

this:

POLITECNICO DI TORINO

Corso di Laurea Magistrale in
Ingegneria Informatica
Tesi di Laurea Magistrale



**Politecnico
di Torino**

Migrazione di una piattaforma di Business
Intelligence da soluzioni proprietarie a tecnologie
open source: un caso di studio nella Pubblica
Amministrazione

Relatore:

Prof. Luca Ardito

Laureando:

Giorgio Rondinone

Anno accademico 2024/2025

Sommario

La presente tesi analizza il processo di migrazione di una piattaforma di Business Intelligence da soluzioni proprietarie a tecnologie open source, valutandone benefici, criticità e implicazioni nel contesto della Pubblica Amministrazione. Il lavoro si inserisce nell'ambito dei sistemi informativi per l'Agricoltura e lo Sviluppo Rurale della Regione Basilicata.

La trattazione si apre con un inquadramento teorico della Business Intelligence, della sua evoluzione dal modello tradizionale centralizzato a quello moderno interattivo e data-driven, e delle specificità legate alla sua adozione nella Pubblica Amministrazione, con particolare attenzione ai temi della sostenibilità economica e dell'indipendenza tecnologica.

Successivamente viene analizzato il sistema di Business Intelligence attualmente in uso (AS-IS), descrivendone il contesto applicativo, l'architettura tecnica e i principali limiti, sia dal punto di vista tecnologico sia in termini di usabilità e flessibilità. Sulla base di tale analisi e della raccolta dei requisiti funzionali e non funzionali, viene definito il modello della nuova piattaforma (TO-BE), individuando criteri di valutazione oggettivi per la selezione della soluzione open source più idonea.

La tesi prosegue con un'analisi comparativa delle principali piattaforme open source di Business Intelligence, tra cui Apache Superset, Metabase, Redash, Pentaho e Jaspersoft, valutate secondo criteri tecnici, funzionali, economici e di governance. A seguito della comparazione, Apache Superset è stato selezionato per la realizzazione di un Proof of Concept.

Viene quindi descritta in dettaglio la fase di implementazione del prototipo, comprendente l'installazione e configurazione dell'ambiente applicativo, l'integrazione con database Oracle mediante viste materializzate, la realizzazione di dashboard

interattive, la gestione dei ruoli e della Row-Level Security, nonché la sperimentazione di funzionalità avanzate quali filtri dinamici e visualizzazioni geografiche. La fase di test ha consentito di verificare la copertura dei requisiti definiti e di validare la fattibilità tecnica della migrazione.

Il lavoro si conclude con una valutazione complessiva dei risultati ottenuti e con l'analisi dei prerequisiti necessari per la messa in produzione, con particolare riferimento agli aspetti di sicurezza e autenticazione istituzionale (SPID o CIE). I risultati dimostrano la sostenibilità tecnica della migrazione e forniscono linee guida operative per un futuro progetto di adozione su larga scala in ambito pubblico.

Indice

Elenco delle figure	7
1 Introduzione e contesto	10
1.1 Introduzione alla Business Intelligence	10
1.2 Come funziona la business intelligence	10
1.3 Vantaggi della business intelligence	11
1.4 Differenza tra Business Intelligence tradizionale e moderna	12
1.5 La Business Intelligence nella Pubblica amministrazione	13
1.6 Scopo della tesi e motivazioni	14
1.7 Software proprietari vs Open-Source	16
1.7.1 Costi iniziali e licenza	16
1.7.2 Manutenzione e supporto	17
1.7.3 Formazione	17
1.7.4 Scalabilità	17
1.7.5 Sicurezza e aggiornamenti	18
1.7.6 Conclusione generale	18
2 Il sistema attuale [AS-IS]	19
2.1 Descrizione del contesto	19
2.2 Architettura attuale	20
2.2.1 Livello dei dati e sorgenti informative	21

2.2.2	Processi di trasformazione e caricamento dei dati	22
2.2.3	Database decisionale e organizzazione dei dati	22
2.2.4	Livello di presentazione e strumenti di reporting	22
2.2.5	Gestione degli accessi e modalità di utilizzo	24
2.3	Limiti e problemi attuali	25
2.3.1	Rigidità architetturale e dipendenza da strumenti proprietari	25
2.3.2	Complessità dei flussi di elaborazione dei dati	25
2.3.3	Limitata usabilità dell'interfaccia grafica	26
2.3.4	Impatti sul contesto della Pubblica Amministrazione	27
3	Requisiti della nuova piattaforma di Business Intelligence(TO-BE)	29
3.1	Requisiti della piattaforma	29
3.1.1	Requisiti funzionali	30
3.1.2	Requisiti tecnici	31
3.2	Criteri di valutazione delle soluzioni	32
4	Analisi comparativa delle soluzioni open source di Business Intelli-	
	gence	35
4.1	Introduzione e obiettivi dell'analisi	35
4.1.1	Contesto e motivi della comparazione	36
4.1.2	Obiettivi della fase di analisi	36
4.1.3	Metodologia di valutazione e criteri adottati	37
4.2	Panoramica delle soluzioni open source considerate	38
4.3	Descrizione generale delle piattaforme analizzate	39
4.3.1	Apache Superset	39
4.3.2	Metabase	41
4.3.3	Redash	43
4.3.4	Pentaho	44
4.3.5	Jaspersoft	46

4.3.6	Tabelle comparative	47
4.3.7	Analisi critica: punti di forza e limiti delle soluzioni	48
4.4	Discussione e selezione della soluzione più idonea	50
5	Realizzazione del proof of concept	51
5.1	Architettura e setup di Apache Superset	51
5.1.1	Scelte architetturali iniziali	51
5.1.2	Installazione nativa su Ubuntu tramite WSL	52
5.1.3	Ambiente software e dipendenze	53
5.2	Connessione ai dati	54
5.2.1	Descrizione del database Oracle GECO	54
5.2.2	Struttura logica dei dati e dominio informativo	55
5.2.3	Preparazione dei dati per l'analisi	56
5.2.4	Configurazione della connessione in Superset	57
5.3	Riproduzione dei casi d'uso reali	59
5.3.1	Mappatura AS-IS e PoC	59
5.3.2	Creazione dei grafici e utilizzo dell'editor	60
5.3.3	Dashboard di monitoraggio e grafici	62
5.3.4	Rappresentazioni tabellari dei dati	64
5.3.5	Visualizzazioni geografiche e mappe	65
5.3.6	Filtri, interazione e navigazione dei dati	67
5.3.7	Gestione degli utenti, ruoli e sicurezza	70
5.3.8	Esportazione e condivisione dei dati	73
5.4	Risultati dei test e valutazioni	73
5.4.1	Test funzionali	74
5.4.2	Test di usabilità	74
5.4.3	Limiti emersi nella fase di test	75

6	Conclusione e sviluppi futuri	77
6.1	Considerazioni conclusive sul lavoro svolto	77
6.2	Prerequisiti per la messa in produzione	78
6.2.1	Sicurezza dell'applicazione	79
6.2.2	Utilizzo di un WSGI server per l'ambiente di produzione . .	79
6.2.3	Autenticazione istituzionale	80
6.2.4	Monitoraggio, logging e auditing	80
6.2.5	Backup, disaster recovery e continuità operativa	80
6.3	Integrazione con sistemi di autenticazione istituzionali (SPID) . . .	81
6.4	Sviluppi futuri e linee guida per una migrazione completa	82

Elenco delle figure

2.1	Diagramma dell'infrastruttura BI	21
2.2	Pannello di modifica query	23
2.3	Esempio di report in versione tabellare	24
2.4	Menù di progettazione e parte dell'interfaccia di consultazione dei report	27
4.1	Esempio di dashboard in superset	40
4.2	Editor sql in superset	41
4.3	Editor di domande in Metabase. Fonte: [8]	43
4.4	Caption	44
4.5	Interfaccia di Pentaho: ambiente di analisi e reportistica	46
5.1	Schermata per la configurazione della connessione del database in Superset	59
5.2	Editor per la creazione dei grafici	62
5.3	Presentazione della dashboard: overview e divisione in tab tematiche	64
5.4	Caption	65
5.5	Editor grafico per la creazione di mappe geografiche	67
5.6	Sezione per la selezione dei filtri strutturati	69
5.7	Operazioni permessi ai ruoli Gamma, da cui sono stati estesi i ruoli personalizzati	71

5.8	Schermata di definizione delle regole per la RLS	72
5.9	Schermata di log in Superset	73

Capitolo 1

Introduzione e contesto

1.1 Introduzione alla Business Intelligence

Parliamo di business intelligence (BI) [2] come un insieme di processi per la raccolta, gestione e analisi dei dati per ottenere insight che forniscano informazioni sulle strategie e le operazioni aziendali.

Viene utilizzata, quindi, per trasformare i dati non elaborati in dati significativi, in modo da guidare un buon processo decisionale all'interno di un'organizzazione. Secondo la rivista CIO: "Sebbene la business intelligence non dica agli utenti business cosa fare o cosa accadrà se seguono un determinato corso, la BI non si limita a generare report, ma offre alle persone un modo per esaminare i dati al fine di comprendere le tendenze e ricavare insight."

1.2 Come funziona la business intelligence

Le piattaforme di BI si affidano ai data warehouse per le informazioni di base. Lo scopo di un data warehouse è quello di aggregare i dati provenienti da più fonti in un unico sistema centrale per supportare l'analytics dei dati aziendali e la reportistica.

La BI presenta i risultati all'utente sotto forma di report, grafici e mappe, che possono essere visualizzati attraverso una dashboard.

I passaggi eseguiti in BI sono solitamente i seguenti:

- **Fonti di dati:** individuare i dati da esaminare e analizzare da, ad esempio, un data warehouse.
- **Raccolta dei dati:** raccolta e pulizia dei dati da varie fonti. Questa preparazione potrebbe consistere nella raccolta manuale di informazioni o in un processo più complesso di estrazione, trasformazione e caricamento (ETL).
- **Analisi:** ricerca di tendenze o risultati inaspettati nei dati. Questo potrebbe richiedere l'utilizzo di strumenti di data mining, data discovery o data modeling.
- **Visualizzazione:** visualizzare i dati tramite grafici o dashboard utilizzate da strumenti di business intelligence come quelli che verranno approfonditi nei successivi capitoli.
- **Piano d'azione:** sviluppo di insight utili basati sull'analisi dei dati storici. Queste azioni potrebbero includere processi più efficienti, cambiamenti nel marketing, adattamento dei problemi nell'ambito dell'esperienza del cliente...

1.3 Vantaggi della business intelligence

- **Reportistica più chiara:** la BI offre la possibilità di porsi delle domande e ottenere delle risposte facili da comprendere. La dashboard consente di dare priorità alle informazioni più importanti, facendo risparmiare tempo agli utenti.
- **Dati consolidati:** la BI è in grado di consolidare i dati da più fonti, interne o esterne, per un'analisi più approfondita e completa. Essa offre alle organizzazioni i mezzi per progettare una strategia aziendale.

- **Processo decisionale più rapido:** Con una visualizzazione e una analisi digitale più intuitiva e completa dei dati, è possibile prendere decisioni informate più rapidamente.
- **Aumento della soddisfazione del cliente:** il personale del servizio clienti può accedere più rapidamente e più efficacemente ai dati.
- **Aumento della soddisfazione dei dipendenti:** anche il resto del personale può svolgere più rapidamente e facilmente il proprio lavoro.

1.4 Differenza tra Business Intelligence tradizionale e moderna

Prima della diffusione dei sistemi informatici e delle piattaforme software dedicate, il concetto di Business Intelligence era principalmente riconducibile ad attività di analisi competitiva e raccolta strutturata di informazioni, svolte da un numero ristretto di figure specializzate. In questa fase iniziale, l'accesso ai dati e la loro interpretazione erano limitati e fortemente dipendenti da competenze specifiche.

L'approccio tradizionale alla Business Intelligence si è poi sviluppato secondo un modello di tipo top-down, in cui i reparti IT ricoprivano un ruolo centrale nella gestione dei dati e nella produzione di report. Le informazioni venivano generalmente distribuite attraverso report statici e predefiniti, con cicli di elaborazione lunghi e poco flessibili. Questo modello, sebbene ancora diffuso per attività di reporting periodico e per interrogazioni standardizzate, rendeva complesso l'utilizzo dei dati a supporto di decisioni rapide e tempestive.

Con l'evoluzione delle tecnologie e delle architetture software, si è affermato un paradigma di Business Intelligence più moderno, orientato all'interattività e all'accessibilità. Le piattaforme di BI contemporanee consentono agli utenti di personalizzare dashboard, esplorare i dati in modo dinamico e creare report in autonomia, riducendo la dipendenza esclusiva dai reparti IT. In questo nuovo scenario, il ruolo

dell'IT rimane fondamentale per garantire la qualità, la sicurezza e la governance del dato, ma viene affiancato da strumenti che favoriscono il **self-service analytics**.

Nei primi anni di diffusione della Business Intelligence, l'utilizzo di dati aggiornati in tempo quasi reale per supportare decisioni operative risultava difficilmente realizzabile. Oggi, al contrario, la Business Intelligence moderna si configura come un **processo continuo e interattivo**, in cui le informazioni sono facilmente accessibili, comprensibili e fruibili anche da utenti non esperti. Questa evoluzione risponde alle esigenze delle organizzazioni moderne, che richiedono strumenti in grado di fornire rapidamente informazioni affidabili e di supportare in modo efficace i processi decisionali strategici [7].

1.5 La Business Intelligence nella Pubblica amministrazione

I sistemi di business intelligence sono determinanti per la crescita economica. In questo ambito gli enti pubblici [1] assumono un'importanza rilevante per i servizi digitali erogati per gli utenti. La trasformazione digitale della Pubblica Amministrazione rappresenta oggi una opportunità per offrire servizi più efficienti, accessibili e orientati alle esigenze dei cittadini. Tuttavia, il vero passo in avanti riguarda la capacità di valorizzare i dati pubblici, rendendoli comprensibili e utilizzabili in un'ottica **data-driven**.

Nonostante i progressi compiuti negli ultimi anni, l'Italia mostra ancora un ritardo rispetto ad altri Paesi in termini di utilizzo di strumenti di BI. Secondo il Digital Government Index (DGI) dell'OCSE, il nostro Paese si colloca al di sotto della media per l'adozione di sistemi capaci di analizzare, classificare e trasformare i dati in informazioni strategiche.

Eppure, fondare le decisioni pubbliche, la gestione dei processi e la pianificazione dei servizi su un uso consapevole e sistematico dei dati rappresenta una condizione indispensabile per vincere la sfida della digitalizzazione e efficienza amministrativa. In questo scenario, la Business Intelligence assume un ruolo centrale, poiché consente alla Pubblica Amministrazione di trasformare i dati grezzi in conoscenza utile, supportando politiche e decisioni basate su evidenze concrete.

1.6 Scopo della tesi e motivazioni

Questa tesi nasce come progetto di innovazione tecnologica volta a favorire la transizione digitale della Pubblica Amministrazione attraverso la sostituzione di soluzioni proprietarie di Business Intelligence con soluzioni software aperte e sostenibili. L'obiettivo principale è la realizzazione di un **Proof of Concept** di una piattaforma di BI con tecnologie open source, con lo scopo di valutare la fattibilità funzionale e tecnica di una migrazione da un sistema proprietario attualmente in uso da parte della Regione Basilicata.

La scelta di approfondire questa tematica è motivata da vari fattori concreti e vantaggiosi per la Pubblica Amministrazione:

- **Riduzione dei costi e sostenibilità economica:** l'utilizzo di software open source permette di eliminare i costi di licenza delle soluzioni proprietarie e ridurre le spese di manutenzione.
- **Allineamento alle politiche nazionali:** le linee guida dell'Agenzia per l'Italia Digitale promuovono la diffusione di software liberi e il riuso delle soluzioni nella Pubblica Amministrazione.
- **Innovazione e modernizzazione:** le piattaforme open source di Business Intelligence, come Apache Superset, Metabase e altre (che verranno viste in seguito) offrono architetture moderne, interfacce intuitive e funzionalità efficienti di visualizzazione e analisi dei dati.

- **Indipendenza e interoperabilità:** adottare soluzioni open source favorisce una migliore integrazione tra diversi sistemi informativi e una maggiore interoperabilità dei dati.

Il lavoro di tesi si occupa quindi di:

1. **analizzare lo stato attuale del sistema** di Business Intelligence utilizzato nel contesto della Pubblica Amministrazione (AS-IS): studio del sistema di BI esistente, identificazione delle principali informazioni e funzionalità fornite agli utenti, raccolta di criticità e punti di miglioramento.
2. **Raccolta requisiti aggiuntivi:** individuare requisiti evolutivi, funzionali e tecnici al fine di definire il software TO-BE.
3. individuare, analizzare e **confrontare le principali alternative open source** disponibili in base a caratteristiche, funzionalità, limiti tecnici, integrazione con DB esistenti, costi, licenze, gestione della sicurezza e usabilità.
4. **selezionare la soluzione più adatta** a costruire la nuova piattaforma, considerando sia i requisiti tecnici sia le esigenze operative della pubblica amministrazione.
5. **Predisposizione dell'ambiente di test e installazione:** configurazione di un ambiente demo per la piattaforma selezionata, installazione, collegamento a dei dataset di prova e verifica della corretta funzionalità dei moduli principali.
6. **Progettazione e sviluppo del prototipo:** creazione di dashboard e report, integrazione di eventuali funzionalità aggiuntive, test di performance e usabilità.
7. **Validazione del prototipo:** presentazione agli utenti finali di scenari reali replicati dal sistema AS-IS, raccolta di feedback e suggerimenti, eventuali correzioni e miglioramenti.
8. **Analisi dei risultati e documentazione finale** tramite una sintesi dei risultati ottenuti, un confronto tra i costi operativi AS-IS e TO-BE e stesura

della documentazione della piattaforma e delle attività svolte.

L'obiettivo finale è fornire un **caso di studio concreto** che possa rappresentare un modello di riferimento per progetti futuri con lo stesso scopo, evidenziando benefici, criticità e impatto economico. In particolare, l'analisi sarà focalizzata sul dominio **Agricoltura e Sviluppo Rurale**, dove la disponibilità di strumenti di analisi appropriati ricopre un ruolo fondamentale.

1.7 Software proprietari vs Open-Source

Nel contesto aziendale, ridurre i costi legati all'infrastruttura IT mantenendo al tempo stesso un buon livello di funzionalità rappresenta un'esigenza fondamentale. Da qui nasce una domanda: conviene adottare una piattaforma open source oppure continuare a utilizzare soluzioni proprietarie soggette a licenze o abbonamenti [9]? In questa sezione vengono messi a confronto i principali vantaggi e svantaggi di entrambe le alternative. Considererem vari aspetti come i costi di gestione, la manutenzione e la possibilità di scalabilità del sistema.

1.7.1 Costi iniziali e licenza

Uno dei maggiori vantaggi delle soluzioni Open-Source è l'assenza di costi di licenza. Al contrario, molte piattaforme proprietarie prevedono il pagamento di licenze onerose o abbonamenti annuali, soprattutto nel caso di software complessi o suite per l'ufficio. L'open source offre quindi un grande risparmio sui costi iniziali, dettaglio che può fare la differenza soprattutto per le piccole imprese.

1.7.2 Manutenzione e supporto

Il risparmio ottenuto eliminando i costi di licenza può essere compensato dalle spese legate alla manutenzione dei software Open-Source. L'installazione e la configurazione di tali sistemi richiedono infatti competenze tecniche specifiche. Sebbene le community open source offrano un valido aiuto, ottenere un'assistenza tempestiva per problemi più critici può risultare più complesso in assenza di un servizio dedicato. Al contrario, i software proprietari includono spesso contratti di supporto che garantiscono interventi rapidi e professionali.

1.7.3 Formazione

La formazione del personale per software open-source può comportare dei costi e richiedere adattamento, soprattutto se i dipendenti sono abituati a soluzioni Microsoft o altre suite proprietarie. In realtà, però, molti software commerciali sono già noti e intuitivi, riducendo il bisogno di una formazione approfondita e favorendo un adattamento più rapido.

1.7.4 Scalabilità

La scalabilità rappresenta un elemento chiave per le aziende in fase di crescita. Le piattaforme open-source offrono ottime prestazioni e una buona capacità di adattamento a contesti in espansione. Tuttavia, la gestione di configurazioni avanzate può diventare impegnativa senza un team IT specializzato. Le soluzioni proprietarie, invece, sono generalmente pensate per garantire una scalabilità più immediata e semplice da implementare, ma comportano un incremento dei costi di licenza proporzionale all'aumento delle dimensioni e delle esigenze dell'organizzazione.

1.7.5 Sicurezza e aggiornamenti

Nel contesto open-source, la trasparenza del codice costituisce un importante vantaggio in termini di sicurezza: la possibilità per la community di analizzare il codice consente di individuare e correggere rapidamente eventuali vulnerabilità. Tuttavia, questo approccio richiede un monitoraggio costante e competenze adeguate per gestire aggiornamenti e applicare patch in modo tempestivo nel caso in cui ce ne fosse bisogno. Le soluzioni proprietarie, al contrario, dispongono di team dedicati alla sicurezza che rilasciano aggiornamenti periodici, garantendo una protezione più centralizzata, sebbene con minore flessibilità e una possibile esposizione a minacce mirate. In sintesi sia i sistemi open-source che quelli proprietari possono offrire standard di sicurezza elevati, ma le piattaforme open source richiedono un controllo continuo e una gestione più proattiva.

1.7.6 Conclusione generale

Concludendo il confronto tra le due tipologie di soluzioni software, si può dichiarare che per le piccole imprese, l'open-source può rappresentare una valida opzione per abbattere i costi di licenza, soprattutto in presenza di risorse o consulenti IT esperti. Tuttavia, senza un team dedicato alla gestione di questa infrastruttura, i risparmi potrebbero ridursi a causa dei vari costi di manutenzione e formazione del personale. Se l'azienda cresce rapidamente, l'open-source potrebbe rimanere economicamente sostenibile, richiedendo però una gestione sempre più complessa e strutturata. Le soluzioni proprietarie, invece, sono più intuitive ma i costi possono aumentare notevolmente con l'espansione aziendale.

In definitiva, la scelta più corretta dipende dalle specifiche necessità di ogni impresa, dal budget a disposizione e dal livello di competenze tecniche interne ad esse.

Capitolo 2

Il sistema attuale [AS-IS]

2.1 Descrizione del contesto

Il sistema di Business Intelligence oggetto di analisi opera all'interno del contesto informativo della Regione Basilicata, con particolare riferimento ai processi di monitoraggio e controllo relativi ai programmi di sviluppo rurale e alle misure di sostegno in ambito agricolo. In tale contesto, la disponibilità di strumenti di analisi affidabili e strutturati rappresenta un elemento fondamentale per supportare le attività di rendicontazione, verifica e pianificazione richieste a livello regionale, nazionale ed europeo.

Il sistema informativo di riferimento è il **SIARB** (Sistema Informativo Agricoltura Regione Basilicata), che raccoglie e gestisce i dati provenienti dai diversi moduli applicativi utilizzati per l'amministrazione degli interventi nel settore agricolo e dello sviluppo rurale. All'interno di questo ecosistema informativo, il modulo di **monitoraggio** svolge un ruolo centrale, in quanto consente di trasformare i dati operativi presenti nei sistemi gestionali in informazioni sintetiche e strutturate, fruibili sotto forma di reportistica a supporto delle attività decisionali e di controllo.

Le esigenze di monitoraggio derivano principalmente dall'attuazione dei **Programmi di Sviluppo Rurale (PSR)**, che prevedono la gestione di un elevato

numero di misure e interventi, caratterizzati da complessità normativa e da un forte impatto economico. In particolare, il monitoraggio delle cosiddette misure a superficie richiede l'elaborazione periodica di dati relativi a domande, beneficiari, superfici dichiarate e pagamenti, al fine di garantire la correttezza delle informazioni trasmesse agli organismi di controllo e agli enti sovraordinati.

In linea con soluzioni già adottate in altre realtà regionali, l'architettura del sistema prevede l'utilizzo di un database decisionale e di strumenti dedicati alla trasformazione e alla presentazione dei dati, con l'obiettivo di fornire una reportistica sintetica e strutturata a partire dai dati presenti nei sistemi operativi del SIARB.

Il contesto operativo è caratterizzato dalla presenza di **differenti tipologie di utenti**, con ruoli e responsabilità eterogenee. Il sistema di monitoraggio è utilizzato sia da personale tecnico-amministrativo incaricato delle attività di controllo e rendicontazione, sia da figure con compiti di supervisione e analisi, che necessitano di una visione aggregata e affidabile dei dati. Questo comporta la necessità di garantire un accesso controllato alle informazioni e di fornire strumenti in grado di supportare diverse modalità di consultazione dei dati.

Nel tempo, il modulo di monitoraggio è stato oggetto di **attività evolutive** finalizzate ad ampliare la copertura dei moduli applicativi analizzati e ad adeguare la reportistica alle nuove esigenze informative. Queste evoluzioni hanno interessato, tra le altre, l'integrazione di nuovi dataset, la predisposizione di procedure di alimentazione del data warehouse e la realizzazione di nuovi report, evidenziando la crescente importanza della Business Intelligence come componente strategica del sistema informativo regionale.

2.2 Architettura attuale

Dal punto di vista generale, l'architettura AS-IS può essere ricondotta a una **struttura multilivello**, composta da un insieme di componenti dedicati alla gestione dei

dati operativi, alla loro trasformazione e consolidamento, e infine alla presentazione delle informazioni sotto forma di report e analisi consultabili dagli utenti finali. Tale architettura è interamente basata su tecnologie proprietarie ed è installata in ambiente on-premise.

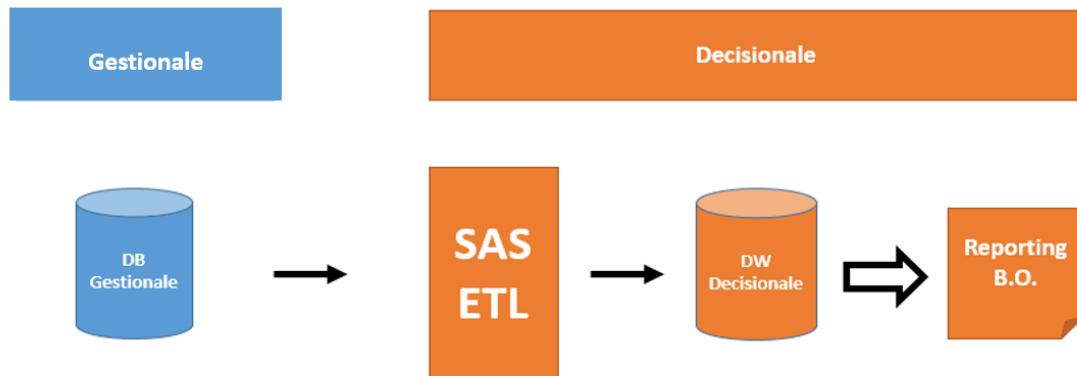


Figura 2.1. Diagramma dell'infrastruttura BI

2.2.1 Livello dei dati e sorgenti informative

Il livello più basso dell'architettura è rappresentato dalle sorgenti dati operative, costituite dai sistemi gestionali che alimentano il SIARB. In particolare, i dati utilizzati per le attività di monitoraggio derivano dai moduli applicativi che gestiscono le domande di aiuto, i beneficiari, le superfici dichiarate e gli importi relativi alle misure del Programma di Sviluppo Rurale.

Una parte significativa dei dati di input è acquisita attraverso file strutturati, prevalentemente in formato Excel, messi a disposizione dai sistemi nazionali di riferimento. Tali file contengono informazioni di dettaglio necessarie per il monitoraggio delle misure e vengono periodicamente elaborati per l'alimentazione del sistema di Business Intelligence. Questa modalità di acquisizione introduce una prima fase di trasformazione dei dati, necessaria per rendere le informazioni coerenti e utilizzabili a fini analitici.

2.2.2 Processi di trasformazione e caricamento dei dati

La trasformazione dei dati e il loro caricamento nel database decisionale avvengono attraverso procedure batch sviluppate utilizzando strumenti dedicati. In particolare, i documenti analizzati evidenziano l'utilizzo di programmi specifici per la gestione delle attività di estrazione, trasformazione e caricamento (ETL), finalizzati alla normalizzazione dei dati e alla loro integrazione nel sistema di monitoraggio. Questi processi includono operazioni di pulizia, aggregazione e riconciliazione dei dati provenienti da sorgenti eterogenee, al fine di garantire la consistenza delle informazioni memorizzate nel database decisionale. Le procedure di caricamento sono eseguite con periodicità definita e risultano fortemente dipendenti da configurazioni tecniche e interventi specialistici, rendendo il sistema poco flessibile rispetto a modifiche o nuove esigenze informative.

2.2.3 Database decisionale e organizzazione dei dati

Il cuore del sistema di Business Intelligence AS-IS è rappresentato dal database decisionale, basato su tecnologia Oracle. Questo database ospita i dati consolidati utilizzati per le attività di analisi e reportistica, organizzati secondo strutture progettate per supportare interrogazioni di tipo analitico.

All'interno del database sono presenti viste e strutture di supporto, incluse viste materializzate, utilizzate per migliorare le prestazioni delle interrogazioni e per semplificare l'accesso ai dati da parte degli strumenti di reporting. L'organizzazione dei dati è orientata alla produzione di indicatori e aggregazioni necessarie per il monitoraggio delle misure.

2.2.4 Livello di presentazione e strumenti di reporting

Il livello di presentazione è realizzato tramite strumenti di Business Intelligence dedicati alla creazione e alla consultazione dei report. L'accesso alle funzionalità

di reporting avviene attraverso un portale centralizzato, che consente agli utenti autorizzati di visualizzare report predefiniti e analisi strutturate.

I report sono generalmente costruiti a partire da modelli semantici che astraggono la complessità del database sottostante, permettendo agli utenti di interrogare i dati senza dover interagire direttamente con il linguaggio SQL. Tuttavia, la creazione e la modifica di tali modelli richiedono competenze tecniche specifiche, limitando l'autonomia degli utenti finali e rendendo necessaria l'intermediazione del personale IT per l'evoluzione della reportistica.

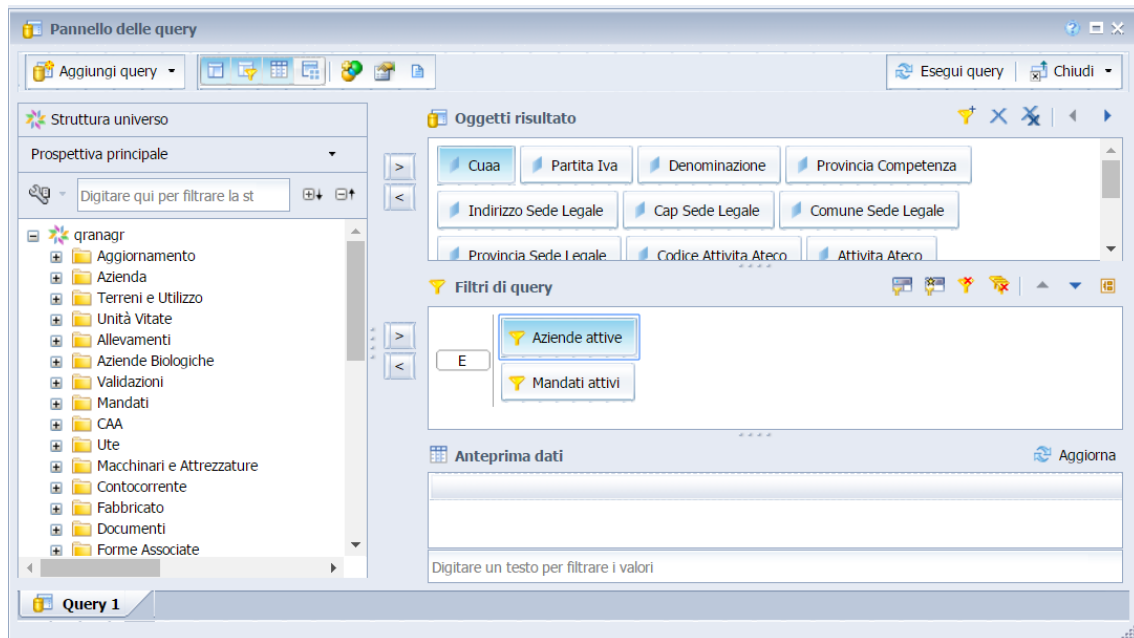


Figura 2.2. Pannello di modifica query

PSR 14/20 - Dettaglio anomalie sulle domande presentate

Anno Campagna	Anno	Mese	Codice Anomalia	Anomalia	Numero Anomalie	Numero Domande
2016	2018	03	A05-4	ASSENZA PEC	24	24
2016	2018	03	C23-03	PARTICELLA CON SUPERFICIE COMPLESSIVAMENTE DICHIARATA MAGGIORE DELLA SUPERFICIE ACCERTATA , TRA PIU' DOMANDE (SUPERO ESTERNO)	74	15
2016	2018	03	C27-01	PARTICELLA CON MANTENIMENTO NON RISCONTRATO	152	32
2016	2018	03	C28-01	PARTICELLA RISCONTRATA CON COLTURA NON ORDINARIA	53	22
2016	2018	03	C29-01	PARTICELLA DI DOMANDA SOTTOPOSTA AI CONTROLLI OGGETTIVI CON LAVORAZIONE NON DISPONIBILE IN LISTA CAMPIONE	9	6
2016	2018	03	F05-01	RITARDO (ENTRO I 25 GIORNI DI CALENDARIO)	3.530	3.530
2016	2018	03	F05-02	RITARDO (OLTRE I 25 GIORNI DI CALENDARIO)	4	4
2016	2018	03	ICO-05	SUPERFICIE NON AMMISSIBILE - SUPERFICI IDENTIFICATE CON "NESSUNA PRATICA" SUL PIANO DI COLTVAZIONE	84	84

Figura 2.3. Esempio di report in versione tabellare

2.2.5 Gestione degli accessi e modalità di utilizzo

L'accesso al sistema di Business Intelligence è regolato da meccanismi di autenticazione e autorizzazione coerenti con l'infrastruttura esistente. Gli utenti accedono alla piattaforma in base al proprio ruolo, con differenti livelli di visibilità sulle informazioni disponibili.

Dal punto di vista operativo, il sistema è prevalentemente orientato alla consultazione di report statici o semi-dinamici, utilizzati per attività di monitoraggio periodico e rendicontazione. Le possibilità di esplorazione interattiva dei dati risultano limitate e fortemente vincolate alla struttura dei report esistenti.

2.3 Limiti e problemi attuali

Nonostante il sistema di Business Intelligence attualmente in uso garantisca la copertura dei requisiti funzionali minimi per il monitoraggio e la rendicontazione delle misure PSR, dall'analisi dell'architettura, dei processi e delle modalità di fruizione emergono una serie di **criticità strutturali e operative**. Tali limiti incidono sia sull'efficienza dei processi interni, sia sull'usabilità del sistema da parte degli utenti finali, rendendo complessa l'evoluzione e la modernizzazione della piattaforma.

2.3.1 Rigidità architetturale e dipendenza da strumenti proprietari

Il sistema AS-IS si basa su una catena di strumenti fortemente integrati ma **dipendenti da tecnologie proprietarie**, in particolare per quanto riguarda:

- l'elaborazione dei dati tramite procedure dedicate;
- la fruizione delle informazioni attraverso una piattaforma di Business Intelligence non open source.

Questa dipendenza comporta:

- **costi elevati di manutenzione e licenza;**
- limitata flessibilità nell'introduzione di nuove funzionalità;
- difficoltà nell'integrazione con strumenti moderni o open source;
- necessità di competenze specialistiche non sempre disponibili.

L'architettura, pur risultando stabile, non è pensata per evolvere rapidamente né per supportare agevolmente approcci di tipo self-service o sperimentazione iterativa.

2.3.2 Complessità dei flussi di elaborazione dei dati

Dalla documentazione emerge come il processo di monitoraggio si basi su **flussi di trasformazione articolati**, che coinvolgono più passaggi intermedi e più livelli di

elaborazione prima di arrivare alla produzione del dato finale.

Questa complessità comporta:

- maggiore difficoltà nel tracciamento delle trasformazioni;
- aumento del rischio di errori o incoerenze;
- tempi di aggiornamento non immediati;
- forte dipendenza da personale tecnico per modifiche o manutenzioni.

Il sistema risulta quindi poco flessibile rispetto a nuove esigenze di analisi o variazioni nei requisiti di monitoraggio.

2.3.3 Limitata usabilità dell'interfaccia grafica

Uno dei limiti più rilevanti del sistema attuale riguarda la scarsa intuitività dell'interfaccia grafica di consultazione dei report.

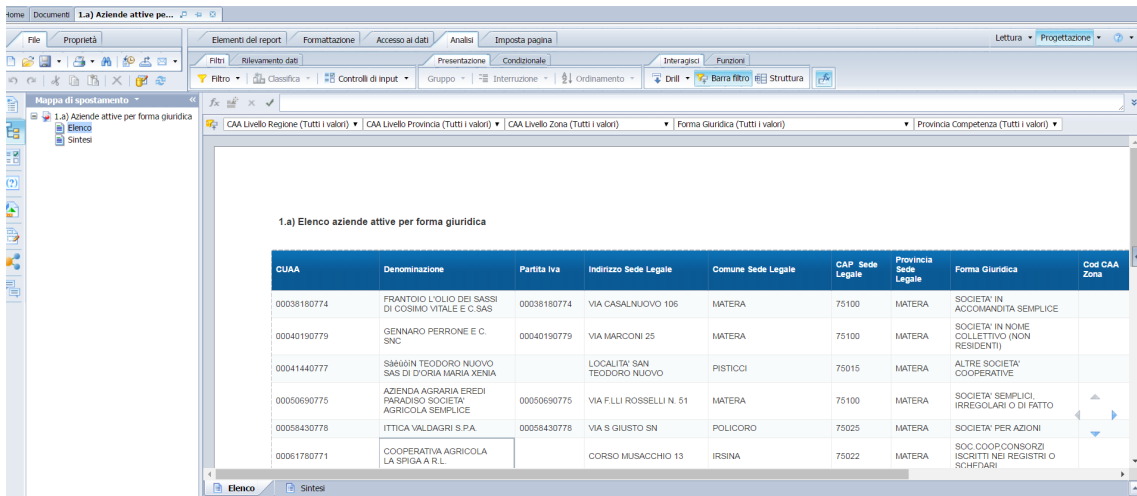
L'accesso e l'utilizzo delle funzionalità di reporting:

- non risultano immediatamente comprensibili per utenti non tecnici;
- necessitano del supporto di documentazione esterna per interpretare correttamente report, filtri e parametri.

L'interfaccia non guida l'utente nella comprensione dei dati, ma presuppone una familiarità con:

- la logica applicativa;
- le convenzioni adottate;
- la struttura sottostante del sistema.

2.3 – Limiti e problemi attuali



1.a) Elenco aziende attive per forma giuridica

CUAUA	Denominazione	Partita Iva	Indirizzo Sede Legale	Comune Sede Legale	CAP Sede Legale	Provincia Sede Legale	Forma Giuridica	Cod CAA Zona
00038180774	FRANTOIO L'OLIO DEI SASSI DI COSIMO VITALE E C SAS	00038180774	VIA CASALNUOVO 106	MATERA	75100	MATERA	SOCIETA' IN ACCOMANDITA SEMPLICE	
00040190779	GENNARO PERRONE E C SNC	00040190779	VIA MARCONI 25	MATERA	75100	MATERA	SOCIETA' IN NOME COLLETTIVO (NON RESIDENTI)	
00041440777	S&B&N TEODORO NUOVO SAS DI D'ORIO MARIANA		LOCALITA' SAN TEODORO NUOVO	PISTICCI	75015	MATERA	ALTRE SOCIETA' COOPERATIVE	
00050690775	AZIENDA AGRARIA EREDI PARADISO SOCIETA' AGRICOLA SEMPLICE	00050690775	VIA F.LLI ROSSELLI N. 51	MATERA	75100	MATERA	SOCIETA' SEMPLICI IRREGOLARI O DI FATTO	
00058430778	ITTICA VALDAGRI S.P.A.	00058430778	VIA S GIUSTO SN	POLICORDO	75025	MATERA	SOCIETA' PER AZIONI	
00061780771	COOPERATIVA AGRICOLA LA SPIGA A R.L.		CORSO MUSACCHIO 13	IRSINA	75022	MATERA	SOC.COOP.CONSORZI ISCRITTI NEI REGISTRI O SOCIETARI	

Figura 2.4. Menù di progettazione e parte dell'interfaccia di consultazione dei report

2.3.4 Impatti sul contesto della Pubblica Amministrazione

Nel contesto della Pubblica Amministrazione, tali limiti si traducono in:

- maggiore rigidità operativa;
- difficoltà di evoluzione del sistema nel tempo;
- minore trasparenza e fruibilità dei dati;
- ridotta capacità di adattamento a nuove normative o requisiti di monitoraggio.

Queste criticità motivano la necessità di valutare soluzioni alternative più flessibili e moderne.

Capitolo 3

Requisiti della nuova piattaforma di Business Intelligence(TO-BE)

3.1 Requisiti della piattaforma

La definizione dei requisiti della nuova piattaforma di Business Intelligence rappresenta una fase fondamentale del processo di progettazione del sistema TO-BE. In un contesto complesso come quello della Pubblica Amministrazione, caratterizzato da sistemi informativi eterogenei, vincoli normativi stringenti e una pluralità di utenti con competenze differenti, l'individuazione accurata dei requisiti consente di guidare in modo strutturato sia la fase di analisi comparativa delle soluzioni disponibili sia le successive attività di implementazione del Proof of Concept.

I requisiti individuati derivano da una combinazione di fattori: l'analisi preliminare delle funzionalità offerte dal sistema di Business Intelligence attualmente in uso (AS-IS), la raccolta di esigenze evolutive espresse dagli utenti finali, nonché le necessità di adeguamento tecnologico e organizzativo legate ai processi di digitalizzazione della Pubblica Amministrazione. L'obiettivo non è quello di replicare fedelmente il sistema esistente, bensì di definirne un'evoluzione capace di migliorare l'esperienza d'uso, aumentare la flessibilità operativa e ridurre i costi complessivi

di gestione nel medio-lungo periodo.

I requisiti della piattaforma TO-BE possono essere suddivisi in due macro-categorie: **requisiti funzionali**, che descrivono le funzionalità attese dal punto di vista degli utenti, e **requisiti tecnici**, che riguardano l'architettura, l'integrazione con i sistemi esistenti e le caratteristiche infrastrutturali della soluzione

3.1.1 Requisiti funzionali

Dal punto di vista funzionale, la nuova piattaforma di Business Intelligence deve essere in grado di supportare in modo efficace le attività di analisi, monitoraggio e reporting tipiche del dominio applicativo considerato, in particolare nell'ambito dell'Agricoltura e dello Sviluppo Rurale. Uno dei requisiti principali riguarda la capacità di realizzare **dashboard interattive**, che consentano agli utenti di visualizzare indicatori chiave di performance, dati aggregati e trend temporali attraverso rappresentazioni grafiche intuitive e personalizzabili. Le dashboard devono poter essere adattate a differenti profili di utenza, in modo da presentare informazioni sintetiche per i decisori e analisi più dettagliate per utenti con competenze tecniche o analitiche avanzate.

Un ulteriore requisito funzionale riguarda il **reporting strutturato**, ossia la possibilità di generare report periodici o ad-hoc a partire dai dati disponibili. Tali report devono poter essere esportati in formati standard (ad esempio PDF, CSV o Excel) per agevolarne la condivisione e l'archiviazione, garantendo al contempo la coerenza delle informazioni e la riproducibilità dei risultati. La piattaforma deve inoltre supportare la creazione di report parametrizzati, che consentano agli utenti di modificare filtri e intervalli temporali senza la necessità di interventi tecnici.

Un aspetto centrale è rappresentato dal supporto alla **analisi interattiva dei dati**, che permetta agli utenti di esplorare le informazioni in modo dinamico, applicando filtri, drill-down e aggregazioni direttamente sulle visualizzazioni. Tale funzionalità risulta particolarmente rilevante per garantire un approccio data-driven

ai processi decisionali, consentendo di individuare rapidamente anomalie, pattern ricorrenti o fenomeni emergenti nei dati. In questo contesto, la piattaforma deve favorire un utilizzo di tipo self-service, riducendo la dipendenza dal personale IT per le attività di analisi ordinaria.

La piattaforma TO-BE deve inoltre prevedere una **gestione strutturata degli utenti e dei ruoli**, consentendo di differenziare i livelli di accesso alle funzionalità e ai dati in base al profilo dell'utente. In un contesto pubblico, è essenziale garantire che ciascun utente possa accedere esclusivamente alle informazioni di propria competenza, nel rispetto dei principi di sicurezza, riservatezza e separazione dei compiti. A tal fine, devono essere supportate politiche di autorizzazione granulari, applicabili sia alle singole dashboard sia ai dataset sottostanti.

3.1.2 Requisiti tecnici

Dal punto di vista tecnico, la nuova piattaforma di Business Intelligence deve integrarsi efficacemente con l'infrastruttura informativa esistente, minimizzando l'impatto sui sistemi già in esercizio. Un requisito fondamentale riguarda la **compatibilità con i database relazionali** attualmente utilizzati, in particolare con sistemi diffusi nella Pubblica Amministrazione come Oracle, nonché con viste e viste materializzate già predisposte a supporto delle attività di analisi. La piattaforma deve essere in grado di interrogare tali sorgenti dati in modo efficiente, evitando duplicazioni non necessarie e garantendo prestazioni adeguate anche in presenza di dataset di grandi dimensioni.

La soluzione TO-BE deve inoltre supportare una **architettura scalabile**, in grado di adattarsi a un incremento progressivo del numero di utenti, del volume dei dati e della complessità delle analisi richieste. Tale requisito risulta particolarmente rilevante in una prospettiva di evoluzione futura del sistema, che potrebbe prevedere l'integrazione di nuove fonti informative o l'estensione della piattaforma ad altri

ambiti applicativi. La scalabilità deve essere garantita sia a livello applicativo sia a livello infrastrutturale, favorendo configurazioni modulari e flessibili.

Un ulteriore requisito tecnico riguarda le **modalità di installazione e gestione** della piattaforma. In linea con i principi di sostenibilità e riuso promossi dalla Pubblica Amministrazione, la soluzione deve poter essere installata su infrastrutture senza vincoli stringenti verso specifici fornitori. È inoltre auspicabile il supporto a modalità di deployment automatizzate e riproducibili, che facilitino le attività di installazione, aggiornamento e manutenzione nel tempo.

Dal punto di vista dell'integrazione applicativa, la piattaforma deve offrire **meccanismi di estensione e personalizzazione**, come API o strumenti di configurazione avanzata, che consentano di adattare il sistema alle esigenze specifiche del contesto organizzativo. Ciò include la possibilità di integrare sistemi di autenticazione centralizzati, adottare politiche di sicurezza coerenti con gli standard della Pubblica Amministrazione e personalizzare l'interfaccia utente in funzione delle necessità operative.

3.2 Criteri di valutazione delle soluzioni

La selezione di una piattaforma di Business Intelligence adeguata al contesto della Pubblica Amministrazione richiede l'adozione di un processo di valutazione strutturato e oggettivo. In presenza di numerose soluzioni open source disponibili, caratterizzate da funzionalità e architetture differenti, risulta infatti fondamentale definire criteri di valutazione chiari e coerenti con i requisiti funzionali e tecnici individuati nella sezione precedente. Tali criteri consentono di confrontare in modo sistematico le alternative disponibili, riducendo il rischio di scelte basate esclusivamente su fattori soggettivi o su preferenze tecnologiche non motivate.

I criteri di valutazione adottati per questo progetto derivano direttamente dai requisiti della piattaforma TO-BE e dalle esigenze operative emerse nel contesto

applicativo di riferimento. In particolare, essi tengono conto sia degli aspetti funzionali legati all'esperienza d'uso e alle capacità analitiche della piattaforma, sia degli aspetti tecnici e organizzativi che influenzano la sostenibilità della soluzione nel tempo. L'obiettivo non è soltanto individuare una piattaforma in grado di soddisfare le esigenze attuali, ma selezionare una soluzione che possa evolvere nel tempo e adattarsi a futuri scenari di crescita e trasformazione digitale.

Un elemento centrale nella definizione dei criteri di valutazione riguarda la necessità di garantire un **equilibrio tra copertura funzionale e semplicità di utilizzo**. In un contesto come quello della Pubblica Amministrazione, la piattaforma di Business Intelligence deve essere accessibile anche a utenti non esclusivamente tecnici, favorendo un utilizzo diffuso degli strumenti di analisi dei dati. Allo stesso tempo, deve offrire funzionalità avanzate per soddisfare le esigenze di utenti esperti, come analisti e personale IT, evitando di limitare le potenzialità analitiche del sistema.

Dal punto di vista tecnico, i criteri di valutazione devono considerare la **compatibilità con l'infrastruttura esistente**, la facilità di integrazione con i sistemi informativi già in uso e la capacità della piattaforma di garantire prestazioni adeguate in presenza di grandi volumi di dati. In particolare, risultano rilevanti aspetti quali il supporto a database relazionali diffusi nella Pubblica Amministrazione, la gestione efficiente delle query e la possibilità di adottare architetture scalabili. Tali elementi incidono in modo significativo sulla fattibilità tecnica della migrazione e sulla complessità delle attività di implementazione e manutenzione.

Un ulteriore insieme di criteri riguarda gli **aspetti economici e di governance** della soluzione. Sebbene le piattaforme open source non prevedano costi di licenza, è necessario valutare attentamente i costi indiretti legati all'installazione, alla formazione degli utenti e alla gestione operativa nel tempo. Allo stesso modo, la presenza di una community attiva, di documentazione aggiornata e di meccanismi di supporto rappresenta un fattore determinante per garantire la continuità e

l'affidabilità della piattaforma nel medio-lungo periodo.

Infine, assumono particolare rilevanza i criteri legati alla **sicurezza e alla gestione degli accessi**, soprattutto in un contesto pubblico in cui la protezione dei dati e il rispetto delle normative vigenti costituiscono requisiti imprescindibili. La piattaforma deve consentire l'adozione di politiche di autenticazione e autorizzazione coerenti con gli standard della Pubblica Amministrazione, garantendo un controllo granulare degli accessi e una gestione efficace dei ruoli utente.

Sulla base di queste considerazioni, nel Capitolo 4 viene presentata una metodologia di valutazione dettagliata, che traduce i requisiti e i principi delineati in questa sezione in criteri operativi applicabili alle singole piattaforme open source analizzate. Tale approccio consente di effettuare un confronto omogeneo e trasparente tra le soluzioni considerate, fornendo un supporto oggettivo alla selezione della piattaforma più idonea per la realizzazione del Proof of Concept.

Capitolo 4

Analisi comparativa delle soluzioni opensource di Business Intelligence

4.1 Introduzione e obiettivi dell'analisi

In questa sezione viene presentata un'analisi comparativa delle principali e possibili soluzioni open source di Business Intelligence disponibili. L'obiettivo di questa analisi è individuare la piattaforma più idonea per la realizzazione del Proof of Concept destinato a sostituire l'attuale sistema proprietario in uso. L'analisi è stata condotta considerando criteri funzionali, tecnici, economici e di governance, in modo da fornire una valutazione completa e oggettiva delle diverse alternative.

Il presente capitolo si concentra esclusivamente sull'analisi comparativa e sulla selezione della piattaforma, mentre le scelte architetturali di dettaglio e l'implementazione del Proof of Concept saranno approfondite nel capitolo 5.

4.1.1 Contesto e motivi della comparazione

Negli ultimi anni, la Pubblica Amministrazione ha avviato un processo di progressiva adozione di soluzioni open source con l'obiettivo di ridurre i costi legati alle licenze software, favorire l'interoperabilità e garantire la trasparenza delle piattaforme informatiche. In questo contesto, la sostituzione di una piattaforma di Business Intelligence basata su prodotti proprietari con una soluzione open source rappresenta un'opportunità strategica per modernizzare i sistemi informativi e migliorare la fruibilità dei dati. La piattaforma attualmente in uso, pur garantendo funzionalità consolidate e una buona stabilità, presenta limitazioni in termini di costi di manutenzione, scalabilità e integrazione.

4.1.2 Obiettivi della fase di analisi

L'obiettivo primario di questa fase di analisi è identificare, tra le soluzioni open source di Business intelligence disponibili, quella più idonea a soddisfare le esigenze tecniche e funzionali del contesto applicativo preso in considerazione. In particolare, si intende valutare le piattaforme in base ai loro criteri funzionali, tecnici, economici e organizzativi. L'obiettivo concreto è quello di supportare la realizzazione di un Proof of Concept (PoC) che possa costituire una base solida per una migrazione del sistema di BI proprietario esistente. Gli obiettivi specifici che guidano questa analisi includono:

- Verificare il grado di copertura funzionale rispetto alle esigenze di reporting, analisi interattiva, dashboarding e esportazione dei dati.
- Comprendere i requisiti tecnici richiesti per l'installazione e la manutenzione delle soluzioni, nonché la loro compatibilità con l'infrastruttura già esistente (database, data-warehouse, ecc...).
- Stimare i costi associati all'implementazione, alla formazione degli utenti e alla manutenzione operativa nel tempo, con il fine di valutare la sostenibilità

economica della migrazione.

- Analizzare le modalità di integrazione e interoperabilità con sistemi esistenti, in particolare con ambienti comuni nella Pubblica Amministrazione (ad esempio Oracle, viste materializzate, autenticazione tramite SPID).
- Valutare le caratteristiche di usabilità, governance e sicurezza, affinché la soluzione possa essere fruibile da utenti finali non esclusivamente tecnici e al contempo rispondere ai requisiti di controllo, ruoli e permessi del contesto pubblico.

Attraverso il soddisfacimento di questi obiettivi, l'analisi mira a fornire una base di conoscenza oggettiva e documentata che consenta di scegliere in modo consapevole la piattaforma su cui sviluppare il PoC, allineandosi alle esigenze operative e strategiche della richiesta.

4.1.3 Metodologia di valutazione e criteri adottati

Le informazioni necessarie all'analisi comparativa sono state raccolte attraverso:

- la documentazione ufficiale delle piattaforme open-source prese in considerazione;
- installazioni e test preliminari in ambiente locale per sperimentare le principali funzionalità;
- lettura di casi d'uso e articoli tecnici.

Nel dettaglio, i criteri di valutazione stabiliti sono i seguenti:

1. **Funzionalità:** ampiezza delle funzionalità offerte (dashboarding, analisi interattiva, personalizzazione, supporto al self-service, facilità d'uso).
2. **Architettura tecnica e integrazione:** compatibilità con database e sistemi eterogenei, facilità di sviluppo, scalabilità, supporto a ETL, viste materializzate e data-warehouse.

3. **Gestione dei dati e compatibilità DB:** supporto alle principali sorgenti di dati, come Oracle, PostgreSQL, MySQL, ecc..., gestione dei dataset, viste, aggregazioni, caching e performance.
4. **Aspetti economici e licenze:** tipo di licenza (open source vs proprietaria), costi di implementazione, formazione, manutenzione e possibili costi nel tempo.
5. **Usabilità e governance:** curva di apprendimento per gli utenti finali, gestione dei vari ruoli e permessi, personalizzazione dell'interfaccia, supporto alla fruizione delle dashboard in contesto PA.
6. **Sicurezza e autenticazione:** supporto ai sistemi di autenticazione (OAuth o SPID, ...), politiche di accesso ai dati.
7. **Community e supporto:** frequenza degli aggiornamenti, vitalità della community, documentazione disponibile, presenza di estensioni o plugin.

Per ogni piattaforma, ogni criterio è stato valutato con una scala descrittiva (Elevato, Medio, Limitato), permettendo così un confronto omogeneo e trasparente tra tutte le alternative.

4.2 Panoramica delle soluzioni open source considerate

Nell'ambito di questa analisi sono state selezionate alcune delle principali piattaforme open-source di Business Intelligence, ritenute adeguate per un contesto come quello della Pubblica Amministrazione. I criteri di selezione delle soluzioni hanno riguardato: la diffusione e la maturità tecnologica, la compatibilità con le infrastrutture, la disponibilità di comunità attive con aggiornamenti recenti, e la capacità di coprire l'intero ciclo di vita dell'analisi dei dati (estrazione, trasformazione, visualizzazione e condivisione).

Le piattaforme analizzate sono le seguenti:

- **Apache Superset**
- **Metabase**
- **Redash**
- **Pentaho Community Edition**
- **Jaspersoft Community Edition**

Questa panoramica consente di avere una visione ampia del panorama open-source della BI, utile a orientare la scelta finale sulla piattaforma più adatta al progetto in questione.

4.3 Descrizione generale delle piattaforme analizzate

4.3.1 Apache Superset

Apache Superset [6] è una piattaforma moderna di esplorazione e visualizzazione dei dati. Superset può sostituire o migliorare gli strumenti di business intelligence proprietari ed è stata progettata per essere leggera, scalabile e si integra bene con svariate sorgenti dati SQL. Superset fornisce:

- Un'**interfaccia senza codice** per creare rapidamente grafici.
- Un potente **editor SQL** basato sul Web per l'esecuzione avanzata di query.
- Un **livello semantico leggero** per definire rapidamente dimensioni e metriche personalizzate
- Supporto pronto all'uso per **quasi tutti i database SQL** o motori dati.
- Un'ampia gamma di **visualizzazioni per mostrare** i dati, dai semplici grafici a barre alle visualizzazioni geospaziali.

- **Livello di memorizzazione nella cache** leggero e configurabile per facilitare il caricamento del database.
- **Ruoli di sicurezza e opzioni di autenticazione** altamente estendibili.
- **Una API** per la personalizzazione programmatica.
- **Un'architettura nativa per il cloud** progettata da zero per la scalabilità.

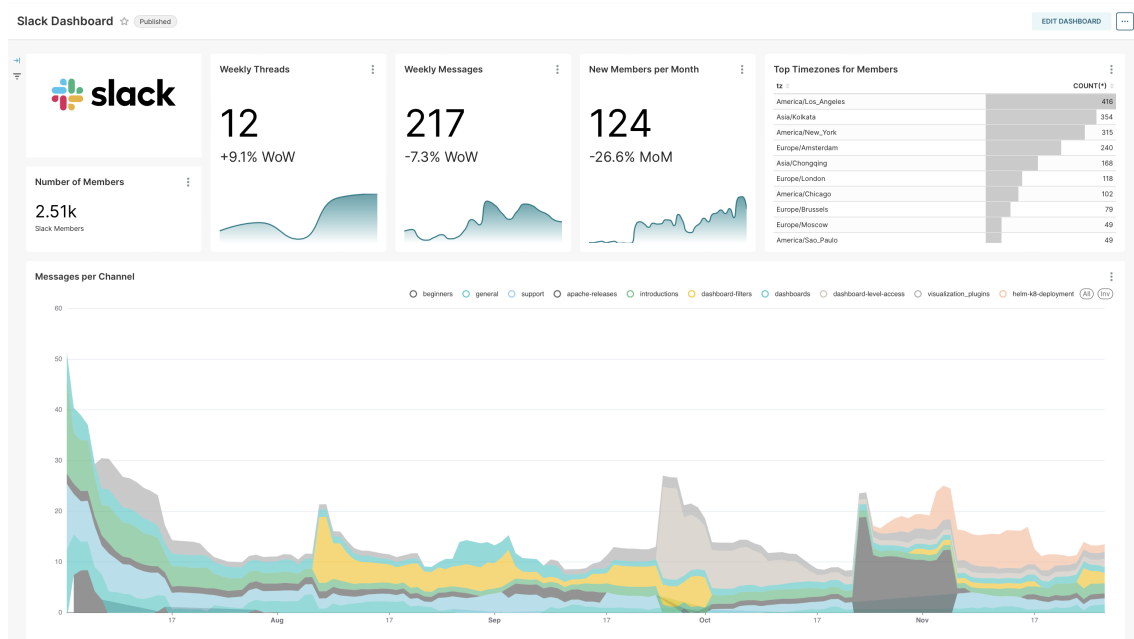


Figura 4.1. Esempio di dashboard in superset

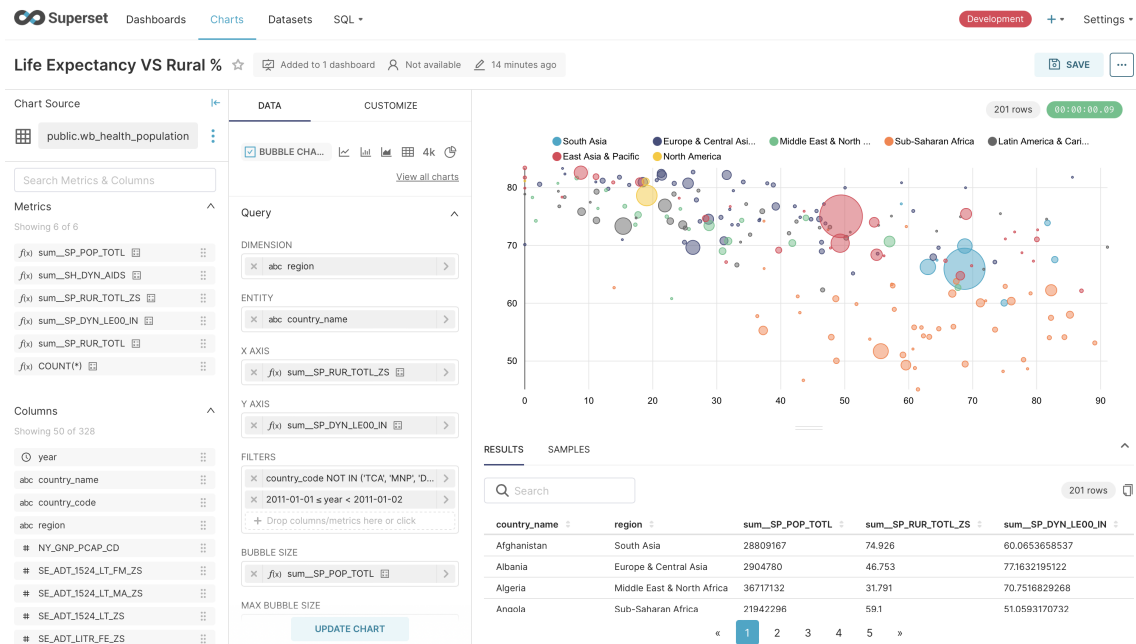


Figura 4.2. Editor sql in superset

4.3.2 Metabase

Metabase [3] è una piattaforma di Business Intelligence che offre una serie di strumenti per comprendere e condividere dei dati. Le aziende in genere utilizzano Metabase per offrire ai propri team un modo semplice per interrogare i dati o per incorporare Metabase nell'applicazione per consentire ai clienti di esplorare i dati da soli. E' possibile utilizzare Metabase per:

- **Interrogare un database** utilizzando un generatore di query grafico, oppure un editor di query nativo.
- **Visualizzare i risultati** sotto forma di tabelle o grafici personalizzabili.
- **Salvare i risultati** come domande che possono essere aggiunte a dashboard o collezioni.
- **Aggiungere filtri** alle dashboard in modo da aggiornare tutte le domande contemporaneamente.

- **Organizzare** le domande e le dashboard in raccolte.
- Creare **modelli** per curare i dataset, semplificando le query sui dati comunemente filtrati.
- Creare **metriche** per standardizzare il modo in cui il team calcola e analizza dati importanti.

Metabase dispone di molti strumenti, ma quelli più importanti sono:

- Le **domande**, ovvero query salvate con visualizzazioni che è possibile aggiungere a dashboard o raccolte. Le domande sono sostanzialmente i grafici che è possibile organizzare.
- Le **dashboard** raggruppano le domande correlate che possono essere filtrate e aggiornate insieme.
- Le **raccolte** sono come cartelle per organizzare e gestire le autorizzazioni per le domande, le dashboard e altri elementi.
- I **modelli** sono simili a delle viste che curano i dati del database.
- Le **metriche** definiscono il modo ufficiale per calcolare dati importanti per il team.

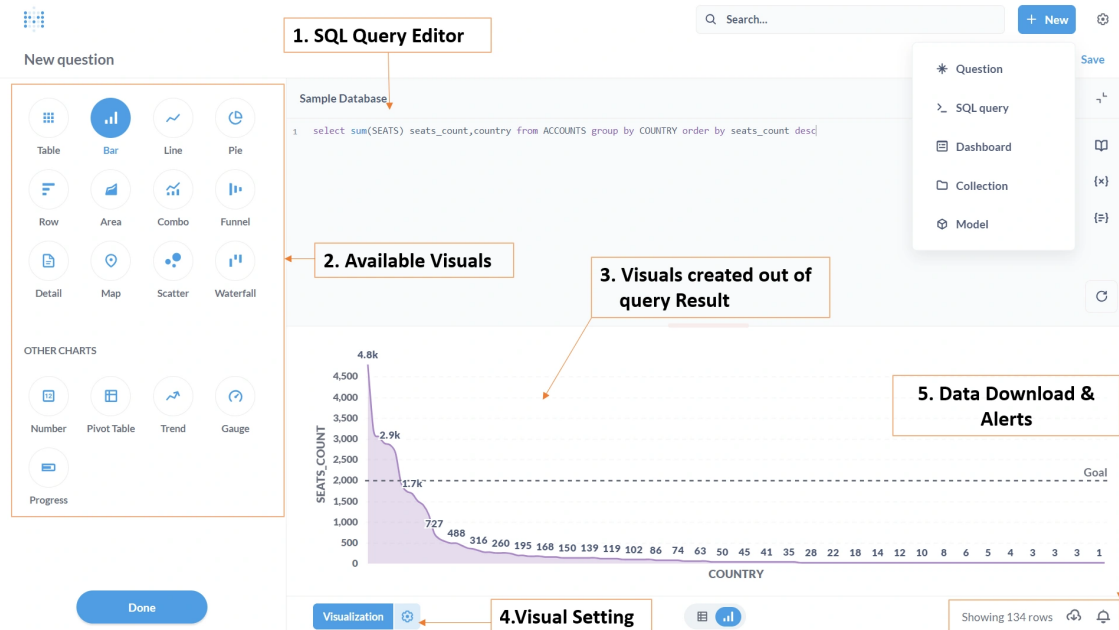


Figura 4.3. Editor di domande in Metabase. Fonte: [8]

4.3.3 Redash

Redash [11] è una piattaforma open source di Business Intelligence orientata principalmente all'esecuzione di query e alla condivisione dei risultati attraverso visualizzazioni semplici e immediate. A differenza di soluzioni più complete e strutturate, Redash adotta un approccio essenziale e fortemente incentrato sull'**utilizzo diretto del linguaggio SQL**, risultando particolarmente adatto a utenti con competenze tecniche e analitiche.

Il cuore della piattaforma è rappresentato dall'editor SQL, che consente di interrogare direttamente numerose sorgenti dati, tra cui database relazionali, data warehouse e API esterne. Le query possono essere salvate, riutilizzate e programmate per l'aggiornamento automatico dei risultati, favorendo la creazione di report dinamici basati su dati sempre aggiornati.

I risultati delle interrogazioni possono essere visualizzati tramite grafici e tabelle essenziali, successivamente organizzabili in dashboard condivisibili. Redash pone

particolare attenzione alla collaborazione, consentendo agli utenti di commentare le query, condividere i risultati e integrare le dashboard all'interno di applicazioni esterne tramite embedding.

Dal punto di vista architetturale, Redash si presenta come una soluzione leggera e relativamente **semplice da installare e mantenere**, caratteristica che ne ha favorito la diffusione in contesti di sviluppo agile e in ambienti dove è richiesto un accesso rapido e diretto ai dati. Tuttavia, **l'assenza di un livello semantico avanzato e di strumenti di modellazione dei dati** limita l'utilizzo della piattaforma in scenari complessi o fortemente orientati al self-service per utenti non tecnici.

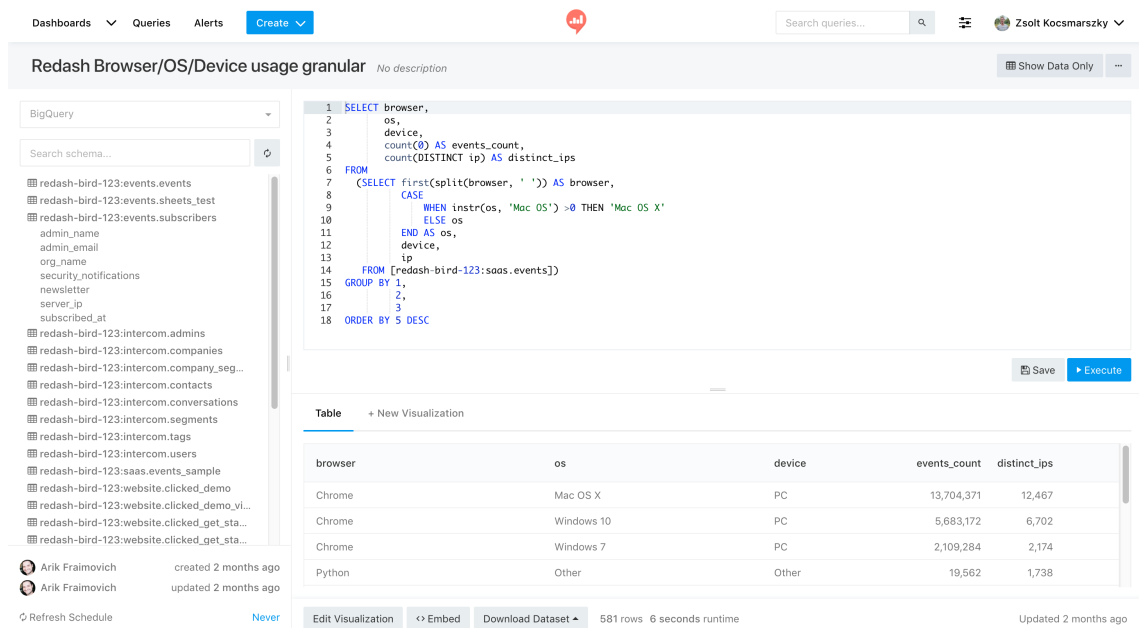


Figura 4.4. Caption

4.3.4 Pentaho

Pentaho Business Analytics [10] è una piattaforma di Business Intelligence che offre una suite integrata di strumenti per l'analisi dei dati, il reporting e la creazione di

dashboard, pensata per affrontare scenari enterprise con esigenze di integrazione, trasformazione e presentazione dei dati.

La soluzione si distingue per il suo approccio modulare e orientato all'intero ciclo di vita dei dati, con componenti pensati per supportare sia utenti tecnici sia analisti all'interno di una stessa piattaforma. Secondo la documentazione ufficiale, Pentaho consente di esplorare e visualizzare informazioni provenienti da sorgenti eterogenee mediante strumenti come l'Analyzer e il Dashboard Designer, che permettono di combinare report, grafici e altri elementi visuali in viste coerenti e fruibili.

A differenza di soluzioni più leggere (come Redash) o con forte orientamento al self-service (come Metabase), Pentaho è progettato per **affrontare esigenze complesse** di Business Intelligence aziendale, spesso integrate con **ampi processi di ETL e data warehousing** [5].

Dal punto di vista architetturale, Pentaho si caratterizza per la sua integrazione con processi di trasformazione dei dati di ampia portata. Questo approccio permette di combinare funzioni di ETL, data blending e analisi all'interno di un'unica piattaforma, favorendo la realizzazione di soluzioni BI che vanno oltre la semplice visualizzazione dei dati.

Tuttavia, nell'ambito di un confronto con strumenti moderni orientati all'esperienza utente e all'adozione self-service, la stessa documentazione evidenzia che l'utilizzo di Pentaho può risultare **più complesso sul fronte della gestione e configurazione**, richiedendo competenze tecniche per la gestione della piattaforma e dei processi sottostanti.

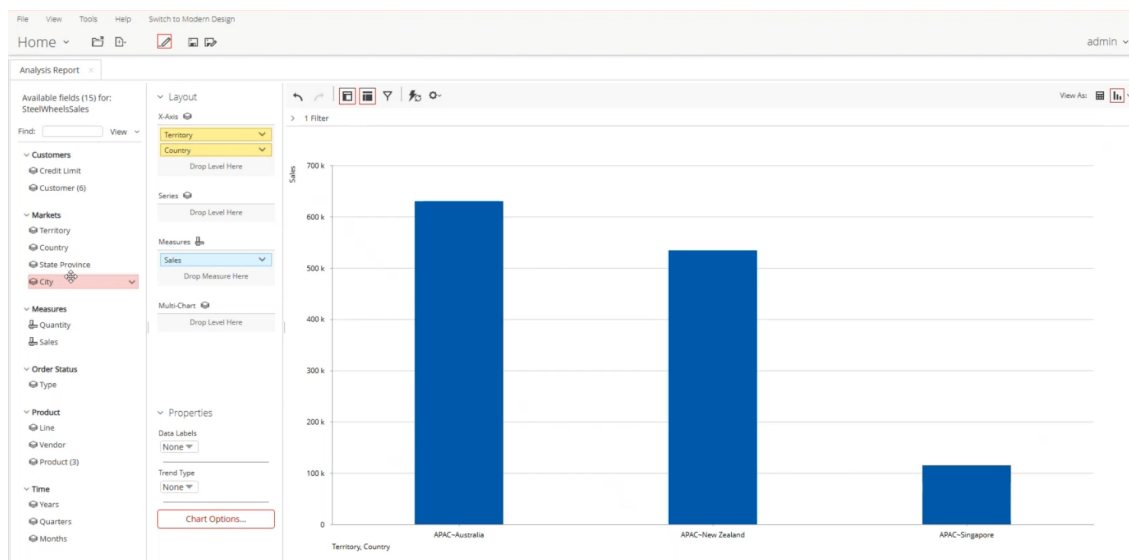


Figura 4.5. Interfaccia di Pentaho: ambiente di analisi e reportistica

4.3.5 Jaspersoft

Jaspersoft [4] è una piattaforma di Business Intelligence orientata alla reportistica avanzata e all'**analisi incorporabile (embedded analytics)**, progettata per consentire alle organizzazioni di **integrare funzioni di BI direttamente all'interno delle applicazioni esistenti** e di fornire report e visualizzazioni ad ampio raggio senza gravare sulle risorse di sviluppo.

Secondo la documentazione ufficiale, Jaspersoft permette di creare report ad alta precisione (pixel-perfect) e visualizzazioni personalizzate basate su un'intuitiva interfaccia grafica, con strumenti pensati per facilitare la generazione sia di report statici formattati sia di report ad hoc più dinamici. La piattaforma mette inoltre a disposizione funzionalità per l'embedding di dashboard e analisi all'interno di altri software, migliorando l'usabilità e l'adozione delle capacità analitiche direttamente nei flussi applicativi di business.

Pur offrendo funzioni di self-service e strumenti per semplificare la creazione di report da parte di utenti non tecnici, Jaspersoft è spesso adottato in scenari enterprise

dove è richiesta una combinazione di reportistiche formattate, analisi embedded e capacità di integrazione avanzata con sistemi applicativi preesistenti. Questo lo rende un'opzione potenzialmente interessante in progetti di BI che privilegiano la distribuzione di insight all'interno di software verticali o piattaforme applicative.

4.3.6 Tabelle comparative

Di seguito una tabella di confronto sintetico tra le piattaforme analizzate, costruita sulla base delle varie documentazioni prese in considerazione.

Tabella 4.1. Confronto tra le principali piattaforme open source di BI analizzate

Critero	Superset	Metabase	Redash	Pentaho CE	Jaspersoft CE
Copertura funzionale BI	Elevato	Medio-Elevato	Medio	Elevato	Medio-Elevato
Dashboarding e visualizzazione	Elevato	Elevato	Medio	Medio	Medio
Analisi interattiva/Self-service	Medio-Elevato	Elevato	Limitato	Medio	Medio
Editor SQL avanzato	Elevato	Medio	Elevato	Medio	Medio
Gestione dataset e livello semantico	Medio	Elevato	Limitato	Elevato	Medio
Supporto DB e sorgenti dati	Elevato	Elevato	Elevato	Elevato	Elevato
Integrazione con data warehouse/viste	Elevato	Medio	Medio	Elevato	Elevato
Usabilità per utenti non tecnici	Medio	Elevato	Limitato	Medio	Medio
Scalabilità e architettura	Elevato	Medio	Medio	Medio	Medio
Sicurezza e gestione ruoli	Elevato	Medio	Medio	Elevato	Elevato
Facilità di installazione e gestione	Medio	Elevato	Medio	Limitato	Medio
Maturità del progetto	Elevato	Elevato	Medio	Elevato	Elevato
Community e documentazione	Elevato	Elevato	Medio	Medio	Medio
Aderenza al contesto PA	Elevato	Medio-Elevato	Medio	Medio	Medio

La tabella 4.1 riassume il confronto tra le principali piattaforme open source di Business Intelligence analizzate, mettendo in evidenza punti di forza e limiti rispetto ai criteri funzionali e tecnici individuati. La valutazione qualitativa mediante scala descrittiva (Elevato, Medio, Limitato) consente di evidenziare le differenze tra le soluzioni in modo immediato, facilitando una lettura comparativa e supportando il

processo decisionale per la selezione della piattaforma più idonea alla realizzazione del Proof of Concept.

4.3.7 Analisi critica: punti di forza e limiti delle soluzioni

L'analisi fatta e la tabella comparativa delle piattaforme open source di Business Intelligence evidenzia come ciascuna soluzione presenti caratteristiche specifiche che la rendono più o meno adatta a determinati contesti applicativi, confermando l'assenza di una piattaforma universalmente migliore e la necessità di una valutazione guidata dai requisiti del dominio di riferimento. I criteri funzionali e tecnici analizzati consentono tuttavia di individuare punti di forza e limiti ricorrenti particolarmente rilevanti.

Apache Superset emerge come una piattaforma moderna e altamente scalabile, caratterizzata da una forte integrazione con sorgenti dati eterogenee e da un potente supporto all'analisi tramite SQL. I principali punti di forza risiedono nell'architettura cloud-native, nella flessibilità delle visualizzazioni e nella gestione avanzata della sicurezza e dei ruoli, aspetti fondamentali in ambienti istituzionali. Di contro, Superset presenta una curva di apprendimento più elevata per utenti non tecnici, richiedendo competenze specifiche per la configurazione iniziale e per la gestione ottimale della piattaforma.

Metabase si distingue invece per l'elevata usabilità e per l'approccio fortemente orientato al self-service analytics. La semplicità dell'interfaccia e la presenza di strumenti visuali per la costruzione delle query rendono la piattaforma particolarmente adatta a utenti con competenze tecniche limitate, favorendo un'adozione rapida e una maggiore autonomia nell'analisi dei dati. Tuttavia, rispetto a Superset, Metabase mostra alcune limitazioni in termini di scalabilità, personalizzazione avanzata e gestione granulare dei ruoli, elementi che possono risultare critici nel contesto di questa tesi.

Redash adotta un approccio fortemente orientato alle query SQL, privilegiando

l'analisi tecnica e l'esplorazione diretta dei dati. Questo lo rende uno strumento efficace per utenti esperti e per scenari di analisi puntuale, ma ne limita l'accessibilità per utenti non tecnici e l'utilizzo come piattaforma di BI generalista. L'assenza di un livello semantico evoluto e di funzionalità avanzate di governance rappresenta un limite significativo.

Pentaho Community Edition e **Jaspersoft Community Edition** rappresentano soluzioni mature, adottate in contesti enterprise e spesso presenti in sistemi informativi legacy. Pentaho si distingue per l'integrazione nativa con processi di ETL e data integration, offrendo una visione completa dell'intero ciclo di vita del dato. Tuttavia, la complessità architetturale e la minore immediatezza dell'interfaccia possono costituire un ostacolo all'adozione da parte di utenti non specializzati. Jaspersoft, dal canto suo, eccelle nella reportistica strutturata e nel supporto all'embedded analytics, risultando particolarmente indicato per la produzione di report formattati e distribuiti all'interno di applicazioni esistenti. Al contempo, la piattaforma risulta meno orientata all'analisi esplorativa e al self-service che sono aspetti abbastanza centrali.

Nel complesso, l'analisi mette in evidenza una tendenza comune: le **soluzioni più moderne privilegiano flessibilità, scalabilità e capacità di analisi interattiva**, mentre le **piattaforme più consolidate offrono maggiore stabilità e funzionalità di reporting strutturato, spesso a fronte di una maggiore complessità di gestione**.

Questi elementi costituiscono la base per la fase successiva di valutazione, nella quale verrà individuata la soluzione più idonea, tenendo conto delle specifiche esigenze operative e tecnologiche del contesto analizzato.

4.4 Discussione e selezione della soluzione più idonea

L'analisi comparativa delle principali piattaforme open source di Business Intelligence ha messo in evidenza come, a fronte di un insieme di funzionalità di base condivise, le soluzioni esaminate presentino differenze significative in termini di complessità architetturale, target di utenti e modalità di utilizzo.

Alcune piattaforme risultano particolarmente complete dal punto di vista funzionale, ma introducono una maggiore complessità nella fase operativa. Al contrario, le altre piattaforme concepite per l'analisi esplorativa e il self-service analytics privilegiano la semplicità d'uso e l'accessibilità, risultando più adatti a contesti in cui l'obiettivo principale è facilitare la consultazione e l'interpretazione dei dati da parte di un'utenza eterogenea.

Nel contesto considerato, caratterizzato dai vincoli già analizzati tipici della Pubblica Amministrazione e dalla necessità di evolvere un sistema esistente senza introdurre ulteriori elementi di rigidità, assumono particolare rilevanza aspetti quali la facilità di integrazione con le sorgenti dati, la scalabilità della soluzione, il supporto a diversi livelli di competenza degli utenti e la possibilità di estensione futura.

Alla luce delle considerazioni emerse dall'analisi comparativa, la piattaforma individuata come più idonea per la realizzazione del proof of concept è **Apache Superset**.

Apache Superset si è dimostrato particolarmente adatto al contesto analizzato grazie alla sua ampia compatibilità con sorgenti dati eterogenee, al supporto nativo per l'esecuzione di query SQL avanzate e alla possibilità di definire un livello semantico leggero, elementi rilevanti per l'integrazione con l'infrastruttura esistente. Inoltre, le funzionalità di gestione dei ruoli e delle autorizzazioni, unite alla scalabilità dell'architettura, risultano coerenti con le esigenze del progetto.

Capitolo 5

Realizzazione del proof of concept

5.1 Architettura e setup di Apache Superset

La realizzazione del Proof of Concept della nuova piattaforma di Business Intelligence si è basata sull'adozione di Apache Superset, selezionato a seguito dell'analisi comparativa presentata nel Capitolo 4. Superset è una piattaforma open source progettata per l'esplorazione, l'analisi e la visualizzazione dei dati, caratterizzata da un'architettura modulare e scalabile, adatta sia a contesti sperimentali sia a scenari enterprise.

L'obiettivo di questa fase è stato quello di **predisporre un ambiente di test stabile e rappresentativo**, in grado di replicare, per quanto possibile, le condizioni operative di un futuro ambiente di produzione, consentendo al contempo lo svolgimento delle attività di analisi, sviluppo e validazione del PoC.

5.1.1 Scelte architetturali iniziali

In una prima fase, seguendo le indicazioni della documentazione ufficiale di Apache Superset, è stata valutata l'installazione dell'applicativo tramite **container Docker**, soluzione generalmente suggerita per rapidità di setup e facilità di riproduzione

degli ambienti.

Tuttavia, nel contesto specifico del progetto, questo approccio ha evidenziato delle significative **criticità**, legate principalmente alla **connessione a database remoti Oracle accessibili tramite rete aziendale e VPN**. In particolare, Superset, eseguito all'interno di un container Docker su un sistema operativo Windows, non riesce a stabilire una connessione stabile con il database Oracle GECO, a causa delle limitazioni di rete introdotte dall'isolamento del container rispetto alla rete aziendale.

Questa problematica ha reso l'approccio basato su Docker **non idoneo per la fase di test**, in quanto impediva la fattibilità di uno degli aspetti fondamentali del PoC: la connessione alle sorgenti dati reali.

5.1.2 Installazione nativa su Ubuntu tramite WSL

A seguito delle criticità riscontrate nell'approccio containerizzato, è stata adottata un'altra strategia basata sull'installazione di **Apache Superset su sistema operativo Windows**, utilizzando **Ubuntu tramite Windows Subsystem for Linux (WSL)**. Questa soluzione ha consentito di superare le limitazioni introdotte dall'isolamento dei container Docker, garantendo la corretta integrazione con la rete aziendale e un accesso diretto alle risorse necessarie al progetto.

L'installazione di Apache Superset è stata eseguita seguendo le linee guida ufficiali adattandole al contesto specifico. In particolare, è stato predisposto un **ambiente Python dedicato**, all'interno del quale sono state installate tutte le dipendenze necessarie al corretto funzionamento dell'applicativo. La procedura di installazione ha previsto:

- la configurazione dell'ambiente Linux e delle dipendenze di sistema richieste (librerie di base, strumenti di compilazione, driver necessari alla connessione con database esterni);

- la creazione di un ambiente virtuale Python dedicato a Superset;
- l'installazione del framework Apache Superset e dei moduli aggiuntivi necessari;
- l'inizializzazione dell'applicativo e la configurazione iniziale del database di metadati interno a Superset.

Durante questa fase è stata compilata una **guida tecnica di installazione**, utilizzata come riferimento operativo, che descrive nel dettaglio i passaggi eseguiti e i comandi utilizzati. Tale documentazione risulta utile non solo per la riproducibilità dell'ambiente di test, ma anche come base per una futura installazione in ambiente di produzione.

L'installazione nativa su Ubuntu ha consentito di ottenere un'istanza di Apache Superset funzionante, accessibile tramite interfaccia web in **locale sulla porta 8088**, sulla quale è stato possibile procedere con la configurazione della connessione al database, la definizione dei dataset e lo sviluppo delle dashboard di analisi.

5.1.3 Ambiente software e dipendenze

Dal punto di vista architetturale, l'ambiente può essere suddiviso in diversi livelli logici. A livello di sistema operativo, abbiamo visto che l'applicativo è stato eseguito su Ubuntu Linux tramite Windows Subsystem for Linux (WSL). Questo approccio ha, tra l'altro, garantito una maggiore compatibilità con le librerie e i framework utilizzati da Apache Superset rispetto ad un'installazione su sistema operativo Windows.

Il livello applicativo è basato su Python, linguaggio su cui Superset è interamente sviluppato. L'utilizzo di un ambiente virtuale dedicato ha permesso di isolare le dipendenze dell'applicativo, semplificando la gestione delle librerie e riducendo il rischio di conflitti con altri software presenti sul sistema.

Per quanto riguarda il livello di persistenza e gestione dei metadati, Superset utilizza

un database interno per memorizzare configurazioni, utenti, dashboard e definizioni dei dataset. Questo database di metadati è separato dalle sorgenti dati analizzate, consentendo una netta distinzione tra configurazione dell'applicativo e dati di dominio.

Un aspetto centrale riguarda la gestione delle **dipendenze per l'accesso ai database esterni**, in questo caso ai sistemi Oracle utilizzati nel contesto applicativo. L'integrazione con i vari database utilizzabili richiedono l'installazione di driver specifici e librerie aggiuntive, indispensabili per garantire la corretta comunicazione tra superset e le sorgenti dati.

L'insieme delle scelte effettuate ha permesso di realizzare un ambiente di test robusto e coerente con gli obiettivi, ponendo le basi per le attività successive di connessione ai dati, sviluppo delle dashboard e comprensione e validazione di tutte le funzionalità offerte dalla piattaforma.

5.2 Connessione ai dati

5.2.1 Descrizione del database Oracle GECO

La sorgente dati utilizzata per la realizzazione del Proof of Concept è rappresentata dal database **Oracle GECO**, sistema informativo impiegato dalla regione Basilicata per la gestione e il monitoraggio di procedimenti amministrativi relativi al dominio della **caccia e delle licenze venatorie**. Il database costituisce il repository centrale delle informazioni relative alle comunicazioni amministrative, alle licenze di caccia, agli utenti coinvolti e ai riferimenti territoriali, ed è già integrato all'interno dell'infrastruttura informatica regionale.

Dal punto di vista funzionale, GECO gestisce dati eterogenei, comprendenti:

- informazioni anagrafiche dei soggetti coinvolti;

- dati amministrativi relativi alle comunicazioni e ai relativi stati di avanzamento;
- informazioni sulle licenze di caccia e sulle stagioni venatorie;
- dati territoriali relativi a comuni, province e sedi territoriali competenti.

Il database è strutturato secondo un modello relazionale normalizzato, in cui le entità principali sono collegate tramite chiavi primarie ed esterne, garantendo coerenza e integrità referenziale. Questa struttura favorisce, ovviamente, la gestione del sistema ma rende necessaria una fase di preparazione e un appiattimento dei dati per consentire un'analisi efficace tramite Superset.

5.2.2 Struttura logica dei dati e dominio informativo

L'analisi della struttura del database GECO, condotta tramite l'utilizzo del client SQL DBeaver, ha evidenziato una suddivisione dei dati in diverse tipologie di tabelle. E' possibile individuare:

- **tabelle transazionali**, contenenti le informazioni principali sulle comunicazioni e sulle licenze di caccia: comunicazioni amministrative, licenze, pagamenti, date di rilascio e di consegna;
- **tabelle anagrafiche**, usate per descrivere soggetti, utenti del sistema e beneficiari, comprensive di dati personali e dati di contatto;
- **tabelle di dominio**, contenenti valori di riferimento e classificazioni, come tipologie di comunicazione, stati delle pratiche e stagioni venatorie;
- **tabelle territoriali**, impiegate per rappresentare la dimensione geografica del dato includendo comuni, province e sedi territoriali competenti.

La presenza di tutti questi dati temporali (date di inserimento, aggiornamento, pagamento, scadenze), amministrativi (stato e tipo delle comunicazioni) e territoriali rende il dataset particolarmente adatto alla costruzione di dashboard di analisi.

Il livello di normalizzazione del database comporta la necessità di aggregare e semplificare i dati per l'analisi. Per questo motivo, è stata progettata una vista materializzata dedicata che integra informazioni provenienti da più tabelle, riducendo la complessità delle interrogazioni e migliorando le prestazioni generali dell'applicativo scelto.

5.2.3 Preparazione dei dati per l'analisi

Al fine di rendere i dati del database GECO fruibili da Superset e adeguati alla costruzione di dashboard e analisi interattive, si è resa necessaria una fase di **preparazione e riorganizzazione dei dati**.

In questo contesto, l'elevata normalizzazione del database, sebbene garantisca coerenza e integrità dei dati, richiede numerose operazioni di join tra tabelle diverse e una conoscenza del modello dati sottostante. Per superare queste criticità e semplificare l'accesso ai dati da parte della piattaforma, è stata progettata una **vista materializzata** dedicata, denominata **MV-COMUNICAZIONE-LICENZA-CACCIA**.

La vista integra informazioni provenienti da più tabelle del sistema GECO, includendo, come visto, dati anagrafici, amministrativi, temporali e territoriali, restituendo un dataset già strutturato le analisi. In particolare, la vista consente di correlare le comunicazioni amministrative con le licenze di caccia, le stagioni venatorie di riferimento, lo stato delle pratiche e le sedi territoriali competenti, offrendo una visione unificata del dominio informativo.

Durante questa fase di progettazione della vista, è stata posta attenzione anche alla **leggibilità dei dati**, soprattutto in relazione alla fruizione da parte di utenti non tecnici. A tal fine, alcuni attributi codificati nel database sorgente, come flag binari e valori simbolici, sono stati trasformati in valori testuali più espliciti (ad esempio la conversione dei valori D/C in DIGITALE/CARTACEO). Questa scelta ha permesso di semplificare la visualizzazione dei grafici in Superset e di rendere i

dati mostrati nelle dashboard immediatamente interpretabili.

Un ulteriore aspetto rilevante della preparazione dei dati riguarda l'introduzione di informazioni a supporto delle visualizzazioni geografiche. Nella vista sono stati, infatti, inclusi campi di **latitudine e longitudine** associati alle sedi territoriali competenti, al fine di consentire la rappresentazione dei dati su mappe tematiche. Tali informazioni, pur non essendo sempre presenti nel database sorgente in forma esplicita, risultano fondamentali per sfruttare le funzionalità di visualizzazione geospaziale offerte da Superset.

Dal punto di vista prestazionale, la scelta di usare una **vista materializzata** consente di separare il carico analitico dalle operazioni transazionali del sistema GECO. La vista è stata definita con un meccanismo di **refresh automatico giornaliero**, tramite schedulazione lato database, così da garantire l'allineamento periodico dei dati senza impattare sulle performance del sistema operativo durante l'orario di lavoro.

5.2.4 Configurazione della connessione in Superset

Una volta completata l'installazione dell'ambiente software e verificato l'avvio dell'applicativo, la fase successiva ha riguardato la **configurazione della connessione tra Superset e il database Oracle GECO**, al fine di consentire l'accesso ai dati necessari per le attività di visualizzazione e analisi.

Apache Superset supporta nativamente la connessione a un'ampia varietà di database relazionali tramite l'utilizzo di driver SQLAlchemy e driver specifici per ogni tecnologia. La configurazione di una nuova sorgente dati avviene attraverso l'interfaccia web dell'applicativo, accedendo alla sezione dedicata alla gestione dei database e specificando i parametri di connessione richiesti.

Nel caso in esame, essendo il database GECO basato su tecnologia Oracle, è stato necessario predisporre in precedenza i **driver e librerie client Oracle** compatibili con l'ambiente python utilizzato da superset. Questa fase risulta fondamentale

affinchè l'applicativo sia in grado di instaurare la comunicazione con il database remoto.

La configurazione della connessione ha richiesto l'inserimento di alcune informazioni di accesso al database, tra cui:

- l'host e la porta del server Oracle;
- il service del database;
- le credenziali di autenticazione;
- il tipo di database e il driver SQLAlchemy appropriato.

Superset mette a disposizione un campo dedicato alla definizione della **stringa di connessione**, che consente di specificare in maniera esplicita il formato di collegamento al database. Una volta completata la configurazione, è possibile eseguire un test di connessione direttamente dall'interfaccia, al fine di verificare la raggiungibilità del database e la correttezza dei parametri inseriti. Tale funzionalità si è rivelata particolarmente utile durante la fase di setup, consentendo di individuare rapidamente problemi di autenticazione o configurazione dei driver.

Una volta stabilita la connessione, Superset consente di esplorare lo schema del database e di selezionare le tabelle o viste da rendere disponibili per l'analisi creando un **dataset**.

The screenshot shows the configuration page for a database in Superset. At the top, it says "oracle" and "Oracle_GECO". There are two tabs: "Basic" (selected) and "Advanced". Under the "Basic" tab, there are two input fields. The first is labeled "Display Name *" and contains the text "Oracle_GECO". Below it is a hint: "Pick a name to help you identify this database." The second input field is labeled "SQLAlchemy URI *" and contains the text "oracle://GECO:XXXXXXXXXX@srv1-oraclesirs.hosting.int:1521/?se". Below it is a hint: "Refer to the [SQLAlchemy docs](#) for more information on how to structure your URI." At the bottom of the form is a large button labeled "Test connection".

Figura 5.1. Schermata per la configurazione della connessione del database in Superset

5.3 Riproduzione dei casi d'uso reali

5.3.1 Mappatura AS-IS e PoC

La realizzazione del Proof of Concept ha previsto una fase preliminare di **mappatura tra il sistema attuale (AS-IS) e le funzionalità offerte dalla nuova piattaforma** di Business Intelligence basata su Apache Superset. L'obiettivo di questa attività è stato quello di verificare se, e in che misura, Superset fosse in grado di **riprodurre i principali casi d'uso** operativi già presenti nel sistema esistente, introducendo al contempo elementi di miglioramento in termini di flessibilità, usabilità e capacità di analisi. A partire da tale contesto, il Proof of Concept è stato progettato con l'obiettivo di replicare i principali flussi informativi del sistema

attuale, utilizzando come sorgente dati il database Oracle GECO e, in particolare, la vista materializzata appositamente predisposta per l'analisi.

La mappatura AS-IS / PoC si è concentrata, in particolare, sul riprodurre le funzionalità del vecchio software non attraverso report statici, bensì mediante **dashboard interattive** e **filtri dinamici**, al fine di valutare l'effettivo valore aggiunto introdotto dalla piattaforma di Business Intelligence rispetto all'approccio tradizionale. Questa fase di mappatura ha avuto un ruolo fondamentale anche nel definire l'ambito del Proof of Concept, permettendo di distinguere tra:

- funzionalità già presenti nel sistema AS-IS e semplicemente riprodotte in Superset;
- funzionalità migliorate o rese più accessibili grazie alle capacità di esplorazione e visualizzazione dei dati offerte dalla piattaforma;
- funzionalità non coperte dal PoC e demandate a eventuali sviluppi futuri.

La mappatura tra sistema attuale e Proof of Concept ha quindi rappresentato il punto di partenza per la successiva realizzazione delle dashboard di monitoraggio, consentendo una valutazione concreta dell'efficacia di Apache Superset nel contesto applicativo oggetto di studio.

5.3.2 Creazione dei grafici e utilizzo dell'editor

Prima della realizzazione delle dashboard di monitoraggio, è stata svolta un'attività di configurazione e creazione dei singoli grafici attraverso l'editor messo a disposizione da Superset. Questo strumento rappresenta uno degli elementi centrali della piattaforma, in quanto consente di definire visualizzazioni analitiche in modo guidato, senza la necessità di scrivere manualmente query SQL complesse.

L'editor dei grafici di Superset si basa su un approccio dichiarativo, in cui l'utente seleziona il **dataset di riferimento**, il **tipo di visualizzazione** e i **campi di**

interesse attraverso un'interfaccia grafica strutturata. A partire da tali configurazioni, Superset genera automaticamente le interrogazioni al database sottostante, sfruttando un **livello semantico** che astrae la complessità del modello dati.

L'interfaccia dell'editor è suddivisa in diverse sezioni principali:

- la selezione del dataset;
- la definizione delle metriche e delle dimensioni di analisi;
- la configurazione dell'asse temporale e del livello di aggregazione;
- le opzioni di personalizzazione della visualizzazione.

Superset mette a disposizione un'ampia gamma di **tipologie di grafici**, tra cui grafici a barre, a linee, ad area, a torta, tabelle pivot, indicatori numerici (KPI), mappe geografiche e visualizzazioni temporali.

Un aspetto particolarmente rilevante è la possibilità di costruire grafici basati su **aggregazioni temporali**, selezionando la granularità desiderata (giorno, mese, anno) e definendo metriche di conteggio o calcolo sui dati.

L'editor consente inoltre di visualizzare in tempo reale il risultato delle configurazioni effettuate, permettendo all'utente di iterare rapidamente sulle impostazioni fino a ottenere la visualizzazione desiderata. Questo approccio facilita l'esplorazione dei dati anche da parte di utenti non strettamente tecnici, riducendo la dipendenza da figure specialistiche per la costruzione delle analisi.

I grafici così definiti possono essere salvati come oggetti riutilizzabili e successivamente inseriti all'interno di una o più dashboard. Questa separazione tra fase di creazione del grafico e fase di composizione della dashboard favorisce la modularità dell'applicativo.

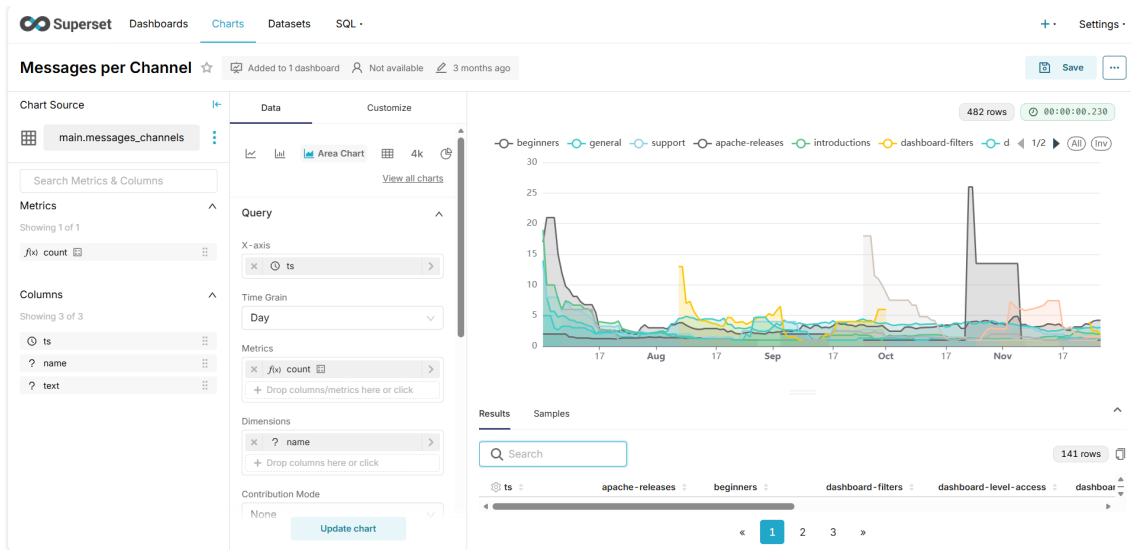


Figura 5.2. Editor per la creazione dei grafici

5.3.3 Dashboard di monitoraggio e grafici

A seguito della fase di mappatura e di creazione dei grafici, è stata realizzata una **dashboard di monitoraggio** all'interno di Apache Superset, con l'obiettivo di fornire una vista sintetica ma al tempo stesso esplorabile sui dati relativi alle licenze di caccia. In Superset, la dashboard consente di **aggregare più visualizzazioni eterogenee** all'interno di un'unica interfaccia, favorendo una lettura immediata dello stato del dominio applicativo.

Nella sezione iniziale della dashboard è stata predisposta un'area di **overview**, pensata per fornire una visione d'insieme dei principali indicatori a scopo di esempio e test, come:

- indicatori numerici relativi al numero totale di licenze rilasciate e al numero di licenze attive;
- grafici a torta per analizzare la distribuzione delle licenze;
- visualizzazioni relative allo stato del pagamento delle tasse associate alle licenze.

Questa configurazione consente all'utente di ottenere, a colpo d'occhio, un quadro generale riassuntivo della situazione corrente.

Al fine di migliorare l'organizzazione delle informazioni e facilitare la navigazione, la dashboard è stata strutturata utilizzando anche delle **tab tematiche**, ciascuna dedicata a un diverso contesto di analisi. Le tab consentono di suddividere la dashboard in sezioni logiche, evitando il sovraffollamento di grafici e rendendo l'interfaccia più chiara e fruibile.

In particolare, sono state definite tab dedicate a:

- analisi territoriale dei tesserini e delle licenze;
- analisi temporale dell'andamento delle comunicazioni e dei rilasci;
- analisi specifica sui cacciatori e sulle caratteristiche delle licenze.

All'interno di ciascuna tab sono presenti visualizzazioni coerenti con il contesto di riferimento, permettendo all'utente di passare rapidamente da una vista generale a un'analisi più dettagliata.

L'utilizzo delle dashboard e delle tab di Superset ha permesso di verificare come la piattaforma sia in grado di supportare **scenari di monitoraggio complessi**, offrendo al contempo un'interfaccia intuitiva e orientata all'utente finale.

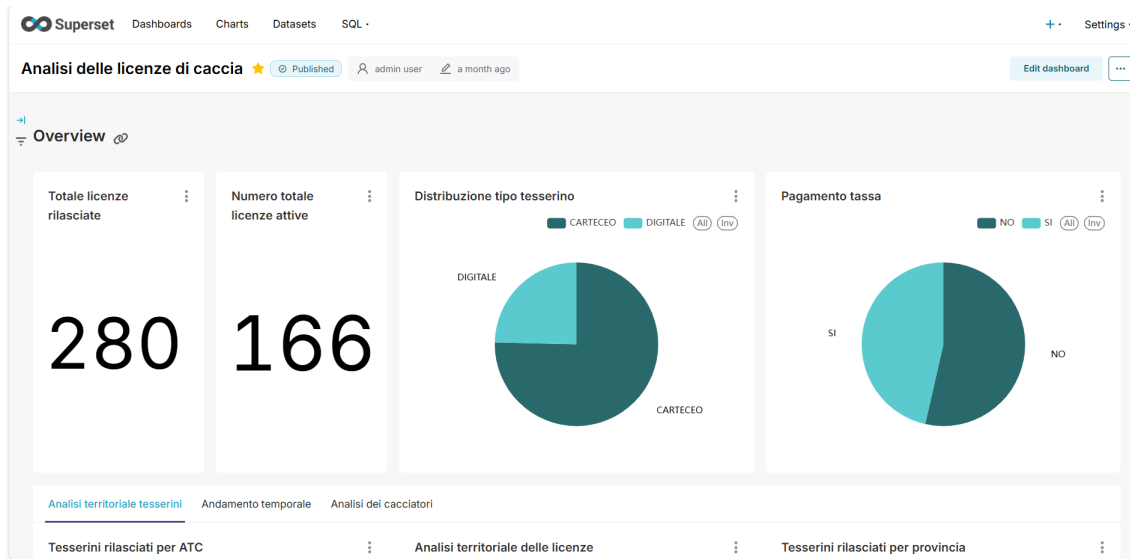


Figura 5.3. Presentazione della dashboard: overview e divisione in tab tematiche

5.3.4 Rappresentazioni tabellari dei dati

Accanto alle visualizzazioni grafiche, nel Proof of Concept è stata data particolare attenzione alle **rappresentazioni tabellari dei dati**, in quanto costituiscono una delle modalità di consultazione più richieste dagli utenti finali. L'esigenza primaria, infatti, non è limitata alla sintesi visiva delle informazioni, ma riguarda la possibilità di accedere all'intero dettaglio delle singole registrazioni, in modo analogo a quanto avviene nel sistema AS-IS.

Superset consente di integrare all'interno delle dashboard **tabelle interattive** che mostrano tutte le colonne e le tuple restituite da un dataset o da una vista del database. Nel PoC, tali rappresentazioni sono state utilizzate per visualizzare i dati derivanti dalla vista materializzata creata sul database GECCO, consentendo la consultazione completa delle informazioni relative alle comunicazioni e alle licenze di caccia.

Le tabelle possono essere configurate per supportare funzionalità come **ordinamento delle colonne e paginazione**, rendendo possibile un'analisi esplorativa

anche su dataset di dimensioni medio-grandi. Inoltre, le rappresentazioni tabellari sono pienamente integrate con i meccanismi di **filtraggio della dashboard**: i filtri applicati a livello globale o tramite interazione con altri grafici influenzano dinamicamente anche il contenuto delle tabelle.

Nel contesto in esame, è stata prevista la presenza di **una tabella per ciascuna vista** del dominio applicativo, in modo da offrire agli utenti un accesso diretto e completo ai dati sottostanti alle analisi grafiche. Sebbene nella fase di test sia stata implementata una sola rappresentazione tabellare, l'approccio adottato risulta facilmente estendibile a scenari più complessi.

Analisi delle licenze di caccia Published admin user a month ago Edit dashboard

Geco

Show 200 entries per page

ID_COMUNICAZIONE	ID_TIPO_COMUNICAZIONE	COGNOME	NOME	DATA_NASCITA	ID_COMUNE_NASCITA	DENOM_COMUNE_EST_NASCITA	ID_COMUNE_RESIDENZA	DENOM_COMUNE_EST_RESID
11747	47	mastrangelo	rosaria	1987-06-18 00:00:00	1914	N/A	1632	N/A
3294	47	Ciancia	Domenico	1989-09-03 00:00:00	1632	N/A	1632	N/A
7958	47	MARINELLI	MARIANO ANTONIO	1997-08-14 00:00:00	1897	N/A	1897	N/A
7947	47	PASCARELLI	PAOLO	1965-01-28 00:00:00	1856	N/A	1856	N/A
7956	47	DI BRIZZI	COSIMO	2004-09-17 00:00:00	1914	N/A	1609	N/A
7952	47	DI BRIZZI	COSIMO	2004-09-17 00:00:00	1914	N/A	1609	N/A
7951	47	ROSSI	MARIO	1952-03-02 00:00:00	1888	N/A	1897	N/A
7950	47	GRIPPO	PASQUALE	1976-05-10 00:00:00	2453	N/A	1930	N/A
7948	47	ROSSI	MARIO	1952-03-02 00:00:00	1888	N/A	1897	N/A
7946	47	GRIPPO	PASQUALE	1976-05-10 00:00:00	2453	N/A	1930	N/A
7930	47	CIPOLLA	MARIO	1956-12-03 00:00:00	1913	N/A	1913	N/A
7929	47	PIZZO	GIOVANNI	2000-05-19 00:00:00	1897	N/A	1616	N/A
7927	47	PIZZO	GIOVANNI	2000-05-19 00:00:00	1897	N/A	1616	N/A

Figura 5.4. Caption

5.3.5 Visualizzazioni geografiche e mappe

Nell'ambito del Proof of Concept è stata dedicata attenzione anche alla sperimentazione delle **visualizzazioni geografiche**, ritenute di elevato interesse per il contesto applicativo analizzato. Sebbene il sistema AS-IS non preveda attualmente una componente cartografica avanzata, la natura dei dati trattati e le possibili evoluzioni future rendono questo tipo di analisi rilevante.

Superset mette a disposizione diverse tipologie di visualizzazioni geospaziali basate sulla **libreria Deck.gl**, che consente la rappresentazione di dati geografici sia tramite coordinate geografiche sia attraverso oggetti cartografici più complessi. Al fine di testare tali funzionalità, è stato necessario intervenire sulla configurazione dell'applicativo, abilitando l'utilizzo di mappe esterne.

In particolare, è stato configurato un file di impostazioni personalizzato per Superset, che consente l'integrazione con servizi di mapping basati su **Mapbox** e l'utilizzo di mappe di base alternative, come **OpenStreetMap**. Questa configurazione ha permesso di verificare la corretta visualizzazione delle mappe e l'interazione con i layer geografici.

Per la realizzazione delle visualizzazioni geografiche sono state utilizzate, a scopo di test, colonne di latitudine e longitudine presenti nel dataset. Tali coordinate, attualmente generate con dati di prova, hanno consentito di simulare la distribuzione territoriale delle licenze e delle comunicazioni, permettendo di validare il corretto funzionamento dei grafici cartografici anche in assenza di dati geografici completi nel database di produzione.

Tra i grafici testati ci sono:

- le **Deck.gl Heatmap**, utilizzate per rappresentare la densità di eventi in determinate aree geografiche sulla base delle coordinate;
- le **Deck.gl GeoJSON**, che consentono di caricare e visualizzare oggetti geografici complessi, come confini amministrativi o aree territoriali, direttamente a partire da dati presenti nel database.

L'utilizzo dei grafici GeoJSON risulta particolarmente interessante per il contesto in esame, in quanto permette di rappresentare graficamente gli **Ambiti Territoriali di Caccia (ATC)**. Gli ATC sono suddivisioni territoriali definite a livello regionale per la gestione e il controllo dell'attività venatoria, e costituiscono un elemento centrale nell'analisi dei dati relativi alle licenze di caccia. La possibilità

di visualizzarne i confini su mappa apre a scenari di analisi avanzata, come il confronto tra aree o l'integrazione con dati ambientali e faunistici.

Sebbene tali visualizzazioni siano state realizzate in una fase sperimentale e con dati parziali, l'attività di test ha permesso di dimostrare la fattibilità dell'integrazione di analisi geografiche all'interno della piattaforma. In **prospettiva futura**, la disponibilità di dati più dettagliati potrebbe consentire la realizzazione di dashboard dedicate al monitoraggio delle aree di caccia maggiormente utilizzate o alla distribuzione geografica degli animali abbattuti, ampliando significativamente le capacità informative del sistema rispetto alla soluzione AS-IS.

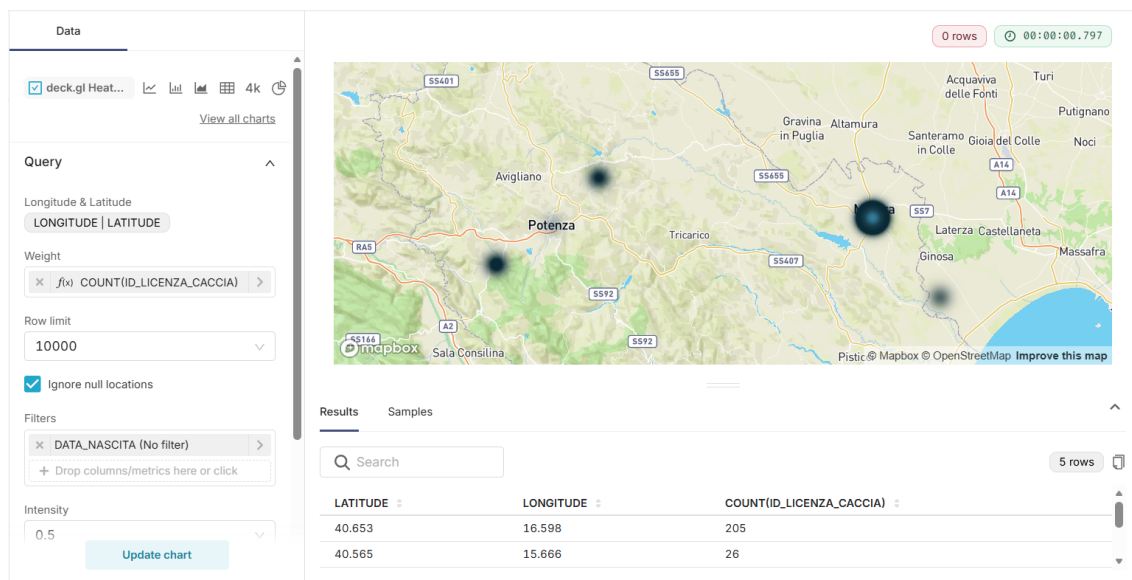


Figura 5.5. Editor grafico per la creazione di mappe geografiche

5.3.6 Filtri, interazione e navigazione dei dati

Un altro degli aspetti rilevanti emersi durante la realizzazione del Proof of Concept riguarda le funzionalità di **interazione e filtraggio dei dati** offerte da Superset, fondamentali per consentire un'esplorazione dinamica delle informazioni e supportare l'analisi decisionale. La piattaforma mette a disposizione diversi meccanismi

di filtro che operano a livelli differenti, permettendo all'utente di navigare i dati in modo intuitivo e flessibile senza la necessità di scrivere query SQL.

Filtri interattivi basati sui grafici

Una prima modalità di filtraggio è rappresentata dall'**interazione diretta con i grafici** presenti all'interno delle dashboard. In questo caso, l'utente può selezionare uno o più elementi di una visualizzazione (ad esempio una barra o una porzione di un grafico a torta, ecc...) per applicare automaticamente un filtro contestuale agli altri grafici collegati alla dashboard.

Questo tipo di interazione consente un'analisi esplorativa immediata, in cui la selezione di un valore o di una categoria aggiorna dinamicamente l'intera vista, permettendo di approfondire rapidamente specifici sottoinsiemi di dati. Questo approccio risulta particolarmente efficace per utenti non tecnici, in quanto non richiede alcuna configurazione preliminare.

Filtri di dashboard configurabili

Accanto ai filtri interattivi, Superset offre un secondo livello di **filtraggio più strutturato**, basato sulla creazione di filtri di dashboard dedicati. Questi filtri vengono definiti tramite un'apposita schermata di configurazione e sono associati esplicitamente a un dataset e a una colonna specifica del modello dati.

Durante la fase di configurazione è possibile personalizzare diversi aspetti del filtro, tra cui:

- il nome e la descrizione del filtro, per migliorarne la comprensibilità per l'utente;
- il tipo di selezione (singolo valore o selezione multipla);
- il comportamento del filtro;
- la possibilità di impostare valori di default o consentire la ricerca testuale all'interno dell'elenco dei valori.

In questo modo l'utente può navigare i dati secondo **dimensioni chiave predefinite**, come periodo temporale, stato di una pratica, area territoriale o tipologia di licenza.



Figura 5.6. Sezione per la selezione dei filtri strutturati

Navigazione e fruizione delle dashboard

L'integrazione dei due meccanismi di filtraggio contribuisce a rendere la navigazione delle dashboard altamente interattiva e orientata all'utente. La combinazione di filtri espliciti e interazioni dirette consente di passare da una visione aggregata a un livello di dettaglio più specifico in modo fluido, migliorando significativamente l'esperienza di utilizzo.

5.3.7 Gestione degli utenti, ruoli e sicurezza

Un aspetto centrale nella valutazione della piattaforma di Business Intelligence riguarda la **gestione degli utenti e dei meccanismi di sicurezza**, elementi particolarmente critici nel contesto della Pubblica Amministrazione, dove l'accesso ai dati deve essere regolamentato in modo rigoroso e tracciabile. Nell'ambito del Proof of Concept, Superset è stato analizzato anche sotto questo profilo, verificando la sua capacità di supportare scenari multi-utente con differenti livelli di autorizzazione.

Superset adotta un modello di sicurezza basato su **ruoli e permessi**, che consente di definire in modo granulare le operazioni consentite a ciascun utente. La piattaforma mette a disposizione una serie di ruoli predefiniti, pensati per coprire i principali scenari di utilizzo del sistema. In particolare, sono previsti:

- ruoli con privilegi completi di amministrazione del sistema (**Admin**);
- ruoli dedicati allo sviluppo e alla creazione di contenuti analitici (**Alpha**);
- ruoli orientati prevalentemente alla consultazione e all'analisi interattiva dei dati (**Gamma**);
- un ruolo minimale per l'accesso pubblico o non autenticato (**Public**);
- un ruolo specifico per l'utilizzo avanzato dell'editor SQL (**SQL Lab**).

Questa articolazione consente di separare in modo chiaro le responsabilità tra chi gestisce la configurazione e la sicurezza della piattaforma, chi realizza dashboard e grafici, e chi accede ai dati esclusivamente in modalità di consultazione.

E' stata inoltre sperimentata la creazione di **ruoli personalizzati**, al fine di adattare il modello di sicurezza alle esigenze specifiche del dominio applicativo. In particolare, è stato definito un ruolo dedicato al contesto GECO, ottenuto come **estensione di un ruolo di tipo Gamma**, che eredita le funzionalità di consultazione dei dataset, dei grafici e delle dashboard, senza consentire modifiche alla configurazione del sistema. Questo ruolo è stato associato a un utente applicativo

creato appositamente, pensato per simulare l’accesso di operatori incaricati della consultazione dei dati relativi al sistema GECCO.

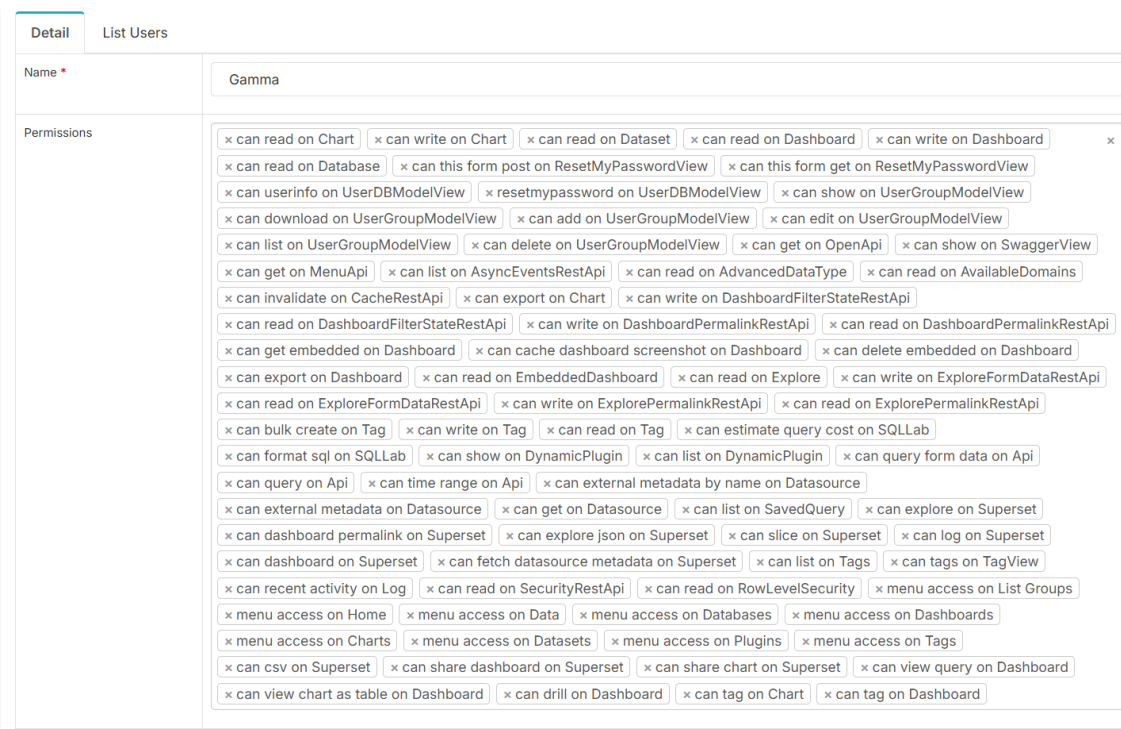


Figura 5.7. Operazioni permessi ai ruoli Gamma, da cui sono stati estesi i ruoli personalizzati

Un altro elemento di particolare importanza è rappresentato dall’utilizzo della **Row Level Security (RLS)**, funzionalità che consente di limitare la visibilità delle righe di un dataset in base all’utente o al ruolo associato. Attraverso la definizione di regole di **sicurezza a livello di riga**, è possibile applicare filtri automatici ai dati restituiti dalle query, impedendo l’accesso a informazioni non autorizzate. Questa funzionalità è stata utilizzata per simulare scenari in cui un utente può visualizzare esclusivamente i dati relativi a **specifici Ambiti Territoriali di Caccia (ATC)**, garantendo così una **separazione logica delle informazioni** all’interno dello stesso dataset. L’adozione della Row Level Security permette di implementare politiche di accesso avanzate senza la necessità di duplicare dataset o creare

viste dedicate per ciascun profilo utente, riducendo la complessità di gestione e migliorando la manutenibilità del sistema.

The screenshot shows a modal window titled "Edit Rule" with a close button (X) in the top right corner. The form contains the following fields:

- Rule Name ***: A text input field containing "ATC_1_2_Only".
- Filter Type**: A dropdown menu with "Regular" selected.
- Datasets ***: A dropdown menu with "geco.mv_comunicaz_licenza_caccia" selected.
- Roles**: A dropdown menu with "Gamma_extension_GECO" selected.
- Group Key**: An empty text input field.
- Clause ***: A text input field containing the SQL clause "DESCRIZIONE_SEDE_TERRITORIALE = 'ATC 1' OR DES".
- Description**: A large empty text area.

At the bottom of the dialog, there are two buttons: "Cancel" (light blue) and "Save" (dark blue).

Figura 5.8. Schermata di definizione delle regole per la RLS

A completamento del modello di sicurezza, Superset mette a disposizione un **sistema di logging e tracciamento delle operazioni**, che consente di registrare le principali azioni effettuate dagli utenti all'interno della piattaforma. Il registro delle attività include informazioni relative agli accessi e alle interrogazioni effettuate. Questo meccanismo di audit rappresenta un elemento essenziale per garantire trasparenza, controllo e conformità alle politiche di sicurezza richieste.

Logs ▼ Filter List

Refresh ↻

	User	Action	dtm
🔍	admin user	RLSRestApi.get_list	2026-01-29 12:24:39.830766
🔍	admin user	RLSRestApi.get_list	2026-01-29 12:17:12.208190
🔍	admin user	RLSRestApi.get	2026-01-29 11:51:41.255573
🔍	admin user	log	2026-01-29 11:51:38.833363
🔍	admin user	RLSRestApi.get_list	2026-01-29 11:51:37.957399
🔍	admin user	RLSRestApi.info	2026-01-29 11:51:37.921468
🔍	admin user	DatabaseRestApi.get_list	2026-01-29 11:51:37.844267
🔍	admin user	DatabaseRestApi.get_list	2026-01-29 11:51:37.518254

Figura 5.9. Schermata di log in Superset

5.3.8 Esportazione e condivisione dei dati

Un'ulteriore funzionalità analizzata nel Proof of Concept riguarda le modalità di **esportazione e condivisione dei dati**, aspetto di particolare rilevanza in questo contesto, in cui spesso è necessario riutilizzare i dati analizzati per attività di reportistica o verifica amministrativa.

Superset consente lo **scarico massivo dei dati** associati a grafici e dashboard direttamente dall'interfaccia web, permettendo l'esportazione dei risultati delle query nei formati più comuni, come CSV ed Excel. Questa funzionalità consente agli utenti autorizzati di ottenere rapidamente una copia strutturata dei dati visualizzati, senza la necessità di accedere direttamente al database o di eseguire interrogazioni manuali tramite strumenti esterni. L'esportazione può essere effettuata sia a livello di singolo grafico sia a partire da viste tabellari derivate dalle dashboard.

5.4 Risultati dei test e valutazioni

La fase finale del Proof of Concept è stata dedicata all'**verifica delle funzionalità implementate e alla valutazione complessiva della piattaforma** rispetto

agli obiettivi definiti nella fase di analisi dei requisiti. I test sono stati condotti in ambiente di test e hanno avuto natura prevalentemente **manuale ed esplorativa**, con l'obiettivo di valutare il comportamento del sistema in scenari d'uso realistici.

5.4.1 Test funzionali

I test funzionali sono stati finalizzati a verificare il corretto funzionamento delle principali componenti della piattaforma. In particolare, sono stati testati:

- la connessione al database Oracle GECO e la corretta interrogazione dei dati;
- la creazione e l'aggiornamento di grafici e dashboard;
- l'applicazione dei filtri interattivi e dei filtri globali di dashboard;
- le rappresentazioni tabellari dei dati;
- le visualizzazioni geografiche basate su coordinate e layer cartografici;
- i meccanismi di esportazione dei dati in formato CSV ed Excel;
- la gestione degli utenti, dei ruoli e delle regole di sicurezza.

I test hanno evidenziato un comportamento coerente con le aspettative, confermando la fattibilità tecnica dell'utilizzo di Apache Superset come piattaforma di Business Intelligence per il dominio applicativo considerato. Le funzionalità testate risultano pienamente operative e allineate ai requisiti individuati nella fase di analisi.

5.4.2 Test di usabilità

La valutazione dell'usabilità della piattaforma è stata condotta considerando l'esperienza di utilizzo dell'interfaccia web di Superset. Sono stati analizzati aspetti come:

- chiarezza dell'interfaccia di navigazione;
- facilità di consultazione delle dashboard;

- comprensibilità delle visualizzazioni;
- interazione con i filtri e con le tabelle;
- accesso alle funzionalità di esportazione dei dati.

Nel complesso, Superset si è dimostrato uno **strumento intuitivo per la consultazione delle dashboard**, soprattutto una volta completata la fase di configurazione iniziale. Le funzionalità di interazione diretta con i grafici e la presenza di filtri favoriscono un'esplorazione autonoma dei dati, riducendo la necessità di competenze tecniche avanzate per l'utente finale.

5.4.3 Limiti emersi nella fase di test

Durante la fase di test sono emerse alcune **criticità e limitazioni**, in particolare:

- la fase di installazione e configurazione iniziale richiede competenze tecniche;
- alcune funzionalità avanzate (ad esempio integrazione con sistemi di autenticazione esterni come SPID) non sono direttamente testabili in ambiente di test, come vedremo nei prossimi capitoli;
- la configurazione di ruoli e permessi richiede una fase iniziale di tuning per adattarsi alle esigenze organizzative;
- le visualizzazioni geografiche dipendono fortemente dalla disponibilità e qualità dei dati territoriali, non sempre presenti nel database di partenza.

Questi limiti non compromettono la validità del Proof of Concept, ma rappresentano aspetti da considerare in un'eventuale fase di messa in produzione, che richiederebbe un'analisi più approfondita e un coinvolgimento diretto delle strutture IT dell'ente.

Capitolo 6

Conclusione e sviluppi futuri

6.1 Considerazioni conclusive sul lavoro svolto

Il lavoro di tesi ha affrontato il tema della migrazione di una piattaforma di Business Intelligence da soluzioni proprietarie a tecnologie open source, attraverso lo studio e la realizzazione di un Proof of Concept applicato a un contesto reale della Pubblica Amministrazione. L'obiettivo principale non era la semplice sostituzione tecnologica di uno strumento esistente, ma la valutazione critica della fattibilità, dei benefici e dei limiti di un approccio flessibile e sostenibile nel medio-lungo periodo. Nel corso del progetto è stata condotta un'analisi approfondita del dominio applicativo, delle esigenze informative e dei requisiti funzionali e non funzionali, con particolare attenzione ai vincoli tipici dei sistemi informativi pubblici. L'analisi comparativa delle principali soluzioni open source di Business Intelligence ha permesso di individuare Apache Superset come piattaforma più idonea per la realizzazione del Proof of Concept, in virtù delle funzionalità offerte e della compatibilità con infrastrutture eterogenee.

La realizzazione del PoC ha consentito di validare concretamente le scelte progettuali effettuate, dimostrando la capacità della piattaforma di supportare la connessione a sorgenti dati reali, la creazione di dashboard interattive, la consultazione

dettagliata dei dati tramite rappresentazioni tabellari e l'applicazione di politiche di sicurezza avanzate. In particolare, l'attenzione posta alla gestione degli utenti, dei ruoli e della sicurezza ha evidenziato come anche una soluzione open source possa rispondere efficacemente alle esigenze di governance del dato richieste.

Pur essendo stato sviluppato in un ambiente di test, il Proof of Concept ha rappresentato un valido strumento di sperimentazione e di comprensione delle implicazioni tecniche e organizzative di una possibile migrazione. Il lavoro svolto ha permesso di individuare con chiarezza i prerequisiti necessari per una futura messa in produzione, evidenziando al contempo i limiti attuali legati sicurezza e integrazione con i servizi istituzionali.

In conclusione, la tesi fornisce un contributo concreto alla valutazione dell'adozione di soluzioni open source di Business Intelligence in ambito pubblico, proponendo un percorso metodologico e tecnico replicabile anche in altri contesti. Il progetto dimostra come un approccio basato su tecnologie aperte possa rappresentare una valida alternativa alle soluzioni proprietarie, ponendo le basi per sistemi informativi più sostenibili.

6.2 Prerequisiti per la messa in produzione

La realizzazione del Proof of Concept ha consentito di validare la fattibilità dell'adozione di Apache Superset in un contesto di Business Intelligence, tuttavia l'ambiente predisposto rimane a livello di test e non rispecchia pienamente i requisiti di sicurezza e affidabilità richiesti per un'effettiva **messa in produzione**. Sulla base dell'analisi della documentazione ufficiale di Superset è stato possibile individuare un insieme di **prerequisiti necessari da soddisfare** per un eventuale sviluppo in ambiente operativo.

6.2.1 Sicurezza dell'applicazione

Prima di essere esposto su un dominio di produzione, Superset deve essere reso sicuro secondo le best practice descritte nella documentazione. Queste includono:

- **Configurazione di HTTPS/TLS** per tutte le comunicazioni tra client e server, al fine di tutelare la riservatezza e l'integrità dei dati trasmessi;
- **Limitazione degli accessi alle interfacce amministrative** mediante firewall, VPN o segmentazione di rete;
- **Gestione delle secret key e di credenziali sensibili** tramite vault o meccanismi crittografici sicuri, evitando valori in chiaro nei file di configurazione;
- **Aggiornamento e patch management** continuo dell'applicazione e delle dipendenze, per ridurre il rischio di vulnerabilità note.

6.2.2 Utilizzo di un WSGI server per l'ambiente di produzione

Un ulteriore prerequisito riguarda l'utilizzo di un **WSGI server** dedicato, in sostituzione del server di sviluppo integrato utilizzato durante la fase di test. Il server WSGI consente di gestire in modo più efficiente le richieste concorrenti, migliorando l'affidabilità e la scalabilità dell'applicazione.

In particolare, viene raccomandato l'utilizzo di **Gunicorn** come WSGI server. Gunicorn permette di avviare Superset come servizio applicativo stabile, configurabile con un numero adeguato di worker e thread, in grado di supportare accessi simultanei da parte di più utenti.

L'adozione di un WSGI server consente inoltre una più agevole integrazione con componenti infrastrutturali tipici come **reverse proxy** (ad esempio Nginx o Apache HTTP Server).

6.2.3 Autenticazione istituzionale

Un requisito strategico per il contesto della Pubblica Amministrazione è l'integrazione con **sistemi di autenticazione esterni e istituzionali**, in particolare SPID. L'obiettivo è consentire l'accesso alla piattaforma utilizzando le credenziali uniche degli utenti pubblici, in linea con le normative vigenti in materia di identità digitale e sicurezza dei servizi web. Per supportare SPID, è necessario prevedere:

- configurazioni SAML (Security Assertion Markup Language) o OAuth;
- gestione dei gruppi e dei ruoli in base agli attributi dell'identità;
- mapping dei ruoli interni di Superset e le identità riconosciute.

6.2.4 Monitoraggio, logging e auditing

Per soddisfare i requisiti di compliance e governance, l'ambiente di produzione deve includere:

- monitoraggio delle performance dell'applicazione;
- log centralizzati delle operazioni degli utenti;
- strumenti di alerting per anomalie o errori;

Superset fornisce meccanismi di logging che possono essere integrati con strumenti esterni, consentendo la tracciabilità delle attività e la verifica retrospettiva delle operazioni critiche.

6.2.5 Backup, disaster recovery e continuità operativa

L'ambiente di produzione deve prevedere procedure per:

- backup regolari del database di metadati di Superset;
- backup delle configurazioni, dashboard, dataset;
- piani di disaster recovery per garantire la continuità del servizio in caso di guasti hardware, software o eventi imprevisti.

6.3 Integrazione con sistemi di autenticazione istituzionali (SPID)

Un ulteriore sviluppo futuro per l'adozione della piattaforma in ambiente di produzione riguarda l'integrazione con **sistemi di autenticazione istituzionali**, in particolare con lo SPID (Sistema Pubblico di Identità Digitale), già ampiamente utilizzato dalla Pubblica Amministrazione per l'accesso ai servizi digitali rivolti a cittadini e operatori.

Nel corso del progetto è stata analizzata una possibile architettura di integrazione con SPID. L'approccio considerato prevede l'utilizzo di un **reverse proxy**, già presente nell'infrastruttura aziendale esistente per la regione Basilicata, con il ruolo di intermediario tra l'utente finale, il sistema di autenticazione SPID e l'applicativo Superset.

In questo scenario, Superset non gestisce direttamente il processo di autenticazione, ma delega tale responsabilità a un componente esterno. Il flusso di autenticazione può essere descritto sinteticamente come segue:

1. l'utente accede all'**indirizzo pubblico dell'applicazione** di Business Intelligence;
2. il **reverse proxy intercetta la richiesta** e verifica l'assenza di una sessione autenticata; in tal caso, l'utente viene reindirizzato alla schermata di login SPID tramite il provider di identità;
3. una volta completata con successo l'autenticazione, il reverse proxy riceve le informazioni relative all'identità digitale e consente l'accesso a Superset, **propagando gli attributi dell'utente autenticato attraverso header HTTP o variabili di ambiente** opportunamente configurate.

Per rendere possibile tale integrazione, è necessario che l'applicazione sia esposta tramite un **indirizzo IP pubblico**, analogamente agli altri servizi digitali già erogati dalla Regione Basilicata. Questo requisito è indispensabile per garantire il

corretto funzionamento del flusso di autenticazione tramite SPID, in particolare per la gestione delle callback e dei meccanismi di sicurezza previsti dal protocollo di autenticazione.

Un elemento centrale dell'architettura analizzata riguarda la **mappatura tra l'identità SPID e gli utenti** interni di Superset. La soluzione ipotizzata prevede l'utilizzo del **codice fiscale come identificativo univoco dell'utente**, impostandolo come nome utente all'interno della piattaforma. Questa scelta consentirebbe di semplificare l'associazione tra identità digitale e profilo applicativo, sfruttando un database aziendale già esistente in grado di stabilire quali codici fiscali siano autorizzati all'accesso al sistema e con quali ruoli. In questo modo, il **processo di autorizzazione rimarrebbe separato dall'autenticazione**: SPID garantirebbe l'identità dell'utente, mentre Superset, sulla base del codice fiscale ricevuto, applicherebbe le politiche di accesso, i ruoli e le eventuali regole di Row Level Security precedentemente definite.

L'integrazione con SPID non è stata implementata nel Proof of Concept, in quanto richiede la disponibilità di un ambiente di produzione, la configurazione di componenti infrastrutturali dedicati e il coinvolgimento diretto delle strutture IT dell'ente. Tuttavia, l'analisi svolta fornisce una base solida per una futura evoluzione del progetto.

6.4 Sviluppi futuri e linee guida per una migrazione completa

Il lavoro svolto consente di delineare con maggiore consapevolezza una serie di **sviluppi futuri e di linee guida** utili per accompagnare l'evoluzione del progetto.

Un primo ambito di sviluppo riguarda, come è già stato analizzato, il completamento delle **attività infrastrutturali e di sicurezza**, attraverso l'adozione delle best practice analizzate per la messa in produzione e l'integrazione di una modalità di

autenticazione istituzionale. In questa fase sarebbe necessario consolidare l'architettura applicativa, definire ambienti separati per sviluppo, test e produzione, e configurare in modo definitivo i meccanismi di monitoraggio, logging e backup, garantendo la continuità operativa del servizio nel tempo.

Dal punto di vista funzionale, la piattaforma potrebbe essere ulteriormente estesa **arricchendo il modello dati e le dashboard** disponibili, introducendo nuovi indicatori, metriche avanzate e visualizzazioni tematiche. In particolare, l'utilizzo di **dati geografici più completi** consentirebbe di potenziare le analisi territoriali. Un altro aspetto interessante riguarda la possibilità di **riuso della soluzione in altri contesti** applicativi della Pubblica Amministrazione. L'approccio adottato, basato su tecnologie open source, favorisce la replicabilità del progetto anche in ambiti diversi da quello analizzato, contribuendo alla diffusione di soluzioni condivise e alla riduzione dei costi complessivi di sviluppo e manutenzione.

Infine, il lavoro svolto fornisce una base metodologica utile per guidare futuri progetti di migrazione, evidenziando l'importanza di una fase preliminare di analisi dei requisiti, di una valutazione comparativa delle soluzioni disponibili e di una sperimentazione controllata. Questi elementi rappresentano fattori chiave per aumentare la qualità delle soluzioni adottate e garantire una transizione graduale ed efficace verso sistemi di Business Intelligence moderni e sostenibili.

Bibliografia

- [1] “Business intelligence chiave di volta per la PA data driven: ecco come usarla”. In: (). URL: <https://www.corrierecomunicazioni.it/pa-digitale/business-intelligence-chiave-di-volta-per-la-pa-data-driven-ecco-come-usarla/>.
- [2] “Che cos’è la Business Intelligence (BI)? — IBM”. In: (). URL: <https://www.ibm.com/it-it/think/topics/business-intelligence>.
- [3] “Concetti relativi a metabase — Metabase Imparare”. In: (). URL: <https://www.metabase.com/learn/metabase-basics/overview/concepts>.
- [4] “Embedded Analytics and Reporting Platform”. In: (). URL: <https://www.jaspersoft.com>.
- [5] “Getting Started — Pentaho Academy”. In: (). URL: <https://academy.pentaho.com/pentaho-business-analytics/business-analytics/getting-started>.
- [6] “intro — Superset”. In: (). URL: <https://superset.apache.org/docs/intro>.
- [7] “Nozioni di base della Business Intelligence: B.I. for Dummies!” In: (). URL: <https://www.smeup.com/magazine/blog/nozioni-di-base-della-business-intelligence/>.

- [8] “Open-Source BI Tool Metabase for Data Visualization and Analytics”. In: (). URL: <https://www.analyticsvidhya.com/blog/2024/01/open-source-bi-tool-metabase-for-data-visualization-and-analytics/>.
- [9] “Open-Source vs Software Proprietario per Piccole Imprese: Risparmio e Scalabilità in Ambiente Business”. In: (). URL: <https://blog.giotech.net/open-source-vs-software-proprietario-piccole-imprese/>.
- [10] “Pentaho Business Analytics — Pentaho”. In: (). URL: <https://pentaho.com/products/pentaho-business-analytics/>.
- [11] “Redash helps you make sense of your data”. In: (). URL: <https://redash.io/>.

Ringraziamenti

Giungere alla fine di questo percorso magistrale mi riempie di una gioia profonda, ma anche di una gratitudine immensa verso tutte le persone che hanno reso possibile questo traguardo, semplicemente facendomi diventare la persona che sono oggi. Queste parole non basterebbero mai a contenere tutto ciò che vorrei dire, ma ci provo ugualmente.

Desidero ringraziare il mio relatore: il professore Luca Ardito per avermi dato l'opportunità di sviluppare questo lavoro di tesi e per il supporto istituzionale fornitomi durante il percorso; il mio tutor aziendale Luca Ottone per l'accoglienza, la disponibilità e il supporto offerto durante l'esperienza in azienda, che si è rivelata formativa sia dal punto di vista professionale che personale; e infine al grande Gerry per la pazienza e la disponibilità dimostrata fin dal primo giorno nei miei confronti, e per avermi fatto sentire parte di qualcosa con grande affetto.

Ringrazio tutta la mia amatissima famiglia, mi avete cresciuto con amore e dei valori incredibili, che mi hanno reso la persona che sono oggi. Dedico questo traguardo a tutti i miei zii e zie; ai miei cugini Claudio e Marco, con cui sono cresciuto tutti i giorni della mia vita e che sono grato di aver avuto in quest'ultima, come fossero gli amici migliori e sempre presenti con cui condividere tutto; a Peppe ed Enzo, i miei piccoli figliocci a cui cerco tutt'oggi di tramandare solo il meglio di me, ho sempre avuto cura di ascoltare ogni vostra lamentela per darvi un consiglio che potesse esservi utile e mi è sempre piaciuto essere per voi un buon amico, un punto di appoggio, o anche una rampa di lancio. Ai miei nonni che ci hanno sempre accolto a braccia aperte e che hanno fatto da secondo paio di genitori per tutti noi cugini, e che più di tutti ci hanno insegnato i valori che una famiglia unita dovrebbe avere.

Ringrazio i miei genitori, una mamma e un papà splendidi che hanno creduto in me e nelle mie capacità, offrendomi un supporto in ogni piccolo passo della mia

vita. Spero possiate essere orgogliosi di me, della persona che sono e dei valori che rappresento, anche perchè è, in parte, anche merito vostro.

Ringrazio la mia sorellona Tea, che si è sempre presa cura di me, nonostante la facessi ridere e sclerare contemporaneamente. Ti auguro di credere di più e con coraggio in te stessa e in quello che puoi offrire e di realizzare i tuoi sogni, perchè ti assicuro che vali tanto, e non solo perchè sei mia sorella.

Ringrazio Marco, un amico, un fratellone o il classico zio che ti fa bere birra e gin tonic durante tutte le festività e poi incolpa qualcun altro se ti senti male. Sei la persona che più di tutte mi fa ridere a crepapelle con i tuoi modi gentilmente goffi e lenti.

Ringrazio con il cuore e con tutto il bene che posso offrire i miei amici, che sono la fonte del mio sorriso e della mia socialità.

A tutti i miei amici di Torino con cui ho condiviso la mia quotidianità, le molte ore di studio, le avventure in montagna o i vari giovedì al coguaro. In tutti questi anni siete stati una costante in una città che non era la mia, ma che adesso lo è un po' di più principalmente grazie a voi.

Ai miei coinquilini Christian e Cristiano, che mi hanno permesso di sentirmi a casa anche in un posto che casa non era e con cui condivido un lungo sentiero percorso per le strade di Torino da cinque anni a questa parte.

Un grazie in particolare ai miei amici di giù, o la mia "famiglia" di giù. Un gruppo composto da persone, seppure diverse tra loro, unite e che si vogliono un bene dell'anima senza paura di dimostrarlo. Voi mi conoscete meglio di chiunque altro e sapete già quanto io sia grato di avervi nella mia vita. Siete la sicurezza e la voglia irrefrenabile che ho di staccare da tutto e godermi tutte le giornate possibili in vostra compagnia, che sia a Natale, a Pasqua o in estate, su un campo da beach volley, su una spiaggia o tra quattro mura, io muoio dalla voglia di divertirmi con voi. Voi vedete in me, più di tutti, l'amore e il bene che spero mi rappresentino, e sono sicuro che questo legame non si scioglierà e saremo sempre pronti a riunirci

ogni volta.

Tra voi, ritengo giusto spendere delle parole in particolare per Alex, che considero un fratello acquisito che letteralmente mi ha accolto nella sua famiglia e che mi ha saputo donare sempre le parole giuste nei momenti sereni ma soprattutto in quelli più duri. Alex, per quanto mi riguarda, ha un dono, che è quello di entrare nella mia testa per dire esattamente ciò che ho bisogno di sentirmi dire, con un giudizio e un'oggettività strepitosi. Una persona davvero speciale e un mix perfetto e pericoloso tra lealtà e disagio.

Ovviamente non potrei non ritagliare un paragrafetto per Biagio: la mia amicizia più longeva tanto che ormai ho quasi dimenticato quando è nata. Penso ci siano poche cose da poter raccontare della mia vita in cui lui non ci sia. L'unica persona che potrebbe dirmi una fake news gigante ma raccontata bene e mi ci farebbe credere con tutto me stesso, tanto da spacciare questa notizia in giro credendo di spargere pillole di cultura. Biagio è non solo un grande amico (anche se bacchettone), ma anche un grande compagno di vita e di viaggio, con cui infatti dividerò quello che spero sarà l'esperienza più bella della mia vita in un lungo mese dall'altra parte del mondo: una semplice idea, una disorganizzata proposta di andarci insieme e uno spensierato sì che ci hanno fatto intraprendere un'altra avventura. Perché, alla fine, possiamo riassumerci così.

Sento un gran bisogno di ringraziare anche Ines, perché sia la tua presenza che la tua assenza mi hanno arricchito e fatto crescere, facendomi capire che, anche quando il mondo sembra crollarmi addosso, posso farcela anche da solo. Mi hai insegnato a credere e a prendermi cura di me stesso, ad essere ambizioso, e a dare alla mia persona l'importanza che devo, in una vita che, alla fine, è prima di tutto la mia. Hai scritto un capitolo bello ed importante della mia vita, che magari immaginavo un po' diverso, ma che mi ha reso quello che sono oggi. Ti vorrò sempre bene.

Infine grazie alla mia Claudina, una persona davvero stupefacente e spaventosamente simile a me. Ammiro con grande stupore tutto l'amore che metti nei tuoi

BIBLIOGRAFIA

piccoli e grandi gesti, tutto ciò che fai per le persone a cui vuoi bene, e il grande coraggio che hai nell'affrontare tutte le tue sfide. Non mi hai mai lasciato solo e hai sempre sacrificato la tua forza per darne un po' a me, aiutandomi e prendendomi la mano nei miei momenti più bui, quelli che ti fanno dubitare di te stesso. Auguro a tutti una Claudina o una persona come te nella propria vita, che ci faccia sentire amati, o più banalmente di valere qualcosa nel cuore di qualcuno.

Grazie, vi voglio bene.

Giorgio