大模型时代软件供应链的效率与安全管理实践

李威 JFrog

科技生态圈峰会 + 深度研习

——1000 + 技术团队的共同选择

KEYLINK ing

K+峰会

上海站
K+全球软件研发行业创新峰会
时间：2024.06.21-22

敦煌站
K+思考周®研习社
时间：2024.10.17-19

香港站
K+思考周®研习社
时间：2024.11.10-12

K+峰会详情

AiDD峰会

上海站
AI+研发数字峰会
时间：2024.05.17-18

北京站
AI+研发数字峰会
时间：2024.08.16-17

深圳站
AI+研发数字峰会
时间：2024.11.08-09

AiDD峰会详情

AiDD

2024 AI+研发数字峰会

AI+ Development Digital summit

深圳站 11/08-09

AI驱动研发变革 促进企业降本增效

2024深圳站-议题设置

AI+产品线	LLM驱动产品创新	LLM驱动需求与业务分析	AI驱动设计与用户体验
AI+开发线	AI 原生应用开发框架与技术	AI Agents在研发落地实践	LLM驱动编程与单测
AI+测试线	LLM驱动测试分析与设计	基于LLM生成测试脚本与数据	LLM和AI应用的评测
AI+工程线	AI+DevOps 与工具 (LLM 时代的平台工程)	大模型对齐与安全	端侧大模型与云端协同
AI+领域线	领域大模型 SFT 与优化	知识增强与数据智能	大厂专场

扫描右侧二维码
查看更多会议详情

早鸟票限时抢购中 (截止到9月30日)

¥3680 早鸟票

¥2800 学生票

目录
CONTENTS

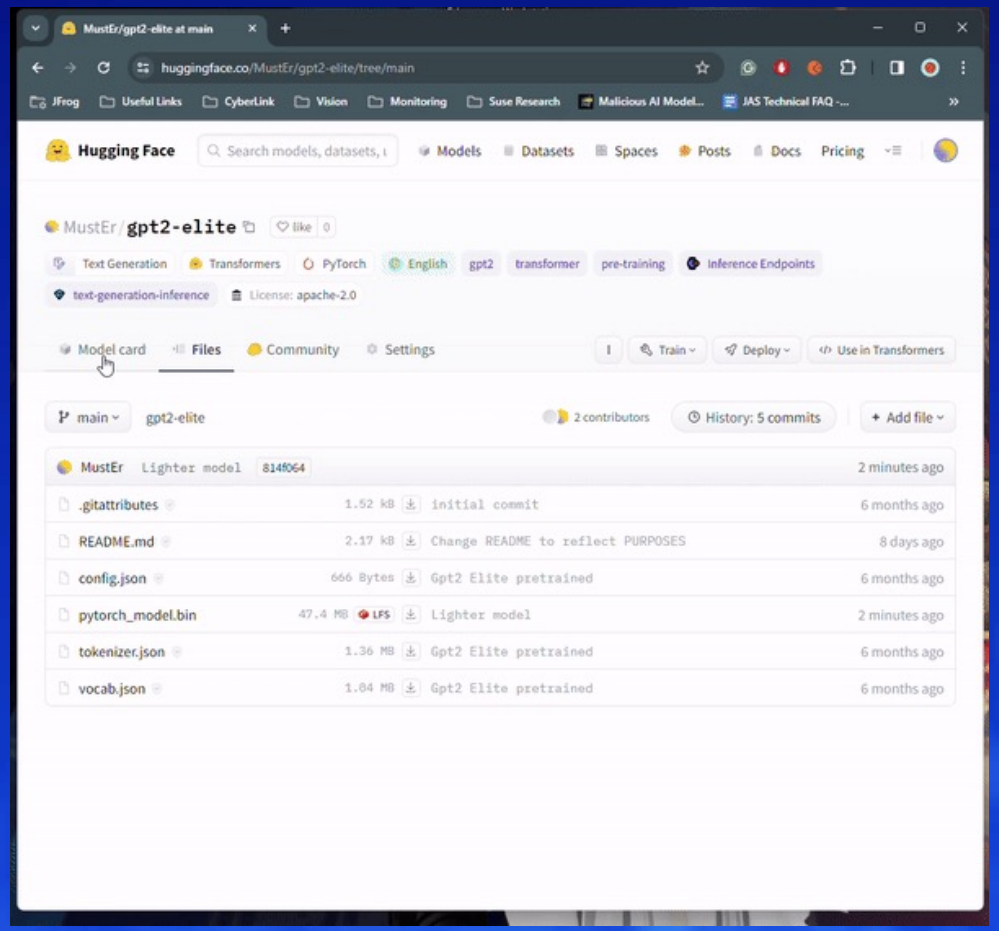
PART 01

软件供应链管理痛点

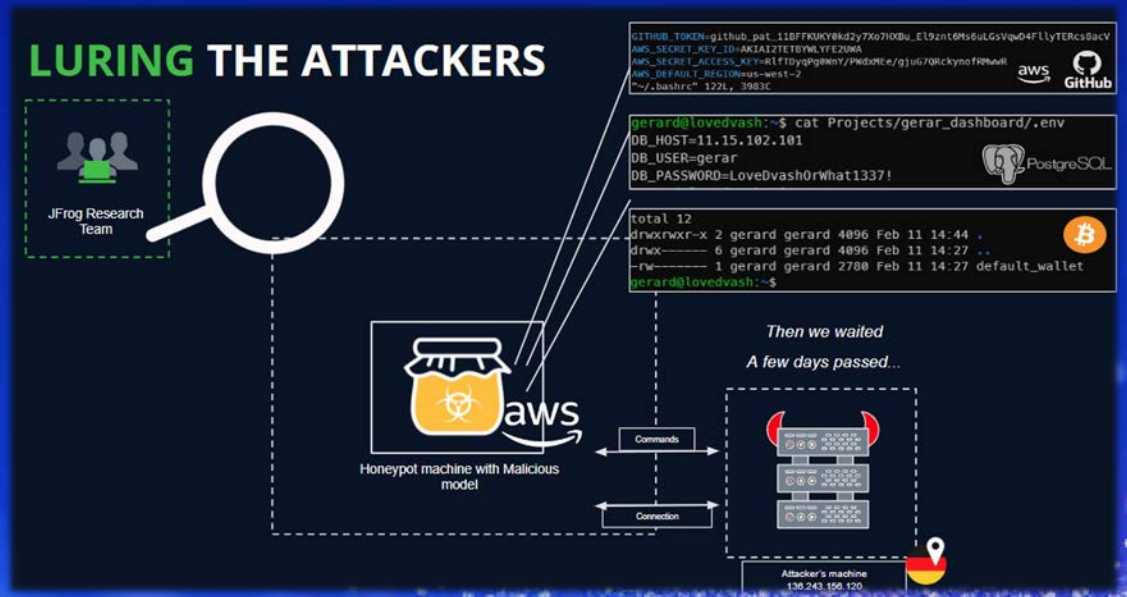
AI的风险

- 幻觉
- 偏见
- 恶意软件
- 数据投毒
- 版权
- 恶意网址
- 越狱
- 间接提示注入
- 恶意指令
- 私人信息
- 伪造来源

数据科学家成为攻击目标



JFrog 安全研究团队开发了扫描环境，每天多次严格检查上传到 Huggingface 的每个新模型



https://jfrog.com/blog/data-scientists-targeted-by-malicious-hugging-face-ml-models-with-silent-backdoor/

AI/ML 与传统软件研发的异同

开发应用程序

Code → Build → Deploy

Software Engineer	DevOps Engineer	Operations Engineer
编写和调试应用程序	管理自动化以构建和部署	部署、监控和维护

开发 ML Models

Data → Model → Deploy

Data Scientist	Research Engineer	DevOps
定义、标注和组织训练数据	开发模型算法，训练并分析模型	实施、部署、监控和维护机器学习模型

如今 AI/ML 模型版本管理的问题

- 使用 S3 存储桶

这会让数据科学家自行命名每个上传，这通常会导致命名不一致、File_Name_Final_Final_Final 难题，甚至丢失文件。

重复存储，占用大量空间

- 使用 Git

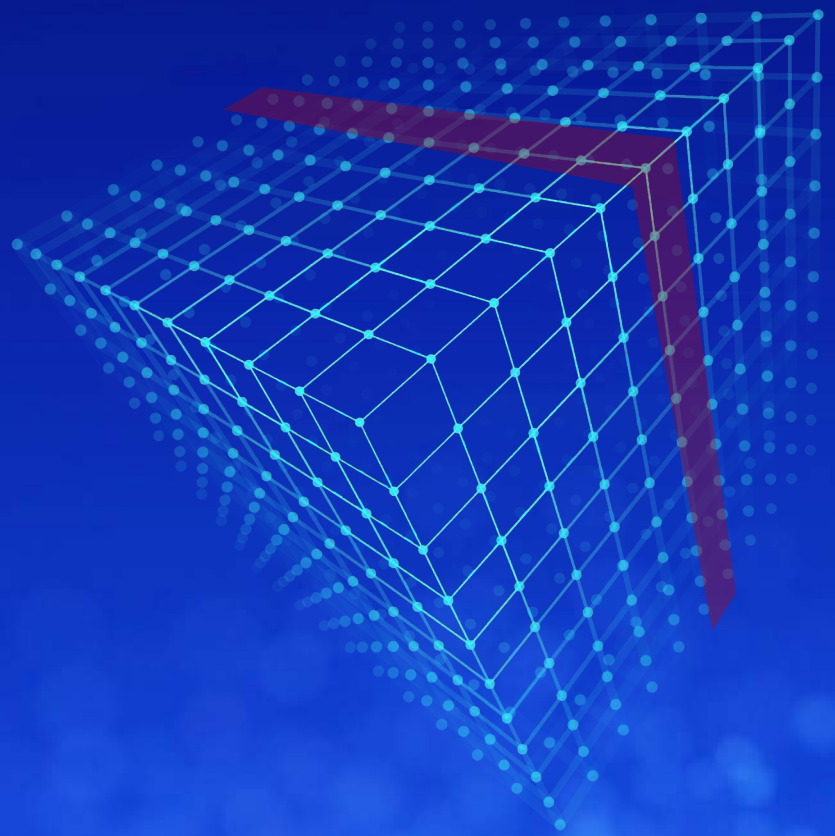
数据科学家和工程师只需在 Main 分支上堆叠 Commit， 利益相关者可以看到以前的提交，但没有简单的方法可以知道他们每次提交会得到什么，因为名称只是一组随机字符。

“基于 FTP/SVN 的手工作坊又回来了”

PART 02

软件供应链的引入与管理最佳实践

Model is a Package!



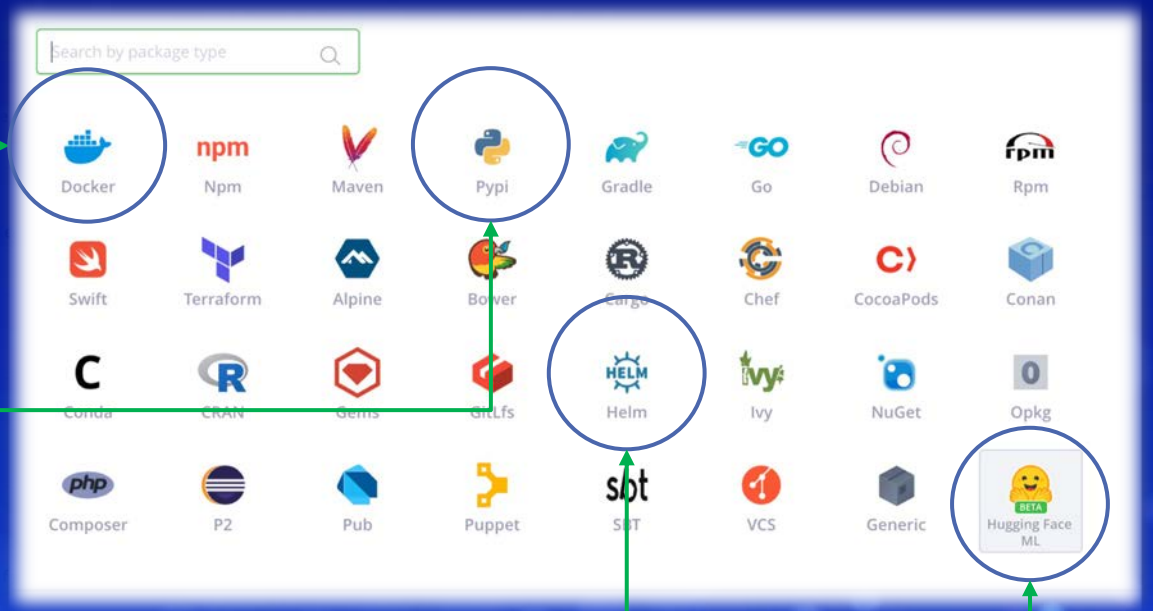
Model是基于算法训练数据生成的二进制文件，用于根据新数据进行推理。

AI软件供应链的单一可信源

JFrog ARTIFACTORY

docker

AI团队使用Docker或者OCI来管理ML的运行环境



python Package Index

AI团队使用 Artifactory 缓存和管理来自 PyPI、Pytorch 的包

HELM

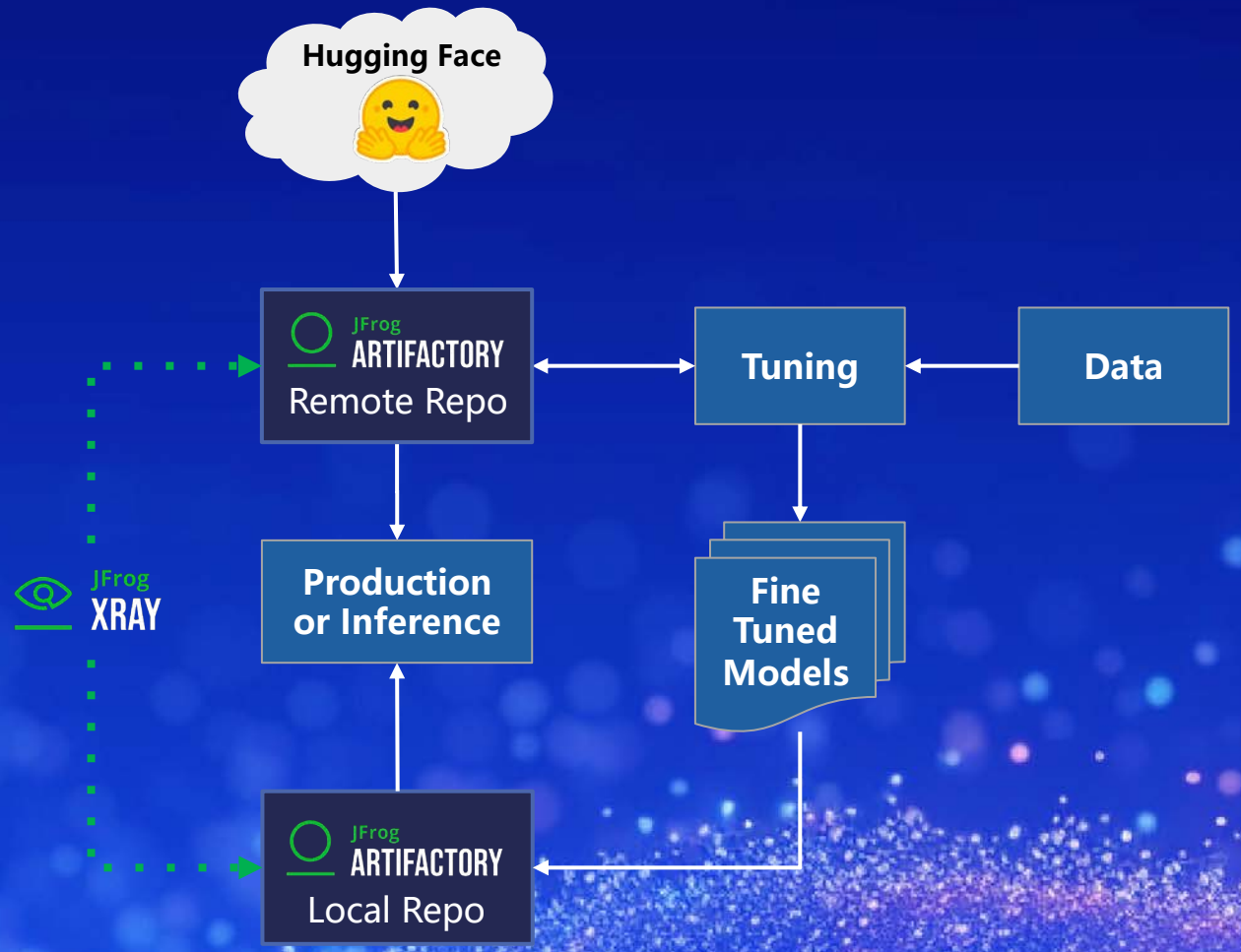
AI团队使用 Artifactory 缓存和管理 Helm Charts

Hugging Face

管理ML模型和其他制品的方式一样简单，仅需在现有流程上扩展一个包管理工具，流程可复用

JFROG 模型管理

- Hugging Face 代理/缓存
- Hugging Face 本地模型存储
- Models 和 Datasets
- 开源协议扫描
- 恶意模型扫描
- 标准化 MLOps
- 单一可信源

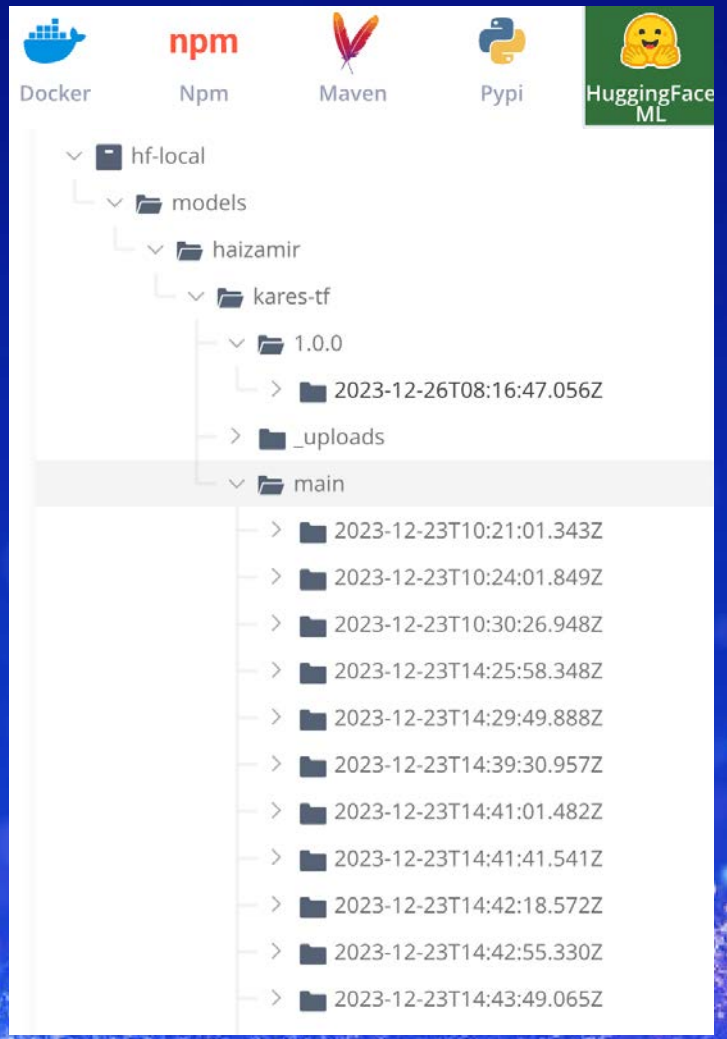


PART 03

大模型版本管理与治理最佳实践

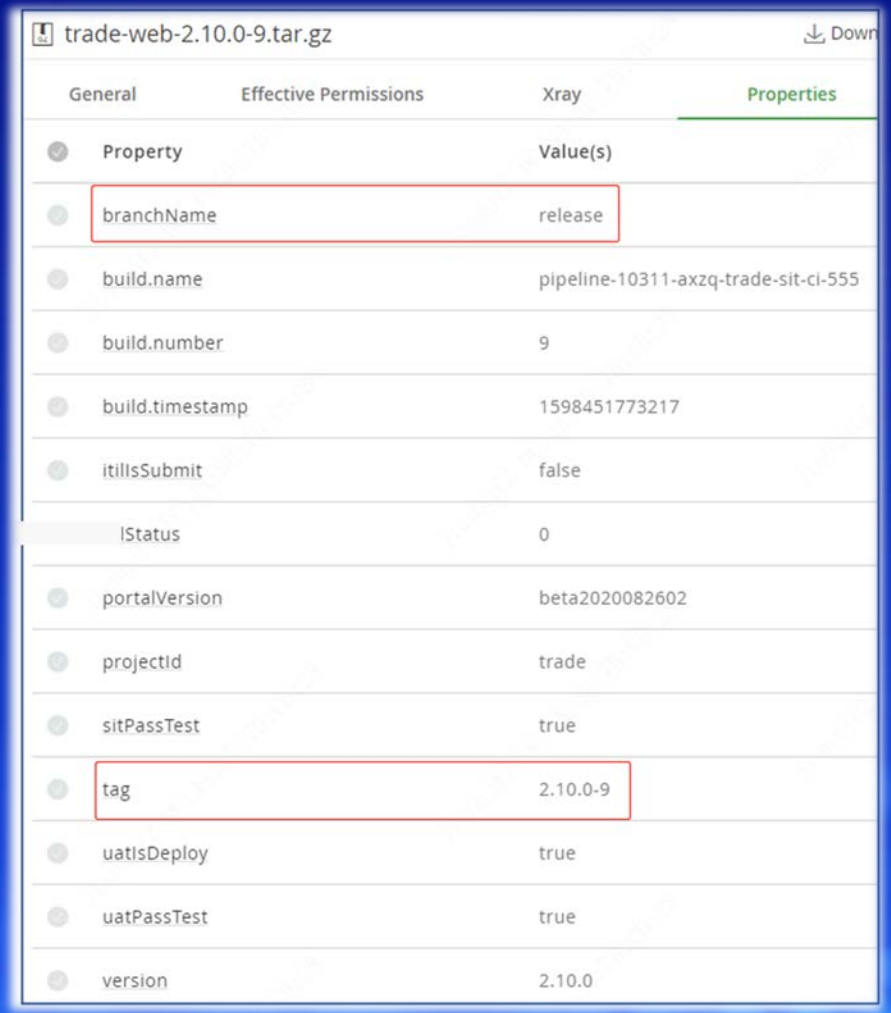
AI/ML 模型版本管理

1. 更好的存储和性能，替换 FTP/S3
2. 模型管理版本化
3. 元数据可视化
4. 存储空间可清理
5. 易于分享模型
6. 晋级模型，而无额外存储成本
7. 同步模型到生产环境，而无额外网络成本
8. 模型安全扫描



https://jfrog.com/blog/ml-model-versioning/

软件供应链可信管理：元数据



- 用于记录软件生命周期信息，解决制品黑盒问题
- 打破部门墙，在上下游传递制品信息
- 支持元数据正向、反向查询
- 可作用于制品筛选、制品清理、制品按需分发等

依赖包

- 准入申请、审批信息
- 生命周期信息

AI Model

- Model原始信息
- 训练数据集信息

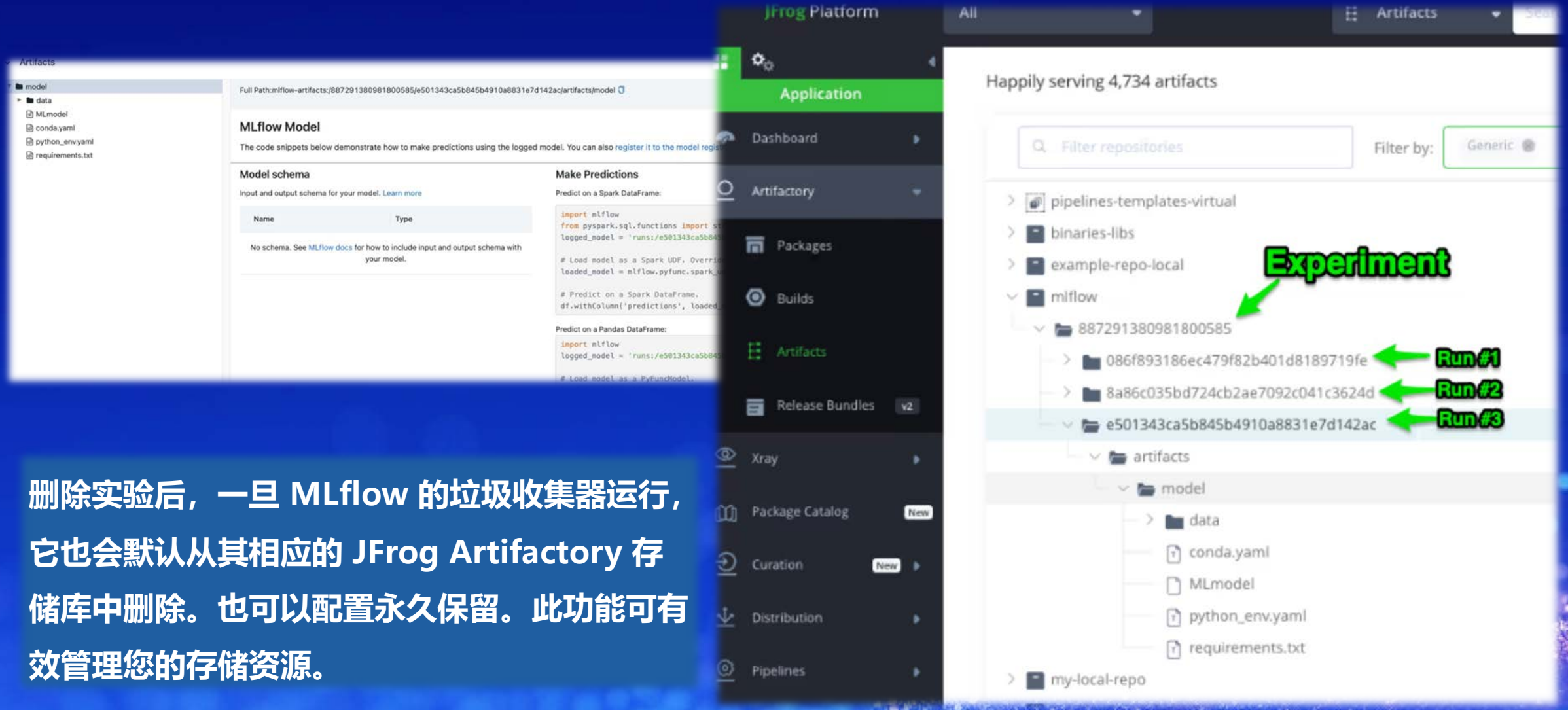
嵌入式软件包

- 匹配设备型号
- 目标客户信息
- 分发同步信息

交付物

- 代码分支、tag
- 需求/task信息
- 开发团队/人员信息
- 构建流水线信息
- 代码扫描结果
- 测试结果
- 供应链扫描结果
- 第三方安全、合规检测信息
- 文档信息
- 审批记录
- 发布信息
- 归档信息

与MLFlow 集成

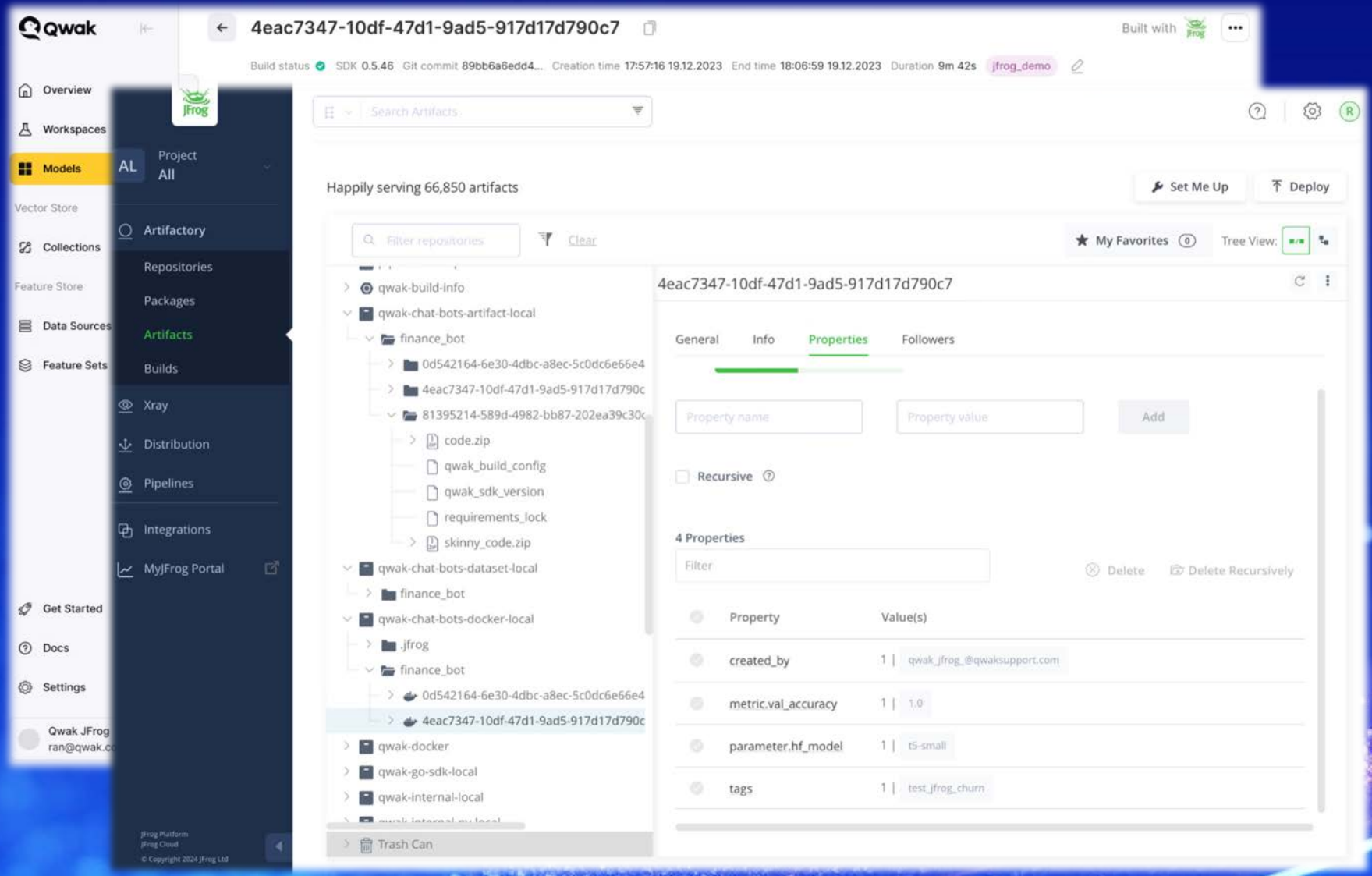


删除实验后，一旦 MLflow 的垃圾收集器运行，它也会默认从其相应的 JFrog Artifactory 存储库中删除。也可以配置永久保留。此功能可有效管理您的存储资源。

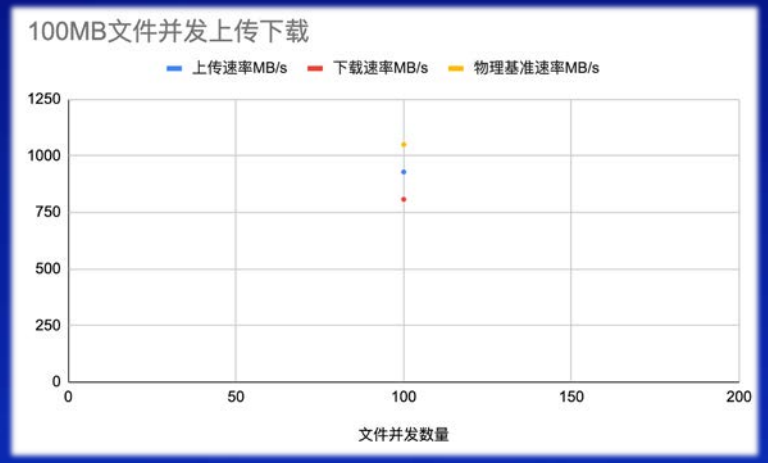
JFrog AISecOps 与Qwak集成

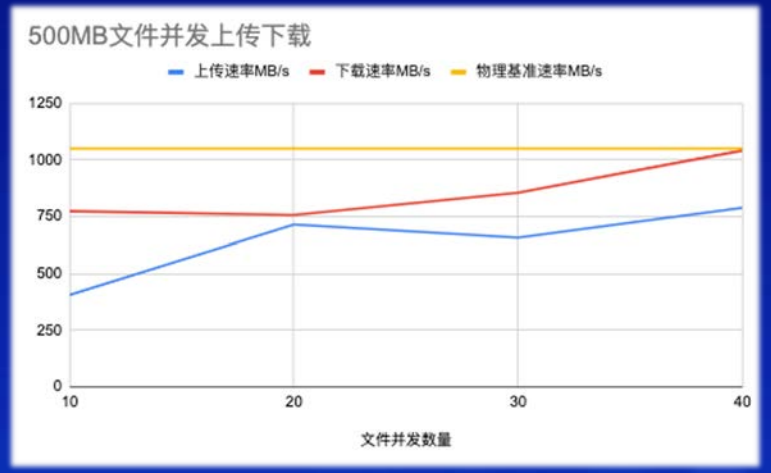
JFrog 与 Qwak 集成的完整 AISecOps 解决方案，可实现团队之间的无缝交叉协作。

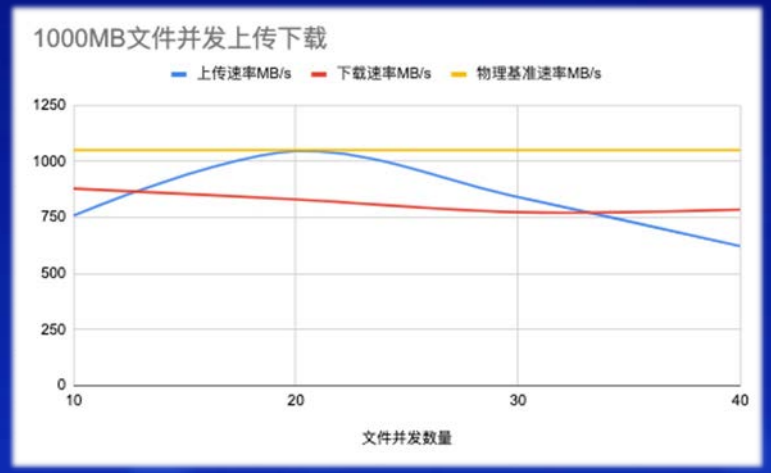
1. 将所有模型、制品集中在唯一可信源中
2. 减少外部服务中断或消除公共存储库中模型或包版本的潜在风险
3. 管理和限制对外部私有或公共存储库的访问，确保用户只能使用经过批准的源
4. 为利益相关者提供有关公司内部使用的内容的全面透明度

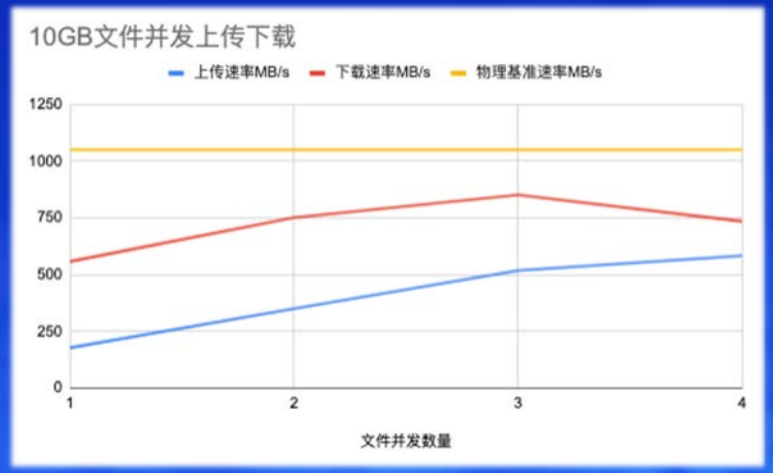


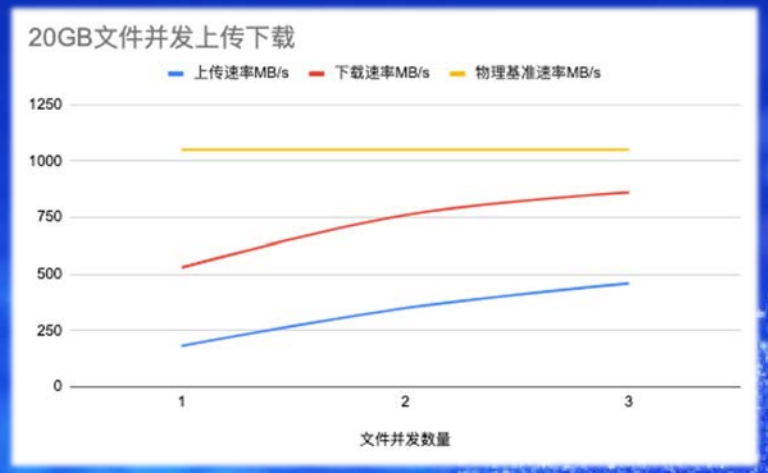
并发下载可打满 80-100% 带宽

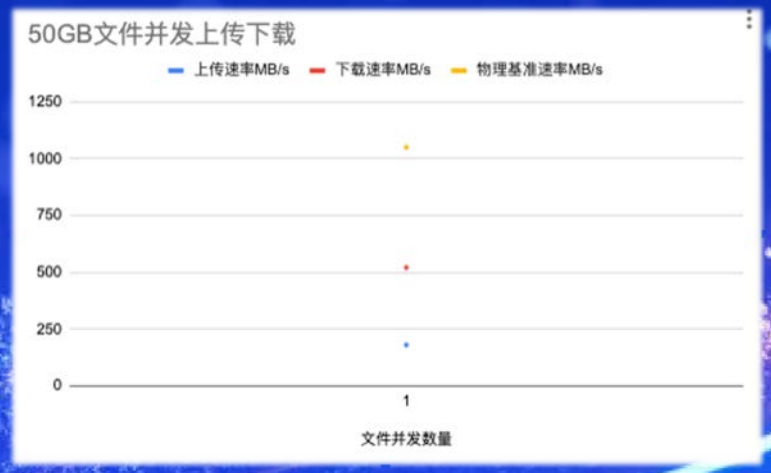






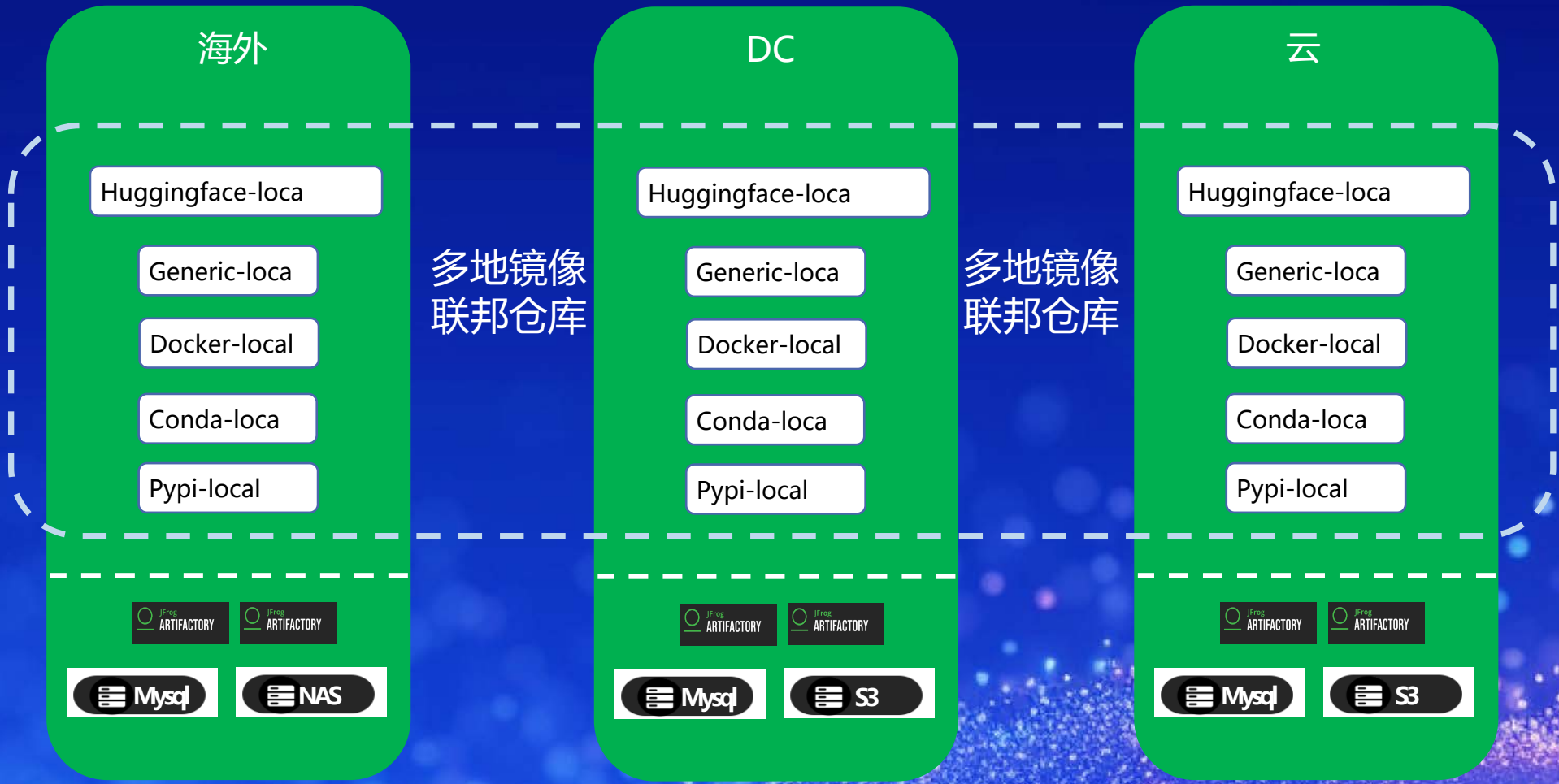






多研发中心Model管理

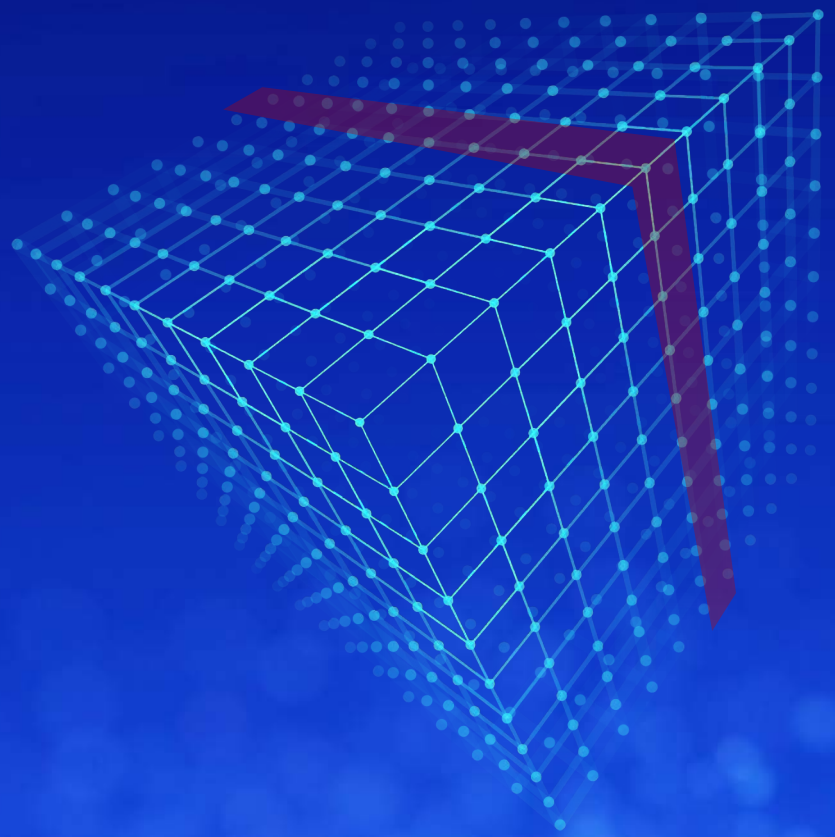
- Model就近下载
- 异地多活
- 计算卡在云上
- 海外计算



PART 04

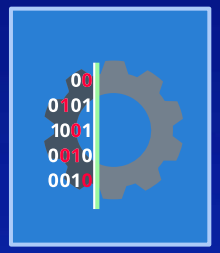
大模型安全风险治理

Model也有安全风险?



攻击者正在针对公共机器学习库进行攻击，以渗透到组织内部。

JFrog对Models风险的调研



480k+
公共机器学习库中
扫描包的数量

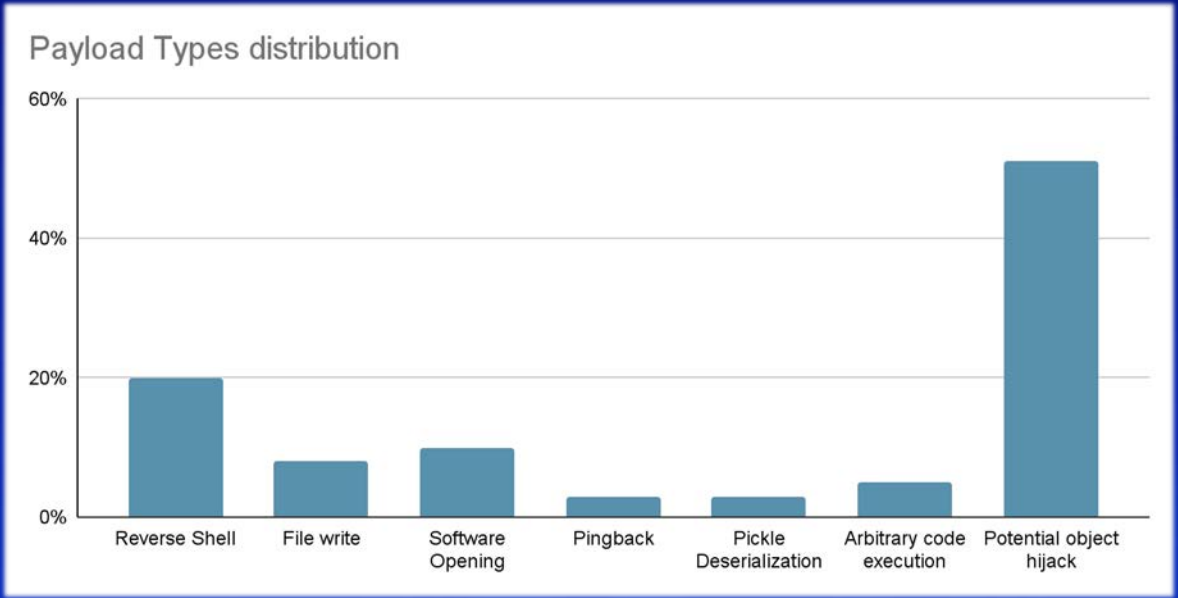
60+
models 有恶意攻击
行为

10+
models中包含与操
作系统无关的反向
Shell恶意软件!

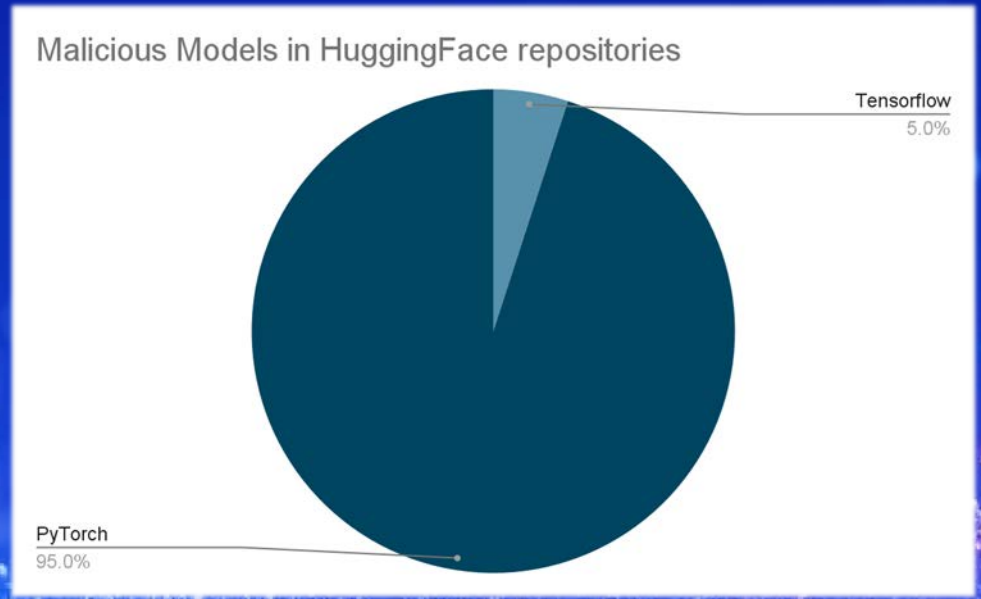
kaggle

Hugging Face

JFrog 对机器学习库进行安全监控

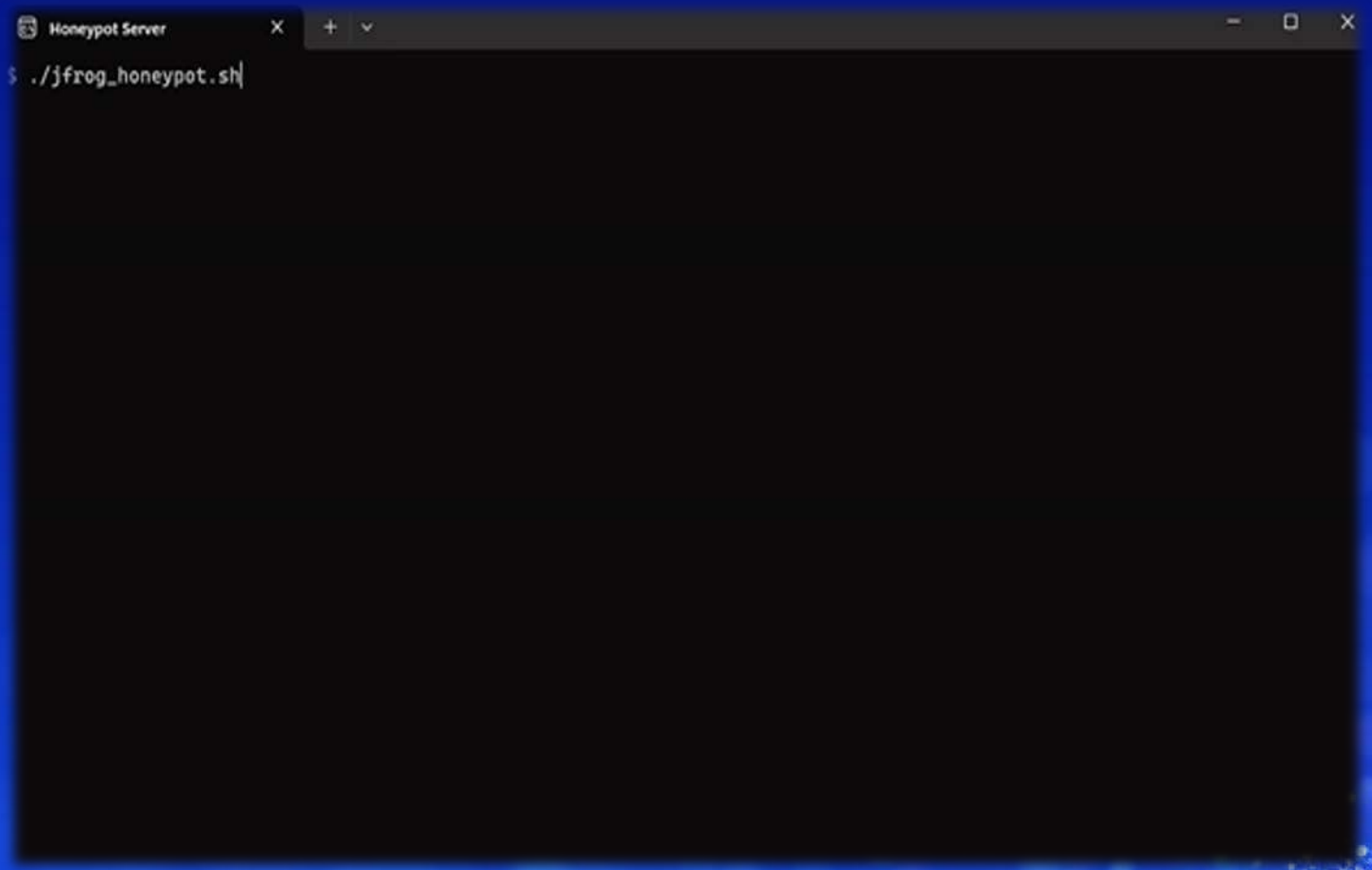


PyTorch 模型的流行率最高，紧随其后的是 Tensorflow Keras 模型。 需要强调的是，当我们提到“恶意模型”时，我们特指那些包含真实、有害有效负载的模型。



PyTorch 模型（大幅）和 Tensorflow Keras 模型（H5 或 SavedModel 格式）构成执行恶意代码的最高潜在风险，因为它们是流行的模型类型，具有已发布的已知代码执行技术。

恶意模型



- 在蜜罐中运行模型
- 发现多个外部的IP连接
- 目前还没有看到攻击者发送哪些指令
- 预测是一个长期的钓鱼项目

加载 ML 模型如何导致代码执行?

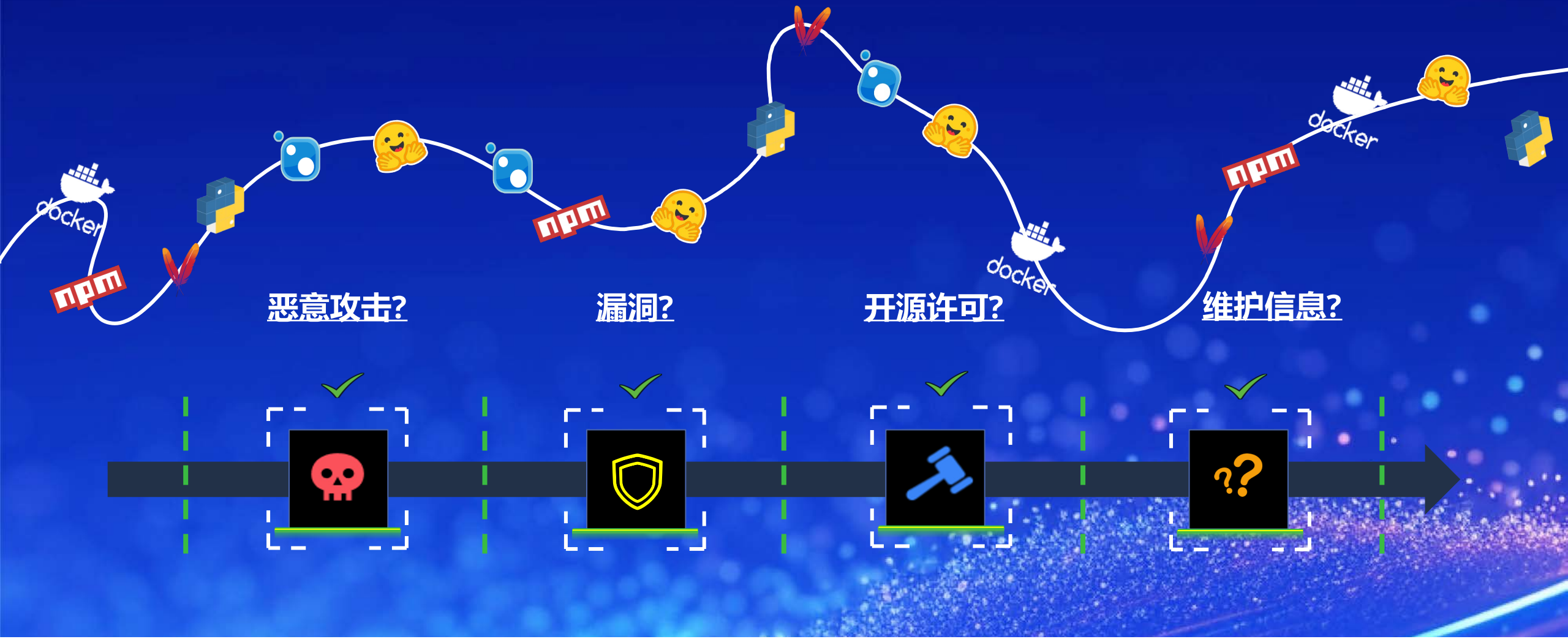
某些模型使用“pickle”格式，这是序列化 Python 对象的常见格式。

但是，pickle 文件还可以包含加载文件时执行的任意代码。

虽然这个问题被Huggingface发现，并开发了一种用于安全存储模型数据的新格式，称为 safetensors，但并没有对pickle类型采取禁止策略

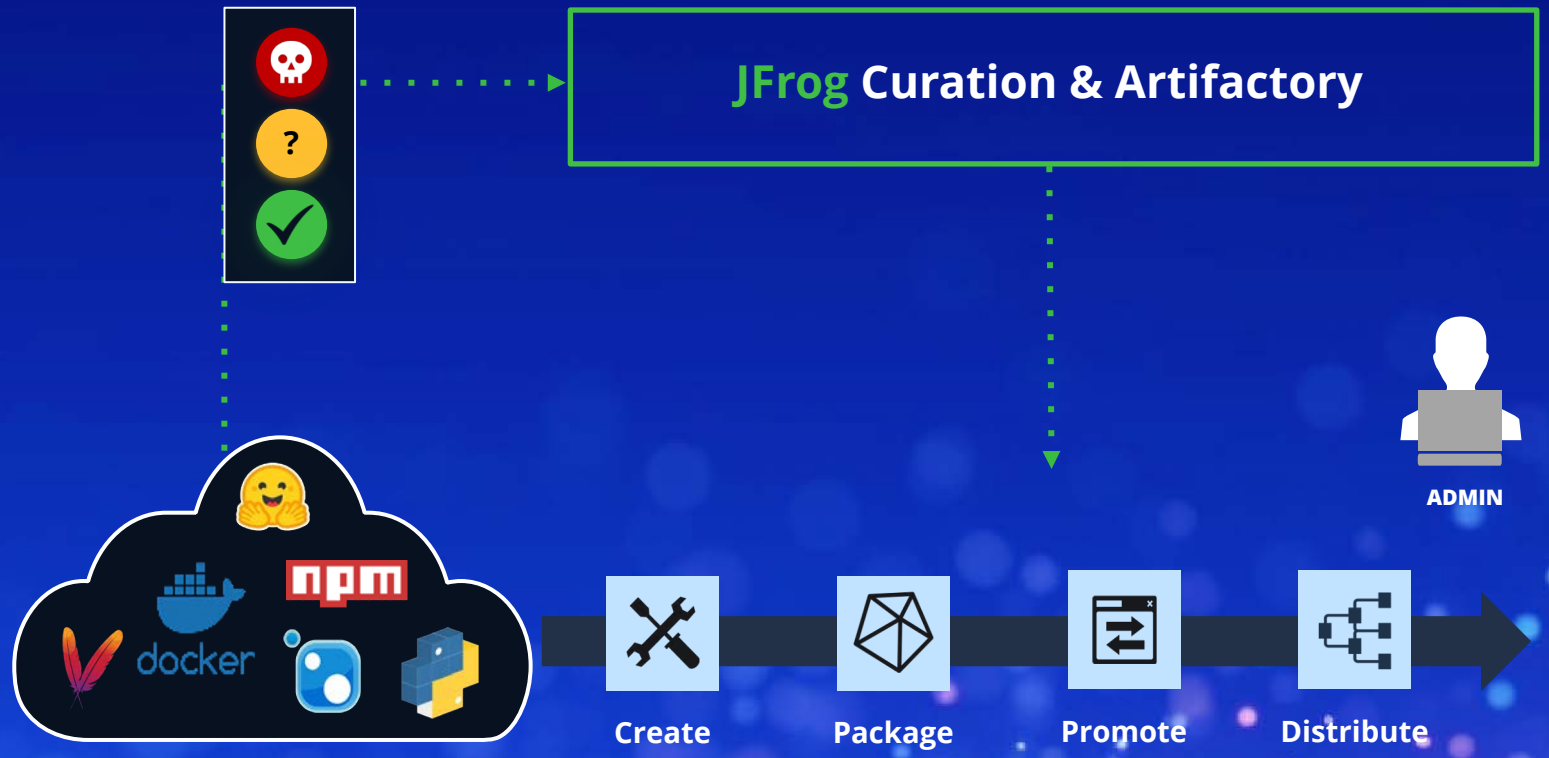
Format	Type	Framework	Code execution?	Description
JSON	Text	Interoperable	—	Widely used data interchange format
PMML	XML	Interoperable	—	Predictive Model Markup Language, one of the oldest standards for storing data related to machine learning models; based on XML
pickle	Binary	PyTorch, scikit-learn, Pandas	🐞	Built-in Python module for Python objects serialization; can be used in any Python-based framework
dill	Binary	PyTorch, scikit-learn	🐞	Python module that extends pickle with additional functionalities
joblib	Binary	PyTorch, scikit-learn	🐞	Python module, alternative to pickle; optimized to use with objects that carry large numpy arrays
MsgPack	Binary	Flax	—	Conceptually similar to JSON, but 'fast and small', instead utilizing binary serialization
Arrow	Binary	Spark	—	Language independent data format which supports efficient streaming of data and zero copy reads
Numpy	Binary	Python-based frameworks	🐞	Widely used Python library for working with data
TorchScript	Binary	PyTorch	🐞	PyTorch implementation of pickle
H5 / HDF5	Binary	Keras	🐞	Hierarchical Data Format, supports large amount of data
SavedModel	Binary	TensorFlow	—	TensorFlow-specific implementation based on protobuf
TFLite/FlatBuffers	Binary	TensorFlow	—	TensorFlow-specific for low resource deployment
ONNX	Binary	Interoperable	🐞 Rare scenarios	Open Neural Network Exchange format based on protobuf
SafeTensors	Binary	Python-based frameworks	—	A new data format from Huggingface designed for the safe and efficient storage of tensors
POJO	Binary	H2O	🐞	Plain Old JAVA Object
MOJO	Binary	H2O	🐞	Model Object, Optimized
Protobuf	Binary	Interoperable	—	Google's protocol buffers

Model可以安全使用么?



AI供应链防火墙

- Curation 隔离仓库
  - 隔离阻断高危漏洞开源组件
  - 漏洞包无法进入内网
- 内网JFrog仓库
  - 连接到隔离区的 JFrog Curation 隔离仓库
  - 持续漏洞扫描
  - 对开源软件包下载的全面监管



阻断恶意Model

tf_model.h5 ×

Edit As: Hex ▾ Run Script ▾ Run Template ▾

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	0123456789ABCDEF
3C00h:	6C	2C	20	22	62	69	61	73	5F	72	65	67	75	6C	61	72	l, "bias_regular
3C10h:	69	7A	65	72	22	3A	20	6E	75	6C	6C	2C	20	22	61	63	izer": null, "ac
3C20h:	74	69	76	69	74	79	5F	72	65	67	75	6C	61	72	69	7A	tivity_regulariz
3C30h:	65	72	22	3A	20	6E	75	6C	6C	2C	20	22	6B	65	72	6E	er": null, "kern
3C40h:	65	6C	5F	63	6F	6E	73	74	72	61	69	6E	74	22	3A	20	el_constraint":
3C50h:	6E	75	6C	6C	2C	20	22	62	69	61	73	5F	63	6F	6E	73	null, "bias_cons
3C60h:	74	72	61	69	6E	74	22	3A	20	6E	75	6C	6C	7D	2C	20	traint": null},
3C70h:	22																ame": "predict
3C80h:	69																ns", "inbound_
3C90h:	6E																des": [[["fc2"
3CA0h:	2C																0, 0, {}]]]},
3CB0h:	7B																class_name": "
3CC0h:	4C																mbda", "config
3CD0h:	22																{"name": "out
3CE0h:	70	75	74	22				74	72	61	69	6E	61	62	6C	65	put", "trainable
3CF0h:	22	3A	20	66			73	65	2C	20	22	64	74	79	70	65	": false, "dtype
3D00h:	22	3A	20	22	66	6C	6F	61	74	33	32	22	2C	20	22	66	": "float32", "f
3D10h:	75	6E	63	74	69	6F	6E	22	3A	20	5B	22	34	77	45	41	unction": ["4wEA
3D20h:	41	41	41	41	41	41	41	41	41	41	41	41	41	41	49	41	AAAAAAAAAAAAAAIA
3D30h:	41	41	41	44	41	41	41	41	51	77	41	41	41	48	4D	57	AAADAAAAQwAAAHMW
3D40h:	41	41	41	41	5A	41	46	6B	41	47	77	41	66	51	46	38	AAAAZAFkAGwAfQF8
3D50h:	41	61	41	42	5A	41	4B	68	41	51	45	41	66	41	42	54	AaABZAKhAQEAfABT
3D60h:	41	43	6B	44	54	75	6B	41	5C	6E	41	41	41	41	2B	67	ACkDTukA\nAAAA+g
3D70h:	68	6A	59	57	78	6A	4C	6D	56	34	5A	53	6B	43	32	67	hjYWxjLmV4ZSkC2g
3D80h:	4A	76	63	39	6F	47	63	33	6C	7A	64	47	56	74	4B	51	Jvc9oGc3lzdGVtKQ
3D90h:	4C	61	41	58	68	79	41	77	41	41	41	4B	6B	41	63	67	LaAXhyAwAAAKkAcg
3DA0h:	59	41	41	41	44	36	56	53	39	6F	62	32	31	6C	4C	32	YAAAD6VS9ob21lL2
3DB0h:	52	68	64	6D	5A	79	5C	6E	4C	30	70	47	55	6B	39	48	RhdmZy\nL0pGUk9H
3DC0h:	58	30	4A	70	64	47	4A	31	59	32	74	6C	64	43	39	68	X0JpdGJ1Y2t1dC9h
3DD0h:	61	53	31	74	62	32	52	6C	62	43	31	79	5A	58	4E	6C	aS1tb2RlbC1yZXN1
3DE0h:	59	58	4A	6A	61	43	39	55	5A	58	4E	30	63	79	39	47	YXJjaC9UZXN0cy9G
3DF0h:	59	57	74	6C	52	47	6C	79	4C	32	4E	79	5A	57	46	30	YWtlRGlyL2NyZWF0
3E00h:	5A	56	39	74	5C	6E	59	57	78	70	59	32	6C	76	64	58	ZV9t\nYWxpY2lvdX
3E10h:	4E	66	56	6B	64	48	4D	54	59	75	63	48	6E	61	42	32	NfVkdHMTYucHnaB2
3E20h:	56	34	63	47	78	76	61	58	51	44	41	41	41	41	63	77	V4cGxvaXQDAAAAcw
3E30h:	59	41	41	41	41	41	41	51	67	43	43	67	45	3D	5C	6E	YAAAAAAQgCCgE=\n
3E40h:	22	2C	20	6E	75	6C	6C	2C	20	6E	75	6C	6C	5D	2C	20	", null, null],
3E50h:	22	66	75	6E	63	74	69	6F	6E	5F	74	79	70	65	22	3A	"function_type":
3E60h:	20	22	6C	61	6D	62	64	61	22	2C	20	22	6D	6F	64	75	"lambda", "modu

Code hidden in the binary
"data"

- 全面的二进制扫描，检测嵌入的恶意代码
- 对被阻止的模型进行全面审计
- 左移策略 - 有问题的Model绝不会进入内部库

JFrog对Model及其他软件供应链进行安全监控

JFrog Research

NVD、Ubuntu、Debian、Redhat等漏洞库

集成全球最完整的商业漏洞数据库VulnDB

JFrog Xray 提供全生命中周期扫描能力，包括安全左移、构建扫描、交付物扫描、本地按需扫描等。

debian ubuntu Red Hat VulnDB NVD

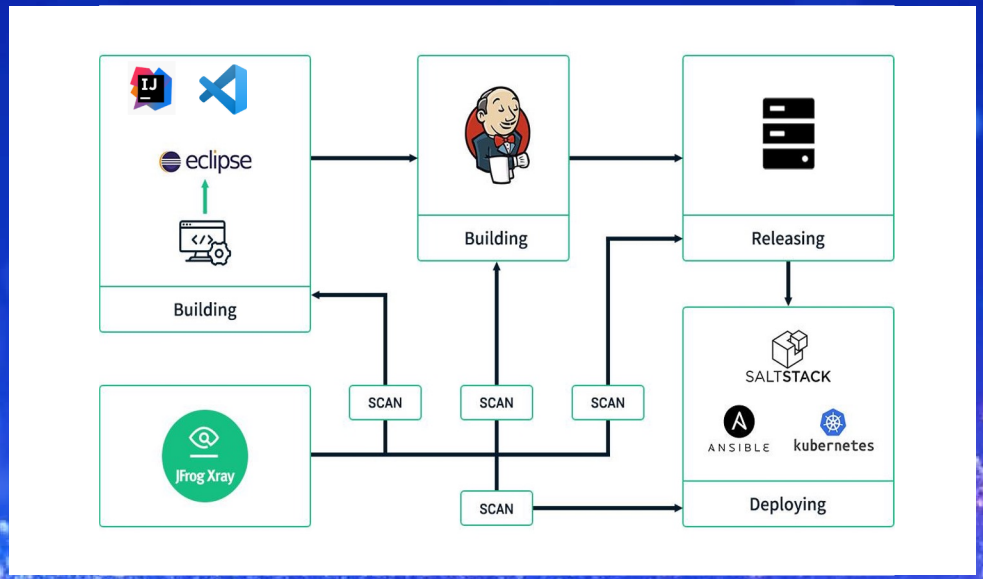
Xray DB

功能特性

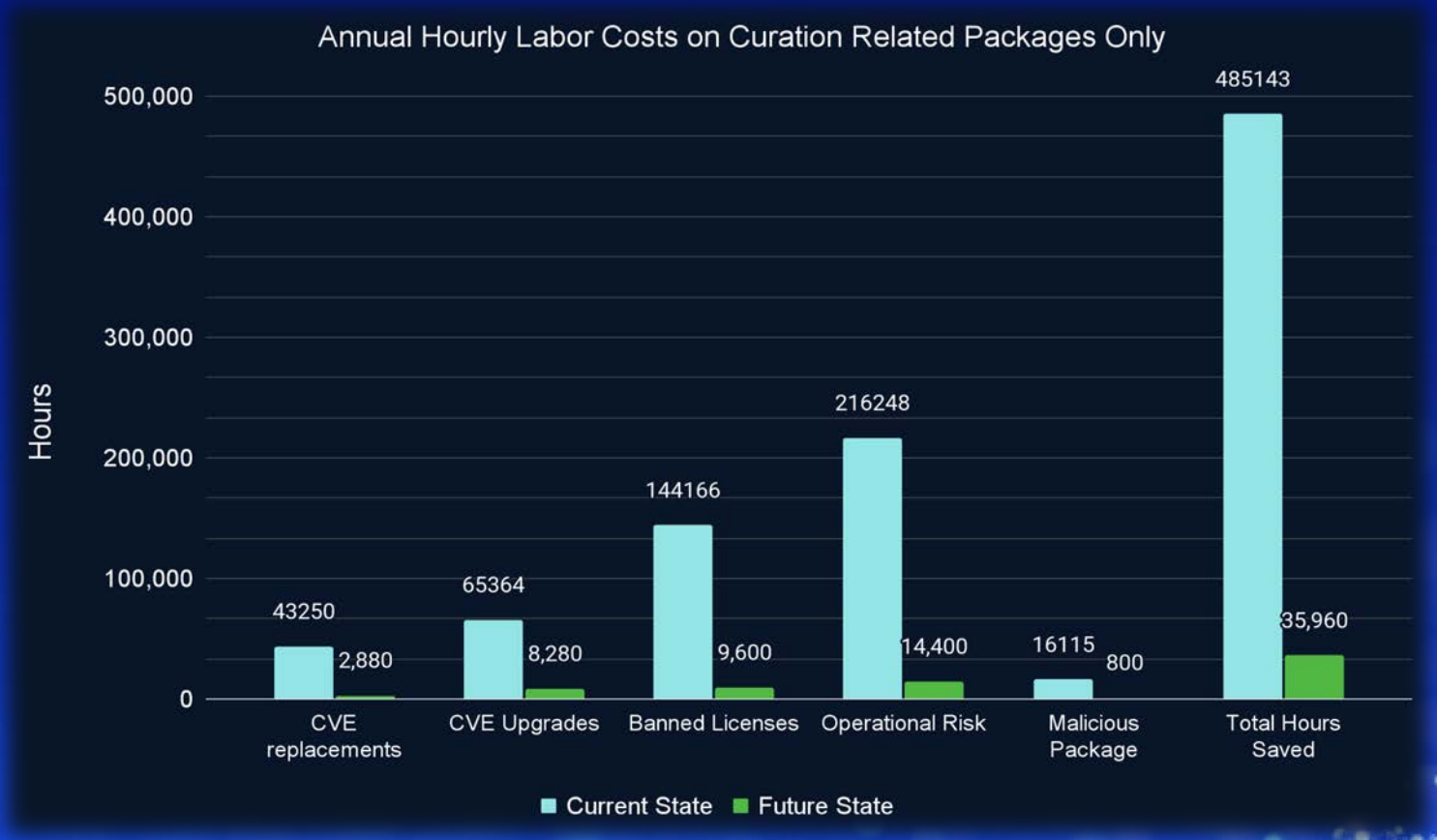
- SCA：漏洞扫描及开源协议扫描
- 提供漏洞修复建议及JFrog调研
- 跨技术栈影响性分析能力
- 与仓库集成，提供单一可信源

易用性

- 扫描看板
- 扫描报告、SBOM报告
- IDE、代码库、构建工具插件
- 基于风险及合规信息阻断包下载



JFrog 能帮开发者节省多少时间?



每年用于修复错误的软件包的总小时数	485,143
每年使用 JFrog Curation 修复软件包的总小时数	35,960
每年节省的总小时数	449,183

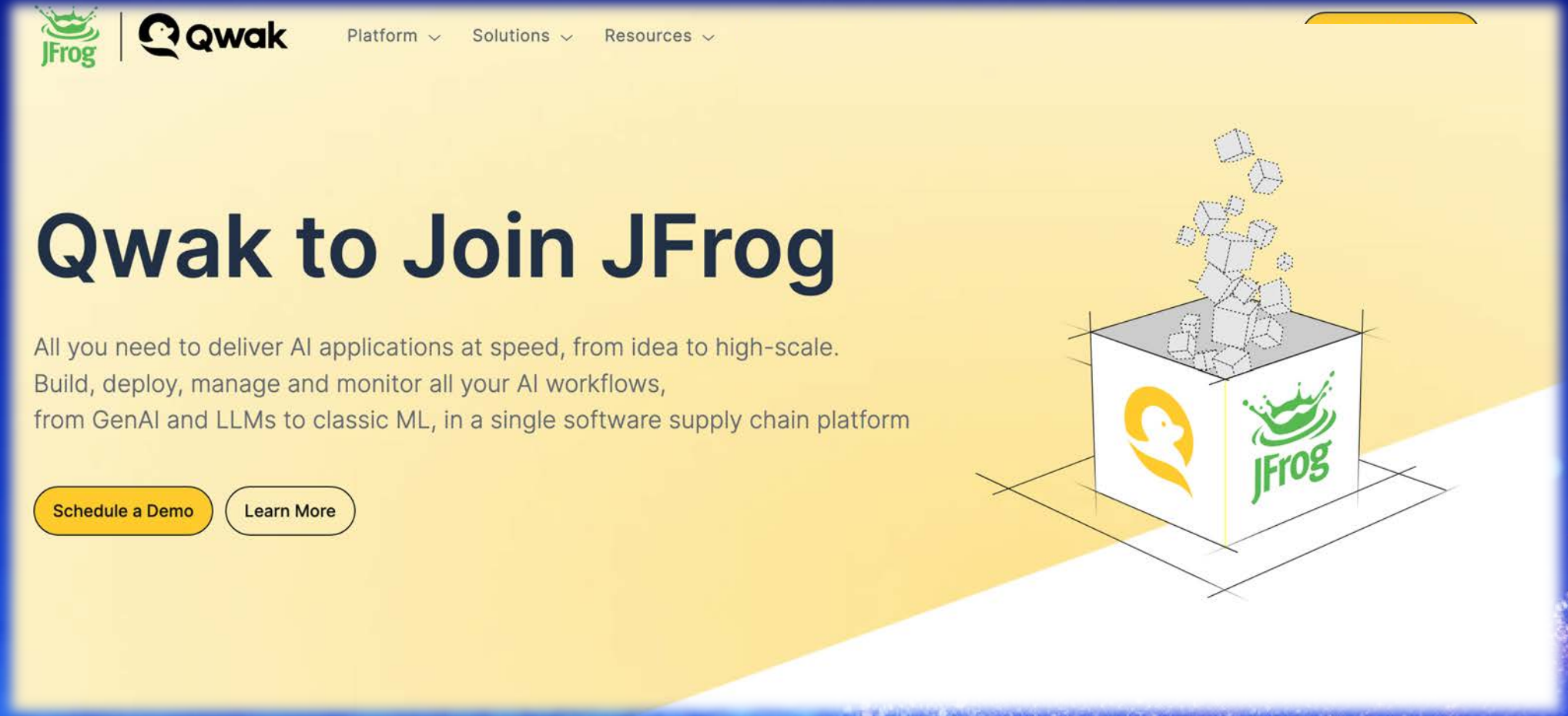
研发人数： 2850 人

PART 05

未来展望

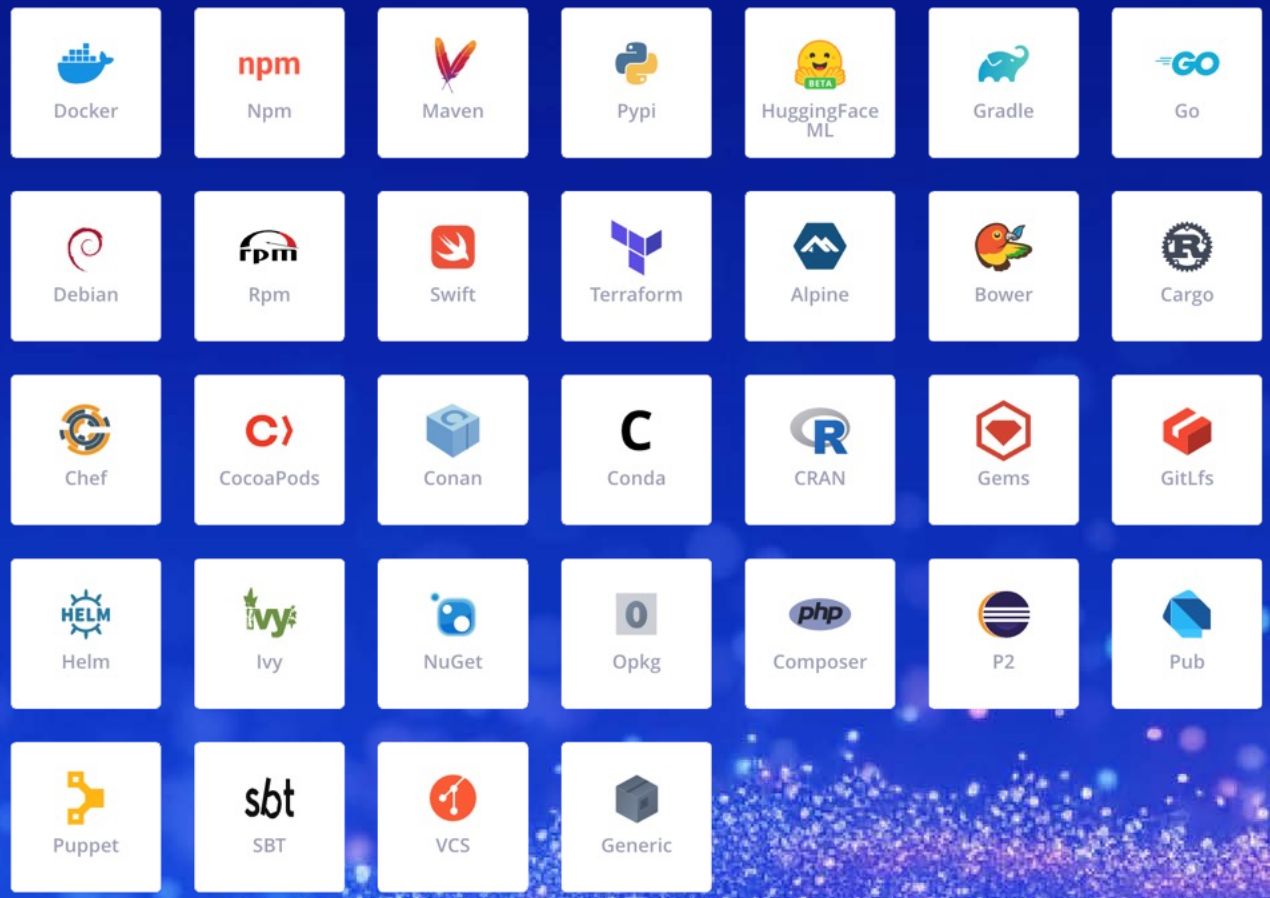
JFrog | Qwak 一体化ML管理平台

Build, Train, Secure, Serve, and Monitor ML Models and GenAI in a Unified Experience

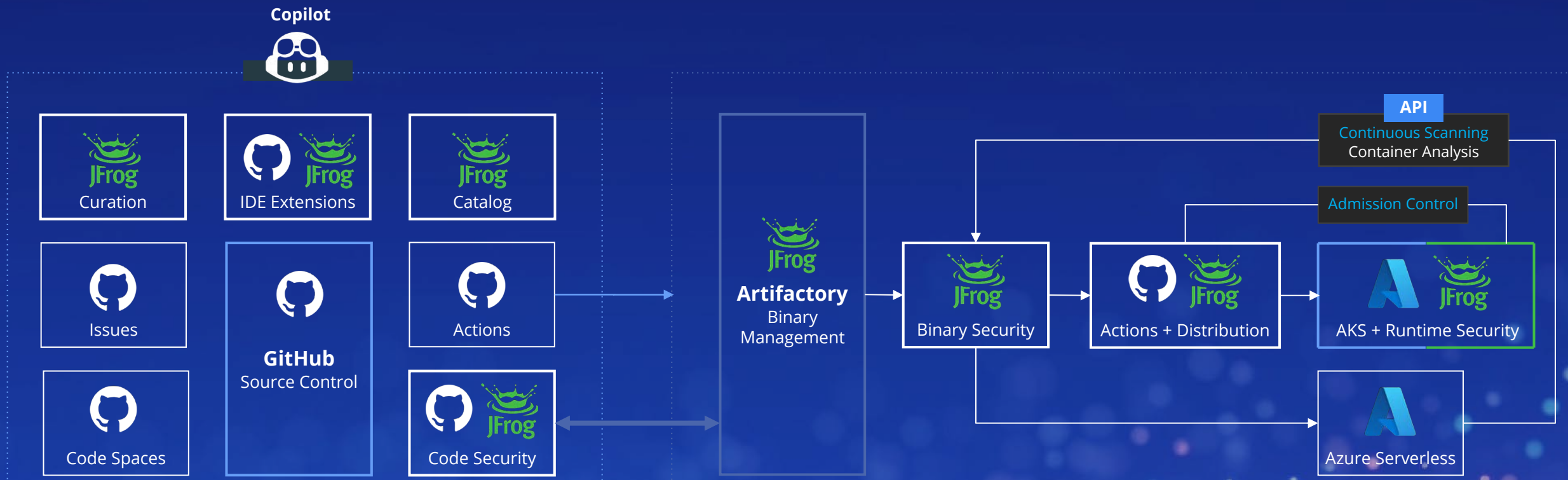


软件供应链单一可信源

- 软件供应链仓库
  - 操作系统软件供应链（yum、apt）
  - 开发语言依赖组件供应链（开发语言私服）
  - 容器供应链（镜像及helm chart）
  - AI供应链（Model、pypi、conda、Docker）
  - 传统制品（tar、zip）
- 企业单一可信制品库是软件供应链仓库，也是企业软件资产仓库。
- 低维护成本、高可用、高性能的仓库
- 没有统一管理，就无法治理。



JFrog | GitHub 深度集成，以AI驱动安全开发



JFrog MLSecOps 解决方案

科技生态圈峰会 + 深度研习

——1000 + 技术团队的共同选择

KEYLINK ing

K+峰会

K+峰会 上海站
K+全球软件研发行业创新峰会
时间：2024.06.21-22

K+峰会 敦煌站
K+思考周®研习社
时间：2024.10.17-19

K+峰会 香港站
K+思考周®研习社
时间：2024.11.10-12

K+峰会详情

AiDD峰会

AiDD峰会 上海站
AI+研发数字峰会
时间：2024.05.17-18

AiDD峰会 北京站
AI+研发数字峰会
时间：2024.08.16-17

AiDD峰会 深圳站
AI+研发数字峰会
时间：2024.11.08-09

AiDD峰会详情

AiDD

2024 AI+研发数字峰会

AI+ Development Digital summit

深圳站 11/08-09

AI驱动研发变革 促进企业降本增效

2024深圳站-议题设置

AI+产品线	LLM驱动产品创新	LLM驱动需求与业务分析	AI驱动设计与用户体验
AI+开发线	AI 原生应用开发框架与技术	AI Agents在研发落地实践	LLM驱动编程与单测
AI+测试线	LLM驱动测试分析与设计	基于LLM生成测试脚本与数据	LLM和AI应用的评测
AI+工程线	AI+DevOps 与工具 (LLM 时代的平台工程)	大模型对齐与安全	端侧大模型与云端协同
AI+领域线	领域大模型 SFT 与优化	知识增强与数据智能	大厂专场

扫描右侧二维码
查看更多会议详情

早鸟票限时抢购中 (截止到9月30日)

¥3680 早鸟票

¥2800 学生票

THANKS

谢谢观看

JFrog 官方网站：www.jfrogchina.com
JFrog 咨询热线：010 - 82023518