

СЕРВІС ДЛЯ ОБРОБКИ ВЕЛИКИХ ДАНИХ НА ОСНОВІ БЕЗСЕРВЕРНИХ ОБЧИСЛЕНЬ

А.К. Яковенко¹, Д.Б. Єльчанінов²

¹ магістрант кафедри інформатики та інтелектуальної власності, НТУ «ХПІ», Харків, Україна

² доцент кафедри інформатики та інтелектуальної власності, канд. техн. наук, НТУ «ХПІ», Харків, Україна

anton.yakovenko@cs.khpi.edu.ua

Безсерверні обчислення завоювали популярність як серед хмарних стартапів, так і великих підприємств. Завдяки ним команди можуть зосередитися на швидшому виведенні ідей на ринок (замість управління інфраструктурою) і при цьому платити лише за ресурси, що використовуються. Це дозволяє підтримувати широкий спектр напрямів використання: від важливих для затримки завдань (обробка веб-запитів) до складних та довготривалих (обробка великих даних) [1]. Саме для вирішення останніх задач є актуальним відповідний сервіс на основі безсерверних обчислень.

Завданням сервісу є обробка великих даних, котрі необхідні користувачеві. Метою роботи є проектування та реалізація такого сервісу.

Основною сервісу є віртуальний сервер під назвою AWS EC2, що знаходиться у хмарному провайдері Amazon Web Services. На цьому сервері розгорнуто нереляційну СУБД MongoDB, що дозволяє зберігати великий об'єм даних типу Sime-structured (напівструктуризована) або Unstructured (неструктуризована). Іншим основним компонентом є безсерверна технологія AWS Lambda, що обробляє великі дані та використовує обчислювальні ресурси лише у момент обробки даних. Використання такого підходу дозволяє значно економити кошти та платити лише за використані фізичні ресурси. Результат обробки та аналізу даних може надсилатися у файли різних форматів (таблиці, текстові документи) та користувачеві на електронну пошту. Наприклад, СУБД MongoDB може зберігати базу даних з великою кількістю документів, що являють собою інформацію про готельні номери онлайн-сервісу Airbnb у вигляді key-value. До цієї інформації входить детальний опис номеру та список усіх клієнтів цього номеру з їхніми відгуками. Безсерверна технологія AWS Lambda за допомогою мови програмування Python дозволяє отримувати, обробляти та аналізувати дані з MongoDB. Приміром, користувачеві сервісу потрібно знайти готельні номери, що відповідають двом критеріям: місто – Нью-Йорк, мінімальна кількість ночей доступна для проживання – 4. На основі отриманих даних користувач може переглянути посилання на доступні номери готелів та вибрати той, що підходить йому найбільше.

Отже, використовуючи даний підхід, можна обробляти та аналізувати велику кількість даних для отримання потрібного результату. У розглянутому прикладі можна визначити оптимальний варіант оренди готельного номеру, задаючи потрібні для користувача критерії. Цей підхід також можна використовувати для аналізу ринку оренди готельних номерів. Функціонал сервісу можна розширювати, додаючи якомога більше критеріїв пошуку, таких як аналіз відгуків відвідувачів номерів на певні слова та словосполучення.

Список літератури:

1. The State of Servless [Електрон. ресурс]. – Режим доступу: <https://www.datadoghq.com/state-of-serverless/> – Назва з титул. екрану.