

Integrazione e governance dei Big Data

*I big data possono costituire un ostacolo all'integrazione?
No, se si seguono le best practice illustrate di seguito*



1

Introduzione

2

**Requisiti di
integrazione
e governance
per i Big Data**

3

**Best practice:
Integrazione
e governance
efficace dei
Big Data**

4

**IBM InfoSphere
offre il livello di
affidabilità
necessario per
sfruttare i Big Data**

5

Perché InfoSphere?

Introduzione

I business leader ambiscono a sfruttare la potenza dei Big Data. Tuttavia, assicurare l'affidabilità e la protezione delle fonti delle informazioni è diventato esponenzialmente più complesso. Se il problema legato all'attendibilità dei dati non viene risolto direttamente, gli utenti finali possono perdere fiducia negli insight generati dai dati, con conseguente impossibilità di cogliere le opportunità e prevenire le minacce.

Per ottenere il massimo vantaggio dai Big Data, il punto di partenza deve essere l'affidabilità dei dati. Tuttavia, il volume e la complessità dei Big Data sono talmente elevati che i metodi manuali tradizionali di discovery, governance e correzione delle informazioni non sono più fattibili. L'integrazione e la governance delle informazioni devono essere implementate all'interno delle applicazioni Big Data, fornendo, sin dall'inizio, una governance adeguata e un'integrazione rapida.

Automatizzando l'integrazione e la governance delle informazioni e implementandole nella fase di creazione dei dati, le aziende possono aumentare notevolmente l'affidabilità dei Big Data.

Un programma di integrazione e governance solido deve includere la discovery, la profilazione e la comprensione di set di dati differenti per fornire un contesto e consentire ai dipendenti di prendere decisioni informate. Questo programma deve essere agile per accogliere una vasta gamma di dati e deve integrarsi perfettamente con le diverse tecnologie, dai data mart ai sistemi Apache Hadoop. Inoltre, deve eseguire automaticamente la discovery, la protezione e il monitoraggio delle informazioni sensibili, come parte delle funzioni delle applicazioni per i Big Data.

I Big Data sono un fenomeno, non una tecnologia

Con tutto il clamore che ruota intorno ai Big Data, è facile essere indotti a pensare che i Big Data rappresentino la panacea di tutti i mali. I Big Data, tuttavia, non sono una tecnologia, ma un fenomeno. Per sfruttarli in modo efficace, bisogna essere in grado di integrare e governare i dati chiave nell'intero ambito aziendale.

Requisiti di integrazione e governance per i Big Data

Quando si parla di Big Data, le discussioni vertono spesso sull'analytics e Hadoop. È interessante notare che la Big Data Analytics si è recentemente spostata verso i dati strutturati, allontanandosi dalle sue origini che risiedevano nei dati non strutturati. Ma se, da un lato, analytics e Hadoop sono importanti per i dati sia strutturati che non, dall'altro rappresentano solo un tassello del puzzle che compone i Big Data.

I professionisti lungimiranti dell'Information Technology (IT) ora si rendono conto che il fenomeno dei Big Data sta avendo un impatto su tutti i sistemi, creando una nuova serie di requisiti che incidono sui risultati delle iniziative di data warehousing, Big Data e analytics. Per

garantire i migliori risultati, i dati provenienti dalle fonti di Big Data devono essere integrati, governati e affidabili.

In realtà, molte delle sfide più comuni associate ai Big Data non sono legate a problemi di analytics. In molti casi, si tratta di problemi di integrazione dati fondamentali, persino "tradizionali", e possono essere evitati o affrontati con una soluzione per l'integrazione e la governance dei dati agile e di classe enterprise.

Inoltre, le nuove fonti di Big Data sono inutili se risiedono in silos e devono essere integrate nell'architettura aziendale. Le soluzioni migliori costituiscono una base solida e integrata

che facilita l'attività di analytics, producendo informazioni aziendali preziose e fruibili.

Le soluzioni appropriate per l'integrazione e la governance dei Big Data devono:

- 1. Essere agili**
- 2. Essere costruite su un'architettura a prestazioni elevate e scalabile**
- 3. Sostenere un'efficienza maggiore**
- 4. Contribuire a creare attendibilità sull'autenticità dei dati**
- 5. Soddisfare le esigenze di flessibilità e agilità di delivery dei dati.**

Best practice: Integrazione e governance efficace dei Big Data

Alcune best practice per l'integrazione e governance possono consentirvi di trarre il massimo valore dai Big Data aziendali.

Introdurre l'agilità nell'IT per migliorare performance e scalabilità

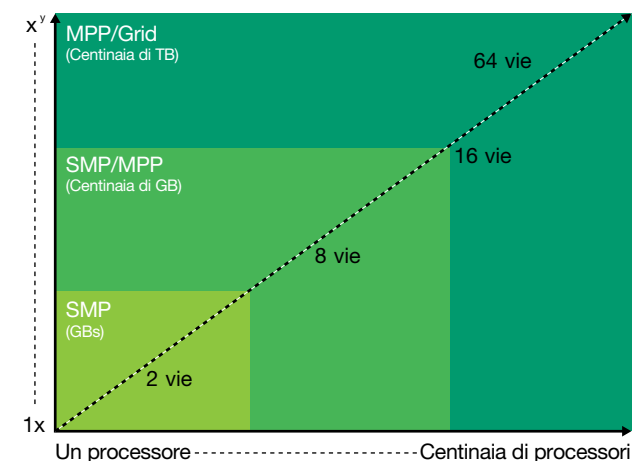
I Big Data affluiscono ad alta velocità, per questo la performance è fondamentale. I dati cambiano velocemente e devono afferire a diverse applicazioni del sistema in tempi rapidi, in modo che i business leader possano reagire alle mutevoli condizioni di mercato il più presto possibile.

Per gestire i Big Data con successo, le aziende hanno bisogno di una soluzione di integrazione dei dati di classe enterprise che sia:

- **Dinamica** per soddisfare i requisiti di performance attuali e futuri
- **Estendibile e partizionata** per una scalabilità facile e veloce
- **Integrata con Hadoop.** Hadoop di per sé non è una piattaforma di integrazione, ma può essere sfruttato come componente di un'architettura di destinazione dei dati per determinarne il valore in un'ottica di ottimizzazione equilibrata.

La scalabilità è uno dei requisiti più complessi di integrazione dei Big Data visto che i requisiti di business possono evolvere molto rapidamente. Di conseguenza, nell'affrontare l'integrazione dei Big Data, è importante disporre di un prodotto in grado di garantire la scalabilità dei dati su tutte le architetture con la stessa funzione e con un aumento lineare delle velocità, scalando senza problemi.

Scalabilità dei dati nelle architetture hardware



- Stessa funzionalità
- Tutte le architetture
- Accelerazione lineare
- Decisione (in fase di esecuzione) di scalare di "N vie"

Lavorare meglio e meno, riducendo i costi

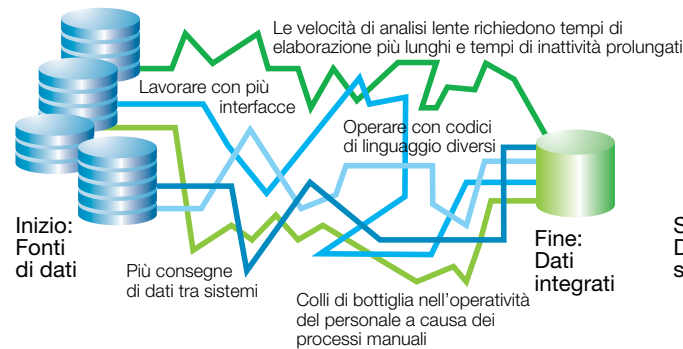
Il tempo dei dipendenti è una risorsa preziosa e costosa. Una soluzione di integrazione per i Big Data che supporti la produttività e l'efficienza dei dipendenti consente di migliorare i profitti dell'impresa, eliminando i colli di bottiglia e aumentando l'agilità.

Per le divisioni IT, gli SLA (Service Level Agreement) risentono spesso delle inefficienze. Mano a mano che i volumi, la varietà, la velocità e la veridicità dei dati aumentano, il tempo necessario per l'elaborazione dei processi di integrazione dei dati supera spesso la finestra consentita dagli SLA, il che significa che non è più conforme alle esigenze dei clienti interni.

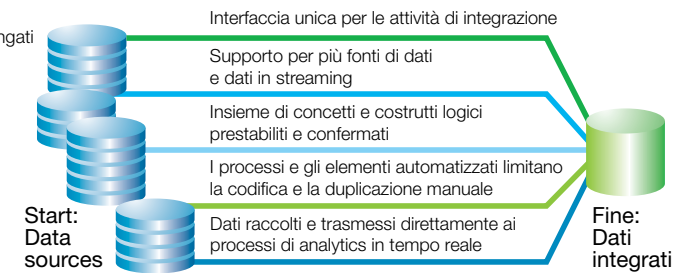
Per migliorare la produttività, è importante creare una logica progettuale dedicata alle iniziative di integrazione dei dati orientate ad Hadoop, utilizzando la stessa interfaccia, concetti e costrutti logici comuni a qualsiasi

altro metodo di implementazione. Questo approccio elimina la necessità di imparare i nuovi linguaggi di codificazione mano a mano che evolvono e di eseguire attività manuali di codifica e duplicazione.

Lavorare di più



Lavorare in modo più produttivo



Per i progetti di Big Data focalizzati sull'elaborazione analitica in tempo reale, è anche fondamentale favorire l'integrazione, in modo semplice e rapido, con i sistemi che supportano i dati di streaming (anche noti come "dati in movimento"). Le soluzioni di integrazione dei Big Data devono essere sufficientemente 'intelligenti' da consentire alle convenzioni di integrazione dei dati standard di acquisire e trasferire i dati direttamente ai processi di analytics in tempo reale.

Creare affidabilità con dati puntuali e accurati

Le aziende di solito gestiscono i Big Data al fine di aumentare e migliorare le proprie capacità di analytics, sia attraverso l'analisi di nuove fonti di dati che affrontando volumi più elevati di dati, obiettivi entrambi non raggiungibili con le tecnologie tradizionali. **Tuttavia, la qualità**



degli insight di analytics equivale alla validità dei dati sottostanti. Se le aziende non alimentano i propri sistemi di analytics con dati di qualità, gli insight che ottengono non sono validi.

Senza la capacità di concordare e sfruttare definizioni comuni per i termini di business, le aziende semplicemente non possono essere reattive e adattabili. Se i reparti dispongono di definizioni incoerenti per i termini chiave,

le decisioni non possono essere prese con la velocità e la precisione necessarie. Ad esempio, cosa succede quando il Marketing richiede i dati del "cliente" ai fini dell'analisi, ma riceve solo un sottoinsieme dei dati effettivamente necessari per prendere una decisione, perché il team IT ha definito "cliente" un nucleo familiare invece che un individuo?

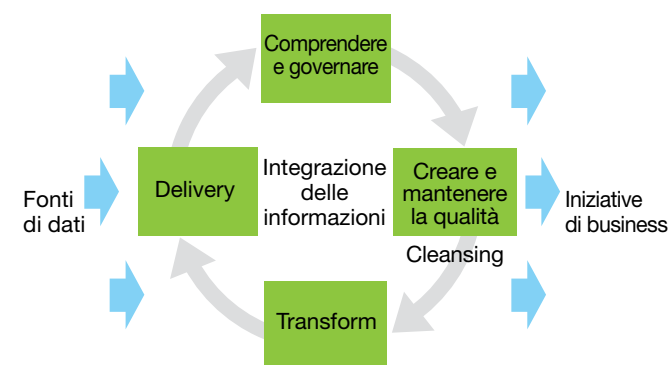
Purtroppo, non basta semplicemente stabilire le definizioni e le politiche relative alle informazioni e poi sperare che le persone seguano le regole. Per essere certi che i propri dati siano affidabili, le aziende devono essere in grado di tracciarne il percorso attraverso i propri sistemi, in modo da poter vedere da dove provengono e in che modo sono stati manipolati. È importante disporre di una soluzione di integrazione di Big Data in grado di supportare questo livello di trasparenza.

Per garantire dati di alta qualità, è anche fondamentale disporre di funzionalità di analisi delle informazioni che consentano ai data steward di verificare la qualità dei dati. Ad esempio, i data steward possono effettuare

una semplice verifica dei null per garantire che tutti i campi e le tabelle che stanno analizzando contengano effettivamente dei dati. In un altro scenario, potrebbero eseguire i dati a fronte di sofisticati algoritmi per determinarne la validità. Queste informazioni raggiungono la massima utilità se visualizzate in una dashboard che consente ai business analyst di identificare rapidamente eventuali problemi e di scendere facilmente nel dettaglio.

È importante applicare il data cleansing a qualsiasi tipo di Big Data si desideri conservare, in modo da poterne stabilire l'attendibilità. La fiducia nella qualità dei dati consente l'attendibilità dei risultati.

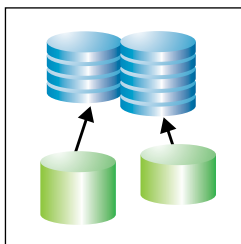
L'applicazione del data cleansing nel flusso di lavoro di integrazione e governance



Il data cleansing, nell'ambito del ciclo di integrazione delle informazioni contribuisce a garantire la qualità dei dati durante il processo.

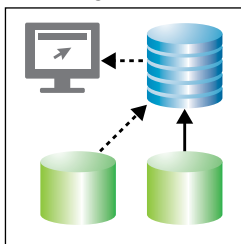
Fornire i dati in modo appropriato

Nell'affrontare i progetti di integrazione dei Big Data, si punta a raggiungere prestazioni e scalabilità elevate per l'elaborazione dei dati in tempo reale e per il trasferimento di dati bulk o in batch. In molti casi, le organizzazioni devono anche sfruttare la duplicazione o virtualizzazione dei dati come parte della soluzione più ampia di integrazione dei dati. Questo vale sia per l'integrazione dei dati tradizionali che dei Big Data. **Di seguito sono riportate alcune soluzioni valide per la delivery dei dati che possono essere implementate su piattaforme per Big Data:**



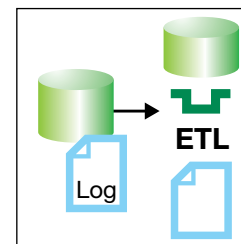
IBM InfoSphere Information Server for Data Integration

Delivery di dati bulk ad alta velocità, inclusi ETL (Extract-Transform-Load), ELT (Extract-Load-Transform) e integrazione dinamica che sfruttano Hadoop per supportare lo scambio di informazioni con le fonti di Big Data.



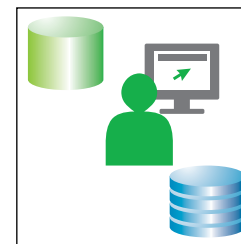
IBM InfoSphere Federation Server

L'accesso virtualizzato da e per la delivery di informazioni diverse e distribuite permette il consolidamento virtuale sia di Big Data che di dati tradizionali.



IBM InfoSphere Data Replication

L'integrazione in tempo reale fornisce la flessibilità necessaria per l'integrità delle transazioni e in più la duplicazione di volumi elevati a bassa latenza per la disponibilità continua del business.



IBM InfoSphere Data Click

L'integrazione self-service dei dati permette agli utenti delle line-of-business e ad altri utenti non tecnici di ottenere informazioni in caso di necessità per alimentare l'analytics.

Sfruttare la duplicazione dei dati

Mano a mano che la quantità e la varietà dei dati presenti nell'ambiente cresce, mantenere pool fisici di dati diventa meno pratico. Per rimanere flessibili e agili nel mondo dei Big Data, le aziende devono sfruttare diverse tecnologie, tra cui la delivery incrementale di dati, per assicurarsi di disporre dei dati di cui hanno bisogno. La trasformazione dei dati e i requisiti di delivery si sono estesi, dal movimento di dati bulk o in batch, fino ad includere anche il trasferimento dei dati in tempo reale sulla base di funzionalità di duplicazione dei dati, in particolare attraverso il Change Data Capture (CDC). Mentre lo spostamento dei dati bulk e in batch avviene relativamente di rado, la trasmissione dei dati in tempo reale si verifica ogni volta che i dati cambiano alla fonte. I dati modificati vengono acquisiti, trasferiti, trasformati e quindi caricati nella sede di destinazione.

I fattori che influenzano le prestazioni e la scalabilità delle trasformazioni dei dati in tempo reale sono tre:

1. L'approccio utilizzato per acquisire una modifica alla fonte.

L'opzione più flessibile ed efficiente per acquisire le modifiche alla fonte consiste nel disporre di un meccanismo CDC che "spinga" le modifiche mano a mano che i dati scorrono. Non appena i dati sorgente vengono modificati, il meccanismo viene a conoscenza della modifica e inoltra i dati

2. Il meccanismo utilizzato. Per il CDC possono essere utilizzati molti meccanismi. Se correttamente implementato, un approccio di acquisizione basato sui log ha spesso un impatto inferiore sul database di origine, migliorando, di conseguenza, la performance complessiva

3. Persistenza temporanea dei dati. Anche la persistenza temporanea dei dati esercita un impatto sulle prestazioni del CDC. Idealmente, l'organizzazione sarebbe in grado di trasmettere le modifiche senza che persistano per aumentare la performance (in quanto i dati non devono essere scritti sul disco al quale poi accede un motore di trasformazione).

Per ulteriori informazioni su come ottenere insight in tempo reale dai Big Data utilizzando la duplicazione dei dati, scarica il [White Paper IBM](#).

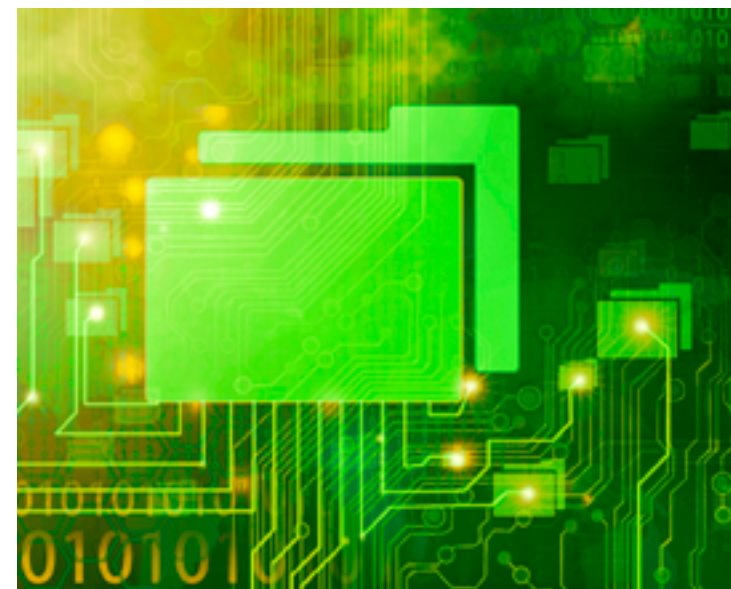
Virtualizzare i dati

Dato il forte aumento del volume, della varietà, della velocità e della veridicità dei dati, la domanda di accesso ai dati ha assunto un'importanza senza precedenti. Le tecnologie di virtualizzazione dei dati possono contribuire a creare il pool di dati necessario per sostenere il vostro business.

La virtualizzazione dei dati è incentrata sulla semplificazione dell'accesso ai dati, isolando i dettagli di storage e recupero e rendendo il processo trasparente per gli utenti.

In questo modo, la virtualizzazione dei dati riduce il tempo necessario per usufruire di dati disparati, rendendo più semplice per utenti e processi ottenere le informazioni richieste, in modo tempestivo.

Esistono due strategie principali per la virtualizzazione dei dati: la federazione dei dati e i data services. In entrambi i casi, i dati vengono esposti per essere resi più consumabili, accessibili e riutilizzabili da utenti, clienti o processi di business dell'azienda.



IBM InfoSphere offre il livello di affidabilità necessario per sfruttare i Big Data

Mentre il termine 'Big Data' è entrato in voga da poco, IBM progetta soluzioni in grado di gestire enormi volumi di dati da decenni. L'azienda ha da tempo spianato la strada con soluzioni di integrazione, gestione, sicurezza e analytics note per la loro affidabilità, flessibilità e scalabilità.

Le funzionalità di integrazione dati end-to-end di IBM InfoSphere Information Server sono progettate per consentire di comprendere, pulire, controllare, trasformare e distribuire i

dati, oltre che per contribuire a colmare il divario tra business e IT. InfoSphere Information Server consente di essere sicuri che le informazioni che guidano il vostro business e le vostre iniziative strategiche, dai big data, alla point-of-impact analytics al master data management e al data warehousing, siano attendibili, coerenti e governate in tempo reale. Infatti, InfoSphere Information Server è di 10-15 volte più veloce di Hadoop per l'integrazione dei dati.¹

Essere veloci e agili

Le organizzazioni che lavorano con i Big Data hanno bisogno di una scalabilità dei dati illimitata garantita dal proprio software di integrazione. Il Software InfoSphere è stato progettato da zero per ottimizzare l'utilizzo delle risorse hardware, consentendo la massima quantità dei dati da elaborare per nodo. Dispone di funzionalità potenti di trasformazione e delivery dei dati, consentendo ai clienti di elaborare su sistemi massively parallel, eliminando i colli di bottiglia e migliorando notevolmente il time-to-value.

Maggiori informazioni su InfoSphere Information Server

Desideri ottenere ulteriori informazioni sulle funzionalità di InfoSphere Information Server che consentono di sostenere l'integrazione agile, la governance orientata al business e la qualità sostenibile? Per il materiale informativo di riferimento consulta:

ibm.com/software/data/integration/info_server/demo.html

1 Introduzione

2 Requisiti di integrazione e governance per i Big Data

3 Best practice: Integrazione e governance efficace dei Big Data

4 IBM InfoSphere offre il livello di affidabilità necessario per sfruttare i Big Data

5 Perché InfoSphere?

La University of Arizona accelera l'accesso ai dati con InfoSphere Information Server

Con oltre 38.000 studenti e docenti, l'infrastruttura della University of Arizona supporta un carico di dati molto elevato. Per rimanere competitiva, aveva bisogno di sostituire i vecchi sistemi informatici amministrativi che non erano in grado di gestire la domanda di informazioni di Business Intelligence (BI). Secondo Manav Mehra, Senior Manager of Information Integration addetta all'integrazione delle informazioni aziendali e all'analytics presso la University of Arizona, l'organizzazione puntava ad ottenere un'unica fonte di dati che consentisse agli utenti di eseguire facilmente le query in base alle proprie esigenze e di ottenere risultati in modo tempestivo.

Il BI Team dell'Università ha utilizzato InfoSphere Information Server per realizzare questa singola fonte di dati attendibile; il team ha impiegato il software per comprendere, ripulire, trasformare e distribuire i dati dai sistemi di origine nel proprio

data warehouse aziendale. La soluzione include strumenti che consentono al personale di BI di:

- Scoprire, modellare, visualizzare, relazionare e standardizzare data set diversi e distribuiti
- Acquisire e definire i requisiti di business in un formato familiare comune per sostenere lo sviluppo delle operazioni di estrazione, trasformazione e caricamento (ETL)
- Ottenere insight sull'analisi di fonti di dati, processi ETL, regole di data quality, terminologia di business, data model e report BI.

“Secondo Mehra: In media, il software InfoSphere Information Server ci consente di risparmiare circa sei ore a sviluppatore in termini di modellazione dei dati e creazione di operazioni ETL. Due studenti laureati del nostro dipartimento MIS ci hanno aiutato a creare job ETL e sono stati in grado di realizzare qualcosa come 9.000 operazioni ETL da un template in tre mesi. A mio avviso,



tuttavia, il fattore ancora più importante è la quantità di tempo che consente di risparmiare per trovare e risolvere i problemi legati ai dati”.

Mehra ha affermato che il team può eseguire più di 22.000 job ETL notturni in 2,5 ore rispetto alle 9 ore necessarie prima dell'introduzione di InfoSphere Information Server. Inoltre, nei sei mesi successivi all'implementazione, l'uso dell'enterprise data warehouse è notevolmente aumentato, segno che gli utenti stanno trovando le informazioni di cui hanno bisogno.

[Per maggiori informazioni sull'esperienza dell'Università fai clic qui.](#)



InfoSphere Information Server for Data Integration

fondamentale quando si tratta di aumentare la precisione e l'efficienza. Guarda questo video e scopri come InfoSphere Information Server consente di riunire le fonti di dati. Scarica il video ibm.co/13jL5mr

Essere efficienti

InfoSphere Information Server include funzionalità che consentono di ottimizzare il tempo di lavoro dei dipendenti. Ad esempio, la Versione 9.1 include InfoSphere Data Click che semplifica notevolmente l'integrazione e il provisioning self-service dei dati. Di conseguenza, il personale delle line-of-business può eseguire queste operazioni autonomamente, mentre gli ingegneri IT specializzati si concentrano sulle iniziative a maggior valore.

InfoSphere Information Server consente inoltre di risparmiare tempo per gli sviluppatori, fornendo un'unica palette di design in un ambiente applicativo condiviso. Gli sviluppatori non devono passare tra diverse interfacce, dal momento che tutto ciò che serve è facilmente accessibile. Inoltre, ogni componente di InfoSphere Information Server utilizza lo stesso

livello di metadati. Ciò consente di tenere traccia dell'avanzamento dei job e di diagnosticare rapidamente i problemi. E' prevista anche una dashboard che fornisce una visione unificata dell'ambiente.

Mano a mano che gli archivi dei Big Data continuano a crescere, queste funzionalità a prestazioni elevate che consentono di risparmiare tempo diventano ancora più importanti. Per le divisioni IT, possono fare la differenza nella conformità o meno con gli SLA, nell'aver tempo per lavorare su nuovi progetti innovativi o dedicarsi solo ad iniziative di scarso valore legate alla gestione dei sistemi esistenti. Per il business, questo può significare un processo decisionale più rapido e informato che porta a profitti più solidi, un servizio migliore per i clienti e un vantaggio competitivo.

Essere fiduciosi

Molte imprese hanno migliorato la qualità dei dati attraverso l'implementazione della governance dei dati. Idealmente, un'iniziativa di governance dei dati comprende tre funzioni: definizione dei termini, cleansing dei dati esistenti e monitoraggio della qualità dei dati.

Per aiutare il personale di tutta l'organizzazione a raggiungere una comprensione condivisa dei termini chiave, InfoSphere Information Server fornisce un glossario dei dati che consente a business e IT di creare e concordare definizioni, regole e politiche. Inoltre sono incluse le funzionalità di data modelling, grazie alle quali i data architect possono stabilire la fonte e la destinazione di ogni singolo dato. Questi

strumenti consentono alle organizzazioni di stabilire la "verità", almeno per quanto riguarda i dati aziendali.

Per essere veramente sicuri che i dati siano affidabili, tuttavia, le organizzazioni devono anche essere in grado di tracciare il percorso dei dati attraverso i propri sistemi. Per sostenere questo livello di trasparenza, InfoSphere Information Server fornisce metadati e funzionalità di lineage che consentono agli utenti di tenere traccia dei dati sin dalla fonte originaria e di visualizzare i calcoli eseguiti su di essi lungo tutto il percorso.

InfoSphere Information Server offre inoltre funzionalità di data quality per supportare il data cleansing e per monitorare la qualità su base costante. Le funzionalità di cleansing includono strumenti sofisticati per funzioni quali investigation, standardizzazione, matching e survivorship, consentendo ai data steward di risolvere gli eventuali problemi che emergono in corso di analisi. Ad esempio, i nomi dovrebbero essere abbinati automaticamente, in modo che 'William Smith' e 'Bill Smith' siano elencati come un unico cliente.

InfoSphere Information Server per la qualità dei dati in azione

Ottieni maggiori informazioni sulle quattro fasi di cleansing e standardizzazione dei dati e scopri come InfoSphere Information Server for Data Quality le riunisce tutte. Scarica il video:

ibm.co/17yIBnC

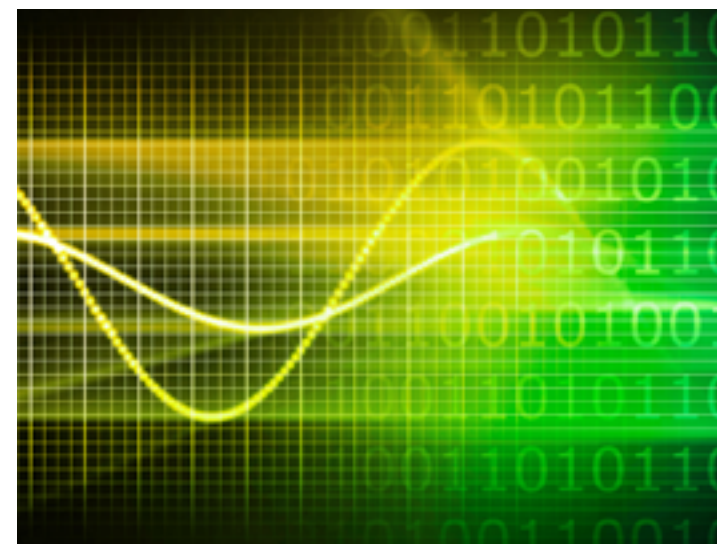
Essere flessibili

In molti casi, la 'velocità' non è sufficiente. Per fornire l'integrazione in tempo reale è necessaria la flessibilità e non solo la velocità. Un modo per ottenere i dati è quello di eseguire query sui database o sulle applicazioni nei quali risiedono. Questo approccio, tuttavia, può rallentare i sistemi transazionali a livelli poco accettabili.

Un approccio migliore: utilizzare una soluzione, come ad esempio InfoSphere Data Replication, che acquisisce rapidamente i dati in continuo cambiamento e li invia alle rispettive destinazioni, fornendo ai business manager una visione aggiornata al secondo delle informazioni cruciali, senza rallentare i processi business-critical. InfoSphere Data Replication utilizza un meccanismo CDC di 'spinta' mano a

mano che i dati fluiscono per fornire flessibilità ed efficienza. Impiega anche l'acquisizione basata sui log per ridurre l'impatto sul database di origine ed esegue lo streaming dei dati modificati senza la persistenza dei dati per migliorare la performance.

A seconda dei requisiti di integrazione dei Big Data, anche la data federation può rispondere alle vostre esigenze associate ai Big Data. IBM InfoSphere Federation Server crea rapidamente una visione consolidata dei dati per supportare i processi e le decisioni di business. È possibile accedere e integrare le fonti di dati e di contenuti differenti, come se fossero un'unica risorsa, indipendentemente da dove risiedono realmente le informazioni.



Le quattro V dei Big Data

In che modo è possibile gestire il volume, la velocità, la varietà e la veridicità dei Big Data? InfoSphere Data Replication offre le funzionalità di near-real-time necessarie per supportare le offerte di prodotti, fornire dati affidabili e molto altro ancora. Scarica il video e scopri maggiori dettagli sulla potenza della flessibilità: ibm.co/11cy27N

Perché InfoSphere?

Come elemento critico di IBM Watson Foundations, la piattaforma per i Big Data e l'analytics di IBM, InfoSphere Information Integration and Governance (IIG) fornisce funzionalità leader di mercato per affrontare le sfide associate ai Big Data. InfoSphere IIG fornisce scalabilità e prestazioni ottimali per enormi volumi di dati, un'integrazione agile e proporzionata, la governance per l'aumento della velocità dei dati e supporto e protezione per un'ampia gamma di tipi di dati e sistemi di Big Data. InfoSphere IIG contribuisce al successo dei progetti di Big Data e analytics fornendo ai business user il livello di confidenza necessario per sfruttare gli insight.

Le funzionalità InfoSphere includono:

- **Metadati, glossario di business e gestione delle policy:** Definizione di metadati, terminologia di business e politiche di governance con IBM InfoSphere Information Governance Catalog
- **Integrazione dei dati:** Gestione di tutti i requisiti di integrazione, incluse funzionalità di trasformazione e trasferimento dei dati in batch (InfoSphere Information Server), duplicazione in tempo reale (InfoSphere Data Replication) e data federation (InfoSphere Federation Server)
- **Qualità dei dati:** Funzionalità di parsing, standardizzare, validazione e matching dei dati aziendali con IBM InfoSphere Information Server for Data Quality
- **Master data management (MDM):** Utilizzare una visione attendibile dei vostri clienti, prodotti, fornitori, sedi e account con InfoSphere MDM
- **Gestione del ciclo di vita dei dati:** Gestire il ciclo di vita dei dati, dalla creazione al data retirement e all'archiviazione con IBM InfoSphere Optim
- **Sicurezza dei dati e privacy:** Monitorare costantemente l'accesso ai dati, proteggere i repository dalle violazioni e supportare la conformità con IBM InfoSphere Guardium. Assicurarsi che i dati sensibili siano mascherati e protetti con InfoSphere Optim.



Risorse aggiuntive

Per maggiori informazioni sul metodo IBM per l'integrazione delle informazioni e la governance dei Big Data, contattare il proprio rappresentante o Business Partner IBM o consultare le seguenti risorse:

- ibm.com/software/data/information-integration-governance
- ibm.com/software/data/infosphere/information-integration-big-data
- ibm.com/software/data/integration/info_server
- [InfoSphere Information Server: A Forrester Total Economic Impact Study](#)
- [Delivering Trusted Information for Big Data and Data Warehousing: A Ventana Research Report](#)
- [Gartner: Hadoop Is Not a Data Integration Solution](#)
- [ITG: Business Case for Enterprise Data Integration Strategy: Comparing IBM InfoSphere Information Server and Open Source Tools](#)



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

IBM, il logo IBM, ibm.com, Guardium, IBM Watson, InfoSphere e Optim 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 all'indirizzo web 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.

Gli esempi di clienti citati sono forniti ai soli fini illustrativi. I risultati di performance effettivi possono variare in base a configurazioni e situazioni operative specifiche. E' responsabilità dell'utente valutare e verificare il funzionamento di qualsiasi altro prodotto o programma con prodotti e programmi IBM. LE INFORMAZIONI CONTENUTE IN QUESTO DOCUMENTO SONO FORNITE "COSI' COME SONO", SENZA GARANZIA DI ALCUN TIPO, ESPLICITA O IMPLICITA, IVI INCLUSA LA GARANZIA DI COMMERCIALIZZABILITA' O DI IDONEITA' AD UNO SCOPO PARTICOLARE O DI NON VIOLAZIONE. I prodotti IBM sono garantiti conformemente ai termini e alle condizioni dei contratti in virtù dei quali vengono forniti

È responsabilità del cliente osservare le disposizioni di legge e i regolamenti applicabili. IBM non fornisce consulenza in materia legale, né dichiara o garantisce che i propri prodotti e servizi assicurino che il cliente sia in conformità a qualunque disposizione di legge.

¹ Test interni IBM.

© Copyright IBM Corporation 2015



Si prega di riciclare