



СЕВАСТОПОЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ

Выпускная квалификационная работа
на тему:

Исследование методов оптимизации распределенного хранилища данных для системы электронного документооборота

магистранта группы ИС/м-ІІІ-І-з
Мжачева Ильи Александровича

г. Севастополь
2021

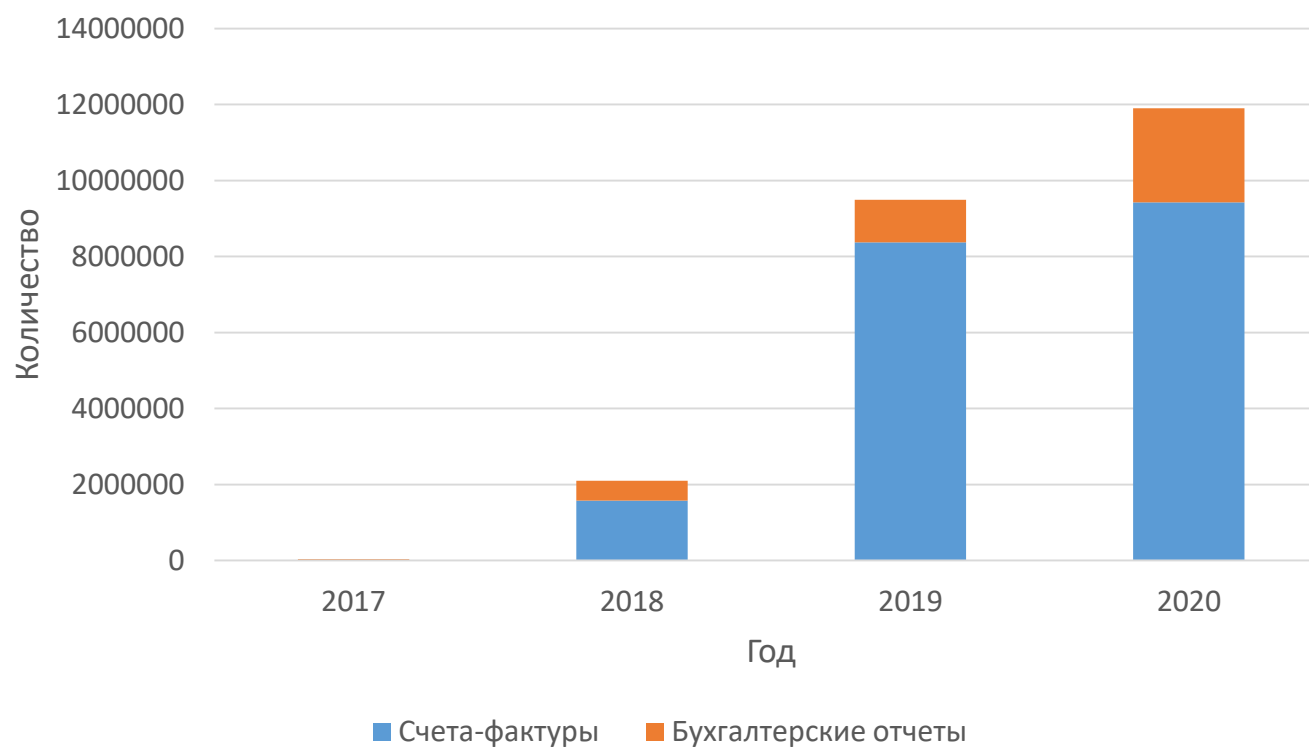
Актуальность

На сегодняшний день системы электронного документооборота (СЭД) встречаются повсеместно. Существует большое количество СЭД, которые могут решать как индивидуальные задачи отдельно взятых предприятий, так и могут решать задачи государственного уровня.

Оператор электронного документооборота – организация, обладающая достаточными технологическими, кадровыми и правовыми возможностями для обеспечения юридически значимого документооборота счетов-фактур в электронном виде и использованием электронной подписи.

Неотъемлемой частью любой системы электронного документооборота является система хранения данных (СХД). С Каждым годом, объем хранимой информации операторами электронного документооборота только увеличивается, в следствии чего, время доступа к данным уменьшается, стоимость поддержки увеличивается.

Статистика системы-оператора электронного документооборота



Цели и задачи

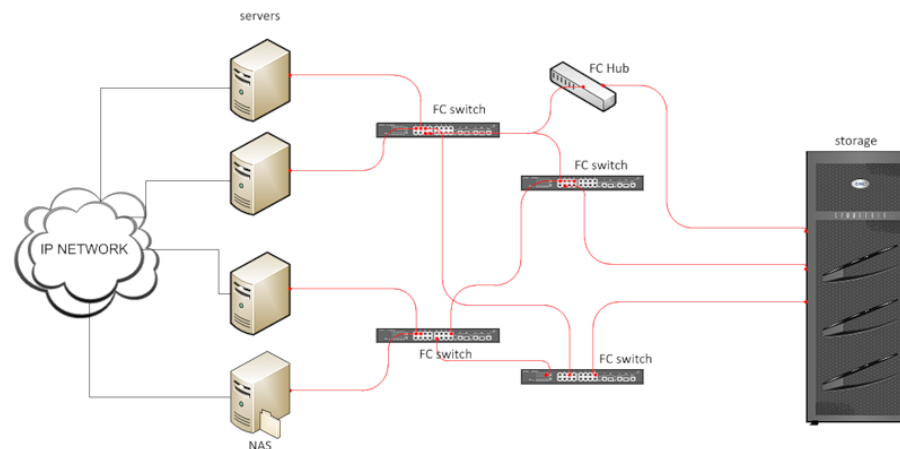
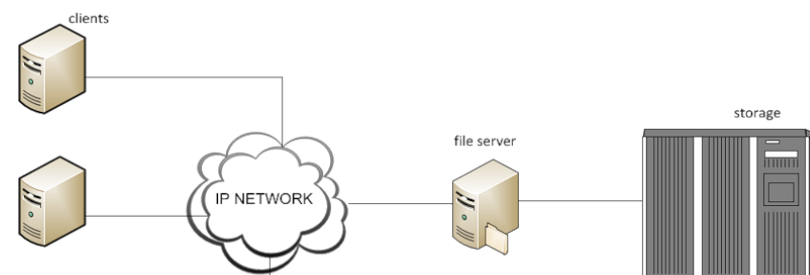
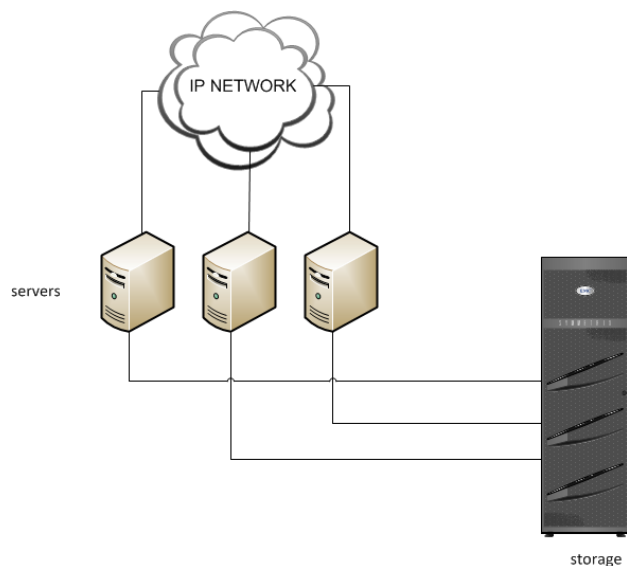
Цель – создание оптимизированной системы хранения данных для системы оператора электронного документооборота, которая будет соответствовать законодательству Российской Федерации, обладает высокими показателями производительности, масштабируемости и надежности.

Задачи:

- рассмотреть существующие подходы к организации систем хранения данных;
- провести анализ предметной области и методов оптимизации хранения данных;
- разработать архитектуру системы хранения данных, отвечающей требованиям системы оператора электронного документооборота;
- разработать систему хранения данных, провести верификацию и тестирование.

Подходы к организации систем хранения данных

- **DAS** (Direct Attached Storage);
- **NAS** (Network Attached Storage);
- **SAN** (Storage Area Network);
- **SDS** (Software-Defined Storage).



Подходы к организации систем хранения данных



Яндекс Облако

Анализ предметной области



Анализ предметной области

Согласно приказу ФНС РФ от 23.10.2020 № ЕД-7-26/775@ определены следующие требования к хранению документов операторами электронного документооборота:

- счета-фактуры, договора, накладные, платежные документы, регистры бухучета – подлежат хранению на протяжении 5 лет;
- бухгалтерская отчетность храниться до тех пор, пока оператор ЭДО осуществляет деятельность;
- документы, как минимум, должны быть сохранены в двух экземплярах на разных физических носителях информации;
- обеспечение надежного режима хранения документов, так, чтобы они не были утрачены, несанкционированным образом распространены, уничтожены или искажены;
- соблюдение федерального закона № 152-ФЗ «О персональных данных».

Методы оптимизации объема

Целью оптимизации по критерию объема данных является увеличение объема хранимых данных и уменьшение занимаемого дискового пространства на запоминающих устройствах. К таким методам можно отнести:

- метод сжатия данных;
- метод дедупликации данных на уровне файлов;
- метод дедупликации данных на уровне блоков данных.

Методы оптимизации надежности

Целью оптимизации по критерию надежности хранения данных является уменьшение риска потери данных, в случае нештатной ситуации (отказ запоминающего устройства). К методам оптимизации надежности хранения данных можно отнести:

- метод синхронной репликации данных;
- метод асинхронной репликации данных;
- метод резервного копирования данных;
- метод помехоустойчивого кодирования данных.

Методы оптимизации производительности

Целью оптимизации по критерию производительности является уменьшение времени доступа к данным, добавления данных в хранилище. К таким методам относятся следующие:

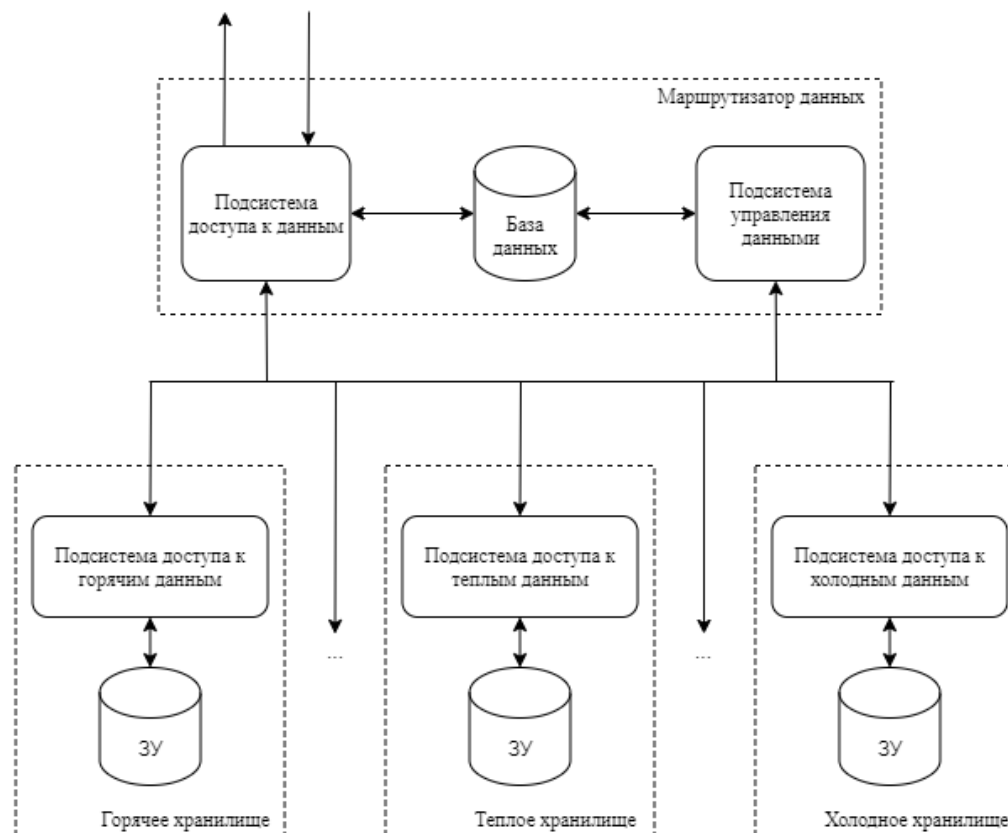
- метод кэширования данных;
- метод шардинга данных.

Методы оптимизации безопасности

Целью оптимизации, с точки зрения безопасности хранения данных, является уменьшение рисков несанкционированного доступа к данным. К таким методам относятся:

- метод симметричного шифрования данных;
- метод асимметричного шифрования данных;
- метод аудита доступа;
- метод ограничения сетевого доступа.

Архитектура системы хранения данных



Подсистема управления данными

Алгоритм определения «температуры» данных

$$A = \max\left(\frac{T_m - T_c}{T_m}, 0\right).$$

где T_m - максимальное время востребованности данных;
 T_c - время, прошедшее с момента добавления данных.

$$P = \frac{V_t}{V}.$$

где V - общее количество обращений к данным;
 V_t - количество обращений к данным за определенный промежуток времени.

$$T = P * A.$$

где P - коэффициент востребованности данных;
 A - коэффициент старения данных.

Реализация системы хранения данных

Выбор технологических средств



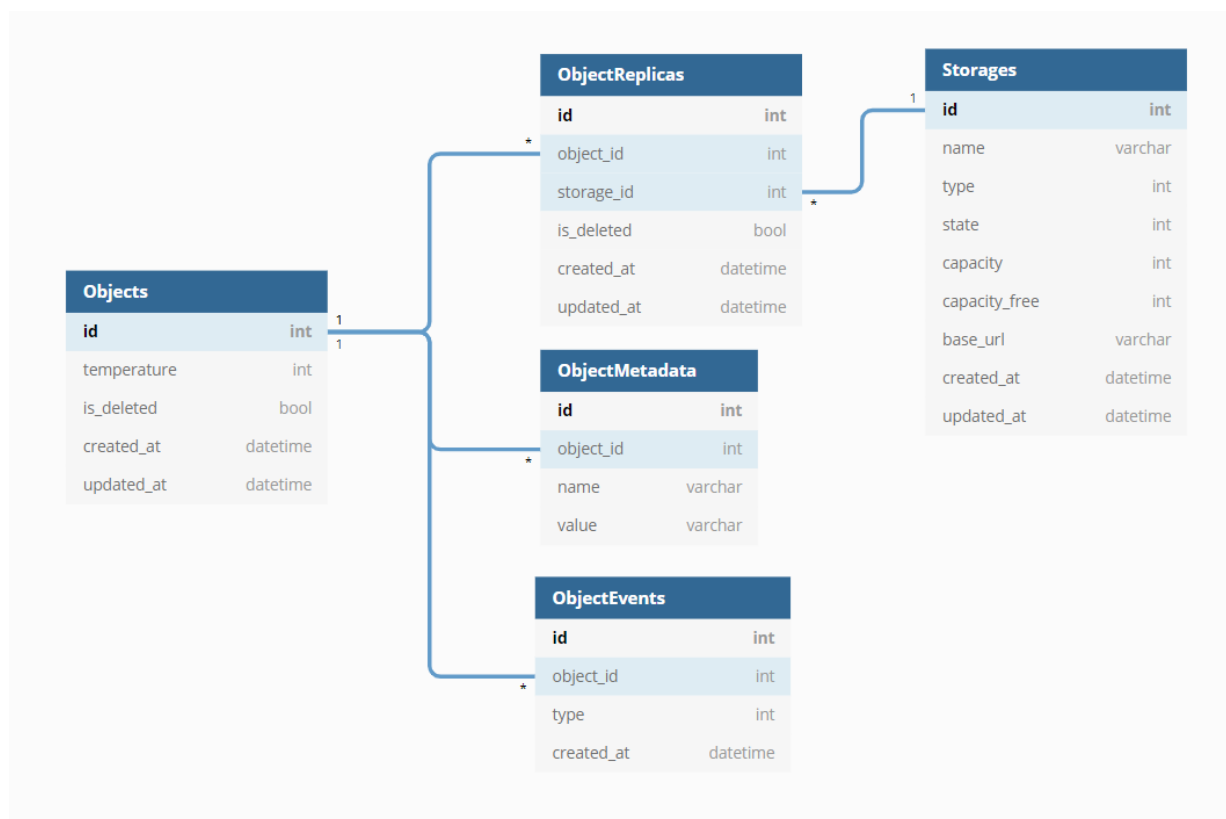
PostgreSQL



Hangfire

Реализация системы хранения данных

Структура базы данных



Реализация системы хранения данных

Верификация и тестирование

В ходе проведения верификации и тестирования разработанной системы хранения данных для системы-оператора электронного документооборота было выявлено:

- прирост производительности на **27%**;
- использование объемов памяти запоминающих устройств на **14%** эффективней;
- масштабирование объема памяти хранилища дешевле на **43%**;
- надежность хранения соответствует требованиям;
- безопасность хранения соответствует требованиям.

СПАСИБО
ЗА ВНИМАНИЕ