Arvados - Bug #16425

[keepstore] should not scan directories it doesn't write in

05/12/2020 02:17 PM - Ward Vandewege

Status:	Resolved		Start date:	06/12/2020
Priority:	Normal		Due date:	
Assigned To:	Tom Clegg		% Done:	100%
Category:			Estimated time:	0.00 hour
Target version:	2020-06-17 Sprint			

Description

Keepstore walks all directories in its mount points, even those it never writes in. This can lead to a situation where keepstore does tons of unnecessary IO (e.g. when .snapshot directories are present, where GPFS makes daily snapshots).

Adding an option to exclude certain directories from what Keep considers would be a good solution.

Subtasks:

Task # 16487: Review 16425-skip-unused-dirs	**Resolved**

Associated revisions

Revision 2cae80e7 - 06/15/2020 03:17 PM - Tom Clegg

Merge branch '16425-skip-unused-dirs'

fixes #16425

Arvados-DCO-1.1-Signed-off-by: Tom Clegg <tom@tomclegg.ca>

History

#1 - 05/12/2020 02:25 PM - Ward Vandewege

- Description updated

#2 - 05/12/2020 02:26 PM - Ward Vandewege

- Description updated

#3 - 05/20/2020 08:34 PM - Tom Clegg

When indexing, it looks like we're already skipping dirs that don't conform to keepstore's storage layout:
source:services/keepstore/unix_volume.go#L355

```
var blockDirRe = regexp.MustCompile(`^[0-9a-f]+$`)
var blockFileRe = regexp.MustCompile(`^[0-9a-f]{32}$`)
...
            if !blockDirRe.MatchString(names[0]) {
                    continue
            }
            ...
            blockdir, err := v.os.Open(blockdirpath)
```

However, the "empty trash" goroutine does walk the entire tree.

```
    err := filepath.Walk(v.Root, func(path string, info os.FileInfo, err error) error {
            if err != nil {
                    v.logger.WithError(err).Errorf("EmptyTrash: filepath.Walk(%q) failed", path)
                    return nil
            }
            todo <- dirent{path, info}
            return nil
    })
```

That walk func should check dirnames and return filepath.SkipDir when appropriate.

#4 - 06/03/2020 04:02 PM - Peter Amstutz

- Target version changed from To Be Groomed to 2020-06-17 Sprint

- Assigned To set to Tom Clegg

#5 - 06/12/2020 08:30 PM - Tom Clegg

16425-skip-unused-dirs @ [b257b5e4fca36172a79db2fd2cf8ab843cf50d1c](#) -- [https://ci.arvados.org/view/Developer/job/developer-run-tests/1915/](#)

#6 - 06/12/2020 08:30 PM - Tom Clegg

- Status changed from New to In Progress

#7 - 06/12/2020 09:00 PM - Ward Vandewege

Tom Clegg wrote:

> 16425-skip-unused-dirs @ [b257b5e4fca36172a79db2fd2cf8ab843cf50d1c](#) -- [https://ci.arvados.org/view/Developer/job/developer-run-tests/1915/](#)

LGTM, thanks!

#8 - 06/15/2020 04:25 PM - Anonymous

- % Done changed from 0 to 100

- Status changed from In Progress to Resolved

Applied in changeset [arvados|2cae80e7746d8fa32df1f819a4a41d29cdc3577a](#).

#9 - 10/07/2020 02:11 AM - Peter Amstutz

- Release set to 25