

MICROSOFT FABRIC & ONELAKE



SOMMARIO

COS'È MICROSOFT FABRIC	PAG. 3
DI COSA SI COMPONE	PAG. 5
APPROCCIO UNIFICATO E PUNTI DI FORZA	PAG. 7
GESTIONE DEGLI UTENTI	PAG. 9
 POTENZIALITÀ DI ONELAKE	 PAG. 11
UN'UNICA COPIA DEI DATI	PAG. 12
L'EQUIVALENTE DI ONEDRIVE PER I DATI	PAG. 13
FORMATO STANDARD DI MEMORIZZAZIONE	PAG. 14
ARCHITETTURA MEDALLION	PAG. 15
SHORTCUTS	PAG. 16
OTTIMIZZAZIONE A PIÙ MOTORI ANALITICI	PAG. 18
MODALITÀ DIRECT LAKE	PAG. 19
 CONVIENE PASSARE A FABRIC?	 PAG. 21
SCENARIO "GREEN-FIELD"	PAG. 23
SCENARIO "FABRIC READY"	PAG. 24
SCENARIO "JOURNEY TO CLOUD DATA PLATFORM"	PAG. 25
MIRRORING DI DATABASE	PAG. 26
 CONCLUSIONI	 PAG. 27
 AUTORI	 PAG. 28

COS'È MICROSOFT FABRIC

Microsoft Fabric è un **ambiente completo ed integrato per la gestione e l'analisi dei dati**. È una soluzione progettata per soddisfare al meglio le esigenze delle aziende che vogliono **sfruttare le moderne soluzioni offerte dal cloud, offrendo un'ampia gamma di funzionalità come ad esempio data management, business intelligence, advanced analytics e data governance**.

Microsoft Fabric non solo fornisce strumenti avanzati per la gestione e l'analisi dei dati, ma anche promuove attivamente la **collaborazione** tra le diverse figure professionali coinvolte nell'ambito dei dati. **Data analysts, data engineers e data scientists** spesso operano in silos separati, lavo-

rando su strumenti e piattaforme differenti e gestendo dati in modo isolato. Tuttavia, Microsoft Fabric favorisce un **approccio integrato**, consentendo a queste figure di lavorare insieme in un **ambiente condiviso**.

La gestione unificata dei dati e l'accesso a diversi sistemi consentono a ciascuna figura professionale di sfruttare le competenze specifiche degli altri membri del team. Ad esempio, i data analysts possono beneficiare dell'infrastruttura fornita dai data engineers per l'elaborazione e la preparazione dei dati, mentre i data scientists possono accedere direttamente ai dati gestiti e curati dagli altri membri del team per condurre analisi avanzate e modellizzazione predittiva.





L'**eliminazione dei silos di dati** e la promozione della collaborazione consentono ai team di lavorare in modo più efficiente e produttivo. Le competenze complementari delle varie figure professionali possono essere integrate in modo sinergico, portando a risultati più accurati e significativi nell'analisi e nell'utilizzo dei dati per supportare le decisioni aziendali.

Microsoft Fabric è **costruito su un'architettura Software as a Services (SaaS)** che rappresenta un balzo in avanti in termini di semplicità e accessibilità. Questa base SaaS unifica e semplifica l'esperienza utente, consentendo agli utenti di accedere ad una **suite completa di servizi di analisi da qualsiasi luogo e dispositivo**, senza la necessità di installare, configurare e gestire software complessi.

In un panorama tecnologico spesso caratterizzato da frammentazione e complessità, Microsoft Fabric si distingue per il suo approccio unificato. Grazie all'integrazione sinergica di componenti chiave **già leader di mercato** tra cui Power BI, Azure Synapse e Azure Data Factory, Fabric offre una **soluzione end-to-end che elimina la necessità di gestire ed integrare servizi disparati** da diversi fornitori.

Questo approccio semplifica notevolmente le operazioni di analisi, consentendo alle **aziende di concentrarsi primariamente sulle proprie esigenze di business** lasciando la complessità della gestione dell'infrastruttura tecnologica sottostante al fornitore dei servizi.



DI COSA SI COMPONE

ONELAKE

rappresenta il cuore pulsante dei servizi offerti da Fabric, offre infatti una soluzione scalabile ed affidabile per soddisfare le esigenze di **archiviazione dati** di qualsiasi dimensione aziendale. Operando come base di tutti i servizi, offre un unico ambiente integrato e facilita la collaborazione. Ne parleremo diffusamente a seguire.

DATA INTEGRATION

con Data Factory all'interno di Microsoft Fabric, sono disponibili funzionalità di copia rapida che consentono di **spostare i dati in modo efficiente sia attraverso flussi di dati che tramite pipeline di dati**.

I flussi di dati consentono di accedere ad oltre 300 trasformazioni direttamente dall'interfaccia di progettazione, semplificando la trasformazione dei

dati con una flessibilità senza precedenti, compresi algoritmi basati sull'intelligenza artificiale.

Le pipeline di dati, d'altra parte, permettono di orchestrare in modo predefinito flussi di lavoro di dati flessibili e adattabili alle esigenze specifiche dell'azienda.

DATA ENGINEERING

Microsoft offre una potente **piattaforma Spark**, garantendo un'esperienza di sviluppo che consente ai data engineer di eseguire trasformazioni su vasta scala e di rendere i dati più accessibili attraverso il concetto di lakehouse. Grazie all'integrazione di Microsoft Fabric Spark con Data Factory, è possibile pianificare e coordinare in modo efficiente sia i notebook che i processi Spark.

DATA WAREHOUSE

con **Azure Synapse** è possibile aumentare o diminuire la capacità di elaborazione e di archiviazione dei dati in modo indipendente. Questo significa che è **possibile aggiungere o rimuovere risorse di calcolo** senza dover necessariamente modificare la capacità di archiviazione e viceversa. Infine, viene utilizzato il formato di archiviazione dei dati **Delta Parquet in modo integrato e nativo** e ciò contribuisce, grazie a funzionalità proprie del formato in oggetto (tra le quali transazioni ACID e supporto per i metadati) a garantire affidabilità e coerenza dei dati memorizzati.

DATA SCIENCE

Fabric offre un **ambiente completo per la Data Science** grazie alla sua **integrazione nativa con Azure Machine Learning**, permettendo agli utenti di gestire workflow end-to-end con facilità e precisione: dall'esplorazione e preparazione dei dati, alla creazione di modelli predittivi, fino all'ottenimento di business insights. La piattaforma offre inoltre una pagina dedicata alla Data Science, dove gli utenti possono creare e gestire esperimenti di machine learning ed integrare notebook esistenti. Questa integrazione permette agli utenti di lavorare in modo più efficiente e di ottenere così risultati rapidi con il fine di supportare le decisioni aziendali.

REAL TIME ANALYTICS

I dati in streaming, raccolti da varie fonti come app, dispositivi IoT, interazioni umane e molto altro ancora, costituiscono attualmente la categoria di dati in più rapida crescita. Questi dati sono spesso semi-strutturati, presentati in formati come JSON o testo, e arrivano in grande quantità con schemi che possono variare nel tempo. Queste caratteristiche rendono difficile il lavoro per le piattaforme tradizionali di data warehousing. In tal senso la componente di Real Time Analytics, implementata tramite Azure Synapse, si conferma **tra le migliori soluzioni sul mercato nell'ambito dell'analisi dei dati in streaming**.

POWER BI

La piattaforma di business intelligence di punta rappresenta un insieme completo di strumenti, applicazioni ed interfacce che lavorano sinergicamente per trasformare dati eterogenei in analisi dettagliate, coerenti, visivamente coinvolgenti ed interattive. **Power BI offre la flessibilità di connettersi ad una vasta gamma di fonti dati**, che possono essere in vari formati, tra cui fogli di calcolo di **Excel** o anche **data warehouse ibridi**, quindi sia nel **cloud** che **on-premise**. Infine, è possibile creare **report personalizzati e dashboard interattive** per visualizzare e condividere le informazioni cruciali con i propri collaboratori o gli stakeholder desiderati.



APPROCCIO UNIFICATO E PUNTI DI FORZA

Un elemento chiave di Microsoft Fabric è il suo **approccio unificato** che consente agli utenti di **accedere in modo intuitivo alle funzionalità** di analisi di cui hanno bisogno, senza dover accedere e spostarsi tra diverse applicazioni o piattaforme.

Microsoft Fabric offre una serie di vantaggi significativi per le aziende, tra cui:

SEMPLIFICAZIONE

elimina la complessità associata alla gestione di una serie di servizi e piattaforme di analisi separati.

INTEGRAZIONE

fornisce un ambiente integrato in cui dati e servizi sono già nativamente connessi e interoperabili.

SCALABILITÀ

supporta **grandi moli di dati e workload analitici crescenti**, consentendo alle aziende di crescere e di adattarsi al progresso.

ACCESSIBILITÀ

offre **un'esperienza utente intuitiva e accessibile da qualsiasi luogo e dispositivo**, senza necessità di installazioni ad-hoc e consentendo agli utenti di accedere a dati, strumenti di analisi quando ne hanno bisogno.

RAPID PROTOTYPING

è una funzionalità che permette di sviluppare in tempi brevi dei prototipi che rappresentano una possibile soluzione al problema o al bisogno dell'utente. Questi **prototipi possono essere condivisi con gli stakeholder aziendali e i key user** per raccogliere feedback e validare le ipotesi di progettazione. Si tratta di uno strumento che consente dunque un miglioramento continuo e un'ottimizzazione del prodotto finale, garantendo che soddisfi appieno le esigenze e le aspettative degli utenti. Il rapid prototyping all'interno di Microsoft Fabric offre un vantaggio significativo alle aziende, consentendo loro di **ridurre i tempi di sviluppo, di raccogliere feedback tempestivi e di creare soluzioni analitiche altamente efficaci**.

DATA ACTIVATOR

una feature che permette di impostare degli **alert su determinati KPI** presenti nelle varie dashboard di Power BI. Questo consente agli utenti di monitorare costantemente i loro indicatori chiave di performance senza dover aprire manualmente i report con una certa frequenza per verificarne il valore. Nel momento in cui un KPI raggiunge un valore tale da attivare la regola di alert impostata, gli utenti possono ricevere una notifica immediata che informa che l'evento monitorato si è verificato. In alternativa, è possibile predisporre dei **flussi di Power Automate** che, al verificarsi del trigger dell'alert, scatenano una serie di azioni predefinite, consentendo una risposta immediata e automatizzata alle variazioni significative nei dati. **Data Activator offre agli utenti un controllo proattivo e tempestivo** sulle prestazioni aziendali, consentendo loro di reagire prontamente a cambiamenti critici nei dati.

COPILOT & AI

Copilot è un potente strumento alimentato dall'**intelligenza artificiale** parte dell'ambiente Microsoft Fabric. È progettato per semplificare e ottimizzare il flusso di lavoro della preparazione dei dati, supportando sia gli utenti tecnici che quelli di business. La sua caratteristica principale è la **generazione di codice intelligente**, automatizzando gran parte delle operazioni neces-

sarie per preparare i dati per l'analisi. Questo riduce il rischio di errori, garantendo dati accurati e coerenti. Inoltre, offre strumenti per la validazione e l'ottimizzazione del codice generato producendo quindi, in modo automatico, documentazione dettagliata per una maggiore trasparenza e tracciabilità nel processo di preparazione dei dati.

INTEGRAZIONE CON ECOSISTEMA DATAVERSE

altra possibilità offerta è quella di integrarsi perfettamente con l'ecosistema **Dataverse e, di conseguenza, di appoggiarsi su Dynamics 365**. **Dataverse offre un framework robusto per la gestione dei dati aziendali**, consentendo di modellare ed organizzare i dati in entità e relazioni che rispecchiano la struttura dell'azienda. **L'integrazione con Dynamics 365, la suite di applicazioni aziendali di Microsoft, amplia ulteriormente le capacità di Microsoft Fabric**. Gli utenti possono beneficiare dell'accesso diretto e dell'interoperabilità con le applicazioni di Dynamics 365, come CRM e ERP, per ottimizzare l'utilizzo dei dati aziendali e ottenere insights più approfonditi.

OMOGENEITA' NELL'ECOSISTEMA MICROSOFT 365

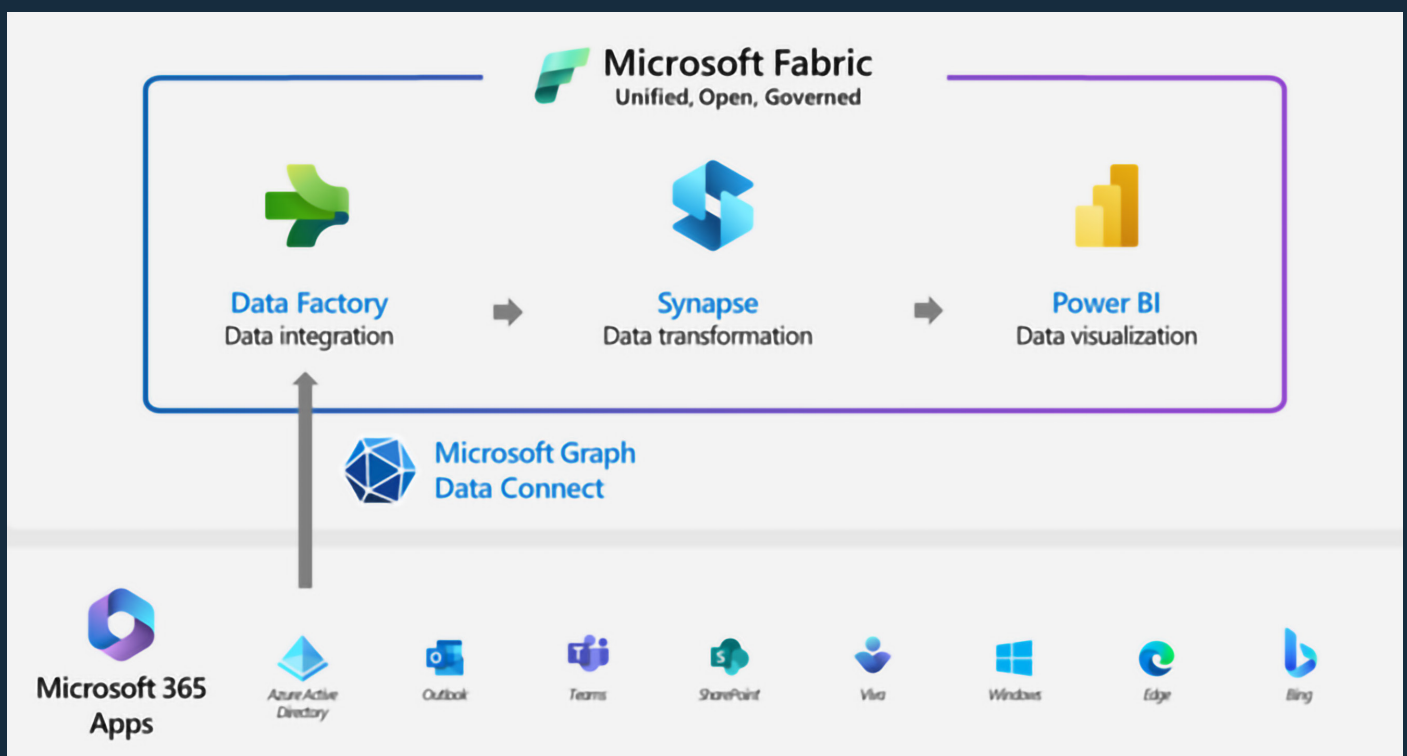
è possibile gestire i dati di Microsoft 365 all'interno di Fabric mediante **Microsoft Graph Data Connect** che ne rappresenta il punto d'accesso.



GESTIONE DEGLI UTENTI

Un'altra nota importante riguarda la **semplicità di gestione delle abilitazioni degli utenti**. Attualmente, per concedere accesso a tutti gli strumenti e servizi della suite BI, è necessario abilitare gli utenti su diversi strumenti separati, come Power BI, Analysis Services, e altri, creando così una serie di autorizzazioni e gruppi di sicurezza dispersi in Active Directory.

Questa frammentazione può comportare una gestione complessa e disorganizzata degli accessi, con la necessità di monitorare e aggiornare costantemente numerosi gruppi di sicurezza e autorizzazioni in modo da garantire che gli utenti abbiano accesso appropriato a tutte le risorse di cui hanno bisogno.



Tuttavia, con Microsoft Fabric, la gestione delle abilitazioni diventa notevolmente semplificata. Infatti, **basta abilitare un utente ad un workspace di Fabric per garantirgli automaticamente l'accesso a tutti gli strumenti e servizi inclusi nella suite.** Questo significa che non è più necessario gestire una serie di autorizzazioni e gruppi di sicurezza separati su diversi strumenti, riducendo così la complessità e semplificando la gestione degli accessi per gli amministratori di sistema.

Potremmo affermare, in ultima analisi, che Microsoft Fabric rappresenta un'evoluzione significativa nel campo dell'analisi aziendale. Con la sua base SaaS, componenti chiave e approccio unificato, Microsoft Fabric si pone come leader nel settore dell'analisi dei dati, consentendo alle aziende di sbloccare il pieno potenziale dei propri dati per ottenere vantaggi competitivi.



POTENZIALITÀ DI ONELAKE

OneLake rappresenta una soluzione progettata per superare le sfide comuni legate alla gestione dei dati. Nell'attuale panorama aziendale, infatti, la mole di dati generata ed utilizzata è in costante aumento e di conseguenza ciò può rappresentare sia un'opportunità che una sfida, poiché le organizzazioni devono trovare modi efficaci per gestire, sfruttare e proteggere questo prezioso asset digitale.

Spesso, infatti, ci si trova di fronte a dati di diverso tipo, complicati processi di integrazione e complessi requisiti

di sicurezza e di conformità. OneLake si propone di affrontare queste sfide fornendo un'**unica piattaforma centralizzata per l'archiviazione e la gestione dei dati aziendali**. Tra le caratteristiche distintive di OneLake vi è la sua natura **software-as-a-service (SaaS)**, che ne semplifica notevolmente l'implementazione e l'uso da parte degli utenti. Con OneLake sia gli utenti di business che i tecnici specializzati possono **accedere ai dati e alle risorse senza dover affrontare complessi e tecnici processi di configurazione**.





UN'UNICA COPIA DEI DATI

OneLake introduce un nuovo paradigma nell'archiviazione e nell'accesso dati, consentendo l'utilizzo di una **singola copia dei dati** per tutti i motori di calcolo. Questa caratteristica rivoluzionaria elimina la duplicazione dei dati e garantisce coerenza ed integrità attraverso l'intero ecosistema aziendale. Con OneLake, i dati sono memorizzati in

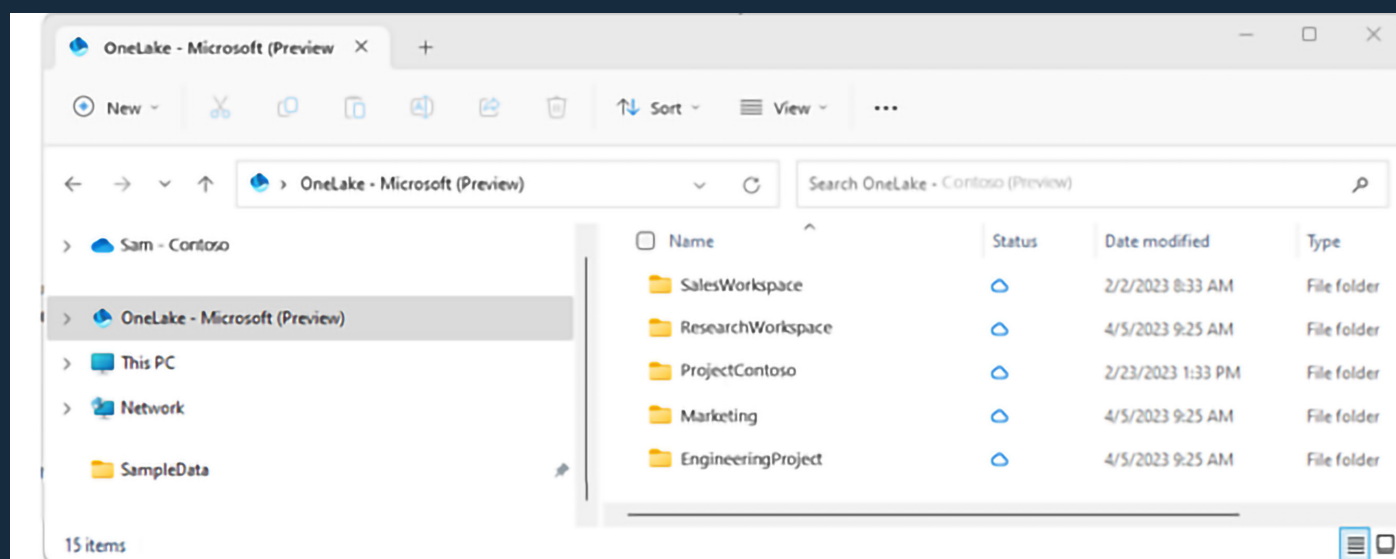
un **unico repository centrale**, eliminando la necessità di gestire copie multiple dei dati su diversi motori di calcolo. Questo non solo semplifica notevolmente la gestione dei dati, ma **riduce anche il rischio di errori** e di **eventuali inconsistenze** causate dalla replicazione dei dati.



L'EQUIVALENTE DI ONEDRIVE PER I DATI

OneLake offre un'esperienza simile a quella di OneDrive per i dati, fornendo dunque un modo intuitivo ed accessibile per esplorare e gestire i dati all'interno dell'organizzazione. **Proprio come OneDrive**, infatti, è possibile **accedere e navigare** facil-

mente nei dati di OneLake utilizzando **OneLake File Explorer** per Windows. Questo strumento consente agli utenti di **esplorare tutte le aree di lavoro, caricare, scaricare o modificare file** con la stessa facilità con cui si lavora su Office.



In aggiunta, **OneLake File Explorer semplifica** sensibilmente l'**utilizzo dei data lake**, offrendo la possibilità anche agli utenti meno tecnici di collaborare e di favorire lo scambio di informazioni all'interno dell'organizzazione o anche semplicemente di visualizzare i file, le cartelle e la gerarchia dei dati in modo chiaro.



FORMATO STANDARD DI MEMORIZZAZIONE

Come precedentemente accennato, OneLake si distingue per l'adozione del formato di archiviazione aperto **Delta Parquet**, riconosciuto come uno standard nel settore dell'archiviazione dati. Questo formato garantisce un ambiente in cui **tutte le informazioni sono conservate in modo uniforme**, facilitando l'interoperabilità e semplificando le analisi.

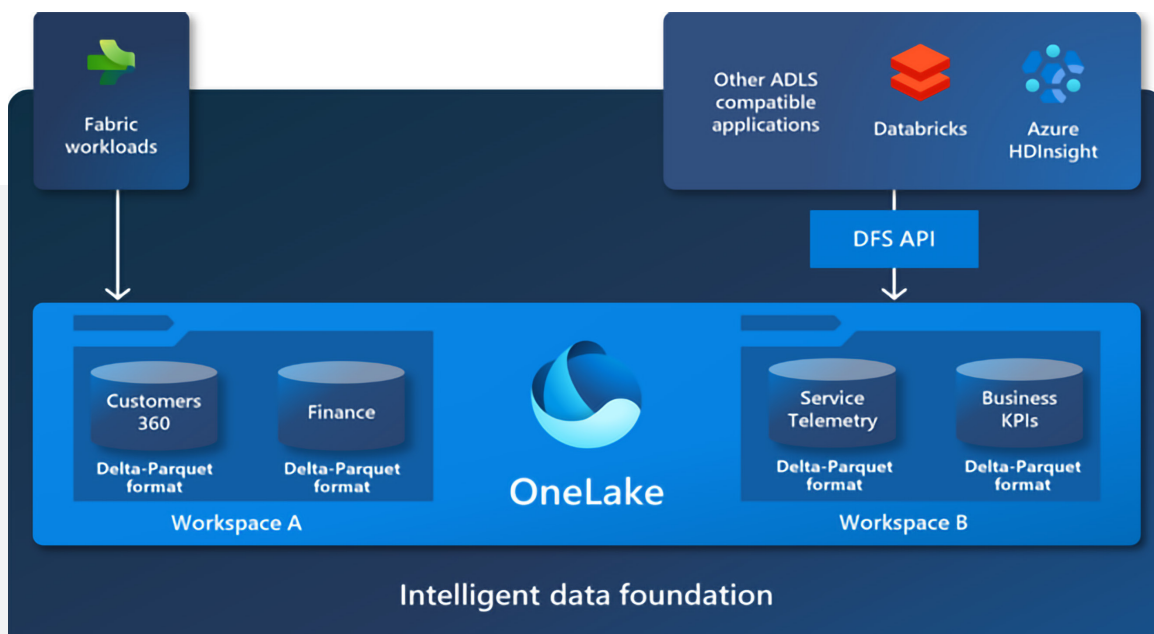
OneLake fa uso del formato Delta Parquet per memorizzare i dati e ciò offre una serie di vantaggi, tra cui la **compressione efficiente dei dati, la partitioning intelligente e ottimizzazioni specifiche per le query**, garantendo prestazioni ottimali durante l'analisi dei dati.

Grazie a questa scelta di architettura, i dati su OneLake sono facilmente **accessibili da tutti i motori di calcolo**, senza necessità di complesse procedure di trasferimento. Ciò non solo semplifica i flussi di lavoro, ma favorisce anche una

collaborazione più efficiente tra i diversi reparti aziendali.

Inoltre, OneLake è basato su **Azure Data Lake Storage (ADLS) Gen2**, una potente infrastruttura di storage che consente di archiviare una vasta gamma di tipologie di file, sia strutturati che no.

La completa compatibilità con le **API** (Application Programming Interfaces) e gli SDK (Software Development Kits) utilizzati con ADLS Gen2 si estende anche a OneLake. Ciò significa che gli sviluppatori possono usufruire delle **stesse librerie di codice e strumenti già in uso con ADLS Gen2**, garantendo un'esperienza fluida e senza interruzioni. Inoltre, questa compatibilità assicura che le applicazioni esistenti che fanno uso di queste tecnologie, inclusi servizi come **Azure Databricks**, possano essere integrate senza problemi con OneLake, consentendo agli utenti di gestire i dati in modo efficace e consolidato.



ARCHITETTURA MEDALLION

Tra i vantaggi chiave di OneLake vi è sicuramente l'adozione dell'architettura Medallion, un **modello di progettazione che organizza logicamente e su più livelli i dati all'interno di un lakehouse**. Questo approccio a più livelli consente di gestire i dati in modo efficiente, garantendo la qualità e l'integrità dei dati stessi. I tre livelli principali dell'architettura Medallion sono: **Bronzo, Argento e Oro**.

- **Livello Bronzo:** questo è il livello dei **dati grezzi**. Qui i dati, strutturati e no, sono ottenuti tramite ingestion, senza alcuna modifica o trasformazione. Strumenti come Azure Data Factory e Azure Data Lake Storage Gen2 vengono utilizzati per caricare i dati nel lakehouse. Il livello Bronzo è **essenziale per conservare una copia completa e non alterata** dei dati originali.

- **Livello Argento:** nel livello Argento, i dati vengono **convalidati e raffinati**. Qui avvengono operazioni come la **pulizia dei dati, l'unione di fonti**

diverse e l'applicazione di **regole di validazione**. I data engineer utilizzano strumenti come Notebook e codice in PySpark, Scala, R o SQL per lavorare con i dati.

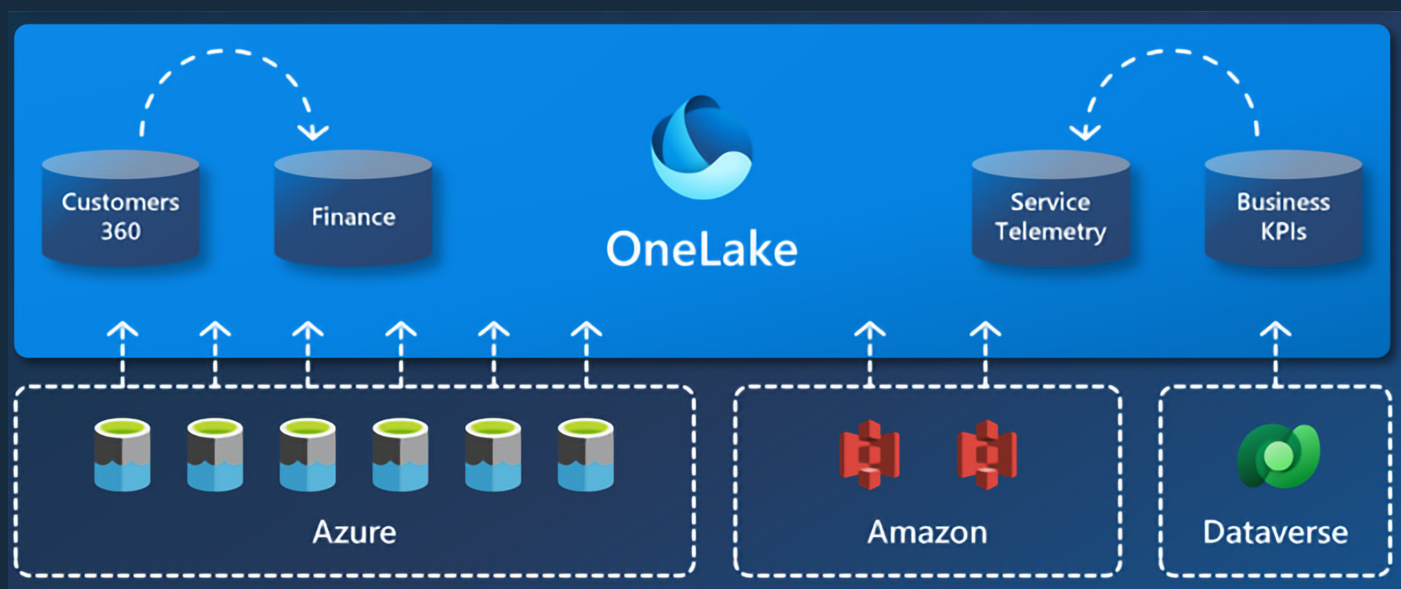
- **Livello Oro:** il livello Oro è il **cuore del processo di arricchimento dei dati**. Qui i dati vengono **arricchiti con informazioni aggiuntive**, come attributi calcolati, aggregazioni e metadati. È il livello da cui i data team, come quelli di **analisi** e di **Machine Learning**, attingono per ottenere e lavorare con dati di alta qualità.



SHORTCUTS

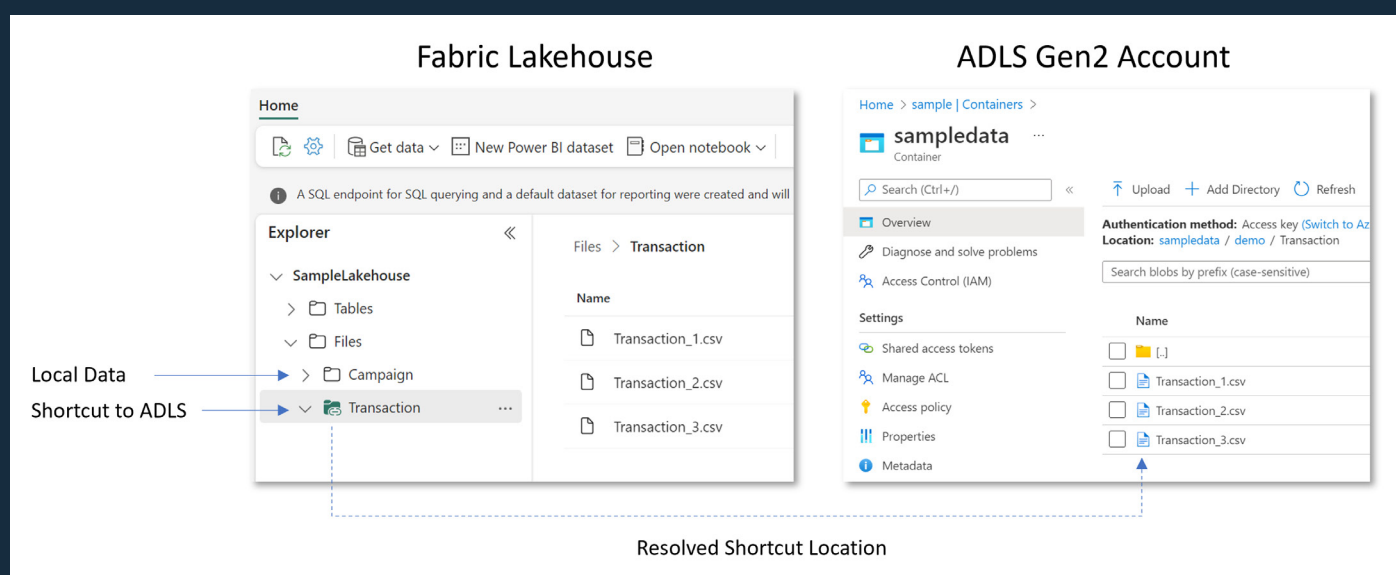
Tra gli altri punti di forza di OneLake, gli shortcuts rappresentano uno strumento essenziale per unificare i dati dell'azienda, consentendo **la creazione di un singolo data lake virtuale che attraversa diversi domini,**

cloud e account. Tutti i servizi ed i motori analitici di Fabric possono facilmente accedere alle fonti dati esistenti, come ADLS, S3 di Amazon Web Services (AWS) e Dataverse, grazie ad uno spazio unificato.



Con OneLake che **gestisce tutte le autorizzazioni e le credenziali**, non è necessario configurare separatamente i servizi di Fabric per connettersi a ciascuna fonte dati. L'uso degli shortcuts consente inoltre di eliminare

le copie superflue dei dati e di ridurre la latenza dei processi legata alle operazioni di staging.



Ma cosa sono esattamente questi shortcuts? Sono semplici **cartelle che in OneLake possono essere utilizzate da qualsiasi servizio che abbia accesso al sistema**. Nell'esempio in figura sopra, è mostrato uno

shortcut ad ADLS. Se si elimina uno shortcut, il percorso di destinazione rimane intatto, ma se si sposta, rinomina o elimina il percorso di destinazione, lo shortcut potrebbe non funzionare correttamente.

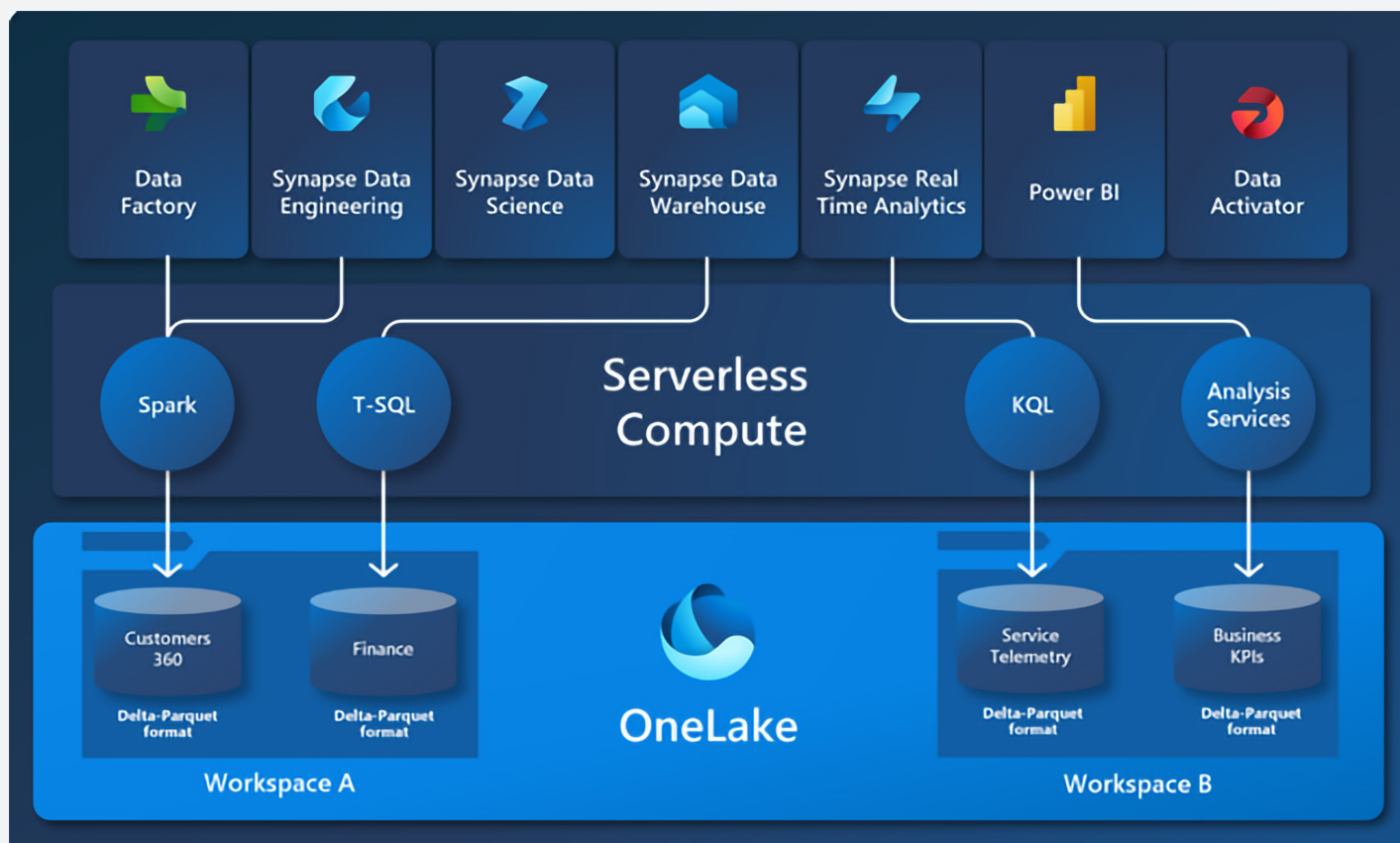


OTTIMIZZAZIONE A PIÙ MOTORI ANALITICI

Sebbene le applicazioni possano prevedere una separazione dello storage e del computing, spesso i dati sono ottimizzati per un singolo motore, il che rende difficile riutilizzare gli stessi dati per più applicazioni. Con Fabric, i **diversi motori analitici (T-SQL, Spark, Analysis Services, ecc.) archiviano i dati nel formato Delta Parquet** al fine di consentire di utilizzare gli stessi dati su più motori.

Non c'è più bisogno di copiare i dati solo per utilizzarli con un altro motore. Si è sempre in grado di scegliere il motore mi-

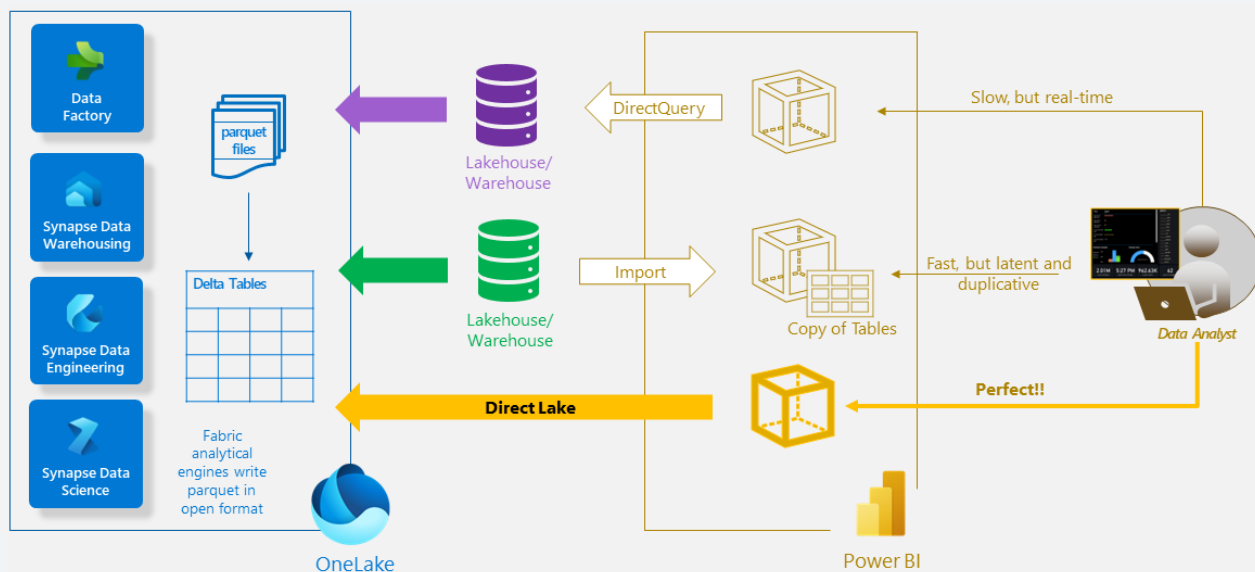
gliore per il compito che si sta cercando di svolgere. Ad esempio, si consideri un team di data engineer che costruiscono un data warehouse completamente transazionale: possono utilizzare il motore T-SQL per creare tabelle, trasformare dati e caricare i dati nelle tabelle. Se un data scientist vuole utilizzare questi dati, non è più necessario utilizzare un driver speciale Spark o SQL: OneLake memorizza tutti i dati nel formato Delta Parquet e i data scientist possono dunque utilizzare tutta la potenza del motore Spark e delle sue librerie open source direttamente sui dati.



MODALITÀ DIRECT LAKE

La modalità Direct Lake rappresenta una svolta nel modello semantico per l'analisi di grandi volumi di dati all'interno di Power BI. Questa innovazione si basa sul **caricamento diretto di file in formato parquet direttamente su OneLake**, senza la necessità di eseguire query su un endpoint Lakehouse o Warehouse, **né di importa-**

re o duplicare i dati all'interno di un modello Power BI. Direct Lake costituisce un'opzione veloce per caricare i dati rendendo di fatto immediata l'analisi via motore di Power BI. Di seguito, viene presentato un diagramma che confronta le modalità precedenti di importazione classica e DirectQuery, con la **nuova modalità Direct Lake**.



In modalità **DirectQuery**, il motore di **Power BI** esegue una query diretta sui dati all'origine, il che può essere lento, ma evita la necessità di copiare i dati come avviene nella import mode. Ogni modifica apportata all'origine dati viene immediatamente riflessa nei risultati della query.

D'altra parte, con l'import mode, le prestazioni possono essere migliori perché i dati vengono memorizzati nella cache e ottimizzati per le query di report DAX e MDX senza dover tradurre ed inviare query SQL o di altri tipi all'origine dati. Tuttavia, il motore di Power BI deve prima copiare tutti i nuovi dati nel modello durante l'aggiornamento. Le modifiche all'origine dati vengono quindi prese in considerazione solo con l'aggiornamento successivo del modello.

La modalità **Direct Lake** elimina il requisito di importazione caricando i dati direttamente su OneLake.

A differenza di DirectQuery, non vi è alcuna conversione da DAX o MDX ad altri linguaggi di query né l'esecuzione di query su altri sistemi di database, garantendo prestazioni simili alla import mode.

Poiché non vi è un processo di importazione esplicito, è possibile incorporare le modifiche all'origine dati man mano che si verificano, combinando i vantaggi di DirectQuery e import mode e mitigando i loro svantaggi. La modalità Direct Lake può essere l'opzione ideale per l'analisi di modelli di grandi dimensioni e con aggiornamenti frequenti all'origine dati.

CONVIENE PASSARE A FABRIC?

La situazione attuale della tecnologia Synapse (Dedicated Pool) ha raggiunto risultati da record, ma non bisogna dimenticare che **ogni prodotto ha un ciclo di vita**, con un inizio, una crescita, una maturazione e, prima o dopo, anche una fine.

Microsoft ha dichiarato esplicitamente che Synapse non è deprecato e certamente continuerà a supportarlo con aggiornamenti di sicurezza, ma il motore ha già diversi anni sulle spalle ed è probabile che non verranno aggiunte nuove funzionalità in futuro.

Inoltre, negli ultimi anni Microsoft è rimasta leggermente indietro rispetto alla concorrenza nei **moderni cloud data warehouse** e ora rispondendo con Fabric ha ripareggiato grandi competitor come Snowflake, BigQuery e Databricks.

Detto ciò, la scelta se migrare o meno una architettura esistente è un tema di grande attualità e la risposta probabilmente è da cercare nel rapporto costo/beneficio che se ne può trarre.

Possono esserci strade diverse da percorrere in funzione della situazione in cui ci si trova.





SCENARIO “GREEN-FIELD”

In uno scenario green-field, in cui si deve partire da zero a costruire una data platform, sicuramente conviene valutare come prima ipotesi lo sviluppo su una piattaforma moderna e su cui Microsoft ha investito tanto e continuerà a farlo in futuro!

Chi ha competenze su Data Factory, Synapse, Analysis Services o altri strumenti simili, può sfruttare tutta la competenza acquisita su questi strumenti anche per lavorare su Fabric.

Sebbene vi siano alcune differenze a livello visuale, tutte le principali logiche rimangono inalterate. Questo consente un passaggio molto più veloce al nuovo prodotto, poiché il gap di conoscenze da colmare è limitato.





SCENARIO “FABRIC READY”

Se si è già nella fortunata posizione di aver creato un Lakehouse utilizzando ad esempio Azure Databricks o Synapse Spark, allora si dispone di una base di partenza perfetta. Con le funzionalità di Shortcut citate in precedenza, il **Delta Lake esistente può essere utilizzato direttamente e senza interruzioni**. Esiste infatti la possibilità di rendere i dati direttamente disponibili ai **numerosi motori di calcolo in Fabric** (data science/ IoT/Power BI, ecc.), senza dover copiare i dati.

Ciò offre un vantaggio immediato: Power BI può utilizzare questi dati in modalità Direct Lake. Ciò significa che **Power BI può accedere direttamente ai file Delta senza dover importare i dati** tramite un modello semantico (es: dataset).

Come detto grazie a questo si otterranno prestazioni della modalità “import” combinate con la freschezza del DirectQuery!





SCENARIO “JOURNEY TO CLOUD DATA PLATFORM”

Microsoft è molto sensibile a questa transizione e per agevolare i propri clienti (ma non solo) ha messo a disposizione un **servizio dedicato di “Mirroring”** proprio per consentire ai propri clienti di beneficiare il prima possibile dei nuovi servizi offerti da Fabric.

Questa modalità consente di **dare continuità di servizio preservando gli investimenti fatti** sui propri database/data warehouse on-premise e contemporaneamente di affrontare in modo graduale la costruzione di un moderno lakehouse su cui far crescere la nuova cloud data platform.



MIRRORING DI DATABASE

Questa funzionalità consente a OneLake in **near real-time di accedere a una ampia gamma di database senza dover passare tra client di database diversi**. Basta fornire le specifiche di connessione ed il database è istantaneamente disponibile.

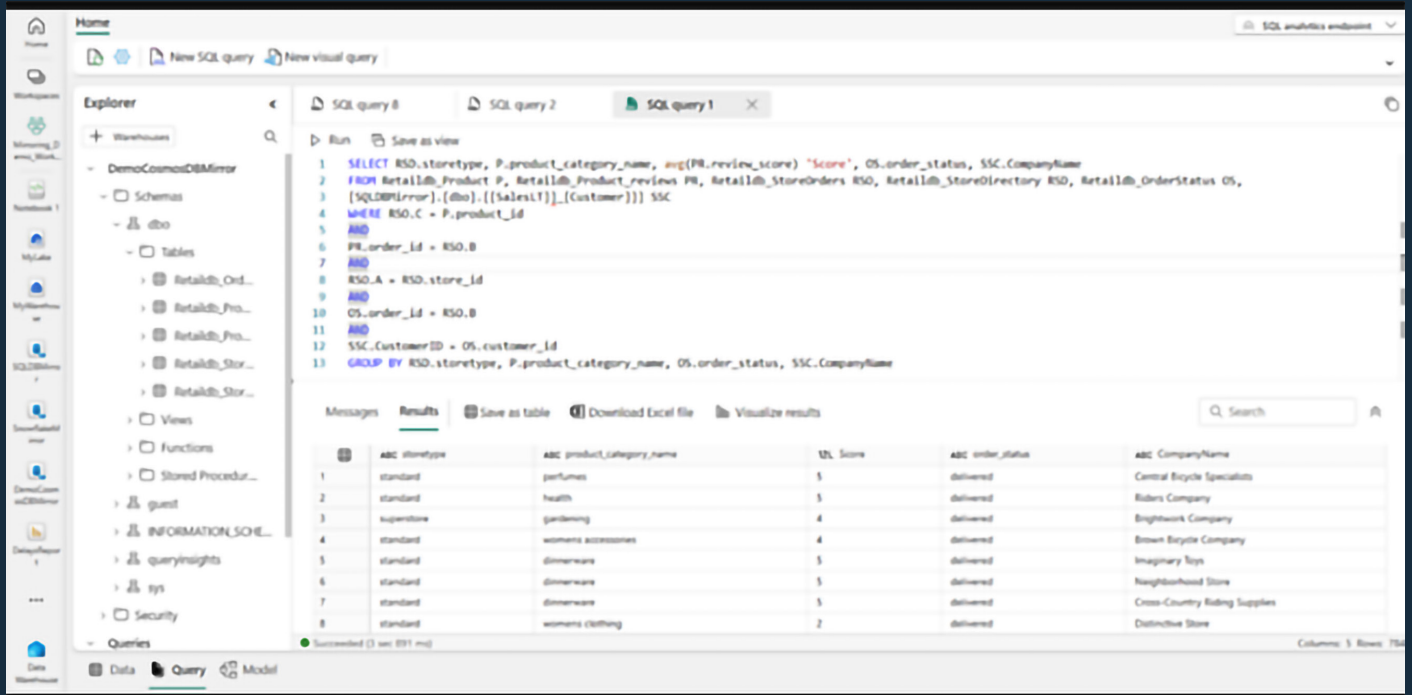
Non è richiesto alcun processo ETL particolarmente complesso per la copia dei dati. Viene creato uno snapshot iniziale, dopo il quale i dati vengono mantenuti sincronizzati, sia che si tratti della creazione di una nuova tabella, sia dell'inserimento/aggiornamento/cancellazione di dati.

Il processo di copia fa uso della **tecnologia Change Data Capture (CDC) del database, trasformando i dati in tabelle "Delta" appropriate e caricandole su OneLake**. Una sistema di caching intelligente aiuta inoltre a deter-

minare quando la sorgente viene alterata, garantendo che non venga attivata inutilmente alcuna computazione, prima di copiare i dati.

Controlli granulari consentono inoltre di configurare cosa viene portato in fase di mirroring su Fabric ed è disponibile anche un monitoraggio dettagliato per ottenere approfondimenti sulle operazioni di mirroring e sull'ultimo refresh della copia dati su OneLake. Da qui in poi, i dati sono pronti per essere utilizzati immediatamente da un qualsiasi workload di Fabric.

Ciascun database portato su Fabric via mirroring viene fornito di un'interfaccia per eseguire query ed analisi dei dati, inclusi i metadati supportati dal formato Delta, che consentono di comprenderne la struttura e le caratteristiche.



Come detto nei precedenti paragrafi, tutti i dati che vengono portati su Fabric, che siano questi aggiunti come Shortcuts in un lakehouse o tramite ingestion attraverso Data Factory o Spark, sono tutti già in formato Delta. In aggiunta, grazie al mirroring, i dati provenienti da database differenti, come Azure Cosmos DB, **Azure SQL DB**, Snowflake, MongoDB, ecc... possono essere messi direttamente in join con tutte le altre fonti semplificando le interrogazioni.

E se il Data Base è on premise?

Uno scenario in cui probabilmente c'è un'opportunità da cogliere è quella del passaggio al cloud partendo dall'on-premise. Infatti, **anche in uno scenario SQL Server on-premise, è possibile portare con uno sforzo ridotto le strutture e i dati presenti verso un database Azure SQL**. In questo modo sarebbe possibile abilitare le funzionalità di mirroring appena descritte, riducendo parte delle complessità insite a questa trasformazione e beneficiando da subito dei moderni servizi offerti da Fabric.

CONCLUSIONI

Microsoft Fabric e OneLake rappresentano il centro di un cambiamento radicale nell'approccio all'analisi dati. In un mondo sempre più complesso ed interconnesso, queste soluzioni puntano ad unire tecnologie all'avanguardia, semplicità d'uso e potenza computazionale senza precedenti.

Questo principio di semplicità ed ampia disponibilità di tecnologia comprende anche il pricing della soluzione. Il **modello di pricing di Fabric è infatti molto semplice ed esplicito**: si sceglie il taglio di potenza necessario e tutti i componenti sono compresi all'interno del taglio di potenza scelto (a parte la quota di storage del Data Lake che solitamente ha una incidenza molto bassa).

È possibile avere evidenza dei prezzi al link: <https://azure.microsoft.com/en-us/pricing/>

Questo cambia le regole del gioco rispetto alle configurazioni attuali, dove è necessario selezionare un taglio di potenza per Power BI, uno per Analysis Services, uno per il motore di database e affrontare la complessità della stima dei costi di elaborazione dati, soprattutto considerando il pricing model non semplice di Data Factory che tiene conto di variabili come la quantità di dati movimentati, i costi di banda, il numero di esecuzioni mensili e altre metriche.

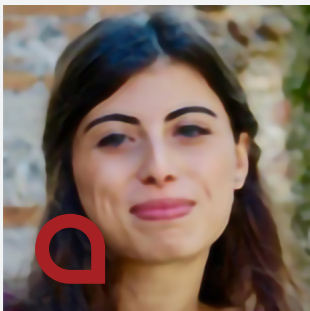
Se questo white paper ha destato il tuo interesse, saremo felici di approfondire ogni eventuale domanda e condividere le nostre esperienze su clienti e risultati concreti ottenuti grazie a questa tecnologia. È inoltre possibile gratuitamente sperimentare e toccare con mano la tecnologia MS Fabric grazie ad una **prova gratuita di 60gg senza limitazioni** collegandosi all'indirizzo <https://app.fabric.microsoft.com/>

GLI AUTORI



Giulio Aragiusto
Data Platform Practice Director

Giulio ha oltre 19 anni di esperienza nella consulenza ed è Data Platform Director per Horsa Insight. Specializzato in processi di Data Digital Transformation e metodologie di Data Management lavora come Advisor nella definizione di Enterprise Data Platforms.



Giovanna Critti
Data Engineer

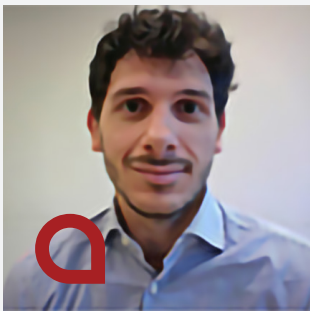
Da oltre tre anni, Giovanna lavora in Horsa come Data Engineer. Si occupa di progetti in ambito Microsoft contribuendo alla definizione e implementazione di architetture dati per soluzioni di Business Intelligence e modern data platform.

GLI AUTORI



Michael Messori
Senior Data Engineer

Michael si occupa da diversi anni di progetti di Data & Analytics su differenti applicativi e su clienti di differenti business. Negli ultimi anni si è focalizzato sulle potenzialità offerte dalla suite Azure di Microsoft sempre nell'ottica di offrire soluzioni che portino valore aggiunto alle sfide di business dei nostri clienti.



Salvatore D'Izzia
Data & Analytics Junior Consultant

Salvatore è un giovane matematico entrato nel mondo dell'analisi dati come consulente junior nell'ambito di soluzioni Microsoft per le aziende. Specializzato in probabilità durante gli studi, possiede competenze pratiche nell'analisi statistica. La sua mentalità orientata ai dettagli e la curiosità intellettuale lo portano ad un costante apprendimento. Oltre al lavoro, è appassionato di scienza e si interessa attivamente al progresso scientifico.

Horsa[®] insight

Horsa Insight è la Business Unit di Horsa che si occupa di Advanced & Predictive Analytics. Un consulente di fiducia che accompagna e guida le aziende nel proprio percorso di data strategy.