

Extract Data Items from ci360:Direct onprem logs

Raja Marla

Last updated: 01/20/2024

Table of Contents

Release History	3
Description	3
Assumptions.....	3
Customization Instructions	3
Program outputs	4
Last-Update-file (Flat file)	4
Outlib.runtask	4
Outlib.export.....	4
Outlib.node	4

Release History

01/20/2024 Improvements

- Capture calc items and their definitions
- Capture the data item selection criteria (in addition to their name)

Bug fixes

- Capture data item names longer than 50 characters. Newer limit is 100 characters

Description

This utility program reads the Engage:Direct onprem logs (but never the current days' log) and produces SAS datasets that help cross-reference each segment map/direct task executed with the data items in the Information Map.

The program is designed to be run once a week (or once a day or another frequency). It saves the last processed file date and continues processing from the next date. At each invocation the program processes a maximum of thirty-one files (configuration options). Processing too many files may result in longer execution times.

The output datasets (described below) are based on the current execution.

Note:

Every execution of each task/segment is preserved in the output dataset. It is normal for a single task/segment to be changed over time and re-executed. Thus, many historical versions may exist for each segment/task. This program does not attempt to remove the historical versions.

Assumptions

1. The onprem logs are named onprem_direct.log.yyyy-mm-dd.
2. All the onprem logs to be read are from the same directory/folder

Customization Instructions

Refer to utl_onpremdiext.sas and update the variables prefixed "c_". Additional description is included in the sas program.

The variable `c_extract_dataitem_values` controls whether to retrieve data items values from criteria nodes. If you are only interested in data item names rather than data item values, you may choose to set

`c_extract_dataitem_values=N`

Program outputs

Last-Update-file (Flat file)

The location/name of file is specified at the top of program as a configuration option. It contains a single record, containing a date specified in the format yyyy-mm-dd. If you wish to process historical files on the initial execution, you will need to create this file with a single record (If you wish to process files from 01-Feb-2022, then specify the date 2022-01-31.)

If the file does not exist, then one would be created after the process termination. In this case, only the one log file for the previous day would be processed.

Outlib.runtask

This SAS dataset is appended after each execution with the data from the logs processed during that run. It contains one row for each task/segment executed in the log files.

Name	Type	Length	Format
task_name	Character	100	
type	Character	15	
externalCode	Character	15	
modifiedByUserName	Character	50	
runtask_id	Character	40	
log_filename	Character	100	
task_runtime	Date	8	E8601DT.
countsOnly	Numeric	8	
businessContextU...	Character	36	
file_record_nbr	Numeric	8	

Outlib.export

This dataset relates each task/segment with the list of data items exported to one or more DM task outputs.

Name	Type	Length	Format
externalCode	Character	15	
runtask_id	Character	40	
dataitem_name	Character	100	
outputName	Character	60	
outputSubjectId	Character	50	
task_runtime	Date	8	E8601DT.

Outlib.node

This dataset relates each DM task/segment nodes with the corresponding data items from the information maps. At the present time, “node” refers to “criteria/segment or split” nodes.

Name	Type	Length	Format
externalCode	Character	15	
runtask_id	Character	40	
segmentMapN...	Character	50	
dataitem_name	Character	100	
node_name	Character	60	
outputSubjectId	Character	50	
task_runtime	Date	8	E8601DT.
dataitem_value	Character	200	
operator_name	Character	20	
node_type	Character	20	

1. Outlib.calc_items

This data captures the calc item name and the expression. Calc item names on the “node” table are represented by an id. The node.id is a foreign key to the calc_items.ci_id.

Name	Type	Length	Format
ci_id	Character	36	
ci_name	Character	100	
ci_type	Character	20	
ci_expression	Character	2048	
ci_related_items	Character	100	

Note – calculated items can contain other calculated items and data items. While we capture the calculated item expressions, we do not recursively search the expressions for other calculated items or data items.

Upgrade considerations

The table structure may change with each new version of this utility. One of the following approaches is recommended:

- Rename the current SAS datasets with the _old suffix, and let the first execution of this process recreate the tables. Afterwards, copy the data from _old tables to the new tables
- Review the new table structure from the above document, and update your old SAS datasets to add the new columns to match the new structure.