

# Otto modi per modernizzare il data management

Come sfruttare l'intelligenza artificiale,  
gestire la crescente diversificazione dei dati  
e integrare al meglio le implementazioni





# Sommario

## **Introduzione**

Le caratteristiche dei dati sono cambiate nel tempo

Pagina 3

## **Uno**

Promuovere ottimizzazione e sviluppo di applicazioni con l'AI

Pagina 4

## **Due**

Ottenere rapidamente report e analisi

Pagina 6

## **Tre**

Ridurre il costo totale di proprietà

Pagina 8

## **Quattro**

Ottenere la flessibilità di scegliere servizi on-premise o cloud secondo necessità

Pagina 10

## **Cinque**

Aumentare la scalabilità e la disponibilità costante

Pagina 12

## **Sei**

Supportare nuove tipologie e fonti di dati

Pagina 14

## **Sette**

Integrare la tecnologia open source

Pagina 16

## **Otto**

Incorporare nuove funzionalità per dati in streaming e fast data

Pagina 19

## **Conclusione**

Come iniziare il processo di modernizzazione del data management

Pagina 21

## Introduzione

Le caratteristiche dei dati sono cambiate nel tempo



## È per questo che anche il data management dovrebbe cambiare

I tempi in cui i dati venivano definiti in modo restrittivo come informazioni altamente strutturate provenienti da poche fonti specifiche sono ormai lontani. Cercare di cambiare questa definizione deve essere un obiettivo per le aziende che intendono sfruttare una più ampia varietà di tipi e strutture di dati provenienti da una molteplicità di fonti. Questo allo scopo di fornire a tutti, dal dipartimento IT ai dipendenti della linea di business (LOB), ai dirigenti di categoria C, degli insight che possano avere un impatto immediato e trasformativo.

L'uso e la gestione efficace delle informazioni è fondamentale per perseguire nuove opportunità di business, attrarre e mantenere i clienti e semplificare le operazioni. Tuttavia, queste esigenze comportano una serie di sfide in termini di carico di lavoro e aumentano le richieste di infrastrutture IT sottostanti e sistemi di database che spesso non sono all'altezza del compito.

La domanda è: in che modo vanno affrontate queste sfide? Assegnando più personale per tenere il passo con le patch, i componenti aggiuntivi e la continua regolazione richiesta dai sistemi esistenti, o semplicemente ignorando i potenziali insight che risiedono in questa immensa quantità di nuovi dati? Molte aziende stanno affrontando questa sfida a testa alta cercando nuove soluzioni che sfruttino anche l'intelligenza artificiale (AI), come le molteplici funzionalità e opzioni di implementazione dei database on-premise e dei cloud pubblici e privati, per innovare l'infrastruttura dei dati e il business aziendale.

Questo ebook presenta otto modi in cui architetti e CIO (Chief Information Officer) dell'azienda possono modernizzare i processi di data management per superare le sfide che l'architettura dell'informazione impone e identificare nuove opportunità di business; sono inoltre incluse opinioni di analisti del settore ed esperienze reali di clienti.



## Uno

# Promuovere ottimizzazione e sviluppo di applicazioni con l'AI

Per essere competitive, le aziende devono estrarre dai loro dati insight che siano il più possibile approfonditi, accurati e fruibili, in modo abbastanza rapido da poterli implementare nel momento in cui sono più utili. Gli insight ottenuti dai dati possono portare innovazione all'intera azienda, permettendo di individuare opportunità per nuovi prodotti o mercati, consentendo agli addetti alle vendite di avere argomenti più significativi e identificando i processi interni che possono essere migliorati. [Per questo motivo, molte aziende si stanno orientando verso soluzioni di data management alimentate dall'AI e progettate per l'AI.](#)

### Alimentato dall'AI

Le soluzioni alimentate da AI e machine learning (ML) ottimizzano in modo diretto le performance del data management, con un effetto che si ripercuote su tutta l'azienda.

Ad esempio, il machine learning può essere utilizzato per velocizzare in modo significativo le tempistiche di elaborazione delle query, migliorando continuamente il percorso che una query compie per arrivare ai dati. Inoltre, può essere utilizzato per ottenere risultati delle query basati sulla fiducia, in modo da rendere più facile ed efficiente il conseguimento dei migliori insight. È inoltre utile integrare il linguaggio naturale nell'elaborazione delle query in modo che gli utenti della LOB (line of business) possano utilizzare un'interfaccia simile a quella della ricerca su Internet per trarre insight direttamente dai dati; in questo modo potranno aiutare a diffondere gli insight in tutta l'azienda, eliminando gli intoppi interni e risparmiando il tempo prezioso dei data scientist e dei business analyst.



## Progettato per l'AI

Le soluzioni progettate per l'AI supportano i framework, i linguaggi e gli strumenti di cui i data scientist e gli sviluppatori di applicazioni hanno bisogno per implementare rapidamente l'AI nei loro progetti evitando i ritardi causati dalle riscritture al momento di portare i prototipi in produzione.

Una soluzione progettata per l'AI include supporto per i linguaggi più diffusi, come Python, GO e JSON, così come per Jupyter Notebooks. Le librerie per il machine learning e altri strumenti aggiuntivi consentono agli utenti di integrare le applicazioni di AI con le proprie soluzioni di data management. Grazie a questo, i data scientist e gli sviluppatori di applicazioni non hanno bisogno di perdere tempo per creare e imparare un nuovo codice. Possono invece iniziare a lavorare immediatamente sui progetti prendendo in prestito un codice preesistente da lavori precedenti o da librerie in modo da ultimare i progetti più velocemente che se fossero partiti da zero.

## Autovalutazione sul data management

- Il vostro ambiente dati supporta i linguaggi e gli strumenti più usati dai data scientist?
- Le vostre query possono essere migliorate con il machine-learning e create con l'uso del linguaggio naturale?
- Quanto velocemente vengono risolte le vostre query, e si basano sulla fiducia nella loro accuratezza?

# Gli analisti annunciano l'arrivo del data management basato sull'AI

Gli analisti riconoscono che basando il data management sull'AI è possibile apportare grandi trasformazioni nelle aziende. Il recente rapporto di 451 Research analizza il motivo per cui l'88% degli intervistati delle aziende più orientate all'utilizzo dei dati ha dichiarato che "AI e ML sono una componente importante della piattaforma dati e delle iniziative di analisi".

Hanno inoltre delineato l'intero insieme dei casi d'uso dell'AI, dall'automazione di compiti ripetitivi all'automazione di compiti complessi permettendo al personale di concentrarsi su tutto il resto.



*"AI e ML sono una componente importante della piattaforma dati e delle iniziative di analisi".*

451 Research: Accelerare l'AI con il Data Management; Accelerare il Data Management con l'AI. [Leggi il report](#) →

88%

## Due

# Ottenere rapidamente report e analisi

Sono disponibili diverse tecnologie avanzate che aiutano ad accelerare il reporting e l'analisi, spaziando da calcolo ed elaborazione a compressione e data skipping. Se combinate tra loro, queste tecnologie possono apportare miglioramenti significativi alla velocità con cui si possono ottenere insight. Queste tecnologie sono particolarmente importanti per i data scientist e i business analyst, poiché aiutano a garantire che le query vengano risolte più rapidamente, consentendo alle aziende di prendere decisioni tempestive.

### **Elaborazione in-memory con tabelle shadow a colonne**

L'elaborazione in-memory ha trovato diffusione grazie ai miglioramenti che apporta alle performance. Tuttavia, non tutti i database in-memory trattano i dati allo stesso modo. A causa della necessità di funzionalità di lettura e scrittura, i database sono solitamente organizzati in righe e richiedono che tutti i dati siano archiviati in memoria per ottenere i miglioramenti di performance desiderati. Il problema è che a causa della loro costante crescita potrebbe non essere possibile inserire tutti i dati in memoria. Per questo motivo sono necessarie tabelle shadow a colonne, che possono essere analizzate in parallelo con le tabelle a righe. La lettura delle colonne risulta più efficiente e le colonne necessarie possono essere messe in memoria per l'analisi. Inoltre, il flusso ottimizzato dei dati dallo storage alla memoria di sistema e alla memoria della CPU che si trova nella tecnologia a colonne in-memory, come IBM® Db2® BLU Acceleration®, accelera la velocità con cui possono essere risolte le query. I carichi di lavoro OLTP e OLAP possono quindi essere eseguiti in parallelo nello stesso database, il che significa che le query possono essere eseguite senza migrare su un data warehouse e che è possibile ricevere risposte più velocemente.



### Compressione attiva

La compressione attiva consente di scansionare o confrontare i dati senza decomprimerli. Quindi la decompressione può essere eseguita solo per i dati necessari e posticipata fino a quando non è assolutamente necessaria. I database ottengono questo risultato comprimendo ogni colonna con i propri dizionari per la compressione e comprimendo i valori dei dati che appaiono più spesso di altri. Ciò consente di risparmiare tempo per gli insight migliorando il throughput degli I/O, conservando al tempo stesso la potenza di elaborazione e riducendo direttamente i costi complessivi di storage.

### Data skipping

Il data skipping aiuta a risparmiare tempo durante l'elaborazione delle query, identificando i dati irrilevanti per la query e ignorandoli. I database ottengono questo risultato tenendo traccia dei valori minimi e massimi delle righe utilizzando oggetti di metadati denominati tabelle sinottiche. Quando si eseguono query, le tecnologie di data skipping analizzano le tabelle sinottiche ricercando intervalli di dati contenenti valori inerenti alla query e saltando i dati che non corrispondono. Solo i dati rilevanti vengono letti e utilizzati, quindi la scansione è più efficiente; vengono così ridotti gli I/O non necessari e l'esecuzione delle query risulta più veloce.

### Autovalutazione sul data management

- Durante un'analisi, quanto tempo impiegate nell'attesa che i dati si decomprimano?
- La velocità con cui le query vengono completate è sufficiente a permettervi di ottenere il massimo dagli insight?
- È possibile analizzare i dati transazionali in modo efficiente senza interrompere le operazioni?

Come un produttore globale ha ridotto il tempo dedicato al reporting del 98% servendosi di alcune tecnologie chiave

Per Knorr-Bremse, produttore di sistemi frenanti, l'espansione della propria presenza a livello globale attraverso acquisizioni e joint venture ha comportato un notevole incremento del volume di dati. Dovevano elaborare fino a 4 TB di dati al giorno. La scelta di tecnologie come la compressione attiva, il data skipping e l'elaborazione in-memory ha fornito notevoli vantaggi.



*“Con Db2 BLU Acceleration abbiamo potuto ridurre i tempi necessari per il processo di reporting da 3 minuti a 3 secondi”.*

– Thomas Brauchle, Director of IT Architecture and Strategy, Knorr-Bremse

98%



## Tre

# Ridurre il costo totale di proprietà

Dato che i budget dedicati al settore IT continuano ad aumentare, è più importante che mai che le aziende utilizzino le risorse finanziarie e il personale in modo efficiente. È necessario che i database forniscano funzionalità di gestione dei dati economicamente vantaggiose e all'avanguardia, soddisfacendo al contempo gli SLA (Service Level Agreement) con un'ampia gamma di servizi completamente gestiti e un supporto di classe enterprise.

**Funzionalità autonome** L'automazione delle attività amministrative aiuta a raggiungere entrambi questi obiettivi, sgravando il personale in modo che possa concentrarsi su iniziative più strategiche e accelerando i flussi di lavoro. È estremamente importante dotarsi di soluzioni di database in grado di automatizzare diverse attività amministrative, come la configurazione e l'implementazione, la gestione del carico di lavoro, l'utilizzo delle risorse e la gestione dello storage, nonché la manutenzione, gli aggiornamenti e l'espansione della capacità.

Scegliendo soluzioni di dati con opzioni cloud e on-premise che condividono una code base comune tra le varie implementazioni (anche su cloud concorrenti) è possibile ridurre il tempo e l'impegno dedicati alle mansioni amministrative, in modo da avere più risorse da impiegare per le riscritture. Un buon esempio è rappresentato dalla famiglia Db2 di prodotti per la gestione dei dati.





## Riduzione dell'impatto dello storage

Lo studio di Spiceworks "2019 State of IT" ha rivelato quanto hanno speso le aziende da oltre 5000 dipendenti per lo storage: L'8% o più del loro budget IT, per hardware, servizi di hosting e cloud e servizi gestiti. La riduzione delle risorse da dedicare allo storage può quindi avere un impatto significativo. Questo può essere ottenuto attraverso le tecnologie di compressione dei dati e di data management multi-temperature, che riducono in modo diretto le esigenze di storage, o scaricando i dati "freddi" su opzioni di storage più economicamente convenienti.

Chi desidera passare da un database all'altro per risparmiare e accelerare i processi senza dover modificare le applicazioni esistenti, può anche considerare la possibilità di cercare soluzioni con la chiarezza della topologia del cluster. Questa soluzione non richiede che le applicazioni siano consapevoli del cluster sottostante e dell'implementazione del database, e quindi aiuta ad accelerare la codifica e il testing e a rendere gli sviluppatori di applicazioni più produttivi.

## Autovalutazione sul data management

- Qual è il livello di automazione del vostro sistema attuale?
- Quanto tempo impiegate nella vostra azienda per le riscritture?
- Avete confrontato i prezzi dei servizi di data management dei vari fornitori?

## Come Owens-Illinois ha ottenuto una riduzione dei costi a sette cifre grazie alla modernizzazione

Per il produttore globale Owens-Illinois, aver ottenuto una riduzione dei costi a sette cifre è un traguardo notevole. Se si aggiungono a questo risparmio le transazioni più rapide e una riduzione quasi del 50% dell'impatto del database e dello storage, è doveroso far conoscere al mondo di questo risultato. Scopri la loro storia completa nel video qui sotto.

*"Abbiamo ottenuto una riduzione a sette cifre del costo totale di proprietà, una riduzione dello storage del 50% e miglioramenti delle prestazioni di almeno il 20-30% nei tempi di risposta delle transazioni".*

– Rod Masney, CIO di Owens-Illinois



50%

## Quattro

# Ottenere la flessibilità di scegliere servizi on-premise o cloud secondo necessità

A seconda delle esigenze specifiche, è possibile scegliere le implementazioni che meglio rispondono alle proprie necessità: on-premise, cloud privato, hosting su cloud, cloud pubblico o un modello ibrido che sfrutta diverse piattaforme. Poiché non tutti i dati sono uguali o vengono usati allo stesso modo, è importante avere accesso a una serie di opzioni di implementazione. La scelta di un fornitore che offra un'ampia gamma di opzioni di implementazione in grado di supportare

tutti gli aspetti della strategia organizzativa offrirà una maggiore flessibilità rispetto a quella di qualsiasi altro fornitore che offra una sola opzione di implementazione. Inoltre, se le implementazioni provengono tutte dalla stessa famiglia di prodotti per il data management, è probabile che si integreranno meglio rispetto a un'architettura costruita a partire da soluzioni indipendenti.

### On-premise

Le implementazioni on-premise forniscono un alto livello di controllo sull'ottimizzazione, l'accesso ai dati e la visibilità, ma richiedono grandi investimenti in hardware e talenti IT. Tali implementazioni vengono utilizzate soprattutto quando la sicurezza è fondamentale o quando le normative le impongono.

### Cloud privato

Le implementazioni di cloud privato consentono di sfruttare la sicurezza del proprio firewall, offrendo al contempo la flessibilità tradizionalmente associata al cloud pubblico. La possibilità di sfruttare la tecnologia container è un vantaggio considerevole per questo tipo di implementazione. Rispetto alle macchine virtuali, i container sono più facili e veloci da implementare e spesso richiedono meno spazio sul server per funzionare.

### Cloud in hosting

Gli ambienti cloud in hosting eliminano la responsabilità dell'acquisto e della manutenzione dell'hardware, consentendo al contempo di gestire in parte il sistema. Le implementazioni cloud in hosting sono ottimali per i progetti di disaster recovery perché non è necessario acquistare infrastrutture aggiuntive e si ha la possibilità di utilizzare le licenze dei database esistenti. Il cloud in hosting si integra perfettamente con i sistemi on-premise.

### Cloud pubblico

Le implementazioni di cloud pubblico forniscono uno scarso controllo diretto su aspetti come la larghezza di banda e il monitoraggio ma hanno una maggiore flessibilità. Il rapido aumento della capacità del cloud pubblico per supportare cloud privati, noto come cloud bursting, permette di separare le esigenze di elaborazione e di storage e di pagare solo per l'utilizzo effettivo, ciò significa che le implementazioni gestite del cloud sono ottime opzioni per i backup e lo sviluppo, o per gli ambienti di testing.

## Autovalutazione sul data management

- Siete in grado di supportare le implementazioni on-premise e cloud sulla stessa code base?
- Al momento siete in grado di aggiungere e integrare con facilità diverse implementazioni di dati?
- Sono disponibili opzioni multicloud? Qual è il loro livello di integrazione?

## Perché è importante avere un ambiente ibrido completo e connesso



Le aziende non sono più vincolate alla falsa dicotomia “on-premise O cloud” e scelgono di cercare fornitori in grado di fornire opzioni on-premise E cloud ben integrate. Scopri di più sulle funzionalità necessarie per colmare il divario tra le diverse implementazioni di data management.

Percorso verso il cloud [Leggi il white paper →](#)





## Cinque

# Aumentare la scalabilità e ottenere una disponibilità costante

Per mantenere operativa un'azienda è necessario che i sistemi di dati siano sempre attivi e sufficientemente solidi da poter supportare la crescente quantità di dati generati. Operatività continua e risposte istantanee sono importanti per clienti e collaboratori, ma anche internamente. Gli errori portano a perdere nuove opportunità di business basate su insight, a un rapporto insoddisfacente con i collaboratori e alla perdita di entusiasmo. Di seguito sono descritte le tre opzioni più utili per la scalabilità e la disponibilità.

### **Clustering dello storage condiviso**

Una delle opzioni è il clustering di storage condiviso, che permette di distribuire il lavoro tra diversi nodi che si basano tutti sulla stessa fonte di dati. Due funzionalità chiave per il clustering dello storage condiviso sono il bilanciamento automatico del carico di lavoro tra i nodi dell'architettura e il failover automatico integrato per utilizzare ogni nodo del cluster dello storage condiviso come nodo di failover per gli altri. Insieme queste funzionalità aumentano la ridondanza ed evitano il sovraccarico di lavoro di alcuni nodi in modo da ridurre le possibilità di guasto e fornire una migliore disponibilità anche in situazioni di emergenza.



### Integrare il cloud on-premise e il cloud privato con il cloud pubblico

Le implementazioni on-premise e di cloud privato possono migliorare sia la scalabilità che la disponibilità con l'aggiunta del cloud pubblico. Questo si ottiene utilizzando le offerte on-premise e di cloud privato come funzionalità di base, utilizzando invece le offerte di cloud pubblico quando sono richieste funzionalità di storage o di elaborazione aggiuntive. Perché il risultato sia efficiente, è necessaria una forte integrazione tra i sistemi cloud e on-premise, in modo che i dati non vengano isolati in base all'implementazione. La creazione di un ambiente cloud pubblico complementare a un'offerta di cloud privato o on-premise esistente può anche fornire una migliore disponibilità in caso di interruzioni pianificate o non all'interno dell'azienda.

### Scalare lo storage del cloud ed elaborare in modo indipendente

Le aziende dovrebbero essere in grado di aumentare o ridurre in modo indipendente lo storage e l'elaborazione all'interno delle offerte cloud per far fronte alla variabilità dei carichi di lavoro specifici del cloud. Con una scalabilità indipendente, il cloud può aumentare l'elaborazione in modo da far fronte ai momenti in cui è richiesto il massimo delle prestazioni per poi ridimensionarla in modo da risparmiare fondi quando non è più necessario mantenere prestazioni elevate. Una maggiore disponibilità richiede anche un minore sforzo grazie ai servizi cloud completamente gestiti, come HADR (High availability disaster recovery), backup giornalieri, manutenzione e supporto aziendale.

### Autovalutazione sul data management

- Qual è la vostra spesa attuale per il trasferimento dei dati?
- Le vostre offerte cloud e on-premise condividono una code base comune, consentendo una migliore integrazione?
- Siete in grado di scalare lo storage del vostro database cloud e di elaborare in modo indipendente?

## Perché il centro di business intelligence per la sanità olandese utilizza cloud scalabili

Vektis, il centro di business intelligence per la sanità olandese, aveva bisogno di creare un'unica versione dei dati reali passando dal fornire i dati stessi, potenzialmente soggetti a modifiche e manipolazioni, al fornire l'accesso a una versione unica dei dati raccolti. Tuttavia, l'azienda sapeva anche quanto fosse importante scalare per poter accogliere più utenti, tutti con la necessità di sottoporre query complesse contemporaneamente.



*“Volevamo migrare su IBM Db2 Warehouse on Cloud per avere la possibilità di poter scalare in modo flessibile. Questa funzionalità ci rende più agili, ma ci dà anche la possibilità di controllare i costi”.*

– Herman Bennema, CEO di Vektis

# Sei

## Supportare nuove tipologie e fonti di dati

Sono molti i tipi e le fonti di dati che ultimamente hanno acquistato rilevanza. Si consideri la notevole crescita dei dati semi-strutturati e non strutturati generati da fonti come i social media, i sensori nei dispositivi collegati all'Internet of Things (IoT) e i repository di contenuti multimediali. Inoltre, molte di queste fonti producono dati in streaming ad alta velocità e in grande quantità.

Mettendo insieme i dati esterni con quelli che l'azienda già raccoglie, si possono generare insight più profondi e accurati in grado di favorire decisioni migliori. Ad esempio, utilizzando solo i dati interni si potrebbero rilevare vendite iniziali consistenti per un nuovo prodotto. Tuttavia la combinazione di questi dati con quelli non strutturati ricavati dai social media e da altre fonti esterne potrebbe portare a conclusioni diverse. Usando la sentiment analysis, un'azienda è in grado di capire se quel prodotto ha fornito esperienze negative nonostante sia stato acquistato frequentemente, permettendo di cogliere il problema e di correggerlo prima che le vendite calino.

Tuttavia, la combinazione di dati interni ed esterni pone una nuova serie di sfide per l'implementazione e lo sviluppo di un'infrastruttura di dati, che deve garantire la flessibilità necessaria per fornire, acquisire, archiviare ed elaborare un'ampia varietà di tipi di informazioni provenienti da diverse fonti e a diverse velocità. Le aziende hanno bisogno di un ambiente di dati in grado di fornire i dati secondo i tempi e le modalità richiesti e di registrare in modo preciso ciò che accade in ogni momento.

Alcuni esempi dei dati non tradizionali più rilevanti:



**I contenuti audiovisivi** offrono alle aziende l'opportunità di attingere a conoscenze che altrimenti potrebbero essere difficili da trovare o rivedere, aggiungendo preziose informazioni contestuali.



**I dati degli eventi** forniscono informazioni spesso molto sensibili al fattore tempo. Se acquisiti e analizzati prima che perdano la loro rilevanza, i dati degli eventi possono fornire input in tempo reale per compiere azioni adeguate.



**I dati dei social media** possono fornire una visione a 360° dei clienti, aiutando il marketing mirato, portando a un aumento delle vendite e identificando velocemente e mitigando eventuali problemi con i clienti o la percezione generale del pubblico.



**I dati all'interno di documenti fisici** possono essere più difficili da acquisire, ma consentono alle aziende di ottenere informazioni registrate manualmente, che possono essere più ricche e avere un contesto più dettagliato rispetto ai dati reperiti online.



**I dati dei sensori IoT** possono fornire notifiche di allerta quando i macchinari o le linee di produzione sono a rischio di malfunzionamento, incoraggiando misure preventive piuttosto che azioni reattive.



**Il contenuto delle email** può includere e spesso include componenti audiovisive, e fornisce inoltre informazioni più importanti rispetto a quanto riportato nel testo, come i dati di indagine, o fenomeni che possono essere osservati, come nel caso delle transazioni.



## Autovalutazione sul data management

- Quanti dati non strutturati siete attualmente in grado di acquisire e analizzare?
- Con quale frequenza combinate dati interni ed esterni per trarre insight?
- Quanti dispositivi IoT vengono utilizzati quotidianamente dalla vostra azienda o dai vostri clienti?

## Come Capitalogix combina dati strutturati e non strutturati per dare un significato contestuale agli hedge fund

Con azioni detenute per una media di 20 secondi, la società di tecnologia finanziaria Capitalogix è costantemente alla ricerca di strategie per raccogliere più velocemente insight significativi e stare al passo con le tendenze del mercato. Per fare ciò, si serve di una varietà di dati strutturati e non strutturati a varie velocità.

*“Abbiamo accesso a dati alternativi che potrebbero essere totalmente non strutturati: sentiment data provenienti dai feed di Twitter e da immagini satellitari”.*

– Howard Getson, CEO di Capitalogix



## Sette

# Integrare la tecnologia open source

La tecnologia open source è una parte fondamentale del panorama moderno del data management. È di fondamentale importanza poter contare sull'impegno di un'intera community di sviluppatori, evitare il vendor lock-in e sfruttare più fonti di dati con opzioni open source. Quando la tecnologia open source è implementata correttamente, può portare a una riduzione dei costi e a una migliore qualità di insight.

### **Sfruttare l'innovazione portata dalla community ed evitare il vendor lock-in**

Le soluzioni open source sono continuamente perfezionate da una community di persone e talvolta di aziende. L'ampia conoscenza tecnica e il gran numero di operatori qualificati in grado di gestire qualsiasi problema o nuova idea vanno oltre le capacità di una singola azienda e conducono verso tecnologie più innovative. Oltre a queste tecnologie all'avanguardia, le soluzioni open source contribuiscono a mitigare il rischio di vendor lock-in in diversi modi. Poiché le stesse tecnologie open source sono accessibili a più fornitori, è più facile selezionare un fornitore diverso senza dover perdere tempo per ricodificare. Inoltre le competenze di codifica open source sono trasferibili, a differenza delle conoscenze proprietarie richieste per i sistemi proprietari. Il superamento del vendor lock-in riduce i costi aumentando la competitività.



## Ottenere l'accesso a più dati

Le soluzioni open source forniscono anche l'accesso a un numero di dati superiore a quello che altrimenti potrebbe essere inserito nei database tradizionali o nei data warehouses. Ad esempio, Hadoop è spesso il repository principale per le informazioni non strutturate, grazie alla sua capacità di evitare l'assegnazione di uno schema durante la memorizzazione dei dati. È anche possibile scaricare dati “freddi” in Hadoop da un data warehouse.

Ma l'accesso una maggiore quantità di dati attraverso la tecnologia open source va oltre capacità di Hadoop. Database come MongoDB e PostgreSQL ampliano i database proprietari. MongoDB, per esempio, è un database di documenti basato su NoSQL. Memorizzando i dati nei documenti JSON, permette una rapida evoluzione dello schema, che in genere non avviene in database relazionali più strutturati. Questo lo rende adatto a casi d'uso che coinvolgono applicazioni mobili, analisi in tempo reale e gestione dei contenuti, tutte operazioni che beneficiano della flessibilità dello schema. Allo stesso modo, sebbene PostgreSQL sia un database relazionale, riesce ad ampliare i database relazionali esistenti in un'organizzazione senza contrapporsi a questi ultimi. Questo perché PostgreSQL è un database relazionale ad oggetti, il che significa che supporta oggetti definiti dall'utente e il modo in cui questi si comportano per creare, memorizzare e recuperare strutture di dati più complesse.

PostgreSQL supporta diversi tipi di dati: UUID, monetari, enumerati, geometrici, binari, indirizzi di rete, stringhe di bit, ricerche di testo, XML, JSON, array, compositi e campi di variazione, così come alcuni tipi di dati interni per l'identificazione di oggetti e l'individuazione di file di log. Anche se non sempre è necessario sfruttare a pieno tutta questa flessibilità, essa rappresenta un utile complemento alle banche dati relazionali standard. In definitiva queste tecnologie open source supplementari consentono l'accesso a più tipi di dati da più fonti, fornendo le informazioni supplementari necessarie per trarre insight in grado di stimolare l'innovazione aziendale e l'ottimizzazione in modo da migliorare i risultati finanziari.

# Gli analisti approvano il valore totale di proprietà della partnership strategica tra IBM e Cloudera

Secondo gli analisti di Cabot Partners, la crescita e l'impatto positivo dell'intelligenza artificiale (AI) sono possibili solo attraverso la giusta architettura di informazioni (IA). Questo comporta l'adozione di soluzioni open source di classe enterprise, come quelle fornite da IBM e Cloudera. Dopo aver condotto uno studio del valore totale di proprietà (Total Value of Ownership, TVO) hanno concluso che “insieme IBM e Cloudera forniscono ai clienti flessibilità, ampia scelta e valore senza precedenti nel loro percorso di sviluppo dell'analisi e dell'AI”. Si invita a leggere il rapporto completo per vedere come la combinazione di questi due leader della tecnologia open source può aiutare a garantire un costo totale di proprietà più basso, una maggiore produttività, un aumento dei ricavi e una diminuzione del rischio.



*“Insieme IBM e Cloudera forniscono ai clienti flessibilità, ampia scelta e valore senza precedenti nel loro percorso di sviluppo dell'analisi e dell'AI”.*

Maggiore scelta e valore per Advanced Analytics e AI. [Leggi il report →](#)



## Scegliere la giusta tecnologia open source

Le aziende devono garantire che le soluzioni open source siano sicure, ben integrate e correttamente gestite. Anche se queste operazioni possono essere svolte internamente all'azienda, è spesso più efficiente affidarsi al supporto di un fornitore di classe enterprise e a soluzioni che portano un valore aggiunto. Prendiamo ad esempio la sicurezza; le soluzioni aziendali preconfezionate che includono la sicurezza consentono agli sviluppatori di concentrarsi su attività più produttive invece di perdere tempo a organizzare controlli di accesso e a integrare il sistema nel firewall aziendale.

Per l'integrazione, un'altra situazione in cui una tecnologia open source aziendale fornisce un notevole valore aggiunto è l'organizzazione di dati attraverso Hadoop e il data warehouse utilizzando un comune motore SQL. Senza questa integrazione, ad esempio, l'accostamento dei dati di vendita della prima settimana con i sentiment data dei social media per fornire una previsione della domanda futura richiederebbe uno sforzo molto maggiore. Una governance di classe enterprise aiuta inoltre a ridurre lo sforzo necessario per raggruppare i dati. Senza i metadati, una catalogazione e un'organizzazione generale di classe enterprise e soluzioni open source che comprendono un data lake, possono rapidamente insorgere difficoltà nell'estrazione dei dati esatti di cui si ha bisogno. Il data lake può diventare quello che alcuni definiscono una palude di dati (data swamp).

## Autovalutazione sul data management

- Siete attualmente bloccati in soluzioni proprietarie che potrebbero essere sostituite dall'open source?
- La vostra azienda ha casi d'uso specifici più adatti alle funzionalità dei database open source?
- Le vostre attuali soluzioni open source garantiscono la sicurezza, l'integrazione e la governance necessarie?



## Otto

# Incorporare nuove funzionalità per dati in streaming e fast data

Fattori come l'incredibile crescita dei dati, la crescente prevalenza di progetti di AI e una maggiore comprensione dei vantaggi dell'open source stanno cambiando il modo in cui le aziende acquisiscono e utilizzano i dati in streaming. Questo cambiamento è di fatto così evidente che alcuni analisti hanno iniziato a fare riferimento a questa categoria del data management con il termine più ampio di “fast data”, invece del più restrittivo “analisi di dati in streaming”. Il passaggio a soluzioni di fast data può aiutare le aziende ad acquisire e utilizzare rapidamente i dati, a integrare meglio le tecnologie open source e ad ottenere analisi e prestazioni di AI più solide.

### Ottenere maggiore velocità, più dati e migliori insight

IoT, clickstream, applicazioni mobili e altre forme di dati che fluiscono rapidamente da eventi diversi sono, per loro natura, fugaci. È per questo che per trarre il massimo da questi preziosi data point, è fondamentale servirsi di un sistema di acquisizione rapida. Ora è possibile memorizzare e analizzare fino a 250 miliardi di eventi al giorno con soli tre nodi su soluzioni come IBM Db2 Event Store. Le tecnologie precedenti richiedono fino a 100 nodi per arrivare allo stesso risultato.

In questo modo le aziende traggono due tipi di vantaggi, il più importante dei quali è il costo. La riduzione del numero richiesto di nodi aiuta le aziende ad eliminare i costi legati all'infrastruttura garantendo una maggiore efficienza, così da poter destinare più risorse economiche a progetti più innovativi e significativi. In secondo luogo, ottimizzare ogni singolo nodo fa sì che con lo stesso numero di nodi si possa acquisire una maggiore quantità di dati, in modo da ottenere una visione più affidabile e completa. Si pensi ai dati clickstream; ogni evento aggiuntivo acquisito è un'opportunità per capire e rispondere meglio a un cliente mentre si trova su un sito web, fornendo la possibilità di adattarsi alle sue esigenze e auspicabilmente di aumentare i profitti.



### Incrementare i fast data con l'open source

Molti dei vantaggi dell'open source discussi nella sezione precedente, come la riduzione del vendor lock-in e l'accesso a un maggior numero di dati, si applicano anche quando l'open source è integrato in soluzioni di fast data. Tuttavia, la velocità con cui vengono erogati i dati di eventi e i dati in streaming impone specifiche sfide. Di recente si è visto un incremento nell'utilizzo di soluzioni in grado di scrivere su uno storage di oggetti in formato aperto, come Apache Parquet, per memorizzare grandi quantità di dati in streaming; questo avviene perché tali soluzioni possono essere più veloci e costano meno rispetto ad altri ambienti, come Hadoop. La capacità di acquisire dati di eventi a velocità incredibilmente elevate dipende tanto dal repository quanto dal motore di assimilazione e analisi dei dati.

## Performance di analisi e AI più efficienti

Le soluzioni di fast data più recenti tengono conto dell'ampia gamma di casi d'uso di analisi e di AI che devono essere affrontati, rendendo più semplice la combinazione di dati storici e di dati in streaming per ottenere insight. Per agire su molti dei tipi di dati che le soluzioni di fast data sono costruite per catturare è importante eseguire analisi quasi in tempo reale, è quindi fondamentale che l'assimilazione dei dati non avvenga in modo casuale. Ottenere una maggiore organizzazione attraverso l'ottimizzazione della memoria può aiutare ad utilizzare i dati più facilmente. Un buon esempio è rappresentato dal monitoraggio degli segnali di allarme. I dispositivi IoT che monitorano i sistemi critici inviano grandi volumi di dati in streaming ad alta velocità a scopo di allerta. La capacità di analizzare i dati che sono stati adeguatamente organizzati durante l'assimilazione rende più facile agire quasi istantaneamente, invece di dover aspettare che qualcuno assegni una struttura. Quando scattano segnali di allarme critici ciò che fa la differenza è l'azione istantanea.

Inoltre, la combinazione di dati in streaming in tempo reale con dati storici o dati che sono stati recentemente trasmessi in streaming consente di ottenere un contesto migliore per l'analisi. I dati appena trasmessi in streaming indicano se un data point segue un certo trend o se è un evento isolato. I dati storici possono essere utilizzati per determinare quali data point si verificano tipicamente prima di un guasto che provoca un segnale di allarme. Questa è anche una grande opportunità per introdurre il machine learning al fine di correggere in continuazione quei fattori che provocano segnali di allarme. In questo modo è possibile rendere più accurato un sistema di monitoraggio degli allarmi, risparmiando il tempo normalmente perso per i falsi allarmi e contribuendo al contempo a prevenire guasti critici.

## Autovalutazione sul data management

- Siete in grado attualmente di combinare dati in streaming e dati storici per prendere decisioni quasi in tempo reale?
- Quanti nodi utilizzate attualmente e quanti eventi vi permettono di elaborare?
- La possibilità di utilizzare un formato di dati aperto per l'archiviazione degli oggetti vi consentirebbe di avere la flessibilità necessaria?

# Il recente studio di Forrester Consulting rivela di più sul panorama in rapida evoluzione dei fast data

Forrester Consulting ha recentemente condotto uno studio sui fast data (inclusi i dati in streaming e i dati di eventi) che ne ha ribadito l'importanza nel contesto attuale. Forrester ha intervistato singolarmente un campione di individui per indagare l'incidenza di utilizzo dei fast data, sia in termini di utilizzo attuale che per quanto riguarda le prospettive future. Hanno anche rilevato ostacoli che possono ridurre l'efficacia delle soluzioni di fast data e hanno fornito alcune importanti raccomandazioni.



*“L'88% degli intervistati afferma di dover eseguire analisi in tempo quasi reale su dati in streaming memorizzati”.*

Non fatevi trovare impreparati dai fast data.

[Leggi lo studio di Forrester Consulting →](#)

88%



## Conclusione

# Come iniziare il processo di modernizzazione del data management

La dipendenza delle aziende dai dati aumenta ogni giorno di più. I dati rappresentano una risorsa in rapida crescita, con sempre più fonti e tipi diversi, e comportano opportunità e sfide. Ecco perché è fondamentale disporre di un'infrastruttura di data management ibrida che risponda a queste esigenze e si evolva con esse.

Per ulteriori informazioni su IBM Hybrid Data Management, contattare il rappresentante IBM o il Business Partner IBM. Puoi anche [iniziare una prova gratuita di IBM Cloud Pak for Data](#).

Se avete domande relative a questo eBook, alla gestione ibrida dei dati, o a IBM Cloud Pak for Data, i nostri esperti sono disponibili per una consulenza gratuita di 30 minuti per rispondere alle vostre domande.

**Chiedi all'esperto** →

### **IBM HDM e IBM Cloud Pak for Data fanno la differenza**

È fondamentale costruire un'architettura informatica orientata al futuro, incorporando gli strumenti di AI, ML e data science necessari per continuare ad eccellere in un ambiente competitivo, offrendo al contempo la possibilità di implementazione on-premise, su un cloud privato o su vari fornitori di cloud pubblici. È inoltre essenziale un'ampia gamma di offerte per il data management, tra cui database, data warehouse, data lake e funzionalità di fast data.

IBM Hybrid Data Management fornisce le funzionalità di AI, le opzioni di implementazione e la gamma di soluzioni necessarie. IBM Db2 è alimentato dall'AI e progettato per l'AI, migliora le prestazioni e fornisce al contempo gli strumenti necessari per costruire applicazioni di AI. Il motore comune SQL su cui sono costruiti i prodotti IBM Db2 fornisce funzionalità di integrazione ottimizzate e la possibilità di scrivere una sola volta per poi funzionare ovunque.

I prodotti IBM Hybrid Data Management sono disponibili come parte di IBM Cloud Pak for Data. Eseguendo soluzioni in un container Kubernetes con Cloud Pak for Data, queste soluzioni diventano indipendenti dall'hardware e dal cloud, in grado di funzionare ovunque vengano supportati rispettivamente i container Linux o RedHat. Come ulteriore vantaggio, in Cloud Pak for Data è inclusa la virtualizzazione dei dati, consentendo di eseguire i carichi di lavoro in modo più efficiente senza dover spostare i dati.

Con Cloud Pak for Data anche aumentare la capacità è semplice. Basta aggiungere ulteriori VPC per accedere a funzionalità aggiuntive. Questo va molto oltre l'estensione della gestione dei dati ibridi e le offerte alla base di soluzioni come IBM Watson Studio Open Scale; in questo modo tutte le componenti della scala dell'AI rimangono collegate, consentendo ai dati di essere automaticamente gestiti e utilizzati senza soluzione di continuità nei progetti di AI.





© Copyright IBM Corporation 2020

IBM Italia S.p.A.  
Circonvallazione Idroscalo  
20090 Segrate (MI)  
Italia

Prodotto negli Stati Uniti d'America  
Marzo 2020

IBM, il logo IBM, **ibm.com**, Db2 e BLU Acceleration sono marchi della International Business Machines Corp., registrata in varie giurisdizioni in tutto il mondo. I nomi di altri prodotti e servizi possono essere marchi di IBM o di altre aziende. L'elenco aggiornato dei marchi IBM è disponibile nella sezione "Copyright and trademark information" all'indirizzo web [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Il presente documento è aggiornato alla data di pubblicazione iniziale e può essere modificato da IBM in qualsiasi momento. Non tutti i prodotti e servizi sono disponibili in tutti i paesi in cui IBM opera.

I dati sulle performance e gli esempi relativi ai clienti menzionati nel presente documento sono riportati solo per scopi illustrativi. I risultati effettivi delle prestazioni possono variare a seconda delle configurazioni specifiche e delle condizioni operative.

LE INFORMAZIONI CONTENUTE NEL PRESENTE DOCUMENTO SONO FORNITE "AS IS", SENZA GARANZIE, ESPLICITE O IMPLICITE, DI QUALSIASI TIPO, IVI INCLUSE LA GARANZIA IMPLICITA DI COMMERCIALIZZABILITÀ O DI IDONEITÀ A UNO SCOPO PARTICOLARE E LA GARANZIA O CONDIZIONE DI NON VIOLAZIONE. I prodotti IBM sono garantiti conformemente ai termini e alle condizioni dei contratti in virtù dei quali vengono forniti.