

# L'architettura di Power BI, il report in cloud e la modalità DirectQuery

di [Francesco Bergamaschi](#)

Pubblicato il 28 Febbraio 2023

Dedichiamo questo intervento all'**architettura del software PowerBi**, puntando il mouse sulle **modalità e possibilità di condivisione in cloud dei report** sfruttando la **DirectQuery**.

Nel precedente articolo dal titolo "[Appunti del primo meeting su Power BI e refresh dei dati sul cloud](#)", il consulente Luca è riuscito ad automatizzare l'**aggiornamento dei dati sul cloud**, dati che alimentano il suo **primo report Power BI**.

In questo modo, non dovrà più preoccuparsi della cosa, il **report vivrà di vita propria, una volta pubblicato**.

Luca è molto sollevato, questa è stata una prima palestra per essere pronti alle richieste dei clienti una volta implementato il sistema Power BI presso di loro.

Luca, tuttavia, vuole fare altra esperienza per essere davvero pronto.

I soci di Luca, Dante e Francesco, gli hanno fatto i complimenti ancora una volta: ogni mattina trovano i dati del report sui margini aggiornati alle 2 del mattino del giorno stesso.

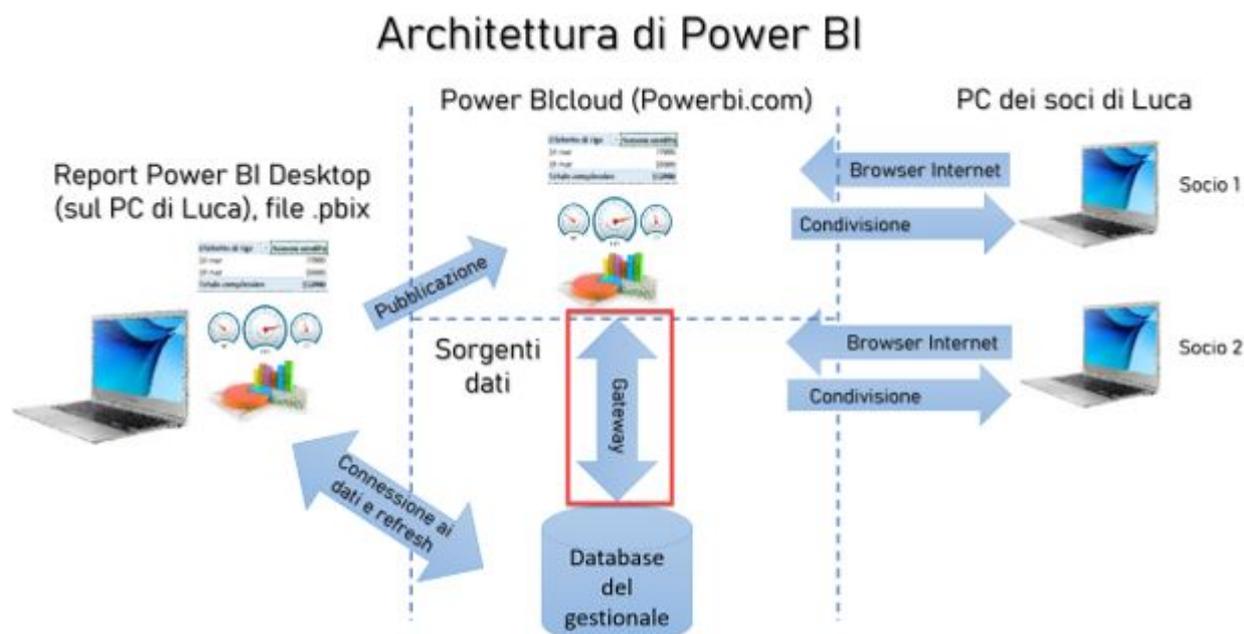
Il **report è consultabile su tablet, cellulare, pc fisso e portatile**.



## L'architettura di Power Bi

Ecco l'architettura che Luca ha implementato e compreso bene, una volta usato il gateway per il refresh dei dati sul cloud, usando il gateway per raggiungere il server dello studio.

Immagine 1



Luca ha usato la **modalità Import per il report**: i dati vengono **copiati in Power BI Desktop** e poi **pubblicati**, cioè copiati sul cloud di Microsoft. Il gateway permette di aggiornarli periodicamente.

Luca riceve però una nuova telefonata: i suoi soci vorrebbero più aggiornamenti al giorno, anzi gli chiedono se sia possibile avere nel report i dati in tempo reale.

## La modalità DirectQuery

Luca voleva già esplorare questa **modalità**, detta **DirectQuery**, **al posto dell'Import**. Adesso ha una **richiesta ufficiale dal suo cliente interno!**

Luca riflette sulla prima richiesta: **è possibile fare più aggiornamenti al giorno?**

Trova subito la risposta nella documentazione Microsoft: con la licenza Pro si possono fare fino a 8 aggiornamenti pianificati al giorno.

Luca aggiunge quindi un refresh alle 13 e uno alle 17, poco prima della fine della giornata lavorativa.

Legge anche che il refresh può essere lanciato manualmente dal cloud in qualunque momento e in questo modo non ci sono limiti al numero di refresh; tuttavia, l'approccio manuale appare scomodo e da usare solo per emergenze.

Si rende anche conto del fatto che il numero di refresh al giorno deve essere deciso sulla base di quanto dura il refresh stesso.

Nel suo caso, avendo qualche decina di migliaia di righe nelle tabelle, il refresh dura 3 o 4 minuti, dunque, il problema non si pone, ma un ipotetico cliente con milioni e milioni di righe potrebbe avere un refresh dai tempi più lunghi e così potrebbe avere senso non andare oltre un certo numero di aggiornamenti al giorno.

Luca comunica ai suoi soci di avere **umentato a tre il numero di refresh**, i quali lo ringraziamo e gli ribadiscono tuttavia di esplorare il real time: i clienti potrebbero chiederlo.

Luca si documenta, quindi sul DirectQuery. Per prima cosa, scopre che **soltanto le sorgenti di tipo database (non file Excel o csv, in altri termini) supportano DirectQuery**.

Quindi **DirectQuery è possibile soltanto se ci si connette a database SQL Server o Oracle e pochi altri**. Del resto, questi sono quelli più diffusi, gli conferma l'informatico.

### **I vantaggi del DirectQuery**

Luca trova DirectQuery molto interessante: sembra abbia soltanto dei più rispetto all'Import. Infatti, **con DirectQuery**:

- i dati sono sempre aggiornati, perché si prelevano sempre dalla sorgente;
- non esiste, dunque, il concetto stesso di refresh;

- il report è un file .pbix piccolissimo, perché è vuoto: ha solo la grafica. Così è un file snello e gestibile;
- i dati restano sui server aziendali e non vengono copiati né su Power BI Desktop né sul cloud di Microsoft, dunque nessun problema di sicurezza né di dimensione dei dati

Luca riflette sulla cosa: in effetti ha sempre pensato che l'aspetto del **copiare i dati sul data center Microsoft potesse essere un aspetto critico.**

I clienti potrebbero innervosirsi.

Ha sempre trovato tuttavia, informazioni che lo hanno tranquillizzato: **Microsoft e Amazon** sono i due gruppi che al mondo **hanno investito di più in sicurezza** e dunque la sua argomentazione verso il cliente sarebbe stata:

*“siete sicuri che i vostri dati siano più protetti nei vostri server che nei data center di Microsoft?”*

Ora però DirectQuery sembra avere eliminato il problema.

Luca crea una copia del suo primo report con un nome differente e modifica con un click la modalità da Import a DirectQuery. Nulla cambia, a prima vista.

Luca si sente contento, ma si chiede: **possibile che l'Import sia così ovviamente superato dal DirectQuery?** In Internet ha sempre letto che Import è da preferire. Qualcosa non torna.

Luca va avanti, controlla il report e funziona tutto. A questo punto, lo pubblica. Tutto OK, a prima vista. Favoloso, pensa, anche perché in DirectQuery non ci sono limiti alla dimensione dei dati a cui ci si può collegare: non dovendoli copiare in Power BI Desktop, Luca potrà collegarsi anche ai dati di un cliente di GDO, con miliardi di righe!!!

### **Attenzione ai possibili errori!**

Tuttavia, una volta aperto il report sul cloud, ottiene un errore: tutti i grafici mostrano una finestra con una x sopra e uno strano messaggio di errore *“non è possibile accedere ai dati necessari per questa*

visualizzazione”.

Luca è perplesso: **usando l'Import, il report pubblicato era funzionante immediatamente. Come mai in DirectQuery non è così?** E cosa dovrà fare per farlo funzionare?

Luca riflette sull'architettura in figura 1:

*“il gateway non dovrebbe entrarci nulla, pensa, perché non devo fare alcun refresh. Però... Ma certo! Non devo fare alcun refresh perché non ho dato sul cloud ma questo vuol dire, al contrario, che devo SEMPRE usare il gateway, in ogni momento, dal cloud o il report non si popolerà perché non avrà dati sottostanti.”*

Luca collega anche il nuovo report al gateway già installato per il refresh del report in Import, prova a rientrare sul report in DirectQuery pubblicato sul cloud e... Tutto funziona! Luca si sente sollevato e sempre più irrobustito da queste prove. Saprà come reagire alle osservazioni dei clienti.

Prova ad interagire col report sul cloud: nota una certa lentezza. Nel report in Import sul cloud, tutto era velocissimo. In questo in DirectQuery, c'è una performance decisamente più lenta. Quasi impercettibilmente, ma c'è un degrado.

Luca ci pensa e si rende conto che, adesso, Power BI Cloud deve sempre connettersi al suo server per aggiornare il report sulla base delle interazioni dell'utente.

In Import, al contrario, i dati sono sul datacenter Microsoft dove si trova anche la grafica e dunque tutto è velocissimo (il report è sul cloud insieme ai dati ivi copiati quando si pubblica).

## **Un primo svantaggio di DirectQuery: le performance in termini di velocità di reazione**

Non gli appare un aspetto trascurabile, anche perché il suo report ha sotto poche migliaia di righe, che dire ad **un cliente che abbia milioni di righe sulla prestazione di un report in DirectQuery?** Sarà indubbiamente **più lento**, sensibilmente, **del fulmine Import**.

Inoltre, Luca nota che i dati sul cloud non cambiano. Strano: è certo che nel mentre sia stata fatta qualche operazione dai suoi impiegati. Dunque, essendo in DirectQuery, si aspettava di vederli apparire.

Chiama un impiegato e gli chiede di inserire una riga nuova di ordine per un cliente dummy. L'impiegato, prontamente, esegue. Luca aspetta e aspetta. Nulla. I dati sul cloud non cambiano.

Prova a cliccare su un oggetto e allora i dati si aggiornano!

Dunque, pensa Luca, **un altro svantaggio**: in realtà non stiamo parlando di **real time**. Affinché Power BI, per un report in DirectQuery, si riconnetta all'origine dati e così mostri i dati aggiornati, l'utente deve interagire col report.

In altri termini, non basta che venga inserita una nuova riga per vedere questi dati sul cloud: l'utente deve chiedere al report qualcosa attraverso un'interazione.

Si dice, in termini informatici, che la modalità è pull (cioè si deve tirare dentro ciò che è nuovo) e non push (cioè le righe nuove nel database automaticamente implicano un riaggiornamento sul cloud).

Dunque, l'eccitazione verso **il real time** cala ancora: **non è un vero real time nel senso che da solo il report non cambia**.

Infine, Luca nota che ci vogliono dei secondi perché i dati si aggiornino in DirectQuery, dunque non è possibile vendere questa soluzione come real time, per intenderci quello che intendono gli operatori di trading.

Luca riapre, infine, Power BI Desktop e prova a modificare alcune colonne del report, per verificare come funziona il motore di connessione e modifica dei dati di Power BI Desktop (detto Power Query) in modalità DirectQuery: nota che soltanto alcune opzioni sono disponibili, molte meno di quando ha usato l'Import.

Dunque, **DirectQuery non appare poi così migliore dell'Import, anzi!**

Luca però deve dirimere un dubbio: come fare ad usare Import quando i dati sono molto estesi e non è possibile importarli per mancanza di sufficiente RAM? Luca sa che in Italia questo capiterà di rado ma vuole essere preparato sull'argomento.

Nel prossimo articolo sveleremo **come si fa ad evitare DirectQuery** anche in questi casi!

A cura Francesco Bergamaschi

Martedì 28 febbraio 2023

## POWER B.I.

### Controllo di gestione con la business intelligence per le aziende

PERCORSO ONLINE ON DEMAND

OLTRE 14 ORE DI LEZIONE DIVISE IN 5 MODULI TEMATICI

#### Introduzione al corso

Un percorso alla scoperta di **POWER BI**, l'ecosistema di **Microsoft per la Business Intelligence**.

Una **guida pratica e operativa** per apprendere come gestire e diventare autonomi in tutto il flusso di lavoro.

**Passo dopo passo**, partendo da zero, si esplorerà dettagliatamente sia la parte Desktop (per la **connessione ai dati**, la modellazione e la **creazione dell'analisi**) che la parte del **Cloud** (pubblicazione, condivisione, gestione delle aree di lavoro, implementazione della sicurezza e dei ruoli).

Il Corso prevede **oltre 14 ore formative** suddivise in 5 moduli tematici e può essere fruito *on demand*, dove e quando vuoi.



[SCOPRI DI PIU' >](#)