

Sveliamo il mistero del data Warehouse

di [Francesco Bergamaschi](#)

Publicato il 13 Dicembre 2022

Come in un libro giallo, i **consulenti aziendali** spesso devono andare a caccia di indizi per **creare il database più adatto** alle loro esigenze.
Oggi vediamo **cosa è e come si costruisce una Data Warehouse**.

Nel precedente articolo dal titolo "[Dove si trovano i dati per la Business Intelligence?](#)", Luca (*il nostro ipotetico Dott. Luca Pacioli, NdR*) ha compreso che le **opzioni di collegamento dal software di Business Intelligence ai dati** sono:

- **opzione ottimale:** collegarsi al database che sta sotto al gestionale e non al gestionale stesso (che tra l'altro potrebbe non avere un connettore al software di BI);
- **opzione sub-ottimale:** collegarsi al gestionale (se il connettore è disponibile);
- **opzione peggiore:** collegarsi a dei file Excel, .csv, .txt o similari estratti dal gestionale (opzione peggiore perché questi file non sono database).

Luca ha dunque, in suo possesso, le **chiavi per connettersi al database SQL Server** che sta sotto al gestionale che usa, SconosciutoGest.

Queste chiavi consistono in:

- indirizzo del server (indirizzo del database, dove il database si trova);
- username (per identificarsi);
- password legata allo username.

Luca le ha ottenute dal suo referente informatico (interno o esterno).



Connettersi al database

In effetti, i database permettono di connettersi ad essi soltanto a seguito di un processo di autenticazione, ecco perché bisogna fornire username e password quando ci si connette ad essi.

Alla fine del precedente articolo, tuttavia, restava un mistero da risolvere, che sarà oggetto di questo articolo: cosa intendeva il referente informatico quando ha chiesto se Luca volesse collegarsi al database SQL Server che sta sotto il gestionale o (nel caso lo avesse a disposizione) ad un Data Warehouse (a volte indicato con DW, a volte con Data Warehouse)?

Andiamo con ordine.

Cosa è un data Warehouse?

Un Data Warehouse non è altro che un database. Avendo già un database sotto al gestionale, per esempio di tipo [SQL Server](#) ma potrebbe essere Oracle o IBM DB2, che cosa c'entra un altro database nella storia?

La risposta è che, nelle aziende che hanno un certo livello di risorse informatiche, si crea in effetti un altro database, il Data Warehouse, in aggiunta al database che sta sotto ai gestionali per le seguenti ragioni:

- il database sotto il gestionale, lo chiameremo d'ora in poi per brevità e generalità DBGest, è quello con cui il gestionale scambia informazioni; dunque, è bene – se possibile, non è un vincolo assoluto, ma gli informatici preferirebbero... – evitare di connettersi ad esso, in modo che venga lasciato “in pace” a fare il suo lavoro per il gestionale;
- DBGest è un database che ha la struttura ideale per servire il gestionale; dunque, ha tantissime tabelle con tante chiavi, in modo da conservare efficientemente i dati minimizzandone le ripetizioni ed offrire flessibilità in caso di necessità di modifiche.
Tuttavia, quello che si fa con la BI non è quello che si fa col gestionale (inserire, modificare, cancellare righe di tabelle) e, dunque, la struttura a tantissime tabelle (si chiama normalizzata)

non è ideale per fare BI che deve analizzare i dati;

- DBGest riflette istantaneamente tutte le righe presenti in esso, che cambiano continuamente perché gli operatori del gestionale fanno operazioni con una certa frequenza. Può essere vantaggioso staccarsi da esso per avere un database, il Data Warehouse, che cambi più lentamente. La BI è, infatti, “lenta” normalmente, è abbastanza comune che i dati coinvolti nella BI abbiano da 12 a 24 ore di età.

Insomma, questo Data Warehouse deve esserci o no?

La risposta è: non deve esserci per forza.

Se c'è, è meglio collegarsi ad esso per creare un report di BI, se non c'è ci si collega a DBGest o si prosegue con le altre opzioni. Ecco, dunque la lista delle opzioni di collegamento ai dati aggiornata:

- **opzione ottimale:** collegarsi al Data Warehouse (se presente);
- **opzione quasi ottimale:** collegarsi al database che sta sotto al gestionale e non al gestionale stesso (che tra l'altro potrebbe non avere un connettore al software di BI);
- **opzione sub-ottimale:** collegarsi al gestionale (se il connettore è disponibile);
- **opzione peggiore:** collegarsi a dei file Excel, .csv, .txt o similari estratti dal gestionale (opzione peggiore perché questi file non sono database).

Bene ma... Come e quando viene creato questo Data Warehouse?

Di solito di notte, ad orari prefissati quando nessuno sta facendo operazioni sul gestionale.

Attraverso procedure informatiche che non è utile specificare qui, si prendono i dati dal DBGest e li si riversa su un database, il Data Warehouse, che ha meno tabelle (sono de-normalizzate, si dice, cioè per esempio si avrà soltanto una tabella dei prodotti e non una gerarchia costituita dalla tabella dei prodotti, dalla tabella delle sottocategorie di prodotti e da quella delle categorie di prodotti e così via – tutte le colonne di queste ultime due saranno riversate in colonne della tabella Prodotti in modo da averne

soltanto una e così per le aree geografiche, i clienti etc.) ed è staccato dal gestionale, così i dati in esso sono fissi fino al prossimo aggiornamento e, infine, non si “disturba” DBGest con la BI.

Quali aziende, normalmente, dispongono di Data Warehouse?

Normalmente aziende medio-grandi, dunque non certo una microimpresa.

Non è detto che un'impresa da 100 persone lo abbia, ma potrebbe darsi. In sostanza, come dovrete imparare ad accettare sempre più, la risposta è: “*dipende*” (It depends)!

Lo scopo di questo articolo era di dare informazioni che permettano a Luca non di diventare un informatico ma, per esempio, di partecipare in un meeting dove si parla di dati con cognizione di causa, cioè capendo un minimo di cosa si tratta per potere dare un senso alla propria partecipazione, essere attivi nella riunione e guadagnarsi così la considerazione di tutti, come è giusto che sia.

Per il resto, ognuno fa il suo mestiere, ma i confini del proprio lavoro nei confronti della terminologia informatica devono essere elastici.

Più si sa, meglio è in un mondo che cambia così velocemente, anche soltanto riferendosi a parole e termini tecnici che, nel caso si tratti di concetti informatici, sono ormai necessari per il proprio lavoro in qualunque ambito.

A cura di Francesco Bergamaschi

Martedì 13 dicembre 2022

POWER BI

Controllo di gestione con la business intelligence per le aziende

PERCORSO ONLINE ON DEMAND

OLTRE 14 ORE DI LEZIONE DIVISE IN 5 MODULI TEMATICI

Introduzione al corso

Un percorso alla scoperta di **POWER BI**, l'ecosistema di **Microsoft per la Business Intelligence**.

Una **guida pratica e operativa** per apprendere come gestire e diventare autonomi in tutto il flusso di lavoro.

Passo dopo passo, partendo da zero, si esplorerà dettagliatamente sia la parte Desktop (per la **connessione ai dati**, la modellazione e la **creazione dell'analisi**) che la parte del **Cloud** (pubblicazione, condivisione, gestione delle aree di lavoro, implementazione della sicurezza e dei ruoli).

Il Corso prevede **oltre 14 ore formative** suddivise in 5 moduli tematici e può essere fruito *on demand*, dove e quando vuoi.



[SCOPRI DI PIU' >](#)