

Loading a Data Warehouse with SSIS

Brian Knight
bknight@pragmaticworks.com
Owner, Pragmatic Works



Agenda

- ▶ Introduction
- ▶ Data Warehousing Concepts
 - Dimensions and Types
 - Fact Tables
 - Inferred members
- ▶ Slowly Changing Dimension Wizard (SCD)
- ▶ Creating Your Own SCD
- ▶ Handling Scalability Issues
- ▶ OLAP Processing



Introduction

- ▶ Biography
 - Based out of Jacksonville, FL
 - SQL Server MVP
 - SQL Server consultant\mentor at Pragmatic works
 - Written 7 SQL Server books, including 2 on SSIS
 - Professional SQL Server 2005 Integration Services
 - Expert SQL Server 2005 Integration Services
 - Co-founder of SQLServerCentral.com
 - Co-founder of JumpstartTV.com.
 - Own training center in Orlando



Introduction

- ▶ Objective
 - The objective of the session is to teach you from the ground up how to load a data warehouse with SSIS and some of the good, the bad and the ugly around this process. You'll also learn how to work around some challenging scalability issues.

200-300 level



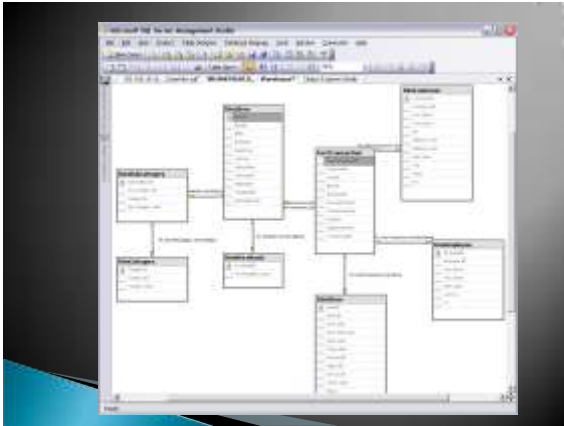
Dimensional Modeling

- ▶ Data Separated into fact and dimension tables
- ▶ Dimension tables answer the pivot or where clause
 - Make as wide and descriptive as possible
 - Surrogate keys operate as unique ID for each row
 - Keep surrogate keys as small as possible
- ▶ Fact tables answer the what or select statement
 - Intersect all dimension tables
 - Surrogate keys from each dimension in this table
 - Measures are the "what" like Price, Quantity, Duration



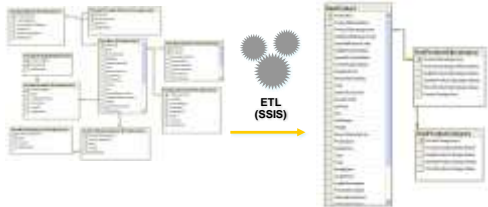
Let's Design One





SSIS ETL Challenge

- AdventureWorks OLTP



Dimension Types

- Known as slowly changing dimensions at a column-level.
- Types of dimension column changes:
 - Type 0 - Fixed, non changing attribute
 - Type 1 - Changing attribute, no history kept
 - Type 2 - Most complex, keeps historical changes
 - Type 3 - Keeps a fixed amount of versions back.

Slowly Changing Dimension Wizard

- SSIS transform that creates many other transforms conditionally
- Handles:
 - Type 0 (fixed attribute)
 - Type 1 (changing attribute)
 - Type 2 (historical attribute)
 - Inferred members
- Typically can address 80% of the business scenarios



SCD Wizard Strengths

- SSIS transform that creates many other transforms conditionally
- Reduces design time of SCD load by 80%-90% to minutes per dimension
- Can be customized easily
- Compares differences between source and destination to find changes and new records
- Outputs:
 - Type 0,1,2 update
 - Inferred members
 - New rows
 - Duplicate rows



Demo: SCD Wizard



SCD Wizard Weaknesses

- Scalability – Generally up to about 50,000 records into the transform but varies based on number of updates
- Maintainability – After you customize, rerunning the wizard recreates all the transforms
- Uses OLE DB Command transforms for updates is row-level. Creates scalability issue here if lots of updates.



Making Your Own SCD Wizard

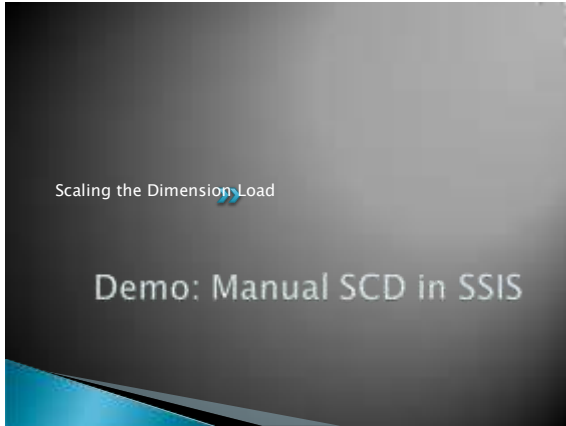
- Can use a Merge Join or Lookup Transform
 - If no match found, it is an insert (Ignore Errors)
- Lookup Transform will scale better than Merge Join but lacks parameterization
- Add a Conditional Split transform after Lookup to direct to insert, duplicate or update path



Additional Scalability

- Watch your Lookup Transformation for scalability issues (don't cache too much!)
 - Potentially cache only the last 1 years worth of data with Partial Caching
 - Only cache columns needed
- Additional scalability can be reached by landing updates into a staging table
 - Then set-based update with an Execute SQL task.
- Checksum Transform can be used to detect changes across many columns
 - Or HASHBYTES T-SQL statement





Fact Table Loads

- Series of Lookup Transforms
 - In Type 2 Dimensions add WHERE EndDate IS NOT NULL
- Measures created Derived Column Transforms
- Aggregate transform to roll up the grain
- Lookup failure would create an inferred member or set to unknown



