Package Performance

Tuning Packages for Optimal Performance

Stacia Misner
blog.datainspirations.com
smisner@datainspirations.com
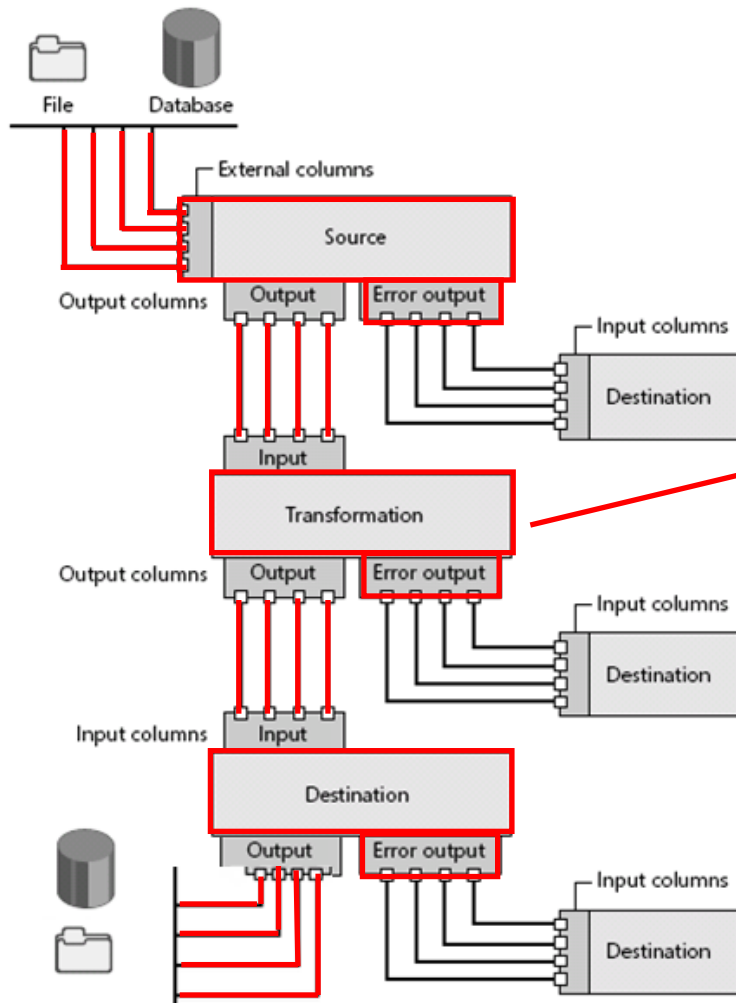
pluralsight
hardcore developer training

Overview

- **Pipeline Buffer Architecture**
- **Execution Trees and Execution Threads**
- **Package Performance Optimization**
- **Performance Monitoring**

Pipeline Buffer Architecture

Whalen, Garcia, Patel, Misner, Isakov, *Microsoft SQL Server 2005 Administrator's Companion* (Microsoft Press, 2007)

Buffers

> Location of BLOB data types storage in buffer not BLOB

DT_I4 4	DT_WSTR 50	DT_WSTR 25	DT_BOOL	DT_BOOL	DT_WSTR 15	DT_I2 2	DT_I2 2	DT_CY 8	DT_CY 8
	ProductID	Name	ProductNumber	MakeFlag	FinishedGoodsFlag	Color	SafetyStockLevel	ReorderPoint	StandardCost	ListPrice

	ProductID	Name	ProductNumber	MakeFlag	FinishedGoodsFlag	Color	SafetyStockLevel	ReorderPoint	StandardCost	ListPrice
1	1	Adjustable Race	AR-5381	0	0	NULL	1000	750	0.00	0.00
2	2	Bearing Ball	BA-8327	0	0	NULL	1000	750	0.00	0.00
3	3	BB Ball Bearing	BE-2349	1	0	NULL	800	600	0.00	0.00
4	4	Headset Ball Bearings	BE-2908	0	0	NULL	800	600	0.00	0.00
5	316	Blade	BL-2036	1	0	NULL	800	600	0.00	0.00
6	317	LL Crankarm	CA-5965	0	0	Black	500	375	0.00	0.00
7	318	ML Crankarm	CA-6738	0	0	Black	500	375	0.00	0.00
8	319	HL Crankarm	CA-7457	0	0	Black	500	375	0.00	0.00

```
4    +  100    +  50    + 1    + 1    + 30 + 2    + 2    + 8    + 8  = 201
                                                                       bytes
```

Column widths affect physical buffer size…

```
201 bytes          *    10,000              = 2MB   < DefaultMaxBufferSize (10MB)
Estimated Row Size  *    DefaultMaxBufferRows
```

Buffers

	ProductID	Name	ProductNumber	MakeFlag	FinishedGoodsFlag	Color	SafetyStockLevel	ReorderPoint	StandardCost	ListPrice
Row 1	x	xxxxxxx	xxxxxxx	x	x	xxxxx	xxx	xxx	xxx	xxxx
	x	xxxxxxx	xxxxxxx	x	x	xxxxx	xxx	xxx	xxx	xxxx
	x	xxxxxxx	xxxxxxx	x	x	xxxxx	xxx	xxx	xxx	xxxx
	x	xxxxxxx	xxxxxxx	x	x	xxxxx	xxx	xxx	xxx	xxxx
	:	:	:	:	:	:	:	:	:	:
Row 50,000	x	xxxxxxx	xxxxxxx	x	x	xxxxx	xxx	xxx	xxx	xxxx

	ProductID	Name	ProductNumber	MakeFlag	FinishedGoodsFlag	Color	SafetyStockLevel	ReorderPoint	StandardCost	ListPrice
Row 50,001	x	xxxxxxx	xxxxxxx	x	x	xxxxx	xxx	xxx	xxx	xxxx
	x	xxxxxxx	xxxxxxx	x	x	xxxxx	xxx	xxx	xxx	xxxx
	x	xxxxxxx	xxxxxxx	x	x	xxxxx	xxx	xxx	xxx	xxxx
	x	xxxxxxx	xxxxxxx	x	x	xxxxx	xxx	xxx	xxx	xxxx
	:	:	:	:	:	:	:	:	:	:
Row 100,000	x	xxxxxxx	xxxxxxx	x	x	xxxxx	xxx	xxx	xxx	xxxx

…To a maximum of 5 buffers receiving data from a source

Non-Blocking Streaming Transformations

Audit	Lookup (full cache)
Character Map	Multicast
Conditional Split	Percent Sampling
Copy Column	Row Count
Data Conversion	Script Component
Derived Column	(without asynchronous output)

ProductID	Name	ProductNumber	MakeFlag	FinishedGoodsFlag
x	ccccccc	xxxxxxx	x	x
x	ccccccc	xxxxxxx	x	x
x	ccccccc	xxxxxxx	x	x
x	ccccccc	xxxxxxx	x	x
:	:	:	:	:
x	ccccccc	xxxxxxx	x	x

These transforms operate on the same buffer from start to finish

OLEDB Source

DC Color

CNV Name

RC Count

OLEDB Stage
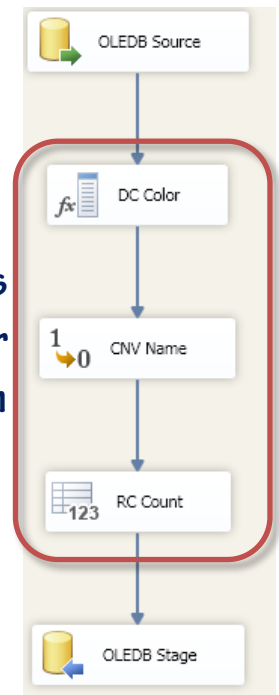
Non-Blocking Row-Based Transformations

Export Column	OLE DB Command
Import Column	Script (with external interaction)
Lookup (no cache or partial cache)	Slowly Changing Dimension

Lookup

Col A	Col B	Col C	Col D	Col E
x	xxxxxxx	xxxxxxx	c	x
x	xxxxxxx	xxxxxxx	c	x
x	xxxxxxx	xxxxxxx	c	x
x	xxxxxxx	xxxxxxx	c	x
:	:	:	:	:
x	xxxxxxx	xxxxxxx	c	x

OLE DB Command

OLEDB Source

49,150 rows

Lookup Price

Lookup Match Output (9,830 rows)

OLE DB Command
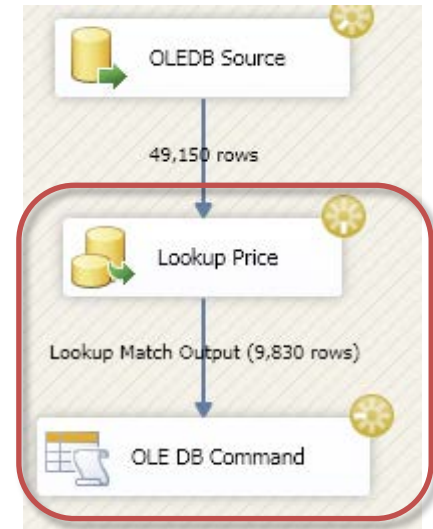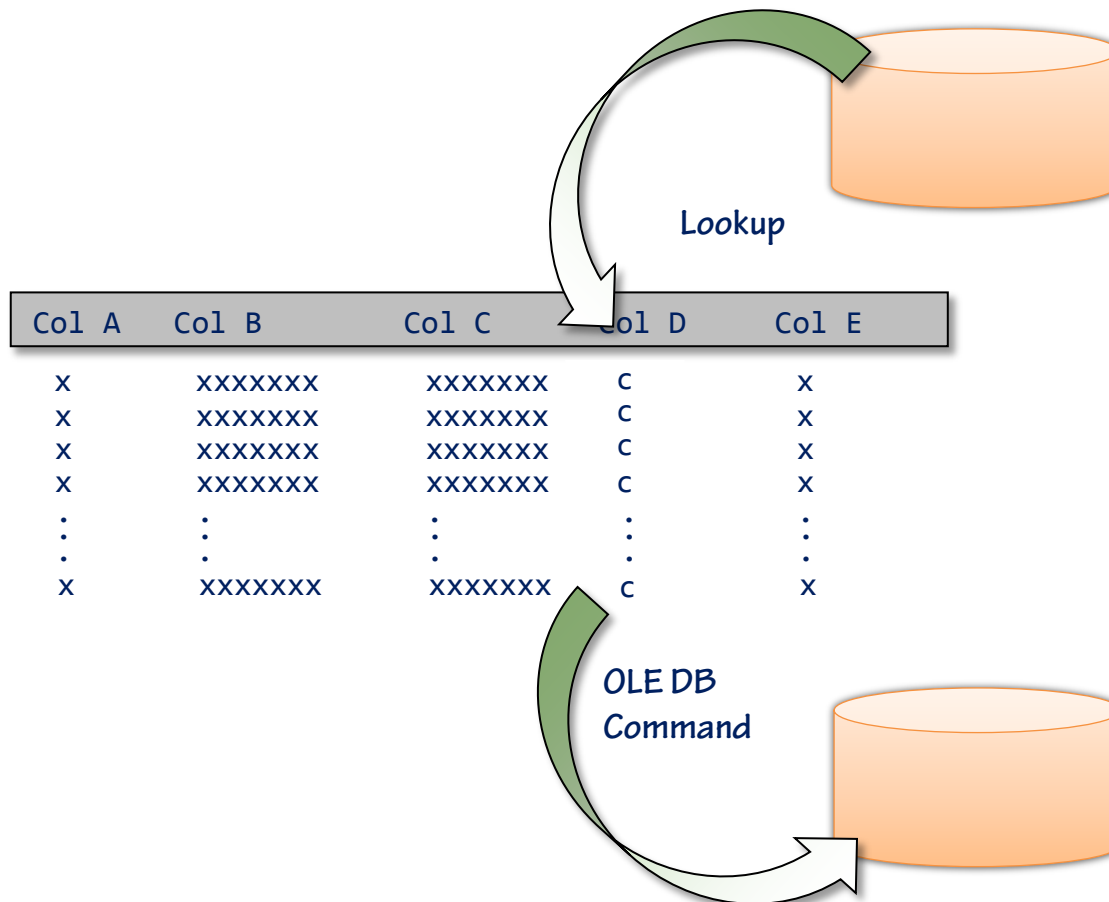
These transforms operate row by row but can't work on same buffer simultaneously

Partially-Blocking Transformations

Data Mining Query	Term Lookup
Merge	Unpivot
Merge Join	Union All
Pivot	

Col A	Col B	Col C	Col D
x	xxxxxxx	xxxxxxx	x
x	xxxxxxx	xxxxxxx	x
x	xxxxxxx	xxxxxxx	x
x	xxxxxxx	xxxxxxx	x
:	:	:	:
:	:	:	:
x	xxxxxxx	xxxxxxx	x

Col A	Col E
x	x
x	x
x	x
x	x
:	:
:	:
x	x

Merge Join

Col A	Col B	Col C	Col D	Col E
x	xxxxxxx	xxxxxxx	x	x
x	xxxxxxx	xxxxxxx	x	x
x	xxxxxxx	xxxxxxx	x	x
x	xxxxxxx	xxxxxxx	x	x
:	:	:	:	:
:	:	:	:	:
x	xxxxxxx	xxxxxxx	x	x

OLEDB Source Detail   OLE DB Source Header

Merge Join
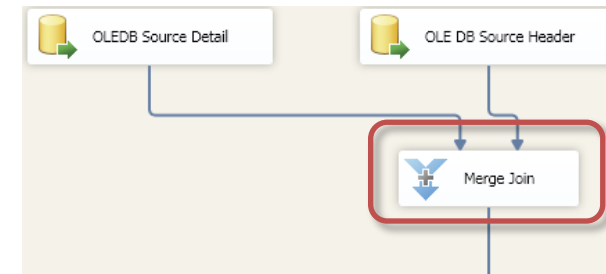
Partially-blocking transforms might wait on other inputs and always produce new buffer

Blocking Transformations

Aggregate	Sort
Fuzzy Grouping	Term Extraction
Fuzzy Lookup	Script Component (if code must read all rows)
Row Sampling	

Aggregate
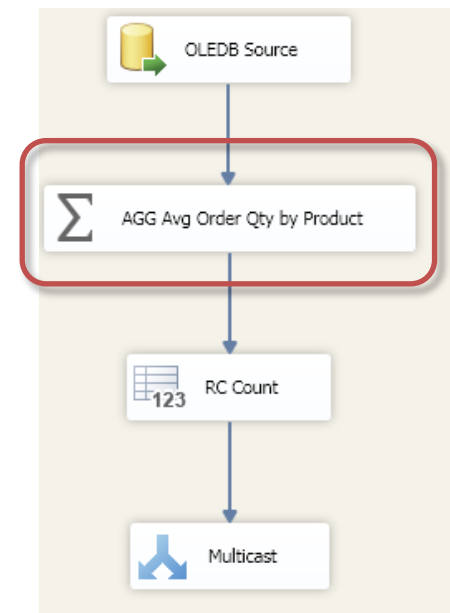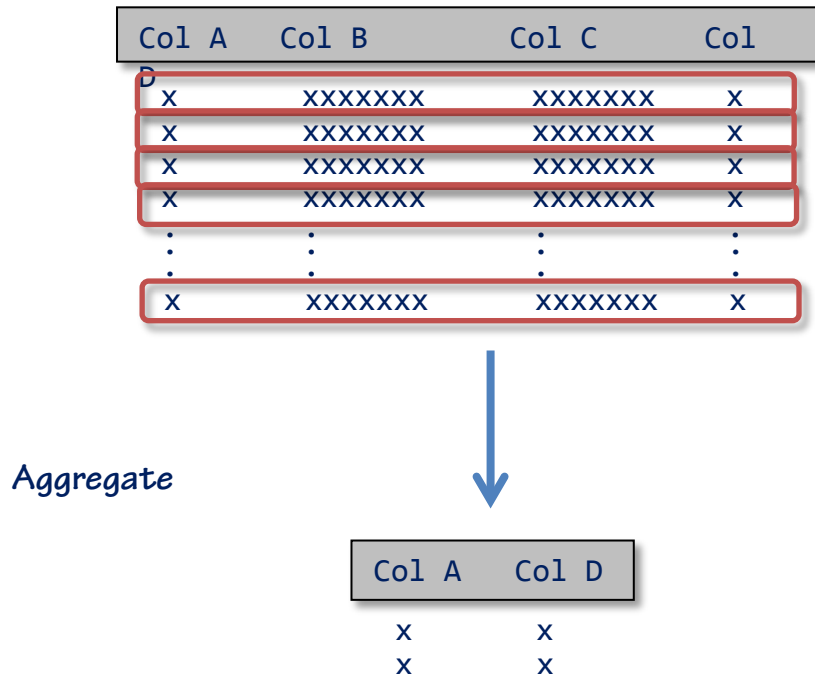
These transforms
stop pipeline until all rows read
and always produce new buffer

Synchronous versus Asynchronous Outputs
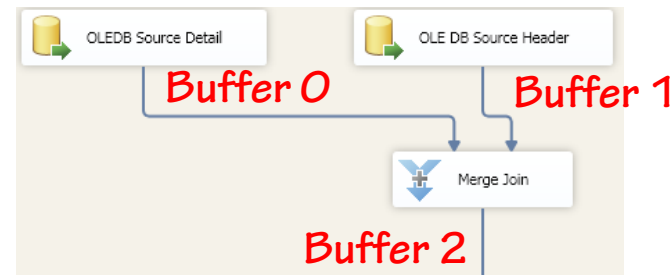
Non-blocking

Synchronous

Partially blocking
Blocking

Asynchronous

Operate on Same buffer

Buffer 0

Input rows = Output rows

Output into New buffer

Buffer 0   Buffer 1

Buffer 2

Buffer 0

Buffer 1
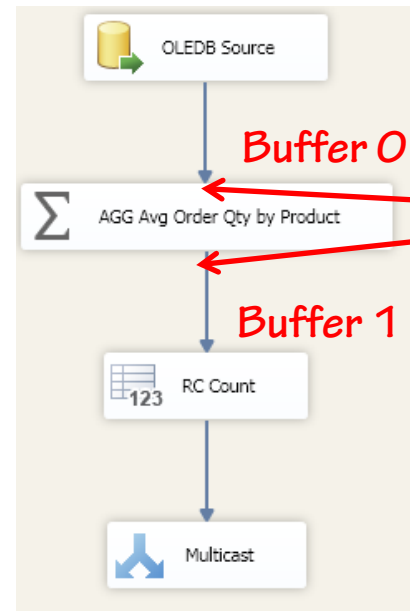
Input rows <> Output rows (usually)
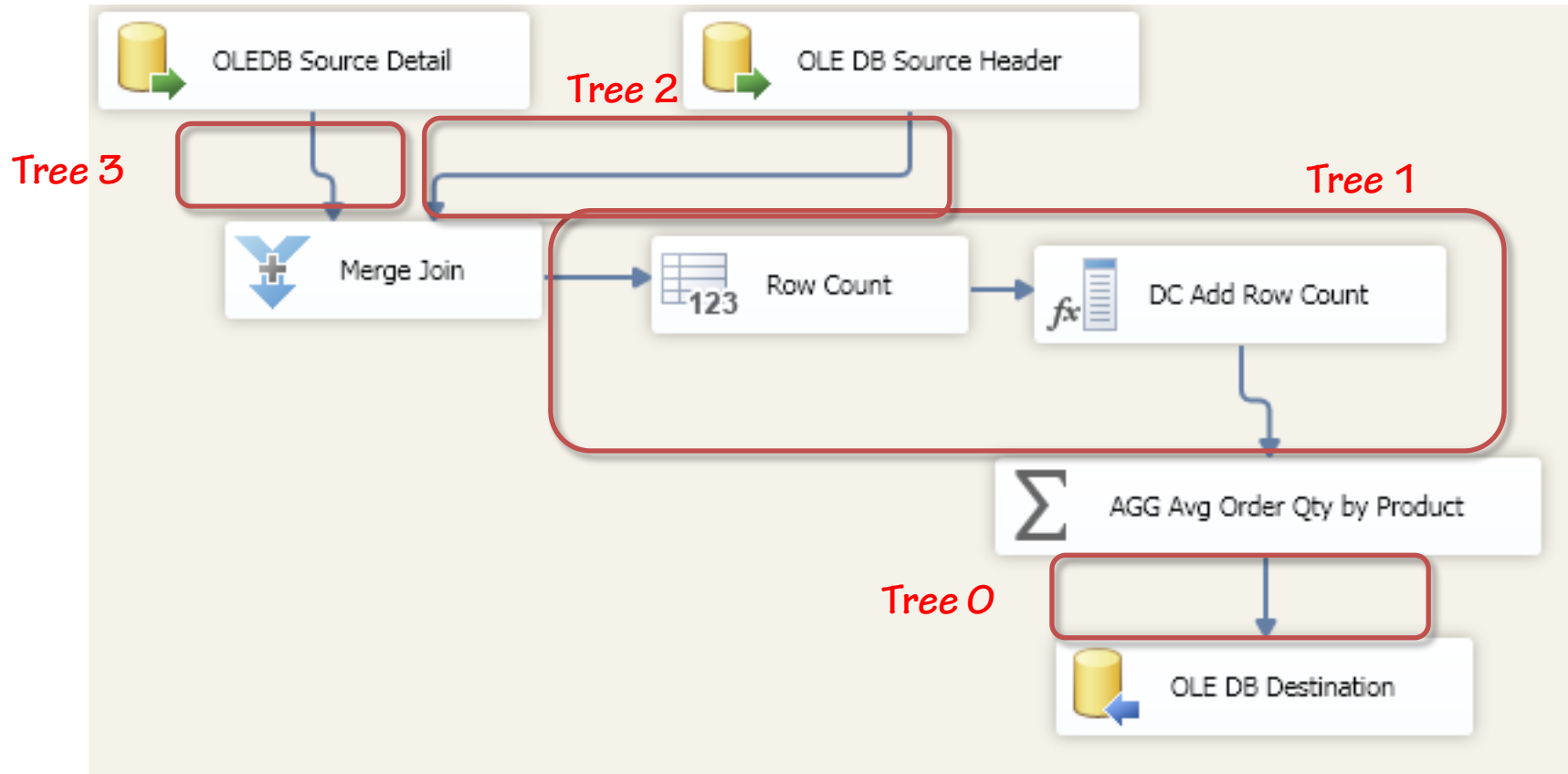
Execution Trees

Execution Trees

OLEDB Source Detail

OLE DB Source Header

Tree 1

Merge Join

Row Count

DC Add Row Count

Buffer size =
greatest number of columns x data length
at any point in execution tree

AGG Avg Order Qty by Product

OLE DB Destination
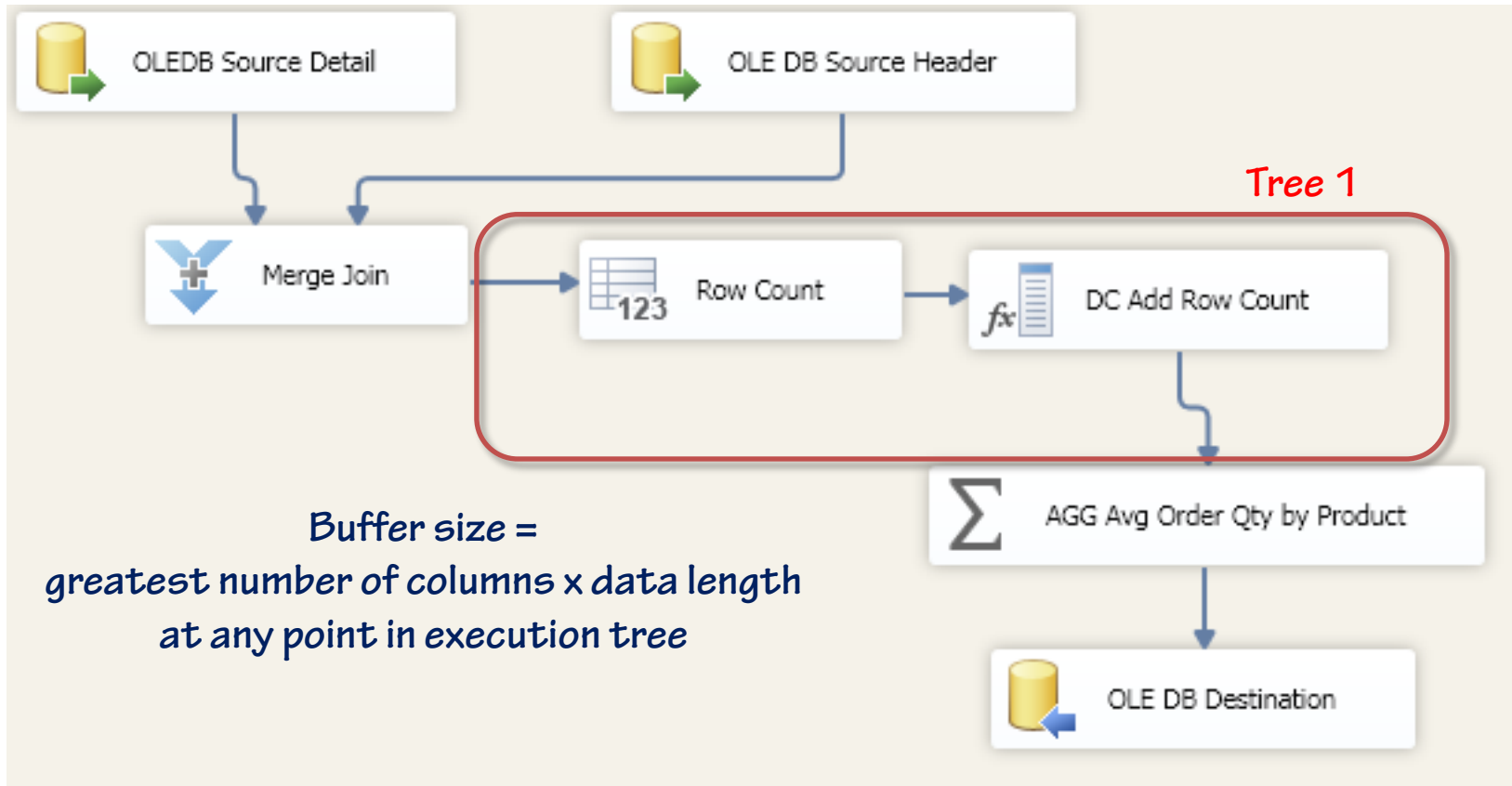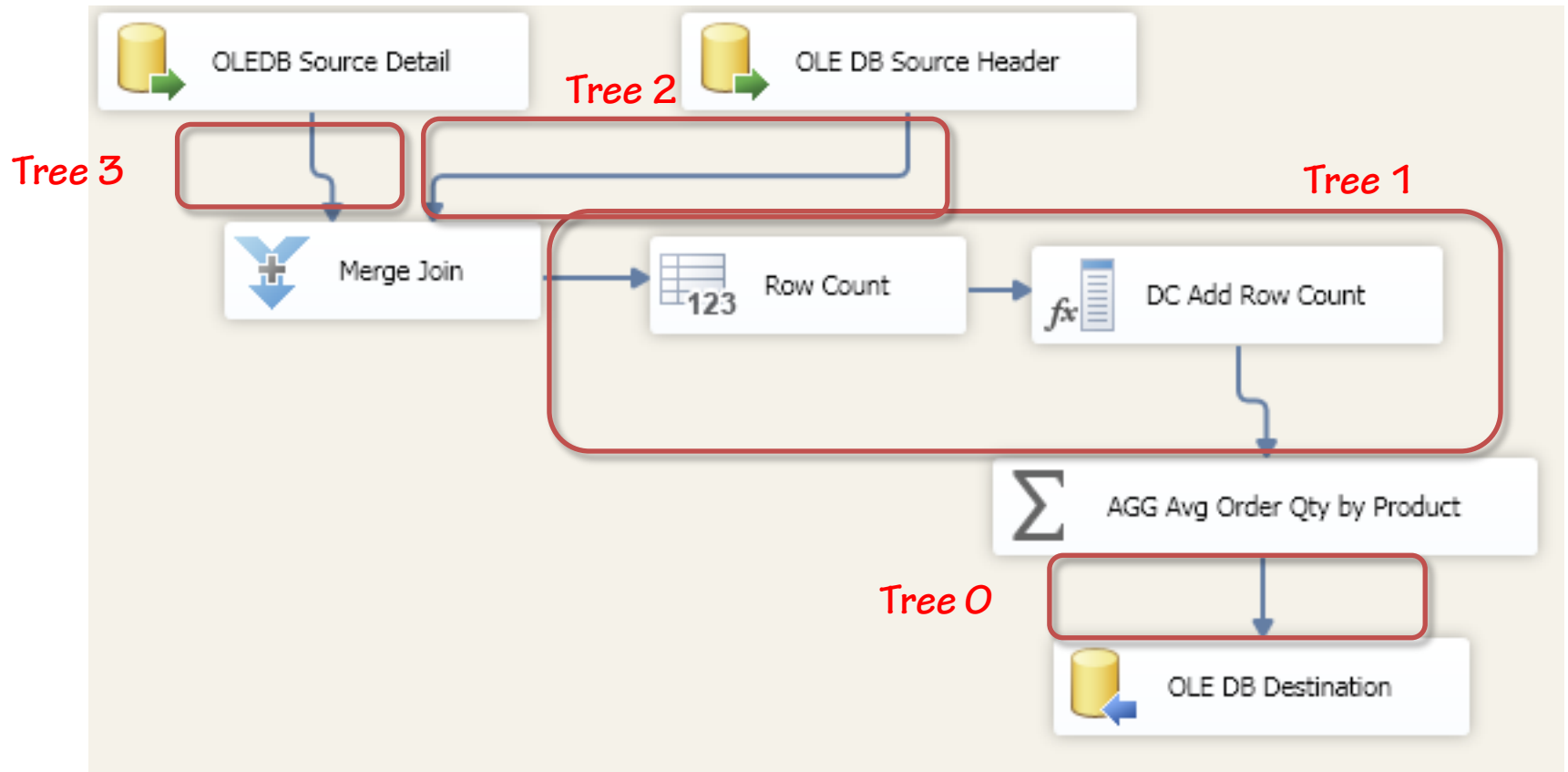
Execution Threads

Thread 0   Thread 3
Thread 1   Thread 4
Thread 2   Thread 5

Increase parallelism by increasing EngineThreads value
(Data Flow Task property)
Minimum: # of Execution Trees + Sources

Performance Optimization: Buffer Sizing

Data Flow Task Properties

Misc	
BufferTempStoragePath	
DefaultBufferMaxRows	10000
DefaultBufferSize	10485760

Estimated Row Size x **Max Buffer Rows** = **Estimated Buffer Size (EBS)**

Is EBS > 65,536 bytes (min) and < DefaultBufferSize? — Yes → **Actual Buffer Size = Multiple of 65,536 bytes**

No ↓

Is EBS > DefaultBufferSize? — Yes → **Fewer rows per buffer, More buffers**

No ↓

Is EBS < 65,536 bytes? — Yes → **More rows per buffer**
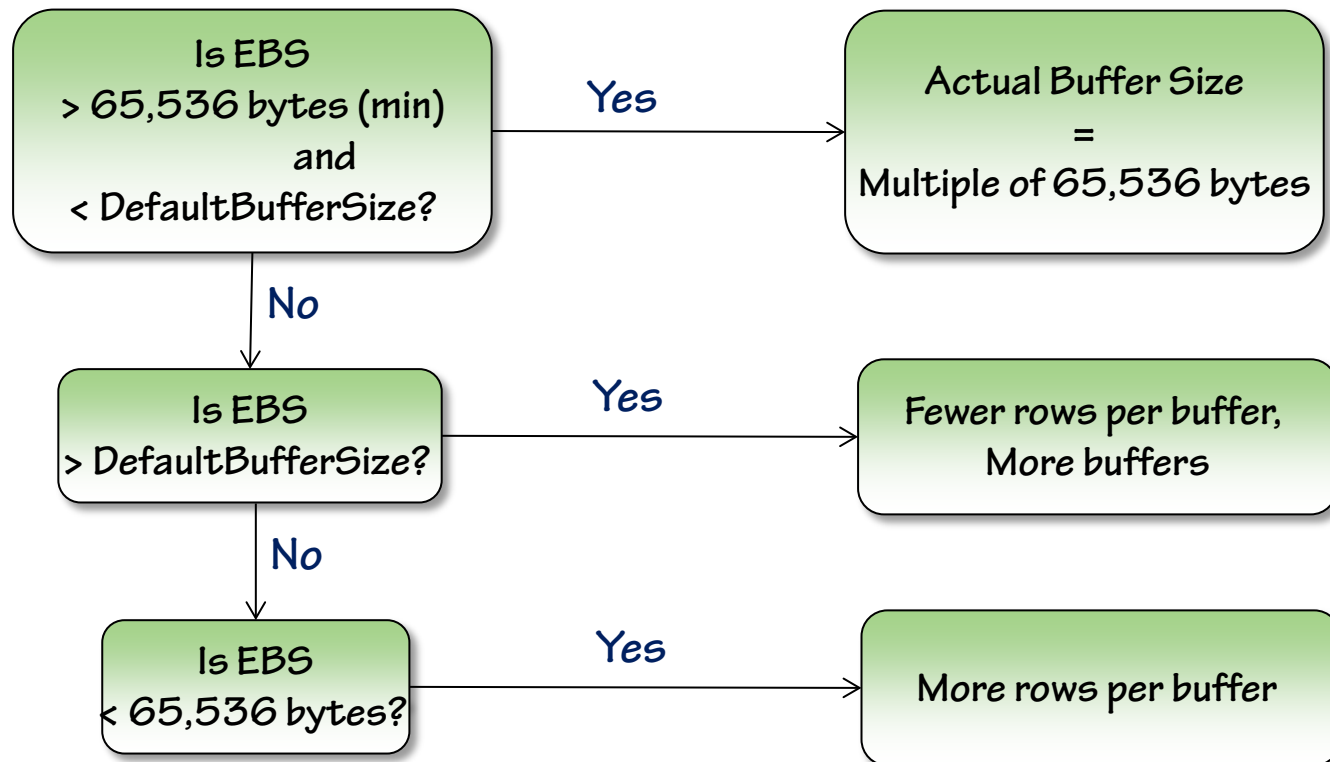
Performance Optimization: Buffer Sizing

- **Reduce the number of columns and use smallest data types possible**
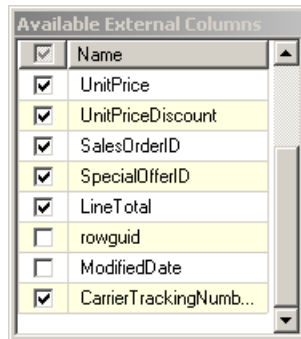
Use query conversions to reduce data length
if possible

- **Warnings alert you to unused columns in data flow**

Performance Optimization: Buffer Sizing

- **Log package with BufferSizeTuning to see adjustment of buffers**

☑ Events	Description
☐ OnPostExecute	Handles post-execution events. Use to define post-...
☐ OnPostValidate	Handles post-validation events. Use to define post-...
☐ OnPreExecute	Handles pre-execution events. Use to define pre-pr...
☐ OnPreValidate	Handles pre-validation events. Use to define pre-pr...
☐ OnProgress	Handles progress notifications. Use to define action...
☐ OnQueryCancel	Handles cancel events. Called periodically to determ...
☐ OnTaskFailed	Handles task failures. Use to define actions to perfo...
☐ OnVariableValueChanged	Handles value changes in variables whose RaiseCha...
☐ OnWarning	Handles warning events. Use to define actions whe...
☑ BufferSizeTuning	The Data Flow changed the size of a buffer away fr...
☐ PipelineBufferLeak	After execution, some buffers were left outstandin...
☐ PipelineComponentTime	Information about each phase of validation and exe...
☐ PipelineExecutionPlan	The execution plan of the data flow.
☐ PipelineExecutionTrees	The input the scheduler had when forming the exec...
☐ PipelineInitialization	Various information from Data Flow initialization.

Log the data flow task

```
User:BufferSizeTuning       DFT Transform-Load
Rows in buffer type 5 would cause a buffer size greater than the configured
maximum.
There will be only 15065 rows in buffers of this type.
```
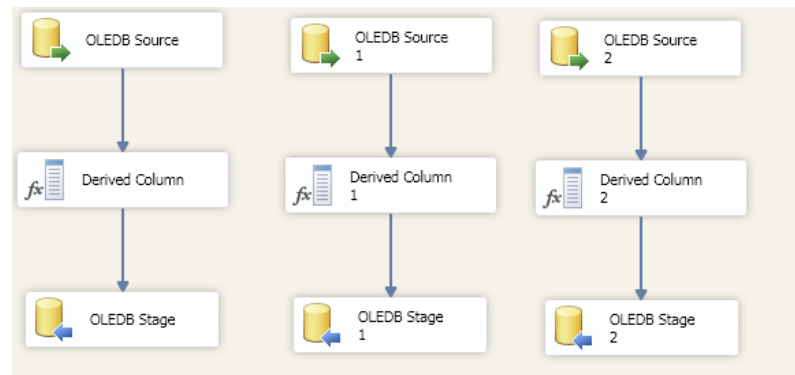
Package Performance: Parallelism

- **Increase MaxConcurrentExecutables on dedicated machine**

-1 = Number of processors +2

- **Use parallel pipelines in data flow with separate data sets**

Package Performance: Transformations

- **Choose transformations carefully**

Best choice: Non-blocking

Audit	Lookup (full *cache*)
Character Map	Multicast
Conditional Split	Percent Sampling
Copy Column	Row Count
Data Conversion	Script Component
Derived Column	(without asynchronous output)

- **Reduce rows with Conditional Split where possible**

Conditional Split
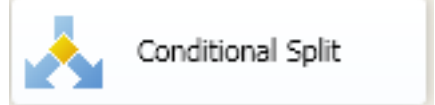
- **Optimize the Lookup cache**

Use results of an SQL query:

select ResellerKey, ResellerAlternateKey from dw.DimReseller
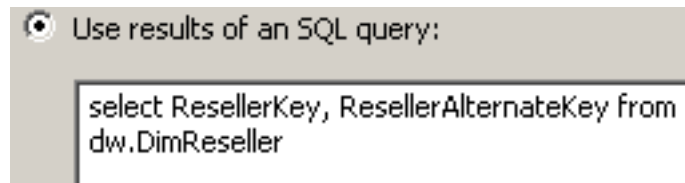
Package Performance: Sources

- **Data Access Mode: SQL Command for relational sources**
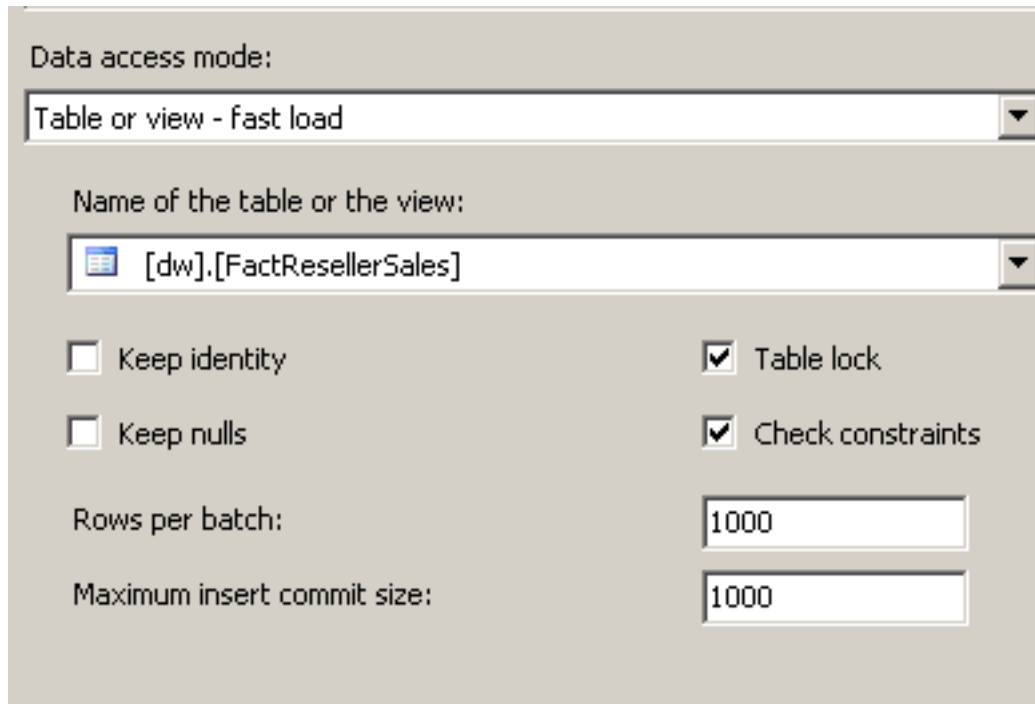
**Uses sp_executesql
(faster than OPENROWSET)**

- **FastParse for flat file sources**

**Use for when data types
Not required to be locale-specific**
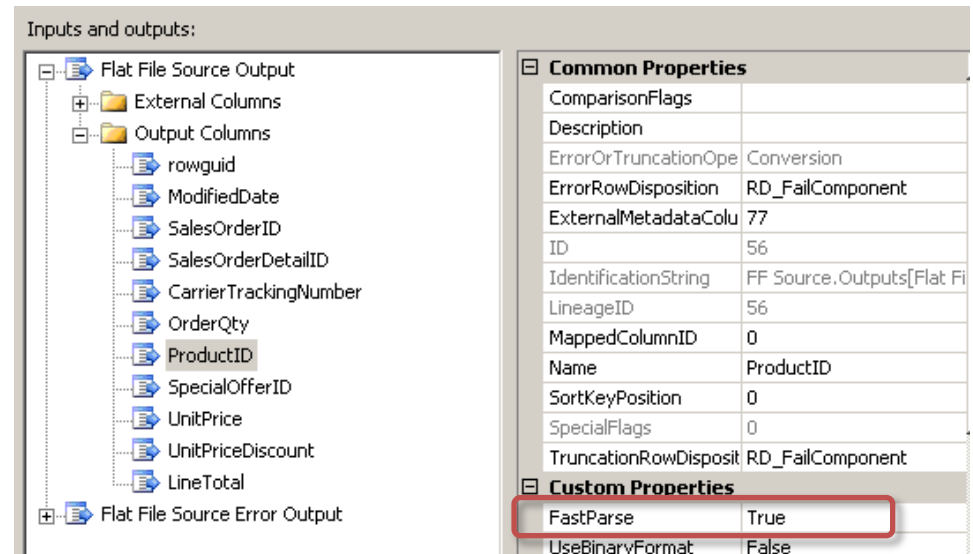
Package Performance: OLE DB Destinations
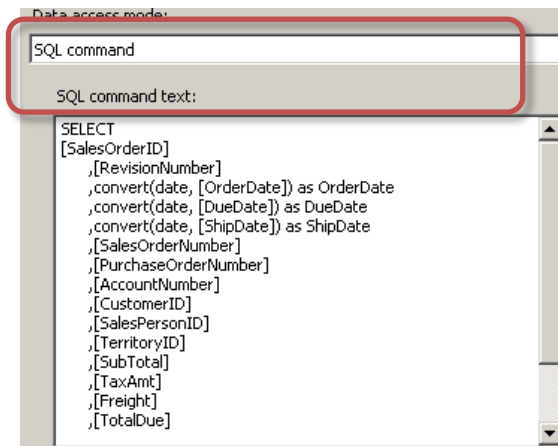
Data access mode:

Table or view - fast load

Name of the table or the view:

[dw].[FactResellerSales]

☐ Keep identity   ☑ Table lock

☐ Keep nulls   ☑ Check constraints

Rows per batch: 1000

Maximum insert commit size: 1000

Bulk insert with Fast Load option

Table lock performs faster

Clustered index?

    Set rows per batch for faster inserts

    Or…

Drop the index, load rows, and rebuild index

Performance Monitoring: Memory

Memory
Page Reads/sec   0.000

Page Reads/sec > 500 = memory pressure

Network Interface	Intel[R] PRO_1000 MT Network Connection _2	
Bytes Total/sec		0.000

Process	MsDtsSrvr	sqlservr
% Processor Time	0.000	0.000
Private Bytes	51,306,496.000	---
Working Set	22,401,024.000	---

SQLServer:Memory Manager

Total Server Memory (KB)	1,264,768.000

SQLServer:SSIS Pipeline 11.0

Buffer memory	13,835,752.000
Buffers in use	17.000
Buffers spooled	0.000
Flat buffer memory	8,232.000
Flat buffers in use	3.000

Performance Monitoring: CPU

Memory
    Page Reads/sec   0.000

Network Interface   Intel[R] PRO_1000 MT Network Connection _2
    Bytes Total/sec   0.000

Process	MsDtsSrvr	sqlservr
% Processor Time	0.000	0.000
Private Bytes	51,306,496.000	---
Working Set	22,401,024.000	---

SQLServer:Memory Manager
    Total Server Memory (KB)   1,264,768.000

SQLServer:SSIS Pipeline 11.0
    Buffer memory   13,835,752.000
    Buffers in use   17.000
    Buffers spooled   0.000
    Flat buffer memory   8,232.000
    Flat buffers in use   3.000

Performance Monitoring: Network

Memory
     Page Reads/sec   0.000

Network Interface   Intel[R] PRO_1000 MT Network Connection _2
     Bytes Total/sec   0.000

Process	MsDtsSrvr	sqlservr
% Processor Time	0.000	0.000
Private Bytes	51,306,496.000	---
Working Set	22,401,024.000	---

SQLServer:Memory Manager
     Total Server Memory (KB)   1,264,768.000

SQLServer:SSIS Pipeline 11.0
     Buffer memory   13,835,752.000
     Buffers in use   17.000
     Buffers spooled   0.000
     Flat buffer memory   8,232.000
     Flat buffers in use   3.000

Performance Monitoring: Integration Services

Memory
 Page Reads/sec   0.000

Network Interface   Intel[R] PRO_1000 MT Network Connection _2
 Bytes Total/sec   0.000

Process	MsDtsSrvr	sqlservr
% Processor Time	0.000	0.000
Private Bytes	51,306,496.000	---
Working Set	22,401,024.000	---

SQLServer:Memory Manager
 Total Server Memory (KB)   1,264,768.000

SQLServer:SSIS Pipeline 11.0
 Buffer memory   13,835,752.000
 Buffers in use   17.000
 Buffers spooled   0.000
 Flat buffer memory   8,232.000
 Flat buffers in use   3.000

Summary

- **Pipeline Buffers**
  - Buffers, transformations, synchronous and asynchronous outputs

- **Execution Trees and Execution Threads**
  - Asynchronous output starts new execution tree and allocation of at least one thread

- **Package Performance Optimization**
  - Buffer size, parallelism, transformation selection, and adapter optimization (sources, destinations)

- **Performance Monitoring**
  - Memory, CPU, network, Integration Services buffers (especially buffer spool)

Resources

- **Top 10 SQL Server Integration Services Best Practices**
  - *http://tinyurl.com/7hasrbm*
- **Integration Services: Performance Tuning Techniques**
  - *http://tinyurl.com/lzaeszr*
- **Integration Services Performance Counters**
  - *http://tinyurl.com/lk57stq*