

Tutto quello che è necessario sapere sulla gestione strategica dei dati



Avete letto abbastanza da sapere che la gestione del dato è un primo passo importante nelle iniziative di Big Data o tra i prerequisiti di qualsiasi progetto analitico.

Non siete, però, così orgogliosi da non ammettere ancora un briciolo di confusione sulle differenze tra Master Data Management e Data Federation. O magari conoscete perfettamente tutti i concetti ed ogni singolo termine, eppure continuate a ripeterli e spiegarli ai vostri utenti oppure al vostro capo, e ogni volta avete la sensazione che non sarà l'ultima.

In ogni caso, abbiamo messo insieme tutte le informazioni sul Data Management in un linguaggio semplice, in modo che chiunque possa capire quello che fate. Non dimenticate però che una buona strategia sui dati non si non si deve mai concentrare su uno solo di questi aspetti. Ma è necessario considerarli tutti insieme, in modo combinato.

Data Access

Cos'è? Il dato è un "asset", lo strumento per arrivare alla conoscenza. L'accesso al dato si riferisce alla capacità di una organizzazione di recuperare informazioni da qualsiasi fonte.

Le tecnologie di Data Access, quali i driver di database o i convertitori di documenti, servono per facilitare e rendere efficiente questa operazione, in modo da impiegare il tempo nell'utilizzo effettivo dei dati senza perdere tempo a cercarli.

Perché è importante? I dati potenzialmente importanti per una organizzazione possono risiedere in luoghi diversi e in vari formati: fogli di calcolo, file di testo, database, email, applicazioni, pagine web/feeds dei social media sono solo alcuni esempi. Senza tecnologie, recuperare informazioni utili può diventare un incubo. Ogni volta. Come doversi districare in una vasta biblioteca, con libri in lingua diversa, senza istruzioni, mappe, e senza nessuno cui potersi rivolgere per chiedere aiuto.

Una valida tecnologia di Data Access elimina il lavoro impiegato a valutare il formato del dato, il tipo di storage, i protocolli di comunicazione, e così via, per ogni possibile fonte a cui accedere. Maggiore è la capacità di ottenere rapidamente nuove tipologie di informazioni, più completo ed accurato sarà l'universo informativo su cui basare il processo decisionale.

Data Quality

Cos'è? E' la pratica che garantisce che il dato sia accurato e utilizzabile per lo scopo previsto. Proprio come nella gestione della qualità ISO 9000 nei processi di produzione, gli strumenti di data quality andrebbero sfruttati in ogni fase del ciclo di produzione e distribuzione dei dati. Tale ciclo inizia nel momento in cui il dato "entra" in azienda, o vi si accede attraverso flussi di caricamento dei datawarehouse, considera i vari punti di integrazione o scambio/arricchimento dei dati, fino al punto immediatamente precedente a dove il dato sarà caricato all'interno dei sistemi di destinazione.

Perché è importante? La qualità dei dati è un elemento imprescindibile per aumentarne l'efficacia nei processi aziendali. Per sfruttare appieno il loro valore, è necessario "codificare" le aspettative e automatizzarne le logiche di realizzazione. Ad esempio, individuare automaticamente il nome di un farmaco all'interno del testo di una ricetta, derivarne il principio attivo, il formato (capsule o soluzione solubile?), e codificare queste informazioni in modo univoco e standard – nella stessa lingua – per tutte le applicazioni e tutti i processi. Quanto è più sfruttabile un dato simile? E quindi il suo valore?

Data Integration

Cos'è? Come si procede una volta che si ha l'accesso ai dati? Il passo successivo piuttosto comune è quello di combinarli con gli altri dati a disposizione per unificare i risultati. La data integration è il processo che definisce i vari passaggi, e gli strumenti di data integration permettono di "disegnare" questi processi, una fase alla volta, e di automatizzarne l'esecuzione. I più comuni strumenti di data integration sono noti come ETL (Extract, Transform, Load) oppure ELT (Extract, Load & Transform). La differenza negli acronimi è in "chi" – quale server - effettua la fase di transform: il server di data integration oppure il sistema di database finale (carico tutti i dati massivamente sul DB e poi li trasformo)?

Oggi l'integrazione dei dati non si limita a movimentarli tra i diversi database, ma può sfruttare anche le nuove tecnologie, come i server in-memory, oppure i dati in streaming.

Perché è importante? L'integrazione dei dati permette di combinare e raccordare informazioni provenienti da sistemi diversi per ottenere una vista complessiva del fenomeno analizzato che sia più esaustiva e più attinente alle esigenze attuali o future. E ogni nuovo dato che "entra" in azienda moltiplica il suo valore laddove lo si combina con i dati attuali. Anche in questo, la semplificazione e l'efficienza è il fine ultimo di queste tecnologie.

E' come realizzare un puzzle a partire da singoli pezzi provenienti da scatole diverse. E' possibile rispondere a domande del tipo "quali sono i nostri migliori clienti?" Per farlo bisognerà comporre opportunamente il puzzle prendendo i pezzi da sistemi che descrivono i clienti (nome, indirizzo, etc.), altri da sistemi che gestiscono gli indicatori di soddisfazione, altri ancora da sistemi che gestiscono la "storia" degli acquisti o delle interazioni del cliente con l'azienda.

Vedi il caso eclatante di acquisizione o fusione di aziende, oppure di aziende che integrano/consolidano i propri sistemi.

Scenari nei quali vi è un disegno "target" che comporterà obbligatoriamente una definizione dei dati e l'integrazione di due (o più) mondi di partenza.

Data Federation

Cos'è? La Data Federation è la capacità di integrare virtualmente i dati che sono archiviati in luoghi diversi senza dover predisporre un nuovo archivio fisico dove mantenere il risultato. In pratica, è una evoluzione o specializzazione della Data Integration. Infatti, mentre la Data Integration muove e combina dati diversi immagazzinandoli in un luogo "terzo" prima del loro utilizzo (ad esempio nei datamart, datawarehouse, etc.), la data federation salva - e gestisce - la sola definizione del processo consentendo di guardare il risultato combinando i dati "al volo". E' un po' come disporre di acqua corrente da acquedotto, senza dover predisporre adeguate cisterne e un sistema di riempimento delle stesse. Le "viste federate" si dicono così perché, contrariamente alle viste definite nei singoli database, coprono contemporaneamente più sistemi e database differenti in natura e gestione.

Perché è importante? Innanzitutto, la necessità di avere dati sempre aggiornati. Per quanto veloci, i processi di data integration sono eseguiti periodicamente (spesso di notte) e rappresentano sempre e solo una fotografia di quanto c'era nei sistemi di origine. Grazie agli strumenti di Data Federation, il risultato viene invece generato al volo, quindi ogni volta il dato è aggiornato "alla fonte".

Altra motivazione, forse più importante, deriva dalla necessità di rispondere velocemente a esigenze di realizzare nuovi dati. Oggi vi sono molte opportunità che hanno una finestra temporale di poche settimane, prodotti lanciati per due-tre mesi, con richieste di analisi e report di dati mai pensati prima, o comunque con cambiamenti e manutenzione sugli archivi. Ogni CIO in procinto di decidere sul progetto del nuovo dwh aziendale deve valutare i tempi di implementazione, se di 6, 8 o 12 mesi: il nuovo dwh conterrà le informazioni che serviranno tra 12 mesi oppure le "novità esterne" che arriveranno nel frattempo renderanno già incompleto oppure obsoleto questo investimento? La data federation promette maggior flessibilità e velocità nel cambiare le cose, almeno alcune. Riducendo i tempi di rilascio di nuovi archivi virtuali senza aumentare il numero di archivi da mantenere, ingrandire, rimpiazzare, dismettere.

Data Governance

Cos'è? La Data Governance è una pratica aziendale volta ad assicurare che tutti i dati rilevanti e strategici vengano gestiti formalmente a livello enterprise. O, per dirla in un altro modo, allineare la strategia sui dati con le strategie aziendali, orchestrando persone, processi e tecnologie di gestione del dato. Gli strumenti di Data Governance semplificano questa "orchestrazione", ad esempio facilitando la definizione e la gestione di regole di controllo sui processi di gestione del dato, automatizzando processi operativi (es. verifica e gestione di anomalie su dati di rischio), ma soprattutto garantendo la tracciabilità univoca della filiera del dato e delle responsabilità dei singoli attori, ad esempio a fini di audit.

Perché è importante? La governance dei dati inizia con domande generali di business e risposte sulle policy di sviluppo: come la vostra organizzazione utilizza i dati? Quali sono i vincoli che devi considerare? Qual è il contesto normativo? Chi ha la responsabilità sui dati? Una volta che le risposte a queste domande sono note, si possono impostare regole appropriate. Ad esempio, decidere a quali dati gli utenti possono accedere, quali utenti possono modificare i dati oppure solo visualizzarli, come gestire le eccezioni. Non solo. Gli strumenti di Data Governance possono essere utilizzati per controllare e gestire le regole, tracciare il modo in cui vengono utilizzate e fornire relazioni per attività di audit. La possibilità di ottenere un controllo accurato è forse l'aspetto più importante, in quanto i vertici delle organizzazioni devono garantire alla fine l'accuratezza delle relazioni finanziarie davanti ai consigli di amministrazione, agli azionisti, ai clienti e alle istituzioni governative. Si tratta di una responsabilità di grande peso e che comporta il rischio di censura, multe onerose e persino azioni legali, se non gestita correttamente.

Master Data Management

Cos'è? Il Master Data Management (MDM) è un insieme di processi e tecnologie che aiutano a definire, uniformare e gestire i dati condivisi da tutte le aree aziendali fornendo una vista unica dei dati relativi a clienti, prodotti, fornitori, asset/location. I Master Data sono gestiti tipicamente in un singolo punto, spesso chiamato master data management hub. L'hub agisce come unico punto di pubblicazione e condivisione di questi dati critici, in modo coerente.

Perché è importante? Assicura che utenti diversi, in aree e con funzioni diverse, utilizzino la stessa versione dei dati. Senza MDM, ad esempio, un cliente che stipula una polizza con una compagnia, potrebbe ricevere un'offerta promozionale per la stessa polizza dalla stessa assicurazione. Questo può capitare quando il dato relativo al cliente è registrato in modo corretto nel sistema di CRM, ma non nella lista dei destinatari delle campagne di marketing. E' quindi fondamentale disporre di un ambiente che permetta alle organizzazioni di disporre di una "vista unica e certificata della realtà" garantendo così la consistenza del dato in tutti i processi "core".



**THE
POWER
TO KNOW.**

SAS INSTITUTE VIA DARWIN, 20/22 20143 MILANO 02 831 341 www.sas.com/italy
WORLD HEADQUARTERS CARY, NC 27513 USA +1 919 677 8000

SAS e tutti gli altri nomi di prodotto o di servizio di SAS Institute Inc. sono marchi registrati o marchi di SAS Institute Inc. negli USA e in altre nazioni. Altri marchi o nomi di prodotti citati sono marchi registrati dei rispettivi produttori. Copyright © 2012, SAS Institute Inc. Tutti i diritti riservati.