

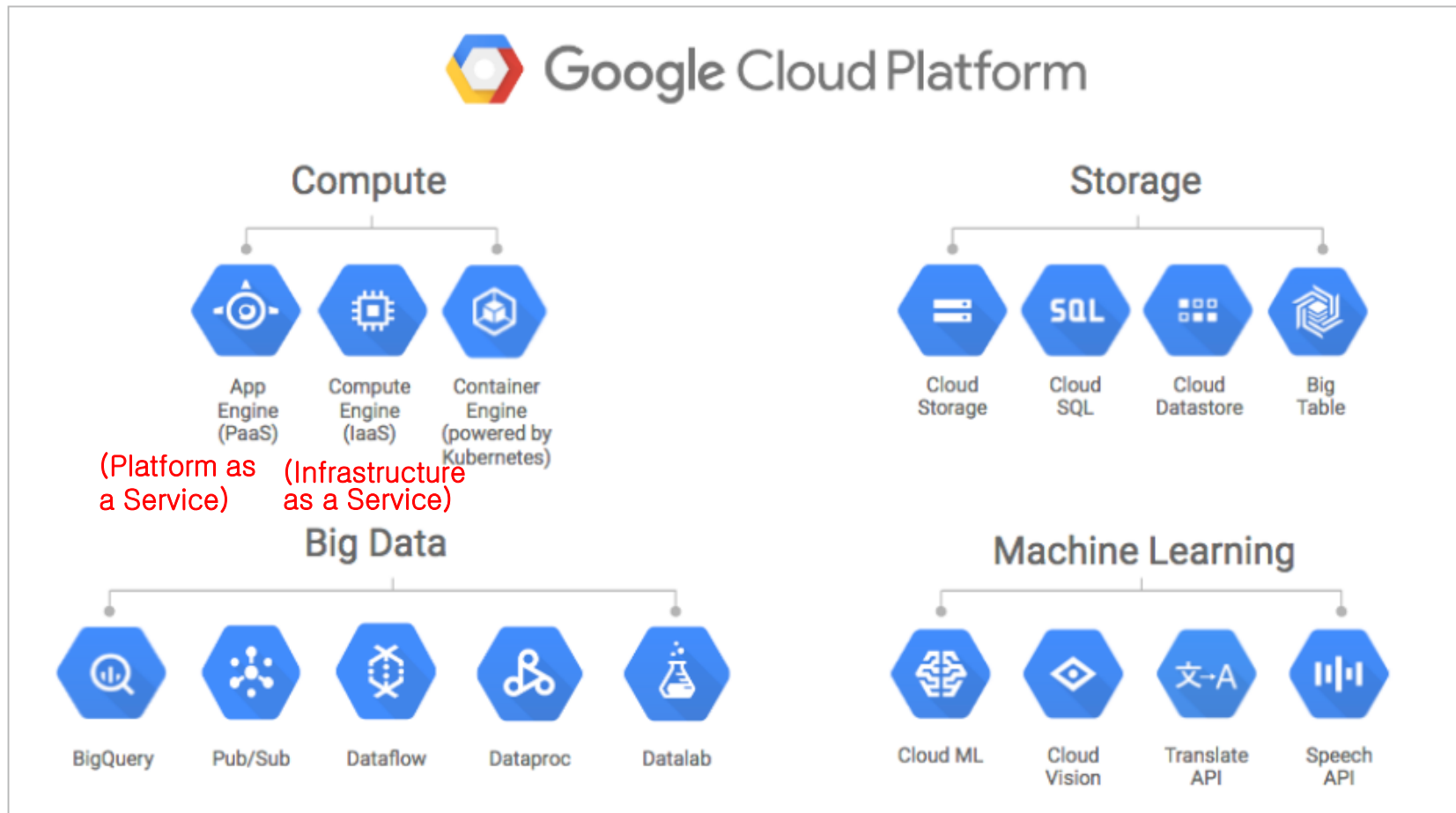


# 데이터 플랫폼 이론-02

엄진영

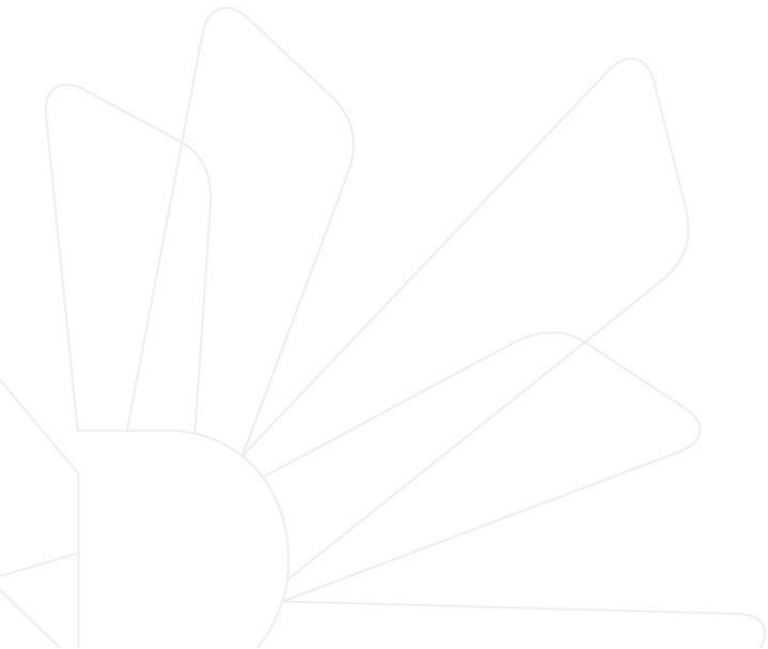
## • 구글 클라우드

- 구글의 데이터 센터 인프라를 기반으로 컴퓨트, 스토리지, 네트워킹, 빅데이터, 머신러닝 등의 서비스를 제공하는 글로벌 클라우드

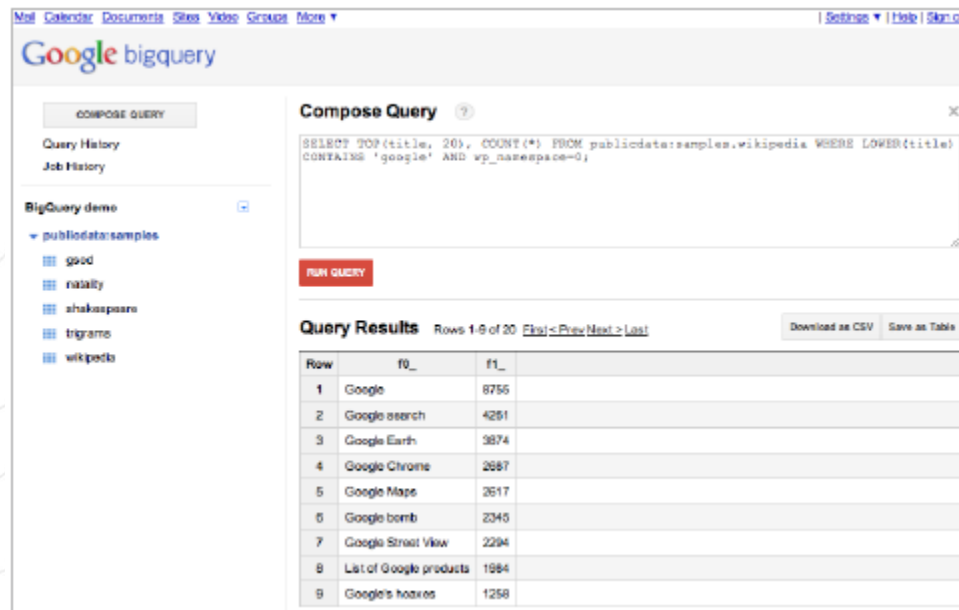


# 구글 클라우드만의 특징

1. 빅데이터와 머신러닝 서비스
2. 구글 전용 네트워크를 이용한 글로벌 커버리지
3. 저렴한 가격 모델



- 데이터의 수집, 가공, 저장기능을 제공하는 서비스
- 빅쿼리(BigQuery)
  - 대규모 데이터 저장 및 분석 플랫폼, 일종의 데이터 웨어하우스
  - 8800개의 CPU와 3600개의 디스크를 사용하는 대규모 인프라를 활용하여 1000억개의 레코드에 의한 질의를 30초 정도에 수행해주며 가격도 저렴
  - SQL과 유사하여 사용하기 매우 쉽다

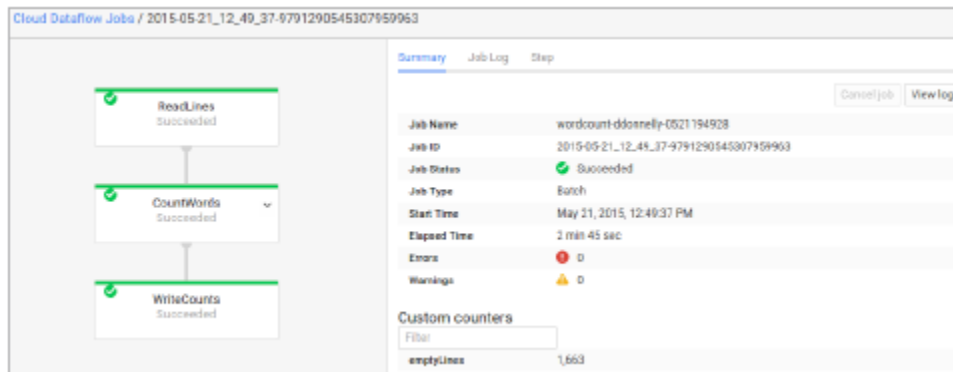


The screenshot displays the Google BigQuery web interface. On the left, there's a sidebar with navigation links like 'Compose Query', 'Query History', and 'Job History'. Below this, a 'BigQuery demo' section shows a tree view of sample datasets including 'good', 'natality', 'shakespeare', 'trigrams', and 'wikipedia'. The main area is titled 'Compose Query' and contains a SQL query: `SELECT TOP(title, 20), COUNT(*) FROM publicdata:samples.wikipedia WHERE LOWER(title) CONTAINS 'google' AND wp_namespace=0;`. A red 'RUN QUERY' button is positioned below the query editor. The 'Query Results' section shows the first 9 rows of the query output, with columns 'Row', 'f0\_', and 'f1\_'. The results list various Google-related terms and their corresponding counts.

Row	f0_	f1_
1	Google	8755
2	Google search	4261
3	Google Earth	3874
4	Google Chrome	2667
5	Google Maps	2617
6	Google bomb	2345
7	Google Street View	2294
8	List of Google products	1984
9	Google's hoaxes	1258

- 데이터플로(Dataflow)

- 실시간 스트리밍 분석 및 배치 분석을 지원하는 플랫폼
- 수집한 데이터를 변환하거나 여러 데이터 소스와 저장소간의 연결(라우팅)을 담당

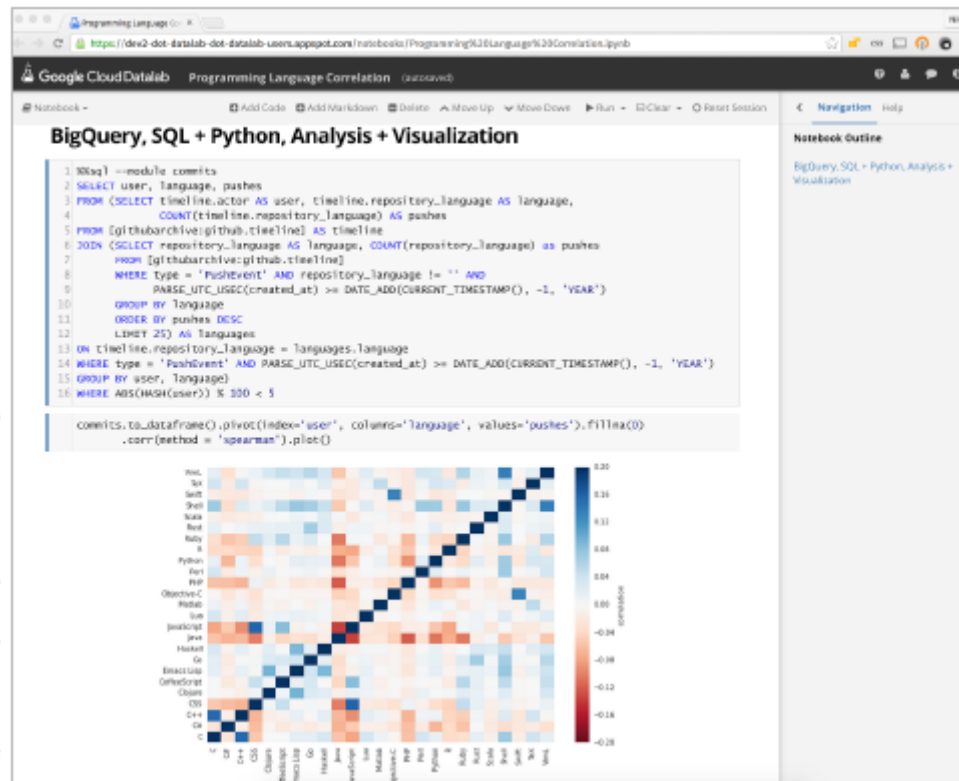


- 펌/섭(pub/sub) 큐

- 대규모 큐잉시스템으로 데이터를 대규모로 수집하는 역할

## 클라우드 데이터 랩(Datalab)

- 과학자나 엔지니어가 여러 가지 데이터 소스에 접속해서 MS워드와 같은 환경에서 작업 내용을 저장하고 빅쿼리 등과 연결하여 바로 질의를 수행할 수 있는 웹기반의 저작도구이다.
- jupyter notebook의 구글 클라우드 버전이라고 생각하면 된다.

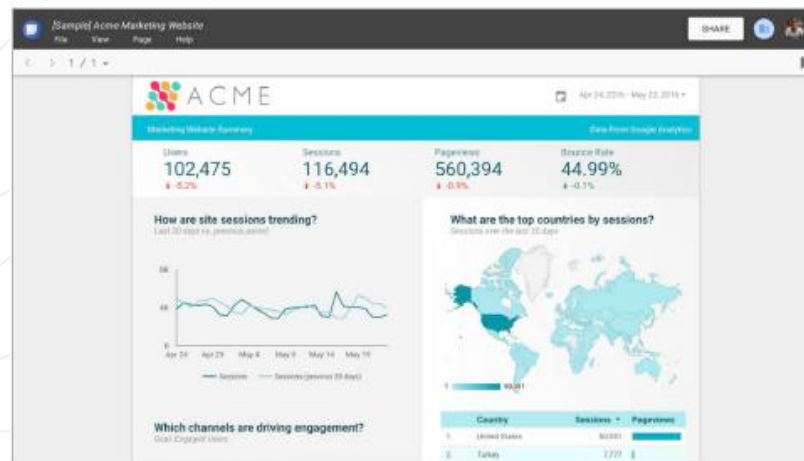


## • 데이터 프로크(Dataprocc)

- 오픈소스 빅데이터 플랫폼인 하둡(Hadoop)과 스파크(Spark)의 매니지드 서비스
- 클러스터를 90초안에 배포할 수 있고 분당 과금을 지원
  - ✓ 빠르고 편리하게 작업하면서 사용한 만큼만 정확하게 비용을 지불하여 비용 낭비를 줄여준다.

## • 데이터 스튜디오(Data Studio)

- SQL이나 빅 쿼리 등의 데이터 소스를 기반으로 시각적인 보고서를 생성해 주는 제품
- 비지오나 파워포인트처럼 간단하게 드래그-앤-드롭 만으로 다음그림과 같이 시각적인 보고서를 생성해낼 수 있다.



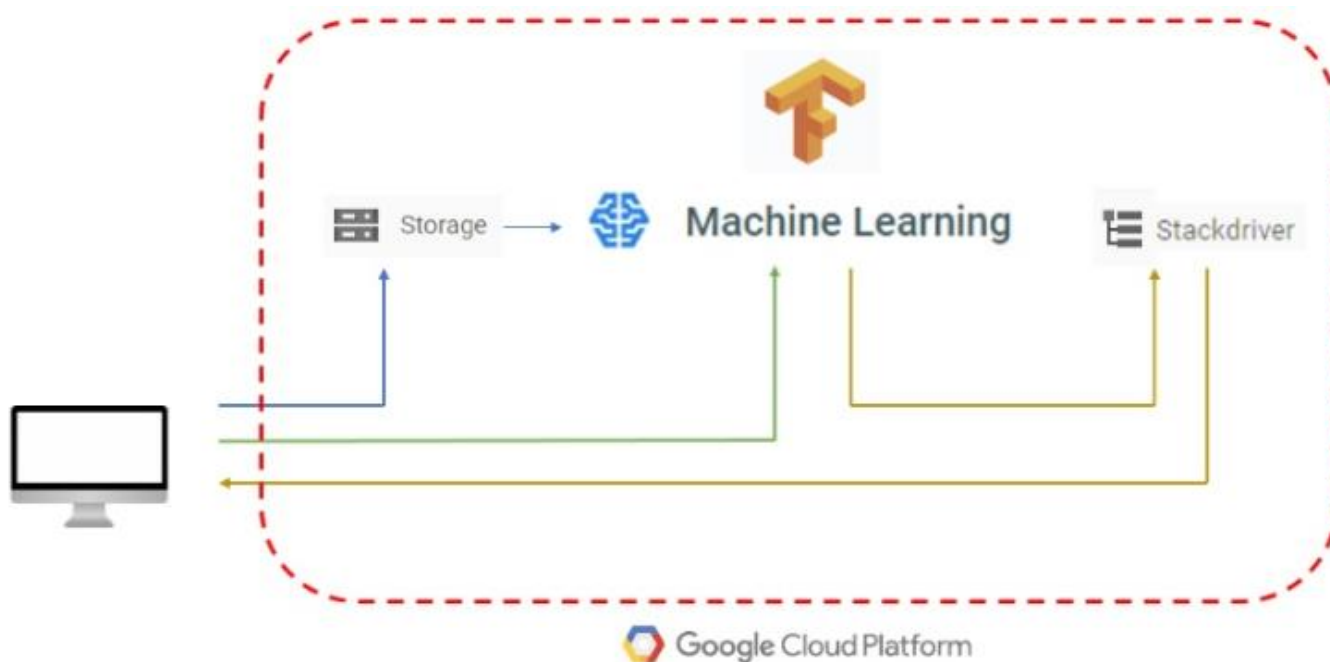
- 머신러닝 기반의 인공지능 기능들을 서비스로 제공
- 머신러닝 API들
  - 음성을 인식하여 텍스트로 변환해주는 스피치(Speech) API
  - 80여 개 나라의 언어를 읽어서 번역해주는 번역(Translate) API
  - 사진 이미지에서 사람의 표정과 사물을 인식해주는 비전(Vision) API
  - API 서비스는 머신 러닝 개념을 모르더라도 20~30 줄의 코드만으로 머신 러닝 기능을 사용할 수 있도록 해준다.





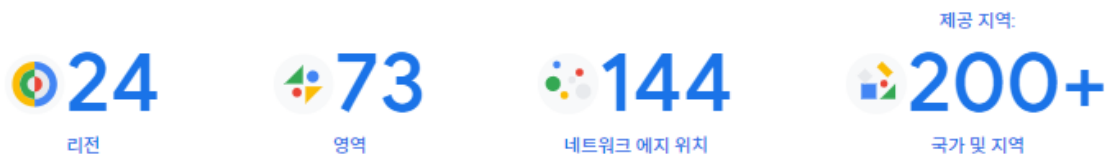
## • CloudML(Cloud Machine Learning) 서비스

- 텐스플로우(TensorFlow)를 지원하며 최신 CPU(Central Processing Unit), GPU(Graphics Processing Unit) 등의 환경을 선택적으로 사용가능
- 자신이 로컬에서 만든 머신러닝 모델을 클라우드에 던져서 학습시킬 수 있도록 하는 것
  - ✓ 로컬 PC는 모델을 클라우드에 던져놓고 자유로워지며 높은 사양도 필요 없음

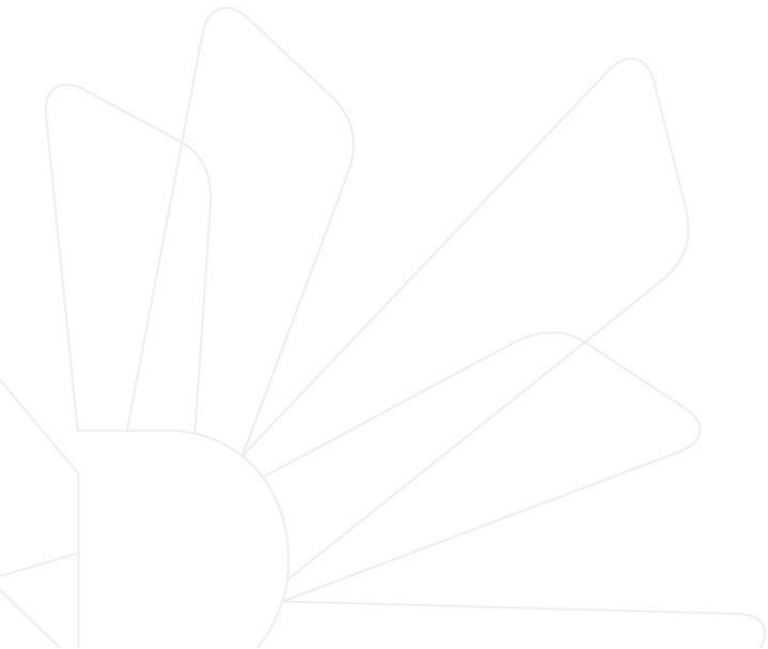


# 구글 전용 네트워크를 이용한 글로벌 커버리지

- 구글은 전 세계에 144개의 데이터 센터를 가지고 있으며, 미국, 유럽, 아시아에 배포된 클라우드 센터와 연결되어 있다.
  - 클라이언트에서 가장 가까운 센터로 접속한 후에 데이터 센터 사이를 연결하는 구글 전용망을 통해서 목표 데이터 센터로 연결된다.



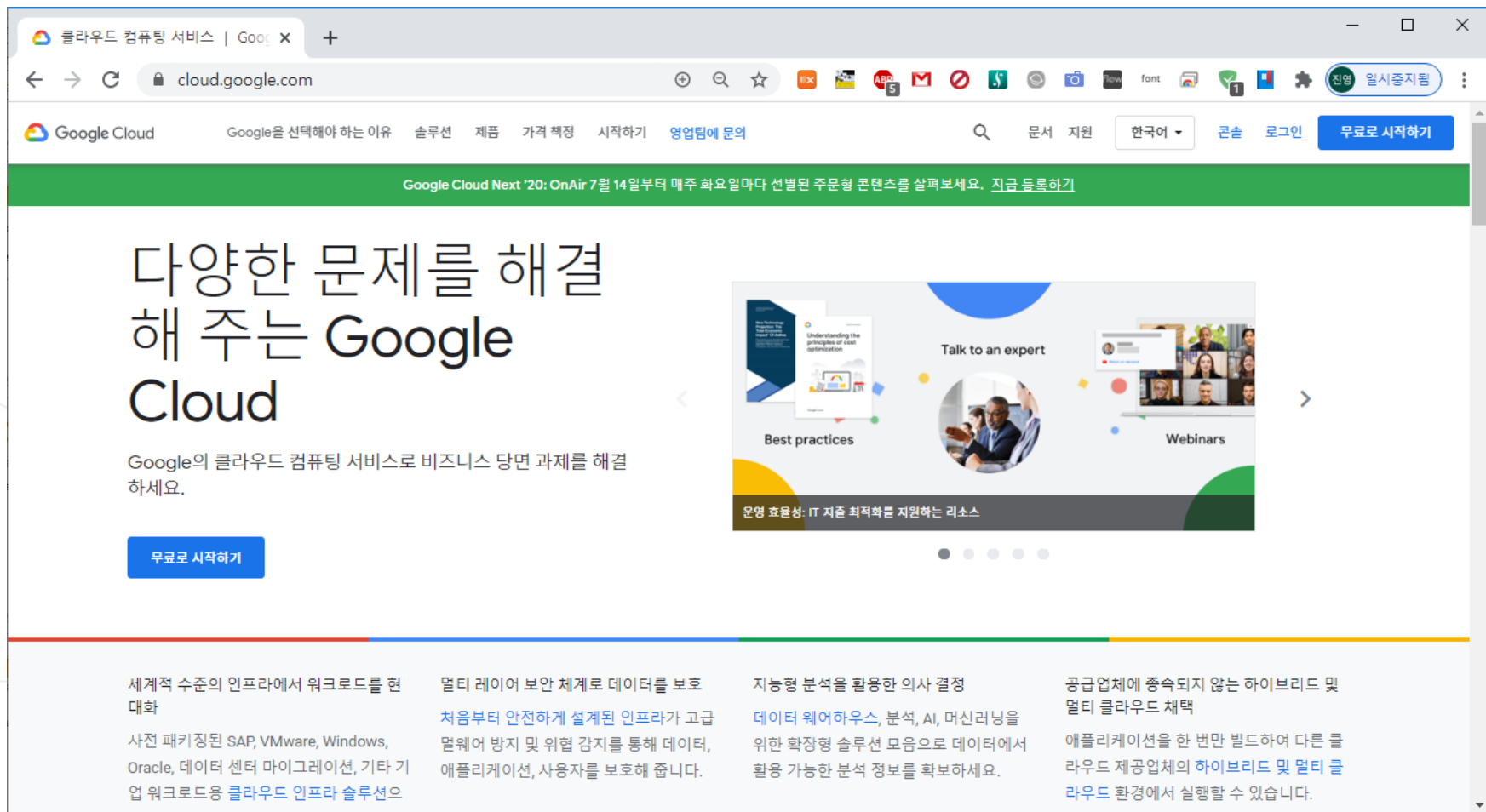
- 구글 클라우드는 타 클라우드보다 가격 정책이 합리적
  - 시간당 과금이 아니라 분당 과금을 한다든지
  - 연 단위 약정이 없더라도 월 단위로 사용량에 따라서 할인
  - 사용자가 직접 필요한 CPU와 메모리 양을 지정하여 자원낭비 없이 딱 필요한 만큼만 사용할 수 있다.



## 1. 계정가입

<https://cloud.google.com/>

<https://cloud.google.com/gcp/getting-started/?hl=ko>



클라우드 컴퓨팅 서비스 | Google x +

cloud.google.com

Google Cloud

Google을 선택해야 하는 이유 솔루션 제품 가격 책정 시작하기 영업팀에 문의

Google Cloud Next '20: OnAir 7월 14일부터 매주 화요일마다 선별된 주문형 콘텐츠를 살펴보세요. [지금 등록하기](#)

# 다양한 문제를 해결해 주는 Google Cloud

Google의 클라우드 컴퓨팅 서비스로 비즈니스 당면 과제를 해결하세요.

[무료로 시작하기](#)

Best practices

Talk to an expert

Webinars

운영 효율성: IT 지출 최적화를 지원하는 리소스

세계적 수준의 인프라에서 워크로드를 현대화

사건 패키징된 SAP, VMware, Windows, Oracle, 데이터 센터 마이그레이션, 기타 기업 워크로드용 [클라우드 인프라 솔루션](#)

멀티 레이어 보안 체계로 데이터를 보호

[처음부터 안전하게 설계된 인프라](#)가 고급 멀웨어 방지 및 위협 감지를 통해 데이터, 애플리케이션, 사용자를 보호해 줍니다.

지능형 분석을 활용한 의사 결정

[데이터 웨어하우스](#), 분석, AI, 머신러닝을 위한 확장형 솔루션 모음으로 데이터에서 활용 가능한 분석 정보를 확보하세요.

공급업체에 종속되지 않는 하이브리드 및 멀티 클라우드 채택

애플리케이션을 한 번만 빌드하여 다른 클라우드 제공업체의 [하이브리드 및 멀티 클라우드](#) 환경에서 실행할 수 있습니다.

2/2단계 - 무료 평가판 - Google x +

console.cloud.google.com/freetrial/signup/billing/KR?\_ga=2.15328...

Google Cloud Platform 무료 평가판 소개

## 2/2단계

고객 정보

계정 유형: 개인

세금 정보: 사업자 유형: 개인

이름 및 주소

**모든 Cloud Platform 제품에 액세스**

Firebase, Google Maps API 등을 포함해 앱, 웹사이트, 서비스를 구축하고 실행하는 데 필요한 모든 기능을 이용할 수 있습니다.

**\$300의 무료 크레딧**

가입하여 Google Cloud Platform에서 12개월간 사용할 수 있는 \$300 크레딧을 받아 보세요.

**무료 체험판 종료 후 자동 청구되지 않음**

신용카드를 요청하는 이유는 자동 가입을 방지하기 위해서입니다. 유료 계정으로 직접 업그레이드하지 않는 한 요금이 청구되지 않습니다.

계속 진행하면 Google이 귀하의 결제 정보를 이 계정에 연결하고 Google 제품 내에서 사용 가능한 모든 기능을 사용할 수 있도록 Google 개인정보처리방침

The screenshot shows the Google Cloud Platform console interface. The browser address bar displays the URL: `console.cloud.google.com/getting-started?project=promising-env...`. The page title is "시작하기 - My First Project - Google Cloud Platform". The main heading is "jinyeong님, 환영합니다" (Welcome, jinyeong님) with the subtitle "Google Cloud Platform 시작하기".

The page is divided into several sections:

- 기본사항으로 시작하기** (Get started with basics): A section with a blue button labeled "체크리스트로 이동" (Go to checklist) and a list of tasks: "결제, 크레딧, 프로젝트 검토", "제품 및 API 찾기", "프로젝트에 리소스 추가", and "가격 책정 이해 및 계산".
- 인기 제품** (Popular products): A section with a button labeled "모두 보기" (View all).
- 컴퓨팅 제품** (Compute products): A section featuring "Compute Engine" (Made by Google) with a button "COMPUTE ENGINE으로 이동" (Go to Compute Engine), "Kubernetes Engine" with a link "Kubernetes Engine", and "App Engine" with a link "App Engine". It also mentions "Cloud Functions" and "Functions".
- 스토리지 및 데이터베이스 제품** (Storage and database products): A section with a button labeled "모두 보기" (View all).

The left sidebar contains a navigation menu with categories like "홈" (Home), "Marketplace", "결제" (Billing), "API 및 서비스" (APIs & Services), "지원" (Support), "IAM 및 관리자" (IAM & Admin), "시작하기" (Get started), "보안" (Security), "Anthos", "reCAPTCHA Enterprise", "컴퓨팅" (Compute), "저장소" (Storage), and "데이터베이스" (Database). The "시작하기" category is currently selected.

## 2. 프로젝트 생성

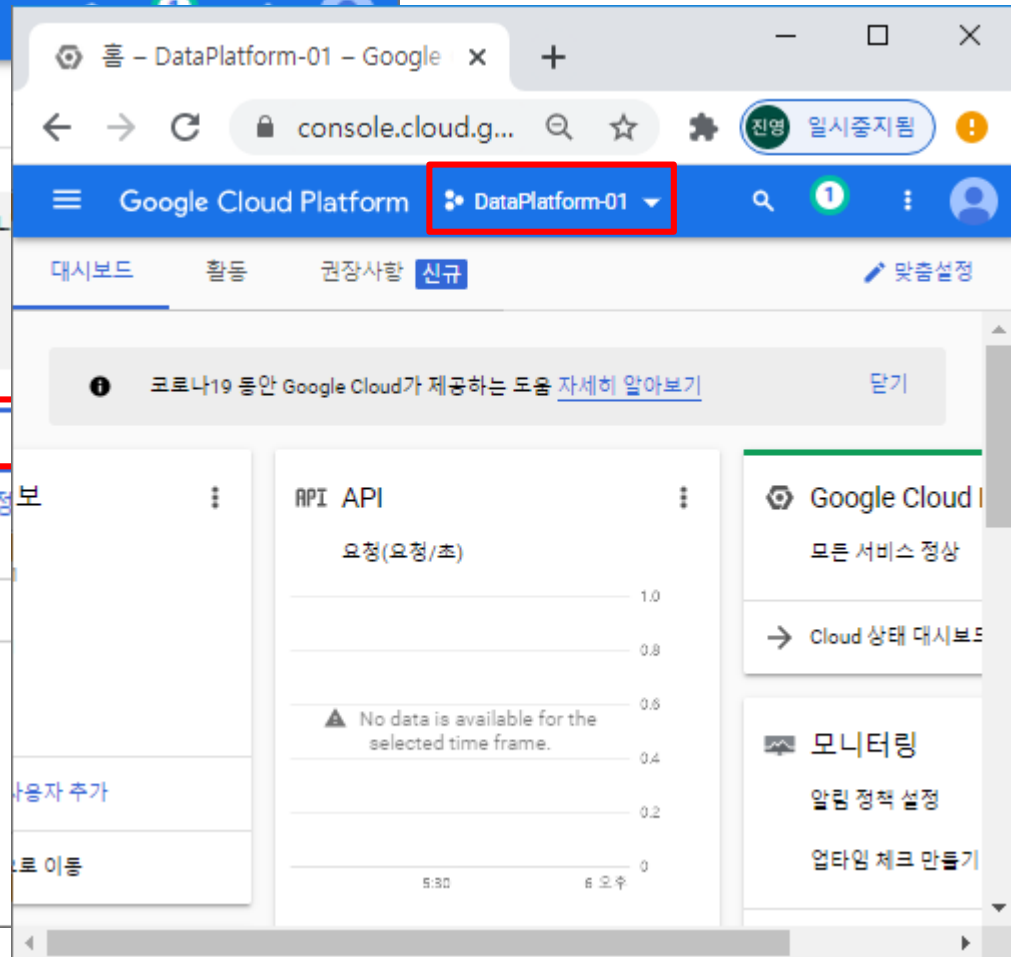
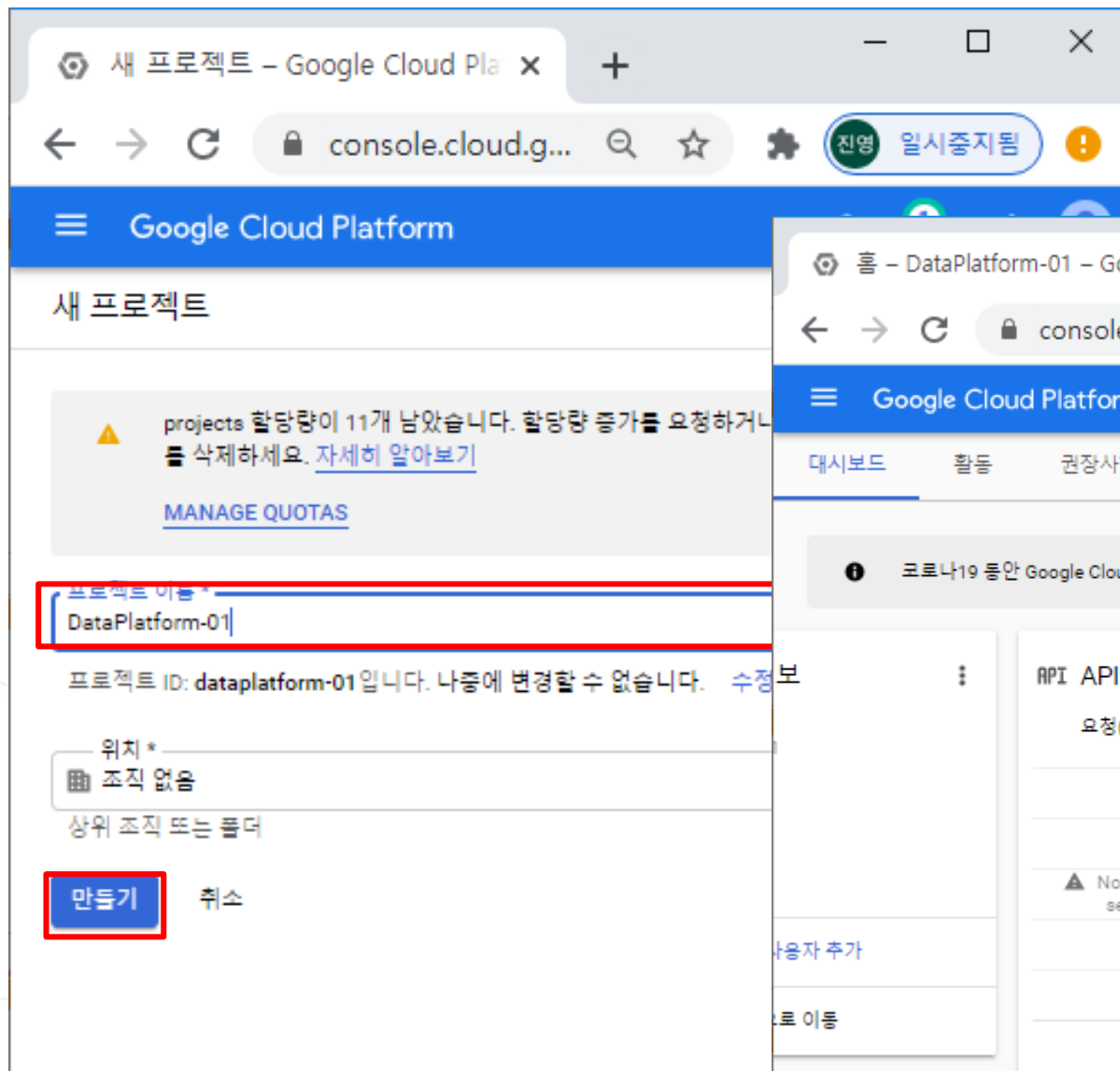
The screenshot shows the Google Cloud Platform console interface. A red box highlights the 'My First Project' dropdown in the top navigation bar. Another red box highlights the '프로젝트 선택' (Project Selection) dialog box. The dialog box has a title bar '프로젝트 선택' and a search bar '프로젝트 및 폴더 검색'. Below the search bar are tabs for '최근' (Recent) and '전체' (All). A table lists the projects:

이름	ID
✓ My First Project ?	promising-env-282408

At the bottom right of the dialog box, there are buttons for '취소' (Cancel) and '열기' (Open). A '새 프로젝트' (New Project) button is also visible in the top right corner of the dialog box.

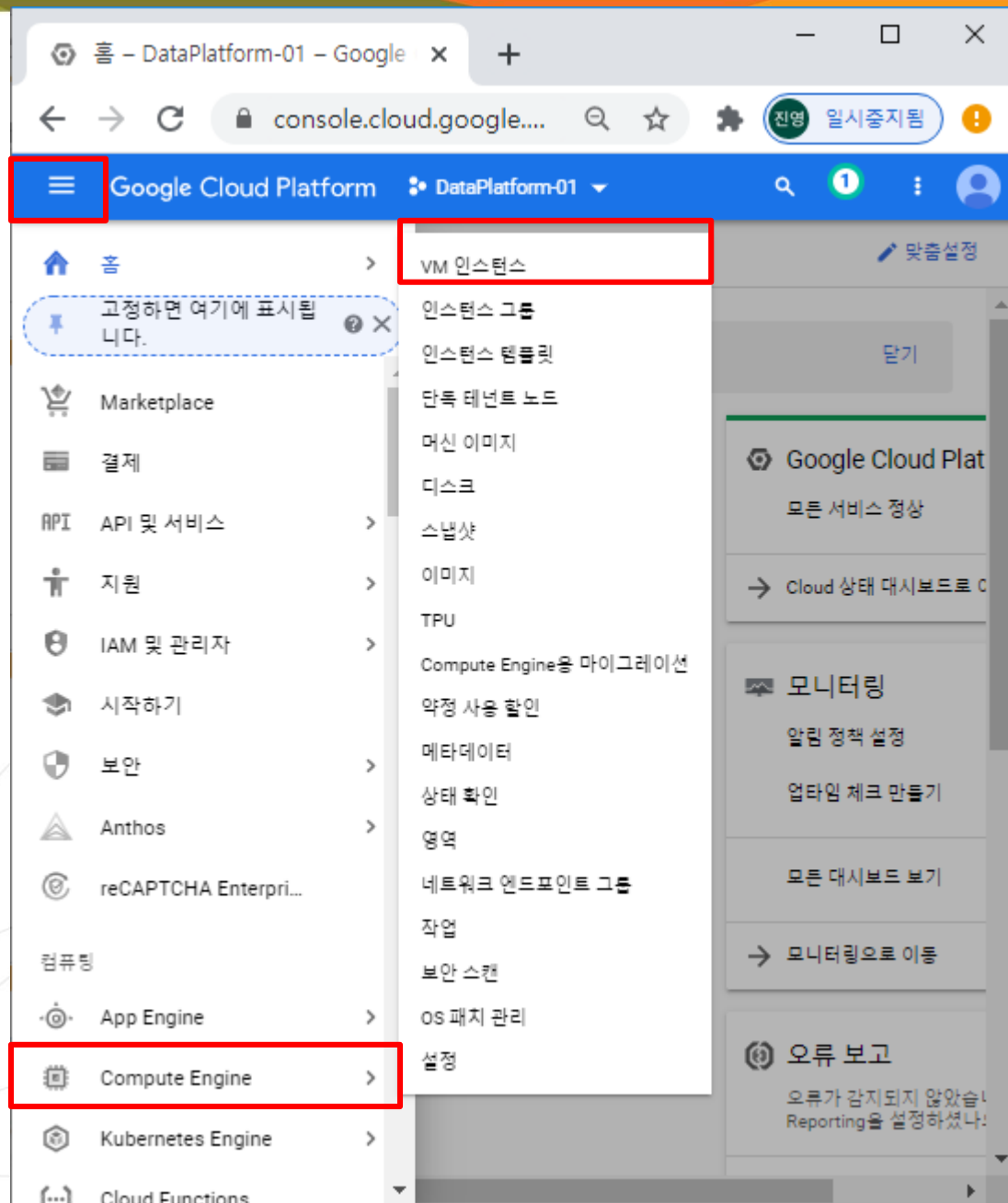
# 구글 클라우드에서 VM 생성하기(5)

16





## 3. VM 생성



# 구글 클라우드에서 VM 생성

Compute Engine - DataPlatform x +

← → ↺ console.cloud.google.... 🔍 ☆

Google Cloud Platform DataPlatform-01

Compute Engine VM 인스턴스

Compute Engine을(를) 준비하는 중이며 1분 이상 걸릴 수 있습니다. [Compute Engine 문서](#)

Compute Engine VM 인스턴스

Compute Engine을 통해 Google의 인프라에서 실행되는 가상 머신을 사용할 수 있습니다. 마이크로 VM은 빠른 Debian, Windows 또는 다른 표준 이미지를 실행하는 대형 인스턴스를 만들 수 있습니다. 첫 번째 VM 인스턴스를 만들거나, 이전 서비스를 사용하여 가져오거나, 빠른 시작을 사용하여 샘플 앱을 제작해 보세요.

만들기

 또는 

가져오기

 또는 

빠른 시작 사용

Compute Engine - DataPlatform x +

← → ↺ console.cloud.google.c... 🔍 ☆

Google Cloud Platform DataPlatform-01 제품 및 리소스

← 인스턴스 만들기

VM 인스턴스를 만들려면 옵션 중 하나를 선택하세요.

새 VM 인스턴스

VM 인스턴스 하나를 처음부터 만듭니다.

템플릿에서 VM 인스턴스 만들기

기존 템플릿에서 VM 인스턴스 하나를 만듭니다.

머신 이미지의 새 VM 인스턴스

기존 머신 이미지에서 VM 인스턴스 하나를 만듭니다.

Marketplace

VM 인스턴스에 바로 사용할 수 있는 솔루션을 배포합니다.

이름은 영구적입니다.

ubuntu-instance

라벨 (선택사항)

+ 라벨 추가

지역은 영구적입니다. asia-northeast3(서울)

영구적입니다. asia-northeast3-a

머신 구성

머신 계열

일반 용도

일반적인 작업 부하에 적합한 머신 유형이며 가격 및 유연성을 위해 최적화되었습니다.

시리즈

N1

Intel Skylake CPU 플랫폼 또는 이전 버전의 플랫폼에서 제공

머신 유형

n1-standard-1(vCPU 1개, 3.75GB 메모리)

vCPU

메모리

1

3.75GB

CPU 플랫폼 및 GPU

컨테이너

☐ 이 VM 인스턴스에 컨테이너 이미지를 배포합니다. 자세히 알아보기

부팅 디스크

새로운 10GB 표준 영구 디스크 이미지

Ubuntu 16.04 LTS

변경

ID 및 API 액세스

서비스 계정

Compute Engine default service account

액세스 범위

☒ 기본 액세스 허용

☒ 모든 Cloud API에 대한 전체 액세스 허용

☐ 각 API에 액세스 허용

필요한 네트워크 규칙을 추가하여 인터넷에서 특정 네트워크 트래픽 허용합니다.

☒ HTTP 트래픽 허용

☒ HTTPS 트래픽 허용

\* 관리, 보안, 디스크, 네트워킹, 단독 임대

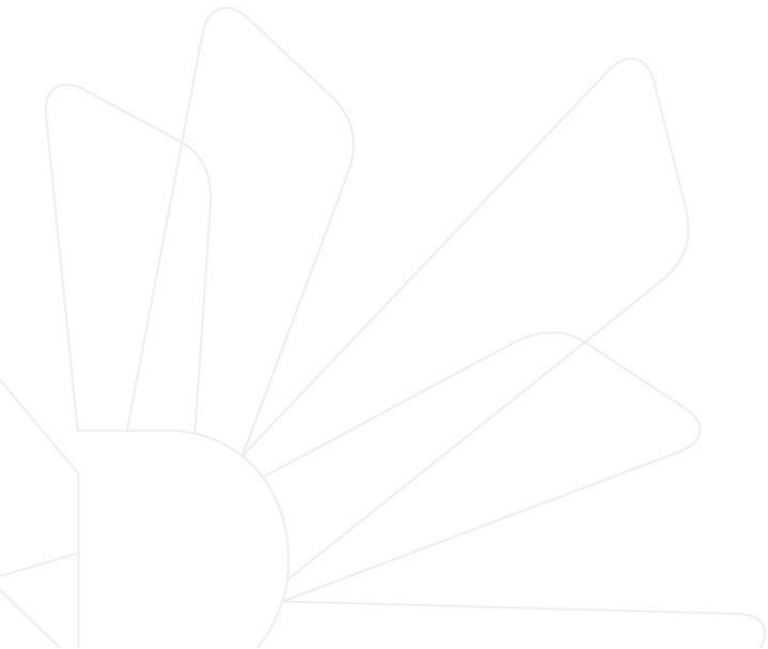
이 VM 인스턴스에 무료 평가판 크레딧이 사용됩니다. GCP 무료 등급

만들기

취소

# 구글 클라우드에서 VM 생성하기(8)

- 인스턴스 이름 : ubuntu-instance
- 영역 : asia-northeast3-a(서울)
- 머신 유형 : n-standard-1(vCPU 1개, 3.75GB 메모리)
- 부팅 디스크 : Ubuntu 16.04 LTS, 표준 영구 디스크 10GB
- ID 및 API 액세스 → 액세스 범위 : 모든 Cloud API에 대한 전체 액세스 허용
- 방화벽 : HTTP/HTTPS 트래픽 허용



# 구글 클라우드에서 VM 생성하기(9)

20

Compute Engine - DataPlatform x +

console.cloud.google.c...

Google Cloud Platform DataPlatform-01

Compute Engine VM 인스턴스 정보 패널 표시

프로젝트 시작하기

이제 VM 인스턴스를 만들었으므로, 활용 방법을 알아봅니다.

[내 인스턴스에 연결](#)

[파일 이전](#)

기존 VM 솔루션 찾기

[Marketplace 탐색](#)

2단계 웹 앱 빌드 방법 빨리 배우기

[방법 보기](#)

안내 가이드

[웹사이트 호스트 이름 설정](#)

[정적 외부 IP 주소 예약](#)

[Windows 인스턴스 만들기 및 관리](#)

도구 및 API 탐색

[Cloud Shell CLI에 대해 알아보기](#)

[클라이언트 라이브러리들 통해 연결](#)

작업 부하 가져오기 또는 이전하기

[Velostrata를 사용하여 VM 이전하기](#)

[가상 디스크 가져오기](#)

가격 책정

[가격 책정 이해](#)

VM 인스턴스 필터링

☐ 이름 ☐ 영역 ☐ 권장사항 ☐ 다음에서 사

☒ ubuntu-instance asia-northeast3-a

관련 작업

[결제 보고서 보기](#)  
Compute Engine  
결제 내역을 확인하고  
관리합니다.

[Stackdriver 로그 모니터링](#)  
VM 인스턴스 로그를 보고  
검색하고 분석하여 다른  
인스턴스도 볼 수 있습니다.

## 4. 브라우저를 통한 SSH 접속

The screenshot shows the Google Cloud Platform console on the left and a terminal window on the right. The console displays the 'VM 인스턴스' (VM Instances) page for the 'DataPlatform-01' project. A list of instances is shown, including 'ubuntu-instance' in the 'asia-northeast3-a' zone. The terminal window shows the SSH connection process, displaying the host fingerprint and the Ubuntu 16.04.6 LTS login prompt. The user 'jinyeongum' is logged in at the 'ubuntu-instance'.

Google Cloud Platform console (left):

- Project: DataPlatform-01
- Section: VM 인스턴스
- Instances list: ubuntu-instance (Status: Running, Zone: asia-northeast3-a)

Terminal window (right):

```
jinyeongum@ubuntu-instance: ~ - Chrome
ssh.cloud.google.com/projects/dataplatform-01/zones/asia-northeast3-a/instances/u...
Connected, host fingerprint: ssh-rsa 0 D6:5E:C9:CB:F7:0F:38:B2:A9:71:C8:67:97:9D:55:99:9D:35:52:93:D0:DC:1A:2E:FA:F9:FB:78:1A:D0:77
Welcome to Ubuntu 16.04.6 LTS (GNU/Linux 4.15.0-1078-gcp x86_64)

* Documentation:  https://help.ubuntu.com
* Management:    https://landscape.canonical.com
* Support:        https://ubuntu.com/advantage

0 packages can be updated.
0 updates are security updates.

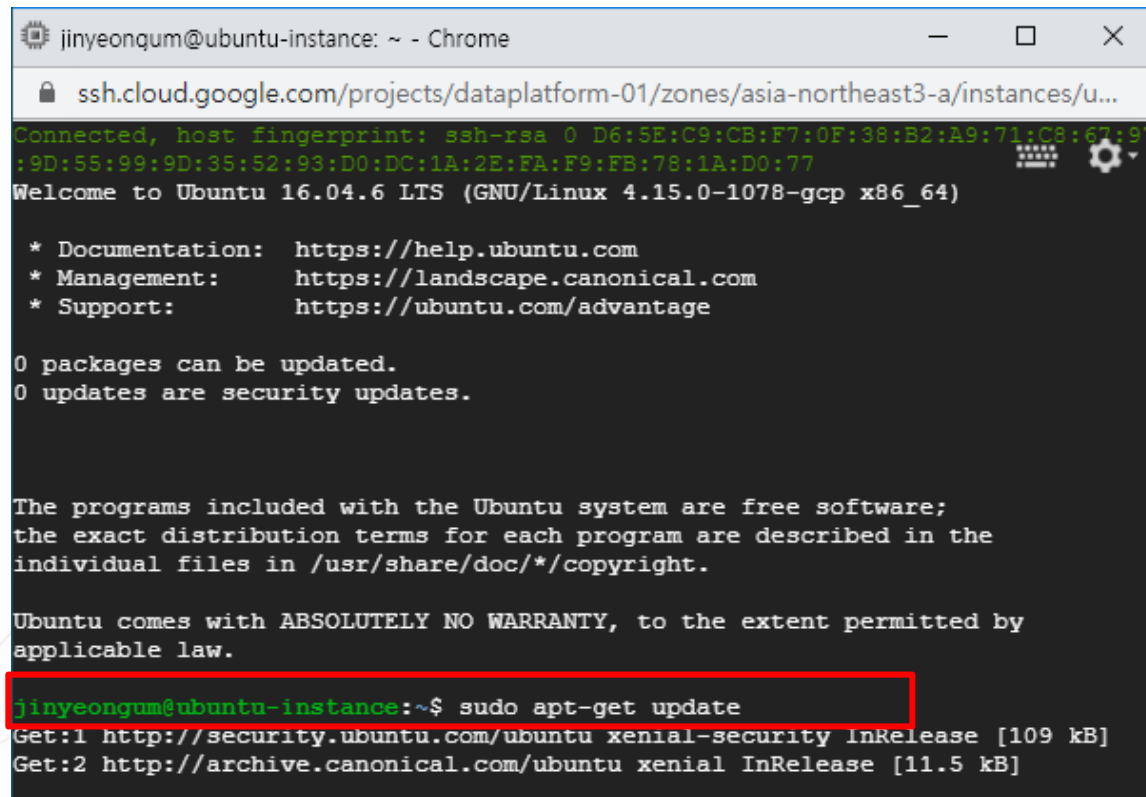
The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

jinyeongum@ubuntu-instance:~$
```

## 5. 아파치 웹 서버 설치

- `sudo apt-get update`
  - ✓ 설치에는 우분투의 apt-get 유틸리티를 사용하기 위해 apt-get 저장소를 최신으로 업데이트



```
jinyeonqum@ubuntu-instance: ~ - Chrome
ssh.cloud.google.com/projects/dataplatform-01/zones/asia-northeast3-a/instances/u...
Connected, host fingerprint: ssh-rsa 0 D6:5E:C9:CB:F7:0F:38:B2:A9:71:C8:67:97
:9D:55:99:9D:35:52:93:D0:DC:1A:2E:FA:F9:FB:78:1A:D0:77
Welcome to Ubuntu 16.04.6 LTS (GNU/Linux 4.15.0-1078-gcp x86_64)

* Documentation:  https://help.ubuntu.com
* Management:    https://landscape.canonical.com
* Support:       https://ubuntu.com/advantage

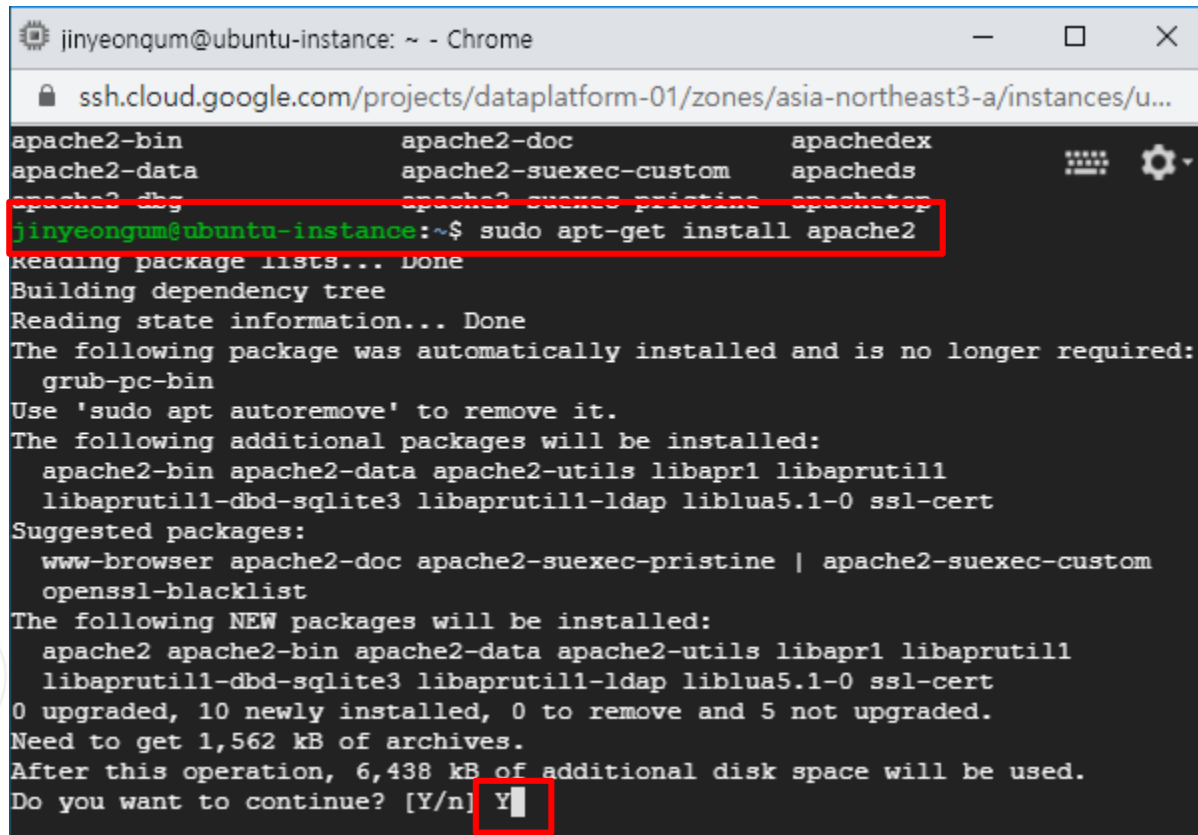
0 packages can be updated.
0 updates are security updates.

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

jinyeonqum@ubuntu-instance:~$ sudo apt-get update
Get:1 http://security.ubuntu.com/ubuntu xenial-security InRelease [109 kB]
Get:2 http://archive.canonical.com/ubuntu xenial InRelease [11.5 kB]
```

- sudo apt-get install apache2



The screenshot shows a terminal window titled 'jinyeonqum@ubuntu-instance: ~ - Chrome'. The browser address bar shows 'ssh.cloud.google.com/projects/dataplatform-01/zones/asia-northeast3-a/instances/u...'. The terminal output lists available Apache packages: apache2-bin, apache2-doc, apachedex, apache2-data, apache2-suexec-custom, apacheds, apache2-dbg, apache2-suexec-pristine, and apachetop. The command 'jinyeonqum@ubuntu-instance:~\$ sudo apt-get install apache2' is entered and highlighted with a red box. The output shows the package lists are read, a dependency tree is built, and state information is read. It notes that 'grub-pc-bin' was automatically installed and is no longer required. The packages to be installed are listed: apache2-bin, apache2-data, apache2-utils, libapr1, libaprutil1, libaprutil1-dbd-sqlite3, libaprutil1-ldap, liblua5.1-0, and ssl-cert. Suggested packages include www-browser, apache2-doc, apache2-suexec-pristine, apache2-suexec-custom, and openssl-blacklist. The new packages to be installed are listed again. The summary shows 0 upgraded, 10 newly installed, 0 to remove, and 5 not upgraded. The total size of archives to be downloaded is 1,562 kB, and the additional disk space required is 6,438 kB. The prompt 'Do you want to continue? [Y/n]' is shown, with 'Y' entered and highlighted by a red box.

```
jinyeonqum@ubuntu-instance:~$ sudo apt-get install apache2
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following package was automatically installed and is no longer required:
  grub-pc-bin
Use 'sudo apt autoremove' to remove it.
The following additional packages will be installed:
  apache2-bin apache2-data apache2-utils libapr1 libaprutil1
  libaprutil1-dbd-sqlite3 libaprutil1-ldap liblua5.1-0 ssl-cert
Suggested packages:
  www-browser apache2-doc apache2-suexec-pristine | apache2-suexec-custom
  openssl-blacklist
The following NEW packages will be installed:
  apache2 apache2-bin apache2-data apache2-utils libapr1 libaprutil1
  libaprutil1-dbd-sqlite3 libaprutil1-ldap liblua5.1-0 ssl-cert
0 upgraded, 10 newly installed, 0 to remove and 5 not upgraded.
Need to get 1,562 kB of archives.
After this operation, 6,438 kB of additional disk space will be used.
Do you want to continue? [Y/n] Y
```

- 설치 후 exist 명령

- 아파치 웹서버 접속을 위해 외부 IP로 접속

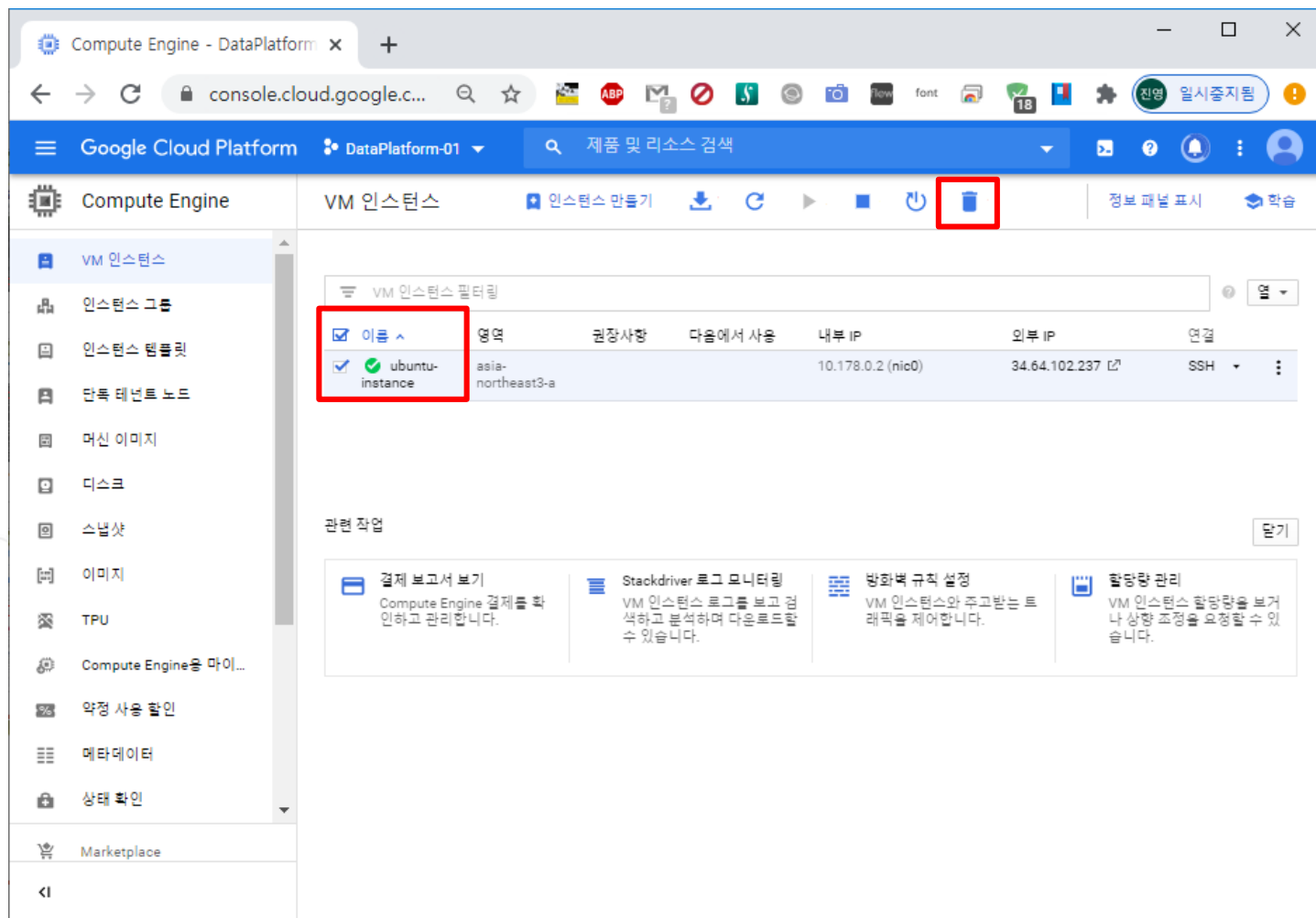
The screenshot shows two overlapping windows. The background window is the Google Cloud Platform console, displaying the 'Compute Engine' section with a list of VM instances. A red box highlights the '외부 IP' (External IP) column, showing the IP address '34.64.102.237'. The foreground window is a web browser displaying the 'Apache2 Ubuntu Default Page' at the IP address '34.64.102.237'. The page content includes the Ubuntu logo, the text 'It works!', and a 'Configuration Overview' section. The configuration overview mentions that the configuration is fully documented in `/usr/share/doc/apache2/README.Debian.gz`.

웹 서버가 자동으로 가동되지 않았다면 다음 명령을 내려 직접 가동  
\$ sudo apachectl start



# 구글 클라우드에서 VM 삭제하기

25



The screenshot shows the Google Cloud Platform console interface. The left sidebar contains the navigation menu with 'Compute Engine' selected. The main content area displays the 'VM instances' page. A red box highlights the 'Delete' button (trash icon) in the top right corner of the VM instances table. Another red box highlights the 'VM instances' table header and the first row, which contains the instance 'ubuntu-instance'.

이름	영역	권장사항	다음에서 사용	내부 IP	외부 IP	연결
ubuntu-instance	asia-northeast3-a			10.178.0.2 (nic0)	34.64.102.237	SSH

관련 작업

- 결제 보고서 보기: Compute Engine 결제를 확인하고 관리합니다.
- Stackdriver 로그 모니터링: VM 인스턴스 로그를 보고 검색하고 분석하며 다음으로 활용할 수 있습니다.
- 장착된 규칙 설정: VM 인스턴스와 주크립 트래픽을 제어합니다.
- 할당량 관리: VM 인스턴스 할당량을 보거나 설정을 조정할 수 있습니다.