

ОНЛАЙН-МОНІТОРИНГ АКТИВНОСТІ КОРИСТУВАЧІВ САЙТУ У РЕАЛЬНОМУ ЧАСІ

Маслов М.К., Добровольський О.О., Філімончук Т.В.

Харківський національний університет радіоелектроніки, Харків, Україна

Розподілені інформаційні системи – це комплекс апаратних та програмних засобів, які організують спільний доступ до обчислювальних ресурсів та даних за допомогою мережних протоколів. На даний час такі системи широко використовуються в сучасному світі в галузі окремих компаній, діяльність яких розподілена просторово. Впровадження інформаційних систем з розподіленою структурою використовується, коли необхідно забезпечити централізований контроль за змінами даних у віддалених підрозділах. Розподілені інформаційні системи з відповідним інструментарієм спрощують аналіз даних та створення звітів із узагальненої бази даних як для всієї компанії в цілому, так і для кожного її структурного підрозділу.

Метою доповіді є наведення структури розподіленої інформаційної системи для реалізації проекту, орієнтованого на збирання та аналіз активності користувачів сайту з подальшим відображенням її в реальному часі. Розгорання такого проекту вимагає продуманого технологічного стеку, який зможе швидко обробляти потоки даних, масштабуватися та візуалізувати результати без затримок для користувача.

Робота проекту організується наступним чином. Користувач заходить на сайт, де здійснює відповідні дії, які відстежуються JavaScript. Всі події, які були здійснено користувачем на сайті, за допомогою REST API приймаються та відправляються у розподілене сховище Kafka [1] для подальшого зберігання та аналізу. Після цього фреймворк Flink звертається до сховища (Kafka) зчитує дані, агрегує відповідні події (наприклад, IP або user ID, з тимчасовими вікнами) та результати (наприклад, "100 кліків на хвилину") направляє на зберігання до бази даних ClickHouse [2], яка орієнтована на аналітичну онлайн-обробку. Далі дані інтерпретуються за допомогою системи візуалізації Grafana [3] через підключення до ClickHouse та відображаються у вигляді графіків активності в реальному часі. Слід зазначити, що на одній панелі можна відстежувати дані з кількох джерел, що робить платформу Grafana універсальним інструментом для підприємництва, IT-розробки та статистичного аналізу.

Список літератури

1. Shapira G. Palino T., Sivaram R., Petty K. Kafka: The Definitive Guide: Real-Time Data and Stream Processing at Scale. Publisher: O'Reilly Media; 2nd edition. 2021. 485 p.
2. Anand R.V. Up and Running with ClickHouse. BPB Publications, 2021. 280 p.
3. Salituro E. Learn Grafana 10.x: A beginner's guide to practical data analytics, interactive dashboards and observability. Publisher: Packt Publishing; 2nd edition. 2023. 542 p.