Pulp - Task #6856

Document Pulp3 Hardware Requirements recommendations

05/28/2020 09:37 PM - bmbouter

Status:	CLOSED - CURRENTRELEASE	**Start date:**	
Priority:	Normal	**Due date:**	
Assignee:	ttereshc	**% Done:**	100%
Category:		**Estimated time:**	0:00 hour
Sprint/Milestone:	3.6.1		
Platform Release:		**Tags:**	Documentation
Groomed:	Yes	**Sprint:**	Sprint 79
Sprint Candidate:	Yes	**Quarter:**	

Description

This question came up in our channel, we should put this info into the docs in the Architecture and Deploying page in a new section called "Hardware Requirements".

Here's some text that was written in the channel about it:

```
13:26 <cognifloyd> For a single VM install of Pulp 3 (using a django-storages backend for artifact
 storage so that artifacts aren't in the VM) how much CPU/RAM/Disk should I expect to need in that
 VM? There will be yum repos for CentOS 6/7/8 + EPEL 6/7/8, and the pypi index, and a few custom f
ile repos. Are there any rule of thumbs to help me initially size this thing?
13:29 <-- lhc130 (~Sam@2a00:23c7:5187:4f00:4e05:623d:9a92:2739) has quit (Ping timeout: 246 second
s)
13:32 <-- shaunm (~shaunm@2600:2b00:9404:7600:2505:1b5f:29f1:a21a) has quit (Quit: shaunm)
13:32 --> shaunm (~shaunm@2600:2b00:9404:7600:c8ea:c1f6:573a:b20) has joined #pulp
13:39 <-- x9c4 (~mdellweg@dslb-002-202-024-226.002.202.pools.vodafone-ip.de) has quit (Quit: Leavi
ng)
13:43 --> pgagne_ (~textual@cpe-76-182-79-82.nc.res.rr.com) has joined #pulp
13:46 <-- pgagne (~textual@cpe-76-182-79-82.nc.res.rr.com) has quit (Ping timeout: 260 seconds)
13:58 <-- orabin (~orabin@31.210.177.133) has quit (Read error: Connection reset by peer)
14:50 --> pombreda_ (~pombreda@host-78-129-33-156.dynamic.voo.be) has joined #pulp
15:05 <cognifloyd> Next question:
15:08 <cognifloyd> Once I get the basic pulp set up, I'll be looking at building a pulp 3 plugin f
or a file-like artifact I have to deal with that has some annoying encryption requirements. ie The
 artifact should be encrypted in the django-storages backend, and pulp must not have the key to de
crypt it. Clients will be given a key to decrypt those artifacts. Has anything like this been done
? I think a plain file repo would work, but I'm wondering if pulp
15:08 <cognifloyd> d need special support since these would be encrypted.
15:26 <bmbouter> cognifloyd: we don't have have sizing recommendations unfortunately, but I can gi
ve some anecdotal info
15:26 <bmbouter> cpu count should equal the number of pulp workers you start, which allows you to
perform N repository operations concurrently
15:26 <bmbouter> so 2 cpus, you can sync 2 repos concurrently
15:28 <bmbouter> RAM tends to hit it's high watermark during sync and then go back down to nominal
 levels, so for N workers I'd say plan on a gig for each and then maybe 1 gig for postgres as a st
art
15:28 <bmbouter> so for 2 workers, 3 gigs total (2 for sync use, 1 for postgresql)
15:28 <bmbouter> our dev machines typically have 2-4 G and we never oom
15:29 <bmbouter> for disk it's the size of the repos you want all added together. pulp de-duplicat
es content so even as you sync those over time they tend not to grow very muh
15:29 <bmbouter> much
15:29 <bmbouter> I'm not sure what centos6/7/8 + el 6/7/8 is these days but maybe 400G?
15:30 <-- ipanova (~ipanova@ip-86-49-115-30.net.upcbroadband.cz) has quit (Quit: Leaving.)
15:30 <cognifloyd> 400G (ish) for the artifacts or the metadata?
15:30 <bmbouter> in terms of the encryption requirements I think pulp_file would work just fine fo
r you, pulp doesn't need to read/parse the binary data it stores ever, it just needs to calculate
the checksums and it can do that on the encrypted data
15:31 <bmbouter> 400G ish for the artifacts
15:31 <bmbouter> the metadata is very small and lives in the db
```

```
15:31 <cognifloyd> I'm not concerned about the filesize of the artifacts as I'll have them stored
in azure blob storage.
15:31 <bmbouter> oh right
15:31 <bmbouter> you said that
15:31  * cognifloyd would prefer to use GCP, but a client demanded we use azure instead. Bummer
15:31 <cognifloyd> ;)
15:32 <bmbouter> your disk can be small enough to provide working storage during sync prior to blo
bs being placed on the backend, so maybe 50G would do it all
15:32 <bmbouter> pulp verifies checksum data locally and artifacts download/verify in parallel so
50G is probably more than you'll need but it's a bit hard to predict
15:32 <cognifloyd> ah. ok. Thanks for some starting point rules of thumb. I should be able to adju
st from there :)
15:33 <bmbouter> yw, if you can share what you find with use we'd love to hear. also let us know i
f anything could be better or doesn't work
15:33 <cognifloyd> will do
15:34 <cognifloyd> I really like the pulp 3 architecture with versioned repos (an entire repo meta
data rollback sounds awesome). And I hate running Java, so a lot of the other artifact repositorie
s left me with a horrible taste in my mouth. Python is awesome.
```

Associated revisions

Revision 784ad5ac - 08/20/2020 02:42 PM - ttereshc

Added hardware requirements.

closes #6856 https://pulp.plan.io/issues/6856

Revision 29c081ff - 09/01/2020 08:16 PM - ttereshc

Added hardware requirements.

closes #6856 https://pulp.plan.io/issues/6856

History

#1 - 05/28/2020 09:51 PM - daviddavis

- Tags Backlog added

#2 - 05/29/2020 12:21 AM - daviddavis

- Groomed changed from No to Yes

- Sprint Candidate changed from No to Yes

#3 - 05/29/2020 03:45 PM - ggainey

We should go to pulp-list@ and ask for anecdotal evidence from anyone who is currently running Pulp3

#4 - 05/29/2020 03:46 PM - rchan

- Sprint set to Sprint 74

#5 - 05/29/2020 07:22 PM - rchan

An additional way to indicate that these numbers are best guess are to add a note in the documentation to ask for feedback and provide a way to do so.

#6 - 06/02/2020 05:03 PM - ipanova@redhat.com

- Tags deleted (Backlog)

#7 - 06/11/2020 10:27 PM - rchan

- Sprint changed from Sprint 74 to Sprint 75

#8 - 06/26/2020 06:04 PM - rchan

- Sprint changed from Sprint 75 to Sprint 76

#9 - 07/10/2020 08:32 PM - rchan

- Sprint changed from Sprint 76 to Sprint 77

#10 - 07/28/2020 12:06 AM - rchan

- Sprint changed from Sprint 77 to Sprint 78

#11 - 08/07/2020 04:33 PM - rchan

- Sprint changed from Sprint 78 to Sprint 79

#12 - 08/13/2020 06:35 PM - ttereshc

- Status changed from NEW to ASSIGNED

- Assignee set to ttereshc

#13 - 08/13/2020 10:58 PM - pulpbot

- Status changed from ASSIGNED to POST

PR: https://github.com/pulp/pulpcore/pull/854

#14 - 08/20/2020 04:37 PM - ttereshc

- Status changed from POST to MODIFIED

- % Done changed from 0 to 100

Applied in changeset pulpcore|784ad5ac4fa78468558117fd1b4089128d4a61a8.

#15 - 09/01/2020 06:31 PM - pulpbot

PR: https://github.com/pulp/pulpcore/pull/878

#16 - 09/01/2020 09:10 PM - dkliban@redhat.com

- Sprint/Milestone set to 3.6.1

#17 - 09/01/2020 10:52 PM - pulpbot

- Status changed from MODIFIED to CLOSED - CURRENTRELEASE