenario è estremamente comune. Molte organizzazioni utilizzano sistemi di elaborazione delle transazioni online (Online Transaction Processing, OLTP), come SQL Server, Oracle o MySQL, per il reporting e l'analisi di base. Anche se questa può essere una soluzione a breve termine, le esigenze di reporting dell'azienda sono in concorrenza con le esigenze operative, sovraccaricando una risorsa fissa e rallentando le prestazioni in entrambi i campi. Una piattaforma veramente elastica elimina questo problema. È abbastanza semplice. Come vedremo più dettagliatamente in seguito, è possibile prendere lo schema transazionale esistente, che di solito è in terza forma normale (3NF), e trasferirlo senza modifiche nel cloud. In questo modo il workload di reporting viene rimosso dal sistema esistente e i dati vengono archiviati in una piattaforma progettata per l'analisi, eliminando i rallentamenti delle prestazioni e, in alcuni casi, infondendo nuova vita al data store operativo.

2. Data warehouse basato su appliance

Molti fornitori di appliance on-premise stanno interrompendo il supporto di questi sistemi legacy. Soprattutto, i loro clienti vogliono sfuggire alle limitazioni in termini di prestazioni, costi ecc. di questi sistemi, che non soddisfano più le esigenze di analisi dei dati delle aziende moderne. Anche i data warehouse basati su appliance richiedono enormi costi iniziali, in genere spese in conto capitale (CapEx).

Al contrario, una piattaforma creata per il cloud è progettata per crescere insieme a voi senza spese iniziali, grazie a un modello pay-as-you-go, per cui le spese saranno operative (OpEx). Eviterai così di pianificare in base al giorno di massimo consumo, per poi pagare più del necessario per un sistema sottoutilizzato per gli altri 364 giorni dell'anno. Allo stesso modo, se dovrai improvvisamente aumentare la capacità di analisi nel corso dell'anno ti servirà un sistema che si adatta dinamicamente in base alle tue nuove esigenze. Una piattaforma scalabile istantaneamente è la soluzione.

Un altro vantaggio è l'accesso ai dati quasi in tempo reale. Poiché le appliance on-premise sono una risorsa fissa, i team di gestione dei dati creano finestre di caricamento notturno per rendere disponibili i dati la mattina successiva. Le attuali tecnologie progettate per il cloud consentono di segmentare i workload e caricare i dati 24/7, senza impatto sull'elaborazione delle query e accelerando il time-to-value dei tuoi dati.

3. Data mart

La maggior parte delle organizzazioni risente della mancanza di un'unica fonte di riferimento e ha dati suddivisi in silos in numerosi repository. Alcune hanno tentato di realizzare un sistema di accesso federato a tutti questi archivi, ma si sono rese conto rapidamente che i costi per creare e mantenere tale accesso non erano sostenibili. Per eliminare queste barriere alla generazione di insight da tutti i loro dati queste organizzazioni hanno bisogno di un repository centralizzato. Una piattaforma nata per il cloud rappresenta una scelta ovvia per il consolidamento dei dati, poiché supporta le proprietà ACID, e quindi le transazioni dei database relazionali, può essere partizionata/segmentata logicamente senza replica e può scalare verticalmente le risorse di elaborazione on demand. Indipendentemente dall'approccio Kimball o Inmon, la piattaforma deve separare la capacità di calcolo dall'archiviazione e offrire agli utenti finali la massima flessibilità per accedere ai data set con strumenti aziendali che sfruttano ANSI SQL e altri linguaggi per eseguire workload analitici avanzati.

4. Data lake

Le iniziative di data lake sono diventate una sorta di buco nero: inserire i dati è facile, estrarli è molto complesso. Molte aziende si sono rese conto che le infrastrutture Apache Hadoop on-premise sono costose e complesse e non soddisfano i loro requisiti di analisi e simultaneità. Fortunatamente, esiste una soluzione. Un approccio a più livelli o "a zone" è un ottimo modo per identificare i data set che è possibile migrare agevolmente al cloud. Questo metodo consente di visualizzare facilmente lo spostamento dei dati in modo controllato e sicuro da un ambiente on-premise a un'infrastruttura di archiviazione cloud come Azure Blob Storage, AWS S3 o Google Cloud Storage.

5. Consolidamento tecnologico

OLTP ed elaborazione analitica online (OLAP) sono sistemi separati a causa dei requisiti diversi per i dati e il loro utilizzo. Quando i dati sono contenuti in un ampio set di database, la presenza di silos rende difficile ottenere analisi, creare applicazioni e agire in tempo reale sfruttando sia i dati transazionali che quelli analitici. Nel tentativo di riunire i dati, le organizzazioni utilizzano processi ETL (Extract, Transform, Load) ed ELT (Extract, Load, Transform) per trasferire i dati dai database OLTP alla propria piattaforma di analisi. Tuttavia, questo approccio vecchia maniera non è sostenibile e crea nuovi problemi: i ritardi nel rendere disponibili i dati ai team, le complessità nella creazione e nell'aggiornamento delle pipeline e la gestione di più sistemi diversi, che rendono difficile scalare e gestire la governance su più sistemi. Le moderne piattaforme dati stanno colmando questa lacuna e creando l'opportunità di riunire i dati transazionali e i dati pronti per la query nello stesso luogo, eliminando così la necessità di mantenere sistemi duplicati e riducendo la complessità e la proliferazione delle tecnologie.

REQUISITI DELLA STRATEGIA

Le migrazioni non sono molto diverse dalla maggior parte dei progetti IT, il che significa che in genere per prima cosa si stabiliscono i requisiti. I requisiti spesso superano numerosi confini interni aziendali, poiché una strategia di migrazione al cloud può essere una decisione a livello di direzione aziendale. Senza l'approvazione dei dirigenti, il progetto avrà un'ampiezza limitata e sarà considerato come una sorta di "shadow IT". Il che potrebbe andar bene per iniziare a portare nel cloud alcune linee di business, ma con il tempo tutta l'organizzazione dovrà essere allineata, dal team che si occupa dell'architettura a quello di sicurezza, dal responsabile dei dati allo stesso CFO.

Strategia organizzativa

Molte organizzazioni hanno il cloud nel proprio DNA e persino nei propri mission statement. Altre, che esistono da più tempo, forse comprendono che la modernizzazione è necessaria, ma non vogliono cambiare troppo rapidamente. I vantaggi aziendali del cloud sono difficili da ignorare, a cominciare dalla maggiore agilità, dai costi inferiori e dalle analisi più approfondite. Tuttavia, il progetto di migrazione al cloud deve procedere a un ritmo coerente con i tuoi obiettivi aziendali. Stabilisci se il progetto dovrà essere una "testa di ponte", un modo per far avvicinare l'azienda al cloud, o se farà parte di un'iniziativa cloud più ampia, con best practice e risorse tecniche condivise. Uno dei fattori critici per il successo sarà stabilire quali data set sono già pronti per lasciare il data center e quali potranno passare al cloud solo in un secondo tempo.



Strategia tecnica

Ogni strategia inizia con la stessa domanda: “cosa stiamo cercando di fare?” Per chi lavora con la tecnologia, la risposta può essere data con i requisiti. Una vecchia massima recita: “Un lavoro può essere veloce, economico o ben fatto. Scegli due cose.” Oggi per chi utilizza lo sviluppo Agile e sfrutta i vantaggi del cloud, questa massima non è più tanto accurata.

Requisiti aziendali/funzionali

Progetta senza perdere di vista il traguardo. Parla degli obiettivi con gli utenti delle analisi esistenti per comprendere le sfide e i requisiti attuali. Il cloud supporta nuove funzionalità, come l'accesso ai dati quasi in tempo reale, la democratizzazione dei dati e l'analisi di livello superiore con accesso a dati dettagliati e non solo aggregati. Crea un piano che evidenzi la capacità di fornire gli strumenti e i controlli di sicurezza appropriati a tutte le linee di business. Utilizza questi obiettivi per allineare gli utenti IT e business e definire la vision di fornire informazioni accurate in modo sicuro alle persone giuste e al momento giusto.

Requisiti non funzionali

Fai il punto delle politiche relative ai contratti di servizio (SLA) tra l'IT e l'azienda, dei requisiti di sicurezza per la protezione dei dati, dei requisiti di usabilità degli utenti finali e molti altri aspetti essenziali. I requisiti non funzionali più comuni sono:

- **Sicurezza**
- **Governance**
- **Affidabilità**
- **Prestazioni**
- **Manutenibilità**
- **Scalabilità**
- **Usabilità**

Nel caso del cloud, non esitare a includere alcuni requisiti ambiziosi. Una volta stilato un elenco soddisfacente, classifica ogni elemento come “essenziale” o “opzionale”. Non farti paralizzare dalla dovizia di opzioni e scegli una finestra di implementazione e una metodologia di sviluppo (Agile, Waterfall e così via) che siano praticabili per la tua organizzazione. Presta particolare attenzione ai requisiti di alta disponibilità e disaster recovery. Una piattaforma moderna può evitare di dover progettare soluzioni eleganti ma costose per soddisfare le esigenze dell'azienda.

TRE APPROCCI ALL'IMPLEMENTAZIONE DELLA MIGRAZIONE

Ora è il momento di stabilire il tipo di migrazione più adatto. Le diverse opzioni includono: trasferimento senza modifiche, trasferimento con migliorie e riprogettazione completa. La scelta della strategia di migrazione sarà guidata dalle attività svolte prima di questa fase, come l'allineamento della migrazione con l'organizzazione e la definizione delle strategie tecniche e aziendali.

Trasferimento senza modifiche

Molti considerano questa opzione, detta anche “lift-and-shift”, il modo più sicuro e immediato per effettuare una migrazione. Il piano è semplice: tutte le operazioni svolte con il sistema esistente dovrebbero rimanere identiche nel nuovo sistema, con modifiche minime. Una strategia di trasferimento senza modifiche è la scelta giusta se:

- **I requisiti sono molto specifici (pochissimi nuovi requisiti)**
- **Il tempismo dell'implementazione è cruciale (si deve abbandonare il vecchio sistema al più presto)**
- **Il nuovo sistema ha tutte le caratteristiche e le funzioni del vecchio sistema**
- **L'ecosistema di strumenti correlati (ETL, BI, gestione del sistema) richiede modifiche minime o nulle**
- **La migrazione non è il fulcro della strategia di migrazione al cloud (vedi di seguito)**

L'ultimo punto è discutibile, poiché molte migrazioni cloud richiedono cambiamenti sia tecnici che alla cultura dell'organizzazione. Se la prima iniziativa di migrazione non fornisce caratteristiche e funzioni

in più rispetto al vecchio sistema, l'organizzazione può considerarla un successo? A volte, dimostrare che è possibile migrare senza rischi per l'azienda, migliorando in qualche modo le prestazioni, può essere sufficiente per gli stakeholder.

Trasferimento con miglorie

Questo è di gran lunga l'approccio più comune. Il concetto è semplice: Mentre si convertono gli asset, si cercano opportunità per semplificare o migliorare la pipeline dei dati, il modo in cui i dati sono organizzati o trasformati, e le modalità di accesso ai dati. Quindi, si trovano i modi per sfruttare le nuove funzionalità/capacità del sistema di destinazione. Il punto non è modificare le funzionalità principali del sistema, ma semplicemente cogliere l'opportunità di semplificare o snellire.

Il vantaggio è la possibilità di dimostrare miglioramenti rispetto al processo esistente senza provocare danni o introdurre troppi rischi. Il team dirigenziale potrà quindi utilizzare la migrazione come prova concreta per l'azienda. Anche se si tratta di un importante cambiamento nella filosofia informatica aziendale, non influisce negativamente sulla performance e offre ulteriori vantaggi aziendali, tra cui:

- **Accesso più rapido a più dati**
- **Analisi dei dati più granulare**
- **Migliori risultati da query e report**
- **Nessuna contesa di risorse di calcolo quasi illimitate**

Riprogettazione, riorganizzazione e consolidamento

Molte organizzazioni non dispongono di un data warehouse o di un data lake aziendale e, in alcuni casi, i tentativi di crearne uno sono stati deludenti. I dati sono contenuti in più sistemi locali: alcuni sono utilizzati per OLTP, altri per OLAP, altri ancora

risiedono semplicemente nei file system in attesa di essere analizzati. Un cambiamento di piattaforma è considerato il momento ideale per riprogettare, o progettare per la prima volta, una piattaforma dati completamente funzionale in grado di scalare con l'azienda.

La piattaforma deve soddisfare i requisiti descritti nelle sezioni precedenti: deve gestire più tipi di dati e deve consentire agli utenti finali di utilizzare i propri strumenti e linguaggi preferiti, come SQL, per creare una strategia di democratizzazione dei dati per tutte le LOB. In alcuni casi, la piattaforma dati crea persino l'opportunità di avviare un nuovo e moderno data sharing all'interno e all'esterno dell'azienda. Progetti come questo sono un ottimo modo per consolidare l'infrastruttura e gestire meglio contratti, sicurezza e shadow IT, producendo risultati incrementali per l'azienda.

La maggior parte delle iniziative di consolidamento inizia combinando più data set sulla nuova piattaforma cloud per dimostrarne il valore analitico. La fase successiva consisterà nel limitare gradualmente l'accesso ai sistemi legacy, aumentando al contempo le funzionalità della piattaforma dati e ottenendo risultati rapidamente. Molte di queste iniziative si concentrano anche sulla creazione di nuovi flussi di ricavi, sulla riduzione o l'eliminazione delle pipeline di dati e sul consolidamento di repository di dati diversi.



SECONDA PARTE

PIANIFICARE LA MIGRAZIONE

Eseguire queste fasi, e in quest'ordine, non è obbligatorio. A seconda della portata della migrazione, potrebbero essere necessari più o meno passaggi. La chiave è progettare un framework e gli elementi di base del piano da cui partire. Valuta le competenze interne, non esitare a sfruttare le best practice delineate dai tuoi fornitori strategici e considera la possibilità di collaborare con esperti di migrazione.

FASE 1: DEFINIRE LA PORTATA

La verità è che ogni migrazione è diversa dalle altre ed è raro che il punto di arrivo sia perfettamente chiaro a priori. L'obiettivo è creare un piano allineato agli obiettivi dell'azienda, che fornisca le capacità desiderate nel più breve tempo possibile e ponga le basi per un miglioramento incrementale. Il punto di arrivo potrebbe essere il trasferimento di un singolo workload nel cloud entro un mese, oppure la migrazione dell'intera piattaforma di analisi entro la fine dell'anno. È ragionevole prevedere il ritorno sull'investimento nel giro di un anno, o addirittura in tempi più brevi in determinati scenari.

FASE 2: DOCUMENTARE LO STATO DI FATTO

Questa non è la parte più affascinante di una migrazione, ma probabilmente è una delle più importanti. Dovrai comunicare sia internamente che esternamente e a ogni livello della struttura decisionale aziendale qual è lo stato dell'implementazione esistente. Gli asset da migrare includono:

- Tutte le fonti che popolano i sistemi esistenti
- Tutti gli oggetti di database (tabelle, viste, utenti e così via)
- Tutte le trasformazioni, con programmazioni per l'esecuzione o criteri di attivazione
- Un diagramma dell'interazione tra sistemi/strumenti

FASE 3: DECIDERE L'APPROCCIO E FORMARE IL TEAM DI IMPLEMENTAZIONE

Qui sono possibili varie opzioni: gli approcci più comuni descritti in precedenza o una combinazione di più approcci. È possibile scegliere un metodo per le funzionalità iniziali e un altro all'approssimarsi della messa in opera completa. La creazione di attività cardine di alto livello in questa fase è un buon modo per segmentare il processo e definire quando sarà disponibile una funzionalità e quali requisiti verranno soddisfatti con le varie release.

FASE 4: DOCUMENTARE L'ARCHITETTURA DESIDERATA

Una volta stabilito cosa migrare e quando dovranno essere disponibili le funzionalità, potrai cominciare a documentare l'architettura desiderata. Tieni presente che schemi e documentazione potranno variare in base alle esigenze. Dovrai comunicare con platee tecniche, di business e direzionali e ciascuna di esse si aspetterà livelli diversi di dettaglio sulla capacità operativa iniziale e finale. Le figure 1-4 illustrano alcune architetture di riferimento per aiutarti a esaminare le considerazioni comuni per la creazione di una piattaforma moderna.

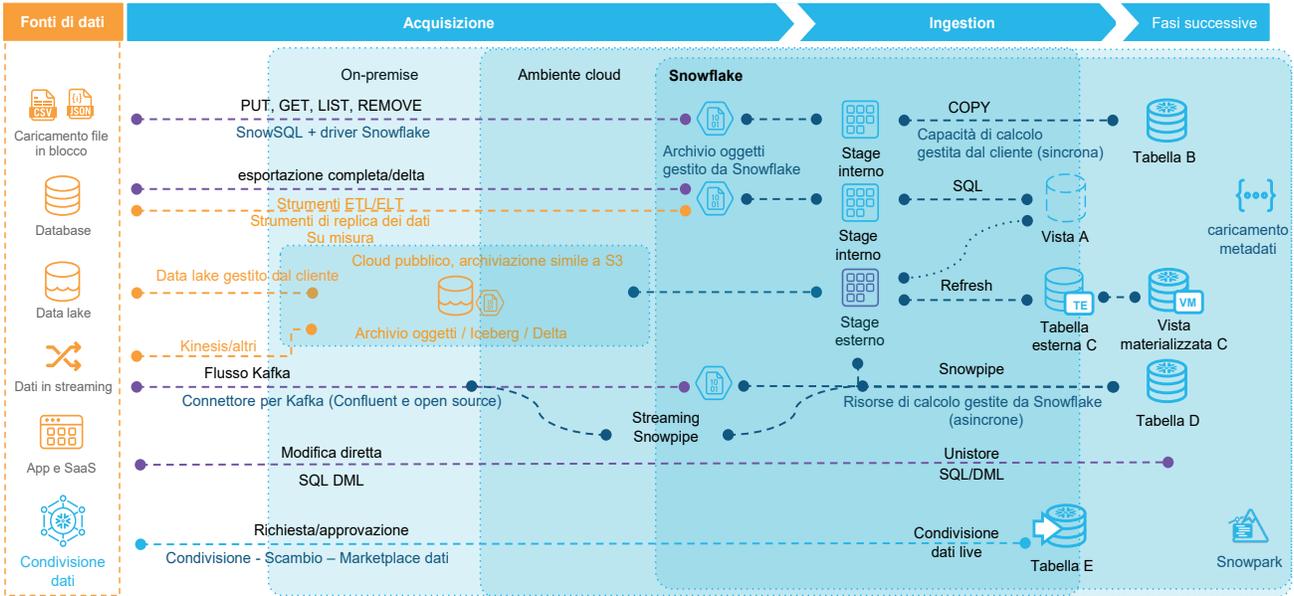


Figura 1: acquisizione e ingestione dei dati

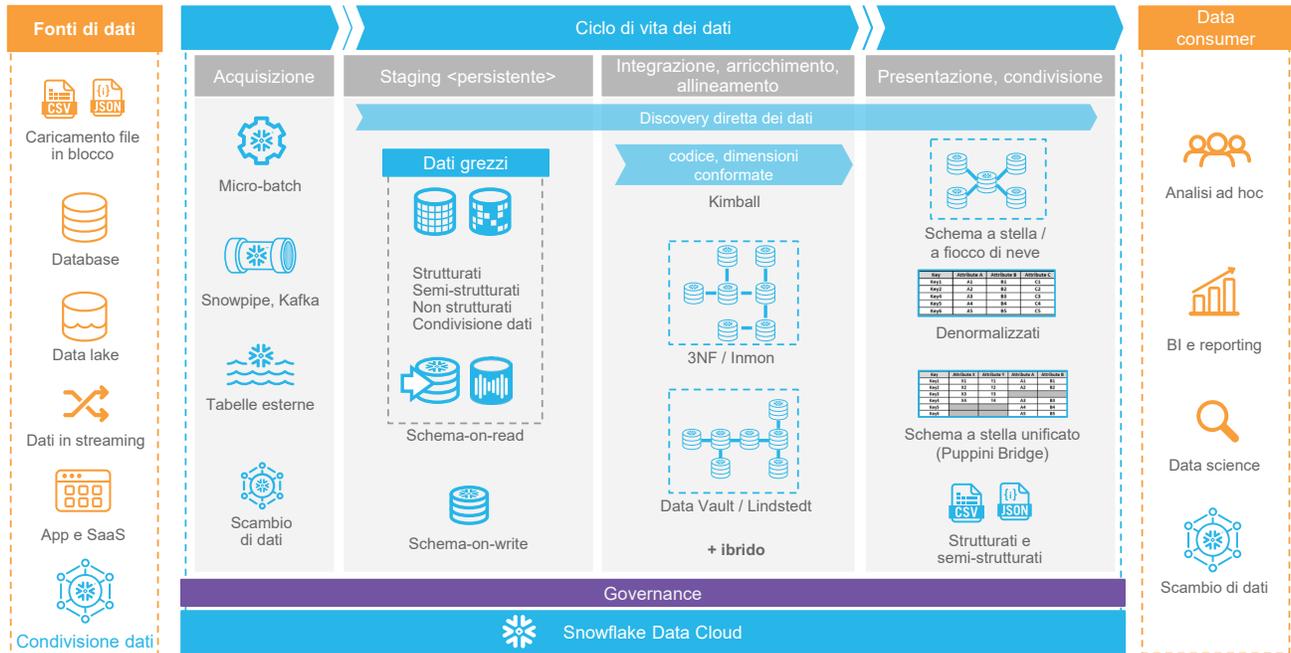


Figura 2: modellazione dei dati

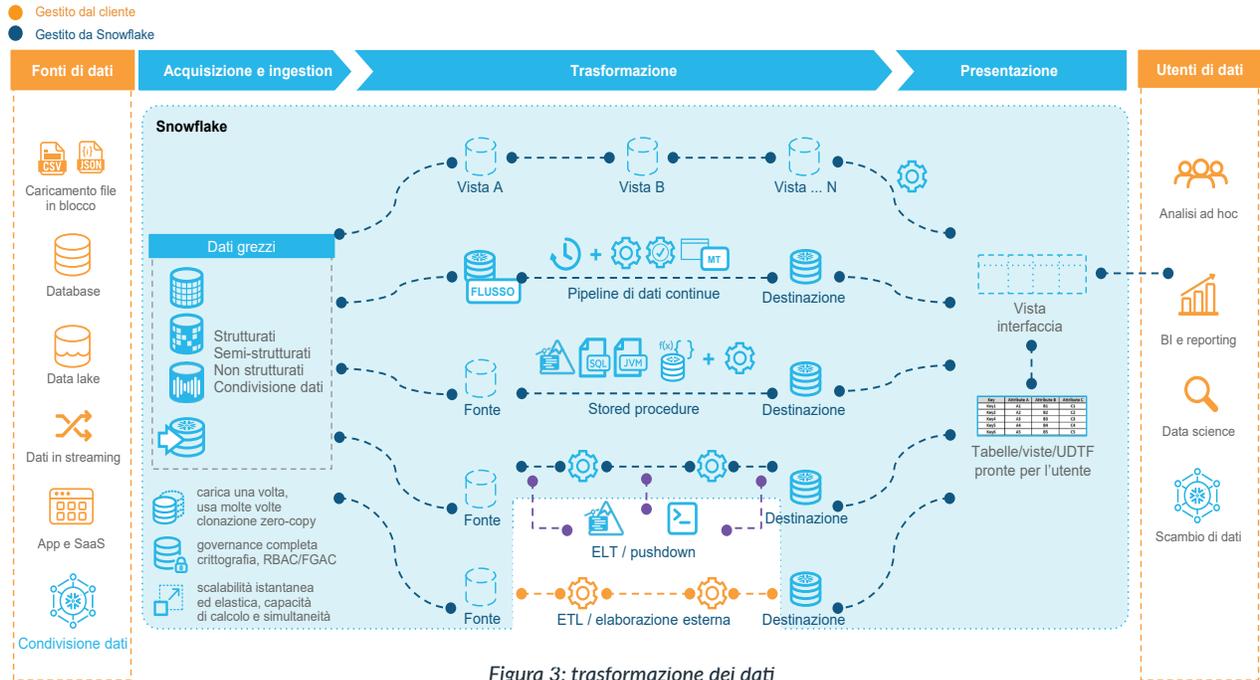


Figura 3: trasformazione dei dati

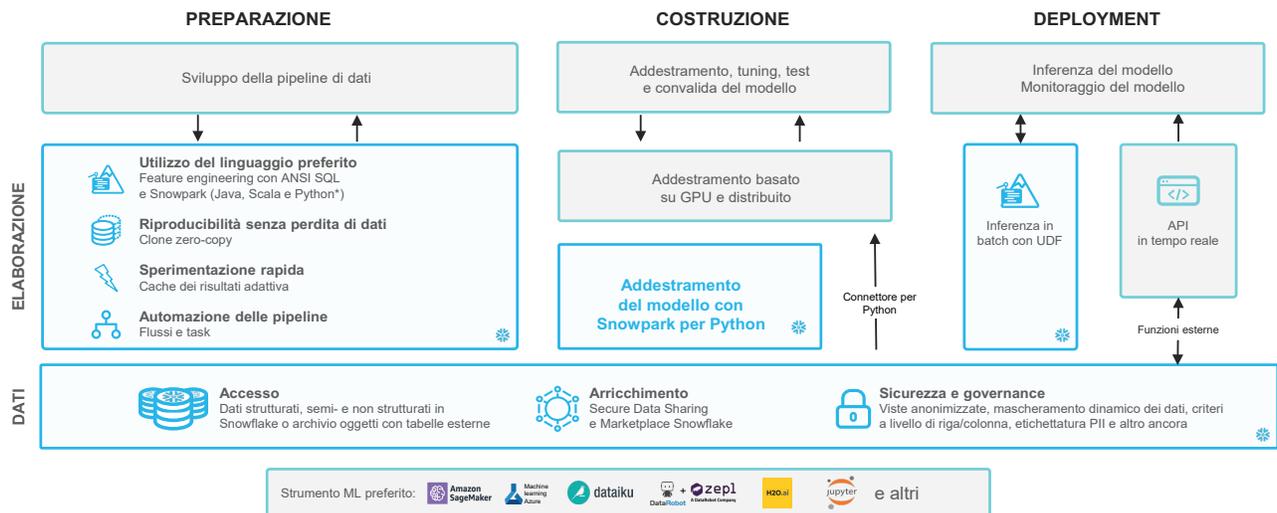


Figura 4: machine learning

FASE 5: PIANIFICARE IL CARICO DI DATI E IL “DIMENSIONAMENTO”

Uno dei compiti più difficili della migrazione al cloud è trasferire i dati e modificare il paradigma per sfruttare al meglio le risorse elastiche del cloud. È un puzzle composto da tre pezzi:

1. Carico iniziale

Il caricamento iniziale può essere complesso, a seconda dei volumi di dati e dei requisiti di sicurezza. Lavora a stretto contatto con il team di sicurezza e le LOB a cui appartengono i dati per non dover effettuare un processo di tokenizzazione/ offuscamento prima di trasferire i dati nel cloud. Quindi deciderai quali data set possono essere spostati, tenendo conto di normative e standard nazionali, locali e di settore in materia di trattamento e privacy dei dati, come GDPR, PII, PCI e HIPAA. Per quanto riguarda il volume dei dati, man mano che il networking migliora e si possono spostare terabyte di dati nel cloud, questo problema inizia a scomparire.

SUGGERIMENTO: molte organizzazioni ricevono dati da partner e fornitori esterni (ad esempio Salesforce). Potrebbe essere prudente conservare questi data set in un servizio di archiviazione oggetti nel cloud per evitare che vengano classificati come asset on-premise. Se i dati arrivano tramite internet, proteggili nel cloud.

2. Aggiornamenti continui

La fonte di dati, la logica ETL e l'integrazione determinano i metodi utilizzati per aggiornare i dati nella piattaforma. Alcuni strumenti ETL e fonti supportano il Change Data Capture (CDC). Altri potrebbero supportare tutti gli inserimenti, mentre altri ancora potrebbero richiedere un aggiornamento completo. Non esiste un approccio unico.

3. Pianificazione dell'utilizzo e dell'archiviazione

La maggior parte delle organizzazioni esegue un progetto pilota e valuta il ROI potenziale prima di procedere alla migrazione. In questa fase, di solito è utile convalidare nuovamente il piano di utilizzo e lavorare su aspetti operativi come le modalità di monitoraggio della disponibilità e di controllo dell'utilizzo del sistema. Poiché alcune piattaforme offrono la possibilità di scalare verticalmente, attivare e disattivare le risorse, segmentare workload e risorse e scalare automaticamente sia la potenza di elaborazione che lo spazio di archiviazione, si passa dal tempo impiegato per suddividere una risorsa fissa (limitando l'accesso degli utenti aziendali) all'allocazione delle risorse in base alle esigenze e al valore aziendali. Non è più necessario un grande lavoro di pianificazione per poter gestire il workload più grande, ma sottoutilizzando il sistema per gli altri 364 giorni dell'anno.



FASE 6: CONVERTIRE GLI ASSET

Questa fase consiste nel definire i data asset che potrebbe essere necessario convertire, tra cui il linguaggio DDL (Data Definition Language), il controllo di accesso basato sui ruoli (RBAC) e il linguaggio DML (Data Manipulation Language) utilizzati negli script. La buona notizia è che la maggior parte dei database relazionali utilizza lo standard ANSI-SQL. La maggior parte delle modifiche si concentrerà sulla corretta conversione dei formati DATE e TIMESTAMP e sulla verifica della conformità delle funzioni SQL utilizzate per accedervi. Tieni presente che non tutti i fornitori implementano le funzioni nello stesso modo. Alcune piattaforme semplificano il DDL eliminando la necessità di partizionare e indicizzare, rendendo il DDL molto più pulito (meno dettagliato).

FASE 7: CONFIGURARE L'AMBIENTE E TESTARE CONNETTIVITÀ E SICUREZZA

Come prevedibile, mentre prepari la migrazione dovrai completare le configurazioni di rete, proxy e firewall. Di solito è utile disporre di qualche diagramma che indichi le porte e gli URL a cui dovrai accedere. Dovrai inoltre collaborare con il gruppo di protezione per scaricare e installare eventuali driver (ODBC, JDBC e così via) o software di supporto, come un'interfaccia della riga di comando (CLI), utilizzata dalla maggior parte degli sviluppatori DBA per interagire con una piattaforma moderna. Occorre anche impostare i parametri dell'account, come l'elenco di IP consentiti e l'RBAC, prima di consentire l'accesso all'ambiente a gruppi più ampi.

FASE 8: TESTARE L'INTERO PROCESSO CON UN SOTTOINSIEME DI DATI

Dopo aver connesso tutti i set di strumenti, la best practice prevede un test del processo end-to-end per verificarne funzionalità e prestazioni.

Quando si migra un sistema esistente, si avrà il vantaggio di conoscere gli SLA per ogni LOB e i requisiti di ogni fase del processo: caricamento, trasformazione/aggregazione, interrogazione e così via. È sempre consigliabile selezionare casi pilota e data set essenziali per la riuscita iniziale del progetto. Questo è anche il momento migliore per apportare le "migliorie" associate alla migrazione, se segui questa metodologia.

FASE 9: MIGRARE I DATI E TESTARE I RISULTATI

Prima di mettere in funzione il sistema, dovrai eseguire nuovamente alcuni o tutti i test delle prestazioni della fase 8 per assicurarti che il sistema sia correttamente configurato per le query e per la simultaneità. Ciascuna linea di business strategica dovrà testare l'accesso al sistema per verificare che tutti i suoi strumenti funzionino e che le prestazioni siano quelle previste. I risultati calcolati dovranno essere uguali dal vecchio al nuovo sistema.

FASE 10: ESEGUIRE I DUE SISTEMI IN PARALLELO

Quando si sostituisce un data warehouse o un data lake aziendale on-premise, è facile caricare entrambi i sistemi ed eseguirli in parallelo. Da qui, potrai spostare gradualmente gli utenti o i gruppi nel nuovo sistema, cercando di non interrompere le operazioni aziendali. Concentrati sui gruppi che hanno espresso le maggiori difficoltà o preoccupazioni durante le fasi di pianificazione e definizione dei requisiti, perché saranno i più ricettivi e probabilmente diventeranno i più entusiasti sostenitori del progetto.

FASE 11: SCEGLIERE LA DATA DEL PASSAGGIO UFFICIALE

I sistemi possono coesistere per un po' di tempo, ma una volta che gli utenti finali avranno sperimentato i nuovi vantaggi, non vorranno più tornare indietro. Probabilmente sarà utile eseguire i sistemi in parallelo per almeno un ciclo di reporting completo: una settimana, un mese o un trimestre. Una volta scelta la data del passaggio ufficiale al nuovo sistema, continua a eseguire il vecchio in parallelo per l'eventualità che si verifichino problemi. Quando tutti saranno soddisfatti, sarà il momento di stappare lo spumante e di mandare in pensione il vecchio sistema.



TERZA PARTE

SFRUTTARE I VANTAGGI DELLA NUOVA PIATTAFORMA DATI NEL CLOUD

Una volta completata la migrazione, è il momento di iniziare a sfruttare le nuove funzionalità.

Una possibile strategia consiste nel puntare a migliorare le prestazioni e rendere disponibili i dati agli utenti finali più in fretta o distribuirli a un maggior numero di utenti.

Alcune opzioni includono:

STABILIRE LE ESIGENZE DELLE LOB E DEGLI UTENTI FINALI

In una fase precedente del processo, è stato compilato un elenco iniziale dei requisiti delle LOB. Ora è il momento di rivisitare quell'elenco e avviare un dialogo con le LOB per informarle delle potenzialità e dei vantaggi del nuovo sistema, tra cui:

- **Più dati, grazie all'accesso a record storici dettagliati riferiti ad anni, non solo mesi o settimane, precedenti**
- **Tipi diversi di dati strutturati, semi-strutturati (JSON, AVRO, XML, PARQUET, ORC) e non strutturati**
- **Dati più puliti**
- **Migliore formattazione/modellazione grazie al passaggio dagli schemi 3NF agli schemi a stella o ad altri modelli di dati**
- **Risultati più veloci, con tabelle di riepilogo o viste materializzate per migliorare le prestazioni**

ACCEDERE AI DATI DI PIÙ BUSINESS UNIT

Oltre ai data set specifici di ogni LOB, molte organizzazioni vogliono anche accedere alle informazioni più recenti e importanti di altre business unit e fonti aziendali. Alcune piattaforme moderne

forniscono vari metodi per controllare l'accesso ai dati e al tempo stesso consentire ad altri utenti finali di utilizzare data set curati. In alcuni casi, le organizzazioni condividono o ricevono set di dati da altre organizzazioni tramite FTP. Questa modalità di trasferimento dei dati è costosa e richiede molto tempo, soprattutto quando i volumi e la frequenza dei dati aumentano. Le piattaforme moderne offrono funzioni di condivisione dei dati che eliminano la necessità di trasferire e trasformare i dati, ottimizzando il data sharing e riducendo costi e complessità del processo ETL. Inoltre, sono più sicure e consentono di condividere dati governati aggiornati in modo da poter rispettare le normative sulla privacy dei dati.

OTTIMIZZARE (RIPENSARE) LA STRATEGIA DI CARICAMENTO

In genere, i data warehouse on-premise utilizzano finestre di caricamento in batch durante la notte per rendere i dati del giorno prima disponibili per l'analisi la mattina successiva. La tecnologia attuale consente di caricare e interrogare i dati senza contese, per cui si possono caricare i dati 24/7 mettendoli a disposizione degli utenti finali più in fretta. Le risorse possono essere aumentate o ridotte in modo istantaneo, il che consente di caricare i dati alla velocità desiderata e allo stesso prezzo, indipendentemente dalle risorse di calcolo che vengono attivate.

Supponiamo, ad esempio, che ogni notte vengano impiegate quattro ore per caricare 1 TB di dati. In alcuni casi si utilizza un singolo nodo per questo workload. Con una scalabilità quasi lineare, se si esegue lo stesso workload con un cluster a due nodi, il caricamento sarà due volte più veloce, ma avrà esattamente lo stesso costo. Raddoppiando nuovamente le dimensioni del cluster, la velocità di caricamento raddoppia di nuovo, ma il costo rimarrà esattamente uguale.

ANALIZZARE LE FONTI PER SEMPLIFICARE/RAZIONALIZZARE

In alcuni casi, gli strumenti e i sistemi utilizzati in passato potrebbero risultare superflui in futuro. Oggi le piattaforme moderne sono in grado di gestire nuovi tipi di dati ed elaborare i dati in modo più efficiente, sfumando le distinzioni tra OLTP, OLAP e data lake. Le grandi organizzazioni tendono ad avere molti strumenti, tutti acquistati per fini o funzionalità specifici.

Questo è un buon momento per razionalizzare gli strumenti ETL/ELT. Alcuni fornitori hanno aggiornato i propri prodotti perché possano funzionare in un ambiente cloud. Non cercare di standardizzare tutto su un unico strumento; offri piuttosto consigli tecnici per l'accesso a fonti specifiche o, se state implementando una strategia di data lake, utilizza strumenti appropriati per accedere a zone o livelli diversi.

Dovresti inoltre analizzare gli strumenti nel contesto della rapidità con cui possono spostare i dati dal sistema originale alla nuova piattaforma. La maggior parte delle organizzazioni sta passando dalle strategie ETL alle strategie ELT e di elaborazione di flussi di dati per sfruttare la scalabilità on-demand. Separando le risorse di calcolo da quelle di archiviazione, è possibile caricare i dati senza impatto sulle prestazioni di query, rendendo praticabile l'elaborazione dei dati quasi in tempo reale su larga scala. Dovrai semplicemente capire l'entità del cambiamento che la tua organizzazione può assorbire in una sola volta.

CARICAMENTO IDEMPOTENTE

Non ne hai mai sentito parlare? Non è un termine nuovo, ma sta iniziando a comparire sempre più spesso in molte conversazioni sull'elaborazione dei dati. L'obiettivo è semplice: caricare continuamente i dati nel sistema, ma se c'è qualche intoppo, vuoi sapere chiaramente cosa è stato elaborato e cosa no. Il caricamento idempotente significa che non importa se si carica un file o un record una sola volta o dieci volte: il risultato finale sarà sempre lo stesso. Questa potente funzionalità può essere ottenuta combinando il comando COPY con la trasformazione mediante i comandi MERGE disponibili nelle piattaforme moderne. È particolarmente utile per lo streaming di dati con aggiornamenti o eliminazioni, ma può essere utilizzata anche in altre situazioni.

ZERO-COPY CLONING, TIME TRAVEL E UNDROP

Alcune piattaforme moderne sono in grado di eseguire query a ritroso nel tempo, nel periodo compreso tra le 24 ore e i 90 giorni precedenti. Molte organizzazioni utilizzano questa funzionalità per trasformare le pipeline ELT. In questo modo, se, ad esempio, qualcosa va storto nel passaggio 45 di un processo in 50 passaggi, non sarà necessario ricominciare da capo. Basta utilizzare Time Travel per tornare a uno stato noto (il passaggio 44 che è stato eseguito correttamente), correggere il passaggio 45 e continuare l'elaborazione. Con la maggior parte delle piattaforme legacy, sarebbe necessario ricaricare o ripristinare la tabella originale e ricominciare da capo. Se ti serve uno snapshot del database o della tabella prima di eseguire un aggiornamento, utilizza Zero-Copy Cloning per creare una copia senza duplicare i dati. È una funzionalità estremamente utile, soprattutto per i processi ETL di grandi dimensioni che trasferiscono terabyte di dati all'interno di data mart per i data scientist o i team che si occupano di test e controllo qualità. Infine, hai eliminato accidentalmente un database o una tabella di produzione durante la modifica notturna? Nessun problema. UNDROP funziona molto bene in questa situazione e fa ricomparire magicamente tutti i dati e gli schemi.



CONCLUSIONE

SEMPRE PIÙ DATI NASCONO NEL CLOUD

Oggi le aziende conoscono i benefici in termini di efficienza e prestazioni dell'archiviazione, dell'analisi e della condivisione dei dati nel cloud. Il cloud elimina numerosi passaggi per le aziende e il caos che si può creare quando dati duplicati e suddivisi in silos sono scollegati dalla fonte originale. Una migrazione attentamente pianificata può apportare significativi vantaggi rispetto a un data warehouse o a un data lake on-premise, comprese un numero maggiore di funzionalità a costi inferiori.

SCOPRI DI PIÙ

Fai [clic qui](#) per scoprire di più su come modernizzare la tua piattaforma dati e per ottenere istruzioni specifiche per la migrazione dal tuo data warehouse o data lake on-premise esistente al Data Cloud di Snowflake.

INFORMAZIONI SU SNOWFLAKE

Snowflake permette a ogni organizzazione di mobilitare i propri dati grazie al Data Cloud. I clienti utilizzano il Data Cloud 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 AI/ML e analitici. 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 590 della classifica 2022 Forbes Global 2000 (G2K) al 30 aprile 2023, utilizzano il Data Cloud di Snowflake per far crescere le loro aziende.

Scoprite di più sul nostro sito [snowflake.com](https://www.snowflake.com).

