

ОБРОБКА ДАНИХ ТА ГЕНЕРУВАННЯ РЕКОМЕНДАЦІЙ В МАРКЕТИНГОВОМУ ТЕСТУВАННІ

Танасієнко Д.О¹., Андрющенко В.О².

¹Український Державний університет науки і технологій, PhD студент, України

²Український Державний університет науки і технологій, к.т.н., доцент, України

Анотація: У цій науковій статті досліджується структурна модель обчислювальних процесів, які є ключовими для ефективної обробки даних та формування точних рекомендацій у контексті маркетингового тестування. Увага приділяється застосуванню цих процесів в умовах нестабільного хмарного середовища, що додає значної складності до управління та оптимізації. Запропонована модель описує систему як сукупність взаємопов'язаних задач, що виконуються на регулярній періодичній основі, забезпечуючи безперервний цикл аналізу та прогнозування. Центральним аспектом дослідження є розробка інноваційних методів управління цими обчислювальними процесами. Зокрема, стаття зосереджується на критично важливій проблемі мінімізації запізнень щодо встановлених директивних термінів виконання задач. Крім того, значна увага приділяється оптимізації використання доступних обчислювальних ресурсів, що дозволяє підвищити ефективність системи та знизити витрати. У статті також детально аналізуються важливі характеристики задач, такі як їх ідемпотентність, класифікація за пріоритетом виконання та за обсягом споживаних ними обчислювальних ресурсів. Нарешті, стаття пропонує моделювання функціонування системи в умовах нестабільності хмарного середовища.

Ключові слова: граф задач, ідемпотентність, хмарні обчислювальні процеси, моделювання, програмне забезпечення, інформаційні технології.

У сучасних умовах маркетингове тестування часто здійснюється в хмарних обчислювальних середовищах. Це створює виклики для ефективного управління обчислювальними процесами, які повинні виконуватися вчасно та з оптимальним використанням ресурсів. Задачі можуть мати директивні терміни, вимагати різних обчислювальних ресурсів та завершуватися невдачами через зовнішні фактори[1]. Таким чином, постає проблема розробки структурної моделі, яка дозволить ефективно управляти цими процесами, мінімізуючи запізнення та оптимізуючи використання ресурсів.

Метою даного дослідження є розробка структурної моделі обчислювальних процесів для маркетингового тестування в нестабільному хмарному середовищі. Це включає в себе створення методів управління, які дозволять мінімізувати запізнення щодо директивних термінів, прогнозувати необхідну кількість обчислювальних ресурсів та забезпечити ідемпотентність задач.

Є система, яка складеться з множини різнотипних задач з наступними властивостями:

- на множені задач існує відношення часткового порядку;
- кожна задача виконується на періодичній основі (кожен день, певні дні тижня);
- кожен запуск задачі потребує певних обчислювальних ресурсів RAM, CPU;
- є окремий SQL кластер, де виконуються запити над великими даними, всі запити в цьому кластері використовують абстрактний ресурс – SQL Slots;
- кожна задача може завершитися невдачею та її потрібно перезапустити;
- деякі задачі мають директивні терміни.

Позначимо цю множину задач як Ψ . Є глобальна задача керування виконанням множиною задач (порядок виконання, виділення ресурсів) та керування ресурсами RAM, CPU та SQL. В данному дослідженні будемо розглядати тільки структуру множини задач та самі задачі.

У випадку курування задачами маємо три задачі:

1. задача розкладів з директивними термінами. Мета: мінімізувати запізнення щодо директивних термінів;
2. задача прогнозування тривалості кожної задачі. Мета: мінімізувати неточність у прогнозах тривалості;
3. задача прогнозування затрат ресурсів для кожної задачі. Мета: Мінімізувати неточність у прогнозах затрат ресурсів. Опис моделі задач. Будемо називати задачі з множини Ψ як EngineTasks. Тоді маємо такі визначення.

EngineTask – ідемпотентна одиниця роботи у системі Ψ . Тобто завдання, яке при повторному виконанні дає той самий результат, навіть якщо попередній запуск був перерваний на середині виконання

Далі в тексті будемо позначати EngineTask як ET для стислості. ET – програмний код на мові Python, оформлений у вигляді модулів [2] і який запускається через CLI (Command Line Interface) [3].

Класифікація за пріоритетом виконання. За пріоритетом виконання завдання можна розділити на два підмножини:

- primary – основні завдання, час виконання яких безпосередньо пов'язаний з директивними термінами. Тобто або ET має директивний термін завершення, або одне з наступних завдань має директивний термін.

- service – завдання, які не мають директивних термінів.

Класифікація за типом ресурсів, які використовує ET. Є 3 типи обчислювальних ресурсів, які використовує ET:

- CPU (англ. central processing unit) – обчислювальні ядра процесора. Приймає речові значення з кроком 0.5. 1 означає одне логічне ядро центрального процесора. Приймає значення від 0,5 до 220.

- RAM (англ. random-access memory) – оперативна пам'ять, яка доступна процесору. Приймає речові значення з кроком 0.25. 1 означає один гігабайт пам'яті. Набуває значення від 1 до 800.

- SQL Slots – абстрактна одиниця обчислення, яка використовується при запитах у Google BigQuery [4]. Приймає речові значення з кроком 1. Набуває значення від 1.

ET за ресурсами ділиться на дві групи:

- які використовують CPU+RAM. Будемо позначати їх як $(CR)^{ET}$, де CR – аббревіатура перших букв ресурсів, які вони використовують;

- які використовують CPU+RAM+SQL Slots. Будемо позначати їх як $(CRS)^{ET}$, де CRS – аббревіатура перших букв ресурсів, які вони використовують.

Граф задач. Представимо множину завдань як ациклічний граф. Тоді набір завдань, які пов'язані між собою, можна записати як

$$DAG_i = \{V_i, E_i\} \quad (1)$$

де V_i – це одна з задач ET , а E_i – це спрямований зв'язок між двома задачами.

У такому разі всю систему Ψ в момент часу t можна записати як:

$$\Psi_t = \{DAG_i, i = 1 \dots n_t\} \quad (2)$$

де n_t – кількість ациклічних графів в момент часу.

Приклади завдань із прикладами послідовності виконання. Для наочності наведемо приклад реального графу задач на рис.1, рис. 2.

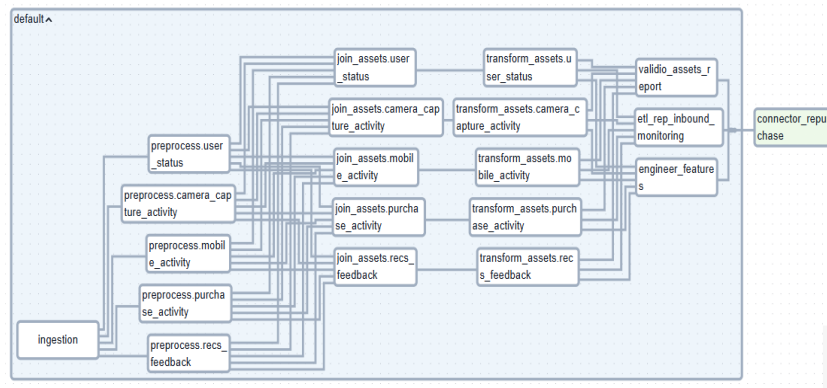


Рисунок 1 – Перша частина графу задач

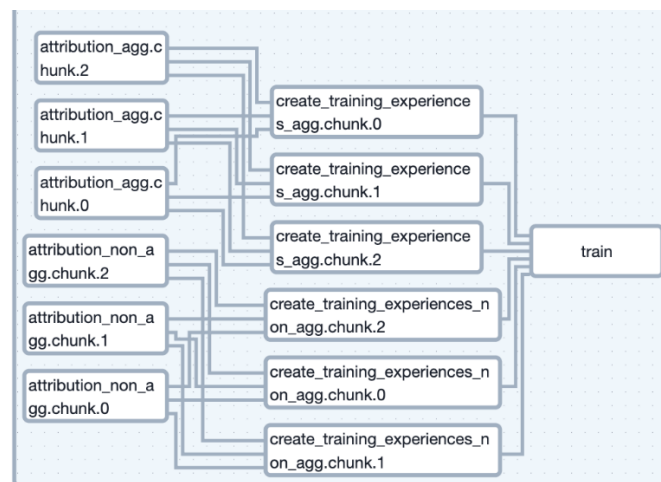


Рисунок 2 – Друга частина графу задач

Важливо відзначити, деякі завдання виконуються паралельно, наприклад Train і Predict. Це пов'язано з тим, що в межах одного дня їх результати не залежать один від одного. Predict використовує модель, що згенерувала через Train за минулий день, а не за поточний.

Висновки. Запропонована в статті структурна модель обчислювальних процесів для маркетингового тестування в нестабільному хмарному середовищі демонструє потенціал для мінімізації затримок виконання задач та оптимізації обчислювальних ресурсів. Важливу увагу приділено забезпеченню ідемпотентності задач, що підвищує стійкість до збоїв. Класифікація задач за пріоритетом і ресурсом дає змогу ефективніше керувати їхнім виконанням відповідно до директивних термінів та доступності ресурсів. Розроблені методи прогнозування термінів виконання та витрат ресурсів дозволяють зменшити неточності. Перспективним напрямком є подальший розвиток цієї

моделі, зокрема інтеграція автоматизованих систем управління обчислювальними процесами в широкій сфері додатків, включаючи інші галузі, що потребують аналогічного підходу до управління ресурсами в хмарних середовищах.

ЛІТЕРАТУРА

1. Танасієнко Д.О., Андрющенко В.О. Обчислювальні процеси у хмарному середовищі. Сучасні інформаційні та комунікаційні технології на транспорті, в промисловості і освіті – 2024. С. 132.
2. Parnas D. L. On the criteria to be used in decomposing systems into modules. Communications of the ACM. – 1972. № 15(12). P. 1053-1058.
3. Richard Blum, Christine Bresnahan. Linux Command Line and Shell Scripting BIBLE. John Wiley & Sons, Inc., Indianapolis, Indiana – 2015. P. 818
4. Melnik S. et al. Dremel: interactive analysis of web-scale datasets. Proceedings of the VLDB Endowment. – 2010. №. 3(12). P. 330-339.

PROCESSING DATA AND GENERATING RECOMMENDATIONS IN MARKETING TESTING

Tanasienko Dmytro, Andriushchenko Vadym

Abstract: *This scientific article explores a structural model of computational processes that are crucial for the effective processing of data and the generation of accurate recommendations in the context of marketing testing. Attention is paid to the application of these processes in unstable cloud environments, which adds significant complexity to management and optimization. The proposed model describes a system as a set of interconnected tasks that are performed on a regular periodic basis, ensuring a continuous cycle of analysis and forecasting. A central aspect of the research is the development of innovative methods for managing these computational processes. In particular, the article focuses on the critically important problem of minimizing delays concerning established deadlines for task execution. Furthermore, significant attention is given to optimizing the use of available computational resources, which allows for increasing system efficiency and reducing costs. The article also analyzes in detail important characteristics of tasks, such as their idempotency, classification by execution priority, and by the volume of computational resources consumed. Finally, the article proposes a simulation of the system's functioning under conditions of cloud environment instability*

Keywords: *tasks graph, idempotence, cloud computing, modeling, software, information technology.*

REFERENCE

1. Tanasienko D.O., Andriushchenko V.O. Obchislyuvalni protsesi u khmarnomu seredovishchi. Cuchasni informatsiini ta komunikatsiini tekhnologii na transporti, v promislovosti i osviti – 2024. S. 132. [in Ukrainian]
2. Parnas D. L. On the criteria to be used in decomposing systems into modules. Communications of the ACM. – 1972. № 15(12). P. 1053-1058.
3. Richard Blum, Christine Bresnahan. Linux Command Line and Shell Scripting BIBLE. John Wiley & Sons, Inc., Indianapolis, Indiana – 2015. P. 818
4. Melnik S. et al. Dremel: interactive analysis of web-scale datasets. Proceedings of the VLDB Endowment. – 2010. №. 3(12). P. 330-339.