

## **Warehouse Architecture and Design Principles**

**21. – 23. máj 2008, SAS Training Centre, Bratislava**

### **Course description:**

Course provides a broad coverage of the architecture and detailed physical design of a data warehouse - the storage, management and exploitation of data in an integrated information architecture.

### **Objectives:**

After attending the course, attendees will be able to create architecture and designs for an integrated information architecture, including:

- the Enterprise storage layer of a data warehouse, based on the most common physical models used
- OLAP data marts
- data mining data marts
- processes (ETL - Extract, Transform and Load) used in loading the Enterprise layer and in maintaining data marts
- management and publishing of metadata
- managing the environment, including archival and backup of data
- basic principles of planning the infrastructure (hardware and software)
- considerations for optimal performance in large data warehouses

### **Prerequisite Skills:**

Essential:

- familiarity with a wide range of SAS software used in data warehousing, SAS data management and data access features, SAS OLAP viewing/reporting tools
- ability to read entity-relationship diagrams
- knowledge of the basic principles of designing applications and data

### **Useful but not essential is experience with:**

- SAS/Warehouse Administrator software
- databases and/or SPD Server
- client/server SAS features
- data mining
- web technology
- data warehouse applications
- projects as warehouse architect (WAC)
- entity-relationship modeling techniques and tools. SAS System Modules Used

### **Lecturer:**

Training will be held by Mr Steve Morton, Independent Data Warehousing Consultant from Applied System Knowledge Ltd. Mr Morton has worked in management information systems and software for 34 years, including data warehousing since the mid-1990's. Mr Morton held various senior technical and management roles with SAS UK over 16 years, and for the past 10 years have been an independent consultant in data warehousing. During that time Mr Morton has worked on data warehouse projects in most industries, including banking, insurance, mobile telco, retail and pharmaceutical manufacturing in roles ranging from architecture, design and planning through project review, tool development and performance analysis. Since 2000 I started to combine consulting with

teaching classes for SAS in data warehouse methodology, architecture and design. He has obtained also SAS Certified Warehouse Architect certificate.

## **Course Topics:**

The course will consist of three parts:

1. Presentations
2. Demonstrations of practical data warehouse techniques
3. Exercises based on a case study

The presentations cover the following:

### Introduction

- Concepts of an integrated information architecture
- Role of the 'Data Warehouse Architect'
- Evolution of a data warehouse environment

### Subject Modelling

- Data models used in planning a data warehouse
- Subject models and how to define them
- Case study exercise: defining a Subject Model

### Designing the Enterprise Layer

- Definition of Logical modelling versus Physical modelling
- Preparing a Logical model
- Preparing a Physical model, choosing the model type
- Defining and using Normalized data models
- Defining and using Dimensional data models
- Specific features of Dimensional models: facts and measures, conforming dimensions, time and history, querying a star schema
- Case study: Dimensional modelling

### Process Architecture and Design

- Planning the Process Architecture
- Considerations for Process Design: Source identification, business rules, data quality, security
- ETL - acquiring, preparing and loading data
- Extraction - business rules, changed data capture
- Transformation - validation & cleansing, integration, enrichment, transformation, transfer
- Load - load techniques including 'slowly changing dimensions'
- Case study: Extract processing, Load processing
- ETL administration
- Use of 'standard techniques' in ETL

### Security Considerations

- Defining Security requirements
- Architecture considerations, main types of Security
- Design considerations for Security

### Metadata

- What is metadata
- Defining and using metadata

#### Data Marts

- Role of data marts in an integrated information architecture
- Structure of data marts
- Data mart physical models
- Aggregates in data marts
- Data mart design

#### Optimizing the Enterprise Layer of the Data Warehouse

- Scalability impacts, performance trade-offs
- Optimization strategy
- Common optimizations used in the data warehouse
- Case study: optimization considerations

#### Process Architecture and Design part 2: Data Warehouse Management

- Considerations for the end-to-end process
- Scheduling
- Distribution of data
- Aging, Archiving and Recovery
- Backup and Restore

#### Infrastructure

- Considerations for hardware choices, software choices
- Estimating size, capacity planning

#### Summary and Close

- The exercises are based on a case study and include group activities.
- Attendees work on requirements derived from real projects to create physical data models and design warehouse processes.

Examples demonstrated during the class include a complete working ETL process for an enterprise-style data warehouse with slowly changing dimension history.