openQA Project - action #109864
coordination # 109846 (Blocked): [epic] Ensure all our database tables accommodate enough data, e.g. bigint for ids

Conduct Five Whys for "All Jobs on OSD are incomplete since 2022-04-12" size:M

2022-04-12 13:20 - okurz

<table>
<thead>
<tr>
<th>Status:</th>
<th>Resolved</th>
<th>Start date:</th>
<th>2022-04-12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Priority:</td>
<td>Normal</td>
<td>Due date:</td>
<td></td>
</tr>
<tr>
<td>Assignee:</td>
<td>cdywan</td>
<td>% Done:</td>
<td>0%</td>
</tr>
<tr>
<td>Category:</td>
<td>Organisational</td>
<td>Estimated time:</td>
<td>0.00 hour</td>
</tr>
<tr>
<td>Target version:</td>
<td>Ready</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty:</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Description

See the original ticket

Acceptance criteria

- AC1: A Five-Whys analysis has been conducted and results documented
- AC2: Improvements are planned

Suggestions

- Bring up in retro
- Conduct "Five-Whys" analysis for the topic
- Identify follow-up tasks in tickets
- Organize a call to conduct the 5 whys (not as part of the retro)

Five Whys

1. Why did all jobs fail unexpectedly?
   - Database contained more data than the design anticipated => Research and investigate how we should anticipate exceeding ID ranges, e.g. re-use IDs within openQA before hitting ID limits (additional advantage that user-readable IDs don't get too big but oldest OSD job is [https://openqa.suse.de/tests/385557](https://openqa.suse.de/tests/385557) so we don't have a non-fragmented free range so likely not feasible for jobs, for other tables maybe); Likely not useful, just use bigint everywhere where it matters -> covered in #109846 already

2. Why didn't we see it coming?
   - We don't have alerts based on the amount of id's or rows
   - There is no log message for this => Add alerts for any IDs nearing a limit: Research the industry standard for postgresQL admins best practices

3. Why isn't everything using bigint?
   - We rather not waste space by changing all data types => We should research how much additional space and performance cost bigint would really mean, maybe negligible

4. Why did we need to apply manual repair?
   - Someone had to login via SSH because a PR wouldn't assure us the problem was fixed
   - We don't have a test deployment
   - Deployment itself took a long time
   - We could not foresee how long a migration would take without having our production database as a realistic example

5. Why did it take 6 hours to fix?
   - From the report to acute action time passed
   - The team isn't currently spread across multiple timezones. => This was communicated when the team composition changed and we should remind about that

History

#1 - 2022-04-14 09:56 - cdywan
- Subject changed from Conduct Five Whys for "All Jobs on OSD are incomplete since 2022-04-12" to Conduct Five Whys for "All Jobs on OSD are incomplete since 2022-04-12" size:M
- Description updated
- Status changed from New to Workable

#2 - 2022-04-20 10:24 - cdywan
- Description updated
- Status changed from Workable to Resolved

Conducted just now. Apologies for cutting the process short.

#3 - 2022-04-20 10:25 - okurz
- Assignee set to cdywan

you did good ;)

#5 - 2022-04-28 09:05 - jstehlik

Thank you okurz for spotlighting the importance of covering multiple timezones. I spoke about it with Ralf and we are looking for someone based in China who could reinforce the tools team at least in the error mitigation topic.