

# Внедрение общебанковской платформы MLOps в банке «Открытие»

## Experience and lessons learned



# DAY 0



# DAY 0

- Несколько Data Science команд в различных бизнесах
- Команды использовали примерно одинаковый технологический стек, но не имели промышленной платформы внедрения ML-моделей
- Потребность в снижении операционного и модельного риска при использовании моделей
- «Продвинутая аналитика» – стратегическая инициатива, направленная на увеличение использования ML для решения бизнес задач
- Наличие массовых продуктов, основанных на ML-моделях

# GOALS

- **Снижение операционного и модельного рисков**
  - Версионирование данных
  - Версионирование моделей
  - Управление моделями
- **Сокращение T2M**
  - Бесшовное внедрение моделей в эксплуатацию и удобная поддержка
- **Возможность оперативного подключения новых Data Scientist's, совместная работа команд разработки моделей**
- **Единая точка обновления ПО**
- **Унифицированный технологический стек и единый центр компетенций**
- **Эффективное использование вычислительных ресурсов**

# ВЫБОР РЕШЕНИЯ

- **Выбор между:**
  - «коробочным решением» от вендора
  - Заказное решение на стеке Open Source
- **Требования к собственной платформе:**
  - Cloud-ready – желание разместить платформу в уже имеющемся kubernetes
  - Независимость от вендора
  - Реализация концепции MAAS
  - Поддержка сред разработки на языках R Python -> ориентация на экспертизу DS на рынок труда

# DAY 1

# MVP



# DAY 1. MVP

- **Выбрали для примера реальную модель: модель СРП МСБ**
- **Адаптировали модель для применения в MLOps-платформе**
- **Создали технологическое приближение MLOps-платформы:**
  - Развернули и убедились в работоспособности основных компонент
  - Реализовали pipeline внедрения batch-модели на выделенных виртуальных машинах
- **Презентовали решение руководству банка**

# DAY 2 PROJECT





# DAY 2. PROJECT

- **Что способствовало старту проекта:**
  - Бизнес уже использовал модели
  - Платформа изначально проектировалась как промышленное решение со статусом mission critical
  - Kubernetes – одобренная и развернутая в Банке платформа
- **Вызовы:**
  - Секвестрование. Сокращение бюджета на развертывание целевой платформы
  - Организация доступов. Много компонент, много новых ролей пользователей в системе
  - Масштаб проекта и количество вовлеченных подразделений

ВЗГЛЯД  
СО СТОРОНЫ  
ТЕХНОЛОГИЧЕСКОГО  
ПАРТНЕРА

Neofie 

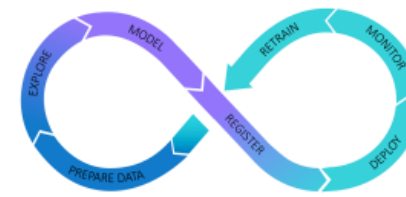
# КОНЦЕПЦИЯ СОЗДАНИЯ MLOPS ПЛАТФОРМЫ



MLOps = DataOps + DevOps + Collaboration & Governance



Feature engineering  
Data versioning



CI/CD pipelines

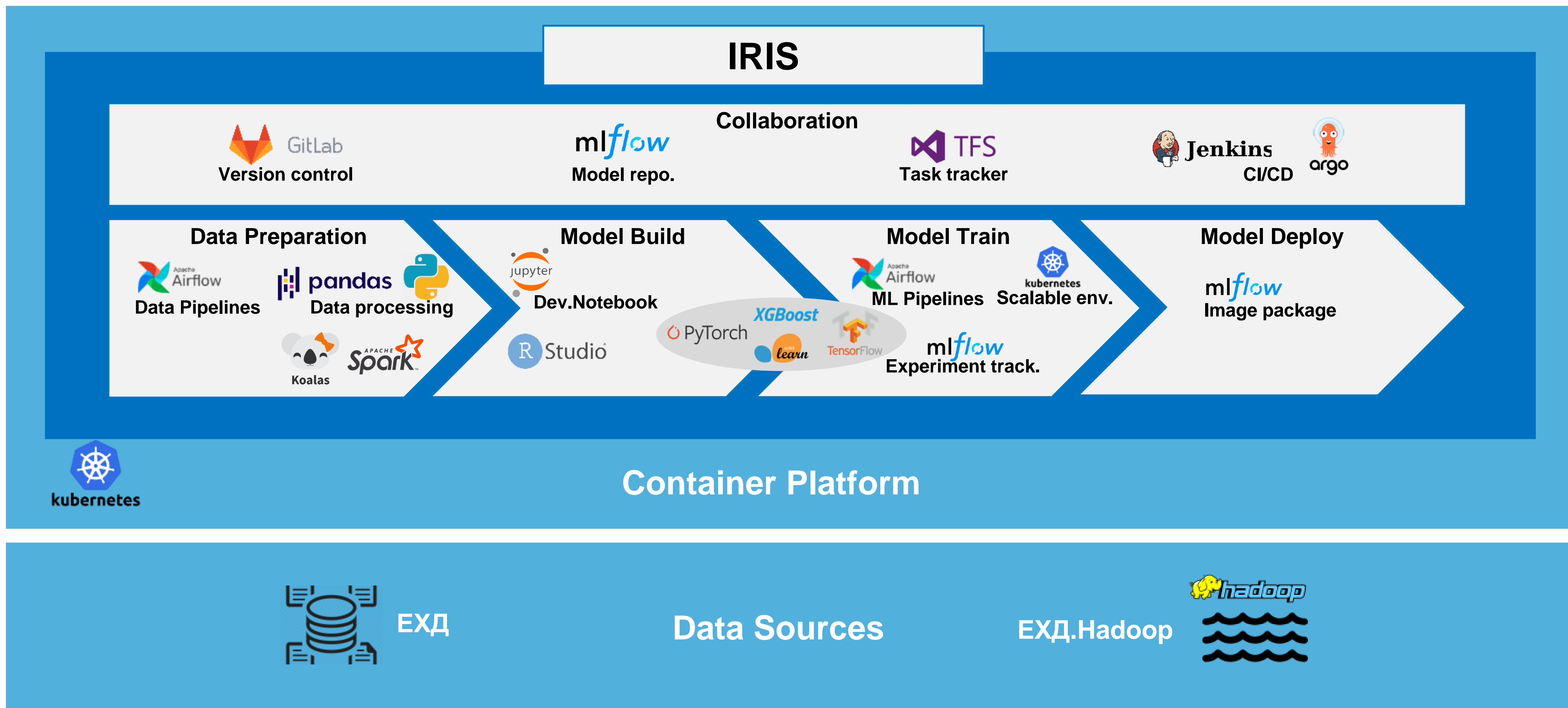


Model registry  
Model lifecycle mgmt.

- **Open Source**
- Возможность использовать любые современные ML библиотеки (Tensorflow, XGBoost, Scikit-learn)
- **Cloud-ready**
- Возможность развернуть платформу: on-premise (Kubernetes, OpenShift), public cloud (AWS, GCP)

- **DataOps**
- Автоматизация feature engineering
- Data versioning
- **DevOps**
- Автоматизация CI/CD процессов
- **ModelOps**
- Реестр моделей и артефактов
- Автоматизация процессов жизненного цикла

# DAY 2. РЕЗУЛЬТАТЫ ПРОЕКТА



# DAY N. ПЛАНЫ ТЕХНОЛОГИЧЕСКОГО РАЗВИТИЯ



Контроль версий данных



Продвинутая платформа  
деплой моделей



Feature Store

# DAY {1..N} ...НЕ ЗАБЫТЬ ВЗЯТЬ В ПРОЕКТ

## Эксперты

- Объяснить концепцию
- Убедить коллег из различных подразделений в правильности подхода
- Научить пользователей пользоваться платформой

## Менеджер

- Платформа на пересечении нескольких подразделений – требуется большое количество усилий, чтобы выстроить коммуникации: Блок ЕХД, IT Operations, IT Architecture, Data Engineers, Data Scientist, Владельцы систем – потребителей моделей

## MLOps-инженеры

- Развернуть и настроить инфраструктуру платформы
- Реализовать CI/CD/CT
- Поддержка Data Engineers и Data Scientist при выносе моделей

**Data Engineers - подготовка данных – самый трудоемкий процесс!**

## Data Scientist

- «от Data Scientist'а слышу» – передача опыта использования технологии Data Scientist'ам Банка
- Служба поддержки
- Пока платформа не находится в ПЭ – решаем проблемы самостоятельно (требуется и 2 и 3 линии)

# DAY 3 & LESSONS LEARNED

# DAY 3. ЧТО ЕСТЬ СЕЙЧАС

## Среда применения:

- Развернута в Kubernetes
- Отказоустойчива
- Передается в ПЭ
- Интегрируется со всеми кредитными конвейерами
- Идет миграция batch и online моделей

## Среда разработки:

- Развернута в Kubernetes
- Настроена ролевая модель
- Повышена отказоустойчивость
- Идет миграция команд разработки
- Выделены стримы развития среды по подразделениям



# LESSONS LEARNED

- Пилотный проект нужен
- Пилотный проект позволит понять, подходит ли выбранная архитектура и ее компоненты под процессы банка и под нужды конкретных специалистов – DS
- Необходимо изначально проработать архитектуру платформы совместно (архитекторы технологического партнера + архитекторы Банка) и регулярно «сверять часы»
- Иметь план B, C, ... на случай секвестирования бюджета
- На старте определиться с уровнем отказоустойчивости системы
- Максимальная централизация на первом этапе проекта, децентрализация на последующих
- требуется немалая просветительско-разъяснительная работа для привлечения специалистов DS различных подразделений к работе в единой среде

# DAY {1..N} ПЛАНЫ РАЗВИТИЯ

- Миграция большинства процессов моделирования на платформу
- Ввод в ПЭ среды разработки
- Реализация инструмента управления жизненным циклом моделей
- Создание промышленных витрин для обучения и применения моделей
- Реализация продвинутых методов ML:
  - Кластер GPU
  - Витрины для графовой аналитики