

## Capitolo 1

# Gestire il patrimonio informativo aziendale: l'informazione a servizio dei *decision maker*

*Maria Palazzo*

Sommario: 1.1. L'Informazione digitale e l'organizzazione aziendale. – 1.2. L'informazione economica. – 1.3. Gli standard tecnologici della *big data analytics*. – 1.4. L'evoluzione dei confini organizzativi e della *Value Chain* ai tempi della trasformazione digitale. – 1.5. Il valore di scambio dell'informazione: l'innovazione per lo sviluppo della crescita aziendale. – 1.6. La *performance* aziendale nell'economia di velocità. – 1.7. La *talent analytics*.

### 1.1. L'informazione digitale e l'organizzazione aziendale

In una società scandita dall'iper-connesione, le *new technology* hanno dato vita a processi organizzativi (Viriyasitavat *et al.*, 2019), generando nuove opportunità economiche (Lee *et al.*, 2019), ridisegnando la struttura dei costi (Chen e Bellavitis, 2020) e modificando gli ambiti spazio-temporali della competizione fra imprese (Balog, 2020).

Per cui, è necessario circoscrivere precisamente il perimetro entro cui utilizziamo il termine azienda e le sue varie accezioni.

Partiamo con il definire l'azienda come una struttura organizzativa di *asset* materiali e immateriali (Hoang e Phang, 2023). Ai primi possiamo ascrivere tutti quei beni che sono effettivamente tangibili, come macchinari, strumenti, forniture ecc. di cui si calcola l'ammortamento<sup>1</sup>. I secondi invece, mancano di una dimensione fisica (esempi includono *software*, brevetti, marchi, ecc.). Il valore di questi ultimi, ad una prima analisi, potrebbe risultare più complicato da quantificare rispetto agli *asset* tangibili.

Nelle scienze manageriali contemporanee dell'ultimo quinquennio, la definizio-

---

<sup>1</sup> In contabilità, l'ammortamento è il procedimento tecnico per ripartire il costo di talune immobilizzazioni a vita definita in un dato numero d'esercizi futuri, mediante quote che incidono sul calcolo del reddito dei vari esercizi (Codice civile art. 2426 c.c.).

ne di azienda assume una connotazione dinamica, ovvero come sistema di forze economiche, che sviluppa relazioni negli ambienti in cui opera e coi sistemi di produzione<sup>2</sup> e/o consumo<sup>3</sup> con cui entra in contatto.

Nonostante esistano diversi parametri attraverso i quali classificare le aziende, i principali criteri idonei a stabilirne la natura sono: il fine perseguito e la natura dei soggetti promotori.

Sulla base della distinzione, nell'ambito delle finalità perseguite, tra processi di consumo e processi produttivi è possibile distinguere tra aziende di produzione ed aziende di consumo.

Nel caso delle aziende di produzione parliamo di sistemi orientati allo svolgimento di attività a fini di lucro, rivolta alla produzione diretta o indiretta di beni e/o di servizi. Tenuto conto della natura dell'attività economica che svolgono, esse sono classificate in imprese di produzione diretta (es. le imprese industriali); imprese di produzione indiretta (es. le imprese commerciali); imprese di servizi (es. imprese di trasporti, di forniture, ecc.)<sup>4</sup>.

Le imprese di consumo invece, utilizzano i beni e servizi disponibili per soddisfare bisogni di svariata natura (es. famiglie, pubbliche amministrazioni, ecc.).

Infine, sulla base del soggetto promotore che le gestisce, le aziende possono essere classificate come private, pubbliche e miste, in cui intervengono soggetti privati ed enti pubblici insieme (es. le aziende a partecipazione statale).

Questi distinguo terminologici aiutano ad orientare studiosi e manager nel prosieguo della trattazione, costituendo un bacino di informazioni che nel corso dei capitoli si arricchirà delle nuove significazioni apportate dal digitale.

Tra i principali fattori che hanno favorito la rivoluzione digitale vi è senza dubbio la globalizzazione che ha impattato la dimensione organizzativa consentendo di ottenere vantaggi sulle economie di scala tanto nei beni materiali, quanto nella produzione di contenuti ad alta densità informativa (Levitt, 1983).

Conseguenza diretta delle implementazioni tecnologiche è la costituzione all'interno delle aziende di patrimoni informati derivanti dai sistemi di raccolta ed anali-

---

<sup>2</sup> Il sistema di produzione è l'insieme delle operazioni, semplici o complesse, attraverso le quali si produce un bene trasformando altri beni. Si veda tra l'altro, Bennett e Balvanera (2007).

<sup>3</sup> Il consumo riguarda il processo o l'atto mediante cui i beni economici sono utilizzati e, nel caso di beni materiali, integralmente o parzialmente distrutti per appagare un bisogno (consumo di godimento) o per produrre nuovi beni (consumo produttivo o riproduttivo). A tal riguardo si veda tra l'altro Lebel e Lorek (2008).

<sup>4</sup> Le imprese di produzione diretta incentrano la loro attività è la trasformazione materiale di ricchezza mediante l'impiego di macchinari e della forza lavoro per ottenere beni economici finiti idonei allo scambio. Le imprese di servizi operano nell'ambito della circolazione della ricchezza, aumentando l'utilità dei beni preesistenti e favorendone lo scambio. Le imprese di servizi operano per la produzione di servizi a scopo di lucro atte a fornire prestazioni che favoriscono lo svolgimento delle attività di produzione diretta o indiretta.

si dei dati sempre più sofisticate, tali sistemi prendono il nome di *knowledge base system* (KBS) (Leo Kumar, 2019).

Un KBS non va interpretato solo come un contenitore di conoscenza su un dominio specifico, bensì come un modello operativo che riproduce il comportamento di un esperto in base ad una serie di osservazioni (*input*) provenienti dai fenomeni esterni o sulla base degli *output* desiderati.

Al fine di chiarire i principi di funzionamento di un KBS si rimanda alla successiva modellizzazione (Fig. 1.1).

Come si evince dalla figura il principio di funzionamento di un sistema di gestione della conoscenza<sup>5</sup> muove anzitutto dall'identificazione dei dati da raccogliere e successivamente dalla loro classificazione e contestualizzazione rispetto al campo in cui opera l'azienda ed agli obiettivi strategici che si è posta di raggiungere.

Il filone di studi sui KBS rappresenta un punto di partenza necessario per i successivi sviluppi nel campo dei *big data* e della *big data analytics*.

Facendone un breve *excursus* è bene ricordare che i sistemi di gestione della conoscenza hanno visto sostanzialmente l'avvicinarsi di due visioni:

1. Apprendimento supervisionato umano. Questo approccio propone che il processo di acquisizione della conoscenza sia ascritto ad un processo di estrazione della conoscenza da un esperto umano per addestrare il KBS. In pratica, si cerca di riprodurre l'esperienza dell'esperto posto dinanzi ad una certa situazione problematica, successivamente si traduce il linguaggio naturale in una serie di istruzioni elaborabili dall'applicativo.

A questo filone sono ascrivibili moltissimi lavori dei primi anni '90, tra questi ricordiamo a titolo esemplificativo, O'Leary (1997), Hayes-Roth e Jacobstein (1994). È bene già sottolineare come alcuni fra questi autori avessero messo in luce la difficoltà a realizzare pienamente questo approccio, sottolineando ad esempio, come fosse difficile trovare accordo nel gestire un problema fra l'esperto ed il KBS;

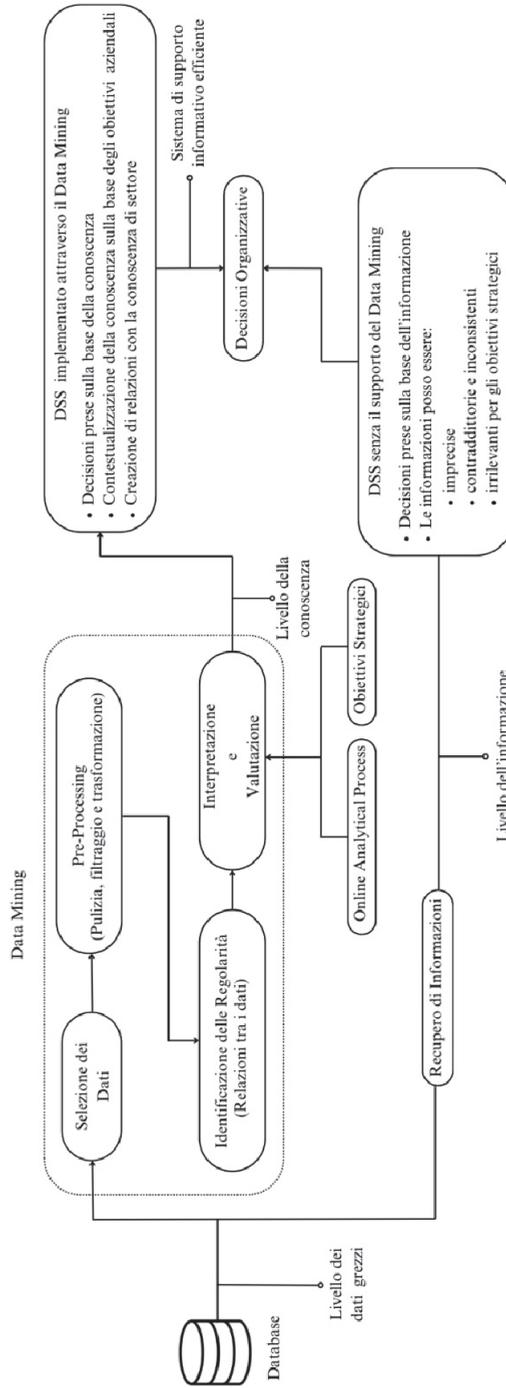
2. Apprendimento per elaborazione automatica. Questa prospettiva considera i KBS come modelli di elaborazione automatica delle informazioni, vale a dire dei veri e propri modelli di linguaggio (Large Language Model, LLM). Il cambio di paradigma ascrive i KBS non più a contenitori pieni di conoscenza mediata dall'esperienza di un esperto, ma ad un modello operativo che restituisce un comportamento desiderato osservando e ragionando di volta in volta sulla natura degli stimoli e dei dati che gli vengono sottoposti.

A questa concezione moderna possiamo ascrivere molti lavori a noi vicini temporalmente, per citarne alcuni, Khan *et al.* (2021), Al Khamisi *et al.* (2019), De-Grancey e Audouy (2023). Si precisa che ad oggi, tale filone è ancora in divenire ed i lavori, seppur numerosi, presentano evidenze discordanti e campi d'applicazione

---

<sup>5</sup> La conoscenza è l'insieme di nozioni teoriche e pratiche risultanti dall'assimilazione ed elaborazione di informazioni durante un processo di apprendimento in un settore.

Figura 1.1. Modellizzazione del funzionamento di un KBS



Fonte: ns. elaborazione.

disomogenei non solo nelle discipline del marketing e del management, ma anche nell'apprendimento, nella formazione, ecc.

Alla luce di quanto detto finora, è chiaro che il processo di gestione delle informazioni mediante KBS pone accento rilevante sulla dimensione del patrimonio informativo aziendale, basti pensare che attraverso una corretta gestione del flusso informativo le aziende di produzione diretta, ad esempio, hanno potuto implementare sistemi di produzione di beni materiali aderenti alle aspettative dei consumatori, differenziando maggiormente i prodotti (Pech e Vrchota, 2022), mentre, le aziende di servizi hanno sviluppato nuovi modelli di gestione della conoscenza (Jin e Choi, 2019) in cui i beni intangibili ad alta densità informativa trovano nuovi canali di diffusione, elaborazione, trasporto e vendita (Van Criekingen *et al.*, 2022).

Tale processo prende il nome di economia dell'informazione, in cui l'informazione diviene sia motore delle decisioni aziendali che *asset* economico. All'interno di questo paragrafo non ci dilungheremo nella spiegazione delle implicazioni che questo fenomeno comporta, trattandolo nei successivi. Per il momento è bene ricordare che il concetto ha vissuto diverse risemantizzazioni teoriche a partire dagli studi di Machlup (1962) e poi, come avvenuto per altri concetti tipici del marketing e del management, grazie all'evoluzione del mondo digitale ed ai lavori di alcuni autori come Trushkina *et al.* (2020) e Vdovenko *et al.* (2019), la tematica è arrivata sino ai giorni nostri carica di nuove accezioni.

Come si evince dai temi sino ad ora trattati, l'informazione diviene un elemento caratterizzante delle strategie aziendali. Tali strategie, come intuibile dalla Figura 1.1, prevedono un processo di gestione dei dati che va dal *data mining* alla costruzione di *dataset* pertinenti agli obiettivi aziendali.

Doveroso anzitutto è fare una distinzione fra ciò che chiamiamo dato e ciò che chiamiamo informazione. Per dare una prima definizione di dati mutuamo due accezioni date dal dizionario di Oxford, la prima qualifica il dato come funzione, "Ciascuno degli elementi di cui si dispone per formulare un giudizio o per risolvere un problema"; la seconda qualifica il dato come unità, "singola informazione codificabile o codificata". Ciò che viene messo in luce da tali definizioni è l'importanza dei dati nell'attività decisionale degli individui, delle organizzazioni e dei sistemi tecnologici.

Nel caso di informazione invece, lo stesso dizionario ci restituisce una definizione che soddisfa le nostre aspettative per quanto concerne l'accezione dell'informazione come "trasmissione e ricezione di messaggi relativi a notizie o nozioni ritenute utili o addirittura indispensabili per l'individuo o la società", mentre, scavando affondo nella definizione, si evince che l'informazione diviene sinonimo di dato. Tale aspetto non deve sorprenderci del tutto, in quanto, come ricordato ci troviamo nell'ambito di una società iper-connessa in cui la maggior parte degli stimoli che ci vengono posti ha una natura digitale, dunque, per loro stessa struttura rappresenta un insieme organizzato di unità minime (dati) in un formato che assume una connotazione informativa (Bornschein *et al.*, 2020).

Il processo sinteticamente descritto prende il nome di *datafication* o datificazione, recuperando l'impostazione data da Sarra (2022), possiamo qualificarla come "fenomeno, tipico della società attuale, attraverso cui mediamo e traduciamo porzioni di realtà (oggetti, esperienze, problemi, ecc.) in rappresentazioni simboliche, i dati, organizzati sulla base delle aspettative di conoscenza sul campo di applicazione selezionato, in forme sempre più sofisticate".

Per comprendere l'importanza di tale prospettiva nelle nostre discipline, basti pensare alle tipologie di dati con cui entriamo in contatto quotidianamente, scambiamo in maniera anche inconsapevole e che le imprese raccolgono sistematicamente.

Di seguito proponiamo una sintetica rassegna delle principali tipologie di formati di dati, precisando che ad esse sarà data più ampia trattazione nei paragrafi dedicati all'analisi degli standard tecnologici. Nello specifico possiamo parlare di (Cunha *et al.*, 2021):

1. Dati demografici; vale a dire informazioni attinenti alla popolazione (età, genere, occupazione, reddito, istruzione, ecc.);
2. Dati economici; vale a dire statistiche economiche di tipo macro (PIL, occupazione, inflazione, ecc.) e micro (imprese, redditività, ecc.);
3. Dati ambientali; vale a dire informazioni su clima, inquinamento, biodiversità, ecc.

Le possibilità fornite dalla *datafication* di generare dati su tutte queste forme di informazione, in formati disomogenei e di estrapolarne conoscenza e relazionarle per definire nuove forme di significazione, grazie alle tecnologie digitali, ci consente di introdurre il concetto di *big data*.

In questa sede, si intende fornire una disamina concisa di cosa intendiamo nelle scienze manageriali come *big data* e quali siano, attraverso le loro caratteristiche, i punti di arrivo di questo fenomeno (Staegemann *et al.*, 2020).

L'obiettivo è sottolineare che nell'ambito della digitalizzazione nascono processi come la *big data analytics* (BDA) attraverso cui le aziende tracciano in maniera puntuale i momenti in cui viene generato valore, sia internamente che esternamente (Mohamed *et al.*, 2020).

I dati trattati all'interno di un processo di BDA divengono informazione necessaria a riorientare interi processi organizzativi che altrimenti risulterebbero difficili da riconfigurare. I processi di BDA nascono in risposta alle sfide dei dati moderni, troppo vasti, veloci e destrutturati per essere gestiti con metodi tradizionali.

Difatti, la capacità di ricavare informazioni significative e vantaggi competitivi da enormi quantità di dati è diventato sempre più importante per le organizzazioni di tutto il mondo per realizzare il pieno valore dei *big data*, migliorare le prestazioni aziendali e aumentare la quota di mercato (Mikalef *et al.*, 2020).

Andando con ordine, iniziamo qualificando prima il soggetto della nostra analisi, i *big data*, e successivamente il processo attraverso cui li analizziamo, cioè la *big data analytics*. Per descrivere i *big data*, di seguito, forniremo ai lettori gli

strumenti necessari a comprendere la natura del problema che richiede alle aziende un processo organizzativo dedicato.

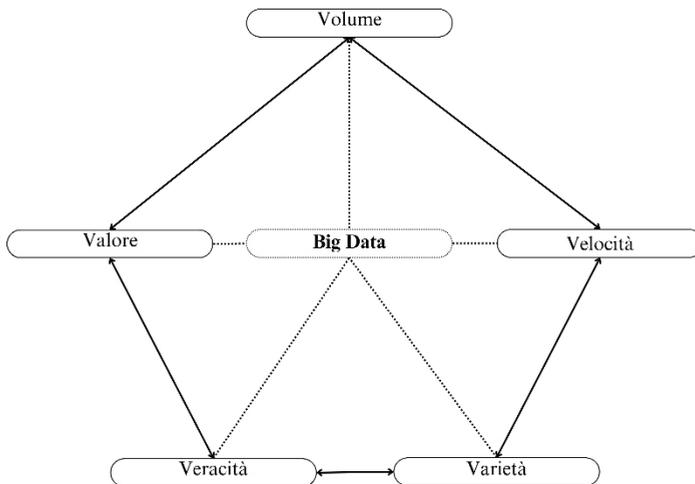
Per quanto il termine *big data* ci evochi modernità, è bene ricordare che esso trova prima applicazione a partire dai primi anni del 2000 all'interno degli scritti di Diebold che fa riferimento a un'imminente esplosione della quantità di dati disponibili e potenzialmente rilevanti (Diebold, 2003).

Pochi anni dopo, Laney svilupperà un modello per spiegare le caratteristiche chiave dei *big data*, comunemente noto come modello delle 3 V (Laney, 2001). Ulteriori approfondimenti in merito sono stati poi offerti dai lavori di Kushwaha *et al.* (2021), Wiener *et al.* (2020) e degli stessi Diebold (2003) e Laney (2017) che nel corso della loro produzione scientifica sono ritornati sul tema.

A partire dai lavori sopracitati abbiamo assistito nel corso dell'ultimo decennio ad una produzione scientifica abbondante sul tema, in particolare da Laney (2001) in poi abbiamo avuto una proliferazione di modelli inerenti alle caratteristiche dei *big data*, le 3V, le 4 V, le 8 V, ecc., fino a spingere alcuni ricercatori a chiedersi "quante "V" ci sono nei *Big Data*?", ipotizzando tramite revisioni della letteratura modelli a 19 V (Marx, 2013).

Per quanto i lavori in merito ai modelli teorici siano di vitale importanza nelle discipline del management e del marketing, in questa sede ci limiteremo a circoscrivere il perimetro di studio dei *big data* nell'ambito delle cinque caratteristiche più rilevanti. Le caratteristiche a cui facciamo riferimento sono schematizzate all'interno della Figura 1.2.

Figura 1.2. Schematizzazione del modello a 5V dei *big data*



Fonte: ns. elaborazione.

Come si evince dalla schematizzazione teorica, ogni elemento costituente la definizione di *big data* si relaziona costantemente con gli altri, dunque, per quanto la dicitura di “grande” (big) qualifichi il concetto da un punto di vista terminologico, è sbagliato dal punto di vista epistemologico porre tali dimensioni in un’ottica gerarchica.

Le caratteristiche vengono sinteticamente riportate di seguito (Khan *et al.*, 2019):

(i) Volume. È una caratteristica che fa riferimento alla quantità dei dati esistenti. Se il volume dei dati è sufficientemente ampio, può essere considerato un *big data-set*. Tuttavia, è bene segnalare che ciò che viene considerato *big data* è relativo e cambia a seconda della potenza di calcolo, della tipologia di dati e del mercato di riferimento (Ghasemaghaei *et al.*, 2020; Lv *et al.*, 2019);

(ii) Velocità. Fa riferimento alla rapidità con cui i dati vengono generati e a quella con cui si muovono. Si tratta di un aspetto importante per le organizzazioni, in quanto ad una rapidità di accesso generalmente corrisponde una maggiore disponibilità del dato stesso, consentendo di prendere decisioni aziendali più efficienti ed efficaci a raggiungere gli obiettivi preposti (Alhroob *et al.*, 2020; Christophides *et al.*, 2020);

(iii) Varietà. Si riferisce alla diversità dei formati in cui si presentano i dati. Infatti, un’organizzazione può ottenere dati da diverse fonti (interne ed esterne), ai quali di volta in volta va attribuito valore in modo diverso. La sfida posta da questa caratteristica pone il problema della standardizzazione (dati strutturati, semi-strutturati, non-strutturati<sup>6</sup>) e della distribuzione dei dati raccolti (Hajjaji *et al.*, 2021, Hassan *et al.*, 2020; Tamiminia *et al.*, 2020);

(iv) Veridicità. Si riferisce alla qualità, all’accuratezza, all’integrità e alla credibilità dei dati. I dati raccolti potrebbero essere mancanti, imprecisi o non in grado di fornire informazioni veritiere e preziose. Dunque, la veridicità può divenire una metrica del livello di fiducia nei dati raccolti. Da un punto di vista organizzativo, spesso esistono delle soglie di veridicità dei dati entro i quali stabilire quali dati sono idonei ad essere inseriti nel KBS aziendale e trattati all’interno di un processo di BDA (Shukla *et al.*, 2020; Reimer *et al.*, 2019; Lozano *et al.*, 2020);

(v) Valore. Fa riferimento ai vantaggi che i *big data* possono offrire e si riferisce direttamente a ciò che l’azienda può tradurre in decisioni strategiche. La capacità di trarre valore dai dati è un requisito fondamentale per i processi di BDA, poiché il va-

---

<sup>6</sup> I dati strutturati sono dati organizzati in un archivio formattato, ciò implica che i dati sono più facilmente indirizzabili per un’efficace elaborazione ed analisi.

I dati semi-strutturati sono dati che non sono stati organizzati in un archivio specializzato, ma hanno informazioni associate ad essi mediante la meta-datazione (questo li rende più facili da elaborare rispetto ai dati non strutturati).

I dati non strutturati non sono organizzati e spesso si presentano in formati disomogenei, essi non sono adatti a un *database* relazionale tradizionale perché non si adattano ai modelli di dati convenzionali.

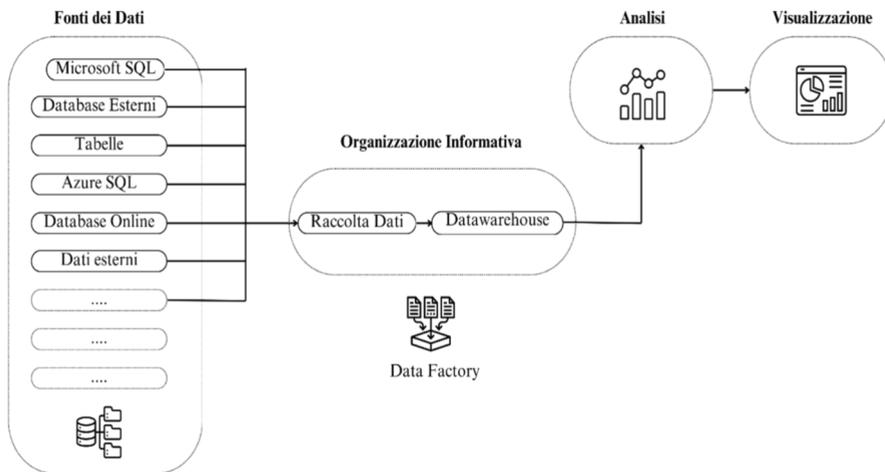
lore dei *big data* aumenta in modo significativo solo a seconda della capacità di porre in relazione i dati tra loro. Esempio concreto di questo processo è la ricostruzione delle aspettative del target sull'impresa, mediante l'accumulazione di dati di varia natura sul comportamento, sui bisogni e sui desideri dei consumatori attraverso il processo di profilazione (Aljumah *et al.*, 2021; Monino *et al.*, 2021; Dong *et al.*, 2020).

Partendo da queste premesse, emerge in maniera chiara la problematica che si riflette sul tema dell'integrazione dei *big data* all'interno dei processi aziendali, e di come i sistemi tradizionali non siano oramai idonei a gestire la rivoluzione digitale che ha accelerato tali processi. Attraverso la rete, infatti, le imprese hanno dovuto ripensare la propria *value chain*, re-ingegnerizzandola, ampliandola e rendendola virtuosa sfruttando i processi di *big data analytics* (Bag *et al.*, 2020).

Il problema più difficile da risolvere per gestire efficacemente i *big data* è rappresentato dal modo in cui essi vengono immagazzinati: non è necessariamente facile gestire grandi quantità e varietà di dati. Il secondo problema da affrontare è costituito dalla tipologia di processi di analisi stessi.

Per comprendere la vitalità di tali processi, si propone nella figura di seguito (Fig. 1.3) uno spaccato del principio di funzionamento di un processo di BDA.

Figura 1.3. Sintetizzazione delle principali attività di *big data analytics*



Fonte: ns. elaborazione.

Nello specifico, la Figura 1.3 riflette sul fatto che, con lo sviluppo tecnologico in costante evoluzione e l'incremento esponenziale dei dati che fluiscono costantemente attraverso le imprese, emerge la necessità di approcci più efficienti e tempestivi per analizzare queste informazioni.

Il semplice accumulo di enormi quantità di dati non è più adeguato a formulare valutazioni puntuali e proposte nel momento opportuno. Le tradizionali metodologie e infrastrutture di gestione e analisi dati faticano, difatti, ad affrontare agevolmente queste vastissime raccolte di informazioni.

Pertanto, diventano indispensabili nuovi strumenti e approcci specializzati per l'analisi dei *big data*, così come sistemi dedicati per la conservazione e la gestione di tali volumi informativi sono ormai irrinunciabili.

L'incremento dei *big data*, mostrato in Figura 1.3, ha conseguenze che si ripercuotono su ogni fase, dalla raccolta all'elaborazione, sino alle scelte finali.

In risposta a questa sfida, Elgandy e Elragal (2016) hanno introdotto il framework conosciuto come *Big – Data, Analytics, and Decisions* (B-DAD), il quale integra strumenti e metodologie per l'analisi dei *big data* nel contesto dei processi decisionali (Elgandy e Elragal, 2016).

Questo quadro teorico connette diverse tecnologie per la memorizzazione, la gestione e l'elaborazione dei *big data*, insieme a strumenti e approcci analitici, nonché strumenti di visualizzazione e valutazione, in tutte le fasi del processo decisionale. Di conseguenza, gli sviluppi nell'analisi dei *big data* possono essere suddivisi in tre categorie: (i) archiviazione e architettura; (ii) elaborazione e analisi; (iii) acquisizione di conoscenze finalizzate al *decision making*.

Più in dettaglio, sono analizzate di seguito le tre categorie succitate.

(i) Archiviazione e gestione dei *big data*.

Quando ci si confronta con i *big data*, una delle prime considerazioni che le aziende devono affrontare riguarda la modalità e il luogo in cui verranno conservati i dati dopo la loro raccolta. I *database* relazionali, i *data mart* e i *data warehouse* rappresentano esempi di approcci tradizionali per l'archiviazione e il recupero dei dati. I dati vengono trasferiti in memoria dagli archivi di dati operativi mediante l'utilizzo di tecnologie *Extract, Convert, Load* (ETL) o *Extract, Load, Transform* (ELT) (Nwokeji *et al.*, 2021). Questi processi estraiono i dati da fonti esterne, li convertono per soddisfare i requisiti operativi e successivamente li caricano nel *database* o nel *data warehouse*. In questo modo, prima di diventare accessibili per il *data mining* e le attività analitiche online, i dati vengono sottoposti a procedure di pulizia, elaborazione e classificazione (Nwokeji *et al.*, 2021). D'altra parte, nell'ambiente dei *big data*, sono necessarie capacità analitiche magnetiche, agili e profonde (MAD), che differiscono dalle esigenze di un tradizionale *Enterprise Data Warehouse* (EDW). Innanzitutto, le pratiche tipiche di un EDW impediscono l'aggiunta di ulteriori fonti di dati prima che siano sottoposte a pulizia e integrazione. Poiché attualmente i dati sono così ampiamente distribuiti, gli ambienti di *big data* devono essere 'magnetici', attingendo da tutte le fonti di dati indipendentemente dalla loro qualità (Côte-Real *et al.*, 2019). Inoltre, considerando il crescente numero di fonti di dati e la complessità delle analisi dati, lo *storage* dei *big data* deve consentire agli analisti di sviluppare e adattare rapidamente i dati. Questa ne-

cessità richiede un *database* agile, con contenuti logici e fisici in grado di adattarsi alla veloce evoluzione dei dati (Guo *et al.*, 2022). In conclusione, dato che le attuali analisi dei dati coinvolgono approcci statistici complessi e gli analisti devono essere in grado di esplorare vaste collezioni di dati in modo dettagliato, un archivio di *big data* deve possedere profondità e fungere da potente motore di *runtime* algoritmico (Côrte-Real *et al.*, 2019).

(ii) Dopo l'archiviazione dei dati, entra in gioco l'elaborazione analitica.

Le esigenze per l'elaborazione di dati di grandi dimensioni sono quattro.

La prima necessità consiste nella rapida fase di caricamento dei dati. Dal momento che il traffico su disco fisico e sulla rete può interferire con l'esecuzione delle *query* durante il processo di caricamento, è essenziale ridurre il tempo necessario per il caricamento dei dati.

Il secondo requisito riguarda la velocità di elaborazione delle *query*. Dato che molte *query* sono cruciali per i tempi di risposta e per soddisfare le esigenze di carichi di lavoro estesi e richieste in tempo reale, la struttura di posizionamento dei dati deve essere in grado di mantenere elevati tassi di elaborazione delle *query* stesse quando il numero di richieste aumenta rapidamente.

Il terzo criterio per l'elaborazione di dati di grandi dimensioni riguarda l'efficiente utilizzo dello spazio di archiviazione. Dato che l'incremento rapido delle attività degli utenti può richiedere una capacità di archiviazione e una potenza di calcolo scalabili, la limitatezza dello spazio su disco impone una gestione attenta della fase di archiviazione dei dati, per garantire un utilizzo ottimale dello spazio.

Infine, è essenziale adattare i modelli a carichi di lavoro estremamente dinamici. Poiché ampi insiemi di dati vengono elaborati da diverse applicazioni e da diversi utenti, caratterizzati da una gamma diversificata di obiettivi e metodi, il sistema sottostante deve mostrare una grande flessibilità per gestire in modo efficace le dinamiche imprevedibili nell'elaborazione dei dati e non essere vincolato a modelli specifici di carico di lavoro.

(iii) Acquisizione di conoscenze finalizzate al *decision making*.

Oggi i manager desiderano comprendere il significato e la rilevanza dei dati e utilizzarli per prendere decisioni, non limitandosi alla mera raccolta degli stessi. L'analisi dei dati rappresenta la pratica di impiegare algoritmi per esaminare vaste quantità di dati al fine di individuare modelli, correlazioni e informazioni preziose (Figueiredo *et al.*, 2019). Inoltre, l'analisi dei dati è utilizzata per estrarre modelli e informazioni rilevanti, valide e occultate all'interno di enormi set di dati, permettendo di scoprire connessioni chiave tra le variabili coinvolte.

Di conseguenza, i processi di analisi hanno avuto un'enorme influenza sulla ricerca e sulla tecnologia, poiché i responsabili delle decisioni sono sempre più interessati a trarre insegnamento dai dati per ottenere un vantaggio competitivo (Chen *et al.*, 2023).

L'importanza dei *big data* per i *decision makers* deriva dal loro potenziale di

fornire informazioni e competenze preziose su cui basare i giudizi. Nel corso degli anni, il processo decisionale manageriale è stato un tema di rilevanza e ampio studio. In tale ambito, i *big data* stanno diventando una risorsa sempre più preziosa per i decisori. Vaste quantità di dati altamente dettagliati provenienti da diverse fonti (tra cui smartphone, carte fedeltà, piattaforme di social media, ecc.) offrono il potenziale per conferire enormi vantaggi alle imprese.

Tuttavia, specialmente nel campo del marketing e del management, ciò è possibile soltanto se i dati vengono elaborati in modo accurato per fornire significativi approfondimenti, consentendo ai decisori di sfruttare le opportunità generate dalla ricchezza di dati, in tempo reale (Alemanno, 2018).

In aggiunta, c'è da dire che le aziende sono già abituate ad esaminare dati interni come vendite, spedizioni e inventari. Tuttavia, sta emergendo una crescente necessità di analizzare dati esterni, e, in tale ambito, i *big data* possono apportare valore e informazioni cumulative importanti. Con l'aumento del numero e della varietà di dati non strutturati disponibili, diventa fondamentale formulare giudizi più informati basati su conclusioni pertinenti tratte dai dati. Per tale motivo di seguito viene analizzata la rilevanza del concetto di informazione con un focus sul suo valore economico.

## 1.2. L'informazione economica

La digitalizzazione è stata identificata come una delle principali tendenze che stanno cambiando la società e le imprese, comportando mutamenti a causa dell'adozione di tecnologie digitali nell'organizzazione (Karimi e Walter, 2015). Difatti, l'interconnessione di rete genera un nuovo *trade-off* informativo, cambiando in modo significativo la velocità di funzionamento dell'economia stessa (Morris *et al.*, 2019). Tale evidenza ha spinto studiosi e ricercatori ad introdurre ed approfondire il concetto di informazione economica in termini di ampiezza e di profondità.

Difatti, è quasi un luogo comune osservare che le imprese si trovino ad affrontare condizioni di crescente complessità: questo concetto – ripreso da molte discipline – è stato utilizzato in molteplici circostanze per indicare diverse tipologie (ad esempio, dinamico o statico) o livelli (ad esempio, sistemi o componenti) di complessità (McDermott *et al.*, 2023).

Pertanto, è necessario chiarire il significato di complessità nel contesto specifico delle imprese e della gestione manageriale. Per farlo, partiamo dalla constatazione che la complessità è spesso intesa in termini di profondità e ampiezza, o considerando entrambe allo stesso tempo.

Un'utile esemplificazione che può essere utilizzata per chiarire il concetto è quella di analizzare il divario esistente tra complessità cognitiva (che attiene la profondità) (Angelakis *et al.*, 2022) e complessità relazionale (che attiene l'ampiezza) (Karyawati *et al.*, 2020).

La complessità in profondità, difatti, si riferisce alla sofisticazione analitica di un argomento, che diventa complesso a causa della difficoltà cognitiva di spingere la questione in analisi ai suoi estremi logici. D'altro canto, la complessità in termini di ampiezza si riferisce alla diversità di ambiti che devono essere esplorati per sviluppare un determinato argomento. Tale complessità sorge dal tentativo di conciliare modalità o temi che sono difficilmente compatibili, anziché dall'astrazione e dall'attenzione – spesso laboriosa – associate alla profondità.

La profondità, dunque, riguarda la quantità di sofisticazione, mentre l'ampiezza si riferisce al grado di variabilità.

Le imprese, come altri sistemi, sono considerate capaci di affrontare la profondità e l'ampiezza della complessità. Questo concetto ha salde radici nelle discipline del management e del marketing. A tal proposito, Galbraith definì l'azienda, indipendentemente dalle sue dimensioni, "organizzazione complessa" (Galbraith, 1977).

L'interconnessione tra la complessità tecnica, intesa come motore di progresso economico e le diverse interpretazioni dei concetti di dati, informazioni e conoscenza, costituisce un campo di studio cruciale nell'ambito dell'economia e dell'analisi dell'informazione.

In primo luogo, considerando la definizione proposta, la complessità tecnica emerge come un elemento chiave nella spinta al progresso economico. La nostra proposta consiste nel delineare uno spazio tecnologico globale, definendo metriche di complessità che classificano le aziende in questo contesto. Tale approccio implica l'utilizzo di micro-informazioni raccolte attraverso tecnologie d'analisi, mettendo in evidenza l'importanza della raccolta dettagliata di dati nel contesto tecnologico stesso (Nepelski e De Prato, 2020).

Successivamente, poniamo l'accento sulla diversità tecnologica quale elemento distintivo che manifesta l'esclusività e l'originalità del portafoglio tecnologico di un'azienda rispetto alle altre. L'approccio *firm-specific* delineato si basa sull'idea che la diversità tecnologica rappresenta non solo una caratteristica distintiva, ma anche un fattore determinante nel posizionamento strategico delle aziende sul mercato globale.

Passando all'analisi delle definizioni di dati, informazioni e conoscenza, emerge una visione articolata di come questi concetti siano interconnessi. La definizione di dati, come tutto ciò che può essere percepito e trasformato in una forma simbolica, spesso digitale, costituisce il punto di partenza. Allo stesso tempo, l'informazione è identificata come un sottoinsieme di dati che possiede qualità specifiche come rilevanza, precisione, tempestività e concisione.

È importante sottolineare la natura intrinsecamente legata dell'informazione sia al mittente che al destinatario della stessa, variando in base al contesto in cui essa viene utilizzata. Questa dinamica si allinea con le prospettive contemporanee sull'informazione economica, che possono essere definite in varie modalità e con differenti gradi di precisione.

Nel cercare di definire l'informazione economica, si può, ad esempio, cercare di

esaminarla prendendo in considerazione i settori chiave a livello industriale (Gault, 2018), le singole imprese (Sardana *et al.*, 2020), le occupazioni (Papay *et al.*, 2020) oppure concentrarci sui singoli processi (Bag *et al.*, 2020). In altri termini, la complessità tecnica emerge come un fattore trainante nel progresso economico, mentre la definizione e l'analisi dei concetti di dati, informazioni e conoscenza evidenziano la complessità intrinseca dell'informazione economica. L'approccio multidimensionale proposto riflette la complessità del contesto attuale, sottolineando la necessità di considerare molteplici prospettive per ottenere una comprensione completa e accurata dei fenomeni economici e informativi che altrimenti risulterebbe estremamente disomogeneo, e dunque, porta con sé un grado molto alto di complessità.

Tale prospettiva ci spinge a scontare la problematica dei paradossi informativi che si creano quando dobbiamo effettuare delle scelte in termini di management e di marketing (Gebauer *et al.*, 2020). Tali paradossi riguardano i seguenti quesiti:

(i) stiamo diventando relativamente meno informati perché lo stock di conoscenza che possiamo conoscere cresce più velocemente di quello che possiamo apprendere?

(ii) stiamo diventando sempre meno informati su ciò che sta diventando sempre più importante?

(iii) stiamo passando da un'economia dell'informazione basata sulla conoscenza a una società della disinformazione?

Per quanto le questioni epistemologiche siano ancora aperte, in questa sede si intende darne una chiave di lettura idonea a qualificare le domande fondanti questa riflessione.

Nello specifico, è chiaro che:

(i) affinché il primo paradosso sia effettivamente verificato, è necessario mantenere uno stato di notevole asimmetria informativa. Questo si verifica ipotizzando che gli attori sociali (ivi comprese le aziende) apprendano solo attraverso la sperimentazione e occasionali fallimenti, e che tale processo di apprendimento si relazioni ad un processo digitale in continua espansione, in cui lo spazio degli stati può estendersi più velocemente di quanto siamo in grado di imparare. Quindi, l'informazione economica – che si collega a tale paradosso – implica che lo spazio in continua evoluzione sia l'economia stessa. Per cui, possiamo affermare che, sfruttando l'ambito delle possibilità dell'economia digitale, si amplierà di molto anche lo spazio economico disponibile.

(ii) Per ciò che attiene il secondo paradosso dell'informazione, quest'ultimo riguarda la crescente intangibilità della qualità degli *input* e degli *output* economici, caratteristiche che diventano sempre più difficili da quantificare, al punto che, i manager si ritrovano ad apprendere sempre meno su ciò che è rilevante per le attività aziendali, rischiando di pregiudicare le decisioni relative all'agire imprenditoriale.

(iii) L'ultimo paradosso dell'informazione è molto probabilmente generato dagli altri due. Infatti, poiché nessuna persona è in grado di valutare integralmente il ventaglio completo delle opportunità connesse alle decisioni, diventa necessario adottare vari dispositivi di semplificazione delle informazioni e di interpretazione, come teorie e modelli. Ciò potrebbe produrre un sistema di disinformazione aziendale (Petratos, 2021).

Il successo, dunque, di un'azienda è connesso alla sua capacità di bilanciare gli aspetti informativi che si trova a gestire, generando e facendo circolare la conoscenza all'interno dei processi aziendali.

### 1.3. Gli standard tecnologici della *big data analytics*

La digitalizzazione è stata identificata come una delle principali tendenze che stanno cambiando la società. Considerando la prospettiva finora analizzata, è chiaro che sono i flussi informativi e la logistica delle informazioni a definire i nuovi confini della catena del valore e della competizione, oltre che a determinare i momenti di generazione del valore che risiedono all'interno di un processo organizzativo.

Molti autori hanno sottolineato che un numero sempre più elevato di aziende sta attuando l'implementazione di piattaforme di *big data analytics* (Bharadiya, 2023; Talaoui *et al.*, 2023). D'altra parte, però, questa informazione solleva interrogativi sul motivo per cui l'adozione dei *big data* non sia ancora una realtà per tutte le aziende.

Da un lato, infatti, ci troviamo di fronte a un'enorme quantità di dati che richiede soluzioni scalabili per estrarre valore.

Dall'altro, le tecnologie legate ai *big data* sono spesso presentate come la principale risposta a questa sfida.

Questa diffidenza di prospettiva è spesso legata alla mancanza di consapevolezza su cosa costituisca effettivamente un problema di *big data analysis* e quali siano i vantaggi e gli svantaggi delle attuali soluzioni *data-driven* disponibili (Choi *et al.*, 2018).

I *big data* sono comunemente associati alla gestione e all'analisi di insiemi di dati su vasta scala, essi rappresentano l'attività di raccolta ed elaborazione di grandi quantità di dati, insieme agli strumenti e alle tecniche utilizzati per analizzare tali enormi insiemi di dati. In altri termini, come sottolineato in precedenza, questi ultimi sono definibili come enormi volumi eterogenei provenienti da fonti indipendenti, con una governance dispersa e decentralizzata, e hanno l'obiettivo di esplorare le relazioni complesse e dinamiche tra i dati.

Poiché l'elaborazione distribuita coinvolge numerosi nodi, la gestione dei *big data* deve affrontare l'eventuale inefficienza dei nodi come un evento comune,

piuttosto che considerarlo un'eccezione nel processo di elaborazione (Tabesh *et al.*, 2019).

La letteratura in merito ci suggerisce che nell'affrontare le sfide dei *big data*, vanno incorporati una serie di elementi dai quali è impossibile prescindere, quali: tolleranza agli errori, scalabilità, elasticità, architettura distribuita, archiviazione generica ed elaborazione di enormi quantità di dati (Torres *et al.*, 2018).

Un sistema di *big data* completo, arricchito da una vasta gamma di servizi e tecnologie, dovrebbe permettere anche il controllo delle risorse, la regolamentazione dei dati e la supervisione della loro integrità.

In questa sede ci concentriamo esclusivamente sulle tecnologie e sui sistemi che favoriscono tali funzionalità. In particolare, verranno confrontate diverse tecnologie *big data* e piattaforme di analisi nelle seguenti categorie: elaborazione, archiviazione, integrazione dei dati, analisi, governance e monitoraggio.

Molti autori confrontano tra loro le diverse tecnologie dei *big data* (Inoubli *et al.*, 2018). Nella maggior parte dei casi, gli autori sottolineano che prima di scegliere la tecnologia o la piattaforma migliore da utilizzare, l'organizzazione ricerca i requisiti specifici dell'applicazione o dell'algoritmo e valuta cosa ciascuna tecnologia o piattaforma può offrire.

È importante notare che obiettivo del lavoro in oggetto non è confrontare le tecnologie e le piattaforme *big data* per diverse applicazioni, come ad esempio il *Cloud Computing* e l'*Internet of Things* (IoT), ma piuttosto valutarle in base alle categorie specifiche di problemi legati ai *big data*.

A tale scopo, esamineremo brevemente, come alcune delle tecnologie *big data* più utilizzate e discusse recentemente, quali YARN/Hadoop, Spark, Flink e Hyracks/ASTERISK, assolvano varie funzioni nella gestione dei *big data*.

Apache Hadoop è diventato uno standard *de facto* per la condivisione e l'accesso ai dati e alle risorse di calcolo in vari scenari *big data*. Hadoop è un framework di calcolo *open-source* scalabile che consente di distribuire le operazioni di calcolo tra un gran numero di *server host* che non sono necessariamente dotati di processori ad alte prestazioni. Hadoop è composto da un motore di esecuzione MapReduce e un file system distribuito (DFS) noto come HDFS – *Hadoop Distributed FileSystem*. I principali vantaggi di Hadoop includono la notevole flessibilità, la scalabilità, il costo contenuto e l'affidabilità, che permettono la gestione e l'elaborazione efficiente di ingenti volumi di dati, sia strutturati che non strutturati<sup>7</sup>.

Apache Spark è un motore di elaborazione dati unificato. Utilizza un'architettura di programmazione simile a MapReduce, ma introduce un concetto di condivisione dei dati noto come *Resilient Distributed Dataset* (RDD). Con questa caratte-

---

<sup>7</sup>Hadoop offre piani di lavoro per il bilanciamento dei carichi di dati, risorse e attività. YARN, acronimo di "Yet Another Resource Negotiator", è derivato da Hadoop. La sua architettura disaccoppia il paradigma di programmazione dall'infrastruttura di gestione delle risorse, affidando a componenti specifici per l'applicazione numerosi compiti di schedulazione.

ristica, Spark è in grado di gestire una varietà di attività di elaborazione che in passato richiedevano l'uso di motori distinti, come SQL, streaming, machine learning ed elaborazione di grafi<sup>8</sup>.

Il progetto Apache Flink è un framework *open-source* per l'elaborazione di stream e batch in applicazioni distribuite e ad alte prestazioni. Esso si basa sull'idea che diversi tipi di applicazioni di elaborazione dati possano essere gestiti come flussi di dati *fault-tolerant*.

Flink utilizza lo *storage in-memory* per migliorare le prestazioni dell'esecuzione del *runtime*<sup>9</sup>.

Il framework software partizionato-parallelo Hyracks/ASTERIX è stato progettato per condurre calcoli ad alta intensità di dati su cluster massivi condivisi. Hyracks presenta un *overlay* compatibile con Yarn per l'esecuzione di processi MapReduce esistenti. Hyracks è un sistema scalabile di gestione delle informazioni in grado di memorizzare, interrogare e analizzare enormi collezioni di dati semi-strutturati annidati<sup>10</sup>.

La breve introduzione fornita in merito a YARN/Hadoop, Spark, Flink e Hyracks/ASTERISK serve a contestualizzare meglio il modo in cui queste tecnologie si avvicinano alle sfide dei *big data*, fra le quali segnaliamo principalmente: (i) gestire l'archiviazione di fonti eterogenee, tra cui dati strutturati, non strutturati e semi strutturati; (ii) estrarre conoscenza da insiemi di dati vasti ed eterogenei, applicando non solo query SQL, ma anche complessi algoritmi di machine learning o calcoli di grafi; (iii) ricevere flussi continui di dati che devono essere elaborati in millisecondi o meno.

Partendo da questi problemi, verrà esaminato, di seguito, se questi framework possono supportare le fasi di: elaborazione in streaming e in batch, l'archiviazione e l'analisi dei dati.

Partiamo con l'elaborazione in batch<sup>11</sup>.

Questo tipo di elaborazione è strettamente correlato all'elaborazione a lungo termine di una grande quantità di dati, tutti in una sola volta, per un determinato

---

<sup>8</sup> Anche Spark è stato sviluppato per superare le limitazioni del processo di input/output (I/O) su disco e migliorare le prestazioni del sistema. Esso consente di memorizzare i dati nella cache, eliminando così il vincolo dei costi dell'I/O su disco associato a YARN per le attività iterative.

<sup>9</sup> Le principali differenze tra Flink e le tecnologie precedenti per i *big data* includono: un runtime distribuito per i flussi di dati che utilizza l'esecuzione in streaming in pipeline per i carichi di lavoro batch e stream; la coerenza dello stato *exactly-once* attraverso un checkpoint leggero; l'elaborazione iterativa nativa; una sofisticata semantica a finestre che supporta l'elaborazione fuori ordine.

<sup>10</sup> Rispetto a MapReduce, Hyracks offre un modello utente più flessibile ed è un'implementazione più efficiente per i carichi di lavoro MapReduce rispetto a Hadoop in diversi casi di utilizzo ad alta intensità di dati.

<sup>11</sup> L'elaborazione in batch è il metodo utilizzato dai computer per completare periodicamente processi di elaborazione di dati ripetitivi e con volumi elevati.

periodo di tempo. Viene comunemente impiegato nelle attività di ETL (*Extract, Transform, and Load*), nell'aggregazione dei dati e nell'addestramento e nell'aggiornamento dei modelli di apprendimento automatico. Hadoop è stato ampiamente utilizzato nell'elaborazione in batch. Anche Hyracks può gestire l'elaborazione dei dati in batch. Tuttavia, Spark è diventato il motore più diffuso per l'elaborazione di enormi quantità di dati in una vasta gamma di settori, grazie alla sua capacità di elaborazione rapida dei dati in memoria, che elimina le spese generali di lettura e scrittura di Hadoop.

Analizziamo ora, il campo dell'elaborazione in tempo reale.

L'elaborazione dei flussi coinvolge il trattamento degli input e la generazione di risultati entro limiti di tempo rigorosi, di solito nell'ordine dei millisecondi o dei microsecondi, a seconda dell'applicazione e dei requisiti dell'utente. Ad esempio, Spark Streaming accetta flussi di dati in ingresso in tempo reale e li suddivide in microbatch, che vengono successivamente elaborati dal motore Spark per creare il flusso finale di risultati in batch. Il micro-batching consente di elaborare un flusso come frammenti di dati. Tuttavia, può introdurre un notevole sovraccarico sotto forma di pianificazione delle attività. Flink, al contrario, è in grado di offrire tutti i vantaggi del buffering senza i costi di programmazione delle attività. Flink può altresì operare in modo efficiente in applicazioni in tempo reale o quasi, dove le informazioni sui dati devono essere fornite quasi immediatamente dopo la loro produzione.

Focalizziamoci, a seguire, sul tema dell'archiviazione.

A tal proposito, c'è da dire che HDFS può ospitare una vasta gamma di tipologie di dati, inclusi quelli strutturati, non strutturati e semistrutturati. Hyracks può consumare dati da HDFS e offre anche la memorizzazione dei dati di AsterixDB per la gestione, l'archiviazione, l'indicizzazione, l'interrogazione e l'analisi di grandi quantità di dati mediante un modello di dati flessibile (ADM). Spark può essere integrato con una vasta gamma di sistemi di file, tra cui HDFS, MapR File System, ecc. Flink consente l'integrazione di diverse fonti di dati, inclusi dati relazionali strettamente strutturati, dati testuali non strutturati e dati semistrutturati.

Né Flink né Spark forniscono una soluzione di archiviazione principale.

Infine, abbiamo, la necessità di esplorare le peculiarità in termini di analisi dei dati. In tale ambito, infatti, YARN/Hadoop supporta attivamente diversi progetti di alto livello, come Giraph, Pig, Hive, Mahout e HBase, per costruire strumenti di sviluppo e gestire il flusso e l'elaborazione dei dati. Spark supporta anche una serie di applicazioni, come ETL, Machine Learning, ecc. Lo stack di Flink fornisce librerie API di alto livello per molti casi d'uso, tra cui Complex Event Processing, ecc. Lo stack software del sistema Hyracks comprende anche interfacce di analisi come SQL (Hivesterix), XQuery (Apache VXQuery) e Graph (Pregelix). Anche Hyracks può condurre complicate operazioni di flusso di dati distribuiti ed esprimere efficacemente intere algebre relazionali.

#### 1.4. L'evoluzione dei confini organizzativi e della *Value Chain* ai tempi della trasformazione digitale

Operare nel mercato digitale comporta per l'azienda il superamento della dimensione spaziale e temporale intesa in modo tradizionale. In particolare, c'è da dire che la tecnologia aumenta la diffusione e la quantità di informazioni accessibili per l'azienda (Schwarz Müller *et al.*, 2018). La tecnologia, dunque, passa dall'essere considerata strumento operativo ad essere un vero e proprio *asset* strategico. In altri termini, se nell'era antecedente i *big data*, si organizzavano i processi aziendali e poi vi si affiancava la tecnologia per supportarli, oggi si tiene conto a monte del processo tecnologico da implementare e della tipologia di valore che essa apporta e poi vi si adattano i processi in maniera sistemica e sinergica.

Studi recenti hanno progressivamente esaminato le varie potenzialità, ma anche le sfide, dell'innovazione digitale, tenendo presente che l'adozione dell'innovazione digitale comporta procedure complesse (come la partecipazione a ecosistemi digitali), senza dimenticare che diversi settori presentano sfide uniche in tale ambito (Subramaniam *et al.*, 2019; Barykin *et al.*, 2020).

Con la diffusione della digitalizzazione che permea tutti i settori, diviene necessario comprendere come sfruttare, efficacemente ed efficientemente, le risorse esistenti (interne ed esterne) in un ambiente in continua evoluzione.

Al fine di rispondere a queste sollecitazioni, molte aziende decidono di standardizzare procedure e pratiche organizzative. Questo iter è noto come processo di ricerca e sistematizzazione delle capacità.

In questo processo vengono create strutture organizzative per completare determinate attività direttamente collegate al prodotto o al servizio che l'azienda propone. In tale ambito, i principali processi implementati si dividono in analitici e sperimentali, tale distinzione è ben descritta dalla teoria delle capacità dinamiche, secondo cui le disparità di capacità a livello aziendale sono influenzate da tre fattori principali, analizzati di seguito.

(i) La capacità di un'azienda di modificare il proprio repertorio futuro di competenze è vincolata dal suo stock di capacità esistenti, secondo un ragionamento economico evolutivo e di *path-dependency*. In tal senso utilizziamo il termine *asset* in modo generico per fare riferimento alle risorse storiche (conoscenze, abilità tecniche, competenze organizzative, ecc.) che influenzano le possibilità di crescita futura delle capacità dell'impresa.

(ii) Le capacità di un'azienda sono fortemente influenzate dai processi aziendali. Difatti, attraverso investimenti e altri interventi manageriali, le imprese possono riconfigurare le loro attività. Tuttavia, la capacità di riconfigurazione di un'azienda è limitata. Essa dipende da un insieme di routine che determinano la flessibilità organizzativa (come, ad esempio, le strutture di governance, le procedure di allocazione delle risorse e i sistemi di gestione). Questa capacità di riorganizzare le risorse

se di un'azienda, e in particolare i processi che ne sono alla base, viene definita capacità dinamica.

(iii) La capacità di un'azienda sono strettamente condizionate dai percorsi attivati per raggiungerle. Poiché la maggior parte delle competenze sono cumulative e crescono nel tempo attraverso una successione di investimenti coordinati, richiedono l'impegno in percorsi prestabiliti. Per cui, una questione strategica cruciale per le aziende è l'identificazione di percorsi di crescita delle capacità che portino a un vantaggio competitivo. La discrezionalità manageriale nella selezione dei percorsi, insieme ai limiti imposti dalle risorse e dalle procedure di riconfigurazione, può portare a disparità nelle capacità aziendali.

I processi sopra descritti si relazionano al concetto di *Value Chain* (Porter, 1985). La catena del valore descrive, infatti, l'intero insieme di attività necessarie per portare un prodotto o un servizio dalla concezione allo smaltimento finale dopo l'uso, compresa la trasformazione fisica e l'apporto di vari servizi del produttore (Walters e Lancaster, 2000). In tal senso, la produzione è solo una delle numerose relazioni che possono riuscire ad aggiungere valore. Inoltre, ogni processo della catena è caratterizzato da una serie di attività specifiche e di informazioni e conoscenze peculiari atte a permettere la creazione del valore.

Come detto in precedenza, la digitalizzazione di tali processi e dei beni genera un'enorme quantità di dati e informazioni che richiedono un'amministrazione e un'analisi approfondite per garantire un'interpretazione accurata dei dati e per addurre valore all'intero processo aziendale.

Alla luce delle reinterpretazioni adottate dal digitale, le aziende moderne considerano la tecnologia parte integrante del processo, dunque, si parla di *digital value chain*. Un valido esempio di come tali processi si modifichino è dettato dall'introduzione del *cloud computing*, una piattaforma *cloud* è una soluzione per mantenere alta la qualità dei dati e gestire tutti i processi digitalizzati. Ciò consentirebbe, qualora fosse implementata fattivamente, la condivisione continua dei dati di processo e l'analisi automatizzata attraverso l'intera catena di processi, nonché una gestione completa della catena del valore attraverso piattaforma.

Le funzioni tipiche di una piattaforma *cloud* includono: capacità *multi-client*, scalabilità e abilità di combinare *database* esterni con un ambiente di sviluppo integrato che supporta diversi linguaggi di programmazione (es. Apache).

Per cui è chiaro che una parte fondamentale dello sviluppo della piattaforma sia quella di garantire che i dati siano raccolti in modo corretto per assicurarne la buona qualità. Questi dati grezzi servono come base per tutte le scelte e le analisi automatizzate. Infatti, per digitalizzare l'intera catena del valore, è necessario digitalizzare tutti i *touch point* dell'azienda, in modo che i loro dati possano essere collegati all'ambiente in cui essa opera e in cui operano gli stakeholder con cui l'organizzazione decide di interfacciarsi (Bousonville, 2017).

Per cui, integrando tutte le interazioni con gli stakeholder chiave, i beni e i ser-