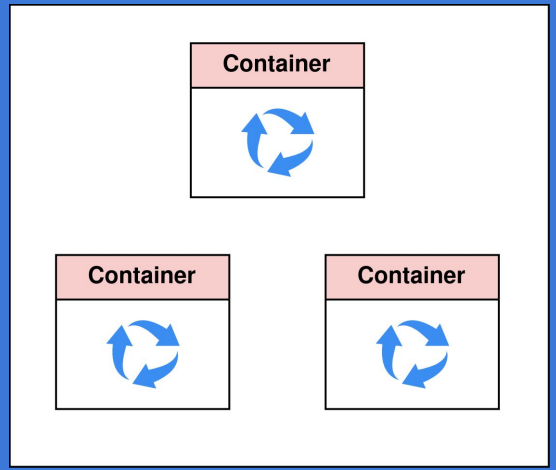
Дизайн и использование MLSecOps платформы

Киранов Д.М. Комаров А.М.

ИСП РАН

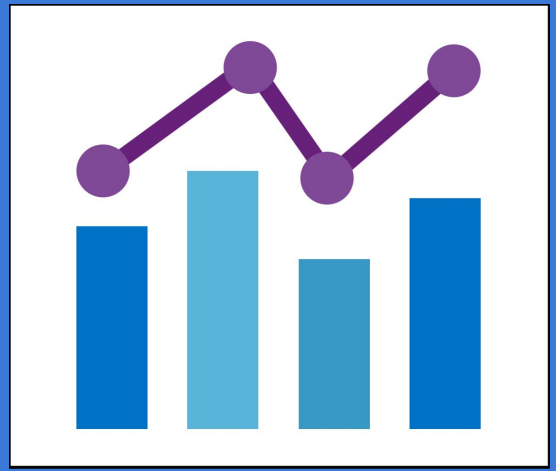
Потребности при ML-разработке: исследования

- **воспроизводимость и изолированность экспериментов**
- аналитика результатов
- версионирование активов
- управление ресурсами
- унификация процессов разработки в команде
- обеспечение доверия



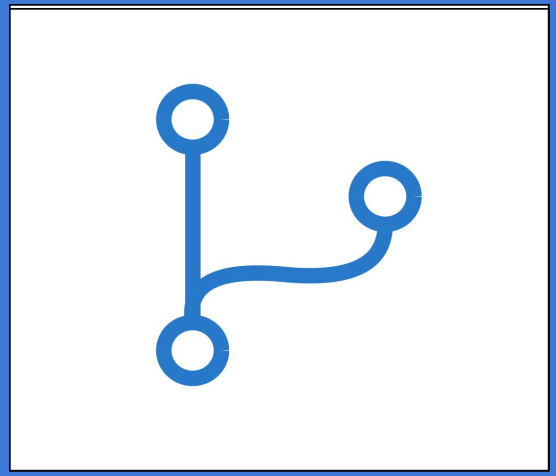
Потребности при ML-разработке: исследования

- воспроизводимость и изолированность экспериментов
- **аналитика результатов**
- версионирование активов
- управление ресурсами
- унификация процессов разработки в команде
- обеспечение доверия



Потребности при ML-разработке: исследования

- воспроизводимость и изолированность экспериментов
- аналитика результатов
- **версионирование активов**
- управление ресурсами
- унификация процессов разработки в команде
- обеспечение доверия



Потребности при ML-разработке: исследования

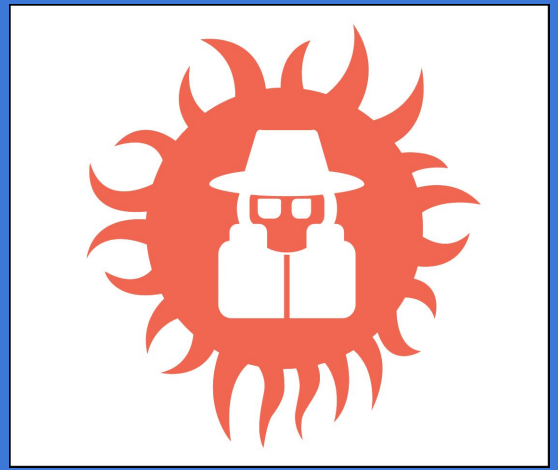
- воспроизводимость и изолированность экспериментов
- аналитика результатов
- версионирование активов
- **управление ресурсами**
- унификация процессов разработки в команде
- обеспечение доверия

Потребности при ML-разработке: исследования

- воспроизводимость и изолированность экспериментов
- аналитика результатов
- версионирование активов
- управление ресурсами
- **унификация процессов разработки в команде**
- обеспечение доверия

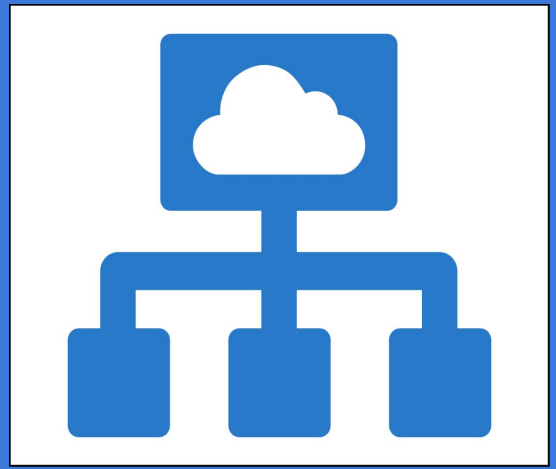
Потребности при ML-разработке: исследования

- воспроизводимость и изолированность экспериментов
- аналитика результатов
- версионирование активов
- управление ресурсами
- унификация процессов разработки в команде
- **обеспечение доверия**



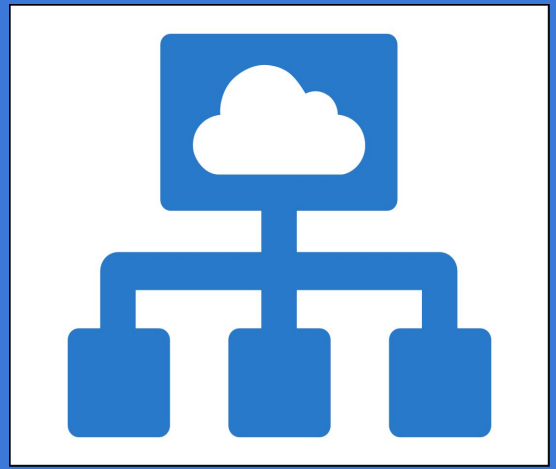
Потребности при ML-разработке: исполнение моделей

- **деплой, масштабирование**
- мониторинг производительности
- мониторинг устаревания



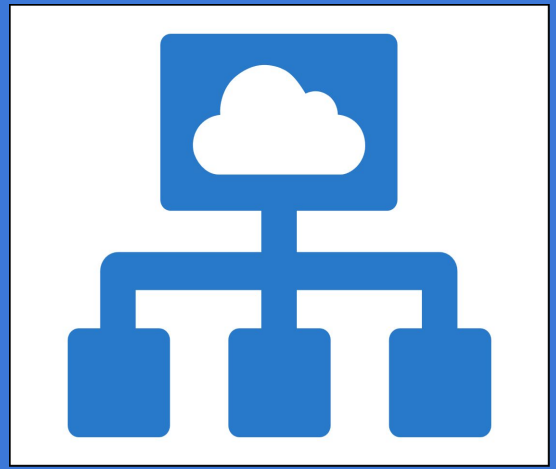
Потребности при ML-разработке: исполнение моделей

- деплой, масштабирование
- **мониторинг производительности**
- мониторинг устаревания



Потребности при ML-разработке: исполнение моделей

- деплой, масштабирование
- мониторинг производительности
- **мониторинг устаревания**



Проблемы при ML-разработке

Исследования

- воспроизводимость и изолированность экспериментов
- версионирование активов
- аналитика результатов
- управление ресурсами
- унификация процессов разработки в команде
- обеспечение доверия

Исполнение

- **деплой, масштабирование**
- **обеспечение доверия**
- **мониторинг**
- **мониторинг устаревания**

Автоматизация

MLOps

- практики и инструменты для автоматизации и упрощения процессов жизненного цикла моделей машинного обучения (ML)

• • •

Ядро платформы ДИИ

- удобство проведения множества экспериментов в команде
- инструменты сравнения результатов
- управление ресурсами
- развертывание и планирование экспериментов в приватных облаках
- эффективность внедрения, эксплуатации и поддержки моделей
- поддержка сценариев доверия и встраивание их в существующие процессы

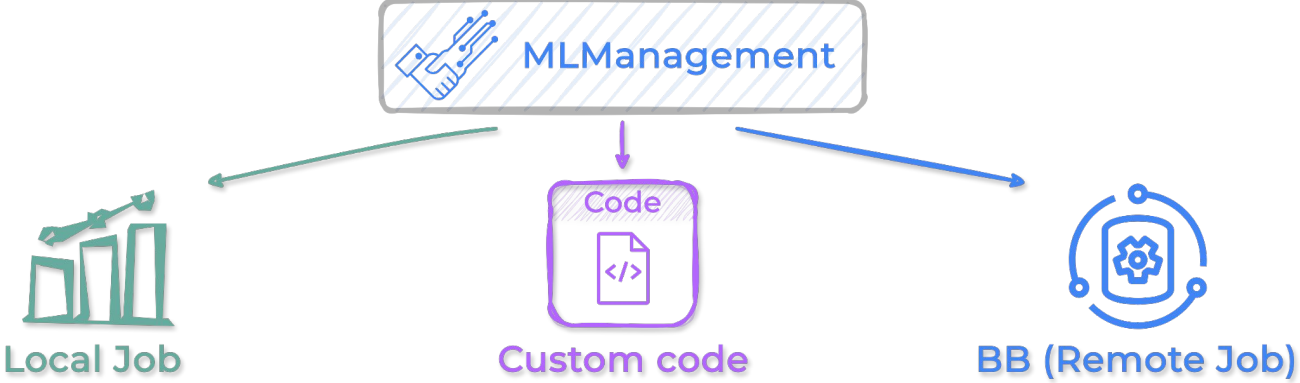
Платформа ДИИ

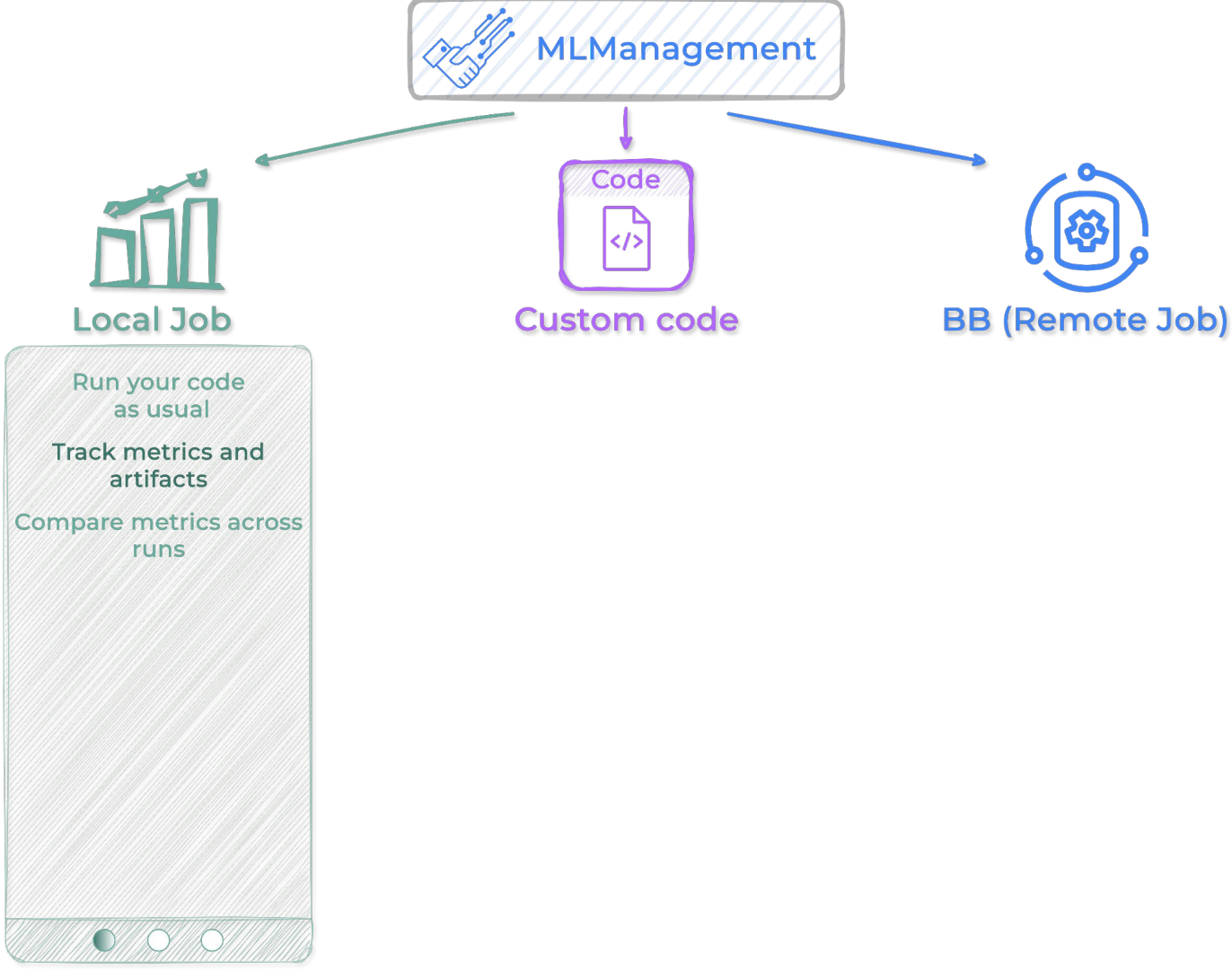
MLManagement +

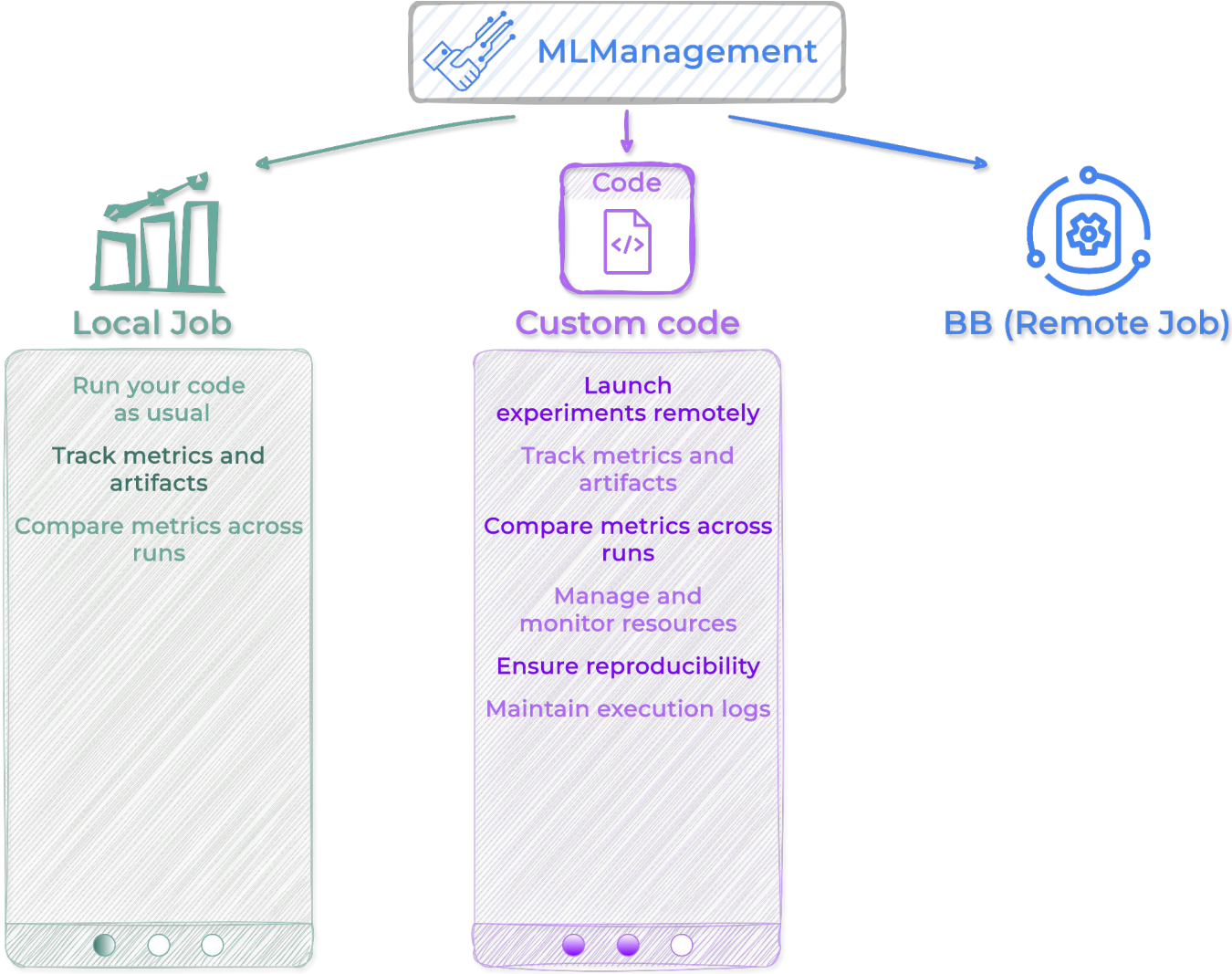
отделимый инструмент оценки доверия =

Платформа ДИИ

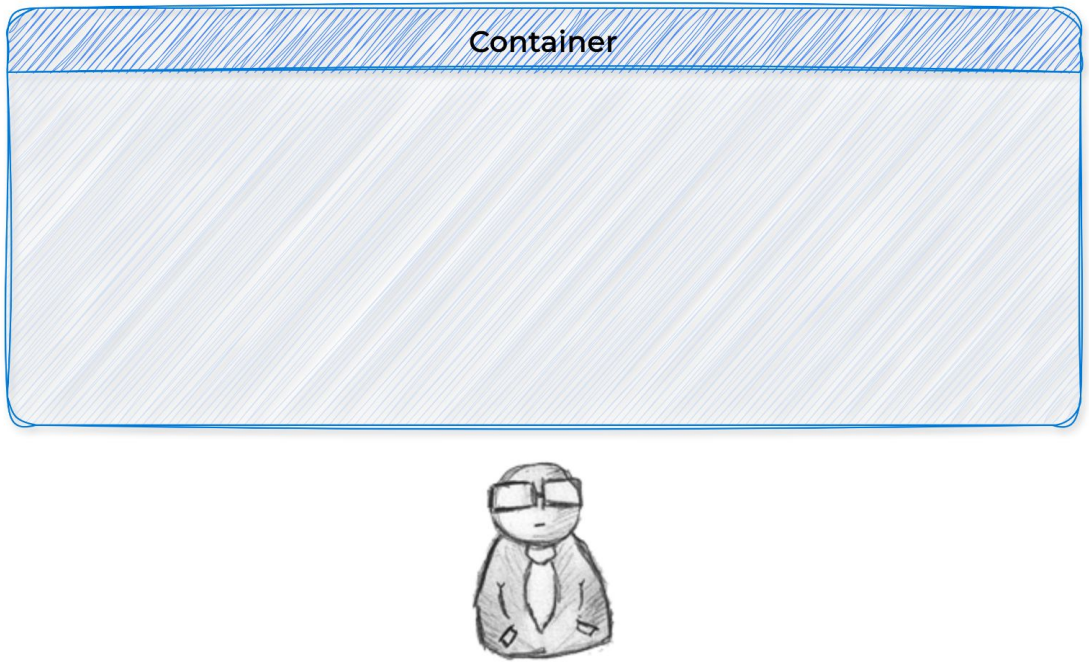
MLManagement


MLManagement
Code
Local Job
Custom code
BB (Remote Job)

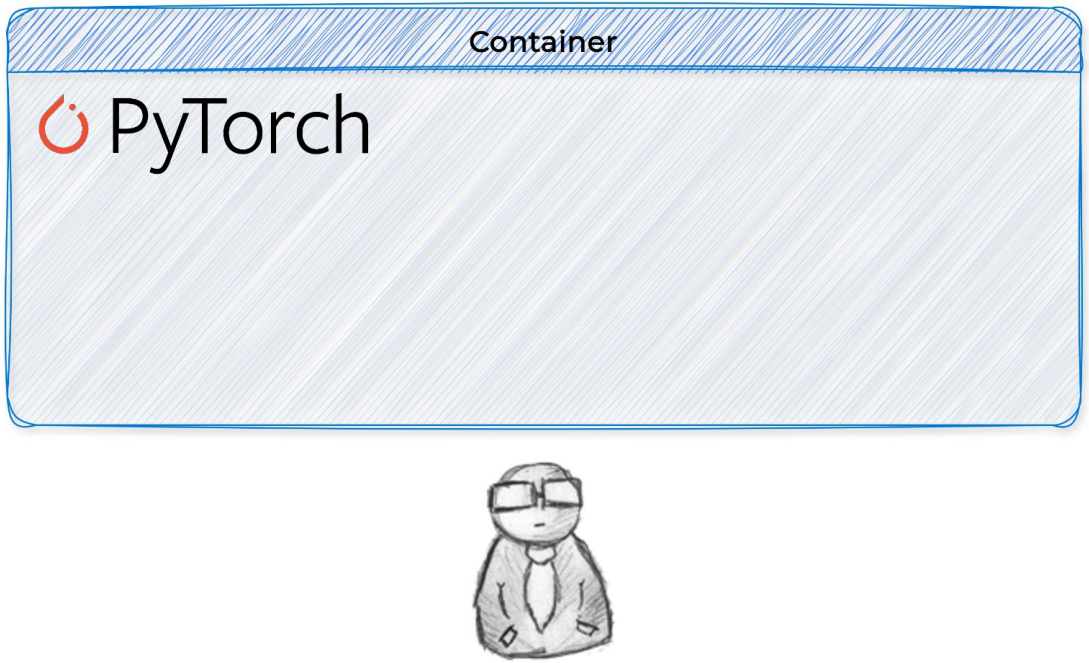

MLManagement
Code
Local Job
Custom code
BB (Remote Job)
Run your code as usual
Track metrics and artifacts
Compare metrics across runs


MLManagement
Code
Local Job
Custom code
BB (Remote Job)
Run your code as usual
Track metrics and artifacts
Compare metrics across runs
Launch experiments remotely
Track metrics and artifacts
Compare metrics across runs
Manage and monitor resources
Ensure reproducibility
Maintain execution logs

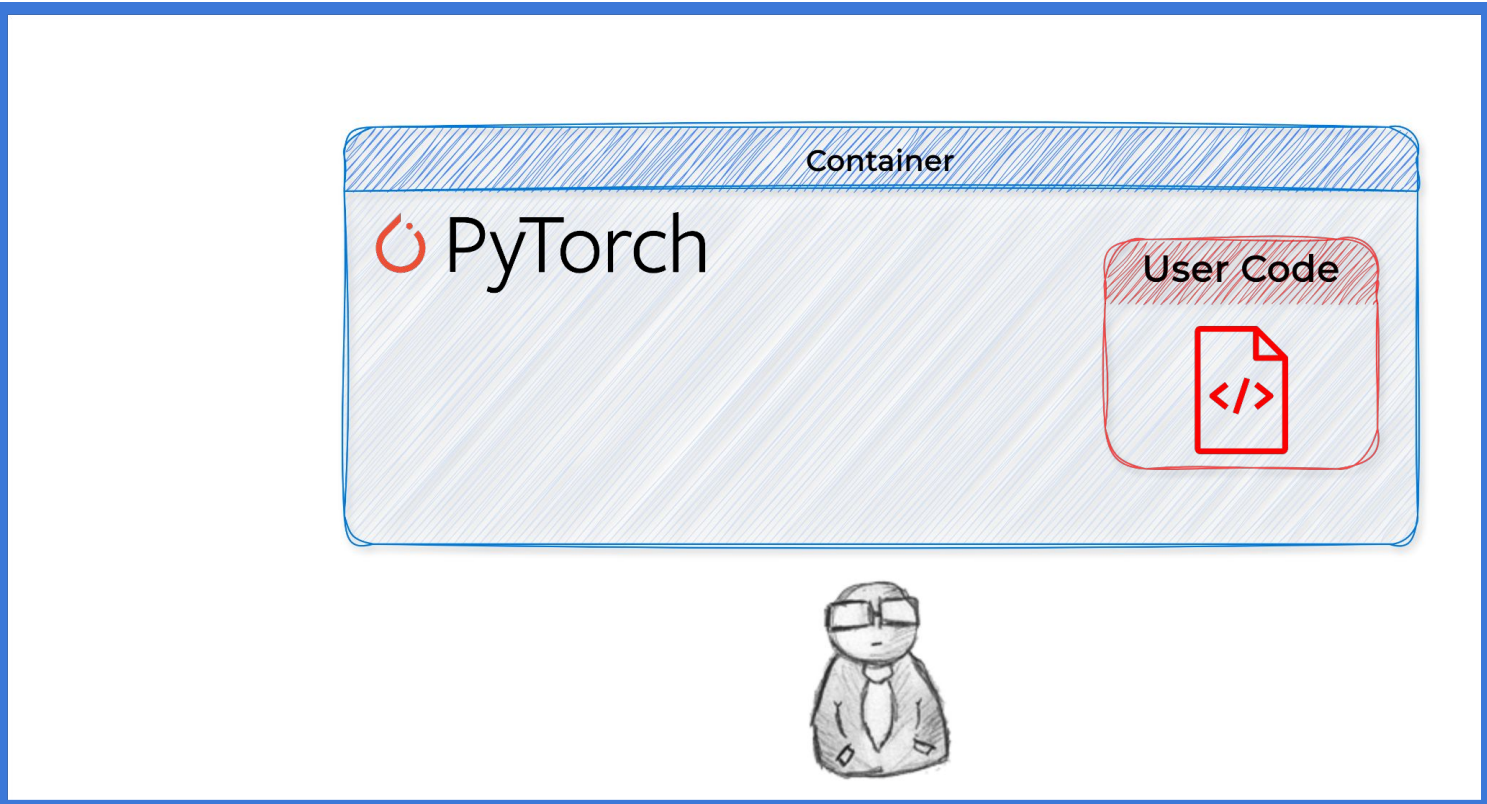
Запуск произвольного кода



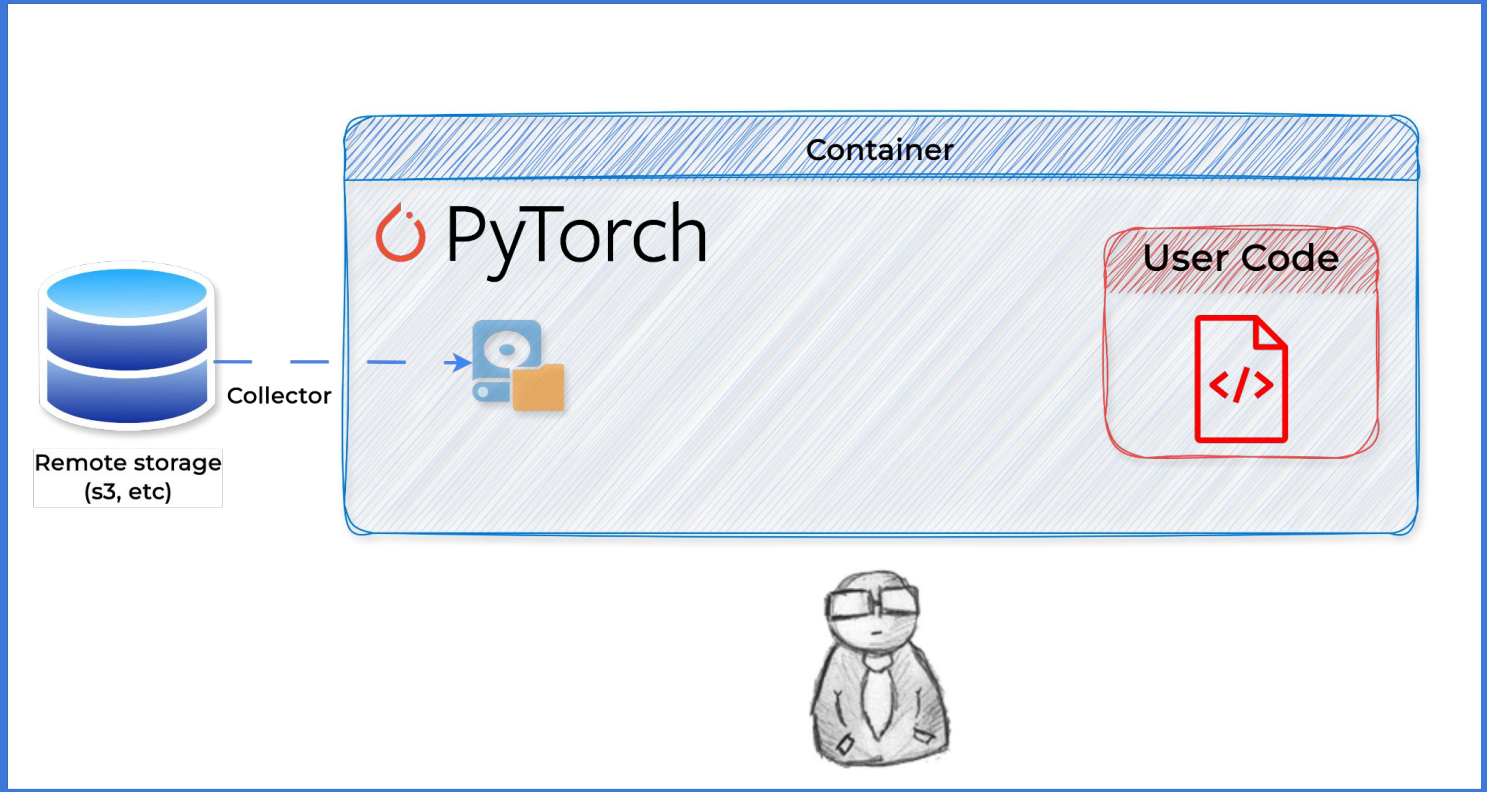
Запуск произвольного кода



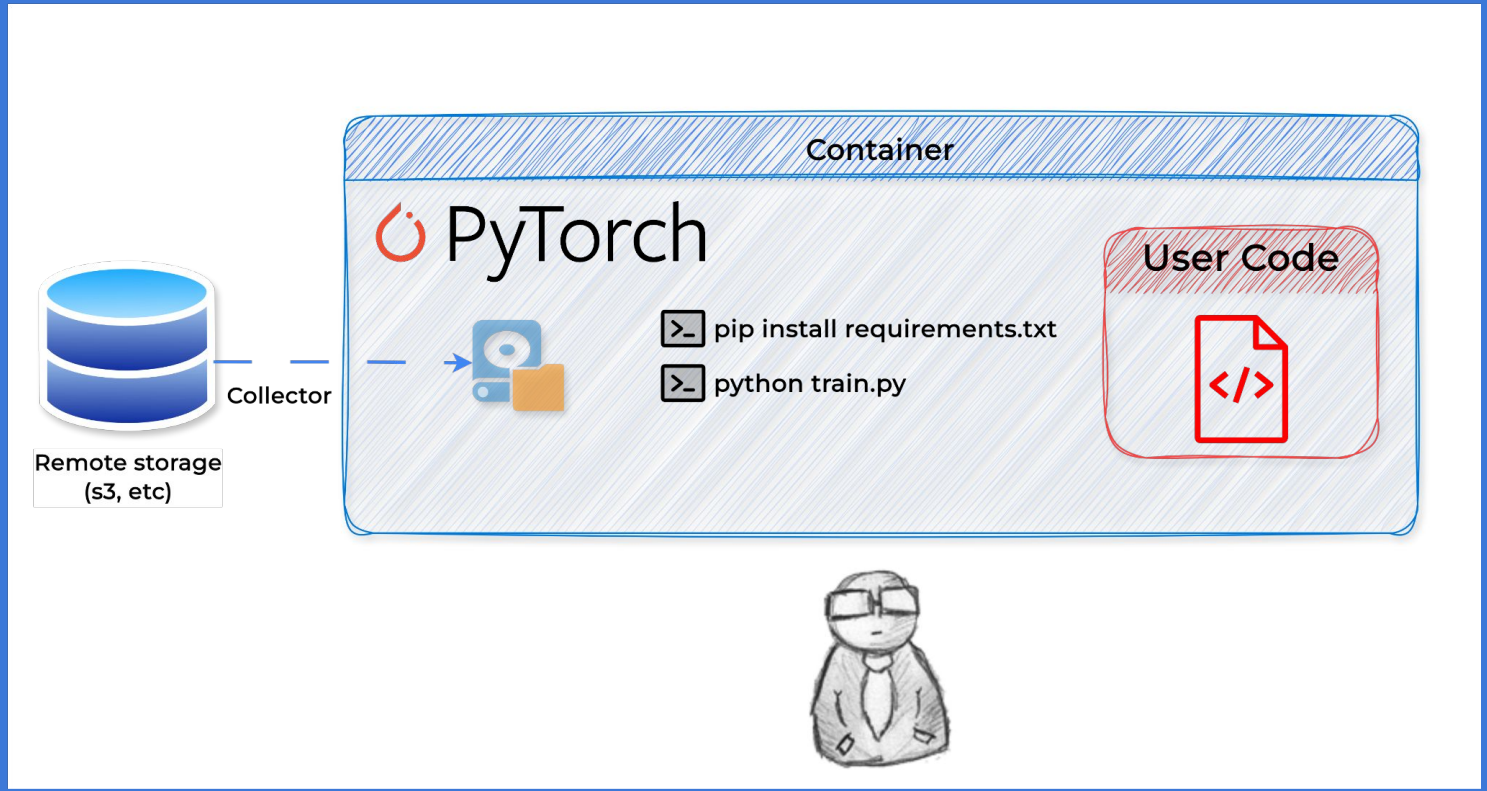
Запуск произвольного кода



Запуск произвольного кода



Запуск произвольного кода

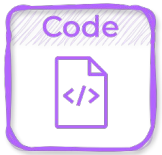


Local Job

Run your code as usual

Track metrics and artifacts

Compare metrics across runs



Custom code

Launch experiments remotely

Track metrics and artifacts

Compare metrics across runs

Manage and monitor resources

Ensure reproducibility

Maintain execution logs

BB (Remote Job)

Launch experiments remotely

Track metrics and artifacts

Compare metrics across runs

Manage and monitor resources

Ensure reproducibility

Maintain execution logs

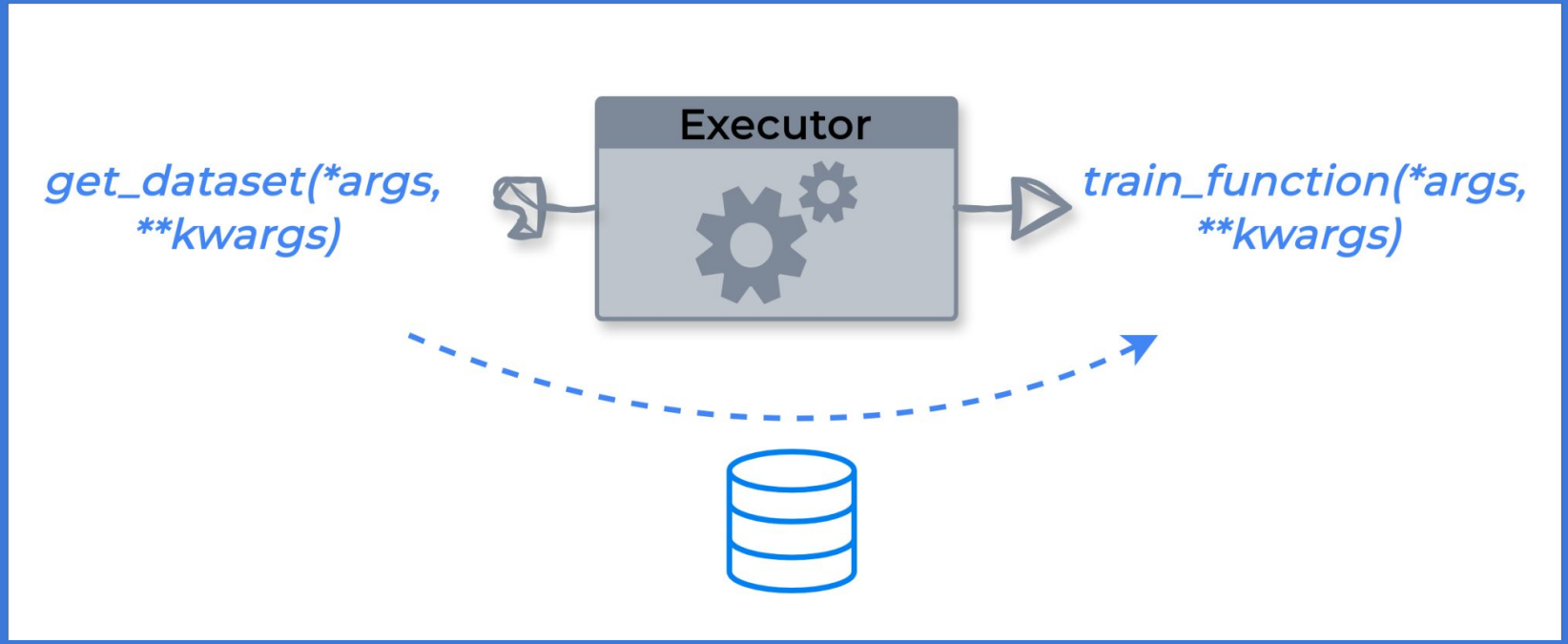
Version models and data

Reuse components

Configure flexibly

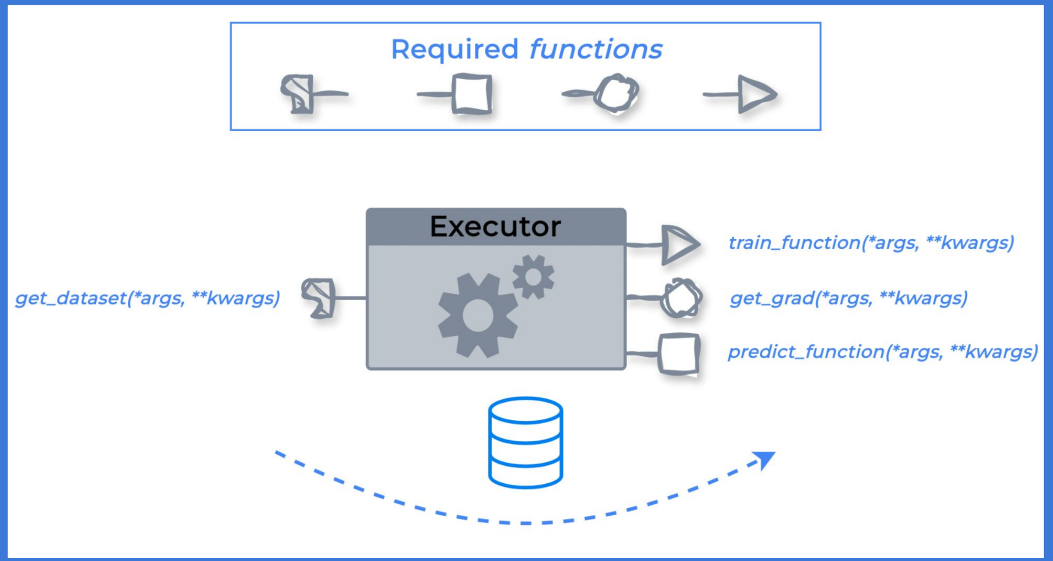
etc.

Train Executor

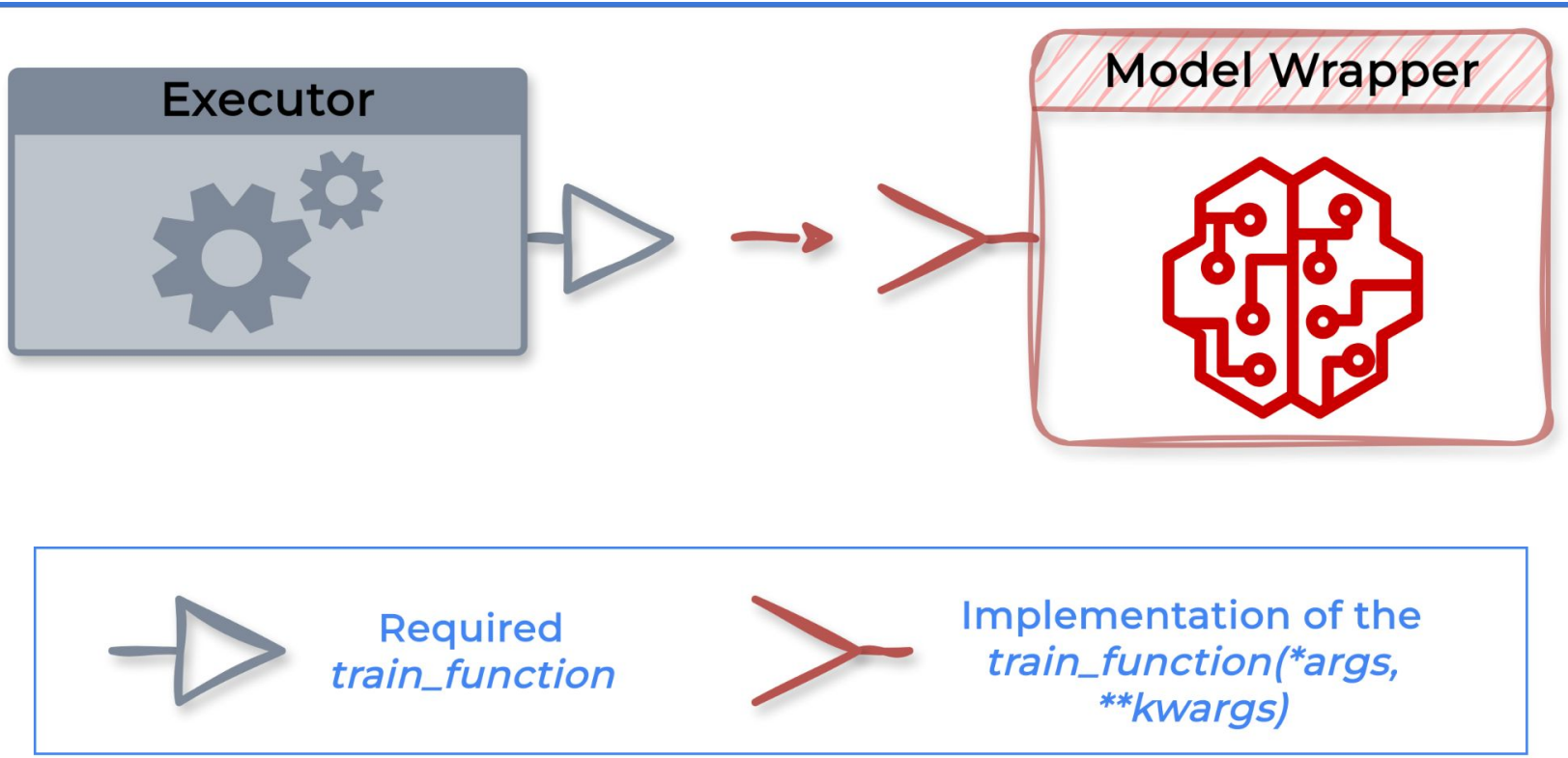


Executor

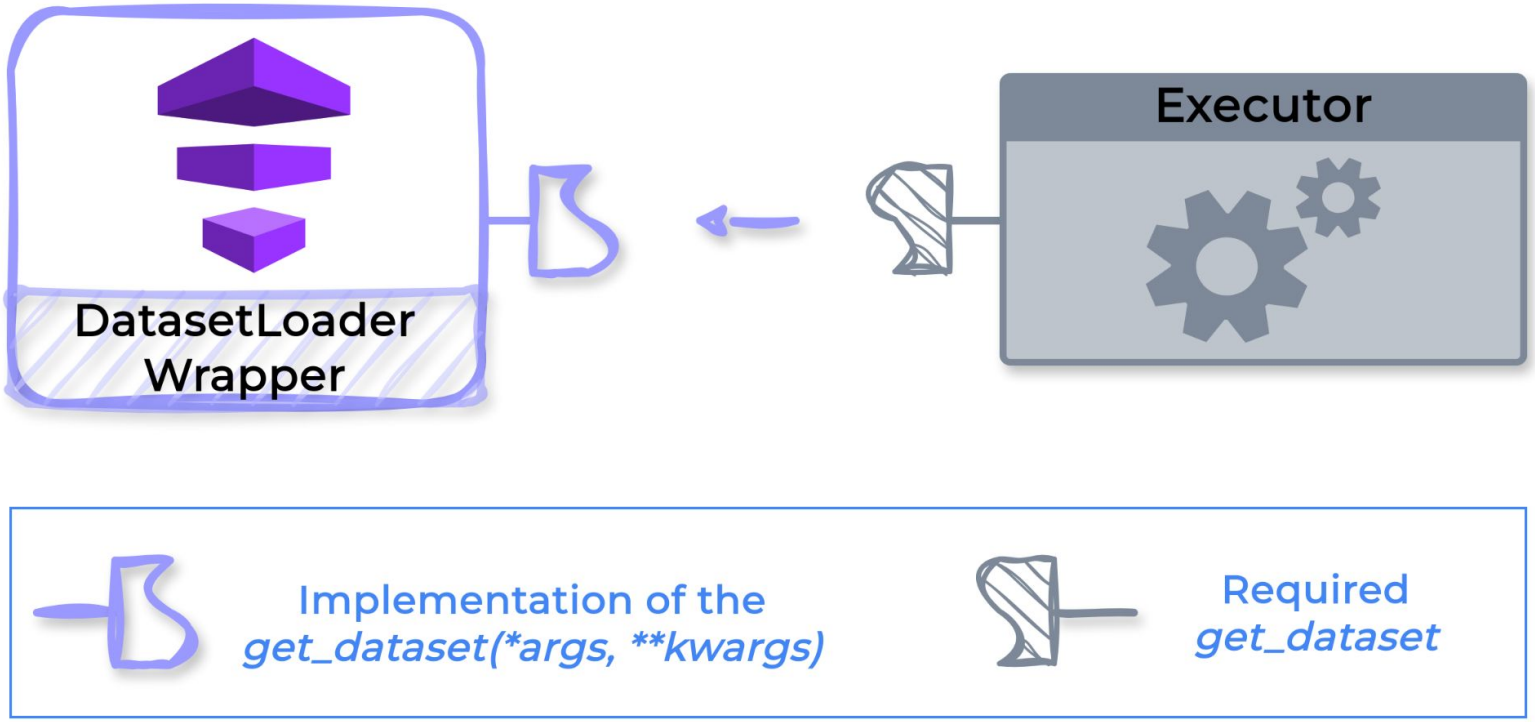
- Определяет дизайн эксперимента
- Определяет необходимые интерфейсы
- Позволяет:
  - сохранять результаты
  - предоставлять данные в модель
  - взаимодействовать с несколькими моделями и наборами данных
  - и др.



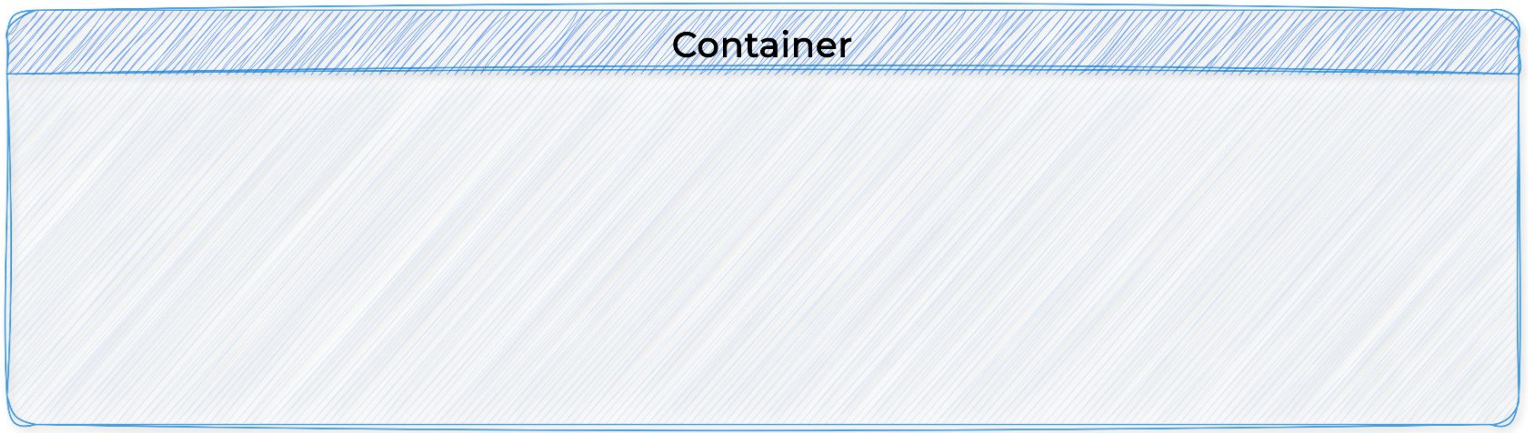
Model


Executor
Model Wrapper
Required
train_function
Implementation of the
train_function(*args,
**kwargs)

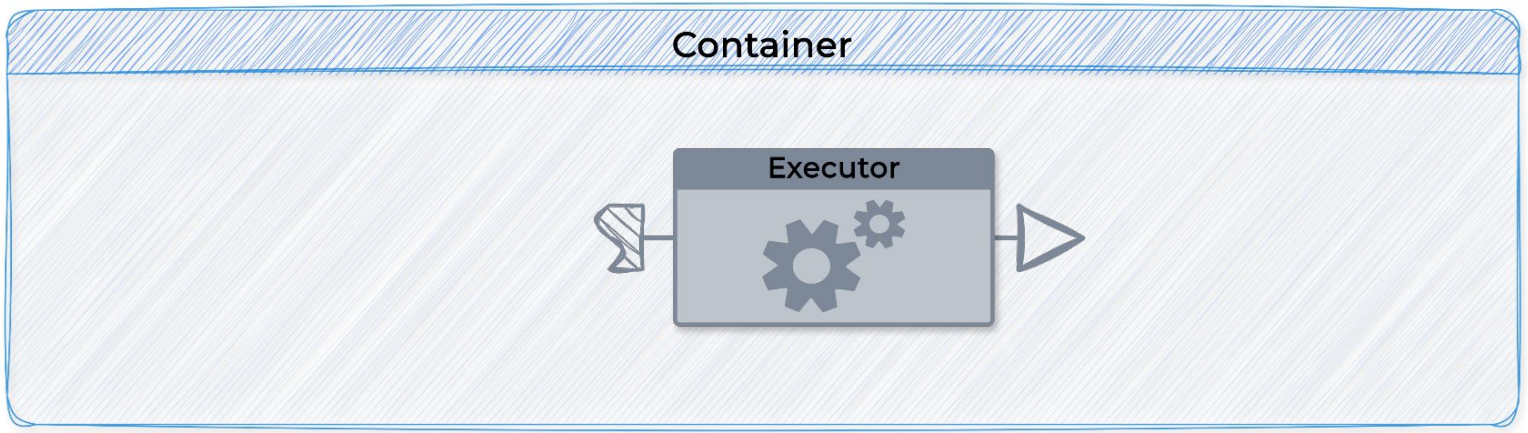
DatasetLoader



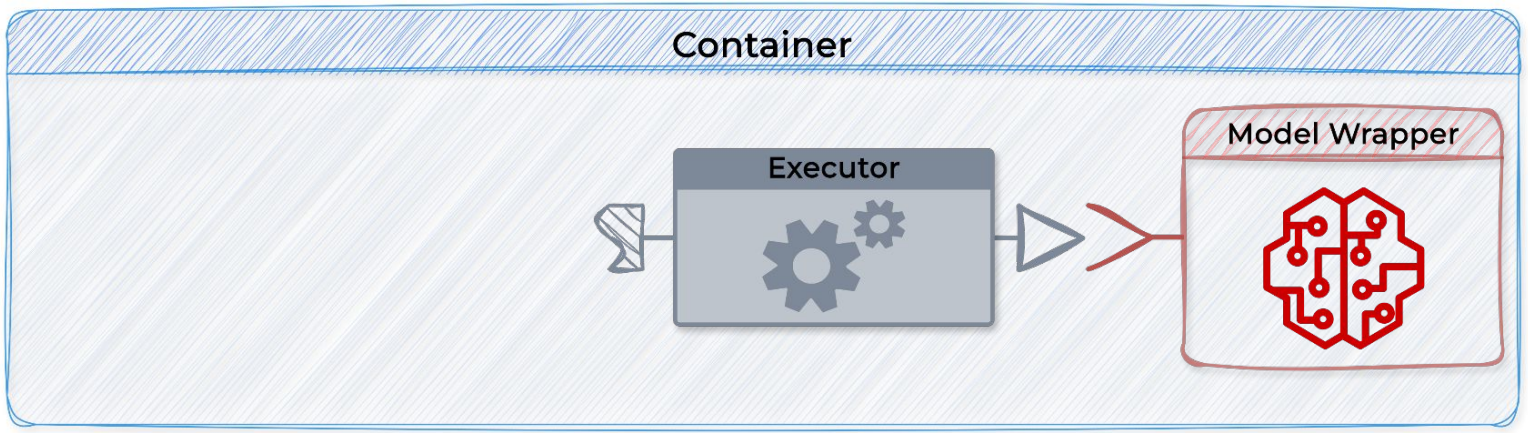
Удаленная задача



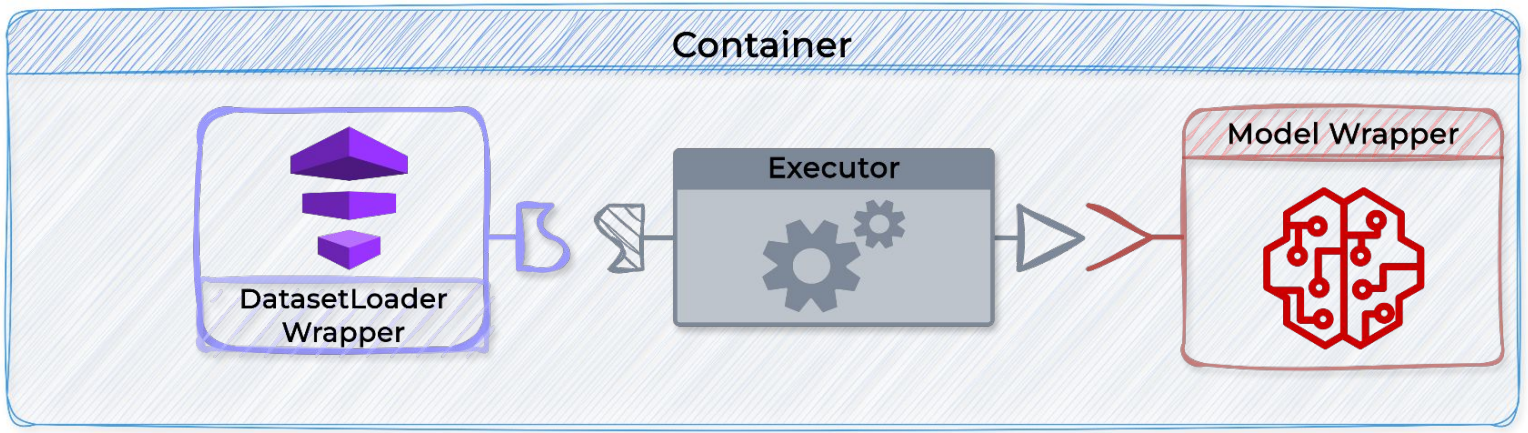
Удаленная задача



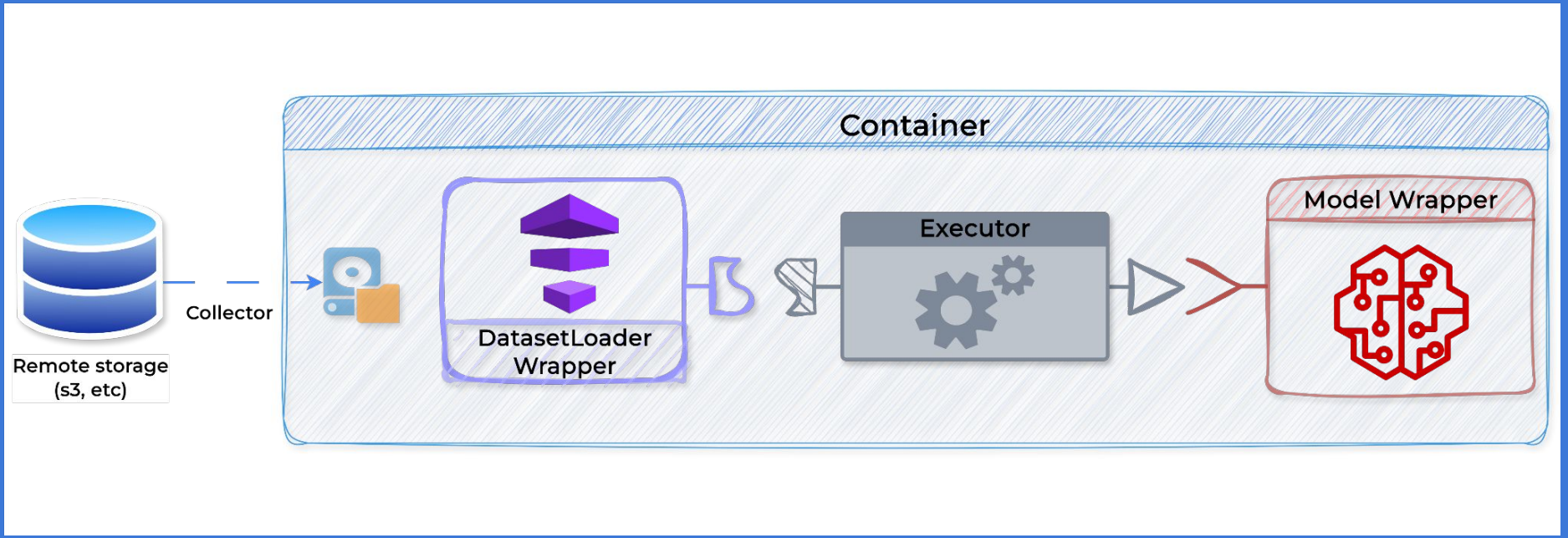
Удаленная задача



Удаленная задача

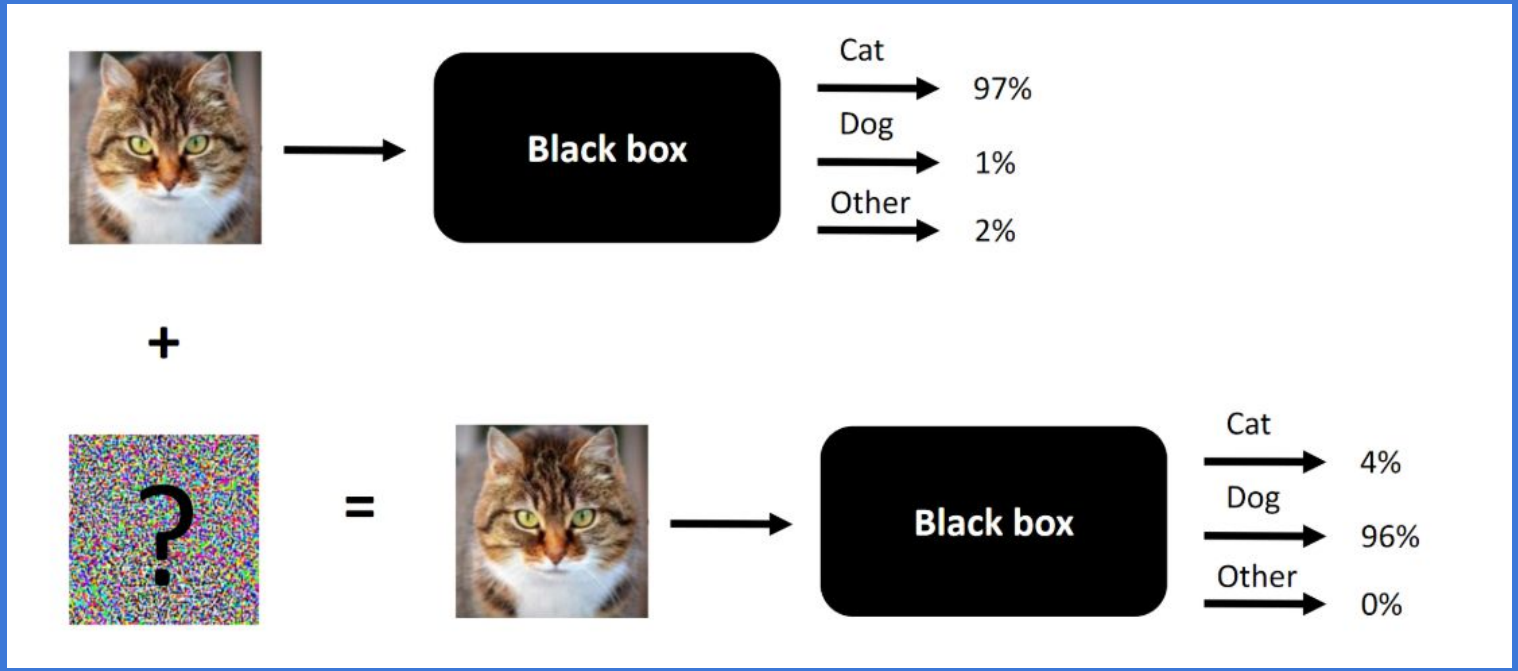


Удаленная задача



Инструмент оценки доверия системам ИИ

Пример атаки



Ключевые задачи инструмента аудита

- Оценка уязвимости модели к атакам уклонения
- Оценка риска атак на приватность
- Анализ устойчивости модели к атакам отравления

Что нужно для использования

- Model Wrapper с необходимыми интерфейсами (get_grad, get_loss, ...)
- DatasetLoader Wrapper с необходимыми интерфейсами (get_train_data, get_test_data, ...)

Поддерживаемые типы данных и задач

- Классификация изображений
- Сегментация изображений
- Детекция объектов
- Классификация на табличных данных
- Регрессия на табличных данных
- Классификация временных рядов

Варианты использования тестов

Уровень	Цель	Конфигурируемость	Формат результата	Тип пользователя / контекст применения
I - Стандартный комплексный тест	Полная оценка устойчивости к конкретному типу уязвимости	Низкая (фиксированный набор тестов)	Расширенный отчет об уязвимости	Для любого пользователя запуск "из коробки"
II - Тесты на устойчивость к конкретным атакам	Быстрый запуск одного конкретного метода атаки	Средняя (частично фиксированные параметры)	Отчет по конкретному методу	Для инженеров, интеграция в CI / CD
III - Полностью настраиваемый запуск	Максимальная свобода параметров запуска	Полная (все параметры могут быть настроены пользователем)	Набор метрик метода без доп. информации	Для исследователей и специалистов по безопасности ML

Платформа ДИИ

MLManagement

Отделимый инструмент оценки доверия