

*Прінц М.В., студентка 2 курсу
спеціальності 122 «Комп'ютерні
науки»*

*Потапова Н. А., к.е.н., доцент,
доцент кафедри інформаційних
технологій*

СЕРВІСИ ДЛЯ АНАЛІЗУ ДАНИХ І НАБЛИЖЕНИХ ОБЧИСЛЕНЬ

Донецький національний університет імені Василя Стуса, м. Вінниця

Застосування сервісів для аналізу даних та наближених обчислень стає все більш актуальним для компаній та дослідницьких організацій у зв'язку з ростом обсягів даних та потребою в їх швидкому та ефективному аналізі. Сервіси для аналізу даних та наближених обчислень дозволяють отримувати доступ до високопродуктивних комп'ютерних ресурсів, що значно полегшує та прискорює обробку великого обсягу даних.

Одним із найпопулярніших сервісів для аналізу даних є Amazon Web Services (AWS). AWS надає великий набір інструментів для аналізу даних, зокрема Amazon EMR (Elastic MapReduce), Amazon Athena, Amazon Redshift, Amazon Kinesis та інші. Amazon EMR є повністю керованим сервісом, який дозволяє запускати Apache Hadoop та Apache Spark на віртуальних машинах, що забезпечує високу продуктивність та масштабованість. Amazon Athena дозволяє здійснювати аналіз даних без необхідності виконувати складні запити та налаштовувати інфраструктуру. Amazon Redshift надає можливість аналізувати великі об'єми даних у режимі реального часу, а Amazon Kinesis дозволяє збирати та аналізувати стрічкові дані.

Інший популярний сервіс для аналізу даних - Google Cloud Platform (GCP). GCP надає великий набір інструментів для обробки та аналізу даних, зокрема Google BigQuery, Google Cloud Dataproc, Google Cloud Dataflow та інші. Google BigQuery дозволяє здійснювати швидкий та ефективний аналіз великих обсягів даних у реальному часі. Google Cloud Dataproc є повністю керованим сервісом, який дозволяє запускати Apache Hadoop та Apache Spark на віртуальних машинах, що забезпечує швидку обробку та аналіз даних. Google Cloud Dataflow дозволяє використовувати потокову обробку даних з використанням Apache Beam, що дозволяє аналізувати дані в режимі реального часу та забезпечує автоматичне масштабування.

Ще один сервіс для аналізу даних, який набуває популярності - Microsoft Azure. Azure надає великий набір інструментів для аналізу даних, зокрема Azure HDInsight, Azure Databricks, Azure Stream Analytics та інші. Azure HDInsight дозволяє запускати Hadoop та Spark на віртуальних машинах з підтримкою кластерів різного розміру, що дозволяє аналізувати великі обсяги даних. Azure Databricks дозволяє проводити розподілені обчислення та розробку моделей

машинного навчання на основі Apache Spark, що забезпечує високу продуктивність та масштабованість. Azure Stream Analytics дозволяє збирати та аналізувати стрічкові дані в режимі реального часу.

Крім великих хмарних сервісів, існують також інші сервіси для аналізу даних та наближених обчислень, зокрема IBM Watson Studio, Oracle Cloud Infrastructure, Alibaba Cloud та інші. IBM Watson Studio дозволяє розробляти та виконувати моделі машинного навчання та аналізувати дані за допомогою Apache Spark та Jupyter Notebooks. Oracle Cloud Infrastructure надає високопродуктивні сервіси для аналізу даних, зокрема Oracle Autonomous Data Warehouse та Oracle Analytics Cloud. Alibaba Cloud надає інструменти для аналізу даних та розробки моделей машинного навчання, зокрема MaxCompute та Machine Learning Platform for AI.

Усі ці сервіси надають користувачам можливість аналізувати великі обсяги даних та розробляти моделі машинного навчання, що допомагає компаніям та дослідникам знайти нові способи використання даних та зробити більш точні прогнози. Важливо пам'ятати, що кожен сервіс має свої особливості та переваги, тому користувачам потрібно обрати той, який найбільш відповідає їхнім потребам та вимогам.

Одним з найважливіших факторів при виборі сервісу для аналізу даних є його безпека. Користувачам потрібно забезпечити безпеку даних, які вони збирають та оброблюють, тому необхідно ретельно перевіряти політики конфіденційності та захисту даних, які надає сервіс. Крім того, важливо враховувати потенційні ризики, пов'язані з використанням хмарних сервісів та передачею даних через мережу Інтернет.

Отже, сервіси для аналізу даних та наближених обчислень надають великий вибір інструментів для обробки та аналізу даних, розробки моделей машинного навчання та зроблення точних прогнозів. При виборі сервісу важливо ретельно аналізувати його функціональні можливості, технічну підтримку, безпеку та фінансові аспекти, щоб знайти найкращий варіант, який найбільш відповідає потребам та вимогам користувача.

Список літератури

1. Dean, J., & Ghemawat, S. (2008). *MapReduce: Simplified data processing on large clusters*. *Communications of the ACM*, 51(1), 107-113.
2. Groves, P., Kayyali, B., Knott, D., & Van Kuiken, S. (2016). *The 'big data' revolution in healthcare: Accelerating value and innovation*. McKinsey & Company.
3. McKinney, W. (2010). *Data structures for statistical computing in Python*. *Proceedings of the 9th Python in Science Conference*, 51-56.
4. Reilly, M. (2015). *Cloud computing and big data*. *International Journal of Ambient Computing and Intelligence (IJACI)*, 6(1), 1-13.