

6. Пат. 1774279 Российская Федерация, МКИ³ G 01 R 23/00. Устройство для измерения кратковременной нестабильности периода / А. И. Вервейко ; заявитель и патентообладатель Производственное объединение "Черниговский радиоприборный завод". – № 4837835/21 ; заявл. 11.05.90 ; опубл. 07.11.92, Бюл. № 41.

7. А. с. 1723563 СССР, МКИ³ G 05 B 23/02. Измеритель переходных характеристик / А. И. Вервейко, А. П. Коваль, С. Н. Макарук и Ю. С. Шмалый (СССР). – № 4798085/21 ; заявл. 08.12.89 ; опубл. 30.02.92, Бюл. № 12.

8. V. Pavlovsky, A. Verveiko, I. Shkola Reconfigured laboratory complexes in educational process of the Chernihiv state technological university / Advanced Computer Systems and Networks: Design and Application // Proceedings of the 1st International Conference ACSN-2003. – Lviv : Publishing House of Lviv Polytechnic National University. – 2003. – С. 46-47.

9. LabVIEW для радиоинженера: от виртуальной модели до реального прибора. Практическое руководство для работы в программной среде LabVIEW. – М. : ДМК Пресс, 2007. – 400 с.

10. Рубичев Н. А. Измерительные и информационные системы : учебное пособие / Н. А. Рубичев. – М. : Дрофа, 2010. – 334 с.

УДК 004.75

А.М. Волокита, канд. техн. наук

С.В. Пих, магістр

В.В. Воробйов, магістр

Д.С. Вітюк, магістр

Національний технічний університет України «Київський політехнічний інститут», м. Київ, Україна

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ ПРИВАТНОЇ ХМАРНОЇ СИСТЕМИ ДЛЯ ОБЧИСЛЕНЬ КРЕДИТНОГО СКОРИНГУ

А.Н. Волокита, канд. техн. наук

С.В. Пих, магістр

В.В. Воробйов, магістр

Д.С. Вітюк, магістр

Национальный технический университет Украины «Киевский политехнический институт», г. Киев, Украина

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ ПРИВАТНОЙ ОБЛАЧНОЙ СИСТЕМЫ ДЛЯ ВЫЧИСЛЕНИЙ КРЕДИТНОГО СКОРИНГА

Artem Volokyta, PhD in Technical Sciences

Sviatoslav Pikh, master

Vitalii Vorobiov, master

Dmytro Vitiuk, master

National Technical University of Ukraine «Kyiv Polytechnic Institute», Kyiv, Ukraine

THE INVESTIGATION OF THE EFFICIENCY OF PRIVATE CLOUD-SYSTEMS FOR CALCULATION THE SCORECARD COEFFICIENT

Запропоновано приватну хмарну систему для розподілених обчислень коефіцієнтів скорингових карт. Проведено експериментальні дослідження ефективності запропонованої системи.

Ключові слова: приватна хмара, скорингові карти, лінійне програмування, регресійний аналіз.

Предложено частную облачную систему для распределенных вычислений коэффициентов скоринговых карт. Проведены экспериментальные исследования эффективности предложенной системы.

Ключевые слова: приватное облако, скоринговые карты, линейное программирование, регрессионный анализ.

Proposed private cloud system for distributed computing coefficients scorecard. Experimental research on the effectiveness of the proposed system.

Key words: private cloud, scorecards, linear programming, regression analysis.

Вступ. Кредитний скоринг – це метод класифікації різних груп потенційних клієнтів в умовах, коли доступна інформація не про параметри, які поділяють ці групи, а тільки про деякі вторинні змінні. Доступна інформація про потенційних позичальників міститься в заповнених ними анкетах. Такі фактори, як річний дохід, розмір непогаше-

ного боргу, володіння нерухомістю або автомобілем, стаж роботи на останньому місці, вік і т. ін. потенційно пов'язані з кредитоспроможністю і тому можуть виявитися вхідними змінними в моделі [1; 2]. Якщо деяким соціальним характеристикам клієнта присвоїти певні ваги, то кожного нового клієнта можна, на основі його анкети, віднести до певної групи. Набір цих характеристик та відповідних вагових коефіцієнтів, називають скоринговою картою, розроблення якої проводиться на основі статистичного оброблення великих масивів історичних даних по прецедентах. Згідно з Fair & Isaac у ході побудови скорингової моделі можуть бути враховані 50-60 змінних, внаслідок кореляцій у кінцеву модель потрапляють не більше 8-12 [3].

Хмарні обчислення являють собою модель для забезпечення зручного мережевого доступу до загального пулу обчислювальних ресурсів (наприклад, мереж, серверів, систем зберігання даних, програм та послуг), що налаштовуються і які можна швидко виділити і надати з мінімальними управлінськими зусиллями або мінімальним втручанням з боку постачальника послуг [4].

Аналіз останніх досліджень і публікацій. Нині для кредитного скорингу використовуються методи статистики (дискримінантний аналіз, лінійна регресія, логістична регресія, дерева класифікації), дослідження операцій (лінійне програмування, нелінійна оптимізація) і штучного інтелекту (нейронні мережі, експертні системи, генетичні алгоритми, методи найближчих сусідів, байєсовські мережі, логіко-імовірнісні методи). Зазначені методи можуть застосовуватися як окремо, так і в різних комбінаціях.

Приватні хмари поєднують у собі гнучкість і цінові переваги «хмари» з контролем за наданням послуг і безпекою (рис. 1).

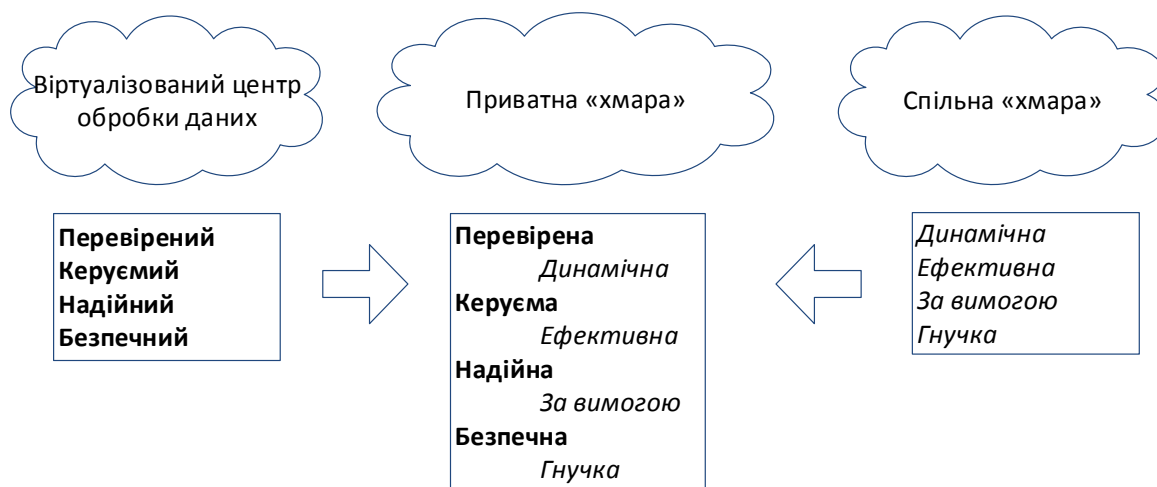


Рис. 1. Характеристики приватної хмарної хмари

Приватні хмари пропонують шлях міграції, дозволяючи віртуалізувати і модифікувати необхідні додатки з потрібним темпом. Це дає можливість захистити інвестиції в інфраструктуру, додатки та інформацію, забезпечивши при цьому ефективно і гнучке використання даних ресурсів [5; 6; 7].

Постановка задачі пошуку коефіцієнтів скорингових карт. Проблему кредитного скорингу можна розглядати як завдання класифікації: знаючи відповіді на питання анкети $x \in A$, визначити, до якої групи належить позичальник: $x \in A_g$ для «хороших клієнтів», і $x \in A_b$ для «поганих». При цьому необхідно розуміти, що абсолютно точна класифікація принципово неможлива.

Згідно з проведеними дослідженнями [8], до теперішнього часу не отримано відомостей про значні переваги якого-небудь з традиційних методів скорингу в точності одержуваних результатів, тобто рівні похибок під час використання цих методів є порі-

вняними. Так, для одних методів (лінійна регресія) бажано використання рівних часток «поганих» і «хороших», тоді як інші методи (дерева класифікації, байєсовські мережі) вимагають, щоб вибірка відображала реальне співвідношення «поганих» і «хороших» клієнтів. Більшість статистичних методів призводять до побудови правила класифікації, заснованого на лінійній скорингової функції.

У процесі реалізації будемо розглядати 2 методи: регресійний аналіз як один з найпоширеніших методів та лінійне програмування, що дає можливість включити в програму додаткові обмеження [9; 10].

Рівняння множинної регресії (з багатьма змінними) у вигляді:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m + \varepsilon,$$

де $X = (X_1, X_2, \dots, X_m)$ – вектор незалежних (пояснюючих) змінних; $\beta = (\beta_1, \beta_2, \dots, \beta_n)$ – вектор параметрів (що визначаються); ε – випадкова помилка (похибка); Y – залежна (пояснювальна) змінна. Для оцінювання параметрів β використовується метод найменших квадратів.

У методі лінійного програмування є набір даних, що включає в себе відповіді на питання x_i для кожного з N клієнтів і індикатори Y_i . Лінійна функцію $s(x)$, яка розділяє «хороших» і «поганих» клієнтів, шукається з міркувань мінімізації помилки a_i .

$$\begin{cases} a_1 + \dots + a_N \rightarrow \min \\ w_1 x_{i1} + \dots + w_m x_{im} \geq c - a_i, & Y_i = 0, \\ w_1 x_{i1} + \dots + w_m x_{im} \leq c + a_i, & Y_i = 1, \\ a_i > 0 \end{cases}$$

Мінімізація проводиться за змінним $(w_1, \dots, w_m, c, a_1, \dots, a_N)$. Додаткові обмеження можна накласти через обмеження відповідних ваг $w_i \geq w_i^0$.

Розроблення системи. У цій роботі використана модель тонкого клієнта, в межах якої вся логіка та управління даними зосереджені на сервері, а клієнтська програма забезпечує тільки функції рівня представлення.

Вимоги до системи. Для виконання обчислень обраними методами розроблена система, яка відповідає таким вимогам:

- 1) заснована на клієнт-серверній архітектурі для доступу через відкрите API;
- 2) повністю віртуальна, щоб підходити під визначення приватної cloud-системи;
- 3) забезпечує можливість виконувати планування та розподілення задачі пошуку коефіцієнтів скорингових карт;
- 4) обмеження: планувальник повинен постійно отримувати інформацію від вузлів про їх стан.

Для створення приватної хмари обрано open source рішення для віртуалізації Openstack 2011.3 [11]. Це рішення дозволяє замінити метод планування для прикладної задачі обчислення коефіцієнтів скорингових карт. Загальна структура системи представлена на рис. 2 і передбачає використання клієнт-серверної архітектури.

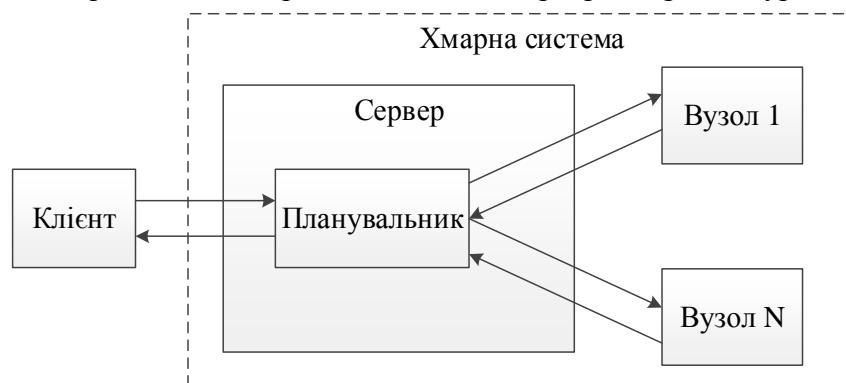


Рис. 2. Структура приватної хмарної системи

Розподілення завдань по ресурсах у неоднорідних розподілених системах належить до класу NP-повних задач. Як метод планування обрано метод покрокового планування [12]. Метод підходить для неоднорідних систем, в яких банки можуть виконувати власні обчислювання. Обладнання можна використати як компоненти неоднорідної системи зі збереженням однорідності обчислювального середовища. Для моделювання використовуватимемо віртуальні вузли на основі ОС Ubuntu 12.04 [13]. Через обмеження моделювання для проведення тестів використовуватимемо 4 віртуальні вузли.

Віртуалізація виконана за допомогою пакета з відкритим кодом Openstack, загальна структура системи віртуалізації представлена на рис. 3.

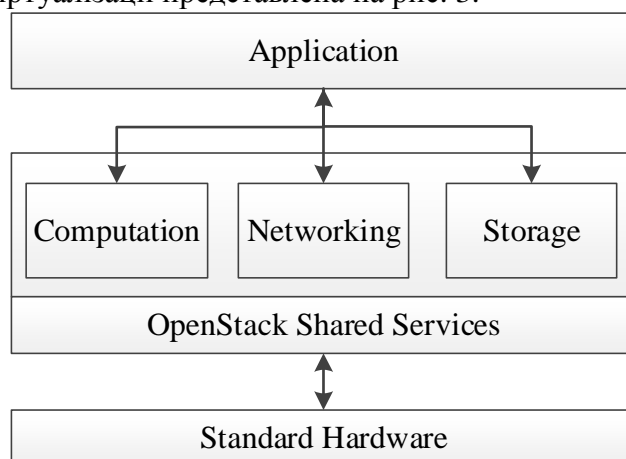


Рис. 3. Структура системи віртуалізації OpenStack

Планувальник потактово виконує такі дії:

1. Перевірка потоку вхідних заявок і формування черг.
2. Оцінювання вимогливості задач до ресурсів.
3. Планування розкладу задач від найбільш вимогливих до найменш вимогливих.
4. Перевірка журналу планування та завантаження задач на ресурси.

Розроблювана система призначена для управління сервером та обчислювальними вузлами: створення API для зовнішнього клієнта, планування задач, управлінням передачею даних, запуску прикладних задач на обчислювальних вузлах та організацію зберігання даних у розподіленій системі.

Для вирішення прикладної математичної задачі побудови скорингової карти використані готові реалізації:

- реалізацію регресійного аналізу у бібліотеці `org.apache.commons.math.stat`;
- реалізацію методу лінійного програмування в бібліотеці `lp_solve 5.5.2.0`.

Бібліотека `Apache commons-math` включає базові методи математичної статистики, частотні розподіли, двоваріантну регресію, функцію хі-квадрат та статистичні тести ANOVA [14].

Бібліотека `lp_solve 5.5.2.0` спеціалізована на методах лінійного програмування, має відкриті коди та API, що підходить для розроблення системи побудови скорингових карт [15].

Експериментальні дослідження. У більшості банків використовуються різні види скорингу: `Application scoring` для оцінювання кредитної заявки, скоринг поведінки, або `Behavioral scoring`, для аналізу дій позичальника та його платоспроможності, `Collection scoring` для роботи з простроченою заборгованістю і `Fraud scoring`, націлений на оцінку ймовірності шахрайських дій з боку позичальника. Крім перерахованих, є такі технології скорингу, які дозволяють аналізувати ефективність спрямованих споживачеві пропозицій (скоринг відгуку, `Response scoring`) і оцінювати ймовірність, що споживач піде до іншого банку (скоринг втрат, `Attrition scoring`).

Для тестування системи використовується вибірка з 15 анкетних питань, кожна з відповідей оцінюється коефіцієнтом, нормалізованим до інтервалу $[-1;1]$. Еталон визначає «правильні» відповіді в анкеті. Керуючись еталоном, згенеровано експериментальну вибірку анкет, які є вхідними даними методів регресійного аналізу і лінійного програмування для отримання скорингових карт.

Оскільки для обчислення рейтингу анкети використовують суму ненормованих коефіцієнтів, то можна побачити, що по кожному з методів похибка симетрична. Метод лінійного програмування має в середньому меншу похибку, ніж регресійний аналіз. Числові дані для порівняння та відносні похибки кожного з методів також наведено в табл.

Таблиця

Дані точності обчислень регресійного аналізу та лінійного програмування

№ коеф.	Еталон	Лінійний метод	Регресійний метод	δ лін. методу	δ регр. методу
1	-0,2	-0,13043	-0,39722	0,347863	0,986077
2	0,4	0,373731	0,075457	0,065673	0,811357
3	0,32	0,235657	0,481272	0,263572	0,503975
4	0,12	0,132875	0,133429	0,107289	0,111906
5	0,3	0,411536	0,300159	0,371788	0,00053
6	0,24	0,221411	0,34919	0,077452	0,454957
7	0,72	0,530287	0,739118	0,263491	0,026552
8	0,8	0,850655	0,174108	0,063318	0,782366
9	0,76	0,479773	0,338125	0,36872	0,555099
10	0,12	0,101027	0,042181	0,158107	0,648493
11	-0,1	-0,08138	-0,11746	0,186249	0,17456
12	0,26	0,238436	0,36683	0,082937	0,410885
13	-0,32	-0,35167	-0,32975	0,09898	0,030473
14	-0,1	-0,10532	-0,01784	0,053152	0,821575
15	-0,12	-0,09338	-0,23706	0,221814	0,975516
16	0,36	0,336559	0,217071	0,065115	0,397025
17	0,3	0,379333	0,549964	0,264442	0,833214
18	-0,06	-0,06213	-0,11403	0,035542	0,900518
19	0,2	0,159438	0,337855	0,202808	0,689276
20	0,14	0,110718	0,039018	0,209157	0,721297
21	-0,12	-0,08504	-0,1227	0,291359	0,022497
22	0,12	0,11183	0,21521	0,06808	0,793418
23	-0,06	-0,07149	-0,01844	0,191557	0,692656
24	-0,08	-0,08918	-0,1161	0,114733	0,451248
25	-0,14	-0,12441	-0,03305	0,111367	0,763939
26	-0,4	-0,50381	-0,6357	0,259525	0,58926
27	-0,3	-0,34352	-0,41649	0,145063	0,388304
28	-0,52	-0,42444	-0,50805	0,183768	0,022977
29	-0,54	-0,62387	-0,44926	0,155318	0,168035
30	-0,62	-0,80196	-0,75622	0,293482	0,21971
Середнє				0,177391	0,498256

На рис. 4. проілюстрована похибка результатів обчислень регресійного та лінійного методу порівняно з еталоном.

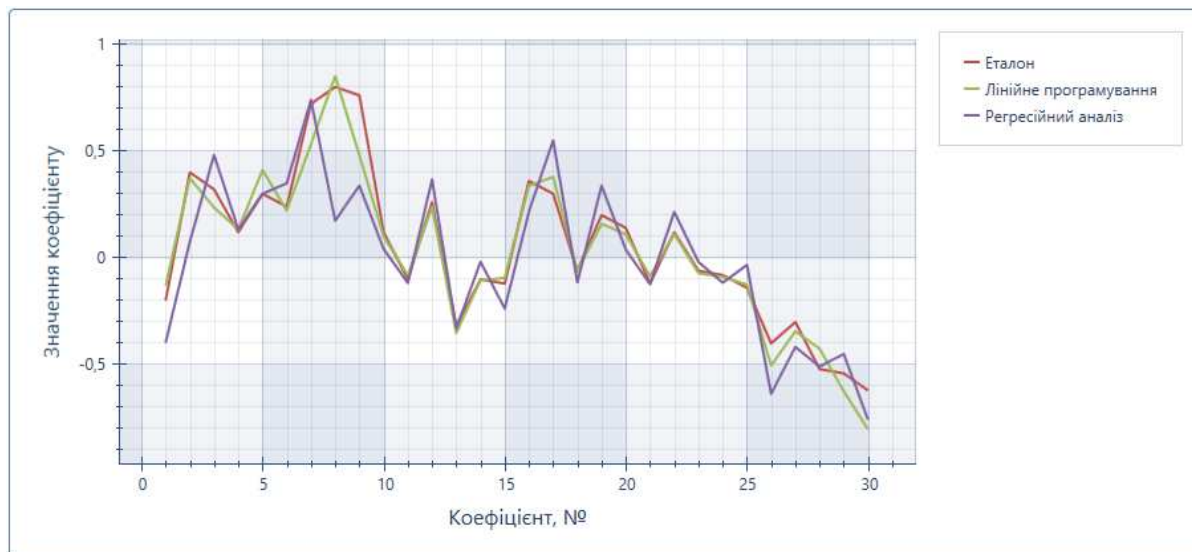


Рис. 4. Порівняння результатів оброблення даних різними методами з еталоном

На рис. 5. наведено порівняння часу, затраченого на виконання обчислень відповідно методами лінійного програмування та регресійного аналізу.

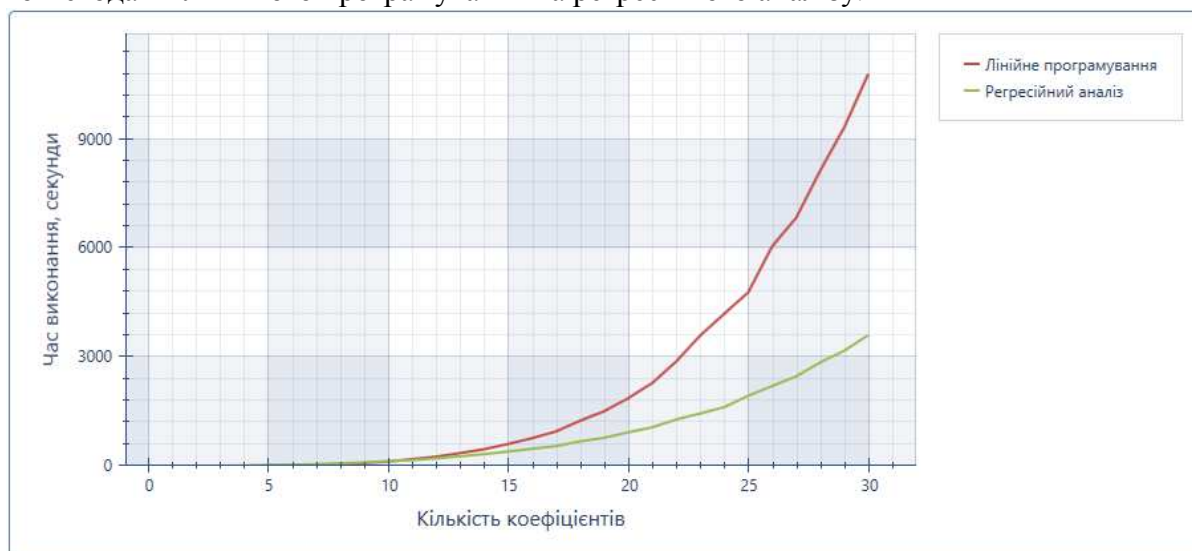


Рис. 5. Порівняння часу виконання методів у хмарній системі (витрати часу на обчислення коефіцієнтів скорингової карти)

Як видно з рис. 5, метод лінійного програмування повільніший, незважаючи на те, що він має кращі можливості для розпаралелювання у хмарі. При цьому необхідно зазначити, що при збільшенні кількості задіяних вузлів метод регресійного аналізу не дасть суттєвого виграшу у продуктивності в порівнянні з методом лінійного програмування.

Висновки. Про ефективність застосування методів у cloud-системах говорить той факт, що збільшення вузлів призводить до зменшення часу на побудову скорингової карти. Метод лінійного програмування має хорошу точність і добре розпаралелюється, але при цьому вимагає набагато більше обчислень. У результаті для великої кількості коефіцієнтів він значно програє регресійному аналізу у продуктивності. Регресійний аналіз потенційно погано розпаралелюється, але має доволі малу обчислювальну складність. З іншого боку, для побудови скорингової карти за допомогою регресійного аналізу не обов'язково мати приватну хмарну систему. При цьому для отримання результатів з тією ж точністю необхідно буде більше часу.

У майбутніх дослідження механізм обчислення коефіцієнтів скорингових карт може бути використаний для розроблення моделі довіри та гарантій безпеки користувачів розподілених комп'ютерних систем.

Список використаних джерел

1. Шакирова А. Р. Скоринг как метод оценки риска неплатежеспособности заемщика / А. Р. Шакирова, Э. В. Саяхова // Экономика и социум. – 2012. – № 2.
2. Уланов С. В. Оценка качества и сравнение скоринговых карт / С. В. Уланов // Экономические науки. – 2009. – № 9 (58). – С. 330-335.
3. Модели кредитного и поведенческого скоринга [Электронный ресурс]. – Режим доступа : <http://masters.donntu.edu.ua/2006/kita/shepeleva/library/metod%20scoring.pdf>.
4. Olivier Brian, Thomas Brunschwiler, Heinz Dill and others. Cloud computing. White paper [Электронный ресурс]. – Режим доступа : https://www.satw.ethz.ch/organisation/tpf/tpf_ict/box_feeder/2012-11-06_2_SATW_White_Paper_Cloud_Computing_EN.pdf.
5. Siani Pearson, Yun Shen, Miranda Mowbray. A Privacy Manager for Cloud Computing // Cloud Computing. Lecture Notes in Computer Science Volume 5931, 2009. – P. 90-106.
6. Wayne Jansen, Timothy Grance. Guidelines on Security and Privacy in Public Cloud Computing // National Institute of Standards and Technology. – U.S. : Department of Commerce. Special Publication 800-144.
7. Bhaskar Prasad Rimal, Eunmi Choi. A taxonomy and Survey of Cloud Computing Systems // School of Business IT, Kookmin University. 2009 Fifth International Joint Conference on INC, IMS and IDC.
8. Гараган С. А. Метод эмпирической скоринговой функции и его использование в кредитном процессе [Электронный ресурс] / Гараган Сергей Александрович. – Режим доступа : http://crosys.org/empirical_scoring_function.html.
9. Бородич С. А. Эконометрика : учебное пособие [Электронный ресурс] / С. А. Бородич. – Минск : Новое знание, 2001. – 408 с. – Режим доступа : http://www.economy.bsu.by/library/Бородич_Эконометрика/Бородич_Эконометрика.pdf.
10. Логистическая регрессия и ROC-анализ - математический аппарат [Электронный ресурс]. – Режим доступа: <http://www.basegroup.ru/library/analysis/regression/logistic>.
11. Gomez-Folgar, F., Garcia-Loureiro, A., Pena, T.F., Valin, R. Performance of the CloudStack KVM Pod Primary Storage under NFS Version 3// IEEE 10th International Symposium on Parallel and Distributed Processing with Applications (ISPA), 2012.
12. Система пространственного распределения заданий в распределенных вычислительных системах / А. В. Симоненко, С. В. Пих, Н. В. Слуцкий, В. В. Воробйов // Вісник НТУУ «КПІ» Інформатика, управління та обчислювальна техніка. – 2013. – № 56. – С. 94-97.
13. Stanik, A., Hovestadt, M. ; Odej Kao. Hardware as a Service (HaaS): The completion of the cloud stack// Computing Technology and Information Management (ICCM), 2012 8th International Conference on (Volume:2).
14. Apache Commons. Commons Math. [Электронный ресурс]. – Режим доступа : <http://commons.apache.org/proper/commons-math/userguide/overview.html>.
15. Introduction to Ip_solve 5.5.2.0 [Электронный ресурс]. – Режим доступа : <http://lpsolve.sourceforge.net/5.5>.