МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМЕНИ Н.Э. БАУМАНА
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)

Метод блочного хранения данных с возможностью доказательства неправомерного доступа на основе хеш-сумм

Студент: Пересторонин Павел Геннадьевич

Группа: ИУ7-83Б

Руководитель: Григорьев Александр Сергеевич

Актуальность

- Защита информации приобретает все большее значение с ростом объема информации в электронном виде.
- СУБД ClickHouse, на основе которой реализуется метод, имеет большое распространение.
- Движок MergeTree СУБД ClickHouse уже имеет функционал, обеспечивающий некоторые виды защиты от неправомерного доступа, который можно улучшить.

Цель и задачи

Цель – разработать метод блочного хранения данных с возможностью доказательства неправомерного доступа с помощью хеш-сумм.

Задачи:

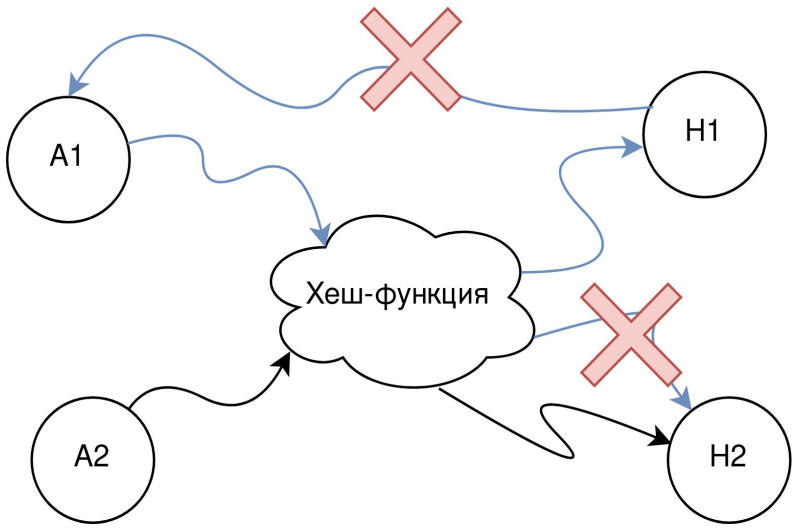
- описать виды защиты информации, классифицировать их;
- рассмотреть базовые элементы и понятия, используемые при проектировании методов хранения информации с возможностью защиты от неправомерного доступа;
- провести анализ существующих методов хранения информации с защитой от неправомерного доступа;
- провести анализ блочного хранения данных в системе, в которой планируется использование метода, на предмет защиты информации от неправомерного доступа;
- спроектировать и реализовать метод блочного хранения данных с возмозможностью доказательства неправомерного доступа;
- исследовать метод на предмет невозможности реализации угроз при различных конфигурациях системы.

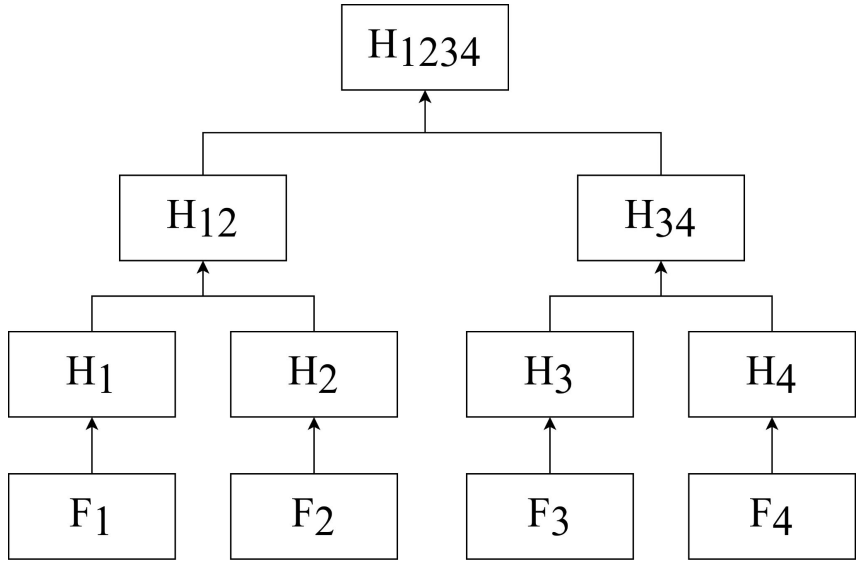
Виды защиты от неправомерного доступа

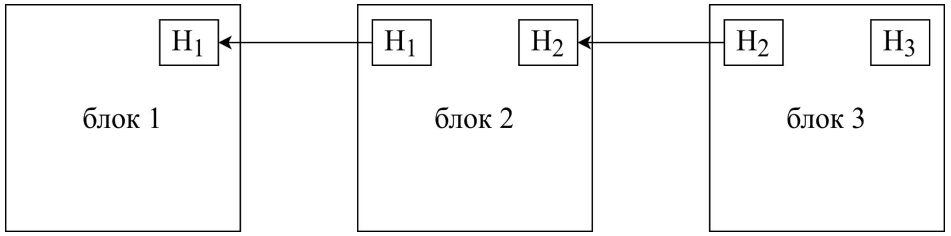
- Исключение:
  - удаления блока;
  - чтения;
  - изменения.
- Возможность восстановления:
  - частичного удаления блока;
  - частичного изменения.
- Возможность доказательства:
  - удаления блока;
  - изменения блока.

Базовые понятия

- Хеш-функция;
- Блокчейн;
- Ациклический направленный граф и дерево Меркла.

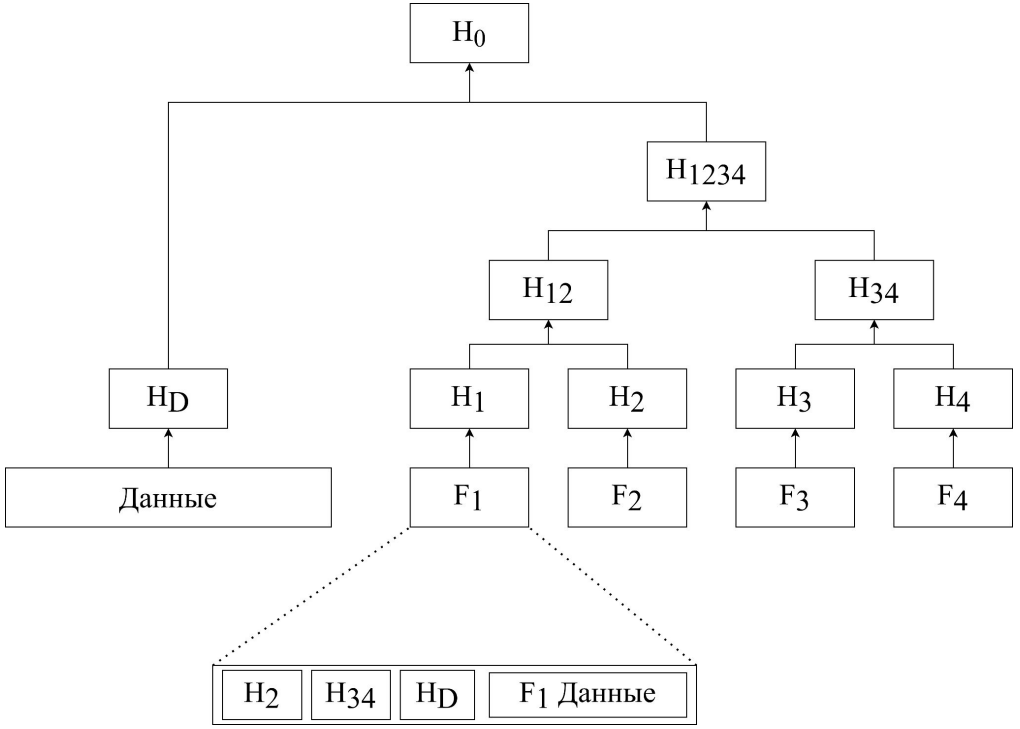






Существующие решения

- PASIS
  - p-m-n схема с использованием стирающих кодов;
  - проверка валидности через кросс чексуммы;
  - не накладывает дополнительных условий по использованию.
- OceanStore
  - дерево Меркла вместо кросс чек-сумм;
  - адресация на основе содержимого;
  - обновление - создание нового объекта.



Существующие решения

- Криптографические файловые системы:
  - ключ, как способ исключения чтения.
- Git:
  - пример зависимости блоков на основе ациклического направленного графа Меркла;
  - адресация по содержимому.
- Bitcoin:
  - корень дерева Меркла как хеш-сумма отдельного блока;
  - блокчейн, как способ сохранения консистентности цепочки блоков;
  - защита от изменения, удаления, добавления на основе концепции proof-of-work.

Существующие решения

НД	PASIS	КФС	OceanStore	Git	Bitcoin
чтение (исключение)	-	+	-	-	-
ЧУ или ЧИ (восстановление)	+	-	+	-	-
изменение (исключение)	-	-	-	-	+
изменение (доказательство)	+	-	+	+/-	+
удаление (доказательство)	-	-	-	+/-	+

Метод хранения данных в СУБД ClickHouse в движке MergeTree

Кусок — блок, в виде которого движок хранит данные.

Партиция — логическая группа кусков, определяемая по заранее известному признаку.

Операция слияния кусков — объединение подмножества кусков в 1 кусок, содержащий все данные объединяемого подмножества.

Гранула — минимальный набор данных (множество элементов), которым оперирует движок при чтении и записи данных.

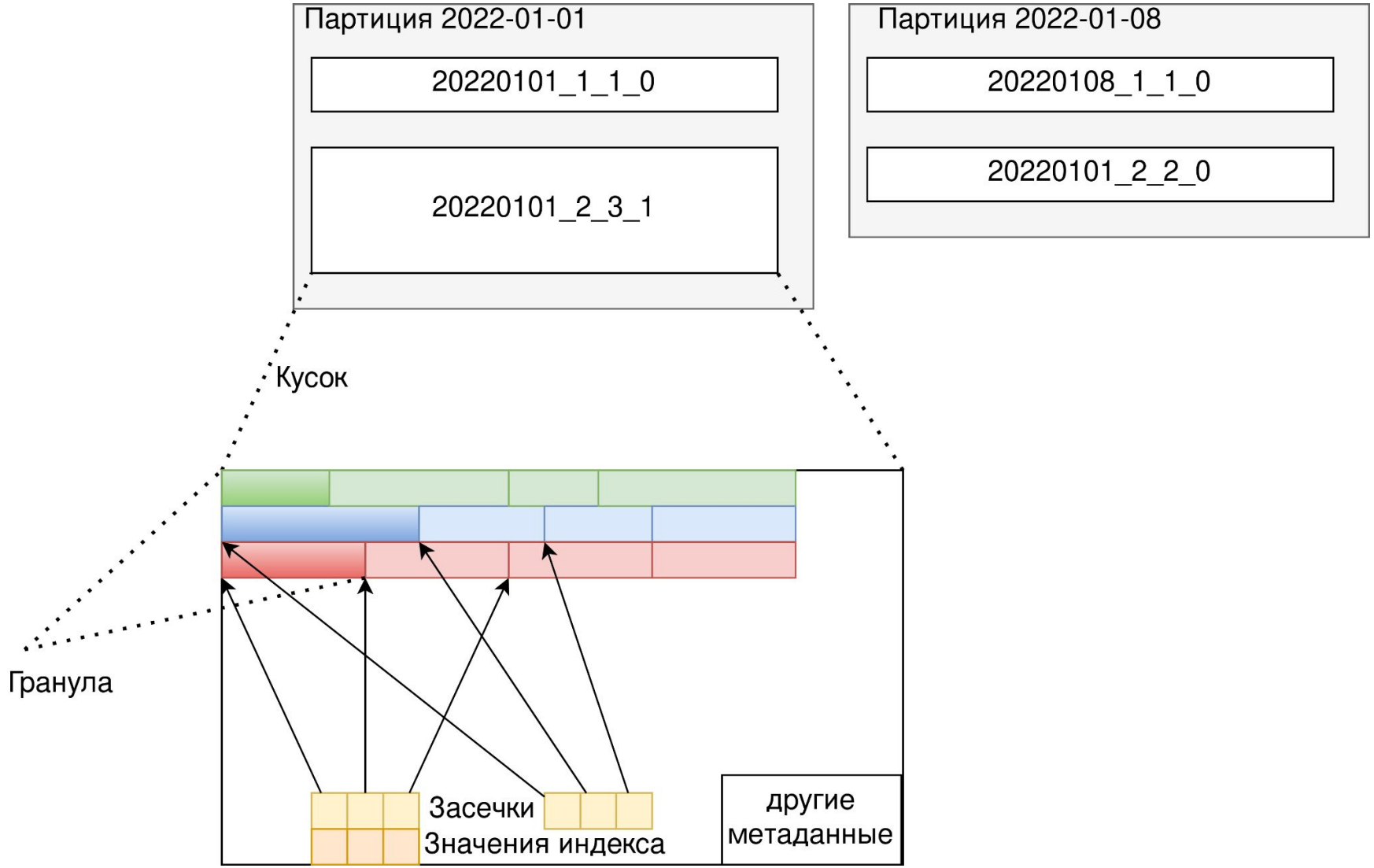
Индекс — отсортированный первичный ключ, по нему сортируются все данные.

Засечки — смещение значений столбцов в файле, соответствующих определенному значению первичного ключа.

2 формата хранения:

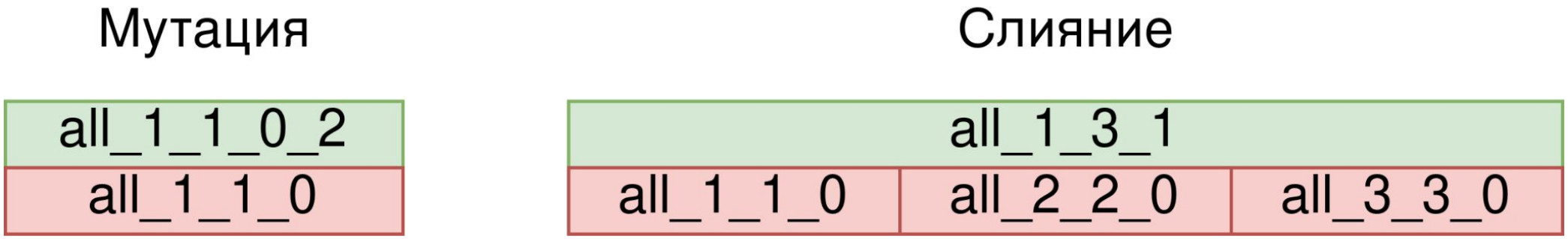
- компактный;
- широкий.

Метод хранения данных в СУБД ClickHouse в движке MergeTree



Вставки, слияния и мутации в MergeTree

- Вставка:
  - каждая вставка — новый блок;
  - атомарность операции за счет использования временного блока и переименовывания.
- Слияния:
  - служат для оптимизации хранения и поиска;
  - на входе произвольное количество кусков;
  - на выходе всегда 1 кусок.
- Мутации:
  - служат для изменения данных;
  - могут применяться по несколько к одному куску;
  - в одной задаче мутации участвует 1 кусок.



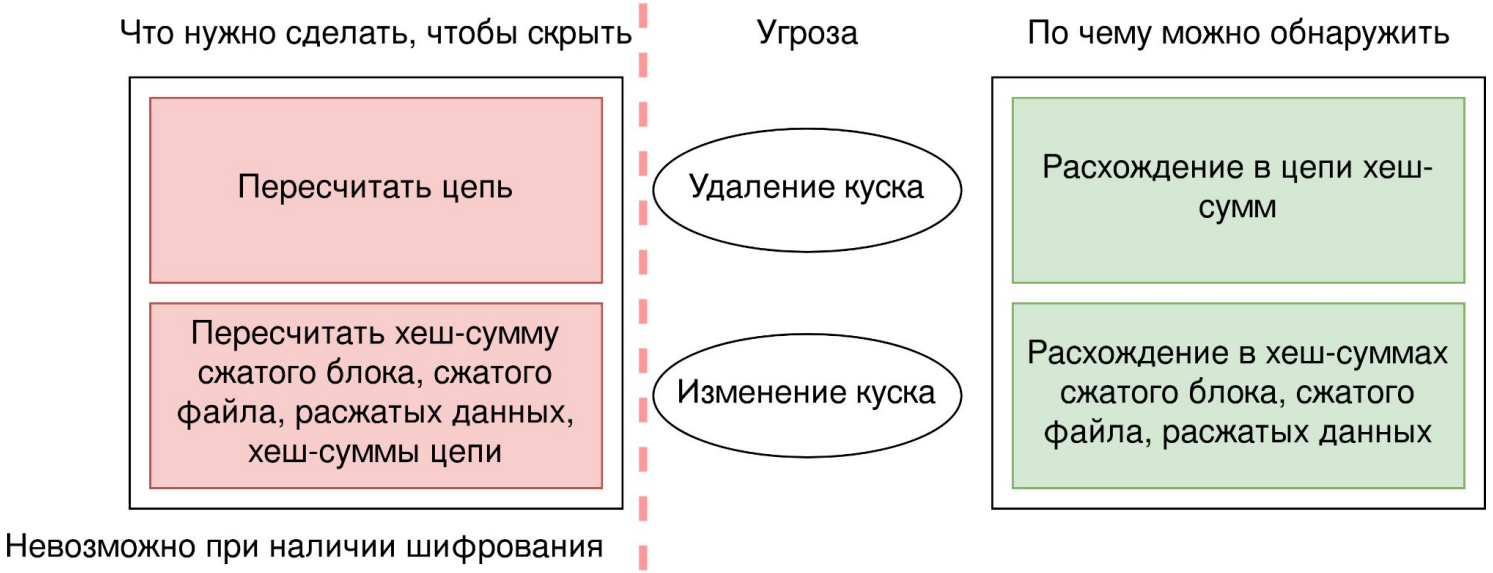
Анализ текущей защиты от неправомерного доступа в MergeTree

- Дополнительные возможности:
  - шифрование данных на уровне директории и столбца;
  - проверка целостности данных на уровне куска.
- Без шифрования не имеет защиты.
- С шифрованием:
  - исключается возможность неправомерного чтения;
  - возможность доказательства неправомерного изменения.
  - **невозможность** доказательства неправомерного удаления блока.

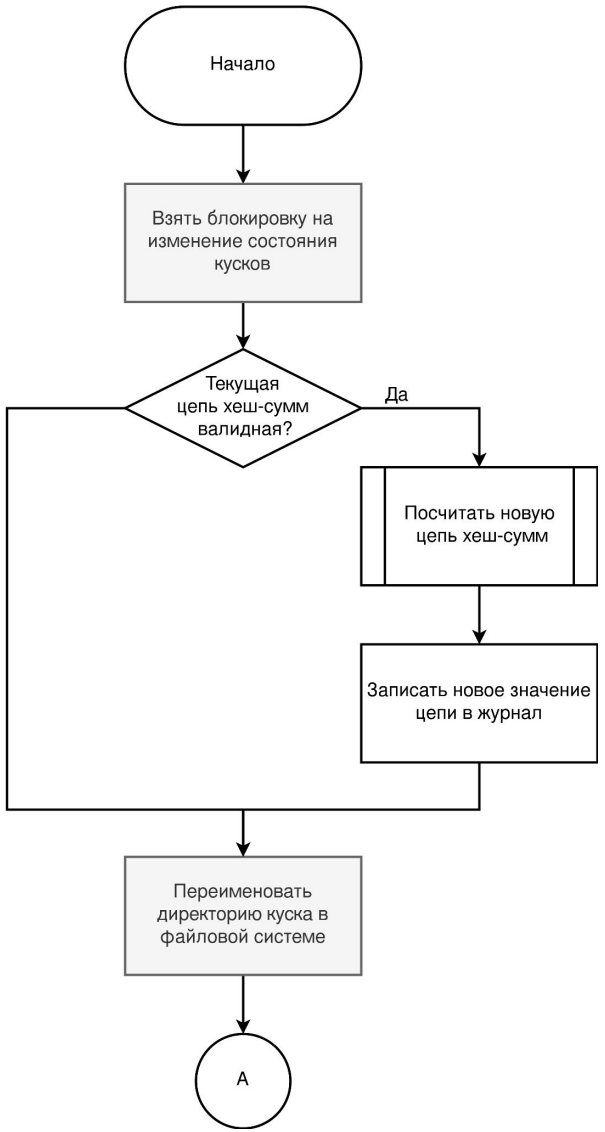
Предлагаемый метод хранения данных с возможностью доказательства неправомерного доступа

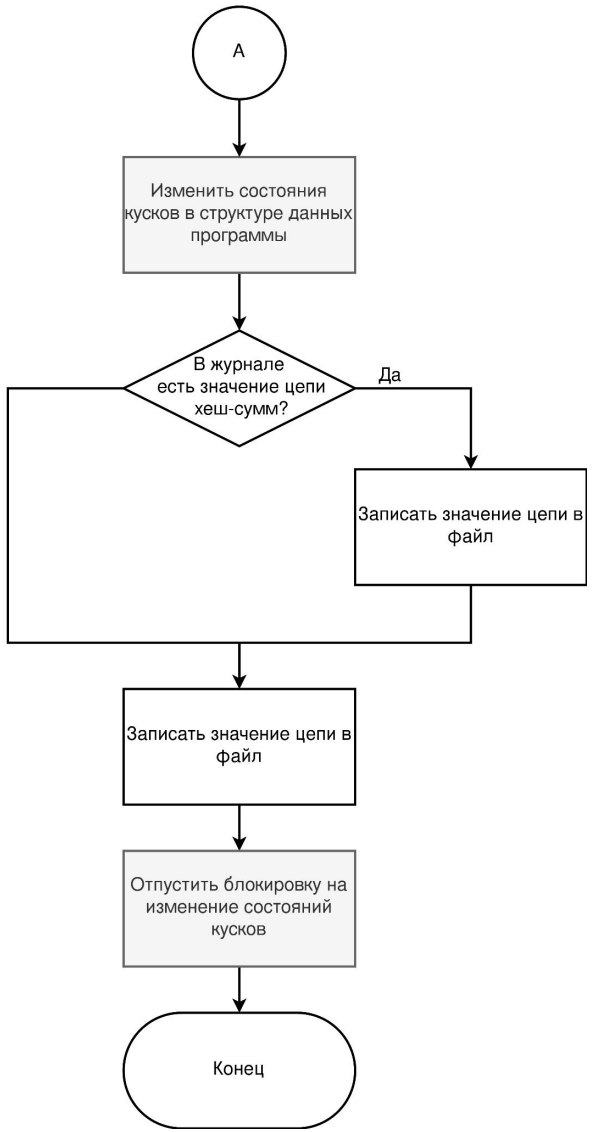
- связывает куски между собой в цепь хеш-сумм — аналог блокчейна;
- поддерживает атомарность за счет использования журнала;
- имеет возможность доказательства неправомерного удаления в отличие от оригинального метода.

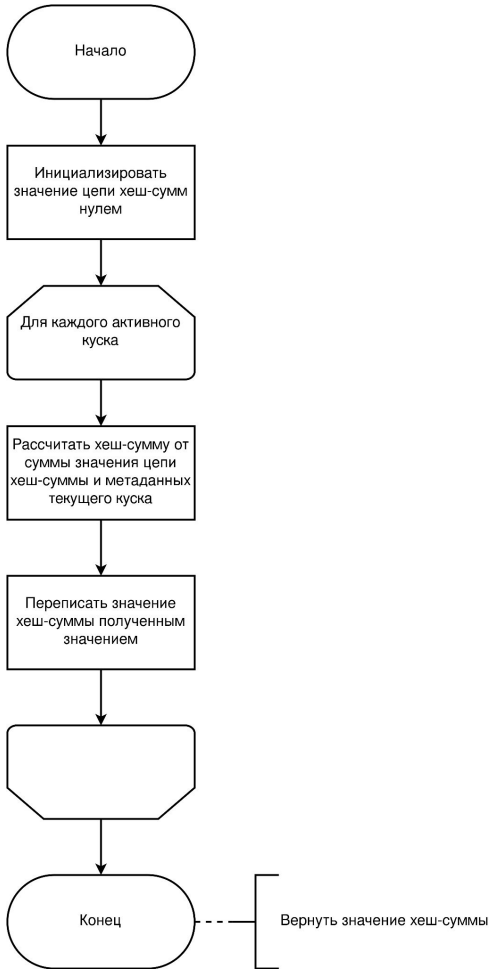
$$y_i = hash(x_i|y_{i-1}), y_1 = hash(x_1)$$



Схемы алгоритмов разрабатываемого метода







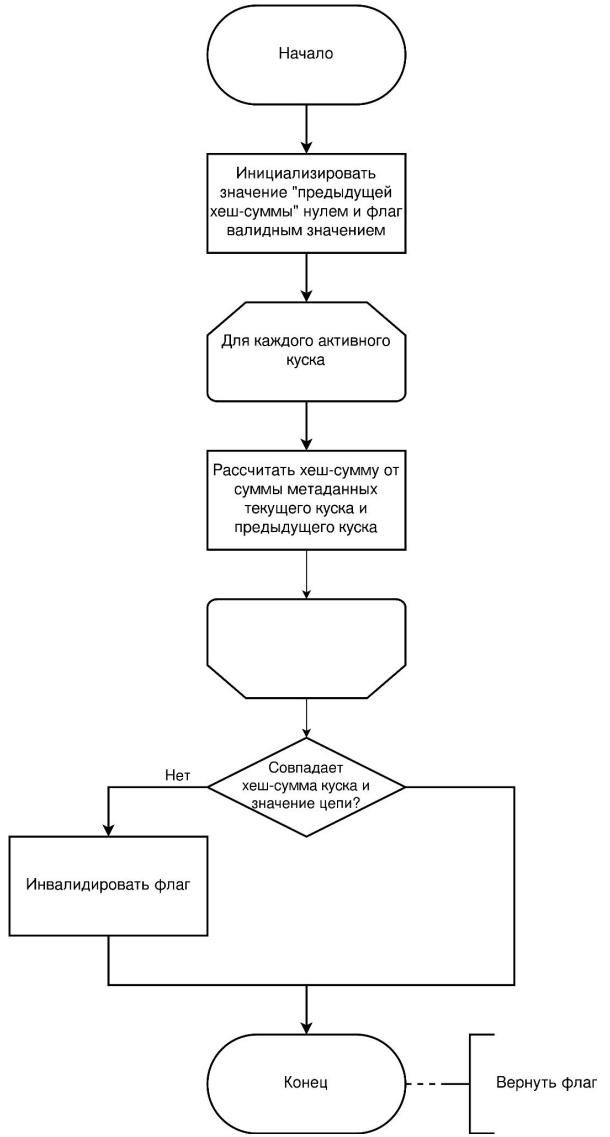
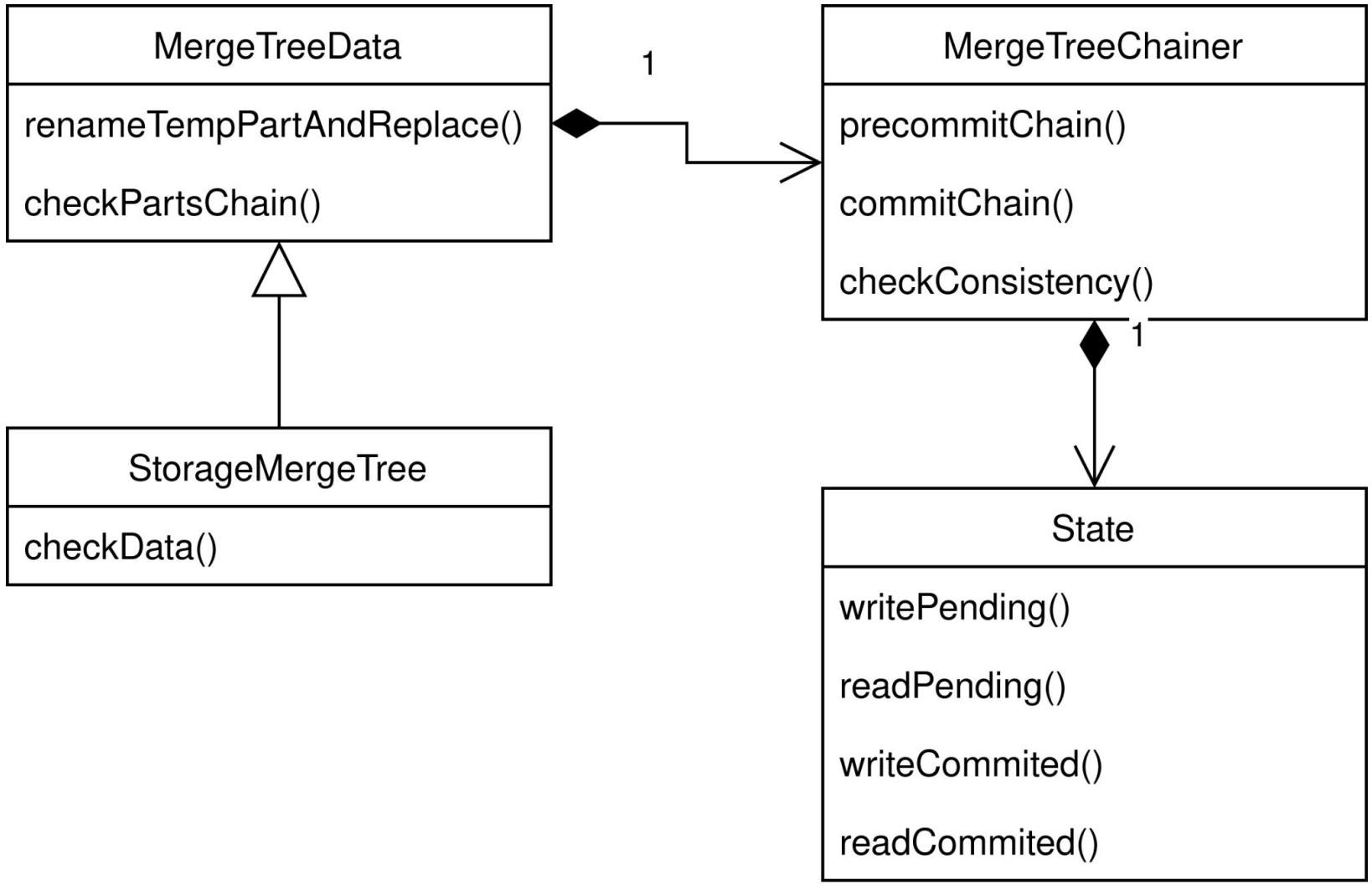


Диаграмма классов, участвующих в новом методе



Результаты экспериментов

- Был проделан обход защиты при отсутствии шифрования на уровне бинарных файлов:
  - если был изменен блок, то пересчитать хеш-суммы измененных сжатых блоков;
  - если был изменен блок, то пересчитать хеш-суммы сжатых и расжатых данных всего файла;
  - пересчитать значение цепи хеш-сумм.
- С шифрованием была проверена:
  - правильность работы метода при вставке, слиянии и мутации;
  - правильность работы при изменении блока (старый метод тоже работает);
  - правильность работы при **удалении** блока (старый метод не работает).

Выводы

В результате выполнения данной работы была достигнута цель работы, а также решены все поставленные задачи, а именно:

- были описаны и классифицированы виды защиты информации;
- были рассмотрены базовые элементы и понятия, используемые при проектировании методов хранения информации с возможностью защиты от неправомерного доступа;
- был проведен анализ существующих методов хранения информации с защитой от неправомерного доступа;
- был проведен анализ блочного хранения данных в СУБД ClickHouse в движке MergeTree, на предмет защиты информации от неправомерного доступа;
- был спроектирован и разработан метод блочного хранения данных с возмозможностью доказательства неправомерного доступа;
- разработанный метод был исследован на предмет невозможности реализации угроз при наличии шифрования системы.

Направления дальнейшего развития

Реализация метода для движка ReplicatedMergeTree.