

RQ1: Can we quantify interest of TD at the functional level? How much is the interest? (Version 5)

Yasutaka Kamei

Feb 13th, 2016

What's new in the version 5?

- Add histograms

What's new in the version 4?

- We used the new dataset that solves one bug that is fixed by Everton.

Data Load

```
setwd("/Users/kamei/Research/techdebt/msr16_td_interest/")
source("./r_scripts/data_read.r")
```

```
## Num Of Raw Data : 3285
## Num Of Tech Debt : 2269
##   Per of linking to v1 : 85.54% (=1941/2269)
##   Per of linking to v2 : 84.44% (=1916/2269)
##   Per of linking to all: 83.91% (=1904/2269)
```

Data Summary

- (Step 1) choose one of duplicated method and version name
- (Step 2) only use technical debt including metrics

```
nrow(data)
```

```
## [1] 3285
```

```
apply(data.s1[,c("version_name", "CountInput_v1", "CountInput_v2")], 2, function(x){sum(x == -1) })
```

```
## version_name CountInput_v1 CountInput_v2
##           2269           328           353
```

Observation

- The number of methods that cannot be linked between Everton's data and metrics data
 - 171 => 221 (introducing)
 - 101 => 94 (being found as last version)

(Step 3) use technical debt including non 0 for division

CountLine

```
summary(data.CountLine.all$Project)
```

```
##      ant  hadoop  jmeter  log4j  tomcat  
##    354    370    460     65    655
```

```
summary(data.CountLine.positive$Project)
```

```
##      ant  hadoop  jmeter  log4j  tomcat  
##    110    105    119     7    135
```

```
summary(data.CountLine.negative$Project)
```

```
##      ant  hadoop  jmeter  log4j  tomcat  
##     45     41     91     6     79
```

```
summary(data.CountLine.positive$Project) / summary(data.CountLine.all$Project) * 100
```

```
##      ant  hadoop  jmeter  log4j  tomcat  
## 31.07345 28.37838 25.86957 10.76923 20.61069
```

```
summary(data.CountLine.negative$Project) / summary(data.CountLine.all$Project) * 100
```

```
##      ant  hadoop  jmeter  log4j  tomcat  
## 12.711864 11.081081 19.782609 9.230769 12.061069
```

CountInput

```
summary(data.CountInput.all$Project)
```

```
##      ant  hadoop  jmeter  log4j  tomcat  
##    332    360    425     57    623
```

```
summary(data.CountInput.positive$Project)
```

```
##      ant  hadoop  jmeter  log4j  tomcat  
##     78    101    124     11    134
```

```
summary(data.CountInput.negative$Project)
```

```
##      ant  hadoop  jmeter  log4j  tomcat  
##     36     45     41     5     61
```

```
summary(data.CountInput.positive$Project) / summary(data.CountInput.all$Project) * 100
```

```
##      ant  hadoop  jmeter  log4j  tomcat  
## 23.49398 28.05556 29.17647 19.29825 21.50883
```

```
summary(data.CountInput.negative$Project) / summary(data.CountInput.all$Project) * 100
```

```
##      ant  hadoop  jmeter  log4j  tomcat  
## 10.843373 12.500000 9.647059 8.771930 9.791332
```

Observation

- The number of all methods is 837
 - (s1) 754
 - (s2) 488
- We use 71, 181, and 236 methods including technical debt.
 - The data set we used had 67 (ant), 169(jmeter) and 268(jruby) technical debt.
- 32.6%-44.2% of technical debt has positive interest.
- 13.8%-28.7% of technical debt has negative interest.

CountLine

```
# interest of CountLine (LOC)
fc <- factor(data.CountLine.all$Project)
interest <- data.CountLine.all$interest
tapply(interest, fc, summary)
```

```
## $ant
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -87.260   0.000   0.000   6.605   4.922 157.400
##
## $hadoop
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -95.000   0.000   0.000   8.824   2.801 283.300
##
## $jmeter
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -83.660   0.000   0.000   4.490   1.786 200.000
##
## $log4j
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -45.000   0.000   0.000   2.491   0.000  63.410
##
## $tomcat
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -95.650   0.000   0.000   5.143   0.000 433.300
```

```
fc <- factor(data.CountLine.positive$Project)
interest <- data.CountLine.positive$interest
tapply(interest, fc, summary)
```

```
## $ant
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.6289   5.9120 16.4600 28.2400 35.8300 157.4000
##
## $hadoop
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.3289   8.6960 25.0000 41.9800 57.1400 283.3000
##
## $jmeter
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.4785   6.6670 18.1800 34.3100 49.4700 200.0000
##
```

```
## $log4j
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   4.651  29.450  40.740  39.010  52.690  63.410
##
## $tomcat
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.4049   5.3000  13.6400  36.3300  38.6600 433.3000
```

```
idx <- order(interest, decreasing = T)
write.csv(file="a.csv", data.CountLine.positive[idx,])
data.CountLine.positive$interest[idx]
```

```
##   [1] 433.3333333 333.3333333 283.3333333 277.7777778 211.3207547
##   [6] 200.0000000 200.0000000 190.9090909 169.5652174 166.6666667
##  [11] 162.5000000 157.4468085 157.1428571 154.1666667 150.0000000
##  [16] 148.1481481 144.5945946 142.1052632 140.0000000 137.1134021
##  [21] 135.3658537 133.3333333 132.0754717 125.0000000 119.5121951
##  [26] 118.5567010 116.6666667 115.7894737 114.7058824 114.2857143
##  [31] 114.2857143 112.9032258 109.6153846 106.2937063 105.0000000
##  [36] 103.3057851 100.0000000 100.0000000 100.0000000 100.0000000
##  [41] 100.0000000 100.0000000  97.5000000  95.1219512  94.5945946
##  [46]  94.2857143  90.6666667  90.4109589  88.8888889  85.7142857
##  [51]  84.3750000  82.7586207  82.6086957  81.8181818  81.8181818
##  [56]  80.8219178  80.0000000  79.3650794  78.4313725  77.6859504
##  [61]  76.9230769  75.4601227  72.4137931  70.8333333  70.0000000
##  [66]  69.6969697  69.6428571  69.4656489  68.4210526  68.0000000
##  [71]  67.7419355  66.6666667  66.6666667  66.6666667  65.2173913
##  [76]  64.6153846  64.2857143  63.6363636  63.6363636  63.4831461
##  [81]  63.4146341  63.2911392  63.1578947  63.1578947  61.5942029
##  [86]  60.9756098  60.4651163  60.0000000  59.2592593  59.1836735
##  [91]  58.3333333  58.1395349  57.1428571  56.6371681  56.4102564
##  [96]  54.5454545  54.0540541  53.8461538  53.4482759  53.3333333
## [101]  52.6315789  52.5423729  52.3809524  52.3809524  52.2388060
## [106]  51.7241379  51.2820513  51.2195122  50.0000000  50.0000000
## [111]  50.0000000  48.9361702  48.8372093  47.6190476  47.4358974
## [116]  47.0588235  45.8333333  45.2380952  44.0000000  44.0000000
## [121]  43.9393939  43.4782609  43.4782609  42.8571429  42.8571429
## [126]  42.8571429  42.8571429  42.8571429  42.6829268  42.3076923
## [131]  42.2222222  42.1052632  41.6666667  41.3793103  41.2698413
## [136]  41.0714286  40.9638554  40.7407407  40.6779661  40.5405405
## [141]  40.0000000  38.8888889  38.4615385  38.4615385  38.2022472
## [146]  37.5000000  37.5000000  36.6666667  36.3636364  36.1111111
## [151]  36.1111111  35.8974359  35.5555556  35.0877193  35.0000000
## [156]  34.5454545  34.3283582  33.9285714  33.3333333  33.3333333
## [161]  33.3333333  33.3333333  33.3333333  32.3943662  32.0512821
## [166]  31.9148936  31.5789474  31.0344828  31.0344828  30.8823529
## [171]  30.7692308  30.4347826  30.0000000  30.0000000  29.4117647
## [176]  29.1666667  28.9855072  28.8135593  28.5714286  28.3018868
## [181]  28.1250000  27.5862069  27.2727273  27.2727273  27.2727273
## [186]  27.0833333  27.0000000  26.0869565  25.7575758  25.5813953
## [191]  25.4237288  25.4237288  25.4237288  25.0000000  25.0000000
## [196]  25.0000000  25.0000000  25.0000000  25.0000000  24.3589744
## [201]  24.2105263  24.0506329  23.8095238  23.2876712  23.2558140
## [206]  22.4358974  22.2797927  22.2222222  22.2222222  22.2222222
## [211]  22.2222222  21.9512195  21.6666667  21.5686275  21.2121212
```

## [216]	20.8333333	20.7317073	20.4918033	20.0000000	20.0000000
## [221]	20.0000000	20.0000000	20.0000000	20.0000000	19.5121951
## [226]	19.5121951	19.4444444	19.2771084	19.1780822	19.0476190
## [231]	18.9189189	18.6915888	18.5714286	18.5185185	18.3006536
## [236]	18.2608696	18.1818182	18.1818182	18.1818182	18.1818182
## [241]	18.0000000	17.3913043	17.3913043	17.2413793	16.8421053
## [246]	16.6666667	16.6666667	16.6666667	16.6666667	16.6666667
## [251]	16.3636364	16.2500000	15.8730159	15.8730159	15.8333333
## [256]	15.7894737	15.7142857	15.3846154	15.3846154	15.3846154
## [261]	15.3846154	15.3846154	15.2173913	15.1515152	15.1515152
## [266]	15.0000000	14.8148148	14.2857143	14.2857143	14.2857143
## [271]	14.0625000	13.8888889	13.8686131	13.6986301	13.6363636
## [276]	13.6363636	13.3333333	13.3333333	13.3333333	13.1578947
## [281]	12.7906977	12.6984127	12.5000000	12.5000000	12.5000000
## [286]	12.5000000	12.5000000	12.5000000	12.3893805	12.2448980
## [291]	12.2448980	12.1951220	12.0000000	11.9565217	11.8881119
## [296]	11.8181818	11.7647059	11.6279070	11.5384615	11.4285714
## [301]	11.1111111	11.1111111	11.1111111	10.9756098	10.9589041
## [306]	10.8433735	10.3448276	10.1694915	10.1010101	10.0000000
## [311]	9.8591549	9.6385542	9.6296296	9.5652174	9.3023256
## [316]	9.0909091	9.0909091	9.0909091	9.0909091	9.0909091
## [321]	9.0909091	8.9108911	8.6956522	8.4745763	8.4337349
## [326]	8.3333333	8.3333333	8.3333333	8.3333333	8.1818182
## [331]	8.1632653	8.1632653	8.1481481	8.0357143	8.0000000
## [336]	7.6923077	7.6923077	7.6923077	7.6923077	7.6923077
## [341]	7.5000000	7.4074074	7.4074074	7.0175439	6.8376068
## [346]	6.6666667	6.6666667	6.6666667	6.5217391	6.4516129
## [351]	6.4516129	6.4516129	6.3829787	6.2500000	6.2500000
## [356]	6.2500000	6.2500000	6.2500000	6.1797753	6.1728395
## [361]	6.0975610	6.0606061	6.0000000	6.0000000	5.9210526
## [366]	5.8823529	5.8823529	5.8823529	5.8823529	5.7803468
## [371]	5.6338028	5.6338028	5.5555556	5.4054054	5.4054054
## [376]	5.2631579	5.2631579	5.2631579	5.1948052	5.1546392
## [381]	5.1282051	5.0000000	4.9549550	4.9382716	4.9019608
## [386]	4.8437500	4.7619048	4.7619048	4.7619048	4.7619048
## [391]	4.6875000	4.6875000	4.6511628	4.6153846	4.5977011
## [396]	4.5801527	4.5454545	4.5454545	4.5454545	4.3956044
## [401]	4.3478261	4.1666667	4.1666667	4.0540541	4.0000000
## [406]	4.0000000	3.9215686	3.8961039	3.8461538	3.8461538
## [411]	3.8461538	3.7878788	3.7735849	3.7383178	3.7037037
## [416]	3.6363636	3.6363636	3.4482759	3.4482759	3.4482759
## [421]	3.4482759	3.2967033	3.0612245	3.0303030	3.0303030
## [426]	3.0303030	2.9629630	2.9411765	2.8571429	2.8571429
## [431]	2.7777778	2.7027027	2.7027027	2.7027027	2.7027027
## [436]	2.7027027	2.6666667	2.6315789	2.6315789	2.6315789
## [441]	2.5641026	2.5641026	2.4096386	2.2598870	2.1739130
## [446]	2.0979021	1.9607843	1.9047619	1.8181818	1.7857143
## [451]	1.7857143	1.7543860	1.6949153	1.6666667	1.6393443
## [456]	1.6129032	1.5873016	1.5384615	1.5384615	1.5151515
## [461]	1.5037594	1.4388489	1.3698630	1.3422819	1.2820513
## [466]	1.2658228	1.1363636	1.0989011	0.9345794	0.8620690
## [471]	0.7751938	0.7042254	0.6289308	0.4784689	0.4048583
## [476]	0.3289474				

```
data.CountLine.positive$Remove_Comment[idx]
```

```
## [1] Implement merge rules for a few more web.xml elements Required some refactoring as some of the
## [2] Drop tomcat-lite module. It never saw a release and has not seen meaningful development for over
## [3]
## [4]
## [5]
## [6] Remove all InterruptTimer changes Using a branch instead
## [7]
## [8] Use JOrphanUtils.split()
## [9] HDFS-3726. If a logger misses an RPC, don't retry that logger until next segment. Contributed by
## [10] s/can not/cannot/
## [11] Extend support for the WebSocket permessage-deflate extension to compression of outgoing messages
## [12] sconolly advises using TODO rather than XXX in comments.
## [13] Remove the 'integration' - ObjectManager and wrapping of DI frameworks. Lite is just a library.
## [14]
## [15] Make TODO comments consistent to make them easier to find
## [16] Drop tomcat-lite module. It never saw a release and has not seen meaningful development for over
## [17]
## [18] Improved InterruptedException handling, especially from Parallel task. Should also solve Pr 4292
## [19] Remove some TODOs that have been implemented
## [20]
## [21] removal of testlet, visual age for java, and icontract optional tasks.
## [22] HDFS-2899. Service protocol changes in DatanodeProtocol to add multiple storages. Contributed by
## [23]
## [24] HDFS-4095. Add some snapshot related metrics. Contributed by Jing Zhao.
## [25]
## [26] Subtargets is a little confusing - change to "Other Targets"
## [27]
## [28]
## [29] CSV Dataset does not handle embedded new lines in quoted data Bugzilla Id: 53807
## [30] Fix https://issues.apache.org/bugzilla/show\_bug.cgi?id=49884
## [31] Another TODO. Inputs are never null.
## [32]
## [33]
## [34]
## [35] - This error shouldn't happen, and it's out of the servlet, so it is not going to be possible to
## [36]
## [37] Bug 27780 (patch 19792) - update POST handling Also updated PostWriterTest to include additional
## [38] Bug 42173 - Let HTTP Proxy handle encoding of request, and decode parameter values
## [39] Code cleanup. No functional change
## [40]
## [41]
## [42] Moved the connection pool to a top level class and started to add more code. Still missing is c
## [43] Fix https://issues.apache.org/bugzilla/show\_bug.cgi?id=56530 Add a web application class loader
## [44] Changes to DefaultConfigurer:
## [45]
## [46] Considered. It is not a good idea to do this generally.
## [47] Use HTTPConstants instead of HTTPSamplerBase
## [48]
## [49]
## [50] Switch to using the Http2Parser for frame processing - remove original frame parsing code from
## [51]
## [52]
```

```

## [53]
## [54]
## [55] WebSocket 1.0 implementation part 11 of many Fix s/ping/pong/ in the API. pings are automatica
## [56]
## [57]
## [58] Update NON-NLS markers
## [59]
## [60] MAPREDUCE-5062. Fix MR AM to read max-retries from the RM. Contributed by *Zhijie Shen.
## [61] Clarify TODOs
## [62] Re-write comment to remove XXX marker
## [63]
## [64] Added option to turn on notification of emacs on a build error so that emacs goes to the file a
## [65]
## [66] Everything related with properties is in PropertyHelper. ProjectHelper delegates for backward c
## [67] YARN-230. RM Restart phase 1 - includes support for saving/restarting all applications on an RM
## [68]
## [69] HDFS-2367. Enable the configuration of multiple HA cluster addresses. Contributed by Aaron T. M
## [70] HDFS-3024. Improve performance of stringification in addStoredBlock. Contributed by Todd Lipcon
## [71]
## [72] Added ANTLRsetGlib(File) Marked ANTLRsetGlib(String) as deprecated, and make it call the new s
## [73] Bug 58980 - JMS Subscriber will return successful as long as 1 message is received. resolve 11
## [74]
## [75]
## [76]
## [77]
## [78]
## [79]
## [80] Remove unnecessary code. Request can not be async at this point.
## [81] And with a click of her heals, Chainsaw v2 heads off home to Kansas.
## [82] Jikes supports -bootclasspath PR: 32609
## [83]
## [84] * Added Drag & Drop (rudimentary) support to Chainsaw. You can drag any File to Chainsaw's
## [85]
## [86]
## [87]
## [88] HDFS-3950. QJM: misc TODO cleanup, improved log messages, etc. Contributed by Todd Lipcon.
## [89] 2.0 merge into main and TestBean fixes
## [90] Make Definer use the new ClasspathUtils, make ClasspathUtils even more reuse-friendly.
## [91] Move code that parses EL expressions within JSP template text from Parser to JspReader class.
## [92] Fix up RE so it picks up versions containing -bnn and -BETAn
## [93]
## [94]
## [95] move embed proposal
## [96] Made the MasterConverter search through the destination classes hierarchy to find a match for c
## [97] Bug 7552 audit of <copy> and <move> - really make sure state is reset, even if exceptions have
## [98] Get this compiling again.
## [99]
## [100]
## [101] Don't save responseData unless it is text
## [102]
## [103] removal of the jprobe and metamata tasks.
## [104] Drop tomcat-lite module. It never saw a release and has not seen meaningful development for ovr
## [105] HDFS-3950. QJM: misc TODO cleanup, improved log messages, etc. Contributed by Todd Lipcon.
## [106]

```

```

## [107]
## [108] Refactor configuration out of Frame into Component Manager
## [109]
## [110]
## [111]
## [112] Clarify issue with BeanShell compilation
## [113] https://issues.apache.org/bugzilla/show\_bug.cgi?id=48137 Fix javadoc typos Patch provided by q
## [114] MAPREDUCE-2793. Corrected AppIDs, JobIDs, TaskAttemptIDs to be of correct format on the web pa
## [115] Test improvements: convert to JUnit4, improved error testing and remove sleeps
## [116] Split HTTP and HTTPClient tests to make it easier to see which fail
## [117] Remove a JDK 1.3 work-round
## [118] Remove a TODO
## [119]
## [120] Drop tomcat-lite module. It never saw a release and has not seen meaningful development for ov
## [121] move embed proposal
## [122]
## [123]
## [124]
## [125]
## [126] Add stacktrace in exception
## [127]
## [128]
## [129] Fix a concurrency issue in the unit test (streams 19 and 21 may not both be blocked by the tim
## [130] Remove some unit tests that are part of container.
## [131]
## [132]
## [133]
## [134] HDFS-3900. QJM: avoid validating log segments on log rolls. Contributed by Todd Lipcon.
## [135] Trial sampler using Commons HTTPClient
## [136]
## [137]
## [138]
## [139]
## [140] Reformatted to JMeter conventions. Minor JavaDoc updates (mostly formatting). Removed an unuse
## [141] Zap aut from main myrmidon tree
## [142] Bug 56539 - Mail reader sampler: When Number of messages to retrieve is superior to 1, Number o
## [143] Remove recorder as should be reimplemented as a configured aspect aka a Facility
## [144] Added option to turn on notification of emacs on a build error so that emacs goes to the file a
## [145] Tidy up Request Data output; fix NPE in HTTPSampler2
## [146]
## [147]
## [148]
## [149] Remove some unused code
## [150] Implement jar index for referenced jars, PR: 14255
## [151]
## [152]
## [153] Add comment Remove now useless code as httpclient 4.2.1 has been integrated which contains fix
## [154] Drop tomcat-lite module. It never saw a release and has not seen meaningful development for ov
## [155]
## [156]
## [157] - New tasks StarTeamCheckin and StarTeamList - Modify StarTeamCheckin, Add ability to check fi
## [158] Antidote officially moved to the CVS module jakarta-ant-antidote.
## [159] add rule that"$"->"$"; retaining "$$"->"$"
## [160] Merge from 2-0 and some fixes to controllers that need to recover running version after no samp

```



```

## [161] HDFS-4144. Create test for all snapshot-related metrics. Contributed by Jing Zhao.
## [162]
## [163] Fix NPE in tests
## [164] HADOOP-8334. HttpServer sometimes returns incorrect port (Daryn Sharp via bobby)
## [165] Fix https://issues.apache.org/bugzilla/show_bug.cgi?id=49884
## [166]
## [167] HDFS-4103. Support 0(1) snapshot creation.
## [168] Zap the yet to be converted ant1 code
## [169] MAPREDUCE-3104. Implemented Application-acls. (vinodkv)
## [170] BeanShell Assertion checker
## [171]
## [172] MAPREDUCE-3942. Randomize master key generation for ApplicationTokenSecretManager and roll it
## [173] PR 56748 Spelling fixes, submitted by Ville Skyttä
## [174]
## [175] Fix https://issues.apache.org/bugzilla/show_bug.cgi?id=56521 Implement a TODO Reuse the async v
## [176]
## [177]
## [178] Stop interrupt() from attempting disconnect whilst we are already disconnecting
## [179] Resolve properties loaded from files before sending them to the project helper for resolution.
## [180]
## [181] New sandbox home for log4jMini
## [182] HDFS-2920. fix remaining TODO items. Contributed by Aaron T. Myers and Todd Lipcon.
## [183]
## [184] Zap framework classes
## [185] Remove the remaining SPDY code and associated references
## [186] Add comment Remove now useless code as httpclient 4.2.1 has been integrated which contains fix
## [187]
## [188]
## [189] Modernize Jikes (unconditionally) - PR 25868 and 26404
## [190] MAPREDUCE-4211. Error conditions (missing appid, appid not found) are masked in the RM app page
## [191] - Added rulespath to Audit to use customized rules. - Fix Audit messages to remove fullpath in
## [192] Remove unused class
## [193] Fix https://issues.apache.org/bugzilla/show_bug.cgi?id=48133 Spelling errors in javadoc Patch p
## [194] merge Zip64 support from Commons Compress
## [195] Zap the yet to be converted ant1 code
## [196] Merge from 2-0 and some fixes to controllers that need to recover running version after no sampl
## [197]
## [198] Revert HDFS-2922 via svn merge -c -1242572
## [199]
## [200] Regular Expression Extractor now deletes all stale variables from previous matches.
## [201] <javadoc> added some options to the not-cloned Commandline potentially adding them several tim
## [202] Proxy spoofing now drops the default HTTPS port (443) when converting https: links to http: Up
## [203] Make Definer use the new ClasspathUtils, make ClasspathUtils even more reuse-friendly.
## [204] Use multicatch to reduce code duplication.
## [205] MAPREDUCE-3849. Change TokenCache's reading of the binary token file (Daryn Sharp via bobby)
## [206]
## [207] Refactoring. Switch from a boolean to an Enum for error state so we can differentiate between
## [208] Rewritten part of the regexp stuff, adding substitution and options. New optional replaceregexp
## [209] Makes more sense to use IOError for file format problems; include last char in logging output.
## [210] Extend Lifecycle/MBean refactoring to ContainerBase
## [211] Further work for https://issues.apache.org/bugzilla/show_bug.cgi?id=56029 Review from kkolinko
## [212] Implemented
## [213]
## [214] Remove some unused code and simplify the test.

```

```

## [215] Add "negate" attribute to <propertyset>s.
## [216] Remove the remaining SPDY code and associated references
## [217]
## [218] Using ClientJMeterEngine.tidyRMI allows server to exit without needing to use System.exit, so
## [219] Zap aut from main myrmidon tree
## [220] Zap the anteater sourcecode after harvesting path todos.
## [221] HDFS-3637. Add support for encrypting the DataTransferProtocol. Contributed by Aaron T. Myers.
## [222] YARN-635. Renamed YarnRemoteException to YarnException. Contributed by Siddharth Seth. MAPREDUC
## [223]
## [224] https://issues.apache.org/bugzilla/show_bug.cgi?id=48137 Fix javadoc typos Patch provided by q
## [225]
## [226] Fix https://issues.apache.org/bugzilla/show_bug.cgi?id=48133 Spelling errors in javadoc Patch p
## [227] Too tightly coupled to ant1 to run in ant2 so zap them for now and come back to it later.
## [228] Do the scanning to count nesting levels and presence of formal parameters just once. Simplify
## [229] Fix https://issues.apache.org/bugzilla/show_bug.cgi?id=27122 Remove a workaround for a very old
## [230]
## [231]
## [232]
## [233] Add cookies field and populate it
## [234] Bug 57193: Add param and return tags to javadoc Bugzilla Id: 57193
## [235]
## [236] Add async-supported impl for Servlets
## [237] move embed proposal
## [238] * Rolled back change to PropertyResolver.resolveProperties(), so that it uses a TaskContext a
## [239] Use Integer.toString instead of concat Cascade exceptions Remove catch on NumberFormatException
## [240] Fix remainder of failing tests and a related TODO spotted in the tests as well.
## [241] Proxy needs to return the headers as well
## [242] No need to call prepare during shutdown
## [243]
## [244]
## [245] Bugzilla report 28865: allow files in classpath for <jdepend>
## [246] Get this compiling again.
## [247] Zap aut from main myrmidon tree
## [248]
## [249]
## [250]
## [251]
## [252] removing starteam code
## [253] Purge of many (but not all) the if-java-1.1 tests in the ant code. I left those in the tests in
## [254]
## [255] Update names of properties.
## [256] add method that compares file contents to FileUtils, use it in FixCrLf.
## [257] Fix a few checkstyle errors
## [258] Changes to <propertyfile>
## [259]
## [260]
## [261]
## [262] Remove ServerLifecycleListener
## [263]
## [264]
## [265]
## [266]
## [267]
## [268]

```

```

## [269]
## [270] Drop tomcat-lite module. It never saw a release and has not seen meaningful development for over
## [271] Cookies are handled; unused import
## [272]
## [273] Bug 47474 - View Results Tree support for plugin renderers
## [274]
## [275] Clean-up
## [276] Drop tomcat-lite module. It never saw a release and has not seen meaningful development for over
## [277]
## [278]
## [279]
## [280]
## [281] Regular Expression Extractor now deletes all stale variables from previous matches.
## [282] Allow typelibs to define custom roles:
## [283] Add Sampler scope selection to XPathExtractor
## [284] Bug 53522 - Cache Manager should not store at all response with header "no-cache" and store other
## [285] New JDBC Connection Pool resource. Revamping JDBC Sampler to use TestBean framework and the new
## [286]
## [287] MAPREDUCE-2896. Simplify all apis to in org.apache.hadoop.yarn.api.records.* to be get/set only.
## [288] Fix https://issues.apache.org/bugzilla/show_bug.cgi?id=49567 Better handling of calls to complete
## [289] Remove the org.apache.jk package as per discussion on dev list. Do this early in Tomcat 7 development
## [290] * Add bean customizer support. * Factor out the bean GUI to a GenericTestBeanCustomizer.
## [291] Add init() and destroy() to the Lifecycle interface Note that this commit breaks the shutdown logic
## [292] Fix https://issues.apache.org/bugzilla/show_bug.cgi?id=45451 Testing for this threw up all sorts of
## [293]
## [294] Fix https://issues.apache.org/bugzilla/show_bug.cgi?id=49884
## [295] Fix PUT proxy handling
## [296] Refactor methods that call read/write listeners out of AsyncContext as CoyoteAdapter has all the
## [297] s/can not/cannot/
## [298]
## [299]
## [300]
## [301] MAPREDUCE-4059. The history server should have a separate pluggable storage/query interface. (JIRA)
## [302]
## [303]
## [304] Fix thread-unsafe local truststore setup Ensure StartTLS can be used with TrustAll and UseLocalCertificate
## [305] Bug 52997 - Jmeter should not exit without saving Test Plan if saving before exit fails
## [306] Use latency for time to establish the connection
## [307] * Added OsFamily enumerated type, to represents OS groupings. * Replaced all usages of Os.isFamily
## [308]
## [309] <javac> rewritten to use a factory.
## [310] Fixed language change handling for menus (does not yet work for TestBeans)
## [311] Bug 45904 - Allow 'Not' Response Assertion to succeed with null sample
## [312] Fix thread-unsafe local truststore setup Ensure StartTLS can be used with TrustAll and UseLocalCertificate
## [313]
## [314] When using Servlet 3.1 non-blocking reads with an AJP connector, make the reads between the AJP connector
## [315] Bugzilla 28773 * added support for nested filesets into DotnetResource * fixed problem when connecting
## [316] revert Bugzilla: 35436, getResources is final in java 1.4
## [317] removal of the jprobe and metamata tasks.
## [318] Add runtime debug switching
## [319]
## [320] Added NIO connector to server.xml parser warnings
## [321] Simplify since minimum Java version is now >= 1.5
## [322]

```

```

## [323] HDFS-2784. Update hftp and hdfs for host-based token support. Contributed by Kihwal Lee.
## [324] Adding default "parse all embedded resources" into HTTP Defaults Config and did some refactori
## [325] Zap the anteater sourcecode after harvesting path todos.
## [326] Remove some unit tests that are part of container.
## [327] resolve 46
## [328] Merging trunk to HDFS-1623 branch
## [329]
## [330]
## [331] A change in ProjectHelper2 - RC doesn't need to depend on SAX2. The attributes are just stored
## [332] Axed a bucketload of classes.
## [333]
## [334]
## [335]
## [336] Get this compiling again.
## [337] * DefaultTaskContext now uses a ServiceManager to locate services, rather than a ComponentMan
## [338] VFS Tidy-ups:
## [339]
## [340]
## [341]
## [342]
## [343] Add support for WebFilter Remove wrappers to implement isAsyncSupported() having found the set
## [344]
## [345] Merging trunk to HDFS-1623 branch
## [346]
## [347]
## [348] Drop tomcat-lite module. It never saw a release and has not seen meaningful development for ov
## [349] Use locale independent methods to identify OS or match attributes and elements in build files.
## [350] Correct menu popup for Listeners
## [351]
## [352] Fix https://issues.apache.org/bugzilla/show_bug.cgi?id=45977 Trivial comment clean up. No plan
## [353]
## [354] Bug 38687 - Module controller does not work in non-GUI mode
## [355]
## [356]
## [357]
## [358] More cookie refactoring - new support class for common elements of parsing and writing - bet
## [359] Pull up the process() method
## [360]
## [361]
## [362] Next round of typo fixes by Larry Shatzer
## [363] Allow to specify a given factory implementation for the transformer (xsltc, xalan, saxon...) a
## [364]
## [365]
## [366] Bring condition into line with Ant2 thinking - make it check the value of condition - not just
## [367] Started to refactor XSLTProcess to be loose the xsltLiason. We can assume that most future pro
## [368]
## [369]
## [370] Merge new resources implementation from sandbox/trunk-resources
## [371] Move over to new tar package
## [372] Antidote officially moved to the CVS module jakarta-ant-antidote.
## [373]
## [374] StreamError needs the stream ID (so we know which stream to close)
## [375]
## [376] Zap some antlibs

```

```

## [377] Integrate type so that tasks/datatypes are access via TypeManager rather than through individual
## [378] Warn user when a property value he entered is not accepted.
## [379] Drop tomcat-lite module. It never saw a release and has not seen meaningful development for over
## [380] Fix http://issues.apache.org/bugzilla/show\_bug.cgi?id=56027 Alternative implementation for the
## [381] Use setResponseData(String, null) where default charset is OK
## [382]
## [383] Complete some error handling TODOs
## [384] MAPREDUCE-2896. Simplify all apis to in org.apache.hadoop.yarn.api.records.* to be get/set only.
## [385]
## [386]
## [387] Bug 58987 - Report/Dashboard: Improve error reporting Throw exception if mismatch between expected
## [388]
## [389]
## [390]
## [391] Zap some antlibs
## [392] Drop the non-JSR356 WebSocket implementation
## [393] Bug 37930: Maven build for log4j 1.3
## [394] More fun and games with nested lambda expressions. Each invocation of a nested expression consists
## [395] sconnolly advises using TODO rather than XXX in comments.
## [396]
## [397] remove input proposal step 1
## [398] No longer need propertyChange handling; also when removing this realised that only one property
## [399] Cookie Manager no longer stores cookies by default
## [400] MAPREDUCE-4266. remove Ant remnants from MR (tgraves via bobby)
## [401] Remove Antcall until it can be determined if we want or need it. (and if we do want/need it - v
## [402]
## [403]
## [404] Add TODO Charset markers
## [405]
## [406] Try removing the beta tag from the NIO2 connector to get more feedback, after the latest round
## [407] Complete some error handling TODOs
## [408]
## [409] Performance: if a task does not have an id, do not bother to replace the unknown element in
## [410] HADOOP-7320. Refactor the copy and move commands to conform to new FsCommand class. Contributed
## [411] Partial fix for https://issues.apache.org/bugzilla/show\_bug.cgi?id=48644 Some Throwables must
## [412]
## [413] Ensure if and for statements always use blocks Other minor tidyups
## [414] Clean-up. No functional change.
## [415] Zap some antlibs
## [416] * Rolled back change to PropertyResolver.resolveProperties(), so that it uses a TaskContext
## [417] Remove the invoker servlet and all references for security reasons. This will not be proposed
## [418] Get this compiling again.
## [419]
## [420]
## [421] Drop tomcat-lite module. It never saw a release and has not seen meaningful development for over
## [422]
## [423] HADOOP-6760. WebServer shouldn't increase port number in case of negative port setting caused by
## [424] Revert logic that performs the <os family="win9x"/> check from "I know all that are not win9x"
## [425]
## [426]
## [427] Javadoc fixes.
## [428] Zap the yet to be converted ant1 code
## [429]
## [430]

```

```

## [431] sconolly advises using TODO rather than XXX in comments.
## [432] Move to english locale when lower casing things, to avoid problems in other locales, like, say
## [433]
## [434] Zap the bootstrap stuff
## [435]
## [436] Remove the remaining SPDY code and associated references
## [437] PR 56748 Spelling fixes, submitted by Ville Skyttä
## [438] No need to create the JLabel() any more
## [439]
## [440] Fix an HTTP/2 TODO for processing URIs. Add support for query strings
## [441] resolve 43 Thanks
## [442] Allow Checkstyle's unused imports test to be used with the o.a.catalina.core package (partial)
## [443]
## [444] MAPREDUCE-4266. remove Ant remnants from MR (tgraves via bobby)
## [445]
## [446]
## [447] Resolve a FIXME and expand unit tests to cover CRLF vs LF checking.
## [448]
## [449] Merge BCEL changes from r1614166 to r1682271 The changes consist of: - code clean-up - Javadoc
## [450]
## [451]
## [452] Remove the ProjectHelper proposal.
## [453] Drop bayuex module. It never saw a release and has not seen meaningful development for over 3
## [454] HDFS-4098. Add FileWithSnapshot, INodeFileUnderConstructionWithSnapshot and INodeFileUnderConst
## [455]
## [456]
## [457]
## [458] remove .NET tasks
## [459] Move the frame header processing into the WsInputStream since the fragmented frame processing v
## [460] HDFS-3044. fsck move should be non-destructive by default. Contributed by Colin Patrick McCabe
## [461] Fix https://issues.apache.org/bugzilla/show_bug.cgi?id=46264 Implement threaded start/stop for
## [462]
## [463]
## [464] Fix https://issues.apache.org/bugzilla/show_bug.cgi?id=46264 Implement threaded start/stop for
## [465] HDFS-3170. Add more useful metrics for write latency. Contributed by Matthew Jacobs.
## [466] MAPREDUCE-4159. Job is running in Uber mode after setting "mapreduce.job.ubertask.maxreduces"
## [467]
## [468] FileDialoger can return null
## [469] HDFS-2868. Expose xceiver counts via the DataNode MXBean. (harsh)
## [470] Remove the remaining SPDY code and associated references
## [471] Remove unused code
## [472]
## [473] Removed Project.getVersion() replaced it with getJavaVersion() (inherited from AbstractTask) w
## [474]
## [475] Remove temporary code. Now that at least one required field is configured via SSLHostConfig th
## [476] HADOOP-11698. Remove DistCpV1 and Logalyzer. Contributed by Brahma Reddy Battula.
## 1062 Levels: ...

```

```

fc <- factor(data.CountLine.negative$Project)
interest <- data.CountLine.negative$interest
tapply(interest, fc, summary)

```

```

## $ant
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -87.260 -21.670  -9.091 -17.060  -5.000  -1.000

```

```
##
## $hadoop
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -95.0000 -45.3500 -16.9000 -27.8700 -5.4050  -0.3984
##
## $jmeter
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -83.6600 -29.2900 -18.7500 -22.1700 -7.6920  -0.6579
##
## $log4j
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -45.000 -26.650 -15.290 -18.530  -6.678  -1.356
##
## $tomcat
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -95.6500 -24.7800 -13.1000 -19.4300 -5.2630  -0.4149
```

CountInput

```
# CountInput
fc <- factor(data.CountInput.all$Project)
interest <- data.CountInput.all$interest
tapply(interest, fc, summary)

## $ant
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -100.00    0.00    0.00   10.65    0.00   600.00
##
## $hadoop
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -100.00    0.00    0.00   18.09   11.11   666.70
##
## $jmeter
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -100.00    0.00    0.00   12.97   11.76   500.00
##
## $log4j
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -33.33    0.00    0.00   11.00    0.00   200.00
##
## $tomcat
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -100.00    0.00    0.00   10.34    0.00   469.20

fc <- factor(data.CountInput.positive$Project)
interest <- data.CountInput.positive$interest
tapply(interest, fc, summary)

## $ant
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      4.00   20.26   33.33   61.72   78.75   600.00
##
## $hadoop
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
## 2.857 20.000 40.000 77.480 100.000 666.700
##
## $jmeter
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 5.882 16.670 29.710 53.300 63.540 500.000
##
## $log4j
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 11.11 27.56 50.00 67.32 100.00 200.00
##
## $tomcat
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 2.03 16.67 28.57 62.75 66.67 469.20
```

```
fc <- factor(data.CountInput.negative$Project)
interest <- data.CountInput.negative$interest
tapply(interest, fc, summary)
```

```
## $ant
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## -100.000 -50.000 -29.170 -35.480 -20.000 -8.333
##
## $hadoop
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## -100.000 -38.460 -25.000 -29.190 -12.500 -4.545
##
## $jmeter
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## -100.00 -39.13 -20.00 -26.81 -11.11 -5.00
##
## $log4j
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## -33.33 -25.00 -25.00 -22.67 -20.00 -10.00
##
## $tomcat
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## -100.00 -41.67 -25.00 -32.23 -12.50 -3.39
```

Plot

```
library(reshape2)
library(ggplot2)

if(0){
  idx <- data.CountLine.positive$Project == "apache-ant"
  a1 <- data.frame(Interest=data.CountLine.positive[idx,"interest"])
  g = ggplot(a1, aes(x=Interest, y=..density.., fill=T), lims(x = c(0,400)))
  g = g + geom_density(alpha = 0.5) + xlim(0, 400) + ylim(0,0.04) + guides(fill=FALSE)
  print(g)
  ggsave(file = "./tex/figures/rq1-ant.pdf", plot = g, width = 8.09, height = 5)

  idx <- data.CountLine.positive$Project == "apache-jmeter"
  a2 <- data.frame(Interest=data.CountLine.positive[idx,"interest"])
```



```

g = ggplot(a2, aes(x=Interest, y=..density.., fill=T, lims(x = c(0,400))))
g = g + geom_density(alpha = 0.5) + xlim(0, 400) + ylim(0,0.04) + guides(fill=FALSE)
print(g)
ggsave(file = "./tex/figures/rq1-jmeter.pdf", plot = g, width = 8.09, height = 5)

idx <- data.CountLine.positive$Project == "jruby"
a3 <- data.frame(Interest=data.CountLine.positive[idx,"interest"])
g = ggplot(a3, aes(x=Interest, y=..density.., fill=T), lims(x = c(0,400)))
g = g + geom_density(alpha = 0.5) + xlim(0, 400) + ylim(0,0.04) + guides(fill=FALSE)
print(g)
ggsave(file = "./tex/figures/rq1-jruby.pdf", plot = g, width = 8.09, height = 5)
}

if(0){
idx <- data.CountInput.positive$Project == "apache-ant"
a1 <- data.frame(Interest=data.CountInput.positive[idx,"interest"])
g = ggplot(a1, aes(x=Interest, y=..density.., fill=T), lims(x = c(0,400)))
g = g + geom_density(alpha = 0.5) + xlim(0, 400) + ylim(0,0.04) + guides(fill=FALSE)
print(g)
ggsave(file = "./tex/figures/rq1-ant-fanin.pdf", plot = g, width = 8.09, height = 5)

idx <- data.CountInput.positive$Project == "apache-jmeter"
a2 <- data.frame(Interest=data.CountInput.positive[idx,"interest"])
g = ggplot(a2, aes(x=Interest, y=..density.., fill=T, lims(x = c(0,400))))
g = g + geom_density(alpha = 0.5) + xlim(0, 400) + ylim(0,0.04) + guides(fill=FALSE)
print(g)
ggsave(file = "./tex/figures/rq1-jmeter-fanin.pdf", plot = g, width = 8.09, height = 5)

idx <- data.CountInput.positive$Project == "jruby"
a3 <- data.frame(Interest=data.CountInput.positive[idx,"interest"])
g = ggplot(a3, aes(x=Interest, y=..density.., fill=T), lims(x = c(0,400)))
g = g + geom_density(alpha = 0.5) + xlim(0, 400) + ylim(0,0.04) + guides(fill=FALSE)
print(g)
ggsave(file = "./tex/figures/rq1-jruby-fanin.pdf", plot = g, width = 8.09, height = 5)
}

```