



KISTI-GSDC Report Sang-Un Ahn **On behalf of KISTI-GSDC**

16-18 April 2024 @ ALICE Tier-1/Tier-2 Workshop in Seoul











100Gpbs became a standard

KISTI ALICE T1 Structure Overview









Active Jobs





Active jobs in KISTI_GSDC Intense ALICE Activity New worker nodes deployed 160 cores (10.78 HS06 / core) 790 GB memory m A Aug Sep Feb 2024 - RUNNING - ASSIGNED - STARTED - SAVING - ZOMBIE





Multicore Jobs



Number of jobs at KISTI_GSDC									
	Series	Last value	Min	Avg	Max				
1.	1	358	0	1069	580				
2.	2	12	0	18.39	87				
3.	4	1	0	15.16	30				
4.	8	119	0	54.35	549				
5.	16	1	0	0.619	1				
6.	64	2	0	4.2	2				

Number of jobs at KISTI_GSDC 8-core jobs + 1 + 2 + 4 + 8 + 16 + 64 + SUM



Jobs Efficiency



SE Status



CDS : Custodial Disk Storage EOS : EOS Disk Storage SE2 : XRootD Disk Storage

	Used (PiB)	Total (PiB)	Usa
CDS	5.766	15.78	3
EOS	1.62	5.948	2
SE2	1.172	1.446	8

> 97% of overall SE availability for writing/reading



Linknows	Dat	a	Individual results of reading tests						
спк пате	Starts	Ends	Successful	Failed	Success ratio	Ava			
KISTI_GSDC::CDS	31 Dec 2022 23:01	13 Apr 2024 03:21	11090	243	97.86%	2			
KISTI_GSDC::EOS	31 Dec 2022 23:17	13 Apr 2024 03:37	11017	236	97.90%	2			
KISTI_GSDC::SE2	31 Dec 2022 23:11	13 Apr 2024 03:31	11145	132	98.83%	>			

Mar



WLCG Data Challeng CDS Participation as a Tape

- Transfer of real Pb+Pb data collected in 2023, 34PB in
- 1PB of data being transferred after the challenge, ETA



		Centre	Target rate GB/s	Average achieved GB/s
		CNAF	0.8	0.98 (+20%)
		IN2P3	0.4	0.6 (+40%)
		KISTI	0.2	0.25 (+22%)
		GridKA	0.6	1.12 (+90%)
		NDGF	0.3	0.35 (+15%)
n total		NL-T1	0.1	0.25 (+150%)
end of March		RAL	0.1	0.58 (+500%)
		CERN	10	14.2 (+40%)
Steady state	Pb+Pb data		sfer cor	Total 1PB tinuing
25 26 27 28 29 1 2	3 4 5	6 7 Mar 2024	8 9 10	11 12 13



Site Availability

	2023													2024			
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb			
Availability	98	97	99	100	98	94	98	98	99	98	95	97	97	97			
Overall	97.5											97	7				

- until recently
 - Close collaboration with the KREONet service team to address the issue
- An unknown timeout issue during the EOS deployment on CentOS 9 Stream degraded availability in February and March of this year
 - Migrating to AlmaLinux 9.3 resolved the issue, attributed to the newer kernel on CS9 and complementary packages accompanying infrastructure provisioning for security, accounting, auditing and monitoring
- Availability re-computation requests vis GGUS tickets were approved and applied to WLCG monthly reports

• Misconfiguration on LHCOPN backup link has affected significantly the site availability, persisting from the second half of last year



Pledges



Installed (Plan)	2021	2022	2023	2024	(2025
CPU (cores)	3,880	3,880	3,880	6,920	8,040
DISK (TB)	4,500	4,500	6,500	7,100	8,200
TAPE (TB)	12,000	12,000	12,000	12,000	12,00



EOS Deployments (1/2) **Disk Storage**



- Transparent transition of MGM and QuarkDB clusters from VMs to Containers
- EOS upgrade from 5.1.22 to 5.2.16 for existing setup, FMD migration from LevelDB completed beforehand
- Expanded to 6.5PB



EOS Deployments (2/2) **Custodial Storage**



- Disk-based Raw Archive storage for ALICE in production since 2021 deployed using Docker Container
- Comparable level of data protection provided by QRAIN Layout (12 stripes + 4 parities + 2 spares)
- Successful upgrade to v5.1.22 from v4.8.82 (May 2023)

[root@jbod-mgmt-07 MGM_MASTER=true /]# eos attr ls /eos/gsdc/grid sys.eos.btime="1612374338.811408574" s.forced.blockchecksum="crc32c" ys.forced.blocksize="1M" sys.forced.checksum="adler" sys.forced.layout="qrain" sys.forced.nstripes="16" sys.forced.space="default"



EOS @ KISTI for ALICE

ALICE::KISTI_GSDC::EOS

KISTI_GSDC - EOS																						
	AliEn SE			C	atalogue st	atistics				Sto	orage-provi	ded info	rmation		Func	tiona	l tests		Last day ad	d tests	Demotion	IP
SE Name	AliEn name	Tier	Size	Used	Free	Usage	No. of files	Туре	Size	Used	Free	Usage	Version	EOS Version	add	get r	m 3rd	Last OK add	Successful	Failed	factor	ac
1. KISTI_GSDC - EOS	ALICE::KISTI_GSDC::EOS	1	5.948 PB	1.639 PB	4.309 PB	27.55%	50,149,564	FILE	5.948 PB	1.74 PB	4.208 PB	29.25%	Xrootd 5.6.7	5.2.16				14.03.2024 14:43	25	0	(ງ
Total			5.948 PB	1.639 PB	4.309 PB		50,149,564		5.948 PB	1.74 PB	4.208 PB				1	1	1 1					

ALICE::KISTI_GSDC::CDS

CDS																						
	AliEn SE			Ca	atalogue s	tatistics				Stora	age-provide	ed inform	ation		Fun	ctiona	l tests		Last day ad	d tests	Demotion	IP
SE Name	AliEn name	Tier	Size	Used	Free	Usage	No. of files	Туре	Size	Used	Free	Usage	Version	EOS Version	add	get r	m 3r	Last OK add	Successful	Failed	factor	ad
1. KISTI_GSDC - CDS	ALICE::KISTI_GSDC::CDS	1	15.79 PB	5.378 PB	10.41 PB	34.06%	10,959,791	FILE	15.76 PB	7.909 PB	7.856 PB	50.17%						14.03.2024 14:27	24	0	4.706%	,
Total			15.79 PB	5.378 PB	10.41 PB		10,959,791		15.76 PB	7.909 PB	7.856 PB				1	1	1 1					1

- IPv4/IPv6 Dual Stack
- ALICE-Specific Token Authentication/Authorization, HTTP(S), Third-Party Copy enabled
- MLSensor (successor of EOS Apmon) deployed for monitoring (not yet for CDS)



Disk storage elements

Custodial storage elements





EOS v5 Container on EL9: Practices (1/3) **Podman container runtime**

- Exploiting native support of EL9 for Podman (daemonless)
- ContainerFile (cf. *Dockerfile*)
 - EL9 base images { CentOS 9 Stream | AlmaLinux 9.3 }
 - EOS v5 EL9 release installation
 - Reused Container entry script for CDS deployment (Docker based)
 - A few modification made to accommodate different monitoring scheme of ALICE: eosapmond →mlsensor

(L)	탐색기	! site.ya	ml	< ContainerFile 🗙	\$ docker-entrypoint
	◇ 열려 있는 편집기	playbook	s > role	s > image-builder > file	s > 👉 ContainerFile > 🛙
Q	! site.yaml playbooks	1	FROM	almalinux:9	
/	🗙 👉 ContainerFile playbooks/roles/i	2	LABEL		ng-Un Ahn, sahn@
የያ	\$ docker-entrypoint.sh playboo	3			
6	! main.yaml playbooks/roles/met	4	# for	- IGTF CA distri	bution
\sim	~ 위제 다 다 ひ &	5	COPY	repos/EGI-trust	anchors.repo /et
ė∕	> inventory	6	RUN C	dnf install -v c	a-policv-eqi-cor
	imes playbooks	7		, , , , , , , , , , , , , , , , , , ,	- p, .g
- <u>-</u> ⊗	\checkmark group_vars/all	, 8	# Cre	eating certifica	te bundle for na
	! vars.yaml	q	# RIIN	l cat /etc/arid_	security/certifi
b D	\checkmark roles	10		a cat /ctc/griu	Security/certifi
	imes image-builder	11	# for		
÷.	> defaults	10		LFLL Anf inctall v b	ttpc.//d] fodoro
	\checkmark files	12	KUN C	ini instatt –y <u>n</u>	
	> image_scripts	13			
	> keytab	14	# TOP	r wLCG repo	
	> repos	15	RUN C	inf install –y <u>h</u>	ttps://linuxsoft
	🗇 ContainerFile	16			
	🐡 ContainerFile.el7	17	# Ins	stall utils	
	> tasks	18	RUN C	dnf install −y w	get bind-utils n
	> message-queue	19			
	> meta-data-server	20	# Ena	albe CRB	
	> quarkdb-observer	21	RUN C	dnf config-manag	er ––enable crb
	> register-filesystem	22	RUN C	dnf install −y e	dg-mkgridmap
	> storage-server	23			
	! site.yaml	24	# for	r EOS installati	on
	🌣 ansible.cfg	25	RUN C	dnf config-manag	eradd-repo <mark>"h</mark>
	\$ cleanup.sh	26	RUN C	dnf config-manag	eradd-repo "h
	≡ memo.txt	27	RUN C	dnf install –y –	-nogpgcheck \
	≡ podman_run_cmd.txt	28	e	eos-server eos-q	uarkdb eos-jemal
	 README.md 	29	e	eos-fusex eos-fu	sex-core_eos-fus
\mathfrak{D}	! requirements.yaml	30	e	eos-ns-inspect \	
ע		31	r	nscd \	
3	> 개요	32	X	<pre>krootd-client \</pre>	
کن	> 타임라인	33	a	alicetokenacc \	
< វ្ត	ி main				



EOS v5 Container on EL9: Practices (2/3) **Automation via Ansible Playbook**

- Playbook structure:
 - *site.yaml* tags defined to perform specific operation (role) in an automated way
 - group_vars
 - *vars.yaml* key-value variables for group parameters such as eos_instance_name, eos_geotag, ports, master/slave MGM FQDNs, QDB cluster and FST data directories
 - roles
 - image-builder | message-queue (MQ) | meta-data-server (MGM) | quarkdbobserver (QDB) | register-filesystem | storage-server (FST)
 - *handlers* defined to invoke *firewalld* policy implementation and *systemd* integration
 - Creating essential configuration files by templating *xrd.cf.*{*qdb*|*mq*|*mgm*|*fst*}, eos_env, scitokens, ALICE-specific (TkAuthz.Authorization & mlsensor), etc.

••	•			$\leftarrow \rightarrow$,							
۴h	탐색기		! site.yan	nl 🗙 🐡 ContainerFile	\$ docker-entrypoint.sh							
	◇ 열려 있는 편집기		playbooks	s > ! site.yaml								
Q	× ! site.yaml playbook	s	1									
	< ContainerFile play	books/roles/i	2	<pre>- name: Build an EOS</pre>	image							
Ŷ٥	\$ docker-entrypoint	.sh playboo	3 become: true									
6	! main.yaml playboo	ks/roles/met	4 hosts: "{{ eos build nodes }}"									
	∽ 무제 [.	ដ្ដីប្រា	5	roles:	<u>-</u>							
₿.	> inventory		6	- role: image-bu	ilder							
	imes playbooks		7	tags: build-im	ade							
L_⊗	\sim group_vars/all		8	vars:								
	! vars.yaml		g	eos build nodes:	localhost							
Ш	\checkmark roles		10									
	> image-builder		11									
.	> message-queue		12	- name: Setup ODB Ob	server container servi							
	> meta-data-server		13	become true	Server concarner Servi							
	> quarkdb-observer		1/	bosts "{{ eos adb	nodes }}"							
	> register-filesystem		15									
	> storage-server		15		bconvor							
	! site.yaml											
	ansible.cfg	~/Library/Mo	bile Docume	nts/com~apple~CloudDocs/업·	무/GitHub/ansible-							
	\$ cleanup.sh	playbook-eo	s/무제/playbo	ooks/site.yaml								
	≡ memo.txt		19	eos_qub_noues: u								
	<pre></pre>		20		tainan aamuiaa							
	(i) README.md		21	- name: Setup MQ con	tainer service							
	I requirements.yaml		22	become: true								
			23	hosts: "{{ eos_mq_i	nodes }}"							
			24	roles:								
			25	- role: message-	queue							
			26	tags: mq								
			27	vars								
			28	eos_mq_nodes: lo	calhost							
			29									
\otimes			30	- name: Setup MGM con	ntainer service							
			31	become: true								
503	> 개요		32	<pre>hosts: "{{ eos_mgm_</pre>	_nodes }}"							
	> 타임라인		33	roles:								
× ł	<u>° main ↔ ⊗ 0 ∆ 0 </u> ⊗ 0											



EOS v5 Container on EL9: Practices (3/3) **Systemd Integration**

- Systemd service file for each of EOS components manipulating podman commands in such a way that it invokes *podman* {*run*|*rm*|*stop*|...}
 - E.g. /etc/systemd/system/{qdb|mq|fst|mgm}-container.service

<...> ExecStart=/usr/bin/podman run < parameters > ExecStop=/usr/bin/podman stop ExecStopPost=/usr/bin/podman rm <...>

- systemdctl {start|stop|restart} {qdb|mq|fst|mgm}-container.service
 - syslog (journalctl) traces container logs (= podman logs)
- Service update as well as roll-back can be quick and easy
 - Update images (pulling from registries or uploading from local one)
 - systemctl restart *-container.service

			\leftarrow	- >	▷ 무제							
Ð	탐색기 ····	! site.yar	nl	< ContainerFile	\$ docker-entrypoint.sh							
	◇ 열려 있는 편집기	playbooks	s > roles	s > meta-data-server	> tasks > ! main.yaml							
Q	! site.yaml playbooks	51	- nam	ne: Setup meta-o	data server							
/	🐡 ContainerFile playbooks/roles/i	55	– n	ame: Creates me	eta-data server container							
و٢	\$ docker-entrypoint.sh playboo	56	56 ansible.builtin.copy:									
ō	imes $$! main.yaml <code>playbooks/roles/met</code>	57		content:								
	~ 무제 다 다 ひ @	58		[Unit]								
₽́	> inventory	59		Description=	=meta-data server contain							
	\checkmark playbooks	60		Documentatio	on=https://eos-docs.web.c							
Ľ_⊗	✓ group_vars/all	61		Wants=netwo	rk-online.target							
	! vars.yaml	62		After=netwo	rk-online.target							
Б	✓ roles	63		RequiresMour	ntsFor=%t/containers							
	> image-builder	64										
	> message-queue	65		[Service]								
	✓ meta-data-server	66		Type=notify								
	> defaults	67		NotifvAcces	s=all							
	> handlers	68		SvslogIdent	ifier=container-mgm							
	✓ tasks	69		Restart=on-1	failure							
		70		RestartSec="	55							
		71			Sec=70							
		72		Environment:	=PODMAN_SYSTEMD_UNIT=%n							
		72		Environment:	=RIN_DTR=%+							
		7/		EvecStartPre	-// hin/rm _f &t/%n ctr_id							
	ansible of a	75			sr/bin/nodman stopigno							
	s cleanun sh	76		ExecStart-/	sr/bin/podman runidf							
	€ memo txt				ver-"journald"log-opt							
	≡ podman run cmd.txt	78		ulimit r	memlock = 1 = 1 = -1							
	(i) README.md	70			$red-true - stop_timeout-3$							
	! requirements.yaml	20 20			-inc locknetwork-bast							
		00 01			sonfig base directory							
		10			[config base directory							
\bigcirc		02			(config base_directory							
8		03			[{ config base_directory							
~~~	· 개요	84 05		#vo tume	<pre>{{ config_base_directory {{ config_base_directory }</pre>							
505	> 타임라이	85		volume -	<pre>({ config_base_directory )</pre>							
<u> </u>		80		volume	a contid base directory							





- Further work on EOS deployment playbook to run on AWX system
- Expanding EOS Disk for ALICE further up to 7.5 PB to meet pledges
  - FST nodes running on bare metals (2PB) to be decommissioned
    - Group draining could help to vacate there FSTs
- Updating EOS CDS to v5 as well as upgrading to EL9 flavour
  - Heavy revisions required on CDS Docker deployment



### HPC for ALICE @ KISTI **Collaboration with KISTI Nurion Team & CBNU (a member of KoALICE)**

### **Central Services**

- TaskQueue
- Catalog



**JA** submission

Monitoring

Monitoring





### Considerations

- A "*Knight Landing*" node has 68 cores and 96GB memory (1.4GB RAM per core)
- Assuming 100MB/s bandwidth per single job slot, (may) start with 1,020 cores (15 nodes) for testing

### **ALICE HPC Requirements**

- Preemption policy allowing backfilling
- Java compatibility (ALICE Job Agent runs on JVM)
- CernVM-FS
- 2GB RAM & 10GB space per single core job (resources proportional to # of cores)

### **Central Services**

- TaskQueue
- Catalog



**JA** submission

Monitoring



- Bypassing external authentication procedure that allows (direct) submission from CE (or VO-Box) @ GSDC
- Pool accounts (or dedicated one to be mapped)
- Scratch storage (spooling input files, storing output)
- autofs for CernVM-FS mount
- Java Runtime to run JVM binaries on CernVM-FS
- Allowing out-going connectivity

- Standalone VO-Box (independent of T1) Unique Site identifier (e.g. KISTI_Nurion)
- Standalone CE (independent of T1) Compatible to HPC batch scheduler PBS Pro + ARC-CE (or direct submission?)
- Accessible SEs @ GSDC
- Standalone (load-balanced) Squid-proxy
- •





### **Prototyping Site Services KISTI_GSDC_Nurion**





## **Current Test-bed Status**



### Successful Jobs





24



# Summary

- KISTI Tier-1 for ALICE experiment has been operating without critical issues
  - Configuration change made for HTCondor to accept multi-core as well as whole-node submission jobs
  - New and powerful machines were deployed to meet CPU pledges
  - Successful participation to WLCG DC24 in early this year
  - OPN backup network and storage OS issues affected site availability (re-computation approved)
- Major EOS disk deployment with container technology on EL9 flavour has been successful
  - With efforts on deployment automation and systemd integration
- KISTI HPC Project has conducted in collaboration with a KoALICE member and KISTI Nurion Team
  - Successful jobs have been observed since March this year