

---

## ОПТИМИЗАЦИЯ АНАЛИЗА ДАННЫХ С ПОМОЩЬЮ GOOGLE CLOUD: ИДЕИ И ИННОВАЦИИ ДЛЯ СПЕЦИАЛИСТОВ ПО ДАННЫМ

### **Маштаков Никита Сергеевич**

Студент бакалавриата 4 курс

МИРЭА-Российский технологический университет (РТУ МИРЭА)

Институт информационных технологий

### **Гайнутдинова Аделя Ришатовна**

Студент бакалавриата 4 курс

МИРЭА-Российский технологический университет (РТУ МИРЭА)

Институт информационных технологий

### **Аннотация**

---

В данной статье описана роль Google Cloud Platform (GCP) для обработки и анализа данных в корпоративной среде. В ней освещаются преимущества GCP, включая сокращение расходов и повышение эффективности ИТ-инфраструктуры. Описываются основные аспекты GCP, такие как безопасность, глобальная доступность и масштабируемость, а также рассматриваются различные модели облачных вычислений (IaaS, PaaS, SaaS) и ключевые компоненты платформы, включая BigQuery и Compute Engine. Статья также включает примеры практического использования GCP в крупных компаниях.

---

**Ключевые слова:** Google Cloud, анализ данных, ИТ-инфраструктура, облачные вычисления, хранилище данных.

---

## OPTIMIZING DATA ANALYTICS WITH GOOGLE CLOUD: INSIGHTS AND INNOVATIONS FOR DATA SCIENTISTS

### **Nikita S. Mashtakov**

Bachelor's degree student 4nd year

MIREA-Russian Technological University (RTU MIREA)

Institute of Information Technology

### **Adelya R. Gainutdinova**

Bachelor's degree student 4nd year

MIREA-Russian Technological University (RTU MIREA)

Institute of Information Technology

---

ABSTRACT

---

This article describes the role of Google Cloud Platform (GCP) for data processing and analysis in an enterprise environment. It highlights the benefits of GCP, including reducing costs and improving the efficiency of the IT infrastructure. The main aspects of GCP, such as security, global availability and scalability, are described, as well as various cloud computing models (IaaS, PaaS, SaaS) and key components of the platform, including BigQuery and Compute Engine. The article also includes examples of practical use of GCP in large companies.

---

**Keywords:** Google Cloud, data analysis, IT infrastructure, cloud computing, data storage.

---

Большинство компаний сталкиваются с проблемой использования и обработки огромных объемов данных. Вычислительные сервисы лишь иногда обладают техническими возможностями для работы с такими масштабами. Доступны облачные ресурсы для облегчения различных вычислений и анализа данных, что может значительно упростить работу ИТ-специалистов и повысить эффективность.

Преимущества облачных вычислений для специалистов по обработке данных

Google Cloud Platform (GCP) - это эффективная облачная платформа, которая может улучшить управление данными, сократить финансовые и людские ресурсы для управления инфраструктурой и обеспечить настройку сети. GCP поддерживает систему выставления счетов с оплатой по мере поступления, поэтому компании взимают плату только за вычислительные ресурсы, используемые в течение каждого цикла выставления счетов. Это позволяет гибко увеличивать или уменьшать масштаб использования по мере необходимости. Доступ к серверам поддерживается круглосуточно, а платформа отличается высокой степенью безопасности [4].

Благодаря этой платформе специалисты по обработке данных могут улучшить управление своим хранилищем данных и различными системами. Кроме того, этот сервис позволяет максимально оптимизировать и настраивать множество вычислительных процессов и создавать облачные решения.

Итак, теперь, когда мы выяснили, что такое GCP, давайте выделим наиболее важные преимущества его использования:

- высокая степень безопасности, которая обеспечивается созданием эффективных методов защиты;
- доступ к глобальной сети – крупнейшей в мире;
- многие сервисы обеспечивают надежное хранение данных, а также их обработку для решения различных задач специалистов;
- снижение финансовых затрат – если сравнивать с другими облачными провайдерами, эта платформа обойдется на 20-25% дешевле – в целом использование GCP позволяет компании экономить более 52% затрат;
- экономия рабочей нагрузки сервисов компании до 80%; в противном случае перегрузки приводят к сбоям программного обеспечения;
- предусмотрена возможность выбора облачной конфигурации и объема памяти (экономия до 50%) [5].

Кроме того, использование GCP обеспечивает искусственный интеллект и мощные аналитические возможности. Учитывая все эти преимущества, облачную платформу Google

можно рассматривать как очень эффективный инструмент для хранения и обработки огромных объемов данных, позволяющий значительно экономить затраты и ресурсы.

#### Типы облачных вычислений

Облачная платформа Google предоставляет несколько вариантов облачных решений, позволяя вам оптимально выбрать наиболее эффективное решение для вашей компании. Каждый вариант требует разного уровня технических знаний в области анализа данных. При выборе облачного варианта вам следует определиться с целями компании и желаемыми результатами.

#### IaaS

Инфраструктура как услуга. Компания располагает инфраструктурой с соответствующим программным обеспечением, сетевыми кабелями, контейнерным хранилищем, процессорами и необходимой оперативной памятью. То есть техническое оборудование предоставляется в аренду. Остальное вы должны сделать самостоятельно (администрирование) – настройка сети, установка операционной системы, выбор процессора для требуемой нагрузки и многое другое.

#### PaaS

Платформа как услуга. В данном случае у компании есть платформа с полноценной средой, которая позволяет ей разрабатывать приложения. Поставщик выполняет весь технический отбор (процессор, оперативная память, хранилище и т.д.).

#### SaaS

Программное обеспечение как услуга. Провайдер предлагает различные сервисы с программным обеспечением, которое они также обслуживают (включая автоматические обновления). Провайдер полностью управляет техническим оборудованием и обеспечивает безопасность.

#### Основные компоненты Google Cloud для анализа данных

GCP включает в себя несколько компонентов, каждый из которых имеет конкретное назначение для анализа данных:

- вычислительные вычисления – позволят вам выполнять различные вычисления любой сложности, независимо от объема используемых данных;
- облако: облачное хранилище с базой данных – обеспечивает надежную сохранность размещенных там данных;
- инструменты управления – позволяют отслеживать, регистрировать, генерировать отчеты об ошибках, отслеживать, отлаживать и многое другое;
- сетевое взаимодействие: виртуальная облачная сеть – доменная система, сеть доставки контента, взаимосвязи, балансировка нагрузки;
- BigData: большие данные;
- Cloud Machine: машинное обучение;
- инструменты облачной разработки (хранилища, конечные точки, развертывание);
- безопасность и идентификация.

Весь этот набор инструментов позволяет эффективно обрабатывать огромное количество данных и выполнять процессы любой сложности. Далее мы подробнее рассмотрим основные компоненты облачной платформы Google.

#### BigQuery

Являясь сервисом облачных вычислений типа IaaS, BigQuery предоставляет возможности хранения, анализа и управления огромными объемами данных. Сюда входят

опции для создания, удаления и импорта данных. Также возможно предоставить доступ к хранилищу данных третьим лицам или команде специалистов. Вы можете интегрировать сохраненные данные с различным программным обеспечением. Кроме того, опция позволяет создавать и запускать машинное обучение. Пользователям этого сервиса предоставляется 10 ГБ облачного хранилища, и они могут выполнять до одного терабайта запросов ежемесячно.

#### App Engine

Облачные вычисления типа PaaS. Этот сервис позволяет разрабатывать и размещать веб-сервисы и мобильные приложения. Эта система обладает большим функциональным набором управления, позволяющим масштабировать программное обеспечение. Кроме того, широкий спектр программных API-интерфейсов может соединять различные приложения, позволяя ускорить разработку продукта.

Сервис поддерживается на различных языках программирования. Основные ресурсы предоставляются бесплатно только в информационных целях. В случае платного набора пользователь платит исключительно за использованный объем.

#### Compute Engine

Облачные вычисления типа IaaS. Предоставляет возможность создавать и запускать облачные машины на основе инфраструктуры Google. Доступ предоставляется через интерфейс. Другие функциональные возможности включают облачные машины point cloud и возможность шифровать данные и оптимизировать ресурсы с помощью автоматических рекомендаций. Пользователю предоставляется одна виртуальная машина в бесплатное пользование.

#### Движок Kubernetes

Основная цель этого сервиса - работа с приложениями с контейнерами. Предоставляет возможность изменять размер и развертывать приложения автоматизированным способом. Он также включает в себя функции безопасности, такие как шифрование данных и сканирование контейнеров для выявления слабых мест. Сервис поддерживает базовые технологии контейнеризации и аппаратную виртуализацию.

#### Облачное хранилище

Облачные вычисления типа IaaS. Облачный сервис, в котором данные не имеют структуры. Включает в себя возможность изменять размер виртуального облака Google наряду с дополнительной функциональностью. Этот сервис обеспечивает хранение до пяти терабайт в контейнерах с назначением индивидуальных ключей. Кроме того, есть возможность оптимизировать данные и сократить количество ненужных данных.

#### Хранилище данных

Предоставляет масштабируемую нереляционную базу данных для приложений. Набор опций позволяет управлять сегментацией и синхронизацией автоматизированным способом. Этот сервис идеально подходит для обработки данных небольшого масштаба.

#### Реестр контейнеров

Это единый реестр контейнеров, который предоставляет возможность управлять образами, выполнять процессы для обнаружения уязвимостей и настраивать доступ. В этом случае нет необходимости арендовать виртуальную машину или находить дисковое пространство для компании.

Вся инфраструктура и необходимые инструменты представлены в виде полностью готового решения. Вы можете работать с контейнерами через панель управления. Этот сервис избавляет разработчиков от проблем с хранением и развертыванием переносных контейнеров. Вы также можете интегрироваться с процессами CI/CD.

#### Облачные функции

С помощью этого сервиса вы можете запускать приложения в защищенной среде, которая может масштабироваться, без создания и обслуживания облачных виртуальных машин. То есть разработанное приложение будет запущено на сервере компании, предоставляющей эту услугу, и нет необходимости запускать и тестировать приложение на ее сервере.

Мы перечислили самые основные сервисы GCP, но, конечно, их в несколько раз больше, не только для обработки данных и разработки программного обеспечения, но и для создания высокого уровня безопасности.

#### Примеры использования Google Cloud

Вот несколько примеров крупных компаний, которые использовали сервис GCP для решения различных программных проблем.

Spotify предлагает множество музыкальных треков и видео. У этой платформы более 75 миллионов подписчиков и около 2 миллиардов плейлистов. Благодаря таким сервисам GCP стало возможным создать надежную инфраструктуру и повысить эффективность сервиса Spotify. Также возможно выполнять запросы, на выполнение которых раньше уходил день, но теперь все происходит за несколько минут. Это позволяет оптимизировать работу сервиса с пользователями [3].

X (ранее Twitter): популярная социальная сеть насчитывает около 330 миллионов активных пользователей. Twitter является хранилищем огромных объемов данных и благодаря использованию GCP значительно повысил безопасность своей платформы и расширил возможности аварийного восстановления [1].

BestBuy: международная компания по торговле бытовой электроникой с более чем 1000 магазинами по всему миру. В свое время они создали свое приложение с использованием App Engine, позволяющее пользователям создавать свой собственный список пожеланий и делиться им с друзьями [2].

Эти и многие другие глобальные компании смогли оптимизировать работу своих платформ, включая Twitter, PayPal, eBay и 20th Century Fox.

#### Вывод

Google Cloud Platform является основным поставщиком облачных услуг, который предоставляет различные вычислительные услуги и услуги по обработке данных для обеспечения аналитики и оптимизации процессов. Кроме того, можно использовать инструменты искусственного интеллекта и машинного обучения.

Благодаря GCP компании смогут развиваться, одновременно экономя значительные средства и ресурсы. GCP предоставляет готовые решения для различных нужд, включая модернизацию инфраструктуры, интеграцию и безопасность. Эти услуги будут эффективны во многих областях деятельности компании, особенно в работе специалистов по обработке данных.

#### Список литературы:

1. «How twitter modernized its data processing with Google Cloud». [Электронный ресурс] / Режим доступа: <https://cloud.google.com/blog/products/data-analytics/how-twitter-modernized-its-data-processing-with-google-cloud> (Дата обращения: 07.01.2024)
2. «Google Cloud for BestBuy» [Электронный ресурс] / Режим доступа: <https://cloud.google.com/files/BestBuy.pdf> (Дата обращения: 07.01.2024)
3. «Innovation at the speed of Spotify» [Электронный ресурс] / Режим доступа: <https://cloud.google.com/customers/featured/spotify> (Дата обращения: 06.01.2024)
4. «Why Google Cloud» [Электронный ресурс] / Режим доступа: <https://cloud.google.com/why-google-cloud/?hl=en> (Дата обращения: 06.01.2024)

5. «Что такое GCP и как вы можете использовать его для своего бизнеса» [Электронный ресурс] / Режим доступа: <https://cloudfresh.com/ru/cloud-blog/cto-takoe-gcp-y-kak-vy-mozhete-yspolzovat-ego-dlya-svoego-byznesa/> (Дата обращения: 06.01.2024)