

Отчёт за 2024 год по стипендии
им. Михаила Григорьевича Мещерякова
для молодых учёных ЛИТ



Пелеванюк Игорь
научный сотрудник

11.12.2024

Объявленные задачи

1. Подготовить диссертацию.
2. Оптимизация работы пилотов DIRAC.
3. Разработка и реализация метода оценки целостности файлов под управлением DIRAC для пользователей ОИЯИ.
4. Включение ленточного робота на базе СТА в инфраструктуру DIRAC.
5. Разработка методики оценки производительности ленточных хранилищ при работе в рамках инфраструктуры DIRAC.
6. Разработка системы мониторинга NICA кластера.

Объявленные задачи

1. Подготовить диссертацию.
2. Оптимизация работы пилотов DIRAC.
3. Разработка и реализация метода оценки целостности файлов под управлением DIRAC для пользователей ОИЯИ.
4. Включение ленточного робота на базе СТА в инфраструктуру DIRAC.
5. Разработка методики оценки производительности ленточных хранилищ при работе в рамках инфраструктуры DIRAC.
6. Разработка системы мониторинга NICA кластера.

Результаты

1. Подготовить диссертацию.

Автореферат написан и прошёл первый раунд исправлений.

Текст диссертации написан на 40%.

Подготовлена презентация представления работы для семинара по материалам кандидатской диссертации.

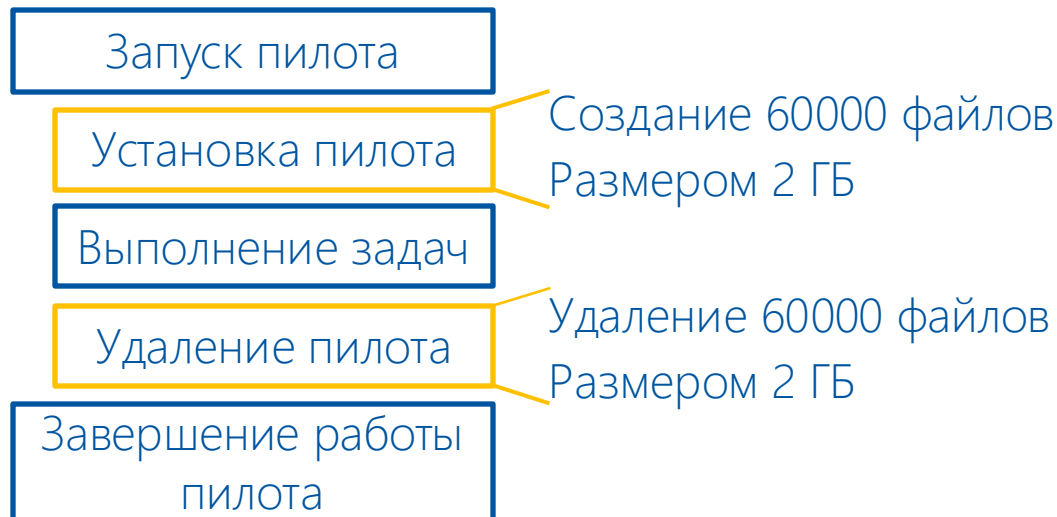
Научный семинар планируется 24 декабря 2024 года.

Результаты

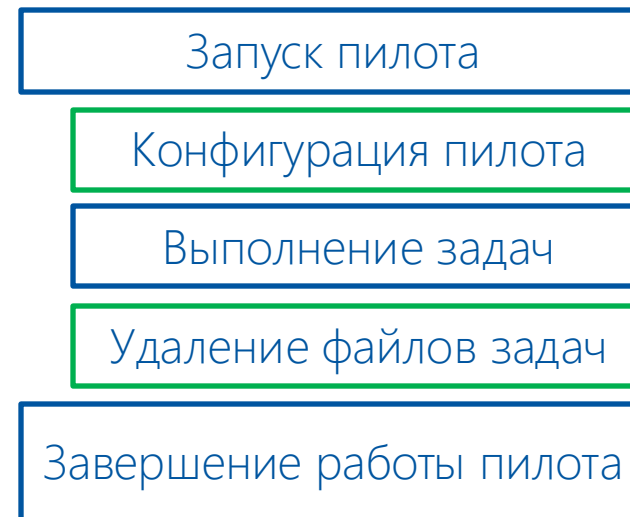
2. Оптимизация схемы работы пилотов DIRAC в ОИЯИ

Речь идет об использовании предустановленной в CVMFS версии пилота.

Было:



Стало:



Потребовалось создать специальный алгоритм конфигурации пилота для работы в рамках системы DIRAC в ОИЯИ.

Результаты

3. Разработка и реализация метода оценки целостности файлов под управлением DIRAC

Шаг 1: Создание в рамках файлового каталога DIRAC нового поля метаданных `status`

Шаг 2: Присвоение всем файлам значение `status=unchecked`

Шаг 3: Массовый запуск в грид специально созданной задачи `FileCheck`:

Данная задача поочерёдно скачивает файлы со статусом `unchecked`.

Если контрольная сумма скачанного файла верна: `status=ok`

Если контрольная сумма скачанного файла не верна: `status=bad_checksum`

Шаг 4: Анализ файлов со статусами `ok` и `bad_checksum`

VM@N: повреждено 90% всех файлов

Восстановлены из резервной копии.

MPD: повреждено 15% всех файлов.

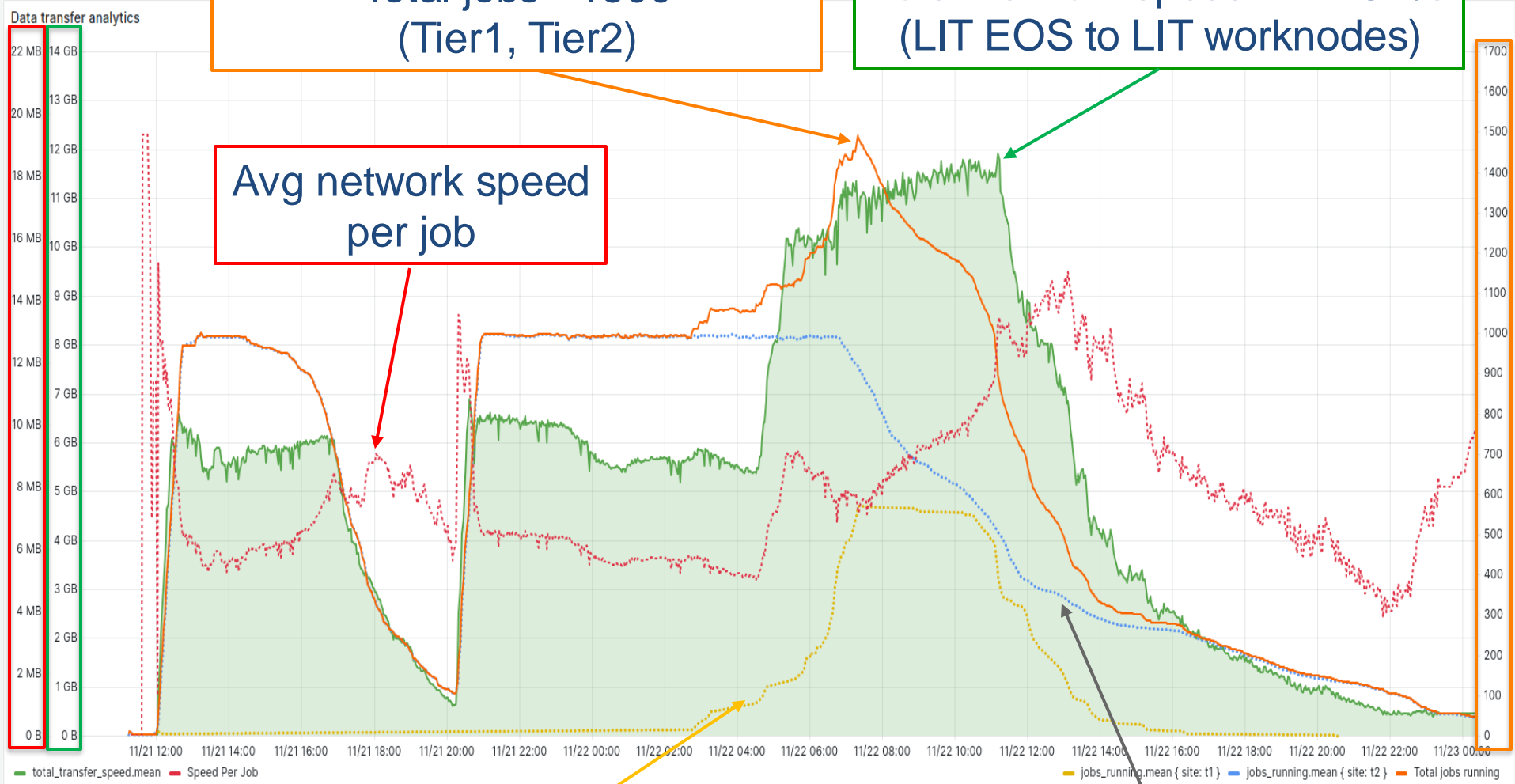
Часть восстановлены из резервной копии. Часть регенерированны.

Результаты

Total jobs - 1500
(Tier1, Tier2)

Total network speed – 12 GB/s
(LIT EOS to LIT worknodes)

Avg network speed
per job



Tier1 jobs

Tier2 jobs

Результаты

4. Включение ленточного робота на базе СТА в инфраструктуру DIRAC.

Ленточный робот был успешно интегрирован в систему DIRAC в ОИЯИ.

На ленточный робот были сохранены данные 8го сеанса эксперимента VM@N. (~430 TB)

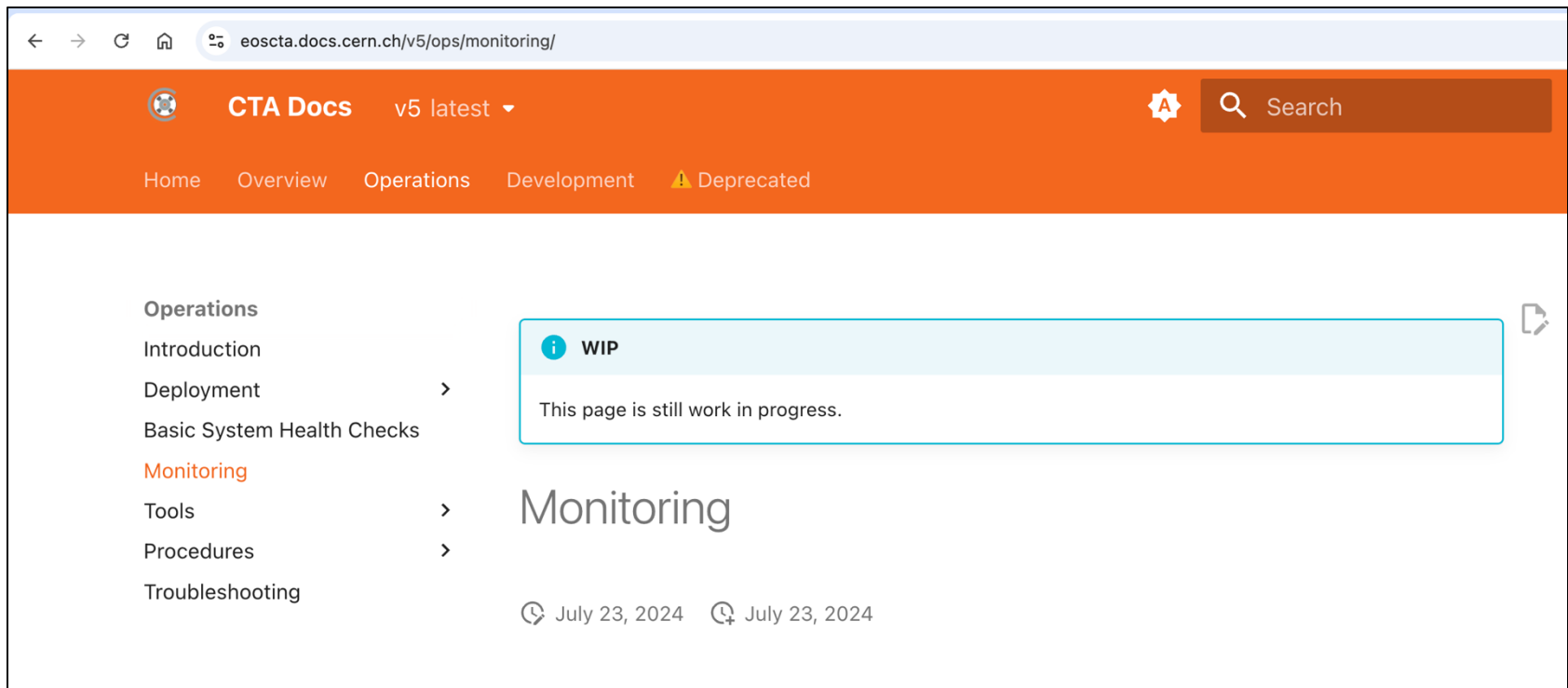
Была сделана резервная копия части данных эксперимента MPD. (~11 TB)

Консультирование эксперимента SRC.

Работа сделана совместно с Владимиром Трофимовым и Валерием Мицыным.

Результаты

5. Разработка методики оценки производительности ленточных хранилищ при работе в рамках инфраструктуры DIRAC.



Официальный сайт с документацией к СТА

Результаты

5. Разработка методики оценки производительности ленточных хранилищ при работе в рамках инфраструктуры DIRAC.

Для решения этой задачи потребовалось изучить существующие инструменты СТА и создать собственные методы оценки производительности

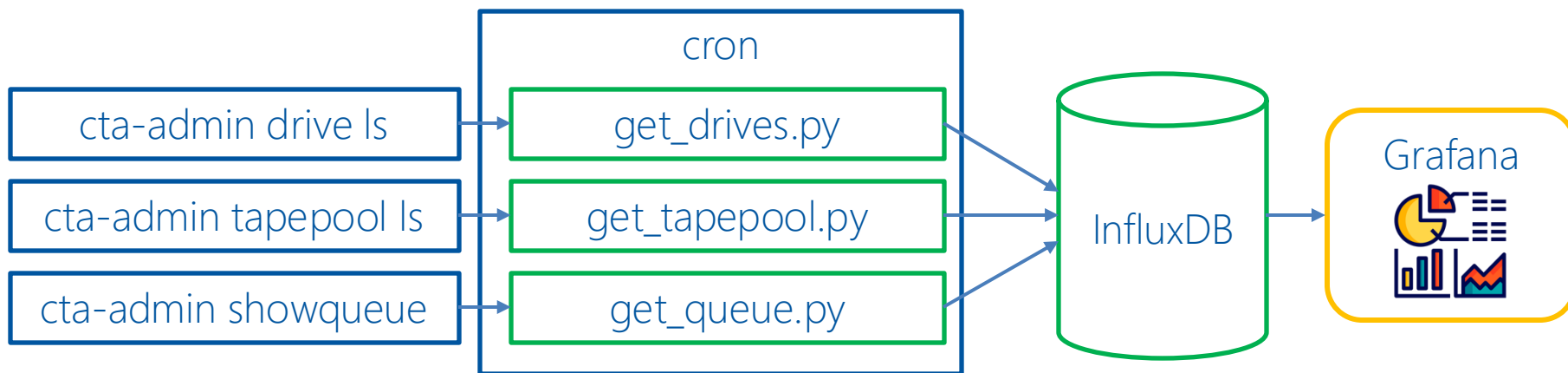
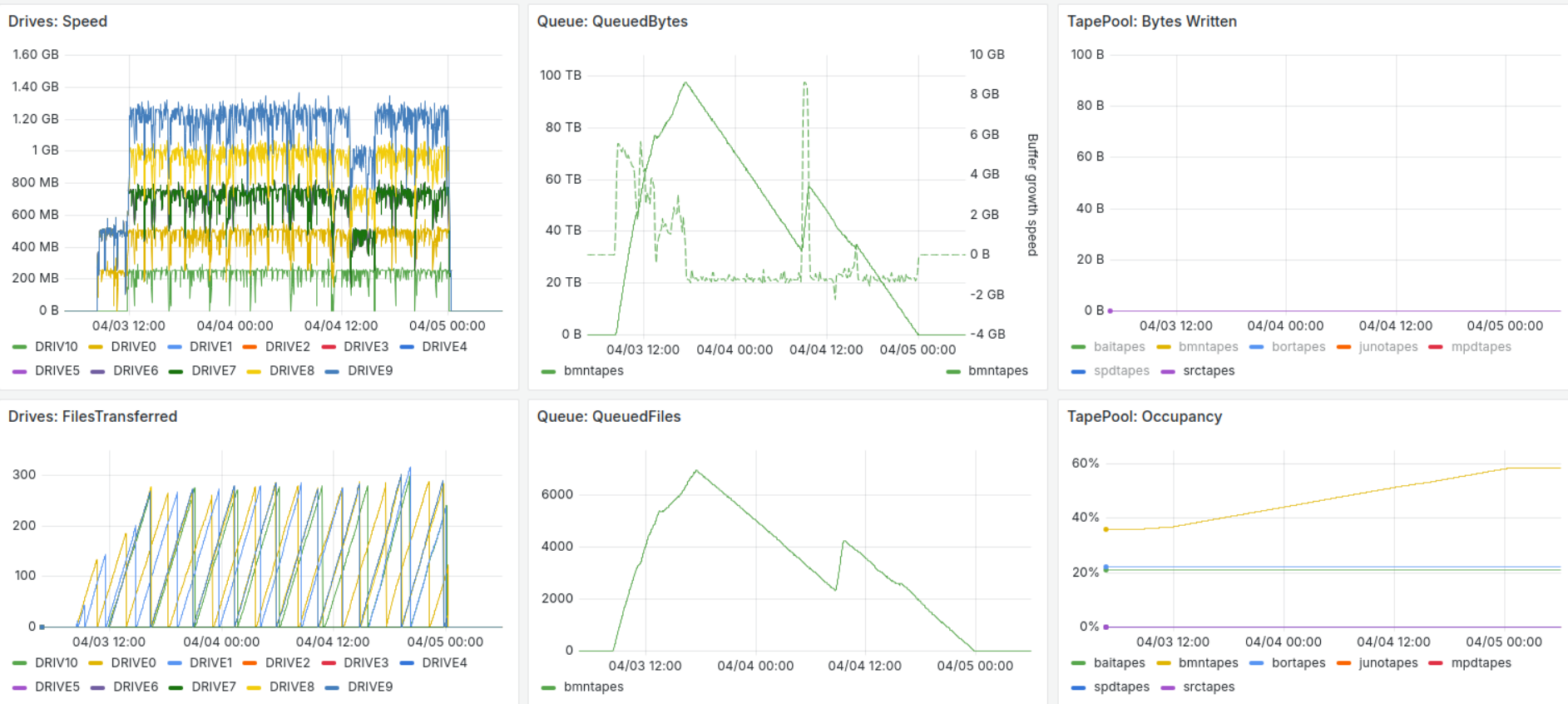


Схема получения данных о работе и состоянии ленточного работа СТА.

Работа сделана совместно с Владимиром Трофимовым и Валерием Мицыным.

Результаты

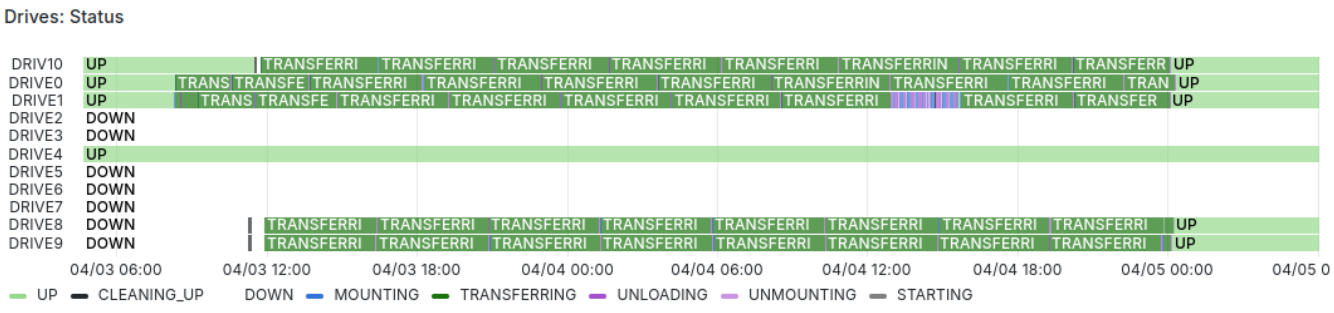
5. Разработка методики оценки производительности ленточных хранилищ при работе в рамках инфраструктуры DIRAC.



Результаты визуализации показателей очереди и драйвов

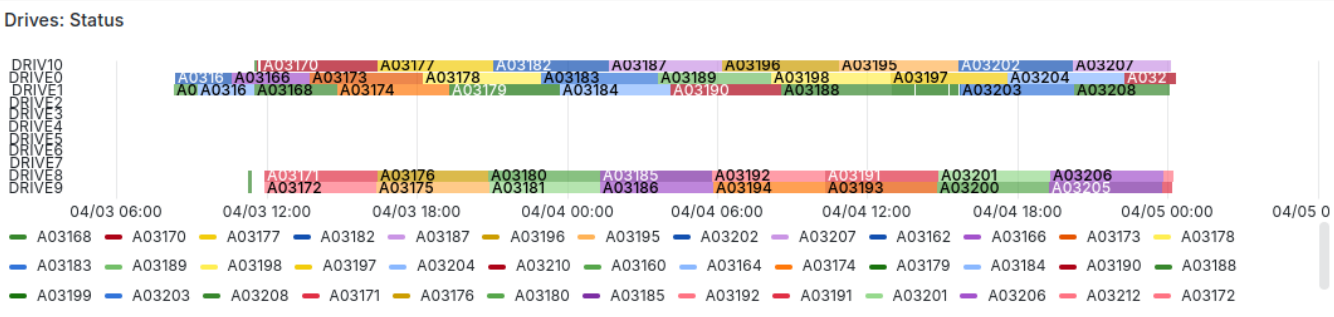
Результаты

5. Разработка методики оценки производительности ленточных хранилищ при работе в рамках инфраструктуры DIRAC.



Drives: Current Status

driveName	logicalLibrary	status
DRIVE10	common	UP
DRIVE0	common	UP
DRIVE1	common	UP
DRIVE2	juno	DOWN
DRIVE3	juno	DOWN
DRIVE4	neutrino	UP
DRIVE5	neutrino	DOWN
DRIVE6	neutrino	DOWN
DRIVE7	neutrino	DOWN
DRIVE8	common	UP
DRIVE9	common	UP

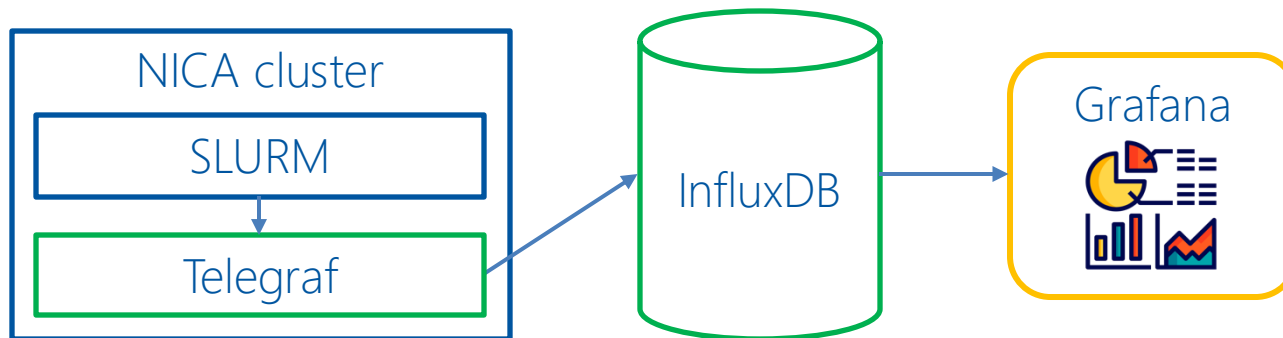


Результаты визуализации состояния лент и драйвов

Результаты

6. Разработка системы мониторинга NICA кластера.

Основная задача: лучше понимать состояние кластера, его загрузку, количество задач в очереди, состояние очереди.



Был найден готовый модуль Telegraf для сбора данных из SLURM.

Созданы закрытый дашборд для администраторов и публичный дашборд для пользователей.

Работа сделана совместно с Иваном Слеповым.

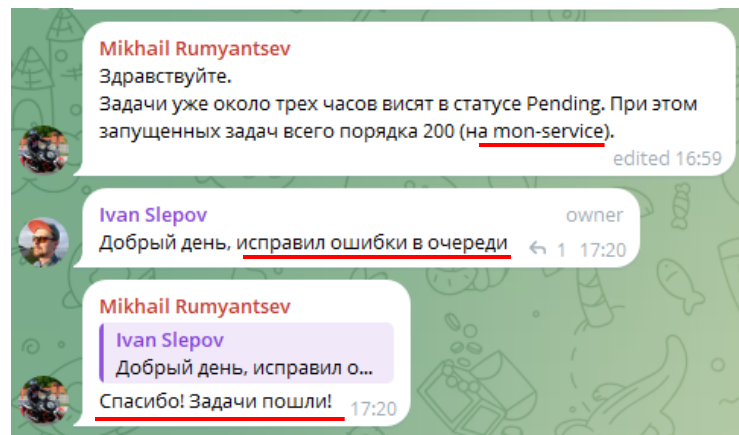
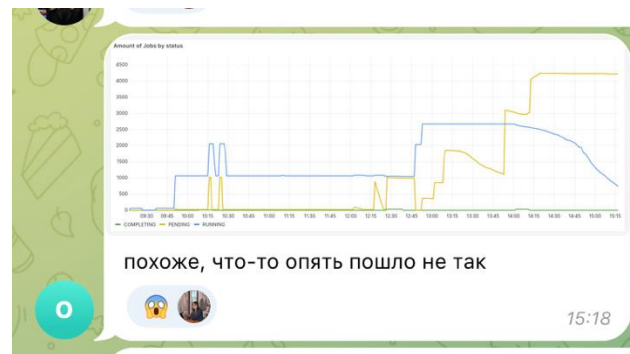
Результаты

6. Разработка системы мониторинга NICA кластера.



Созданный и открытый для пользователей мониторинг показывает количество задач в очереди и количество работающих задач. Помимо этого показывается занятое место на разных хранилищах NICA кластера.

Сообщения в чате пользователей NCX в Телеграм (06.03.2024).



Планы на 2025 год

из заявки на грант молодых ученых и специалистов ОИЯИ

Защита кандидатской диссертации.

Интеграция системы FTS в платформу DIRAC для автоматизации передач данных между дисковыми и ленточными хранилищами.

Разработка и реализация методов анализа информации о данных, находящихся под управлением платформы DIRAC:

- Разработка и реализация методов сбора информации о данных, находящихся под управлением платформы DIRAC, как из системы управления данными DIRAC, так и напрямую с систем хранения данных.
- Разработка подходов сравнения собранной информации.
- Разработка методов визуализации позволяющих производить анализ собранной информации.

Публикации

1. Gertsenberger, K.V., Pelevanyuk, I.S.
BM@N Run 8 Data Processing on a Distributed Infrastructure with DIRAC.
Phys. Part. Nuclei Lett. 21, 778–781 (2024). <https://doi.org/10.1134/S1547477124701334>
2. Gertsenberger, K., Pelevanyuk, I., Klimai, P. & A. Chebotov
Computing Software Architecture for the BM@N Experiment.
Phys. Part. Nuclei 55, 338–342 (2024). <https://doi.org/10.1134/S1063779624030407>
3. Campis, D., Ilina, A. & Pelevanyuk, I.
System for Analysis of the Performance of Scientific Jobs in Distributed Systems.
Phys. Part. Nuclei 55, 401–403 (2024). <https://doi.org/10.1134/S1063779624030262>
4. Balashov, N., Kuprikov, I., Kutovskiy, N., Pelevanyuk, I., et al.
Changes and Challenges at the JINR and Its Member States Cloud Infrastructures.
Phys. Part. Nuclei 55, 366–370 (2024). <https://doi.org/10.1134/S1063779624030092>
5. Baginyan, A., Balandin, A., Dolbilov, A., Pelevanyuk, I., et al.
JINR Grid Infrastructure: Status and Plans.
Phys. Part. Nuclei 55, 355–359 (2024). <https://doi.org/10.1134/S1063779624030079>
6. The BM@N collaboration
The BM@N spectrometer at the NICA accelerator complex,
Nuclear Instruments and Methods in Physics Research Section A, V. 1065, 2024, 169532
<https://doi.org/10.1016/j.nima.2024.169532>

Выступления

1. ПКК по частицам (постер) – BM@N Run 8 data processing on a distributed infrastructure with DIRAC
2. Суперкомпьютерные дни в России – (соавтор) Распределённая параллельная файловая система Lustre для обработки и анализа данных экспериментов физики высоких энергий
3. AYSS Conference 2024 – BM@N Run 8 data processing on a distributed infrastructure with DIRAC

4. Научная школа Липня – Полезные веб-сервисы и как они могут сделать нашу работу эффективнее

5. Рабочее совещание МИФИ-ОИЯИ Компьютинг для мегапроекта NICA – Распределённая обработка и генерация данных экспериментов MPD и BM@N
6. BM@N collaboration meeting 12th – BM@N distributed computing status and analytics
7. JUNO weekly meeting – Analysis of DIRAC Jobs' performance
8. China-JINR Workshop on Software and Computing for Future HEP Experiments – DIRAC for NICA experiments
9. BM@N collaboration meeting 13th – BM@N distributed data processing using DIRAC
10. MPD collaboration meeting 14th – MLIT resources and services for the MPD experiment
11. Семинар ЛИТ – Распределённая гетерогенная вычислительная среда для обработки данных NICA

Другое



Вторая премия за постер на ПКК по частицам в 2024 году с темой



Член орг комитета AYSS-2024, Семинары в доме ученых



Работа на стенде ОИЯИ: Наука 0+, Весенний день открытых дверей МФТИ 2024, Ярмарка вакансий ОЭЗ Дубна



Научно-популярные лекции: Наука 0+, Весенний день открытых дверей МФТИ 2024, Школа учителей ОИЯИ (2 раза), Летняя школа "Физика. Математика. Информатика"



Преподавание в Университете Дубна курса «Распределённые системы и облачные технологии»



Экскурсии по МИВК: 15 экскурсий (с сайта JINRex.jinr.ru)

Спасибо за внимание!