

# **Метод резервного копирования с контролируемым размером избыточной информации в распределенном файловом хранилище**

Студент: Варин Дмитрий Владимирович

Научный руководитель: Кузнецов Дмитрий Александрович

Москва

2023

# Цель и задачи

Цель работы – разработка метода резервного копирования с контролируемым размером избыточной информации в распределенном файловом хранилище и его программная реализация.

## Задачи:

- провести анализ существующих методов резервного копирования;
- разработать метод резервного копирования с контролируемым размером избыточной информации в распределенном файловом хранилище;
- разработать программное обеспечение, реализующее данный метод;
- исследовать эффективность разработанного метода и выявить количество избыточной информации, хранящейся при резервировании файлов.

# Классификация резервного копирования по способу хранения данных

Характеристика	Полное	Инкрементное	Дифференциальное
<i>Объем данных</i>	Большой	Больше полного	Больше инкрементного
<i>Время копирования</i>	Длительное	Малое	Среднее
<i>Восстановление данных</i>	Возможно	Невозможно без полного копирования	Невозможно без полного копирования
<i>Гибкость восстановления</i>	Все данные в одной копии	Зависимость от последовательности копий, риск потери копий	Зависимость от последовательности копий, риск потери копий
<i>Частота выполнения</i>	Единоразово	Периодически	Периодически

# Классификация резервного копирования по месту хранения данных

Характеристика	Локальное копирование	Удаленное копирование
Скорость доступа	Высокая	Ограничена интернет-соединением
Надежность	Уровень надежности зависит от используемых носителей	Высокий уровень надежности
Масштабируемость	Ограничена объем хранилища носителей	Поддерживает большие объемы
Восстановление данных	Быстрое с локальных носителей	Восстановление через интернет
Стоимость	Высокие затраты на оборудование	Ежемесячная плата

# IDEF0-диаграмма метода резервного копирования



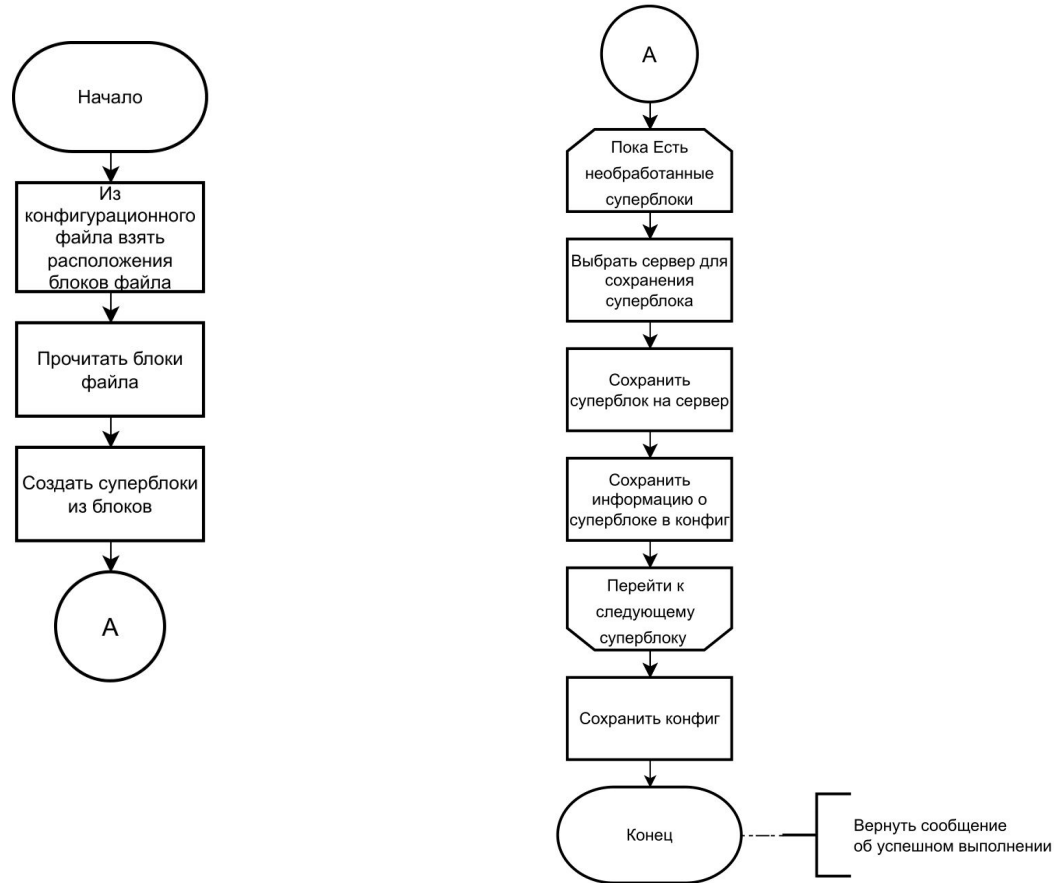
# Механизмы создания избыточности в РФХ

Механизм	Плюсы	Минусы
Репликация	<ul style="list-style-type: none"><li>● Увеличение доступности данных</li><li>● Быстрый доступ к данным</li></ul>	<ul style="list-style-type: none"><li>● Высокие затраты на оборудование</li></ul>
Разделение информации	<ul style="list-style-type: none"><li>● Экономия места</li><li>● Высокая надежность</li></ul>	<ul style="list-style-type: none"><li>● Некорректное разделение может привести к потере данных</li><li>● Увеличение нагрузки на сеть</li></ul>
<b>Избыточные блоки данных</b>	<ul style="list-style-type: none"><li>● Экономия места</li><li>● Высокая надежность</li></ul>	<ul style="list-style-type: none"><li>● Вычислительная сложность</li></ul>

# Характеристика разработанного метода

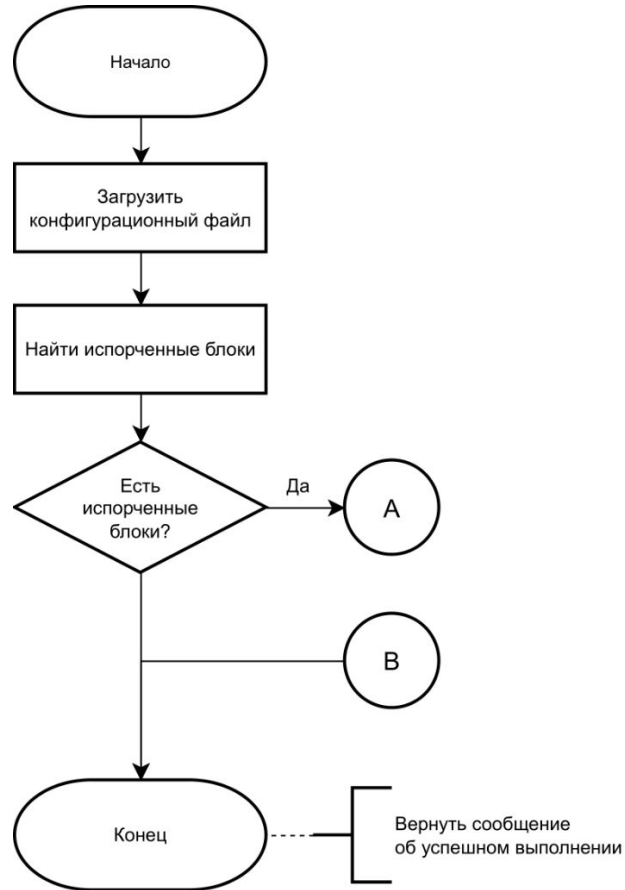
1. Метод применяется в *распределенном файловом хранилище*.
2. По способу хранения и обновления данных будет использоваться механизм полного копирования.
3. По месту хранения данных – удаленное копирование и локальное.
4. Контролируемый размер избыточной информации будет обеспечиваться с помощью механизма создания избыточных блоков данных (кодирование с использованием **XOR**).

# Алгоритм создания суперблока

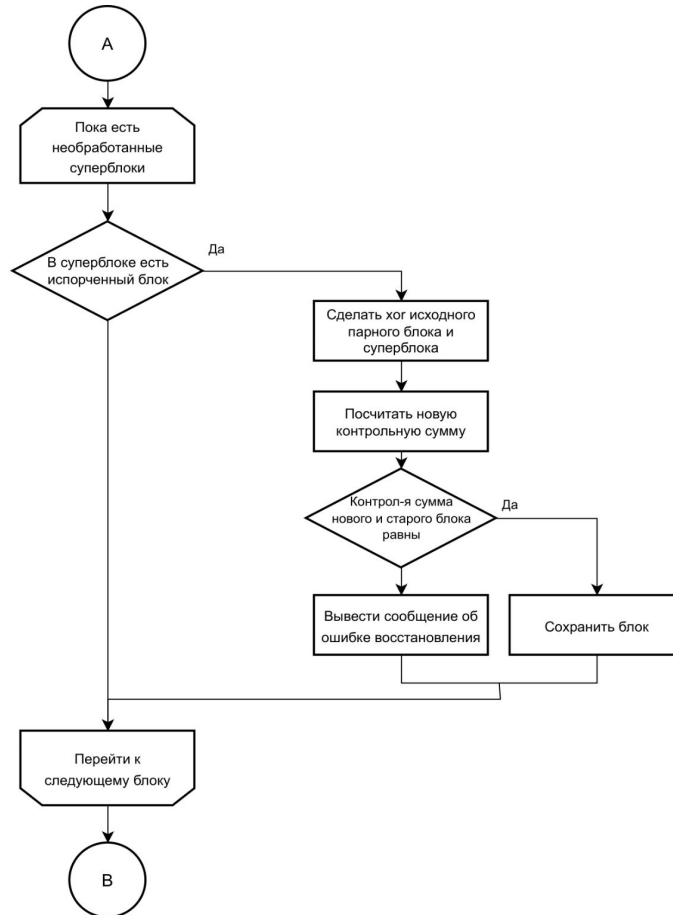




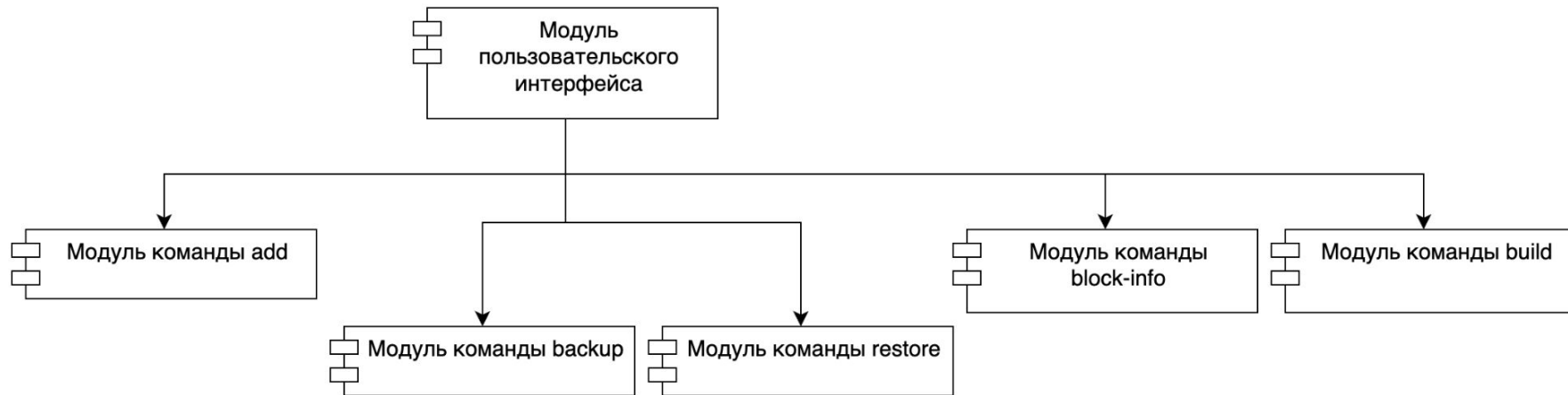
# Алгоритм восстановления суперблока – часть 1



# Алгоритм восстановления суперблока – часть 2



# Диаграмма компонентов



# Исследование изменения размера избыточной информации от размера суперблока

Алгоритмы сжатия:

- GZIP;
- ZSTD;
- LZMA.

Типы файлов:

- текстовые *.txt*;
- документы *.pdf*;
- изображения *.jpg*;
- видео *.mp4*, *.mov*.

Этапы исследования:

1. Разбить файл на блоки.
2. Создать резервную копию.
3. Сжать резервную копию.
4. Посчитать, какой процент от исходной резервной копии составляет сжатая.

# Исследование изменения размера избыточной информации от размера суперблока

## Текстовые файлы и документы

Название	Исходный размер (Байт)	% с LZMA	% с GZIP	% с ZSTD
book.txt	3 202 332	64.76	62.87	63.67
book1.pdf	63 950 013	76.11	75.93	75.96
book.pdf	136 542 190	88.72	88.13	88.46

## Видеозаписи

Название	Исходный размер (Байт)	% с LZMA	% с GZIP	% с ZSTD
video.mp4	3 891 031	0.29	0.24	0.08
video1.mp4	69 075 086	0.05	0.02	0.04
<i>video2.mov</i>	63 639 323	26.71	3.25	20.61

# Результаты исследования

1. Сжатие суперблоков позволяет уменьшить их размер:
  - в худшем случае размер суперблока равен исходному;
  - в лучшем – уменьшится на десять и более процентов.
2. Сжатие хорошо работает на файлах, имеющих малый процент сжатия, таких как текстовые файлы, документы, некоторые форматы видео, например, mov.
3. Добавление в разработанный метод алгоритма сжатия позволит уменьшить размер суперблоков и сделает метод более эффективным по памяти.

# Заключение

**Цель работы была достигнута** – был разработан *метод резервного копирования с контролируемым размером избыточной информации в распределенном файловом хранилище.*

Все поставленные **задачи выполнены:**

- проведен анализ существующих методов резервного копирования;
- разработан метод резервного копирования с контролируемым размером избыточной информации в распределенном файловом хранилище;
- разработано программное обеспечение, реализующее данный метод;
- исследована эффективность разработанного метода и выявлено количество избыточной информации, хранящейся при резервировании файлов.

## Дальнейшее развитие

- добавить альтернативу консольному интерфейсу – графический;
- интегрировать метод в распределенное файловое хранилище.