

# BIG DATA, LA NUOVA FRONTIERA DELL'IT

CARMELA DE PADOVA\*

## SOMMARIO

1. Introduzione. - 2. Perché i Big Data. - 3. Cosa sono i Big Data. - 4. Database server per Big Data. - 5. Big Data e Analytics per la gestione dei rischi di infortunio e malattie professionali. - 6. Conclusioni.

## 1. Introduzione

L'odierna società dell'informazione e della conoscenza (c.d. dai sociologi "società post-industriale") è caratterizzata essenzialmente da un bene immateriale, l'informazione, che aggiunge "valore" ad imprese e a pubbliche amministrazioni.

Il progresso tecnologico ha poi consentito non solo l'accesso alle informazioni ma anche la possibilità di metterle in relazione tra di loro ed elaborarle per ottenerne di nuove, rendendole, pertanto, un bene sempre più "prezioso".

Parlando dello sviluppo della società della conoscenza nel Piano triennale per l'informatica dell'Istituto 2007-2009 scrivevo: "L'evoluzione dell'Information Technology, peraltro, ha provocato negli anni una duplicazione e frammentazione dei dati. Per ovviare a tale inconveniente, le aziende di software hanno introdotto nel mercato strumenti di Business Intelligence (BI)<sup>1</sup> che collezionano, sistematizzano ed interpretano dati ed informazioni. Secondo quanto previsto da Gartner, inoltre, è in costante aumento l'interesse delle aziende per l'insieme delle informazioni che costituiscono la propria conoscenza (Knowledge) e per le modalità che consentono di gestirla più efficacemente (Management), affinché

\* Consulenza per l'Innovazione Tecnologica - Direzione Generale Inail.

<sup>1</sup> La Business Intelligence consente la razionalizzazione e l'utilizzo di dati che vengono trasformati in informazioni, permette di ottenere una visione globale e completa dell'andamento di business attraverso strumenti di sintesi, indicatori di prestazioni aziendali, capaci cioè di rappresentare le informazioni provenienti dai diversi processi aziendali, senza implicare una esplorazione diretta degli stessi, permettendo agli organi direttivi di prendere decisioni ottimali per il raggiungimento degli obiettivi.

sia possibile riutilizzarla, fornire un ampio accesso ai propri dipendenti e incorporarla nelle attività aziendali, nonché commercializzare il proprio capitale intellettuale in servizi, prodotti e proprietà intellettuale. Non si tratta di semplice gestione delle informazioni, ma di un nuovo obiettivo di portata avanzata che comporta l'integrazione del Knowledge Management (KM)<sup>2</sup> nella gestione delle informazioni”.

Da allora abbiamo assistito a molti mutamenti a livello globale. L'accesso alle informazioni, ad esempio, avviene tramite sistemi tecnologici sempre più intelligenti e sempre più interconnessi.

Secondo IBM<sup>3</sup> “stiamo assistendo a grandi cambiamenti nel mondo: organizzazioni di ogni tipo stanno investendo in nuove tecnologie e nuovi processi che le rendono più efficienti, agili e competitive. Il mondo sta diventando sempre più intelligente, nel senso che la tecnologia e le interconnessioni stanno determinando un'esplosione di dati. Sistemi potenti per l'analisi e l'utilizzo di questi dati forniranno al mondo un nuovo tipo d'intelligenza: un'intelligenza che non solo permetterà di sviluppare il business, ma ci aiuterà a risparmiare energia, a migliorare il rendimento dell'agricoltura e a ridurre l'impatto dei disastri naturali”.

Da quando abbiamo a disposizione un'immensa mole di dati da analizzare e condividere e sui quali basare il nostro approccio alla conoscenza, la nostra vita è cambiata in modo significativo.

È il cosiddetto universo dei Big Data, la nuova frontiera dell'Information Technology che sta rivoluzionando la società dell'informazione e della conoscenza.

Questo immenso archivio, oltre a raccontarci con precisione chi siamo e come operiamo nel mondo, con basso margine d'errore, ha cambiato il nostro modo di creare, condividere e sviluppare conoscenza, in particolare quella scientifica.

## 2. Perché i Big Data

Come sopra detto, la digitalizzazione a livello mondiale, la diffusione dell'utilizzo delle nuove tecnologie e l'interconnessione stanno determinando una esplosione di dati.

Oggi una grande quantità di imprese che operano nel web, quali ad es. i gestori di social network, le società che offrono servizi e contenuti sui media digitali, le banche e le compagnie telefoniche, le autostrade e i produttori di navigatori GPS, le compagnie aeree e quelle delle carte di credito, raccolgono enormi quantità di dati sulle persone.

2 Per Knowledge Management si intende la sistematica, esplicita e deliberata costruzione, applicazione e rinnovamento della conoscenza per massimizzare l'efficacia della base conoscitiva di un'azienda ed i relativi benefici.

3 IBM, *Un pianeta più intelligente è anche più sicuro*.

Le *smart city*<sup>4</sup>, le pubbliche amministrazioni, i servizi *machine-to-machine*<sup>5</sup> con sensori, le etichette intelligenti basate su tecnologia RFID<sup>6</sup> collezionano a loro volta enormi quantità di dati.

Perché tanto interesse verso i Big Data? Il tema, inizialmente relegato al settore della ricerca e della scienza, ambiti in cui sono gestiti grandi database di immagini, archivi sismici e così via, è divenuto di interesse anche per le grandi aziende. La possibilità di catturare e sfruttare, compatibilmente con le leggi vigenti<sup>7</sup>, queste nuove fonti eterogenee di dati, quali ad es. video, audio, e-mail, sms, social network, dati GPS, RFID, per offrire servizi tempestivi legati al contesto e, soprattutto, la possibilità di utilizzo di nuove tecnologie per l'analisi dettagliata di tutti i dati raccolti con ottenimento dei risultati in tempo reale possono creare indubbi vantaggi per ogni tipo di organizzazione.

In un periodo in cui l'economia è in evidente difficoltà, la possibilità di prendere decisioni basandosi su informazioni e analisi aggiornate può portare enormi benefici alle imprese per i propri processi decisionali.

Le applicazioni di Big Data, sviluppate per migliorare le capacità analitiche di un'organizzazione, sono nella maggior parte dei casi guidate dagli *Analytics*<sup>8</sup>.

I Big Data consentono, ad esempio, alle organizzazioni di ottimizzare le loro performance e i processi di business, aumentare la fedeltà dei clienti e il customer care, ridurre i costi di sviluppo e, infine, migliorare la monetizzazione dei loro dati.

### 3. Cosa sono i Big Data

Ad oggi non esiste una definizione precisa, univoca e concorde di Big Data. Tutti sappiamo che i Big Data sono sistemi che memorizzano o elaborano enormi quantità di dati, strutturati e non, cosa impensabile sino a qualche anno fa.

4 La città intelligente è un ambiente urbano in grado di migliorare la qualità della vita dei propri cittadini anche grazie all'impiego diffuso e innovativo delle tecnologie dell'informazione e comunicazione, in particolare nell'ambito della comunicazione, della mobilità, dell'ambiente e dell'efficienza energetica (fonte Wikipedia).

5 In generale ci si riferisce a tecnologie ed applicazioni di telemetria e telematica che utilizzano le reti wireless. Indica anche un insieme di software ed applicazioni che migliorano l'efficienza e la qualità dei processi tipici di ERP, CRM e asset management (fonte Wikipedia).

6 Radio Frequency Identification è una tecnologia per l'identificazione e/o memorizzazione automatica di oggetti, animali o persone basata sulla capacità di memorizzazione di dati da parte di particolari dispositivi elettronici (detti tag) e sulla capacità di questi di rispondere all'"interrogazione" a distanza da parte di appositi apparati fissi o portatili chiamati per semplicità "lettori" (in realtà sono anche "scrittori") a radiofrequenza comunicando (o aggiornando) le informazioni in essi contenute (fonte Wikipedia).

7 Codice privacy (D.Lgs. n. 196/2003), Legge sulla trasparenza (L. n. 241/90, D.Lgs. n. 33/2013).

8 Gli Analytics sono strumenti di analisi dei dati che hanno una potenza significativamente elevata e che abilitano anche analisi di tipo predittivo, capaci, cioè, di proporre un'azione immediata sul cliente in base ai risultati ottenuti.

Gartner definisce Big Data<sup>9</sup>: *high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.*

Wikipedia definisce Big Data “grandi aggregazioni di dati, la cui mole richiede strumenti differenti da quelli tradizionali in tutte le fasi del processo (dalla gestione, alla *curation*<sup>10</sup>, passando per condivisione, analisi e visualizzazione)”. La dimensione dei Big Data può raggiungere anche l’ordine di 10<sup>21</sup> Byte (detto Zettabyte) e, quindi, di miliardi di Terabyte (già 10<sup>12</sup> Byte).

Le caratteristiche principali<sup>11</sup> dei Big Data si possono riassumere nelle 3 “V”:

- **Volume:** perché per ottenere analisi dettagliate con un alto grado di confidenza servono grandi quantità di dati da analizzare;
- **Velocità:** perché è essenziale effettuare analisi e ottenere risultati in tempo reale;
- **Varietà:** perché oltre alla quantità, servono numerose fonti diverse (strutturate e non).

Secondo Colin White<sup>12</sup>, i Big Data sono una modalità di trattamento dei dati il cui volume aumenta esponenzialmente, grazie anche alle nuove tecnologie *Cloud*<sup>13</sup>, e che non possono, quindi, essere gestiti utilizzando applicazioni analitiche e sistemi Data Warehouse tradizionali (un esempio è costituito dai dati correlati alle informazioni di Facebook e Twitter) per l’inadeguatezza delle tecnologie o per i costi troppo elevati.

Esistono diverse forme e dimensioni di Big Data, così come esistono varie tipologie di utilizzo, tra cui, ricerca in tempo reale di tentativi di frode, analisi competitiva su web, ottimizzazione delle attività dei call center, analisi dei social media, gestione intelligente del traffico<sup>14</sup>.

Grazie all’evoluzione tecnologica è stato possibile realizzare negli ultimi anni ciò che un tempo sarebbe stato impensabile per la gestione di questi enormi archivi di dati provenienti da fonti così eterogenee.

9 Big Data è un patrimonio informativo ad alto volume, velocità e varietà che richiede forme innovative e vantaggiose di elaborazione delle informazioni atte a migliorare capacità analitiche dettagliate e processo decisionale.

10 Data Curation è la gestione dei dati atta a scoprire, recuperare, mantenere, aggiungere valore, permettere il riutilizzo dei dati garantendone contemporaneamente l’autenticazione, l’archiviazione, la gestione e la preservazione (definizione della Graduate School of Library and Information Science dell’Illinois).

11 Fonte Wikipedia.

12 Fonte Colin White, presidente e fondatore della BI Research, analista, educatore e scrittore statunitense noto per la sua profonda conoscenza della Business Intelligence, della gestione dei dati e delle tecnologie di integrazione delle informazioni.

13 Cloud è un insieme di tecnologie informatiche che permettono l’utilizzo di risorse hardware (storage, CPU) o software distribuite in remoto, senza che l’utente ne percepisca la complessità.

14 Fonte Colin White.

In sintesi, per lavorare con i Big Data occorre utilizzare le seguenti tecnologie<sup>15</sup>:

- **nuovi sistemi di gestione dei dati** in grado di gestire un'ampia varietà di dati rilasciati da sensori, web, social media (detti anche dati multi-strutturati);
- **nuovi strumenti di analisi dei dati** (chiamati anche *Advanced Analytics*) di tipo predittivo e di analisi dei testi;
- **hardware più veloce e performante** ottenibili grazie a processori multi-core e spazi di indirizzamento memoria più grandi così come dischi allo stato solido e storage multilivello per la gestione di dati caldi e freddi.

Come tutti sanno, a decorrere dagli anni '90 la maggior parte delle applicazioni analitiche utilizzavano dati strutturati (in generale database relazionali) estratti dai sistemi operazionali usati per popolare i Data Warehouse.

Come sopra evidenziato, la caratteristica dei Big Data è, invece, la crescita esponenziale del numero di sorgenti di dati, la varietà dei contenuti, il significativo volume dei dati nonché l'utilizzo di dati multi-strutturati per distinguerli dai dati operazionali strutturati.

Il problema dei sistemi di Big Data è la complessità che si cela dietro la loro gestione, complessità che supera i limiti attuali delle piattaforme e dei sistemi dedicati alla elaborazione e analisi dei dati.

#### 4. Database server per Big Data

Con l'avvento dei Big Data negli ultimi anni è cambiato fortemente il mercato dei database server.

Database management systems (DBMS) relazionali classici quali IBM DB2, Microsoft SQL Server, MySQL e Oracle dominanti sino a cinque anni fa oggi non sono più adeguati per la gestione e l'analisi dei dati dei Big Data. La mole dei dati dei Big Data richiede, infatti, strumenti differenti da quelli tradizionali in tutte le fasi del processo.

Come già detto, i Big Data comprendono anche l'interrelazione di dati provenienti da sorgenti eterogenee e, quindi, non solo da database relazionali ma anche da file di testo o altro.

Lo sviluppo dei Big Data con database classici, oltre ad essere molto complesso, comporterebbe costi troppo elevati, probabilmente insostenibili e con scarse prestazioni.

Questo segnale è stato colto da molti dei nuovi vendor che hanno realizzato database server più adatti ai sistemi di Big Data e a prezzi più ragionevoli.

<sup>15</sup> COLIN WHITE, *What is Big Data and why do we need it?*.

Questi nuovi database server possono essere suddivisi in tre gruppi: NoSQL, NewSQL e Analytical SQL.

Si riporta di seguito la descrizione di ciascun gruppo di database server fornita dall'analista Rick van der Lans<sup>16</sup> con elenco dei principali prodotti oggi presenti nel mercato che consentono di sviluppare e gestire sistemi di Big Data:

- **NoSQL.** Sono database server che non supportano né SQL, né concetti relazionali puri. Questo gruppo può essere suddiviso nelle seguenti quattro sottocategorie:
  - memorizzazione per valori chiave (MemcachedDB, Hamsterdb, Oracle Berkeley DB, LevelDB, Redis, AreospikeDB e Riak),
  - memorizzazione di documenti (Apache CouchDB, Apache Jackrabbit, MongoDB, OrientDB, Terrastore e RavenDB),
  - memorizzazione a colonne (Apache Cassandra, HBase, Hypertable (Google's Bigtable) e Amazon SimpleDB),
  - memorizzazione a grafo (AllegroGraph, FlockDB, HyperGraphDB, InfiniteGraph e Neo4j).

Molti di questi supportano API e un modello dati che rendono più facile alle applicazioni la memorizzazione e il recupero dei dati. La maggior parte supporta un modello di dati aggregati, in cui i dati possono essere memorizzati in strutture a nido e gerarchiche, mentre i record nelle tabelle possono avere strutture differenti. Grazie al modello non relazionale, a una architettura snella e semplice, a un differente meccanismo transazionale, tali server offrono prestazioni elevate, alta scalabilità e disponibilità e costi di archiviazione contenuti, rendendoli così molto adatti per particolari sistemi di Big Data.

Un prodotto che vale la pena menzionare è Hadoop, un file system distribuito a elevata scalabilità e architettura che poggia su MapReduce che, pur non essendo un database server, è alla base di molti dei citati server di database NoSQL.

- **NewSQL.** Sono database server che supportano SQL tradizionale e sono detti "nuovi" server di database, in quanto internamente diversi; esempi di prodotti NewSQL sono Akiban, Clustrix, GenieDB, JustOneDB, MemSQL, NuoDB, ScaleBase, TransLattice, VMware SQLFire e VoltDB.
- **Analytical SQL.** Sono database server che mirano ad analizzare i Big Data, al contrario dei server New SQL che si concentrano su operazioni con analisi semplici in tempo reale; alcuni di questi server sono progettati specificatamente per dati scientifici in grado di eseguire le query più complesse su massicce quantità di dati mantenendo però buone prestazioni. Esempi di prodotti che appartengono a questa categoria sono: Actian VectorWise, Exasol,

<sup>16</sup> RICK VAN DER LANS, *Big Data is changing the database server market*. Rick Van der Lans è un analista consulente olandese indipendente specializzato in Data Warehousing, Business Intelligence, Service Oriented Architecture e tecnologie database.

EMC/Greenplum, HP/Vertica, IBM/Netezza, InfoBright, Information Builders HyperStage, Kognitio, Oracle ExaLytics, ParAccel, SAP HANA e le Appliances Teradata Aster Database. Questi sono tutti i database server SQL progettati e ottimizzati per l'elaborazione analitica su grandi insiemi di dati.

La necessità di sistemi di Big Data ha avviato un mercato completamente nuovo di database server: alcuni forse non sopravvivranno alla sfida, altri si affermeranno e diventeranno noti in futuro come lo sono oggi alcuni dei DBMS relazionali classici.

## 5. Big Data e Analytics per la gestione dei rischi di infortunio e malattie professionali

Il responsabile dei sistemi informativi dell'Inail, dr. Stefano Tomasini, ha presentato nel maggio 2013 a Forum PA un progetto di Big Data che l'Istituto intende avviare per la gestione dei rischi di infortunio e malattia professionale.

Tale progetto prevede l'utilizzo di strumenti di *Advanced Analytics* per analisi descrittive della situazione storica, analisi diagnostiche della situazione attuale, analisi predittive.

Per monitorare l'efficacia delle policy assicurative, verificandone il livello di attuazione e gli esiti, l'Inail intende individuare nuovi e più evoluti strumenti che consentano di:

- **definire il modello** del patrimonio informativo di riferimento per tutti gli operatori del Welfare, a livello nazionale e internazionale;
- **realizzare la banca dati** che costituisca il **Welfare Information Capital Center (WICC)** censendo, qualificando e integrando tutte le fonti che si renderanno disponibili;
- **pubblicare i contenuti della WICC**, corredati da glossari e metadati esplicativi, per condividere dati e informazioni, la cui qualità è certificata secondo metodi e misure pubblicamente condivisi (WikiWelfare<sup>17</sup>).

Per svolgere in modo ottimale i propri compiti istituzionali l'Inail necessita del più ampio dominio di informazioni su "Lavoratori" e "Aziende", ma anche di dati estesi ad aspetti socio-ambientali.

L'obiettivo è quello di effettuare analisi che evidenzino le correlazioni fra infortuni/malattie professionali e contesto locale, comportamenti aziendali, generatori di stress, stile di vita del lavoratore, azioni di prevenzione e quant'altro.

<sup>17</sup> Cittadino che sia anche stakeholder attivo del welfare, ossia contribuisca a formarlo, plasmarlo, in un'ottica win-win di condivisione (wiki) della prestazione e del benessere sociale che vada oltre-il-Pil.

## 6. Conclusioni

Big Data e Network Knowledge sembrano essere i paradigmi che regoleranno il funzionamento del nostro universo nei prossimi anni, trasformando il nostro modo di vivere, lavorare, pensare e le nostre prese di decisioni.

Il report<sup>18</sup> della McKinsey Global Institute del maggio 2011 sottolinea che i Big Data hanno la possibilità di incrementare il valore del settore sanitario degli Stati Uniti di 300 miliardi di dollari, di aumentare quello del settore pubblico europeo di 250 miliardi di dollari, di ridurre del 50% i costi di produzione relativi a sviluppo e assemblaggio, di far crescere di 100 milioni di dollari il fatturato dei service provider e, infine, di incrementare del 60% il margine netto del commercio al dettaglio negli Stati Uniti.

In un'inchiesta<sup>19</sup> commissionata da Capgemini agli inizi del 2012, che ha coinvolto 607 dirigenti di alto livello, i partecipanti hanno evidenziato che, nei processi in cui è stato applicato un approccio analitico su Big Data, le organizzazioni hanno ottenuto un miglioramento medio del 26% delle prestazioni rispetto ai tre anni precedenti e prevedono di ottenere un ulteriore miglioramento del 41% nei successivi tre anni.

Da un survey realizzato per Fujitsu da Coleman Parker Research è emerso che il 49% dei senior manager aziendali intervistati si è detto convinto che entro il 2015 i Big Data avranno un ruolo fondamentale nel pianificare il business aziendale e valutarne i risultati e costituiranno un elemento chiave nel modificare il ruolo stesso in cui le aziende opereranno sul mercato e realizzeranno i propri affari<sup>20</sup>. Di certo i Big Data e i modelli di previsione basati sulla statistica stanno già cambiando i processi decisionali e diventeranno sicuramente la chiave di volta per la crescita dell'innovazione, competizione e produttività.

Ma i Big Data stanno anche diventando un problema importante da risolvere al più presto. Man mano che vengono raccolte più informazioni e che aumentano i collegamenti alla rete internet, tutto diventa un "problema da Big Data", anche il semplice controllo della rete stessa.

È evidente che i Big Data hanno un grosso potenziale, ma è altrettanto evidente che le criticità da superare sono tante e, in particolare, quelle relative alla complessità di gestione che non è legata solamente alla mole dei dati ma anche alla loro localizzazione, alla loro eterogeneità, alla estrema frammentazione delle fonti informatiche.

<sup>18</sup> MCKINSEY GLOBAL INSTITUTE, *Big Data: the next frontier for innovation, competition, and productivity*.

<sup>19</sup> ECONOMIST INTELLIGENCE UNIT, *The Deciding Factor: Big Data & Decision Making*.

<sup>20</sup> Fonte Fujitsu.



## RIASSUNTO

La digitalizzazione a livello mondiale, la diffusione dell'utilizzo delle nuove tecnologie e l'interconnessione stanno determinando una forte esplosione di dati. Sistemi potenti per l'analisi e l'utilizzo di questi dati forniranno un nuovo tipo d'intelligenza che non solo permetterà di sviluppare il business, ma ci aiuterà a risparmiare energia, a migliorare il rendimento dell'agricoltura e a ridurre l'impatto dei disastri naturali.

È il cosiddetto universo dei Big Data, la nuova frontiera dell'Information Technology che sta rivoluzionando la società dell'informazione e della conoscenza.

Perché tanto interesse verso i Big Data? Il tema, inizialmente relegato al settore della ricerca e della scienza, ambiti in cui sono gestiti grandi database di immagini, archivi sismici e così via, è divenuto di interesse anche per le grandi aziende. La possibilità di catturare e sfruttare nuove fonti eterogenee di dati, quali ad es. video, audio, e-mail, sms, social network, dati di GPS, RFID, per offrire servizi tempestivi legati al contesto e, soprattutto, la possibilità di utilizzo di nuove tecnologie (*Analytics*) per l'analisi dettagliata di tipo predittivo di tutti i dati raccolti con ottenimento dei risultati in tempo reale possono creare indubbi vantaggi per ogni tipo di organizzazione.

In un periodo in cui l'economia è in evidente difficoltà, la possibilità di prendere decisioni basandosi su informazioni e analisi aggiornate può portare enormi benefici alle imprese per i propri processi decisionali.

Cosa sono i Big Data? Ad oggi non esiste una definizione precisa, univoca e concorde di Big Data. Tutti sappiamo che i Big Data sono sistemi che memorizzano o elaborano enormi quantità di dati, strutturati e non, cosa impensabile sino a qualche anno fa.

Caratteristica dei Big Data è la crescita esponenziale del numero di sorgenti di dati, la varietà dei contenuti, il significativo volume dei dati nonché l'utilizzo di dati multi-strutturati per distinguerli dai dati operazionali strutturati (database).

Il problema dei Big Data è la complessità che si cela dietro la loro gestione, complessità che supera i limiti attuali delle piattaforme e dei sistemi dedicati alla elaborazione e analisi dei dati.

## SUMMARY

Digitising at a global level, the spread of the use of new technologies and interconnection are causing a powerful explosion of data. Powerful systems for analysis and the use of this data will provide a new type of intelligence that will not only assist the development of business but will help us to save energy, improve agricultural yield and reduce the impact of natural disasters.

This is the so-called Big Data universe, the new frontier of Information Technology that is revolutionising the world of information and knowledge.

Why so much interest in Big Data? The theme, initially relegated to the field of science and research, areas in which large image databases, seismic archives and so on are managed, is also becoming popular with large companies.

The ability to capture and exploit new heterogeneous sources of data, such as e.g., video, audio, e-mail, sms, social networks, GPS data, RFID to provide timely services related to context, and, above all, the possibility of using new technologies (Analytics) for detailed predictive analysis of all the data collected resulting in the achieving of real time results can lead to undoubted advantages for all types of organisation.

In a period in which the economy is experiencing significant crisis, the ability to make decisions based on updated analysis and information can bring enormous benefits to companies and organisations in terms of their decision-making processes.

What is Big Data? To date, there is no precise, unambiguous and commonly agreed definition of Big Data. We all know that Big Data are systems that store or process huge amounts of data, structured or otherwise, a notion that would have been unthinkable up until a few years ago.

A feature of Big Data is the exponential growth of the number of data sources, the variety of content, the significant volume of data and the use of multi-structured data, distinguishing it from structured operational data (databases).

The problem with Big Data is the complexity involved in its management, complexity that exceeds the current limits of systems and platforms dedicated to the processing and analysis of data.