

УДК 681.5.017

Научная статья

Особенности выбора архитектуры программного комплекса для организации сбора, обработки, анализа данных томской обсерватории радиоактивности и ионизирующих излучений

Г. А. Яковлев

Томский государственный университет, 634050, г. Томск, пр. Ленина, 36, Россия
E-mail: vsyakovleva@tpu.ru

В данной работе рассматривается планирование и реализация архитектуры программного комплекса для обеспечения оптимальной работы информационной системы физической обсерватории. Также описываются основные моменты, касающиеся выбора ключевых компонент комплекса, и их преимущества. Показан пример работы программного комплекса и его компонент, и сделаны выводы об архитектурных решениях и их полезность в прикладных задачах информатики по организации работы информационных систем обработки и хранения данных.

Ключевые слова: программный комплекс, большие данные, информационные системы, автоматизация процессов, обработка данных.

DOI: 10.26117/2079-6641-2020-33-4-150-156

Поступила в редакцию: 21.11.2020

В окончательном варианте: 17.12.2020

Для цитирования. Яковлев Г. А. Особенности выбора архитектуры программного комплекса для организации сбора, обработки, анализа данных томской обсерватории радиоактивности и ионизирующих излучений // *Вестник КРАУНЦ. Физ.-мат. науки.* 2020. Т. 33. № 4. С. 150-156. DOI: 10.26117/2079-6641-2020-33-4-150-156

Контент публикуется на условиях лицензии Creative Commons Attribution 4.0 International (<https://creativecommons.org/licenses/by/4.0/deed.ru>)

© Яковлев Г. А., 2020

Введение

Процесс исследования окружающего мира, его практическое познание никогда не заканчиваются. Наоборот, постоянно возникает большое количество научных проектов направленных на исследование естественных процессов и систем. Кроме сложных этапов анализа проблем, постановки эксперимента и подбора оборудования учёные рано или поздно придут к вопросу организации сбора, обработки и анализа полученных данных. Многие из таких проектов реализуются на базе уже имеющегося оборудования научных коллективов, которое могло быть использовано ранее, но также может быть заменено в ходе эксплуатации. Обычно для исследования сложных многопараметрических процессов требуется большое количество разнородного оборудования. Однако в большинстве случаев им присущи

Финансирование. Работа выполнялась без финансовой поддержки.

следующие недостатки, такие как: закрытость, отсутствие документации о работе внутренних компонентов оборудования и различных протоколов связи с приборами, отсутствие легко обнаружимой индикации об ошибках в их работе, а также ограниченность в использовании лишь программного обеспечения, предоставленного производителем.

Такая ситуация сложилась и в Томской обсерватории радиоактивности и ионизирующего излучения, где и было необходимо организовать работу с приборным парком и обеспечить возможность оперативного получения информации как о состоянии оборудования, так и непосредственно самих измерений. Для решения проблем томских ученых был разработан специальный программный комплекс, который по мере реализации адаптировался под задачи и совершенствовался для упрощения его дальнейшего обслуживания.

Описание программного комплекса

В качестве способа разработки была выбрана методика RAD [1], т.к. были предоставлены требования, реализация которых должна была обеспечить работу обсерватории и мониторинг в режиме near real time. Отсутствие автоматизации в работе с данными и необходимость реализации большого числа подзадач также привела к их решению, воплотившемуся в модули получившегося комплекса.

В рамках выбранной методики (и доступности ресурсов), разработка осуществлялась соответственно решаемым задачам, а программный комплекс был разделен на 3 компоненты – агенты, серверы, клиенты (рис. 1).

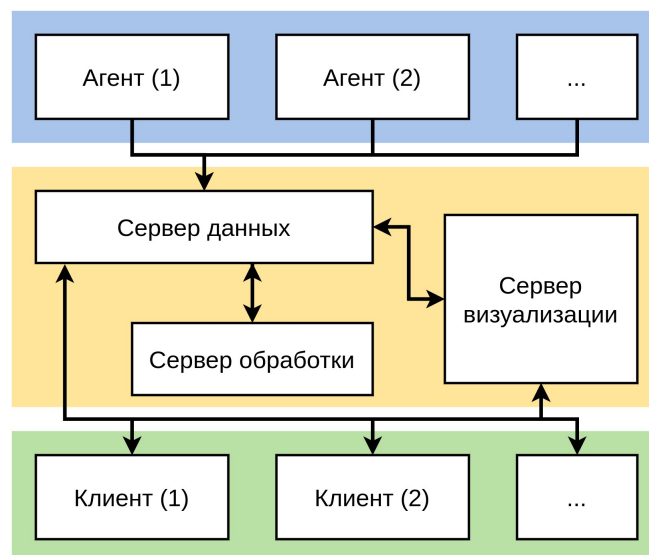


Рис. 1. Диаграмма архитектуры программного комплекса

Агенты — программы, установленные на компьютерах использующихся для сбора данных измерений в обсерватории. Они, в свою очередь, состоят из агента отправки данных мониторинга, агента проверки и наблюдения за работоспособностью детекторов и информирующего агента для отправки сообщений. Первый имел минимум функционала и собственную реализацию. Для реализации второго потребовалось выяснить как обнаруживать сбои в работе детектора, были написаны отдельные скрипты выполняющие данную функцию (для различного оборудования

они могут быть выполнены отдельными модулями). Так, например, заводская программа детекторов АТОМТЕХ записывала данные без внутренней буферизации сразу в файл после указанного интервала измерения, это позволило узнать с точностью до 1-го интервала измерения о неисправностях, для этого был написан специальный модуль, следящий за состоянием записи в файлы (рис. 2).

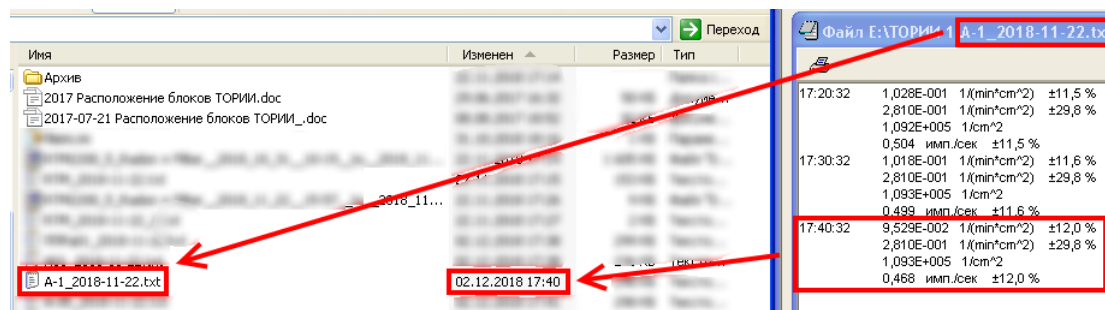


Рис. 2. Запись измерений в файл заводской программой

В дополнение был реализован функционал дублирования, для случаев недоступности головного сервера информирующий агент отправит сообщение о сбое на электронную почту, т.к. такая ситуация, зачастую, возможна и встречалась ранее (во время разработки и тестирования сервера).

Серверы – реализованы как независимые программы, которые могут работать на различных компьютерах. Они используются для организации и хранения данных мониторинга (в т.ч. проверки целостности данных), а также информации, посылаемой агентами (рис. 3).

Computer	D-11	D-12	D-13	D-14	D-15	D-16	D-17	D-18	Time difference	Last seen time
<input type="checkbox"/> PC1									5 seconds ahead	05.11.2019, 17:33:28
<input type="checkbox"/> PC2									1 seconds ahead	05.11.2019, 17:34:44
<input checked="" type="checkbox"/> PC3									11 seconds behind	05.11.2019, 17:31:50
<input type="checkbox"/> PC4									13 seconds behind	05.11.2019, 17:38:08
<input type="checkbox"/> PC5									3 seconds ahead	05.11.2019, 17:36:29

Рис. 3. Пример построения таблицы по данным агентов (по информации с сервера данных)

Наиболее гибким решением было использование отдельного сервера для обработки данных и для визуализации данных. Обработка рядов данных различных приборов производится с помощью простых скриптов, которые обрабатывают данные в каждом файле, объединяют данные за все периоды и выдают необходимые поля, которые уже сохраняются в широко распространенном формате CSV (TSV), пригодном для дальнейшего использования в любых средах для анализа данных.

Такая серверная архитектура дает возможность независимой работы отдельных компонентов, которые уже были реализованы и не требуют вмешательства в их работу, а также позволяет в дальнейшем строить распределенные вычислительные системы, т.к. система хранения отделена от вычислений. Процессы обработки данных разделены для различных типов данных и каждому формату соответствует свой алгоритм преобразования. К тому же это позволяет легко заменить или организовать дублирование серверов для стабильной работы всего комплекса

в целом. Использование скриптов позволяет свести к минимуму трудозатраты, т.к. требуется лишь однажды написать и протестировать программу. Поэтому была создана Low-code платформа [2], которая дает возможность параллельной обработки больших массивов данных. Такой же подход использовался для системы визуализации и экспорта, который заключался в описании формата обработанных данных, что позволяло более абстрактно работать с ними и упростить работу со всей системой для неспециалистов, без требования знаний о работе системы на более низком уровне.

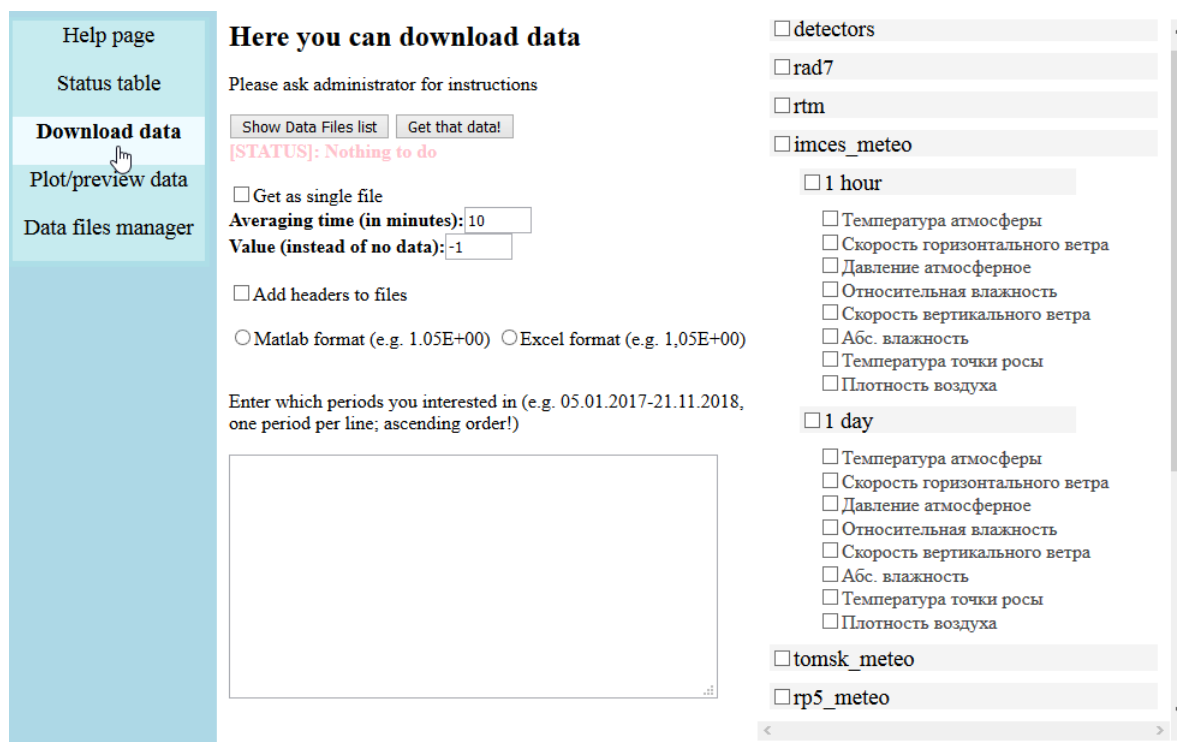


Рис. 4. Возможности экспорта

Пример построения древовидного списка с различными типами данных приведен на рис. 4 справа, а слева представлена форма для экспорта выбранных данных, где может быть применено усреднение или приведение данные к эквидистантному виду (что ускоряет процесс дальнейшей работы с ними).

Клиенты – в данном случае подразумеваются пользователи предоставленных серверами ресурсов. Они могут использовать данные (т.е. файловый клиент) или же обращаться к программе визуализации (т.е. клиент – браузер). В системе имеется web-сервер, который представляет агрегированный интерфейс ко всем функциям системы контроля. Предоставляемый web-интерфейс позволяет получить выбранные данные в указанном формате, для работы в других программах, когда они после этапов обработки будут доступны файловому клиенту (соответственно через сервер данных).

Одной из основных возможностей программного комплекса является возможность быстрого визуального анализа. За это отвечает сервер визуализации, который предоставляет клиентам доступ к данным. После различных итераций программы с использованием средств визуализации таких как Matlab, matplotlib, gnuplot, GNU Octave, как самое оптимальное было выбрано построение данных на стороне клиента.

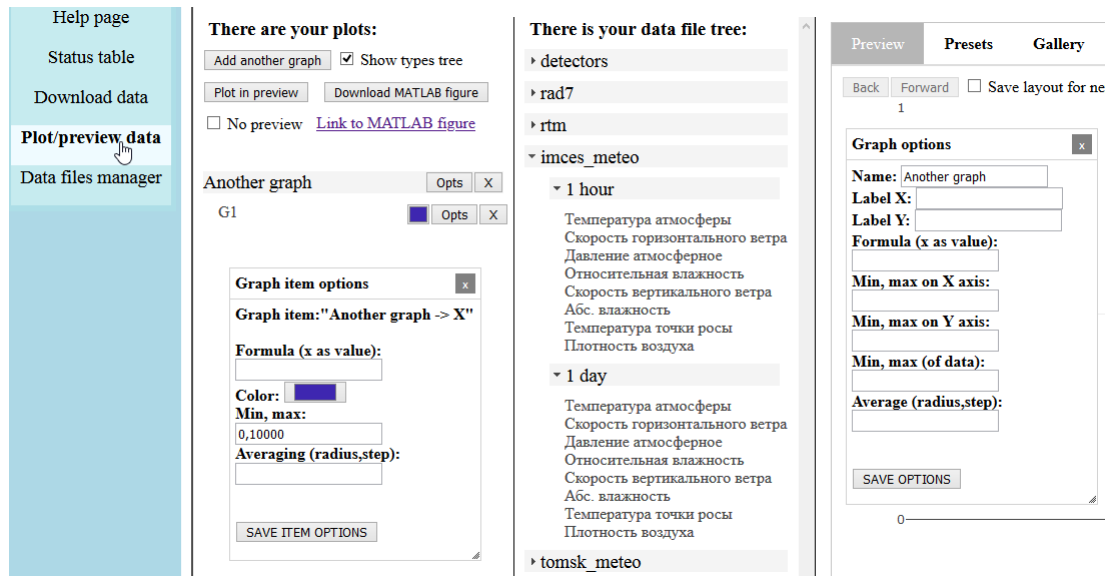


Рис. 5. Возможности визуализации

Для него были реализованы базовые функции работы с данными (см. рис.5, базовые возможности такие как: фильтрация ошибок, линейные преобразования значений, скользящее среднее, и другие параметры визуализации такие как цвета и подписи).

За саму визуализацию и динамическое изменение масштаба рассмотрения данных использована библиотека Plotly.js, которая позволяла работать с большими объемами данных на графиках, при этом в скорости работы не уступала Matlab (самому быстрому из серверных вариантов визуализации, но также не поддерживавшему динамическое изменение), и использовалась для просмотра данных мониторинга сразу, по мере их поступления в систему (см. рис.6).

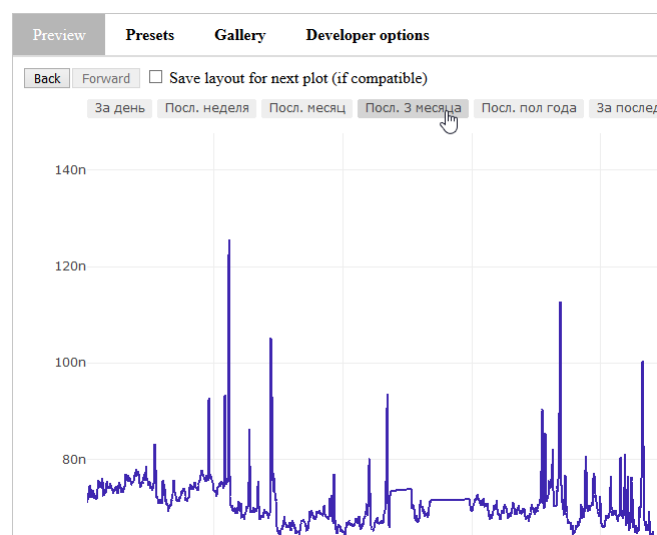


Рис. 6. Пример построения графика с возможностью выбора временного диапазона, для режима работы в near real time

Простота, удобство и относительное быстродействие позволяют сразу определить необходимые параметры фильтрации и усреднения, чтобы упростить работу

в сторонних средах. В дополнение была включена возможность построения и в формате Matlab, который позволит избежать лишних вычислений после установления всех других параметров построения (а также позволяет сохранить сами данные внутри бинарного файла).

Заключение

Описанный в данной работе программный комплекс был реализован и внедрен в Томской обсерватории радиоактивности и ионизирующего излучения, был испытан в различных условиях, как совместной параллельной работы всех компонент комплекса в режиме реального времени, так и независимой работы каждого модуля отдельно, сохраняя при этом всю необходимую функциональность. Выбранная архитектура позволит легко адаптироваться под возможное расширение приборного парка другим оборудованием. За счет своей модульности обслуживание и модернизация требуют минимальных затрат: скрипты легко заменять или реализовать на других языках программирования, а элементы пользовательского интерфейса и конфигурации доступны и понятны для любого сотрудника обсерватории. Дальнейший рост объемов данных и их разнородности не мешает объединить новые данные со старыми. В дополнение к этому открыта возможность перехода на распределенные системы как хранения, так и обработки информации для построения больших баз экспериментальных данных на благо ученым.

Конкурирующие интересы. Конфликтов интересов в отношении авторства и публикации нет.

Авторский вклад и ответственность. Автор участвовал в написании статьи и полностью несет ответственность за предоставление окончательной версии статьи в печать.

Список литература/References

- [1] Gerber A. Van Der Merwe A.; Alberts R., “Practical Implications of Rapid Development Methodologies”, *Proceedings of the Computer Science and Information technology Education Conference, CSITEd*, Computer Science and IT Education Conference, Mauritius, 2007, 233–245.
- [2] Яковлев Г. С., Иванов Ф. Ф., “Использование Low-code платформ в период перехода на процессный подход в создании автоматизированных систем”, *Вестник КРАУНЦ. Физ.-мат. науки*, 2020, № 1, 120–126. [Yakovlev G. S., Ivanov F. F., “Ispol’zovaniye Low-code platform v period perekhoda na protsesssnyy podkhod v sozdanii avtomatizirovannykh sistem”, *Vestnik KRAUNTS. Fiz.-mat. nauki*, 2020, № 1, 120–126 (in Russian)].

Features of the software complex architecture for collecting, processing and analysis of data in the Tomsk observatory of radioactivity and ionizing radiation

G. A. Yakovlev

Tomsk State University, 634050, Tomsk, Lenin Ave., 36, Russia

E-mail: vsyakovleva@tpu.ru

This paper considers the planning and implementation of the architecture of the software package to ensure the optimal operation of the information system of the physical observatory. It also describes the main points regarding the choice of key components of the complex, and their advantages. An example of the operation of the software complex and its components is shown, and conclusions are drawn about architectural solutions and their usefulness in applied problems of informatics for organizing the work of information systems for processing and storing data.

Key words: software package, big data, information systems, process automation, data processing.

DOI: 10.26117/2079-6641-2020-33-4-150-156

Original article submitted: 21.11.2020

Revision submitted: 17.12.2020

For citation. Yakovlev G. A. Features of the software complex architecture for collecting, processing and analysis of data in the Tomsk observatory of radioactivity and ionizing radiation. *Vestnik KRAUNC. Fiz.-mat. nauki.* 2020, **33**: 4, 150-156. DOI: 10.26117/2079-6641-2020-33-4-150-156

Competing interests. The author declare that there are no conflicts of interest regarding authorship and publication.

Contribution and Responsibility. The author contributed to this article. The author is solely responsible for providing the final version of the article in print. The final version of the manuscript was approved by the author.

The content is published under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/deed.ru>)

© Yakovlev G. A., 2020

Funding. The work was carried out without financial support.