

Maximizing SSIS Package Performance

17 quick tips for speeding up your SSIS packages



Maximizing SSIS Package Performance

1

About Me

Tim Mitchell



- Data architect, consultant, and author
- Microsoft Data Platform MVP
- Dallas, Texas area
- TimMitchell.net
- @Tim_Mitchell
- TimMitchell.net/Newsletter

Tyleris Data Solutions

- Data architecture
- Business intelligence
- Data warehousing
- ETL
- Analytics
- Training
- Tyleris.com



Maximizing SSIS Package Performance

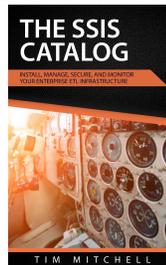
2

About Me

Recently published:

The SSIS Catalog

Install, Manage, Secure, and Monitor your Enterprise ETL Infrastructure



Maximizing SSIS Package Performance

4

TIP #1: TUNE YOUR SOURCES AND DESTINATIONS



Maximizing SSIS Package Performance

4

Tip #1: Tune your sources and destinations

- Many performance problems in SSIS aren't related to SSIS at all
- Sources and destination issues are often to blame



Tip #1: Tune your sources and destinations

- Inefficient data retrieval queries
- Index issues
- Network speed/latency
- Disk I/O



TIP #2: USING OPTION (FAST <N>)



Maximizing SSIS Package Performance

7

Tip #2: Using OPTION (FAST <n>)

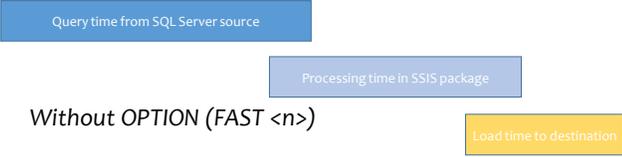
- Directs the query to return the first <n> rows as quickly as possible


```
SELECT FirstName, LastName, Address, City, State, Zip
FROM dbo.People
OPTION (FAST 10000)
```
- Not intended to improve the overall performance of the query



Tip #2: Using OPTION (FAST <n>)

- Useful for packages that spend a lot of time processing data in data flow

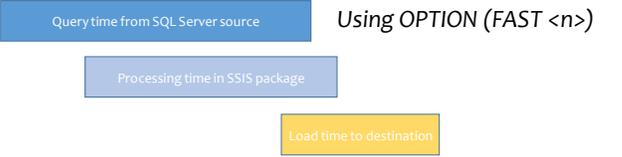


Without OPTION (FAST <n>)



Tip #2: Using OPTION (FAST <n>)

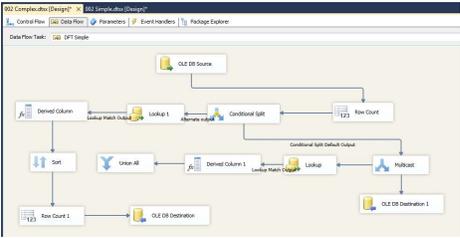
- Useful for packages that spend a lot of time processing data in data flow



Using OPTION (FAST <n>)



Tip #2: Using OPTION (FAST <n>)




TIP #3: BLOCKING TRANSFORMATIONS



Maximizing SSIS Package Performance

12

Tip #3: Blocking transformations

- Know the blocking properties of transformations
 - Nonblocking
 - Partially blocking
 - Fully blocking



Tip #3: Blocking transformations

- Nonblocking – no holding buffers
 - Row count
 - Derived column transformation
 - Conditional split



Tip #3: Blocking transformations

- Partially blocking – some buffers can be held as needed
 - Merge Join
 - Lookup
 - Union All



Tip #3: Blocking transformations

- Fully blocking – everything stops until all data is received
 - Sort
 - Aggregate
 - Fuzzy grouping/lookup



Tip #3: Blocking transformations

- Partially or fully blocking transforms are not inherently bad! Pick the right tool for every job.



TIP #4: SLOW TRANSFORMATIONS



Tip #4: Slow transformations

- Some transformations are often slow by nature
 - Slowly Changing Dimension wizard
 - OleDB Command
 - Asynchronous script components



Tip #4: Slow transformations

- Useful for certain scenarios, but should not be considered go-to tools for transforming data



TIP #5: AVOID THE TABLE LIST

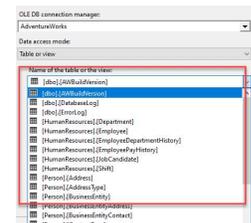


Maximizing SSIS Package Performance

21

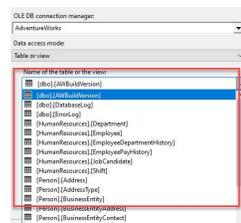
Tip #5: Avoid the table list

- Table drop-down list on data sources (OLEDB, ADO.NET)
- Uses SELECT *



Tip #5: Avoid the table list

- Can result in unnecessary columns
- A narrow buffer is happy buffer



Tip #5: Avoid the table list

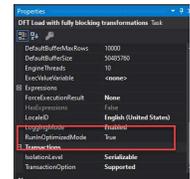
- Use the query window to select from table, specifying only the required columns
- Writing the query can be a reminder to apply filtering via WHERE clause, if appropriate



TIP #6: RUNINOPTIMIZEDMODE

Tip #6: RunInOptimizedMode

- Data flow setting to prevent the allocation of memory for unused columns
- Note that you'll still get a warning from SSIS engine when executing the package



TIP #7: TRANSFORM DATA IN SOURCE

Tip #7: Transform data in source

- When dealing with relational data, consider transforming in source
- Let the database engine do what it already does well

Tip #7: Transform data in source

- Filtering
- Sorting
- Lookups/joins
- Aggregations



TIP #8: CONCURRENT EXECUTABLES

Tip #8: Concurrent executables

- Package-level setting to specify how many executables (tasks) can be running at once in a given package
- Default = -1
 - Number of logical processors + 2



Tip #8: Concurrent executables

- For machines with few logical processors or potentially many concurrent executables, consider increasing this value



Tip #8: Concurrent executables

Demo – Concurrent Executables



TIP #9: GO PARALLEL



Maximizing SSIS Package Performance

34

Tip #9: Go parallel

- Many operations in SSIS are done serially
- For nondependent operations, consider allowing processes to run in parallel



Tip #9: Go parallel

- Dependent on machine configuration, network environment, etc.
- Can actually *hurt* performance
- Testing, testing, testing!



Tip #9: Go parallel

Demo – Show Parallel Design



TIP #10: CONSIDER NON-SSIS OPTIONS



Maximizing SSIS Package Performance

38

Tip #10: Consider non-SSIS options

- Sometimes a non-SSIS solution will perform better than SSIS
- If all you have is a hammer...



Tip #10: Consider non-SSIS options

- Some operations are better suited for T-SQL or other tools:
 - MERGE upsert
 - INSERT... SELECT
 - Third party components
 - External applications (via Execute Process Task)



TIP #11: WATCH YOUR SPOOLS



Maximizing SSIS Package Performance

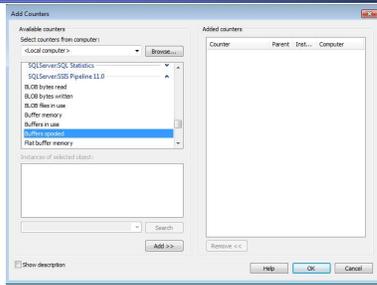
41

Tip #11: Watch your spools

- Buffers spooled = writing to physical disk
- Usually indicates memory pressure
- Keep an eye on PerfMon counters for SSIS: Buffers Spooled



Tip #11: Watch your spools



Tip #11: Watch your spools

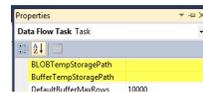
- Any value above zero should be investigated
- Keep it in memory!



TIP #12: PREPARE FOR PAGING BUFFERS

Tip #12: Prepare for paging buffers

- Plan for no paging of buffers to disk. But...
- Build for it if (when?) it happens
 - BLOBTempStoragePath
 - BufferTempStoragePath
- Fast disks if possible



TIP #13: BUFFER SIZING

Tip #13: Buffer sizing

- Data flow task values:
 - DefaultBufferSize
 - DefaultBufferMaxRows
- Ideally, use a small number of large buffers
- Generally, max number of active buffers = 5

Tip #13: Buffer sizing

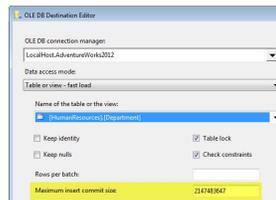
- Buffer size calculation:
 - $(\text{Row size} * \text{est num of rows}) / \text{DefaultMaxBufferRows}$



TIP #14: MAXIMUM INSERT COMMIT SIZE

Tip #14: MaximumInsertCommitSize

- OleDbDestination setting
- Controls how buffers are committed to the destination database



Tip #14: MaximumInsertCommitSize

MICS > buffer size	Setting is ignored. One commit is issued for every buffer.
MICS = 0	The entire batch is committed in one big batch.
MICS < buffer size	Commit is issued every time MICS rows are sent. Commits are <i>also</i> issued at the end of each buffer.

Source: Data Loading Performance Guide [http://technet.microsoft.com/en-us/library/dd425079\(v=sql.10e\).aspx](http://technet.microsoft.com/en-us/library/dd425079(v=sql.10e).aspx)

Tip #14: MaximumInsertCommitSize

- Note that this does NOT impact the size of the buffer in SSIS
- In most cases, a small number of large commits is preferable to a large number of small commits

Tip #14: MaximumInsertCommitSize

- Larger commit sizes = fewer commits
- Good:
 - Potentially less database overhead
 - Potentially less index fragmentation
- Bad:
 - Potential for log file or TempDB pressure/growth

TIP #15: MANAGE YOUR LOOKUPS



Maximizing SSIS Package Performance

55

Tip #15: Manage your lookups

- Lookup cache modes
 - Full (default)
 - Partial
 - None



Tip #15: Manage your lookups

- Full cache:
 - Entire contents of lookup query is written to memory before processing begins



Tip #15: Manage your lookups

- Partial cache:
 - Lookup rows are written to cache after initial use, and then reused if the same value is encountered



Tip #15: Manage your lookups

- No cache:
 - Every row is a separate lookup
 - RBAR!



Tip #15: Manage your lookups

Demo – Lookups and caching



TIP #16: SHARE THE ROAD



Maximizing SSIS Package Performance

61

Tip #16: Share the road

- Strategically schedule packages to avoid contention with other packages, external processes, etc.
 - Consider a queuing pattern in SSIS



Maximizing SSIS Package Performance

62

Tip #16: Share the road

- Resource contention is a key player in SSIS performance issues
 - Other SSIS or ETL processes
 - External processes
 - The operating system itself



Maximizing SSIS Package Performance

63

TIP #17: STAGE YOUR DATA



Maximizing SSIS Package Performance

64

Tip #17: Stage your data

- SQL-to-SQL operations perform well
- Can't do direct SQL to SQL with nonrelational or external data



Maximizing SSIS Package Performance

65

Tip #17: Stage your data

- Staging the data can allow for faster, direct SQL operations
 - *Remember: Let the database engine do what it does well.*
- Updates in particular



Maximizing SSIS Package Performance

66

Tip #17: Stage your data

Demo – Staged Update



Additional Resources

- SSIS Performance Design Patterns whitepaper:
<https://www.tyleris.com/publications/>



Additional Resources

- Rob Farley demonstrates FAST hint:
<http://bit.ly/eYNIMg>
- Microsoft data loading performance guide:
<http://bit.ly/17xjgbw>



Thanks!

Tyleris.com

TimMitchell.net/Newsletter

Twitter: @Tim_Mitchell



Maximizing SSIS Package Performance