Source Data Extract

# Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Date | Author | Remark |
| 1.0 | 31/08/2011 | Dirk Vermeylen | Initial Release |
| 1.1 | 02/09/2011 | Dirk Vermeylen | email required for ESL Availability report |
| 1.2 | 08/09/2011 | Dirk Vermeylen | Add Applications and Application Relations |
| 1.3 | 09/09/2011 | Dirk Vermeylen | Add OVSD Server Relations |
| 1.4 | 22/02/2012 | Dirk Vermeylen | Update after review |
| 1.5 | 24/04/2012 | Dirk Vermeylen | Add Security Category and Security Class load |
| 1.6 | 14 may 2012 | Philip Van Daele |  |

# Introduction

This document describes the procedure to extract source data and convert the data to the template format for the different components and the different source systems.

# System Preparation

The extract process needs following software applications:

* Perl – ActiveState Perl v5.10.0 build 1005 is used
* MySQL – Server version 5.1.36-community is used
* Microsoft Office Access 2007

Following software applications are helpful for a user-friendly mysql

* Apache Web Server
* phpmyadmin

## MySQL

2 databases are required in MySQL:

* alu\_cmdb: to store source system data. Do not create any tables.
* cim: to store extracted/converted data.

Review MySQL access info in Perl script dbParams\_aluCMDB.pm. The database names can be changed if done consistently both in the module and on the MySQL database.

Create an ODBC driver to MySQL alu\_cmdb database. This driver will be used to convert data from Microsoft Access to MySQL. Find MySQL ODBC driver, configure using Administrative Tools, Data Sources, go to System DNS, add MySQL ODBC 5.1 driver. Data source name is free to choose. Database is alu\_cmdb.

The CIM database must be created from cim.sql.

The CIM database has some tables that need to be populated with master data. These tables are:

* person
* translate
* os\_translation

The CIM database and the master tables are available on SVN, artifacts\trunk\analysisdeliverables\sourceprep\Database

Before a data extract, it is best to restore the CIM database to the current version from SVN and load the master tables into CIM.

Review that strict mode is disabled for MySQL. Strict mode does not allow ‘’ as NULL values for integers.

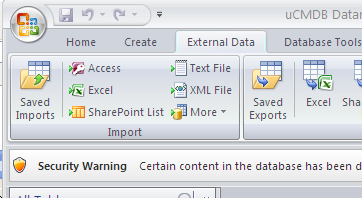
This can be done by changing the my.ini file from  
sql-mode="STRICT\_TRANS\_TABLES,NO\_AUTO\_CREATE\_USER,NO\_ENGINE\_SUBSTITUTION"  
to  
sql-mode="NO\_AUTO\_CREATE\_USER,NO\_ENGINE\_SUBSTITUTION"  
and restart MySQL.

## Microsoft Access

Access is used as a bridge to import \*.csv and .xls(x) files into MySQL. Access allows an import with little effort and without quality loss.

Microsoft Access allows to convert the (Assetcenter and OVSD) excel files and ESL csv files.

This is done by selecting “External Data” – Import “Excel” or “Text File” (for csv):



### Import Excel Files

Excell Macro preparation :

* How to add the developer ribbon - <http://www.itechtalk.com/thread10128.html>
* Create a macro - <http://www.itechtalk.com/thread10139.html>
* Add macro to quick access toolbar - <http://www.dummies.com/how-to/content/how-to-add-an-excel-2010-macro-to-the-quick-access.html>

Prepare Excel Files by converting long integers to strings. This avoids that long integers are converted into scientific formats.

Select the area with the long integers and run the macro:

Dim cv

For Each xcell In Selection

If Not xcell.Value = "" Then

If (IsNumeric(xcell.Value)) Then

cv = xcell.Value

cv = "'" & cv

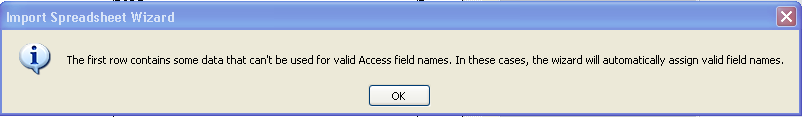
xcell.Value = cv

End If

End If

Next xcell

For Excel files, browse to the file, select the appropriate worksheet if required, always select “First Row contains Column Headings” (scripts do use Column Headings as labels). Ignore this warning message if it pops up:



Select “Let Access add primary key”.

Select the table name to import as specified in appropriate section.

Review if there are any Import errors. If there are import errors, then change the field setting to ‘text’ for the appropriate field.

### Import csv Files

Select “Import Text File” and browse to the file.

Select “Import the source data into a new table in the current database”. In the “Import Text Wizard”, select “Advanced...”. Change the Decimal Symbol to a . (dot) if required. Change the Text Qualifier to “ (double quote). This allows multi-line fields to be imported into a single cell if required. Accept Advanced settings.

Choose Comma as a delimiter and “First Row Contains Field Names”. Review to make sure that Text Qualifier is the double quote. Let Access add primary key and import into the appropriate Table name.

## Importing Files

It has been seen that the import of a file from large Excel files causes Microsoft Access to hang. Terminating Microsoft Access and restarting seems to recover the import completely.

On some occasions the error message “File sharing lock count exceeded...” occured. A solution for the error message is documented on <http://support.microsoft.com/kb/815281>.

The temporary solution requires these steps:

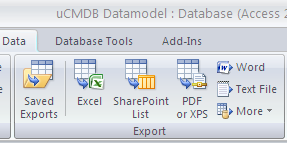
In the database,

* Press Alt + F11 to launch Visual Basic Editor
* In Visual Basic Editor, click the Immediate Window menu
* In Immediate Window, enter DAO.DBEngine.SetOption dbmaxlocksperfile,15000
* Press Enter to execute the statement.

Another work-around that seems to work is to start the import from a new database.

## Export to MySQL

Select the table to be exported, select “Export – More”:



and find the MySQL ODBC driver. Review the Table name and make sure that the table does not exist in MySQL. Export will go smoothly.

## Perl

Store the scripts from directory ALU in a script directory. Store the modules in a Perl module directory. A full data extract is done by running the all\_all.bat file. This will call all required scripts in a valid sequence.

Review module Log.pm to make sure that the log directory exists, or update to an existing log directory.

Review module dbParams\_aluCMDB.pm for MySQL connectivity information (username, password, port, ...)

Review the extract Perl scripts for the Query – Where statements, to guarantee that the required set of data is extracted. Maintain documentation in the script.

Documentation is available in the scripts in pod format. This can be extracted and formatted using the htmldoc.pl and tocgen.pl scripts from the TablePreparation subdirectory.

A requirement is to have all output files in UTF-8. Set environment variable

PERL\_UNICODE=2

to ensure STDOUT will be UTF-8. Do not set STDIN to UTF-8, since this will result in all kind of error messages.

## Directory Structure

The Data Extract scripts require specific directories.

Log files are in d:\temp\log. This setting can be changed in Log.pm

Extract files \*.csv are in directory d:\temp\alucmdb. This is a hardcoded setting in every script that creates \*.csv files.

The file ‘runFileSplitter.bat’ requires additional directories (to be modified in this bat file. These directories are:

d:\temp\FileRewriter\inputDir

d:\temp\FileRewriter\outputDir

# Sources of the data

## CMO Assetcenter and OVSD

Extracts are available on Sharepoint:

<https://external1.collaboration.hp.com/external/ALUandHP/GovernanceBusinessOps/Reporting/Asis%20Reports%20Schedule%2013/Forms/AllItems.aspx>

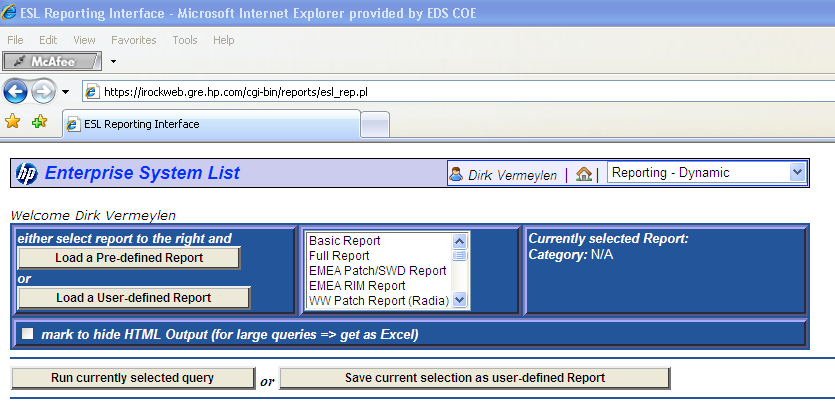
Assetcenter files are in the directory [ASSET CENTER CMDB (Configuration Management Database) Reports - Infrastructure - EMEA CALA[Use SHIFT+ENTER to open the menu (new window).](https://external1.collaboration.hp.com/external/ALUandHP/GovernanceBusinessOps/Reporting/Asis%20Reports%20Schedule%2013/Forms/AllItems.aspx?RootFolder=/external/ALUandHP/GovernanceBusinessOps/Reporting/Asis%20Reports%20Schedule%2013/ASSET%20CENTER%20CMDB%20(Configuration%20Management%20Database)%20Reports%20-%20Infrastructure%20-%20EMEA%20CALA&FolderCTID=&View=%7b217FC9AB-A61B-4374-8196-36238EACCCD0%7d)](https://external1.collaboration.hp.com/external/ALUandHP/GovernanceBusinessOps/Reporting/Asis%20Reports%20Schedule%2013/Forms/AllItems.aspx?RootFolder=%2fexternal%2fALUandHP%2fGovernanceBusinessOps%2fReporting%2fAsis%20Reports%20Schedule%2013%2fASSET%20CENTER%20CMDB%20%28Configuration%20Management%20Database%29%20Reports%20%2d%20Infrastructure%20%2d%20EMEA%20CALA&FolderCTID=&View=%7b217FC9AB%2dA61B%2d4374%2d8196%2d36238EACCCD0%7d).

OVSD files are in the directory [OVSD CMDB (Configuration Management Database) Reports - Infrastructure - AMS APJ[Use SHIFT+ENTER to open the menu (new window).](https://external1.collaboration.hp.com/external/ALUandHP/GovernanceBusinessOps/Reporting/Asis%20Reports%20Schedule%2013/Forms/AllItems.aspx?RootFolder=/external/ALUandHP/GovernanceBusinessOps/Reporting/Asis%20Reports%20Schedule%2013/OVSD%20CMDB%20(Configuration%20Management%20Database)%20Reports%20-%20Infrastructure%20-%20AMS%20APJ&FolderCTID=&View=%7b217FC9AB-A61B-4374-8196-36238EACCCD0%7d)](https://external1.collaboration.hp.com/external/ALUandHP/GovernanceBusinessOps/Reporting/Asis%20Reports%20Schedule%2013/Forms/AllItems.aspx?RootFolder=%2fexternal%2fALUandHP%2fGovernanceBusinessOps%2fReporting%2fAsis%20Reports%20Schedule%2013%2fOVSD%20CMDB%20%28Configuration%20Management%20Database%29%20Reports%20%2d%20Infrastructure%20%2d%20AMS%20APJ&FolderCTID=&View=%7b217FC9AB%2dA61B%2d4374%2d8196%2d36238EACCCD0%7d).

Assetcenter extracts are monthly, OVSD extracts are weekly. Always use the most recent data.

## ESL

Extracts are available from the ESL Dynamic Reporting Interface:



and select “Load a User-defined Report”.

Select user “Datamigration ALU” and “Show User-Reports”.

Available Reports have a name and a date/time label (YYMMDDhhmm). Select the most recent report. Select Report Details and “Load Report”. Do NOT run the report until the Parameter screen is displayed.

Select ‘mark to hide HTML Output’ and “Run currently selected query”. Download the generated csv file.

# Hardware Component

## Assetcenter

### Hardware CIs

Source file in Sharepoint: HP\_ALU\_SLM\_*yyyy\_mm\_dd*\_SRV\_V2.xlsx

Columns for long integer to string conversion:

* Z: Billing Request Number
* BF: \*° Laste Change #

Target file in MySQL: a7\_servers

Indexes:

* Asset ID - key
* Asset Tag - unique
* Hostname / inst - not unique

alter table a7\_servers add key (`Asset Id`);

alter table a7\_servers add unique (`Asset tag`);

alter table a7\_servers add index (`\*° Hostname / inst`)

Script: hw\_from\_a7.pl.

### Hardware Types and Relations

There is no known procedure to understand which hardware boxes are blade servers or blade enclosures.

## OVSD

### Hardware CIs

Source file in Sharepoint: HP\_ALU\_SLM\_*yyyy-mm-dd*-Server Report\_V1.zip, extract excel file from the zip. In the Excel file, the source is worksheet Server\_Attributes.

Target file in MySQL: ovsd\_servers

Indexes:

* CIID – unique

alter table ovsd\_servers add unique (`CIID`);

alter table ovsd\_servers add index(`ASSETTAG`);

Script: hw\_from\_ovsd.pl

### Hardware Types and Relations

Server Relations: from “Server Report on August 17th,2011.xls”, worksheet: “Server\_CI\_Relations”. Remove columns K (FROM-STATUS) up til CE (FROM-COVERAGE End Date). Remove all columns after “FROM-CATEGORY” and before “RELATIONSHIP”.

Target table: ovsd\_server\_rels

Verify that all FROM CIs are also in the ovsd\_server table:

SELECT \* FROM `ovsd\_server\_rels` WHERE `FROM-CIID` not in (select ciid from ovsd\_servers)

## ESL

### Hardware CIs

Source file for User “Datamigration ALU”: Hardware Component *YYMMDDhhmm*

Target file in MySQL: esl\_hardware\_extract

Script: hw\_from\_esl.pl

### Hardware Types and Relations

Source for for User “Datamigration ALU”: Relations *YYYYMMDDhhmm*

## Extract Hardware data

Script: create\_hw.pl.

Review output directory.

# Computersystem Component

## Assetcenter

No action required, same table as A7 Hardware Component is used.

Script: computersystem\_from\_a7.pl

## OVSD

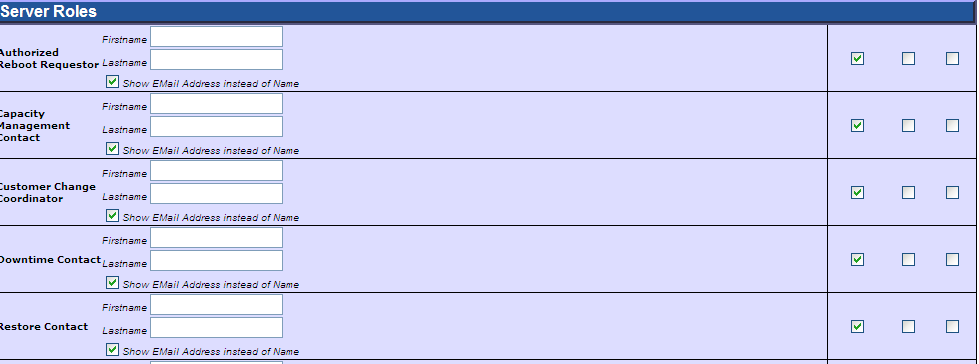
No action required, same table as OVSD Hardware Component is used.

Script: computersystem\_from\_ovsd.pl.

## ESL

|  |  |  |
| --- | --- | --- |
| Source File | Target File | Script |
| CS Techn gen YYMMDDhhmm | esl\_cs\_techn\_gen | cs\_techn\_gen\_from\_esl.pl |
| CS Techn IP YYMMDDhhmm | esl\_cs\_techn\_ip | cs\_techn\_ip\_from\_esl.pl |
| CS Techn Cons YYMMDDhhmm | esl\_cs\_techn\_cons | cs\_techn\_cons\_from\_esl.pl |
| CS Availability EMAIL YYMMDDhhmm | esl\_cs\_availability | cs\_availability\_from\_esl.pl |
| CS Admin YYMMDDhhmm | esl\_cs\_admin | cs\_admin\_from\_esl.pl |
| CS Funtions YYMMDDhhmm | esl\_cs\_functions | cs\_function\_from\_esl.pl |
| CS Usage YYMMDDhhmm | esl\_cs\_usage | cs\_usage\_from\_esl.pl |

CS Availability report: do flag to request email addresses instead of names for Contact Roles (e.g. Technical Owner). This setting is not saved in reports.



esl\_cs\_techn\_cons: errors with Console Notes – Field Truncation. To be handled later.

esl\_cs\_admin: errors with Contract Notes – Field Trunction. To be handled later.

The script cs\_techn\_gen\_from\_esl.pl needs to run first, to extract the master list of computersystems.

The script cs\_availability\_from\_esl.pl needs to run BEFORE cs\_admin\_from\_esl.pl.

The BAT file cs\_from\_esl.bat will run all scripts consecutively.

## Extract Computersystem data

Script: create\_computersystem.pl

# Solution Components

## ESL

Source: ‘Product EMAIL YYYYMMDD’. Don’t forget to check email addresses for selected contact details. This file has the Components and the Relations.

Target: esl\_instances – run script esl\_instance\_handling.pl. This will create the table esl\_instance\_work.

## Assetcenter

Source: HP\_ALU\_SLM\_YYYY\_MM\_DD\_SOL\_V2.xlsx

V2 of the extract has the User name and the CI Ownership information.

Target: a7\_solutions

Script: solutions\_from\_a7.pl

## OVSD

### Applications

Source: Application\_Report\_On\_Month\_day.xls, worksheet “Application\_Attributes”

Target: ovsd\_applications

### Interfaces information

Source: Application\_Report\_On\_Month\_day.xls, worksheet “Interfaces\_Report”

Target: ovsd\_interface

# Application Relations

## ESL

No specific handling.

## OVSD Applications

Source HP\_ALU\_SLM\_yyyy\_mm\_dd-Application\_Report\_V1.xls, worksheet “Application\_CI\_Relations”.

The worksheet cannot be imported into Access, since the records are too long. Modify the sheet by removing FROM columns, as they do not contain new information.

Note that the TO columns do contain additional information, about CI types that were not available before and that require separate handling.

Remove columns:

* Column D (FROM-ALIAS NAMES (4000)) to Column AN (FROM-RESOURCE UNIT)
* Enable Filter – to show columns after the last useful column (TO-RESOURCE UNIT). Remove the empty columns.

There are a lot of empty lines as well, but they can be handled by the scripts.

In Access Import, modify these columns to type ‘TEXT’:

* TO-MEMORY SIZE
* TO-CAPITAL ASSET TAG
* TO-REMOTE ACCESS

or use an easy work-around to force all columns to TEXT type (Access scans the first 25 rows of any column to determine the type):

* copy the first row in excel , to create a duplicate header row.
* import in Access
* Find and delete the the row in access, since it is the header row.

This works of course only when all column headers are available. This is the case in all known situations (maybe not always unique, but that is not the problem here).

Import to “ovsd\_apps\_rels”.

Script ovsd\_other\_ci\_types works on table ovsd\_apps\_rels. Extract ‘other’ CI Types from TO fields in the table.

## OVSD DB

Source: Database CI Report On MonthDD.xls

Target: ovsd\_db AND ovsd\_db\_rels

## Assetcenter

Source: YYYY\_MM\_DD\_ALL\_RELATIONSHIP.xlsx

This file is on the internal sharepoint <http://teams3.sharepoint.hp.com/teams/ALUInternal/ITO/SvM/CfgM/CMR/Forms/AllItems.aspx?RootFolder=%2Fteams%2FALUInternal%2FITO%2FSvM%2FCfgM%2FCMR%2FMonthly%20Asset%20Center%20CMDB%20Report&View=%7b662BB155%2d3969%2d41A2%2dB67C%2d43E54D515208%7d>.

Target: a7\_all\_relations.

Then run script a7\_all\_relations.pl to prepare work table.

~~Source: HP\_ALU\_SLM\_2011\_09\_03\_RELATION\_SOL\_SRV\_V2.xlsx~~

~~Target: a7\_rel\_sol\_srv~~

~~Source: Global\_EMEA\_Mapping\_SOL\_LOG to physical SRV\_2908.xlsx~~

~~Avoid Type Conversion Errors, by copying the title line to the second line on the spreadsheet. This will let Microsoft Access believe that all fields are Text fields.~~

~~Do not forget to remove first data line then in Microsoft Access.a~~

~~Target: a7\_sol\_log\_srv~~

~~Scripts:~~

* ~~a7\_sol\_log\_srv\_handling.pl – to add fieldnames, and to extract solution list a7\_sol\_extract~~
* ~~a7\_sol\_log\_srv\_relations.pl – to extract all relations, into table a7\_glob\_rel~~

# Portfolio Handling

## Source of Information

The Solution Portfolio file snapshot has a number of worksheets. The current snapshot (Week 3 – 2012) has following worksheets:

* IS-IT MGD => ‘Normal’ ALU Business Applications, import into pf\_is\_it\_mgd. Remove columns AF [Active User Count] – BB [MS Office Integration] to allow import.
* Business MGD => ALU Retained Business Applications, import into pf\_business\_mgd
* IS-IT Other =>Applications not in one of the classes above?, import into pf\_it\_other
* Working Portfolio => Ignore for now, do not import.

Note that ALU Business Application / ALU Retained Application does not come from Solution Portfolio File, but from OVSD / Assetcenter information.

## Workbook Modifications

In each workbook, “App ID” format must be TEXT, not double.

Do not add Primary Key.

Import Errors on ‘Service Window Comment’. This can be ignored for now, as this field is not used.

In directory TablePreparation, run the script pf\_handling.pl. This will consolidate all portfolio information into the PF table.

# Security File Handling

## Source of Information

The source of information is the excel sheet ‘Archer extract\_Application\_Data Classification Ratings\_120216 with Formula.xls’. The file is sent by Ken Hughes (ALU) in an email.

## Handling

Load the Excel file into table alu\_pf\_security.

# Locations

Source: ‘Location Information’

# Target: esl\_locationsComponent Template Review Process

## Introduction

This process describes the steps that are required to review Component Templates. Each time a component template has been changed, a series of steps are required. These steps are described in the process in the next sesssion.

The Component Template focuses on the Transition Model. The Protege Class model focuses on the Source and Target systems. Attributes discovered during Source and target system analysis are documented in Protege.

Although a full mapping between Component Template fields and Protege attributes is not mandatory, the fields and attributes that are not mapped must be very well understood to ensure that no information is lost or no important data is ommitted.

## Process

1. Start from the Component Template, map each template field to a Protege Attribute. Compare template field description with Protege Attribute comment. Ensure that equal fields only are mapped. This needs to be documented in a copy of the Component\_Interface\_Template file.
   * Result: each component template field on each template worksheet can be mapped to a Protege Attribute. Each attribute that is required for the transition model is known in the Protege model.
   * A component template field that is not mapped, will never get data from the source systems. This may be OK, but review carefully.
2. Extract the Protege Class Model for the Component. Review to use OntoViz to extract Class Model. In Ontoviz:
   * Select the Top Class for the Component Model
     + PhysicalBox for Hardware Component
     + ComputerSystem for Computersystem Component.
     + Application for Product Component
   * Enable two options
     + slx – slot extensions, to list classes linked to from the top class
     + sle – slot edges, to draw the lines from the top class to the extension classes.
   * Modify the options, set maximum depth for slot extension to “Inf” to see all extension classes.
   * For ComputerSystem, “configure slots” (S-symbol) to hide attributes that have dependencies from ComputerSystem to PhysicalBox and to Installed Product.
   * Launch the picture.
   * To see full picture:
     + select Application\_Instance
     + select options sub, slx and sle
     + Set option ‘max depth for slot extension (classes)’ to maximum
3. Make sure that each Protege Attribute in each class can be mapped to a Component Template field.
   * Result: each attribute can be mapped to a component template field.
   * An attribute that cannot be mapped, will never get into the transition model. Review that this is acceptable for the attribute.
4. Modify CIM Model to be in line with Protege Class Model.
5. For each Source System
   * Map each source system attribute to a Protege Attribute. This needs to be documented in the (ESL) Data Model Repository workbook. Add the mapping to the Component Template file, for easy reference.
   * Modify Extract Report handling: report definition in ESL, report handling for all source systems.
   * Modify Source System Extract script.
6. Modify Component CREATE script, to move data from CIM model to template.

# Appendix

## Overview of ESL Reports

|  |  |  |
| --- | --- | --- |
| ESL Report | Target File | Table |
| CS Techn gen YYMMDDhhmm | cs\_techn\_gen.csv | esl\_cs\_techn\_gen |
| CS Techn IP YYMMDDhhmm | cs\_techn\_ip.csv | esl\_cs\_techn\_ip |
| CS Techn Cons YYMMDDhhmm | cs\_techn\_cons.csv | esl\_cs\_techn\_cons |
| CS Availability EMAIL YYMMDD | availability.csv | esl\_cs\_availability |
| CS Admin YYMMDDhhmm | admin.csv | esl\_cs\_admin |
| CS Funtions YYMMDDhhmm | functions.csv | esl\_cs\_functions |
| CS Usage YYMMDDhhmm | usage.csv | esl\_cs\_usage |
| Product EMAIL YYYYMMDD | product.csv | esl\_instance |
| Hardware Component YYYMMDD | hardware\_extract.csv | esl\_hardware\_extract |
| Location Information | location.csv | esl\_locations |
| Relations YYYYMMDD | relations.csv | esl\_relations |