

BIG DATA E ANALYTICS: IL PATRIMONIO DELLE INFORMAZIONI

di MARIA MORO

AVERE A DISPOSIZIONE UNA GRANDE QUANTITÀ DI DATI NON OFFRE DI PER SÉ LA SICUREZZA DI OTTENERE LA RISPOSTA CHE SI CERCA NEI TEMPI ATTESI. LE NUOVE TENDENZE DELL'ANALYTICS PUNTANO SU UNA MAGGIORE VELOCITÀ DI ELABORAZIONE E SOPRATTUTTO SULLA QUALITÀ DEL RISULTATO

La sfida in atto è quella di mettere in campo i migliori strumenti tecnologici con le migliori competenze per poter governare in modo quanto più ampio ed efficace i dati che già oggi sono a disposizione delle aziende e presenti all'interno dei *data warehouse* con quelli che la rete raccoglie dagli oggetti collegati.

Queste enormi mole di dati, ovvero big data, possono essere definiti come un'importante quantità di risorse informative raccolte, immagazzinate e gestite da tecnologie differenti, che richiedono modalità di elaborazione veloci ed economiche e che permettono di ottenere informazioni dettagliate utili ai processi decisionali e all'automazione delle attività.

Diventa così essenziale la gestione del ciclo di vita del dato per poter rispondere alle esigenze del business e dare un significato a quelli che costituiscono il patrimonio informativo di un'azienda, senza dimenticare gli adempimenti normativi in tema di presidio della qualità dei dati e di privacy ai quali oggi tutte le aziende sono chiamate a rispondere.

A OGNI RICHIESTA IL DATO GIUSTO

Gartner offre una sintesi di come dovrebbe essere un sistema di big data, ovvero quelle che in gergo vengono chiamate le 5 V: *volume*, *velocità*, *varietà* (asset tradizionali), *valore*, e *veridicità* (asset nuovi).

Intorno a questi asset si costruisce la visione dei big data di **Kirey Group**, che ha implementato una struttura per gestire e monitorare il ciclo di vita dei dati, dalla nascita sui sistemi sorgente fino all'arrivo sul database e all'elaborazione delle informazioni.

“Il termine utilizzato per indicare la massa diversificata dei dati non strutturati che vengono raccolti è *datalake*”, spiega **Valter Cavallino**, senior manager information management, application analytics, BI tools di Kirey Group. “Un'immagine – osserva – che concretizza il concetto di un contenitore che raccoglie tutte le informazioni delle aziende in un silos di dati disgregati, ancora senza forma, liquidi. Servono competenze specialistiche per disegnare quest'area di dati destrutturati e per far affio-

rare le informazioni e portarle in un data warehouse”. Proprio come la sabbia in un lago, i dati si depositano e rimangono in attesa di essere fatti emergere. Il compito dei data engineers di Kirey Group è di identificare i dati utili a fornire delle risposte a specifiche domande e di dare loro una forma organizzata: “facciamo affiorare le informazioni rilevanti per gli utenti dell’azienda a seconda dell’interesse, il livello di strutturazione del dato viene definito in base alle esigenze dei vari settori e livelli aziendali, e le informazioni vengono aggregate in layer utilizzabili dall’operatore dell’impresa”, rivela Cavallino.

LA VIRTUALIZZAZIONE E IL CLOUD

Le tradizionali tecnologie di big data possono trovare ulteriore vigore, minimizzando ulteriormente il time-to-market delle soluzioni, attraverso l’adozione di tecnologie di *data virtualization* che consentono di modellare uno o più layer dati della soluzione di big data in maniera più snella rispetto alle tecnologie tradizionali. Il dato risiede in un unico punto e resta di fatto certificato. Da non trascurare, infine, l’opportunità dei servizi in cloud che garantiscono al cliente elevati standard di sicurezza e *disaster recovery*, oltre a un’altissima scalabilità che diventa a portata di click.

COMPETENZE PER LEGGERE GLI ADVANCED ANALYTICS

La grande mole di dati presente all’interno di un datalake, la loro complessità e la velocità necessaria per processare ed estrapolare informazioni e fenomeni all’interno di esso, richiedono sempre di più l’intervento di figure professionali con competenze specialistiche diverse dalle tradizionali figure che si possono trovare all’interno di un dipartimento IT.

Le società che offrono servizi di information management si trovano oggi ad ampliare la propria offerta anche a forme di servizio che favoriscano la capacità dell’impresa cliente di utilizzare quanto meglio possibile le informazioni raccolte: “ci siamo accorti dell’esigenza dei nostri interlocutori di poter contare su un partner in grado di gestire innanzitutto strategicamente un sistema di informazioni, attraverso l’analisi della situazione e l’implementazione di un’architettura tecnologicamente innovativa in grado di fornire un pieno controllo del dato. Abbiamo iniziato, con il nostro centro di innova-



zione **Kubris**, a formare persone con background matematico e statistico, e a utilizzare nei nostri progetti tecnologie innovative come quella del machine learning”. Figure come il *data scientist* o l’esperto di *machine learning* lavorano in collaborazione con chi all’interno dell’impresa conosce le logiche aziendali di business per ricercare nel datalake le informazioni utili all’obiettivo e dare loro la forma corretta e facilmente leggibile.

L’INFORMAZIONE DEVE ESSERE VERA

Il cerchio dei big data si chiude con la veridicità e il valore del dato. Garantire la *data quality* è una sfida complessa: utilizzare il dato non porta a una soluzione se l’informazione di cui è costituito non è corretta. Il dato, quindi, prima di essere utilizzato deve essere assoggettato a un “sistema di controllo che fornisce degli indici per verificare la robustezza di tutto il processo”, precisa Cavallino. Una modalità di analisi che è in grado di ricostruire la storia di quel dato e delle sue componenti. Grazie all’esperienza maturata, il team di information management di Kirey Group ha sviluppato una suite di prodotti, *DG Suite*, al fine di fornire un unico prodotto che supporti l’utente nei processi di monitoraggio, gestione e reportistica dei dati. Basata sulle metriche di IQF, la suite permette di progettare i controlli necessari al monitoraggio dei dati e raccogliere gli esiti delle analisi per calcolare la robustezza di un impianto di controllo; permette inoltre di attuare azioni di miglioramento ai processi di gestione dei dati e di produrre una reportistica compliant alle normative e regolamenti vigenti.