

LA TECHNOLOGY TRANSFER PRESENTA

MIKE FERGUSON

**Progettare, costruire e gestire
un Multi-purpose DATA LAKE**

ONLINE LIVE STREAMING

12-13 NOVEMBRE 2020



info@technologytransfer.it
www.technologytransfer.it

DESCRIZIONE

Molte aziende oggi devono trattare con molti silos di informazione come sistemi Cloud e sistemi on-premise, diversi Data Warehouses, Data Marts, sistemi MDM (Master Data Management), databases NoSQL, Cloud storage, sistemi Hadoop e Content Management in aggiunta le sorgenti dati stanno aumentando vertiginosamente.

Non sorprende che molte aziende stiano lottando per conoscere quali dati sono disponibili, se sono affidabili e come si integrano. Molte aziende stanno smettendo di gestire queste informazioni in silos con tools differenti usati per preparare e integrare i dati. Per di più sia gli utenti dell'IT che quelli di Business stanno ora integrando i dati e tutto questo genera caos.

A questo punto la domanda è: Dobbiamo continuare a operare così o c'è un altro modo per governare e unificare molto più facilmente i dati in questo panorama sempre più complesso che può abbreviare il time to value?

Questo seminario esamina come il nuovo approccio DataOps propone una nuova Architettura dati in cui i dati sono inseriti e organizzati in un Data Lake e gli utenti IT e di Business collaborano per unificare e costruire dati affidabili che sono pubblicati in un Data Market Place e disponibili agli altri. L'obiettivo è ridurre il time to value, evitare la reinvenzione quando governiamo e unifichiamo velocemente i dati in un multi-Cloud, Data Store multipli e ambiente di Computing ibrido.

In particolare i partecipanti impareranno:

- Come definire una strategia per produrre Data-as-a-Service in un ambiente distribuito di multipli Data Stores e sorgenti multiple di dati
- Come organizzare i dati in un ambiente centralizzato o distribuito per superare la complessità e il caos
- Come progettare, costruire e gestire un Logical o centralizzato o distribuito Data Lake
- L'importanza di un Catalogo dell'Informazione per produrre Data-as-a-Service
- Come la standardizzazione dei dati e i glossari di business possono aiutare a rendere i dati più comprensibili
- Un modello operativo per una efficace governance dell'informazione distribuita
- Le tecnologie e le metodologie di implementazione di cui avete bisogno per tenere tutto sotto controllo e produrre "prodotti Dati" pronti e attendibili
- Quali metodologie applicare per tenere sotto controllo Master e Reference Data, Big Data, Data Warehouse e dati non strutturati, sia che si trovino su Cloud o su sistemi on-premise

Partecipanti

Business Data Analysts, Data Architects, Chief Data Officers, Professionisti di Master Data Management, Professionisti di Content Management, Database Administrators, Professionisti di Big Data, Sviluppatori di Data Integration Compliance Managers che sono responsabili di EIM, Data Quality Managers.

PROGRAMMA

1. Mettere a punto una Strategia Dati per una rapida unificazione di Dati affidabili

Questa sessione introduce il Data Lake insieme alla necessità di avere una strategia per i dati e definisce perché le aziende ne hanno bisogno. Spiega quale dovrebbe essere nella vostra strategia il modello operativo per implementarla, i tipi di dati che dovete gestire e lo scopo dell'implementazione.

- I dati distribuiti in continua crescita
- L'approccio a silos per la gestione e la governance dei dati
- Integrazione dei dati IT, self-service dei dati. Data Governance o caos dei dati?
- Principali requisiti per il Data Management
- Trattare con le nuove sorgenti dati: dati del Cloud, dati dei sensori, dati dei social media, prodotti smart (Internet of Things)
- Capire lo scopo del vostro Data Lake
- Costruire un caso di business per il Data Management
- Definire una strategia per i dati
- Un nuovo approccio collaborativo per governare, gestire e curare i dati
- Introdurre il Data Lake e la Data Refinery
- Configurazioni del Data Lake: quali sono le opzioni? o Centralizzato, distribuito o Logico
- Mettere su un Data Lake Multi-Purpose e una Information Supply Chain per produrre "prodotti dati" per l'azienda
- DataOps: un approccio a componenti per curare prodotti dati affidabili
- L'importanza sempre crescente di un Catalogo dell'Informazione e il suo ruolo come un mercato dei dati
- Principali componenti tecnologiche in un Data Lake e Information Supply Chain includendo Data Fabric Software
- Usare il Cloud o Hadoop come una Data Staging Area e perché non è abbastanza
- Opzioni di implementazione Run-Time: la necessità

di esecuzione in ambienti multipli

- Integrare un Data Lake nella vostra Enterprise Analytical Architecture

2. Metodologie di Produzione dell'Informazione

Avendo capito la strategia, questa sessione esamina le diverse metodologie e le tecnologie necessarie a tenere i vostri dati sotto controllo.

- Produzione dell'Informazione e consumo dell'Informazione
- Una metodologia pratica di Data Governance
- Perché la metodologia deve cambiare per dati semi-strutturati e non strutturati
- Metodologie per dati strutturati e multi-strutturati

3. Standardizzazione dei dati e il Glossario del Business

Questa sessione spiega perché è necessaria la standardizzazione dei dati per i dati strutturati e i nuovi modi di fare il processing dei dati non strutturati. Per fare questo bisogna creare nomi e definizioni di dati comuni per costruire uno Shared Business Vocabulary (SBV). SBV dovrebbe essere definito e memorizzato in un Glossario di Business.

- Standardizzazione semantica dei dati usando uno SBV dentro un Catalogo dell'Informazione
- Il ruolo di un vocabolario comune in MDM, RDM, SOA, DW e Data Virtualization
- Perché un vocabolario comune è rilevante nel Data Lake e nel Logica Data Warehouse?
- Approcci per creare un vocabolario comune
- Prodotti di Business Glossary come ad esempio: Alteryx Connect Glossary, ASG, Collibra, Informatica Axon, IBM Information Governance Catalog, Microsoft Azure Data Catalog Business Glossary, SAS Business Data Network
- Pianificare un Glossario di Business
- Organizzare definizioni di dati in un Glossario

di Business

- Principali ruolo e responsabilità: attivare il giusto modello operativo per creare e gestire un SVB
- Formalizzare la governance dei nomi dei dati di business
- Coinvolgimento del business nella creazione di SBV
- Oltre i dati strutturati: dal Glossario di Business all'Information Catalog
- Cosa è un Information Catalog?
- Perché l'Information Catalog è un fattore critico del Data Management?
- Tecnologie di Information Catalog: Alation, Alteryx Connect, Amazon Glue, Apache Atlas, Collibra Catalog, Cambridge Semantics ANZO Data Catalog, Denodo Data Catalog, Google Data Catalog, IBM Information Governance Catalog & Watson Knowledge Catalog, Informatica EDC & Live Data Map, Microsoft Azure Data Catalog, Qlik Data Catalyst, Waterline Data, Zaloni Data Platform
- Capabilities dell'Information Catalog

4. Organizzare il Data Lake

Questa sessione spiega come organizzare i dati in modo da poterli gestire in questo ambiente complesso. Esamina zoning e versioning, l'uso di un information Catalog nella gestione dei dati e la necessità di collaborazione fra Business e IT.

- Organizzare i dati in un Data Lake Centralizzato o Logico
- Creare zone per gestire i dati
- Nuovi requisiti per gestire dati in Data Lakes Centralizzati e Logici
- Creare progetti di Data Lake collaborativi
- Hadoop o memoria Cloud come staging area per la pulizia e integrazione dei dati aziendali
- Processi fondamentali nelle operazioni di Data Lake
- Il processo di Data Ingestion
- Tools e tecniche di Data Ingestion
- Implementazione di dati sistematici e individuazione di relazioni fra dati diversi mediante il software

dell'Information Catalog

- Usare domini e Machine Learning per accelerare il discovery e il tagging
- Alation, IBM Watson Knowledge Catalog, Informatica CLAIRE, Silwood, Waterline Data Smart Data Catalog
- Profiling, tagging e catalogazione automatica dei dati
- Data mapping automatico e lineage discovery
- La classificazione della governance dei dati e la politica di definizione dei processi
- Classificazione della governance dei dati manuale e automatica per permettere la governance
- Usare politiche basate su tag per governare i dati

5. Il processo di Data Refinery

Questa sessione esamina il processo di raffinare e curare i dati in una Information Supply Chain per produrre "prodotti dati" affidabili.

- Cosa è una Data Refinery?
- I Principali Requisiti per affinare i dati
- Opzioni per la raffinazione dei dati
- Approcci chiave per una integrazione scalabile dei dati usando Apache Spark
- Self-Service Data Preparation tools per Spark e Hadoop: Alteryx Designer, Azure Data Factory Wrangling Flows, Informatica Intelligent Data Lake, IBM Data Refinery, Paxata, Tableau (Project Maestro), Tamr, Talend, Trifacta
- Data Profiling automatico usando analitica nei tools di preparazione dei dati
- Eseguire compiti di Data Refinery in un Data Lake distribuito usando Apache Beam
- Approcci per integrare ETL e Self-Service Data Preparation
- ODPi Egeria per la condivisione di Metadati
- Unire il processing analitico da ETL ai pipelines analitici
- Pubblicare dati nell'Information Catalog
- Fare il mapping fra i dati scoperti e il vostro DW e vo-

cabolario di Business

- Data Provisioning: pubblicare prodotti Dati già pronti e affidabili in un Enterprise Data Marketplace
- Enterprise Data Marketplace: permettere ai consumatori dell'informazione di comprare Dati
- Approvvigionare dati sicuri usando la virtualizzazione dei dati, Logical Data Warehouse e servizi informativi on-demand
- Un Data Management consistente fra sistemi Cloud e sistemi on-premise

6. Unificare Big Data, Master Data e Data Warehouses per ottenere valori di Business

Questa sessione guarda a come i processi di raffinamento dei dati devono essere applicati per governare, unificare e fornire dati attraverso Big Data, MDM e Data Warehouse per produrre nuovo valore di Business. Come relazionarsi con grandi volumi e grandi varietà di dati? Come caricare e processare i dati in Hadoop? Come dovrebbero essere gestiti i dati low-latency?

- Come creare una single customer view
- Tipi di Big Data e Small Data necessari per creare una customer view
- Connettersi alle sorgenti di Big Data: Web logs, click stream, sensor data, contenuto non strutturato e semi strutturato
- Immettere e analizzare dati clickstream
- La sfida di catturare dati esterni dei clienti dai social networks
- Trattare la qualità dei dati non strutturati in un ambiente Big Data
- Usare la graph analysis per identificare nuove relazioni
- La necessità di combinare Big Data, Master Data e dati nel vostro Data Warehouse
- Fare il match fra Big Data e Customer Master Data
- Governare i dati in un ambiente Data Science

7. Audit e Protezione dell'Informazione: Governare i Dati in un ambiente di Dati distribuito

Questa sessione definisce il problema importante dell'Audit e della Protezione dell'informazione ed esamina quali tecnologie sono disponibili per aiutarvi a integrare questi aspetti nella vostra strategia dati.

- Cosa sono la sicurezza e l'audit dei dati e cosa è coinvolto nella loro gestione?
- Status check: a che punto siamo oggi sull'audit dei dati aziendali, la sicurezza degli accessi e la protezione?
- Quali sono i requisiti per l'audit dei dati aziendali, la sicurezza degli accessi e la protezione?
- Cosa è necessario considerare quando trattiamo questi aspetti?
- Data Discovery automatico e il Catalogo dell'Informazione: un grande aiuto per identificare i dati sensibili
- Cosa sugli utenti privilegiati?
- Utilizzo di una piattaforma di gestione dati e di un Catalogo di Informazioni per governare i dati tra più archivi di dati
- Mettere in sicurezza e proteggere i dati usando politiche basate sui tag in un Information Catalog
- Quali tecnologie sono disponibili per affrontare questo problema: Apache Knox, Cloudera Sentry, Dataguise, Hortonworks Ranger, IBM (Watson Data Platform, Knowledge Data Catalog, Optim & Guardium), Immuta, Informatica Secure@Source, Imperva, Micro Focus, Okera, Privitar
- Queste tecnologie possono aiutare il GDPR?
- Come si integrano con i programmi di Data Governance?
- Come cominciare per mettere in sicurezza, fare l'audit e proteggere i vostri dati?

INFORMAZIONI

<p>QUOTA DI PARTECIPAZIONE</p> <p>Quota senza Traduzione Simultanea € 1100(+iva)</p> <p>Quota con Traduzione Simultanea € 1200(+iva)</p> <p>La quota di partecipazione comprende documentazione.</p> <p>DURATA ED ORARIO</p> <p>2 giorni: 9.30-13.00 14.00-17.00</p> <p>Il seminario sarà tenuto in lingua inglese con la possibilità di avere la Traduzione Simultanea</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: Cariparma 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>entro il 20 Maggio 2020</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

PROGETTARE, COSTRUIRE E GESTIRE UN MULTI-PURPOSE DATA LAKE

12-13 Novembre 2020

Quota di iscrizione:

Quota senza Traduzione Simultanea
€ 1100(+iva)

Quota con Traduzione Simultanea
€ 1200(+iva)

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

nome

cognome

funzione aziendale

azienda

partita iva

codice fiscale

indirizzo

città

cap

provincia

telefono

fax

e-mail



Timbro e firma

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 36 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.