

Математические модели и методы мониторинга и прогнозирования состояния глобально распределенных вычислительных комплексов

Д.А. Щемелинин¹ *

¹Центр научных исследований «Три Би»
Санкт-Петербург, 195277, Российская Федерация
*Адрес для переписки: dshchmel@gmail.com

Информация о статье

Поступила в редакцию 07.06.2021

Принята к публикации 18.08.2021

Ссылка для цитирования: Щемелинин Д.А. Математические модели и методы мониторинга и прогнозирования состояния глобально распределенных вычислительных комплексов // Труды учебных заведений связи. 2021. Т. 7. № 3. С. 73–78. DOI:10.31854/1813-324X-2021-7-3-73-78

Аннотация: Мониторинг событий и прогнозирование поведения динамической информационной системы приобретают все большую актуальность в связи с глобализацией облачных сервисов и резким увеличением объема обрабатываемых данных. Для своевременного обнаружения и оперативного исправления аномалии используются известные системы мониторинга, которые нуждаются в новых более эффективных и проактивных средствах прогнозирования. На конференции CMG-2013 был представлен метод прогнозирования утечки памяти в Java-приложениях, который позволяет автоматически высвободить ресурсы путем безопасной перезагрузки сервисов при достижении некоторого критического порогового значения. В данном решении реализована простая линейная математическая модель описания функции исторического тренда. Однако на практике деградация памяти и других вычислительных ресурсов может происходить не постепенно, а очень быстро в зависимости от рабочей нагрузки, и поэтому решение проблемы прогнозирования линейными методами недостаточно эффективно.

Ключевые слова: мониторинг, большие данные, моделирование, функция прогнозирования, горизонт прогнозирования, облачные технологии.

Математические модели прогнозирования состояния вычислительных систем

В представленном исследовании рассматриваются нелинейные математические модели, критерии оценки их эффективности, алгоритм использования в задачах прогнозирования, главные сложности, возникающие при обработке больших наборов данных, и пути их решения. За основу математического решения была принята методика интерполяции полиномами Лагранжа (1), дающей минимальное отклонение рассчитываемой функции $y = f(x)$, на заданных точках $X = \{x_i | i = 1, \dots, n\}$, находящихся в некоторой области D .

Было принято следующее допущение, что значения функции $f(x)$ определено в известных точках X , которые, в свою очередь, названы узлами интерполяции, отстоящими друг от друга на расстояние шага интерполяционной сетки: $\Delta x_i = x_i - x_{i-1}$, т. е. задача интерполяции сводится к поиску заданного

класса функции $y_i = F(x_i)$, максимально приближенную к текущей ситуации происходящей в информационной системе:

$$P_n(x) = \sum_{i=0}^n y_i Q_{n,i}(x), \quad (1)$$

где $Q_{n,i}(x)$ – полиномы степени вида n :

$$Q_{n,i}(x) = \prod_{j=0, j \neq i}^n \frac{(x - x_j)}{(x_i - x_j)}. \quad (2)$$

Так как полином (2) принимает значение единицы в точке x_i и нуля в остальных узлах интерполяции, следовательно, в точке x_i исходный полином принимает значение y_i . Это доказывает, что построенный полином (1) является интерполяционным для функции $y = f(x)$ на сетке X .

Однако при увеличении числа узлов интерполяции приходится перестраивать весь полином в форме Лагранжа заново, что является неудобным

подходом для использования в системах непрерывного мониторинга глобально распределенных вычислительных комплексов (ГРВК), где поведение системы и отдельных компонент изменяется по непериодической закономерности, поэтому полином Лагранжа был представлен в виде выражения:

$$P_n(x) = P_0(x) + \sum_{i=1}^n (P_i(x) - P_{i-1}(x)), \quad (3)$$

где $P_i(x)$ – полиномы Лагранжа степени $i \leq n$.

Допустим:

$$Q_i(x) = P_i(x) - P_{i-1}(x). \quad (4)$$

Полином $Q_i(x)$ имеет степень i и будет обращаться в нуль, при следующем условии:

$$x = x_0, x = x_1, \dots, x = x_{i-1}.$$

Следовательно, полином $Q_i(x)$ может быть представлен в следующем виде:

$$Q_i(x) = A_i(x - x_0) \dots (x - x_{i-1}), \quad (5)$$

где A_i – коэффициент при x_i , который, в свою очередь, не входит в $P_{i-1}(x)$, а, следовательно, A_i совпадает с коэффициентом при x_i в полиноме $P_i(x)$.

Таким образом, из определения полинома (3) получаем следующее выражение для коэффициента A_i :

$$A_i = \sum_{k=0}^i \frac{f(x_k)}{w_{k,i}}, \quad (6)$$

где $w_{k,i} = (x_k - x_0) \dots (x_k - x_{k-1})(x_k - x_{k+1}) \dots (x_k - x_i)$, что дает возможность переписать формулу (4) в следующем виде:

$$P_n(x) = P_{n-1} + A_n(x - x_0) \dots (x - x_{n-1}). \quad (7)$$

Далее, рекуррентно представляем полином $P_i(x)$ в следующем виде:

$$P_n(x) = A_0 + A_1(x - x_0) + \dots + A_n(x - x_0) \dots (x - x_{n-1}). \quad (8)$$

Выражение для полинома (8) является удобным для вычисления интерполяционной функции, так как последующее увеличение числа узлов на единицу требует только добавления одного слагаемого. Точность интерполяции зависит от правильного выбора узлов, которые минимизируют значение w_{n+1} в оценке погрешности математического вычисления аппроксимирующей функции.

Эту задачу можно решить при помощи применения многочлена Чебышева:

$$T_{n+1}(x) = \frac{(b-a)}{(2^{2n-1})} \cos((n+1) \arccos \frac{2x - (b+a)}{(b-a)}), \quad (9)$$

где в качестве узлов интерполяции необходимо использовать корни следующего многочлена:

$$x_k = \frac{a+b}{2} + \frac{b-a}{2} \cos \frac{(2k+1)\pi}{2(n+1)}. \quad (10)$$

Методы и алгоритмы мониторинга больших данных ГРВК

Мониторинг систем больших данных имеет ряд особенностей. Одной из них является раздельное хранение статистики реального времени и архивированной истории с целью улучшения производительности системы управления базами данных. Если заданный период анализа охватывает архивный период, то качество прогноза снижается по причине погрешности, вызванной ограниченностью усредненных архивных данных [1].

В этом случае целесообразно использовать рекуррентную модель вычисления средних значений A_k и среднеквадратических отклонений Q_k :

$$A_0 = 0, \quad A_k = A_{k-1} + \frac{x_k - A_{k-1}}{k}, \quad (11)$$

где $k = 0, \dots, n$.

$$Q_0 = 0, \\ Q_k = Q_{k-1} + \frac{k-1}{k} (x_k - A_{k-1})^2 = \\ = Q_{k-1} + (x_k - A_{k-1})(x_k - A_k), \quad (12)$$

если $Q_1 = 0$, т. к. $k - 1 = 0$ или $x_1 = A_1$.

Совокупная дисперсия:

$$q_n^2 = \frac{q_n}{n}. \quad (13)$$

Другой особенностью систем больших данных является выбор параметров функции прогнозирования. Например, при интерполяции полиномами высокого порядка появляется эффект нежелательных осцилляций на концах интервала, называемый феноменом Рунге [2], который ухудшает корреляцию данных, как это показано на рисунке 1.

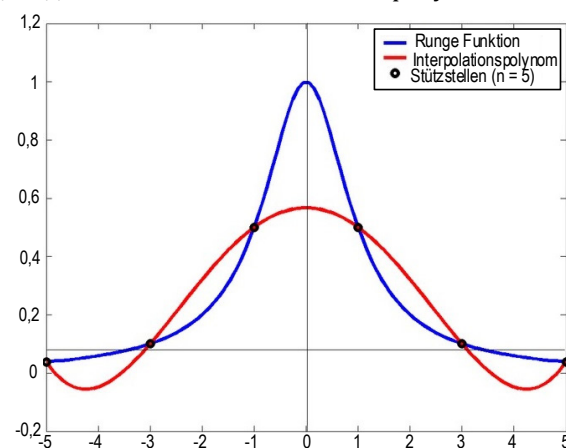


Рис. 1. Феномен Рунге

Fig. 1. The Runge Phenomenon

Ключевой задачей прогнозирования является выбор математической модели, наиболее близкой к реальным статистическим данным, накопленным в системе мониторинга. В дополнение к линейной модели $y(t) = a_0 + a_1t$, являющейся частным случаем полинома 1-го порядка, в рамках данной ста-

тью, рассматривается набор следующих нелинейных функций, которыми можно описать до 100 % встречающихся на практике тенденций изменения контролируемых метрик:

– полиномиальная:

$$y(t) = a_0 + a_1t + a_2t^2 + \dots + a_nt^n; \quad (14)$$

– логарифмическая:

$$y(t) = a + b \log(t); \quad (15)$$

– экспоненциальная:

$$y(t) = a \exp(bt); \quad (16)$$

– степенная:

$$y(t) = a t^b. \quad (17)$$

С другой стороны, использование близкого горизонта прогнозирования совместно с быстрорастущими функциями, такими как экспоненциальная (16) или степенная (17) или даже полиномиальная (14) порядка 2-3, показанных на рисунке 2, может привести к тому, что результатом вычисления функции прогнозирования будет бесконечность, т. е. арифметическое переполнение, что на практике означает событие никогда не произойдет либо наоборот, уже произошло.

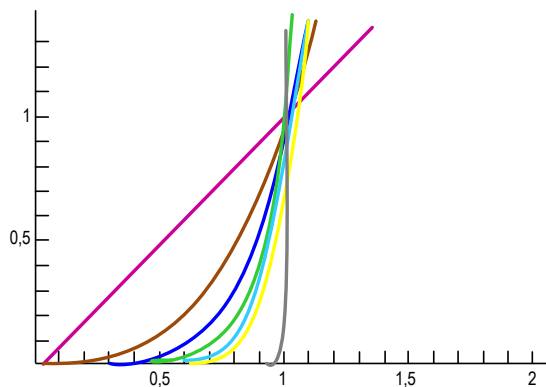


Рис. 2. Примеры быстрорастущих функций (14–17)

Fig. 2. Examples of Fast-Growing Functions (14–17)

Таким образом, применение полиномов высоких степеней и других сложных функций увеличивает время вычисления, но не дает ощутимого преимущества в точности прогнозирования, а наоборот, повышает риск появления аномалии при больших наборах статистических данных.

На практике, в системах мониторинга статистика реального времени хранится в интервале от двух недель до месяца, а исторические данные интерполируются и хранятся в архиве в виде усредненных, максимальных и минимальных значений. Такой метод организации СУБД не позволяет строить долговременный прогноз на временные промежутки более недели.

В качестве критерия корреляции статистических данных и выбранной математической модели использованы стандартное среднее квадратическое

отклонение и коэффициент детерминации R^2 модели зависимости V случайной величины y от факторов x [3]:

$$Q^2 = \sum_{t=1}^n (y_t - \hat{y}_t)^2, \quad (18)$$

$$R^2 = 1 - \frac{V(y|x)}{V(y)} = 1 - \frac{Q^2}{Q_y^2}.$$

В идеальном случае коэффициент детерминации $R^2 = 1$, что означает 100-процентное совпадение с фактическими данными [4]. Например, для прогнозирования тенденций поведения метрики, отображенные на рисунке 2, использованы следующие полиномиальные функции с оптимальной корреляцией по критерию (18):

$$y(t) = -24,67t^2 + 2,106t - 4,101; \quad R^2 = 0,98,$$

$$y(t) = 8,57t^2 - 141,91t + 667,27; \quad R^2 = 0,96, \quad (19)$$

$$y(t) = -10,14t^3 + 106t^2 - 5,101t + 7,101;$$

$$R^2 = 0,89.$$

В ходе выполнения исследования была разработана и внедрена следующая пошаговая методика прогнозирования вычислительных ресурсов в облачных ГРВК и алгоритм ее программной реализации.

Шаг 1. Определение наиболее важных системных и/или бизнес метрик, мониторинг и прогнозирование которых принесет наибольший эффект при наименьших трудозатратах. Например, для Java-приложений облачных сервисов такими метриками являются объем виртуальной памяти для оценки ресурсов и частота веб-запросов, как индикатор рабочей нагрузки на систему.

Шаг 2. Выбор математической модели, наиболее близкой к реальным данным. Данный процесс является итерационным и на каждом шаге итерации требует уточнения параметров выбранной модели вплоть до достижения необходимого уровня точности прогноза, включая следующие шаги:

шаг 2а. Экспорт статистических данных системы мониторинга той метрики, для которой будет строиться прогноз и загрузка их в одну из известных аналитических систем [5]: WolframAlpha, Mathematica, R, Matlab и т. д. При сравнительно небольшом объеме данных можно воспользоваться стандартными средствами MS Excel;

шаг 2б. Разделить весь временной интервал на участки нормальной работы, период деградации и аномалии. Участки разной природы лучше анализировать отдельно, как показано на рисунке 3 с применением математических моделей (19);

шаг 2в. Выбрать тип функции прогнозирования и построить линию тренда на указанный период, который должен быть меньше, чем имеющийся (фактических данных). На практике при анализе ежемесячной статистики можно построить достаточно точный прогноз, максимально на неделю;

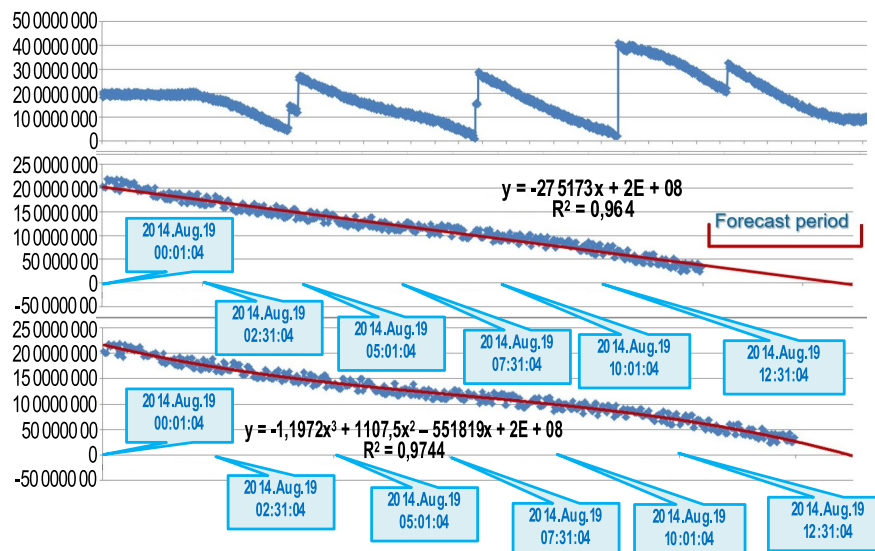


Рис. 3. Пример статистических данных и вычисления функции прогнозирования

Fig. 3. An Example of Statistical Data and Calculation of the Forecasting Function

шаг 2г. Вычислить оптимальные параметры выбранной математической модели для достижения максимальной корреляции по критерию (18);

шаг 2д. Повторить шаги 2б–2г до получения необходимого уровня точности прогноза;

шаг 2е. Полученную в результате оптимизации математическую функцию указать в качестве модели для построения прогноза данных в системе мониторинга.

Шаг 3. Создать дополнительную вычисляемую метрику в системе мониторинга на основе аппроксимации статистических данных, с использованием выбранной функции прогнозирования, для которой можно создать специальные триггеры, которые будут реагировать на прогноз аномального события раньше, чем оно произойдет в реальном времени.

Производительность современных вычислительных систем дает возможность повторять шаги 2–3 и пересчитывать прогностический тренд в реальном времени, варьировать параметрами математической модели, уточнять пороговые значения реакции триггеров и в конечном итоге повысить дальность и точность прогноза [6, 7].

Выводы

Предложенные математические модели и алгоритм прогнозирования загрузки вычислительных ресурсов в облачных ГРВК реализованы с использованием мониторинговой системы Zabbix, которая входит в пятерку мировых лидеров по результатам опроса 60 крупных международных IT-компаний [8].

Точность прогноза с использованием нелинейной модели значительно выше, чем линейной. В то же время сложные выражения и полиномы высоких порядков не увеличивают точность прогноза, а, напротив, могут привести к нежелательным результатам, таким как переполнение или феномен Рунге. Согласно опыту, полиномы второго и третьего порядка дают высокую производительность вычислений и достаточную точность прогноза.

Корректный горизонт прогнозирования имеет важное значение, т. к. данные меняются в реальном времени, и чем он короче, тем точнее прогноз. Баланс между историческим периодом и горизонтом прогноза рекомендуется находить экспериментально, начиная с пропорции 5:1 [9–16].

Список используемых источников

1. Shchemelinin D., Mescheryakov S. Capacity Management of Java-based Business Applications Running on Virtualized Environment // Annual International Conference of the Computer Measurement Group (CMG 2013, London, UK, 14–15 May 2013). Computer Measurement Group Inc., 2013. PP. 193–197.
2. Шарый С.П. Курс вычислительных методов. Новосибирск: Институт вычислительных технологий СО РАН, 2012. URL: <http://mathscinet.ru/files/SShary.pdf> (дата обращения 28.07.2021)
3. Kucheroва K., Mescheryakov S., Shchemelinin D. Cloud monitoring – focusing on forecasting // Internet, Mobile, Performance and Capacity, Cloud and Technology Conference (IMPACT 2016, La Jolla, USA, 07–10 November 2016). Computer Measurement Group Inc., 2016.
4. Гайсенюк И.С., Закревская И.С. Оценка качества прогнозов // Бизнес-образование как инструмент устойчивого развития экономики: Сб. трудов науч.-практич. конф. Иркутск: Иркутский государственный университет, 2012.
5. Wolfram Mathematica Computations // Wolfram. URL: <https://www.wolfram.com/mathematica> (дата обращения 28.07.2021)

6. Тихонов Э.Е., Бурдо А.И. К вопросу совершенствования автоматизированных систем прогноза // Материалы межрегиональной конференции «Студенческая наука – экономика научно-технического прогресса». Ставрополь: Сев-Кав ГТУ, 2000. С. 30–31.
7. Сизов А.А. Модели, способы и программные средства поддержки принятия решений на основе прогнозирования временных рядов с переменной структурой. Дис. ... канд. техн. наук. М.: Национальный исследовательский университет «МЭИ», 2014. 141 с.
8. Gildeh D. Monitoring Nightmares for DevOps // SlideShare. 2014. URL: <https://www.slideshare.net/outlyer/david-41050819> (дата обращения 28.07.2021)
9. Bhandarkar M. Future of Data Intensive Applications // Proceedings of the 2nd ASE International Conference on Big Data Science and Computing. Stanford, 2014. URL: <https://www.kdnuggets.com/2014/08/ase-conference-big-data-science-day1.html> (дата обращения 28.07.2021)
10. 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
11. Логунов Д.Г. Применение методов и моделей прогнозирования временных рядов для оценки циклической нагрузки в облачных системах // Научно-практическая конференция с международным участием «Неделя науки СПбПУ» (Санкт-Петербург, Россия, 18–23 ноября 2019). СПб: Политех-Пресс, 2020. С. 165–168.
12. Михайлов К.И. Анализ программных средств мониторинга высоконагруженных облачных информационных систем // XXIII международная научно-практическая конференция «Системный анализ в проектировании и управлении». Ч. 1. (SAEC-2019, Санкт-Петербург, Россия, 10–11 июня 2019) СПб: Политех-Пресс, 2019. С. 359–362.
13. Михайлов К.И. Мониторинг и визуализация ресурсов смешанной природы в облачной среде // Научно-практическая конференция с международным участием «Неделя науки СПбПУ» (Санкт-Петербург, Россия, 19–24 ноября 2018). Институт машиностроения, металлургии и транспорта. Ч. 2. СПб: Политех-Пресс, 2018. С. 234–236.
14. Яковлев К.А. Аналитическая обработка и визуализация мониторинговых данных глобально распределенной информационной системы // Научно-практическая конференция с международным участием «Неделя науки СПбПУ» (Санкт-Петербург, Россия, 13–19 ноября 2017). СПб: Федеральное государственное автономное образовательное учреждение высшего образования "Санкт-Петербургский политехнический университет Петра Великого", 2017. С. 203–205.
15. Кучерова К.Н. Мониторинг и прогнозирование серверных ресурсов баз данных облачной архитектуры // Научно-практическая конференция с международным участием «Неделя науки СПбПУ» (Санкт-Петербург, Россия, 13–19 ноября 2017). СПб: Федеральное государственное автономное образовательное учреждение высшего образования "Санкт-Петербургский политехнический университет Петра Великого", 2017. С. 199–201.
16. Кучерова К.Н., Ефимов В.В. Повышение эффективности мониторинга облачной инфраструктуры на основе прогностического моделирования // Научно-практическая конференция с международным участием «Неделя науки СПбПУ» (Санкт-Петербург, Россия, 14–19 ноября 2016). СПб: Федеральное государственное автономное образовательное учреждение высшего образования "Санкт-Петербургский политехнический университет Петра Великого", 2016. С. 117–119.

* * *

Mathematical Models and Methods for Monitoring and Predicting the State of Globally Distributed Computing Systems

D. Shchemelinin¹ 

¹Research Center "3B"
St. Petersburg, 195277, Russian Federation

Article info

DOI:10.31854/1813-324X-2021-7-3-73-78

Received 7th June 2021

Accepted 18th August 2021

For citation: Shchemelinin D. Mathematical Models and Methods for Monitoring and Predicting the State of Globally Distributed Computing Systems. *Proc. of Telecom. Universities*. 2021;7(3):73–78. (in Russ.) DOI:10.31854/1813-324X-2021-7-3-73-78

Abstract: *Monitoring events and predicting the behavior of a dynamic information system are becoming increasingly important due to the globalization of cloud services and a sharp increase in the volume of processed data. Well-known*

monitoring systems are used for the timely detection and prompt correction of the anomaly, which require new, more effective and proactive forecasting tools. At the CMG-2013 conference, a method for predicting memory leaks in Java applications was presented, which allows IT teams to automatically release resources by safely restarting services when a certain critical threshold value is reached. Article's solution implements a simple linear mathematical model for describing the historical trend function. However, in practice, the degradation of memory and other computational resources may not occur gradually, but very quickly, depending on the workload, and therefore, solving the forecasting problem using linear methods is not effective enough.


Keywords: monitoring, big data, modeling, forecasting function, forecasting horizon, cloud technologies.

References

1. Shchemelinin D., Mescheryakov S. Capacity Management of Java-based Business Applications Running on Virtualized Environment. *Annual International Conference of the Computer Measurement Group, CMG 2013, 14–15 May 2013, London, UK*. Computer Measurement Group Inc.; 2013. p.193–197.
2. Sharyy S.P. *Computational Methods Course*. Novosibirsk: Institute of Computational Technologies SB RAS Publ.; 2012. (in Russ.) URL: <http://mathscinet.ru/files/SShary.pdf> [Accessed 28th July 2021]
3. Kucherova K., Mescheryakov S., Shchemelinin D. Cloud monitoring – focusing on forecasting. *Internet, Mobile, Performance and Capacity, Cloud and Technology Conference, IMPACT 2016, 07–10 November 2016, La Jolla, USA*. Computer Measurement Group Inc.; 2016.
4. Gaysenok I.S., Zakrevskaya D.V. Assessment of the Quality of Forecasts. *Proceedings of the Scientific and Practical Conferences on Business Education as a Tool for Sustainable Economic Development*. Irkutsk: Irkutsk State University Publ.; 2012. (in Russ.)
5. *Wolfram*. Wolfram Mathematica Computations. URL: <https://www.wolfram.com/mathematica> [Accessed 28 July 2021]
6. Tikhonov E.Ye., Burdo A.I. On the Issue of Improving Automated Forecasting Systems. *Proceedings of the Interregional Conference on Student Science – Economics of Scientific and Technological Progress*. Stavropol: North Caucasus Federal University Publ.; 2000. p.30–31. (in Russ.)
7. Sizov A.A. *Models, Methods and Software for Decision Support Based on Forecasting Time Series with Variable Structure*. PhD Thesis. Moscow: National Research University "Moscow Power Engineering Institute" Publ.; 2014. p. 141. (in Russ.)
8. Gildeh D. Monitoring Nightmares for DevOps. *SlideShare*. 2014. URL: <https://www.slideshare.net/outlyer/david-41050819> [Accessed 28 July 2021]
9. Bhandarkar M. Future of Data Intensive Applications. *Proceedings of the 2nd ASE International Conference on Big Data Science and Computing*. Stanford; 2014. URL: <https://www.kdnuggets.com/2014/08/ase-conference-big-data-science-day1.html> [Accessed 28 July 2021]
10. 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
11. Logunov D.G. Application of Methods and Models for Forecasting Time Series for Assessing Cyclical Load in Cloud Systems. *Proceedings of Scientific and Practical Conference with International Participation on Week of Science SPbPU, 18–23 November 2019, St. Petersburg, Russia*. St. Petersburg: Politekh-Press Publ.; 2020. p.165–168. (in Russ.)
12. Mikhailov K.I. Analysis of Software Tools for Monitoring High-Load Cloud Information Systems. *Proceedings of the XXIIIrd International Scientific and Practical Conference on System Analysis in Design and Management, SAEC-2019, 10–11 June 2019, St. Petersburg, Russia. Part 1*. St. Petersburg: Politekh-Press Publ.; 2019. p.359–362. (in Russ.)
13. Mikhailov K.I. Monitor and Visualize Mixed Assets in the Cloud. *Proceedings of Scientific and Practical Conference with International Participation on Week of Science SPbPU, 19–24 November 2018, St. Petersburg, Russia. Institute of Mechanical Engineering, Metallurgy and Transport. Part 2*. St. Petersburg: Politekh-Press Publ.; 2018. p.234–236. (in Russ.)
14. Yakovlev K.A. Analytical Processing and Visualization of Monitoring Data of a Globally Distributed Information System. *Proceedings of Scientific and Practical Conference with International Participation on Week of Science SPbPU, 13–19 November 2017, St. Petersburg, Russia*. St. Petersburg: Peter the Great St.Petersburg Polytechnic University Publ.; 2017. p.203–205. (in Russ.)
15. Kucherova K.N. Monitoring and Forecasting Server Resources of Cloud Architecture Databases. *Proceedings of Scientific and Practical Conference with International Participation on Week of Science SPbPU, 13–19 November 2017, St. Petersburg, Russia*. St. Petersburg: Peter the Great St.Petersburg Polytechnic University Publ.; 2017. p.199–201. (in Russ.)
16. Kucherova K.N., Efimov V.V. Improving the Efficiency of Monitoring Cloud Infrastructure Based on Predictive Modeling. *Proceedings of Scientific and Practical Conference with International Participation on Week of Science SPbPU, 14–19 November 2016, St. Petersburg, Russia*. St. Petersburg: Peter the Great St.Petersburg Polytechnic University Publ.; 2016. p.117–119. (in Russ.)

Сведения об авторе:

ЩЕМЕЛИНИН
Дмитрий Александрович

кандидат технических наук, старший научный сотрудник Центра научных исследований «3В», dshchmel@gmail.com
 <https://orcid.org/0000-0003-3032-130X>