|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| BODI/DS Checklist | | | | | | | | | | | | KILN | |
| Reference: [Kiln BI Standards Document](http://bsd/business_intelligence/Shared%20Documents/Standards/Kiln%20BI%20Standards.doc) | | | | | | | | | | | | | |
| Document ID: | |  | | Project Name: | | | | | | | |  | |
| Self Review Date: | |  | | Self Review Reference Documents: | | | | | | | |  | |
| Peer Review Date: | |  | | Peer Review Reference Documents: | | | | | | | |  | |
| SQA Review Date: | |  | | SQA Review Reference Documents: | | | | | | | |  | |
| Review Readiness Checklist  Note: If any of the checks in this section are ‘No’ then further review has to be stopped and the review stands rejected. | | | | | | | | | (Yes/NO) | | | Remarks | |
|  | | |  | |
| **Revision History:** | | | | | | | | | | | | | |
| **Version** | **Date** | | **Updated by** | | **Reviewed by** | | | **Updated Sections** | | | | | |
| V0.1 | 20-Mar-2012 | | Venkata Sudheer Induru | |  | | | Initial Version | | | | | |
| V0.2 | 27-April-2012 | | Yousuf K | | Dushyant Salathia | | | Checklist | | | | | |
| V0.3 | 30-April-2012 | | Dushyant Salathia | |  | | | Peer Review, SQA Review and Checklist | | | | | |
| V0.4 | 09-May-2012 | | Dushyant Salathia | |  | | | Appendix | | | | | |
| V0.5 | 30-May-2012 | | Dushyant Salathia | |  | | | Checklist | | | | | |
| V0.6 | 13-Jul-2012 | | Yousuf K | |  | | | Added remarks section | | | | | |
| For Peer Review | | | | | | | | | | | | | Remarks |
| Is Self Review Checklist available with all the checks passed? | | | | | | |  | | |  | | |  |
| Is Peer Review done on the same version of code which is marked for Release? | | | | | | |  | | |  | | |  |
| For SQA Review | | | | | | | | | | | | | |
| Is Peer Review Checklist available with all the checks passed? | | | | | | |  | | |  | | |  |
| Is the Peer Review DMS available and is in ‘Accepted’ status? | | | | | | |  | | |  | | |  |
| Has Peer review completed for the code which is marked for SQA Review? | | | | | | |  | | |  | | |  |
| For Peer & SQA Review | | | | | | | | | | | | | |
| Are all the reference documents (including standards/guidelines if any) of the current work item available? | | | | | | |  | | |  | | |  |
| Are all the reference documents reviewed and approved (except for client supplied documents)? | | | | | | |  | | |  | | |  |
| Checklist | | | | | | Self Review  (Yes/No) | Peer Review  (Yes/No) | | | SQA Review  (Yes/No) | | |  |
| Always check-out an object before starting the development. This will help other developers to know that someone is working on it hence would avoid any conflicts. | | | | | |  |  | | |  | | |  |
| The name of the Project/Jobs/Workflows/Data Flows/ Transformations/Tables/File Names/Directory structure/Column Names should have proper naming conventions as mentioned in the BI Standards Documents or [Appendix](#_TABLE_Appendix_-) below. Talk to Dushyant in case of any concerns. | | | | | |  |  | | |  | | |  |
| Jobs should have standard load management format. | | | | | |  |  | | |  | | |  |
| If there is any kind of filtering or sorting to be done while selecting the records from the source table, incorporate this filter or sorting condition near by the source system to improve pushdown operation. | | | | | |  |  | | |  | | |  |
| The validation of Dataflows/Workflows/Jobs should not give any Error messages or Warnings.  PS: Existing warnings can be ignored depending upon the criticality of the warnings e.g. warning about type casting can be ignored while ‘not enough tables’ joined need to be looked into. | | | | | |  |  | | |  | | |  |
| The WFs, DFs and Transformations should have proper descriptions in the properties about the role of the WF /DF/Transformations. | | | | | |  |  | | |  | | |  |
| Add proper annotation on the code. Refer to Kiln BI standards for details. Version number in annotation should not be the Release number.  Refer to following format while adding annotation comments.  # REPT-452: Changed the mapping to use TMP\_REP\_LANDED\_FORECAST\_ACTUAL\_INW\_CLAIM table.  #nvl(TMP\_REP\_LANDED\_INW\_CLAIM.CLAIM\_REFERENCE\_KILN,'')  nvl(TMP\_REP\_LANDED\_INW\_CLAIM.CLAIM\_REFERENCE\_KILN,TMP\_REP\_LANDED\_FORECAST\_ACTUAL\_INW\_CLAIM.INW\_CLAIM\_REFERENCE) | | | | | |  |  | | |  | | |  |
| During modification, do not delete existing code, always comment it with proper justification or reason and add the code. | | | | | |  |  | | |  | | |  |
| Ensure that the run time of DF/WF/Job after the change is better than before. In Worst case, it should be same. If not, then tune the code. | | | | | |  |  | | |  | | |  |
| Does ETL functionality match with the requirement? | | | | | |  |  | | |  | | |  |
| Are target tables deleted before they are loaded? If not then add an explicit annotation near the target table. | | | | | |  |  | | |  | | |  |
| Are’ select’ operations have done correctly while restricting on deleted\_ind? | | | | | |  |  | | |  | | |  |
| Is the use of ‘CURRENT\_IND’ or ‘Effective Start/End Dates’ is in accordance to the existing/new requirements of the sub-system? | | | | | |  |  | | |  | | |  |
| Does the CDC handles ‘delete’ operation from source correctly? | | | | | |  |  | | |  | | |  |
| Use parallel flows whenever it is possible. | | | | | |  |  | | |  | | |  |
| Insert the LOAD\_ID, INSERTED\_DT and LAST\_UPDATED\_DT columns just before the Target table | | | | | |  |  | | |  | | |  |
| Use the Table Comparison transformation only on the final Target table (Not on the Temp tables). | | | | | |  |  | | |  | | |  |
| If the code fails, will it start again from a known position or will rows be transformed twice? | | | | | |  |  | | |  | | |  |
| To develop any new dataflow as similar to the existing dataflow, make sure all the naming conventions and functionality implemented as per the new requirement. | | | | | |  |  | | |  | | |  |
| Make sure that variables (Global and Local) are printed whenever they are initialized or value is re-assigned. This will help in debugging the job. | | | | | |  |  | | |  | | |  |
| Design the code to group all source column name changes as much as possible and avoid scattering them unless absolutely required. Provide an explanation if that’s the case. | | | | | |  |  | | |  | | |  |
| Design the code to group all Summary/Aggregations as much as possible and avoid scattering them unless absolutely required. Provide an explanation if that’s the case. | | | | | |  |  | | |  | | |  |
| Design the code to group all lookup conditions as much as possible and avoid scattering them unless absolutely required. Provide an explanation if that’s the case. | | | | | |  |  | | |  | | |  |
| Merge Transform columns in the tables to be merged and to be in same order. Data types also to be matched. | | | | | |  |  | | |  | | |  |
| Make sure that the code is written optimally and no unnecessary code is written.   * Remove unused variables and columns in the code. * Avoid loading of unnecessary columns data into the target tables and check the dependency. * Create indexes on source table columns. | | | | | |  |  | | |  | | |  |
| Key columns should be marked as Key’s | | | | | |  |  | | |  | | |  |
| Avoid manipulating data using scripts. Limit usage to set variables, update a few records or only when there is a valid reason for not using dataflow. Document such reasons explicitly in the script for later reference. | | | | | |  |  | | |  | | |  |
| Do not drop/re-create tables/ indexes inside a Job. | | | | | |  |  | | |  | | |  |
| Code should not have any Template tables. All tables should be physically present in the ‘correct’ database with naming convention adhering to existing standards. Refer to BI Standards document and/or existing table naming convention. | | | | | |  |  | | |  | | |  |
| DDL Scripts - Ensure that database object names are free of any leading or trailing spaces and script comment black should be as per BI standards | | | | | |  |  | | |  | | |  |
| ‘Monitor sample rate’ of the job object should be set to 64000 if the job object was created or modified. | | | | | |  |  | | |  | | |  |
| ‘Print all trace messages’ should be unchecked if the job object was created or modified. | | | | | |  |  | | |  | | |  |
|  | | | | | |  |  | | |  | | |  |
| 1. **Configuration Management** | | | | | | | | | | | | |  |
| Has this code base lined from Production code? If it’s a merge with some other project then has TL & SQA been notified? | | | | | |  |  | | | |  | |  |
| If multiple developers have modified or plan to modify this code as part of the same release, then has the necessary steps been taken to ensure that one single version of the code is being released and shared with SQA? | | | | | |  |  | | | |  | |  |
| Is the code checked into the BODS project? Is SVN updated with the latest version of the code? | | | | | |  |  | | | |  | |  |
| While check-in the codes in BODS Central repository add appropriate Check-in comments in the following format.  <Project Name> - <JIRA Number> - <Release Number> - < Brief Change Comment >  E.g.  FC4 – FORE-575 - 4.2.1 - This dataflow has been modified to pull the YOA column from the source into the target table. Also the target table has been modified to include this new column.  Note: The comment should not be short & incomplete statement. It should be a clear and crisp sentence explaining why, how and where exactly the change has been made. | | | | | |  |  | | | |  | |  |
| ATL file name in SVN should conform to following standard –  If there is only one object(DF, WF, Function etc) modified as part of a Jira\Job, then export ATL file name should be  <Jira No >\_<Object Name>.atl  However if there are more number of Objects modified as part of a Jira for the same job, then instead of creating ATL file for each object, group them all at the Job level and the name of the file should be  <Jira No >\_<Job Name>.atl  Note: JIRA Number itself uses hyphen (-) and not underscore(\_) so here is an example of good and bad file name  Good: FORA-860\_WF\_FC\_Recalc\_FC\_Attr\_Inw.atl  Bad: FORA\_860\_WF\_FC\_Recalc\_FC\_Attr\_Inw.atl | | | | | |  |  | | | |  | |  |
| 1. **Performance Optimization** | | | | | | | | | | | | |  |
| Has the code been checked for Performance tuning? | | | | | |  |  | | | |  | |  |
| Multiple queries in the “Optimized SQL” are discouraged to improve push down, has steps been taken to address this? | | | | | |  |  | | | |  | |  |
| Do data flows ‘push down’ as much as possible? (Are sourced from multiple databases?) | | | | | |  |  | | | |  | |  |
| When loading large amount of data, is the bulk loader used appropriately? | | | | | |  |  | | | |  | |  |
| Is the lookup caching strategy correct? | | | | | |  |  | | | |  | |  |
| Is Table comparison transformation using optimal Comparison strategy? Eg. Cached Comparison Method | | | | | |  |  | | | |  | |  |
| Are there any Cartesian products in the dataflow? If yes, then are they properly documented explaining the reason? | | | | | |  |  | | | |  | |  |
| Has ‘distinct’ been used unnecessarily? Could distinct be avoided and replaced with a filter condition – where clause or join condition? | | | | | |  |  | | | |  | |  |
| While using Map Operation to implement ‘Update’ strategy, ‘Key Columns’ should be used appropriately. | | | | | |  |  | | | |  | |  |
| Have steps been taken to consolidate all ‘similar’ logic at a single place where applicable to improve reusability and avoid redundant processing? | | | | | |  |  | | | |  | |  |
| When loading huge amount of test data, only then increase the value of “Rows per commit” to 10K or higher. While doing so, give valid explanation. | | | | | |  |  | | | |  | |  |
| Do not use scripts instead use DF’s wherever possible. | | | | | |  |  | | | |  | |  |

# Appendix

Refer to Kiln BI Standards document for details however table below is here for quick and handy reference.

[Business Services Division > Business Services Division - Business Intelligence > Shared Documents > Standards](http://bsd/business_intelligence/Shared%20Documents/Standards/Kiln%20BI%20Standards.doc)

**Kiln BI Standards.doc**

|  |  |
| --- | --- |
| **BODI/DS Object Type** | **Naming Convention** |
| Project | <Project or Datamart Name>\_PROJECT |
| Jobs | JOB\_<Project>\_<LogicalOrder>\_<Description>\* |
| Workflows | WF\_<Project>\_<Description> |
| Dataflows | 1. DF\_<Project>\_<Description> |
| Query Transforms | QRY\_<Narrative> |
| Table Comparison | TC\_<Compare Table Name> |
| History Preserving | HP\_<History Table Name> |
| Key Generation | KG\_<Key Name> |
| Map Operations | MAP\_<Narrative> |
| Map CDC Operation | MAP\_CDC\_<Narrative> |
| Case | CASE\_<Narrative> |
| Merge | MERGE\_<Narrative> |
| Pivot | PIVOT\_<Narrative> |
| Reverse Pivot | RP\_<Narrative> |
| Row Generation | RG\_<Narrative> |
| SQL | SQL\_<Narrative> |
| Validation | VAL\_<Narrative> |
| File Names | <Project>\_<Description> |
| Table Names | |  |  | | --- | --- | | **Table Type** | **Name Suffix** | | Landing | TMP\_<data source>\_LANDED\_<name> | | Working | WRK\_<data source>\_<name> | | Control | CTRL\_<description of purpose> | | Reporting Dimension | DIM\_<description of dimension> | | Reporting Ready To Load | RTL\_DIM | FACT\_<description> | | Reporting Fact | FACT\_<description of fact> | | Landed for flat file sources | TMP\_<data source>\_VALIDATED\_<description> | |
| 1. Directory Structure | {SVN}\BI\<Application>\Patch\<Release\_Number> \*\* |
| Column Name | Refer to Kiln BI Standards |
| 1. Data\_Transfer | DT\_<Narrative> |
| 1. Date\_Generation | DG\_<Narrative> |
| Effective\_Date | EFF\_DT\_<Narrative> |
| 1. Hierarchy\_Flattening | HF\_<Narrative> |
| 1. XML\_Pipeline | XMLP\_<Narrative> |
| Conditional | CND\_<Narrative> |
| While loop | WHILE\_<Narrative> |
| Try - Catch | Used to enclose the ETL code for a job as below: Try transform with the job ending with a Catch\_Error transform |
| Script Operations | SCR\_<Short Description of Script> |
| Custom Functions | fn\_<Narrative> |

|  |  |
| --- | --- |
| **Parameters and Variables** | **Naming Standards** |
| Global Variable | $G\_<Narrative> |
| Local Variable | $LV\_<Narrative> |
| Workflow Variable | $WF\_<Narrative> |
| Parameter | $p\_<Narrative>fora function or  $DF\_Param\_<Narrative>for a dataflow |

\* Job Logical Order needs to be assigned by Team Leads after discussion with BI Architects.

\*\*Release Number needs to be assigned by Team Leads after discussion with Configuration Manager.

The above appendix table is based on the following version of the Kiln BI Standard from Kiln SharePoint.

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Last Modified On** | **Last Modified By** | **Modification Comments** |
| 0.45 | 21-Feb-2012 16:18 | Ravinder Dosanjh | 21/02/2012 RD: Added section 2.3.3 dataflow layout |