



INTELLIGENT BUSINESS STRATEGIES 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

www.intelligentbusiness.biz
training@intelligentbusiness.biz

OVERVIEW

In today's digital economy, the customer is all powerful. They can switch loyalty in a single click while on the move from a mobile device. The internet has made loyalty cheap and many CEOs want new data to enrich what they already know about customers in order to keep them loyal and offer them a more personalised service. In addition, companies are capturing new data using sensors in to gain sight of what's happening and to optimise business operations. This new data is causing many companies with traditional data warehouses and data marts to realise that this is not enough for analytics. Other systems are needed and with the pace of change quickening, lower latency data and machine learning is in demand everywhere. All of it is needed to remain competitive. So 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 accommodate change, lower data latency and integrate with other analytical workloads to provide a new modern data warehouse for the digital enterprise? This 2-day seminar looks at why you need to do this. It discusses the tools and techniques needed to capture new data types, establish new data pipelines across cloud and on-premises system and how to produce re-usable data assets, modernise your data warehouse and bring together the data and analytics needed to accelerate time to value.

AUDIENCE

CDOs, CIO's, IT Managers, CTOs, Business Analysts, data scientists, BI Managers, data warehousing professionals, enterprise architects, data architects, solution architects

LEARNING OBJECTIVES

After two days attendees will:

- 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 modernise ETL processing by building data products using Data Fabric
- 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 your data warehouse with data stored in cloud storage data lakes
- Understand Lakehouse and open table formats like Delta Lake and Iceberg
- Understand how to modernise your data architecture
- Understand how data warehouse automation tools
- 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 create a common semantic layer for multiple BI and data science tools
- 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

MODULE 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, Spark, 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
 - Big Data analytics on Spark, cloud storage and Hadoop
 - Real-time streaming data analytics
 - Graph analysis in Graph Databases
- New challenges brought about by:
 - Data complexity
 - Data management siloes
 - Managing data in a distributed and hybrid computing environment
 - 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

MODULE 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 re-usable 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

MODULE 3: MODERN DATA MODELLING TECHNIQUES FOR AGILE DATA WAREHOUSING

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

MODULE 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?
 - Data Mesh
 - Offloading staging data to a data lake and use Spark for scalable ELT processing
 - Data Lake Vs Lakehouse
 - The real-time data option
 - Using data warehouse automation software to generate ETL processing
- Pros and cons of these options
- Data architecture implications of modernising and democratising data engineering

MODULE 5: ACCELERATING ETL PROCESSING USING A DATA CATALOG, DATA FABRIC, DATA PRODUCTS AND A DATA MARKETPLACE

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
 - Types of streaming data - IoT data, weblogs, OLTP system change data capture, ...
 - Key technologies for processing streaming data - Kafka, streaming analytics and event stores
 - Turning OLTP change data capture into Kafka data streams
 - Using Kafka as a data source to process data in real-time
 - 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 lakehouse
- Using ETL data pipelines to produce reusable data products for use in your data warehouse, data science and other analytical data stores
- Publishing reusable data products in a data marketplace ready for consumption
- Using Data Science to develop new analytical models to run in your data warehouse

MODULE 6: RAPID DATA WAREHOUSE DEVELOPMENT USING DATA AUTOMATION

In addition 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 Automation?

- Using Data Warehouse Automation Tools for rapid data warehouse and data mart development
 - Generating Data Vault, E/R and Star Schema design
 - Data Pipeline generation
 - Processing streaming data using Data Warehouse Automation
 - Integrating big data with a data warehouse using Data Warehouse Automation
 - Integrating cloud Data Warehouses with data lakes using Data Warehouse Automation
 - Integrating business glossaries with Data Warehouse Automation Tools
 - Using Data Warehouse Automation to migrate data warehouses
 - Using Data Virtualisation to shield existing BI tools from changes in design
- The Data Warehouse Automation Tools market, e.g. IDERA WhereScape, biGENIUS, Qlik Compose, TimeXtender, Varigence BIMLStudio, VaultBuilder & more
- Metadata driven data warehouse maintenance
- Separating storage from compute for elasticity and scalability
- Managing and integrating cloud and on-premises data
- Using Data Fabric software to integrate data in cloud ETL processing, e.g. AWS Glue, Azure Data Factory, Google Cloud Dataplex and Data Fusion, IBM Cloud Pak for Data, Informatica IDMC, Qlik (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
- Migrating DW schema, data, ETL, and security
- Dealing with cloud DW migration issues like data types, SQL differences, privilege differences, data volumes
- Managing access to cloud-based data warehouses
- Integrating cloud based BI tools with on-premise systems

MODULE 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
- Amazon Redshift, Google BigQuery, IBM Db2 Warehouse on Cloud, Microsoft Azure Synapse Analytics, Oracle Autonomous Data Warehouse, Snowflake, SAP Data Warehouse Cloud, Teradata VantageCloud Lake, Kinetica,
- Lakehouses – Apache Iceberg, Databricks Delta Lake, Apache Hudi open table formats

MODULE 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

MODULE 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?

PRESENTER



Mike Ferguson is Managing Director of Intelligent Business Strategies Limited. As an independent IT industry analyst and consultant, he specialises in BI / analytics and data management. With over 40 years of IT experience, Mike has consulted for dozens of companies on BI/Analytics, data strategy, technology selection, data architecture, and data management. Mike is also conference chairman of Big Data LDN, the fastest growing data and analytics conference in Europe and a member of the EDM Council CDMC Executive Advisory Board. He has 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. He teaches popular master classes in:

- Big Data Architecture & Technology
- Building a Competitive Data Strategy for A Data-Driven Enterprise
- Data Catalogs – Governing and Provisioning Data in a Data Driven Enterprise
- Data Virtualisation & the Logical Data Warehouse
- Data Warehouse Modernisation
- Embedded Analytics, Intelligent Apps & AI Automation
- How to Govern Data Across a Distributed Data Landscape
- Migrating your Data Warehouse to the Cloud
- Modern Data Architecture
- Practical Guidelines for Implementing a Data Mesh (Data Catalog, Data Fabric, Data Products, Data Marketplace)
- Real-Time Analytics

ONSITE TRAINING

All training can be given as onsite education, tailored to your company's requirements. For further details please contact us at training@intelligentbusiness.biz

