LUNTRY

Безопасность AI, ML и Big Data кластеров

Дмитрий Евдокимов

Основатель и технический директор Luntry

Обо мне

> Я не верю в то, что систему можно сделать надежной и безопасной, не понимая того, как она устроена.

Основатель и технический директор Luntry

Более 15 лет опыта в ИБ

Специализация — безопасность контейнеров и Kubernetes

Автор ТГ-канала k8s(in)security

Эксперт в сфере безопасности контейнерных сред

- Организатор конференции «БеКон» по БЕзопасности КОНтейнеров
- Бывший редактор рубрик в журнале «ХАКЕР», автор серии статей
- Автор курса «Cloud Native безопасность в Kubernetes»
- Член программного комитета CFP DevOpsConf и KUBER CONF

Спикер

VK Kubernetes	Confidence	ZeroNights	БеКон	HITB
DevOpsConf	HackInParis	KuberConf	BlackHat	PHDays
Kazhackstan	HighLoad++	OFFZONE	DevOops	SAS

LUNTRY

О компании Luntry

Luntry — это Комплексная Защита
на всем жизненном цикле
контейнерных приложений
и средств оркестрации на базе Kubernetes

Продукт в реестре Минцифры
https://reestr.digital.gov.ru/reestr/1057835/

Получение сертификата ФСТЭК планируется во втором квартале 2026 года.

LUNTRY.RU

Функциональность Luntry

Домены безопасности Kubernetes

Контроль Kubernetes-ресурсов

Контроль состояния Kubernetes-кластеров

Контроль соответствия кластера стандартам

Управление безопасностью образов

Сетевая безопасность

Анализ прав доступа

Защита Runtime

План вебинара

LUNTRY

01	Введение
02	Специфика в каждом из 7 доменов безопасности Kubernetes
03	Как Luntry помогает решить задачи
04	Выводы

LUNTRY

Введение

Разноплановая AI безопасность

LUNTRY

Безопасность AI	Безопасность от AI	Безопасность с AI
защищаем AI-системы и их lifecycle	защищаем компанию/пользователей от злоупотребления AI	используем AI как инструмент безопасности

Безопасность AI

- AI Infrastructure Security
  - hardware, GPU, cluster, storage, network
- AI Supply Chain & MLOps Security
  - pipelines, registries, checkpoints, dependencies
- AI Data & Privacy Security
  - datasets, labeling, provenance, DLP, retention
- Model / LLM Security
  - adversarial ML, extraction, robustness, safety bypass
- Inference / App / RAG / Agent Security
  - serving, prompts, retrieval, tools, action controls
- AI-Generated Code / SDLC Security
  - secure use of AI in engineering
- Governance / Third-party / Compliance
  - policy, audit, vendor management

LUNTRY

Kubernetes (K8s)

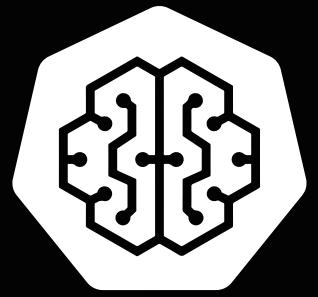
LUNTRY

Оркестратор

Стандарт де-факто для управления контейнеризированной нагрузкой

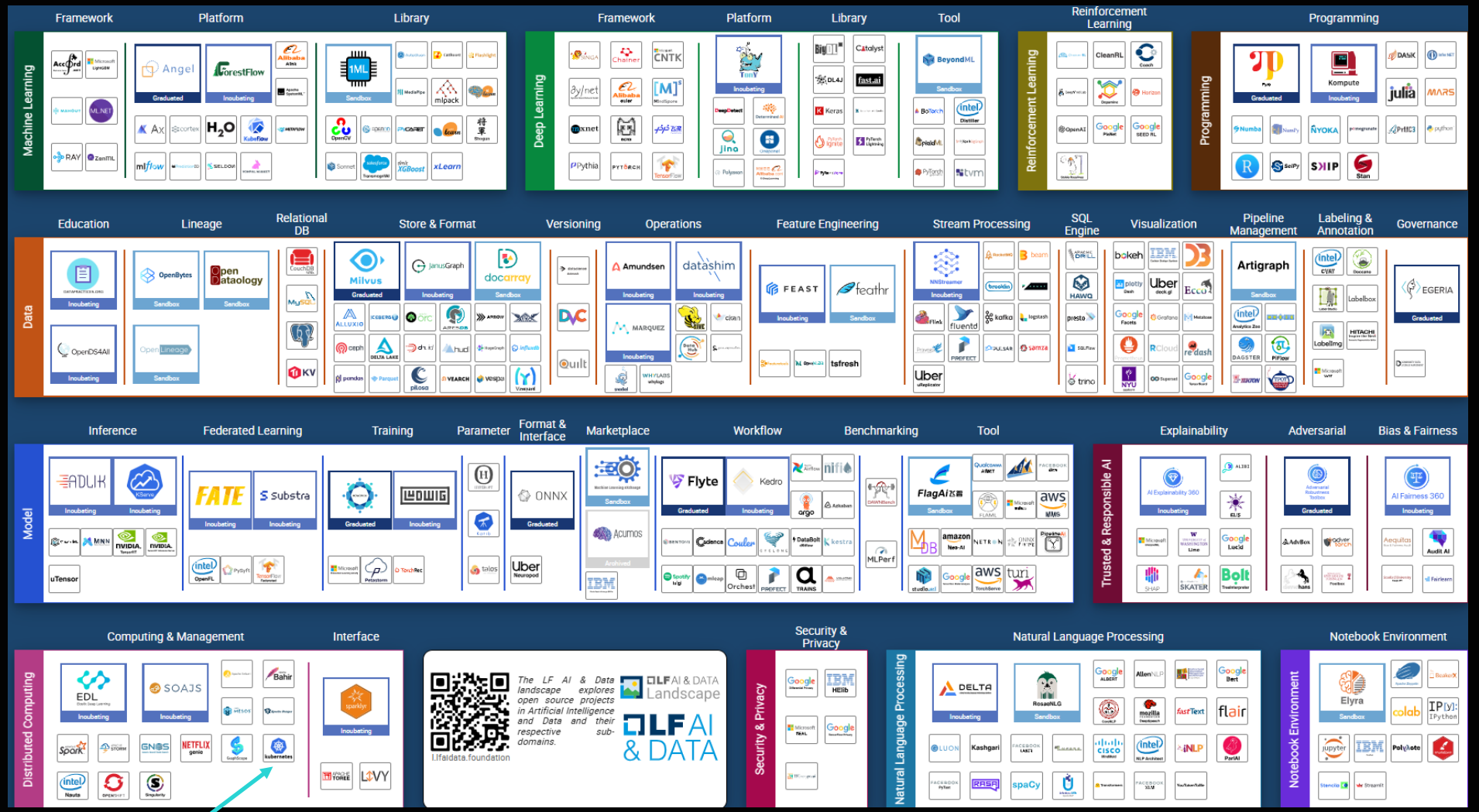
Обеспечивает надежную и масштабируемую среду для обработки больших данных

- Big Data-ворклоады (например, с использованием Hadoop, Spark или AI/ML-пайплайнов) в Kubernetes увеличивают сложность системы
  - Volume (Объем)
  - Velocity (Скорость)
  - Variety (Многообразие)



Kubernetes AI DAY

Kubernetes для задач AI и Data



Kubernetes

[LF AI & Data Foundation Interactive Landscape]

LUNTRY

K8s продолжает адаптироваться для AI

LUNTRY

Node Feature Discovery (NFD)

Данный проект призван помочь определить совместимость образа с конкретной конфигурацией хостовой ОС.

Ведь Nodes в кластере могут отличаться друг от друга:
- Драйверами
- Библиотеками и специализированным ПО
- Фичами ядра

Container Device Interface (CDI)

Спецификация для container runtimes, для поддержки разных сторонних устройств через общепринятую, понятную всем систему плагинов.

Механизм Checkpoint/Restore

Ускорение старта тяжёлых приложений, пауза и возобновление длительных вычислений (ML, batch-задачи), а также более эффективное использование ресурсов кластера.

Проблемы и вызовы для ИБ в Kubernetes

LUNTRY

01 Окружение становится все сложнее

- "Complexity is the worst enemy of security, and our systems are getting more complex all the time.", Bruce Schneier
- "The only thing that ever yielded real security gains was controlling complexity.“,
  Thomas Dullien “Halvar Flake”

02 Разработка стремительно развивается

- Старые подходы к безопасности не работают
- Департаменты разработки, поддержки и безопасности должны работать вместе

03 Kubernetes – это фреймворк

- Уникальные модели нарушителя, модель угроз и поверхность атаки

04 Атакующий по-прежнему на шаг впереди

- 0day, unfixed 1day, backdoors, APT и т.д.

Риски контейнеризации

Смотря как подойти к безопасности ☺

LUNTRY

01

На уровне образов

- уязвимость в компонентах образа
- небезопасная конфигурация
- наличие чувствительных данных
- вредоносное ПО и ПО двойного назначения
- отсутствие подлинности образов

02

На уровне контейнеров

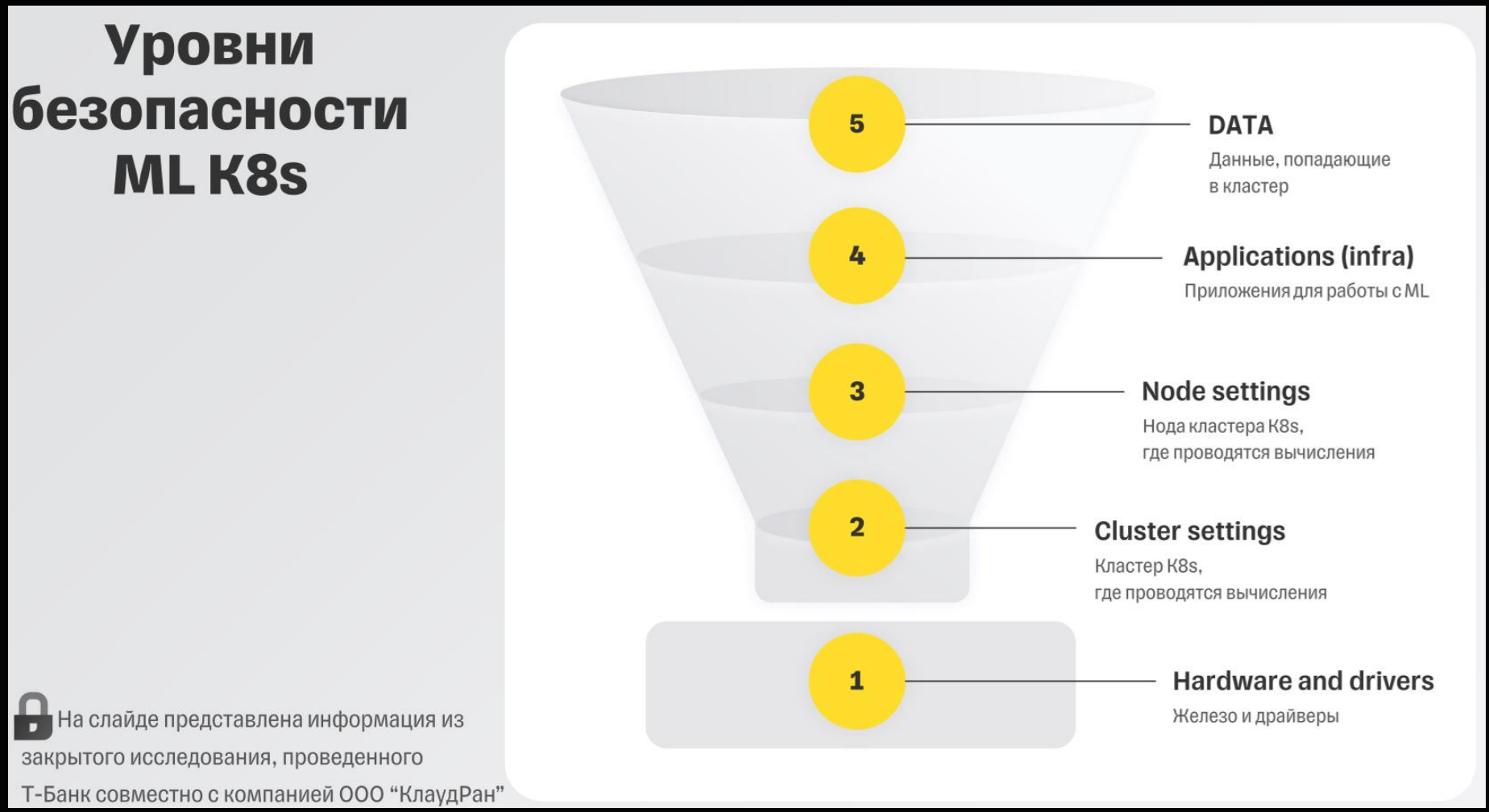
- недекларированные возможности
- небезопасная конфигурация
- избыточные привилегии

03

На уровне оркестратора

- неподдерживаемые системные компоненты
- уязвимость в системных компонентах
- отсутствие изоляции
- отсутствие сетевой сегментации
- небезопасная конфигурация
- отсутствие контроля доступа
- избыточные привилегии сущностей

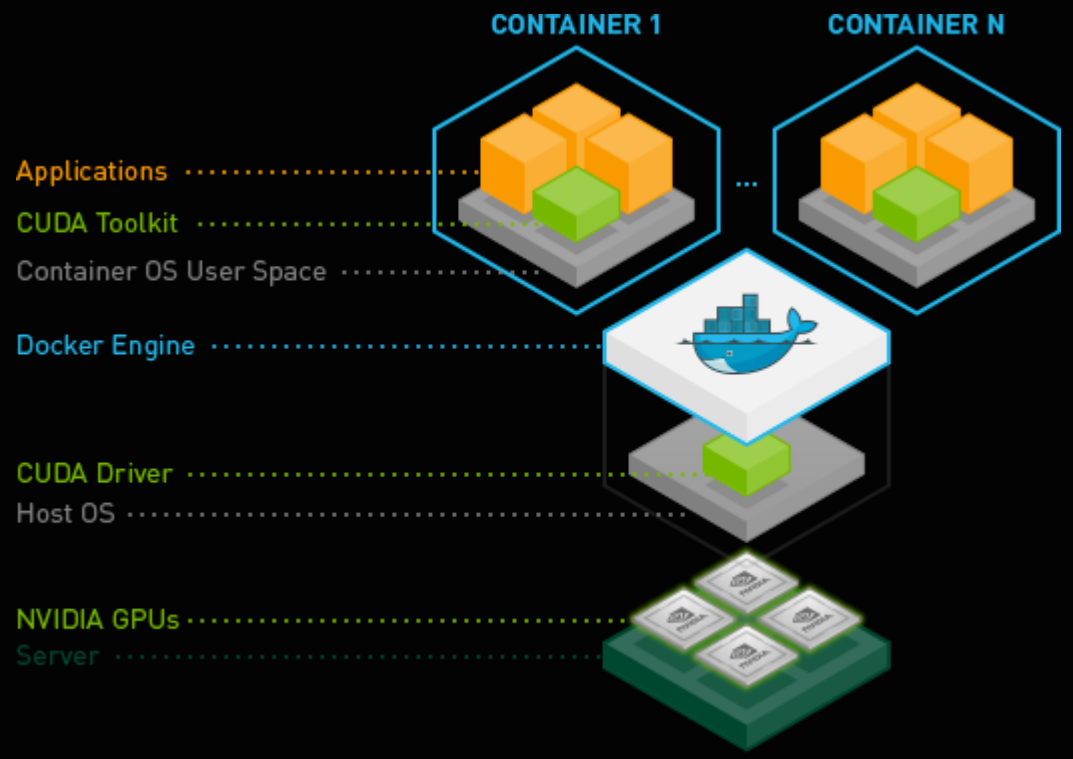
Безопасность AI кластеров Kubernetes



"Чеклист безопасности ML-кластеров", Николай Панченко

Уязвимости в NVIDIA Container Toolkit

- NVIDIA Container Toolkit:
  - CVE-2025-23359
  - CVE-2025-23266
  - CVE-2024-0132
- NVIDIA CUDA driver:
  - CVE-2025-23282
  - CVE-2025-23332



LUNTRY

Контроль состояния Kubernetes-кластеров

Контроль состояния Kubernetes-кластеров

- Контроль актуальности используемых версий системных компонент
- Проверка системных компонент на уязвимости
- Специфика отсутствует

Release	Released	Active Support	Maintenance Support	Latest
1.34	2 months and 2 weeks ago (27 Aug 2025)	Ends in 9 months (27 Aug 2026)	Ends in 11 months (27 Oct 2026)	1.34.1 (09 Sep 2025)
1.33	6 months and 3 weeks ago (23 Apr 2025)	Ends in 5 months (28 Apr 2026)	Ends in 7 months (28 Jun 2026)	1.33.5 (09 Sep 2025)
1.32	11 months ago (11 Dec 2024)	Ends in 1 month and 2 weeks (28 Dec 2025)	Ends in 3 months and 2 weeks (28 Feb 2026)	1.32.9 (09 Sep 2025)
~~1.31~~	1 year and 3 months ago (13 Aug 2024)	Ended 2 months and 2 weeks ago (28 Aug 2025)	Ended 2 weeks ago (28 Oct 2025)	~~1.31.13 (09 Sep 2025)~~

Kubernetes Documentation / Reference / Kubernetes Issues and Security / CVE feed

Official CVE Feed

FEATURE STATE: Kubernetes v1.27 [beta]

This is a community maintained list of official CVEs announced by the Kubernetes Security Response Committee. See Kubernetes Security and Disclosure Information for more details.

The Kubernetes project publishes a programmatically accessible feed of published security issues in JSON feed and RSS feed formats. You can access it by executing the following commands:

Контроль соответствия кластера стандартам

Соответствие стандартам

- Big Data может добавить сложность для compliance из-за обработки чувствительных данных (PII, регуляторные требования как GDPR или PCI-DSS)

"Чеклист безопасности ML-кластеров", Николай Панченко

Контроль Kubernetes-ресурсов

Контроль Kubernetes-ресурсов (custom)

LUNTRY

- Big Data-ворклоады усиливают необходимость строгого контроля ресурсов из-за их ресурсоёмкости: они часто используют GPU, большие объёмы памяти и CPU для обработки датасетов, что приводит к рискам перегрузки кластера и DoS-атак.
- Big Data проекты это внешние решения/операторы со своими Custom Resources, которые также надо контролировать
- Необходимо использовать Policy Engine (Kyverno, OPA Gatekeeper) или Validating Admission Policy с политиками для данных Custom Resources

```yaml
- name: validate-hadoop-pod-limits
  match:
    any:
    - resources:
        kinds:
        - Pod
        selector:
          matchLabels:
            app.kubernetes.io/component: hadoop
  validate:
    message: "Hadoop pods must have CPU and memory limits defined."
    pattern:
      spec:
        containers:
        - resources:
            limits:
              cpu: "?*" # Requires a CPU limit
              memory: "?*" # Requires a memory limit
```

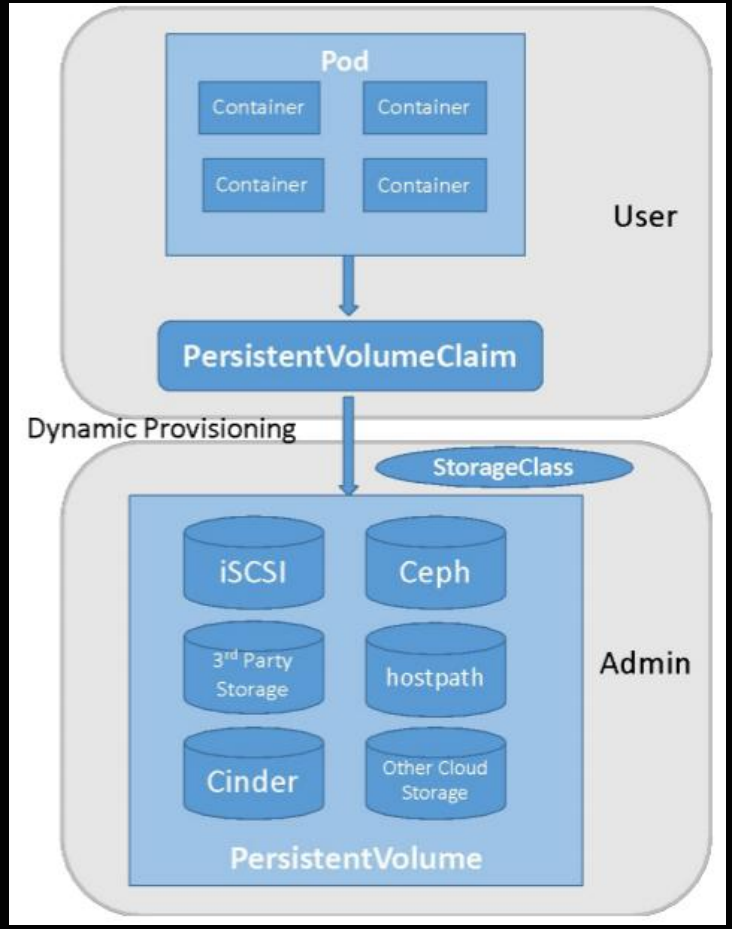
Контроль Kubernetes-ресурсов (native)

Persistent Volume Protection

Protecting access to volumes is critical to ensure only authorized containers and workloads may leverage volumes provided. It is imperative to define trust boundaries for namespaces to cordon access to volumes. Leverage existing or create new security policies that prevent groups of containers from accessing volume mounts on worker nodes and ensure only appropriate worker nodes have access to volumes. It is especially critical as privileged containers can gain access to a mounted volume in a different namespace, so additional precautions are needed.

Specifying the UID or GID of the volume still permits access by container in the same namespace and will not provide data protection. Network file system version 3 (NFSv3) assumes the client has already performed authentication and authorization and does not perform validation. It is critical to consider where authentication and authorization occur and whether validation of that action exists when implementing protections.

Cloud Native Security Whitepaper



LUNTRY

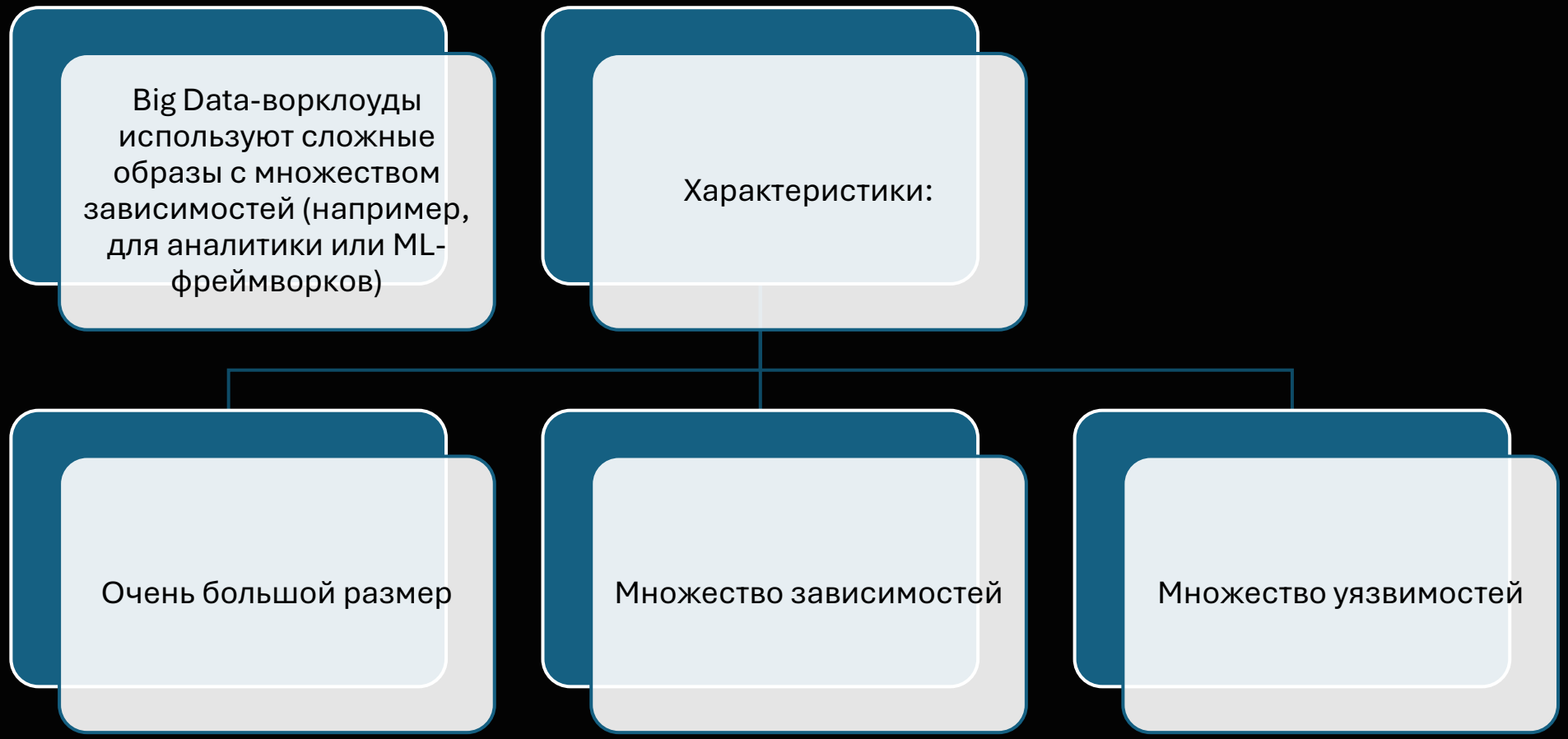
Kubernetes Audit Log

LUNTRY

- SOC мониторить специфику данных кластеров
- Данные кластера должны иметь соответствующую Audit Policy
- SOC должен иметь соответсвующие playbooks для данных кластеров

Управление безопасностью образов

Управление безопасностью образов



Хорошо известная проблема

AI / ML containers / Kubernetes

The difficult problems...

- Size
  - Containers - small, minimal, reduced set of dependencies
  - AI / ML - big and ships everything - several GB images are normal
- Combinatorial explosion of supported versions
  - Applications - which features to ship, driver version support
  - Multiple versions of Python
  - GPU and CPU based versions of software
  - Hardware support - driver versions, leveraging compatibility layers

chainguard

5

["Securing AI/ML Workflows: Optimizing Container Images in Kubernetes Environments", Wojciech Kocjan]

KEP-4639: VolumeSource: OCI Artifact and/or Image

LUNTRY

С версии 1.31 (alpha) можно монтировать образы как volume (источники данных) с любой информацией в нем.

- В 1.33 уже статус beta
- В 1.35 включен по умолчанию

Можно хранить и распространять любой контент через OCI реджистри, а потом подмонтировать к основному контейнеру с бизнес логикой.

- На пример, конфиги, базы знаний, feeds, нейронки, LLM модели, датасеты и т.д.

ImageVolume feature gate

```yaml
kind: Pod
spec:
  containers:
    - …
      volumeMounts:
        - name: my-volume
          mountPath: /path/to/directory
  volumes:
    - name: my-volume
      image:
        reference: my-image:tag
```

"Kubernetes 1.31: Read Only Volumes Based On OCI Artifacts (alpha)“

Уязвимости NVIDIA Container Toolkit

LUNTRY

Очень много уязвимостей:

- CVE-2025-23266 (CVSS Score 9.0)
- CVE-2025-23359 (CVSS Score 8.3)
- CVE-2024-0132 (CVSS Score 9.0)
- …

Характеристики:

- Очень простая и стабильная эксплуатация для побега из контейнера.
- Специально сформированный Dockerfile.

Нужно проверять как можно раньше!

Сетевая безопасность

Сетевая безопасность

LUNTRY

- Big Data генерирует огромный east-west трафик
  - Pod-to-Pod коммуникации для data ingestion, preprocessing и синхронизации моделей
- Big Data-ворклоады часто развертываются в нескольких кластерах
  - На пример, отдельные для тренинга и инференса AI

- Сторонние сетевые решения по безопасности, которые стоят в разрыве, очень негативно влияют на производительность!
  - Никаких DPI, DLP и т.д.
- Native NetworkPolicy (Layer 3/4) хороши с точки зрения производительности, но их может не хватить для гранулярности
  - На пример, недостаточны для контекстного контроля. Например, FQDN-фильтрации для внешних API.
  - Нужны Custom Network Policy от Cilium, Calico, Antrea

Защита Runtime

Необычные атаки и “фичи”

What are Shadow Vulnerabilities?

A pattern we found in many popular open source libraries.

Vulnerabilities that do not have a CVE assigned
- Or the CVE that was “Disputed”

Usually known to the maintainer
- “It is your responsibility to use it right”
- Intended behavior: “Vulnerable-by-design”

Silent misconfigurations

Mentioning them in the doc
- (which no one reads)

"Shadow Vulnerabilities in AI/ML Data Stacks - What You Don’t Know CAN Hurt You", Avi Lumelsky, Nitzan Mousseri

Мониторинг за контейнером

LUNTRY

Big Data подразумевает long-running процессы (аналитика, ML-джобы).

Следим как обычно за:

- Процессными событиями
- Сетевыми операциями
- Файловыми операциями

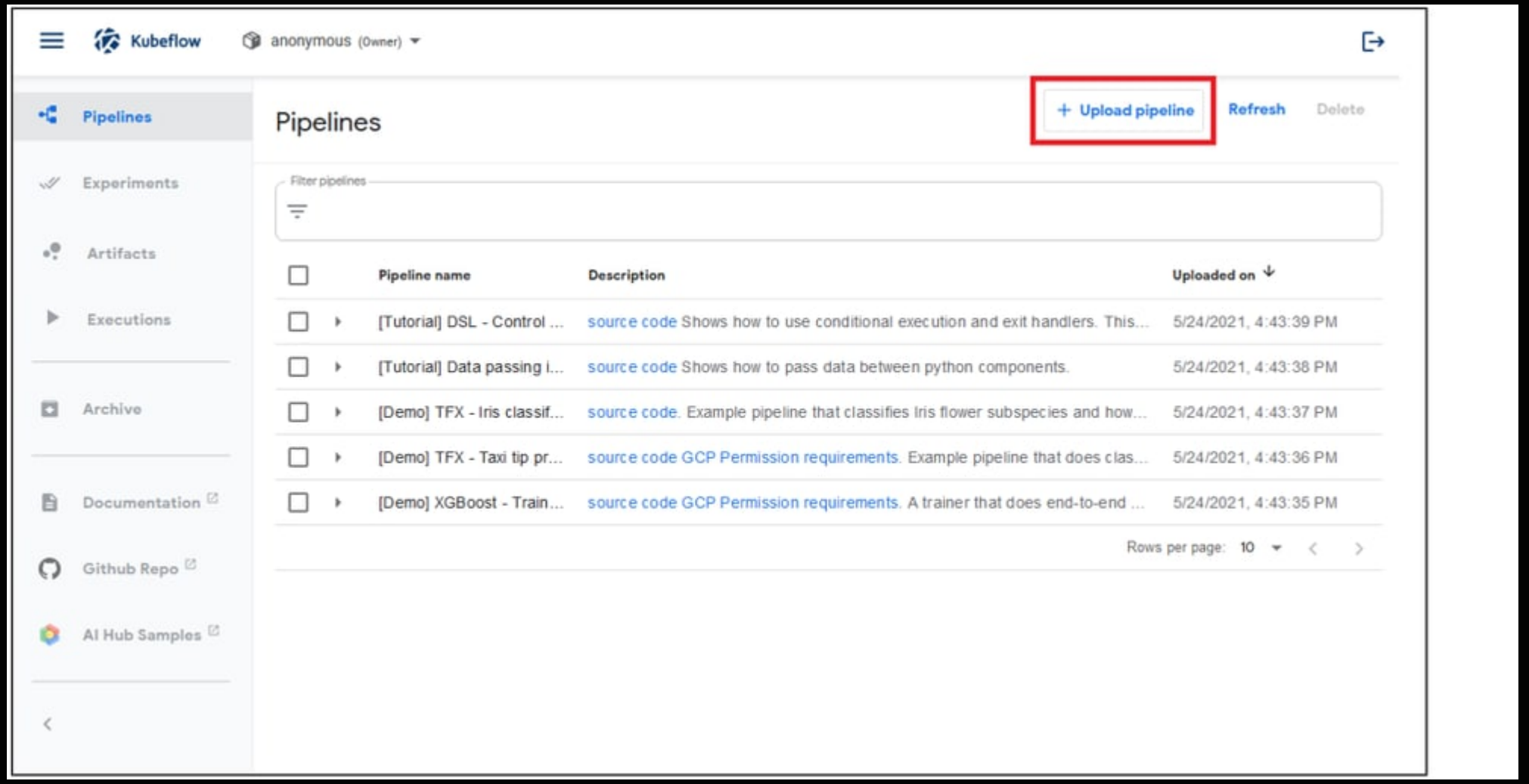
Строим профили нормального поведения.

Специфика только в количестве и в объеме.

Это требует особой обработки, иначе сенсор просто не справиться с обработкой событий для детектирования атаки и будет потреблять большое количество ресурсов.

Анализ прав доступа

Не забывайте про аутентификацию



"New large-scale campaign targets Kubeflow"

LUNTRY

Анализ прав доступа

LUNTRY

- Нужно следовать принципу наименьших привилегий
- Контролировать права как к native Kubernetes ресурсам, так и к custom.
- Особой специфики нет, все крутиться вокруг классического RBAC Kubernetes

LUNTRY

КАК LUNTRY ПОМОГАЕТ РЕШИТЬ ЗАДАЧИ

Возможности Luntry

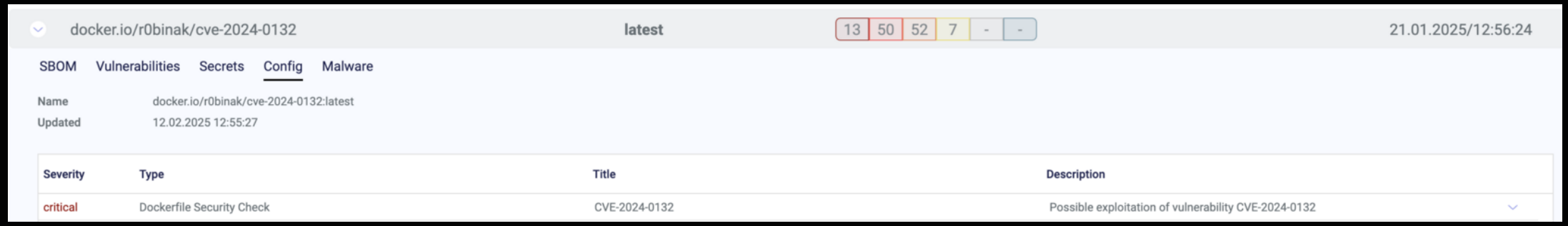
Специфика:

- Стек решений
- Объем данных и трафика
- Требования к производительности

Специальные проверки для AI/ML стека

LUNTRY

Детекты для побегов через Nvidia GPU Operator



"Ломаем ваши видеокарты: распаковка эксплойта для CVE-2024–0132 под NVIDIA Container Toolkit"

LUNTRY

ВЫВОДЫ

Выводы

LUNTRY

01 Big Data действительно влияет на безопасность в Kubernetes, в основном усиливая существующие вызовы и вводя новые риски.

02 Big Data не только усиливает преимущества Kubernetes, но и усугубляет security-риски, требуя проактивного подхода.

03 Big Data не отменяет то, что делается для классических кластеров — необходим комплексный подход к безопасности.

04 Luntry учитывает специфику и помогает защищать во всех доменах.

LUNTRY

luntry_official
luntrysolution
luntrysolution

luntry.ru
info@luntry.ru

ДМИТРИЙ ЕВДОКИМОВ

Основатель и технический директор Luntry

de@luntry.ru
Qu3b3c
k8security

СПАСИБО ЗА ВНИМАНИЕ!