



QUEST FOR KNOWLEDGE

ETL ARCHITECTURE IN DEPTH

DATE

17 - 20 May 2010

LOCATION

Amsterdam, The Netherlands

INFORMATION AND REGISTRATION

www.Q4K.com

A KIMBALL UNIVERSITY course

**Taught by RALPH KIMBALL
and BOB BECKER**

Bestselling authors **Ralph Kimball** and **Bob Becker** show you how a properly designed ETL system extracts the data from the source systems, enforces data quality and consistency standards, conforms the data so that separate sources can be used together and finally delivers the data in a presentation-ready format.

A must for every Data Warehouse and Business Intelligence specialists!

Join the industry experts as they lead you through four days of learning on how to:

- Plan and design your ETL System
- Choose the appropriate architecture from the many possible options
- Build a comprehensive data cleaning subsystem
- Tune the overall ETL process for optimum performance
- And much more!

Organized by



With the support of



Kimball University

Kimball University (KU), operated by the Kimball Group, is the definitive source for dimensional data warehouse education. The KU goal is to provide the highest quality and most practical education consistent with KU instructors' books and extensive experience in the dimensional approach. All our class content is vendor neutral. You'll learn from the best in the business. Quest For Knowledge has partnered with KU for several years to bring the highly rated KU classes to Europe.

RALPH KIMBALL



Ralph Kimball is known worldwide as an innovator, writer, educator, speaker and consultant in the field of data warehousing. He has remained steadfast in his long-term conviction that data warehouses must be designed to be understandable and fast. His books on dimensional design techniques have become the all time best sellers in data warehousing. His books include *The Data Warehouse Toolkit - 2nd Edition*, *The Data Warehouse Lifecycle Toolkit - 2nd Edition*, *The Data Webhouse Toolkit*, *The Data Warehouse ETL Toolkit* and *The Microsoft Data Warehouse Toolkit*. To date Ralph has written more than 100 articles and columns for Intelligent Enterprise and its predecessors, winning the Readers' Choice Award five years in a row.

After receiving a Ph.D. in 1972 from Stanford in electrical engineering (specializing in man-machine systems), Ralph joined the Xerox Palo Alto Research Center (PARC).

At PARC Ralph co-invented the Xerox Star Workstation, the first commercial product to use mice, icons and windows.

Ralph then became vice president of applications at Metaphor Computer Systems, pioneering decision support software and services provider. As a hands-on manager, he developed the Capsule Facility in 1982. The Capsule was a graphical programming technique which connected icons together in a logical flow, allowing a very visual style of programming for non-programmers. The Capsule was used to build reporting and analysis applications at Metaphor.

Ralph founded Red Brick Systems in 1986, serving as CEO until 1992. Red Brick Systems, now owned by IBM, was known for its lightning fast relational database optimized for data warehousing. Ralph Kimball Associates incorporated in 1992 to provide data warehouse consulting and education.

BOB BECKER



Bob Becker has worked with business managers and IT professionals to prioritize, justify and implement large-scale decision support and data warehousing systems since 1990. Regardless of the industry, he is highly skilled at identifying business requirements, facilitating organizational consensus and designing dimensional data models. Bob leverages these consulting experiences when teaching the Kimball University on-site courses. He also co-authored *The Data Warehouse Lifecycle Toolkit - 2nd Edition*.

Before co-founding DecisionWorks, Bob worked at Metaphor. He also held various sales and management positions with Oracle, Tandem Computers, IBM and Data General. He graduated from the University of Minnesota School of Business with a BSB in Marketing.

Course Outline

(Numbered items refer to the 34 subsystems taught in this course)

Surrounding The Requirements

- Business needs
- Compliance
- Data profiling
- Security latency
- Archiving
- End user profiles
- Skills
- Licenses
- Coding vs. tool choice
- The restaurant analogy
- Data types used in ETL systems

- (1) Data profiling
- Source to target map
- Access methods, source types
- Software, techniques
- (2) Change data capture
- (3) Extract window
- (3) Immediate transformations
- (3) Extract staging table designs, table types, retention, backup
- (22) Job scheduler
- (22) Exception handling architecture
- (23) Backup, (24) recovery, (24) restart
- Historical versus incremental load
- Team responsibilities

Cleaning

- (4) Data quality architecture
- (4) Data quality screens
- (5) Error event fact table
- (6) Audit dimension, compliance tracking
- (28) Sorting
- Module designs: (7) customer deduplication, address validation, ...
- Final clean data table designs

(8) Conforming

- Definition of conformed dimensions and facts
- Using the matrix
- Master data management
- Mapping incompatible structures into common structure
- (25) Version control
- (26) System and version migration, testing and regression

- (27) Workflow monitor
- (23) Job scheduler
- (29) Lineage and dependency analyzer
- (30) Problem escalation system

Modifying your ETL architecture for Real Time data warehousing

- The Hot Partition
- Streaming ETL vs. batch ETL
- Streaming extract
- Streaming cleaning and conforming
- Streaming delivery, query, reporting, dashboards, notifications
- EII architecture (Enterprise Information Integration)
- CTF architecture (Capture, Transform, and Flow)
- EAI architecture (Enterprise Application Integration)
- MBETL architecture (Micro Batch ETL)

Delivering Dimension Tables

- Referential integrity
- (9) Time variance designs (Slowly Changing Dimensions)
- (10) Surrogate key generator
- (15) Multi-valued dimensions, bridge tables I
- Special cases (extreme dimensionality, extreme dimension width, many incompatible members)
- (11) Hierarchical dimensions (fixed, variable, ragged), bridge tables II
- (12) Special dimensions (mini, junk, outrigger, shrunken, step, text fact)

Delivering Fact Tables

- (13) Fact table builder (transaction, periodic, and accumulating grains)
- (14) Surrogate key pipeline
- Graceful extensibility (add attributes, add facts, add dimensions to existing schemas)
- Handling structure changes
- (16) Late arriving data design

- (17) The dimension manager, responsibilities and procedures, real time complexities
- (18) The fact provider, responsibilities and procedures, real time complexities
- Distributed, federated data warehouses
- Delivering remote dimensions and attributes
- Delivering remote facts
- (19) Aggregations
- (20) Feeding OLAP cubes
- (21) DI (Data Integration) manager (feeding data mining, presentation layer extracts, 3rd party flat files)

Development and Operations

- (31) Parallel processing and pipelining
- (32) Security
- (33) Compliance

(34) Metadata

- Process metadata
- Run results, exception handling, immediate schedule
- Technical metadata
- System inventory, data models, data definitions,
- Business rules, ETL jobs, transformations, batch parameters
- Business metadata
- Business definitions, source system info, DW data dictionary

- Responsibilities
- Team roles
- Stepping back from the details
- The important decisions
- The important deliverables

Why attend

This course makes sure that you understand all the factors necessary for effectively designing the back room of a data warehouse that can gracefully evolve over time as your needs mature and new technologies become available. The course is intended for the data warehouse designer who has identified the sources of data and the target end users and is ready to start implementing.

Above all, this course tries to guarantee that you don't overlook a critical requirement. For example you dare not design your data warehouse while ignoring:

- Compliance
- Integration of diverse sources
- Increasingly demanding real-time pressures
- The time variance of your major dimensions such as customer and product
- Being able to resume or back out a partially completed load
- Having a 100% certainty that you have captured all the changes in the source systems
- And a host of other requirements that you will learn about in this course

Even if you don't have an immediate qualified need for every item on our list, over time it is likely that you will. At the end of this course you will understand how your data warehouse ETL system can be built to anticipate all of the possible requirements. This is not a microscopic code-oriented implementation course. Rather, it is an architecture course for the designer who must keep a broad perspective, and who needs to know what the latest technologies and techniques make possible. The course is organized around 34 necessary ETL subsystems which are developed in detail as the course progresses. (See the course outline for the names of the 34 subsystems). In this course, you will circle around a series of design issues starting with the first steps of extraction, on through to the final steps of delivery of properly formatted data suitable for your BI tool. In this four day course, each student builds on paper a comprehensive ETL system based on a realistic complex example. All 34 subsystems are included.

Prerequisites

Familiarity with the basic principles of dimensional modeling is helpful since dimensional models are designed as the ultimate ETL deliverables. The student can gain this familiarity by reading the first four articles in the Fundamentals series of articles found on the Kimball University website (<http://www.kimballgroup.com/html/articlesFundamental.html>). The course will include selected brief reviews of the principles of dimensional modeling so that everyone has the same vocabulary.

Registration fee

The fee for the 4-day course "ETL Architecture in Depth" is €2.695 per person. This includes four days of instruction, refreshment breaks, lunch, course materials and a Certificate of completion. Students also receive a copy of *The Data Warehouse ETL Toolkit*.

We offer the following discounts. Only one discount may be used.

- 10% Early Bird Discount for students registering before the Early Bird cut off date (26 March 2010). Payment must be received before the cut off date to receive the discount.
- 10% discount for groups of three or more students from the same company registering at the same time. Groups of three or more students that register at a discounted rate must retain the minimum group size or the discount will be revoked.

Who should attend

This course is designed for data warehouse implementers, who are responsible for building the back room, or ETL portion, of a data warehouse environment. This would include ETL developers, ETL architects, data warehouse operational staff, compliance tracking data warehouse professionals and real time data warehouse designers.

Venue

Dorint Hotel Amsterdam Airport is located on the outskirts of the Amsterdam Forest. It is a very comfortable hotel which was opened in August 2003. Only 7 minutes from Schiphol Airport and a few minutes away from the highway. The free shuttle bus drives you to the Airport between 06.00 - 23.45 o'clock, two times an hour. The hotel's underground car park can accommodate up to 266 vehicles.

Dorint Hotel Amsterdam Airport

Stationsplein ZW 951

1117 CE SCHIPHOL

T: +31 20 5400777

F: +31 20 5400700

E: www.dorintamsterdam.com

17 - 20 May 2010

ETL ARCHITECTURE IN DEPTH

Quest For Knowledge

The Netherlands

Hoge Schouw 1H | 4817 BZ Breda

T: +31 (0)76 572 21 99 | F: + 31 (0)76 572 21 96

Belgium

Uitbreidingstraat 84-3 | 2600 Antwerp

T: +32 (0)3 877 93 39 | F: + 32 (0)3 877 93 41

Online

www.Q4K.com | info@Q4K.com

Organized by

Quest For Knowledge

Architecting an IT environment that stands the test of time begins with a sharp vision on the durability of all of its components. Quest for Knowledge (Q4K) concentrates on education and training on software and concepts that have a bright future in one of these interrelated disciplines: Data Warehousing, Business Intelligence and Customer Relationship Management. The Q4K Data Warehouse and Business Intelligence curriculum provides in the most comprehensive education and training available in the Benelux. With in depth Data Warehouse courses and a series of product oriented training classes for leading Business Intelligence solutions, Q4K training provides you with the best knowledge transfer and a sound foundation to make your projects successful. Visit our website www.Q4K.com or request our training catalog for a complete overview

Kimball University

Kimball University (KU) is the definitive source for dimensional data warehouse education. KU provides the highest quality and most practical education consistent with KU instructors' books and extensive experience in the dimensional approach. You'll learn from the best in the business. Kimball University offers public classes in venues around the US and internationally. In addition, KU teaches classes on-site at client locations. All class content is vendor neutral.

With the support of

CIBER The Netherlands

CIBER is an international IT service provider, specializing in building, integration and management of critical business solutions and systems. CIBER's solutions are available in both custom-tailored and package form. Their services are provided for all technological platforms, operating systems, and existing infrastructures. In the Netherlands, CIBER has now been operating for over 10 years with approximately 450 employees in the offices in Eindhoven, Nieuwegein, Den Bosch, Rotterdam and Amsterdam. CIBER Nederland focuses on the (high-tech) industry, utilities, government, telecom, media, financial and business services sectors. Committed to the limit is a promise that is inextricably bound up with CIBER and signifies 'dedication' or 'unlimited involvement'. The corresponding core values such as client-focus, passion and a deal is a deal, support this message. CIBER has consultants with an above-average level of experience. They express themselves clearly and get genuinely involved. That is why all CIBER employees stand behind the agreements they make! CIBER The Netherlands can be found at www.ciber.nl.

Oracle

Oracle's business is information—how to manage it, use it, share it, protect it. For nearly three decades, Oracle, the world's largest enterprise software company, has provided the software and services that let organizations get the most up-to-date and accurate information from their business systems. Today, Oracle is the world's largest business software company, with more than 320.000 customers. For more information, please visit www.oracle.com.

REGISTRATION FORM Register by fax : +31 76 572 21 96**Course Details**

ETL Architecture in Depth ■ 17-20 May 2010 ■ Amsterdam, The Netherlands ■ € 2695 (excl. VAT)

Company Details

Company Name: _____

E-mail: _____

Contact Name: _____

Phone: _____

Address: _____

Fax: _____

Postal Code: _____

Website: _____

City: _____

Invoice Address: _____

Country: _____

Postal Address: _____

VAT Number: _____

Purchase Order no.: _____

Student Details

First Name: _____

Language: NL FR EN

Last Name: _____

Gender: Male Female

Job Title: _____

E-Mail: _____

Authorization

Name: _____

Job Title: _____

Date: _____

Signature: _____

Registration Information

Confirmation and Invoicing: upon receipt of your registration our customer service department will send you a customer information pack including details of payment and hotel information. Full payment is due prior to the course start date.

Cancellations and Substitutions: Cancellations must be received in writing 20 working days prior to the course start date and are subject to a 20% administration fee. Otherwise the full registration fee remains due. As an alternative to cancellation you may transfer your place for the course to a colleague without extra costs, but Quest For Knowledge has to be informed about this transfer in advance. Quest For Knowledge reserves the right to cancel any course at anytime without any liability whatsoever, safe for the refund of the registration fee.