

02 – SSIS

Whats New in 2008

Bob Duffy

Senior – Consultant

Microsoft Certified Database Architect | SQL Ranger



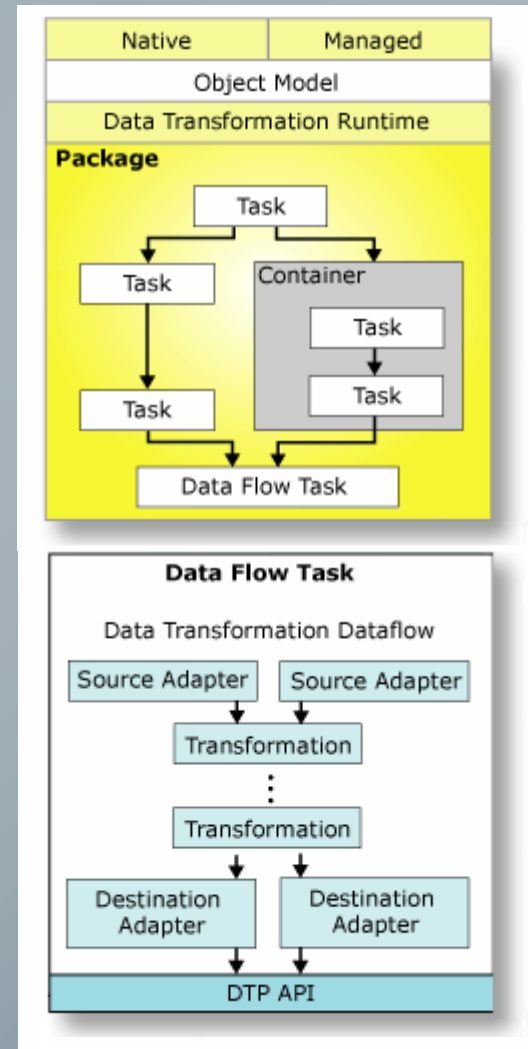
Agenda

- Level 100 recap
- What's New in 2008
- Best Practices
- Loading 1TB in under 30 mins
- Performance Tuning



Level 100 recap – What is SSIS

- Package
 - Tasks
 - Precedence Constraints
 - Connection Managers
 - Containers
- Data Flow Task
 - Components – Source Adapters, Transformations, Destination Adapters
 - Paths



Is SSIS for you ?

- Volume: 10's of MB to 10's of TB
- Latency: minutes to days
- Lower data volumes? SSIS might be overkill.
- Need lower latency? Consider BizTalk or SSB
- Key distinctions:
 - Data Integration -> SSIS
 - Business Process Integration -> BizTalk
 - Application Integration -> Biztalk, SSB
 - You need 1TB in 30 mins or 450 million rows/sec?
 - Needs work for meta data approach



New Features in SQL 2008

Perf and Scale

- Pipeline perf (dataflow scheduling and parallelism)
- Lookup with persistent cache
- Upsert/Merge
- CDC

Non-Events

- DTS is still supported
- 2005→2008 upgrade will be automatic.

Usability & Design

- Data Profiling Tools
- Real .NET Scripting and C#
- I/E Wizard: simpler, more flexible
- SuperDump

Data Sources

- ADO.NET
- Office 12
- SAP



The Data Profiling Task

The screenshot displays the SQL Server Data Profiling Task Editor and the Data Profile Viewer. The Data Profiling Task Editor on the left shows a tree view of the database schema with the 'city' column selected. The Data Profile Viewer on the right shows the 'Column Pattern Profile' for the 'city' column. The 'Pattern Distribution' table shows the following data:

Pattern	Pattern Percentage
p{L}+	68
p{L}+ \p{L}+	24
N/A	7

The 'Column Pattern Profile' window at the bottom shows a list of city names: EL SEGUNDO, SANTA ANA, NEW YORK, SAN DIEGO, DES MOINES, SAINT LOUIS, and NEW YORK.

Import/Export Wizard is Much Smarter

SQL Server Import and Export Wizard

Review Data Type Mapping

Select a table to review how its data types map to those in the destination. Also, select how the wizard handles conversion issues.

Table:

	Source	Destination
⚠	[Random Tests].[dbo].[Customer]	'Customer'
✓	[Random Tests].[dbo].[CustomerImport]	'CustomerImport'
✓	[Random Tests].[dbo].[CustomerStage]	'CustomerStage'

Data type mapping:

	Source Column	Source Type	Destination Co...	Destination Ty...	Convert	On Error	On Truncati...
✓	state_province	nvarchar	state_province	VarChar			
✓	postal_code	nvarchar	postal_code	VarChar			
✓	country	nvarchar	country	VarChar			
⚠	customer_regi...	int	customer_regi...	Short	✓	Use Global	Use Global
✓	phone1	nvarchar	phone1	LongText	✓	Use Global	Use Global
✓	phone2	nvarchar	phone2	LongText	✓	Use Global	Use Global
✓	birthdate	smalldatetime	birthdate	Date Time			
✓	marital_status	nvarchar	marital_status	LongText	✓	Use Global	Use Global
✓	yearly_income	nvarchar	yearly_income	LongText	✓	Use Global	Use Global
✓	gender	nvarchar	gender	LongText	✓	Use Global	Use Global
✓	total_children	smallint	total_children	Short			

To view conversion details, double-click the row that contains the column source type to be converted.

On Error (global) Fail On Truncation (global) Fail

Help < Back Next > Finish >> Cancel

Column Conversion Details

Review the conversion information for the selected column.

[Source Information]
Source Location : (local)
Source Provider : SQLNCLI10
Table : [Random Tests].[dbo].[Customer]
Column : customer_region_id
Column Type : int
SSIS Type : four-byte signed integer [DT_I4]
Mapping file (to SSIS type) : C:\Program Files\Microsoft SQL Server\100\DTS\Binn\MSQLToSSIS10.XML

[Destination Information]
Destination Location : C:\WINDOWS\system32\ias\dna\dna.mdb
Destination Provider : Microsoft.Jet.OLEDB.4.0
Table : 'Customer'
Column : customer_region_id
Column Type : Short
SSIS Type : two-byte signed integer [DT_I2]
Mapping file (to SSIS type) : C:\Program Files\Microsoft SQL Server\100\DTS\Binn\JetToSSIS.xml

[Conversion Steps]
Step 1: DT_I4 to DT_I2
SSIS conversion file: C:\Program Files\Microsoft SQL Server\100\DTS\Binn\DTwTypeConversion.xml

Scripting: VSA -> VSTA

2005: VSA

- Limited designer experience
- Visual Basic only
- Limited set of "reference-able" assemblies
- It is dead

2008: VSTA

- Seamless upgrade of packages
- Much improved designer
- C# as a language
- Can reference all .Net assemblies.
- Can use .Net 3.5, C# 3.0 features



ADO.Net Support

- Custom UI for data for Source Adapter
 - Drop down for list of tables
 - Invoke query builder
 - “expression-able” for dynamic changes
- A destination adapter
 - Custom UI
 - ODBC
- Source and Destination available in I/E Wizard



New Lookup with Caches

Explicit control over sharing and lifetime of reference data

- A connection manager stores the cache.
- All similar lookups in the same package can share the cache.
- Cache can be populated in one dataflow task and used in another
- Cache can be populated from any part of a dataflow (before being used for a lookup)
- Cache can be persisted to a file and re-used later
 - Get most used data once a week

Other nice features

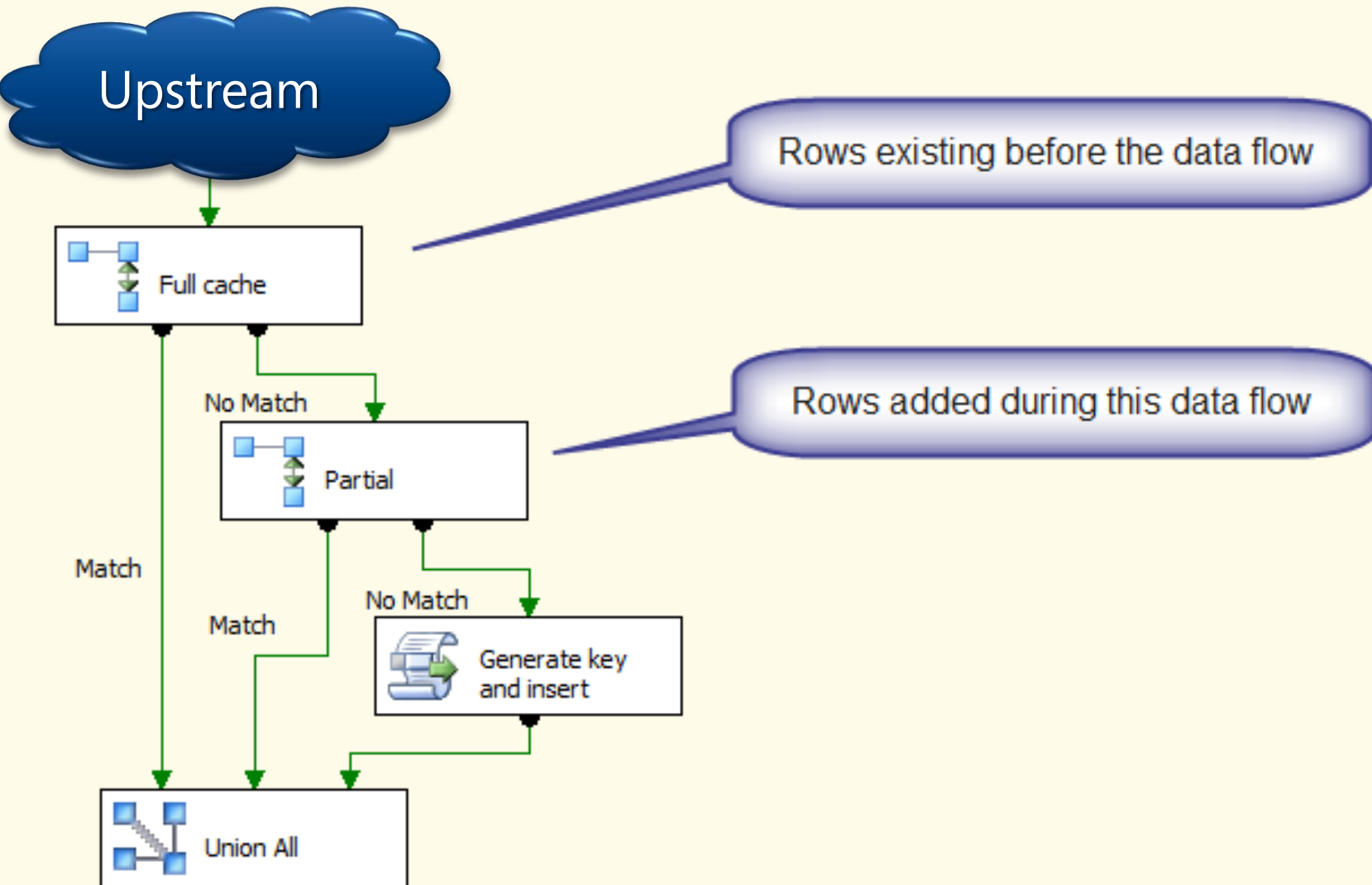
- A Cache for misses. Don't keep querying for rows that don't exist
- A missed row output, not the error output.
- Persisted cache files look just like RawFile format

Benefits

- Just retrieve and cache reference data once when many lookups use the same reference table
- Use ADO.Net or flat file as a source.
- Save caches locally until they get stale.
- Cache most common data (80/20 rule) and save it. Use cascaded lookups to get the infrequent rows via query.
- Use a complex data flow to fill the cache used by other data flows



Lookup – Cascading with NoMatch



Relevant SQL enhancements

New SQL 'MERGE' syntax

- Does UPDATE or INSERT based on key match
 - Or DELETE or combinations
- Handles SCD1 and SCD2 scenarios
- MERGE <table> USING <query> ON <key condition>
WHEN MATCHED THEN UPDATE SET <set columns>
WHEN NOT MATCHED THEN INSERT <values>
- Could replace SSIS SCD Wizard and Transform where appropriate.

Change Data Capture

- Find all of the changes to a table or set of tables
- Fully Integrated (SQL Syntax) part of SQL Server
- Log-based (no triggers)



demo

SSIS 2008 New Features

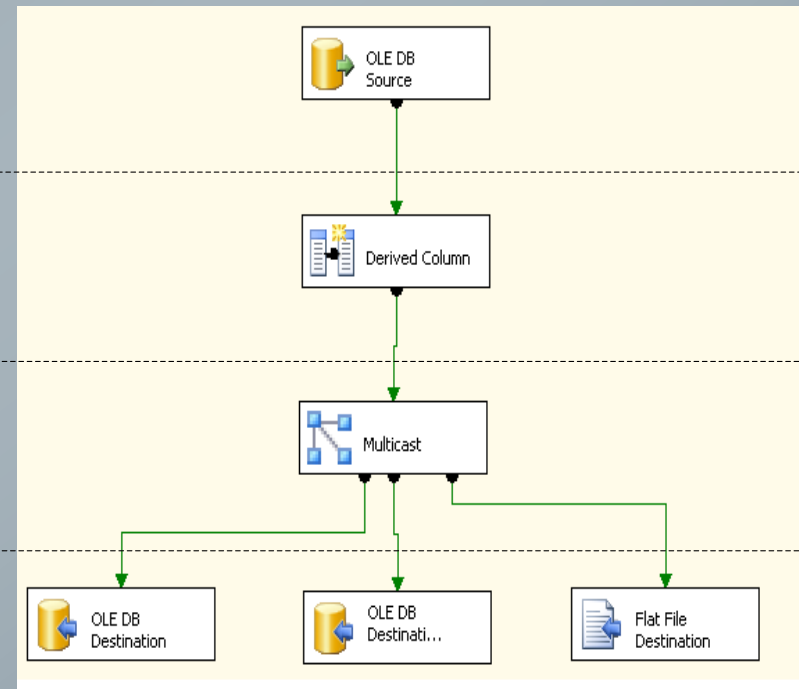
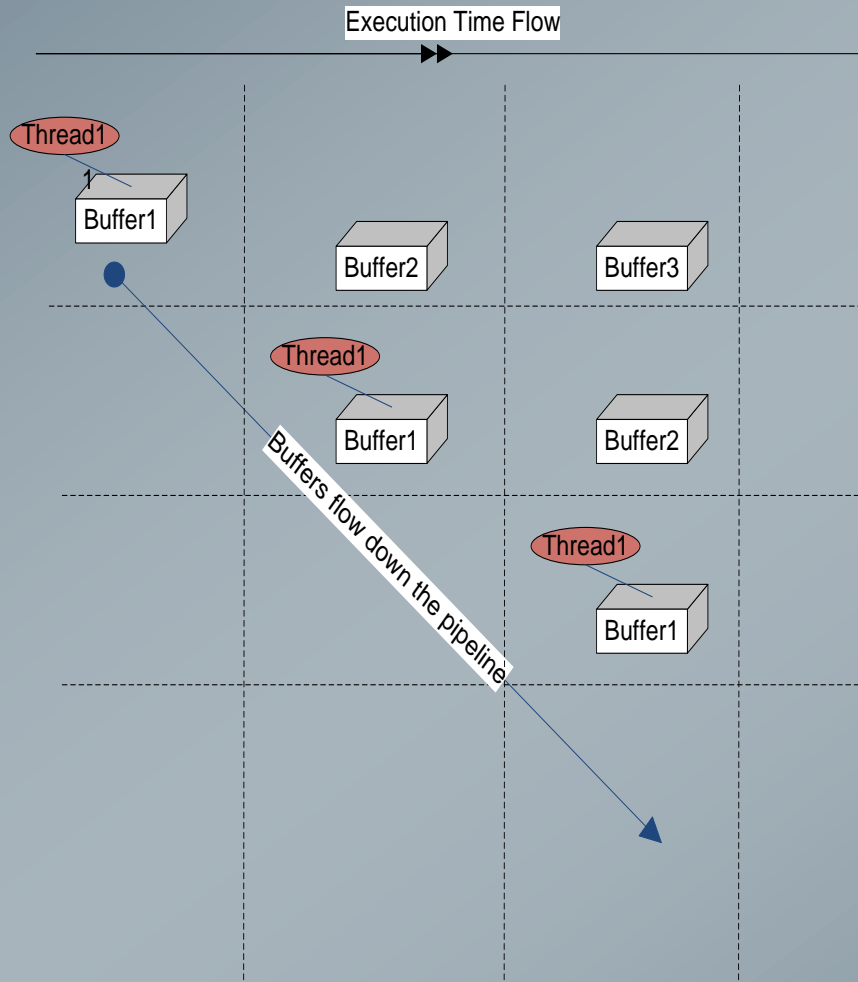
- ado.net
- c# and web services
- cache transform
- lookups



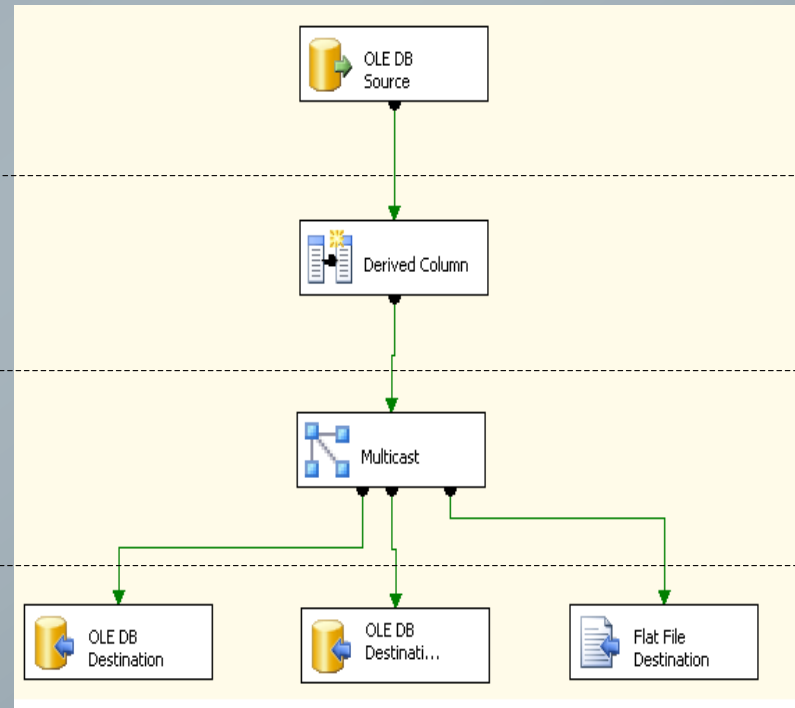
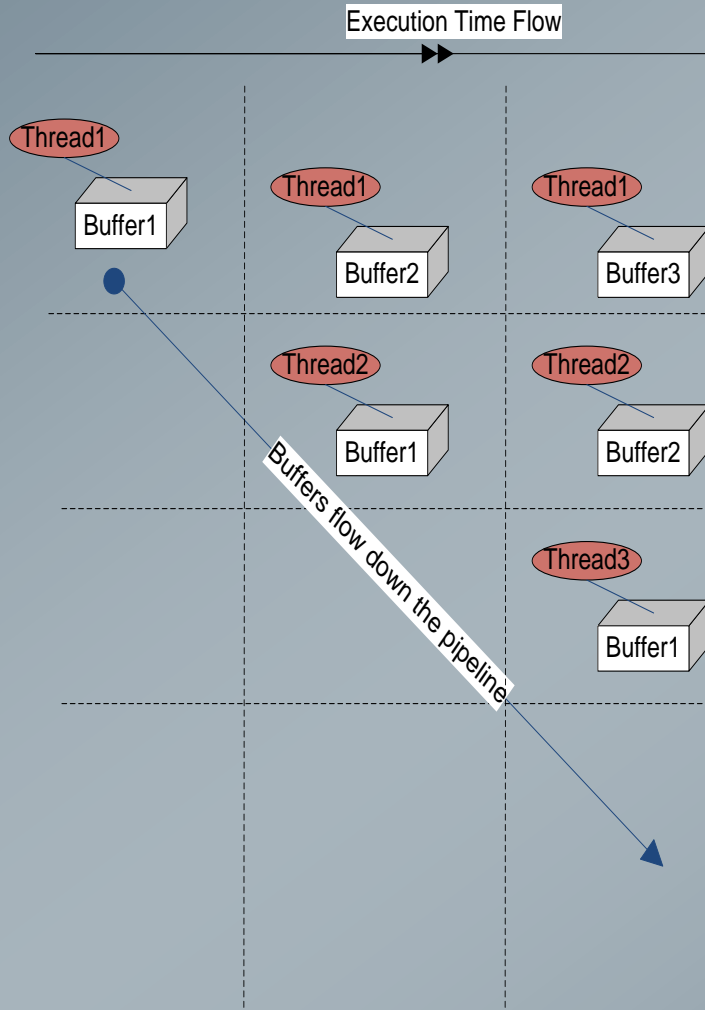
Data Flow Engine Scale-Up

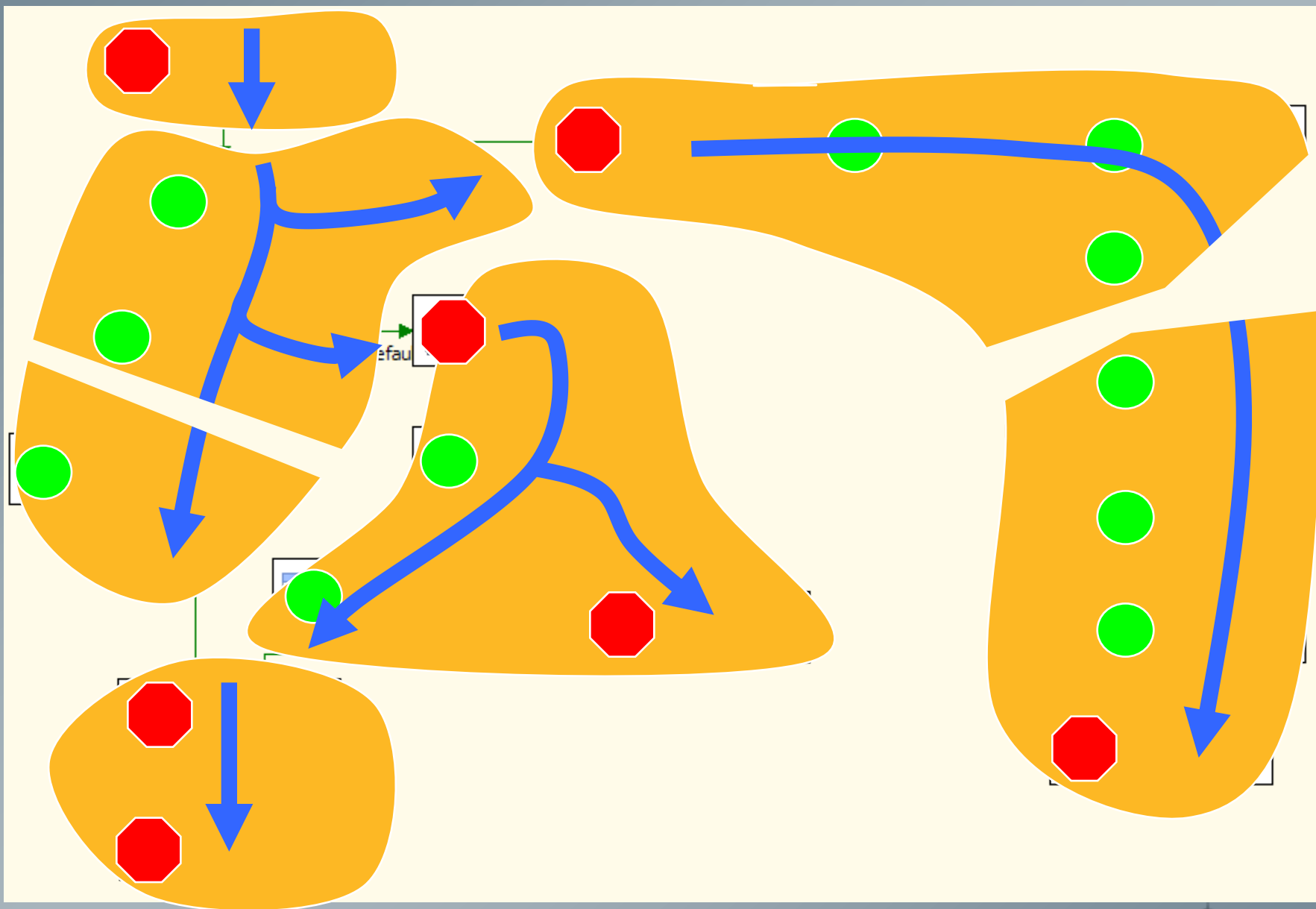


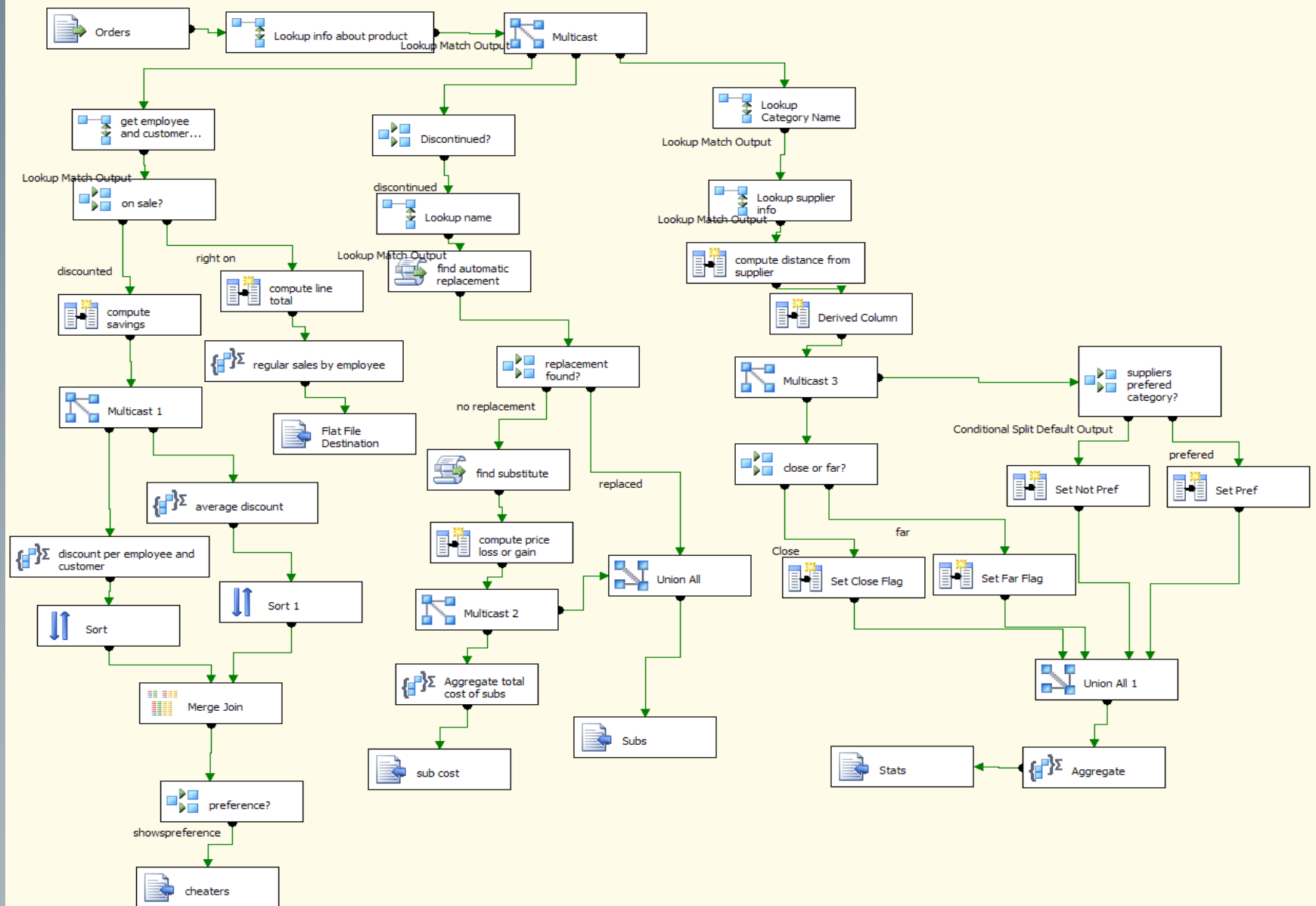
Thread Allocation - 2005



Thread Allocation - 2008

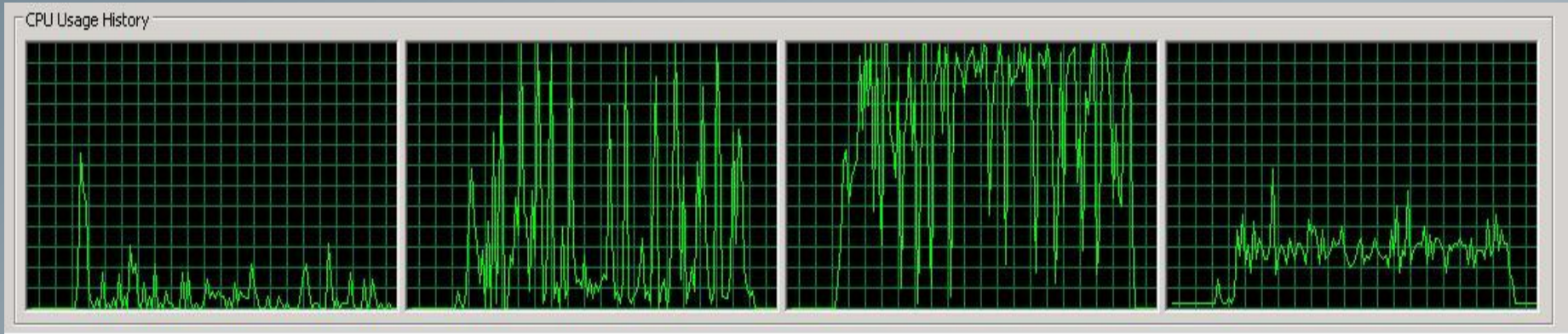




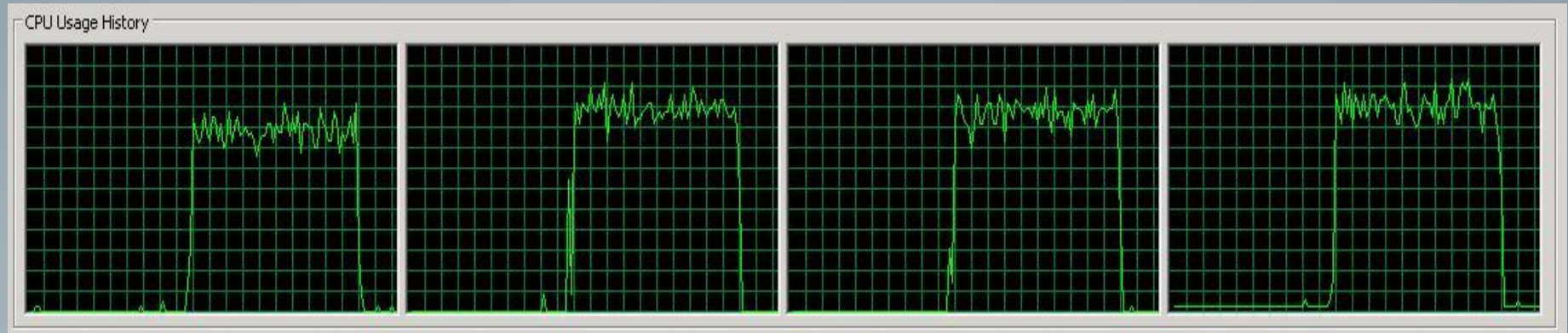


How Well Can it Work?

Yukon



Katmai



80% faster*

* Your mileage may vary



demo

SSIS Data Flow Scale Up



Business Intelligence →

Questions?



Resources

- Introductory Webcasts on SSIS
<http://www.microsoft.com/sql/technologies/integration/tours.mspix>
- Migrating Data Transformation Services Packages
<http://msdn2.microsoft.com/en-us/library/ms143501.aspx>
- Project REAL: Business Intelligence ETL Design Practices
<http://www.microsoft.com/downloads/details.aspx?familyid=705B03F3-1BBF-417F-9E63-92A00A4744E6&displaylang=en>
- SQL Server Integration Services site – links to blogs, training, partners, etc.:
<http://msdn.microsoft.com/SQL/sqlwarehouse/SSIS/default.aspx>
- SSIS MSDN Forum: <http://forums.microsoft.com/MSDN/ShowForum.aspx?ForumID=80&SiteID=1>
- SSIS MVP community site:
<http://www.sqlis.com>



Resources (cont)

- [TechNet Webcast: From DTS to SSIS: Making the Move to SQL Server Integration Services \(Level 200\)](#)
- [TechNet Webcast: Integrating Non-Standard Data Sources and Adding Value to Data with SQL Server 2005 Integration Services \(Level 200\)](#)
- [Extending SQL Server Integration Services with Custom Components \(Level 300\)](#)
- [Integrating Hand-coded ETL with SQL Server Integration Services \(Level 300\)](#)
- [TechNet Webcast: Troubleshooting SQL Server 2005 Integration Services \(Level 300\)](#)
- [TechNet Webcast: Deploying, Managing and Securing Integration Services \(Level 300\)](#)
- [Introducing SQL Server Integration Services for SQL Server 2005 \(Level 200\)](#)
- [TechNet Webcast: Scalability, Performance and Optimization in SSIS \(Level 300\)](#)
- [More SQL Server 2005 Webcasts](#)





Microsoft[®]

Your potential. Our passion.[™]

© 2008 Microsoft Corporation. All rights reserved. Microsoft, Windows, Windows Vista and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries. The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation.

MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.