Universitas Mercatorum

I BIG DATA

Carlo De Matteo

Indice

1. DEFINIZIONI

Big data ("grandi dati" in inglese) è un termine adoperato per descrivere una raccolta di dati eterogenei, strutturati e non strutturati, definita in termini di volume, velocità e varietà. Per la gestione di tale mole di dati sono richieste tecnologie e metodi analitici specifici adatti a sviluppare la ricerca per supportare differenti tipi di analisi.

I Big Data sono dati che vanno oltre i limiti dei database tradizionali, ma con questo termine si intendono anche le tecnologie finalizzate ad estrarre dai dati stessi conoscenze e valore.

In considerazione della loro enorme estensione in termini di volume, ma anche delle loro intrinseche caratteristiche velocità e varietà, i Big Data richiedono tecnologie e metodi analitici specifici indirizzati all'estrazione di valori di interesse.

Il progressivo aumento della dimensione dei dataset è legato alla necessità di analisi su un unico insieme di dati, con l'obiettivo di estrarre ulteriori informazioni rispetto a quelle che si potrebbero ottenere analizzando piccole serie, con la stessa quantità totale di dati.

E' possibile, per esempio generare un'analisi per sondare gli "umori" dei mercati e del commercio, e quindi dell'andamento complessivo della società e del fiume di informazioni che attraversano Internet.

Per poter parlare di Big Data il volume dei dati deve essere correlato alla capacità del sistema di acquisire le informazioni così come arrivano dalle differenti sorgenti dati che si sono utilizzate. Quindi, un sistema diventa big quando aumenta il volume dei dati e allo stesso tempo aumenta la velocità/flusso di informazioni che il sistema deve poter acquisire e gestire per ogni secondo.

Data la complessità di una definizione univoca del termine Big Data, sono state proposte differenti definizioni da varie organizzazioni.

Nel 2011, Teradata afferma che "Un sistema di big data eccede/sorpassa/supera i sistemi hardware e software comunemente usati per catturare, gestire ed elaborare i dati in un lasso di tempo ragionevole per una comunità/popolazione di utenti anche massiva".

Un'ulteriore definizione di big data è stata data dal McKinsey Global Institute: "Un sistema di Big Data si riferisce a dataset la cui taglia/volume è talmente grande che eccede la capacità dei sistemi di database relazionali di catturare, immagazzinare, gestire ed analizzare".

Big Data rappresenta anche l'interrelazione di dati provenienti potenzialmente da fonti eterogenee, quindi non soltanto da dati strutturati, come quelli contenuti in un database. Infatti i sistemi gestionali tradizionali trattano esclusivamente dati strutturati o strutturabili utilizzando tabelle tra loro relazionabili.

I Big Data comprendono, invece, dati generati da molte fonti eterogenee come quelli del web e che sovente sono rappresentate da

dati semi-strutturati o non strutturati affatto come i post sui blog, le informazioni prese dai social network, i documenti di testo, audio, video disponibili in diversi formati, immagini, email, dati GPS, e così via.

La rivoluzione Big Data e, in generale, il termine Big Data si riferisce proprio a cosa si può fare con tutta questa quantità di informazioni, ossia con algoritmi capaci di trattare così tante variabili in poco tempo e con poche risorse computazionali.

Il paragone è presto e fatto: fino a qualche tempo fa, uno scienziato per analizzare una montagna di dati che oggi definiremmo small o medium data avrebbe impiegato molto tempo e si sarebbe servito di computer mainframe da oltre 2 milioni di dollari. Oggi, con un semplice algoritmo, quelle stesse informazioni possono essere elaborate nel giro di poche ore, magari sfruttando un semplice laptop per accedere al software di analisi.

Questa è la rivoluzione Big Data: nuove capacità di collegare fra loro le informazioni fornendo anche un approccio "visual" dei dati e suggerendo pattern e modelli di interpretazione fino ad ora inimmaginabili.

I Big Data non interessano solo il settore IT. Infatti, se l'Information Technology rappresenta, da una parte, per i Big Data la grande partenza da cui si è iniziata l'opera con strumenti come il cloud computing, gli algoritmi di ricerca, dall'altra i Big Data sono necessari e utili nei mercati business più disparati, dall'automotive, alla medicina, dal commercio all'astronomia, dalla biologia alla

chimica farmaceutica, dalla finanza al gaming. In sostanza nessun settore in cui esiste la possibilità di produrre una grande quantità di dati da analizzare può dirsi escluso dalle applicazioni Big Data.

Il flusso informativo non accenna ad arrestarsi, anzi secondo gli esperti si avrà un aumento del 4300 per cento nella generazione di dati annuali, entro il 2020. Oggi si parla spesso di big data in relazione all'Internet of Things, in quanto gli oggetti "connessi" generano un prezioso ed esteso patrimonio informativo.

2. STORIA E ORIGINI

Non è recente il tentativo di quantificare la dimensione del flusso informativo, in quanto già nella prima metà del secolo scorso furono condotti i primi esperimenti che miravano ad individuare il tasso di crescita del volume di dati, noto a tutti come "esplosione di informazioni", termine apparso per la prima volta nel 1941, secondo l'Oxford English Dictionary. Con il tempo è diventata sempre più inderogabile l'esigenza di archiviare questo patrimonio di dati in modo razionale ed organizzato, utilizzando una metodologia che potesse portare alla loro comprensione in modo automatico.

La prima volta che un articolo scientifico ha parlato di big data è stato quando i ricercatori Steve Bryson, David Kenwright, Michael Cox, David Ellsworth e Robert Haimes, nel 1999, sulla rivista "Communications of the ACM", hanno sottolineato l'importanza di fare uno "sforzo" significativo per comprendete questo enorme ammontare di dati.

Negli anni '60, grazie alle nuove tecnologie, un gran quantitativo di dati venne per la prima volta digitalizzato e archiviato, grazie anche alla crescita esponenziale degli stessi, in un singolo ed enorme computer, il "Main Frame".

Il Main Frame era grande come un palazzo di 4 piani e per accedere ai dati bisognava recarsi fisicamente sul posto e operare con la macchina attraverso linguaggi rudimentali e laboriosi.

Nel 1967: Marron e de Maine svilupparono il primo algoritmo di compressione dei dati e poco dopo, nel 1970 l'evoluzione dei primi

computer permise di archiviare una grande quantità di dati su più sistemi, creando network indipendenti favorendo la nascita del primo sistema distribuito, quello che oggi si ricorda come CERNET.

Nel 1975 il Ministero delle Poste e Telecomunicazioni Giapponese realizzò nel 1975 il primo censimento del flusso informativo nazionale, introducendo la misura di «numero di parole».

All'inizio degli anni 80 e con la grande espansione dei networks e della tecnologia consumer, si diffuse il Personal Computer e con questo la possibilità di accedere da remoto ed in modo del tutto autonomo a qualsiasi fonte di dati.

J. R Masey nel 1998 con un articolo dal titolo «Big Data … and the next wave of Infrastress», ad introdusse il termine «Big Data» teorizzando che questi sarebbero stati il prossimo elemento di stress delle infrastrutture informatiche.

Nel 2001 Laney indentificò per la prima volta le tre dimensioni che sono generalmente accettate come gli elementi necessari a definire i Big Data ovvero le tre V: Volume dei dati, Varietà dei dati e Velocità di raccolta e di utilizzo.

All'inizio degli anni 2000, data l'elevata quantità di Big Data presenti nel mondo, fu necessario organizzare interi edifici adibiti all'archiviazione.

3. TIPOLOGIE

Una prima caratterizzazione venne fornita da Douglas Laney, analista di Gartner, in un suo articolo del 2001, nel quale egli stabilì un modello di crescita dei dati orientato intorno tre aspetti fondamentali:

Volume: si riferisce alla quantità di dati, strutturati e non strutturati, generati ogni secondo.

Tali dati possono essere generati da sorgenti eterogenee di ogni genere quali: sensori, log, eventi, email, social media e ovviamente database tradizionali.

Con i Big Data poi, elementi come transazioni bancarie, movimenti sui mercati finanziari o informazioni prodotte dagli utenti sui social media, assumono naturalmente valori mastodontici che non possono in alcun modo essere gestiti con i tradizionali database. Per dare una dimensione al problema possiamo ricondurre il volume dei dati generati da una azienda di media grandezza all'ordine di terabyte o petabyte ma quando valutiamo dei Big Data, per darne una dimensione reale, si deve ricondurre la mole dei dati all'ordine degli zettabyte, ovvero di miliardi di terabyte.

È facile immaginare che per analizzare questi volumi è richiesta una potenza di calcolo parallelo e massivo con strumenti di data processing dedicati ed eseguiti su decine, centinaia o anche migliaia di servers.

Varietà: si riferisce alla differente tipologia dei dati che vengono generati, collezionati ed utilizzati. In precedenza i dati erano

fortemente strutturati e la loro manipolazione veniva eseguita mediante uso di tabelle o database relazionali. Con l'avvento dei Big Data e le tecnologie a questi applicati, i dati possono essere di qualsiasi genere: non strutturati, semi strutturati (non hanno una strutturazione logica precisa) o strutturati.

La varietà dei Big Data è dovuta infatti soprattutto alla loro mancata strutturazione. Tra questi infatti vengono inclusi documenti di vario genere come ad esempio: file txt, csv, PDF, Word, Excel, blog post, commenti sui social network o sulle piattaforme di micro-blogging come Twitter.

I Big Data sono vari anche nelle fonti: alcuni sono generati automaticamente da macchine, come i dati provenienti da sensori o i log di accesso a un sito web o quelli del traffico su un router, altri sono generati dagli stessi utenti del web nella loro quotidiana navigazione.

Velocità: usualmente si riferimento a due concetti di velocità. Primo concetto: la velocità con cui i dati vengono generati ovvero come speed of generation. Secondo concetto: come velocità nella loro disponibilità, ovvero quanto velocemente questi dati possono essere reperiti in tempo reale al fine di effettuare analisi su di essi.

Tra le tecnologie capaci di gestire i dati "ad alta velocità" vi sono ad esempio i "database historian", per l'automazione industriale e quelle denominate "streaming data o complex event processing (CEP)", che consentono di monitorare più fonti di dati, analizzando questi ultimi in modo incrementale con una bassissima latenza.

Le applicazioni CEP sono applicate con successo in vari ambiti come quello industriale, scientifico, finanziario e in quello relativo all'analisi degli eventi generati sul web.

Con il passare del tempo, ulteriori caratteristiche si sono aggiunte al modello, quali ad esempio la **Variabilità** che rappresenta un problema sempre più tipico e si riferisce alla possibilità di inconsistenza dei dati "data inconsistency" e la **Complessità**: maggiore è la dimensione del dataset, maggiore è la complessità dei dati da gestire. In questi casi il compito più difficile è collegare le informazioni, ed ottenerne output interessanti.

4. TECNOLOGIE PER LA GESTIONE DEI BIG DATA

La crescente mole di dati che vengono generati da sorgenti diverse ha posto l'attenzione su come collezionarli, archiviarli ed utilizzarli ai fini di business.

Uno dei problemi manifestatisi circa la gestione dei Big Data è offerta anche dalla natura degli stessi, che cambia di volta in volta, aumentando sempre di più la dimensione dell'informazione da gestire. Il problema che si è riscontrato è stato dovuto principalmente alla difficoltà di gestirli con database tradizionali, sia in termini di costi, sia in termini di volume.

L'insieme di questi elementi ha portato allo sviluppo di nuovi modelli di elaborazione, che ha permesso alle aziende di diventare più competitive, sia attraverso una riduzione dei costi, sia perché i nuovi sistemi, sono in grado di archiviare, spostare e combinare i dati con maggiore velocità e in maniera agile.

In considerazione ai problemi legati al volume, velocità e varietà che rappresentano, per poter gestire i Big Data, si adoperano appositi sistemi e soluzioni tecnologiche che hanno come peculiarità la capacità di distribuire sia risorse che servizi. Tali architetture si definiscono "distribuite" e utilizzano gruppi di computer denominati "clusters", connessi tra loro al fine di cooperare al raggiungimento di un obiettivo comune realizzando la cosiddetta scalabilità orizzontale ovvero, per ottenere più potenza di calcolo non è necessario aggiungere memoria ad un solo computer ovvero in regime di

”scalabilità verticale” ma si può ottenere la stessa potenza connettendo più computer che tra di loro condividono risorse di calcolo in modo bilanciato.

Un tipico esempio di soluzioni distribuite è il “grid computing”. Il termine "griglia", in inglese grid, è stato coniato intorno alla metà degli anni novanta. Il vero e specifico problema alla base del concetto di griglia è la condivisione coordinata di risorse all'interno di una organizzazione virtuale (Virtual Organization, brevemente indicata con VO). Nel Grid Computing, la condivisione non è limitata solo allo scambio dei file, ma si estende all'accesso diretto a computer, a software, in generale a tutto l'hardware necessario alla risoluzione di un problema scientifico, ingegneristico o industriale. Gli individui e le istituzioni, che mettono a disposizione della griglia le loro risorse per la medesima finalità, fanno parte della stessa VO. Caratteristica comune è la necessità di disporre un ambiente di calcolo data-intensive, all'interno del quale le applicazioni hanno il bisogno di accedere a grandi quantità di dati geograficamente distribuiti in maniera veloce e affidabile ed è proprio l'onere della grid far operare tali applicazioni nel miglior modo possibile.

È facile osservare che nessun computer attualmente in commercio sarebbe in grado, da solo, di elaborare simili moli di dati in tempi ragionevoli; tuttavia la condivisione di risorse quali CPU e dischi opportunamente coordinati può dare l'impressione all'utente di accedere ad un supercomputer virtuale, con una incredibile potenza computazionale e capacità di memorizzazione in grado di sopportare grandi carichi di lavoro. Dall'idea di far apparire tutta l'architettura

di un grid come un unico supercomputer virtuale, celando all'utilizzatore tutta la complessità interna e mostrandogli solo i benefici, nasce l'esigenza di progettare e realizzare uno schedulatore di risorse, il cosiddetto Resource Broker.

Una delle caratteristiche delle architetture distribuite è che devono essere progettate per essere tolleranti ai guasti, il così detto "fault tolerance" e per questo le risorse e i servizi sono replicati sulle differenti macchine che compongono il "cluster". Infine il modello di elaborazione è distribuito in modo da poter sfruttare la potenza elaborativa del "cluster" progettato allo scopo quindi di processare più calcoli e più velocemente per lo stesso quantitativo di dati: il così detto "distributed processing".

Per gestire un così grande quantitativo di dati sono nate diverse metodologie.

In primo luogo il "Data Mining" ovvero il processo di estrazione di conoscenza da banche dati di grandi dimensioni tramite l'applicazione di algoritmi che individuano le associazioni "nascoste" tra le informazioni e le rendono visibili. In altre parole, col nome data mining si intende l'applicazione di una o più tecniche che consentono l'esplorazione di grandi quantità di dati, con l'obiettivo di individuare le informazioni più significative e di renderle disponibili e direttamente utilizzabili. L'estrazione di conoscenza (informazioni significative) avviene tramite individuazione delle associazioni, o "patterns", o sequenze ripetute, o regolarità nascoste nei dati. In questo contesto un "pattern" indica una struttura, un modello, o, in generale, una rappresentazione sintetica dei dati.

Il termine data mining è utilizzato come sinonimo di knowledge discovery in databases (KDD), anche se sarebbe più preciso parlare di knowledge discovery quando ci si riferisce al processo di estrazione della conoscenza, e di data mining come di una particolare fase del suddetto processo (la fase di applicazione di uno specifico algoritmo per l'individuazione dei "patterns").

A supporto di queste metodologie sono state proposte le alcune nuove tecnologie e linguaggi di programmazione tra queste, degne di nota è in particolare NoSQL utilizzato da Cassandra, MongoDB e Hadoop Framework (HDFS, Spark, Storm, Mout ed altri.

I database NoSQL (Not Only SQL) rappresentano una valida alternativa ai database relazionali (RDMBS – Relational Database Management System) che vengono interrogati da un linguaggio chiamato SQL (Structured Query Language). Per spiegare i Data Base NoSQL è necessario comprendere innanzitutto quali sono gli elementi principali e le caratteristiche di un database SQL che ruota attorno al concetto rigido di tabella.

In un database SQL, di tabelle ne esiterà una per ogni tipo di informazione da trattare, ed ognuna di queste sarà costituta da colonne una per ogni caratteristica dei dati. Una tabella infatti dovrebbe avere una o più colonne che svolgono il ruolo di chiave primaria, una sorta di indice che permette di riconoscere univocamente quella riga rispetto a tutte le altre. Tra le tabelle di un database relazionale, inoltre, possono esistere alcune relazioni, ad esempio, una riga di una tabella o una colonna.

L'impostazione iniziale condizionerà l'evoluzione di tutto il database, dall'inserimento dei dati alla loro interrogazione. La strutturazione rigida dei contenuti, tipica dei database relazionali. è invece totalmente assente nei database NoSQL e tale assenza è uno degli aspetti che maggiormente ne hanno garantito il successo in ambito Big Data. I database NoSQl, sono orientati infatti al documento e non memorizzano i dati in tabelle con campi uniformi per ogni record, ma ogni record è memorizzato come un documento che possiede determinate caratteristiche. Qualsiasi numero di campi con qualsiasi lunghezza può essere aggiunto o tolto al documento in modo estremamente flessibile.

Con l'avvento dei Data Base non relazionali nasce anche il concetto di "Data Hub" che permette di inserire dei dati opportunamente selezionati e validati in un grande repository senza preoccuparsi della modellazione del dataset sottostante.

Uno dei più famosi software di raccolta e analisi dei Big Data è Hadoop, un progetto open-source di Apache che consente ad uno sviluppatore o a un matematico di inserire nel framework migliaia di dati ed effettuare analisi di ogni tipo.

Alla base di Hadoop c'è una libreria di software per sistemi di computer, scalabili, affidabili e distribuiti e capaci di gestire una grande quantità di Big Data e fornisce anche una piattaforma per l'analisi dei dati.

Hadoop distribuisce su più "cluster" di server, il salvataggio "storage" e la lavorazione "processing" di grandi insiemi di dati "data sets" utilizzando un semplice modello di programmazione. Il numero

di server in un "cluster" può crescere facilmente, a seconda dalle necessità, da 50 a 2000 o più computer.

Se le tradizionali soluzioni a larga scala affidano queste attività a server fisici ad alto costo e con una alta fault tollerance, Hadoop invece rileva e compensa eventuali falle dell'hardware o altri problemi legati ai sistemi, a livello puramente applicativo.

Questo consente l'erogazione del servizio in continuità su molte macchine appartenenti allo stesso "cluster", dove la singola macchina, potenzialmente incline al disservizio, può essere trascurata.

La lavorazione di una grande quantità di dati su una grande ed economica infrastruttura di computer distribuiti diventa quindi una proposizione necessaria per la lavorazione di grandi quantitativi di dato.

Tecnicamente Hadoop è composto da due elementi chiave. Il primo è il suo sistema di archiviazione l'Hadoop Distributed File System (HDFS) che consente il salvataggio dei dati ad alta velocita su un cluster di computer, essenziale per la computazione dei dati. Il secondo elemento di Hadoop è il framework di data processing chiamato MapReduce che prende origine dalla tecnologia di ricerca di Google. MapReduce distribuisce o meglio "mappa", grandi set di dati su molteplici server.

Ciascuno di questi server esegue poi solo la parte di lavorazione del data set che le è stata assegnata e ne crea un sommario. Ciascun sommario generato sui vari server viene poi aggregato ad uno stadio logico applicativo di livello più alto chiamato appunto "reduce". Questo approccio consente ad un set di dati incredibilmente esteso di

venire rapidamente pre-processato prima che vengano applicati successivamente i tool di analisi più tradizionali.

Grazie a questi strumenti, gli analisti non hanno solo i dati su cui lavorare, ma hanno anche l'opportunità di gestire un immenso numero di record con moltissimi attributi. Questa grande quantità di dati avvantaggia l'analisi, che da statistica diventa predittiva, potendo contare su un numero di record e un numero di attributi per record fino a ora inimmaginabile.

5. APPLICAZIONI DEI BIG DATA NEI SETTORI DELL'ECONOMIA

Le funzioni aziendali che beneficiano maggiormente dei Big Data sono il marketing e le vendite, i team di finanza e di controllo, i sistemi informativi, gli acquisti, la produzione e la supply chain.

Le aziende per sfruttare appieno i dati e le analisi hanno bisogno di sviluppare almeno tre capacità:

1. essere in grado di identificare, unire e gestire più fonti di dati;
2. poter costruire modelli avanzati di analisi per la previsione e l'ottimizzazione dei risultati;
3. saper creare una strategia chiara per sfruttare i dati, in modo che questi effettivamente portino a decisioni migliori.

I Big Data possono essere impiegati nelle analisi di tipo interpretativo o predittivo che le organizzazioni sviluppano a supporto dei processi decisionali.

I modelli di business basati sui dati, si stanno concentrando per lo più su come le aziende possono utilizzare la grande quantità di dati raccolti per ottenerne un vantaggio competitivo sul mercato ed evolvere il proprio modello di business.

Le aziende che si dotano di competenze e tecnologie utili per estrapolare "insight" strategici dal patrimonio informativo aziendale sono quelle maggiormente predisposte alla crescita economica e sono

anche quelle più vicine ad avere risultati positivi sul miglioramento della propria situazione finanziaria e riguardo alle proprie prospettive di crescita, rispetto a quelle aziende che non sono ancora invece in grado di aggregare e interpretare i dati.

Qualsiasi settore può trarre valore dai Big Data e sono già numerose le realtà settoriali che hanno testato con successo le tecnologie legate ai Big Data e che le stanno sfruttando per crescere e migliorarsi.

Nell'**industria finanziaria** i Big Data sono molto utilizzati in particolare in alcuni ambiti:

Prevenzione delle frodi - Ogni anno aumentano le perdite causate da frodi e crimini finanziari come il riciclaggio di denaro, i cyber-attacchi e le minacce interne. Risulta sempre più necessario avere a disposizione una soluzione che consenta alle organizzazioni di acquisire una migliore visibilità e adottare un approccio completo e proattivo per analizzare in tempo reale gli impulsi e i segnali delle transazioni commerciali e contrastare le frodi.

Analisi del rischio - Per una banca è fondamentale poter avere a disposizione dei dati costantemente aggiornati che indichino il rischio di determinate operazioni in portafoglio. Più velocemente i manager del rischio hanno accesso ad una panoramica completa ed esauriente degli indicatori, più possono avvisare tempestivamente il front office che allinea le attività con la policy di propensione al rischio della banca.

Nell'industria manifatturiera stanno prendendo sempre di più spazio le applicaizoni dei Big Data alle funzioni di:

Prevenzione di guasti e rotture o "Anomaly detection". Per un'azienda i costi di manutenzione delle macchine e delle altre componenti sono spesso molto alti ed è importante conoscere il ciclo di vita delle attrezzature per poter intervenire in tempo programmando interventi e sostituzione delle componenti.

Ottimizzazione del processo produttivo - Analizziamo il caso di un'impresa che produce energia elettrica e che ha la necessità di distribuirla in modo adeguato ai suoi utenti. Si rende indispensabile un'analitica avanzata dei dati disponibili, in grado di fornire un modello predittivo per la gestione dell'energia.

Gestione della produzione e del magazzino - Un'azienda che opera nel settore dei prodotti freschi deve individuare il giusto quantitativo di articoli da avere in magazzino in maniera sufficiente per riuscire a rispondere alle esigenze di mercato ma non eccessivo per evitare inutili sprechi di risorse e di denaro per lo stoccaggio delle merci.

Largo utilizzo per tutte le aziende rispetto alla propria capacità competitiva è la "Customer Intelligence" potenziata dall'analisi dei Big Data allo scopo di ottimizzare la relazione con i clienti e attrarne di nuovi. Le aziende hanno la necessità di accrescere l'engagement dei clienti o espandere la propria presenza nei mercati di riferimento. Per

farlo devono comprendere il comportamento dei clienti, aumentare la soddisfazione, ridurre il "churn rate" (ovvero la percentuale di clienti che smettono di acquistare o consumare un prodotto o un servizio) e incrementare gli acquisti.

Social media e analisi del "sentiment" - La maggior parte delle aziende comunica con il proprio mercato di riferimento anche attraverso i social network, oppure opera in settori in cui i potenziali clienti si scambiano pareri e informazioni attraverso forum e magazine online. Importante è capire come "capitalizzare" queste interazioni, analizzandole per definire strategie di engagement sempre più efficaci.

Muoversi verso modelli di business basati sui dati corrisponde ad un sempre maggiore uso qualificato del dato ma per definire le strategie di business è necessario un cambiamento che sia prima di tutto culturale e solo successivamente strumentale. Questo processo di trasformazione digitale (o "digital transformation") in azienda interessa non solo ogni ambito e settore industriale, produttivo e amministrativo, ma anche tutti i contesti organizzativi e gestionali all'interno di una stessa impresa.

Tale trasversalità rende auspicabile l'ingresso anche nelle aziende del "Chief Digital Officer". Nonostante l'espansione continua della digitalizzazione ad ogni livello aziendale, tuttavia, non è immediato il consolidamento di questa professione che sembra assumere piuttosto un ruolo di momentanea transizione per estendere a tutta la forza lavoro le competenze digitali necessarie per affrontare

i cambiamenti in atto. Molte aziende, ad esempio, hanno privilegiato la formazione dei propri dirigenti piuttosto che acquisire sul mercato la nuova professionalità del “Chief Digital Officer”.