

Немного теории SRE



Что такое SRE?

- Site Reliability Engineering is a discipline that incorporates aspects of software engineering and applies them to infrastructure and operations problems.

https://en.wikipedia.org/wiki/Site_Reliability_Engineering

- SRE's approach to this is to apply a software engineering mindset to system administration topics

<https://www.scalyr.com/blog/what-is-sre/>

- ... what happens when you ask a software engineer to design an operations function

Ben Treynor, founder of Google's Site Reliability Team



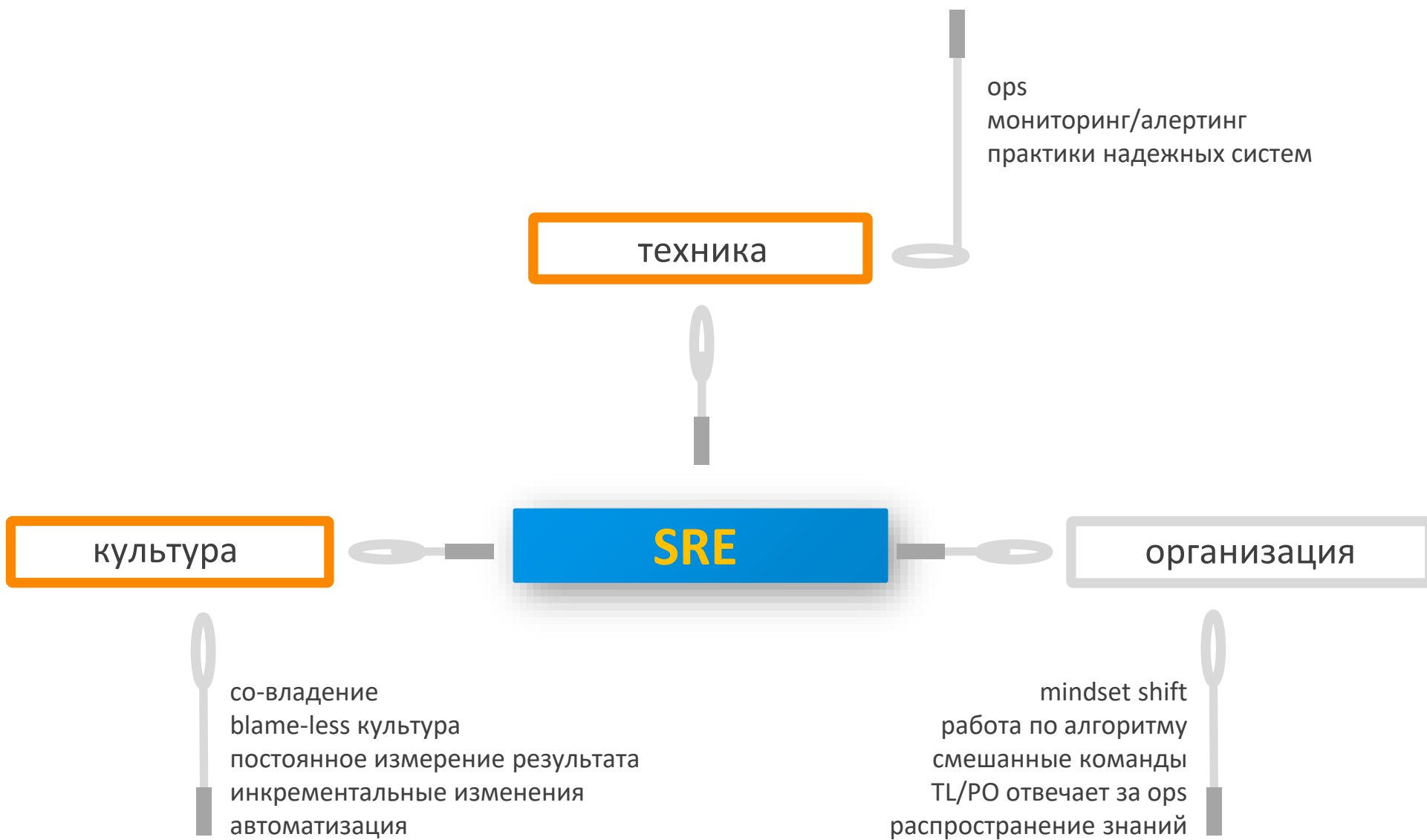
Как мы понимаем SRE

- SRE – дисциплина, которая помогает вместе думать об одной цели – как сделать так, чтобы наш продукт приносил пользу клиентам;
- metrics-driven ops & development.



всё далее —
это видение **SRE**
авторов курса





SRE vs DevOps

DevOps говорит, куда нужно
двигаться, но не говорит как

SRE отвечает на вопрос
(частично)

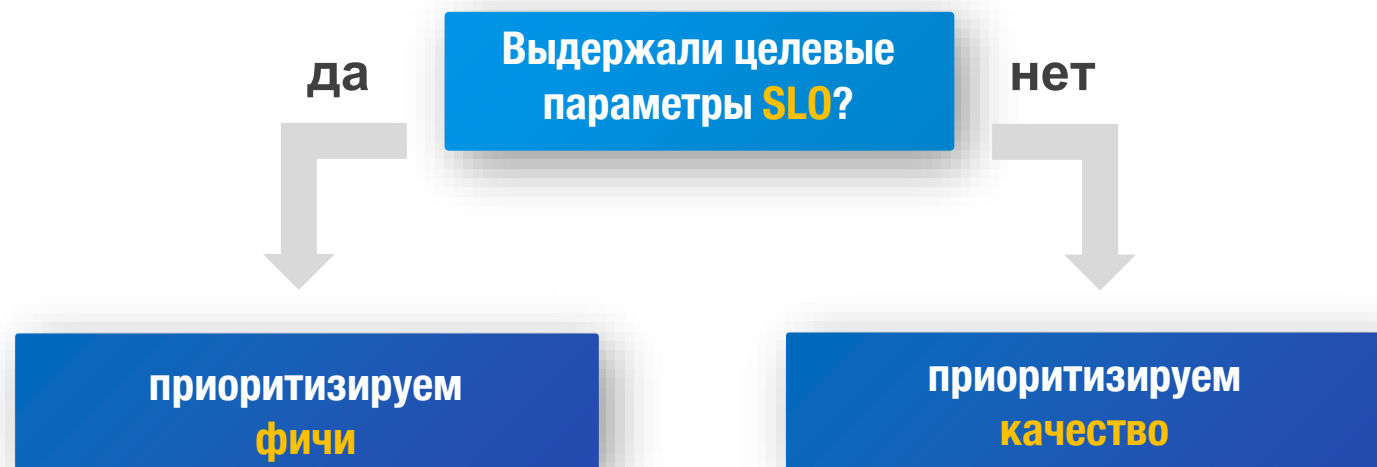
SLO SLI SLA

SLO SLI SLA



SLO – Service Level Objective

- Задаёт **целевой** уровень качества предоставляемой услуги;





SLO – Service Level Objective

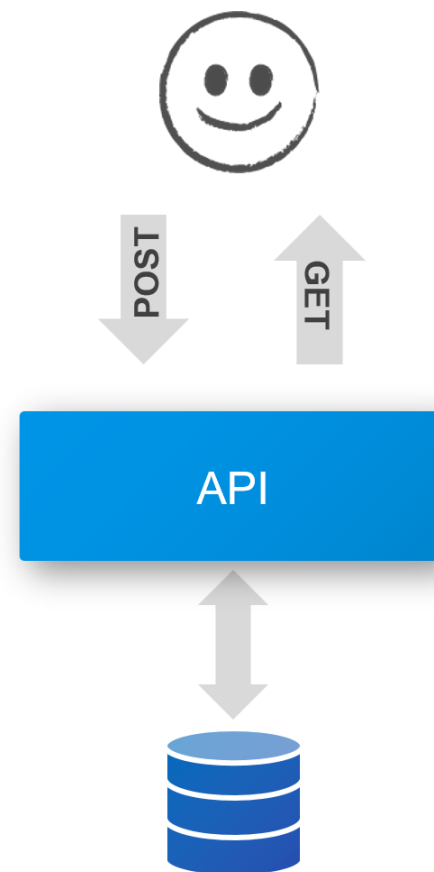
- Задаёт **целевой** уровень качества предоставляемой услуги;
- Инструмент **data-driven** приоритизации.



provider_backend сервис

API

GET /movies
GET /seances
POST /bookings

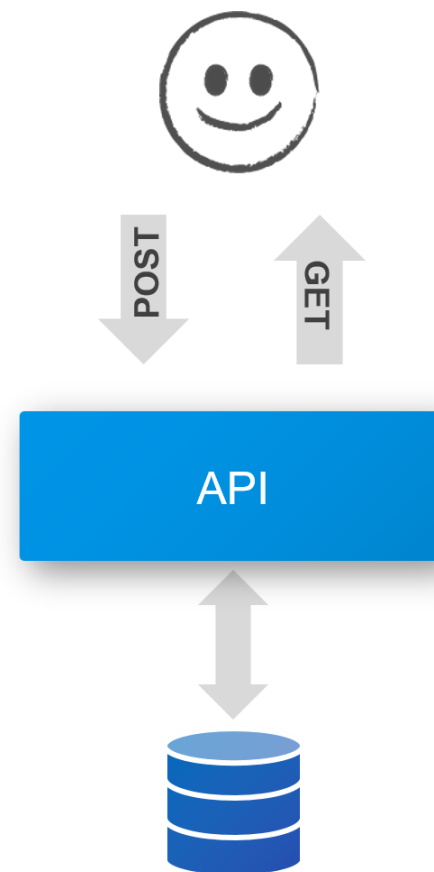




Гипотетический SLO

Сервис должен быстро и бесперебойно оказывать услуги поиска и бронирования мест, а также надежного хранения и извлечения броней.

да – добавим фичу поиска
нет – делаем, чтобы сервис работал



SLO SLI SLA

SLO SLI SLA



SLI – Service Level Indicator

- SLO – **задает** целевой **уровень** качества услуги;
- SLI – **измеряет** качество;
- $SLI = \text{кол-во хороших событий} / \text{общее кол-во событий}$



Service Level Indicators

Тип	Измерение	Описание
API	availability	соотношение запросов, на которые был успешно получен ответ, и общего количества ответов
API	latency	соотношение запросов, на которые ответ был получен быстрее определенного порога, и общего количества ответов
API	quality	соотношение запросов, на которые ответ был получен с пониженным качеством, и общего количества ответов
storage	durability	соотношение сохраненных записей, которые могут быть успешно прочитаны, и общего количества записей
pipeline	freshness	соотношение данных, которые были обновлены раньше определенной точки во времени, и общего количества данных
pipeline	correctness	соотношение записей, обработка которых выдала корректный результат, и общего количества записей
pipeline	coverage	для групповой обработки - соотношение обработанных записей для потоковой обработки - соотношение записей обработанных за определенный временной интервал

source: Site Reliability Workbook