

BIBLIOTECA

Big Data per una PA data driven

Dati delle PA come patrimonio digitale

I dati delle pubbliche amministrazioni sono uno dei principali patrimoni digitali della PA.

Per questo il Piano Triennale per l'Informatica nella Pubblica Amministrazione di AgID include tra gli obiettivi strategici da perseguire la valorizzazione del patrimonio informativo pubblico nazionale.

La Piattaforma Digitale Nazionale Dati

Uno degli strumenti pensati per il raggiungimento di questo obiettivo è la Piattaforma Digitale Nazionale Dati (PDND), precedentemente nota con il nome di Data & Analytics Framework (DAF), definita all'art. 50-ter del Codice dell'Amministrazione Digitale.

La Piattaforma Digitale Nazionale Dati/2

Con questa piattaforma si intende sviluppare e semplificare l'interoperabilità dei dati pubblici tra PA, sfruttando le opportunità offerte dalla gestione e l'analisi dei big data.

Cosa sono i Big Data

Ma cosa sono i Big Data? La definizione di Big Data non è univoca, perché cambia a seconda degli ambiti. La più comune si riferisce all'enorme quantità di dati che l'attività contemporanea, di un cittadino o di un'organizzazione produce.

Alcuni esempi sono i

dati sul clima, le immagini satellitari, le immagini e i video digitali, le registrazioni di operazioni o i segnali GPS.

Oggi quando si parla di Big Data, ci si riferisce anche all'insieme di tecnologie e metodi analitici di cui si ha bisogno per estrarre valori di interesse.

Le 5 V

Per descrivere i Big Data partiamo da 5 caratteristiche che li contraddistinguono. Le cosiddette 5 V: Volume, Velocità, Varietà, Variabilità, Veridicità.

Vediamole nel dettaglio.

Le 5 V: Volume

Il volume è sicuramente la caratteristica più famosa associata ai big data e si riferisce all'enorme quantità di dati che vengono prodotti ogni giorno. Per trattare e gestire questi dati spesso i requisiti di elaborazione superano le capacità di un singolo computer. Si crea quindi una nuova sfida per mettere in comune, allocare e coordinare le risorse da gruppi di computer.

Le 5 V: Velocità

Con velocità si intende la rapidità con cui le informazioni si muovono attraverso il sistema. Talvolta questi dati, che arrivano da più fonti, hanno bisogno di essere elaborati in tempo reale.

Le 5 V: Varietà

La varietà è la caratteristica che si riferisce alle numerose fonti da cui i dati provengono: applicazioni e server, social media, dispositivi ecc. Ma non importa da dove essi provengano: le tecnologie big data cercano di gestirli tutti con un unico sistema.

Le 5 V: Variabilità

Per variabilità si intende la difficoltà che a volte può sorgere nel gestire i grandi flussi di dati che arrivano. L'interpretazione dei dati può infatti cambiare in base ai contesti di raccolta e analisi.

Le 5 V: Veridicità

Molteplici fonti e elaborazioni complesse possono comportare difficoltà nel valutare i dati come veri. Questo rimane quindi uno degli aspetti più complicati da gestire soprattutto nel caso di dati sensibili.

L'evoluzione della B: Data Warehouse e Data Lake

L'avvento dei Big Data ha semplificato la raccolta di nuove tipologie di informazioni, sia all'interno che all'esterno dell'organizzazione.

Questo passaggio ha portato all'evoluzione del Data Warehouse tradizionale nei cosiddetti Data Lake.

Cos'è il Data Warehouse

Il Data Warehouse è un componente di sistema più convenzionale. Un grande archivio di dati puliti, ordinati e integrati con altre fonti, che possono essere utilizzati per l'analisi e/o la reportistica.

Il Data Warehouse viene perciò solitamente considerato come la fonte di dati certificati all'interno dell'organizzazione.

Cos'è il Data Lake

Il Data Lake è solitamente un grande deposito di dati raccolti a uno stato relativamente grezzo. Questo termine viene spesso utilizzato per fare riferimento ai dati che potrebbero essere non strutturati e spesso soggetti a variazioni.

Opportunità offerte dai Big Data

Detto questo, indipendentemente dalla definizione specifica di big data e delle loro caratteristiche, è fondamentale sottolineare quanto i dati siano diventati una vera e propria ricchezza per la possibilità di poterne ricavare informazioni di valore, sia per efficientare processi interni che per impostare nuove politiche di sviluppo.

Big Data per una PA Data Driven

I Big data consentono di creare conoscenza per coloro che devono prendere decisioni e ridurre drasticamente i tempi di analisi.

La scalabilità orizzontale di queste tecnologie permette di estrarre informazioni dall'incrocio di molteplici basi di dati e di processare dati in tempo reale consentendo di avere più prospettive di analisi su un dato fenomeno.

La sesta V: Valore

In questo senso è giusto aggiungere una sesta V alla definizione di Big Data. E' Valore.

Lo scopo ultimo dei Big Data infatti è quello di fornire valore aggiunto

Conclusioni

Riassumendo:

- I big data sono una grande mole di dati, che se elaborati possono aiutare le pubbliche amministrazioni a definire strategie di sviluppo
- I big data si caratterizzano per 5 dimensioni: Volume, Velocità, Varietà, Variabilità, Veridicità. A queste si aggiunge Valore come sesta dimensione

- L'evoluzione dei big data ha portato ai data lake ovvero a un deposito di dati raccolti a uno stato relativamente grezzo.