На конкурс научных работ ОИЯИ за 2024 год

ПРЕДСТАВЛЕНИЕ

Представляется работа:

“Разработка комплекса программных систем для реализации единой архитектуры распределенной обработки и хранения данных эксперимента BM@N/NICA”

Раздел: **Научно-методические и научно-технические работы**

Коллектив соавторов:

1. Александров Е. И.
2. Александров И. Н.
3. Балашов Н. А.
4. Герценбергер К. В.
5. Мошкин А. А.
6. Пелеванюк И. С.
7. Филозова И. А.
8. Чеботов А. И.
9. Шестакова Г. В.
10. Климай П. А.

В представленный цикл работ входят 25 публикаций:

1. K. Gertsenberger, I. Pelevanyuk, P. Klimai, and A. Chebotov, “*Computing Software Architecture for the BM@N Experiment*”, Phys. Part. Nuclei **55**, 338–342 (2024)
2. K. Gertsenberger, I. Pelevanyuk, “*BM@N Run 8 Data Processing on a Distributed Infrastructure with DIRAC*”, Phys. Part. Nucl. Lett. **21**. 778–781 (2024)
3. N. Balashov, “*JINR Container Distribution Service*”, Phys. Part. Nuclei **55**, 482–484 (2024)
4. K. Gertsenberger, P. Klimai, O. Nemova, “*Development of Monitoring Service for BM@N Information Systems*”, Phys. Part. Nucl. Lett. **21**, 793–796 (2024)
5. E. Alexandrov, I. Alexandrov, A. Chebotov, K. Gertsenberger, I. Filozova, D. Priakhina, G. Shestakova, and A. Yakovlev, “*Development of the Online Configuration System for the BM@N Experiment* “, Phys. Part. Nuclei **55**, 433–436 (2024)
6. E. Alexandrov, I. Alexandrov, A. Chebotov, A. Degtyarev, I. Filozova, K. Gertsenberger, P. Klimai, and A. Yakovlev, “*Implementation of the Event Metadata System for physics analysis in the NICA experiments*”, Journal of Physics: Conference Series **2438**, 012046 (2023)
7. E. Alexandrov, I. Alexandrov, A. Chebotov, K. Gertsenberger, I. Filozova., D. Priakhina, and G. Shestakova, “*Configuration Information System for online processing and data monitoring in the NICA experiments*”, Journal of Physics: Conference Series **2438**, 012019 (2023)

8. A. Chebotov, A. Degtyarev, K. Gertsenberger, and P. Klimai, "*REST API and Web Interface for the Event Metadata System of the BM@N Experiment*", Phys. Part. Nucl. Lett. **20**, 1527–1530 (2023)
9. A. Chebotov, K. Gertsenberger, A. Moshkin, and I. Slepov, "*Common Deployment Complex for the Information Systems of the BM@N Experiment*", Phys. Part. Nucl. Lett. **20**, 1269–1271 (2023)
10. K. Gertsenberger, P. Klimai, M. Zelenyi, "*Auxiliary Services for the Condition Database of the BM@N Experiment at NICA*", Phys. Part. Nucl. Lett. **20**, 1217–1219 (2023)
11. E. Alexandrov, I. Alexandrov, A. Chebotov, K. Gertsenberger, I. Filozova, D. Priakhina, and G. Shestakova, "*Status of the Configuration Information System for the NICA experiments* ", Phys. Part. Nucl. Lett. **19**, 543–546 (2022)
12. A. Degtyarev, K. Gertsenberger, P. Klimai, "*Usage of Apache Cassandra for Prototyping the Event Metadata System of the NICA Experiments*", Phys. Part. Nucl. Lett. **19**, 562–565 (2022)
13. A. Chebotov, K. Gertsenberger, P. Klimai, and A. Moshkin, "*Information System Based on the Condition Database for the NICA Experiments, User WEB Application, and Related Services*", Phys. Part. Nucl. Lett. **19**, 558–561 (2022)
14. K. Gertsenberger, I. Alexandrov, I. Filozova, E. Alexandrov, A. Moshkin, A. Chebotov, M. Mineev, D. Pryahina, G. Shestakova, A. Yakovlev, A. Nozik, and P. Klimai, "*Development of Information Systems for Online and Offline Data Processing in the NICA Experiments*", Phys. Part. Nuclei **52**, 801-807 (2021)
15. A. Chebotov, K. Gertsenberger, I. Slepov, and A. Moshkin, "*Electronic Logbook platform for NICA experiments*", AIP Conf. Proc. **2377**, 040003 (2021)
16. E. Akishina, E. Alexandrov, I. Alexandrov, I. Filozova, K. Gertsenberger, and V. Ivanov, "*Development of a Geometry Database and Related Services for the NICA experiments*", Phys. Part. Nuclei **52**, 842-846 (2021)
17. E. Alexandrov, I. Alexandrov, A. Degtyarev, K. Gertsenberger, I. Filozova, P. Klimai, A. Nozik, and A. Yakovlev, "*Design of the Event Metadata System for the Experiments at NICA*", Phys. Part. Nucl. Lett. **18**, 603-616 (2021)
18. K. Gertsenberger, A. Chebotov, P. Klimai, I. Alexandrov, E. Alexandrov, I. Filozova, and A. Moshkin, "*Implementation of the Condition Database for the Experiments of the NICA Complex*", CEUR Workshop Proceedings **3041**, 128–132 (2021)
19. E. Akishina, E. Alexandrov, I. Alexandrov, I. Filozova, K. Gertsenberger, V. Ivanov, D. Priakhina, and G. Shestakova, "*Development of the Geometry Database for the BM@N Experiment of the NICA Project*", EPJ Web Conf. **226**, 03001 (2020)
20. K. Gertsenberger, A. Chebotov, I. Alexandrov, I. Filozova., and E. Alexandrov, "*Design of the Condition Database for online and offline data processing in experimental setups of the NICA complex*" (in russian), Izvestiya SFedU. Engineering Sciences **217**, no.7, 172-180 (2020)
21. A. Chebotov, K. Gertsenberger, "*Development of web-service for Unified Database of the BM@N experiment at NICA*", AIP Conf. Proc. **2163**, 040002 (2019)
22. K. Gertsenberger, A. Moshkin, A. Chebotov, "*Development of the Electronic Logbook for the BM@N Experiment at NICA*", CEUR Workshop Proceedings **2507**, 175–179 (2019)
23. E. Alexandrov, I. Alexandrov, K. Gertsenberger, M. Mineev, A. Moshkin, D. Pryahina, I. Filozova, A. Chebotov, G. Shestakova, and A. Yakovlev, "*Information Systems for Online and Offline Data Processing in Modern High-energy Physics Experiments*" (in

russian), International scientific journal «Modern Information Technologies and IT-Education» **15**, no.3, 654-671 (2019)
24. K. Gertsenberger, O. Rogachevsky, "*The Unified Database for BM@N experiment data handling*", EPJ Web Conf. **177**, 05001 (2018)
25. E. Akishina, E. Alexandrov, I. Alexandrov, I. Filozova, V. Friese, K. Gertsenberger, V. Ivanov, and O. Rogachevsky, "*Geometry Database for the CBM experiment and its first application to the experiments of the NICA project*", CEUR Workshop Proceedings **2267**, 504–508 (2018)

Выдвинутая на конкурс работа представляется в виде цикла статей, опубликованных с 2018 г. по 2024 г. в рецензируемых научных журналах в рамках темы 1065: "Комплекс NICA: создание комплекса ускорителей, коллайдера и экспериментальных установок на встречных и выведенных пучках ионов для изучения плотной барионной материи, спиновой структуры нуклонов и легких ядер, проведения прикладных и инновационных работ", 02-1-1065-2007/2026, проект BM@N (02-1-1065-2-2012/2026).

Эксперимент BM@N – первый, уже идущий эксперимент на ускорительно-накопительном комплексе NICA, предназначенный для изучения взаимодействия пучков релятивистских тяжелых ионов с энергией до 6 ГэВ на нуклон с неподвижными мишенями. Физическая программа эксперимента направлена на изучение плотной барионной материи, образованной в результате таких столкновений, включая изучение уравнения состояния материи при экстремальных плотностях, материи с сильной изоспиновой асимметрией, смешанной фазы при фазовом переходе первого рода, рождения гиперонов и гиперядер в данном диапазоне энергий, адронной фемтоскопии, пособытийных флуктуаций.

Начиная с 2015 года, было проведено 7 технических сеансов, в которых пучки дейтронов, углерода, аргона и криптона сталкивались с различными типами мишеней, а зимой 2022-2023 гг. был успешно проведен первый физический сеанс эксперимента, в котором пучки ионов ксенона сталкивались с мишенью цезий-йод при энергиях 3.8 и 3 ГэВ на нуклон. Только за последний сеанс было набрано около 600 миллионов событий объемом около 400 ТБ необработанных ("сырых", *raw*) данных. Более того, когда эксперимент достигнет проектных параметров, ожидается, что объем получаемых данных увеличится на порядок. Поскольку все события столкновения частиц эксперимента такого большого объема данных должны быстро обрабатываться и проводиться необходимый физический анализ, то для реализации физической программы BM@N требуется комплексный подход с правильно спроектированной архитектурой программного обеспечения для распределенной обработки данных на предоставляемых эксперименту вычислительных платформах. Важную роль в такой архитектуре, помимо, собственно, самих систем и сервисов, непосредственно связанных с распределенной обработкой, занимают информационные системы, обеспечивающие сбор, хранение, организацию удобного доступа и управление информацией, необходимой для обработки и анализа данных эксперимента.

Результаты, полученные в рамках представленного цикла работ:

1. Разработанная комплексная архитектура программных систем для распределенной обработки данных эксперимента BM@N [1]. Для определения программных систем, которые должны были быть реализованы для автоматизации

выполнения задач обработки больших данных эксперимента, на первом этапе проведено исследование потока обработки экспериментальных данных установки (Рис. 1), разработана модель обработки событий BM@N. В результате анализа определены требуемые программные системы и сервисы, включая информационные системы, необходимые для проведения обработки данных как в режиме онлайн, то есть во время сеансов эксперимента, для контроля качества поступающих данных, так и после сеансов (офлайн) для декодирования данных, реконструкции событий и их физического анализа. Стоит особо отметить, что коллектив авторов проводил разработку комплекса программных систем для распределенной обработки данных после сбора и записи на онлайн кластер эксперимента, то есть не занимался разработкой программных систем DAQ.

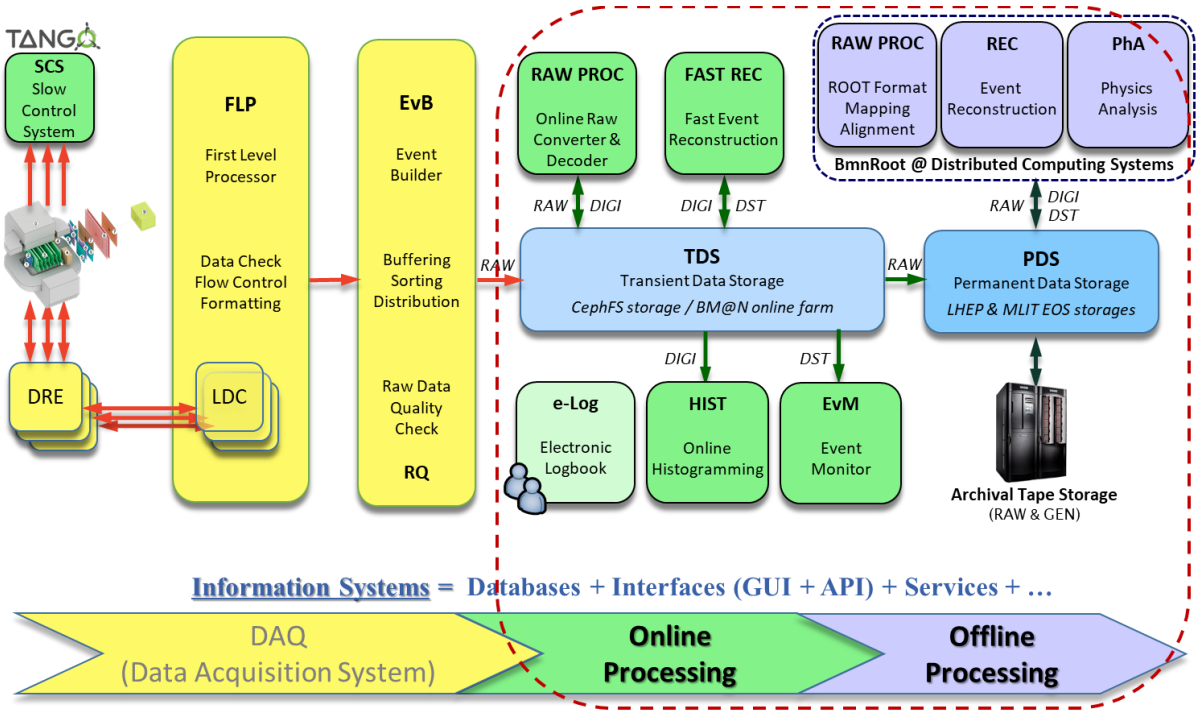

Рис. 1. Условная схема потока обработки экспериментальных данных BM@N (красным прямоугольником выделена часть, автоматизацией которой занимался коллектив авторов)

В итоге для решения задачи быстрой и качественной обработки как экспериментальных, так и моделированных данных событий эксперимента с целью своевременного получения научных результатов была разработана комплексная архитектура программных систем для распределенной обработки данных BM@N, представленная на Рис. 2. Особенностью предоставляемых эксперименту аппаратных вычислительных платформ, а именно: кластера NICA в ЛФВЭ, Центрального информационно-вычислительного комплекса (ЦИВК) ЛИТ, платформы HybriLIT с суперкомпьютером Говорун и онлайн кластера BM@N (DAQ Data Center, DDC), – является разделенность их вычислительных ресурсов и хранилищ данных. Для решения данной проблемы представленная архитектура включает набор программных систем для объединения всех распределенных вычислительных ресурсов и хранилищ данных в единую вычислительную систему с единым пространством хранения, позволяющих управлять потоком обработки больших данных на всех доступных ресурсах, тем самым сокращая время необходимое для получения физических результатов. Данный комплекс включает центральный менеджер управления вычислительной нагрузкой (Workload Management System) DIRAC [A], систему управления данными (Data Management

System) – единый каталог файлов данных DIRAC File Catalog, сервис удаленной передачи данных и сервис конфигурации и управления потоком задач (Workflow Management Service), реализованный на базе решения Apache Airflow [B].

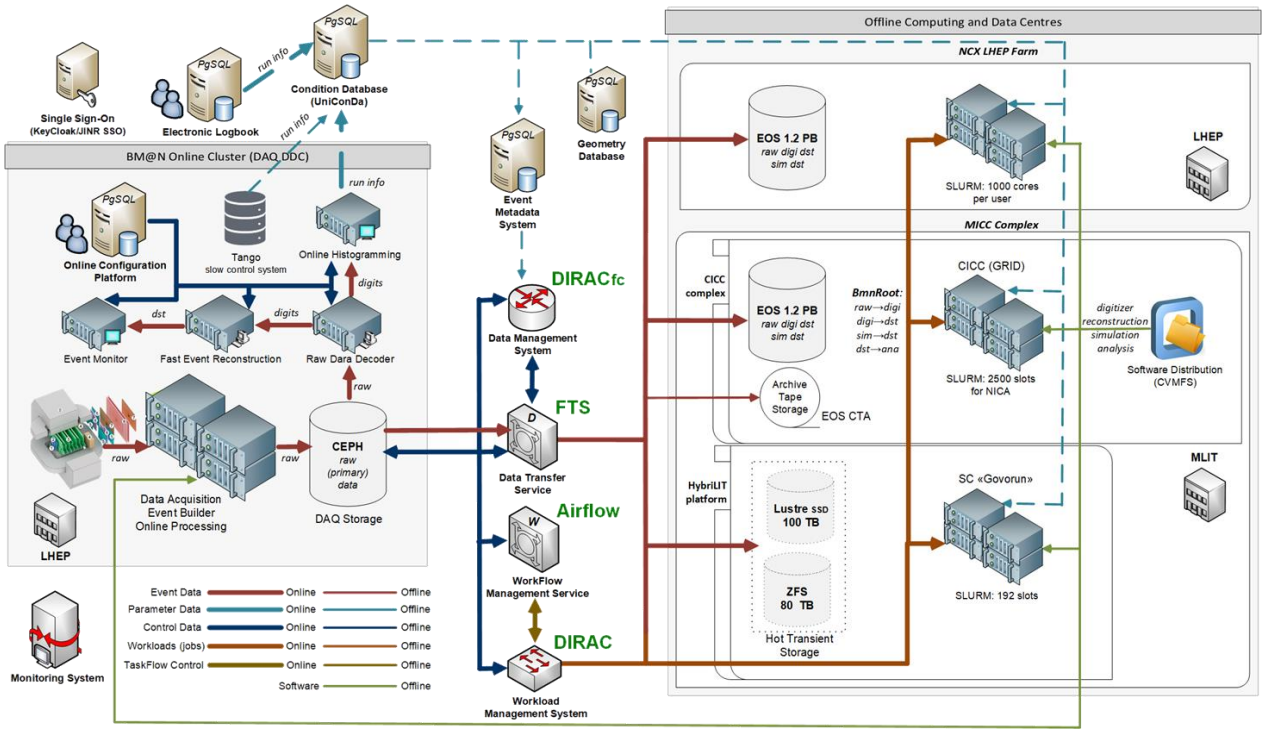

Рис. 2. Разработанная комплексная архитектура программных систем для распределенной обработки данных эксперимента BM@N

Важным моментом для обработки данных как во время работы эксперимента, так и после него является необходимость использования большого количества различных параметров и информации об эксперименте, поэтому разработанная архитектура также содержит комплекс реализованных информационных систем и сервисов, обеспечивающих качественное управление и предоставление требуемой для обработки информации программным системам эксперимента для ее использования на всех этапах, включая декодирование “сырых” данных, реконструкцию событий и их физический анализ, а также моделирование работы установки. Комплекс [14, 23] включает систему электронного журналирования сеансов, конфигурационную онлайн платформу, геометрическую информационную систему, информационную систему на основе базы данных состояний и условий работы, а также систему метаданных событий. Разработанные информационные системы построены на современных базах данных и предоставляют пользовательские сервисы для прозрачного доступа и управления хранимыми данными и информацией об эксперименте, позволяя одновременно обслуживать большое количество запросов от программных систем и пользователей, а также обеспечивают автоматическое резервное копирование данных на случай возникновения сбоя в работе программного обеспечения или оборудования. Также представленные на рисунке система дистрибуции программного обеспечения, сервис единой аутентификации и авторизации и сервис мониторинга программных систем повышают эффективность и надежность разработанной архитектуры.

2. Комплекс программных систем по автоматизации выполнения потока задач обработки больших данных эксперимента BM@N.

2.1. Центральный менеджер управления работами DIRAC Interware для распределенной обработки данных эксперимента [1, 2]. Для распределенной обработки экспериментальных и смоделированных данных BM@N внедрена и успешно используется в качестве центрального менеджера работ платформа DIRAC Interware, которая предоставляет необходимые компоненты для построения инфраструктуры обработки данных, связывая вычислительные ресурсы разного типа. Система DIRAC развернута на вычислительной инфраструктуре ОИЯИ и объединяет доступные эксперименту BM@N ресурсы кластера NICA, центров Tier1 и Tier2 ЦИВК, суперкомпьютера Говорун, а также онлайн кластера BM@N, успешно используемого также для обработки данных между сеансами эксперимента. Помимо вычислительных ресурсов эксперимент BM@N использует интегрированные в DIRAC системы хранения: элемент хранилища на файловой системе EOS и систему ленточного хранения CTA. Кроме того, при появлении у коллаборации облачных ресурсов или вычислительных ресурсов внешних организаций, DIRAC позволит провести подключение и их в общую систему обработки. Центральный менеджер работ включает графический веб-интерфейс и интерфейс командной строки, используемые для запуска и управления тысячами заданий по обработке данных эксперимента на всех предоставляемых вычислительных ресурсах.

В результате в настоящее время менеджер работ DIRAC активно используется коллаборацией для массовой обработки экспериментальных и смоделированных данных эксперимента BM@N, позволяя, например, как декодировать все “сырые” данные физического сеанса за 35 часов, так и реконструировать события столкновения частиц последнего сеанса с энергией пучка 3.8 ГэВ на нуклон за 7 дней. С момента появления данных 8го сеанса для задач связанных с обработкой этих данных было использовано более 200 лет процессорного времени в пересчете на одно вычислительное ядро (Рис. 3). Также система демонстрирует высокую скорость передачи данных по сети по протоколу XRootD, которая достигает 2 ГБ/с, и справляясь в пике с нагрузкой почти в 8 ГБ/с.

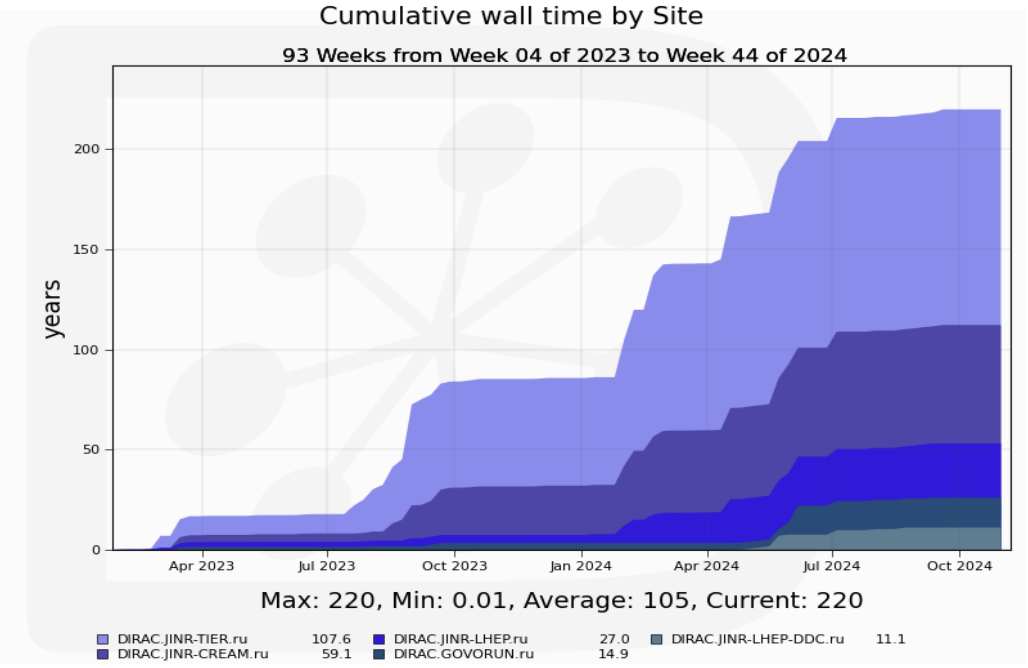

Рис. 3. Накопленное время обработки данных последнего сеанса на центрах Tier1 и Tier2 ЦИВК (DIRAC.JINR-TIER.ru и DIRAC.JINR-CREAM.ru), кластере NICA (DIRAC.JINR-LHEP.ru), суперкомпьютере Говорун (DIRAC.GOVORUN.ru), онлайн кластере (DIRAC.JINR-LHEP-DDC.ru)

С февраля 2023 года было проведено 13 больших кампаний по обработке и генерации данных для коллаборации BM@N и более 20 тестовых запусков на небольших выборках. Апробированы системы мониторинга и аналитики, которые позволяют сравнивать разные кампании по обработке данных, следить за корректностью работы программного обеспечения, анализировать эффективность использования ресурсов.

2.2. Система каталога файлов DIRAC File Catalog для создания единого пространства хранения в эксперименте BM@N [1, 2]. Проблема использования экспериментом нескольких разделенных хранилищ данных привела к потребности в специализированной системе управления данными, называемой также системой единого каталога файлов, которая сопоставляет логические имена файлов с конечными физическими на хранилищах данных, позволяя пользователю видеть разделенные хранилища как единый каталог файлов. Для решения данной задачи в эксперименте BM@N успешно реализовано решение на базе встроенного каталога файлов данных платформы DIRAC – DIRAC File Catalog. Интегрированный каталог файлов предоставляет широкую функциональность, включая репликацию данных и внесение и использование метаданных, то есть атрибутов, содержащих суммарную информацию о файле.

Развернутый каталог файлов эксперимента BM@N обеспечивает единое пространство имен файлов, используемое при распределенной обработке данных, и в настоящее время содержит список реконструированных файлов физического сеанса с большим набором соответствующих метаданных: номером запуска, временем набора, типом частицы пучка и мишени, энергией пучка, значением магнитного поля, количеством событий в файле и других. Разработанный прикладной программный интерфейс (Рис. 4) в виде сервиса на архитектуре REST [F] предоставляет возможность поиска, используя заданные метаданные, и получения списка только тех файлов, которые необходимы для конкретного физического анализа реконструированных данных.

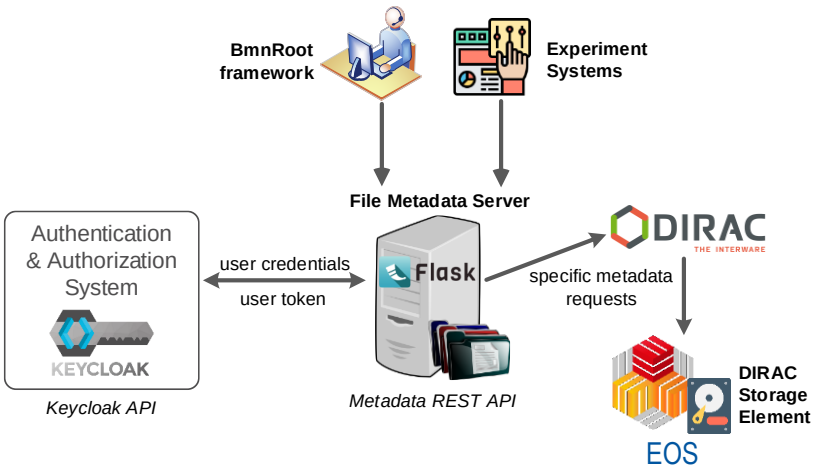

Рис. 4. Обработка запроса прикладным-программным интерфейсом каталога файлов

2.3. Разработанный менеджер конфигурации и управления потоком задач эксперимента на базе решения Apache Airflow [1]. Для автоматизации выполнения повторяющихся наборов задач по обработке больших данных BM@N был реализован менеджер конфигурации и управления потоком задач эксперимента на базе решения Apache Airflow, которое предоставляет платформу для создания, планирования и мониторинга выполнения заранее определенного набора рабочих процессов. Такие решения также

называются системами оркестрации, потому что управляют выполнением стандартных задач эксперимента, использующих различные программные системы, собирая все вместе в единый поток исполнения. В соответствии с описанным набором задач с установленными зависимостями между ними в виде направленного ациклического графа менеджер формирует цепочки рабочих процессов, необходимых для обработки данных, запускает, управляет и мониторирует их выполнение. Развернутое решение (Рис. 5) уже использовалось для передачи во время сеанса приходящих экспериментальных данных с онлайн фермы DDC на основные офлайн хранилища в ЛФВЭ и ЛИТ ОИЯИ с проверкой целостности переданных файлов. В дальнейшем планируется расширить использование системы оркестрации для запуска массовой обработки поступающих данных через платформу DIRAC, а также автоматизированное архивирование полученных “сырых” данных эксперимента на надежное ленточное хранилище BM@N в ЛИТ.

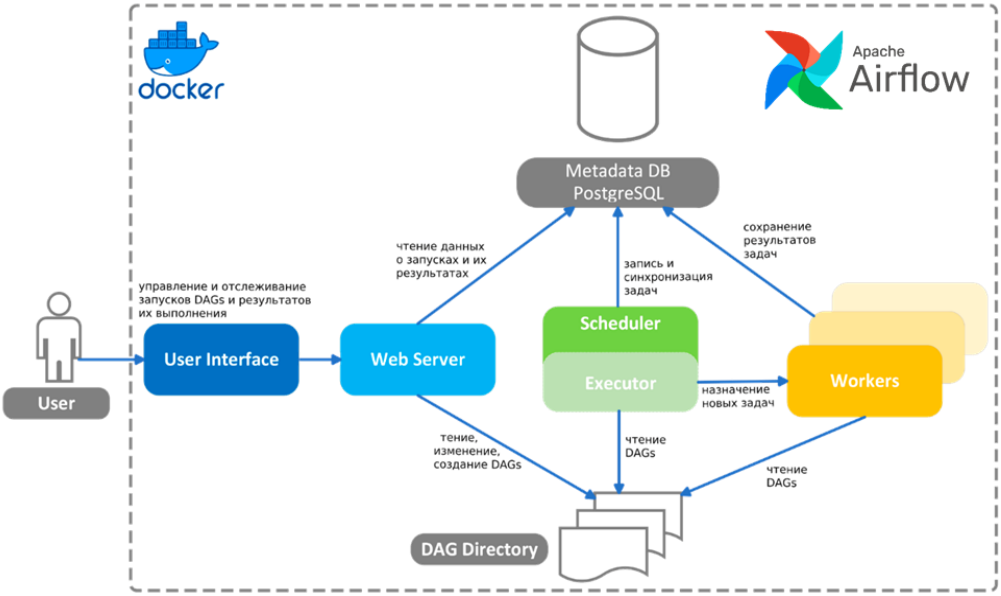

Рис. 5. Взаимосвязь компонент менеджера конфигурации и управления потоком задач

2.4. Система сборки, тестирования и распространения программного обеспечения эксперимента BM@N [1, 3]. Чтобы обеспечить доступность программного обеспечения эксперимента на всех используемых вычислительных платформах, были организованы два хранилища: центральное программное хранилище, основанное на сетевой файловой системе CernVM-FS [C], и реестр контейнеров. Контейнеризация среды выполнения программного обеспечения эксперимента была применена для упрощения процесса подготовки пользовательских приложений к работе в условиях разнородного программного окружения на рабочих узлах вычислительных кластеров и персональных компьютерах. Кроме того, изолирование среды выполнения от программного окружения вычислительных узлов снижает нагрузку на системных администраторов вычислительных инфраструктур, поскольку от них требуется лишь поддержка системы контейнеризации.

Система сборки, тестирования и публикации программного обеспечения была реализована с помощью средств GitLab CI/CD [D]: был разработан конвейер задач, который автоматически осуществляет эти операции в автоматическом режиме при каждом внесении изменений в кодовую базу основного фреймворка эксперимента – BmnRoot. Поскольку все эти задачи являются весьма ресурсоемкими, для их выполнения было задействовано облако

ОИЯИ в качестве вычислительного ресурса. В рамках этого конвейера также осуществляется сборка и публикация образов контейнеров в формате Docker [E], который может использоваться в различных системах контейнеризации, таких как Apptainer на вычислительных кластерах и Docker на персональных компьютерах. Как образы контейнеров, так и программное обеспечение распространяются на вычислительные кластера через репозиторий эксперимента, размещенный в CernVM-FS, используемой при распределенной обработке данных всеми вычислительными платформами.

3. Разработанный комплекс информационных систем, обеспечивающих сбор, хранение и предоставление информации, требуемой для обработки данных.

3.1. Система электронного журналирования сеансов эксперимента [15, 22]. Во время сеансов эксперимента важное значение для понимания происходящих событий имеют не только данные, собираемые с детекторов, но и записи операторов смен в журналах, описывающие режимы работы и состояния различных подсистем, детекторов и сами события. Для решения данной задачи разработана и успешно используется система электронного журнала, предоставляющая операторам смен веб-интерфейс (*bmn-elog.jinr.ru*) для записи во время сеансов эксперимента информации о текущих параметрах и режимах работы подсистем, о текущих событиях, возникших проблемах и предпринятых действиях, а также предназначенная для удобного просмотра, корректировки и поиска требуемой информации в журнале участниками коллаборации. Она обеспечивает автоматическое резервирование данных журнала и предоставляет следующие возможности в зависимости от роли аутентифицированного пользователя: просмотр, добавление или изменение записей в журнале, прикрепление текстовых и графических файлов, сортировку и фильтрацию данных, удобный поиск по параметрам, а также важную во время сеансов функцию – подписку в личном кабинете на уведомления при появлении записей выбранного типа.

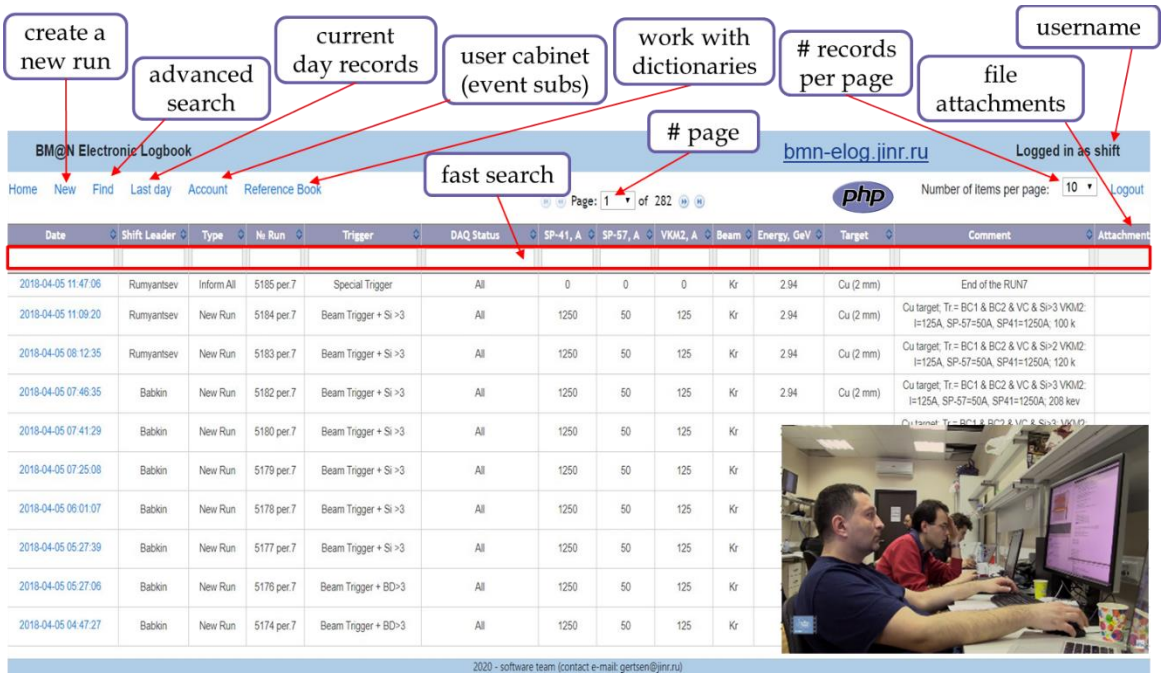

Рис. 6. Интерфейс электронного журнала сеансов эксперимента BM@N с отмеченной функциональностью (на фото внизу операторы смен сеансов BM@N)

Реализованная система электронного журналирования включает базу данных для хранения журнала эксперимента и работы с ним, а для последующего использования внесенных данных журнала другими программными системами, в том числе в алгоритмах декодирования, обработки и анализа событий столкновения частиц, разработан прикладной программный интерфейс, а также набор вспомогательных сервисов. Например, один из сервисов позволил автоматически формировать в последнем сеансе эксперимента BM@N ежедневную статистику по собранным событиям столкновения частиц разного типа.

3.2. Конфигурационная онлайн платформа эксперимента [5, 7, 11]. Другая проблема, потребовавшая реализации новой программной системы, – это конфигурация и управление требуемым набором задач онлайн обработки во время сеанса эксперимента. Для проверки качества поступающих экспериментальных данных в режиме онлайн должны непрерывно выполняться и обмениваться между собой данными программные задачи по декодированию “сырых” данных, построению контрольных гистограмм, быстрой реконструкции и монитору событий столкновения частиц. Управление одним онлайн процессом, в принципе, возможно вручную, но если речь идет о полном наборе процессов, которые должны работать одновременно на распределенных ресурсах и зависимо друг от друга, то эффективно управлять ими вручную невозможно. Для автоматизации онлайн обработки данных BM@N реализована конфигурационная онлайн платформа (Рис. 7), предназначенная для хранения и предоставления данных о конфигурации аппаратных систем и программных задач эксперимента, выполняющихся в режиме онлайн.

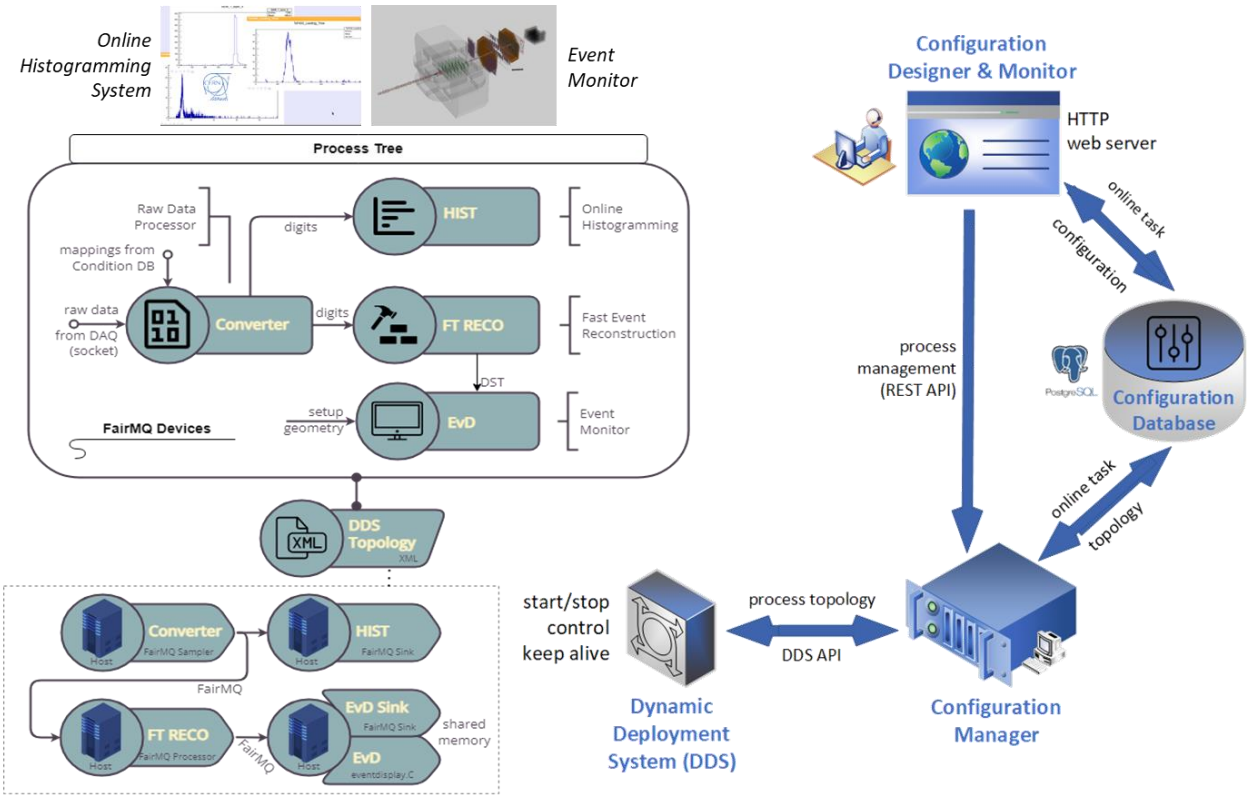

Рис. 7. Проверка качества данных при помощи конфигурационной онлайн платформы

В разработанной конфигурационной базе данных хранится как набор требуемых конфигурационных параметров, так и описание последовательности задач, которые должны запускаться и выполняться во время сеанса. Оператор описывает в дизайнере

конфигураций требуемые онлайн задачи, система загружает данную последовательность в конфигурационную базу данных, а центральный менеджер считывает сформированную топологию, запускает все необходимые процессы с требуемыми параметрами на указанных распределенных узлах, управляет, а также автоматически перезапускает их в случае сбоя. Для запуска на распределенных ресурсах и управления задачами онлайн обработки, а также обеспечения их взаимодействия друг с другом выбрана система динамического развертывания DDS [G]. Веб-интерфейс платформы (*bmn-online.jinr.ru*) помимо удобного задания и управления топологией процессов также включает монитор активных задач, предоставляющий информацию по их статусу и потоковый вывод.

3.3. База данных состояний и условий работы эксперимента BM@N [13, 18, 20, 21, 24]. Другой важной проблемой при решении задач обработки полученных данных является необходимость использования большого количества различных параметров подсистем эксперимента, поэтому в рамках данной архитектуры реализована информационная система, основанная на параметрической базе данных, которую также называют база данных состояний и условий работы (Condition Database). Данная система разработана и активно используется, предоставляя членам коллаборации унифицированный доступ, поиск и управление параметрической информацией об эксперименте, необходимой для обработки экспериментальных и смоделированных данных, обеспечивая актуальность, согласованность и автоматическое резервирование данных. Важным свойством хранимых параметрических данных является то, что они характеризуются временным интервалом их действия и используются для обработки данных, собранных только в течение действия соответствующего параметра. Кроме того, разработанная архитектура обеспечивает хранение параметров произвольной структуры.

База данных состояний BM@N содержит следующие основные части: хранение информации о модельных файлах, полученных генераторами событий, о проведенных запусках эксперимента (метаданные ранов) и соответствующих файлах данных, а также главную часть – хранение значений требуемых (конфигурационных, калибровочных, алгоритмических и других) параметров подсистем произвольного формата, необходимых для проведения обработки данных. Для удобного просмотра пользователями коллаборации через браузер соответствующих данных разработан специальный веб-интерфейс (*bmn-uniconda.jinr.ru*), как отображающий сводную статистику по хранимым данным (Рис. 8), так и предоставляющий удобную форму просмотра, управления и поиска параметрической информации об эксперименте. Главные реализованные интерфейсы базы данных состояний – это прикладные программные (REST API и C++), используемые, например, для получения в задачах BmnRoot параметров, необходимых для моделирования, декодирования, реконструкции событий и их физического анализа.

Помимо разработки в рамках данной архитектуры крупных программных систем, также в эксперименте реализован большой набор различных вспомогательных сервисов [10]. В качестве примера, на языке Python разработана служба проверки доступности и целостности (File Inspector) экспериментальных файлов и файлов с модельными данными, описанными в базе данных состояний, что обусловлено большим количеством данных и длительностью их хранения, а также необходимостью оперативно реагировать на потерю или повреждение данных. Разработанный сервис выполняет регулярную проверку файлов с событиями эксперимента на распределенных кластерах. Он проверяет, что файлы

присутствуют, доступны для чтения и не были повреждены, а вся информация по проверкам выводится на централизованный веб-сайт службы.

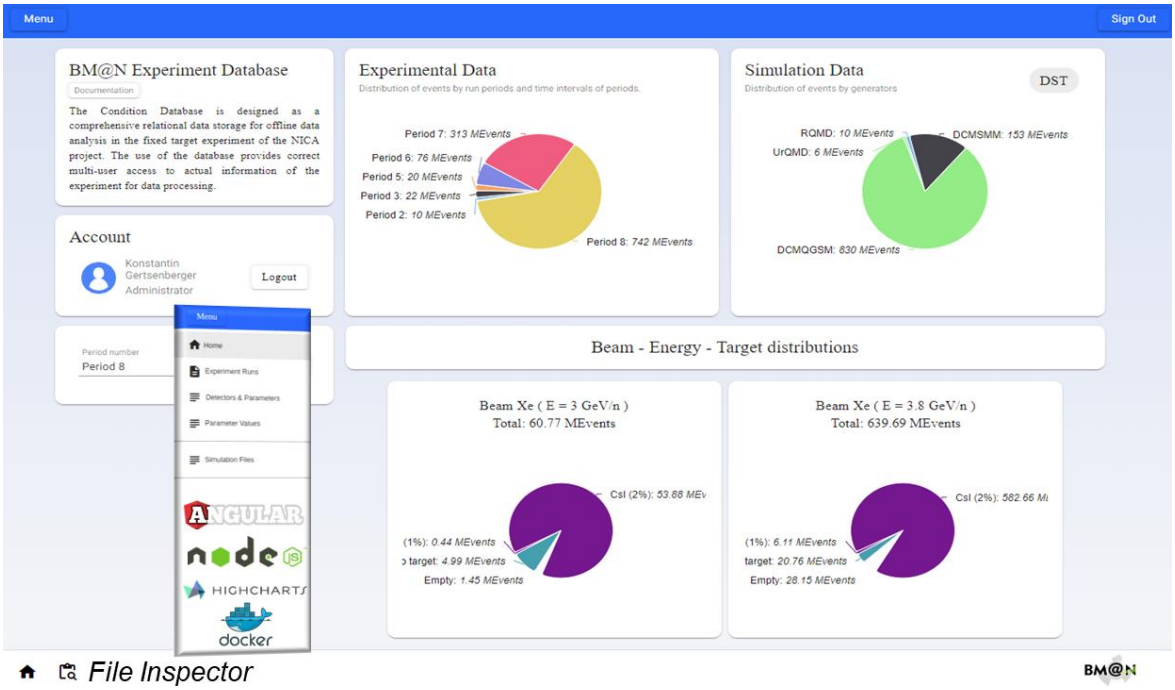

Рис. 8. Веб-интерфейс базы данных состояний и условий эксперимента BM@N

3.4. Геометрическая информационная система BM@N [16, 19, 25]. Для обработки данных событий столкновения частиц важное значение имеют используемые версии геометрических компонент установки, задающие реализацию объектов геометрии в программном обеспечении эксперимента. Поэтому еще одним реализованным программным решением стала информационная система для работы с геометрией детекторов установки, базирующаяся на разработанной геометрической базе данных. Геометрическая система предназначена для хранения, обработки и управления информацией о геометрической модели детекторов, которая в дальнейшем используется для обработки и анализа моделированных и экспериментальных данных. Для каждого геометрического модуля хранится идентификатор, версия, матрица преобразования и ссылка на родительский модуль. Версии полной геометрии установки определяются в виде комбинации составляющих геометрических модулей детекторов, описания магнитного поля и используемых материалов и среды, сохраняемых в базе данных. Разработанная геометрическая информационная система предоставляет централизованное хранилище геометрической информации и набор удобных инструментов для управления как различными версиями отдельных модулей, так и сборками версий полной установки.

Геометрическая информационная система построена по модели взаимодействия “клиент-сервер” (Рис. 9), где серверная часть представлена центральным хранилищем геометрий на СУБД PostgreSQL, обеспечивающем все функциональные возможности, а клиентская часть предусматривает работу с локальными пользовательскими репликами центральной базы данных на СУБД SQLite. Реализованы основные интерфейсы работы с геометрической системой: прикладной программный интерфейс для выбора и загрузки геометрии установки и ее составляющих в программное обеспечение эксперимента для их учета в алгоритмах моделирования, реконструкции и физического анализа полученных данных, а

также графический интерфейс в виде веб-сервиса (*bmn-geometry.jinr.ru*) для доступа ко всем функциям системы в зависимости от категории пользователя.

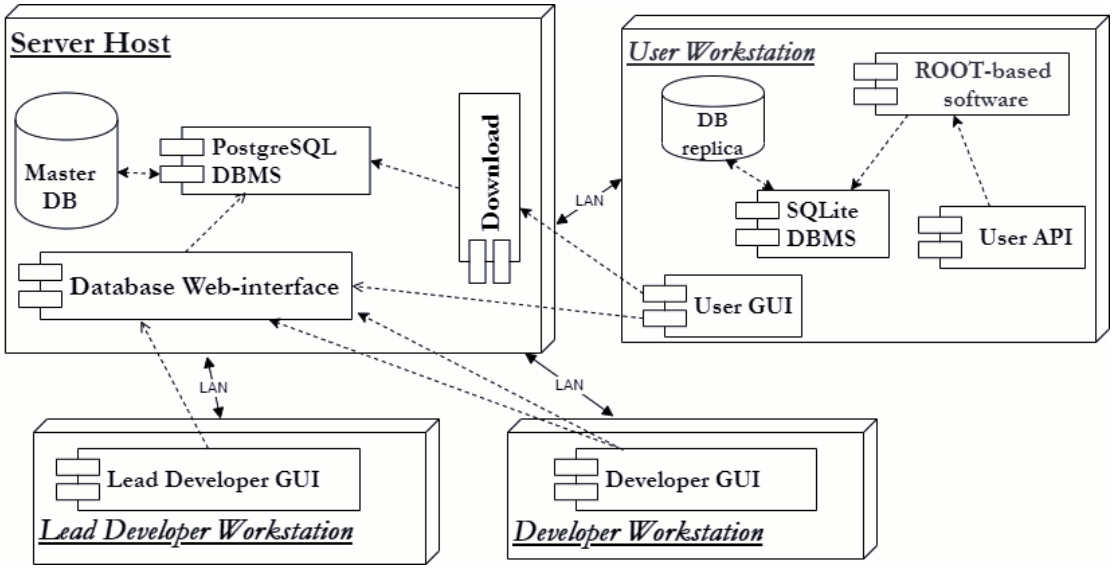

Рис. 9. Архитектура геометрической базы данных эксперимента BM@N

3.5. Система метаданных событий (каталог событий) эксперимента [6, 8, 12, 17]. Другой актуальной проблематикой, которая решается в эксперименте BM@N, является поиск и отбор только тех событий столкновений частиц, которые требуются для проведения конкретного физического анализа, для чего в рамках данной работы была реализована система метаданных событий, содержащая сводные параметры событий, необходимые для отбора, называемые метаданными, например, число реконструированных треков. Система основана на разработанной базе данных, называемой каталогом событий для хранения метаданных событий BM@N, версий используемых при обработке программ, адресов событий в распределенных хранилищах, и позволяет пользователю искать и отбирать с использованием различных критериев набор только тех событий, которые необходимы для конкретного физического анализа. Соответствующие признакам события идентифицируются уникальной ссылкой, представляющей собой комбинацию ссылки на файл данных в распределенном хранилище и номером события в этом файле.

Архитектура системы метаданных событий является достаточно всеобъемлющей (Рис. 10) и включает, в частности, разработанный пользовательский веб-интерфейс (*bmn-event.jinr.ru*) для просмотра и поиска членами коллаборации метаданных событий, хранящихся в каталоге, а также для запроса и получения только необходимых событий, удовлетворяющих заданным параметрам, для их физического анализа. Для повышения эффективности процесса отбора событий, если возможно, используется информация о запусках эксперимента, хранимая в базе данных состояний, для предварительного отбора событий перед выполнением поиска в самом каталоге. Реализован прикладной программный интерфейс в виде сервиса REST API, используемый для записи информации о полученных новых событиях эксперимента и выполнения запросов от пользователей и других программных систем на требуемые события с использованием различных критериев поиска. Реализованная система является конфигурируемой для поддержки произвольного набора метаданных событий эксперимента, обеспечивает управление доступом на основе ролей членов коллаборации и мониторинг работы компонент системы.

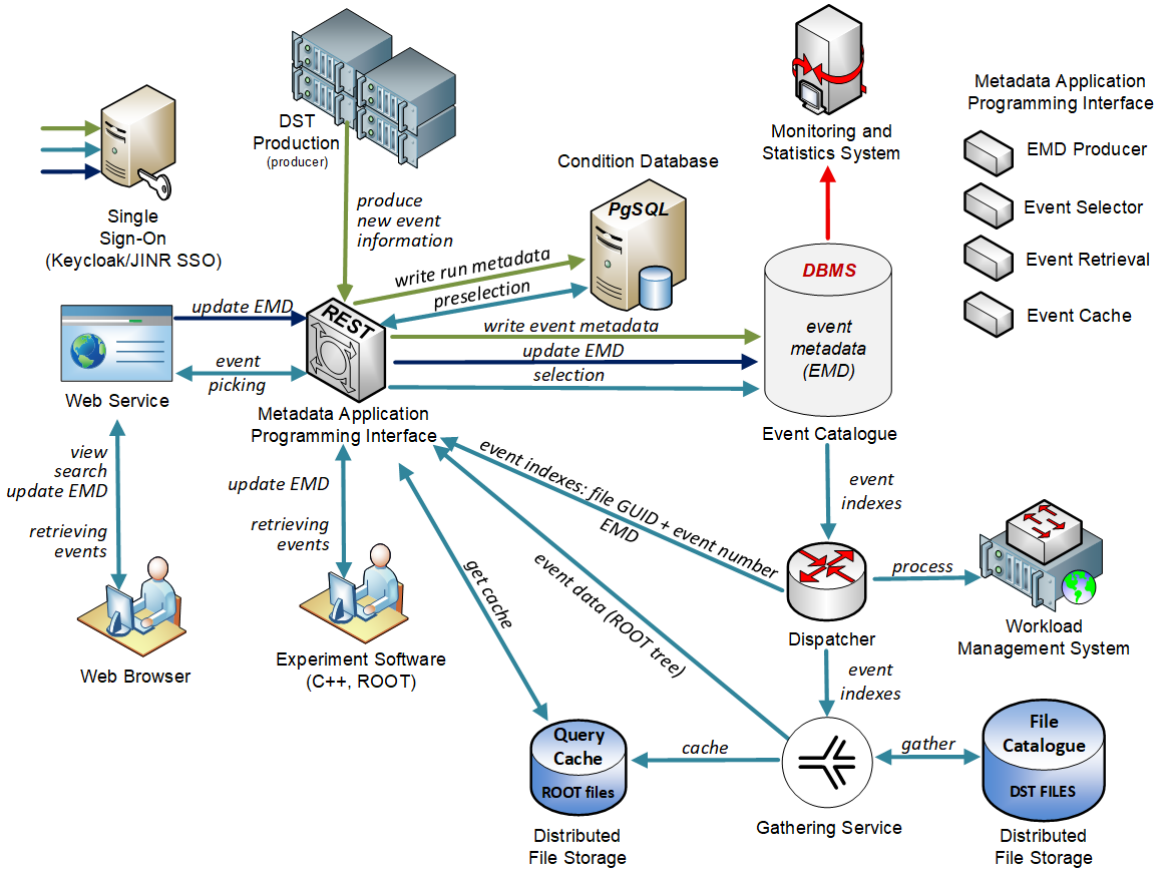

Рис. 10. Архитектура системы метаданных событий эксперимента BM@N

3.6. Сервис унифицированного развертывания разработанных информационных систем [9]. Представленные выше информационные системы могут быть востребованы и в других экспериментах по столкновению частиц, поэтому они были сделаны конфигурируемыми для возможности использования в экспериментах со схожими требованиями, и реализован сервис унифицированной конфигурации и развертывания, обеспечивающий настраиваемую установку приведенных систем. В разработанных системах при помощи внешних параметров задаются как простые элементы, включая логотип и контактные данные, так и более сложные, например, поля электронного журнала или состав метаданных каталога событий. Основная концепция созданной системы развертывания заключается в использовании конфигурационного файла с настраиваемыми параметрами и установочного скрипта, который разворачивает и настраивает все компоненты информационной системы в соответствии со спецификой эксперимента, включая центральную базу данных и необходимые интерфейсы для работы с использованием мультиконтейнерной архитектуры на нескольких серверах. Результатом конфигурируемой установки является созданная база данных, интерфейсы и сервисы, развернутые в Docker-контейнерах на заданных машинах, настроенное регулярное автоматическое обновление из репозитория и резервное копирование хранимых данных, после чего коллаборанты могут приступать к работе с готовой системой.

4. Разработанные вспомогательные сервисы, повышающие эффективность и надежность реализованной архитектуры.

4.1. Система единой аутентификации и авторизации членов коллаборации [1]. Разработанная архитектура программных систем для организации распределенной обработки данных BM@N также включает в себя набор важных вспомогательных систем, одной из которых в эксперименте является система централизованной аутентификации и авторизации для администрирования и разграничения прав членов коллаборации в программных системах и информационных сервисах. Разработанная система основана на современном решении Keycloak, которое позволяет участникам иметь одну учетную запись для всех программных систем эксперимента, а также использовать ее для перемещения по системам без повторного ввода данных аккаунта. Она обеспечивает централизованную аутентификацию и авторизацию и хранит информацию об учетных записях и группах пользователей, которые разграничивают доступ членов коллаборации и определяют доступные им действия. Запросы на данные представленных информационных систем эксперимента идут через разработанные интерфейсы, взаимодействующие с этой системой для аутентификации пользователей и выполнения проверки их роли и разрешения на запрошенную операцию.

4.2. Сервис мониторинга программных систем эксперимента [1, 4]. Для минимизации времени реакции в случае аппаратных или программных сбоев в работе реализованных программных систем эксперимента важно непрерывно следить за их состоянием, что особенно критично во время проведения сеансов. Для решения данной задачи реализован сервис мониторинга, предназначенный для отслеживания существующей инфраструктуры программного обеспечения BM@N, включая задействованные сервера, базы данных и веб-интерфейсы систем, хранения параметров их состояния во временной базе данных и визуализации при помощи пакета Grafana [H] на центральном веб-сервисе ОИЯИ (Рис. 11). В случае сбоев в работе или превышения критических значений параметров систем сервис мониторинга рассылает соответствующие уведомления ответственным лицам, которым необходимо оперативно отреагировать на проблему и предпринять необходимые действия, на указанные адреса электронной почты и специальный Telegram-канал.

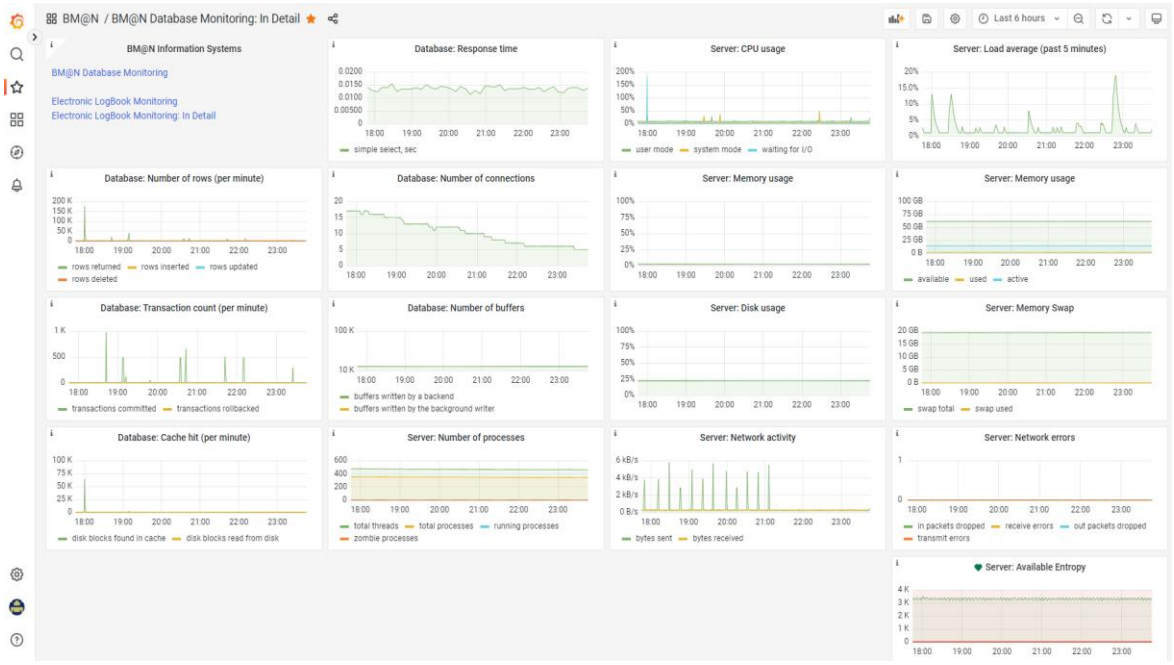

Рис. 11. Веб-интерфейс системы мониторинга программных систем эксперимента BM@N

Таким образом, в рамках представляемой на конкурс работы коллективом авторов была спроектирована и реализована архитектура, включающая комплекс современных программных решений для организации распределенной обработки данных эксперимента BM@N на предоставляемых вычислительных платформах, а также хранения, управления и предоставления унифицированного доступа к информации, требующейся для проведения обработки данных событий на всех этапах, включая декодирование необработанных данных, реконструкцию, физический анализ, как и моделирования работы установки.

Реализованная архитектура включает как программные системы (такие как центральный менеджер управления работами, единый каталог файлов данных, менеджер конфигурации и управления потоком обработки), решающие задачу объединения распределенных ресурсов эксперимента в единую систему обработки и хранения для автоматизации выполнения потока задач, так и оригинальные информационные системы (электронный журнал, конфигурационную онлайн платформу, геометрическую информационную систему, базу данных состояний и систему метаданных событий), обеспечивающие сбор, хранение, управление и организацию доступа к информации, необходимой для обработки и анализа полученных данных, на протяжении жизненного цикла научных исследований эксперимента BM@N. Кроме того, набор вспомогательных сервисов, таких как система дистрибуции программного обеспечения, сервис единой аутентификации и авторизации и сервис мониторинга программных систем, повышает эффективность и надежность разработанной архитектуры.

Реализованный комплекс программных систем развернут на существующей инфраструктуре эксперимента BM@N, успешно используется при решении задач распределенной обработки, хранения и физического анализа собранных экспериментальных (как и модельных) данных и является необходимым элементом для качественного управления данными и своевременного получения физических результатов в условиях работы с большими данными.

Полученные результаты представлялись на коллаборационных совещаниях эксперимента BM@N, совещании консорциума RDIG-M, а также были представлены в 25 докладах на различных конференциях.

Литература

[A] F. Stagni, A. Tsaregorodtsev, A. Sailer and C. Haen, *"The DIRAC interware: current, upcoming and planned capabilities and technologies",* EPJ Web Conf. **245**, 03035 (2020)

[B] B. Harenslak and J. Rutger de Ruiter, Data Pipelines with Apache Airflow, 480 p. (2021)

[C] J. Blomer, P. Buncic, R. Meusel, G. Ganis, I. Sfiligoi and D. Thain, "*The evolution of global scale filesystems for scientific software distribution*", CiSE **17**, no.6, 61-71 (2015)

[D] Ya. Jani, "*Implementing continuous integration and continuous deployment (ci/cd) in modern software development*", IJSR **12**, no.6, 2984-2987 (2023)

[E] D. Moreau, K. Wiebels and C. Boettiger, "*Containers for computational reproducibility*" Nat Rev Methods Primers **3**, no.1, 50 (2023)

[F] C. Pautasso, E. Wilde and R. Alarcon, REST: Advanced Research Topics and Practical Applications, 222 p. (2014)

[G] A. Lebedev, A. Manafov, “*DDS: The Dynamic Deployment System*”, EPJ Web Conf. **214**, 01011 (2019)

[H] E. Salituro, Learn Grafana 7.0: A beginner's guide to getting well versed in analytics, interactive dashboards, and monitoring, 410 p. (2020)

Председатель НТС ЛФВЭ Е. А. Строковский

Ученый секретарь НТС ЛФВЭ С. П. Мерц