СИСТЕМА ЗБЕРІГАННЯ ТА АНАЛІЗУ ДАНИХ

Климова І.М., Шмарін М.Д.
Харківський національний університет радіоелектроніки, Харків, Україна

У сучасних умовах стрімкого зростання обсягів цифрової інформації критичною задачею стає ефективне зберігання та обробка великих обсягів даних, що надходять з різноманітних джерел: сенсорних мереж, інформаційних систем підприємств, онлайн-сервісів тощо [1]. Здатність своєчасно накопичувати, структуровано зберігати й аналітично опрацьовувати ці дані визначає конкурентоспроможність організацій та якість прийнятих управлінських рішень. У зв'язку з цим побудова комп'ютерних систем, здатних забезпечити надійне зберігання й високопродуктивний аналіз даних, є важливою науково-технічною проблемою.

Метою доповіді є висвітлення архітектурних принципів побудови комп'ютерної системи, призначеної для тривалого зберігання та обробки великих обсягів даних [2]. Зокрема, досліджені методи організації розподіленого зберігання, балансування навантаження та застосування інструментів аналізу з використанням паралельних обчислень.

У доповіді наведено результати впровадження гібридного сховища даних на базі HDFS та обчислювального кластеру Spark, що використовується для обробки логів клієнтських запитів і фінансових транзакцій. Аналіз показав, що використання схеми розподілу даних із попередньою фільтрацією суттєво знижує затримку при виконанні аналітичних запитів і дозволяє досягти масштабування системи без істотного зниження продуктивності.

Важливим компонентом є впровадження механізмів резервного копіювання та індексації, що гарантують як відмовостійкість системи, так і прискорене виконання складних запитів. Зокрема, у випадку використання Elasticsearch для аналізу логів за часовими мітками забезпечено продуктивність до 100 тис. запитів/хв на одному вузлі.

Таким чином, сучасна комп'ютерна система для зберігання та аналізу даних виступає як інтегрована платформа, що поєднує засоби горизонтального масштабування, адаптивного кешування, обробки даних у реальному часі й можливості когнітивного аналізу.

Список літератури

1. Marz, N., Warren J. Big Data: Principles and best practices of scalable real-time data systems. Manning Publications, 2021. P. 328.

2. Guller M. Big Data Analytics with Spark: A Practitioner's Guide to Using Spark for Large Scale Data Analysis. Apress, 2020. P. 235.