**Big Data**

**Che cos’è Big data:**

È la **quantità massiva di dati che vengono utilizzati da aziende/organizzazioni per creare servizi/prodotti** per i clienti e/o creare modelli di business. È caratterizzato dal volume, la varietà e la complessità dei dati.

**L’importanza:**

Il valore dei Big data dipende dalla misura in cui si lasciano analizzare e comprendere.

**L'AI**, il **machine learning** e le moderne tecnologie di database consentono l'analisi e la rappresentazione grafica dei Big Data per ottenere insight su cui agire, in tempo reale.

**Tipologie di Big Data:**

* **Dati strutturati**: è il più semplicemente da organizzare e consultare (sql). Essendo questa tipologia la più facile da organizzare e consultare. Non si qualificano necessariamente come Big data
* **Dati non strutturati:** i dati di questo tipo non possono essere inquadrabili nei normali database relazionali di righe e colonne. I dati di questo tipo vengono archiviati in database NoSql
* **Dati semi-strutturati:** è una forma ibrida delle due precedenti. Per esempio una mail contiene dati strutturati come il mittente, destinatario e il contenuto della mail che è un tipo di dato non strutturato

**Fonti dei Big data:**

* **Dati dei social:** sono generati da post, commenti, video che circolano sui social media
* **Dati macchina:** I dispositivi e le macchine collegati a una rete IoT sono dotati di sensori e hanno la capacità di trasmettere e ricevere dati digitali. I sensori IoT permettono alle aziende di raccogliere ed elaborare i dati macchina provenienti da diversi dispositivi
* **Dati transazionali**: sono quelli in più rapida evoluzione e crescita al mondo. Per esempio tutte le transazioni di acquisto e bancarie del mondo, otteniamo volume sbalorditivo di dati che vengono generati.

**Le 5 V che definiscono il Big data:**

* **Volume:** sono necessari algoritmi avanzati e guidati dall’AI
* **Velocità**: oggi la tecnologia dei Big Data consente ai database di elaborare, analizzare e configurare i dati nel momento stesso in cui vengono generati, a volte nel giro di millisecondi.
* **Varietà:** I Big Data sono tipicamente costituiti da combinazioni di dati strutturati, non strutturati e semi-strutturati (maggiore flessibilità).
* **Veridicità**: per un’azienda, i Big data sono utili solo ed esclusivamente se accurati, pertinenti e puntuali.
* **Valore:** l'esito di tali analisi deve essere l'emersione di conoscenze che possano aiutarle a diventare più competitive e resilienti, e a servire meglio i propri clienti.

**Vantaggi dei Big data:**

* **Esperienza del cliente**
* **Manutenzione predittiva**
* **Sviluppo di prodotti e servizi**
* **Risparmi sui costi e maggiori efficienza**
* **Miglioramenti della competitività**

**Tecnologie per l’archiviazione di dati semi e non strutturati:**

* **Data lake:** vasta quantità di dati che non sono stati elaborati
* **Data warehouse:** è un database relazionale in cui i dati sono strutturati o pre-modellati in tabelle. La struttura e lo schema dei dati sono stati progettati per ottimizzare le query SQL veloci
* **No Sql:** schema flessibile che può essere modificato per adattarsi alla natura dei dati da elaborare.

**Come funzionano i Big data:**

I Big data una volta analizzati, si traducono in insight pertinenti e fruibili, capaci di migliorare il business in modo misurabile. I tre passaggi principali per l'utilizzo dei Big Data:

1. **Raccogliere i Big Data:** la maggior parte di questi dati sono semi o non strutturati e quindi sarebbe molto difficile poter integrare questi dati nei tradizionali database
2. **Archiviare i Big Data:** molte aziende utilizzano i cloud per archiviare i Big data
3. **Analizzare i Big Data:** utilizzo delle AI e machine learning