Analysis of pipeline success rate

Wouter Klijn 01-02-2013

**Abstract: An analysis of the pipeline success rate was performed for LOFAR software version 1.9+. Causes for failures were tallied and grep commands were constructed matching on the log files with specific causes. The pipeline framework displayed a failure rate of 1.4%, due to (currently solved) bugs. 25.7% of all pipelines failed due to incorrect specification or missing input data. The analysis was performed with a script which can be used to perform this analysis on single observation. In this mode it displays detailed error information and the matching grep command to perform additional investigations manually.**

## Data set

The analysis is limited to pipeline runs performed with LOFAR release 1.9 and higher. The parsets for these pipelines are located in /opt/cep/lofar/var/run, The log files in /opt/cep/lofar/var/run/pipeline. Older pipeline version logfiles are located in /opt/cep/lofar/var/run/pipeline/jobs. These logs are **not** analyzed.

## Method

A bash script is created parsing all the available log files. Upon encountering a failed pipeline notification in the log, the script attempts to match a regular expression with the contents of a log file. Currently 16 different failure causes can be recognized by the script accounting for more than 99% of all failures.

Two types of failures are covered. Firstly the simple grep which will triggers when a single instance of the ‘grepped’ regular expression is found. Secondly the counted grep which triggers if more than 20 instances are found, typically for missing files.

## Results

Output of script (11:50 17-01-2013)

$ pipeline\_log\_analyzer.sh

-------------------------------------------------------------

logfiles containing unknown error:

------------------------------------------------

number of parsets : 1170

total number of pipeline logfiles: 1220

Failing pipelines : 349

pipeline type : total , error , explained

-------------------------------------------------

unknown\_pipeline : 0 , 0 , 0

msss\_imager\_pipeline : 67 , 22 , 21

msss\_calibrator\_pipeline : 413 , 110 , 109

msss\_target\_pipeline : 720 , 210 , 210

calibration\_pipeline : 7 , 5 , 5

preprocessing\_pipeline : 13 , 2 , 2

------------------------------------------------

detailed information simple grep

count : msg

8 : NPPP parset: incorrect values in the parset: Demix average

3 : NPPP parset: time window incorrect (itsNTimeChunk \\* itsNTimeAvg) % itsNTimeAvgSubtr !=0

1 : GSM:An error occured in a call to the Global Sky model

3 : FIXED: Incorroct parset or mapfile implementation: tuple object has no attribute

2 : PARSET: 0 Output\_SkyImage data products specified

13 : FIXED: Incorrect number of input data products

27 : BBS PARSET: <x>.skymodel is an invalid value for FileField skymodel

2 : BBS DATA: Key xx contains no solvable parameters for step xx: bbs aborted on a step

10 : GSM/SOURCEDB: Patch name x multiply defined. Incorrect local sky model from GSM

46 : VDSMAKER MISSING INPUTS: Exception caught: No such file or directory

12 : FIXED: Race issue between thread and socked connection: solved

29 : FIXED: incorrect deletion of vds files, possible duplicate input dataproducts

5 : MANUALLY ABORTED: ctrl-c (-or- abort from mac )

2 : OTHER: Failing pipelines with known singular cause of failure

------------------------------------------------

detailed information counted greps

count : msg

175 : MISSING INPUT FILES: A large number of copier steps failed

9 : MISSING INPUT FILES: vdsmaker reported A large number of missing input measurement sets

2 : MISSING INPUT FILES: copier: rsync: No such file or directory

-----------------------------------------------

Table 1: List of encountered errors, sorted on number of occurrences. Percentages are calculated on the total amount of pipelines ran. The third column contains the rough error type, with the fourth column a short, technical, description. Percentages might not add to a 100% due to rounding errors.

|  |  |  |  |
| --- | --- | --- | --- |
| N. | % | Error type | Detailed error information |
| 175 | 14,3% | Data | MISSING INPUT FILES. A large number of copier steps failed |
| 46 | 3,8% | Data | VDSMAKER MISSING INPUTS, Exception caught No such file or directory |
| 29 | 2,4% | Specification | FIXED incorrect deletion of vds files possible duplicate input dataproducts |
| 27 | 2,2% | Specification | BBS PARSET <x>.skymodel is an invalid value for FileField skymodel |
| 13 | 1,1% | Specification | FIXED Incorrect number of input data products |
| 12 | 1,0% | BUG | FIXED Race issue between thread and socked connection solved |
| 10 | 0,8% | Other | GSM/SOURCEDB. Incorrect local sky model from GSM |
| 9 | 0,7% | Data | MISSING INPUT FILES. vdsmaker reported A large number of missing input |
| 8 | 0,7% | Specification | NPPP parset incorrect values in the parset Demix average |
| 5 | 0,4% | Other | MANUALLY ABORTED ctrl-c (-or- abort from mac ) |
| 3 | 0,2% | Specification | NPPP parset time window incorrect ( |
| 3 | 0,2% | BUG | FIXED Incorrect parset or mapfile implementation tuple object has no attribute |
| 2 | 0,2% | Specification | PARSET 0 Output\_SkyImage data products specified |
| 2 | 0,2% | Other | BBS DATA Key xx contains no solvable parameters for step xx bbs aborted |
| 2 | 0,2% | BUG | OTHER Failing pipelines with known singular cause of failure |
| 2 | 0,2% | Data | MISSING INPUT FILES. copier rsync No such file or directory |
| 1 | 0,1% | Other | GSM An error occurred in a call to the Global Sky model |
| 349 | 28,6% | ERRORS |  |
| 1220 | 100,0% | TOTAL |  |

Table 2: Listing of the encountered error types, including the absolute and relative percentages. Percentages might not add to a 100% due to rounding errors.

|  |  |  |  |
| --- | --- | --- | --- |
| Error type | Occurances | Absolute percentage | Percentage relative to the number of errors |
| Data | 232 | 19,0% | 66,5% |
| Specification | 82 | 6,7% | 23,5% |
| Other | 18 | 1,5% | 5,2% |
| BUG | 17 | 1,4% | 4,9% |

Of all the pipeline runs 28.6% ended in an error state. The script recognizes 99.8% of these failures. Missing input is the most common cause (66.5% of all errors), then incorrect specification(23.5%) and problems in internal executables (5.2%), known and solved bugs account for 4.9% of the encountered errors. A single failed run was caused by a known and analyzed low priority bug.

Conclusions  
The pipelines and called executables had a failure rate of 2.9%. The failure rate due to bugs in the pipeline software itself was 1.4%. The cause of all failed pipelines is known. With the majority of failures due to missing input data and/or incorrect specification.

## HOWTO Observation specific analysis

The script written for this analysis can be applied to a single observation or range of observations. The script can be found at :

lhn001:/opt/cep/operations/bin/pipeline\_log\_analyzer.sh  
The command line argument –help will display a detailed usage message. Ranges of Observation ids can be analysis using default command line astrixes:

pipeline\_log\_analyzer.sh 8083\*

-------------------------------------------------------------

logfiles containing unknown error:

---------------------------------------------------

/opt/cep/lofar/var/run/pipeline/Observation80830/logs/2012-12-17T16:01:41/pipeline.log

FIXED: Incorrect number of input data products

Manually: grep -c "msss\_target\_pipeline.bbs\_reducer: Validation of input data mapfiles failed" <file>

---------------------------------------------------

/opt/cep/lofar/var/run/pipeline/Observation80833/logs/2012-12-17T16:32:41/pipeline.log

FIXED: Incorrect number of input data products

Manually: grep -c "msss\_target\_pipeline.bbs\_reducer: Validation of input data mapfiles failed" <file>

---------------------------------------------------

/opt/cep/lofar/var/run/pipeline/Observation80836/logs/2012-12-17T17:03:41/pipeline.log

FIXED: Incorrect number of input data products

Manually: grep -c "msss\_target\_pipeline.bbs\_reducer: Validation of input data mapfiles failed" <file>  
  
All log files for the matching observation ids will be displayed in order of creation. The last logfile will be the latest run.