

ІМПЛЕМЕНТАЦІЯ ТА РЕЙТИНГ РІШЕНЬ WEB-АНАЛІТИКИ НА ОСНОВІ ХМАРНИХ СЕРВІСІВ

Знахур Л. В., Знахур С. В.

Харківський національний економічний університет ім. С. Кузнеця

Анотація. Вартість володіння інфраструктури для WEB-аналітики e-commerce сайтів є важливим аспектом ефективності інтернет-бізнесу. У роботі проведено аналіз сучасних рішень, архітектур щодо розв'язання наскрізного завдання створення та імплементації WEB-аналітики e-commerce сайтів. Проблемою, яку вирішує дослідження, є визначення оптимальної архітектури хмарних сервісів для реалізації WEB-аналітики e-commerce сайтів на основі використання інтегрального показника, який містить адитивну функцію показників функціональності рішення та вартості володіння обраних хмарних сервісів для побудови рішення.

Ключові слова: e-commerce, ML, WEB-аналітика, інтегральна оцінка, імплементація.

Вступ

Криза 2020–2021 рр. показала, що функціональність та вартість володіння інфраструктури для WEB-аналітики e-commerce сайтів є дуже важливим аспектом ефективності інтернет-бізнесу. Ціна володіння інфраструктури провідних провайдерів хмарних сервісів має свої особливості та залежить від обраної архітектури, кількості та специфіки сервісів, часу їхнього використання. Також необхідно враховувати можливість масштабування та інтеграції рішення до власних сервісів або даних за рахунок використання нативних конекторів, API та різних додаткових служб, що дозволяють підключати інші зовнішні сервіси, необхідні для побудови наскрізної WEB-аналітики e-commerce сайтів.

Аналіз публікацій

Дослідженню питань побудови наскрізних рішень щодо аналізу трафіку сайтів e-commerce на основі інструментів вебаналітики присвячено праці таких учених, як Х. Карджалото, П. Сінгх, К. Родден, Р. Томас, Х. Хатчінсон, Б. Янсен та ін. Проте є необхідність практичного дослідження сучасних рішень та розроблення методичних підходів щодо комплексного обґрунтованого вибору архітектур та інтеграції рішень WEB-аналітики.

Згідно з дослідженнями [3, 4], сучасний інтернет-бізнес потребує вирішення таких завдань:

1. Створення та актуалізація системи наскрізної аналітики (dwh, etl, дашборди).
2. Створення оперативного маркетингового аналізу діяльності бізнесу та товарної аналітики на основі даних із різних джерел.

3. Автоматизованого впровадження рекомендаційної системи на основі ML у сфері аналітики та сегментації аудиторії.

Інтерес до ML зростає щороку. Лідери e-commerce залучені до R&D та застосування машинного навчання та поглибленого аналізу даних для завдань бізнесу та WEB-аналітики. Упроваджуються результати успішних проектів на рівні промислової експлуатації. Виникає потреба й у методології вибудовування бізнес-процесів навколо машинного навчання та в інструментах автоматизації цих процесів для повного циклу WEB-аналітики. Цей напрям отримав назву ModelOps WEB-аналітики. На ринку вже є понад 70 рішень та інструментів, так чи інакше пов'язаних із завданнями ModelOps, серед яких безкоштовні, відкриті рішення (наприклад MLFlow, KubeFlow) та рішення від лідерів у галузі Data Science, таких як AWS, GCP, Azure та IBM. Архітектурні рішення мають аналогічні підходи щодо методології управління життєвим циклом моделей машинного навчання, але імплементації моделей залежать від обраної концепції загальної архітектури. Досить популярним рішенням є мікросервісна архітектура або використання API для моделей.

Проблемою, яку вирішує дослідження, є визначення оптимальної архітектури хмарних сервісів для реалізації WEB-аналітики e-commerce сайтів на основі використання інтегрального показника, що містить адитивну функцію показників функціональності рішення та вартості володіння обраних хмарних сервісів для побудови рішення. Для вирішення проблеми вибору та побудови оптимальної архітектури було визначено такі завдання:

1. Дослідити необхідні інструменти, стандарти та показники WEB-аналітики (у роботі розглянуто тільки Google Analytics).

2. Проаналізувати сучасні архітектури, технології та сервіси щодо впровадження сайтів та інтеграції Google Analytics, рішень ML.

3. Реалізувати різні хмарні архітектури та підключити різні аналітичні індикатори Google Analytics.

4. Побудувати рейтинг архітектурних рішень на основі інтегрального показника.

Використання Google Analytics 4 (GA4) суттєво змінило сучасний облік WEB-аналітики. Однією з основних переваг GA4 є машинне навчання та NLP-функції, які застосовуються для того, щоб:

передбачати можливість конверсії та створювати на основі цих прогнозів аудиторії для Google Ads;

попереджати про важливі тенденції в даних. Наприклад, про товари, на які зростає попит через те, що змінюються потреби користувачів;

знаходити аномалії;

передбачати можливість відтоку клієнтів.

Детальний аналіз наявних рішень показав, що проблема інтеграції Google Analytics до хмарних рішень провайдерів має декілька рівнів. Перше: 2020 р. було змінено стандарт та концепцію Google Analytics, тобто зроблено перехід на стандарт GA4. Це суттєво зменшило кількість сервісів конекторів до даних GA4, оскільки більшість комерційних рішень були реалізовані для Google Analytics (universal). Друге: рішення на основі сервісів AWS або Azure потребують додаткові налаштування, дозволи та ролі для підключення до GA. Також рішення AWS та Azure потребують додаткові процедури ETL (вони можуть бути реалізовані на основі лямбда-функцій) для процесингу даних GA. Ці особливості використання даних Google Analytics суттєво збільшують трудомісткість та вартість володіння архітектур, які не базуються винятково на сервісах GCP [6].

Мета та постановка завдання

Метою роботи є аналіз сучасних рішень, архітектур щодо вирішення наскрізного завдання створення та імплементації WEB-аналітики e-commerce сайтів.

Завдання роботи – визначити необхідні інструменти, сервіси та показники для WEB-аналітики; проаналізувати сучасні архітектури, технології та сервіси щодо впро-

вадження сайтів та інтеграції Google Analytics, упровадження рішень ML на основі даних Google Analytics; здійснити реалізацію різних архітектурних рішень щодо оперативного аналізу даних Google Analytics, формування аналітичної звітності, побудови рейтингу архітектурних рішень.

Об'єктом дослідження є впровадження інформаційних технологій, сервісів для побудови рішень щодо реалізації ML-моделей на основі даних WEB-аналітики e-commerce сайтів.

Предметом дослідження є побудова та обґрунтування оптимальної архітектури WEB-аналітики e-commerce сайтів для оперативного та ML-аналізу даних Google Analytics.

Наукова новизна роботи полягає в розробленні методики щодо рейтингу рішень відомих провайдерів, які будуть зіставлені згідно з параметрами функціональності та вартості володіння.

Виклад основного матеріалу

Для виконання завдань роботи щодо побудови рейтингу рішень WEB-аналітики e-commerce сайтів було здійснено дослідження особливостей побудови та імплементації WEB-аналітики на основі хмарних сервісів AWS, GCP, Azure. Загальна концепція вибору оптимальної архітектури рішення передбачає розрахунок компонент оцінки функціональності WEB-аналітики та компонент TCO (total cost of ownership) – вартості володіння рішень. На рис. 1 наведено контекстну діаграму «Проведення експерименту щодо імплементації та рейтингу рішень WEB-аналітики для e-commerce сайтів», на рис. 2 наведено її декомпозицію.

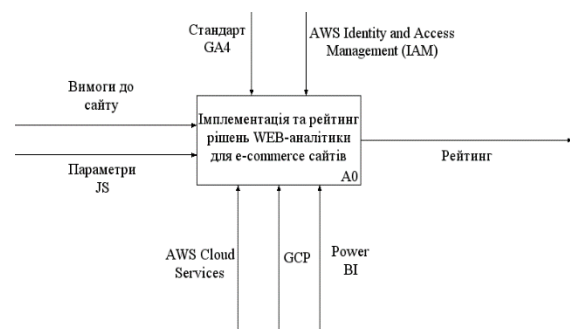


Рис. 1. Контекстна діаграма «Проведення експерименту щодо імплементації та рейтингу рішень WEB-аналітики для e-commerce сайтів»

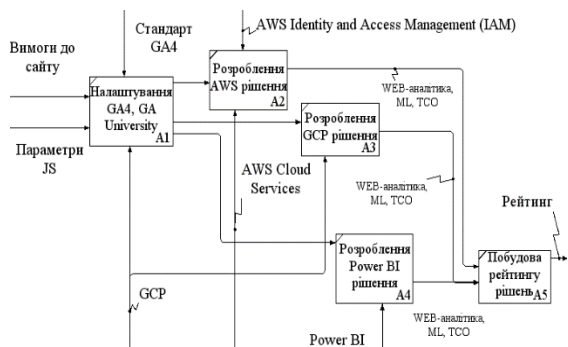


Рис. 2. Декомпозиція процесу «Проведення експерименту щодо імплементації та рейтингу рішень WEB-аналітики для e-commerce сайтів»

Запропонований у роботі підхід щодо побудови рейтингу рішень WEB-аналітики для e-commerce сайтів містить такі етапи:

1. Розроблення (або отримання дозволу щодо адміністрування) e-commerce сайту.

2. Створення GA-акаунтів для e-commerce сайту. Пропонується використати стандарт GA (universal) та GA4. Це дозволить отримати різні дані аналітики та використати різні конектори в архітектурі та сервісах.

3. Реалізація архітектури AWS для аналізу даних GA, GA4. Варіант реалізації необхідно здійснити в межах одного часового інтервалу трафіку на e-commerce сайт. Розраховується час виконання запитів та побудови дашбордів. Для аналізу часу отримання результатів моделей ML пропонується реалізувати класичну модель ML (Tree Decision) на основі даних GA.

4. Розрахунок вартості володіння архітектурою AWS на базі AWS Pricing Calculator (<https://calculator.aws/#/estimate>).

5. Реалізація обраної архітектури GCP для аналізу даних GA, GA4. Варіант реалізації необхідно здійснити в межах одного часового інтервалу трафіку на e-commerce сайт. Розраховується час виконання запитів та побудови дашбордів. Для аналізу часу отримання результатів моделей ML пропонується реалізувати класичну модель ML (Tree Decision) на основі даних GA.

6. Розрахунок вартості володіння архітектурою GCP на базі <https://cloud.google.com/products/calculator>.

7. Побудова моделі для дашборду GA на основі використання конектора GA (universal) PowerBI. Конектор GA4 не реалізовано в PowerBI. Для аналізу часу отримання результатів моделей ML пропонується ре-

алізувати класичну модель ML (Tree Decision) на основі даних GA.

8. Формування показників за методикою побудови інтегрального показника.

9. Розрахунок рейтингу для кожної архітектури рішення на основі значень інтегрального показника.

Для побудови інтегрального показника архітектурних рішень було запропоновано використати такі показники функціональності WEB-аналітики, складності архітектури та компонент TCO (табл. 2). Вибір показників є результатом анкетування WEB-аналітиків, DevOps та MLOps інженерів.

Таблиця 2 – Показники функціональності WEB-аналітики, складності архітектури та компонент TCO

Показник	Опис
Кількість сервісів, використаних в архітектурі	Складність архітектури визначається кількістю хмарних сервісів, що необхідно використати для побудови масштабованого рішення. Збільшення кількості сервісів збільшує складність архітектури за рахунок інтеграції, налаштування IAM (ролі, повноваження). Еталонне значення – максимальне
Час оновлення даних	Технологічні особливості оновлення даних GA. Еталонне значення – мінімальне
Вартість рішення	TCO – розраховується на основі відповідних калькуляторів хмарних провайдерів. Еталонне значення – мінімальне
Кількість показників аналітики	Максимальна кількість показників, які підтримує обрана технологія інтеграції. Еталонне значення – максимальне
Джерело даних	Старий GA (0) або новий стандарт GA4 (1). Еталонне значення – максимальне
Інтеграція (або наявність) ML	Нативна підтримка або інтеграція ML (так чи ні) (0 або 1). Еталонне значення – максимальне
Масштабування рішення	Можливість масштабування (так чи ні) (0 або 1). Еталонне значення – максимальне
Імплементація аналітики у WEB	Можливість імплементації у Web на основі iframe (так чи ні) (0 або 1). Еталонне значення – максимальне

Розрахунок інтегрального показника базується на формулі:

$$I_j = \sum_{i=1}^m P_{ij}^n, \quad (1)$$

де P_{ij}^n – i -ий нормований показник моделі для j -архітектури; m – кількість архітектур.

$$P_{ij}^n = P_{ij} \div P_{ij\max}, \quad (2)$$

де P_{ij} – i -ий показник моделі для j -архітектури; $P_{ij\max}$ – максимальне значення i -го показника j -ої архітектури.

Максимальне значення показника розраховується відповідно до напрямку дії показника. Так, еталонний показник може бути максимальним (наприклад, кількість показників WEB-аналітики в рішенні), або мінімальним (наприклад, час розрахунку або трансферу даних). Якщо використовується як еталон мінімальне значення, тоді P_{ij} розраховується, як $1 \div P_{ij}$.

Для оцінювання та рейтингу рішень використано підхід розрахунку інтегральних показників із нормуванням. Результати розрахунку показників для отримання інтегрального показника наведено в табл. 1.

Таблиця 1 – Результати розрахунку показників архітектур

Показники	Оцінка / AWS	Оцінка / GCP	Оцінка / Azure
Сервіси, використані в архітектурі	0.16/S3, AppFlow, Amazon Athena, Amazon QuickSight	0.3/Data Studio, BigQuery, AppEngine	1/Microsoft Power BI
Час оновлення даних (до 1 Гб)	0,2/10 хв	1/2 хв	0,01/24 год
Вартість рішення	0,37	0,38	1
Кількість показників	1/50 спеціальних параметрів і 50 спеціальних показників	1/50 спеціальних параметрів і 50 спеціальних показників	0,4/20 показників
Джерело даних	0/GA4	0/GA4	0/GA (universal)
Інтеграція ML	0/QuickSight, SageMaker	0/python (function), BigQuery	0/Python Script, API
Масштабування рішення	1/існує	1/існує	0/не існує
Імплементация	1/iframe	1/iframe	1/iframe
Інтегральна оцінка	3,73	4,68	3,41

Для розрахунку показників реалізовано скрипт оцінки часу виконання запитів та розрахунок вартості володіння архітектурами.

особливості імплементации рішення або ML-моделей можуть змінити рейтинг.

Висновки

Формальний підхід щодо оцінки архітектурних рішень показав, що фаворитом є архітектурне рішення на основі GCP із максимальним результатом 4,68 балів. Архітектурне рішення AWS має 3,73 бали, а рішення Azure – 3,41. Але варто врахувати, що в аналізі не було використано ваг для параметрів оцінок. Так, наприклад, час оновлення даних зазвичай має максимальну вагу та може суттєво впливати на результат рейтингу. Необхідно також зазначити, що вибір сервісів для архітектур GCP та AWS містить тільки базові сервіси з мінімальними характеристиками щодо інстанс, зберігання даних, безпеки та масштабування. Збільшення вимог щодо продуктивності промислових архітектур та

Література

1. AWS: быстрая разработка, сборка и развертывание приложений на AWS. URL: <https://aws.amazon.com/ru/codestar/>
2. BigQuery. URL: <https://cloud.google.com/bigquery>.
3. Burby, J., & Brown, A. Web Analytics Definitions – Version 4.0. URL: <http://www.digitalanalyticsassociation.org/>
4. Data Age 2025: The digitization of the world – From edge to core. White paper. Farmingham, MA. URL: <https://www.seagate.com/files/www-content/our-story/trends/files/idcseagate-dataage-whitepaper.pdf>.
5. Google Analytics (n.d.). Features. URL: <http://www.google.com/analytics/features/>
6. GCP. URL: <https://cloud.google.com/solutions/retail>.

References

1. AWS: fast development, assembly and deployment of applications on AWS. URL: <https://aws.amazon.com/ru/codestar/>
2. BigQuery. URL: <https://cloud.google.com/bigquery>
3. Burby, J., & Brown, A. Web Analytics Definitions – Version 4.0. URL: <http://www.digitalanalyticsassociation.org/>
4. Data Age 2025: The digitization of the world - From edge to core. White paper. Farmingham, MA. URL: <https://www.seagate.com/files/www-content/our-story/trends/files/idcseagate-dataage-whitepaper.pdf>.
5. Google Analytics (n.d.). Features. URL: <http://www.google.com/analytics/features/>
6. GCP. URL: <https://cloud.google.com/solutions/retail>.

Знахур Людмила Володимирівна, ст. викл. каф. інформаційних систем, Харківський національний економічний університет ім. С. Кузнеця, тел. +38 050-65-11-282, razinalv@gmail.com,
Знахур Сергій Вікторович, к.е.н., доц. каф. інформаційних систем, Харківський національний економічний університет ім. С. Кузнеця, тел. +38 050-30-04-390, serhii.znakhur@gmail.com.

Implementation and rating of Web-analysis solutions based on cloud services

Abstract. *The total cost of ownership WEB-analytics of e-commerce sites is a very important aspect of the Internet business effectiveness. The thesis includes the modern solutions, architectures for solving the cross-cutting problem of creating and implementing*

WEB-analytics of e-commerce sites. Goal. The goal is to determine the optimal architecture of cloud services for the implementation of WEB-analytics of e-commerce sites. The object of research is the introduction of information technology, services for building and implementation of ML models based on data WEB-analytics of e-commerce sites. The subject of the research is the architecture development for WEB-analytics and ML analysis of Google Analytics data. The scientific novelty of the work is to develop a methodology for rating cloud providers solutions which are compared according to the parameters of functionality and cost of the ownership.

Methodology. *The system analysis and analytical and empirical methods of research were used for development and testing of cloud solutions. Results.* *Solving includes investigation of the necessary tools, standards and indicators of WEB-analytics, architectures, technologies, services for the implementation of sites and integration of Google Analytics, ML solutions. Cloud solutions of different architectures were developed and tested. Practical value.* *Rating methodology was used for the best Cloud solutions estimation.*

Key words: *e-commerce, ML, WEB-analytics, TCO, architecture, cloud service.*

Znakhur Liudmyla, Lecturer of Information Systems Department, Simon Kuznets Kharkiv National University of Economics, tel. +38 050-65-11-282, razinalv@gmail.com,
Znakhur Serhii, Assoc. Prof. Information Systems Department, Simon Kuznets Kharkiv National University of Economics, tel. +38 050-300-43-90, serhii.znakhur@gmail.com.