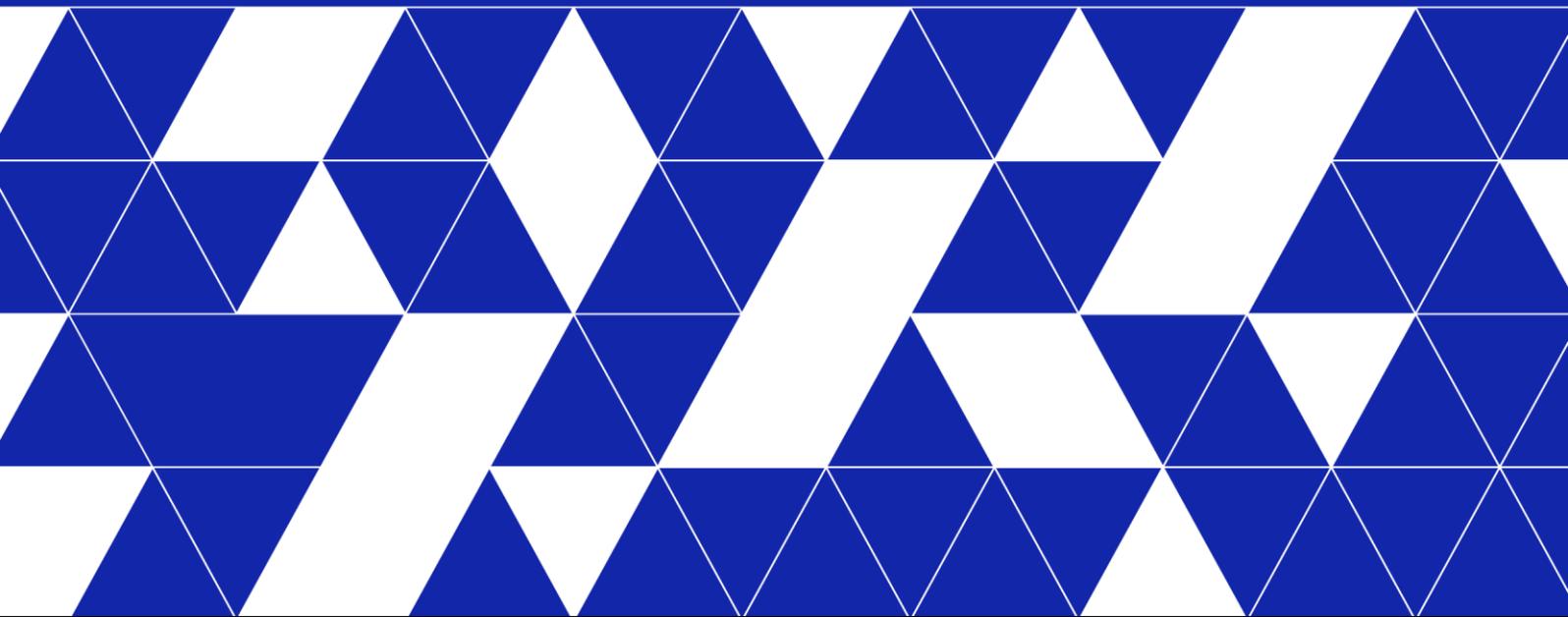




**ТЕХНИКО-КОМЕРЧЕСКОЕ ПРЕДЛОЖЕНИЕ
НА СОЗДАНИЕ ГОСУДАРСТВЕННОЙ
ИНФОРМАЦИОННОЙ СИСТЕМЫ
«НАЦИОНАЛЬНАЯ БАЗА ГЕНЕТИЧЕСКОЙ
ИНФОРМАЦИИ»**

Версия: релиз 1.27 ревизия 10.06.2024



Содержание

1. НАЗНАЧЕНИЕ ДОКУМЕНТА.....	3
2. ВВЕДЕНИЕ.....	4
3. РЕЗЮМЕ ДЛЯ РУКОВОДСТВА	5
4. ТРЕБОВАНИЯ К АРХИТЕКТУРЕ	7
5. ВЫСОКОУРОВНЕВАЯ АРХИТЕКТУРА.....	9
6. АРХИТЕКТУРА КОМПЛЕКСА ТЕХНИЧЕСКИХ СРЕДСТВ	11
7. ЭТАПНОСТЬ ПРОЕКТА	16
7.1. ЭТАП 1	16
7.2. ЭТАП 2.....	16
8. ПРЕИМУЩЕСТВА РЕШЕНИЯ ОБЪЕКТНОГО ХРАНЕНИЯ	17
8.1. ОБЩИЕ ПРЕИМУЩЕСТВА	17
8.2. ПРЕИМУЩЕСТВА TATLIN.OBJECT В СРАВНЕНИИ С КЛАСТЕРОМ CEPH	20
8.2.1 ПРЕИМУЩЕСТВА ИСПОЛЬЗОВАНИЯ TATLIN.OBJECT	20
8.2.2 РИСКИ ИСПОЛЬЗОВАНИЯ CEPH	24
8.2.2.1 РЕПУТАЦИОННЫЕ РИСКИ	24
8.2.2.2 САНКЦИОННЫЕ И КОРПОРАТИВНЫЕ РИСКИ.....	25
8.2.2.3 РИСКИ СВЯЗАННЫЕ С ОТКРЫТЫМ ПО	27
8.2.2.4 РИСКИ СВЯЗАННЫЕ С АРХИТЕКТУРОЙ CEPH	28
9. ПРЕИМУЩЕСТВА СИСТЕМ ХРАНЕНИЯ ДАННЫХ YADRO.....	31
9.1. TATLIN.UNIFIED GEN 2.....	31
9.2. TATLIN.BACKUP.M	33
10. ВЫВОДЫ	36
11. ПРИЛОЖЕНИЯ	37
11.1. ОСНОВНЫЕ НОРМАТИВНЫЕ ПРАВОВЫЕ И ТЕХНИЧЕСКИЕ ДОКУМЕНТЫ.....	37
11.2. СОСТАВ ОБОРУДОВАНИЯ YADRO	42
11.2.1. СОСТАВ ОБОРУДОВАНИЯ. ПЕРВЫЙ ЭТАП.....	42
11.2.2. СОСТАВ ОБОРУДОВАНИЯ. ВТОРОЙ ЭТАП.	43
11.2.3. СОСТАВ ОБОРУДОВАНИЯ. КОНТУР БЕЗОПАСНОСТИ.....	44

1. НАЗНАЧЕНИЕ ДОКУМЕНТА

Данный документ является описанием решения компании YADRO для создания ИТ-инфраструктуры государственной информационной системы в области генетической информации «Национальная база генетической информации» (ГИС НБГИ).

При разработке технического решения специалисты компании YADRO руководствовались нормативными правовыми и техническими актами, приведёнными в разделе 9 настоящего документа «Основные нормативные правовые и технические документы».

Авторское право

Компания YADRO полагает, что информация в данном документе является достоверной на момент публикации. Информация может быть изменена без предварительного уведомления.

Вся информация, изложенная в данном документе, является конфиденциальной и не должна публиковаться или раскрываться полностью или частично никакой стороне без письменного согласия компании YADRO.

Все прочие товарные знаки являются собственностью соответствующих владельцев.

©2024, ООО «КНС ГРУПП» (YADRO). Все права защищены.

2. ВВЕДЕНИЕ

ГИС НБГИ должна стать единым интегратором генетической информации, представлять собой совокупность взаимосвязанных цифровых систем, подсистем и сервисов, предназначенных для работы с единым набором генетической информации научным, образовательным, медицинским, сельскохозяйственным сообществом и другими организациями, ведущими работы в области генетики во всех регионах Российской Федерации.

Целью проекта является построение информационной системы, обеспечивающей хранение данных, оперативность, доступность и качество полученной цифровой генетической информации по всему многообразию биообразцов, включая растения, животных, микроорганизмы дикой природы и метагеномы экосистем, сельскохозяйственные растения и сельскохозяйственных животных, промышленные микроорганизмы, вирусы, исключая персонифицированные генетические данные человека, а также особо опасные патогенные микроорганизмы и вирусы.

Данный документ представляет собой описание решения от компании **YADRO** по построению инфраструктуры для единой централизованной системы хранения цифровой информации Заказчика и является составной частью общего проекта.

Основным назначением системы хранения данных является надёжное централизованное хранение, резервирование и синхронизация информационных ресурсов ГИС НБГИ. Целью создания подсистемы хранения является повышение доступности информационных ресурсов, защиты от аппаратных и программных сбоев, выполнение требований по гарантированному хранению данных.

3. РЕЗЮМЕ ДЛЯ РУКОВОДСТВА

В настоящем документе дано описание решения компании YADRO для создания ИТ-инфраструктуры ГИС НБГИ. Предлагаемое решение обеспечивает надежную эффективную обработку и хранение информационных ресурсов ГИС НБГИ, выполнение требований к производительности и доступности информационных ресурсов, защиты от аппаратных и программных сбоев, гарантированному хранению данных и информационной безопасности.

Решение предполагает использование исключительно отечественных аппаратных и программных систем, внесенных в реестр Минпромторга. В состав предложения включены пуско-наладочные работы, инсталляция и настройка, а также техническая поддержка от компаний-производителей.

Решение построено на основе хорошо зарекомендовавших себя серверов, сетевого оборудования и систем хранения, используемых крупнейшими российскими государственными и частными компаниями:

- ✓ TATLIN.UNIFIED – высокопроизводительная надежная СХД общего назначения (хранение оперативных данных ГИС НБГИ).
- ✓ TATLIN.OBJECT – специализированная система хранения с объектным доступом, предназначенная для хранения больших объемов информации с минимальными усилиями по обслуживанию и администрированию (хранение генетической информации и архивное хранение ГИС НБГИ).
- ✓ TATLIN.BACKUP – специализированная система хранения, обеспечивающая надежное эффективное хранение и быстрое восстановление из резервных копий (защита оперативных данных ГИС НБГИ).

Предлагаемая архитектура подсистем хранения на базе продуктов YADRO обладает рядом неоспоримых технологических преимуществ, которые обеспечивают максимально надёжное и эффективное применение в составе ГИС НБГИ:

- ✓ Линейное масштабирование и минимизация затрат на обслуживание.
- ✓ Обеспечение высокой доступности и целостности данных при долгосрочном хранении.
- ✓ Использование полностью российских законченных аппаратно-программных решений, дополненных пуско-наладочными работами и профессиональной технической поддержкой.
- ✓ Минимизация санкционных рисков.
- ✓ Минимизация рисков, связанных со сложной архитектурой и проблемами сопровождения решения, построенного на открытом ПО.

Реализацию проекта предлагается выполнить в два этапа. На первом этапе будет выполнена поставка основного комплекса технических средств для всех трёх контуров, с учётом полного объёма вычислительной и сетевой инфраструктуры и ограниченного бюджетом объёма основной подсистемы хранения. На втором этапе будет выполнено расширение основной подсистемы хранения генетической информации.

Данный подход позволит построить полнофункциональную инфраструктуру ГИС НБГИ уже на первом этапе и приступить к разработке системы в полном объёме, масштабируя ресурсы хранения по мере необходимости.

4. ТРЕБОВАНИЯ К АРХИТЕКТУРЕ

В ходе обсуждений с представителями Заказчика были выявлены следующие основные требования к архитектуре системы хранения данных:

- ✓ обеспечить централизованное хранение данных для снижения эксплуатационных расходов на поддержание подсистемы хранения данных;
- ✓ обеспечить доступ к сохранённой генетической информации по открытым протоколам внутри ИТ-инфраструктуры;
- ✓ повысить эффективность хранения цифровой информации;
- ✓ обеспечить достаточный уровень надёжности хранения данных и защиту данных от логических ошибок;
- ✓ применяемые технологии должны поддерживать построение катастрофоустойчивого решения с использованием асинхронной репликации информации между центрами обработки данных;
- ✓ обеспечить защиту данных от логических ошибок;
- ✓ упростить ежедневное администрирование и обслуживание системы, сведя к минимуму задачи по ручному управлению системой;
- ✓ обеспечить возможность простого горизонтального расширения системы в соответствии с ростом объема производственной информации и количества используемых систем;
- ✓ обеспечить возможность увеличения ёмкости системы путём добавления единых блоков, содержащих дополнительные вычислительные ресурсы, дисковую ёмкость и сетевые интерфейсы, без остановки доступа к данным и без потери производительности системы;
- ✓ обеспечить возможность добавления и удаления узлов/контроллеров системы без остановки доступа к хранимой информации с автоматической ребалансировкой данных.

Необходимо обеспечить хранение генетической информации ГИС НБГИ, располагающейся в трех контурах со следующим объемом продуктивных данных:

- открытый контур ГИС НБГИ 32 ПБ;
- конфиденциальный контур ГИС НБГИ 20 ПБ;
- специальный контур ГИС НБГИ 1,5 ПБ.

Для каждого контура предусмотреть архивный слой для долгосрочного хранения генетической информации ГИС НБГИ следующего объёма:

- открытый контур ГИС НБГИ 10 ПБ;
- конфиденциальный контур ГИС НБГИ 6 ПБ;
- специальный контур ГИС НБГИ 1,5ПБ.

Для обеспечения требований Заказчика компания YADRO предлагает использовать объектное хранение данных, реализованный в решении TATLIN.OBJECT (реестровая запись 7473 \ 1 \ 2023).

Доступ к данным осуществляется с использованием протоколов:

- gRPC (основной, высокопроизводительный доступ к данным);
- S3 (альтернативный доступ к данным, режим совместимости).

В каждом контуре необходимо обеспечить хранение оперативных данных объёмом 150 ТБ на высокоскоростном хранилище данных, а также, резервное копирование оперативных данных на хранилище данных СРК (для резервного копирования подсистем виртуализации, СУБД, управления, инфраструктурных сервисов и т.п.).

Для обеспечения требований Заказчика компания YADRO предлагает рассмотреть блочное хранение данных и построить решение с использованием протоколов iSCSI на базе TATLIN.UNIFIED GEN2 (реестровая запись 3361 \ 1 \ 2023), а для резервного копирования оперативных данных решение со сжатием и дедубликацией данных на всем объёме – TATLIN.BACKUP (реестровая запись 10535769).

5. ВЫСОКОУРОВНЕВАЯ АРХИТЕКТУРА

Предлагаемая архитектура ГИС НБГИ задействует выделенное аппаратное обеспечение для каждого из трёх уровней хранения данных (Tier) для каждого контура – Открытого, Конфиденциального и Специального. Такое решение обеспечивает максимальную надёжность хранения критичных данных (Рисунок 1):

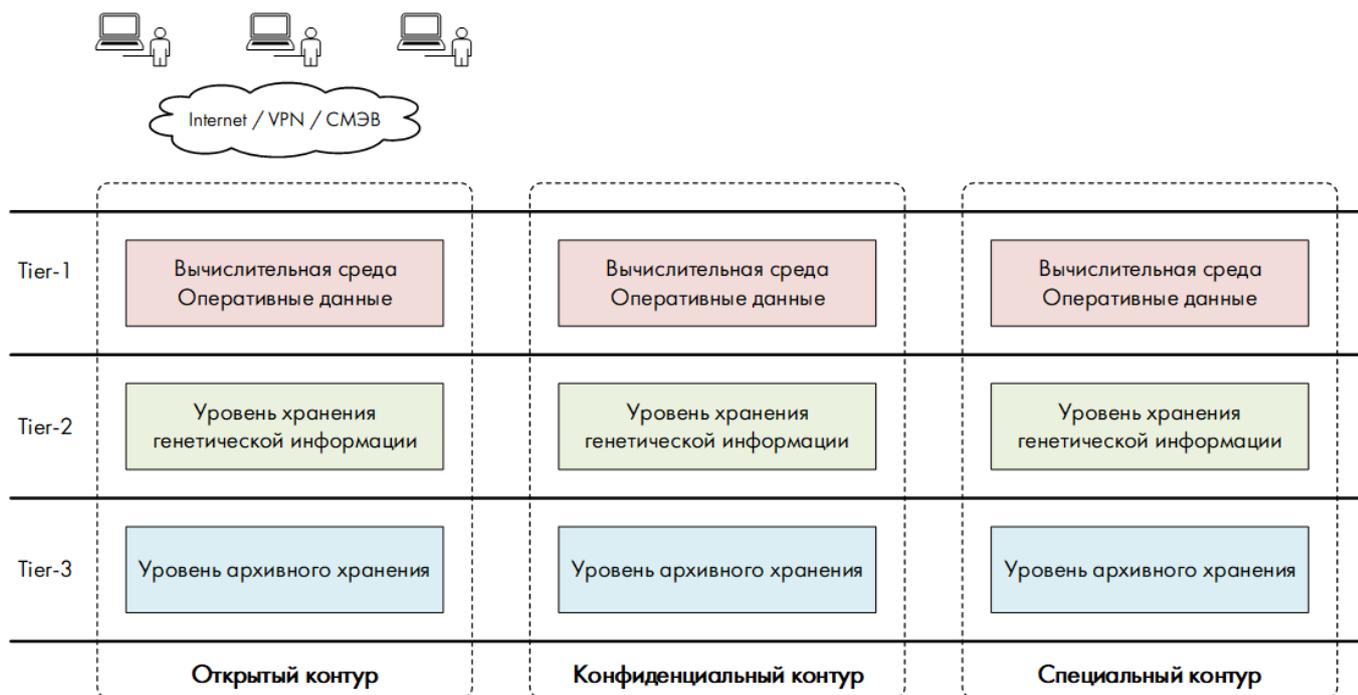


Рисунок 1 – Архитектура хранения данных

— Уровень оперативных данных (Tier-1) создаётся для хранения и обработки следующей информации:

- ✓ данные системы управления виртуализацией, загрузочные диски виртуальных машин, persistent-тома контейнеров;
- ✓ данные WEB-портала, приложений ГИС НБГИ, обработчиков данных, оперативные данные подсистемы высокопроизводительных вычислений;
- ✓ данные инфраструктурных сервисов (сетевые сервисы, балансировщики нагрузки, система управления СРК, система мониторинга и прочие вспомогательные подсистемы, и сервисы);

- Уровень хранения генетической информации (Tier-2) создаётся для хранения и обработки следующей информации:
 - ✓ все основные генетические данные, загружаемые пользователями в систему и обрабатываемые приложениями ГИС НБГИ.
- Уровень архивного хранения (Tier-3) создаётся для хранения и обработки следующей информации:
 - ✓ архивные генетические данные, обработанные приложениями, ГИС НБГИ, классифицированные и подлежащие долговременному хранению (в соответствии с выбранной политикой хранения).
 - ✓ резервные копии данных оперативного уровня, создаваемые СРК.

6. АРХИТЕКТУРА КОМПЛЕКСА ТЕХНИЧЕСКИХ СРЕДСТВ

Каждый контур (Открытый, Конфиденциальный, Специальный) изолирован на сетевом уровне и размещается на независимом комплексе технических средств (Рисунок 2).

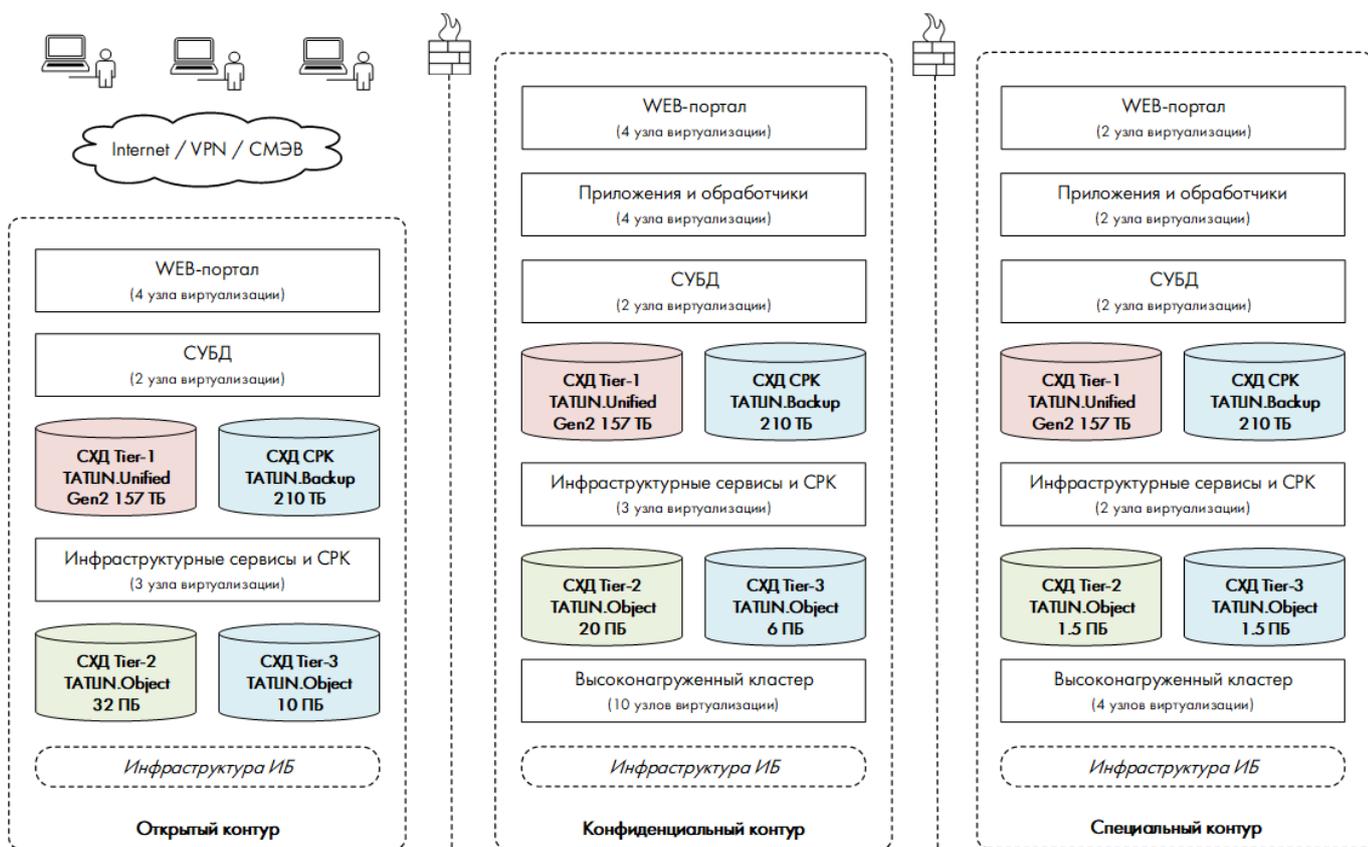


Рисунок 2 - Верхнеуровневая схема КТС

В состав каждого контура входит следующее оборудование:

— Серверы подсистемы виртуализации:

- ✓ серверы VEGMAN R120 G2 (WEB-портал/инфраструктурные сервисы);
- ✓ серверы VEGMAN R220 G2 (СУБД, серверы приложений/обработчики);
- ✓ серверы VEGMAN R220 G2 (серверы высоконагруженных кластеров).

— Системы хранения данных:

- ✓ СХД TATLIN.UNIFIED Gen2 (данные Tier-1);
- ✓ СХД TATLIN.OBJECT (данные Tier-2);
- ✓ СХД TATLIN.OBJECT (данные Tier-3);
- ✓ СХД TATLIN.BACKUP (данные Tier-3).

Сводная таблица конфигураций и количества серверов:

Описание	Кол-во	CPU	RAM	Storage	LAN
Открытый контур					
WEB-сервер - VEGMAN R120 G2	4	2 x Xeon 6326 (всего 32 ядра, 2,9 ГГц)	256	2 x 480ГБ SSD	4 x 25G
СУБД - VEGMAN R220 G2	2	2 x Xeon 6348 (всего 56 ядер, 2,6 ГГц)	1024	2 x 480ГБ SSD	4 x 25G
Infra - VEGMAN R120 G2	3	2 x Xeon 6326 (всего 32 ядра, 2,9 ГГц)	512	2 x 480ГБ SSD	4 x 25G
Конфиденциальный контур					
WEB-сервер - VEGMAN R120 G2	4	2 x Xeon 6326 (всего 32 ядра, 2,9 ГГц)	512	2 x 480ГБ SSD	4 x 25G
Приложения - VEGMAN R220 G2	6	2 x Xeon 6348 (всего 56 ядер, 2,6 ГГц)	512	2 x 480ГБ SSD	4 x 25G
СУБД - VEGMAN R220 G2	2	2 x Xeon 6348 (всего 56 ядер, 2,6 ГГц)	1024	2 x 480ГБ SSD	4 x 25G
HighPerf - VEGMAN R220 G2	10	2 x Xeon 6348 (всего 56 ядер, 2,6 ГГц)	512	2 x 480ГБ SSD	4 x 25G 2 x 100G
Infra - VEGMAN R120 G2	3	2 x Xeon 6326 (всего 32 ядра, 2,9 ГГц)	512	2 x 480ГБ SSD	4 x 25G
Специальный контур					
WEB-сервер - VEGMAN R120 G2	2	2 x Xeon 6326 (всего 32 ядра, 2,9 ГГц)	512	2 x 480ГБ SSD	4 x 25G
Приложения - VEGMAN R220 G2	3	2 x Xeon 6348 (всего 56 ядер, 2,6 ГГц)	512	2 x 480ГБ SSD	4 x 25G
СУБД - VEGMAN R220 G2	2	2 x Xeon 6348 (всего 56 ядер, 2,6 ГГц)	1024	2 x 480ГБ SSD	4 x 25G
HighPerf - VEGMAN R220 G2	6	2 x Xeon 6348 (всего 56 ядер, 2,6 ГГц)	512	2 x 480ГБ SSD	4 x 25G 2 x 100G
Infra - VEGMAN R120 G2	2	2 x Xeon 6326 (всего 32 ядра, 2,9 ГГц)	512	2 x 480ГБ SSD	4 x 25G

Сводная таблица конфигураций и количества СХД:

Описание	Кол-во	Узлы	Полезная ёмкость
Открытый контур			
TATLIN.UNIFIED Gen2 (Tier-1)	1	-	157 ТБ
TATLIN.OBJECT (Tier-2)	1	26	32 000 ТБ
TATLIN.OBJECT (Tier-3)	1	8	10 000 ТБ
TATLIN.BACKUP (Tier-3)	1	-	214 ТБ

Описание	Кол-во	Узлы	Полезная ёмкость
Конфиденциальный контур			
TATLIN.UNIFIED Gen2 (Tier-1)	1	-	157 ТБ
TATLIN.OBJECT (Tier-2)	1	16	20 000 ТБ
TATLIN.OBJECT (Tier-3)	1	8	6 000 ТБ
TATLIN.BACKUP (Tier-3)	1	-	214 ТБ
Специальный контур			
TATLIN.UNIFIED Gen2 (Tier-1)	1	-	157 ТБ
TATLIN.OBJECT (Tier-2)	1	8	1500 ТБ
TATLIN.OBJECT (Tier-3)	1	8	1500 ТБ
TATLIN.BACKUP (Tier-3)	1	-	214 ТБ

Все предлагаемые системы хранения данных позволяют обеспечить синхронную (Tier-1) или асинхронную (Tier-2, Tier-3) репликацию данных в удалённый ЦОД для реализации выбранной Заказчиком DR-стратегии.

Для обеспечения георезервирования с целью полной сохранности данных рекомендуется предусмотреть, как минимум, репликацию резервных копий данных оперативного уровня и архивных данных генетической информации в локацию, удалённую географически от основного центра обработки данных. Предлагаемые СХД готовы к расширению решения и реализации такой репликации.

Одной из важнейших подсистем комплекса является построение вычислительной сети центра обработки данных, обеспечивающая надёжную и высокопроизводительную связь между вычислительной инфраструктурой и системами хранения данных. Предлагаемая схема сети представлена на Рисунке 3.

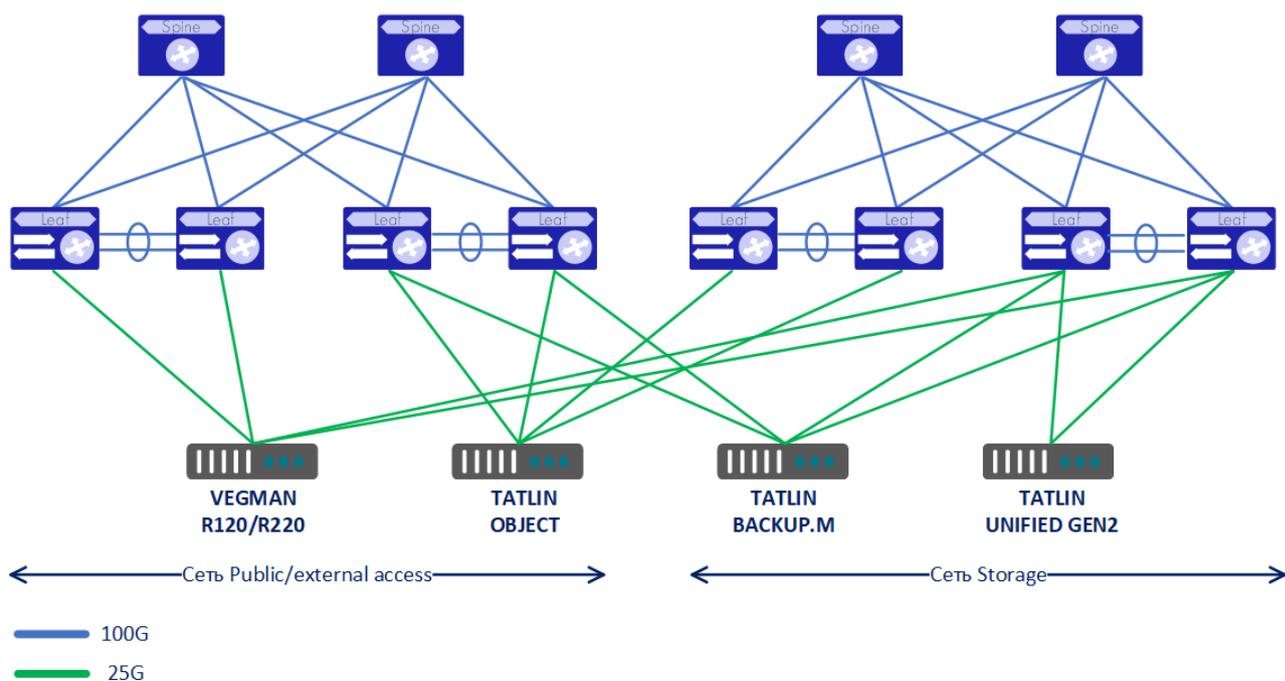


Рисунок 3 – Концептуальная схема сети передачи данных в ЦОДе

Вычислительную сеть центра обработки данных предлагается реализовать в соответствии с CLOS-архитектурой Leaf-Spine, с высоким уровнем надёжности и оптимизированной пропускной способностью. Архитектура обеспечивает минимальную задержку и неблокируемую связность между вычислительными узлами и системами хранения данных.

Узлы вычисления и хранения подключаются к уровню коммутаторов leaf по линкам 25G. Spine-коммутаторы обеспечивают транспортную связность для уровня leaf. Коммутаторы leaf подключаются к уровню spine по отказоустойчивым линкам 100G.

Предусмотрены выделенные сетевые фабрики Leaf-Spine для внешнего доступа и для IP сети хранения данных. Для блочного хранения уровня Tier-1 резервирование подключений осуществляется с помощью iSCSI multipath. Обеспечиваются резервированные подключения MLAG для серверов вычислений и узлов объектного хранения.

Сводная таблица конфигураций и количества коммутаторов:

Описание	Кол-во	Порты
Открытый контур		
Коммутатор KORNFELD D1156	26	48 x 25G + 8 x 100G
Конфиденциальный контур		
Коммутатор KORNFELD D1156	30	48 x 25G + 8 x 100G
Специальный контур		
Коммутатор KORNFELD D1156	20	48 x 25G + 8 x 100G

7. ЭТАПНОСТЬ ПРОЕКТА

Реализацию проекта предлагается выполнять в два этапа:

7.1. ЭТАП 1

Закупка и ПНР основного комплекса технических средств для всех трёх контуров, с учётом полного объёма вычислительной и сетевой инфраструктуры и ограниченного бюджетом объёма основных объектных СХД для хранения генетической информации.

7.2. ЭТАП 2

Расширение основных объектных СХД для хранения генетической информации комплектом модернизации, обеспечивающим оптимизированную стоимость хранения.

Данный подход позволит построить полнофункциональный прототип всех подсистем вычислительного комплекса на первом этапе и приступить к разработке программного обеспечения без каких-либо ограничений в плане производительности, архитектуры применяемых программных средств и методик обеспечения отказоустойчивости.

По мере реализации программных модулей и накопления реальных данных будет обеспечена возможность точной оценки размера данных и эффект от их компрессии (сжатия). Данный подход позволит рассчитать параметры целевой конфигурации основного объектного хранилища данных (как самого дорогостоящего компонента комплекса) и обеспечить максимальную эффективность.

В предложение включены все пуско-наладочные работы и настройка инфраструктуры. Спецификации на оборудование и программное обеспечение включают в себя техническую поддержку уровня «оптимальная» на один год.

8. ПРЕИМУЩЕСТВА РЕШЕНИЯ ОБЪЕКТНОГО ХРАНЕНИЯ

8.1. ОБЩИЕ ПРЕИМУЩЕСТВА

Предлагаемое решение **TATLIN.OBJECT** для хранения и обработки данных генетической информации от компании **YADRO** учитывает современные тенденции, когда в области геномики и биоинформатики технологические достижения в области секвенирования следующего поколения (NGS) позволяют участникам процесса генерировать большее количество геномных данных и делать это намного быстрее.

Решение от компании **YADRO** учитывает, что за последнее десятилетие объем геномных данных в мире удваивался каждые 7 месяцев, а быстрое развитие геномики выявляет проблему организации высокоскоростного, распределённого, долгосрочного доступа к данным. В последнее время традиционные технологии организации систем хранения неструктурированных данных с файловым доступом начали быстро вытесняться объектными хранилищами, которые обладают рядом важных преимуществ:

— **Линейное масштабирование.**

Быстрый рост объёмов хранения геномных данных диктует необходимость организации инфраструктуры хранения, масштабирующейся до десятков и сотен петабайт. При этом важна не только ёмкость, но и обеспечение производительности при параллельной работе сотен и тысяч пользователей, возможность организации обмена данными между большими и распределёнными системами. Современные объектные системы хранения обеспечивают возможность построения гибкой технологической инфраструктуры практически неограниченного масштаба в широком спектре сетевых и географических топологий, что делает их идеально подходящими для современных больших геномных хранилищ.

— **Долгосрочное хранение и целостность данных.**

Как правило, политика хранения геномных данных формулируется как «десятки лет» или «хранить вечно». Для обеспечения соответствия таким требованиям необходима инфраструктура, защищающая целостность данных на протяжении всего жизненного цикла без дорогостоящих ручных проверок. Объектные хранилища отлично справляются с обнаружением и исправлением битовых

ошибок, особенно по сравнению с традиционными хранилищами, благодаря проверке целостности данных, выполняемой в фоновом режиме.

Долгосрочное хранение обеспечивается за счёт интегрированных возможностей обновления узлов хранения и ПО «на лету». При этом необходимо предусмотреть эволюцию прикладного программного обеспечения, развитие сетевых технологий и кратное усложнение инфраструктуры. Объектные системы хранения обеспечивают простое взаимодействие с приложениями по современным стандартизированным протоколам доступа gRPC и S3, поддерживая долгосрочные проекты геномных исследований.

— **Высокая доступность и распределённый доступ.**

В современных объектных хранилищах используются технологии избыточности способные восстановить целые пакеты данных в случае их потери (как пример, коды Рида-Соломона). Данные технологии позволяют распределять данные между различными дисками, узлами хранения и стойками, а также между центрами обработки данных и географическими регионами. Если конкретный диск или узел хранения выходит из строя, данные можно считать с оставшихся дисков и узлов.

Такая устойчивость к отказам устройств работает и в том случае, если весь центр обработки данных разрушен в результате стихийного бедствия. Технология рассредоточения данных в объектных хранилищах даёт два больших преимущества: готовность к незапланированным простоям и упрощение обмена данными между географически распределёнными локациями без дополнительных накладных расходов.

Использование объектных хранилищ для хранения и обработки данных генетической информации на примере мирового опыта:

- проект австралийского Центр Исследований генома (AGRF) с 2021 года использует S3-совместимое объектное хранилище, позволяющее AGRF обрабатывать и хранить быстро растущие объёмы данных секвенирования генома;
- проект, реализованный в 2020 году организацией Genomics England. Этот институт был создан в 2013 году Министерством здравоохранения и

социального обеспечения Великобритании для поддержки проекта «100 000 геномов» - работы по секвенированию целых геномов пациентов с редкими заболеваниями и распространёнными видами рака. В 2018 году проект был значительно расширен, что потребовало от Genomics England масштабирования инфраструктуры хранения данных и замены, существующего NAS-хранилища, которое не справлялось с растущими объёмами данных, на специализированное S3-совместимое решение;

— проекты, реализованные в университете Стэнфорда (Genetics Bioinformatics Service Center), компании PetaGene и многих других организациях, занимающихся геномными исследованиями.

8.2. ПРЕИМУЩЕСТВА TATLIN.OBJECT В СРАВНЕНИИ С КЛАСТЕРОМ CEPH

Ниже приведены преимущества и недостатки реализации системы хранения генетической информации с использованием технологий объектного доступа с использованием TATLIN.OBJECT и недостатки реализации системы хранения генетической информации с использованием файлового доступа CEPH.

8.2.1 ПРЕИМУЩЕСТВА ИСПОЛЬЗОВАНИЯ TATLIN.OBJECT

1. TATLIN.OBJECT – полностью российский продукт, разработанный и поддерживаемый российскими разработчиками. Права на исходный код принадлежат российской компании, продукт находится в реестре

— Описание на сайте:
[https://yadro.com/ru/tatlin/object/;](https://yadro.com/ru/tatlin/object/)

— Реестровая запись 5197\1\2022:
<https://gisp.gov.ru/goods/#/product/3444468;>

2. TATLIN.OBJECT представляет собой законченное решение, устанавливаемое и настраиваемое сотрудниками производителя. Заказчику доступны курсы администрирования и сопровождение от разработчика систем, что максимально упрощает сопровождение системы;

3. TATLIN.OBJECT предоставляет высокую надёжность решения:

- Отсутствие единой точки отказа;
- Самовосстановление после сбоев;
- Настраиваемая репликация;
- Механизм быстрой эвакуации данных с аварийного узла хранения;
- Эффективная работа в режиме деградации;

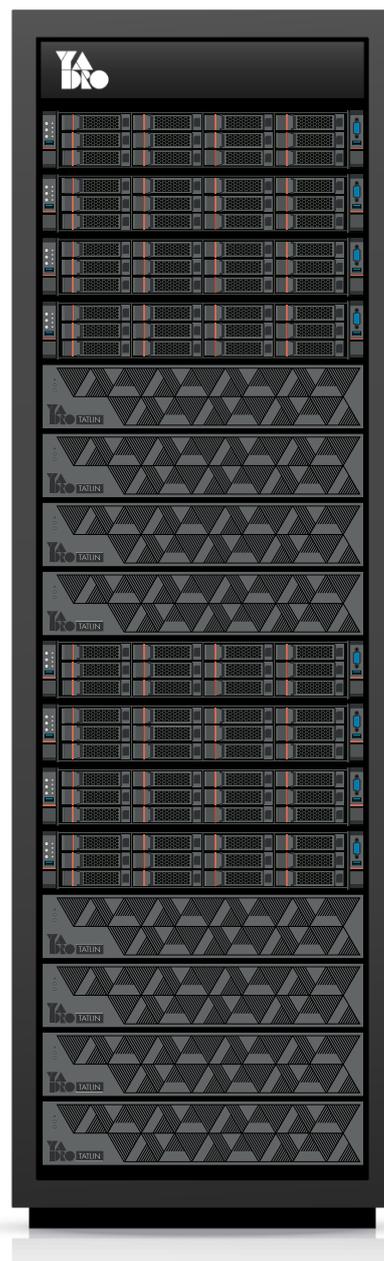


Рисунок 4 – размещение Tatlin.Object в стойке

— Формат данных, пригодный для ручного извлечения и восстановления;

4. **TATLIN.OBJECT** предоставляет гибкие настройки безопасности:

— Гибкие политики хранения на каждый контейнер позволяют описать правила размещения объектов для удовлетворения требований регламентов и законов;

— Поддержка IAM-совместимых методов для разграничения доступа к ресурсам хранилища;

— Управление TLS-сертификатами из Web UI.

5. **TATLIN.OBJECT**, кроме собственного протокола на базе gRPC, поддерживает S3, HTTP(S). В наличии возможность интеграции с другими приложениями через SDK (Go);

6. **TATLIN.OBJECT** позволяет использовать различные схемы размещения данных, в том числе технологии Erasure Coding, обеспечивающие максимально экономичное и надёжное хранение данных;

7. В **TATLIN.OBJECT** нет выделенного слоя метаданных, утеря или повреждение которого приведёт к потере работоспособности кластера. Метаданные хранятся вместе с объектами и в случае потери кэширующего слоя на узлах хранения все индексы могут быть заново получены пересканированием объектов;

8. В **TATLIN.OBJECT** Хранящиеся данные могут быть эвакуированы с неисправного узла и доступны для восстановления даже при отказе множества компонентов и даже узлов целиком;

9. Архитектура **TATLIN.OBJECT** позволяет осуществлять простое масштабирование как внутри площадки, так и между площадками (репликация на уровне контейнеров средствами СХД, либо логики прикладного ПО);

10. При использовании **TATLIN.OBJECT** приложение обращается к объектам по их идентификаторам. В результате, смена поколений аппаратных компонент **TATLIN.OBJECT** и перемещение данных внутри системы является абсолютно прозрачной для интегрированного приложения;

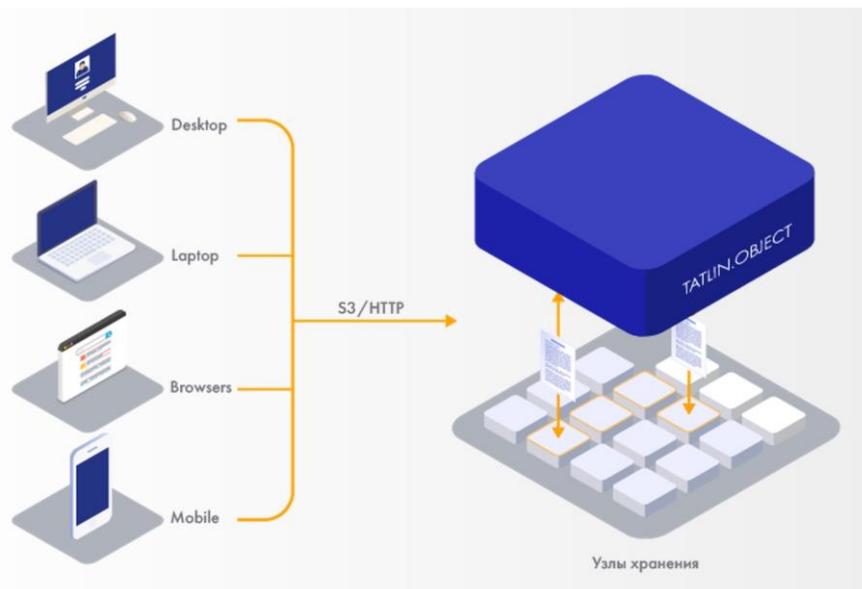


Рисунок 5 – Работа в режиме современного gRPC, S3 хранилища

11. TATLIN.OBJECT позволяет линейно расширять сеть от одного узла до сотен в разных ЦОД без простоев и деградации производительности;

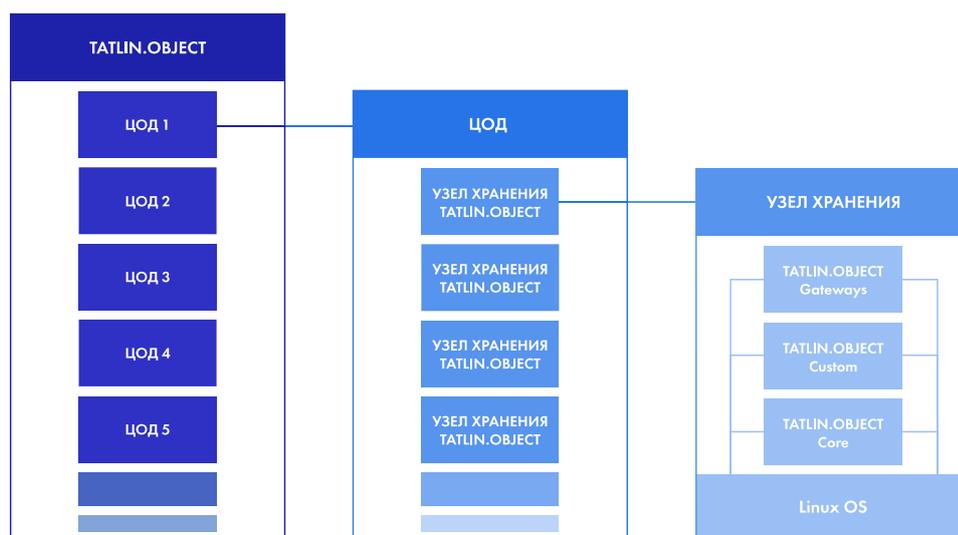


Рисунок 6 – Глобальная географически распределенная система хранения данных

12. TATLIN.OBJECT позволяет задать правила хранения на уровне каждого бакета/контейнера и гарантирует их соблюдение:

- объекты хранятся на каждом узле TATLIN.OBJECT на HDD дисках («шардах»);
- каждый объект хранится в контейнере, созданном пользователем;

- контейнеру задается ряд параметров, например: ACL, описывающий права доступа, и политика хранения, определяющая количество копий и местоположение объекта в системе;
- при сохранении объекту присваивается уникальный идентификатор (hash), по которому его можно найти и получить из системы;

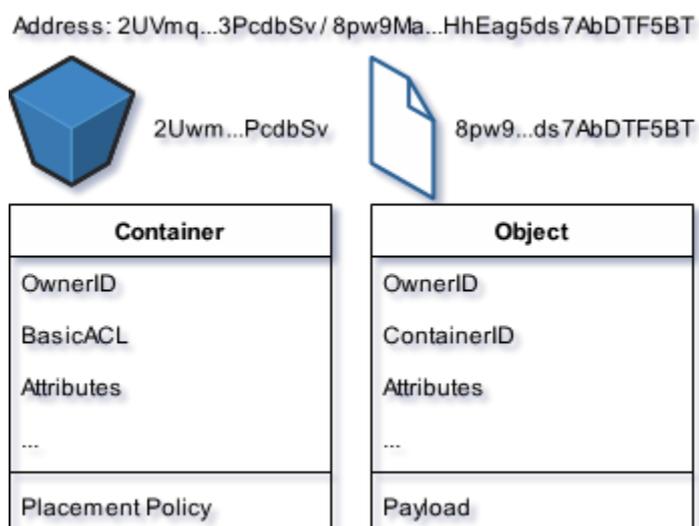


Рисунок 7 – Хранение объектов

13. Ведутся работы по сертификации ФСТЭК решения TATLIN.OBJECT.

8.2.2 РИСКИ ИСПОЛЬЗОВАНИЯ СЕРН

8.2.2.1 РЕПУТАЦИОННЫЕ РИСКИ

Согласно федеральному закону от 29 декабря 2022 г. № 643-ФЗ «О внесении изменений в Федеральный закон «О государственном регулировании в области генно-инженерной деятельности» ГИС НБГИ создаётся в целях обеспечения национальной безопасности, охраны и жизни здоровья граждан, а также обеспечения обмена содержащейся в ней информацией между федеральными государственными органами, государственными органами субъектов РФ, органами местного самоуправления и обладателями генетических данных при их взаимодействии в рамках осуществления генно-инженерной деятельности. Оператором данной системы будет выступать ФГБУ НИЦ «Курчатовский институт».

Права интеллектуальной собственности на систему в целом принадлежат Российской Федерации. Потеря данных является критическим событием государственного уровня и недопустима. Оператор системы несет прямую ответственность перед государством за обеспечение сохранности данных.

Зафиксировано несколько публичных случаев, когда информация о невозможности восстановления данных с СЕРН привела к потере данных и простоям:

- Широко известен случай с **многодневным простоем** ИС Росреестра, связанным с использованием ПО СЕРН <https://www.kommersant.ru/doc/3731094>;
- обращение пользователя по причине **падения хостинга VK Cloud** (<https://www.tinkoff.ru/invest/social/profile/flexxx777/eeec589d-10a1-4a7e-90de-bce9532ae4e0/>). VK Cloud использует инфраструктуру хранения данных на основе кластеров СЕРН;
- **потеря данных** на **Cloudmouse** (<https://habr.com/ru/articles/250097/>);
- **потеря данных** у **Yandex.Cloud** (<https://habr.com/ru/news/452238/>).

Подробно вопросы восстановления кластеров СЕРН рассмотрены в материалах <https://habr.com/ru/companies/oleg-bunin/articles/431536/> и <https://devopsconf.io/moscow/2024/abstracts/11659>. Обращает на себя внимание сложность процедур восстановления, необходимость тщательно планировать действия по

восстановлению работоспособности CERN и то, с какой легкостью можно сделать ошибку, приводящую к потере данных, даже высококвалифицированным специалистам.

ВЫВОД: CERN без должной поддержки корпоративных разработчиков: Red Hat, Intel, CERN, Cisco, Fujitsu, SanDisk, Canonical и SUSE, - может привести к потере данных без возможности их восстановления.

8.2.2.2 САНКЦИОННЫЕ И КОРПОРАТИВНЫЕ РИСКИ

CERN выпускается под лицензией LGPLv2.1 (<https://www.gnu.org/licenses/old-licenses/lgpl-2.1.en.html>), согласно данной лицензии:

- When we speak of free software, we are referring to freedom of use, not price (перевод: «Когда мы говорим о свободном программном обеспечении, мы имеем в виду свободу использования, а не цену»)
- For example, if you distribute copies of the library, whether gratis or for a fee, you must give the recipients all the rights that we gave you (перевод: «Например, если вы распространяете экземпляры библиотеки бесплатно или за плату, вы должны предоставить получателям все права, которые мы вам предоставили»)
- 12. If the distribution and/or use of the Library is restricted in certain countries either by patents or by copyrighted interfaces, the original copyright holder who places the Library under this License may add an explicit geographical distribution limitation excluding those countries, so that distribution is permitted only in or among countries not thus excluded. In such case, this License incorporates the limitation as if written in the body of this License. (перевод: «12. Если распространение и/или использование Библиотеки ограничено в определенных странах либо патентами, либо интерфейсами, защищёнными авторским правом, первоначальный владелец авторских прав, который размещает Библиотеку под настоящей Лицензией, может добавить явное ограничение географического распространения, исключая эти страны, чтобы распространение разрешено только в странах или среди стран, не

исключённых таким образом. В таком случае настоящая Лицензия включает ограничение, как если бы оно было написано в тексте настоящей Лицензии);

- 16. In no event unless required by applicable law or agreed to in writing will any copyright holder, or any other party who may modify and/or redistribute the library as permitted above, be liable to you for damages, including any general, special, incidental or consequential damages arising out of the use or inability to use the library (including but not limited to loss of data or data being rendered inaccurate or losses sustained by you or third parties or a failure of the library to operate with any other software), even if such holder or other party has been advised of the possibility of such damages. (перевод: «16. Любой обладатель авторских прав или любая другая сторона, может изменять и/или распространять библиотеку, как разрешено выше, не несет перед вами ответственности за ущерб, включая любые общие, специальные, случайные или косвенные убытки, возникшие в результате использования или невозможности использования библиотеки (включая, но не ограничиваясь, потерю данных или неточными отображениями данных или потери, понесенные вами или третьими лицами, или невозможность библиотеки работать с любым другим программным обеспечением) , даже если такой держатель или другая сторона была предупреждена о возможности такого убытка»).

Данная лицензия не защищает от прекращения распространения библиотеки в страны, а также позволяет организовать платный доступ к исходным кодам.

Несмотря на то, что оригинальными авторами являлись следующие специалисты: Sage Weil, Yehuda Sadeh Weinraub, Gregory Farnum, Josh Durgin, Samuel Just, Wido den Hollander, - они работали в компании **Inktank Storage**, которая была ведущим участником разработки и финансовым спонсором распределённой файловой системы Ceph с открытым исходным кодом. **Red Hat** приобрела Inktank Storage за 175 миллионов долларов в апреле 2014 года.

На данный момент основными разработчиками CEPH являются Red Hat, Intel, CERN, Cisco, Fujitsu, SanDisk, Canonical и SUSE. Данные компании наложили ограничения на распространения собственной продукции для компаний и государственных органов Российской Федерации.

Компания Red Hat была уже замечена в ограничениях к исходному коду RHEL (дистрибутив Linux от компании Red Hat) - <https://servernews.ru/1088917>, <https://www.redhat.com/en/blog/furthering-evolution-centos-stream>, https://www.cnews.ru/news/top/2023-06-27_kod_rhel_linux_ischez_iz_svobodnogo, https://yandex.ru/q/article/red_hat_nanosit_sokrushitelnyi_udar_po_d83b3fc4/, а Broadcom также ограничивала доступ к исходному коду (https://www.comnews.ru/content/233630/2024-06-07/2024-w23/1007/zelenaya-sliva-upala-nezreloy-rossiyskie-subd-baze-greenplum-okazalis-zone-riska??utm_source=telegram&utm_medium=general&utm_campaign=general)

ВЫВОД: программное обеспечение с открытым исходным кодом под лицензией LGPLv2.1 не защищено от санкционных рисков, как и возможность прекращения публикации отдельных компонентов данного ПО компаниями разработчиками.

8.2.2.3 РИСКИ СВЯЗАННЫЕ С ОТКРЫТЫМ ПО

Разработка CEPH напрямую зависит от компании IBM, в которую была переведена вся команда разработчиков CEPH из приобретённой IBM компании Red Hat. Компания IBM на коммерческой основе предлагает клиентам сопровождение решений на CEPH, без данного сопровождения невозможно гарантировать обеспечение сохранности данных.

Согласно информации [, в настоящее время для CEPH зарегистрировано 7351 неисправленных программных ошибок, из них 10 со статусом Immediate, 129 со статусом Urgent и 329 со статусом High. Часть ошибок имеет статус критичных более 2 лет.](https://tracker.ceph.com/issues?utf8=%E2%9C%93&set_filter=1&sort=id%3Adesc&f%5B%5D=status_id&op%5Bstatus_id%5D=o&f%5B%5D=tracker_id&op%5Btracker_id%5D=%3D&v%5Btracker_id%5D%5B%5D=1&f%5B%5D=&c%5B%5D=project&c%5B%5D=tracker&c%5B%5D=status&c%5B%5D=priority&c%5B%5D=subject&c%5B%5D=assigned_to&c%5B%5D=updated_on&c%5B%5D=category&c%5B%5D=fixed_version&c%5B%5D=cf_3&group_by=&t%5B%5D=)

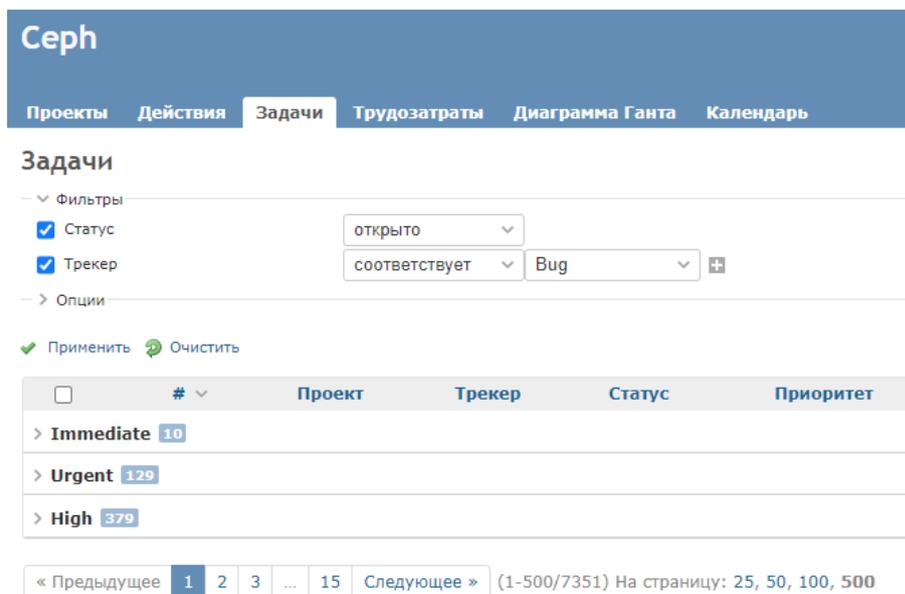


Рисунок 8 – Статус ошибок с CEPH

Документация CEPH не содержит информации о настройке решения и особенностях сопровождения действительно больших инсталляций (полноценная документация предоставляется компанией Rad Hat и IBM, что несёт риск сокрытия информации для всех пользователей без платной подписки). По опыту зарубежных и российских заказчиков известно, что поддержание системы в работоспособном состоянии требует штата высококвалифицированных специалистов в несколько десятков для сопровождения инфраструктуры и оперативного внесения изменений в код продукта.

ВЫВОД: использование ПО с открытым исходным кодом не означает, что сообщество будет консультировать и исправлять выявленные недочёты. Распространение исходных кодов – это один из способов инвестиции компании с целью получения прибыли на сопровождении данных систем.

8.2.2.4 РИСКИ СВЯЗАННЫЕ С АРХИТЕКТУРОЙ CEPH

- разработчики решения CEPH не рекомендуют использовать алгоритмы EC для файловых систем (Erasure coding, although supported for all types of storage (block, file and object), is not recommended for block and file as it delivers lower performance. – см. IBM Storage Ceph Concepts and Architecture Guide (REDP-5721-00);

- Нет гарантии быстрого восстановления данных, данные от Red Hat и IBM закрыты (IBM Storage Ceph does not restrict the OSD drive sizes to be used or the number of OSD drives per node. However, it is required that the cluster can recover from a node failure in less than 8 hours. Official recovery calculator tool to be used to determine node recovery time can be found here: <https://access.redhat.com/labs/rhsrc/>);
- восстановление данных – это крайне трудоемкая и сложная процедура, требующая близкой к 100% утилизации CPU на всех узлах восстановления, что означает необходимость глубокого анализа и изменения конфигурации решения;
- в текущем решении необходимо экспертное проектирование архитектуры CEPH, так как вендор, предлагающий своё собственное решение на основе CEPH не даёт рекомендации по нагрузке на сетевую подсистему;
- кластер CEPH сложно обновлять (затрагивает процессы обновления ОС, прикладного программного обеспечения, обновление самого продукта и т.д.). В частности, изменение топологии CEPH приводит либо к долговременной просадке в производительности, либо к даунтайму (<https://object-storage-ca-ymq-1.vexxhost.net/swift/v1/6e4619c416ff4bd19e1c087f27a43eea/www-assets-prod/summits/27/presentations/24003/slides/OpenStack-201911-EN-final-20191112.pdf>, “The error of Ceph cluster and recovery failure occur when we adjust the configuration of network equipment in the running state of Ceph cluster. It is suggested that if you need to adjust the physical network, stop all services of the Ceph cluster first.”);
- кластер CEPH имеет логический выделенный слой, при утере / повреждении которого весь кластер становится неработоспособным;
- при использовании CEPH помещаемые в хранилище данные адресуются по местоположению (логический том, файловая система, каталог и т.п), что требует соответствующего конфигурирования настроек серверов приложений и в целом достаточно трудозатратно и подвержено ошибкам;
- кластер CEPH использует сторонние библиотеки ЕС и не гарантирует их полную совместимость.

ВЫВОД: СЕРН сложное решение, которое требует высококвалифицированных специалистов, обладающих, как минимум, набором следующих навыков:

- Разбираться в аппаратных решениях на архитектурном уровне взаимодействия всех компонентов;
- Разбираться в алгоритмах работы всех подсистем СЕРН;
- Владеть навыками программирования на С и т.п.;
- Владеть навыками работы с ЕС;
- Владеть навыками конфигурирования сети ЦОД.

Без сопровождения разработчика решения СЕРН сопровождение данного продукта возможно только с внесением изменений в код продукта.

9. ПРЕИМУЩЕСТВА СИСТЕМ ХРАНЕНИЯ ДАННЫХ YADRO

9.1. TATLIN.UNIFIED GEN 2

TATLIN.UNIFIED GEN2 — надёжная система хранения данных класса midrange уровня для решения разноплановых задач с исключительными показателями плотности и стоимости владения:

- базы данных;
- системы виртуализации;
- аналитика данных
- и т.п.

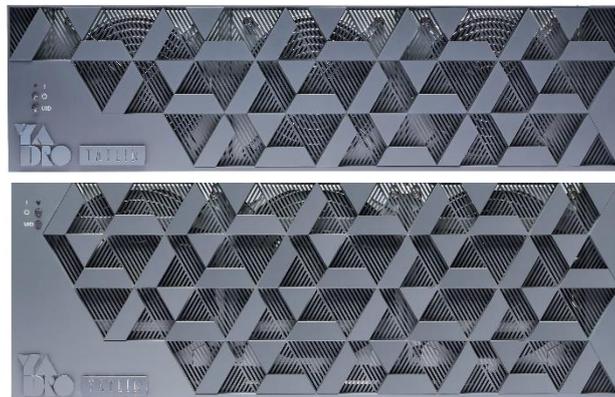


Рисунок 9 – внешний вид Tatlin.Unified GEN 2

TATLIN.UNIFIED GEN2 является высоко масштабируемым решением:

- Возможности расширения:
 - до 20 портов FC 32 Гб/с
 - до 40 портов FC 16 Гб/с
 - до 20 портов Ethernet 10/25 Гб/с
- До 574 SAS SSD и HDD накопителей
- Гибридное или all-flash хранилище по выбору
- Поддержка «тонких» томов (thin provisioning)

TATLIN.UNIFIED GEN2 предоставляет надёжные решения хранения данных:

- Защита данных T-RAID
 - Алгоритм защиты данных Erasure coding;
 - Возможность одновременной потери до 8 дисков в рамках единого пула;
 - Быстрое восстановление. При выходе из строя диска, его данные восстанавливаются на всех дисках одновременно;

- Отсутствие барьеров. Возможность собрать сотни дисков в единый логический пул благодаря технологиям EC;
 - Гибкие политики защиты. ПО позволяет подбирать необходимую модель защиты в соответствии с TCO;
 - Дополнительные настройки. При создании логического пула можно выбрать требуемый stripe size;
- Синхронная репликация данных;
 - Энергонезависимая кэш-память;
 - Работа контроллеров в режиме Symmetric Active-Active;
 - Резервирование и горячая замена ключевых компонентов;
 - Обновление системы без потери доступа к данным.

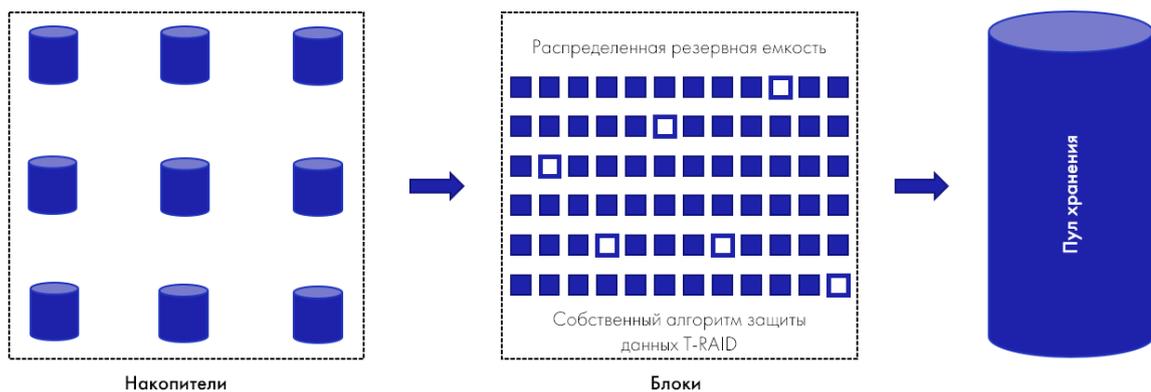


Рисунок 10 – защита данных T-RAID в Tatlin.Unified Gen 2

TATLIN.UNIFIED GEN2 – это система хранения данных с высоким уровнем сервиса и поддержки:

- Полная локальная поддержка на всех уровнях от L0 до L3;
- Возможности расширенного пакета поддержки;



Рисунок 11 – внешний вид интерфейса Tatlin.Unified GEN 2

- Курсы по системе на базе профессионального учебного центра;
- Документация на русском в печатном и веб-формате.

TATLIN.UNIFIED GEN2 – это собственная разработка и производство в Российской Федерации.

9.2. TATLIN.BACKUP.M

TATLIN.BACKUP — системы для хранения резервных копий:

- оптимальная производительность – это оптимизированная под задачу хранения и восстановления существующих массивов данных аппаратная платформа;
- надёжность хранения данных – это специальные механизмы защиты хранения данных: верификация записи end-to-end, аудит хранения и другие;
- минимизация ТСО – это эффективное сжатие данных с помощью механизмов глобальной дедупликации и компрессии;
- восстановление данных - соблюдение RTO (Recovery Time Objectives) & RPO (Recovery Point Objectives).



Рисунок 12 – внешний вид
Tatlin.Backup.M

Применяемые технологии сжатия данных:

- дедупликация блоками переменной длины позволяет существенно сократить объем хранения. Применяемая технология «Content Defined Chunking» позволяет разбивать данные на блоки переменной длины, где граница блока определяется на основе хеш-алгоритмов. Такие блоки можно найти в потоке данных, независимо от того, какие новые данные были добавлены в любом месте;
- глобальная inline дедупликация («на лету») позволяет искать повторяющиеся блоки по всему массиву данных, что обеспечивает оптимизацию используемого дискового пространства и дает возможность хранить данные многочисленных клиентов на одной СХД;

— компрессия по алгоритму ZSTD обеспечивает сжатие до 2.5:1 в рамках фрагмента.

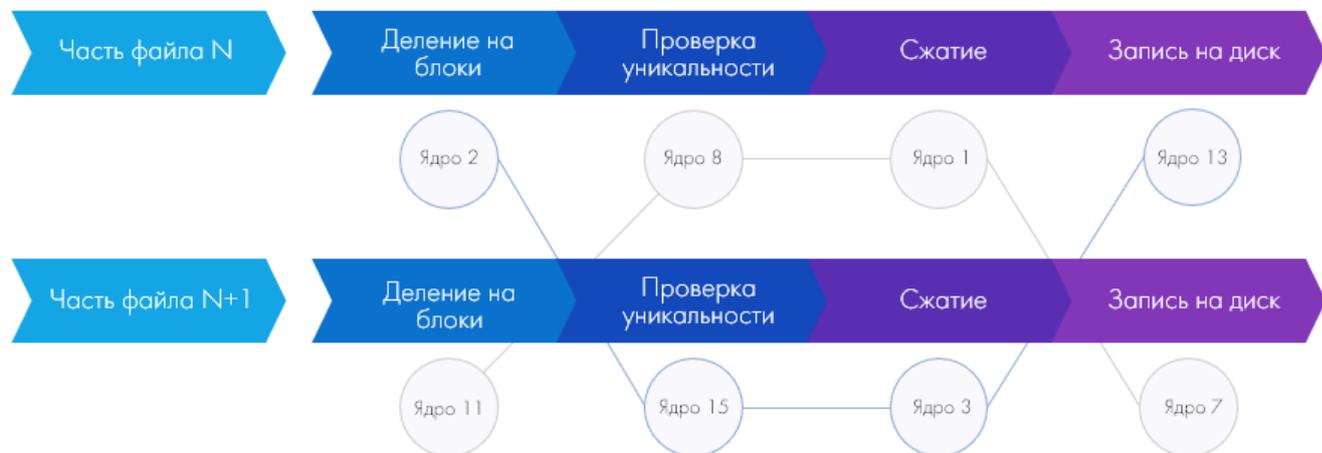


Рисунок 13 - многопоточная дедупликация блоками переменной длины

TATLIN.BACKUP использует продвинутые механизмы защиты данных:

— верификация записи end-to-end;

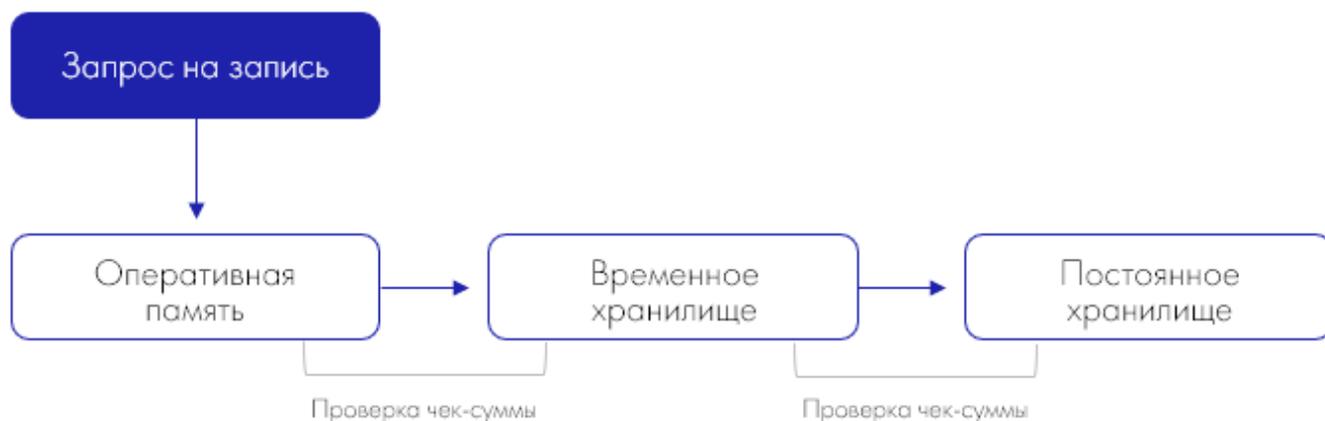


Рисунок 14 - Верификация записи end-to-end

— копирование при записи. Старые данные всегда сохранены, применение новой записи всегда создаёт новое консистентное состояние файловой системы, от которого можно стартовать в случае outage;



Рисунок 15 - Копирование при записи

— проверка целостности записанных данных:

- RAID Scrubbing . Процесс периодического считывания данных для проверки чек-сумм. При несовпадении чек-сумм происходит восстановление данных. Scrubbing применяется как к основным данным, так и к мета-данным;
- Garbage collector. Бэкграунд-процесс проверки всего пространства хранения для освобождения хранилища от неиспользуемых блоков, также процесс проверяет блоки на консистентность;

— упреждающая журнализация. Для защиты данных при сбоях используется техника упреждающей журнализации (Write-Ahead Logging). Информация об изменениях вносится в журнал предзаписи (Write-Ahead log) и фиксируется перед записью;

— программный RAID6 защищает от одновременного выхода одного или двух дисков из строя;

— при чтении данных хэш блоки одновременно являются чек-суммой читаемых данных, что обеспечивает консистентность прочитанных данных.

10. ВЫВОДЫ

Предлагаемая архитектура подсистем хранения, передачи и резервирования данных на базе продуктов **YADRO** обладают рядом неоспоримых технологических преимуществ, которые обеспечивают максимально надёжное и эффективное применение в составе решений ГИС НБГИ.

Изменение этапности проведения проекта позволяет сразу реализовать все три контура хранения и приступить к разработке системы в полном объёме, масштабируя ресурсы хранения по мере необходимости.

11. ПРИЛОЖЕНИЯ

11.1. ОСНОВНЫЕ НОРМАТИВНЫЕ ПРАВОВЫЕ И ТЕХНИЧЕСКИЕ ДОКУМЕНТЫ

- Указ Президента Российской Федерации от 1 декабря 2016 г. № 642 «О Стратегии научно-технологического развития Российской Федерации»;
- Указ Президента Российской Федерации от 7 мая 2018 г. № 204 «О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года»;
- Указ Президента Российской Федерации от 21 июля 2020 г. № 474 «О национальных целях развития Российской Федерации на период до 2030 года»;
- Указ Президента Российской Федерации от 2 июля 2021 г. № 400 «О Стратегии национальной безопасности Российской Федерации»;
- Указ Президента Российской Федерации от 30 марта 2022 г. № 166 «О мерах по обеспечению технологической независимости и безопасности критической информационной инфраструктуры российской федерации»;
- Федеральный закон от 5 июля 1996 г. № 86-ФЗ «О государственном регулировании в области генно-инженерной деятельности»;
- Федеральный закон от 27 июля 2006 г. № 149-ФЗ «Об информации, информационных технологиях и о защите информации»;
- Федеральным закон от 26 июля 2017 г. № 187-ФЗ «О безопасности критической информационной инфраструктуры Российской Федерации»;
- Федеральный закон от 31 июля 2020 г. № 247-ФЗ «Об обязательных требованиях в Российской Федерации»;
- Федеральный закон от 29 декабря 2022 г. № 643-ФЗ «О внесении изменений в Федеральный закон «О государственном регулировании в области генно-инженерной деятельности»;
- постановление Правительства Российской Федерации от 8 сентября 2010 г. № 697 «О единой системе межведомственного электронного взаимодействия»;

- постановление Правительства Российской Федерации от 8 июня 2011 г. № 451 «Об инфраструктуре, обеспечивающей информационно-технологическое взаимодействие информационных систем, используемых для предоставления государственных и муниципальных функций в электронной форме»;
- постановление Правительства Российской Федерации от 22 декабря 2012 г. № 1382 «О присоединении информационных систем организаций к инфраструктуре, обеспечивающей информационно-технологическое взаимодействие информационных систем, используемых для предоставления государственных и муниципальных услуг и исполнения государственных и муниципальных функций в электронной форме»;
- постановление Правительства Российской Федерации от 6 июля 2015 г. № 676 «О требованиях к порядку создания, развития, ввода в эксплуатацию, эксплуатации и вывода из эксплуатации государственных информационных систем и дальнейшего хранения содержащейся в их базах данных информации»;
- постановление Правительства Российской Федерации от 16 ноября 2015 г. № 1236 «Об установлении запрета на допуск программного обеспечения, происходящего из иностранных государств, для целей осуществления закупок для обеспечения государственных и муниципальных нужд» в части, касающейся, единого реестра российских программ для электронных вычислительных машин и баз данных»;
- постановление Правительства Российской Федерации от 5 мая 2016 г. № 392 «О приоритетных направлениях использования и развития информационно-коммуникационных технологий в федеральных органах исполнительной власти и органах управления государственными внебюджетными фондами и о внесении изменений в некоторые акты Правительства Российской Федерации»;
- постановление Правительства Российской Федерации от 8 февраля 2018 г. № 127 «Об утверждении Правил категорирования объектов критической информационной инфраструктуры Российской Федерации, а также перечня

- показателей критериев значимости объектов критической информационной инфраструктуры Российской Федерации и их значений»;
- постановление Правительства Российской Федерации от 10 июля 2019 г. № 878 «О мерах стимулирования производства радиоэлектронной продукции на территории Российской Федерации при осуществлении закупок товаров, работ, услуг для обеспечения государственных и муниципальных нужд, о внесении изменений в постановление Правительства Российской Федерации от 16 сентября 2016 г. № 925 и признании утратившими силу некоторых актов Правительства Российской Федерации»;
 - постановление Правительства Российской Федерации от 22 апреля 2019 г. № 479 «Об утверждении Федеральной научно-технической программы развития генетических технологий на 2019 - 2030 годы»;
 - постановление Правительства Российской Федерации от 14 мая 2021 г. № 733 «Об утверждении положения о федеральной государственной информационной системе «Единая информационная платформа национальной системы управления данными» и о внесении изменений в некоторые акты Правительства Российской Федерации»;
 - постановление Правительства Российской Федерации от 31 января 2024 г. № 87 «О государственной информационной системе в области генетической информации «Национальная база генетической информации»;
 - распоряжение Правительства Российской Федерации от 26 июля 2016 г. № 1588-р «Об утверждении плана перехода органов исполнительной власти и государственных внебюджетных фондов на использование отечественного программного обеспечения»;
 - приказ Минкомсвязи России от 27 июня 2013 г. № 149 «Об утверждении Требований к технологическим, программным и лингвистическим средствам, необходимым для размещения информации государственными органами и органами местного самоуправления в сети «Интернет» в форме открытых данных, а также для обеспечения ее использования»;
 - приказ Минкомсвязи России от 3 мая 2014 г. № 120 «Об утверждении требований, обеспечивающих технологическую совместимость

- информационных систем организаций, подключаемых к инфраструктуре, обеспечивающей информационно-технологическое взаимодействие информационных систем, используемых для предоставления государственных и муниципальных услуг в электронной форме с указанной инфраструктурой, к каналу связи и используемым для его защиты средствам криптографической защиты информации, а также особенностей использования стандартов и протоколов при обмене данными в электронной форме между информационными системами указанных организаций и инфраструктурой»;
- приказ Минкомсвязи России от 23 июня 2015 г. № 210 «Об утверждении Технических требований к взаимодействию информационных систем в единой системе межведомственного электронного взаимодействия»;
 - приказ Минцифры России и ФСО России от 4 декабря 2020 г. № 667/233 «Об утверждении Требований к организационно-техническому взаимодействию государственных органов и государственных организаций»;
 - приказ ФСБ России и ФСТЭК России от 31 августа 2010 г. № 416/489 «Об утверждении Требований о защите информации, содержащейся в информационных системах общего пользования»;
 - приказ ФСБ России от 24.07.2018 № 367 «Об утверждении Перечня информации, представляемой в государственную систему обнаружения, предупреждения и ликвидации последствий компьютерных атак на информационные ресурсы Российской Федерации и Порядка представления информации в государственную систему обнаружения, предупреждения и ликвидации последствий компьютерных атак на информационные ресурсы Российской Федерации»;
 - приказ ФСБ России от 19.06.2019 № 282 «Об утверждении Порядка информирования ФСБ России о компьютерных инцидентах, реагирования на них, принятия мер по ликвидации последствий компьютерных атак, проведенных в отношении значимых объектов критической информационной инфраструктуры Российской Федерации»;

- приказ ФСБ России от 24.10.2022 № 524 «Об утверждении Требований о защите информации, содержащейся в государственных информационных системах, с использованием шифровальных (криптографических) средств»;
- приказ ФСБ России от 13.02.2023 № 77 «Об утверждении порядка взаимодействия операторов с государственной системой обнаружения, предупреждения и ликвидации последствий компьютерных атак на информационные ресурсы Российской Федерации, включая информирование ФСБ России о компьютерных инцидентах, повлекших неправомерную передачу (предоставление, распространение, доступ) персональных данных»;
- приказ ФСТЭК России от 11 февраля 2013 г. № 17 «Об утверждении Требований о защите информации, не составляющей государственную тайну, содержащейся в государственных информационных системах»;
- приказ ФСТЭК России от 18 февраля 2013 г. № 21 «Об утверждении Составы и содержания организационных и технических мер по обеспечению безопасности персональных данных при их обработке в информационных системах персональных данных»;
- приказ ФСТЭК России от 25 декабря 2017 г. № 239 «Об утверждении Требований по обеспечению безопасности значимых объектов критической информационной инфраструктуры Российской Федерации»;
- меры защиты информации в государственных информационных системах, утверждёнными ФСТЭК России от 11 февраля 2014 г..

11.2. СОСТАВ ОБОРУДОВАНИЯ YADRO

11.2.1. СОСТАВ ОБОРУДОВАНИЯ. ПЕРВЫЙ ЭТАП.

Артикул	Наименование	Количество
Открытый контур		
Y05X81U2S101A_5D1E94	Открытый контур // Сервер Тип 1 // Web-сервер/балансировщик (Angie pro/NginX) - Узел виртуализации	4
Y05X82U2S101A_5A370C	Открытый контур // Сервер Тип 2 // СУБД (PosgressPro или аналог) - Узел виртуализации	2
Y16SYSBAS201A_19011B	Открытый контур // Сервер Тип 3 // СХД общего назначения	1
Y05X81U2S101A_C0118D	Открытый контур // Сервер Тип 4 // Инфраструктурные	3
Y14X22052U00A_6AA036	Открытый Контур // Т.О. Головы (32ПБ ECC 6+2)	1
Y14X22052U00A_B3202A	Открытый Контур // Т.О. - СРК. Головы (10ПБ ECC 6+2)	1
Y19K2S128W2P7_4A662F	Открытый контур // TATLIN.BACKUP	1
Y40SYSCOM101U_E20D2B	Открытый контур // Коммутатор KORNFELD D1156	26
Конфиденциальный контур		
Y05X81U2S101A_169BC6	Конфиденциальный контур // Сервер Тип 1 // Web-сервер/балансировщик (Angie pro/NginX) - Узел виртуализации	4
Y05X82U2S101A_BF707D	Конфиденциальный контур // Сервер Тип 5 // Обработчик (сервер приложений) - Сервер	6
Y05X82U2S101A_70B684	Конфиденциальный контур // Сервер Тип 2 // СУБД (PosgressPro или аналог) - Узел виртуализации	2
Y16SYSBAS201A_5B440F	Конфиденциальный контур // Сервер Тип 3 // СХД общего назначения	1
Y05X82U2S101A_06F20E	Конфиденциальный контур // Сервер Тип 6 // Высоконагруженный кластер	10
Y05X81U2S101A_2814B8	Конфиденциальный контур // Сервер Тип 4 // Инфраструктурные	3
Y14X22052U00A_2C2FBE	Конфиденциальный Контур// СХД -Головы (20ПБ ECC 6+2)	1
Y14X22052U00A_9F7BFC	Конфиденциальный Контур//СХД-ПК. Головы (6ПБ ECC 6+2)	1
Y19K2S128W2P7_FDD53C	Конфиденциальный контур//СРК - TATLIN.BACKUP	1

Y40SYSCOM101U_E20D2B	Конфиденциальный контур // Коммутатор KORNFELD D1156	30
Специальный контур		
Y05X81U2S101A_1346EE	Специальный контур // Сервер Тип 1 // Web-сервер/балансировщик (Angie pro/NginX) - Узел виртуализации	2
Y05X82U2S101A_0148EF	Специальный контур // Сервер Тип 5 // Обработчик (сервер приложений) - Сервер	3
Y05X82U2S101A_F8354A	Специальный контур // Сервер Тип 2 // СУБД (PostgressPro или аналог) - Узел виртуализации	2
Y16SYSBAS201A_B03293	Специальный контур // Сервер Тип 3 // СХД общего назначения	1
Y05X82U2S101A_A150F8	Специальный контур // Сервер Тип 6 // Высоконагруженный кластер	6
Y05X81U2S101A_D6B696	Специальный контур // Сервер Тип 4 // Инфраструктурные	2
Y19K2S128W2P7_A78D15	Специальный контур // СРК - TATLIN.BACKUP	1
Y40SYSCOM101U_E20D2B	Специальный контур // Коммутатор KORNFELD D1156	20
Y14X22052U00A_288528	Специальный Контур // СХД - 1.5ПБ R2	1
Y14X22052U00A_F4F902	Специальный Контур // СХД -РК 1.5ПБ R2	1

11.2.2. СОСТАВ ОБОРУДОВАНИЯ. ВТОРОЙ ЭТАП.

Артикул	Наименование	Количество
Открытый контур		
Y16UPG0100001_65F9EA	Открытый контур // Комплект модернизации до 32ПБ	26
Y16UPG0100001_65F9EA	Открытый контур // Комплект модернизации до 10ПБ	8
Конфиденциальный контур		
Y16UPG0100001_65F9EA	Конфиденциальный контур // Комплект модернизации до 20ПБ	16
Y16UPG0100001_65F9EA	Конфиденциальный контур // Комплект модернизации до 6ПБ	8

11.2.3. СОСТАВ ОБОРУДОВАНИЯ. КОНТУР БЕЗОПАСНОСТИ.

Артикул	Наименование	Количество
Y05X81U2S101A_23E2EA	Сервер VEGMAN R120 G2 (Сервер, SIEM для высоконагруженных систем (для компонента Core, до 50K EPS))	1
Y05X81U2S101A_65ECD0	Сервер VEGMAN R120 (Сервер, SIEM для распределенных систем (для дополнительного высоконагруженного компонента Collector))	1
Y05X82U2S101A_C92C5A	Сервер VEGMAN R220 G2 Сервер, SIEM для высоконагруженных систем (для компонента Storage, до 15K EPS, до 2 недель распределённого хранения событий (до 11000 ГБ + 27000 ГБ хранилища))	1
Y15SYSBAS304R_AC1B35	Дисковый модуль расширения L78	1
Y05X82U2S101A_E60F22	Сервер VEGMAN R220 G2 (Сервер, Sandbox (для дополнительного узла к компоненту Server, для высоконагруженного поведенческого анализа))	2
Y05X82U2S101A_8EA501	Сервер VEGMAN R220 G2 Сервер , Sandbox (до 1000 Мб/с, статический и динамический анализ объектов, до 11000 ГБ встроенного хранилища просканированных объектов)	1
Y05X82U2S101A_30FA58	Сервер VEGMAN R220 G2 Сервер , для компонента Server, хранение оперативных метаданных до 17500 ГБ)	8
Y05X82U2S101A_9F92BB	Сервер VEGMAN R220 G2 (Сервер для компонента Capturer, производительность захвата трафика до 10000 Мб/с, хранение данных сырого трафика до 80000 ГБ)	4
Y05X81U2S101A_04C79F	Сервер VEGMAN R120 G2 (Сервер Base node, производительность до 4000 Мб/с, до 40000 RPS)	2
Y05X81U2S101A_4EFA63	Сервер VEGMAN R120 G2(Сервер Base node, производительность до 10000 Мб/с, до 100000 RPS)	4