GEN3 DATA COMMONS

Security and Compliance in Gen3 Data Commons

Gen3 Community Forum
6 September 2023

THE UNIVERSITY OF CHICAGO CENTER FOR TRANSLATIONAL DATA SCIENCE

OCC

KRUMWARE

Australian BioCommons

NeSI New Zealand eScience Infrastructure

The Agenda

- Introduction
- Presentations
  - **A Project Owner's Overview of Security & Compliance for Gen3 Data Commons** (Robert Grossman - Center for Translational Data Science, University of Chicago)
  - **Security and Compliance Practices at CTDS** (Clint Malson - Center for Translational Data Science)
  - **Security practices for Gen3 and applications** (Plamen Martinov - Open Commons Consortium)
  - **Securing Cloud-Native and Kubernetes** (Colin Griffin - Krumware)

GEN3 DATA COMMONS

A Project Owner’s Overview of Security & Compliance for Gen3 Data Commons

Robert Grossman

THE UNIVERSITY OF CHICAGO
CENTER FOR TRANSLATIONAL DATA SCIENCE

TenQuestions

?

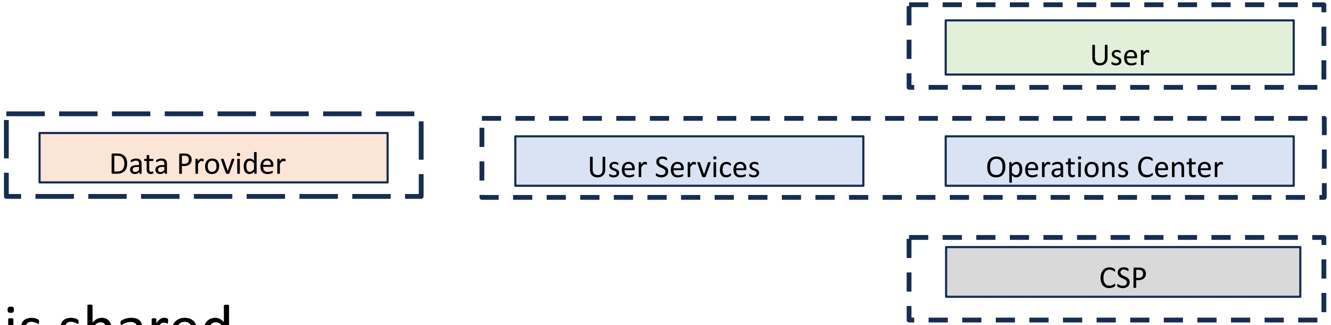
- Perspective of this talk
- If you are the project owner, but not an expert on security and compliance, what are the most important questions to ask when setting up a Gen3 data commons?

Q1. Who reviews approves and signs off on security and compliance?

- The project maybe organized as a research project without a single organization in charge.
- There are may also be multiple organizations involved with overlapping responsibilities and different frameworks.

Q2. What security and compliance framework is being used?

- A security plan is usually long and detailed and follows a proscribed format for describing the policies, procedures and controls.
- You need to know what questions need to be answered and what format is required.



User

Data Provider

User Services

Operations Center

CSP

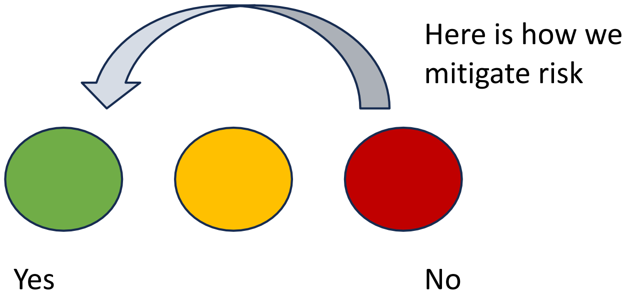
Q4. What is shared responsibility model?

- What are the responsibilities of your organization versus the responsibilities of your partners and service providers?

Q5. Do you know the difference between security / compliance and operational security?

- In addition to your security and compliance plan, you need a team that can focus very practically and effectively on day to day operational security.

Q6. How well do you understand risk management?

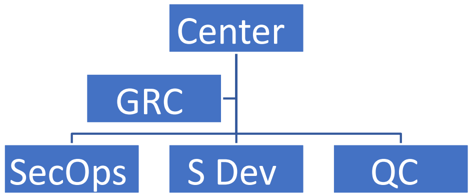


Here is how we mitigate risk

Yes

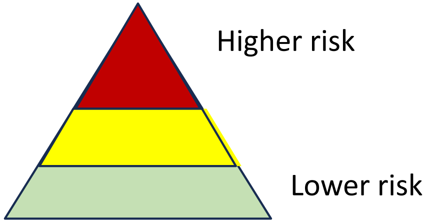
No

- The easiest way to do security and compliance is simply to say no.
- This produces systems that are impossible to use.
- You need someone who can implement processes to reduce and mitigate risk so you can say yes enough so the system is usable.



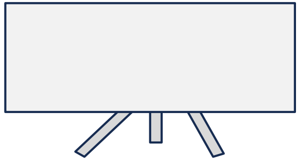
Q7. Do you have a good working internal organizational structure?

- QC, DevOps, SecDevOps, SecOps Governance / Risk / Compliance



Q8. Can you reframe the problem in order to reduce risk?

- Can you work with de-identified data versus health care information?
- Can you shift responsibility from the researcher to the researcher’s organization.



Q9. How good is your training?

- How often do you practice data recovery?
- How often do you do table top exercise with real surprises and real challenges?
- How often do you practice looking for threat

Q10. Do you have people to compare notes with?

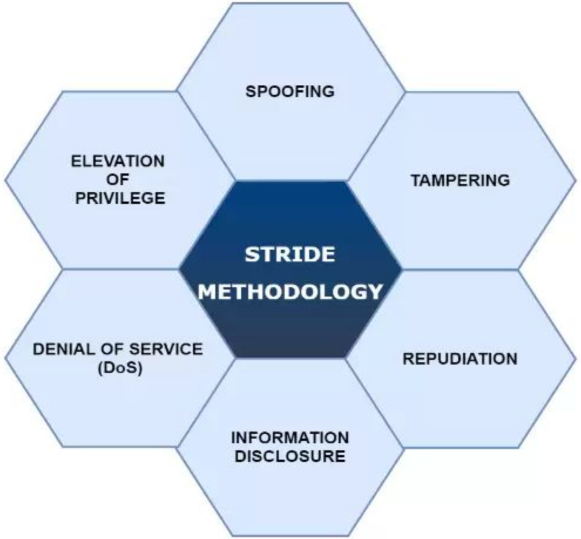
- Make sure to leverage the Gen3 Community forum to exchange information about best practices for setting and operating Gen3 data commons with security and compliance.

Security and Compliance Practices at CTDS

Clint Malson

Deploying a secure Gen3

- Cloud automation / Helm only gets you so far
  - Know your requirements
- Think past your Frameworks and compliance
  - Get in the mind of an attacker or malicious actor
  - I like to start with a Threat Model
  - STRIDE
    - Periodically retest the risk model



Deploying a secure Gen3

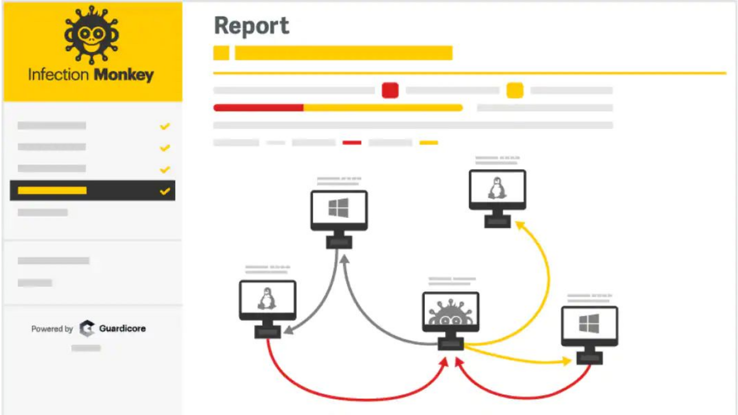
- Utilize that threat model to build your controls and minimize risk
- Boundary Controls
  - Firewalls, WAF’s, VPN’s proxies and more
- Cloud security posture management
  - Set your security target and make risk based decisions on what to focus on
- Encryption
  - Often dictated by your compliance frameworks or the level of data you ingest
- Always keep up to date on Gen3
  - Always updating libraries, patching vulnerabilities

Deploying a secure Gen3

- Behavioral analytics, especially in the cloud, is key!
  - Knowing when and how connections or processes go outside their normal bounds
  - Get a baseline of traffic, connections, usage patterns
- Use this to filter out false positives
- Is data moving in a different way or connecting to a new source
- Helps know when new features are deployed or something has changed
- Sources:
  - Flow logs, login data, IDP’s, web traffic logs, ids/ips, data protection systems,

Deploying a secure Gen3

- Rotate / redeploy your containers frequently
  - Ensure they are running in a non-privileged / read only mode
- Test your security
  - Tools like Infection Monkey help validate your security
  - Launch a container that purposefully does a connection to a unknown source
  - Trust but verify your have built a secure environment.
- Run tabletops where you pose what if's
  - What if we get ddos'd or what if we are attacked in X way what do we do and what tools will tell us / help us.



Security at CTDS

- Behavioral analytics
- SOAR – Security Orchestration Automation and Response
  - Building Automation and workflows
  - Helps determine false positives
  - Threat intelligence feeds
- Focusing on more open source and API-first tooling
- Empowering people and training to make our apps secure from the start
- FedRamp & FISMA
  - Compliance vs. Operational Security
  - Bridging that gap or overhead with tools and open-source

Security practices for Gen3 and applications

Plamen Martinov

About Open Commons Consortium

- The Open Commons Consortium (OCC) is a 501(c)(3) Nonprofit organization, which is a division of the Center for Computational Science Research Inc.
- OCC manages and operates cloud computing platforms, data commons, and data ecosystems to advance scientific, medical, health care, and environmental research for human and societal impact.
- OCC works closely with the University of Chicago, Center for Translational Data Science (CTDS) for many years.
- OCC and CTDS work together on several Gen3 Commons, such as Veterans Administration Oncology Data Commons (VPODC), Blood-profiling Atlas in Cancer (BLOODPAC) and the Pandemic Response Commons (PRC).

OCC Gen3 Applications

Payment Portal

Credit Card Payments for using computing on Gen3

Governance Portal

Onboard users on Gen3 with appropriate permissions

Cost Summary

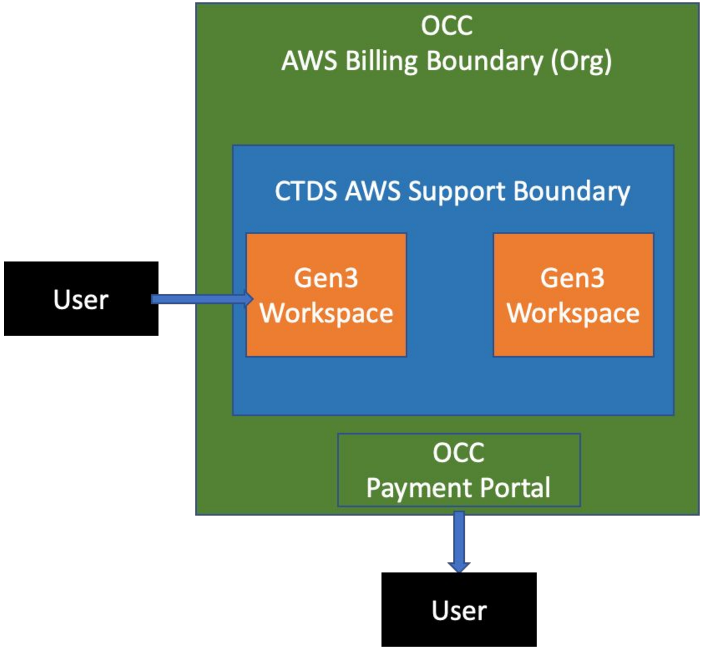
Cost summary reporting for blanket billing on Gen3

Strategic Considerations for Securing Data Commons

- **Identify the Authorizing Official (AO) and organizational requirements**
  - What/who is the legal entity and or individual that will make decisions about the system
- **Perform assessment on what policies, plans and practices that exist or need to be created**
  - Security standards can be used to identify the proper security alignment
  - Security policies
  - System Security Plan
  - Interoperability and MOUs
- **Plan out resources and technical expertise and skills**
  - Define roles and responsibilities
  - Establish a security budget and multiple-year roadmap
- **Start with a good security foundation**
  - Establish risk management, which is a process of identifying vulnerabilities vs threats (*pick an existing framework, don't have to start from scratch*)
  - Develop a CONMON, which is a process to review and prioritized both technical and process shortcoming
  - Establish continuous training and testing for Incident and Disaster management

Secure interoperability models

Outside FedRAMP



OCC
AWS Billing Boundary (Org)

CTDS AWS Support Boundary

Gen3 Workspace

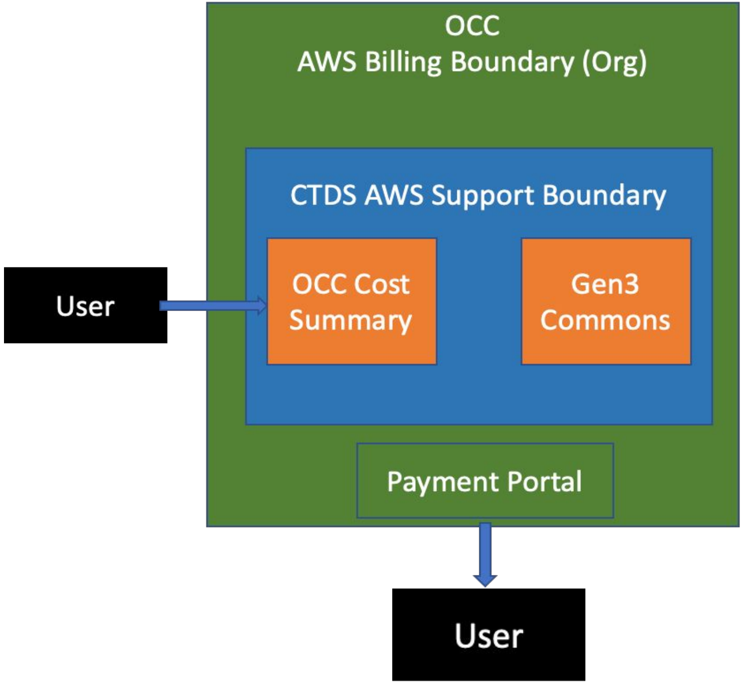
Gen3 Workspace

OCC
Payment Portal

User

User

Inside FedRAMP



OCC
AWS Billing Boundary (Org)

CTDS AWS Support Boundary

OCC Cost Summary

Gen3 Commons

Payment Portal

User

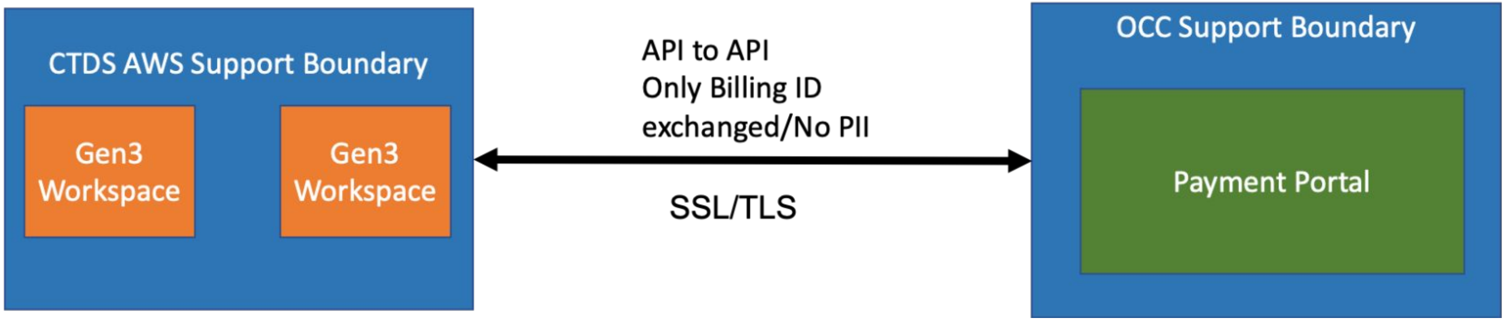
User

Secure interoperability models (cont)

When systems exchange information and there is no PII or other sensitive information both systems can have their own support boundaries

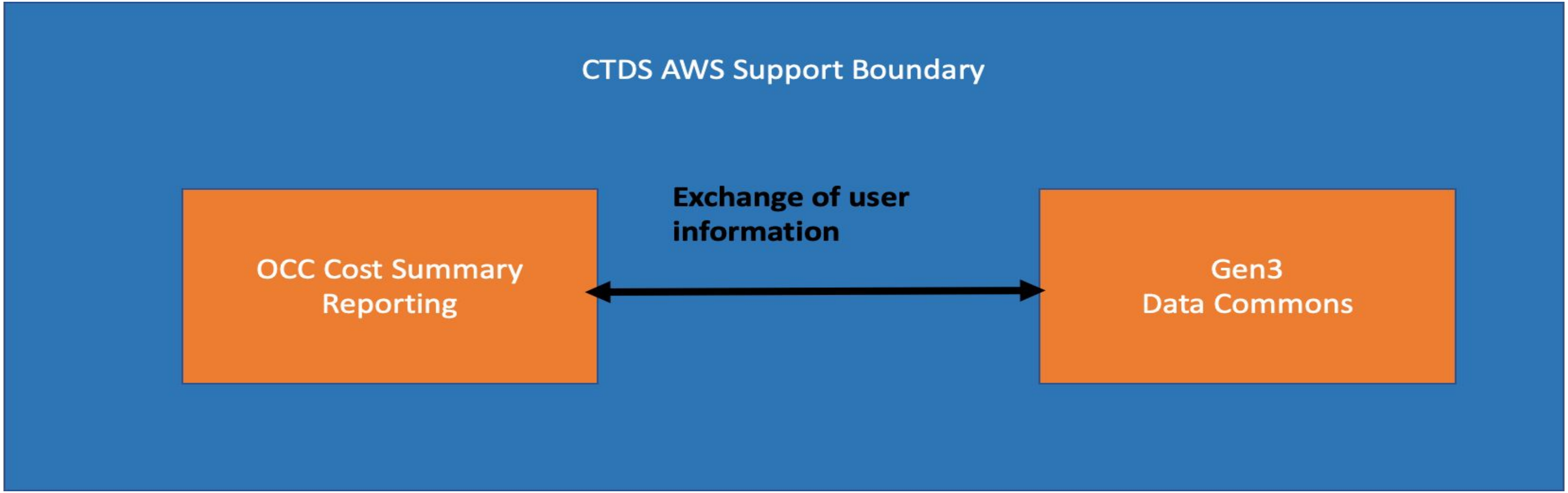
- User signs on via oauth
- User checks for billing ID
- User requests Workspace

- User signs on via OTP
- Completes enrollment documents
- User makes a payment
- User receives BillingID



Secure interoperability models (cont)

When systems exchange PII or other sensitive information both systems must be in the ATO’s authoritative boundy unless both system have like ATOs (i.e. FedRAMP ATO)


CTDS AWS Support Boundary
OCC Cost Summary Reporting
Exchange of user information
Gen3 Data Commons

General Security Considerations

- Before building or connecting your application, work with the organization’s security team to discuss support boundaries and data exchanged between systems.
- Code or configure application in alignment to the support boundary requirements (i.e. AO, security team or security standards).
- When building use similar backend systems that can support FedRAMP or other security ATO processes.
- Routinely set time aside for staff to perform security hygiene practices.
  - Empower staff to take action when they see something wrong
- Enable encryption on staff devices including MFA on each management system.
- Provide general but also role based security training to staff routinely.

General Security Considerations

- Ensure while initially developing and testing code, repos are private.
  - To open code repo scan for keys, passwords and other sensitive variables initially and routinely thereafter
- Design or utilize a risk management process to identify common sense security technology to protect systems and staff from threats.
- Allow time in the development process to scan and remediate vulnerabilities, scan there after routinely
  - Software comp analysis
  - Static code scan
  - Web Application scan
  - Vulnerability scan (infrastructure)
- Implement proactive security controls such as Application Firewall, Runtime Application Self-Protection etc (use cloud native tech where possible).

Securing Cloud-Native and Kubernetes

Colin Griffin

About Krumware

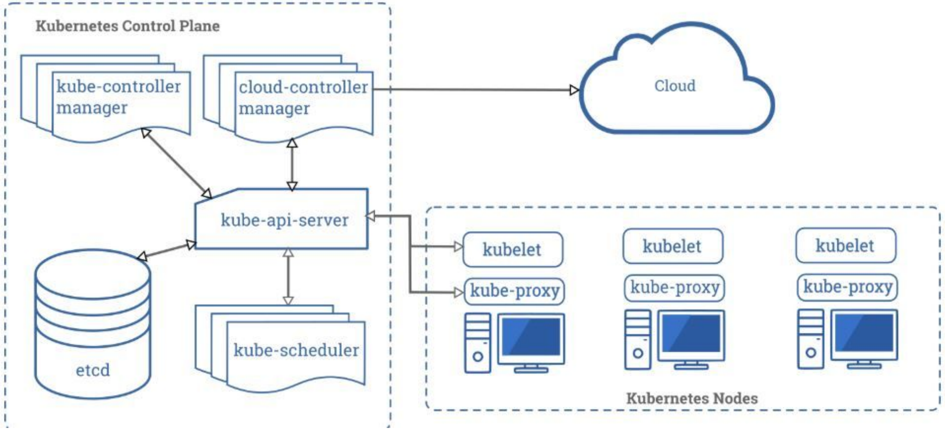
- Cloud-Native software and platform developers
- Members of Cloud-Native Computing Foundation working groups
  - App Delivery
  - Platforms
  - Cloud-Native Maturity Model
- Began involvement through Wake Health to assist with kubernetes

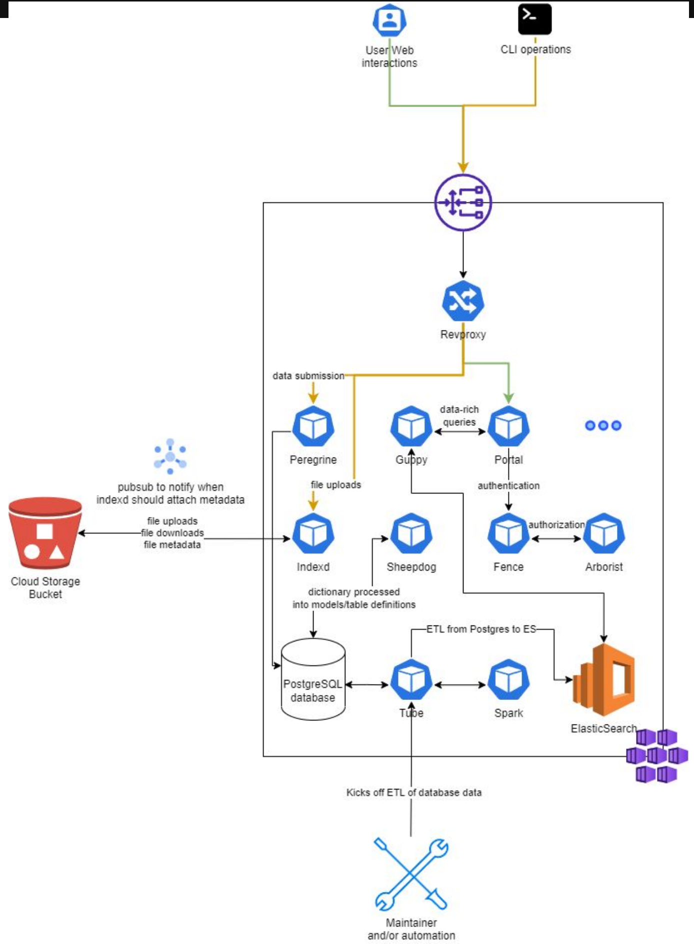
Kubernetes Security Considerations

Kubernetes has many useful features/advantages which can help to remediate issues, but there are risks:

- It enables microservices and containers, which means many small independent machines that are at risk
- Is highly configurable, which can also mean unsecured
- System components of Kubernetes must also be secured
- Securing the network boundary is not enough, each pod is a network boundary
- Traditional security tools may not be able to protect against breaches that come from the inside (man-in-the-middle)
- Many different teams may have access to kubernetes, which is like having access to the entire network if not properly segmented

Kubernetes Security Considerations





Kubernetes Security Considerations

- Tools to help you secure Kubernetes
- OWASP Cheatsheet
  - https://cheatsheetseries.owasp.org/cheatsheets/Kubernetes_Security_Cheat_Sheet.html
- Security tools needed
  - Container scanning
  - Network monitoring and protection
  - Policy management and RBAC (Role Based Access Control)
- Tools we use for Wake Health
  - Rancher
  - Neuvector
  - Cloud-provider tools, like GCloud CLI to authenticate with k8s
  - Revproxy + Fence + Arborist, etc
  - Evaluating: Istio

Shift Security Left

NeuVector provides key features that allow you to shift security & compliance concerns left, covering your entire CI/CD pipeline without obstructing your development flow.

- Image & Registry Scanning
  - NeuVector can be configured to scan your container registry & individual image layers to compare against an actively maintained CVE database.
- Security Policy as Code
  - NeuVector provides Custom Resource Definitions (CRDs) to declaratively define your security rules, segmentation, admission control, Web Application Firewall (WAF) & Data Loss Prevention (DLP) sensors, etc.
  - These may be stored in source control as part of your chart definition and graduated as part of your CI/CD pipeline
- Continuous Scanning & Behavioral Learning
  - NeuVector will actively scan your deployed containers & their communication patterns to detect vulnerabilities and provide generated security profiles and rules based on its detection.
  - Provides real-time threat/abnormality detection & protection.

Open Source Security tools - NeuVector

NeuVector ships with a dashboard reporting crucial details across your network & container deployments

- Network Activity
  - An interactive display which visualizes network traffic outside, into, and amongst your deployments
  - Helps identify unwanted events, communication patterns, & behavior
  - Performs Deep Packet Inspection to analyze content not just connections
- Assets
  - A breakdown of assets within your environment and their scan statuses, including compliance and vulnerability details, statuses, & severities
    - Most notably, Nodes, Containers, & Registries
- Policy
  - Provides a breakdown of various policies, rules, & sensors that were either predefined, user created, or generated as part of NeuVector's behavioral learning capabilities
  - Includes drilldown details of these rules, administration capabilities (add/remove), and import/export functionality in order to update any pertinent declarative definitions

Open Source Security tools - NeuVector

NeuVector Dashboard (Continued)

- Security Risks
  - Detected vulnerabilities detected across your container deployments
    - Detected against a maintained CVE database
  - Manage Vulnerability Profiles
  - Compliance summaries
    - Impacted containers & nodes per compliance item
  - Manage Compliance Profiles
- Notifications
  - Security Event tracking
    - For example, suspicious processes or file permissions within a container
  - Non-critical Event tracking
  - Risk Reporting
    - Down to the container level

Open Source Security tools - NeuVector

- Show NeuVector dashboard (sandbox environment)
- Workflow for AppDev and SecOps collaboration
  - Security Policies as Code
  - Policy generation -> CICD Resource Addition -> Automated deployment/graduation
- Gen3 Security Policy resource file donation
  - Krumware will be donating optional NeuVector security profile (kubernetes resource file) to Gen3
  - End of September

Acknowledgements

- Speakers
  - Robert Grossman - Center for Translational Data Science, University of Chicago
  - Clint Malson - Center for Translational Data Science, University of Chicago
  - Plamen Martinov - Open Commons Consortium
  - Colin Griffin - Krumware
- Gen3 Forum Steering Committee
  - Robert Grossman - Center for Translational Data Science, University of Chicago
  - Steven Manos - Australian BioCommons
  - Claire Rye - New Zealand eScience Infrastructure
  - Plamen Martinov - Open Commons Consortium
  - Michael Fitzsimons - Center for Translational Data Science, University of Chicago