

Sfruttare tutto il potenziale del tuo ecosistema di dati

Comprendere le differenze tra motori di Data Lake, Data Virtualization e Data Fabric





Introduzione

In un mondo che sempre più ruota intorno ai dati, le aziende ne generano in grandi quantità e alcuni possono offrire informazioni preziose se analizzati in modo efficace. A tal fine, le aziende necessitano di una solida infrastruttura dati in grado di gestirne l'integrazione, la gestione, l'erogazione e il consumo.

I principali paradigmi per fornire agli utenti un accesso rapido ai dati sono, a oggi, i motori di data lake, la virtualizzazione dei dati e il data fabric. Sebbene i motori di data lake offrano funzionalità di query SQL comode e scalabili, potrebbero non essere in grado di gestire tutte le esigenze aziendali, motivo per cui occorre essere lungimiranti, dato che investire in un motore di data lake senza considerare le esigenze future potrebbe portare a rimpiangere la decisione in futuro. È pertanto essenziale confrontare le funzionalità dei motori di data lake, della virtualizzazione dei dati e del data fabric, prima di selezionare una soluzione che possa crescere insieme alle esigenze di un'organizzazione.

L'ascesa dei data lake

I data lake hanno acquisito una notevole popolarità negli ultimi anni, poiché consentono alle aziende di archiviare dati strutturati e non strutturati provenienti da un'ampia gamma di fonti. A differenza dei data warehouse tradizionali, che necessitano di dati strutturati e puliti prima dell'archiviazione, i data lake possono acquisire dati non elaborati, consentendo agli utenti di archiviare e analizzare le informazioni senza la necessità di una preelaborazione dispendiosa in termini di tempo e risorse.

Con i volumi di dati in continua crescita, le aziende hanno bisogno di soluzioni capaci di elaborare e analizzare in modo efficiente i dati per ottenere informazioni fruibili. I data lake forniscono questa flessibilità consentendo alle aziende di acquisire dati nel loro formato grezzo, senza uno schema predefinito. I dati vengono archiviati così come sono, consentendo un'acquisizione più rapida e i data lake si sono evoluti per aggiungere funzionalità, come la conformità ACID (Atomicità, Coerenza, Isolamento e Durabilità), per cui ora inglobano quelle dei data warehouse, tanto che spesso l'espressione "data lakehouse" si utilizza per riferirsi a questa architettura avanzata del data lake.

Perché spesso si ha bisogno di più di un data lake o data lakehouse

I data lake o data lakehouse si basano sulla filosofia secondo cui le aziende devono raccogliere e consolidare fisicamente i dati prima di poterne ottenere valore aggiunto. Per molti casi d'uso, questo approccio può funzionare, tuttavia:

- **Non esiste un'unica soluzione** che vada bene per tutti i casi e benché esistano molti tipi di piattaforme dati, come sistemi di gestione di database relazionali operativi, fonti NoSQL, data lake e data warehouse aziendali (ad esempio, Amazon Web Services (AWS) offre oltre 15 motori sviluppati appositamente per diversi modelli di dati, ciascuno specializzato in casi d'uso specifici), difficilmente un'unica piattaforma consentirà di gestire tutti i possibili casi d'uso.
- **Il consolidamento dei dati potrebbe non essere permesso**, ad esempio perché le norme sulla privacy e sulla sovranità dei dati si stanno evolvendo e alcuni stati hanno già limitazioni sulla relativa trasmissione dei dati al di fuori del paese di origine, mentre altri hanno leggi sulla privacy che limitano la divulgazione dei dati personali a terzi. Tutte queste limitazioni possono di fatto rendere impossibile trasferire i propri dati o inviarli a un fornitore terzo per archivarli o elaborarli.

Una moderna architettura dei dati deve prevedere un ambiente dati distribuito

Da oltre 30 anni, le aziende utilizzano data warehouse basati su architetture di dati centralizzate, tuttavia, in questo periodo, poche di queste società, per non dire nessuna, sono riuscite a consolidare con successo tutti i propri dati in un'unica piattaforma. Ciò potrebbe essere dovuto a uno o più dei motivi sopra indicati o a molti altri fattori e, sebbene le moderne versioni nel cloud di queste piattaforme rimuovano i vincoli delle tradizionali implementazioni on-premise, semplifichino la scalabilità e siano componenti chiave di una moderna architettura dei dati, gli stessi motivi che in passato rendevano poco pratico consolidare tutto su un'unica piattaforma valgono ancora oggi. Per riuscire a modernizzare un'architettura di dati, le aziende devono quindi aggiungere una gestione logica dei dati alle proprie strategie.



L'elemento mancante: la gestione logica dei dati

Una moderna architettura di dati deve trovare un equilibrio tra la raccolta fisica dei dati, come avviene con un'architettura centralizzata, e la connessione logica ai dati, che avviene invece con un'architettura decentralizzata. La gestione logica dei dati fornisce un livello di accesso unificato a tutte le risorse di dati aziendali, consentendo l'accesso immediato a qualsiasi dato senza doverlo prima copiare o replicare. I vantaggi della gestione logica dei dati includono:

- **facilità d'uso.** Gli utenti accedono a qualsiasi dato da un unico posto;
- **agilità nell'integrazione dati.** Un'unica piattaforma permette l'intera gamma di opzioni di integrazione dei dati, tra cui replica e trasformazioni complete, memorizzazione nella cache e federazione dei dati in tempo reale;
- **sicurezza e governance centralizzate.** Il controllo degli accessi e l'implementazione delle policy vengono applicati operando a livello logico di accesso ai dati, garantendone la coerenza complessiva;
- **soluzione a prova di futuro.** Separando l'accesso ai dati dalle caratteristiche tecniche delle sorgenti dove questi risiedono, il dipartimento ICT è libero di modificare le relative tecnologie sottostanti senza che ciò si rifletta sugli utilizzatori dei dati. Quando si apportano modifiche tecniche, infatti, sarà sufficiente riconnettere la rappresentazione logica dei dati interessati alla nuova sorgente, senza che tale cambiamento sia percepito dagli utilizzatori. In tal modo, le aziende possono mantenere la velocità operativa di cui hanno bisogno, mentre il dipartimento ICT può apportare le modifiche necessarie secondo i propri ritmi.

Come evolve l'ecosistema della gestione logica dei dati

La gestione logica dei dati non è un tema nuovo, con molte tecnologie già ben consolidate e nuove categorie di prodotti. In questo whitepaper esamineremo i punti di forza e i limiti di tre comuni approcci all'integrazione e gestione dei dati distribuiti, così da avere tutti gli elementi per poter prendere la decisione migliore.

- **I motori di data lake** consentono agli utenti di eseguire query SQL sui dati archiviati in un data lake e una delle loro caratteristiche principali è la capacità di migliorare le prestazioni delle query eseguite su grandi insiemi di dati con query SQL interattive. Per questo motivo, alcuni analisti chiamano questi sistemi "acceleratori di query", in contrapposizione alla prima generazione di data lake su Hadoop che prevedevano un'elaborazione batch più lenta. Questi sistemi utilizzano in genere un'architettura di query simile alle classiche soluzioni di gestione di database con elaborazione massicciamente parallela e spesso richiedono competenze specializzate per garantirne il funzionamento e la gestione.

Alcuni motori di data lake, inoltre, forniscono anche accesso in tempo reale a fonti esterne, sebbene spesso con funzionalità limitate. È quindi importante assicurarsi di capire cosa offrono i venditori e quali potrebbero essere i limiti di quanto offerto.

- **La virtualizzazione dei dati** consente agli utenti di accedere ai dati da molteplici fonti eterogenee mediante una vista logica unificata, capace di nascondere la complessità sottostante delle fonti di dati e di fornire agli utenti una visione semplificata e standardizzata delle informazioni. Ciò permette agli utenti di interrogare e analizzare i dati a partire da un modello semantico unificato, che astrae da come i dati sono fatti e da come sono distribuiti, cosa che consente una facile e non ambigua comprensione dei dati da parte di chiunque e senza richiedere particolari competenze.



- Un **data fabric** integra i dati in ambienti ibridi/multi-cloud e automatizza le attività di ingegneria e consumo dei dati, in modo da ridurne i tempi e aumentarne la semplicità.

Tra queste funzionalità troviamo l'integrazione dei dati, la relativa preparazione ed elaborazione, catalogazione ed individuazione, nonché governance e sicurezza dei dati. È opportuno notare che la virtualizzazione dei dati è spesso alla base della strategia dei data fabric, poiché consente l'accesso logico ai dati in tutto l'ambiente, piuttosto che optare sempre per una replica fisica, tanto che si parla appunto di logical data fabric, che per essere implementato in modo pieno, necessita comunque di ulteriori funzionalità rispetto a quelle offerte da un sistema di virtualizzazione.

Nel corso di questo whitepaper, ti sarà possibile acquisire una comprensione più approfondita di queste tecnologie chiave e dei vantaggi che ne conseguono. L'obiettivo del whitepaper è fornire elementi per scegliere al meglio come gestire un ecosistema dati e come estrarne il massimo valore.

L'approccio è quello di esaminare nel dettaglio i tre paradigmi già introdotti, mostrando come il passaggio dall'uno all'altro preveda l'introduzione di nuove funzionalità, fino a raggiungere quello che si ritiene sia il miglior modo di gestire, in modo olistico, il patrimonio informativo di un'organizzazione.

Le funzionalità necessarie per un approccio logico alla gestione dei dati

Le architetture di data fabric stanno riscuotendo sempre più interesse, principalmente perché forniscono un approccio più completo e flessibile alla gestione dei dati rispetto alle architetture tradizionali, alla base dei data warehouse e dei data lake. I data fabric, capaci di unificare i dati provenienti da più fonti e offrire funzionalità avanzate di elaborazione, consultazione e uso dei dati, consentono alle aziende di ottenere informazioni più dettagliate e di prendere decisioni più adeguate. Inoltre, i data fabric possono fornire controlli di governance e sicurezza più completi, semplificando il rispetto delle normative sulla privacy e sulla conformità dei dati.

Come accade spesso nel settore tecnologico, quando le nuove architetture diventano popolari, si crea molta concorrenza tra i fornitori nella nicchia di mercato dove queste si collocano, anche se non tutti i fornitori sono in grado di offrire lo stesso livello delle funzionalità richieste. In questa sezione, prendiamo quindi in considerazione i criteri stabiliti da una serie di analisti del settore per suddividere le principali funzionalità delle varie categorie di prodotti, seguiti da una descrizione di ciascuna funzionalità, in modo che sia possibile:

- comprendere le funzionalità disponibili;
- considerare le tue specifiche esigenze attuali e future;
- definire i criteri di valutazione.

NOTA: nella sezione Letture consigliate, alla fine di questo documento, sono indicati i link ai report degli analisti che sono stati consultati per compilare la seguente tabella.





1. Il **livello di accesso ai dati** semplifica l'accesso e l'integrazione dei dati da più fonti usando il linguaggio SQL, indipendentemente dal formato con cui i dati sono rappresentati sulle fonti originarie (astrazione sui dati). Offre inoltre una visione coerente e unificata dei dati ai data consumer, che disaccoppia il significato dei dati dalle loro caratteristiche tecniche e dalla loro distribuzione. Il livello di accesso ai dati include connettori per moltissime fonti dati, come database, applicazioni, API o servizi web. Nel valutare diverse tecnologie, è opportuno esaminare i connettori disponibili, la complessità delle loro funzionalità di pushdown e i sistemi compatibili, evitando di investire in soluzioni che potrebbero non essere compatibili con tutte le tue fonti di dati, attuali o future.
2. L'**ottimizzazione delle query del data lake** migliora l'efficienza delle query eseguite su un data lake, garantendo che il processo avvenga in modo efficiente e con prestazioni accettabili. Le tecniche includono l'elaborazione parallela massiva e diversi tipi di accelerazione delle query. L'ottimizzazione delle query del data lake è una caratteristica rilevante quando si seleziona uno di questi strumenti, perché può ovviamente migliorare le prestazioni delle query, ridurre i costi di trasferimento dei dati sulla rete e, in sostanza, migliorare l'efficienza complessiva del sistema. Va tenuto presente che i motori di data lake sono generalmente pensati per funzionare in modo molto efficiente quando un data lake è la fonte primaria di informazioni, ma tali prestazioni potrebbero non essere adeguate in ambienti distribuiti o

quando sia necessario integrare anche altre fonti dati, motivo per cui, nella scelta, è importante valutare la tipologia di dati che, al momento e nel futuro, potrebbero essere oggetto di integrazione, tenendo presente quanta parte di questi risiedono sul data lake, perché ciò potrebbe avere implicazioni sulle prestazioni complessive.

3. I **motori di query federate e distribuite** permettono agli utenti di accedere ed eseguire query sui dati da più fonti, come se facessero parte di un unico database, elaborando query che utilizzano più fonti di dati e combinare i risultati in un unico set di risultati da mostrare all'utente. Possono consultare e integrare dati provenienti da diversi tipi di fonti, tra cui database, applicazioni, servizi web e data lake, essendo in grado di identificare fonti di dati rilevanti, recuperare i dati richiesti e combinare i risultati in un'unica vista. Oltre a eseguire molte delle stesse tecniche impiegate dagli acceleratori di query stand-alone, i motori di query federate offrono altre tecniche, sviluppate e ottimizzate per funzionare con più fonti di dati, come il pushdown avanzato delle query, la riscrittura e l'accelerazione delle query, e la gestione della cache.
4. La **replica dei dati (salvataggio in memoria)** si riferisce alla creazione e alla gestione di copie di dati. La replica può assumere molte forme: processi di estrazione, trasformazione e caricamento (ETL) per la replica point-to-point, processi di estrazione, caricamento e trasformazione (ELT) per operazioni all'interno di un data lake o data lakehouse, micro-batching da API esterne, raccolta da canali di streaming, strutture di accelerazione e memorizzazione nella cache. Ogni caso d'uso è diverso e un'architettura moderna deve offrire quella flessibilità in grado di gestire ogni caso d'uso nel modo migliore, ottimizzando sia le prestazioni che le risorse utilizzate.
5. Il **repository di metadati** è l'elemento centralizzato che archivia i metadati che caratterizzano le fonti di dati, nei termini, ad esempio, il loro schema, i tipi di dati, le relazioni reciproche e le statistiche che le caratterizzano. Fornisce inoltre informazioni sui modelli dati degli utenti, come le loro mappature sulle fonti, la loro origine e le trasformazioni delle quali sono stati oggetto. Il repository può anche archiviare metadati relativi alle attività fatte (si veda il punto 11. Metadati attivi) e costituisce la base, non solo per l'ottimizzazione delle query, ma anche per processi di governance dei dati, come la gestione delle modifiche, le impostazioni di sicurezza e alcune funzionalità self-service descritte in modo approfondito in altre sezioni.
6. Il **livello di sicurezza e governance** offre meccanismi avanzati per gestire l'autenticazione e l'accesso a dati specifici. Ad esempio, se la colonna è contrassegnata come informazioni personali, è possibile impostare le autorizzazioni per non condividere tali dati o condividerli solo con un sottoinsieme di utenti autorizzati. I dettagli sono importanti in questo ambito. È essenziale determinare se la soluzione presa in considerazione permette funzionalità avanzate come un sistema di controllo degli accessi basato su attributi (ABAC) per garantire l'applicazione di politiche di sicurezza per accedere ai dati. Questo livello definisce e implementa politiche di sicurezza che si possono configurare in base a vari fattori, tra cui la riservatezza dei dati, ruolo dell'utente e requisiti aziendali. È anche utile capire se lo strumento che si sta considerando si integri o meno con gli strumenti di sicurezza e governance già in uso.
7. Il **livello di consumo dei dati** fornisce ai data consumer un accesso centralizzato a tutte le informazioni, permettendo loro di utilizzare gli strumenti che preferiscono. La maggior parte di questi strumenti è in genere compatibile con JDBC o ODBC, ma è importante considerare tutti i potenziali data consumer, in modo da avere una visione completa degli strumenti e delle modalità di consumo che saranno effettivamente utilizzati, che potrebbero richiedere la gestione di protocolli quali ADO.NET, MDX, servizi web RESTful, OData, GraphQL, GeoJSON, export verso Microsoft Excel/SQL, sottoscrizione a Kafka e code di messaggi JMS.
8. Il **modello semantico unificato** fornisce funzionalità per definire i dati che i data consumer possono esplorare, comprendere e utilizzare con facilità. Esistono numerose funzionalità legate alla modellazione semantica, come la definizione di modelli di dati derivati che hanno nomi parlanti e descrizioni associate, la definizione di metriche aziendali, l'assegnazione di tag e categorie di classificazione, i controlli di governance (come sostenere o deprecare un dato), l'attribuzione di etichette di qualità dei dati e molto altro ancora. Tutte queste funzionalità hanno in comune l'obiettivo di democratizzare l'accesso ai dati, rendendoli più comprensibili e più semplici da usare e, se queste sono esigenze e priorità sentite, allora la semantica dovrebbe essere senza dubbio un criterio determinante nella valutazione della soluzione più adatta.
9. L'**interfaccia e strumenti di design** aiutano nella gestione del sistema. Le funzionalità variano notevolmente a seconda dei fornitori, da strumenti che prevedono solo un'interfaccia da riga di

comando, a quelli con interfacce web evolute, che consentono ai data engineer e ai data consumer di usare il sistema in maniera intuitiva e senza richiedere competenze particolari. La valutazione sul tipo di interfaccia dipende ovviamente dalle specifiche esigenze di un'organizzazione, ma va tenuto presente che, se la democratizzazione dei dati è un aspetto prioritario, elementi come intuitività e facilità diventano elementi chiave nella scelta.

10. I **cataloghi di dati arricchiti** forniscono un inventario centralizzato e consultabile dei dati disponibili, dove ogni dato è “arricchito” con metadati, lineage e altre informazioni che ne facilitano la comprensione e l'uso. I cataloghi di dati arricchiti classificano ed etichettano automaticamente i dati in base al loro contenuto e alla loro struttura. Consentono inoltre agli amministratori dei dati e agli analisti di taggare e annotare manualmente i dati con informazioni aggiuntive, come definizioni aziendali, metriche sulla qualità e informazioni sull'utilizzo. Permettono infine agli utenti di scoprire, comprendere e usare, in velocità e semplicità, i dati disponibili, cosa che si traduce in una migliore collaborazione e nel prendere decisioni migliori.
11. I **metadati attivi** possono automatizzare le attività di ingegneria dei dati per ridurre la complessità e ottimizzare i risultati e il valore per l'azienda. Tradizionalmente, quando le persone parlano di “metadati”, si riferiscono a dati che descrivono altri dati. Nella moderna gestione dei dati, però, si usano metadati tecnici aggiuntivi, tra cui elementi come registri delle transazioni, accessi utente e piani di ottimizzazione delle query, con il fine, appunto, di creare metadati attivi. Questi ultimi sono dati che descrivono parametri quali le prestazioni del sistema e il modo in cui gli utenti vi interagiscono. I metadati attivi vengono utilizzati anche come parte delle iniziative FinOps, per promuovere la responsabilità condivisa riguardo all'infrastruttura dei dati e ai costi che questa ha per l'organizzazione.
12. I **motori di raccomandazione** sfruttano algoritmi di IA e machine learning per analizzare i metadati attivi, come le query effettuate, i dati acceduti e utilizzati, per identificare modelli di comportamento e fornire raccomandazioni al data engineer su come ottimizzare il funzionamento del sistema, e al data consumer su altri dati che potrebbero essere di suo interesse. In questo modo, le attività manuali di gestione dei dati vengono ridotte e i data consumer possono accedere velocemente ai dati di cui hanno bisogno.
13. I **livelli di preparazione e delivery dei dati** consentono agli utenti di accedere al livello semantico in un ambiente simile a una sandbox, nella quale possono esplorare e trasformare i dati per crearne di nuovi. Si tratta di una funzionalità fondamentale per abilitare il self-service e garantire che i data consumer possano eseguire quelle trasformazioni necessarie per adattare i dati alle loro specifiche esigenze, catturando al tempo stesso il loro feedback, in modo da poter evolvere la soluzione in una direzione guidata dai suoi utilizzatori.
14. **Orchestrazione e DataOps** gestiscono pipeline di dati e flussi di lavoro, automatizzando i processi che sono relativi ai dati. L'orchestrazione si riferisce al coordinamento di più attività e flussi di lavoro legati ai dati, spesso su più sistemi o piattaforme, che può comportare la gestione di pipeline di dati che integrano informazioni provenienti da fonti diverse, coordinando attività di elaborazione e analisi dei dati. DataOps, invece, fa riferimento all'applicazione alla gestione dei dati dei principi e delle pratiche DevOps, enfatizzando la collaborazione e l'automazione, migliorando la velocità, la qualità e l'affidabilità dei processi relativi ai dati.

I motori di data lake sono stati progettati affinché i data engineer li utilizzino per elaborare grandi volumi di dati strutturati e non strutturati mantenuti in un repository centralizzato. Con il tempo, sono poi stati migliorati per essere compatibili con ulteriori fonti esterne di dati. La virtualizzazione dei dati, invece, è stata sviluppata per consentire agli utenti di accedere ed eseguire query con dati provenienti da qualsiasi fonte, come se fossero tutti archiviati in un unico database. Sebbene i motori di data lake siano molto efficienti per le query quando tutti i dati si trovano nel data lake, le soluzioni di virtualizzazione dei dati in genere hanno prestazioni superiori quando le query interessano motori esterni, ad esempio quello di un data warehouse, o a dati che risiedono su fonti di dati distribuite.

Le esperienze degli utenti tendono a differire tra le due categorie di prodotti. Gli utenti tradizionali dei motori di data lake hanno generalmente un profilo tecnico e una familiarità con interfacce che spesso sono basate su riga di comando. Al contrario, la virtualizzazione dei dati è nata per aiutare gli utenti aziendali, a prescindere dalle loro competenze tecniche, nel trovare, comprendere e usare i dati di cui hanno bisogno. Per questo motivo, gli strumenti di virtualizzazione dei dati in genere dispongono di interfacce più intuitive da usare.

Infine, il data fabric, che utilizza le migliori funzionalità delle due categorie precedenti per fornire una vista unificata e integrata dei dati di più fonti e sistemi, comprese funzionalità avanzate aggiuntive in termini di governance, sicurezza, metodi di integrazione dei dati e molto altro ancora. Un'altra differenza tra il data fabric e le altre categorie di prodotti menzionate è l'attenzione posta sull'automazione di quelle attività che, in sua mancanza, vengono normalmente fatte in modo manuale.



Fornendo tutte le funzionalità sopra indicate, un data fabric costituisce una base solida per una moderna architettura dei dati, in grado di adattarsi ed evolversi per soddisfare le esigenze di quelle aziende data-driven, promuovendo e sostenendo agilità, flessibilità e innovazione, così che le aziende possano sfruttare al meglio tutto il potenziale dei propri dati.

Denodo lavora con una grande varietà di aziende, da società di medie dimensioni con implementazioni semplici, fino ad alcune delle più grandi aziende al mondo in diversi settori di mercato, con migliaia di utenti che, giornalmente, eseguono milioni di query su ecosistemi distribuiti di dati, come il caso di LeasePlan, leader globale nel car-as-a-service, con una flotta di circa 1,8 milioni di veicoli in 29 paesi, descritto nel prossimo paragrafo.

La Denodo Platform come elemento fondante del data fabric di LeasePlan

LeasePlan desiderava che il proprio modello di business analogico diventasse quello di un'azienda car-as-a-service completamente digitale, ampliando il proprio portafoglio di servizi e promuovendo attività che incrementassero i ricavi con prodotti e servizi innovativi. La difficoltà era che LeasePlan disponeva di grandi volumi di dati comportamentali, di marketing, provenienti dai social network, informazioni sul traffico, nonché su servizi e manutenzione, che erano distribuiti su molteplici fonti eterogenee (SAP, Salesforce, IBM DB2, Snowflake, ecc.) e suddivise in silos. Pertanto, l'integrazione e la distribuzione dei dati rappresentavano una sfida per un sano processo decisionale aziendale, per l'ottimizzazione dei processi, per la creazione di nuovi modelli di business e per il rispetto delle normative europee. LeasePlan ha quindi scelto la Denodo Platform, una piattaforma di gestione logica dei dati che sfrutta la virtualizzazione dei dati e che ha consentito all'azienda di creare un logical data fabric con un unico punto di accesso a tutte le sue principali fonti di dati, moderne, legacy e nel cloud che fossero. Il logical data fabric, che ricorre alla virtualizzazione dei dati per nascondere agli utenti le complessità dell'accesso diretto, ha inoltre consentito a LeasePlan di connettersi a un'ampia gamma di nuove fonti e applicazioni che utilizzano dati. Ciò ha consentito a LeasePlan di cambiare facilmente tecnologia, in caso di necessità, senza condizionare l'operatività aziendale. Subito dopo l'implementazione, il logical data fabric ha sostituito i data warehouse preesistenti dell'azienda ed è diventato il livello centrale di accesso ai dati, fornendo un quadro centralizzato per la governance dei dati, il logical data fabric, inoltre, ha consentito di proteggere i dati dei clienti e garantire il massimo livello di sicurezza dei dati.

Grazie al logical data fabric, LeasePlan ha potuto migliorare l'esperienza del cliente, poiché i dati disponibili hanno permesso ai gestori della flotta di ottenere informazioni più dettagliate. Combinando le informazioni di un veicolo con lo storico del resto della flotta, LeasePlan è ora in grado di prevedere in modo molto più preciso quando a un'auto potrebbe necessitare di un servizio di manutenzione e, di conseguenza, indirizzare in tempo i propri clienti verso le officine meccaniche che preferiscono o persino programmare loro appuntamenti di manutenzione. LeasePlan è ora in grado di migrare i dati in locale al cloud senza che ciò implichi tempi di inattività nell'uso e nell'analisi dei dati e gli ha consentito di migrare i data mart preesistenti al cloud, creando una vista unica tra varie unità aziendali in diversi paesi.

“Il nostro nuovo logical data fabric, sviluppato con la tecnologia di Denodo Platform, è un elemento fondamentale del nostro Global Data Hub: costituisce la base per il self-service dei dati presso LeasePlan e ci ha permesso di sviluppare nuovi servizi aziendali e supportare al meglio le persone alla guida della nostra flotta in ogni fase dei loro spostamenti”



Sumit Arya, direttore degli Enterprise & Data Analytics per LeasePlan Digital.

Riguardo la Denodo Platform

La Denodo Platform è una soluzione di gestione logica, integrazione e distribuzione dei dati che crea le basi per una moderna strategia di data fabric. I nostri clienti la usano per una business intelligence self-service, analisi avanzate, integrazione ibrida/multi-cloud e soluzioni di data service aziendali. Offrendo tutte le funzionalità identificate dagli analisti del settore ed evidenziate nei precedenti paragrafi, Denodo Platform permette alle aziende di sviluppare un logical data fabric su molteplici fonti di dati, per un accesso unificato e senza interruzioni operative.

Dispone inoltre di funzionalità aggiuntive che la rendono la soluzione leader per creare un logical data fabric:



Catalogo dati integrato, che semplifica l'accesso ai dati con raccomandazioni automatiche, opzioni avanzate di collaborazione, profilatura e funzionalità di ricerca intelligente.



Connettività universale, che facilita la connettività alla più ampia gamma di fonti di dati, con oltre 200 connettori a sistemi di terzi.



Accelerazione intelligente delle query, che migliora l'esecuzione di query da più fonti tramite la memorizzazione intelligente nella cache, l'ottimizzazione in base ai costi e un motore di suggerimenti basato su algoritmi di intelligenza artificiale.



Motore di data lake integrato con elaborazione a parallelismo massivo, che fornisce agli utenti un'interfaccia altamente scalabile e un efficiente motore SQL, basato sulla tecnologia di Presto, con cui accedere ai dati del data lake, rendendo più semplice l'accesso ai dati mediante un'interfaccia grafica intuitiva.



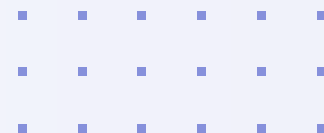
Controllo degli accessi basato su attributi (ABAC), che combina il livello semantico, l'assegnazione di tag e attributi di sessione per permettere politiche di accesso altamente granulari, definite a livello globale secondo diversi fattori, tra i quali la riservatezza dei dati, i ruoli utente e i requisiti definiti dalle politiche aziendali.



FinOps, che fornisce visibilità sui costi operativi, per gestire al meglio i costi dell'infrastruttura cloud.



Ampia gamma di opzioni di delivery dei dati, offrendo alternative ai semplici protocolli JDBC o ODBC, affinché gli sviluppatori dispongano di una scelta più ampia possibile, tra cui API REST e GraphQL, Kafka e code di messaggi JMS.



Riconosciuto come soluzione leader dagli analisti del settore, come il **Forrester Wave: Enterprise Data Fabric**, secondo trimestre 2022 e il **Gartner Magic Quadrant 2023**, nella categoria di strumenti di integrazione dei dati.

I clienti, inoltre, hanno confermato Denodo come “Customers’ Choice” nella categoria strumenti di integrazione dati secondo “Voice of the Customer” di Gartner® Peer Insights™ 2023. Denodo è l'unico fornitore di data management ad aver vinto questo premio per tre anni consecutivi.

Come iniziare

Con questo whitepaper speriamo di aver reso evidenti i vantaggi del logical data fabric: offrire la flessibilità e la scalabilità del cloud con la facilità d'uso derivante dall'aver un unico punto in cui gli utenti possono trovare, comprendere, integrare e condividere i dati in modo sicuro, per prendere decisioni aziendali nel modo più veloce ed informato possibile.

Conclusioni

I dati saranno sempre distribuiti, ma il logical data fabric fornisce un modo per connettersi in tempo reale, secondo le necessità e senza doverli spostare fisicamente in un archivio centralizzato.

Nel whitepaper sono state analizzate 14 funzionalità chiave che gli analisti del settore hanno identificato come prioritari per la scelta tra le soluzioni di gestione logica dei dati attualmente disponibili, ognuna delle quali dispone di un mix leggermente diverso di queste funzionalità.

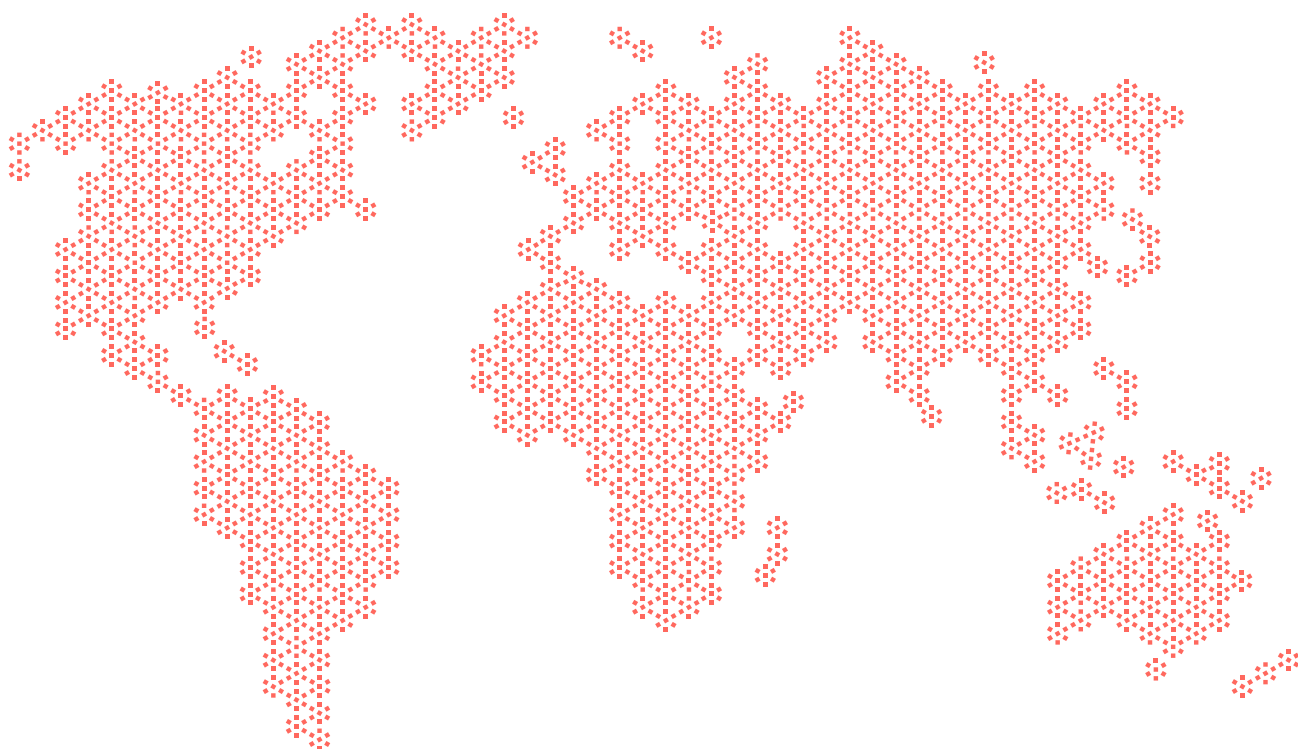
Benché alcune aziende con esigenze altamente specifiche e che non prevedono grandi cambiamenti in futuro potrebbero trovarsi bene con un motore di data lake o uno strumento classico di virtualizzazione dei dati, la necessità di massimizzare la flessibilità e rendere gli investimenti a prova di futuro, si ritiene valga la pena considerare una soluzione capace di offrire tutte le 14 funzionalità citate prima, secondo il paradigma di un logical data fabric, in modo da assicurarsi la massima resilienza in caso di evoluzioni future, sfruttando i vantaggi di un'infrastruttura dati moderna, che nel contesto aziendale dinamico e altamente competitivo di oggi, può implicare la differenza futura tra successo e oblio.



Lettere consigliate:

1. **Gartner - Demystifying the Data Fabric**
2. **Gartner - Assessing the Relevance of Data Virtualization in Modern Data Architectures**
3. **The Forrester Tech Tide™: Enterprise Data Integration, Q4 2021**
4. **Denodo - Logical Data Fabric Whitepaper**

Nota: Potrebbe essere necessario un abbonamento per consultarne i contenuti.



Denodo è leader nella gestione dei dati. La pluripremiata Denodo Platform è la piattaforma leader per l'integrazione, la gestione e la distribuzione dei dati che utilizza un approccio logico per una business intelligence self-service, data science, integrazione dei dati ibrida/multi-cloud e data service aziendali. Con un ritorno sull'investimento superiore al 400% e profitti di milioni di dollari, i clienti di Denodo di grandi e medie imprese in oltre 30 settori hanno ammortizzato l'investimento in meno di sei mesi.

Visita www.denodo.com | E-mail info@denodo.com | Scopri community.denodo.com