

TECHNOLOGY TRANSFER PRESENTS

# MIKE FERGUSON

## DATA WAREHOUSE MODERNISATION

Modern Data Modelling Techniques, Automated Data Discovery, Data Catalog,  
DataOps Pipelines, Cloud DW Development and Migration, Data Warehouse Automation,  
Data Lakehouse, Virtual Data Marts, Real-time Data Warehouse,  
Machine Learning, Logical Data Warehouse and External Tables

ONLINE LIVE STREAMING

MAY 15-16, 2023



info@technologytransfer.it  
www.technologytransfer.it

## ABOUT THIS SEMINAR

Today, with most people connected to the Internet, the power of the customer is almost limitless. They can browse your competitors' Web sites. They can compare prices, view sentiment about your business, and switch loyalty in a single click any-time, anywhere all from a mobile device. In addition, social media has given customers a voice to express opinion and sentiment about products and brands and to create social networks by attracting followers, and following others. For many CEOs, customer retention, loyalty, service and growth are top of their agenda. Therefore, they want access to new data to enrich what they already know about customers.

In addition, COOs are adding telemetry to capture new data to optimise operations. Yet at the same time, regulations like GDPR, KYC, MiFiD are everywhere, making governance and risk management also a priority.

Given these new requirements, many companies running traditional Data Warehouses and data marts are realising that just recording historical transaction activity is not enough. The pace of change is quickening, business is demanding lower latency data, the backlog of changes to Data Warehouses and data marts is growing rapidly while testing remains slow and complicated. Also with business unit autonomy, new technology available on the Cloud, and pent up demand for Machine Learning everywhere, shadow IT is springing up in business units fracturing the analytical effort and building new analytical silos that are not integrated with Data Warehouses. With so much pressure to remain competitive, how then do you modernise your analytical setup, to improve governance and agility, bring in new data, re-use data assets, modernise your Data Warehouse to easily accommodating change, lower data latency and integrate with other analytical workloads to provide a new modern logical Data Warehouse for the digital enterprise?

This 2-day seminar looks at the business case as to why you need to do this, discusses the tools and techniques needed to capture new data types, establish a data pipeline to produce re-usable data assets, modernise your Data Warehouse and bring together the data and analytics needed to accelerate time to value, deliver new insights to foster growth, reduce costs, improve effectiveness and enable competitive advantage.

### WHO SHOULD ATTEND

- Understand why Data Warehouse modernisation is needed to help improve decision making and competitiveness
- Have the ingredients to know how to modernise your Data Warehouse to improve agility, reduce cost of ownership, facilitate easy maintenance
- Understand modern data modelling techniques and how to reduce the number of data stores in a Data Warehouse without losing information
- Understand how to exploit Cloud computing at lower cost and how to migrate to the cloud
- Understand how to reduce data latency in your data warehouse
- Understand how to integrate you data warehouse with data stored in cloud storage data lakes
- Know how to migrate from a waterfall based Data Warehouse and data marts to a lean, modern Logical Data Warehouse with virtual data marts that integrates easily with cloud storage and other analytical systems
- Know how to use data virtualisation to simplify access to a more comprehensive set of insights available on multiple analytical platforms running analytics on different types of data for precise evidence-based decision making
- Understand the role of a modern Data Warehouse in a data-driven enterprise

# OUTLINE

## 1. The Traditional Data Warehouse and why it needs Modernised

For many organisations today, their Data Warehouse is still based on a waterfall style architecture with data flowing from source systems into operational data stores, staging areas, then on to Data Warehouses under the management of batch ETL jobs. However, analytical landscape has changed. New data sources continue to grow with data now being collected in edge devices, Cloud storage, Cloud or on-premises NoSQL data stores, Hadoop systems as well as Data Warehouse staging. Cloud storage, streaming data platforms and Graph databases are also now used in data analysis. Also many business units are using the Cloud to quickly exploit these new analytical technologies at lower cost.

This opening session looks at these new activities and explains why Data Warehouses have to change not only to speed up development, improve agility, reduce costs but also to exploit new data, enable self-service data preparation, utilise advanced analytics and integrate with these other analytical platforms.

- The traditional Data Warehouse
- Multiple data stores, waterfall data architecture and data flows
- New data entering the enterprise
- The changing face of analytics - new analytical data stores and platforms
  - o Big Data analytics on Spark, Cloud storage and Hadoop
  - o Real-time streaming data analytics
  - o Graph analysis in Graph Databases
- New challenges brought about by:
  - o Data complexity
  - o Data management siloes
  - o Managing data in a distributed and hybrid computing environment
  - o Self-service data prep vs ETL/DQ
- Problems with existing Data Warehouse architecture and development techniques
- The need to avoid silos, accommodate new data, accommodate change quickly and integrate analytical workloads to deliver value

## 2. Modern Data Warehouse Requirements

This session looks at the key building blocks of modern Data Warehouse that need to be in place for flexibility and agility.

- Modern data modelling techniques
- Accelerating ETL processing using automated data discovery, a data catalog DataOps pipelines and reusable data products
- Cloud based analytical DBMS
- External tables Lakehouses and in-database analytics
- Shortening development time using Data Warehouse automation
- Data Virtualisation for data independence, flexibility and to integrate new analytical data stores into a Logical Data Warehouse
- Incorporating fast streaming data, prescriptive analytics, embedded and operational BI

## 3. Modern Data Modelling Techniques for Agile Data Warehouse

In order to improve agility, change friendly data modelling techniques have emerged and are becoming increasingly popular in designing modern data warehouses. This session looks at data modelling and asks Is Star Schema dead? Which Data Warehouse modelling technique is best suited to handling change? Should you use Data Vault? Does Data Warehouse design need to change? Does Data Mart design need to change? It also looks at the disadvantages of such techniques and how you can overcome these.

- Data Warehouse modelling approaches - Inmon Vs Kimball Vs Data Vault
- The need to handle change easily
- What is Data Vault?
- Data Vault modelling components - hubs, links and satellites
- Pros and cons of data modelling techniques
- Using data virtualisation to improve agility in data marts while reducing cost
- Using governance processes in data standardisation
- Enterprise Data Modelling using a SBV

#### 4. Modernising your ETL Processing

This session looks at the challenges posed by new data on ETL processing. Also, what options are available to modernise ETL processing, where should it run and what are the pros and cons of each option? How does this impact on your Data Architecture?

- New data and ETL processing - high volume data, semi-structured data, unstructured data, streaming data (e.g., IoT data)
- What are the implications and challenges of this new data on ETL processing?
- Should all this data go into a Data Warehouse or not?
- What options are available to modernise Data Warehouse ETL processing?
  - o Data Lake Vs Lakehouse Vs Data Mesh?
  - o Offloading staging data to a Data Lake and use Spark for scalable ETL processing
  - o The real-time data option
  - o Using Data Warehouse automation software to generate ETL processing
- Pros and cons of these options
- Data Architecture implications of modernising and democratising data engineering

#### 5. Accelerating ETL Processing using Data Products a Multi-Purpose Data Lake, a Lakehouse or Data Mesh

This session looks at how you can use a multi-purpose Data Lake, Lakehouse and Data Mesh to accelerate ETL processing and integration of data for your Data Warehouse

- How can you accelerate ETL processing and self-service data preparation?
- Decentralising data engineering
- Using Data Fabric to build DataOps pipelines to create reusable data products in a Data Mesh
- Options for Implementing Data Mesh - data lakes, lakehouses, cloud data platforms, Kafka and data virtualisation
- Directly accessing data sources Vs ingesting and staging your data in readiness for ELT processing

- Using a data catalog to automatically discover, classify, data quality profile, catalog and map data to a business glossary
- GDPR - Detecting sensitive data during automatic data discovery
- Masking GDPR sensitive data during ingestion or ETL pipeline execution
- Using Machine Learning in DataOps ELT pipelines to process unstructured data
- Pipeline processing streaming data in a real-time Data Warehouse
  - o Types of streaming data, weblogs OLTP system change data capture,
  - o Key technologies for processing streaming data - Kafka, streaming analytics and event stores
  - o Turning OLTP change data capture into Kafka data streams
  - o Using Kafka as a data source to process data in real-time
  - o Ingesting streaming data into your Data Lakehouse or Data Warehouse
- Real-time Data Warehouse - Integrating your Data Warehouse with streaming data - external tables, data virtualisation and Data Lake OR using a lake house
- Using ETL data pipelines to produce re-usable data products for use in your Data Warehouse, Data Science and other analytical data stores
- Publishing reusable data products in a data market place ready for consumption
- Using Data Science to develop new analytical models to run in your Data Warehouse

#### 6. Rapid Data Warehouse Development using Data Warehouse Automation

To modernising ETL processing, this session looks at how you can use metadata driven Data Warehouse automation tools to rapidly build, change and extend modern Cloud and on premises Data Warehouses and data marts. It looks at how these tools help you adopt new modern data modelling techniques quickly, how they generate schemas and data integration jobs and how they can help you migrate to your new Data Warehouse systems on the Cloud.

- What is Data Warehouse Automation?
- Using Data Warehouse Automation Tools for rapid data Warehouse and Data Mart development
  - o Generating Data Vault, E/R and Star Schema design
  - o Data Pipeline generation
  - o Processing streaming data using Data Warehouse Automation
  - o Integrating Big Data with a Data Warehouse using Data Warehouse Automation
  - o Integrating Cloud Data Warehouses with Data Lakes using Data Warehouse Automation
  - o Using Data Warehouse Automation to migrate Data Warehouses
  - o Using Data Virtualisation to shield existing BI tools from changes in design
- The Data Warehouse Automation Tools market e.g. IDERA, WhereScape, biGENIUS, Attunity Compose, TimeExtender, Varigence BIMLStudio, Vault Builder & more
- Metadata driven Data Warehouse maintenance

## 7. Building a modern Data Warehouse in a Cloud Computing Environment

A key question for many organisations is what do you do with your existing Data Warehouse? Should you try to change the existing set-up to make it more modern or re-develop it in the Cloud? This session looks at the advantages of building modern Data Warehouses in a Cloud computing environment using a Cloud based analytical Relational DBMS.

- Why use Cloud Computing for your Data Warehouse?
- Cloud based Data Warehouse development - what are the options?
- Cloud based analytical relational DBMSs
  - o Amazon Redshift, Google BigQuery, IBM Db2 Warehouse on Cloud, Microsoft Azure Synapse Analytics, Oracle Autonomous Data Warehouse, Snowflake, SAP Data Warehouse Cloud, Teradata, Kinetica
- Lakehouses - Apache Iceberg/Databricks, HPE Ezmeral

- Separating storage from compute for elasticity and scalability
- Managing and integrating Cloud and on-premises data
- Using iPaaS software to integrate data in Cloud ETL processing - Informatica IICS, Boomi, SnapLogic
- Non-iPaaS Cloud ETL tools, e.g. AWS Glue, Azure Data Factory, Google Cloud Dataplex and Data Fusion, IBM Cloud Pak for Data, Talend, Software AG StreamSets
- Managing streaming data in the Cloud
- Integrating Big Data analytics into a Cloud based Data Warehouse
- Train and deploying Machine Learning models in your analytical database for in-Warehouse analytics
- Tools and techniques for migrating an existing Data Warehouse to the Cloud
- Dealing with Cloud DW migration issues
- Managing access to Cloud based Data Warehouses
- Integrating Cloud based BI tools with on-premise systems

## 8. Simplifying Data Access - Creating Virtual Data Marts and a Logical Data Warehouse Architecture to Integrate Big Data with your Data Warehouse

This section looks at how you can make use of data virtualisation software to modernise your Data Warehouse architecture, and simplify access to and integrate data in your Data Warehouse and Big Data underlying data stores and improve agility.

- What is data virtualisation?
- How does data virtualisation work?
- How can data virtualisation reduce cost of ownership, improve agility and modernise your Data Warehouse architecture?
- Simplifying your architecture by using data virtualisation to create Virtual Data Marts

- Migrating your physical data marts to virtual data marts to reduce cost of ownership
- Layering virtual tables on top of virtual marts to simplify business user access
- Publishing virtual views and queries as services in a data catalog for consumption
- Integrating your Data Warehouse with your Data Lake and low latency data using external tables and data virtualisation
- Enabling rapid change management using data virtualisation
- Creating a logical Data Warehouse architecture that Integrates data from Big Data platforms, graph databases, streaming data platforms and your Data Warehouse into a common access layer for easy access by BI Tools and applications
- Using a business glossary and data virtualisation to create a common semantic layer with consistent common understanding across all BI tools

## 9. Getting Started with Data Warehouse Modernisation

This final session looks at what you have to do to get started with a Data Warehouse modernisation initiative. In particular it looks at:

- Data Warehouse Modernisation options
  - Change Vs rebuild?
- What order do you do this in?
- How do you minimise impact on the business while you modernise?
- How to you deal with a backlog of change when you are also trying to modernise?
- Pros and cons of build Vs automating Data Warehouse development
- What new skills are needed?
- Delivering new business value whilst in the progress of modernising
- How do you involve business professionals in the modernisation effort?

## AUDIENCE

- CDOs
- CIO's
- IT Managers
- CTOs
- Business Analysts
- Data Scientists
- BI Managers
- Data Warehousing Professionals
- Enterprise Architects
- Data Architects

# INFORMATION

<p><b>PARTICIPATION FEE</b></p> <p>€ 1100</p> <p>The fee includes all seminar documentation.</p> <p><b>SEMINAR TIMETABLE</b></p> <p>9.30 am - 1.00 pm 2.00 pm - 5.00 pm</p>	<p><b>HOW TO REGISTER</b></p> <p>You must send the registration form with the receipt of the payment to: <b>TECHNOLOGY TRANSFER S.r.l.</b> Piazza Cavour, 3 - 00193 Rome (Italy) Fax +39-06-6871102</p> <p><b>PAYMENT</b></p> <p>Wire transfer to: Technology Transfer S.r.l. Banca: Credit Agricole Agenzia 1 di Roma IBAN Code: IT 03 W 06230 03202 000057031348 BIC/SWIFT: CRPPIT2P546</p>	<p><b>GENERAL CONDITIONS</b></p> <p><b>DISCOUNT</b></p> <p>The participants who will register 30 days before the seminar are entitled to a 5% discount.</p> <p>If a company registers 5 participants to the same seminar, it will pay only for 4. Those who benefit of this discount are not entitled to other discounts for the same seminar.</p> <p><b>CANCELLATION POLICY</b></p> <p>A full refund is given for any cancellation received more than 15 days before the seminar starts. Cancellations less than 15 days prior the event are liable for 50% of the fee. Cancellations less than one week prior to the event date will be liable for the full fee.</p> <p><b>CANCELLATION LIABILITY</b></p> <p>In the case of cancellation of an event for any reason, Technology Transfer's liability is limited to the return of the registration fee only.</p>
---	---	---

MIKE FERGUSON

## DATA WAREHOUSE MODERNISATION

May 15-16, 2023

Registration fee: € 1100

first name .....

surname .....

job title .....

organisation .....

address .....

postcode .....

city .....

country .....

telephone .....

fax .....

e-mail .....



Stamp and signature

Send your registration form with the receipt of the payment to:  
**Technology Transfer S.r.l.**  
Piazza Cavour, 3 - 00193 Rome (Italy)  
Tel. +39-06-6832227 - Fax +39-06-6871102  
info@technologytransfer.it  
www.technologytransfer.it

If anyone registered is unable to attend, or in case of cancellation of the seminar, the general conditions mentioned before are applicable.

## **SPEAKER**

**Mike Ferguson** is Managing Director of Intelligent Business Strategies Limited. As an analyst and consultant he specialises in business intelligence and enterprise business integration. With over 41 years of IT experience, Mike has consulted for dozens of companies. He has spoken at events all over the world and written numerous articles. Mike is Chairman of Big Data LDN – the fastest growing Big Data conference in Europe. Formerly he was a principal and co-founder of Codd and Date Europe Limited – the inventors of the Relational Model, a Chief Architect at Teradata on the Teradata DBMS and European Managing Director of Database Associates. He teaches popular master classes in Analytics, Big Data, Data Governance a Distributed Data Landscape, Data Warehouse Modernisation and Creating Data Products in a Data Mesh, Data Lake or Lakehouse for Use in Analytics.