

Міністерство освіти і науки України
Вінницький національний технічний університет

КОЗАЧУК АНДРІЙ ВАЛЕРІЙОВИЧ

УДК 519.216.3:004.9

**ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ МАСШТАБУВАННЯ ХМАРНОГО
ЗАСТОСУНКУ, ЩО БАЗУЄТЬСЯ НА ПРОГНОЗУВАННІ ЙОГО СТАНУ**

Спеціальність 05.13.06 – інформаційні технології

Автореферат
дисертації на здобуття наукового ступеня
кандидата технічних наук

Вінниця – 2016

Дисертацією є рукопис.

Роботу виконано у Вінницькому національному технічному університеті Міністерства освіти і науки України.

Науковий керівник: кандидат технічних наук, доцент
Савчук Тамара Олександрівна,
Вінницький національний технічний університет,
професор кафедри комп'ютерних наук.

Офіційні опоненти: доктор технічних наук, професор
Поворознюк Анатолій Іванович,
Національний технічний університет «Харківський
політехнічний інститут», професор кафедри
обчислювальної техніки та програмування


доктор технічних наук, професор
Фісун Микола Тихонович,
Чорноморський державний університет імені Петра
Могили, м. Миколаїв, завідувач кафедри
інтелектуальних інформаційних систем.

Захист відбудеться «01» грудня 2016 р. о 15:00 годині на засіданні спеціалізованої вченої ради Д 05.052.01 у Вінницькому національному технічному університеті за адресою: 21021, м. Вінниця, вул. Хмельницьке шосе, 95, ГНК, ауд. 210.

З дисертацією можна ознайомитися в бібліотеці Вінницького національного технічного університету за адресою: 21021, м. Вінниця, вул. Хмельницьке шосе, 95, ГНК.

Автореферат розісланий «29» жовтня 2016 р.

Учений секретар
спеціалізованої вченої ради



С. М. Захарченко

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Використання хмарних обчислень дозволяє значно полегшити підтримку серверної інфраструктури, збільшити швидкість розгортання застосунку, а також адаптувати хмарний застосунок під режим навантаження зі змінними піками. Нерівномірність інтенсивності використання хмарного застосунку визначає актуальною задачею оптимізації обчислювальних ресурсів, що виділяються для підтримки його роботи. Ця задача розв'язується завдяки масштабуванню хмарного застосунку на основі оцінки його стану – адаптації кількості виділених обчислювальних ресурсів до навантаженості хмарного застосунку.

На сьогоднішня спостерігається стійка тенденція до збільшення популярності хмарних обчислень. За 2015 рік світовий ринок хмарних обчислень у цілому зріс на 28%, а ринок хмарних обчислень типу "Платформа як послуга" - на 51%. За прогнозом агентства TBR сумарний дохід від надання послуг хмарних обчислень збільшиться з 80 мільярдів доларів США в 2015 році до 167 мільярдів у 2020 році. Національний інститут стратегічних досліджень прогнозує, що багаторазове збільшення українського ринку хмарних обчислень найближчими роками приведе до виникнення нового специфічного і значущого сектору української економіки та інфраструктури.

Системи масштабування хмарного застосунку здебільшого використовують реактивне масштабування – нарощування або зниження обчислювальних потужностей при досягненні певного порогу використання наявних обчислювальних ресурсів. Даний підхід є ефективним у цілому, але у випадку коротких та частих піків навантаження спостерігаються проблеми в роботі застосунку в проміжку часу між початком процесу масштабування та виділенням обчислювальних ресурсів. Для уникнення такого недоліку застосовують предиктивне масштабування, що ґрунтується на прогнозі інтенсивності навантаження хмарного застосунку. Такий прогноз може бути побудований за допомогою методів прогнозування часових рядів із застосуванням їх до часового ряду мережевих запитів, які надходять до хмарного застосунку. При цьому інформація про можливі піки навантаження хмарного застосунку може бути використана для підвищення точності прогнозу відповідними методами в залежності від стану, в якому перебуває хмарний застосунок. Більшість сучасних систем здійснюють лише горизонтальне масштабування на базі реактивних правил. Поєднання горизонтального та вертикального масштабування з використанням як реактивних правил так і прогнозу стану хмарного застосунку збільшить кількість можливих станів серверної інфраструктури та ефективність масштабування.

Сучасні методи масштабування хмарного застосунку розглянуті в роботах М. Нето, Л. Мур, З. Гонга, С. Сальвадора, М. Мао, З. Шена, З. Факіда, М. Седагат, Е. Елмрота, Р. Кунча, Ж. Сімсон, Х. Канг. Застосування методів прогнозування станів об'єктів представлено в роботах В. Н. Афанасьєва, М. М. Юзбашева, Х. Діа, А. І. Петренка, В. М. Дубового, А. А. Ярового, С. В. Юхимчука, Р. Н. Кветного, Б. І. Мокіна, І. В. Кузьміна, С. Д. Штовби.

Класичні підходи до масштабування хмарного застосунку у випадку змінних піків навантаження часто дають незадовільні результати, оскільки це не відповідає шаблону його нормального функціонування та може призвести до стрімкої зміни потреби в обчислювальних ресурсах.

Більшість з існуючих підходів до масштабування хмарного застосунку не враховують нерівномірність часу проведення масштабування, деякі з них мають обмежене застосування через високі технічні вимоги, наприклад, потребу у можливості зміни черги виконання запитів або внесення змін до коду хмарного застосунку. Крім того, часто ігнорується можливість проведення вертикального масштабування за обмеженням підтримки хмарними платформами.

Отже, актуальною є задача розробки нової інформаційної технології масштабування хмарного застосунку, що використовує прогноз його стану при прийнятті рішень щодо масштабування, з метою скорочення часу реакції на зміну завантаженості та збільшення ефективності функціонування хмарного застосунку.

Зв'язок роботи з науковими програмами, планами, темами

Дисертаційна робота виконана відповідно до напрямку наукових досліджень кафедри комп'ютерних наук Вінницького національного технічного університету «Моделі, методи, технології та пристрої інтелектуальних інформаційних систем управління, економіки, навчання та комунікацій», відповідно до переліку пріоритетних тематичних напрямів наукових досліджень і науково-технічних розробок, де здобувач брав участь в якості виконавця окремих розділів.

Мета і задачі дослідження. Метою дисертаційного дослідження є підвищення ефективності функціонування хмарного застосунку за рахунок використання інформаційної технології, що базується на поєднанні реактивного та предиктивного підходів до здійснення його горизонтального та вертикального масштабування.

Для досягнення поставленої мети необхідно розв'язати такі задачі:

1. Дослідити методи та засоби масштабування хмарного застосунку.
2. Розробити інформаційну модель масштабування хмарного застосунку, що описує хмарний застосунок як систему масового обслуговування, включає у себе критерій класифікації поточного режиму роботи хмарного застосунку та граф станів його інфраструктури.
3. Удосконалити метод прийняття рішень щодо масштабування хмарного застосунку, що базується на поєднанні реактивного та удосконаленого предиктивного підходів до горизонтального та вертикального масштабування з урахуванням інформації про можливі піки навантаження хмарного застосунку.
4. Розробити інформаційну технологію масштабування хмарного застосунку, яка поєднує горизонтальне та вертикальне масштабування з використанням як реактивних правил так і прогнозу стану хмарного застосунку для збільшення ефективності його функціонування.
5. Розробити інформаційну систему автоматизованого масштабування хмарного застосунку.

6. Розробити симулятор роботи хмарного застосунку, що моделює функціонування хмарного застосунку із заданим профілем навантаження.

7. Провести порівняння ефективності функціонування хмарного застосунку при використанні розробленої інформаційної технології та за технологією, що використовує класичні підходи до масштабування.

8. Дослідити результати застосування інформаційної системи автоматизованого масштабування хмарного застосунку.

Об'єктом дослідження є процес масштабування хмарного застосунку.

Предметом дослідження є інформаційні технології масштабування хмарного застосунку.

Методи дослідження. Використано теоретичні та змішані методи, серед них: методи системного аналізу та синтезу для розробки інформаційної моделі масштабування хмарного застосунку, методи авторегресії–проінтегрованого ковзного середнього та експоненційного згладжування – для прогнозування часового ряду мережевих запитів до хмарного застосунку, теорію систем масового обслуговування – для оцінки часу виконання мережевих запитів, метод Пейджа–Хінклі – для класифікації стану роботи хмарного застосунку, методи предиктивного та реактивного масштабування – для регулювання обчислювальних потужностей хмарного застосунку, статистичні методи – для визначення рівня значущості отриманих результатів, теорія інформаційних систем – при розробці архітектури системи автоматизованого масштабування хмарного застосунку, метод ідеалізації – для побудови моделі процесу функціонування хмарного застосунку, метод формалізації – при побудові моделі інфраструктури хмарного застосунку, метод об'єктно-орієнтованого програмування – при розробці модулів системи автоматизованого масштабування хмарного застосунку, метод комп'ютерного моделювання – для симуляції роботи хмарного застосунку.

Наукова новизна одержаних результатів полягає в тому, що:

1. Вперше запропоновано інформаційну модель масштабування хмарного застосунку, яка відрізняється моделлю хмарного застосунку як системи масового обслуговування, критерієм класифікації поточного режиму роботи хмарного застосунку та графом станів його інфраструктури, що дозволяє здійснювати оцінювання поточного і прогнозування майбутніх станів хмарного застосунку.

2. Удосконалено метод прийняття рішень щодо масштабування хмарного застосунку, що на відміну від існуючих базується на поєднанні реактивного та удосконаленого предиктивного підходів до горизонтального та вертикального масштабування з урахування інформації про можливі піки навантаження хмарного застосунку, збільшує потужність множини рішень щодо масштабування хмарного застосунку та підвищує їх ефективність.

3. Вперше розроблено інформаційну технологію масштабування хмарного застосунку, яка забезпечує аналіз інформації про стан хмарного застосунку та реалізує удосконалений метод прийняття рішень щодо його масштабування, що дозволило підвищити ефективність функціонування хмарного застосунку.

Практичне значення одержаних результатів полягає у тому, що:

1. Розроблено алгоритм прогнозування часу виконання мережевого запиту до хмарного застосунку.

2. Розроблено алгоритм визначення локального максимуму часового ряду кількості мережевих запитів до хмарного застосунку.

3. Розроблено інформаційну систему автоматизованого масштабування хмарного застосунку, що включає в себе модулі збору статистичної інформації, прогнозування кількості мережевих запитів та прийняття рішень щодо масштабування, а також інтерфейс роботи з хмарною платформою.

4. Розроблено симулятор роботи хмарного застосунку.

Результати дисертаційного дослідження впроваджено в ТОВ "СДМ Україна" у вигляді інформаційної системи масштабування хмарного застосунку автоматизації проведення мозкових штурмів «BrainTank» (№ 47/4 від 28.10 2015 р.).

Особистий внесок здобувача.

У друкованих працях, опублікованих у співавторстві, автору дисертації належать: аналіз ефективності застосування моделей часових рядів для прогнозування стану хмарного застосунку [1], визначення залежності втрати хмарним застосунком користувачів від завантаженості його інфраструктури, оцінювання ефективності функціонування хмарного застосунку в залежності від технології його масштабування [2], алгоритм автоматизованого прийняття рішень щодо масштабування хмарного застосунку [3], розробка архітектури інформаційної технології масштабування хмарного застосунку, граф станів інфраструктури хмарного застосунку [4, 14], побудова архітектури симулятора хмарного застосунку [5], інформаційна модель масштабування хмарного застосунку, критерій класифікації поточного режиму його роботи [6], використання моделі хмарного застосунку як системи масового обслуговування для прогнозування часу виконання мережевих запитів [7], тестування роботи симулятора хмарного застосунку [8], аналіз систем масштабування хмарних застосунків, модель хмарного застосунку як системи масового обслуговування [9], огляд способів опису девіантних законів розподілу, побудова функції генерації випадкових величин на основі графіка функції щільності розподілу [11, 12].

Апробація результатів дисертації. Результати проведеного дослідження були апробовані на 3-ій міжнародній конференції з автоматичного управління та інформаційних технологій ICASIT-2015 11-15 грудня 2015 р., м. Київ, НТУУ «КПІ»; конференції «Наукові підсумки 2015 року», 22 грудня 2015 р., м. Харків, ПП Технологічний центр; 4-й міжнародній науково-практичній конференції «Сучасні інформаційні технології», 25-27 травня 2016 р., м. Суми, СумДУ; XVI міжнародній науково-технічній конференції «Вимірювальна та обчислювальна техніка в технологічних процесах» 10 – 15 червня 2016 р., м. Одеса, Одеська національна академія зв'язку ім. О.С.Попова, XLIV та XLV науково-технічних конференціях професорсько-викладацького складу, співробітників та студентів ВНТУ, 2015, 2016 р.р., Вінниця, ВНТУ.

Публікації. Матеріали досліджень опубліковані у 14 друкованих працях, з них 7 статей у виданнях що входять до переліку фахових видань, в тому числі 1 стаття у науковому журналі, що входить до наукометричної бази SCOPUS, 1 стаття у іноземному фаховому журналі, 2 свідоцтва на реєстрацію авторського права на комп'ютерну програму, 5 матеріалів та тез доповідей науково-технічних та науково-практичних конференцій.

Структура і обсяг дисертації. Дисертаційна робота складається зі вступу, 4-х розділів, основних висновків по роботі, переліку використаних джерел (161 бібліографічне посилання, 16 сторінок) та додатків (20 сторінок). Загальний обсяг роботи, в якому викладено основний зміст, складає 131 сторінку і містить 34 рисунки, 11 таблиць. Повний обсяг дисертації – 173 сторінки.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі обґрунтовано актуальність теми дисертації, зазначено її зв'язок з науковими програмами, планами та темами, сформульовано мету та напрямок досліджень, відзначено наукову новизну та практичне значення одержаних результатів, надано інформацію про їх впровадження, апробацію та публікації.

У першому розділі проведено аналіз методів та засобів масштабування хмарного застосунку. Встановлено, що найпоширенішим методом масштабування хмарного застосунку є масштабування на основі реактивних правил, яке дозволяє змінювати кількість обчислювальних ресурсів при фіксуванні зміни навантаження на хмарний застосунок та реалізовано в більшості хмарних платформ. Поширення набули методи предиктивного масштабування, що базуються на прогнозуванні стану хмарного застосунку. У випадку змінних піків навантаження розглянуті підходи показують незадовільні результати масштабування, це пояснюється стрімкою зміною потреби у обчислювальних ресурсах, що не відповідає шаблону нормального функціонування хмарного застосунку. Більшість з розглянутих підходів не враховують залежності часу проведення масштабування від його умов, деякі з них мають обмежене застосування через високі технічні вимоги, наприклад, потребу у можливості зміни черги виконання запитів або можливість внесення змін до коду хмарного застосунку. Також, часто ігнорується можливість проведення вертикального масштабування через те, що не всі хмарні платформи його підтримують. Таким чином, означені методи та підходи до масштабування при їх застосуванні до хмарного застосунку потребують удосконалення.

Проведено аналіз платформ за допомогою яких може бути здійснено прогнозування часового ряду мережеских запитів до хмарного застосунку. Зроблено висновок про те, що пакет SAS/ETS найкраще підходить для вибору моделі, так як він дозволяє використовувати велику кількість потенційних моделей та містить гнучку систему автоматизованого вибору моделі часового ряду та її параметрів. Проведено огляд методів прогнозування часових рядів, що використовуються при прогнозуванні стану хмарного застосунку. Проаналізовано методи класифікації режимів роботи хмарного застосунку, що вирішують такі задачі як ідентифікація піків навантаження та визначення

напряму тренду. Зроблено висновок про те, що для розв'язання задачі класифікації режимів роботи хмарного застосунку слід застосовувати методи, що не потребують навчальної вибірки та оптимізовані для швидкого виявлення піків навантаження. Досліджено основні методи оцінювання ефективності функціонування хмарного застосунку. Показано, що для оцінювання ефективності здебільшого використовуються одиничні метрики роботи хмарного застосунку. Проте, використання таких метрик не дозволяє безпосередньо побудувати критерій оцінювання ефективності, що визначається на комбінації швидкодії роботи хмарного застосунку та вартості утримання інфраструктури хмарної платформи. Тому, доцільною є задача отримання співвідношення, яке дозволяє оцінити економічний ефект зміни швидкодії роботи хмарного застосунку з метою побудови критерію ефективності функціонування хмарного застосунку.

Отже, проведено аналіз методів та засобів масштабування хмарного застосунку, існуючих платформ прогнозування кількості мережевих запитів, методів класифікації режимів роботи хмарного застосунку та методів оцінювання ефективності функціонування хмарного застосунку. На основі проведеного аналізу визначено основні задачі дослідження.

У другому розділі запропоновано інформаційну модель масштабування хмарного застосунку, яка відрізняється моделлю хмарного застосунку як системи масового обслуговування, критерієм класифікації поточного режиму роботи хмарного застосунку та графом станів його інфраструктури, що дозволяє здійснювати оцінювання поточного і прогнозування майбутніх станів хмарного застосунку.

Сформовано критерій класифікації поточного режиму роботи хмарного застосунку, що на основі методу Пейжда–Хінклі та календаря подій, пов'язаних з роботою хмарного застосунку, відносить поточний режим роботи до одного з трьох класів з метою підвищення точності прогнозування завантаженості хмарного застосунку. Календар подій при цьому представлений множиною пар

$$C = \{c_1, c_2, \dots, c_n\}, \quad (1)$$

де $c_i = (t_i, d_i)$ – подія, представлена часом початку t_i та тривалістю d_i . Настання події за допомогою календаря може бути ідентифіковане при виконанні умови ε_i :

$$\varepsilon_i = (T \geq t_i) \wedge (T < t_i + d_i), \quad (2)$$

$$(t_i, d_i) = c, \exists c \in C,$$

де T – відлік часу, для якого здійснюється класифікація.

Критерій класифікації базується на використанні двох величин магнітуди визначення події за допомогою календаря:

$$m_{T,c0} = \sum_{t=t_0}^T (x_t - \bar{x}_T - \partial_{c0}), \quad (3)$$

$$m_{T,c1} = \sum_{t=t_0}^T (x_t - \bar{x}_T - \partial_{c1}),$$

$$\begin{aligned}
m_{T,c0} - M_{T,c0} &> \pi_{c0} | \neg \varepsilon, \\
m_{T,c1} - M_{T,c1} &> \pi_{c1} | \varepsilon, \\
\partial_{c0} &> \partial_{c1}, \\
\varepsilon &= \varepsilon_1 \vee \varepsilon_2 \vee \varepsilon_3 \vee \dots,
\end{aligned}$$

де \bar{x}_T – середнє значення часового ряду в момент часу T , ($t_0 = 1$) – індекс першого елемента часового ряду, x_t – значення часового ряду в момент часу t , $m_{T,c0}$, $M_{T,c0}$, π_{c0} – поточне агреговане значення часового ряду, мінімальне агреговане значення та порогове значення для будь-якої точки T часового ряду, яка не потрапляє в календар подій, $m_{T,c1}$, $M_{T,c1}$, π_{c1} – поточне агреговане значення часового ряду, мінімальне агреговане значення та порогове значення для будь-якої точки T часового ряду, яка потрапляє в календар подій. ∂_{c0} та ∂_{c1} – магнітуди для двох вищезгаданих режимів.

Розроблено модель системи масового обслуговування, що описує роботу хмарного застосунку, на основі якої було отримане співвідношення для розрахунку середнього часу виконання мережевого запиту. Багатоканальна система масового обслуговування з необмеженою чергою містить нескінченну кількість станів ($D_0, D_1, \dots, D_\Lambda, D_{\Lambda+1}, \dots$), де перші Λ станів відповідають хмарному застосунку у якого є незадіяні процесорні ядра та черга запитів пуста. Граф станів хмарного застосунку як системи масового обслуговування зображений на рисунку 1.

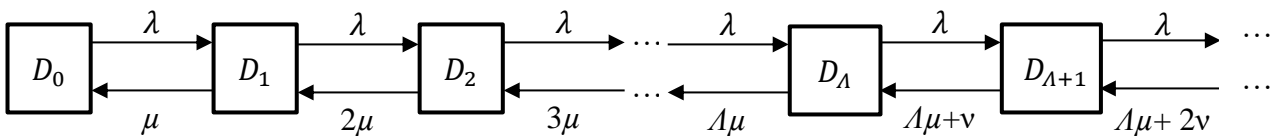


Рисунок 1 – Граф станів хмарного застосунку як системи масового обслуговування

Потік запитів послідовно переводить систему з будь якого стана у стан, що знаходиться правіше на рисунку з постійною інтенсивністю λ . Інтенсивність потоку обслуговування зростає зі збільшенням кількості задіяних каналів від μ до $\Lambda\mu$, коли задіяні всі канали. Також на систему діє потік відхилених запитів, який переводить систему зі стану D_i у стан $D_{(i-1)}$. Інтенсивність потоку відхилених запитів зростає при збільшенні черги на v для кожного наступного стану СМО. Відповідно до граничної теореми про складання потоків, сумарний потік обслуговування становить $i\mu$ при $i \leq \Lambda$ та $(\Lambda\mu + (i - \Lambda)v)$ при $i > \Lambda$, де i – номер стану СМО.

Знайдено імовірність того, що через малий проміжок часу Δt починаючи з момента t система буде знаходитись в стані D_0 , цю імовірність позначено за $p_0(t + \Delta t)$. Система може перейти в стан D_1 при надходженні нового мережевого запиту з імовірністю $(\lambda\Delta t + o(\Delta t))$. Імовірність протилежної події

(новий запит не надійде) складає $(1 - \lambda\Delta t + o(\Delta t))$. Також система може перейти зі стану D_1 в стан D_0 з імовірністю $(\mu\Delta t + o(\Delta t))$ відповідно:

$$p_0(t + \Delta t) = p_0(t)(1 - \lambda\Delta t) + p_1(t)\mu\Delta t + o(\Delta t), \quad (4)$$

$$\frac{p_0(t + \Delta t) - p_0(t) - o(\Delta t)}{\Delta t} = p_1(t)\mu - p_0(t)\lambda. \quad (5)$$

За рахунок того, що проміжок часу Δt малий, а потік вхідних запитів – ординарний можна вважати що інші імовірності не перевищують порядок малості $o(\Delta t)$. При $\Delta t \rightarrow 0$ отримуємо диференціальне рівняння

$$\frac{dp_0}{dt} = \mu p_1 - \lambda p_0. \quad (6)$$

Отримані співвідношення об'єднано в систему диференціальних рівнянь:

$$\left\{ \begin{array}{l} \frac{dp_0}{dt} = \mu p_1 - \lambda p_0; \\ \frac{dp_1}{dt} = 2\mu p_2 - (\lambda + \mu)p_1 + \lambda p_0; \\ \dots \\ \frac{dp_i}{dt} = (i + 1)\mu p_{(i+1)} - (\lambda + i\mu)p_i + \lambda p_{(i-1)}, i < \Lambda; \\ \dots \\ \frac{dp_i}{dt} = (\Lambda\mu + (i - \Lambda + 1)\nu)p_{(i+1)} - (\lambda + \Lambda\mu + (i - \Lambda)\nu)p_i + \lambda p_{(i-1)}, i \geq \Lambda. \\ \dots \end{array} \right. \quad (7)$$

Знайдено середній час перебування мережевого запиту у СМО хмарного застосунку, що складається з часу обробки мережевого запиту та часу перебування в черзі, який в свою чергу дорівнює часу обробки одного мережевого запиту помноженому на кількість запитів в черзі:

$$\bar{u} = \bar{T}_{об} + \eta \bar{T}_{оч} = \frac{1}{\mu} + \bar{\eta} \frac{1}{\mu} = \frac{\bar{\eta} + 1}{\mu}, \quad (8)$$

де $\bar{\eta}$ – середня кількість мережевих запитів у черзі, \bar{u} – середній час виконання мережевого запиту.

Параметр η може приймати нескінченну кількість дискретних випадкових значень $0, 1, 2, \dots$ з імовірностями p_0, p_1, p_2, \dots . Математичне сподівання цієї випадкової величини може бути задане таким виразом:

$$\bar{\eta} = 0p_0 + 1p_1 + 2p_2 + \dots + ip_i + \dots, \quad (9)$$

цей ряд сходиться та може бути наближено обчислений як сума перших m членів ряду та доданку остачі

$$\bar{\eta} = \sum_{i=0}^m ip_i + \frac{(\lambda/\mu)^A (\lambda/\nu)^m}{\Lambda! m!} e^{\frac{\lambda}{\nu}}. \quad (10)$$

Удосконалено метод прийняття рішень щодо масштабування хмарного застосунку, що на основі інформації про можливі піки навантаження на хмарний застосунок, його стан, побудованого на основі цього стану прогнозу роботи хмарного застосунку, набору правил реактивного масштабування та оцінки ефективності функціонування хмарного застосунку здійснює вибір оптимального стану інфраструктури хмарного застосунку та, за потреби, генерує рішення щодо вертикального або горизонтального масштабування з метою переведення хмарного застосунку в розрахований стан. Удосконалений метод дозволяє враховувати інформацію про можливі піки навантаження та збільшити потужність множини доступних способів масштабування хмарного застосунку. Розроблено оціночну функцію варіантів масштабування дозволяє отримати оцінку ефективності функціонування хмарного застосунку, що базується на вартості утримання інфраструктури хмарного застосунку та на оцінці кількості користувачів, що перестануть користуватися хмарним застосунком через його перевантаженість.

$$\begin{aligned} Y(e_{ij})|t_{ij} \leq h = & P_i t_{ij} + P_j (t_{\text{спок}} - t_{ij}) + \\ & + P_{\text{втрат}} \left(\sum_{t=1}^{t_{ij}} v_t F(M(u_{it})) + \sum_{t=t_{ij}+1}^h v_t F(M(u_{jt})) + \right. \\ & \left. + (t_{\text{спок}} - h) v_h F(M(u_{jh})) \right), \end{aligned} \quad (11)$$

де $Y(e_{ij})$ – значення оціночної функції, e_{ij} – дуга графу станів інфраструктури хмарного застосунку, h – горизонт прогнозування, u_{ij} – час виконання мережевого запиту хмарним застосунком, t_{ij} – тривалість операції масштабування, $M(u)$ – математичне сподівання часу виконання мережевого запиту, $t_{\text{спок}}$ – проміжок часу спокою після масштабування впродовж якого нові операції масштабування не здійснюються, $F(\tau)$ – функція розподілу імовірностей, що представляє залежність імовірності втрати користувача від часу виконання мережевого запиту, $P_{\text{втрат}}(e)$ – вартість втрати одного користувача, v_t – кількість користувачів хмарного застосунку в період часу $(t;t+1)$.

Для випадку коли $t_{ij} > h$:

$$Y(e_{ij})|t_{ij} > h = P_i t_{ij} + P_j (t_{\text{спок}} - t_{ij}) + \quad (12)$$

$$+P_{\text{втрат}} \left(\sum_{t=1}^h v_t F(M(u_{it})) + (t_{ij} - h)v_h F(M(u_{ih})) \right. \\ \left. + (t_{\text{спок}} - t_{ij})v_h F(M(u_{jh})) \right).$$

Формальний опис правила масштабування матиме такий вигляд:

Якщо $t > t_{\text{гран}}$, то здійснити масштабування $e_{\min} | Y(e_{\min}) = \min_{e_{ij}} Y(e_{ij})$.

Якщо $e_{\min} \cdot S' \neq S_i$ або $e_{\min} \cdot \Delta N \neq 0$, то $t_{\text{гран}} \leftarrow t + t_{\text{спок}}$.

Розроблено критерій ефективності функціонування хмарного застосунку на основі інформації про час виконання мережевого запиту, кількість користувачів та вартість утримання інфраструктури. Запропонований критерій дозволяє співставляти та комбінувати різні метрики роботи хмарного застосунку. За допомогою отриманого критерію можна проводити оцінку ефективності функціонування хмарного застосунку при використанні різних стратегій масштабування та порівнювати доцільність використання методів масштабування.

Остаточне співвідношення, що дозволяє розрахувати значення критерія ефективності функціонування хмарного застосунку має вигляд:

$$W = \frac{P_{\text{інфр}} + P_{\text{втрат}} * \sum_{t=t_0}^{t_{\text{кінц}}} v_t q_{\text{втрат}}}{t_{\text{кінц}} - t_0 + 1}, \quad (13)$$

де $P_{\text{інфр}}$ – втрати на утримання інфраструктури хмарного застосунку, $P_{\text{втрат}}$ – вартість втрати одного користувача, $q_{\text{втрат}}$ – імовірність того, що користувач перестане користуватися хмарним застосунком.

У третьому розділі розроблено інформаційну технологію масштабування хмарного застосунку, яка забезпечує аналіз інформації про стан хмарного застосунку, а також реалізує удосконалений метод прийняття рішень щодо його масштабування, що дозволило підвищити ефективність функціонування хмарного застосунку. Основні компоненти розробленої технології зображені на рисунку 2.

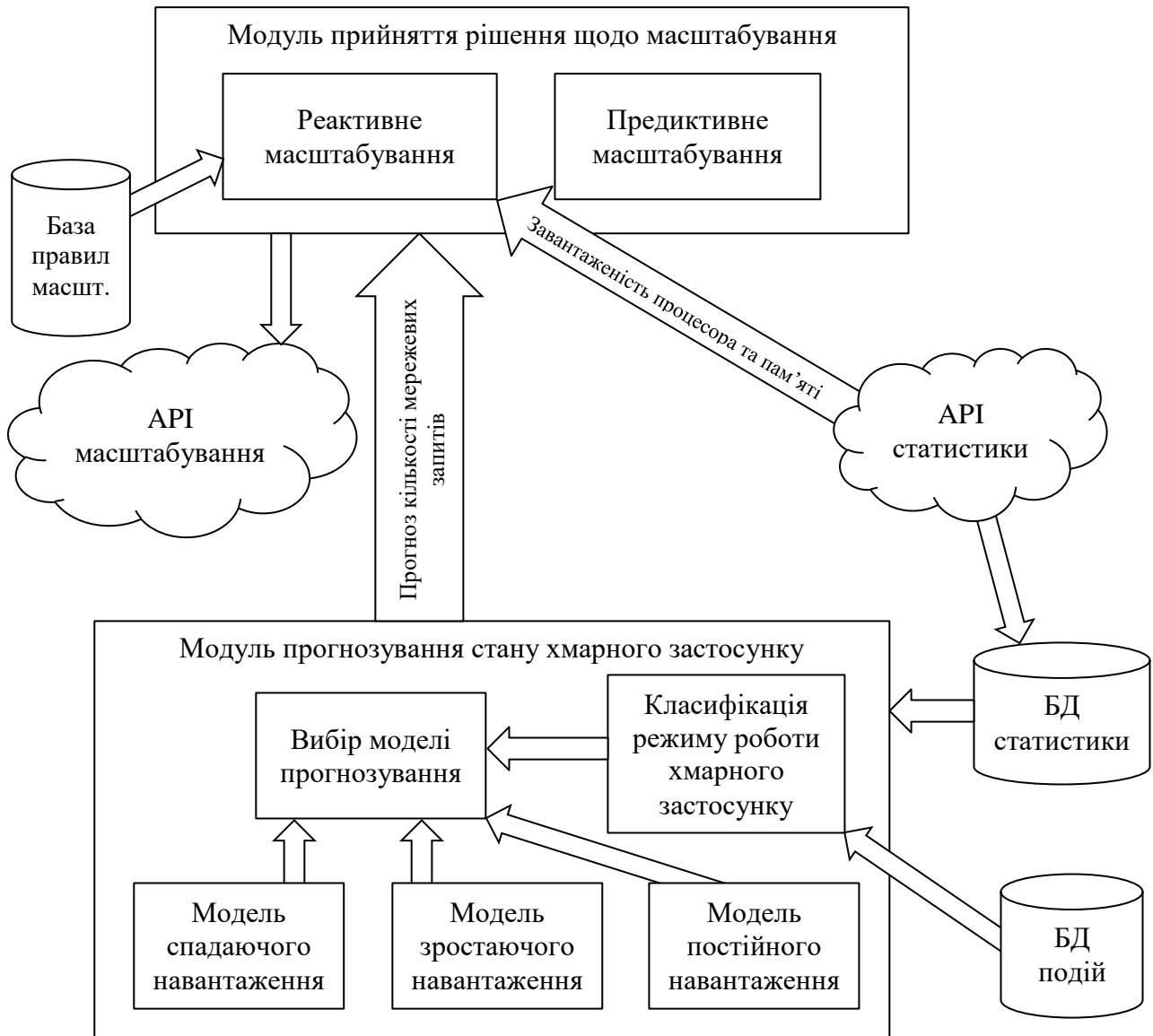


Рисунок 2 – Архітектура інформаційної технології масштабування хмарного застосунку

Проведено проектування архітектури модулів інформаційної технології масштабування хмарного застосунку, що здійснюють прогнозування кількості мережевих запитів та прийняття рішення щодо масштабування хмарного застосунку. Визначено основні складові модулів, сценарії їх взаємодії та зовнішні залежності необхідні для їх функціонування. Сформовано рекомендації до формату збереження налаштувань роботи модулів. Обмін даними між блоками модуля масштабування хмарного застосунку зображено на рисунку 3.

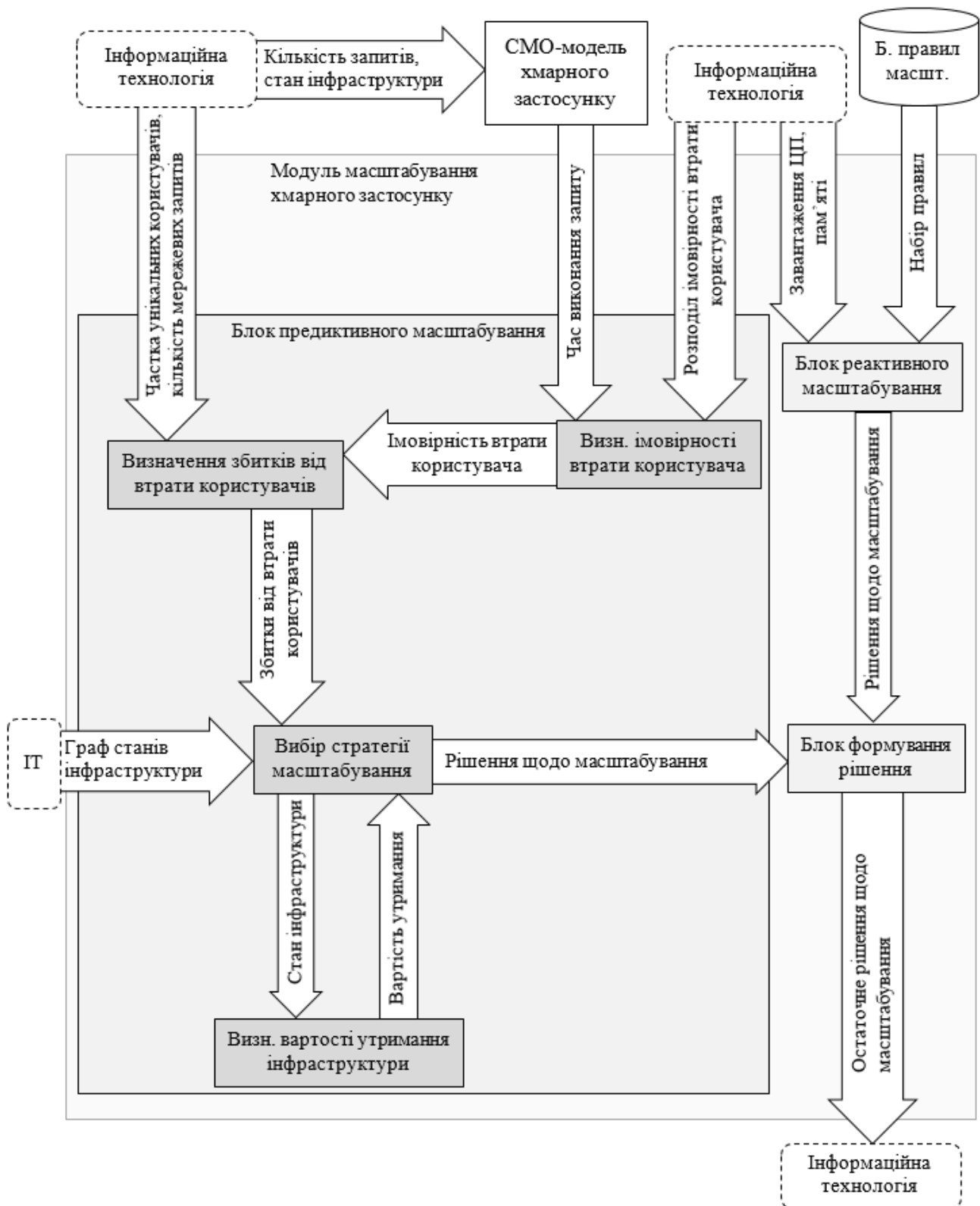


Рисунок 3 – Обмін даними між блоками модуля масштабування хмарного застосунку

Розроблено алгоритм прогнозування часу виконання мережевого запиту до хмарного застосунку на основі моделі системи масового обслуговування. Проведено аналіз стаціонарності часового ряду мережевих запитів до хмарного застосунку під час піку навантаження за допомогою розрахунку коефіцієнтів автокореляції. Визначено, що ряд не є стаціонарним, проте стаціонарною є перша

різниця часового ряду, що обмежує кількість моделей прогнозування часових рядів, що можуть бути застосовані для прогнозування кількості мережових запитів до хмарного застосунку.

Визначено вимоги до функціональних можливостей симулятора роботи хмарного застосунку, що дозволило пришвидшити та здешевити проведення порівняння ефективності використання різних методів масштабування за рахунок моделювання роботи хмарного застосунку та можливості відмовитись від спеціального обладнання для проведення тестів та використання платних сервісів навантажувального тестування. Взаємодія компонентів симулятора роботи хмарного застосунку зображена на рисунку 4.

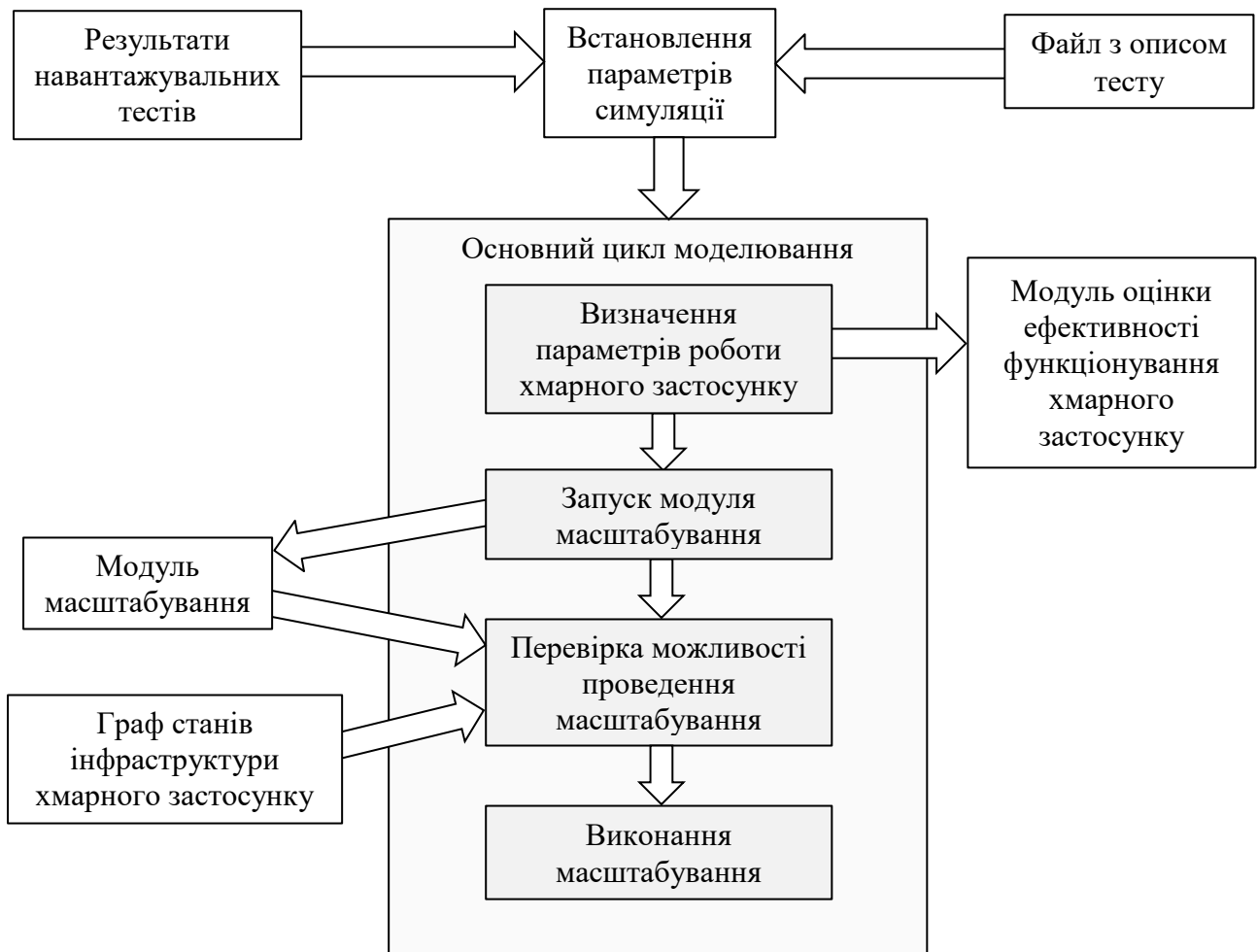


Рисунок 4 – Взаємодія компонентів симулятора роботи хмарного застосунку

Проведено порівняння точності моделей прогнозування часових рядів при прогнозуванні часового ряду кількості мережових запитів до хмарного застосунку в умовах піків навантаження та в умовах проміжків між піками. Визначено, що в зростаючій фазі піку навантаження доцільно використовувати модель експоненційного згладжування, при спадаючій фазі піку навантаження – модель експоненційного згладжування з урахуванням тренду, при відсутності піків навантаження – модель авторегресії–проінтегрованого ковзного середнього з урахуванням періодичності.

У четвертому розділі наведено практичну реалізацію інформаційної технології масштабування хмарного застосунку та здійснено вибір цільової платформи для її розміщення. При цьому враховано такі фактори як простота розгортання та можливість програмного масштабування хмарних застосунків. Визначено, що технологія Microsoft Azure App Service (Web apps) найкраще відповідає поставленим вимогам, тому інформаційну систему автоматизованого масштабування хмарного застосунку доцільно реалізовувати саме в контексті цієї технології.

На базі технології .NET Framework було розроблено модулі прогнозування кількості мережевих запитів та масштабування як складові частини інформаційної системи автоматизованого масштабування хмарного застосунку, що на основі даних про використання хмарного застосунку, отриманих за допомогою бібліотеки Azure Web Sites Management та бази даних подій, пов'язаних з піками навантаження дозволяють вибрати стратегію масштабування хмарного застосунку. Діаграма класів модулю прогнозування кількості мережевих запитів, що надходять до хмарного розрахунку наведена на рисунку 5. Модуль розроблено мовою C# з використанням .Net Framework 4.5.1. Розроблено інтерфейс користувача інформаційної системи автоматизованого масштабування хмарного застосунку, що надає можливості налаштування в графічному режимі основних параметрів роботи системи за допомогою використання технології ASP.NET MVC та програми GenGraph.

Проведено порівняння середньоквадратичного відхилення прогнозу кількості мережевих запитів до хмарного застосунку отриманого за допомогою класичних моделей прогнозування часових рядів та за допомогою комбінованого застосування моделей на основі класифікації режиму роботи хмарного застосунку. Показано, що комбіноване застосування моделей дозволяє збільшити точність прогнозування часового ряду мережевих запитів до хмарного застосунку.

Розроблено симулятор роботи хмарного застосунку, що моделює функціонування хмарного застосунку із заданим профілем навантаження. В основі симулятора лежать функціональні залежності метрик роботи хмарного застосунку від стану його інфраструктури та кількості мережевих запитів, отримані шляхом апроксимації експериментальних даних, що базуються на тестах навантаження. Розроблений симулятор дозволяє пришвидшити процес порівняння стратегій масштабування хмарного застосунку та дає можливість проведення більшої кількості експериментів.

Проведено порівняння ефективності функціонування хмарного застосунку при використанні розробленої інформаційної технології та за технологією, що використовує класичні підходи до масштабування. Результати тесту, що відтворює роботу хмарного застосунку з низькою частотою мережевих запитів та зі змінними піками навантаження показали, що сумарне значення критерію ефективності склало 13,6 (0,00125 за хв.) для класичного підходу та 12,51 (0,00118 за хв.) для інформаційної технології масштабування хмарного застосунку.

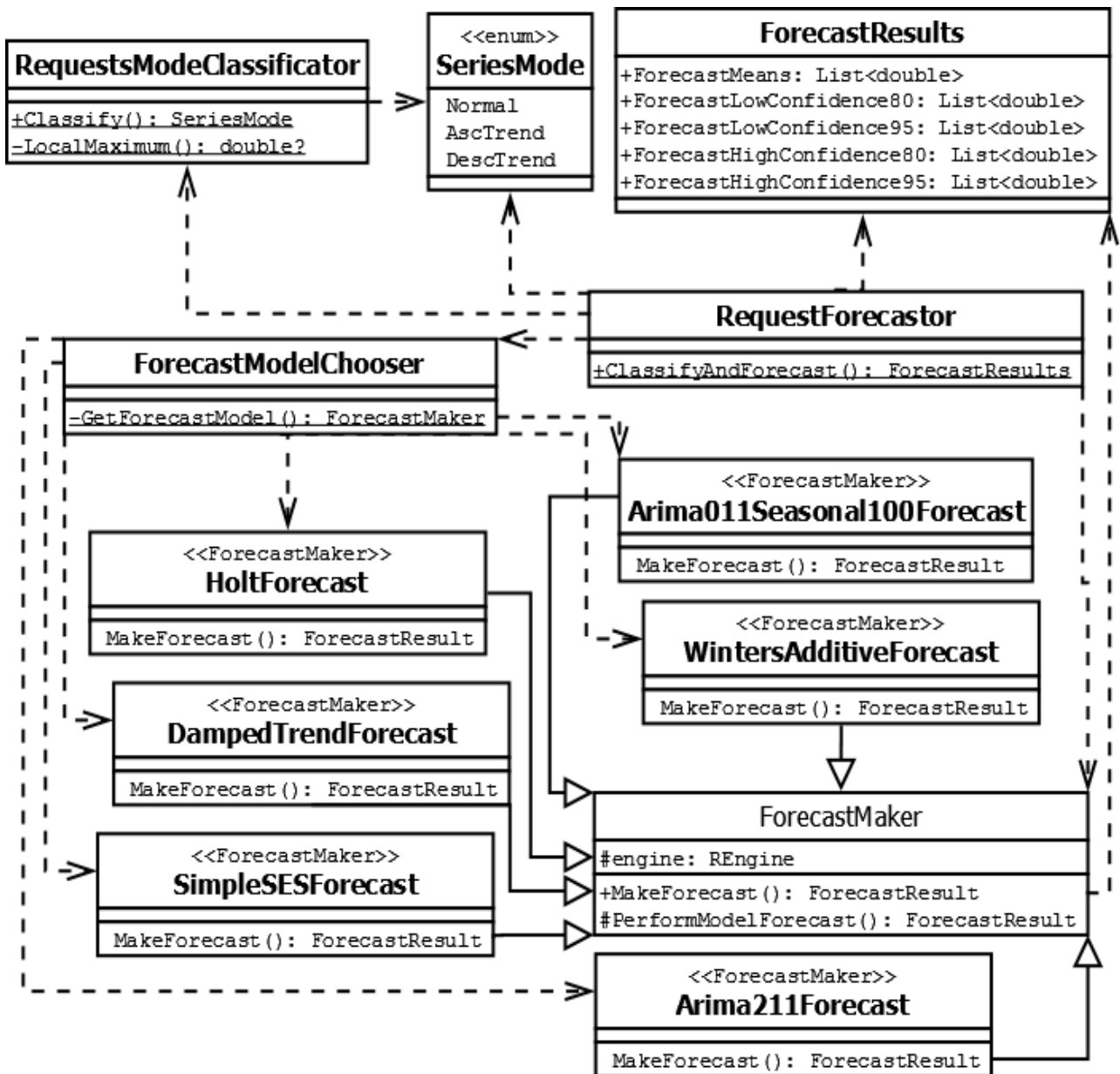


Рисунок 5 – UML-діаграма класів модуля прогнозування кількості мережевих запитів

Для тесту, що відтворює інтенсивні пікові навантаження сумарне значення критерію ефективності склало 3.12 (0,0044 за хв.) для класичного підходу та 2,75 (0,0039 за хв.) для інформаційної технології масштабування хмарного застосунку.

Таким чином, використання інформаційної технології масштабування хмарного застосунку дозволяє зменшити експлуатаційні витрати на 8% в режимі нормального функціонування та на 12% в режимі інтенсивних піків навантаження. Результати інших тестів наведені в таблиці 1.

Таблиця 1 – Порівняння результатів тестів навантаження

Опис тесту навантаження	Інформаційна технологія масштабування		Реактивне масштабування		Приріст ефективності з використанням інформаційної технології
	Значення критерію ефективності	Середнє за 1 хв.	Значення критерію ефективності	Середнє за 1 хв.	
1	2	3	4	5	6
Усі дані	12,51	0,00118	13,6	0,00125	8%
Подвійний пік	2,75	0,0039	3,12	0,0044	12%
Подвійний пік №2	1,27	0,00253	1,77	0,0035	28%
Малий подвійний пік	0,758	0,00094	0,74	0,00093	-2%
Режим з високою частотою запитів	2,34	0,00086	2,34	0,00086	0%
Одиничний пік	1,25	0,00313	1,67	0,00417	25%
Один день	2,14	0,00149	2,56	0,00178	16%
Режим з низькою частотою запитів	2,24	0,00075	2,24	0,00075	0%

Як видно з таблиці 1, застосування інформаційної технології масштабування хмарного застосунку збільшує ефективність використання хмарного застосунку, при чому в найбільшій мірі ефективність зростає під час піків навантаження. В середньому, використання інформаційної технології масштабування збільшує ефективність функціонування хмарного застосунку на 8%, а при високій частоті піків навантаження – до 28%.

Досліджено результати застосування інформаційної системи автоматизованого масштабування хмарного застосунку у ТОВ «СДМ Україна» для масштабування системи проведення мозкових штурмів BrainTank. Проведено порівняння роботи запропонованої інформаційної системи з системою автоматизації CloudMonix. Показано, що розроблена інформаційна система автоматизованого масштабування хмарного застосунку дозволяє збільшити ефективність функціонування хмарного застосунку на 10,5%.

У додатках наведено результати оцінювання моделей прогнозування часового ряду мережевих запитів до хмарного застосунку, порівняння значень критерію ефективності функціонування хмарного застосунку із застосуванням розробленої інформаційної технології та системи Cloudmonix, лістинг програмної реалізації розробленої інформаційної технології, інструкція користувача інформаційної системи масштабування хмарного застосунку, акти впровадження наукових розробок.

ЗАГАЛЬНІ ВИСНОВКИ ПО РОБОТІ

1. Проведено аналіз методів та засобів масштабування хмарного застосунку. Встановлено, що найпоширенішим методом масштабування хмарного застосунку є масштабування на основі реактивних правил, що дозволяє змінювати кількість обчислювальних ресурсів при фіксації зміни навантаження на хмарний застосунок, цей метод реалізовано в більшості хмарних платформ. Також використовуються методи предиктивного масштабування, що базуються на прогнозуванні стану хмарного застосунку. Класичні підходи до масштабування хмарного застосунку у випадку змінних піків навантаження часто дають незадовільні результати, оскільки це не відповідає шаблону його нормального функціонування та може призвести до стрімкої зміни потреби в обчислювальних ресурсах. Таким чином, розробка інформаційної технології масштабування хмарного застосунку, що базується на прогнозуванні його стану є актуальною теоретичною і прикладною задачею.

2. Запропоновано інформаційну модель масштабування хмарного застосунку, яка відрізняється моделлю хмарного застосунку як системи масового обслуговування, критерієм класифікації поточного режиму роботи хмарного застосунку та графом станів його інфраструктури. Інформаційна модель масштабування хмарного застосунку дозволяє отримати інформацію, необхідну для прогнозування стану хмарного застосунку, а саме: прогноз середнього часу виконання мережевого запиту та особливості поточного режиму роботи хмарного застосунку.

3. Удосконалено метод прийняття рішень щодо масштабування хмарного застосунку, що на основі інформації про можливі піки навантаження на хмарний застосунок, його стан, побудованого на основі цього стану прогнозу роботи хмарного застосунку, набору правил реактивного масштабування та оцінки ефективності функціонування хмарного застосунку здійснює вибір оптимального стану інфраструктури хмарного застосунку та, за потреби, генерує рішення щодо вертикального або горизонтального масштабування з метою переведення хмарного застосунку в розрахований стан. Удосконалений метод дозволяє враховувати інформацію про можливі піки навантаження та збільшити потужність множини доступних способів масштабування хмарного застосунку. Розроблений критерій ефективності масштабування дозволяє отримати оцінку ефективності функціонування хмарного застосунку, що базується на вартості утримання інфраструктури хмарного застосунку та на оцінці кількості користувачів, що перестануть користуватися хмарним застосунком через його перевантаженість.

4. Розроблено інформаційну технологію масштабування хмарного застосунку, яка забезпечує аналіз інформації про стан хмарного застосунку, реалізує удосконалений метод прийняття рішень щодо його масштабування шляхом використання реактивного та предиктивного підходів до горизонтального та вертикального масштабування, а також використовує список очікуваних піків навантаження та прогноз стану хмарного застосунку, отриманий на основі історичної інформації про його завантаженість, що дозволило підвищити ефективність функціонування хмарного застосунку.

5. Розроблено інформаційну систему автоматизованого масштабування хмарного застосунку на базі запропонованої інформаційної технології, що підтримує інтеграцію з новими та існуючими хмарними застосунками на платформі Microsoft Azure web apps та не потребує модифікації бінарного коду самого хмарного застосунку.

6. Розроблено симулятор роботи хмарного застосунку, що моделює функціонування хмарного застосунку із заданим профілем навантаження. В основі симулятора лежать функціональні залежності метрик роботи хмарного застосунку від стану його інфраструктури та кількості мережевих запитів, отримані шляхом апроксимації експериментальних даних, що базуються на тестах навантаження. Розроблений симулятор дозволяє пришвидшити процес порівняння стратегій масштабування хмарного застосунку та дає можливість проведення більшої кількості експериментів.

7. Проведено порівняння ефективності функціонування хмарного застосунку при використанні розробленої інформаційної технології та за технологією, що використовує класичні підходи до масштабування. Показано, що застосування інформаційної технології масштабування хмарного застосунку збільшує ефективність функціонування хмарного застосунку, при чому в найбільшій мірі ефективність зростає під час піків навантаження. В середньому, використання інформаційної технології масштабування збільшує ефективність функціонування хмарного застосунку на 8%, а при високій частоті піків навантаження – до 28%.

8. Досліджено результати застосування інформаційної системи автоматизованого масштабування хмарного застосунку у ТОВ «СДМ Україна» для масштабування системи проведення мозкових штурмів BrainTank. Проведено порівняння роботи запропонованої інформаційної системи з системою автоматизації CloudMonix. Показано, що розроблена інформаційна автоматизована система масштабування хмарного застосунку дозволяє збільшити ефективність функціонування хмарного застосунку на 10,5%.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Савчук Т. О. Прогнозування кількості мережевих запитів до хмарного застосунку / Т. О. Савчук, А. В. Козачук // Вісник Національного університету «Львівська політехніка». Серія: Інформаційні системи та мережі. – 2015. – № 1 – С. 139-146. – ISSN 0321-0499.
2. Савчук Т. О. Розробка критерія оцінювання ефективності роботи хмарного застосунку / Т. О. Савчук, А. В. Козачук // Східно-Європейський журнал передових технологій. – 2015. – № 2(77). – С. 20 – 26. – ISSN 1729-3774.
3. Савчук Т. О. Автоматизоване прийняття рішень щодо масштабування хмарного застосунку / Т. О. Савчук, А. В. Козачук // Інформаційні технології та комп'ютерна інженерія. – 2015. – №2. – С. 15 – 22. – ISSN 1999-9941.
4. Савчук Т. О. Інформаційна технологія масштабування хмарного застосунку зі змінними піками навантаження / Т. О. Савчук, А. В. Козачук // Технологічний аудит та резерви виробництва. – 2015. – № 5 (25). – С. 4 – 11. – ISSN 2226-3780.
5. Савчук Т. О. Масштабування хмарного додатку з використанням симулятора навантаження його роботи / Т. О. Савчук, А. В. Козачук // Вимірювальна та обчислювальна техніка в технологічних процесах. – 2015. – № 4. – С. 222 – 226. – ISSN 2219-9365.
6. Савчук Т. О. Система моделювання неперервних випадкових величин з графічно заданими особливостями «GenGraph» / Т. О. Савчук, В. В. Колодний, А. В. Козачук // Інформаційні технології та комп'ютерна інженерія. – 2011. – №2. – С. 49-55. – ISSN 1999-9941.
7. Savchuk T. O. Information model of a cloud application scaling with variable load peaks/ Т. О. Savchuk, А. V. Kozachuk// EUREKA: Physics and Engineering. – 2016. – № 3 (4). – С. 38 – 45. – ISSN 2461-4254
8. Савчук Т. О. Розробка симулятора роботи хмарного застосунку / Т. О. Савчук, А. В. Козачук // Технологический аудит и резервы производства. – 2015. – №6/7 (26). – С. 4 – 6. – ISSN 2226-3780
9. Савчук Т. О. Інформаційна технологія масштабування хмарного застосунку з неперіодичними піками навантаження / Т. О. Савчук, А. В. Козачук // Тези XLV науково-технічної конференції професорсько-викладацького складу, співробітників та студентів університету. – 2016. – ВНТУ. – С. 1 – 2.
10. Савчук Т. О. Information technology of cloud app scaling / Т. О. Савчук, А. В. Козачук // ICACIT-2015: Матеріали 3-ї Міжнародної конференції з автоматичного управління та інформаційних технологій, м. Київ, 11-15 грудня 2015 р. – 2015. – С. 108 – 111.
11. Козачук А. В. Cloud Application Working Mode Classification Criteria // Матеріали 4-ї міжнародної науково-практичної конференції «Сучасні інформаційні технології», м. Суми, 25-27 травня 2016 р. – Суми, 2016. – С. 96 – 97. – ISSN 2311-8504.
12. Козачук А. В. Модель хмарного застосунку як системи масового обслуговування // Матеріали XVI міжнародної науково-технічної конференції

«Вимірювальна та обчислювальна техніка в технологічних процесах», м. Одеса, 10 – 15 червня 2016 р. – С. 164 – 166. – ISBN 978-966-330-236-2.

13. Колодний В. В. Комп'ютерна програма "GenGraph"/ В. В. Колодний, А. В. Козачук // Свідоцтво про реєстрацію авторського права на комп'ютерну програму №35767 від 28.02.2011.

14. Савчук Т. О. Система автоматичного масштабування хмарних застосунків з періодичними піками навантаження / Т. О. Савчук, А. В. Козачук // Свідоцтво про реєстрацію авторського права на комп'ютерну програму №62386 від 05.11.2015.

АНОТАЦІЯ

Козачук А. В. Інформаційна технологія масштабування хмарного застосунку, що базується на прогнозуванні його стану. – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук зі спеціальності 05.13.06 – інформаційні технології. Вінницький національний технічний університет, Вінниця, 2016.

Дисертаційна робота присвячена підвищенню ефективності функціонування хмарного застосунку зі змінними піками навантаження за рахунок використання інформаційної технології, що поєднує горизонтальне та вертикальне масштабування з використанням реактивних правил та прогнозу стану хмарного застосунку. Удосконалено метод прийняття рішень щодо масштабування хмарного застосунку, що на відміну від існуючих при поєднанні реактивного та предиктивного підходів дозволяє здійснювати горизонтальне та вертикальне масштабування. Це дозволило враховувати інформацію про змінні піки навантаження та збільшити потужність множини доступних способів масштабування хмарного застосунку. Розроблено інформаційну модель масштабування хмарного застосунку, яка відрізняється моделлю хмарного застосунку як системи масового обслуговування, критерієм класифікації поточного режиму роботи хмарного застосунку та графом станів його інфраструктури, що дозволяє здійснювати оцінювання поточного і прогнозування майбутніх станів хмарного застосунку.

Ключові слова: хмарні обчислення, масштабування хмарного застосунку, інформаційна модель хмарного застосунку, PaaS, ефективність функціонування хмарного застосунку, прогнозування стану хмарного застосунку.

АННОТАЦИЯ

Козачук А. В. Информационная технология масштабирования облачного приложения, основанная на прогнозировании его состояния. - На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.06 - информационные технологии. Винницкий национальный технический университет, Винница, 2016.

Диссертационная работа посвящена повышению эффективности функционирования облачного приложения за счет использования информационной технологии, сочетающей горизонтальное и вертикальное масштабирование с использованием реактивных правил и прогноза состояния облачного приложения.

Разработана модель системы массового обслуживания, описывающая работу облачного приложения, на основе которой было получено соотношение для расчета среднего времени выполнения сетевого запроса.

Усовершенствован метод принятия решений о масштабировании облачного приложения, который на основе информации о возможных пиках нагрузки на облачное приложение; его состоянии; прогнозе работы облачного приложения, построенного на основе этого состояния; правил реактивного масштабирования и с учетом эффективности функционирования облачного приложения осуществляет выбор оптимального состояния инфраструктуры облачного приложения и, при необходимости, генерирует решение о вертикальном или горизонтальном масштабировании с целью формирования облачного приложения необходимого масштаба. Усовершенствованный метод позволяет также увеличить мощность множества доступных способов масштабирования облачного приложения.

Разработанная оценочная функция вариантов масштабирования позволяет определить эффективность функционирования облачного приложения с учетом стоимости содержания инфраструктуры облачного приложения и количества пользователей, которые перестанут им пользоваться как перегруженным.

Разработан критерий эффективности функционирования облачного приложения, учитывающий время выполнения сетевого запроса, количество пользователей и стоимость содержания инфраструктуры. Предложенный критерий позволяет сопоставлять и комбинировать разнородные метрики работы облачного приложения. С помощью полученного критерия можно проводить оценку эффективности функционирования облачного приложения при использовании различных стратегий масштабирования и оценивать целесообразность использования методов масштабирования.

Разработана информационная технология масштабирования облачного приложения, которая обеспечивает анализ информации о состоянии облачного приложения, а также реализует усовершенствованный метод принятия решений о его масштабировании, что позволило повысить эффективность функционирования облачного приложения.

Проведено проектирование архитектуры модулей информационной технологии масштабирования облачного приложения, осуществляющих прогнозирование количества сетевых запросов и принятие решения о масштабировании облачного приложения. Определены основные составляющие модулей, сценарии их взаимодействия и внешние связи, необходимые для их функционирования. Сформированы рекомендации к формату сохранения настроек работы модулей.

Разработан алгоритм прогнозирования времени выполнения сетевого запроса к облачному приложению на основе модели системы массового

обслуживания. Разработан симулятор работы облачного приложения, моделирующий функционирование облачного приложения с заданным профилем нагрузки. В основе симулятора лежат функциональные зависимости метрик работы облачного приложения от состояния его инфраструктуры и количества сетевых запросов, полученные путем аппроксимации экспериментальных данных, основанных на тестах нагрузки. Разработанный симулятор позволяет ускорить процесс сравнения стратегий масштабирования облачного приложения.

Проведено сравнение эффективности функционирования облачного приложения при использовании разработанной информационной технологии и технологии, использующей классические подходы к масштабированию. Результаты теста, который воспроизводит работу облачного приложения с низкой частотой сетевых запросов и со сменными пиками нагрузки показали, что суммарное значение критерия эффективности составило 12,51 (0,00118 за мин.) при использовании предложенной информационной технологии масштабирования облачного приложения и 13,6 (0,00125 за мин.) – при использовании технологии с классическим подходом к масштабированию.

Исследованы результаты применения информационной системы автоматизированного масштабирования облачного приложения в ООО «СДМ Украины» для масштабирования системы проведения мозговых штурмов BrainTank. Проведено сравнение работы предлагаемой информационной системы с системой автоматизации CloudMonix. Показано, что разработанная информационная система автоматизированного масштабирования облачного приложения позволяет повысить эффективность функционирования облачного приложения на 10,5%.

Ключевые слова: облачные вычисления, масштабирование облачного приложения, информационная модель облачного приложения, PaaS, эффективность функционирования облачного приложения, прогнозирования состояния облачного приложения.

ABSTRACT

Kozachuk A. V. Information technology of scaling of a cloud application with variable load peaks. – A manuscript.

The thesis for the getting of degree of Ph.D. in specialty 05.13.06 – information technology. Vinnytsia National Technical University, Vinnytsia, 2016.

The thesis is dedicated to improving the effectiveness of cloud application functioning by usage of an information technology that combines reactive and predictive scaling of a cloud application. Information technology of a cloud application scaling was developed. The method of making decisions about cloud application scaling was proposed, this method unlike existing combines reactive and predictive approach allowing both horizontal and vertical scaling, which made it possible to take into account information about the variable load peaks and to increase the number of available scaling methods. Developed information model of cloud application scaling, is a formal description of cloud application infrastructure and possible transitions

between them and the criterion for classification of cloud application mode with variable load peaks.

Keywords: cloud computing, cloud application scaling, cloud application information model, PaaS, cloud application efficiency, forecasting of cloud application state.

Підписано до друку 25.10.2016 р. Формат 29.7×42 ¼

Наклад 100 прим. Зам. № 2016-230

Віддруковано в комп'ютерному інформаційно-видавничому центрі

Вінницького національного технічного університету
м. Вінниця, вул. Хмельницьке шосе, 95. Тел.: 59-81-59