

Gen³ Webinar

NCI DCFS

Quality Assurance

Processes

Monday, August 24, 2020
12:30 PM - 1:30 PM (CST)



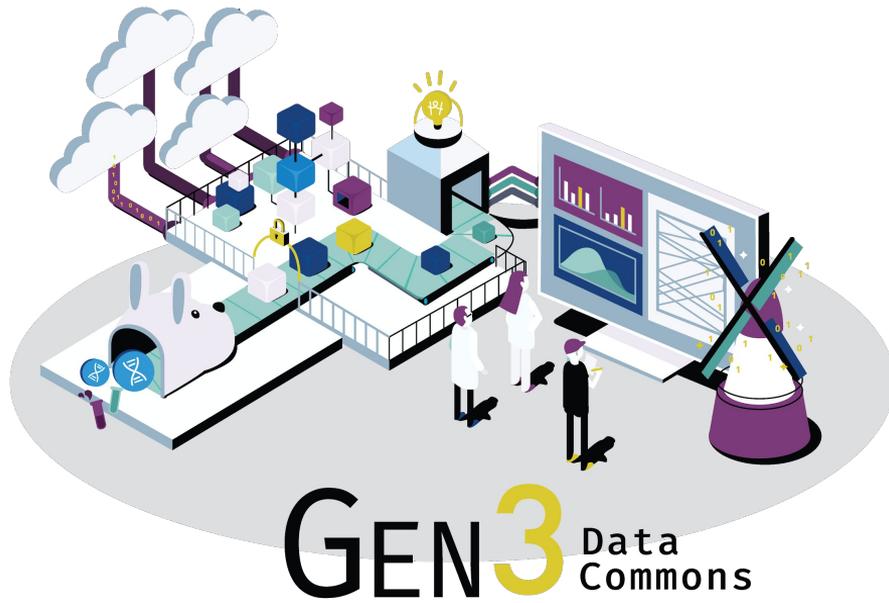
NCI DCFS Quality Assurance Processes

Jiaqi Liu

Marcelo Costa

Andre Paredes

Center for Translational Data Science,
University of Chicago



GEN³ Data Commons

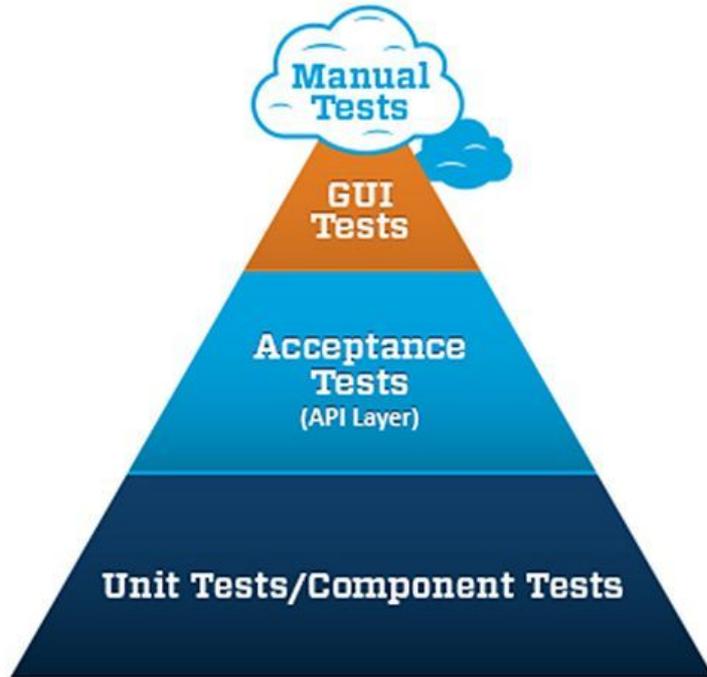
Because Framework Services serve controlled access data on our platforms, we want to ensure we provide reliable and scalable solutions to our users. This presentation comprises an overview of the processes we follow before a release to ensure that DCFS provides a reliable, scalable and secure way to access data.

Robert L. Grossman, Allison Heath, Mark Murphy, Maria Patterson and Walt Wells, A Case for Data Commons Towards Data Science as a Service, IEEE Computing in Science and Engineer, 2016. Source of image: The CDIS, GDC, & OCC data commons infrastructure at the University of Chicago Kenwood Data Center.

- Gen3 is an open source solution for operating data commons.
- NCI Data Commons Framework Services (DCFS) runs an instance of Gen3 Framework Services.
- Gen3 is built and operated by the Center for Translational Data Science at the University of Chicago



- Intro to QA
 - Testing Pyramid / Observability
- How Code is Released
 - Our Release Process / Monthly releases
- How Data is Released
 - Data Indexing / GDC Data Replication
- Testing
 - Automation
 - Observability
 - Independent Infrastructure Tests
 - Load testing
 - Long Running Tests
- Conclusion
 - Managing Quality



- Feature Freeze every 2nd Friday
- Code freeze every 4th Friday
 - The Release is published and PROD PRs are produced.

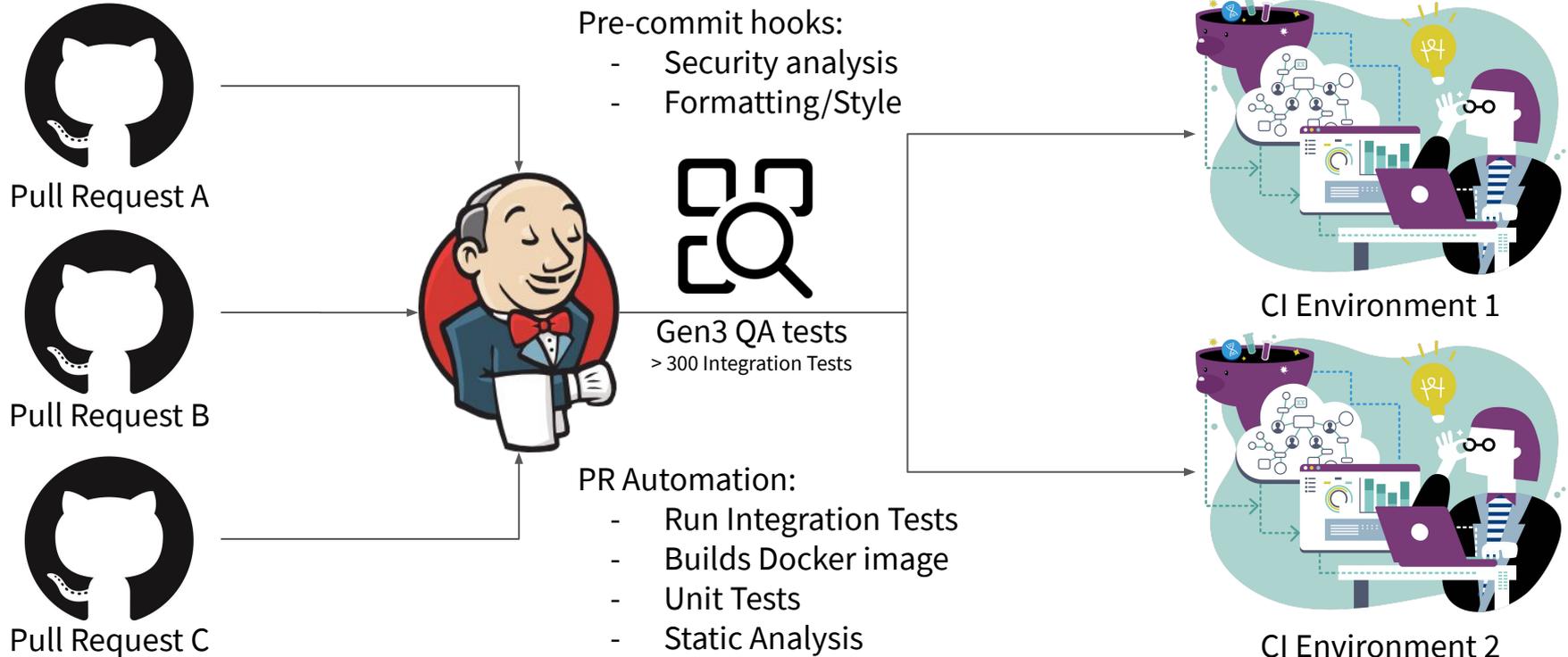
Month A				Month B		
Release Management version N	Cut testing branch version N	Testing version N	Publish Release N	Release Management version N+1	Cut testing branch version N+1

- <https://github.com/uc-cdis/cdis-manifest/tree/master/releases/2020>

NCI CRDC Data Releases

- 08/03/2020: GDC Data Release v25.0 available through DCFS. See more at [GDC](#).
- 07/22/2020: PDC CPTAC3-Discovery CPTAC LUAD Discovery Study - Acetylome available through DCFS. See more at [PDC](#).
- 07/22/2020: PDC CPTAC3-Discovery CPTAC LUAD Discovery Study - CompRef Acetylome available through DCFS. See more at [PDC](#).
- 07/21/2020: ICDC GLIOMA01 available through DCFS. See more at [ICDC](#).
- 06/08/2020: GDC Data Release v24.0 available through DCFS. See more at [GDC](#).
- 06/02/2020: CPTAC CCRCC Discovery Study DIA Proteome available through DCFS. See more at [PDC](#).
- 06/02/2020: PDC CPTAC HNSCC Discovery Study available through DCFS. See more at [PDC](#).
- 06/02/2020: PDC ICPC Academia Sinica LUAD-100 available through DCFS. See more at [PDC](#).
- 05/15/2020: PDC ICPC Early Onset Gastric Cancer dataset available through DCFS. See more at [PDC](#).
- 05/08/2020: CDS LCCC 1108 available through DCFS. See more at [LCCC 1108](#).
- 05/05/2020: GDC Data Release v23.0 available through DCFS. See more at [GDC](#).
- 04/08/2020: CDS PPTC available through DCFS. See more at [PPTC](#).

- Continuous Integration



<https://dcf.gen3.org/software-releases>

NCI DCFS Test Suites - release 2020.07

Status: 0 0 28 0 0 Marks:  

1. Linking accounts - DCF Staging testing for release sign off - PXP-3836:

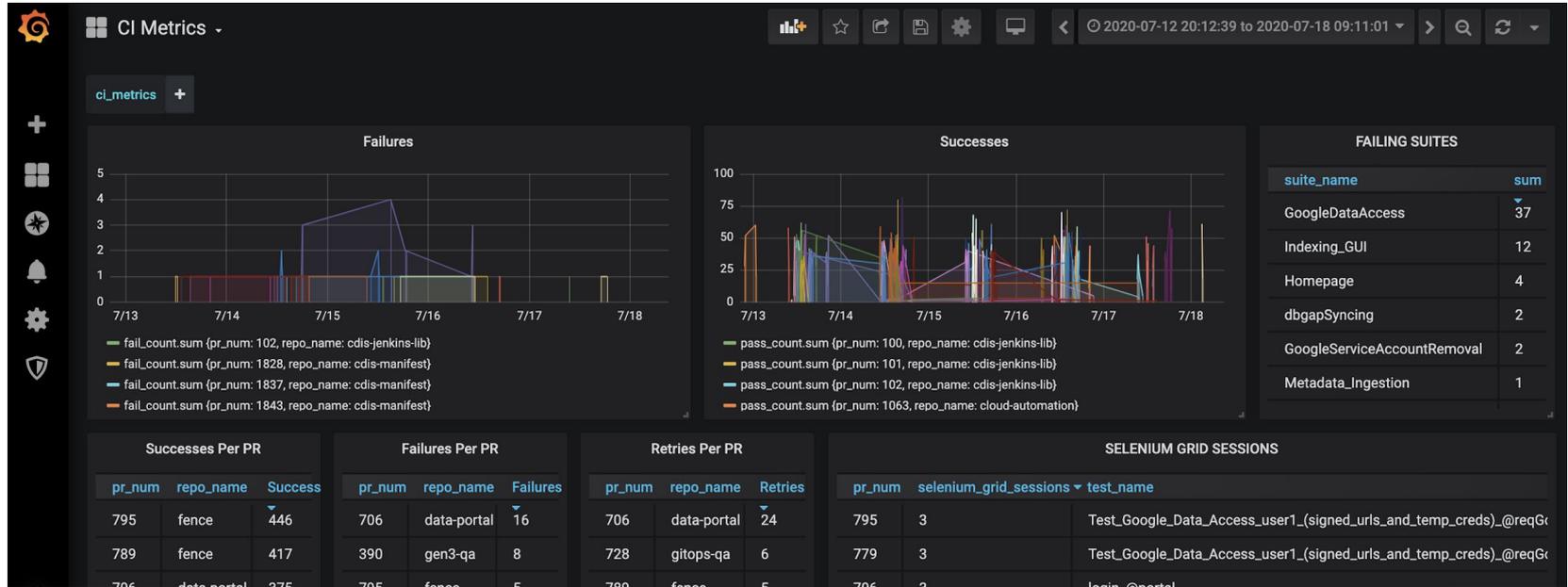
- ✔ #1 Login to <https://nci-crdc-staging.datacommons.io> and check the Project Access list under the Profile page @manual 3s 073
- ✔ #2 Link Google identity to the NIH user @manual 1s 243
- ✔ #3 Unlink Google identity from the NIH user @manual 5s 122
- ✔ #4 Adjust the expiration date of the Google account that has been linked @manual 8s 550

2. Service accounts registration - DCF Staging testing for release sign off - PXP-3836:

- ✔ #1 Register Google IAM Service Account: invalidSvcAccount @manual 8s 576
- ✔ #2 Register Google IAM Service Account: invalidPrjAccess @manual 3s 065
- ✔ #3 Register Google IAM Service Account: existingSvcAccount @manual 6s 915
- ✔ #4 Register Google IAM Service Account: dryRunRegistration @manual 5s 494
- ✔ #5 Get details from registered service account @manual 7s 195
- ✔ #6 Get the ID of the GCP monitor ("fence-service" account) @manual 7s 651
- ✔ #7 Update existing service account @manual 20s 483
- ✔ #8 Get billing GCP projects @manual 6s 530
- ✔ #9 Delete existing service account @manual 7s 764
- ✔ #10 Register Google IAM Service Account: validSvcAccount @manual 11s 254

Example Negative Tests

Testing: Observability

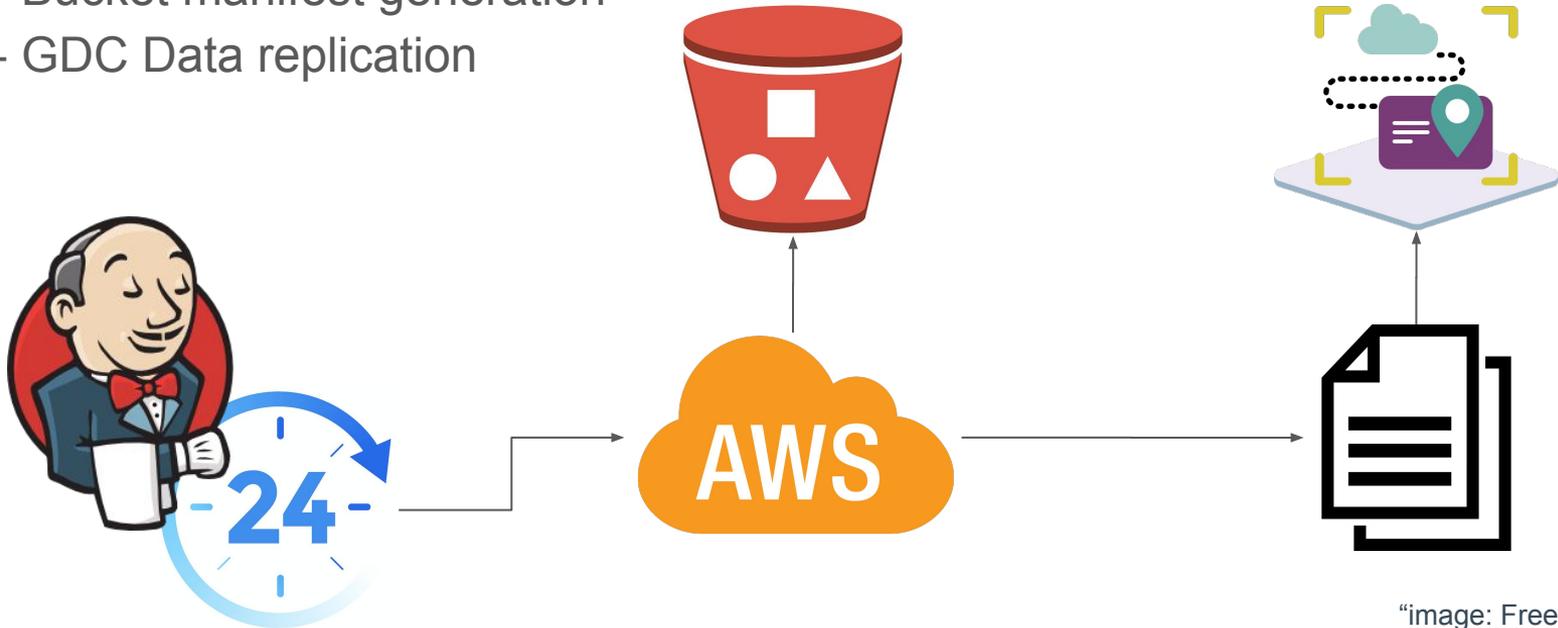


Successes Per PR			Failures Per PR			Retries Per PR			SELENIUM GRID SESSIONS		
pr_num	repo_name	Success	pr_num	repo_name	Failures	pr_num	repo_name	Retries	pr_num	selenium_grid_sessions	test_name
795	fence	446	706	data-portal	16	706	data-portal	24	795	3	Test_Google_Data_Access_user1_(signed_urls_and_temp_creds)_@reqGt
789	fence	417	390	gen3-qa	8	728	gitops-qa	6	779	3	Test_Google_Data_Access_user1_(signed_urls_and_temp_creds)_@reqGt
706	data-portal	275	706	fence	5	789	fence	5	706	3	login_@portal



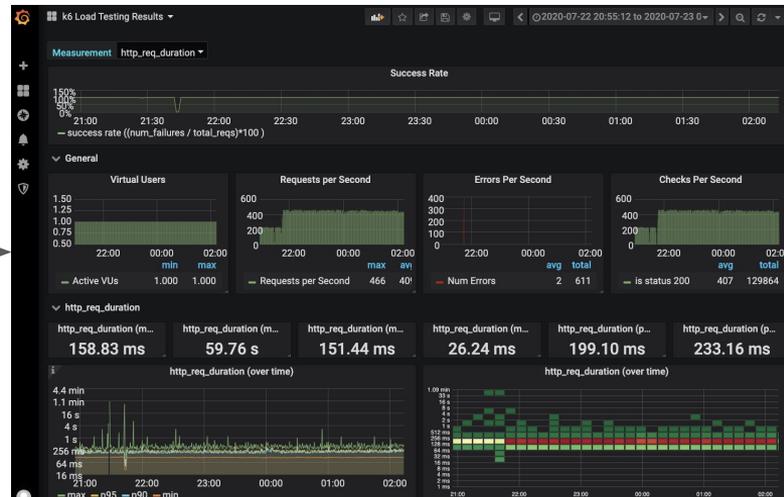
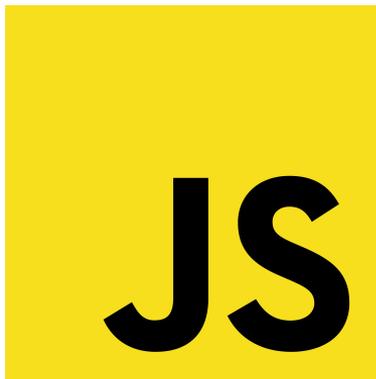
Testing: Independent Infra Tests

- Tests that require Complex / Extensive infrastructure provisioning
- Not suitable for PR checks
- e.g.,
 - Bucket manifest generation
 - GDC Data replication

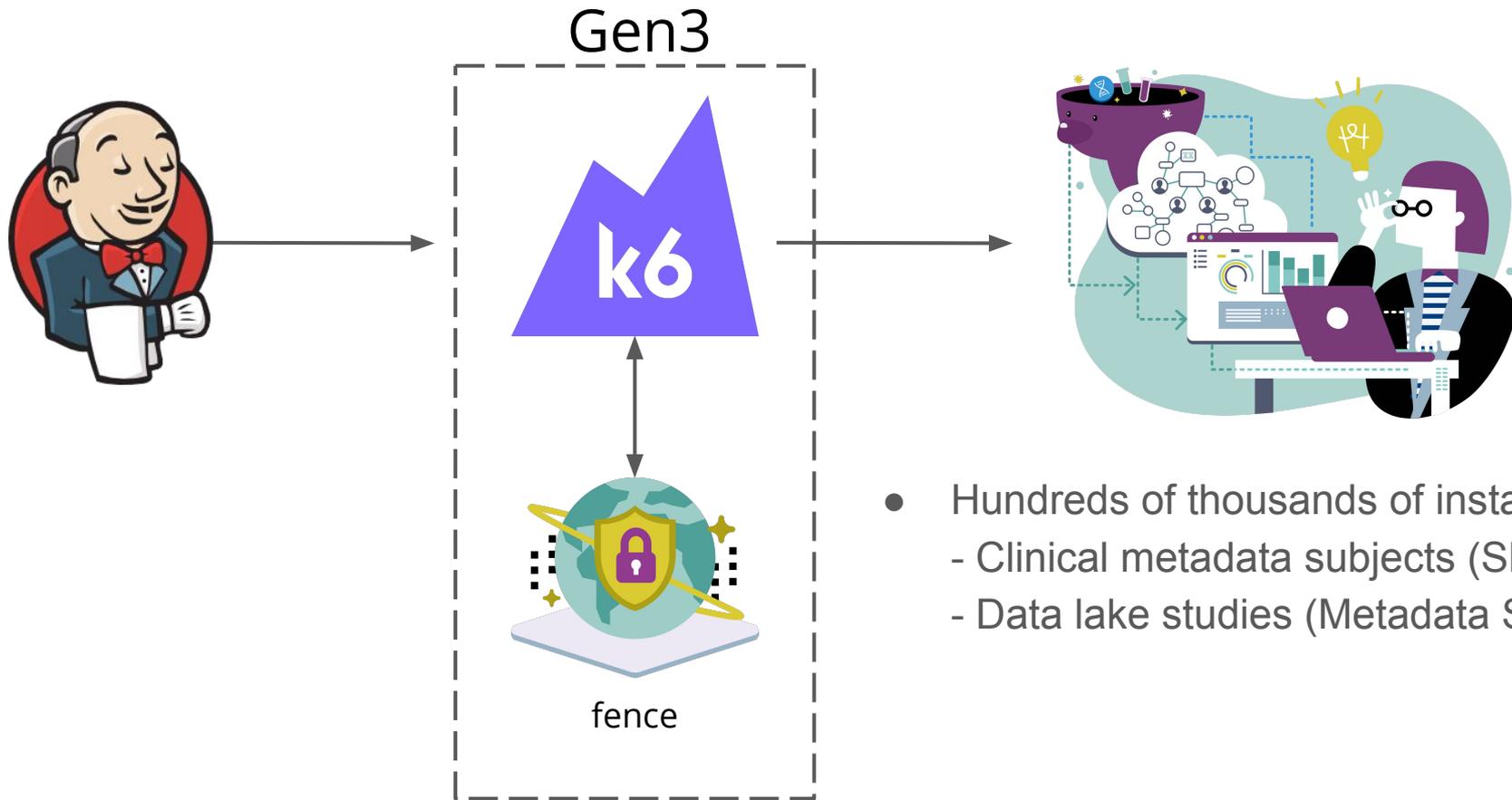


Testing: Load Testing

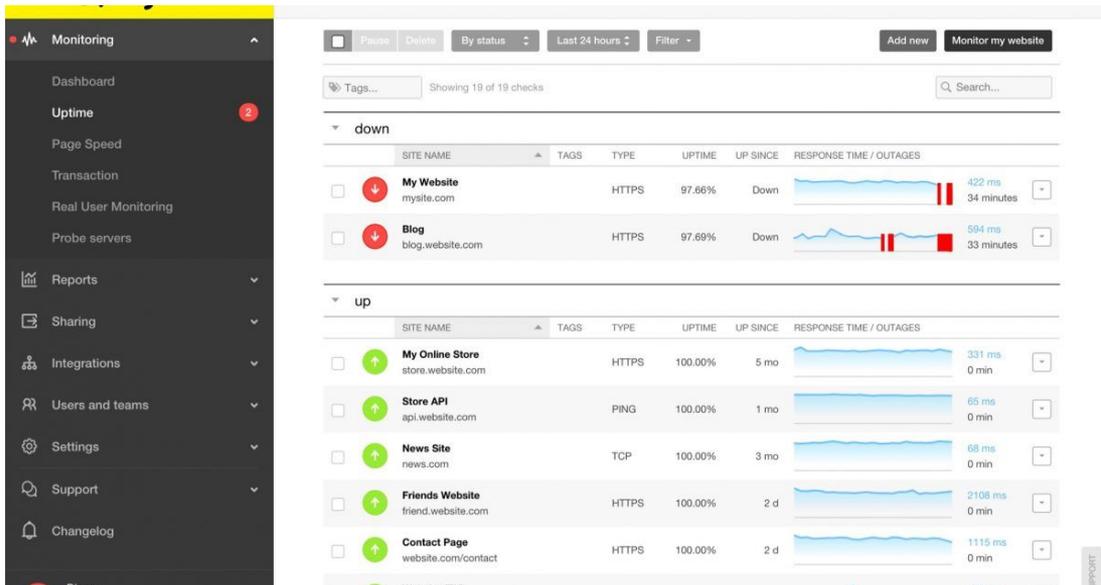
- Benchmarking
- Performance bottleneck investigations
 - HAR (HTTP Archive) testing
 - PreSigned URL Transactions



QA or Staging Environment



- Hundreds of thousands of instances:
 - Clinical metadata subjects (Sheepdog)
 - Data lake studies (Metadata Service)



- Continuous Monitoring
 - Healthchecks
 - Monitor for Kubernetes Pod Health
- Alerting
 - NCI DCFS has alerts for if services are unavailable
- Logging
 - Logs for triaging bugs and issues
- Rollbacks/Hotpatches
 - If an issue is identified in production, we can rollback to a previous working version of the code
 - Bug fixes and hotpatches go through the QA process before they are released to production



- github.com/uc-cdis



- Gen3.org



- Gen3 Community on Slack

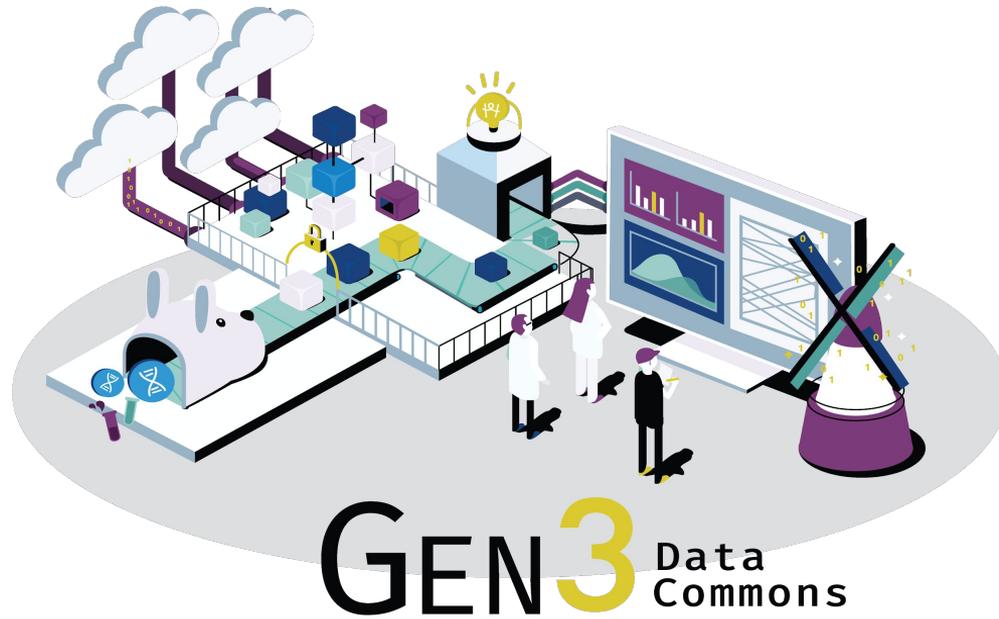


- dcf-support@datacommons.io



- ctds.uchicago.edu

Questions?



GEN3 Data Commons