

УДК 519.876

**МЕТОД ПРОГНОЗУВАННЯ НАВАНТАЖЕННЯ НА КЛАСТЕРНУ СИСТЕМУ***Москвін Олексій, Деркач Анна*

Вінницький національний технічний університет

**Анотація**

*На основі аналізу існуючих засобів прогнозування навантаження на кластерну систему був розроблений двоетапний метод, що базується на використанні логістичної регресії та нечіткої бази знань. Дослідження розробленого методу та результати тестування програмного забезпечення підтвердили його ефективність.*

**Abstract**

*Having analyzed the existing means for predicting the load on the cluster system a two-stage method based on the use of logistic regression and fuzzy knowledge was developed. The effectiveness of the developed method and software confirmed the correctness of the results.*

**Вступ**

На етапі проектування програмного забезпечення орієнтованого на роботу у мережі Інтернет все частіше постає задача гарантування безвідмовної його роботи в умовах стрибкоподібних змін інтенсивності вхідних запитів. Найбільш перспективний та економічно обґрунтований підхід до вирішення даної задачі полягає у використанні еластичних хмарних середовищ. Найбільш відомі реалізації даного підходу доступні у сервісах Amazon EC2, Azure, Google Cloud. Основним принципом роботи таких систем є автоматичне розгортання додаткових обчислювальних ресурсів за умов зростання навантаження на існуючі. Системи, що використовують горизонтальне масштабування можна розділити на еластичні та нееластичні. Особливістю еластичних систем є здатність автоматично збільшувати кількість обчислювальних ресурсів відповідно до навантаження за рахунок використання хмарних ресурсів середовища. Нееластичні системи позбавлені такої можливості і можуть бути адаптовані до більшого навантаження лише в ручному режимі. Стрибкоподібне збільшення навантаження на нееластичну систему призводить до її миттєвого перевантаження та відмови у обслуговуванні. Оскільки в таких системах основним показником, що впливає на вартість їх обслуговування є процесорний час, що лінійно залежить від кількості одночасно функціонуючих обчислювальних вузлів, аналіз та оптимізація методу, що використовується для прийняття рішень щодо управління розгортанням обчислювальних ресурсів, необхідних для роботи системи, є актуальною задачею.

Аналіз існуючих засобів в зазначених еластичних хмарних системах показав, що механізми, які контролюють навантаження у еластичних системах є однопараметричними, та задаються правилами ЯКЦО-ГО, на основі яких керуючий компонент здійснює додавання або видалення вузлів. Основними параметрами, які використовуються в таких правилах є миттєві значення навантаження на процесор та інтенсивності вхідних запитів. Такий підхід не враховує профіль навантаження системи, особливості її функціонування та призводить до відмов у пікові моменти або до більших витрат на обслуговування.

Для вирішення зазначених недоліків в роботі було розроблено метод прогнозування навантаження на кластерну систему, який дозволяє керувати її розміром не лише на основі миттєвих даних навантаження, але і на основі історичних даних інтенсивності вхідних запитів, навантаження на процесор, дискову підсистему, розмір вільної оперативної пам'яті отриманих в реальному часі з підсистеми балансування навантаження та всіх розгорнутих вузлів. Запропонований метод побудовано з

використанням логістичної регресії, методів машинного навчання [2] та нечітких баз знань.

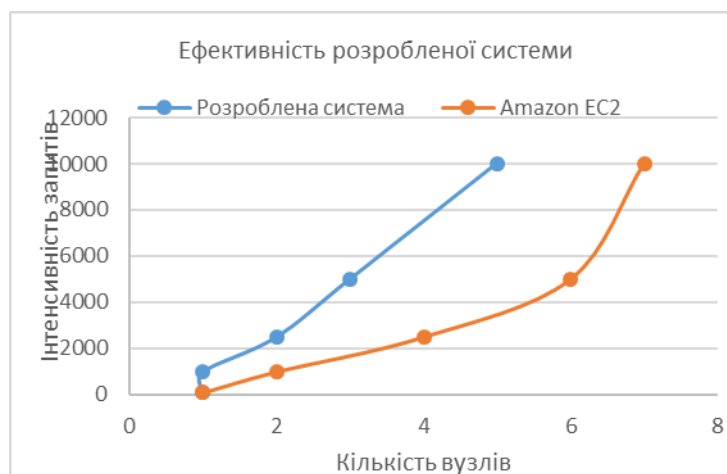


Рисунок 1 – Результати дослідження ефективності системи прогнозування навантаження

Метод прогнозування навантаження дозволяє визначити ймовірність перевантаження системи. Bazуючись на отриманому результаті прогнозування, підсистема планування приймає рішення про необхідність розгортання додаткових апаратних ресурсів в кластерній системі. У випадку якщо ймовірність стрибкоподібної зміни навантаження менше 50%, метод передбачає здійснення додаткового уточнення прогнозу для прийняття рішення про навантаження [3]. Для цього в роботі використано нечітку базу знань надану експертами, розмір якої складає 125 правил.

В роботі було здійснено перевірку ефективності методу прогнозування навантаження на кластерну систему. Результати порівнювалися з аналогічним хмарним сервісом Amazon EC2. При подачі однакового навантаження на кластерну систему, та за однакових критеріїв до мінімальної пропускної здатності системи, Amazon EC2 створює на 38% більшу кількість вузлів кластера ніж система, побудована з використанням розробленого методу (рисунок 1).

**Висновки.** Результати тестування реалізованої автоматизованої системи прогнозування навантаження на кластерну систему на основі розробленого методу з використанням інструментарію Apache Mahout підтвердило ефективність розробленого методу. Впровадження власного методу дозволило здійснити більш ефективне використання наявних апаратних ресурсів та відповідно знизити витрати на розгортанні додаткових вузлів кластера при виявленні приросту навантаження.

### Список використаних джерел

1. Москвін О. М. Про підвищення стійкості методів моделювання жорстких систем. / С. М. Москвіна, О. М. Москвін, Т. О. Голубева. // Вісник Вінницького політехнічного інституту. - Вінниця: ВНТУ. – 2007. - №1. / <http://visnyk.vntu.edu.ua/article/view/419/3109> [електронний ресурс] - С. 84-88.
2. Larsen K.R.T. A Cost and Performance Model for Web Service Investment / K. R. T. Larsen and P. A. Bloniarz. – ACM, Feb. 2000, pp. 109-116.
3. Menasce D. A. Capacity Planning and Performance Modeling: From Mainframes to Client-Server Systems / D. A. Menasce, V. A. F. Almeida, and L. W. Dowdy. – Prentice Hall, Upper Saddle River, New Jersey, 1994.