TATA CONSULTANCY SERVICES
Experience certainty.

Lessons Learnt

7/28/15

How it happened

Process Adopted

1Understanding the project requirements

. 2.Designing the jobs accordingly

3.Extraction of columns from source tables or files

4.Applying transformation logic on columns (if required)

5.Loading result to target table/file

Issues & Analysis

Issues Faced

- Some of the jobs took a lot of time to execute
- Jobs got aborted due to space issue in UNIX Server
- Inconsistency in job requirements

Causal

- Space taken up by large number of temporary files
- Large number of joins with tables led to consume more Time.
- Data extraction from multiple tables in a single job took long execution time.

Results

Justification

Joining multiple tables and using number of transformer stages led to performance issue.

Extracting data into datasets and breaking up single job reduced the execution time.

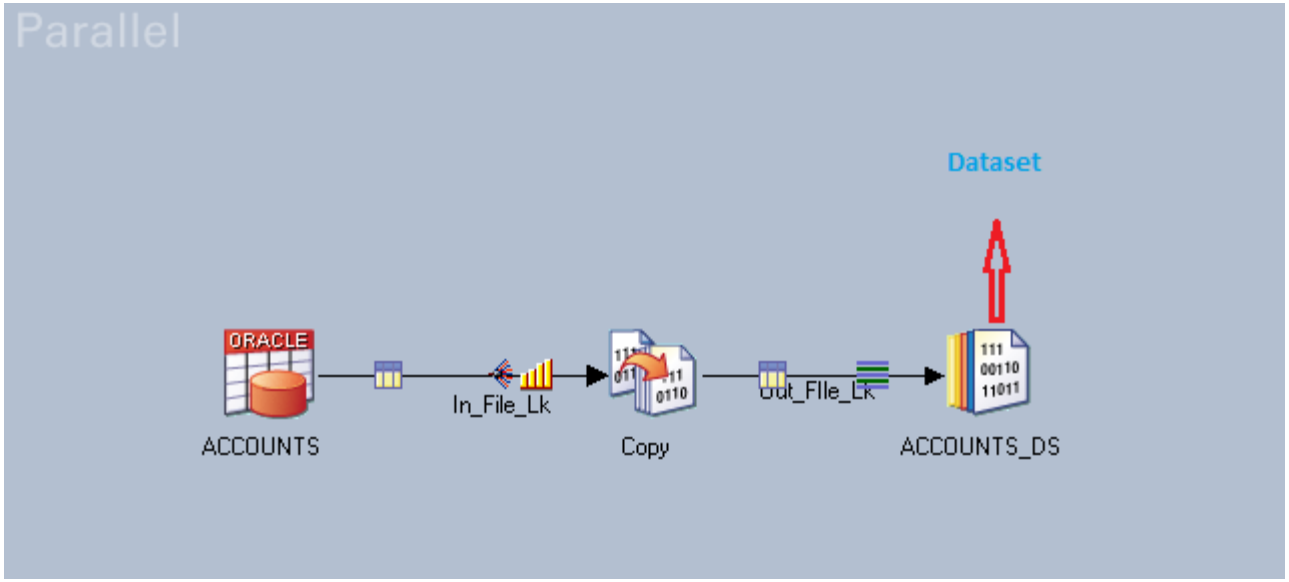
Benefits

- Execution time reduced
- Less re-work

Solutions & lessons

Lessons Learnt

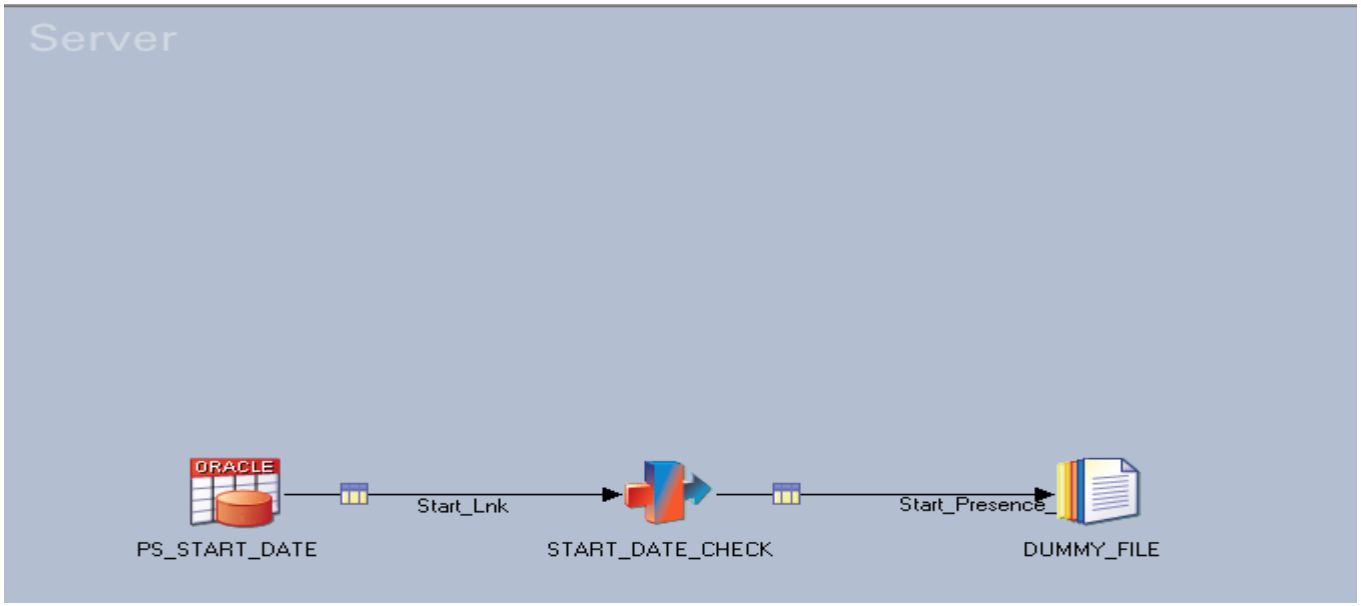
- Datasets were created extracting only the required columns and reused where ever required.
- It reduced extraction time.



Solutions & lessons

Lessons Learnt

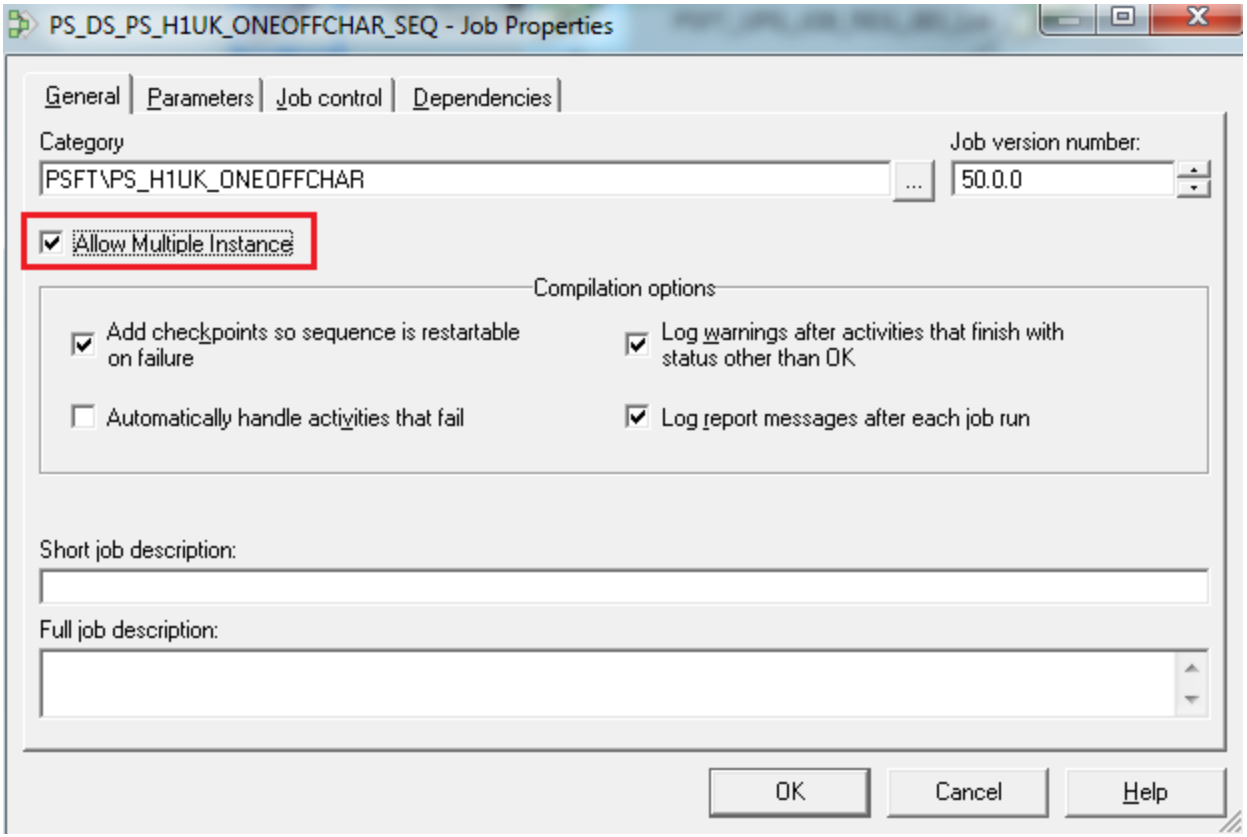
- Reuse of Common Job reduced total job development time.



Solutions & lessons

Lessons Learnt

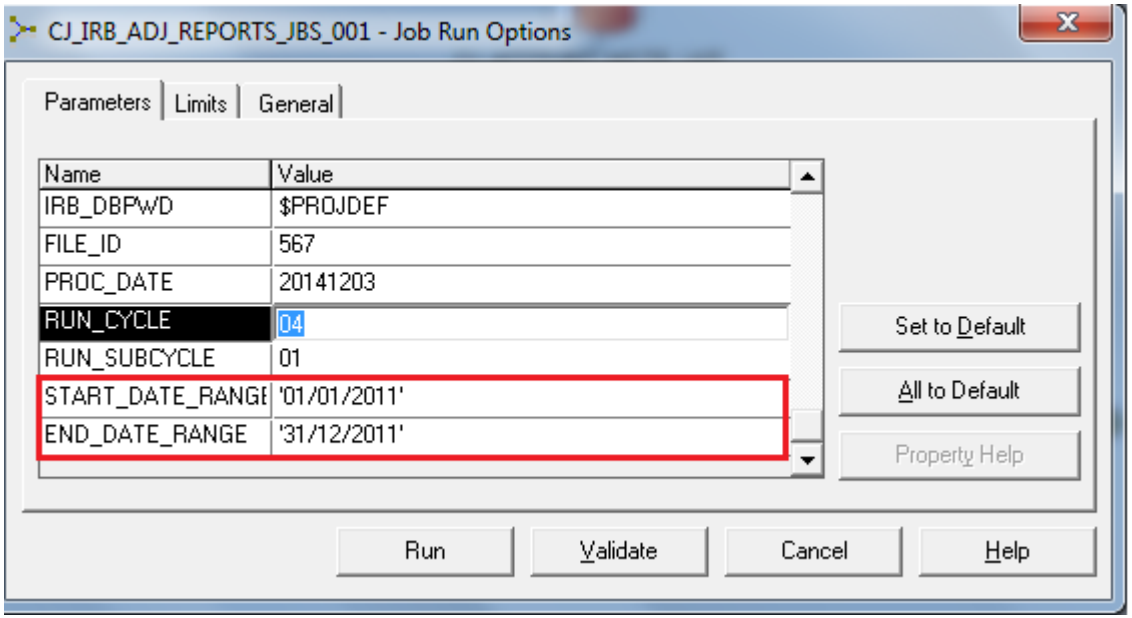
- Enabling ‘Allow Multiple Instances’ under Job Parameters tab helped to run the same job with different invocation id for different range of data.



Solutions & lessons

Lessons Learnt

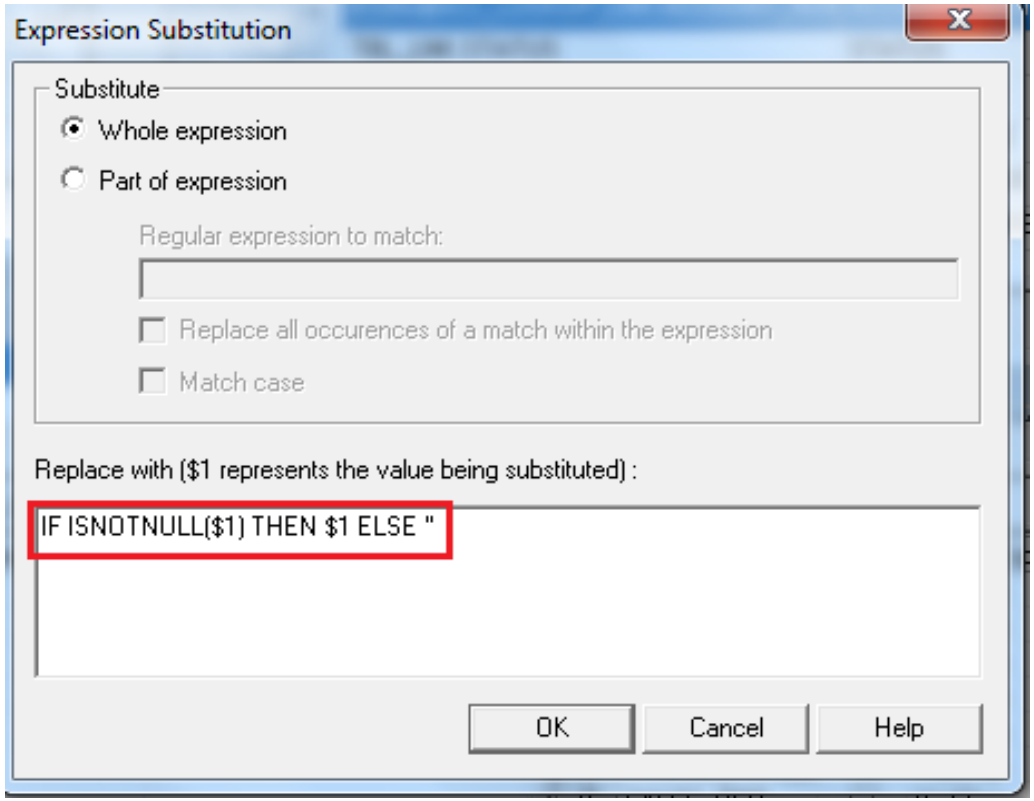
- Some tables had huge volume of data so if the job is run in a single step there is a chance for the job to get aborted due to server space issue.
- To avoid that jobs are run based on year wise or by partitioning the table into small batches.



Solutions & lessons

Lessons Learnt

- Derivation substitution technique helped to apply the same transformation logic to a set of columns.



Solutions & lessons

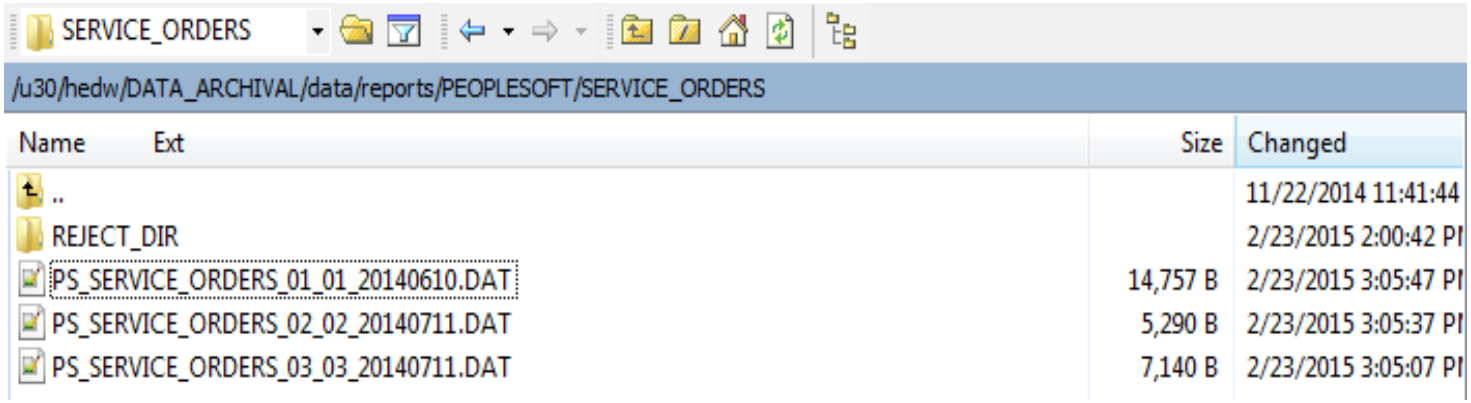
Lessons Learnt

- Adapted a standard naming convention for resultant filename

Syntax:
#pPROJECT_DIR##pDAT_DIR##pREPORT_DIR##pSRC_SYS#/#pSUB_AREA#/#pOUT_FILE_NAME##pRUN_CYCLE#_#pRUN_SUB_CYCLE#_#pPROC_DATE##pFILE_EXT#

Example:
/u30/hedw/DATA_ARCHIVAL/data/reports/PEOPLESOFT/SERVICE_ORDERS/

Output_Directory:



How this may be prevented elsewhere

Prevention

- Less time is taken to join tables in data extracting and keeping in a dataset when reusing it.
- Large transformations in transformer stage should be replaced by
stage variables.

Contact Info

Sunil S(684804)
Mailto: s.sunil2@tcs.com

TATA CONSULTANCY SERVICES
Experience certainty.

Thank You

IT Services
Business Solutions
Consulting