



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Курсовая работа

ОПТИМИЗАЦИЯ ХРАНЕНИЯ ДАННЫХ В CLICKHOUSE ДЛЯ ГЕТЕРОГЕННЫХ СИСТЕМ

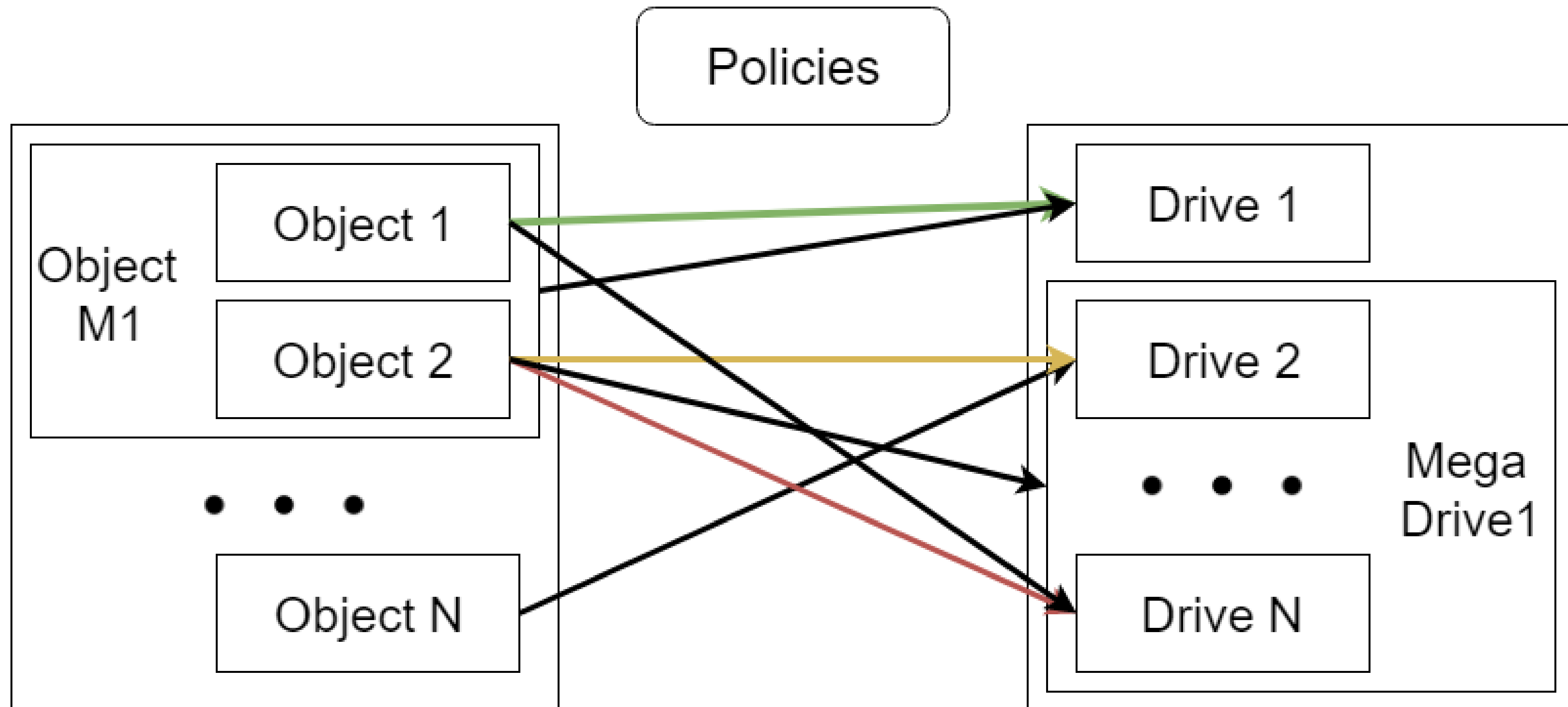
Минеев Игорь Евгеньевич, БПМИ166

Научный руководитель:

Руководитель группы разработки СУБД ClickHouse, Яндекс
Миловидов Алексей Николаевич

Москва, 2019

STORAGE POLICIES





АКТУАЛЬНОСТЬ

Активное развитие устройств долговременного хранения

Решение будет добавлено в open source DBMS ClickHouse

Отсутствуют достаточно функциональные альтернативные решения этой проблемы в рамках простого использования и развертывания СУБД ClickHouse

ЦЕЛИ И ЗАДАЧИ

Цель:

- Создать возможность гибкого использования систем хранения

Задачи:

- Создать прототип описания различных хранилищ («путей» хранения)
- Рассмотреть пути пассивного перемещения (в фоне)
- Предложить идеи для осуществления User Space RAID
- Написать тесты

ПОСТАНОВКА ЗАДАЧИ

Входные данные:

- Пути в UNIX системе
- Набор приоритетов и ограничений дисков, кусков, и т.д.
- Предполагаемый размер «куска» данных
- План хранилища

Выходные данные:

- Опциональны



СУЩЕСТВУЮЩИЕ РЕШЕНИЯ

Подходы хранения

- Nothing
- MDRAID
- Physical RAID controllers

Другие СУБД:

- Vertica
- Apache Druid

- 1) [Документация Vertica Storage Policies](#)
- 2) [Документация Druid Deep Storage](#)



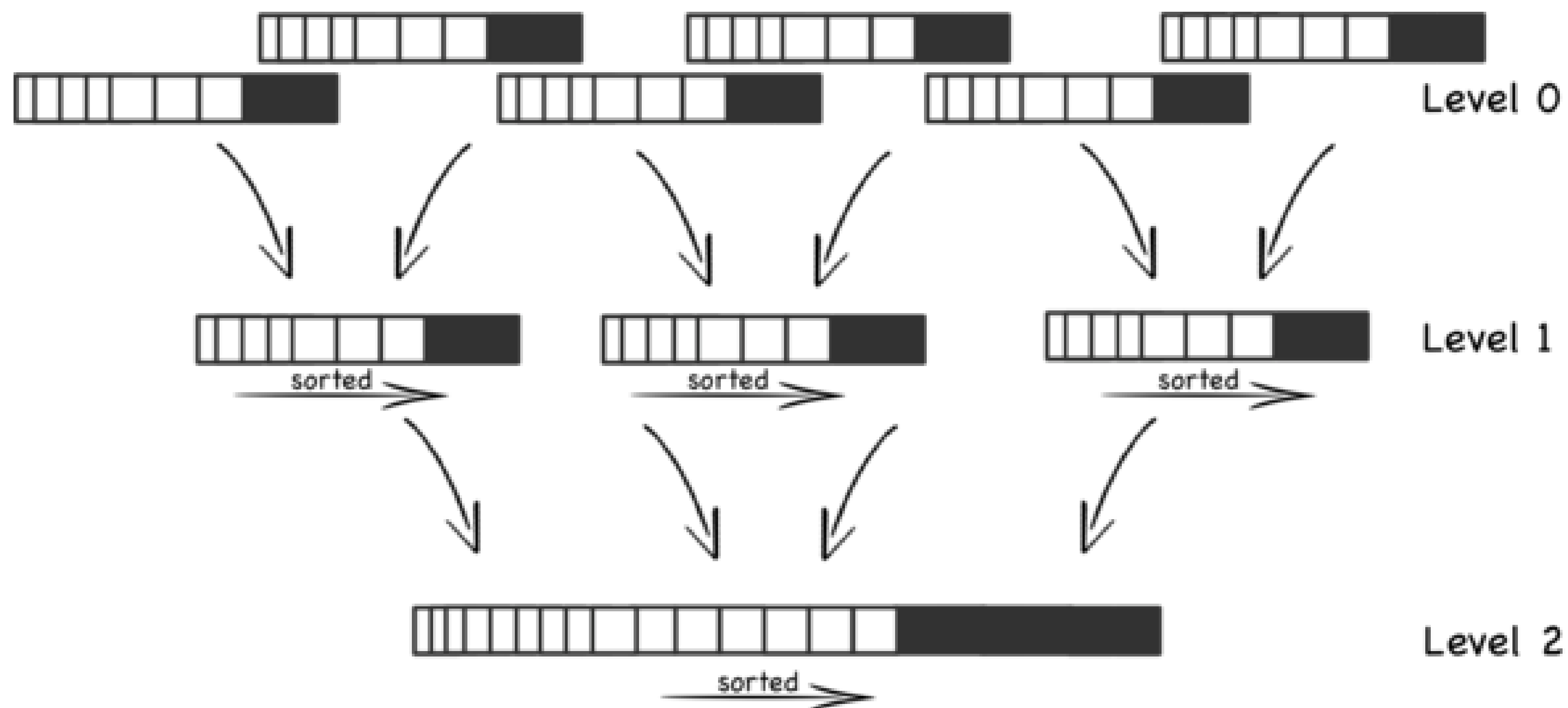
БАЗОВЫЙ АЛГОРИТМ

Предварительно выполняем деление дисков на «холодные» и «горячие»

Для «актуальных» данные используем «горячие» диски
Для прочих используем «холодные» диски

Вручную или автоматически решаем нехватку места
перетеканием данных между хранилищами

CLICKHOUSE MERGETREE



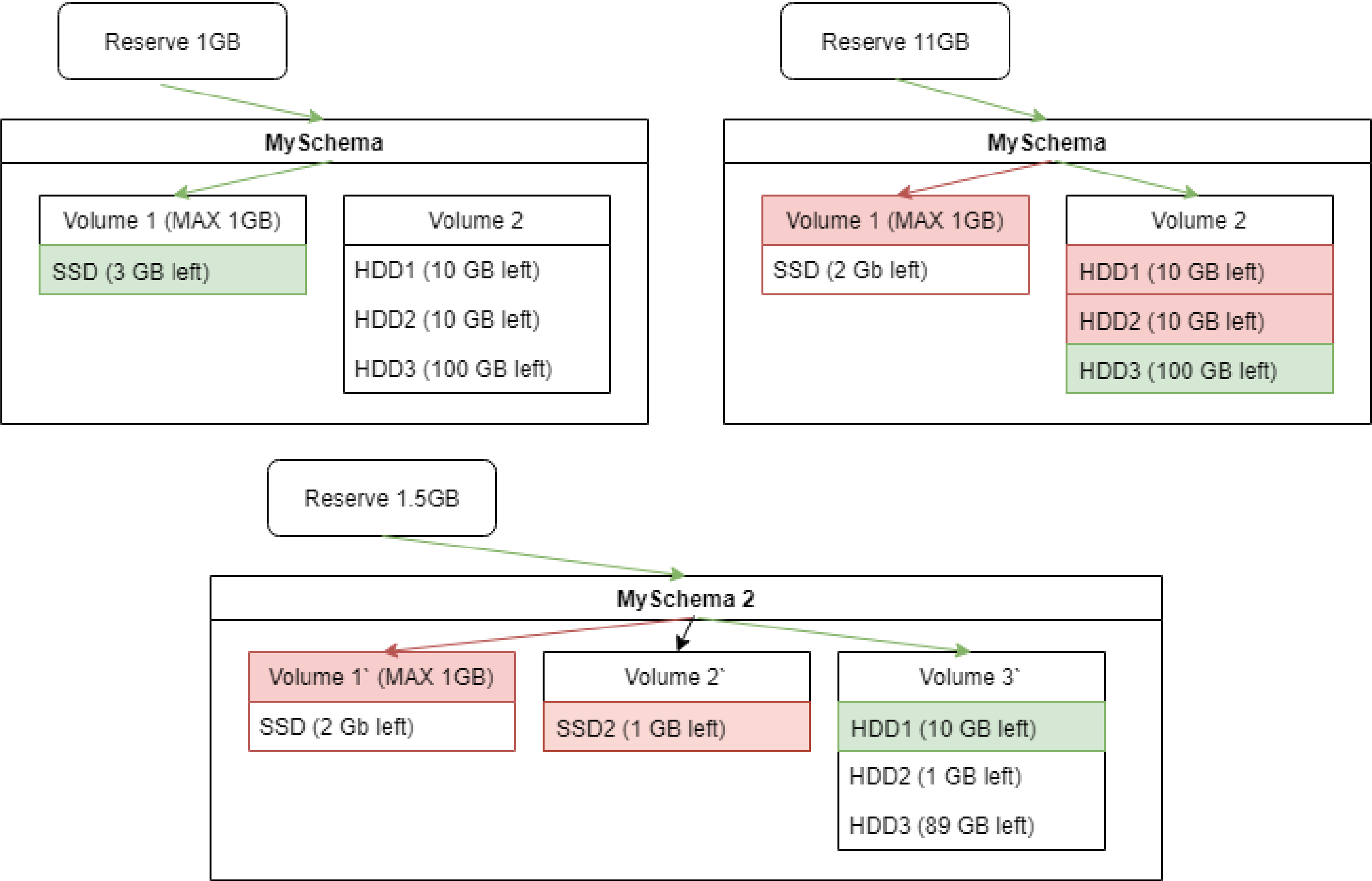
Compaction continues creating fewer, larger and larger files

СХЕМА ХРАНЕНИЯ

- Disk – путь системы хранения
- Volume – «неупорядоченный» набор из Disks
- Policy – упорядоченный набор из Volumes
- Одна Policy для таблицы

В базовом варианте:

- Ограничение на размер куска для Volume
- Ограничение на занятое пространство для Disk

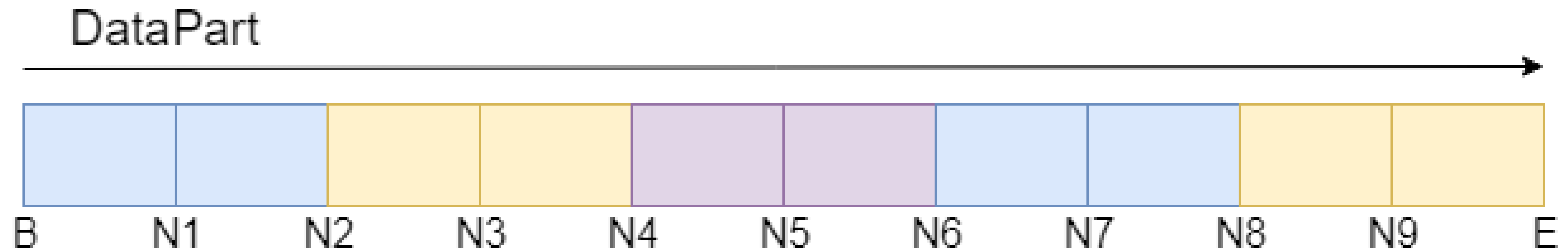


МЕЛОЧИ И МАСШТАБНОСТЬ

- 40 файлов и 2000 строк C++
- Совместимость протоколов передачи кусков
- Система резервации пространства
- Система перемещения кусков в фоне по заполненности
- Политики по умолчанию
- System Tables disks and policies
- Производительность не важна (*)
- Локальность Policy для серверов

RAID 0. STRIPE. И ДРУГИЕ...

- Удобное хранение данных
- Путь для дальнейшего развития
- Работоспособность из коробки



РЕЗУЛЬТАТЫ РАБОТЫ

- Гибкая настройка политик хранения
- Pull request ожидающий review и доработки тестов
- Удовлетворенность работой другими разработчиками и пользователями
- Задел для будущих доработок
- Опыт написания промышленного кода в «большом» проекте



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ