





ATRESYS

Automated Tape REpacking SYStem

v2

Richard Bachmann (developer)
Vladimír Bahyl (speaker)
IT-ST-TAB



Motivation

- There are at least 3 reasons why it is necessary to move data between tapes:
 - Recuperate free space after deletion
 - In case of a hard medium error
 - Migration from one technology to another

- CTA has an internal engine capable of repacking: `cta-admin repack`
 - Works well for set of tapes
 - Missing other steps of operational workflows
 - Limited ways to manage overload (only with tape expansion)

ATRESYS

- Builds a layer on top for:
 - Identification of tapes that require repack (by occupancy; tape pool)
 - Handling of large number of tapes and controlled submission to the main CTA repack engine
 - Periodic checks of the repack progress
 - Moving tapes into quarantine area and wait for specified delay before reclaim (resets counters; no longer to undelete)
 - Reassignment of the tapes to the label or scrutinize tape pools as per configuration

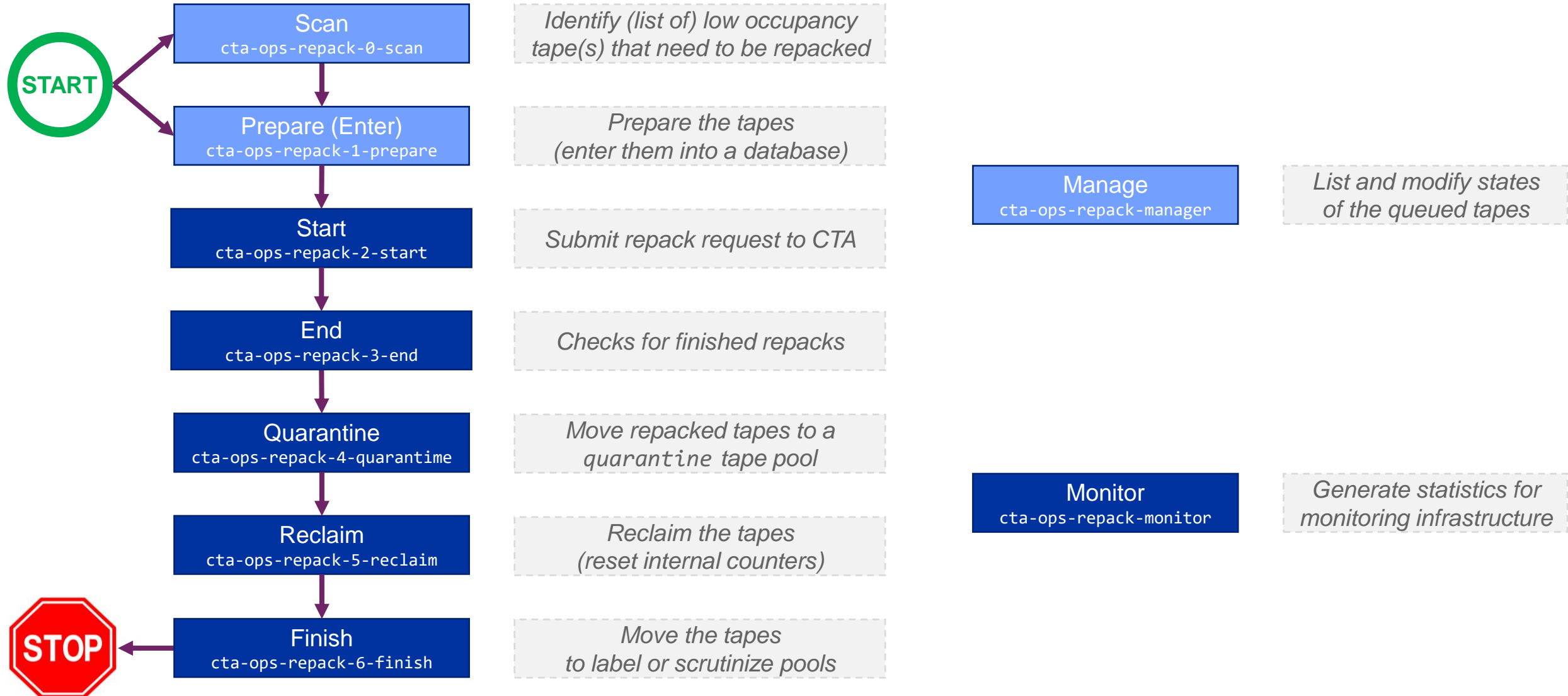
- Implemented in Python

Database schema

- Using separate (CTA tape operations related) PostgreSQL database

Column Name	Type	Nullable
(PK) tape	character varying(20)	not null
entered	timestamp	not null
started	timestamp	
repacked	timestamp	
quarantined	timestamp	
reclaimed	timestamp	
finished	timestamp	
files	bigint	not null
bytes	bigint	not null
usage	smallint	not null
last_write_date	timestamp	
mode	[auto, manual, error]	not null
priority	smallint	not null
mediatype	character varying(20)	not null
tapepool	character varying(100)	not null

Process architecture



cta-ops-repack-manager

```
[root@ctaproductiofrontend02 ~]# cta-ops-repack-manager ls
```

Tape	Status	Media	Pool	Bytes	Usage	Last Written	Mode	Priority	
I72148	6/6 Finished	2023-03-21 09:12	3592JE20T	tolabel_IBM3JE	3.49T	17%	2022-07-14 05:31	auto	low
L87414	6/6 Finished	2023-03-21 03:12	LT09	tolabel_SPC1L9	1.36T	8%	2022-07-31 23:40	auto	low
L87119	6/6 Finished	2023-03-21 03:12	LT09	tolabel_IBM1L	1.09T	6%	2022-07-21 20:01	auto	low
L86107	6/6 Finished	2023-03-20 21:12	LT09	tolabel_SPC1L9	3.61T	20%	2022-08-29 21:47	auto	low
I74408	4/6 Quarantined	2023-03-21 09:12	3592JE20T	quarantine	331.45G	2%	2022-12-11 04:38	auto	low
I73297	2/6 Started	2023-03-16 14:51	3592JE20T	r_backup_afs_2	3.61T	18%	2022-07-24 10:23	auto	low
I73195	2/6 Started	2023-03-16 14:51	3592JE20T	r_backup_afs_2	207.33G	1%	2022-06-09 22:59	auto	low
L87387	2/6 Started	2023-03-16 14:51	LT09	r_backup_afs_1	6.55T	36%	2022-07-08 11:34	auto	low
L85813	2/6 Started	2023-07-19 09:52	LT09	r_backup_afs_1	6.98T	39%	2022-08-24 23:02	auto	low
L87457	2/6 Started	2023-03-16 14:53	LT09	r_backup_afs_1	7.0T	39%	2022-08-27 00:02	error	low
L87993	1/6 Entered	2023-03-16 14:51	LT09	r_backup_afs_1	793.74G	4%	2022-07-23 20:13	auto	low
L87571	1/6 Entered	2023-03-16 14:51	LT09	r_backup_afs_1	4.13T	23%	2022-07-07 15:08	auto	low
L87131	1/6 Entered	2023-03-16 14:51	LT09	r_backup_afs_1	86.19G	0%	2022-06-16 21:47	auto	low
I74430	1/6 Entered	2023-03-16 14:51	3592JE20T	r_backup_afs_2	76.12G	0%	2022-06-05 19:21	auto	low
I75403	1/6 Entered	2023-03-16 14:51	3592JE20T	r_backup_afs_2	6.99T	35%	2022-08-14 19:22	auto	low
I75399	1/6 Entered	2023-03-16 14:51	3592JE20T	r_backup_afs_2	7.13T	36%	2022-08-18 20:14	auto	low
I74375	1/6 Entered	2023-03-16 14:51	3592JE20T	r_backup_afs_2	6.27T	31%	2022-07-25 04:38	auto	low
I74346	1/6 Entered	2023-03-16 14:51	3592JE20T	r_backup_afs_2	291.11G	1%	2022-06-23 22:11	auto	low
I74310	1/6 Entered	2023-03-16 14:51	3592JE20T	r_backup_afs_2	720.69G	4%	2022-07-07 22:11	auto	low

Tape in quarantine

Tapes transition between states

6 states in total

Error situation detected, but not blocking

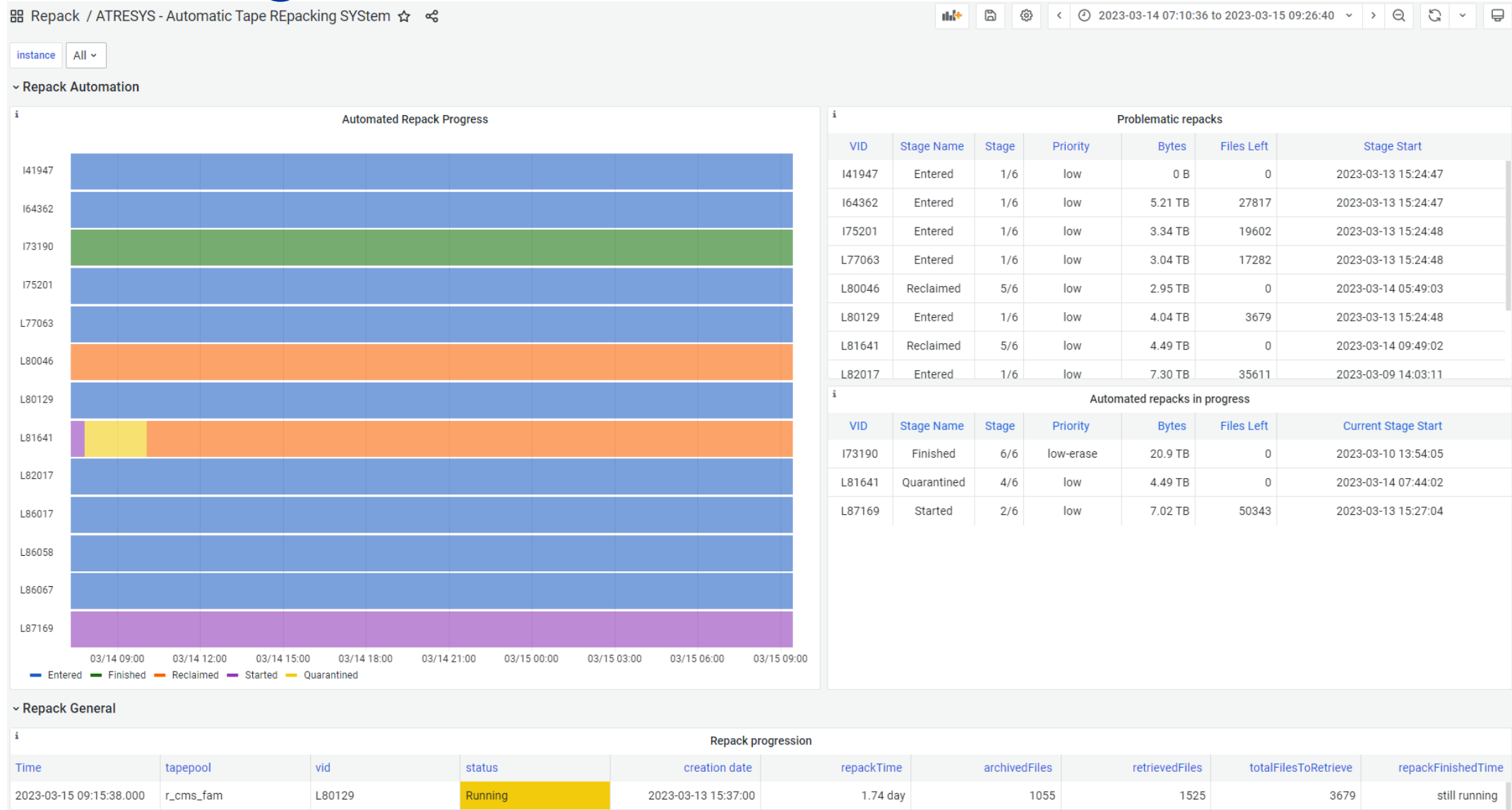
Orchestration with Rundeck

- Rundeck = cron equivalent
- Jobs cta-ops-repack-2-, -3-, -4-, -5-, -6- run:
 - Every 2 hours
 - 10 minutes apart

The screenshot displays the Rundeck interface for a job named 'CTA_repack'. The left sidebar contains navigation options: DASHBOARD, JOBS, NODES, COMMANDS, ACTIVITY, and WEBHOOKS. The main content area shows a list of jobs under the 'CTA Repack Automation' category, including steps like 'cta-repack-0-scan', 'cta-repack-1-prepare', 'cta-repack-2-start', 'cta-repack-3-end', 'cta-repack-4-quarantine', 'cta-repack-5-reclaim', and 'cta-repack-6-finish'. Below this, the 'Activity for Jobs' section shows a table of recent executions.

Execution ID	Time	Status	Duration	Author	Job Name	Config Path
03/21/2023 2:18 PM	Today at 2:18 PM	2 ok	3 sekundy	rbachman	Repack Monitor	
03/21/2023 2:03 PM	Today at 2:03 PM	2 ok	3 sekundy	rbachman	Repack Monitor	
03/21/2023 1:52 PM	Today at 1:52 PM	2 ok	24 sekúnd	rbachman	cta-repack-2-start	/etc/cta-op:
03/21/2023 1:48 PM	Today at 1:48 PM	2 ok	3 sekundy	rbachman	Repack Monitor	
03/21/2023 1:42 PM	Today at 1:42 PM	2 ok	3 sekundy	rbachman	cta-repack-4-quarantine	/etc/cta-op:

Monitoring in Grafana



Conclusion

- ATRESYS significantly facilitates day-to-day repack operations on large number of tapes
 - It follows CERN tape operations' logic but is fully configurable
 - Easily modifiable as written in Python

- Available from the public CTA repository :

<https://gitlab.cern.ch/cta/cta-operations>

- In production at CERN since March 2023