



International Science Group

ISG-KONF.COM

XXXVI

**INTERNATIONAL SCIENTIFIC
AND PRACTICAL CONFERENCE
"CURRENT TRENDS IN THE DEVELOPMENT OF YOUTH
THEORIES"**

Ankara, Turkey

September 12 - 15, 2023

ISBN 979-8-89145-199-5

DOI 10.46299/ISG.2023.1.36

CURRENT TRENDS IN THE DEVELOPMENT OF YOUTH THEORIES

Proceedings of the XXXVI International Scientific and Practical Conference

Ankara, Turkey
September 12 – 15, 2023

UDC 01.1

The 36th International scientific and practical conference “Current trends in the development of youth theories” (September 12 – 15, 2023) Ankara, Turkey. International Science Group. 2023. 212 p.

ISBN – 979-8-89145-199-5

DOI – 10.46299/ISG.2023.1.36

EDITORIAL BOARD

<u>Pluzhnik Elena</u>	Professor of the Department of Criminal Law and Criminology Odessa State University of Internal Affairs Candidate of Law, Associate Professor
<u>Liudmyla Polyvana</u>	Department of Accounting and Auditing Kharkiv National Technical University of Agriculture named after Petr Vasilenko, Ukraine
<u>Mushenyk Iryna</u>	Candidate of Economic Sciences, Associate Professor of Mathematical Disciplines, Informatics and Modeling. Podolsk State Agrarian Technical University
<u>Prudka Liudmyla</u>	Odessa State University of Internal Affairs, Associate Professor of Criminology and Psychology Department
<u>Marchenko Dmytro</u>	PhD, Associate Professor, Lecturer, Deputy Dean on Academic Affairs Faculty of Engineering and Energy
<u>Harchenko Roman</u>	Candidate of Technical Sciences, specialty 05.22.20 - operation and repair of vehicles.
<u>Belei Svitlana</u>	Ph.D., Associate Professor, Department of Economics and Security of Enterprise
<u>Lidiya Parashchuk</u>	PhD in specialty 05.17.11 "Technology of refractory non-metallic materials"
<u>Levon Mariia</u>	Candidate of Medical Sciences, Associate Professor, Scientific direction - morphology of the human digestive system
<u>Hubal Halyna Mykolaiivna</u>	Ph.D. in Physical and Mathematical Sciences, Associate Professor

37.	Коробко Т.О. АНАЛІЗ СУЧАСНИХ ПІДХОДІВ ДО ВИВЧЕННЯ МЕТОДІВ ФОРМУВАННЯ КОМУНІКАТИВНИХ УМІНЬ СТУДЕНТІВ ТРАНСПОРТНОЇ ГАЛУЗІ ОСВІТИ	167
38.	Чжао Жуйчень ФОРМУВАННЯ МІЖКУЛЬТУРНОЇ КОМПЕТЕНТНОСТІ МАЙБУТНІХ УЧИТЕЛІВ: МІЖНАРОДНИЙ ДОСВІД	170
TECHNICAL SCIENCES		
39.	Брусило Д.О., Гладких І.В. ВЗАЄМОДІЯ МІЖ НАУКОВЦЯМИ, ІНЖЕНЕРАМИ ТА ІНШИМИ СТЕЙКХОЛДЕРАМИ У ПРОЦЕСІ ІНВЕСТИЦІЙНО- ІННОВАЦІЙНОГО РОЗВИТКУ ЗАЛІЗНИЧНОЇ ГАЛУЗІ	174
40.	Корчак М.М., Лісевич О.В. ОБҐРУНТУВАННЯ ПРОЦЕСУ УЩІЛЬНЕННЯ ДЛЯ РІВНОМІРНОГО ПОДРІБНЕННЯ ЗАЛИШКІВ КУКУРУДЗИ	177
41.	Кіяшко Д. АНАЛІЗ ВИКОРИСТАННЯ МІКРОСЕРВІСІВ У ВЕБЗАСТОСУНКАХ	185
42.	Матківський С. ОПТИМІЗАЦІЯ САЙКЛІНГ-ПРОЦЕСУ ПРИ РОЗРОБЦІ ГАЗОКОНДЕНСАТНИХ РОДОВИЩ ІЗ ЗНАЧНИМИ ЗАПАСАМИ КОНДЕНСАТУ	189
43.	Матїїшин Л., Норка В. ВИКОРИСТАННЯ СУЧАСНИХ ПРОГРАМНИХ ПРОДУКТІВ ДЛЯ ПРОЄКТУВАННЯ ОПТИМАЛЬНОЇ СІТКИ ЕКСПЛУАТАЦІЙНИХ СВЕРДЛОВИН	194
44.	Самойленко Б. АРХІТЕКТУРА DATA PIPELINE ДЛЯ ЗБОРУ ДАНИХ ПРО АКТИВНІСТЬ КОРИСТУВАЧІВ	197
45.	Шишацький А.В., Гаман О.В., Шкнай О.В., Налапко О.Л., Ляшенко Г.Т. АНАЛІЗ ДОЦІЛЬНОСТІ ВИКОРИСТАННЯ ЕКСПЕРТНИХ СИСТЕМ В СИСТЕМАХ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ	202

АРХІТЕКТУРА DATA PIPELINE ДЛЯ ЗБОРУ ДАНИХ ПРО АКТИВНІСТЬ КОРИСТУВАЧІВ

Самойленко Богдан,
магістрант кафедри інформатики
Харківський національний університет радіоелектроніки

Метою дослідження є розробка архітектури data pipeline (конвеєр даних) для збору даних про активність користувачів із подальшим використанням цих даних для прийняття бізнес рішень та побудови звітів. Розглянуто концепцію побудови конвеєру даних, основні етапи та за що вони відповідають, розроблено архітектуру побудови конвеєру за допомогою хмарних технологій.

Конвеєр даних – це набір елементів обробки даних, які взаємопов’язані між собою, для автоматизованого збору, перетворення та передачі даних з одного джерела до іншого або з одного етапу обробки до іншого [1-3]. Необроблені дані збираються з різних джерел та подаються до сховища даних, такого як data lake (сховище необроблених даних) або data warehouse (сховище структурованих даних), для подальшого аналізу та історичного зберігання. Перед тим, як дані надійдуть до сховища, вони зазвичай піддаються певній обробці. Це може включати в себе перетворення даних, такі як фільтрація, маскування та агрегація, для забезпечення їх інтеграції та стандартизації. Це особливо важливо, коли дані призначені для реляційної бази даних, оскільки вона вимагає певної структури і відповідності типів даних [1].

Організовані конвеєри даних надають підґрунтя для різноманітних проєктів з аналізу даних, включаючи пошуковий аналіз даних, візуалізацію даних, машинне навчання, комп’ютерний зір [4-9].

Загалом конвеєр даних складається із наступних етапів (рис. 1) [2, 3]:

– збір (Collection). На цьому етапі дані збираються з різних джерел. Збір може включати в себе регулярні запити для отримання нових даних або періодичні завантаження;

– поглинання (Ingestion). Дані переносяться із джерела до сховища даних, це може бути інтеграція в реальному часі або пакетні завантаження;

– зберігання даних (Data storage). Часто необроблені дані зберігаються в data lake, потім вони очищаються, дублікати та аномалії видаляються та трансформуються відповідно до схеми. Потім ці готові до використання дані зберігаються в data warehouse [9];

– обчислення (Computation). На цьому етапі дані піддаються аналізу, обчисленням і витягненню цінної інформації. Це може включати в себе створення звітів, агрегацію даних, виконання аналізу рядів часу, машинне навчання та інші обчислювальні операції, такі як розпізнавання [10-15];

– презентація (Presentation). На останньому етапі дані відображаються для користувача або інших систем. Це може включати в себе застосування business intelligence платформ, створення звітів, візуалізацію даних, створення інформаційних панелей, вебсайтів або API для доступу до даних. Представлення

даних допомагає користувачам отримувати доступ до аналітичної інформації та приймати рішення на основі оброблених даних [16, 17].

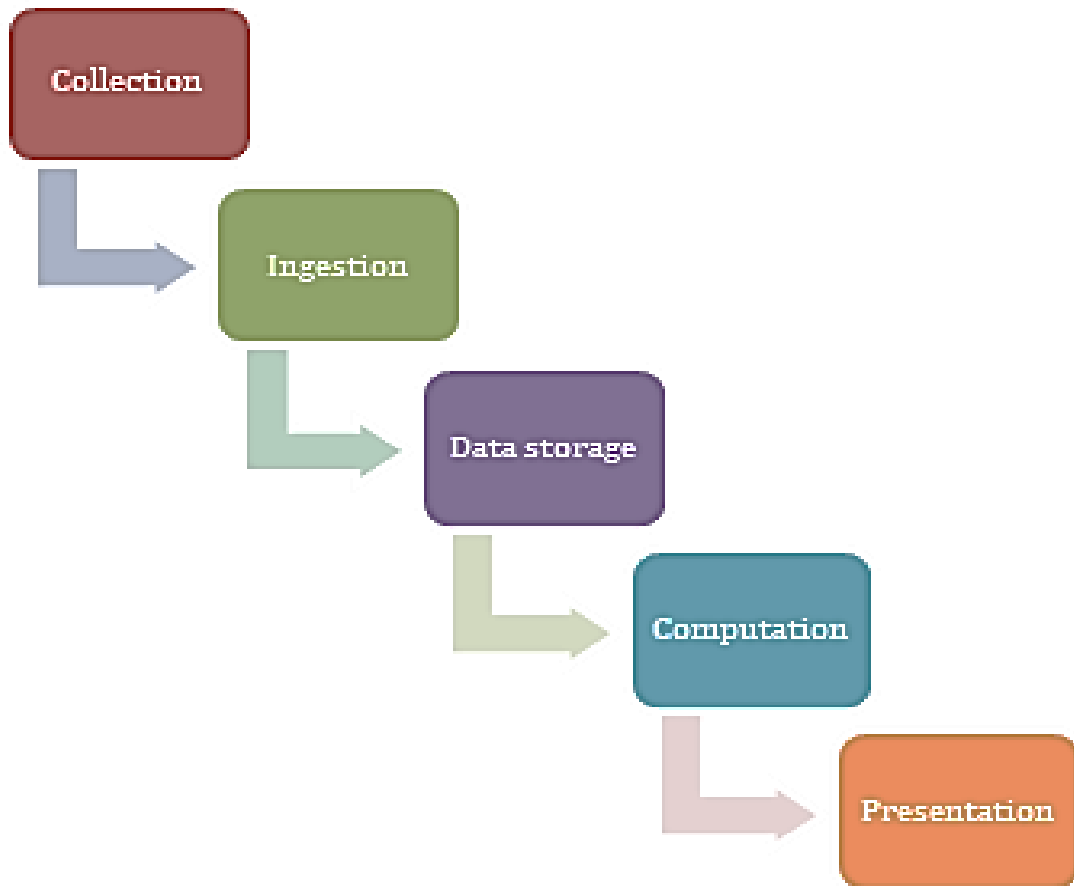


Рисунок 1 – Загальна модель конвеєру даних

Етапи конвеєру даних можуть варіюватися залежно від конкретного використання і потреби в аналізі даних, але загальна ідея полягає в тому, щоб забезпечити послідовний та ефективний потік даних від джерела до результату.

Описана модель була використана для створення конвеєру, який може бути застосований у сфері освіти. Наприклад в рамках платформи дистанційного навчання запропонований підхід допомагає визначити, скільки часу студенти витратили на проходження курсу, що для деяких дисциплін є показником того, чи надавати студенту сертифікат про успішне проходження.

Архітектура конвеєру даних була розроблена за допомогою хмарних технологій, які надає AWS (Amazon Web Services) (рис. 2) [18, 19].

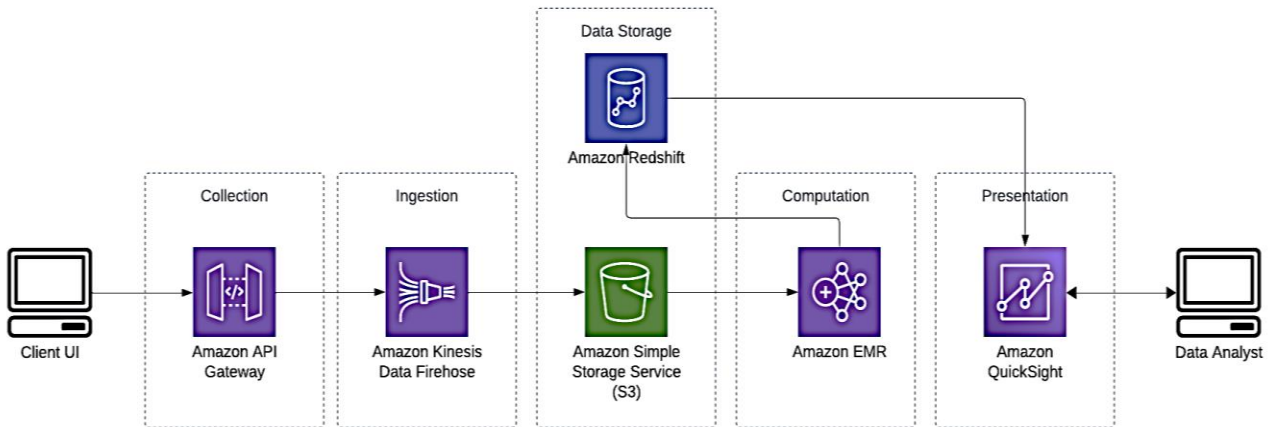


Рисунок 2 – Архітектура data pipeline для збору даних про активність користувачів

Процес даного конвеєру полягає в наступному:

– коли користувач виконує якусь дію на UI (User Interface), то клієнтський код відправляє HTTP запит на API (Application Programming Interface) з тілом, яке має корисні дані про цю дію, наприклад, коли вона була зроблена, та унікальне її ім'я. API розвернуто за допомогою Amazon API Gateway сервісу, який дозволяє створювати Serverless API, тобто розробникам не потрібно підтримувати належну інфраструктуру (наприклад, сервери) для роботи даного API. Управління інфраструктурою бере на себе сам AWS;

– API Gateway, в свою чергу, інтегрований з Amazon Kinesis Data Firehose, що дозволяє завантажити потокові дані, для даного конвеєру – це повідомлення про виконану дію користувачем, до сховища даних (data lake);

– сховищем неструктурованих даних (data lake) виступає Amazon S3, даний сервіс дозволяє зберігати необмежений обсяг даних;

– неструктуровані дані потрібно привести до певної схеми, яку очікує data warehouse, для цього завдання використовується Amazon EMR. Цей сервіс дозволяє виконувати розподілені обчислення, тобто використовувати для обробки даних декілька серверів у мережі. Для використання потужностей кластера код для цього завдання потрібно розробити за допомогою Apache Spark – фреймворк, який виконує розподілені обчислення на кластері. Зі заданою періодичністю на Amazon EMR запускається завдання, реалізоване Apache Spark, яке завантажує неструктуровані дані із Amazon S3, трансформує їх відповідно до схеми та завантажує структуровані дані до data warehouse, для розгортання якого використовується Amazon Redshift [20];

– дані, завантажені до data warehouse, готові для використання. Аналітики даних можуть використовувати Amazon QuickSight для побудови звітів, які допоможуть встановити функціональність, що в свою чергу дозволяє компанії приймати ті чи інші бізнес-рішення.

У даного конвеєра є переваги перед іншим готовим рішенням, таким як Google Tag Manager (GTM), для збору даних про активність:

– GTM запити блокуються розширеннями браузерів для блокування реклами, тому в сферах, де точність даних дуже важлива, такі як освіта, фінанси, розробка власного конвеєру даних є необхідною;

– дані обробляються набагато швидше. Затримка від моменту, коли користувач виконав якусь дію, до моменту, коли дані звантажені до data warehouse, під час використання GTM може досягати декілька годин. Завдяки пропонованому рішенню вона може бути зменшена до декількох хвилин.

Список літератури:

1. Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences*, 30(4), 431-448.

2. Alwan, H. B., & Ku-Mahamud, K. R. (2020, February). Big data: Definition, characteristics, life cycle, applications, and challenges. *In IOP Conference Series: Materials Science and Engineering* (Vol. 769, No. 1, p. 012007). IOP Publishing.

3. Oleghe, O., & Saloniis, K. (2020). A framework for designing data pipelines for manufacturing systems. *Procedia CIRP*, 93, 724-729.

4. Гороховатский, В.А., Пуятин, Е.П., Столяров, В.С. (2017) Исследование результативности структурных методов классификации изображений с применением кластерной модели данных. *Радиоэлектроника, информатика, управление*, №3 (42). С. 78-85.

5. Гороховатський В.О., Пупченко Д.В., Солодченко К.Г. (2018) Аналіз властивостей, характеристик та результатів застосування новітніх детекторів для визначення особливих точок зображення. *Системи управління, навігації та зв'язку*, С. 93-98.

6. Gorokhovatskyi V., Gadetska S., Ponomarenko R. (2020) Recognition of Visual Objects Based on Statistical Distributions for Blocks of Structural Description of Image. Proc. of the XV Int. Scientific Conference “Intellectual Systems of Decision Making and Problems of Computational Intelligence” (ISDMCI'2019), Ukraine, May 21–25, 2019, pp. 501-512.

7. Pomazan V., Tvoroshenko I., and Gorokhovatskyi V. (2023) Development of an application for recognizing emotions using convolutional neural networks, *International Journal of Academic Information Systems Research*, 7(7), pp. 25-36.

8. Гороховатский В.А. Структурный анализ и интеллектуальная обработка данных в компьютерном зрении: монография, Комп. СМИТ, 2014. 316 с.

9. Kleppmann, M. (2019). *Designing Data-Intensive Applications*.

10. Gorokhovatskyi, V., Peredrii, O., Tvoroshenko, I., Markov, T. (2023) Матриця відстаней для множини компонентів структурного опису як інструмент для створення класифікатора зображень. *Advanced Information Systems*, 7(1), С. 5-13.

11. Gadetska S., Gorokhovatskyi V., Stiahlyk N., Vlasenko N. (2022) Aggregate Parametric Representation of Image Structural Description in Statistical Classification Methods. *In CEUR Workshop Proceedings: Computer Modeling and Intelligent Systems (CMIS-2022)*, 3137, pp. 68-77.

12. Gorokhovatskyi, V., Tvoroshenko, I., Kobylin, O., Vlasenko, N. (2023) Search for Visual Objects by Request in the Form of a Cluster Representation for the Structural Image Description. *Advances in Electrical and Electronic Engineering*, 21(1), pp. 19-27.
13. Gorokhovatskyi, V., Vlasenko, N. (2021). Редукція опису зображення у складі множини дескрипторів на основі метричного критерію інформативності. *Advanced Information Systems*, 5(4), pp. 10-16.
14. Гороховатский, В.А., Путятин, Е.П. (2008) Структурное распознавание изображений на основе моделей голосования признаков характерных точек. *Реєстрація, зберігання і обробка даних*, Т. 10, № 4. С.75-85.
15. Gorokhovatskyi, O., Peredrii, O., Gorokhovatskyi, V., Vlasenko, N. (2023) Explanation of CNN Image Classifiers with Hiding Parts. In: J. Benois-Pineau, R. Bourqui, D. Petkovic, G. Quenot (eds), *Explainable Deep Learning Artificial Intelligence*, pp. 125-146, Academic Press, 346 p.
16. Tvoroshenko I. S. and Tabashnyk V. A. (2018) Development of a spatial model of geoinformation support for people with disabilities in wheelchairs in Kharkiv. *Collection of scientific works of KhNUPS*, vol. 1, no. 55, pp. 122-128.
17. Творошенко, І. С. (2018). Особливості застосування сучасних принципів штучного інтелекту до розробки ефективних механізмів моделювання складних систем. *Science and Technology of the Present Time: Priority Development Directions of Ukraine and Poland*, 118-121.
18. Srivastava, P., & Khan, R. (2018). A review paper on cloud computing. *International Journal of Advanced Research in Computer Science and Software Engineering*, 8(6), 17-20.
19. AWS Documentation. URL: <https://docs.aws.amazon.com/> (дата звернення 04.09.2023).
20. Lu, T., Hoyer, S., Wang, Q., Hu, L., & Chen, Y. F. (2021, July). Distributed data processing for large-scale simulations on cloud. In *2021 IEEE International Joint EMC/SI/PI and EMC Europe Symposium* (pp. 53-58). IEEE.