

SPECIALE Big Data in Sanità

Big data:

UN ESEMPIO
D'USO

**Come utilizzare
un approccio big
data per ottenere
maggiore efficienza nel
contesto sanitario?**

**Di seguito uno
schema progettuale
di applicazione,
pensato per una
struttura sanitaria
di riabilitazione
post operatoria, che
mostra con chiarezza i
benefici che si possono
ottenere**

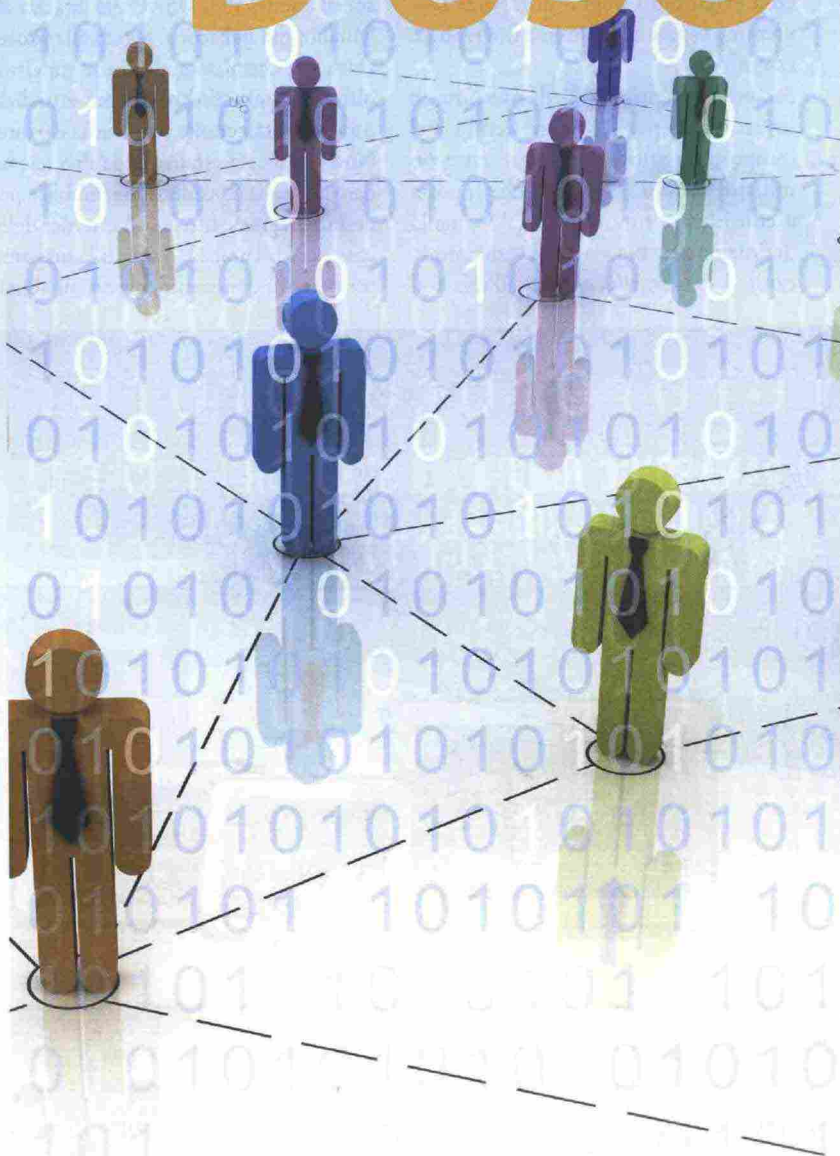
Di Luca Rodolfi (*)

(*) Luca Rodolfi si occupa a tempo pieno di business intelligence dal 1996. Ha maturato esperienze lavorative che gli hanno consentito di sviluppare numerose e diverse competenze sia in ambiti progettuali e tecnologici che in ambiti organizzativi e manageriali, tutti accomunati dal "filo rosso" della business intelligence. Attualmente è responsabile della divisione di business intelligence di Sb Italia e collabora con l'università Cattolica del Sacro Cuore su tematiche di corporate performance management. luca.rodolfi@sbitalia.com

RIFERIMENTI

Il presente documento è una sintesi ragionata di diverse pubblicazioni oltre che di valutazioni personali dell'autore. Tra le pubblicazioni che hanno contribuito alla definizione dei contributi si ricordano:

Il rapporto Big Data Analytics di Phil Russom
Understanding Big Data di Chris Eaton e altri
Diversi IBM Redbooks relativi alle tematiche Big Data





Il mondo della sanità è uno dei settori in cui i big data potrebbero dare i contributi più rilevanti nel prossimo futuro. Il caso proposto riguarda un'ipotetica struttura ospedaliera esperta nella riabilitazione di pazienti che hanno subito eventi traumatici; di fatto il paziente dopo

le cure e gli interventi effettuati in regime ospedaliero viene assegnato a questa struttura perché intraprenda un percorso di riabilitazione attraverso sedute, esercizi, fisioterapia, applicazioni medicali, ecc, naturalmente seguito da medici, infermiere, fisioterapisti, ecc.

Poiché tale struttura ottiene dei rimborsi sulla base dei costi che deve sopportare rispetto alla tipologia di azione di cura (immaginiamo rimborsi del SSN se pensiamo all'Italia o a sistemi di assicurazione se pensiamo ad altre realtà come gli Stati Uniti) il compito del progetto di big data è raccogliere dati per comprendere se il costo forfettario che viene riconosciuto sulla base di una sorta di tariffario per prestazione è corretto o se invece i costi che devono essere sostenuti sono superiori di quelli ipotizzati.

Semplificando, un progetto del genere dovrebbe avere tre macrofasi:

1. **Predisporre un sistema in grado di generare informazioni utili;**
2. **Predisporre un sistema in grado di "catturare" tali informazioni;**
3. **Predisporre un sistema in grado di analizzare i dati e individuare possibili indicazioni utili al business.**

GENERAZIONE DELLE INFORMAZIONI

Nel nostro ipotetico caso si presuppone che il costo maggiore che la struttura ospedaliera deve sopportare sia rappresentato dal costo dei professionisti che vi lavorano: medici, infermiere, fisioterapisti, tecnici di laboratorio, ecc... L'ipotesi forte è che

ogni paziente, occupando per un certo tempo ogni singolo professionista, genera un costo pari al costo per quella frazione di tempo dedicata a lui del professionista. Per estremizzare se il costo mensile di un medico fosse di 8.000 euro e questi dedicasse tutto il suo tempo per un mese ad un singolo paziente questo avrebbe già generato per se stesso un costo di 8.000 euro.

Nel progetto quindi si dotano tutte le persone (pazienti e personale medico) di un rilevatore che indica gli spostamenti all'interno dell'ospedale (potrebbe essere un dispositivo GPS o un dispositivo RFID), questo rilevatore ogni secondo genera alcune informazioni:

1. luogo fisico dove si trova la persona (in forma di longitudine e latitudine, o attraverso una mappatura dei locali dell'ospedale);
2. ruolo della persona (medico, infermiere, ecc...);
3. data - ora - minuto - secondo della rilevazione;
4. ...

Ipotizziamo che nella struttura di riabilitazione ci siano 100 pazienti e 500 professionisti dell'ambito sanitario, e che ogni singola informazione prodotta abbia una dimensione di 1 kilobytes.

Una semplice operazione ci consente di valutare in un giorno quanti dati vengono raccolti tenendo conto delle dimensioni, della frequenza e del numero delle persone. Indicativamente il sistema produce circa 21 gigabytes al giorno (ipotizzando 10 ore di attività ospedaliera), un monitoraggio di 6 mesi si traduce in circa 4 terabytes di dati raccolti.

A questi dati poi si potrebbero aggiungere anche una serie di informazioni di tipo più anagrafico e clinico per completare il quadro di riferimento di ogni paziente in cura.

Come Volume siamo probabilmente nel limite inferiore per considerare il sistema

SPECIALE Big Data in Sanità

come un sistema di big data, ma valutando il rapporto tra numero di soggetti coinvolti (600) e il volume generato, possiamo certamente considerare che quasi 6 giga di dati "posizionali" a persona sia una quantità sterminata di dati pro capite.

CATTURA DELLO STREAM DI DATI

Avere 600 dispositivi che ogni secondo generano informazioni rappresenta un punto di attenzione nella definizione dell'architettura che dovrà poter raccogliere tali informazioni. Nel caso specifico i volumi e le frequenze possono essere, con una certa facilità, raccolte e archiviate in un sistema ad hoc: la "soluzione" tecnologica individuata (in connessione con i dispositivi stessi) catturerà ed archiverà le informazioni fornite.

DATAMINING, ANALISI PREDITTIVA E ALGORITMI "INTELLIGENTI"

L'intero progetto ha senso solo se la fase di analisi dei dati raccolti è correttamente progettata e il sistema in grado di individuare pattern di dati e informazioni di rilievo per il business. Nel caso specifico è possibile individuare due distinte tipologie di analisi. Una tipologia quasi "classica" in cui i dati raccolti possono

essere analizzati quantitativamente per fare analisi, ad esempio, sui tempi medi, per tipologia di problema del paziente, erogati dai professionisti rispetto ad un singolo paziente: è un'analisi fondamentale, non sofisticata ma di grande rilievo. Consente di validare o meno le assunzioni sul fatto che una particolare attività di riabilitazione per uno specifico problema sia correttamente valorizzata come costo. Si potrebbe ad esempio scoprire che alcuni tipici problemi (come la riabilitazione dopo incidenti che hanno causato gravi problemi al bacino del paziente) necessitano di molte più cure di quanto ipotizzato o della presenza costante di diverse figure professionali, ecc.

Una tipologia avanzata, più da "big data", in cui un'analisi esplorativa del dato attraverso strumenti e modelli di datamining potrebbe evidenziare informazioni non note e nemmeno individuabili, piuttosto che proporre soluzioni di analisi predittiva per specifiche situazioni grazie alle serie storiche raccolte, scoprendo magari che la probabilità di uno specifico impegno in termini di cure da erogare per situazioni particolari (ad esempio pazienti donne, con età superiore a 60 anni, con problemi di salute come diabete, artrite, ecc.) il sistema potrebbe infatti individuare in anticipo soggetti candidati a cure supplementari perché rientrano in un pattern individuato dall'analisi "estesa" dei dati raccolti. Naturalmente il modello di riferimento dovrebbe definire delle regole per capire quando un professionista si sta dedicando ad uno specifico paziente. Potrebbero essere definite regole per cui se un medico si trova entro X metri dal paziente per più di Y minuti il suo costo viene attribuito a quel paziente, ecc.

I vantaggi sarebbero rilevanti, la struttura ospedaliera potrebbe dimostrare con i dati raccolti quali specifiche cure sono sottovalutate a livello dei costi fornendo così informazioni che potrebbero facilmente essere generalizzate e estese ad altre strutture. L'investimento in un simile progetto

avrebbe un ROI facilmente calcolabile nel momento in cui le correzioni successive alla valutazione dei dati raccolti consentirebbero un più equo riconoscimento dei costi per le prestazioni e d'altro canto un simile progetto verrebbe innescato nel momento in cui fosse palese il rapporto sconveniente per l'istituto ospedaliero tra attività e costi sostenuti e rimborsi concordati ricevuti.

CONCLUSIONE: COMPNDERE IL FENOMENO BIG DATA

Nell'ambito IT, big data è uno dei termini più alla moda del momento: carico di aspettative sia in ambito tecnologico che economico, complesso per sua natura da comprendere e da spiegare, mistificato come nuova ventata rivoluzionaria nell'ambito della business intelligence. È insomma un fenomeno che va compreso: questo ci siamo proposti di fare con questo articolo, soprattutto con l'esempio citato, di ottimizzazione delle performance in un contesto sanitario. Ora vogliamo chiarire ulteriormente il concetto, tornando ancora per un poco nella parte teorica.

Probabilmente non esiste peggior nome di "big data" per definire l'insieme di dati che, grazie ad alcune peculiarità, rientrano in tale ambito. Questo perché il termine "big" nell'ambito informatico è quanto di più "relativo" si possa concepire: quello che è big oggi sarà semplicemente small fra alcuni mesi (o pochi anni).

Tale termine va quindi considerato in una modalità dinamica legata allo scorrere del tempo e al miglioramento delle tecnologie, come a dire per big s'intende per quell'unità di tempo.

Oggi, alla fine del 2015, possiamo ipotizzare che in presenza di centinaia di terabytes di dati si entri nell'ambito dei big data, domani, in una data imprecisata del prossimo futuro, potremo magari ipotizzare che l'ambito dei big data riguardi volumi di centinaia di petabytes.

Fatta questa premessa quasi terminologica, definire cosa è big data non è poi

Il termine "big" nell'ambito IT è relativo: quello che è big oggi sarà semplicemente small fra qualche anno

così complesso ed è strettamente legato all'aumento esponenziale di informazioni disponibili "nell'etere" grazie alla fortuna di molti dispositivi come gli smartphone, i tablet, i sistemi gps, i computer (questi ultimi, per la verità, da molti anni protagonisti nella generazione di dati).

Non si deve comunque pensare ai big data essenzialmente come ai dati "globali" accessibili da tutti attraverso la rete. È importante invece focalizzare l'attenzione sul concetto di flusso dati, perché è nella cattura e analisi di questo flusso che risiede il vantaggio della gestione dei big data. Si pensi ad esempio a tutte le informazioni che ogni secondo vengono generate in un aeroporto (che certamente non possiamo considerare come dati "accessibili" a tutti), quali vantaggi potrebbero nascere dalla possibilità di catturare queste informazioni e (senza la preoccupazione di dover gestire grandissime quantità di dati) innescare con algoritmi "intelligenti" una

ricerca che possa fornire indicazioni di rilievo nel miglioramento dei processi legati al funzionamento dell'aeroporto stesso.

Ma quali informazioni di rilievo potranno mai celarsi all'interno di milioni o miliardi di "pacchetti" di dati? Non c'è il rischio di fare investimenti per raccogliere tali dati (strutturati e non) per poi scoprire che non hanno un valore utile per il business? Non è meglio insistere con l'analisi del "dato classico" organizzato e pensato per tale scopo (ossia il data warehouse con gli strumenti di querying e reporting)?

Il volume "Understanding big data" dà una definizione interessante relativamente alla differenza tra il dato classico che siamo abituati ad analizzare e i big data. L'idea è immaginare una miniera d'oro dalla quale (guarda caso) si vuole estrapolare tutto l'oro presente. Ci saranno pepite mediamente grandi che è possibile ricercare in una modalità classica: il minatore e il suo "piccone", egli sa cosa cercare, più o meno

dove e riconosce la pepita quando la trova, tali pepite sono le informazioni classiche della business intelligence e hanno un grande valore rispetto alla "dimensione" della pepita stessa.

Accanto alle pepite ci sarà anche una "polvere" d'oro, sparsa e distribuita un po' ovunque, che non si può recuperare nel modo classico usando un piccone e soprattutto avrà un valore molto ridotto se pensiamo ad un minuscolo granello d'oro. Tale valore però potrebbe aumentare moltissimo se esistesse un modo di estrapolare da tutta la miniera tutta la polvere d'oro, e milioni o miliardi di minuscoli granelli potrebbero diventare un'enorme pepita. Il paradigma dei big data è tutto qui (più o meno): non si sa esattamente cosa cercare, ma si sa che il numero spropositato di dati, se fosse catturato e opportunamente "setacciato" (con algoritmi di analisi e data-mining) potrebbero rivelare informazioni rilevanti per le performance aziendali.