

Foreman - Feature #22876

Change message digest from SHA1/VARCHAR(40) to XXHASH/BIGINT

03/13/2018 09:03 AM - Lukas Zapletal

Status:	Rejected	Fixed in Releases: Found in Releases: Red Hat JIRA:
Priority:	Normal	
Assignee:	Lukas Zapletal	
Category:	Performance	
Target version:		
Difficulty:		
Triaged:	No	
Bugzilla link:		
Pull request:	https://github.com/foreman/foreman/pull/5319	
Description Using 120bits SHA1 for detecting dupes is ultra-overkill. Simple 64bit number with hash function like CRC64 (https://github.com/postmodern/digest-crc) can do the trick, there will be no collisions for tens or even hundreds of millions of strings imported. But index on a number on 64bit system is much faster than index on VARCHAR, also this will save a lot of memory/space on the SQL server. This would need: <ul style="list-style-type: none">• changing the digest from string to int64• rehashing all entries• code changes• benchmark to verify it performs better (I will setup production instance with real data to get real numbers)		
Related issues: Related to Foreman - Refactor #22875: Limit digest fields to 40 characters		
		Closed 03/13/2018

History

#1 - 03/13/2018 09:03 AM - Lukas Zapletal

- Related to Refactor #22875: Limit digest fields to 40 characters added

#2 - 03/13/2018 02:03 PM - The Foreman Bot

- Status changed from New to Ready For Testing

- Assignee set to Lukas Zapletal

- Pull request <https://github.com/foreman/foreman/pull/5319> added

#3 - 03/13/2018 02:38 PM - Lukas Zapletal

- Subject changed from Change message digest from SHA1/VARCHAR(40) to CRC64/BIGINT to Change message digest from SHA1/VARCHAR(40) to XXHASH/BIGINT

Changed from CRC64 to XXHASH which serves the same purpose, but it's faster and there are many more Ruby implementation than CRC64 (I only found pure Ruby one which was super slow).

#4 - 06/26/2018 09:21 AM - Lukas Zapletal

- Status changed from Ready For Testing to Rejected

- Triaged set to No

Unable to prove it improves the performance.