

Dapper, a Large-Scale Distributed Systems Tracing Infrastructure

В данной статье сотрудники компании *Google* описывают систему для трассировки выполнения вызовов RPC, HTTP запросов и прочих межсетевых коммуникаций связанных с работой сервисов.

Предложенный ими метод должен опираться на несколько целевых метрик:

1. Низкие накладные расходы – ввиду того, что *Google* оперирует сервисами в таких масштабах, что даже небольшие дополнительные расходы повлекут за собой огромные траты на физические ресурсы и крайне значительные суммарные задержки;
2. Невидимость на уровне приложения – разработчики не должны уделять внимание трассировке. Другими словами разработчики не создают свое взаимодействие с инфраструктурой трейсинга;
3. Масштабируемость – система должна быть масштабируема как минимум на протяжении нескольких лет(внутри компании Google).

Помимо этого трэйсинг должен удовлетворять такому свойству как причинно-следственная связь.

Основная часть(второй раздел) статьи описывает несколько техник помогающих в создании такой инфраструктуры – *trace context*, аннотации и функции помощники(*helpers*). Вместе они создают инфраструктуру которая помогает выполнять трейсинг на масштабах *Google*.

В четвертой часть авторы акцентируют внимание на том как их система работает на конкретном железе и как она влияет на перфоманс системы.

Личное мнение: статья написана хорошо, читается легко. Однако тема весьма специфична для больших компаний и ввиду того, что у меня не хватает экспертизы в этой области. Помимо этого большая проблема заключается в том, что в виду того, что это статья от компании нам не дан ни код, ни реальная возможность протестить эту систему.