УНИВЕРСИТЕТ ИТМО

Проектирование БД

Лекция 6.

Распределенные данные. Репликация

Репликация

- Репликация (replication) означает хранение копий одних и тех же данных на нескольких машинах, соединенных с помощью сети.

Причины репликации данных

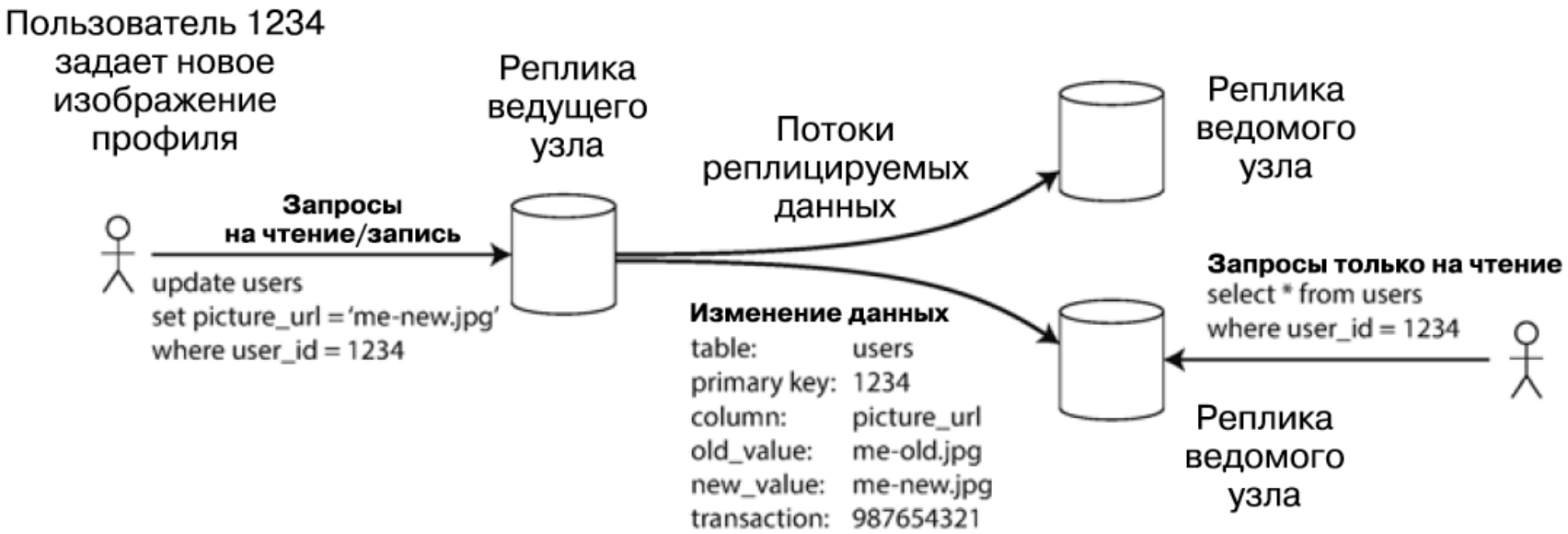
- ради хранения данных **географически близко** к пользователям
- чтобы система могла продолжать работать **при отказе некоторых ее частей**
- для **горизонтального масштабирования** количества машин, обслуживающих запросы на чтение

Алгоритмы репликации изменений

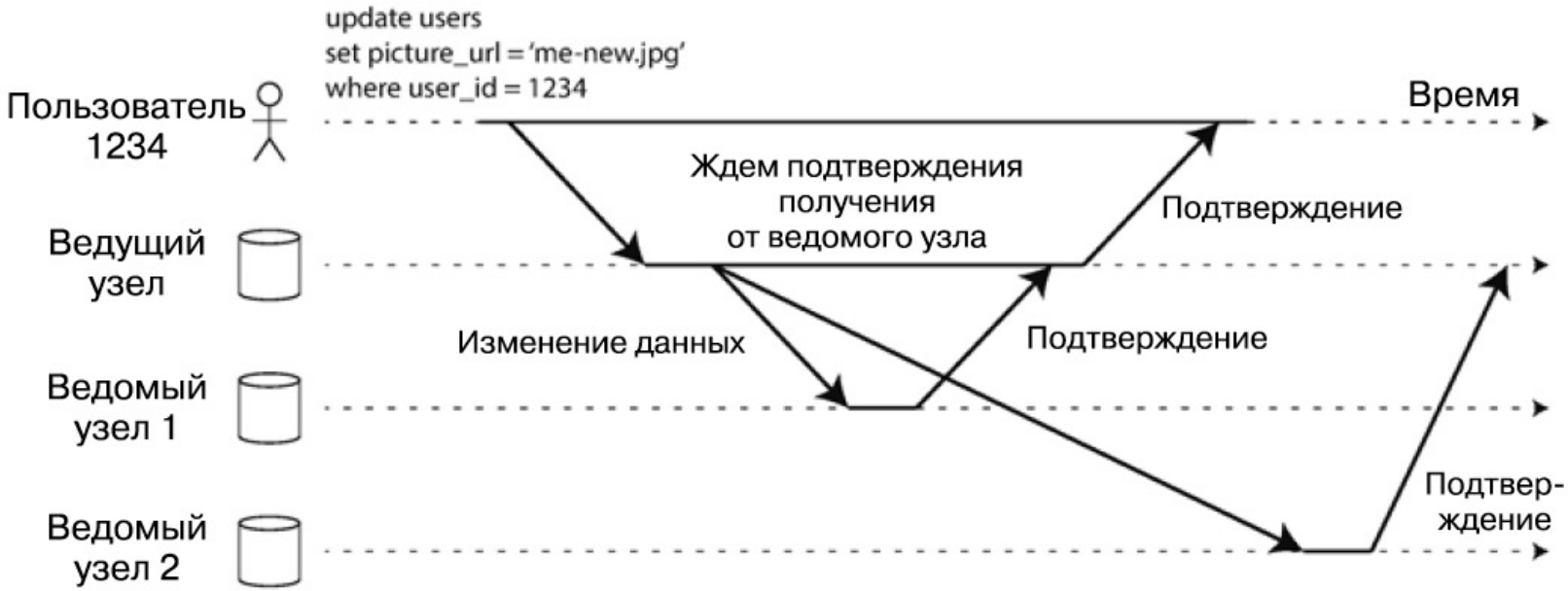
- репликацию с одним ведущим узлом (**single-leader**)
- с несколькими ведущими узлами (**multi-leader**)
- без ведущего узла (**leaderless**)

Репликация с одним ведущим узлом

Репликация с одним ведущим узлом



Синхронная и асинхронная репликация



Создание новых ведомых узлов

1. Сделать согласованный снимок состояния БД ведущего узла на определенный момент времени
2. Скопировать снимок состояния на новый ведомый узел.
3. Ведомый узел подключается к ведущему и запрашивает все изменения данных, произошедшие с момента создания снимка.
4. Когда ведомый узел завершил обработку изменений данных, произошедших с момента снимка состояния, говорят, что он наверстал упущенное

Перебои в обслуживании узлов

- **Отказ ведомого узла**: наверстывающее восстановление
- **Отказ ведущего узла**: восстановление после отказа:
  1. установить отказ ведущего узла
  2. выбрать новый ведущий узел
  3. настроить систему на использование нового ведущего узла

Реализация журналов репликации

- Операторная репликация
- Перенос журнала упреждающей записи (WAL)
- Логическая (построчная) журнальная репликация
- Триггерная репликация

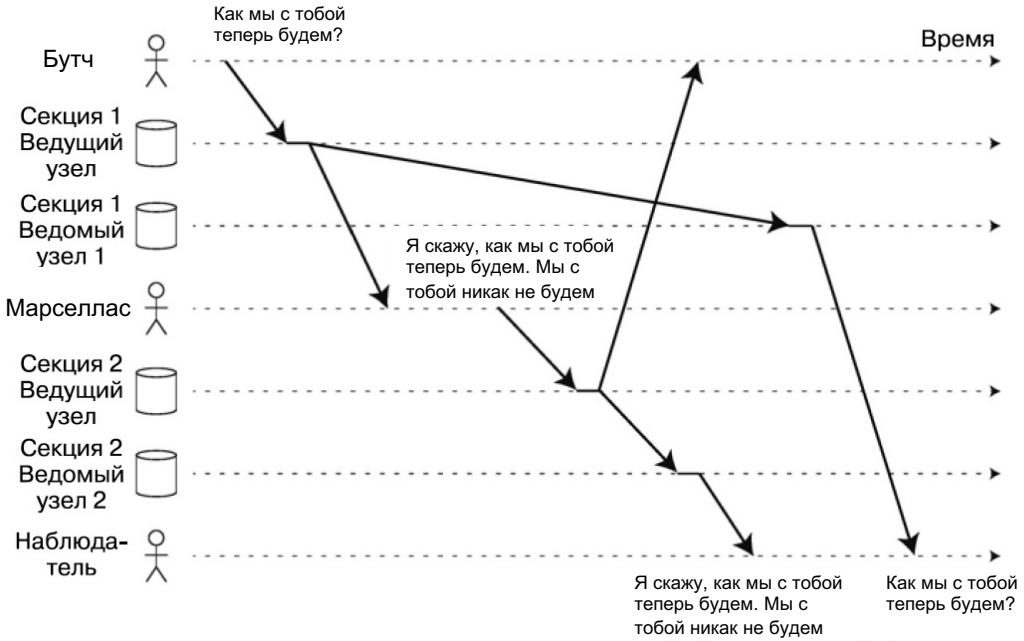
Проблемы задержки репликации

- Чтение своих же записей
- Монотонные чтения
- Согласованное префиксное чтение

Читаем свои же записи

Монотонные чтения

Согласованное префиксное чтение

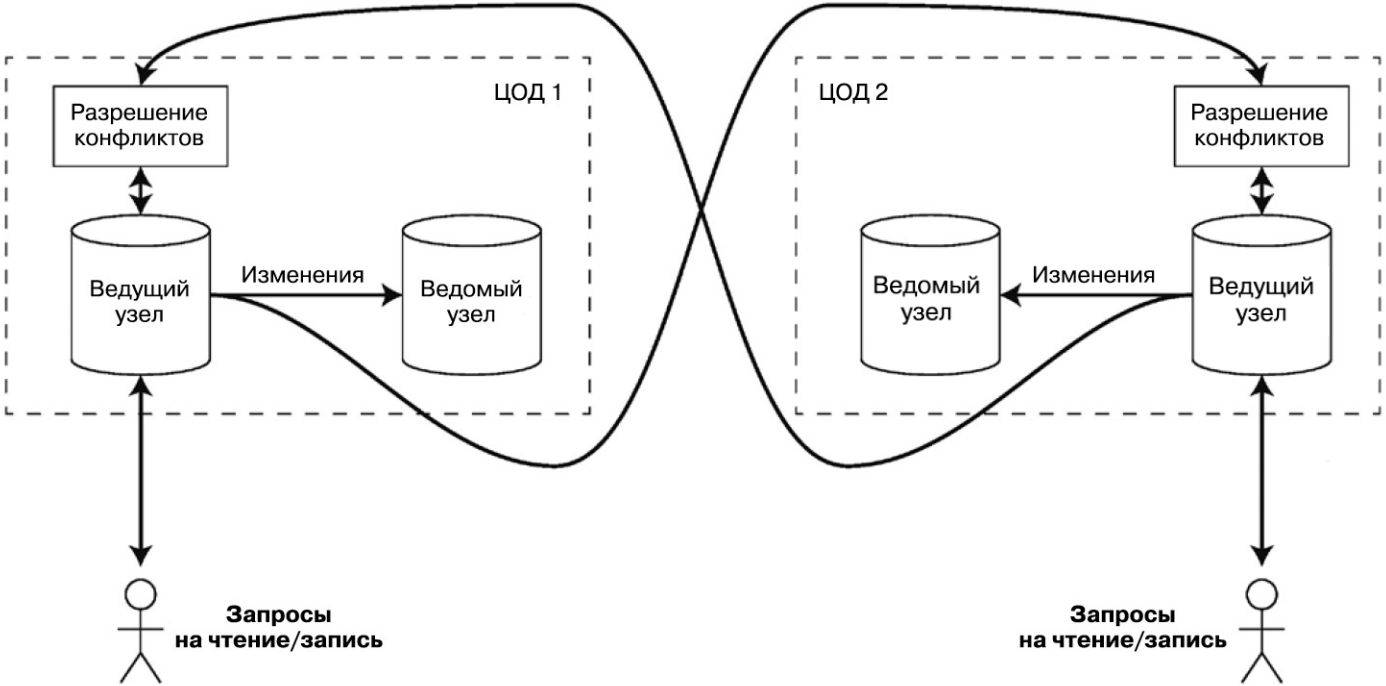


Репликация с несколькими ведущими узлами

Репликация с несколькими ведущими узлами

- разрешение приема запросов на запись более чем одному узлу
- multi-leader replication / master — master replication / active/active replication
- каждый из ведущих узлов одновременно является ведомым для других ведущих.

Эксплуатация с несколькими ЦОДами



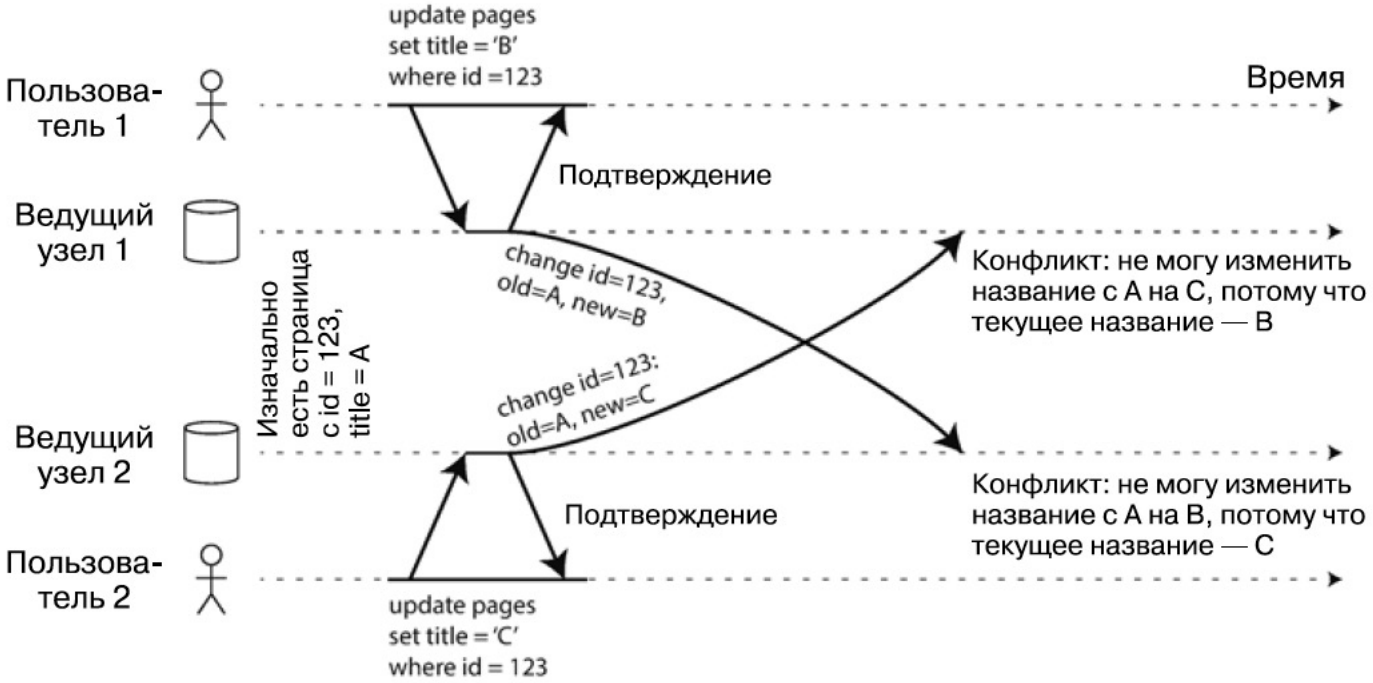
Офлайн-клиенты

- у каждого устройства есть своя локальная база данных, служащая ведущим узлом (она принимает запросы на запись)
- по сути, то же самое, что и репликация с несколькими ведущими узлами между ЦОДами
- для подобного режима эксплуатации создана СУБД CouchDB

Совместное редактирование

- предоставляют возможность нескольким людям редактировать документ одновременно
- прежде чем пользователь сможет отредактировать документ, запросить на этот документ блокировку
- Такая модель совместной работы эквивалентна репликации с одним ведущим узлом и выполнением транзакций на ведущем узле

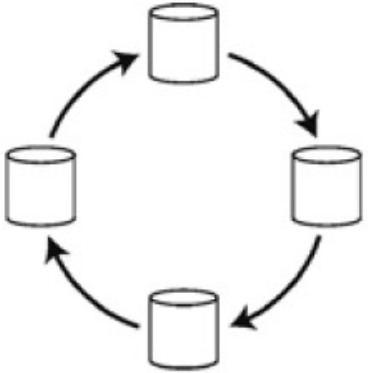
Обработка конфликтов записи



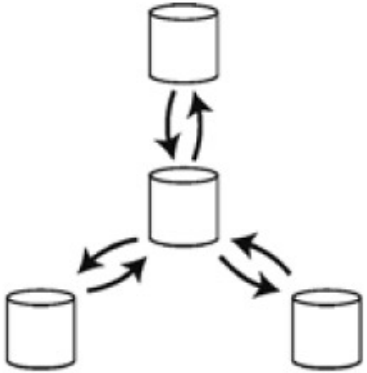
Способы конвергентного разрешения конфликтов

- Присвоить каждой операции записи уникальный идентификатор, после чего просто выбрать операцию («победителя»)
- Присвоить уникальный идентификатор каждой реплике и считать, что у исходящих от реплик с большим номером операций записи есть приоритет перед теми, которые исходят от реплик с меньшим.
- Каким-либо образом слить значения воедино, например, выстроить их в алфавитном порядке, после чего выполнить их конкатенацию
- Написать код приложения, который бы разрешал конфликты позднее

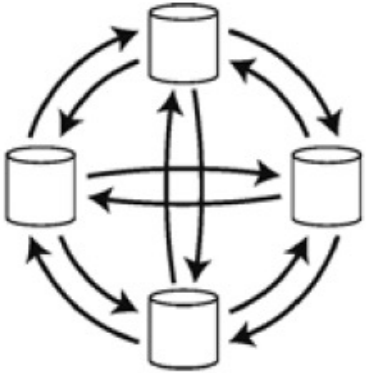
Топологии репликации с несколькими ведущими узлами



А. Топология типа «кольцо»

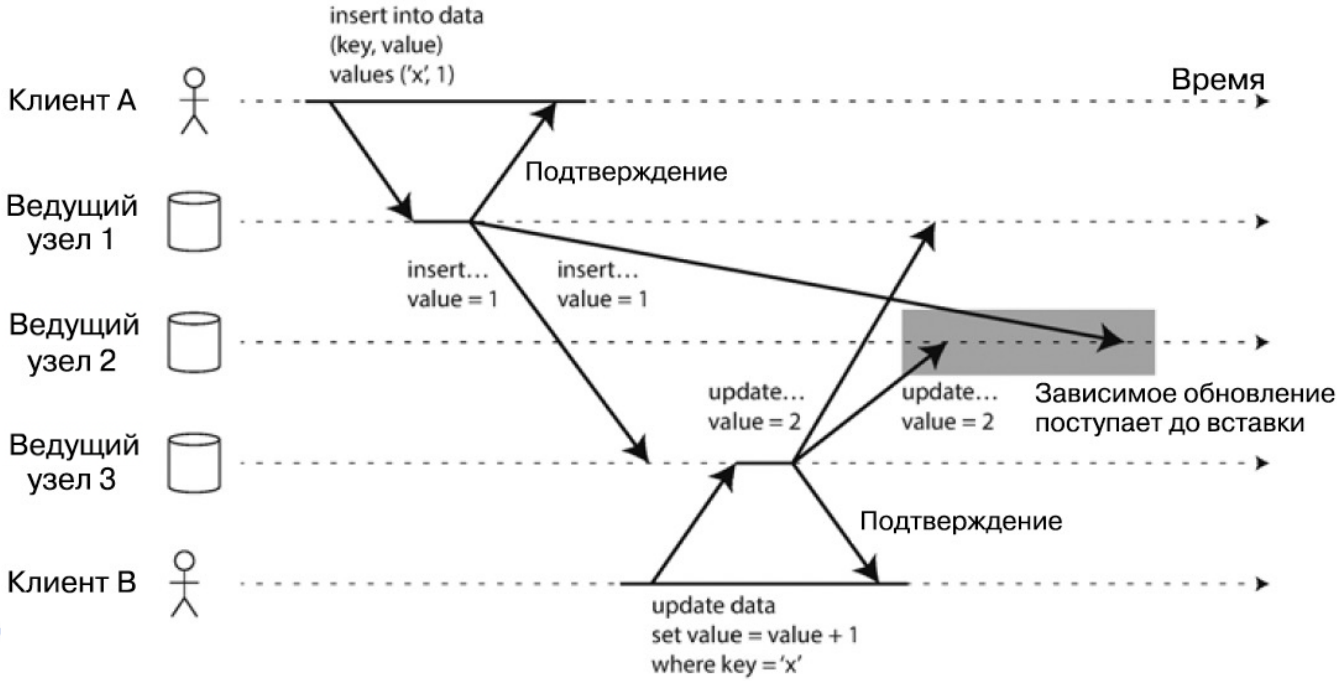


Б. Топология типа «звезда»



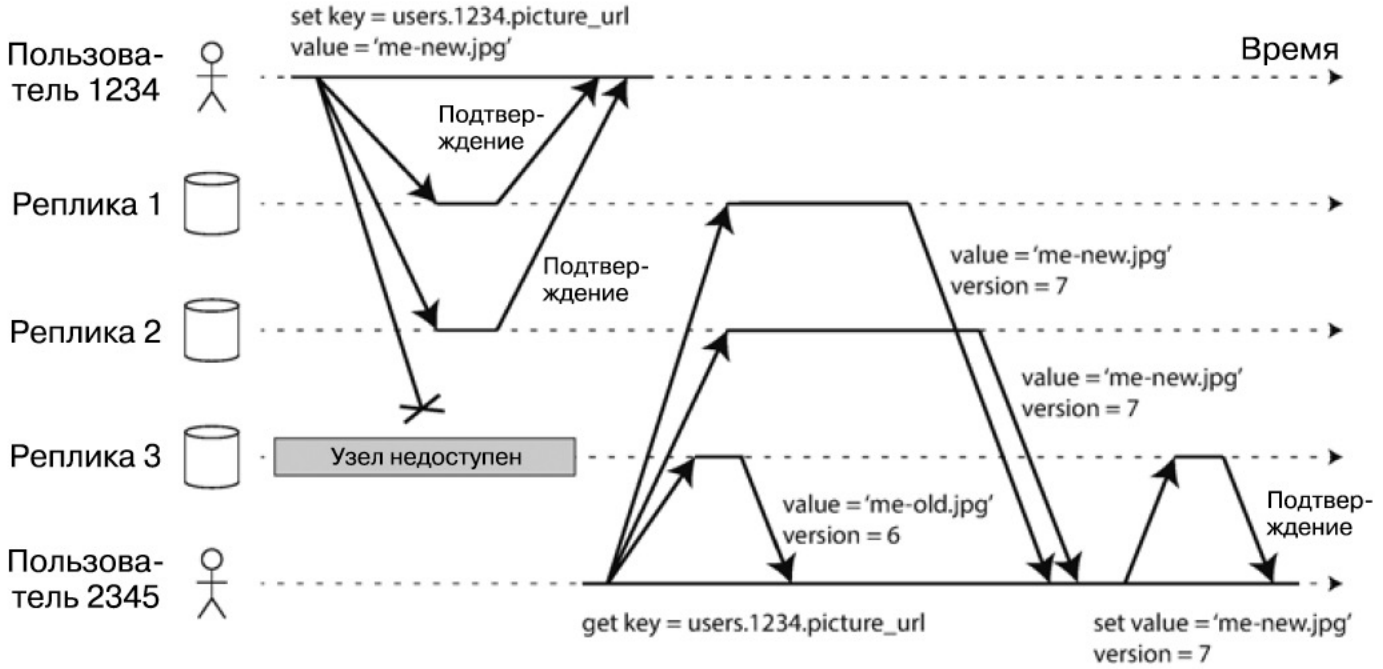
В. Топология типа «каждый с каждым»

Replication overtakes



Репликация без ведущего узла

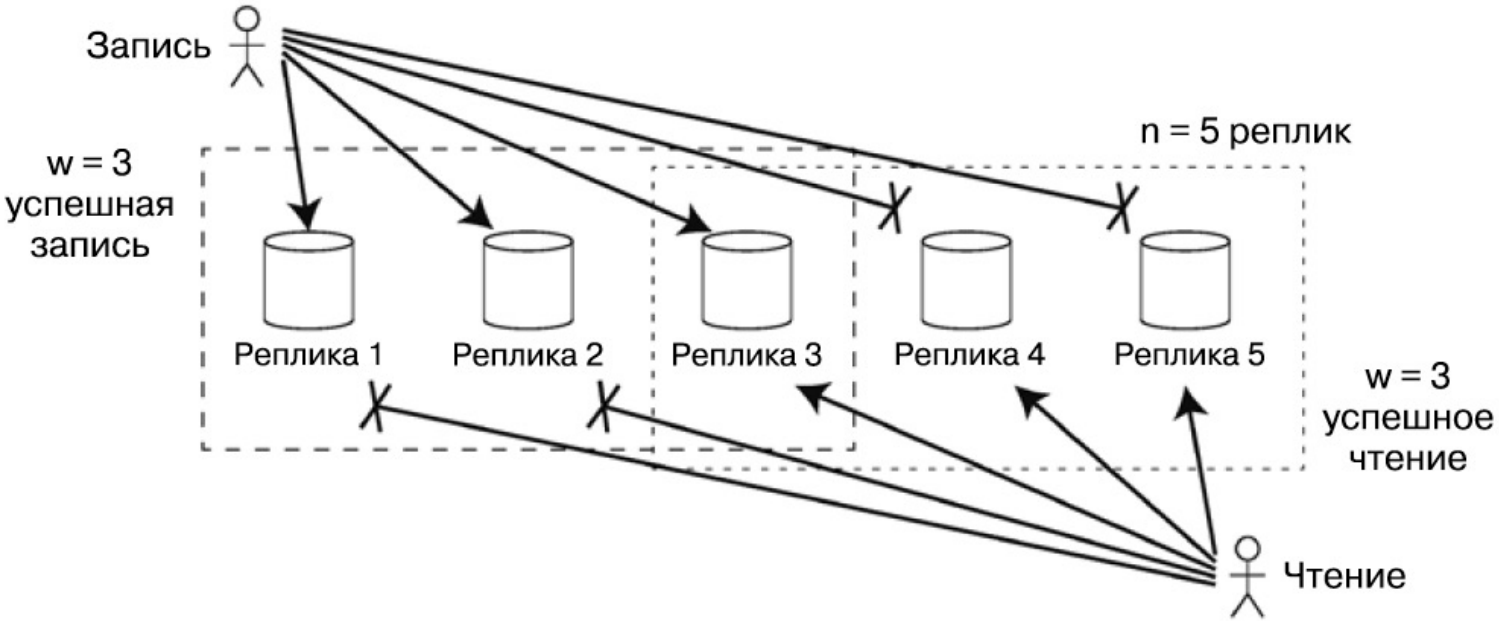
Запись в базу данных при отказе одного из узлов



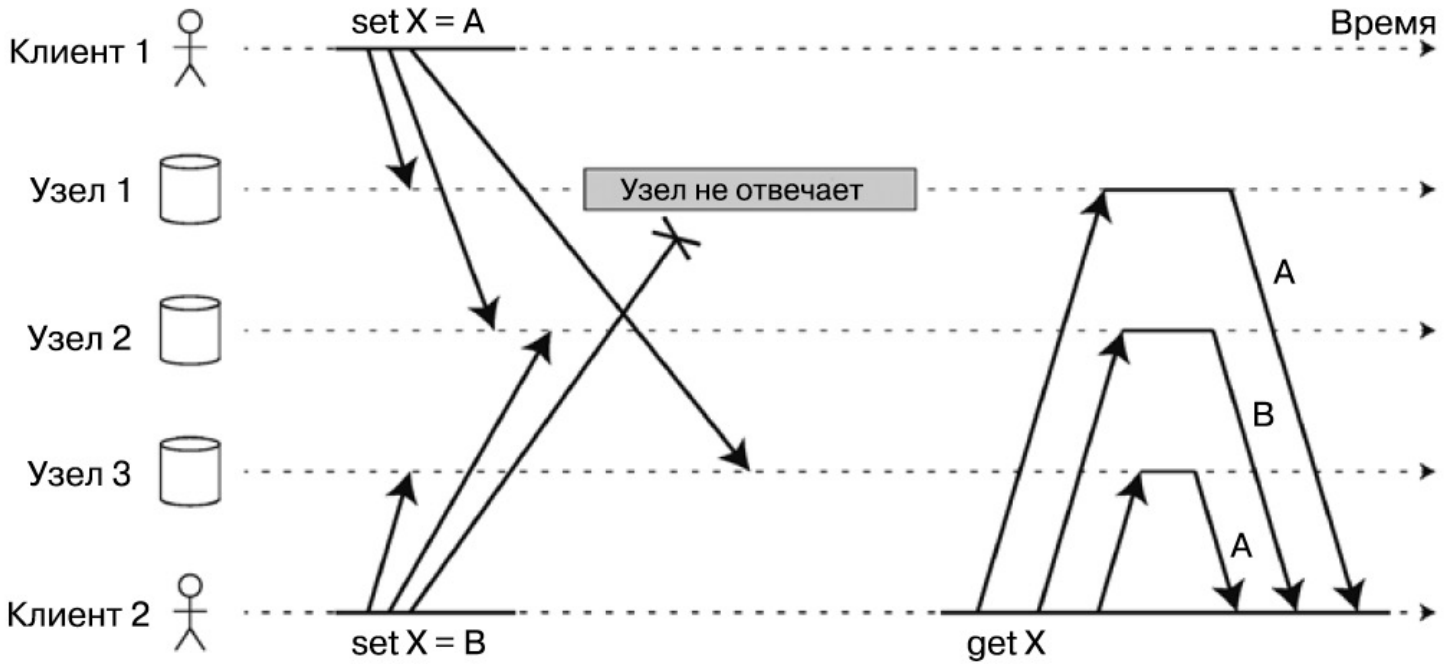
Операции записи и чтения по кворуму

- при наличии **n** реплик операция записи, чтобы считаться успешной, должна быть подтверждена **w** узлами, причем мы должны опросить как минимум **r** узлов для каждой операции. Если **w + r > n**, то можно ожидать: полученное при чтении значение будет актуальным, поскольку хотя бы один из **r** узлов, из которых мы читаем, должен оказаться актуальным.

Устойчивость к недоступности узлов



Обнаружение конкурентных операций записи



Спасибо за внимание!

www.ifmo.ru

ITMOre than a UNIVERSITY