

LA TECHNOLOGY TRANSFER PRESENTA

MIKE FERGUSON

Linee guide pratiche per implementare un Data Mesh

Data Catalog, Data Fabric, Data Product, Data Marketplace

ONLINE LIVE STREAMING

26-27 APRILE 2023



info@technologytransfer.it
www.technologytransfer.it

DESCRIZIONE

La maggior parte delle aziende oggi archivia dati ed esegue applicazioni in un ambiente multi-Cloud ibrido. I sistemi analitici tendono a essere centralizzati e isolati come Data Warehouse e Data Mart per BI, Hadoop o Data Lake memorizzati su Cloud per Data Science e sistemi analitici di streaming stand-alone per l'analisi in tempo reale.

Questi sistemi centralizzati si affidano a Data Engineer e Data Scientist che lavorano all'interno di ciascun silos per acquisire dati da molte fonti diverse, pulirli e integrarli per l'utilizzo in uno specifico sistema analitico o in modelli di Machine Learning.

Ci sono molti problemi con questo approccio centralizzato e in silos, tra cui più strumenti per preparare e integrare i dati, reinvenzione delle pipeline di integrazione dei dati in ogni silos e ingegneria dei dati centralizzata con scarsa comprensione dei dati di origine che non permettono di tenere il passo con le richieste del Business che richiedono sempre nuovi dati. Anche i Master Data non sono ben gestiti.

Per affrontare questi problemi è emerso un nuovo approccio che tenta di accelerare la creazione di dati da utilizzare in più workload analitici, l'approccio è il Data Mesh. Data Mesh è un approccio decentralizzato orientato al dominio del business, alla data ownership e al data engineering per creare una rete di prodotti di dati riutilizzabili che possono essere creati una sola volta e condivisi per servire su più sistemi analitici e workload. Un Data Mesh può essere implementato in diversi modi. Questi includono l'utilizzo di uno o più account di archiviazione Cloud su Cloud Storage, su un Data Lake organizzato, su un Lakehouse, su un Data Cloud di dati, utilizzando Kafka o utilizzando la virtualizzazione dei dati e i prodotti dati possono quindi essere consumati in altre pipeline per essere usate nello Streaming Analytics, Business Intelligence, Data Science e altri workload analitici.

Questo seminario esamina in dettaglio il Data Mesh evidenziando i suoi punti di forza e di debolezza. Quale è la migliore Architettura per implementarlo? Come coordinare più team orientati al dominio e usare una comune infrastruttura software dati, come Data Fabric, per creare prodotti dati di qualità, conformi e riusabili in un Data Mesh e come usare i Data Marketplace per condividere i prodotti dati? L'obiettivo è abbreviare il time to value, garantendo al contempo che i dati siano gestiti correttamente in un ambiente decentralizzato.

Questo corso si sofferma inoltre sulle implicazioni organizzative del Data Mesh e su come creare prodotti dati condivisibili Master Data Management e per l'uso nell'analisi multi-dimensionale su un Data Warehouse, Data Science, Graph Analysis e real time streaming Analytics per creare valore di business. Le tecnologie discusse includono cataloghi di dati, Data Fabric per lo sviluppo collaborativo di pipeline di integrazione dati per creare prodotti dati, DataOps per velocizzare il processo, automazione della Data Orchestration, marketplace di dati e piattaforme di governance dei dati.

PARTECIPANTI

- Analisti dati di business
- Architetti dati
- Chief Data Officer
- Professionisti di Master Data Management
- Data Scientist
- Sviluppatori IT ETL
- Professionisti di Data Governance

Si assume che i partecipanti abbiano una conoscenza base dei principi del Data Management e di Architettura dati.

Cosa Imparerete

- Punti di forza e di debolezza delle Architetture dati centralizzate usate nell'Analitica
- I problemi causati nei sistemi analitici esistenti dagli ambienti multi-Cloud ibridi
- Cosa è un Data Mesh e come differisce da un Data Lake e da un Data Lakehouse?
- Quali benefici offre il Data Mesh e quali sono le opzioni di implementazione?
- Quali sono i principi, i requisiti e le sfide dell'implementazione di questi approcci?
- Come organizzarsi per creare prodotti dati in un ambiente decentralizzato senza creare caos?
- L'importanza di un Catalogo dati per capire quali dati sono disponibili come un servizio
- Come i glossari di business possono aiutare ad assicurarci che i prodotti dati siano capiti e semanticamente collegati
- Un modello operativo per una efficace Data Governance Federata
- Quale infrastruttura software è richiesta per operare e governare un Data Mesh di prodotti dati da essere usati in un Data Lake, in un Data Lakehouse o in un Data Warehouse?
- Cosa è il software Data Fabric, come si integra con i cataloghi dati e come si connette al vostro patrimonio dati?
- Una metodologia di implementazione per produrre prodotti dati affidabili e riusabili
- Sviluppo collaborativo orientato al dominio di pipeline distribuite DataOps per creare prodotti dati
- Come un Catalogo dati e software di automazione possono essere usati per generare pipeline di DataOps
- A gestire la qualità dei dati, privacy, sicurezza dell'accesso, versioning e il ciclo di vita dei prodotti dati
- A mettere a disposizione prodotti dati collegati semanticamente in un mercato dati dove altri possono consumarli e usarli
- A consumare prodotti dati in un sistema MD
- A consumare e assemblare prodotti dati in sistemi analitici multipli come Data Warehouse, Lakehouse e Graph database per ridurre il time to value

PROGRAMMA

1. Cosa è un Data Mesh, un Data Lake e un Lakehouse? Perché usarli?

Questa sessione esamina le sfide che devono affrontare le aziende che cercano di diventare data-driven e i punti di forza e di debolezza delle attuali Architetture dati centralizzate utilizzate nell'Analitica. Introduce quindi Data Lakes, Data Lakehouse e Data Mesh come potenziali modi per affrontare i problemi attuali. Esplora i pro e i contro di ciascuno, spiega come funzionano e come consentono la creazione di prodotti dati affidabili e riutilizzabili in un Data Mesh usando diverse opzioni architetturali come account di Cloud Storage, Data Lakes, piattaforme di dati Cloud eccetera da utilizzare in diversi workload analitici. Si chiede anche se combinare questi approcci di architetture multiple sia vantaggioso o meno.

- Complessità dei dati in un ambiente ibrido, multi-Cloud
- La crescita nelle nuove sorgenti dati
- Architetture dati centralizzate in uso negli attuali sistemi analitici - Data Warehouse, Data Lakes
- Pro e contro del Data Lake
- La fusione del Data Warehouse e del Data Lake
- Il passaggio dalla sola Data Science al Data Lake multiuso
- Cosa è un Data Lakehouse?
- Come lavora un Data Lakehouse?
- Pro e contro del Data Lakehouse
- Cosa è un Data Mesh?
- Principi del Data Mesh
- Come lavora il Data Mesh decentralizzato?
- Cosa è un prodotto dati?
- Quali tipi di prodotti dati si possono costruire?
- Sviluppo decentralizzato dei prodotti dati
- Pro e contro del Data Mesh
- Quali sono le sfide con questo approccio decentralizzato?
- Il software di Data Management è pronto per il Data Mesh?
- Come il Data Mesh impatterà sulla vostra organiz-

zazione IT e sulla cultura dati?

- E' possibile una Data Governance federata?
- Implementare Data Mesh su Data Cloud o su Kafka o su Cloud Storage Lakehouse o Data Virtualization
- Requisiti di implementazione per creare prodotti dati
 - o Modello operativo federato
 - o Vocabolario di business comune
 - o Produttori e consumatori di dati
 - o Indipendenza dell'Architettura
 - o Una piattaforma dati unificata per costruire qualsiasi pipeline per processare qualsiasi dato
 - o DataOps: sviluppo basato su componenti di pipeline
 - o Esecuzione di pipeline distribuita
 - o Prodotti dati riusabili e semanticamente collegati
 - o Governance di un panorama dati distribuito
- Principali tecnologie: Data Fabric, Data Catalogs, Data Classifiers, Data Marketplace, Data Warehouse Automation tools
- Offerta dei vendors nel mercato: Alation, AWS, BigID, Cambridge Semantics, Collibra, Dremio, Global IDs, Google, IBM, Informatica, Microsoft, Oracle, Qlik, Talend, SAP, SAS, Software AG, Starbust Data, StreamSets, IDERA WhereScape, Truist Zoloni

2. Metodologie per creare Prodotti Dati

Questa sessione esamina come produrre prodotti di dati riutilizzabili e pronti per il business per l'utilizzo da parte dei consumatori di dati in molteplici casi d'uso di Analitica per generare valore di business. Inoltre, esamina come i prodotti Master Data possono essere prodotti anche per l'utilizzo nel Master Data Management.

- Creare un Program Office
- Sviluppo decentralizzato di prodotti dati in un Data Mesh usando un Data Lake su Cloud Storage, un

Lakehouse, Data Cloud o Data Virtualization

- Il caso speciale e critico del Master Data
- Una metodologia pratica step-by-step per costruire prodotti dati riusabili
- Come i dati strutturati, semistrutturati e non strutturati impattano sulla metodologia
- Applicare pratiche di sviluppo DataOps allo sviluppo di prodotti dati?

3. Usare un Glossario di Business per definire Prodotti Dati

Questa sessione esamina come creare nomi di dati e definizioni comuni per i prodotti dati in un glossario di business in modo che i consumatori di dati possano comprendere il significato dei dati prodotti e disponibili in un Data Mesh. Esamina anche come i glossari di business sono diventati parte di un Catalogo di dati.

- Perché un vocabolario comune è molto rilevante?
- Cataloghi dati e il glossario di business
- Il mercato dei Cataloghi Dati: Alation, Amazon Glue, Cambridge Semantics ANZO Data Catalog, Collibra Catalog, data.world, Denodo Data Catalog, Google Data Catalog, Oracle, Hitachi Vantara Lumada, IBM Watson Knowledge Catalog, Informatica Axon and EDC, Microsoft Purview Data Catalog, Qlik Catalog, SAP, Truist Zaloni Data Platform
- Ruolo, responsabilità e processi necessari per gestire un glossario di business
- Avviare un glossario di business con un modello concettuale di dati
- Definire i prodotti dati usando termini del glossario
- Usare un catalogo e un glossario per assicurarsi che i prodotti dati siano semanticamente collegati

4. Standardizzare sviluppo e Operation in un Data Mesh, Data Lake o Lakehouse

Questa sessione esamina come standardizzare l'impostazione in ciascun dominio aziendale per ottimizzare lo sviluppo di prodotti dati in un Data Mesh.

- L'importanza di un Program Office
- Implementare il Data Mesh su un singolo Cloud o in un ambiente ibrido multi-Cloud
- Implementare un Data Lake o Lakehouse
- Standardizzare il processo di implementazione del dominio
- Creare zone in un Cloud Storage account, un Data Lake o Lakehouse per produrre e conservare prodotti dati in un Data Mesh
- Usare Kafka come una opzione per conservare prodotti dati in un Data Mesh
- Selezionare software di Data Fabric come una piattaforma per team orientati al dominio per costruire prodotti dati
- Sviluppo step-by-step di prodotti dati
 - o Registrazione della sorgente dati
 - o Data discovery automatizzato, data quality profiling, sensitive data detection, classificazione della Governance
 - o Data Ingestion
 - o Creazione di una policy globale e di dominio per Governance federata di dati classificati
 - o Sviluppo di pipeline di prodotti dati
 - o Standardizzare le Best Practice e togliere complessità ai data engineer
 - o Publishing del Prodotto dato per il consumo

5. Costruire Pipeline di DataOps per creare Prodotti Dati multi-uso

Questa sessione esamina la progettazione e lo sviluppo di pipeline DataOps modulari per produrre prodotti di dati affidabili utilizzando il software Data Fabric.

- Sviluppo collaborativo di pipeline e orchestrazione per produrre prodotti dati
- Progettare pipeline DataOps basato su componenti per produrre prodotti dati
- Usare CI/CD per accelerare sviluppo, testing e deployment
- Progettare in sensitive data protection nelle pipeline
- Processing streaming data in una pipeline
- Gestione della deriva dello schema in una pipeline
- Processing di dati non strutturati in una pipeline usando ML
- Generare pipeline di dati usando i tool di Data Automation
- Rendere i prodotti dati disponibili per il consumo in un Data Mesh usando un Data Marketplace
- The Enterprise Data Marketplace: mettere in condizione i consumatori di avere informazioni per comprare i dati
- Fornire prodotti dati affidabili da utilizzare in più sistemi analitici e in MDM
- Consumare prodotti dati in altre pipeline da usare nei Data Warehouse, Lakehouse, Data Science sandbox, graph analysis, Kafka e MDM

6. Implementare una Data Governance federata per produrre e usare Prodotti Dati Compliant

Con i dati altamente distribuiti su così tanti archivi di dati e applicazioni, on-premise, in più Cloud e perimetrali, molte aziende stanno lottando per governare i dati durante il loro ciclo di vita. Questo è di fondamentale importanza in un Data Mesh in cui la governance dei dati computazionali federati è un principio fondamentale, lo sviluppo dei prodotti dati è decentralizzato e i prodotti dati sono condivisi e consumati in tutta l'organizzazione. È anche fondamentale nell'intero panorama di dati multi-Cloud ibrido. Questa sessione esamina come questo può essere ottenuto.

- Cosa è coinvolto in una Data Governance federata?
- Come implementarla in un panorama dati distribuiti ibridi, multi-Cloud?

- Capire gli obblighi della Compliance
- Tipi di politiche di Data Governance
- Capire politiche globali e locali quando si crea un Data Mesh
- Definire tipi di dati sensibili
- Usare il Catalogo dati per automatizzare il data profiling, il grado di qualità e la classificazione dei dati sensibili
- Definire e attaccare politiche ai dati classificati in un Catalogo dati
- Creare prodotti dati Master Data e Reference data condivisibili per MDM e RDM
- Assicurare la qualità dei dati in uno sviluppo di prodotti dati
- Proteggere i dati sensibili nello sviluppo di prodotti dati per la Compliance con la Data Privacy
- Governare la gestione delle versioni dei prodotti dati
- Governare l'accesso ai prodotti dati contenenti dati sensibili
- Prevenire la condivisione eccessiva accidentale di prodotti con dati sensibili utilizzando DLP
- Gestione della conservazione dei dati dei prodotti dati in linea con la conformità e gli aspetti legali
- Monitoraggio e gestione dei dati per garantire l'applicazione della policy
- Tecnologie di Data Catalog e Data Fabric che aiutano a governare i dati nello scenario di dati distribuiti
 - o Tipi di offerte di Data Governance
 - o Alation, Ataccama, Collibra, Dataguisse, Confluent Schema Registry and Catalog
 - o Google Cloud IAM, Data Catalog, BigQuery, Dataplex e DLP
 - o IBM Cloud Pak for Data, Watson Knowledge Catalog, Optim & Guardium
 - o Hitachi Vitara, Immuta, Imperva
 - o Informatica EDC and Axon
 - o Microsoft Purview
 - o Okera, OneTrust Data Governance Suite
 - o Oracle Enterprise Data Management Cloud
 - o Privitar
 - o SAP Data Intelligence
 - o Software AG StreamSets DataOps Platform
 - o Talend, TopQuadrant

INFORMAZIONI

<p>QUOTA DI PARTECIPAZIONE</p> <p>€ 1200 (+iva)</p> <p>DURATA E ORARIO</p> <p>2 giorni: 9.30-13.00 14.00-17.00</p> <p>Il seminario sarà tenuto in lingua inglese con il servizio di Traduzione Simultanea incluso.</p>	<p>MODALITÀ D'ISCRIZIONE</p> <p>Il pagamento della quota, IVA inclusa, dovrà essere effettuato tramite bonifico, codice IBAN: IT 03 W 06230 03202 000057031348 Banca: Credit Agricole Agenzia 1 di Roma intestato alla Technology Transfer S.r.l. e la ricevuta di versamento inviata insieme alla scheda di iscrizione a: info@technologytransfer.it</p> <p>TECHNOLOGY TRANSFER S.r.l. Piazza Cavour, 3 00193 ROMA (Tel. 06-6832227)</p> <p>Vi consigliamo di far precedere la scheda d'iscrizione da una prenotazione telefonica.</p>	<p>CONDIZIONI GENERALI</p> <p>In caso di rinuncia con preavviso inferiore a 15 giorni verrà addebitato il 50% della quota di partecipazione, in caso di rinuncia con preavviso inferiore ad una settimana verrà addebitata l'intera quota. In caso di cancellazione del seminario, per qualsiasi causa, la responsabilità della Technology Transfer si intende limitata al rimborso delle quote di iscrizione già pervenute.</p> <p>SCONTI</p> <p>I partecipanti che si iscriveranno al seminario 30 giorni prima avranno uno sconto del 5%.</p> <p>Se un'azienda iscrive allo stesso evento 5 partecipanti, pagherà solo 4 partecipazioni. Gli sconti per lo stesso evento non sono cumulabili fra di loro.</p>	<p>TUTELA DATI PERSONALI</p> <p>Ai sensi dell'art. 13 della legge n. 196/2003, il partecipante è informato che i suoi dati personali acquisiti tramite la scheda di partecipazione al seminario saranno trattati da Technology Transfer anche con l'ausilio di mezzi elettronici, con finalità riguardanti l'esecuzione degli obblighi derivati dalla Sua partecipazione al seminario, per finalità statistiche e per l'invio di materiale promozionale dell'attività di Technology Transfer. Il conferimento dei dati è facoltativo ma necessario per la partecipazione al seminario. Il titolare del trattamento dei dati è Technology Transfer, Piazza Cavour, 3 - 00193 Roma, nei cui confronti il partecipante può esercitare i diritti di cui all'art. 13 della legge n. 196/2003.</p>
---	--	--	---



MIKE FERGUSON

LINEE GUIDA PRATICHE PER IMPLEMENTARE UN DATA MESH

26-27 Aprile 2023

Quota di iscrizione:
€ 1200 (+iva)

nome

cognome

funzione aziendale

azienda

partita iva

codice fiscale

indirizzo

città

cap

provincia

telefono

fax

e-mail



Timbro e firma

In caso di rinuncia o di cancellazione dei seminari valgono le condizioni generali riportate all'interno.

Da restituire compilato a:
Technology Transfer S.r.l.
Piazza Cavour, 3 - 00193 Roma
Tel. 06-6832227 - Fax 06-6871102
info@technologytransfer.it
www.technologytransfer.it

DOCENTE

Mike Ferguson è Managing Director di Intelligent Business Strategies Limited. Come analista e consulente è specializzato in Business Intelligence e Enterprise Business Integration. Ha più di 41 anni di esperienza di IT, ha operato consulenze per importanti aziende ed è stato frequentemente speaker in molti eventi e seminari in tutto il mondo. E' stato Principal e co-fondatore della Codd & Date UK e Chief Architect alla Teradata su DBMS Teradata e Managing Director Europeo di Database Associates. Insegna delle master classes che hanno molto successo su Analitica, Big Data, Data Governance & MDM, Data Warehouse Modernization e Data Lake Operations.