

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

Студент: Пересторонин Павел Геннадьевич

Руководитель: Григорьев Александр Сергеевич

Москва

2022

Цель и задачи

Цель — разработать метод блочного хранения данных с возможностью доказательства неправомерного доступа на основе хеш-сумм.

Задачи:

- 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;
- спроектировать и реализовать метод блочного хранения данных с возможностью доказательства неправомерного доступа;
- исследовать метод на предмет защищенности при различных условиях.

Защита информации ограниченного доступа

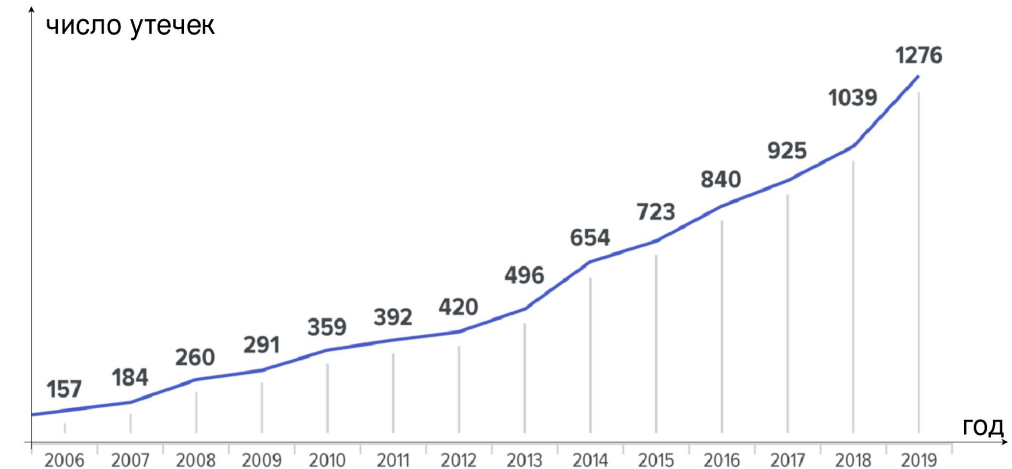
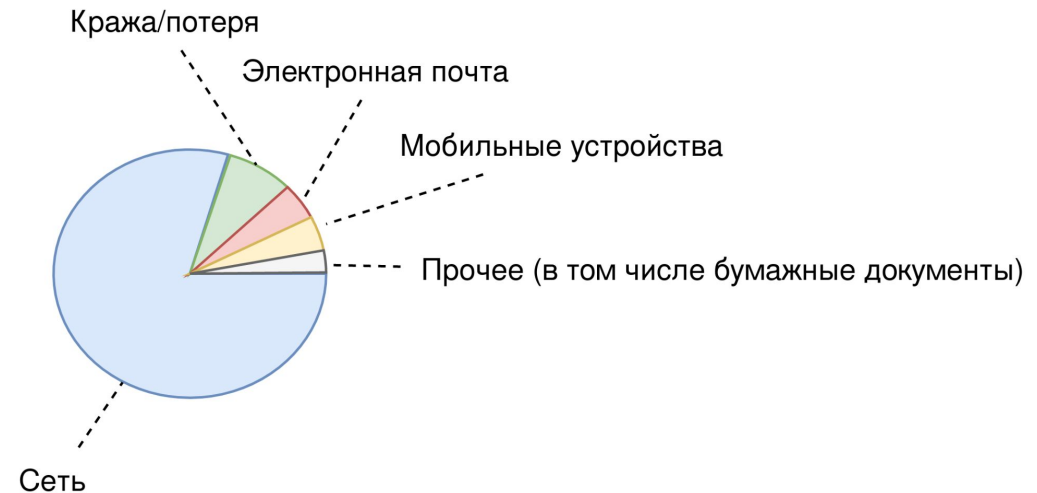
Информация ограниченного доступа:

- коммерческая тайна;
- персональные данные;
- служебная тайна;
- секрет производства;
- другое.

Интерес к теме повышается, потому что:

- количество утечек увеличивается;
- большую долю составляет сеть (76,6%);
- видна тенденция на ужесточение регулирования в сфере информационной безопасности со стороны государств в мире.

Каналы утечек данных



Виды защиты от неправомерного доступа



Анализ существующих решений на предмет защиты от неправомерного доступа

неправомерное действие (защита)	Криптографические ФС	Git	PASIS	OceanStore	Bitcoin
чтение (исключение)	+	-	-	-	-
удаление/изменение (исключение)	-	-	-	-	+
частичное удаление/изменение (восстановление)	-	-	+	+	+
удаление блока (доказательство)	-	+/-	-	-	+
изменение (доказательство)	-	+/-	+	+	+

Методы блочного хранения данных

Файловая система (ФС) и СУБД

Дополнительные возможности СУБД в сравнении с ФС:

- сжатие данных;
- возможность контроля размеров и количества блоков;
- ускоренный поиск по условию и извлечение информации;
- обеспечение целостности данных.

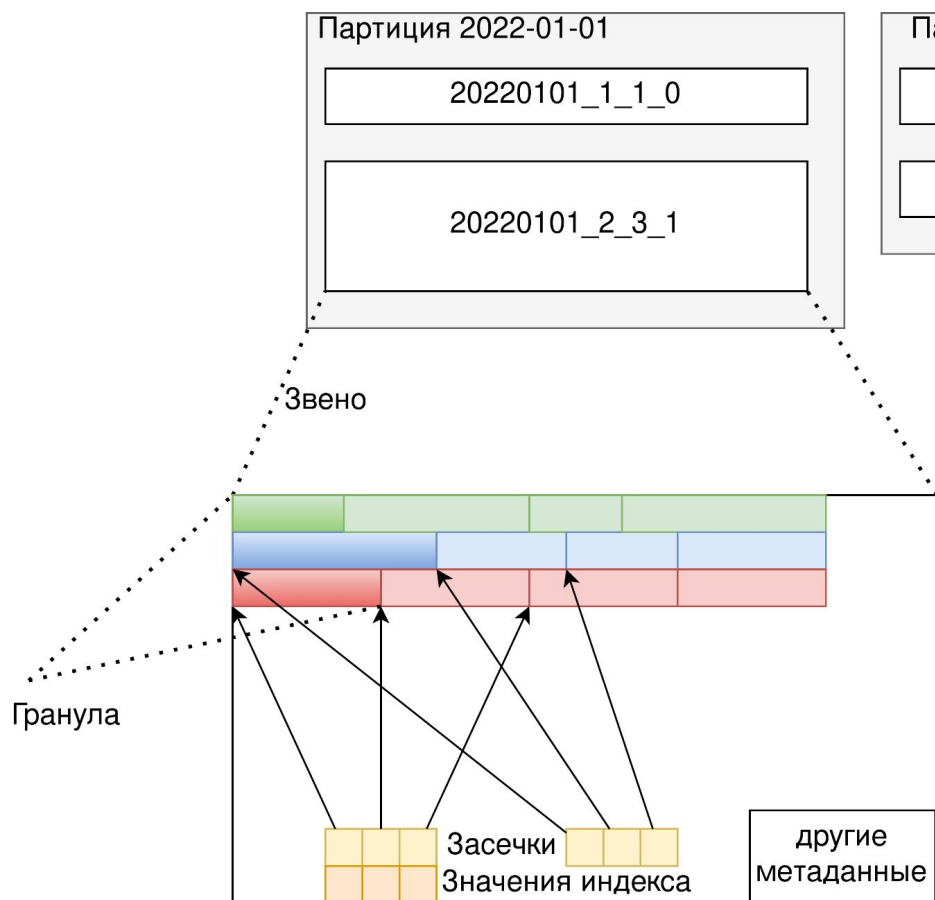
Требования к СУБД:

- блочное хранение данных;
- открытый исходных код.

СУБД	блочное хранение	Открытый код
Vertica	+	-
Infobright	+	+/-
ClickHouse	+	+

В ClickHouse рассматривается подсистема хранения MergeTree.

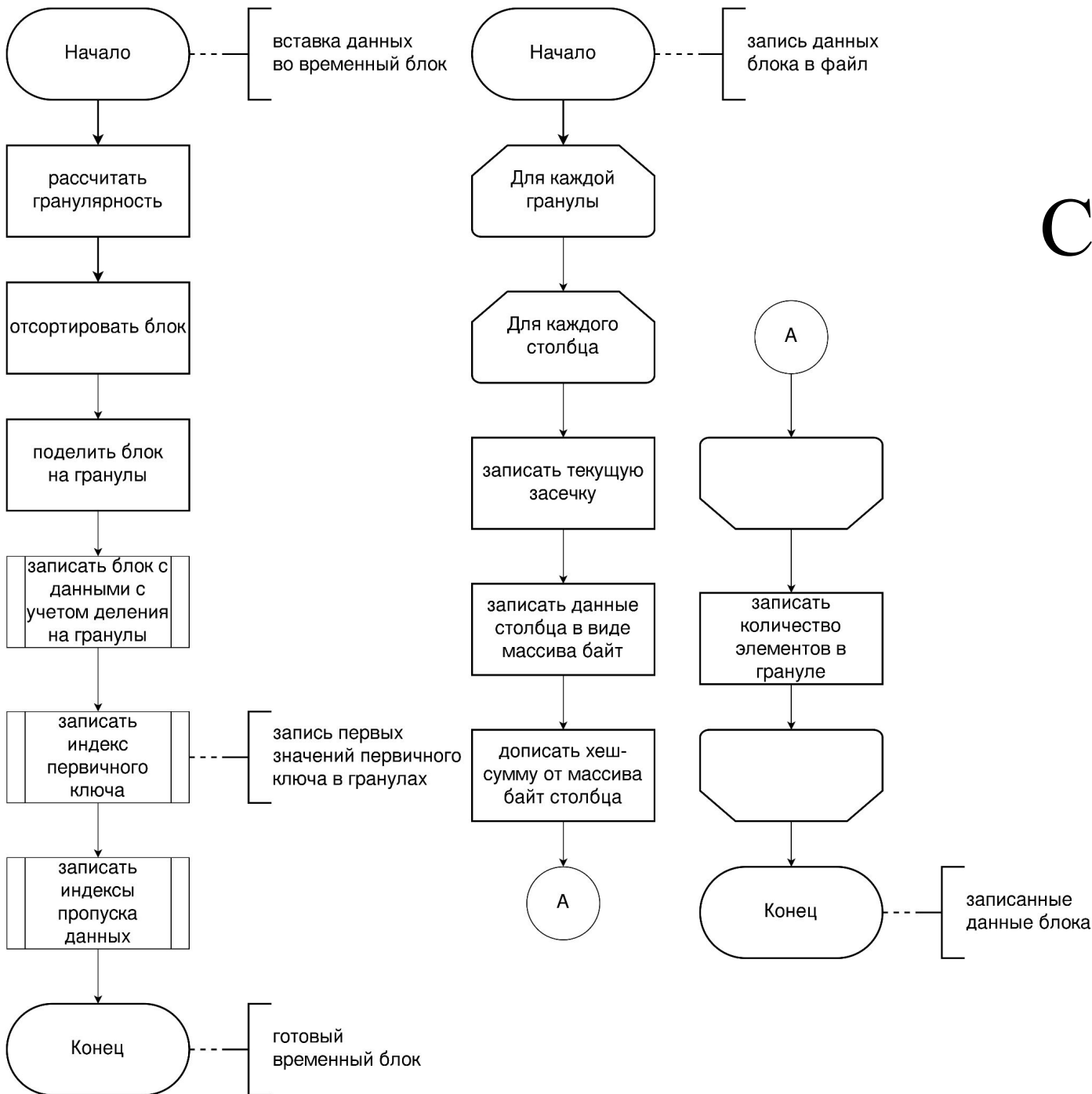
Метод блочного хранения данных в СУБД ClickHouse в подсистеме хранения MergeTree



Компоненты СУБД ClickHouse подсистемы хранения MergeTree:

- Звено — блок, единица хранения информации.
- Партиция — логическая группа звеньев.
- Гранула — единица записи и чтения данных.
- Индекс — отсортированные значения первых в гранулах первичных ключей.
- Засечки — смещение столбцов в файле для значений индекса.

Операции вставки и слияния в СУБД ClickHouse в MergeTree



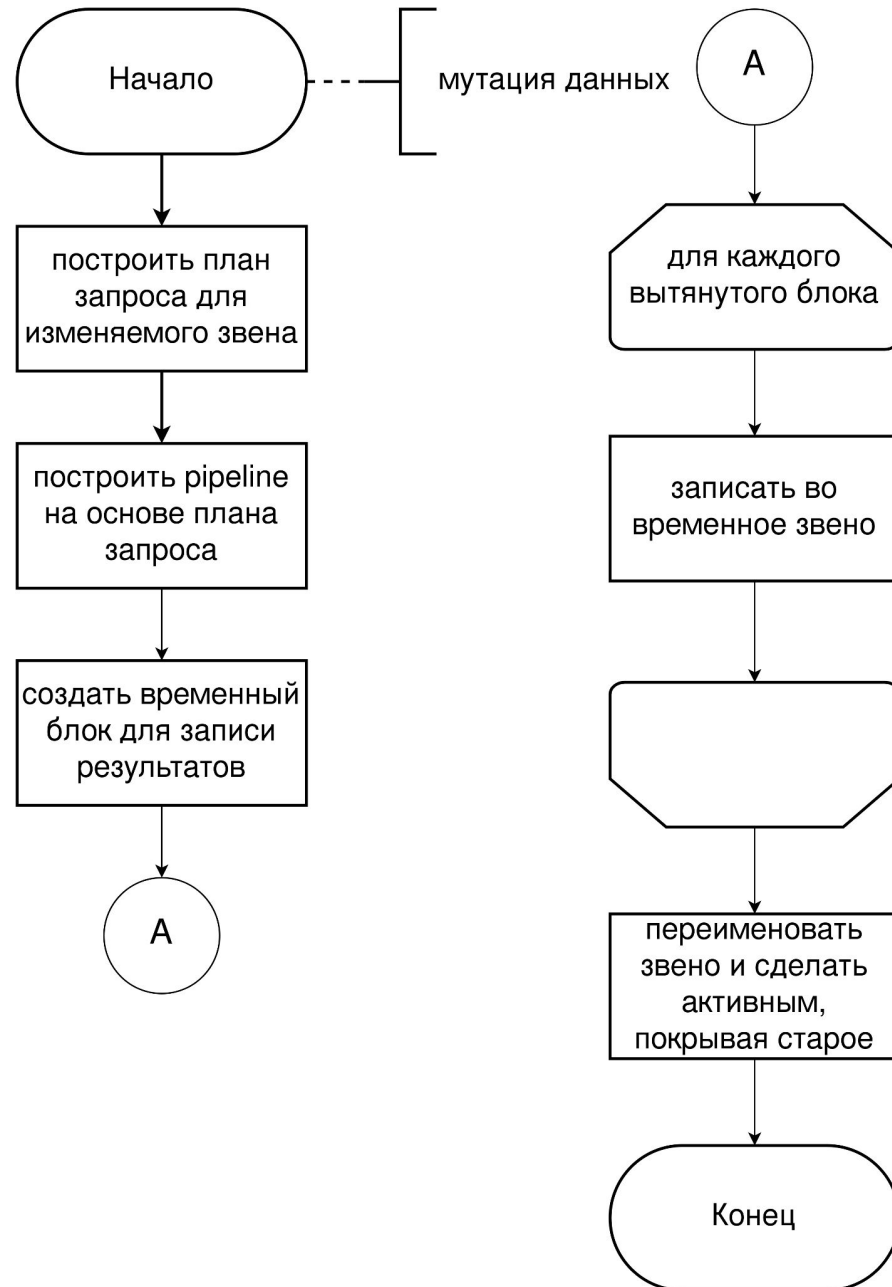
Пример добавления звена:

all_1_3_1	all_4_4_0
-----------	-----------

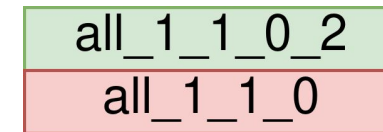
Пример слияния звеньев:

all_1_3_1		
all_1_1_0	all_2_2_0	all_3_3_0

Операция мутации в СУБД ClickHouse в MergeTree



Пример мутации звена:



Свойства мутаций:

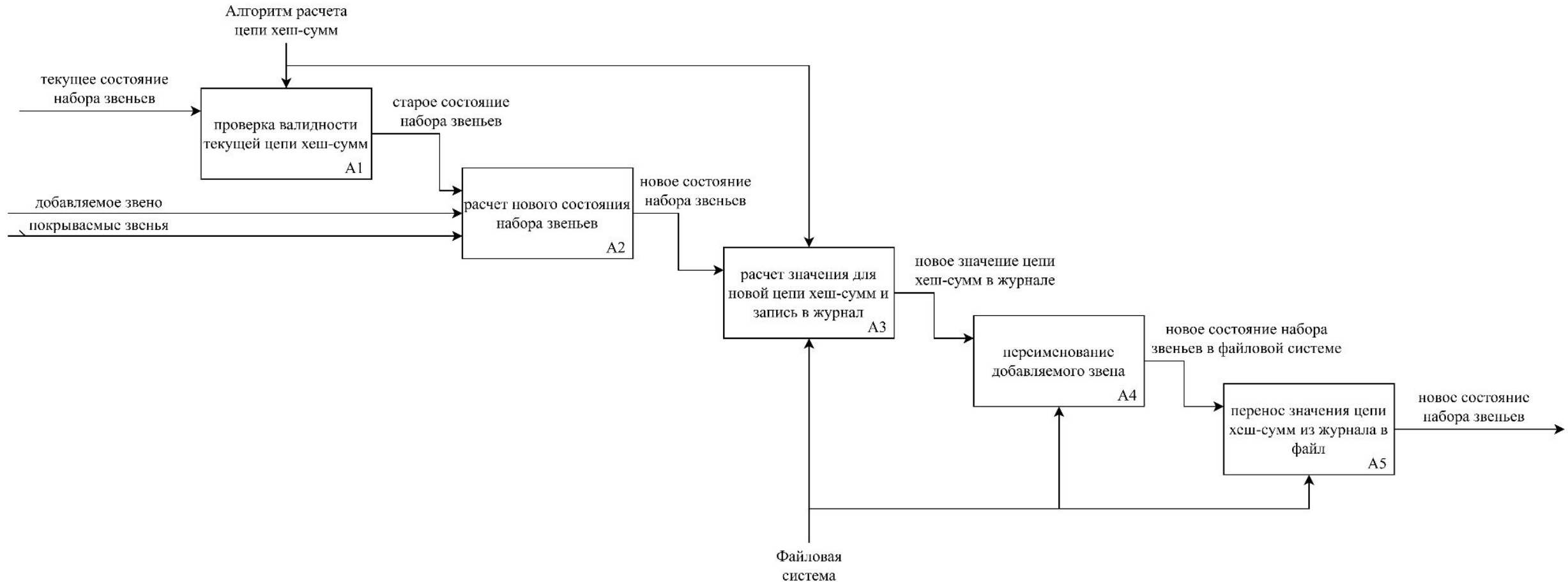
- могут применяться одновременно несколько к 1 звену;
- версия разделяется с версией звеньев, используется в выборе звеньев для мутаций.

Анализ существующей защиты данных от неправомерного доступа в движке MergeTree

неправомерное действие (защита)	без шифрования	с шифрованием
чтение (исключение)	-	+
удаление/изменение (исключение)	-	-
частичное удаление/изменение (восстановление)	-	-
удаление блока (доказательство)	-	-
изменение (доказательство)	+/-	+

- Дополнительные возможности:
 - шифрование данных на уровне директории и столбца;
 - проверка целостности данных на уровне звена.

Функциональная модель программного комплекса



Предлагаемый метод хранения данных с возможностью доказательства неправомерного доступа

Сценарии возможных угроз:

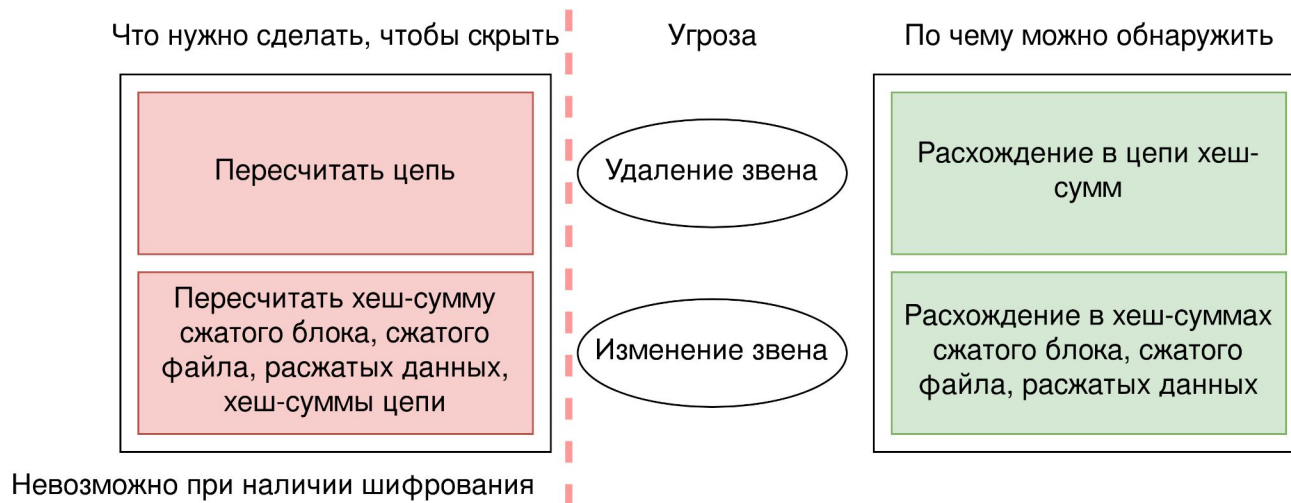
Расчет цепи хеш-сумм:

$$y_1 = \text{hash}(x_1),$$

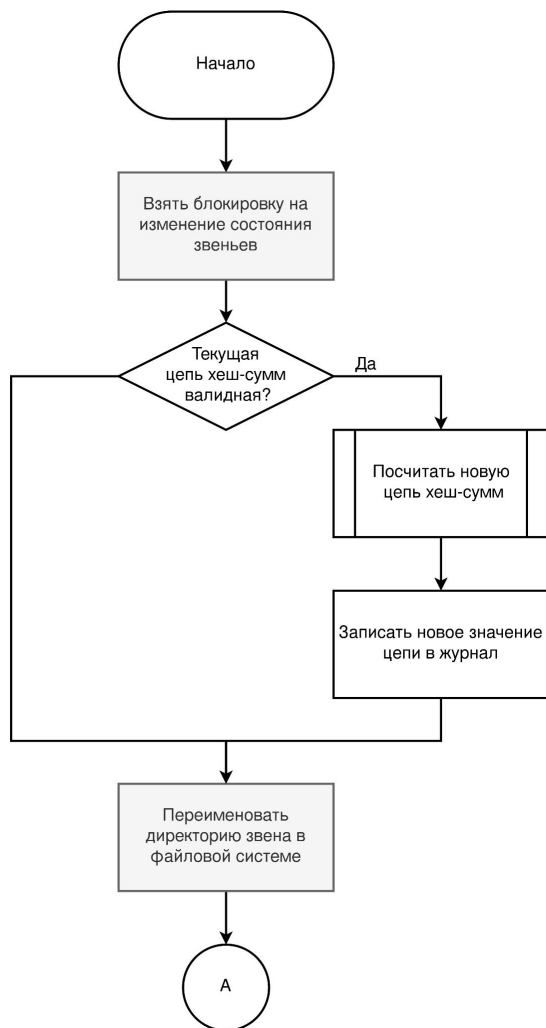
$$y_i = \text{hash}(x_i | y_{i-1}).$$

где:

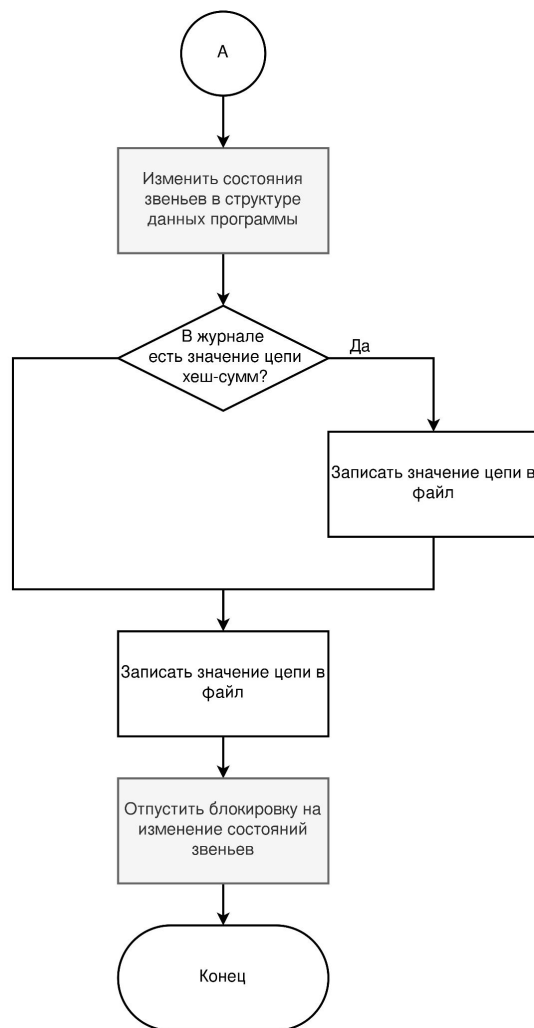
- i — индекс звена;
- hash — хеш-функция;
- $|$ — операция конкатенации байтовых массивов;
- x — метаданные звена в байтовом представлении.



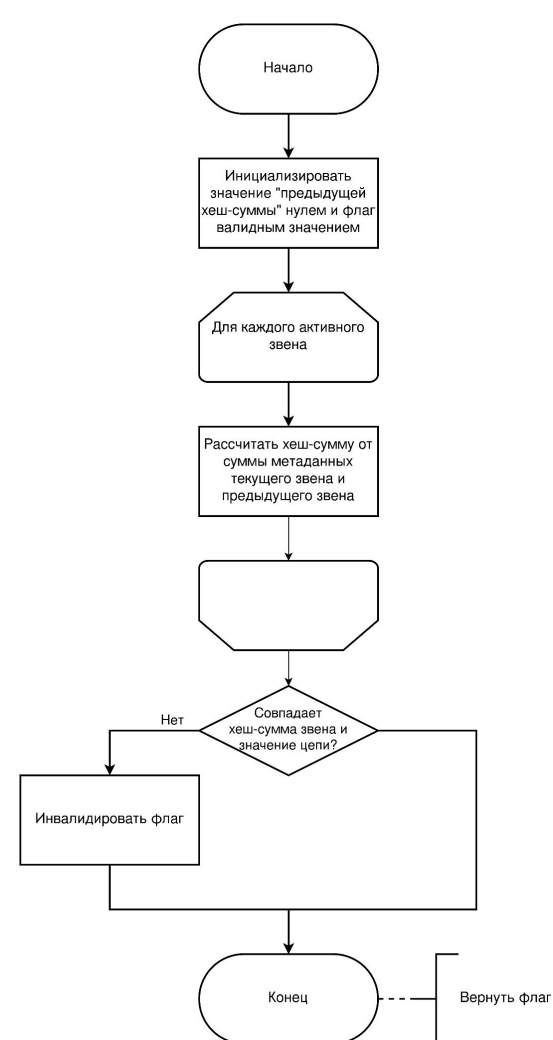
Методы расчета и валидации цепи хеш-сумм и обновления состояния звеньев



обновление состояния звеньев

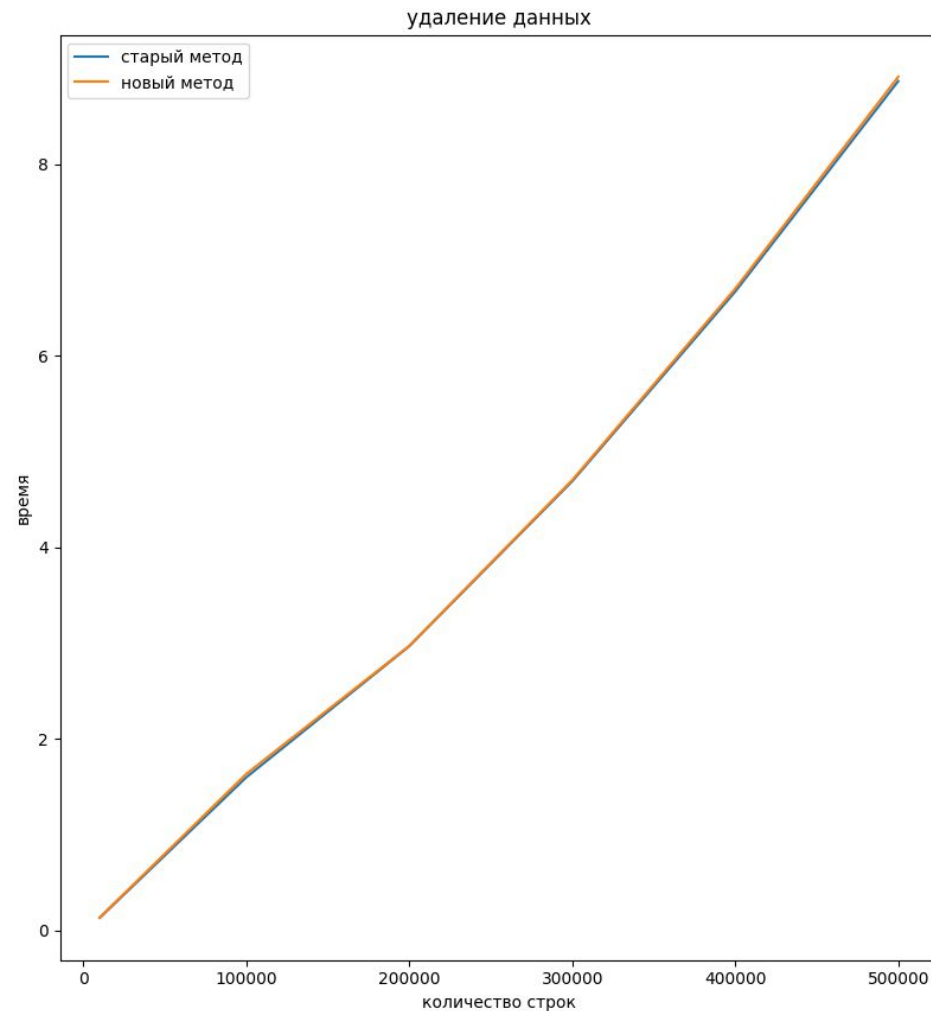
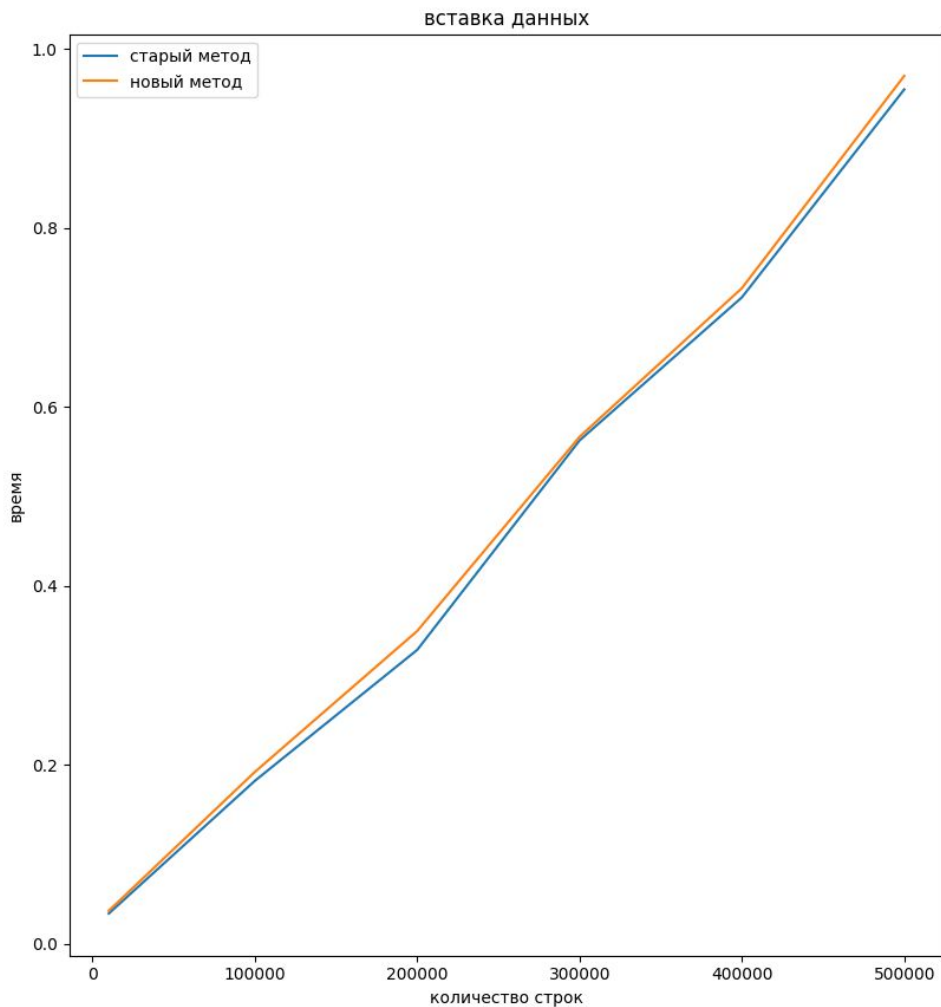


расчет цепи хеш-сумм



валидация цепи хеш-сумм

Зависимость времени выполнения операций вставки и удаления данных от количества строк в блоке и используемого метода



Заключение

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

- 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 спроектирован и реализован метод блочного хранения данных с возможностью доказательства неправомерного доступа;
- метод был исследован на предмет защищенности при различных условиях.

Направления дальнейшего развития

- реализация метода блочного хранения данных с возможностью доказательства неправомерного доступа для подсистемы хранения ReplicatedMergeTree СУБД ClickHouse;
- реализация возможности восстановления после частичного удаления или изменения звена в движке ReplicatedMergeTree СУБД ClickHouse.

ReplicatedMergeTree — аналогичная с точки зрения физического хранения данных MergeTree подсистема хранения, обладающая возможностью репликации звеньев.