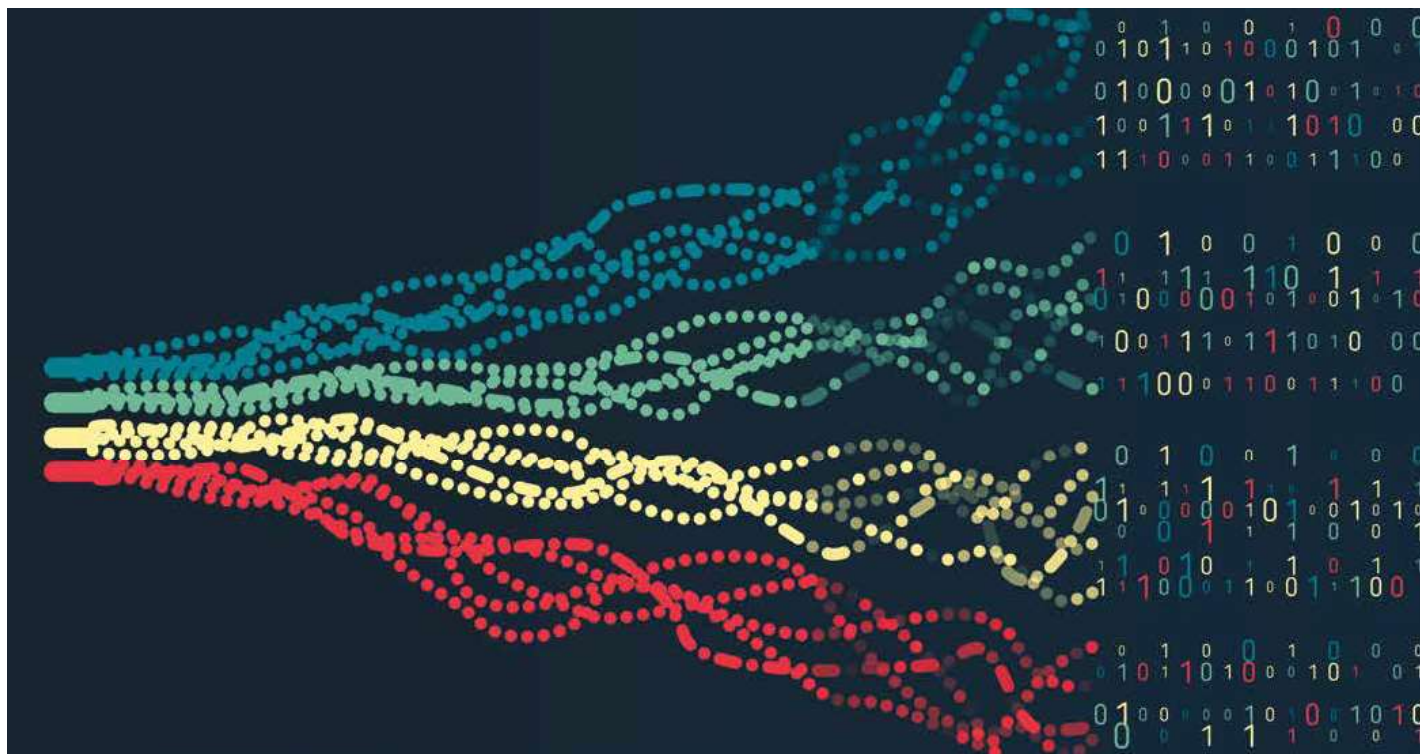


OPEN INNOVATION



Alla scoperta dei Big Data

La gestione e analisi di enormi quantità di dati come opportunità di business



DI VINCENZO GERMANO*

Big Data rappresentano una delle più significative evoluzioni nella gestione e nell'analisi dei dati nell'era digitale. Anche se il termine stesso è emerso solo negli ultimi decenni, gli "antecedenti" dei Big Data possono essere fatti risalire a molto tempo fa. Basti pensare che negli anni '60-'70 l'elaborazione dei dati inizia a diventare più comune con l'avvento dei primi computer, le organizzazioni iniziano a raccogliere e archiviare dati in modo elettronico ma le dimensioni dei dati rimangono relativamente piccole rispetto agli standard moderni. Nei successivi anni '80-'90 l'espansione dell'uso dei database relazionali permette alle aziende di gestire grandi quantità di dati strutturati in modo più efficiente, tuttavia la maggior parte dei dati rimane relativamente gestibile. Negli anni 2000 la crescita esplosiva di Internet, dei dispositivi mobili e delle transazioni online inizia a generare enormi quantità di dati, le prime aziende tech iniziano a raccogliere e sfruttare questi dati per offrire servizi migliori, dando vita al concetto di Web 2.0. Tuttavia dopo il 2010, con l'aumento dei social media, degli smartphone e dei dispositivi IoT (Internet of Things), i dati non strutturati, come testi, immagini e video, diventano più diffusi, con una conseguente crescita esponenziale di volumi.

Quando si parla di Big Data spesso non si riesce a dare una defi-

nizione precisa di quello che realmente sono: infatti, pur sapendo che possono essere utilizzati da aziende, enti governativi e qualsiasi tipologia di impresa per realizzare servizi/prodotti basati sulle abitudini dei consumatori, rimane difficile riuscire a trovare una giusta definizione senza approfondirne il concetto.

A grandi linee, si potrebbe dire che i Big Data sono grandi quantità di dati, organizzati in database e raccolti in vari modi come possono essere le ricerche web, i social network, gli smartphone, i dispositivi smart IoT e via dicendo. Tali dati, una volta raccolti, vengono processati con algoritmi specifici (o intelligenza artificiale) per ricavare informazioni utili agli scopi più disparati, come, ad esempio, comprendere le abitudini dei consumatori e le preferenze; oppure vengono utilizzati per "educare" ulteriori algoritmi come può accadere nei sistemi di riconoscimento facciale. Tuttavia non ne abbiamo ancora dato una definizione specifica, per fare ciò e quindi definire i Big Data ci viene in aiuto il concetto delle tre "V".

Il termine viene utilizzato per descrivere il crescente volume, velocità e varietà dei dati generati e raccolti da varie fonti digitali, come sensori, social media, dispositivi mobili, transazioni online e molto altro; anche perché questi dati, in molti casi, sono troppo grandi e complessi per essere gestiti ed elaborati con le tradizionali tecniche di gestione dei dati. Approfondiamone i dettagli e le componenti principali.

LE TRE "V"

Per comprendere il concetto del "volume" facciamo un esempio, considerando le foto salvate dagli utenti su Facebook; si stima che da quando sia stato creato abbia raccolto oltre 260 miliardi di foto, per quanto è molto probabile che il numero sia anche maggiore. Questo se parliamo di foto, ma considerando i post, la quantità di dati legata ai suoi iscritti è almeno 10 volte superiore. Solo per dire che quando parliamo di volume dei Big Data dobbiamo relazionarci con questi numeri, ossia una mole di dati enorme tra i più disparati. Prendendo d'esempio gli oggetti legati all'Internet of Things (IoT), esistono sensori sempre connessi che forniscono informazioni della nostra casa, auto o ufficio, e ogni secondo registrano in tutto il mondo milioni di dati. Basti pensare che in pochi anni i numeri incredibili di Facebook stanno diventando "piccoli" paragonati a quelli che riusciamo a ricevere dai vari dispositivi sempre connessi. Ecco, il volume dei Big Data si riferisce alla quantità massiccia di dati generati quotidianamente, che possono essere strutturati, come database relazionali o non strutturati, come testo, immagini, audio e video.

Passando oltre e considerando la "velocità", bisogna pensare che per poter salvare, archiviare, catalogare e riutilizzare queste informazioni così pesanti serve tale aspetto; infatti, se non si avesse la giusta velocità, sarebbe inutile avere tante informazioni senza poi poterle gestire nell'im-

mediato. Perciò essa riguarda la rapidità con cui i dati vengono generati, raccolti e analizzati. Ad esempio, i social media generano continuamente nuovi dati attraverso post, tweet e commenti, infatti gli operatori devono essere in grado di acquisire, elaborare e analizzare questi dati in tempo "quasi reale" per prendere decisioni rapide. Considerando che l'obiettivo finale dei Big Data è nell'analisi ed estrazione delle informazioni significative e utili dalle masse di dati disponibili utilizzando tecniche di analisi avanzate, tra cui l'apprendimento automatico, la data mining, l'analisi predittiva e l'analisi dei sentimenti; senza velocità di raccolta ed elaborazione delle informazioni, le varie aziende non sarebbero interessate ad acquistare dati riguardanti potenziali consumatori o attuali clienti dai vari servizi online.

Infine, la "varietà" dei Big Data si riferisce alla diversità dei tipi di dati disponibili: infatti, ogni dato è molto diverso dall'altro, oltre a quelli testuali e numerici, ci sono dati multimediali, geospaziali, dati generati da sensori IoT e molti altri. Gestire questa varietà richiede la capacità di trattare dati strutturati e non strutturati, che in ogni caso può creare problemi di gestione e unione dei vari dati per ottenere delle informazioni più complete. Tuttavia le varie tecniche applicate ai Big Data aiutano le organizzazioni a prendere decisioni più informate, a scoprire nuove opportunità di business e a ottimizzare i processi

operativi.

Ulteriori caratteristiche dei Big Data sono la **veridicità** e la **variazione**. La prima si riferisce alla qualità e all'affidabilità dei dati, infatti a causa della varietà delle fonti e della velocità di generazione i dati possono essere "sporchi", incompleti o inaccurati. Pertanto, è fondamentale avere processi di controllo di qualità dei dati per garantire che i dati siano affidabili e accurati. La variazione invece si riferisce alla costante evoluzione e cambiamento dei dati nel tempo, questo perché le informazioni raccolte oggi potrebbero non essere rilevanti domani.

Pertanto, le organizzazioni devono essere in grado di adattarsi rapidamente ai cambiamenti nei dati e nelle condizioni del mercato. L'analisi in tempo reale e l'uso di modelli dinamici sono strategie chiave per affrontare questa sfida.

CONCLUSIONI

In conclusione, i Big Data rappresentano una tecnologia che sta rivoluzionando la gestione e l'analisi di enormi quantità di dati. Affrontano sfide legate al volume, alla velocità, alla varietà e alla veridicità dei dati, offrendo opportunità per estrarre valore da questa mole di informazioni. L'adozione di tecnologie e strategie adeguate è essenziale per sfruttare appieno il potenziale dei Big Data in vari settori.

*INGEGNERE ELETTRONICO E VICECOORDINATORE COMMISSIONE INNOVATION & PROJECT MANAGEMENT ORDINE DEGLI INGEGNERI DELLA PROVINCIA DI TORINO