# Project Standards:

### Each table (dimensions & facts) need to record:

* SOR\_ID
* *DI\_Job\_ID*
* *DI\_Create\_Date*
* *DI\_Modified\_Date*



(from BI Guidebook - Figure 12-16: Data Integration Table - Job Audit Columns)

* *SOR\_ID* – This is the SOR identifier that will tie this row to a particular SOR. Use this when the table the row is sourced from multiple systems of records and enables the row to be tied to the specific SOR.
* *DI\_Job\_ID* – the data integration job identifier is be the job id that the data integration tool generated. This identifier is a foreign key to the data integration tool’s processing metadata. If that metadata is available, then this link provides a powerful mechanism to analyze data integration processing and performance down to the level of a table’s row. *Note: The datatype for this column is specific to the data integration product used.*
* *DI\_Create\_Date* – This is the date and time that this row was originally created in this table. Often a database trigger is used to insert the current time, but the data integration job could also insert the current time directly.
* *DI\_Modified\_Date* - This is the most recent date and time that this row has been modified in this table. Often a database trigger is used to insert the current time, but the data integration job could also insert the current time directly. It is often a standard practice to populate this column with an initial value far in the future such as “9999-12-31” rather than leaving it a NULL to avoid queries with NULLs when analyzing this column.

For SQL Server the SQL syntax for these columns is:

* DI\_Job\_ID **nvarchar**(**20**) NULL
* DI\_Create\_Date **datetime** NOT NULL DEFAULT (getdate())
* DI\_Modified\_Date **datetime** NOT NULL DEFAULT (getdate())

### Every Talend job needs to tract job stats & errors

The steps to completing this:

1. Insert Joblet (supplied) that tracks job stats & errors (below)
2. Create columns as specified above
3. Insert values into those columns (below)

#### Joblets

* + Every job created must have a Joblet that tracks stats via tStatCatcher, tLogCatcher & tAssetCatcher.
  + For this project the Joblets provided:
    - CNTL\_Job\_Tracking\_Stats\_ROOT – placed in top level job also referred to as the Root job
    - CNTL\_Job\_Tracking\_Stats\_CHILD – Placed in every non-root job

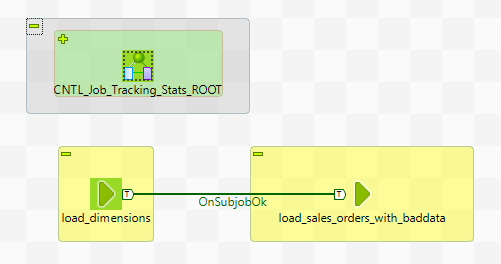


Figure : Joblet to track stats & errors in Root job

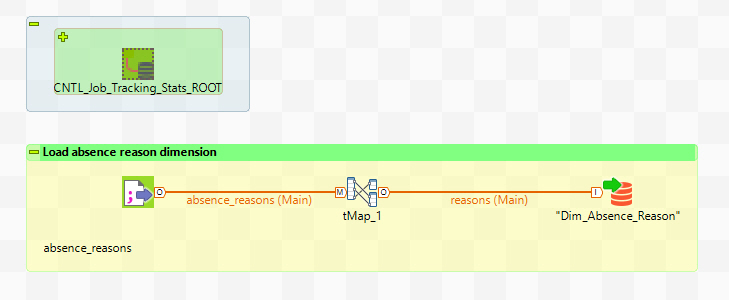


Figure : Joblet to track stats & errors in Child job

#### Insert values into standard columns

* + DI\_Job\_ID - context.getProperty("vRootPID")
  + DI\_Create\_Date - TalendDate.getCurrentDate()
  + DI\_Modified\_Date - TalendDate.getCurrentDate()



Figure : tMap inserting values into target tables

### Converting a date to a date surrogate key:

In the target table use the following function:

Integer.parseInt(routines.TalendDate.formatDate("yyyyMMdd",row1.TheSourceDate))

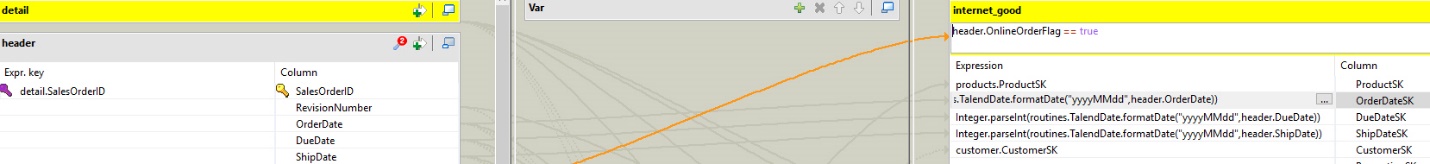


Figure : Converting a date to a date surrogate key

The example:

* Source column: header.OrderDate (date or datetime)
* Target column: internet\_good.OrderDate\_SK (integer)
* Expression:
  + Integer.parseInt(routines.TalendDate.formatDate("yyyyMMdd",header.OrderDate))