Raft Consensus Algorithm

Обзор Raft

Raft — алгоритм

для решения задач консенсуса

в сети надёжных вычислений.

Обзор Raft

Rластер, состоящий из нескольких узлов.

Все узлы содержат копию
одного и того же состояния

(данные и бизнес-логика).

Обзор Raft

часть узлов может падать (и возвращаться)

может теряться связь между узлами

могут добавляться новые узлы в кластер

Обзор Raft

Нужно как-то гарантировать,

что состояния на всех узлах идентичны.

Raft предлагает способ,
как этого можно добиться.

Обзор Raft

Это алгоритм общего назначения,

на основе которого можно построить

разные прикладные системы.

Обзор Raft

"In Search of an Understandable Consensus Algorithm"

Diego Ongaro and John Ousterhout

Stanford University

http://ramcloud.stanford.edu/raft.pdf

Обзор Raft

https://raft.github.io/

Интерактивная модель

Ссылки на статьи по теме

Ссылки на реализации

Основы Raft

В обычном режиме:

один узел выполняет роль **Leader**,

все остальные узлы роль **Follower**.

Основы Raft

Leader

принимает все запросы от клиентов,
применяет их к своему состоянию,

и рассылает эти запросы
всем остальным узлам,

чтобы они тоже применили их
к своему состоянию.

Основы Raft

Follower-узлы не взаимодействуют с клиентами и друг с другом,

а только получают запросы от **Leader**.

Основы Raft

Leader может упасть
или потерять связь с остальными узлами.

Тогда оставшиеся узлы
должны выбрать нового лидера
и вернуться в штатный режим работы.

Основы Raft

Существуют промежутки времени,
когда в кластере нет лидера.

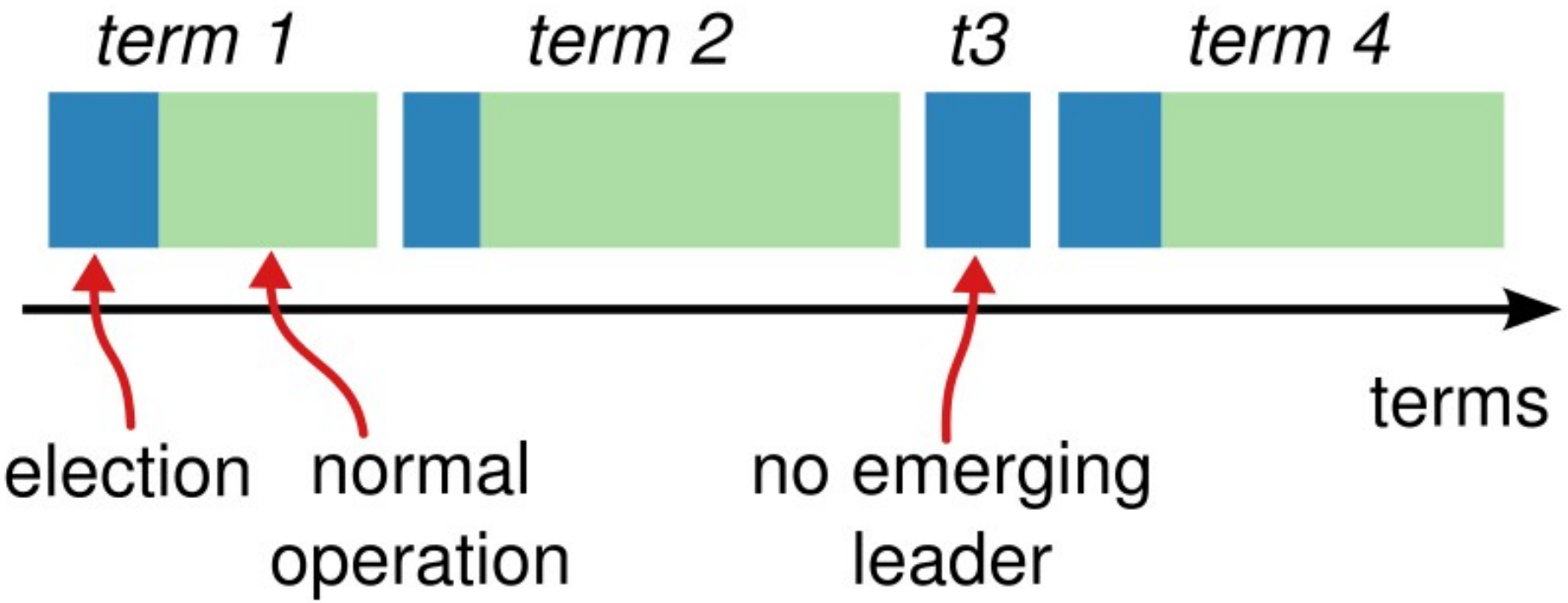
В это время кластер
не может обслуживать запросы клиента
(или работает только на чтение).

Основы Raft

Период работы
от одного выбора лидера
до следующего выбора
называется **Term**.

Каждый Term имеет свой номер.

Основы Raft


term 1
term 2
t3
term 4
terms
election
normal operation
no emerging leader

Основы Raft

Каждый запрос от клиента
в терминах Raft называется **Log**,

и каждый Log имеет свой id.

Основы Raft

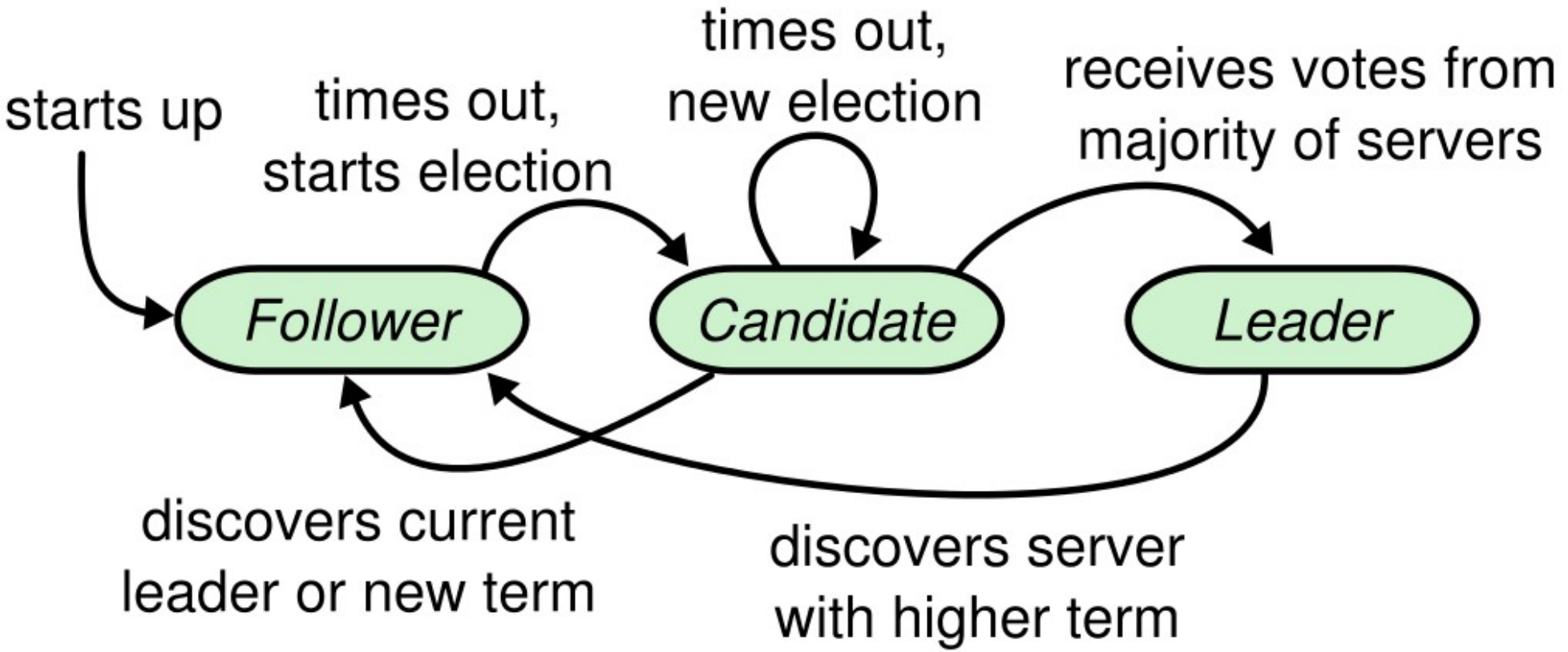
Алгоритм Raft
делится на отдельные задачи:

Выбор лидера

Репликация логов

Изменение размеров кластера

Выбор лидера



Выбор лидера

https://raft.github.io/

Репликация логов

Клиент не знает, какой из узлов в кластере является лидером.

Репликация логов

Запросы от клиента
обрабатываются в два этапа:

Append и **Commit**.

Изменение размеров кластера

Добавление или удаление узлов,

без остановки кластера.

Изменение размеров кластера

Информация о появлении нового узла

распространяется по кластеру

не мгновенно.

Изменение размеров кластера

Информация о появлении нового узла

распространяется по кластеру

не мгновенно.

Изменение размеров кластера

Двухфазный механизм выбора лидера

Live Coding ...