HL++ 2022 HighLoad++ FOUNDATION

Трек Яндекс

Миллион RPS в YDB: история одного переезда Метрики

Александр Прудаев, разработчик

Яндекс Метрика

- Самый популярный сервис веб-аналитики в России
- Третий по популярности сервис веб-аналитики в мире — по данным W3Techs.com

		Usage, %	Change since 1 December 2021, %	Market share, %	Change since 1 December 2021, %
1	Google Analytics	56,8	+0,1	86,5	+0,2
2	Facebook Pixel	11,4	+0,1	17,3	+0,1
3	Yandex Metrica	6,4	−0,1	9,8	−0,1
4	WordPress Jetpack	5,1		7,8	
5	Hotjar	3,7	+0,1	5,6	+0,1

w3techs.com

Яндекс Метрика

>1,6 млн

запросов в секунду

>100 млрд

событий в сутки

Визит

Visit 1

Visit 2

Движок сейчас

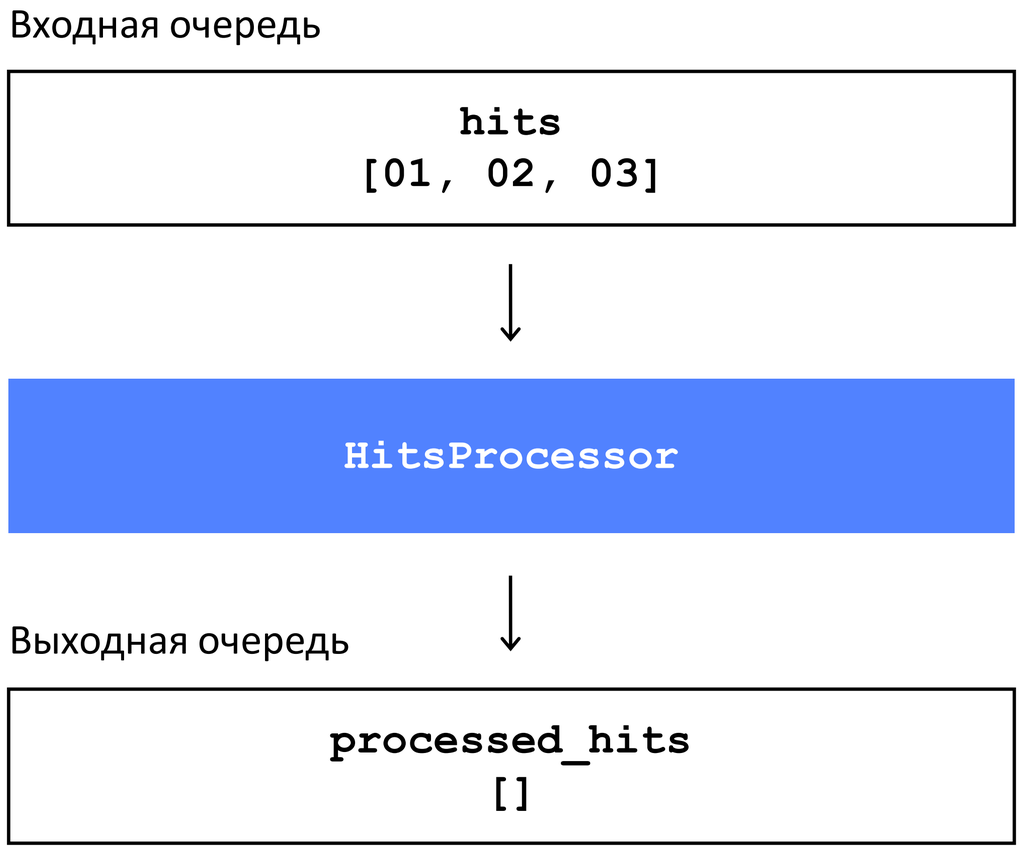
Переживает потерю одного дата-центра

Обеспечивает exactly-once

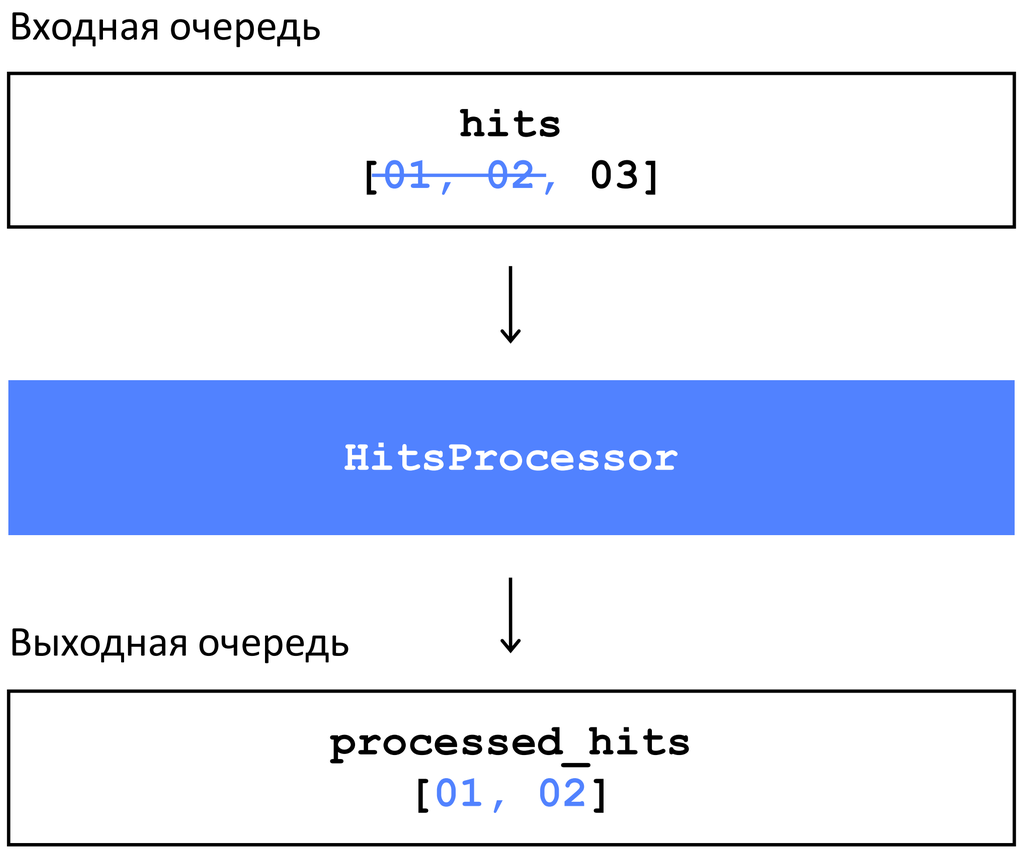
1,6 млн

событий в секунду

Очереди в ZooKeeper

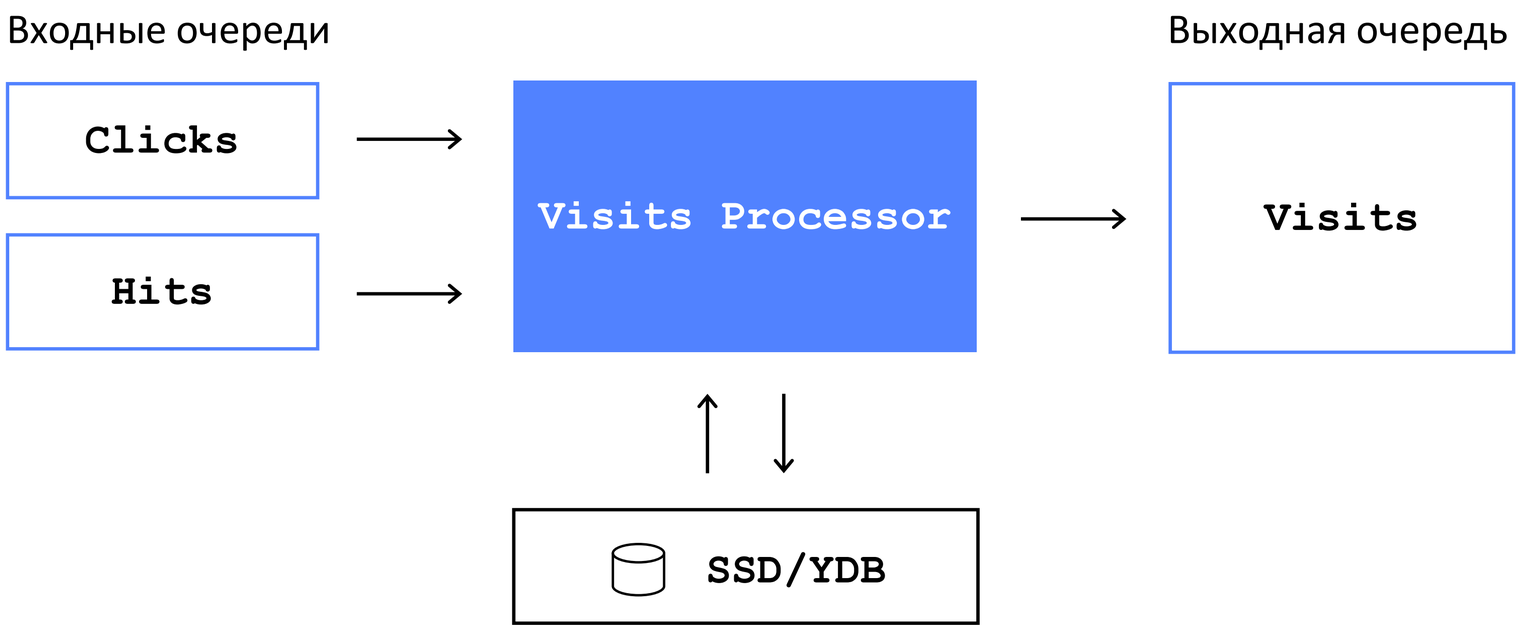


Очереди в ZooKeeper



Очереди в ZooKeeper

Сервис сборки визитов



Сервис сборки визитов раньше

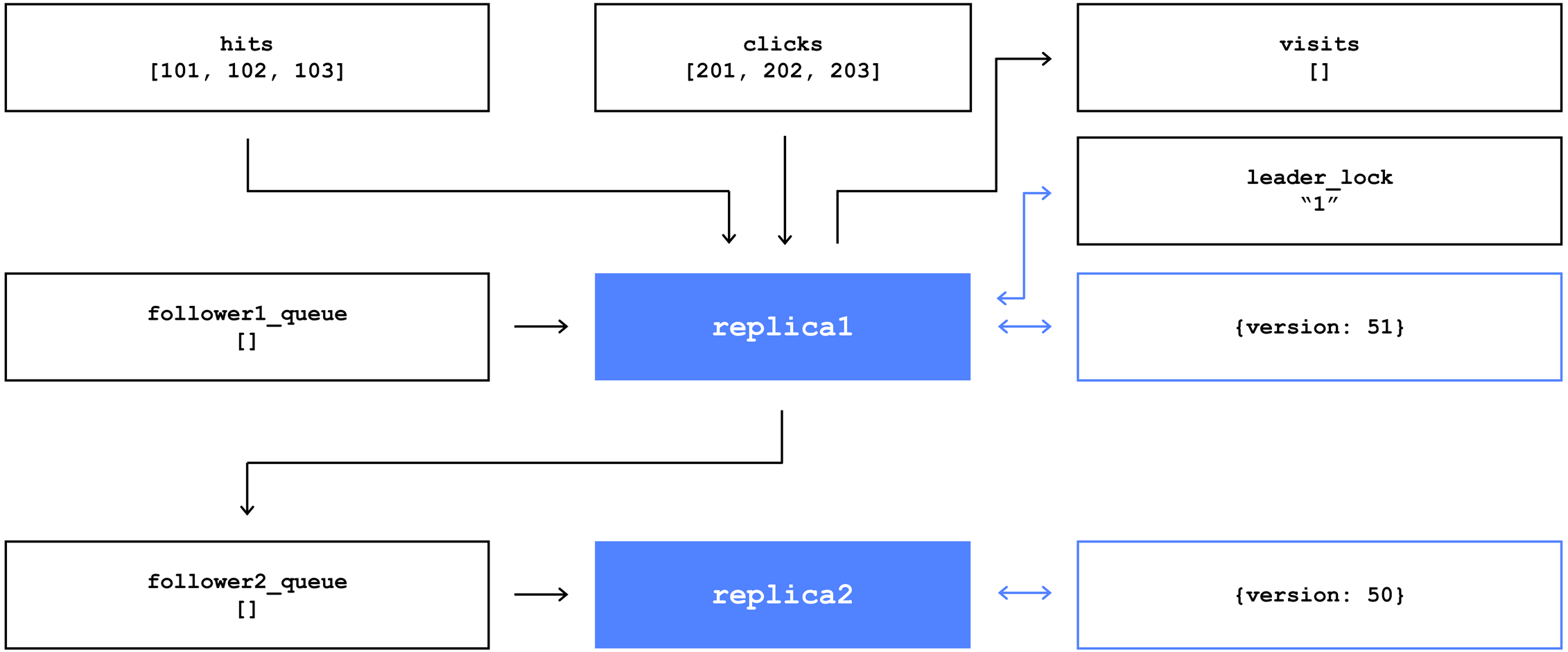
Своя проприетарная БД

~100 серверов с SSD

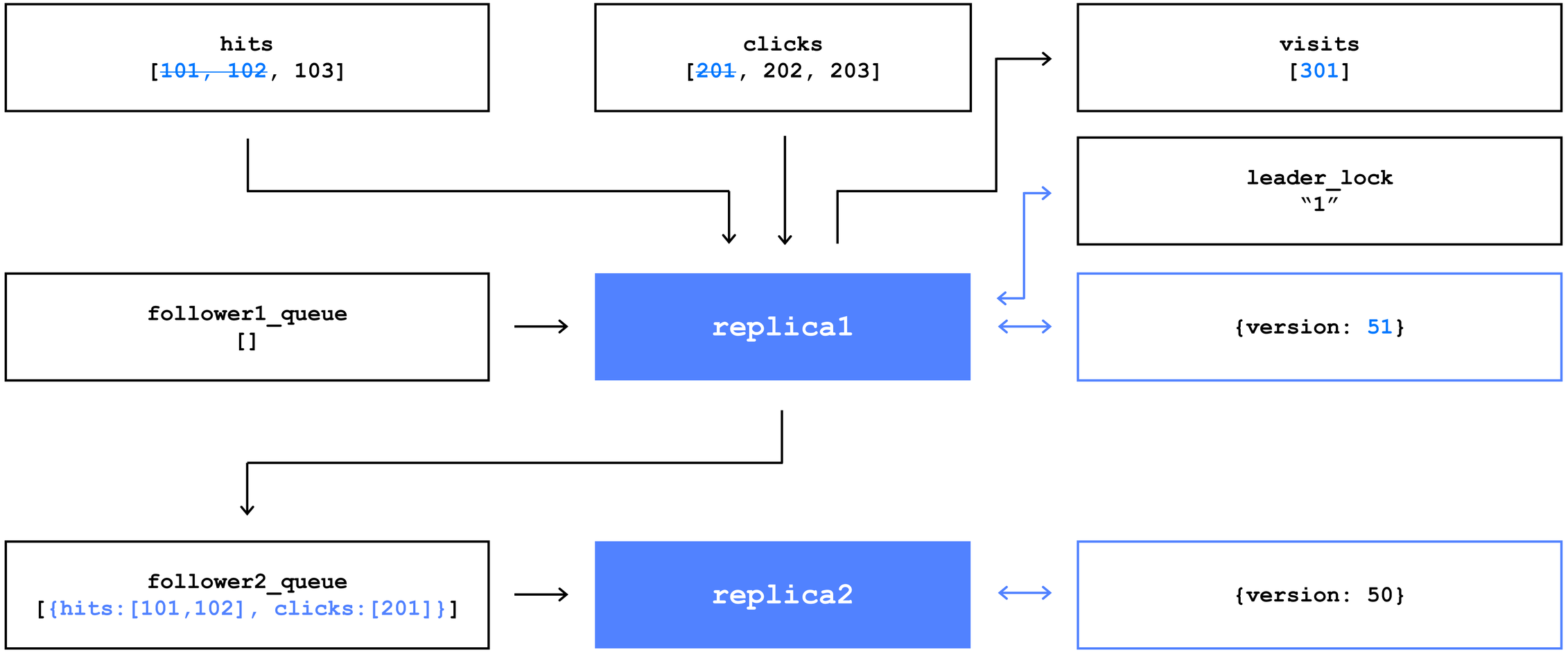
Leader-follower-репликация
с зеркалированием потоков

Шардирование по счётчикам

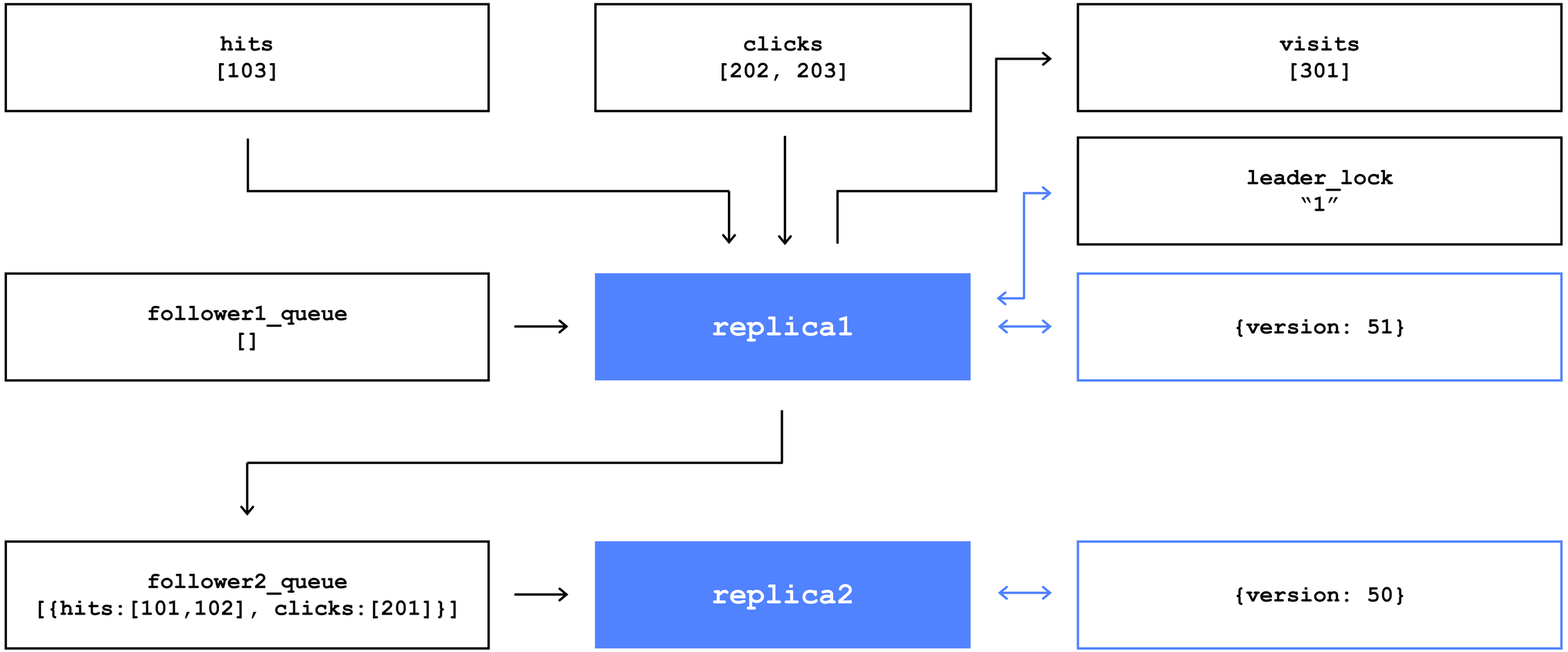
Репликация и мультиверсионность БД



Репликация и мультиверсионность БД



Репликация и мультиверсионность БД



Проблемы

Плохая масштабируемость и шардирование

Система репликации

Своя БД

Что хотели иметь после

Отказ
от сложного
механизма
репликации

Stateless —
вычислительные
ноды в облаке

Стейт
в облачном
хранилище

Возможность
быстро расти

Требования к базе данных

- Синхронная репликация leader-leader
- Отказоустойчивость, zero downtime, безостановочные обновления
- Бесконечное масштабирование

Реальная нагрузка, близкая к оценке

400 тыс.
запросов в секунду

80 ТБ
данных

5 ГБ/с
чтение

1 ГБ/с
запись

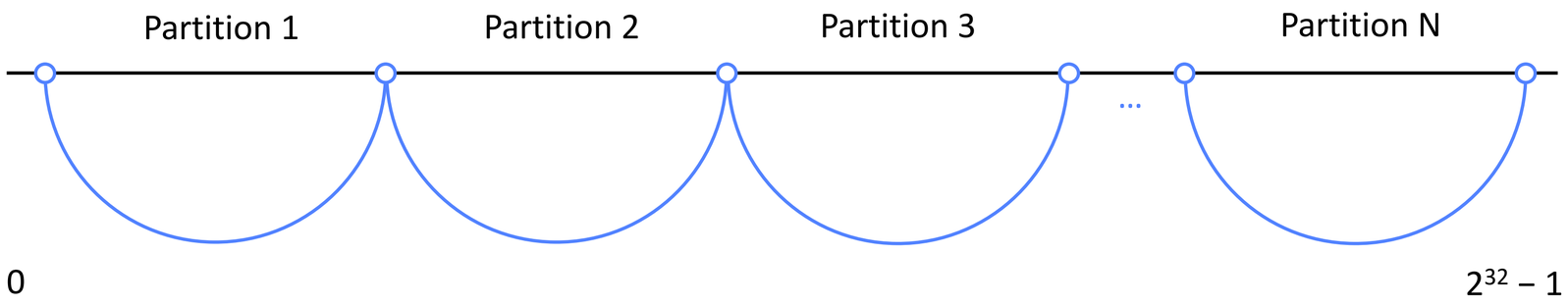
YDB

- ACID-транзакции
- Диалект SQL для запросов (YQL)
- Синхронная leader-leader-репликация
- Высокая доступность

Партиции и шардирование

Первичный ключ:

Hash(UserID, CounterID), UserID, CounterID, EventTime, EventID

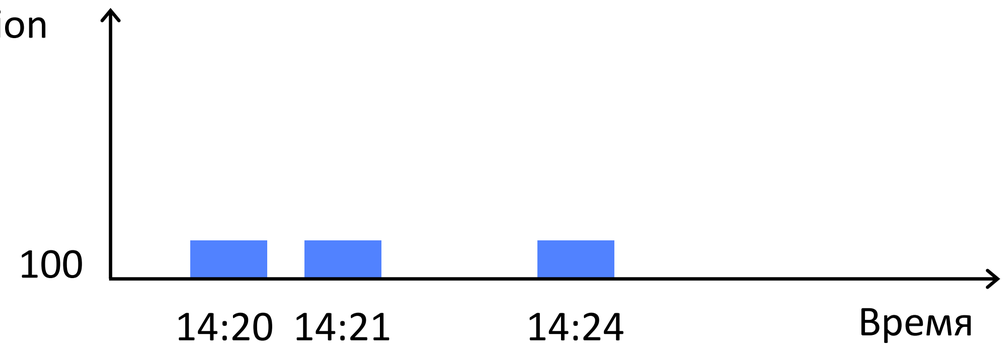


Первичный ключ и версия

Первичный ключ:

Hash(UserID, CounterID), UserID, CounterID, EventTime, EventID, Version

События в базе

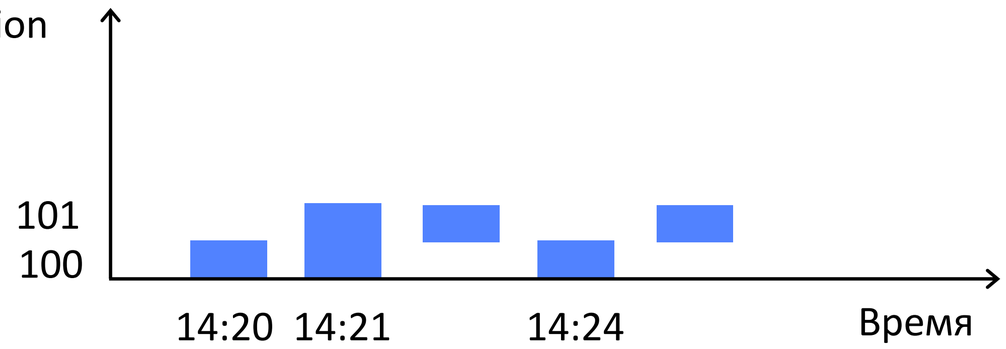


Первичный ключ и версия

Первичный ключ:

Hash(UserID, CounterID), UserID, CounterID, EventTime, EventID, Version

События в базе



Переезд: План А

1

Скопировать данные

2

Убедиться, что истории, читаемые
из обеих БД, совпадают

3

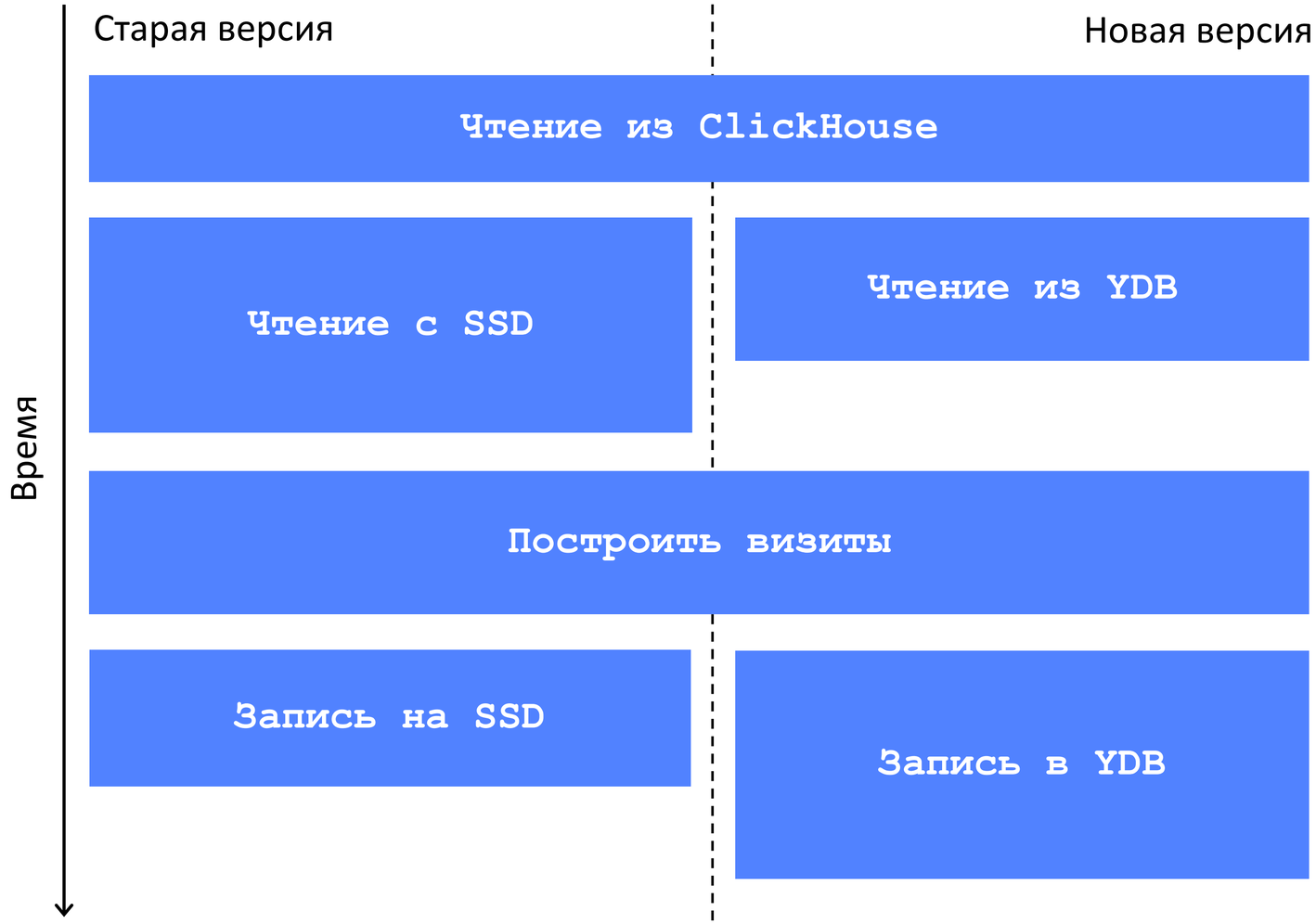
Отказаться
от локальной БД

Проблемы

Производительность

Данные отличаются

Время этапов итерации



Переезд: План Б

1

Скопировать данные

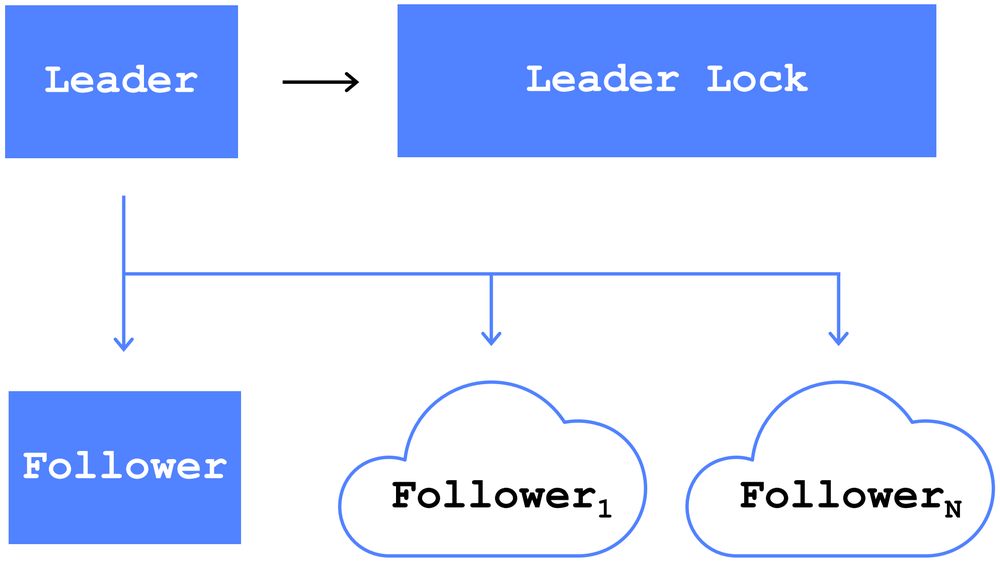
2

Обрабатывать зеркальный поток данных на YDB

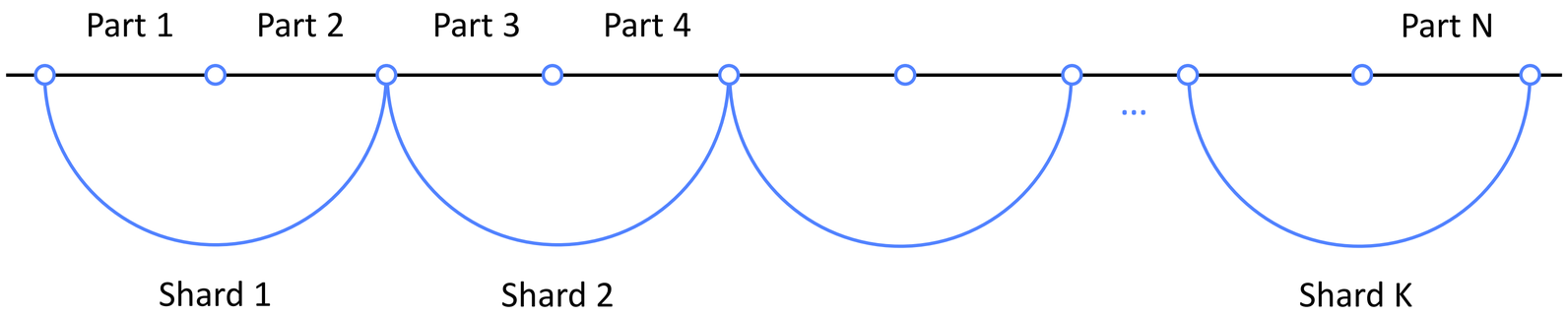
3

Сравнение выходных визитов железо vs облако

Репликация и облачные ноды



Шардирование



Итерация копирования

- Копирование данных
- Запуск новой системы на зеркалированном потоке входных данных
- Сравнение выходных данных
- Поиск причин расхождения
- Багфикс, разработка
- Повторить

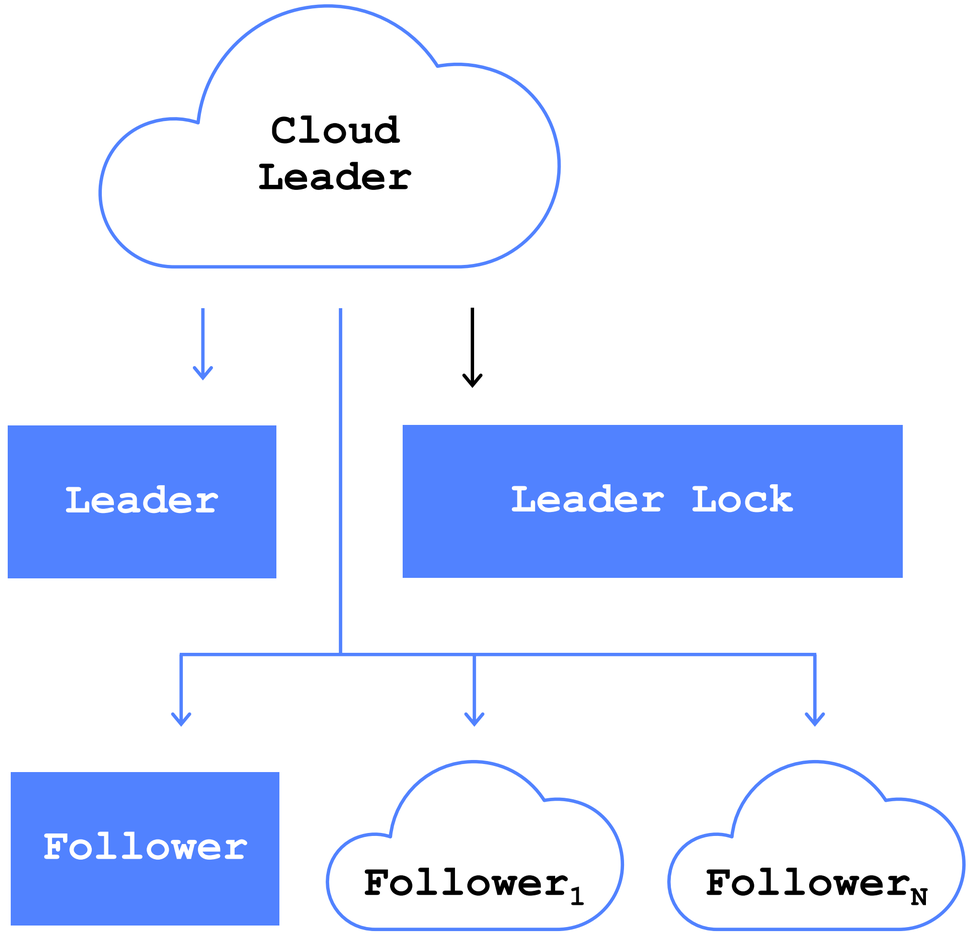
Проблемы

Ошибки копирования

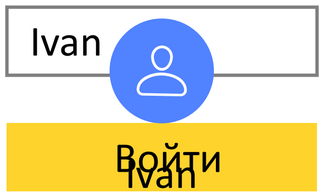
Ошибки перешардирования

Производительность

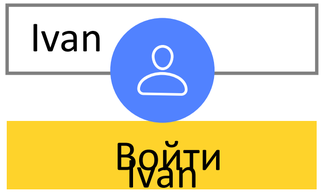
Облачный Leader



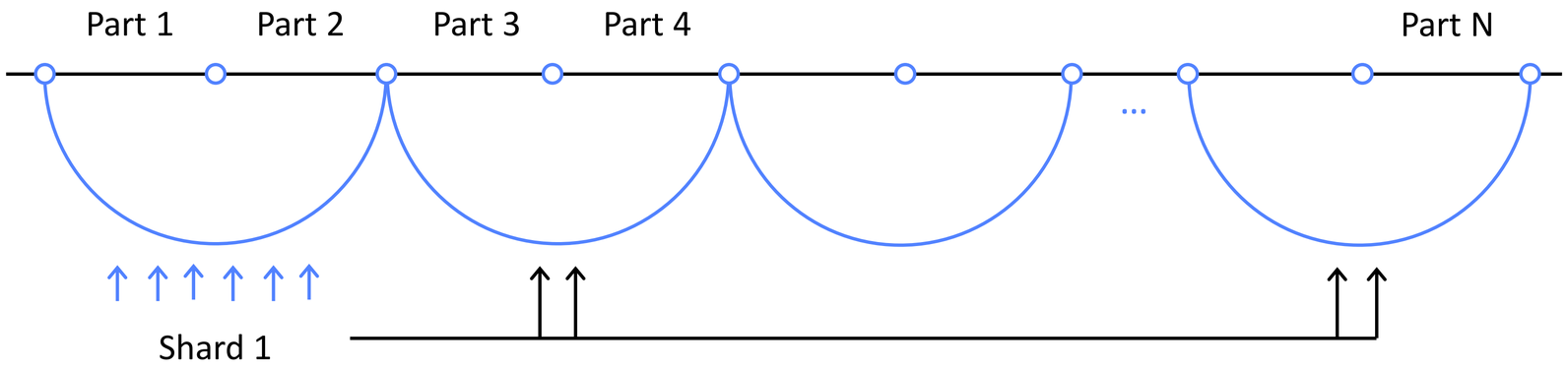
Кросс-девайс



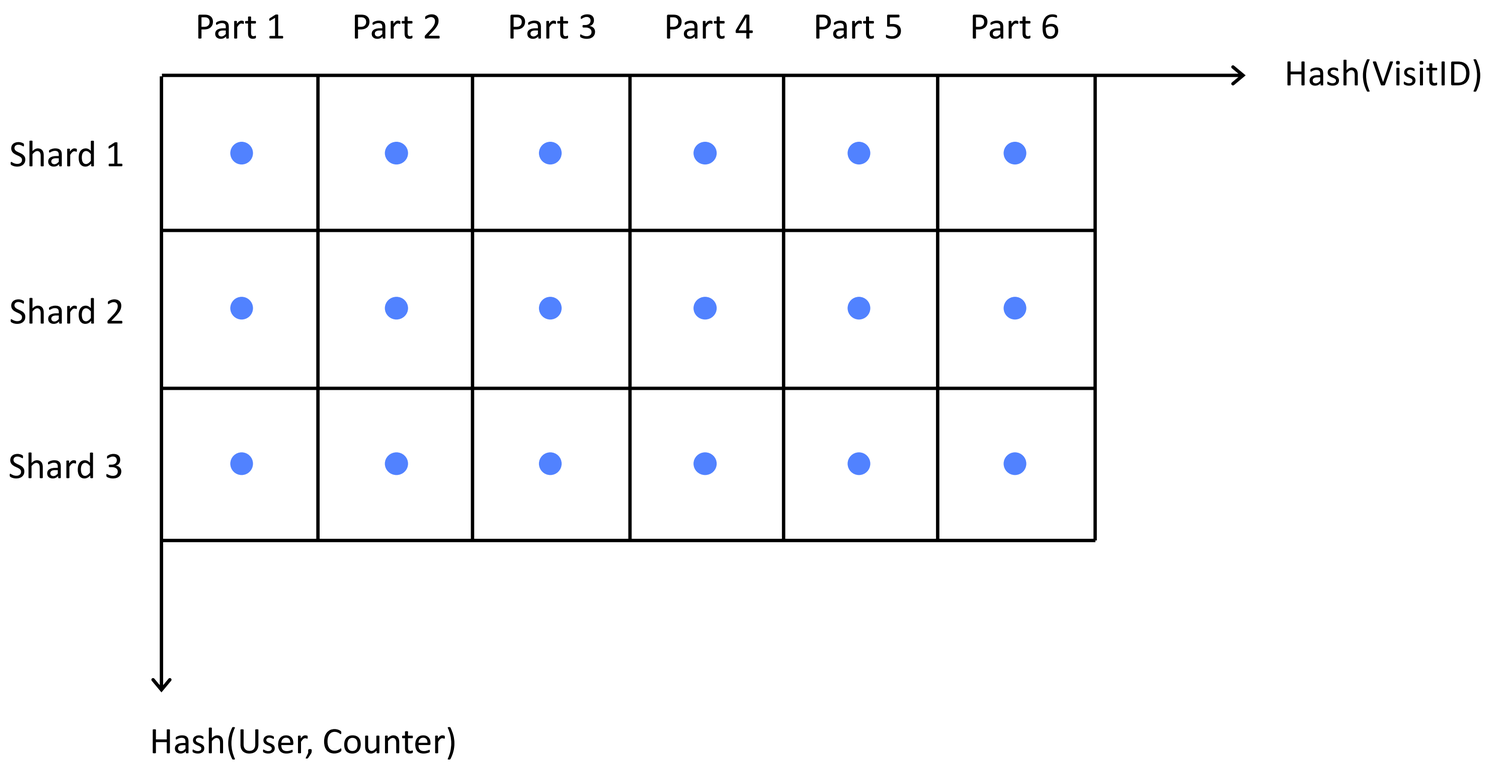
+



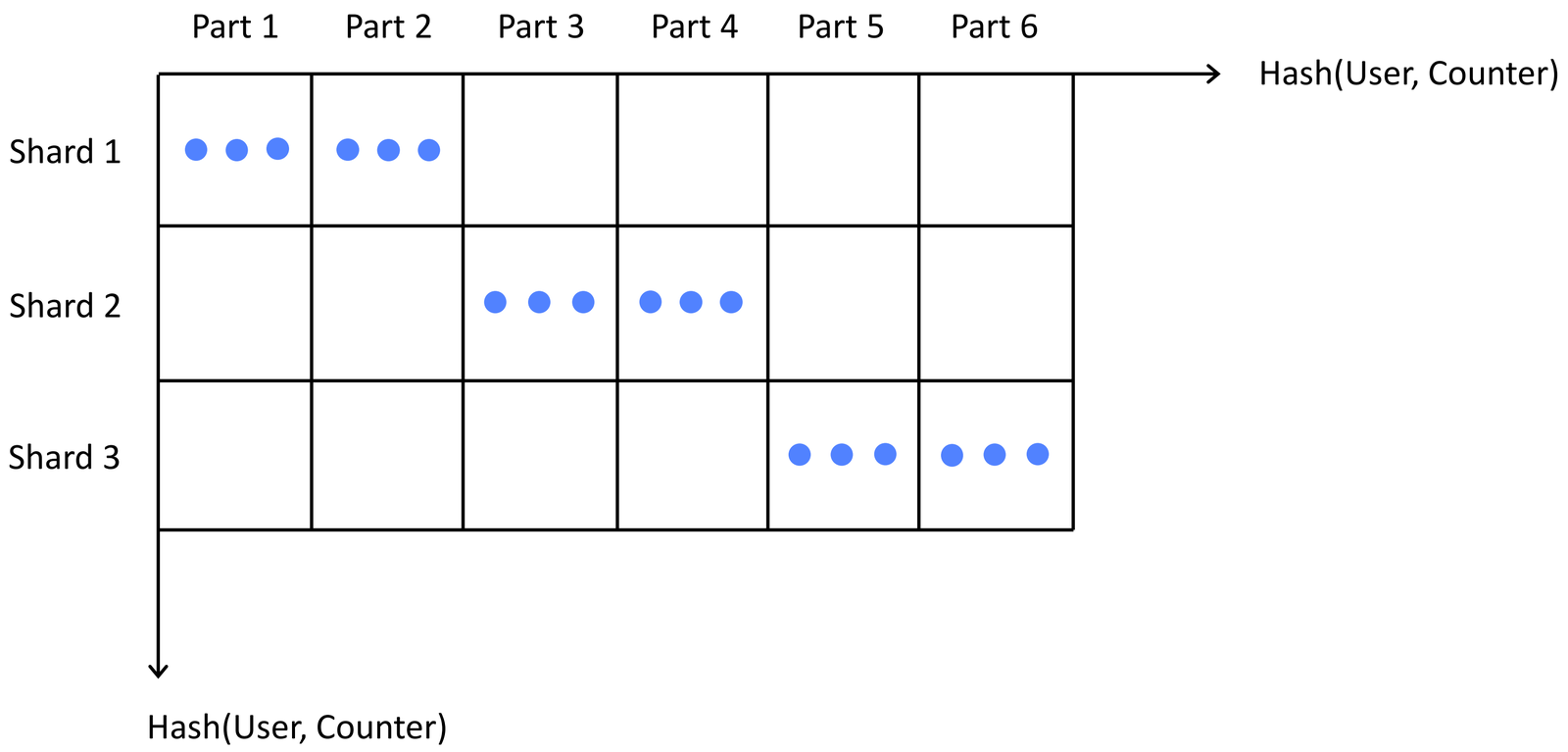
Кросс-девайс



Зависимость RPS от шардирования



Зависимость RPS от шардирования



YDB в Open source

— исходный код

— документация

— SDK

и все инструменты для работы
с базой опубликованы на GitHub под
лицензией Apache 2.0

github.com/ydb-platform/ydb


HL++
2022
HighLoad++
FOUNDATION
Трек Яндекс
Спасибо!
Александр Прудаев
Разработчик
aprudaev@yandex-team.ru
@aprudaev