

УДК 378.096:004.738.5

О.Ю. Недоснований¹, О.І. Черняк¹, В.В. Голінко²

ПОРІВНЯЛЬНИЙ АНАЛІЗ ХМАРНИХ СЕРВІСІВ ДЛЯ ОБРОБКИ ГЕОІНФОРМАЦІЙНИХ ДАНИХ

¹Вінницький національний технічний університет, Вінниця²Київський національний університет технологій та дизайну, Київ

Анотація. Стаття присвячена порівняльному аналізу хмарних сервісів для обробки геоінформаційних даних. У ній детально розглядаються сервіси - Google Cloud, Amazon Web Services та Microsoft Azure, що надають засоби для зберігання, обробки та аналізу великих обсягів географічних даних. Також наведено параметри геоінформаційних сервісів, алгоритм доступу та приклади програмного коду для обробки супутникових даних. У статті описуються такі можливості та обмеження використання хмарних сервісів, як автоматизованість, безпека та масштабованість. Надано висновки та рекомендації для подальшого розвитку систем обробки геоінформаційних даних на основі хмарних сервісів. Сервіси Amazon Web Services (AWS), Microsoft Azure та Google Cloud Platform (GCP), пропонують різноманітні рішення для зберігання геоданих. Ці рішення включають об'єктні сховища, такі як Amazon S3, Azure Blob Storage та Google Cloud Storage, а також геопросторові бази даних, наприклад, Amazon RDS, Azure Cosmos DB та Google Cloud Firestore. Крім того, кожен з цих сервісів надає набір сервісів для аналізу та обробки геоінформаційних даних. Наприклад, AWS пропонує такі сервіси, як Amazon Athena, Amazon Redshift та AWS Glue, які дозволяють виконувати SQL-запити, проводити аналітику та інтегрувати геодані з іншими сервісами. Azure має в своєму арсеналі сервіси, такі як Azure SQL Database, Azure Databricks та HDInsight, які забезпечують можливості для обробки та аналізу геоінформаційних даних. GCP також надає сервіси, такі як BigQuery, Dataflow та Dataproc, які дозволяють виконувати аналітичні операції та обробку великих обсягів геоданих. Підтримка сервісами інтеграції з різноманітними геоінструментами є важливою для аналізу, наприклад, AWS, Amazon Location Service, Amazon Ground Truth та Amazon Rekognition - дозволяють працювати з геоданими на різних рівнях складності. Azure має Azure Maps, який надає сервіси для геокодування, маршрутизації та візуалізації геоданих. GCP також пропонує Google Maps Platform, який надає широкі можливості для інтеграції з географічними технологіями, такими як маршрутизація, геокодування та візуалізація карт. Усі ці процеси дозволять ефективніше обробляти дані.

Ключові слова: хмарні сервіси, геоінформаційні системи, обробка геоінформаційних даних, аналіз зображень, обробка супутникових даних, вартість, безпека, масштабованість, ефективність, продуктивність, рекомендації.

Abstract. The article is devoted to a comparative analysis of cloud services for processing geographic data. It describes in detail the services - Google Cloud, Amazon Web Services and Microsoft Azure - that provide tools for storing, processing and analyzing large amounts of geographic data. The article also describes the parameters of geoinformation services, the access algorithm, and examples of program code for processing satellite data. The article describes such opportunities and limitations of using cloud services as automation, security and scalability. The conclusions and recommendations for further development of geographic information systems based on cloud services are provided. Services. Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) offer a variety of geodata storage solutions. These solutions include object storage, such as Amazon S3, Azure Blob Storage, and Google Cloud Storage, as well as geospatial databases, such as Amazon RDS, Azure Cosmos DB, and Google Cloud Firestore. In addition, each of these services provides a set of services for analyzing and processing geographic information data. For example, AWS offers services such as Amazon Athena, Amazon Redshift, and AWS Glue, which allow you to run SQL queries, conduct analytics, and integrate geodata with other services. Azure offers services such as Azure SQL Database, Azure Databricks, and HDInsight, which provide capabilities for processing and analyzing geographic data. GCP also provides services such as BigQuery, Dataflow, and Dataproc, which allow you to perform analytical operations and process large amounts of geodata. Support for integration with various geo-tools is important for analysis, such as AWS, Amazon Location Service, Amazon Ground Truth, and Amazon Rekognition, which allow you to work with geodata at different levels of complexity. Azure has Azure Maps, which provides services for geocoding, routing, and visualization of geodata. GCP also offers Google Maps Platform, which provides extensive integration with geographic technologies such as routing, geocoding, and map visualization. All these processes will allow for more efficient data processing.

Keywords: cloud technologies, geographic information systems, geographic information data processing, cloud services, image analysis, satellite data processing, machine learning, cost, security, scalability, testing and validation, efficiency, performance, recommendations.

DOI: <https://doi.org/10.31649/1999-9941-2023-57-2-50-57>.

Вступ

Зростання обсягів геоінформаційних даних та потреба у їх ефективній обробці породжують нові виклики та необхідності у їх покращенні. Одним із потужних інструментів для обробки геоінформаційних даних є хмарні сервіси. Вони забезпечують гнучкість, масштабованість та високу доступність, що дозволяє ефективно працювати з великими обсягами даних та розв'язувати складні географічні задачі.

Важливо підкреслити, що реалізація системи обробки геоінформаційних даних на основі хмарних сервісів вимагає ретельного аналізу вимог та вибору відповідних технологій. Окрім того, забезпечення безпеки даних, моніторинг та оптимізація продуктивності є важливими аспектами успішної реалізації хмарних сервісів.

Наші висновки та рекомендації отримані в результаті аналізу досвіду найкращих практик у цій області.

У статті розглянуто різні аспекти використання хмарних сервісів, включаючи обробку геоінформаційних даних, платформи, методи та алгоритми обробки даних а також надано висновки та рекомендації для подальшого розвитку цих систем.

Актуальність

Розширення галузі використання геоінформаційних систем вимагає швидкої та ефективної обробки великих обсягів даних. Проте можливість реалізації їх традиційними методами є обмеженою через недостатній обсяг ресурсів. У такому контексті хмарні сервіси надають потужні інструменти для вирішення цих задач, забезпечуючи гнучкість, автоматизованість, масштабованість та практично необмежену доступність ресурсів.

Мета

Метою статті є детальний аналіз та порівняння хмарних сервісів, здатних обробляти геоінформаційні дані та надання практичних рекомендацій з їх реалізації.

Задачі дослідження використання хмарних сервісів для обробки геоінформаційних даних включають такі аспекти:

- проведення огляду різних хмарних сервісів, їх функціональних можливостей, пропонувананих сервісів та особливостей інтеграції з геоінформаційними системами.
- аналіз методів та алгоритмів обробки геоінформаційних даних: дослідження існуючих методів та алгоритмів для обробки даних у сервісі на основі геопроектингу.
- формулювання висновків та рекомендацій щодо використання хмарних сервісів у геоінформаційному аналізі.

Ця робота спрямована на підвищення розуміння використання хмарних сервісів для обробки геоінформаційних даних і на забезпечення практичних рекомендацій для дослідників у цій області.

Основна частина

Хмарні сервіси та геоінформаційні системи (ГІС) є двома пов'язаними областями.

Хмара (Cloud) - це великий пул обчислювальних ресурсів (сервери, сховища даних, мережі), доступних через Інтернет. Хмарні платформи надають інфраструктуру для зберігання та обробки даних без необхідності наявності власного обладнання [2].

Обробка геоінформаційних даних в хмарних сервісах, таких як Amazon Web Services (AWS), Microsoft Azure та Google Cloud Platform (GCP), відкриває безліч можливостей для аналізу та використання геопросторової інформації. Основні компоненти обробки геоінформаційних даних включають зберігання, аналіз, візуалізацію та інтеграцію з іншими сервісами.

Усі три хмарні сервіси (AWS, Azure та GCP) надають різноманітні інструменти для роботи з геоданими.

Платформи пропонують різні сховища для зберігання геоданих, такі як об'єктні сховища (S3 в AWS, Blob Storage в Azure, Cloud Storage в GCP) або геопросторові бази даних (Amazon RDS, Azure Cosmos DB, Google Cloud Firestore).

Кожна платформа надає різні сервіси для аналізу та обробки геоінформаційних. Наприклад, AWS має такі, як Amazon Athena, Amazon Redshift та AWS Glue, які дозволяють виконувати SQL-запити, проводити аналітику та інтегрувати геодані з іншими сервісами. Azure має сервіси, такі як Azure SQL Database, Azure Databricks та HDInsight, які надають можливості для обробки та аналізу геоінформаційних даних. GCP має сервіси, такі як BigQuery, Dataflow та Dataproc, які дозволяють виконувати аналітичні операції та обробку великих обсягів геоданих.

Кожна платформа надає інтеграцію з різноманітними геоінструментами. Наприклад, AWS пропонує широкий вибір сервісів, таких як Amazon Location Service, Amazon Ground Truth та Amazon Rekognition, які дозволяють працювати з геоданими на різних рівнях складності. Azure має Azure Maps, який надає сервіси для геокодування, маршрутизації та візуалізації. GCP має Google Maps Platform, який надає широкі можливості для інтеграції з географічними технологіями, такими як маршрутизація, геокодування та візуалізація карт.

Загалом, обробка геоінформаційних даних в хмарних сервісів надає широкі можливості для аналізу, використання та інтеграції геопросторової інформації. Користувачі можуть ефективно працювати з геоданими, використовуючи потужні технології та інструменти, що надаються цими хмарними сервісами.

У таблиці 1 наведено короткий опис відмінностей у віртуалізації між Amazon Web Services (AWS), Microsoft Azure та Google Cloud Platform (GCP):

Кожна платформа пропонує свій власний сервіс віртуальних машин, які дозволяють розгортати та керувати віртуальними обчислювальними ресурсами.

Хмарні сервіси можуть легко змінювати масштаб, збільшуючи або зменшуючи обсяг обчислювальних ресурсів залежно від потреби дослідника.

У таблиці 2 буде наведено порівняльні характеристики масштабованості. Ці характеристики є важливими для хмарних сервісів і визначає їх здатність змінювати масштаб обчислювальних ресурсів в залежності від потреб користувачів.

Таблиця 1 – Характеристики відмінностей віртуалізації хмарних сервісів для обробки геоінформаційних даних

Характеристика	AWS	Azure	GCP
Віртуальні машини	Elastic Compute Cloud (EC2)	Virtual Machines	Compute Engine
Контейнери обробки	Elastic Container Service	Kubernetes Service (AKS)	Kubernetes Engine (GKE)
Сервіси безсерверного обчислення	AWS Lambda	Azure Functions	Cloud Functions
Сервіси для розгортання	Elastic Beanstalk	App Service	App Engine
Приватні мережі	Amazon VPC	Azure Virtual Networks	Google VPC
Інтеграція з геоданими	Широкий вибір геоінструментів та сервісів	Інтеграція з Azure Maps	Інтеграція з Google Maps
Масштабованість	Гнучке масштабування ресурсів	Масштабування за потребою	Автоматичне масштабування
Швидкість обробки	Висока швидкість обробки даних	Ефективна обробка даних	Швидка обробка даних

Таблиця надає порівняльну характеристику трьох провідних хмарних сервісів: Amazon Web Services (AWS), Microsoft Azure та Google Cloud Platform (GCP) за такими підсервісами всередині: віртуальні машини, контейнери обробки, сервіси безсерверного обчислення, додатки для розгортання, інтегратори геоданих, масштабованість та швидкість обробки.

Розглянемо більш детально ці характеристики.

Віртуальні машини. Усі три платформи мають інструменти для візуалізації геоданих. Наприклад, AWS має Amazon QuickSight, Azure має Power BI, а GCP має Data Studio. Ці сервіси дозволяють створювати динамічні графіки, картограми та інші візуалізації на основі геоданих. Також вони дозволяють розділити фізичні ресурси хмарної інфраструктури на віртуальні машини. Кожна віртуальна машина може працювати незалежно і мати свою операційну систему та програмне забезпечення.

Контейнери обробки. Контейнер Elastic Container Service (ECS): дозволяє розгортання й керування засобом Docker на інфраструктурі AWS. Kubernetes Service (AKS) – це сервіс, що пропонується Azure. Він дозволяє розгортати та керувати контейнерами з використанням оркеструвальника Kubernetes. Kubernetes Engine (GKE) - це сервіс, що пропонується Google Cloud Platform (GCP). Він надає можливість розгортати й керувати контейнерами з використанням оркеструвальника Kubernetes.

Сервіси безсерверного обчислення. AWS Lambda – це сервіс, що пропонується AWS. Він дозволяє виконувати код без необхідності управління серверами. Azure Functions – це сервіс, що пропонується Azure. Він дозволяє виконувати код без управління інфраструктурою. Cloud Functions – це сервіс, що пропонується Google Cloud Platform. Він також дозволяє виконувати код без необхідності управління інфраструктурою.

Додатки для розгортання. Elastic Beanstalk – це сервіс, що пропонується AWS, для швидкого розгортання і масштабування веб-додатків та служб, заснованих на контейнерах або платформах. App Service – це сервіси для розгортання додатків, що пропонується Azure. Він дозволяє розгортати веб-додатки та мобільні додатки, що мають зручний доступ.

Інтеграція з геоданими вимагає розуміння та обробки геоінформаційних даних для використання їх у різних додатках та системах. Ця інтеграція може включати отримання, зберігання, аналіз та візуалізацію геоданих.

Масштабованість означає здатність системи або додатку працювати ефективно з різними обсягами геоданих. Масштабованість важлива, оскільки геодані можуть бути дуже об'ємними, особливо при обробці великих територій або деталізованих картографічних даних. Система повинна мати можливість обробляти великі обсяги геоданих швидко і ефективно, щоб забезпечити продуктивність та зручність використання.

Геодані можуть бути складними для обробки через їх розмір та внутрішню структуру. Операції, такі як обчислення шляху, аналіз просторових взаємозв'язків або обчислення геостатистичних характеристик, можуть потребувати значних обчислювальних ресурсів. Ефективна обробка геоданих вимагає оптимізації алгоритмів та використання спеціалізованих інструментів для прискорення обчислень.

Існує кілька способів досягнення масштабованості та покращення швидкості обробки геоінформаційних даних. Одним з них є використання розподілених систем, які дозволяють обробляти геодані одночасно на кількох вузлах або серверах. Це дозволяє прискорити обробку.

Для досягнення масштабованості та високої швидкості обробки геоінформаційних даних потрібно враховувати можливість розширення ресурсів інфраструктури, таких як обчислювальна потужність, мережева пропускну здатність та використання сховищ даних. Оптимізація архітектури системи, використання кешування, компресії та інших технологій також можуть покращувати продуктивність та швидкість обробки геоданих. У таблиці 2 наведено порівняння масштабованості хмарних сервісів.

Таблиця 2 – Порівняння масштабованості хмарних сервісів для обробки геоінформаційних даних

AWS	Azure	GCP
Пропонує гнучке масштабування ресурсів, що означає, що користувачі можуть збільшувати або зменшувати обсяг обчислювальних ресурсів, таких як віртуальні машини, залежно від потреб своїх додатків або проєктів.	Надає масштабування за потребою, що означає, що хмарні ресурси можуть автоматично збільшуватись або зменшуватись в залежності від навантаження. Це дозволяє забезпечити оптимальне використання ресурсів та ефективне розподілення обчислювального навантаження.	Пропонує автоматичне масштабування, що означає, що хмарні ресурси автоматично реагують на зміну навантаження та адаптуються, щоб забезпечити потрібну продуктивність. Це спрощує управління ресурсами та забезпечує гнучкість при розгортанні додатків та послуг.

Таким чином, всі три платформи, AWS, Azure та GCP, надають різні підходи до масштабованості, що дозволяє користувачам ефективно управляти обчислювальними ресурсами та забезпечувати високу продуктивність своїх додатків у хмарному сервісі [5].

З таблиці 2 видно, що кращим з точки зору масштабування є сервіс GCP, оскільки завдяки автоматичному масштабуванню він дозволяє зменшувати витрати часу людської праці.

Самообслуговування (Self-Service): Користувачі можуть самостійно керувати та налаштовувати ресурси хмарних сервісів, використовуючи веб-інтерфейс або API. В таблиці 3 наведено переліки технологій самообслуговування хмарних сервісів найбільш відомих геоінформаційних систем.

Таблиця 3 – Технології самообслуговування хмарних сервісів для обробки геоінформаційних даних

Характеристика	AWS	Azure	GCP
Самообслуговування	AWS Management Console, AWS CLI, AWS SDK	Azure Portal, Azure CLI, Azure SDK	Google Cloud Console, Cloud SDK, API

Самообслуговування є важливою характеристикою хмарних сервісів, вона дозволяє користувачам самостійно керувати та налаштовувати ресурси хмарних сервісів. Кожен з сервісів - AWS, Azure та GCP надає користувачам можливість автоматичної обробки даних.

У AWS, користувачі можуть скористатися AWS Management Console - веб-інтерфейсом для керування та налаштування ресурсів. Крім того, доступні AWS CLI (Command Line Interface) та AWS SDK (Software Development Kit) для автоматизації операцій через командний рядок або програмні скрипти.

GCP має Google Cloud Console - веб-інтерфейс, який дозволяє користувачам керувати та налаштовувати ресурси. Додатково, доступні Google Cloud SDK та API, які надають можливість автоматизувати операції та керувати інфраструктурою з допомогою командного рядка або програмного інтерфейсу [9].

Завдяки цим інструментам, користувачі можуть самостійно керувати та налаштовувати ресурси у хмарній інфраструктурі в зручний спосіб, використовуючи веб-інтерфейс або API, що спрощує процес розгортання та управління сервісами у хмарі.

Гнучкість (Flexibility): Хмарні сервіси можуть легко змінюватись або адаптуватись до потреб користувачів. Вони можуть включати різні режими обробки даних, рівні безпеки та налаштування ресурсів.

Забезпечення безпеки (Security): Хмарні сервіси надають механізми для захисту даних, такі як шифрування, автентифікація та контроль доступу. Вони також можуть забезпечувати резервне копіювання та відновлення даних для запобігання втраті інформації [2].

У таблиці 4 наведено характеристики режимів обробки даних у хмарних сервісах.

Гнучкість є важливою характеристикою хмарних сервісів, вона дозволяє змінювати та адаптувати ресурси відповідно до змінних потреб користувачів. Кожна з систем - AWS, Azure та GCP, надає різні можливості для гнучкості ресурсів [4].

Таблиця 4 - Характеристики режимів обробки даних у хмарних сервісах.

Характеристика	AWS	Azure	GCP
Режими обробки даних	Можливість використання різних типів інстансів та служб для оптимізації обробки даних	Різноманітність послуг для обробки даних, включаючи Azure Data Lake, Azure Databricks, Azure HDInsight та інші	Широкий набір інструментів для обробки даних, включаючи BigQuery, Dataflow, Dataproc та інші.
Рівні безпеки	Механізми безпеки, такі як AWS Identity and Access Management (IAM), вбудовані контролю доступу та шифрування даних	Широкий набір інструментів для управління безпекою, включаючи Azure Active Directory, Azure Security Center, Azure Key Vault та інші	Інструменти безпеки, такі як Identity and Access Management (IAM), Resource Manager, Cloud Key Management Service (KMS) та інші
Налаштування ресурсів	Гнучкість масштабування ресурсів, включаючи зміну розміру інстансів, горизонтальне та автоматичне масштабування	Різноманітність типів ресурсів та можливість гнучко налаштувати їх характеристики, включаючи розмір, кількість та налаштування.	Широкий набір ресурсів, що можуть бути гнучко налаштовані, включаючи розмір інстансів, кількість ресурсів та налаштування мережі

У AWS, користувачі мають можливість використовувати різні типи служб для оптимізації обробки даних. Крім того, AWS забезпечує гнучкість масштабування ресурсів, дозволяючи змінювати розмір інстансів, виконувати горизонтальне та автоматичне масштабування в залежності від потреб.

Azure надає широкий набір послуг для обробки даних, що дозволяють користувачам вибрати найбільш оптимальні рішення для своїх потреб. Також, в Azure можна гнучко налаштовувати ресурси, включаючи їх розмір, кількість та налаштування, що дозволяє пристосовувати їх до конкретних вимог.

GCP має широкий набір інструментів для обробки даних, таких як BigQuery, Dataflow та інші, які забезпечують гнучкість при роботі з даними. Крім того, ресурси в GCP можуть бути гнучко налаштовані, включаючи розмір інстансів, кількість ресурсів та налаштування мережі.

Завдяки гнучкості цих систем, користувачі можуть легко змінювати та адаптувати ресурси хмарної інфраструктури, що дозволяє їм ефективно використовувати ресурси відповідно до своїх потреб [8].

Представлено основні характеристики, які роблять їх унікальними, а в описі наведено короткий огляд цих особливостей. Дані в таблиці допомагають порівняти різні аспекти кожного сервісу, а опис доповнює таблицю, пояснюючи, чому ці особливості є важливими, та як вони можуть бути вигідними для користувачів [13].

У таблиці 5 наведено підсумкові переваги хмарних сервісів AWS, Azure, GCP.

Таблиця 5 - Підсумкові переваги хмарних сервісів AWS, Azure, GCP

Перевага	AWS	Azure	GCP
Масштаб	Широкий вибір сервісів та регіонів	Глобальна поширеність та гнучкість масштабування	Широка географічна розподіленість сервісів
Широкий функціонал	Широкий набір сервісів та інструментів	Широкий вибір послуг та інтеграція з Microsoft	Широкий набір інструментів та інтеграція з Google
Модульність	Структура на основі служб та сервісів	Модульна архітектура та різні рівні послуг	Модульна структура та гнучкість конфігурації
Гнучкість	Гнучкість масштабування та налаштування	Гнучкість вибору та налаштування ресурсів	Гнучкість налаштування та розширення сервісів
Широкі можливості інтеграції	Широкий вибір партнерів та інтеграція з іншими сервісами	Інтеграція з Microsoft та партнерськими рішеннями	Інтеграція з екосистемою Google та партнерськими рішеннями

Функціонал AWS пропонує широкий вибір сервісів та регіонів, Azure має глобальну присутність та гнучкість масштабування, а GCP відрізняється широкою географічною розподіленістю сервісів.

Широкий функціонал: AWS, Azure та GCP надають набір сервісів та інструментів. AWS має багатофункціональний набір, Azure відрізняється інтеграцією з Microsoft, а GCP пропонує інтеграцію з екосистемою Google [11].

Модульність у AWS та GCP базується на структурі сервісів, а Azure має модульну архітектуру та різні типи глибини досліджень.

Гнучкість у AWS, Azure та GCP надає масштабування та налаштуваність ресурсів. Користувачі можуть вибирати та налаштувати ресурси відповідно до своїх потреб.

Широкі можливості інтеграції: AWS, Azure та GCP дозволяють інтегруватися з іншими сервісами та партнерськими рішеннями. Кожна платформа має свої власні партнерські програми та екосистему [14].

Серед наведених технологій, Google Cloud і Microsoft Azure є вибором для високоточної обробки геоданих.

Google Cloud надає розширені сервіси для обробки геоданих, такі як Google BigQuery для аналізу великого обсягу даних, Google Cloud Machine Learning Engine для моделювання геологічних явищ, а також інструменти візуалізації та інтерактивного аналізу даних. Багато компаній та організацій вже використовують Google Cloud для рішення завдань геоінформатики та геоаналітики [9].

Microsoft Azure також надає широкий спектр інструментів для аналізу та обробки геоданих. Azure також пропонує інтеграцію з іншими продуктами Microsoft, що сприяє зручному та потужному аналізу геоданих [7].

Обидва ці хмарні сервіси мають ресурси для роботи з геоданими, включаючи розгортання високопродуктивних обчислювальних ресурсів та велику кількість доступних сервісів. Остаточний вибір між ними буде залежати від конкретних вимог проекту.

На ринку існує багато сервісів для автоматичної обробки даних, і вибір найкращого залежить від конкретних потреб та вимог вашого проекту [6].

AWS Glue є повністю керованим сервісом для автоматичної обробки, аналізу та переміщення даних. Для використання AWS Glue підключено акаунт AWS та доступ до консолі керування AWS. Наведено кілька кроків та практичних рекомендацій для автоматичної обробки AWS Glue:

Крок 1 створення бази даних та таблиці

```
import boto3
# Ініціалізуємо клієнта AWS Glue
glue_client = boto3.client('glue', region_name='your_region')
# Створення бази даних
glue_client.create_database(DatabaseInput={'Name': 'your_database_name'})
# Створення таблиці
glue_client.create_table(
    DatabaseName='your_database_name',
    TableInput={
        'Name': 'your_table_name',
        'StorageDescriptor': {
            'Location': 's3://your_bucket/your_table_data',
            'InputFormat': 'org.apache.hadoop.mapred.TextInputFormat',
            'OutputFormat': 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat',
            'SerdeInfo': {
                'SerializationLibrary': 'org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe',
                'Parameters': {
                    'serialization.format': '1' } },
            'Columns': [
                { 'Name': 'column1', 'Type': 'string' },
                { 'Name': 'column2', 'Type': 'int' }
                # Додайте інші стовпці за необхідністю },
            'TableType': 'EXTERNAL_TABLE' })
```

Крок 2: Визначення ETL-скрипту

Встановлення залежностей для ETL-скрипту

```
import sys
!{sys.executable} -m pip install --upgrade awsglue
from awsglue.transforms import *
from awsglue.utils import getResolvedOptions
from pyspark.context import SparkContext
from awsglue.context import GlueContext
from awsglue.job import Job
```

```

# Ініціалізуємо GlueContext та SparkContext
sc = SparkContext()
glueContext = GlueContext(sc)
spark = glueContext.spark_session
job = Job(glueContext)
# Визначення ETL-скрипту
datasource = glueContext.create_dynamic_frame.from_catalog(database = "your_database_name",
table_name = "your_table_name")
apply_mapping = ApplyMapping.apply(frame = datasource, mappings = [("column1", "string",
"new_column1", "string"), ("column2", "int", "new_column2", "int")])
datasink = glueContext.write_dynamic_frame.from_catalog(frame = apply_mapping, database =
"your_database_name", table_name = "new_table_name", transformation_ctx = "datasink")
job.commit()
# Ініціалізуємо клієнта AWS Glue
glue_client = boto3.client('glue', region_name='your_region')
# Запуск задачі ETL
response = glue_client.start_job_run(
    JobName='your_job_name',
    Arguments={
        '--TempDir': 's3://your_bucket/temp',
        '--extra-py-files': 's3://your_bucket/your_dependencies.zip' })
job_run_id = response['JobRunId']

```

Ці фрагменти коду демонструють процес використання Amazon Web Services Glue для автоматичної обробки даних.

Висновки

Хмарні сервіси є потужним інструментом для обробки геоінформаційних даних. Вони надають гнучкість, автоматизованість та високу доступність, що дозволяє ефективно обробляти великі обсяги даних та вирішувати складні географічні задачі без людського ресурсу.

Провідні хмарні сервіси, такі як Google Cloud, Amazon Web Services та Microsoft Azure, надають багатий набір інструментів та сервісів для розробки, зберігання та обробки геоінформаційних даних. Вони підтримують різні алгоритми, включаючи аналіз зображень та обробку супутникових даних.

Досліджено та наведено порівняльний аналіз трьох провідних хмарних сервісів для обробки геоінформаційних даних: Amazon Web Services (AWS), Microsoft Azure та Google Cloud Platform (GCP).

Виявлено унікальні особливості кожної платформи та показано що сервіс AWS Glue є найбільш автоматизованим для користувачів.

За допомогою таблиць і описів ми ретельно розглянули характеристики кожної платформи, зосередившись на таких аспектах, як віртуалізація, базові програмні рішення, самообслуговування, масштабованість та гнучкість. Зроблені порівняння надали нам глибше розуміння особливостей кожного сервісу та дозволили визначити їх переваги.

AWS вирізняється своїм широким набором віртуальних машин, сервісами контейнеризації та автоматичним обчисленням. Azure надає потужні інструменти для розгортання додатків, інтеграції з геоданими та масштабованості. GCP пропонує ефективну обробку геоданих та їх швидку передачу.

Загальним висновком нашого дослідження є те, що кожна з розглянутих хмарних платформ має свої унікальні переваги та можливості. Вибір конкретної платформи залежить від конкретних потреб та вимог користувача. AWS, Azure та GCP можуть бути ефективними інструментами для геоінформаційної обробки даних, проте вони мають різні особливості, які варто врахувати при виборі найбільш підходящої платформи для конкретного проекту.

Таким чином у статті надано огляд найбільш відомих хмарних сервісів та унікальних особливостей кожного з них у контексті обробки геоінформаційних даних. Це дозволяє зробити обґрунтований вибір та використовувати найбільш відповідний сервіс для потреб геоінформаційної обробки даних. В результаті обрано автоматизовану платформу – AWS Glue, оскільки вона дозволяє отримувати такий самий результат, як і інші сервіси, зберігаючи при цьому людський час. Це зменшує фінансові витрати на аналіз даних.

Список літератури

- [1] Carillo, K. D., & White, D. (2020). *Cloud Computing: Concepts, Technology, and Architecture*. Morgan Kaufmann.
- [2] Kuo, M. F. (2020). *Geographic Information Science: Fundamentals and Applications*. CRC Press.
- [3] Chen, M., & Zhang, Y. (2020). *Cloud Computing for Geospatial Data Analysis and Applications*. Springer.

- [4] Kolodziej, J., & González-Vélez, H. (Eds.). (2020). *Cloud Computing for Science and Engineering*. Wiley.
- [5] Li, X., & Karimi, H. A. (2020). *Geospatial Data Science Techniques and Applications*. Springer.
- [6] Fan, Y., et al. (2021). *Geospatial Machine Learning for Urban Research and Planning*. Springer.
- [7] Ahmed, N., & Sheng, Q. Z. (2021). *Security and Privacy in Cloud Computing for Geospatial Data Services*. Springer.
- [8] Deng, Z., et al. (2020). *Cloud Computing for Geoengineering: Concepts, Applications, and Perspectives*. Springer.
- [9] Laplante, P. A. (2020). *Requirements Engineering for Software and Systems*. CRC Press.
- [10] Vögler, M., & Gartner, G. (2021). *User-Centric Requirements Engineering in the Cloud*. Springer.
- [11] Li, M., & Zhang, Z. (2020). *Cloud Testing: Fundamentals, Techniques, and Tools*. Springer.
- [12] Balalaie, A., et al. (2020). *Microservices Architecture: Aligning Principles, Practices, and Culture*. O'Reilly Media.
- [13] Frappier, M., & Bahrami, M. (2020). *Advances in Geo-Information Science*. Springer.
- [14] Liu, C., & Zhou, M. (2020). *Intelligent Big Geospatial Data Analytics: Deep Learning and Neural Networks*. CRC Press.

Стаття надійшла: 29.04.2023.

References

- [1] Carillo, K. D., & White, D. (2020). *Cloud Computing: Concepts, Technology, and Architecture*. Morgan Kaufmann.
- [2] Kuo, M. F. (2020). *Geographic Information Science: Fundamentals and Applications*. CRC Press.
- [3] Chen, M., & Zhang, Y. (2020). *Cloud Computing for Geospatial Data Analysis and Applications*. Springer.
- [4] Kolodziej, J., & González-Vélez, H. (Eds.). (2020). *Cloud Computing for Science and Engineering*. Wiley.
- [5] Li, X., & Karimi, H. A. (2020). *Geospatial Data Science Techniques and Applications*. Springer.
- [6] Fan, Y., et al. (2021). *Geospatial Machine Learning for Urban Research and Planning*. Springer.
- [7] Ahmed, N., & Sheng, Q. Z. (2021). *Security and Privacy in Cloud Computing for Geospatial Data Services*. Springer.
- [8] Deng, Z., et al. (2020). *Cloud Computing for Geoengineering: Concepts, Applications, and Perspectives*. Springer.
- [9] Laplante, P. A. (2020). *Requirements Engineering for Software and Systems*. CRC Press.
- [10] Vögler, M., & Gartner, G. (2021). *User-Centric Requirements Engineering in the Cloud*. Springer.
- [11] Li, M., & Zhang, Z. (2020). *Cloud Testing: Fundamentals, Techniques, and Tools*. Springer.
- [12] Balalaie, A., et al. (2020). *Microservices Architecture: Aligning Principles, Practices, and Culture*. O'Reilly Media.
- [13] Frappier, M., & Bahrami, M. (2020). *Advances in Geo-Information Science*. Springer.
- [14] Liu, C., & Zhou, M. (2020). *Intelligent Big Geospatial Data Analytics: Deep Learning and Neural Networks*. CRC Press.

Відомості про авторів

Недоснований Олександр Юрійович – аспірант кафедри обчислювальної техніки ВНТУ

Черняк Олександр Іванович – кандидат наук, доцент кафедри обчислювальної техніки ВНТУ

Голінко Віталій Володимирович – аспірант Київського національного університету технологій та дизайну (КНУТД), Київ.

O.Y. Nedosnovanyi¹, O.I. Cherniak¹, V.V. Golinko²

COMPARATIVE ANALYSIS OF CLOUD SERVICES FOR GEOINFORMATION DATA PROCESSING

¹Vinnitsia National Technical University, Vinnitsia

²Kyiv National University of Technologies and Design