

УДК 004.622

Т. Патральський

(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

ЗБЕРІГАННЯ ТА ТРАНСФОРМАЦІЯ ДАНИХ У ХМАРНОМУ СЕРЕДОВИЩІ GOOGLE CLOUD BIGQUERY

UDC 004.622

T. Patralskyi

DATA STORAGE AND TRANSFORMATION IN THE CLOUD ENVIRONMENT GOOGLE CLOUD BIGQUERY

BigQuery – це безсерверне мультихмарне сховище даних, яке спрощує процес роботи з усіма типами даних, тож можна зосередитися на швидкому отриманні цінної бізнес-аналітики. BigQuery, який є ядром хмари даних Google, дозволяє спростити інтеграцію даних, економічно й безпечно масштабувати аналітику, ділитися багатими даними за допомогою вбудованої бізнес-аналітики, а також навчати та розгортати моделі ML за допомогою простого інтерфейсу SQL [1].

До особливостей BigQuery відноситься:

- керування даними – можна створювати та видаляти такі об'єкти, як таблиці, подання та визначені користувачем функції, імпортувати дані з Google Storage у таких форматах, як CSV, Parquet, Avro або JSON;
- запити, які виражаються стандартним діалектом SQL, а результати повертаються у форматі JSON із максимальною довжиною відповіді приблизно 128 МБ або необмеженим розміром, якщо ввімкнено великі результати запиту;
- інтеграція - BigQuery можна використовувати з Google Apps Script (наприклад, як зв'язаний сценарій у Google Docs) або з будь-якої мови, яка може працювати з REST API або клієнтськими бібліотеками.

BigQuery має багато інтегрованих платформ, найпопулярніші та часто застосовані це Google spreadsheet та REST API, де з останнього можна отримувати дані в реальному часі з різних платформ які дозволяють це робити.

Щоб використати API, потрібно використати одну з мов програмування, краще всього підійде python, за допомогою якого можна написати скрипт для витягування даних, обробки та агрегації даних з потрібної вам платформи та записати її у хмарне сховище Bigquery, для того щоб оновлювалися дані потрібно це скрипт у Google cloud function, щоб дані постійно були актуальними. Після цього потрібно записати робоче місце, де можна написати SQL запит для бази даних Bigquery та зробити потрібні операції для кращого виводу даних.

У доповіді розглянуто основні етапи трансформації відкритих даних та їх зберігання і оновлення, наведено приклади практичної реалізації Google Cloud Bigquery.

Література

1. Огляд Google Cloud BigQuery. G2. – 2022. URL: <https://www.g2.com/products/google-cloud-bigquery/reviews>.
2. Trino 403 Documentation. trino.io. URL: <https://trino.io/docs/current/connector/bigquery.html>.