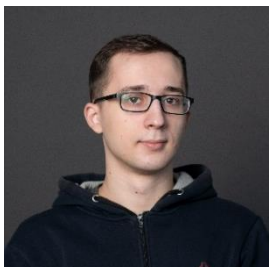


УДК 004.75

ИСПОЛЬЗОВАНИЕ ОБЛАЧНЫХ СЕРВИСОВ ДЛЯ РЕШЕНИЯ ЗАДАЧ, СВЯЗАННЫХ С ПРИМЕНЕНИЕМ ТЕХНОЛОГИИ BIG DATA



М.И. Одиноченко
студент БГУИР



С.Н. Нестеренков
доцент кафедры
программного обеспечения
информационных технологий,
кандидат технических
наук, доцент

Белорусский государственный университет информатики и радиоэлектроники, факультет компьютерных систем и сетей, кафедра программного обеспечения информационных технологий, Республика Беларусь
E-mail: s.nesterenkov@bsuir.by

М.И. Одиноченко

Студент 4 курса специальности «Программное обеспечение информационных технологий» БГУИР.

С.Н. Нестеренков

Окончил БГУИР в 2007 году по специальности «Программное обеспечение информационных технологий», окончил магистратуру БГУИР в 2008 по специальности «Системный анализ, управление и обработка информации», окончил аспирантуру БГУИР в 2013 по специальности «Системный анализ, управление и обработка информации», окончил магистратуру БГУИР в 2013 по специальности «Экономика и управление народным хозяйством», в 2017 защитил диссертацию на соискание ученой степени кандидата технических наук по специальности «Системный анализ, управление и обработка информации».

Аннотация. Целью работы являлось анализ возможностей и актуальности наиболее популярных решений облачных вычислений, описание их преимуществ и недостатков. Были определены наиболее популярные сервисы, причины их использования для обработки больших объемов информации, сделаны выводы об эффективности использования.

Ключевые слова: Big Data, облачный сервис, анализ данных, AWS, Azure, Google Cloud.

Введение.

Big Data[1] имеет популярность в современном мире технологий. Её использование нашлось во многих сферах: медицина, наука, реклама, анализе различных закономерностей. И это неслучайно, с каждым годом количество данных и скорость их появления только увеличивается, интернет всё больше входит в мельчайшие повседневные вещи. Однако выделить какую-либо осмысленную информацию из данных становится тяжело, они ценны, если из них можно выделить сведения, сделать анализ этой информации и применить для каких-либо действий.

Различные данные о человеке или его деятельности чаще всего записываются машиной в больших объёмах, зачастую являющихся бесполезными [2]. По этой причине нужно уметь правильно фильтровать имеющиеся данные, повышать их информативность при сборе и анализе. Обработка Big Data представляет собой довольно тяжёлый и

трудозатратный процесс. Исследование возможностей упрощения анализа данных имеет большую важность, любое улучшение или оптимизация, помноженные на огромный массив данных для обработки, может сэкономить очень много времени.

На сегодняшний день можно выделить основные задачи, связанные с обработкой Big Data[3]:

- сбор данных для анализа;
- хранение данных;
- выполнение облачных вычислений и анализ;
- визуализация и использование результатов обработки.

Сбор данных – первая проблема в сфере Big Data. Необходимо собирать максимальное количество данных с наибольшим потенциалом для анализа. Правильное решение для работы с Big Data облегчает работу, предоставляя весь необходимый функционал по сбору разработчику. Могут пригодиться самые разные данные: реального времени, структурированные различным образом, прошедшие первичную обработку.

Хранение данных. Для больших объемов данных в первую очередь требуются большие хранилища. Так же не лишним будет простой интерфейс работы с ним и возможности по взаимодействию с исполняемым кодом. В таких хранилищах могут храниться как исходные собранные данные, так и промежуточные результаты.

Вычисления и анализ результатов. Данные обрабатываются развёрнутыми в облаке программами нужным образом и преобразованными в нужный для последующего использования формат. Далее результаты выполнения алгоритмов сохраняются в хранилище для последующей обработки более высокого уровня или предоставляются для визуализации.

Использование и визуализация результатов обработки. Данные, полученные в предыдущем пункте, уже представляют некую коммерческую ценность. Из них можно делать выводы, продавать, в общем, извлекать выгоду. Конечно, эти результаты неточны и имеют только некую степень достоверности, зависящую от качества анализа. Информация так же зачастую визуализируется и упрощается, ведь человек не способен осознать по-настоящему большие объёмы данных. Для этого используются графики, диаграммы и таблицы, основываясь на которых можно сделать презентацию данных как товара или найти закономерность.

Используя облачные сервисы можно быстро разворачивать приложения, адаптивно менять их мощность и конфигурации, а затем отключать, не тратя деньги на поддержание. В современных реалиях оплачивается не время или домен, а только затраченные сервером ресурсы, что является оптимальным вариантом для обеих сторон. Большие данные можно обработать всеми ресурсами сервера за короткий срок и меньшую стоимость, что и является ключевым моментом в выборе облачных решений. Для сравнения будем использовать наиболее популярных провайдеров решений.

Перейдём к разбору конкретных преимуществ каждого из производителей облачных решений. Будем рассматривать по следующим пунктам: возможности по хранению данных, вычислительные мощности, возможности анализа, стоимость, доступность.

Комплекс решений **AWS** имеет сервис Simple Storage Service (S3) для хранения файлов и Amazon Glacier для их архивирования, а также множество более узкоспециализированных сервисов, позволяющих минимизировать расходы на необходимый функционал, оплачивая только его. В этом году AWS анонсировала новую функцию хранилищ данных, Redshift Spectrum, которая позволит заказчикам выполнять запросы к эксабайтным объемам данных.

Вычислительная мощность[4] – основа коммерческого использования Big Data, AWS имеет основную сетевую службу на основе Elastic Compute Cloud (EC2) и решение AWS Lambda для запуска программных файлов в облаке. В EC2 интегрированы такие

подсервисы, как AWS Cloud Watch, Amazon EC2 Container Service, AWS Auto Scaling и многие другие полезные инструменты.

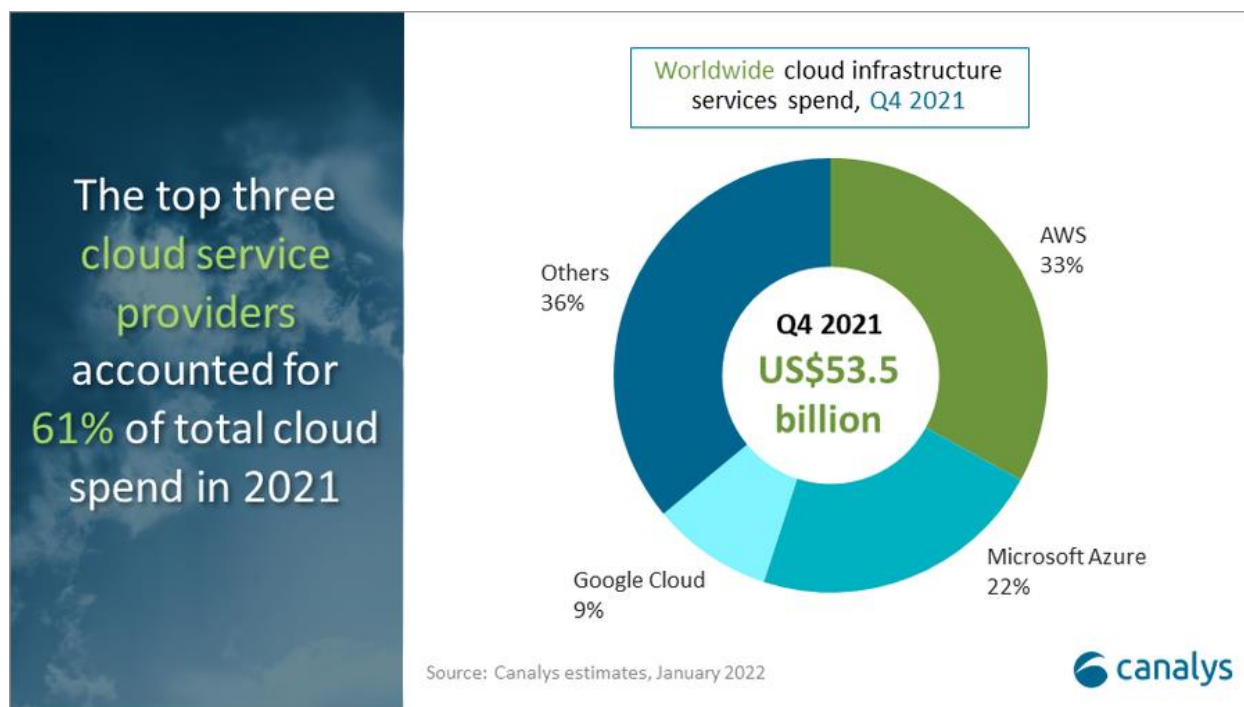


Рисунок 1. Расходы на услуги облачной инфраструктуры в четвертом квартале 2021

В практически каждое крупное решение встроены свои инструменты анализа. Они используют передовые технологии: искусственный интеллект[5] (ИИ), машинное обучение (ML), высокопроизводительные вычисления (HPC) и приложения для обработки мультимедиа для всех облачных данных. AWS недавно открыла к использованию прорывной аналитический сервис Quick Sight. Он включает в себя готовые шаблоны, а сама услуга значительно дешевле по сравнению с классическими решениями BI.

Имеется 3 варианта стоимости услуг: оплата используемых ресурсов, пробная версия до 3 лет, уменьшение стоимости с увеличением нагрузки. Оплата в каждом случае почасовая.

Сервера AWS расположены практически по всему миру, что минимизирует задержку до них.

В общем, AWS не зря является лидером рынка, его возможности очень велики, что, впрочем, повышает входной порог сервисов, ведь взаимодействие большого количества сервисов может вызвать множество проблем у начинающего пользователя.

Microsoft Azure имеет свои хранилища, обладающие высокой надёжностью, а так же активно разрабатывает возможности резервного копирования и восстановления файлов, создание бэкапов. Сервис StorSimple, используется в качестве гибридного сервиса хранения для корпоративных клиентов, значительно повысил эффективность компании.

В плане вычислительных мощностей, Azure основывается на работе множества виртуальных машин с загруженными программами (Virtual Machine Scale Sets) в качестве центра обработки данных.

Аналитические инструменты менее разнообразны чем у AWS, однако за последнее время компания улучшила свои инструменты аналитики[6] и машинного обучения, а также создала подсистему Data Lake Analytics и представила машинное обучение.

Azure имеет другой подход и к ценообразованию. Оплата ресурсов поминутная, с возможностью скидки в случае больших объёмов.

Особенностью решений от Microsoft является поддержка Linux-систем и общая универсальность. Однако в среднем они менее надёжны и масштабируемы, цена же наибольшая из рассматриваемых компаний.

Google Cloud имеет своё хранилище, имеющее неплохую надёжность и опции по безопасности использования.

Для вычислений Google Cloud Platform использует свой собственный сервис Compute Engine для исполнения программ и анализа. Он поддерживает большую часть основных инструментов облачной платформы. Так же имеется сервис Cloud Dataproc для работы с Hadoop[7], MapReduce, Spark, Pig и Hive.

В настоящее время доступна специальная область аналитики Big Data. Проекты Cloud Vision API, Cloud Speech API и Google Translate API могут быть интегрированы в сторонние ресурсы. В инструменты аналитики интегрированы результаты собственных исследований Google в области машинного обучения.

GCP монетизирует свои ресурсы по тем же принципам, что и Azure, но за исключением того, что итоговая стоимость округляется каждые 10 минут использования.

Заключение.

Пользование услугами провайдеров облачных решений не ставится под сомнение, слишком простым и мало затратным оно является. Крупные компании имеют наиболее выгодные тарифы использования и счёт объёмов. Выбор состоит из AWS, Microsoft Azure и Amazon Web Services. AWS имеет явное преимущество по покрытию территорий серверами для доступа, охватывает значительную часть от общей доли рынка и является лидером по количеству сервисов. Google Cloud является наиболее доступной платформой по критерию цены. При выборе подходящего поставщика облачных услуг стоит учитывать свои особенные потребности и рабочую нагрузку компании. Microsoft Azure - одна из самых универсальных платформ, так как поддерживается интеграция с ОС Windows, являющейся самой распространённой на данный момент.

Список использованных источников

[1] Жалейко, Д. А. Применение облачных технологий в дистанционном обучении / Д. А. Жалейко, С. Н. Нестеренков, Д. В. Басак // Качество образовательного процесса: проблемы и пути развития = Quality of the educational process: challenges and ways of development: материалы II Междунар. науч.-практ. конф., Минск, 30 апреля 2021 г. / Белорус. гос. ун-т информатики и радиоэлектроники ; редкол.: Ю.Е. Кулешов [и др.]. - Минск, 2021. - С. 161-162.

[2] Мигалевич, С.А. Облачные решения в сфере IT-образования / С.А. Мигалевич, С.Н. Нестеренков, А.Н. Марков // Проблемы повышения эффективности образовательного процесса на базе информационных технологий : материалы XI Междунар. науч.-практ. конф., Минск, 27 апреля 2018 г. / Белорус. гос. ун-т информатики и радиоэлектроники ; редкол.: Ю.Е. Кулешов [и др.]. - Минск, 2018. - С. 57-59.

[3] Ваганова О.И., Дворникова Е.И., Кутепов М.М., Лунева Ю.Б., Трутанова А.В. Возможности облачных технологий в электронном обучении // Международный журнал прикладных и фундаментальных исследований. – 2017. – № 6-2. – С. 183-187.

[4] Maryam Farsi, Hosseinian-Far A., Alireza Daneshkhah, Tabassom Sedighi. Strategic Engineering for Cloud Computing and Big Data Analytics // Springer International Publishing. - 2017, - С. 3-27.

[5] Нестеренков, С.Н. Применение искусственных нейронных сетей в информационной системе учреждения высшего образования / С.Н. Нестеренков, Н.В. Ющенко, А.Д. Радкевич // Актуальные вопросы профессионального образования = Actual issues of professional education : тезисы докладов II Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 11 апреля 2019 г.) / редкол. : С. Н. Анкуда [и др.]. - Минск : БГУИР, 2019. - С. 197-198.

[6] Нестеренков, С. Н. Модифицированный генетический алгоритм для обучения нейронной сети / С. Н. Нестеренков, К. П. Белов // Информационные технологии и системы 2017 (ИТС 2017) : материалы междунар. науч. конф., Минск, 25 окт. 2017 г. / Белорус. гос. ун-т информатики и радиоэлектроники ; редкол.: Л. Ю. Шилин [и др.]. - Минск, 2017. - С. 204-205.

[7] Ms. D.Prema swarupa rani. "Real-time Big Data Analytics and parallel processing using Hadoop on Remote Sensing data." IOSR Journal of Computer Engineering (IOSR-JCE) 19.4 (2017): 29-32.

USING CLOUD SERVICES FOR SOLVING PROBLEMS RELATED TO THE APPLICATION OF BIG DATA TECHNOLOGY

M.I. ADZINCHANKA
*Student of Belarusian State
University of Informatics and
Radioelectronics*

S.N. NESTERENKOV,
*PhD Associate professor of department of the software
of information technologies*

*Belarusian State University of Informatics and Radioelectronics, Faculty of Computer Systems and Networks,
Department of Information Technology Software, Republic of Belarus
E-mail: s.nesterenkov@bsuir.by*

Abstract. The purpose of the work was to analyze the capabilities and relevance of the most popular cloud computing solutions, to describe their advantages and disadvantages. The most popular services were identified, the reasons for their use for processing large amounts of information, conclusions were drawn about the effectiveness of the use.

Keywords: Big Data, cloud service, data analysis, AWS, Azure, Google Cloud.