



TECHNOLOGY TRANSFER PRESENTS

Rome, June 21-22, 2018  
Residenza di Ripetta  
Via di Ripetta, 231

INTERNATIONAL  
SUMMIT  
2018

# DATA & ANALYTICS

## PRE E POST-SUMMIT WORKSHOPS

- Enterprise Data Governance & MDM
- Data Vault Modeling Workshop

For many years companies have been extracting operational transaction data from OLTP systems into classic Data Warehouses for query, reporting and analysis to help them make decisions in a relatively passive way. However, in recent years things have been changing quite dramatically with organisations now demanding more data and the use of analytics in almost every part of the business. There are several reasons for this. One of the most important is the fact that the mobile device has made prospects and customers very powerful. Everywhere you go, people are looking down at a mobile device that is almost always connected to the Internet. The result is that loyalty has become cheap and companies are now having to work harder to keep their customers by collecting as much data as possible to understand them better and using it to personalize products and services to meet their needs.

At the same time, Digital Transformation is changing the way companies operate. Prior to digitalization customers, suppliers and partners interacted with employees who then used applications to transact business. In the new digital enterprise customers, suppliers and partners are interacting with organizations via self-service mobile applications and on-line computer systems. Therefore, the chance to engage them is very limited. Also, companies are deploying sensors in operations to collect data to enable them to improve efficiency and reduce cost.

The bottom line is that Data and Analytics have shifted to the centre of the enterprise. Data is now coming in everywhere and the need for analytics is in every part of the business.

New NoSQL analytical platforms have emerged, we have gone from batch based data extraction, a Data Warehouse and passive analysis to a more complex world of multiple analytical systems and workloads. In addition, there is massive pressure to manage, govern and integrate data and across these systems and to simplify access to delivering integrated actionable insights into to every part of the business.

- How can this be achieved?
- How can you do this in a hybrid computing environment when you have to manage data across multiple Cloud and on-premises systems?
- How can you improve agility while also ensuring data is governed to remain compliant with new EU legislation like GDPR?
- How can you improve design and data modelling to deliver more value?
- How do you leverage new technologies like Machine Learning, Artificial Intelligence and advanced Analytics while simplifying access so that self-service BI users can take advantage of it all?

In this, Enterprise Data and Analytics Summit 2018, we address these issues. We will look at the technologies, techniques and architectures needed and how to integrate analytics and insights across analytical systems to deliver competitive advantage.

## Day 1

### Session 1

#### Chairman's Introduction - Building an Intelligent Enterprise in a Hybrid Computing Environment

Mike Ferguson

In this short introductory session, we will set the scene for the Summit in a year where Data Governance and Data Management dominates the European market. The session will define a vision for Intelligent Business with data and analytics at the centre of the enterprise connected into every part of the business. It will look at the challenges ahead both in terms of data and analytics in an environment made more complex by more data sources, more data stores and the adoption of hybrid computing as Cloud adoption accelerates in European enterprises. It will look at data and compliance, Cloud adoption, modern Data Warehouse design and the need to manage both data and analytics in a seamless way while remaining Agile in a hybrid computing set-up.

### Session 2

#### Governing Data Across Multiple Data Stores - The Critical Importance of an Information Catalog

Mike Ferguson

With so much new data being captured across the enterprise and multiple self-service and Data Science initiatives being undertaken, something has to know and track what's going on and what's available in an increasingly complex data landscape. In addition, people need the ability to publish what data and what artefacts (ETL jobs, data preparation jobs, analytical models, dashboards, etc) currently exist to encourage re-use and prevent re-invention. This session shows how information catalogue software can be used to publish data and artefacts to manage and organise a multi-platform analytical environment.

- What is an Information Catalogue and why do you need one?
- What does an Information Catalogue do?
  - Information Catalogue capabilities, e.g. automatic data profiling, automatic tagging and data classification, automatic data indexing, faceted search, data marketplaces, artefact publishing
- How can it help in sharing data and analytics?
- Information Catalogue technology offerings
- How does a Information Catalogue help govern a Data Lake?
- Creating a governed information value chain using an Information Catalogue
- Key roles and responsibilities - Information producers, information consumers and governance

### Session 3

#### Understanding the GDPR

Daragh O'Brien

2018 is the year that the European Union General Data Protection Regulation (GDPR) becomes law across all EU member countries but many companies are unlikely to be ready by the activation date in May. This session gives delegates an overview of the GDPR from a metadata and data governance perspective.

- What is personal data - understanding the importance of the context of processing (and why it's more than just names and addresses)
- Understanding Consent Management in the context of MDM and Metadata - right data, right purpose, right permission, at the right time
- Running a Data Privacy Programme (some do's and don'ts)
- Taming the data jungle (governing Data Privacy across Data Lakes, Data Warehouses, and other technologies)
- The role of Data Management in GDPR Risk Management
- The emerging role of AI and Deep Learning in data discovery and operational processes
- Data Virtualisation as a Privacy Enhancing Tech-

nology

- The Vendor Landscape - What is out there? What to watch for "Lawyer tools" vs "Data Tools"
- The challenges of the Cloud...
- The Post-GDPR landscape - how to survive
- KPIs to consider monitoring
- Using analytics to detect non-compliance
- The marketing and analytics landscape post-GDPR
- The future of analytics and marketing post GDPR and the "Legitimate Interests" basis
- The ePrivacy Regulation
- What do you need to do to do electronic marketing in a way that won't get you in trouble post GDPR...

## Session 4

### Big SQL Solutions for Big Data Systems

Rick van der Lans

There is a tendency to think that big data systems can't be developed with SQL technology. Wrong! Development of SQL technology hasn't stood still the last 10 years. On the contrary, many analytical SQL database servers (e.g. Aster, SnowflakeDB, Edge Intelligence) are more than suited for analyzing big databases, NewSQL products (e.g. Clustrix, VoltDB) can handle massive transactional systems with serious data ingestion rates, and GPU-based SQL databases (e.g. BlazingDB, MapD, Kinetica) are optimized to run complex forms of analytics on massive amounts fast. In addition, SQL technology is combined with Hadoop increasingly. The competitive market of SQL-on-Hadoop engines pushes development, even the combination of the new Hadoop file system called Kudu has created new Big Data opportunities for SQL. And don't forget the SQL-based streaming database server allowing analytics at the edge. In this session, the market of new SQL technologies developed for storing, processing, and analyzing Big Data are discussed. The limitations, use cases, pros, and cons of all these SQL solutions are critically dissected.

## Session 5

### Machine Learning in the Enterprise

Jen Underwood

Despite the unprecedented speed and ease of creating predictive models today, the human mind is still essential for generating good models. From selecting the right solution architecture and right problems to solve to preventing algorithm bias, Machine learning is still an art and a science. To reap the benefits of machine Learning across the enterprise, we will walk-through the entire machine learning project life-cycle showcasing real-time, Big Data, Cloud and in-database analytical environments. We'll explore common use cases, technical architecture options, typical mistakes and proven best-practices to help you build and operationalize better models.

## Day 2

## Session 6

### Data Modelstorming: From Business Models to Analytical Models

Lawrence Corr

Have you ever been disappointed with the results of traditional data requirements gathering, especially for BI and data analytics? Ever wished you could 'cut to the chase' and model the data with the people who know it and want to use it. But that's not a realistic alternative, is it? Business people don't do data modeling! What if that wasn't the case.

This session discusses collaborative modeling techniques - popularized in books such as 'Business Model Generation' and 'Agile Data Warehouse Design' - for successfully engaging stakeholders using BEAM (Business Event Analysis and Modeling) and the Business Model Canvas for value-driven BI requirements gathering and star schema design. Learn how visual thinking, narrative, 7Ws and lots of Post-it (tm) notes can get your BI customers thinking dimensionally and capturing their own data requirements with agility.

## Session 7

### Practical Analytics Innovation

Jen Underwood

This session will explore a wide variety of useful advancements happening across the analytics ecosystem. It will discuss key transformational trends, explore "blue ocean" opportunities in a new world of data, and inspire you to try something new. Topics include but are not limited to Analytics Automation, Natural Language, Citizen Data Science, Real-Time, Cloud, Hybrid and Immersive Analytics Experiences. Don't miss the demos in this thought-provoking session that illustrate how easy it is to get started enjoying these innovations to gain a competitive advantage.

## Session 8

### Dimensional Design Pattern Recognition and Usage: Advanced Star Schema Modeling

Lawrence Corr

Design patterns are used to represent some of the best practices adopted by experienced modelers and developers. A pattern is a template which systematically names and explains a general design that addresses a recurring problem. It describes the problem (motivation for the pattern), the solution, when to apply the solution with hints and examples, and its consequences (when not to apply it).

**BEAM** (Business Event Analysis and Modeling) not only provides a simple 7W (who, what, when, where, how many, why and how) script for collaborative modeling, it also offers a powerful way of categorising dimensional design patterns by type of data to help you recognise their applicability and adapt them to your BI use cases.

In this session Lawrence Corr uses the BEAM canvas, memorable data values and worked examples of client billing and website visit analytics to bring alive three powerful design patterns for solving common customer/employee and product/service relat-

ed BI problems:

- **Hybrid Slowly Changing Dimensions** - Supporting conflicting requirements to provide historically correct (type 2 SCDs), current (type 1) and point-in-time (e.g. Year End) descriptive values
- **Multi-level Dimension** - Measuring business events with naturally varying levels of descriptive detail

**(Variable-Depth) Hierarchy Map** – Querying recursive customer relationships and organisation structures; product and service decomposition for component (bill of materials) and product unbundling analysis

## Session 9

### Integrated Data and Analytics or Unmanaged Silos?

Panel with Speakers and Vendors

With the increasingly complex data landscape of thousands of data sources, multiple Cloud and on-premises analytical data stores, the adoption of self-service data preparation, self-service BI and self-service Data Science, what chance has any enterprise of co-ordinating and integrating initiatives to keep things integrated while remaining compliant with tough regulations? Will the pressures of autonomous business units, data complexity and self-service result in un-coordinated silos and missed opportunity or is there a way to keep it all integrated to benefit the enterprise as a whole? This Panel session will discuss this complex challenge.

## Session 10

### Data Management in a Cloud Computing Environment

Mike Ferguson

As the adoption of Cloud Computing continues to grow and we are now at the point where many companies may have deployed applications both off-premise on public Clouds and on-premises on private Clouds.

They may even be using off-premise infrastructure to extend their private Cloud environments. As this investment grows, there is now a demand to seamlessly manage and govern data in a consistent way irrespective of its location in a Cloud Computing and hybrid environment.

This session looks in detail at the challenge of consistently managing data in a Cloud Computing environment and explores concerns about the added complexity that off-premises data brings. It looks at what is needed to manage data consistently across off-premises and on-premises systems. In particular it looks at the complexity of APIs to connect to data sources and targets and at important Data Management disciplines such as maintaining Data Privacy, data access security, data quality, data consolidation, data virtualisation, replication, Master Data Management and data synchronisation across on-premises and off-premises Clouds and what is possible today. It also looks at integration Platform as a Service (iPaaS), why Cloud based Data Integration Hubs are a good idea and how Data Management platforms can create a 'data fabric' that spans the corporate firewall to manage a Logical Data Lake in a hybrid computing environment. In addition it will highlight problems that still need to be solved to get to a point where companies can confidently and freely manage off-premises and on-premises data in a seamless manner.

- Pros and cons of deploying on the Cloud?
- Deploying systems on public and private Clouds - what are the options
- Options for storing data on the Cloud e.g. Cloud object storage, NoSQL data stores, SQL databases for Data Warehouse, Hadoop and MDM systems in the Cloud
- Moving data between on-premises and Cloud based systems
- Ingesting data in the Cloud – streaming and batch ingestion
- Cloud-based and on-premises data integration solutions, such as, Data Virtualisation

- Data Management technologies for hybrid computing environments - iPaaS, Cloud based Data Integration, Hubs Data Management Platforms
- Managing Data Governance across Cloud and on-premises systems
- Managing Data Privacy and data access security in a hybrid Cloud Computing environment

## Session 11

### Unifying Data Lakes, Data Marketplaces, and Data Warehouses

Rick van der Lans

To support their Business Intelligence needs, most organizations have developed a classic Data Warehouse system. But for a new group of business users with Data Science and data investigative needs, a new architecture is developed called the Data Lake. Most of the time, Data Lakes are developed as stand-alone systems with almost no relationship with the existing Data Warehouse. Currently, a third system for delivering data to specific business users has been introduced: the data marketplace. And again, organizations are developing this third system that delivers data, a third analytical island. The fourth data delivery system relates to the world of streaming analytics.

Developing all these data delivery systems independently is far from ideal. Development-wise wheels are invented over and again, low productivity, meta data replication, and inconsistencies across reports and analysis are just a few of the many drawbacks of this approach. It's crucial for an organization to somehow bring these developments together. One solution is by deploying a data delivery architecture based on Data Virtualization technology. Such an architecture can support a wide range of business users, ranging from users demanding a highly agile environment, such as the marketplace, up to those who require governable and auditable reports. This session discusses how these environments can be merged into one unified architecture.

# WORKSHOP 1

## Enterprise Data Governance & Master Data Management

Rome 19-20 June, 2018

### Speaker



Mike Ferguson

He is Managing Director of Intelligent Business Strategies Limited. As an analyst and consultant he specialises in Business Intelligence, Analytics, Big Data, and Data Management. With over 33 years of IT experience, Mr. Ferguson has consulted for dozens of companies, spoken at events all over the world and written numerous articles. 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.

**Mike Ferguson is also Chairman of:  
BIG DATA & DATA SCIENCE  
INTERNATIONAL CONFERENCE**

### About This Seminar

Most organisations 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 integrating data. Business users are also getting involved with new self-service data preparation tools. The question is, is this the only way to manage data? Is there another level that we can reach to allow us to more easily manage and govern data across an increasingly complex data landscape consisting of multiple data stores?

This 2-day seminar looks at the challenges faced by companies trying to deal with an exploding number of data sources, collecting data in multiple data stores (cloud and on-premises), multiple analytical systems and at the requirements to be able to define, govern, manage and share trusted high quality information in a distributed and hybrid computing environment. It also explores a new approach of how IT data architects, business users and IT developers can collaborate together in building and managing a Logical data lake to get control of your data. This includes data ingestion, automated data discovery, data profiling and tagging and publishing data in an information catalog. It also involves refining raw data to produce enterprise data services that can be published in a catalog available for consumption across your company. We also introduce multiple data lake configurations including a centralised data lake and a 'logical' distributed data lake as well as execution of jobs and governance across

multiple data stores. It emphasises the need for a common collaborative approach to governing and managing data of all types.

### **What you will learn**

- How to define a strategy for producing trusted data as-a-service in a distributed environment of multiple data stores and data sources
  - How to organise data in a centralised or distributed data environment to overcome complexity and chaos
  - How to design, build, manage and operate a Logical or centralised Data Lake within their organisation
  - The critical importance of an Information Catalog in understanding what data is available as a service
  - How data standardisation and business glossaries can help make sure data is understood
  - An operating model for effective distributed information governance
  - What technologies and implementation methodologies they need to get their data under control.
  - How to apply methodologies to get master and reference data, big data, data warehouse data and unstructured data under control irrespective of whether it be on-premises or in the cloud.
- 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 preparation or both? – data governance or data chaos?
  - Key requirements for data management
    - o Structured data – master, reference and transaction data
    - o Semi-structured data – JSON, BSON, XML
    - o Unstructured data - text, video
    - o 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 scope of your data lake
    - o OLTP system sources
    - o Data Warehouses
    - o Big Data systems, e.g. Hadoop
    - o MDM and RDM systems
    - o Data virtualisation
    - o Streaming data
    - o Enterprise Content Management
  - Building a business case for data management
  - Defining an enterprise data strategy
  - A new inclusive approach to governing and managing data
  - Introducing the data lake and data refinery
  - Data lake configurations – what are the options?
    - o Centralised, distributed or Logical
  - The rising importance of an Information catalog
  - Integrating a data lake into your enterprise analytical architecture

## **Outline**

### **1. Strategy & Planning**

This session introduces the data lake together with the need for a data strategy and looks at the reasons why companies need it. It looks at what should be in your data strategy, the operating model needed to implement, the types of data you have to manage and the scope of implementation. It also looks at the policies and

## 2. Methodology & Technologies

Having understood strategy, this session looks at multiple methodologies and the technologies needed to help apply it to your structured and multi-structured data to bring it under control. It also looks at how platforms like Hadoop and common data services provide the foundation to manage information across the enterprise

- Information production and information consumption
- Data Lake use cases
- The role of data management technology platforms, in managing data across multiple data stores e.g. Cask, Cambridge Semantics, Global IDs, IBM Watson Data Platform, Informatica, Oracle, SAP, SAS, Talend
- A best practice step-by-step methodology structured data governance
- Why the methodology has to change for semi-structured and unstructured data
- Methodologies for structured Vs multi-structured data
- Technology components in the new world of distributed data
- Hadoop as a data staging area and why it is not enough
- Implementation run-time options – the need to execute in multiple environments

## 3. Data Standardisation, the Business Glossary and the Information Catalog

This session looks at the need for data standardisation of structured data and of new insights from processing unstructured data. 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 and is important for information consumers to understand published data in

a Data Lake. It also looks at the emergence of more powerful Information Catalog software and how business glossaries have become part of what a Catalog offers

- Semantic data standardisation using a shared business vocabulary within an information catalog
- The role of a SBV in MDM, RDM, SOA, DW and data virtualisation
- Why is an SBV relevant in a Data Lake and a Logical Data Warehouse?
- Approaches to creating an SBV
- Business glossary products storing SBV business data names Alteryx Connect Glossary, ASG, Collibra, Global IDs, Informatica, IBM Information Governance Catalog, SAP Information Steward Metapedia, SAS Business Data Network, Tibco Information Server
- Planning for a business glossary Organising data definitions in a business glossary
- Key roles and responsibilities - getting the operating model right to create and manage an SBV
- Formalising governance of business data names, e.g. the dispute resolution process
- Business involvement in SBV creation
- Beyond structured data - from business glossary to Information Catalog
- What is an Information Catalog?
- Why are Information Catalogs becoming critical to Data Management?
- Information Catalog technologies, e.g. Alation, Alteryx Connect, Amazon Glue, Apache Atlas, Collibra Catalog, IBM Information Governance Catalog & Watson Data Platform Data Catalog, Informatica Live Data Map, Microsoft Azure Data Catalog, Podium Data, Waterline Data, Zaloni Mica
- Information Catalog capabilities

## 4. Organising and Operating 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 centralised or distributed data lake
- Creating zone to manage data
- New requirements for managing data in centralised and distributed data lakes
- Creating collaborative data lake projects
- Hadoop as a staging area for enterprise data cleansing and integration
- Core processes in Data Lake operations
- The data ingestion process
- Tools and techniques for data ingestion
- Implementing systematic disparate data and data relationship discovery using Information Catalog software
- Using domains and machine learning to automate and speed up data discovery and tagging
- IBM Watson Data Catalog, Informatica CLAIRE, Silwood, Waterline Data Smart Data Catalog
- Automated profiling and tagging and cataloguing of data
- Automated data mapping
- The data classification and policy definition processes
- Manual and automated data classification to enable governance
- Using tag based policies to govern data

## 5. The Data Refinery Process

This session looks at the process of refining data to get produce trusted information

- What is a data refinery?
- Key requirements for refining data

- The need for multiple execution engines to run in multiple environments
- Options for refining data - ETL versus self-service data preparation
- Key approaches to scalable ETL data integration using Apache Spark
- Self-service data preparation tools for Spark and Hadoop e.g. Alteryx Designer, Informatica Intelligent Data Lake, IBM Data Refinery, Paxata, Tableau (Project Maestro), Tamr, Talend, Trifacta
- Automated data profiling using analytics in data preparation tools
- Executing data refinery jobs in a distributed data lake using Apache Beam to run anywhere
- Approaches to integrating IT ETL and self-service data preparation
- Apache Atlas Open Metadata & Governance
- Joined up analytical processing from ETL to analytical workflows
- Publishing data and data integration jobs to the information catalog
- Mapping produced data of value into your DW and business vocabulary
- Data provisioning – provisioning consistent information into data warehouses, MDM systems, NoSQL DBMSs and transaction systems
- Provisioning consistent refined data using data virtualisation, a Logical Data Warehouse and on-demand information services
- Governing the provisioning process using rules-based metadata
- Consistent data management across cloud and on-premise systems

## 6. Refining Big Data & Data For Data Warehouses

This session looks at how the data refining 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 do you load and process data in Hadoop?

How should low-latency data be handled? Topics that will be covered include:

- A walk through of end-to-end data lake operation to create a Single Customer View
- Types of big data & small data needed for single customer view and the challenge of bringing it together
- Connecting to Big Data sources, e.g. web logs, clickstream, sensor data, unstructured and semi-structured content
- Ingesting and analysing clickstream data
- The challenge of capturing external customer data from social networks
- Dealing with unstructured data quality in a Big Data environment
- Using graph analysis to identify new relationships
- The need to combine big data, master data and data in your data warehouse
- Matching big data with customer master data at scale
- Governing data in a Data Science environment

## **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 the same 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 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?
- Automatic data discovery and the information catalog – a huge help in identifying sensitive data
- What about privileged users?
- Using a Data Management platform and Information Catalog to govern data across multiple data stores
- Securing and protecting data using tag based policies in an Information Catalog
- What technologies are available to protect data and govern it? – Apache Knox, Cloudera Sentry, Dataguise, Hortonworks Ranger, IBM (Watson Data Platform, Data Catalog, Optim & Guardium), Imperva, Micro Focus, Privitar
- Can these technologies help in GDPR?
- How do they integrate with Data Governance programs?
- How to get started in securing, auditing and protecting your data

## WORKSHOP 2

### Data Vault Modeling Workshop

Rome June 25-26, 2018

#### Speaker

Hans Hultgren



He is a DWBI, Data Modeling & Big Data Advisor, Trainer & Author. He is President at Genesee Academy, LLC and a Principal at Top Of Minds AB.

Hans works in the areas of Data Modeling, Data Warehousing, Business Intelligence and Big Data as an educator, author, speaker, and advisor. Currently Hans is working on Business Intelligence and Enterprise Data Warehousing (EDW) with a focus on Ensemble Modeling and Data Vault. Hans has spoken at events internationally in the Nordics, EU, USA and AU/NZ. He is primarily located in Stockholm, Amsterdam, and Denver, Colorado. Hans was a professor at the University of Denver for 20 years and while there started a masters of science program in Data Warehousing and Business Intelligence. For the past 10 years, since it was formed, Hans has been working on global training programs at Genesee Academy, LLC. Hans published the Data Modeling book Modeling the Agile Data Warehouse with Data Vault which is available on Amazon websites in both print and Kindle.

#### About this Class

Today the Data Warehouse needs to be Agile. While there are many barriers to achieving Data Warehouse program agility, one of the issues has been the Data Modeling approach. With traditional techniques, the Data Warehouse quickly becomes hardened and difficult to change. Moving to Agile development techniques, based on incremental builds, is almost impossible. Why? Because we ultimately need

to do a great deal of re-engineering of the data structures.

Enter Data Vault (DV). For the past 15 years companies around the world have been using a new Data Modeling technique that greatly improves agility - the Data Vault data modeling approach. The premise behind Data Vault is Unified Decomposition - basically this means we separate the things that change from the things that don't change. How this works: The existence of a person named "Hans" (for example) is always going to be true. So, the instances of Core Business Concepts (CBCs) are placed in their own data structures (Customer Hub, for example). Next the innate relationship that the Customer has with a Sale is also something that is not generally subject to change. This is a business-driven, foundational relationship that we capture in unique table structures (Links).

Since no relationships are embedded they can be added without any re-engineering impact. Lastly the way we describe our CBCs can vary over time, vary by source, vary by type of data and also by rate of change. We use a set of separate tables to capture this context (Satellites). Because new attributes introduced in later iterations can be included in new Satellites, the Data Warehouse can accept new attributes without re-engineering. There are over 1500 Data Vault models in organizations today and the technique is growing rapidly.

The Data Vault modeling pattern is also very applicable to Big Data, Cloud, Virtual and Streaming deployments. Because the context is separated from the way CBCs and Relationships are stored, the context can take any form.

## Outline

This course is delivered through lectures, exercises, and small group modeling cases. Plan to spend 50% time in lectures, 15% time with individual exercises and 35% time with small group cases and interactive discussions.

### DAY 1

1. Welcome & Introductions
2. Data Warehouse Modeling
3. Ensemble Modeling
4. Data Vault Core Constructs
5. Comparing Modeling Methods
6. Tracking History
7. Information Modeling
9. Data Vault Modeling Process
10. First Small Group Case Exercise
11. Continue First Case
12. Class Discussion & Groups Present
13. Link Design & Exceptions
14. SAL & HAL
15. Business Key Topics
16. Disconnected & Reference Tables
17. Data Vault Unit of Work
18. Second Small Group Case Exercise
19. Wrap up for the Day dispute

### DAY 2

1. Q&A from Day One
2. Continue Second Case
3. Class Discussion & Groups Present
4. Satellite Design & Exceptions
5. Expanded UOW & Keyed Instance
6. Modeling Address
7. Bike Shop Example
9. Applying the DV Ensemble
10. Standard and Expanded Attributes
11. Managing Dates/Bi-temporal Data
12. Third Small Group Case Exercise
13. Continue Third Case
14. Class Discussion & Groups Present
15. Architecture: RAW & BDV Layers
16. Hyper Forms and Unstructured
17. Deploying Big Data and Streaming
18. Virtualization
19. Wrap up for the Day

### Attendees to this seminar will learn:

- The foundational pillars of the Data Vault modeling approach
- Through small group modeling exercises learn to translate requirements to DV models
- How to model a Data Warehouse using Data Vault (DV)  
To distinguish between encapsulated (3NF & Dimensional) versus ensemble (DV, etc.) modeling patterns
- To identify modeling scenarios that are best addressed by Data Vault
- To review and critique DV models for best practice compliance & optimal performance



**Mike  
Ferguson**

He is Managing Director of Intelligent Business Strategies Limited. As an analyst and consultant he specialises in Business Intelligence, Analytics, Big Data, and Data Management. With over 35 years of IT experience, Mr. Ferguson has consulted for dozens of companies, spoken at events all over the world and written numerous articles. He teaches classes on Machine Learning and Advanced Analytics, Enterprise Data Lakes, Data Governance and MDM and Big Data. 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.

**Lawrance  
Corr**



He is a leading international BI consultant and former Ralph Kimball Associate. He has worked on Data Warehousing projects in the US, Europe, the Middle East and Africa within healthcare, telecoms, broadcasting, higher education, financial services and retail, helping organizations benefit from simpler, more inclusive requirements modeling techniques. He is the author of Agile Data Warehouse Design: Collaborative Dimensional Modeling, from Whiteboard to Star Schema, an Amazon #1 bestseller in Data Warehousing and database design. Mr. Corr's new book Data Modelstorming: Using BEAM to Design the Data Everyone Wants will be published in Spring 2019.



**Daragh  
O'Brien**

He was recently rated the 24th most influential person in Information Security worldwide on Twitter. Mr. O'Brien, FICS, is a leading consultant, educator, and author in the fields of Information Privacy, Governance, Ethics, and Quality. After over a decade in a leading telco, he now works with clients in a range of sectors on a range of Information Management challenges. Mr. O'Brien is a Fellow of the Irish Computer Society and a Privacy Officer for DAMA-I. He teaches Data Privacy Law and Practice at the Law Society of Ireland. Castlebridge Assois is a commercial partner of the Adapt Centre in Trinity College Dublin and collaborates with the Insight Centre for Digital Analytics, Europe's largest Analytics research group.

## SPEAKERS

**Mike  
Ferguson**

**Daragh O'Brien**      **Lawrance  
Corr**

**Rick  
van der Lans**

**Jen  
Underwood**

# REGISTRATION FORM



Once filled to be given to:  
**Technology Transfer**  
Piazza Cavour, 3 - 00193  
Roma  
Tel. 06-6832227  
Fax 06-6871102  
[www.technologytransfer.it](http://www.technologytransfer.it)  
[info@technologytransfer.it](mailto:info@technologytransfer.it)

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

**PARTICIPATION FEE**

**Data & Analytics SUMMIT 2018**  
Participation fee, Euro 1.500

**Enterprise Data Governance &  
Master Data Management**  
Participation fee, Euro 1.300

**Data Vault Modeling Workshop**  
Participation fee, Euro 1.300

**Discounted shares**

Summit + Enterprise Data  
Governance & Master Data  
Management

Euro 2.700

Summit + Data Vault Modeling  
Workshop: Euro 2.700

Summit + both workshops  
Euro 3.800

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

**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

**PAYMENT**

Wire transfer to: Technology Transfer

first name \_\_\_\_\_

surname \_\_\_\_\_

job title \_\_\_\_\_

organisation \_\_\_\_\_

address \_\_\_\_\_

postcode \_\_\_\_\_

city \_\_\_\_\_

country \_\_\_\_\_

telephone \_\_\_\_\_

fax \_\_\_\_\_

e-mail \_\_\_\_\_  
Stamp and signature \_\_\_\_\_

S.r.l.  
Banca: Cariparma  
Agenzia 1 di Roma  
Iban Code:  
IT 03 W 06230 03202  
000057031348

**within June 6, 2018**

**GROUP 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.

**EARLY REGISTRATION**

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

**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.

**SEMINAR TIMETABLE**

2 days: 9.30 am - 1.00 pm  
2.00 pm - 5.00 pm

**Jen**

**Underwood**



She is founder of Impact Analytix, LLC, is a recognized analytics industry expert. She has a unique blend of product management, design and over 20 years of "hands-on" development of Data Warehouses, reporting, visualization and advanced analytics solutions. In addition to keeping a constant pulse on industry trends, she enjoys digging into oceans of data. Mrs. Underwood is honored to be an IBM Analytics Insider, SAS contributor, former Tableau Zen Master, and active analytics community member.

In the past she has held worldwide product management roles at Microsoft and served as a technical lead for system implementation firms. She has launched new analytics products and turned around failed projects. Today she provides industry thought leadership, advisory, strategy, and market research. She has been mentioned by KD Nuggets, Information Management and Forbes. She writes for InformationWeek, O'Reilly Media, and other tech industry publications.

Mrs. Underwood has a Bachelor of Business Administration - Marketing, Cum Laude from the University of Wisconsin, Milwaukee and a post-graduate certificate in Computer Science - Data Mining from the University of California, San Diego.



**Rick  
van der Lans**

He is a highly-respected independent analyst, consultant, author, and internationally acclaimed lecturer specializing in Data Warehousing, Business Intelligence, Big Data, and database technology. He has presented countless seminars, webinars, and keynotes at industry-leading conferences. He also helps clients worldwide to design their Data Warehouse, Big Data, and Business Intelligence architectures and solutions and assists them with selecting the right products. He has been influential in introducing the new Logical Data Warehouse architecture worldwide which helps organizations to develop more agile Business Intelligence systems. Over the years, Mr. van der Lans has written hundreds of articles and blogs for newspapers and websites and has authored many educational and popular white papers for a long list of vendors. He was the author of the first available book on SQL, entitled including Introduction to SQL, which has been translated into several languages with more than 100,000 copies sold. More recently, he published his book Data Virtualization for Business Intelligence Systems. He presents seminars, keynotes, and in-house sessions on Big Data and Analytics, Data Virtualization, the Logical Data Warehouse, Data Warehousing and Business Intelligence.



**TECHNOLOGY TRANSFER**

Piazza Cavour, 3 - 00193 Roma

Tel. 06.6832227 - Fax 06.6871102

[www.technologytransfer.it](http://www.technologytransfer.it)

[info@technologytransfer.it](mailto:info@technologytransfer.it)