

ПРОЕКТИРОВАНИЕ СИСТЕМЫ МОНИТОРИНГА КАЧЕСТВА ДАННЫХ ДЛЯ ЭКСПЕРИМЕНТА BM@N

Е. И. Александров^{1,2}, И. Н. Александров^{1,2,},
А. И. Чеботов¹, К. В. Герценбергер^{1,2}*

¹ Объединенный институт ядерных исследований, Дубна, Россия

² Государственный университет «Дубна», Дубна, Россия

Представлен дизайн системы мониторинга качества данных (DQM) для эксперимента BM@N проекта NICA, включающий как описание целей создания системы, краткий обзор подобных систем, работающих в экспериментах LHC в ЦЕРН, так и общие подходы к их созданию. Представлены архитектура системы, объектная модель и схема базы данных, а также описана структура конфигурационных файлов, которые будут использоваться для тонкой настройки работы системы.

The paper presents the design of the Data Quality Monitoring (DQM) system for the BM@N experiment of the NICA project, including a description of the system's objectives, a brief overview of similar systems that operate in the CERN LHC experiments, and general approaches to creating the systems. The system architecture, object model, and database scheme are presented, and the structure of the configuration files that will be used for fine-tuning the work of the DQM is described.

PACS: 89.20.Ff; 07.05.Kf

ВВЕДЕНИЕ

Контроль качества собираемых физических данных как в онлайн-режиме с целью своевременного и быстрого реагирования в случаях потери качества данных, так и в режиме офлайн для более детальной проверки качества и сертификации данных является важной частью систем сбора информации в экспериментах физики высоких энергий. Так, во всех четырех экспериментах LHC в ЦЕРН эти системы были введены после запуска промышленного сбора данных. Несмотря на схожесть задач, все системы имеют отличия, связанные со специфическими характеристиками их детекторов и потоков собираемых данных. Так, например, в ATLAS [1] упор делался на гибкость системы и наполненность возможностями графического интерфейса. В эксперименте ALICE [2]

* E-mail: alexand@jinr.ru

создана одна из самых больших систем с теоретически неограниченными возможностями ее масштабируемости. Своевременное обнаружение и локализация ошибок и проблем, возникающих в аппаратном обеспечении детектора или программном обеспечении реконструкции, с достаточной точностью и ясностью для достижения хорошей эффективности детектора и его работы были основной целью такой системы в CMS [3]. Во всех экспериментах увеличивались возможности автоматического анализа качества данных, как, например, в эксперименте LHCb [4], где при создании системы этому уделялось меньше внимания. BM@N [5] — первый эксперимент на комплексе коллайдера мегапроекта NICA [6], который уже приступил к сбору данных. Таким образом, задача мониторинга качества данных является важной в рамках работы детектора. Настоящая работа посвящена описанию дизайна системы мониторинга качества данных эксперимента BM@N.

ОСНОВНЫЕ ЦЕЛИ СОЗДАНИЯ СИСТЕМЫ

Система мониторинга и обеспечения качества данных для эксперимента BM@N должна предоставлять одинаковым образом контрольные гистограммы в трех режимах: онлайн декодирование и реконструкция (онлайн мониторинга данных), качественная оценка новых версий BmnRoot после GIT Merge Request и ручной запуск для проверки пользовательских версий программного обеспечения.

Архитектура DQA (т.е. онлайн DQM и офлайн QA (Quality Assurance)) BM@N должна обеспечивать предопределенные проверки данных и графическое представление контрольных гистограмм на центральном веб-интерфейсе, который получает отображаемые распределения данных от производителя гистограмм BM@N (центрального менеджера DQA). Кроме того, система DQA должна предоставлять возможность легко добавлять новые контрольные гистограммы с предопределенными проверками и оповещениями в формате, подходящем для физиков и членов детекторных рабочих групп, не являющихся разработчиками программного обеспечения, например, с использованием JSON-описания.

ОБЩАЯ АРХИТЕКТУРА СИСТЕМЫ МОНИТОРИНГА КАЧЕСТВА ДАННЫХ

Архитектура системы разработана с учетом того, что эксперимент BM@N имеет относительно небольшие потоки данных по сравнению с LHC. Поэтому главными принципами при создании архитектуры были не масштабируемость системы, а в первую очередь гибкость, надежность и простота в использовании, возможность удобно добавлять новые алгоритмы создания и анализа гистограмм и другие параметры контроля качества данных. В соответствии с целями создания системы учитывалась возможность ее использования в офлайн-режиме. На рис. 1 представлена диаграмма общей архитектуры системы мониторинга качества данных

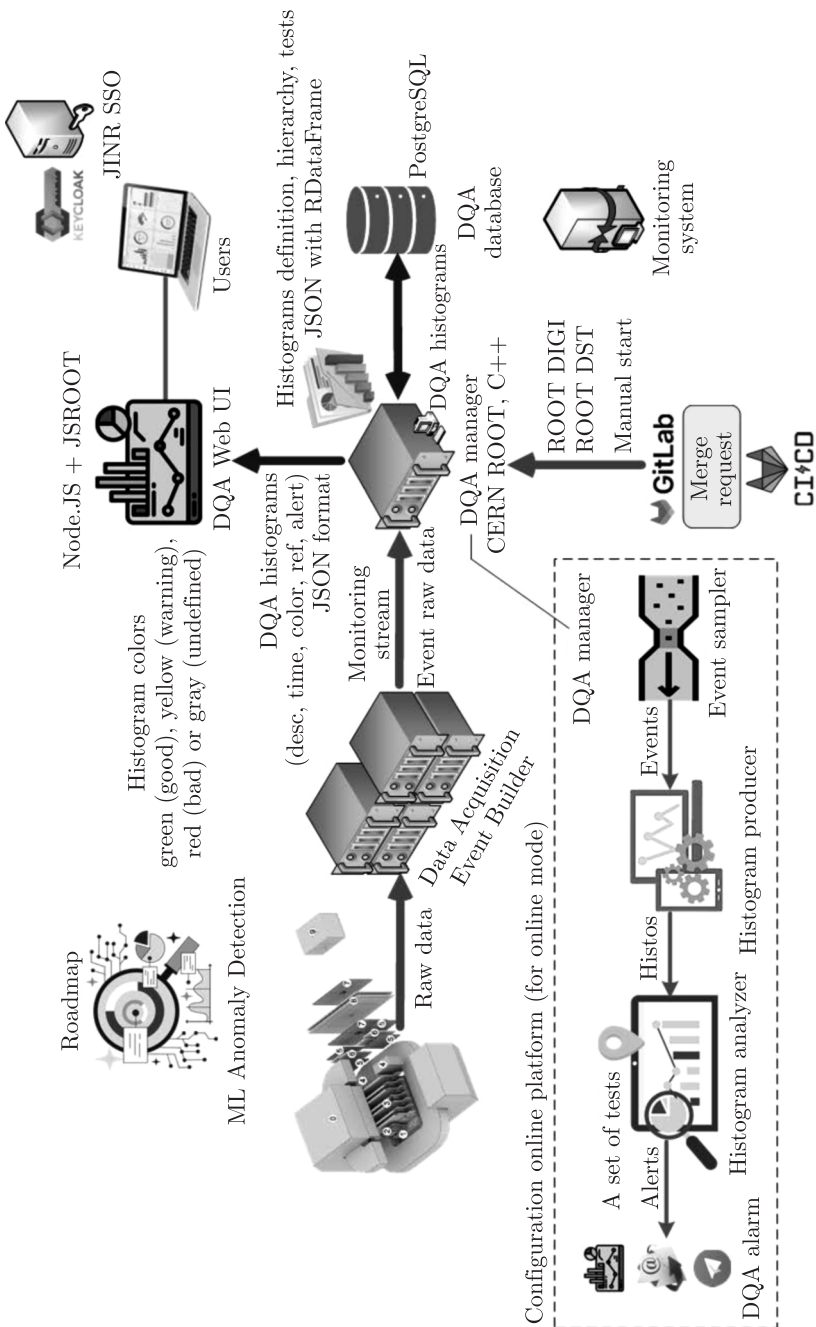


Рис. 1. Общая архитектура системы мониторинга и обеспечения качества данных

BM@N в рамках общей системы сбора и обработки физических событий. «Сырые» данные после детектора посылаются в Event Builder, и далее отдельный поток событий идет в DQA Manager сэмплер, который в соответствии с заданной скоростью отбора событий для анализа, заложенной в него и загружаемой из конфигурационного файла, отбрасывает некоторые из них. Конфигурация для работы DQM хранится в формате JSON в файле, содержащем всю необходимую информацию о сэмплере, производителях и анализаторах гистограмм, их базе данных и веб-интерфейсе. Описание конфигурации хранится в отдельных файлах с тем, чтобы каждая детекторная группа могла описывать необходимые гистограммы только в рамках своих задач. Настройки, не связанные с контролем данных для этой группы, не содержатся в соответствующем файле конфигурации. При запуске системы DQM загружаются все JSON-файлы с описанием всех контрольных гистограмм. Содержимое JSON-файлов соответствует заданной объектной модели. Из сэмплера события поступают в Histogram Producer, где в соответствии с загруженной конфигурацией гистограммы производятся и отправляются в Histogram Analyzer (анализатор). Поступающие в анализатор кумулятивные гистограммы добавляются в уже имеющиеся. Некумулятивные и обновленные кумулятивные гистограммы обрабатываются непосредственно в загруженных алгоритмах анализа в соответствии с заданной конфигурацией. Результатом анализа является качество данных со значением good, bad или undefined, причем при значении bad аварийное сообщение генерируется и отправляется на пользовательский веб-интерфейс (Web UI) DQA. Произведенные на конец сеанса кумулятивные и некумулятивные с результатом анализа bad гистограммы сохраняются в базе данных вместе с результатом анализа и источником поступивших данных. Эти данные поступают также на DQA WEB UI.

ОБЪЕКТНАЯ МОДЕЛЬ КОМПОНЕНТОВ СИСТЕМЫ

Диаграмма основных классов, предназначенных для построения и анализа контрольных гистограмм, представлена на рис. 2. Классы Sampler, Producer и Analyzer реализуют основные функции в запускаемых процессах. Sampler отбирает события и передает их набору экземпляров класса Producer для создания соответствующих гистограмм с использованием объектов класса BMNEvent. Producer создает гистограммы из поступающих событий, используя алгоритмы класса ProduceAlgorithm, загруженные в коллекцию produce_algorithms с ключом tree_path из библиотеки, прописанной в конфигурации. Созданные гистограммы поступают в объект класса Analyzer, для кумулятивных гистограмм они добавляются к имеющимся гистограммам коллекции cumulative_histograms, которые используются как входные данные к алгоритмам анализа, хранящимся в коллекции analyze_algorithms с ключом имя алгоритма. Результат работы алгоритма записывается в объект

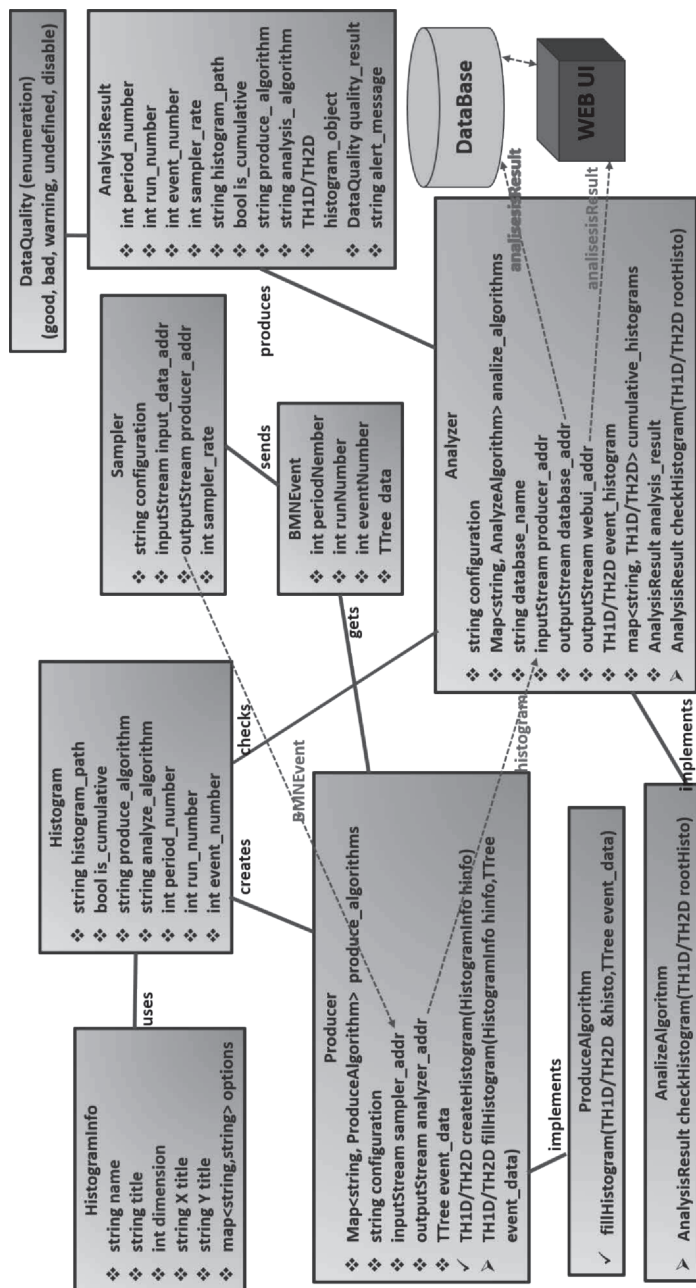


Рис. 2. Диаграмма основных классов создания и анализа гистограмм

класса `AnalysisResult`, содержащий информацию об источнике события, типе и данных гистограммы, алгоритме ее создания и обработки, а также результате анализа.

ЗАКЛЮЧЕНИЕ

Важность систем DQA для мониторинга данных физических событий подтверждается успешным применением таких систем, позволяющим оперативно реагировать на проблемы во время сеансов, например, в экспериментах на LHC в ЦЕРН. Представлены основные цели создания системы, общая архитектура и диаграмма классов ее компонентов. Затронуты вопросы гибкого подхода к загрузке компонентов системы в виде отдельных файлов с контекстом в формате JSON. В архитектуре учитываются потребности визуализации контрольных гистограмм на рабочем месте оператора смены, необходимость автоматического контроля качества данных и выдачи соответствующих сообщений и оповещений в графическом интерфейсе в случае ухудшения их качества или серьезного сбоя.

СПИСОК ЛИТЕРАТУРЫ

1. *Corso-Radu A., Hadavand H., Illchenko Y., Kolos S., Okawa H., Slagle K., Taffard A.* Data Quality Monitoring Framework for the ATLAS Experiment: Performance Achieved with Colliding Beams at the LHC // *J. Phys.: Conf. Ser.* 2011. V. 331. 022027.
2. *von Haller B., Konopka P.* The ALICE Data Quality Control // *Eur. Phys. J. Web Conf.* 2024. V. 295. 02026.
3. *Tuura L., Meyer A., Segoni I., Della Ricca G.* CMS Data Quality Monitoring: Systems and Experience // *J. Phys.: Conf. Ser.* 2010. V. 219. 072020.
4. *Adinolfi M., Archilli F., Baldini W., Baranov A., Derkach D., Panin A., Pearce A., Ustyuzhanin A.* LHCb Data Quality Monitoring // *J. Phys.: Conf. Ser.* 2017. V. 898. 092027.
5. *Baranov D., Kapishin M., Mamontova T., Pokatashkin G., Ruffanov I., Vasendina V., Zinchenko A.* The BM@N Experiment at JINR: Status and Physics Program // *KnE Energy Phys.* 2018. V. 3. P. 291–296.
6. *Trubnikov G. et al.* NICA Heavy-Ion Collider at JINR (Dubna). Status of Accelerator Complex and First Physics at NICA // *J. Phys.: Conf. Ser.* 2023. V. 2586. 012013.