DATA WAREHOUSE MODERNISATION

FROM PASSIVE DATA WAREHOUSE TO LIVE ANALYTICAL ECOSYSTEM
2-3 MAY 2019 - ETC.VENUES MARBLE ARCH
Today, with most people connected to the Internet, the power of the customer is almost limitless. They can browse your competitors’ web sites. They can compare prices, view sentiment about your business, and switch loyalty in a single click anytime, anywhere all from a mobile device. In addition, social media has given customers a voice to express opinion and sentiment about products and brands and to create social networks by attracting followers, and following others. For many CEOs, customer retention, loyalty, service and growth are top of their agenda. Therefore, they want access to new data to enrich what they already know about customers.

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

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

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

AFTER 2 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 exploit cloud computing at lower cost
- Understand how to reduce data latency
- 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 other analytical systems
- Know how to use data virtualisation to simplify access to a more comprehensive set of insights available on multiple analytical platforms running analytics on different types of data for precise evidence-based decision making
- Understand the role of a modern data warehouse in a data-driven enterprise

MIKE FERGUSON - TRAINER

Mike Ferguson is Managing Director of Intelligent Business Strategies Limited. As an analyst and consultant he specialises in business intelligence and enterprise business integration. With over 35 years of IT experience, Mike has consulted for dozens of companies. He has spoken at events all over the world and written numerous articles. He is Chairman of Big Data LDN – the fastest growing Big Data conference in Europe, chairman of the UK CDO Exchange, the Italian Big Data Conference and the Italian Data and Analytics Conference. Formerly he was a principal and co-founder of Codd and Date Europe Limited – the inventors of the Relational Model, a Chief Architect at Teradata on the Teradata DBMS and European Managing Director of Database Associates. He teaches popular master classes in Machine Learning and Advanced Analytics, Big Data Fundamentals, Data Governance & MDM, and Data Lake operations.
1. THE TRADITIONAL DATA WAREHOUSE AND WHY IT NEEDS TO BE MODERNISED

For most organisations today, their data warehouse is 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. Hadoop, 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 and integrate to deliver value

2. MODERN DATA WAREHOUSE REQUIREMENTS

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

> Modern data modelling techniques
> Accelerating ETL processing using data lake, automated data discovery, an information catalog and re-usable data assets
> Cloud based analytical DBMS
> External tables 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,
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

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?
5. ACCELERATING ETL PROCESSING USING A MULTI-PURPOSE DATA LAKE & DATA CATALOG

This session looks at how you can use a multipurpose data lake to accelerate ETL processing and integration of data for your data warehouse.

> What is a data lake?

> How can it accelerate ETL processing and self-service data preparation?

> Ingesting and staging your data in a data lake

> Using an information catalog to automatically discover, profile, catalog and map data

> GDPR - Detecting sensitive data during automatic data discovery

> Creating an information supply chain to process data in a data lake

> Using Spark or Hadoop for scalable big data ETL processing

> Masking GDPR sensitive data during ingestion or ETL processing

> Is using ETL tools for processing unstructured data a good idea?

> ETL processing for streaming data in a real-time data warehouse

  - What is streaming data?
  - Types of streaming data - IoT data, OLTP system change data capture, weblogs...

> Key technologies for processing streaming data – Kafka, streaming analytics and event stores

> Turning OLTP change data capture into Kafka data streams

> Linking Kafka and ETL tools to process data in real-time

> Running ETL processing at the edge Vs on the cloud or the data centre

> Future proofing streaming ETL processing using Apache Beam

> Ingesting streaming data into your data lake

> Real-time data warehouse - Integrating your data warehouse with streaming data – external tables, data virtualisation and data lake

> Using ETL data pipelines to produce reusable data assets for use in your data warehouse and other analytical data stores

> Publishing reusable data in a catalog ready for consumption

> Using Data Science to develop new analytical models to run in your data warehouse
6. RAPID DATA WAREHOUSE DEVELOPMENT USING DATA WAREHOUSE AUTOMATION

In addition to a data lake, this session looks at how you can use metadata driven data warehouse automation tools to rapidly build, change and extend modern cloud and on premises Data Warehouses and data marts. It looks at how these tools help you adopt new modern data modelling techniques quickly, how they generate schemas and data integration jobs and how they can help you migrate to your new data warehouse systems on the cloud.

> What is Data Warehouse Automation?
> Using Data Warehouse Automation Tools for rapid data warehouse and data mart development
  - Generating Data Vault, E/R and Star Schema design
  - ETL job 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
- 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. WhereScape, Trivadis BIGenius, Attunity Compose, Balanced Insight, BIML and more
> Metadata driven data warehouse maintenance

7. BUILDING A MODERN DATA WAREHOUSE IN A CLOUD COMPUTING ENVIRONMENT

A key question for many organisations is what do you do with your existing data warehouse? Should you try to change the existing set-up to make it more modern or re-develop it in the cloud? This session looks at the advantages of building modern data warehouses in a cloud computing environment using.

> Why use Cloud Computing for your Data Warehouse?
> Pros and cons of deploying on the cloud?
> Cloud based data warehouse development – what are the options?
> Cloud based analytical relational DBMSs
  - Amazon Redshift, Google BigQuery, Microsoft Azure SQL Warehouse, Snowflake, Teradata, Kinetica, IBM Db2 Warehouse on Cloud
> Separating storage from compute for elasticity and scalability
> The power of GPUs and In-memory caching in an analytical DBMS
> Managing and integrating cloud and on - premises data
> Using iPaaS software to integrate data in cloud ETL processing – Informatica IICS, Dell Boomi, SnapLogic, Talend, StreamSets…..
> Managing streaming data in the cloud
> Integrating big data analytics into a cloud based data warehouse
> Train and deploying machine learning in your analytical database for in-warehouse analytics
> Tools and techniques for migrating an existing data warehouse to the cloud
> Dealing with cloud DW migration issues
> Managing access to cloud based data warehouses
> Integrating cloud based BI systems with on-premise systems
8. SIMPLIFYING DATA ACCESS – CREATING VIRTUAL DATA MARTS AND A LOGICAL DATA WAREHOUSE ARCHITECTURE TO INTEGRATE BIG DATA WITH YOUR DATA WAREHOUSE

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

- What is data virtualisation?
- How does data virtualisation work?
- How can data virtualisation reduce cost of ownership, improve agility and modernise your data warehouse architecture?
- Simplifying your architecture by using data virtualisation to creating 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 catalog for consumption
- Integrating your data warehouse with your data lake and low latency data using external tables and data virtualisation
- Enabling rapid change management using data virtualisation
- Creating a logical data warehouse architecture that Integrates data from big data platforms, graph databases, streaming data platforms and your data warehouse into a common access layer for easy access by BI Tools and applications
- Using a business glossary and data virtualisation to create a common semantic layer with consistent common understanding across all BI tools

9. GETTING STARTED WITH DATA WAREHOUSE MODERNISATION

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

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

Big Data LDN Training

For more information on Big Data LDN Training including future dates and locations head to BigDataLDN.com/Training or call 020 3291 4014

PRICING

<table>
<thead>
<tr>
<th>Delegation Type</th>
<th>Single Delegate Admission</th>
<th>2 Day Delegate Admission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Delegate Admission</td>
<td></td>
<td>£1,395.00+VAT</td>
</tr>
<tr>
<td>2 Day Delegate Admission</td>
<td></td>
<td>£1,255.00+VAT</td>
</tr>
<tr>
<td>2-3 Delegates Admission</td>
<td></td>
<td>£1,116.00+VAT</td>
</tr>
<tr>
<td>2 Day Delegate Admission (10% Discount)</td>
<td></td>
<td>£1,046.00+VAT</td>
</tr>
<tr>
<td>4-5 Delegates Admission</td>
<td></td>
<td>£1,116.00+VAT</td>
</tr>
<tr>
<td>2 Day Delegate Admission (20% Discount)</td>
<td></td>
<td>£1,046.00+VAT</td>
</tr>
<tr>
<td>6+ Delegates Admission</td>
<td></td>
<td>£1,046.00+VAT</td>
</tr>
<tr>
<td>2 Day Delegate Admission (25% Discount)</td>
<td></td>
<td>£1,046.00+VAT</td>
</tr>
</tbody>
</table>

IN PARTNERSHIP WITH:

Intelligent Business Strategies