

# CREARE VALORE CON I BIG DATA

Gli strumenti, i processi,  
le applicazioni pratiche

scritto da  
Leonardo Camiciotti  
Christian Racca

Con il supporto del  
Consorzio TOP-IX

Illustrazioni di  
Laura Pippinato





# Creare valore con i Big Data

*Gli strumenti, i processi,  
le applicazioni pratiche*

*scritto da*

*Leonardo Camiciotti*

*Christian Racca*

*Con il supporto del*

*Consorzio TOP-IX*

*Illustrazioni di*

*Laura Pippinato*

EDIZIONI  
LSWR



Creare valore con i Big Data | Gli strumenti, i processi, le applicazioni pratiche

**Autori:** Leonardo Camiciotti, Christian Racca

**Con il supporto del:** Consorzio TOP-IX

**Illustrazioni di:** Laura Pippinato

**Collana:**

MDB - Modelli di Business

**Editor in Chief:** Marco Aleotti

**Progetto grafico:** Roberta Venturieri

**Immagine di copertina:** © Rawpixel Ltd| Thinkstock

**ISBN:** 978-88-6895-188-7

© 2015 Edizioni Lswr\* – Tutti i diritti riservati

*I diritti di traduzione, di memorizzazione elettronica, di riproduzione e adattamento totale o parziale con qualsiasi mezzo (compresi i microfilm e le copie fotostatiche), sono riservati per tutti i Paesi. Le fotocopie per uso personale del lettore possono essere effettuate nei limiti del 15% di ciascun volume dietro pagamento alla SIAE del compenso previsto dall'art. 68, commi 4 e 5, della legge 22 aprile 1941 n. 633.*

*Le fotocopie effettuate per finalità di carattere professionale, economico o commerciale o comunque per uso diverso da quello personale possono essere effettuate a seguito di specifica autorizzazione rilasciata da CLEARedi, Centro Licenze e Autorizzazioni per le Riproduzioni Editoriali, Corso di Porta Romana 108, 20122 Milano, e-mail autorizzazioni@clearedi.org e sito web www.clearedi.org.*

*La presente pubblicazione contiene le opinioni dell'autore e ha lo scopo di fornire informazioni precise e accurate. L'elaborazione dei testi, anche se curata con scrupolosa attenzione, non può comportare specifiche responsabilità in capo all'autore e/o all'editore per eventuali errori o inesattezze.*

*L'Editore ha compiuto ogni sforzo per ottenere e citare le fonti esatte delle illustrazioni. Qualora in qualche caso non fosse riuscito a reperire gli aventi diritto è a disposizione per rimediare a eventuali involontarie omissioni o errori nei riferimenti citati.*

*Tutti i marchi registrati citati appartengono ai legittimi proprietari.*

EDIZIONI  
**LSWR**

Via G. Spadolini, 7  
20141 Milano (MI)  
Tel. 02 881841  
www.edizionilswr.it

Printed in Italy

Finito di stampare nel mese di novembre 2015 presso "Press Grafica" s.r.l., Gravellona Toce (VB)

(\*) Edizioni Lswr è un marchio di La Tribuna Srl. La Tribuna Srl fa parte di LSWR GROUP.

# Sommario

|   |    |  |     |
|---|----|--|-----|
| Sommario  | 5  | <b>CAPITOLO 5:</b> Data Science tra presente e passato | 76  |
| Un grande tuffo nei dati                                      | 8  | <i>Definizioni a confronto</i>                         | 76  |
| <i>A chi si rivolge questo libro?</i>                         | 13 | <i>Dalla statistica alla Data Science</i>              | 78  |
| <i>Booksite</i>   | 13 | <b>CAPITOLO 6:</b> Le competenze del Data Scientist    | 86  |
| <b>Parte 1 - Dati dalla complessità</b>                       |    | <i>Development</i>                                     | 86  |
| <b>CAPITOLO 1:</b> L'abbraccio della complessità              | 16 | <i>Data Science</i>                                    | 100 |
| <b>CAPITOLO 2:</b> Il mondo nei dati                          | 32 | <i>Visualization</i>                                   | 109 |
| <i>Big Data = buzzword</i>                                    | 32 | <i>Big Data "framework and platforms"</i>              | 121 |
| <i>Dati, dati e ancora dati</i>                               | 39 | <i>Aspetti legali dei Big Data</i>                     | 131 |
| <i>Una definizione per i Big Data</i>                         | 47 | <b>CAPITOLO 7:</b> Data Scientist tra mito e realtà    | 138 |
| <i>Dati - modelli - decisioni - impatto</i>                   | 53 | <b>Parte 3 - Il Data Ring</b>                          |     |
| <i>The Dark Side of Big Data</i>                              | 58 | <b>CAPITOLO 8:</b> Metodo                              | 144 |
| <b>CAPITOLO 3:</b> Il Data Scientist come ponte tra due mondi | 64 | <i>Goal</i>  | 148 |
| <b>Parte 2 - Il Data Scientist</b>                            |    | <i>Strumenti</i>                                       | 150 |
| <b>CAPITOLO 4:</b> Dai dati alla Data Science                 | 70 | <i>Competenze</i>                                      | 150 |
| <i>Dati, informazioni, conoscenza</i>                         | 73 | <i>Processo</i>  | 151 |
|   |    | <i>Valorizzazione</i>                                  | 151 |
|   |    | <b>CAPITOLO 9:</b> Obiettivi                           | 152 |

|                                    |     |  |     |
|------------------------------------|-----|--|-----|
| <b>CAPITOLO 10:</b> Strumenti      | 160 | <i>Secondo pivot: l'incontro con Cerved</i>    | 218 |
| <i>Dati</i>                        | 161 | <i>Lessons learnt e considerazioni</i>         | 220 |
| <i>Infrastruttura</i>              | 174 |  |     |
| <b>CAPITOLO 11:</b> Competenze     | 178 |  |     |
| <b>CAPITOLO 12:</b> Processo       | 186 |  |     |
| <b>CAPITOLO 13:</b> Valorizzazione | 194 | <b>CAPITOLO 16:</b> Il caso di aizoOn          | 222 |
| <i>Interpretazione</i>             | 195 | <i>Contesto</i>                                | 224 |
| <i>Valutazione d'impatto</i>       | 196 | <i>Goal</i>                                    | 225 |
| <i>Data input</i>                  | 198 | <i>Strumenti</i>                               | 226 |
| <i>Struttura</i>                   | 200 | <i>Competenze</i>                              | 228 |
|                                    |     | <i>Processo</i>                                | 230 |
|                                    |     | <i>Valorizzazione</i>                          | 232 |
|                                    |     | <b>CAPITOLO 17:</b> Il caso di Iconsulting     | 234 |
|                                    |     | <i>Il contesto e il problema</i>               | 235 |
|                                    |     | <i>Implementazione</i>                         | 238 |
|                                    |     | <i>Output</i>                                  | 241 |
|                                    |     | <i>Lessons learnt</i>                          | 242 |
|                                    |     | <b>CONCLUSIONI:</b> C'era una volta un dato... | 246 |
|                                    |     | <i>Complessità e dati</i>                      | 246 |
|                                    |     | <i>Data Scientist</i>                          | 247 |
|                                    |     | <i>Data Ring</i>                               | 248 |
|                                    |     | <i>"Use Cases"</i>                             | 248 |
|                                    |     | <i>Ringraziamenti</i>                          | 249 |
|                                    |     | <b>BIBLIOGRAFIA</b>                            | 251 |
|                                    |     | <i>Sitografia</i>                              | 252 |

## Parte 4 - La realtà dei dati

|  |     |
|--|-----|
| <b>CAPITOLO 14:</b> Dalla teoria ai casi pratici         | 206 |
| <i>Come usare il Data Ring?</i>                          | 206 |
| <i>Validare il Data Ring</i>                             | 210 |
| <b>CAPITOLO 15:</b> Il caso di SpazioDati                | 212 |
| <i>Il problema e la vision</i>                           | 213 |
| <i>Dalla value proposition al primo prodotto</i>         | 214 |
| <i>Primo pivot: oltre ai dati... anche uno strumento</i> | 215 |



## Introduzione

# Un grande tuffo nei dati

Nel marzo 2014, quando l'editore ci contattò per proporci di scrivere un libro sul tema Big Data, si era conclusa da pochi mesi la terza edizione del corso Big Dive e avevamo quindi avviato i lavori di progettazione per il quarto episodio.

Big Dive è un programma di formazione immersivo di cinque settimane finalizzato al training di nuove figure professionali capaci di gestire ed estrarre valore da grandi volumi di dati. Il concept iniziale prese vita a seguito di un viaggio di lavoro negli USA (Silicon Valley) nel febbraio-marzo del 2012, nel corso del quale toccammo con mano l'incredibile richiesta di sviluppatori software evidenziata dal mercato sia start-up sia corporate con particolare riferimento ad alcune tecnologie emergenti. Tornati in Italia, constatammo che il problema era in realtà diffuso e generalizzato: l'incedere dei trend tecnologici e l'offerta formativa tradizionale (intesa prevalentemente

in senso accademico) erano disallineati, evidenziando un gap di domanda-offerta forte e quindi un'opportunità. Quest'ultima affermazione non è da leggersi in tono necessariamente polemico nei confronti delle università: seppur con alcuni limiti oggettivi dovuti alla rigidità di rinnovamento dei corsi, il ruolo di fornire le basi logiche e teoriche era ed è tuttora svolto in modo soddisfacente. D'altra parte, il settore dell'ICT è stato caratterizzato negli ultimi anni da una ramificazione di contenuti e una specializzazione tali da essere praticamente inconciliabile con qualsiasi percorso formativo pensato su larga scala. Risultano invece ormai indispensabili momenti di approfondimento intensivo, focalizzati sulle soluzioni (hardware, software, architetturali ecc.) in voga (in costante rinnovamento) e spiccatamente orientati a un approccio fattuale e concreto: teoria ma soprattutto tanta pratica sul campo e confronto con problemi reali.



Il professionista del futuro, a nostro avviso, dovrà quindi comporre il suo background di competenze sempre più come mosaico “dinamico” di discipline e progettualità, piuttosto che confidare troppo su percorsi lineari di apprendimento.

L’idea di attivare un percorso di training passò rapidamente dallo stadio di “chiacchiere da viaggio di rientro” a obiettivo per l’annualità, da sviluppare internamente all’azienda (Consorzio TOP-IX). D’altra parte un corso di formazione specifico richiedeva un ambito di studio preciso onde evitare di inciampare nello stesso problema evidenziato come fattore scatenante. Sgonfiatosi (o forse sarebbe meglio dire consolidatosi come prassi) l’hype del Cloud Computing, il trend che, a detta di tutti, avrebbe segnato gli anni successivi dell’*Information and Communication Technology* vedeva al centro l’unità atomica dei sistemi informativi: il dato.

In effetti la cosa non avrebbe dovuto stupire nessuno: la globalizzazione, lo sviluppo delle tecnologie “machine-to-machine” e il moltiplicarsi delle interazioni sociali virtuali stavano delineando, in maniera sempre più chiara, i tratti di quella che oggi è spesso definita come “l’era della complessità”. In tale contesto il volume di dati prodotti dai social network, dalle reti di sensori intelligenti e dai log dei sistemi informativi aziendali, a cui si aggiunge l’output del processo costante di digitalizzazione della conoscenza, si configura come asset imprescindibile per tutti i settori di attività (pubblica amministrazione, industria, start-up innovative, NGO ecc.).

Sui dati, TOP-IX stesso aveva tra l’altro già maturato solide basi nel suo ruolo di innovatore per le politiche pubbliche. Il percorso che aveva portato la Regione Piemonte a essere pioniera sul fronte del rilascio dei cosiddetti Open Data ci aveva visti infatti tra i protagonisti in prima linea.

Definito il perimetro, il passo successivo fu costituito dall'individuare, nel network aziendale, i partner con cui avviare il numero zero. Per la componente scientifica la scelta non poteva che ricadere sulla Fondazione ISI, istituto di ricerca di altissimo profilo, leader di settore nello studio dei sistemi complessi. Le competenze tecnico-pratiche necessitavano, come detto, di una forte impronta innovativa, così da intercettare al meglio gli ultimi trend del settore, e si decise pertanto di puntare su una giovane azienda torinese, Axant, sufficientemente geek per interpretare le esigenze manifestate. Ma i dati, nella loro fredda dimensione numerica e digitale, richiedono anche una forte componente visuale che consenta di interpretarli al meglio: gli amici dello studio di interaction design Todo furono, così, arruolati nel team di progetto.

La squadra era al completo e nell'autunno 2012 il format "Big Dive: Hacking Development Visualization and Data Science" prese il via con quel pizzico di ingenuità e ambizione che contraddistingue le nuove sfide (tra cui anche la scrittura di questo libro). In qualche modo il risultato andò oltre le aspettative e ben 20 partecipanti di 8 diverse nazionalità costituirono la prima classe di (Big) "Divers". Per il team organizzatore il fatto di avere studenti provenienti da Germania, Irlanda, Svizzera, Polonia, Iran, USA e Brasile, oltre ovviamente a un nutrito

manipolo di italiani, fu il risultato più stimolante e una chiara testimonianza del fatto che il trend dei Big Data era prossimo a esplodere.

A valle della quarta edizione del corso, con circa 70 Divers formati, l'hype mediatico attorno al tema si è in qualche modo mitigato e la buzzword Big Data entra ora necessariamente nella sua fase di maturità e concretezza, volta a quantificarne le ricadute economiche ma anche sociali. La "ricetta" per l'estrazione del valore a partire dai dati grezzi si è negli anni consolidata e vede tra i suoi ingredienti un importante mix di componenti: **contesto, relazioni, multisettorialità e interdisciplinarietà, consapevolezza della complessità e visione a lungo termine**. Tali elementi si inseriscono in un framework che, a partire dai dati, consente lo sviluppo di modelli e attraverso questi abilita la creazione di policy e strumenti di decisione basati sull'osservazione effettiva della realtà ("**data-driven decision making**"). L'attenzione, in qualche modo, si è spostata così dalla disponibilità dei dati e dagli strumenti tecnologici alle **competenze professionali** necessarie, all'**approccio** ("data-driven") e alle dinamiche che preludono alla **generazione di impatto**.

Lo spirito narrativo che abbiamo cercato di seguire in questo lavoro di scrittura prova a trasferire "su carta" esattamente il concetto esposto nell'ultimo paragrafo, così

come tutta l'esperienza maturata grazie a Big Dive, nonché in alcuni progetti di tipo consulenziale svolti in sede aziendale. Sfogliando le prossime pagine, non troverete pertanto una celebrazione dei Big Data come rivoluzione dogmatica (molto, infatti, si è scritto su tale fronte), bensì una sorta di guida pratica finalizzata a illustrare i punti chiave su cui impostare la propria strategia "data-driven", minimizzando così le probabilità di incorrere negli errori più comuni legati alla scienza dei dati sia in un contesto pubblico sia in uno privato. A tal fine - e solo grazie ai vostri feedback scopriremo se saremo stati in grado di raggiungere l'obiettivo - abbiamo cercato di fare largo uso di spunti concreti e casi pratici, così come di organizzare il materiale secondo una prospettiva operativa e progettuale. Più precisamente, il libro è organizzato in quattro sezioni principali, di seguito descritte.

**La parte 1 - Dati dalla complessità** definisce il contesto e l'ambito generale d'azione: non potevamo in tal senso esimerci dall'introdurre e illustrare il concetto di Big Data, anche se noi stessi, *in primis*, non siamo particolarmente affezionati a tale termine. Si tratta, infatti, di un neologismo troppo spesso usato e abusato impropriamente, che ha poca rilevanza oggettiva e subisce invece molta influenza lato marketing. D'altro canto è la cornice che fa da sfondo al fenomeno che stiamo descrivendo, ed effettivamente la

grande disponibilità di dati, così come la presenza di alcuni fattori abilitanti, ha pesantemente modificato l'approccio risolutivo a determinati problemi di natura tecnica, piuttosto che business. Parleremo dunque di Big Data, ma anche del suo alter-ego (perdonateci la semplificazione in questa prima introduzione) scientifico, ossia la scienza della complessità. Illustreremo il rapporto che li lega e daremo evidenza di come l'uno abbia condizionato il successo dell'altro, portando in qualche modo a un avvicinamento tra il mondo della ricerca scientifica e la dimensione più operativa-produttiva, troppo spesso separati da un muro fatto di preconcetti, assunzioni e dinamiche di risultato differenti (fatturato, pubblicazioni ecc.).

**La parte 2 - Il Data Scientist** sposta pesantemente il baricentro narrativo sulla questione delle competenze necessarie all'analisi ma soprattutto alla valorizzazione dei (Big) Data. Il cosiddetto Data Scientist in questo senso è diventato rapidamente una tra le figure professionali più ambite, ricercate e di conseguenza remunerate degli ultimi anni, proprio in funzione della sua capacità di saper analizzare i problemi in una dimensione olistica e complessa e della sua sensibilità nel saper estrarre conoscenza dai dati accumulati.

Anche attorno ai Data Scientist vi è molta confusione informativa: la sensibilità, di cui sopra, nei confronti del

dato è necessariamente un mix di nozioni tecnico-pratiche che richiedono una forte base di esperienza sul campo, in una disciplina comunque ancora agli albori, e che difficilmente possono essere interamente accentrate in un singolo individuo. La dimensione del team di lavoro è quindi fondamentale in tal senso, a patto però che ogni elemento del gruppo abbia una chiara e completa visione della catena del valore. Il ruolo auspicato per il Data Scientist assume quasi la rilevanza di mediatore e collante culturale in attesa che l'approccio definito come "data-driven" estenda la sua sfera di influenza dall'ambito tecnico fino a quello umanistico, così come dal lato produttivo/operativo alle posizioni manageriali/decisionali.

**La parte 3 - Il Data Ring** intende invece raccogliere tutta l'esperienza degli autori in uno strumento pratico che possa essere utilizzato prima, dopo e durante la realizzazione di un progetto basato sui dati. Ritornano ancora una volta le competenze, ma entrano necessariamente in gioco gli strumenti, il metodo e il contesto in cui il progetto viene definito, misurato e valorizzato: il tutto a partire dalla definizione degli obiettivi e in un'ottica di validazione iterativa delle ipotesi (da qui *ring* inteso come anello).

Ispirati da altri celebri lavori in letteratura, anche noi abbiamo quindi provato a formalizzare una sorta di "data-

canvas" che sia di supporto al manager e al decisore, e più in generale a chi approccia un progetto "data-driven". Come spesso accade in questi casi, l'uso effettivo decreterà la bontà o meno dello strumento, concepito, in ogni caso, per essere leggero (e quindi non di ostacolo) e adattabile ai vari contesti e casi applicativi puntuali.

**La parte 4 - La realtà dei dati** assolve infine il ruolo di validazione concreta dei principi, delle teorie, degli strumenti e degli assunti delineati nel corso delle tre precedenti sezioni. La maturità dei Big Data è infatti dettata dall'impatto generato a livello imprenditoriale e/o sociale: casi d'uso concreti diventano quindi il modo migliore per dare senso e sostanza a un impianto teorico che rischierebbe altrimenti di risultare sterile e pretenzioso. A tal fine sono stati scelti tre casi rappresentativi del contesto italiano, per certi aspetti ancora acerbo (nel suo complesso) rispetto ad altre realtà fuori dai confini nazionali: il primo relativo a una start-up, il secondo al mondo corporate, il terzo nell'ambito della pubblica amministrazione. Un doppio filo sottile accomuna queste tre storie: in primo luogo, la consapevolezza del fatto che i Big Data sono un fenomeno attuale e non marginale; in seconda battuta, la necessità di valorizzare i dati partendo dall'esplorazione degli "atomi digitali" fino ad arrivare all'integrazione con le policy e i modelli di business.

## **A chi si rivolge questo libro?**

Si tratta di una domanda che ci ha posto l'editore in sede di brief e che inseriamo in queste pagine introduttive, provando a formulare una non facile risposta.

La trattazione nel suo complesso è di alto livello, introducendo molti concetti tecnico-specifici senza però avere la pretesa di illustrarli in dettaglio. Laddove possibile, anche compatibilmente con le nostre capacità, ci avvalremo di box di approfondimento; in tutti gli altri casi invitiamo il lettore a usare questo volume come una sorta di "indice" per poi approfondire su fonti più autorevoli e settoriali gli ambiti di interesse e rilevanza.

Riteniamo poi sia senz'altro un utile strumento per manager o più in generale per figure responsabili all'interno di aziende, pubblica amministrazione, soggetti no-profit, settore culturale, che si accingano (per volontà o per richiesta) a progettare la transizione a organizzazione/

realtà "data-driven", strutturando l'acquisizione di dati ex-novo o trasformando il patrimonio informativo esistente in un asset strutturale.

Più in generale, confidiamo che molti dei nostri lettori siano appassionati, come noi, dei progressi tecnologici e dell'innovazione aperta, che negli ultimi anni ha visto il tema dei dati quale propulsore quantico verso nuovi paradigmi di applicazione della scienza all'evoluzione della società moderna.

Indipendentemente dal profilo dei nostri lettori, confidiamo infine che la lettura di questo volume sia di ispirazione e invitiamo a contattarci personalmente al fine di fornirci feedback, critiche costruttive e spunti per migliorarci.

## **Booksite**

Sul sito [www.dataring.eu](http://www.dataring.eu) potete scaricare il Data Ring e altri materiali a supporto del libro.