

openQA Project - action #98604

Provide data about ratio of automatically approved SLE Maintenance incidents size:M

2021-09-14 12:20 - VANASTASIADIS

Status: Resolved	Start date: 2021-09-14
Priority: Normal	Due date: 2021-10-12
Assignee: VANASTASIADIS	% Done: 0%
Category: Organisational	Estimated time: 0.00 hour
Target version: Ready	
Difficulty:	

Description

User story

In order to have a better overview of our automatic testing procedures and improve our workflows, we need to know what percentage of tests fail and why. This should ideally be <10% under normal circumstances, so if we have a lot of test failures that then need manual intervention we should investigate why this happens and how to fix it.

Acceptance criteria

- **AC1::** data about the percentage of automatically approved incidents exist

Suggestions

- read data from smelt, i.e. <https://tools.io.suse.de/smelt/user/api/>
- gather the data in human readable format

Related issues:

Related to openQA Project - action #96191: Provide "fail-rate" of tests, espe...	Resolved	2021-07-28	2021-09-29
--	-----------------	-------------------	-------------------

History

#1 - 2021-09-14 12:30 - okurz

I would like to give it a more positive vibe. The goal is to motivate test maintainers to reach a lower false-positive rate and for that I would start with a higher rate of automatically accepted SLE maintenance release requests. As openQA job results are soon pruned, especially passed tests reading such data from openQA might give a too negative impression. Would it be possible to read the number of incidents from smelt where the qam-openqa review was approved automatically and not a human reviewer?

#2 - 2021-09-14 12:41 - cdywan

This seems a lot like [#96191](#). Can you disambiguate where this is different? Should this perhaps be an epic?

#3 - 2021-09-14 12:49 - VANASTASIADIS

cdywan wrote:

This seems a lot like [#96191](#). Can you disambiguate where this is different? Should this perhaps be an epic?

[#96191](#) is more concerned about mm tests, as a response to a specific problem (mm tests failing more frequently) while this one concerns overall failure rate and additional data for more general use.

I think turning this to an epic is a good idea. [okurz](#) what's your take?

#4 - 2021-09-14 19:46 - okurz

VANASTASIADIS wrote:

cdywan wrote:

This seems a lot like [#96191](#). Can you disambiguate where this is different? Should this perhaps be an epic?

[#96191](#) is more concerned about mm tests, as a response to a specific problem (mm tests failing more frequently) while this one concerns overall failure rate and additional data for more general use.

I think turning this to an epic is a good idea. [okurz](#) what's your take?

My take on it is [#98604#note-1](#) . Not sure if cdywan has read it or we wrote comments at roughly the same time. So the differences to [#96191](#) I see are:

- [#96191](#) is about failure rate of multi-machine tests, [#98604](#) is about rate of automatically approved SUSE SLE QA maintenance incident release requests
- The goal of [#96191](#) is to monitor the infrastructure with the tools team as target audience, goal of [#98604](#) is to motivate test maintainers to reduce rate of false positives
- [#96191](#) is likely best solved with SQL queries ran by telegraf on OSD, [#98604](#) is likely best solved by reading data from smelt

#5 - 2021-09-15 07:55 - cdywan

okurz wrote:

VANASTASIADIS wrote:

cdywan wrote:

This seems a lot like [#96191](#). Can you disambiguate where this is different? Should this perhaps be an epic?

[#96191](#) is more concerned about mm tests, as a response to a specific problem (mm tests failing more frequently) while this one concerns overall failure rate and additional data for more general use.

I think turning this to an epic is a good idea. [okurz](#) what's your take?

My take on it is [#98604#note-1](#) . Not sure if cdywan has read it or we wrote comments at roughly the same time. So the differences to [#96191](#) I see are:

- [#96191](#) is about failure rate of multi-machine tests, [#98604](#) is about rate of automatically approved SUSE SLE QA maintenance incident release requests
- The goal of [#96191](#) is to monitor the infrastructure with the tools team as target audience, goal of [#98604](#) is to motivate test maintainers to reduce rate of false positives
- [#96191](#) is likely best solved with SQL queries ran by telegraf on OSD, [#98604](#) is likely best solved by reading data from smelt

Thank you, Oli. I think it would help a lot if we can not refer to both as "tests". Meaning give this ticket a better title and also be clear in the description, which currently talks about "tests" in a way that can apply to both tickets.

#6 - 2021-09-16 09:42 - okurz

- Related to action [#96191](#): Provide "fail-rate" of tests, especially multi-machine, in grafana size:M added

#7 - 2021-09-16 09:46 - cdywan

- Subject changed from Gather data about the number of openQA test failures to Provide data about ratio of automatically approved SLE Maintenance incidents size:M

- Description updated

- Status changed from New to Workable

#8 - 2021-09-16 09:50 - cdywan

- Description updated

#9 - 2021-09-16 11:41 - VANASTASIADIS

- Assignee set to VANASTASIADIS

#10 - 2021-09-27 13:04 - VANASTASIADIS

Update and results so far:

Before posting the results, after talking with some of the reviewers, one important note: for the past couple of months, the workflow of openqa-reviewers has changed, resulting in drastic changes in the results as well. Namely:

- Before, they would review failing tests and then approve them manually
- Now, they ping the test developers when failures appear, so that they fix the tests, and wait for the bot to automatically approve them in the next run. They only approve manually in special cases

That drastically reduces the amount of manually approved requests. It doesn't necessarily mean that the failures are reduced, but since the reviewers notify the test developers (or even possibly fix the tests themselves) and then wait for the automatic approval, it's harder to see how many tests are actually failing.

Case in point:

```

PERIOD 01/01/2021 - 01/08/2021
-----
REQUESTS | APPROVED_BY
-----
1000      real users (manual approval)
 699      sle-qam-openqa (automatic approval)

```

```

PERIOD 01/08/2021 - 24/09/2021
-----
REQUESTS | APPROVED_BY
-----
105       others
444       sle-qam-openqa

```

The new approach is potentially better, as it drives people to fix and improve tests when failures appear. However, looking at the percentage of auto-approvals no more hints to the state of the tests, since they don't happen anymore when failures appear, but are only used as a last resort.

Further data upcoming.

#11 - 2021-09-27 13:23 - VANASTASIADIS

- Status changed from Workable to In Progress

#12 - 2021-09-28 04:10 - openqa_review

- Due date set to 2021-10-12

Setting due date based on mean cycle time of SUSE QE Tools

#13 - 2021-09-28 09:33 - okurz

VANASTASIADIS wrote:

Update and results so far:
[...]

thank you for providing such detailed explanation and discussion of the results

Case in point:

```

PERIOD 01/01/2021 - 01/08/2021
-----
REQUESTS | APPROVED_BY
-----
1000      real users (manual approval)
 699      sle-qam-openqa (automatic approval)

PERIOD 01/08/2021 - 24/09/2021
-----
REQUESTS | APPROVED_BY
-----
105       others
444       sle-qam-openqa

```

is "others" corresponding to real users then?

So the auto-approval ratio before 2021-08-01 is 0.70 (699/1000) and after that 4.23 . That seems like a way too high improvement :) It's interesting to see that there had been nearly as many sle-qam-openqa approvals in just 2 months compared to 7 months in before?

#14 - 2021-09-28 09:53 - VANASTASIADIS

[okurz](#) thank you, yes, 'others' is referring to real users as in the first example.

About your last paragraph, indeed the change is big. Keep in mind however that the much higher rate of auto approval lately doesn't necessarily mean that tests fail less per se, but that now instead of manually approving requests, the tests are fixed and then the requests get auto-approved in the next run. This probably helps with fixing tests (and makes the reviewing procedure more streamlined). On the other hand it also means auto-approvals won't reflect the test failure rate accurately anymore.

#15 - 2021-09-29 08:25 - okurz

VANASTASIADIS wrote:

[...] the much higher rate of auto approval lately [means] that now instead of manually approving requests, the tests are fixed and then the

requests get auto-approved in the next run.

That would be lovely and is what I want to reach as well but I doubt it :)

On the other hand it also means auto-approvals won't reflect the test failure rate accurately anymore.

Same as in before the "manual/automatic approval"-ratio would tell us because in the end if tests are not fixed then someone asks for manual approval. This for example just happened the past days in <https://suse.slack.com/archives/C02D16TCP99/p1632753957114700> regarding <https://smelt.suse.de/incident/21254/>. Despite me fixing some tests there was at least one blocking test failure. Then a reviewer manually approved the update and after asking back they stated that the decision was made in a verbal meeting with no written down information about that decision (IMHO suboptimal).

#16 - 2021-09-29 12:56 - VANASTASIADIS

- File `auto_approvals_query` added
- File `auto_approvals_list` added
- File `2021_auto_approvals.png` added
- File `2020-21_auto_approvals.png` added

All results attached as files.
Changing this to "Feedback" and then "Resolved" is the data are deemed sufficient.

#17 - 2021-09-29 12:57 - VANASTASIADIS

- Status changed from *In Progress* to *Feedback*

#18 - 2021-09-29 13:22 - okurz

ok, for convenience I can link in the images for display directly:

- 2020-2021
2020-21_auto_approvals.png
- 2021
2021_auto_approvals.png

The graph 2020-2021 looks like the auto-approval ratio was increasing already since the mid of 2020, interesting.

Thank you for providing the query in a way reusable way.

- What would be necessary to provide the access to the smelt DB to other users, e.g. members of SUSE QE Tools?
- As SMELT seems to be connected somewhat to <https://maintenance-statistics.dyn.cloud.suse.de/browse/3> already, how about we try to include queries in there?

EDIT: I tried to include it in https://maintenance-statistics.dyn.cloud.suse.de/question/329?group=qam-openqa&start_date=2021-01-01&end_date=2021-09-30 but I get "ERROR: permission denied for table auth_user", maybe we can find a way to not need that table, maybe just hardcode the id?

#19 - 2021-09-29 13:44 - VANASTASIADIS

- You are correct, indeed the SMELT database can be accessed via metabase.
- You are also correct that some tables (for example the `auth_user` one) are not accessible there, for privacy/security reasons. Most tables are accessible, though.

For most queries this won't be a problem. For this specific query, the users were needed since the `openqa-review` bot exists as a user in smelt, and to see what it does you need to be able to at least filter by it.

Maybe there's a way for that to be bypassed, I didn't try much tbh - once I saw I couldn't access the table in metabase I just queried directly. I will look at it.

#20 - 2021-09-29 15:12 - okurz

if you could just lookup the user id of the bot user then I guess we can use that

#21 - 2021-09-29 15:28 - VANASTASIADIS

[okurz](#) here is an implementation using the ids of `qam-openqa` group and `sle-qam-openqa` user: <https://maintenance-statistics.dyn.cloud.suse.de/question/330>

#22 - 2021-09-29 15:54 - okurz

- Status changed from Feedback to Resolved

this is awesome! Thanks a lot. I have updated the query with two date-pickers so that one can select different time spans. Seems I could save and overwrite your code. I hope you don't mind. I have provided a comment on <https://progress.opensuse.org/issues/96543#note-6> . With this I think we did more than enough and can call this ticket resolved.

#23 - 2021-09-30 14:14 - VANASTASIADIS

Thanks Oliver, no problem at all. You can alter and use the queries however you see fit. Thank you for updating the other ticket.

Files

auto_approvals_list	2.44 KB	2021-09-29	VANASTASIADIS
auto_approvals_query	1.33 KB	2021-09-29	VANASTASIADIS
2021_auto_approvals.png	53.5 KB	2021-09-29	VANASTASIADIS
2020-21_auto_approvals.png	61.6 KB	2021-09-29	VANASTASIADIS