Analysis of pipeline success rate

Wouter Klijn 3-09-2015

**Abstract: An analysis of the pipeline logfiles was performed for pipeline runs between the data 2014-04-10 and 2015-07-22. In total 33685 logfiles where analyzed. For a total of 33118 parsets. For repeated runs all logfiles were analyze. In total 1.8% of all pipelines in the time range failed. 27% of all errors were in the Pulsar pipeline. Another 35% due to missing input data. 11% due to specification errors.**

## Data set

The analysis is limited to pipeline runs performed with LOFAR release pipeline runs between the data 2014-04-10 and 2015-07-22. 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 34 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 8 instances are found, typically for missing files.

## Results

Output of script (14:44 2015-07-22)

$ pipeline\_log\_analyzer.sh

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

Analysis of large number of files is time consuming, no progress indication possible in the first steps

logfiles containing unknown error:

/opt/lofar/var/run/pipeline/Observation341156/logs/2015-04-28T08:44:28/pipeline.log.gz

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

number of parsets : 33118

total number of pipeline logfiles: 33685

Failing pipelines : 620

pipeline type : total , error , explained

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

unknown\_pipeline : 0 , 0 , 0

msss\_imager\_pipeline : 22 , 17 , 17

msss\_calibrator\_pipeline : 553 , 32 , 30

msss\_target\_pipeline : 9180 , 86 , 85

calibration\_pipeline : 0 , 0 , 0

preprocessing\_pipeline : 13648 , 191 , 190

long\_baseline\_pipeline : 8881 , 89 , 89

pulsar\_pipeline : 1361 , 184 , 183

imaging\_pipeline : 40 , 21 , 21

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

detailed information simple grep

count : msg

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

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

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

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

3 : PARSET: 0 Output\_SkyImage data products specified

32 : FIXED: Incorrect number of input data products

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

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

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

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

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

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

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

42 : Nr. antennas in data is different to meta data bug #8157

59 : The number of input and output dataproducts is incorrect

172 : An error occured in a pulsar pipeline, inform Vlad Condratu

1 : Suspected connectivity problems on locus nodes, test ssh connection to affected nodes

1 : The pipeline was stopped due to reboot of system, (stopday?)

3 : Incorrect executable configuration, warn pipeline developer

5 : Casacore problem, warn pipeline developer or Ger van Diepen

2 : An input data set is empty, check predecessor

1 : Major error in on of the executables. Check input data, else warn pipeline dev.

6 : Network reset while communicating with node. If problem persist for the same node suspect faulty hardware, else restart

4 : Known and solved bug in pulp/pipeline integration

7 : Known and solved bug in pulp/pipeline integration

6 : Executable has problem finding librairy, suspect deployment error

2 : Internal pipeline problem warn pipeline developer

95 : Makevds failed on all nodes, suspect corrupt input data.

1 : NDPPP specification error

1 : Error in BBS, check inputdata and specification

1 : suspect incorrect specification of input or output products

4 : OTHER: Failing pipelines with unique singular cause of failure

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

detailed information counted greps

count : msg

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

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

0 : 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.

|  |  |  |  |
| --- | --- | --- | --- |
| N. | % | Error Type | Detailed error information |
| 172 | 27.8% | Unknown | An error occured in a pulsar pipeline. inform Vlad Condratu |
| 98 | 15.8% | Data | MISSING INPUT FILES: A large number of copier steps failed |
| 95 | 15.3% | Data | Makevds failed on all nodes. suspect corrupt input data. |
| 59 | 9.5% | Specification | The number of input and output dataproducts is incorrect |
| 42 | 6.8% | BUG | Nr. antennas in data is different to meta data bug #8157 |
| 40 | 6.5% | Other | MANUALLY ABORTED: ctrl-c (or abort from mac ) |
| 33 | 5.3% | Data | VDSMAKER MISSING INPUTS: Exception caught: No such file or directory |
| 32 | 5.2% | Specification | FIXED: Incorrect number of input data products |
| 7 | 1.1% | BUG | Known and solved bug in pulp/pipeline integration |
| 6 | 1.0% | BUG | Network reset while communicating with node. If problem persist for the same node suspect faulty hardware. else restart |
| 6 | 1.0% | BUG | Executable has problem finding librairy. suspect deployment error |
| 5 | 0.8% | BUG | Casacore problem. warn pipeline developer or Ger van Diepen |
| 4 | 0.6% | BUG | Known and solved bug in pulp/pipeline integration |
| 4 | 0.6% | Other | OTHER: Failing pipelines with unique singular cause of failure |
| 3 | 0.5% | Specification | PARSET: 0 Output\_SkyImage data products specified |
| 3 | 0.5% | BUG | Incorrect executable configuration. warn pipeline developer |
| 2 | 0.3% | Other | An input data set is empty. check predecessor |
| 2 | 0.3% | BUG | Internal pipeline problem warn pipeline developer |
| 1 | 0.2% | BUG | Suspected connectivity problems on locus nodes. test ssh connection to affected nodes |
| 1 | 0.2% | Other | The pipeline was stopped due to reboot of system (stopday?) |
| 1 | 0.2% | BUG | Major error in on of the executables. Check input data. else warn pipeline dev. |
| 1 | 0.2% | Specification | NDPPP specification error |
| 1 | 0.2% | Specification | Error in BBS. check inputdata and specification |
| 1 | 0.2% | Specification | suspect incorrect specification of input or output products |
| 619 | 100.0% |  |  |

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

|  |  |  |
| --- | --- | --- |
| Error type | Occurances | As percentage of errors |
| Data | 226 | 36.5% |
| Unknown (Pulsar) | 172 | 27.8% |
| Specification | 97 | 15.7% |
| BUG | 77 | 12.4% |
| Other | 47 | 7.6% |

Conclusions  
A summary of the types of errors can be found in table 2. The pipelines and called executables had a failure rate of 1.8% compared to the previous 2.9%. The failure rate due to bugs in the pipeline software itself was 0.22%. With the majority of failures due to missing input data and/or incorrect specification or in the unsupported Pulsar pipeline

## 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.