

TECHNOLOGY TRANSFER PRESENTS

MIKE FERGUSON

Big Data and Analytics

**From Strategy
to Implementation**

JUNE 5-6, 2017

Enterprise

**Data Governance
& Master**

Data Management

JUNE 7-8, 2017

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



info@technologytransfer.it
www.technologytransfer.it

ABOUT THIS SEMINAR

This new two day workshop is aimed at getting Data Scientists, Data Warehousing and BI Professionals up to scratch on Big Data, Hadoop, other NoSQL DBMSs and Multi-Platform Analytics. What is this new phenomenon? How can you make use of it? How does it fit within our traditional analytical environment? What skills do you need to develop for Big Data? All of these questions are addressed in this new knowledge packed workshop.

AUDIENCE

- IT Directors
- CIO's
- IT Managers
- BI Managers
- Data Warehousing Professionals
- Data Scientists
- Enterprise Architects
- Data Architects

LEARNING OBJECTIVES

Attendees to this seminar will learn:

- What Big Data is
- How Big Data creates several new types of analytical workload
- The different Big Data technology platforms
- Big Data Analytical techniques and front-end tools
- How to analyse un-modelled, multi-structured data using Hadoop and MapReduce & Spark
- How to integrate Big Data with traditional Data Warehouses and BI systems
- How to clearly understand business use cases for different Big Data technologies
- How to set up and organise Big Data projects
- How to make use of Big Data to deliver business value

OUTLINE

1. An Introduction to Big Data

This session defines Big Data and looks at business reasons for wanting to make use of this new area of technology. It looks at Big Data use cases and what the difference is between traditional BI and Data Warehousing versus Big Data.

- What is Big Data?
- Types of Big Data
- Why analyse Big Data?
- The need to analyse new more complex data sources
- Industry use cases – Popular Big Data analytic applications
- What is Data Science?
- Data Warehousing and BI Versus Big Data
- Popular patterns for Big Data technologies

2. An Introduction to Big Data Analytics

This session looks at Big Data Analytics, the tools and techniques involved in it and how you can integrate this new environment into an existing DW/BI environment to enrich business insight. It also looks at how to get more value out of existing Data Management tools and BI tools across DW and Big Data platforms.

- Types of Big Data analytical workloads
- Streaming data analytics at high velocity
- Exploratory analysis for multi-structured data
- Complex analysis of structured data
- Graph analytics
- Challenges when managing and analysing Big Data
- Key components in a Big Data Analytics environment
- Preserving existing BI/DW investments
- The Big Data Extended Analytical Ecosystem

3. Big Data Platforms and Storage Options

This session looks at Platforms and Data storage options for Big Data Analytics.

- The new Multi-Platform Analytical Ecosystem
- Beyond the Data Warehouse: Hadoop NoSQL and analytical RDBMSs, NewSQL DBMSs
- NoSQL DBMSs

- An introduction to Hadoop and the Hadoop Stack
- HDFS, MapReduce, Pig & Hive
- Apache Spark Framework
- SQL on Hadoop options
 - Impala, Hive, SparkSQL, HawQ, HP Vertica SQL on Hadoop, IBM BigSQL, CitusDB JethroData, Splice Machine, Actian Analytics Platform, Oracle Big Data SQL, Teradata QueryGrid
- The Big Data Marketplace
 - Hadoop distributions – Cloudera, HortonWorks, MapR, IBM BigInsights Open Data Platform, Microsoft HD Insight, PivotalHD
 - Big Data Appliances – Oracle Big Data Appliance, IBM PureData System for Hadoop, HP HAeN, Teradata Aster Discovery Server
 - NoSQL databases e.g. Datastax, Neo4J, Cassandra, MongoDB, Riak
- The Cloud deployment option – Microsoft Windows Azure (HDInsight Data Lake & Data Factory), IBM Bluemix, Amazon Elastic MapReduce, Altiscale Data Cloud, Qubole, Oracle Analytics Cloud
- Creating a Multi-Platform analytical ecosystem

4. Big Data Integration and Governance in a Multi-Platform Analytical Environment

This session will look at the challenge of dealing with the integration of Big Data and the unique issues it raises. 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 analytical relational databases? What about NoSQL databases? How should low-latency data be handled?

- Types of Big Data
- Connecting to Big Data sources, e.g. Web logs, clickstream, sensor data, unstructured and semi-structured content
- Supplying consistent data to Multiple Analytical Platforms
- Loading Big Data – what's different about loading HDFS, Hive & NoSQL Vs analytical relational databases
- Change data capture – What's possible
- Data Warehouse offload
- Tools for ELT processing on Hadoop – The Enterprise Data Refinery

- ETL tools Vs Pig Vs self-service DI/DQ
- Dealing with Data Quality in a Big Data environment
- Parsing unstructured data
- Governing data in a Data Science environment
- Joined up analytical processing from ETL to analytical workflows
- The impact of data scientist and end user self-service DQ/DI – Paxata, Trifacta, MS Excel, MicroStrategy, Tableau
- Mapping discovered data of value into your DW and business vocabulary
- Big data audit, protection and security – Cloudera Sentry, Dataguise, Hortonworks Ranger, IBM Guardium, Protegrity

5. Tools and Techniques for Analysing Big Data

This session looks at tools available for both data Scientists and also traditional DW/BI Professionals. It looks how both types of developers can exploit Big Data platforms such as Hadoop and NoSQL databases using programming techniques and traditional BI tools as well as how vendors are making it easier to gain access both the NoSQL/Hadoop world and the Analytical RDBMS world by using data virtualisation.

- Data Science projects
- Creating Sandboxes for Data Science projects
- Options for analysing unstructured content – Text analytics, custom MapReduce code and MapReduce developer tools
- Using R as an analytical language for Big Data
- Text analysis and visualisation, Sentiment analysis and visualisation
- Clickstream analysis and visualisation
- Analysing Big Data using MapReduce and applications for Hadoop, e.g. ClearStory Data, Datameer, FICO Karmasphere, Platfora
- Exploratory graph analysis and visualisations
- Using search to analyse multi-structured data
 - Creating search indexes on multi-structured data
 - Building dashboards and reports on top of search engine indexed content
 - The integration of search with traditional BI platforms
 - Guided analysis using multi-faceted search
 - The marketplace: Apache Solr, Attivio, Cloudera Search, Connexica, DataRPM, HP IDOL, IBI WebFocus Magnify, IBM Watson Explorer, LucidWorks, Microsoft, Oracle Endeca Quid, Splunk

- Analysing Big Data using Self-Service BI Tools, e.g. Dell Statistica, Excel, IBM Watson Analytics, Tableau, Qlik, RapidMiner, TIBCO Spotfire, SAS Visual Analytics, MicroStrategy, SAP Lumira, Zoomdata
- Big Data analytics – query performance enablers
- Managing stream computing in a Big Data environment
- Tools and techniques for streaming analytics

6. Integrating Big Data Analytics into the Enterprise

This session looks at how new Big Data platforms can be integrated with traditional Data Warehouses and Data Marts. It looks at Stream Processing, Hadoop, NoSQL databases, Data Warehouse appliances and shows how to put them together to maximize business value from Big Data.

- Integrating Big Data Platforms with traditional DW/BI environments – What's involved
- Integrating Stream Processing with Hadoop and Analytical DW Appliances
- Integrating Hadoop with DW Appliances and Enterprise Data Warehouses
- Tying together front end tools
- Options for implementing Multi-Platform Analytics
- Cross-Platform Analytical workflows
- The role of Data Virtualisation in a Big Data environment
- Multi-Platform optimisation

ABOUT THIS SEMINAR

This two-day in-depth seminar is intended for Chief Data Offices, Enterprise Architects, Data Architects, Master Data Management Professionals, Business Professionals, Database Administrators, Data Integration Developers, and Compliance Managers who are responsible for Management and Governance of Enterprise Data.

The seminar takes a detailed look at the business problems caused by poorly managed data including inconsistent identifiers, data names and policies, poor data quality, poor information protection, and piecemeal project oriented approaches to data integration. It also defines the requirements that need to be met for a company to confidently define, manage and share reference data, master data and transactional data across operational and analytic applications and processes both on-premise and in the Cloud.

Having understood the requirements, you will learn what should be in an Data Management strategy and what you need in terms of people, processes, methodologies, and technologies to bring your data under control. In addition we will look at how to introduce Governance into different Data Management disciplines including Data Naming, Enterprise Metadata Management, Data Modelling, Data Relationship Discovery, Data Profiling, Data Cleaning, Data Integration, Data Service (Data-as-a-service) Provisioning, Reference Data Management and Master Data Management.

During the seminar we take an in-depth look at the technologies needed in each of these areas as well as best practice methodologies and processes for Data Governance and Master Data Management.

LEARNING OBJECTIVES

Attendees will learn how to set up an Enterprise Data Governance program and to determine what technologies they need for Enterprise Data Governance, Data Integration and Master Data Management (MDM). In addition they will learn when to use certain technologies over others and methodologies to use for Metadata Management, Data Integration, and designing and implementing Data Governance and MDM solutions.

WHO SHOULD ATTEND

This seminar is intended for business and IT Professionals responsible for Enterprise Data Governance including 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 at least a high level of understanding of the concepts of Data Migration, Data Replication, Metadata, Data Warehousing, Data Modelling, Data Cleansing etc.

OUTLINE

1. An Introduction To Enterprise Data Governance

This session defines what Enterprise Data Governance is and looks at why companies need to invest in provisioning trusted, commonly understood, high quality information services across the enterprise to guarantee consistency. It also looks at why Data Integration and Data Management should now be a core competency for any organisation.

- An introduction to Enterprise Data Governance
- The impact of unmanaged data on business profitability and ability to respond to competitive pressure
- Is your data out of control?
- Key requirements for Enterprise Data Governance (EDG)
- Establishing a strategy for Data Governance
- Getting the organisation and operating model right
- Key roles and responsibilities - data stewards and data owners
- Formalising EDG processes e.g. the dispute resolution process
- Types of policies needed to govern data
 - Data Integrity rules
 - Data Validation rules
 - Data Cleansing rules
 - Data Integration rules
 - Data Provisioning rules
 - Data Privacy rules
 - Data Access security

2. Data Governance Methodology & Technologies

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

- A best practice step-by-step methodology for Data Governance and Data Management
 - Define, Identify, Assess, Integrate, Provision, Monitor, Protect and Secure

- The Data Management technology platform
- The Data Management Marketplace: Actian, Global IDs, IBM InfoSphere, Informatica, Oracle, SAP, SAS, Talend
- The Data Management platform in your Enterprise Architecture
- Data governance and data management implementation options
 - Centralised, distributed or federated
- The impact of Self-Service BI and Self-Service Data Integration – the need for Data Governance in our business units
- Data Management on-premise and on the Cloud

3. Data Standardisation & The Business Glossary

This session looks at the first step in Data Management – 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.

- Data standardisation using a shared business vocabulary
- SBV vs. taxonomy vs. ontology
- The role of a SBV in Master Data Management, Reference Data Management, 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 Information Governance Catalog, 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
- Enterprise Data Modelling using a SBV

4. Data Discovery, Data Quality Profiling, Cleansing & Integration

Having defined your data, this session looks at the next steps in an a Data Governance methodology, discovering where your data is and how to get it under control.

- Implementing systematic disparate data and data relationship discovery
- Data Discovery tools Global IDs, IBM InfoSphere Discovery Server, Informatica, Silwood, Sypherlink, SAS DataFlux
- 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
- 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

5. Master Data Management Design & Implementation

This session introduces Master Data Management and looks at why businesses are serious about introducing it. It also looks at the components of an MDM and RDM system and the styles of implementation.

- Reference Data vs. Master Data
- What is Master Data Management
- Why is MDM needed? - benefits
- Components of a MDM solution
- How does MDM fit into a SOA?
- MDM implementation options
 - Master Data Synchronisation vs. Virtual MDM.
 - Single Entity Hub vs. Enterprise MDM
- Identifying candidate entities
- Understanding Master Data creation and maintenance
- Master Data implementation
- Defining an SBV for master data entities
- Hierarchy Management
- Master Data Modelling
- Data discovery – identifying where your disparate Master Data is located
- Mapping your disparate Master Data
- Profiling disparate Master Data to understand Data Quality
- Creating trusted Master Data entities using data cleaning and data integration
- Implementing outbound Master Data synchronisation
- Identifying and re-designing Master Data business processes
- The MDM solution marketplace
 - IBM, Informatica, Kalido, Microsoft, Oracle, SAP, SAS DataFlux, Stibo, Talend, Teradata, Tibco
- Evaluating MDM products
- Integration of MDM solutions with Data Management platforms
- Implementing MDM on Hadoop e.g. IBM Big Match and MDM Server
- NoSQL Graph DBMSs and MDM
- Integrating MDM with Enterprise Portals
- Sharing access to master data via master data services in a Service Oriented Architecture (SOA)
- Leveraging SOA for data synchronisation

- Integrating MDM with operational applications and process workflows
- Using Master Data to tag unstructured content

6. Transitioning to Enterprise MDM – The Change Management Process

This session looks at the most difficult job of all – the Change Management process needed to get to Enterprise Master Data Management. It looks at the difficulties involved, what really needs to happen and the process of making it happen.

- Starting a MDM Change Management program
- Changing data entry system data stores
- Changing application logic to use shared MDM services
- Changing user interfaces
- Leveraging portal technology for user interface re-design
- Leveraging a Service Oriented Architecture to access MDM shared services
- Changing ETL jobs to leverage Master Data
- Hierarchy Change Management in MDM and BI systems
- Transitioning from multiple data entry systems to one data entry system
- Transitioning change to existing business processes to take advantage of MDM
- Planning for incremental Change Management

7. Information Audit & Protection – The Forgotten Side of Data Governance

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 Data Governance 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
- How do they integrate with Data Governance programs?
- How to get started in securing, auditing and protecting your data

INFORMATION

PARTICIPATION FEE

Big Data and Analytics

€ 1300

Enterprise Data Governance & Master Data Management

€ 1300

Special price for the delegates who attend both seminars:

€ 2500

The fee includes all seminar documentation, luncheon and coffee breaks.

VENUE

Roma, Residenza di Ripetta
Via di Ripetta, 231
Rome (Italy)

SEMINAR TIMETABLE

9.30 am - 1.00 pm
2.00 pm - 5.00 pm

HOW TO REGISTER

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

within
May 22, 2017

PAYMENT

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

GENERAL CONDITIONS

DISCOUNT

The participants who will register 30 days before the seminar are entitled to a 5% discount.

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.

CANCELLATION POLICY

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.

CANCELLATION LIABILITY

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.

MIKE FERGUSON

BIG DATA AND ANALYTICS

Rome June 5-6, 2017
Residenza di Ripetta - Via di Ripetta, 231
Registration fee: € 1300

**ENTERPRISE DATA GOVERNANCE
& MASTER DATA MANAGEMENT**

Rome June 7-8, 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

