



Abbonamento n. 1074 valido dal 20/03/2013 al 19/03/2014 - Licenza esclusiva a MIMESI SRL

ESPERIENZE - REPLY

L'evoluzione dell'IT ed i Big Data: un unico percorso!

I BIG DATA HANNO CATTURATO L'ATTENZIONE DELL'IT NEGLI ULTIMI ANNI, CON UN ECO PARAGONABILE A QUELLO AVUTO DA INTERNET 10 ANNI PRIMA. UNA DELLE RAGIONI PER CUI QUESTO FENOMENO È COSÌ IMPORTANTE È PERCHÉ LE ARCHITETTURE CHE STANNO EMERGENDO CON ESSO CONSENTONO DI GESTIRE, IN MODO EFFICACE, DATI DI DIVERSI FORMATI E STRUTTURA SUPERANDO I VINCOLI POSTI DAI SISTEMI E DAI DATABASE TRADIZIONALI

Che cosa sono veramente i Big Data? Talvolta il termine identifica il proliferare delle informazioni dovuto all'aumento delle attività sulla rete, o la generazione di dati attraverso Social Network, Tablet e Smartphone, o le interazioni Machine to Machine. Ma dobbiamo riconoscere come per molti aspetti il termine "Big Data" sia infelice. Difatti il valore che quest'ampia gamma di tecnologie, di nuova generazione, porta alle organizzazioni è la capacità di ampliare le tipologie di informazioni e velocizzare i processi con cui operare, decidere, ridurre i rischi e servire efficacemente i propri clienti. Secondo noi la vera essenza del valore introdotto dalle tecnologie Big Data è la varietà, la velocità e la flessibilità da applicare ai dati; più che il solo volume di

dati da processare. Non bisogna essere una pure-internet-company alla "Facebook", per avere vantaggi dall'utilizzo di questo nuovo paradigma.

La prospettiva Tecnologica: la crisi delle architetture tradizionali

Il tema della gestione ed interpretazione dei dati è sempre stato critico, ed è stato affrontato negli anni costruendo un ecosistema di soluzioni di Information Technology basate sul cosiddetto "modello relazionale".

Questo ecosistema sembra oggi non essere più in grado di evolvere per gestire (anche) i Big Data. Ciò soprattutto per:

- L'esplosione dei volumi, che ha messo a nudo l'incapacità delle architetture tradizionali di fornire dei tempi di risposta adeguati rispetto agli investimenti richiesti;

- La varietà dei dati (strutturati, multi-strutturati, destrutturati), che si scontra con la rigidità del modello relazionale, tipicamente basato su schemi prefissati.

Proprio la necessità di gestire questa complessità sta mettendo in discussione, dopo trent'anni di dominio assoluto, il modello relazionale: la cui struttura basata su tabelle e relazioni tra tabelle (Schema on Write) mal si adatta ad un'effi-

Guido Guerrieri, Executive Partner, Reply



cace gestione dei "Big Data", spesso destrutturati o provenienti da fonti eterogenee. I nuovi modelli, identificati genericamente come No-SQL database, forniscono la flessibilità necessaria a caricare i dati nella loro forma atomica, differendo l'applicazione di uno schema al momento dell'interrogazione (Schema on Read), con un evidente incremento della flessibilità di utilizzo e con delle prestazioni garantite da architetture a parallelismo massivo basati su hardware standard e non specializzato (quindi tendenzialmente a basso costo).

L'ascesa di Hadoop

Tra i dominatori della scena dei Big Data, vi è Hadoop: un framework open source del-

la APACHE Open Software Foundation che supporta applicazioni distribuite in grado di lavorare con migliaia di nodi e grandi moli di dati. Hadoop è stato ispirato da quanto realizzato da Google, Yahoo ed altre internet-company per la raccolta e l'indicizzazione di milioni e milioni di link e contenuti web accessibili dai motori di ricerca. Oggi Hadoop è la componente principale di molte offerte nell'ambito Big Data, incluso come componente architeturale standard da quasi tutti i brand di rilevanza mondiale nel settore IT. Il successo crescente di Hadoop è testimoniato dall'elenco delle organizzazioni che lo hanno adottato: non solo grandi Internet Company come Facebook o Goo-

gle, molte delle aziende della classifica Fortune 500.

Questa architettura ha anche rilanciato il mercato del lavoro professionale, creando opportunità per architetti Big Data (su infrastrutture, sviluppi software, gestione dei flussi informativi) e dei Data Scientist (specialisti verticali nella progettazione di modelli statistici e matematici per l'interpretazione e lo sfruttamento dei dati). In Italia soprattutto sulla seconda community di professionisti siamo leggermente in ritardo rispetto ai paesi di stampo anglosassone, ma è evidente che nel futuro prossimo soprattutto per i giovani queste specializzazioni rappresenteranno nuove concrete opportunità.

TECNOLOGIE BIG DATA E CONTROLLO ANTIRICICLAGGIO

Gli Obiettivi

- 1 Collezionare i numerosi dati dei movimenti dei clienti titolari di Carta di Credito
- 2 Costruire KPI e dashboard per monitorare le attività dei clienti in ottica di Compliance con la normativa d.lgs. 231/2007.



Approccio Seguito

- 1 Costruzione di un'Architettura Big Data basata su un cluster Hadoop
- 2 Caricamento dei dati relativi a transazioni, Clienti e Carte
- 3 Messa a disposizione degli utenti di uno strumento di Data Discovery, tramite il quale definire dinamicamente le soglie e il periodo temporale oggetto dell'analisi.



I Benefici

- 1 Miglioramento delle performance, la navigazione è caratterizzata da bassissimi tempi di latenza.
- 2 La potenza di calcolo e l'utilizzo di algoritmi di compressione offre la possibilità di mantenere online i dati storici anziché cancellarli o metterli in un archivio off-line
- 3 Riduzione dei costi rispetto all'architettura esistente:
 - "Commodity Hardware"
 - Nessun costo di licenza Software, solo costi di sottoscrizione



Abbonamento n. 1074 valido dal 20/03/2013 al 19/03/2014 - Licenza esclusiva a MIMESI SRL

ESPERIENZE - REPLY

Casi d'uso nel Finance: Enterprise Risk Management

L'ambito del Risk Management è da sempre caratterizzato dalla forte presenza di competenze di tipo Analytics. Le soluzioni di tipo Big Data consentono, ad esempio, di gestire grossi volumi di dati sia qualitativi che quantitativi necessari per la valutazione del rischio di credito, accorciando i tempi di analisi alle velocità oggi richieste dal processo di valutazione del merito di credito.

Un secondo ambito di applicazione è sicuramente il calcolo del rischio di controparte, in virtù della potenza computazionale che si renderà necessaria all'aumentare del numero e della complessità dei modelli interni e delle esigenze di calcolo di scenario e stress test: a titolo di esempio si pensi che un processo di revaluation coinvolgente 1.000 Controparti, 40 trade medie per ciascuna, una time grid di 50 passi e l'adozione di 1.000 scenari comporta l'esecuzione di 2 Miliardi di operazione di re-pricing.

Fraud Detection

Le tecnologie Big Data vedono il tema della Fraud Detection come ambito naturale di applicazione, a causa della natura dei dati e dei requisiti posti in termini di tempi di elaborazione. Infatti si tratta di processa-

re in tempo reale dati caratterizzati da:

- Un alto volume (ad es. le serie storiche delle transazioni);
- Un'alta varietà.

Le soluzioni di nuova generazione si concentrano sull'identificare in tempo reale le transazioni sospette tramite algoritmi di pattern recognition in grado di lavorare su più regole e dimensioni di scoring e di fornire uno score cumulativo, da confrontarsi con le soglie definite, come ad esempio:

- Il conto in questione è Delinquent o charge-off;
- Il saldo è vicino o oltre il limite di fido;
- Uno o più pagamenti sono stati rifiutati;
- Il Cliente non può essere localizzato;
- Le condizioni sopra elencate esistono in più di un conto presso diversi Istituti.

Customer Enrichment e marketing e cross-selling

Un'altra area primaria di applicazione è quella della raccolta di informazioni destrutturate, quali ad esempio quelle reperibili dai Social Media, al fine di:

- Migliorare il processo di Customer Profiling con informazioni quali il Social Value del cliente, per arricchire i modelli tradizionali basati su dati demografici, geografici e comportamentali;

- Per avere un quadro maggiormente significativo della Customer Experience, tramite tecniche quali ad esempio la Sentiment Analysis.

Social Risk Monitoring & Mitigation

Collegato ai temi di Social Customer Caring, i Big Data sono la tecnologia eletta per il presidio dei rischi derivanti dall'uso dei Social Network, e in generale la protezione del brand collegata alla crescente esposizione sulla rete (web, account Facebook, social dei dipendenti/clienti, etc.). Tecniche di ascolto delle community, early watching di possibili attacchi informatici, monitoring del customer sentiment sono tutte tipologie di prevenzione e mitigazione dei possibili rischi di immagine introdotti, sempre più frequentemente, dalla esposizione delle organizzazioni sul web.

Guido Guerrieri
Executive Partner
Reply