Data Integrator XI: Extracting, Transforming & Loading Data

Data Integrator enables the implementation of ETL (Extract, Transform, and Load) projects from disparate data sources to deliver more timely and accurate data that end users in an organisation can trust. In this four-day course, you will learn about batch data transformation jobs, techniques for capturing changes in data, handling errors, multi-user environment tasks, administering server and migration basics. The course also shows you how to use tools to audit, profile data, and manage metadata to assist in the implementation of an ETL project.

Smaller activities in this course focus on the tools and features discussed in each lesson and enable you to create dimensions and a fact table. You will have a chance to apply different concepts learned from a few combined lessons in two 1 hour comprehensive workshops. At the end of the course, you will be able to put into practice data flow design concepts in a final 2-3 hour workshop.

As a business benefit, by being able to create efficient ETL projects, you can use the transformed data to help improve operational and supply chain efficiencies, enhance customer relationships, create new revenue opportunities, and optimise return on investment from enterprise applications.

On completion, delegates will be able to:

- Make effective use of techniques during ETL to improve efficiency and gain real business benefits

Who should attend?

This course is designed for individuals responsible for implementing ETL projects (batch-mode), administering and managing projects that involve Data Integrator.

Practical work

Practical sessions will reinforce many of the topics covered and are used extensively throughout this course. These allow delegates to have direct hands-on practice using BusinessObjects Data Integrator.

Prerequisites

Experience with these products or technologies will be helpful:

- Knowledge of data warehousing and ETL concepts, SQL language
- Experience with Microsoft SQL Server
- Experience using functions, elementary procedural programming and flow-of-control statements, for example: If then Else, and While Loop statements.

It is recommended you review these articles prior to attending the course:
http://www.rkimball.com/html/articles.html

- Data Warehouse Fundamentals: *TCO Starts with the End User* and *Fact Tables and DimensionTables*
- Data Warehouse Architecture and Modeling: *There Are No Guarantees*
- Architecture/Modeling: Advance Dimension Topics Surrogate Keys: *It's Time for Time and Slowly Changing Dimensions*
- Architecture/Modeling - Industry - and Application-Specific Issues: *Think Globally, Act Locally*
- Data Staging and Data Quality**:** *Dealing with Dirty Data*

Course Duration

3 days

Related and Follow-On Courses

ETL (2 days)

0870 789 4333
information@acuma.co.uk
www.acuma.co.uk

Course Contents

Data Warehousing Concepts
- Describe dimensional modeling

Understanding Data Integrator
- Describe components, management tools, and the development process
- Explain object relationship

Defining Source and Target Metadata
- Create a database datastore and import metadata
- Create a new file format and handle errors in file formats

Validating, Tracing and Debugging Jobs
- Use descriptions and annotations
- Validate and trace jobs
- Use View Data and the Interactive Debugger

Creating a Batch Job
- Create a project, job, work flow, and data flow
- Use the Query transform in a data flow
- Use template tables

Using Built-in Transforms and Nested Data
- Use the case, merge, and validation transforms
- Import metadata from XML documents
- Use the XML_Pipeline in a data flow

Using Built-in Functions
- Use date and time functions and the date generation transform to build a dimension table
- Use the lookup functions to look up status in a table
- Use match pattern functions to compare input strings to patterns
- Use database type functions to return information on data sources

Using Data Integrator Scripting Language and Variables
- Explain differences between global and local variables
- Create global variables and custom functions
- Use strings and variables in Data Integrator scripting language

Capturing Changes in Data
- Use Changed Data Capture (CDC) with time-stamped sources
- Create an initial and delta load job
- Use history preserving transforms

Handling Errors and Auditing
- Recover a failed job
- Create a manual, recoverable work flow
- Define audit points, rules and actions on failure

Supporting a Multi-user Environment
- Describe terminology and repository types in a multi-user environment
- Create and activate the central repository
- Work with objects in the central repository

Migrating Projects
- Create multiple configurations in a datastore
- Work with projects in the central repository
- Create a secure central repository
- Implement and modify group permissions

Using the Administrator
- Add a repository and user roles
- Set the job status interval and log retention period
- Execute, schedule, and monitor batch jobs
- Understand architecture, load balance index, and job execution in server groups

Profiling Data
- Set up the Data Profiler and users
- Submit a profiling task
- Monitor profiling tasks in the administrator