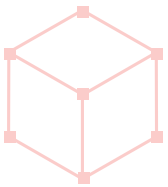
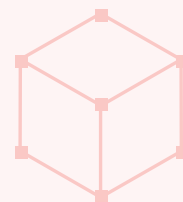
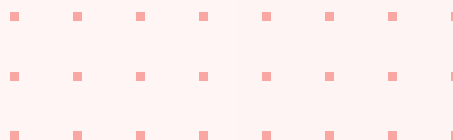


L'integrazione dei dati è un'avventura semantica e non una sfida tecnologica





Sommario

	3
I Premessa	3
I Introduzione	5
I La semantica guida. La tecnologia supporta	6
I Verso un'integrazione dei dati guidata dal significato: un cambiamento di paradigma per le organizzazioni	9
I Semantica sì, ma anche indipendenza e separazione	11
I Dalla semantica all'operatività, per dare vita all'integrazione	12
I Il ruolo delle architetture logiche	14
I Il valore aggiunto della Denodo Platform	15
I Conclusioni	16
I Allegati	17



Integrazione dei dati non significa solo unire sistemi, ma creare comprensione. Senza una semantica condivisa, i dati restano frammenti scollegati. È il linguaggio, chiaro, coerente e accessibile, ad abilitare l'interoperabilità, perché prima della tecnologia, serve l'intesa.

Premessa

Nel contesto attuale, caratterizzato da una crescente complessità dei sistemi informativi e dalla moltiplicazione delle fonti dati, l'integrazione efficace delle informazioni rappresenta una sfida cruciale per le organizzazioni.

Tuttavia, troppo spesso l'approccio all'integrazione si limita ad operare su di un piano tecnico, trascurando un aspetto fondamentale, ovvero la dimensione semantica. Affinché i dati possano essere realmente utili, devono innanzitutto rappresentare una concettualizzazione corretta e coerente del mondo osservato. Ciò significa che la loro definizione, classificazione e modellazione devono riflettere fedelmente i concetti chiave del dominio di riferimento, cogliendone gli elementi essenziali. In assenza di questa aderenza semantica, ogni sforzo di integrazione rischia di produrre informazioni distorte o incoerenti.

È in questa prospettiva che le architetture logiche per l'integrazione dei dati – basate su modelli concettuali condivisi e sulla separazione tra livello semantico e livello fisico – si configurano come la scelta più efficace per garantire una visione integrata, affidabile e comprensibile dei dati. Esse permettono di superare le frammentazioni e le ambiguità, offrendo un'infrastruttura flessibile e scalabile che sostiene una governance dei dati basata sul significato, e non solo sulla sintassi.

Sono questi i temi che saranno affrontati e approfonditi nel presente documento, con l'intento di fornire quegli elementi che consentano di affrontare con successo le sfide insite in un mondo nel quale i dati sono sempre più elemento imprescindibile di comprensione, analisi e decisione.



EXECUTIVE SUMMARY

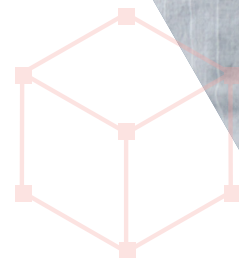
Per troppo tempo, l'integrazione dei dati è stata trattata come una questione esclusivamente tecnica, come insieme di procedure, strumenti e linguaggi volti a collegare tra loro sistemi eterogenei, allineare formati, trasformare strutture. Ma oggi, in un contesto in cui le organizzazioni vivono e si trasformano attraverso i dati, questo approccio mostra tutti i suoi limiti e un dato integrato ma non compreso, accessibile ma non interpretabile, rischia di alimentare rumore invece che conoscenza.

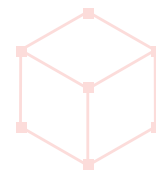
È tempo di cambiare prospettiva e l'obiettivo di questo documento è quello di proporre un nuovo modo di concepire l'integrazione dei dati, partendo dalla semantica, ovvero dal significato che i dati intendono rappresentare, con l'intento, non solo di "mettere insieme" le informazioni, ma di costruire una comprensione condivisa del contesto, dei concetti, delle relazioni che compongono la realtà osservata. L'integrazione, in questa visione, diventa un processo concettuale prima che tecnologico, fondato sull'integrare ciò che si comprende; collegare ciò che si sa descrivere; rendere accessibile ciò che è stato prima pensato.

Il cuore di questa nuova visione è il modello semantico, che prende forma all'interno del livello logico delle architetture informative ed è dove i concetti fondamentali del dominio aziendale vengono formalizzati, strutturati, messi in relazione. Non si tratta di un semplice glossario tecnico, ma di una vera e propria mappa del significato, che permette di connettere in modo intelligente i dati distribuiti nei diversi sistemi aziendali. Questo modello ha la caratteristica essenziale di isolare il significato dalle caratteristiche tecniche dei dati, garantendo che ogni cambiamento infrastrutturale – nel formato, nella fonte, nella struttura – non si ripercuota su chi consuma i dati.

Grazie al livello logico, i Data Consumer, siano essi analisti, decisori o, knowledge worker, possono accedere alle informazioni tramite concetti chiari e stabili, senza dover conoscere le complessità tecniche sottostanti. In questo modo, si protegge l'esperienza dell'utente, si preserva la coerenza informativa nel tempo, e si abbatte il debito cognitivo legato alla gestione dei dati. Per abilitare questa visione, le architetture logiche rappresentano la scelta tecnologica più efficace, consentendo di separare i livelli, federare le fonti, mantenere allineato il modello semantico con l'evoluzione del contesto. Strumenti come la Denodo Platform rendono possibile questo nuovo modo di intendere l'integrazione dei dati, grazie alla virtualizzazione dei dati, alla creazione di viste orientate al business e a un catalogo attivo, ricco e facilmente interrogabile.

In definitiva, questo documento racconta una trasformazione profonda, che vede l'integrazione non più come un'operazione tecnica, ma come atto di comprensione e di collaborazione. È un processo culturale prima che ingegneristico, perché solo quando il significato è ben costruito, ben condiviso e ben protetto, i dati possono davvero generare valore.





Introduzione

A volte, o forse spesso, la fascinazione per la tecnologia ci porta a focalizzare l'attenzione sul dettaglio piuttosto che sull'essenza, con il rischio di vedere la tecnologia non come il mezzo, ma come il fine, cosa ovviamente sbagliata, in tutti i campi e, quindi, anche quando parliamo di integrazione dati.

Per comprendere meglio il fine dobbiamo allora andare al cuore della questione, capire cioè perché dobbiamo integrare i dati, prendendo atto che il nostro obiettivo è la rappresentazione in digitale di qualcosa che accade nel mondo fisico, mondo che dobbiamo osservare per comprenderlo, perché senza comprensione ogni decisione che sarà presa al riguardo sarà intrinsecamente aleatoria.

Comprensione e rappresentazione, quindi e anche come leva per la condivisione, perché ogni decisione è sempre un'azione corale, dove la visione privata di ognuno deve armonizzarsi con quella aziendale e dove entrambe hanno pari dignità, dato che in ogni azienda ci sono obiettivi e responsabilità diverse che, pur operando per un fine comune devono necessariamente tener conto dei differenti punti di vista che caratterizzano i singoli ruoli organizzativi.

Se accettiamo questo punto di vista, allora, benché la componente tecnologica sia importante – se non riesco a connettermi ai dati ben difficilmente potrò pensare a cose come rappresentazione e significato – vien da sé che lo sforzo maggiore sarà quello di creare una rappresentazione universalmente condivisa, in grado di dare concretezza alla concettualizzazione di ciò che deve essere osservato e che, inoltre, sia viva nel senso di legare la componente intensionale a quella estensionale, perché una rappresentazione ha senso solo se renda facile passare dal significato dei dati alle loro occorrenze e, viceversa, ricondurre ciò che viene rilevato e misurato a un modello che ne dia una coerenza semantica complessiva.

Il significato prima di ogni altra cosa, quindi, perché è lì che si gioca tutto e perché senza di esso ogni integrazione sarà priva di senso, rimanendo un esercizio fine a sé stesso o, peggio, portando a una rappresentazione – e alle conseguenti decisioni – che potrebbero condurre in una direzione diversa da quella che ci si aspetterebbe.

Non bisogna però fare l'errore di sottovalutare la complessità del compito, di pensare al significato come a un qualcosa di scontato, assumendo che la visione privata di chi ha questo arduo compito equivalga a quella collettiva, più sfaccettata, articolata e fluida, confidando che quello che normalmente viene indicato come Semantic by Design sia qualcosa che, una volta posto in essere, rimanga immutabile nel tempo, quando invece deve essere qualcosa di profondamento adattabile a un mondo che, per sua stessa natura, è mutevole.

Un modello semantico è infatti una cosa viva, in grado di adattarsi al contesto – e non potrebbe essere altrimenti – di coglierne i diversi punti di vista, armonizzandoli, così da costituire non solo una rappresentazione coerente e accurata, ma soprattutto di essere la base per la condivisione e la comunicazione tra soggetti diversi, che proprio grazie a un modello comune e condiviso potranno dare valore a ciò che devono fare, certi che la profonda tessitura propria di un modello di rappresentazione consenta di connettere queste visioni individuali con quelle complessive.

Quando operiamo una concettualizzazione di qualcosa con la quale interagiamo, inoltre, agiamo secondo due direzioni diverse: per ciò che ci è familiare, definiamo i concetti e li mappiamo con le osservazioni; per ciò che non conosciamo, ci lasciamo guidare dalle osservazioni provando a definirne i concetti che ne danno una visione coerente. Entrambi i modi hanno senso e pari dignità: il primo si fonda sull'esperienza (Semantic by Design); il secondo sulla sperimentazione (Semantic by Experiment) e la scoperta e la storia ci insegna che il progresso è sempre una combinazione dei due, cosa che implica che la costruzione di un modello di rappresentazione debba necessariamente consentire i due approcci.

Che si proceda per esperienza o per sperimentazione, si dovrà comunque affrontare il tema – o il problema – di chi sia il giudice ultimo, ammesso che possa esistere, della rappresentazione che si vuole costruire. È esperienza comune, infatti, che il modo in cui ciascuno di noi osserva e concettualizza il mondo in cui vive si fonda su due elementi fondamentali: (1) l'essere parte di una comunità che condivide quegli elementi che ne consentono la mutua interazione e (2) avere ciascuno di noi una propria individualità, che ci porta a vedere e interpretare ciò che osserviamo secondo quella componente privata legata all'esperienza che nel tempo accumuliamo (a tale proposito è di fondamentale importanza quando ci racconta George Lakoff nella sua celebre opera "Women, Fire and Dangerous Things: What Categories Reveal About the Mind", ma anche l'affermazione di Ludwig Wittgenstein che, nelle sue "Ricerche filosofiche", ci ricorda che "se un leone potesse parlare, noi non lo capiremmo").



Potremmo quindi dire che, se la prima componente dà la direzione, la seconda ne traccia il percorso e l'una non vive senza l'altra, perché la speranza di una visione comune a tal punto che non ci sia mai difformità di vedute è, appunto, una speranza, purtroppo vana.

Naturalmente, se è pur vero che i confini di un'azienda sono tali da ridurre l'ampiezza del mondo osservato, è anche vero che questa riduzione non sarà mai tale da appiattire e conformare le diverse visioni private a un modello universale, suscettibile di un'accettazione da parte di tutti, senza sé e senza ma. All'interno di un'azienda, infatti, ci saranno sempre dei micromondi, a volte identificabili con diversi dipartimenti, altre volte da esigenze contingenti, che porteranno a interpretare le cose in modo differenti, perché come ci ha insegnato Wittgenstein, "il significato è l'uso" e l'uso è sempre "qui e adesso".

In sostanza, per liberare tutto il potenziale dei dati dobbiamo integrarli, riconoscendo però che tale operazione, seppur abbia una dimensione tecnologica, è in realtà una vera e propria avventura semantica, dove il significato di dati diversi, creati da soggetti diversi, in epoche diverse, deve essere ricondotto a un modello unico che li rappresenti, armonizzando i diversi "punti di vista" e connettendo gli uni agli altri, legando i dati alle loro osservazioni che, nel tempo, sono state fatte, perché, se il significato è l'uso e l'uso è sempre qui e adesso, è anche vero che l'uso senza rappresentazione e comprensione è qualcosa di effimero, di fatuo, che non porta a nulla, soprattutto quando l'uso è finalizzato a guidare le decisioni che devono essere prese.

La semantica guida. La tecnologia supporta

Come già detto nell'introduzione, integriamo i dati affinché questi siano poi usati per comprendere e analizzare i fenomeni che accadono, per prendere decisioni su come gestirli, per definire le azioni future e adattare la strategia dell'organizzazione a un mondo fluido e mutevole. In altre parole, integriamo e usiamo i dati per estrarre da essi il massimo valore possibile, un valore che non è una componente strutturale dei dati, ma dipende da come questi vengono utilizzati.

Questa diversa lettura del ruolo dei dati sposta inevitabilmente il focus su coloro che devono utilizzare i dati e, quindi, su aspetti quali, ad esempio:

- la chiarezza rappresentativa, nel senso di adottare un formalismo semplice e auto-esplicativo, così che a chi deve usare i dati – il Data Consumer – non siano richieste attività formative al solo fine di poter usare i dati;
- la facilità di trovare i dati dei quali si ha bisogno, adottando paradigmi vicini a come l'uomo comunica, così che si possano cercare i dati secondo meccanismi cognitivi simili – se non identici – a come normalmente comunichiamo e agiamo;
- la qualità della rappresentazione, nel senso di quanto bene un dato rappresenti ciò che deve rappresentare, elemento più importante della qualità della misura, ovvero di quanto i dati rilevati lo siano con la giusta precisione e accuratezza.

Queste considerazioni implicano che bisogna adottare un approccio diverso, che si allontani dal vedere l'integrazione dei dati secondo una prospettiva puramente tecnica, che si focalizza su strumenti, formati, protocolli e architetture. Una visione tecnica utile, certo, ma profondamente limitata.

La vera integrazione non può prescindere da una riflessione più profonda, che consideri i dati secondo quanto detto più sopra e per fare ciò è imprescindibile partire da una base semantica condivisa, in grado di sostenere l'intero impianto informativo.



LA SEMANTICA COME FONDAMENTO DELL'INTEGRAZIONE

La semantica è ciò che dà significato ai dati e integrare informazioni prive di una coerenza semantica equivale a mettere insieme elementi che non parlano la stessa lingua, motivo per cui, prima ancora di essere memorizzati o trasportati, i dati devono essere concepiti come rappresentazioni fedeli del mondo che intendono descrivere.

È un processo che richiede consapevolezza e intenzionalità, perché ciò che viene rappresentato deve riflettere non solo ciò che è misurabile, ma anche ciò che ha senso nel contesto dell'organizzazione.

In questo scenario, il significato non è mai solo oggettivo, legato ai numeri o alle unità di misura. È anche soggettivo, perché i dati vivono in un contesto di interpretazione, intenzione e uso e un modello semantico che sia in grado di ospitare e valorizzare entrambe le dimensioni è la condizione necessaria per una vera integrazione.

IL MODELLO SEMANTICO COME ARCHITRAVE

Il modello semantico non è più, quindi, un semplice glossario, né una lista di definizioni tecniche. È un sistema strutturato che modella le entità, le relazioni e le regole che governano il dominio di interesse, diventando il linguaggio condiviso attraverso il quale sistemi e persone possono comunicare.

In un'organizzazione, funge da strato fondamentale di astrazione: separa il significato dal dato fisico, mantenendone però il legame, proteggendo così chi usa i dati dai cambiamenti tecnologici sottostanti, cambiamenti che, qualora accadessero, sarebbero totalmente gestiti da questo livello di disaccoppiamento, senza riverberarsi sui Data Consumer ed è proprio questa capacità che rende il modello semantico così centrale, perché in un ecosistema tecnologico in continua evoluzione, dove sistemi, database e strumenti cambiano rapidamente, il significato dei dati deve restare stabile.

Se questa stabilità non viene garantita, ogni cambiamento tecnico rischia di propagarsi verso i Data Consumer, generando discontinuità, inefficienze e perdita di fiducia.



IL RISCHIO DI UN APPROCCIO PURAMENTE TECNICO

Quando l'integrazione è affrontata senza un impianto semantico o, comunque, delegando questo a una sorta di elemento accessorio, emergono problematiche significative:

- diverse aree aziendali possono attribuire significati differenti agli stessi termini, creando incomprensioni e conflitti;
- dati apparentemente simili proliferano in più sistemi, ma con variazioni sottili e spesso non documentate, che rendono difficile distinguerli o aggregarli;
- la dipendenza dai dettagli tecnici diventa insostenibile e ogni modifica richiede interventi lungo tutta la catena del dato, con costi crescenti e perdita di efficienza;
- in assenza di un linguaggio condiviso, i dati faticano a essere riutilizzati, perché mancano di contesto e di una semantica accessibile e condivisa.

È necessario allora, in una sorta di simmetria funzionale, partire da quello che normalmente, secondo l'approccio che sovente viene seguito, è il punto di finale dell'integrazione e cioè l'uso dei dati integrati e modellati, per arrivare da dove normalmente si è tentati di cominciare, cioè le caratteristiche tecniche dei dati che sono stati individuati come necessari alla concettualizzazione della realtà con la quale un'organizzazione si confronta.

LA CONCETTUALIZZAZIONE COME PRIMO PASSO

Per evitare questi rischi, è fondamentale che ogni iniziativa di integrazione cominci con un'analisi accurata del contesto. È necessario chiedersi quale sia il dominio da rappresentare, quali concetti lo popolano, quali relazioni li collegano, e come questi concetti si riflettano nei processi, nelle attività e negli obiettivi dell'organizzazione. Solo una volta costruita questa base concettuale, è possibile procedere alla definizione dei dati che concretizzeranno quella visione.

I dati, in questa prospettiva, non sono creati in modo arbitrario o casuale. Sono l'espressione operativa di una visione condivisa, derivata da un lavoro di modellazione concettuale che garantisce coerenza, validità e significato.

IL RUOLO DEI METADATI

Una volta definiti i dati in coerenza con il modello semantico, è altrettanto importante arricchirli con metadati di business che ne guidino l'interpretazione e l'utilizzo.

Questi metadati, diversi dai metadati tecnici, che normalmente sono generati automaticamente dai sistemi che possiedono i dati oggetto di integrazione, includono definizioni precise, criteri di calcolo, classificazioni, riferimenti all'origine del dato, informazioni sul ciclo di vita e sui responsabili, oltre a indicazioni sulla qualità e sulle modalità di validazione.

È proprio grazie a questa documentazione, il dato diventa trasparente, accessibile, comprensibile, anche da parte di utenti che non lo hanno prodotto direttamente. Si costruisce così un ecosistema informativo solido, collaborativo e democratico, in cui le informazioni sono davvero patrimonio condiviso e, complessivamente, danno vita a una vera e propria Data Economy.

RIEPILOGANDO

Integrare i dati significa prima di tutto comprenderli e la comprensione nasce da un linguaggio chiaro, esplicito e condiviso. Un'integrazione moderna e sostenibile deve partire da un modello semantico che dia senso e coerenza alle informazioni, disaccoppi le tecnologie dai significati e protegga l'intero ecosistema informativo dai cambiamenti non governati.

Solo partendo da una concettualizzazione profonda e consapevole si possono costruire dati che abbiano valore reale per l'organizzazione. E solo con una semantica ben definita e sostenuta da metadati adeguati si può garantire che i dati siano cercabili, interpretabili e utilizzabili da tutti.

L'integrazione non è solo un esercizio tecnico. È un atto di cultura organizzativa, di progettazione condivisa, e di responsabilità nel costruire un linguaggio comune che trasformi i dati in conoscenza.

Verso un'integrazione dei dati guidata dal significato: un cambiamento di paradigma per le organizzazioni

Come più volte rimarcato, quando si parla di integrazione dei dati, benché partire dagli aspetti tecnici sia una prospettiva comprensibile ma parziale, le organizzazioni che desiderano governare in modo efficace i propri dati dovrebbero compiere un vero e proprio cambio di paradigma e mettere al centro il significato, non l'infrastruttura, il che implica partire da una domanda essenziale ma spesso trascurata: che cosa rappresentano davvero questi dati?

Questo cambio di prospettiva, da un lato facile a pronunciarsi, dall'altro ha però un impatto profondo sulle tre dimensioni fondamentali: organizzazione, cultura e tecnologia.

IMPATTO ORGANIZZATIVO – IL RUOLO DELLE PERSONE CAMBIA

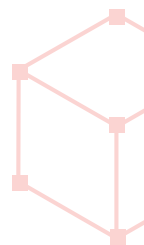
In un approccio semantico all'integrazione, il punto di partenza non sono i dati, ma le persone e costruire un modello semantico significa prima di tutto costruire una rappresentazione condivisa del contesto, definendo i concetti chiave e le relazioni fondamentali e questo non può essere frutto esclusivo di un gruppo tecnico o IT, ma richiede il coinvolgimento attivo di chi conosce il dominio, i processi, le ambiguità linguistiche, le eccezioni operative.

Ne deriva un profondo ripensamento dei ruoli e gli esperti di dominio, gli analisti di business e i knowledge worker diventano co-protagonisti nella progettazione del significato, non semplici fruitori finali. Il loro contributo non si limita più a "usare i dati", ma diventa essenziale per definirli in modo corretto.

IMPATTO CULTURALE – DALLA TECNOLOGIA AL SENSO

Il passaggio da un'integrazione basata sulla tecnologia a una basata sul significato comporta una trasformazione culturale. La cultura organizzativa deve evolversi verso una maggiore consapevolezza semantica: ciò che importa non è solo "avere" i dati, ma sapere cosa rappresentano, come sono stati costruiti, perché sono rilevanti. Questo significa sviluppare un nuovo tipo di alfabetizzazione, che non è solo Data Literacy in senso tecnico, quanto piuttosto Semantic Literacy, ovvero la capacità di leggere il contesto dietro il dato.

Questo tipo di cultura ha un effetto positivo anche sulla collaborazione interfunzionale, perché di fatto costringe a un confronto tra linguaggi diversi, a trovare un terreno comune, a esplicitare convenzioni implicite. Il risultato è una maggiore coerenza interna e una più forte capacità di prendere decisioni informate.



IMPATTO TECNOLOGICO – LA TECNOLOGIA AL SERVIZIO DEL SIGNIFICATO

In questo nuovo scenario, la tecnologia non scompare, ma si riposiziona. Non è più l'elemento trainante, ma l'infrastruttura di supporto e le architetture logiche, basate su modelli semantici condivisi, diventano strumenti per rendere il significato accessibile, navigabile, interrogabile. Il loro scopo è semplificare l'esperienza dell'utente finale, che deve poter cercare, comprendere e utilizzare i dati con naturalezza, senza perdersi nella complessità tecnica.

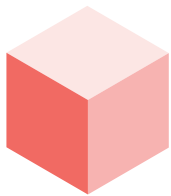
I Data Consumer non devono essere esperti di modellazione semantica, ma devono beneficiare della sua presenza, grazie a un ambiente informativo dove è facile trovare ciò che serve, capirne il senso, fidarsi della sua coerenza. Solo così, infatti, si evita il sovraccarico cognitivo e si preserva la concentrazione sull'obiettivo finale, che è quello di usare i dati per generare valore.

RIEPILOGANDO

L'integrazione dei dati guidata da considerazioni semantiche non è solo un'evoluzione tecnica, ma una trasformazione profonda del modo in cui le organizzazioni pensano, collaborano e agiscono.

Significa mettere il significato al centro, riconoscendo che i dati non sono mai neutri, ma portatori di una visione del mondo ed è solo attraverso la partecipazione consapevole delle persone che questa visione può essere costruita in modo solido, condiviso e utile.

mettere a disposizione dei cosiddetti portatori di interesse, con un evidente rimando al Data Governance Act, Nel prosieguo del documento saranno analizzati, con maggior dettaglio, gli aspetti e le componenti di un'architettura logica – e nello specifico della Denodo Platform – che più di ogni altro sono abilitatori essenziale per cogliere lo spirito del EDA, sia per quanto riguarda gli aspetti regolamentatori che quelli votati a creare una vera e propria economia dei dati, consapevole, governata e portatrice di valore per la comunità.



Semantica sì, ma anche indipendenza e separazione

Quando diciamo che l'integrazione dei dati deve partire dalla semantica, non stiamo semplicemente scegliendo un diverso punto di partenza, ma stiamo costruendo un diverso modo di pensare e organizzare l'intero sistema informativo.

In questo nuovo paradigma, il livello logico – il luogo dove prende vita il modello semantico – diventa non solo uno strumento di progettazione, ma un elemento di mediazione strategica, avendo tale livello: da un lato è ciò che rappresenta il significato, la concettualizzazione condivisa del dominio di interesse; dall'altro è ciò che isola il significato dalle variabili tecniche che caratterizzano i dati nelle loro molteplici forme e collocazioni.

ISOLARE IL SIGNIFICATO PER PROTEGGERE CHI USA I DATI

In un ecosistema digitale in continua evoluzione, le infrastrutture cambiano; si aggiornano i database; si migrano i sistemi; si modificano i formati; si ridefiniscono le API e si sostituiscono sorgenti. Sono interventi naturali, necessari, spesso inevitabili.

Il problema nasce quando ogni cambiamento tecnico si riflette direttamente sull'esperienza dell'utente finale e, in assenza di un livello logico intermedio, ogni minima variazione nella fonte dati impone un adattamento a valle, costringendo i Data Consumer a rincorrere aggiornamenti, decifrare nuove strutture, comprendere nuovamente ciò che prima era chiaro.

Il livello logico, invece, svolge la funzione opposta, smorzando, assorbendo e filtrando. Costruito intorno a un modello semantico, questo livello diventa un'interfaccia tra il mondo mutevole dei sistemi e l'universo stabile della conoscenza. I Data Consumer non accedono più direttamente ai dati grezzi, ma ai concetti che ne sono astrazione e che sono espressi nel linguaggio del business, rendendo i dati comprensibili, navigabili, coerenti. È una forma di astrazione protettiva, dove il significato resta integro anche se la tecnologia cambia.

CONTINUITÀ PER CHI DECIDE, FLESSIBILITÀ PER CHI GESTISCE I DATI

Questo isolamento non è solo una comodità, ma una condizione essenziale per garantire continuità cognitiva all'interno delle organizzazioni.

Le decisioni aziendali si fondano sulla fiducia nei dati, una fiducia che crolla ogni volta che cambia il modo in cui i dati vengono presentati, ogni volta che la semantica si frantuma perché l'infrastruttura si è evoluta.

Il livello logico protegge quella fiducia e permette a chi gestisce i dati – i Data Steward – di operare liberamente sugli strati fisici, ottimizzando le performance, aggiornando le sorgenti e reingegnerizzando i flussi, senza impattare sull'esperienza di chi analizza, interpreta, decide. È un patto implicito tra chi costruisce e chi utilizza e ciò che cambia sotto non deve far rumore sopra.

IL VALORE DELLE ARCHITETTURE LOGICHE

È per questo che le architetture logiche sono la scelta più adatta in un approccio semantico all'integrazione dei dati, perché riconoscono e valorizzano il ruolo del livello logico come cuscinetto semantico tra instabilità tecnica e stabilità concettuale.

Al loro interno, il modello semantico non è un semplice dizionario di campo, ma una struttura dinamica e viva, capace di astrarre, federare, comporre e ricomporre i dati fisici secondo il linguaggio del dominio.

In questo modo, le architetture logiche rendono il sistema informativo resiliente e capace di cambiare in profondità senza perdere la propria identità, senza scaricare la complessità sull'utente finale.

RIEPILOGANDO

Mettere la semantica al centro dell'integrazione non significa solo partire dal significato: significa anche proteggerlo, mantenerlo stabile e accessibile nel tempo. Il livello logico, in questa visione, non è un passaggio tecnico tra due strati, ma un custode del significato, un garante di continuità semantica, un alleato silenzioso dei Data Consumer.

Grazie a esso, l'organizzazione può evolversi tecnicamente senza perdere coerenza informativa, potendo essere agile sotto e affidabile sopra. Può cambiare tutto, senza cambiare ciò che conta davvero, ovvero la capacità di comprendere il mondo attraverso i dati.

Dalla semantica all'operatività, per dare vita all'integrazione

Aver compreso che l'integrazione dei dati deve partire dalla semantica è già un primo, fondamentale passo, che sancisce il superamento dell'idea, tanto diffusa quanto fuorviante, che l'integrazione sia solo un esercizio tecnico, fatto di connettori, formati e protocolli.

Riconoscere il ruolo centrale del significato impone però anche un'assunzione di responsabilità: una volta che questa visione è accettata, bisogna darle forma, bisogna darle vita, e per farlo, serve un modo di lavorare strutturato, riflessivo e coerente.

L'integrazione semantica non è un atto istantaneo, ma un processo articolato, che si sviluppa per gradi. È una costruzione progressiva, che parte da una scelta consapevole e si sviluppa attraverso un modello concettuale fino a raggiungere i dati reali, presenti nei sistemi operativi.

Questo processo può essere descritto in tre passaggi fondamentali, che rappresentano il cuore dell'operatività di un'integrazione guidata dal significato.

SELEZIONARE IL CONTESTO DA RAPPRESENTARE

Ogni progetto di integrazione inizia da una delimitazione consapevole del dominio di interesse, perché non si può né si deve – e probabilmente non avrebbe neanche senso – modellare tutto in un colpo solo. Occorre scegliere quella parte del contesto organizzativo che è rilevante rispetto agli obiettivi di analisi, comprensione o decisione.

Si tratta di una fase esplorativa, quasi di osservazione attiva, durante la quale si guarda al mondo operativo e si decide cosa è necessario concettualizzare, quali fenomeni sono significativi, quali entità vale la pena descrivere e mettere in relazione. Questo primo atto non è solo tecnico né solo analitico, ma è una scelta strategica, in quanto rappresenta l'identificazione del perimetro semantico dentro il quale prenderà forma l'integrazione.

DUE APPROCCI PER DEFINIRE I CONCETTI

Una volta scelto il campo da modellare, si passa alla costruzione concettuale vera e propria ed è qui che il modello semantico prende forma, definendo le entità, le relazioni, le proprietà e le regole che descrivono il contesto selezionato.

Questa costruzione può seguire due approcci distinti, entrambi validi, spesso complementari:

Approccio Top-Down: si parte da concetti generali, astratti, e li si scompone progressivamente in elementi più semplici e operativi. Ad esempio, si può partire dal concetto di "Cliente" e decomporlo in anagrafica, comportamenti, contratti, preferenze. Questo approccio si basa su una visione strutturata e spesso normativa del dominio.

Approccio Bottom-Up: si parte invece da elementi semplici, presenti nel mondo operativo – un identificativo, un evento, un valore – e li si aggrega progressivamente in concetti più complessi, che acquisiscono significato solo nel loro insieme. È un processo più esplorativo, che si fonda sull'osservazione e sull'esperienza.



In realtà, i due approcci convivono e si rafforzano a vicenda, in quanto la costruzione del significato non è mai lineare, ma nasce dall'interazione tra astrazione e osservazione, tra intuizione e dettaglio.

RENDERE IL MODELLO VIVO CONNETTENDO I CONCETTI AI DATI

Una volta costruiti i concetti, questi devono essere connessi ai dati che li rappresentano nei sistemi informativi. È il passaggio che dà vita al modello, che lo rende operativo ed è anche ciò che distingue una pura modellazione teorica da una vera architettura logica.

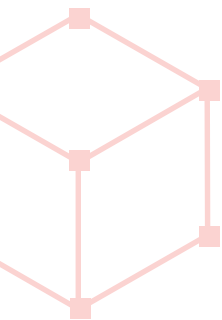
In questa fase, si stabilisce il legame tra i concetti definiti nel modello semantico e le occorrenze presenti nei sistemi aziendali (database, applicativi, fogli di calcolo, servizi esterni, ...). È il momento in cui la semantica si ancora alla realtà operativa, permettendo – come nelle migliori architetture logiche – di passare dal concetto all'istanza, dal significato al dato concreto, e viceversa.

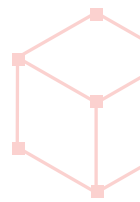
Questa connessione non avviene una volta per tutte, ma è un legame dinamico, aggiornabile, tracciabile, ed è proprio questo che rende il sistema informativo vivo, interrogabile, comprensibile, perché ogni dato può essere interpretato alla luce del concetto che rappresenta, e ogni concetto può essere esplorato attraverso i dati che lo incarnano.

RIEPILOGANDO

Il vero valore dell'integrazione semantica sta nel suo modo di procedere: intenzionale, modulare, partecipato. Si parte da ciò che è rilevante, si costruisce un linguaggio condiviso, e si stabiliscono connessioni concrete con i dati reali. In questo modo, non si costruisce solo un'infrastruttura tecnica, ma una struttura di significato, capace di sostenere analisi, decisioni e trasformazioni.

È in questo spazio tra significato e operatività che prende forma l'integrazione moderna: un processo in cui la conoscenza guida la tecnologia, e non il contrario.





Il ruolo delle architetture logiche

Nel contesto della trasformazione digitale e della crescente complessità informativa, l'integrazione dei dati non può più quindi essere affrontata unicamente come un problema tecnico di connessione tra fonti eterogenee, ma è necessario adottare un approccio che metta al centro il significato dei dati, la loro capacità di rappresentare fedelmente i concetti del mondo reale e del dominio applicativo.

In questo scenario, le architetture logiche svolgono un ruolo cruciale, diventando il ponte tra la comprensione del contesto e l'implementazione tecnica dell'integrazione.

CHE COSA SONO LE ARCHITETTURE LOGICHE PER L'INTEGRAZIONE DATI

Le architetture logiche sono strutture che separano il livello semantico-concettuale da quello fisico-tecnologico. Invece di integrare direttamente i dati provenienti da sistemi eterogenei (attraverso il ben noto paradigma ETL – Extract, Transform and Load), esse si fondano sulla definizione di un modello logico comune che rappresenta i concetti chiave, le relazioni e le regole del dominio.

Questo modello logico funge da interfaccia semantica tra le fonti dati fisiche e i Data Consumer, siano essi applicazioni o persone e i dati non vengono forzati in uno schema rigido unico, ma sono mappati dinamicamente su un livello superiore, che ne esplicita il significato.

PERCHÉ SONO CENTRALI NELL'INTEGRAZIONE SEMANTICA

Il valore delle architetture logiche sta nella loro capacità di disaccoppiare il significato dal supporto fisico, creando una base coerente, estendibile e interrogabile semanticamente. In particolare:

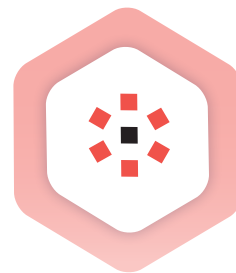
- **preservano la complessità delle fonti**, evitando trasformazioni distruttive o semplificazioni forzate
- **permettono l'evoluzione del modello semantico** nel tempo, senza impattare direttamente le fonti;
- **supportano la federazione dei dati**, dato che le query vengono eseguite dinamicamente su sistemi distribuiti attraverso un'interfaccia semantica comune;
- **favoriscono la trasparenza**, in quanto i Data Consumer possono comprendere e interrogare i dati facendo riferimento a concetti noti (e non a nomi di tabelle, colonne o identificatori tecnici).

BENEFICI ORGANIZZATIVI E OPERATIVI

L'adozione di architetture logiche porta benefici tangibili su vari fronti:

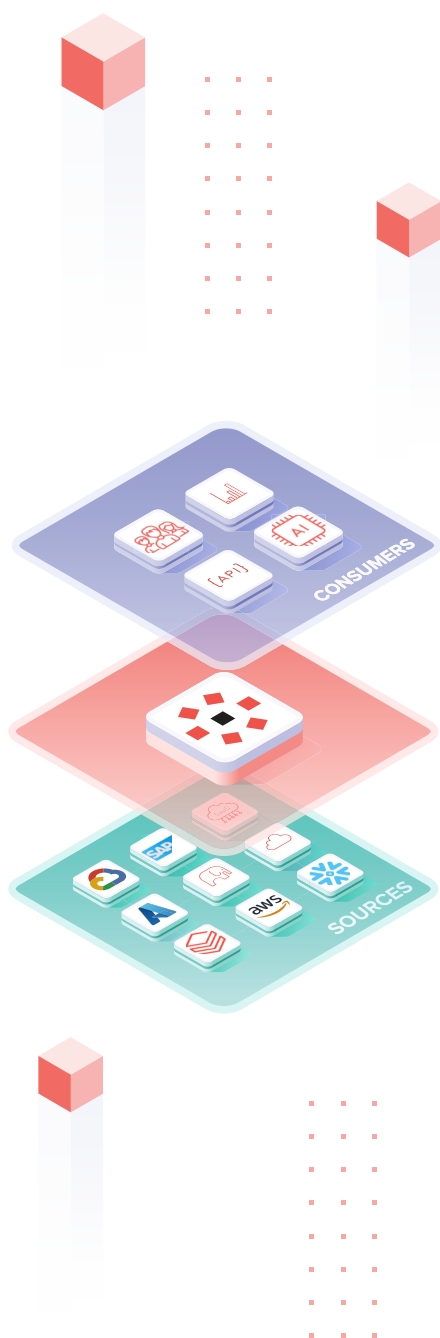
- **governance dei dati**: la centralità del modello semantico facilita il tracciamento del significato, della provenienza e della qualità del dato;
- **agilità operativa**: le fonti possono evolvere indipendentemente, purché le mappature logiche siano aggiornate;
- **riduzione del debito tecnico**: si evitano duplicazioni massive di dati, riducendo i costi di manutenzione e sincronizzazione;
- **empowerment dei Data Consumer**: gli utenti possono accedere ai dati tramite interfacce semantiche (es. motori di ricerca semantici, cataloghi intelligenti) che facilitano la scoperta e l'uso consapevole delle informazioni, senza dover conoscere le strutture fisiche sottostanti.

Il valore aggiunto della Denodo Platform



Tra le soluzioni disponibili per abilitare un'integrazione semantica tramite architetture logiche, Denodo Platform si distingue per completezza, robustezza e maturità tecnologica.

Si tratta di una Data Platform, basata sulla Data Virtualization, che consente di accedere, unificare, modellare e pubblicare dati provenienti da fonti eterogenee senza doverli replicare o spostare fisicamente.



INTEGRAZIONE SEMANTICA NATIVA

Denodo consente la costruzione di un modello logico semantico che mappa i dati delle fonti fisiche a concetti comprensibili e condivisi. I dati vengono esposti agli utenti finali in forma di viste di business, modellate secondo il linguaggio e le esigenze del dominio, riducendo la distanza tra dati grezzi e significato.

VIRTUALIZZAZIONE DEI DATI DISTRIBUITI

Grazie al suo motore di data virtualization avanzato, Denodo permette di interrogare fonti dati eterogenee (relazionali, Big Data, cloud, NoSQL, API, Excel, ecc.) come se fossero un unico repository coerente, senza replicare i dati. Questo approccio migliora l'agilità e riduce i costi infrastrutturali.

CATALOGO SEMANTICO E DATA GOVERNANCE

Denodo offre un catalogo semantico attivo integrato che facilita l'esplorazione e la comprensione dei dati da parte dei Data Consumer. Il catalogo è arricchito con metadati, glossari di business, descrizioni semantiche e lineage. In questo modo, gli utenti possono comprendere il significato, la provenienza e la qualità dei dati in autonomia.

SICUREZZA E CONTROLLO GRANULARE

La piattaforma permette di definire politiche di sicurezza centralizzate e granulari, con controllo puntuale su colonne, righe, funzioni e interi concetti, cosa che garantisce una protezione coerente anche quando le fonti si moltiplicano e l'accesso è distribuito.

USER EXPERIENCE SEMPLIFICATA

Gli utenti finali possono accedere ai dati tramite SQL, REST, GraphQL o strumenti di BI, senza dover conoscere la complessità sottostante.

La Denodo Platform si pone come facilitatore dell'esperienza utente, minimizzando il carico cognitivo sui Data Consumer e permettendo loro di concentrarsi sull'analisi, non sull'accesso o sull'estrazione.

SUPPORTO ALL'EVOLUZIONE E ALLA SCALABILITÀ

Il modello logico e le viste virtuali possono essere modificate, estese e gestire versioni diverse in modo agile, senza necessità di reingegnerizzare fisicamente i sistemi sottostanti, cosa particolarmente utile in contesti dinamici e in progetti di lungo respiro.

VERSO UNA NUOVA VISIONE DELL'INTEGRAZIONE

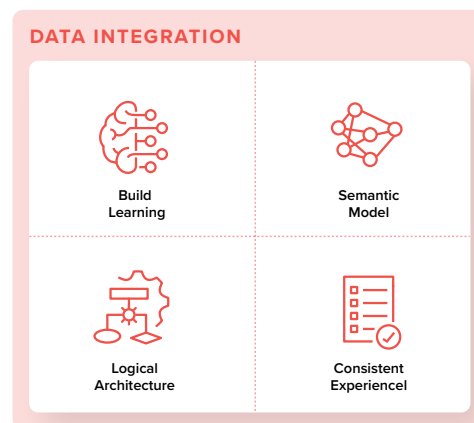
Con le architetture logiche, l'integrazione dei dati non è più un'operazione puntuale o statica, ma un processo continuo, che si adatta all'evoluzione del contesto e delle esigenze informative. Il focus si sposta dal collegare sistemi al connettere significati, abilitando un'organizzazione a diventare veramente data-driven, non nel senso della quantità di dati disponibili, ma nella qualità del pensiero che li sostiene.

RIEPILOGANDO

Le architetture logiche rappresentano il cuore tecnico di una strategia che pone il significato al centro dell'integrazione dei dati.

Offrono un'infrastruttura flessibile, sostenibile e orientata al lungo termine, capace di valorizzare la conoscenza distribuita nell'organizzazione e di semplificare la fruizione dei dati da parte di chi ne ha realmente bisogno.

In questo nuovo paradigma, la tecnologia non è più protagonista, ma abilitatore silenzioso di una comprensione condivisa del mondo



I quattro pilastri dell'integrazione dei dati

Conclusioni

Integrare i dati non significa più – e non deve più significare – collegare tecnologie, replicare tabelle o costruire ponti tra sistemi. Significa, piuttosto, costruire significato, generare comprensione, creare coerenza e questo è un passaggio epocale, che sposta il baricentro dell'integrazione dall'infrastruttura al linguaggio, dalla sintassi alla semantica, dal dato al contesto.

I quattro pilastri dell'integrazione dei dati:

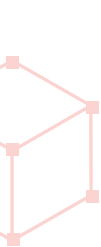
In questo nuovo paradigma, il **modello semantico** diventa la vera architrave dell'ecosistema informativo, non limitandosi più e solo a definire che cosa un dato rappresenta, ma struttura la conoscenza aziendale in modo che sia interrogabile, accessibile, condivisibile.

Il livello logico, dove questo modello prende forma, non è più un semplice strato di astrazione, ma un custode del significato, ma media tra l'instabilità tecnica e la stabilità concettuale, tra ciò che cambia sotto e ciò che deve rimanere comprensibile sopra.

È qui che le **architetture logiche** si rivelano decisive, perché non solo abilitano l'integrazione semantica, ma proteggono l'esperienza del Data Consumer da ogni turbolenza infrastrutturale. Permettono all'organizzazione di evolvere senza perdere coerenza; di innovare senza sacrificare la fiducia; di crescere, sapendo che il significato – ciò che davvero conta – resterà intatto, accessibile e condiviso.

Abbiamo quindi bisogno di una **nuova alleanza, tra chi conosce i processi, chi governa i dati e chi li interpreta**. Un'alleanza fondata non più su connessioni tecniche, ma su comprensioni comuni, perché solo quando il significato diventa patrimonio condiviso, i dati diventano davvero strumenti di trasformazione.

L'integrazione dei dati, allora, non è più una sfida solo tecnica, ma diventa un progetto culturale; un atto collettivo di comprensione; una nuova forma di linguaggio, che mette in dialogo sistemi, persone e decisioni. Proprio per questo, può guidarci verso una conoscenza più profonda, più affidabile, più utile del mondo che vogliamo osservare, comprendere e trasformare.



Allegati

Le parti che seguono sono approfondimenti di quanto descritto nel documento con particolare riferimento alle già menzionate architetture logiche, che saranno approfondite in quelle loro componenti che, più di ogni altre, sono collegate al Data Act, nel senso di rappresentare funzionalità utili a far sì che quanto definito in termini organizzativi e culturali, trovi poi nella tecnologia un valido ed efficiente supporto.

COMPRENDERE LE ARCHITETTURE LOGICHE DEI DATI

A differenza delle architetture fisiche dei dati, principalmente basate sul ben noto paradigma ETL (Extract, Transform and Load) e che si basano su una duplicazione dei dati, trasformandoli e concentrandoli in un unico posto, come ad esempio un Data Warehouse o un Data Lake, le architetture logiche si occupano di come i dati sono collegati e di come possono essere utilizzati in modo efficiente indipendentemente dalla loro posizione fisica di archiviazione.

Le architetture logiche dei dati si riferiscono ai modelli concettuali, che definiscono il significato dei dati e il modo in cui questi sono logicamente organizzati, un'astrazione che garantisce flessibilità e scalabilità. Inoltre, il livello logico, all'interno del quale prende vita il modello semantico, funge anche da isolante rispetto alla tecnologia sottostante, i cui eventuali cambiamenti saranno assorbiti dal tale livello, consentendo di implementare un'architettura estremamente resiliente.

COMPONENTI DELLE ARCHITETTURE LOGICHE DEI DATI



Integrazione: il processo di combinazione di dati provenienti da fonti diverse per fornire una visione unificata.



Astrazione: creare un livello che nasconda le caratteristiche tecniche dei dati, concentrandosi invece sul modo in cui i dati vengono rappresentati, compresi e utilizzati.



Governance: stabilire politiche e procedure per gestire la qualità, la sicurezza e la conformità dei dati.



Modellazione: progettare modelli semantici che rappresentino le entità dei dati, le relazioni e i vincoli, in modo da creare la miglior concettualizzazione possibile del contesto di riferimento dell'amministrazione e, di conseguenza, progettare, erogare e misurare i servizi forniti.



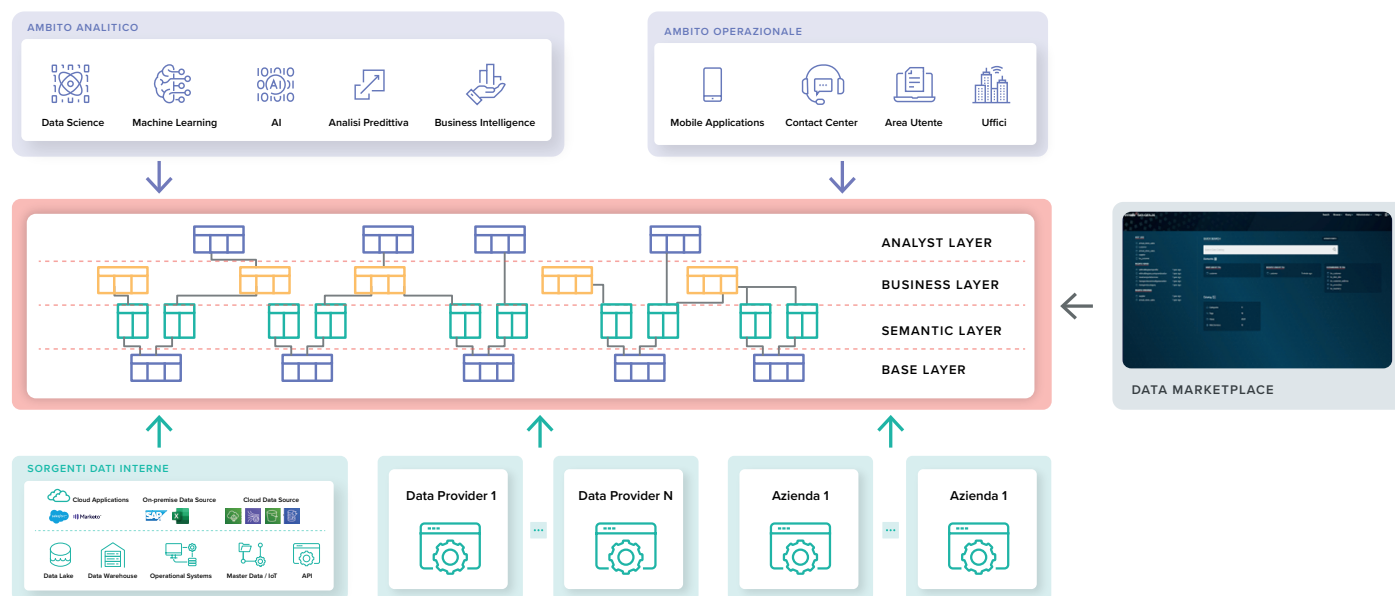
Accesso, consultazione e consumo: rendere effettiva la democratizzazione dei dati attraverso un Data Catalog o Data Marketplace, specificatamente rivolto ai Data Consumer, tipicamente competenti su ciò che i dati rappresentano e non sulle loro caratteristiche tecniche, in modo che questi possano facilmente esplorare il patrimonio informativo disponibile, comprenderlo e usarlo, senza doversi preoccupare degli aspetti tecnici dell'integrazione e della consegna di tali dati.

I modelli logici o semantici consentono quindi di dare concretezza alla concettualizzazione del contesto di riferimento, modellando i dati di interesse con un unico formalismo di rappresentazione, indipendentemente dalle caratteristiche tecniche dei dati, dalla loro distribuzione in rete e dal soggetto che li detiene e ne è responsabile, in modo da poter sfruttare dati di terze parti, come altre amministrazioni, aziende o data provider, senza dover duplicare, copiandoli, i dati da loro gestiti, con evidenti benefici sulla razionalizzazione complessiva e sulla loro governance.

Allo stesso tempo, il livello logico rappresenta un elemento di disaccoppiamento tra chi utilizza i dati – il Data Consumer – e l'infrastruttura tecnica necessaria per gestire tali dati. Questo disaccoppiamento diventa essenziale per gestire le modifiche a quest'ultima, poiché ogni cambiamento in questo senso sarà gestito dal livello logico, senza alcun riverbero sugli utenti, che continueranno a consumare i dati attraverso il modello semantico.



Graficamente, questo duplice ruolo del livello logico – rappresentazione del significato e disaccoppiamento della tecnologia – è rappresentato dalla figura seguente:



Panoramica dell'architettura di alto livello per l'integrazione logica dei dati

COME LE ARCHITETTURE LOGICHE DEI DATI SEMPLIFICANO LA GOVERNANCE DEI DATI

Le LDA migliorano significativamente la governance dei dati, consentendo la definizione di regole di sicurezza, visibilità e mascheramento dei dati esclusivamente a livello di astrazione logica. Questo approccio comporta notevoli vantaggi, in particolare la possibilità di centralizzare la gestione delle regole in un unico punto di controllo logico. In questo modo, le organizzazioni possono implementare sofisticate politiche di governance dei dati che si adattano dinamicamente ai ruoli e ai profili degli utenti, garantendo un accesso coerente e appropriato ai dati in tutta l'organizzazione.

Un vantaggio fondamentale della gestione logica della governance è la completa indipendenza dalle specifiche fisiche e tecniche delle fonti di dati sottostanti. Di conseguenza, se un'organizzazione intraprende cambiamenti tecnici, come la migrazione delle fonti di dati da ambienti on-premise a piattaforme cloud, le regole di governance stabilite rimangono inalterate. Poiché queste regole sono disaccoppiate dal livello fisico di archiviazione e sono definite in modo logico, si applicano immediatamente alla nuova fonte di dati senza richiedere modifiche o un'ampia convalida.

Le LDA offrono quindi soluzioni di governance robuste, flessibili e a prova di futuro, semplificando in modo significativo la gestione dei dati e migliorando l'agilità organizzativa e la conformità.

LA DEMOCRATIZZAZIONE DEI DATI COME ELEMENTO CHIAVE PER LA DATA DRIVEN TRANSFORMATION

Per diventare ed essere realmente data-driven, un'azienda deve necessariamente abbracciare la democratizzazione dei dati, il principio secondo cui i dati devono essere accessibili, comprensibili e utilizzabili da tutti i membri dell'organizzazione. Questo concetto va oltre la semplice concessione dell'accesso, affermando che l'utilizzo dei dati è un diritto, non un privilegio. Se l'uso dei dati è complesso o limitato, questi diritti rischiano di diventare dominio di pochi eletti, minando le fondamenta stesse di una cultura data-driven.

La democratizzazione implica che tutti – non solo gli analisti o il management – devono essere messi in grado di utilizzare i dati per l'uso che intendono farne. Questa accessibilità deve essere accompagnata dalla semplicità: strumenti intuitivi, governance chiara e supporto all'uso, che riducano le barriere all'ingresso e incoraggino un utilizzo dei dati diffuso.

Tuttavia, ai diritti devono corrispondere le responsabilità e deve essere altrettanto facile per gli individui adempiere ai propri doveri, senza che questi diventino un ostacolo all'uso dei dati, ma ne siano solamente strumento per un uso responsabile. Ma soprattutto, un'azienda veramente data-driven è una comunità, dove i dati non vengono solo consumati ma anche condivisi, arricchendo l'intelligenza collettiva dell'organizzazione.

La circolazione dei dati – la libertà che questi possano fluire all'interno dell'organizzazione – è essenziale per sbloccare il loro potenziale e quanto intuito o scoperto in un'area organizzativa può e deve alimentare o suggerire opportunità in altre aree e ciò tendenzialmente accade solo se i dati circolano in modo libero e responsabile.

In questo contesto, la tecnologia svolge un ruolo abilitante fondamentale e un'architettura logica dei dati ben progettata consente di:

- Avere un accesso facile e sicuro ai dati rilevanti;
- Definire e applicare chiaramente e semplicemente diritti, doveri e responsabilità;
- Garantire un supporto scalabile per la collaborazione e la condivisione dei dati.

La democratizzazione dei dati non è solo un obiettivo tecnico o procedurale. È una pietra miliare culturale per le aziende che cercano di trasformare i dati in un patrimonio condiviso e in un motore di progresso collettivo.

L'IMPORTANZA DI UN MODELLO SEMANTICO

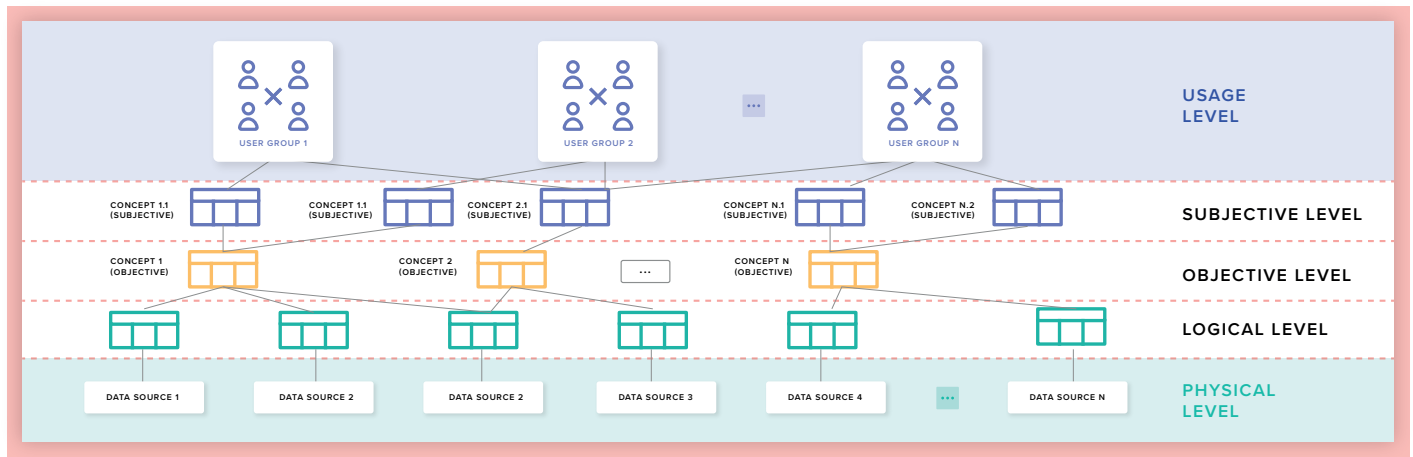
Il livello logico, che come già detto ospita i modelli semantici, è uno degli elementi più importanti, se non il più importante in assoluto, di una architettura logica, in quanto rappresenta il significato dei dati, nei termini di cosa rappresentano, di come lo fanno e di quali sono le relazioni reciproche.

La possibilità di realizzare tali modelli è fondamentale per rendere agevole a chi deve usare i dati – i Data Consumer – la loro ricerca e comprensione, ovviamente nel pieno rispetto delle regole che ne governano visibilità e uso.

Un modello semantico, inoltre, funge da collante semantico tra dati che appartengono a soggetti differenti, in quanto diventa il punto di armonizzazione di dati che, pur rappresentando lo stesso concetto, potrebbero farlo in modo differente, in virtù della diversa interpretazione che ne dà il soggetto che ne ha curato la genesi, interpretazione che dipende dal ruolo e dalle responsabilità del soggetto stesso.

Considerando che oggi praticamente ogni settore di mercato si trova a gestire processi caratterizzati da una forte condivisione di dati, posseduti e gestiti da soggetti diversi, con ruoli e responsabilità differenti, la possibilità – o meglio, la necessità – di avere questo punto di armonizzazione diventa fondamentale, poiché è lì che si sprigiona il potenziale insito nei dati, che può essere espresso solo all'interno di una rappresentazione che, da un lato, sia coerente, completa e univoca e, dall'altro, sia comunque in grado di rispettare quella soggettività rappresentativa necessaria a un'organizzazione per poter efficacemente gestire il proprio modello di business.

Volendo dare una rappresentazione di quanto detto, la figura seguente schematizza un modello semantico che rappresenta sia la componente oggettiva che quella soggettiva



Una esemplificazione del modello semantico

La figura identifica diversi livelli, secondo un percorso di astrazione crescente:

- il **livello fisico** rappresenta le sorgenti dati oggetto di integrazione;
- il **livello logico** è dove si effettua la separazione logico-fisica delle sorgenti dati, elemento fondante delle architetture logiche;
- il **livello oggettivo** è dove vengono modellati quei concetti in modo tale che la loro rappresentazione sia considerata oggettiva all'interno dell'organizzazione, cosa che normalmente si traduce nel definire un concetto caratterizzandolo solo con quegli attributi che hanno un significato condiviso tra tutte le strutture dell'organizzazione;
- il **livello soggettivo** rappresenta la derivazione, a partire dal livello oggettivo, dei dati che catturano lo specifico punto di vista dei diversi Data Consumer, punto di vista che dipenderà dal loro particolare ruolo e dall'uso che ne devono fare in relazione all'obiettivo che stanno perseguendo;
- il **livello dell'uso**, infine, rappresenta i diversi Data Consumer che usano i dati, tipicamente usando quelli definiti secondo il loro punto di vista, che ragionevolmente meglio si adattano a ciò che devono fare.

Naturalmente, quelli che nella figura sono descritti genericamente come concetti, assumeranno una specifica caratterizzazione quando il modello sarà realizzato all'interno di una data organizzazione e con fini ben determinati, potendo ad esempio rappresentare il cliente, il cittadino, il punto vendita, il prodotto o l'impianto di produzione. Allo stesso modo, i gruppi di utenti potranno corrispondere, ad esempio, alle Vendite, al Customer Care, al Marketing, all'Amministrazione o alla Progettazione.

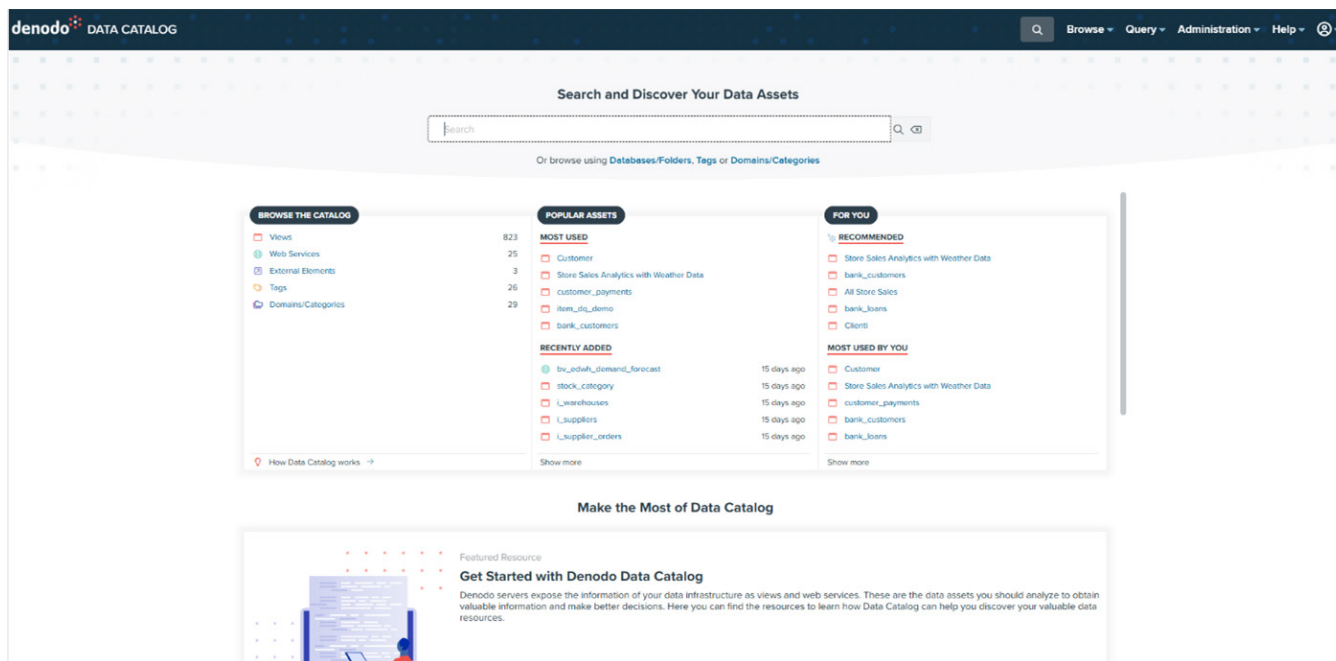
Va anche sottolineato che, sempre per non rendere troppo complessa la figura, si è deciso di evidenziare, da parte dei diversi utenti, solamente l'uso dei concetti del livello soggettivo, cosa che ovviamente non è un vincolo, dato che in situazioni reali ogni utente potrà usare uno qualsiasi dei concetti presenti nel modello semantico, indipendentemente dal livello nel quale è collocato.

L'IMPORTANZA DI UN CATALOGO DATI

In un mondo sempre più orientato e fondato sui dati, è essenziale disporre di un catalogo di dati specificamente concepito per i Data Consumer. Tale catalogo semplifica il processo di esplorazione, comprensione e utilizzo efficace dei dati, astraendo dalla complessità tecnica sottostante. Offrendo un'interfaccia intuitiva e facile da usare che non richiede competenze specifiche, il catalogo dei dati elimina le barriere tradizionali che spesso ostacolano il processo di democratizzazione dei dati.

Questa facilità d'uso è fondamentale per promuovere una cultura orientata ai dati in cui tutti gli utenti, indipendentemente dalla loro formazione tecnica, possano accedere autonomamente ai dati rilevanti per il loro ruolo e sfruttarli al meglio. Quando i Data Consumer hanno la possibilità di interagire in modo indipendente con i dati, la dipendenza da un Team che gestisce centralmente i dati diminuisce, liberandolo e consentendogli di concentrarsi su iniziative strategiche di maggior valore. Inoltre, gli utenti che possono accedere autonomamente ai dati e gestirli responsabilmente sono maggiormente propensi ad essere partecipanti attivi, informati e responsabili nella più ampia comunità di dati dell'organizzazione.

In definitiva, un catalogo dei dati accessibile e orientato al consumatore è una pietra miliare per le organizzazioni che mirano a democratizzare i dati in modo efficace. Assicura che ogni consumatore di dati possa adempiere alle proprie responsabilità in modo efficiente e sicuro, contribuendo positivamente all'intelligenza collettiva e alla competitività dell'intera organizzazione.

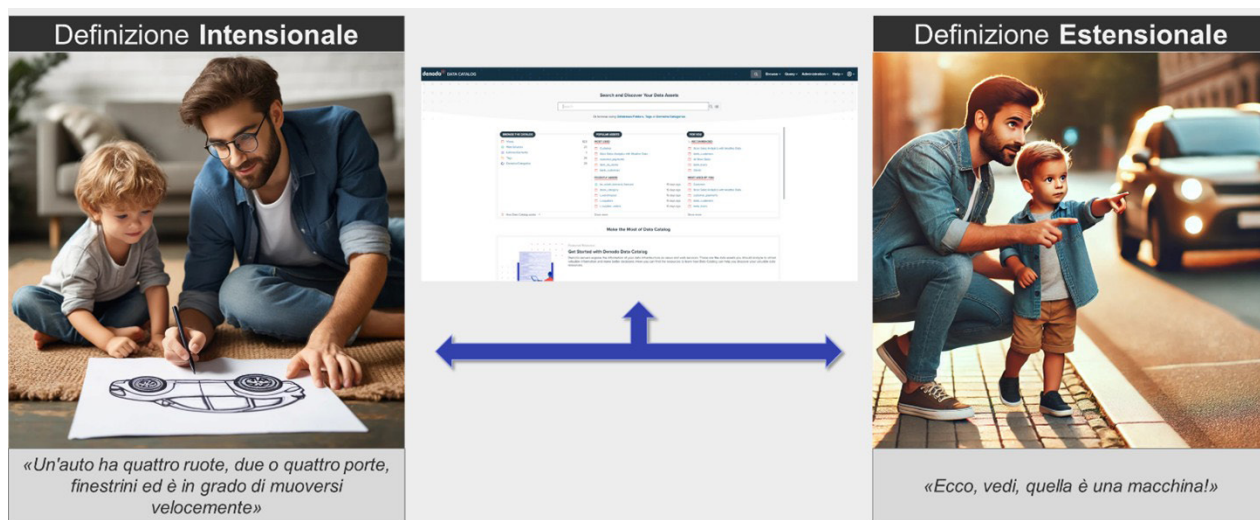


Il Catalogo dei dati, parte integrante della Piattaforma Denodo

RENDERE ATTIVE LE ONTOLOGIE CON LE ARCHITETTURE LOGICHE DEI DATI

Il modello semantico, cuore pulsante delle architetture logiche per l'integrazione dei dati, non solo mette al riparo da potenziali incoerenze nella definizione dei dati, riducendo – se non eliminando – il rischio di avere dati con gli stessi nomi che rappresentano concetti diversi, ma è anche lo strumento ideale, grazie al modo in cui il modello viene gestito in un contesto logico, per rendere viva un'ontologia, potendo, da un lato, importarla se è già stata definita altrove, in modo che il modello semantico sia automaticamente popolato dai concetti precedentemente definiti, e dall'altro collegarla alle fonti di dati che rappresentano la realizzazione di questi concetti. Questa connessione sarà responsabilità del Data Steward che, grazie alle sue competenze, sarà in grado di definire le cosiddette “viste di interfaccia”, arricchendo ciò che è rappresentato dall'ontologia con i dati che ne costituiscono la realizzazione.

In questo modo sarà possibile passare dalla componente logica (il significato dei dati) a quella fisica (le loro occorrenze) per fornire a chi deve utilizzare i dati la possibilità di valutarne, in fase di ricerca ed esplorazione, sia la correttezza della rappresentazione che la qualità dei dati raccolti, rafforzando così la fiducia che si può riporre in tali dati.



La connessione tra significato e occorrenze realizzata dal Data Catalog

In conclusione, le architetture logiche, con il loro concetto di modello semantico attivo, sono il ponte tra la componente intensionale dei dati e quella estensionale, una connessione continua e armoniosa che non richiede duplicazioni o spostamenti di dati.

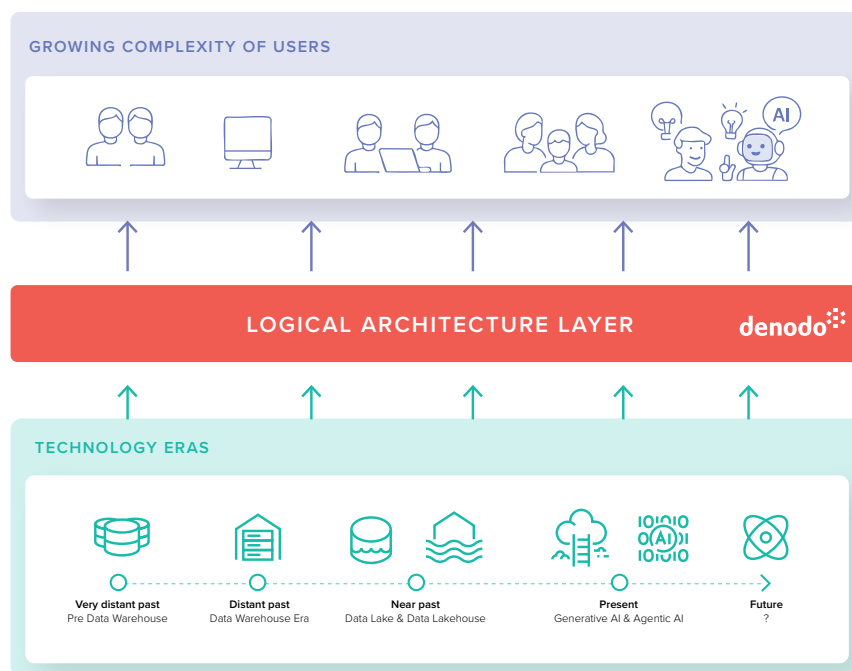
LA SEMANTICA NON BASTA, SERVE ANCHE LA RESILIENZA

Se il ruolo dei dati è tutto sommato invariante nel tempo, visto che il loro compito è quello di essere lo strumento per una concettualizzazione del mondo osservato, sia esso quello con cui un'azienda si relaziona o quello interno, che definisce il modo in cui essa si organizza e funziona, lo stesso non si può dire della tecnologia che ne permette la gestione, decisamente più mutevole, probabilmente in accelerazione e soggetta a salti quantici in corrispondenza di quelle innovazioni che possono essere considerate rivoluzionarie.

Una tecnologia soggetta a cambiamenti, a volte entropici, se da un lato permette di abilitare l'innovazione e il progresso, dall'altro crea inevitabili perturbazioni su chi deve utilizzare tale tecnologia, dove ogni cambiamento, ogni evoluzione, se non ben gestita, si riverbera inevitabilmente sugli utilizzatori di tale tecnologia, rallentando il loro lavoro e costringendoli a un continuo percorso di adattamento e formazione, con potenziali impatti negativi sull'efficienza complessiva e sulla capacità di governare quel tempo che, come detto all'inizio del documento, è ormai la nuova unità di misura.

Questo disturbo indotto dai cambiamenti tecnologici è efficacemente ridotto, se non addirittura eliminato, dalle architetture logiche per l'integrazione dei dati, che disaccoppiano l'accesso ai dati dai livelli fisici di storage e tecnologici sottostanti, creando un'astrazione flessibile che migliora significativamente la resilienza ai cambiamenti tecnologici. A differenza degli approcci tradizionali strettamente accoppiati, le architetture logiche garantiscono che gli aggiornamenti, le migrazioni o addirittura le sostituzioni delle tecnologie sottostanti avvengano in modo trasparente, senza interrompere l'esperienza del Data Consumer. Di conseguenza, gli utenti continuano ad accedere ai dati e a utilizzarli senza soluzione di continuità, al riparo da qualsiasi cambiamento, il che consente alle organizzazioni di evolvere tecnologicamente senza influire sulle operazioni quotidiane o sui flussi di lavoro degli utenti.

La figura seguente riassume graficamente la capacità delle architetture logiche di gestire un mondo che si muove a due velocità, dove all'invarianza delle motivazioni che guidano l'utilizzo dei dati ha corrisposto una significativa evoluzione tecnologica, che ha introdotto nuovi paradigmi per gestire uno scenario in cui i dati continuano a crescere in volume, eterogeneità e velocità.



Come un livello logico isola l'utilizzo dei dati dalla tecnologia di backend

IL RUOLO DELL'INTELLIGENZA ARTIFICIALE

Non va dimenticato che ogni algoritmo di intelligenza artificiale, tradizionale o generativo, utilizza i dati esattamente come fa ogni altro Data Consumer nello svolgimento delle sue attività, per cui anche in questo caso le architetture logiche sono lo strumento ideale, centralizzando la rappresentazione dei dati in modo che, da un lato, sia agevole costruire i dataset necessari agli algoritmi di intelligenza artificiale e, dall'altro, prelevando e consegnando i dati solo quando richiesto.

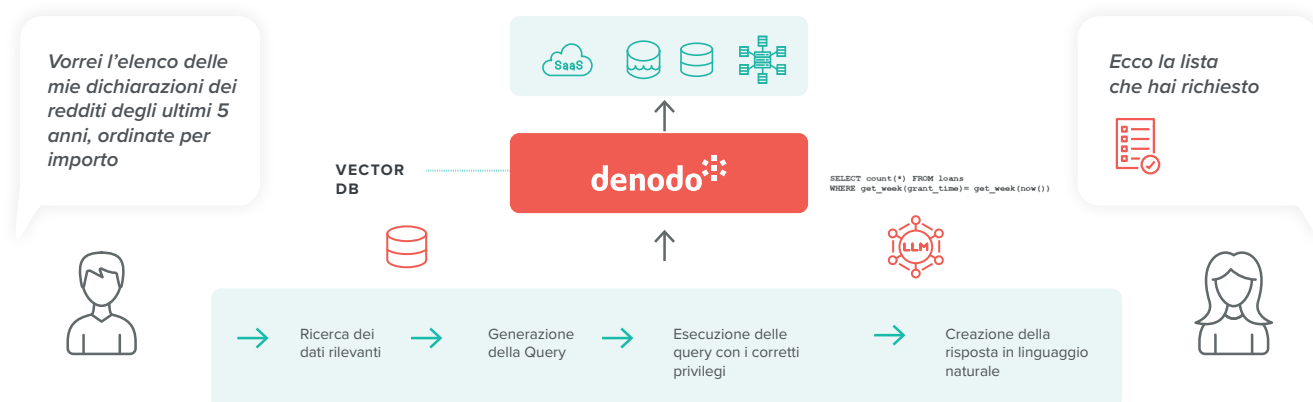
RETRIEVAL AUGMENTED GENERATION (RAG)

L'IA generativa riveste un ruolo sempre più centrale per la valorizzazione delle informazioni di natura testuale, rappresentando la base per assistenti virtuali di nuova generazione, in grado di sostenere conversazioni articolate, fornendo così a clienti, cittadini, collaboratori e dipendenti informazioni dettagliate, sia per un uso immediato, che per fornire più rapidamente le risposte.

Tuttavia, affinché questi assistenti virtuali siano in grado di rispondere in modo preciso e contestualizzato, è necessaria un'integrazione tra i Large Language Model (LLM) e le fonti di dati strutturati e non strutturati, necessari sia alla corretta comprensione del contesto nel quale le domande sono poste, che a reperire i dati che rappresentano l'oggetto della richiesta, dati che saranno utilizzati dai LLM per comporre la risposta.

Affinché un sistema di RAG possa associare correttamente gli elementi linguistici di una query in linguaggio naturale ai dati sottostanti, deve comprendere non solo i valori dei dati, ma anche il loro significato e il contesto al quale si riferiscono. Questa comprensione si basa su metadati ricchi e ben strutturati, che definiscono, con una terminologia vicina al business, le relazioni, le gerarchie e il significato dei dati.

Le Logical Data Architectures, in particolare quelle costruite sulla Denodo Platform, sono ideali per fornire questi metadati attraverso i loro modelli semantici, che offrono una visione astratta dei dati, orientata al business e indipendente dalle fonti fisiche, costituendo di fatto un ponte tra il linguaggio naturale e i dati strutturati. Di conseguenza, consentono ai sistemi RAG di interpretare con maggiore precisione l'intento dell'utente e di recuperare o generare risposte in linea con le aspettative dell'utente e con la vera semantica del panorama dei dati aziendali.



Un esempio di come il RAG consenta di rispondere a domande complesse

L'INTELLIGENZA ARTIFICIALE PER IL DATA MANAGEMENT

L'intelligenza artificiale, sia tradizionale che generativa, offre vantaggi sostanziali alla gestione dei dati. I metodi tradizionali di intelligenza artificiale, come gli algoritmi di apprendimento automatico, semplificano i processi automatizzando la classificazione, la convalida e la garanzia di qualità dei dati, alleggerendo notevolmente il carico di lavoro dei Data Steward. L'IA generativa espande queste capacità consentendo interazioni in linguaggio naturale e sintesi dei dati, facilitando notevolmente la scoperta, la comprensione e la fruibilità dei dati da parte dei Data Consumer.

GESTIONE SEMPLIFICATA DEI DATI

Le tecnologie di intelligenza artificiale automatizzano le attività di gestione, tra cui l'estrazione dei metadati, il monitoraggio della qualità dei dati e il rilevamento delle anomalie. Per i Data Steward, questa automazione riduce l'impegno manuale, assicura la coerenza delle pratiche di gestione dei dati e migliora l'accuratezza delle attività di governance. Il risultato è un approccio più efficiente, affidabile e agile alla gestione di ecosistemi di dati ampi e diversificati.

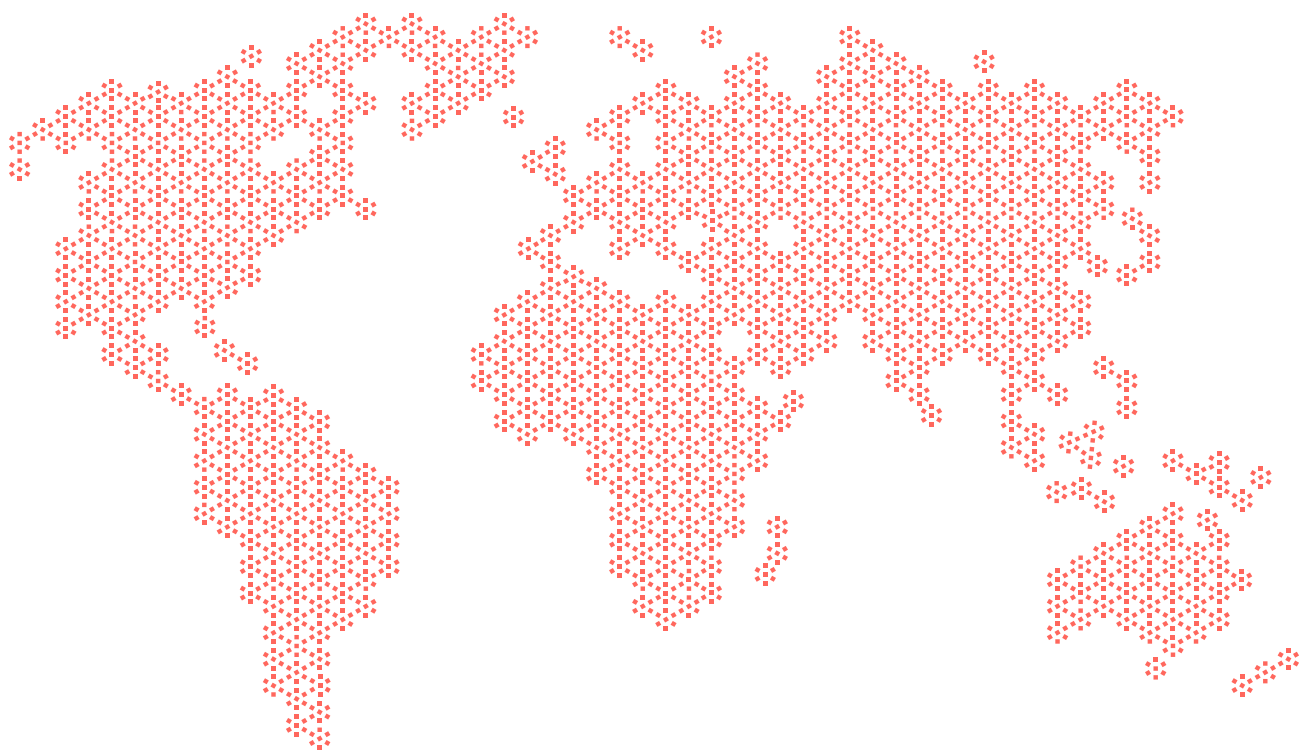
Con specifico riferimento alla piattaforma Denodo, il Denodo AI Assistant offre supporto, tra l'altro, per:

- Ottimizzazione delle query. La piattaforma Denodo utilizza da tempo l'intelligenza artificiale per l'ottimizzazione delle query, tra cui l'ordine di join, le associazioni e le aggregazioni. Ciò contribuisce a garantire prestazioni ottimali con costi ridotti, riducendo la richiesta di competenze specifiche, necessarie quando tale lavoro viene fatto manualmente.
- Preparazione e arricchimento dei dati. La piattaforma Denodo dispone di funzionalità di preparazione dei dati che sfruttano le potenzialità dell'intelligenza artificiale. Ad esempio, al Data Steward che sta lavorando su uno specifico insieme di dati, viene presentata una serie di trasformazioni consigliate in base ai suoi comportamenti precedenti e di quelli di utenti simili che hanno utilizzato lo stesso set di dati o altri dati simili, una capacità che migliora con l'uso, dato che la piattaforma Denodo apprende e aggiorna i metadati sulla base del loro effettivo uso.

MIGLIORAMENTO DELL'ACCESSIBILITÀ DEI DATI E DELL'ESPERIENZA UTENTE

L'intelligenza artificiale generativa consente di creare interfacce intuitive e di facile utilizzo, facilitando notevolmente la ricerca e l'interpretazione di dati complessi da parte dei consumatori di dati. Traduce i metadati tecnici in approfondimenti comprensibili, consentendo agli utenti privi di competenze tecniche approfondite di sfruttare i dati in modo efficace, accelerando così il processo decisionale informato in tutte le funzioni aziendali.

Con specifico riferimento alla piattaforma Denodo, il Denodo AI Assistant consente a ogni utente di formulare una domanda in linguaggio naturale, che viene poi convertita in codice SQL ed eseguita dalla piattaforma. Si tratta di un grande passo avanti nella democratizzazione dei dati, che consente ai Data Consumer di esplorare, comprendere e usare i dati con facilità.



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

