

# МЕТОД ФОРМИРОВАНИЯ КОМПЛЕКСА ТЕХНОЛОГИЙ ДЛЯ ОРГАНИЗАЦИИ ХРАНЕНИЯ ДАННЫХ КОРПОРАТИВНЫХ СОБЫТИЙ

Даниелян А.Г.

Российский государственный гуманитарный университет, Москва, Россия

artem.dan5@mail.ru

*Аннотация. В данной статье предложен метод формирования комплекса технологий для организации хранения данных корпоративных событий с оптимизацией по заданному набору критериев. Основу метода составляет новая модель процесса принятия решения по выбору совместимых технологий с учетом многих критериев оценивания их эффективности при обработке больших данных, а также введения ограничений на последовательную их совместимость при использовании, что составляет новизну предлагаемого метода.*

*Ключевые слова: обработка больших данных, технологии хранения корпоративных данных, оптимизация комплекса технологий, многокритериальная оптимизация.*

## Введение

В условиях стремительного роста цифровизации и увеличения числа информационных потоков современные ИТ-инфраструктуры формируют значительные объемы событийной информации – включая журналы регистрации (логи), показатели мониторинга (метрики) и данные аудита. Эти данные являются отражением ключевых аспектов функционирования информационных систем и процессов организации. Их надежное, масштабируемое и безопасное хранение становится критически важным как с точки зрения обеспечения устойчивости корпоративных сервисов, так и в контексте соблюдения требований регуляторных органов, анализа инцидентов и информационной безопасности.

Повышенное внимание к вопросам защиты информации, а также необходимость быстрого доступа к актуальным данным для принятия управленческих решений подчеркивают значение создания эффективной архитектуры хранения событий. Такая архитектура должна обеспечивать не только хранение и доступность информации, но и соответствовать современным требованиям к адаптивности и интеграции с действующими компонентами цифровой экосистемы предприятия.

Процесс выбора технологических решений для построения событийного хранилища представляет собой многопараметрическую задачу, требующую комплексного анализа. Среди ключевых характеристик, подлежащих обязательному учету, можно выделить: уровень отказоустойчивости, способность к масштабированию, показатели производительности при записи и чтении данных, совокупную стоимость владения, эффективность использования ресурсов хранилища, наличие механизмов шифрования, а также степень совместимости с другими элементами ИТ-ландшафта организации.

Дополнительные сложности в процессе выбора обусловлены быстро меняющимся технологическим ландшафтом, приоритетом использования отечественных решений в ряде отраслей и необходимостью интеграции с уже существующими системами. Перечисленное формирует актуальную задачу формирования комплекса технологий для организации хранения данных корпоративных событий, оптимального с точки зрения критериев оценивания его эффективности.

Проведенный анализ источников, представленных в открытом доступе, позволил выделить работы, наиболее близкие по смыслу.

Так, в работе [1] предложена гибридная архитектура хранения: транзакции в блокчейне и off-chain хранилище (IPFS), реализованы smart-контракты для автоматизации верификации и контроля доступа. Показана и оценена масштабируемость при обработке миллионов записей с низкой задержкой.

Аналогично, модель в работе [2] интегрируется с публичным блокчейном Ethereum и демонстрирует высокую устойчивость к атакам.

В статье [3] разработано распределённое решение на базе PBFT-blockchain, позволяющее преобразовать традиционные audit-логи в защищённую, отказоустойчивую систему с оценкой производительности по задержке, объёму payload и размерам сети.

Работа [4] сочетает неизменность логов с криптографической защитой конфиденциальной информации (с применением zero-knowledge proofs), сохраняя возможность публичной верификации без раскрытия чувствительных данных.

Архитектуры вроде LogStore (Alibaba, SIGMOD 2021) показывают реальные промышленные примеры облачных лог-баз, позволяющие обрабатывать десятки миллионов лог-записей в секунду и обслуживать сотни тысяч клиентов с эффективным хранением и распределением нагрузки [5]. Также

исследование концепций LogBase иллюстрирует лог-ориентированное хранение с оптимизацией записи для систем с интенсивной нагрузкой [6].

Представленные в открытом доступе обзоры по интеграции блокчейна и логирования, например, [7] подчеркивают недостаток комплексных критериев выбора: большинство систем фокусируются на одном аспекте – безопасности или производительности, но не охватывают многокритериальный комплекс: отказоустойчивость, совместимость, стоимость, нормативное соответствие.

Кроме того, не были обнаружены работы, позволяющие осуществлять обоснованный выбор тех или иных технологий с учетом их совместимости.

Большинство исследований либо развивают отдельные технологии (например, blockchain-логирование или облачные системы) [1-3], либо ограничиваются упрощенной моделью оценки [8].

В данной работе предложен метод формирования комплекса технологий для организации хранения данных корпоративных событий. На основании анализа литературных источников выбраны основные показатели эффективности функционирования хранилища данных, сформированы критерии оптимизации комплекса технологий, также сформулирована оптимизационная задача целочисленного программирования, решение которой позволяет получить оптимальную комбинацию технологий для хранения данных. Проведенный натурный эксперимент позволяет сделать вывод о целесообразности использования предлагаемого метода при проектировании хранилищ данных корпоративных событий.

## 1. Показатели эффективности хранения данных корпоративных событий

Выберем показатели, на основании которых мы будем определять эффективность хранения событий.

При анализе эффективности архитектурных решений для систем хранения событий ключевыми показателями, подлежащими оценке, являются следующие:

- отказоустойчивость;
- масштабируемость;
- пропускная способность при записи данных;
- производительность выборки и извлечения информации;
- совокупная стоимость владения;
- эффективность хранения;
- возможности шифрования данных.

Отказоустойчивость описывает степень защищенности системы от сбоя отдельных компонентов, обеспечивая непрерывный доступ к данным, а также сохранение их целостности и конфиденциальности. Данный критерий учитывает наличие механизмов репликации, резервирования и автоматического восстановления.

Масштабируемость отражает способность системы адекватно реагировать на увеличение объема и скорости поступления информации, а также обеспечивать сохранение производительности при росте числа источников и длительности хранения.

Параметр скорости записи характеризует возможность системы в режиме реального времени обрабатывать высокообъемные потоки событий, достигающие сотен тысяч единиц в секунду (events per second, EPS), что является типичным для крупных корпоративных сетей.

Скорость извлечения данных включает в себя быстродействие выполнения аналитических, агрегационных и поисковых операций, как правило, выполняемых пользователями с аналитическим доступом.

Экономический аспект, включающий стоимость реализации решения, учитывает затраты на инфраструктуру, администрирование, поддержку, а также лицензионные отчисления при использовании проприетарного ПО.

Эффективность хранения представляет собой степень снижения объема занимаемого пространства за счёт механизмов сжатия и оптимального хранения информации на физических носителях.

Шифрование данных рассматривается в контексте обеспечения информационной безопасности как при хранении («данные в покое»), так и в процессе передачи («данные в движении»), с акцентом на поддержку современных криптографических стандартов.

Для инструментов извлечения, преобразования и загрузки данных (ETL) в контексте событийных хранилищ выделим следующие ключевые параметры:

- отказоустойчивость;
- масштабируемость;

- гибкость трансформационных операций;
- производительность;
- совокупная стоимость.

Данные параметры обусловлены ролью ETL-компонента в системной архитектуре: он обеспечивает агрегацию, фильтрацию и трансформацию данных, поступающих от распределённых источников. Отказоустойчивость и масштабируемость остаются критически важными ввиду необходимости непрерывной обработки больших потоков информации.

Производительность определяется объемом событий, обрабатываемых при заданных характеристиках вычислительной инфраструктуры, и напрямую влияет на своевременность поступления информации в хранилище.

Гибкость трансформаций особенно актуальна в случаях, когда данные подвергаются глубокому обогащению, что характерно для поведенческих аналитических систем и систем обнаружения аномалий.

Финансовые издержки включают расходы на внедрение, сопровождение и адаптацию ETL-инструмента, а также возможные затраты на лицензирование и обучение персонала.

В отличие от хранилищ данных и ETL-инструментов, системы визуализации несут меньшую нагрузку и менее критичны к отказам, однако требуют оценки по следующим показателям:

- функциональная гибкость;
- производительность;
- стоимость внедрения и эксплуатации.

Гибкость в данном случае определяется поддержкой различных способов представления данных: графиков, диаграмм, дашбордов, тепловых карт (heatmaps), а также наличием механизмов оповещения и адаптации интерфейса под задачи аналитиков.

Производительность измеряется временем генерации визуальных элементов при обработке больших объемов данных, что особенно важно для интерактивного анализа.

Экономическая составляющая аналогична предыдущим компонентам и охватывает затраты на установку, поддержку, масштабирование и лицензирование.

Обоснование выбранных показателей эффективности опирается на специфику систем класса Centralized Log Management (CLM), ориентированных на централизованный сбор и долговременное хранение событий от различных корпоративных источников: автоматизированных рабочих мест, серверов, сетевой инфраструктуры, корпоративных приложений и пр. При этом особое значение приобретает неизменяемость событийных данных после их записи в хранилище, что накладывает ограничения на выбор СУБД и архитектурных решений.

Следует отметить, что универсальной приоритизации показателей эффективности не существует – она определяется нормативно-правовой базой, корпоративной политикой информационной безопасности, специфическими требованиями к аналитике и допустимыми бюджетами. Таким образом, оценка и выбор решений должны проводиться индивидуально для каждой организации.

Для формализации задачи выбора оптимального стека технологий для создания корпоративного хранилища событий, с учетом выбранных показателей эффективности, можно подойти к ней как к многокритериальной задаче.

Рассмотрим 3 технологии, которые являются базовыми для организации хранения событий:

- ETL-инструмент.
- Хранилище событий.
- Инструмент визуализации данных.

В нашем случае, технологии включают в себя множества элементов:

ETL-инструменты: Logstash, Fluentd, Vector, NiFi.

Хранилища событий: ClickHouse, Hadoop, Elasticsearch, VictoriaMetrics, Loki, TimescaleDB.

Визуализация данных: Grafana, Apache Superset, Kibana, Redash.

Множества показателей эффективности для технологий включают в себя следующие элементы:

Хранилища данных: отказоустойчивость, масштабируемость, скорость записи, скорость чтения, стоимость, эффективность хранения, шифрование.

ETL-Инструменты: отказоустойчивость, масштабируемость, гибкость трансформации, производительность, стоимость

Инструменты визуализации: гибкость, производительность, стоимость

Оценки технологий формируются путем создания и опроса экспертной комиссии. Оценки проставляются в нормализованном виде от 0 до 10, где 10 – наилучший показатель по критерию, 0 – наихудший.

Таблица 1. Оценки хранилищ

	Отказоустойчивость	Масштабируемость	Скорость записи	Скорость чтения	Стоимость	Эффективность хранения	Шифрование
<b>Hadoop</b>	9.5	9.5	8	6	5	8	9
<b>CH</b>	8	9	9.5	9	7	8.5	7
<b>ES</b>	8	8	7.5	8.5	7	6	7
<b>Loki</b>	7	8	8	7	9.5	8	8
<b>TSDB</b>	8	7	7	8.5	6	7.5	8
<b>VM</b>	7.5	8	9	8.5	9	9.5	6

В таблице 2 представлены оценки для ETL-инструментов.

Таблица 2. Оценки ETL-инструментов

	Отказоустойчивость	Масштабируемость	Гибкость трансформаций	Производительность	Стоимость
<b>Logstash</b>	8	8	8	8	6
<b>Fluentd</b>	8	9	8	8	6
<b>Vector</b>	9	8	8	9	8
<b>NiFi</b>	9.5	9.5	9.5	7	5

В таблице 3 представлены оценки для инструментов визуализации данных.

Таблица 3. Оценки инструментов визуализации

	Гибкость	Производительность	Стоимость
<b>Grafana</b>	9	9	9.5
<b>Superset</b>	9.5	8	9
<b>Kibana</b>	8.5	8.5	7
<b>Redash</b>	8	7.5	8.5

Значения в данных таблицах нормируются. Таким образом, сумма оценок по каждому критерию будет равна единице.

## 2. Ограничения на совместимость технологий

Далее составим матрицу совместимостей технологий. Данный аспект важно учитывать при проектировании, так как из-за сложности совместимости технологий и их многообразия, стоимость разработки, внедрения и поддержки конечного решения можеткратно возрасти.

В рамках нашей задачи, рассматривая три основных компонента для хранения корпоративных данных, составим две матрицы совместимости: ETL-инструментов с хранилищами и инструментов визуализации с хранилищами.

Для проставления оценок совместимости будем использовать бинарные оценки. В данной системе оценивания 1 будет означать возможность интеграции решений встроенными средствами ПО («из коробки»), наличие распространенных плагинов для интеграции или известных простых способов интеграции, не требующих дополнительных серьезных затрат ресурсов на разработку. Оценка 0 будет подразумевать сложность интеграции решений, случаи, в которых нет простого способа интегрировать технологии, требующие дополнительных затрат на разработку.

Таблица 4. Совместимость ETL-инструментов и хранилищ данных

	Logstash	Fluentd	Vector	NiFi
<b>Hadoop</b>	1	1	1	1
<b>CH</b>	1	1	1	1
<b>ES</b>	1	1	1	1
<b>Loki</b>	1	0	1	1
<b>TSDB</b>	0	1	1	1
<b>VM</b>	0	1	0	1

Таблица 5. Совместимость инструментов визуализации и хранилищ

	Grafana	Superset	Kibana	Redash
Hadoop	1	1	0	1
CH	1	1	0	1
ES	1	1	1	1
Loki	1	0	0	0
TSDB	1	1	0	1
VM	1	0	0	0

### 3. Математическая модель процесса принятия решения по выбору совместимых технологий

Для каждой технологии может быть выбран лишь один вариант. Требуется осуществить выбор вариантов технологии, при котором суммарная оценка эффективности будет максимальна, при этом совместимость технологий не должна быть равна нулю.

Пусть  $A_{ij}$ ,  $i1 = 1, 2, \dots, n1$ ;  $j1 = 1, 2, \dots, m1$ ; – матрица, содержащая нормированные значения из таблицы 1;

$B_{ij}$   $i2 = 1, 2, \dots, n2$ ;  $j2 = 1, 2, \dots, m2$ ; – матрица, содержащая нормированные значения из таблицы 2;

$C_{ij}$   $i3 = 1, 2, \dots, n3$ ;  $j3 = 1, 2, \dots, m3$ ; – матрица, содержащая нормированные значения из таблицы 3;

$F_{ij}$   $i1 = 1, 2, \dots, n1$ ;  $j2 = 1, 2, \dots, m2$ ; – матрица, содержащая значения из таблицы 4;

$G_{ij}$   $i1 = 1, 2, \dots, n1$ ;  $j3 = 1, 2, \dots, m3$ ; – матрица, содержащая значения из таблицы 5.

$$F = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$

$$G = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

Введем матрицы бинарных переменных:

Матрицу-столбец  $X$ ,  $x_{ij} \in \{0,1\}$ , по количеству строк совпадающую с количеством строк матрицы  $A$ ,  
 Матрицу-столбец  $Y$ ,  $y_{ij} \in \{0,1\}$ , по количеству строк совпадающую с количеством строк матрицы  $B$ ,  
 Матрицу-столбец  $Z$ ,  $z_{ij} \in \{0,1\}$  с количеством строк, соответственно, как матрица  $C$ .

Значение 1 в перечисленных матрицах означает выбор той или иной технологии в исходных матрицах  $A$ ,  $B$ ,  $C$ .

Также введем дополнительные матрицы ограничений на допустимые значения показателей качества технологий:  $A_{constr}$ ,  $B_{constr}$ ,  $C_{constr}$ , размерностью, соответственно, как матрицы  $A$ ,  $B$ ,  $C$ . Значения в данных матрицах будут отражать минимально допустимые значения

Представим обобщенную критериальную функцию при помощи аддитивной свертки.

Для этого суммы оценок для каждой технологии взвешиваются некоторым  $\omega$ , определяющим важность показателя.

$$u(a_{i_1}) = \sum_{j_1=1}^{m_1} A_{i_1 j_1} \omega_{j_1} \quad (1)$$

$$u(b_{i_2}) = \sum_{j_2=1}^{m_2} B_{i_2 j_2} \omega_{j_2} \quad (2)$$

$$u(c_{i_3}) = \sum_{j_3=1}^{m_3} C_{i_3 j_3} \omega_{j_3} \quad (3)$$

Где

$$\sum_{j_1=1}^{m_1} \omega_{j_1} = 1; \sum_{j_2=1}^{m_2} \omega_{j_2} = 1; \sum_{j_3=1}^{m_3} \omega_{j_3} = 1;$$

$$\forall \omega_{j_1} \geq 0; \forall \omega_{j_2} \geq 0; \forall \omega_{j_3} \geq 0$$

В результате нам необходимо найти максимальную сумму оценок для каждого типа технологий так, чтобы результат удовлетворял ограничениям минимального значения и совместимости. В рамках данной многокритериальной задачи также необходимо найти взвешенную сумму, с учетом весов критериев  $\alpha$  (в данном случае – вес технологии). Также проведем нормализацию итоговых сумм, которые также должны быть меньше единицы.

Для решения многокритериальной задачи можем воспользоваться методом аддитивной свертки критериев [7], так как оценки по нашим критериям являются соизмеримыми и нормированными. В первую очередь для решения подобной задачи необходимо найти взвешенную сумму критериев, так как используется весовой метод [8].

Тогда, необходимо найти такие  $X, Y, Z$  которые обеспечивают.

$$\alpha_1 \sum_{i_1=1}^{n_1} u(a_{i_1}) x_{i_1} / \max \left( \sum_{i_1=1}^{n_1} u(a_{i_1}) \right) + \alpha_2 \sum_{i_2=1}^{n_2} u(b_{i_2}) y_{i_2} / \max \left( \sum_{i_2=1}^{n_2} u(b_{i_2}) \right) + \alpha_3 \sum_{i_3=1}^{n_3} u(c_{i_3}) z_{i_3} / \max \left( \sum_{i_3=1}^{n_3} u(c_{i_3}) \right) \rightarrow \max \quad (4)$$

И удовлетворяют системе ограничений

$$\alpha_1 + \alpha_2 + \alpha_3 = 1$$

$$\alpha_1 \geq 0; \alpha_2 \geq 0; \alpha_3 \geq 0$$

$$\sum_{i_1=1}^{n_1} x_{i_1} = 1, \sum_{i_2=1}^{n_2} y_{i_2} = 1, \sum_{i_3=1}^{n_3} z_{i_3} = 1,$$

$$f_{i_1 j_2} \neq 0, g_{i_1 j_3} \neq 0,$$

$$\forall x_{i_1}, a_{i_1 j_1} \geq a_{constr\_i_1 j_1};$$

$$\forall y_{i_2}, b_{i_2 j_2} \geq b_{constr\_i_2 j_2};$$

$$\forall z_{i_3}, c_{i_3 j_3} \geq c_{constr\_i_3 j_3}.$$

#### 4. Метод формирования комплекса технологий для организации хранения данных корпоративных событий

В предыдущем разделе была сформулирована многокритериальная задача оптимизации, в рамках которой необходимо максимизировать обобщенную критериальную функцию (4).

Соответственно, метод формирования комплекса технологий для организации хранения корпоративных событий будет иметь следующую последовательность действий:

- Сбор данных о специфике проектируемого хранилища
- Сбор экспертных оценок для множества используемых технологий
- Принятие решения о минимальных требуемых значениях
- Решение задачи (4)

Для решения подобной задачи могут быть применены различные подходы. Один из наиболее очевидных — метод полного перебора, обеспечивающий гарантированный просмотр всех допустимых

комбинаций параметров. Достоинством данного метода является его концептуальная простота и прозрачность реализации. Однако при увеличении размерности задачи и объёмов входных данных, данный подход становится крайне ресурсоёмким и плохо масштабируется.

Альтернативным методом является использование генетических алгоритмов [9], относящихся к классу эволюционных эвристик. Эти алгоритмы демонстрируют высокую эффективность при решении задач с большой размерностью поискового пространства. Их преимуществом является способность приближаться к оптимальному решению за счёт итеративного улучшения популяции решений на основе механизмов естественного отбора, мутаций и скрещивания. Основным ограничением при их применении является относительная сложность реализации и необходимость настройки параметров алгоритма.

## 5. Практический пример применения метода

В ходе решения задачи, описанной формулой (4) был получен следующий набор технологий: ETL-инструмент Vector, СУБД Clickhouse, инструмент визуализации данных Grafana. Веса критериев были взяты равные, однако в каждом конкретно рассматриваемом случае весовые коэффициенты должны расставляться исходя из задач, которые должно решать хранилище.

На основании данных программных продуктов был подготовлен тестовый стенд, включающий в себя конечный хост, на котором мы хотим собирать логи, в нашем случае это виртуальная машина с операционной системой Windows 10, центральный сервер Vector, реализующий обработку и отправку данных в хранилище, СУБД Clickhouse, и инструмент визуализации – Grafana.

В случае нашего стенда мы будем собирать журнал безопасности Windows. Для этого событий журнала будут собираться и обрабатываться с помощью утилиты evtx\_dump.exe, затем передаваться на вход агентской части vector. Агентская часть vector без дополнительных обработок будет отправлять данные по защищенному с помощью SSL каналу связи на центральный сервер vector в формате json. Центральный сервер vector будет получать данные в стандартной структуре, после чего обрабатывать их и отправлять в хранилище – Clickhouse также, по защищенному каналу связи. На сервере с Clickhouse будет установлена Grafana для визуализации данных, информации о мониторинге Clickhouse и т.д. В рамках данного стенда Grafana не будет использовать SSL для ускорения работы, так как расположена на одной машине с Clickhouse.

Архитектура практической реализации представлена на рисунке 1

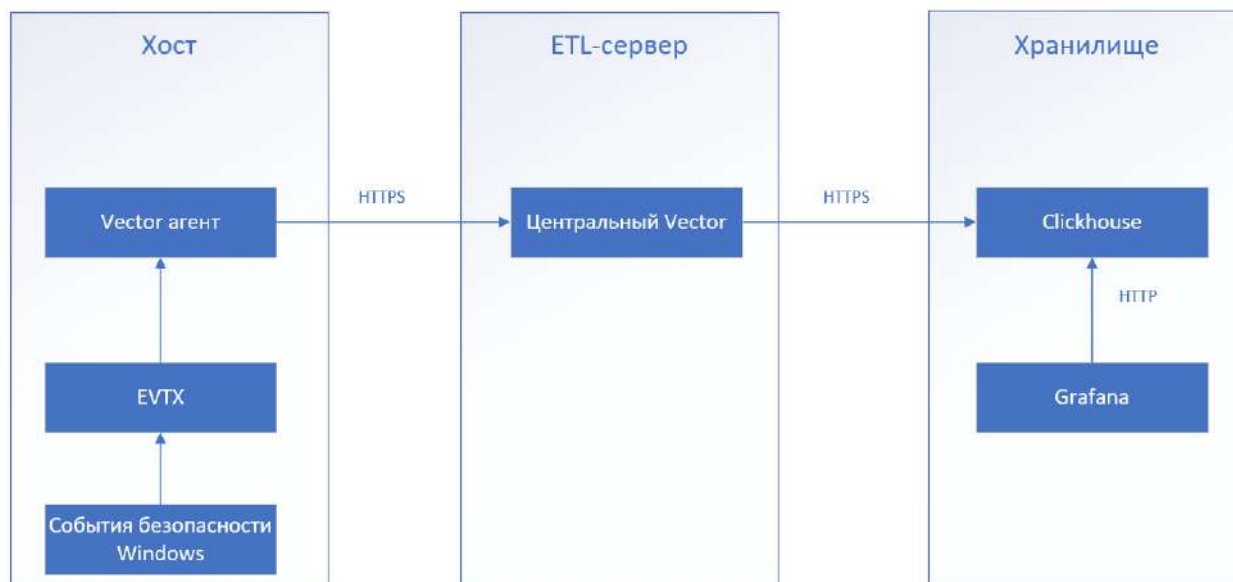


Рис. 1. Архитектура хранилища

Для распределения данных из текста события был разработан парсер на языке vector remap language (VRL), целью которого является обработать получаемой в формате json событие и распределить данные по колонкам в СУБД Clickhouse. Так как стенд является тестовым и преследует цель исследовать эффективность хранилища, а не разработать конечный вариант для внедрения, то в парсере будем вычленять события безопасности windows со следующими eventId:

4697 – в системе была установлена новая служба. Это событие фиксируется, когда пользователь или система создают новую службу или модифицируют существующую.

4648 – успешная попытка входа в систему с использованием явных учетных данных. Это событие возникает, когда процесс пытается войти в систему под определенной учетной записью, указывая ее явные данные. Как правило, это происходит в сценариях, когда процесс выполняется от имени другого пользователя, например, при запуске программ в режиме "Запуск от имени".

4688 – создание нового процесса

4625 – неудачная попытка входа в систему на локальном компьютере

4624 – успешный вход пользователя в систему.

Также, для создания искусственной нагрузки на серверную часть Vector и СУБД журнал на конечном устройстве будет циклично перечитываться с начала, то есть фактически, данный одной машины будут загружаться множество раз в СУБД.

В результате, был получен объем данных в 121.7 млн строк. С использованием Clickhouse данные заняли 3.2 ГБ на диске. Был получен и обработан стабильный поток данных, скорость записи в хранилище составила 3500 событий в секунду. Задержка агрегационного запроса `select count(*)` в хранилище Clickhouse составила 92 мс.

На идентичных аппаратных мощностях был развернут стенд с наиболее популярным стеком технологий для хранения логов, а именно ELK-стек. Данный стек состоит из набора инструментов Elasticsearch (хранилище), Logstash – ETL-Инструмент и Kibana – инструмент визуализации данных.

Архитектурная схема для данного стенда была идентична предыдущей, за исключением того, что на конечном устройстве данные собирались с помощью Filebeat. Соответственно роль ETL-инструмента выполнял Logstash, хранилища – Opensearch, инструмента визуализации – Kibana.

На этом стенде была проделана идентичная работа по сбору и отправке в Opensearch (являющийся проектом с открытым исходным кодом, основанным на Elasticsearch) большого количества данных, в результате чего был получен такой же объем данных в 121.7 млн строк.

Результаты работы показали, что хранения такого же объема данных на диске заняло 60.4 ГБ, что примерно в 19 раз больше, чем в Clickhouse. Задержка агрегационного запроса составила 457 мс, что в 5 раз больше, чем в Clickhouse. Был получен и обработан стабильный поток событий, скорость записи в Opensearch составила 2100 событий в секунду, что приблизительно в 1.7 раза меньше, чем в связке Vector+Clickhouse.

## 6. Заключение

В данной статье решается актуальная задача формирования комплекса технологий для организации хранения корпоративных событий. Предложен метод, позволяющий произвести обоснованный выбор технологий, с учетом ограничений на совместимость и различных показателей эффективности их функционирования, что составляет новизну относительно аналогов. По результатам натурального эксперимента было получено улучшение значений, относительно одного из самых распространенных технологических стеков для работы с событийными данными. Данные улучшения составляют: эффективность хранения (занятое место на жестком диске) в 19 раз, скорость записи данных в 1.7 раза, задержка агрегационного запроса в 5 раз, что подтверждает целесообразность применения разработанного метода.

## Литература

1. *Islam M.R., Rahman M.A.* LogStamping: a blockchain-based log auditing approach for large-scale systems [Электронный ресурс]. – 2025. – URL: [https://www.researchgate.net/publication/392085649\\_LogStamping\\_A\\_blockchain-based\\_log\\_auditing\\_approach\\_for\\_large-scale\\_systems](https://www.researchgate.net/publication/392085649_LogStamping_A_blockchain-based_log_auditing_approach_for_large-scale_systems) (дата обращения 03.08.2025).
2. *Pourmajidi A., Takabi H.* Immutable Log Storage as a Service (Logchain) [Электронный ресурс] – 2019 – URL: [https://www.researchgate.net/publication/335364012\\_Immutable\\_Log\\_Storage\\_as\\_a\\_Service](https://www.researchgate.net/publication/335364012_Immutable_Log_Storage_as_a_Service) (дата обращения 03.08.2025).
3. *Ahmad I., Shah M.A., Wahid A., Maple C.* Secure and Transparent Audit Logs with BlockAudit [Электронный ресурс] – URL: [https://www.researchgate.net/publication/334668830\\_Secure\\_and\\_Transparent\\_Audit\\_Logs\\_with\\_BlockAudit](https://www.researchgate.net/publication/334668830_Secure_and_Transparent_Audit_Logs_with_BlockAudit) (дата обращения 03.08.2025).
4. *Thazhath R., Mo Y., Chaki S., Hsiao M.* Harpocrates: Privacy-Preserving and Immutable Audit Log for Sensitive Data Operations [Электронный ресурс]. – 2022. – URL: <https://arxiv.org/abs/2211.04741> (дата обращения 03.08.2025).
5. *Cao Y., Tang Y., Yu J.X., et al.* LogStore: A Cloud-Native and Multi-Tenant Log Database [Электронный ресурс] // Proceedings of the 2021 ACM SIGMOD International Conference on Management of Data. – 2021. – С. 2672–2685. –

- URL: [https://www.researchgate.net/publication/352526159\\_LogStore\\_A\\_Cloud-Native\\_and\\_Multi-Tenant\\_Log\\_Database](https://www.researchgate.net/publication/352526159_LogStore_A_Cloud-Native_and_Multi-Tenant_Log_Database) (дата обращения 03.08.2025).
6. ResearchPublish. An In-Depth Review of Blockchain for Logging and Auditing [Электронный ресурс]. – 2025. – URL: [https://www.researchgate.net/publication/352526159\\_LogStore\\_A\\_Cloud-Native\\_and\\_Multi-Tenant\\_Log\\_Database](https://www.researchgate.net/publication/352526159_LogStore_A_Cloud-Native_and_Multi-Tenant_Log_Database) (дата обращения 03.08.2025).
  7. *Аристова Е.М.* Установление взаимосвязи между методами аддитивной свертки и метрики // Вестник ДГТУ. Технические науки. – 2017. – № 2. [Электронный ресурс]. – URL: <https://cyberleninka.ru/article/n/ustanovlenie-vzaimosvyazi-mezhdu-metodami-additivnoy-svertki-i-metriki> (дата обращения 03.08.2025).
  8. *Котенко И.В., Кулешов А.А., Ушаков И.А.* Система сбора, хранения и обработки информации и событий безопасности на основе средств Elastic Stack, Тр. СПИИРАН, 2017. – Выпуск 54. – С. 5–34.
  9. *Безрук В.М.* Методы решения многокритериальных задач оптимизации информационных систем // Радиоэлектроника и информатика. – 1999. – № 2 (7). [Электронный ресурс]. – URL: <https://cyberleninka.ru/article/n/metody-resheniya-mnogokriterialnyh-zadach-optimizatsii-informatsionnyh-sistem> (дата обращения 03.08.2025).
  10. *Гольдштейн А.Л.* Многокритериальный генетический алгоритм // Вестник ПНИПУ. Электротехника, информационные технологии, системы управления. – 2013. – № 8. [Электронный ресурс]. – URL: <https://cyberleninka.ru/article/n/mnogokriterialnyy-geneticheskiy-algoritm> (дата обращения 03.08.2025).