

# BUSINESS INTELLIGENCE

Questo articolo è rilasciato sotto licenza "Creative Commons Attribuzione-StessaLicenza 2.5 Italy License" da **Simone Brunozi** ([www.ubuntu.it](http://www.ubuntu.it) | [www.simpler.it](http://www.simpler.it) | [simone.brunozzi@gmail.com](mailto:simone.brunozzi@gmail.com))

12 dicembre 2005

## Introduzione

Mi è stato chiesto di informarmi su possibili software GPL / Open Source per gestire la Business Intelligence di una azienda.

Avendo solo sentito parlare di Business Intelligence (BI), e non avendo alcuna reale competenza in merito, mi sono sentito in dovere di documentarmi prima di esaminare qualsiasi soluzione.

Questa prima fase di raccolta informazioni, pur rischiando di fornire dati già conosciuti, è stata qui riportata per completezza.

Ho raccolto informazioni unicamente attraverso internet, partendo dalla classica ricerca su <http://www.wikipedia.org> in lingua inglese, per poi seguire i link e i documenti più interessanti.

Molte fonti saranno omesse per pura brevità.

## Business Intelligence: raccolta informazioni

[[http://en.wikipedia.org/wiki/Business\\_intelligence](http://en.wikipedia.org/wiki/Business_intelligence)]

La Business Intelligence (termine coniato da Howard Dresner, analista del Gartner, nei primi anni '90) è un insieme di modelli, metodi e strumenti rivolti:

- A) alla raccolta sistematica del patrimonio di informazioni generate da un'azienda;
- B) alla loro aggregazione e analisi;
- C) alla loro presentazione in forma semplice.

Questa conoscenza è utilizzabile in processi decisionali e di analisi da parte dei cosiddetti knowledge workers (termine coniato da Peter Drucker nel 1959: si dice di una persona che lavora principalmente con l'informazione, o che sviluppa e usa conoscenza nel suo spazio lavorativo).

I dati generati dai vari sistemi (contabilità, produzione, R&D (Research & Development), CRM (Customer Relationship Management)) vengono archiviati in sistemi di basi di dati.

Il termine Business Intelligence (a volte abbreviato con BI) può riferirsi a:

- A) Un insieme di processi di business;
- B) la tecnologia utilizzata in tali processi;
- C) le informazioni ottenute da tali processi.

Un processo di business è una ricetta per ottenere un risultato commerciale. Ogni processo ha degli input, dei metodi, e degli output. Gli input sono prerequisiti che devono sussistere per l'applicazione del metodo, a cui poi seguono come risultati gli output.

Vengono spesso utilizzati i Key Performance Indicators (**KPI**), conosciuti anche come Key Success Indicators (**KSI**).

Si tratta di metriche, finanziarie e non, usate per riflettere i fattori critici di successo di una organizzazione. Aiutano a misurare il progresso verso gli obiettivi organizzativi.

[[http://en.wikipedia.org/wiki/Key\\_performance\\_indicators](http://en.wikipedia.org/wiki/Key_performance_indicators)]

Lo scopo di un software che supporta la Business Intelligence è di aiutare le persone a prendere decisioni migliori, rendendo disponibili, in maniera accurata e rilevante, le informazioni giuste quando necessario.

## Business Intelligence: strumenti utilizzabili

Sono stati sviluppati strumenti che facilitano il lavoro della BI, specialmente nei confronti di grandi quantità di dati da raccogliere, normalizzare e analizzare.

I principali strumenti comprendono:

1) **OLAP** (On Line Analytical Processing) basati sulla analisi dimensionale e sul concetto di ipercubo (OLAP è stato coniato da E.F. Codd & Associates in un paper del 1994);

Dal paper del 1994 di Codd è emersa la necessità di organizzare i dati in array, per permettere una analisi più veloce. Questi array sono chiamati "cubi".

Tale organizzazione permette di evitare le limitazioni di un normale database relazionale, che di solito non è indicato per l'analisi immediata di grandi quantità di dati, mentre è più adatto a creare record da una serie di transazioni (OLTP: On Line Transaction Processing).

Nonostante esistano molti software di report per database relazionali, la loro lentezza limita la dimensione massima dei dati analizzati.

I cubi OLAP possono essere immaginati come delle estensioni degli array bidimensionali di un foglio di calcolo.

Per esempio, una compagnia può voler analizzare i dati finanziari per prodotto, per periodo di tempo, per città, per tipo di guadagno e costo, e comparando i dati con il budget stabilito.

Questi metodi addizionali di analisi dei dati sono conosciuti come "dimensioni".

Dato che possono esserci più di tre dimensioni in un sistema OLAP, il termine "hypercube" (ipercubo) viene talvolta usato.

2) **Scorecarding** (letterale: segnapunti)

[[http://en.wikipedia.org/wiki/Balanced\\_scorecard](http://en.wikipedia.org/wiki/Balanced_scorecard)]

Nel 1992 Robert S. Kaplan e David Norton hanno introdotto la "balanced scorecard" (BSC), un metodo per misurare le attività di una azienda in termini di visione e strategie. Fornisce ai manager una visuale completa delle prestazioni di un business.

Lo "scorecarding" è un sistema di gestione strategica che forza i manager a focalizzare la loro attenzione sulle metriche di prestazione che portano al successo. Equilibra la prospettiva finanziaria col cliente, il processo interno, e le prospettive di apprendimento e di crescita.

Il sistema consiste in quattro processi:

1. Tradurre la visione in obiettivi operativi;
2. Comunicare la visione e collegarla a prestazioni individuali;
3. Pianificare del business;
4. Ottenere feedback, e conseguente apprendimento e aggiustamento della strategia.
- 5.

3) **Dashboarding** (letterale: cruscotto)

Presentazione dei dati attraverso interfacce grafiche modellate ad hoc. Ecco alcuni esempi:

[[http://www.oracle.com/technology/pub/images/rittman\\_dash\\_f15.gif](http://www.oracle.com/technology/pub/images/rittman_dash_f15.gif)]

[<http://www.informationbuilders.com/products/webfocus/images/business-intelligence.jpg>]

[[http://www.microsoft.com/presspass/images/press/2005/10-23dashboard-scorecard\\_lg.jpg](http://www.microsoft.com/presspass/images/press/2005/10-23dashboard-scorecard_lg.jpg)]

4) **Information Visualization** (visualizzazione dati)

[[http://en.wikipedia.org/wiki/Information\\_visualization](http://en.wikipedia.org/wiki/Information_visualization)]

La visualizzazione dati è un'area di ricerca complessa. Si basa su teorie di information design, computer graphic, interazione uomo-macchina, scienze cognitive.

L'applicazione pratica della visualizzazione dati coinvolge la selezione, la trasformazione e la rappresentazione di dati astratti in una forma che facilita l'interazione con l'uomo per la sua esplorazione e comprensione. Forme dinamiche di visualizzazione, come le animazioni didattiche, hanno il potenziale di migliorare la comprensione di sistemi che cambiano col tempo.

Tecniche particolari permettono all'utente di modificare la visualizzazione in tempo reale, permettendo una percezione migliore dei campioni e delle relazioni strutturali tra i dati in oggetto. Infine, non ci si limita solamente ad aspetti visivi, ma anche uditivi.

## 5) **Data Modeling** (modellazione dati)

[[http://en.wikipedia.org/wiki/Data\\_modeling](http://en.wikipedia.org/wiki/Data_modeling)]

Nella creazione di sistemi informativi, il Data Modeling è l'analisi e il design della informazione all'interno del sistema, incentrata sulle entità logiche e le dipendenze logiche tra di esse.

È una attività astratta in cui i dettagli dei valori delle singole osservazioni di dati sono ignorati in favore delle strutture, relazioni, nomi e formati dei dati di interesse, nonostante una lista di valori validi è solitamente inclusa. Il modello dei dati non deve solo definire la struttura dei dati, ma anche la semantica (il significato) degli stessi.

Nonostante un termine comune per questa attività sia "analisi dei dati", essa è più collegata alla sintesi (mettere insieme) dei dati stessi, dato che l'attività si sforza di unire le strutture dati di interesse in un "tutto" coesivo, inseparabile, eliminando tutte le ridondanze non necessarie e collegando le strutture dati con delle relazioni.

## 6) **Strumenti per definire la Logica del Business**

[[http://en.wikipedia.org/wiki/Business\\_logic](http://en.wikipedia.org/wiki/Business_logic)]

La logica del business si riferisce alla logica che incarna le business rules (regole del business) anziché la visualizzazione o l'immagazzinamento dei dati. È un termine usato nella programmazione software per intendere uno strumento che effettua certi tipi di operazioni su dati passati attraverso di esso.

La business logic è solitamente associata con una architettura software three-tier (a tre strati).

## 7) **Strumenti per definire le Regole del Business**

[[http://en.wikipedia.org/wiki/Business\\_rules](http://en.wikipedia.org/wiki/Business_rules)]

Le regole del business, molto semplicemente, descrivono le operazioni, definizioni e costrizioni che si applicano in una organizzazione nella realizzazione dei suoi obiettivi.

## 8) **Data Warehouse** (magazzino dati)

[[http://en.wikipedia.org/wiki/Data\\_warehouse](http://en.wikipedia.org/wiki/Data_warehouse)]

Un magazzino dati (data warehouse) contiene tutte le informazioni passate di una azienda, immagazzinate in un Database disegnato per favorire efficienti analisi dei dati e reporting dei dati. Non è inteso per gestire dati correnti.

I magazzini dati solitamente conservano grandi quantità di informazioni che vengono talvolta suddivise in unità logiche più piccole, chiamate "Dependent Data Mart".

Solitamente, due idee di base guidano la creazione di un magazzino dati:

Integrazione dei dati da database distribuiti e strutturati diversamente, il che facilita una visione globale e una analisi complessiva.

Separazione dei dati utilizzati nelle operazioni giornaliere dai dati (questi ultimi presenti nel magazzino dati) usati per reporting, supporto decisionale, analisi e controllo.

Periodicamente si importano dati da sistemi **ERP** (Enterprise Resource Planning) e altri sistemi software collegati al business dell'azienda.

## 9) **Document Warehouse** (magazzino documenti; la traduzione letterale è un po' infelice)

Un Document Warehouse contiene informazioni sui motivi ("why things have happened") per cui certe cose sono successe, invece di avere informazioni su cosa sia successo ("what happened") come nel caso del data warehouse.

Tali informazioni possono essere combinate insieme per avere uno strumento davvero utile. Spesso sono accompagnati da vari software adatti ad analizzare ed utilizzare tali dati.

## 10) **Data mining** (estrazione di dati)

[[http://en.wikipedia.org/wiki/Data\\_mining](http://en.wikipedia.org/wiki/Data_mining)]

Il Data Mining (estrazione dei dati), conosciuto anche come KDD (Knowledge Discovery in Databases) consiste nelle tecniche per ricercare automaticamente dei campioni in grandi archivi di dati, usando tecniche computazionali derivate dalla statistica e dal pattern recognition (riconoscimento di modelli).

11) **Text mining** (estrazione di testo)

[[http://en.wikipedia.org/wiki/Text\\_mining](http://en.wikipedia.org/wiki/Text_mining)]

La estrazione di testo è conosciuta anche come "intelligent text analysis" (analisi intelligente del testo), o come "Knowledge Discovery in Text" (KDT) (Scoprire conoscenza nel testo).

Si riferisce al processo di estrarre informazioni e conoscenze interessanti e non banali a partire da testo non strutturato. ECHELON è il più famoso esempio di applicazione di Text Mining.

12) **EIS - Executive Information Systems** (sistemi informativi di esecuzione)

[[http://en.wikipedia.org/wiki/Executive\\_Information\\_Systems](http://en.wikipedia.org/wiki/Executive_Information_Systems)]

Sono sistemi informatici per facilitare e supportare le esigenze decisionali di manager di alto livello, fornendo facile accesso ad informazioni interne ed esterne importanti per raggiungere gli obiettivi strategici della azienda. Sono spesso considerati una forma specializzata dei DSS.

L'enfasi degli EIS è nelle visualizzazioni grafiche e nelle facili interfacce utente. Offrono resoconti dettagliati e capacità di approfondimento.

In generale, i sistemi EIS sono dei DSS che abbracciano l'intera azienda, aiutando i dirigenti ad analizzare, confrontare e sottolineare le tendenze di variabili importanti, permettendo loro di controllare le prestazioni e identificare opportunità e problemi.

I sistemi EIS e i sistemi di Data Warehouse stanno attualmente convergendo nel mercato.

13) **DSS - Decision Support Systems** (sistemi di supporto alle decisioni)

[[http://en.wikipedia.org/wiki/Decision\\_Support\\_Systems](http://en.wikipedia.org/wiki/Decision_Support_Systems)]

I sistemi DSS sono una classe di sistemi informativi computerizzati che supportano attività decisionali. Esistono vari tipi di DSS:

Model-driven (basato sul modello): enfatizza accesso e manipolazione di modelli statistici, finanziari, di ottimizzazione o di simulazione. Usano dati e parametri forniti dagli utenti DSS per aiutare le persone incaricate di prendere decisioni ad analizzare una situazione.

Un DSS model-driven in ambito Open Source è Dicodess [<http://dicodess.sourceforge.net>].

Communication-driven (basato sulla comunicazione): supporta più di una persona che lavora ad un compito condiviso.

Esempi di tali software sono Microsoft NetMeeting, o Groove [<http://www.groove.net>].

Data-driven (basato sui dati): enfatizza l'accesso e la manipolazione di serie temporali di dati interni alla azienda e, a volte, dati esterni.

Document-driven (basato sui documenti): gestisce, recupera e manipola informazioni non strutturate in una varietà di formati elettronici.

Knowledge-driven (basato sulla conoscenza): fornisce perizie specializzate nella risoluzione dei problemi; tali perizie sono immagazzinate come fatti, regole, procedure, o strutture simili.

14) **MIS - Management Information Systems** (sistemi di gestione dell'informazione)

[[http://en.wikipedia.org/wiki/Management\\_Information\\_Systems](http://en.wikipedia.org/wiki/Management_Information_Systems)]

I MIS sono sistemi informativi utilizzati all'interno di una organizzazione, definiti come "sistema costituito dalla rete di tutti i canali di comunicazione usati all'interno di una organizzazione".

Possono essere anche definiti come "sistema che raccoglie e processa informazioni e le fornisce ai manager a tutti i livelli, i quali le usano per prendere decisioni, pianificare, implementare programmi, e controlli".

Purtroppo sembra che in molti casi la definizione di MIS si allarghi a dismisura.

#### 15) **GIS - Geographic Information Systems** (sistemi di informazioni geografiche)

[[http://en.wikipedia.org/wiki/Geographic\\_information\\_system](http://en.wikipedia.org/wiki/Geographic_information_system)]

Un GIS è un sistema per creare e gestire dati geografici e attributi associati.

#### 16) **Analisi dei dati**

Questa sezione è da iniziare.

#### 17) **Analisi del flusso di Business**

Questa sezione è da iniziare.

#### 18) **Business Performance Management** (gestione delle performance del business)

[[http://en.wikipedia.org/wiki/Business\\_performance\\_management](http://en.wikipedia.org/wiki/Business_performance_management)]

Il Business Performance Management (BPM) è un insieme di processi che aiutano le organizzazioni ad ottimizzare le prestazioni del business. BPM è visto come una Business Intelligence di nuova generazione. Si focalizza nei processi di business come la pianificazione e la previsione. Aiuta a scoprire utilizzi efficienti delle unità di business, finanziarie, umane, e risorse materiali.

BPM coinvolge il consolidamento dei dati da varie sorgenti, la loro interrogazione, e l'analisi dei dati, mettendo in pratica i risultati.

BPM migliora i processi creando dei migliori cicli di feedback. Revisioni continue e in tempo reale aiutano ad identificare ed eliminare i problemi prima che crescano. Le abilità di previsione del BPM aiutano l'azienda ad intraprendere azioni correttive in tempo per rispettare le proiezioni di guadagni. La previsione è caratterizzata da un alto grado di predicibilità, la quale viene usata per rispondere a scenari what-if (cosa succede se...). BPM è utile nell'analisi del rischio e nel predire esiti di unioni e di acquisizioni, e fornisce piani per superare problemi potenziali.

BPM fornisce KPI (Key Performance Indicators) che aiutano le compagnie a misurare l'efficienza dei progetti e degli impiegati, confrontata con obiettivi operativi.

### **Software Open Source che supporta Business Intelligence**

Di seguito vengono indicati i nomi dei software, o i nomi delle aziende che li hanno creati e che li stanno supportando. A fianco del nome viene anche indicata la Google relevance, ovvero il numero di risultati che si ottiene cercando il software (o l'azienda) con Google.

Tale risultato fornisce una stima approssimativa della popolarità del termine ricercato.

- **Pentaho** [Google relevance: 224.000 risultati]

[<http://www.pentaho.org>]

[<http://en.wikipedia.org/wiki/Pentaho>]

Pentaho fornisce capacità di reporting, analisi, dashboard, data mining e workflow.

- **Eclipse software** [Google relevance: non applicabile]

[<http://www.eclipse.org>]

[[http://en.wikipedia.org/wiki/Eclipse\\_\(software\)](http://en.wikipedia.org/wiki/Eclipse_(software))]

[[http://en.wikipedia.org/wiki/Eclipse\\_Foundation](http://en.wikipedia.org/wiki/Eclipse_Foundation)]

Eclipse fornisce applicazioni "rich-client" (ovvero: pesantemente appoggiate sul lato client, un approccio opposto alle applicazioni "thin-client", solitamente web-based).

Finora Eclipse è stato utilizzato per sviluppare IDE (Integrated Development Environment) come il Java Development Toolkit (JDT), il client Azureus (per file torrent) e lo stesso compilatore che viene fornito come parte di Eclipse stesso.

Eclipse è stato originariamente sviluppato da Object Technology International (in seguito acquistata da IBM), ma è ora supportato dalla Eclipse Foundation, un consorzio non-profit di aziende software di livello internazionale.

Alcune aziende, come Borland e IBM Rational, stanno adottando Eclipse come futura piattaforma di sviluppo per i loro IDE.

- **OpenI** [Google relevance: 92.900]

[<http://openi.sourceforge.net>]

[<http://en.wikipedia.org/wiki/OpenI>]

OpenI è una applicazione web-based, basata su J2EE (Java Enterprise Edition), la cui installazione di base funziona sul Java server Tomcat.

OpenI fornisce soluzioni per costruire e pubblicare rapporti da risorse di dati OLAP come Microsoft Analysis Services o Mondrian, e provvede analisi consolidate per tutti i componenti chiave di una applicazione intelligente, includendo:

- OLAP fonti di dati OLAP;
- database relazionali;
- modelli di dati statistici;
- modelli di Data Mining.

OpenI™ è posseduto da Loyalty Matrix, Inc. [<http://www.loyaltymatrix.com>]

- **Bizgres** [Google relevance: 49.900]

[<http://www.bizgres.org>]

[<http://www.greenplum.com>]

Bizgres è un database server focalizzato nel supportare applicazioni di Business Intelligence. Offre una robusta e ricca piattaforma per supporto alle decisioni.

Bizgres offre numerose caratteristiche critiche per le applicazioni di Business Intelligence:

Facilità di installazione  
supporto multi-piattaforma  
caricatore di grandi quantità di dati (bulk data)  
connettività aperta agli strumenti di Business Intelligence

Bizgres si basa sull'ottimo database open source PostgreSQL [<http://www.postgresql.org>].

## Considerazioni personali

Eclipse sembra essere il software più popolare e più supportato.

Bizgres è uno strumento di supporto alla BI, e sembra impostato molto bene; la partecipazione alla comunità di sviluppo è tuttavia ancora scarsa.

A proposito di Greenplum e di Bizgres:

[[http://news.com.com/Firms+release+open-source+business-intelligence+tools/2100-7344\\_3-5818769.html](http://news.com.com/Firms+release+open-source+business-intelligence+tools/2100-7344_3-5818769.html)]

OpenI sembra un framework molto macchinoso.

Pentaho è più misterioso, e non è facile farsi una opinione al riguardo in poco tempo.

A proposito del recente finanziamento di Pentaho:

[<http://www.networkworld.com/news/2005/120905-open-source-bi.html>]

Ancora da esaminare: [<http://tinyp.com>] [<http://extendb.com>]

Questo articolo è rilasciato sotto licenza "Creative Commons Attribuzione-StessaLicenza 2.5 Italy License"  
da **Simone Brunozi** ([www.ubuntu.it](http://www.ubuntu.it) | [www.simpler.it](http://www.simpler.it) | [simone.brunozzi@gmail.com](mailto:simone.brunozzi@gmail.com))