

A short vertical green line is positioned to the left of the text.

QUALITÀ E GOVERNANCE DEI DATI

Massimizzare il valore dei dati nell'epoca dell'IA

I prodotti di dati permettono di ottenere risultati più velocemente, rafforzando la governance

Nell'attuale ambiente aziendale, caratterizzato da ritmi frenetici, i dati sono un asset strategico fondamentale, il cardine per ottimizzare operazioni e favorire innovazione e crescita. I dati non si limitano a rappresentare numeri e statistiche, veicolano anche insight, trend e pattern che mostrano la via per il successo. Oggi più che mai, con i riflettori puntati su IA generativa e machine learning, le aziende si pongono una questione fondamentale:

come sfruttare rapidamente il pieno potenziale dei dati per ottenere risultati più velocemente, aumentando al contempo la fiducia e la governance?

La velocità è l'elemento fondamentale.

Secondo un recente studio di IDC, il 75% dei dirigenti afferma che i dati perdono valore nel giro di pochi giorni¹.

La vera sfida non si limita alla sola gestione di enormi volumi di dati. Implica anche la trasformazione delle informazioni in asset curati, di alta qualità, immediatamente accessibili e adattati a specifici settori aziendali. E come sa chiunque lavori nel settore, è più facile a dirsi che a farsi. I leader di dati e analytics affrontano molte sfide che minano la piena realizzazione delle loro iniziative:

- **La gestione dei dati basata su progetti manca di scalabilità.** Ogni team di progetto agisce secondo un proprio set di parametri, tra cui scopi, obiettivi e stakeholder. Questo approccio frammentato mina la collaborazione interfunzionale e limita l'estrazione di valore dai dati. E i rischi sono vari: disallineamento degli stakeholder nei vari progetti, sforzi duplicati, lacune nella governance, utilizzo inefficiente dei dati, costi elevati e, di conseguenza, risultati aziendali ritardati o non raggiunti.

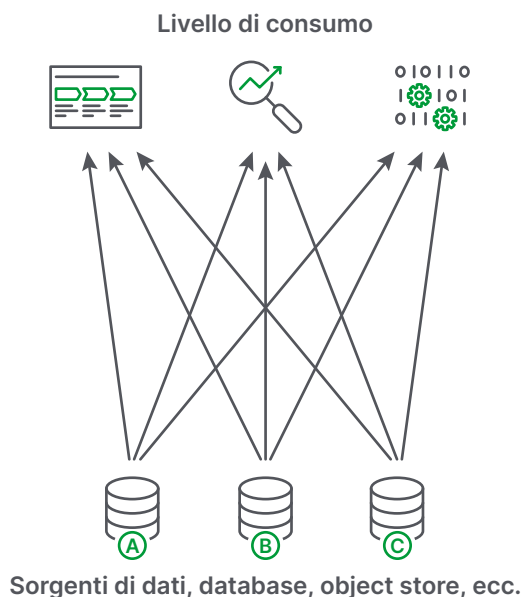
- In azienda **c'è una mancanza di consistenza nella qualità dei dati dei vari sistemi**, che inficia i processi decisionali e riduce la fiducia negli insight guidati dai dati.
- Nei dati e nella catena di fornitura di valore **non c'è chiarezza su proprietà e responsabilità** tra produttori di dati e loro fruitori, il che mina i tentativi di estrarre valore alla velocità necessaria all'azienda.
- **Basarsi su soluzioni frammentate e approcci ad-hoc e improvvisati** alla gestione dei dati allunga il time-to-value, mentre manutenzione e scalabilità sono costose se manca il personale adeguatamente formato.

Queste sfide aumentano il divario tra produttori e consumatori di dati. Da una parte, i produttori di dati, che si concentrano sull'IT, danno priorità alle piattaforme anziché capire le modalità di utilizzo e le esigenze dei fruitori. Dall'altra, i fruitori di dati, chiusi nei propri reparti, faticano a far comprendere le proprie esigenze e non si fidano dei dati forniti. **Per colmare questo divario sempre più ampio, le organizzazioni hanno bisogno di un approccio strategico e lineare alla gestione dei dati.**

¹Blog di IDC, "Navigating the Planes of Enterprise Intelligence Architecture", giugno 2023.

Cambiare gli approcci alla gestione dei dati

Tradizionalmente, le organizzazioni hanno utilizzato due approcci distinti per estrarre valore dai dati: l'approccio **frammentato** e quello **centralizzato**.



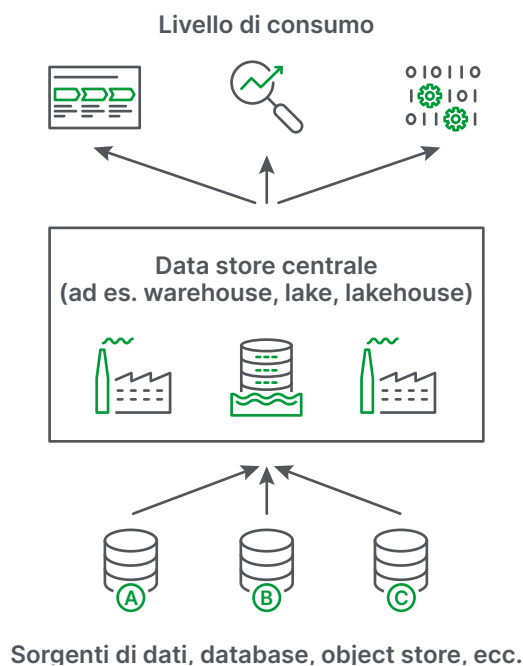
L'APPROCCIO FRAMMENTATO

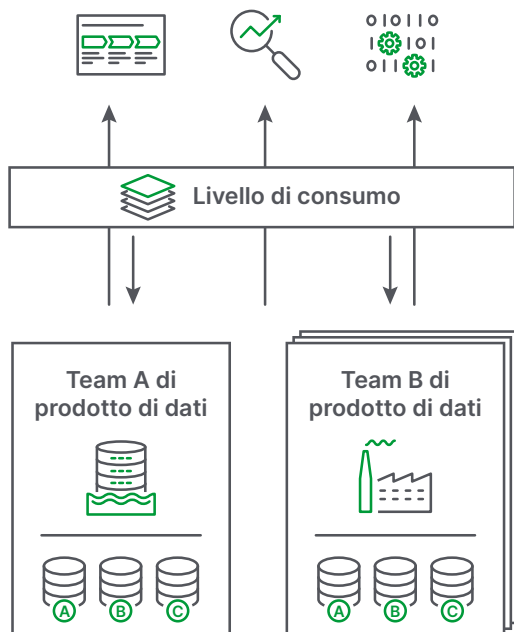
Con l'approccio frammentato, i vari team di "data-project" raccolgono dati da diversi sistemi sorgente per affrontare casi d'uso specifici dell'azienda. Ogni team usa i dati in modo indipendente, con strumenti e tecnologie proprie per estrarre, integrare e analizzare i dati e soddisfare le proprie esigenze. Se questo metodo genera valore immediato sul breve periodo, non è però scalabile per le grandi aziende sul lungo periodo, perché crea più silos di dati, raddoppia gli sforzi, aumenta i costi e ritarda i risultati aziendali.

L'APPROCCIO CENTRALIZZATO

Al contrario, l'approccio centralizzato dà il compito a un team di dati centrale di estrarre, pulire e aggregare dati su larga scala. L'obiettivo di questo team è soddisfare le esigenze di analisi dell'intera organizzazione. Implementa complesse pipeline per raccogliere dati da vari sistemi, pulirli, consolidarli in data lake in cloud, warehouse o lakehouse (come Databricks Delta Lake, Snowflake, BigQuery, S3 ecc.).

La centralizzazione può sembrare la soluzione alla frammentazione, e in un certo senso lo è, ma presenta un punto debole: richiedendo una grande coordinazione tra i reparti per apportare variazioni o aggiornamenti, frena l'agilità organizzativa. Inoltre, concentrandosi sulle esigenze comuni a tutta l'organizzazione, il team di dati centrale spesso ignora le necessità specifiche e le preferenze di stakeholder e fruitori aziendali, creando una soluzione non ottimale che non soddisfa le esigenze di tutti.





L'APPROCCIO FEDERATIVO

Entrambi gli approcci tradizionali, frammentato e centralizzato, presentano dei limiti e non soddisfano appieno le esigenze di fruibilità di dati dell'azienda. In risposta a queste sfide, si sta facendo largo un cambio di prospettiva. Sempre più popolare, l'**approccio federativo** distribuisce strategicamente la fornitura dei dati sfruttando i prodotti di dati e gestendo in modo centralizzato le funzioni di ingegneria della piattaforma.

Diversi team di prodotti dei dati possono lavorare in parallelo, e ognuno si concentra sui prodotti di dati specifici per il proprio reparto. Questo approccio velocizza i risultati sfruttando la collaborazione interfunzionale, coinvolgendo stakeholder sia economici sia tecnici.

L'approccio federativo presenta diversi vantaggi importanti:



Agilità e flessibilità

Snellendo la gestione dei dati end-to-end, l'approccio federativo consente alle organizzazioni di **ottenere velocemente valore** dai dati. Promuove l'agilità sostenendo un approccio basato sul prodotto in cui la fornitura e la creazione di valore sono distribuiti tra i team cross-funzionali, mentre le funzionalità di ingegneria della piattaforma sono gestite a livello centrale. In questo modo si ottengono **risposte più rapide alle mutevoli necessità aziendali**, assicurando allo stesso tempo integrità dei dati e responsabilità lungo tutto il processo.



Contenimento dei costi

Il modello federativo promuove il **riutilizzo**, aiutando le organizzazioni a massimizzare l'investimento e ridurre i costi operativi. Inoltre, la condivisione dall'alto di componenti fondamentali ai vari team di prodotti **riduce il lavoro ridondante e favorisce la sostenibilità sul lungo periodo**.



Distribuzione più rapida dei casi d'uso

La collaborazione facilitata tra team interfunzionali porta a **risultati rapidi e specifici per il settore**, personalizzati per rispondere alle esigenze mutevoli dell'azienda e a nuove iniziative guidate dai dati. L'approccio federativo aiuta la **scalabilità e favorisce la personalizzazione di casi d'uso specifici** attraverso flussi di sviluppo parallelo tra i vari team di prodotti di dati.



Aumento della fiducia

Il modello federativo **promuove la fiducia migliorando governance, trasparenza e conformità**, infondendo sicurezza nei processi decisionali guidati dai dati. Una chiara proprietà dei domini di dati favorisce l'allineamento e la responsabilità tra stakeholder di dati e analisi, rafforzando la governance dei dati dell'organizzazione.

Introduzione ai prodotti di dati

L'approccio federativo è la base di partenza migliore per i prodotti di dati perché per molte organizzazioni rappresenta un cambiamento rivoluzionario: dal controllo alla collaborazione, dal progetto al prodotto, dalla rigidità all'agilità, il tutto fornendo valore in scala. I [data mesh](#) e [data fabric](#), diversi nel loro approccio alla gestione dei dati, creano una potente sinergia per sfruttare appieno il potenziale dei prodotti di dati. Nel paradigma del data mesh, la proprietà dei dati è decentralizzata, posseduta e gestita dai team che la usano. I prodotti di dati sono l'aspetto più tangibile ed efficace di questo paradigma, mentre il data fabric diventa l'architettura fondamentale di gestione dei dati, che consente la fornitura ottimale di prodotti di dati ai team di reparto².



Secondo il sondaggio Gartner CDAO Agenda Survey for 2024, il 50% delle aziende che ha partecipato al sondaggio ha già usato prodotti di dati e un altro 29% li sta testando o ne sta considerando l'utilizzo entro il prossimo anno³.

Ma cosa sono i prodotti di dati? I prodotti di dati forniscono dataset di alta qualità, curati, pronti all'uso anche da parte dell'IA, facilmente accessibili e applicabili a diversi casi d'uso all'interno di un'organizzazione. Un prodotto di dati è una specie di microservizio per i dati. Come i microservizi rendono disponibili le funzionalità delle applicazioni tramite unità più piccole, contestuali e composite, un prodotto di dati dà accesso a dati rilevanti e destinati a uno specifico reparto attraverso interfacce di servizi. Un prodotto di dati risponde a diversi criteri fondamentali:



Orientamento e proprietà del dominio

Ogni prodotto di dati deve essere allineato a uno specifico dominio e avere un proprietario (persona o team) responsabile del suo successo, ad es. generare valore e aumentarne l'adozione.



Product thinking

I prodotti di dati vengono creati applicando il product thinking ai dati, tenendo in considerazione proprietà, controllo delle versioni e consumatori (in questo caso i fruitori di dati). I prodotti sono potenziati tramite la comprensione degli schemi di utilizzo e i feedback dei fruitori. Hanno un loro ciclo di vita, proprio come i prodotti reali.



Standardizzato e interoperabile

I prodotti di dati seguono formati standardizzati, schemi e convenzioni relative ai metadata. Facilitano l'interoperabilità e la semplicità di integrazione con altri componenti dell'ecosistema di dati.



Trovabile e individuabile

I prodotti di dati sono registrati in un catalogo e sono facilmente individuabili dai fruitori. Un prodotto di dati si autodescrive: quali dati rappresenta, come accedervi, la semantica aziendale, la provenienza e la qualità.



Riutilizzabile

I prodotti di dati vengono costruiti una volta e poi usati in vari casi d'uso. Altri prodotti di dati vengono creati usando prodotti di dati esistenti.

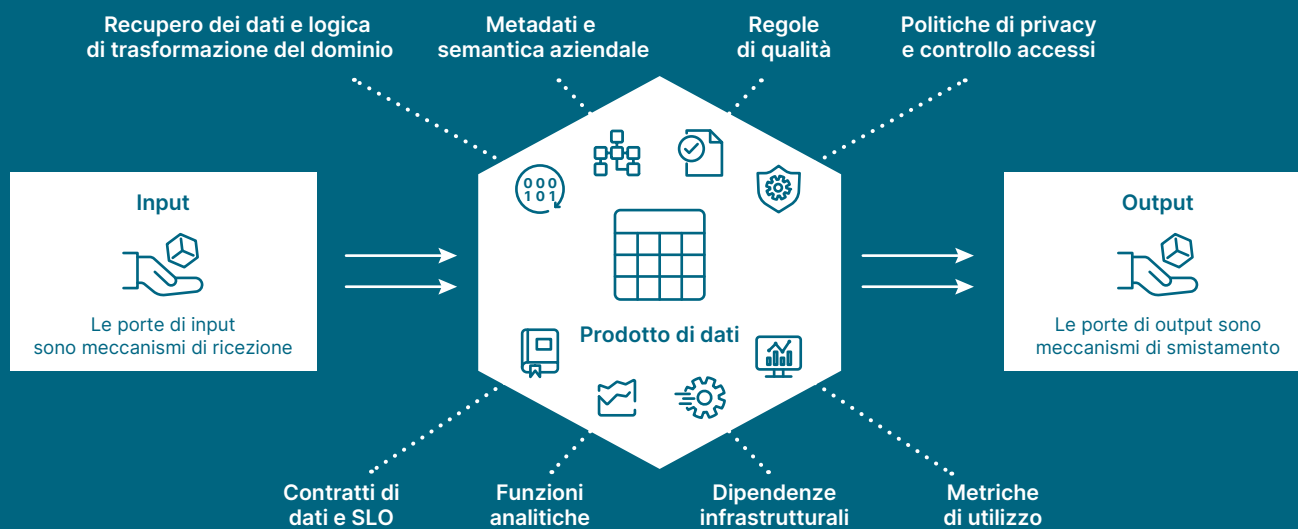


Accessibile

I prodotti di dati sono accessibili tramite un set standard di interfacce come SQL o REST API che sono definite e presentate quando vengono creati i prodotti di dati. L'interfaccia nasconde la complessità del prodotto di dati.

² Gartner: 6 Lessons Data Leaders Can Learn from Adopters of Data Mesh by Robert Thanaraj et al., 4 marzo 2024. GARTNER è un marchio registrato e un marchio di servizio di Gartner, Inc. e/o delle sue filiali negli Stati Uniti e nel mondo. L'utilizzo del marchio nel presente documento è stato autorizzato. Tutti i diritti riservati.

³ Gartner: Gartner Chief Data and Analytics Officer Agenda Survey for 2024. 12 marzo 2024. I risultati di questo studio non rappresentano una conclusione globale o il mercato nel suo intero, ma riflettono il sentiment dei partecipanti e delle aziende intervistate. GARTNER è un marchio registrato e un marchio di servizio di Gartner, Inc. e/o delle sue filiali negli Stati Uniti e nel mondo. L'utilizzo del marchio nel presente documento è stato autorizzato. Tutti i diritti riservati.



Un prodotto di dati è un quantum autosufficiente che racchiude in un'unità dati, trasformazioni, codice, politiche di gestione, SLA, schemi di accesso e dipendenze infrastrutturali. Il diagramma qui sopra mostra l'anatomia dei prodotti di dati.

- 1 Dataset:** esiste come centro del prodotto. Può essere un set di tabelle, file, grafici o rappresentazioni vettoriali. All'interno di un prodotto di dati possono essere presenti vari livelli: un set grezzo con i dati così come sono acquisiti da un sistema sorgente, un set curato con dati puliti e integrati e un dataset pronto all'uso modellato per la semplicità di utilizzo.
- 2 Input:** rappresenta le porte di input o la sorgente di dati in entrata. Possono essere tabelle di un'applicazione SAP da acquisire, tabelle in Snowflake da usare come sorgente, un altro prodotto di dati da cui si ricava quello nuovo o una richiesta di ricezione di feed di dati in tempo reale.
- 3 Output:** rappresenta i meccanismi di smistamento dei dati per il consumo. Un prodotto di dati può collegarsi a una varietà di interfacce standardizzate come SQL, API, ricerca vettoriale, messaggi e file. Anche le porte di output variano in base al tipo di servizio fornito dal prodotto di dati. Un output può fornire accesso a dati storici, un altro all'ultima istantanea sui dati.
- 4 Recupero dei dati e logica di trasformazione del dominio:** rappresenta il codice e la logica della pipeline per acquisire i dati (ad es. per spostarli da un database sorgente a una piattaforma cloud o per leggere un set di tabelle) e applicare una logica di trasformazione per passare dal formato sorgente a quello del dataset target, che può essere rappresentato da tabelle, viste o integrazione di vettori.
- 5 Metadata e semantica aziendale:** contengono i metadata tecnici (nome tabelle, nome file, definizioni colonne) e aziendali (dominio, nome prodotto, tipo di prodotto, scopo, proprietà, tag ecc.) per consentire di comprendere meglio il prodotto di dati.
- 6 Regole di data quality:** definiscono i controlli di qualità da eseguire sul prodotto di dati. Ogni regola contiene informazioni di configurazione su quando eseguirla, il suo scopo, se eseguirla su tutti i dati o in modo incrementale ecc. Le regole di data quality sono eseguite in momenti prestabiliti per assicurare l'accuratezza e aumentare la fiducia nei dati.
- 7 Politiche di privacy e controllo accessi:** stabiliscono le policy per il controllo degli accessi ai dataset del prodotto di dati. Per esempio, se il dataset contiene informazioni di identificazione personale (PII), la politica di privacy può mascherarle a tutti gli utenti o garantire l'accesso solo a un determinato gruppo.
- 8 Contratti di dati e SLO:** Definiscono i valori target per gli obiettivi di servizio che i prodotti di dati devono adempiere. Per esempio, un SLO può richiedere un livello di accuratezza della qualità dei dati maggiore del 95% o un aggiornamento ogni 5 minuti.
- 9 Funzioni analitiche:** ulteriori logiche di business personalizzate per rendere riutilizzabili le funzioni. Una funzione di un prodotto di dati per un ordine di acquisto può riguardare gli ordini inadempienti e includerà la complessa logica di dominio che determina quali ordini siano inadempienti. Le funzioni analitiche si possono esporre tramite una porta di output, come un'API.
- 10 Dipendenze infrastrutturali:** informazioni sui requisiti infrastrutturali dei dataset associati a un prodotto di dati. Ad esempio, i dataset sono conservati in una piattaforma (come Snowflake o Databricks) con un codice associato (es. logica di trasformazione) che richiede di eseguire un ambiente di calcolo (es. SQL).

Ciclo di vita del prodotto di dati

I prodotti di dati non sono entità statiche. Come i prodotti materiali, richiedono manutenzione e monitoraggio costanti. Ogni prodotto di dati è posseduto e gestito dal data team di quel dominio, che è responsabile della sua capacità di generare valore, di aumentarne l'adozione e di mantenerne il ciclo vitale.



Di seguito una panoramica delle tappe fondamentali del ciclo di vita di un prodotto di dati:

- 1 Concettualizzazione.** Il processo inizia con l'identificazione delle necessità specifiche del settore, inclusa la comprensione degli utenti, dei loro casi d'uso e del modo in cui solitamente consumano i dati.
- 2 Progettazione.** Si passa poi a creare una soluzione che includa l'individuazione dei dataset necessari e delle loro fonti e a definire le caratteristiche del prodotto di dati in termini di funzionalità, interattività, approccio di presentazione ecc.
- 3 Sviluppo.** Dopo la fase di design, è il momento di costruire l'infrastruttura tecnica e le funzionalità del prodotto di dati. Ciò include la raccolta e l'integrazione di dataset grezzi dalle sorgenti e la loro trasformazione, assicurandone la qualità.
- 4 Distribuzione.** È la fase per rendere accessibili i dati, che include il completamento della documentazione del prodotto, l'istituzione della proprietà, l'integrazione con i sistemi aziendali esistenti o l'utilizzo di workflow per un accesso immediato e l'attivazione del prodotto di dati o la sua preparazione all'uso.
- 5 Marketizzazione.** Dopo l'attivazione è importante far conoscere il prodotto e promuoverne l'uso. Questa fase include la formazione dei fruitori e la creazione di canali di comunicazione per recepire i loro feedback.
- 6 Uso.** Dopo un lancio di successo, i consumatori inizieranno a interagire con il prodotto di dati per cercare, capire, comprendere e ottenere accesso. Questa fase si concentra sul monitoraggio dell'utilizzo e sulla raccolta di feedback. Solitamente i dati raccolti riguardano le forme di interazione degli utenti, le prestazioni e la capacità di rispondere alle necessità dell'azienda in quel settore.
- 7 Mantenimento.** I prodotti di dati non sono entità statiche, devono essere aggiornati per preservarne il valore. Basandosi sui feedback degli utenti e sulle mutate esigenze dell'azienda, vengono proposte nuove dimensioni, funzionalità e caratteristiche da aggiungere nella nuova versione del prodotto di dati.

Seguendo queste tappe, le organizzazioni possono creare prodotti di dati che forniscono insight validi, portano a decisioni guidate dai dati e consentono il raggiungimento degli obiettivi aziendali.

Cosa NON è un prodotto di dati

Il mondo dei dati è pieno di termini il cui utilizzo non è sempre coerente. Ad esempio: l'espressione "prodotto di dati" spesso è confusa con "dataset" o "catalogo di dati", tuttavia non sono la stessa cosa. Questa sezione illustra alcune differenze fondamentali tra questi concetti.



Anche se un dataset è un componente fondamentale di molti prodotti di dati, è importante comprendere che è solo uno dei vari componenti. Un prodotto di dati va ben oltre un semplice dataset grezzi.

DADASET: LA MATERIA PRIMA

Un dataset è solitamente un insieme organizzato di data point, strutturato o meno. Può contenere informazioni grezze come dettagli sui clienti, rapporti finanziari o letture provenienti da sensori. I dataset sono come una materia prima (ad esempio una catasta di legno) che ha un potenziale, ma deve essere prima lavorata. Un aspetto fondamentale che distingue un dataset da un prodotto di dati è la mancanza di un contesto specifico per quel settore aziendale.

I PRODOTTI DI DATI: LA MERCE FINITA

I prodotti di dati generano valore prendendo dataset meticolosamente curati ed elaborati, spesso uniti ad altri artefatti discussi nella sezione precedente. Riprendendo la metafora del legname, i prodotti di dati sono bellissimi mobili costruiti con quel legno (ovvero i dataset).

Ecco cosa differenzia i prodotti di dati dai dataset:

- **Contesto e centralità del settore.** I prodotti di dati sono progettati per soddisfare esigenze specifiche dell'azienda o richieste degli utenti. Vanno oltre i dati grezzi, includendo informazioni contestuali come definizioni, provenienza e controlli di qualità.
- **Usabilità e accessibilità.** I prodotti di dati sono progettati per aumentare il valore e facilitare l'uso. Sono curati e assemblati per essere pronti al consumo da parte di fruitori di un settore specifico con limitate competenze tecniche.
- **Insight utilizzabili.** I prodotti di dati non si limitano a presentare i dati, ma offrono insight operazionali, analitici, predittivi e prescrittivi. Tali insight possono essere forniti tramite dashboard di visualizzazione, modelli analitici precostituiti o link di integrazione ad altri strumenti tramite SQL o API.
- **Proprietà e responsabilità.** I prodotti di dati presentano chiare indicazioni di proprietà, garantendo così la responsabilità riguardante produzione, utilizzo e feedback all'interno dell'organizzazione. La proprietà di un prodotto dei dati dipende dal contesto aziendale, il proprietario può essere chi li ha prodotti, il fruitore del team di settore responsabile per quei determinati casi d'uso, oppure una figura emergente nota come Data Product Manager.



I prodotti di dati trasformano i dataset in risorse di valore fruibili che favoriscono l'assunzione di decisioni informate e la crescita aziendale. L'Harvard Business Review stima che "le aziende che trattano i dati come un prodotto riducono il tempo necessario a implementarli in nuovi casi d'uso fino al 90%, abbattano i costi di gestione (tecnologia, sviluppo e mantenimento) fino al 30% e abbassano il rischio e il peso della governance dei dati"⁴.

⁴ "A Better Way to Put Your Data to Work", Harvard Business Review, luglio-agosto 2022.

IL CATALOGO DEI PRODOTTI DI DATI: L'HUB PER I PRODOTTI FINITI

Un catalogo di prodotti di dati è un archivio complessivo che ospita i prodotti di dati e i metadati associati (descrizioni, tag, versioni, etichette di proprietà ecc.) per rendere più facile trovare e comprendere i prodotti di dati. Si rivolge sia ai produttori di dati sia ai fruitori, con soluzioni su misura per soddisfarne le esigenze. Per i Data Product Manager, il catalogo fornisce strumenti per la configurazione, l'attivazione e la gestione dell'intero ciclo di vita del prodotto di dati, dalla creazione, all'attivazione fino al fine vita. Per i fruitori, il catalogo è un marketplace che consente esperienze simili al processo di acquisto e ne promuove l'utilizzo.

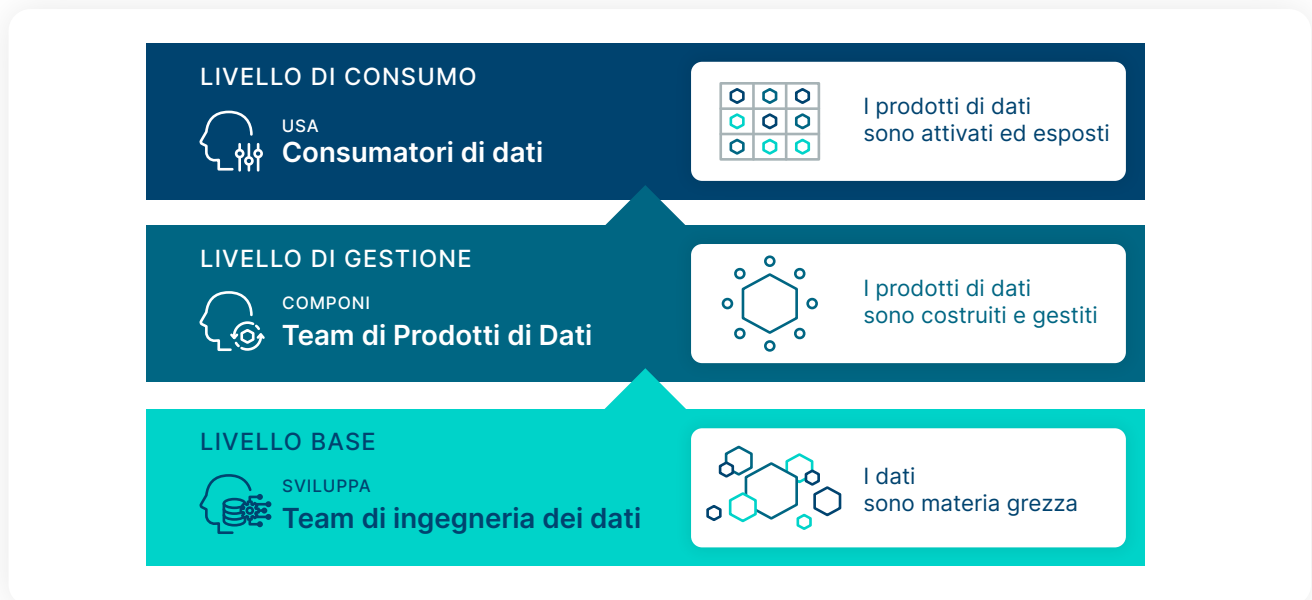
Qui sotto gli elementi che distinguono un catalogo di prodotti di dati da un data catalog:

CARATTERISTICA	CATALOGO DATI	CATALOGO PRODOTTI DI DATI
Mission	Rafforzare la governance dei dati	Aumentare il valore dei dati
Ambito di applicazione	Centrato sull'azienda	Centrato sul settore
Modello di gestione	Centralizzato	Federativo
Finalità	Accessibilità, conformità, trasparenza	Usabilità, fiducia, fattibilità
Contenuto	Metadati di vario tipo: tecnici, operativi e aziendali collegati a ogni asset di dati	Prodotti di dati affidabili, metadati sul ciclo di vita dei prodotti di dati, documentazione dei prodotti di dati e proprietà
Entità	Nell'ordine di centinaia di migliaia	Nell'ordine di centinaia
Pubblico	Soprattutto governance dei dati	Produttori e fruitori di dati

Nel contesto della gestione dei dati, la sfida non è solo creare piattaforme di dati ma anche assicurarne l'effettiva adozione. Molto spesso le organizzazioni investono risorse importanti (solitamente milioni di dollari) per creare infrastrutture di dati e poi si trovano ad affrontare un grande problema: **le barriere nel data-sharing tra produttori e fruitori**. Perché? Perché i risultati generati da molte piattaforme tecnologiche hanno poca affidabilità e non riescono ad allinearsi agli specifici obiettivi aziendali dei fruitori. **In conclusione, l'azienda ha bisogno sia di un catalogo di dati sia di un catalogo di prodotti di dati**. Sono come le corsie di un supermercato: in una si trovano le materie prime come farina, burro e zucchero, nell'altra i prodotti finiti come torte e biscotti.

Il catalogo di prodotti di dati: livelli e attori

Il catalogo di prodotti di dati è l'hub fondamentale per organizzare i prodotti di dati di un'organizzazione. È strutturato su tre livelli, ognuno pensato per una diversa categoria di stakeholder. Nel caso di ingegneri di dati, steward, analisti o figure emergenti come il Data Product Manager, ogni gruppo interagisce complessivamente sui livelli, in base ai propri obiettivi e requisiti. Questo approccio garantisce la separazione dei problemi basati sulle necessità degli utenti, con il giusto livello di astrazione per facilitare l'adozione e l'utilizzo.



IL LIVELLO BASE

La base del catalogo di prodotti di dati è un livello fondamentale che stabilisce l'infrastruttura tecnica e fornisce le funzionalità che supportano i livelli successivi.

Le funzionalità principali di questo livello sono:

- **Data engineering.** Spostare ed elaborare i dati da varie fonti, tra cui file, data warehouse in cloud, data lake e flussi di dati in tempo reale.
- **Gestione dei metadati.** Raccogliere e organizzare informazioni preziose sui dati, come la provenienza, i glossari di dati, gli schemi ecc.
- **Sicurezza e controllo degli accessi.** Avviare controlli di sicurezza per proteggere i dati e assicurare l'accesso solo agli utenti autorizzati.
- **Tracciamento di provenienza e data quality.** Monitorare la qualità dei dati per assicurare accuratezza e tracciare la provenienza per una maggiore affidabilità.

Il livello di base è rivolto principalmente al team della piattaforma di dati, incluse le figure tecniche e pratiche come ingegneri di dati, steward e architetti. L'obiettivo è costruire le funzionalità del livello come self-service in modo che anche gli utenti non tecnici possano utilizzarle. Le funzionalità richieste dal livello base si trovano in vari strumenti di gestione dei metadati, di acquisizione e trasformazione dei dati, di data quality e di sicurezza.

IL LIVELLO GESTIONALE

Il livello gestionale dei prodotti di dati, un concetto nuovo, è in grado di soddisfare i bisogni della figura emergente del Data Product Manager (o dell'intero team). Questo livello prevede funzionalità di sviluppo, gestione e mantenimento della fiducia nei prodotti di dati orientati al settore. Si basa sui servizi del livello base per fare l'inventario degli asset di dati, spostarli, trasformarli e metterli al sicuro. Fa da ponte tra i metadati grezzi che si trovano nel livello base e gli insight utilizzabili che i fruitori cercano nei dati.

Le funzionalità principali di questo livello sono:

- **Composizione del prodotto di dati.** Ingegnerizzare un prodotto di dati, definire input e output e configurarne tutti i vari artefatti.
- **Utilizzo del prodotto di dati.** Gestire le versioni e il flusso di lavoro per attivare un prodotto di dati quando è pronto per essere utilizzato dagli utenti finali.
- **Gestione del prodotto di dati.** Tracciare l'utilizzo, raccogliere feedback, cambiamenti e aggiornamenti e cogliere tutte le fasi del ciclo di vita del prodotto di dati descritte nelle sezioni precedenti.

Il livello gestionale del prodotto di dati serve principalmente al team di product management, soprattutto ai Data Product Manager, che guidano il ciclo di vita dei prodotti di dati per soddisfare gli obiettivi aziendali.

IL LIVELLO DI CONSUMO (OVVERO DEL MARKETPLACE)

Il livello del marketplace si trova in cima al catalogo dei prodotti di dati ed è la parte del sistema rivolta ai fruitori. Consente l'esplorazione e l'utilizzo autonomo dei dati, a supporto di una cultura guidata dai dati all'interno dell'organizzazione.

Le funzionalità principali di questo livello sono:

- **Ricerca e scoperta.** Le interfacce intuitive consentono agli utenti di esplorare e scoprire facilmente i prodotti di dati più adatti alle loro esigenze.
- **Accessi e permessi.** Questo livello controlla chi può accedere a determinati prodotti di dati, assicurando sicurezza e conformità alle politiche di governance dei dati.
- **Apprendimento e fiducia.** I fruitori di dati comprendono i prodotti di dati guardando i metadata e i dati campione, imparando a usarli osservando il codice di esempio e la documentazione, e in questo modo vi ripongono fiducia rivedendo la qualità e la provenienza dei dati.
- **Interfacce per l'utilizzo dei dati.** API, strumenti di visualizzazione o anche access point diretti al database che forniscono agli utenti i mezzi per utilizzare i prodotti di dati. A seconda del prodotto, sono disponibili diverse interfacce per rispondere alle varie necessità e capacità degli utenti.

Il livello di consumo dei dati è rivolto soprattutto ad analisti aziendali, scienziati di dati, utenti delle unità aziendali e altri dirigenti.

BENEFICI DEI PRODOTTI DI DATI PER I VARI ATTORI DELL'ECOSISTEMA

I dati sono la maggiore risorsa di un'organizzazione, ma il loro vero potenziale può essere sfruttato solo se velocemente accessibili, ben compresi e usati in modo efficace. I prodotti di dati aiutano in questo senso, trasformando dati grezzi in risorse curate e fruibili a supporto degli stakeholder in tutta l'organizzazione.

Vediamo l'impatto dei prodotti di dati sugli attori principali dell'ecosistema



Per i **team di ingegneria dei dati**, i prodotti di dati sono un linguaggio comune (un accordo ben definito) che consente una comunicazione strutturata e chiara tra i produttori e i fruitori di dati. Ciò consente ai produttori di sviluppare prodotti di dati riutilizzabili in modo iterativo, riducendo i costi totali di gestione. Inoltre, il team di data engineering non deve costruire grandi pipeline di dati centralizzate e complesse: può concentrarsi sullo sviluppo di pipeline indipendenti e più piccole dedicate a ciascuna implementazione di prodotto.



Per i **team di data governance** i prodotti di dati sono un punto di audit fondamentale, e dato che i prodotti sono affidabili e gestiti, è più facile ridurre i rischi. Il team di gestione definisce policy e standard a livello centrale, vari team di prodotti di dati rendono operative le policy in modo federativo.



Per i **Data Product Manager** i prodotti di dati aiutano a trovare un punto di incontro tra produttori e fruitori migliorando la collaborazione e la responsabilità in tutto l'ecosistema. Ciò garantisce proprietà e responsabilità dei dati.



Per i **fruitori di dati** i prodotti di dati sono risorse curate e specifiche che forniscono esattamente quello che serve per massimizzare il valore aziendale. Si riducono le rielaborazioni e si ottengono risultati più velocemente.

Tutti insieme, questi stakeholder possono sfruttare un catalogo di prodotti di dati per sfruttare il valore delle risorse di dati in modo efficiente, prendendo decisioni data-driven e raggiungendo gli obiettivi aziendali.

“ Nel settore dei mutui, i prodotti di dati non sono meri numeri e algoritmi: li consideriamo un fattore di trasformazione. Definendo dei prodotti di dati curati e self-service, vogliamo fornire ai nostri addetti ai prestiti la soluzione giusta per ogni persona che richiede il mutuo, mentre il team finanziario può valutare meglio costi e rischi. L'utilizzo dei prodotti di dati in azienda sta realmente ridefinendo il futuro del settore mutui. ”

Julia Fryk

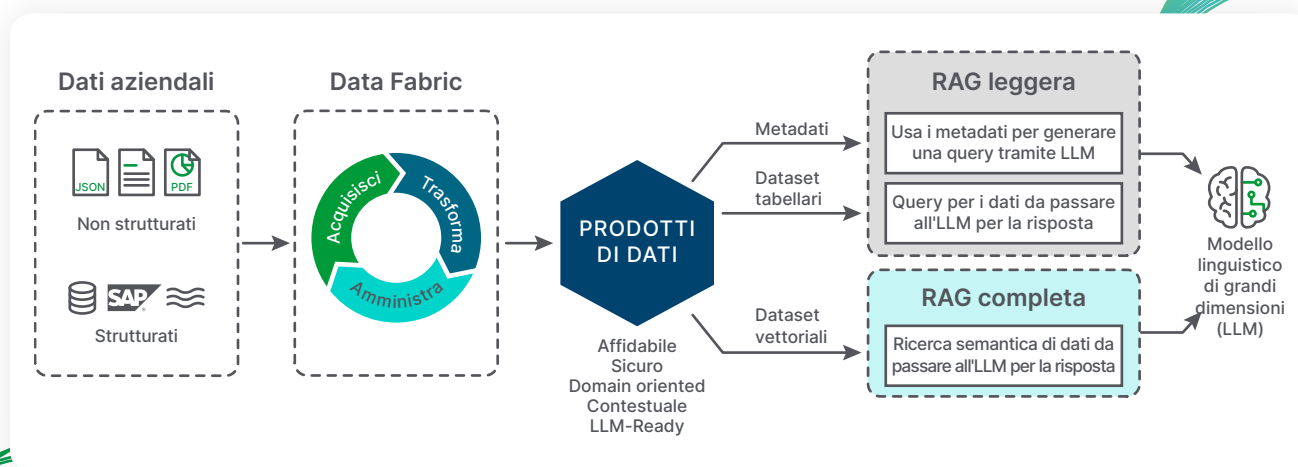
Principal Data Architect, Waterstone Mortgage

Fornire dati contestuali ai LLM con i prodotti di dati

L'IA generativa sta creando enormi opportunità per aumentare la produzione in qualsiasi area aziendale. Il cuore della GenAI sono i modelli di base (LLM come GPT-3, Anthropic, Llama e BERT) che vengono pre-addestrati con ampi corpora di dati testuali (pagine web, documenti ecc.). Tuttavia, i modelli di base non comprendono i dati interni e specifici dell'azienda. Per questo devono essere arricchiti con dati aziendali.

Una delle tecniche più efficaci per fornire dati aziendali ai LLM è un sistema detto RAG o generazione potenziata dal recupero. Come suggerisce il nome, i dati contestuali relativi a un prompt (domanda) dell'utente vengono **recuperati**, il prompt viene **potenziato** con questi dati e passato al LLM per generare una risposta. Esistono due versioni di RAG, RAG leggera e RAG completa.

- Con la RAG leggera (per dati strutturati), l'applicazione RAG passa i metadati al LLM per generare una query SQL. Questa query viene eseguita in un dataset a tabella e il risultato viene fornito al LLM per generare la risposta. Questo metodo si usa quando è richiesta una risposta esplicita dai dati.
- Con la RAG completa (per dati strutturati e non), i dati vengono prima resi vettoriali. Poi l'applicazione RAG esegue una ricerca semantica su questi dati vettoriali per recuperare i dati. Tali dati sono poi passati al LLM per generare una domanda.



In entrambi i sistemi è importante avere dati di qualità, puliti e contestuali disponibili per la RAG. **I prodotti di dati forniscono dati pronti all'uso contestuali e specifici in molti modi.** Come detto in precedenza, i prodotti di dati contengono vari metadati (aziendali, tecnici e specifici) e dataset altamente curati (file, tabelle o vettori) pronti all'uso. Dunque, i prodotti di dati diventano un veicolo perfetto per le applicazioni basate su RAG.

Per costruire un'applicazione RAG si inizia cercando nel catalogo di prodotti di dati o il marketplace il giusto prodotto di dati da usare in base alla domanda posta. Ad esempio, per una domanda riguardante le vendite potrebbe servire un prodotto di dati "Opportunity" proveniente dal CRM, basato su metadati dettagliati associati al prodotto di dati. Oppure una domanda relativa a un ordine potrebbe aver bisogno di un prodotto di dati "Sales Order" proveniente dal SAP, basato su metadati dettagliati associati a quel prodotto di dati. Una volta identificato il prodotto di dati, l'applicazione RAG può interrogare direttamente il dataset del prodotto di dati (in tabella o vettore) basato sul caso d'uso.

Conclusione

Mentre i dati continuano a guidare le aziende moderne, i prodotti di dati diventano sempre più importanti per **sfruttare appieno il valore dei dati** e ottenere risultati più velocemente, **rafforzando al contempo la governance dei dati e la fiducia**. Qlik si impegna a fornirti assistenza nel percorso per costruire la tua base di dati con funzionalità self-service. Per ottenere questo risultato, aiutiamo i tuoi team di dati a costruire e gestire prodotti di dati e garantiamo ai fruitori la possibilità di cercare, scoprire e utilizzare facilmente i prodotti di dati, per guidare l'azienda verso il successo.

Un'ultima considerazione: i prodotti di dati sono strumenti potenti, ma non sono la soluzione a tutti i problemi. Il vero potenziale dei prodotti dei dati risiede nell'identificare le esigenze dell'azienda, avere una strategia di dati federativa ben definita, assicurare una governance efficace e collaborare tra team.

Vuoi sfruttare appieno il potenziale dei tuoi dati?

Scopri di più

“ Le organizzazioni cercano sempre più di creare prodotti di dati specifici che trascendano l'integrazione e la qualità dei dati. Qlik si concentra su qualità e governance dei dati per infondere nel concetto di prodotti di dati una maturità che risponda alle attuali esigenze del cliente. ”

Stewart Bond

VP, Data Intelligence and Integration Software, IDC



Informazioni su Qlik

Qlik converte ambienti di dati complessi in informazioni fruibili, favorendo insight strategici per il business. Con oltre 40.000 clienti nel mondo, il suo portafoglio si avvale di AI/ML avanzati di livello enterprise che permettono di migliorare la data quality. Leader nell'integrazione e nella governance dei dati, Qlik offre soluzioni complete che funzionano con fonti di dati differenti. Le analisi intuitive e in tempo reale di Qlik scoprono modelli nascosti, consentendo ai team di affrontare sfide complesse e cogliere nuove opportunità. Gli strumenti IA/ML, pratici e scalabili, consentono di prendere decisioni migliori e più tempestive. Grazie ad una tecnologia avanzata e forti competenze del settore, indipendenti dalla piattaforma, Qlik è un partner strategico che rende i propri clienti più competitivi.

qlik.com