

UNIVERSITÀ DEGLI STUDI DI BARI

DIPARTIMENTO DI INFORMATICA

Caso di studio del corso di Integrazione e test di sistemi software

Business Intelligence: dai dati alla conoscenza

Gruppo: 4Designers

Federica Lisco, Davide Di Pierro, Domenico Gigante, Graziano Castellano

**Abstract**

Al giorno d’oggi le aziende di ogni calibro sono sempre alla ricerca di strumenti, strategie e soluzioni per migliorare la propria capacità competitiva nei confronti della concorrenza. Le sorti del business sono influenzate fortemente, oltre che dalle idee, dalla risolutezza e dalla brillantezza di chi compie le scelte. I mezzi aiutano ad assumere le corrette valutazioni e le corrette decisioni.

La disciplina che si occupa della creazione di conoscenza è la Business Intelligence. Questa consente alle organizzazioni di diventare competitive prendere decisioni in maniera consapevole attraverso l’analisi delle informazioni.

Quindi “*After the information has been extracted from the data the information is yet to be interpreted the process used to interpret and derive value from information is often called as information value chain. The first step in the value chain is the extraction of data from different sources; applying different logics and business contexts to this data creates information; information is then consumed by BI users; Based on these information different decisions are made and executed; thus increasing the business value.*” [1]

Nel nostro caso di studio si è deciso di fare BI sugli incidenti stradali avvenuti in Lombardia tra gli anni 2000 e 2011.

I risultati ottenuti sono i seguenti:

//VALUTARE

“*The collected data is used to form the Data Warehouse of the enterprise and will be analyzed by the BI applications. The analysis from the BI applications is used to understand business behavior and make strategic decisions.*”[2]

**Indice**

1. Introduzione
2. Letteratura a supporto
3. Procedura ETL
4. Strumenti utilizzati
5. Esperienza
6. Conclusione
7. Bibliografia

**Introduzione**

Compiere decisioni può risultare complicato specialmente quando c’è una considerevole mole di dati alla base. Il principale problema da affrontare diventa, quindi, quello di estrapolare le informazioni a partire da questi dati. Sebbene dato e informazione siano due termini spesso utilizzati come sinonimi, hanno un significato differente, così come sottoscrive Drucker affermando che *“Information is data endowed with relevance and purpose”*[3].

Questa differenza viene evidenziata anche da altri tra cui Hicks che definisce il dato come: “*A representation of facts, concepts or instructions in a formalised manner suitable for communication, interpretation, or processing by humans or by automatic means*.”[4], o ancora è definito come *“Raw facts that can be shaped and formed to create information.”.*[5]

Mentre l’informazione è *“data that has been processed so that it is meaningful to a decision maker to use in a particular decision. ”*[4], oppure *“Data that have been shaped or formed by humans into a meaningful and useful form.”*[5]

Tuttavia, dopo aver estrapolato le informazioni dai dati, il problema consiste nel non riuscire ad usare questa ingente mole di informazioni per ottenere conoscenza e valutare le prestazioni di un determinato processo aziendale. Definiamo i processi aziendali come *“quite simply, the processes a business must execute in order for the business to operate.”*[6]

Per conoscenza, invece, si intende *“a fluid mix of framed experience, values,contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information.”*[7]

Il processo che trasforma i dati e le informazioni in conoscenza è detto Business Intelligence ed è definito come: *“the intelligence as information valued for its currency and relevance. It is expert information, knowledge and technologies efficient in the management of organizational and individual business. Therefore, in this sense, business intelligence is a broad category of applications and technologies for gathering, providing access to, and analyzing data for the purpose of helping enterprise users make better business decisions.”* [8]

Un’altra definizione di Business Intelligence (BI) è “*the techniques or practices which utilize different technologies to create different methods or applications which analyze the business data available with the organization to help the enterprise to take decisions based on the predictions made by the data.*”[9]

(Integrare con la frase precedente).È possibile ottenere la suddetta conoscenza attraverso l’uso di un complesso strumento chiamato Business Intelligence System (BI System), cioè *“an integrated set of tools, technologies and programmed products that are used to collect, integrate, analyse and make data available”*[10]

È importante sottolineare che la BI non comprende soltanto le tecnologie di elaborazione e analisi dei dati ma anche molte pratiche e metodi di business che possono essere applicati a vari settori (e-governance, assistenza sanitaria, sicurezza, …). [1], [9]

Il BI System elabora informazioni a partire da una sorgente, chiamata Data Warehouse definita come: “*an integrated and time-varying collection of data primarily used in strategic decision making by means of online analytical processing (OLAP) techniques. It is essentially a database that stores integrated, often historical, and aggregated information extracted from multiple, heterogeneous, autonomous, and distributed information sources.”*[11]

Costruire un grande Data Warehouse spesso porta a un maggiore interesse nell'analisi e nell'utilizzo dell'accumulo di dati storici. Una soluzione sarebbe analizzare i dati in un Data Warehouse mediante analisi analitica on-line con strumenti di elaborazione (OLAP). Per OLAP (On-Line Analysis Processing) si intende “*software for manipulating multidimensional data from a variety of sources that has been stored in a data warehouse. The software can create various views and representations of the data. OLAP software provides fast, consistent, interactive access to shared, multidimensional data. These systems are used to discover trends, analyze critical factors and perform statistical analysis*.[12]

L’OLAP nasce però grazie al mondo OLTP (On-Line Transaction Processing) che è definito come: “*a class of systems that supports or facilitates high transaction-oriented applications. OLTP’s primary system features are immediate client feedback and high individual transaction volume.”[13]*

I sistemi informativi si possono tassonomizzare nelle seguenti quattro tipologie utili al fine di costituire quel valore aggiunto indispensabile per un’azienda che opera nella società dell’informazione:

* TPS (Transaction Processing System). Sistema informativo a livello operativo di elaborazione delle transazioni.
* MIS (Management Information System). Sistema informativo di gestione delle informazioni per i middle manager.
* DSS (Decision Support System). Sistema informativo di supporto alle decisioni.
* ESS (Executive Support System). Sistema informativo a livello strategico e di supporto direzionale.

I primi due fanno riferimento al mondo OLTP, mentre i restanti al mondo OLAP.

Per TPS si intende “*an information processing system for business transactions involving the collection, modification and retrieval of all transaction data. Characteristics of a TPS include performance, reliability and consistency.*” [14]

Per MIS si intende “*a computerized database of financial information organized and programmed in such a way that it produces regular reports on operations for every level of management in a company.*” [15]

Per DSS si intende “*a computer program application that analyzes business data and presents it so that users can make business decisions more easily. It is an informational application that collects the*[*data*](https://searchdatamanagement.techtarget.com/definition/data)*in the course of normal business operation.*” [16]

Un’ altra definizione concerne il DSS è: “*Decision support systems (DSS) are computer technology solutions that can be used to support complex decision making and problem solving*.”[17].

Per ESS si intende “*a software that allows users to transform enterprise data into quickly accessible and executive-level reports, such as those used by billing, accounting and staffing departments. It enhances decision making for executives.*” [18]

Riassumendo, possiamo affermare che lo scopo principale della tecnologia dei Data Warehouse è proprio quello di riorganizzare e sintetizzare le informazioni immagazzinate dai sistemi operazionali, permettendo di condurre analisi immediate sull’andamento di determinati processi.

Definiamo Data warehousing come *“the process of constructing and using data warehouses”*[19]*.* Completata la fase di datawarehousing, si può passare all’analisi dei dati contenuti nel DW. Questa fase è detta OLAP: *“OLAP (On-Line Analytical Processing) refers to a set of data analysis techniques developed for analyzing data in data warehouses since 1990s.”*[20]

I componenti di un Data warehouse sono:

* Sistemi Sorgenti operazionali – Data Source
* Area di staging dei dati - Sistema di alimentazione;
* Area di presentazione dei dati – Data Warehouse Server
* Strumenti di accesso ai dati

La componente di Data Source è costituita dall’insieme delle sorgenti di dati che un’azienda dispone. Questi possono essere correlati o meno ed è per questo necessaria una fase di integrazione iniziale.

L’area di staging è un’area di memorizzazione ma anche un insieme di processi ETL (Extraction Transform and Loading).

Per procedura ETL si intende: “*a process in data warehousing responsible for pulling data out of the source systems and placing it into a data warehouse. ETL involves the following tasks:*

* *Extracting the data;*
* *Transforming the data;*
* *Loading the data;”* [21]

L’area di presentazione dei dati descrive la modalità con cui sono risolti i problemi legati ai database relazionali. L’area di presentazione si può basare su sistema relazione (ROLAP) o multidimensionale (MOLAP).

Per ROLAP si intende “*a form of online analytical processing that performs dynamic multidimensional analysis of data stored in a*[*relational database*](https://searchdatamanagement.techtarget.com/definition/relational-database)*.*” [22] Per MOLAP si intende “*an online analytical processing that indexes directly into a*[*multidimensional database*](https://searchoracle.techtarget.com/definition/multidimensional-database).” [23]

**Strumenti utilizzati e ETL**

**Bibliografia**

[1] P. Wazurkar, R. S. Bhadoria, and D. Bajpai, “Predictive analytics in data science for business intelligence solutions,” in *2017 7th International Conference on Communication Systems and Network Technologies (CSNT)*, 2017, pp. 367–370.

[2] A. D. Patil and N. D. Gangadhar, “OLaaS: OLAP as a Service,” in *Proceedings - 2016 IEEE International Conference on Cloud Computing in Emerging Markets, CCEM 2016*, 2017, pp. 119–124.

[3] P. F. Drucker, “The Coming of the New Organization Harvard Business Review,” 1987.

[4] J. O. Hicks, *Management Information Systems: A User Perspective*. West Publishing Company, 1993.

[5] K. C. Laudon and J. P. Laudon, “Business information systems: a problem solving approach.” Dryden Press, Chicago, p. xxiii, 631 p., 1991.

[6] A. Elkin and S. Opitz, “Method and system for top-down business process definition and execution.” Google Patents, 2007.

[7] T. H. Davenport, P. D. M. I. S. M. T. H. Davenport, L. Prusak, and I. York University - NetLibrary, *Working Knowledge: How Organizations Manage what They Know*. Harvard Business School Press, 1998.

[8] J. Ranjan, “Business intelligence: Concepts, components, techniques and benefits,” *J. Theor. Appl. Inf. Technol.*, vol. 9, no. 1, pp. 60–70, 2009.

[9] G. Shmueli and O. R. Koppius, “PREDICTIVE ANALYTICS IN INFORMATION SYSTEMS RESEARCH 1.”

[10] W. Yeoh and A. Koronios, “Critical success factors for business intelligence systems,” *J. Comput. Inf. Syst.*, vol. 50, no. 3, pp. 23–32, 2010.

[11] B. Hüsemann, J. Lechtenbörger, and G. Vossen, *Conceptual data warehouse design*. Universität Münster. Angewandte Mathematik und Informatik, 2000.

[12] “On-line Analytical Processing (OLAP).” [Online]. Available: http://dssresources.com/glossary/142.php. [Accessed: 01-Dec-2018].

[13] “What is Online Transaction Processing (OLTP)? - Definition from Techopedia.” [Online]. Available: https://www.techopedia.com/definition/24436/online-transaction-processing-oltp. [Accessed: 09-Dec-2018].

[14] “What is a Transaction Process System (TPS)? - Definition from Techopedia.” [Online]. Available: https://www.techopedia.com/definition/707/transaction-process-system-tps. [Accessed: 09-Dec-2018].

[15] “Management Information Systems (MIS) - Encyclopedia - Business Terms | Inc.com.” [Online]. Available: https://www.inc.com/encyclopedia/management-information-systems-mis.html. [Accessed: 09-Dec-2018].

[16] “What is decision support system (DSS)? - Definition from WhatIs.com.” [Online]. Available: https://searchcio.techtarget.com/definition/decision-support-system. [Accessed: 09-Dec-2018].

[17] J. P. Shim, M. Warkentin, J. F. Courtney, D. J. Power, R. Sharda, and C. Carlsson, “Past, present, and future of decision support technology $.”

[18] “What is an Executive Support System (ESS)? - Definition from Techopedia.” [Online]. Available: https://www.techopedia.com/definition/543/executive-support-system-ess. [Accessed: 09-Dec-2018].

[19] W. H. Inmon, “What is a data warehouse,” *Prism Tech Top.*, vol. 1, no. 1, 1995.

[20] J. Han, “OLAP mining: An integration of OLAP with data mining,” in *Proceedings of the 7th IFIP*, 1997, vol. 2, pp. 1–9.

[21] “ETL.” [Online]. Available: https://www.datawarehouse4u.info/ETL-process.html. [Accessed: 09-Dec-2018].

[22] “What is relational online analytical processing (ROLAP)? - Definition from WhatIs.com.” [Online]. Available: https://searchoracle.techtarget.com/definition/relational-online-analytical-processing. [Accessed: 09-Dec-2018].

[23] “What is MOLAP (multidimensional online analytical processing)? - Definition from WhatIs.com.” [Online]. Available: https://searchsqlserver.techtarget.com/definition/MOLAP. [Accessed: 09-Dec-2018].