

# Errors in Database Systems, Eventual Consistency, and the CAP Theorem

Sergey Serebryakov, 545

Source: [blog post](#) on acm.org + comments

# I'm 12 years old and what is this

2/9

- **Теорема CAP:** в одной распределенной системе невозможно одновременно обеспечить свойства согласованности, высокой доступности и устойчивости к разделению сети

# Теорема CAP

3/9

- **C: Consistency (согласованность)**
  - ▣ логическая целостность в смысле ACID
  - ▣ реплики в согласованном состоянии
- **A: Availability (доступность)**
  - ▣ если происходит какой-либо сбой, система должна оставаться работоспособной, переключаясь при надобности на какую-нибудь реплику
- **P: Partition-tolerance (устойчивость к разделению)**
  - ▣ если возникает сбой в работе сети, приводящий в разделению узлов обработки на две группы, которые не могут общаться, то должно быть можно продолжать обработку в обеих этих группах

# Eventual consistency

4/9

- Пусть сеть разделилась на несколько участков, в каждом несколько обрабатывающих узлов; тогда реплики *когда-нибудь* придут к одному и тому же состоянию, а именно, когда:
  - связность сети восстановится
  - пройдет достаточное время для очистки реплик

# Представим архитектуру

5/9

- Типичная модель аппаратуры
- Узлы обработки и хранения данных объединяются в кластер с использованием средств построения локальной сети (LAN)
- Кластеры, в свою очередь, связываются посредством технологии территориально-распределенных сетей (WAN)

# Ошибочные ситуации

6/9

1. Ошибки в приложениях - некорректные операции обновления данных, нужен откат базы
2. Повторяющиеся ошибки СУБД – на каком-то узле обработки СУБД падает, и это повторяется на реплике
3. неповторяющиеся ошибки СУБД – СУБД падает, но на реплике норм (мб асинхронные операции)
4. Ошибки операционной системы – на каком-то узле падает вся ОС
5. Сбой аппаратуры в локальном кластере (сбои памяти, дисков и т.д.) – аварийная остановка ОС или СУБД
6. Сбой локальной сети в кластере – узлы кластера не могут больше общаться
7. Стихийное бедствие – кластер уничтожается в результате наводнения
8. Сбой сетевой аппаратуры, связывающей кластеры в WAN – кластеры не могут больше общаться

# Применимость теоремы CAP

7/9

- Случаи 1 и 2 – нельзя обеспечить доступность системы
- Случай 7 – фиксируем транзакцию только после передачи её на другой кластер? Задержка
- Случаи 3, 4, 5 – локальные, сводятся к вырожденному случаю ошибки 6 (минус один узел) – легко обходятся, в LAN лучше жертвовать P
- Случай 8 – редкий из-за избыточности WAN или тривиальный
- Вывод: неразумно всё время жертвовать C в пользу ES: есть неприменимые случаи, а во многих применимых это негодный компромисс

# Случай замедления

8/9

- Причины
  - ▣ Неравномерность нагрузки
  - ▣ Проблемы с пулом буферов
- Можно рассматривать как сбой
  - ▣ Но отказываться от части аппаратуры?
  - ▣ Всё равно потом восстанавливать
- Правильно:
  - ▣ Мониторинг
  - ▣ Добавлять вычислительную мощность
  - ▣ Auto-provisioning

# Welcome to the desert of the real

9/9

- Можно поддерживать репликацию и «буфер согласованности»
  - ▣ Нет сбоев – согласованность
  - ▣ Сбой – часть записей согласована «строго», часть согласована «eventually»
- Ошибка 8
  - ▣ Синхронная запись в несколько узлов – будет задержка
  - ▣ Eventual consistency – асинхронная запись
  - ▣ В случае проблем не писать вообще
- Настраиваемость характеристик CAP
- Consensus Quorum
- Двухфазная фиксация – слишком большие задержки