детский мир

дм тех

Hadoop Is Not Dead Just Secure!

Александров Антон

Содержание

детский мир

- Обзор текущих трендов
- Типичные кластера Hadoop
- Наш путь
- Итоги

01 О докладчике

О себе

детский мир

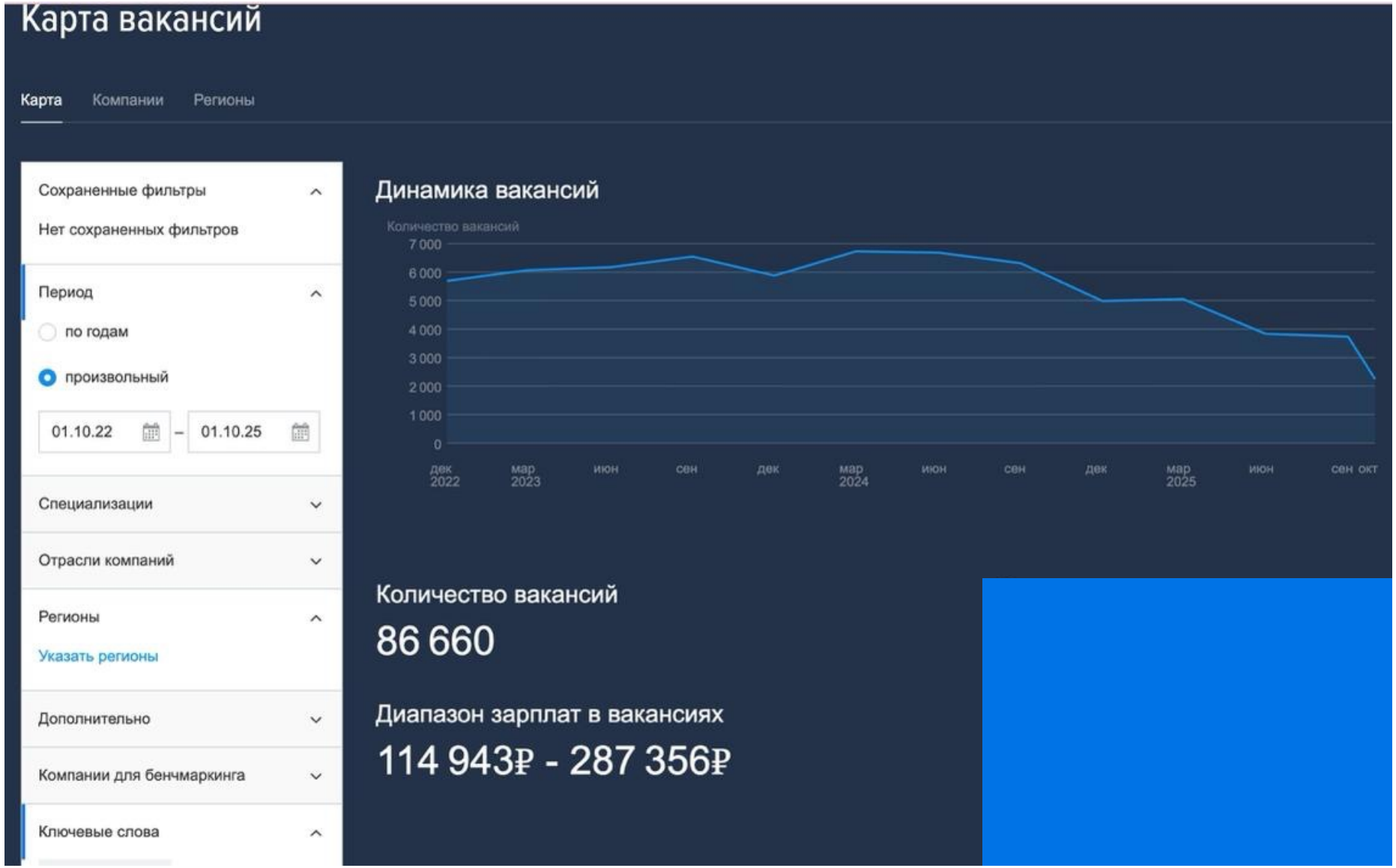
Александров Антон

- Head of Data Engineering
- Физтех
- Знаком с Hadoop и всем что его окружает уже 10 лет

02 Обзор трендов

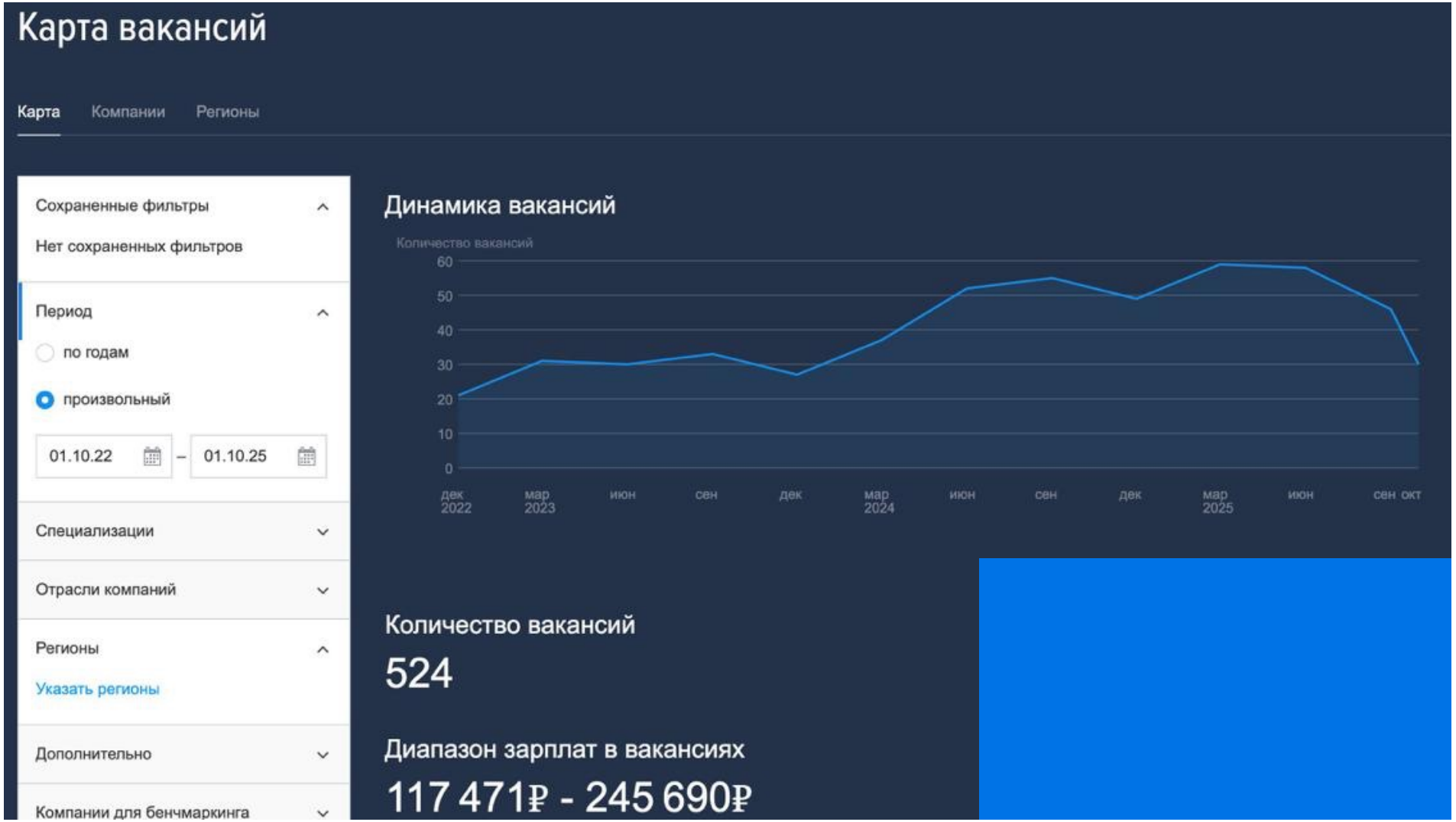
Вакансии на hh / Postgresql

детский мир

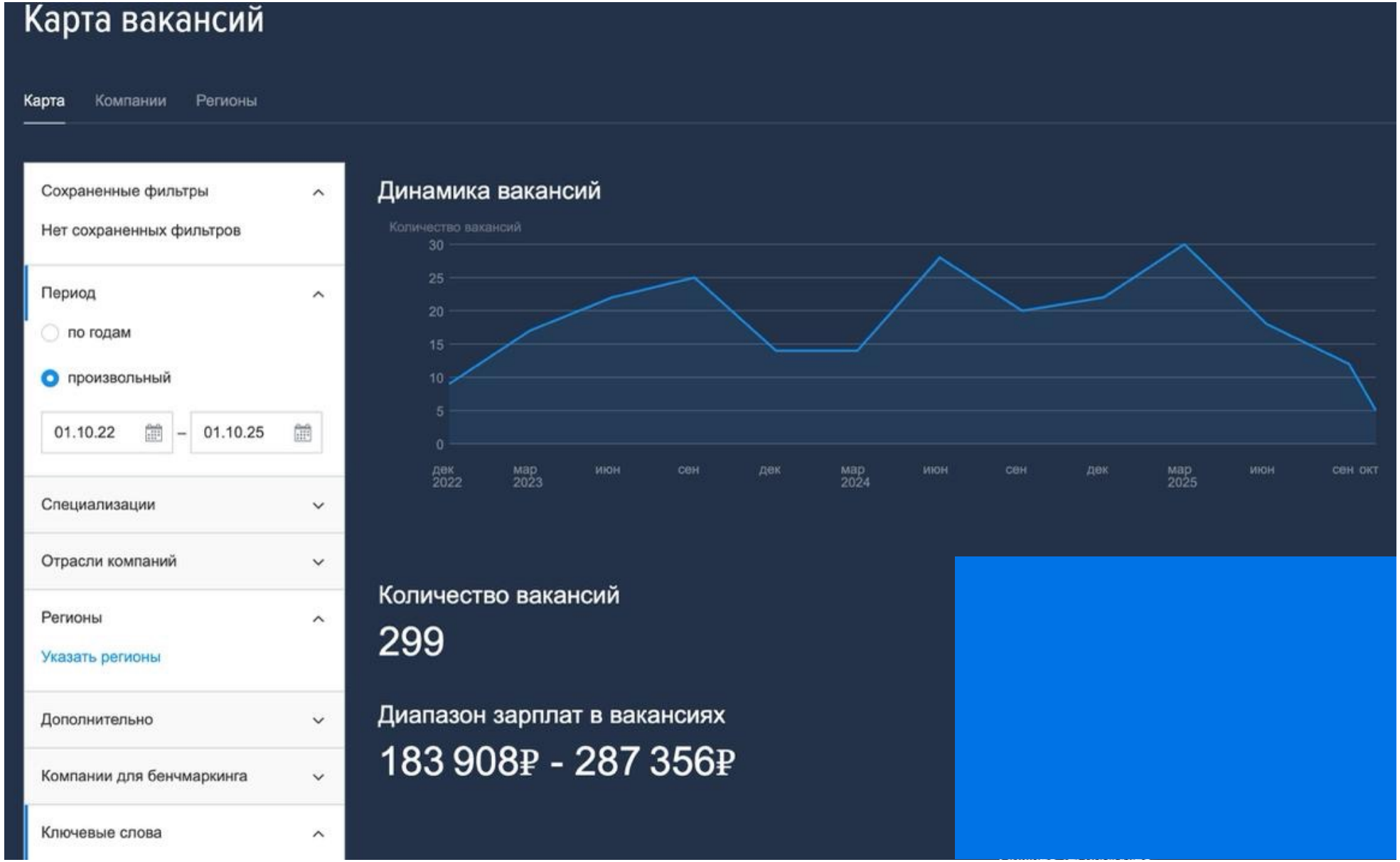


Вакансии на hh / Ceph

детский мир

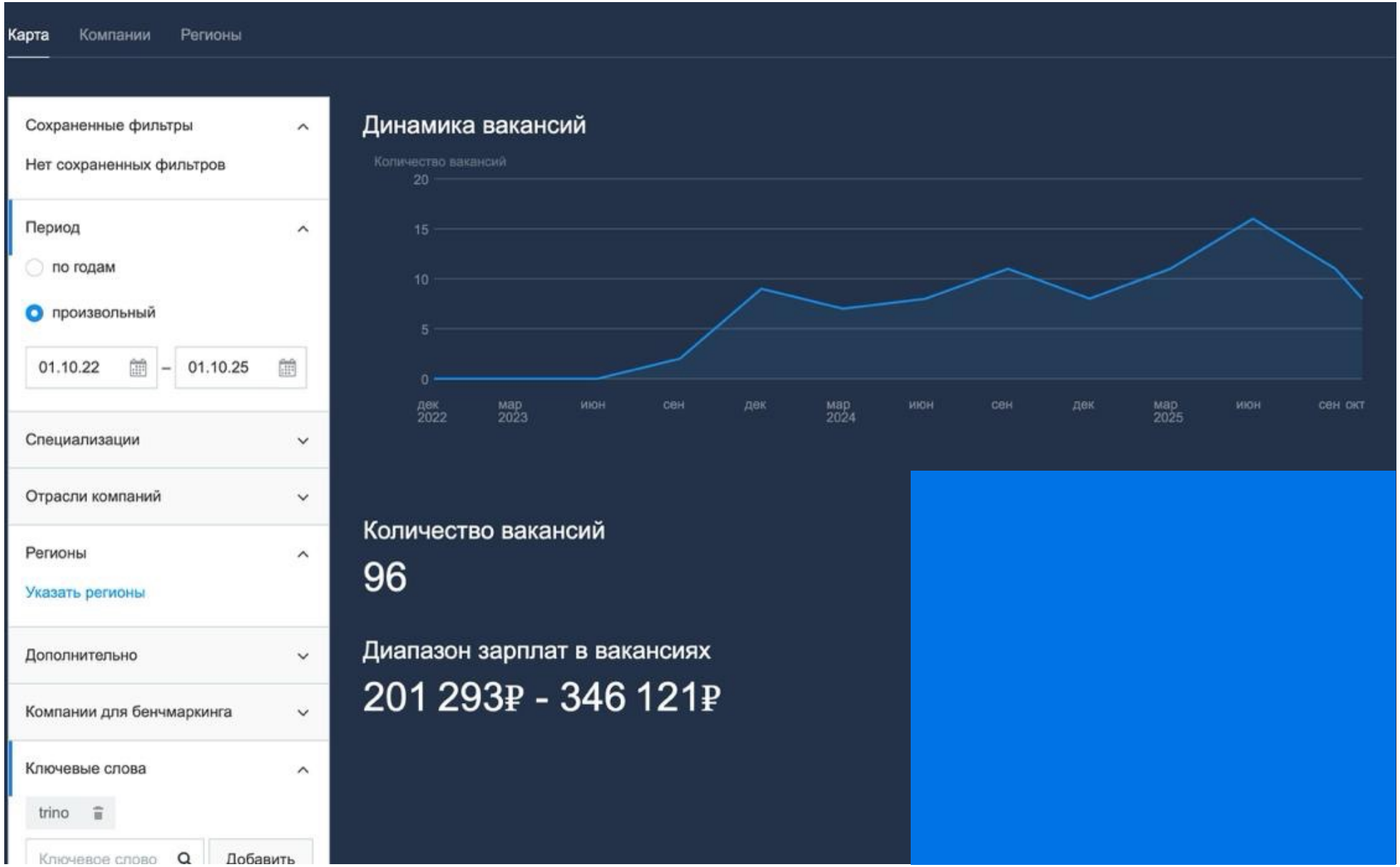


Вакансии на hh / Minio



Вакансии на hh / Trino

детский мир



Вакансии на hh / Hadoop

детский мир

Вакансии на hh / Собираем вместе

детский мир

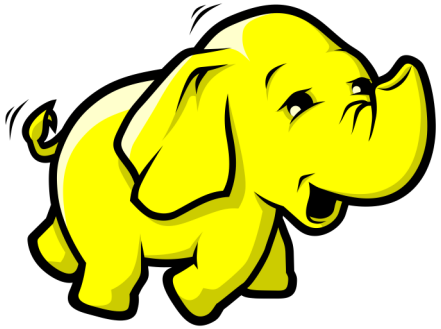
Ключ	Вилки	Количество	Открыто на 05.10.25
postgresql	114-287	86660	5 866
ceph	117-245	524	135
minio	183-287	299	187
trino	201-346	96	92
hadoop	149-344	7193	502

Hadoop

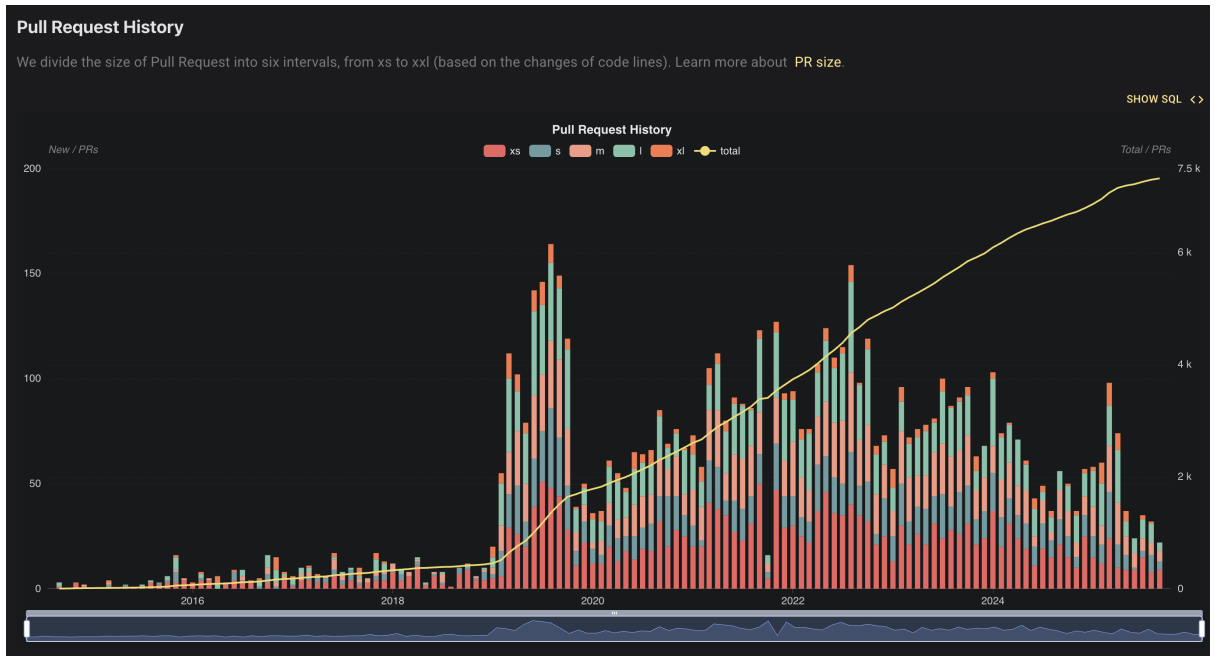
детский мир

Java 17 https://issues.apache.org/jira/browse/HADOOP-17177
Java 25 https://issues.apache.org/jira/browse/HADOOP-19486

Задачи делаются, Hadoop **медленно,** но развивается

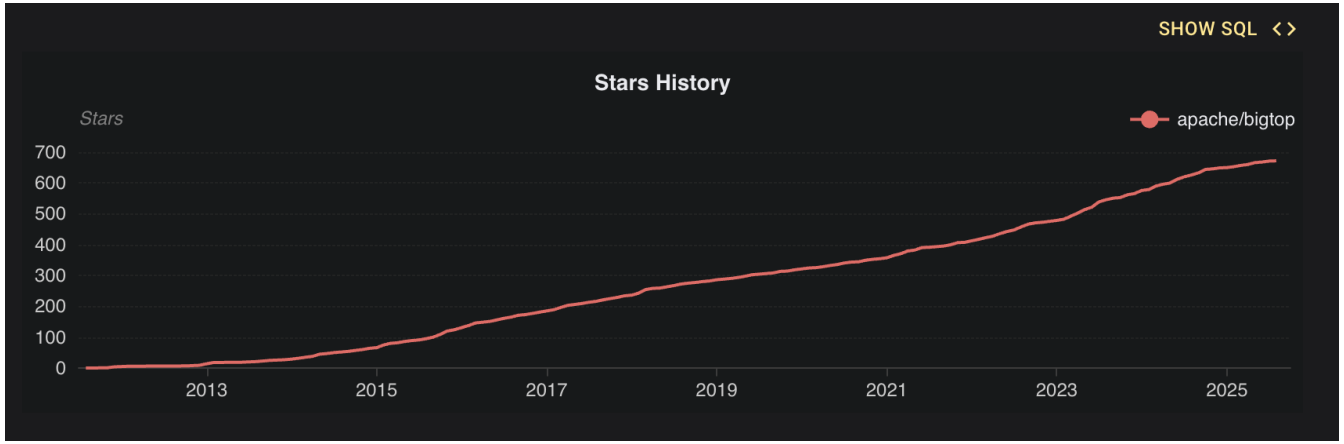


PR 2014-2025



Line of code 2019-2025

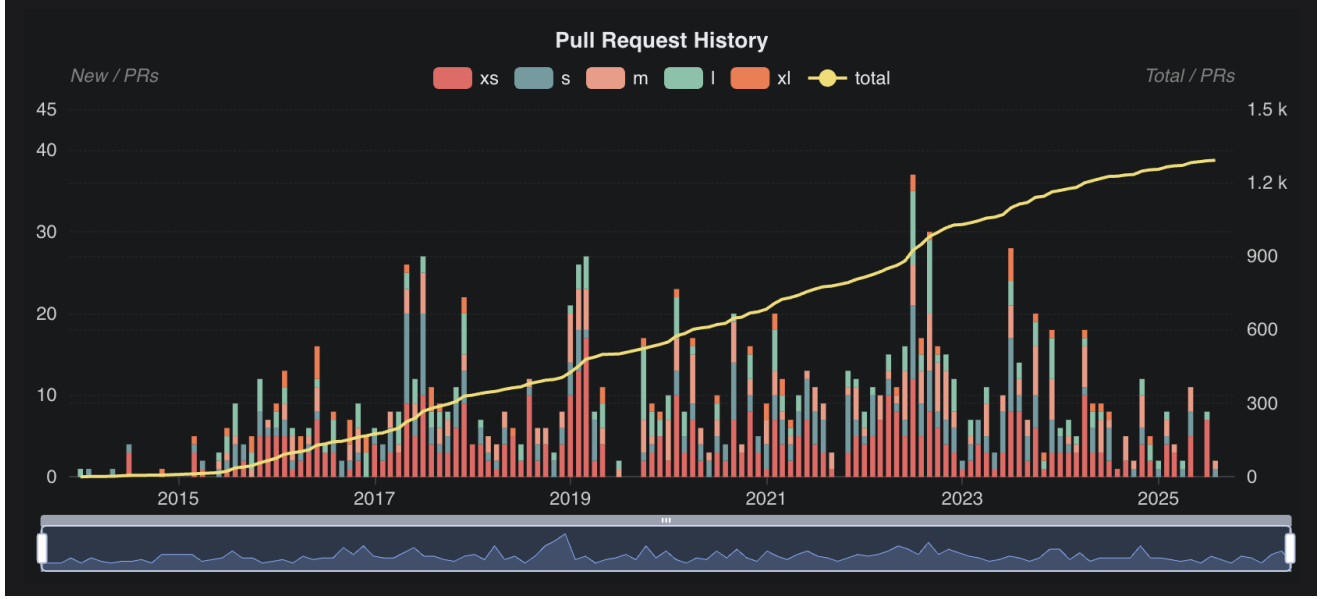
Apache Bigtop

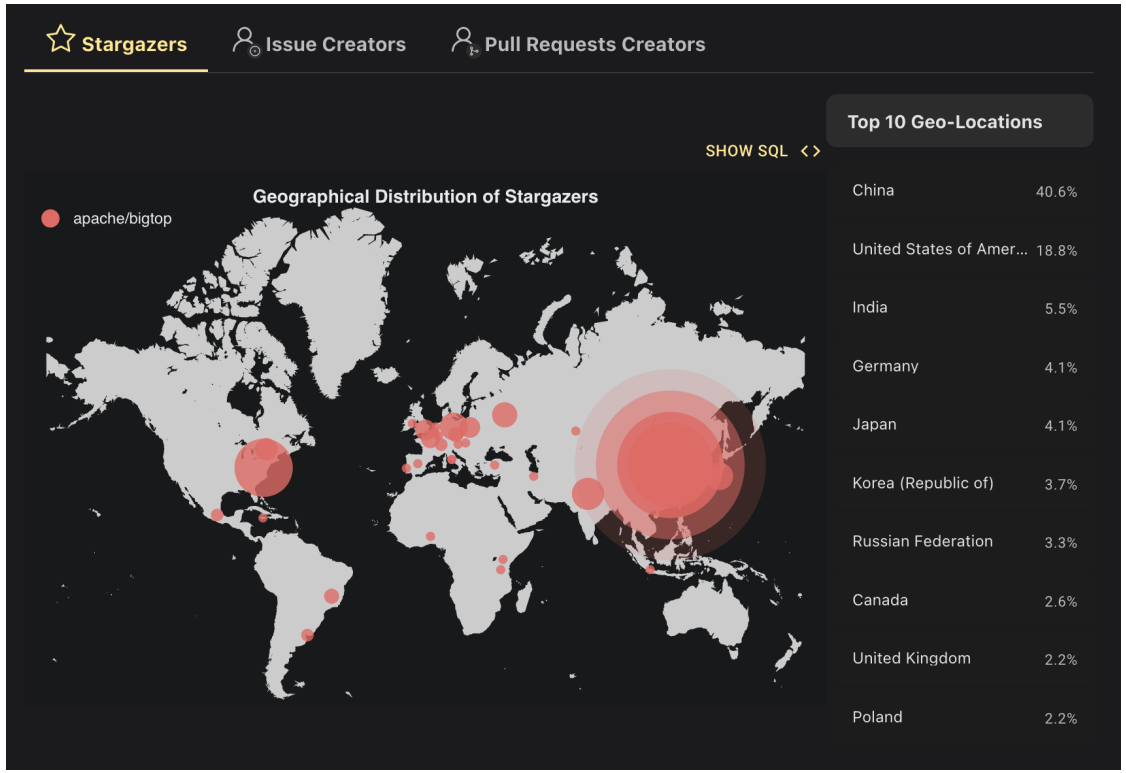


детский мир

Pull Request History

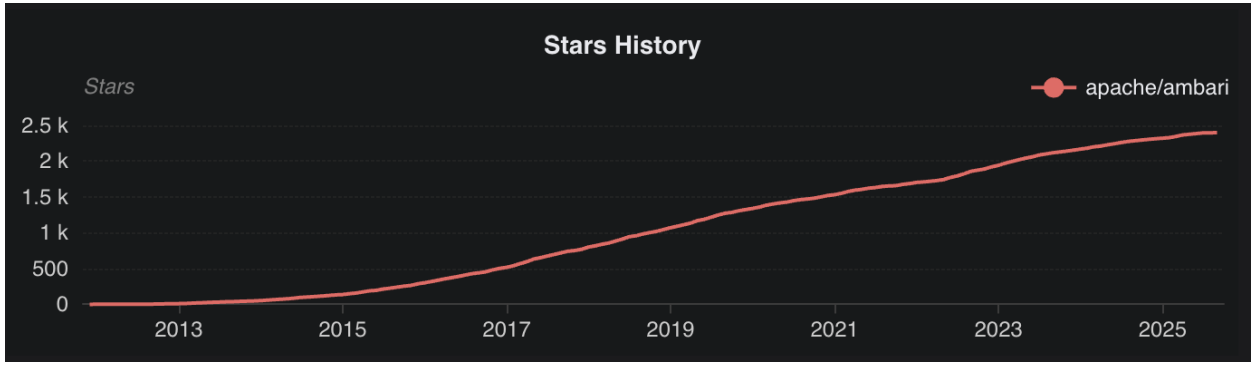
We divide the size of Pull Request into six intervals, from xs to xxl (based on the changes of code lines). Learn more about PR size.

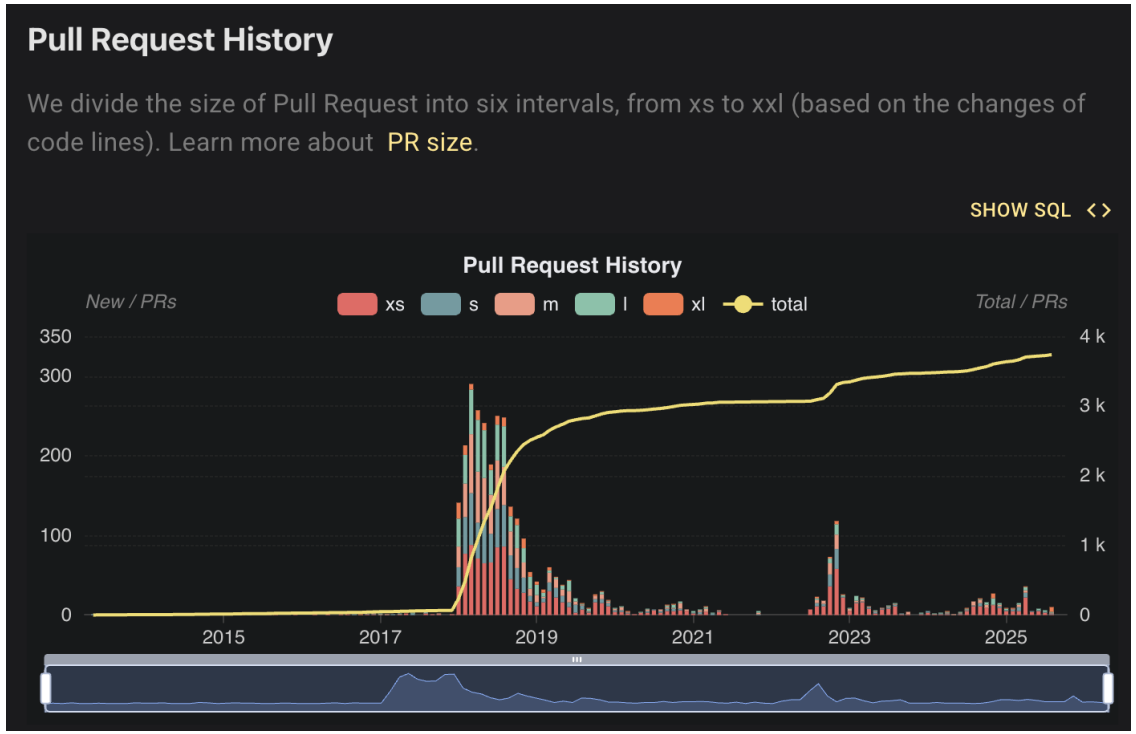




Apache Ambari 3.0.0

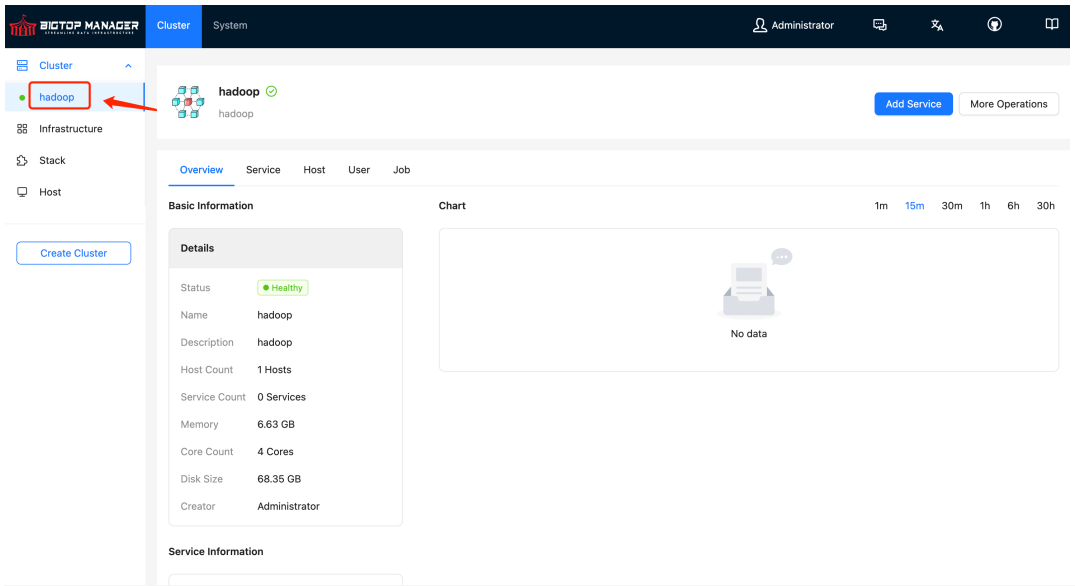
детский мир

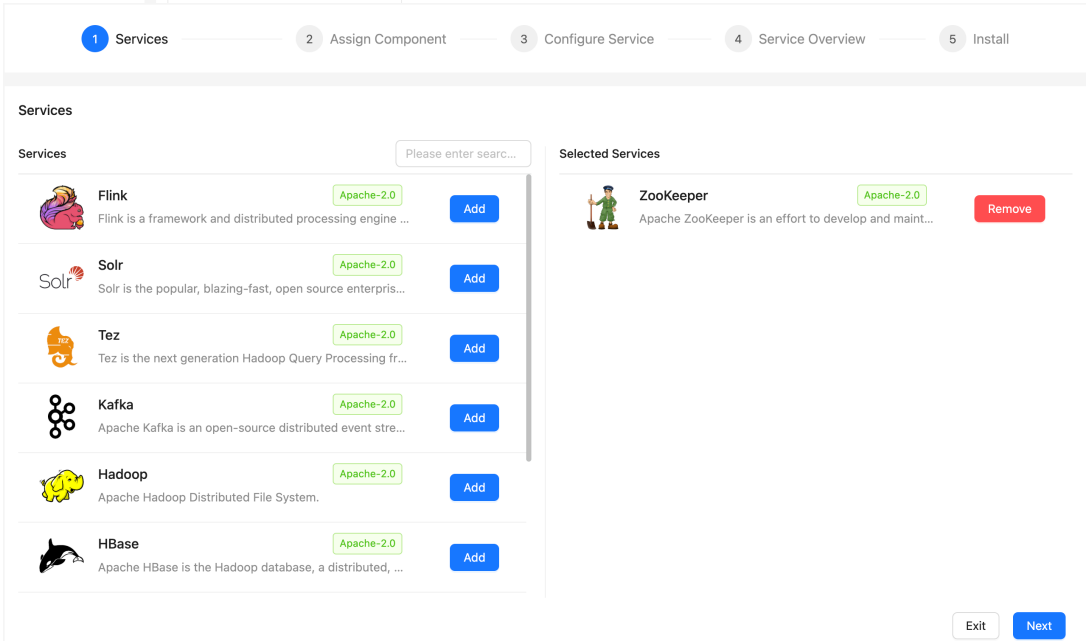


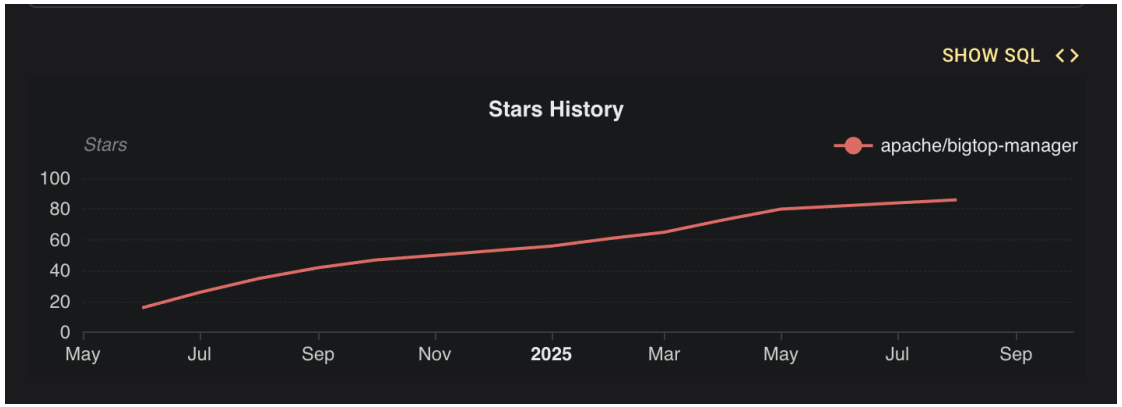


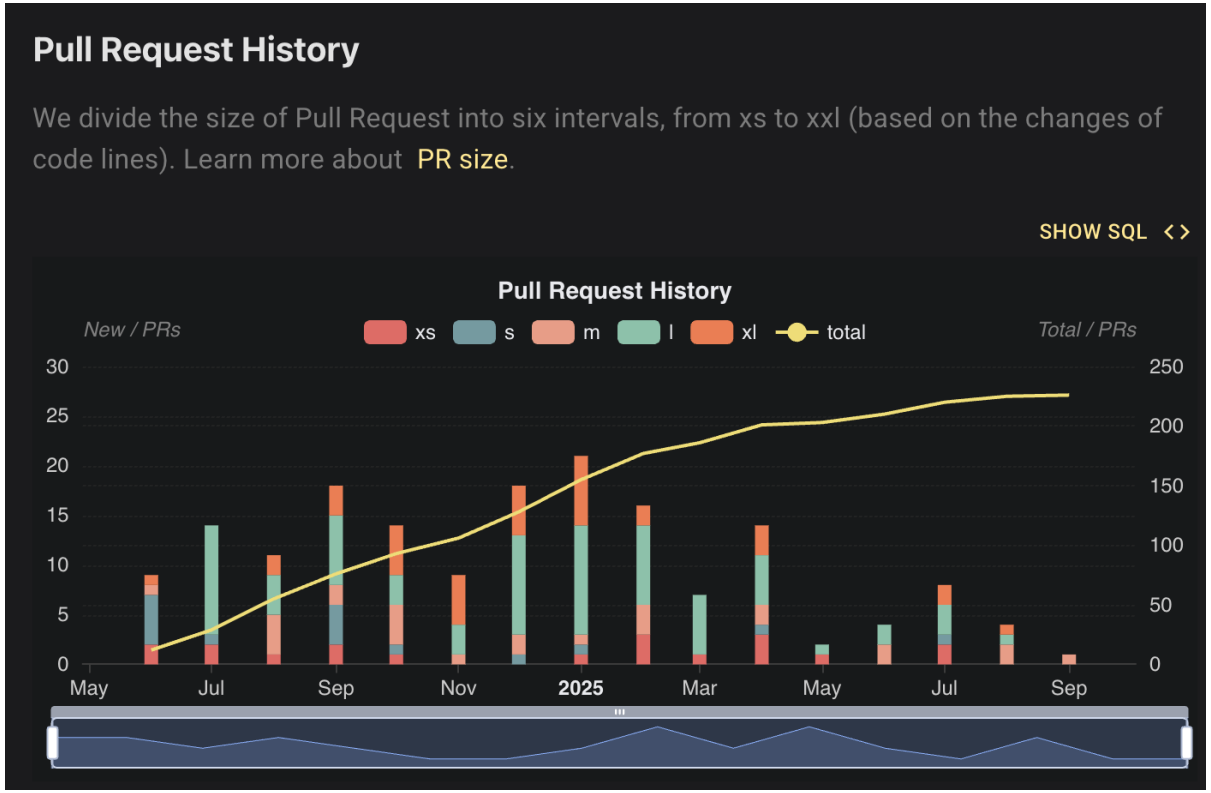
- 📦 Apache Bigtop is now our default packaging system
- 🔧 Bigtop Stack serves as the default project stack
- 🚀 This integration provides a more sustainable and community-driven approach to package management

Bigtop manager









03 Как обычно разворачивают Hadoop?

Обычная история Hadoop

детский мир

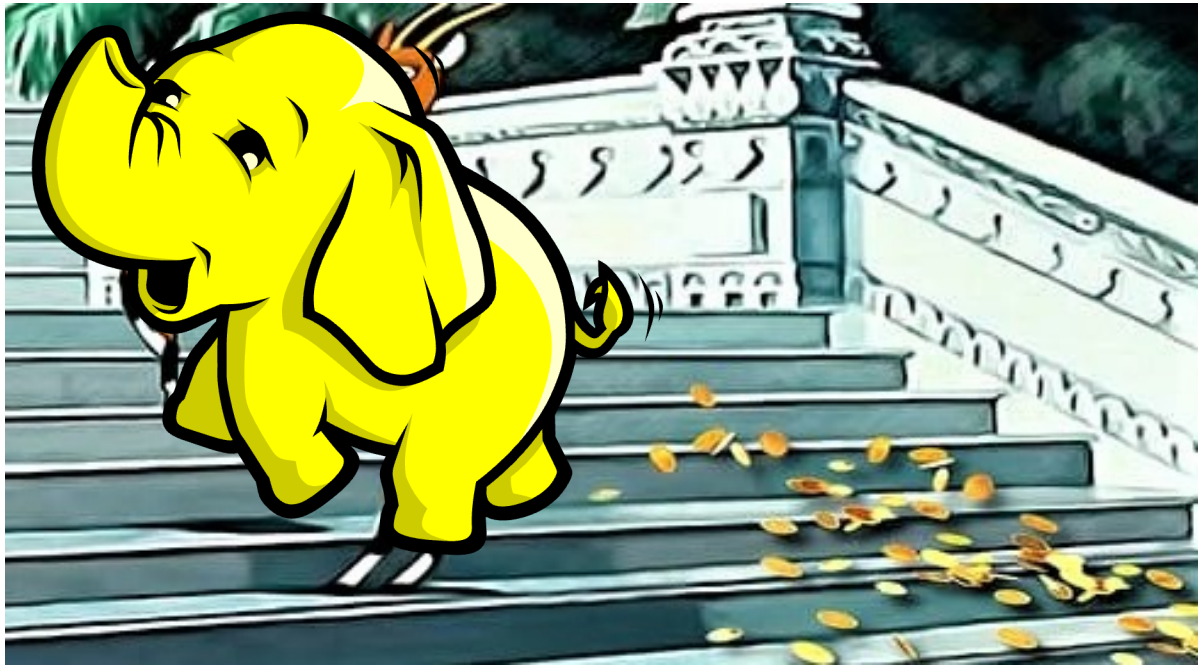
Развернули Hadoop

Запустили Spark

Загрузили данные

Посчитали витрины

Вы прекрасны



детский мир

Храним **МНОГО** данных за мало денег
Не платим за лицензии
Разгрузили другие системы
Открыли возможность для аналитических проектов

Но где то грустит отдел безопасности

- Доступ не контролируется
- Все данные в одном месте
- Любой может все скачать
- Любой может все удалить

С 2025 года стало особенно важно

Оборотные штрафы

Утечка данных	Количество пострадавших	Штраф
Первая	1 000 – 10 000	От 3 до 5 миллионов рублей
Первая	10 000 – 100 000	От 5 до 10 миллионов рублей
Первая	От 100 000	До 15 миллионов рублей
Повторная	–	До 3% годовой выручки, от 15 до 500 миллионов рублей

А потом грустят дата инженеры

детский мир

- Библиотеки начинают отставать по версиям
- Работают с Legacy
- Недоступны новые технологии
- Невозможность обновления и улучшения ситуации

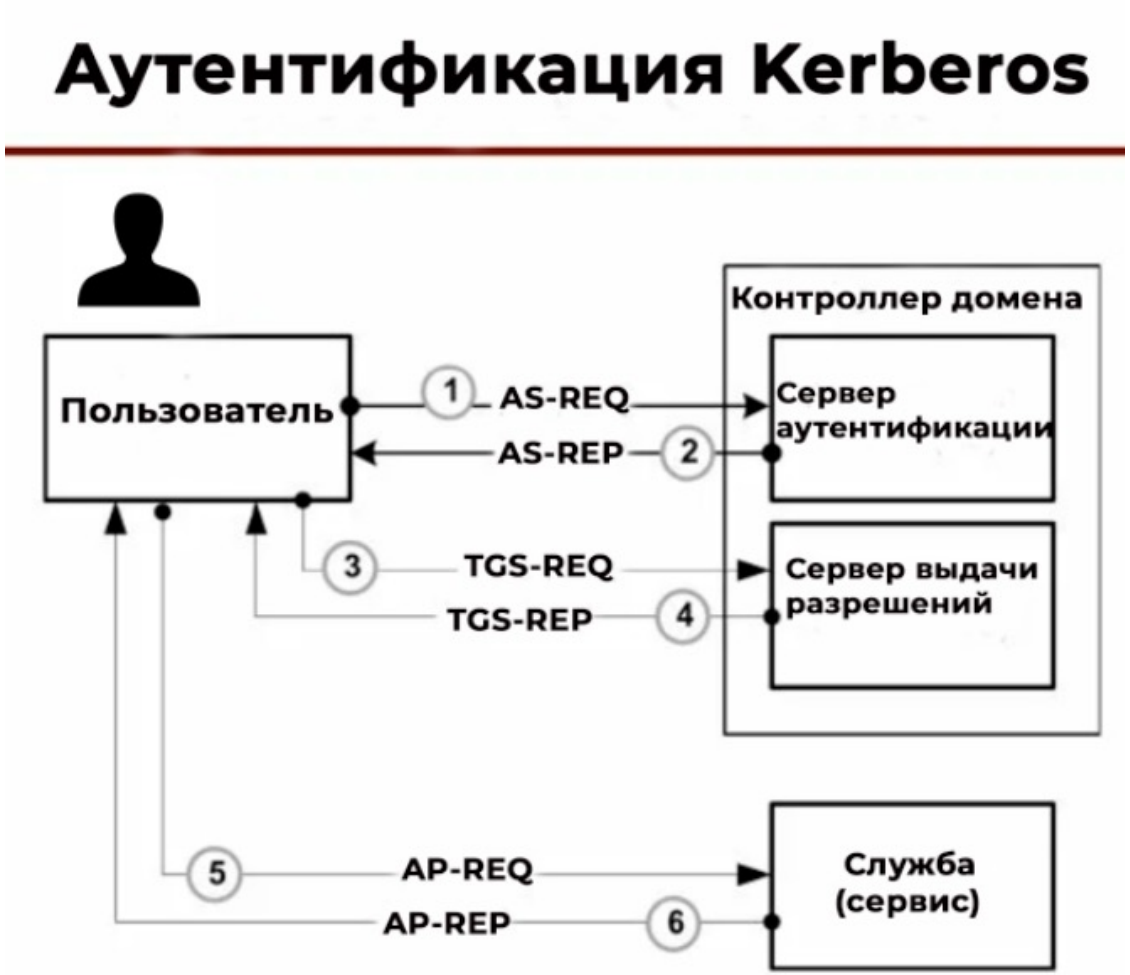
04 Kerberos

Для чего нужен и как устроен

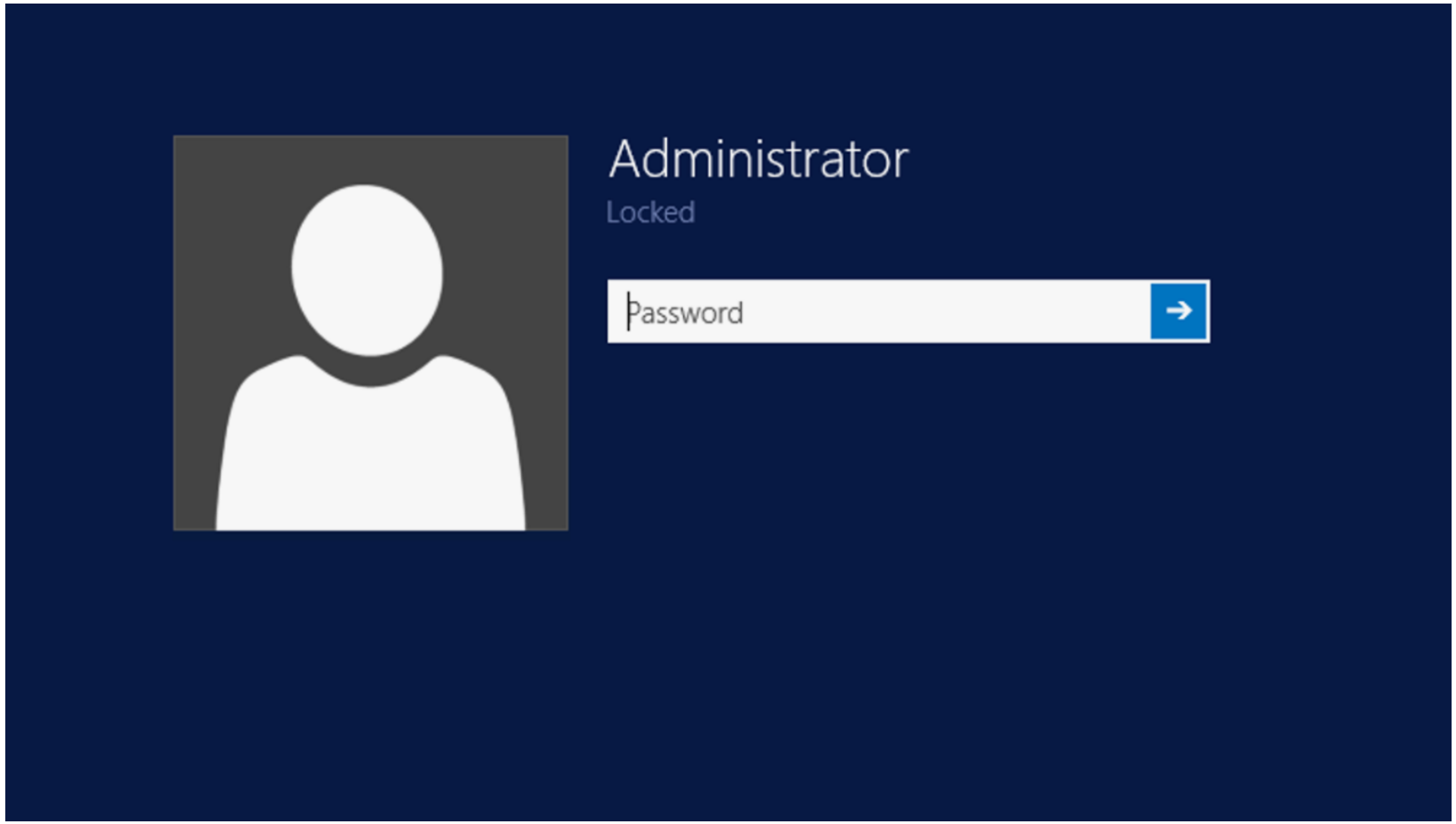
детский мир

- Единая авторизация
- Чтобы сервисы взаимодействовали между собой

- После получения тикетов, сервисы могут взаимодействовать между собой пока не протухнут тикеты



А пользовались ли вы им когда нибудь?

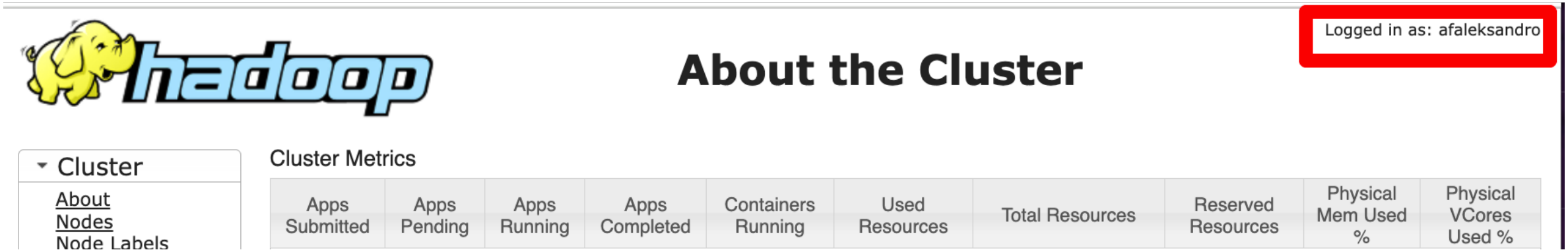


Как это выглядит

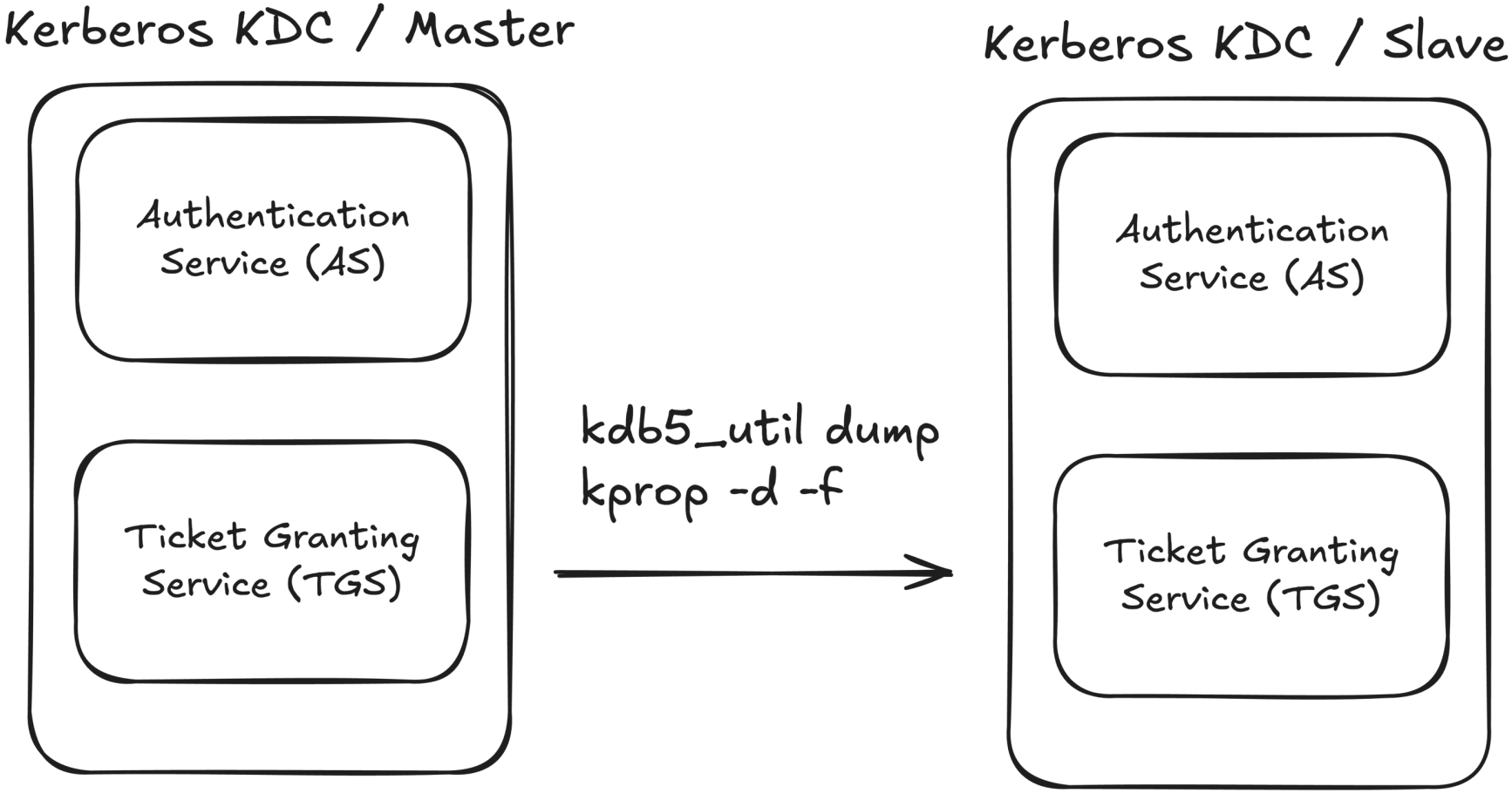
детский мир

```
base ~  (4.741s)
kinit afaleksandro
Password for afaleksandro@BIGDATA-HADOOP:

base ~  (0.388s)
klist
Ticket cache: KCM:501
Default principal: afaleksandro@BIGDATA-HADOOP

Valid starting     Expires            Service principal
09/14/25 11:38:01  09/14/25 21:38:01  krbtgt/BIGDATA-HADOOP@BIGDATA-HADOOP
        renew until 09/15/25 11:37:56
```

```
base ~  (0.08s)
kinit -kt afaleksandro.keytab afaleksandro@BIGDATA-HADOOP

base ~  (0.04s)
klist
Ticket cache: KCM:501
Default principal: afaleksandro@BIGDATA-HADOOP

Valid starting     Expires            Service principal
09/14/25 11:38:44  09/14/25 21:38:44  krbtgt/BIGDATA-HADOOP@BIGDATA-HADOOP
        renew until 09/15/25 11:38:44
```



Как обычно разворачивается



05 Как было в дм

Пул Big Data проектов

детский мир

Snowplow

OMNI

Customer Value Management (CVM)

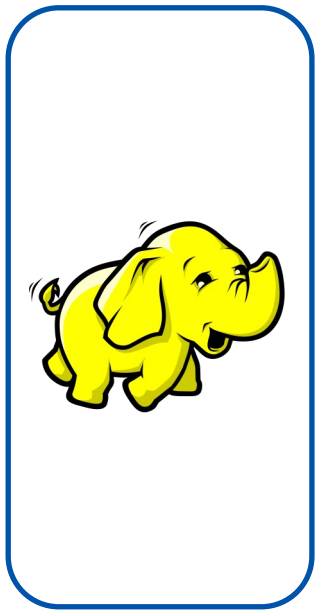
Omni Dashboard

Источники данных

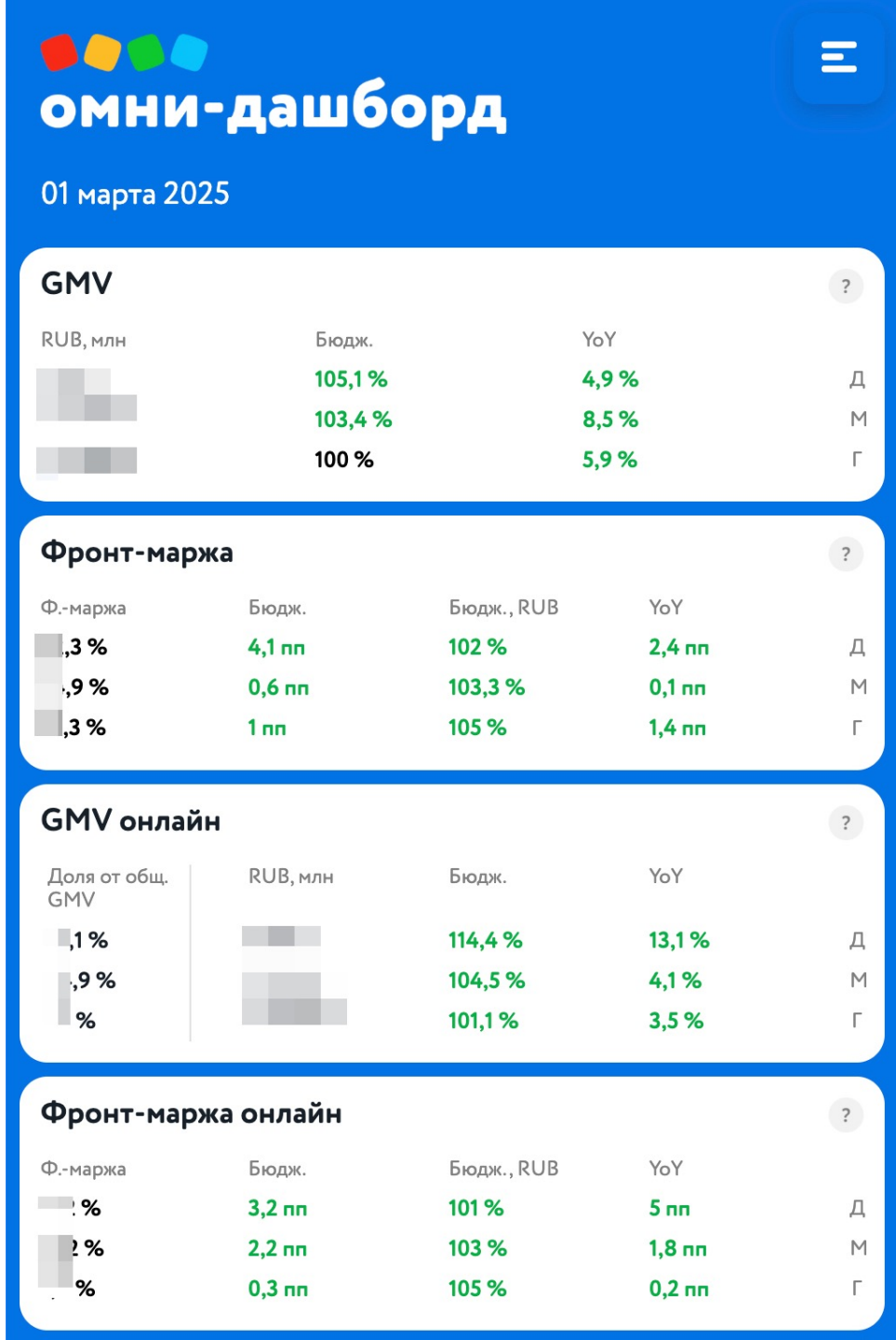
Загрузка

Хранение

Манипуляции с данными



Оркестрация



Snowplow / Мы ушли от Google Аналитики

Сбор метрик со всех клиентов ДМ (All web, Android, Apple)

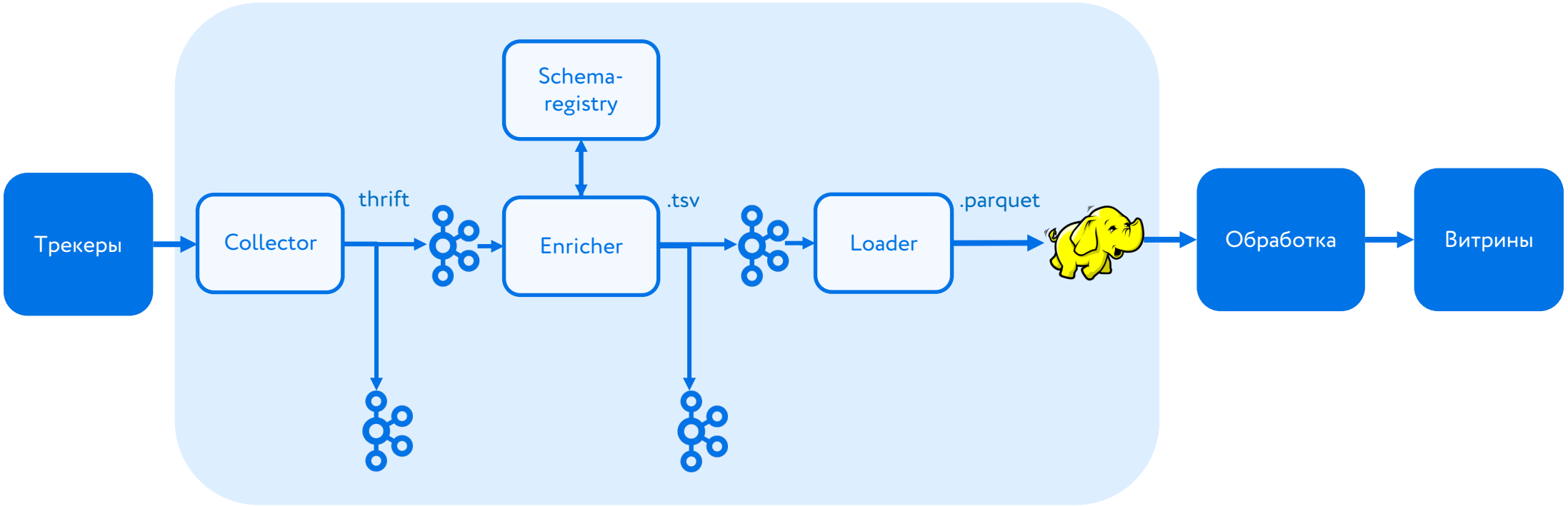
Open Source

Популярен среди аналитиков

Лучшие инженерные практики уже в продукте

- ~Collector load 10к rps
- Количество событий непрерывно растет.

  Сейчас это от 800 млн – до 1.1 млрд



MVP stage

детский мир

1

В качестве MVP

ArenaData Hadoop CE

2

Страшно что то делать на кластере

DevOps не знаю что в под капотом, не могут давать гарантии на работы и просто не берутся за них

3

Попытались прикрутить Ranger

Но для кластера без Kerberos, пользователи могли обманывать с HADOOP_USER_NAME

4

CentOs

EOL , знаем что будут DS и нужна свежая Ubuntu

Взаимодействие CVM с системами ДМ

детский мир

Инструментарий:

- IT-системы, базы данных, математические модели,
- Онлайн и оффлайн каналы для применения промокодов

Обогащение списка эффективными кампаниями

Выбор сегментов

Обучение ML-моделей

CVM кампейнинг

Запуск маркетинговых кампаний

Сервисы ПРОМО и нотификаций

Промокоды

Список клиентов

Сервисы рассылок

QR-промокод

EMAIL, промокод

PUSH, промокод

Кассы магазинов

Сайт detmir.ru + Личный кабинет клиента

Мобильное приложение «ДМ»

Атмосфера накаляется

детский мир

Большинство знало про силу HADOOP_USER_NAME

Аварии из за смены прав на папки

Пришли DS начали запускать Python

На проектах стало много подрядчиков

06 Построим все с нуля?

Собираем все вместе

детский мир

Управление доступом

Контроль над системой

Включение фичей

Высокая доступность

DE большой опыт без Kerberos

DevOps Уже разворачивали Kerberos Hadoop

Готова попробовать

Если сейчас то никогда

Надо усиливать DevOps ставкой

Общаемся с руководством

- Роль BigData растет в компании
- Презентуем текущие проблемы
- Рассказываем про инвестиции временем
- Презентуем решение и получаем одобрение
- Делаем без влияние на бизнес задачи

Big Data Platform / Карта платформы

детский мир

User Workspace

jupyter

HUE

DataHub

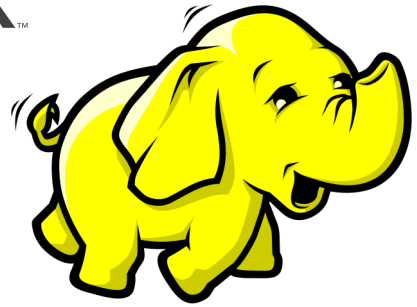
Транспорт

kafka

MINIO

Хранение / Обработка

APACHE Spark



Ad-hoc

Визуализация

Superset

Grafana

APACHE nifi

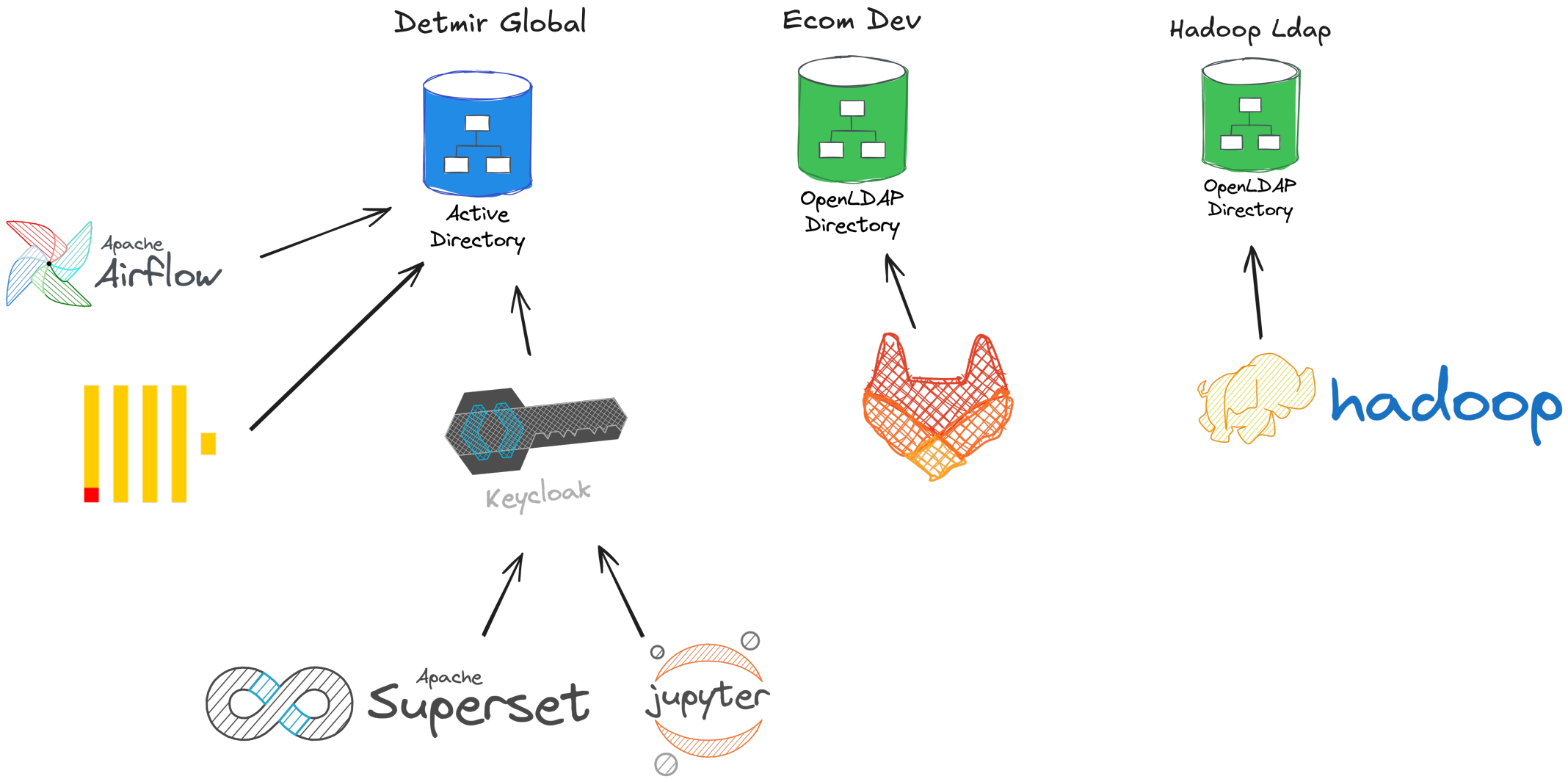
Apache Airflow

dbt Core

VICTORIA METRICS

Оркестрация / Мониторинг

Откуда брать пользователей?

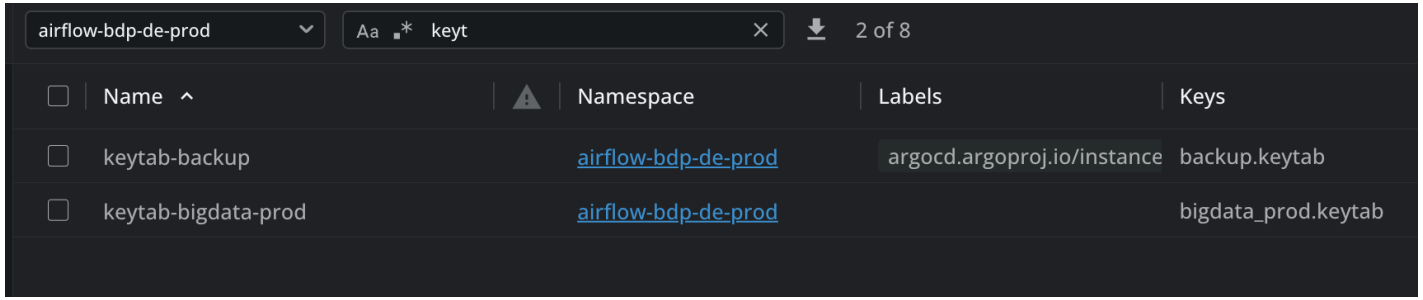


Где хранить секреты?

Можно запустить с тикетом Spark
Например в Gitlab CI CD

```
script:
 - echo $HADOOP_BDP_USER_CI_PASSWORD | kinit $HADOOP_BDP_USER_CI_PRINCIPAL
 - |
   spark-submit --master yarn \
   --deploy-mode cluster \
   --name "dbt_omni / $GITLAB_USER_LOGIN / $CI_COMMIT_BRANCH" \
```

Для Kubernetes кейтабы храним в секретах



Как поменялась оркестрация

детский мир

Apache Airflow

Основная сложность

Kubernetes executor каждая задача в изолированном контейнере

В каждом контейнере надо сделать kinit

Пришлось переписывать все взаимодействия с hdfs

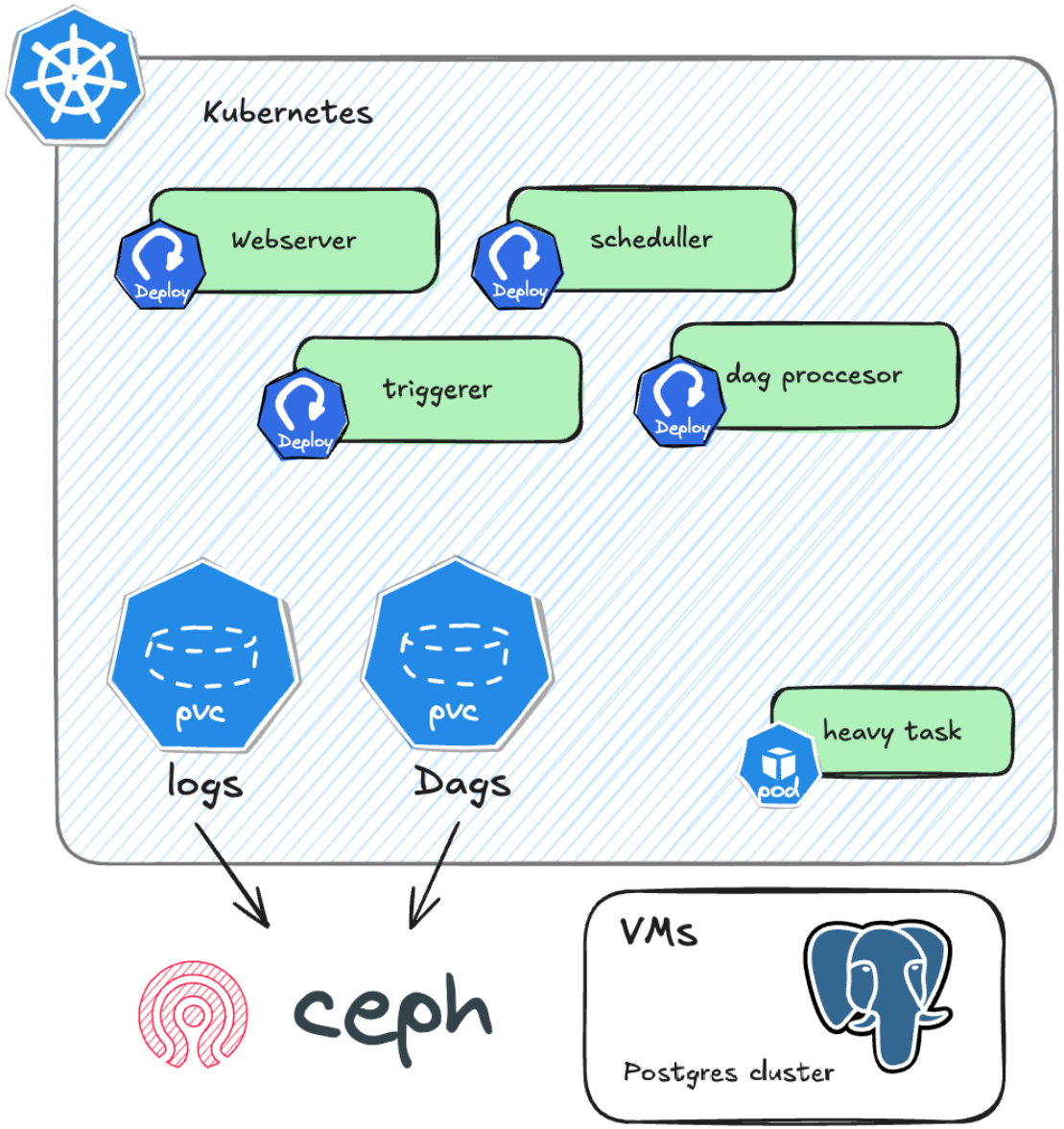
Чтобы перед действием был kinit

Для spark проблем нету

spark.kerberos.keytab=/etc/security/keytabs/bigdata_prod.keytab

spark.kerberos.principal=bigdata_prod@BIGDATA-HADOOP

spark.kerberos.krb5.path=/etc/krb5.conf



Остальные инструменты

детский мир

Нужно пользователю разрешить представляться другими (proxy-users) в hdfs-site.xml

DataHub

Раньше мета собиралась датахабом, пришлось перейти на сбор через отдельный процесс где выполняется кинит и есть доступ к hive

Все из коробки работает хорошо

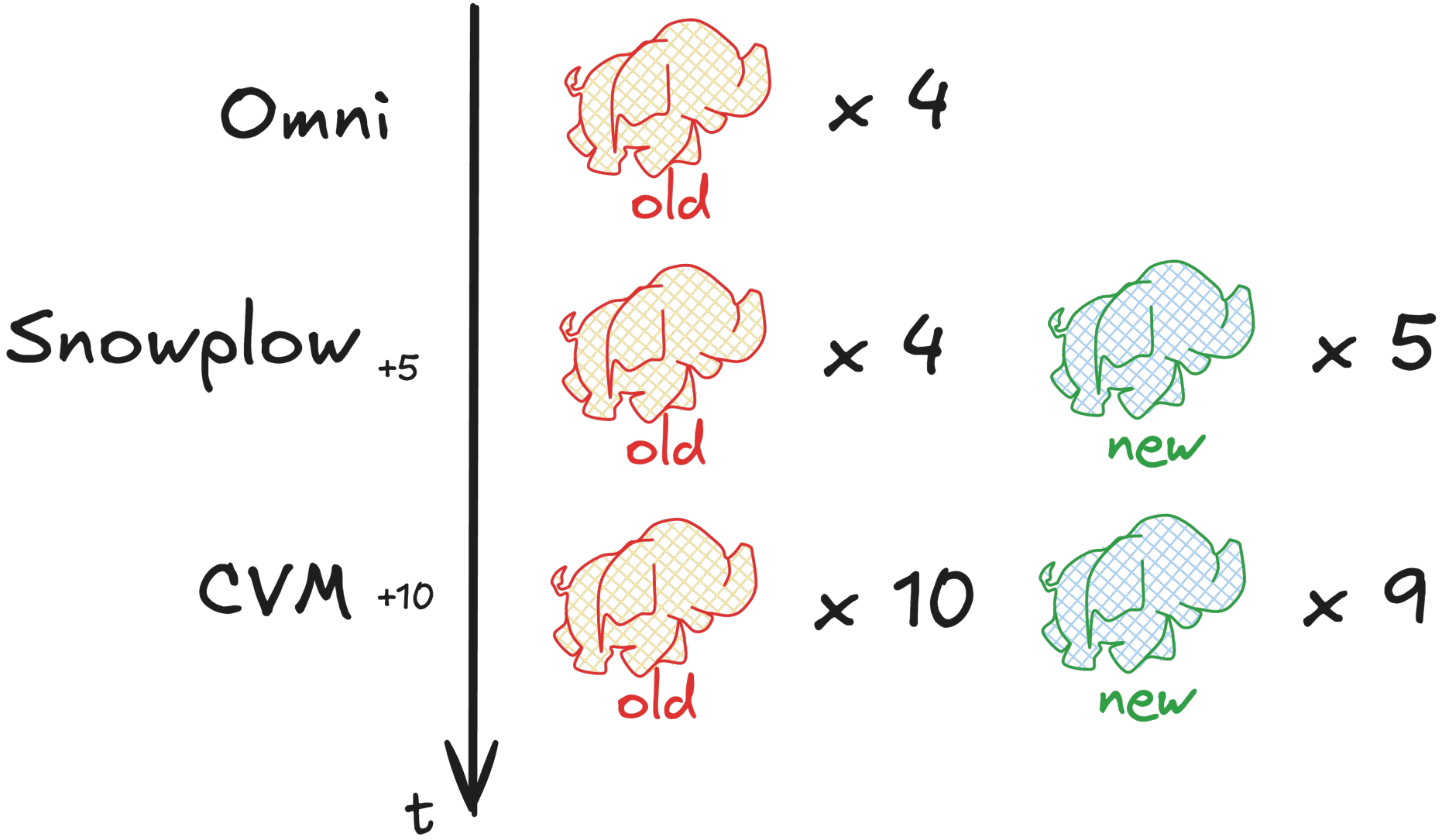
jupyter

Долго думали как пользователям упросить жизнь, поняли что kinit сделать 1 раз в день не проблема

APACHE nifi

Из коробки хорошая интеграция

А где будет Kerberos Hadoop?



Переосмысление ci cd для dbt Core

детский мир

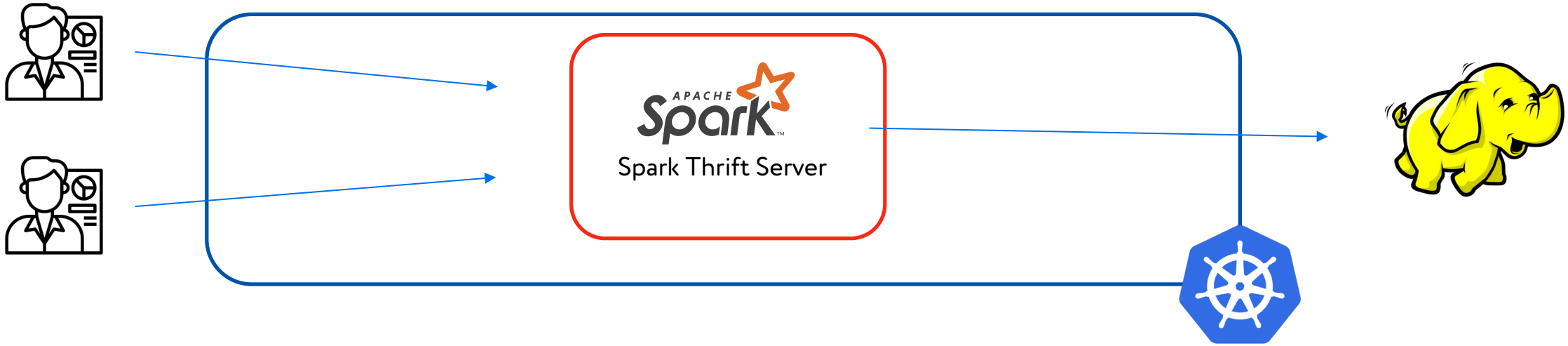
```
test run:
  stage: test
  image: dbt:rel_1.8.4_u1
  variables:
    DBT_PROFILES_DIR: "${CI_PROJECT_DIR}/omni_1"
  script:
    - python validation/validation.py --models_directory
${CI_PROJECT_DIR}/omni_1/models
    - cd omni_1
    - dbt debug -t dev_view --project-dir ../ci_cd
    - dbt deps -t dev_view --project-dir ../ci_cd
    - dbt compile -t dev_view --project-dir ../ci_cd
    - dbt -x ls -t dev_view --project-dir ../ci_cd
    - dbt run -t dev_view --project-dir ../ci_cd
    - dbt test -t dev_view --project-dir ../ci_cd
```

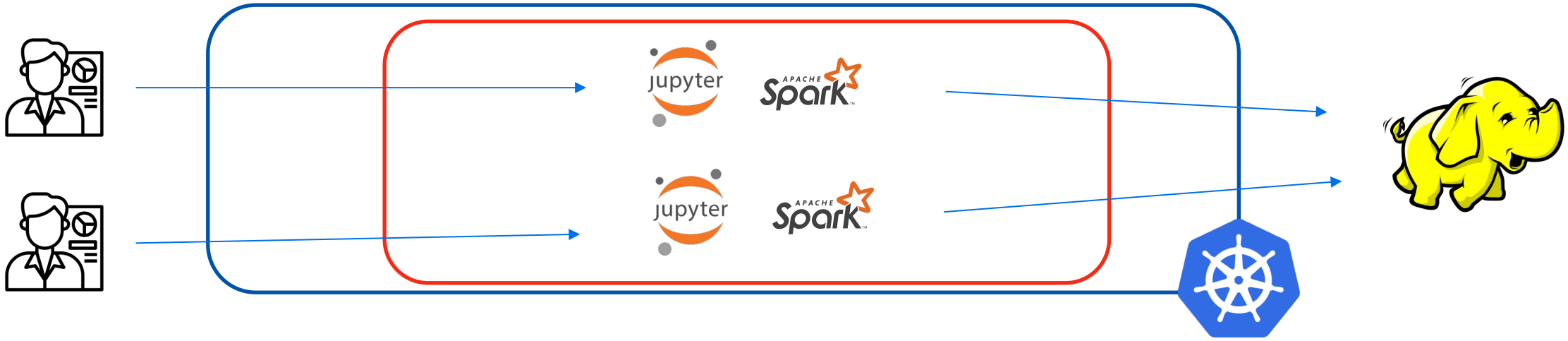

```
script:
  - echo $HADOOP_BDP_USER_CI_PASSWORD | kinit
$HADOOP_BDP_USER_CI_PRINCIPAL
  - |
    spark-submit --master yarn \
    --deploy-mode cluster \
    --name "dbt_omni / $GITLAB_USER_LOGIN / $CI_COMMIT_BRANCH" \
    --num-executors 6 \
    --conf spark.kerberos.keytab=$KEYTAB_FILE \
    --conf spark.yarn.maxAppAttempts=2 \
    --files "$MODULE_DIR/dbt_analyst.zip" \
    --py-files=sql_runner.py
```

Переосмысление для аналитиков

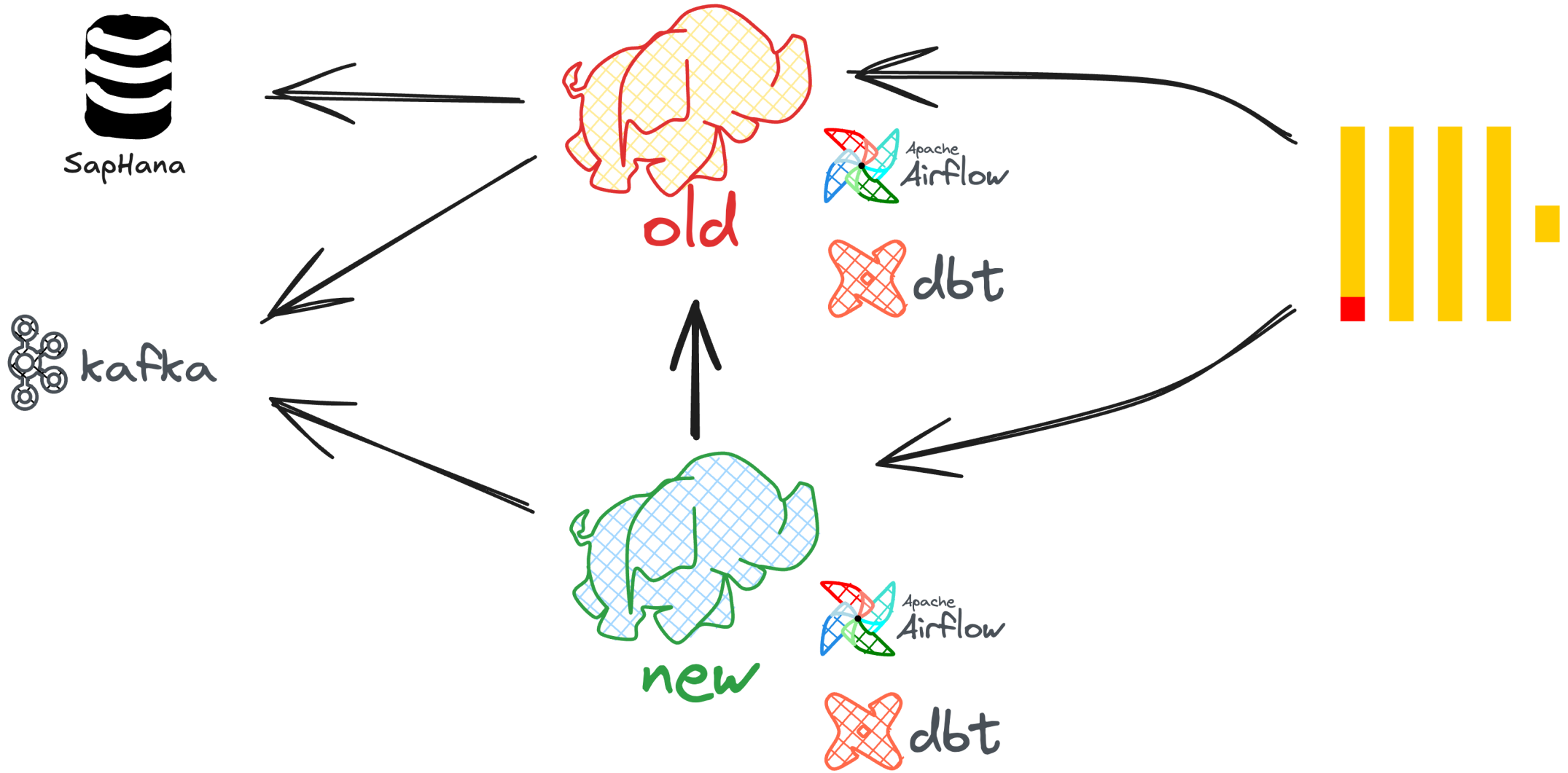
dbt Core

детский мир





Потоки на время миграции

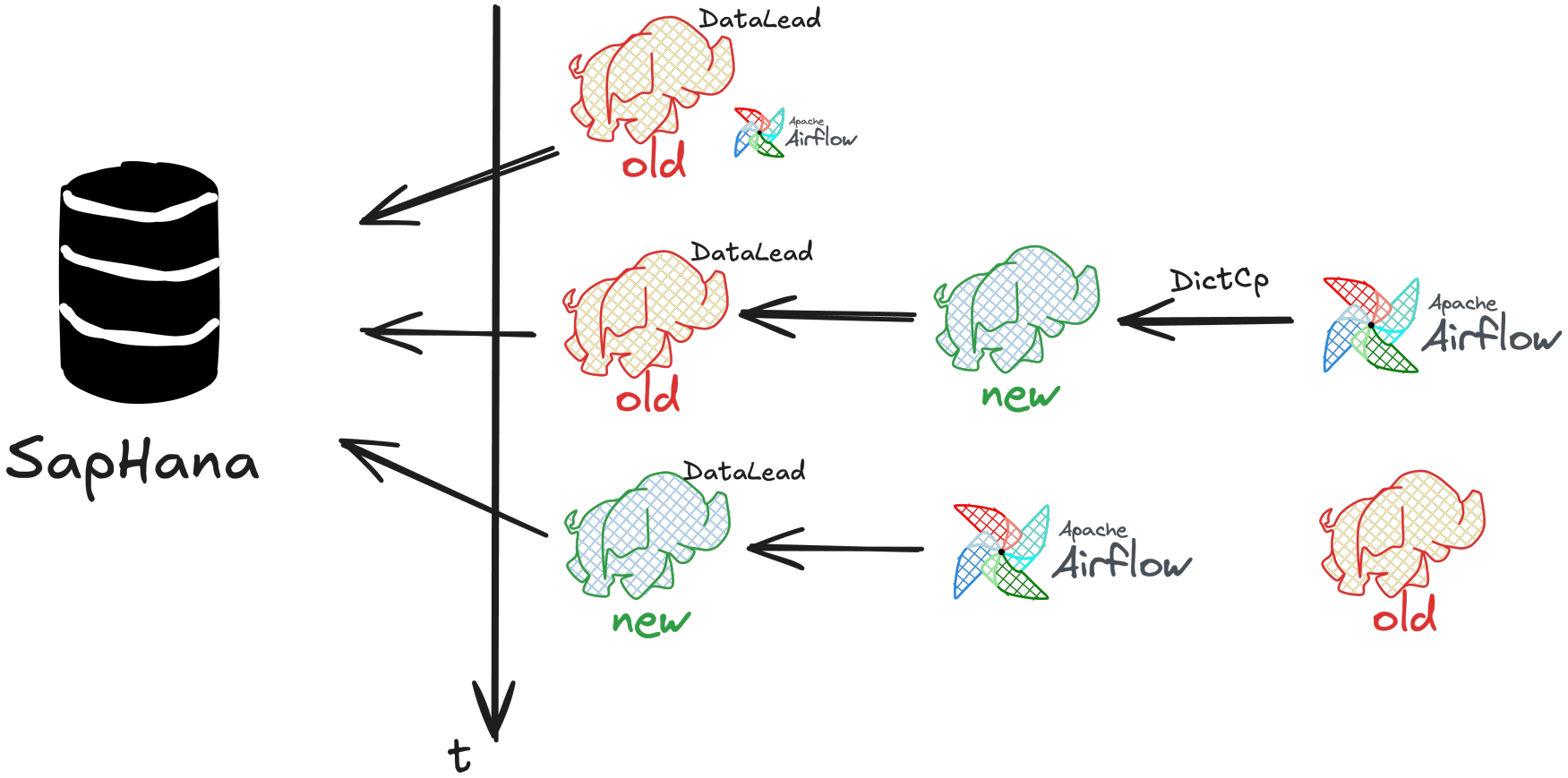


Меняем мастер кластер

15 сентября 2024 «Омни отчет» пришел с Нового кластера

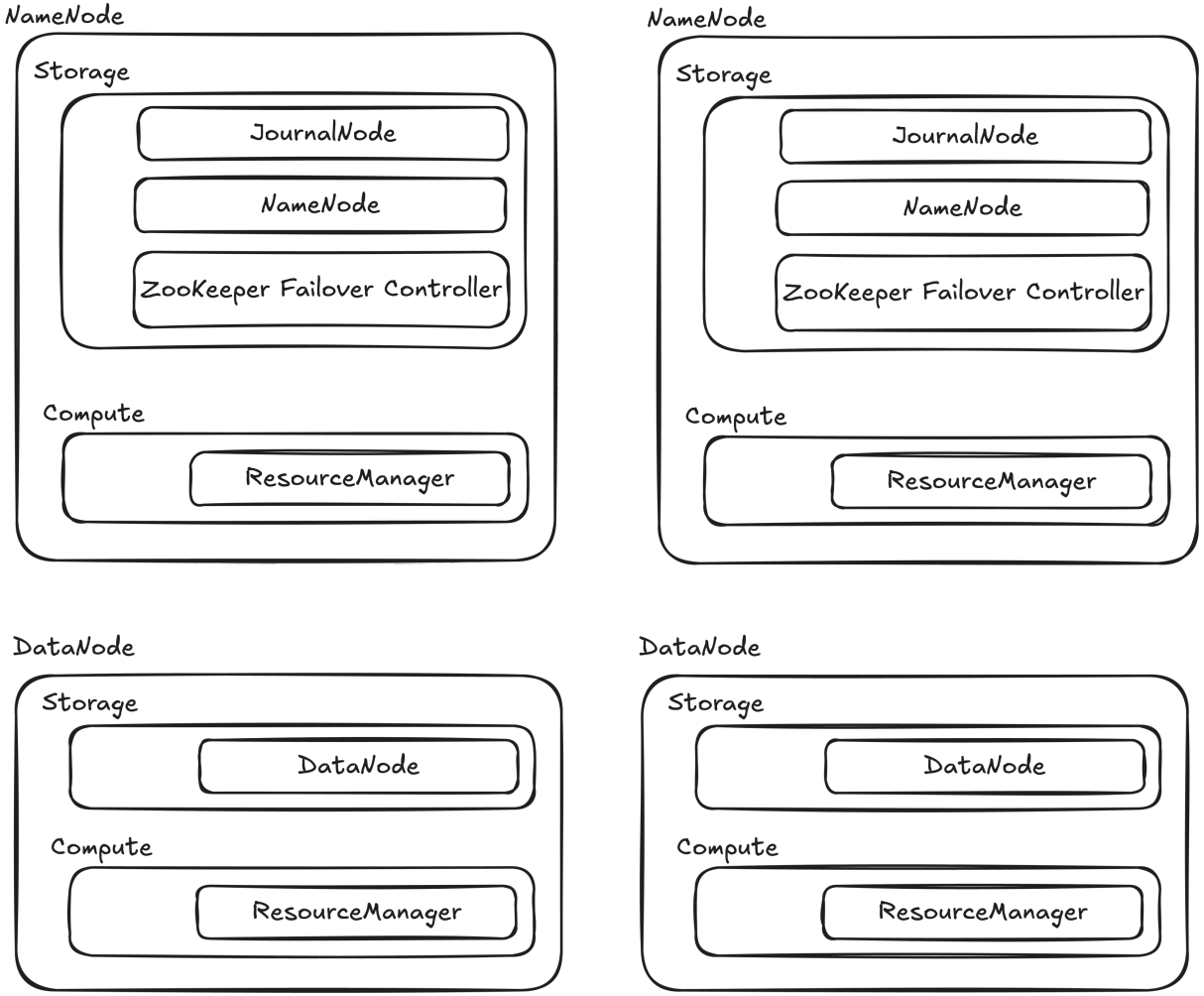
оставили 3 ноды и объявили что есть месяц чтобы забрать свои данные

5 декабря 2024 полностью выключили старый Hadoop



07 Идеальный шторм

Простой на рабочий день 05.11.2024



05.11.2024 12:00 рестарт namonode для применения конфигов (изменяли настройки ldap). Активная неймнода не может достучаться до Journal Node, чтобы стянуть свежие изменения. В логах ошибки про авторизацию.

05.11.2024 14:30 - созвон, решаем попробовать переключить активную ноду вручную. Это не работает так как у нас включено autofailover через zkfc. Оказывается, что 2 из 3 zkfc демонов лежали уже несколько месяцев. Так же в zkfc сломан fencing, он не может по ssh сходить и кильнуть процесс.

Простой на рабочий день 05.11.2024

05.11.2024 16:00 В итоге мы понимаем, что у нас есть лишь одна journal node, на которой доступны последние журналы транзакций.

Выпускаем заново keytab для spnego только для namenodes.

Перезапускаем journal nodes и они могут уже с тикетом друг в друга сходить, однако теперь другая ошибка, что нет нужных файлов.

Руками перекладываем файлы.

После того как все journal nodes ссинкались - уже можно сделать failover вручную, hdfs начинает работать.

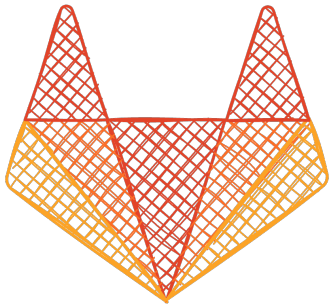
```
afaleksandro@bdp-test-nn-01:/etc/hadoop/keytabs$ sudo klist -kt jn.keytab
Keytab name: FILE:jn.keytab
KVNO Timestamp         Principal
---- ----------------- --------------------------------------------------------
   3 11/13/24 16:01:10 jn/bdp-test-nn-01.detmir-infra.ru@BIGDATA-HADOOP
   3 11/13/24 16:01:10 jn/bdp-test-nn-01.detmir-infra.ru@BIGDATA-HADOOP
   3 11/13/24 16:01:10 host/bdp-test-nn-01.detmir-infra.ru@BIGDATA-HADOOP
   3 11/13/24 16:01:10 host/bdp-test-nn-01.detmir-infra.ru@BIGDATA-HADOOP
```

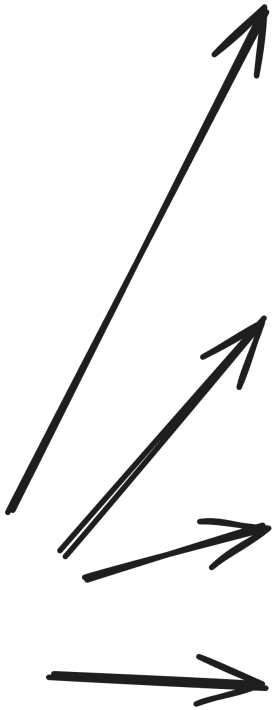
```
KVNO Timestamp
---- -----------------
   3 11/13/24 16:01:10
   3 11/13/24 16:01:10
   3 11/13/24 16:01:10
   3 11/13/24 16:01:10
```

08 Мы вообще ВСЁ поняли

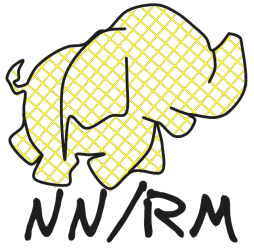
Поднимаем dev стенд

- Поняли что экономили на спичках
- Спас очень много раз потом
- В hadoop очень много настроек, конфликтуют или не дают не запускать сервис



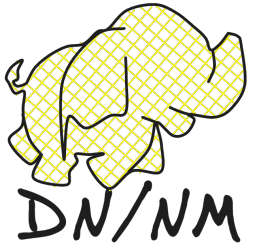


2vcpu
16ram
50gb fs



x 3

8vcpu
48ram
150gb fs

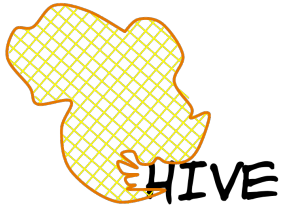


x 3

2vcpu
6ram
20gb fs

x 1

2vcpu
16ram
30gb fs



x 1

09 Но была еще 1 проблема

Запуск python job

```
spark-submit --master yarn \
--deploy-mode cluster \
--archives hdfs:///distribution/pyspark/dbt.zip#env \
--conf spark.yarn.appMasterEnv.PYSPARK_PYTHON=./env/bin/python \
--conf spark.yarn.appMasterEnv.PYSPARK_DRIVER_PYTHON=./env/bin/python \
"$MODULE_DIR/ci_cd/sql_runner.py"
```

```
DEBUG [ContainerLocalizer Downloader] org.apache.hadoop.security.UserGroupInformation:
PrivilegedActionException as: ci (auth:SIMPLE)
org.apache.hadoop.yarn.exceptions.YarnException: Download and unpack failed
```

И в один прекрасный день

Самое просто окружение где установлено только dbt
Во время расчетов и когда диски под нагрузкой
Распаковывается 30 минут
А это надо на всех нодах

```
time unzip dbt.zip

real    27m10.029s
user    0m56.287s
sys     0m26.047s

find . |wc -l
105986
```

```
spark-submit --master yarn \
--deploy-mode cluster \
--conf spark.pyspark.driver.python=/opt/conda-envs/dbt-v1.7.18-u1/bin/python3 \
--conf spark.pyspark.python=/opt/conda-envs/dbt-v1.7.18-u1/bin/python3 \
--py-files=sql_runner.py \
"$MODULE_DIR/ci_cd/sql_runner.py"
```

А что еще веселого у нас было?

Кластер уходил в консервацию по нехватке места

Кластер переезжал в другой ДЦ

Ломались диски на которых ОС

Поставили на все ноды SSD

09 Подводим итоги

Что мы получили

Технологический контроль

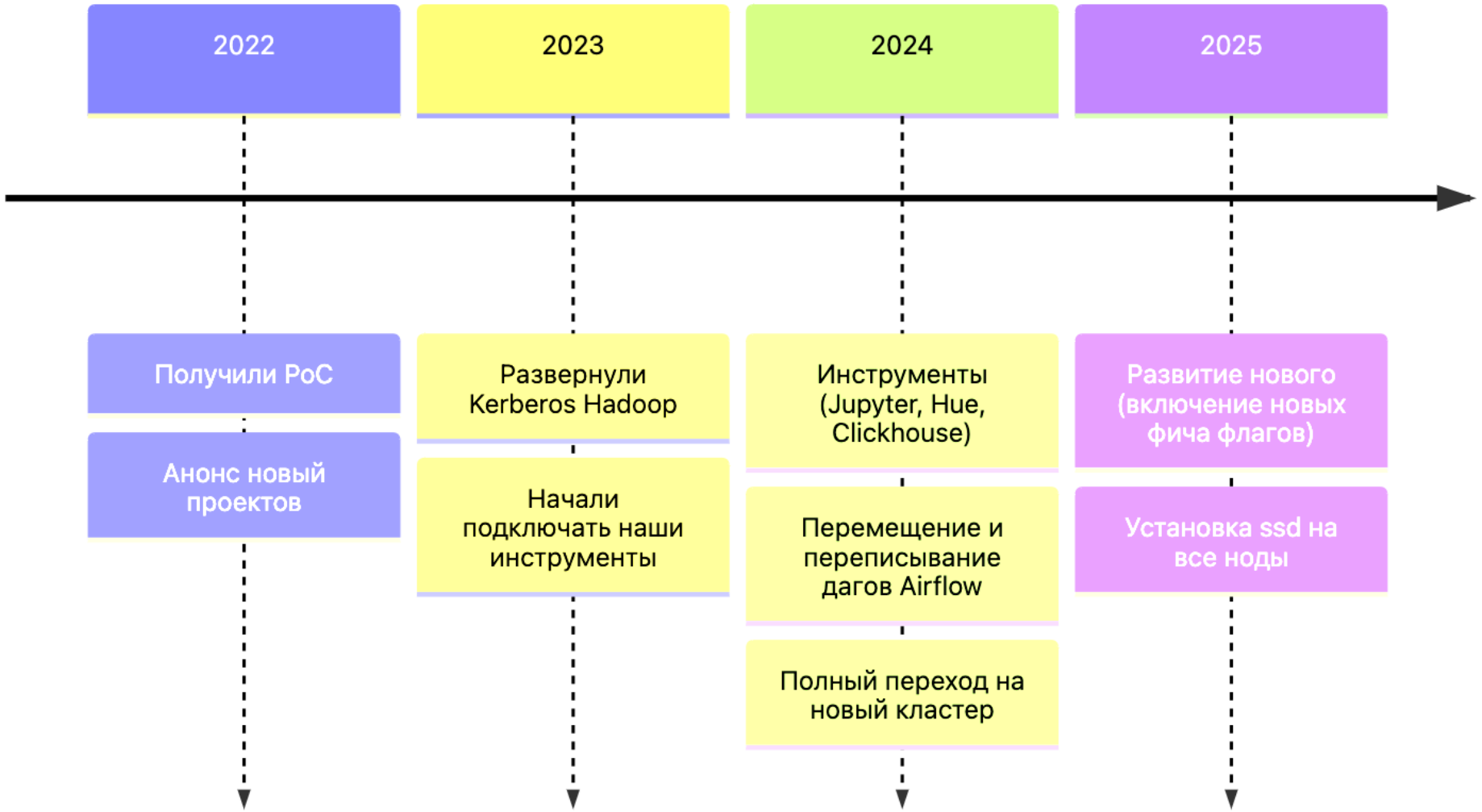
Финансовая экономия

Масштабируемость и гибкость

Безопасность

Какова цена внедрения

детский мир



Спасибо за внимание!

Tg: bioqwer