The back of how we build database

Agenda

- What is the most important of database
- The layers
- What did we need
- How did we do it

What is the most important of database?

- performance ?
- usability ?
- stability ?
- functional ?
- else ?

What is the different of the beginning. the order? why?

What we deliver. Just TiDB? with libs?

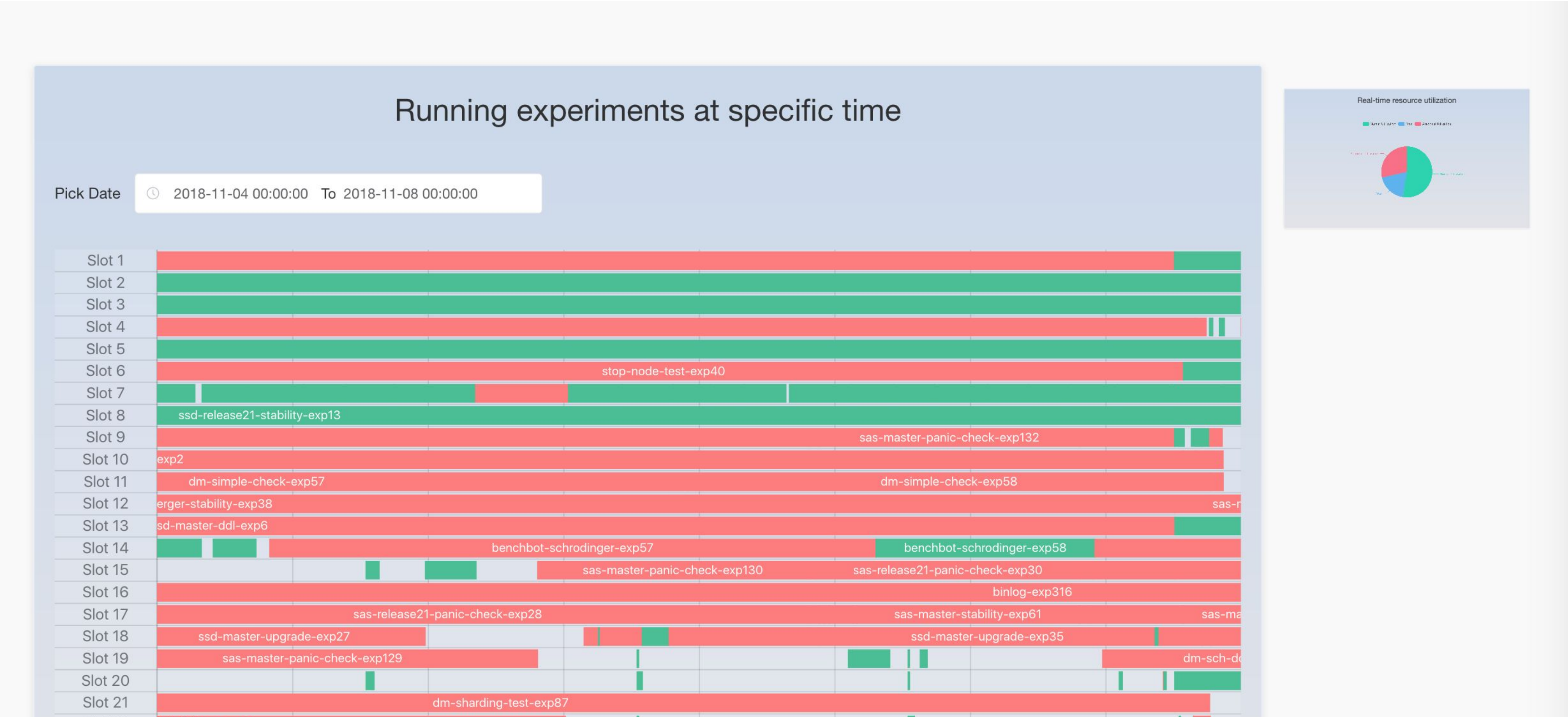
The layers

- TiDB(Parser/Prepare/Plan/xxx)
- TiKV(Engine/Coprocessor/Rocksdb/Titan/xxx)
- Fs/Kernel

What did we need

- A monkey. destroy things random. Named chaos. run in box
- A balance. measure on you
- A god. watching on you
- A cop. Monitor for you
- A bot. assisted you

The monkey



Running experiments at specific time

Pick Date 2018-11-04 00:00:00 To 2018-11-08 00:00:00

The origion

- the base linux cmd (tc/iptables/cgroup…)
- fuse
- jepsen

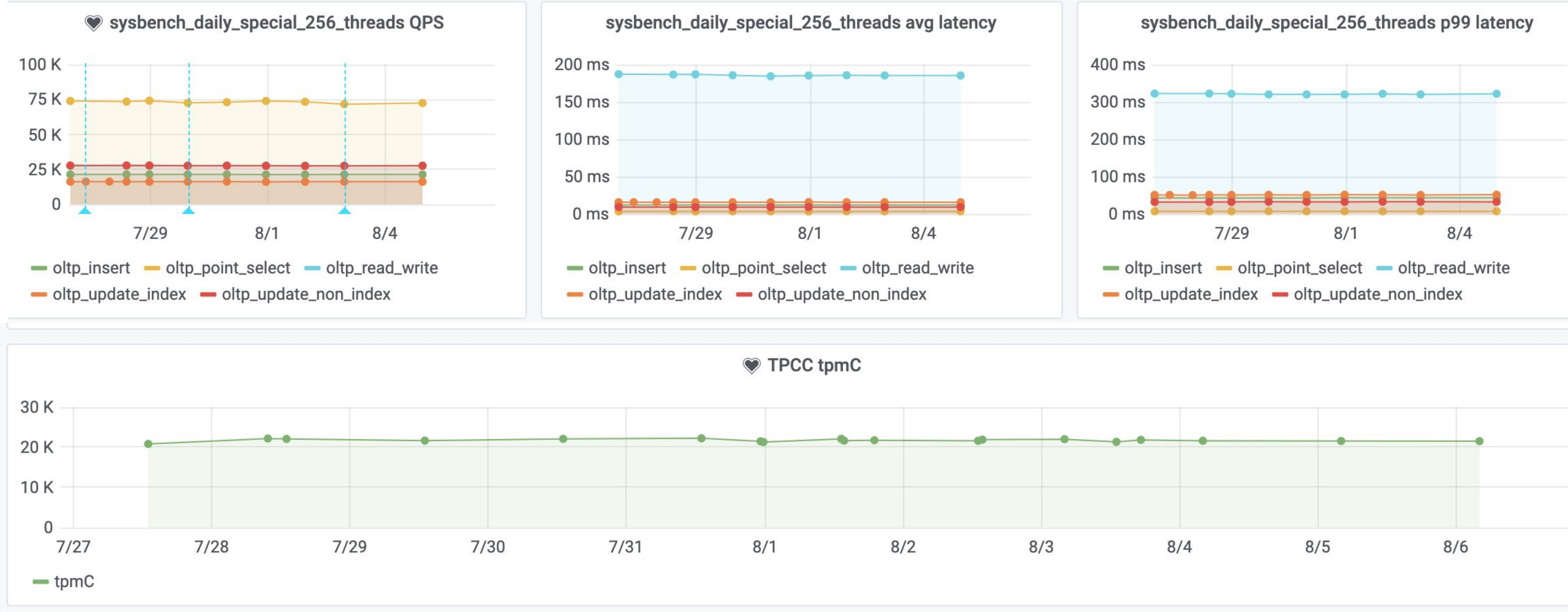
future

- chaos operator. cloud native

The key

- test -> oberserve(how.what) -> result -> test

The god

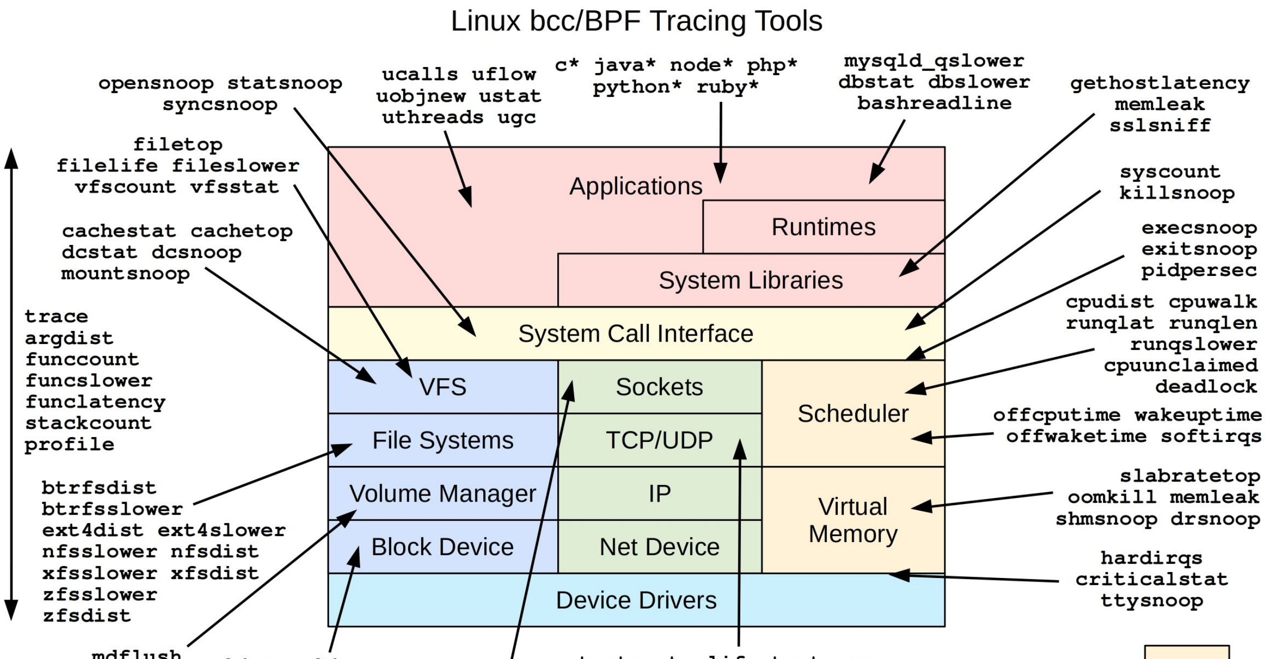

sysbench_daily_special_256_threads QPS
100 K
75 K
50 K
25 K
0
7/29
8/1
8/4
oltp_insert
oltp_point_select
oltp_read_write
oltp_update_index
oltp_update_non_index
sysbench_daily_special_256_threads avg latency
200 ms
150 ms
100 ms
50 ms
0 ms
7/29
8/1
8/4
oltp_insert
oltp_point_select
oltp_read_write
oltp_update_index
oltp_update_non_index
sysbench_daily_special_256_threads p99 latency
400 ms
300 ms
200 ms
100 ms
0 ms
7/29
8/1
8/4
oltp_insert
oltp_point_select
oltp_read_write
oltp_update_index
oltp_update_non_index
TPCC tpmC
30 K
20 K
10 K
0
7/27
7/28
7/29
7/30
7/31
8/1
8/2
8/3
8/4
8/5
8/6
tpmC

benchmark as a service

- A large mount of workloads
- the key point
  - what should we do when code is unstable?
  - when average value almost same. how did you track it

The kernel trace tools

- ebpf
- heaptrack (http://github.com/ethercflow/eheaptrack)



Linux bcc/BPF Tracing Tools

opensnoop statsnoop
syncsnoop

ucalls uflow
uobjnew ustat
uthreads ugc

c* java* node* php*
python* ruby*

mysqld_qslower
dbstat dbslower
bashreadline

gethostlatency
memleak
sslsniff

filetop
filelife fileslower
vfscount vfsstat

cachestat cachetop
dcstat dcsnoop
mountsnoop

trace
argdist
funccount
funcslower
funclatency
stackcount
profile

btrfsdist
btrfsslower
ext4dist ext4slower
nfsslower nfsdist
xfsslower xfsdist
zfsslower
zfsdist

mdflush

Applications

Runtimes

System Libraries

System Call Interface

VFS

Sockets

Scheduler

File Systems

TCP/UDP

Volume Manager

IP

Virtual
Memory

Block Device

Net Device

Device Drivers

syscount
killsnoop

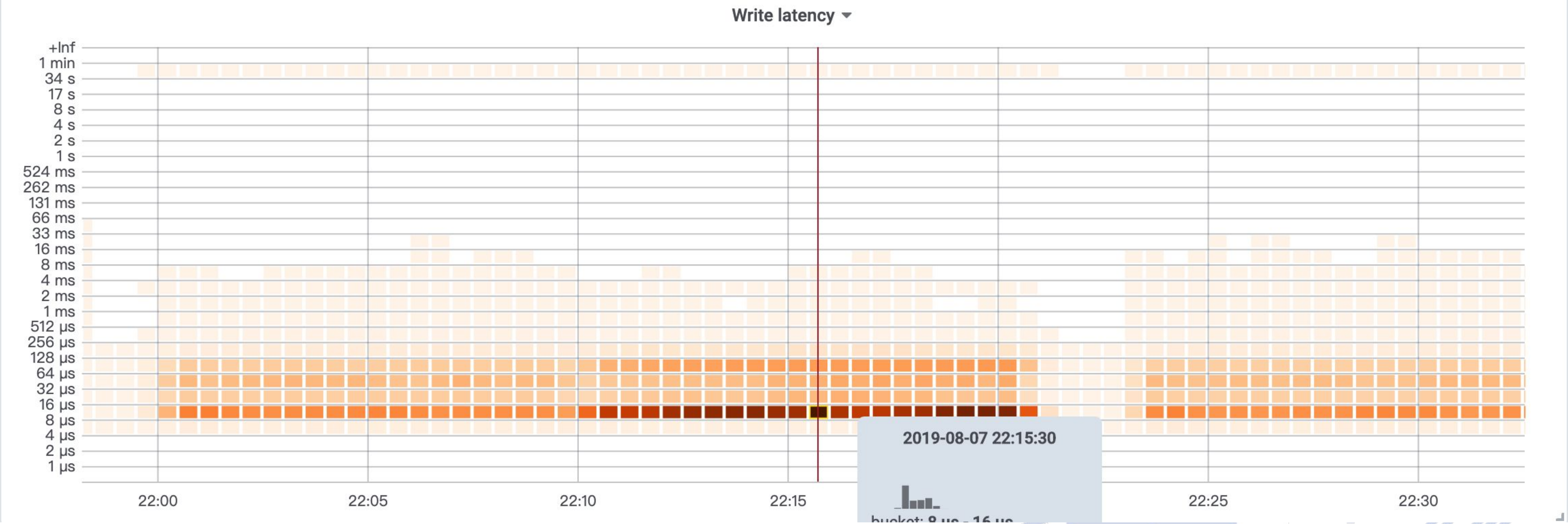
execsnoop
exitsnoop
pidpersec

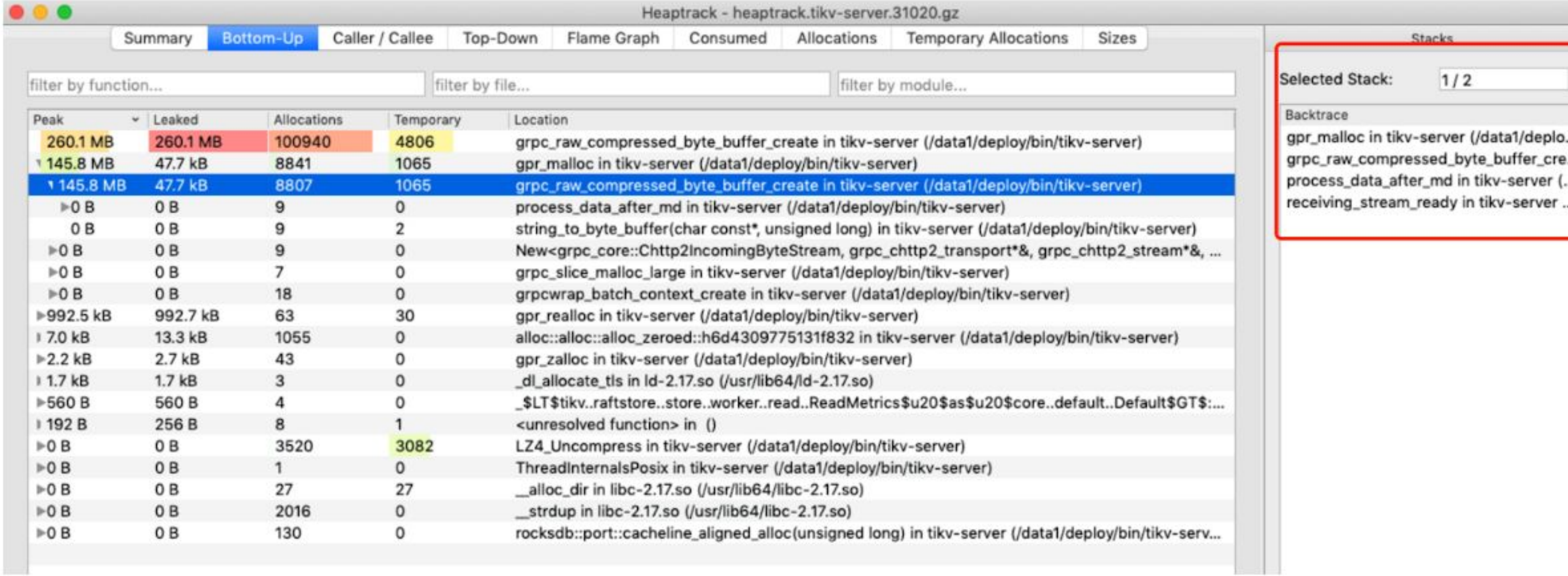
cpudist cpuwalk
runqlat runqlen
runqslower
cpuunclaimed
deadlock

offcputime wakeuptime
offwaketime softirqs

slabratetop
oomkill memleak
shmsnoop drsnoop

hardirqs
criticalstat
ttysnoop


Write latency
+Inf
1 min
34 s
17 s
8 s
4 s
2 s
1 s
524 ms
262 ms
131 ms
66 ms
33 ms
16 ms
8 ms
4 ms
2 ms
1 ms
512 µs
256 µs
128 µs
64 µs
32 µs
16 µs
8 µs
4 µs
2 µs
1 µs
22:00
22:05
22:10
22:15
22:25
22:30
2019-08-07 22:15:30
bucket: 8 µs - 16 µs



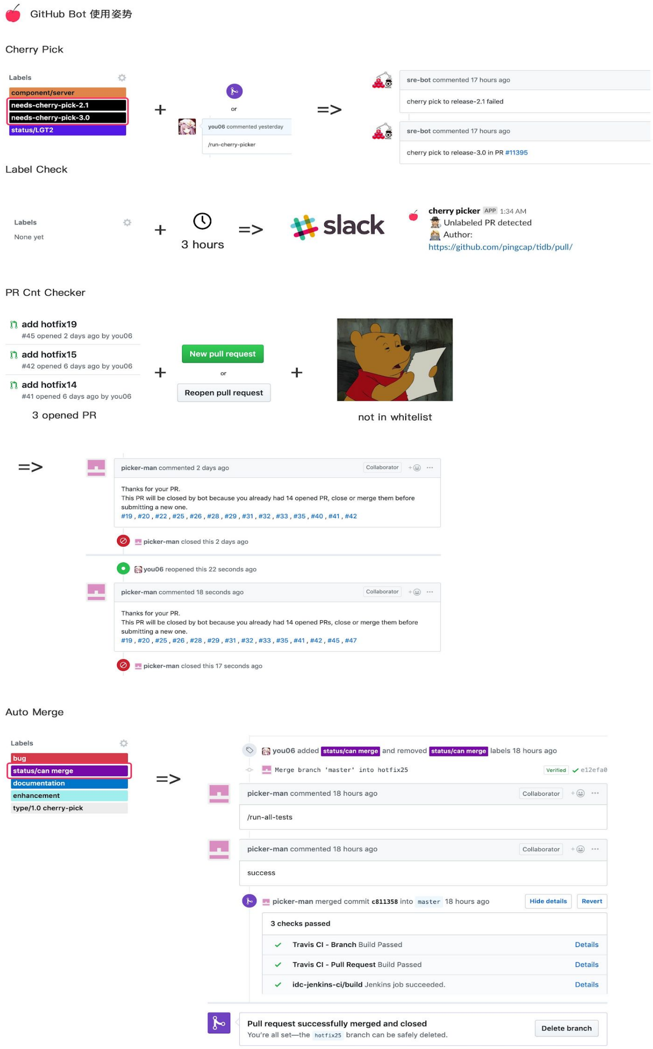
Peak	Leaked	Allocations	Temporary	Location
260.1 MB	260.1 MB	100940	4806	grpc_raw_compressed_byte_buffer_create in tikv-server (/data1/deploy/bin/tikv-server)
▾ 145.8 MB	47.7 kB	8841	1065	gpr_malloc in tikv-server (/data1/deploy/bin/tikv-server)
▾ 145.8 MB	47.7 kB	8807	1065	grpc_raw_compressed_byte_buffer_create in tikv-server (/data1/deploy/bin/tikv-server)
▸ 0 B	0 B	9	0	process_data_after_md in tikv-server (/data1/deploy/bin/tikv-server)
0 B	0 B	9	2	string_to_byte_buffer(char const*, unsigned long) in tikv-server (/data1/deploy/bin/tikv-server)
▸ 0 B	0 B	9	0	New<grpc_core::Chttp2IncomingByteStream, grpc_chttp2_transport*&, grpc_chttp2_stream*&, ...
▸ 0 B	0 B	7	0	grpc_slice_malloc_large in tikv-server (/data1/deploy/bin/tikv-server)
▸ 0 B	0 B	18	0	grpcwrap_batch_context_create in tikv-server (/data1/deploy/bin/tikv-server)
▸ 992.5 kB	992.7 kB	63	30	gpr_realloc in tikv-server (/data1/deploy/bin/tikv-server)
▸ 7.0 kB	13.3 kB	1055	0	alloc::alloc::alloc_zeroed::h6d4309775131f832 in tikv-server (/data1/deploy/bin/tikv-server)
▸ 2.2 kB	2.7 kB	43	0	gpr_zalloc in tikv-server (/data1/deploy/bin/tikv-server)
▸ 1.7 kB	1.7 kB	3	0	_dl_allocate_tls in ld-2.17.so (/usr/lib64/ld-2.17.so)
▸ 560 B	560 B	4	0	_$LT$tikv..raftstore..store..worker..read..ReadMetrics$u20$as$u20$core..default..Default$GT$:...
▸ 192 B	256 B	8	1	<unresolved function> in ()
▸ 0 B	0 B	3520	3082	LZ4_Uncompress in tikv-server (/data1/deploy/bin/tikv-server)
▸ 0 B	0 B	1	0	ThreadInternalsPosix in tikv-server (/data1/deploy/bin/tikv-server)
▸ 0 B	0 B	27	27	__alloc_dir in libc-2.17.so (/usr/lib64/libc-2.17.so)
▸ 0 B	0 B	2016	0	__strdup in libc-2.17.so (/usr/lib64/libc-2.17.so)
▸ 0 B	0 B	130	0	rocksdb::port::cacheline_aligned_alloc(unsigned long) in tikv-server (/data1/deploy/bin/tikv-serv...

Selected Stack: 1 / 2

Backtrace
gpr_malloc in tikv-server (/data1/deplo
grpc_raw_compressed_byte_buffer_cre
process_data_after_md in tikv-server (.
receiving_stream_ready in tikv-server .

The bot

- key
  - the webhook
  - the label system
  - the command system



GitHub Bot 使用姿势

Cherry Pick

Labels

component/server

needs-cherry-pick-2.1

needs-cherry-pick-3.0

status/LGT2

+ or

you06 commented yesterday

/run-cherry-picker

=>

sre-bot commented 17 hours ago

cherry pick to release-2.1 failed

sre-bot commented 17 hours ago

cherry pick to release-3.0 in PR #11395

Label Check

Labels

None yet

+ 3 hours => slack

cherry picker APP 1:34 AM

Unlabeled PR detected

Author:

https://github.com/pingcap/tidb/pull/

PR Cnt Checker

add hotfix19

#45 opened 2 days ago by you06

add hotfix15

#42 opened 6 days ago by you06

add hotfix14

#41 opened 6 days ago by you06

3 opened PR

+ New pull request or Reopen pull request +

not in whitelist

=>

picker-man commented 2 days ago · Collaborator

Thanks for your PR.
This PR will be closed by bot because you already had 14 opened PR, close or merge them before submitting a new one.
#19 , #20 , #22 , #25 , #26 , #28 , #29 , #31 , #32 , #33 , #35 , #40 , #41 , #42

picker-man closed this 2 days ago

you06 reopened this 22 seconds ago

picker-man commented 18 seconds ago · Collaborator

Thanks for your PR.
This PR will be closed by bot because you already had 14 opened PRs, close or merge them before submitting a new one.
#19 , #20 , #25 , #26 , #28 , #29 , #31 , #32 , #33 , #35 , #41 , #42 , #45 , #47

picker-man closed this 17 seconds ago

Auto Merge

Labels

bug

status/can merge

documentation

enhancement

type/1.0 cherry-pick

=>

you06 added status/can merge and removed status/can merge labels 18 hours ago

Merge branch 'master' into hotfix25 · Verified ✓ e12efa0

picker-man commented 18 hours ago · Collaborator

/run-all-tests

picker-man commented 18 hours ago · Collaborator

success

picker-man merged commit c811358 into master 18 hours ago · Hide details · Revert

3 checks passed

✓ Travis CI - Branch Build Passed · Details

✓ Travis CI - Pull Request Build Passed · Details

✓ idc-jenkins-ci/build Jenkins job succeeded. · Details

Pull request successfully merged and closed

You're all set—the hotfix25 branch can be safely deleted. · Delete branch

the more

- foresight.

Thank You!

Any Questions ?

关注 PingCAP 官方微信

了解更多技术干货