KubeVirt

squash the flakes!

FOSDEM 2024

Daniel Hiller

KubeVirt

agenda

- about me
- about flakes
- impact of flakes
- flake process
- tools
- the future
- Q&A
- want to help?

about me

- Software Engineer @ Red Hat OpenShift Virtualization team
- KubeVirt CI, automation in general

KubeVirt

about flakes

a question:

who can explain what a flake is?

KubeVirt

about flakes

a ***flake***

...

...

...

KubeVirt

about flakes

a ***flake***

is a **test** that

without any code change

will either **fail** or **pass** in successive runs

about flakes

another question:

who thinks handling flakes is important?

about flakes

another question:

who has had to deal with flakes?

about flakes

another question:

who has to deal with flakes on a regular basis?

about flakes

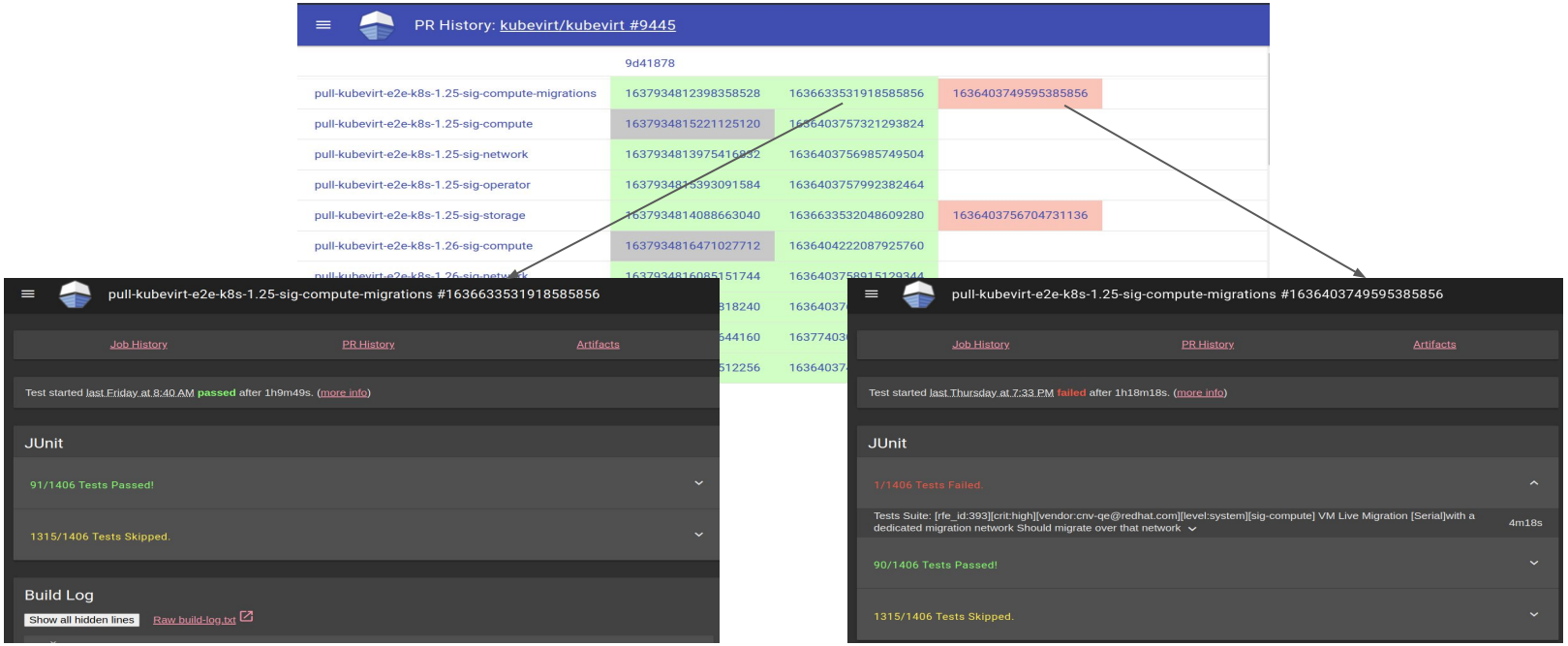
flakes are caused

either by production code (a bug)

or flaky test code (also a bug, but handled differently)

KubeVirt

about flakes



source: https://prow.ci.kubevirt.io/pr-history/?org=kubevirt&repo=kubevirt&pr=9445

impact of flakes

from “a survey of flaky tests”:

- **97%** of flakes were **false alarms**, and
- more than **50%** of flakes could not be **reproduced in isolation**

this leads to the conclusion:
“ignoring flaky tests is ok”

source: “A survey of flaky tests”

impact of flakes

impact of flakes

in CI automated testing must give a **reliable signal of stability**

any failed test run signals that the product is **unstable**

test runs failed due to flakes do not give this reliable signal

they only waste time

impact of flakes

Flaky tests waste everyone’s time - they cause

- longer feedback cycles for developers
- slowdown of merging pull requests - “retest trap”
- reversal of acceleration effects (i.e. batch testing)

impact of flakes

Flaky tests also cause trust issues - they make people

- lose trust in automated testing
- ignore test results

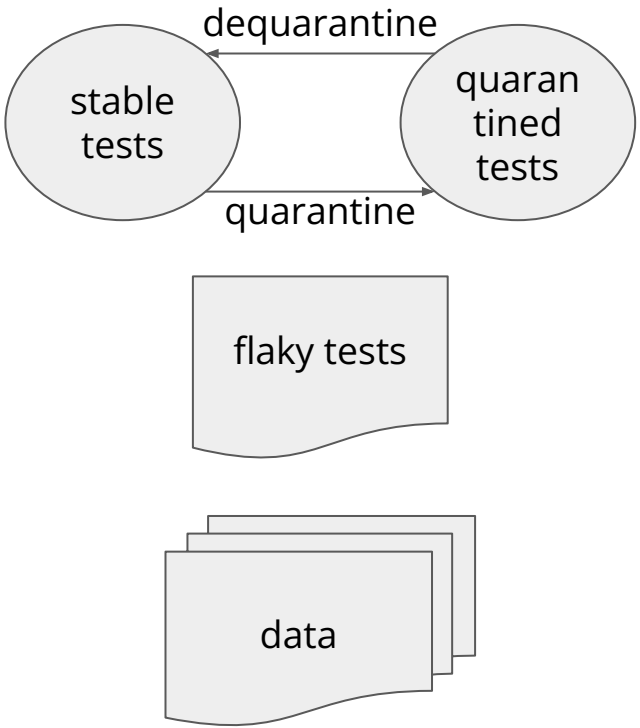
minimizing the impact

tldr; exclude (aka quarantine) a flaky test from test runs **as early as possible**, but **only as long as necessary**

minimizing the impact

what do we need?

- ability to move a test from **set of stable** tests into **set of quarantined** tests and back
- a report over possible flaky tests
- enough runtime data to triage flakes
  - devs decide whether we quarantine right away or they can fix them in time



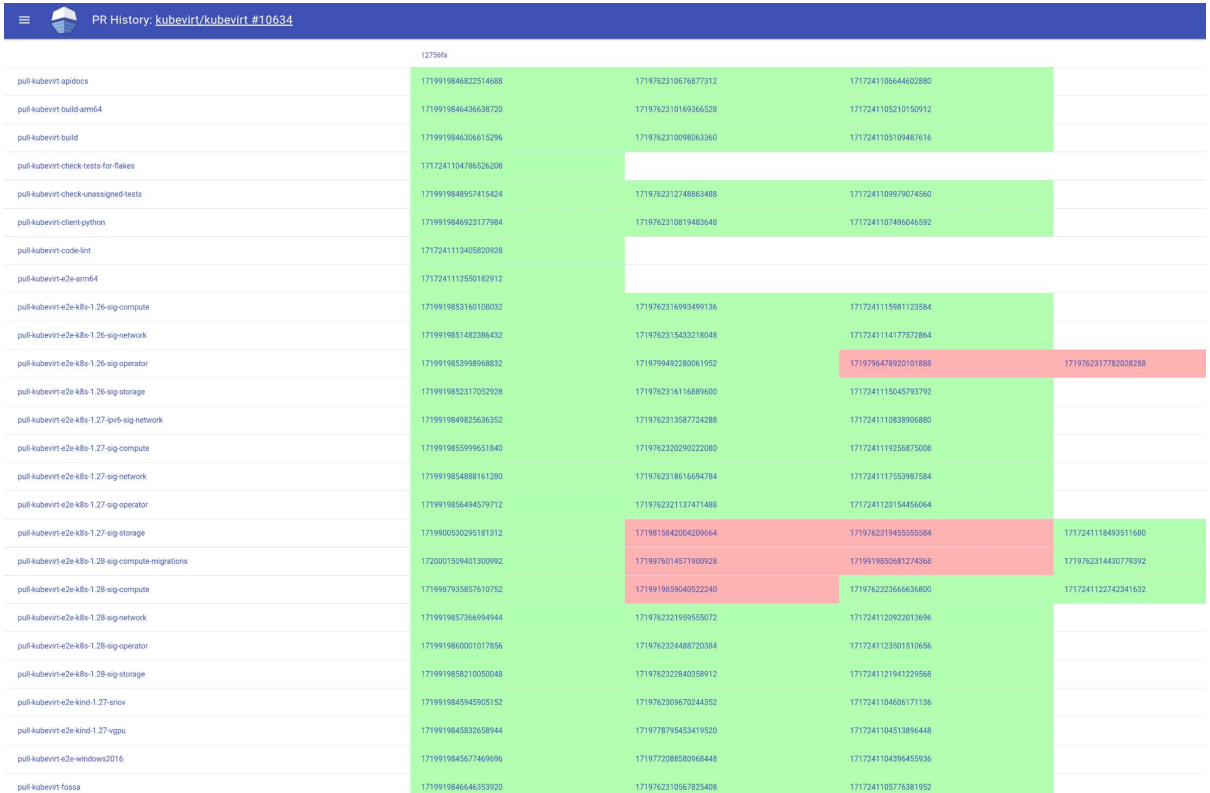
KubeVirt

minimizing the impact

how can we find flaky tests?

any merged PR had all tests succeeding in the end,

thus any test run with test failures from that PR *might* contain execution of flaky tests



KubeVirt

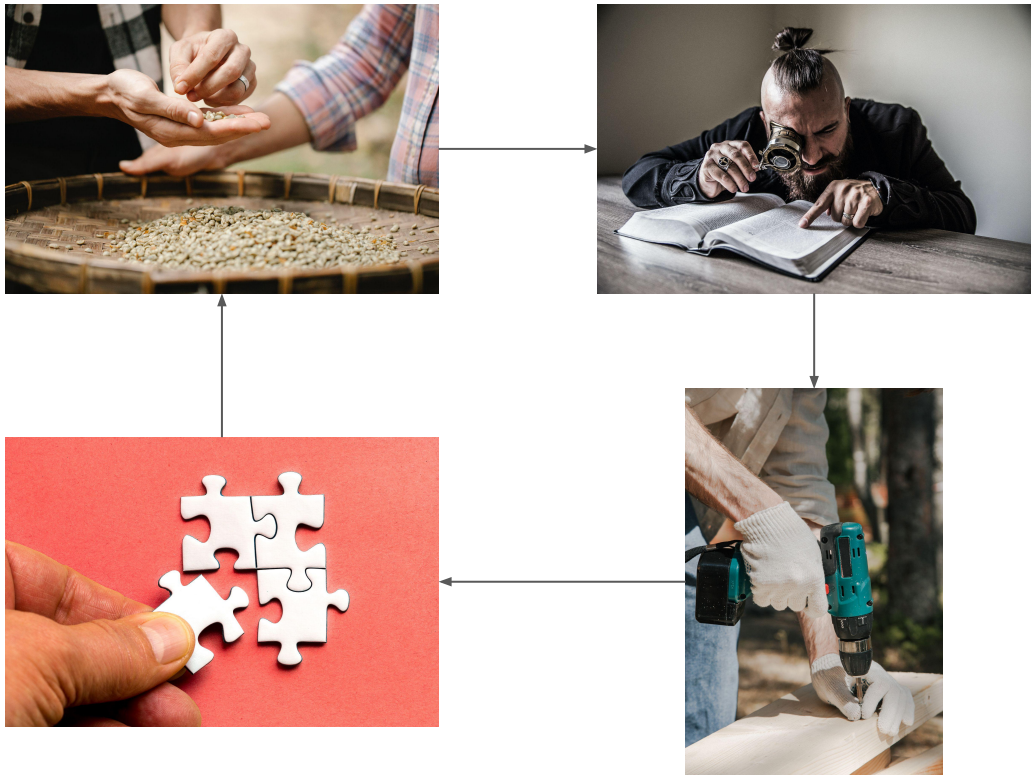
the flake process

regular meeting

- look at flakes
- decide: fix or quarantine?
- hand to dev
- bring back in

emergency quarantine

source: QUARANTINE.md



tools

ci honoring QUARANTINE label

```
# If KUBEVIRT_QUARANTINE is not set, do not run quarantined tests. When it is
# set the whole suite (quarantined and stable) will be run.
if [ -z "$KUBEVIRT_QUARANTINE" ]; then
    if [ -n "$KUBEVIRT_E2E_SKIP" ]; then
        export KUBEVIRT_E2E_SKIP="${KUBEVIRT_E2E_SKIP}|QUARANTINE"
    else
        export KUBEVIRT_E2E_SKIP="QUARANTINE"
    fi
fi
```

- presubmits **skip** quarantined tests
- periodics **execute** quarantined tests to check their stability

```
176
177         It("[QUARANTINE]should successfully upgrade virt-handler", func() {
178                 var expectedEventsLock sync.Mutex
179                 expectedEvents := []string{
180                         "maxUnavailable=1",
181                         "maxUnavailable=10%",
182                         "virt-handler=ready",
183                         "maxUnavailable=1",
184                 }
185
186                 ds, err := virtCli.AppsV1().DaemonSets(flags.KubeVirtInstall
```

sources:

- https://github.com/kubevirt/kubevirt/blob/38c01c34acecfafc89078b1bbaba8d9cf3cf0d4d/automation/test.sh#L452
- https://github.com/kubevirt/kubevirt/blob/38c01c34acecfafc89078b1bbaba8d9cf3cf0d4d/hack/functests.sh#L69
- https://github.com/kubevirt/kubevirt/blob/38c01c34acecfafc89078b1bbaba8d9cf3cf0d4d/tests/canary_upgrade_test.go#L177

KubeVirt

tools

flake stats report

the high level overview

(source)

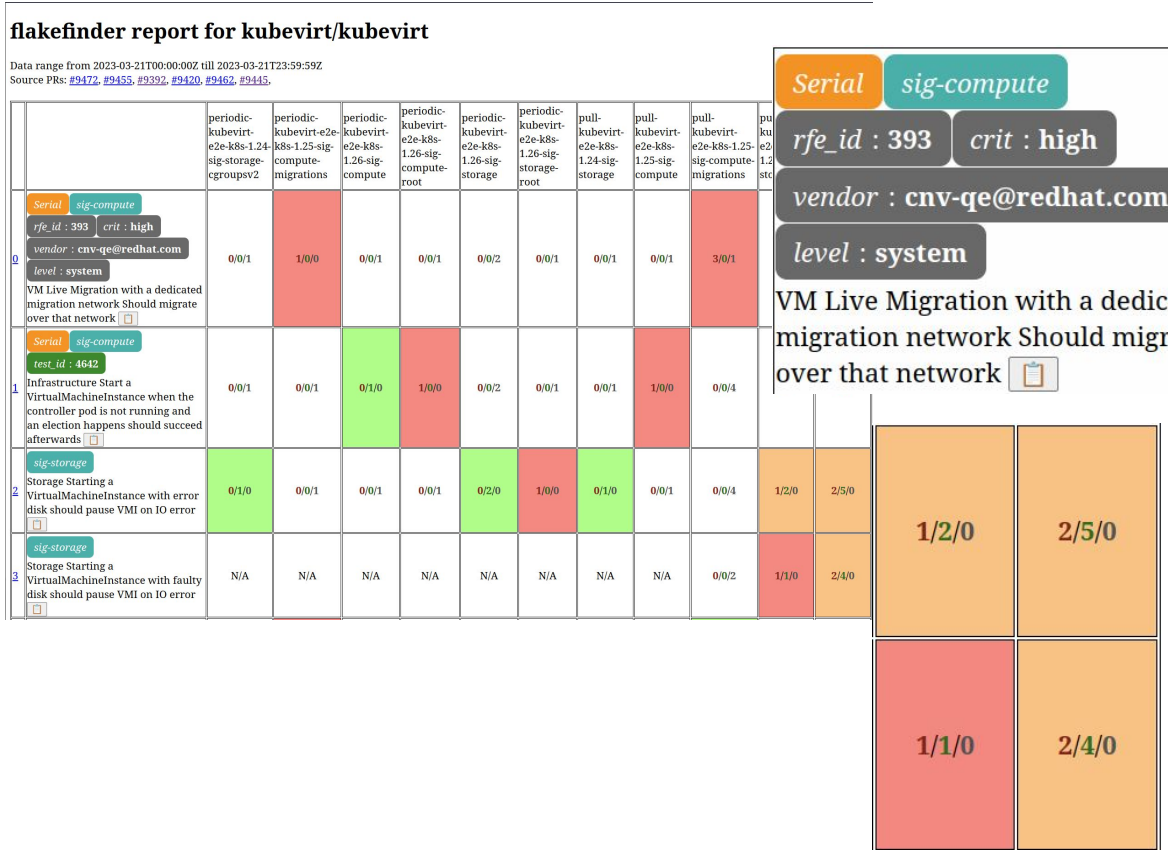
KubeVirt

tools

[flakefinder](#) report

the detail overview

gives an overview of the current flaky tests



flakefinder report for kubevirt/kubevirt

Data range from 2023-03-21T00:00:00Z till 2023-03-21T23:59:59Z
Source PRs: #9472, #9455, #9392, #9420, #9462, #9445,

		periodic-kubevirt-e2e-k8s-1.24-sig-storage-cgroupsv2	periodic-kubevirt-e2e-k8s-1.25-sig-compute-migrations	periodic-kubevirt-e2e-k8s-1.26-sig-compute	periodic-kubevirt-e2e-k8s-1.26-sig-compute-root	periodic-kubevirt-e2e-k8s-1.26-sig-storage	periodic-kubevirt-e2e-k8s-1.26-sig-storage-root	pull-kubevirt-e2e-k8s-1.24-sig-storage	pull-kubevirt-e2e-k8s-1.25-sig-compute	pull-kubevirt-e2e-k8s-1.25-sig-compute-migrations	pu ku e2 1.2 sto	
0	Serial sig-compute rfe_id : 393 crit : high vendor : cnv-qe@redhat.com level : system VM Live Migration with a dedicated migration network Should migrate over that network	0/0/1	1/0/0	0/0/1	0/0/1	0/0/2	0/0/1	0/0/1	0/0/1	3/0/1		
1	Serial sig-compute test_id : 4642 Infrastructure Start a VirtualMachineInstance when the controller pod is not running and an election happens should succeed afterwards	0/0/1	0/0/1	0/1/0	1/0/0	0/0/2	0/0/1	0/0/1	1/0/0	0/0/4		
2	sig-storage Storage Starting a VirtualMachineInstance with error disk should pause VMI on IO error	0/1/0	0/0/1	0/0/1	0/0/1	0/2/0	1/0/0	0/1/0	0/0/1	0/0/4	1/2/0	2/5/0
3	sig-storage Storage Starting a VirtualMachineInstance with faulty disk should pause VMI on IO error	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0/0/2	1/1/0	2/4/0

Serial sig-compute rfe_id : 393 crit : high vendor : cnv-qe@redhat.com level : system VM Live Migration with a dedicated migration network Should migrate over that network

1/2/0 2/5/0 1/1/0 2/4/0

tools

ci-health

record metrics over merge-queue-length, time-to-merge, retests-to-merge and merges-per-day

README.md

kubevirt/kubevirt

avg merge queue length 1.29 ± std 1.67 | days to merge 0.42 ± std 0.23

retests to merge 0.63 ± std 0.70

Merged PRs with 0 retest vs. Merged PRs 8 / 16

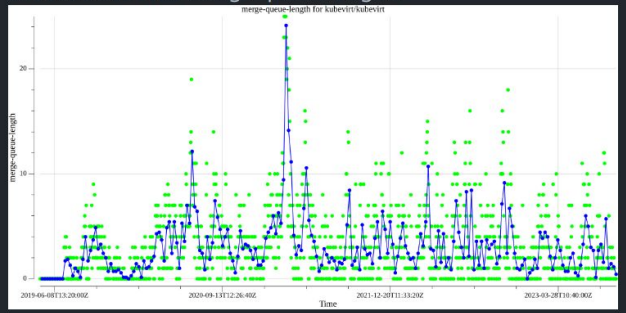
Latest execution data

Latest weeks data

Historical data evolution

These plots will be updated every week.

- kubevirt/kubevirt merge queue length:



Data available here.

tools

quarantine overview

(source)

Overview of Quarantine tests

Total: 1 tests

https://github.com/kubevirt/kubevirt/tree/c35be138d85864b17946ae1fe07f99a75445d501/tests/operator/operator.go

```
var _ = Describe("[Serial][sig-operator]Operator", Serial, decorators.SigOperator, func() {
        Describe("[rfe_id:2897][crit:medium][vendor:cnv-qe@redhat.com][level:component]Dynamic feature detection", func() {
                It("[test_id:3153][QUARANTINE] Ensure infra can handle dynamically detecting DataVolume Support", func() {
```

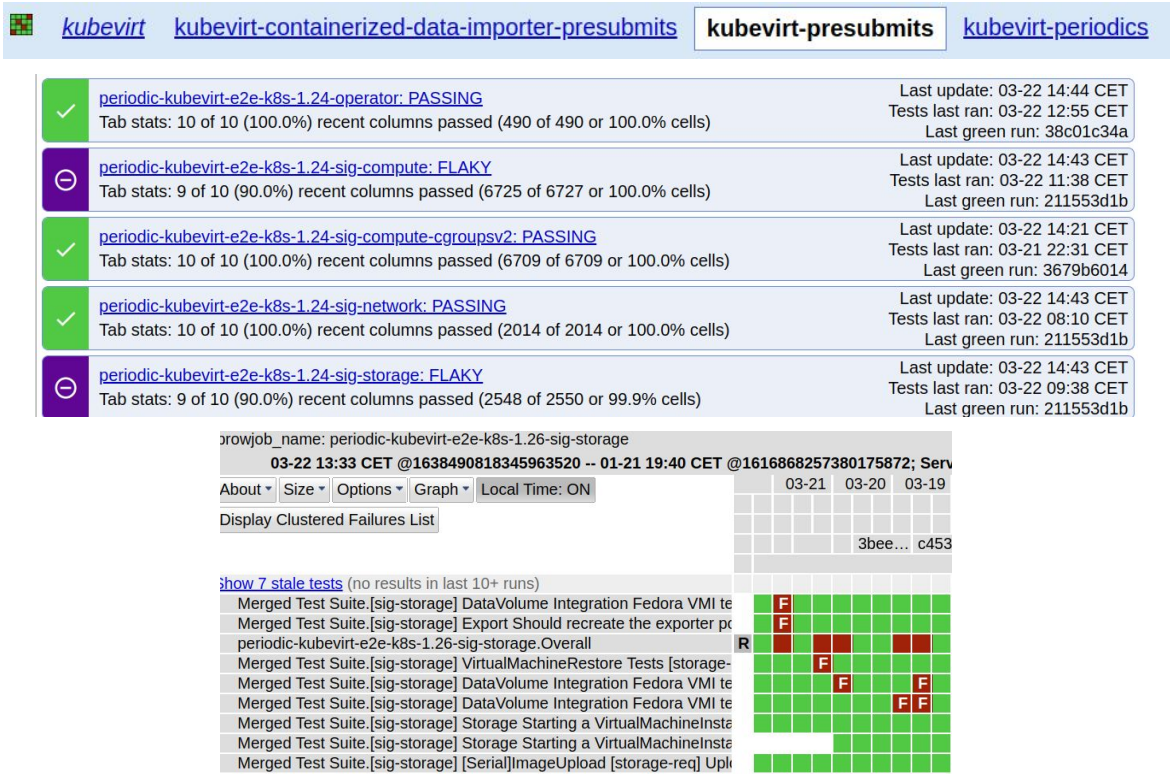
2 months

Last updated: 2023-09-08 10:18:02.458076498 +0000 UTC m=+3.056989100

tools

[testgrid]

drill down on all jobs for kubevirt/kubevirt that are running inside KubeVirt Prow



tools

check-tests-for-flakes test lane

tries to catch flakes before they enter the codebase

(source)

a test lane that

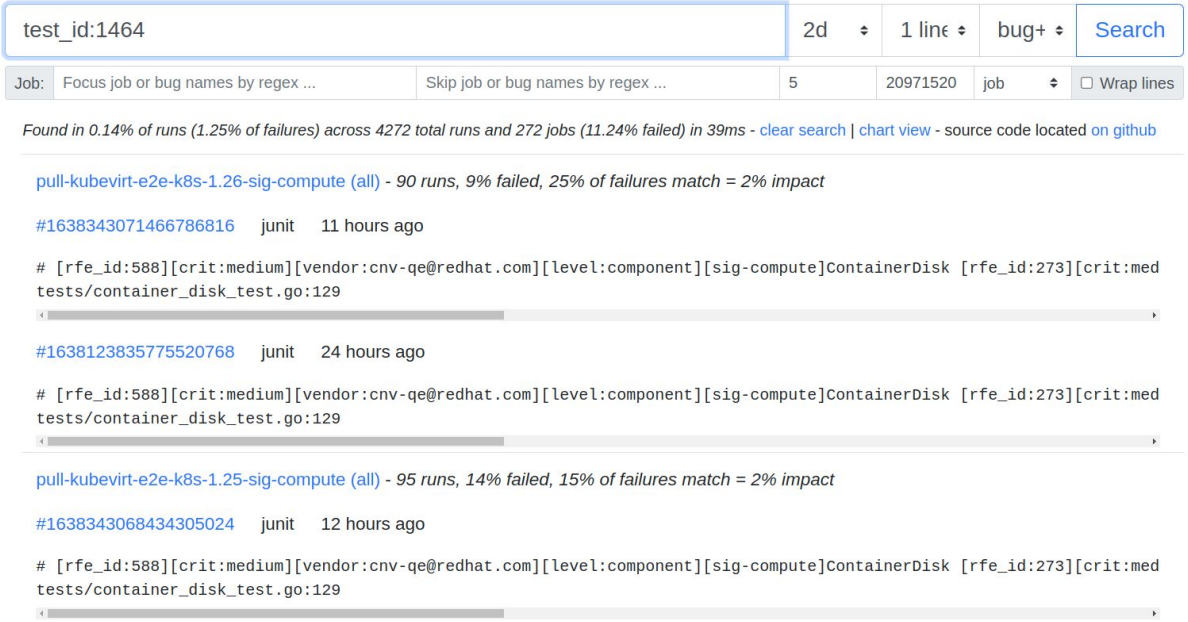
- selects the changed e2e test files from the commit set
- runs changed e2e tests five times
- runs in random execution order to catch order-dependent test

KubeVirt

tools

[ci-search]

search for terms in prow job logs (see [openshift ci-search])



in a nutshell

In regular intervals:

- follow up on previous action items
- derive action items from data available
- hand action items over to dev teams
- revisit and dequarantine quarantined tests

the future - more data, more tooling

gaps we want to close:

- collect more data - run the majority of tests frequently
- close the retest gap - stop retesting after a certain threshold
- get better in detecting new flakes
- long term - automatic quarantine PRs when new flakes have entered the codebase

Q&A

Any questions?

Who else is trying to tackle this problem?

What have you done to solve this problem?

KubeVirt

Thank you for your patience!

Feel free to send questions and comments:

mailto:	dhiller@redhat.com
k8s slack:	@dhiller
mastodon:	@dhiller@fosstodon.org
web:	www.dhiller.de

KubeVirt

interested in kubevirt.io? want to help?

- join #kubevirt-dev Slack channel
- join kubevirt-dev Google group
- fix flakes on kubevirt/kubevirt