УДК 539.12

DOI:10.31854/1813-324X-2020-6-3-100-106

Прогнозирование ресурсов облачных сервисов на основе мониторинговой системы с открытым кодом

К.Н. Кучерова1*

¹Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, 195251, Российская Федерация

Информация о статье

Поступила в редакцию 05.08.2020 Принята к публикации 22.09.2020

Ссылка для цитирования: Кучерова К.Н. Прогнозирование ресурсов облачных сервисов на основе мониторинговой системы с открытым кодом // Труды учебных заведений связи. 2020. Т. 6. № 3. С. 100-106. DOI:10.31854/1813-324X-2020-6-3-100-106

Аннотация: В статье описан универсальный подход к мониторингу хранилищ данных глобально распределенных вычислительных комплексов, что позволяет автоматизировать создание новых метрик в системе и прогнозировать их поведение для конечного пользователя. Так как существующие мониторинговые программные продукты обеспечивают готовую схему для мониторинга только системных метрик, таких как использование оперативной памяти, процессора, внешних дисков и сетевого траффика, но не предлагают решения для бизнес-функций, то IT-компаниям приходится проектировать специализированные структуры баз данных. Предложенная в статье структура данных для хранения мониторинговой статистической информации универсальна и позволяет экономить ресурсы при организации мониторинга баз данных в масштабе глобально распределенных вычислительных комплексов. Целью работы является разработка универсальной модели для мониторинга и прогнозирования хранилищ данных глобально распределенных вычислительных комплексов и оценка ее адекватности реальным условиям эксплуатации.

Ключевые слова: облачная система, база данных, мониторинг, системная метрика, прогнозирование, корреляционная функция.

Введение

Как известно, контролировать системные метрики нужно для того, чтобы быстро реагировать на технические проблемы в системе, например, на отсутствие свободного места на диске, или перегруженность центального процессора (также, центрального процессорного устройства – ЦПУ), и тем самым поддерживать высокий уровень доступности облачных сервисов для конечных пользователей. Однако далеко не все понимают значимость мониторинга бизнес-метрик, которые позволяют обнаружить проблему в бизнес-логике сервисов, в то время как с технической точки зрения встроенные системные метрики будут показывать, что с системой все в порядке.

Например, такие бизнес-метрики, как «Количество Интернет-соединений клиентов» и «Количество транзакций в минуту», могут выявить неожиданное сочетание большого количества пользователей и подозрительно маленького количества

операций, что, скорее всего, будет указывать на неработоспособность сервисов у большей части клиентов. Эту проблему и подобные ей невозможно обнаружить с помощью стандартных системных метрик, а только с помощью специализированных бизнес-метрик.

Правильно подобранные бизнес-метрики позволяют обнаружить ошибки в логике работы вебприложения или, наоборот, указать на удачные решения, когда потребности конечных пользователей растут после выпуска новой версии программного продукта.

В данной статье предложен универсальный подход к мониторингу баз данных (БД) в составе глобально распределенных вычислительных комплексов (ГРВК), который позволяет автоматизировать добавление новых бизнес-метрик без изменения структуры данных и написания новых процедур взаимодействия между БД и системой мониторинга.

^{*}Адрес для переписки: kristina.mylife@gmail.com

Анализ существующих подходов и постановка задачи мониторинга ГРВК

Объектом исследования является международная телекоммуникационная компания Distillery, США [1], где в качестве БД используется MS SQL Server 2016 Enterprise Edition. Предметом исследования является система мониторинга с открытым исходным кодом Zabbix 3.0 [2], позволяющая вносить изменения в программный продукт без ущерба последующей миграции на новые версии. Но предложенный универсальный подход может быть использован для любой другой системы мониторинга и БД реляционного типа.

Существующий подход к созданию новых метрик в Zabbix и других аналогичных мониторинговых системах имеет крупный недостаток. Он заключается в отсутствии автоматизации и необходимости ручного создания специальных таблиц БД для хранения описания метрик и разработке SQL-запросов либо хранимых процедур для расчета их значений. При таком подходе каждое добавление новой метрики связано с модификацией структуры БД и написанием специализированных SQL-запросов, а также интеграцией с существующей системой мониторинга.

Заказчиком и потребителем системных метрик обычно являются только технические отделы компании, и их мониторинг используется для проверки доступности и работоспособности всей системы в целом в режиме 24/7 [3]. В случае с мониторингом бизнес-метрик ситуация другая, т. к. их наблюдают и анализируют другие специалисты компании. Это накладывает дополнительные требования к задаче, в том числе необходимость более гибко и быстро добавлять в систему новые бизнес-метрики по требованию отдельных групп заказчиков. Например, аналитический отдел может попросить в дополнение к существующей метрике количества транзакций в день добавить новое распределение по времени суток, отдельным заказчикам, регионам, типам операций и другим параметрам.

Системный мониторинг является стандартизированным, т. к. все системы используют серверы похожей архитектуры, где есть одинаковые компоненты: дисковое пространство, оперативная память, быстродействие и количество ядер ЦПУ, в том числе распространенные метрики, которые можно получать из операционной системы (ОС). По этой причине системный мониторинг поставляется в стандартном пакете, причем с использованием встроенных шаблонов ОС в случае Zabbix, и достаточно просто установить систему мониторинга и получить базовые системные метрики автоматически без дополнительной настройки [4]. С бизнес-метриками все иначе, поскольку все бизнес-системы отличаются друг от друга по архитектуре и реализации, нет никаких общих стандартов для получения метрик внешними системами. Кроме того, методы расчета бизнес-метрик могут отличаться от сервиса к сервису, поэтому на данном этапе развития IT-индустрии и стандартов шаблонный мониторинг бизнес-метрик не кажется реализуемым.

Идеальным подходом для бизнес-мониторинга была бы возможность добавлять новые метрики без необходимости внесения изменений в саму систему, а именно через настройку мониторинга без модификации структуры БД. К сожалению, в существующих мониторинговых решениях такая возможность отсутствует.

Интересным примером гибкой реализации подхода к бизнес-мониторингу является система AppDynamics [5], которая позволяет добавлять нужные метрики по бизнес-транзакциям. Т. к. AppDynamics осуществляет мониторинг веб-приложений, то в системе можно настроить сбор метрик по событию и параметрам приложения и таким образом создавать необходимые срезы для просмотра состояния бизнес-транзакций без внесения изменений в код приложения, а также без необходимости дорабатывать интеграцию с системой мониторинга. Тем не менее, продукт AppDynamics не является полностью универсальным решением для всех типов веб-приложений, а затраты на его использование трудно согласовать между заказчиком и исполнителем.

Многие IT-компании признают необходимость отслеживания бизнес-метрик и описывают их преимущества, но сами не предоставляют готовую схему реализации. Так, система PagerDuty [6] является ведущим решением по оповещению и отслеживанию инцидентов в реальном режиме времени 24/7 с привлечением мобильных устройств обслуживающего персонала. Но предоставление информационных услуг ограничивается веб-интерфейсом и его связью с любыми типами компьютеров клиентов, а бизнес и прочие метрики наблюдения и оценки работоспособности систем остаются на совести конечных пользователей.

Популярное решение для мониторинга бизнесметрик завоевывает Prometheus [7], т. к. имеет высокую производительность и масштабируемость, а также готовые шаблоны для мониторинга системных метрик. Норвежская компания FINN рассказывает о своем опыте применения Prometheus для мониторинга бизнес-метрик [8], из которого видно, что несмотря на удобства настройки метрик, сам бизнес-мониторинг требует реализации и адаптации для каждой компании.

На российском рынке телекоммуникаций и технологий Internet of Things (IoT) лидером мониторинга бизнес-метрик является разработка компании Saymon [9]. Ее преимущество состоит в возможности конфигурации объекта и выборки данных из других систем мониторинга, создании

экранов с консолидированной информацией и настройки связей между объектами, что позволяет получить более полную картину изменения статусной модели объекта в системе с течением времени.

Таким образом, ни одна из существующих мониторинговых систем не является универсальной с точки зрения контроля бизнес-метрик, которые в разных ІТ-компаниях имеют свою бизнес-логику сбора и анализа данных. Для дальнейшего решения поставленной задачи необходимо разработать информационную модель в виде универсальной структуры данных БД для хранения любых типов бизнес-метрик, провести анализ ее адекватности условиям эксплуатации ГРВК и определить область применения для мониторинга ІТ-объектов.

Описание мониторинга системных метрик

Универсальная система мониторинга системных и бизнес-метрик разработана и внедрена в международной компании Distillery, США [1], которая входит в аналитический отчет Gartner [10] в своем секторе по синхронизации файловых ресурсов (рисунок 1), имеет несколько миллиардов файлов на обслуживании и корпоративных клиентов с численностью более 300 тыс. человек в США и Европе.



Рис. 1. Квадрант Гартнер по корпоративным системам обмена и синхронизации файлов (июнь 2018)

Fig. 1. Gartner Magic Quadrant for Content Collaboration Platforms (as of June 2018)

Актуальность разработки вызвана возросшими требованиями корпоративных клиентов к показателям доступности и работоспособности сервисов. Жесткое соглашение Service Level Agreement (SLA) предполагает доступность сервисов для клиентов 99,99 % в режиме 24/7, в противном случае в ре-

зультате каждого отказа в обслуживании компания вынуждена выплачивать своим контрагентам неустойки.

Важным условием новой разработки является ее интеграция с существующим мониторингом системных метрик Zabbix, который допускает изменение исходного кода и модификацию структуры БД. Для этого использованы технологии Microsoft, серверы С# .net, БД MS SQL, 14 активных серверов и 14 – в режиме зеркалирования данных. Постоянная нагрузка в пиковые часы составляет порядка 400 тыс. запросов в минуту.

На рисунке 2 приведен пример системной метрики мониторинга свободного места на дисках серверов БД [11], а на рисунке 3 показана логика соблюдения соотношения исторического периода и горизонта прогнозирования для своевременного реагирования на окончание свободного места на диске [12].

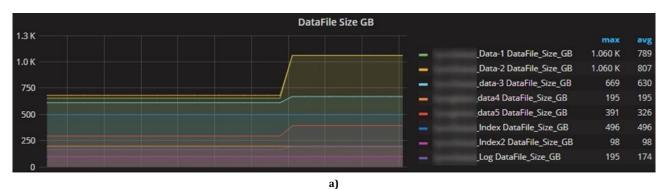
Разработка универсальной структуры БД мониторинга бизнес-метрик

Список бизнес-метрик ГРВК, которые необходимо хранить в БД и добавить в мониторинг в дополнение к стандартным системным метрикам, представлен в таблице 1.

ТАБЛИЦА 1. Список бизнес-метрик мониторинга ГРВК

TABLE 1. The List of Business Metrics for Monitoring Globally Distributed Cloud Computing System

Nº π/π	Название	Ед. изм.	Нижний порог срабатывания	Верхний порог срабатывания
1.	Количество загруженных файлов	%	1	20
2.	Количество удаленных файлов	%	5	30
3.	Всего файлов в системе	шт.	-	1
4.	Всего папок и файлов с распре- деленным доступом	шт.	-	ı
5.	Количество компаний-клиентов	шт.	-	-
6.	Количество компаний-клиентов с несколькими пользователями	%	50	-
7.	Общее количество уникальных пользо- вателей с корпора- тивной подпиской	%	40	-
8.	Общее количество папок в системе	шт.	-	-
9.	Количество пользователей в разрезе по серверам	шт.	-	-
10.	Количество папок по серверам	шт.	-	-



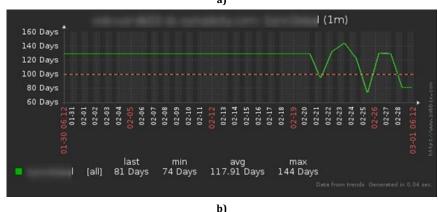
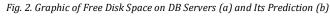


Рис. 2. График использования свободного места на дисках серверов (a) БД и их прогноз (b)



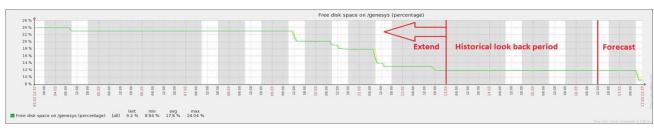


Рис. 3. Оценка соотношения исторического периода и горизонта прогнозирования

Fig. 3. Evaluation of Relationship between Historical Period and Forecast Horizon

Бизнес-метрики «Количество компаний-клиентов» и «Количество компаний-клиентов с несколькими пользователями» кажутся одинаковыми, но крайне важно собирать различные сведения по такому параметру, как количество пользователей. Если пользователь зарегистрировался, но потом не пользоватся сервисом, тогда их фактическое количество не соответствует формальному, и нагрузка на систему будет меньше. Для реализации мониторинга бизнес-метрик таблицы 1 создана универсальная структура из четырех таблиц БД.

1) MetricDescription – таблица со списком метрик, которые необходимо собирать.

Структура таблицы:

MetricDescriptionId – идентификатор описания; MetricShortName – сокращенное название, используемое для обозначения метрики в процедурах сбора метрик;

MetricUserName – название, которое нужно отображать пользователю;

MetricDescription – описание бизнес-смысла метрики;

MetricGroupByColumn – поле группировки, но для метрик, состоящих из одного параметра, оно не задается;

MetricDBSource – БД-источник данных по метрике;

MetricSelectType – тип выборки либо выборка по расписанию, либо метрика, которая собирается в связи с каким-то событием в системе;

MetricOrder – порядок сортировки метрики.

2) MetricData – таблица для хранения данных по собранным метрикам.

Поля таблицы:

MetricDataId – идентификатор строки;

MetricServerId – идентификатор сервера, на котором была собрана метрика;

MetricGroupId – идентификатор группы, если метрика групповая;

MetricDescriptionId – ссылка на идентификатор описания метрики;

MetricGroupByColumn – значения идентификатора группирующего поля (заполнено, только если в MetricDescription задан MetricGroupByColumn);

MetricGroupByColumnDescription – название группирующего поля (заполнено, только если в MetricDescription задан MetricGroupByColumn);

MetricDataBigInt – значения метрики; DateCreatedUTC – дата сбора метрики; OrderData – порядок сортировки.

Также в структуре БД есть две таблицы для настройки отчета по значениям метрик:

- 3) MetricReport заголовок отчета;
- 4) MetricReportContent список метрик, которые входят в отчет.

Сбор метрик осуществляется по расписанию, которое задает администратор БД. В соответствии с этим расписанием вызывается процедура MetricsCalculateAll, в которой идут обращения к базам данных, указанным в MetricDBSource. Для каждой БД создана специальная процедура spMetricsCalculate, в которую передается список собираемых метрик. Внутри процедуры осуществляется подсчет только тех метрик, которые включены в список. Для определения наличия метрики в списке используется ShortName, так как это более читабельно, чем просто идентификатор метрики, что позволяет уменьшить количество ошибок.

Если метрика есть в списке, она рассчитывается заранее определенным способом, и результаты

складываются во временную таблицу БД, которая возвращается в качестве результата в конце процедуры spMetricsCalculate, и таким образом результаты передаются в вызывающую процедуру MetricsCalculateAll. Когда результаты процедуры приходят в БД для хранения метрик, они записываются в таблицу MetricData.

Затем из мониторинговой системы Zabbix периодически вызывается выдающая данные по метрикам и их краткое описание в систему мониторинга процедура MetricsGetForMonitoring. Далее в системе мониторинга можно настроить необходимый вариант отображения метрик в виде графика или таблицы.

Практическая реализация системы мониторинга бизнес-метрик

На рисунке 4 приведен пример мониторинга бизнес-метрик «Количество загруженных файлов» в процентах от «Всего файлов в системе». Для удобства визуализации полученный процент умножается на 1 000 000. В результате на графике отмечено падение загрузки файлов в систему, в то время как все системные метрики были в норме. Как видно из графика, такое количество загруженных файлов не характерно для этого периода суток и свидетельствует о какой-то возможной аномалии в системе. Позднее было выявлено, что это произошло из-за ошибки в логике приложения.

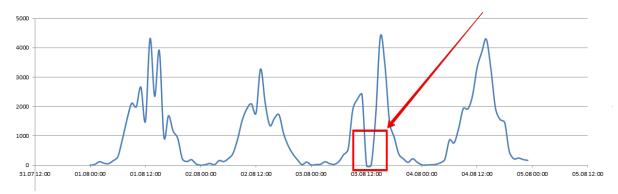


Рис. 4. Пример обнаружения аномалии в бизнес-метрике «Количество загруженных файлов, %»

Fig. 4. Example of an Anomaly Detection in the Business Metric "Uploaded Files, %"

Описанный подход к мониторингу бизнес-метрик реализован на практике в международной компании Distillery, США [1]. Это позволило добавить в мониторинг новые бизнес-метрики и упростило обнаружение и анализ причин появления аномальных ситуаций. В то же время такой мониторинг активно используется бизнес-заказчиками компании, для того чтобы анализировать изменения в активности пользователей и успешность новых продуктов компании.

Универсальность подхода к мониторингу ГРВК и расчету метрик позволяет добавлять новые бизнес-метрики без модификации структуры БД, а только через внесение необходимой настройки в

таблицу описания метрик и добавление SQL-запроса в процедуру сбора метрик. Все остальные элементы системы не требуют доработки.

Внедрение данного подхода позволяет повысить быстроту и надежность внесения изменений в систему, снизить затраты на создание системных и особенно бизнес-метрик, упростить и ускорить процесс обнаружения аномалий, сократить время простоев при плановой профилактике. Для созданных метрик можно настраивать пороги срабатывания триггеров и давать прогноз изменения бизнес-параметров на определенный период времени в будущем, используя известные математические модели прогнозирования [13].

Заключение

В результате выполненной работы создана универсальная структура данных и подход к мониторингу бизнес-метрик в различных областях применения. Предложенное решение не зависит от конкретных метрик, что позволяет унифицировать хранение в БД и мониторинг новых метрик с возможностью их динамического изменения без модификации структуры БД.

В статье показаны преимущества предложенного подхода и пример практической реализации бизнес-метрики «Количество загруженных фай-

лов» в международной телекоммуникационной компании. Внедрение универсальной структуры БД позволило кардинально сократить затраты на организацию новых бизнес-метрик и ускорить оперативное обнаружение возможных аномалий в системе.

Предложенное решение универсально по отношению к используемой системе мониторинга, типу БД, программным средствам практической реализации, а также математическим моделям и подходам к прогнозированию вычислительных ресурсов ГРВК.

Список используемых источников

- 1. Distillery. URL: https://distillery.com (дата обращения 12.08.2020)
- 2. Dhyani A. Create a Custom Metric in Zabbix. 2016. URL: https://www.tothenew.com/blog/create-a-custom-metric-in-zabbix (дата обращения 12.08.2020)
- 3. Efimov V.V., Mescheryakov S.V., Shchemelinin D.A., Yakovlev K.A. Integration and Continuous Service Delivery in Globally Distributed Computing System // Университетский научный журнал. 2017. № 30. С. 13–20.
- 4. Gildeh D. What We Learnt Talking to 60 Companies about Monitoring // Dataloop.IO. 2014. URL: https://dataloopio.wordpress.com/2014/01/30/what-we-learnt-talking-to-60-companies-about-monitoring (дата обращения 12.08.2020)
- 5. Levey T. Introducing Real-Time Business Metrics // AppDynamics. 2013. URL: https://www.appdynamics.com/blog/news/introducing-real-time-business-metrics (дата обращения 12.08.2020)
- 6. Cliffe D. Monitoring Business Metrics and Refining Outage Response // PagerDuty. 2015. URL: https://www.pagerduty.com/blog/monitoring-business-metrics (дата обращения 12.08.2020)
- 7. Prometheus. Open-source Systems Monitoring and Alerting Toolkit. URL: https://prometheus.io (дата обращения 12.08.2020)
- 8. Nesvold H. Getting into Business with Prometheus. 2016. URL: https://schibsted.com/blog/business-with-prometheus (дата обращения 12.08.2020)
 - 9. Saymon. URL: https://saymon.info/en-version (дата обращения 12.08.2020)
- 10. Gartner Magic Quadrant for EFSS (Enterprise File Sharing and Sync) // Content Collaboration Platforms. 2018. URL: https://www.gartner.com/en/documents/3881863/magic-quadrant-for-content-collaboration-platforms (дата обращения 12.08.2020)
- 11. Kucherova K., Mescheryakov S., Shchemelinin D. Prediction Experience and New Model // Proceedings of the 7th Annual International Zabbix Conference (Riga, Latvia, 15–17 September 2017). URL: http://www.zabbix.com/conf2017_agenda.php
- 12. Kucherova K.N., Mescheryakov S.V., Shchemelinin D.A. Using Predictive Monitoring Models in Cloud Computing Systems. Communications in Computer and Information Science // Proceedings of the 21st International Conference on Distributed Computer and Communication Networks (DCCN 2018, Moscow, Russia, 17–21 September 2018). Cham: Springer, 2018. Vol. 919. PP. 341–352. DOI:10.1007/978-3-319-99447-5_29
- 13.Mescheryakov S., Shchemelinin D., Efimov V. Adaptive control of cloud computing resources in the internet telecommunication multiservice system // Proceedings of the 6th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT-2014, St.-Petersburg, Russia, 6-8 October 2014). IEEE, 2014. PP. 287–293. DOI:10.1109/ICUMT.2014.7002117

* * *

Prediction of Cloud Computing Resources Based on the Open Source Monitoring System

K. Kucherova¹

¹Peter the Great Saint-Petersburg Polytechnic University, St. Petersburg, 195251, Russian Federation

Article info

DOI:10.31854/1813-324X-2020-6-3-100-106 Received 05th August 2020 Accepted 22th September 2020 **For citation:** Kucherova K. Prediction of Cloud Computing Resources Based on the Open Source Monitoring System. *Proc. of Telecom. Universities.* 2020;6(3):100–106. (in Russ.) DOI:10.31854/1813-324X-2020-6-3-100-106

Abstract: The paper describes the universal approach for monitoring the data storage of a globally distributed cloud computing system, which allows you to automate creation of new metrics in the system and predict their behavior for the end users. Since the existing monitoring software products provide built-in scheme only for system metrics like RAM, CPU, disk drives, network traffic, but don't offer solutions for business functions, IT companies have to design specialized database structure (DB). The data structure proposed in this paper for storing the monitoring statistics is universal and allows you to save resources when orginizing database monitoring on the scale of the GDCCS. The goal of the research is to develop a universal model for monitoring and forecasting of data storage in a globally distributed cloud computing system and its adequacy to real operating conditions.

Keywords: cloud computing system, database, monitoring, system metrics, prediction, correlation function.

References

- 1. Distillery. Available from: https://distillery.com [Accessed 12th August 2020]
- 2. Dhyani A. *Create a Custom Metric in Zabbix*. 2016. Available from: https://www.tothenew.com/blog/create-a-custom-metric-in-zabbix [Accessed 12th August 2020]
- 3. Efimov V.V., Mescheryakov S.V., Shchemelinin D.A., Yakovlev K.A. Integration and Continuous Service Delivery in Globally Distributed Computing System. *Humanities and Science University Journal*. 2017;30:13–20.
- 4. Gildeh D. What We Learnt Talking to 60 Companies about Monitoring. *Dataloop.IO.* 2014. Available from: https://dataloopio.wordpress.com/2014/01/30/what-we-learnt-talking-to-60-companies-about-monitoring [Accessed 12th August 2020]
- 5. Levey T. Introducing Real-Time Business Metrics. *AppDynamics*. 2013. Available from: https://www.appdynamics.com/blog/news/introducing-real-time-business-metrics [Accessed 12th August 2020]
- 6. Cliffe D. Monitoring Business Metrics and Refining Outage Response. *PagerDuty*. 2015. Available from: https://www.pagerduty.com/blog/monitoring-business-metrics [Accessed 12th August 2020]
- 7. *Prometheus.* Open-source Systems Monitoring and Alerting Toolkit. Available from: https://prometheus.io [Accessed 12th August 2020]
- 8. Nesvold H. *Getting into Business with Prometheus*. 2016. Available from: https://schibsted.com/blog/business-with-prometheus [Accessed 12th August 2020]
 - 9. Saymon. Available from: https://saymon.info/en-version [Accessed 12th August 2020]
- 10. Content Collaboration Platforms. Gartner Magic Quadrant for EFSS (Enterprise File Sharing and Sync). 2018. Available from: https://www.gartner.com/en/documents/3881863/magic-quadrant-for-content-collaboration-platforms [Accessed 12th August 2020]
- 11. Kucherova K., Mescheryakov S., Shchemelinin D. Prediction Experience and New Model. *Proceedings of the 7th Annual International Zabbix Conference, 15–17 September 2017, Riga, Latvia.* Available from: http://www.zabbix.com/conf2017_agenda.php [Accessed 12th August 2020]
- 12. Kucherova K.N., Mescheryakov S.V., Shchemelinin D.A. Using Predictive Monitoring Models in Cloud Computing Systems. Communications in Computer and Information Science. *Proceedings of the 21st International Conference on Distributed Computer and Communication Networks. DCCN 2018, 17–21 September 2018, Moscow, Russia.* Cham: Springer; 2018. vol.919. p.341–352. DOI:10.1007/978-3-319-99447-5_29
- 13. Mescheryakov S., Shchemelinin D., Efimov V. Adaptive control of cloud computing resources in the internet telecommunication multiservice system. *Proceedings of the 6th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops, ICUMT-2014, 6-8 October 2014, St.-Petersburg, Russia.* IEEE; 2014. p.287–293. DOI:10.1109/ICUMT.2014.7002117

Сведения об авторе:

КУЧЕРОВА Кристина Николаевна аспирант высшей школы креативной индустрии и дизайна института машиностроения, материалов и транспорта Санкт-Петербургского политехнического университета Петра Великого, kristina.mylife@gmail.com