

ЗБІР ТА ОБРОБКА КОНФІГУРАЦІЙ ГІБРИДНИХ ХМАР**Малініч І. П., Іванчук Я. В.**

(malinich@vntu.edu.ua, ivanchuck@vntu.edu.ua)

Вінницький національний технічний університет (Україна)

При проведенні досліджень гібридних хмар необхідно зібрати різні хмарні конфігурації та переконвертувати їх в уніфікований формат. Це можуть бути як складні конфігурації Google Anthos, Azure Arc, Rancher та Terraform, так і статті чи навіть зображення на веб-сайтах Інтернету. Якщо до появи загальнодоступних LLM-моделей їх було складно знаходити, обробляти та конвертувати в уніфікований формат, то тепер це стало набагато простіше зробити. В процесі пошуку можна знайти різні джерела використовуючи пошукову базу таких сервісів як GitHub, Google та можливості ChatGPT deep research. Для конвертації конфігурацій в уніфікований формат можна використати API-доступ до LLM-моделі OpenAI GPT-4.1. Іншим інструментом за допомогою якого зі картинок статей можна отримати більше структурних даних є Google Vertex AI Vision.

Для успішних досліджень у сфері гібридних хмар необхідно знайти та підготувати якісні початкові дані об'єкта дослідження – гібридних хмар [1, 2]. Найпростіший спосіб їх знайти – це виконати пошук сценаріїв розгортання Rancher, Terraform, Google Anthos чи Azure Arc [3] на сервісі онлайн-репозиторіїв GitHub, що власне і було зроблено при початку роботи над інформаційною базою дослідження. Серед знайдених сценаріїв було обрано лише ті, що мають стосунок до гібридних хмар та мультихмарного розгортання. Для видобування даних з джерел у вигляді сценаріїв розгортання, статей, форумів та зображень використано зв'язку хмарних інструментів з LLM-моделей та машинного зору (рис. 1). Часто в подібних матеріалах структура гібридної хмари не описана явно у вигляді взаємозв'язної структури яку можна просто переконвертувати.

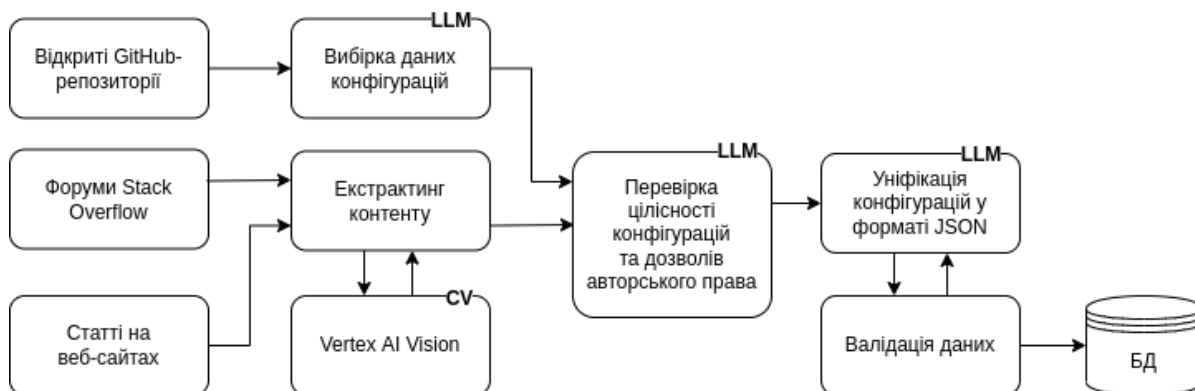


Рисунок 1 – Зв'язка хмарних інструментів штучного інтелекту для видобування

В статтях та обговореннях можуть бути описані різні сервіси, зв'язки між ними, їх функціонал, а також в якій частині гібридної хмари вони знаходяться. Для екстракції основного контенту із сторінки сайту використано бібліотек Readability від Mozilla, яка дозволяє відсіяти вторинне наповнення сторінки та залишити лише основний контент. Даний інструмент застосовується у браузері Mozilla Firefox для активації режиму читання, прибираючи сторонні елементи веб-сайту. Іншою проблемою є те в самій статті може не бути детально описано структуру хмари, а сама структура може бути зображена на картинках, як до прикладу це зроблено у статті блогу [4]. Сервіс розпізнавання Vertex AI Vision дозволяє виокремити окремі об'єкти на зображенні та виявити між ними зв'язки у вигляді ліній чи стрілок [5]. Також він може розпізнати текстові підписи під об'єктами за допомогою OCR. На жаль даний сервіс не орієнтується у семантиці хмарних структурних діаграм, тому у зв'язці з ним необхідно використовувати LLM-модель, яка зможе інтерпретувати виявлені Vertex AI Vision об'єкти та пояснити компоненти і типи з'єднань [5].

Наступним чином потрібно проводити перевірку кожного екстрактованого матеріалу з текстового та/або графічного контенту на предмет наявності в ньому цілісного опису хмарної інфраструктури, а також на відсутність обмежень авторського права на повторне використання чи обробку засобами штучного інтелекту. У різних матеріалах може бути різний рівень абстракції хмарної інфраструктури. Дане дослідження зосереджує фокус на платформному та сервісному рівнях абстракції хмар, в той час як інфраструктурних рівень хмар не містить корисних даних для даного дослідження. Тому при застосуванні LLM-моделей для відбору матеріалів для подальшої обробки необхідно задати інструкції для виявлення хмарних сервісів, виявлення між ними зв'язків та належності до приватної чи публічної частини хмари.

Незважаючи на простоту перевірки їх цілісності, обробка сценаріїв Terraform також є досить складним завданням, оскільки різні процеси сценарію розгортання можуть бути розосереджені по різних файлах. Але виявлення зв'язків між сервісами може бути менш очевидним, оскільки в коді сценаріїв розгортання може не бути описано характеру зв'язків, а виявлені вони можуть бути лише в процесі виконання. Потенційними ознаками зв'язку може слугувати визначення змінних оточення для того чи іншого сервісу, яких описуються параметри з'єднання з іншими сервісами, оскільки у Docker та Kubernetes зв'язки задаються через змінні оточення. При розгортанні створюються контейнери, які отримують певне визначене сценарієм розгортання ім'я на внутрішньому DNS-сервері кластеру [6]. Наприклад при розгортанні веб-додатку є чотири складові: фронтенд, бекенд, база даних та кеш, кожна складова має своє ім'я. Фронтенд матиме відповідні змінні для зв'язку із бекендом та кешем, бекенд у свою чергу матиме зв'язок із базою даних за наперед визначеними параметрами доступу. Таким чином можна встановити зв'язки між ними із зазначенням залежностей. Виявлення таких зв'язків може бути визначено у вигляді окремої інструкції для LLM-моделей, що працюватимуть з git-репозиторіями. Крім того навіть у такому випадку може порушуватись зв'язність між окремими сервісами. Для встановлення відсутніх зв'язків у випадку появи острівних сервісних осередків можуть бути використані можливості більш складних LLM-моделей для їх з'єднання. При перевірці цілісності варто ввести абстракції таких об'єктів як хмарні сервіси, зв'язки між ними (залежності), мережі, а також

мережеві шлюзи та екрани як їх складові, які після збору та обробки даних будуть структуровані у вигляді наборів даних JSON (рис. 2).

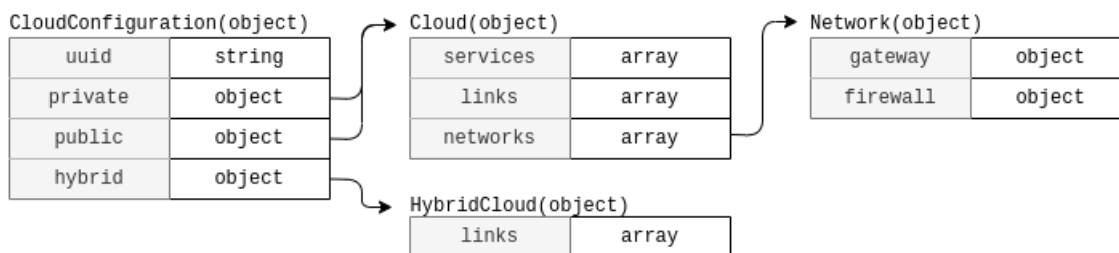


Рисунок 2 – Ієрархія об'єктів набору даних JSON

Іншою стороною застосування LLM-моделей для уніфікації хмарних конфігурацій є ризик спотворення даних – як внаслідок галюцинацій [7] та ігнорування контексту [8], так і прикладних помилок, наприклад створення помилкових залежностей або конфліктуючих налаштувань. Крім того та чи інша модель LLM-модель може покладатись на певні шаблони із своїх навчальних даних і таким чином може ускладнити хмарну конфігурацію при її уніфікації. Це можна виправити на етапі роботи алгоритму валідації даних (рис. 1). Для кожної помилки можна створити набір інструкцій для LLM-моделей, які дозволять їх виправити.

Висновок. В результаті дослідження в Інтернеті знайдено 401 хмарну конфігурацію, які задовільняли базові вимоги вибірки. З них було відкинуто 31 конфігурацію через ліцензійні обмеження. Зібрано структуровані дані про 158 хмарних конфігурацій, в яких сумарно знаходяться дані про 1207 хмарних сервісів. Решта 212 хмарних конфігурації на момент написання публікації не мали достатньої цілісності для подальшої роботи. Метадані та джерела хмарних конфігурації розміщено у репозиторії GitHub [9]. В подальших дослідженнях планується виконати більш

Список використаної літератури

- [1]І. П. Малініч та Я. В. Іванчук, "Особливості розміщення мікросервісів систем управління навчанням у гібридних хмарах," System technologies, no. 3(158), 2025, pp. 157-170.
- [2]І. П. Малініч та Я. В. Іванчук, "Показники оцінки мережевої доступності та зв'язності інформаційних систем у обчислювальних хмарах," у Матеріали 2-ої Міжнародної науково-практичної конференції «Інформаційні системи та технології: результати і перспективи», Київ, Україна, 5 березня 2025 р. – Київ: ФІТ КНУТШ, 2025, с. 248-251.
- [3]C. Luca, "Architecture of Multi-Cloud Kubernetes Environments," Preprint, Mar. 2024. [Online]. Available: https://www.researchgate.net/publication/388527469_Architecture_of_Multi-Cloud_Kubernetes_Environments [Accessed: Apr. 16, 2025].
- [4]B. Downey, "Building a Kubernetes Hybrid Cloud with Terraform," ITNEXT. [Online]. Available: <https://itnext.io/building-a-kubernetes-hybrid-cloud-with-terraform-fe15164b35fb> [Accessed: June 9, 2025].
- [5]G. S. Manikanta, D. Likitha, B. Hareesh, G. Ajay Kamal, and A. Subha, "Object Detection Using Vertex AI AutoML," International Journal of Innovative Research in Science Engineering and Technology, vol. 14, no. 4, pp. 9304–9312, 2025.
- [6]M. Waseem, A. Ahmad, P. Liang, M. A. Akbar, A. A. Khan, I. Ahmad, M. Setälä, and T. Mikkonen, "Containerization in Multi-Cloud Environment: Roles, Strategies, Challenges, and Solutions for Effective Implementation," arXiv preprint arXiv:2403.12980, 2025. [Online]. Available: <https://arxiv.org/abs/2403.12980> [Accessed: Sep. 26, 2025].
- [7]A. Yarovy, D. Kudriavtsev, S. Baraban, V. Ozeranskyi, L. Krylyk, A. Smolarz, and G. Karnakova, "Information technology in creating intelligent chatbots," Proc. SPIE 11176, Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments, 2019, Art. no. 1117627, doi: 10.1117/12.2537415.
- [8]A. Yarovii, D. Kudriavtsev, and P. Olena, "Improving the accuracy of text message recognition with an intelligent chatbot information system," in Proc. 2020 IEEE 15th Int. Conf. Comput. Sci. Inf. Technol. (CSIT), Zbarazh, Ukraine, 2020, pp. 76–79, doi: 10.1109/CSIT49958.2020.9322036.

Proceedings of the XVIII International scientific and practical conference «Information technologies and automation– 2025»

[9]Hybrid Cloud Configurations Metadata Collection. GitHub repository [Online]. Available: <https://github.com/linuxation/hybrid-scan/> [Accessed: Sep. 26, 2025].