SUCCESS STORY

# Migrazione DWH da Oracle a BigQuery

## Come migrare un DWH Legacy con un approccio Cloud Native e Serverless

### SETTORE

Retail, Luxury Goods

### COMPETENZE

- Data Architecture
- App Modernization
- Cloud Migration
- Data Engineering

### SERVICE PROVIDER

Google Cloud

### LA SFIDA

Ogni azienda è una data company, per questo motivo il Data Warehouse (DWH) è una parte importante, se non la più importante, dell'infrastruttura IT. Il nostro cliente, il quale opera nel settore dei beni di lusso, ha iniziato a costruire il proprio DWH più di 20 anni fa. Questo processo di evoluzione incrementale ha portato ad un accumulo di hotfix ed eccezioni, senza un adeguato rinnovamento dell'infrastruttura per adeguarsi all'innovazione tecnologica e alle nuove esigenze di business del mercato.
La principale sfida, quindi, è stata risolvere problemi di stabilità e prestazioni e allo stesso tempo eseguire il refactoring del codice e documentare parti dell'infrastruttura che, a causa della loro natura legacy, erano sconosciute ai più. In particolare, numerose erano le pipeline critiche che influivano sulla generazione di report chiave in caso di errore. L'esecuzione di queste pipeline nella vecchia infrastruttura poteva richiedere ore, il che era inaccettabile data la natura urgente dei report.

### LA NOSTRA SOLUZIONE

BIP xTech ha svolto una prima fase di valutazione, necessaria per perimetrare l'ambito da rinnovare e scegliere le tecnologie adeguate. Abbiamo deciso di creare il nuovo DWH su Google Cloud Platform (GCP) utilizzando BigQuery e Dataform.
BigQuery ci offre prestazioni migliori (rispetto al vecchio hardware on-premise), si integra facilmente con gli strumenti di Business Intelligence, è conforme agli standard di sicurezza del settore e consente il controllo granulare degli accessi, e offre un modello di prezzi "pay-as-you-go" (particolarmente utile nella fase di migrazione, quando non si conosce con precisione la quantità di dati da trattare).
Dataform è l'agente di orchestrazione che esegue le query da BigQuery a BigQuery. È stata scelto perché, nel sistema legacy, la maggior parte delle trasformazioni erano già implementate in linguaggio SQL. Ciò ha consentito a BIP xTech di eseguire una pulizia completa del logic e del refactoring senza dover riscrivere tutte le pipeline e sfruttando l'attuale know-how del team DWH interno dell'azienda. L'adozione di questo strumento ha richiesto solo la formazione per la gestione della console GCP e l'utilizzo degli strumenti di controllo della versione (Git). Oltre a mantenere la continuità con il passato, Dataform ci ha permesso di sfruttare tutte le innovazioni portate negli anni dall'ingegneria del software, essendo basato sul concetto di data transformations as code (con pieno supporto di versioning delle trasformazioni). Inoltre, integra nativamente asserzioni (controlli di qualità dei dati, utilizzati per garantire che i dati soddisfino le aspettative) e unit test, in modo tale da avere una vera e propria unica fonte di verità. Altrettanto fondamentale è la capacità di documentare lo schema dei dati e le trasformazioni aggiungendo documentazione al codice Dataform. La documentazione viene aggiunta automaticamente al Data Catalog all'interno


© Copyright BIP xTech 2023

di Dataform e può essere successivamente esportata in altri strumenti. Infine, la visualizzazione dependency tree di tutte le dipendenze rilevate automaticamente da Dataform consente al visualizzatore di comprendere facilmente la relazione tra ciascuna entità del progetto e il loro ordine di esecuzione.
Portare l'infrastruttura nel cloud ci ha anche permesso di sfruttare tutte le funzionalità native offerte da GCP, che avrebbero richiesto molti sviluppi ad-hoc da implementare nel vecchio ambiente on-premise. BIP xTech ha utilizzato "table snapshot" per salvare uno stato ultraleggero della tabella in un determinato momento, consentendo il rollback istantaneo dei dati o il calcolo semplice del delta tra due momenti diversi. Abbiamo utilizzato Cloud Workflow, consentendo la gestione dell'integrazione tra Dataform e altri servizi GCP, e abbiamo monitorato l'intera infrastruttura tramite Cloud Alerting, che avvisa in tempo reale il personale preposto in caso di errori.

## I RISULTATI

L'adozione di BigQuery ha permesso di ridurre drasticamente i tempi di esecuzione della pipeline. Nella prima pipeline analizzata la riduzione è stata da due ore a pochi minuti. Abbiamo inoltre ridotto i costi, sfruttando le potenzialità del modello "pay-as-you-go" scrivendo query ottimizzate che caricano solo la porzione di dati strettamente necessaria (partition pruning) ed eliminando i costi di gestione dell'infrastruttura. Questa architettura completamente gestita è anche scalabile automaticamente e si adatta alle crescenti esigenze dell'azienda.
Grazie a Dataform abbiamo prodotto documentazione integrata nel progetto permettendo una sostanziale velocizzazione del processo di onboarding di nuovi membri del team. Inoltre, i processi di pulizia dei dati prima delle trasformazioni ei controlli dopo le trasformazioni hanno permesso di abbassare il numero di errori imprevisti e di avere dei log descrittivi quando si verificano.
L'enorme capacità di archiviazione del cloud consente di rieseguire una pipeline come se fosse in un preciso momento nel passato, permettendo di correggere rapidamente eventuali errori nei dati causati da bug introdotti nelle nuove versioni del codice.
Infine, l'uso di strumenti di controllo della versione migliora la collaborazione tra i membri del team e regola il ciclo di vita del software seguendo le best practice del settore, facilitando futuri e continui interventi di refactoring e modernizzazione, per non tornare alla situazione precedente.

## I BENEFICI

Riduzione del 90% dei tempi di esecuzione della pipeline

Eliminazione dei costi di gestione delle infrastrutture

Cronologia completa delle modifiche di tutte le pipeline disponibili

Data Quality Assurance

Documentazione integrata nello strumento


© Copyright BIP xTech 2023

## Contattaci

**Davide Tresoldi**

Senior Cloud Architect @ BIP xTech

davide.tresoldi@bip-group.com

---

**Simeone Vilardo**

Specialist @BIP xTech

simeone.vilardo@bip-group.com

bip. xTech

www.bipxtech.com

© Copyright BIP xTech 2023