



Санкт-Петербургский государственный университет
Кафедра системного программирования

Интеграция микросервисов в систему логов в формате OpenTelemetry

Евгений Владимирович Бакаев, группа 20.Б11-мм

Научный руководитель: ст. преп. Сартасов С.Ю.

Консультант: руководитель групп разработки DataCamp и технологий индексации, ООО «Яндекс.Технологии», Хвастунов А.П.

Санкт-Петербург
2023

- Мониторинги, оповещения и логи – инструменты для реагирования и отлаживания
- Телеметрия – сбор и анализ данных о работе различных систем
- OpenTelemetry – набор инструментов для телеметрии

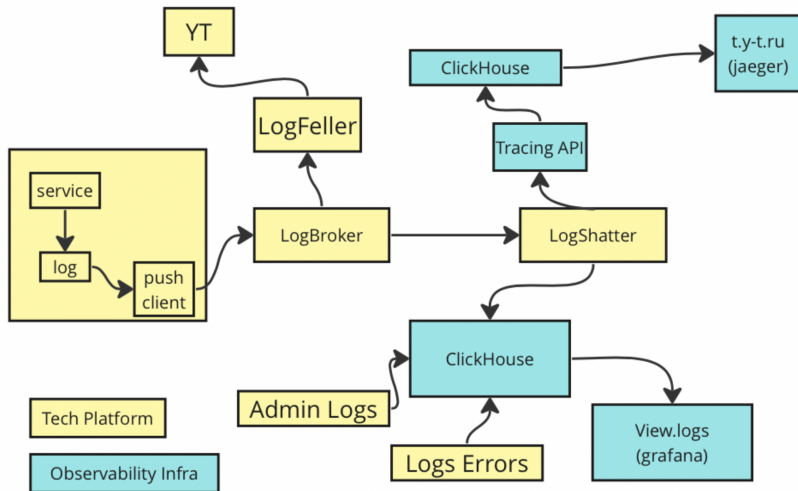
При решении инцидентов логи помогают:

- Оценить масштаб инцидента
- Найти время начала инцидента
- Локализовать проблему и начать ее отлаживать

Для микросервисной архитектуры важно иметь:

- Хранилище с логами всех сервисов в одном месте
- Возможность агрегировать записи по нужным меткам
- Возможность просматривать логи с минимально возможной задержкой

Поставка логов через Logbroker

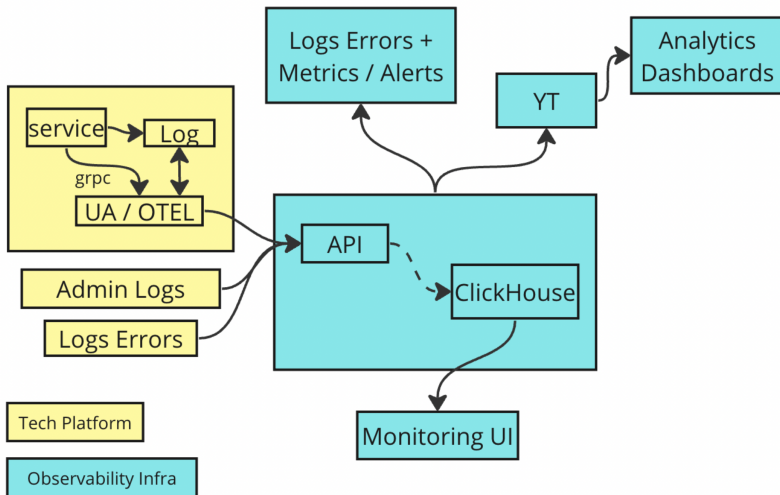


Постановка задачи

Цель данной работы – реализовать прямую поставку логов в ClickHouse для всех микросервисов внутри DataCamp. Для достижения цели были поставлены следующие задачи:

- Выполнить обзор существующих инструментов для поставки и агрегации логов
- Выполнить обзор всех форматов логов, использующихся в DataCamp
- Поддерживать поставку логов для нового микросервиса
- Перенести логи с системы поставки через Logbroker на прямую
- Протестировать доставку логов

Поставка логов напрямую в ClickHouse



Формат логов в DataCamp

- tskv (Tab-Separated Key-Value). Строка лога начинается с "tskv" и состоит из пар ключ-значение, разделенных \t
- BigRT.
2023-09-04 15:22:29,578979\tI\tBigRT\tSomeMessage\tthread
\tSomeLabel\tAnotherLabel
- Cpp.
INFO: 2023-09-04 00:00:27.974 +0300 someFile.cpp:1 2 Some
Message
- Py.
2023-09-04 15:19:05,289 INFO [Module.File:123] some message

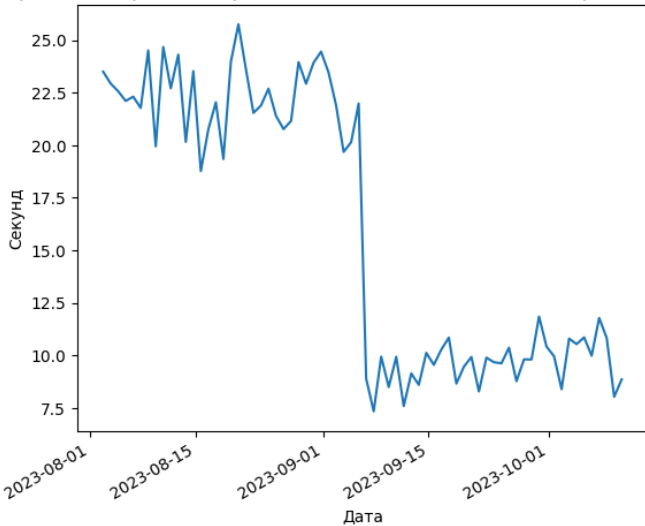
Переход на прямую поставку логов

Для каждого микросервиса внутри DataCamp были проделаны следующие шаги:

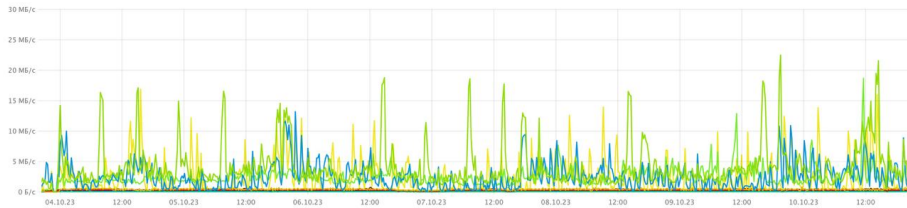
- Внутри каждого контейнера обновлен агент поставщик (Unified Agent) до версии, поддерживающей OpenTelemetry
- Настроен конфиг на прямую поставку логов
- Протестирована доставка логов в каждом окружении (testing, prestable, production)

Тестирование

95-перцентиль среднего времени доставки логов по всем сервисам DataCamp



Объем поставляемых логов



Заключение

В ходе работы были выполнены все поставленные задачи:

- Выполнен обзор существующих инструментов для поставки и агрегации логов
- Выполнен обзор всех форматов логов, использующихся в DataCamp
- Для всех сервисов DataCamp выполнен переход с системы поставки логов через Logbroker на прямую поставку
- Протестирована доставка логов для всех сервисов