

UNIVERSITÀ DEGLI STUDI DI BARI

DIPARTIMENTO DI INFORMATICA

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

Antonio Calabrese, Marco Casamassima, Aleksandar Kostic

Abstract

Le aziende sono interessate allo studio di metodologie per conoscere l’andamento della richiesta di prodotti al fine di poter acquisire un vantaggio competitivo sul mercato. In particolare, il loro interesse è acquistare stock di merci a prezzi vantaggiosi, ridurre la giacenza in magazzino e non creare fondi di magazzino da dover in seguito svendere.

*“Sales forecasting is among the fundamental inputs for planning decisions throughout the supply chain. Estimating future demand more accurately is critical for meeting it, while minimising inventory and other related costs. These demand estimates are often modelled based on historical patterns in the data.” [1]*

Il caso di studio preso in considerazione si occupa di fare **Business Intelligence** sulle vendite di un negozio e-commerce dall’anno 2010 in poi.

“Implicit in this definition is the idea (perhaps the ideal) that business intelligence systems provide actionable information delivered at the right time, at the right location, and in the right form to assist decision makers. The objective is to improve the timeliness and quality of inputs to the decision process, hence facilitating managerial work.” [2]

La Business Intelligence consente di analizzare i dati conservati in un data warehouse per produrre informazioni di supporto alle decisioni di mercato.

Sommario

[Abstract 1](#_Toc3139088)

[Introduzione 3](#_Toc3139089)

[Caso di studio 7](#_Toc3139090)

[Procedura ETL 11](#_Toc3139091)

[Software utilizzati 12](#_Toc3139092)

[Microsoft Power BI vs Pentaho Business Analytics 12](#_Toc3139093)

[Report 14](#_Toc3139094)

[Progettazione casi di test 16](#_Toc3139095)

[Conclusioni 24](#_Toc3139096)

[Bibliografia 25](#_Toc3139097)

Introduzione

Con l’espressione “Business Intelligence” (d’ora in poi BI) ci riferiamo ai modelli, metodi e tecniche che aiutano le compagnie a migliorare le loro performance per conseguire migliori opportunità di sviluppo e di crescita.

Questo sistema utilizza software che processano i dati per creare profili di interesse di ogni punto d’azione di una organizzazione.  
  
BI è definita come “*the process that transforms data into information and then into knowledge*” [3], ovvero il processo che trasforma il dato prima in **informazione** e poi in **conoscenza**.

Una delle risorse più importanti per ogni azienda è **l’informazione ottenibile dal dato**. Da ciò possiamo dedurre che esiste una sostanziale differenza tra il concetto di dato e di informazione.

I dati sono definiti come *“facts or figures — bits of information, but not information itself.”* [4]La relazione che esiste tra dati e informazione, invece, è definita nel seguente modo*: “When data are processed, interpreted, organized, structured or presented so as to make them meaningful or useful, they are called information. Information provides context for data.”* [4]

In particolare, una informazione assume un particolare significato in un determinato **contesto**. L’informazione si trasforma in seguito in conoscenza. La **conoscenza** è “*understanding of or information about a subject that you get by experience or study, either known by one person or by people generally*” [5].

Ovvero la conoscenza è l’utilizzo delle informazioni apprese ed elaborate dall’individuo combinate con le esperienze, le idee e le informazioni derivate dal contesto.

I dati e le informazioni vengono trasformati in conoscenza dal processo di Business Intelligence, il quale non include soltanto gli strumenti per effettuare l’analisi dei dati ma è un sistema completo che fornisce un vero e proprio supporto al management per prendere decisioni.

*“BI can be presented as an architecture, tool, technology or system that gathers and stores data, analyzes it using analytical tools, and delivers information and/or knowledge, facilitating reporting, querying, and, ultimately, allows organizations to improve decision making.”* [3]

I dati utilizzati per svolgere BI provengono da sorgenti eterogenee e devono essere prima integrati in un’unica grande raccolta che si chiama **data warehouse**. Imnon definisce un data warehouse come “*a subject-oriented, integrated, non-volatile, time-variant collection of data organized to support management needs*” [6].

Un Data warehouse, dunque, contiene dati storici provenienti da repository diversi ed è formato da un numero limitato di tabelle. Il Data warehouse è un database “non in linea”, ovvero raccoglie dati storici ed è aggiornato con una frequenza definita dall’azienda. I dati “on-line” vengono raccolti all’interno di altri database aziendali che supportano le operazioni di vendita.

Definiamo dunque i concetti di OLAP e OLTP.

Gli strumenti **OLAP** (On-Line Analysis Processing) si occupano di analizzare i dati presenti in un Data warehouse per supportare le attività decisionali e previsionali. All’interno di questa sfera rientrano i DSS (Decision Support System) ed i ESS (Executive Support System).

La loro definizione è riassunta in: *“OLAP software enables analysts, managers, and executives to gain insight into an enterprise performance through fast interactive access to a wide variety of views of data organized to reflect the multidimensional aspect of the enterprise data.”* [7]

I sistemi **OLTP** (On-Line Transaction Processing) utilizzano la tecnologia delle basi di dati per memorizzare e gestire i dati “in linea”. All’interno della sfera OLTP sono presenti i sistemi informativi TPS (Transaction Processing System) e MIS (Management Information System).

I sistemi OTLP vengono, dunque, definiti come *“Software products devoted to the operations of a business, principally large-scale database systems, have come to be known as Online Transaction Processing Systems, or OLTP Systems.”* [8]

Il mondo OLAP si basa su quello OLTP. I sistemi OLAP (On-Line Analytical Processing) sfruttano la tecnologia dei Data warehouse per consentire l’analisi dei dati.

*“Data warehouses differ from operational databases in that they are subject oriented, integrated, time variant, non volatile, summarized, larger, not normalized, and perform OLAP.”* [9]

In sintesi, possiamo affermare che un Data warehouse è un grande database decisionale, orientato all’analisi delle informazioni.  
Esso è composto da diversi componenti: data source, sistema di alimentazione (area di staging dei dati), data warehouse server (area di presentazione dei dati) e strumenti di accesso ai dati.

I **data sources** sono le varie sorgenti di dati di cui l’azienda dispone. Questi possono essere correlati o meno ed è pertanto necessaria una fase di analisi ed integrazione iniziale.

L’area di staging dei dati è sia una area di memorizzazione che di processi **ETL** (Extraction, Transform, Loading).

Il sistema ETL viene definito come “the fundation of the data warehouse. A properly designed ETL system extracs data from the source systems, enforce data quality and consistency standards, conforms data so that separate sources can be used together, and finally delivers data in a presentation-ready format”. [10]

Ovvero il sistema si occupa di estrarre i dati provenienti dai vari repository, di trasformarli per adeguarli ad un unico formato e caricarli all’interno del Data warehouse.

L’area di presentazione dei dati descrive la modalità con cui sono risolti i problemi dei database relazionali. Essa si può basare su sistemi **ROLAP** o **MOLAP**.

Per ROLAP si intende “a form of online analytical processing (OLAP) that performs dynamic multidimensional analysis of data stored in a relational database” [11], mentre MOLAP si definisce come “online analytical processing (OLAP) that indexes directly into a multidimensional database” [11].

# Caso di studio

Per il nostro caso di studio la sorgente dei dati era costituita da un file csv contenente record di transazioni di una azienda e-commerce fornitoci dal professore.  
I dati sono stati inseriti all’interno di un Data warehouse secondo l’approccio MOLAP.

Il database contiene un’unica tabella composta dai seguenti campi:

|  |  |  |
| --- | --- | --- |
| IdOrdine | Misura: Quantitˆ | PagamentoOrdine |
| DataOrdine | Misura: Prezzo | ValoreTagliaEffettivo |
| CodStatoFattura | Outlet | NomeCat |
| CodProvinciaFattura | Brand | NomeMac |
| ComuneFatturazione | Collezione |  |
| pImponibile Ordine | Colore |  |
| Sex | NomeSes |  |

Sono stati aggiunti in seguito i seguenti campi per arricchire il Data warehouse con altre informazioni che potrebbero tornare utili nell’esecuzione delle query.

* Semestre
* Mese
* Anno
* Sesso Articolo
* Età Articolo

Inoltre, alcune colonne sono state rinominate per cercare di rendere auto esplicativo il loro nome:

|  |  |
| --- | --- |
| Nome originale colonna | Nuovo nome colonna |
| Sex | SessoCliente |
| Misura: Quantit^ | Quantità |
| Misura: Prezzo | Prezzo |
| ValoreTagliaEffettivo | Taglia |
| NomeCat | Categoria |
| NomeMac | Macro Categoria |
| NomeSes | SessoArticolo |
| PagamentoOrdine | Metodo Pagamento |

Si è proceduto stabilendo cosa costituisse i f*atti* e cosa le *dimensioni*.

Nella seguente tabella sono riportate le colonne con il loro significato.

|  |  |  |  |
| --- | --- | --- | --- |
| Nome colonna | Tipo | Significato | Dominio |
| IdOrdine | Dimensione | Identifica un ordine ed è usato per raggruppare più prodotti appartenenti allo stesso | Numero intero maggiore di 0 |
| DataOrdine | Dimensione | Identifica la data in cui è stato effettuato l’ordine | Stringa in formato gg/mm/aaaa |
| CodStatoFattura | Dimensione | Identifica il codice dello stato della fattura | Stringa di due caratteri alfabetici |
| CodProvinciaFattura | Dimensione | Identifica il codice della provincia della fattura | Stringa di due caratteri alfabetici |
| SessoCliente | Dimensione | Indica il sesso di chi ha effettuato l’ordine | Carattere m, f o stringa alfabetica “altro” |
| Quatità | Fatto | Indica la quantità dell’articolo ordinato | Numero intero maggiore di 0 |
| Prezzo | Fatto | Indica il prezzo dell’articolo | Numero reale maggiore di 0 |
| Outlet | Dimensione | Indica se l’articolo è stato acquistato in un outlet oppure no | Intero, 0 se non è stato acquistato in outlet, 1 altrimenti |
| Brand | Dimensione | Indica il brand dell’articolo acquistato | Stringa alfanumerica |
| Collezione | Dimensione | Indica di quale collezione è l’articolo | Stringa alfanumerica in formato  “Autunno – Inverno ####” o “Primavera – Estate ####”. Con #### che indica l’anno |
| Colore | Dimensione | Indica il colore dell’articolo | Stringa alfanumerica |
| SessoArticolo | Dimensione | Indica se l’articolo è destinato a pubblico maschile o femminile o se è unisex | Stringa alfabetica |
| EtaArticolo | Dimensione | Indica se l’articolo è per bambini, ragazzi o adulti | Stringa alfabetica |
| MetodoPagamento | Dimensione | Indica il metodo di pagamento utilizzato al momento dell’acquisto | Stringa alfanumerica |
| Taglia | Dimensione | Indica la taglia dell’articolo | Stringa alfanumerica |
| Categoria | Dimensione | Nome della categoria di appartenenza dell’articolo | Stringa alfabetica |
| MacroCategoria | Dimensione | Nome della macrocategoria di appartenenza della categoria | Stringa alfabetica |
| TotaleImponibile Ordine | Fatto | Costo totale dell’ordine | Numero reale maggiore di 0 |

## Procedura ETL

La procedura ETL realizzata prende in input un file sorgente in formato CSV e scrive i dati in un file CSV di output. Il file di output viene creato solo se non è esistente, altrimenti viene aggiornato con i nuovi dati.

In più, la procedura genera un file di report in formato HTML facilmente consultabile all’interno del quale vengono illustrati i risultati dell’elaborazione, evidenziando eventuali inconsistenze, duplicati ed errori all’interno dei dati analizzati.

La procedura esegue diversi controlli su tutti i campi per verificare la correttezza del file dato in input.

Inizialmente viene controllata la prima riga di intestazione per essere sicuri che essa contenga tutte le colonne necessarie e che queste siano nell’ordine corretto. Se questo requisito non è rispettato, il programma si interrompe segnalando l’errore all’utente nel file “report.html”.  
Il controllo del numero di campi viene fatto anche per ogni altra tupla presente all’interno del file CSV. Tuple non contenenti il numero corretto di campi vengono scartate e questo comportamento viene notificato nel file di “report.html”.

In seguito, l’applicazione si occupa di effettuare dei controlli sul dominio dei dati. Se le colonne non appartengono al dominio, verranno notificati sul file “report.html” la riga e la colonna che hanno riscontrato il problema, accettandoli nel nuovo file csv. Dopo aver verificato la correttezza del record, si verifica che non sia duplicato nel file sorgente e nel file csv di destinazione. Nel caso in cui lo sia, il record non verrà accettato.

Viene utilizzata una chiave primaria composta dai seguenti campi: IdOrdine, Prezzo, EtaArticolo, Quantità, Brand, Collezione, Colore, SessoArticolo, Taglia, Categoria. Poichè il prodotto non ha un vero e proprio identificatore, e la colonna Taglia contiene valori che dipendono dal tipo di articolo acquistato e dallo stato in cui è stato acquistato, i valori di questa colonna sono usati solo per migliorare la ricerca dei duplicati.

La procedura esegue dei controlli di consistenza per verificare che tutti gli ordini con lo stesso IdOrdine abbiano uguali CodStatoFattura, CodProvinciaFattura, ComuneFatturazione, TotaleImponibile, MetodoPagamento, DataOrdine e SessoCliente, notificando nel file di report eventuali problemi. I record inconsistenti verranno comunque aggiunti nel file di destinazione.

La procedura ETL impiega circa **6 secondi** per analizzare **100.000 record**. Consultando il file degli ordini, si è stimata una media di 1000 ordini al mese. Per questa quantità la procedura impiega meno di un secondo ad analizzare i nuovi dati se eseguita mensilmente, tempo più che accettabile.

## Software utilizzati

Gli strumenti utilizzati sono stati:

* IntelliJ IDEA per la realizzazione della procedura ETL in codice Java
* Gitlab per consentire ai vari membri del gruppo di lavorare contemporaneamente sul sorgente dell’applicazione
* Microsoft Power BI
* Pentaho Business Analytics

## Microsoft Power BI vs Pentaho Business Analytics

Questi due software sono stati testati e valutati dal gruppo prima di scegliere quale fosse il più adeguato al nostro caso di studio e, quindi il più completo e flessibile.

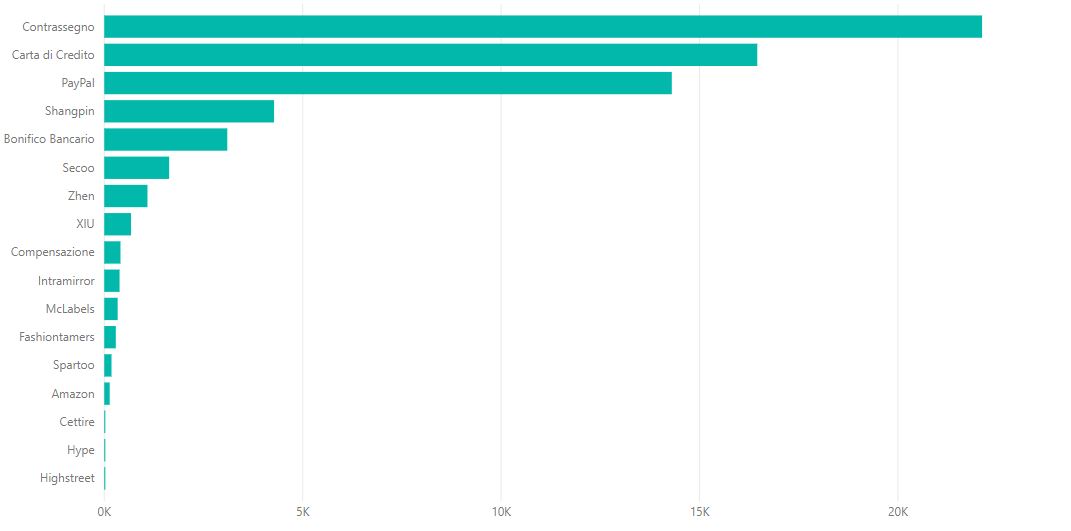
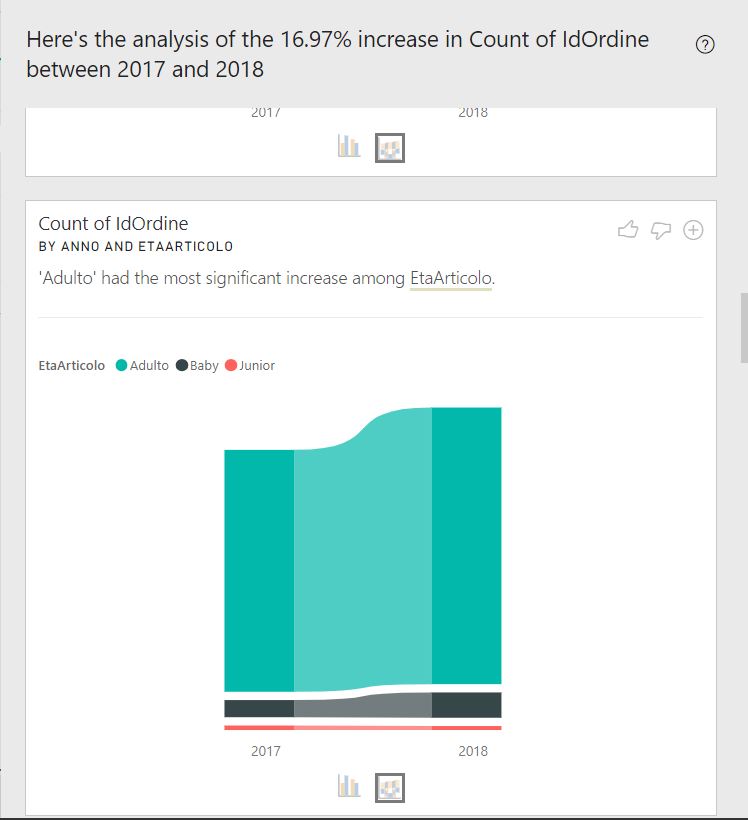
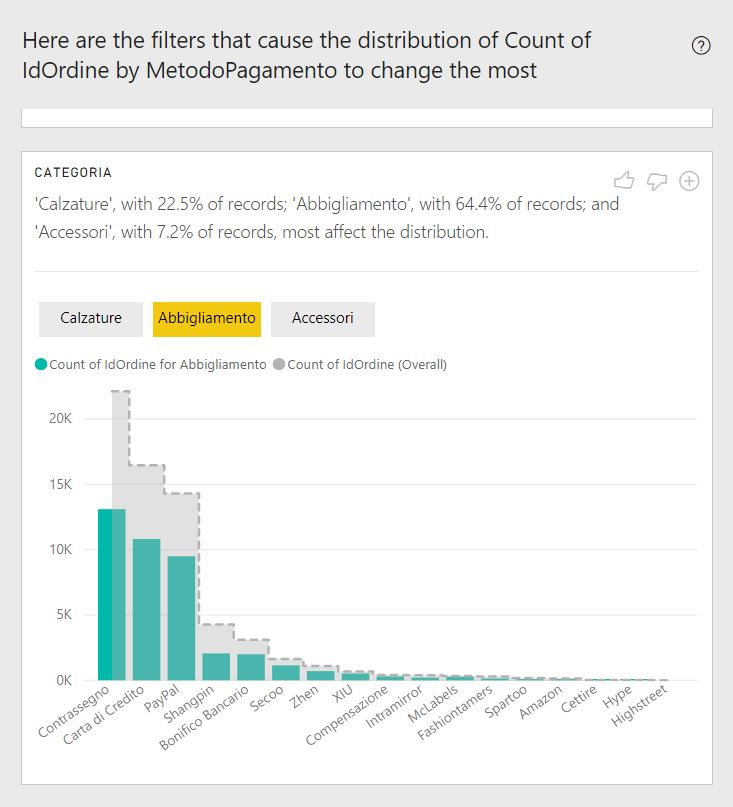
Power BI si presenta come una valida alternativa a strumenti di pagamento che operano nello stesso settore. Dispone di una licenza gratuita, è possibile connetterlo con diverse data sources (tra cui anche sql server) e dispone di tutorial abbastanza dettagliati per introdurre l’utente all’utilizzo del software, nonostante sia molto intuitivo e facile da usare.  
Power BI è una soluzione molto adatta a piccole e medie imprese che non hanno necessità sofisticate. I limiti del tool vengono mostrati quando si cerca di ottenere qualcosa in più. Ad esempio, manipolare grosse sorgenti di dati risulta complicato in quanto ricaricare i contenuti può portare a lunghi tempi di attesa.

Pentaho è uno strumento molto versatile ed efficace, ma ha una curva di apprendimento molto più ripida. La presentazione del report è pulita e professionale. Le visualizzazioni e i grafici sono piacevoli alla vista. E’ noto che l'aspetto grafico è importante nelle presentazioni di vendita e marketing.  
Pentaho risulta fornito di una serie piuttosto ampia di funzioni grafiche e di creazione di grafici che non sono invece presenti in Power BI.  
Lo strumento di integrazione dei dati è molto valido. Non sono presenti limiti alle opzioni di trasformazione. È possibile trasformare virtualmente qualsiasi file di testo, file excel o script in una tabella, un foglio di calcolo o in un rapporto. Inoltre, è in grado di processare larghe quantità di dati velocemente con l’hardware appropriato.

Possiamo concludere che Pentaho è rivolto ad un pubblico più professionale, mentre Power BI si addice maggiormente ad una impresa di modeste dimensioni.

## Report

Per la generazione del report si è scelto di utilizzare Microsoft Power BI perché più adatto a business di dimensioni contenute.

Di seguito sono riportati degli esempi di report.

## Progettazione casi di test

Attraverso il **test** si è verificato il corretto funzionamento della procedura ETL, ma possiamo evidenziare comportamenti anomali in alcune situazioni. Data la semplicità dei moduli di controllo, si è deciso di eseguire test white box.

I casi di test progettati ed eseguiti sono riassunti qui di seguito.

|  |  |
| --- | --- |
| Oggetto | Test |
| Intestazione file input | * L’intestazione del file presenta un numero corretto di campi, ma nell’ordine errato * L’intestazione del file presenta campi completamente errati * L’intestazione del file ha un campo mancante |
| Record | * Inserimento di una riga con valori non numerici nei campi che prevedevano un valore numerico (es: quantità, prezzo) * Inserimento di una riga con data non valida * Inserimento di una riga con valori nulli * Inserimento di una riga con una stringa alfabetica nel campo TotaleImponibileOrdine * Inserimento di un duplicato |
| Campi | * Inserimento di una data con mese inesistente in DataOrdine * Inserimento di una stringa che non corrisponde ad una data in DataOrdine * Inserimento di una data nel formato aaaa/mm/gg in DataOrdine * Inserimento di una data con giorno inesistente in DataOrdine * Inserimento di una data in formato gg-mm-aaaa in DataOrdine * Inserimento di una stringa alfabetica invece dell’anno in formato gg/mm/aaaa in DataOrdine * Inserimento di un carattere speciale nel mese in formato gg/mm/aaaa in DataOrdine * Inserimento di una stringa vuota in DataOrdine * Inserimento di una stringa con soli spazi in DataOrdine * Inserimento di un valore null in DataOrdine * Inserimento di “M” in SessoCliente * Inserimento di “F” in SessoCliente * Inserimento di “m” in SessoCliente * Inserimento di “f” in SessoCliente * Inserimento di un carattere alfabetico in SessoCliente * Inserimento di “Maschio” in SessoCliente * Inserimento di 0 in SessoCliente * Inserimento di una stringa vuota in SessoCliente * Inserimento di un carattere speciale in SessoCliente * Inserimento di una stringa con soli spazi in SessoCliente * Inserimento di “altro” in SessoCliente * Inserimento di un valore null in SessoCliente * Inserimento di 0 in Outlet * Inserimento di 1 in Outlet * Inserimento di “si” in Outlet * Inserimento di un carattere speciale in Outlet * Inserimento di una stringa vuota in Outlet * Inserimento di una stringa alfabetica in Outlet * Inserimento di una stringa con soli spazi in Outlet * Inserimento di un valore null in Outlet * Inserimento di 0 in Prezzo * Inserimento di un numero reale negativo con la virgola in Prezzo * Inserimento di un numero reale con il punto in Prezzo * Inserimento di un numero intero * Inserimento di un numero reale con la virgola in Prezzo * Inserimento di una stringa alfabetica * Inserimento di “!!!.!” in Prezzo * Inserimento di un numero reale negativo con il punto in Prezzo * Inserimento di una stringa vuota in Prezzo * Inserimento di una stringa con soli spazi nulli in Prezzo * Inserimento di una stringa composta da un numero reale e un carattere alfabetico in Prezzo * Inserimento di una stringa composta da numeri e “!” invece di “.” in Prezzo * Inserimento di una stringa composta da un numeri e due virgole in Prezzo * Inserimento di un valore null in Prezzo * Inserimento di un intero negativo in IdOrdine * Inserimento di 0 in IdOrdine * Inserimento di un numero intero positivo in IdOrdine * Inserimento di una stringa alfabetica in IdOrdine * Inserimento di un numero reale con la virgola in IdOrdine * Inserimento di un numero reale con il punto in IdOrdine * Inserimento di un carattere alfabetico in IdOrdine * Inserimento di una stringa vuota in IdOrdine * Inserimento di una stringa formata da caratteri alfanumerici in IdOrdine * Inserimento di un valore null in IdOrdine * Inserimento di una stringa formata solo da spazi in IdOrdine * Inserimento di una stringa formata da un numero in CodProvinciaFattura * Inserimento di una stringa formata da due lettere alfabetiche in CodProvinciaFattura * Inserimento di una stringa formata da due lettere divise da uno spazio in CodProvinciaFattura * Inserimento di una stringa formata da una lettera minuscola e una maiuscola in CodProvinciaFattura * Inserimento di una stringa vuota in CodProvinciaFattura * Inserimento di un valore null in CodProvinciaFattura * Inserimento di un carattere alfabetico e uno spazio in CodProvinciaFattura * Inserimento di due caratteri alfabetici e uno spazio in CodProvinciaFattura * Inserimento di una stringa alfanumerica in Colore * Inserimento di una stringa alfabetica in Colore * Inserimento di una stringa con due colori in Colore * Inserimento di una stringa vuota in Colore * Inserimento di un numero in Colore * Inserimento di una stringa alfabetica e un carattere speciale in Colore * Inserimento di una stringa formata solo da spazi in Colore * Inserimento di un valore null in Colore * Inserimento di una stringa alfanumerica in Quantità * Inserimento di un numero intero positivo in Quantità * Inserimento di una stringa numerica con degli spazi in Quantità * Inserimento di una stringa vuota in Quantità * Inserimento di un valore numerico reale con una virgola in Quantità * Inserimento di un valore numerico reale con un punto in Quantità * Inserimento di un valore numerico negativo in Quantità * Inserimento di un valore numerico con un carattere speciale in Quantità * Inserimento di una stringa con soli spazi in Quantità * Inserimento di un valore null in Quantità * Inserimento di un valore null in Brand * Inserimento di una stringa con soli spazi in Brand * Inserimento di una stringa alfabetica in Brand * Inserimento di una stringa alfanumerica in Brand * Inserimento di una stringa vuota in Brand * Inserimento di una stringa alfabetica con degli spazi in Brand * Inserimento di una stringa alfanumerica in Categoria * Inserimento di una stringa alfabetica in Categoria * Inserimento di una stringa alfabetica con degli spazi in Categoria * Inserimento di una stringa alfabetica con caratteri speciali in Categoria * Inserimento di una stringa alfabetica con “-“ in Categoria * Inserimento di una stringa vuota in Categoria * Inserimento di una stringa con soli spazi in Categoria * Inserimento di un valore numerico in Categoria * Inserimento di un valore null in Categoria * Inserimento di una stringa formata da un numero in CodStatoFattura * Inserimento di una stringa formata da due lettere alfabetiche in CodStatoFattura * Inserimento di una stringa formata da due lettere divise da uno spazio in CodStatoFattura * Inserimento di una stringa formata da una lettera minuscola e una maiuscola in CodStatoFattura * Inserimento di una stringa vuota in CodStatoFattura * Inserimento di un valore null in CodStatoFattura * Inserimento di un carattere alfabetico e uno spazio in CodStatoFattura * Inserimento di due caratteri alfabetici e uno spazio in CodStatoFattura * Inserimento di una stringa alfabetica con più di 2 caratteri in CodStatoFattura * Inserimento di una stringa alfanumerica in ComuneFatturazione * Inserimento di una stringa alfabetica in ComuneFatturazione * Inserimento di una stringa alfabetica con caratteri speciali in ComuneFatturazione * Inserimento di una stringa con soli spazi in ComuneFatturazione * Inserimento di una stringa vuota in ComuneFatturazione * Inserimento di un valore null in ComuneFatturazione * Inserimento di una stringa vuota in SessoArticolo * Inserimento di “Kids boy” in SessoArticolo * Inserimento di “Junior boy” in SessoArticolo * Inserimento di “Uomo” in SessoArticolo * Inserimento di “Donna” in SessoArticolo * Inserimento di “Baby boy” in SessoArticolo * Inserimento di “Kids girl” in SessoArticolo * Inserimento di “Junior girl” in SessoArticolo * Inserimento di “Baby boy” in SessoArticolo * Inserimento di una stringa diversa da quelle inserite nei precedenti 8 test * Inserimento di un valore null in SessoArticolo * Inserimento di una stringa vuota in Collezione * Inserimento di “Autunno – Primavera 2019” in Collezione * Inserimento di una stringa con soli spazi in Collezione * Inserimento di “Autunno – Inverno 2000” in Collezione * Inserimento di “Primavera – Estate 2002” in Collezione * Inserimento di una collezione senza anno in Collezione * Inserimento di una stringa alfabetica in Collezione * Inserimento di un numero in Collezione * Inserimento di un valore null in Collezione * Inserimento di un numero reale con la virgola in TotaleImponibileOrdine * Inserimento di un numero reale con il punto in TotaleImponibileOrdine * Inserimento di una stringa alfabetica in TotaleImponibileOrdine * Inserimento di una stringa con soli spazi in TotaleImponibileOrdine * Inserimento di un numero reale negativo in TotaleImponibileOrdine * Inserimento di un numero intero in TotaleImponibileOrdine * Inserimento di un numero reale con una lettera in TotaleImponibileOrdine * Inserimento di un valore null in TotaleImponibileOrdine * Inserimento di una stringa alfanumerica in MacroCategoria * Inserimento di una stringa alfabetica in MacroCategoria * Inserimento di una stringa alfabetica con degli spazi in MacroCategoria * Inserimento di una stringa alfabetica con caratteri speciali in MacroCategoria * Inserimento di una stringa alfabetica con “-“ in MacroCategoria * Inserimento di una stringa vuota in MacroCategoria * Inserimento di una stringa con soli spazi in MacroCategoria * Inserimento di un valore numerico in MacroCategoria * Inserimento di un valore null in MacroCategoria |

I test sono stati effettuati utilizzando il framework di unit testing per Java Junit 5. Il gruppo, in un primo momento, ha preso in considerazione un file csv contenenti le città con le relative nazioni di tutto il mondo. Tramite alcuni avii della procedura, si è notato che nel file report.html, si registravano degli warning con alcune città esistenti. Non potendo verificare la reale correttezza del file csv, il gruppo ha deciso di non utilizzarlo.

# Conclusioni

I dati inseriti durante le normali azioni dei livelli operativi aziendali hanno il compito di produrre informazioni e conoscenza e quindi di consentire ad una azienda di crescere.  
Molti strumenti sono fruibili per questo scopo e aiutano l’utente finale nella costruzione di un knowledge di qualità, consentendo di integrare adeguatamente i dati provenienti da diverse sorgenti e di ottenere dei report validi e fruibili nell’atto di esplorare soluzioni vantaggiose per il proprio business.

# Bibliografia

[1] Y. R. Sagaert, E. H. Aghezzaf, N. Kourentzes, and B. Desmet, “Tactical sales forecasting using a very large set of macroeconomic indicators,” *Eur. J. Oper. Res.*, vol. 264, no. 2, pp. 558–569, 2018.

[2] S. Negash, “Business intelligence (BI),” *Commun. Assoc. Inf. Syst.*, vol. 13, pp. 177–195, 2004.

[3] A. A. Santos and F. Manuel, “BUSINESS INTELLIGENCE State of the Art, Trends, and Open Issues,” pp. 296–300, 2011.

[4] Diffen, “Data vs. Information.” [Online]. Available: https://www.diffen.com/difference/Data\_vs\_Information.

[5] COLLINS, “Knowledge Definition.” .

[6] W. H. Inmon and D. Linstedt, “What a Data Warehouse is Not,” 2015, pp. 127–132.

[7] G. Colliat and George, “OLAP, relational, and multidimensional database systems,” *ACM SIGMOD Rec.*, vol. 25, no. 3, pp. 64–69, 1996.

[8] E. Thomsen, *OLAP Solutions: Building Multidimensional Information Systems*. .

[9] S. H. A. El-Sappagh, A. M. A. Hendawi, and A. H. El Bastawissy, “A proposed model for data warehouse ETL processes,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 23, no. 2, pp. 91–104, 2011.

[10] K. Ralph and C. Joe, *The Data Warehouse ETL Toolkit : Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. 2011.

[11] M. Rouse, “Techtarget.com.” [Online]. Available: https://searchsqlserver.techtarget.com/definition/MOLAP.