

An ETL Framework for Operational Metadata Logging

Quite often there is requirement to track runtime information about your ETL jobs such as record count, error count, job run time etc. In general it is a non-functional requirement, required by the IT team to have such information for reconciliation purposes, calculate performance statistics etc. It is important to have a framework, which can capture all the operational metadata you need without adding too much time to your development cycle.

This framework is to capture Operational Metadata by leveraging the capabilities provided by Informatica PowerCenter.

Framework Components

Our Framework for Operational Metadata logging will include three components.

1. **A Relational Table** :- To store the metadata.
2. **Pre/Post Session Command Task** :- Command task to collect the metadata.
3. **Reusable Session** :- Session to log the metadata details into the relational table.

I. Relational Table

A relation table will be used to store the operational metadata with the structure as below. Data in this table will be retained for historical analysis.

- ETL_JOB_NAME : ETL job name or Session name.
- ETL_RUN_DATE : ETL job execution date.
- SRC_TABLE : Source table used in the ETL job.
- TGT_TABLE : Target table used in the ETL job.
- ETL_START_TIME : ETL job execution start timestamp.
- ETL_END_TIME : ETL job execution end timestamp.
- SRC_RECORD_COUNT : Number of records read from source.
- INS_RECORD_COUNT : Number of records inserted into target.
- UPD_RECORD_COUNT : Number of records updated in target.
- ERR_RECORD_COUNT : Number of records error out in target.
- ETL_STATUS : ETL Job status, SUCCESS or FAILURE.
- ETL_CREATE_TIME : Record create timestamp.
- ETL_UPDATE_TIME : Record update timestamp.

II. Pre/Post Session Command Task

Pre/Post session command task will be used to generate a comma-delimited file with session run details. This file will be stored into `$PMSourceFileDir\` directory with a name `$PMWorkflowName_stat.txt`.

Note :

- **\$PMSourceFileDir**, **\$PMWorkflowName** are the session parameter, which gives the source file directory and name of workflow.
- File name generated will always be **<WorkflowName>_stat.txt**

The comma-delimited file will have the structure as below.

- ETL Start time
- ETL End time
- ETL Job name
- Source table name
- Target table name
- Source record count
- Records inserted count
- Records updated count
- Error record count
- ETL Job status

We will be using the built-in session parameters to collect session run details.

- **\$PMSessionName** : Name of the Informatica session.
- **\$PMSourceName@TableName** : Name of the source table name.
- **\$PMTargetName@TableName** : Name of the source table name.
- **\$PMSourceQualifierName@numAffectedRows** : Number of records returned from source.
- **\$PMTargetName@numAffectedRows** : Number of record inserted/updated into the target table.
- **\$PMTargetName@numRejectedRows** : Number of records error out in target.

Note : *SourceName, TargetName, SourceQualifierName* will be replaced by corresponding transformation instance name used in the mapping.

Pre Session Command Task

Pre session command task will be used to create the file with the session start time stamp.

```
echo %DATE:~10,4%-~4,2%-~7,2%
%TIME:~0,2%:~3,2%:~6,2%,
> $PMSourceFileDir\PMWorkflowName_stat.txt
```

Post Session Success Command Task

Post session success command task will be used to append the file, which is created in the pre session command with session run details. This will capture the SUCCESS status along with other session run details.

```
echo %DATE:~10,4%-~4,2%-~7,2%
%TIME:~0,2%:~3,2%:~6,2%,PMSessionName,
PMSTG_CUSTOMER_MASTER@TableName,
```

```

$PMINS_CUSTOMER_MASTER@TableName,
$PMSQ_STG_CUSTOMER_MASTER@numAffectedRows,
$PMINS_CUSTOMER_MASTER@numAffectedRows,
$PMUPD_CUSTOMER_MASTER@numAffectedRows,
$PMINS_CUSTOMER_MASTER@numRejectedRows,
SUCCESS,
>> $PMSourceFileDir\$PMWorkflowName_stat.txt

```

Post Session Failure Command Task

Post session failure command task will be used to append the file, which is created in the pre session command with session run details. This will capture the FAILURE status along with other session run details.

```

echo %DATE:~10,4%- %DATE:~4,2%- %DATE:~7,2%
%TIME:~0,2%: %TIME:~3,2%: %TIME:~6,2%, $PMSessionName,
$PMSTG_CUSTOMER_MASTER@TableName,
$PMINS_CUSTOMER_MASTER@TableName,
$PMSQ_STG_CUSTOMER_MASTER@numAffectedRows,
$PMINS_CUSTOMER_MASTER@numAffectedRows,
$PMUPD_CUSTOMER_MASTER@numAffectedRows,
$PMINS_CUSTOMER_MASTER@numRejectedRows,
FAILURE,
>> $PMSourceFileDir\$PMWorkflowName_stat.txt

```

Note :

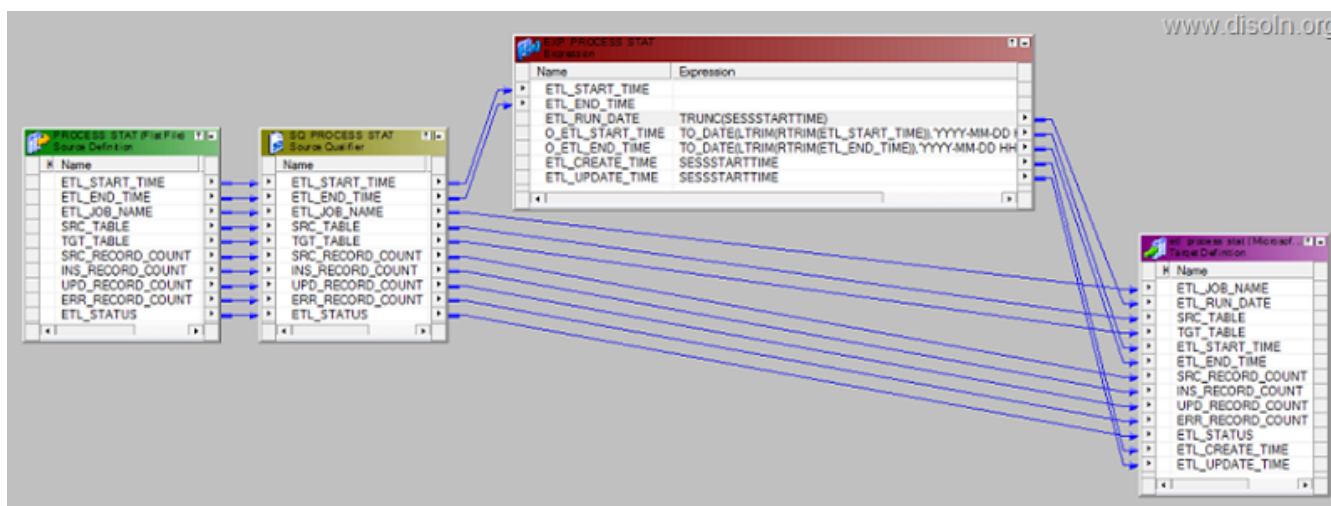
- Pre/Post session commands need to be changed based on Informatica server operating system.
- Highlighted part of the script need to be change based on the source, target table instance name used in the mapping.

III. Reusable session

Reusable Mapping

A reusable mapping will be created to read data from the comma delimited file generated by the pre/post session command task. Below is the mapping created with an expression transformation to populate additional columns required in the target table '**ETL_PROCESS_STAT**'

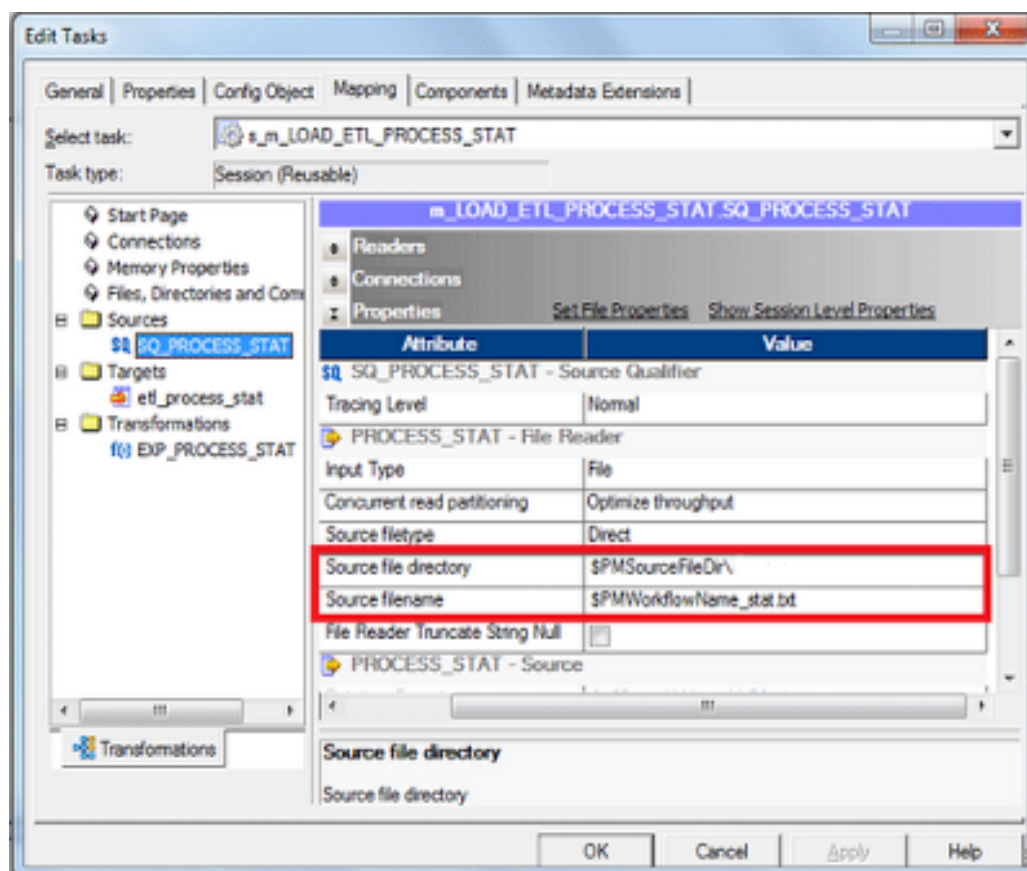
- ETL_RUN_DATE :- TRUNC(SESSSTARTTIME)
- ETL_CREATE_TIME :- SESSSTARTTIME
- ETL_UPDATE_TIME :- SESSSTARTTIME
- ETL_START_TIME :- TO_DATE(LTRIM(RTRIM(ETL_START_TIME)), 'YYYY-MM-DD HH24:MI:SS')
- ETL_END_TIME :- TO_DATE(LTRIM(RTRIM(ETL_END_TIME)), 'YYYY-MM-DD HH24:MI:SS')



Reusable Session

A reusable session will be created based on the mapping created in the last step. Session will be configured to read data from the file created by the pre/post session command as shown below.

Note : Make sure the Source File Directory and Source File name are given correctly based on the file generated by pre/post session command



Framework implementation in a workflow

Shown below is a workflow using Operational Metadata logging framework. Pre/Post session command to generate file with session run details will be in the first session, The reusable session will be connected after to read data from the file and load into ETL_PROCESS_STAT table.

