

Introduction to DWH Using SQL Server Integration Services

Mgr. Vladimír Mužný

MVP: Data Platform | MCSE: Data Platform | MCT

vladimir.muzny@dropman.cz

 @VladimirMuzny

Agenda

- What the data warehouse/data warehousing is?
- What SSIS are?
- Architecture of SSIS
- Demo time

What the Data Warehouse Is

- We can think about relational databases as of
- OLTP – On-line Transactional Processing
 - Well normalized
 - Optimized for highly concurrent writes
- **OLAP – On-line Analysis Processing**
 - Denormalized
 - Optimized for big scans
 - Typically follow „star“ or „snowflake“ design pattern
- HTAP – Hybrid Transactional Analysis Processing
 - Almost in cloud

What the Data Warehousing Means?

- *It's many different disciplines*
- OLTP and OLAP contention is not easy to handle in the same database
- Data warehousing is huge set of procedures reading data from many sources, transforming the data and finally loading the data to the DWH
 - ETL
 - ELT

Data Warehousing Patterns

- Source-Sink
 - Not so reliable, just „copy-paste“ approach
- Source-Stage-Sink
 - A compromise between Source-Sink and other approaches
 - Trustworthy source data
- Source-Landing-Stage-Sink
 - Untrusted source data
 - A need for reliable transfer of data from many heterogenous sources
 - Many transformations

SQL Server Integration Services

- A technology offered by Microsoft since SQL Server 2005
- A Successor of DTS (Data Transformation Services)
- Used mostly for:
 - ETL projects (data warehousing)
 - Migration projects
- But also for:
 - Administration automation (Maintenance Plans)
 - Data Collection

SSIS Architecture

- Windows Service SSIS
 - Just monitoring, for backward compatibility purposes
 - ◆ And for SSISDB installation
- Two cooperating engines:
 - Control Flow
 - Data Flow
- Development tools
 - SQL Server Data Tools (SSDT)
 - ◆ Add-in into Visual Studio
 - ◆ SSDT versions are not the same as versions of SQL Server (SSIS)

SSIS Objects

- SSIS Package
 - Control Flow
 - ◆ Containers
 - ◆ Tasks
 - ◆ Precedence Constraints
 - ◆ Variables
 - Data Flow
 - ◆ Source
 - ◆ {Transforms}
 - ◆ Destination
 - Connection Manager
 - Event Handlers: Control Flow triggered by some event

Non-Functional Capabilities of SSIS

- Logging
 - Error handling
 - Configurations
-
- Until 2008 decentralized
 - Since 2012 centrally controlled in SSIS Catalog

SSIS Project/Package Development

- SQL Server Management Studio: Import/Export Data
- SQL Server Management Studio: Maintenance Plan
- SQL Server Data Tools: SQL Server Integration Services Project

SSIS Catalog

- SQL Server database called SSISDB
 - Storage for packages
 - Environments
 - Logging
 - Diagnostics
 - Performance

Enterprise features of SSIS

- Master Data Services
 - Data mastering
 - Master Data HUB
- Data Quality Services
 - Data deduplication
 - Column values correction
 - Since 2008
 - Very slow and poor performing

Open Discussion

Mgr. Vladimír Mužný

MVP: Data Platform | MCSE: Data Platform | MCT

vladimir.muzny@dropman.cz

 @VladimirMuzny