EOS migration tools
motivated by AMS experiment migration

Dr. Jaroslav Guenther
on behalf of EOS PDS team
(CERN IT-ST-PDS)
Objective and Tools

EOS space migration:

- data: directories, files, links
- metadata: permissions, attributes
- instance and space settings (quotas, access rights etc.)
- sync, freeze source and redirect all clients to use new location

Low level tools:

- eos cp, xrdcp, rsync, fusex, eos-ns-inspect
- Incomplete/no attribute or permission transfer
- requires special care per migration item type
- need for higher level tool automating the migration process
Challenges

Procedure pitfalls:

- very diverse file occupancy among directories
  (e.g. one can find $O(1M)$ very small files in 1 dir)
- user permission inconsistencies (posix/ACL)
  (users not allowed to browse or write to the directory where they own files)
- interference with durability tools (fsck)
  (problems during migration process may be difficult to spot)
- random data/metadata corruption at source and destination
  (HW/SW problems)
- human factor

  Do you really want to run this from master headnode?

Some of the automation pitfalls:

- scale (e.g. metadata volume to keep in sync)
- sync time (in case of no source freeze)
  (AMS 60M files, 400k directories, 13PB)
- problem catching, reporting and fixing
  (many migration jobs)
- low level tool issues
  (e.g. string byte decoding in python)
Existing tools

Per username migration tools:

- repository
- migration item = user
- `/eos/user/<letter>/<username>`
- no long term automatic sync functionality
- philosophy: migrate and switch “in one go”

General migration tools:

- different approaches being developed
- `eos-instance-migration tool (docs)`
  * migrating any path/dir/file/link/quota
  * philosophy:
    → migrate
    → keep in sync
    → switch whenever convenient
- EOS fusex could provide a general solution as well
  (extended attribute support in development)
New migration tool

**Migration steps:**
1. migrate all directories
2. migrate eos quota settings
3. directory sync and file sync
4. automatic sync replays

**Tool Configuration:**
- run options (emailing, file names)
- access to EOS instances
- attribute check settings
- dir/file/link removal behaviour
- quota node list
- sync optimisation options
- Thread settings
- Timeouts
- check the docs

---

**EOS instance migration tool**

Check the documentation for more info.

**Dependencies**
- python (v3.7+)
- eos-ns-inspect tool (part of EOS v4.8.57 and later only is optimised for this use case)
- optional: pip3 install iljson

**Repository structure**
- docs: content for documentation for more info.
- code:
  - migration_daemon.py: main executable
  - migrate.py: prepare and run migration tasks
  - namespace.py: query namespaces
  - xcheck.py: migration item cross checking functionality between source and destination
  - transfer.py: data transfer functionality (for directories/files/links)
  - utils.py: utility functions
  - quota.py: quota migration methods
  - decode_json_lines.py: if namespace output requested as json, this module will be called, requires pip3 install iljson
  (allowing to read big json line by line)
  - accounting.py: general accounting functionality, consolidation of job reports
  - logsetup.py: helper methods for logging
  - config.py: configure the migration tool (expected to be modified according to use-case)
- tests: testing scripts (unit tests supposed to come here)
Functionality

Operator Input:
- configuration file
- root directory to migrate
- migration script output path

Migration Daemon:
1. xcheck namespaces
2. remove/fix @ dest
3. bunch per process
4. transfers via threads
5. replay [1 - 5]

Error reporting:
- in categorised files
  1. namespace check and filter categories
  2. transfer crash categories
- check the docs
Directory migration (hands-on)

- Migrating from EOSPILOT to EOSAMS02
- Instances setup for migration (TPC, keytab files, access)
- Directory order is important - processing sequential processing for the moment
- 1 process for namespace xcheck and 1 for directory processing

```
[root@eosams02-ip554 (mgm:slave mq:slave) migtest]$ workdir=/var/migtest/
[root@eosams02-ip554 (mgm:slave mq:slave) migtest]$ codedir=/var/migtest/code/
[root@eosams02-ip554 (mgm:slave mq:slave) migtest]$ rundir=/var/migtest/run/
[root@eosams02-ip554 (mgm:slave mq:slave) migtest]$ migrate=$codedir/eos-instance-migration/code/migration_daemon.py
[root@eosams02-ip554 (mgm:slave mq:slave) migtest]$ cd run

[root@eosams02-ip554 (mgm:slave mq:slave) run]$ rootpth=/eos/pilot/cms/tier0/store/backfill/1/data/;
[root@eosams02-ip554 (mgm:slave mq:slave) run]$ python3.7 $migrate --root-nsdir $rootpth --sync-dirs-only --output-path $rundir > $rundir/ams.log 2>&1 &

[root@eosams02-ip554 (mgm:slave mq:slave) run]$ less ams.log
```

```
1 directories were queued for processing, from these:
- 0 initially matched the source (no jobs ran for them)
- 1 were (at least partially) migrated
- 1 were completely migrated
0 directories have a timestamp of complete "file" migration

Subdirectories of the directories above were migrated, their counts were:
- 244 queued for processing
- 244 migrated successfully
- 244 part of completely migrated directories
- 0 crashed
- 0 left to be processed (from the queued ones at the moment)
```

```
[root@eosams02-ip554 (mgm:slave mq:slave) code]$ ll ../*run/
total 232
-rw-r--r--. 1 root root 2227 Mar  6 05:59 ams.log
-drwxr-xr-x. 6 root root  4096 Mar  6 04:48 dir_bunches
-rw-r--r--. 1 root root  3400 Mar  6 04:48 DIR_migration_summary_report.json
-drwxr-xr-x. 2 root root  4096 Mar  6 03:56 dir_reports
-rw-r--r--. 1 root root    2 Mar  6 03:55 dir_sync_timestamps.json
-rw-r--r--. 1 root root 216542 Mar  6 04:48 file_processing_dirs.json
```
Directory migration (hands-on)

```
[root@eosams02-ip554 ~]# eos ls -l /eos/pilot/cms/tier0/store/backfill1/data/tier0_REPLAY_2021/HLTPhysics

drwxr-x--- 1 cmst0 zh 0 Mar 6 04:46 ALCARECO
drwxr-x--- 1 cmst0 zh 0 Mar 6 04:46 AOD

drwxr-x--- 1 cmst0 zh 0 Mar 6 04:46 DQM10

drwxr-x--- 1 cmst0 zh 0 Mar 6 04:46 MINIAOD

drwxr-x--- 1 cmst0 zh 0 Mar 6 04:46 RAW

drwxr-x--- 1 cmst0 zh 0 Mar 6 04:46 RAW-RECO

drwxr-x--- 1 cmst0 zh 0 Mar 6 04:46 USER
```

```
[root@eospilot-ns-ip700 ~]# eos ls -lh /eos/pilot/cms/tier0/store/backfill1/data/tier0_REPLAY_2021/HLTPhysics

drwxr-x--- 1 cmst0 zh 7.03 G Nov 22 16:15 ALCARECO

drwxr-x--- 1 cmst0 zh 8.55 G Nov 22 16:26 AOD

drwxr-x--- 1 cmst0 zh 1.42 G Nov 22 16:15 DQM10

drwxr-x--- 1 cmst0 zh 1.93 G Nov 22 16:29 MINIAOD

drwxr-x--- 1 cmst0 zh 109.39 G Nov 16 12:35 RAW

drwxr-x--- 1 cmst0 zh 35.56 G Nov 16 14:44 RAW-RECO

drwxr-x--- 1 cmst0 zh 44.89 M Nov 22 16:07 USER
```

```
[root@eosams02-ip554 ~]# eos attr ls /eos/pilot/cms/tier0/store/backfill1/data/tier0_REPLAY_2021/HLTPhysics

sys.aci="g!1399:rx,u:103931:rxw,u:8959:rw,u:22014:rw"

drs.esi=bitme=1646358325.27947950

sys.forced.atomic=1

sys.forced.blocksize=4K

sys.forced.checksum="adler"

sys.forced.layout="replica"

sys.forced.nstripes=2

sys.forced.space=delete

sys.lru.expire.match=*,root:105d
```

```
[root@eospilot-ns-ip700 ~]# eos attr ls /eos/pilot/cms/tier0/store/backfill1/data/tier0_REPLAY_2021/HLTPhysics

sys.aci="g!1399:rx,u:103931:rxw,u:8959:rw,u:22014:rw"

drs.esi=bitme=1637062517.44869590

sys.forced.atomic=1

sys.forced.blocksize=4K

sys.forced.checksum=""adler"

sys.forced.layout=""replica"

sys.forced.nstripes=2

sys.forced.space=""delete"

sys.lru.expire.match=*,root:105d
```

---

Initial namespace exploration found the following items:
- 244 files [desc] totals seen:
  - 0 [0] [desc] items from directories exclusively explained by operator
  - 0 [0] [desc] failed prefiltering (see below [1])
    - 0 [0] failed prefiltering and were removed
      - not on storage, ignored, decode errors, zero size/len corrupted
      - 0 [0] remained failed (removal omissions)
    - 244 [0] [desc] items passed the initial filter (see below [1])
      - 0 [0] [desc] items with matching AT charm in the source and destination
      - 0 [0] [desc] had bad metadata (attribute, ownership, mode perm) crosscheck size and checksum were OK.
      - 0 [0] removed due to bad size or len
      - 0 [0] remained bad (removal omissions)
      - 0 [0] not present at source:
      - 0 [0] removed
    - 0 [0] [desc] were selected for processing input

[1] Failed cases at source:
  - ignored: 0
    - decode errors: 0
    - zero size/len corrupted: 0
    - no storage only in namespace: 0

Failed cases at destination:
  - ignored: 0
    - decode errors: 0
    - zero size/len corrupted: 0
    - no storage only in namespace: 0

End of migration summary.
Quota migration (hands-on)

```bash
[root@eosams02-ip554 (mgm:slave mq:slave) run] $ python3.7 ${migrate} --root-nsdir $rootpth --set-quota --output-path ${rundir}
> ${rundir}/ams.log 2>&1 &

[root@eospilot-ns-ip700 (mgm:master mq:slave) ~]$ eos quota ls /eos/pilot/cms/tier0/

> Quota Node: /eos/pilot/cms/tier0/

<table>
<thead>
<tr>
<th>user</th>
<th>used bytes</th>
<th>logi bytes</th>
<th>used files</th>
<th>aval bytes</th>
<th>aval logib</th>
<th>aval files</th>
<th>filled[%]</th>
<th>vol-status</th>
<th>ino-status</th>
</tr>
</thead>
<tbody>
<tr>
<td>cmst0</td>
<td>111.16 TB</td>
<td>55.58 TB</td>
<td>233.20 K</td>
<td>0 B</td>
<td>0 B</td>
<td>0</td>
<td>100.00 %</td>
<td>ignored</td>
<td>ignored</td>
</tr>
<tr>
<td>nobody</td>
<td>0 B</td>
<td>0 B</td>
<td>0</td>
<td>0 B</td>
<td>0 B</td>
<td>0</td>
<td>100.00 %</td>
<td>ignored</td>
<td>ignored</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>group</th>
<th>used bytes</th>
<th>logi bytes</th>
<th>used files</th>
<th>aval bytes</th>
<th>aval logib</th>
<th>aval files</th>
<th>filled[%]</th>
<th>vol-status</th>
<th>ino-status</th>
</tr>
</thead>
<tbody>
<tr>
<td>nobody</td>
<td>0 B</td>
<td>0 B</td>
<td>0</td>
<td>0 B</td>
<td>0 B</td>
<td>0</td>
<td>100.00 %</td>
<td>ignored</td>
<td>ignored</td>
</tr>
<tr>
<td>zh</td>
<td>111.16 TB</td>
<td>55.58 TB</td>
<td>233.20 K</td>
<td>300.00 TB</td>
<td>150.00 TB</td>
<td>10.00 M</td>
<td>37.05 %</td>
<td>ok</td>
<td>ok</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>summary</th>
<th>used bytes</th>
<th>logi bytes</th>
<th>used files</th>
<th>aval bytes</th>
<th>aval logib</th>
<th>aval files</th>
<th>filled[%]</th>
<th>vol-status</th>
<th>ino-status</th>
</tr>
</thead>
<tbody>
<tr>
<td>All users</td>
<td>111.16 TB</td>
<td>55.58 TB</td>
<td>233.20 K</td>
<td>0 B</td>
<td>0 B</td>
<td>0</td>
<td>100.00 %</td>
<td>ignored</td>
<td>ignored</td>
</tr>
<tr>
<td>All groups</td>
<td>111.16 TB</td>
<td>55.58 TB</td>
<td>233.20 K</td>
<td>300.00 TB</td>
<td>150.00 TB</td>
<td>10.00 M</td>
<td>37.05 %</td>
<td>ok</td>
<td>ok</td>
</tr>
</tbody>
</table>

[root@eospilot-ns-ip700 (mgm:master mq:slave) ~]$ eos quota ls /eos/pilot/cms/tier0/

> Quota Node: /eos/pilot/cms/tier0/

<table>
<thead>
<tr>
<th>group</th>
<th>used bytes</th>
<th>logi bytes</th>
<th>used files</th>
<th>aval bytes</th>
<th>aval logib</th>
<th>aval files</th>
<th>filled[%]</th>
<th>vol-status</th>
<th>ino-status</th>
</tr>
</thead>
<tbody>
<tr>
<td>nobody</td>
<td>0 B</td>
<td>0 B</td>
<td>0</td>
<td>0 R</td>
<td>0 R</td>
<td>0</td>
<td>100.00 %</td>
<td>ignored</td>
<td>ignored</td>
</tr>
<tr>
<td>zh</td>
<td>0 B</td>
<td>0 B</td>
<td>0</td>
<td>300.00 TB</td>
<td>150.00 TB</td>
<td>10.00 M</td>
<td>0.00 %</td>
<td>ok</td>
<td>ok</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>summary</th>
<th>used bytes</th>
<th>logi bytes</th>
<th>used files</th>
<th>aval bytes</th>
<th>aval logib</th>
<th>aval files</th>
<th>filled[%]</th>
<th>vol-status</th>
<th>ino-status</th>
</tr>
</thead>
<tbody>
<tr>
<td>All users</td>
<td>0 B</td>
<td>0 B</td>
<td>0</td>
<td>0 B</td>
<td>0 B</td>
<td>0</td>
<td>100.00 %</td>
<td>ignored</td>
<td>ignored</td>
</tr>
<tr>
<td>All groups</td>
<td>0 B</td>
<td>0 B</td>
<td>0</td>
<td>300.00 TB</td>
<td>150.00 TB</td>
<td>10.00 M</td>
<td>0.00 %</td>
<td>ok</td>
<td>ok</td>
</tr>
</tbody>
</table>
File migration

- directory loop as before to sync with source if changes were made
- prepares list of dirs for file processing
  (to be used if --limit-depth is used, otherwise namespace queried recursively with top dir)

- Tool configured with:
  - 1 process running namespace exploration with
    * 5+5 threads parallel queries to src and dest
    * 5 threads for xchecks
  - 5 migration processes with 10 threads each (50 transfers in parallel)

```
[root@eosams02-ip554 (mgmslave mgmslave) run]$ python3.7 $(migrate) --root-nsdir $rootpth --limit-depth --max-job-files 20 --max-retries 1 --output-path ${rundir} > ${rundir}/ams.log 2>&1 &
watch tail -n 40 ${rundir}/ams.log
```

```bash


```


```
```
File migration

> namespace file entry xcheck → entries to process bunched to files
> main process picks up prepared bunches and dispatches them
Sync replays (hands-on)

- replays `<n>` to run `<n>` sync loops
- timestamps after full directory migration recorded to json
- directories changes at later date are taken for md sync and/or file migration

```bash
[root@eosams02-ip554 (mgm:slave mq:slave) run]$ python3.7 $(migrate) --root-nsdir $rootptr --replays 2 --only-dir-changes --max-job-files 20 --max-retries 1 --output-path $(rundir) > $(rundir)/ams.log 2>&1

watch tail -n 40 $(rundir)/ams.log

[INFO] 2022-03-06 14:13:38,916 [migration_daemon.py:528] 122856 (MainThread) Starting migration replay 0
[INFO] 2022-03-06 14:13:39,188 [migrate.py:420] 122867 (MainThread) DIR: PID 0 RETRY 0: 0 items skipped
[INFO] 2022-03-06 14:13:39,568 [migration_daemon.py:412] 122856 (MainThread) MP: Finished all processing jobs, saving stat output and exiting unless more migration replays will be run.
[INFO] 2022-03-06 14:16:05,138 [migration_daemon.py:528] 122856 (MainThread) Starting migration replay 1
[INFO] 2022-03-06 14:16:05,213 [migration_daemon.py:227] 122856 (MainThread) MP DIR: queued namespace MD sync and item bunching job (id 0).
[INFO] 2022-03-06 14:16:05,384 [migrate.py:416] 135763 (MainThread) DIR: PID 0 RETRY 0: Accounting total items to process after prefilter:0 processing.
[INFO] 2022-03-06 14:16:05,384 [migrate.py:420] 135763 (MainThread) DIR: PID 0 RETRY 0: 43 items skipped
[INFO] 2022-03-06 14:16:05,384 [migrate.py:440] 135763 (MainThread) DIR: PID 0 RETRY 0: 7 directories collected for file migration
[INFO] 2022-03-06 14:16:07,016 [migrate.py:200] 135808 (MainThread) FILE: PID 0 RETRY 0: Preparing list of entries for 5 paths ['/eos/pil/t/cms/tier0/store/backfill/1/data/Tier0_REPLAY_2021/HLTHPhysics/ALCARECO/TkAlMinBias-PromptReco-v1/000/344/063/00000/', ...]...
[INFO] 2022-03-06 14:16:08,026 [migrate.py:416] 135808 (MainThread) FILE: PID 0 RETRY 0: Accounting total items to process after prefilter:0
[INFO] 2022-03-06 14:16:08,026 [migrate.py:475] 135808 (MainThread) FILE: PID 0 RETRY 0: Finished xchecks and filtering out entries which exist at source and destination.
[root@eosams02-ip554 (mgm:slave mq:slave) run]$ wc -l dir_sync_timestamps.json 44 dir_sync_timestamps.json
```
Migration done (hands-on)

```
[root@eosams02-ip554 (mgm:slave mq:slave) ~]$ ll /var/migtest/run/
total 356
-rw-------. 1 root root 314214 Mar 6 14:16 ans.log
drwxr-xr-x. 2 root root 4096 Mar 6 14:16 dir_bunches
drwxr-xr-x. 2 root root 4096 Mar 6 14:16 dir_reports
drwxr-xr-x. 2 root root 5572 Mar 6 14:16 dir_sync_timestamps.json
drwxr-xr-x. 2 root root 4096 Mar 6 14:16 file_bunches
-rw-------. 1 root root 10596 Mar 6 14:16 FILE_migration_summary_report.json
drwxr-xr-x. 2 root root 5978 Mar 6 14:16 file_processing_dirs.json
drwxr-xr-x. 2 root root 4096 Mar 6 14:14 file_reports
```

43 directories were queued for processing, from these:
- 36 initially matched the source
- 7 were at least partially migrated
- 7 were completely migrated
36 directories have a timestamp of complete "file" migration

Files of the directories above were migrated, their counts were:
- 105 queued for processing
- 105 migrated successfully
- 0 crashed
- 0 left to be processed

```
43 directories were queued for processing, from these:
- 7 initially matched the source
- 0 were at least partially migrated
- 0 were completely migrated
43 directories have a timestamp of complete "file" migration
```

Files of the directories above were migrated, their counts were:
- 0 queued for processing
- 0 migrated successfully
- 0 part of completely migrated directories
- 0 crashed
- 0 left to be processed (from the queued ones at the moment)
Migration done (hands-on)

```
[root@eosans02-ip554 (msg:slave msg:slave) ~]$ eos ls -lh /eos/pilot/cms/tier0/store/backfill/1/data/Tier0_REPLAY_2021/HLTPhysics/
-dwax--- 1 cmst0  zh  7.03 M Mar  6 13:31:1 ALCARECO
dwax--- 1 cmst0  zh  8.55 G Mar  6 13:31:1 AOD
dwax--- 1 cmst0  zh  1.42 G Mar  6 13:31:1 DOMIO
-dwax--- 1 cmst0  zh  1.93 G Mar  6 13:31:1 MINIAOD
-dwax--- 1 cmst0  zh  109.39 G Mar  6 13:31:1 RAW
-dwax--- 1 cmst0  zh  35.56 G Mar  6 13:31:1 RAW-RECO
-dwax--- 1 cmst0  zh  44.09 M Mar  6 13:31:1 USER

[root@eospilot-ns-ip700 (msg:master msg:slave) ~]$ eos ls -lh /eos/pilot/cms/tier0/store/backfill/1/data/Tier0_REPLAY_2021/HLTPhysics/
-dwax--- 1 cmst0  zh  7.03 M Nov 22 16:15:1 ALCARECO
-dwax--- 1 cmst0  zh  8.55 G Nov 22 16:26:1 AOD
-dwax--- 1 cmst0  zh  1.42 G Nov 22 16:15:1 DOMIO
-dwax--- 1 cmst0  zh  1.93 G Nov 22 16:29:1 MINIAOD
-dwax--- 1 cmst0  zh  109.39 G Nov 16 13:35:1 RAW
-dwax--- 1 cmst0  zh  35.56 G Nov 16 14:44:1 RAW-RECO
-dwax--- 1 cmst0  zh  44.09 M Nov 22 16:57:1 USER
```

```
[root@eosans02-ip554 (msg:slave msg:slave) ~]$ eos file info /eos/pilot/cms/tier0/store/backfill/1/data/Tier0_REPLAY_2021/HLTPhysics/USER/
LogErrorMonitor-PromptReco-v1/000/344/063/000000/078495ac-1e43-433f-aa3f-beb6ae7696e21.root
File: /eos/pilot/cms/tier0/store/backfill/1/data/Tier0_REPLAY_2021/HLTPhysics/USER/LogErrorMonitor-PromptReco-v1/000/344/063/000000/
078495ac-1e43-433f-aa3f-beb6ae7696e21.root' Flags: 0644
Size: 15062478
Modify: Mon Dec 13 08:32:03 2021 Timestamp: 1639380723.458453749
Change: Sun Mar 6 14:15:20 2022 Timestamp: 1646572520.349817905
Birth: Mon Dec 13 08:32:03 2021 Timestamp: 1639380723.458453749
CUID: 13833 CGUID: 1399 File: 074ac769 Fid: 25668385 Pid: 833735 Pid: 000cc4f
XType: adler Xs: 0e 67 60 lc ETAEs: "89068943448473680e67691c"
Layout: replica Stripes: 2 Blocksize: 4k LayoutId: 00100112 Redundancy: d2:t0
```

```
[root@eospilot-ns-ip700 (msg:master msg:slave) ~]$ eos file info /eos/pilot/cms/tier0/store/backfill/1/data/Tier0_REPLAY_2021/HLTPhysics/USER/
LogErrorMonitor-PromptReco-v1/000/344/063/000000/078495ac-1e43-433f-aa3f-beb6ae7696e21.root
File: /eos/pilot/cms/tier0/store/backfill/1/data/Tier0_REPLAY_2021/HLTPhysics/USER/LogErrorMonitor-PromptReco-v1/000/344/063/000000/
078495ac-1e43-433f-aa3f-beb6ae7696e21.root' Flags: 0644
Size: 15062478
Modify: Mon Dec 13 08:32:03 2021 Timestamp: 1639380723.458453749
Change: Sun Mar 6 14:15:20 2022 Timestamp: 1646572520.349817905
Birth: Mon Dec 13 08:32:03 2021 Timestamp: 1639380723.458453749
CUID: 13833 CGUID: 1399 File: 074ac769 Fid: 25668385 Pid: 833735 Pid: 000cc4f
XType: adler Xs: 0e 67 60 lc ETAEs: "751049712926728e67691c"
Layout: replica Stripes: 2 Blocksize: 4k LayoutId: 00100112 Redundancy: d2:t0
```

---

EOS Workshop 8th March 2022

Dr. Jaroslav Guenther (CERN)
Sync of AMS at scale

> every 30 min sync between EOSPUBLIC with EOSAMS02

we also have in place EOS Quota monitoring for IT-ST-PDS instances
(showing here that EOSAMS02 need quota sync ;) )
Thank you

Questions ?