

ETL

The purpose of this course is to provide IT professionals with an in-depth understanding of the steps involved in the Extract, Transform and Load (ETL) process used principally for building a data warehouse. This is typically the most difficult and challenging part of any data warehouse project.

On completion, delegates will be able to:

- Apply techniques to extract data from various source systems, transform data including cleaning and conforming, and load data into a data warehouse.
- Appreciate the practical issues and steps involved within ETL, covering design, development, architecture, operations and metadata.
- Design, develop and implement an ETL process to successfully obtain, prepare and publish data in a data warehouse.

Who should attend?

This course is aimed at an audience of IT professionals who will be involved in any project where data is being extracted from one or more source systems, consolidated and transformed, and subsequently loaded into a destination database such as a data warehouse. The techniques and steps covered are also applicable to data integration and data migration projects.

Prerequisites

A basic awareness of the following database concepts:

- Relational database concepts such as tables, attributes, joins, primary and foreign keys.
- Dimensional database concepts such as dimensions, measures, artificial keys, aggregate and snapshot tables, and slowly changing dimension techniques.
- Physical database concepts such as referential integrity, indexes, partitions and rollback logs.

Course Duration

2 days.

Related and Follow-On Courses

Data Quality, Dimensional Modelling for the Data Warehouse, SQL, Data Integrator.

Practical work

Group discussions form an essential part of the course theory as topics are introduced. Practical exercises will reinforce the learning points and are used extensively throughout the course to allow delegates to consolidate their understanding of the ETL process.

Course Contents

Introduction to ETL

Why, how and where it is used, the types of source and destination systems involved and the effect of their stability, and the importance of a robust ETL design.

ETL Design and Architecture

This covers production of a logical data map, the role of the Staging Area, discusses the different flavours of ETL architecture and compares the use of an ETL tool with hand-crafting the ETL process.

Understanding and Extracting the Source Data

This concentrates initially on data analysis and discovery to identify our source data, and data profiling to gain an understanding of the various attributes of our source data, with a look at the role of data profiling tools. Then the issues surrounding the actual extract of data from the source systems are discussed, documenting these via the Source System Tracking Report, and the importance of identifying the master data - the system of record.

Transformation - Data Cleansing

This covers the basic concepts of data quality including its objectives, how errors should be handled, an overview of the data cleansing process, audit dimensions and error event fact tables, what types of checks can be performed, and approaches to dealing with problem data.

Course Contents (cont.)

- **Transformation - Conformance**

Covering the issues and steps involved in conforming both dimensions and facts, and loading the conformed records. This includes slowly changing dimensions and the various types of fact table. The three key steps in conforming dimensions of standardisation, matching and de-duplication, and survivorship are discussed.

- **Metadata**

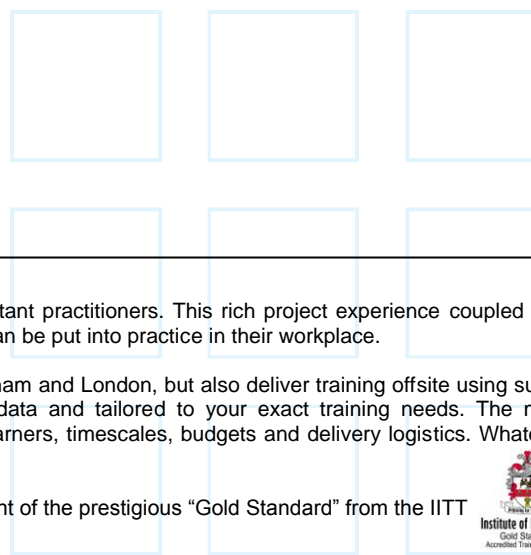
Covering the definition of metadata, the three categories of business, technical and process metadata, data lineage and impact analysis. Standards and best practice for metadata are discussed.

- **Loading Data & Operational Considerations**

This provides an overview of the main factors which influence optimal loading, and the operational considerations when implementing and running the ETL system. The main considerations for real time ETL are discussed.

- **Approach to ETL Projects**

Discussing the main issues associated with planning and managing ETL projects and the recommended approach. Further examples of ETL principles and best practice are discussed.



Acuma Education

All Acuma courses are delivered by trainers who are highly experienced consultant practitioners. This rich project experience coupled with professional training skills helps delegates to focus on how the course contents can be put into practice in their workplace.

Acuma delivers courses from dedicated training facilities in Manchester, Birmingham and London, but also deliver training offsite using suites of laptops, with the option to customise courses to use your own business data and tailored to your exact training needs. The most appropriate training format depends upon your training objectives, number of learners, timescales, budgets and delivery logistics. Whatever your needs, Acuma will find the approach most suited to them.

Acuma Education's seal of quality is demonstrated through consistent achievement of the prestigious "Gold Standard" from the IITT (Institute of IT Trainers).



Acuma is an Information Management (IM) Specialist. Acuma is part of the Saksoft group and provides solutions, which are unique, flexible and cost-effective service blending local high value consultancy and global high quality project delivery. Acuma delivers business improvements by drawing together strategy, technology and methods of Information Management into a single philosophy called the Information Value Model (IVM).