**Cognizant Technology Solutions**

**SSIS Best Practices**

**Version 1.0**



**Controlled Copy**



|  |  |  |  |
| --- | --- | --- | --- |
|  | **Prepared By** | **Reviewed by** | **Approved By** |
| **Name** | Venkatesh, V | Vasantha Madhavan,  Srividya |  |
| **Role** | Team Member | Team Lead |  |
| **Signature** |  |  |  |
| **Date** | 10/21/2010 | 10/22/2010 |  |

Project ID: *1000043700 <SCI.ID. > /* Ver: *1.0*

Release ID: **C3: Protected**

## Table of Contents

[1.0 Purpose 1](#_Toc24524)

[2.0 Scope 1](#_Toc24525)

[3.0 Intended Audience 1](#_Toc24526)

[4.0 Software Components 1](#_Toc24527)

[5.0 SSIS Best Practices 1](#_Toc24528)

[6.0 Naming Conventions 4](#_Toc24529)

[7.0 Config Files Parameters: 7](#_Toc24530)

[8.0 Change Log 8](#_Toc24531)

**Controlled Copy Technical Proof of Concept Document**



2

# 1.0 Purpose

The purpose of this document is to provide information on the Best Practices to be followed during the design and development of an SSIS package.

# 2.0 Scope

The scope of the document is to identify a set of best practices to be followed during the design and development of an SSIS package.

# 3.0 Intended Audience

The intended audience is:

 Cognizant Project Managers and Project Teams.

# 4.0 Software Components

SQL Server Integration Services.

# 5.0 SSIS Best Practices

1. If the package is running on the same server as your destination SQL Server 2005 relational database instance, use the **SQL Server Destination** component instead of the OLE DB connection for SQL Server. This component runs in-process and thus avoids the overhead of a connection. It can be up to 25 percent faster than the OLE DB connection.
2. Remove unnecessary columns from the data flow to improve performance. The data flow engine warns the user about the output columns not used. Removing these columns saves the engine from allocating space and processing data that is not used.
3. The **EngineThreads** property on a task sets the number of threads used by the task. The default value is **5** but on a multi-processor server, this value can be set at a higher value to improve performance. The number of threads for optimal performance can be determined on the basis of testing.
4. On a **multi-processor server**, execute independent tasks in parallel. For example, for a Data Warehouse project, once the dimension tables in Data Warehouse are loaded, if the

fact tables access different dimension tables and can be loaded simultaneously, consider loading the data in parallel.

1. Though the correct number of tasks executed in parallel for optimal performance can be determined only after testing, you can start by setting the number of parallel tasks equal to the number of processors.
2. If you know that data in a source is sorted, set **IsSorted=TRUE** on the source adapter output. This may save unnecessary SORTs later in the pipeline which can be expensive. Setting this value does not perform a sort operation; it only indicates that the data it sorted.
3. Rename all Name and Description properties from the default. This will help when debugging particularly if the person doing the debugging is not the person that built the package.
4. Only select columns that you need in the pipeline to reduce buffer size and reduce OnWarning events at execution time
5. Always use a SQL statement in an OLE DB Source component or LOOKUP component rather than just selecting a table. Selecting a table is akin to "SELECT \*..." which is universally recognized as bad practice.
6. Use Sequence containers to organize package structure into logical units of work. This makes it easier to identify what the package does and also helps to control transactions if they are being implemented.
7. Where possible, use expressions on the SQLStatementType property of the Execute SQL Task instead of parameterized SQL statements. This removes ambiguity when different OLE DB providers are being used. It is also easier.
8. If you are implementing custom functionality try to implement custom tasks/components rather than use the script task or script component. Custom tasks/components are more reusable than scripted tasks/components. Custom components are also less bound to the metadata of the pipeline than script components are.
9. Use caching in your LOOKUP components where possible. It makes them quicker. Watch that you are not grabbing too many resources when you do this though.
10. LOOKUP components will generally work quicker than MERGE JOIN components where the 2 can be used for the same task.
11. Always use DTExec to performance test your packages. This is not the same as executing without debugging from SSIS Designer.
12. If you want to conditionally execute a task at runtime use expressions on your precedence constraints. Do not use an expression on the "Disable" property of the task.
13. Don't pull all configurations into a single XML configuration file. Instead, put each configuration into a separate XML configuration file. This is a more modular approach and means that configuration files can be reused by different packages more easily.
14. If you need a dynamic SQL statement in an OLE DB Source component, set AccessMode="SQL Command from variable" and build the SQL statement in a variable that has EvaluateAsExpression=TRUE.
15. When using checkpoints, use an expression to populate the CheckpointFilename property which will allow you to include the value returned from System::PackageName in the checkpoint filename. This will allow you to easily identify which package a checkpoint file is to be used by.
16. When using raw files and your Raw File Source Component and Raw File Destination Component are in the same package, configure your Raw File Source and Raw File Destination to get the name of the raw file from a variable. This will avoid hardcoding the name of the raw file into the two separate components and running the risk that one may change and not the other.
17. Variables that contain the name of a raw file should be set using an expression. This will allow you to include the value returned from System::PackageName in the raw file name. This will allow you to easily identify which package a raw file is to be used by. N.B. This approach will only work if the Raw File Source Component and Raw File Destination Component are in the same package.
18. Use variables to store your expressions. This allows them to be shared by different objects and also means you can view the values in them at debug-time using the Watch window.
19. If you can, filter your data in the Source Adapter rather than filter the data using a Conditional Split transform component.
20. Some Component properties that can be set for better performance are:

|  |  |
| --- | --- |
| **Component Name** | **Property** |
| Oledb Source, Lookup, Fuzzy Lookup | Remove columns which are not used downstream. |
| Oledb Source, Lookup, Fuzzy Lookup | Use a select statement when fetching data from a view instead of using the table drop down. |
| OLEDB Command Destination | For large number of rows, this component might not scale because it sends one SQL statement per row. Persist data to a temporary table and use set based SQL. |
| SCD | For large number of rows that will not exist in the dimension table, consider using a lookup before the SCD. |
| OLE DB Destination | If the server is local, consider using SQL Server destination. |
| OLEDB Destination | Review the fast load settings on the destination adapter. |
| Flat file source | Turn on fast parse for columns that have types that fast parse understands. |
| Conditional Split | For transforms that use conditions based on columns coming straight from OLEDB or ADO.Net sources, consider using the native filtering from the relational database to remove rows before they come in to the pipeline. |
| Flat file Source | For transforms that have all columns with the default column type and size, consider using the ‘Suggest Type’ functionality. |
| SQL Server | For OLEDB Destinations/SQL Server  Destinations that perform bulk insert into a database, verify that the logging mode is appropriate for performance. |

# 6.0 Naming Conventions

The following acronyms can be used as prefix to identify the tasks being used:

|  |  |
| --- | --- |
| **Task** | **Prefix** |
| For Loop Container | FLC |
| Foreach Loop Container | FELC |
| Sequence Container | SEQC |
| ActiveX Script | AXS |
| Analysis Services Execute DDL | ASE |
| Analysis Services Processing | ASP |
| Bulk Insert | BLK |
| Data Flow | DFT |
| Data Mining Query | DMQ |
| Execute DTS 2000 Package | EDPT |
| Execute Package | EPT |
| Execute Process | EPR |
| Execute SQL | EST |
| File System | FSYS |
| FTP | FTP |
| Message Queue | MSMQ |
| Script | SCR |
| Send Mail | SMT |
| Transfer Database | TDB |
| Transfer Error Messages | TEM |
| Transfer Jobs | TJT |
| Transfer Logins | TLT |
| Transfer Master Stored Procedures | TSP |
| Transfer SQL Server Objects | TSO |
| Web Service | WST |
| WMI Data Reader | WMID |
| WMI Event Watcher | WMIE |
| XML | XML |

The following acronyms can be used as prefix to identify the components being used:

|  |  |
| --- | --- |
| **Component** | **Prefix** |
| DataReader Source | DR\_SRC |
| Excel Source | EX\_SRC |
| Flat File Source | FF\_SRC |
| OLE DB Source | OLE\_SRC |
| Raw File Source | RF\_SRC |
| XML Source | XML\_SRC |
| Aggregate | AGG |
| Audit | AUD |
| Character Map | CHM |
| Conditional Split | CSPL |
| Copy Column | CPYC |
| Data Conversion | DCNV |
| Data Mining Query | DMQ |
| Derived Column | DER |
| Export Column | EXPC |
| Fuzzy Grouping | FZG |
| Fuzzy Lookup | FZL |
| Import Column | IMPC |
| Lookup | LKP |
| Merge | MRG |
| Merge Join | MRGJ |
| Multicast | MLT |
| OLE DB Command | CMD |
| Percentage Sampling | PSMP |
| Pivot | PVT |
| Row Count | CNT |
| Row Sampling | RSMP |
| Script Component | SCR |
| Slowly Changing Dimension | SCD |
| Sort | SRT |
| Term Extraction | TEX |
| Term Lookup | TEL |
| Union All | ALL |
| Unpivot | UPVT |
| Data Mining Model Training | DMMT\_DST |
| DataReader Destination | DR\_DST |
| Dimension Processing | DP\_DST |
| Excel Destination | EX\_DST |
| Flat File Destination | FF\_DST |
| OLE DB Destination | OLE\_DST |
| Partition Processing | PP\_DST |
| Raw File Destination | RF\_DST |
| Recordset Destination | RS\_DST |
| SQL Server Destination | SS\_DST |
| SQL Server Mobile Destination | SSM\_DST |

The following acronyms can be used as prefix to identify the connection managers being used:

|  |  |
| --- | --- |
| **Component** | **Prefix** |
| Excel Connection | EX\_SRC OR EX\_DST |
| Flat File Connection | FF\_DST or FF\_SRC |
| OLE DB Connection | OLE\_<databasename> |
| Raw File Connection | RF\_DST or RF\_SRC |
| SQL Server Connection | SS\_ |
| SQL Server Mobile Connection | SSM\_ |
| FTP Connection Manager | FTP\_<InterfaceName> |

# 7.0 Config Files Parameters:

|  |  |  |
| --- | --- | --- |
| **Type** | **Parameter** | **Properties** |
| Flat File Connections | Connection  String |  |
| OLEDB Connections | ServerName |  |
| UserName |  |
| Password |  |
| FTP Connection | Connection  String |  |
| ServerName |  |
| ServerPassword |  |
| Variables | Value |  |
| Execute Process task |  | Delay Validation = True |
| Excel Connection | ExcelFilePath |  |

# 8.0 Change Log

Please note that this table needs to be maintained even if a Configuration Management tool is used.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Version Number** | **Changes made** | | | |
| V1.0 | *<First version>* | | | |
| V1.1 | *<If the change details are not explicitly documented in the table below, reference should be provided here>* | | | |
| Page no | Changed by | Effective date | Changes effected |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| V1.2 | *<If the change details are not explicitly documented in the table below, reference should be provided here>* | | | |
| Page no | Changed by | Effective date | Changes effected |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |