



РЕКОМЕНДАЦИИ ПО РАСЧЕТУ ЦЕЛЕВЫХ УСТРОЙСТВ ДЛЯ СИСТЕМЫ РЕЗЕРВНОГО КОПИРОВАНИЯ.

Как рассчитать необходимую емкость целевых устройств хранения резервных копий (дедупликатора) и ленточной библиотеки. Для обеспечения хранения резервных копий в течение года для 200 ТБ пользовательских данных с использованием методологии ITIL и правил хранения 3-2-1 (с учетом airgap).

Описание подходов к хранению резервных копий.

Предлагаемая архитектура резервного копирования соответствует принципам ITIL и классической схеме GFS (Grandfather-Father-Son) в её оптимизированной реализации для современных систем с дедупликацией и ленточным хранением.

1. Разделение уровней хранения и целей (методология ITIL):
 - Дисковый дедупликатор обеспечивает быстрое восстановление за счёт оперативного дискового хранилища, обеспечивающее глубину хранения в 42 дня - это соответствует стандартным RTO и RPO целевого уровня SLA.
 - Ленточное хранилище используется для долгосрочного хранения и защиты от киберинцидентов.
2. GFS схема хранения:
 - В базовом варианте применяется ежедневное инкрементальное копирование с дедупликацией на диск (оперативное окно 42 дня).
 - Для ленточной подсистемы предусмотрены два сценария:
 - Годовая политика: 12 месячных и 1 годовая копия (off-site air-gap).
 - Квартальная политика: хранение всех ежедневных копий в течение 90 дней на ленте без извлечения картриджей, с учётом линейного роста данных (15 %/год).
3. Соответствие правилу 3-2-1 (с air-gap):
 - **3 копии данных:** рабочая (production), on-site (дедуп-пул), off-site (air-gap-ленты).
 - **2 типа носителей:** диск и лента.
 - **1 копия вне площадки:** физически изолированные картриджи, обеспечивающие защиту от ransomware, логических сбояв, а также уменьшает риски от внешних катастроф.

- При необходимости допускается on-site ленточный пул для квартального хранения без извлечения носителей - не нарушает логику 3-2-1, оставаясь в зоне локального восстановления.

Исходные данные (можно подставить свои)

- Объём пользовательских данных (Source), **S = 200 ТБ**;
- Суточная изменчивость (daily change rate), **c = 2 % = 0.02**.
 - подразумевается доля уникальных изменений исходных данных (новые + модифицированные блоки) за сутки;
- Окно для оперативного хранения на дисковом дублированном хранилище: **R^{on} = 42 дня**.
 - Для возможности и перекрытия двух бухгалтерских циклов;
- Годовая политика: *дни - на диске, месяцы/год — на ленте*.
- Эффективный коэффициент хранения данных на дедупликаторе (дедупликация + компрессия), **F = 20:1**.
 - коэффициент исходя из консервативного подхода;
- Технологический запас на метаданные/служебные нужды/резерв: **o = 20 %**;
- Коэффициент годового прироста данных, **G_γ = 15 %**;
- Количество полных копий **N_{full} = 12** месячных + 1 годовая = **13 копий**
- Лента: **LTO-10 = 30 ТБ «native»** на картридж.
 - компрессия не учитывается, т. к. резервные копии уже сжаты до записи на ленту.

Формулы и расчёты

1. Дедуплицированное хранилище для оперативного хранения и восстановления резервных копий

I. Уникальный логический объем данных за окно хранения на дедупликаторе:

$$U_{\{unique\}} = S \times (1 + c) \times R^{\{on\}} \times (1 + g)$$

S - исходный объем данных,
 c - суточная изменчивость (уникальные изменения),
 R^{on} - окно хранения (в днях),
 g - коэффициент роста (в долях).

II. Физическая емкость дедупликатора:

$$C_{\{unique\}} = \frac{U_{\{unique\}}}{F} \times (1 + o)$$

Пример:

$$U_{\{unique\}} = 200 \times (1 + 0.02 \times 42) \times (1 + 0.15) = 423.2 \text{ ТБ}$$

368 ТБ уникальных логических данных с учетом годового роста в 15%

$$C_{\{dedupe\}} = \frac{368}{20} \times 1.2 = 25.392 \text{ ТБ}$$

Таким образом, для 200ТБ исходных данных, коэффициента суточной изменчивости 2% при коэффициенте дедупликации 20 с учетом округления имеем 26 ТБ.

2. Годовое хранение на лентах offsite:

Принята классическая GFS-схема: 12 ежемесячных полных копий + 1 годовая (12М + 1Y). Дополнительно классической подход 3-2-1 подразумевает физическое извлечение картриджей из ленточной библиотеки и хранение в изолированной от текущей ИТ инфраструктуры, что закрывает вопрос катастрофоустойчивости для площадки.

I. Логический объем, хранимый на ленте (с учетом роста данных):

$$U_{\{tape\}} = S \times (1 + G_{\{y\}}) \times N_{\{full\}}$$

где $N_{\{full\}}$ это количество полных копий (12 месяцев + 1 годовая).

II. Количество картриджей исходя из объема 1 картриджа LTO-10 в 30 ТБ:

$$N_{\{tape\}} = \frac{U_{\{tape\}}}{30} \times N_{\{service\}}$$

Где $N_{\{service\}}$ это значение для картриджей типов scratch/spare, которое считается из расчёта в 10 процентов от целевого количества.

Пример:

$$U_{\{tape\}} = 200 \times 1.15 \times 13 = 2990 \text{ ТБ} = 2.99 \text{ ПБ}$$

$$N_{\{tape\}} = \frac{2990}{30} \times 1.1 = 109.67 \text{ картриджей} \cong 110 \text{ картриджей}$$

Получаем 110 картриджей на годовую 12 ежемесячных и 1 годовую копию.

3. Годовое хранение на лентах onsite в хранилище включая ежедневные копии (опционально):

В ряде инфраструктур реализуется резервное копирование, при котором все данные, включая ежедневные копии, сохраняются на ленточных носителях. Картриджи в этом случае не извлекаются из библиотеки, а хранятся в хранилище и перезаписываются по окончании установленного срока хранения.

Период хранения принимается равным одному кварталу (90 дней).

I. Логический объем данных на лентах с учетом ежегодного роста данных в 15%:

$$U_{\{tape_{90}\}} = S + (S \times c \times 90) \times (1 + 0.5 \times G_{\{y\}} \times \frac{90}{365})$$

II. Количество картриджей исходя из объема 1 картриджа LTO-10 в 30 ТБ:

$$N_{\{tape_90\}} = \frac{U_{\{tape_90\}}}{30} \times N_{\{service\}}$$

Пример:

$$U_{\{tape_90\}} = 200 + (200 \times 0.02 \times 90 \times (1 + 0.5 \times 0.15 \times 90/365)) = 566.7 \text{ ТБ}$$

$$N_{\{tape_90\}} = \frac{566.7}{30} \times 1.1 = 20.8 \text{ картриджей} \cong 21 \text{ картридж};$$

Таким образом необходим 21 картридж исходя из хранения ежедневных резервных копий в течение 90 дней.

Короткий вывод

Для 200 ТБ исходных данных, 42-дневного окна хранения на дублицированном хранилище(дедупликаторе), 1 годовой и 12 ежемесячных полных копий, а также ежедневных копий на ленты, с учётом 15% роста данных в год.

- При коэффициенте эффективного хранения 20:1 требуется ≈ 26 ТБ физической ёмкости на дедупликаторе.
- Для хранения 12 полных и 1 годовой резервных копий требуется 110 картриджей LTO-10.
- Для хранения всех ежедневных копий на лентах в течение 90 дней требуется ≈ 567 ТБ логически, 21 картридж LTO-10.
- Архитектура соответствует правилу 3-2-1 (с air-gap) и принципам ITIL/GFS.
- 15% рост данных в год считается из консервативного подхода к оценке роста данных
- Рекомендуется закладывать увеличенные объёмы хранения (ленты и дисковое хранилище) - с запасом не менее 30%, с учётом специфики бизнес-систем и индивидуальных требований заказчика.