

LA TECHNOLOGY TRANSFER PRESENTA

MIKE FERGUSON

**Progettare, costruire e gestire un
Enterprise DATA LAKE**

5-6 DICEMBRE 2018

RESIDENZA DI RIPETTA - VIA DI RIPETTA, 231



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 RDM (Reference Data Management), sistemi MDM (Master Data Management), sistemi ECM (Enterprise Content Management), piattaforme NoSQL per Big Data come Hadoop o altri databases NoSQL. In aggiunta le sorgenti dati stanno aumentando vertiginosamente, in particolare quelle esterne all'azienda.

Trovandosi questa situazione, non sorprende che molte aziende stiano smettendo di gestire queste informazioni in silos con tools differenti usati per preparare e gestire i dati attraverso questi sistemi con vari gradi di governance. Per di più oggi non è solo l'IT che gestisce i dati. Gli utenti del business cominciano ad avere i propri dati.

A questo punto la domanda è: ma è questo il solo modo per gestire i dati? Esiste un altro livello che possiamo raggiungere che ci permetta di gestire e governare molto più facilmente i dati in questo panorama sempre più complesso?

Questo seminario esamina le problematiche che le aziende si trovano ad affrontare relativamente alla proliferazione delle sorgenti dati, alla difficoltà di raccogliere questi dati in diversi data stores (Cloud e on-premise), in diversi sistemi analitici e spiega quali sono i requisiti affinché un'azienda sia in grado di definire, governare, gestire e condividere un'informazione sicura e di qualità in questo ambiente ibrido di computing.

Viene descritto un nuovo approccio affinché gli architetti dei dati, gli utenti del business e gli sviluppatori IT possono collaborare per costruire e gestire un Logical Data Lake e riprendere il controllo dei propri dati.

Questo include data ingestion, automated data discovery, data profiling, tagging e pubblicazione dei dati in un catalogo dell'informazione disponibile per tutta l'azienda. Vengono anche introdotte configurazioni multiple di Data Lake, includendo un Data Lake centralizzato e un "Logical" e distribuito Data Lake. In fine si enfatizza la necessità di avere un processo collaborativo comune e un comune approccio per governare e gestire tutti i tipi di dati.

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
- 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

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

PROGRAMMA

1. Strategia e Pianificazione

Questa sessione introduce il Data Lake insieme alla necessità di avere una strategia per i dati e 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
 - o Dati strutturati: master, reference e transazioni
 - o Dati semi-strutturati: JSON, BSON, XML
 - o Dati non strutturati: testi, video, audio
 - o Servizi riusabili per gestire i dati
- 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
 - o Sistemi OLTP
 - o Data Warehouses
 - o Sistemi di Big Data
 - o Sistemi MDM e RDM
 - o Virtualizzazione dei dati
 - o Streaming Data
 - o Enterprise Content Management
- Costruire un caso di business per il Data Management
- Definire una strategia per i dati
- Un nuovo approccio inclusivo per gestire e governare i dati
- Introdurre il Data Lake e la Data Refinery
- Configurazioni del Data Lake: quali sono le opzioni?
 - o Centralizzato, distribuito o Logico
- Casi d'uso di Information Supply Chain
- L'importanza sempre crescente di un catalogo dell'informazione
- Principali componenti tecnologiche in un Data Lake

- Hadoop come una Data Staging Area e perché non è abbastanza
- Opzioni di implementazione Run-Time: la necessità di esecuzione in ambienti multipli
- Principali ruoli e responsabilità
- Tipi di politica per governare i dati
- Formalizzare i processi di governance
- Integrare un Data Lake nella Vostra Enterprise Analytical Architecture

2. Metodologia e Tecnologie

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, Global IDs, Informatica, IBM Information Governance Catalog, Microsoft Azure Data Catalog Business Glossary, SAP Information Steward Metapedia, SAS
- 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, IBM Information Governance Catalog & Watson Knowledge Catalog, Informatica Live Data Map, Microsoft Azure Data Catalog, Podium Data, Waterline Data, Zaloni Mica
- 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 distribuito
- Creare zone per gestire i dati
- Nuovi requisiti per gestire dati in Data Lakes centralizzati e distribuiti
- Creare progetti di Data Lake collaborativi
- Hadoop 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
- La classificazione dei dati e la politica di definizione dei processi
- Classificazione 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 scoprire dove sono i Vostri dati e come raffinarli per tenerli sotto controllo.

- 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, 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
- Apache Atlas Open Metadata & Governance
- Unire il processing analitico da ETL ai workflows

analitici

- Pubblicare dati nell'Information Catalog
- Fare il mapping fra i dati scoperti e il Vostro DW e vocabolario di business
- Data Provisioning: immettere consistente informazione nei Data Warehouses, sistemi MDM, data bases NoSQL e sistemi transazionali
- Approvvigionare dati raffinati usando la virtualizzazione dei dati, Logical Data Warehouse e servizi in formativi on-demand
- Governare il processo di approvvigionamento usando metadati basati su regole
- Un Data Management consistente fra sistemi Cloud e sistemi on-premise

6. Raffinare i Big Data e i dati per i Data Warehouses

Questa sessione guarda a come i processi di raffinamento dei dati devono essere applicati per gestire, governare e fornire dati in un ecosistema analitico di Big Data e in un tradizionale Data Warehouse. 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: l'aspetto dimenticato della Data Governance

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, Dataguisse, Hortonworks Ranger, IBM (Watson Data Platform, Knowledge Data Catalog, Optim & Guardium), Imperva, Informatica Secure@Source, Micro Focus, 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

QUOTA DI PARTECIPAZIONE € 1300 (+iva) La quota di partecipazione comprende documentazione, colazioni di lavoro e coffee breaks. LUOGO Roma, Residenza di Ripetta Via di Ripetta, 231 DURATA ED ORARIO 2 giorni: 9.30-13.00 14.00-17.00 È previsto il servizio di traduzione simultanea	MODALITÀ D'ISCRIZIONE 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: TECHNOLOGY TRANSFER S.r.l. Piazza Cavour, 3 00193 ROMA (Tel. 06-6832227 Fax 06-6871102) entro il 20 Novembre 2018 Vi consigliamo di far precedere la scheda d'iscrizione da una prenotazione telefonica.	CONDIZIONI GENERALI 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. SCONTI I partecipanti che si iscriveranno al seminario 30 giorni prima avranno uno sconto del 5%. 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.	TUTELA DATI PERSONALI 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.
--	--	---	--

MIKE FERGUSON

PROGETTARE, COSTRUIRE E GESTIRE UN ENTERPRISE DATA LAKE

Roma 5-6 Dicembre 2018
Residenza di Ripetta - Via di Ripetta, 231
Quota di iscrizione: € 1300 (+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.

È previsto il servizio di traduzione simultanea

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 Europeo di Intelligent Business Strategies Limited. Come consulente, è specializzato in sistemi database, Business Intelligence, Enterprise Application Integration, Corporate ed e-Business Portals, soluzioni intelligenti di CRM e di Supply Chain. Ha più di 30 anni di esperienza nell'IT, ha operato consulenze per importanti aziende ed è stato frequentemente speaker in molti eventi e seminari in tutto il mondo. È stato Principal e co-fondatore della Codd & Date UK e Chief Architect alla NCR sul DBMS Teradata.