

Capitolo 1: BI come strumento di supporto alle decisioni

1.1: Business Intelligence

Nel mondo moderno il valore dell'informazione sale incessantemente in quanto è necessario per analizzare l'operato dell'azienda, monitorare eventuali progressi verso obiettivi previsti (vendite previste, costi stimati ecc.) ed organizzare attività aziendali in maniera efficiente. Tuttavia, il reperimento dell'informazione è un processo non banale e laborioso quando si devono trattare moli di dati di dimensioni enormi quali possono essere i dati di vendita annuali di un'azienda o i costi di produzione relativi ad uno specifico prodotto. Inoltre, le aziende necessitano di tali informazioni in tempi brevi, spesso organizzate in modo personalizzato e di conservare i dati in modo centralizzato e non volatile.

Con *Business Intelligence* si intende l'insieme degli strumenti e dei procedimenti per selezionare, aggregare, correggere e trasformare i dati grezzi in conoscenza utile per supportare processi decisionali. Qui di seguito sono riportate alcune delle varie tecniche di Business Intelligence quali:

- Data Mining
- Reporting
- Analisi Statistica
- Analisi Descrittiva
- Query
- Metriche e Benchmarking delle prestazioni
- Ecc...

1.2: Data Warehousing

Il Data Warehousing è *una collezione di metodi, tecnologie e strumenti di ausilio al knowledge worker (dirigente, amministratore, gestore, analista) per condurre analisi dei dati finalizzate all'attuazione dei processi decisionali e al miglioramento del patrimonio informativo.* [1]

È un sistema di data management di supporto alla BI che permette di centralizzare e consolidare dati da diverse origini, mantenere grandi quantità di dati storici e su cui possono essere eseguite query e analisi. Il processo di Data Warehousing offre i seguenti vantaggi:

- Accessibilità a utenti con scarse conoscenze informatiche
- Integrazione dei dati sulla base di un modello standard dell'impresa
- Flessibilità di interrogazione per ottenere il massimo dalle informazioni presenti
- Sintesi per analisi mirate ed efficaci
- Rappresentazione multidimensionali per offrire all'utente una visione intuitiva ed efficacemente manipolabile delle informazioni
- Correttezza e completezza dei dati integrati

Il fulcro di questo processo è il Data Warehouse, *una collezione di dati di supporto per il processo decisionale che presenta le seguenti caratteristiche: orientato ai soggetti, integrato e consistente, rappresentativa dell'evoluzione temporale, non volatile.* [2]

1.2.1: Orientato ai soggetti

A differenza dei database relazionali, progettati con il focus sulle applicazioni che andranno ad utilizzarli, il Data Warehouse è progettato in vista degli utenti finali che andranno ad usufruirne.

[1]: Slide 9 Corso Business Intelligence, S.Rizzi

1.2.2: Integrato e consistente

Appoggiandosi a più fonti di dati eterogenee provenienti da basi di dati relazionali, sistemi informativi esterni o addirittura documenti non strutturati, è necessario unificare e assicurarsi che questi dati siano uniformi e renderli tali qualora non lo fossero.

Questo è il ruolo degli strumenti di ETL (Extraction, Transformation and Loading).



Figura 1: funzioni degli strumenti di ETL

1.2.3: Rappresentativa dell'evoluzione temporale

A differenza dei DB relazionali, in un DW il tempo è parte delle chiavi e i dati contenuti non possono essere aggiornati o sovrascritti.

1.2.4: Non volatile

In un DB relazionale i dati sono soggetti ad operazioni di Insert, Update e Delete, il che rende i dati non persistenti. In un DW i dati vengono caricati una volta e non verranno più modificati o rimossi. Il problema si sposta dalla gestione delle transazioni al query-throughput.

1.3: Architetture Data Warehouse

Affinché un'architettura di Data Warehouse funzioni, sono necessari dei requisiti:

- Separazione: l'elaborazione analitica e transizionale devono essere mantenute il più possibile separate
- Scalabilità: l'architettura deve poter essere facilmente ridimensionata a fronte della crescita nel tempo dei volumi di dati e del numero di utenti da gestire
- Estendibilità: deve essere possibile adottare nuove tecnologie e applicazioni senza riprogettare integralmente il sistema
- Sicurezza: controllo sugli accessi
- Amministrabilità: è necessario limitare la complessità dell'attività amministrativa

A fronte di questi requisiti ci sono varie implementazioni, differenziate dal numero di livelli di cui si compongono. Qui di seguito è descritta l'architettura a due livelli.

1.3.1 Architettura a due livelli

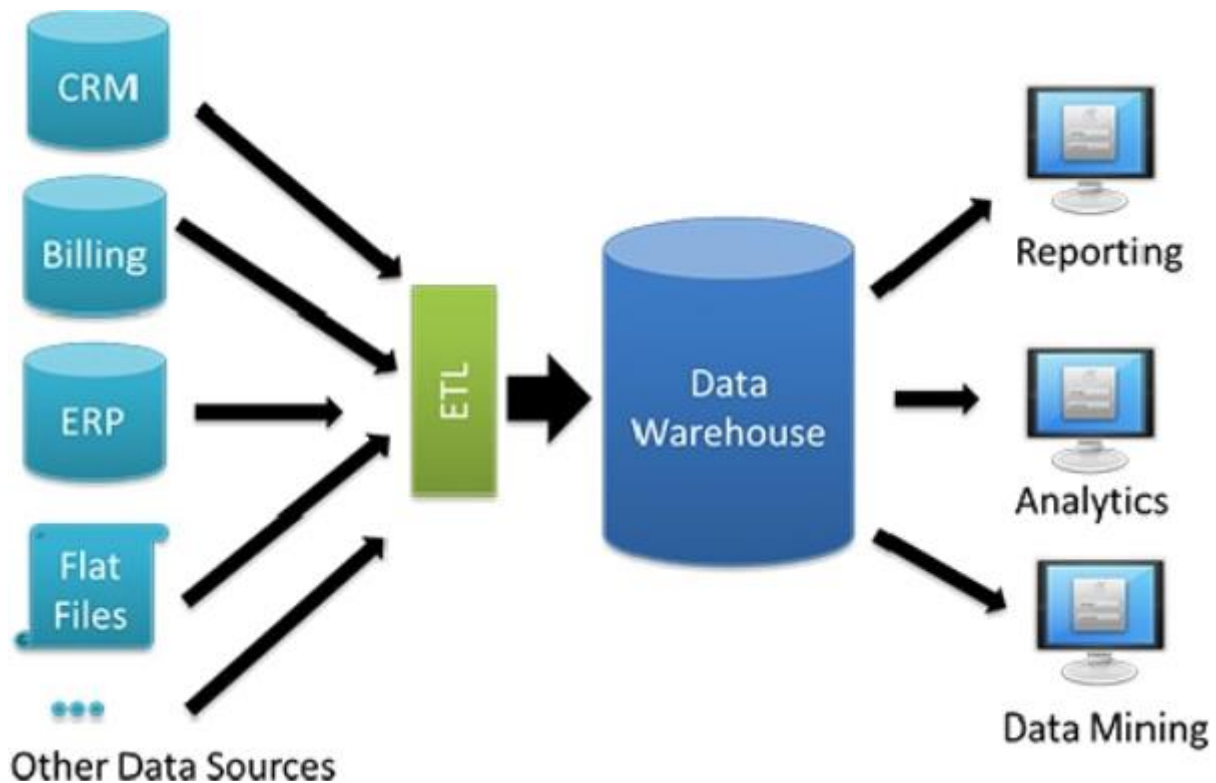


Figura 2: Architettura a due livelli

I due livelli che compongono questa architettura sono il livello sorgenti, composto da tutte le fonti di dati, e il livello Data Warehouse, dove il DW è composto da ulteriori “sotto DW” chiamati Data Mart, ossia *sottoinsiemi o aggregazioni di parte dei dati presenti nel DW primario*. Essi possono essere dipendenti, se usati per suddividere un DW che sarebbe altrimenti troppo grande per essere gestito agilmente e migliorare le prestazioni, o indipendenti, se alimentato direttamente da una fonte dati e sfruttato per facilitare la fase progettuale rendendo più difficile lo schema di accesso ai dati.

L'utilizzo di un'architettura a due livelli comporta numerosi vantaggi quali:

- Informazioni di buona qualità continuamente disponibili a livello di DW anche quando è temporaneamente precluso l'accesso alle sorgenti.
- Le interrogazioni analitiche effettuate sul DW non interferiscono con la gestione delle transazioni a livello operativo.
- L'organizzazione logica del DW è basata su un modello multidimensionale mentre le sorgenti offrono modelli relazionali o semi-strutturati.
- Discordanza temporale e di granularità tra sistemi OLTP, che trattano dati correnti e al massimo livello di dettaglio, e sistemi OLAP che operano su dati storici e di sintesi.
- A livello del warehouse è possibile impiegare tecniche specifiche per ottimizzare le prestazioni per applicazioni di analisi e reportistica

Capitolo 2: Descrizione Software utilizzati

In questo capitolo si tratteranno i software che sono stati impiegati per l'elaborazione del progetto di tesi.

Segue un elenco dei vari software per poi approfondire ognuno di essi nello specifico:

- *SQL Server Management Studio (SSMS)*: è un ambiente integrato per la gestione di qualsiasi infrastruttura SQL. SSMS integra un'ampia gamma di strumenti grafici con numerosi editor di script avanzati per offrire accesso a SQL Server per gli sviluppatori e gli amministratori di database qualsiasi sia il livello di competenza.
- *SQL Analysis Services (SSAS)*: è un motore dati analitici usato nel supporto decisionale e nell'analisi aziendale. Fornisce funzionalità del modello di dati semantico di livello aziendale per le applicazioni di business intelligence (BI), di analisi dei dati e di Reporting, ad esempio Power BI, Excel, Reporting Services e altri strumenti di visualizzazione dei dati.
- *Visual Studio Analysis Services projects*: è un'estensione di Visual Studio che fornisce la possibilità di progettare e costruire modelli tabulari e multidimensionali dispiegati poi in SQL Server Analysis Services, Power BI o Azure Analysis Services.
- *Visual Studio Reporting Services*: è un'estensione di Visual Studio che fornisce la possibilità di progettare e creare (anche tramite wizard) report professionali.
- *SQL Server Reporting Services (SSRS)*: offre un set di servizi e strumenti locali per creare, distribuire e gestire report impaginati e per dispositivi mobili anche a partire da modelli SSAS.

Tutti i software impiegati sono proprietari Microsoft ed è stata utilizzata la distribuzione 2019 per Visual Studio e la 2017 per i servizi relativi alla suite SQL.

2.1 SQL Server Management Studio



Figura 3: logo di Microsoft SQL Server Management Studio

Microsoft SQL Server Management Studio è uno degli ambienti di sviluppo e gestione per tipo di infrastruttura SQL più usati. Ciò è dovuto alla sua facilità d'uso per utenti meno esperti e dalla flessibilità e potenza se usato da mani esperte.

La peculiarità di SSMS è la possibilità di mantenere aperte più connessioni e navigare liberamente tra di esse mentre agisce in tempo reale su oggetti e tabelle.

L'ambiente include, oltre a tool grafici, editor testuali per scrivere script in vari linguaggi quali SQL, DAX, DMX, MDX ecc.

Siccome SSMS è stato pensato per essere affiancato a Visual Studio, un progetto sviluppato in tale ambiente gode di numerose estensioni e funzioni integrate per permettere agli sviluppatori di lavorare in maniera agile ed efficace.

2.2 SQL Analysis Services



Figura 4: logo di Microsoft SQL Analysis Services

Microsoft SQL Server Analysis Services è uno strumento del pacchetto Business Intelligence di SQL Server insieme a Reporting Services ed Integration Services e la sua funzione è offrire supporto per il data mining e OLAP (online analytical processing).

SSAS aggrega informazioni disseminati in più DataBase o tabelle in modelli tabulari o multidimensionali e supporta il linguaggio XML come DDL (data definition language) e MDX, LINQ, SQL (una parte), DMX e DAX come DML (data Manipulation Language).

2.3 Visual Studio Analysis Services projects

Estensione di Visual Studio che permette di creare un modello tabulare o multidimensionale a partire dai dati contenuti in DataBase. È inoltre possibile ritoccare, selezionare, ordinare e aggregare i dati prima di inserirli nel modello.

Dopodiché è possibile creare relazioni tra vari campi, gerarchie, misure sulle varie tabelle e dividere i dati in diverse partizioni o creare gruppi di utenti che potranno accedere a specifiche informazioni del modello.

Infine, il modello può essere distribuito su un server di SQL Analysis Services.

2.4 Visual Studio Reporting Services

Estensione di Visual Studio che permette di creare dei report impaginati a partire da modelli tabulari o multidimensionali reperiti da un server di SQL Analysis Services.

Reporting Services consente di generare report anche tramite wizard e di gestire tutti gli aspetti di un report quali parametri (visibili o no all'utente finale), set di dati estratti dai modelli SSAS usando query scritte in SQL, DMX, DAX o MDX ed elementi grafici quali indicatori, grafici di ogni tipo.

Ultimato il report è possibile pubblicarlo su SQL Reporting Services o un'altra piattaforma di visualizzazione report.

2.5 SQL Server Reporting Services



Figura 5: logo di Microsoft SQL Server Reporting Services

Microsoft SQL Server Reporting Services permette sia di visualizzare progetti di report creati esternamente sia di crearli. È possibile creare e gestire separatamente le origini dati, i set di dati e le pagine di report con impostazioni di sicurezza, piani di aggiornamento cache e simulazioni di esecuzione con calcolo dei tempi.

SSRS si presenta come un sito web molto intuitivo e strutturato come un filesystem dove l'utente seleziona la cartella contenente il report desiderato e visionarlo (è compito degli admin nascondere le parti non pensate per essere viste dagli utenti).

Capitolo 3: Progettazione modello

In questo capitolo si tratterà delle fasi di progettazione che riguardano il modello di Analysis Services su cui verrà poi basato il report.

3.1 Analisi dei Requisiti

Questa fase consiste nello studio degli elementi alla base del progetto e nella comprensione di quello che dovrà essere il risultato.

Inizialmente l'azienda ha illustrato nel dettaglio i procedimenti reali che stanno dietro ai dati raccolti nei Database, per poi andare sulle tabelle in questione. Segue lo schema ER riassuntivo delle tabelle principali (è stato scelto di mostrare uno schema riassuntivo e non esaustivo in quanto queste tabelle sono state create per essere usate nell'applicativo online dell'azienda e non per un progetto di BI, per il quale è stata invece precedentemente creata una "tabellona" unica contenente già le informazioni delle varie tabelle in join tra loro e un numero di attributi tale da non permettere una facile comprensione della stessa)

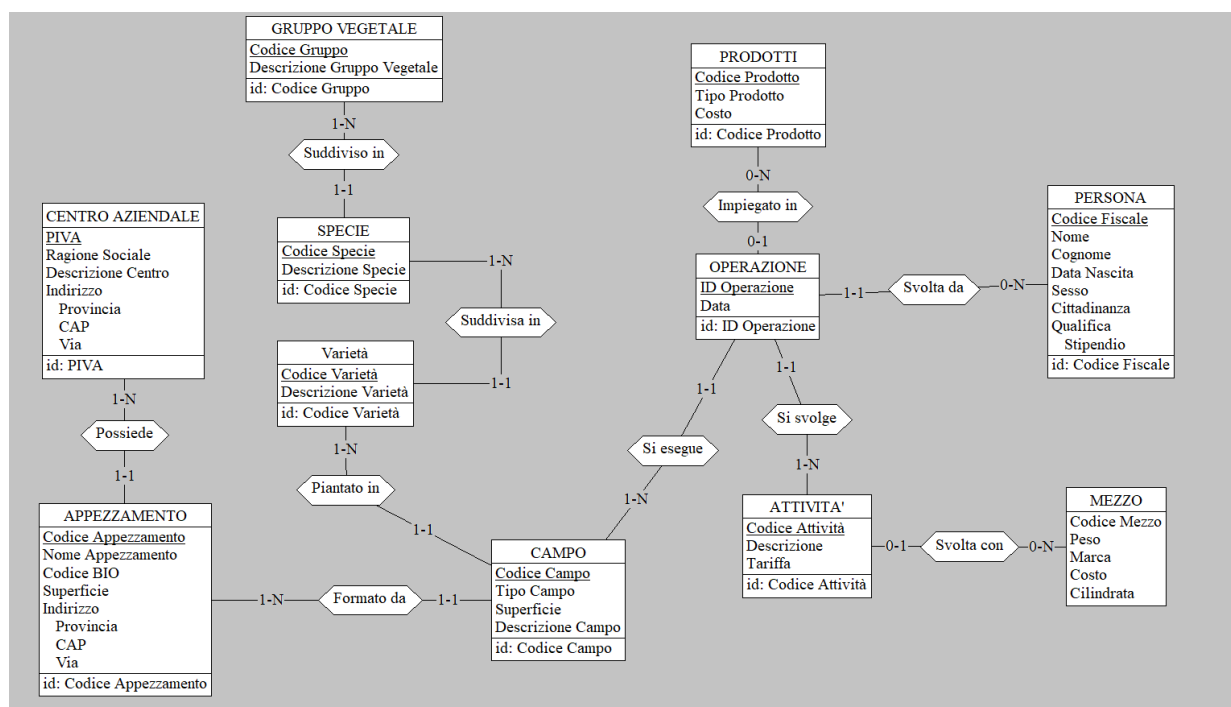


Figura 6: schema ER delle principali tabelle (semplificate)

3.1.1 Analisi e comprensione delle tabelle

Partendo da sinistra verso destra e dall'alto verso il basso si ha:

Centro Aziendale, *fabbricato o complesso di fabbricati situati nel perimetro dei terreni aziendali*

- PIVA: partita iva del centro in questione e identificatore della tabella
- Ragione sociale: indica il nome e il tipo della società
- Descrizione centro: una descrizione testuale che spesso indica l'ubicazione del centro in modo informale
- Indirizzo: ubicazione del centro aziendale
 - Provincia
 - CAP
 - Via

Appezamento, *porzione di terreno destinata a determinate coltivazioni*

- Codice appezzamento: codice usato come identificatore dell'appezzamento
- Nome Appezzamento: una descrizione testuale che spesso indica l'ubicazione dell'appezzamento in modo informale
- Codice BIO: codice usato per identificare se l'appezzamento è coltivato con metodi biologici ed eventualmente quali.
- Superficie: superficie in ettari dell'appezzamento
- Indirizzo: ubicazione dell'appezzamento
 - Provincia
 - CAP
 - Via

Campo, ulteriore suddivisione dell'appezzamento

- Codice Campo: codice usato come identificatore univoco del campo
- Tipo Campo: attributo binario per indicare se il campo si trova in pianura o in montagna.
- Superficie: superficie in ettari del campo (minore o uguale della superficie dell'appezzamento di appartenenza)
- Descrizione Campo: una descrizione testuale che spesso indica l'ubicazione del campo all'interno dell'appezzamento in modo informale

Gruppo Vegetale, distingue le colture in erbacee, arboree e prodotti di orto-floro vivaismo

- Codice gruppo: codice usato come identificatore univoco del gruppo vegetale
- Descrizione gruppo: descrizione testuale del gruppo vegetale

Specie, identifica ogni coltura all'interno di un gruppo vegetale (esempio: grano, menta ecc....)

- Codice Specie: codice usato come identificatore univoco della specie
- Descrizione specie: descrizione testuale della specie

Varietà, identifica ogni coltura all'interno della stessa specie con la sua varietà (esempio: grano tenero, grano saraceno ecc....)

- Codice Varietà: codice usato come identificatore univoco della varietà
- Descrizione varietà: descrizione testuale della varietà

Operazione, vengono eseguite sui campi e sono il centro dell'analisi in questo progetto di Business Intelligence. Identificato dall'attributo Codice operazione e la Data in cui è stata eseguita, tutti gli altri attributi sono poco significativi ai fini del progetto oppure ottenuti tramite join con le tabelle seguenti

Prodotti, qualunque cosa di esterno venga usato nelle operazioni viene fatto rientrare nella tabella prodotti, ad esempio fertilizzanti, sementi o pesticidi (poco presenti in quanto la maggior parte delle operazioni dell'azienda per la quale è rivolto il progetto usa metodi perlopiù biologici in ogni suo campo)

- Codice prodotto: codice usato come identificatore univoco del prodotto
- Tipo prodotto: tipologia di prodotto (sementi, fertilizzanti, ecc....)
- Costo: prezzo unitario per il prodotto in questione (unitario inteso come una confezione di prodotto o una unità di prodotto, dipendentemente dalla natura dello stesso)

Persona, ogni persona che lavora per l'azienda viene registrato nella tabella Persona (personalmente avrei gestito questa tabella con una gerarchia, ma non è questo il focus)

- Codice Fiscale: Codice fiscale utilizzato come identificatore univoco
- Nome: nome della persona
- Cognome: cognome della persona
- Data di Nascita: data di nascita della persona
- Sesso: maschile o femminile
- Cittadinanza:
- Qualifica: ruolo ricoperto dalla persona all'interno dell'azienda
 - Stipendio: associato alla qualifica

Attività, tutto ciò che può essere svolto in un'operazione è registrato nella tabella Attività

- Codice Attività: codice usato come identificatore univoco dell'attività
- Descrizione: descrizione testuale dell'attività, spesso associata alla macchina con cui viene svolta
- Tariffa: costo dell'attività

Mezzo, ogni mezzo registrato che può essere impiegato per eseguire un'attività

- Codice mezzo: codice usato come identificatore univoco del mezzo
- Marca: produttore del mezzo
- Peso: peso in kg del mezzo
- Cilindrata: dimensioni del motore del mezzo mm³
- Costo: prezzo per un'ora di utilizzo del mezzo

3.1.2 Discussione con il cliente

La discussione è avvenuta in azienda con Stefano Scattolin, tutor aziendale di tirocinio e leader del team di sviluppo software di Agronica, che prendeva il posto del cliente.

Lo scopo primario del progetto è ricreare un set di report relativi ad una specifica azienda (divisa in più centri aziendali), limitatamente alle operazioni svolte in campagna (ossia escludendo tutto il comparto zootecnico) utilizzando gli strumenti di supporto al Business Intelligence di Visual Studio e SQL Server per scoprire le potenzialità e quali pregi può portare all'azienda rispetto all'uso di Crystal Report, software ben rodato ed integrato nell'applicativo online di Agronica per visualizzare report.

Il fulcro del report dovrà composto da una tabella (previa selezione di vari filtri quali centro aziendale, anno, membri del personale specifici, ecc....) avente sulle righe le specie coltivate, con possibilità di andare in drill-down sulle varietà, sull'*impianto* (altro nome per Appezzamento), sull'*esercizio*, sull'operazione e infine sull'id del singolo record, e sulle colonne l'anno dove è possibile andare in drill-down

L'Esercizio è identificato da un codice univoco e consiste in un insieme di operazioni per ogni anno di attività dell'appezzamento. Questo è fatto in modo tale da constatare meglio i costi su programmi pluriennali sui vari appezzamenti, per esempio sapere se il costo di una singola operazione è riferita ad uno o più esercizi e quindi ammortizzare il costo su più anni.

Ogni operazione è composta da più record nel "tabellone", così come visto precedentemente nello schema ER che la tabella Operazione possiede associazioni verso varie tabelle (una per il personale, una per l'attività ed eventualmente il mezzo, una per ogni prodotto impiegato) ogni operazione viene sintetizzata a partire da tanti record riferiti ad un singolo aspetto dell'operazione messi insieme sul campo ID Agenda (che indica il codice dell'operazione nell'"Agenda", componente dell'applicativo online di Agronica per monitorare, appunto, le operazioni di campagna svolte).

In aggiunta a queste specifiche, sono richiesti anche alcuni grafici per arricchire il report di elementi visuali immediati.

Per conseguire questo obiettivo, si inizia creando un nuovo progetto tabulare su Visual Studio Analysis Service Projects.

3.2 Progettazione ETL

Siccome il modello verrà alimentato dal "tabellone" e non direttamente dalle singole tabelle (e anche date le mie scarse conoscenze sull'argomento e al tempo limitato), questa fase risulta un po' alterata dalla sua forma standard.

Inizialmente è stata creata una nuova partizione rispetto alla totalità della tabella.

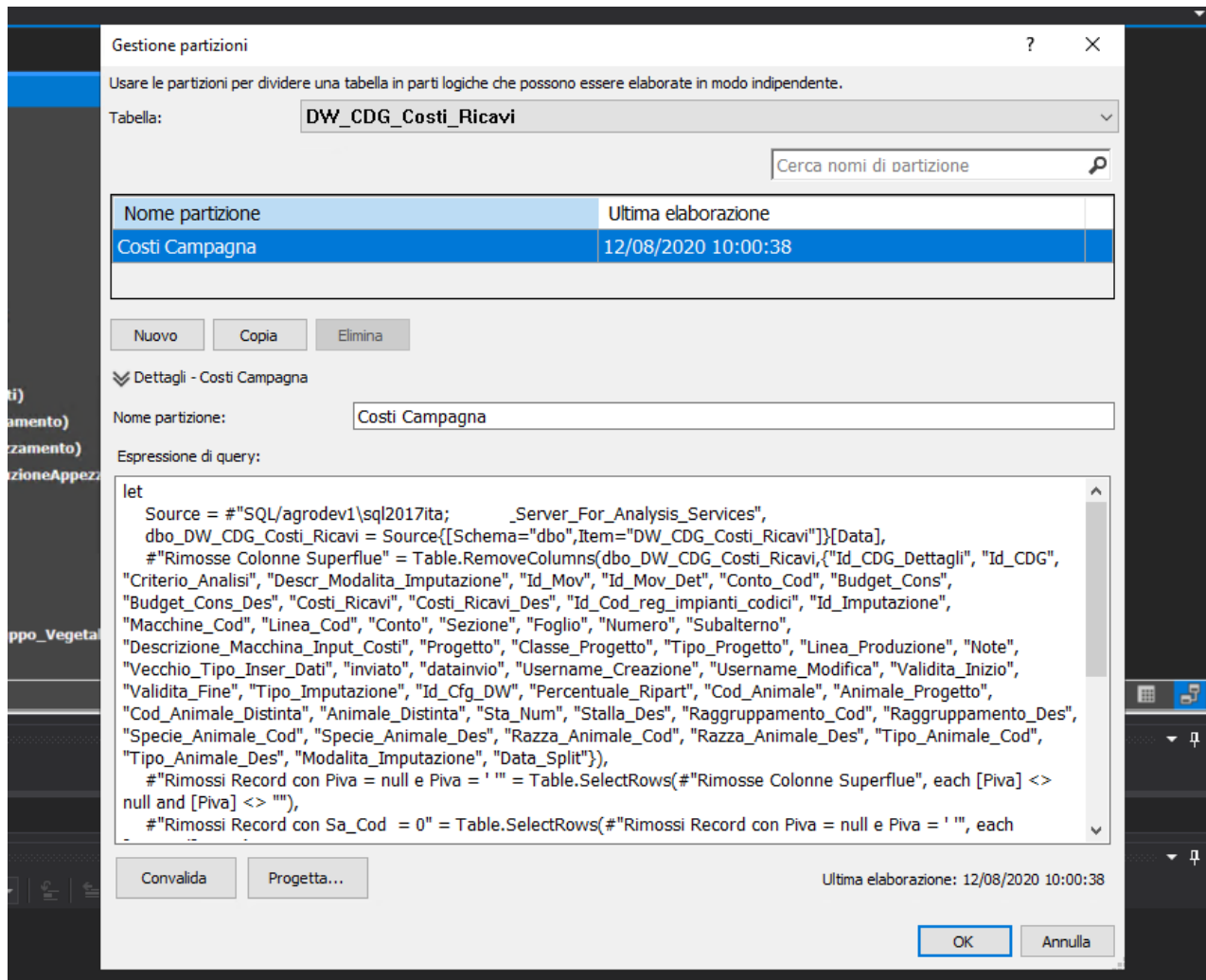


Figura 7: creazione partizione Costi Campagna

Utilizzando lo strumento di gestione partizioni fornito dall'estensione Analysis Service Projects, a partire dalla tabella “DW_CDG_Costi_Ricavi” sono state rimosse numerosi attributi come quelli relativi alle operazioni zootecniche o altri poco importanti per il progetto in questione.

Alla rimozione degli attributi superflui segue la fase di pulitura dei dati, ossia rimuovere tutti i record relativi ad operazioni zootecniche o con dati mancanti (come la partita IVA uguale a null).

Non mostrato in figura, i dati sono stati ordinati secondo ID_DW, ossia l'identificativo all'interno della tabella DW_CDG_Costi_Ricavi, e poi sono stati eliminati selettivamente alcuni record in cui sono stati riscontrati problemi con gli importi registrati durante lo sviluppo (il tempo a disposizione non era sufficiente per contattare l'azienda e correggere gli errori, dunque è stato scelto di rimuovere direttamente i record interessati).

Le istruzioni di query per eseguire queste operazioni sulla partizione sono state scritte automaticamente replicando ciò che viene eseguito nella schermata di progettazione della partizione, che fornisce un approccio più diretto sulla manipolazione dei dati.

Query [1]
= Table.SelectRows(#"Rimosse prime righe", each [Specie_impianto] <> null and [Specie_impianto] <> "")

	Data_Inserimento	Id_Agenda	Mac_Cod	Cod_RisUm	Elem_Cod	Pro_Cod	Mat_Cod	Id_Attivita	Qualifica_Cod
1	01/01/2020 00:00:00	71050	0	0	10	0	1	54	
2	04/01/2020 00:00:00	18291	0	0	10	0	120	149	
3	16/04/2020 00:00:00	57231	0	0	3	5348	0	45	
4	05/07/2019 00:00:00	14918	0	0	10	0	120	149	
5	12/07/2019 00:00:00	18177	0	0	10	0	173	149	
6	30/04/2019 00:00:00	5511	0	0	10	0	65	54	
7	02/05/2019 00:00:00	5727	0	0	10	0	65	54	
8	20/12/2019 00:00:00	18560	0	0	10	0	173	149	
9	20/12/2019 00:00:00	18560	0	0	10	0	173	149	
10	25/05/2020 00:00:00	63633	0	0	10	0	9	54	
11	09/09/2019 00:00:00	28724	0	0	10	0	5	54	
12	18/04/2020 00:00:00	57745	0	0	10	0	173	149	
13	15/10/2019 00:00:00	35312	0	0	10	0	198	149	
14	19/03/2020 00:00:00	52291	0	0	3	5348	0	45	
15	11/05/2019 00:00:00	6513	0	0	10	0	5	54	
16	04/03/2020 00:00:00	50025	0	0	3	8665	0	45	
17	20/03/2020 00:00:00	52421	0	0	3	5348	0	45	
18	12/03/2020 00:00:00	51429	0	0	3	8665	0	45	
19	14/06/2019 00:00:00	10793	0	0	3	8661	0	45	
20	14/06/2019 00:00:00	10677	0	0	10	0	65	54	
21	13/06/2019 00:00:00	10421	0	0	10	0	65	54	
22	16/05/2019 00:00:00	7989	0	0	10	0	65	54	
23	24/04/2020 00:00:00	58335	0	0	10	0	45	54	
24	24/07/2019 00:00:00	18563	0	0	10	0	173	149	
25	10/04/2020 00:00:00	56254	0	0	3	8661	0	45	
26	10/04/2020 00:00:00	56239	0	0	3	8661	0	45	
27	01/04/2020 00:00:00	63633	0	0	3	5348	0	45	

Impostazioni query

PROPRIETÀ

Nome
DW_CDG_Costi_Ricavi/Costi Campagna

Tutte le proprietà

PASSAGGI APPLICATI

Source

Navigazione

Rimosse Colonne Superflue

Rimosse Record con Piva = nul...

Rimosse Record con Sa_Cod = ...

Rimosse Record con Apprezza ...

Rimosse Record con Id_Reg = 0

Rimosse Record con Progetto...

Ordinate righe

Rimosse prime righe

Rimosse prime righe1

Filtrate righe

Figura 8: finestra progettazione della partizione

Questa finestra permette di lavorare sulla tabella in questione in maniera molto simile a Microsoft Excel, consentendo un approccio intuitivo e abbastanza user friendly.

Al lato destro si può notare una lista di azioni eseguite sulla tabella, ognuna annullabile o personalizzare a posteriori (con possibili ripercussioni sulle azioni eseguite dopo di essa), feature ampiamente usata in fase di sviluppo per aggiustare dati importati.

