

УДК 004.942

DOI 10.17513/snt.39892

ОСОБЕННОСТИ РЕАЛИЗАЦИИ СИСТЕМЫ МОНИТОРИНГА ГЕТЕРОГЕННОЙ ВЫЧИСЛИТЕЛЬНОЙ СРЕДЫ

Костромин Р.О.

*ФГБУН «Институт динамики систем и теории управления имени В.М. Матросова»
Сибирского отделения Российской академии наук, Иркутск, e-mail: kostromin@icc.ru*

В статье обсуждаются вопросы мониторинга гетерогенных распределенных вычислительных сред, включающих как выделенные ресурсы вычислительных кластеров, так и облачные ресурсы. Особенностью мониторинга кластерных систем является необходимость мониторинга не только вычислительных узлов, но и сетевых компонентов и инженерной инфраструктуры. Инженерная инфраструктура кластера, как правило, состоит из систем электроснабжения, вентиляции и охлаждения. Данные компоненты не всегда имеют техническую возможность предоставлять доступ к показателям работы по сетевым и цифровым протоколам. Кроме того, в системах мониторинга необходимо обеспечивать методы длительного, надежного и эффективного хранения накопленных данных с возможностью быстрого и гибкого доступа к ним. К сожалению, готового решения данных проблем не существует, что подчеркивает актуальность исследований, связанных с разработкой и внедрением специализированных систем мониторинга. В работе предложена архитектура и особенности реализации прототипа модульной системы мониторинга, предназначенной для сбора, обработки, визуализации метрик и автоматизированного мониторинга состояния гетерогенной вычислительной среды. Практические эксперименты подтвердили работоспособность прототипа и определили дальнейшие направления развития системы мониторинга, предполагающие использование методов искусственного интеллекта для обнаружения и прогнозирования сбоев в компонентах вычислительной среды.

Ключевые слова: системы мониторинга, вычислительный кластер, облачные вычисления, гетерогенные вычислительные среды

Исследование выполнено при поддержке Министерства науки и высшего образования Российской Федерации, проект № FWEW-2021-0005 «Технологии разработки и анализа предметно-ориентированных интеллектуальных систем группового управления в недетерминированных распределенных средах» (рег. № 121032400051-9).

ASPECTS OF IMPLEMENTING A MONITORING SYSTEM IN A HETEROGENEOUS COMPUTING ENVIRONMENT

Kostromin R.O.

*Matrosov Institute for System Dynamics and Control Theory of the Siberian Branch
of the Russian Academy of Sciences, Irkutsk, e-mail: kostromin@icc.ru*

The paper discusses the issues of monitoring heterogeneous distributed computing environments, including both dedicated computing cluster resources and cloud resources. A feature of monitoring cluster systems is the need to monitor not only computing nodes, but also network components and engineering infrastructure. The engineering infrastructure of the cluster, as a rule, consists of power supply, ventilation and cooling systems. These components do not always have the technical ability to provide access to performance indicators via network and digital protocols. In addition, monitoring systems must provide methods for long-term, reliable and efficient storage of accumulated data, with the ability to quickly and flexibly access them. Unfortunately, there is no ready-made solution to these problems, which emphasizes the relevance of research related to the development and implementation of specialized monitoring systems. The paper proposes the architecture and implementation features of a prototype modular monitoring system designed for collecting, processing, visualizing metrics and automated monitoring of the state of a heterogeneous computing environment. Practical experiments confirmed the functionality of the prototype and determined further directions for the development of the monitoring system, involving the use of artificial intelligence methods to detect and predict failures in components of the computing environment.

Keywords: monitoring systems, computational clusters, cloud computing, heterogeneous computing environments

The research was carried out with the support of the Ministry of Science and Higher Education of the Russian Federation, project No. FWEW-2021-0005 “Technologies of development and analysis of subject-oriented intellectual systems of group control in non-deterministic distributed environments” (Reg. No. 121032400051-9).

В современных научных исследованиях вычислительные кластеры играют большую роль, предоставляя высокопроизводительные ресурсы при решении сложных научных и инженерных задач [1, 2]. Однако для обеспечения эффективной и беспере-

бойной работы таких кластеров необходима специализированная система мониторинга [3], которая способна следить за различными параметрами их функционирования и своевременно предупреждать о нештатных ситуациях.

При этом мониторинг вычислительных кластеров выходит за рамки простого наблюдения, превращаясь в важную составляющую управления инженерной и вычислительной инфраструктурой. Эффективная система мониторинга позволяет оперативно выявлять потенциальные проблемы, оптимизировать распределение ресурсов, обеспечивать высокую производительность и стабильную работу кластеров за счет автоматического устранения части неисправностей.

В ходе вычислительных экспериментов возникает необходимость не только в использовании существующих вычислительных кластеров, но и в создании облачных (виртуализированных) [4]. Этот переход к облачным решениям обусловлен не только расширением требований к вычислительным мощностям, но и необходимостью обеспечения гибкости и масштабируемости вычислений. В связи с этим возникает неизбежная задача развертывания системы мониторинга, которая отвечала бы требованиям высокопроизводительных облачных кластеров. Выбор и тонкая адаптация системы мониторинга к имеющейся инфраструктуре становится важным компонентом успешной реализации облачной инфраструктуры, предоставляя не только полный контроль над состоянием кластера, но и обеспечивая инструменты для быстрого выявления и решения возможных проблем.

Существующие системы мониторинга не предоставляют готовых решений, которые были бы универсальными для развертывания в различных типах инфраструктур [3, 5]. Это ставит перед исследователями задачи не только инженерного, но и научно-технического характера. В число задач входит разработка критериев сбора и хранения метрик состояния кластера, а также алгоритмов для определения аномалий в текущем состоянии, прогнозирование и устранение неисправностей, анализ больших объемов данных, использование методов машинного обучения для предсказания возможных сбоев, а также разработку новых подходов к автоматизированному реагированию на изменения в инфраструктуре.

С целью решения озвученных выше проблем ведется разработка модульной системы мониторинга, ориентированной на специфику различных типов инфраструктур и требований к современным системам мониторинга. В статье представлена архитектура такой системы, обеспечивающей мониторинг облачных ресурсов, вычислительных кластеров и их инженерного оборудования.

Материалы и методы исследования

Характерной особенностью вычислительного кластера по сравнению с облачной средой является более локализованная и высокооптимизированная инфраструктура. В отличие от облачных вычислений, которые предоставляют вычислительные ресурсы через интернет, кластер может быть развернут внутри организации. Это может быть обусловлено рядом причин, в том числе и с позиций информационной безопасности. С другой стороны, кластеры являются менее гибкими и масштабируемыми по сравнению с облачными средами, где ресурсы могут моментально добавляться в зависимости от потребностей пользователей [4]. Другой важной чертой кластера является способ организации совместной работы через систему очередей заданий с использованием менеджера ресурсов, в то время как облачные ресурсы выделяются по запросу пользователя в монопольное пользование.

В ходе исследований методов интеграции облачных систем в гетерогенную вычислительную среду был предложен подход к представлению вычислительных заданий пользователям в виде расширенного описания, содержащего спецификацию и требования к контейнеру или виртуальной машине [6], запускаемых в облачной среде. Кроме того, ранее в работе [7] был представлен подход к классификации заданий пользователей для их распределения на более подходящие ресурсы. Таким образом была организована гетерогенная распределенная вычислительная среда, но без единой системы контроля за ее состоянием.

Как правило, мониторинг включает в себя широкий спектр операций, таких как наблюдение, анализ и поддержание работоспособности системы, процессов или предоставляемых услуг. Кроме того, он может охватывать мониторинг производительности, безопасности, доступности и других параметров. Набор этих параметров, а также их полнота определяются на основе экспертного опыта системных администраторов ресурсов и их технических возможностей по предоставлению информации о своем текущем состоянии.

В области мониторинга существует обширный набор различных технологий и программных систем, предназначенных для эффективного наблюдения, анализа и оценки аспектов наблюдаемых систем и процессов [8], таких как Zabbix, Nagios, Ganglia. Подобные в большей степени пригодны для мониторинга отдельных серверов, для кластерных систем существует более современная

система NetData [9], но и ее эффективность на крупных кластерных системах не нашла подтверждения, несмотря на обширный набор собираемых метрик. С целью решения проблемы мониторинга гетерогенной вычислительной среды предложена архитектура прототипа системы мониторинга (рис. 1), которая обеспечивает доступ к мониторингу как классических кластеров и их инфраструктуры, так и облачных ресурсов.

Пользователям предоставляется веб-ориентированный доступ для отображения текущего состояния вычислительной среды в виде таблиц, диаграмм, графиков и т.д. Кроме того, реализована поддержка интерфейса командной строки (CLI) в браузере для выполнения SQL-подобных запросов на языке PromQL для работы с необходимыми метриками. Для обеспечения доступа к состоянию вычислительной

среды внешним системам, например средствам управления вычислениями, реализован соответствующий API для поддержки как REST, так и SOAP сервисов. Через этот API представляются актуальные сведения о загрузке ресурсов, что позволяет повысить эффективность распределения потоков заданий между узлами среды. С другой стороны, API обеспечивает доступ к ретроспективным данным мониторинга для их применения в специализированных системах моделирования работы вычислительных сред.

Доступ к определенным видам метрик определяется политиками групп доступа через службу LDAP. За проверку доступа через LDAP и маршрутизацию запросов отвечает реверс-прокси сервер на базе Traefik. В изолированной сети Traefik осуществляет запросы к остальным подсистемам.

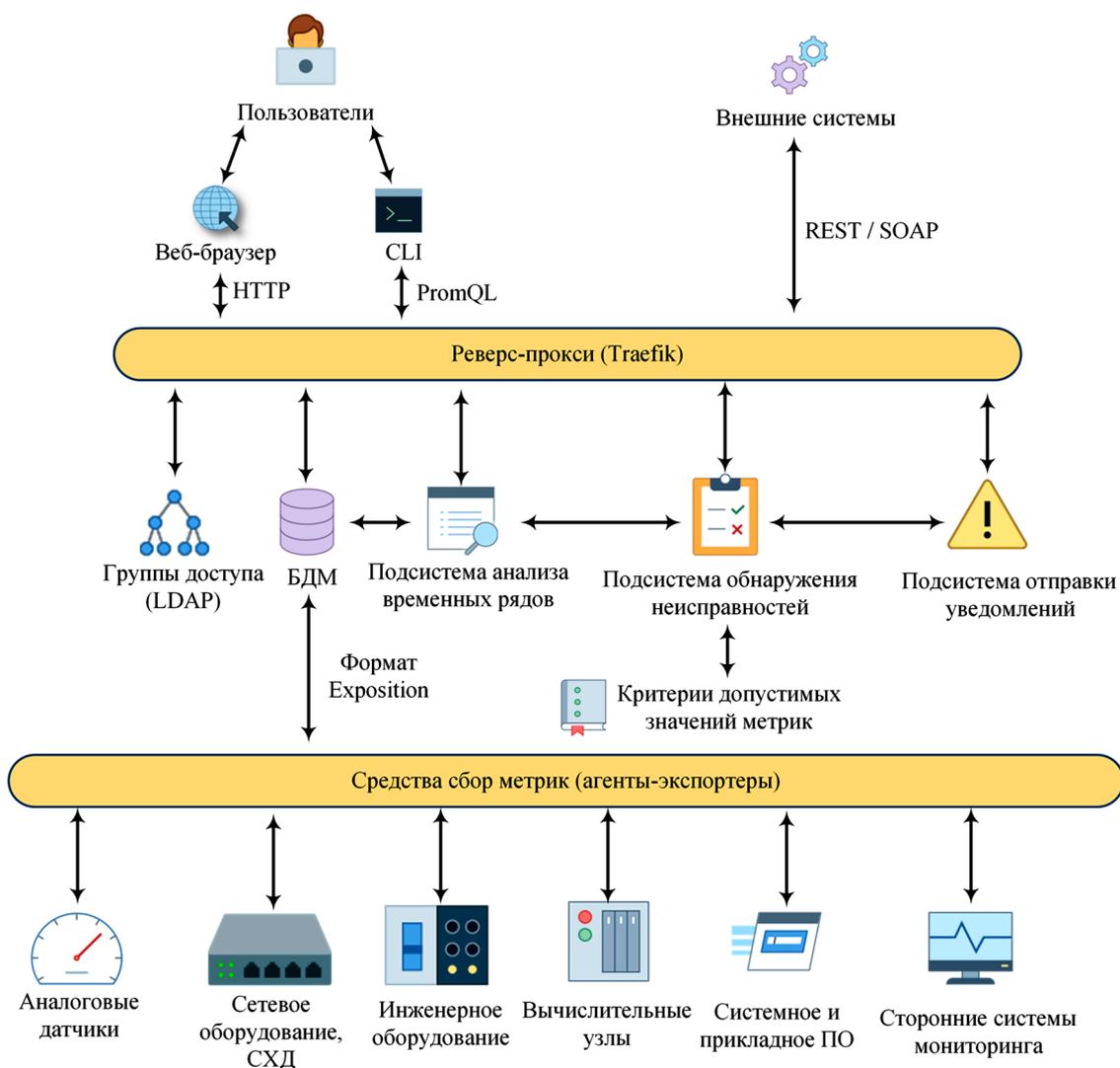


Рис. 1. Архитектура системы мониторинга

Эти подсистемы получают данные из базы данных метрик (БДМ) (реляционных, объектных и других специализированных баз данных). Формат хранения метрик определяется критериями БДМ и ориентирован на представление в виде временных рядов. В качестве БДМ выбрана система Prometheus, которая является в настоящее время индустриальным стандартом для сбора и хранения метрик [10]. Также Prometheus обеспечивает сжатие данных и быстрый доступ к ним на языке запросов PromQL.

В свою очередь за наполнение БДМ отвечает комплекс средств сбора метрик (агентов-экспортеров), которые выгружают специализированный набор метрик из вычислительной среды и конвертируют их в формат хранения БД метрик. Вычислительная среда включает в себя вычислительные и управляющие ресурсы (физические узлы и облачные), сетевую инфраструктуру (сервисные и управляющие сети, коммутаторы Ethernet и InfiniBand), системы хранения данных (СХД), инженерную инфраструктуру, системное и прикладное программное обеспечение (ПО), вычислительные задания пользователей.

При таком модульном подходе перво-степенной задачей является определение набора собираемых метрик и разработка (настройка) агентов-экспортеров. Агент-экспортер является отдельной службой, доступ к которой осуществляется через HTTP-запросы (API). Он может быть как частью системы, поддерживаемой БДМ, так и работающим независимо. Отдельный экспортер позволяет реализовать сбор данных с датчиков устройств (температуры, влажности, напряжения, оборотов вентиляторов и т.д.), работающих по специализированным протоколам (UART или через GPIO), в том числе и аналоговым. В качестве платформы для сбора данных может быть использован

микрокомпьютер [11], например Raspberry Pi или Arduino. Аналогично мониторингу аппаратных средств, с помощью экспортеров выполняется сбор данных о приложениях, контейнерах и виртуальных машинах.

Каждый экспортер зарегистрирован в БДМ для обеспечения сбора данных из него. Регистрация экспортеров возможна в автоматическом режиме, поэтому такой подход применим как для мониторинга статичных сущностей кластера, так и ресурсов облачной среды. В нашей системе мониторинга реализуется сбор метрик с вычислительных узлов, контейнеров и виртуальных машин с помощью Node Exporter [5, 9] и NetData. Они обеспечивают гибкую конфигурацию набора метрик, собираемых с узлов (загрузка и частота процессора, загрузка оперативной памяти, загрузка скорости сети и дисков, сетевые задержки, промахи кэша и т.д.) и их гранулярность. Сенсоры системы электропитания, вентилирования и охлаждения кластера имеют собственные протоколы и текстовые форматы представления данных, для них реализованы соответствующие экспортеры на языке Node.js. Все экспортеры по запросу БДМ отправляют данные в формате Exposition (Листинг).

В Листинге представлены следующие поля источника бесперебойного питания (ИБП):

- ups_battery_charge – гистограмма, представляющая уровень заряда батареи ИБП;
- ups_input_voltage – гистограмма, представляющая входное напряжение ИБП;
- ups_output_voltage – гистограмма, представляющая выходное напряжение ИБП;
- ups_load_percentage – гистограмма, представляющая загрузку ИБП в процентах от его емкости;
- ups_status – гистограмма, представляющая операционный статус ИБП (0 – ОК, 1 – Предупреждение, 2 – Критическое состояние).

```
# HELP ups_battery_charge Battery charge level of the UPS
# TYPE ups_battery_charge gauge
ups_battery_charge{ups_id="ups1"} 92.5
# HELP ups_input_voltage Input voltage provided to the UPS
# TYPE ups_input_voltage gauge
ups_input_voltage{ups_id="ups1"} 230.5
# HELP ups_output_voltage Output voltage provided by the UPS
# TYPE ups_output_voltage gauge
ups_output_voltage{ups_id="ups1"} 220.0
# HELP ups_load_percentage UPS load as a percentage of its capacity
# TYPE ups_load_percentage gauge
ups_load_percentage{ups_id="ups1"} 35.7
# HELP ups_status UPS operational status (0 - OK, 1 - Warning, 2 - Critical)
# TYPE ups_status gauge
ups_status{ups_id="ups1"} 0
```

Листинг. Пример данных в формате Exposition от системы электропитания

Каждая метрика имеет уникальный идентификатор (`ups_id`), который может использоваться для идентификации конкретных ИБП в системе. Эти метрики могут быть использованы БДМ для мониторинга и анализа работы системы бесперебойного электропитания.

Следующим компонентом системы мониторинга является комплекс визуализации метрик и отправки уведомлений администратору вычислительной системы. В нашей системе мониторинга используется система Graphana [9], которая наряду с Prometheus является стандартизированным решением [10]. Graphana имеет встроенную поддержку метрик из Prometheus, что позволяет использовать единую точку мониторинга вычислительной инфраструктуры в виде приборной панели. Приборная панель является инструментом для группировки метрик и настройки способа их представления (тепловые карты, гистограммы, шкалы и т.д.). Для каждой задачи визуализации (общее состояние кластера, состояние конкретного узла, загрузка сети, состояние инженерной инфраструктуры) администратор создает отдельную приборную панель. Кроме того, в Graphana выполняется формирование набора наблюдаемых метрик и их пороговых значений, конфигурирование методов отправки оповещений через e-mail, SMS и системы мгновенного обмена сообщениями.

Отличительной особенностью Graphana от других подобных систем является поддержка создания сложного набора правил для определения неисправностей. Система создания правил является достаточно гибкой и позволяет описывать нестандартные ситуации, которые могут привести к сбою. В правилах имеется поддержка описания

зависимостей между группами метрик. Однако создание таких правил требует экспертных знаний системных администраторов. В качестве инструментального средства для расширенного анализа метрик (выполнение запросов и фильтрация на языке PromQL, применение алгоритмов анализа временных рядов) системными администраторами разработан соответствующий JupierNotebook на языке Python. Из JupierNotebook выполняется вызов сервисов и библиотек [12, 13] для многокритериального анализа временных рядов, получаемых в результате выполнения запросов на языке PromQL из БДМ.

Результаты исследования и их обсуждение

На основе выбранных технологий и экспертного опыта администрирования вычислительных кластеров, а также с учетом требований новых вычислительных задач был разработан прототип системы мониторинга гетерогенной вычислительной (облачных и кластерных ресурсов) и инженерной инфраструктур. В процессе интеграции существующих инженерных систем разработаны экспортеры для сбора наиболее важных метрик их состояния. Данные экспортеры были зарегистрированы в БДМ и полученные метрики были успешно визуализированы с помощью Graphana. Также проведена настройка приборной панели мониторинга администратора, что обеспечило единую точку мониторинга ключевых компонентов вычислительной среды (рис. 2). Проведено тестирование системы оповещений и отправки уведомлений, что в целом подтверждает работоспособность прототипа и позволяет продолжить его разработку.

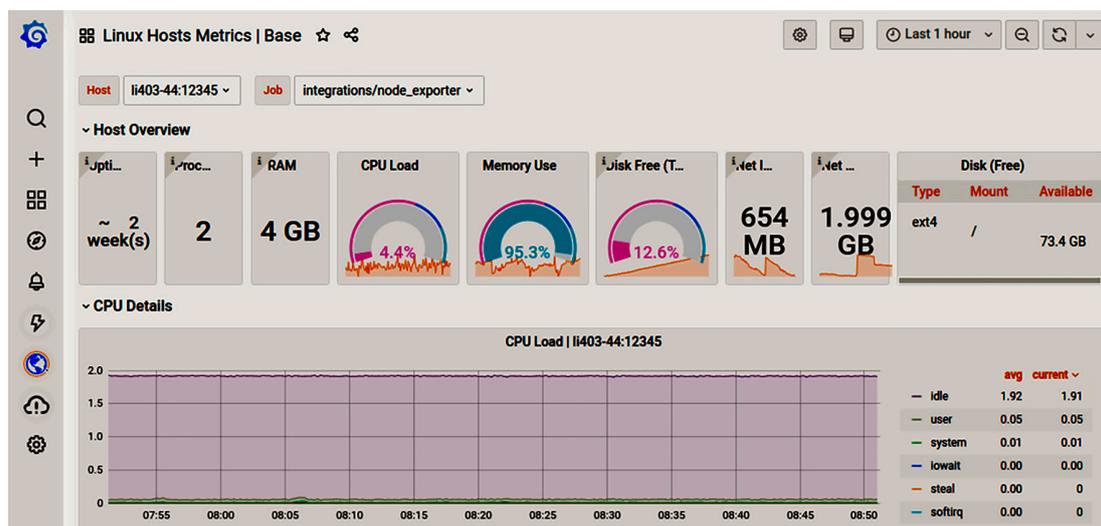


Рис. 2. Интерфейс приборной панели системы Graphana

Заключение

В статье представлены подход и архитектура разрабатываемой системы мониторинга гетерогенной вычислительной среды. Представлены основные компоненты и особенности их реализации. В настоящее время ведется активная разработка специализированных экспортеров и их интеграция с БДМ. Кроме того, продолжают исследования, связанные с применением методов искусственного интеллекта для обнаружения аномалий во временных рядах метрик, а также методов обнаружения и исправления сбоев. В качестве обучающей выборки для нейронной сети выступают ретроспективные данные мониторинга, хранимые в БДМ. На основе этих данных планируется выполнить разметку с указанием нештатных ситуаций. Предполагается, что методы искусственного интеллекта позволят обнаруживать корреляции и взаимосвязи между временными рядами метрик и смогут прогнозировать аварийные ситуации заранее, что даст возможность провести профилактические мероприятия. Как следствие, это увеличит эффективность функционирования вычислительной среды, ее отказоустойчивость и надежность.

Прототип новой системы мониторинга функционирует параллельно с уже существующей системой мониторинга кластера для обеспечения преемственности разработок. Этот подход позволяет поэтапно внедрять и проверять новые возможности, снижая воздействие на текущую эксплуатацию кластера.

Список литературы

1. Collier N., Wozniak J.M., Stevens A., Babuji Y., Binois M., Fadikar A., Würth A., Chard K., Ozik J. Developing Distributed High-performance Computing Capabilities of an Open Science Platform for Robust Epidemic Analysis // Proceeding of the IEEE International Parallel and Distributed Processing Symposium Workshops (St. Petersburg, FL, 15–19 May, 2023). IEEE, 2023. P. 868–877. DOI: 10.1109/IPDPSW59300.2023.00143.
2. Li Z. Geospatial Big Data Handling with High Performance Computing: Current Approaches and Future Directions // High Performance Computing for Geospatial Applications. Geotechnologies and the Environment. 2020. Vol. 23. P. 53–76. DOI: 10.1007/978-3-030-47998-5_4.
3. Voevodin V.V., Antonov A.S., Nikitenko D.A., Shvets P.A., Sobolev S.I., Sidorov I.Yu., Stefanov K.S., Voevodin V.V., Zhumatiy S.A. Supercomputer Lomonosov-2: Large Scale, Deep Monitoring and Fine Analytics for the User Community // Supercomputing Frontiers and Innovations. 2019. Vol. 6, Is. 2. P. 4–11. DOI: h10.14529/jsfi190201.
4. Beltre A.M., Saha P., Govindaraju M., Younge A., Grant R.E. Enabling HPC Workloads on Cloud Infrastructure Using Kubernetes Container Orchestration Mechanisms // Proceeding of the IEEE/ACM International Workshop on Containers and New Orchestration Paradigms for Isolated Environments in HPC (Denver, Colorado, November 18, 2019). IEEE, 2019. P. 11–20. DOI: 10.1109/CANOPIE-HPC49598.2019.00007.
5. Stefanov K.S., Pawar S., Ranjan A., Wandhekar S., Voevodin V.V. A Review of Supercomputer Performance Monitoring Systems // Supercomputing Frontiers and Innovations. 2021. Vol. 8, Is. 3. P. 62–81. DOI: 10.14529/jsfi210304.
6. Черных А.Н., Бычков И.В., Феоктистов А.Г., Горский С.А., Сидоров И.А., Костромин Р.О., Еделев А.В., Зоркальцев В.И., Аветисян А.И. Смягчение неопределенности при разработке научных приложений в интегрированной среде // Труды Института системного программирования РАН. 2021. Т. 33. № 1. С. 151–172. DOI: 10.15514/ISPRAS-2021-33(1)-11.
7. Feoktistov A., Tchernykh A., Dmitriev V. Mitigating uncertainty of resource allocation in heterogeneous computing environment // Информационные, вычислительные и управляющие системы для распределенных сред: материалы IV Международного семинара (Иркутск, 4–8 июля 2022 г.). Иркутск: Издательство ИДСТУ СО РАН, 2022. P. 74–78.
8. Сидоров И.А., Кузьмин В.Р. Обзор современных средств для комплексного мониторинга высокопроизводительных вычислительных систем // Фундаментальные исследования. 2016. № 1–9. С. 62–67.
9. Verginadis Y. A Review of Monitoring Probes for Cloud Computing Continuum // Lecture Notes in Networks and Systems. 2023. Vol. 655. P. 631–643. DOI: 10.1007/978-3-031-28694-0_59.
10. Bastos J., Araujo P. Hands-On Infrastructure Monitoring with Prometheus // Packt Publishing, 2019. 430 p.
11. Sidorov I.A., Kostromin R.O., Feoktistov A.G. System for monitoring parameters of functioning infrastructure objects and their external environment // Proceedings of the 2nd International Workshop on Information, Computation, and Control Systems for Distributed Environments (Иркутск, 6–7 июля 2020 г.). CEUR Workshop Proceedings, 2020. Vol. 2638. P. 252–264. DOI: 10.47350/ICCS-DE.2020.23.
12. Гришко А.К., Кочегаров И.И., Лысенко А.В. Многокритериальный выбор оптимального варианта сложной технической системы на основе интервального анализа слабоструктурированной информации // Измерение. Мониторинг. Управление. Контроль. 2017. № 3 (21). DOI: 10.21685/2307-5538-2017-3-14.
13. Феоктистов А.Г., Костромин Р.О., Башарина О.Ю. Сервис анализа временных рядов природно-климатических показателей окружающей среды инфраструктурных объектов Байкальской природной территории // Оптика атмосферы и океана. Физика атмосферы: материалы XXVII Международного симпозиума (Москва, 5–9 июля 2021 г.). Томск: Издательство ИОА СО РАН, 2021. Т. F. С. 96–99.