

Core - Task # 54730

Epic # 55070 (Accepted): Workpackages

Epic # 54260 (Accepted): WP: FAL Missing Issues / Features / API

Story # 54266 (New): As an User I want FAL to be performant

Task # 51094 (Resolved): SQL-Optimize the FAL

Status:	Resolved	Priority:	Should have
Author:	Ingo Schmitt	Category:	File Abstraction Layer (FAL)
Created:	2014-01-03	Assigned To:	Ingo Schmitt
Updated:	2015-01-30	Due date:	
TYPO3 Version:	6.2		
PHP Version:			
Complexity:			
Sprint Focus:	On Location Sprint		
Subject:	sys_file_processedfile.checksum shorten DB field		
Description			
<p>The contents for sys_file_processedfile.checksum are created by \TYPO3\CMS\Core\Resource\Processing\AbstractTask\getConfigurationChecksum by calling \TYPO3\CMS\Core\Utility\GeneralUtility::shortMD5(implode(' ', \$this->getChecksumData())).</p> <p>Since shotMD5 will always return at maximum 10 Characters, the size of the database field could be lowered to 10 characters.</p>			

Associated revisions

Revision fd19c522 - 2015-01-30 15:30 - Mathias Schreiber

[BUGFIX] Reduced sys_file_processedfile.checksum to correct size

Since only a shortMD5 (of 10 characters length) is used in this field
the size is changed to 10 characters and the field type has been set to
char instead of varchar.

Resolves: #54730

Releases: master, 6.2

Change-Id: I8e846786230b55d42464f6ea791202579e6d7873

Reviewed-on: <http://review.typo3.org/36388>

Reviewed-by: Ingo Schmitt <is@marketing-factory.de>

Tested-by: Ingo Schmitt <is@marketing-factory.de>

Reviewed-by: Cedric Ziel <cedric@cedric-ziel.com>

Reviewed-by: Frans Saris <franssaris@gmail.com>

Tested-by: Frans Saris <franssaris@gmail.com>

Reviewed-by: Michael Oehlhof <typo3@oehlhof.de>

Reviewed-by: Mathias Schreiber <mathias.schreiber@wmdb.de>

Tested-by: Mathias Schreiber <mathias.schreiber@wmdb.de>

History

#1 - 2014-01-08 10:10 - Steffen Ritter

- Parent task set to #51094

#2 - 2014-01-08 11:32 - Steffen Ritter

is there any benefit to use a shortened MD5 instead of a real one? would rather opt to change that.

#3 - 2014-01-08 12:32 - Ingo Schmitt

A short MD5 uses a smaller field size, thus index is smaller and faster.

While reading the code, I had a similar thought, normally we can use the full md5, since the calculation is fast and the byte size is not the problem at B-Trees, more the differences of keys.

When using a MD5, the size could be 32 characters, so set size to 32 varchar

#4 - 2014-01-10 23:38 - Christoph Dörfel

Ingo Schmitt wrote:

A short MD5 uses a smaller field size, thus index is smaller and faster.

While reading the code, I had a similar thought, normally we can use the full md5, since the calculation is fast and the byte size is not the problem at B-Trees, more the differences of keys.

When using a MD5, the size could be 32 characters, so set size to 32 varchar

While looking through the code yet another idea came to my mind. Why use sha1, md5 or shortMd5 when the only thing we are looking for is "change", not "uniqueness". The generated hashes are not used as unique identifiers but to make sure that data is valid and doesn't have to be recalculated. So we could just use a simple checksum, crc32 in this case. crc32 is used in a lot of cases, where e.g. file transfers have to be checked for errors.

The benefit of using crc32 is that it's a simple 32 bit integer. Comparisons and DB searches can't get faster than that :)

Your opinions?

Edit:

Also in <http://forge.typo3.org/issues/54729> with file checksums, "originalfilessha1" and similar.

#5 - 2014-01-11 10:19 - Ingo Schmitt

If this field is used only to detect the "change", then we should name it accordingly. By looking at the field from outside right now, it seems that a file hash is stored. So for an extension developer this field could be used for detecting duplicates or similar.

@Steffen: What do you think about it?

#6 - 2015-01-02 11:04 - Mathias Schreiber

- Status changed from New to Accepted
- Target version changed from 6.2.0 to 7.1 (Cleanup)
- Sprint Focus set to On Location Sprint

#7 - 2015-01-12 09:50 - Mathias Schreiber

- Category changed from Performance to File Abstraction Layer (FAL)

#8 - 2015-01-29 22:25 - Gerrit Code Review

- Status changed from Accepted to Under Review

Patch set 1 for branch **master** of project **Packages/TYPO3.CMS** has been pushed to the review server.

It is available at <http://review.typo3.org/36388>

#9 - 2015-01-30 15:46 - Gerrit Code Review

Patch set 1 for branch **TYPO3_6-2** of project **Packages/TYPO3.CMS** has been pushed to the review server.

It is available at <http://review.typo3.org/36433>

#10 - 2015-01-30 16:00 - Mathias Schreiber

- *Status changed from Under Review to Resolved*

- *% Done changed from 0 to 100*

Applied in changeset commit:fd19c5221722c4ed56a4b67b149c7d049e2d1189.