Treat OSD non-root mounts as non-critical for boot

Status: Resolved
Priority: Normal
Assignee: nick singer
Category: Feature requests
Target version: Ready

Description

Motivation

#162332-7 OSD was encountered as unbootable, possibly due to filesystem problems in the big volumes like vdb,vdc,vdd,vde. We should treat those mount points as non-critical for boot but make critical services like openQA depend on them so that openQA would only start up with a consistent set of mount points. If possible still trigger the automatic filesystem checks but make them non-critical for making the system reachable over network.

Acceptance criteria

- AC1: OSD is reachable over SSH after reboot regardless of problems with vda[a-e]
- AC2: openQA services are prevented from starting if not all relevant mount points are available

Suggestions

- DONE Add "nofail" to all non-root partitions in /etc/fstab
- DONE Add RequiresMountsFor for all relevant high-level services, e.g. openQA, to prevent openQA starting and writing incomplete data
- Ensure that a system like that can boot with all partitions in good state
- Ensure that a system like that can boot at least reachable over ssh with at least one partition in non-good or absent state, e.g. simulate on OSD directly with qemu-system-x86_64 -m 8192 -snapshot -hda /dev/vda -nographic -serial mon:stdio -smp 4

Related issues:

Copied from openQA Infrastructure - action #162353: Ensure consistent known root password on all OSD webUI+workers added
Copied to openQA Infrastructure - action #162359: Change OSD root to more modern filesystem mount options added

History

#1 - 2024-06-17 10:18 - okurz
- Copied from action #162353: Ensure consistent known root password on all OSD webUI+workers added

#2 - 2024-06-17 10:19 - okurz
- Copied to action #162359: Change OSD root to more modern filesystem mount options added

#3 - 2024-06-17 10:22 - okurz
- Assignee set to nick singer
- Priority changed from Normal to High
- Target version changed from Tools - Next to Ready

I assume @nick singer is working on that.

#4 - 2024-06-17 14:10 - okurz
- Due date set to 2024-07-01
- Status changed from New to Feedback

This is how we tested from OSD how the system handles if e.g. additional disks are completely absent

gemu-system-x86_64 -m 8192 -snapshot -hda /dev/vda -nographic -serial mon:stdio -smp 4
I haven't conducted a full test-matrix with different disks missing but can at least ensure that my changes somehow bring up the machine if nothing besides the root disk is present. The biggest other problem seems to be auditd.service which tries to start 30 times and always failing at this step:

which seems to happen because our /var/log is a symlink pointing into /srv/log/. I have to think about a solution or maybe we just keep it because eventually auditd aborts and the system boot continues until login (what we actually want to archive).

Maybe adding a new systemd dependency on /srv/log being mounted for auditd would help?

I've created an MR to adjust auditd. I think we now have the main parts in place to lower the prio as we can now safely restart OSD if any disk fails.

However, we still need to adjust how systemd behaves if this happens. Currently the system starts up fine but only puts some services (nginx, postgres) into their failed state (systemctl --failed). This is confusing and not really helpful for the uninitiated person debugging this. systemd clearly realizes that something went wrong: openqa-webui.service: Job openqa-webui.service/start failed with result 'dependency'. but the service is put into inactive (dead). I failed to understand why this happens and how this behavior can be changed (without blocking the complete bootchain again).
So both ACs are fulfilled as far as we can verify without needing to artificially impact OSD itself. I verified with virtual machines and could see that relevant services are not started and I think the current behaviour of services not being started and ending up "inactive" or "dead" is what we can get but no better. That should suffice. Resolving and assigning back to nicksinger who did most of the relevant work.