

Università Politecnica delle Marche

Facoltà di Ingegneria

Dipartimento di Ingegneria dell'Informazione

Corso di Laurea Magistrale in Ingegneria Informatica e dell'Automazione



Tesi di Laurea

Progettazione e implementazione di un sistema di Business Intelligence a supporto delle attività di una compagnia finanziaria

Design and implementation of a Business Intelligence system to support the activities of a financial company

Relatore

Prof. Domenico Ursino

Candidato

Laura Fratini

Anno Accademico 2020-2021

Indice

Introduzione	9
1 La Business Intelligence	13
1.1 Verso la Business Intelligence	13
1.1.1 Storia della Business Intelligence	14
1.2 Data-oriented e data-driven	15
1.2.1 I sistemi informativi	15
1.2.2 I database	17
1.2.3 I Database Management System	18
1.3 Struttura della Business Intelligence	18
1.3.1 Obiettivi della Business Intelligence	19
1.3.2 Caratteristiche della Business Intelligence	19
1.3.3 Sorgenti dei dati	19
1.3.4 Processi di ETL	20
1.3.5 Data Warehouse	21
1.4 Design di un sistema di Business Intelligence	23
1.4.1 Progettazione concettuale	23
1.4.2 Progettazione logica: OLAP	25
1.4.3 Funzionalità	26
1.4.4 Analisi What-If	26
1.5 Data Mining e Data Visualization	27
1.6 Vantaggi della Business Intelligence	28
1.7 Software a confronto	28
2 Contesto di riferimento	33
2.1 BigQuery	33
2.2 Qlik	34
2.2.1 Qlik Sense	34
2.2.2 QlikView	35
2.3 Il progetto	36
2.3.1 Architettura	36
2.3.2 Alimentazione	37

3	Analisi dei requisiti	43
3.1	Raccolta dei requisiti	43
3.1.1	Casi d'uso	43
3.2	Requisiti funzionali	44
3.2.1	Workflow operativi	46
3.3	Requisiti non funzionali	48
4	Progettazione	51
4.1	Flusso delle operazioni	51
4.1.1	Semafori indici	51
4.1.2	Semaforo Go/Not Go	52
4.1.3	Export dei dati	52
4.2	Casi d'uso	52
4.2.1	Esito positivo	52
4.2.2	Esito negativo	52
4.2.3	Esito negativo ripetuto	53
4.3	Data Model	53
5	Implementazione	57
5.1	Applicazione Quality Check in Qlik Sense	57
5.1.1	Data model	57
5.2	Flussi in BigQuery	68
6	Principali dashboard realizzate	71
6.1	Overview	71
6.2	Check 1	71
6.3	Check 2	72
6.4	Check 3	73
6.5	Check 4	73
6.6	Check 9	74
6.7	Check 12	74
6.8	Check 15	75
7	Discussione in merito al lavoro svolto	77
7.1	Considerazioni riguardo l'esperienza di tirocinio	77
7.2	Considerazioni personali	78
8	Conclusioni e uno sguardo al futuro	79
8.1	Conclusioni	79
8.2	Sviluppi futuri	79
	Riferimenti bibliografici	81
	Ringraziamenti	83

Elenco delle figure

1.1	Piramide DIKW	16
1.2	Ampliamento della piramide DIKW	17
1.3	Processo di ETL	21
1.4	Schema di organizzazione e funzionamento di un Data Warehouse in ambito Business Intelligence	23
1.5	Rappresentazione multidimensionale dei dati (fonte: <i>Data Warehouse, teoria e pratica della progettazione</i> , Golfarelli, Rizzi)	24
1.6	Esempio di gerarchia dei dati (fonte: <i>Data Warehouse, teoria e pratica della progettazione</i> , Golfarelli, Rizzi)	25
1.7	Magic Quadrant di Gartner (fonte: <i>Gartner, Febbraio 2021</i>	29
1.8	Qlik (fonte: <i>QLik</i>)	30
1.9	Tableau (fonte: <i>Tableau</i>)	31
1.10	Microsoft (fonte: <i>Microsoft</i>)	31
2.1	Architettura del sistema di Business Intelligence	37
2.2	Entità Agente	38
2.3	Entità Cliente	38
2.4	Entità Commissioni	39
2.5	Entità Consulenza	39
2.6	Entità Contratto	39
2.7	Entità Incentivi	40
2.8	Entità Portafoglio	40
2.9	Entità Transazione	41
3.1	Early requirement Analysis	45
3.2	Workflow operativo con esito positivo	46
3.3	Workflow operativo con esito di warning	47
3.4	Workflow operativo per l'esito negativo	47
4.1	Data model sistema di Quality Check	54
4.2	Diagramma di flusso del controllo bloccante	54
4.3	Diagramma di flusso del controllo con alert	55
5.1	Modello dei dati relativo ai Quality Check in Qlik Sense	69

5.2	Esempio della tabella <i>TRAFFICLIGHTS</i> in BigQuery	69
6.1	Overview dell'applicazione Quality Check.....	72
6.2	Check 1 dell'applicazione Quality Check.....	72
6.3	Check 2 dell'applicazione Quality Check.....	73
6.4	Check 3 dell'applicazione Quality Check.....	74
6.5	Check 4 dell'applicazione Quality Check.....	74
6.6	Check 9 dell'applicazione Quality Check.....	75
6.7	Check 12 dell'applicazione Quality Check.....	75
6.8	Check 15 dell'applicazione Quality Check.....	76

Elenco delle tabelle

3.1	Elenco dei requisiti funzionali	44
3.2	Elenco dei requisiti non funzionali	48
3.3	Elenco dei requisiti non funzionali	49
5.1	“Valorizzazione” dei campi per il check 1	58
5.2	“Valorizzazione” dei campi per il check 2	59
5.3	“Valorizzazione” dei campi per il check 3	59
5.4	“Valorizzazione” dei campi per il check 4	60
5.5	“Valorizzazione” dei campi per il check 5	61
5.6	Valorizzazione dei campi per il check 6	61
5.7	“Valorizzazione” dei campi per il check 7	62
5.8	Valorizzazione dei campi per il check 8	62
5.9	“Valorizzazione” dei campi per il check 9	63
5.10	“Valorizzazione” dei campi per il check 10	64
5.11	“Valorizzazione” dei campi per il check 11	64
5.12	“Valorizzazione” dei campi per il check 12	65
5.13	“Valorizzazione” dei campi per il check 13	66
5.14	“Valorizzazione” dei campi per il check 14	66
5.15	“Valorizzazione” dei campi per il check 15	66
5.16	“Valorizzazione” dei campi per il check 16	67
5.17	“Valorizzazione” dei campi per il check 17	68
5.18	“Valorizzazione” dei campi per il check 18	68

Introduzione

Negli ultimi decenni l'innovazione tecnologica ha contribuito in maniera esponenziale a facilitare la produzione, la gestione, il potenziamento, la diffusione ed il riutilizzo di tutto ciò che è connesso al patrimonio conoscitivo che un'azienda ha a disposizione.

Gli elementi che consentono alle organizzazioni di acquisire costante perfezionamento nella gestione di tale ricchezza sono:

- La costante implementazione e innovazione delle infrastrutture informatiche, finalizzate a creare un ambiente sempre più interconnesso atto ad agevolare la comunicazione e la condivisione delle informazioni.
- Lo sviluppo continuo di strumenti sempre più ricettivi ed attenti ad un'adeguata considerazione di quanto prodotto, trattato ed utilizzato dalle organizzazioni in termini di conoscenza.
- L'utilizzo di algoritmi sempre più mirati, specifici ed efficaci nella selezione dell'informazione, così come nella razionalizzazione dei relativi processi di memorizzazione e recupero dei dati.

La crescente disponibilità di risorse informatiche ad alto potenziale informativo e conoscitivo, di facile e veloce reperibilità, ha accelerato tale processo, inducendo le aziende a cercare valore aggiunto nelle informazioni provenienti dall'esterno dei propri confini, ed integrandole con quanto prodotto all'interno. Grazie a tale processo partendo da semplici blocchi, rappresentanti i dati grezzi, si può arrivare, passando per i vari stati, ad ottenere costruzioni utilizzabili, ovvero dati utili. Si tratta di un procedimento irreversibile ed in costante aumento, divenuto ormai vitale per la stessa sopravvivenza aziendale, sempre più tesa a fronteggiare la concorrenza ed a mantenere alti livelli di competitività.

Ciò ha necessariamente spinto le società a rivedere il ruolo svolto dalla Business Intelligence, arricchendolo di nuovi strumenti e procedure e creando ulteriori figure professionali.

Infatti, una mancata o cattiva gestione del dato può generare impatti negativi su almeno quattro aspetti della vita aziendale: costi, "decision making", produttività e reputazione. Una profonda conoscenza delle informazioni in possesso permette di trasformare i punti di debolezza in punti di forza, ottenendo così notevoli benefici.

La Business Intelligence, in questo senso, consente alle aziende di prendere decisioni migliori, mostrando dati attuali e storici all'interno del contesto aziendale. Può essere sfruttata per fornire benchmark su prestazioni e concorrenti e per consentire all'organizzazione di funzionare in modo più fluido e più efficiente. Inoltre, si possono individuare facilmente i trend di mercato, per aumentare le vendite o gli introiti. Se usati in modo efficace, i dati giusti possono essere utili per qualsiasi attività, dalla conformità alle assunzioni.

L'obiettivo della presente tesi è quello di esporre un'applicazione di Business Intelligence legata al contesto aziendale. In particolare, tale applicazione è parte di un progetto più ampio che ha lo scopo di:

- rafforzare l'attuale infrastruttura architettuale dei dati presenti;
- certificare la qualità del patrimonio informativo;
- fornire evidenza del valore e del contenuto del patrimonio informativo ai principali stakeholder;
- estendere la base dati sulle nuove fonti disponibili.

L'applicazione si colloca, all'interno del data lineage, a livello di data quality; ha, infatti, il fine di effettuare controlli qualitativi sui dati grezzi che vengono archiviati in modalità "mirror" rispetto ai flussi inviati dai sistemi sorgenti.

Tali controlli di data governance sono volti a garantire, in merito al dato:

- la *completezza*;
- l'*unicità*, poichè nessuna istanza dell'entità deve essere considerata più di una volta in base al modo in cui viene identificata;
- la *validità*, dal momento che i dati, per essere validi, devono rispettare i criteri imposti;
- l'*accuratezza*, ovvero il livello in cui i dati descrivono correttamente l'oggetto o l'evento considerato;
- la *consistenza*, ossia la significatività e l'effettiva utilizzabilità degli stessi.

Le principali attività che compongono, a livello generale, un processo di Data Quality Management in grado di garantire una puntuale misurazione della qualità e di gestire gli esiti di tali misurazioni sono: la verifica dell'efficacia delle attività di miglioramento svolte, la definizione, l'aggiunta e la correzione delle regole definite in base alle necessità di business, l'esecuzione di processi a breve termine per correggere eventuali errori, l'analisi dettagliata dei risultati delle misurazioni e l'esecuzione di queste ultime secondo le regole di Data Quality tracciate.

Nell'elaborato verranno analizzati ed implementati nello specifico tali aspetti.

La struttura della presente tesi è la seguente:

- Nel Capitolo 1 verrà introdotto il mondo della Business Intelligence, con un focus sui Data Warehouse.
- Nel Capitolo 2 si parlerà degli applicativi BigQuery, Qlik e dei principali dataset utilizzati.
- Nel Capitolo 3 si descriverà il processo di raccolta e analisi dei requisiti.
- Nel Capitolo 4 verrà illustrato l'approccio alla progettazione delle nuove funzionalità.

- Nel Capitolo 5 sarà analizzata la loro implementazione.
- Nel Capitolo 6 verranno mostrate le principali dashboard realizzate.
- Nel Capitolo 7 verranno espresse delle considerazioni riguardo al lavoro svolto.
- Nel Capitolo 8 saranno tratte le conclusioni, mostrando dei possibili sviluppi futuri.

La Business Intelligence

Il presente capitolo ha il compito di introdurre le basi dei processi data-driven e, in particolare, della Business Intelligence, elencandone caratteristiche, obiettivi, struttura e progettazione. Nella parte conclusiva verranno messi a confronto e descritti i principali software di BI presenti al momento sul mercato.

1.1 Verso la Business Intelligence

Di fronte alla costante crescita delle informazioni disponibili ed utilizzabili, si è assistito alla nascita di tecnologie, tecniche e strumenti in grado di raccogliere, analizzare, condividere e gestire tali informazioni in modo efficiente e, al tempo stesso, proficuo. Tali strumenti si pongono come oggetti di valido supporto in ambito decisionale, sia per quanto riguarda i processi operativi che le scelte strategiche.

Per i motivi elencati, la Business Intelligence (BI) si è posta come soluzione necessaria ed essenziale per sostenere processi di misurazione, controllo ed analisi dei risultati e delle performance aziendali.

Il reporting direzionale fa parte di quest'ultima e riassume l'insieme di rendiconti, tabelle e grafici, strutturati opportunamente in aree di responsabilità, che mettono a confronto i dati consuntivati con quelli programmati, al fine di: *evidenziare* il rispetto di obiettivi dettati dalla direzione e dal management; *identificare* eventuali cause di scostamento; *favorire* l'adozione di azioni correttive e valutare, di conseguenza, le prestazioni dei gruppi di lavoro.

Il sistema di Business Intelligence comporta, dunque:

- la raccolta dei dati del patrimonio aziendale;
- la loro pulizia, validazione e integrazione;
- la successiva elaborazione, aggregazione e analisi;
- l'utilizzo fondamentale di questa mole di informazioni nei processi decisionali.

Nell'ultimo punto si collocano sistemi di previsione, predizione, simulazione e costruzione di scenari alternativi.

1.1.1 Storia della Business Intelligence

Il termine “Business Intelligence” trova il suo primo utilizzo già nel XIX secolo; Richard M. Devens lo coniò nel 1865 all’interno della sua opera “Cyclopaedia of Commercial and Business Anecdotes” per descrivere il segreto del successo del banchiere Sir Henry Furness, capace di ottenere profitti raccogliendo informazioni politiche e di mercato ed agendo di conseguenza prima dei suoi concorrenti.

Ufficialmente il termine, però, nasce nel 1958 ad opera di Hans Peter Luhn, che pubblicò l’articolo “A Business Intelligence System”. Al tempo Luhn lavorava per la IBM (International Business Machines Corporation), l’importante azienda statunitense appartenente al settore informatico.

In questo articolo egli espose un sistema automatico, il “KWIC” (Key Word in Context), in grado di costruire rapidamente e automaticamente una sorta di indice, considerando in input un gran numero di testi, in genere articoli da 500 a 5.000 parole. All’epoca la classificazione e l’indicizzazione delle informazioni scritte avvenivano molto lentamente e fu per questo che la dimostrazione del KWIC proposta da Luhn ad un raduno a Washington di bibliotecari e scienziati dell’informazione venne accolta con grande clamore ed entusiasmo.

Un notevole apporto all’avanzamento dei sistemi di Business Intelligence fu dato dalla stessa IBM per cui lavorava Luhn, in particolar modo con la creazione di hard disk nel 1956, floppy disk e altri strumenti di memorizzazione dati. Questi contribuirono in maniera considerevole al passaggio dall’archiviazione cartacea a quella digitale, dando impulso alla diffusione di sistemi computerizzati.

L’avvento di questi dispositivi innescò la creazione del primo Database Management System, e, successivamente anche del primo Decision Support System (DSS).

Nel primi anni ’70 la crescente competizione tra i progettisti delle diverse aziende produttrici di sistemi operativi portò allo sviluppo di software sempre più sofisticati e alla creazione dei primi Data Warehouse, ovvero raccolte di dati “integrati, orientati al soggetto, variabili nel tempo e non volatili”, dalla definizione che ne diede William H. Inmon, l’inventore del termine.

Il Data Warehouse permise di comprimere di gran lunga il tempo necessario ad accedere alle informazioni richieste, collocando i dati in un’unica postazione, invece che memorizzandoli in diversi database. Di fatto migliorarono il flusso di dati dai sistemi operazionali ai sistemi di supporto alle decisioni.

Di fianco a questo primo sviluppo sorsero altri elementi che sono attualmente alla base della Business Intelligence, come, ad esempio, i tool di Extract, Transform and Load (ETL) e i software Online Analytical Processing (OLAP).

Tra il 1980 e il 2000 si è assistito ad una consistente diffusione della Business Intelligence nei sistemi aziendali, anche grazie alla definizione data da Howard Dresner nel 1989. Nell’era cosiddetta della “Business Intelligence 1.0” le funzioni svolte da quest’ultima erano principalmente due: produrre report ed organizzare e visualizzare dati in maniera presentabile al board dell’azienda. Ma continuavano a persistere due problemi legati alla tecnologia: l’esclusività dell’utilizzo dei software e dei tool soltanto da parte di utenti esperti e l’eccessivo tempo necessario all’elaborazione dei dati e alla redazione dei report. Per questo motivo i tool iniziarono lentamente ad evolversi e ad orientarsi sempre più verso utenti non esperti.

L'avvento del XXI secolo fu accompagnato da innovativi sviluppi tecnologici che hanno portato ad un vero e proprio punto di svolta e all'ingresso nell'era della "Business Intelligence 2.0". Questa si differenzia dalla precedente per l'introduzione di nuove e differenti tecnologie, quali, ad esempio, processi in real-time, che includevano informazioni provenienti da eventi così come apparivano nel Data Warehouse, permettendo alle aziende di prendere decisioni basandosi sulle informazioni più recenti. Un'altra innovazione riguarda l'estensione degli accessi anche ad utenti non esperti, che portò ad un notevole ampliamento del personale in grado di completare progetti senza l'ausilio dell'area IT.

Da questo punto in poi la Business Intelligence non si identificò più come una semplice utility aggiuntiva ma iniziò a diventare una necessità delle aziende per restare competitive e, persino per rimanere a galla in un mondo nuovo, interamente guidato dai dati.

Ad oggi i principali punti di forza della BI sono i tool specifici messi a disposizione, l'espansione di opzioni self-service e l'aumento delle visualizzazioni. Tali tool sono, nella maggior parte dei casi, "vertical market software" e, quindi, tarati per qualsivoglia azienda facente parte di un settore o mercato specifico.

1.2 Data-oriented e data-driven

I dati da cui vengono estratte le informazioni, il loro immagazzinamento, la loro manipolazione, l'analisi e la sintesi divengono campi sempre più determinanti, proporzionalmente allo sviluppo della digitalizzazione, per ottenere o mantenere un vantaggio competitivo, arrivando ad essere espedienti strategici nell'ambito decisionale e di monitoraggio.

Tale concetto è riassunto dai termini data-oriented e data-driven che mostrano il voler improntare la direzione di un apparato sulla fruizione dei dati a disposizione, per vincolarsi all'oggettività e all'empiricità che questi rappresentano.

Un volta compresa l'importanza del patrimonio informativo si può analizzare l'utilizzo che ne viene fatto, facendo una particolare distinzione tra *Business Analytics* e *Business Intelligence*. La prima cerca di anticipare eventi e situazioni, di determinare nuovi modelli e correlazioni; prima di effettuare aggregazioni analizza nel dettaglio i dati, applica formule statistiche e solo dopo trova correlazioni. Principalmente viene utilizzata a livello di marketing e di previsione del comportamento del cliente.

La Business Intelligence, invece, si focalizza sulla valutazione delle serie storiche e dell'attimo presente, evidenziandone i trend e gli aspetti, così da realizzarne benchmark attinenti. È improntata sul reporting con KPI, grafi e misure, al monitoraggio e all'analisi descrittiva e diagnostica con dashboard e analisi delle relazioni multidimensionali. Ha in comune con la Business Analytics la collezione e gestione dei dati, il Data Mining e le visualizzazioni.

1.2.1 I sistemi informativi

Un'organizzazione, in base ai processi interni corrispondenti alle funzioni, agli scopi, e al contesto con il quale interagisce, si dispone in differenti livelli ed ambiti

che necessitano, per la loro comprensione e gestione, di determinate e specifiche informazioni e conoscenze.

Inoltre maggiore è la complessità dell'organizzazione e maggiore sarà il rigore verso il dettaglio e la compattezza di cui le informazioni dovranno disporre. La complessità per la composizione di un adeguato e ben strutturato sistema informativo può di fatto dipendere da molteplici fattori. Per ovviare a queste complicità vengono utilizzati diversi strumenti e alternative che consentono di identificare e classificare i meccanismi dediti alla realizzazione e alla veicolazione dell'informazione. Si definiscono, inizialmente, l'ambiente in cui agisce il sistema, gli attori interni ed esterni, oltre che le influenze, e poi le fasi e gli elementi costituenti.

Le informazioni, in generale, possono essere viste in maniera gerarchica tramite la piramide DIKW, in Figura 1.1. Questa mostra come i dati possono essere arricchiti con il contesto per creare informazioni, a queste può essere associato un significato per creare conoscenza e la conoscenza può essere integrata per formare la saggezza.

D'altro canto anche una qualsiasi organizzazione può essere vista come costituita dai seguenti strati:

- *strato strategico*: in cima alla piramide; è occupato dai dirigenti e dai gruppi consultivi;
- *strato tattico o manageriale*: cerca di pilotare l'organizzazione per mantenerla in linea con la strategia;
- *strato operativo*: dove, l'azienda, di fatto esegue i suoi processi fondamentali e rilascia valore al cliente.

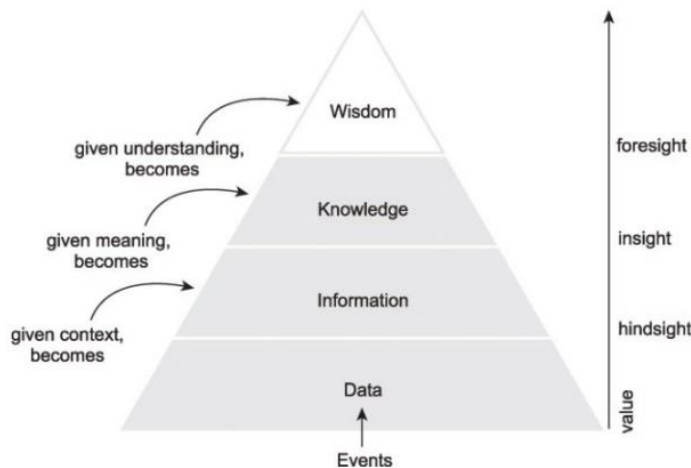


Figura 1.1. Piramide DIKW

I Big Data hanno un collegamento con la business architecture ad ogni livello, come si vede in Figura 1.2. Nel livello operativo vengono generate metriche che riportano cosa sta succedendo nell'organizzazione; a livello manageriale quest'informazione viene esaminata per rispondere a domande in merito a come viene condotto

il business; a livello strategico possono emergere altre intuizioni su quale strategia debba essere adottata o cambiata.

A questo punto il livello strategico guida la risposta tramite l'applicazione di giudizi prendendo decisioni riguardanti la strategia aziendale, la politica e gli obiettivi che vengono comunicati come vincoli al livello tattico. A sua volta quest'ultimo utilizza tale conoscenza per generare priorità e azioni verso il livello operativo, che acquisisce esperienza.

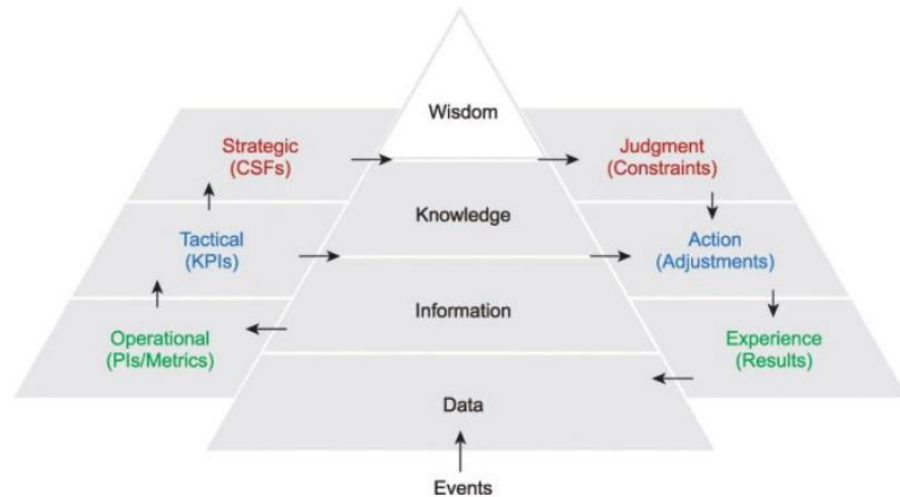


Figura 1.2. Ampliamento della piramide DIKW

1.2.2 I database

Indispensabili, per quanto detto fino ad ora, risultano essere le abilità in relazione all'archiviazione dei dati ed alla loro memorizzazione. Infatti costantemente si incrementano i volumi, comportando un continuo e crescente bisogno di capacità fisiche e conoscitive.

Al volume si aggiungono, poi, altre caratteristiche fondamentali e non meno importanti, quali, la velocità di memorizzazione, la varietà dei dati, la veridicità e il valore.

In questo contesto possono essere introdotti i dataset che, quindi, risultano essere collezioni o gruppi di dati correlati. Ciascun gruppo o membro di un dataset condivide gli stessi attributi o proprietà con gli altri gruppi o membri all'interno del dataset.

Le soluzioni di Big Data elaborano una grande varietà di dati; questi possono essere suddivisi in:

- *Dati strutturati*: si conformano a un modello o ad uno schema dei dati e vengono spesso memorizzati in tabelle all'interno database relazionali. Vengono generati dalle applicazioni delle imprese e dai sistemi informativi ERP e CRM.

- *Dati non strutturati*: non si conformano ad un modello o ad uno schema; possono essere testuali o binari e spesso sono registrati in file che sono auto-contenuti e non-relazionali. Questi dati non possono essere elaborati o interrogati direttamente per mezzo di SQL. Essi possono essere memorizzati in un DBMS relazionali come Binary Large Object (BLOB) oppure in un database Not-only-SQL (NoSQL).
- *Dati semi-strutturati*: hanno un livello di struttura e consistenza definito, ma sono per natura non-relazionali. A causa della natura testuale e alla conformità ad una struttura, vengono elaborati più facilmente rispetto ai dati non-strutturati.

1.2.3 I Database Management System

I database così costituiti necessitano di sistemi software di organizzazione, ovvero Database Management System o DBMS. Questi consentono la creazione, la manipolazione e l'interrogazione efficiente dei database. Essi si appoggiano a kernel che supportano nativamente il multitasking e il collegamento in rete. Oltre alle funzionalità di organizzazione, memorizzazione e reperimento dati, i DBMS si occupano, anche, della sicurezza e dell'integrità del database.

Anche i DBMS stessi possono essere classificati a seconda del fine ultimo dei relativi dati a disposizione. Nell'ambito della Business Intelligence si possono distinguere due differenti tipologie, che utilizzano le due seguenti tecniche:

- *Sistemi On-Line Transaction Process (OLTP)*: l'utilizzo è, principalmente, volto alla transazione ed elaborazione di query, in particolare alla lettura e alla modifica parziale del dato. I dati risultano essere di tipo corrente, dettagliati e aggiornati, senza riferimenti a dati storici o dati di diversi apparati, in quanto l'interesse è incentrato sulle attività quotidiane, operazionali. Per quanto riguarda la progettazione, i sistemi OLTP adoperano il modello Entità-Relazione mentre lo sviluppo del database è orientato all'applicazione. Le modalità di accesso al DBMS richiedono meccanismi di controllo, ripristino e concorrenza, in quanto, comunemente, un database sarà utilizzato contemporaneamente da un elevato numero di utenti. L'intento principale di questi sistemi è proporre soluzioni predefinite con poco margine di errore, in modo tale da velocizzarne le rispettive operazioni.
- *Sistemi On-Line Analytical Process (OLAP)*: si ha un'aggregazione dei dati che vengono sintetizzati e riassunti e coprono un arco di tempo sufficientemente ampio da poterne evidenziare variazioni. In questo caso non vengono gestite operazioni ordinarie, ma si ha l'esigenza di assumere direttive quando risulta necessario. Essendo l'arco temporale molto più ampio, lo è anche la numerosità, ma i dati vengono archiviati su diversi livelli al fine di garantirne una più facile estrapolazione. L'accesso è, principalmente, basato sulla sola lettura, ma le query in questione risultano essere complesse, e l'uso in questo caso è ad "hoc".

1.3 Struttura della Business Intelligence

1.3.1 Obiettivi della Business Intelligence

Principalmente, gli obiettivi che si pone la Business Intelligence sono i seguenti:

- aumentare la qualità delle informazioni;
- restituire informazioni significative;
- diffondere informazioni;
- condividere informazioni.

1.3.2 Caratteristiche della Business Intelligence

La Business Intelligence presenta le seguenti peculiarità:

- *facilità d'uso*: presentazione dei dati in un formato facilmente leggibile ed interpretabile
- *velocità*: possibilità di trattare grandi volumi di dati con tempi di risposta quasi istantanei grazie all'uso di tecniche di modellazione, memorizzazione e indicizzazione dei dati orientate all'analisi;
- *integrazione*: integrazione dei dati provenienti da fonti differenti, sia interne che esterne all'azienda. Tale processo deve essere affidabile e testato, in modo che gli utenti possano fare affidamento sui dati presenti nel Data Warehouse. Nel caso fosse necessario, prima di essere inseriti nel DW, i dati devono passare attraverso processi di pulizia e certificazione;
- *storicizzazione*: memorizzazione della storia dei cambiamenti subiti da determinati attributi selezionati, al fine di mantenere analisi storiche contestualizzate;
- *identificazione di trend e anomalie*: identificazione facile di trend nei dati, confrontando periodi e prodotti differenti, possibili solo con l'utilizzo di strumenti interattivi che permettono di effettuare operazioni di drill down/roll up (visualizzazione di un dato a diversi livelli di dettaglio) e di slice & dice (cambiamento delle dimensioni di analisi sui due assi);
- *subject orientation*: presentazione dei dati in modo da fornire una visione completa del processo aziendale superando i confini delle singole aree dei sistemi gestionali;
- *simulazione scenari*: possibilità di impostare scenari e di confrontarli con i valori reali, in particolare per applicazioni di budgeting, forecasting e planning;
- *indipendenza dal reparto IT*: possibilità per gli utenti finali di creare in autonomia i report di cui necessitano;
- *adattabilità nel tempo*: capacità di affrontare e sfruttare le inevitabili evoluzioni della realtà aziendale, dei sistemi operazionali e delle esigenze di analisi;
- *sicurezza*: possibilità di controllare, al tempo stesso in maniera stretta e flessibile, l'accesso ai dati, che, in molti casi, includono informazioni altamente riservate.

1.3.3 Sorgenti dei dati

Le fonti di dati da cui si parte per realizzare i sistemi complessi di cui si è parlato in precedenza possono provenire sia da sorgenti esterne che da sorgenti interne all'azienda.

- Le possibili *sorgenti interne* riguardano: sistemi di Customer Relationship Management (CRM), per la gestione delle relazioni con i clienti, sistema di Content Management (CMS), per la gestione dei contenuti, sistema di Enterprise Resource Planning (ERP), per l'integrazione tutti i processi di business rilevanti e di tutte le funzioni aziendali, sistema di Supply chain Management (SCM), per la gestione della catena di distribuzione, banche dati operative, flat file, e-mail, etc;
- Le possibili *sorgenti esterne* riguardano: applicazioni e sistemi di fornitori esterni, siti web, sistemi di e-commerce, social media, servizi cloud, etc.

Nella maggior parte dei casi i sistemi ricevono dati strutturati e semi-strutturati, vale a dire file Excel, CSV, JSON, XML o file di testo. A questi si aggiungono poi i metadati prodotti che apportano notevoli benefici.

La varietà dei dati, continuamente in crescita, comporta sfide per le imprese in termini di integrazione, trasformazione, elaborazione e memorizzazione dei dati.

1.3.4 Processi di ETL

I dati in arrivo subiscono procedure di estrazione, trasformazione e caricamento, che, nell'insieme, vengono chiamati processi di ETL (Extraction, Transformation and Loading). Tramite queste procedure i dati provenienti dalle diverse sorgenti, vengono uniformati affinché risultino coerenti, robusti, corretti ed omogenei prima dell'invio all'archivio che li conterrà, ad esempio il Data Warehouse.

Quindi l'attività di ETL non si limita a raccogliere ed organizzare i dati, ma agisce al fine di rendere le informazioni adeguate e disponibili per le future analisi. Per comprenderne a pieno l'importanza, Ralph Kimball e Joe Caserta, due esperti in materia, hanno stimato che i processi in questione possano raggiungere il 70% di impiego di tutte le risorse necessarie alla realizzazione del Data Warehouse.

La loro rilevanza cresce all'aumentare della complessità delle fonti, delle dipendenze reciproche dei dati, della loro qualità e dello spazio di archiviazione a disposizione.

In dettaglio i processi sono:

- *Estrazione*: durante questa prima fase i dati provenienti dalle sorgenti eterogenee vengono estrapolati; vi è, anche, una parte di data filtering, in cui i dati acquisiti sono soggetti al filtraggio automatico per la rimozione di dati corrotti o di dati inutili all'analisi. Tale procedura, nella maggior parte dei casi, avviene in background o in momenti di scarsa attività, ad esempio nelle ore notturne.
- *Pulizia*: Una volta estratti, i dati subiscono procedure volte a migliorarne la qualità in termini di correttezza e consistenza. Vengono rimossi i dati duplicati, errati, non concreti o mancanti, si correggono eventuali usi impropri di un campo o l'inconsistenza tra i valori; infine, si eliminano errori e incongruenze nei formati.
- *Trasformazione*: In questa fase vi è la conversione dei dati, affinché questi possano essere interpretati dal Management aziendale. I dati vengono, poi, uniformati nello schema del database di destinazione, integrati e consolidati.
- *Caricamento*: in questo ultimo passaggio viene popolato il Data Warehouse.

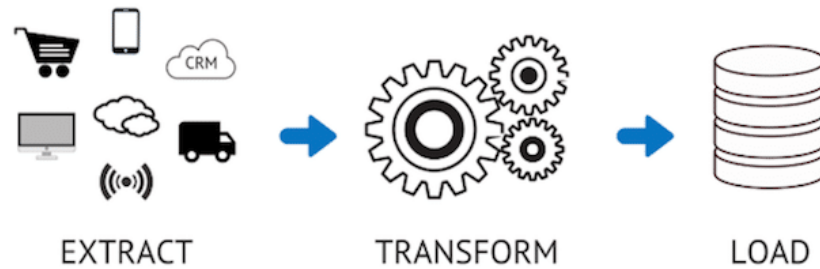


Figura 1.3. Processo di ETL

1.3.5 Data Warehouse

I dati così trattati e sottoposti ai processi di ETL sopra esposti confluiscono in collezioni o aggregazioni di dati strutturati, chiamati Data Warehouse o DW. Questi permettono ai cosiddetti knowledge worker, ovvero analisti, gestori e amministratori, di eseguire query, effettuare analisi e generare report, permettendo alle organizzazioni di ricavare un valore significativo da quei dati, oltre a mantenerne un record storico. In ogni caso, la caratteristica principale dei DW rispetto agli altri sistemi di supporto alle decisioni è l'integrazione dei dati stessi.

Inmon, il primo a coniare il termine, definisce tale raccolta:

- *Integrata:* i Data Warehouse creano coerenza tra i dati provenienti da più sistemi transazionali e fonti esterne. L'obiettivo dell'integrazione può essere raggiunto in vari modi, ad esempio mediante metodi di codifica uniforme, di omogeneità semantica delle variabili o delle unità di misura;
- *Orientata al soggetto:* Il Data Warehouse è orientato a specifici temi aziendali, a particolari applicazioni o aree funzionali. I dati, quindi, vengono archiviati in modo tale da facilitarne la lettura e l'elaborazione da parte degli utenti. Non si intende più minimizzare le ridondanze ma fornire dati organizzati in modo da favorire la produzione di informazioni. C'è il passaggio da una progettazione per funzione ad una modellazione dei dati che consenta una visione multidimensionale degli stessi;
- *Variable nel tempo:* rispetto ai dati immagazzinati in un sistema operativo, quelli di un DW coprono un orizzonte temporale molto più ampio. Questo implica che i dati presenti siano aggiornati fino ad una certa data, che, nella maggior parte dei casi, è antecedente a quella in cui l'utente interroga il sistema. Anche per questo essi si discostano dai sistemi tradizionali che, solitamente, non forniscono un quadro logico del fenomeno analizzato;
- *Non volatile:* i dati contenuti nel Data Warehouse non sono modificabili; infatti l'accesso è consentito soltanto in lettura e questo semplifica notevolmente la progettazione.

L'architettura di un Data Warehouse ricopre un ruolo estremamente importante, poichè esso deve avere la capacità di eseguire operazioni complesse e pesanti in termini di elaborazioni.

Di seguito vengono riportate le caratteristiche essenziali per un sistema di DW:

- *separazione*: elaborazione analitica ed elaborazione operativa devono essere mantenute distinte;
- *scalabilità*: l'architettura deve essere facilmente ridimensionabile, a fronte della crescita nel tempo dei volumi di dati da gestire ed analizzare e del numero di utenti da soddisfare;
- *estendibilità*: nuove tecnologie e applicazioni devono poter essere integrate senza dover riprogettare il sistema;
- *sicurezza*: il controllo degli accessi è essenziale, per via della natura strategica dei dati trattati e memorizzati;
- *amministrabilità*: la complessità delle attività di amministrazione non deve risultare eccessivo.

Le architetture sono, principalmente, di quattro tipologie:

- *Ad un livello*: l'obiettivo principale è quello di minimizzare la memorizzazione dei dati, eliminando le ridondanze. Questa tipologia di DW è virtuale, ovvero implementata come una vista multidimensionale dei dati operazionali. Questa architettura, però, non rispetta la separazione tra l'elaborazione analitica e quella operativa;
- *A due livelli o Multi-Tier*: nonostante la nomenclatura, questa architettura si suddivide in quattro livelli distinti, che descrivono stadi successivi del flusso di dati. Si parte dal livello sorgente, con le varie fonti dati, per poi passare al livello dell'alimentazione, con i processi di ETL. Successivamente vi è il livello del Data Warehouse, in cui vengono raccolte le informazioni; questo può essere consultato direttamente o utilizzato come sorgente per la costruzione di Data Mart, ovvero repliche parziali orientate verso specifiche aree del business, verso particolari divisioni dell'impresa o verso specifiche categorie di soggetti. Accanto al DW, il repository dei metadati mantiene informazioni sui meccanismi di accesso, sulle sorgenti, sulle procedure di pulizia e alimentazione, sugli utenti etc. Infine vi è il livello di analisi che permette la consultazione efficiente e flessibile dei dati integrati per simulazioni, analisi e stesura di report.
- *A due livelli con Data Mart indipendenti*: si tratta di un particolare tipo di DW in cui i Data Mart sono alimentati direttamente dalle sorgenti e sono, quindi, detti indipendenti.
- *A tre livelli*: viene introdotto un terzo livello a valle delle operazioni di ETL. Tale livello viene denominato Operational Data Store e materializza i dati operazionali ottenuti dai processi di integrazione e ripulitura.

In figura 1.4 un esempio di architettura di un Data Warehouse.

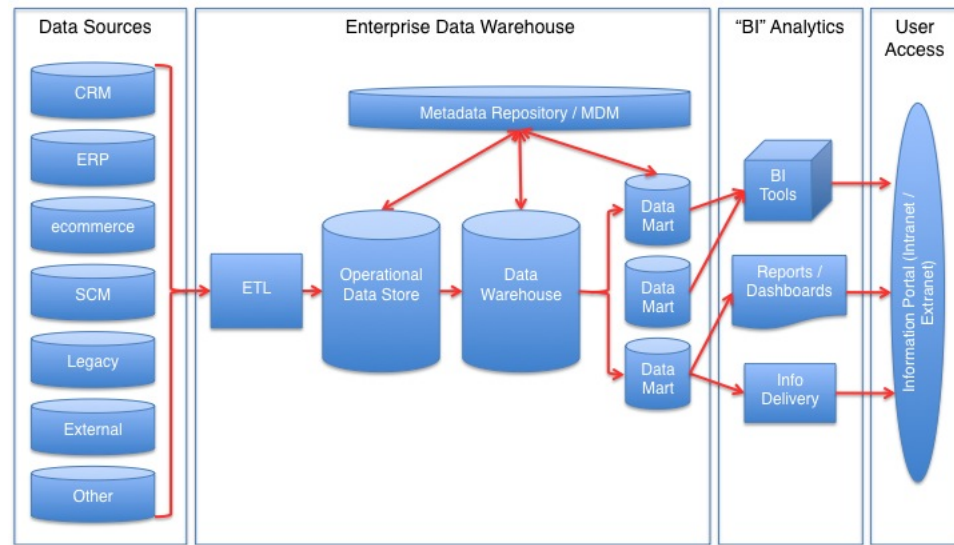


Figura 1.4. Schema di organizzazione e funzionamento di un Data Warehouse in ambito Business Intelligence

1.4 Design di un sistema di Business Intelligence

1.4.1 Progettazione concettuale

Il modello utilizzato per la progettazione, in prima fase concettuale, del sistema di Business Intelligence risulta essere il *Dimension Fact Model*, che definisce, per uno specifico fatto, schemi che modellano dimensioni, gerarchie e fatti.

Questo metodo permette di avere un modello grafico a supporto della progettazione concettuale e una documentazione utile sia prima della revisione dei requisiti con gli utenti, sia a valle.

In particolare, i concetti fondamentali di questo metodo sono i seguenti:

- *Fatto*: modella un insieme di eventi di interesse, ad esempio vendite, spedizioni etc, ed evolve nel tempo. Si tratta del gruppo di eventi su cui è incentrata l'analisi e corrisponde all'intero cubo n-dimensionale in figura 1.5.
- *Dimensione*: descrive le coordinate di analisi di un fatto, è caratterizzata da numerosi attributi di tipo categorico. Rappresenta le proiezioni del fatto verso aspetti di interesse, descrive, quindi, le coordinate di analisi
- *Misura*: descrive una proprietà numerica di un fatto, spesso oggetto di operazioni di aggregazione. È rappresentata dalla celle all'interno del cubo

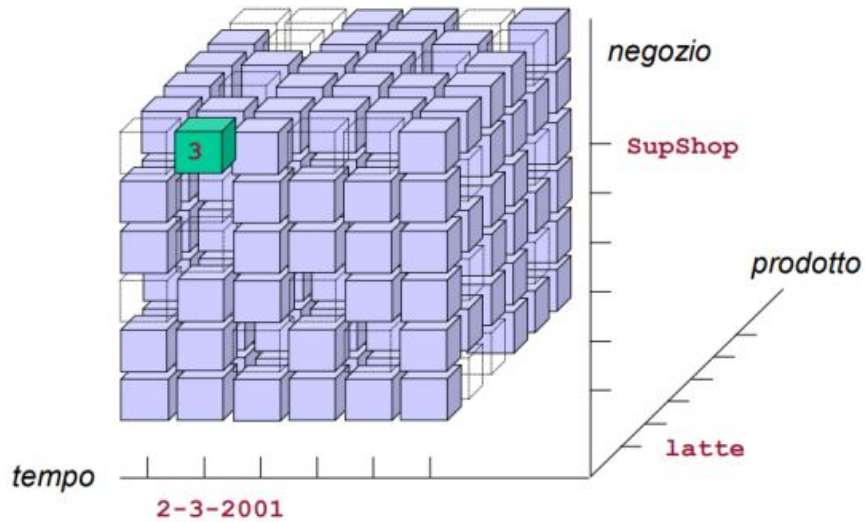


Figura 1.5. Rappresentazione multidimensionale dei dati (fonte: *Data Warehouse, teoria e pratica della progettazione*, Golfarelli, Rizzi)

Al fine di conoscere i valori delle misure contenuti in ogni cubo, è necessario fissare il valore di ogni dimensione e tracciare, per ogni asse, le perpendicolari; il punto di intersezione rappresenterà l'evento da considerare.

Nel caso riportato in figura 1.5, il fatto è rappresentato dalla vendita, le dimensioni sono rappresentate dalla data, dal prodotto e dal negozio, mentre le misure sono quantità di vendita, di incasso, di numero di clienti e di prezzo unitario.

Si parla, anche, di gerarchie, intendendo relazioni di generalizzazione tra un sottoinsieme di attributi di una dimensione, ovvero dipendenze funzionali. In merito all'esempio riportato, una rappresentazione delle possibili gerarchie è mostrata in figura 1.6.

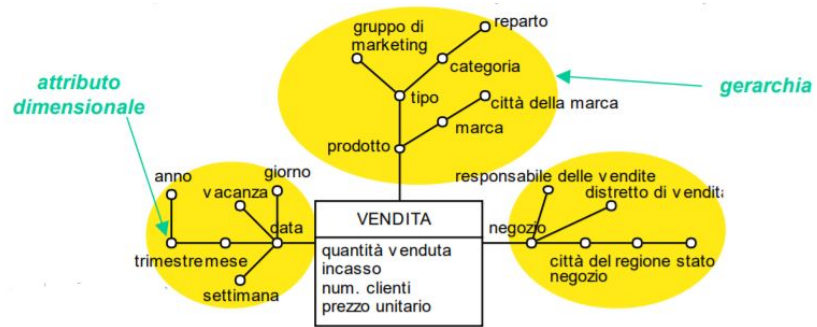


Figura 1.6. Esempio di gerarchia dei dati (fonte: *Data Warehouse, teoria e pratica della progettazione*, Golfarelli, Rizzi)

1.4.2 Progettazione logica: OLAP

Una volta ottenuto il modello concettuale, è possibile, tramite la progettazione logica, definire uno schema di base, detto schema logico, nel modello implementabile dal DBMS. In particolare, questa fase include l'insieme di passi che, partendo dallo schema dei fatti, porta al tracciamento dello schema logico di ogni Data Mart. Nella fase precedente non veniva espressa nessuna informazione in merito all'organizzazione dei dati, cosa che avviene nella fase successiva.

I modelli logici che permettono la rappresentazione della struttura multidimensionale dei dati sono di tipo OLAP (On-Line Analytical Processing) e comprendono un elevato numero di operazioni, rappresentate da query complesse eseguite su grandi volumi di dati; queste tecniche consentono un rapido accesso alle diverse informazioni, utilizzando una visione dimensionale dei dati aggregati. A seconda delle tecnologie adoperate, si distinguono due soluzioni, a cui se ne aggiunge una terza intermedia. Esse sono:

- *ROLAP (Relational On-Line Analytical Processing)*: questa tipologia opera con database relazionali. I dati e le tabelle delle dimensioni sono memorizzati come tabelle relazionali, e nuove tabelle sono create al fine di contenere le informazioni di aggregazione. Ha il vantaggio di essere scalabile e presenta requisiti di archiviazione e memorizzazione minori rispetto alle altre. Di contro, risulta essere computazionalmente onerosa, in particolar modo nella creazione delle tabelle e nella produzione di report. Essendo un'estensione dei database relazionali, è gestibile attraverso l'uso del linguaggio query standard SQL.

Tale modellazione si basa sul cosiddetto schema a stella, in cui le tabelle vengono classificate come dimensioni o fatti.

Le tabelle delle dimensioni descrivono le entità aziendali e sono costituite da una o più colonne chiave, con funzione di identificatore univoco, e da colonne descrittive.

Le tabelle dei fatti presentano l'archiviazione di osservazioni o eventi, e contengono colonne chiave correlate alle tabelle delle dimensioni e colonne di misure numeriche. La dimensionalità di una tabella dei fatti è data dalla dimensione delle colonne chiave, mentre i valori chiave della dimensione determinano la granularità. In tale schema la tabella dei fatti è correlata con le tabelle delle dimensioni.

Esiste anche lo schema del fiocco di neve, in cui vi è una tabella dei fatti composta da varie tabelle delle dimensioni e delle sotto-dimensioni collegate tramite chiave primaria e chiave esterna alla tabella dei fatti. La suddivisione dei dati in tabelle aggiuntive avviene tramite normalizzazione, e permette di ridurre le ridondanze e prevenire lo spreco della memoria.

- *MOLAP (Multidimensional On-Line Analytical Processing)*: in questo caso l'allocazione avviene tramite vettore multidimensionale ad accesso posizionale e i dati vengono memorizzati e gestiti dagli MDBS del DBMS proprietario. La vista multidimensionale non è realizzata dinamicamente, ma risulta già presente, e pertanto statica. Esso costituisce la tipologia più utilizzata poiché ha delle prestazioni molto efficienti e una maggiore velocità di accesso ai dati, in quanto non sono necessarie complesse query, ma sono sufficienti le specifiche coordinate per risalire alla combinazione di interesse in modo diretto. Questa tecnica presenta, però, la problematica della sparsità dei dati, ovvero la presenza di un elevato numero di celle con valore nullo, e comporta la creazione di ingenti quantità di dati intermedi con conseguente spreco di memoria.
- *HOLAP (Hybrid On-Line Analytical Processing)*: è una tipologia ibrida, che utilizza tabelle relazionali per la memorizzazione dei dati e tabelle multidimensionali per le aggregazioni "speculative". Presenta una velocità maggiore rispetto all'approccio ROLAP e una scalabilità maggiore rispetto all'approccio MOLAP.

1.4.3 Funzionalità

Il sistema generale OLAP mette a disposizione particolari funzionalità, quali:

- *Roll up*: si tratta di una semplificazione dei dati mediante la riduzione del livello di dettaglio di una delle dimensioni presenti, con l'aumento di livello in gerarchia, o l'eliminazione di una delle dimensioni presenti.
- *Drill down*: è un incremento del dettaglio dei dati, ottenuto tramite l'aumento del livello di dettaglio di una delle dimensioni presenti, o l'aggiunta di una nuova dimensione.
- *Pivoting*: è l'operazione di rotazione delle dimensioni di analisi.
- *Slicing*: permette di selezionare e proiettare una sola dimensione del cuboide, riducendo il volume dei dati.
- *Dicing*: in modo del tutto simile all'operazione precedente, consente la selezione di più dimensioni del medesimo cuboide.

1.4.4 Analisi What-If

I Data Warehouse hanno svolto un ruolo fondamentale all'interno delle piattaforme di Business Intelligence nel supporto alle decisioni, analizzando dettagliatamente

i dati passati. Tuttavia questi sistemi non sono in grado di dare anticipazioni sui trend futuri.

Per coprire questa necessità si è iniziato a parlare di analisi What-If, che può essere vista come una simulazione dei dati con l'obiettivo di ispezionare un sistema complesso sotto alcune ipotesi date, chiamate scenari. In particolare si misura come i cambiamenti di variabili indipendenti influiscano su un insieme di variabili dipendenti con riferimento ad un dato modello di simulazione, che è una rappresentazione semplificata del business costituita sui dati storici.

Per quanto possano apparire simili, analisi What-If e previsioni, presentano delle differenze importanti. Infatti, mentre le seconde sono normalmente eseguite estrapolando tendenze dalle serie storiche memorizzate nei sistemi, le prime richiedono la simulazione di fenomeni complessi, i cui effetti non possono essere determinati come una proiezione dei dati passati. Per il fatto di richiedere un'approssimazione soddisfacente del comportamento di un'organizzazione non vi sono piattaforme in commercio dedicate. Tuttavia software di Data Warehousing e previsione sono stati estesi con le caratteristiche dell'analisi What-If. Tra questi software si possono citare SAS BPS, SAS Forecast Server o QlikView.

1.5 Data Mining e Data Visualization

Fino a questo momento è stato esposto il processo che porta all'ottenimento di un dato pulito, integrato ed elaborato analiticamente. Una volta ottenuto un dato utilizzabile, però, sono necessari processi di analisi approfondita che permettono di trasformare i dati grezzi in informazioni rilevanti.

L'analisi dei dati esplorativa è un approccio induttivo strettamente associato al Data Mining; con questo termine si indica la capacità di estrarre conoscenze significative da informazioni utili, ma non evidenti.

I tool di Data Mining si occupano di coprire il gap tra i dati e le informazioni, cercando anche eventuali correlazioni nascoste. I principali task del Data Mining sono la scoperta di regole associative, la classificazione, la predizione, il clustering, l'analisi degli "outlier", l'analisi evolutive, il Web Mining e la Social Network Analysis.

La capacità di analizzare grandi quantità di dati e trovare intuizioni utili ha un valore limitato se i risultati possono essere interpretati solo da analisti esperti. Al fine di comunicare graficamente i risultati ottenuti, si introduce lo stadio di Data Visualization, che consente un'interpretazione efficace da parte degli utenti aziendali.

Le informazioni possono essere rappresentate in vari modi, ed è necessario utilizzare la tecnica di visualizzazione più adatta tenendo in considerazione il dominio del business.

Una volta rese disponibili le informazioni agli utenti per supportare il decision making ci possono essere ulteriori opportunità per analizzare i risultati dell'analisi; questi potrebbero produrre dei "modelli". Nella fase di Utilization of Analysis Results vengono esplorati input per sistemi enterprise, si procede con l'ottimizzazione dei Business Process e con l'invio di eventuali alert.

La rappresentazione delle informazioni può essere effettuata tramite:

- *Report ad-hoc*: comportano l'elaborazione manuale dei dati; solitamente il focus è su un'area specifica del business.
- *Dashboard*: forniscono una versione olistica delle aree di business più importanti; l'informazione viene generata ad intervalli periodici in tempo reale o quasi. In tale visualizzazione sono riportati istogrammi, grafici a torta e indicatori.

1.6 Vantaggi della Business Intelligence

In generale la Business Intelligence fornisce i seguenti vantaggi:

- *ottimizzazione dei processi*, mediante l'analisi, lo studio e la condivisione di informazioni con clienti, fornitori e partner;
- *controllo del mercato*, in particolare, il monitoraggio del mercato, l'individuazione di trend e delle opportunità di sviluppo del business;
- *semplicità*, ovvero l'utilizzo di strumenti senza la necessità di conoscenze tecniche di business;
- *flessibilità*, grazie alla sperimentazione di modelli o soluzioni differenti e alle analisi dinamiche;
- *condivisione delle informazioni*.

1.7 Software a confronto

Ad oggi esistono molteplici piattaforme di Business Intelligence che favoriscono l'elaborazione di enormi quantità di dati. Di fatto, vi è stata una grande evoluzione dei tool e dei sistemi, che, inizialmente, riportavano i dati in tabelle strutturate di fogli di calcolo e database. Un decisivo passo avanti venne fatto con l'invenzione della tabella pivot introdotta per la prima volta nel 1991 in Lotus Improv e resa nota da Microsoft Excel dal 1994. Questa, poi, nel 2000 subì un'estensione portata da Polaris, che consentiva un riepilogo grafico di grandi database multidimensionali, e ciò pose le basi del software Tableau. Successivamente Microsoft ha potenziato il programma Excel con funzionalità simili, rilasciando, nel 2015, il prodotto indipendente Microsoft Power BI.

Attualmente l'utilizzo di tali piattaforme è diventato convenzionale e ha investito i settori più svariati, dalla scienza e dall'ingegneria fino all'istruzione e al governo.

I molteplici software presenti oggi sul mercato sono confrontati annualmente in termini di prestazione da Gartner, che genera un particolare report chiamato Magic Quadrant. Questa visualizzazione è disposta su due dimensioni: l'abilità di esecuzione e la completezza visiva. A seconda delle varie grandezze è possibile individuare quattro quadranti che rappresentano, rispettivamente, i niche players, i visionaries, i challengers e i leaders. In Figura 1.7 la rappresentazione del diagramma relativo a Febbraio 2021.



Figura 1.7. Magic Quadrant di Gartner (fonte: *Gartner, Febbraio 2021*)

Come si evince dalla figura, nel quadrante dei leader, sono presenti tre tool. Essi sono:

- *Qlik*: la principale soluzione di Qlik, Qlik Sense, funziona sull'esclusivo Qlik Associative Engine, il motore cognitivo con cui sono state aggiunte le funzionalità di Intelligenza Artificiale e Machine Learning. Qlik offre flessibilità di implementazione con opzioni SaaS aziendali e ospitate dal cliente. Qlik continua a migliorare l'architettura aperta della sua piattaforma e le capacità multicloud. Insight Advisor, ora, fornisce una gamma completa di esperienze utente in Qlik Sense con analisi visiva basata sulla ricerca, sull'analisi conversazionale, sulle intuizioni associative, sulla creazione accelerata e sulla preparazione dei dati.
 - *Punti di forza*: Qlik offre la possibilità di implementare "on-premises" o con qualsiasi fornitore di cloud principale, con multicloud, o di utilizzare una

combinazione di questi approcci. Vi è, anche, un’alfabetizzazione dei dati e un’attenzione al cliente.

- *Punti di debolezza:* Qlik Sense offre funzionalità di base della piattaforma analitica e di BI in una singola licenza, ma offre anche una serie di funzionalità aggiuntive, come Qlik Catalog, Qlik Insight Advisor Chat, per esperienze chatbot, e Qlik NPrinting, per il reporting in modalità 1. Questi comportano licenze e costi aggiuntivi, se distribuiti “on-premises”. Qlik presenta, inoltre, un basso slancio di mercato; infatti, nonostante l’impegno della casa madre nella fornitura di supporto e risorse dedicate ai clienti che passano da QlikView a Qlik Sense, tale passaggio, negli utenti, desta preoccupazioni. In figura 1.8 il logo.



Figura 1.8. Qlik (fonte: *Qlik*)

- *Tableau:* offre un’esperienza di esplorazione “visual-based” che permette agli utenti aziendali di accedere, preparare, analizzare e presentare i risultati dei loro dati. Ha potenti capacità di marketing e di prodotti aziendali. Nel 2020, Tableau ha migliorato le sue capacità di preparazione e gestione dei dati, rilasciando funzionalità di modellazione dei dati migliorate, che rendono più facile l’analisi dei dati stessi su più tabelle a diversi livelli di dettaglio costruendo relazioni con una semplice esperienza visiva in-browser.
 - *Punti di forza:* Anche se Tableau continua ad aggiungere nuove capacità, mantiene sempre un’esperienza elegante per gli utenti, in modo che possano eseguire analisi senza vincoli. Anche se l’esplorazione visuale è altamente mercificata, Tableau può ancora differenziarsi offrendo un’esperienza di analisi intuitiva, con capacità più ricche basate sul suo motore brevettato Viz-QL. Gli utenti riportano entusiasmo e soddisfazione promuovendo molto positivamente l’esperienza.
 - *Punti di debolezza:* Tableau non ha un’architettura cloud-native per i clienti “on-premises” che permetta di trarre tutti i benefici del cloud. Risulta, inoltre, essere una soluzione premium pricing. Per quanto riguarda le funzionalità di analisi aumentata, gli utenti affrontano un’esperienza frammentata poichè tale caratteristica è ancora in fase di sviluppo. In figura 1.9 il logo.



Figura 1.9. Tableau (fonte: *Tableau*)

- *Microsoft*: ha una massiccia portata di mercato attraverso Microsoft Office e una roadmap di prodotti completa e visionaria. Microsoft consente la preparazione dei dati e l'esplorazione “visual-based” degli stessi, cruscotti interattivi e l'analisi aumentata in Power BI. Microsoft rilascia un aggiornamento settimanale al suo servizio Power BI basato sul cloud, che ha guadagnato centinaia di funzionalità nel 2020. Le aggiunte degne di nota includono più analisi aumentate sotto forma di esperienze di AI.
 - *Punti di forza*: L'inclusione di Power BI nella SKU E5 di Office 365 ha fornito un enorme canale per la diffusione della piattaforma. La crescente integrazione di Power BI in Microsoft Teams, con le sue decine di milioni di utenti attivi giornalieri, aumenterà ulteriormente la portata di tale tool nel mondo del lavoro remoto. L'influenza di Power BI ha ridotto drasticamente il prezzo degli strumenti nel mercato della piattaforma ABI dal suo lancio. In questo caso, però, il prezzo basso non equivale ad una funzionalità limitata, infatti il servizio cloud di Power BI è estremamente ricco nelle sue capacità, che includono un set allargato di analisi aumentate e capacità di Machine Learning automatizzate. Microsoft continua a investire in un'ampia serie di funzionalità visionarie e a integrarle con Power BI.
 - *Punti di debolezza*: Rispetto al servizio cloud Power BI, l'offerta “on-premises” di Microsoft ha significative lacune funzionali. Il modo in cui Power BI gestisce la promozione e la pubblicazione dei contenuti può portare a un significativo sovraccarico amministrativo per i clienti. Il fatto che ci sia una relazione uno-a-uno tra le app di Power BI pubblicate e i Workspace (l'ambiente di “sviluppo” collaborativo di Power BI) significa che le organizzazioni potrebbero dover affrontare una situazione in cui si trovano a gestire manualmente molte centinaia di Workspace. In figura 1.10 il logo.



Figura 1.10. Microsoft (fonte: *Microsoft*)

Contesto di riferimento

Nel seguente capitolo si esplorerà il contesto di riferimento, approfondendo i principali tool utilizzati e fornendo una panoramica sul progetto in esame.

2.1 BigQuery

Google BigQuery si presenta come un data warehouse multi-cloud serverless, a scalabilità elevata e progettato per l'agilità aziendale. Esso permette un'analisi integrativa di grandi set di dati, lavorando con Google Storage.

Le principali caratteristiche e funzionalità sono:

- *Scalabilità*: Uno dei vantaggi intrinseci del cloud computing è la capacità di ampliare l'infrastruttura su richiesta, garantendo una scalabilità dinamica della capacità applicativa, in base all'aumento delle esigenze. BigQuery ottimizza le prestazioni dei dati per archiviare ed analizzare, da petabyte, fino a exabyte di dati con facilità.
- *Interattività*: È in grado di eseguire query di selezione o di raggruppamento su miliardi di record in pochi secondi.
- *Familiarità*: Utilizza un dialetto SQL per la scrittura delle query. Infatti, BigQuery supporta un dialetto SQL standard conforme ad ANSI:2011, che riduce la necessità di riscrivere il codice. Esso fornisce, inoltre, driver ODBC e JDBC gratuiti che assicurano l'interazione di altre applicazioni con il suo potente motore. L'ambiente permette, inoltre, la materializzazione delle viste.
- *Utilità lato BI*: Costituisce la struttura portante per le moderne soluzioni di BI nel cloud e consente di eseguire senza difficoltà le attività di integrazione, di trasformazione, di analisi, di visualizzazione e di reporting dei dati, grazie agli strumenti offerti.
- *Sicurezza*: L'integrazione con i servizi per la sicurezza e per la privacy di Google Cloud fornisce una sicurezza efficace e controlli di governance granulari fino ai livelli di colonna e riga.

2.2 Qlik

Qlik è una piattaforma di analisi dei dati che supporta un portafoglio completo di soluzioni, che forniscono analisi avanzate, per soddisfare le diverse esigenze di Business Intelligence. I principali prodotti di data analytics sono Qlik Sense e QlikView.

2.2.1 Qlik Sense

Qlik Sense è un software della famiglia Qlik e si presenta come un tool estremamente potente con due funzioni principali:

- *data discovery*: consente un'esplorazione del dato profonda e trasversale, con feedback istantanei, che permettono di lavorare;
- *reporting*: mette a disposizione un numero ampio di oggetti grafici al fine di presentare e visualizzare i dati, facendo emergere relazioni ed evidenze.

La maggior parte dei tool di Business Intelligence sul mercato condividono essenzialmente la stessa architettura, che poggia su database relazionali e utilizza approcci query-based.

Questa potrebbe essere una limitazione, perché SQL non è stato progettato per supportare analisi interattive dei dati e pone dei limiti su come questi vengono utilizzati.

Nei tool query-based le sorgenti dati sono riportate tramite delle join SQL secondo la visualizzazione richiesta dall'utente.

Nel caso in cui quest'ultimo volesse indagare ulteriormente un aspetto rivelatosi di interesse, dovrebbe ri-costruire query complesse, e, quindi fare una richiesta agli sviluppatori. Questo ciclo viene detto "ask, wait, answer".

Qlik offre un Associative Engine, progettato specificatamente per analisi dei dati interattive, ottenute combinando diverse sorgenti di dati, ed indicizzandole in modo da poter dedurre le associazioni. Il motore associativo è in grado di fare calcoli ed aggregazioni "on-the-fly", per poter aggiornare le analisi ed evidenziare le relazioni tra i dati.

Alcune delle caratteristiche principali sono la possibilità di integrare dati provenienti da qualsiasi tipo di sorgente, l'utilizzo esteso ad utenti di qualsiasi livello, l'assenza di attesa propria dei processi "ask, wait, answer".

Di seguito vengono riportati i principali termini utilizzati all'interno del tool:

- Script di caricamento dei dati: lo script si connette ad una sorgente e recupera i dati.
- Modello dati: permette la strutturazione dei dati caricati.
- Misure: sono calcoli ed espressioni da utilizzare nelle visualizzazioni. Vengono create da un'espressione composta da funzioni di aggregazione, ad esempio sum o max, combinate con uno o più campi.
- Dimensioni: sono campi da utilizzare nelle visualizzazioni e determinano il modo in cui i dati vengono raggruppati.
- Fogli: raggruppano punti di raccolta di visualizzazioni di dati, ad esempio grafici e tabelle. Si crea una struttura nell'applicazione raggruppando le visualizzazioni con differenti scopi su fogli diversi.

- Segnalibri: sono collegamenti a determinate sezioni su un particolare foglio.
- Racconti: sono basati su snapshot di visualizzazione. Si presenta il dato mediante la creazione di un racconto che guida l'utente a nuove informazioni.
- Hub: sono componenti in cui sono contenute tutte le applicazioni a cui l'utente può accedere; possono essere locali o su cloud.

Quindi, per la creazione di un'applicazione, inizialmente si strutturano i fogli ancora vuoti, per avere una panoramica di ciò che verrà sviluppato e vedere se rispecchia l'obiettivo. Successivamente vengono caricati i dati di interesse, tramite il tab gestione dei dati, oppure mediante l'editor di caricamento dei dati. Le sorgenti possono essere in-app, percorsi file, connessioni dati e contenuto dati, ovvero si selezionano i dati normalizzati di Qlik DataMarket provenienti da database pubblici e commerciali. All'interno del caricamento dei dati è possibile visualizzare il modello dei dati.

Una volta caricati i dati all'interno dell'applicazione si può passare alla generazione delle visualizzazioni, che permette l'interpretazione e l'esplorazione di questi ultimi da parte dell'utente.

Qlik NPrinting

Qlik NPrinting consente alle organizzazioni di creare, gestire e condividere report di grande impatto, in modo rapido, utilizzando i dati e le analytics di Qlik Sense e QlikView per raggiungere il massimo valore.

I report prodotti possono essere di tutti i formati più diffusi, tra cui Microsoft PowerPoint, Excel e Word. È sufficiente eseguire il drag-and-drop di tabelle, grafici e oggetti esistenti, per poter combinare in un unico report oggetti di Qlik Sense o QlikView. NPrinting permette quindi di creare analytics una volta sola e da queste costruire i report, senza necessità di rielaborarli. I report realizzati presentano una qualità professionale con layout avanzati grazie ad una libreria PixelPerfect, con la risoluzione di testo e grafica, la formattazione e il dimensionamento desiderato.

Infine, si può disporre di soluzioni di analisi e reportistica per esplorare ulteriormente, approfondire in dettaglio e scoprire l'intera storia dei dati. Qlik NPrinting è realizzato in scala ed è sicuro per le aziende; esso consente di eseguire le attività simultaneamente anche durante la creazione di molti report. Vi è attenzione alla sicurezza a livello di riga, tramite i filtri combinati con l'accesso basato sul controllo.

I report possono essere attivati dalle persone indicate e inviati ad orari personalizzati o quando determinate condizioni risultano essere soddisfatte. Qlik NPrinting permette di pianificare, generare e inviare centralmente i report personalizzati tramite e-mail, directory aziendale, web, portale Qlik NPrinting NewsStand o l'hub Qlik Sense.

2.2.2 QlikView

QlikView è un software di visualizzazione e Business Intelligence della famiglia Qlik, che permette lo sviluppo di dashboard personalizzabili, in grado di fornire informazioni utili sui dati a disposizione. Lavora sulla stessa base dati di Qlik Sense, e ciò permette di scambiare i contenuti.

QlikView presenta sia la componente server che quella client e lavora su tre estensioni:

- *qvw*: si tratta di un'applicazione contenente dati, script e front;
- *qvd*: si tratta di un'applicazione contenente soltanto dati;
- *qvs*: si tratta di un'applicazione contenente soltanto script.

View ha sia un'ambiente di sviluppo sia un'ambiente di collaudo. Di fatto si tratta di un interprete che compila ed esegue le istruzioni una per volta.

Nel momento in cui si vuole creare una nuova applicazione, per prima cosa si importano i dati nello script, tramite file Excel o *qvd* di interesse o, se la fonte dati è un database esterno, si indica il connettore da cui essi devono essere presi. Tra i vari connettori è presente anche quello verso BigQuery.

2.3 Il progetto

Il progetto a cui abbiamo preso parte e da cui prende spunto questo elaborato è incentrato sulla progettazione e implementazione di un sistema di Business Intelligence a supporto delle attività di una holding in campo finance, per questo si farà riferimento ad essa come "Finance".

Gli obiettivi che tale progetto si pone sono:

- potenziare la data platform dell'azienda, rendendola un elemento centrale nell'architettura operativa presente ed abilitando le capacità analitiche;
- abilitare la creazione e la messa a disposizione di un patrimonio informativo completo secondo un disegno unitario;
- far evolvere l'architettura dati in maniera sinergica con le necessità evolutive degli utenti;
- creare una forte cooperazione con gli altri applicativi al fine di migliorare i processi elaborativi;
- attivare il processo di BI demand management con il fine di razionalizzare e prioritizzare i business requirement ed indirizzarli verso le migliori soluzioni applicative;
- formalizzare il processo di data governance con il fine di definire le best practice, le metriche di data quality e gli strumenti a supporto;
- identificare la data platform, come la Single Source of Truth del dato.

2.3.1 Architettura

Per quanto riguarda l'architettura BI, questa risulta essere strutturata su tre livelli, che formano una catena di elaborazione del dato, come mostrato in Figura 2.1.

Per questioni di riservatezza commerciale i nomi di alcuni database e moduli non possono essere indicati. In questi casi è stato utilizzato nomi fittizi.

Tali livelli sono:

- principali fonti dati: Alpha, Beta, Gamma;

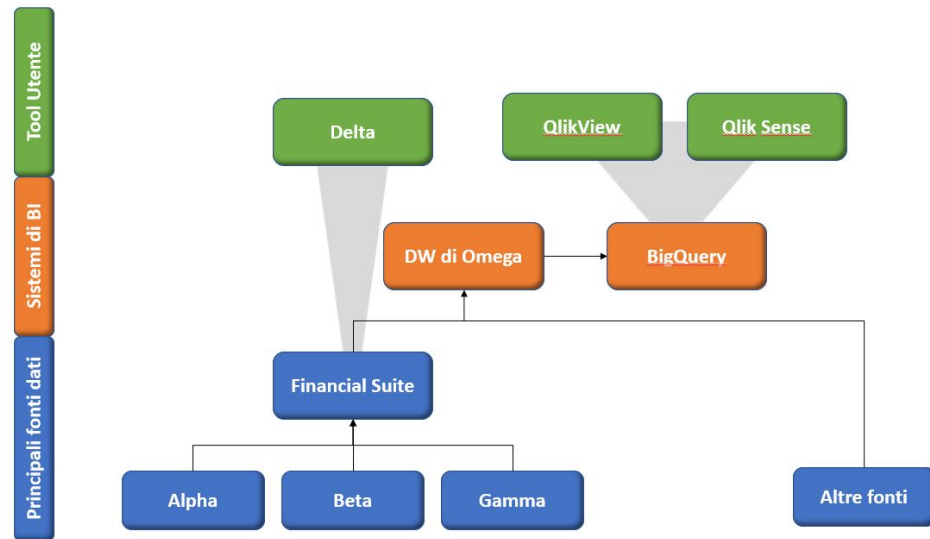


Figura 2.1. Architettura del sistema di Business Intelligence

- sistemi di Business Intelligence: Data Warehouse di Omega, società che si occupa di software in ambito finanziario, e BigQuery;
- Tool Utente: Delta, QlikView e Qlik Sense.

Utilizzando un approccio bottom-up, le principali fonti dati che costituiscono i sistemi sorgente sono:

- Alpha: relativo alla parte amministrata e ai fondi;
- Beta: relativo a gestioni, conti correnti e altre attività;
- Gamma: relativo a banche partnership o altri enti partner.

I dati provenienti da queste tre entità confluiscono in una Financial Suite, che svolge il ruolo di database principale, armonizzando e aggregando le informazioni dei sistemi sottostanti. Questo dataset è il principale alimentatore del Data Warehouse di Omega, ed è anche il sistema di interfaccia con l'applicativo Delta. Il DW di Omega è alimentato anche da altre fonti esterne.

2.3.2 Alimentazione

Il Data Warehouse di Omega è il principale alimentatore di BigQuery; infatti, da esso proviene il dataset master, che contiene la maggior parte delle tabelle presenti ed eredita le logiche dal Data Warehouse.

Nel dettaglio, il flusso di dati parte dal DW per arrivare al Google Cloud Storage, che, a sua volta, alimenta BigQuery. In particolare, i dati sono salvati in file .csv in due differenti repository.

I dati vengono caricati in BigQuery da un applicativo esterno, che elabora i file ed esegue procedure di ETL specifiche. Tali procedure sono suddivise in blocchi eseguiti a cascata, e prevedono l'aggregazione dei dati presenti nei file .csv e il confronto con

i file di quadratura. Questi ultimi permettono di effettuare una verifica dell'integrità del dato e fungono da trigger per l'elaborazione in sequenza.

In BigQuery i dati sono strutturati in un unico progetto, organizzato in dataset rappresentanti le principali macrocategorie degli stream progettuali.

Il linguaggio attualmente utilizzato è il Legacy SQL.

Le tabelle presenti, relative a dimensioni o fatti, possono essere raggruppate in entità, come segue:

- *Agente*: presenta le informazioni anagrafiche, storiche e non, relative agli agenti finanziari che operano per tale azienda. In Figura 2.2 sono riportate le entità con cui tali tabelle sono collegate.

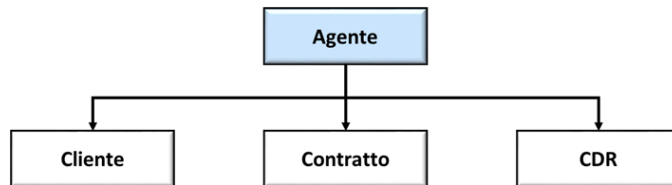


Figura 2.2. Entità Agente

- *Calendario*: presenta l'informazione del calendario solare.
- *Cliente*: presenta le informazioni anagrafiche dei clienti. In Figura 2.3 sono riportate le entità con cui tali tabelle sono collegate.

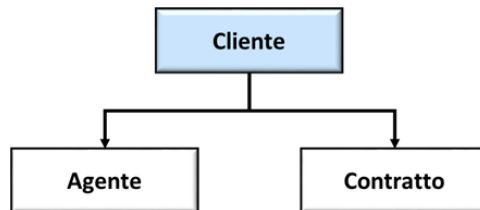


Figura 2.3. Entità Cliente

- *Commissioni*: presenta le informazioni in merito alle commissioni che vengono pagate ai promotori finanziari. In Figura 2.4 sono riportate le entità con cui tali tabelle sono collegate.
- *Consulenza*: presenta le informazioni relative ai servizi di consulenza. In Figura 2.5 sono riportate le entità con cui tali tabelle sono collegate.

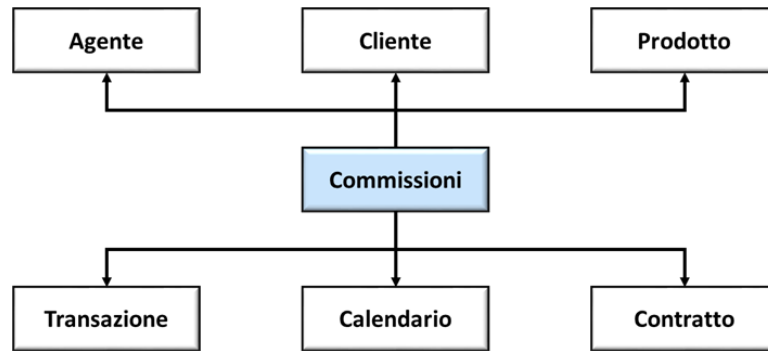


Figura 2.4. Entità Commissioni

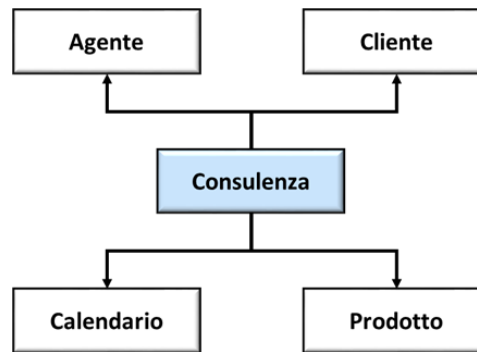


Figura 2.5. Entità Consulenza

- *Contratto*: presenta le informazioni relative ai contratti stipulati con i clienti. In Figura 2.6 sono riportate le entità con cui tali tabelle sono collegate.

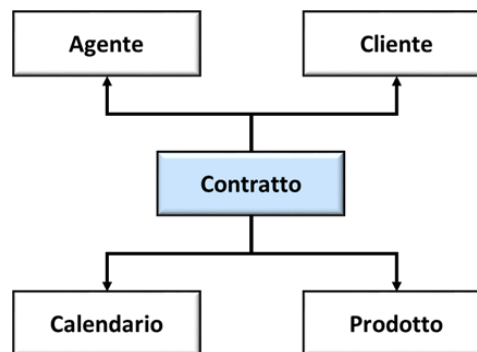


Figura 2.6. Entità Contratto

- *Incentivi*: presenta le informazioni relative agli incentivi legati ai promotori

finanziari. In Figura 2.7 sono riportate le entità con cui tali tabelle sono collegate.

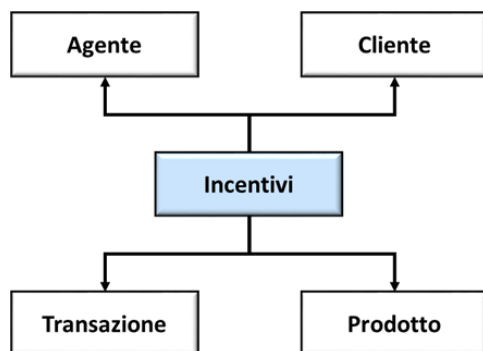


Figura 2.7. Entità Incentivi

- *Portafoglio*: presenta le informazioni legate al portafoglio del cliente. In Figura 2.8 sono riportate le entità con cui tali tabelle sono collegate.

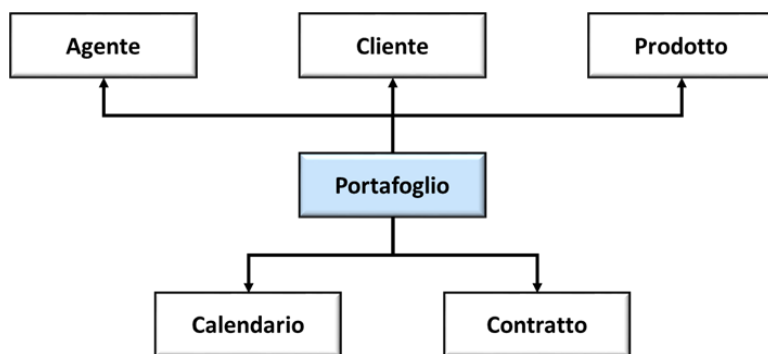


Figura 2.8. Entità Portafoglio

- *Prodotto*: presenta le informazioni relative ai prodotti messi a disposizione dall'azienda.
- *Transazione*: presenta le informazioni legate alle movimentazioni dei clienti. In Figura 2.9 sono riportate le entità con cui tali tabelle sono collegate.

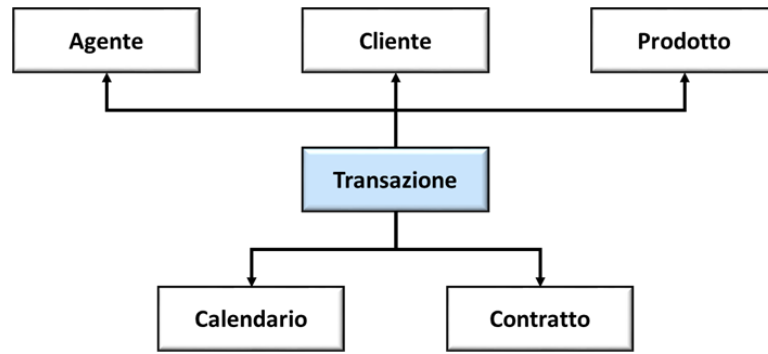


Figura 2.9. Entità Transazione

Analisi dei requisiti

Questo capitolo presenta il processo di analisi e raccolta dei requisiti in merito ad un particolare stream progettuale di back-end in ambito data quality. In particolare, verranno elencati i principali requisiti funzionali e non funzionali.

3.1 Raccolta dei requisiti

Lo stream progettuale di cui si effettueranno l'analisi dei requisiti, la progettazione e l'implementazione è volto ad effettuare un controllo sulla qualità e sulla coerenza del dato.

La raccolta dei requisiti necessari si è basata, principalmente, sullo studio della documentazione messa a disposizione dall'azienda di cui si gestisce il sistema di Business Intelligence. In tali documenti vengono presentate soglie e range che particolari dati devono rispettare, al fine di evidenziare squadrature, disallineamenti o errori nei caricamenti.

In base al tipo di dato che non rispetta il range o la soglia prestabiliti, vengono generati warning o viene bloccato l'invio dei dati ai sistemi di front-end.

3.1.1 Casi d'uso

Inizialmente è stata svolta un'analisi ad alto livello della documentazione disponibile, al fine di individuare gli attori coinvolti nel processo in esame e delineare le principali operazioni. In Figura 3.1, è stato elaborato il diagramma rappresentante l'analisi dei requisiti, in particolare la cosiddetta Early Requirement Analysis, che considera il sistema analizzato come attore. Questa analisi è stata realizzata seguendo i criteri tipici del linguaggio di modellazione I*.

In questo linguaggio compaiono gli attori, rappresentati da cerchi, i goal o obiettivi, rappresentati da ovali, e le risorse, rappresentate da rettangoli. I campi di azione degli attori vengono tracciati dai relativi boundary.

Come si può notare, sono presenti quattro attori principali che interagiscono; essi sono il Data Warehouse di Omega, BigQuery, il sistema di Quality Check e i sistemi di front-end, QlikView e Qlik Sense.

L'attore Data Warehouse di Omega ha l'obiettivo, in questo caso, di inviare i dati grezzi che ha immagazzinato a BigQuery. Questo, a sua volta, riceve i dati grezzi dal Data Warehouse di Omega, li elabora applicando regole definite dal Business, e invia i dati elaborati ai sistemi di front-end.

Il sistema di Data Quality riceve i dati elaborati da BigQuery e su questi attua i controlli prestabiliti. Una volta ottenuti i dati coerenti, li invia ai sistemi di front-end.

Questi, ricevono il dato corretto dal sistema di Quality Check e lo espongono sulle piattaforme Qlik Sense e Qlik View.

3.2 Requisiti funzionali

Dopo una prima analisi, è stato possibile definire puntualmente i principali requisiti richiesti, visibili in Tabella 3.1.

CODICE	DESCRIZIONE
RF1	La soluzione deve prevedere un controllo giornaliero.
RF2	La soluzione deve prevedere la gestione dell'esito positivo del controllo.
RF3	La soluzione deve prevedere la gestione dell'esito warning del controllo.
RF4	La soluzione deve prevedere la gestione dell'esito negativo del controllo.
RF5	La soluzione deve prevedere, in caso di esito negativo, il blocco degli scarichi dei sistemi di front-end.
RF6	La soluzione deve tenere conto dell'esito del controllo effettuato al giorno precedente $t - 1$.
RF7	La soluzione deve prevedere la possibilità di un intervento manuale.
RF8	La soluzione deve prevedere la notifica automatica, dove necessario, agli utenti.

Tabella 3.1. Elenco dei requisiti funzionali

- *RF1* rappresenta il fulcro dell'intero sistema, il requisito fondamentale. Il sistema ha il fine di implementare un controllo giornaliero sul dato, la cui tipologia spazia dal tecnico-formale all'integrità referenziale, e verificare che gli standard di qualità definiti a livello aziendale vengano rispettati.
- *RF2* dipende dall'esito del controllo effettuato, e comporta un'azione di scarico da parte dei sistemi di front-end.
- *RF3* comporta, comunque, lo scarico dei dati, ma il warning viene notificato agli utenti.
- *RF4* e *RF5* prevedono che i dati vengano, di fatto, bloccati e venga inviata la comunicazione agli utenti.
- *RF6* prevede di dover considerare nel giorno t l'esito del Quality Check al giorno $t - 1$, così da garantire il massimo controllo sul dato esposto.
- *RF7* tiene conto della possibilità di un intervento manuale, che porti al risanamento o alla comunicazione di quanto avvenuto.

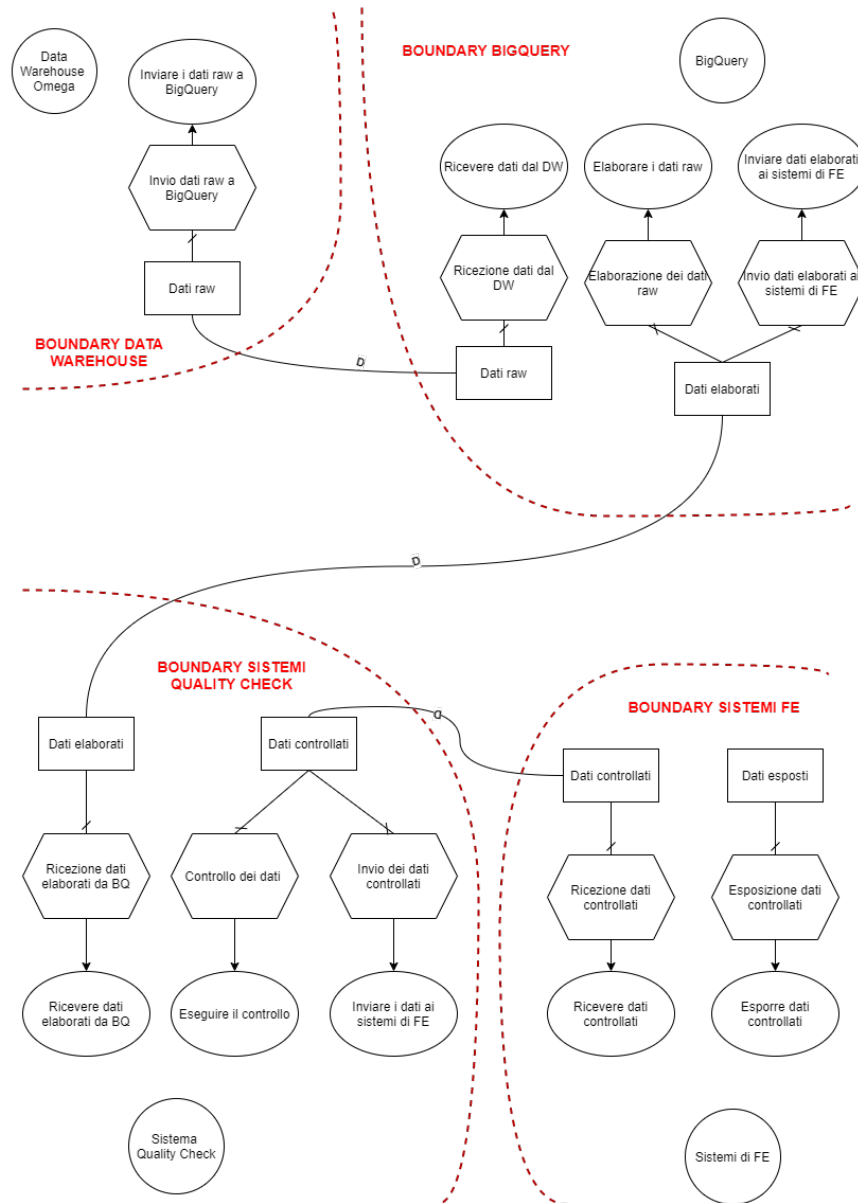


Figura 3.1. Early requirement Analysis

- *RF8* afferma che, qualora richiesto, in caso di particolari esiti, venga notificato l'accaduto all'utente in maniera automatica.

3.2.1 Workflow operativi

Per esplicitare al meglio i requisiti che il sistema di Data Quality in oggetto deve rispettare, vengono illustrati i workflow operativi per i vari esiti ammessi.

Nel caso in cui l'esito del controllo sia positivo, i file giornalieri vengono inviati a Qlik Sense e QlikView, e sono pronti per essere esposti, come riporta la Figura 3.2.



Figura 3.2. Workflow operativo con esito positivo

Se, invece, l'esito del check è “warning”, i file giornalieri vengono comunque correttamente inviati ai sistemi di front-end. Viene, poi, inviata una notifica all'utente e, in seguito, viene richiesta un'analisi da parte del gruppo di lavoro di Business Intelligence per identificare la tipologia di errore. Successivamente, viene individuato l'owner del dato e vengono calcolate le tempistiche di risoluzione. In Figura 3.3 viene riportato il relativo diagramma.

Quando l'esito del controllo risulta essere negativo, il processo si complica. In tal caso, non viene dato il “go” per l'invio dei file giornalieri ai sistemi Qlik e viene notificato all'utente il mancato invio dei dati. Anche in questo caso viene effettuata un'analisi da parte del gruppo di lavoro di Business Intelligence, con l'ausilio del Business, per individuare la tipologia di errore, che può essere tecnico, dei dati o relativo al Business stesso. A questo punto, vi è una diramazione, poichè dall'analisi può emergere che il caso in esame sia un falso negativo o un vero negativo. Il primo caso comporta il riavvio del caricamento per lo sblocco dei sistemi di front-end, mentre nel secondo caso si individua la macro categoria dell'errore, si condivide con gli stakeholder di competenza il risultato dell'analisi e si stimano le tempistiche di risoluzione. Se queste non rientrano nella giornata stessa, allora viene bloccato il caricamento fino a risoluzione. Se, al contrario, la risoluzione avviene in giornata, viene riavviato il caricamento per lo sblocco dei sistemi di front-end.

Nel caso in cui l'individuazione dell'errore non avvenisse in giornata, il sistema risulta bloccato fino alla risoluzione/individuazione del perimetro dell'errore, anche se l'esito del giorno successivo risulta positivo.

In Figura 3.4 viene mostrato il workflow operativo per l'esito negativo.

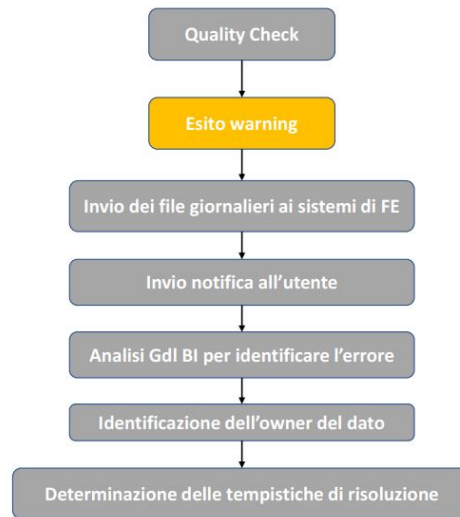


Figura 3.3. Workflow operativo con esito di warning

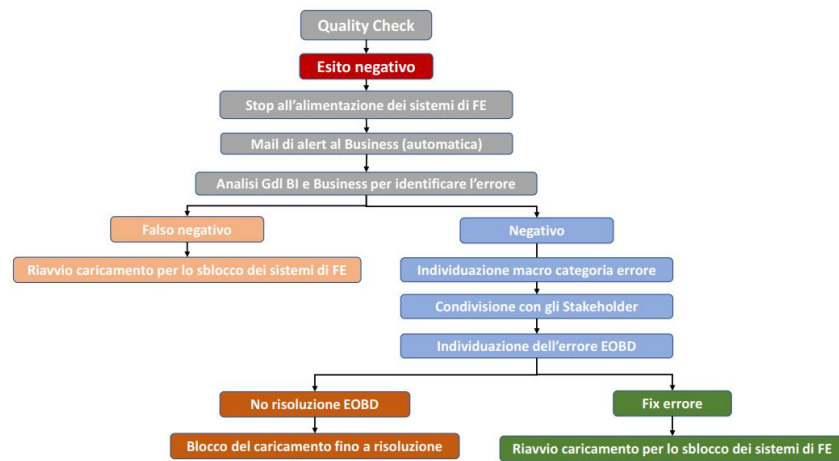


Figura 3.4. Workflow operativo per l'esito negativo

3.3 Requisiti non funzionali

I requisiti non funzionali sono stati tracciati in base alle specifiche degli standard di qualità espressi dallo stakeholder Finance. In base a questo, i requisiti non funzionali sono riportati nella Tabella 3.2 e 3.3.

CODICE	DESCRIZIONE
RNF1	La soluzione deve prevedere che la differenza tra l'indice, relativo alle transazioni effettuate, calcolato a t e $t - 1$, rispetti un range delineato
RNF2	La soluzione deve prevedere che, per ogni consulente finanziario, la differenza tra l'indice, relativo alle transazioni effettuate, calcolato a t e $t - 1$ rispetti un range delineato.
RNF3	La soluzione deve prevedere che, per ogni cliente, la differenza tra l'indice, relativo alle transazioni effettuate, calcolato a t e $t - 1$ rispetti un range delineato.
RNF4	La soluzione deve prevedere che, per ogni prodotto, la differenza tra l'indice, relativo alle transazioni effettuate, calcolato a t e $t - 1$ rispetti un range delineato.
RNF5	La soluzione deve prevedere che, per ogni agente, il controvalore di portafoglio sia pari a quello calcolato con l'indice.
RNF6	La soluzione deve prevedere che, per ogni agente, il controvalore relativo alle transazioni, sia pari a quello calcolato con l'indice.
RNF7	La soluzione deve prevedere che il rapporto, tra l'indice giornaliero relativo alle transazioni e l'indice relativo al portafoglio, rispetti il range delineato.
RNF8	La soluzione deve prevedere che la differenza tra l'indice, relativo al portafoglio, calcolato a t e $t - 1$ rispetti un range delineato.
RNF9	La soluzione deve prevedere che, per ogni consulente, la percentuale dell'indice, relativo al portafoglio, rispetti un determinato range delineato.
RNF10	La soluzione deve prevedere che, per ogni consulente finanziario, la differenza tra l'indice, relativo al portafoglio, calcolato a t e $t - 1$ rispetti un range delineato.
RNF11	La soluzione deve prevedere che, per ogni cliente, la differenza tra l'indice, relativo al portafoglio, calcolato a t e $t - 1$ rispetti un range delineato.
RNF12	La soluzione deve prevedere che, per ogni prodotto, la differenza tra l'indice relativo al portafoglio, calcolato a t e $t - 1$ rispetti un range delineato.
RNF13	La soluzione deve prevedere che, per ogni prodotto, la percentuale dell'indice, relativo al portafoglio, rispetti il range delineato
RNF14	La soluzione deve prevedere che venga verificata l'assenza di movimenti duplicati nel giorno.

Tabella 3.2. Elenco dei requisiti non funzionali

CODICE	DESCRIZIONE
RNF15	La soluzione deve prevedere che la differenza tra il conteggio del numero di clienti a t e il conteggio del numero di clienti a $t - 1$ rispetti il range delineato.
RNF16	La soluzione deve prevedere che la differenza tra il conteggio del numero di consulenti a t e il conteggio del numero di consulenti a $t - 1$ rispetti il range delineato.
RNF17	La soluzione deve prevedere che la differenza tra il conteggio del numero di prodotti a t e il conteggio del numero di prodotti a $t - 1$ rispetti il range delineato.
RNF18	La soluzione deve prevedere che venga verificata l'assenza di posizioni in portafoglio duplicate nel giorno.

Tabella 3.3. Elenco dei requisiti non funzionali

Progettazione

Nel capitolo corrente verrà trattata la fase di progettazione delle funzionalità individuate durante l'analisi dei requisiti. Inizialmente, verrà presentato il flusso operativo che il dato dovrà seguire e, nella seconda parte, il modello dei dati su cui poggierà il sistema dei Quality Check.

4.1 Flusso delle operazioni

La soluzione identificata prevede la creazione di semafori su livelli diversi. I requisiti non funzionali sono riportati in singoli controlli giornalieri, che, poi, determinano l'esito del controllo finale. I due principali indici su cui si effettuano le verifiche riguardano le tabelle dei fatti relativi al portafoglio e alle transazioni.

L'output finale del sistema Quality Check, che si comporta, di fatto, come un semaforo, prevede due possibili risultati:

- *Go*: le tabelle giornaliere relative al portafoglio e alle transizioni vengono inviate al sistema Delta per la visualizzazione, e l'applicativo viene aggiornato;
- *Not Go*: l'invio al sistema Delta, delle tabelle giornaliere relative al portafoglio e alle transizioni, viene bloccato e l'applicativo non viene aggiornato.

4.1.1 Semafori indici

Per quanto riguarda i singoli requisiti non funzionali, essi dipendono dall'esito di controlli puntuali sul dato. Per i principali due indici che vengono testati, relativi ai dati di portafoglio e transazione, si considerano, rispettivamente, due semafori.

Una volta generate le due tabelle principali di portafoglio e transazione, vengono scritti i due semafori, che garantiscono il corretto processo di caricamento dei due indicatori.

L'esito dei due semafori viene scritto in una tabella, per tenerne traccia. Nel caso in cui le due tabelle dei fatti di portafoglio e transazione debbano essere rigenerate nella stessa giornata per varie motivazioni, verranno inseriti due record per i nuovi esiti dei semafori, senza sovrascrivere i precedenti, al fine di garantire il mantenimento dello storico del dato.

4.1.2 Semaforo Go/Not Go

In base all'esito dei singoli Quality Check descritti nella Sezione 4.1.1, viene calcolato il semaforo finale di Go/Not go, che porta all'invio dei dati verso Qlik.

L'esito di questo semaforo viene scritto nella stessa tabella in cui sono presenti gli esiti dei singoli semafori e gli output possibili sono soltanto Go/Not go (1 o 0). Per facilità d'uso la tabella dei semafori deve poter contenere un campo note dove aggiungere eventuali commenti del Business, in particolar modo in presenza di falsi negativi.

4.1.3 Export dei dati

Le tabelle giornaliere relative ai due indici vengono esportate automaticamente in una cartella temporanea. Esse verranno, poi, spostate nella cartella ufficiale solo se il semaforo riportato nella Sezione 4.1.2 restituisce esito positivo (Go).

In caso contrario (Not go), i file non devono essere spostati dalla cartella temporanea; in questo modo, non essendo i dati presenti nella cartella di export, l'invio è bloccato.

È previsto un sistema di controllo automatico, capace di leggere con frequenza stabilita l'ultimo record relativo all'ultimo esito del semaforo Go/Not go, scritto nell'apposita tabella. Questo controllo permette di ottenere un risultato costantemente aggiornato e in linea con eventuali modifiche o nuovi caricamenti.

4.2 Casi d'uso

Di seguito vengono mostrati alcuni scenari che possono verificarsi e il comportamento previsto dalle specifiche delineate nel Capitolo 3.

4.2.1 Esito positivo

Nel caso in cui i Quality Check restituiscano esito positivo alla prima elaborazione, i passaggi da effettuare sono i seguenti:

1. scrittura dei semafori relativi ai dati di portafoglio e transazione nell'apposita tabella;
2. export dei dati nel bucket temporaneo;
3. scrittura dell'esito positivo del semaforo Go/Not go;
4. caricamento nel bucket di esportazione e cancellazione dei file nella cartella temporanea;
5. caricamento dei file nel sistema di front-end Delta.

4.2.2 Esito negativo

Nel caso in cui il sistema restituisca esito negativo alla prima elaborazione, i passaggi da effettuare sono i seguenti:

1. scrittura dei semafori relativi ai dati di portafoglio e transazione nell'apposita tabella;
2. export dei dati nel bucket temporaneo;
3. scrittura dell'esito negativo del semaforo Go/Not go;
4. blocco del file nel bucket temporaneo e monitoraggio continuo della presenza di un nuovo semaforo Go/Not go con esito positivo Go.

Nel momento in cui il monitoraggio rilava un esito positivo, i punti 4 e 5 della Sezione 4.2.1 devono essere ripetuti.

4.2.3 Esito negativo ripetuto

Se l'errore non viene risolto durante la giornata, e l'esito del sistema di Quality Check risulta positivo il giorno successivo, il caricamento dei dati non è comunque autorizzato.

Tale scenario deve, comunque, essere gestito dal semaforo Go/Not go descritto nella Sezione 4.1.2.

Questa scelta ha il vantaggio di effettuare un'analisi profonda dell'errore, di assegnare quest'ultimo allo stakeholder di pertinenza e di attendere la risoluzione.

Di contro, vi è un possibile ritardo nella risoluzione dell'errore, la necessità di un presidio costante che monitori la situazione, e il blocco dell'aggiornamento dei dati del giorno successivo, anche in caso di esito positivo. L'ultima scelta è motivata dalla necessità di una verifica manuale dipendente dal tipo di errore verificatosi il giorno precedente.

4.3 Data Model

Per quanto riguarda i singoli controlli, i cui risultati portano poi alla scrittura del semaforo finale, viene creata un'applicazione in Qlik Sense in cui ogni foglio rappresenta un particolare check.

I dati provenienti da BigQuery vengono inseriti, tramite l'apposito connettore, in file .qvd, manipolabili da Qlik.

Per ogni requisito non funzionale, ovvero per ogni tipo di check relativo ai dati di portafoglio e transizione, viene creata una tabella con chiave univoca. Queste tabelle vengono, poi, collegate attraverso la creazione di una Common Table, composta dalla concatenazione delle varie chiavi per la fruizione della vista univoca finale.

I singoli controlli che agiscono sia sul dato di portafoglio che su quello delle transazioni, possono essere di due tipologie, ovvero bloccanti o con alert.

In Figura 4.1 è rappresentato graficamente il modello dei dati e i due tipi di controlli che possono essere attuati. Ogni check riceve il dato da utilizzare dalla Common Table e lo elabora in base al fine del check stesso.

Il controllo bloccante contribuisce in modo determinante all'esito del semaforo Go/Not go, mentre il controllo con alert genera warning.

Il primo controllo, come è visibile in Figura 4.2, produce soltanto due risultati, ovvero positivo e negativo.

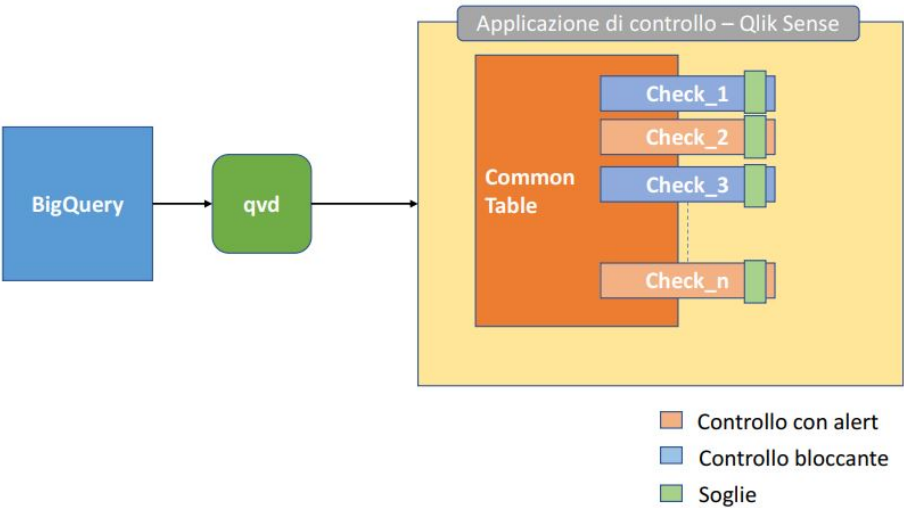


Figura 4.1. Data model sistema di Quality Check

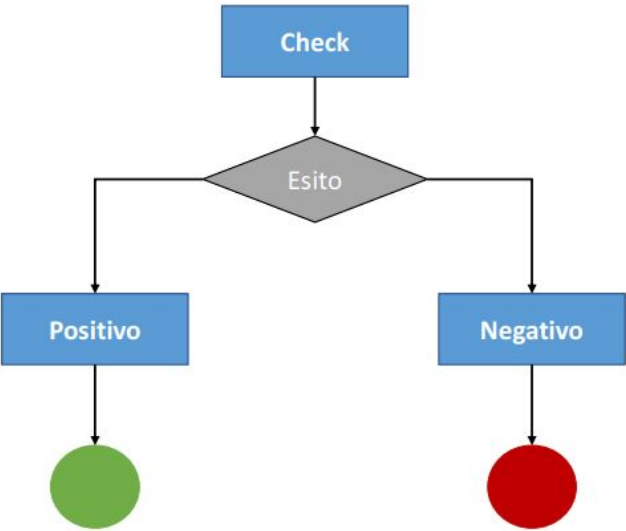


Figura 4.2. Diagramma di flusso del controllo bloccante

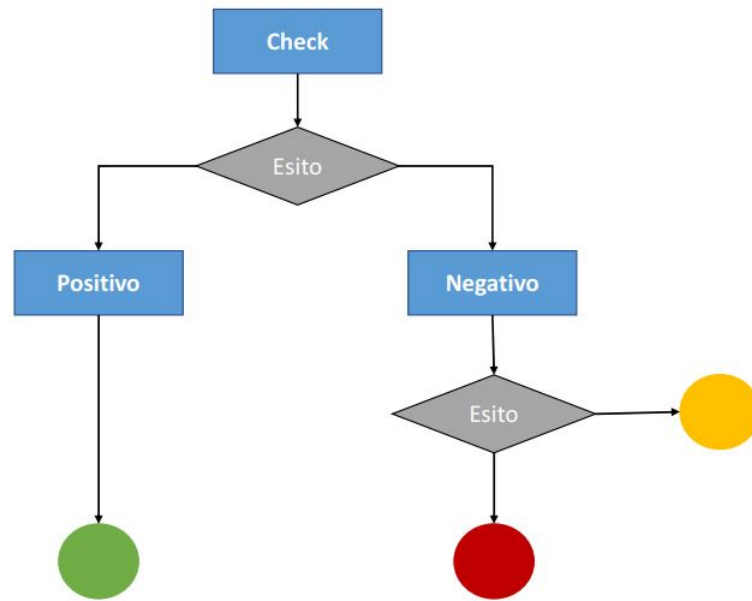


Figura 4.3. Diagramma di flusso del controllo con alert

Il controllo con alert, invece, prevede tre risultati: positivo, negativo e warning. In Figura 4.3 è visibile il diagramma di flusso.

Una volta ottenuti gli esiti di tutti i singoli check si possono generare tre scenari:

- Caso Green, in cui tutti i singoli check hanno dato esito positivo e si procede al caricamento dei dati. Corrisponde all'esito Go del semaforo Go/Not go.
- Caso Yellow, in cui sono presenti uno o più esiti warning e si procede comunque al caricamento dei dati, ma viene notificato all'utente il warning presente mediante il tool Qlik NPrinting. Anche in questo caso l'esito del semaforo Go/Not go sarà positivo.
- Caso Red, in cui sono presenti uno o più esiti negativi e non si procede al caricamento dei dati. Viene notificato agli utenti l'esito negativo dei Quality Check mediante il tool Qlik NPrinting e il semaforo finale avrà esito Not go.

Implementazione

In questo capitolo, dopo aver analizzato e compreso le procedure relative ai flussi e il modello dei dati, verranno esposte le implementazioni svolte. Saranno illustrate la costruzione dell'applicazione che attua il controllo sul dato e le procedure per la gestione dei flussi ricevuti e inviati lato BigQuery.

5.1 Applicazione Quality Check in Qlik Sense

L'implementazione della soluzione proposta al Capitolo 4 prevede la creazione di un'applicazione in Qlik Sense, costituita da:

- Una overview riassuntiva in cui viene esposto il risultato finale dei singoli semafori Go/Not go per i due indici analizzati, con relativa panoramica di aggiornamento dei dati.
- Un foglio per ogni controllo implementato.

5.1.1 Data model

Il modello dei dati da cui vengono estrapolate le informazioni per il controllo si basa su una Common Table. Lo script implementato nel data load editor di Qlik Sense permette la creazione della tabella, avente come campi:

- *ID*: codice univoco di ogni check realizzato;
- *TIPO*: identificativo del controllo specifico;
- *CATEGORIA*: indice a cui fa riferimento il controllo, relativo al portafoglio o alle transazioni;
- *ESITO*: risultato del singolo check.

Questi campi sono presenti in ciascuno dei controlli e, attraverso il data loader, vengono caricati tramite concatenazione nella tabella comune.

Le soglie di minimo e massimo e di alert, su cui vengono effettuati i check, sono prese da un file in Google Drive, che ne facilita la modifica.

Di seguito vengono analizzati singolarmente i vari controlli.

Check 1

Nel primo controllo viene valutata la differenza dell'indice relativo alle transazioni tra la giornata in esame t e quella precedente $t - 1$.

Tramite una query in linguaggio SQL vengono estratti da BigQuery:

- il valore numerico dell'indice sulle transazioni ricavato dalla tabella giornaliera t ;
- il valore numerico dell'indice sulle transazioni ricavato dalla tabella giornaliera $t - 1$;
- la differenza calcolata tra i due indici.

Nel linguaggio previsto da Qlik, successivamente, i campi che verranno “mappati” nella tabella comune vengono “valorizzati” con la logica esposta nella Tabella 5.1.

CAMPO	VALORIZZAZIONE
ID	1
CATEGORIA	TRANSAZIONE
TIPO	INDICE TRANSAZIONE Giornaliero
ESITO	0 Controllo positivo
	2 Controllo negativo

Tabella 5.1. “Valorizzazione” dei campi per il check 1

La differenza ottenuta da BigQuery viene testata con un *if* che verifica se è maggiore del valore di Go minimo e minore del valore di Go massimo. Nel caso in cui la condizione sia vera restituisce il valore 0, altrimenti, ritorna il valore 2.

Check 2

Nel secondo check si verifica per ogni consulente finanziario la differenza dell'indice relativo alle transazioni tra la giornata in esame t e la giornata precedente $t - 1$.

Tramite una query in linguaggio SQL vengono estratti da BigQuery:

- il codice identificativo del singolo consulente finanziario;
- il nominativo del singolo consulente finanziario;
- il valore numerico dell'indice sulle transazioni ricavato dalla tabella giornaliera t ;
- il valore numerico dell'indice sulle transazioni ricavato dalla tabella giornaliera $t - 1$;
- la differenza calcolata tra i due indici.

I campi così estratti vengono riportati, poi, nella Common Table, come presentato nella Tabella 5.2

L'output è ottenuto con la condizione di *if* sul valore differenza, che deve essere maggiore del valore di Go minimo e minore del valore di Go massimo.

CAMPO	VALORIZZAZIONE
ID	2
CATEGORIA	TRANSAZIONE
TIPO	INDICE TRANSAZIONE Giornaliero per consulente finanziario
ESITO	0 Controllo positivo 2 Controllo negativo

Tabella 5.2. “Valorizzazione” dei campi per il check 2

Check 3

Il terzo check verifica che il valore dell’indice relativo alle transazioni, calcolato a t e $t - 1$, per ogni cliente rispetti le soglie individuate.

Da BigQuery, tramite query in SQL, vengono estratti i seguenti campi:

- codice identificativo del singolo cliente;
- nominativo del singolo cliente;
- valore numerico dell’indice sulle transazioni ricavato dalla tabella giornaliera t ;
- valore numerico dell’indice sulle transazioni ricavato dalla tabella giornaliera $t - 1$;
- differenza calcolata tra i due indici.

Nella Common Table i campi verranno “mappati” come riportato in Tabella 5.3. L’esito è positivo nel caso il cui la differenza calcolata sia maggiore del valore di Go minimo e minore del valore di Go massimo.

CAMPO	VALORIZZAZIONE
ID	3
CATEGORIA	TRANSAZIONE
TIPO	INDICE TRANSAZIONE Giornaliero per cliente
ESITO	0 Controllo positivo 2 Controllo negativo

Tabella 5.3. “Valorizzazione” dei campi per il check 3

Check 4

Il quarto controllo testa il valore dell’indice relativo alle transazioni tra il giorno t e il giorno $t - 1$ per ogni prodotto finanziario messo sul mercato dall’azienda.

BigQuery, tramite query, fornisce a Qlik i seguenti campi:

- codice identificativo del singolo prodotto finanziario;
- nominativo del singolo prodotto finanziario;
- valore numerico dell’indice sulle transazioni ricavato dalla tabella giornaliera t ;

- valore numerico dell'indice sulle transazioni ricavato dalla tabella giornaliera $t - 1$;
- differenza calcolata tra i due indici.

Si controlla che la differenza calcolata ricada all'interno del range che va dal valore minimo di Go al valore massimo di Go. La mappatura per i campi comuni alla tabella master è visibile in Tabella 5.4.

CAMPO	VALORIZZAZIONE
ID	4
CATEGORIA	TRANSAZIONE
TIPO	INDICE TRANSAZIONE Giornaliero per prodotto
ESITO	0 Controllo positivo 2 Controllo negativo

Tabella 5.4. “Valorizzazione” dei campi per il check 4

La differenza estrapolata da BigQuery viene valutata in una condizione di *if* che restituisce esito positivo nel caso in cui il dato calcolato sia maggiore del valore minimo di Go e minore del valore massimo di Go.

Check 5

Il check 5 verifica se il controvalore di portafoglio calcolato dalla tabella su cui vengono applicate le regole di business per il calcolo dell'indice e quello della tabella base giornaliera coincidono.

I seguenti campi vengono estratti tramite query da BigQuery:

- data di osservazione del portafoglio;
- codice identificativo del singolo consulente finanziario;
- nominativo del singolo consulente finanziario;
- controvalore numerico del portafoglio calcolato dalla tabella su cui sono applicate le regole di business;
- controvalore numerico del portafoglio calcolato dalla tabella giornaliera di base.

In particolare, i dati vengono presi dalla tabella originale ed uniti in left join con la tabella calcolata. In questo modo è sufficiente vedere i valori che dal legame creato presentano controvalore base nullo.

Nel caso in cui venga riportato almeno un controvalore di portafoglio base che non viene “mappato” nel ricalcolo, ovvero nullo, l'output sarà negativo, altrimenti positivo.

Anche in questo caso i campi da “mappare” nella tabella comune sono esposti in Tabella 5.5.

CAMPO	VALORIZZAZIONE
ID	5
CATEGORIA	PORTAFOGLIO
TIPO	INDICE PORTAFOGLIO allineato con portafoglio base
ESITO	0 Controllo positivo 2 Controllo negativo

Tabella 5.5. “Valorizzazione” dei campi per il check 5

Check 6

Nel check 6 viene verificato che il controvalore delle transazioni della tabella, che implementa il calcolo dell’indice relativo alle transazioni, coincida con il controvalore delle transazioni della tabella di partenza.

I seguenti campi sono estrapolati tramite query in BigQuery:

- data di competenza;
- codice identificativo del singolo consulente finanziario;
- nominativo del singolo consulente finanziario;
- controvalore numerico delle transazioni calcolato dalla tabella su cui sono applicate le regole di business;
- controvalore numerico delle transazioni calcolato dalla tabella giornaliera di base.

La logica di funzionamento è la stessa del Check 5, con la differenza che, se il controvalore non è “mappato” correttamente, l’output è 1, altrimenti 0. Nella Tabella 5.6 vengono riportati i campi che saranno “mappati”.

CAMPO	VALORIZZAZIONE
ID	6
CATEGORIA	TRANSAZIONE
TIPO	INDICE TRANSAZIONE allineato con controvalore base
ESITO	0 Controllo positivo 1 Controllo negativo

Tabella 5.6. Valorizzazione dei campi per il check 6

Check 7

In questo caso viene verificato il rapporto tra l’indice relativo alle transazioni e quello relativo al portafoglio.

Da BigQuery vengono derivati i seguenti campi:

- data di competenza;
- valore dell’indice relativo alle transazioni;

- valore dell'indice relativo al portafoglio;
- rapporto dei due espresso in percentuale.

Se la percentuale ottenuta è maggiore del valore minimo di Go o minore del valore massimo di Go allora l'esito del controllo sarà positivo.

In Tabella 5.7 viene riportata la “mappatura” dei campi.

CAMPO	VALORIZZAZIONE
ID	7
CATEGORIA	TRANSAZIONE
TIPO	INDICI a confronto
ESITO	0 Controllo positivo 1 Controllo negativo

Tabella 5.7. “Valorizzazione” dei campi per il check 7

Check 8

Nell'ottavo check viene valutata la differenza dell'indice relativo al portafoglio tra la giornata in esame t e quella precedente $t - 1$.

Grazie ad una query in linguaggio SQL vengono estratti da BigQuery:

- il valore numerico dell'indice sul portafoglio ricavato dalla tabella giornaliera t ;
- il valore numerico dell'indice sul portafoglio ricavato dalla tabella giornaliera $t - 1$;
- la differenza calcolata tra i due indici.

Successivamente, i campi che saranno “mappati” nella tabella comune vengono “valorizzati” con la logica esposta nella Tabella 5.8.

CAMPO	VALORIZZAZIONE
ID	8
CATEGORIA	PORTAFOGLIO
TIPO	INDICE PORTAFOGLIO Giornaliero
ESITO	0 Controllo positivo 2 Controllo negativo

Tabella 5.8. Valorizzazione dei campi per il check 8

La differenza ottenuta da BigQuery viene testata con un *if* che verifica se il calcolo è maggiore del valore di Go minimo e minore del valore di Go massimo. Nel caso in cui la condizione sia vera restituisce il valore 0, altrimenti, ritorna il valore 2.

Check 9

Nel check 9 viene calcolata la percentuale dell'indice relativo al portafoglio, per ogni consulente finanziario, a t e $t - 1$.

In particolare, da BigQuery, vengono estrapolati:

- il codice identificativo del singolo consulente finanziario;
- il nominativo del singolo consulente finanziario;
- la percentuale massima dell'indice sul portafoglio calcolata a t ;
- la percentuale minima dell'indice sul portafoglio calcolata a $t - 1$;
- la differenza tra le due percentuali.

Si verifica che la differenza percentuale rispetti il range delineato dai valori di Go minimo e massimo. Nella Tabella 5.9 viene riprotata la mappatura rispetto ai campi della Common Table.

CAMPO	VALORIZZAZIONE
ID	9
CATEGORIA	PORTAFOGLIO
TIPO	%INDICE PORTAFOGLIO per consulente finanziario
ESITO	0 Controllo positivo 1 Controllo negativo

Tabella 5.9. “Valorizzazione” dei campi per il check 9

Check 10

Il decimo check testa, per ogni consulente finanziario, la differenza dell'indice relativo al portafoglio tra la giornata in esame t e la giornata precedente $t - 1$.

Da BigQuery vengono estratti i seguenti campi:

- codice identificativo del singolo consulente finanziario;
- nominativo del singolo consulente finanziario;
- valore numerico dell'indice sul portafoglio ricavato dalla tabella giornaliera t ;
- valore numerico dell'indice sul portafoglio ricavato dalla tabella giornaliera $t - 1$;
- differenza calcolata tra i due indici.

I campi così estratti vengono riportati, poi, nella Common Table come mostrato nella Tabella 5.10.

L'output è ottenuto con la condizione di *if* sul valore differenza, che deve essere maggiore del valore di Go minimo e minore del valore di Go massimo.

CAMPO	VALORIZZAZIONE
ID	10
CATEGORIA	PORTAFOGLIO
TIPO	INDICE PORTAFOGLIO Giornaliero per consulente finanziario
ESITO	0 Controllo positivo 2 Controllo negativo

Tabella 5.10. “Valorizzazione” dei campi per il check 10

Check 11

L’undicesimo check verifica che il valore dell’indice, calcolato a t e $t - 1$, per ogni cliente rispetti le soglie individuate.

Da BigQuery, tramite query in SQL, vengono estratti i seguenti campi:

- codice identificativo del singolo cliente;
- nominativo del singolo cliente;
- valore numerico dell’indice sul portafoglio derivato dalla tabella giornaliera t ;
- valore numerico dell’indice sul portafoglio derivato dalla tabella giornaliera $t - 1$;
- differenza calcolata tra i due indici.

Nella Common Table i campi verranno “mappati” come riportato in Tabella 5.11. L’esito è positivo nel caso il cui la differenza calcolata sia maggiore del valore di Go minimo e minore del valore di Go massimo.

CAMPO	VALORIZZAZIONE
ID	11
CATEGORIA	PORTAFOGLIO
TIPO	INDICE PORTAFOGLIO Giornaliero per cliente
ESITO	0 Controllo positivo 2 Controllo negativo

Tabella 5.11. “Valorizzazione” dei campi per il check 11

Check 12

Il dodicesimo check testa il valore dell’indice relativo alle transazioni tra il giorno t e il giorno $t - 1$ per ogni prodotto finanziario immesso sul mercato dall’azienda.

BigQuery, tramite query fornisce a Qlik i seguenti campi:

- codice identificativo del singolo prodotto finanziario;
- nominativo del singolo prodotto finanziario;
- valore numerico dell’indice sul portafoglio ricavato dalla tabella giornaliera t ;
- valore numerico dell’indice sul portafoglio ricavato dalla tabella giornaliera $t - 1$;

- differenza calcolata tra i due indici.

Si controlla che la differenza calcolata sia all'interno del range che va dal valore minimo di Go al valore massimo di Go. La “mappatura” per i campi comuni alla tabella master è visibile in Tabella 5.12.

CAMPO	VALORIZZAZIONE
ID	12
CATEGORIA	PORTAFOGLIO
TIPO	INDICE PORTAFOGLIO Giornaliero per prodotto
ESITO	0 Controllo positivo 2 Controllo negativo

Tabella 5.12. “Valorizzazione” dei campi per il check 12

La differenza estrapolata da BigQuery viene valutata in una condizione di *if* che restituisce esito positivo nel caso in cui il dato calcolato sia maggiore del valore minimo di Go e minore del valore di massimo di Go.

Check 13

Nel check 13 viene calcolata la percentuale dell'indice relativo al portafoglio, per ogni prodotto finanziario, sulle transazioni effettuate.

In particolare, da BigQuery, vengono estrapolati:

- il codice identificativo del singolo prodotto finanziario;
- il nominativo del singolo prodotto finanziario;
- la percentuale massima dell'indice sul portafoglio calcolato sulle transazioni effettuate;
- la percentuale massima dell'indice sul portafoglio calcolata sulle transazioni effettuate;
- la percentuale minima dell'indice sul portafoglio calcolata sulle transazioni effettuate;
- la causale della transazione.

Si verifica che la percentuale massima e la minima rispettino i valori di Go massimi e minimi. Nella Tabella 5.13 viene riportata la mappatura rispetto ai campi della Common Table.

Check 14

In questo check viene verificata l'assenza di duplicati nei dati relativi alle transazioni.

In BigQuery viene calcolato il numero di codici di transazioni presenti nella relativa tabella e vengono estratti i duplicati.

Se il conteggio è nullo allora il risultato è positivo.

La “valorizzazione” dei campi della Common Table viene presentata in Tabella 5.14.

CAMPO	VALORIZZAZIONE
ID	13
CATEGORIA	PORTAFOGLIO
TIPO	% INDICE PORTAFOGLIO sulle transazioni
ESITO	0 Controllo positivo 2 Controllo negativo

Tabella 5.13. “Valorizzazione” dei campi per il check 13

CAMPO	VALORIZZAZIONE
ID	14
CATEGORIA	TRANSAZIONE
TIPO	INDICE TRANSAZIONI duplicate
ESITO	0 Controllo positivo 2 Controllo negativo

Tabella 5.14. “Valorizzazione” dei campi per il check 14

Check 15

Il quindicesimo check verifica il numero di clienti a t e $t - 1$.

Da BigQuery vengono estratti:

- il numero di clienti della fotografia di portafoglio alla data t ;
- il numero di clienti della fotografia di portafoglio alla data $t - 1$;
- la data a t ;
- la data a $t - 1$;
- la differenza percentuale tra i due conteggi.

In questo caso si verifica l'appartenenza della differenza percentuale calcolata al range delimitato dal valore di Go minimo e dal valore di Go massimo. Se tale differenza non rispetta la condizione stabilita si valuta l'appartenenza ad un altro range compreso tra il valore di alert minimo e il valore di alert massimo.

In Tabella 5.15 viene mostrata la “valorizzazione” dei campi che saranno mappati nella Common Table.

CAMPO	VALORIZZAZIONE
ID	15
CATEGORIA	PORTAFOGLIO
TIPO	NUMERO clienti
ESITO	0 Controllo positivo 1 Controllo negativo

Tabella 5.15. “Valorizzazione” dei campi per il check 15

Check 16

Il check 16 verifica il numero di consulenti finanziari a t e $t - 1$.

Da BigQuery vengono estratti:

- il numero di consulenti finanziari della fotografia di portafoglio alla data t ;
- il numero di consulenti finanziari della fotografia di portafoglio alla data $t - 1$;
- la data a t ;
- la data a $t - 1$;
- la differenza percentuale tra i due conteggi.

Nel controllo viene valutata l'appartenenza della differenza percentuale calcolata al range delimitato dal valore di Go minimo e dal valore di Go massimo. Se la condizione non è vera viene verificato se la differenza rispetta un altro range delineato dai valori di alert minimo e massimo.

In Tabella 5.16 viene illustrata la “valorizzazione” dei campi che andranno “mappati” nella Common Table.

CAMPO	VALORIZZAZIONE
ID	16
CATEGORIA	PORTAFOGLIO
TIPO	NUMERO consulenti finanziari
ESITO	0 Controllo positivo
	1 Controllo negativo

Tabella 5.16. “Valorizzazione” dei campi per il check 16

Check 17

Il check 17 verifica il numero di prodotti finanziari a t e $t - 1$.

Tramite BigQuery vengono estrapolati:

- il numero di prodotti finanziari della fotografia di portafoglio alla data t ;
- il numero di prodotti finanziari della fotografia di portafoglio alla data $t - 1$;
- la data a t ;
- la data a $t - 1$;
- la differenza percentuale tra i due conteggi.

La condizione da valutare è l'appartenenza della differenza percentuale calcolata al range delimitato dal valore di Go minimo e dal valore di Go massimo. Se la condizione non è verificata, viene testato se la differenza rispetta un range delineato dai valori di alert minimo e massimo.

In Tabella 5.17 viene illustrata la “valorizzazione” dei campi che andranno “mappati” nella Common Table.

CAMPO	VALORIZZAZIONE
ID	17
CATEGORIA	PORTAFOGLIO
TIPO	NUMERO prodotti finanziari
ESITO	0 Controllo positivo 1 Controllo negativo

Tabella 5.17. “Valorizzazione” dei campi per il check 17

Check 18

Nell’ultimo check viene verificata l’assenza di duplicati nelle posizioni di portafoglio.

In BigQuery viene calcolato il numero di codici relativi alle posizioni di portafoglio presenti nella tabella giornaliera e vengono estratti quelli che sono duplicati, ovvero il cui conteggio è maggiore di uno.

Se il conteggio è nullo allora il risultato è positivo.

La “valorizzazione” dei campi della Common Table è presentata in Tabella 5.18.

CAMPO	VALORIZZAZIONE
ID	18
CATEGORIA	PORTAFOGLIO
TIPO	INDICE PORTAFOGLIO duplicati
ESITO	0 Controllo positivo 2 Controllo negativo

Tabella 5.18. “Valorizzazione” dei campi per il check 18

Graficamente, quindi, il modello dei dati è quello presentato in Figura 5.1.

La Common Table considera come output per ogni controllo il valore massimo dei singoli esiti.

Infine, come output totale del sistema di Quality Check, si ottengono i corrispondenti valori.

5.2 Flussi in BigQuery

Una volta ottenuto l’esito dell’intero sistema, Go o Not Go, tramite il connettore verso BigQuery, viene scritto un record nell’apposita tabella *TRAFFICLIGHTS*.

Il sistema che si occupa dell’invio dei dati verso Delta legge il campo “*ESITO*” della tabella, “valorizzato” a 0 in caso di esito negativo del semaforo e 1 altrimenti.

Per tenere conto dell’esito del sistema di Quality Check del giorno precedente, viene letto il relativo record dalla tabella.

In Figura 5.2, viene riportato un esempio di esito positivo e negativo mostrato nella tabella *TRAFFICLIGHTS*.

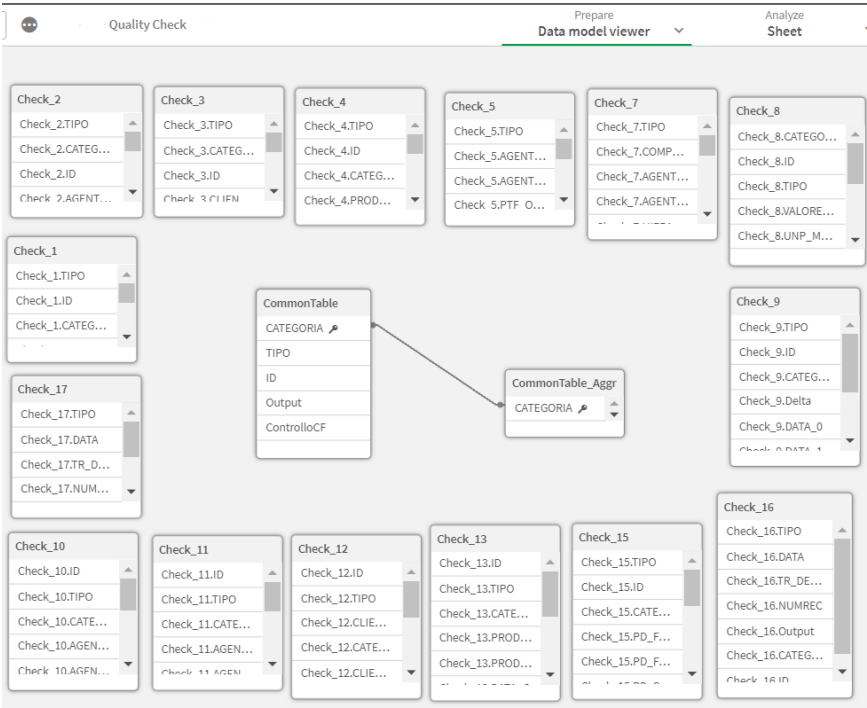


Figura 5.1. Modello dei dati relativo ai Quality Check in Qlik Sense

Riga	ID	DATA_LOAD	VALORE	DATA_TIMESTAMP	EVENTO	OWNER	ANNOTATION
101	QC_ALL	2021-09-29	1	2021-09-29 09:31:25 UTC	1	BI_QLIK_QUALITYCHECK	qlik sense
102	QC_ALL	2021-09-27	1	2021-09-27 06:11:42 UTC	1	BI_QLIK_QUALITYCHECK	qlik sense
103	QC_ALL	2021-09-28	1	2021-09-28 06:09:43 UTC	1	BI_QLIK_QUALITYCHECK	qlik sense
104	QC_ALL	2021-09-30	1	2021-09-30 07:11:40 UTC	1	BI_QLIK_QUALITYCHECK	qlik sense
105	QC_ALL	2021-10-01	1	2021-10-01 07:12:09 UTC	1	BI_QLIK_QUALITYCHECK	qlik sense
106	ALTRO	2021-10-02	1	2021-10-02 05:51:55 UTC	0	BIGLOADER	null

Figura 5.2. Esempio della tabella *TRAFFICLIGHTS* in BigQuery

Principali dashboard realizzate

In questo capitolo verranno mostrate le principali dashboard dell'applicazione Quality Check realizzate in Qlik Sense. Le visualizzazioni legate all'applicazione, avendo come scopo principale il controllo dei dati, hanno strutture basilari. Inizialmente verrà analizzata la schermata principale; dopo di ciò si passerà ad alcuni dei controlli più significativi. Negli esempi di controlli riportati sono stati oscurati alcuni dati per il rispetto delle policy di privacy e riservatezza.

6.1 Overview

Nel foglio introduttivo di overview viene presentato il quadro generale con l'esito complessivo dei due semafori legati agli indici relativi alle transazioni e al portafoglio.

In Figura 6.1 viene mostrata la dashboard relativa.

In particolare, nella parte superiore vengono visualizzate della figura, sotto forma di semplici caselle di testo, la panoramica dell'aggiornamento dei dati e gli esiti finali dei semafori.

La prima informazione è utile a capire se i dati esposti dall'applicazione sono correttamente aggiornati, perchè vi potrebbero essere blocchi sul dato di partenza che non permettono all'applicazione di girare.

La seconda informazione risulta essere di fondamentale importanza poichè, mostrando l'esito finale dei controlli attuati, permette di comprendere la qualità dei dati in esame.

Nella parte inferiore del foglio *Overview* è riportata la Common Table, con gli esiti dei singoli check che, in questo caso, vengono esposti con bullet verdi o rossi, a seconda del risultato. Nella Figura 6.1 viene riportato un esempio, in cui i risultati dei singoli check sono tutti positivi e i semafori mostrano come esito Go.

6.2 Check 1

In Figura 6.2 è riportata la visualizzazione del primo controllo. Nella tabella vengono visualizzati i valori principali su cui si basa il check, ovvero l'indice relativo alle

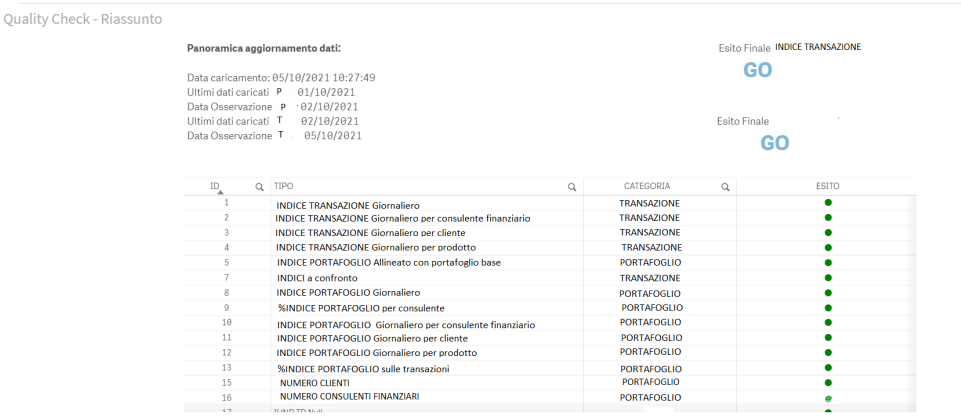


Figura 6.1. Overview dell'applicazione Quality Check

transazione a t e a $t - 1$ e la relativa differenza. L'esito è espresso tramite un bullet, in questo caso di colore verde ad indicare la positività del test.

A lato della tabella è visibile un grafico gauge, in cui il valore dell'ago viene letto e confrontato con il campo dati colorato. In questo caso la parte in verde, rappresenta il range ammesso, mentre quella in rosso denota il range non consentito, che porterà alla generazione dell'esito negativo.



Figura 6.2. Check 1 dell'applicazione Quality Check

Nel check 8 vengono presentati i medesimi dati ma relativi all'indice sul portafoglio.

6.3 Check 2

Nel secondo check, oltre alla tabella contenente il dettaglio dell'indice relativo alle transazioni per consulente finanziario, sono stati inseriti anche due indicatori chiave

di prestazione (KPI).
Il KPI *Check positivi* riporta il numero di consulenti finanziari che hanno superato il test e la relativa percentuale.
Il KPI *Check negativi* segnala il numero di consulenti finanziari per cui l'esito del controllo è risultato negativo, e la relativa percentuale.

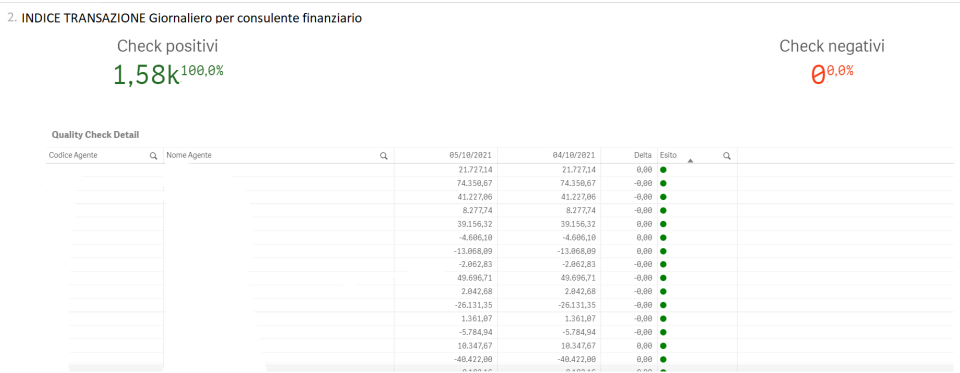


Figura 6.3. Check 2 dell'applicazione Quality Check

Il check 10 mostra le stesse informazioni del secondo controllo, ma relative all'indice sul portafoglio.

6.4 Check 3

Il terzo foglio, che riporta il controllo attuato sull'indice relativo alle transazioni per cliente, ha la stessa struttura del Check 2 presentato nella Sezione 6.3.
In questo caso i KPI sono relativi rispettivamente al numero dei clienti per cui l'esito del test è positivo e al numero di clienti per cui l'esito del test risulta negativo.
Questa struttura è replicata per il check 11 con i dati relativi al portafoglio.

6.5 Check 4

Per il quarto check i dati visualizzati sono visibili in Figura 6.5.
In tale figura viene visualizzata la tabella con le informazioni relative all'indice sulle transazioni prese al giorno t e $t - 1$ e viene valutato il valore della differenza.
Anche in questo caso l'esito positivo è rappresentato da un bullet verde.
In alto sono presenti i due KPI relativi al numero di prodotti con esito positivo e al numero di prodotti con esito negativo.

3. INDICE TRANSAZIONE Giornaliero per cliente

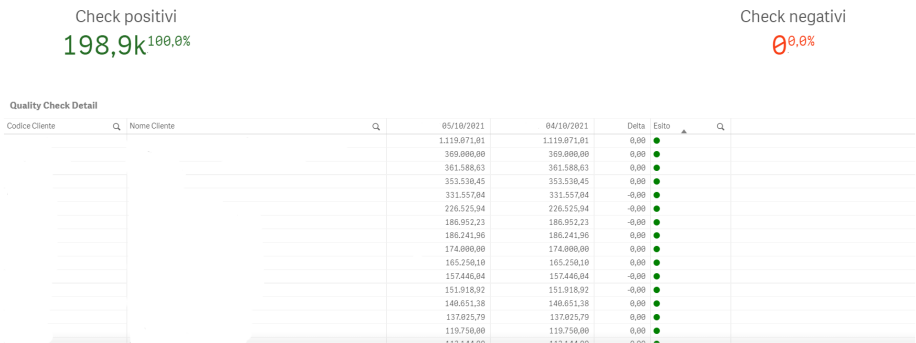


Figura 6.4. Check 3 dell'applicazione Quality Check

4. INDICE TRANSAZIONE Giornaliero per prodotto

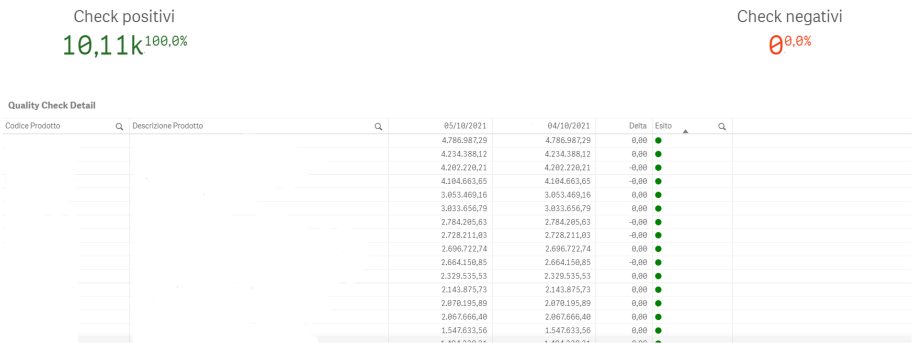


Figura 6.5. Check 4 dell'applicazione Quality Check

6.6 Check 9

Il nono check è volto a verificare la percentuale dell'indice relativo al portafoglio, per ogni consulente finanziario.

In questo caso, come è riportato in Figura 6.6, si può ottenere come esito alert. La tabella mostra che il primo record non rispetta il range delineato e, per questo, presenta un bullet giallo.

I check che restituiscono alert, in questo preciso scenario, sono due e rappresentano lo 0.1% del totale.

6.7 Check 12

Il dodicesimo check prevede la verifica dell'indice relativo al portafoglio per ogni prodotto finanziario.

Come si vede in Figura 6.7, la struttura è la medesima dei controlli visti in precedenza, ma presenta tre KPI; oltre ai due sull'esito positivo e negativo viene, infatti, mostrato il numero di prodotti che hanno generato alert e la relativa percentuale.

9. %INDICE PORTAFOGLIO per consulente

Check positivi
1,57k^{99,9%}

Check alert
2^{0,1%}

Quality Check Detail						
Codice Agente	Q	Nome Agente	Q	01/10/2021	30/09/2021	Delta
				0,005100%	0,005300%	15,09%
				0,007500%	0,007000%	7,14%
				0,002000%	0,001900%	5,25%
				0,003300%	0,003200%	3,13%
				0,003500%	0,003400%	2,94%
				0,008000%	0,007800%	2,50%
				0,004100%	0,004000%	2,50%
				0,008600%	0,008400%	2,38%
				0,004300%	0,004200%	2,38%
				0,004500%	0,004400%	2,27%
				0,004800%	0,004700%	2,13%
				0,005200%	0,005100%	1,95%
				0,010700%	0,010500%	1,90%
				0,005400%	0,005300%	1,89%
				0,010900%	0,010700%	1,87%

Figura 6.6. Check 9 dell'applicazione Quality Check

12. INDICE PORTAFOGLIO Giornaliero per prodotto

Check positivi
9,7k^{100,0%}

Check alert
0^{0,0%}

Check negativi
0^{0,0%}

Quality Check Detail						
Codice Prodotto	Q	Descrizione Prodotto	Q	01/10/2021	30/09/2021	Delta
				113.265,11	114.014,92	-749,81
				108.851,22	109.496,80	-645,58
				92.493,75	92.383,06	110,70
				80.633,36	80.693,19	-59,83
				69.558,04	69.272,54	277,50
				63.663,21	63.369,10	294,10
				61.340,31	61.284,38	55,93
				40.750,62	40.737,08	13,53
				38.760,55	38.680,01	80,54
				37.553,12	37.620,50	-67,38
				36.321,16	36.267,11	54,05
				35.999,88	35.841,78	158,10
				33.424,82	33.418,48	6,34
				33.261,07	33.258,48	2,59
				31.973,61	31.982,34	-8,73

Figura 6.7. Check 12 dell'applicazione Quality Check

6.8 Check 15

Per la rappresentazione del numero di clienti presenti a t e $t - 1$ è stato utilizzato un grafico gauge, visibile in Figura 6.8. In verde è riportato il range di accettazione. I Check 16 e 17 presentano la stessa visualizzazione.

15. NUMERO CLIENTI

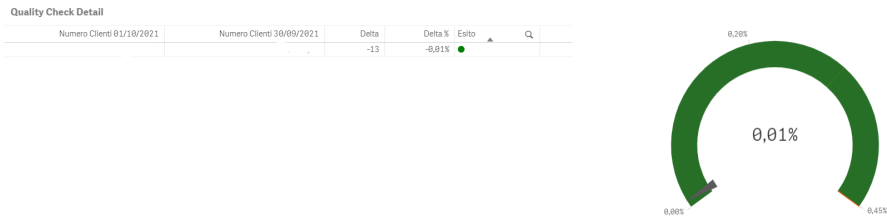


Figura 6.8. Check 15 dell'applicazione Quality Check

Discussione in merito al lavoro svolto

Nel capitolo corrente verrà proposto un riepilogo di quanto svolto nel periodo di tirocinio, per poi passare a delle considerazioni di carattere personale.

7.1 Considerazioni riguardo l'esperienza di tirocinio

Il tirocinio curriculare che ho svolto presso l'azienda di consulenza informatica ha avuto una durata complessiva di sei mesi. Durante questo periodo i task che mi sono stati assegnati hanno toccato svariati ambiti. Fin dal primo giorno ho avuto modo di confrontarmi con la realtà progettuale e di team.

Inizialmente, mi sono occupata dell'handover di un componente del team con grande esperienza, che giornalmente trattava un aspetto del progetto. Questo periodo è stato molto utile per la comprensione della struttura, dell'organizzazione e delle nozioni tecniche relative alle varie attività da svolgere.

Nel mentre, ho iniziato ad occuparmi delle attività di application maintenance, seguita passo passo dai miei colleghi. Queste attività, svolte in BigQuery, riguardavano principalmente l'estrazione di particolari set di dati tramite query in SQL.

In seguito mi è stata assegnata l'attività discussa nel corso della trattazione. Dalla fase di analisi dei requisiti, si è passati alla fase di progettazione e, infine, alla fase d'implementazione vera e propria. Sebbene tali attività siano state esposte in questa sede come ben distinte, in realtà sono state portate avanti con un certo grado di parallelismo, dovuto, soprattutto, al continuo aggiornamento dei requisiti e agli adattamenti dettati dal business.

Tuttavia, è stata la fase di progettazione e modellazione dei dati a richiedere maggior effort, viste le diverse fonti di dati considerate.

Successivamente ho svolto altri task in Qlik, relativi a diverse applicazioni, con gradi di difficoltà sempre diversi.

Nell'ultimo periodo ho iniziato a lavorare ad un serie di sviluppi inerenti ad un importante stream progettuale.

7.2 Considerazioni personali

L'argomento di tirocinio e di tesi si è rivelato molto interessante, in quanto parte di una serie di sviluppi del tutto nuovi all'interno dell'applicativo. Inoltre, questa prima esperienza in un ambiente lavorativo — che ad oggi ancora prosegue — è stata per me estremamente valorizzante: ho, *in primis*, messo in gioco le mie potenzialità e avuto l'occasione di accrescere le mie conoscenze tecniche nell'ambito dei dati.

Questo periodo mi ha consentito, inoltre, di crescere personalmente in un contesto sfidante, creativo e dinamico.

I sei mesi di tirocinio sono stati svolti totalmente in modalità smart working; questo non ha influito negativamente sull'esperienza vissuta, se non in parte per l'ambito umano.

Nonostante ciò, ho avuto modo di rapportarmi con colleghi altamente preparati e da loro ho avuto modo di apprendere nuove competenze e conoscenze che vanno oltre lo specifico aspetto progettuale.

Lavorare in team, inizialmente, si è rivelato complesso per via della distanza; tuttavia, superata già la prima settimana, si è instaurato un rapporto di complicità con ogni membro che ha apportato un valore aggiunto all'esperienza e che mi ha spinto a continuare il percorso.

Conclusioni e uno sguardo al futuro

In questo ultimo capitolo saranno tratte le conclusioni riguardanti il lavoro svolto e si tratteranno gli sviluppi futuri.

8.1 Conclusioni

In questa tesi è stato illustrato lo sviluppo di un'applicazione in Qlik Sense volta al controllo della qualità dei dati esposti in un applicativo di front-end. Tale dati sono relativi ad un progetto in ambito finanziario, gestito dall'azienda di consulenza informatica in cui è stato svolto il tirocinio.

Nell'introduzione sono state trattate le nozioni di base per la comprensione dei temi toccati nel corso della tesi, con particolare attenzione ai concetti di Business Intelligence e Data Warehouse.

Nella prima parte, poi, è stata esaminata la struttura del sistema di Business Intelligence dell'azienda finanziaria da cui provengono i dati utilizzati nell'applicazione. In particolare, si è posta l'attenzione sull'alimentazione del Data Warehouse utilizzato, riportando le entità principali dei dataset presenti.

Successivamente si è passati all'analisi dell'argomento centrale del lavoro di tesi, iniziando dalla stesura dei requisiti, effettuata secondo la documentazione fornita dal cliente, per poi proseguire con la progettazione dell'applicazione vera e propria e dei flussi interessati dal processo di data quality richiesto. È stata, quindi, descritta l'implementazione di quanto progettato mediante il tool Qlik Sense.

Infine, dopo aver illustrato i meccanismi di funzionamento, gli scenari, l'invio e la ricezione dei flussi verso il sistema di front-end, sono state esposte le principali dashboard realizzate per l'applicazione in esame.

8.2 Sviluppi futuri

Il lavoro esaminato in questa sede rappresenta una piccola parte di un processo ben più ampio che ha come obiettivo il mantenimento e l'innovazione del sistema di Business Intelligence dell'azienda. Si tratta di un progetto ambizioso e sfidante, sia per

la mole di dati coinvolti sia per i molteplici aspetti da monitorare e implementare, che riguardano sia la parte di back-end sia quella di front-end.

A tendere verranno introdotti, in particolari ambiti, algoritmi di Intelligenza Artificiale.

Riferimenti bibliografici

1. BigQuery. <https://cloud.google.com/>, 2021.
2. Bismart, Datawarehouse. <https://blog.bismart.com/en/what-do-we-do-etl>, 2021.
3. Database management system. https://it.wikipedia.org/wiki/Database_management_system, 2021.
4. Digital4, Data-Driven. <https://www.digital4.biz/marketing/big-data-e-analytics/sei-regole-d-oro-per-un-data-driven-marketing-di-successo/>, 2021.
5. Elena Baralis, Big data. <https://dbdmg.polito.it/wordpress/wp-content/uploads/2018/10/1-DSIntro.pdf>, 2021.
6. Elena Baralis, Progettazione di Datawarehouse. https://dbdmg.polito.it/wordpress/wpcontent/uploads/2020/05/2-DWprog_x6pdf.pdf, 2021.
7. Extrasys, Che cos'è la Business Intelligence. <https://www.extrasys.it/it/redblog/che-cosa-%C3%A8-business-intelligence>, 2021.
8. Gartner. <https://www.gartner.com/en>, 2021.
9. Microsoft. <https://www.microsoft.com/it-it>, 2021.
10. Qlik. <https://www.qlik.com/it-it>, 2021.
11. Storia della Business Intelligence. <https://www.inside.agency/storia-business-intelligence/>, 2021.
12. Storia della Business Intelligence. <https://www.slideshare.net/lmauri/business-intelligence-v03>, 2021.
13. Tableau, Business Intelligence. <https://www.tableau.com/it-it>, 2021.
14. Tutto sulla Business Intelligence. <https://www.datamaze.it/risorse/tutto-sulla-business-intelligence>, 2021.
15. Wikipedia, Business Intelligence. https://it.wikipedia.org/wiki/Business_intelligence, 2021.
16. Manuel Filipe Santos João Lopes, Tiago Guimarães. *Adaptive Business Intelligence: A New Architectural Approach*. Procedia Computer Science, Volume 177, 2020.
17. Liu Yu-His Liang, Ting-Peng. *Research Landscape of Business Intelligence and Big Data analytics: A bibliometrics study*. Expert Systems with Applications, Volume 111, 2018.
18. Stefano Rizzi M. Golfarelli. *Data Warehouse: teoria e pratica della progettazione*. McGrawHill, seconda edizione, 2006.
19. Arnaldo Coelho Nuno Caseiro. *The influence of Business Intelligence capacity, network learning and innovativeness on startups performance*. Journal of Innovation and Knowledge, 2019.

20. João Guerreiro Iis Tussyadiah Sandra Maria Correia Loureiro. *Artificial intelligence in business: State of the art and future research agenda*. Journal of Business Research, 2020.
21. Zhijun Lin Yansheng Chen. *Business Intelligence Capabilities and Firm Performance: A Study in China*. International Journal of Information Management, Volume 57, 2021.

Ringraziamenti

Il primo ringraziamento va alla mia famiglia, che mi ha dato la possibilità di intraprendere questo percorso e che è stata fonte di supporto costante e valorizzante confronto; a loro dedico il traguardo raggiunto.

Vorrei ringraziare, inoltre, il mio relatore, il Professore Domenico Ursino, per la straordinaria disponibilità e umanità mostrata durante la stesura di questa tesi e per l'interesse che ha fatto scaturire in me con le Sue materie; questo interesse è diventato oggi passione e lavoro.

Ringrazio di cuore i miei colleghi: Teresa, Massimiliano, Fabiano, Roberta ed Eugenio, per l'attenzione, la cura e l'impegno che hanno rivolto nei miei confronti e che tutt'ora continua.

Ringrazio i miei coinquilini, senza i quali non avrei ricordi così preziosi degli anni trascorsi ad Ancona, divenuta ormai una seconda casa.

Infine, vorrei ringraziare tutti i miei colleghi universitari e i miei amici di sempre, per essere stati al mio fianco e aver condiviso con me ogni momento importante.