

ВИКОРИСТАННЯ ХМАРНИХ ОБЧИСЛЕНЬ ДЛЯ АНАЛІЗУ ЯКОСТІ ПОВІТРЯ

П. О. Гридін^{1,а}, Г. О. Яйлимова^{2,б}

^{1, 2} Навчально-науковий Фізико-технічний інститут

Анотація

Ця публікація містить огляд переваг використання хмарних обчислень для аналізу якості повітря. Тут пояснюється, як хмарні обчислення можуть зменшити вартість аналізу даних, підвищити точність результатів аналізу даних і зробити аналіз даних швидше. Вона також охоплює різні типи хмарних служб, доступних для аналізу даних.

Ключові слова: аналіз якості повітря, аналіз даних, хмарні обчислення, AWS, GCP, Azure

Вступ

Забруднення повітря є серйозною екологічною проблемою в усьому світі, яка має значні наслідки для здоров'я та економіки. За даними Всесвітньої організації охорони здоров'я, забруднене повітря є причиною передчасної смерті 7 мільйонів людей щороку. Для боротьби з цією проблемою уряди та організації впровадили різні заходи, зокрема моніторинг якості повітря. В останні роки хмарні постачальники відіграють дедалі важливішу роль в аналізі якості повітря, надаючи послуги зберігання, обробки та аналізу даних. Дослідимо, як хмарні постачальники допомагають аналізувати якість повітря, переваги та недоліки цього підходу.

Хмарні обчислення надають дослідникам якості повітря широкий спектр переваг. Використовуючи хмарні обчислення, дослідники даних можуть отримати доступ до обчислювальних ресурсів на вимогу з мінімальними початковими витратами та без потреби у внутрішній IT-інфраструктурі. Хмарні обчислення дозволяють дослідникам даних швидко та легко масштабувати свої потреби в обробці та зберіганні даних вгору або вниз залежно від вимог свого проекту. Крім того, хмарні обчислення надають дослідникам даних доступ до широкого спектру передових аналітичних інструментів, таких як машинне навчання та штучний інтелект, які можуть допомогти їм виявляти ідеї та створювати потужні моделі. Хмарні обчислення також пропонують дослідникам підвищену доступність і безпеку їхніх даних, оскільки дані зберігаються на хмарних серверах і доступ до них здійснюється практично з будь-якої точки світу. Використання хмарних обчислень дослідниками даних може значно скоротити час і витрати, одночасно забезпечуючи необхідні ресурси для ефективного проведення досліджень.

1. Хмарні сервіси для аналізу якості повітря

Хмарні постачальники пропонують ряд послуг, які можна використовувати для аналізу даних про якість повітря. Ці послуги включають зберігання даних, обробку даних, машинне навчання та інструменти візуалізації даних.

- **Зберігання даних.** Хмарні постачальники пропонують послуги зберігання даних, які можна використовувати для зберігання великої кількості даних про якість повітря, зібраних із різних джерел. Ці дані можуть включати дані зі стаціонарних станцій моніторингу, мобільних датчиків і супутникових зображень. Хмарне сховище надає масштабоване та безпечне рішення, яке може вмістити великі обсяги даних і зробити їх доступними для дослідників, політиків та громадськості.
- **Обробка даних.** Хмарні постачальники пропонують послуги обробки даних, які можна використовувати для попередньої обробки даних про якість повітря, очищення та перевірки даних, а також аналізу даних для отримання значущої інформації. Ці послуги включають сховища даних, конвеєри даних і інфраструктуру обробки великих даних, такі як Hadoop і Spark. Ці інструменти можуть обробляти великі обсяги даних і виконувати складні обчислення швидко й ефективно.
- **Машинне навчання.** Хмарні постачальники пропонують послуги машинного навчання, які можна використовувати для розробки прогнозних моделей якості повітря. Ці моделі можна використовувати для прогнозування рівня якості повітря та визначення потенційних джерел забруднення. Алгоритми машинного навчання також можна використовувати для виявлення закономірностей у даних про якість повітря та надання

^аipavlo47@mail.com

^бyailymova.hanna@iit.kpi.ua

інформації, яка може обґрунтувати політичні рішення.

- Візуалізація даних. Хмарні постачальники пропонують інструменти візуалізації даних, які можна використовувати для представлення даних про якість повітря в зручному форматі. Ці інструменти можна використовувати для створення інтерактивних карт, діаграм і графіків, які можуть допомогти політикам і громадськості зрозуміти дані про якість повітря та приймати обґрунтовані рішення.

1.1. Переваги хмарних провайдерів для аналізу якості повітря

Використання хмарних провайдерів для аналізу даних про якість повітря має кілька переваг:

1. Масштабованість. Хмарні постачальники пропонують масштабовані рішення, які можуть обслуговувати великі обсяги даних і обробляти складні обчислення. Це дає змогу аналізувати дані про якість повітря з багатьох джерел і з різною просторовою та часовою роздільною здатністю.
2. Економічна ефективність. Хмарні постачальники пропонують економічно ефективні рішення для зберігання, обробки та аналізу даних про якість повітря. Це пов'язано з тим, що хмарні послуги зазвичай платять за використання, тобто користувачі платять лише за ті послуги, якими вони користуються.
3. Доступність. Хмарні постачальники пропонують доступні рішення для аналізу даних про якість повітря. Це означає, що дані можуть бути доступні для дослідників, політиків і громадськості через веб-програми, API та інші інтерфейси.
4. Гнучкість. Хмарні постачальники пропонують гнучкі рішення для аналізу даних про якість повітря. Це означає, що користувачі можуть вибрати інструменти та послуги, які найкраще відповідають їхнім потребам, і можуть збільшувати чи зменшувати масштаб за потреби.

1.2. Недоліки хмарних провайдерів для аналізу якості повітря

Є також деякі недоліки використання хмарних провайдерів для аналізу даних про якість повітря:

1. Конфіденційність і безпека. Хмарні постачальники можуть зберігати конфіденційні дані про якість повітря, що може становити загрозу конфіденційності та безпеці. Користувачі повинні переконатися, що для захисту даних застосовано відповідні заходи безпеки.
2. Залежність від постачальників послуг. Користувачі хмарних служб можуть стати залежними від постачальника послуг щодо зберігання, обробки та аналізу. Це може призвести до блокування постачальника, коли користувачі не зможуть перейти до альтернативного постачальника послуг без значних витрат і зусиль.

3. Затримка. Хмарні служби можуть мати затримку, що може вплинути на програми моніторингу якості повітря в реальному часі. Це означає, що деякі програми можуть не підходити для хмарних рішень.

2. Огляд хмарних сервісів

Хмарні обчислення можуть зменшити вартість аналізу даних, підвищити точність результатів аналізу даних і зробити аналіз даних швидшим, надаючи доступ до величезних обсягів обчислювальної потужності та сховища та дозволяючи підприємствам платити лише за те, що їм потрібно [1, 2].

Одним із прикладів хмарної служби, яка може зменшити вартість аналізу даних, є Amazon Web Services (AWS). AWS пропонує ряд послуг, таких як Amazon Elastic Compute Cloud (EC2) і Amazon Simple Storage Service (S3), які дозволяють компаніям швидко та безпечно зберігати, обробляти та аналізувати великі обсяги даних. За допомогою EC2 компанії можуть розкручувати обчислювальні ресурси за запитом і платити лише за ресурси, які вони використовують, тоді як S3 забезпечує безпечний і надійний спосіб зберігання великих обсягів даних.

Іншим прикладом хмарної служби, яка може підвищити точність результатів аналізу даних, є Google Cloud Platform (GCP). GCP пропонує набір інструментів, таких як BigQuery та Cloud Machine Learning Engine, які дозволяють компаніям швидко та легко аналізувати великі набори даних і отримувати з них висновки. BigQuery надає інтерактивне середовище аналізу, яке дозволяє компаніям досліджувати набори даних, а Cloud Machine Learning Engine надає потужний набір алгоритмів машинного навчання, які можуть допомогти компаніям робити точніші прогнози на основі їхніх даних.

Нарешті, прикладом хмарної служби, яка може пришвидшити аналіз даних, є Microsoft Azure. Azure пропонує ряд послуг, таких як Azure Databricks і Azure Stream Analytics, які дозволяють компаніям швидко й легко обробляти великі обсяги потокових даних. Azure Databricks допомагає компаніям швидко створювати та розгортати конвеєри даних, а Azure Stream Analytics дозволяє компаніям аналізувати поточкові дані в режимі реального часу.

Таким чином, хмарні обчислення можуть зменшити вартість аналізу даних, підвищити точність результатів аналізу даних і зробити аналіз даних швидшим, надаючи доступ до величезних обсягів обчислювальної потужності та сховища та дозволяючи підприємствам платити лише за те, що їм потрібно. Прикладами хмарних служб, які можна використовувати для досягнення цих цілей, є Amazon Web Services, Google Cloud Platform і Microsoft Azure.

2.1. Amazon Web Services (AWS)

AWS [3] надає різноманітні сервіси для аналізу даних. Ці служби можна використовувати для широкого спектру цілей, включаючи керування та аналіз

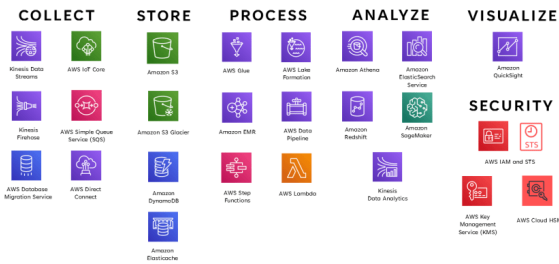


Рис. 1. Сервіси AWS для аналізу даних

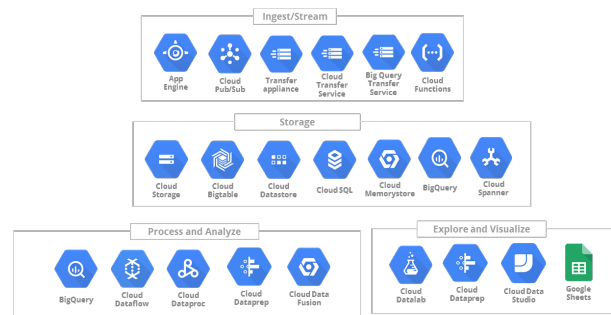


Рис. 2. Сервіси GCP для аналізу даних

великих наборів даних, створення моделей і виконання складних аналітичних завдань.

Наприклад, Amazon Relational Database Service (RDS) – це керована служба бази даних, яку можна використовувати для зберігання та запитів до великих наборів даних. За допомогою RDS користувачі можуть легко налаштувати та керувати базами даних для своїх завдань аналізу даних. Крім того, Amazon Redshift – це сховище даних, яке можна використовувати для зберігання та аналізу великих наборів даних. Його можна використовувати для створення моделей даних і виконання аналітичних завдань на великих наборах даних.

AWS також надає такі послуги, як Amazon SageMaker, який є керованим сервісом машинного навчання. Його можна використовувати для створення, навчання та розгортання моделей машинного навчання. Крім того, Amazon Athena – це безсерверний сервіс запитів, який можна використовувати для запитів до великих наборів даних, що зберігаються в S3, RDS та інших сховищах даних.

AWS також надає інструменти візуалізації даних, такі як Amazon QuickSight, які можна використовувати для візуалізації даних, що зберігаються в S3, RDS та інших сховищах даних. Amazon QuickSight можна використовувати для створення різноманітних візуалізацій для аналізу великих наборів даних.

Крім того, AWS надає такі послуги, як Amazon Kinesis, які можна використовувати для обробки поточкових даних і виконання аналітичних завдань. Amazon Kinesis можна використовувати для керування та аналізу великих наборів даних у режимі реального часу.

Нарешті, Amazon EMR можна використовувати для налаштування керованого кластера Hadoop для завдань аналізу даних. За допомогою Amazon EMR користувачі можуть легко налаштувати та керувати кластерами Hadoop для завдань аналізу даних.

Тому AWS надає різноманітні сервіси для аналізу даних. Ці служби можна використовувати для керування та аналізу великих наборів даних, створення моделей і виконання складних аналітичних завдань.

2.2. Google Cloud Platform (GCP)

Google Cloud Platform (GCP) [4] – це служба хмарних обчислень, яка дозволяє компаніям зберігати, обробляти та аналізувати дані. GCP надає набір служб, спеціально розроблених для аналізу даних,

зокрема BigQuery, Cloud Dataflow, Cloud Dataproc і Cloud ML Engine. Ці служби можна використовувати для аналізу великих наборів даних, розробки моделей машинного навчання та створення конвеєрів даних.

BigQuery – це безсерверне кероване сховище даних, яке дозволяє користувачам надсилати запити до великих наборів даних за лічені секунди. BigQuery добре підходить для аналізу даних і бізнес-аналітики, його можна використовувати для аналізу даних кліків, даних про дії користувачів та інших великих наборів даних. BigQuery також інтегровано з Google Analytics, що дозволяє користувачам легко запитувати та аналізувати дані свого веб-сайту.

Cloud Dataflow – це повністю керований сервіс для потокової та пакетної обробки даних. Він дозволяє користувачам розробляти та виконувати конвеєри даних у безсерверному середовищі, а також забезпечує вбудовану підтримку обробки даних і алгоритмів машинного навчання. Cloud Dataflow можна використовувати для обробки поточкових даних у режимі реального часу та для створення конвеєрів даних для перетворення та аналізу даних.

Cloud Dataproc – це повністю керована служба для виконання завдань Apache Hadoop і Apache Spark на GCP. Він добре підходить для обробки даних, ETL і задач машинного навчання, а також може використовуватися для створення конвеєрів даних і обробки великих наборів даних. Завдання Cloud Dataproc можна легко планувати та керувати ними, що дозволяє користувачам легко масштабувати свої завдання аналізу даних.

Cloud ML Engine – це керований сервіс для розробки, навчання та розгортання моделей машинного навчання на GCP. Він надає доступ до ряду інструментів і послуг для створення, навчання та розгортання моделей машинного навчання, а також може використовуватися для розробки прогнозних моделей, моделей розпізнавання зображень та інших програм машинного навчання.

Таким чином, GCP надає набір послуг, спеціально розроблених для аналізу даних і машинного навчання, включаючи BigQuery, Cloud Dataflow, Cloud Dataproc і Cloud ML Engine. Ці служби можна використовувати для аналізу великих наборів даних, розробки моделей машинного навчання та створення конвеєрів даних.

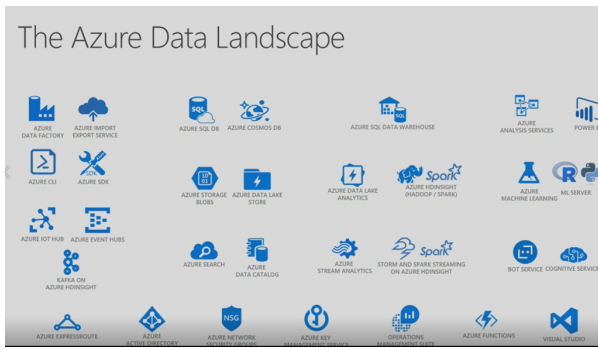


Рис. 3. Сервіси GCP для аналізу даних

2.3. Azure

Azure [5] надає комплексну хмарну платформу аналітики даних, яка дозволяє організаціям отримувати доступ, аналізувати та ділитися статистикою даних у своєму бізнесі. Ця платформа дозволяє аналізувати великі набори даних у реальному часі та розробляти прогностичні моделі та алгоритми машинного навчання.

Azure надає ряд інструментів і служб для аналізу даних, як-от Azure Data Lake, Azure Machine Learning і Azure HDInsight. Ці сервіси дозволяють організаціям зберігати, обробляти та аналізувати великі набори даних у хмарі, щоб отримувати розуміння та приймати обґрунтовані рішення.

Наприклад, Azure Data Lake дозволяє організаціям зберігати й аналізувати великі набори даних у хмарі. Це повністю керована хмарна платформа даних, яка дозволяє організаціям зберігати й обробляти будь-які типи даних, у тому числі структуровані та неструктуровані дані. Крім того, Azure Data Lake надає інтегрований набір інструментів і служб для обробки даних, аналітики та машинного навчання, що робить його ідеальною платформою для аналізу великих наборів даних.

Машинне навчання Azure – це ще одна платформа, яка дозволяє організаціям розробляти, розгорнути та керувати прогностичними моделями та алгоритмами машинного навчання в хмарі. Він надає інтегрований набір інструментів і послуг для науки про дані та машинного навчання, що полегшує створення, розгортання та керування прогностичними моделями.

Нарешті, Azure HDInsight – це повністю керований хмарний сервіс для обробки й аналізу великих наборів даних. Він надає інтегрований набір інструментів і служб для обробки даних, аналітики та машинного навчання, що робить його потужною платформою для аналізу великих наборів даних.

Підсумовуючи, Azure надає комплексну хмарну платформу аналітики даних, яка дозволяє організаціям отримувати доступ, аналізувати та ділитися статистикою даних у своєму бізнесі. Він надає інтегрований набір інструментів і служб для обробки даних, аналітики та машинного навчання, що робить його ідеальною платформою для аналізу великих

наборів даних.

Висновки

Підсумовуючи, хмарні технології пропонують численні переваги для аналізу даних про якість повітря. Масштабованість, економічність, доступність і гнучкість хмарних сервісів роблять їх ідеальним рішенням для зберігання, обробки та аналізу великих обсягів даних про якість повітря. Використовуючи хмарні сервіси, дослідники, політики та громадськість можуть отримати уявлення про тенденції якості повітря, визначити джерела забруднення та прийняти обґрунтовані рішення щодо зменшення забруднення повітря.

Крім того, хмарні постачальники пропонують ряд інструментів і послуг, які можна використовувати для попередньої обробки, очищення та перевірки даних про якість повітря. Алгоритми машинного навчання можна використовувати для розробки прогностичних моделей, які можуть прогнозувати рівень якості повітря та визначати потенційні джерела забруднення. Інструменти візуалізації даних також можна використовувати для представлення даних про якість повітря в зручному для користувача форматі, легко зрозумілому політикам і громадськості.

Хоча є певні недоліки використання хмарних служб для аналізу даних про якість повітря, як-от питання конфіденційності та безпеки, залежність від постачальників послуг і проблеми із затримкою, їх можна пом'якшити, впровадивши відповідні заходи безпеки, уникнувши блокування постачальника та вибравши відповідні хмарні сервіси для конкретних програм.

Загалом, хмарні технології мають потенціал революціонізувати спосіб моніторингу та аналізу даних про якість повітря. Оскільки забруднення повітря продовжує залишатися серйозною проблемою для навколишнього середовища та охорони здоров'я, використання хмарних сервісів може надати цінну інформацію про причини та наслідки забруднення повітря та обґрунтувати політичні рішення, спрямовані на зменшення його впливу на здоров'я людини та навколишнє середовище.

Перелік використаних джерел

1. Erl T. Cloud Computing: Concepts, Technology & Architecture. — Pearson, 02.05.2013. — 528 с. — ISBN B00CM9V7Q8.
2. Erl T. Cloud Computing Design Patterns. — Pearson, 23.05.2015. — 590 с. — ISBN B00YF0ORCS.
3. Amazon Web Services. — URL: <https://aws.amazon.com/>.
4. Google Cloud Platform. — URL: <https://cloud.google.com/>.
5. Azure. — URL: <https://azure.microsoft.com/>.