

# EMF Surrogate Tool Case

## Surrogate Tool Case Overview

The Surrogate Tool Case is a relatively simple Case that encapsulates all the information necessary to run an installation of the Surrogate Tool, allowing an EMF user to create and QA surrogates. This is the main paradigm employed by the EMF: through the use of Jobs, Inputs, Parameters, and scripts, an EMF user is able to integrate with and run external applications (e.g., the Surrogate Tool, SMOKE, etc.).

The Surrogate Tool Case predefines 4 Jobs:

- 2 “Create Surrogates” Jobs for creating surrogates using a 16km (coarse) grid and an 8km (fine) grid
- 2 “QA Surrogates” Jobs for QAing already created surrogates using a 16km (coarse) grid and an 8km (fine) grid

## Install Spatial Allocator (contains the Surrogate Tool)

Use [Surrogate Tool User Guide](#) to download and install the Surrogate Tool (Section 2), noting the directory that is referred to as SA\_HOME. You do not need to modify your .cshrc file to include SA\_HOME because these environmental variables will be dealt with by the EMF Case. This directory should be available from both the EMFServer and the SMOKEServer.

In addition to the Spatial Allocator application, you will need to download the “Sample Emission Surrogate Data Shapefiles” from the CMAS website, extracting the tar.gz file into the same location as the Spatial Allocator tar.gz. Here is an example directory structure:

```
$ mv sa_03_2009.tar.gz /data4/surg_tool/  
$ mv sample_srgtools_data.tar.gz /data4/surg_tool  
$ cd /data4/surg_tool  
$ tar xvzf sa_03_2009.tar.gz  
$ tar xvzf sample_srgtools_data.tar.gz
```

In this example, SA\_HOME is:

```
/data4/surg_tool/sa_03_2009
```

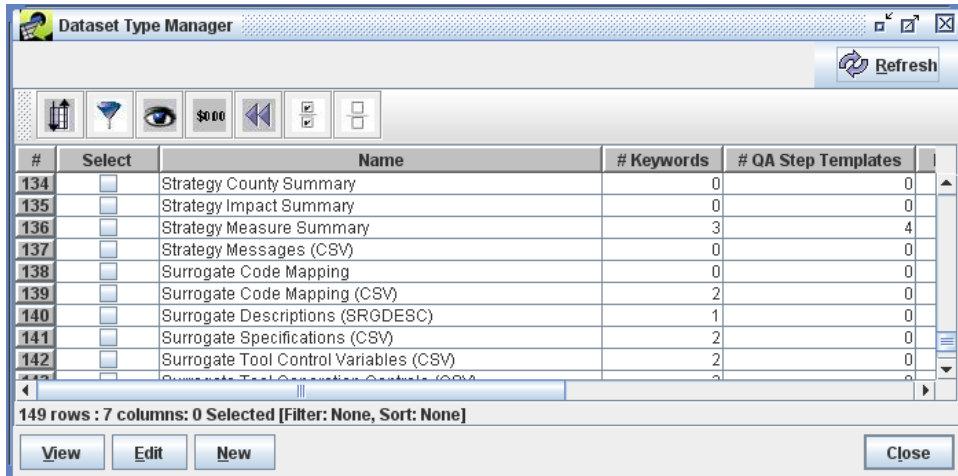
Make sure that SA\_HOME and the rest of the directory tree have the proper permissions (see [Directory Permissions](#)).

## Create New Dataset Type

For the Surrogate Tool Case, two new Dataset types will need to be created (Note: This step requires the EMF client user to have administrative privileges).

1. From the EMF client, launch the *Dataset Type Manager* (*Manage > Datasets* from the main menu)
2. From the *Dataset Type Manager* window, click the **New** button.
3. In the *Create New Dataset Type* window:
  - a) Enter “Surrogate Code Mapping” in the **Name** field. The name needs to match this exactly.

- b) Select “Line-based File” from the **Derived From** field
  - c) Click the **Save** button
4. From the *Dataset Type Manager* window, click the **Refresh** button. You should now see a new Dataset Type “Surrogate Code Mapping”:



5. From the *Dataset Type Manager* window, click the **New** button again.
6. In the *Create New Dataset Type* window:
  - a) Enter “Surrogate Tool QA Reports (External Multifile)” in the **Name** field. The name needs to match this exactly.
  - b) Select “External File” from the **Derived From** field
  - c) Click the **Save** button
7. From the *Dataset Type Manager* window, click the **Refresh** button. You should now see a new Dataset type “Surrogate Tool QA Reports (External Multifile)”.
8. Now we are going to add some keywords that control the behavior of the datasets when they are exported.
  - a) Select the dataset type "Surrogate Code Mapping" and click the **Edit** button.
  - b) Click the **Add** button at the bottom-of the keywords section
  - c) In the ‘Keyword’ cell of the new table row, select the ‘EXPORT\_COLUMN\_LABEL’ from the pull-down
  - d) In the ‘Value’ cell, enter ‘ALL\_CAPS’
  - e) Click the **Add** button again
  - f) In the ‘Keyword’ cell of the new table row, select the ‘EXPORT\_HEADER\_COMMENTS’ from the pull-down
  - g) In the ‘Value’ cell, enter ‘false’. Your *Edit Dataset Type* window should now look like this:

**Edit Dataset Type: Surrogate Code Mapping \***

Name:

Description:

Default Sort Order:

Select	Keyword	Value
<input type="checkbox"/>	EXPORT_COLUMN_LABEL	ALL_CAPS
<input type="checkbox"/>	EXPORT_HEADER_COMMENTS	false

**QA Step Templates**

Select	Name	Program	Arguments	Required	Order
--------	------	---------	-----------	----------	-------

h) Click the **Save** button

## Configuring the Case

1. Download [surgtool.EMF.tar.gz](http://surgtool.EMF.tar.gz) and untar the file.

```
$ tar xvzf surgtool_EMF.tar.gz
```

2. Move (or copy) and extract the additional shapefiles to the shapefiles data directory under the SA\_HOME directory:

```
$ cp surgtool_EMF/addl_shapefiles/*
  /data4/surg_tool/sa_03_2009/data/emiss_shp2003/us
$ cd /data4/surg_tool/sa_03_2009/data/emiss_shp2003/us
$ unzip pophu2k_tnnc.zip
$ unzip tnnc_usrds_2000.zip
```

3. Move (or copy) the surgtool\_EMF/scripts/emf directory and contents to your shared disk (so that they can be accessed by both the EMFServer and SMOKEServer). Make sure the emf directory and the individual scripts are 'rx' for the emiss group. These are the surrogate tool driver scripts that have been EMF enabled. An example directory for these scripts is:

```
/data4/surg_tool/scripts/emf
```

4. Move (or copy) the case to be imported, `surgtool_EMF/surgtool_case_setup`, to your import directory on the EMFServer. An example directory for the case imports is:

`/data3/imports/cases/surgtool_case_setup`

5. As with the previous case when the EMF job header was imported, this Case requires the import of some additional input files. You may need to move the files to an import area so that they can be read on the EMFServer (e.g. `/data3/imports`). Use the *Dataset Manager* to import the following files:

- a) `surgtool_EMF/in/surrogate_generation_grid.csv` should be imported as dataset type “Surrogate Tool Generation Controls (CSV)”, changing the dataset name to “`surrogate_generation_grid_beta`”
- b) `surgtool_EMF/in/GRIDDESC.txt` should be imported as dataset type “Grid Description (Line-based)”, changing the dataset name to “`GRIDDESC_beta`”
- c) `surgtool_EMF/in/shapefile_catalog.csv` should be imported as dataset type “Shapefile Catalog (CSV)”, changing the dataset name to “`shapefile_catalog_beta`”
- d) `surgtool_EMF/in/surrogate_codes` should be imported as dataset type “Surrogate Code Mapping”, changing the dataset name to “`surrogate_codes_beta`”
- e) `surgtool_EMF/in/surrogate_specification.csv` should be imported as dataset type “Surrogate Specifications (CSV)”, changing the dataset name to “`surrogate_specification_beta`”

6. Use the *Case Manager* to import the Surrogate Tool Case. The files to import are:

`surgtool_case_setup/Surrogate_Tool_Example_Inputs.csv`  
`surgtool_case_setup/Surrogate_Tool_Example_Jobs.csv`  
`surgtool_case_setup/Surrogate_Tool_Example_Summary_Parameters.csv`

Once the import has completed, you should see a new Case under the “Demo” category called “Surrogate Tool Example”. You may need to refresh the manager to see the new Case.








Select the new Case and click **Edit**, which will launch the *Case Editor*. You should see, among other things;

- 4 Jobs (2 “Create Surrogates” Jobs and 2 “QA Surrogates” Jobs):

Case Editor: Surrogate Tool Example

Summary Jobs Inputs Parameters Outputs History

Output Job Scripts Folder: \$SRG\_ROOT/cases/\$CASE Browse

#	Select	Name	Order	Region	Sector	Run Status	Running User	Last Message
1	<input type="checkbox"/>	Create Surrogates Coarse	0.00	NASH_16KM	All sectors	Completed		201
2	<input type="checkbox"/>	Create Surrogates Fine	0.00	NASH_8KM	All sectors	Completed		201
3	<input type="checkbox"/>	QA Surrogates Coarse	1.00	NASH_16KM	All sectors	Completed		201
4	<input type="checkbox"/>	QA Surrogates Fine	1.00	NASH_8KM	All sectors	Completed		201

4 rows : 23 columns: 0 Selected [Filter: None, Sort: Order(+), Sector(+), Name(+), Executable(+)]

Add Remove Edit Copy Modify Run Validate Set Status Cancel

Refresh Load Export Save View Parent View Related Close








- 6 Inputs:

Case Editor: Surrogate Tool Example

Summary Jobs Inputs Parameters Outputs History

Input Folder: /orchid/share/surgtool/cases/\$CASE/in Browse

Sector: Select a Sector Environment Variable Contains:

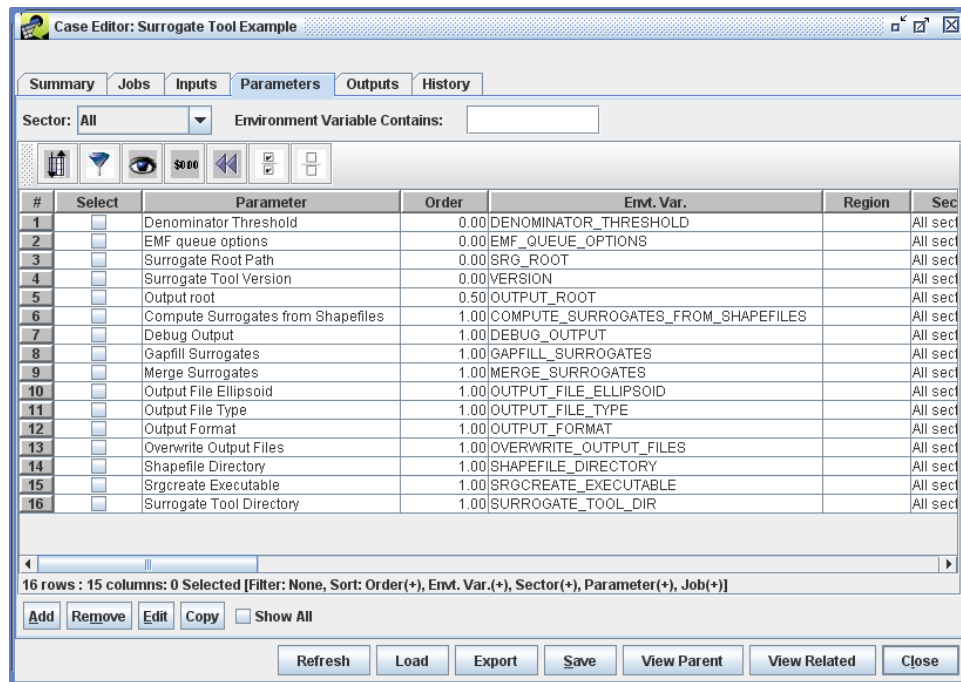
#	Select	Input	Env. Var.	Region	Sector	Job	Progr
1	<input type="checkbox"/>	EMF Job Header	EMF_JOBHEADER		All sectors	All jobs for sector	All progra
2	<input type="checkbox"/>	Generation Control File	GENERATION_CONTROL_FILE		All sectors	All jobs for sector	All progra
3	<input type="checkbox"/>	Grid Description List	GRIDDESC		All sectors	All jobs for sector	All progra
4	<input type="checkbox"/>	Shapefile Catalog	SHAPEFILE_CATALOG		All sectors	All jobs for sector	All progra
5	<input type="checkbox"/>	Surrogate Code File	SURROGATE_CODE_FILE		All sectors	All jobs for sector	All progra
6	<input type="checkbox"/>	Surrogate Specification File	SURROGATE_SPECIFICATION_FILE		All sectors	All jobs for sector	All progra

6 rows : 15 columns: 0 Selected [Filter: None, Sort: Env. Var.(+), Sector(+), Input(+), Job(+)]

Add Remove Edit Copy View Dataset Export Inputs Find ☐ Show All

Refresh Load Export Save View Parent View Related Close

- 16 Parameters:



7. As with the previous test Case, you will need to modify the imported case to match your system:

From the *Jobs* tab:

- a. Modify the **Output Job Scripts Folder** field. An example directory is:

`/data4/surg_tool/scripts/cases/$CASE`

The EMF server will replace the variable \$CASE by the case abbreviation. It will create this subdirectory when the first job is run. Make sure that the rest of the directory tree has been created and has the proper permissions (see [Directory Permissions](#)).

- b. For each Job, you will need to change the **Host** and **Queue Options** field. Note that you can do all of them simultaneously by selecting the jobs and then click the **Modify** button.
- c. For each “Create Surrogates” Job, you will need to change the **Executable** field to point to the *surrogate\_wrapper.csh* script on your system. This can be done with the **Modify** button.
- d. For each “QA Surrogates” Job, you will need to change the **Executable** field to point to the *surrogate\_qa\_wrapper.csh* script on your system. Again, this can be done with the **Modify** button.

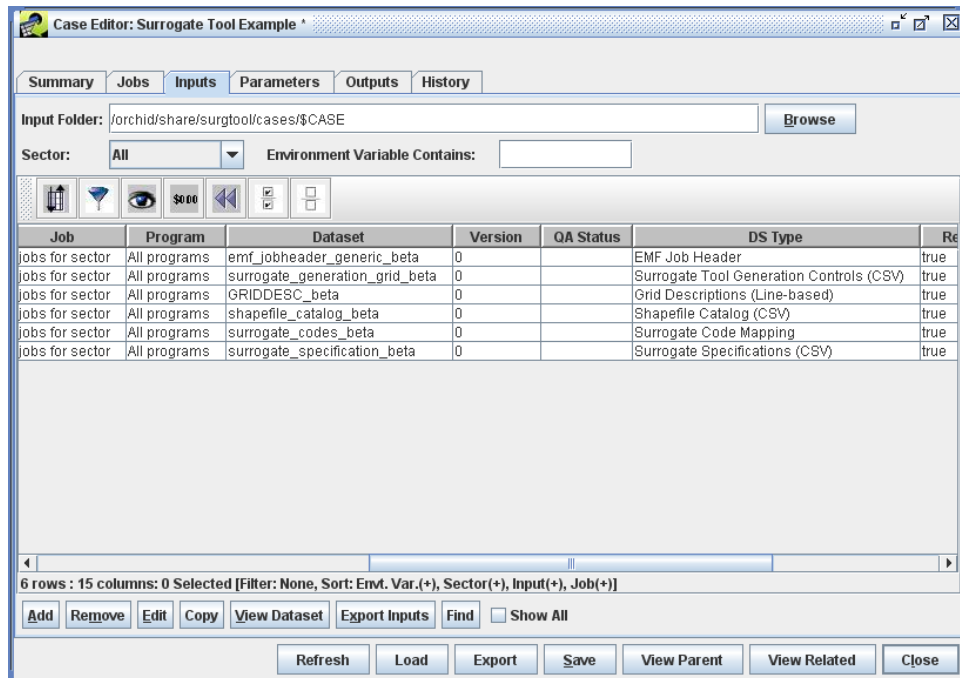
From the *Inputs* tab:

- a. Modify the **Input Folder** field. An example directory is:

```
/data4/surg_tool/inputs/$CASE
```

The EMF server will replace the variable \$CASE by the case abbreviation. It will create this subdirectory when the first job is run. Make sure that the rest of the directory tree has been created and has the proper permissions (see [Directory Permissions](#)).

- b. During the import of the Case, each Input should have found a matching dataset. In other words, the Dataset and Version columns should be filled in for each Input. If they are not filled in, the importer could not find a dataset that exactly matched the name and version in the import files. If your Inputs are missing any datasets, follow the appropriate step below:
- i. Modify the “EMF Job Header” Input to use the same emf job header dataset from the previous test Case. Select the Input and click **Edit**.
  - ii. Modify the “Generation Control File” Input to use the “surrogate\_generation\_grid\_beta” (from above)
  - iii. Modify the “Grid Description List” Input to use the “GRIDDESC\_beta” (from above)
  - iv. Modify the “Shapefile Catalog” Input to use the “shapefile\_catalog\_beta” (from above)
  - v. Modify the “Surrogate Code File” Input to use the “surrogate\_codes\_beta” (from above)
  - vi. Modify the “Surrogate Specification File” Input to use the “surrogate\_specification\_beta” (from above)
  - vii. Click the **Refresh** button to see the changes:



From the *Parameters* tab:

- Select “All” from the **Sector** pull-down near the top of the tab
- A few parameters will need to be modified to match your system:
  - Modify the ‘EMF queue options’ parameter as in the previous test Case. As a reminder: If you are using a real process queue, select the parameter and click **Edit**. If you are using the fake queue, the value of the 'EMF queue options' does not matter.
  - Modify the “Surrogate Root Path” parameter to match the path to the Surrogate Tool (referred to as SA\_HOME above). For example:

/data4/surg\_tool/sa\_03\_2009

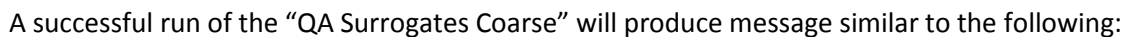
## Running the Case

The Jobs can be run as they were in the previous test Case, with the exception that the each “QA Surrogates” Job depends on the output from its “Create Surrogates” partner. What this means, for example, is that prior to running the “QA Surrogates Coarse” Job, there will need to have been a successful run for the “Create Surrogates Coarse” Job.

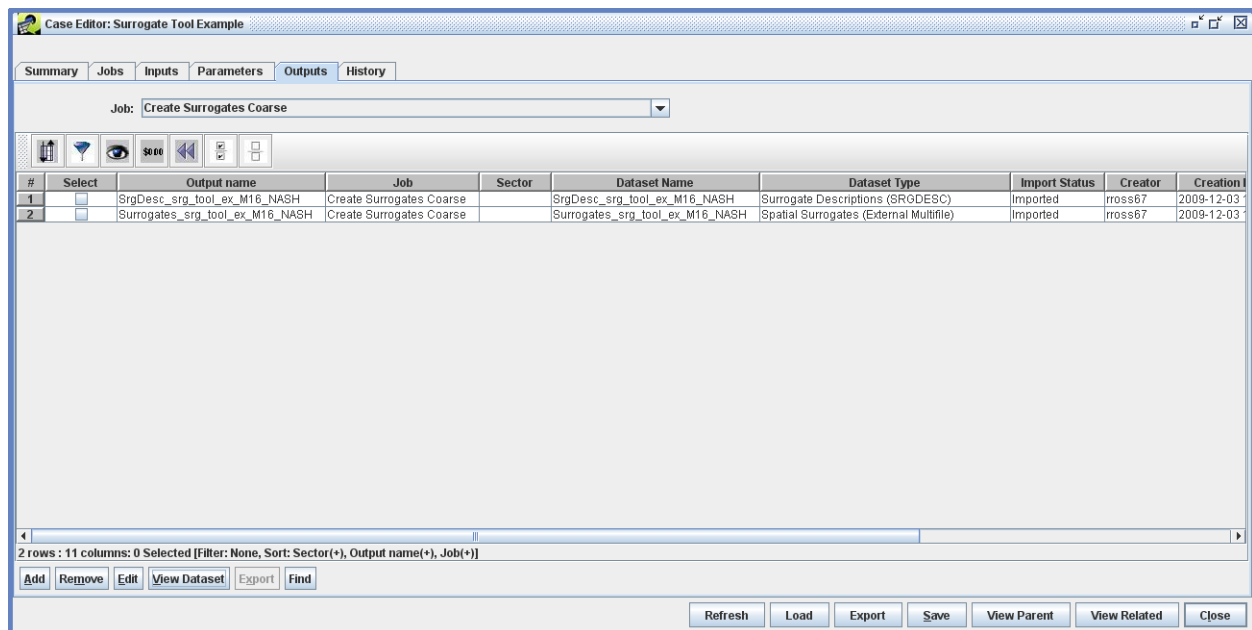
## Messages

A successful run of the “Create Surrogates Coarse” will produce message similar to the following:

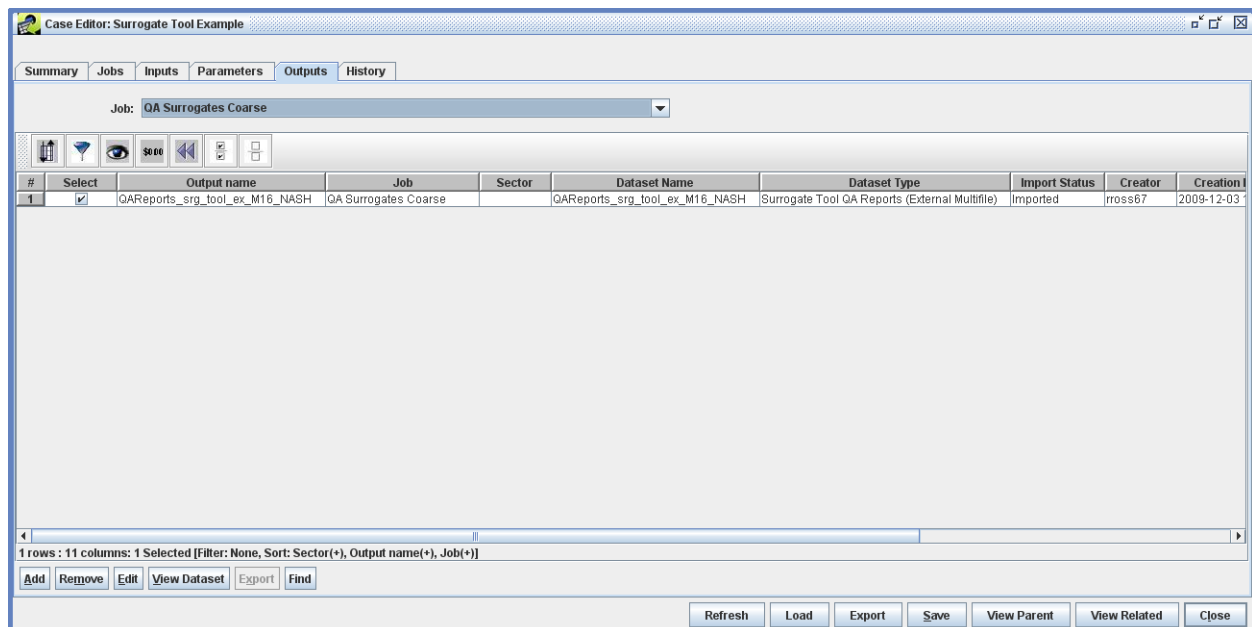




A successful run of the “Create Surrogates Coarse” will produce the following outputs:



A successful run of the “QA Surrogates Coarse” will produce the following output:



## Logs

In addition to the EMF logs that are located in `<Output Job Scripts Folder>/logs` and the Tomcat `catalina.out` log file located in `<tomcat home>/logs`, the Surrogate Tool will produce logs in `$OUTPUT_ROOT/$OUTPUT_GRID_NAME/logs`, where `$OUTPUT_ROOT` and `$OUTPUT_GRID_NAME` are defined Case Parameters. These three locations will be invaluable in troubleshooting this test Case.

## Trouble-shooting

If a Job fails (i.e., **Run Status** is “Failed”) or did not produce the expected Outputs, the first place to look is in the History tab of the Case. For some errors, the EMF wrapper scripts will send back meaningful error messages. If the History doesn't adequately diagnose the problem, look at the EMF job log file corresponding to the specific run (Note: Each run of a Job is distinct and has a unique timestamp in the name. If no other jobs have been run since the failure, use `ls -lrt` to get a time-ordered listing of the log files, the one you are looking for being the last one listed).

If the Job fails and there is a series of messages in the EMF Job log file similar to:

```
EMF cmd client: Python client done Wed Dec  2 12:49:59 2009
Your Operating system is:  linux
```

```
Using Java version of gapfilling and merging
```

```
Run Date: Wed Dec 02 12:49:59 EST 2009
```

```
End Date: Wed Dec 02 12:49:59 EST 2009
```

```
Elapsed time in minutes: 0.0011833333333333333
```

```
ERROR -- The Program Run Stopped. See log file for details.
```

```
EMF cmd client: python client starting: Wed Dec  2 12:49:59 2009
```

This means that the script successfully launched the Surrogate Tool, but the Surrogate Tool failed for some reason. You will now want to look in the Surrogate Tool log file. A successful run of the Surrogate Tool will have “SUCCESS -- The Program Run Completed” as the last line of the log file.

If an error cannot be found in either the EMF or the Surrogate Tool log files, you should look for errors in the Tomcat `catalina.out` log file.