

TECHNOLOGY TRANSFER PRESENTS

MIKE FERGUSON

**Building an Enterprise
Data Lake
for Enterprise
Data-as-a-Service**

MAY 2-3, 2017

**Predictive and
Advanced Analytics
for BI Professionals
and Business Analysts**

MAY 4-5, 2017

RESIDENZA DI RIPETTA - VIA DI RIPETTA, 231
ROME (ITALY)



info@technologytransfer.it
www.technologytransfer.it

ABOUT THIS SEMINAR

Most organizations today are dealing with multiple silos of information. These include cloud and on-premises based transaction processing systems, multiple data warehouses, data marts, reference data management (RDM) systems, master data management (MDM) systems, content management (ECM) systems and more recently Big Data NoSQL platforms such as Hadoop and other NoSQL databases. In addition the number of data sources is increasing dramatically especially from outside the enterprise. Given this situation it is not surprising that many companies have ended up managing information in silos with different tools being used to prepare and manage data across these systems with varying degrees of governance. In addition, it is not only IT that is now managing data. Business users are also getting involved with new self-service data wrangling tools. The question is, is this the only way to manage data? Is there another level that we can get reach to allow us to more easily manage and govern data across an increasingly complex data landscape?

This 3-day seminar looks at the business problems caused by poorly managed information. It looks at reference data, master data, transaction data, metrics, big data and unstructured content (e.g. documents, email, etc). It looks at the requirements to be able to define, govern, manage and share trusted high quality information in a hybrid computing environment. It also explores a new approach getting control of your data that includes participation from IT data architects, business users and IT developers. This includes creating and organising data in reservoirs and introduces data refineries in an enterprise approach to managing data. It emphasises the need for a common collaborative process and common data services to govern and manage data.

AUDIENCE

This seminar is intended for data Architects, chief data Officers, master data management Professionals, content management Professionals, database Administrators, Big Data Professionals, data integration Developers, and compliance Managers who are responsible for Enterprise Information Management (EIM). This includes metadata management, data integration, data quality, master data management and enterprise content management. It assumes that you have an understanding of basic data management principles as well as a high level of understanding of the concepts of data migration, data replication, metadata, data warehousing, data modelling, data cleansing, etc.

LEARNING OBJECTIVES

Attendees will learn how to define a strategy for Enterprise Information Management and how to implement it within their organisation. They will also learn the importance of data standardisation and business glossaries when defining data to be managed and the operating model for effective information governance. They will learn what technologies they need and an EIM implementation methodology to get their data under control. They will learn how to apply this methodology to get master and reference data, Big Data, Data Warehouse data and unstructured data under control whether it be on-premise or in the Cloud.

OUTLINE

1. Strategy & Planning

This session introduces Enterprise Information Management (EIM) and looks at the reasons why companies need it. It looks at what should be in your EIM strategy, the operating model needed to implement EIM, the types of data you have to manage and the scope of EIM implementation. It also looks at the policies and processes needed to bring your data under control.

- The ever increasing distributed data landscape
- The siloed approach to managing and governing data
- IT data integration, self-service data wrangling or both? – data governance or data chaos?
- Key requirements for EIM
 - Structured data – Master, reference and transaction data
 - Semi-structured data – JSON, XML, email
 - Unstructured data - Text, video, audio
 - Re-usable services to manage data
- Dealing with new data sources - Cloud data, sensor data, social media data, smart products (the internet of things)
- Understanding EIM scope
 - OLTP systems
 - Data Warehouses
 - Big Data systems
 - MDM and RDM systems
 - Data virtualisation
 - Messaging, ESBs and process data flows
 - Enterprise Content M'gmt
- Building a business case for EIM
- Defining a strategy for EIM
- A new inclusive approach to governing and managing data
- Introducing the data reservoir and data refinery
- Key roles and responsibilities
- Getting the organisation and operating model right
- Types of EIM policy
- Formalising governance processes, e.g. the dispute resolution process
- EIM in your Enterprise Architecture

2. Methodology & Technologies

Having understood strategy, this session looks at methodology for EIM and the technologies needed to help apply it to your data to bring it under control. It also looks at how EIM platforms provide the foundation in your Enterprise Architecture to manage information across the enterprise

- A best practice step-by-step methodology for EIM and Data Governance
 - Define, Identify, Assess, Integrate, Provision, Monitor, Protect and Secure
- The role of Hadoop as a data reservoir
- The EIM Marketplace: Actian, Global IDs, IBM InfoSphere, Informatica, Oracle, SAP, SAS, Talend
- Self-service data wrangling tools e.g. Paxata, Trifacta, Tamr, ClearStory Data
- EIM implementation options
 - Centralised, distributed or federated
 - Self-service BI – The need for Data Governance at the edge
 - EIM on-premise and on the Cloud
 - Common Data Services for Service-Oriented Data Management

3. EIM Implementation – Data Standardisation & the Business Glossary

This session looks at the first step in EIM – The need for data standardisation. The key to making this happen is to create common data names and definitions for your data to establish a Shared Business Vocabulary (SBV). The SBV should be defined and stored in a business glossary.

- Semantic data standardisation using a Shared Business Vocabulary
- SBV vs. taxonomy vs. ontology
- The role of a SBV in MDM, RDM, SOA, DW and data virtualisation
- Approaches to creating an SBV
- Enterprise Data Models & the SBV
- Business glossary products
 - ASG, Collibra, Global IDs, Informatica, IBM InfoSphere Governance Directory, SAP Information Steward Metapedia, SAS Business Data Network

- Planning for a business glossary Organising data definitions in a business glossary
- Business involvement in SBV creation
- Using governance processes in data standardisation

4. Organising the Data Lake

This session looks at how to organise data to still be able to manage it in a complex data landscape. It looks at zoning, versioning, the need for collaboration between business and IT and the use of an information catalog in managing the data

- Organising data in a distributed data reservoir
- Data ingestion zones, data exploration zones, data archive zones, trusted refined data zones
- New requirements for managing data in a distributed data environment
- Collaboration
- Hadoop as a staging area for enterprise data cleansing and integration
- Beyond structured data - from business glossary to information catalog
- Information catalog technologies e.g. Waterline Data, Alation, Informatica 'Project Sanoma' Live Data Map, IBM Information Governance Catalog
- The power of a graph database for storing metadata – dynamic tracking of data and data relationships in real-time
- The semantic web INSIDE THE ENTERPRISE – dynamics taxonomies of data in a distributed data reservoir

5. The Data Refinery Process

This session looks at the process of discovering where your data is and how to refine it to get it under control.

- Implementing systematic disparate data and data relationship discovery
- Data Discovery tools Global IDs, IBM InfoSphere Discovery Server, Informatica, Silwood, SAS
- Automated data mapping
- Data Quality Profiling

- Best Practice Data Quality metrics
- Key approaches to data integration – Data virtualisation, data consolidation and data synchronisation
- Generating Data Cleansing and integration services using common metadata
- Taming the distributed data landscape using enterprise Data Cleansing and integration
- The Enterprise Data Refinery - Hadoop as a staging area for enterprise Data Cleansing and Integration
- Data Provisioning – Provisioning consistent information into Data Warehouses, MDM systems, NoSQL DBMSs and transaction systems
- Achieving consistent data provisioning through reusable data services
- Provisioning consistent refined data using data virtualisation and on-demand information services
- Smart provisioning and governance using rules-based data services
- Provisioning consistent on-demand information services using data virtualisation
- Achieving consistent Data Provisioning in a SOA
- Consistent Data Management across Cloud and on-premise systems
- Data Entry – Implementing an enterprise Data Quality firewall
 - Data Quality at the keyboard
 - Data Quality on inbound and outbound messaging
 - Integrating Data Quality with Data Warehousing & MDM
 - On-demand and event driven Data Quality Services
- Monitoring Data Quality using dashboards
- Managing Data Quality on the Cloud

6. Refining Big Data & Data For Data Warehouses

This session looks at how EIM methodology and processes can be applied to managing, governing and provisioning data in a Big Data analytical ecosystem and in traditional Data Warehouses. How do you deal with very large data volumes and different varieties of data? How does loading data into Hadoop differ from loading data into a Data Warehouse? What about NoSQL databases? How should low-latency data be handled? Topics that will be covered include:

- Types of Big Data
- Connecting to Big Data sources, e.g. Web logs, clickstream, sensor data, unstructured and semi-structured content
- The role of Information Management in an extended analytical environment
- Supplying consistent data to multiple analytical platforms
- Best Practices for integrating and governing multi-structured and structured Big Data
- Dealing with Data Quality in a Big Data environment
- Loading Big Data – What’s different about loading Hadoop files versus NoSQL and analytical relational databases
- Data Warehouse offload – Using Hadoop as a staging area and data refinery
- Governing data in a Data Science environment
- Data Wrangling tools for Hadoop
- Mapping Discovered Data of value into your DW and Business Vocabulary
- Big Data protection

7. Information Audit & Protection – The Forgotten Side of EIM

Over recent years we have seen many major brands suffer embarrassing publicity due to data security breaches that have damaged their brand and reduced customer confidence. With data now highly distributed and so many technologies in place that offer audit and security, many organisations end up with a piecemeal approach to information audit and protection. Policies are everywhere with no single view of the policies associated with securing data across the enterprise. The number of administrators involved is often difficult to determine and regulatory compliance is now demanding that data is protected and that organisations can prove this to their auditors. So how are organisations dealing with this problem? Are data privacy policies enforced everywhere? How is data access security co-ordinated across portals, processes, applications and data? Is anyone auditing privileged user activity? This session defines this problem, looks at the requirements needed for Enterprise Data Audit and Protection and then looks at what technologies are available to help you integrate this into your EIM strategy.

- What is Data Audit and Security and what is involved in managing it?
- Status check - Where are we in data audit, access security and protection today?
- What are the requirements for enterprise data audit, access security and protection?
- What needs to be considered when dealing with the data audit and security challenge?
- What about privileged users?
- What technologies are available to tackle this problem? – IBM Optim and InfoSphere Guardium, Imperiva, EMC RSA, Cloudera Gazzang, Hapache Knox
- How do they integrate with Data Governance programs?
- How to get started in securing, auditing and protecting your data

ABOUT THIS SEMINAR

Today, with most people connected to the Internet, the power of the customer is almost limitless. The Internet has given them freedom to choose in a way that business could never have imagined. They can browse your competitors' web sites with ease. They can compare prices, they can view sentiment about your business, and they can switch loyalty in a single click any time anywhere all from a mobile device. In addition, the emergence of social media sites means that customers also have a voice. They can express opinion and sentiment about products and brands on Twitter Facebook, and review web sites and create social networks by attracting followers, and following others. For many CEOs, customer retention, loyalty, service and growth are top of their agenda. In addition improving operational effectiveness is also high on their priority list. The only way they can achieve this is to acquire more data. CMOs also want access to new data to enrich what they already know about customers. New data is needed to provide insight on customer on-line behaviour for better segmentation and to understand the value of a customers' social network and not just the customer. In addition, COOs want more data to become more effective in operations. Instrumentation is therefore being added so that operations can capture new data. With so much demand we are now in an era where data has never before been so important to business in helping to create competitive advantage. This 2-day seminar looks at the need to capture new data sources and perform exploratory analytics for competitive advantage. It looks at new data sources and at data discovery and visualisation tools to show how these technologies can be used to provide new insight that helps foster growth, reduce costs and improve effectiveness.

AUDIENCE

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

LEARNING OBJECTIVES

Attendees to this seminar will learn:

- How data and analytical characteristics can dictate the approach taken and tools needed to conduct exploratory analytics
- To distinguish between the types of data discovery and visualisation tools on the market
- How to distinguish data discovery and visualisation tools from other BI tools
- Tools and techniques for discovery, analysis and visualisation of multi-structured data
- Build reports and dashboards quickly and easily without the need for IT
- How to publish insights for others to access over the web and mobile devices

OUTLINE

1. An Introduction to Data Exploration, Discovery and Visualisation

This session introduces the relatively area of Data Discovery and Visualisation and looks at why businesses now need.

- New data sources - Structured versus multi-structured data
- What are the different analytical workloads that dictate the need for Data Discovery and Visualisation?
- The Data Discovery and Visualisation process
- What is exploratory analysis?
- What is Data Discovery and Visualisation?
- Why do businesses need this new capability? - Example use cases
- Skills required for Data Discovery and Visualisation
- Types of Data Discovery and Visualisation tools?

2. Deploying and Using Data Discovery and Visualisation Tools

Data Discovery and Visualisation tools are frequently sold into business departments so that local business analysts can start building their own BI applications without having to wait for IT. These new tools offer the attraction of agile development and much faster time to value. When business areas buy them it often means that development starts without any IT guidance and quickly spreads to other parts of the business with little thought for integration or re-use. The result is that inconsistency and chaos can quickly set in. This session looks at best practices in deploying Data Discovery and Visualisation tools to maximise business benefit through data management, re-use and integration with existing BI/DW environments to facilitate consistency.

- The Data Discovery and Visualisation tools marketplace – Tableau, SAS Visual Analytics, SAP Lumira, Platfora, MicroStrategy Visual Intelligence, Qlik, Information Builders WebFOCUS Visual Discovery, etc.
- Key features of Data Discovery and Visualisation tools
- Automated charting, visual exploration and analysis and advanced visualisation

- Automated data discover versus manual data discovery
- Outside-in Versus Inside-out BI application development
- Personal and team based self-service development
- Key requirements for successful self-service BI development using Data Discovery and Visualisation
- Best practice steps in deploying self-service BI applications
 - Simplifying data access and understanding via data management, data governance and information services
 - Removing complexity of data access using data virtualisation
 - Steps to developing self-service BI applications
 - Types of self-service analytical processing
 - Using templates and components for rapid self-service BI application development
 - Ensuring aggregate consistency
 - Prototyping and bookmarking valuable insight
 - Simplifying information delivery and making content easy-to-consume
 - Building report components and dashboards
 - Publishing dashboards and self-service BI applications for business use
 - Handing over self-service applications to IT for hardening
 - Securing access to dashboard based self-service BI applications

3. Getting Started with Predictive Analytics and Machine Learning

As we move into the era of smart business, looking back in time is not enough to make good decisions. Companies have to also model the future to forecast and predict so that they can anticipate problems and act in a timely manner to compete. Predictive Analytics is a therefore a key part of any BI initiative and should be integrated into analysis, reporting and dashboards. This session introduces Predictive Analytics and how shows how it can be used in analysis and in business optimisation.

- What is Predictive Analytics?
- Technologies and methodologies developing Predictive Analytical models

- Using supervised learning to develop predictive models for automatic classification
- Popular predictive algorithms, e.g. Linear regression, decision trees, random forest, neural networks
- Implementing in-Hadoop, in-memory analytics using Spark and SAS LASR server
- Data Science Workbooks using Databricks Cloud and Apache Zeppelin
- Accessing data in HDFS using SQL to build models
- Accessing in Hadoop machine learning algorithms from data mining tools
- Deploying Predictive Analytical models in analytical databases and in Hadoop
- Integrating Predictive Analytics with event stream processing for automated analysis of high velocity events in every-day business operations
- Accessing Predictive Analytics from self-service BI tools and spread sheets
- Clustering data using unsupervised learning algorithms

4. Exploratory Analytics for Multi-Structured Data

This session looks at emerging analytical technologies for multi-structured data and explores how you can use them to improve business insight. Not all analytical projects are implemented using relational database technology, especially when it comes to very large data volumes with unstructured content, sensor data, and clickstreams. This session looks at the emergence of Big Data Analytics using NoSQL Platforms like Hadoop. It looks at the approaches to analysing complex unstructured and social content and the challenges of creating valuable business insight from multiple sources of unstructured content.

- Techniques for producing insight from unstructured content
- Tools and techniques for analysing text
- Voice of the customer and social Media analytics
- Examples of content analytics products in the marketplace
- Clickstream analysis
- Streaming analytics
- Graph analysis

5. Search, BI & Big Data

This session will examine the growing role of Search in an analytical environment both as an information consumer tool for self-service BI and as a way of analysing both structured and unstructured data. Search has been incorporated into BI tools for some time, but with the emergence of Big Data as a platform for analysing unstructured information, it is taking on a major new role. Search is a simple mechanism that is familiar to most people, and opening up the interactive use of BI via Search can have enormous business benefits. Search can be used to grow the use of BI to a much wider group of users and also provide a way to extract additional insight from unstructured content. Topics that will be covered include:

- Why Search and BI?
- The growing importance of analysing unstructured content
- The implications of Big Data on Search and BI
- Creating Search indexes on multi-structured data
- Building dashboards and reports on top of Search engine indexed content
- Using Search to analyse multi-structured data
- The integration of search with traditional BI platforms
- Using Search to find BI content and metrics
- Guided analysis using multi-faceted search
- The search based analytical tools marketplace: Apache Solr (Lucene), Attivio, Connexica, HP IDOL, IBI WebFocus Magnify, IBM Watson Explorer, Microsoft, Oracle Endeca, Quid, SAP Lumira, Splunk

INFORMATION

<p>PARTICIPATION FEE</p> <p>Building an Enterprise Data Lake for Enterprise Data as a Service € 1300</p> <p>Predictive and Advanced Analytics for BI Professionals and Business Analysts € 1300</p> <p>Special price for the delegates who attend both seminars: € 2500</p> <p>The fee includes all seminar documentation, luncheon and coffee breaks.</p> <p>VENUE</p> <p>Roma, Residenza di Ripetta Via di Ripetta, 231 Rome (Italy)</p> <p>SEMINAR TIMETABLE</p> <p>9.30 am - 1.00 pm 2.00 pm - 5.00 pm</p>	<p>HOW TO REGISTER</p> <p>You must send the registration form with the receipt of the payment to: TECHNOLOGY TRANSFER S.r.l. Piazza Cavour, 3 - 00193 Rome (Italy) Fax +39-06-6871102</p> <p>within April 17, 2017</p> <p>PAYMENT</p> <p>Wire transfer to: Technology Transfer S.r.l. Banca: Cariparma Agenzia 1 di Roma IBAN Code: IT 03 W 06230 03202 000057031348 BIC/SWIFT: CRPPIT2P546</p>	<p>GENERAL CONDITIONS</p> <p>DISCOUNT</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.</p> <p>Those who benefit of this discount are not entitled to other discounts for the same seminar.</p> <p>CANCELLATION POLICY</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>CANCELLATION LIABILITY</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

BUILDING AN ENTERPRISE DATA LAKE FOR ENTERPRISE DATA AS A SERVICE

Rome May 2-3, 2017
Residenza di Ripetta - Via di Ripetta, 231
Registration fee: € 1300

PREDICTIVE AND ADVANCED ANALYTICS FOR BI PROFESSIONALS AND BUSINESS ANALYSTS

Rome May 4-5, 2017
Residenza di Ripetta - Via di Ripetta, 231
Registration fee: € 1300

BOTH SEMINARS

Special price for the delegates who attend both seminars: € 2500

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

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

