

**Цифровые двойники для решения задач
управления и развития распределенных
центров сбора, хранения и обработки данных**

Пряхина Дарья Игоревна

Научная специальность 2.3.1. Системный анализ, управление и обработка информации, статистика

соискание ученой степени кандидата технических наук

Научный руководитель:
доктор технических наук
Кореньков Владимир Васильевич

Актуальность темы исследования



Data Center



Institute of High Energy Physics
Chinese Academy of Sciences



Распределенные центры для сбора, хранения и обработки данных (РЦОД):

- наиболее актуальные и применимые вычислительные инфраструктуры для масштабных научных экспериментов;
- географически распределенная инфраструктура, объединяющая множество ресурсов разных типов, доступ к которым осуществляется из любой точки, независимо от места их расположения.

Актуальность темы исследования



Data Center



Institute of High Energy Physics
Chinese Academy of Sciences



Проблемы при решении задач проектирования, управления и развития РЦОД:

- проверка сценариев масштабирования РЦОД, учитывая требования к потокам данных и задач;
- формирование и улучшение стратегии управления потоками задач для эффективного распределения ресурсов РЦОД при обработке данных;
- анализ используемых ресурсов и оценка необходимого количества ресурсов для конкретных задач согласно предъявленным требованиям к РЦОД.

Актуальность темы исследования



Data Center



Institute of High Energy Physics
Chinese Academy of Sciences



К решению перечисленных задач могут применяться:

– проблемно-ориентированные системы управления и принятия решений на основе цифровых двойников (ЦД).

Представляется актуальным:

– разработка метода и реализация алгоритмов построения цифровых двойников РЦОД для решения задач проектирования, управления и развития РЦОД, в том числе для проверки эффективности и надежности их функционирования.

Цель и задачи исследования

Цель

Разработка метода и алгоритмов создания цифровых двойников для описания распределенных систем, принятия решений по выбору конфигураций оборудования в рамках задачи масштабирования и развития распределенных центров сбора, хранения и обработки данных, управления ресурсами и процессами, происходящими в них.

Задачи

1. Провести анализ предметной области.
2. Разработать метод построения цифровых двойников РЦОД, который выполняет моделирование распределенных центров с учетом характеристик потоков задач и данных для хранения и обработки, а также вероятностей появления изменений в процессах, происходящих в РЦОД.
3. Реализовать алгоритмы, структуру базы данных (БД) и веб-интерфейс для построения и запуска ЦД, предоставления графической информации о результатах работы. Разработать специальное программное обеспечение (СПО), позволяющее сравнивать эффективность функционирования РЦОД в зависимости от конфигураций оборудования.
4. Провести верификацию и опытную эксплуатацию СПО.

Глава 1. Анализ задачи построения цифровых двойников распределенных центров сбора, хранения и обработки данных

Недостатки существующих средств для построения ЦД центров обработки данных (Future Facilities, Sunbird DCIM):

- ЦД – виртуальная копия центра обработки данных (ЦОД), расположенного в одном физическом пространстве;
- ЦОД рассматриваются только с точки зрения инженерной инфраструктуры;
- ЦД не позволяют формировать стратегии управления потоками задач.

Цифровой двойник (ЦД):

- отражает архитектуру распределенных центров сбора, хранения и обработки данных (РЦОД); процессы, происходящие в системе и связанные с потоками данных и задач;
- позволяет моделировать и тестировать новые конфигурации оборудования для проектирования и развития РЦОД.

Главный компонент ЦД (ядро) – компьютерная модель.

Глава 1. Анализ задачи построения цифровых двойников распределенных центров сбора, хранения и обработки данных

Недостатки существующих средств моделирования РЦОД (библиотеки Bricks, OptorSim и GridSim; комплекс SyMSim)

– узкоспециализированные устаревшие библиотеки не учитывают облачные структуры, суперкомпьютеры, системы иерархической памяти и др.;

– не рассматриваются важные критерии функционирования системы (потери данных, вероятности сбоев, отказов и изменений в производительности различных ресурсов);

– детальное описание параметров в БД;

– не предусмотрен поиск конфигурации оборудования по заданным критериям;

– требуется адаптация программного кода для каждой моделируемой инфраструктуры;

– отсутствует пользовательский интерфейс.

Глава 1. Анализ задачи построения цифровых двойников распределенных центров сбора, хранения и обработки данных

Выводы

Необходимо разработать метод и алгоритмы построения цифровых двойников РЦОД:

- для проведения исследований эффективности и надежности РЦОД как при проектировании, так и в процессе функционирования;
- для управления РЦОД, формирования стратегий управления потоками данных и задач;
- для проверки различных сценариев масштабирования, учитывая требования к потокам данных и задач.

Глава 2. Модели, методы и алгоритмы построения цифровых двойников распределенных центров сбора, хранения и обработки данных

Требования к цифровым двойникам РЦОД

К ядру	К пользовательскому интерфейсу
<ul style="list-style-type: none">– учитывать возможные вероятности изменения процессов функционирования системы (например, потери данных, изменение производительности и т.д.);– предусмотреть применимость для моделирования процессов любых РЦОД;– придерживаться принципа разработки универсального программного пакета.	<ul style="list-style-type: none">– предусмотреть раздел для построения, описания, изменения инфраструктуры РЦОД;– предоставлять возможность задавать параметры базовой конфигурации оборудования РЦОД, характеристики потоков данных и задач для обработки;– реализовать раздел для создания вычислительных экспериментов для поиска требуемой конфигурации оборудования;– предусмотреть возможность формировать сценарии масштабирования РЦОД;– реализовать функции запуска ЦД и просмотра результатов в виде графиков.

Глава 2. Модели, методы и алгоритмы построения цифровых двойников распределенных центров сбора, хранения и обработки данных

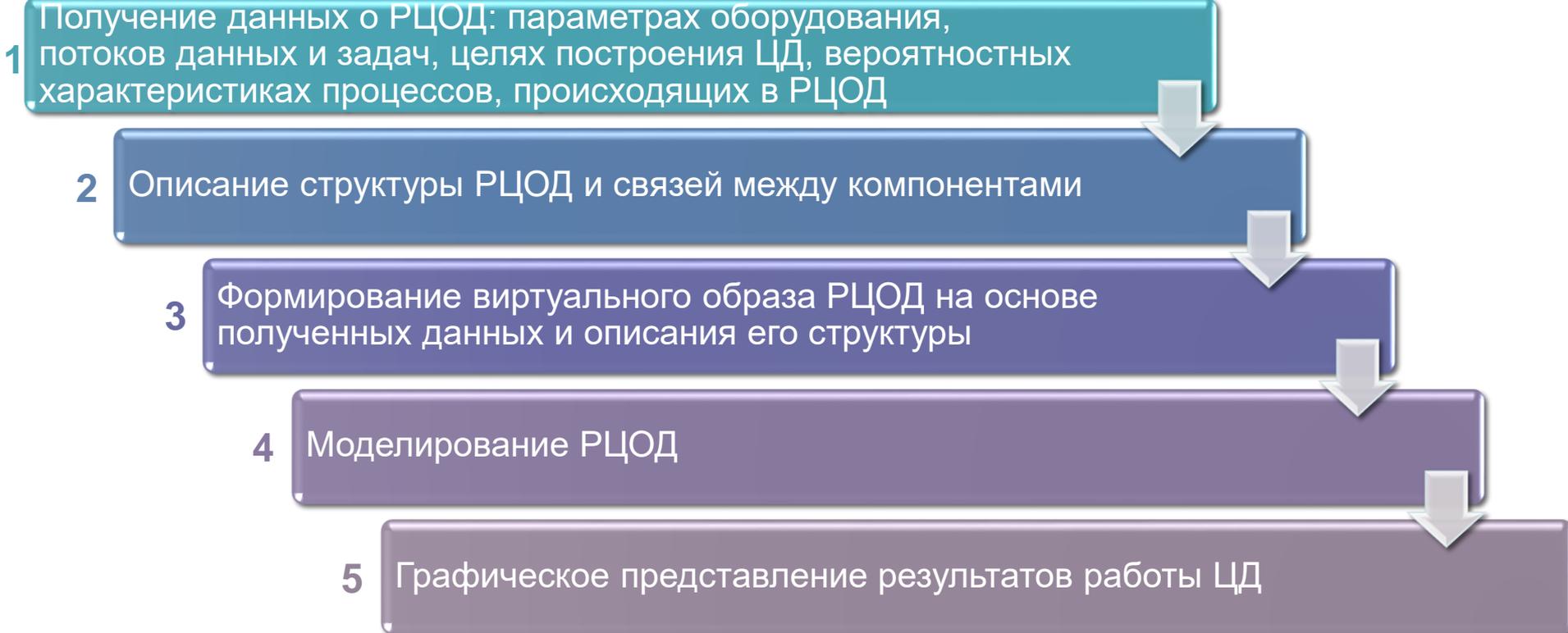
Требования к цифровым двойникам РЦОД

Нефункциональные требования к методу построения ЦД:

- различные параметры и результаты работы ЦД должны храниться в БД;
- скорость работы ЦД должна определяться исходя из масштаба РЦОД и процессов, происходящих в нем, но **время от момента запуска до получения результатов не должно превышать 24 часов**;
- адекватность ЦД должна быть проверена результатами работы существующего РЦОД по разным параметрам, которые **могут отличаться не более, чем на 20%**;
- работоспособность системы не должна зависеть от стороннего программного обеспечения.

Глава 2. Модели, методы и алгоритмы построения цифровых двойников распределенных центров сбора, хранения и обработки данных

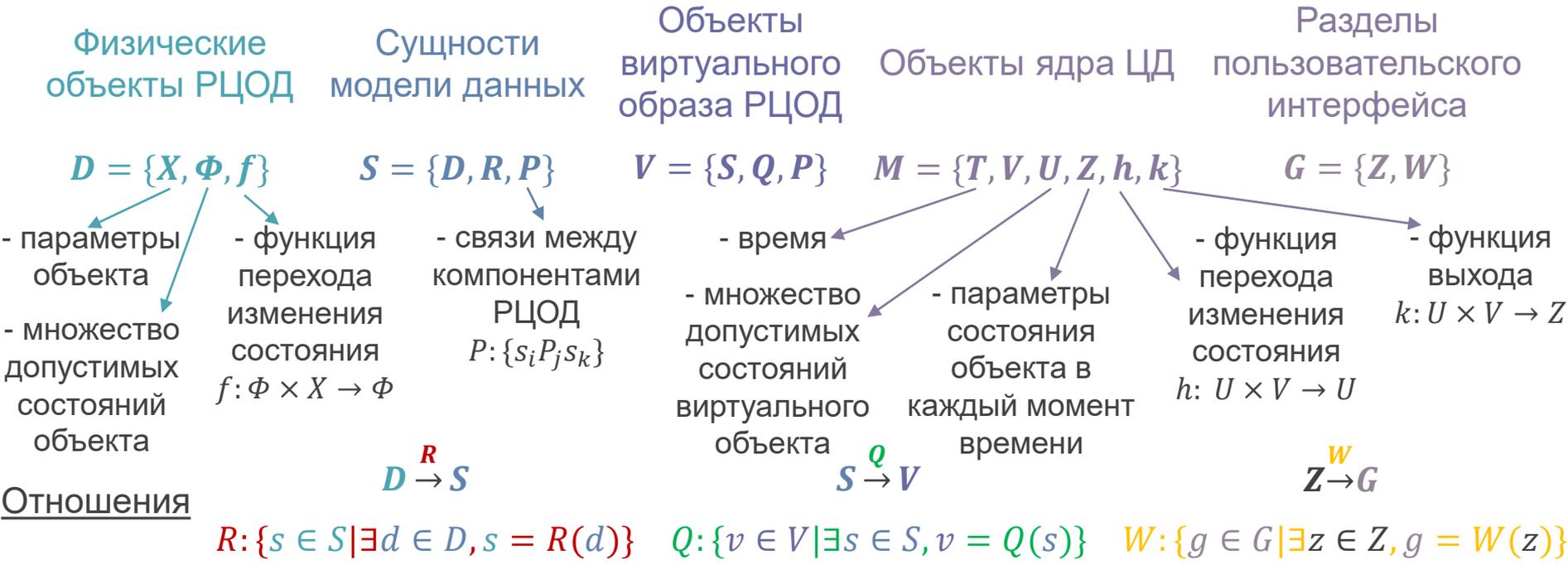
Метод построения цифровых двойников РЦОД



Глава 2. Модели, методы и алгоритмы построения цифровых двойников распределенных центров сбора, хранения и обработки данных

Система для реализации метода построения цифровых двойников РЦОД

Формальное представление: $MDT\{D, S, V, M, G, R, Q, W\}$



Глава 2. Модели, методы и алгоритмы построения цифровых двойников распределенных центров сбора, хранения и обработки данных

Система для реализации метода построения цифровых двойников РЦОД

Задача принятия решения:

$$\{\{\psi\}, \Omega\} \rightarrow \psi^*$$

$\{\psi\}$ — множество альтернатив

ψ^* — выбранная альтернатива

Ω — принцип выбора

Пример: поиск конфигурации оборудования и стратегии управления потоком задач, которая обеспечит обработку всех данных

- за минимальное количество времени;
- с использованием наименьшего количества ресурсов;
- и др.

Показатель эффективности:

$$\alpha = \langle H_{ц}, T \rangle$$

$H_{ц}$ — целевые эффекты

T — затраты времени

Поиск альтернативы:

- наиболее предпочтительный критерий

$$criteria_1(\psi) \rightarrow \min;$$

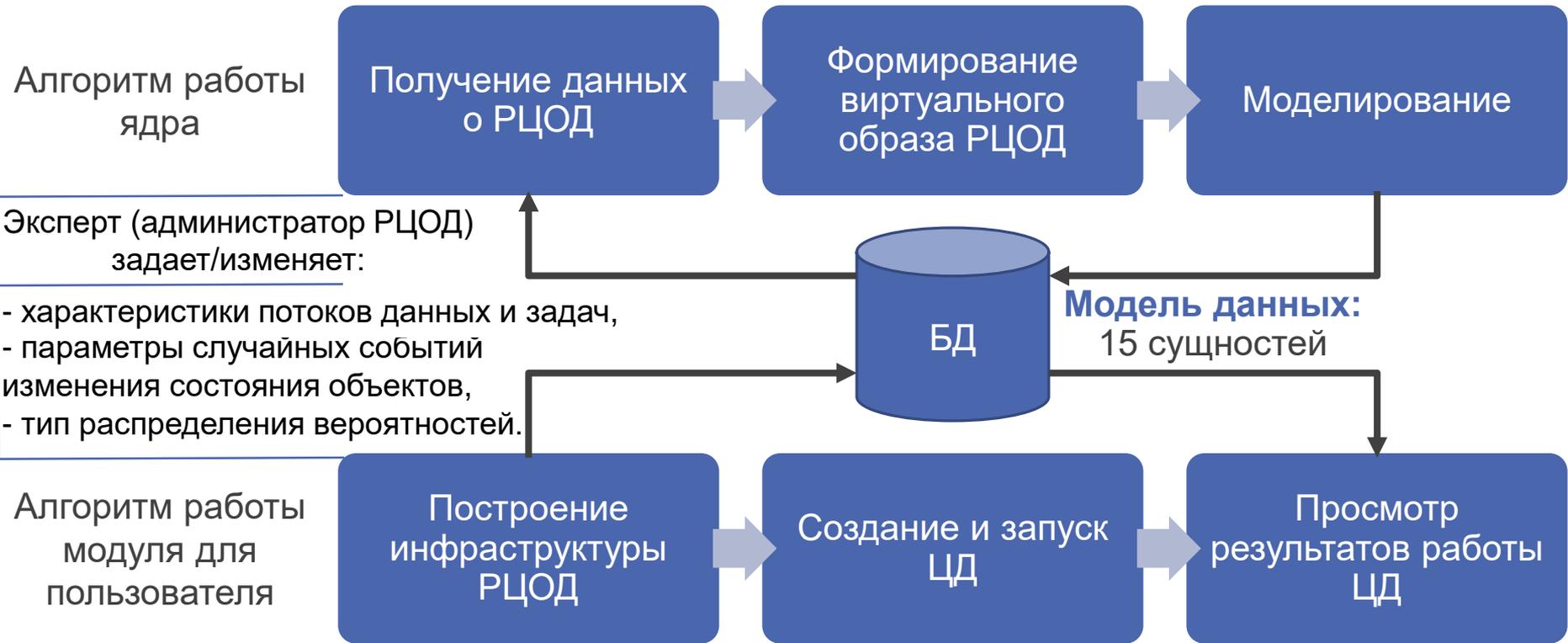
$$criteria_j(\psi) \leq c_j; j = \overline{2, N} \quad (N \in \mathbb{N})$$

- множество критериев

- заданное допустимое значение j-го критерия

Глава 2. Модели, методы и алгоритмы построения цифровых двойников распределенных центров сбора, хранения и обработки данных

Алгоритмы для реализации метода построения цифровых двойников РЦОД



Глава 2. Модели, методы и алгоритмы построения цифровых двойников распределенных центров сбора, хранения и обработки данных

Алгоритмы для реализации метода построения цифровых двойников РЦОД

Формирование потока событий
(закон Пуассона):

$$\tau = -\frac{1}{\lambda} * \ln(r)$$

- интервал между случайными событиями

- среднее число событий в единицу времени

- равномерно распределенное случайное число [0;1]

Генерация параметров объектов

(объемы данных, время выполнения задач и т.д.):

$$pr(x) = \frac{1}{b-a}$$

$$pn(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- функция плотности вероятности равномерного распределения, где [a, b) — интервал изменения значения параметра

- функция плотности вероятности нормального распределения, где μ — среднее значение случайной величины, σ — стандартное отклонение



Выводы

- Определены функциональные и нефункциональные требования к цифровым двойникам РЦОД.
- Сформулирован новый метод построения и использования ЦД для решения задач управления и развития РЦОД, в том числе улучшения их технических характеристик.
- Особенность метода: возможностью моделировать процессы обработки и хранения данных, учитывая характеристики потоков данных и задач, вероятностей сбоев, отказов и изменений в производительности оборудования и других процессов, происходящих в системе.
- Для реализации метода разработаны алгоритмы работы ядра и модуля для пользователя, создана модель данных.

Глава 3. Реализация алгоритмов и разработка специального программного обеспечения для создания цифровых двойников и взаимодействия с ними

Реализация	Ядро ЦД	Модуль для пользователя
Состав программы	<ul style="list-style-type: none">– абстрактный класс и его реализации для описания объектов модели данных;– классы планировщиков перемещения данных и управления задачами;– дополнительные классы для взаимодействия с БД, генерации вероятностных событий и случайных значений по требуемым распределениям.	<ul style="list-style-type: none">– раздел создания структуры РЦОД, настройки базовой конфигурации оборудования, потоков данных и задач;– раздел создания вычислительного эксперимента;– раздел для запуска ЦД;– раздел просмотра результатов.
Особенности	<ul style="list-style-type: none">– объектно-ориентированный подход;– универсальная программа (не требует изменений кода под каждую инфраструктуру РЦОД).	<ul style="list-style-type: none">– реализация адаптивных веб-страниц;– возможно расширение функционала.
Средства реализации	Python + библиотеки psycopg2, configparser, argparse, NumPy.	HTML, CSS, Bootstrap, JavaScript + библиотеки Cytoscape.js, Plotly.

Глава 3. Реализация алгоритмов и разработка специального программного обеспечения для создания цифровых двойников и взаимодействия с ними

Разработка специального программного обеспечения (СПО)



Состав СПО:

- хранение данных (БД);
- моделирование (ядро ЦД);
- представление результатов (модуль для пользователя).

Средства реализации:

- фреймворк Django;
- система управления БД PostgreSQL.

Глава 3. Реализация алгоритмов и разработка специального программного обеспечения для создания цифровых двойников и взаимодействия с ними

Построение инфраструктуры центра сбора, хранения и обработки данных

Настройка базовой конфигурации оборудования
* Обязательное поле для заполнения

- Название*:
- Описание:
- Объем: ТБ

Добавить устройство

1

2

3

Генератор данных

Вычислительная компонента

Хранилище данных

Роботизированная библиотека

Сохранить изображение

Завершить построение инфраструктуры

Настроить каналы связи

Настроить потоки данных

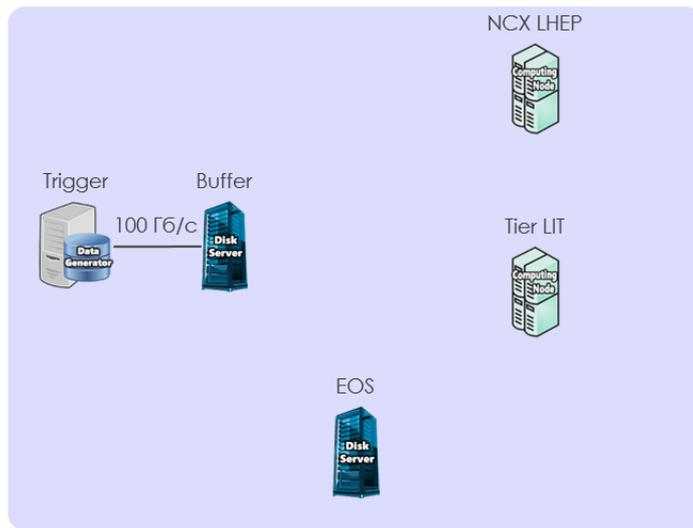
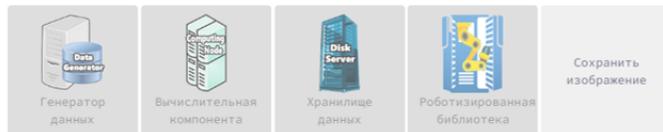
Настроить потоки задач

Создать цифровой двойник

Построение инфраструктуры и задание базовой конфигурации оборудования

Глава 3. Реализация алгоритмов и разработка специального программного обеспечения для создания цифровых двойников и взаимодействия с ними

Построение инфраструктуры центра сбора, хранения и обработки данных



Настройка каналов связи

Редактировать инфраструктуру

Настроить каналы связи

Настроить потоки данных

Настроить потоки задач

Создать цифровой двойник

Настройка каналов связи

* Обязательное поле для заполнения

Выберите существующий канал из списка, чтобы изменить параметры

link1

• Название*: link1

• Описание: trigger-buffer

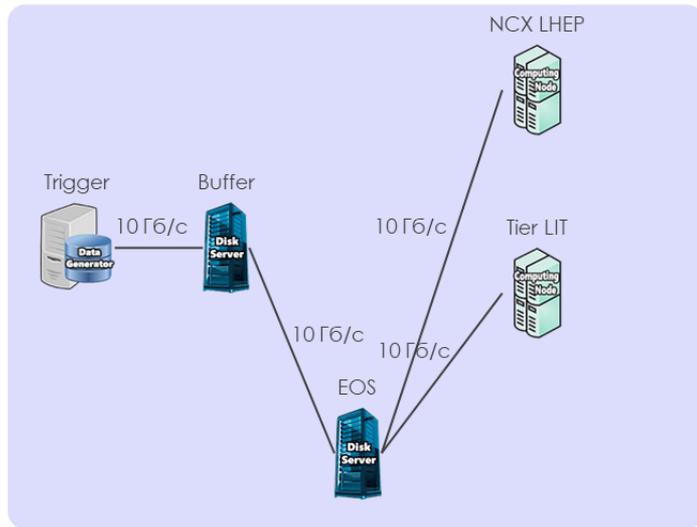
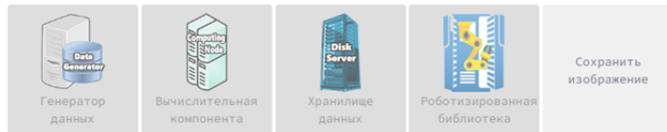
• Канал от*: Trigger до*: Buffer

• Пропускная способность: 10 Гб/с

Сохранить изменения

Глава 3. Реализация алгоритмов и разработка специального программного обеспечения для создания цифровых двойников и взаимодействия с ними

Построение инфраструктуры центра сбора, хранения и обработки данных



Редактировать инфраструктуру

Настроить каналы связи

Настроить потоки данных

Настроить потоки задач

Создать цифровой двойник

Настройка потоков задач

* Обязательное поле для заполнения

- Название*:
- Описание:
- Тип обрабатываемых данных*:
- Объем входных данных для задачи*: Гб
- Объем входных данных (допустимое отклонение): Гб
- Тип результирующих данных*:
- Объем выходных данных*: Гб
- Объем выходных данных (допустимое отклонение): Гб
- Распределение вероятности для генерации данных:
- Среднее время выполнения задачи*: с
- Среднее время выполнения (допустимое отклонение):
- Распределение вероятности для генерации задач:
- Общее количество задач *

Добавить

Настройка потоков данных и задач

Глава 3. Реализация алгоритмов и разработка специального программного обеспечения для создания цифровых двойников и взаимодействия с ними

Добавление эксперимента

Заполните поля формы, чтобы добавить новый эксперимент для поиска оптимальной конфигурации оборудования

* Обязательное поле для заполнения

Название эксперимента *

Test 1

Описание эксперимента

Поиск оптимального количества ресурсов для хранения данных

Параметры моделирования

- Продолжительность работы моделируемой инфраструктуры – ч.
- Ускорение процесса моделирования в раз.

Параметры логирования

Выберите объекты и события, о которых необходимо сохранять информацию во время моделирования

• Объекты моделируемой инфраструктуры

- Хранилища данных
- Вычислительные компоненты
- Каналы связи

• События

- Генерация данных
- Потери данных
- Работа с файлами
- Генерация, запуск, выполнение задач

Добавить

Очистить

Отмена

Эксперименты

Выберите существующий эксперимент или добавьте новый эксперимент для поиска оптимальной конфигурации оборудования

Test 4

Test

Дата создания: 9 марта 2023 г. 15:04

Test 3

Исследование загрузки каналов связи
Дата создания: 7 февраля 2023 г. 10:42

Test 2

Поиск оптимального количества вычислительных ресурсов
Дата создания: 7 февраля 2023 г. 10:38

Test 1

Поиск оптимального количества ресурсов для хранения данных
Дата создания: 7 февраля 2023 г. 10:36

Добавить

Глава 3. Реализация алгоритмов и разработка специального программного обеспечения для создания цифровых двойников и взаимодействия с ними

Информация об эксперименте

Дата создания: 7 февраля 2023 г. 10:36

Название эксперимента

Test 1

Описание эксперимента

Поиск оптимального количества ресурсов для хранения данных

Параметры моделирования

- Продолжительность работы моделируемой инфраструктуры – 800 ч.
- Ускорение процесса моделирования в 1000 раз.

Параметры логирования

• Объекты моделируемой инфраструктуры

- Хранилища данных
- Вычислительные компоненты
- Каналы связи

• События

- Генерация данных
- Потери данных
- Работа с файлами
- Генерация, запуск, выполнение задач

Посмотреть результаты

Выбрать другой эксперимент

Базовая конфигурация

Хранилища данных

Название	Описание	Объем (ТБ)
trigger	Trigger BM@N	1000,0
buffer	Data reception buffer	5400,0
eoslhep	Main storage LHEP	1000,0
eoslit	Main storage LIT	1000,0
dcach	pp	1000,0

Вычислительные компоненты

Название	Описание	Количество ядер
t2lit	LIT T2 farm	500
ncxlhep	LHEP main farm	1200
super	Govoron	190

Каналы связи

Название	Описание	Пропускная способность (Гб/с)
raw0	trigger - buffer	100,0
raw1	buffer - lhep	10,0
raw2	buffer - lit	10,0
compute0	lhep - farm lhep	10,0
compute1	lit - Govoron	10,0
compute2	lit - farm lit	10,0
dataeoslhepLit	eoslhep - eoslit	10,0
dataeosLitlhep	eoslit - eoslhep	10,0

Добавить модификацию

Список модификаций

№	Статус	Дата обновления	
16	NEW	9 марта 2023 г. 14:52	Просмотр Запуск Результаты
15	NEW	9 марта 2023 г. 14:52	Просмотр Запуск Результаты

Список событий

Название	Описание	
decrease	уменьшение количества ядер	Подробнее

Добавить событие

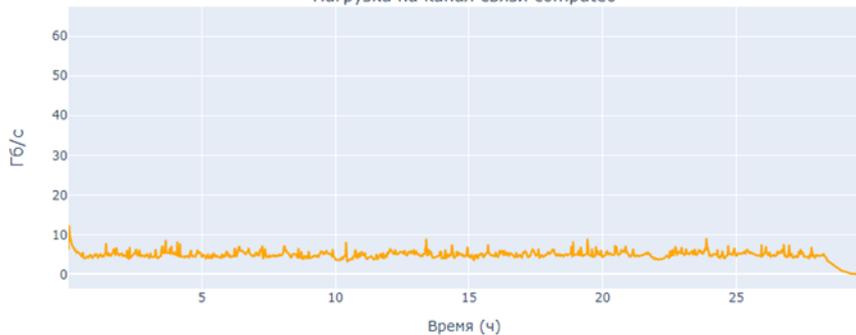
Глава 3. Реализация алгоритмов и разработка специального программного обеспечения для создания цифровых двойников и взаимодействия с ними

Результаты эксперимента Test 1

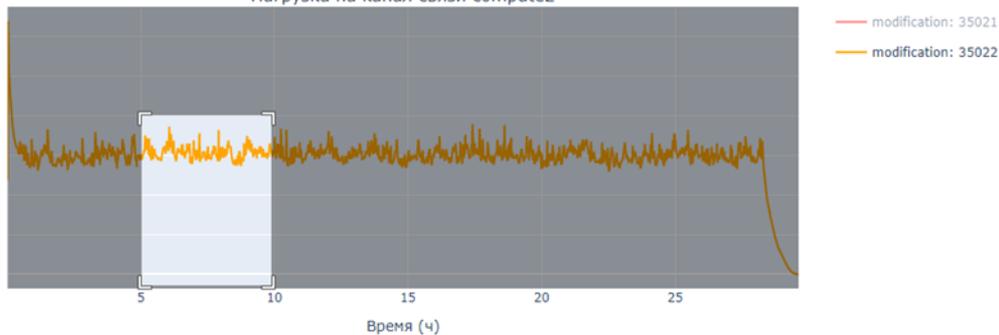
Выберите вкладку для просмотра результатов

Хранилища данных Вычислительные компоненты **Каналы связи** Очереди задач Распределения файлов

Нагрузка на канал связи compute0



Нагрузка на канал связи compute2



Глава 3. Реализация алгоритмов и разработка специального программного обеспечения для создания цифровых двойников и взаимодействия с ними

Выводы

- Реализованы алгоритмы построения цифровых двойников РЦОД с использованием современных архитектурных решений.
- На основе алгоритмов разработано СПО.
- **Получено свидетельство о государственной регистрации программы для ЭВМ «Программный комплекс для создания цифровых двойников распределенных центров сбора, хранения и обработки данных» от 14 августа 2023 года.**



Глава 4. Верификация и опытная эксплуатация специального программного обеспечения для создания цифровых двойников

Ускорительный комплекс NICA

BM@N



NICA
MPD

SPD
NICA

- Физика высоких энергий
- Изучение барионной материи
- Спиновая физика

Эксперименты
класса
«мегасайенс»



NICA

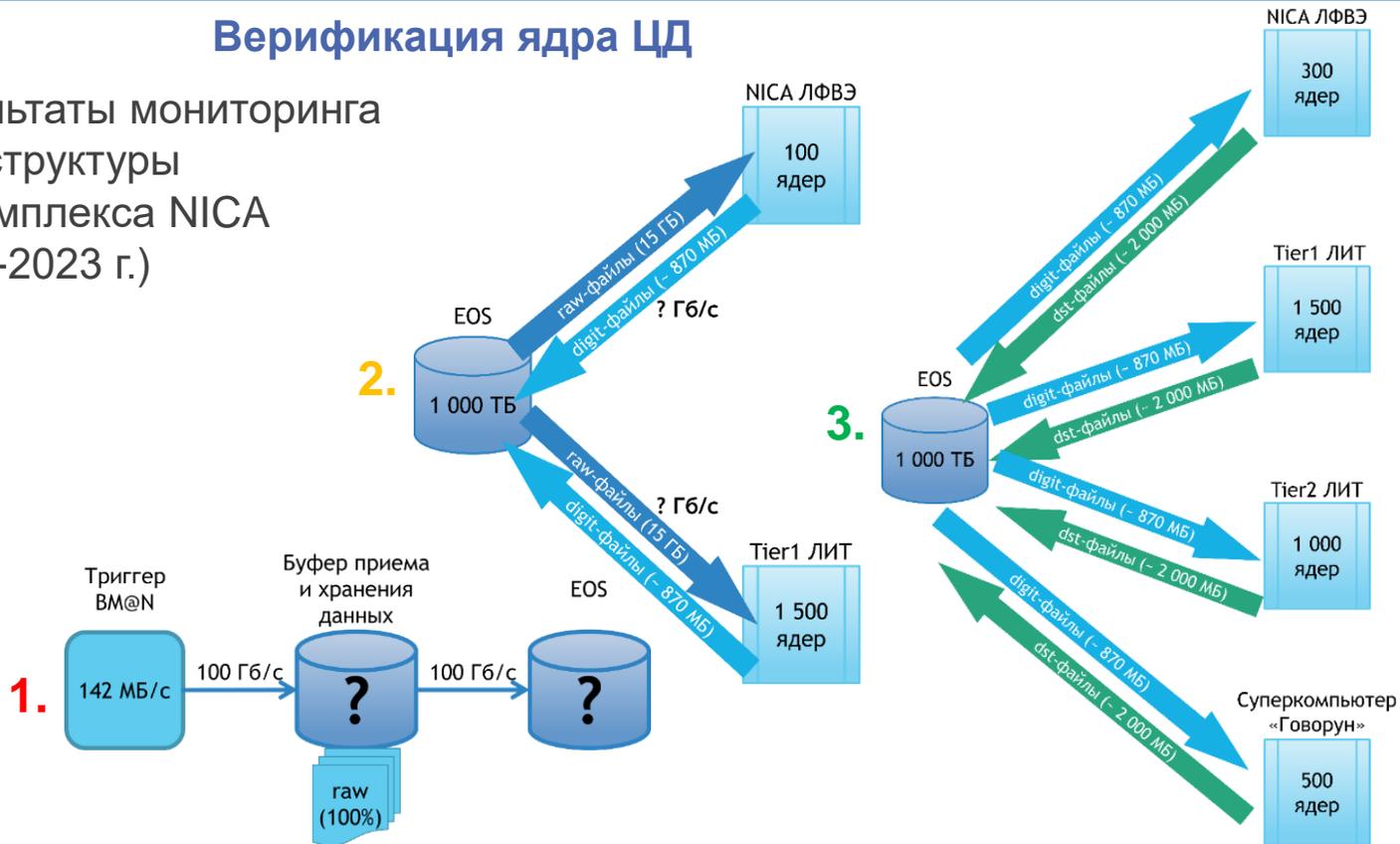
Глава 4. Верификация и опытная эксплуатация специального программного обеспечения для создания цифровых двойников

Верификация ядра ЦД

Исходные данные: результаты мониторинга вычислительной инфраструктуры эксперимента VM@N комплекса NICA (физический сеанс 2022-2023 г.)

Задачи:

1. Получение и хранение экспериментальных данных.
2. Выполнение задач преобразования экспериментальных данных.
3. Выполнение задач реконструкции событий.



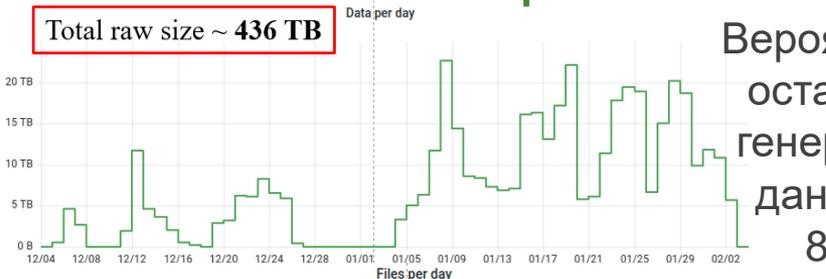
Глава 4. Верификация и опытная эксплуатация специального программного обеспечения для создания цифровых двойников

Верификация ядра ЦД

1. Получение и хранение экспериментальных данных.

Мониторинг

Total raw size ~ 436 TB



Вероятность
остановки
генератора
данных –
80%

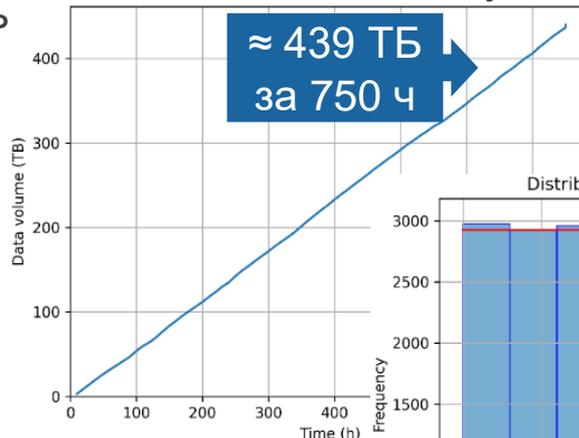
Total files: 31306



- Неравномерное распределение.
- В некоторые периоды данные не поступают.

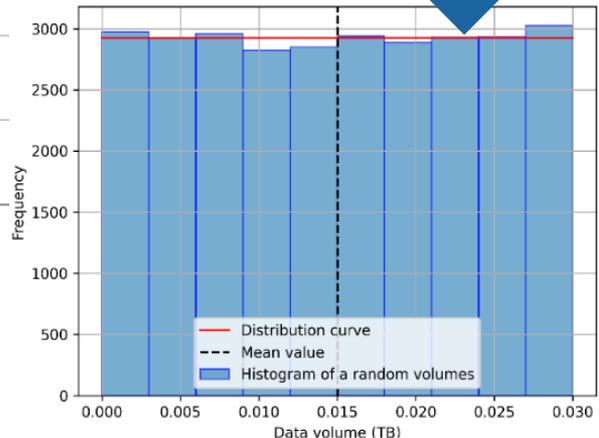
Моделирование

Total data volume on the Online storage



~ 29 241
raw файлов

Distribution of Raw Files on the



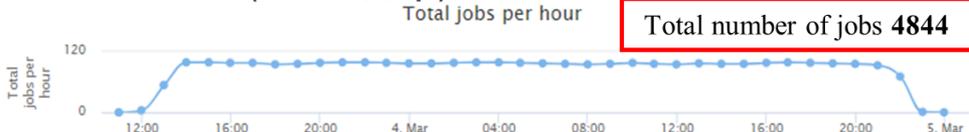
Глава 4. Верификация и опытная эксплуатация специального программного обеспечения для создания цифровых двойников

Верификация ядра ЦД

2. Выполнение задач преобразования экспериментальных данных.

Мониторинг

NCX LHEP (100 ядер)



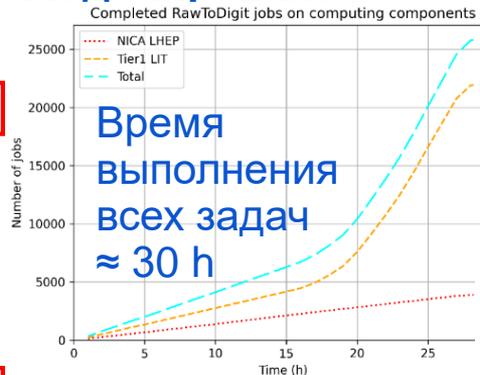
- Равномерное распределение ресурсов
- ≈ 100 задач / ч

Tier LIT (от 200 до 1500 ядер)

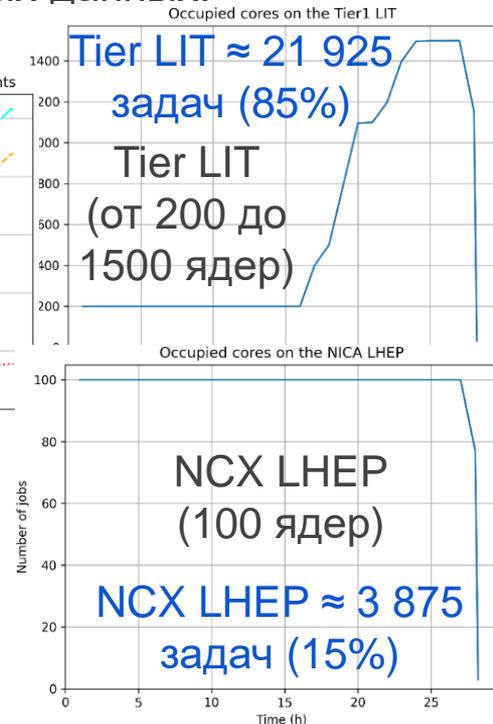


- Неравномерное использование ресурсов
- 200 – 1 500 задач / ч
- 81% задач от общего кол-ва

Моделирование



Сравнение по загрузке вычислительных ресурсов и времени выполнения задач.

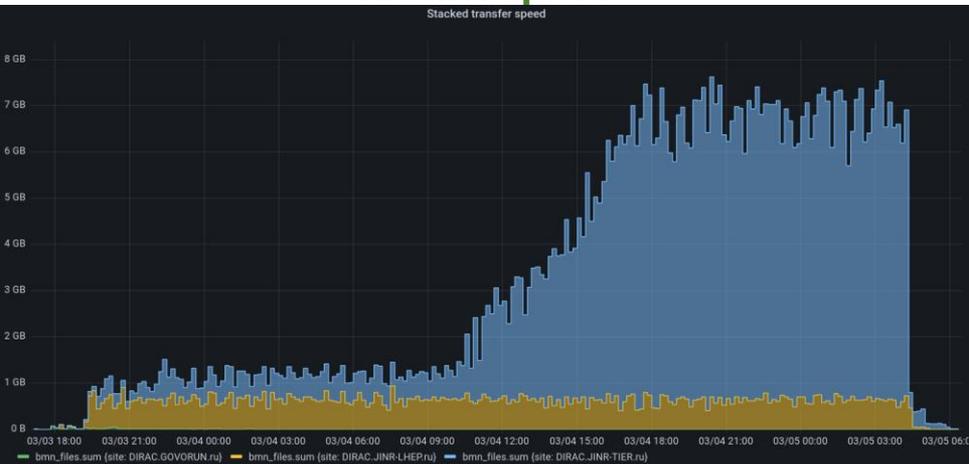


Глава 4. Верификация и опытная эксплуатация специального программного обеспечения для создания цифровых двойников

Верификация ядра ЦД

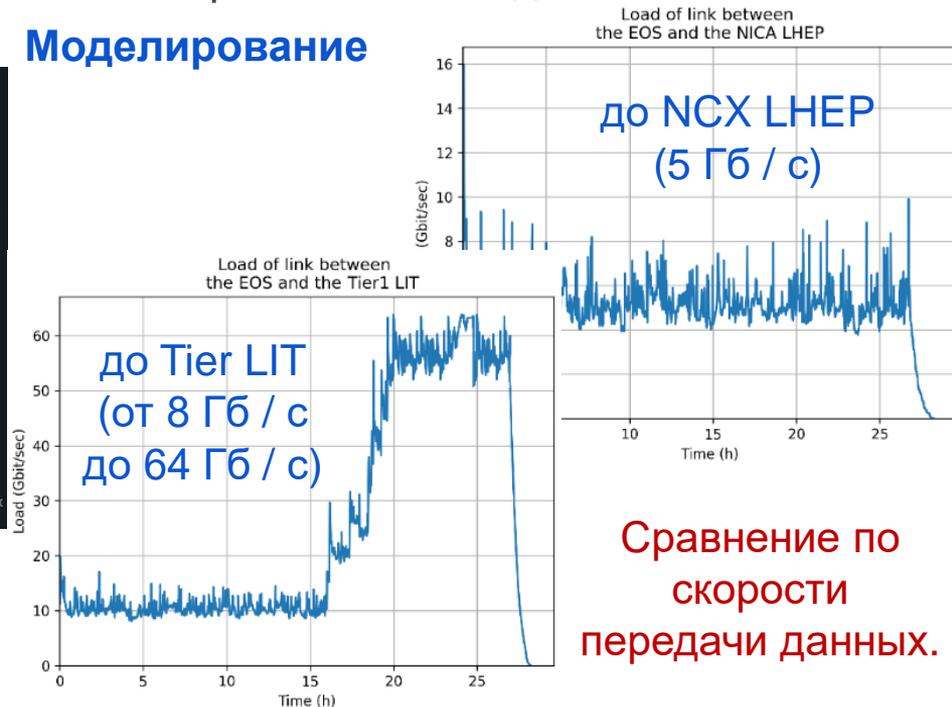
2. Выполнение задач преобразования экспериментальных данных.

Мониторинг



- Средняя скорость передачи данных до NCX LHEP ≈ 4 Гб / с
- Средняя скорость передачи данных до Tier LIT от 8 Гб / с до 64 Гб / с

Моделирование



Сравнение по скорости передачи данных.

Глава 4. Верификация и опытная эксплуатация специального программного обеспечения для создания цифровых двойников

Верификация ядра ЦД

	Мониторинг	Моделирование	Точность
Объем экспериментальных данных, накопленных за 720 часов работы установки (ТБ)	379	363	96%
Общее время выполнения всех RawToDigit задач (ч.)	36	30	83%
Количество выполненных задач на кластере NICA ЛФВЭ / % от общего числа задач	4 844 / 19%	3 875 / 15%	80%
Количество выполненных задач на Tier1 ЛИТ / % от общего числа задач	20 956 / 81%	21 924 / 85%	96%
Скорость передачи данных между EOS и кластером NICA ЛФВЭ (ГБ/с)	0.5	0.63	80%
Скорость передачи данных между EOS и Tier1 ЛИТ (ГБ/с)	от 1 до 8	от 1 до 8	100%

Результаты работы ЦД будут отличаться не более, чем на 20%

Глава 4. Верификация и опытная эксплуатация специального программного обеспечения для создания цифровых двойников

Верификация ядра ЦД

	Мониторинг	Моделирование	Точность
Суммарный объем digit-файлов (ТБ)	23	22	96%
Общее время выполнения всех DigitToDst задач (ч.)	73	80	91%
Количество выполненных задач на кластере NICA ЛФВЭ / % от общего числа задач	5 315 / 21%	5 906 / 24%	88%
Количество выполненных задач на Tier1 ЛИТ / % от общего числа задач	9 289 / 36%	8 872 / 34%	94%
Количество выполненных задач на Tier2 ЛИТ / % от общего числа задач	9 016 / 35%	8 598 / 33%	94%
Количество выполненных задач на Суперкомпьютере / % от общего числа задач	2 180 / 8%	2 424 / 9%	88%
Суммарный объем dst-файлов (ТБ)	53	52	98%

Результаты работы ЦД будут отличаться не более, чем на 20%

Глава 4. Верификация и опытная эксплуатация специального программного обеспечения для создания цифровых двойников

Результаты верификации ядра ЦД

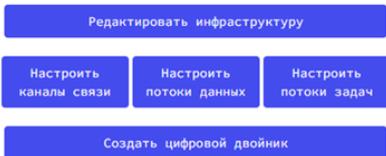
1. Верификация проводилась по следующим показателям:

- количество ресурсов, необходимых для хранения поступающих данных;
- загрузка вычислительных ресурсов при преобразовании экспериментальных данных и получении по ним реконструированных данных событий;
- общее время выполнения задач;
- скорость передачи данных между объектами инфраструктуры.

2. Доказана адекватность ЦД, результаты работы которого будут отличаться от результатов работы реальных РЦОД не более, чем на 20%.

Глава 4. Верификация и опытная эксплуатация специального программного обеспечения для создания цифровых двойников

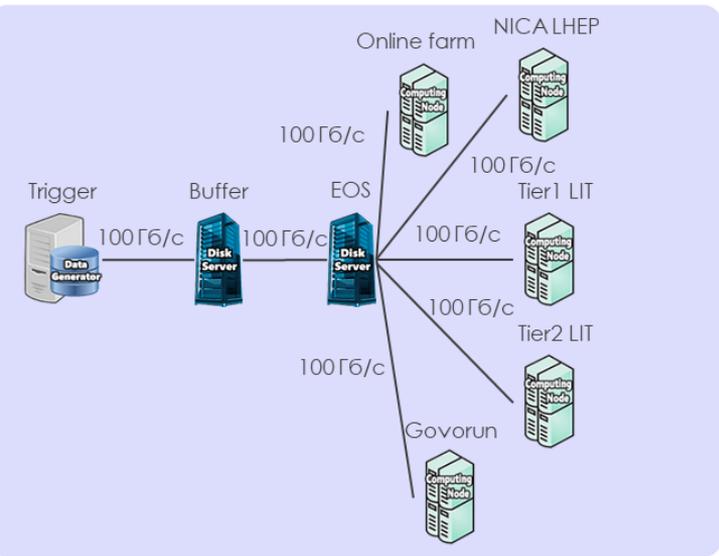
Построение инфраструктуры центра сбора, хранения и обработки данных



Применение СПО для построения ЦД вычислительной инфраструктуры эксперимента VM@N комплекса NICA

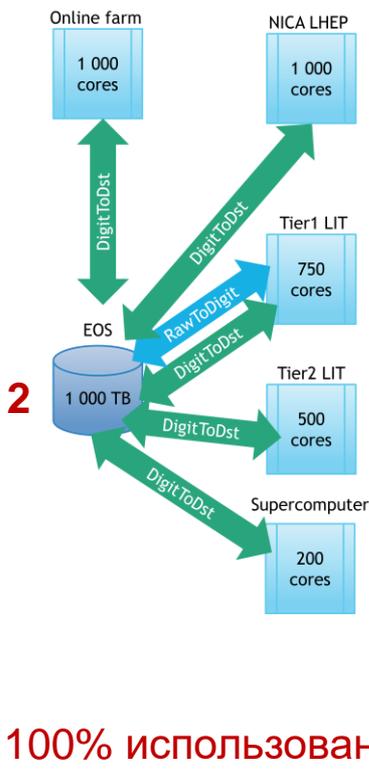
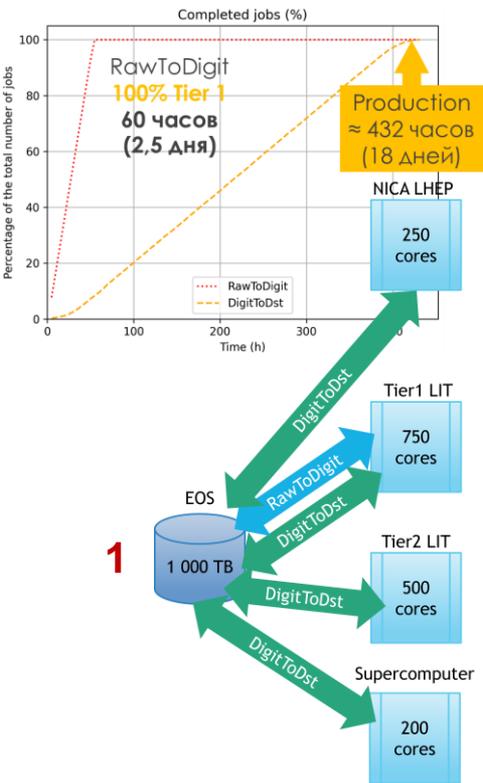
Цели:

- оценить потребности в ресурсах для хранения и обработки данных с учетом планируемых параметров потоков данных будущих сеансов эксперимента;
- получить примерное время, которое потребуется на обработку данных по окончании сеанса эксперимента.

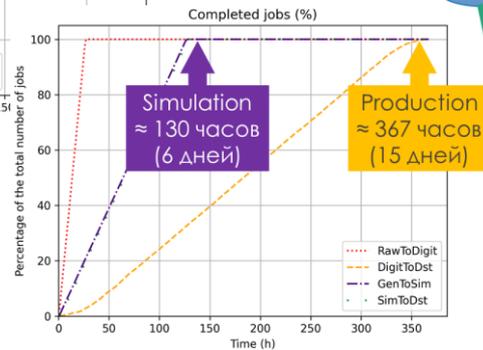
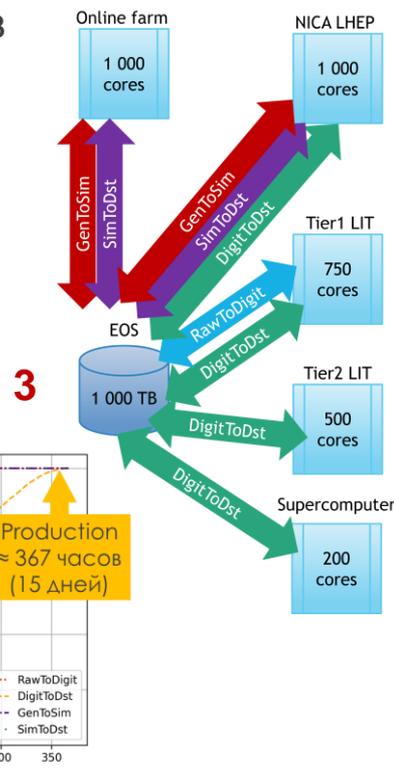
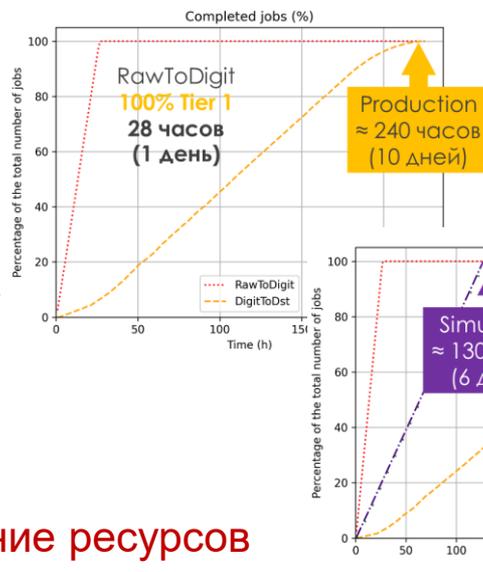


Глава 4. Верификация и опытная эксплуатация специального программного обеспечения для создания цифровых двойников

Сравнение различных конфигураций инфраструктуры для обработки данных



Увеличено количество ресурсов NICA LHEP до 1 000 ядер.
Добавлены ресурсы Online farm (1 000 ядер).



100% использование ресурсов

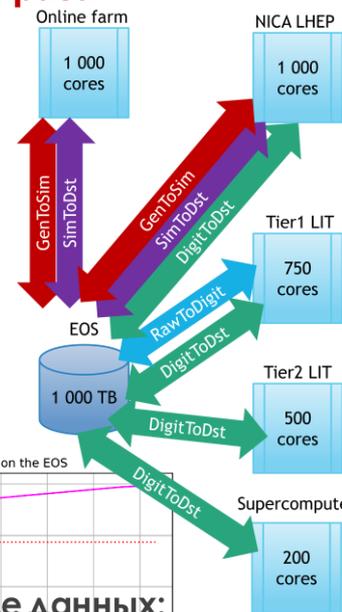
Глава 4. Верификация и опытная эксплуатация специального программного обеспечения для создания цифровых двойников

Результаты

Процесс преобразования raw-данных,

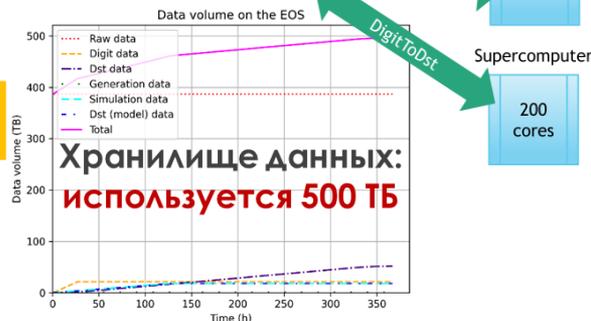
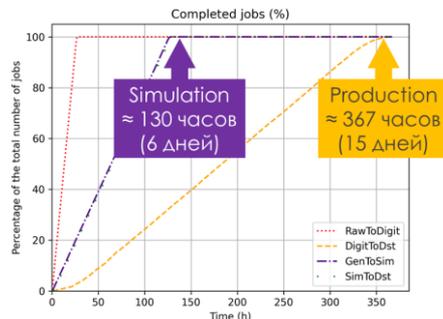
ускорен в 2 раза

	1	2	3
Преобразование всех raw-данных в формат digit (Raw-ToDigit задачи)	60 часов (2,5 дня)	28 часов (1 день)	28 часов (1 день)
Время полной обработки всех экспериментальных данных	432 часов (18 дней)	240 часов (10 дней)	367 часов (15 дней)
Время полной обработки всех модельных данных	---	---	130 часов (6 дней)



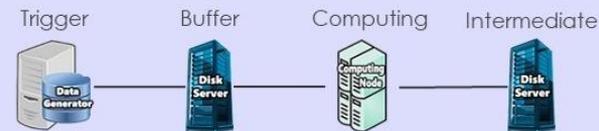
1. Найдена конфигурация для обработки экспериментальных и смоделированных данных за наименьшее количество времени с использованием всех имеющихся ресурсов.

2. Общее время работы ЦД для всех вычислительных экспериментов - 20 ч.



Глава 4. Верификация и опытная эксплуатация специального программного обеспечения для создания цифровых двойников

Построение инфраструктуры центра сбора, хранения и обработки данных



Применение СПО для построения ЦД вычислительной системы онлайн-фильтра данных эксперимента SPD комплекса NICA

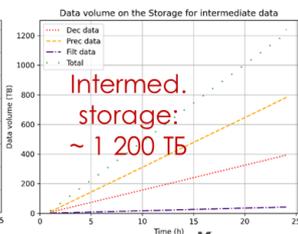
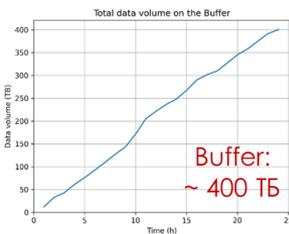
Цели:

- спроектировать эффективную вычислительную систему для обеспечения быстрой реконструкции событий и фильтрации данных в режиме реального времени;
- предусмотреть возможности для дальнейшего развития и оптимизации проектируемой системы.

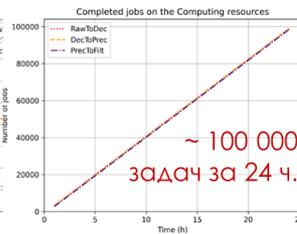
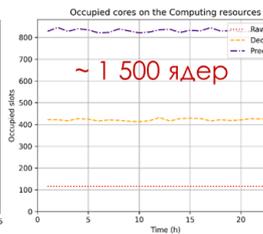
Глава 4. Верификация и опытная эксплуатация специального программного обеспечения для создания цифровых двойников

Эффективность генерации данных – 20%

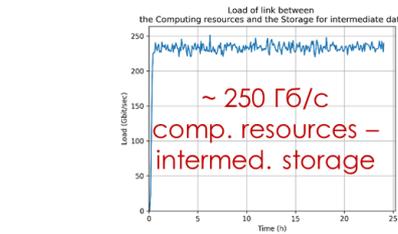
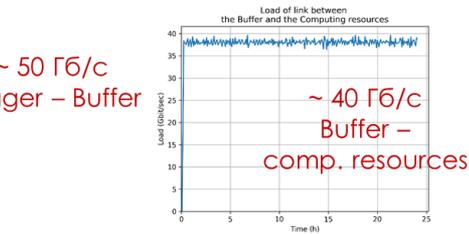
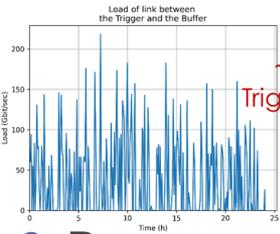
Хранилища данных



Вычислительные ресурсы



Каналы связи



Результаты применения СПО для построения ЦД вычислительной системы онлайн-фильтра данных эксперимента SPD комплекса NICA

1. Даны оценки по требуемому оборудованию для решения функциональных задач онлайн-фильтра в режиме реального времени с учетом планируемых характеристик потоков данных эксперимента.

2. Результаты доказали возможность использования СПО для решения задачи проектирования РЦОД.

3. Время работы ЦД составило 40 минут.

4. Дальнейшее применение ЦД обеспечит качественную проверку эффективности и надежности функционирования будущей системы.

Глава 4. Верификация и опытная эксплуатация специального программного обеспечения для создания цифровых двойников

Выводы

- Результаты верификации доказали корректную работу программы моделирования: **результаты работы ЦД будут отличаться** от результатов работы реальных РЦОД **не более, чем на 20%**.
- Подтверждено требование к скорости работы ЦД: **время получения результатов не превышает 24 часов**.
- Проведена опытная эксплуатация СПО на примерах построения ЦД вычислительной инфраструктуры эксперимента VM@N и ЦД системы онлайн-фильтра данных эксперимента SPD комплекса NICA.
- По результатам работы ЦД были найдены характеристики оборудования, необходимого для хранения, обработки и передачи данных с учетом планируемых параметров потоков данных будущих сеансов экспериментов, что позволило принять правильное решение в процессе управления потоками задач и более корректно распределить нагрузку на вычислительные ресурсы.

Защищаемые положения

- 1.** Разработан метод построения цифрового двойника, который позволяет описывать распределенные центры сбора, хранения и обработки данных, учитывая потоки данных и задач, а также процессы, происходящие в РЦОД.
- 2.** Разработаны алгоритмы, на основе которых создано специальное программное обеспечение, используемое для принятия решений по выбору конфигурации оборудования распределенных центров сбора, хранения и обработки данных согласно заданным требованиям.
- 3.** Доказана адекватность построенных методов и алгоритмов на примере вычислительной инфраструктуры существующего эксперимента.

Научная новизна полученных результатов

- 1.** Предложен и разработан новый метод построения и использования цифровых двойников РЦОД, отличительной особенностью которого является возможность моделирования процессов обработки и хранения данных, учитывая характеристики потоков данных и задач, вероятности сбоев, отказов и изменений в производительности оборудования и других процессов, происходящих в моделируемой системе.
- 2.** Разработаны и реализованы алгоритмы описания инфраструктуры распределенной системы и формирования ее виртуального образа.
- 3.** Разработана проблемно-ориентированная система принятия решений для задач управления и оптимизации с целью улучшения технических характеристик РЦОД на основе моделей ЦД, адекватность которых подтверждена на примере вычислительной инфраструктуры эксперимента VM@N комплекса NICA.
- 4.** Научно обоснована конфигурация РЦОД для экспериментов VM@N и SPD комплекса NICA.

Теоретическая значимость

1. Развитие метода моделирования распределенных вычислительных систем, которые оперируют очень большими объемами данных, требующих надежного хранения и сложной системы обработки.
2. Доказана применимость разработанной в диссертации методологии для создания цифровых двойников РЦОД на основе результатов верификации модели вычислительной инфраструктуры существующего эксперимента.

Практическая значимость

1. Применение результатов для повышения эффективности функционирования, качества и надежности сложных систем сбора, хранения и обработки данных.
2. Разработанное СПО может применяться для широкого класса задач в области проектирования, построения и развития РЦОД.
3. Получено свидетельство о государственной регистрации программы для ЭВМ.
4. Решена задача поиска конфигурации оборудования для системы сбора, хранения и обработки данных эксперимента VM@N комплекса NICA в ОИЯИ.
5. Результаты работы применяются при проектировании вычислительной системы онлайн-фильтра данных эксперимента SPD комплекса NICA в ОИЯИ.
6. Результаты работы используются в учебном процессе ФГБОУ ВО «Университет «Дубна» в курсе «Распределенные вычисления и облачные технологии» для подготовки магистров по направлению 27.04.03 Системный анализ и управление по профилю «Цифровые платформы и аналитика больших данных».

Апробация результатов

1. Ежегодные совещания коллаборации эксперимента BM@N комплекса NICA (6).
2. Совещание коллаборации эксперимента SPD комплекса NICA (1).
3. Российские и международные научные мероприятия (13), в том числе:
 - The International Conference “Distributed Computing and Grid-technologies in Science and Education” (GRID 2018, 2021, 2023)
 - The International Conference “Mathematical Modeling and Computational Physics” (MMCP 2019)
 - The Conference (with international participation) “Information and Telecommunication Technologies and Mathematical Modeling of High-Tech Systems” (ITTMM 2023)
 - Международная конференция «Математика. Компьютер. Образование» (МКО 2019)
 - The International Scientific Conference of Young Scientists and Specialists (AYSS 2020, 2022, 2023)
 - JINR PACs for Particle Physics (2022, 2023, 2024)
4. Гранты для молодых научных сотрудников ОИЯИ (2); стипендии им. М.Г. Мещерякова (2) и им. Н.Н. Говоруна (2) в ЛИТ ОИЯИ.
5. Лауреат II степени Премии ОИЯИ для молодых ученых и специалистов за 2023 г. в номинации «Научно-технические прикладные работы».

Публикации (основные)

1. Пряхина Д.И., Кореньков В.В. Актуальность создания цифрового двойника для управления распределенными центрами сбора, хранения и обработки данных // Современные информационные технологии и ИТ-образование. 2023. Т. 19. № 2. С. 262–271.
2. Пряхина Д.И., Кореньков В.В., Трофимов В.В. Метод построения цифровых двойников для решения задач эффективного управления и развития распределенных центров сбора, хранения и обработки данных // Современные информационные технологии и ИТ-образование. 2023. Т. 19. № 2. С. 272–281.
3. Пряхина Д.И., Кореньков В.В., Трофимов В.В., Герценбергер К.В. Верификация программы моделирования для создания цифровых двойников распределенных центров сбора, хранения и обработки данных // International Journal of Open Information Technologies. Январь 2024. Т. 12. № 1. С. 118–128.
4. Свидетельство о государственной регистрации программы для ЭВМ №2023667305 «Программный комплекс для создания цифровых двойников распределенных центров сбора, хранения и обработки данных» от 14 августа 2023 г.

Публикации (дополнительные)

5. Трофимов В.В., Нечаевский А.В., Ососков Г.А., Пряхина Д.И. Вероятностно-стоимостной подход к оптимизации распределенных систем хранения данных физических экспериментов // CEUR Workshop Proceedings. 2018. Т. 226. С. 393–399.
6. Кореньков В.В., Пряхина Д.И., Нечаевский А.В., Ососков Г.А., Трофимов В.В. Моделирование центров хранения и обработки данных с учетом экономических составляющих // Системный анализ в науке и образовании. 2018. № 4.
7. Korenkov V., Nechaevskiy A., Ososkov G., Priakhina D., Trofimov V. A probabilistic approach of the simulation of data processing centers // European Physical Journal Web of Conferences. January 2020. Vol. 226. P. 03012.
8. Priakhina D., Trofimov V., Ososkov G., Gertsenberger K. Data center simulation for the BM@N experiment of the NICA project // AIP Conference Proceeding. 2021. Vol. 2377. P. 040007.
9. Priakhina D., Korenkov V., Gertsenberger K., Trofimov V. Simulation of Data Processing for the BM@N Experiment of the NICA Complex // CEUR Workshop Proceedings. 2021. Vol. 3041. pp. 483–487.
10. Priakhina D., Korenkov V., Trofimov V., Gertsenberger K. Simulation Results of BM@N Computing Infrastructure // Physics of Particles and Nuclei Letters. 2023. Vol. 20. No. 5. pp. 1272–1275.

СПАСИБО ЗА ВНИМАНИЕ!

Пряхина Дарья Игоревна

Научная специальность 2.3.1.
Системный анализ, управление и
обработка информации, статистика

Научный руководитель:
доктор технических наук
Кореньков
Владимир Васильевич