SSIS 2012, Part 05 of 11: Join Transformations

Meet the expert: Don Kiely is a featured instructor on many of our SQL Server and Visual Studio courses. He is a nationally recognized author, instructor, and consultant specializing in Microsoft technologies. Don has many years of teaching experience, is the author or co-author of several programming books, and has spoken at many industry conferences and user groups. In addition, Don is a consultant for a variety of companies that develop distributed applications for public and private organizations.

Prerequisites: This course assumes that you have a basic familiarity with the concept of relational databases and a basic understanding of what SQL Server is and the high-level tools in it, as well as how to create and manage objects using Management Studio. You should also have a basic understanding of how SQL Server implements security, including its authentication and authorization schemes, and how to assign permissions on securable objects to principals. You should know the fundamentals of Transact-SQL to write queries to retrieve data and join data from multiple tables, and how to execute scripts using the query editor in Management Studio. You must also know how to connect to an instance of SQL Server 2012 using the various connection dialog boxes in Management Studio and development tools. It will be very helpful, but not absolutely necessary, to have experience with .NET development using Visual Studio 2012 or later for the portions of the course that deal with SQL Server Data Tools (which is a lot of it). At the very least, we'll assume that you are well familiar with the Visual Studio user interface.
This course assumes no prior knowledge of SQL Server Integration Services.

Runtime: 01:27:24

Course description: Trivial Integration Services packages can take data straight from a data source and dump it unchanged into a destination. That's certainly a valued use of a package but it doesn't take advantage of all that you can do to data as it moves from place to place. More often you're likely to join data for multiple sources together into a single data flow. For example, in an online transactional source data base you might have normalized product information spread across several tables, including product, product category, product description and others. That can be an efficient structure for day to day online operations but if you need to insert the data into a database warehouse for analyses, normalized data is very inefficient. So a package you create to populate the data warehouse would need to join several sources together in to a single data flow and that's where the join transformations are useful. In this course you'll learn about two of the join transformations built into data flows. The Merge join transformation merges two data flows into a single output flow, performing a join operation on the data. This is the same operation you can perform in T-SQL using the inner-join or outer-join syntax in a select statement. The major difference is that the transformation operates on a stream of data coming in from the data flow pipeline. Whereas the T-SQL type of join is a set operation. The other join transformation is the Lookup transformation. This performs an inner join on its single input data flow using a data set from a source defined within the transformation instead of using another data flow. What really sets this component off is its ability to cash look up data in very flexible and powerful ways. The join transformations provide powerful features to process data in an Integration Services package and this course will get you started understanding on how to put them to use.

Course outline:

Merge Join Transformation
- Introduction
- The Merge Join Transformation
- Demo: IS Package
- Demo: Database Connection
- Demo: Data Sources
- Demo: Merge Join
- Demo: Sort Transformation
- Demo: Input Columns
- Demo: Union All
- Summary

Merge Join Transformation Cont.
- Introduction
- Demo: Missorted Data
- Demo: Category Data Source
- Demo: Merge Join
- Demo: More Data Flows
- Summary

The Lookup Transformation
- Introduction
- The Lookup Transformation
- Demo: Package Setup
- Demo: Lookup Transformation
- Demo: Lookup Columns
- Demo: Union All
- Demo: More Lookups
- Summary

Lookup Transformation Caching
- Introduction
- Caching
- The Cache Connection Manager
- Cache Connection Manager (Cont)
- Demo: Cache Modes
- Summary