

**BỘ THÔNG TIN VÀ TRUYỀN THÔNG
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN
THÔNG**



BÁO CÁO THỰC TẬP CƠ SỞ

**BÁO CÁO TÌM HIỂU VỀ ETL VÀ DATA
PIPELINE**

Giảng viên hướng dẫn : Kim Ngọc Bách

Họ và tên : Ngô Vũ Minh Quý

Mã sinh viên : B22DCVT427

Lớp : E22CQCN02-B

MỤC LỤC

- I. Giới thiệu về Google Cloud Platform
- II. Lý do sử dụng GCP trong phát triển và vận hành hệ thống
- III. Cài đặt và làm quen với GCP
- IV. Các dịch vụ tiêu biểu trong GCP
- V. Đánh giá và kết luận

I. Giới thiệu về Google Cloud Platform

Google Cloud Platform (GCP) là một bộ công cụ điện toán đám mây được phát triển bởi Google, cung cấp hạ tầng, nền tảng và dịch vụ phần mềm dưới dạng dịch vụ (IaaS, PaaS, SaaS). GCP cho phép các tổ chức xây dựng, triển khai và mở rộng các ứng dụng hiện đại bằng cách sử dụng hạ tầng tương tự như Google sử dụng cho các sản phẩm như Gmail, YouTube và Google Search.

Được thiết kế với hiệu năng, bảo mật và khả năng mở rộng cao, GCP hiện là một trong ba nền tảng điện toán đám mây lớn nhất thế giới, cùng với Amazon Web Services (AWS) và Microsoft Azure.

Các đặc điểm nổi bật của GCP:

- **Hiệu năng mạnh mẽ:** Chạy trên hạ tầng của Google – một trong những hệ thống nhanh và đáng tin cậy nhất thế giới.
- **Mô hình định giá linh hoạt:** Thanh toán theo mức sử dụng thực tế (pay-as-you-go), tiết kiệm chi phí cho doanh nghiệp.
- **Tích hợp AI/ML mạnh mẽ:** Hỗ trợ các công cụ như Vertex AI, BigQuery ML giúp xây dựng mô hình học máy nhanh chóng.
- **Bảo mật cao:** Hệ thống bảo mật cấp độ doanh nghiệp, đáp ứng nhiều tiêu chuẩn toàn cầu (ISO, GDPR,...).

II. Lý do sử dụng GCP trong phát triển và vận hành hệ thống

1. Tối ưu chi phí và hiệu suất

GCP cung cấp nhiều lựa chọn về tài nguyên như máy ảo, lưu trữ đám

mây, cơ sở dữ liệu quản lý,... với chi phí linh hoạt. Các tính năng như tự động mở rộng (autoscaling), khởi tạo máy chủ theo nhu cầu (serverless) giúp tiết kiệm tối đa tài nguyên.

2. Hỗ trợ phát triển nhanh chóng

- Với các dịch vụ như App Engine (nền tảng không máy chủ) hoặc Cloud Run (chạy container), lập trình viên có thể triển khai ứng dụng mà không cần quản lý hạ tầng.
- Tích hợp sẵn công cụ DevOps: Cloud Build, Artifact Registry, Cloud Source Repositories,...

3. Phân tích và xử lý dữ liệu lớn

- GCP cung cấp BigQuery – một hệ quản trị CSDL phân tích cực kỳ mạnh, cho phép chạy các truy vấn SQL trên hàng tỷ dòng dữ liệu chỉ trong vài giây.
- Hỗ trợ ETL bằng Dataflow, Dataproc (Apache Spark), và tích hợp với Looker cho trực quan hóa.

4. Khả năng mở rộng toàn cầu

Hệ thống mạng lưới trung tâm dữ liệu toàn cầu (global infrastructure) giúp người dùng triển khai ứng dụng trên nhiều khu vực khác nhau, đảm bảo hiệu năng và độ trễ thấp.

III. Cài đặt và làm quen với GCP

Bước 1: Tạo tài khoản Google Cloud

- Truy cập: <https://cloud.google.com>
- Đăng nhập bằng tài khoản Google và kích hoạt dịch vụ GCP.
- GCP thường cung cấp khoản tín dụng miễn phí \$300 cho người dùng mới.

Bước 2: Kích hoạt Billing và tạo Project

- GCP yêu cầu gắn phương thức thanh toán để sử dụng hầu hết dịch vụ.
- Tạo một “Project” – đơn vị tổ chức các dịch vụ, tài nguyên riêng biệt.

Bước 3: Làm quen với giao diện Cloud Console

- Giao diện trực quan cho phép điều khiển dịch vụ.
- Có thể sử dụng Cloud Shell – môi trường terminal tích hợp sẵn CLI (gcloud).

Bước 4: Cài đặt SDK cục bộ (tuỳ chọn)

- Tải Google Cloud SDK để điều khiển GCP từ máy tính cá nhân:

IV. Các dịch vụ tiêu biểu trong GCP

1. Compute Engine (IaaS)

- Tạo máy ảo (VM) linh hoạt để chạy ứng dụng, có thể cài Linux, Windows,...
- Hỗ trợ snapshot, load balancing, auto-scaling.

2. App Engine (PaaS)

- Chạy ứng dụng web không cần quản lý máy chủ.
- Hỗ trợ nhiều ngôn ngữ như Python, Java, Node.js, Go, PHP...

3. Cloud Storage

- Lưu trữ file dạng object (tương tự AWS S3).
- Hỗ trợ backup, CDN, chia sẻ quyền truy cập.

4. BigQuery

- Cơ sở dữ liệu phân tích lớn (OLAP).
- Cho phép truy vấn dữ liệu dung lượng hàng TB đến PB trong vài giây.

5. Cloud Functions / Cloud Run

- Chạy code theo sự kiện (event-driven) hoặc container nhỏ gọn.
- Thích hợp cho ứng dụng serverless và microservices.

6. Firebase (dành cho mobile/web)

- Cung cấp dịch vụ realtime database, authentication, cloud messaging, analytics – phù hợp cho app di động.

7. Looker & Data Studio

- Tạo báo cáo, dashboard BI dựa trên dữ liệu từ BigQuery, Google Sheets,...

2. Truy vấn dữ liệu

- Sử dụng “Explore” để chọn bảng, lọc dữ liệu, nhóm và tạo biểu đồ.
- Looker sẽ tự động tạo truy vấn SQL tương ứng và gửi đến CSDL.

3. Tạo dashboard và báo cáo

- Thêm các biểu đồ: cột, đường, donut, bảng.
- Cấu hình filter tương tác (ví dụ: chọn ngày/tháng, người dùng).
- Lên lịch gửi định kỳ qua email, webhook, hoặc Google Sheets.

4. Quản lý quyền truy cập

- Gán quyền xem/sửa với từng dashboard.
- Phân quyền theo vai trò (admin, developer, viewer).

V. Đánh giá và kết luận

Qua thời gian tìm hiểu và thực hành với Google Cloud Platform (GCP), em nhận thấy GCP là một nền tảng điện toán đám mây toàn diện, đáng tin cậy, và giàu tính năng. Với hệ sinh thái rộng lớn và tích hợp chặt chẽ, GCP đáp ứng tốt các yêu cầu về xây dựng hệ thống, phát triển ứng dụng, phân tích dữ liệu, và triển khai AI/ML.

Một trong những ưu điểm nổi bật của GCP là khả năng truy cập và triển khai toàn cầu – nhờ hệ thống mạng riêng của Google – đảm bảo độ trễ thấp, tốc độ cao và hiệu suất ổn định. Ngoài ra, GCP còn nổi bật với các dịch vụ dành riêng cho xử lý dữ liệu lớn và AI như BigQuery, Vertex AI, Dataflow,... là lựa chọn lý tưởng cho các doanh nghiệp chuyển đổi số.

Với sinh viên ngành công nghệ thông tin, việc làm quen và thành thạo nền tảng GCP sẽ giúp mở ra nhiều cơ hội nghề nghiệp, đặc biệt trong các lĩnh vực như DevOps, Cloud Engineer, Data Analyst, Machine Learning Engineer,... GCP không chỉ là một kỹ năng mà còn là một công cụ không thể thiếu trong thời đại điện toán đám mây.

Việc học và áp dụng GCP ngay từ sớm sẽ tạo tiền đề vững chắc cho sự nghiệp phát triển phần mềm và phân tích dữ liệu chuyên sâu sau này.