calnt 0017 - Why is my study stuck in the 'Processing' status hours after I deployed it

Question: Study stuck in the 'Processing' status hours after I deployed it?

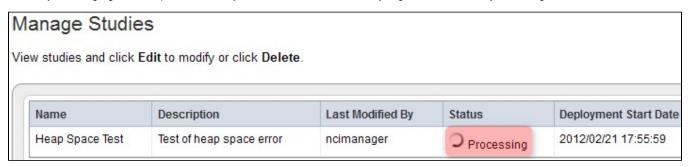
Topic: calntegrator usage

Release: all versions

Date entered: 2/22/2012

Details About the Question

When you deploy a study, you may find its status showing as 'Processing' on the 'Manage Studies' page hours or even days later (see screenshot below with study status highlighted in red). The status may remain stuck like this indefinitely, regardless of how fast your calntegrator server is.



Answer

The most common cause of this problem is an out of memory error caused by limited heap space in the Java Virtual Machine on the JBoss server instance running calntegrator. If a study deployment fails due to this error, calntegrator does not notify the user explicitly and instead logs the error in the server.log file located at the following path:

[MATKC:installation root]\caintegrator2\jboss-4.0.5.GA\server\default\log

Note that the study's status will continue to show as 'Processing' on the 'Manage Studies' page even after the deployment has failed and the error has been logged.



Note

Studies may show a Status error indicating a timeout after 48 hours when in fact the study is still properly deploying, As a result, a study showing a timeout error should not be deleted or edited. In such a case, the server log correctly indicates that no error or failure has occurred.



Warning

calntegrator is not able to deploy studies using Affymetrix CEL files. calntegrator is able to deploy studies using Affymetrix CHP files loaded as parsed data in caArray or Affymetrix TXT files loaded as imported (not parsed) data in caArray.

In Windows, the heap size is set in the 'run.bat' file located at the following path:

[MATKC:installation root]\caintegrator2\jboss-4.0.5.GA\bin

In Linux, the heap size is set in the 'run.conf' file located at the following path:

[MATKC:installation root]/caintegrator2/jboss-4.0.5.GA/bin

By default, the heap size, which is dynamically allocated, is set at a minimum of 256 MB and a maximum of 512 MB, which is not nearly enough when deploying studies with large datasets. For instructions on how to modify the heap size by editing 'run.bat', please refer to the following page from the calntegrator local installation guide:

https://wiki.nci.nih.gov/display/caIntegrator/caIntegrator+1.3+Local+Installation+Guide#caIntegrator1.3LocalInstallationGuide-ConfiguringJBoss

The minimum heap space should be set to 4096 MB (4 GB), assuming that your calntegrator server has this amount of physical memory available.

The recommended heap size varies greatly depending on the size of your dataset and the amount of available physical memory on your calntegrator server. For reference, for a dataset containing 500 Affymetrix CEL files that are approximately 16GB in combined size, the minimum heap size required for the study deployment to complete successfully is 15 GB.

Ideally, calntegrator should be run on a dedicated server, with the heap size set as close as possible to the amount of available physical memory without destabilizing the underlying operating system.

The tables below shows the results of extensive testing of calntegrator study deployments on different hardware configurations with varying amounts of heap space.

REFERENCE INFORMATION:

- Trials #1 and #2 were performed on a Dell Optiplex 755 workstation running Windows XP Professional
- The workstation runs on an Intel Core2 Quad Q6600 processor at 2.40 Ghz
 The total installed physical memory is 3.25 GB, with approximately 1.75 GB available at the time of testing before launching calntegrator

Trial #1 (The heap space setting as specified in run.bat is -Xms256m -Xmx512m)

| # of samples mapped | Total size of samples (MB, uncompressed) | Deployment Status | Time to deploy or fail (minutes:seconds) | |
|---------------------|--|-------------------|--|------------------|
| 1 | 2 | SUCCESS | 1:00 | * time not exact |
| 2 | 4 | SUCCESS | 0:47 | * time not exact |
| 4 | 7.8 | SUCCESS | 1:15 | * time not exact |
| 8 | 15.5 | SUCCESS | 1:50 | * time not exact |
| 16 | 31.2 | SUCCESS | 3:15 | |
| 64 | 124.8 | SUCCESS | 13:55 | |
| 128 | 249.6 | FAIL | 21:16 | |
| 192 | 374.4 | FAIL | 23:47 | |
| 224 | 436.8 | FAIL | 25:44 | |
| 256 | 499.2 | FAIL | 1h 5:02 | |

Trial # 2 (The heap space setting as specified in run.bat is -Xms256m -Xmx1024m)

| # of samples mapped | Total size of samples (MB, uncompressed) | Deployment Status | Time to deploy or fail (minutes:seconds) |
|---------------------|--|-------------------|--|
| 1 | 2 | SUCCESS | 0:10 |
| 4 | 7.8 | SUCCESS | 0:21 |
| 16 | 31.2 | SUCCESS | 1:08 |
| 64 | 124.8 | SUCCESS | 5:12 |
| 128 | 249.6 | SUCCESS | 12:48 |
| 192 | 374.4 | SUCCESS | 21:35 |
| 224 | 436.8 | FAIL | 27:59 |
| 256 | 499.2 | FAIL | 34:08 |

- Trial #3 was performed on a Dell Poweredge server running Linux
- The server runs on a quad-core 2.33 Ghz Intel(R) Xeon(R) 5148 CPU
- The total installed physical memory is 16 GB

Trial #3 (The heap space setting as specified in run.bat is -Xms2048m -Xmx2048m)

| # of samples mapped | Total size of samples (MB, uncompressed) | Deployment Status | Time to deploy or fail (minutes:seconds) |
|---------------------|--|-------------------|--|
| 192 | 374.4 | SUCCESS | 18:12 |
| 208 | 405.6 | SUCCESS | 41:16 |
| 224 | 436.8 | SUCCESS | 27:21 |
| 256 | 499.2 | SUCCESS | 33:13 |
| 512 | 998.4 | FAIL | 4h 47:23 |

Have a comment?

Please leave your comment in the calntegrator End User Forum.