

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 KPI basato su Qlik e Google BigQuery a supporto delle attività di una società di gestione del risparmio

Design and implementation of a KPI system based on Qlik and Google BigQuery to support the activities of an asset management company

Relatore

Prof. Domenico Ursino

Candidato

Antonio Monaco

ANNO ACCADEMICO 2021-2022

Sommario

Negli ultimi anni la Business Intelligence ha assunto un'importanza sempre maggiore, in quanto, analizzando i dati dei clienti, è possibile prendere decisioni di impatto strategico e tattico. Tali decisioni apportano un valore aggiunto all'azienda che ne fa uso. In questa tesi sono stati analizzati i dati relativi ai clienti e ai promotori finanziari facenti parte di una società di gestione del risparmio. In particolare, sono stati realizzati un sistema di KPI in supporto delle attività aziendali. Questi indici sono stati realizzati mediante l'utilizzo di tool di Business Intelligence leader del mercato, ovvero, Qlik e Google BigQuery.

Keyword: Business Intelligence, Data Analytics, Data Visualization, Qlik View, Qlik Sense, KPI, Google BigQuery, Gestione del Risparmio, Finance

Introduzione	1
1 Introduzione alla Business Intelligence	3
1.1 Il valore dei dati	3
1.1.1 Le 5 V dei Big Data	4
1.1.2 I benefici dei Big Data	5
1.1.3 Come analizzare i Big Data	6
1.1.4 La nascita della Business Intelligence	6
1.2 I protagonisti della BI	8
1.2.1 Le figure coinvolte nella BI	8
1.2.2 I sistemi utilizzati nella BI	9
1.2.2.1 Piramide DIKW	9
1.2.2.2 I Database	11
1.2.2.3 I DBMS	11
1.2.2.4 I Data Warehouse	12
1.3 Le Fasi della BI	13
1.4 Stato dell'arte della BI con Gartner	16
1.4.1 Come si utilizza un MQ di Gartner?	16
1.4.2 Come si legge un MQ di Gartner?	16
1.4.3 I Magic Quadrant di BI e Cloud DBMS	17
1.5 Costi e benefici della BI	17
1.5.1 Quanto mi costa la BI?	18
1.5.2 Quanto mi costa non avere la BI?	18
1.5.3 Benefici diretti della BI	19
2 Introduzione a Qlik e a BigQuery	20
2.1 Google BigQuery	20
2.1.1 Funzionalità e vantaggi di BigQuery	20
2.2 Qlik	21
2.2.1 Qlik Sense	22
2.2.2 Qlik View	26
2.2.3 Qlik NPrinting	26

3	Descrizione del contesto di riferimento	28
3.1	Introduzione agli obiettivi di progetto	28
3.2	L'Architettura della BI di progetto	28
3.3	Sorgenti di alimentazione del progetto	29
4	Analisi dei requisiti	34
4.1	Proposta e obiettivi finali	34
4.1.1	Architettura di riferimento	35
4.1.2	Approccio adattato per lo stream BI	36
4.1.3	Fasi di progetto	36
4.2	Requisiti funzionali	38
4.2.1	Dati propedeutici al calcolo	38
5	Progettazione	40
5.1	Procedura di ripristino dati	40
5.2	Regole di calcolo degli indici Alfa e Beta	42
5.3	Architettura degli indici Alfa e Beta	44
5.3.1	Architettura indice Alfa Giornaliero	44
5.3.2	Architettura indice Alfa MTD	44
5.3.3	Architettura indice Beta Giornaliero	45
5.3.4	Architettura indice Beta36	45
6	Implementazione	47
6.1	Controllo di qualità dei dati	47
6.2	Calcolo Degli indici Alfa e Beta	48
7	App realizzate e Mashup	52
7.1	Dashboard relativa all'indice Alfa	52
7.2	Dashboard relativa all'indice Beta	54
7.3	Mashup	56
8	Discussione in merito al lavoro svolto	58
8.1	Punti di forza del progetto	58
8.2	Punti di debolezza	59
8.3	Lezioni apprese	59
9	Conclusioni	61
	Bibliografia	63
	Ringraziamenti	65

Elenco delle figure

1.1	Le caratteristiche dei Big Data	4
1.2	Big Data: le quattro principali tipologie di analisi	6
1.3	Hans Peter Luhn: il padre della Business Intelligence	7
1.4	Paradigma organizzativo IT-Driven BI	8
1.5	Paradigma organizzativo Business/IT Hybrid BI	9
1.6	Piramide DIKW: dai dati grezzi alla saggezza	10
1.7	DBMS: OLTP e OLAP	12
1.8	Schema di organizzazione a due livelli di un Data Warehouse in BI	14
1.9	Processo di ETL: Extract, Transform, Load	15
1.10	Logo della Gartner S.p.A	16
1.11	Magic Quadrant di Gartner per la Business Intelligence e Analytics	17
1.12	Magic Quadrant di Gartner per i Cloud DBMS	18
2.1	Il Logo di Google BigQuery	20
2.2	Il Logo di Qlik	21
2.3	Struttura delle App in Qlik Sense	22
2.4	Hub di Qlik Sense	23
2.5	Le due modalità di associazione in Qlik Sense	24
2.6	Preview del modello dati in Qlik Sense a seguito di associazioni	25
2.7	Dashboard di Qlik NPrinting	26
3.1	Architettura del sistema di BI di Society	29
3.2	L'entità Agente	30
3.3	L'entità Cliente	30
3.4	L'entità Commissioni	31
3.5	L'entità Consulenza	31
3.6	L'entità Contratto	32
3.7	L'entità Incentivi	32
3.8	L'entità Portafoglio	33
3.9	L'entità Transazione	33
4.1	Architettura di riferimento per la costruzione dei KPI	36
4.2	Approccio ibrido adottato per le attività	37
5.1	Architettura riguardante il ripristino dei dati	42
5.2	Sistema di KPI Alfa/Beta	43

5.3	Architettura su BigQuery per il KPI Alfa Giornaliero	45
5.4	Architettura su BigQuery per il KPI Alfa MTD	45
5.5	Architettura su BigQuery per il KPI Beta Giornaliero	46
5.6	Architettura su BigQuery per il KPI Beta36	46
6.1	Tabelle di controllo della qualità dei dati	48
6.2	Flusso di controllo degli esiti sul controllo dei dati	48
7.1	App e relativi sheet relativi all'indice Alfa	52
7.2	Dashboard Overview relativa all'indice Alfa	53
7.3	Dashboard Dettagli relativa all'indice Alfa	53
7.4	Dashboard Analisi relativa all'indice Alfa	54
7.5	App e relativi sheet relativi all'indice Beta	54
7.6	Dashboard Overview relativa all'indice Alfa	55
7.7	Dashboard Analisi relativa all'indice Alfa	55
7.8	Dashboard Mashup relativa all'indice Alfa	56
7.9	Dashboard Mashup relativa all'indice Beta	56
7.10	App Mashup di Qlik Sense	57

L'emergere di una nuova ondata di dati provenienti da fonti quali l'Internet of Things, Sensor Networks, Open Data, dati delle applicazioni mobili e dei social network, insieme alla naturale crescita dei dataset all'interno delle organizzazioni, crea la necessità di una nuova gestione dei dati. I Big Data sono un campo emergente poiché possono far fronte a nuove "scale" di dati; inoltre, sono un campo in cui la tecnologia è innovativa ed offre nuovi modi per riutilizzare ed estrarre valore dalle informazioni.

La capacità di gestire efficacemente le informazioni e di estrarre conoscenza è oggi considerata un vantaggio competitivo fondamentale, e molte organizzazioni stanno costruendo il loro core business sulla capacità di raccogliere e analizzare informazioni per raggiungere tale scopo. Oggigiorno, l'adozione della tecnologia dei Big Data nei settori industriali non è più un lusso, ma un'esigenza imprescindibile per la maggior parte delle organizzazioni al fine di guadagnare un vantaggio competitivo sul mercato.

Per consentire alle aziende di sfruttare il proprio patrimonio informativo, in vista di decisioni tattico-strategiche, sono stati creati numerosi strumenti che prendono il nome di "strumenti di Business Intelligence (BI)". Grazie alla BI è possibile raccogliere e uniformare tutti i dati per, poi, gestirle e trasformarle in informazioni, utili, appunto, al supporto decisionale.

La Business Intelligence è un sistema di persone, modelli, metodi, processi e strumenti che rendono possibile la raccolta organizzata del patrimonio dei dati generati da un'azienda. Essa, inoltre, permette la trasformazione dei dati grezzi in informazioni utili e ne assicura la loro conservazione, reperibilità e presentazione in una forma semplice, flessibile ed efficace, tale da costituire un supporto alle decisioni di cui oggi nessuna azienda, media o grande, può più fare a meno.

In questa tesi, grazie ai tool di BI leader del mercato, quali Qlik e Google BigQuery, verrà illustrato il processo che porterà alla creazione di un sistema di KPI utili al supporto delle attività di una società di gestione del risparmio. Questi indicatori permetteranno di monitorare le prestazioni di clienti e dei consulenti finanziari al fine di delineare delle possibili scelte strategiche da intraprendere per apportare vantaggi competitivi alla società.

La presente tesi è strutturata come di seguito specificato:

- Nel Capitolo 1 introdurremo il lettore alla Business Intelligence e al valore dei dati al giorno d'oggi.
- Nel Capitolo 2 parleremo dei tool di BI a supporto del progetto di tesi svolto, ovvero Qlik e Google BigQuery.

- Nel Capitolo 3 descriveremo il contesto di riferimento aziendale analizzando la struttura gerarchica e il sistema informativo.
- Nel Capitolo 4 verranno descritte la raccolta e l'analisi dei requisiti funzionali e non funzionali.
- Nel Capitolo 5 si discuterà della progettazione delle nuove funzionalità.
- Nel Capitolo 6 descriveremo l'implementazione delle funzionalità progettate nel capitolo precedente.
- Nel Capitolo 7 mostreremo le principali App realizzate e discuteremo della possibilità, da parte di Qlik Sense, di realizzare un Mashup.
- Nel Capitolo 8 esprimeremo le considerazioni in merito al lavoro svolto analizzando i punti di forza e di debolezza del progetto ed elencando quali sono state le lezioni apprese.
- Nel Capitolo 9 trarremo le conclusioni e proporremo dei possibili sviluppi futuri.

Introduzione alla Business Intelligence

In questo capitolo verrà introdotta la Business Intelligence (BI) e in particolare si discuterà dell'importanza dei dati nel contesto storico odierno, cercando di comprendere e analizzare chi sono i protagonisti in gioco, si parlerà quindi delle figure che utilizzano la BI e dei mezzi di cui questi si servono. Si esporranno poi le fasi della BI e si analizzerà lo stato dell'arte discutendo di quali siano i benefici che questa apporta al mondo della Data Science.

1.1 Il valore dei dati

“Se torturi i dati abbastanza, alla fine confesseranno quello che vuoi”

È di Darrell Huff la citazione risalente all'anno 1954 ma che, al giorno d'oggi, risulta essere attuale più che mai, nell'era in cui le grandi moli di dati, anche conosciuti come “Big Data”, la fanno da padrone per le aziende che li generano quasi in modo automatico, e di cui spesso non ne conoscono le potenzialità.

anche se è vero che viviamo in un periodo storico che ci sommerge di informazioni, è vero allo stesso modo che non tutte queste apportino una visione chiara di quella che è la realtà, in quanto molte di queste informazioni sono spesso rumore di fondo che può distrarre; è per questo che nasce la necessità di applicare dei filtri.

L'analisi dei dati rappresenta, ad esempio, uno di questi filtri, poiché ha come compito quello di aiutarci a far emergere le informazioni che sono veramente importanti, in modo da analizzare meglio la realtà dei dati che abbiamo a disposizione.

Estrarre informazioni di valore non è affatto banale; sono richieste, infatti, molteplici attività a supporto come raccolta, conservazione, elaborazione, analisi, modellazione e rappresentazione visiva. Inoltre, per ognuna di queste, sono necessari competenze e strumenti ad hoc.

Per l'organizzazione è, quindi, uno sforzo notevole trattare i Big Data al fine di effettuare una selezione di quelle che sono le informazioni utili rispetto a quelle non significative; diventa, quindi, essenziale valutare, in primo luogo, il ritorno in termini di patrimonio informativo che si vuole ottenere facendo riferimento al proprio modello di impresa. Sostanzialmente, prima di avviare una iniziativa di tipo “Big Data”, è utile aver ben chiaro quali decisioni ci permetterebbero di prendere le informazioni estratte dall'analisi, e gli impatti che queste avrebbero sull'azienda.

Velocità La velocità di lettura e scrittura dei dati non è applicabile a tutti i settori, e nemmeno a tutte le aziende ed è, per questo, una V critica nelle 5 V dei Big Data. Con alta velocità dei dati si intende lavorare con dati in tempo reale, aggiornati al minuto, che mostrano gli approfondimenti più rilevanti.

Alcuni esempi di dati ad alta velocità sono:

- un negozio online che tiene traccia dei click su una schermata per ogni cliente che naviga sul sito;
- un’applicazione di navigazione che raccoglie e invia dati a ogni viaggiatore;
- una rete di social media che tiene traccia di quanto tempo un utente guarda un post e dei tipi di post con cui interagisce.

Varietà Un’elevata varietà di dati aiuta a scoprire sottili sfumature all’interno del database. Grazie alla varietà dei dati, è possibile mantenere un’esperienza personalizzata per i clienti. I dati raccolti possono provenire da diverse fonti, possono essere sia strutturati che non strutturati, possono essere interni o esterni all’azienda, in modo che il loro contenuto informativo sia in grado di rappresentare la complessità della realtà in esame, e non “una sola dimensione” della stessa.

Veridicità I dati raccolti devono essere affidabili, “integri” nella loro potenzialità di descrivere la realtà. Dati falsati da errori di rilevazione, da distorsioni dovute a malintesi oppure a inganni, costituiscono un grave danno, come osservato da chi opera nel settore: “Bad data is worse than no data”: meglio non avere informazioni, che utilizzare dati errati.

Valore Questa variabile fondamentale descrive che l’importanza dei Big Data stessi sta nella possibilità di poter introdurre informazioni utili per le aziende, e quindi di apportare dei benefici.

I dati fine a sé stessi, infatti, non hanno alcun valore. Per essere davvero utili devono poter essere convertiti in informazioni preziose, che permetteranno all’organizzazione di verificare, ed eventualmente modificare, la scelta delle sue mosse strategiche.

1.1.2 I benefici dei Big Data

Una scelta di strategia efficace è un fattore chiave per il successo aziendale. Esistono diversi casi d’uso dei Big Data, uno di quelli di maggiore impatto è l’utilizzo di questi per il miglioramento aziendale.

Ma perché proprio i Big Data sono vantaggiosi per le aziende? I vantaggi dei Big Data per le aziende sono principalmente quattro:

- *Migliorare l’esperienza dei clienti.* Quando si raccolgono dati sui clienti, si possono tracciare le tendenze per poter offrire loro il miglior servizio durante tutto il loro percorso con l’azienda.
- *Problem Solving.* I Big Data possono essere utilizzati per identificare i processi che generano dei colli di bottiglia e porvi rimedio con l’analisi diagnostica dei dati.
- *Aumento dei ricavi.* I Big Data consentono alle aziende di aumentare i ricavi anche iterando i prodotti esistenti e progettando nuovi lanci in base alle esigenze comprovate dei clienti.
- *Riduzione dei costi.* Per loro natura, i Big Data funzionano su scala, essi eliminano la necessità di raccogliere e analizzare manualmente i dati, il che, insieme alla riduzione del personale adibito, comporta un notevole risparmio sui costi.

1.1.3 Come analizzare i Big Data

Esistono quattro tipi principali di analisi dei dati che possono essere utili per le aziende quando si parla di Big Data: descrittiva, diagnostica, predittiva e prescrittiva (Figura 1.2).

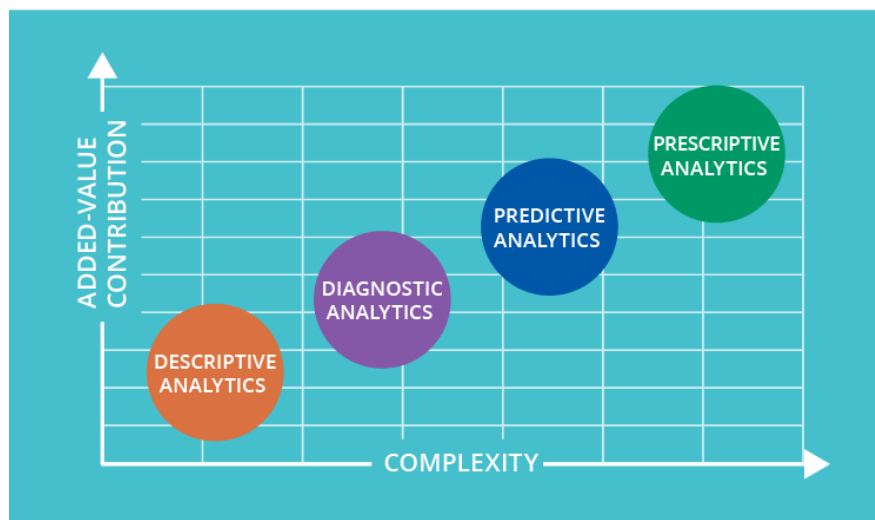


Figura 1.2: Big Data: le quattro principali tipologie di analisi

Vediamoli più in dettaglio:

- *Analisi Descrittiva*: mostra ciò che sta accadendo in un determinato momento in base ai dati in arrivo, spesso tramite una dashboard in tempo reale o mediante l'invio di report.
- *Analisi Diagnostica*: esamina delle informazioni passate e memorizzate per determinare cosa è successo e perché, spesso sotto forma di una dashboard analitica.
- *Analisi Predittiva*: prevede scenari probabili di ciò che potrebbe accadere.
- *Analisi Prescrittiva*: rivela quali azioni devono essere intraprese a seguito di un'analisi predittiva e si traduce in regole e passi raccomandati.

Queste analisi sono di diversa complessità come mostrato in Figura 1.2. Tuttavia, è anche vero che, al crescere di questa, cresce anche il valore aggiunto apportato all'azienda.

1.1.4 La nascita della Business Intelligence

Note, quindi, le grandi potenzialità dei Big Data c'è bisogno di un insieme di metodologie, processi, architetture e tecnologie che da un input di dati strutturati o non, restituiscano in output informazioni utili all'azienda in termini di Decision Making, ovvero un insieme di decisioni strategiche che ne migliorino il business, e quindi i profitti.

È proprio a questo scopo che nasce la Business Intelligence (BI), un importante strumento che si pone come obiettivo quello di proporre un cambiamento di approccio e cultura aziendale per poterla al meglio valorizzare. Una volta in possesso dei dati ed effettuate le analisi descrittive e diagnostiche, per la BI sarà possibile avviare anche analisi predittive e prescrittive, cercando, quindi, di prevedere l'andamento dell'azienda, con dei processi conosciuti come Business Analytics Process.

In breve, la Business Intelligence è essenziale per interpretare i dati e da questi estrapolare informazioni per capire come sta andando la nostra azienda e quali scelte prendere per un futuro più competitivo sul mercato.

Si potrebbe pensare che il termine “Business Intelligence” sia un neologismo; in realtà non è così. Infatti, questo veniva utilizzato già nel 1800; comparve per la prima volta nell’opera pubblicata da Richard M. Devens nel 1865: “Cyclopaedia of Commercial and Business Anecdotes” nella quale si utilizzava il termine “Business Intelligence” per descrivere il segreto del banchiere Sir Henry Furnese, il quale, raccogliendo informazioni politiche e di mercato al fine di proporre delle scelte strategiche prima dei suoi concorrenti, riusciva ad aumentare i suoi guadagni.

È nel 1958 che, però, verranno riconosciuti i benefici della Business Intelligence, grazie alla pubblicazione dell’articolo “A Business Intelligence System” opera di Hans Peter Luhn (Figura 1.3) che, in quel tempo, lavorava per IBM.

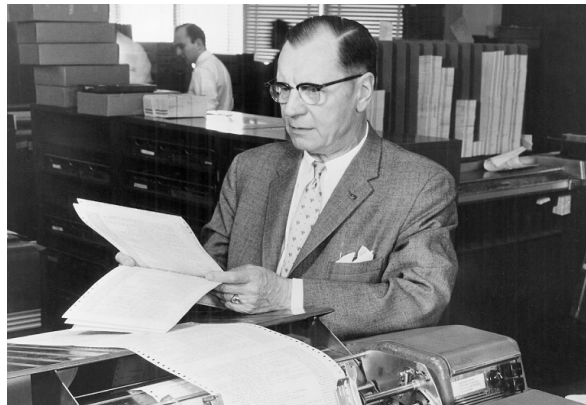


Figura 1.3: Hans Peter Luhn: il padre della Business Intelligence

Egli espose, nel suo articolo, un sistema automatico in grado di trasmettere informazioni in diverse aree di organizzazioni industriali, scientifiche e governative. Egli si concentrò su questi particolari organismi poiché erano quelli maggiormente coinvolti nella terza rivoluzione industriale. Si necessitava, quindi, di affiancare alla crescente innovazione tecnologica un sistema di organizzazione di dati che li rendeva più facilmente fruibili.

Nei primi anni '70 i vendor cominciarono a proporre dei sistemi di supporto alle decisioni (DSS, da Decision Support System) alle aziende. Nacque sin da subito una certa competizione che portò a migliorare man mano questi software e alla creazione dei primi Data Warehouse, che permisero il raggruppamento di dati prima ripartiti in diversi database. Tutto ciò diede il via all’ideazione di diversi tool e software utilizzati nella BI.

Nell’ultimo ventennio del secolo scorso la Business Intelligence veniva sempre più adottata dalle aziende, tuttavia, erano due i principali problemi riscontrati:

- i software e i tool, per via della loro difficoltà di utilizzo da parte di utenti meno esperti, non erano fruibili al massimo;
- le tempistiche necessarie ad elaborare i dati e generare i report risultavano proibitive.

Per tali ragioni iniziarono a nascere tool di BI facilmente utilizzabili dagli utenti meno esperti e venne migliorata la velocità di elaborazione dei dati. Le aziende utilizzavano tantissimo la Business Intelligence costringendo, così, le aziende che, invece, non la utilizzavano a decidere se investire su questa o essere surclassate dai competitor più innovativi.

Allo stato dell’arte i tool e i software di BI sono mirati a soddisfare le esigenze di qualsiasi azienda facente parte di un settore o mercato specifico, che sia per questioni di sicurezza aziendale e/o di crescita economica.

1.2 I protagonisti della BI

Come si evince dalla sezione precedente, l'obiettivo della Business Intelligence è quello di generare un ambiente che contenga informazioni acquisite dalle diverse fonti aziendali. In questo modo, i dati ricavati saranno in grado di supportare le attività di analisi e di reportistica in modo semplice, efficace e intuitivo, attraverso i Decision Maker Aziendali.

Nelle prossime sottosezioni illustreremo, quindi, chi sono le figure che fanno BI e i mezzi informatici di cui si servono per farlo al meglio.

1.2.1 Le figure coinvolte nella BI

Le varie figure coinvolte nei processi di Business Intelligence possono essere sia interne che esterne all'associazione.

Storicamente la BI è sempre stata uno spin-off dell'IT, poiché, per metterla in pratica, occorre essere esperti di informatica. Questa tipologia di idea, anche conosciuta come *IT-driven BI* (Figura 1.4), porta ad avere due distinti gruppi di professionisti coinvolti in un dato progetto:

- il reparto Business nei suoi vari reparti funzionali, guidato dal cosiddetto *Business Project Sponsor*;
- il reparto IT, guidato dal *BI Solution Architect*.

Il primo determina *cosa* deve offrire la BI, quindi di cosa il business ha bisogno; il secondo decide *come* costruirla, quindi quali sono i requisiti e qual è la miglior soluzione.

Il BI Solution Architect possiede un team ben strutturato, formato da figure tecniche specializzate, come il *Model Designer*, che definisce il modello, ovvero la struttura semantica finale accessibile al business per svolgere le attività di analisi. Questo modello non è null'altro che la fonte dati da cui leggono i report e le dashboard creati dal *Visual Designer*. Tale strato semantico poggia sul Data Warehouse che, a sua volta, è disegnato dal *Data Architect* il quale decide struttura e nomenclatura in contesto rigorosamente globale, corporate e caricato dall'*ETL Specialist*. Quest'ultimo pone l'attenzione a temi quali la performance dell'ETL, il registro degli errori, l'auditing dei dati e il checkpointing del flusso. Il paradigma appena descritto viene riportato in Figura 1.4.

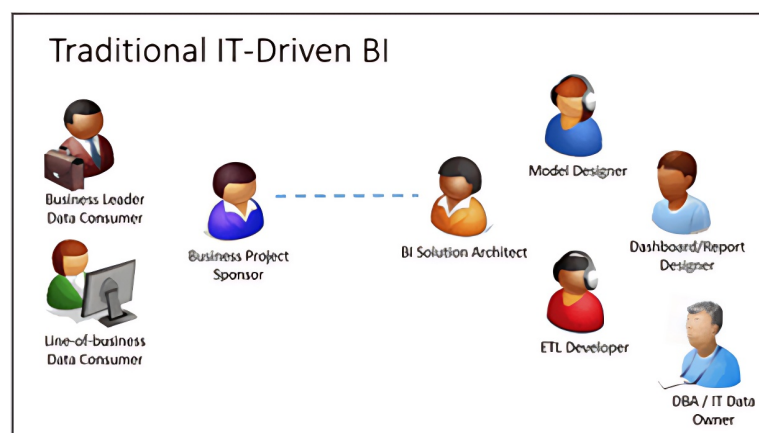


Figura 1.4: Paradigma organizzativo IT-Driven BI

Esso demanda al reparto IT l'onere di dover sviluppare per intero lo Stack BI. Potenzialmente il progetto di Business Intelligence potrebbe non avere mai fine e, quindi, occupare

gran parte del tempo dell'IT in progetti trasversali. Ciò potrebbe non essere proficuo. Nello stesso tempo, ciò costringe, incarica gli utenti di business a formarsi al fine di permettere loro di sviluppare autonomamente ciò di cui hanno bisogno secondo le regole funzionali aziendali. Tali utenti, produttori e consumatori di dati, non sono più solo utenti finali e vengono chiamati *Data Analyst*.

Ultima figura del paradigma IT-Driven BI è il *Data Steward*, che supervisiona la qualità, gestione e distribuzione del dato aziendale.

Un altro tipo di organizzazione è la *Business/IT Hybrid BI* (Figura 1.5) nella quale, invece, "ci si sporca un po' tutti le mani" assieme.

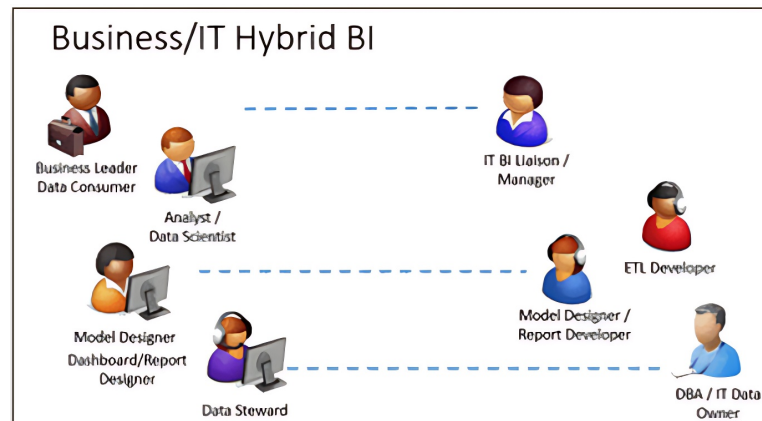


Figura 1.5: Paradigma organizzativo Business/IT Hybrid BI

Il Business mantiene il proprio potere decisionale e un certo livello di autonomia, ma è affiancato dall'IT che, costantemente, lo supporta e lo consiglia. Nelle prime fasi di progetto l'IT è maggiormente presente. Tale presenza gradualmente diminuisce nel tempo senza mai arrivare allo zero.

In questo paradigma si ha una duplicazione di alcune figure descritte nel modello precedente; in particolare, il Model Designer e il Visual Designer trovano un loro corrispettivo nei vari reparti aziendali di business. Ruoli tecnici come l'ETL Specialist e il DBA rimangono esclusiva del reparto IT.

Questo è un paradigma che trova la sua massima espressione quando l'azienda opta di adottare dei *Consulenti BI esterni*, in quanto questi sollevano l'IT da oneri eccessivi. Esso, inoltre, permette di ottimizzare i tempi del business dal momento che gli esperti del settore dovranno intervenire solo in caso di necessità. In questo campo spicca la figura del *BI Analyst* che è dotato di tutte le abilità necessarie dei due mondi IT e funzionale, e che è capace di comunicare con tutti gli agenti in gioco al giusto livello di dettaglio tecnico.

1.2.2 I sistemi utilizzati nella BI

Affinché le figure appena descritte svolgano il loro compito al meglio devono affidarsi a mezzi che permettano di estrarre, immagazzinare, manipolare e analizzare i dati. Questi riguardano principalmente i sistemi informativi, i database, i DBMS e le loro varie evoluzioni.

1.2.2.1 Piramide DIKW

Sono conoscenza e saggezza ciò che i soggetti della Business Intelligence vogliono ottenere partendo da dati e informazioni. Questi quattro elementi fanno parte della piramide DIKW (Figura 1.6), un modello che raffigura il processo che porta dai dati grezzi alla saggez-

za, passando attraverso un processo di aggregazione-contestualizzazione (informazione) e applicazione-sperimentazione (conoscenza).

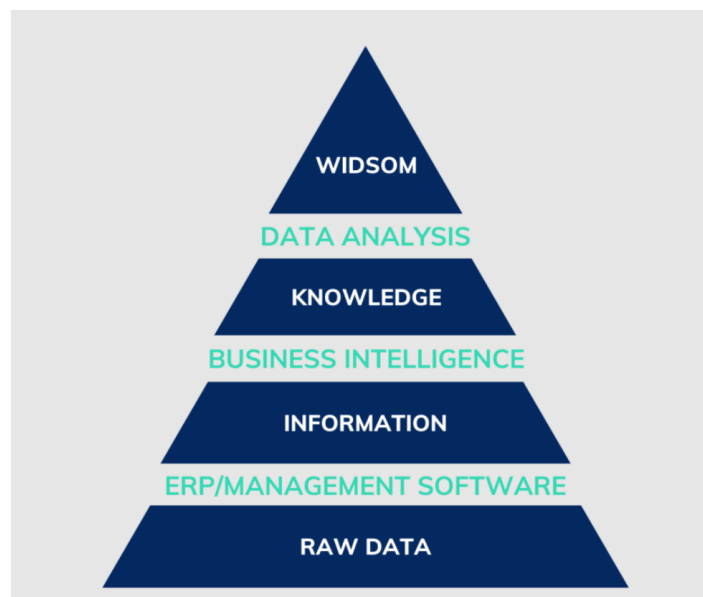


Figura 1.6: Piramide DIKW: dai dati grezzi alla saggezza

Nell'ambito aziendale, quindi, tale piramide lascia comprendere come i dati posseduti dalle aziende possano essere gestiti al fine di ottenere informazione per poi ricavarne conoscenza e permettere ai *Decision Maker* di fare scelte il più competitive possibile.

Per riuscire, quindi a diventare un'azienda che utilizza i dati per prendere decisioni vincenti a livello operativo/strategico è necessario partire da una corretta acquisizione, gestione e archiviazione dei dati grezzi. Analizziamo i blocchi della gerarchia:

- *Dati Grezzi.* Sono i dati non ancora processati, frutto di una raccolta di massa e senza criterio, sostanzialmente sono simboli che ancora non hanno una semantica ma sono la base di partenza verso l'acquisizione di conoscenza. È quindi, di fondamentale importanza che l'azienda ne sia in possesso. Tali dati saranno irrilevanti finché l'azienda non possederà strumenti per analizzarli o per trarne informazioni; quindi, fino ad allora, non contribuiranno a generare valore. È proprio in questa fase che entra in gioco la Business Intelligence.
- *Utilizzo della BI.* La Business Intelligence utilizza strumenti propedeutici all'elaborazione dei dati. Essa si pone questo in un contesto che è circondato da domande del tipo: "cosa?", "dove?", "quando?", dando un primo significato alle informazioni raccolte. Il lavoro del *Consulente* che utilizza software di BI consiste proprio in questo: aggregare e contestualizzare i dati per trasformarli in informazioni. Per arrivare a prendere delle scelte più consapevoli c'è bisogno di salire ancora nella piramide DIKW.
- *Dall'informazione alla conoscenza.* L'output della BI è rappresentato da report, grafici o dashboard che aggregano grandi quantità di dati e restituiscono tabelle che collegano informazioni diverse. Sarà a questo punto che i decision maker cercheranno di comprendere cosa rappresentano i dati ed effettuare delle scelte che porteranno a delle azioni aziendali.
- *dalla conoscenza alla saggezza.* Una volta raccolte le informazioni dai dati storici si può pensare di passare ad una progettazione delle azioni future che l'azienda prenderà; tale

attività viene detta, anche, forecasting. È quest'ultimo, infatti, a rappresentare la sfida più grande nel mondo della Data Analysis.

1.2.2.2 I Database

Per quanto detto finora, risulta indispensabile un modo di memorizzare i dati che abbia determinate caratteristiche. In particolare, saranno necessarie le seguenti proprietà:

- *Consistenza*. I dati devono essere utilizzabili nelle applicazioni aziendali.
- *Sicurezza*. È necessario impedire danni irreversibili alle informazioni memorizzate.
- *Integrità*. È necessario garantire la conservazione del dato senza perdite.

Inoltre, avendo a che fare con i Big Data, saranno presenti le seguenti caratteristiche:

- grandi volumi di dati;
- alte velocità di memorizzazione;
- varietà dei dati;
- veridicità dei dati;
- valore da trarre dal dato.

Vengono a crearsi quindi delle collezioni di dati correlati dette Dataset; ciascun membro all'interno di un dataset condivide gli stessi attributi e proprietà.

I dati elaborati in ambito Big Data sono di vario tipo e sono suddivisibili in:

- *Dati Strutturati*: sono dati che si conformano a un modello o ad uno schema e sono memorizzati all'interno di tabelle relazionali.
- *Dati Non Strutturati*: non sono conformi ad uno schema e sono quindi di natura di diversa. Esse, quindi, non possono essere interrogati tramite SQL, ma vengono memorizzati in database No-SQL.
- *Dati Semi-Strutturati*: sono per loro natura non relazionali ma hanno, comunque, un certo livello di struttura definito.

1.2.2.3 I DBMS

I database appena descritti necessitano di essere organizzati mediante dei software, detti Database Management System, o DBMS, che consentono le operazioni CRUD: creazione, manipolazione e interrogazione efficiente dei database. Essi, inoltre, si occupano della sicurezza e dall'integrità dei database stessi.

Nell'ambito della BI i DBMS possono essere classificati in due differenti tipologie (Figura 1.7):

- *OLTP*: sta per *On-Line Transactional Process*; è volto alla lettura e alla modifica parziale del dato. I dati sono aggiornati; non fanno riferimento, quindi, a dati storici, ponendo l'interesse sulle attività quotidiane. Inoltre essendo il database utilizzato contemporaneamente da un vario numero di utenti, il DBMS deve gestire meccanismi di controllo, ripristino e concorrenza.

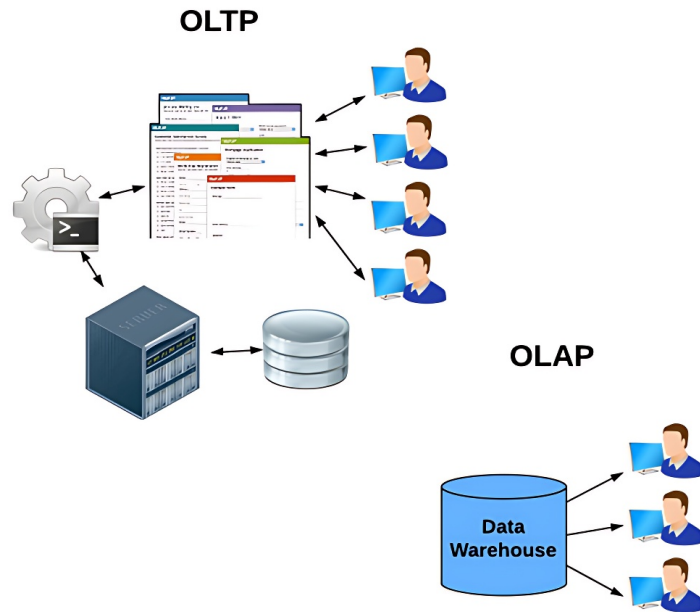


Figura 1.7: DBMS: OLTP e OLAP

- *OLAP*: sta per *On-Line Analytical Process*. In questo caso i dati sono aggregati e storici in modo da coprire un arco temporale sufficiente a sottolineare delle variazioni. Il volume dei dati cresce maggiormente e anche se le operazioni sono di sola lettura, le query risulteranno molto più complesse al punto da richiedere per la loro formazione degli utenti specializzati e formati.

1.2.2.4 I Data Warehouse

I dati a seguito di elaborazione da parte della Business Intelligence confluiscono in collezioni di dati strutturati chiamati Data Warehouse (DW). Questi permetteranno ad analisti o gestori di vario tipo di eseguire query, effettuare analisi e generare report in modo che le aziende possano ricavare valore da tali informazioni.

Ci sono delle caratteristiche che differenziano i DW dagli altri tipi di raccolte di dati. Infatti, i Data Warehouse sono:

- *Integrati*: essendo i dati di un Data Warehouse ottenuti come unione di dati provenienti da fonti esterne, c'è bisogno di creare coerenza, ciò può avvenire, ad esempio mediante metodi di codifica uniforme od omogeneità semantica delle variabili.
- *Orientati al soggetto*: i DW sono orientati a specifiche aree aziendali, funzioni o applicazioni. I dati, quindi, sono archiviati al fine di facilitarne la lettura e l'elaborazione da parte degli utenti. Lo scopo non è più minimizzare le ridondanze ma fornire dati organizzati per meglio produrre informazioni utili. Si passa da una progettazione per funzione a una modellazione di dati che consente una visione multidimensionale del problema.
- *Variabili nel tempo*: i dati memorizzati in un DW coprono un orizzonte temporale molto più alto dei singoli sistemi operazionali. Quindi, i dati disponibili sono, di norma, antecedenti alla data in cui l'utente li interroga.
- *Non volatili*: i dati contenuti nel Data Warehouse non sono modificabili, infatti, l'accesso è disponibile in sola lettura, semplificando di molto la progettazione.

Dovendo il DW compiere operazioni complesse e di pesante elaborazione, risulta chiaro che la sua architettura risulta essere di fondamentale importanza. Per questo i DW devono possedere caratteristiche essenziali, come:

- *separazione*: occorre mantenere separate logicamente l'elaborazione analitica e l'elaborazione operativa;
- *scalabilità*: i volumi crescono velocemente e, di conseguenza, l'architettura deve essere facilmente ridimensionabile in termini di utenti da soddisfare;
- *estensibilità*: non ci deve essere bisogno di riprogettare il sistema nel caso di ingresso di nuove tecnologie o software;
- *sicurezza*: per via della natura strategica dei dati contenuti, è essenziale tenere un rigido controllo degli accessi;
- *amministrabilità*: è utile che la complessità delle mansioni amministrative non sia eccessiva.

Le architetture, quindi, sono di quattro tipologie principali:

- *Ad un livello*: tale architettura ha come obiettivo quello di minimizzare la memorizzazione di dati, eliminando le ridondanze. È una tipologia virtuale di DW, implementata come una vista multidimensionale dei dati. Tale architettura non rispetta, però, la separazione tra l'elaborazione analitica e operativa.
- *A due livelli*: si suddivide in quattro livelli distinti, avremo il livello sorgente con varie fonti di dati, il livello dell'alimentazione, con i processi classici di BI, il livello di DW, in cui vengono raccolte le informazioni; da quest'ultimo è possibile leggere direttamente i dati oppure creare Data Mart, ovvero repliche parziali orientate verso specifiche aree del business. C'è inoltre, accanto al DW, il repository dei meta-dati che mantiene informazioni sui meccanismi di accesso, sulle sorgenti, sugli utenti, etc. Infine troviamo il livello di analisi, che permette la consultazione efficiente dei dati integrati per simulazione, analisi e stesura report. Questa tipologia di architettura è mostrata in Figura 1.8.
- *A due livelli con Data Mart indipendenti*: in questo caso i Data Mart sono alimentati direttamente dalle sorgenti e quindi risultano indipendenti.
- *A tre livelli*: in questo caso viene introdotto un terzo livello a valle delle operazioni di Business Intelligence. Tale livello materializza i dati operazionali ottenuti da processi di integrazione e ripulitura.

1.3 Le Fasi della BI

Come spiegato in precedenza, nella Business Intelligence sono impiegate figure specifiche e mezzi specifici. Questi, però, per poter essere utilizzati al meglio hanno bisogno di regole, o meglio, passi da seguire per poter arrivare da un insieme di dati apparentemente non correlati a delle informazioni che diano vantaggio competitivo.

Esistono quindi delle fasi ben definite nei processi di BI. Questi sono:



Figura 1.8: Schema di organizzazione a due livelli di un Data Warehouse in BI

- *Fase Zero: comunicare con l'utente.* In BI la figura dell'analista, deve essere in grado di interagire con le figure di tutti i dipartimenti aziendali, comprendere ed omologare l'uso di una terminologia che sia univoca e condivisibile con il resto dell'azienda, in modo che lo stesso concetto non sia espresso con termini diversi evitando, così, di creare ambiguità.
- *Fase Uno: comprendere la richiesta.* Se il vocabolario dei termini aziendali non è ben definito, potrebbe essere molto complicato comprendere il bisogno del cliente che, di solito, conoscendo bene i propri processi, omette dei dettagli di fondamentale importanza poiché per lui risultano ovvi. In questo caso il lavoro potrebbe essere ostacolato da incomprensioni e generare errori che allungheranno i tempi, e quindi i costi. Ecco perché risulta fondamentale sviluppare una terminologia chiara e condivisa tra i membri dell'associazione.
- *Fase Due: identificare l'origine dei dati.* Le fonti dei dati nelle aziende sono di vario genere; queste possono arrivare da basi di dati, fogli di calcolo, documenti di testo, e tante altre fonti. È di fondamentale importanza che l'analista BI conosca la provenienza dei dati che dovrà, poi, rielaborare.
- *Fase Tre: elaborazione dei dati.* Spesso il lavoro degli analisti BI viene riassunto con la parola *rielaborazione*, termine troppo riduttivo per spiegare, in realtà, a cosa vengono sottoposti i dati estratti dal sistema informativo. Come spiegato in precedenza, i dati vengono memorizzati in basi di dati specifiche dette Data Warehouse; quest'ultimo è il grande protagonista dietro ad ogni progetto di Business Intelligence. Infatti è in base alla sua qualità di che dipende la buona riuscita di un progetto.

È, quindi, questa la fase in cui si concentra la maggior parte del lavoro, attraverso quelli che sono i passi più famosi della BI, ovvero i passi di *ETL*, dall'acronimo inglese *Extract, Transform and Load* (Figura 1.9).

Obiettivo dell'estrazione è quello di unire tra loro dati provenienti da varie fonti, come, ad esempio, database, registri, report e altre attività transazionali.

La fase di trasformazione è la più critica, infatti, in questa fase ai dati si applicano le regole aziendali necessarie a soddisfare i requisiti. Gli standard che guidano la trasformazione, al fine di garantire una certa qualità del dato sono: *Standardizzazione*, ovvero definizione dei dati, della modalità di memorizzazione e altri fattori che definiranno le fasi successive del processo; *Deduplicazione*, che esclude o elimina dati ridondanti; *Verifica*, che mette a confronto informazioni simili al fine di eliminare ulteriori ridondanze; *Ordinamento*, che utilizza le regole di trasformazione per determinare come ogni singolo dato viene classificato e dove sarà collocato successivamente. Spesso, infatti i processi di ETL hanno come output tabelle di aggregazione per report riepilogativi, lavoro che richiede un certo ordinamento e una successiva aggregazione dei dati.

L'ultima fase del processo di ETL prevede il caricamento dei dati estratti e trasformati in quella che sarà una nuova destinazione, di solito il DW, tramite il caricamento completo o incrementale.

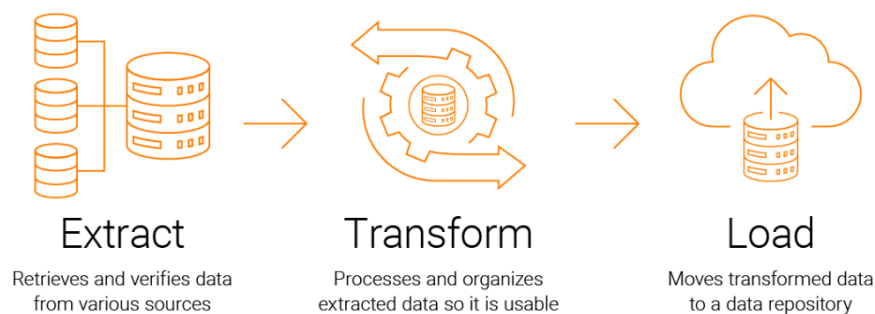


Figura 1.9: Processo di ETL: Extract, Transform, Load

- *Fase Quattro: visualizzare l'informazione.* Gli utenti della BI hanno accesso solo ed esclusivamente al prodotto finale scaturito da questa fase; esso è rappresentato da un insieme di documenti che visualizzano a livello grafico, in diverse forme e modalità di accesso o interazione, dati da esso generati. Tale insieme di dati è chiamato, anche, *reportistica*.

Da non sottovalutare l'importanza di presentare dei dati il più corretti possibile, nella maniera più intuitiva ed appetibile possibile, in modo tale che tutti possano comprendere i vantaggi legati all'adozione della BI. Se l'utente non capisse i dati e facesse delle scelte sbagliate, questo andrebbe a discapito della propria azienda.

La BI fornisce come output dati pronti per essere analizzati. I report servono solo a presentare i dati in un certo modo che risulti facile da comprendere. Il reporting può essere di 3 livelli, ovvero:

- 1 - *il livello operativo;* si tratta del livello più lontano dai vertici dell'azienda dove, solitamente, il dettaglio è molto alto e i grafici non sono strettamente necessari; i dati, quindi, vengono presentati sotto forma di tabelle o matrici.
- 2 - *il livello direzionale;* in questo caso c'è meno dettaglio di informazioni, ma la visione è più ampia; i grafici cominciano ad avere un ruolo importante e c'è la possibilità di mostrare previsioni sul futuro imminente.
- 3 - *il livello strategico;* il dettaglio è minimo. In questo caso viene utilizzato un diverso strumento della BI, ovvero il dashboarding.

I dashboard (o cruscotti, in italiano) perdono livelli di dettaglio e presentano misure che riassumono, attraverso degli indicatori, l'andamento dell'intera azienda. Queste visualizzazioni possono essere statiche o dinamiche. Nel primo caso avremo uno

snapshot non navigabile; di solito si trova in messaggi di e-mail inviati agli utenti finali; nel secondo caso, invece, c'è un certo livello di interattività con le informazioni, applicando filtri e agendo sul dettaglio di visualizzazione.

1.4 Stato dell'arte della BI con Gartner

Ad oggi sono molteplici gli strumenti che permettono alle aziende di praticare la Business Intelligence. Non tutti, però, hanno le stesse caratteristiche e permettono di prendere le stesse decisioni. Gartner (Figura 1.10) è una multinazionale che si occupa di consulenza strategica, ricerca e analisi nel campo della tecnologia dell'informazione.



Figura 1.10: Logo della Gartner S.p.A

Ogni anno questa società produce i cosiddetti *Magic Quadrant*, o *MQ*, che hanno il compito di analizzare i principali player su mercato riguardanti un determinato settore o servizio, utilizzando dei metodi proprietari di analisi qualitativa, studiando le tendenze di mercato, come direzione, maturità e partecipanti.

Graficamente un Magic Quadrant si presenta come una tabella il cui valore competitivo delle principali aziende fornitrici di tecnologie in mercati che sono in rapida crescita. Viene suddiviso in quattro categorie: *Leader*, *Visionari*, *Giocatori di nicchia* e *Sfidanti*.

1.4.1 Come si utilizza un MQ di Gartner?

Grazie a questa tabella si può, a prima vista, comprendere quali sono i fornitori più affidabili e su cui si potrebbe decidere di investire.

Tuttavia, concentrarsi soltanto sul quadrante dei leader non è sempre la scelta migliore. Ad esempio, ci potrebbero essere buone ragioni per scegliere gli sfidanti, o un player di nicchia, poiché potrebbe supportare le esigenze della specifica azienda meglio di un leader; tutto dipende da quali sono gli obiettivi di business di una società.

1.4.2 Come si legge un MQ di Gartner?

Come spiegato in precedenza un competitor viene inserito in una delle 4 categorie dei MQ, che sono:

- *Leader*: qui troviamo i competitor che hanno i punteggi più alti per quanto riguarda completezza di visione e capacità di esecuzione. Qui i fornitori dimostrano un'adeguata comprensione delle esigenze del mercato e hanno piani ben articolati che i clienti già possono utilizzare quando progettano le loro strategie. Inoltre, sono presenti nelle cinque principali aree geografiche del mondo e hanno capacità finanziarie ottime; pertanto, promettono un ottimo supporto della piattaforma.
- *Challengers o Sfidanti*: in quest'area troviamo i competitor che possono rappresentare una minaccia per i leader, grazie ai loro prodotti forti e alla loro continua crescita.

- *Visionari*: sono i competitor che forniscono i prodotti più innovativi, ma che non hanno ancora dimostrato di poter acquisire una fetta di mercato sostenibile poiché sono, spesso, società private con minori obiettivi di acquisizione.
- *Player di nicchia*: qui, i fornitori sono spesso concentrati strettamente su specifici segmenti di mercato, effettuando un'economia verticale. Il quadrante comprende fornitori che stanno adottando i loro prodotti per entrare nel mercato o fornitori maggiori che hanno difficoltà a sviluppare ed eseguire la loro visione.

1.4.3 I Magic Quadrant di BI e Cloud DBMS

Nell'ambito dello sviluppo della presente tesi sono principalmente due i campi in cui ci interessa analizzare i MQ di Gartner; essi sono:

- Il Magic Quadrant di Gartner per la Business Intelligence e Analytics, riportato in Figura 1.11.



Figura 1.11: Magic Quadrant di Gartner per la Business Intelligence e Analytics

- Il Magic Quadrant di Gartner per i Cloud DBMS, riportato in Figura 1.12.

Sono, infatti, due le soluzioni per la Business Intelligence adottate al fine di sviluppare il progetto descritto nella presente tesi, e sono Google BigQuery e Qlik. Entrambi sono posizionati nel settore dei leader nei rispettivi Magic Quadrant di Gartner. Questi software verranno analizzati nel prossimo capitolo.

1.5 Costi e benefici della BI

Ovviamente scegliere la Business Intelligence per la propria azienda presenta dei costi. Nelle prossime sottosezioni analizzeremo costi e benefici a cui si va incontro quando si adotta un progetto di BI.



Figura 1.12: Magic Quadrant di Gartner per i Cloud DBMS

1.5.1 Quanto mi costa la BI?

Per un progetto di Business Intelligence i costi risultano essere di vario genere:

- *Software:* È necessario l'acquisto di software per affrontare le varie fasi, dall'implementazione alla visualizzazione passando per l'ETL. Potrebbe essere necessario, inoltre, munirsi di licenze. Se l'azienda è di piccole dimensioni e questi costi possono essere non sostenibili, esiste la possibilità di programmi gratuiti, rinunciando ovviamente alla completezza.
- *Organizzazione:* È importante quantificare anche il tempo come risorsa di costo, in quanto questo è necessario per coordinare le diverse risorse umane coinvolte nel progetto, ad esempio, per una corretta formazione.
- *Implementazione:* La progettazione, lo sviluppo e la validazione di un progetto è un costo inevitabile in quanto validare comprende, ad esempio, un lavoro di Quality Assurance che va remunerato.
- *Addestramento:* Il training, che sia rivolto agli sviluppatori o agli utenti finali va condotto nei metodi e nei tempi migliori e anche uno studio del genere richiede un costo.

1.5.2 Quanto mi costa non avere la BI?

Quantificare il profitto di un progetto non è mai un processo certo, in quanto ci sono vari fattori che potrebbero influenzarlo anche dall'esterno, tenendo conto sia di rischi di possibilità ma anche di rischio di risparmio.

Va quindi eseguita una manovra inversa: quanto si risparmia implementando la BI? Non esiste neanche in questo caso una formula certa ma, statisticamente, possiamo dire che si basa su fattori come il numero di stakeholder con cui l'azienda ha a che fare, il numero medio

di documenti, il numero medio di ore mensili impiegate per lo sviluppo e la validazione in ambito tradizionale e lo stipendio medio per dipendente. Si vede subito che questo risparmio può facilmente raggiungere numeri considerevoli, senza contare il valore aggiunto per l'azienda.

1.5.3 Benefici diretti della BI

Conoscere tutti i benefici della BI non è possibile dato che il beneficio che possono trarre diverse aziende che operano in settori diversi non è lo stesso. Ci sono però dei vantaggi diretti di cui tutti possono godere, Questi sono:

1. *Automatizzazione.* Non c'è più necessità, da parte del business, di prelevare e integrare manualmente i dati, operazione che, tra l'altro, è soggetta ad errori umani. Grazie agli strumenti della BI, il business è sollevato da queste scomode attività.
2. *Flessibilità.* I Data Warehouse permettono di rimediare alle restrizioni imposte. Con gli strumenti odierni di visualizzazione ed analisi l'attività del business è semplificata e potenziata.
3. *Puntualità.* Il business conduce ogni giorno le analisi prendendo decisioni su un dato che è costantemente aggiornato.
4. *Produttività.* L'IT è sollevato da alcune attività potendosi, così, concentrare su attività produttive di creazione di contenuti.
5. *Qualità dei dati.* Le informazioni usate dall'utente finale sono generate da procedure validate.
6. *Indipendenza.* Con la BI a regime, il reparto IT e il business sono indipendenti, e ciascuno può concentrarsi sulle proprie attività per la maggior parte del tempo.
7. *Sguardo al futuro.* Un progetto di BI forma i propri dipendenti e li rende più liberi di ricercare nuovi aspetti e problematiche prima ignorate generando, così, miglioramenti nel business aziendale.

In questo capitolo verranno analizzati i principali strumenti utilizzati nello svolgimento del progetto relativo alla presente tesi ovvero, BigQuery di Google e Qlik. In particolare, saranno descritti quelli che sono i loro principali punti di forza e le loro debolezze, nonché i tool che mettono a disposizione degli utenti di Business Intelligence.

2.1 Google BigQuery

Google BigQuery (Figura 2.1) è un Data Warehouse cloud di proprietà dell'azienda statunitense, in grado di archiviare e recuperare dati attraverso l'utilizzo di un ambiente SQL. Grazie a ciò, le aziende possono acquisire e analizzare i dati in tempo reale, evitando l'acquisto di macchinari e la creazione di reti ad hoc. BigQuery si mette a disposizione di qualunque tipo di azienda, grande o piccola, che abbia grandi set di dati da analizzare attraverso la Business Intelligence.



Figura 2.1: Il Logo di Google BigQuery

2.1.1 Funzionalità e vantaggi di BigQuery

Sono svariati i vantaggi offerti da BigQuery; essi si possono riassumere in:

- *Importazione dati e condivisione.* Possiamo trasmettere a BigQuery migliaia di righe al secondo contenenti, in modo da poter offrire un'analisi real time, condividendole, istantaneamente, con gli utenti.
- *Serverless.* Non ci sono costi di manutenzione poiché la gestione dell'infrastruttura e la sicurezza dei dati sono garantite dall'azienda che fornisce il servizio e, quindi, da Google.

- *SQL Standard.* La standardizzazione del linguaggio SQL evita alle aziende di dover riscrivere codice da zero. Queste aziende, già in possesso di database, scelgono di adottare un servizio cloud come Google BigQuery.
- *Backup e ripristino.* La posizione dei dati viene diversificata, salvando le informazioni in diversi data center, di cui esistono vari backup. In questo modo i dati sono al sicuro ed il ripristino è semplice e veloce.
- *Data transfer service.* BigQuery offre la possibilità di trasferire automaticamente i dati da strumenti esterni, come Marketing Platform, Google Ads o YouTube.
- *Piattaforma per l'intelligenza artificiale.* BigQuery permette, grazie agli ultimi aggiornamenti, di testare algoritmi di machine learning.
- *Piattaforma per la Business Intelligence.* BigQuery Permette, grazie alla trasformazione e all'analisi dei dati, di applicare la Business Intelligence.
- *Controllo dei costi.* Per ogni azienda i costi sono personalizzabili in base al tipologia di mercato in cui opera e al consumo che essa genera.

Direttamente dal sito ufficiale troviamo elencate le funzionalità di BigQuery; tra le più importanti troviamo i tool di:

- *Data QnA:* permette l'elaborazione del linguaggio naturale (Natural Language Processing, o NLP). Utile per creare, ad esempio, chatbot.
- *BigQuery GIS:* combina l'architettura serverless con il supporto nativo per l'analisi geospaziale.
- *Integrazione programmatica:* offre un'API REST che facilita l'integrazione di applicazioni mettendo a disposizione diverse librerie in Java, Python, Go, etc.

2.2 Qlik

Qlik (Figura 2.2) è un'impresa di software statunitense. Essa offre software per la Business Intelligence; I suoi prodotti sono:

- Qlik Sense;
- Qlik View;
- NPrinting.



Figura 2.2: Il Logo di Qlik

Questi software permettono lo sviluppo di dashboard completamente personalizzabili, in grado di fornire velocemente informazioni utili sui dati che si hanno a disposizione.

2.2.1 Qlik Sense

È la principale e più utilizzata soluzione di Qlik; essa permette ciò che in gergo viene chiamata la «Self-service Analytics», ovvero la possibilità per l'utente finale di esplorare i dati e di utilizzarli per creare delle visualizzazioni, quindi report e dashboard dinamiche, personalizzandole come preferisce.

Il prodotto finale generato da Qlik Sense è dato da raccolte di elementi riutilizzabili chiamate *App*. Queste sono composte da dati, fogli e racconti, come mostrato in Figura 2.3. L'*App*, quindi, è un'entità autonoma che include dati in un modello strutturato per l'analisi.

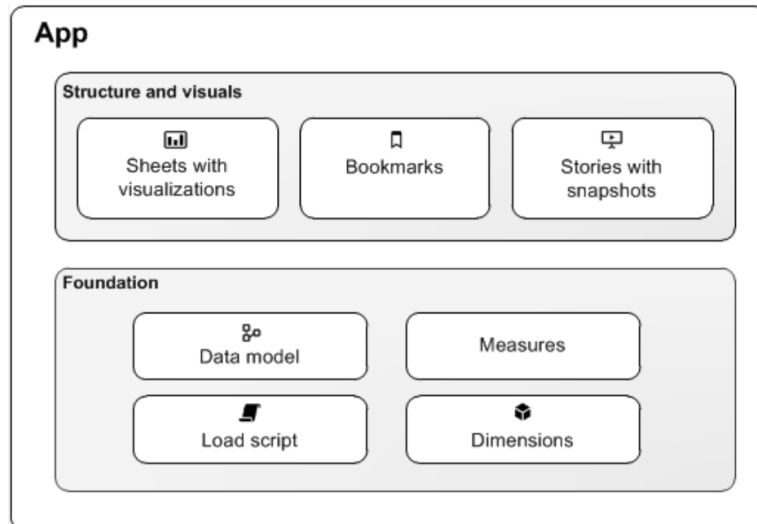


Figura 2.3: Struttura delle App in Qlik Sense

Una particolarità di Qlik Sense, rispetto ad altri competitor è quella di possedere un motore associativo al fine di evitare il processo «Ask, Wait, Answer». Questo processo deriva dai tool query-based, nei quali se un utente vuole indagare dei dati, avrebbe bisogno di chiedere aiuto agli sviluppatori, che conoscono bene l'SQL.

Il Motore associativo di Qlik Sense è progettato specificatamente per analisi di dati interattive; infatti, esso riesce a combinare diverse sorgenti di dati, deducendole in autonomia.

Analizziamo la terminologia definita da Qlik che ci risulterà utile a comprendere come questa tecnologia è stata usata per trattare il progetto relativo alla presente tesi. Nei successivi capitoli potremmo imbatterci in termini del tipo:

- *Script di caricamento dei dati.* È uno script che consente la connessione ad una sorgente dati, che sia derivante da database (ad esempio Google BigQuery), fogli Excel etc., nonché il recupero dei dati di interesse a partire da essa.
- *Modello dati.* I dati appena caricati vengono strutturati in un modello dei dati. Viene, comunque, data la possibilità di modificare lo script di caricamento, al fine di modificare il modello dei dati, in modo da ottenere la versione più appropriata per la propria applicazione.
- *Misure.* Queste non sono nient'altro che i calcoli e le espressioni che vengono mostrati nelle visualizzazioni. Sono frutto di un'espressione composta da funzioni di aggregazione, come la somma o il conteggio, ottenute combinando uno o più campi.
- *Dimensioni.* Esse rappresentano i campi da utilizzare nelle visualizzazioni e determinano il modo in cui i dati vengono raggruppati.

- *Fogli*. Sono il punto di raccolta dove i dati vengono visualizzati e mostrati attraverso grafici, tabelle o altri oggetti. Di solito esistono vari fogli, e ciascuno di questi presenta delle informazioni intrinsecamente pertinenti tra loro.
- *Segnalibri*. Permettono un collegamento a determinate selezioni su un particolare foglio.
- *Racconti*. Sono degli snapshot di visualizzazioni che possono essere combinati tra di loro, guidando l'utente a nuove informazioni, tenendo conto del momento in cui queste vengono derivate.

L'utente può creare delle visualizzazioni nelle App utilizzando dimensioni e misure, al fine di rispondere ad eventuali domande che egli potrebbe porsi.

All'avvio di Qlik Sense, viene visualizzato l'Hub (Figura 2.4). Qui sono contenute tutte le App a cui l'utente può accedere. Esistono due tipi possibili di Hub, ovvero:

- *Hub locale*;
- *Hub cloud*: qui gli utenti possono accedere alle app in lettura o in modifica, ma solo se in possesso di determinati autorizzazioni.

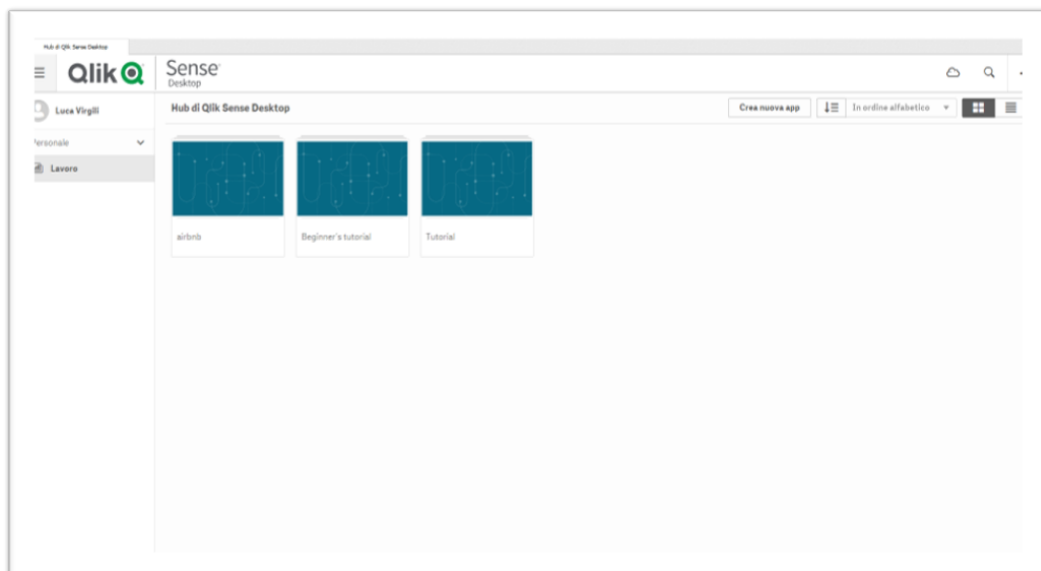


Figura 2.4: Hub di Qlik Sense

Una volta che l'App è stata creata, è necessario aggiungere i dati da esplorare e analizzare. Ci sono due metodi per fare questo, ovvero:

- *Gestione dati*: è possibile aggiungere dati dalle proprie sorgenti o da sorgenti esterne senza la necessità di scrittura di codice, sfruttando le potenzialità di associazione di Qlik.
- *Editor di caricamento dati*: è possibile realizzare un modello dati attraverso processi di ETL, utilizzando il linguaggio proprietario di Qlik, per il caricamento dei dati. Questo strumento è avanzato e potrebbe non essere adatto ai non esperti.

Per quanto riguarda la gestione dei dati abbiamo due modalità:

- *Associazioni*: è possibile modificare l'associazione esistente tra le tabelle.

- *Tabelle*: vengono mostrate sia le tabelle aggiunte grazie al tool di Qlik, sia quelle caricate tramite script. In queste tabelle è mostrato il nome, il numero dei campi e il nome della sorgente.

Inoltre, i dati possono essere aggiunti da varie sorgenti:

- *In-App*: l'App mette a disposizione sorgenti di dati tra cui poter scegliere, come file già allegati all'App.
- *Percorsi file*: esiste la possibilità di selezione tra i file che sono presenti in locale o in un'unità di rete.
- *Connessione dati*: consente di selezionare tra le connessioni dati a servizi cloud, come Google BigQuery.
- *Contenuto dati*: consente di scegliere dati provenienti da database pubblici o commerciali.

Qlik Sense esegue la profilatura dei dati caricati, al fine di assistere l'utente nella creazione delle associazioni delle tabelle. Quindi vengono mostrate le associazioni non valide e quelle potenzialmente valide, al fine di offrire assistenza per la selezione dei campi da associare in base all'analisi dei dati forniti.

Come mostrato in Figura 2.5, sono due le modalità di associare le tabelle:

- *Nella vista Associazioni di Gestione Dati* è possibile associare in base ai suggerimenti oppure creare associazioni personalizzate in base a uno o più campi;
- *Con l'opzione Associa nell'editor delle tabelle* è possibile associare per campi creando chiavi composte.

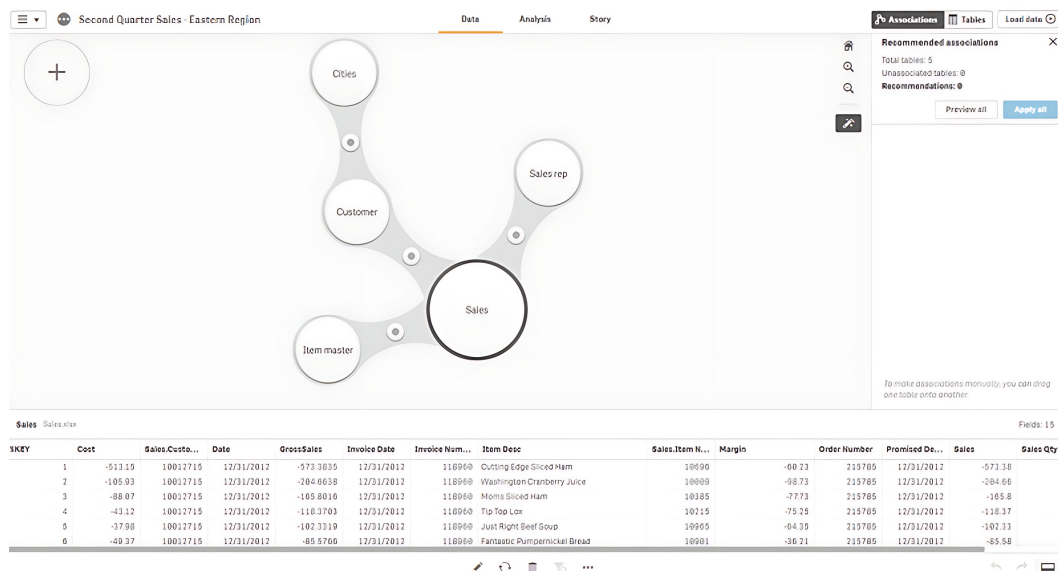


Figura 2.5: Le due modalità di associazione in Qlik Sense

Una volta scelta la modalità di associazione, Qlik Sense permette di visualizzare il modello dati (Figura 2.6), il quale fornirà una panoramica della struttura dei dati nell'App. Qui sarà possibile creare rapidamente dimensioni e misure.

Qlik Sense utilizza uno script di caricamento dei dati; questo è gestito nell'editor apposito e permette di connettersi ai dati nonché di recuperarli dalle diverse sorgenti.

The screenshot shows the Qlik Sense data model editor. It features several tables: 'Sales rep', 'Sales', 'Customer', 'Item master', and 'Cities'. Lines indicate associations between fields in different tables. For example, 'Sales rep Number' in 'Sales rep' is associated with 'Sales Rep Number' in 'Sales'. 'Customer Number' in 'Sales' is associated with 'Customer Number' in 'Customer'. 'Item Number' in 'Sales' is associated with 'Item Number' in 'Item master'. 'City Code' in 'Customer' is associated with 'City Code' in 'Cities'. A 'Preview' panel at the bottom shows a table of data with columns for 'Sales Rep Number', 'Customer Number', 'Item Number', 'XKEY', 'Cost', 'Date', and 'GrossSales'. The 'Customer Number' column is highlighted with a green circle 'C'. The 'Sales' table is highlighted with a green circle 'A', and the 'Customer' table is highlighted with a green circle 'B'.

Customer Number		Preview of data						
		Sales Rep Number	Customer Number	Item Number	XKEY	Cost	Date	GrossSales
Density	100%	180	10012226	10696	3428	-513.15	1/12/2012	-573.3835
Subset ratio	93.1%	180	10012226	10809	3429	-105.93	1/12/2012	-204.6638
Has duplicates	true	180	10012226	10385	3430	-88.07	1/12/2012	-165.8016
Total distinct values	683	180	10012226	10215	3431	-43.12	1/12/2012	-118.3703
Present distinct values	636	180	10012226	10965	3432	-37.98	1/12/2012	-102.3319
Non-null values	96466	180	10012226	10901	3433	-49.37	1/12/2012	-85.5766
Tags	Skey \$numeric \$integer	180	10012226	10681	3434	-45.81	1/12/2012	-68.4399

Figura 2.6: Preview del modello dati in Qlik Sense a seguito di associazioni

Nello script troveremo specificati i campi e le tabelle da caricare. Questo ci dà la possibilità di modificarne la struttura, ad esempio aggiungendo istruzioni ed espressioni.

Una volta caricati i dati, si può passare alla visualizzazione. Questa permette di presentare i dati in maniera grafica, ad esempio attraverso un grafico a barre piuttosto che una tabella, per consentire l'interpretazione e l'esplorazione da parte degli utenti. Per creare visualizzazioni efficienti sarebbe utile:

- *Comprendere le sorgenti dati.* Questo è possibile farlo rispondendo a domande del tipo: qual è il tipo di dati? Nominale, ordinale o un intervallo?; In che modo sono tra loro correlati parti diverse dei dati?; Cosa voglio comunicare?; Ci sono dimensioni o misure da riutilizzare in diverse visualizzazioni?
- *Creare visualizzazioni in base allo scopo.* Bisogna tenere a mente che ogni visualizzazione ha un suo scopo preciso. L'obiettivo deve essere, quindi, quello di creare visualizzazioni che permettano di raggiungere lo scopo nel minor tempo possibile. Quindi, la scelta di oggetti di visualizzazione risulta fondamentale.
- *Aggiornare le visualizzazioni.* Deve essere possibile modificare i filtri della nostra applicazione in modo che i dati mostrati al loro interno possano adattarsi a ciò che l'utente sta cercando.

Quando in Qlik Sense si creano le visualizzazioni sono disponibili le seguenti risorse:

- *Campi:* contengono i dati caricati, quindi uno o più valori corrispondenti alle colonne delle tabelle di un database. I campi vengono utilizzati per creare dimensioni e misure.
- *Misure:* sono create a partire da un campo, aggregando quest'ultimo con funzioni come somma, media etc. combinate con uno o più campi.
- *Dimensioni:* determinano il modo in cui i campi vengono raggruppati. Queste possono anche essere frutto di un'espressione di calcolo.

2.2.2 Qlik View

Qlik View, come specificato da Qlik stesso, «è la nostra piattaforma di analisi di prima generazione». View si basa sull'analisi guidata. Le applicazioni sono create dagli sviluppatori, i quali analizzano il modello dati, il layout, i grafici e le formule. Poi questi forniscono l'applicazione agli analyst, che hanno totale libertà di esplorare i dati ma non di modificarli, di solito.

Qlik View presenta sia la componente server che quella client e permette di creare file con tre estensioni diverse:

- `.qvw`: è un'applicazione che contiene dati, script e front-end;
- `.qvd`: contiene soltanto dati;
- `.qvs`: contiene soltanto script.

Per creare visualizzazioni in View, come in Sense, è necessario innanzitutto importare i dati nello script, tramite excel, `.qvd` o collegamenti con fonti dati esterne come potrebbe essere Google BigQuery.

Di fatto, View è un antecedente di Sense e permette, sostanzialmente, di ottenere risultati simili. In futuro tutte le applicazioni View verranno convertite in applicazioni Sense, scegliendo, poi, attraverso le apposite autorizzazioni, se dare o meno la possibilità all'utente finale di modificare le visualizzazioni.

2.2.3 Qlik NPrinting

NPrinting (Figura 2.7) è la soluzione di Qlik utile a realizzare la reportistica aziendale, consentendo di creare, gestire e condividere report in modo rapido, utilizzando i dati derivanti da Qlik Sense e Qlik View.

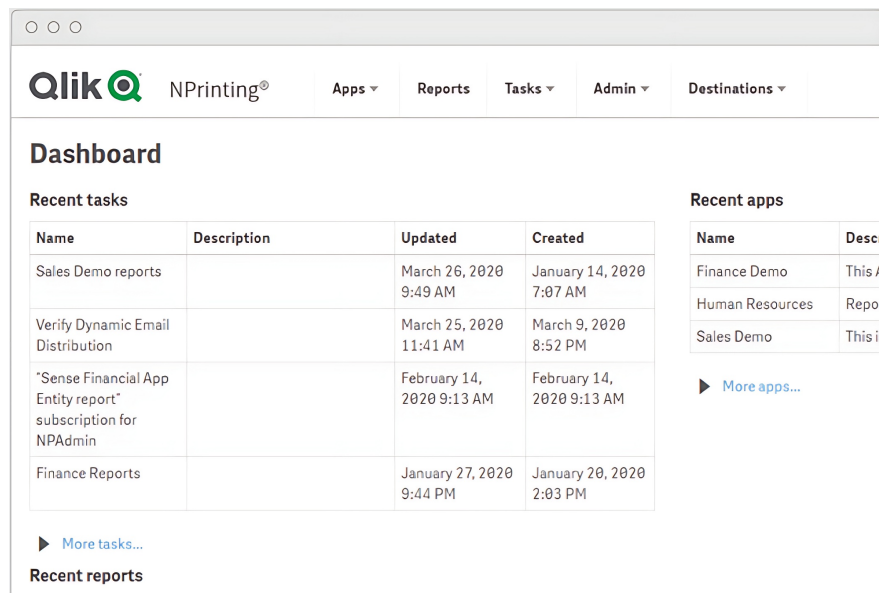


Figura 2.7: Dashboard di Qlik NPrinting

Le funzionalità chiave di NPrinting sono:

- *Creare report*. NPrinting consente di creare report di grande effetto in moltissimi formati, come PowerPoint, Excel e Word o, addirittura, PDF. Questo dà la possibilità di creare

report combinando diversi oggetti di Qlik Sense o View in un unico spazio, in modo da poter creare la propria analytics, senza necessità di rielaborare i dati. Esso permette, anche, di utilizzare la libreria PixelPerfect per la creazione di layout avanzati.

- *Gestire tutti i report in un'unica piattaforma.* Essendo Qlik NPrinting realizzato in scala risulta sicuro per le aziende. È possibile eseguire le attività simultaneamente, anche durante la creazione di molti report. Esso, inoltre, mette a disposizione dei filtri, basati sull'accesso combinato, per garantire la massima tranquillità.
- *Condividere direttamente i report.* Qlik NPrinting è in grado, di far arrivare i report alle persone giuste, attraverso i canali giusti e al momento desiderato, distribuendoli a orari personalizzati o solo al verificarsi di determinate condizioni. Si possono pianificare, generare e inviare i report tramite e-mail, web, all'Hub di Sense o attraverso altri modi. Inoltre, il tool, mette a disposizione il tool NewStand, che permette all'utente finale di utilizzare una modalità self-service per esplorare, visualizzare, scaricare e abbonarsi ai report mediante l'utilizzo di un portale sicuro.

In questo capitolo introduttivo sono state elencate le funzionalità più generali dei software utilizzati nel processo di sviluppo del progetto relativo alla presente tesi. Nei prossimi capitoli, quando necessario, seguiranno approfondimenti sui tool in modo da poter comprendere al meglio le tecniche utilizzate.

Descrizione del contesto di riferimento

In questo capitolo, analizzeremo il contesto di riferimento in cui la presente tesi è stata sviluppata. Si descriverà il progetto in generale, inquadrando l'architettura e come questa viene alimentata, descrivendo le principali entità presente nel modello dei dati sottostante.

3.1 Introduzione agli obiettivi di progetto

Questa tesi ha come obiettivo quello di descrivere in che modo abbiamo progettato e implementato, attraverso l'utilizzo di Google BigQuery e Qlik, un sistema di KPI (Key Performance Indicator), ovvero degli "indicatori chiave di prestazione", quindi, un insieme di indici che descrivono quantitativamente l'andamento di un processo aziendale.

Questi indici sono stati generati a supporto di una società di gestione del risparmio che, per motivi di privacy, nel contesto della presente tesi, verrà indicata come "Society".

Gli obiettivi che ci siamo posti per tale progetto sono:

- potenziare, attraverso l'utilizzo della Business Intelligence, le capacità analitiche dell'azienda, incrementandone il patrimonio informativo;
- razionalizzare il patrimonio informativo in ottica di definizione di una roadmap strategica di medio periodo per l'evoluzione del framework e il miglioramento della qualità del dato;
- realizzazione di due indicatori di performance: raccolta requisiti, analisi funzionale, supporto al test e go-live con coordinamento degli stakeholder;
- fornire la documentazione e gli accessi agli strumenti informativi aziendali necessari per la fase di analisi del patrimonio informativo.

Nelle prossime sezioni analizzeremo l'architettura del progetto e, successivamente, le principali fonti del dato e le entità che descrivono il modello dati di "Society".

3.2 L'Architettura della BI di progetto

L'architettura di Business Intelligence di "Society" è strutturata su 3 livelli. Questi formano una catena di elaborazione del dato, come mostrato in Figura 3.1.

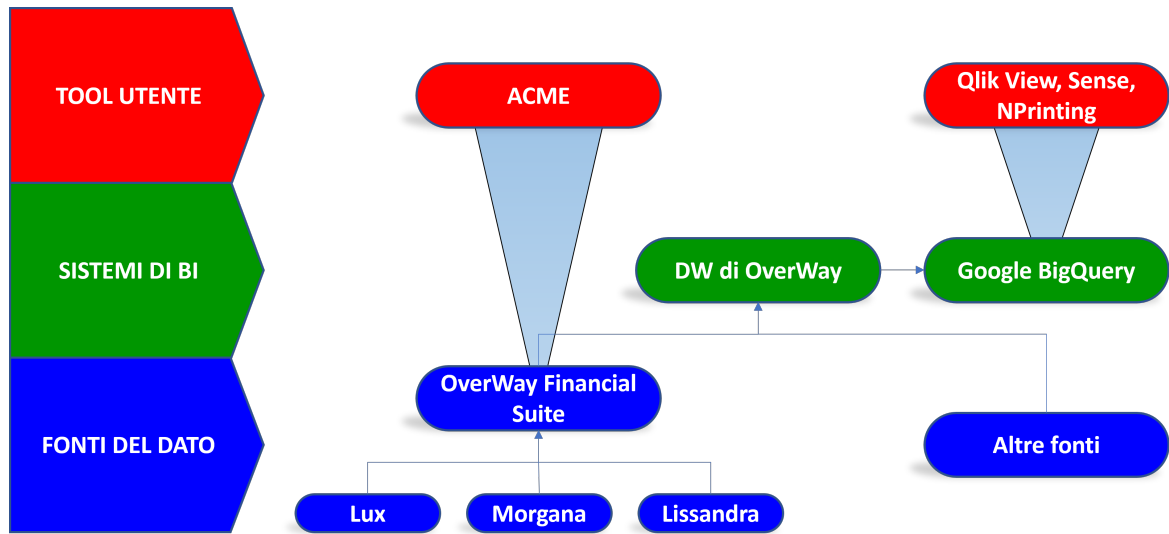


Figura 3.1: Architettura del sistema di BI di Society

Per questioni di privacy commerciale, i nomi di database e i nomi dei moduli non possono essere indicati. Pertanto, questi verranno indicati nel corso della nostra tesi con dei nomi fittizi.

I livelli sono:

- *fonti del dato*: Lux, Morgana, Lissandra;
- *sistemi di Business Intelligence*: Data Warehouse di OverWay, ovvero, la società che si occupa di software in ambito finanziario, e Google BigQuery;
- *tool Utente*: ACME, Qlik View, Qlik Sense e Qlik NPrinting.

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

- *Lux*: fonte della parte amministrativa e parte relativa ai fondi;
- *Morgana*: fonte di gestioni patrimoniali, conti correnti e altre attività;
- *Lissandra*: fonte di dati relativi a banche partnership e altri enti partner.

I dati che provengono da queste tre entità vengono immessi in una Financial Suite, la quale svolge il ruolo di database principale aggregando e conciliando le informazioni dei sistemi sottostanti. Questo dataset alimenta, in modo principale, il Data Warehouse di OverWay, e rappresenta, anche, il sistema di interfaccia con l'applicativo ACME. Inoltre, altre fonti esterne alimentano il DW di OverWay.

3.3 Sorgenti di alimentazione del progetto

Il DW di OverWay è la principale fonte alimentante di BigQuery; da questo, infatti, deriviamo il dataset maestro, il quale eredita le logiche dal Data Warehouse e contiene la maggior parte delle tabelle presenti.

In maniera dettagliata, possiamo dire che l'insieme dei dati parte dal Data Warehouse per arrivare al Google Cloud Storage il quale, a sua volta, alimenta BigQuery. I dati, poi, verranno memorizzati in file di formato `.csv` in due differenti repository.

Di caricare i dati in BigQuery, se ne occupa un applicativo esterno, il quale elabora i file ed esegue l'ETL specifico. Tali procedure, suddivise in blocchi, vengono eseguite a cascata e hanno come scopo l'aggregazione dei dati presenti nei file in formato .csv e il confronto con dei file, detti di quadratura. Questi ultimi permettono di eseguire una verifica dell'integrità dei dati e consentono l'elaborazione in sequenza.

A livello di organizzazione, in BigQuery, i dati sono contenuti all'interno di un unico progetto. Quest'ultimo è organizzato in dataset che rappresentano le principali categorie degli stream di progetto.

Al momento corrente, il linguaggio in uso su BigQuery per il nostro progetto è il Legacy SQL.

Le tabelle all'interno del dataset, relative a dimensioni o fatti, possono essere distinte in entità. Le entità principali sono:

- *Agente*: contiene le informazioni anagrafiche degli agenti finanziari che operano per "Society". Le tabelle a cui questa entità è collegata sono mostrate in Figura 3.2.

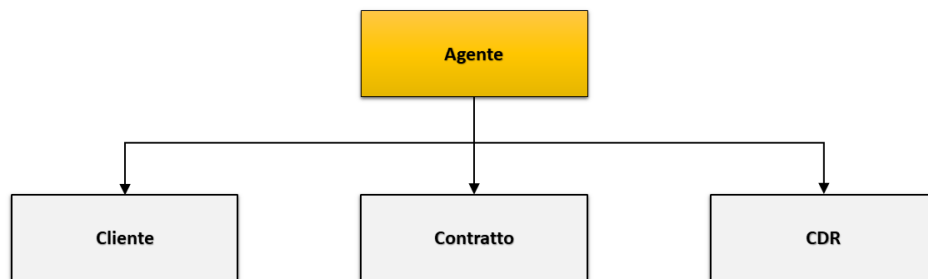


Figura 3.2: L'entità Agente

- *Calendario*: contiene le informazioni relative al calendario solare.
- *Cliente*: contiene le informazioni anagrafiche dei clienti. Le tabelle a cui questa entità è collegata sono mostrate in Figura 3.3.

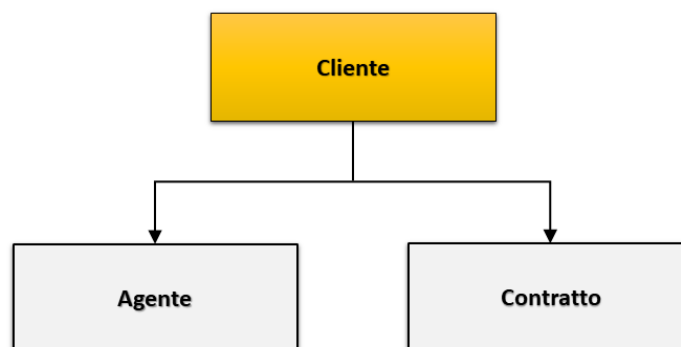


Figura 3.3: L'entità Cliente

- *Commissioni*: contiene le informazioni relative alle commissioni che vengono remunerate ai promotori finanziari. Le tabelle a cui questa entità è collegata sono mostrate in Figura 3.4.

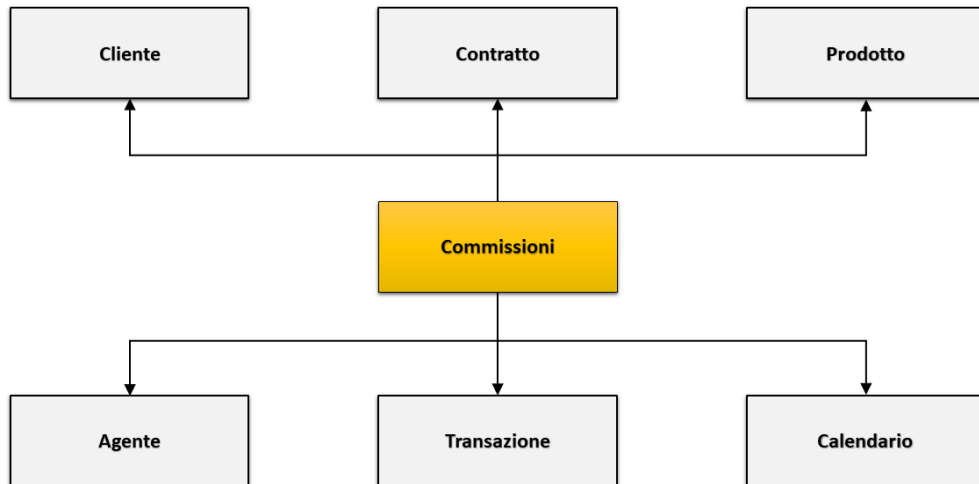


Figura 3.4: L'entità Commissioni

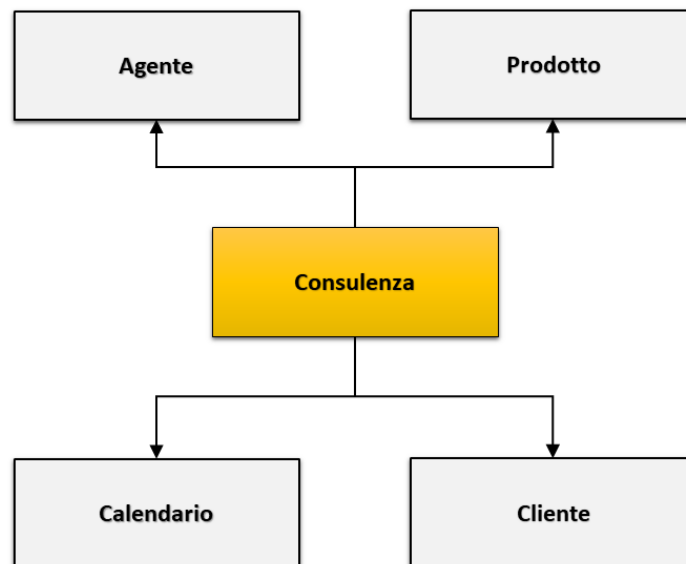


Figura 3.5: L'entità Consulenza

- *Consulenza*: contiene le informazioni relative ai servizi di consulenza. Le tabelle a cui questa entità è collegata sono mostrate in Figura 3.5.
- *Contratto*: contiene le informazioni relative ai contratti con i clienti. Le tabelle a cui questa entità è collegata sono mostrate in Figura 3.6.
- *Incentivi*: contiene le informazioni relative agli incentivi dei promotori finanziari. Le tabelle a cui questa entità è collegata sono mostrate in Figura 3.7.
- *Portafoglio*: contiene le informazioni relative portafoglio del cliente. Le tabelle a cui questa entità è collegata sono mostrate in Figura 3.8.
- *Prodotto*: contiene le informazioni relative ai prodotti offerti da "Society".
- *Transazione*: contiene le informazioni relative ai movimenti dei clienti. Le tabelle a cui questa entità è collegata sono mostrate in Figura 3.9.

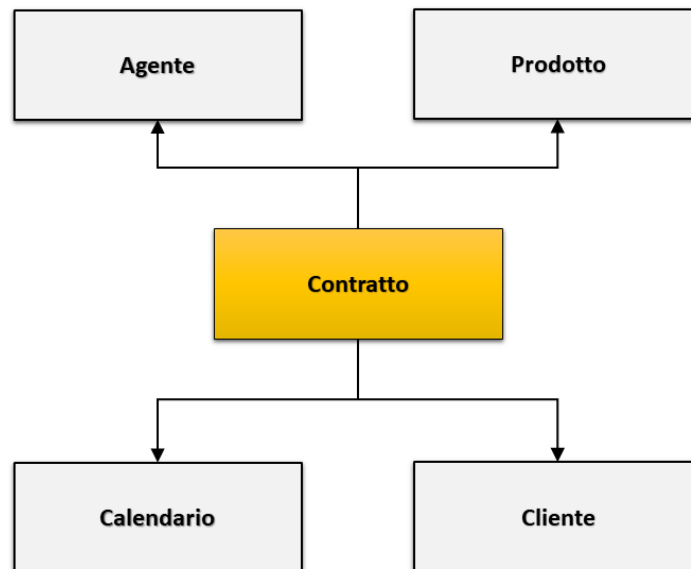


Figura 3.6: L'entità Contratto

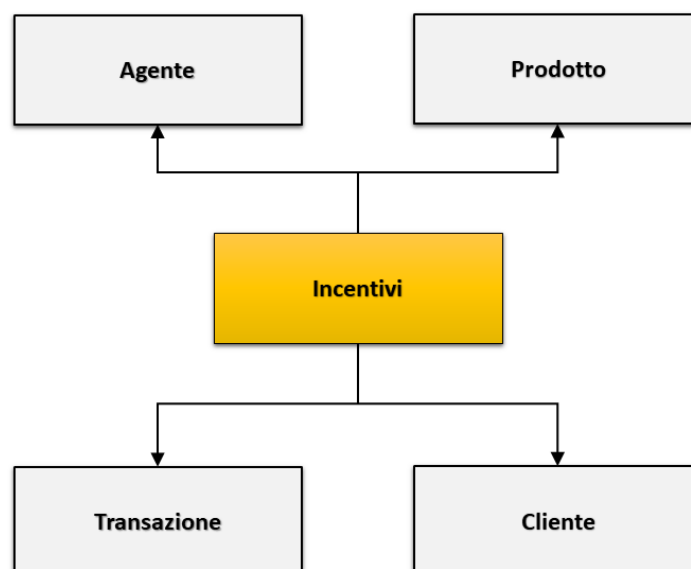


Figura 3.7: L'entità Incentivi

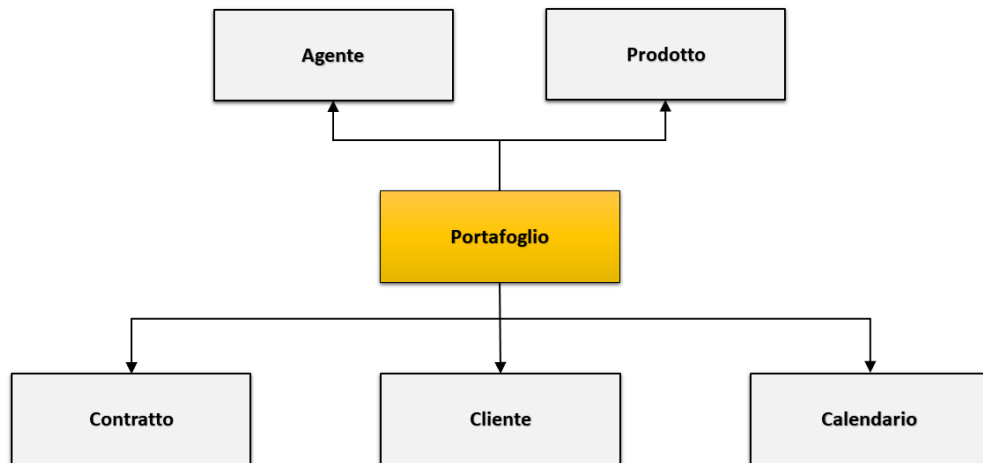


Figura 3.8: L'entità Portafoglio

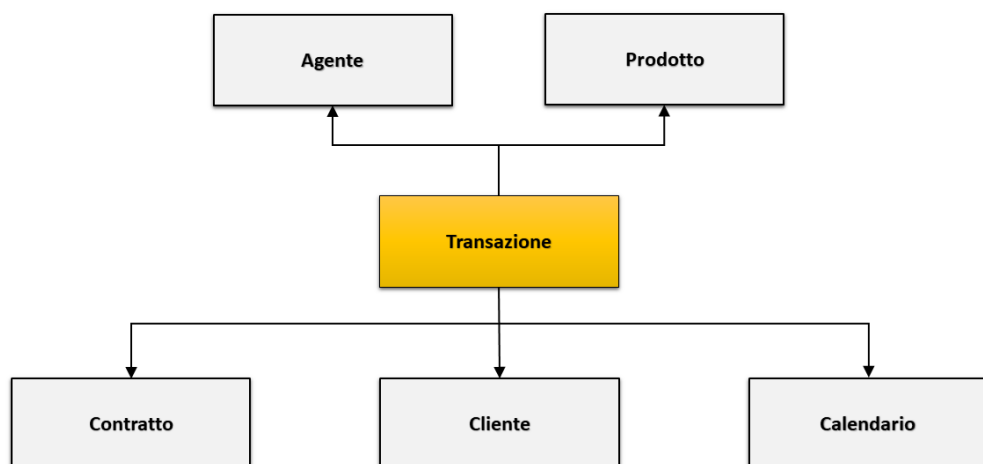


Figura 3.9: L'entità Transazione

In questo capitolo verrà presentato il processo di analisi e raccolta dei requisiti che porteranno alla creazione dei KPI utilizzati dalla "Society" al fine di descrivere l'andamento del processo aziendale trattato nella presente tesi.

4.1 Proposta e obiettivi finali

Questa tesi nasce, dalla necessità da parte del gruppo "Society", di introdurre due nuovi indicatori, i quali verranno indicati, per motivi di riservatezza aziendale, con dei nomi fittizi. Questi nomi sono:

- *Alfa*;
- *Beta*.

Questi due KPI verranno analizzati successivamente nello specifico. Per il momento possiamo affermare che questi indici di performance sono particolarmente rilevanti per l'azienda, e la loro introduzione avrà impatti su applicazioni aziendali, base di dati e sistemi BI a supporto.

Come gruppo di consulenza aziendale abbiamo deciso di avviare questa iniziativa orientata principalmente a raggiungere i seguenti obiettivi:

- *Ambito Alfa/Beta*: realizzazione di due indicatori di performance.
- *Ambito BI*: razionalizzazione del patrimonio informativo enterprise con l'obiettivo di definire una roadmap strategica di medio periodo per l'evoluzione del framework informativo e di Business Intelligence e il miglioramento della qualità del dato.

Tale iniziativa delinea la soluzione individuata dal gruppo di consulenza per rispondere agli obiettivi del gruppo "Society" appena citati. Essa prevede due stream progettuali paralleli, ovvero:

1. *Stream Alfa/Beta*: divide i compiti in raccolta requisiti, analisi funzionale, supporto al test, go-live e coordinamento degli stakeholder.
2. *Stream BI*: si occupa della realizzazione del patrimonio informativo enterprise con un approccio per use case, nonché della definizione della nuova architettura di BI.

A loro volta, tali stream sono organizzati in tre distinte fasi progettuali, ovvero:

- a. *Assessment*: è la fase di studio preliminare per la perimetrazione dell'ambito, dei principali impatti IT e dei vincoli.
- b. *Design*: è la fase dedicata al design tecnico/funzionale delle soluzioni da apportare.
- c. *Execution*: questa fase è dedicata alla realizzazione delle soluzioni, al test e al go live delle stesse.

Come team di consulenza abbiamo supportato il cliente nelle varie attività di stream sopra riportate, al fine di garantire la coerenza tra le reali esigenze interne e l'evoluzione di funzionalità svolgendo attività di:

- Management e governance del progetto (stream *Alfa/Beta*).
- Stesura dei requisiti di business e analisi funzionali propedeutici alla realizzazione degli indici *Alfa/Beta* (stream *Alfa/Beta*).
- Coordinamento delle attività di formalizzazione e raccolta dei requisiti di business, con supporto alla formalizzazione dei casi di test per le fasi di User Acceptance Testing ("UAT") (stream *Alfa/Beta*).
- Reporting periodico sull'avanzamento progettuale (SAL, Tracking Punti Aperti/Issue, Verbali, work-shop, etc.) (stream *Alfa/Beta*).
- Coordinamento della fase di UAT e Go-Live (stream *Alfa/Beta*).
- Definizione degli use case relativi alle esigenze informative (stream BI).
- Identificazione del perimetro informativo (stream BI).
- Disegno del modello dati (stream BI).
- Disegno dell'architettura logica del framework BI (stream BI).
- Definizione della Roadmap (stream BI).

4.1.1 Architettura di riferimento

Per quanto riguarda lo stream di creazione degli indici *Alfa/Beta*, l'architettura usata come riferimento è stata quella mostrata in Figura 4.1. Quest'ultima indica come il calcolo dei due indici avvenga in arie differenti, in primo luogo, attraverso informazioni provenienti da varie forme, ad esempio fonti interne o esterne, le quali si occupano delle fasi di ETL sui dati, fornendo a Google BigQuery il compito della sola applicazione della formula definita per il calcolo degli indici *Alfa/Beta* e mostrandoli negli appositi sistemi software proprietari della "Society".

Infine, questi indici verranno utilizzati sul sistema Qlik sia per fornire informazioni sulla tempistica della richiesta e sia per tener traccia dello storico, in modo da poter calcolare altri indici, i quali permetteranno, nel medio termine, di effettuare anche analisi di tipo predittivo.

Nella prossima sottosezione analizzeremo l'approccio che abbiamo deciso di adottare, relativamente alle attività suddivise per i due stream, descrivendone le fasi progettuali.

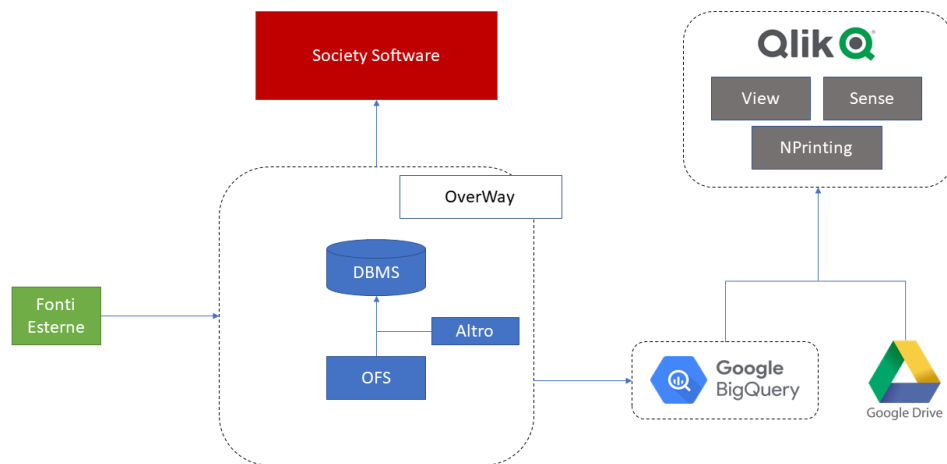


Figura 4.1: Architettura di riferimento per la costruzione dei KPI

4.1.2 Approccio adattato per lo stream BI

Lo stream BI prevede due fasi ed un approccio ibrido di tipo top down, per le attività relative agli indicatori di performance *Alfa/Beta*, e di tipo bottom up, per il design del modello dei dati, dell'architettura BI e la roadmap. Ciò viene meglio mostrato in Figura 4.2. In essa identifichiamo due blocchi, ovvero Design e Assessment, che distinguono i loro task come di seguito specificato:

- Top down Assessment:
 - analisi degli impatti in relazione allo stream *Alfa/Beta*.
- Design; è la fase dedicata al design delle soluzioni da apportare in termini di:
 - design funzionale *Alfa/Beta*;
 - modello dei dati concettuale a livello enterprise;
 - architettura logica target del framework BI a livello enterprise;
 - roadmap evolutiva.
- Bottom up Assessment, essa prevede:
 - la definizione dello scope dello stream BI;
 - lo studio dell'architettura logica BI;
 - l'identificazione del patrimonio informativo in termini di flussi, reportistica e macro entità dei dati.

4.1.3 Fasi di progetto

Inoltre, dividendo per fase progettuale, riportiamo le principali attività nonché i risultati di queste ultime.

Per quanto riguarda la fase di Assessment, le attività saranno le seguenti:

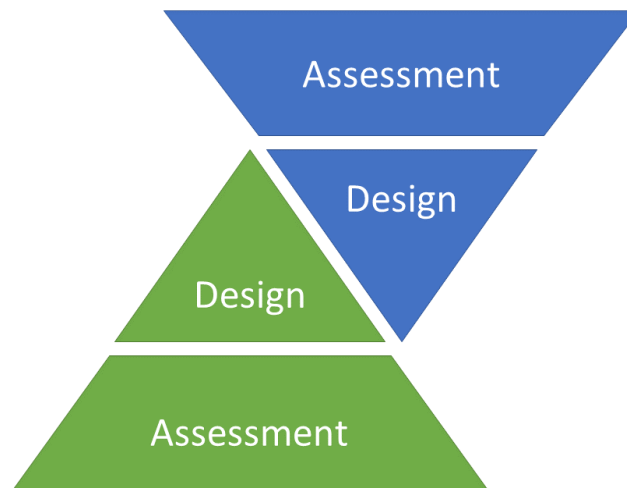


Figura 4.2: Approccio ibrido adottato per le attività

- Interviste strutturate con i colleghi della società di consulenza già dapprima impegnate con “Society” al fine di identificare l’architettura del framework BI allo stato dell’arte e i principali flussi alimentanti.
- Acquisizione della documentazione disponibile utile per l’esecuzione dell’assessment del patrimonio informativo.
- Identificazione del perimetro informativo.

Tali attività produrranno come output:

- la formalizzazione dell’architettura del framework BI allo stato dell’arte;
- l’identificazione e la definizione del patrimonio informativo allo stato dell’arte, in termini di report, macro entità e flussi alimentanti.

Per quanto riguarda la fase di Design, le attività saranno le seguenti:

- definizione del glossario dei dati che uniformi la terminologia di business a livello enterprise;
- interviste con i principali stakeholder di business per identificare need/pain relativi al patrimonio informativo, funzionalità degli applicativi di BI, analisi dei dati, etc;
- design del modello concettuale dei dati, con evidenza delle entità e delle relazioni tra di esse;
- incontri con OverWay per identificare le sorgenti dati e la descrizione dei campi in relazione alle informazioni presenti nel framework di BI;
- disegno logico della soluzione tecnologica target in termini di framework architetturale ed infrastrutturale;
- definizione delle priorità e della timeline progettuale;
- effort economico necessario per raggiungere il disegno target.

Tali attività produrranno come output:

- il macro modello dei dati concettuale relativo al perimetro informativo;
- l'enterprise data glossary;
- il disegno dell'architettura logica del framework di BI per intraprendere le iniziative tattico/strategiche;
- il documento di macro per la stima dei costi da sostenere.

4.2 Requisiti funzionali

L'iniziativa *Alfa/Beta* è stata avviata per dare seguito all'esigenza del gruppo "Society" di introdurre due nuovi indicatori per un maggior controllo sull'andamento della gestione, sia in termini di raccolta sia per le masse. L'iniziativa servirà, anche, per il calcolo degli incentivi da somministrare alla rete.

I due nuovi indicatori rappresentano un nuovo metodo di calcolo del portafoglio e della raccolta, e saranno implementati nei sistemi. Inoltre, questi KPI rappresenteranno la base di calcolo per la determinazione di altri strumenti partecipativi.

In questa sezione dettaglieremo i requisiti funzionali riguardanti i dati propedeutici al calcolo dei due indicatori di performance rispettivamente.

Alla base della buona riuscita di questo progetto c'è il controllo sui requisiti, i quali verranno elencati a breve. Qualora anche uno solo dei requisiti non restituisse come esito OK durante il processo di calcolo, quest'ultimo verrebbe bloccato e si dovrà procedere alla rigenerazione del dato. Soltanto quando il controllo degli esiti sui requisiti avrà un risultato positivo si procederà alla memorizzazione o, in gergo, scrittura, dell'informazione sul DW di Google BigQuery.

4.2.1 Dati propedeutici al calcolo

La soluzione architetturale per l'esecuzione del calcolo degli indici *Alfa/Beta* identifica in BigQuery il sistema target di elaborazione. Di seguito si riportano i controlli di consistenza del dato di input e le regole per implementare il calcolo degli indicatori. Si evidenzia, al solo fine esplicativo, che la nomenclatura dei dati in input segue le regole del mercato in cui la "Society" opera, ovvero, il mercato finanziario. Tuttavia, al fine di preservarne la riservatezza commerciale, alcuni nomi dei dati non possono essere rappresentati, in tali casi verranno utilizzati dei nomi fittizi.

Il controllo di consistenza del dato prevede l'implementazione di un set di controlli sui dati che sono necessari per eseguire il calcolo degli indicatori *Alfa/Beta*. Nello specifico, i controlli possono essere classificati come di seguito riportato:

- 1 Si evidenzia una mancanza del derivato associato ad un determinato prodotto sottoscritto dal cliente; il controllo restituirà KO come esito. In caso contrario, il risultato restituirà come esito OK.
- 2 Si evidenzia una mancanza della percentuale di guadagno sul catalogo dei prodotti; il controllo restituirà KO come esito. In caso contrario, risultato restituirà come esito OK.
- 3 Si evidenzia una mancanza dell'informazione sulle risorse patrimoniali facciali e sottostanti; il controllo restituirà KO come esito. In caso contrario, il risultato restituirà come esito OK.

- 4 Si evidenzia una mancanza delle tasse e degli sconti applicati ai prodotti acquistati; il controllo restituirà KO come esito. In caso contrario, il risultato restituirà come esito OK.
- 5 Si evidenzia una mancanza delle tasse applicate a determinati clienti. Il controllo restituirà KO come esito. In caso contrario, il risultato restituirà come esito OK.
- 6 Si evidenzia una mancanza dell'aliquota da applicare ai prodotti di consulenza; il controllo restituirà KO come esito. In caso contrario, il risultato restituirà come esito OK.
- 7 Si evidenzia una incoerenza tra i dati dei prodotti derivati; il controllo restituirà KO come esito. In caso contrario, il risultato restituirà come esito OK.

Per tutti i casi descritti, qualora si verificassero le suddette condizioni, BigQuery produrrà un report che ha come scopo quello di informare il team per attuare le modifiche richieste. Tale meccanismo verrà esplicito in maniera approfondita nei successivi capitoli, riguardanti progettazione e implementazione.

In questo capitolo descriveremo il processo di progettazione degli indicatori di performance di interesse per la presente tesi. Verranno mostrati i diagrammi di flusso che riguardano il processo di coerenza dei requisiti richiesti. Tali diagrammi rappresenteranno la base, poi, delle scelte implementative che analizzeremo nel capitolo successivo.

5.1 Procedura di ripristino dati

Per tutti i casi descritti nel capitolo precedente, qualora si verificassero le corrispondenti condizioni, BigQuery produrrà un report con cadenza pari a due volte per settimana (sempre a valle dell'ultimo giorno del mese) e a $t+5$ e $t+10$ del mese. È prevista, inoltre, in aggiunta alla schedulazione sopra citata, la possibilità di lanciare l'elaborazione su richiesta in modalità manuale, allo scopo di contenere anche i casi anomali dell'ultimo giorno del mese e al fine di dare evidenza, ai gruppi di lavoro coinvolti, di quali siano i record per cui sono state riscontrate le anomalie a partire dal primo giorno del solo mese corrente.

Il report conterrà le seguenti informazioni:

- codice e nome del cliente;
- codice e nome del promotore;
- codice del contratto;
- classificazione del prodotto;
- descrizione del prodotto;
- controvalore, ovvero il valore di mercato di una posizione finanziaria, determinato dall'attrattività dei titoli, dalle ultime performance e dalle aspettative future;
- codice del dato che ha generato l'errore;
- codice della causale.

Al fine di ripristinare i dati che hanno restituito come esito KO è stata sviluppata una procedura di "verifica e rettifica dei dati", la quale sarà eseguita prima della procedura di ricalcolo dei due indici *Alfa/Beta* di fine mese. Essa avrà come obiettivo quello di ricalcolare i

KPI giornalieri per tutti i casi anomali in cui il calcolo è stato rettificato attraverso l'utilizzo dei valori di default.

Questa attività potrà avere effetto soltanto nel caso in cui i dati reali mancanti siano stati forniti a BigQuery a monte della data di esecuzione della stessa procedura. Nei rimanenti casi il valore dei KPI (*Alfa/Beta*) rimarrà approssimato sulla base dei dati di default applicati.

Prima di lanciare la procedura di ricalcolo mensile (il quale avverrebbe tra il 15 e il 20 del mese successivo a quello di chiusura) si dovrà prevedere una procedura di verifica delle posizioni da rettificare (ovvero delle casistiche in cui si è applicato un valore di default) e si dovranno ricalcolare tutte quelle per cui il dato reale è adesso disponibile. Nello specifico la procedura dovrà occuparsi:

1. Dell'identificazione dei casi anomali di assenza del dato per cui è stata adottato un valore di default (ovvero lancio procedura per produzione report BI).
2. Dell'identificazione dei casi anomali per cui sono presenti i dati reali da sostituire ai dati di default.
3. Del ricalcolo dei KPI giornalieri per i casi al punto 2 per tutta la finestra temporale del mese per cui è stato applicato il valore di default e della sostituzione dei KPI calcolati con valore di default con i valori dei KPI ricalcolati.
4. Della produzione del report "post elaborazione rettifiche" contenente due macro categorie di dati, ovvero:
 - a. La categoria dei dati ricalcolati (corretti): essa comprende quei record per cui i KPI sono stati ricalcolati rispetto ai dati reali (dando evidenza del dato di default applicato, dei nuovi indicatori e del dato reale sostitutivo del dato di default).
 - b. La categoria dei dati non ricalcolati (approssimativi): essa comprende quei record per cui i KPI non sono stati ricalcolati per assenza del dato reale.

Tale report sarà disponibile su una dashboard, realizzata in Qlik, che riguarderà il controllo della qualità al fine di inizializzare la procedura di ricalcolo mensile. Il report sarà inviato al reparto Controllo di gestione o al reparto Direzione Commerciale.

A titolo esemplificativo, la dashboard dovrà dare evidenza dello stato dell'arte della qualità dei dati mostrando le posizioni che sono state ricalcolate e quelle che non sono state ricalcolate (per le quali si applica il dato di default), ovvero mostrare i dati del report sopra descritto. Inoltre, il cruscotto dovrà prevedere un apposito comando per approvare l'esecuzione della procedura di ricalcolo del fine mese (il cui comando è a disposizione del reparto Controllo di Gestione e di quello responsabile della Business Strategy) o, eventualmente, rilanciare la procedura o uscire dalla dashboard.

Una volta presa visione dei dati ricalcolati e di quelli non ricalcolati, e dopo aver estratto eventualmente il report, l'utente potrà:

- Rilanciare la procedura "verifica e rettifica dati" in questione con nuova produzione del report post elaborazione rettifiche.
- Accettare lo stato dell'arte dei dati ricalcolati e di quelli non ricalcolati. Questa opzione abilita la possibilità di eseguire la procedura di ricalcolo di fine mese (il cui comando per l'esecuzione resta a disposizione del reparto Controllo di Gestione e del reparto Business Strategy).
- Uscire dalla dashboard senza accettare lo stato dell'arte dei dati (in tal caso non potrà essere attivato il comando per l'esecuzione della procedura di ricalcolo).

- Attivare il comando per l'esecuzione della procedura di ricalcolo mensile (subordinato al punto due di cui sopra).

Una volta dato l'ok alla procedura di ricalcolo mensile, gli indici *Alfa/Beta* verranno calcolati su base mensile come da regole previste. Il funzionamento appena descritto è mostrato in Figura 5.1. Una volta che la procedura di ricalcolo dei KPI mensili è stata eseguita, i dati si considerano consolidati e non più modificabili.

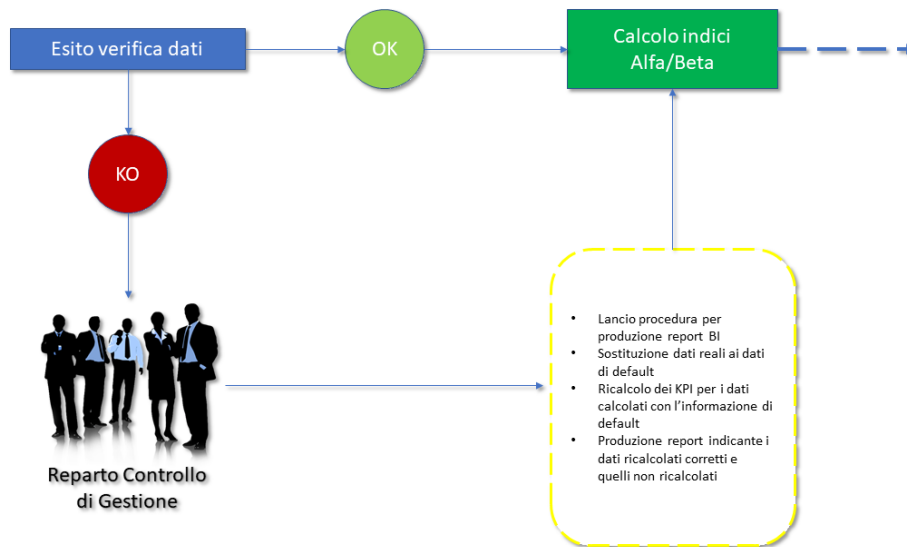


Figura 5.1: Architettura riguardante il ripristino dei dati

Inoltre, nel caso in cui dei record non sono stati oggetto di rettifica per mancanza del dato reale ed è stata già eseguita la procedura di chiusura mese, nel mese successivo sarà possibile inserire manualmente la data posizione, con il valore delta *Alfa/Beta* (tale valore è manualmente determinato dal gruppo Controllo di Gestione ed il sistema carica il dato rettificato *Alfa/Beta*), il quale compensa il valore approssimato calcolato nel mese precedentemente chiuso (alla data di competenza del mese in cui si propone di effettuare la rettifica).

5.2 Regole di calcolo degli indici Alfa e Beta

Il sistema di calcolo che provvederà all'esecuzione degli indicatori *Alfa/Beta* è BigQuery (soluzione centralizzata al fine di evitare che il calcolo venga effettuato su più sistemi a disposizione della compagnia). Alla luce di quanto anticipato nella sezione precedente, le regole di calcolo dei principali KPI (Figura 5.2) oggetto dell'analisi sono le seguenti:

- *%Alfa*: è un valore percentuale che rappresenta il ricavo annualizzato al netto dei "rebates" pagati alla rete, dei costi generali in Italia e delle tasse generate; è censita sul catalogo prodotti (tabella dei dati).

* *Periodicità di aggiornamento/calcolo*: la tabella viene aggiornata a evento ogni qualvolta si crea un nuovo prodotto o si verifica una variazione sostanziale, oppure se il prodotto è oggetto di rivisitazione della gamma d'offerta

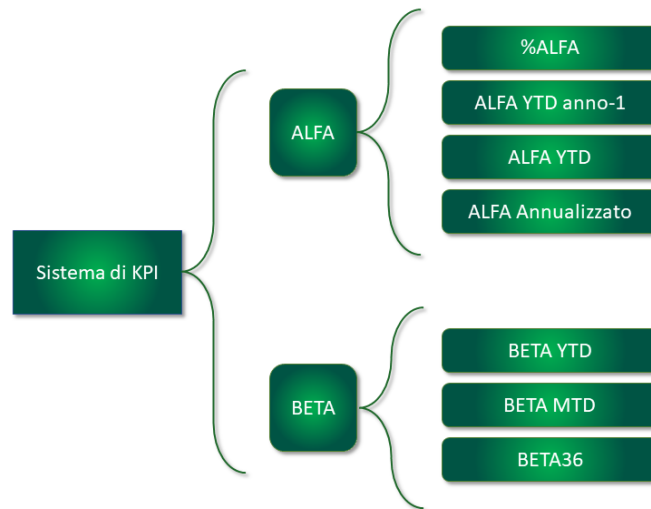


Figura 5.2: Sistema di KPI Alfa/Beta

- *Alfa*: è calcolato a partire dal valore del portafoglio giornaliero/medio mensile dei prodotti collocabili da parte del gruppo “Society” sul mercato. Questo indicatore sarà calcolato su quattro diversi orizzonti temporali.
 - * *Periodicità di aggiornamento/calcolo*: il calcolo è effettuato a livello giornaliero provvisorio, con il successivo ricalcolo definitivo su patrimonio medio mensile con i dati consolidati di fine mese.
 - a. *Alfa YTD anno-1*: dove YTD sta per Year To Date. È calcolato a partire dal valore storicizzato dell’*Alfa* dei mesi dell’anno precedente e il calcolo del PTF medio viene effettuato come sommatoria dei giorni NAV (Net Asset Value, deriva dall’ambito finanziario) utilizzando un calendario standard da applicare a tutti i prodotti.
 - * *Periodicità di aggiornamento/calcolo*: è il calcolo dei valori *Alfa* dell’anno precedente per i mesi chiusi meno uno e della ripartizione dei giorni NAV.
 - b. *Alfa YTD*: è la somma dell’*Alfa* al mese chiuso e l’*Alfa* MTD giornaliero, dove MTD sta per Month To Date.
 - * *Periodicità di aggiornamento/calcolo*: Calcolo giornaliero provvisorio con ricalcolo definitivo su patrimonio medio mensile con i dati consolidati di fine mese usando i dati dell’ultimo giorno del mese.
 - c. *Alfa Annualizzato*: è la somma dell’*Alfa* al fine mese precedente fino al mese in corso, moltiplicato per il numero dei mesi futuri mancanti
 - * *Periodicità di aggiornamento/calcolo*: è un calcolo giornaliero provvisorio, con ricalcolo definitivo su patrimonio medio mensile con i dati consolidati di fine mese usando i dati dell’ultimo giorno del mese.
- *Beta YTD*: è il dato del movimento senza effetto di mercato moltiplicato per la percentuale *Alfa* del singolo prodotto presente sul catalogo prodotti.
 - * *Periodicità di aggiornamento/calcolo*: è un calcolo giornaliero provvisorio

- *Beta MTD*: è il dato della sommatoria del movimento senza effetto mercato dal primo giorno del mese in oggetto fino alla data di riferimento, moltiplicato per la percentuale *Alfa* del singolo prodotto presente sul catalogo prodotti.
 - * *Periodicità di aggiornamento/calcolo*: è un calcolo giornaliero provvisorio, con ricalcolo definitivo sul valore della raccolta mensile sui giorni del mese fino alla data di riferimento.
- *Beta36*: rappresenta il valore per il calcolo degli incentivi che dovranno essere pagati ai consulenti.
 - * *Periodicità di aggiornamento/calcolo*: è un calcolo giornaliero provvisorio sui giorni del mese fino alla data di riferimento.

5.3 Architettura degli indici Alfa e Beta

Al fine di calcolare il valore degli indici *Alfa/Beta* è chiaro che bisogna accedere ad un set di informazioni rilevanti, le quali risiedono sul Data Warehouse aziendale; quest'ultimo si basa, come detto svariate volte, su Google BigQuery.

Per comprendere meglio i passi che porteranno al calcolo dei suddetti KPI, in questa sezione descriveremo l'architettura delle tabelle più importanti utilizzate, localizzate su BigQuery. Vedremo, come queste, interagendo tra di loro, permetteranno di arrivare al risultato finale della anche, presente tesi.

Come già accaduto precedentemente, al fine di preservare la riservatezza commerciale dell'azienda per cui questo elaborato viene redatto, i nomi delle tabelle non saranno fedeli ma subiranno delle variazioni per renderli fittizi.

5.3.1 Architettura indice Alfa Giornaliero

L'architettura per la generazione dell'indice *Alfa Giornaliero*, come mostrato in Figura 5.3, si divide in tre fasi, ovvero:

- prenderemo dalla tabella Portafoglio sempre l'ultima fotografia del giorno;
- assegneremo la categoria dell'*Alfa* e della relativa *%Alfa* per ogni tipologia di prodotto e di sottoprodotto nell'omonima tabella;
- materializzeremo la tabella come vista giornaliera del portafoglio.

5.3.2 Architettura indice Alfa MTD

L'architettura per la generazione dell'indice *Alfa MTD*, come mostrato in Figura 5.4, si divide in quattro fasi, ovvero:

- prenderemo tutte le fotografie giornaliere dell'ultimo mese non consolidato;
- calcoleremo una media del portafoglio mensile;
- materializzeremo la tabella delle medie mensili;
- materializzeremo la vista mensile del portafoglio consolidato;

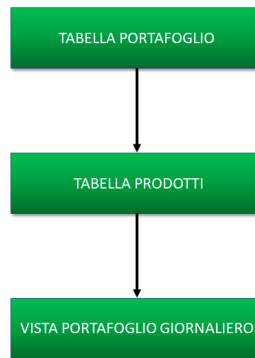


Figura 5.3: Architettura su BigQuery per il KPI Alfa Giornaliero



Figura 5.4: Architettura su BigQuery per il KPI Alfa MTD

5.3.3 Architettura indice Beta Giornaliero

L'architettura per la generazione dell'indice *Beta Giornaliero*, come mostrato in Figura 5.5 si divide in tre fasi, ovvero:

- prenderemo dalla tabella Transazioni sempre l'ultima fotografia del giorno.
- assegneremo la categoria dell'*Alfa* e della relativa *%Alfa* per ogni tipologia di prodotto e di sottoprodotto nell'omonima tabella.
- materializzeremo la tabella come vista giornaliera del portafoglio.

5.3.4 Architettura indice Beta36

L'architettura per la generazione dell'indice *Beta Giornaliero*, come mostrato in Figura 5.6, usa come tabella BigQuery lo storico delle transazioni, e da questa preleverà l'ultimo mese.

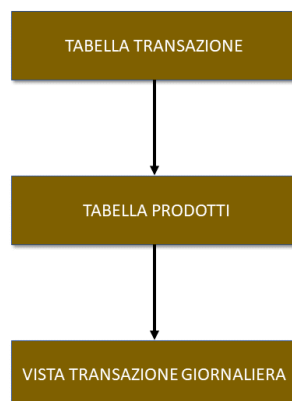


Figura 5.5: Architettura su BigQuery per il KPI Beta Giornaliero

Inoltre, verranno materializzate tre viste e sarà, poi, uno degli enti esterni, ovvero l'ente soprannominato in modo fittizio *LUX*, che si occuperà di integrare tutti i dati insieme per poi rimandarli al reparto di interesse, dopo un lavoro preventivo di ETL.

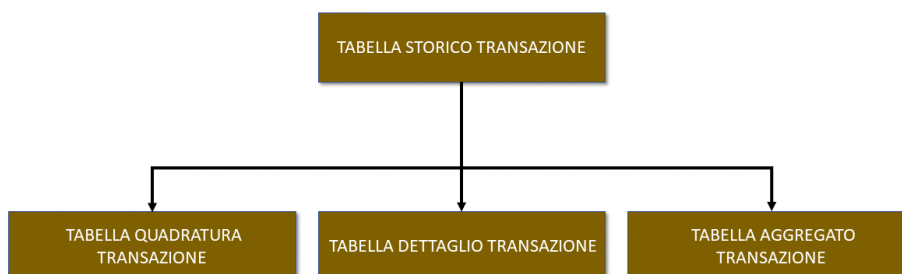


Figura 5.6: Architettura su BigQuery per il KPI Beta36

In questo capitolo verranno esposte le implementazioni degli indicatori di performance argomento della presente tesi. In particolare, verranno illustrate, mediante pseudocodice, le costruzioni delle applicazioni riguardanti tali KPI.

6.1 Controllo di qualità dei dati

Per comprendere meglio l'implementazione ricapitoleremo quali sono gli obiettivi da raggiungere. I principali KPI da calcolare sono:

- *Alfa*, considerando il set di prodotti presenti nel portafoglio.
- *Beta*, considerando le variazioni dovute alle transazioni.

Come esposto nei capitoli precedenti, al fine di poter calcolare questi due indici, è necessario dapprima un controllo sui dati in input.

Questo controllo è stato implementato direttamente su BigQuery attraverso la creazione di una tabella con chiave univoca. Ciò genererà diverse tabelle che saranno collegate tra loro attraverso la creazione di una tabella comune detta `Common Table`, composta dalla concatenazione delle varie chiavi per la fruizione della vista univoca finale.

Al fine di comprenderne l'implementazione si considerino le tabelle mostrate in Figura 6.1, le quali rappresentano le tre diverse tipologie di output ottenibile, ovvero:

- 0: equivale ad un esito `Positivo`.
- 1: equivale ad un esito `Warning`.
- 2: equivale ad un esito `Negativo`.

Infatti, nel processo di caricamento dei dati, il dato sottoposto al controllo verrà processato. L'esito, poi, sarà aggiunto alla tabella stessa, disponibile per la visualizzazione specifica del controllo, mentre il risultato aggregato sarà presente in una tabella secondaria.

Come mostrato in Figura 6.2, i diversi output che vediamo negli esempi delle tabelle fittizie della Figura 6.1 si ripropongono nel flusso di controllo degli esiti.

Ad esempio, nel caso in cui il campo `output` sia posto a 1, il controllo dei dati ritornerà come esito `Warning` e l'aggregazione dei controlli ritornerà per tutti i campi relativi a quell'agente esito `Warning` per il controllo finale.

Check ID	Agent_Key	%Alfa_t	%Alfa_t-1	Delta	GoMin	GoMax	AlertMin	AlertMax	Output
1	AAABBB00	0,025	0,070	0,045	0	0,025	0,025	0,05	1
1	AAABBB00	0,013	0,011	0,002	0	0,025	0,025	0,05	0
1	AAABBB00	0,017	0,023	0,023	0	0,025	0,025	0,05	0

Check ID	Agent_Key	%Alfa_t	%Alfa_t-1	Delta	GoMin	GoMax	AlertMin	AlertMax	Output
1	AAABBB00	0,025	0,080	0,055	0	0,025	0,025	0,05	2
1	AAABBB00	0,013	0,011	0,002	0	0,025	0,025	0,05	0
1	AAABBB00	0,017	0,023	0,023	0	0,025	0,025	0,05	0

Check ID	Agent_Key	%Alfa_t	%Alfa_t-1	Delta	GoMin	GoMax	AlertMin	AlertMax	Output
1	AAABBB00	0,025	0,035	0,005	0	0,025	0,025	0,05	0
1	AAABBB00	0,013	0,011	0,002	0	0,025	0,025	0,05	0
1	AAABBB00	0,017	0,023	0,023	0	0,025	0,025	0,05	0

Figura 6.1: Tabelle di controllo della qualità dei dati

Nel caso in cui il campo output sia posto a 2, quindi ritornando un esito negativo, anche l'aggregazione dei controlli per quel cliente ritornerà un esito finale *Negativo*. Solo qualora tutti i campi relativi allo stesso cliente abbiano output a 0 l'aggregazione restituirà esito *Positivo*, permettendo, così, il calcolo degli indici.

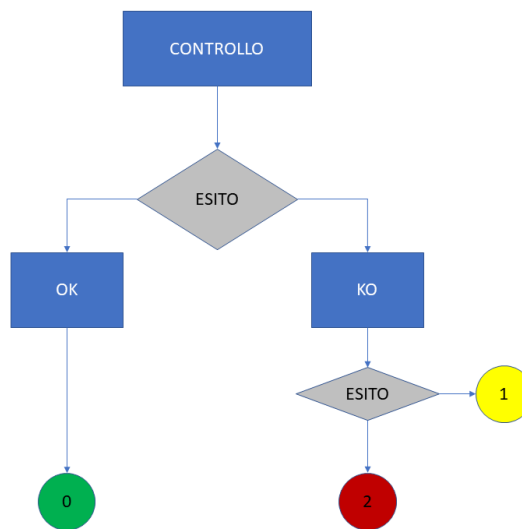


Figura 6.2: Flusso di controllo degli esiti sul controllo dei dati

6.2 Calcolo Degli indici Alfa e Beta

Per quanto riguarda l'indice Alfa partiamo dall'assunzione che un cliente abbia un deposito su un prodotto, che sia questo un fondo, un'assicurazione o altro. Il gruppo "Society" guadagna su questo prodotto una percentuale del suo valore.

Sul prodotto abbiamo accesso a diverse informazioni, come consulente finanziario, nome, data di apertura, etc. L'indice *Alfa* è calcolato come il valore del portafoglio per la percentuale applicata dall'azienda stessa.

Un primo problema potrebbe essere quello di trovare la percentuale corretta da applicare. Nella pratica ci si collega ad una tabella contenente queste informazioni. Tale tabella è mantenuta dagli utenti del gruppo "Society" ed è localizzata su un foglio Google. Ogni categoria di prodotto ha una sua specifica tabella contenente le informazioni sulle percentuali da applicare.

Il portafoglio cambia durante l'anno. Quindi, *Alfa* non è un valore statico, ma viene calcolato tutti i giorni. Il suo valore può fluttuare per effetto del mercato, oppure per qualche decisione del cliente. Ad ogni esecuzione della procedura, non si calcola l'*Alfa* totale, ma l'*Alfa* giornaliero, ovvero il valore del portafoglio attuale diviso il numero di giorni non festivi dell'anno. I giorni non festivi sono identificati da un calendario (anche questo fornito dall'azienda stessa). Nei giorni festivi, l'*Alfa* non viene calcolato.

La formula risulta, quindi, essere:

$$Alfa_{giornaliero} = \frac{Valore * \%Alfa}{giorni_{tot}} \quad (6.2.1)$$

Nel caso di nuovi prodotti, è possibile che manchino delle informazioni rilevanti ai fini del calcolo dell'indice *Alfa*. In questi casi, quest'ultimo viene comunque calcolato usando valori di default.

Per permettere agli utenti di raccogliere queste informazioni, prima di avere il risultato a consuntivo, intorno al 15 del mese viene avviata una procedura per ricalcolare l'*Alfa* mensile (quindi, diverso da quello giornaliero) del mese precedente. Per questo calcolo, si considerano tutte le variazioni del valore del portafoglio del mese precedente e si fa una media aritmetica. Questa media viene moltiplicata per la percentuale applicata dalla società e per un rapporto dato tra i giorni non festivi del mese considerato e tutti i giorni non festivi dell'anno.

La formula per quest'ultimo caso risulta, quindi, essere:

$$Alfa_{mensile} = \frac{\sum_{i=1}^{giorni_{mese}} Valore_i}{giorni_{mese}} * \%Alfa * \frac{giorni_{mese}}{giorni_{tot}} \quad (6.2.2)$$

Non potendo mostrare la codifica implementativa per ragioni di policy di riservatezza, descriveremo i passi e le scelte che sono poi state trasformate in codice su Qlik.

Per il calcolo dell'indice *Alfa*, la tabella dei fatti di riferimento è denominata (in modo fittizio nella presente tesi) FACT_PORTFOLIO. Questa tabella viene aggiornata tutti i giorni, per considerare la variazione del portafoglio clienti; quindi esiste una vista di supporto chiamata LAST_FACT_PORTFOLIO_V (alias PTF) che considera i record solo dell'ultimo caricamento (storicizzazione dei dati effettuata in ingresso). La PTF contiene solo le chiavi, ovvero i codici, dei fatti. Le descrizioni sono poste sulle tabelle delle dimensioni. La PTF presenta il seguente dettaglio di spaccatura: Agente – Cliente – Dossier – Contratto – Prodotto – Sottoprodotto. Per il calcolo del dell'indice *Alfa*, la PTF è la tabella di partenza; e questa subisce diverse trasformazioni distinte in tre fasi:

1. Aumento di dettaglio, con l'aggiunta di descrizioni e informazioni mancanti da tabelle di supporto. Lo scopo principale di questa vista è quello di effettuare tutte le `join` necessarie per poter calcolare il valore della percentuale *Alfa* corretto in base alla categoria del prodotto. Tutte queste procedure sono implementate in `V_ALFA_PTF_STEP1`.
2. In `V_ALFA_PTF_STEP2` viene calcolato il valore corretto della percentuale *Alfa*. Questo valore tiene conto dei valori reali delle percentuali (ovvero valori che esistono nelle

tabelle gestite dal gruppo “Society”), dei valori di default (nel caso in cui non esista una percentuale associata ad un prodotto specifico) e di eventuali maggiorazioni che devono essere applicate nel caso in cui ci siano dei sottoprodotti associati al prodotto. Infine, vengono definiti una serie di `flag` per l’individuazione di valori mancanti.

3. Gli ultimi passaggi e la `select` finale avvengono in `V_ALFA_PTF_STEP3`. Tutti i `flag` vengono raggruppati in un’unica colonna e il valore *Alfa* viene calcolato moltiplicando la percentuale *Alfa* per il valore del portafoglio e dividendo il risultato per il numero di giorni non festivi dell’anno in questione. Per effettuare questo calcolo è necessario legarsi a una tabella calendario.

Nello Step1, si effettuano una serie di `join` sequenziali e si applicano una serie di logiche nella `select` per arricchire il dettaglio del dato in ingresso. In particolare, la `PTF` viene elaborata come segue:

- `join` con `TAB_DIM_CLIENT` per la descrizione del cliente.
- `join` con `TAB_DIM_PRODUCT` per la descrizione del prodotto principale e del sottoprodotto.
- `join` con `TAB_DIM_CONTRACT` per la descrizione del contratto.
- `join` con `VIEW_MAP_CLUSTER` (alias `VPC`) da utilizzare per la definizione della categoria. A questa vista di supporto ci si lega per prodotto, contratto e attività di consulenza; essa restituisce il cluster di appartenenza di un prodotto. Il cluster (ma non solo) serve per calcolare la categoria. Quest’ultima (in tabella il nome tecnico è `ALFA_CATEGORY`) è importante perché è il campo chiave utilizzato per recuperare la logica del motore di calcolo, il quale restituisce la percentuale *Alfa* da applicare al portafoglio. Per definire la categoria si applicano una serie di logiche che sono strettamente legate alla tipologia di prodotto.

L’*Alfa* mensile viene gestito da altre viste, con la nomenclatura `VIEW_ALFA_PTF_STEPn`. Il motore di calcolo delle percentuali è gestito nello Step2 della vista `VIEW_ALFA_PTF_STEP2`.

Il progetto considera tutti i prodotti gestiti dal gruppo “Society” e richiede dei risultati precisi dal momento che va a trattare gli incentivi per i consulenti finanziari. Pertanto, se ci sono degli errori, bisogna gestirli, aggiungendo eccezioni alle regole e nuove procedure. In alcuni casi è sufficiente aggiungere dei dettagli, in altri bisogna necessariamente rivedere la struttura delle viste. Questo impatta soprattutto i dati in ingresso. A conclusione dello Step1, la tabella dei fatti centrale viene messa in `join` successive per legarsi alle percentuali delle singole categorie identificate dalla colonna `ALFA_CATEGORY`.

Lo Step2 è organizzato in quattro fasi successive, gestite tramite `join nested`. Queste fasi si possono riassumere come segue:

1. Identificazione del caso di default.
2. Assegnazione della percentuale *Alfa* ad ogni prodotto in base all’appartenenza ad una specifica categoria. In generale, ad ogni prodotto si assegna il primo valore non nullo tra la percentuale corretta e quella di default.
3. Per alcuni prodotti societari, è necessario considerare una maggiorazione rispetto alla percentuale *Alfa* definita precedentemente. Questo vale a prescindere dal fatto che la percentuale *Alfa* è stata correttamente calcolata oppure no. In particolare, questa maggiorazione si ha quando c’è un’attività di consulenza e l’aliquota (aggiunta nello

Step1 delle viste) è correttamente definita. In particolare, i prodotti che appartengono a determinate categorie aggiungono alla percentuale *Alfa* il valore della percentuale *Alfa* del sottoprodotto. Se non c'è il sottoprodotto si assegna una maggiorazione di default.

Per il calcolo dell'indice *Beta* il procedimento è duale, con l'unica differenza che la tabella dei fatti di partenza è la `FACT_TRANSACTION`, quindi il calcolo non avviene sul portafoglio, come per l'*Alfa*, ma avviene per transazione.

L'elaborazione dei dati è gestita con tre viste di supporto con la seguente nomenclatura: `V_BETA_TRN_STEPn`. La gestione delle transazioni storiche viene effettuata sulle viste con la nomenclatura `VIEW_BETA_TRN_STEPn`.

Nel prossimo capitolo riporteremo le principali dashboard realizzate in Qlik.

App realizzate e Mashup

In questo capitolo verranno mostrate le principali dashboard realizzate in Qlik Sense. Queste visualizzazioni hanno lo scopo di mostrare come gli indici di performance argomento della presente tesi sono stati integrati nel modello dei dati già presente. Inoltre, verrà presentato un lavoro di Mashup, funzionalità offerta da Qlik che permette di dare una panoramica sulle varie dashboard da un unico entry point.

Nelle figure riportate sono state oscurate alcune informazioni per il rispetto delle policy di sicurezza e riservatezza aziendale.

7.1 Dashboard relativa all'indice Alfa

In questa sezione verranno mostrate le dashboard relative al primo dei due indici trattati nella presente tesi, ovvero il KPI *Alfa*.

In Figura 7.1 viene mostrata la pagina dell'App relativa all'indice *Alfa* con tutti i suoi fogli disponibili. Le parti oscurate riguardano le informazioni sulla posizione assoluta nel quale questi sheet sono localizzati nel server di Qlik e le immagini proprietarie dell'azienda utilizzate per identificare i vari fogli.

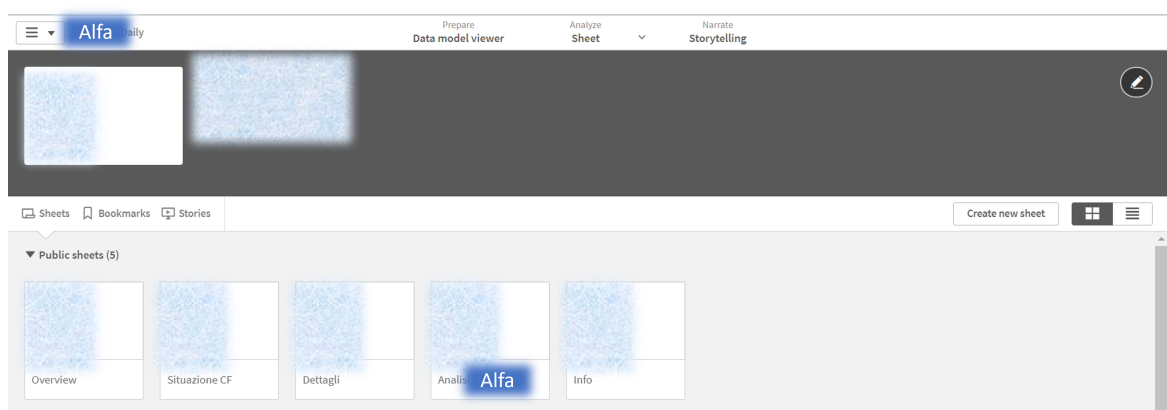


Figura 7.1: App e relativi sheet relativi all'indice Alfa

La prima dashboard realizzata, relativa allo sheet denominato *Overview*, è mostrata in Figura 7.2. Questo foglio riporta, attraverso degli oggetti KPI situati in alto, le informazioni riepilogative relative all'indice *Alfa*, come l'indice alla data odierna, l'*Alfa* percentuale e il valore totale stimato raggiungibile a fine anno.

Inoltre, tale dashboard riporta diverse tabelle, le quali indicano, ad esempio, il valore dell'indice raggruppato per le varie aree di divisione aziendale, o, per mese o area.

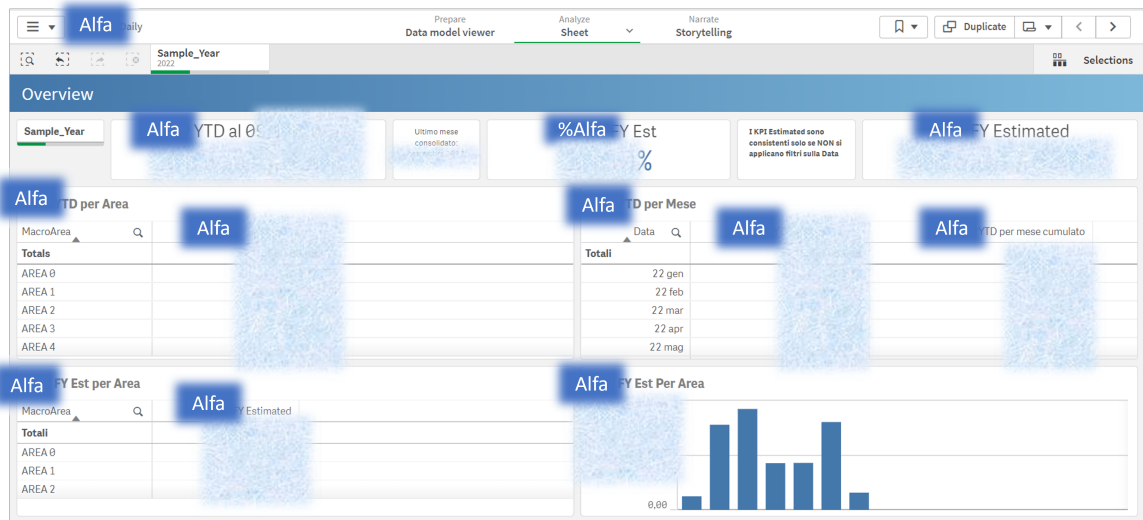


Figura 7.2: Dashboard Overview relativa all'indice Alfa

Un'altra dashboard realizzata, mostrata in Figura 7.3, è quella relativa al foglio *Dettagli*, utile ai promotori finanziari non solo per avere un'idea generale del valore degli indici, anche ricavabile da altri fogli, ma, soprattutto, per poter indagare su quali aree generino determinati profitti e a quali consulenti siano questi attribuibili.

Nell'area di destra del foglio i promotori hanno, inoltre, la possibilità di applicare dei filtri in modo autonomo, grazie al pannello costruito appositamente.

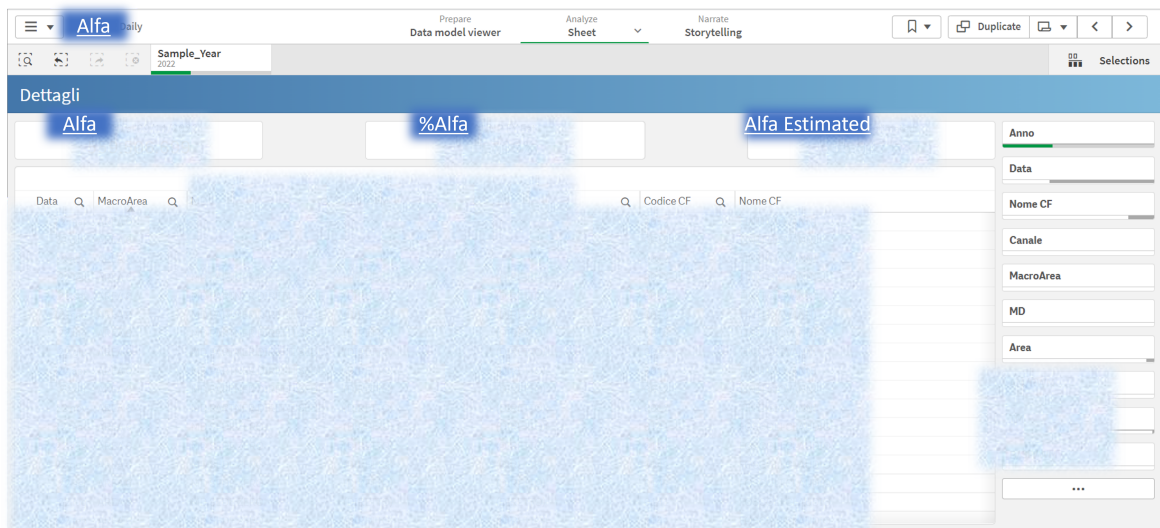


Figura 7.3: Dashboard Dettagli relativa all'indice Alfa

La dashboard relativa all'*Analisi*, in Figura 7.4, mostra, attraverso l'utilizzo di un diagramma a barre, una classifica dei consulenti finanziari i cui clienti associati hanno il miglior valore dell'indice *Alfa*. Questo foglio viene anche utilizzato a fini di concorsi che vedono sfidarsi tra di loro i vari consulenti, ai quali verrà riconosciuta una ricompensa a fine ciclo lavorativo.

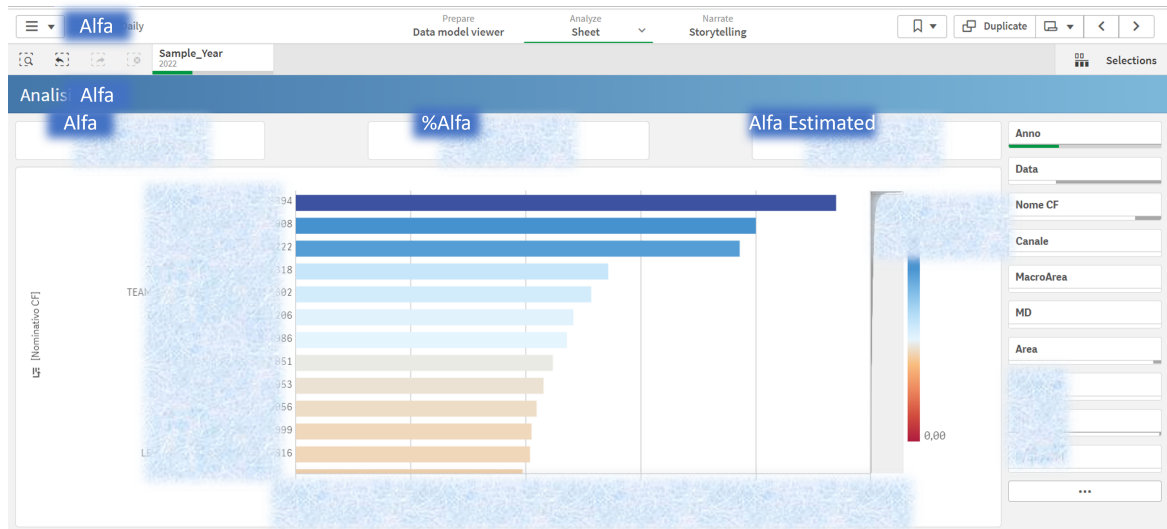


Figura 7.4: Dashboard Analisi relativa all'indice Alfa

7.2 Dashboard relativa all'indice Beta

In questa sezione verranno mostrate le dashboard relative al secondo dei due indici trattati nella presente tesi, ovvero il KPI *Beta*.

Come discusso nei capitoli precedenti, questi due indici differiscono soltanto per l'ambito di applicazione, il primo (*Alfa*) è calcolato a partire dalle informazioni sul portafoglio, il secondo (*Beta*) è calcolato sulla base delle transazioni, motivo per cui le dashboard saranno pressoché duali.

Come per l'indice *Alfa* anche l'indice *Beta* possiede una sua App e le dashboard relative a quest'ultimo sono mostrate in Figura 7.5. Anche in questo caso le parti oscurate riguardano le informazioni sulla posizione assoluta nella quale questi sheet sono localizzate nel server di Qlik e le immagini proprietarie dell'azienda utilizzate per identificare i vari fogli.

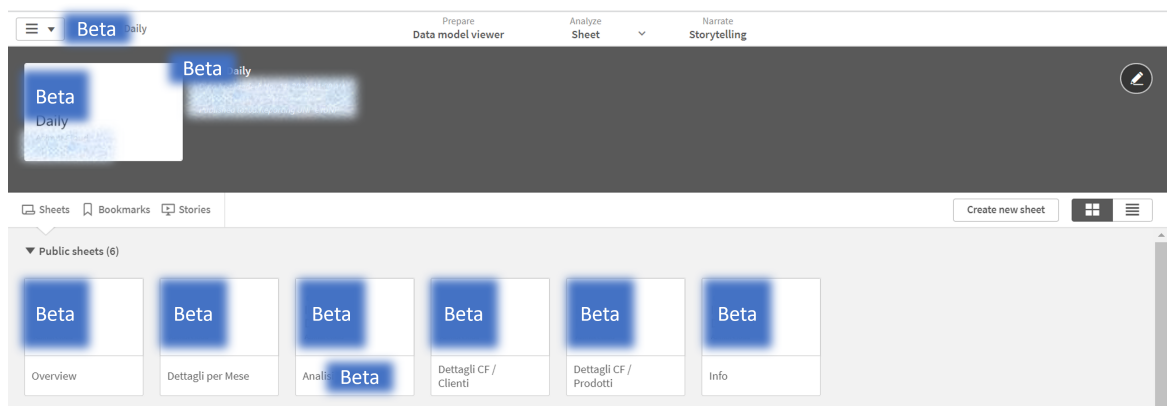


Figura 7.5: App e relativi sheet relativi all'indice Beta

Il primo sheet, *Overview*, contiene la dashboard rappresentata in Figura 7.6. Questo foglio riporta, attraverso degli oggetti KPI situati in alto, le informazioni riepilogative relative all'indice *Beta*, proprio come avviene per la dashboard relativa all'*Alfa*.

Inoltre, tale dashboard riporta diverse tabelle, le quali indicano, ad esempio, il valore dell'indice raggruppato per le varie aree di divisione aziendale, o, per mese.

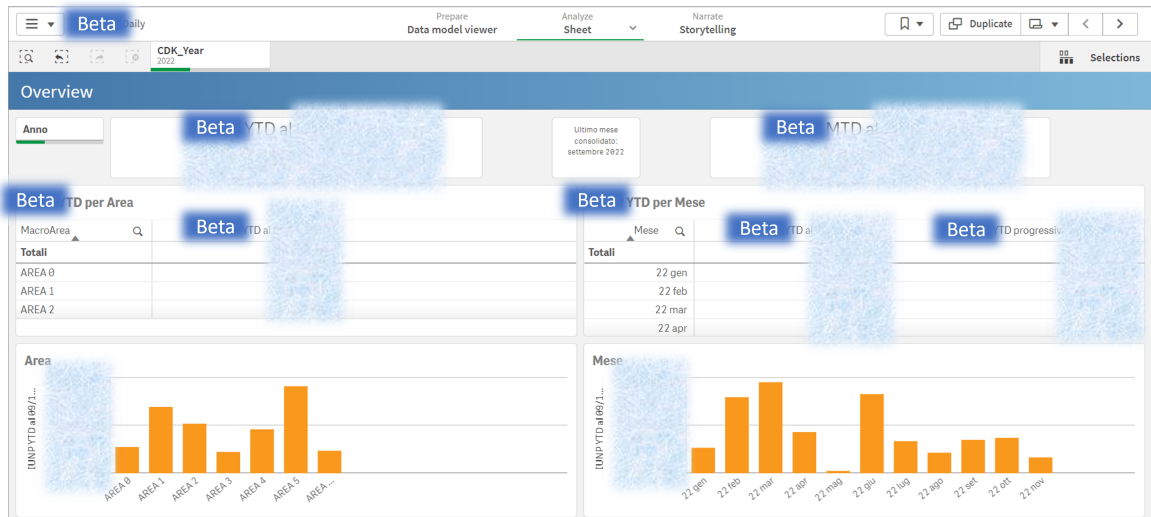


Figura 7.6: Dashboard Overview relativa all'indice Alfa

In Figura 7.7 viene mostrata un'altra dashboard realizzata, ovvero, quella relativa al foglio *Analisi*, utile ai promotori finanziari non solo per avere un'idea generale del valore degli indici, anche ricavabile da altri fogli, ma, soprattutto, per poter indagare su quali aree generino determinati profitti e a quali consulenti siano questi attribuibili. Successivamente, tali dati potranno essere utilizzati al fine di generare una classifica di consulenti finanziari per poi concorrere all'attribuzione di premi aziendali.

Nell'area destra del foglio i promotori hanno, inoltre, la possibilità di applicare dei filtri in modo autonomo, grazie al pannello costruito appositamente.



Figura 7.7: Dashboard Analisi relativa all'indice Alfa

Un'ultima dashboard che analizzeremo relativa all'indice *Beta* è quella relativa allo sheet *Dettaglio CF/Clienti*. In questo foglio, oltre agli oggetti KPI che ormai siamo abituati a riconoscere, vengono mostrate, mediante l'utilizzo di una tabella, le informazioni sui clienti, quali il loro consulente di riferimento e il calcolo del KPI specifico per ogni cliente, in modo da poter indagare su delle future azioni da poter intraprendere con questi ultimi.

7.3 Mashup

Qlik Sense è sviluppato in modo che ogni App creata abbia un suo spazio sul server e i dati contenuti al suo interno siano fruibili solo ed esclusivamente all'interno di questa. Potrebbe, a volte, capitare la necessità di confrontare diverse visualizzazioni presenti in diverse App. Qlik mette a disposizione uno strumento denominato "Mashup", ovvero, una pagina web nella quale poter richiamare oggetti presenti in diverse applicazioni e farli coesistere, a livello di visualizzazione, tra di loro.

Nel progetto argomento della presenti tesi, i due indici *Alfa/Beta* sono stati utilizzati in diverse applicazioni dopo la loro creazione. Partendo dal presupposto che gli indici vengano calcolati allo stesso modo, può esistere una possibilità nella quale i risultati in diverse App siano diversi, magari dovuti a ritardi negli aggiornamenti, errori non voluti da parte degli operatori o altre motivazioni. Il Mashup permette di tenere sotto controllo, in un'unica area di visualizzazione, tutti i vari indici. Ed è proprio per questo motivo che nasce, all'interno del nostro progetto, un'App di Mashup.

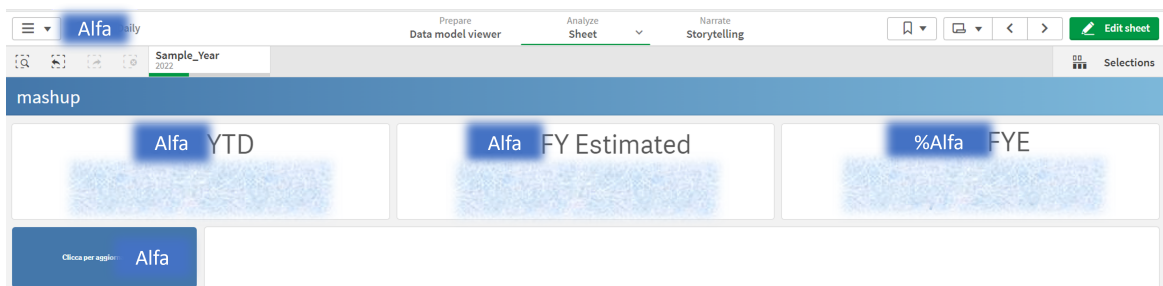


Figura 7.8: Dashboard Mashup relativa all'indice Alfa

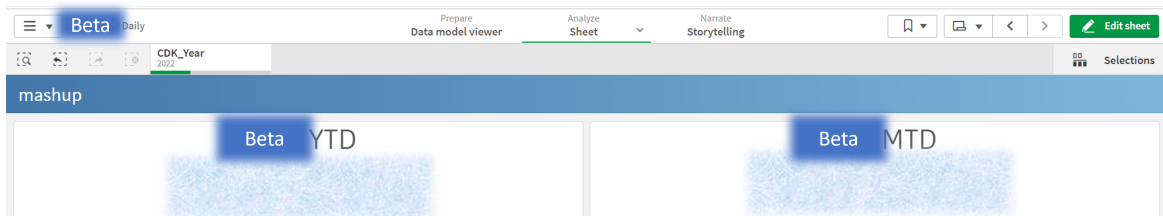


Figura 7.9: Dashboard Mashup relativa all'indice Beta

Come mostrato in Figura 7.8 e in Figura 7.9, sono stati creati due sheet locali nelle App degli indici *Alfa* e *Beta*, che riprendano i valori dei KPI specifici di quell'App. Inoltre, anche all'interno delle altre applicazioni, non discusse nella presente tesi, le quali utilizzano tali indici, sono stati creati i relativi fogli per il Mashup.

Tutti questi oggetti di Qlik vengono richiamati nell'App di Mashup per poter verificare che i dati siano tra loro allineati; in caso contrario si procederà alla correzione delle App disallineate.

Il Mashup permette, attraverso l'utilizzo di vari linguaggi, di creare una vera e propria applicazione web fruibile dal Controllo di Gestione che si occuperà di informare il gruppo di lavoro in caso di presenza di squadrature.

L'App di Mashup è riportata in Figura 7.10.

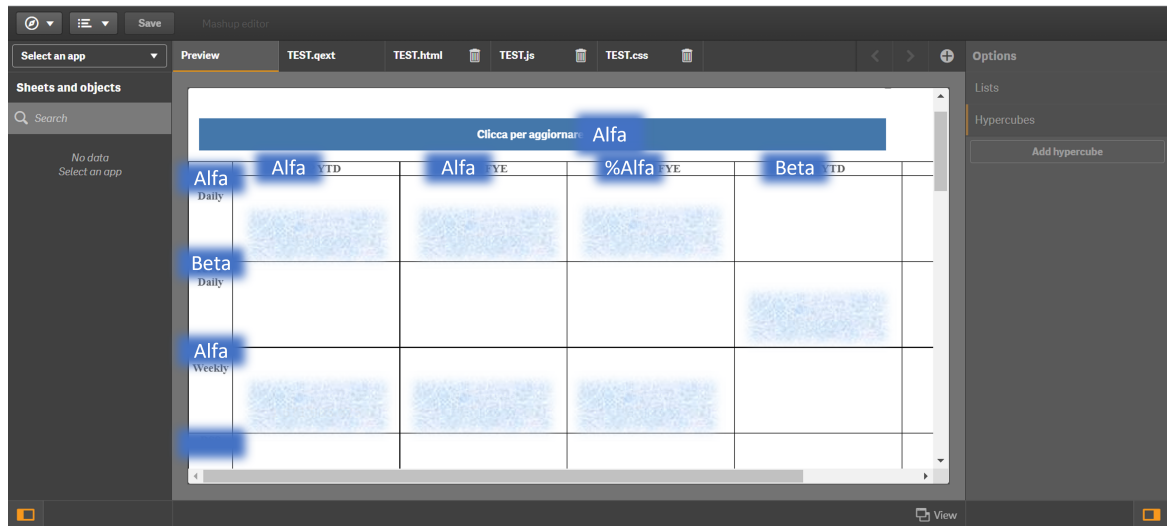


Figura 7.10: App Mashup di Qlik Sense

Discussione in merito al lavoro svolto

Nel capitolo corrente verrà proposto un riepilogo in merito al lavoro svolto nella presente tesi sotto un aspetto non tecnico ma, puramente, esperienziale. Forniremo il nostro punto di vista su quali siano i punti di forza e di debolezza del progetto e discuteremo brevemente sulle lezioni apprese durante questo percorso.

8.1 Punti di forza del progetto

La presente tesi è stata svolta in ambito di tirocinio aziendale presso una famosa azienda di consulenza informatica. Il periodo considerato ha avuto una durata di sei mesi, durante i quali abbiamo svolto svariati task.

Inizialmente, affiancati da figure con anzianità maggiore sul progetto, siamo stati guidati alla conoscenza generale dello stesso, quindi, ci siamo informati sul cliente per cui svolgevamo quotidianamente le mansioni, nell'architettura generale del progetto, sui sistemi di BI utilizzati e sui task da svolgere quotidianamente.

Durante il periodo di tirocinio ogni giornata lavorativa iniziava con lo stand-up meeting durante il quale si ricapitolavano i task di ogni componente. Il gruppo era, ed è tuttora, diviso in due parti, ovvero backend e frontend. Il primo sottogruppo si occupa delle parte riguardante le estrazioni di set di dati per il cliente; il secondo sottogruppo si occupa della visualizzazione delle informazioni. Inoltre, il venerdì, a fine giornata, si teneva il weekly meeting, per discutere dei task da svolgere la settimana successiva.

Dopo un primo periodo di ambientamento, abbiamo ricevuto i primi task, che riguardavano prevalentemente un controllo sulle applicazioni Qlik Sense già esistenti. Una volta presa padronanza delle tecniche utilizzate, abbiamo ricevuto il task argomento della presente tesi. Tutto ciò di cui abbiamo discusso costituisce dei punti di forza, ovvero:

- Partecipare allo stand-up meeting costituisce parte fondamentale del progetto, in quanto permette di delineare i task dei vari membri e restare sempre aggiornati per quanto ne concerne gli sviluppi.
- La divisione del gruppo in sottogruppi, ognuno adibito ad una fase dello stream, permette di concentrare le proprie energie sulla soluzione di problematiche specifiche.
- Vedersi assegnato un task di dimensioni importanti, nonostante la poca esperienza, è stimolante e sfidante, nonché permette di apprendere molti skill tecnici.

- Le soluzioni di Business Intelligence adottate nell'ambito di questo progetto, ovvero Qlik e Google BigQuery, sono al vertice delle migliori soluzioni di BI. Utilizzare questi tool permette di sviluppare skill tecnici facilmente vendibili su mercato.

8.2 Punti di debolezza

Come si può immaginare, spesso, i punti di forza hanno il risvolto della medaglia e nascondono dei punti di debolezza che, se comunque usati a proprio vantaggio, possono diventare delle opportunità.

Ad esempio, quando un task importante viene assegnato ad un nuovo componente, questo può generare una prima fase di timore e sconforto, se non si impara a chiedere aiuto alle figure senior, questo può essere vissuto come uno scoglio insuperabile.

Prendere parte ad un progetto già in essere potrebbe far pensare di essere partirti svantaggiati, e la voglia di recuperare, a volte, potrebbe portare a svolgere i task in maniera veloce, non prestando attenzione ai dettagli, generando, così, un risultato non adatto e non vendibile al cliente, a causa della sua scarsa qualità.

Inoltre, quando un ambiente è suddiviso in vari componenti, ciascuno con i suoi compiti, si potrebbe essere costretti a collaborare; questo, in parte, viene visto come un punto di debolezza poiché delle volte è necessario attendere delle tempistiche anche troppo ampie, generando, così, ritardi nelle consegne.

8.3 Lezioni apprese

Questa esperienza ha avuto i suoi pro e i suoi contro come appena descritto. Per certo, possiamo affermare che abbia contribuito alla crescita personale sotto aspetti riguardanti sia i soft skill che gli skill tecnici. Per quanto riguarda questi ultimi possiamo evidenziare i seguenti aspetti appresi:

- Nell'avere contatti con il cliente è utile utilizzare un linguaggio non troppo tecnico, ma comunque informale. Tale linguaggio è, comunemente, conosciuto come "linguaggio business", e risulta essere importante poiché delle volte potremmo trovarci al cospetto di persone che non conoscono i tecnicismi, e si potrebbero generare delle incomprensioni o, peggio, delle ambiguità, col rischio di non riconoscerle come tali.
- Nell'indicare le tempistiche di progetto è sempre utile considerare degli eventuali imprevisti, e non sottovalutare il lavoro assegnato al solo fine di sembrare più abili, poiché potrebbe creare sfiducia nel cliente quando poi non si rispetteranno le scadenze.
- Prima di consegnare un prodotto è sempre utile testarlo più volte, e farlo testare ad altre persone, in quanto la loro visione potrebbe essere diversa dalla nostra, permettendo, così, di individuare malfunzionamenti e poterle correggere.
- Ogniquale volta venga creato un applicativo è di fondamentale importanza aggiornare il manuale, utile sia ad un eventuale persona esterna che voglia continuare o modificare il lavoro, ma anche a noi stessi nel caso in cui tale applicativo debba essere modificato a distanza di tempo significative

Per quanto riguarda i soft skill appresi possiamo riassumerli come segue:

- *Autonomia*: è utile per svolgere le proprie mansioni senza che sia necessaria una supervisione da parte di altri componenti, non sottraendo, così, del tempo importante.

-
- *Resistenza allo stress*: permette di essere in grado di tenere testa alla pressione lavorativa mantenendo il controllo sulle proprie priorità.
 - *Precisione e attenzione per i dettagli*: questo poiché il cliente è alla ricerca costante di perfettibilità.
 - *Intraprendenza*: essere pro-attivi nel progetto permette di apprendere in maniera più veloce.
 - *Team Work*: è la capacità di lavorare in gruppo senza attriti, ma collaborando al fine di raggiungere gli obiettivi prefissati.

In questa tesi è stato illustrato lo sviluppo di un'applicazione in Qlik Sense, col supporto di Google BigQuery, volta alla creazione e visualizzazione di un sistema di KPI utili a descrivere la situazione finanziaria di un cliente all'interno di una società di gestione del risparmio.

Il percorso che ci ha portato alla definizione di questo sistema di indici parte, dapprima, da uno studio su quanto sia impattante, a livello strategico, l'utilizzo della Business Intelligence all'interno dei processi aziendali. Successivamente, è stato condotto lo studio dell'architettura aziendale che ci ha consentito di comprendere in quali aree sarebbe stato utile impiegare i mezzi messi a disposizione della BI. Una volta individuate tali aree si è passati alla raccolta dei requisiti, effettuata attraverso svariate interviste al cliente, al fine di poter, poi, definire i piani di progettazione e implementazione.

Per quanto riguarda la fase di progettazione, sono stati definiti i passi di ETL da intraprendere al fine di ottenere una struttura della base di dati che permettesse di ottenere le informazioni desiderate eliminando il "rumore" che generalmente contraddistingue le grandi moli di dati. In un secondo momento sono state definite le tecniche, che saranno poi oggetto di implementazione, utili a creare un meccanismo di quality check al fine di ottenere dei dati corretti per il calcolo degli indici di interesse.

Una volta chiare le tecniche di progettazione si è passati alla loro implementazione. Quest'ultima è avvenuta prevalentemente attraverso l'utilizzo di uno dei tool di Business Intelligence leader del mercato, ovvero Qlik, precisamente, nella sua versione più recente e potente, ovvero Qlik Sense. A supporto di quest'ultimo, è stato utilizzato un altro tool, leader del mercato per quanto riguarda i DBMS basati su cloud, ovvero Google BigQuery, il quale, mettendo a disposizione potenti mezzi di ETL, ha permesso, in combinazione con Qlik Sense, di raggiungere i risultati prefissati.

Sono state poi create, al fine della visualizzazione dei dati, diverse App in Qlik Sense con l'obiettivo di descrivere, attraverso il sistema di KPI in gioco, la situazione finanziaria dei vari clienti e il potere aziendale dei loro consulenti finanziari. Tutte queste informazioni sono state, poi, utilizzate dall'azienda per prendere delle decisioni strategiche in merito alla figura del cliente, e anche relativamente al suo consulente finanziario.

Una volta create le applicazioni in Qlik Sense si è passati alla fase di UAT (User Acceptance Testing), ovvero la fase in cui si effettua un test formale per scoprire se un sistema software soddisfa i suoi criteri di accettazione e per fare in modo che l'acquirente esamini se il sistema può essere accettato o meno. Per quanto riguarda il progetto oggetto della presente tesi il

cliente ha espresso la volontà di utilizzare quotidianamente tali indici ritenendosi soddisfatto del lavoro svolto.

Tale progetto è stato svolto in ambito di ETL e Data Visualization, ciò non toglie però, data l'enorme quantità dei dati immagazzinati, di progettare, come sviluppo futuro, un sistema di Forecasting. Quest'ultimo permetterebbe, attraverso la previsione di informazioni, di apportare un vantaggio competitivo ancora maggiore all'azienda, utilizzando tali informazioni per effettuare delle scelte strategiche di prevenzione o, perché no, utilizzandole in modo "offensivo", ovvero anticipando i principali competitor e avanzando nel mercato in maniera innovativa.

- BERGAMASCHI, M. (2021), *Dalla Business Intelligence al Data Warehouse*, Youcanprint.
- BERTHOLD, M. e HAND, D. (2007), *Intelligent Data Analysis: An Introduction*, Springer Berlin Heidelberg.
- BLOKDYK, G. (2018), *DIKW Pyramid: Complete Self-Assessment Guide*, CreateSpace Independent Publishing Platform.
- CADY, F. (2017), *The Data Science Handbook*, Wiley.
- CHAUDHURI, S. e DAYAL, U. (1997), «An overview of data warehousing and OLAP technology», *ACM Sigmod record*, vol. 26 (1), p. 65–74.
- FRIENDLY, M. (2008), «A brief history of data visualization», in «Handbook of data visualization», p. 15–56, Springer.
- GARCÍA, M., HARMSSEN, B., REDMOND, S. e POVER, K. (2018), *QlikView: Advanced Data Visualization: Discover deeper insights with Qlikview by building your own rich analytical applications from scratch*, Packt Publishing.
- ILACQUA, C., CRONSTROM, H. e RICHARDSON, J. (2015), *Learning Qlik Sense®: The Official Guide*, Packt Publishing.
- INMON, W. e LINSTEDT, D. (2014), *Data architecture: a primer for the data scientist: big data, data warehouse and data vault*, Morgan Kaufmann.
- KENNY, D., KASHY, D. e BOLGER, N. (1998), «Data analysis», in «The handbook of social psychology: Vols. 1 and 2», p. 233–265, McGraw-Hill New York.
- LABBE, P., ANJOS, C., SOLANKI, K. e DIMASO, J. (2019), *Hands-On Business Intelligence with Qlik Sense: Implement self-service data analytics with insights and guidance from Qlik Sense experts*, Packt Publishing.
- LAKSHMANAN, V. e TIGANI, J. (2019), *Google BigQuery: The Definitive Guide: Data Warehousing, Analytics, and Machine Learning at Scale*, O'Reilly Media.
- LIM, E.-P., CHEN, H. e CHEN, G. (2013), «Business intelligence and analytics: Research directions», *ACM Transactions on Management Information Systems (TMIS)*, vol. 3 (4), p. 1–10.
- NAVA, O. (2021), *Le SGR. Società di gestione del risparmio*, Minerva Bancaria.

- NEGASH, S. e GRAY, P. (2008), «Business intelligence», in «Handbook on decision support systems 2», p. 175–193, Springer.
- REZZANI, A. (2012), *Business intelligence*, PerCorsi di studio, Apogeo Education.
- SAGIROGLU, S. e SINANC, D. (2013), «Big data: A review», in «2013 international conference on collaboration technologies and systems (CTS)», p. 42–47, IEEE.
- SCHEPS, S. (2011), *Business Intelligence For Dummies, –For dummies*, Wiley.
- SCHWABISH, J. (2021), *Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks*, Columbia University Press.
- SULLIVAN, D. (2020), *Official Google Cloud Certified Professional Data Engineer Study Guide*, Wiley.
- TROYANSKY, O., GIBSON, T. e LEICHTWEIS, C. (2015), *QlikView your business: an expert guide to business discovery with QlikView and Qlik Sense*, John Wiley & Sons.
- TSAI, C.-W., LAI, C.-F., CHAO, H.-C. e VASILAKOS, A. V. (2015), «Big data analytics: a survey», *Journal of Big data*, vol. 2 (1), p. 1–32.
- VITANTONIO, J. e MAHLER, M. (2018), *Mastering Qlik Sense: Expert techniques on self-service data analytics to create enterprise ready Business Intelligence solutions*, Packt Publishing.

Siti consultati

- Magic Quadrant di Gartner – <https://coresistemi.it>
- Piramide Dikw Business – <https://premono.com>
- The Dynamics of Data Roles & Teams – <https://towardsdatascience.com>
- What's The Real Value of Big Data For Business? – <https://blog.hubspot.com>
- Le 5V dei Big Data – <https://www.flyip.it>
- La Business Intelligence... dalla BI alla Z! – <https://www.datamaze.it>
- Che cos'è un processo ETL – <https://www.talend.com>
- Se torturi i dati abbastanza, alla fine confesseranno quello che vuoi – <https://www.intesys.it>
- Il valore dei Big Data per la performance aziendale – <https://iris.luiss.it>
- Cos'è Google BigQuery – <https://www.tagmanageritalia.it>
- Qlik NPrinting – <https://www.qlik.com>

Ringraziamenti

Al professor Ursino, dopo l'esperienza triennale non ho avuto dubbi su chi avrei voluto come relatore della mia tesi magistrale; per la dedizione, la disponibilità, il rispetto e i consigli che mi ha dato dal primo momento all'ultimo.

Alla mia famiglia, che ha sempre creduto in me sostenendomi in ogni mia scelta; per avermi lasciato sempre libero di provare a volare.

Ai miei amici e colleghi, senza i quali, forse, non sarei fuori corso ma non sarei neanche felice.

*A tutti voi,
Grazie.*