DE LA RECHERCHE À L’INDUSTRIE

RobinHood v4 progress report

September 27th 2022
Sebastien Gougeaud
No news, good news?

- The last update at LAD was two years ago
- RobinHood v4 is still in development by the CEA and former members of the team
- Developments can be followed on gerrithub and github (https://github.com/cea-hpc/librobinhood)
- Need for having it soon in production on our clusters
Focus of RobinHood

Providing efficient and easy to use means to replicate and query any filesystem’s metadata
## Why a version 4 for RobinHood?

<table>
<thead>
<tr>
<th></th>
<th>version 3</th>
<th>version 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>scale up</td>
<td>SQL paradigm</td>
<td>NoSQL paradigm</td>
</tr>
<tr>
<td></td>
<td><em>MariaDB</em></td>
<td><em>MongoDB</em></td>
</tr>
<tr>
<td>code genericity</td>
<td>software specialised for Lustre filesystems</td>
<td>generic tools calling specific backends</td>
</tr>
<tr>
<td>inclusion to Linux repositories</td>
<td>expert system</td>
<td>library of features/applications</td>
</tr>
<tr>
<td>code refactoring</td>
<td>heavy code caused by Lustre behavior evolution</td>
<td>clean design to better correspond to current filesystems</td>
</tr>
</tbody>
</table>
RobinHood components – ready to use

RobinHood v4 progress report

September 27th 2022
RobinHood components – in development

- posix
- lustre
- mongo
- rbh-lfind
- rbh-find
- rbh-fsevents
- rbh-sync

librobinhood

BACKEND

mongo

lustre

posix
RobinHood components – future sight

rbh-find

rbh-lfind

rbh-fsevents

rbh-sync

rbh-report

posix

lustre

librobinhood

mongo
Support of custom extended attributes
- Attach key-value attributes to entries
- For instance: the path, the name, Lustre attributes

Conversion between RBH FSEvents and YAML
- FSEvent stands for File System Event: file creation, metadata change, etc.
- Handling a new backend only need to convert specific log events into FSEvents
Additions to rbh-find

- Filter options are added as long as they may be needed to fetch metadata or define data placement policies:
  - Size – -size
  - Permissions – -perm
  - User-defined extended attributes – -xattr

- Entries may be sorted using -sort/-rsort options

- Actions on entries are added:
  - Listing – -ls
  - Printing to files – -fprint/-fprint0/-fls
Additions to rbh-find – demonstration

```bash
# rbh -sync rbh:posix:/mnt/lustre rbh:mongo:test
# rbh-find rbh:mongo:test -type f -sort name
# rbh-find rbh:mongo:test -type f -sort size
# rbh-find rbh:mongo:test -type f -ls
# rbh-find rbh:mongo:test -type f -rsort size -ls
# rbh-find rbh:mongo:test -type f -sort size -ls
# rbh-find rbh:mongo:test -type f -sort name -ls
```

RobinHood v4 progress report September 27th 2022
Lustre backend

- Lustre is the file system used in our clusters
- Storing lustre relevant information in the mirror backend, such as:
  - FID
  - File layout
  - OST
  - HSM state
- Those information are retrieved during the synchronization and converted to namespace attributes of the entry stored in the database backend
- Filtering entries using those information
- Overload of rbh-find tool to add lustre-related options
Lustre backend – demonstration (1/2)

1. # mkdir /mnt/lustre/dir-{0..15}
2. # touch /mnt/lustre/dir-{0..15}/file-{0..15}
3. # for i in {0..5}; do
   4.   lfs hsm_set --archived /mnt/lustre/dir-${i}/file-12
   5.   lfs hsm_set --norelease /mnt/lustre/dir-${i}/file-11
   6. done
7. # rbh-sync rbh:lustre:/mnt/lustre rbh:mongo:test
8. # rbh-find rbh:mongo:test | head
   9. /dir-14/file-13
   10. /dir-14/file-14
   11. /dir-14/file-15
   12. /dir-0
   13. /dir-15/file-0
   14. /dir-1
   15. /dir-15/file-1
   16. /dir-2
   17. /dir-15/file-2
   18. /dir-3
9. # mongo --quiet --eval "db.entries.find({"statx.type": 32768})" | head -n 1
10. { "id" : BinData(0,"lwA...AAA"), "statx" : { "blksize" : 4194304, "nlink" : 1, "uid" : 0, "gid" : 0, "type" : 32768, "mode" : 420, "ino" : NumberLong("14411520527502451"), "size" : NumberLong(0), "blocks" : NumberLong(0), "attributes" : { "immutable" : false, "append" : false }, "atime" : { "sec" : NumberLong(1663145146), "nsec" : 0 }, "ctime" : { "sec" : NumberLong(1663145146), "nsec" : 0 }, "mtime" : { "sec" : NumberLong(1663145146), "nsec" : 0 }, "rdev" : { "major" : 0, "minor" : 0 }, "dev" : { "major" : 1273, "minor" : 181606 } }, "xattrs" : { "lustre" : { "lov" : BinData(0,"0Av...AA=") }, "trusted" : { "link" : BinData(0,"3/H...S0w"), "lma" : BinData(0,"AAA...AAA"), "lov" : BinData(0,"0Av...AAA"), "som" : BinData(0,"BAA...AAA") } }, "ns" : [ { "parent" : BinData(0,"lwA...AAA"), "name" : "file-0", "xattrs" : "path" : "/dir-10/file-0", "fid" : BinData(0,"AQ...A="), "hsm_state" : 0, "hsm_archive_id" : 0, "flags" : 0, "magic" : "LOV_USER_MAGIC_V1", "gen" : 0, "stripe_count" : [], "stripe_size" : [], "pattern" : [], "comp_flags" : [], "pool" : [], "ost" : [ NumberLong(0), "mdt_index" : 0 ] } ]}
Lustre backend – demonstration (2/2)

```bash
# rbh-lfind rbh:mongo:test -sort name -ost 0 | head
/dir-15/file-0
/dir-0/file-0
/dir-1/file-0
/dir-2/file-0
/dir-3/file-0
/dir-4/file-0
/dir-5/file-0
/dir-6/file-0
/dir-7/file-0
/dir-8/file-0

# lfs path2fid /mnt/lustre/dir-0/file-0
[0x200000401:0x13:0x0]

# rbh-lfind rbh:mongo:test -fid [0x200000401:0x13:0x0] | head
/dir-0/file-0

# rbh-lfind rbh:mongo:test -hsm-state archived | head
/dir-0/file-12
/dir-1/file-12
/dir-2/file-12
/dir-3/file-12
/dir-4/file-12
/dir-5/file-12

# rbh-lfind rbh:mongo:test -hsm-state norelease | head
/dir-0/file-11
/dir-1/file-11
/dir-2/file-11
/dir-3/file-11
/dir-4/file-11
/dir-5/file-11
```

RobinHood v4 progress report

September 27th 2022

13/18
Lustre changelog reader

First events are processed:
- CREAT
- ATIME/CTIME/MTIME
- CLOSE

Others are following

Still thinking about the right moment to acknowledge the events:
- When the changelog events are converted to FSEvents?
- When the events are recorded in the backend?
# Lustre changelog reader – demonstration

1. # touch /mnt/lustre/old
2. # lctl --device lustre-MDT0000 changelog_register
3. # touch /mnt/lustre/new

1. 18CTIME 10:45:21.574804672 2022.09.13 0x5

2. t=[0x200000401:0x12:0x0] j=touch.0 ef=0xf u=0:0 nid=0@lo
3. 2 ICLOSE 10:45:21.588901555 2022.09.13 0x42

4. t=[0x200000401:0x12:0x0] j=touch.0 ef=0xf u=0:0 nid=0@lo
5. 3 ICREAT 10:45:25.915058526 2022.09.13 0x0

6. t=[0x200000401:0x13:0x0] j=touch.0 ef=0xf u=0:0 nid=0@lo
7. p=[0x200000007:0x1:0x0] new

8. t=[0x200000401:0x13:0x0] j=touch.0 ef=0xf u=0:0 nid=0@lo
9. 4 ICLOSE 10:45:25.922082667 2022.09.13 0x42

10. # rbh -fsevents --enrich /mnt/lustre lustre-MDT0000 -

   --- !upsert
   "id": "!binary lwAAAAEAAAAACAAAAEwAAAAAAAAAAAAAAAAAAAAAA
   "xattrs":
   "ns":
   - "parent": "!binary lwAAAAcAAAAAAAAAAAAQQQAAAAAAAAAAAAAA
   "name": "new"
   "xattrs":
   "path": "new"
   "id": "!binary AQQQAAATAAAAAAAAAAAAAA=
   "statx":
   "type": "file"
   "uid": 0
   "gid": 0
   "atime":
   "sec": 1663065925
   "nsec": 0
   
   --- !upsert
   "id": "!binary lwAAAAEAAAAACAAAAEwAAAAAAAAAAAAAAAAAAAAAA
   "xattrs":
   "ns":
   "parent": "!binary lwAAAAcAAAAAAAAAAAAQQQAAAAAAAAAAAAAA
   "name": "new"
   "xattrs":
   "path": "new"
   "id": "!binary AQQQAAATAAAAAAAAAAAAAA=
   "statx":
   "type": "file"
   "uid": 0
   "gid": 0
   "atime":
   "sec": 1663065925
   "nsec": 0
   
   --- !upsert
   "id": "!binary lwAAAAEAAAAACAAAAEwAAAAAAAAAAAAAAAAAAAAAA
   "xattrs":
   "ns":
   "parent": "!binary lwAAAAcAAAAAAAAAAAAQQQAAAAAAAAAAAAAA
   "name": "new"
   "xattrs":
   "path": "new"
   "id": "!binary AQQQAAATAAAAAAAAAAAAAA=
   "statx":
   "type": "file"
   "uid": 0
   "gid": 0
   "atime":
   "sec": 1663065925
   "nsec": 0
   
   --- !upsert

RobinHood v4 progress report

September 27th 2022
What’s next?

- Development of a non-lustre related backend, for the IO-SEA European project, using CORTX-Motr and Phobos as object store
- FSEvent deduplication
- Pre-production tests on our systems for a production during 2023
## Repositories

<table>
<thead>
<tr>
<th>project</th>
<th>repository</th>
</tr>
</thead>
<tbody>
<tr>
<td>librobinhood</td>
<td><a href="https://github.com/cea-hpc/">https://github.com/cea-hpc/</a></td>
</tr>
<tr>
<td>rbh-find</td>
<td></td>
</tr>
<tr>
<td>rbh-find-lustre</td>
<td></td>
</tr>
<tr>
<td>rbh-sync</td>
<td></td>
</tr>
<tr>
<td>rbh-fsevents</td>
<td></td>
</tr>
</tbody>
</table>
Thanks for your attention

RobinHood v4 progress report
Sebastien Gougeaud