

ЭВОЛЮЦИЯ ПОКЕМОНА ТРЕЙСИНГА В ЕСОМ.ТЕСН

ЧЕТЫРЕ ФОРМЫ ОДНОЙ ПЛАТФОРМЫ: GRAFANA TEMPO,
GIGAPIPE, VICTORIATRACES И ТРАНСПОРТ С ГАРАНТИЯМИ

Валерий Евдокимов

esom.tech



Кто я

ВАЛЕРИЙ ЕВДОКИМОВ

Tech Lead, Observability platform

5 лет строю платформу метрик, логов
и трейсов. Всё, что помогает понять,
почему сервис работает так, как работает.

ecom.tech



есom.tech в цифрах

9 ЦОД

Москва и Санкт-Петербург + Облака

4 000+

железных серверов

12 000+

виртуальных серверов

240+

стоек

37 k8s

кластер k8s, воркеры bare metal 96 CPU / 756 GB RAM

~23

инсталляций платформы трейсинга

Это – Спанни,
наш проводник



ПОВТОРИМО

Opensource



С ГРАБЛЯМИ

Каждая форма показана
вместе с тем, где она
сломалась.



КАРТА РЕШЕНИЙ

В конце по ней
можно пройти самим
и повторить наш опыт.



Какие вопросы закрывает трейсинг



КАКОЙ
СЕРВИС
ТОРМОЗИТ

(01)

ГДЕ УЗКОЕ
МЕСТО

База, сеть или кривой
цикл в коде

(02)

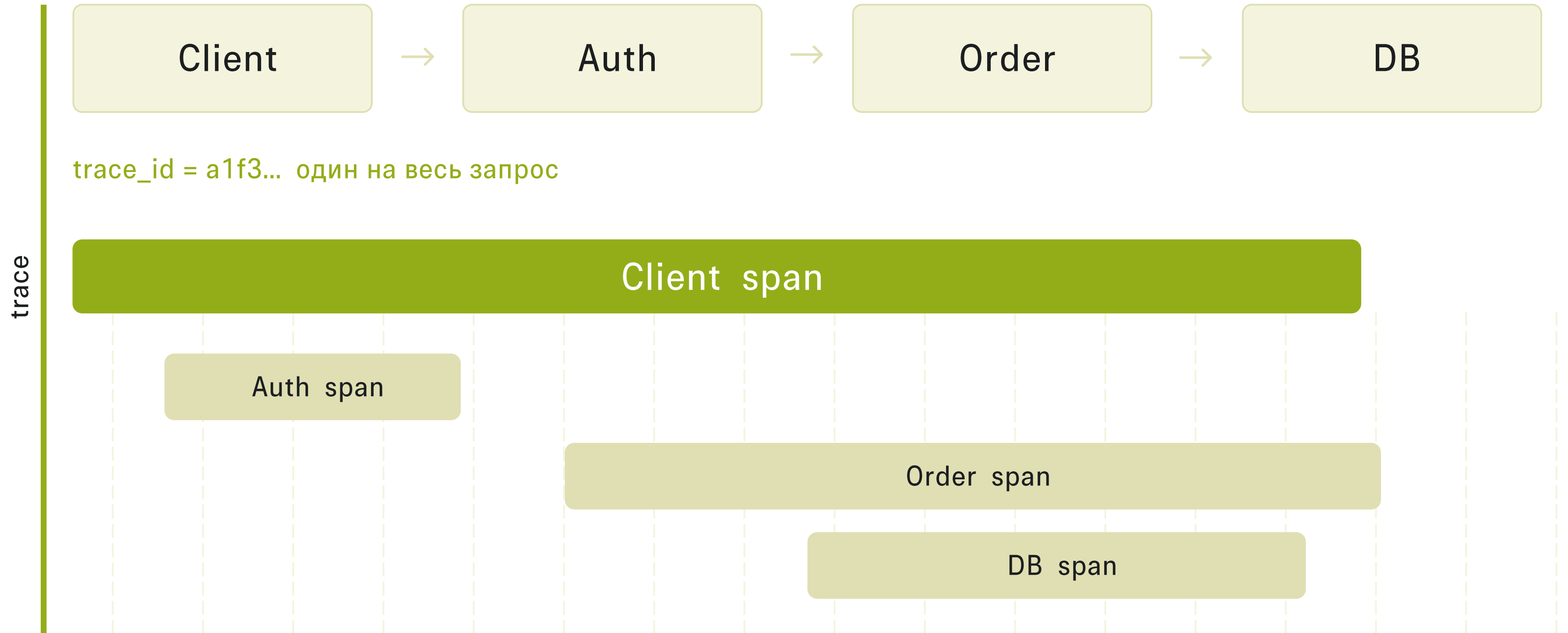
До внедрения
трейсинга



После
внедрения
трейсинга



База: Span и trace



Требования заказчика

СОБИРАТЬ И ХРАНИТЬ
ВСЕ ТРЕЙСЫ

Ни head, ни tail. Полнота
важнее стоимости.

(01)

ХРАНИТЬ
ОТ НЕДЕЛИ

Полное сохранение данных
от недели до двух недель
в зависимости от кластера.

(02)

КОД ПРОДУКТОВ
ТРОГАТЬ МИНИМАЛЬНО

У команд свои приоритеты,
нужно делать общую платформу.

(03)

...И СДЕЛАТЬ НАДО
БЫЛО ВЧЕРА



(04)

Ландшафт: формат и хранилища

OpenTelemetry: формат как основа гибкости.



ИНСТРУМЕНТ	ХРАНИЛИЩЕ	КЛАСТЕРИЗАЦИЯ	ЗРЕЛОСТЬ	МЫ
Jaeger	Elasticsearch, ClickHouse	Через бэкенд	Ветеран	
Grafana Tempo	Объектное хранилище S3	Микросервисный режим	Зрелый	Форма 1
Gigapipe (qryn)	ClickHouse	Single-node из коробки, дальше нюансы	Зрелый	Форма 2
VictoriaTraces	Своё и LSM-подобное	Vminsert / vmselect / vmstorage	Свежий, но вендор хорошо себя зарекомендовал	Форма 3

Форма 1. Grafana

Темро: почему?

СТЕК GRAFANA УЖЕ БЫЛ

Единая точка входа: метрики, логи, трейсы в одном окне

(01)

S3 ПОД РУКОЙ

MinIO в ландшафте, вендорское облако в доступе

(02)

НАСТРАИВАЕТСЯ ЗА ВЕЧЕР

Зашли догоняющим трендом, первые месяцы всё хорошо

(03)

ЧТО МЫ ХОТЕЛИ:

Один вендор, одна точка входа. Разработчик открывает Grafana и получает всю информацию



Проблема 1: резкий рост количества данных

~40 ТБ

Данных в неделю на один
ДЦ. 1 дц = 1 бакет



Вендорский S3
отдаёт таймауты



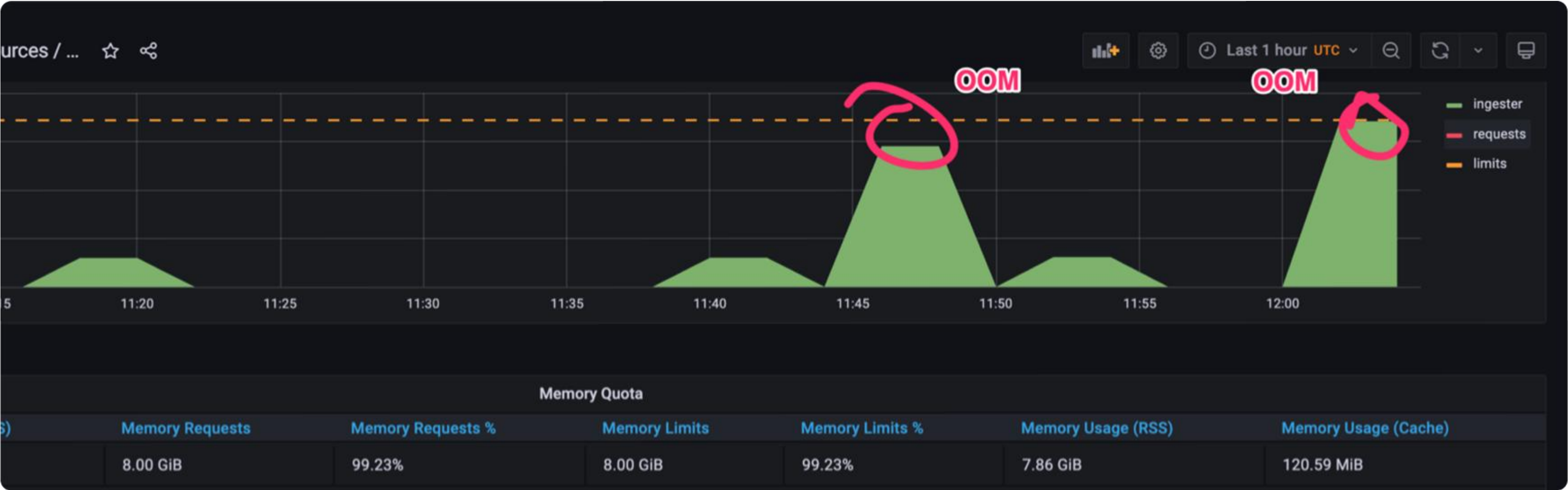
«ТРЕЙС ЗАПИСАЛИ, А ПРОЧИТАТЬ НЕЛЬЗЯ»



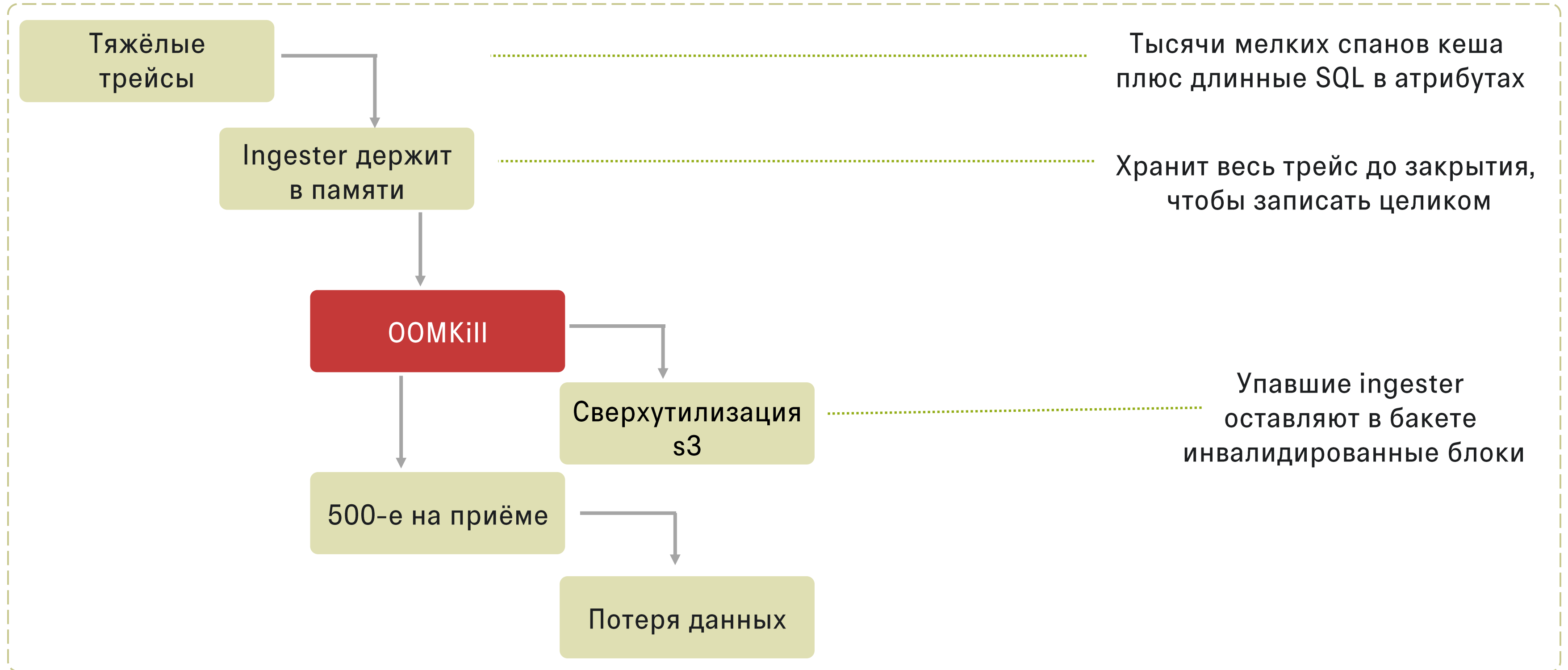
Проблема 2: слишком большие трейсы

INGESTER

лопается по памяти



Из-за чего проблемы?



Защита есть

Темпо просто
выбрасывает трейс.



`max_bytes_per_trace =`
5 МБ по умолчанию

Штатная защита Tempo: трейс
больше лимита отбрасывается
целиком. grafana / tempo PR
#612, issue #3690



```
2024-11-04 12:22:14.438 compactor level=warn ts=2024-11-04T12:22:14.383351302Z caller=compactor.go:
rootSpanName= rootServiceName= discarded_span_count=35515

{
  insertId: "hzcg13kelpgokw5j"
  labels: {12}
  logName: "projects/samaya-prod-403612/logs/stderr"
  receiveTimestamp: "2024-11-04T12:22:17.760464686Z"
  resource: {2}
  severity: "ERROR"
  textPayload:
    "level=warn ts=2024-11-04T12:22:14.383351302Z caller=compactor.go:247 msg="max size of trace exceeded" tenant
    discarded_span_count=35515"
  timestamp: "2024-11-04T12:22:14.438646943Z"
}
```

ТЕРЯЕМ
ДАННЫЕ!



Нужно что-то делать

Нет локальных выделенных S3-кластеров и качественной услуги в облаке



Разработка к трейсингу уже привыкла



Времени на перестроение мало



Форма 1



Прокачать

ФОРМА 2



ФОРМА 3



ФОРМА 4



Форма 2. Gigarire: почему?

ИНСТРУМЕНТАЦИЯ НЕ МЕНЯЕТСЯ

Приложения как отдавали
OpenTelemetry, так и отдают

(01)

ИНТЕРФЕЙС ОСТАЁТСЯ

Разработчик по-прежнему
в Grafana

(02)

CLICKHOUSE УМЕЕМ

В команде есть
компетенции и железо

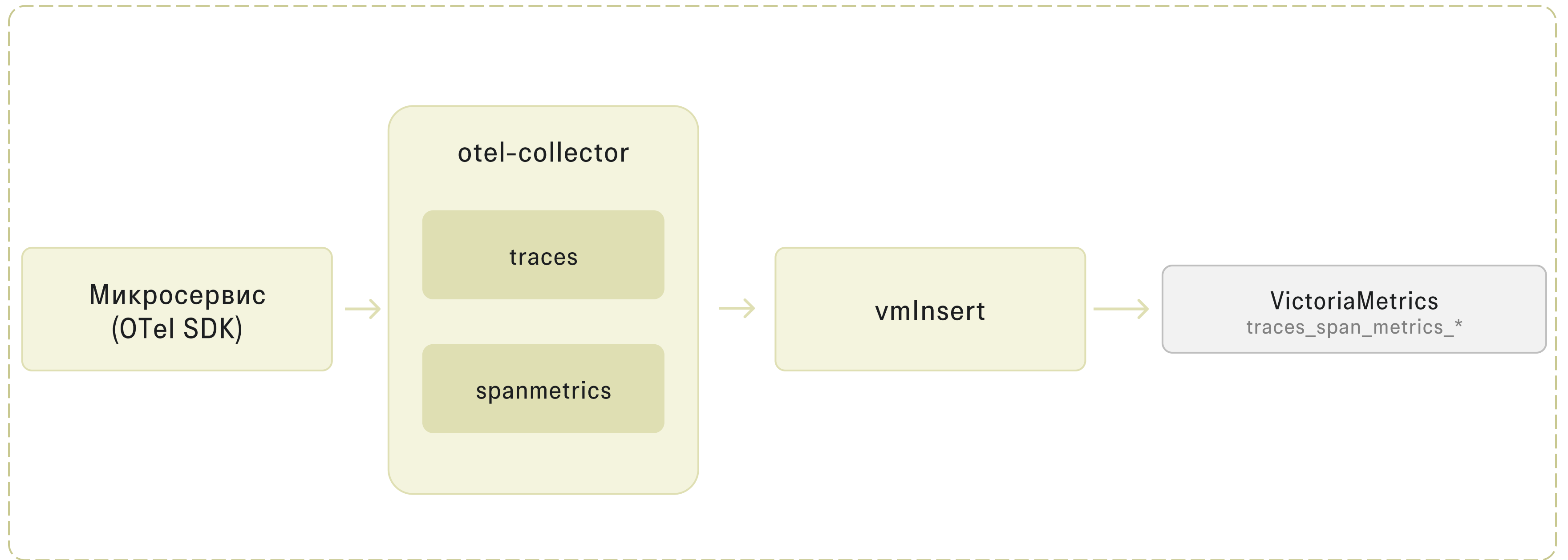
(03)

ГИПОТЕЗА 2:

распределённый ClickHouse
раскладывает трейсы лучше,
чем S3. Качество сервиса
вырастет.



Добавили: span-метрики



*Вызовы, ошибки, латенсы по сервисам и операциям.
Отправляются в VictoriaMetrics.

На базе новых метрик получили

HELICOPTER-VIEW ДАШБОРД

Как сервис работает в целом после релиза.

АЛЕРТЫ НА ДЕГРАДАЦИЮ

Видно раньше, чем кто-то пойдёт искать трейс руками.

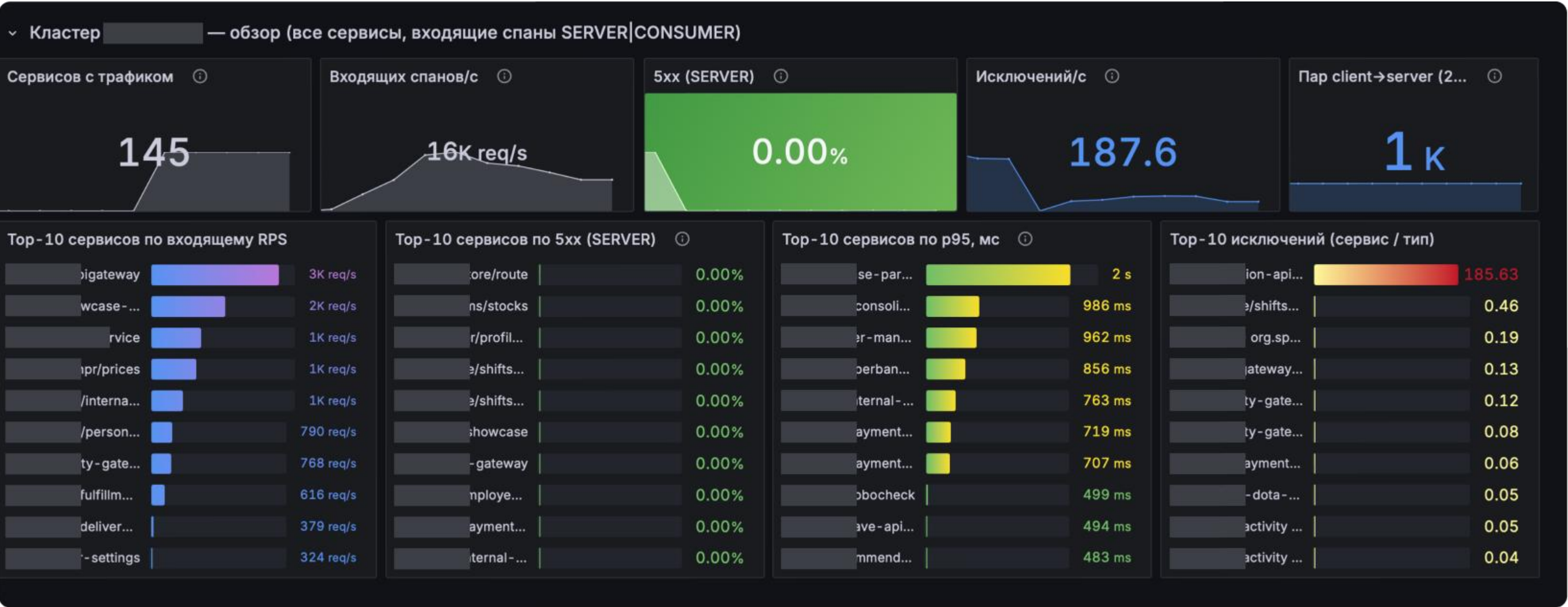
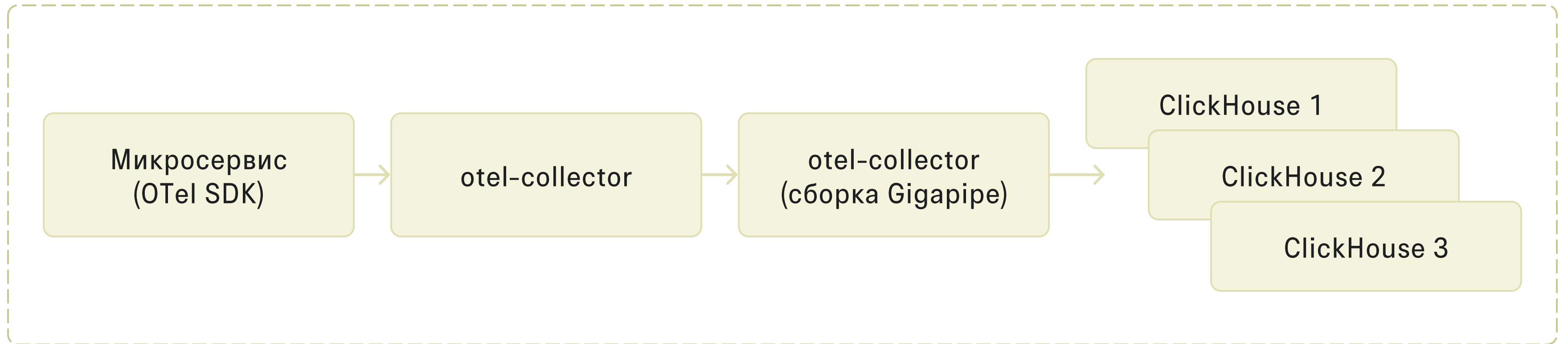


Схема формы 2



КЛАСТЕРИЗАЦИЯ В ДОКУМЕНТАЦИИ:
«ПОДДЕРЖИВАЕТСЯ, НО ...



Антипаттерн Gіgarіre: Кастомная сборка коллектора

GIGAPIRE ПОСТАВЛЯЕТ
СОБСТВЕННЫЙ ДИСТРИБУТИВ
OPENTELEMETRY COLLECTOR

Внутри экспортер
под их схему ClickHouse.



ИХ README:

образ v0.154 перепрыгнул
46 минорных версий
апстрима за один шаг.



ОБЯЗАН ТАЩИТЬ ИХ СБОРКУ

Свой otel-collector собрать нельзя. Issue #81
«error decoding processors», help wanted.

(01)

ЭКСПОРТЕР НЕ В CONTRIB

Issue #23 «Contributing as an Exporter to Contrib»
открыт с мая 2023, до сих пор help wanted.

(02)

«БЕСКОНЕЧНОЕ» ОЖИДАНИЕ ОБНОВЛЕНИЙ

46 минорных версий OTel за один прыжок.

(03)

Проблема 1. Singlenode Clickhouse

GIGAPIRE РАБОТАЛ ТОЛЬКО
С SINGLENODE, ФЛАГИ
КЛАСТЕРИЗАЦИИ НИЧЕГО НЕ
ДЕЛАЛИ*



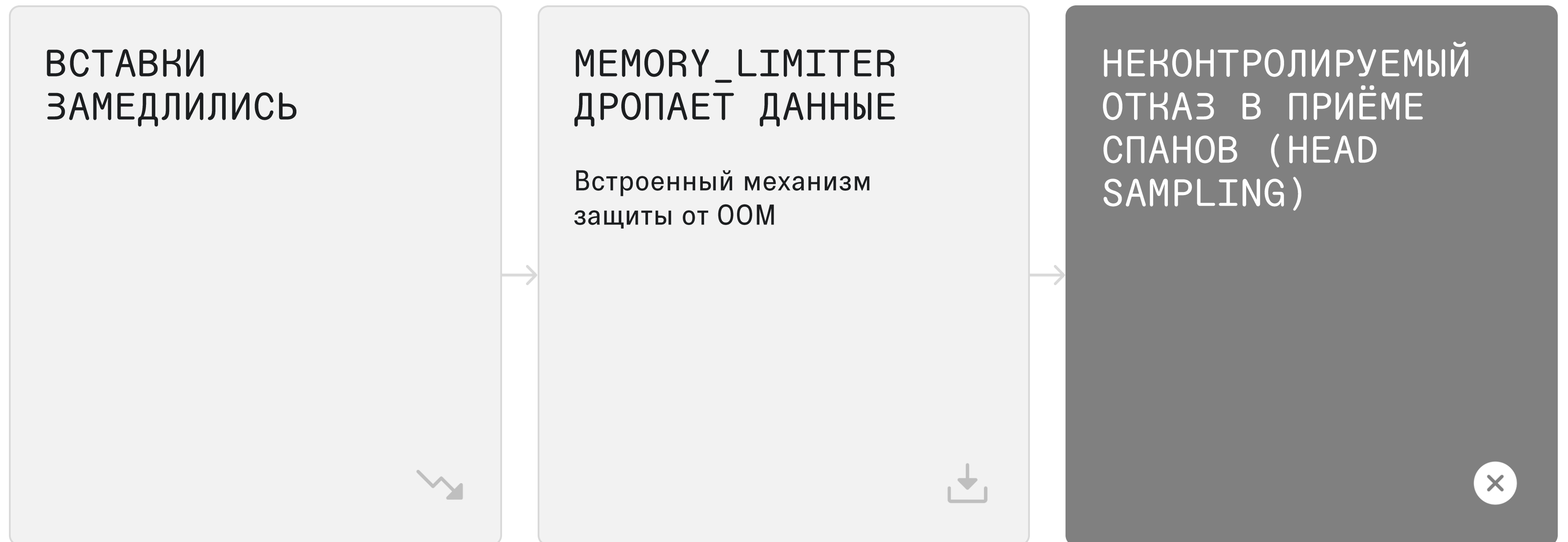
РЕШИЛИ – DISTRIBUTED
TABLES ПОВЕРХ
REPLICATEDMERGETREE



*СКЕЙЛИНГ КЛАСТЕРОВ ТОЛЬКО
ЧЕРЕЗ РУЧНУЮ РАСКАТКУ DDL,
ВОЗМОЖНО СЕЙЧАС УЖЕ ВСЁ ОК



Проблема 2. Backpressure на доставку



Вывод: Gіgarіре не подходит

МОЖЕМ МАСШТАБИРОВАТЬ
ХРАНИЛИЩЕ



МЕДЛЕННАЯ РЕПЛИКАЦИЯ
ПО ЖИРНЫМ НОДАМ



ТРАНСПОРТ ТЕРЯЕТ ДАННЫЕ



Вывод формы 2



УЗКОЕ МЕСТО –
НЕ ТОЛЬКО БЭКЕНД,
НО И ТРАНСПОРТ

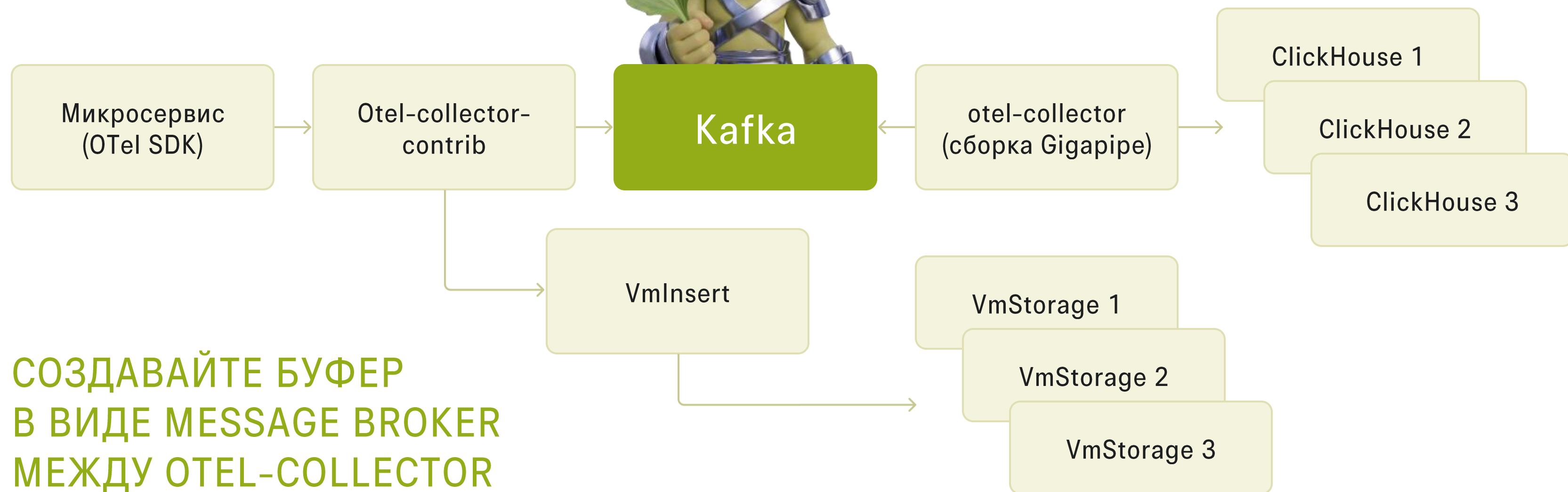
Как полечить доставку?

Sanity-check:

Как можно решить
проблему с потерей
данных на транспорте?



Предлагаю вариант



Форма 3. VictoriaTraces

ПРОСТО
В ЭКСПЛУАТАЦИИ



УДОБНО
МАСШТАБИРОВАТЬ



ХОРОШО
СЖИМАЕТ ДАННЫЕ



При 10 000 спанов/сек
и лимитах 4 CPU / 8 GiB

CPU составил:

0,50

vCPU

0,69

CPU y ClickHouse

Память примерно одинаковая:

1,15

GiB

1,12

GiB

Диск:

3,27

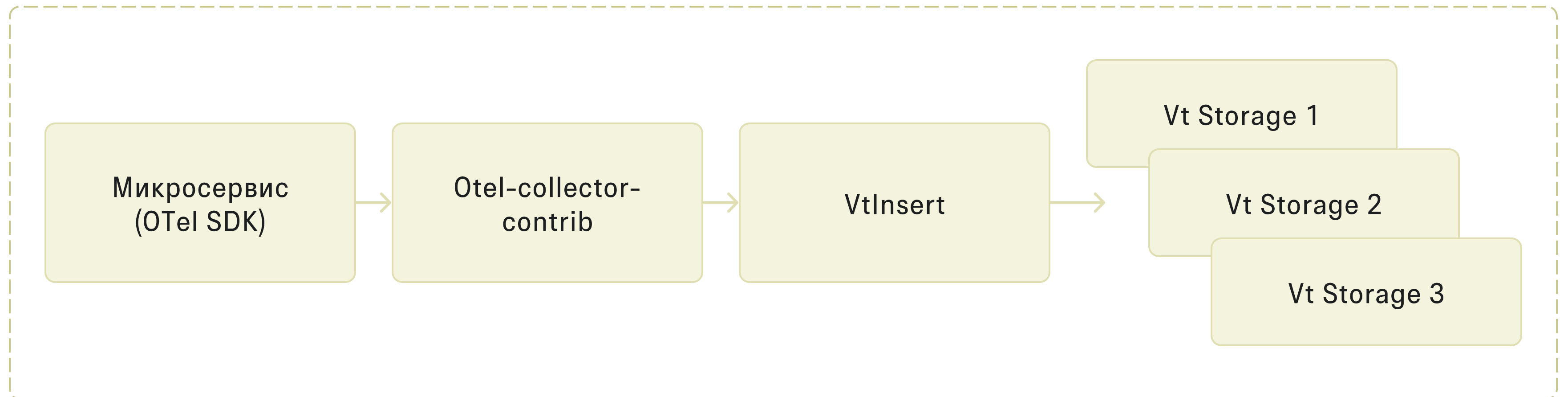
GiB

5,86

GiB



Сначала делали так



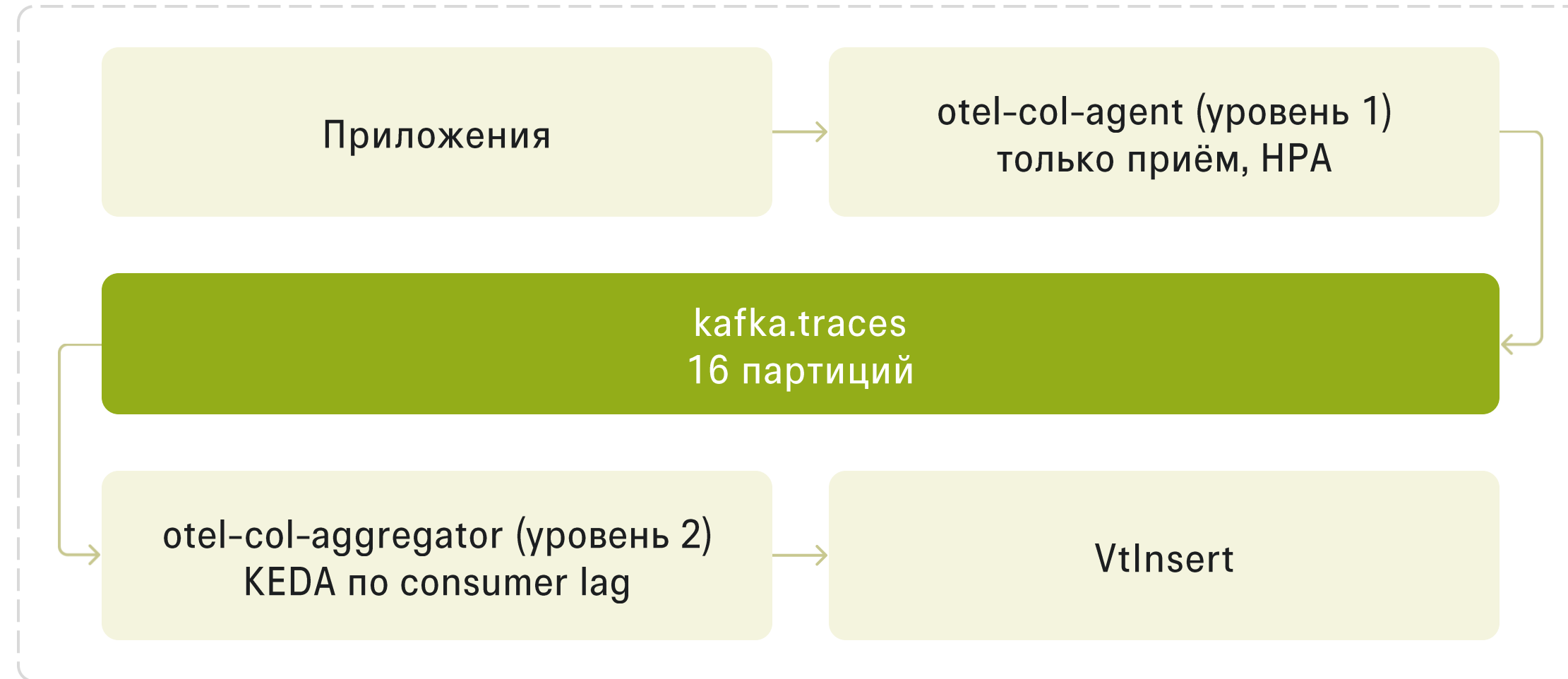
ПЕРЕНЕСЛИ ДОСТАВКУ
КАК ЕСТЬ



BACKPRESSURE
ОТ ХРАНИЛИЩА УШЁЛ



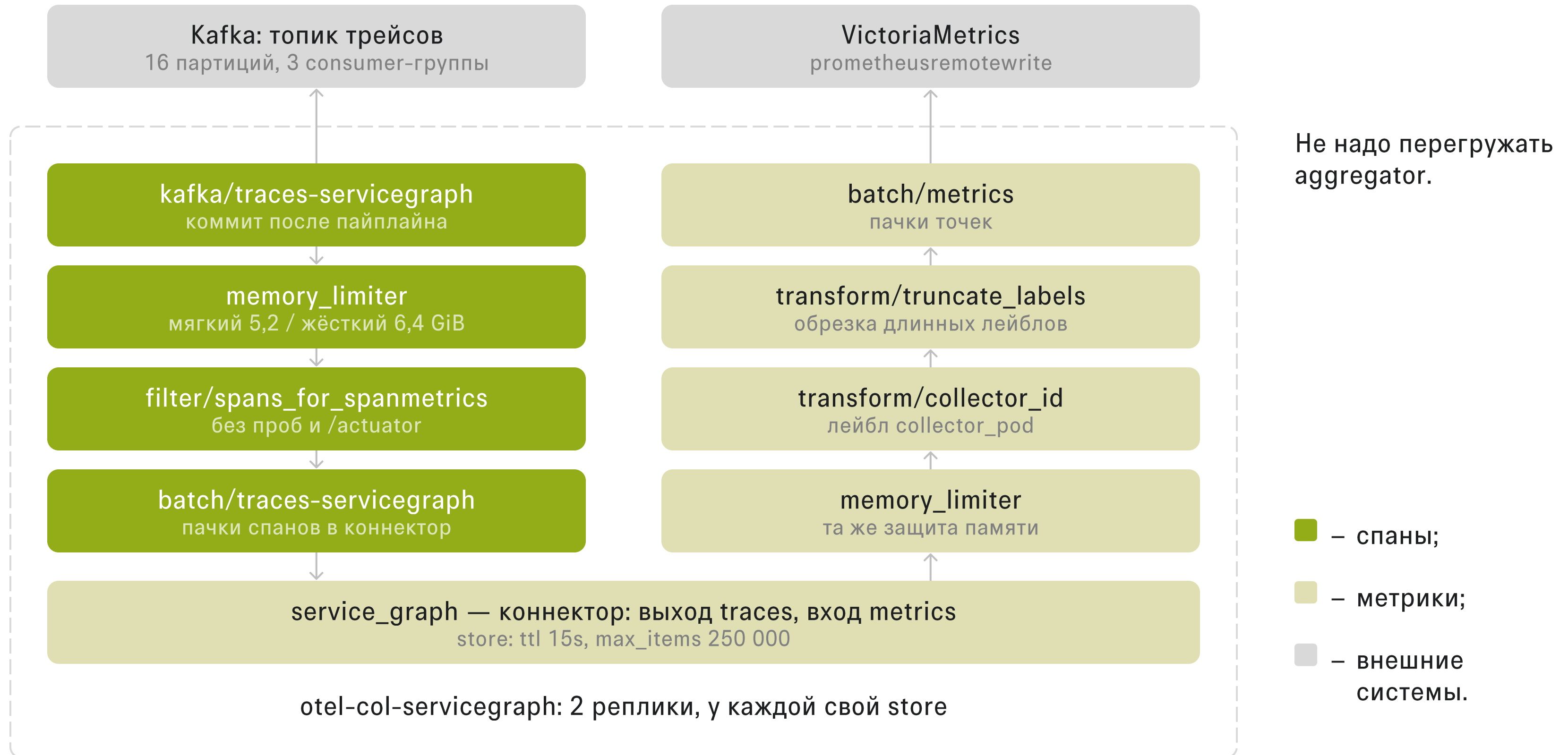
Потом добавили буфер



Приём не зависит от скорости бэкенда. Хранилище тормозит, обновляется или лежит: приложения этого не видят. Всё, что раньше дропал `memory_limiter`, лежит в Kafka и ждёт.



Дополнительно — выделяем serviceGraph



Выделили 3 роли

AGENT

- Принимает opentelemetry-сигналы из внешних источников.
- Складывает в Kafka.

(01)

AGGREGATOR

- Вычитывает трейсы из Kafka.
- Отправляет трейсы в VtInsert.
- Генерирует метрики на базе трейсов.

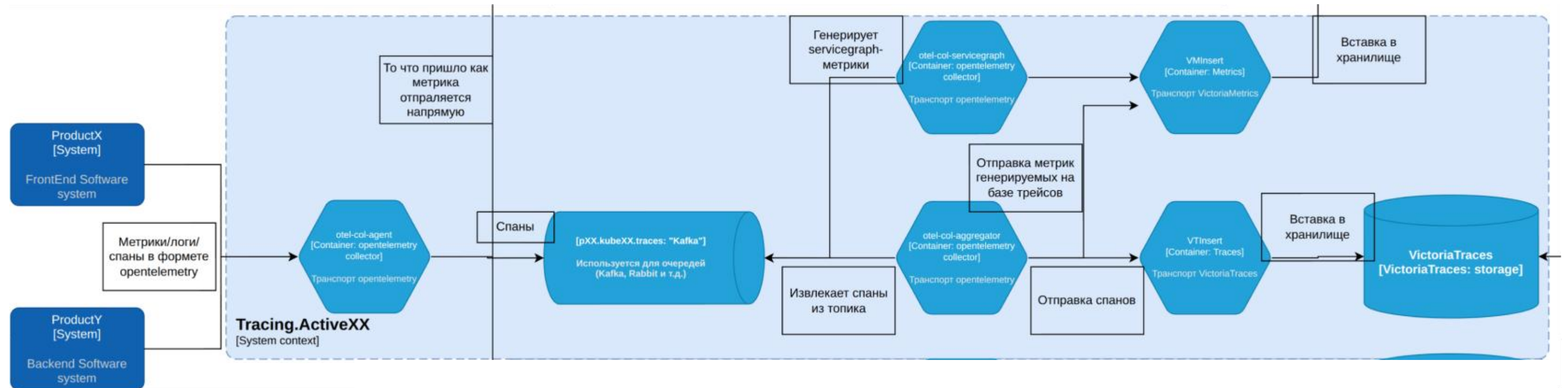
(02)

SERVICEGRAPH

- Вычитывает трейсы из Kafka.
- Генерирует метрики для карты связей.

(03)

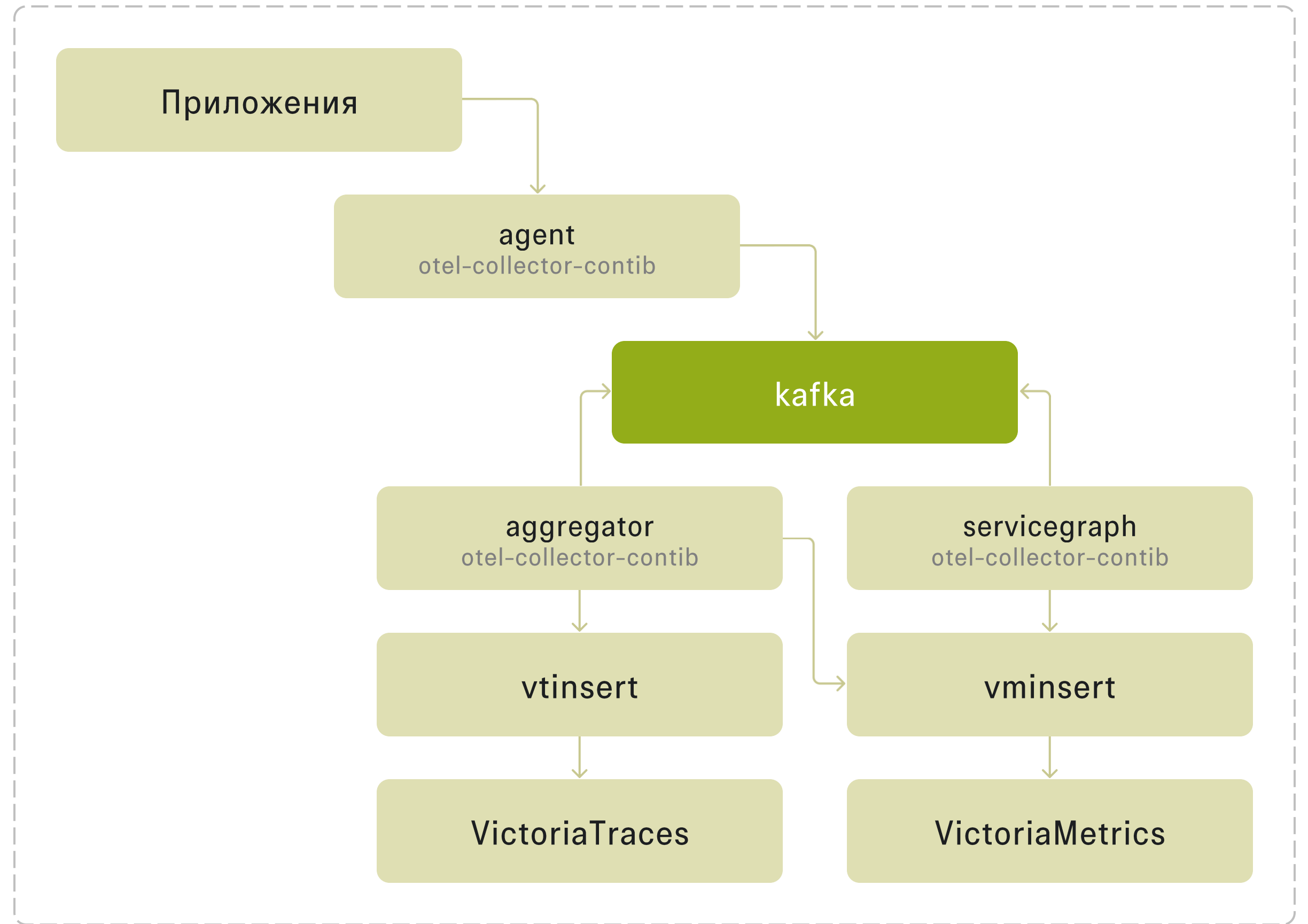
Страшная схема: Путь трейса



Как часть схемы доставки выглядела во внутренних планах



Схема для людей



КАК ОНО ОЩУЩАЕТСЯ, КОГДА ВОШЛИ ВО ВКУС

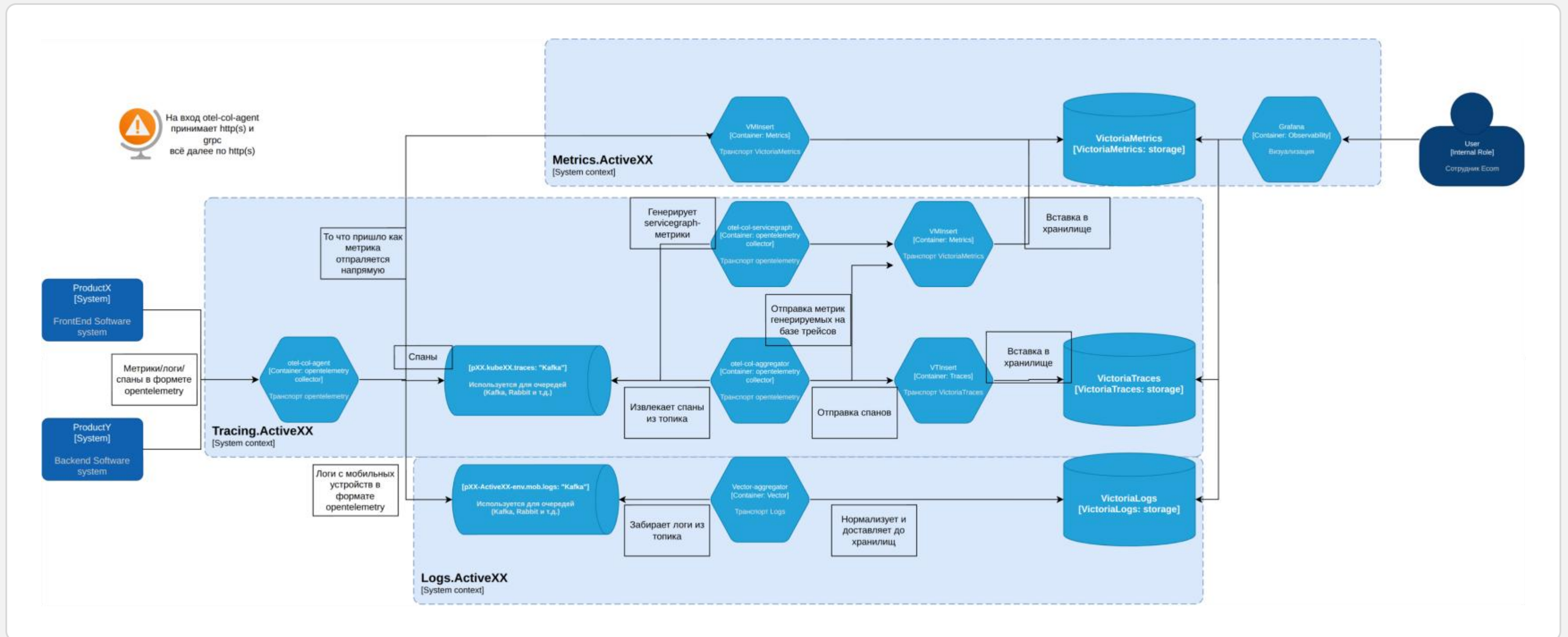
otelcol с одним
exporter debug
и ГОТОВО



400 строк otel-
collector-config.yaml,
contrib, connectors,
transform
и k8sattributes



Добавили логи и метрики для фронтов



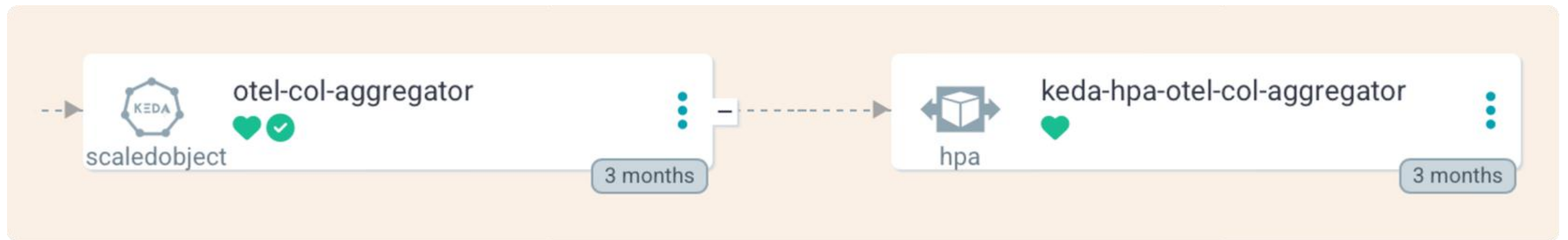
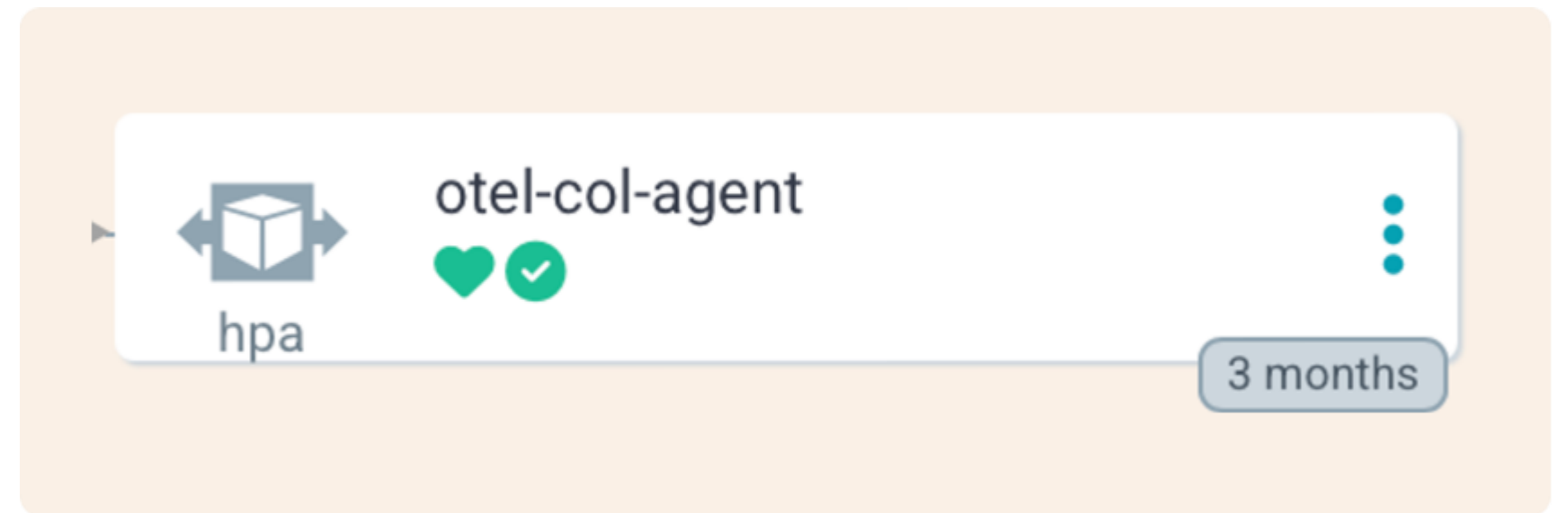
Один транспорт на три сигнала. Трейсы: otel-col-agent, kafka, otel-col-aggregator, VictoriaTraces.
Метрики: vminsert, VictoriaMetrics, то, что пришло как метрика, идёт напрямую. Логи: kafka, vector, VictoriaLogs.

Автоскейлинг для экономии ресурсов

HPA ПО РЕСУРСАМ НА AGENT,
VMINSERT, VTINSERT



KEDA ПО ЛАГУ ТОПИКА
НА AGGREGATOR



Цифры по Kafka в одной инсталляции

5

брокеров
Kafka

16

партиций
в топике

~64

буфера, retention
ограничен
только диском

~40млн

сообщений
в буфере

Прочитали из топика

Отправили в VtInsert

Коммит оффсета

Что нам это стоило по ресурсам: Доставка

servicegraph (01)

2 replica

2 CPU

8 RAM

aggregator (02)

2–16 replica

2 CPU

8 RAM

agent (03)

2–8 replica

2 CPU

4 RAM

vtinsert/vminsert (04)

2–4 replica

2 CPU

4 RAM

Ресурсы на 1 pod

Что нам это стоило по ресурсам: Хранилища

VictoriaTraces

3–11 nodes

16 CPU

32 RAM

2TB disk

(01)

Kafka

5 nodes

8–16 CPU

16 RAM

512GB – 1TB disk

диск под ~6 часов буфера

(02)

Люди (хранители знаний)

команда внедрения
и поддержки 1.5 FTE

(03)

Масштаб сейчас

~1 млн

Span/sec



2 недели

retention на трейсы



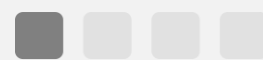
~23

инсталляции



Четыре формы Спанмона

БЫСТРОЕ
РЕШЕНИЕ



КЛАСТЕРИЗАЦИЯ



БЫСТРЫЕ
ХРАНИЛИЩА



ОТКАЗОУСТОЙЧИВАЯ
ДОСТАВКА



Выводы

НЕ ТАЩИ ВЕНДОРСКИЕ ДИСТРИБУТИВЫ OTEL COLLECTOR

OTel Collector Builder и contrib.
Вендор даёт сборку: проверь,
как и с какой задержкой он её
обновляет. Gigaripе: 46 версий
апстрима за один прыжок.

форма 2

(01)

ОТКАЗОУСТОЙЧИВЫЙ ТРАНСПОРТ ВАЖНЕЕ ИДЕАЛЬНОГО БЭКЕНДА

Kafka, коммит после записи,
автоскейлинг по лагу.

формы 2 и 4

(02)

СТРОЙ НА САМОМ ПОПУЛЯРНОМ ФОРМАТЕ

Трижды сменили бэкенд
и ни разу не тронули код
продуктов. Формат даёт
свободу менять решение.

все формы

(03)

ЕСЛИ БЫ ВЫ РЕШИЛИ
РЕАЛИЗОВЫВАТЬ TAIL-
СЭМПЛИНГ ДЛЯ ЗООПАРКА
СИСТЕМ, ТО КАК
БЫ ВЫ ЭТО ДЕЛАЛИ?



ВОПРОСЫ

Валерий Евдокимов

Tech Lead, Observability platform

ecom.tech



📍 @franda1

