

Science and Technology Facilities Council

Evolving Monitoring and Operations services at RAL

<u>Christos Nikitas</u> Will Furnell Adam Laverack

About me

- My name is Christos
- Studying Computer science at the University of Manchester.
 - Completed 2 Years Undergraduate work, and now doing a year in industry at STFC
- Systems Administrator in Tier 1 Production team at RAL
- Currently working remotely from Cyprus 4.
 - Started work here in July so I have never gone to RAL to work.





Motivation

- There are an ever-increasing number of services running.
 Increasing service level expectations from users.
- Our operations and monitoring services are becoming outdated and cannot meet up with the demands.
 - Our time series monitoring service was still running with spinning disks, which made query speeds very slow



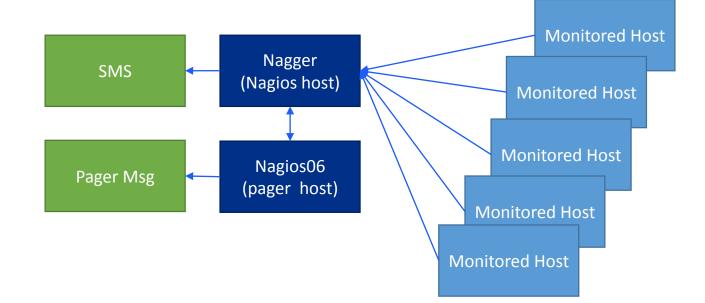
Changes Overview

- Nagios to Icinga
 - Alert Monitoring
- Chenda -> OpsGenie
 - Callout System
- InfluxDB
 - Time Series Database
- RT -> JIRA
 - Project management and Issue Tracking



Alert Monitoring

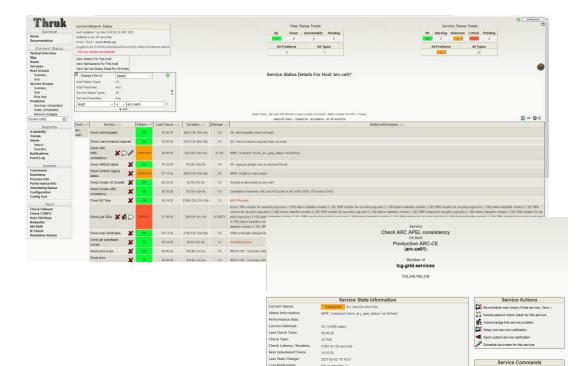
- Nagios Current Alerting Tool.
 - Perform periodic checks on hosts, alerting if something is wrong
- Provides SMS and Pager notification.
- Provides alerting support for the Tier-1 On-call (Production) team.





Nagios

- VERY old version -Nagios(v3.5.1) released in 2013. (This is when iPhone 5s first introduced Touch ID and the galaxy s4 came out)
- Hardware of similar vintage (probably uncomfortably close to selfdecommissioning)
- Single point of failure for alerting (runs on a single host)



Is This Service Flat

In Notification Peris

Last Update:

Notifications

Enable active checks of this service

¥

Stop accepting passive checks for this e

Disable notifications for this servi

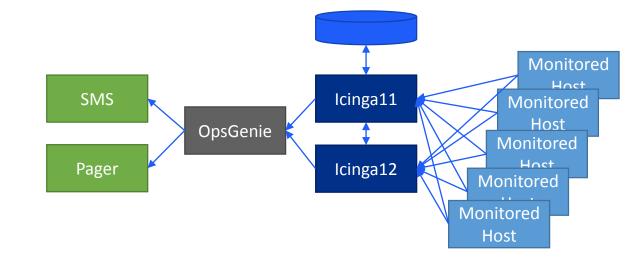
Disable event handler for this servi



lcinga

- Not antiquated! (latest version came out in 2020)
- Integration with SCD configurations management system (Aquilon)
- Consolidation of multiple services (Provides a single SCD wide available alerting service)
- Consolidation of Functionality (Removal of "check creep"/duplication)
- HA cluster for reliability (No single point of failure)
- OpsGenie providing SMS/Pager alerting





		Report Ander Strong Ander Strong Ander Strong Ander Strong		
		C. Search X Name Author Timeframe Date Created Date Modified Hostname	Day	SLA in %
		III Deshboard	2019-04-18	100
		Problems Company Kings Services KingsAdmin Last Day 2019-04-17 13:24 2019-04-17 14:19 c1-mysql-1	2019-04-18	100
Servi	ces History 🗸 O 🗙	Host Service Services History V O X 2019-04-17 18:23 c1-web-1	2019-04-18	100
UP	icinga2.vagrant.demo.icinga.com	UP icinga2.vagrant.demo.icinga.com 62-mysql-2	2019-04-18	12.06
5m 28s	:1 127.0.0.1	07 11 for 6m 27s 11 1270.0.1 c2-web-1	2019-04-18	12.11
ices: 🚺		OK Service: lead cube-app-stage-0	2019-04-18	97.95
	disk DISK (SITICAL - free space: / 38608 MB (92.03)	Check now OComment Q Notification #Downtime cube-app-stage-1	2019-04-18	99.59
TICAL m 1s	DISK CRITICAL - free space: / 34664 NB (92.63%	Plugin Output cube app-stage-2	2019-04-18	98.28
	<pre>inode=100%); /tmp/vagrant-puppet/modules=0dbcabbc6c141041daad0321ce9f= f712 15983 NB (6.60% inode=100%);</pre>	OK - load average: 0.54, 0.21, 0.29 cube-app-stage-3	2019-04-18	100
OK m 56s	disk / DISK OK - free space: / 38608 NB (92-03% inode=100%);	Graphs cube-spp-stage-4	2019-04-18	98.29
ок	http	5 minutes * Hours * Days * Weeks * Months * Years * 0 640 cube-app-stage-5	2019-04-18	100
m 10s	HTTP OH: HTTP/1.1 301 Moved Permanently - 458 bytes in 0.001 second response time	0.50 cube-app-stage-6	2019-04-18	99.33
OK im 8s	icings Icings 2 has been running for 5 minutes and 32 seconds. Version: v2.11.0-rcl=143-g3560931	6.40 cube-upp-trape-7	2019-04-18	100
OK (m 9s	load 0K - load average: 0.56, 0.21, 0.29	0.20 cube-app-stage-8	2019-04-18	99.01
ОК	ping4	0.20 (ute-app-stape-9	2019-04-18	99.93
n 23s OK	pING OF - Packet loss = 0%, HIA = 0.05 ms	0 11.25 11.26 11.27 11.28 cube-d8-0	2019-04-18	99.31
n 32s OK	FING OK - Packet loss = 0%, BZA = 0.07 ms	Load 15 Load 5 Load 1 cube-db-1	2019-04-18	99.75
m 18s OK	FROCE OE: 183 processes	Problem handling comment cube db-2	2019-04-18	99.33
im 23s	SSH OK - OpenSSH_7.4 (protocol 2.0)	Dewntimes Fixed downtime by Icingaadmin scheduled on Sep 17 02:00 expires on	1010-04-10	
OK m 39s	swap SHAP OX - 94% free (1915 NB out of 2047 NB)	Schadule downtime Sep 17 03:00 created 4m 55s ago Scheduled downtime for backup		
OK im 34s	USERS OK - 1 users currently logged in	Actions Business Impact Inspect		
		Performance data		
		Label Value Warning Critical		
		load1 0.56 5.00 10.00		
		lead5 0.21 4.00 6.00		

On Call System

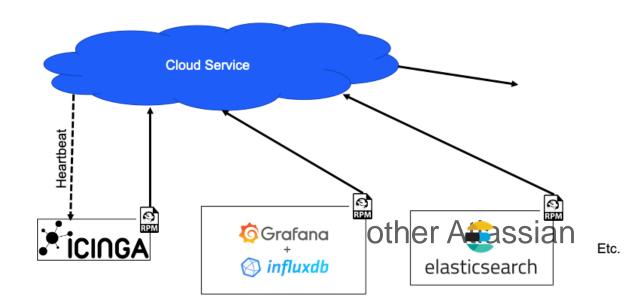
- Our Callout system needs to be integrated with Icinga
 - Sends SMS, emails, pager notifications whenever an alert pops up in icinga.
- Nagios's bespoke call-out functionality (Chenda[1]) is being replaced.
 - It was written by us and is difficult to maintain.
 - Involves a lot of scripting and going into individual files to get it working.
 - Difficult to use/change mutated rather than evolved!
 - Seriously lacking in features

[1] The Chanda is a cylindrical percussion instrument originating in the state of Kerala and widely used in Tulu Nadu of Karnataka and Tamil Nadu in India.



On Call Evaluation

- In Q4 2020, we did an evaluation of cloud based on call systems.
 - OpsGenie
 - PagerDuty
- Both were very capable.
 - OpsGenie was cheaper
 - Potentially better integrated services





OpsGenie

- Provides a nice UI
- Can implement escalation policies
- On-call Rota
- Can integrate with many services(Slack, Jira)
- Mobile app
- Actually cheaper than the previous pager system.

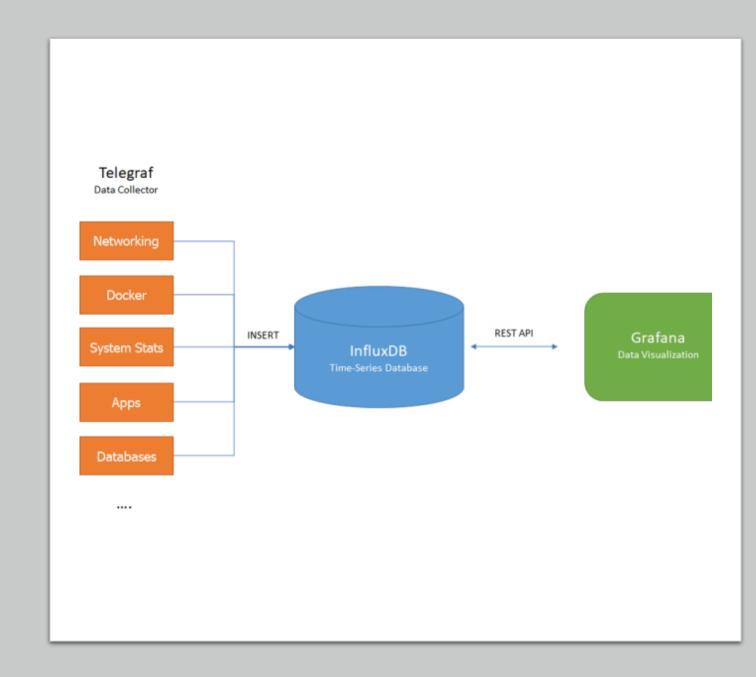


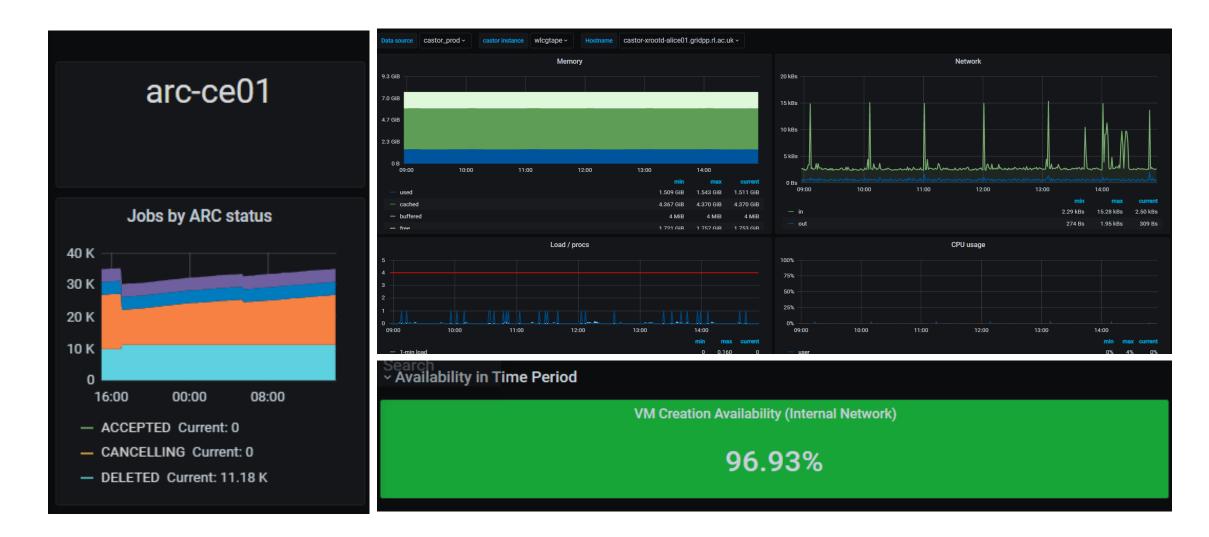
{₽} s1			Select		? Search	Save	
	x1	#11 x1	DBA Team	foL foL	Ack Close ••• Ack Close ••• Assign Escalate to Next Add Team Add Recipient Snooze Delete		
			Error rates increased significantly in the API layer. See rolled up alerts for details. API BA Team		OPEN Mar 27, 2019 11:19 AM		
	0	#7 x1			OPEN Mar 26, 2019 6:00 PM		

	Schedule	(-04:00) EI	OT Eastern Ti	me (US & Ca	nada)									• ~
Today <	> Au	g 12 - Au	g 25				1 Day 1	Week 2	Weeks	1 Month	Ca	lendar	Tin	neline
Rotations	Add rota	tion												
	8/12 Mo	8/13 Tu	8/14 We	8/15 Th	8/16 Fr	8/17 Sa	8/18 Su	8/19 Mo	8/20 Tu	8/21 We	8/22 Th	8/23 Fr	8/24 Sa	8/25 Su
Busin Ø	U	U	U	U	U			U	U	U	U	U		
# After Ø	U	U5	U5	U5	U5	l	Jser 5		U6	U6	U6	U6 [Use	er 6
Overrides -	Add over	rride Tak	e on-call f	or an hour										
	8/12 Mo	8/13 Tu	8/14 We	8/15 Th	8/16 Fr	8/17 Sa	8/18 Su	8/19 Mo	8/20 Tu	8/21 We	8/22 Th	8/23 Fr	8/24 Sa	8/25 Su
Final schedul	e													
	8/12 Mo	8/13 Tu	8/14 We	8/15 Th	8/16 Fr	8/17 Sa	8/18 Su	8/19 Mo	8/20 Tu	8/21 We	8/22 Th	8/23 Fr	8/24 Sa	8/25 Su
Business's f	U	U	U	U	U			U	U	U	U	U		
		U5	U5	U5	U5		Jser 5		U6	U6	U6	U6	Use	- C

- InfluxDB is our Time Series Monitoring Database
- Used as part of the TIG stack
 - Telegraf: Collects metrics and sends them to influx
 - InfluxDB: Stores the metrics sent out by telegraf
 - Grafana: Displays the metric data in a nice UI. Can also make alerts.





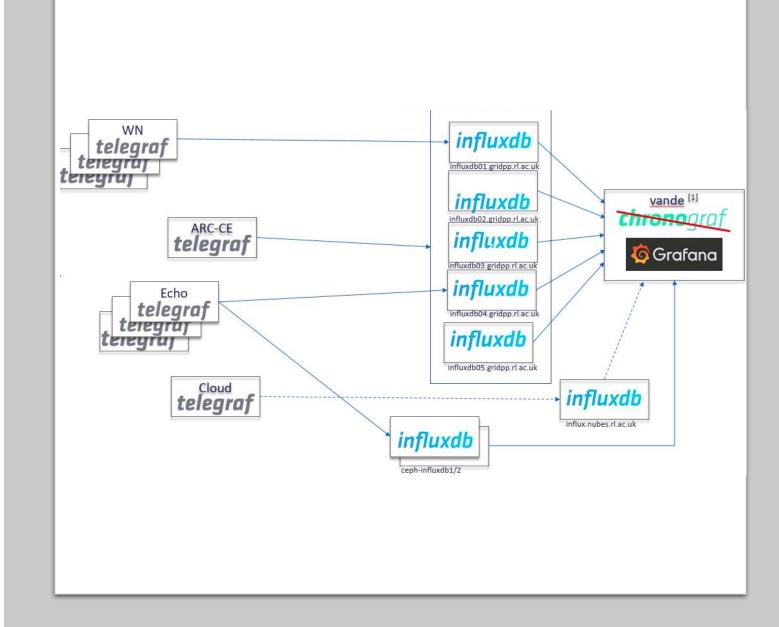




Some Examples

- Up until recently we had 5 influxdb servers running on "influxdb{01-05}.gridpp.rl.ac.uk".
 - Every database was only stored once in the hosts.
 - Some hosts had 80GB worth of data while others had 3GB
 - Some hosts had 200GB storage capacity while others had 1TB
 - Some hosts use spinning disk drives

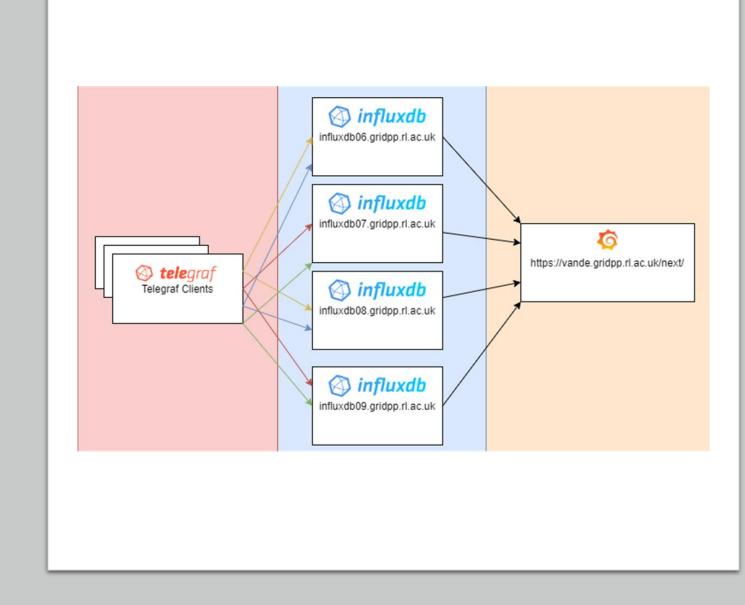




- We are replacing the old servers with 5 new ones. "influxdb{06-10}.gridpp.rl.ac.uk".
 - Store every database twice to increase high availability
 - Each has 900GB of storage
 - All hosts use SSDs
 - Data is spread out more evenly



Technology **Facilities** Council



- Another major benefit from this migration is the change of retention policies and data granularity. (How long data is kept and its resolution)
 - We used to store data at 1-minute intervals for 31 days.
 - Now we keep data like this
 - raw data for 24h
 - Im summary for 31d
 - 10m summary for 6months
 - 60m summary for 5 years
 - More accurate metrics and stored for a much greater amount of time.



RT -> JIRA

- STFC Cloud team did an evaluation of ticket systems.
- RT wasn't able to cater for their growing use case:
 - Serviceable, but becoming outdated
 - Limited metrics production
 - No feedback functionality
 - No Integration with tools like Slack, JIRA, Confluence
 - Meh UI
- JIRA Service Desk was chosen
 - On premise license currently for 25 users (with academic discount)
 - Other services (e.g. Tier-1) will be migrating in 2021.



Homepage Search Reports Articles Tools Logged in as chrisNik						RT for Gridpp	>> PFR	
RT at a glance						New ticket in 3d	~	Search
								Edit
Bookmarked Tickets				Edit	∧ My reminders			
 10 highest priority tickets I own 				Edit	∧ Queue list			Edit
					Queue	new	open	stalled
					3d	-	-	-
▲ 10 newest unowned tickets				Edit	Castor	21	6	
					Ceph	20	31	-
# Subject	Queue	Status	Created		Ceph-Disk	99	81	3
398677 UPS: Passed a self-test.	Fabric	new	11 minutes ago	Take	Ceph-Tasks	1	4	6
398676 IUPS: Started a self-test	Fabric	new	12 minutes ann	Taka	Certificate Requests			-

Create a new ticket in Production

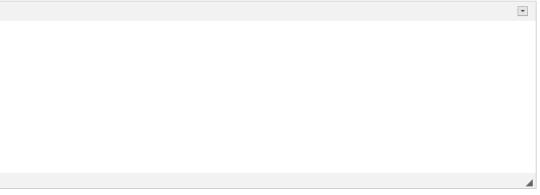
398672 [ral-tier1][PROBLEM] check_service ncm-cdispd on xrootd-cms-uk.gridpp.rl.ac.uk is CRITICAL!	
398671 Nagios issued service alarm Tier1_service_OPN_CERN_Link_Cluster_check_on_host_nagger [NoCallout]	
398669 [ral-tier1] lcinga issued and cleared ha-check_castor_ns_showq_down on fac-central-services01.fds.rl.ac.uk	Requestors
398667 [ral-tier1] lcinga issued and cleared ha-check_castor_ns_showq_down on fac-central-services01.fds.rl.ac.uk	
398666 [ral-tier1][PROBLEM] ha-load db server on dbsdb11.fds.rl.ac.uk is CRITICAL!	Co
Quick ticket creation	Admin Co
Subject:	
Queue: 3d V Owner: Me V	
Requestors: christos nikitas@stfc.ac.uk	

rs:	christos.nikitas@stfc.ac.uk]
Cc:		
	(Sends a carbon-copy of this update to a comma-delimited list of email addresses. These people will receive future updates.)	r
Cc:		
	(Sends a carbon-copy of this update to a comma-delimited list of administrative email addresses. These people will receive futu	ure updates.)

Subject:

Sign 🗌 using Queue's key Encrypt 🗌

Describe the issue below:



Attach:

Drop files here or click to attach



398675 [ral-tier1][PROBLEM] ha-icinga on xrootd-cms-uk.gridpp.rl.ac.uk is WARNING! 398674 [ral-tier1][PROBLEM] ha-load db server on dbsdb11.fds.rl.ac.uk is CRITICAL! 398673 [ral-tier1][PROBLEM] ha-load db server on dbsdb11.fds.rl.ac.uk is WARNING!

Create

Content:

STFCCLOUD-2 💉 🕜 🗘 🚺 👍 Jira Service Desk QUEUES Dashboards 🗙 Projects 🖌 Issues 👻 Create STFC Cloud Support Return to queue John needs help setting up a demo All open 6 STFC Cloud Support Unassigned issues 6 × Workload < 🏠 Export 🖌 Queues TEAM Assigned to me 0 ✓ Details ✓ SLAs Customers 6 Workload Find agents by name or email Q l, Waiting on me 0 -3mo () Time to first response Queues Type: Service Request Status: WAITING FOR SUPP... Reports SLA goals within 1w 🛅 Priority: (View Workflow) 0 Medium Incidents Issues in progress Customers 8 Agent Satisfaction Resolution: Component/s: None Unresolved I, Reported in the la... 0 Raise a request **Beports** Tint, Michael (STFC,RAL,SC) 0 Labels: None People l, Critical 0 KNOWLEDGE BASE Knowledge base ? Assignee Raise a request Article usage Prosser, Andrew (STFC,RAL,SC) ADMIN 0 Service requests 5 Description Unassigned Customer channels Click to add description Assign to me Knowledge base Article effectiveness Invite team l, Due in 24h 0 🚯 Good, Good, John (STFC,RAL,SC) Reporter: 0 Change Customer channels 0 Ø Welcome guide Attachments John CUSTOM (STFC,RAL,SC) I, Ready for implem... 0 Invite team Ward, Jacob (STFC,RAL,SC) 0 Created vs Resolved Drop files to attach, or browse. PROJECT SHORTCUTS Request participants: None L, Emergency change 0 Ø Welcome guide Add a link to useful information for your Time to resolution Organizations: None Dibbo, Alexander (STFC,RAL,SC) 0 whole team to see. Problem 0 0 Votes: Activity SLA met vs breached PROJECT SHORTCUTS + Add link 1 Stop watching Watchers: L, Completed last 3... 0 All Comments Work Log History Activity Add a link to useful information for your Incidents by priority 0 Summers, Martin (STFC,RAL,SC) this issue whole team to see. Recently resolved 0 ✓ ¹⁰ Good, John (STFC,RAL,SC) added a comment - 19/Jan/21 10:59 AM SLA success rate REPORTER + Add link 0 Service Desk request JSD Admin ADMIN Parked 7 Service requests 😝 Quota Request type: This has now been created. + New queue Management Problems by priority 1-7 of 7 agents Waiting for support Customer status: Change by type Channel: Portal Click to add comment T: 1 View customer request @



JIRA Service Desk

- Nicer / More modern UI (Very easy to use)
- Templated tickets
- Integrated Knowledgebase
- Integration with Slack, JIRA and Confluence



CURRENT SERVICE STATUS - 16 March, 2021 UTC Service is operating normally, no recent events Help Center STFC Cloud Support Portal Welcome! You can raise a STFC Cloud Support request from the options provided. What do you need help with? Q Search Request Access **Common Requests** Request access to an existing project. Access STFC Cloud Quota Management Quota Management Request a change to a project quota. Contact Us General Enquiry Get assistance for general problems and guestions FAQ



Science and Technology Facilities Council

Than Kyou

Science and Technology Facilities Council

@STFC_matters

Science and Technology Facilities Council