

Data Warehousing with SSIS (SQL Server 2008 Integration Services)

Introduction

In a data warehouse (DW) or business intelligence (BI) project the ETL (extract, transform and load) or data integration phase usually requires the largest amount of enterprise resources.

The use of SQL Server Integration Services (SSIS), a fully functional, high performance, **graphical ETL tool** with built-in data cleansing and transformation functions, can lead to tremendous savings in time & effort. It is a **productivity tool** that makes development & maintenance of ETL processes **faster & much easier than hand coding**.

This course provides students with the knowledge and skills necessary to successfully design, develop and deploy efficient & manageable ETL solutions for DW/BI applications using SSIS.

Note: This course is not intended to teach students how to design data warehouses. Rather, it focuses on the population and maintenance of data warehouse structures with SSIS.

Course Objectives

The student will learn:

- SSIS features and functionality
- How to use SSIS to process fact & dimension tables in a data warehouse/business intelligence application
- How to design, develop and deploy SSIS solutions
- Best practices for designing and implementing SSIS solutions
- Performance tuning techniques

Duration

3 days

Format

Lecture and hands-on

Audience

ETL and data warehouse/business intelligence architects and developers

Prerequisites

Relational database concepts, dimensional modeling concepts

Course Materials

Course materials are yours to keep. You will be provided with the following software for use in the classroom:

- Microsoft SQL Server 2008

Course Content

SSIS Overview & Core Features

- Course Introduction
 - SSIS definition
 - DEMO: SSIS Tour
 - The role of SSIS in data warehousing/business intelligence
 - ETL processing & SSIS ETL features
- SSIS Overview
 - Architecture
 - Tools & utilities
 - Objects & concepts (incl. packages, control flows, data flows, precedence constraints, variables, event handlers, log providers)
 - Design environment: BIDS (Business Intelligence Development Studio), SSIS Designer

- Solutions & Projects
 - Definition
- Packages
 - Package objects
 - Execution and management (incl. storage, import/export)
 - LAB: Creating an SSIS Solution, Project & Package
- Control Flow
 - Definition
 - Control flow objects incl. tasks, containers & precedence constraints
 - LAB: Working with Control Flows
- Data Flow
 - Definition
 - Data flow objects incl. transformation objects, data sources, data destinations
 - LAB: Working with Data Flows

Data Warehousing with SSIS

- Extracting data
 - Data sources & data source views
 - Connection managers
 - Full & incremental data extraction
- Dimension table processing
 - Dimension table processing with data flow transformations incl. Slowly Changing Dimension, Character Map, Conditional Split
 - Handling inferred dimension members
 - Auditing
 - LAB: Dimension table processing
- Fact table processing
 - Fact table processing with data flow transformations incl. Lookup, Derived Column, Union All
 - LAB: Fact table processing
- Checkpoint/Restart
 - Definition
 - DEMO: Checkpoint/Restart
- Package execution & management
 - Package scheduling with SQL Server Agent
 - LAB: Package scheduling
- Debugging & event handling
 - Debugging (incl. data viewer, breakpoints)
 - Event handling and logging
 - LAB: Debugging
 - LAB: Event Logging
- Package configurations and deployment
 - Package configurations
 - Package deployment (e.g. to production environment)
- XML
 - Processing XML files (incl. SSIS looping and expressions)
 - LAB: Processing XML files
- Extending SSIS functionality with scripting
 - Halting package execution
 - Surrogate key generation

Best Practices, Performance Tuning, Et Al

- Best Practices
 - Modular development
 - Master/control packages
 - Template packages
 - Package configurations
 - LAB: Using master/control packages

- Performance Tuning
 - Parallelism
 - Queries
 - Bulk loading
 - Partitioning
- Additional Topics
 - Data cleansing (fuzzy matching)
 - DEMO: Fuzzy matching
 - Converting DTS packages to SSIS (Optional section)

Student Labs

Our labs use Microsoft Project REAL concepts and data - the reference implementation for SQL Server BI that Microsoft developed using Barnes and Noble data. We use Project REAL in order to leverage the best practices cultivated during the development of the reference implementation.