Модуль "Мониторинг"

Модуль 4

Бочаров Филипп

Руководитель центра мониторинга и наблюдаемости в МТС Digital

Занимаюсь разработкой платформы Наблюдаемости. Помогаю продуктовым командам сделать работу сложных распределенных систем понятной и прозрачной.

Спикер Highload++, Dotnext, TechleadConf ...

Тема 1
Теория наблюдаемости и мониторинга

Базовые понятия

Наблюдаемость – возможность отвечать на вопросы о работе системы

Мониторинг - возможность оперативно и превентивно реагировать на изменения в работе системы

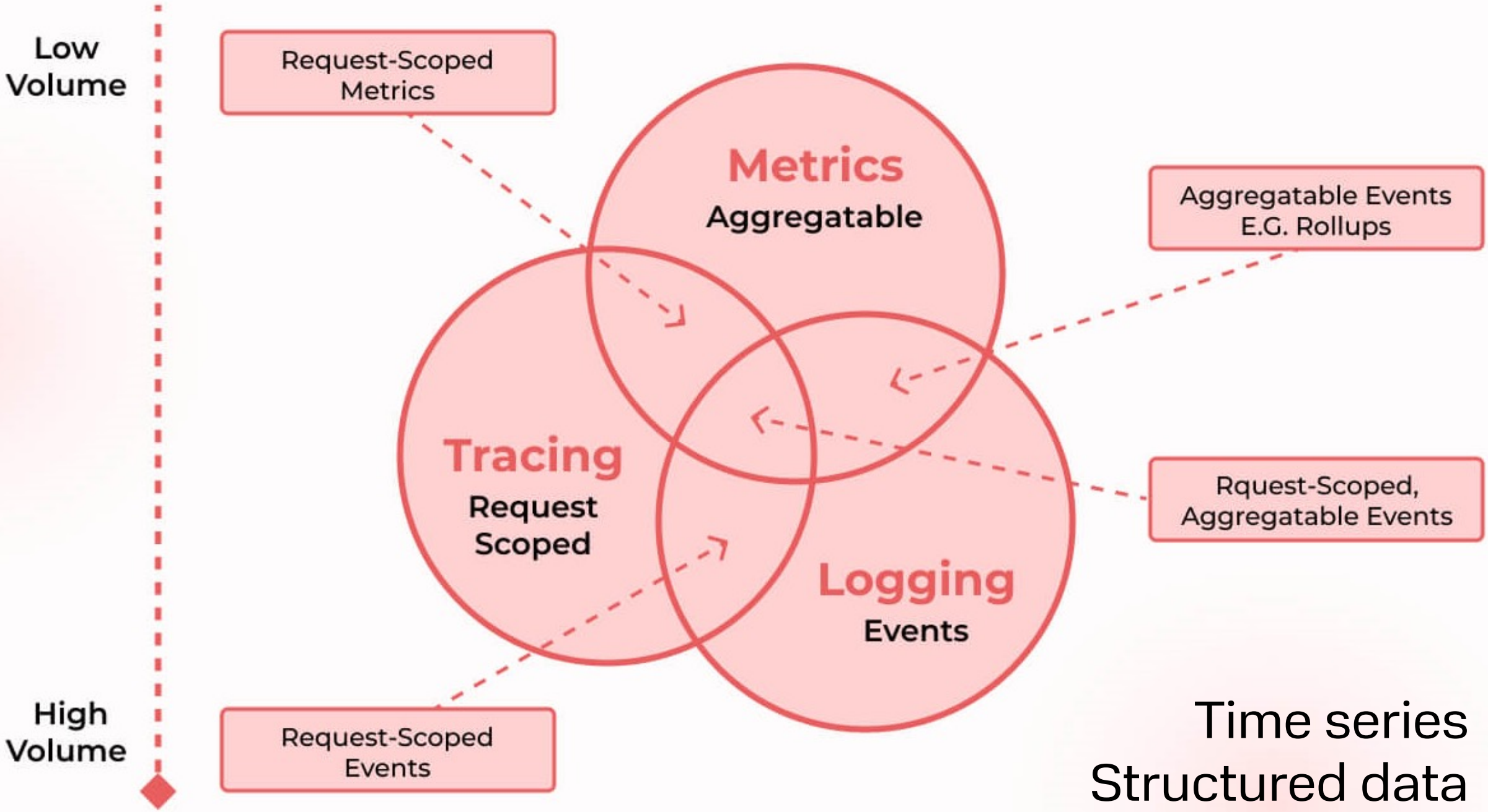
Зачем SRE наблюдаемость и мониторинг?

Невозможно управлять тем, что не можешь измерить (с)

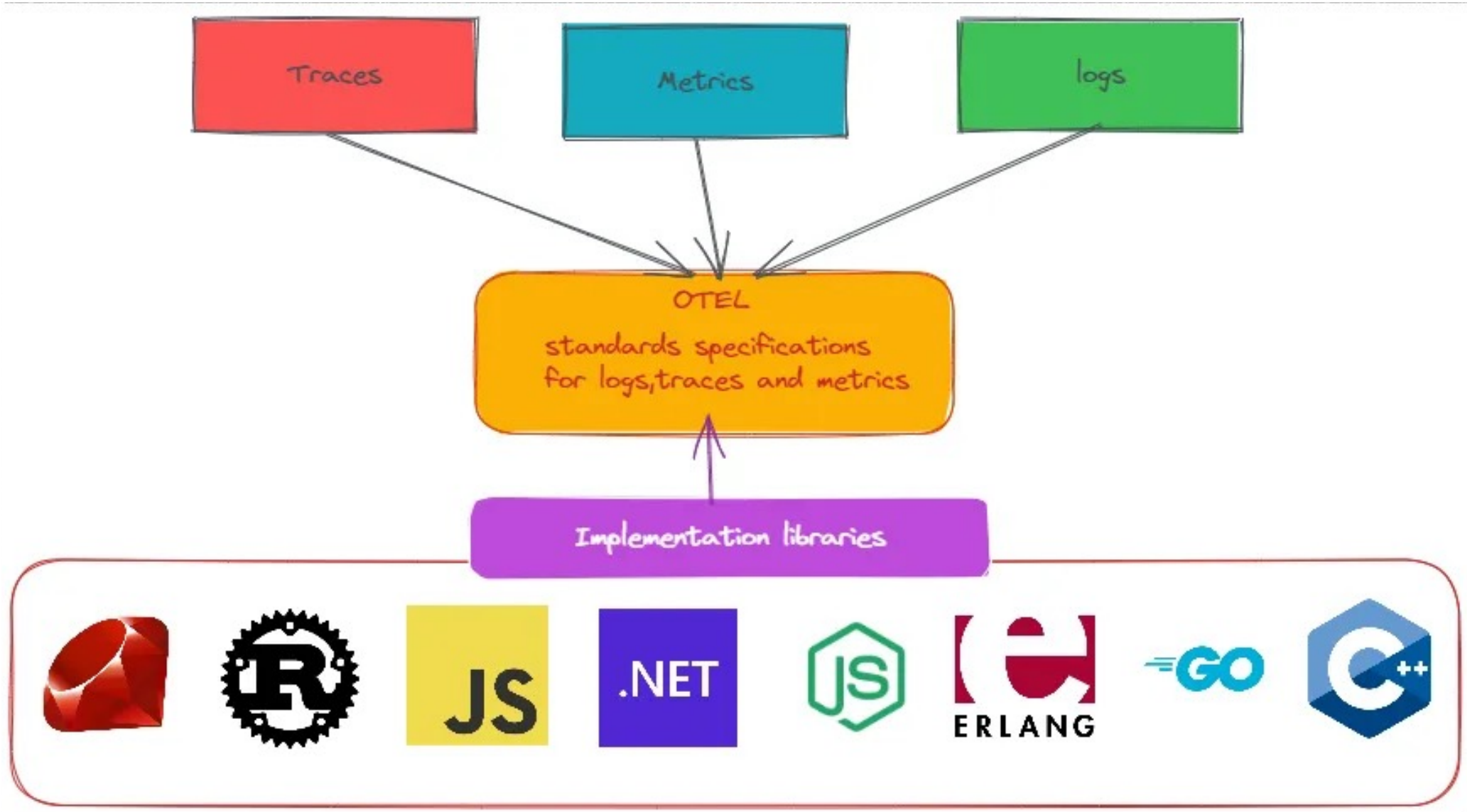
Для SRE:

1. Контроль качества продукта: доступности и производительности
2. Выявление проблем **ДО** их влияния на бизнес
3. Быстрая реакция на аварию

Связь понятий


Low Volume
High Volume
Request-Scoped Metrics
Metrics
Aggregatable
Aggregatable Events E.G. Rollups
Tracing
Request Scoped
Rquest-Scoped, Aggregatable Events
Logging
Events
Request-Scoped Events
Time series
Structured data

Единый стандарт OpenTelemetry

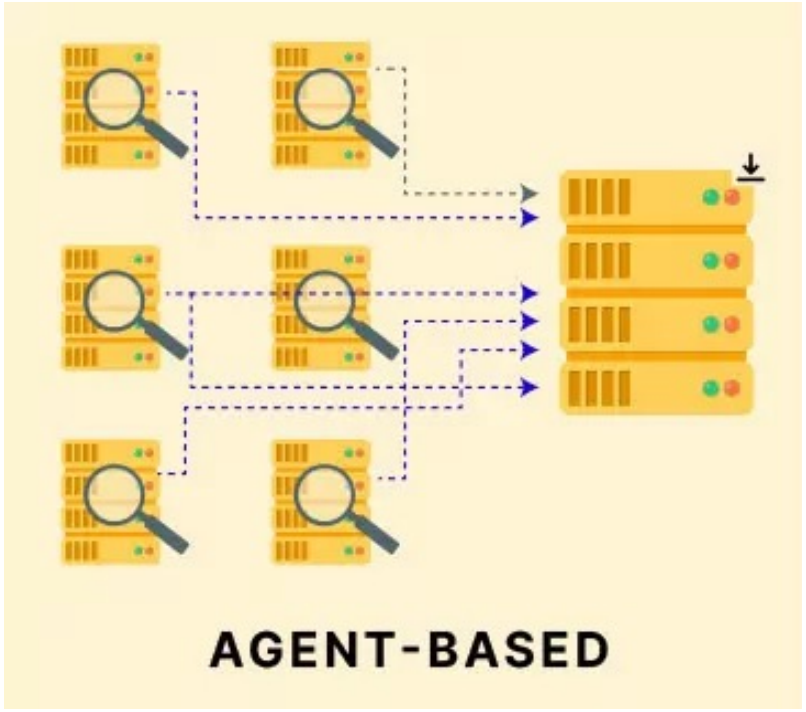


Тема 2
Компоненты мониторинга

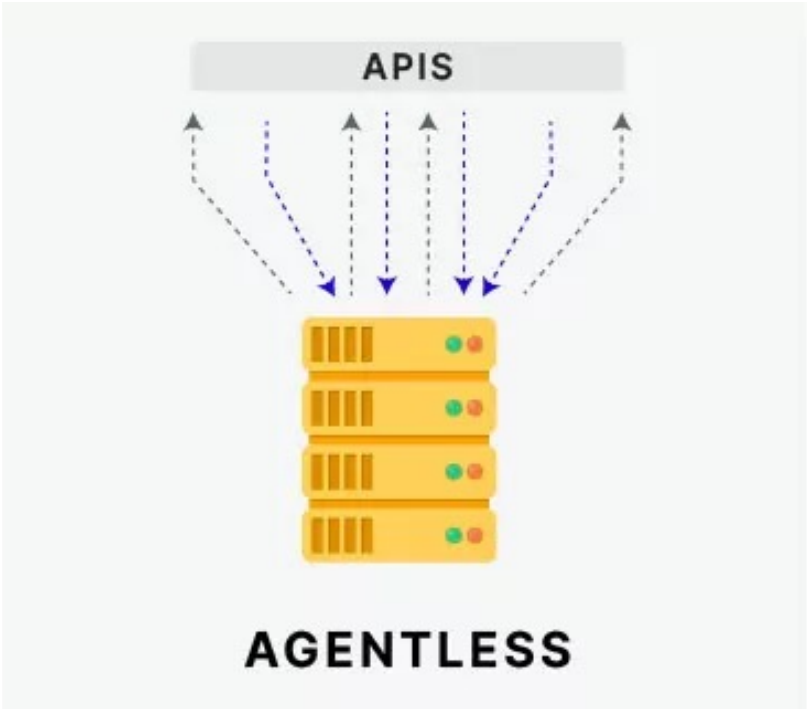
Необходимые компоненты

1. Агенты сбора метрик
2. Горячее и холодное хранилище метрик
3. Визуализация метрик (дашборды)
4. Генерация событий по правилам
5. Корреляция, дедупликация, обогащение событий и алертинг

Агенты сбора: агентский и безагентский мониторинг



VS

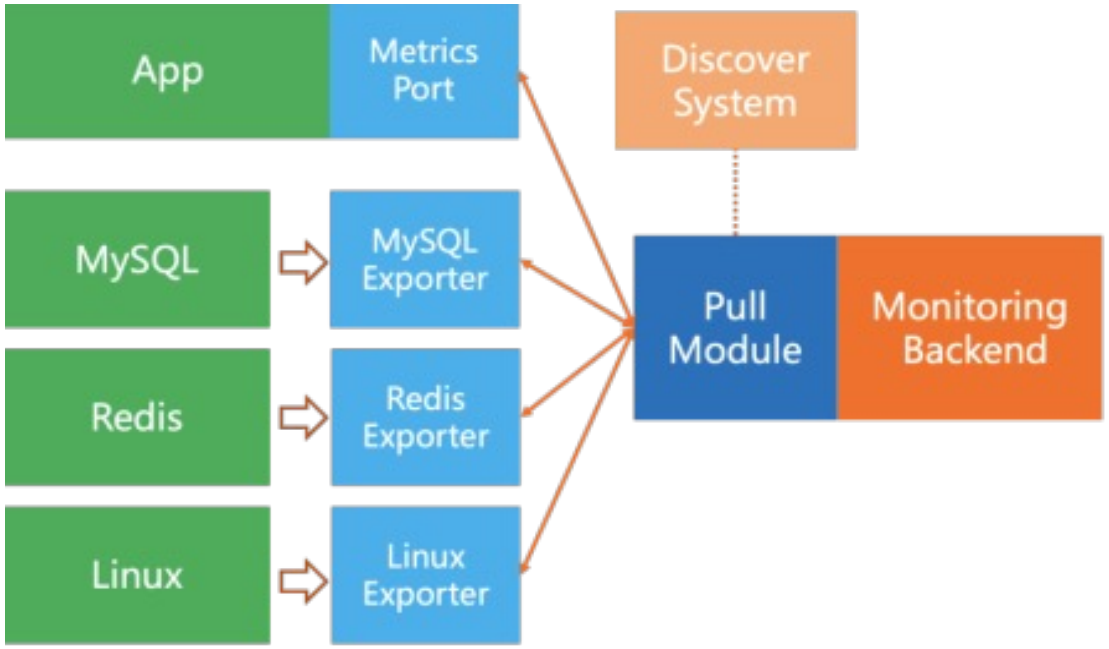


Агент устанавливается непосредственно на хост и передает данные в хранилище

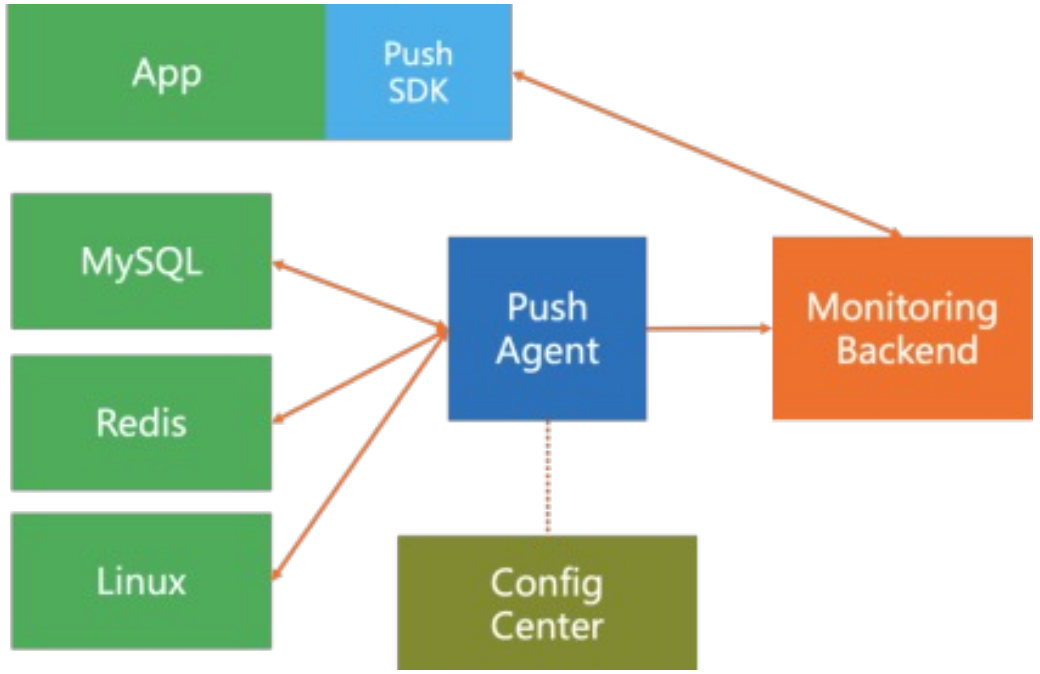
Агент собирает данные удаленно по протоколам:

- SSH (Secure Shell)
- SNMP (Simple Network Management Protocol)
- WMI (Windows Management Instrumentation)
- HTTP/S (Hypertext Transfer Protocol Secure)
- JMX (Java Management Extensions)

Агенты сбора: модель pull и push



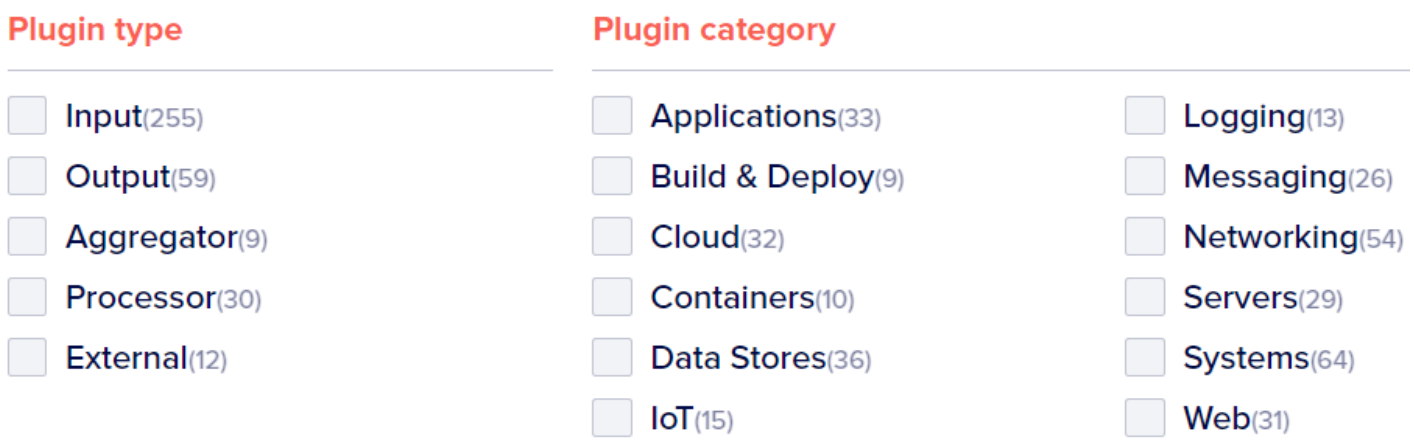
- Для короткоживущих процессов/джобов все равно приходится применять push
- Сложно настраивать для окружения с множеством подсетей и сетевых сегментов



- Выше риск превышения нагрузки на backend и потерь метрик
- Лучше подходит для real-time метрик

Пример агента: telegraf

- Модель push
- Простой текстовый конфиг
- Множество готовых плагинов для сбора метрик с различного ПО
- Конвейер обработки позволяет агрегировать и фильтровать данные
- Поставляется в виде единого binary файла
- Поддержка множества протоколов для отгрузки метрик



Горячее и теплое хранилище метрик

Горячее хранилище

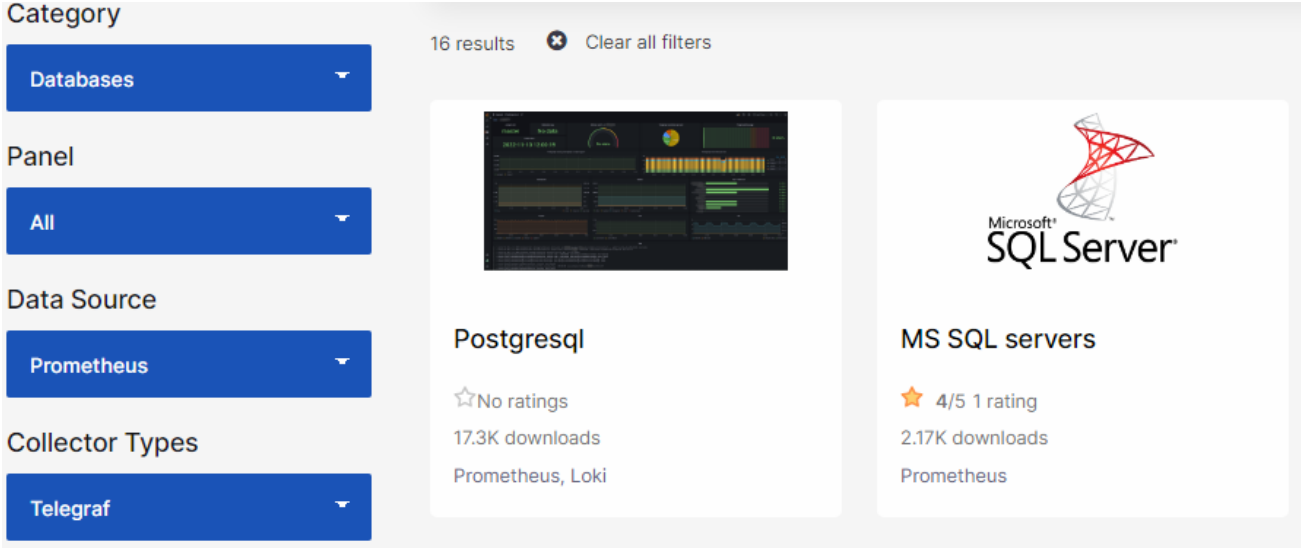
- Хранение последних X периодов
- Дорогое железо
- Быстрые ответы
- Высокая гранулярность данных

Теплое хранилище

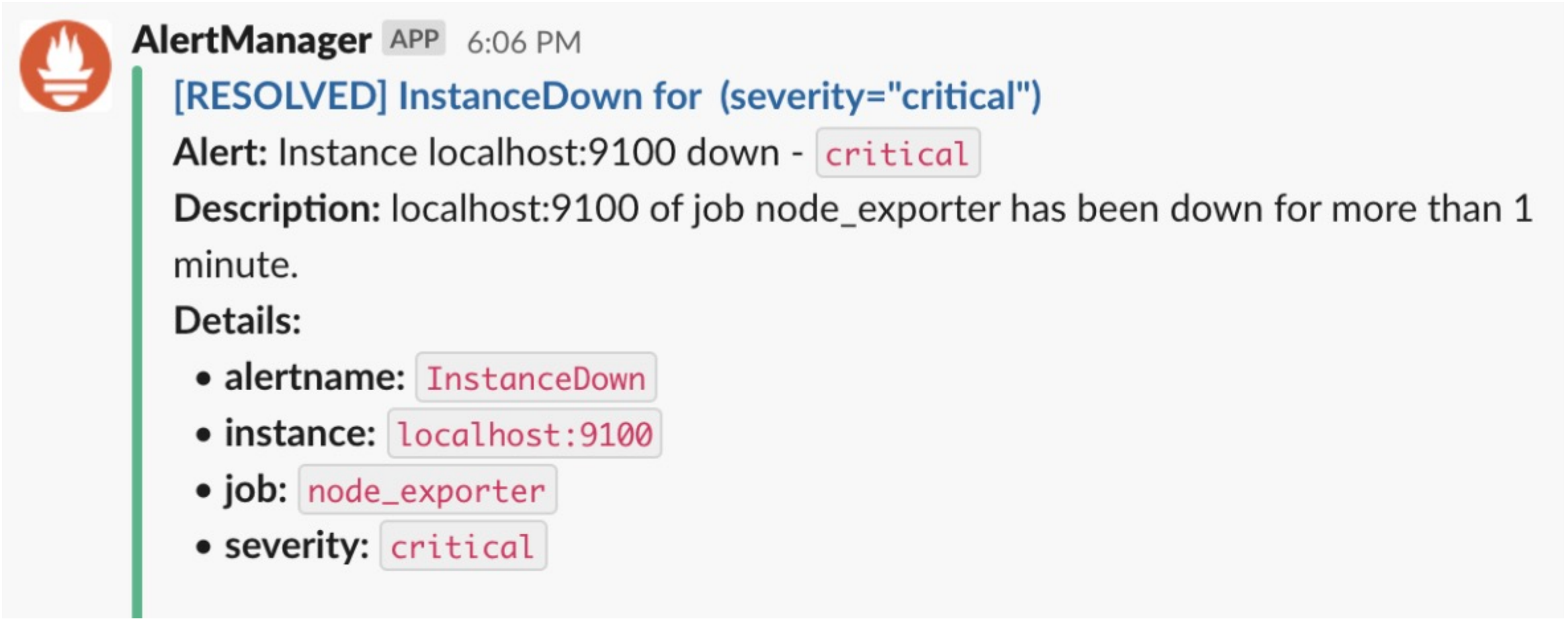
- Хранение долгосрочных трендов
- Дешевое железо
- Медленные ответы
- Низкая гранулярность данных

Визуализация метрик и дашборды: Grafana

- Поддержка любых источников данных
- Множество плагинов для визуализации
- Библиотека готовых дашбордов
- Экспорт дашбордов и данных
- Собственный алертинг
- https://grafana.com/grafana/plugins/
- https://grafana.com/grafana/dashboards



Алерты

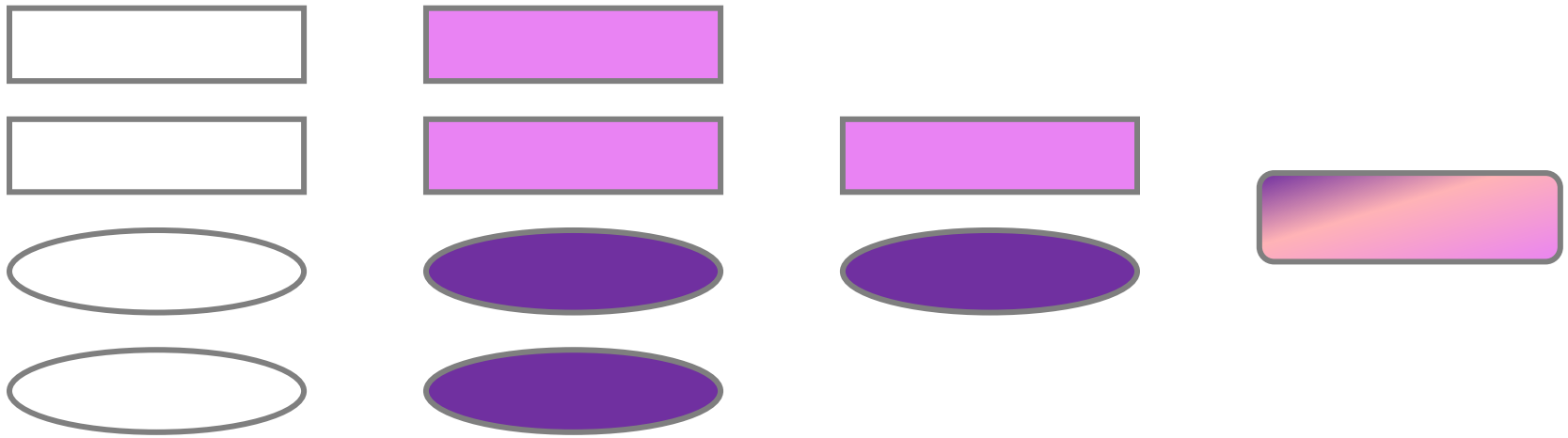

AlertManager APP 6:06 PM
[RESOLVED] InstanceDown for (severity="critical")
Alert: Instance localhost:9100 down - critical
Description: localhost:9100 of job node_exporter has been down for more than 1 minute.
Details:
• alertname: InstanceDown
• instance: localhost:9100
• job: node_exporter
• severity: critical

Признаки плохих алертов

- Постоянно приходит открывающее и закрывающее событие
- Команда поддержки НЕ знает, что делать с событием
- Не проводится исследование с системным решением проблемы
- Одна проблема порождает множество алертов

Обработка событий

- **Обогащение** – добавление новых полей и контекста к событию
- **Дедупликация** – отбрасывание дублирующихся событий
- **Корреляция** – связывание нескольких событий в одну цепочку



Классический стек	Под высокие нагрузки	Наблюдаемость 3 в 1	Old but gold	InfluxDB
Prometheus	Victoria Metrics	Elasticsearch	Zabbix	InfluxDB
Timeseries	**Timeseries**	Объектно ориентированное	Реляционное	**TimeSeries**
PromQL	MetricsQL (расширение Prom QL)	Query DSL		SQL
• Скрейпит самостоятельно • Pushgateway* • **OpenTelemetry**	• telegraf • Vmagent • prometheus • **OpenTelemetry**	• Metricbeat • **OpenTelemetry**	• Zabbix agent	• telegraf • **OpenTelemetry**
Expression browser / PromLens / **Grafana**	vmui / **Grafana**	Kibana / **Grafana**	Zabbix web / **Grafana**	InfluxDB UI / **Grafana**
alertmanager	vmalert	Kibana alerting	Built in trigger	Built in checks and notification rules

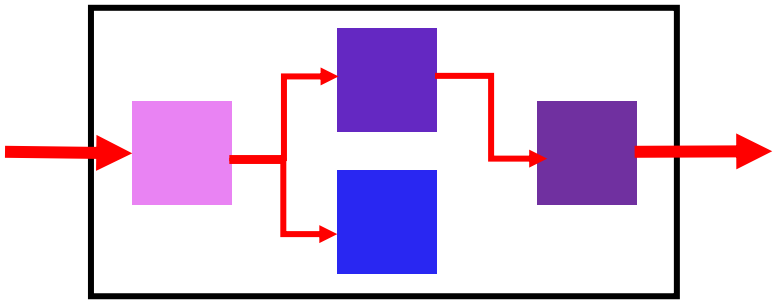
Тема 3
Типы мониторинга или какие метрики нужны

Black box

Видим только вход и выход из системы.
Внутреннее устройство неизвестно

- ✓ **Не требуется изменение исходного кода**
- ✓ Мониторинг "**глазами пользователя**"
- ○ Только **симптомы** проблемы
- ○ **Синтетические транзакции от робота**, а не реальный клиентский опыт

White box



Видим все внутренние
компоненты системы и связи
между ними

- ✓ **Дает представление о корневой причине проблемы**
- ✓ **Оценивает реальный опыт** всех клиентов
- ○ **Требует инструментирования** кода системы

Бизнес

Ключевые сценарии	SLI/SLA
Продуктовые метрики	MAU/DAU, Conversion Rate, Bounce Rate

Прикладное ПО

Метрики собственного кода	Размер внутреннего буфера
Сторонние компоненты: базы данных, веб-сервера, очереди ...	Lag очередей, отставание реплик БД, количество 500х ответов, ...
Внешние зависимости	Время и коды ответа внешней системы

Инфраструктура

Kubernetes	Потребление RAM/CPU подов
Виртуализация / ОС	Потребление RAM/CPU VM
Коммунальные сервисы	S3, SSO, AD, ...
Сеть	WAF, Load balance, сетевые устройства
Инженерка ЦОД	Электропитание, температура

4 золотых сигнала мониторинга

Задержка / Latency

Время ответа системы - blackbox или whitebox

Отдельно замеряем успешные и ошибочные запросы. Используем перцентили.

Насыщение / Saturation

Процент утилизации ресурсов системы - whitebox

На основе
метрик потребления инфраструктуры: CPU, RAM, Disk, Network, ...

Нагрузка / Traffic

Количество входящих запросов (RPS/RPM) - whitebox

Ошибки / Errors

Процент успешных ответов - blackbox или whitebox


The Four Golden Signals
Zoom Out
Last 6 hours
LATENCY
800 ms
600 ms
400 ms
200 ms
0 ms
10:00
11:00
12:00
13:00
14:00
15:00
2017-04-04 12:00:40
OK: 8 ms
Errors: 1 ms
OK Avg: 110 ms
Errors Avg: 6 ms
TRAFFIC
50 ops
40 ops
30 ops
20 ops
10 ops
0 ops
Requests/s Avg: 25 ops
ERRORS
6
5
4
3
2
1
0
code_400 Total: 2
code_500 Total: 38
SATURATION
100%
80%
60%
40%
20%
0%
DB CPU Avg: 53%

Метрики веб-страниц: Web Vitals

Largest Contentful Paint (LCP)
Скорость загрузки основного контента <= **2,5 с**

First Input Delay (FID)
Время ожидания до первого взаимодействия с контентом <= **100 мс**

Cumulative Layout Shift (CLS)
Совокупное смещение макета <= **0,1**.

75% пользователей должны укладываться в норматив

Overview | Web Vitals | Tags

Search for events, users, tags, and more

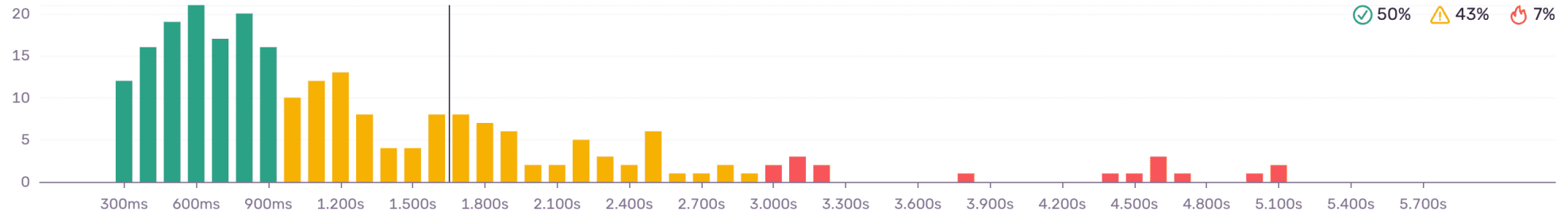
Outliers: Exclude | Reset View

First Paint (FP)

1.76s

Render time of the first pixel loaded in the viewport (may overlap with FCP).

Open in Discover

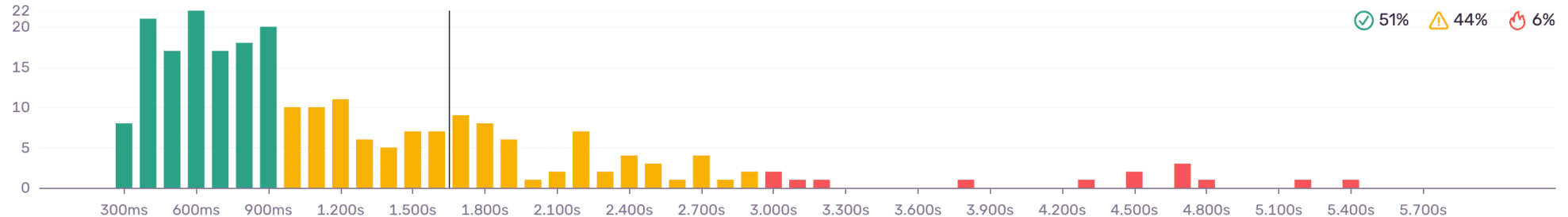


First Contentful Paint (FCP)

1.73s

Render time of the first image, text or other DOM node in the viewport.

Open in Discover



Largest Contentful Paint (LCP)

2.22s

Render time of the largest image, text or other DOM node in the viewport.

Open in Discover