The HERD computing model: status and exploration of CLOUD solutions for cosmic ray data analysis

D. Ciangottini, S. Dal Pra, M. Duranti, <u>V. Formato</u>, N. Mori, D. Spiga

Workshop CCR - Loano - 22/05/2023

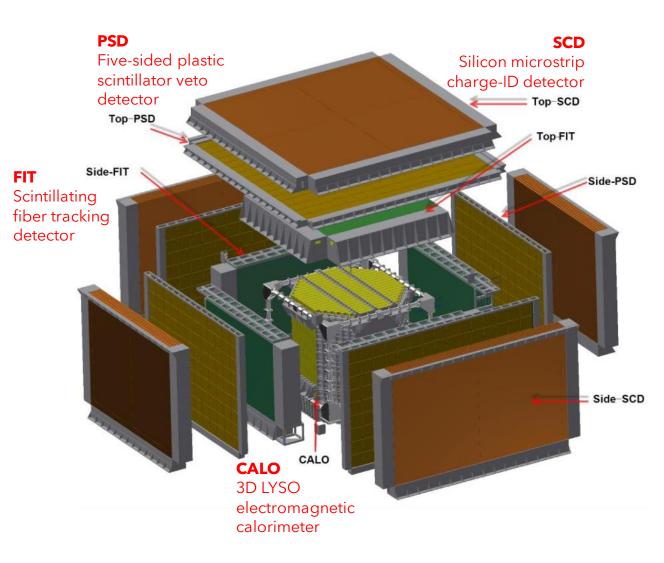






#### HERD

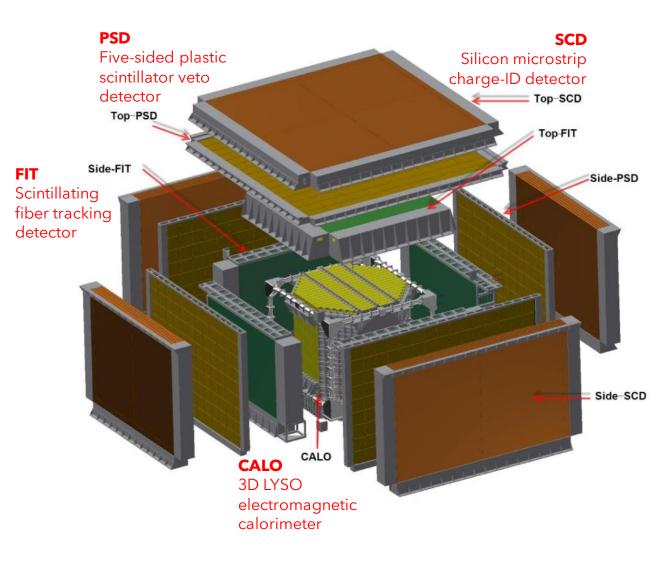
- Flagship and landmark scientific experiment, China-led large international collaboration
- Main Scientific Objectives:
  - <u>Dark matter</u>: dark matter search with unprecedented sensitivity
  - <u>Cosmic-ray</u>: Precise cosmic ray spectrum and composition measurements up to the knee energy
  - <u>Gamma-ray</u>: Gamma-ray monitoring and full sky survey
- Foreseen to operate in space, on board the Chinese Space Station
- Charged CR physics but also  $\gamma$ -ray physics
- ~ O(1M) read-out channels
- The detector is designed to be "isotropic" and accept CR from all (5) the sides





#### HERD

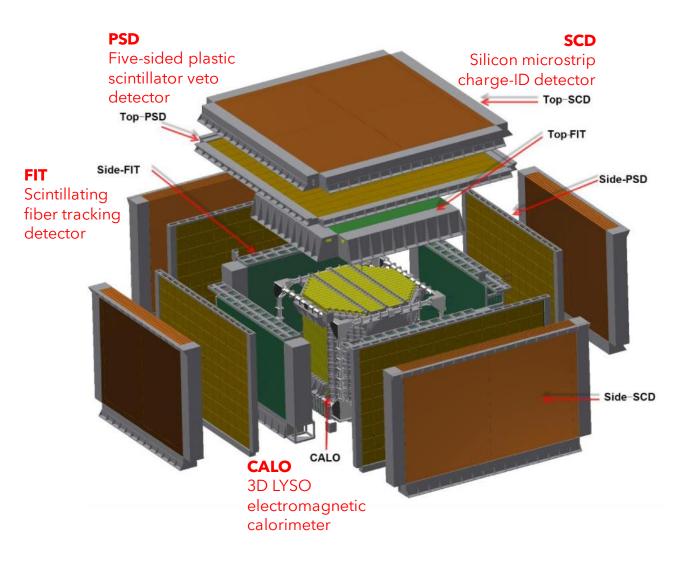
- Currently in advanced design and prototyping stage
  - Possibility to start developing code and computing model from scratch
- Facts about computing models for cosmic-ray experiments:
  - Exp. "size" << LHC
    - "Easy", but...
  - High demand of raw computing power in some scenarios
    - E.g. upper energy limit for MC simulations in the PeV region
  - Difficulties in exploiting some computing optimization techniques
    - E.g. optimized workloads like fast MC simulations may not be suitable for assessing the e/p rejection power with sufficient accuracy
    - Small community with relatively little manpower to devote to this topic
  - Highly dynamic use of computing resources (shortperiod bursts)
    - Due to e.g. small number of involved people leading to significant usage fluctuations

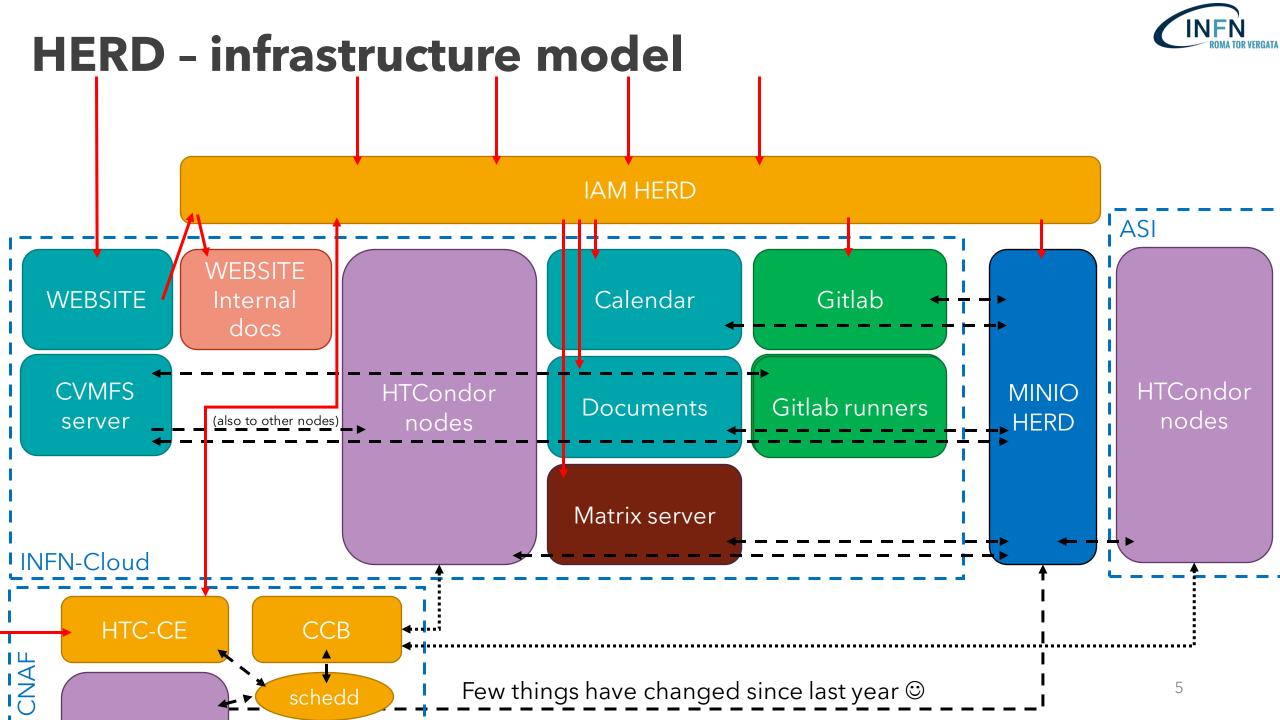




#### HERD

- The cloud can provide some solution:
  - Add/remove resources with minimal overhead to dynamically cope with high/low demand periods
  - Efficiently exploit opportunistic resources with cloud-native solutions
  - Deploy self-hosted, self-managed services
- The computing model must be designed to fully profit of the cloud
  - E.g. to maximize the usage of opportunistic resources by minimizing the set-up period after the resources are made available in the cloud
- Towards a fully-cloud based model for HERD:
  - R&D and prototyping work
  - Based on INFN Cloud infrastructure and resources







## **2022: Feedback and next steps**

#### • Needed:

• <u>CVMFS as-a-service</u>.

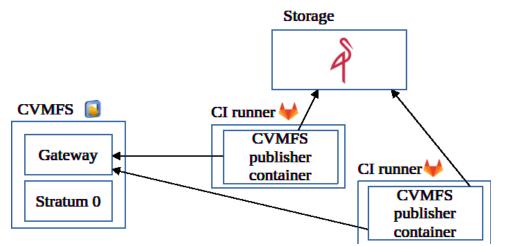
To continue testing the HTCondor service (and eventually use it in production) we need a CVMFS server for software distribution.

- Desirable:
  - <u>Self-managed DNS on select subdomains</u> To ease the workflow of deploying new services or maintaining existing ones
- Feedback:
  - <u>Support team with high availability for critical, non-self-managed parts of the infrastructure</u>
     We had several issues with our Minio backend which we couldn't debug/fix on our own. We relied on a single contact person during the test phase but this solution clearly doesn't scale to production.
  - <u>The whole INFN-Cloud infrastructure feels a bit "user-centric" rather than "team-centric"</u> For example: within the HERD tenant each user can see and manage only his own deployments
- Next:
  - <u>Continue testing and improve the HTCondor workflow for analysis</u>
     Now testing S3 storage solutions ✓, need CVMFS integration ✓, flocking, user-mapping, ...
  - <u>Backport hand-made deployments in INFN-Cloud dashboard</u> (e.g. Gitlab as-a-service)



## **CVMFS**

- Self-deployed CVMFS server on INFN Cloud resources
- Activities synergic with DataCloud WP6
- Using S3 storage as a backend
- Fully integrated with Gitlab CI/CD for deployment of new SW releases / development branch
  - Triggered by commits on release tags (x.y.z) or on master branch



🦪 🚍 🔍 Search GitLab	Γ				o v D'2 % <mark>1</mark> v ⊡5 @	▶ 🍪 -
HerdSoftware	herd > 💿 HerdSoftware > Pipelines > #7049					
<ul> <li>Project information</li> <li>Repository</li> </ul>	passed Pipeline #7049 triggered 12 hours ago by Sicola Mori					
D' Issues         44           11 Merge requests         2           2 <sup>0</sup> CUCD         Pipelines           Editor         Jobs           Schedules         Schedules	Merge branch '374-fix-psd- Resolve "Fix PSD WC geometry regress Closes #374 See merge request 1233	mc-geometry-regression-an	d-upda	te-doc' into 'master'		
<ul> <li>Executive and Compliance</li> <li>Deployments</li> </ul>	() 21 jobs for master in 28 minutes and 56 seconds and was queued for 10 seconds					
Packages and registries	C Scheduled latest					
	↔ bibl0c8d @					
l≟ Analytics ♀ Wiki	No related merge requests found.					
X Snippets ፼ Settings	Pipeline Needs Jobs 21 Tests 0					
, cominge	Group jobs by Stage Job dependencies					
	code-quality-reports	deploy		Downstream		
	Code-quality-html-report	deploy-installer-centos7-scheduled		HerdSoftwareInst #7052 < Multi-project	deploy	
	Code-quality-json-report	deploy-installer-ubuntu20.04-scheduled      Trigger job	C		deploy-hs-cnaf-centos7     deploy-hs-cvmfs-centos7	3
		deploy-installer-ubuntu22.04-scheduled      Trigger Job			eploy-hs-recas-centos7	3
				HerdSoftwareInst #7051 >		
				HerdSoftwareInst #7050		

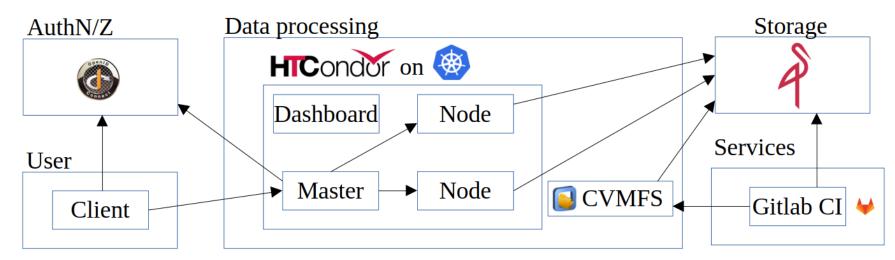
The HERD computing model: status and exploration of CLOUD solutions for cosmic ray data analysis

## **Batch system**



We continued the tests with the on-demand HTCondor cluster on K8S

- Realized a docker image for a HERD HTCondor client
  - Based on the htcondor/submit image
  - Additional dependencies added:
    - Rclone
    - Oidc-agent
    - Boto3-sts
    - Cvmfs client
    - Devtools (gcc, cmake, boost, etc...)
  - The user should be able to build and submit its SW using this image

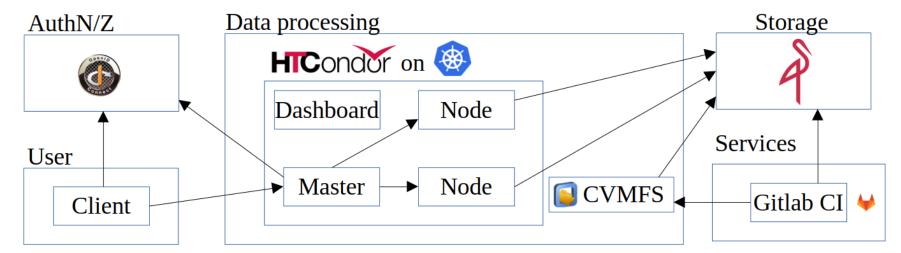


## **Batch system**



We continued the tests with the on-demand HTCondor cluster on K8S

- Jobs perform I/O from/to the S3 storage using pre-signed URLs
  - psURLs are retrieved at job submission using the user's token and are stored on the local storage of the WN
  - Currently the psURLs cannot last more than 7 days, so a solution is needed for long-lasting jobs (e.g. ~PeV simulations)
- WNs mount the HERD CVMFS server automatically
  - On each K8S node of the underlying cluster a daemonset is added exposing the CVMFS mount as a volume to all the pods running on that node

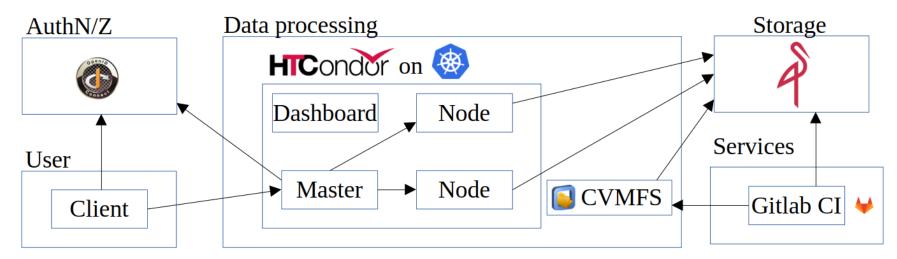




## **Batch system**

Tested real life pipelines:

- A launch script retrieves the psURLs for all input and output files, prepares the Condor submit file and the job script, and submits the job cluster
  - Simulation jobs: standard Geant4 simulation, output file is copied on S3 with curl using a psURL
  - Digitization jobs: the job script reads the simulation input file from S3 using curl and a psURL, executes the digitization, and stores the digitized output file on S3 using a psURL
  - Reconstruction: same workflow as digitization, with a digitized file as input
  - Analysis: same workflow as reconstruction, with a reconstructed file as input

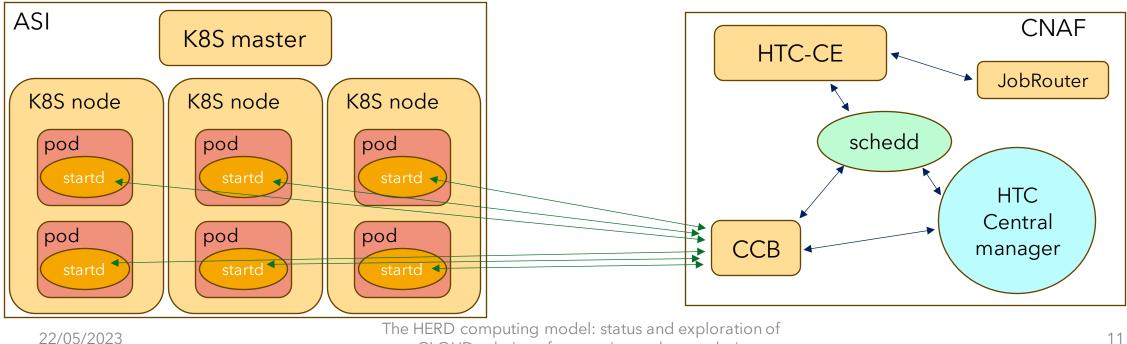


## Batch system: WIP



Currently working on:

- Include opportunistic resources and merge them with resources @ T1
  - Same workflow as in CNAF-Reloaded (see <u>talk by S. Dal Pra @ CHEP 2023</u>): HTCondor-CE element managing additional external nodes
    - Submission to HTC-CE with custom attribute.
    - Attribute is translated to relevant ClassAd by JobRouter
    - Remote WN spawn startd processes that join the CNAF POOL through the CCB. All traffic to remote nodes goes through CCB.
  - Tested with VM @ Recas running a dockerized WN
  - Setting up a dedicated K8S cluster in ASI (activity under ASI-INFN Agreement No. 2021-43-HH.0) where working nodes will be spawned



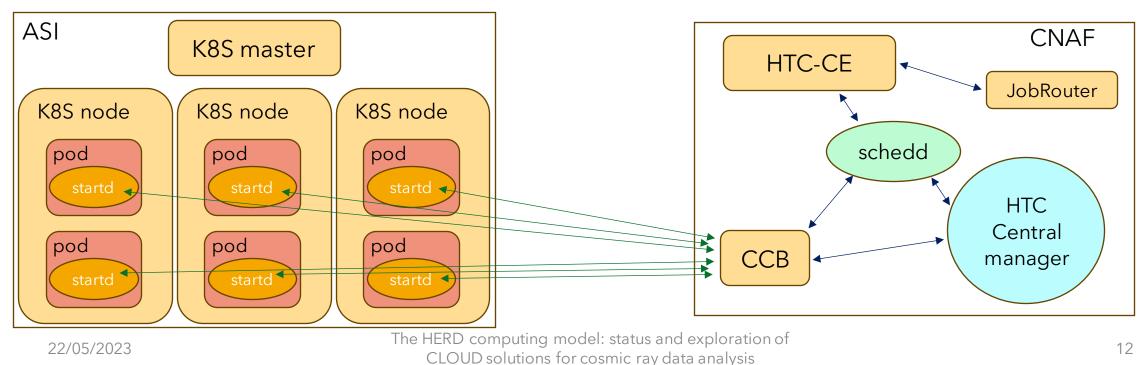
CLOUD solutions for cosmic ray data analysis



## Batch system: WIP

Currently working on:

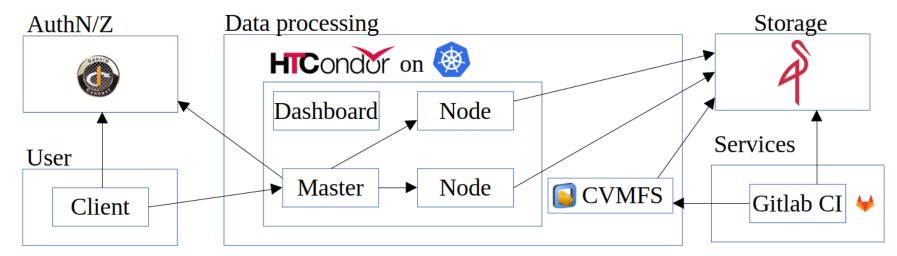
- Include opportunistic resources and merge them with resources @T1
  - Since all traffic goes through CCB it's recommended to minimize the usage of HTCondor file transfer mechanism
    - Most of the SW stack is on cvmfs, but a small part of the job payload (user code) is currently transferred in this way
    - Exploring automated solutions to deploy also user code on cvmfs or on a remote host (and maybe copy via gfal)
  - In the next future we could adapt the auto-scale machinery used for M100 to handle multi-tenant setup (where each experiment requires a custom WN image)





# **Batch system: Open issues**

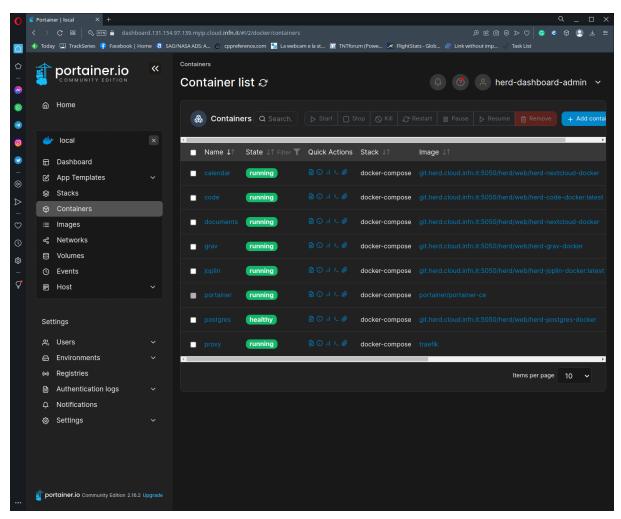
- Performance still to be investigated
  - Sub-optimal storage testbed (MinIO on VMs with Ceph backend storage)
  - Single-site setup, no multi-site optimizations yet
- S3 access control for long-queueing/long-running jobs
  - psURLs validity currently limited to 1 week
- Distribution of user's code (CVMFS?)
  - Using HTCondor file transfer at the moment
- Integration with "conventional" pledged resources: ongoing





## Web infrastructure

- Over the time we have set up an extensive set of web applications in support of the activities of the collaboration.
- Peculiarities of the HERD collaboration:
  - Cosmic-ray experiment → no umbrella organization like CERN → no institutional services available to every collaboration member
  - Chinese institutions → no access to free tools from e.g. Google
- Solution: self-host the needed services at collaboration level
  - The cloud approach offers a valid approach for this purpose
- Fully-containerized web applications running on cloud VMs
  - Currently planning a migration to a K8S cluster to allow automatic re-deploy and load balancing



# Authentication

Goal: everyone <u>in the collaboration</u> should have access to the experiment resources.

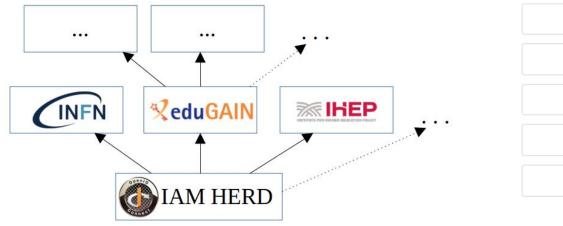
lssues:

- Not everyone has a CERN account
- Not everyone has a INFN-AAI account
- Not everyone has access to Google

Solution: Single Sign On based on a dedicated INDIGO-IAM instance (IAM HERD)

Federated with institutional SSO services directly and through EduGAIN

OpenID Connect protocol





Welcome to **herd** 

Sign in with





Requirement: we need several services to ease scheduling of day-to-day operations and/or meetings, activities, as well as document sharing, and more.

Issue: Almost every mainstream tool is unavailable to colleagues in China.

Solution: self-hosting of all services:

• Calendar (based on Nextcloud)

< > C 8	몸   🔍 🔒 calendar.h	erd.cloud.infn.it/apps	/calendar/dayGridMo	onth/now	K @ Ø > ♡   (	≥ (2) (2) 🕹
🌒 Today 💷 Tra	ackSeries 👍 Facebook   Ho	me (2) SAO/NASA ADS:	AD 💮 cppreference.	com 🛛 👷 La webcam e	la staz 🄟 TNTforum (Po	wered
						Q 😃 🤇
	Mon	Tue	Wed	Thu	Fri	Sat
1	2	3	4	5	6	7
	• 10:00 AM <b>SCD</b>				● 9:30 AM <b>PSD M</b>	
8	9 ● 9:00 AM <b>CALO</b>	10 • 6:00 PM <b>Riunio</b>	11	12	13 ● 9:30 AM <b>HERD</b> ● 12:00 PM <b>ILB m</b>	14
15	16	17 ● 10:00 AM <b>Analy</b>	18	19	20	21
22	23	24	25	26	27	28

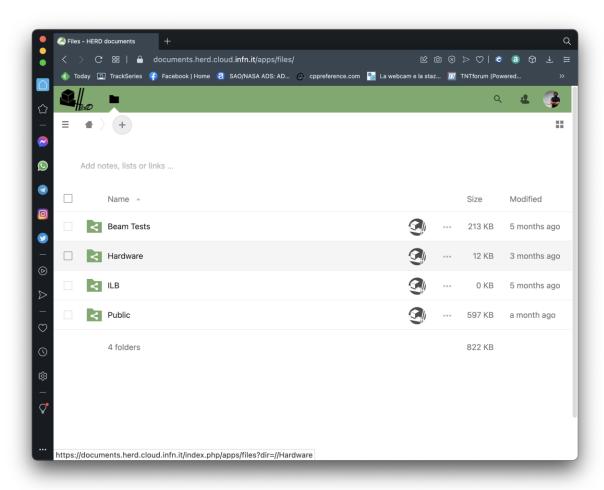


Requirement: we need several services to ease scheduling of day-to-day operations and/or meetings, activities, as well as document sharing, and more.

Issue: Almost every mainstream tool is unavailable to colleagues in China.

Solution: self-hosting of all services:

- Calendar (based on Nextcloud)
- Document server (based on Nextcloud + Collabora)



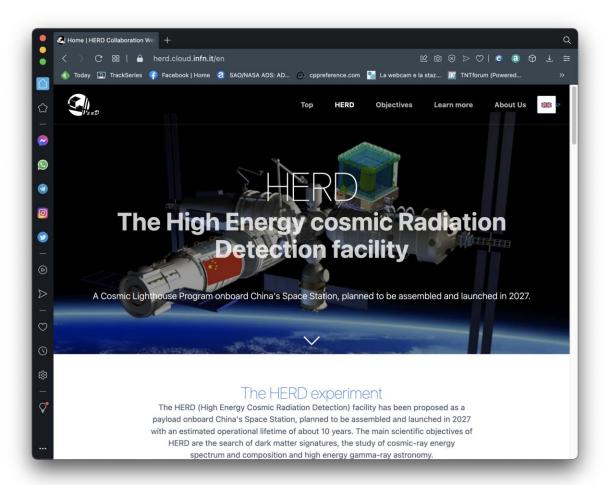


Requirement: we need several services to ease scheduling of day-to-day operations and/or meetings, activities, as well as document sharing, and more.

Issue: Almost every mainstream tool is unavailable to colleagues in China.

Solution: self-hosting of all services:

- Calendar (based on Nextcloud)
- Document server (based on Nextcloud + Collabora)
- Experiment website (with dedicated restricted area for internal documentation, behind IAM login)





Requirement: we need several services to ease scheduling of day-to-day operations and/or meetings, activities, as well as document sharing, and more.

Issue: Almost every mainstream tool is unavailable to colleagues in China.

Solution: self-hosting of all services:

- Calendar (based on Nextcloud)
- Document server (based on Nextcloud)
- Experiment website (with dedicated restricted area for internal documentation, behind IAM login)
- Gitlab instance (with runners)

0	Projects · GitLab × +		Q _ 🗆 X
	< > C 🔐 💵 🗎 git.h	erd.cloud.i <b>nfn.it</b>	ଝ๏⊗⊳♡ © ° ଡ଼ .9 ± ≆
	🚸 Today 🛄 TrackSeries 👎 Facebo	ok   Home 🔞 SAO/NASA ADS: A 🕞 cppreference.com 🧏 La webcam e la st 🔟 TNTforum (Powe 🌌 Fl	ightStats - Glob   kink without imp »
<b>_</b>	🦪 🔳 🔍 Search GitLab	0	× D°2 ╏┧× ☞5 @•× 🍪 × 🔺
—	🔁 Your work	Your work > Projects	
	Projects	Duciante	
۵	Se Groups	Projects	Explore projects New project
	D Issues (2)	Yours 22 Starred 0 Filter by name	Language  VDpdated date  V
0	እ Merge requests	All Personal	
	☑ To-Do List 5	<u> </u>	
<b>&gt;</b>	⑦ Milestones ✗ Snippets	herd / HerdSoftware ① <u>Owner</u> Simulation, reconstruction and analysis framework	Image: Second states of the second states of th
0	S Shippets		
⊳		Wiki documentation for HerdSoftware.	⊘ ✿ 0 撃 1 ເືມ2 D≯ 0 Updated 6 days ago
$\heartsuit$		H herd / HERD_TROC_firmware 🔒 Owner	☆ 0   \$* 0   \$* 0   D* 0 Updated 1 week ago
0		herd / CaloCubeBTsoftware ① Owner	*0 %0 %0 D0
ţĝ;		C Repository for the acquisition and online analysis software used for beam test of CaloCube and	Updated 1 week ago
_		HERD prototype	
ç		H herd / web / Herd Grav 👌 Owner Web site	☆0 ¥0 \$1,0 D•2 Updated 3 weeks ago
		G herd / analysis / benchmarks / GeometricFactor 🔒 Owner	☆0 ¥0 \$30 DP0 Updated 3 weeks ago
		herd / HerdSoftwareInstaller () (Owner Installer scripts for HerdSoftware and environment.	
		H herd / HerdHTCWorkerNode 🔒 Owner	ি ☆ ০
		H Valerio Formato / Htcondor Cvmfs K8s config D Owner	☆ 0  ¥ 0  \$1 0  D 0 Updated 2 months ago
		H herd / web / Herd Grav Docker 🔒 Owner	
	« Collapse sidebar	H herd / HerdHTCFrontend () (Owner)	Updated 3 months ago
	« Conapse sidebai		



Requirement: we need several services to ease scheduling of day-to-day operations and/or meetings, activities, as well as document sharing, and more.

Issue: Almost every mainstream tool is unavailable to colleagues in China.

Solution: self-hosting of all services:

- Calendar (based on Nextcloud)
- Document server (based on Nextcloud)
- Experiment website (with dedicated restricted area for internal documentation, behind IAM login)
- Gitlab instance (with runners)

Currently testing also:

- Experiment-wide chat service (based on Matrix)
- INDICO

•••	Element   General	
Q Filte 18 K		
Settings		
<ul> <li>General</li> <li>Appearance</li> <li>Notifications</li> <li>Preferences</li> <li>Keyboard</li> <li>Sidebar</li> <li>Voice &amp; Video</li> <li>Security &amp; Privacy</li> <li>Labs</li> <li>Help &amp; About</li> </ul>	Ceneral   Profie   valerio Formato   @vformato:chat-server.herd.cloud.infn.it   Cancel Save   Account   Email addresses   Email Address   Add   Phone numbers	e e e e e e e e e e e e e e e e e e e
	Add Language and region	
o	Send a message	e @ …



## HERD - storage

- S3-compatible storage based on MinIO
  - Well-established cloud storage technology
  - Available storage backends for many components (e.g. CVMFS)
- Used for data, software and services
- Testbed based on 4 MinIO instances running on Kubernetes
  - 100 TB raw capacity provided via Ceph (DICE project, grant agreement ID: 101017207)
    - 3+1 erasure coded storage
  - 100 TB on INFN-Cloud resources, to be integrated with existing resources
- Access control:
  - OIDC tokens managed via oidc-agent (users)
  - Access keys (services)

-			
	A MinIO Console +		م
•		l-console.cloud.cnaf. <b>infn.it</b> /buckets	⊠ © > ♡   ♥ ₫ ∰ ± ≡
	🔹 Today 🛄 TrackSeries 存 Facebook	Home 🔞 SAO/NASA ADS: AD 💮 cppreference.com 📓 La webcarn e la staz 🌆 TNTforum (Powered 🗷 Flight	tStats - Global 🚢 Link without improvi
ය –	CONSOLE 4	gitlab-artifacts Created: 2022-03-21T09-22:41Z Access: R/W	Manage ⓓ Browse →
	Dashboard		
©	Buckets	G Usage @ Objects 19 <sub>GB</sub> 1,392	
	<b>a</b> Users		
<b>9</b> -	Sroups	gitlab-backup Created: 2022-03-06T19:16:56Z Access: R/W	Manage ۞ Browse →
- ©	Service Accounts	O Usage O Objects 252 <sub>MB</sub> 40	
⊳	1 IAM Policies		
0	I AM POLICIES		
©	🔅 Settings	gitlab-dependency-proxy Created: 2022-03-21T09:22:42Z Access: R/W	Manage ③ Browse →
(\$)	λ Notification Endpoints	Usage 🟵 Objects	
_ ♀	😸 Tiers	$ \begin{array}{cccc}  & & & & & & \\  & & & & & \\  & & & & & $	
	Tools	gitlab-external-diffs	Manage ⓓ Browse →
	E License	Created: 2022-03-21T09:22:41Z Access: R/W	
	Documentation	$ \begin{array}{c c} & & & & & \\ \hline & & & \\ & & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & $	
	R		

# 2023: Feedback and next steps



- Needed:
  - <u>Cannot reach HTCondor-CE element from VMs hosted by Cloud@CNAF</u> Since our resources are primarily hosted on Cloud@CNAF
- Desirable:
  - <u>Self-managed DNS on select subdomains</u> To ease the workflow of deploying new services or maintaining existing ones
- Feedback:
  - <u>The whole INFN-Cloud infrastructure feels a bit "user-centric"</u> rather than "team-centric" For example: within the HERD tenant each user can see and manage only his own deployments
- Next:
  - <u>Continue testing and improve the HTCondor workflow for analysis</u> Finish work on resource merging, user-mapping, ...
  - Backport hand-made deployments in INFN-Cloud dashboard (e.g. Gitlab as-a-service)
  - Assess the fate of the DICE storage once the project ends



## Conclusions

- The "astro-particle in space" community is eager of resources and poor in terms of manpower for computing: we will always be willing to test any solution to increase our pool of resources and to keep up with the software infrastructure developments, given the limited man-power available
- Given the nature of the partners for the various projects (ASI, Chinese collaborators, ...) we can have small and/or temporary resources at our disposal: merging them in a single batch system is a big added value. <u>We are getting really close now</u>.
- As a new experiment in a design phase, HERD is proving to be the perfect opportunity to migrate towards a deeper integration with INFN-Cloud provided services and we are continuously improving in both creating new services and maintaining them.