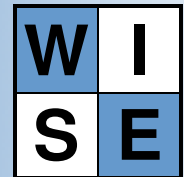


# Advanced SQL Server 2005/2008 Integration Services (SSIS) for Developers



WISE LTD.

- COURSE CODE:** SQSAIS
- COURSE TITLE:** Advanced SQL Server 2005/2008 Integration Services (SSIS) for Developers
- LEVEL:** Advanced
- AUDIENCE:** SQL Server 2005/2008 developers in particular and DBAs in general.
- PREREQUISITES:** Completion of our course 'SQL Server 2005/2008 Integration Services Fundamentals' (SQSSIS) or equivalent knowledge.
- DURATION:** 4 days
- SUMMARY:** As SSIS is mainly an Extract, Transform and Load (ETL) tool and since ETL has been accepted as being between 70-80% of the data warehouse maintenance challenge, this course is accordingly dedicated. After discussing advanced features of the SSIS control and data flow engines, delegates will be taught how to design, develop and implement an SSIS-based ETL system, accounting for some of the more common and complex requirements, including:
- Source-to-destination mapping
  - Change Data Capture (CDC)
  - Extraction (e.g., incremental)
  - Staging tables
  - Slowly Changing Dimensions
  - Late arriving (inferred) dimensions
  - Late arriving facts
  - Fact aggregation
  - Data profiling
  - Audit
  - Pipeline (data flow) performance & tuning
- OBJECTIVES:** Upon completion of this course, the delegate should be able to use a variety of SSIS control and data flow facilities to design, develop and maintain simple-to-complex ETL solutions.
- FORMAT:** Lecture and hands-on

## 1. REVIEW OF BI TERMS & CONCEPTS

- The data warehousing (DW) environment
- The DW computing context
- What is a data warehouse?
- What is a data mart?
- What is Business Intelligence (BI)?
- How do OLTP & OLAP differ?
- What is data mining?
- Operational vs. historical data
- Dimensional modelling
- Some dimensional modelling characteristics
- What is a star schema?
- What is a snowflake schema?
- What is metadata?

## 2. REVIEW OF SSIS BASICS

- Architecture
- Tools
- Wizards
- Packages
- Control Flows
- Data Flows
- SSIS in operational/data warehousing environments
- SSIS data types
- SSIS transformation types
- Containers
- Precedence constraints
- Variables
- Expressions

## 3. REVIEW OF PACKAGE OBJECTS

- Tasks
- Connection Managers
- Log providers
- Containers
- Precedence constraints

## 4. SSIS DATA TYPES & OPTIMISATION

- Types
- How to alter and map
- Parsing
- Optimisation

## 5. SSIS VARIABLES

- Definition
- Properties
- Scope
- Type
- Namespaces
- System
- User
- Execution behaviour

## 6. SSIS EXPRESSIONS

- Definition
- Elements
- Property expressions
- Variable expressions
- Precedence constraint expressions
- For Loop expressions

## 7. CONTAINERS

- Concepts
- Types (e.g., Foreach Loop)
- Scoping
- Properties

## 8. ADVANCED CONTROL FLOW TASKS

- File System
- Send Mail
- Message Queue
- WMI
- Execute Package
- Execute Process
- Bulk Insert
- Analysis Services
- Data Mining
- Execute SQL Agent Job
- Update Statistics
- Script Task

## 9. DATA PROFILING

- Scope and limits
- Invocation
- Assessing output of
- Reconciling with ETL

## 10. ADVANCED DATA FLOW COMPONENTS

- Source adapters (e.g., Raw Files)
- Destination adapters (e.g., Data Reader)
- Conditional Split
- Multicast
- Fuzzy Lookup/Fuzzy Grouping
- Script component
- OLE DB command
- Union All
- Sort
- Merge/Merge Join
- Lookup
- Pivot/Unpivot (Excel worksheet)
- Aggregate
- Slowly Changing Dimension

## 11. CACHE TRANSFORM

- Scope and limits
- Cache connection manager
- Lookup transform vis-à-vis Cache transform

## 12. OPTIMISING DATA FLOW TRANSFORMATIONS

- Buffer architecture
- Pipeline considerations
- Synchronous vs. asynchronous
- Blocking vs. semi-blocking vs. non-blocking
- Execution trees interpretation and usage

## 13. CHECKPOINT RESTART (WITH/WITHOUT TRANSACTIONS)

- Definition
- When to consider
- Checkpoints – how to implement and manage
- Checkpoint files
- Single Container and checkpoints
- Multiple Containers and checkpoints
- Single vs. multiple transactions
- How to configure
- How to set properties
- Distributive transaction considerations
- Scoping per container
- Isolation levels
- Snapshot isolation

## 14. ADVANCED DEPLOYMENT CONSIDERATIONS

- Package configuration
- What does the Deployment Wizard do?
- When should the deployment utility be used vs. import/Export, XCopy, etc?
- How to override configurations
- How to deploy and account for security

## 15. ADVANCED ETL DESIGN & DEVELOPMENT WORKSHOP

- Top-level planning
- Source-to-destination mapping
- Data profiling
- Load frequency
- Historical vs. incremental loads
- Extraction design
- Staging area
- Recovery considerations
- Dimension table processing
- Slowly Changing Dimensions (SCD)
- Late arriving (inferred) dimensions
- Fact table processing
- Extracting fact updates and deletes
- Transforming dimension/fact data
- Fact data aggregation
- Confirmed dimensions/facts
- Surrogate key maintenance
- Fact table load
- Periodic vs. accumulating snapshot fact considerations
- Audit considerations
- Parent package/child package considerations
- Event handling
- Unit testing

## 16. CHANGE DATA CAPTURE

- Scope and limits
- Advantages
- Planning
- Application interfaces

## 17. SSIS PERFORMANCE & TUNING

- How to optimise sources
- How to optimise logging
- How to optimise data flows
- How to optimise error handling
- How to optimise insert performance
- Useful naming conventions