# Описание функциональных характеристик Arenadata Hadoop Platform T4 (ADH T4)

Москва
2024

## Содержание:

# 1 Требования к программному обеспечению

| Компонент | Требования |
|---|---|
| Платформа | Intel x86_64 |
| Операционная система | • Centos 7.x.<br>• RHEL 7.6+.<br>• AltLinux 8 SP (версия ADH Enterprise).<br>Для AltLinux 8 минимальная версия ADH — 2.1.6.b1.<br>• AstraLinux 1.7 "Орел" SE с Axiom JDK (версия ADH Enterprise).<br>Минимальная версия ADH — 3.2.4. Параметру LC_TIME должно быть присвоено значение en_US.utf8. |
| Браузер | • Internet Explorer<br>• Firefox<br>• Google Chrome<br>• Safari |
| Программное обеспечение | • RPM + YUM / DPKG + APT<br>• scp<br>• curl<br>• unzip<br>• tar |
| Java | OpenJDK 8u252 или более поздняя |

# 2 Требования к оборудованию

Планирование оборудования не может быть полноценным без прогнозирования рабочей нагрузки. Когда вы планируете кластер Hadoop, очень важно как можно точнее оценить объем данных, а также тип и количество задач. При проведении пилотного проекта применяйте различные метрики для оценки реальной нагрузки на кластер. Это позволит в дальнейшем масштабировать пилотную среду, избегая критических изменений в существующей инфраструктуре.

Количество узлов и их спецификация зависит от нескольких факторов:

- общий объем данных;
- политика хранения данных (по умолчанию коэффициент репликации равен 3);
- тип предполагаемой нагрузки;
- способы хранения данных (контейнерные данные, использование сжатия).

Каждый кластер Hadoop содержит, по крайней мере, следующие типы узлов:

- NameNode
- DataNode
- YARN

Обратите также внимание на обеспечение пилотного кластера.

## 2.1 Электропитание и потребление

Потребляемая мощность представляет собой важнейший фактор при разработке кластера. Прежде чем приобретать самые большие и быстродействующие физические серверы для реализации узлов кластера, проведите анализ мощности, потребляемой вашим установленным оборудованием. Мы наблюдаем существенную экономию стоимости и потребляемой энергии, если избегать приобретения самых быстрых процессоров, резервных источников питания и другого оборудования.

В настоящее время поставщики предлагают оборудование для облачных центров обработки данных, предусматривающего снижение цены, потребляемой мощности и веса. Компании Supermicro, Dell и HP поставляют подобные линейки продуктов для провайдеров облачных услуг. Поэтому при покупке оборудования для больших кластеров обращайте внимание на такие минимальные конфигурации "облачных серверов".

Для узла DataNode достаточно единственного источника питания, но для узла NameNode используйте вариант с резервированием. Решение с использованием общих источников питания для нескольких серверов повышает надежность без существенного увеличения стоимости.

Некоторые центры размещения оборудования (co-location sites) требуют оплату из расчета максимальной потребляемой мощности, не учитывая реальную потребленную энергию. В таких местах экономичность по мощности последних версий процессоров не очень заметна. Поэтому рекомендуем вам ознакомиться с ценообразованием в таких центрах заранее.

## 2.2 Компьютерная сеть

Это также одна из наиболее значимых систем, поскольку нагрузка в Hadoop очень изменчивая. Очень важны разумные затраты при обеспечении достаточной скорости взаимодействия между узлами. Типовым является подключение каждого узла к коммутатору в рамках одной стойки на 20 позиций с помощью двухканального интерфейса по 1 Гбит/с и межсоединение (interconnect) 2*10 Гбит/с всех стоек через пару центральных коммутаторов.

Хороший сетевой проект принимает в расчет возможность превышения допустимой пропускной способности в критических узлах сети под реальной нагрузкой. Общепринятые коэффициенты превышения возможности сети (oversubscription ratios) составляют приблизительно 4:1 на уровне доступа к серверу и 2:1 между уровнем доступа и уровнем агрегации или ядром сети. При высоких требованиях к производительности снижайте эти коэффициенты. Кроме того, между стойками возможно превышение на 1 Гбит/с.

Критически важно установить выделенные коммутаторы для кластера вместо попыток организовать виртуальные соединения (virtual connect) в уже существующих коммутаторах. В противном случае нагрузка от кластера Hadoop будет влиять на остальных пользователей этого оборудования. Также критически важно работать совместно с командой обслуживания сети при выборе коммутаторов, удовлетворяющих требованиям и со стороны Hadoop, и со стороны средств мониторинга.

Проектируйте сеть таким образом, чтобы оставалась возможность добавлять стойки с серверами для Hadoop. Исправление сетевых ошибок стоит дорого. Заявленная пропускная способность коммутатора аналогична возможности покрытия расстояния автомобилем (измеряемой в километрах на литр) в том, что вы вряд ли достигнете заявленного показателя. В коммутаторах глубокая буферизация (Deep buffering) предпочтительнее уменьшения задержки. Применение больших пакетов (Jumbo Frames) в пределах кластера повышает пропускную способность за счет использования более эффективных контрольных сумм и может также улучшить обеспечение целостности пакетов.

## 2.3 Требования к оборудованию в зависимости от типа ноды

### 2.3.1 NameNode

Для обеспечения высокой доступности кластера планируйте два узла NameNode: первичный (primary) и вторичный (secondary). Они являются критически важными частями кластера Hadoop. Важно обеспечить надежный доступ к этим узлам. Оба сервера хранят состояние HDFS в файле *fsimage* и журнал операций в файле *edits*.

NameNode выполняет следующие действия:

- операции с файлами в HDFS;
- разделение файлов на блоки, находящиеся в узлах DataNode;
- хранение метаданных о файлах и папках HDFS;
- хранение информации о расположении блоков в узлах DataNode;
- управление репликацией данных.

Вторичный узел NameNode используют в качестве резервного хранилища метаданных HDFS в файле *fsimage* и операций в файле *edits*. Поскольку он менее всего нагружен, он периодически обновляет файл *fsimage* с помощью журнала *edits*. При этом происходит очищение журнала, предотвращающее таким образом бесконтрольное увеличение последнего.

#### 2.3.1.1 Средства хранения

В обоих узлах NameNode необходимо иметь надежное хранилище для метаданных и журнала операций. Обычно аппаратный массив RAID или надежное хранилище в сети оправданы для этой цели.

Для узлов NameNode, независимо от количества узлов DataNode, характеристики хранилища неизменны. Устанавливайте четыре дисковых устройства SAS объемом около 1 ТБ каждый вместе с контроллером RAID HDD, настроенным на RAID 1+0. Устройства SAS дороже SATA и обладают меньшей емкостью, но они быстрее и гораздо надежнее.

Объединение устройств SAS в массив RAID обеспечивает сервисы управления Hadoop достаточно надежным средством хранения с резервированием критически важных данных.

#### 2.3.1.2 Оперативная память

Требования по памяти зависят существенно от размера кластера Hadoop. Оперативная память (RAM) является критическим фактором для обоих типов узлов, поскольку они используют RAM как для хранения метаданных HDFS, так и для обработки запросов. По этой причине используйте память с коррекцией ошибок (Error Correction Code, ECC). Обычно каждому узлу NameNode необходимо от 64 ГБ до 128 ГБ памяти.

Требование к объему памяти прямо зависит от количества файловых блоков в HDFS. Как правило, NameNode занимает примерно 1 ГБ памяти на миллион блоков HDFS.

HDFS реплицирует не файлы, а блоки, образуя несколько копий (обычно 3) каждого блока.

#### 2.3.1.3 Процессоры

Рекомендуем материнские платы с двумя сокетами под процессоры, каждый процессор с восемью ядрами и тактовой частотой 2.5-3 ГГц. Обычно используют архитектуру Intel.

#### 2.3.1.4 Подключение к сети

Высокоскоростное подключение к сети крайне необходимо для узлов NameNode. Поэтому рекомендуем использовать пару связанных (bonded) каналов по 10 Гбит/с. Она обеспечивает достаточную избыточность и также увеличивает пропускную способность до 20 Гбит/с. Для кластеров небольших размеров (менее 50 узлов) достаточно соединения 1 Гбит/с.

### 2.3.2 DataNode

Узлы DataNode являются главными средствами хранения данных и обеспечения вычислительными ресурсами. Многие полагают, что по этой причине для них необходимо наилучшее оборудование. Однако в основу Hadoop положен принцип использования этих узлов в качестве расходного материала. В этом смысле их производительность должна быть достаточной для выполнения задач кластера, но при невысокой стоимости, так чтобы при выходе из строя их можно было заменить на другие без значительных затрат. Разработчики Hadoop принимают во

внимание частоту отказов в больших кластерах как одну из наиболее важных характеристик. В Hadoop решение проблемы отказа оборудования перенесено с оборудования на программное обеспечение.

Узел DataNode выполняет следующие операции:

- запись и чтение блоков данных по запросам от клиентов;
- выполнение задач MapReduce;
- посылка периодических контрольных пакетов (heartbeat) и отчетов о состоянии блоков (block reports) в узел NameNode;
- участие в процессе репликации блоков данных.

#### 2.3.2.1 Необходимость в RAID

Hadoop обеспечивает резервирование на различных уровнях. Каждый узел DataNode хранит некоторые блоки файлов HDFS, которые имеют множество реплик на других узлах. Поэтому в случае отказа одиночного узла данные остаются доступными. Кластер допускает даже отказ большего количества узлов в зависимости от выбранной вами настройки. Более того, Hadoop позволяет указать распределение узлов по стойкам с тем, чтобы сохранять реплики данных в серверах, входящих в разные стойки. В таком случае существенно увеличивается вероятность сохранности данных даже при выходе из строя целой стойки (хотя строгая гарантия отсутствует). Такой подход исключает необходимость инвестировать в контроллеры RAID для узлов DataNode.

#### 2.3.2.2 Средства хранения

Обычно один узел DataNode содержит 12-24 дисков с емкостью по 1 ТБ. Вместо объединения их в массив RAID лучше использовать их как простой набор дисков (Just a Bunch of Disks, JBOD). Это обеспечивает более высокую производительность и снижает расходы на оборудование. Не следует беспокоиться по поводу отказа отдельного диска, поскольку HDFS обеспечивает резервирование.

Допустимое количество узлов DataNode зависит от объема оперативной памяти NameNode. Помните, что узел NameNode задействует примерно 1 ГБ памяти на миллион блоков HDFS. По умолчанию размер блока HDFS составляет 128 МБ.

Обычно такие узлы организуют в виде серверов со стандартным форм-фактором (высотой 1U-2U и шириной 19 дюймов) в составе стоек или кабинетов.

#### 2.3.2.3 Оперативная память

Кроме хранения данных, узел DataNode обрабатывает данные, в том числе, исполняет задачи MapReduce. Задания MapReduce разделены на множество задач, которые исполняют параллельно множество узлов DataNode. Чтобы задание получило согласованный логичный результат, все его задачи должны быть выполнены.

#### 2.3.2.4 Процессоры

Рекомендуем материнские платы с двумя сокетами под процессоры, каждый процессор с восемью ядрами и тактовой частотой 2.5-3 ГГц. Обычно используют архитектуру Intel.

#### 2.3.2.5 Подключение к сети

Как и для NameNode, высокоскоростное подключение к сети необходимо для узлов DataNode. Поэтому рекомендуем использовать пару связанных каналов 10 Гбит/с. Эта связанная пара обеспечивает достаточную избыточность и также увеличивает пропускную способность до 20 Гбит/с. Для кластеров небольших размеров (менее 50 узлов) достаточно соединения 1 Гбит/с.

### 2.3.3 YARN

YARN означает Yet Another Resource Negotiator. Основным назначением YARN является предоставление вычислительных ресурсов приложениям, обеспечивающее при этом разделение функций управления ресурсами и планирования заданий.

В основном рекомендации по оборудованию для менеджера ресурсов YARN такие же, как и для NameNode. В случае, если оба узла NameNode выйдут из строя, можно временно использовать узел менеджера ресурсов для активизации на нем NameNode.

### 2.3.4 Ноды PostgreSQL, используемые с Hive Metastore

Планирование размера внешней базы данных для Apache Hive Metastore требует тщательного учета различных факторов, таких как количество таблиц, партиций (partitions), пользователей и общая рабочая нагрузка. Несмотря на то, что определить точный размер внешней базы данных достаточно сложно, следующие рекомендации помогут вам принять обоснованное решение:

- Оценка размера базы данных
  - Количество таблиц. Для каждой таблицы требуется хранилище для схемы и метаданных, которое может быть от нескольких килобайт до нескольких мегабайт на таблицу, в зависимости от сложности схемы таблицы.
  - Количество партиций. Требования к объему памяти для хранения партиций зависят от типа данных ключа партиции и количества столбцов партиции, и в среднем может быть оценено примерно как 1 КБ на партицию.
  - Информация о пользователях и разрешениях. Для этого требуется небольшой объем памяти, обычно несколько килобайт на пользователя.
- Темпы роста данных, сложившиеся исторически и планируемый рост. Учитывайте исторические темпы роста ваших данных и ожидаемый будущий рост. Регулярно отслеживайте размер базы данных и масштабируйте ее по мере необходимости.
- Репликация и резервное копирование. Выделите дополнительный объём хранилища для резервного копирования и репликации, чтобы обеспечить избыточность и надежность данных. Рекомендуется оставлять 50% от общего размера базы данных для резервного копирования.
- Нагрузка. Оцените количество параллельных запросов и пользователей, которые будут обращаться к системе одновременно. Более высокая рабочая нагрузка может потребовать больше ресурсов и дополнительный объём хранилища данных для кеширования, планирования запросов и временного хранения.
- Буфер для оптимизации производительности. Оставьте дополнительное пространство для индексирования, кеширования и

других операций оптимизации производительности. Обычно следует оставлять 20-30% от общего размера базы данных.

Хотя эти рекомендации могут помочь вам оценить размер внешней базы данных для Hive Metastore, крайне важно продолжать отслеживать использование базы данных и соответствующим образом корректировать ее размер.

Системные требования, приведенные ниже, являются минимальными. Целевой сайзинг необходимо рассчитывать исходя из требований вашей организации.

Минимальные требования к оборудованию для хостов внешней базы данных перечислены в таблице ниже.

| **Требование** | **Небольшой кластер** | **Средний кластер** | **Большой кластер** |
|---|---|---|---|
| Количество одновременных подключений (сессий пользователей) | 5-10 | 20-50 | 50+ |
| Процессор | 64bit рекомендуется *, 4+ выделенных ядер | 64bit, 4-8 ядер | 64bit, 8+ ядер<br>Процессоры с большим кешем L3 лучше работают с большим набором данных |
| RAM | 8-16 ГБ | 16-32 ГБ | 32+ ГБ |
| Дисковое пространство<br>(минимум один отдельный диск, SSD, RAID 1 или RAID 10) | 50+ ГБ | 100 ГБ | 200+ ГБ |

* Если вам нужно использовать 32-битную версию сервера, установите в `LDR_CNTRL` значение `MAXDATA=0xn0000000`, где 1 <= n <= 8, перед запуском сервера PostgreSQL. Попробуйте подобрать подходящее значение и параметры *postgresql.conf*, чтобы найти конфигурацию, работающую удовлетворительно.

## 2.4 Требования к оборудованию в зависимости от типа нагрузки

Рабочие нагрузки на Hadoop стоят в основе множества вопросов по управлению ресурсами и возникновению конфликтов. Конфликты могут быть разного рода, например:

- между долгосрочными заданиями, интенсивно использующими ресурсы, и краткосрочными интерактивными запросами;
- между нагрузкой, создаваемой Hadoop, и нагрузкой от других систем, установленных на узлах одного кластера.

Можно выделить три основных типа (шаблона) нагрузки:

- Интенсивное вычисление (Compute intensive);
- Интенсивное хранение (Storage intensive);
- Сбалансированная нагрузка (Balanced).

Бывают ситуации, когда сначала сложно отнести характер нагрузки к одному из перечисленных выше типов. Начальные действия в кластере обычно значительно отличаются от тех заданий, которые будут выполняться в производственной среде. Поэтому мы рекомендуем следовать рекомендациям по настройке на сбалансированную нагрузку в период развития пилотного кластера Hadoop. Затем вы можете пересмотреть настройки в зависимости от реальной нагрузки.

### 2.4.1 Паттерны нагрузки

#### 2.4.1.1 Равномерность нагрузки

Узлы Hadoop подобны колесам транспортного средства, которые передвигают все остальное. Если колеса одинаковые, то и движение будет равномерным, без рывков. Если они разные, то становится сложным добиться равномерного движения. Поэтому мы рекомендуем обеспечить одинаковую конфигурацию всех узлов кластера с минимальными различиями между узлами одного типа, то есть между узлами DataNode и между узлами NameNode.

Характеристики следующих компонентов должны быть одинаковыми для всех узлов кластера:

- Процессоры
- Оперативная память
- Сетевой интерфейс

#### 2.4.1.2 Интенсивное вычисление

Этот тип нагрузки связан с работой процессоров и требует большого количества ядер процессора и большого объема оперативной памяти для хранения данных в процессе их обработки. Такой тип нагрузки характерен для процессов обработки естественной речи и других высоконагруженных систем (High-Performance Computing Cluster, HPCC). Для обеспечения высокой скорости вычислений необходимо устанавливать не менее 10 ядер процессоров в каждый узел.

#### 2.4.1.3 Интенсивное хранение

При таком типе нагрузки рекомендуем больше инвестировать в дисковые устройства внутри каждого узла. Количество узлов и состав дисковых устройств определяется объемом данных для хранения и обработки. Обычно количество устройств хранения одинаково для всех узлов.

Для серверов с низкой нагрузкой (low density server) основной целью является обеспечение низкой стоимости при установке большого количества узлов. 8 ядер процессора в сервере удовлетворяет этим требованиям и предоставляет достаточную мощность обработки. Каждая задача map или reduce выполняется одним ядром, но поскольку некоторое время уходит на ожидание выполнения операций ввода-вывода, имеет смысл планировать превышение нагрузки на ядра процессора (oversubscription). При наличии 8 ядер в узле можно запланировать создание около 12 слотов под задачи map и reduce.

Оптимальный объем жестокого диска 2-3 ТБ. В среднем устанавливают 12 дисков на узел.

#### 2.4.1.4 Сбалансированная нагрузка

Логика выбора оборудования не зависит от того, обладает ли кластер высокой плотностью или низкой плотностью. Например, можно установить в узел 12 дисковых устройств по 2-3 ТБ или 24 устройства по 1 ТБ. Использование дисков с меньшим объемом предпочтительнее, так как это повышает пропускную способность ввода-вывода и отказоустойчивость. Для обеспечения достаточной вычислительной мощности используют 8 ядер процессора и 128-256 ГБ оперативной памяти в каждом узле.

#### 2.4.1.5 Общие рекомендации

Системные требования, приведенные ниже, являются минимальными. Целевой сайзинг необходимо рассчитывать исходя из требований вашей организации.

Следующая таблица содержит рекомендации по подбору оборудования, базируясь на типовых нагрузках кластера ADH T4.

| **Сервер** | **Типовая нагрузка** | **Хранилище данных** | **Процессор** | **Оперативная память** | **Сетевое оборудование** |
|---|---|---|---|---|---|
| DataNode | Сбалансированная нагрузка | 12 обычных дисков по 2-3 ТБ | 8 ядер | 128-256 ГБ | 1 Гбит/с для подключения, 2x10 Гбит/с межсоединение |
| | Интенсивные вычисления | 12 обычных дисков 1-2 ТБ | 10 ядер | 128-256 ГБ | 10 Гбит/с для подключения, 2x10 Гбит/с межсоединение |
| | Преимущественно хранение данных | 12 обычных дисков 4+ ТБ | 8 ядер | 128-256 ГБ | 10 Гбит/с для подключения, 2x10 Гбит/с межсоединение |
| NameNode | NameNode | NameNode | NameNode | NameNode | NameNode |
| Сбалансированная нагрузка | Сбалансированная нагрузка | Сбалансированная нагрузка | Сбалансированная нагрузка | Сбалансированная нагрузка | Сбалансированная нагрузка |

# 3 Требования к файловой системе

## 3.1 Поддерживаемые файловые системы

Файловая система Hadoop Distributed File System (HDFS) предназначена для работы поверх базовой файловой системы операционной системы. Поддерживаются следующие операционные системы:

- ext3 — наиболее протестированная базовая файловая система для HDFS;
- ext4 — является масштабируемым расширением ext3;
- XFS — файловая система по умолчанию в RHEL 7.

Если вы выбираете между ext3 и ext4, рекомендуется ext4.

## 3.2 Использование опции noatime для повышения производительности

Файловые системы Linux хранят метаданные, которые записывают время доступа к каждому файлу. Это означает, что каждая операция чтения также осуществляет запись на диск. Отключите эту функцию, чтобы ускорить чтение файлов. Для этого добавьте опцию монтирования `noatime` в каждую строку, которая определяет файловую систему в файле */etc/fstab*, например:

```
/dev/sdb1 /data1 xfs defaults,noatime 0
```

Используйте следующую команду, чтобы применить изменения без перезагрузки:

```
$ mount -o remount /data1
```

## 3.3 Ограничения использования параметров монтирования файловой системы

- Опция монтирования `sync` позволяет осуществлять запись синхронно. Использование синхронизации снижает производительность сервисов, записывающих данные на диски (например, HDFS и YARN). В ADH большинство операций записи реплицируются, поэтому синхронная запись на диск не является необходимой, она ресурсозатратна и не дает заметного улучшения стабильности работы кластера. Использовать `sync` не рекомендуется.
- Опции `nfs` и `nas` не поддерживаются при монтировании каталога данных DataNode.

- Монтирование */tmp* как файловой системы с опцией `noexec` не поддерживается. Этот способ используется для предотвращения выполнения хранящихся файлов.

### 3.4 Опция umask

UNIX-системы используют umask (user file-creation mode mask), чтобы устанавливать разрешения по умолчанию для создаваемых файлов и каталогов. В большинстве дистрибутивов Linux значение umask по умолчанию — `0022 (022)` или `0002 (002)`. Базовые права для каталога — `0777 (rwxrwxrwx)`, а для файла — `0666 (rw-rw-rw)`. Чтобы определить права доступа после применения umask, вычтите значение umask из базовых прав. Значение umask `0002` используется для обычного пользователя. При использовании этого значения права по умолчанию для каталога — `775`, а для файла — `664`. Для суперпользователя (root) маска по умолчанию — `0022`. При использовании этой маски права по умолчанию для каталога — `755`, а для файла — `644`.

Вы можете установить umask в файле *etc/bashrc* или */etc/profile*.

| **Значение umask** | **Файл** | | | | **Каталог** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Результат** | **Владелец** | **Группа** | **Остальные** | **Результат** | **Владелец** | **Группа** | **Остальные** |
| 0022 (рекомендовано) | 644 | rw- | r-- | r-- | 755 | rwx | r-x | r-x |
| 0002 | 664 | rw- | rw- | r-- | 775 | rwx | rwx | r-x |
| 0000 | 666 | rw- | rw- | rw- | 777 | rwx | rwx | rwx |

## 4 Сетевые порты ADH T4

Сопоставление сетевых портов компонентов и серверов сервисов ADH T4 приведены в таблице на странице https://docs.arenadata.io/ru/ADH/current/planning/adh-port-mapping.html.

## 5 Типовые конфигурации. Пилотный кластер

Ниже приведен пример планирования пилотного кластера Hadoop на основе сбалансированного типа нагрузки (balanced workload pattern). Даже небольшой и простой кластер требует не менее трех узлов DataNode

и один NameNode. Можно использовать один физический сервер (или виртуальную машину) или различные машины.

## 5.1 Параметры

Для пилотного проекта выбираем следующие начальные условия:

- Объем данных — около 500 ТБ.
- Коэффициент репликации равен трём.
- Период сохранения данных равен одному году.
- Характер нагрузки — сбалансированный.
- Формат данных: 20% — простой текст, AVRO, Parquet, Jason, ORC, и другие; 80% — сжатые данные в формате GZIP и Snappy.

## 5.2 Оборудование

Спецификация на узлы DataNode зависит от объема хранимых и анализируемых данных.

## 5.3 Объем данных

В соответствии с коэффициентом репликации, равным трём, нам необходимо хранить данные объемом 500 ТБ * 3 = 1500 ТБ в течение одного года. Предположим, что 20% данных находятся в контейнерном формате, а 80% — архивированные данные в формате Parquet, сжатые с помощью Snappy. Эффективность сжатия составляет 70-80%, примем 80%. Расчет объема хранилища выглядит следующим образом:

```
требуемый объем хранилища = объем данных * % контейнерные данные + объем данных * % в сжатом виде * ожидаемая компрессия
```

При заданных условиях требуемый объем равен

```
1500 * 0.2 + 1500 * 0.8 * (1 - 0.8) = 540 ТБ
```

Кроме учтенных хранимых данных, необходимо также пространство для обработки данных и выполнения других задач. Поэтому следует предусмотреть дополнительное пространство. Предположим, что в среднем каждый день кластер обрабатывает 10% всех хранимых данных и процесс обработки создает в три раза больше временных данных. Таким образом, нам необходимо около 30% дополнительного объема.

Требуемый объем хранилища для данных и различных видов обработки будет 540 + 540 * 0.3 = 702 ТБ.

Для узлов DataNode рекомендуются обычные дисковые накопители. Файловая система требует для служебных целей около 20% от пространства хранения данных. Поэтому необходимо увеличить требования к пространству еще на 20%. Теперь мы имеем окончательное значение 702 * (1 + 0.2) = 842.4 ТБ, которое округлим до 845 ТБ.

## 5.4 Количество узлов DataNode

Вычислим количество узлов DataNode, необходимых для образования хранилища объемом 845 ТБ. Предположим, что в каждый узел мы устанавливаем 12 обычных дисковых устройств объемом по 4 ТБ. Следовательно, объем хранилища в одном узле будет 48 ТБ.

Необходимое количество узлов DataNode будет 845 / 48 ~ 18.

Нет необходимости устанавливать весь кластер в первый же день. Можно наращивать его размер, начиная с установки 20% от требуемого количества узлов и постепенно увеличивая до 100%.

## 5.5 Процессор и память

Согласно рекомендациям, следующие характеристики удовлетворяют требованиям пилотного кластера:

- 8 ядер процессора;
- 128 ГБ оперативной памяти.