

# R 프로그래밍 중급으로 나아가기 위한 Level UP Skill

Step2. 하나의 문법으로 다양한 소스의 데이터를 다뤄보기

[https://mrchypark.github.io/dabrp\\_classnote3/class2](https://mrchypark.github.io/dabrp_classnote3/class2)

[[pdf다운로드](#)] [[문의하기](#)] [[피드백하기](#)]

박찬엽

2017년 9월 21일

# 목차

1. 과제 질답
2. 수업의 목표
3. 데이터 공유와 탐색
  - 데이터를 공유하는 방법
  - 실습에 사용할 데이터 소개
  - 실습
4. 데이터 베이스와 R의 연결
  - DBI 소개와 SQLite
  - DBI 함수의 동작
  - 실습
5. 본격 DB와 R의 연결
  - 클라우드 서비스 이용하기

# 과제 질답

# 수업의 목표

1. 데이터를 잘 공유하는 법을 알고 설명할 수 있다.
2. DBI 패키지의 구조를 이해하고 설명할 수 있다.
3. 클라우드 서비스에 테이블을 저장하고 R로 연동할 수 있다.

# 데이터를 공유하기

# 데이터 공유

Leek group에서 소개하는 데이터 공유 가이드

- 원시 데이터(raw data) : 최초에 수집된 데이터를 뜻하며 센서나 인터넷 서비스의 로그 데이터나 사용자가 직접 작성한 설문 데이터 등을 의미함
- 정제후 데이터(tidy data) : 사용하기 좋은 상태로 정리한 데이터를 뜻하며 보통 설문조사에서는 코딩한다고 표현함
- 코드북(code book) : 공개하는 데이터에 대해서 설명하는 문서로 단위나 각 변수명이 의미하는 바 등을 작성함
- 자료형(data type) : 변수의 종류와 저장법에 대해서 조언하며 엑셀등 특별한 프로그램에 종속적인 기능은 사용하지 말 것을 권고함
- 재현성(reproducibility) : 다시 결과를 만들 수 있는 것을 뜻하며 2010년대 이후 과학적 방법의 중요한 축

# 원시 데이터

최초 획득한 당시 그대로의 데이터

- 어떤 식으로든 수정을 가하지 않은 상태
- 수정을 가하는 과정을 함께 기록함으로써 신뢰성 확보
- 위 두 가지가 없는 경우 상황을 상상해야 함

# 정제후 데이터

해들리 위컴이 설명한 **tidy data**의 요건에 맞게 가공하여 데이터를 쉽게 다룰 수 있게 만든 상태

- 측정하는 각 변수는 하나의 열에 있어야 함
- 측정하는 각 관찰은 하나의 행에 있어야 함
- 각 종류의 변수에 대해 각 하나의 테이블이 있어야 함
- 여러 개의 테이블이 있는 경우 테이블에 합치기 위한 기준 열을 포함해야 함

# 코드북

데이터셋에 대해 필요한 설명을 담은 문서

- 정제후 데이터에 대해 추가적으로 필요한 설명이나 정보(단위 등)
- 정제 과정에서 사용한 방법의 설명과 사용한 이유
- 데이터가 사용된 분석에 대한 정보

# 데이터 원본

확보할 당시의 원시 데이터나, 항상 최신 상태를 유지하여 신뢰할 수 있는 데이터. 기계적 생성 데이터가 아닌 사람이 원본이 되므로 작성된 내용이 원본이라는 합의를 함으로써 데이터 관리의 체계를 세울 수 있음

- 커뮤니케이션 비용 감소
- 의사결정 및 활동의 기준
- 가공된 데이터의 신뢰성 확보

# 데이터 훑어보기

- `head()`: 최초 6행의 데이터를 보여줌(행갯수 조절 가능)
- `tail()`: 마지막 6행의 데이터를 보여줌(행갯수 조절 가능)
- `summary()`: 각 컬럼의 자료형과 숫자라면 대표값을 함께 보여줌
- `str()`: 각 컬럼의 자료형과 초기 값을 보여줌
- `length()`: 데이터의 길이 출력(vector)
- `nrow()`: 행 갯수 출력(data.frame)
- `is.na()`: NA 인지 확인
- `complete.cases()`: 값이 모두 있는지 행단위로 검사
- `tibble()`: 최근 기법으로 재구성한 tidyverse 패키지 군에서 사용하는 data.frame의 일종
- `data.table()`: DB에서 사용하는 기법 등과 메모리 효율성을 극대화한 data.frame의 일종

# 데이터 소개

# 실습 데이터 설정하기

```
# data.table 패키지가 있는지 확인하고 없으면 설치합니다.  
if (!requireNamespace("data.table")) install.packages("data.table")  
# data.table 패키지를 사용할 수 있게 불러옵니다.  
library(data.table)  
  
# tran.csv 파일이 data/recomen 폴더에 있는지 확인 합니다.  
chk<-dir("./data/recomen", pattern = "tran.csv")  
# 없으면 다운로드합니다.  
if(identical(chk, character(0))){  
  recoment<-http://rcoholic.duckdns.org/oc/index.php/s/jISrPutj4ocLci2/download  
  download.file(recoment, destfile="./data/recomen/tran.csv", mode='wb')  
}  
# fread 함수를 이용해서 csv 파일을 R 객체로 불러옵니다.  
chennel<-fread("./data/recomen/chennel.csv")  
competitor<-fread("./data/recomen/competitor.csv")  
customer<-fread("./data/recomen/customer.csv")  
item<-fread("./data/recomen/item.csv")  
membership<-fread("./data/recomen/membership.csv")  
tran<-fread("./data/recomen/tran.csv")
```

# 추천 데이터 - chennel

- cusID : 5자리 숫자조합으로 구성된 고객ID
- chennel: 접속 체널
- useCnt: 사용횟수(건)

```
str(chennel, max.level=1)
```

```
## Classes 'data.table' and 'data.frame': 8824 obs. of 3 variables:  
## $ cusID : int 7 14 42 74 74 94 112 112 122 123 ...  
## $ chennel: chr "A_MOBILE/APP" "A_MOBILE/APP" "B_MOBILE/APP" "A_MOBILE/APP" ...  
## $ useCnt : int 4 1 23 1 30 14 16 1 27 10 ...  
## - attr(*, ".internal.selfref")=<externalptr>
```

# 추천 데이터 - competitor

- cusID : 5자리 숫자조합으로 구성된 고객ID
- partner : 제휴사
- competitor: 경쟁사
- useDate : 이용년월(YYYYDD)

```
str(competitor, max.level=1)
```

```
## Classes 'data.table' and 'data.frame': 28159 obs. of 4 variables:
## $ cusID     : int  2 51 77 77 77 77 77 88 88 110 ...
## $ partner   : chr  "D" "D" "D" "D" ...
## $ competitor: chr  "D02" "D01" "D02" "D02" ...
## $ useDate   : int  201507 201504 201503 201506 201507 201508 201511 201510 201511 201508 ...
## - attr(*, ".internal.selfref")=<externalptr>
```

# 추천 데이터 - customer

- cusID: 5자리 숫자조합으로 구성된 고객ID
- sex: 성별
- age: 연령
- area: 거주지역

```
str(customer, max.level=1)
```

```
## Classes 'data.table' and 'data.frame': 19383 obs. of 4 variables:
## $ cusID: int 1 2 3 4 5 6 7 8 9 10 ...
## $ sex : chr "M" "M" "M" "F" ...
## $ age : chr "60세이상" "60세이상" "60세이상" "60세이상" ...
## $ area : int 60 100 33 16 100 240 36 10 100 24 ...
## - attr(*, ".internal.selfref")=<externalptr>
```

# 추천 데이터 - item

- partner: 재휴사
- cate\_1: 대분류
- cate\_2: 중분류
- cate\_3: 소분류
- cate\_2\_name: 중분류명
- cate\_3\_name: 소분류명

```
str(item, max.level=1)
```

```
## Classes 'data.table' and 'data.frame': 4386 obs. of 6 variables:  
## $ partner : chr "A" "A" "A" "A" ...  
## $ cate_1  : int 1 1 1 1 1 1 1 1 1 ...  
## $ cate_2  : int 101 101 101 101 101 101 101 102 102 102 ...  
## $ cate_3  : chr "A010101" "A010102" "A010103" "A010104" ...  
## $ cate_2_name: chr "일용잡화" "일용잡화" "일용잡화" "일용잡화" ...  
## $ cate_3_name: chr "위생세제" "휴지류" "뷰티상품" "일용잡화" ...  
## - attr(*, ".internal.selfref")=<externalptr>
```

# 추천 데이터 - membership

- cusID : 5자리 숫자조합으로 구성된 고객ID
- memberShip: 멤버십명
- regDate : 가입년월

```
str(membership, max.level=1)
```

```
## Classes 'data.table' and 'data.frame':    7456 obs. of  3 variables:
## $ cusID      : int  11 21 37 43 44 61 65 69 76 79 ...
## $ memberShip: chr  "하이마트" "하이마트" "하이마트" "하이마트" ...
## $ regDate    : int  201512 201506 201306 201403 201411 201312 201506 201404 201406 201311 ...
## - attr(*, ".internal.selfref")=<externalptr>
```

# 추천 데이터 - tran

```
summary(tran)
```

```
##   partner           receiptNum        cate_1        cate_2
## Length:28593030      Min.    : 1       Min.    : 1.00     Min.    : 101
## Class  :character   1st Qu.: 3922474  1st Qu.: 4.00     1st Qu.: 401
## Mode   :character   Median  : 7167787  Median  :11.00     Median  :1102
##                   Mean   : 6447881  Mean   :18.37     Mean   :1840
##                   3rd Qu.: 9116336  3rd Qu.:18.00     3rd Qu.:1808
##                   Max.   :11096601  Max.   :92.00     Max.   :9206
##   cate_3            cusID          storeCode       date
## Length:28593030      Min.    : 1       Min.    : 1.00     Min.    :20140101
## Class  :character   1st Qu.: 5206    1st Qu.: 16.00    1st Qu.:20140711
## Mode   :character   Median  :10104    Median  : 44.00    Median  :20150110
##                   Mean   : 9904    Mean   : 92.26    Mean   :20145817
##                   3rd Qu.:14638    3rd Qu.:110.00   3rd Qu.:20150703
##                   Max.   :19383    Max.   :593.00    Max.   :20151231
##   time              amount
## Min.   : 0.00      Min.   : 1
## 1st Qu.:14.00      1st Qu.: 2050
## Median :17.00      Median : 4290
## Mean   :16.71      Mean   : 23678
## 3rd Qu.:19.00      3rd Qu.: 9900
## Max.   :23.00      Max.   :101330000
```

# 추천 데이터 - tran

```
str(tran, max.level=1)
```

```
## Classes 'data.table' and 'data.frame': 28593030 obs. of 10 variables:
## $ partner : chr "B" "B" "B" "B" ...
## $ receiptNum: int 8664000 8664000 8664000 8664000 8664001 8664001 8664001 8664002 8664002 8664002 ...
## $ cate_1   : int 15 16 16 18 5 15 10 43 54 5 ...
## $ cate_2   : int 1504 1601 1602 1803 509 1501 1003 4301 5403 504 ...
## $ cate_3   : chr "B150401" "B160101" "B160201" "B180301" ...
## $ cusID    : int 17218 17218 17218 17218 17674 17674 14388 14388 14388 15773 ...
## $ storeCode: int 44 44 44 44 44 44 44 44 44 44 ...
## $ date     : int 20140222 20140222 20140222 20140222 20140222 20140222 20140222 20140222 20140222 20140222 ...
## $ time     : int 20 20 20 20 22 22 23 23 23 21 ...
## $ amount   : int 2420 1070 8060 6000 1120 1200 5290 5960 9900 970 ...
## - attr(*, ".internal.selfref")=<externalptr>
```

# 실습

1. dabrp\_classnote3 프로젝트를 pull하여 data 폴더를 받고 확인하세요
  2. 13p의 실습데이터 설정 스크립트를 실행하여 tran.csv(1.4G) 파일을 받으세요
  3. 11p의 함수들을 활용하여 데이터를 탐색해보세요
- \* tibble이나 data.table은 다음 시간에 다룰 예정입니다

# 데이터베이스

# 데이터란

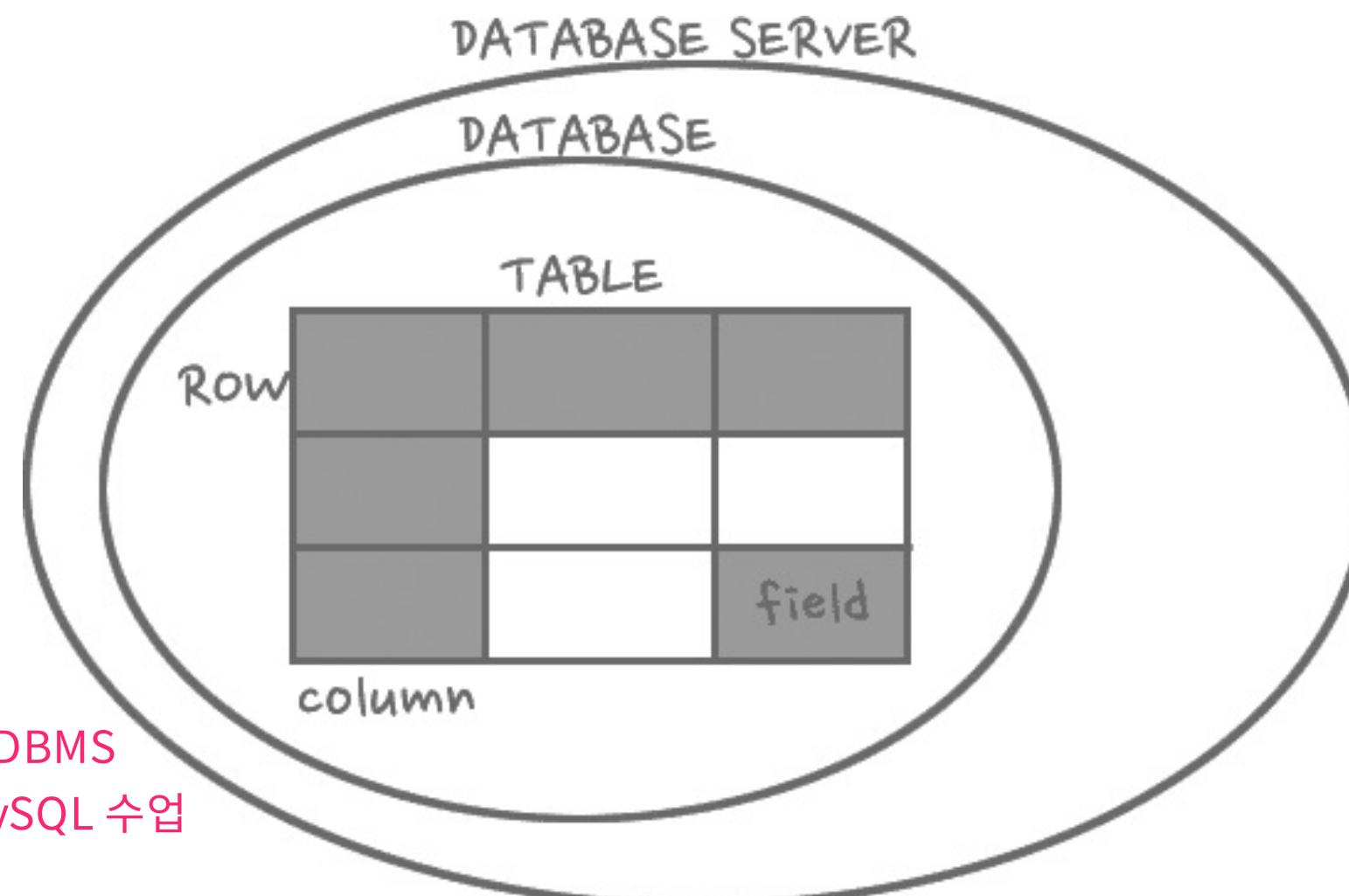
단순한 관찰이나 측정 등의 수단을 통해 현실 세계로부터 수집된 사실이나 값. 의미있게 사용하기 위해서 구조화<sup>\*</sup>가 필요함



\* 구조화: 체계적으로 조직하는 것

# Database Management System(DBMS)<sup>1</sup>

DBMS는 데이터베이스를 관리하는 응용프로그램의 일종으로 중복 제어, 접근 권한 관리, 무결성 제약, 데이터 관리 등의 기능을 제공하며 제품마다 특성이 매우 다양함

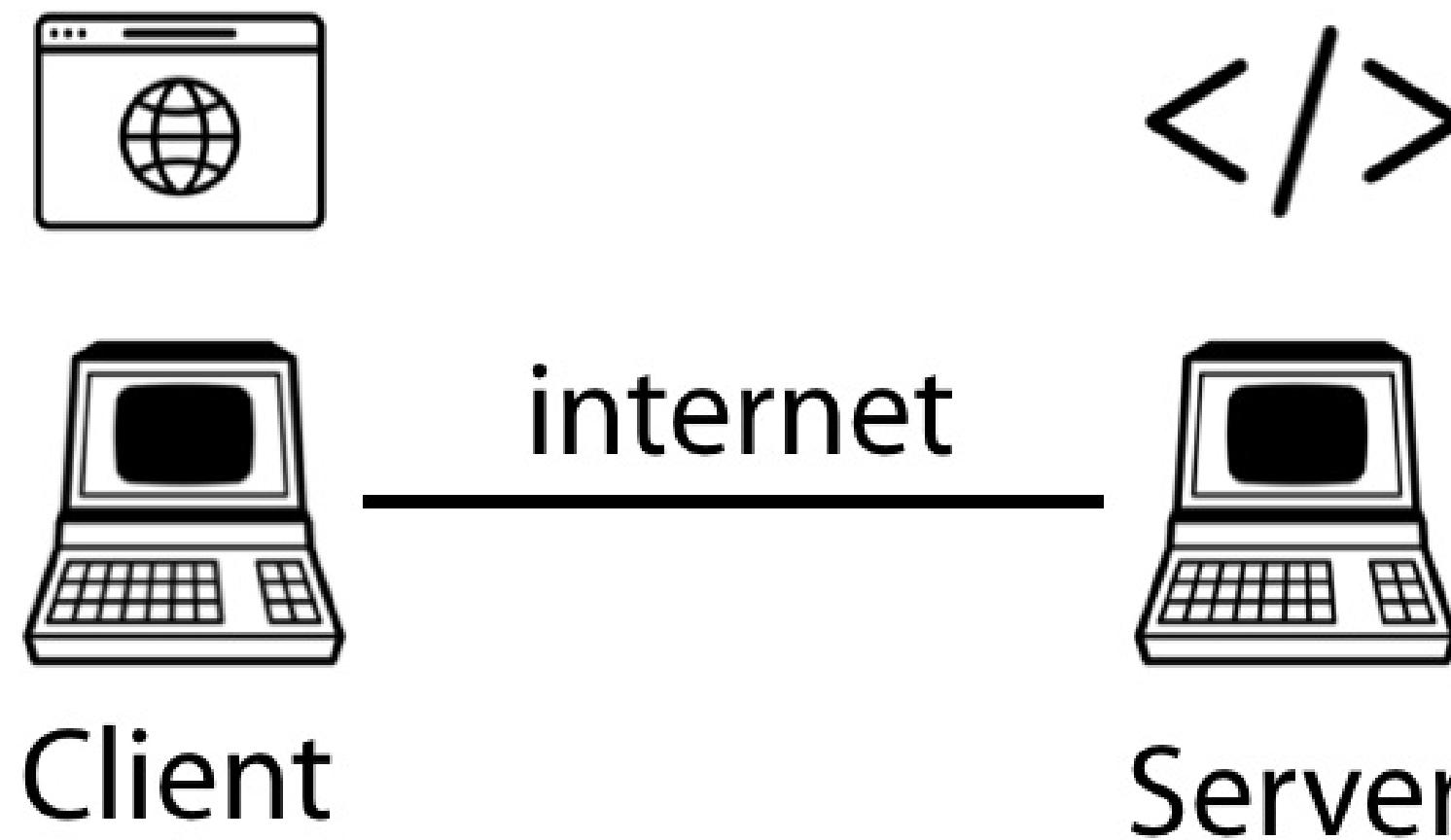


1 <https://namu.wiki/w/DBMS>

\* 이미지 출처: 생활코딩 MySQL 수업

# 서버와 클라이언트

유명하게 사용되는 DBMS는 잘 활용되는 GUI가 있기도 하나 그 기능 구현이 서버와 클라이언트로 되어 있어 연결시 ip 주소가 필요하는 등 관련 지식이 필요함



# 데이터베이스 클라이언트

- 대표적인 클라이언트
  - MySQL monitor
  - PHPmyAdmin
  - Navicat
  - HeidiSQL

# 테이블(table)

- 2차원의 자료형으로 table과 같이 행(row)과 열(column)이 있는 엑셀의 시트와 비슷한 모양

```
print(head(iris))
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

\* R 데이터 객체인 iris의 data.frame 예시

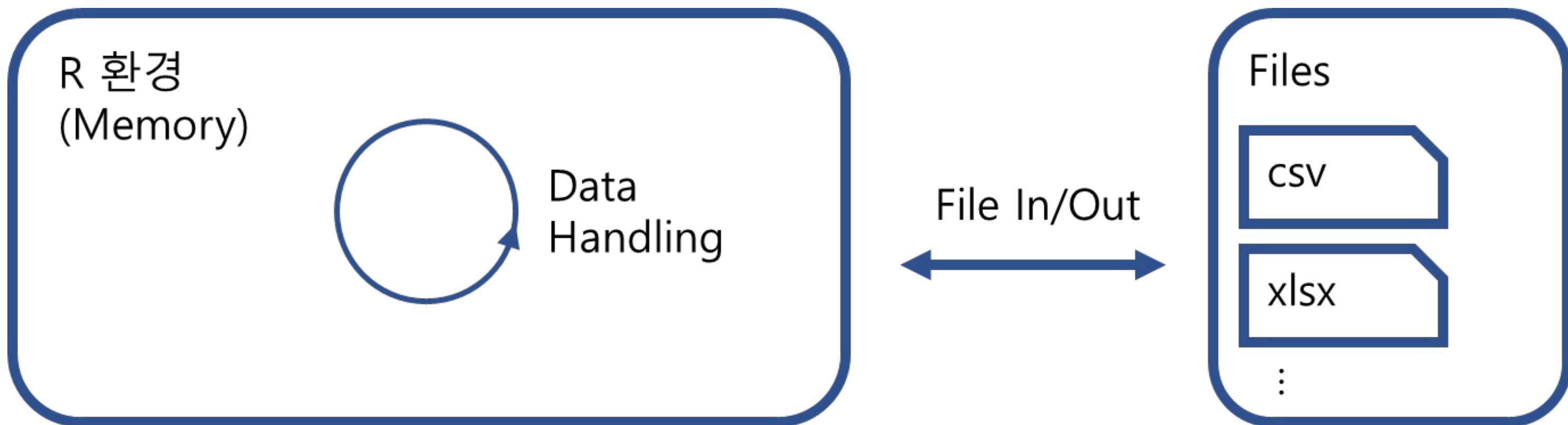
# DBI(database interface)

DBI는 odbc나 java의 jdbc, 파이썬의 db-api와 같이 R에서 데이터베이스를 컨트롤하기 위한 인터페이스 패키지

```
if (!require(devtools)) install.packages("devtools")
if (!require(DBI)) devtools::install_github("rstats-db/DBI")
if (!require(RSQLite)) devtools::install_github("rstats-db/RSQLite")
```

# 파일의 입출력

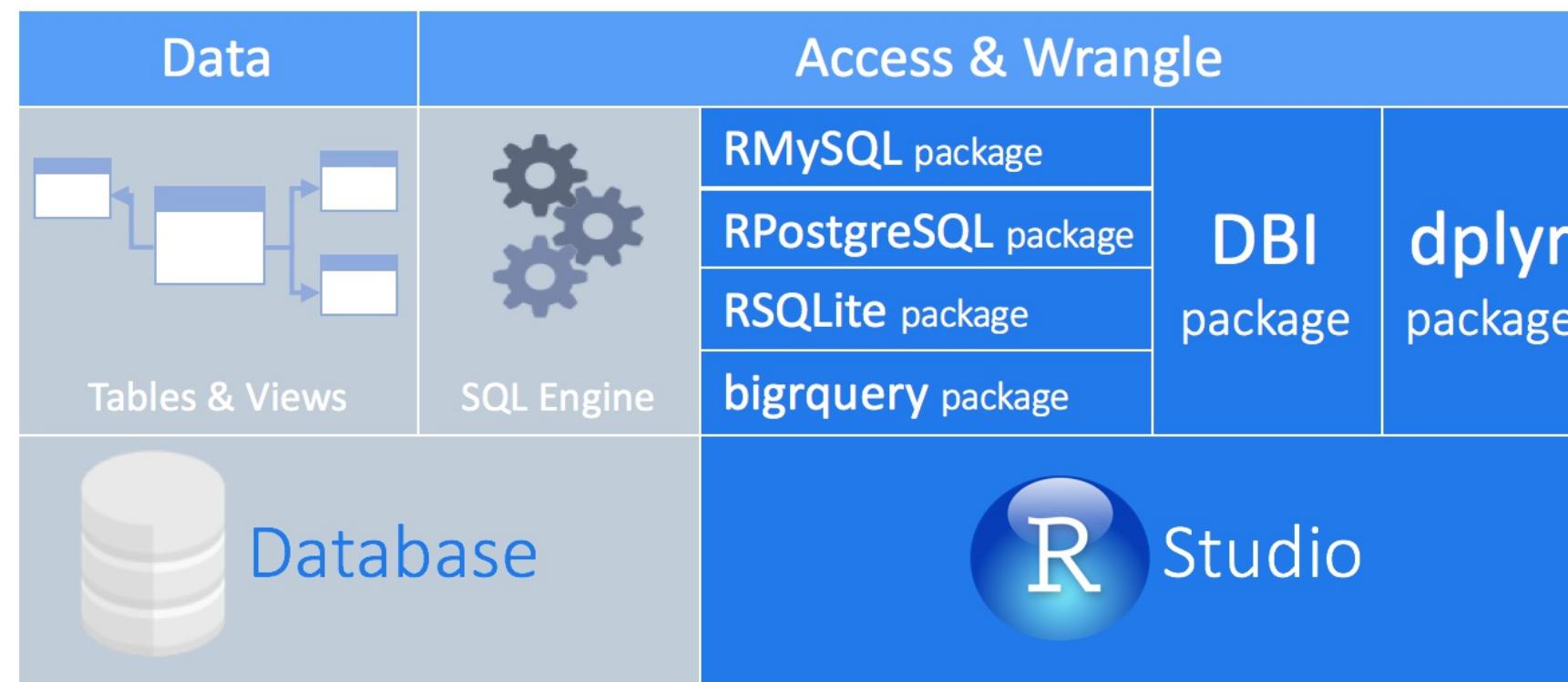
R은 데이터를 사용하기 위해 메모리로 구성된 R 환경에 파일의 내용을 R 객체 형태의 데이터로 변환해서 사용



# DBI의 DBMS 연결

DBI(Database Interface) 패키지는 각 DBMS에 해당하는 드라이버 패키지와 연동하여 DBMS를 통해 데이터를 사용

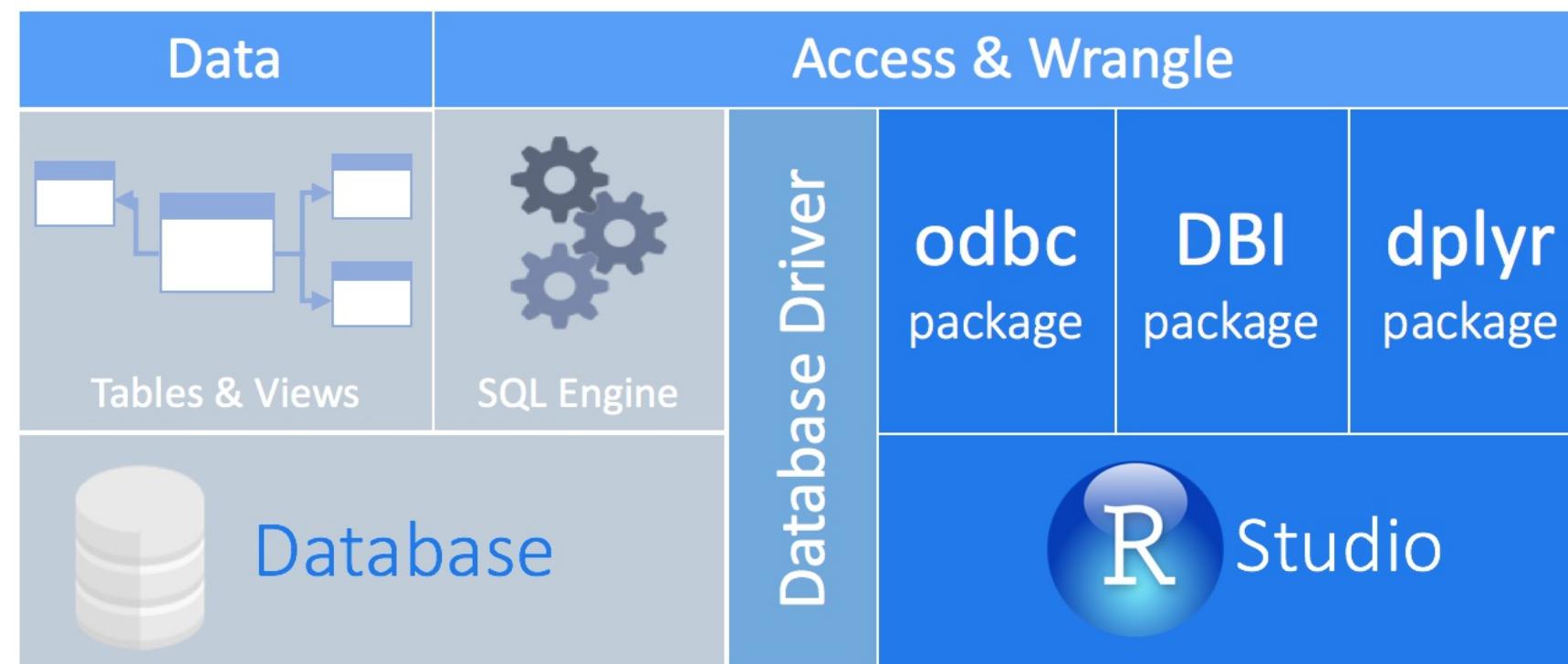
## Open Source Databases



# ODBC를 통한 연결

ODBC를 직접 지원하기도 함

## Commercial Databases



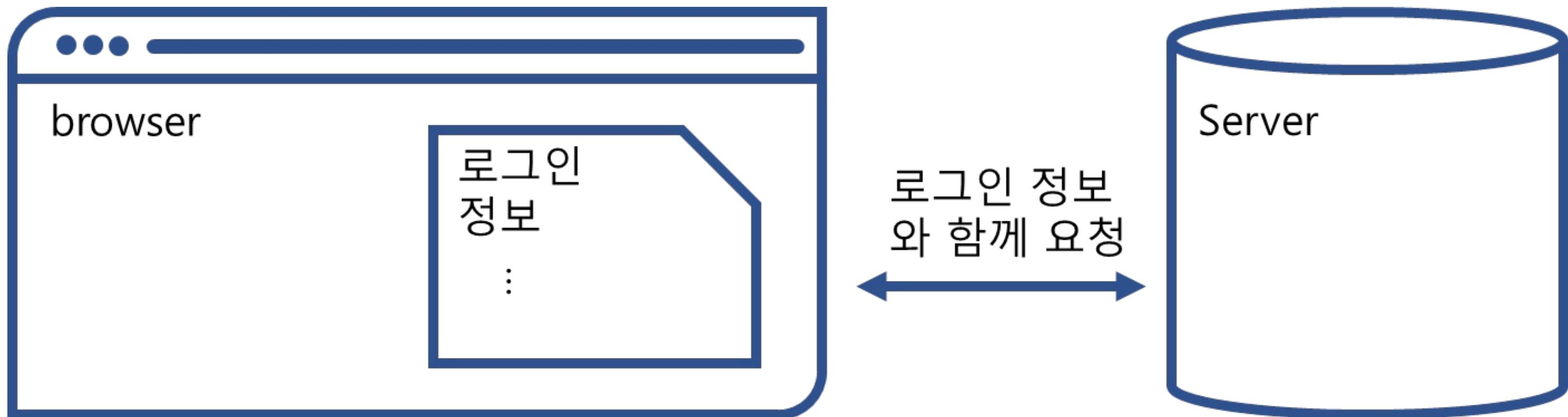
# 데이터 베이스와 연결

파일과 다르게 데이터베이스는 DBMS를 통해서 내부의 데이터에 접근해야 하기 때문에 DBMS의 연결 정보를 가진 conn 객체를 만들고 그것을 활용하는 명령 구조를 따름



# 데이터 베이스와 연결

정확히 같진 않지만 conn은 브라우저가 로그인 정보등을 바탕으로 서버와 통신하는 것과 비슷하게 동작함



# SQLite

SQLite is a self-contained, high-reliability, embedded, full-featured, public-domain, SQL database engine.

파일 하나로 구성하는 작고, 무료인 sql db

# DBI 함수

패키지 불러오기. 패키지가 설치되어 있지 않으면 28p 참조

```
library(DBI)
library(RSQLite)
library(data.table)
```

# SQLite 와 연결하기

conn은 관례적으로 연결 정보를 저장하는 객체의 이름이며 SQLite는 DB가 파일로 구성되어 있어서 권한(ID나 비밀번호) 설정이 없고 연결 정보도 단순(host 없음)함

```
conn <- dbConnect(RSQLite::SQLite(), dbname = ".../FolderForClass2/class2.sqlite")
conn
```

```
## <SQLiteConnection>
##   Path: C:\Users\mrchypark\Documents\project\dabrp_classnote3\FolderForClass2\class2.sqlite
##   Extensions: TRUE
```

```
# DB내에 있는 테이블의 이름을 확인함
dbListTables(conn)
```

```
## character(0)
```

# DBI 명령으로 table 만들기

DBI 패키지에는 테이블을 작성하고 지우는 등 원래 SQL로 작성하는 Query의 몇 가지 유용한 경우를 함수로 만들어 제공.  
dbwriteTable() 함수는 R 객체를 바로 테이블로 작성하는 기능

```
dbwriteTable(conn, "mtcars", mtcars, overwrite = T)  
dbListTables(conn)  
  
## [1] "mtcars"
```

# DBI 명령으로 table 불러오기

dbReadTable()은 명령어에서도 이해하기 쉽듯 DB내에 있는 특정 테이블을 이름으로 지정하여 데이터를 불러오는 함수

```
dbReadTable(conn, "mtcars")
```

```
##   mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## 1 21.0   6 160.0 110 3.90 2.620 16.46 0  1    4    4
## 2 21.0   6 160.0 110 3.90 2.875 17.02 0  1    4    4
## 3 22.8   4 108.0  93 3.85 2.320 18.61 1  1    4    1
## 4 21.4   6 258.0 110 3.08 3.215 19.44 1  0    3    1
## 5 18.7   8 360.0 175 3.15 3.440 17.02 0  0    3    2
## 6 18.1   6 225.0 105 2.76 3.460 20.22 1  0    3    1
## 7 14.3   8 360.0 245 3.21 3.570 15.84 0  0    3    4
## 8 24.4   4 146.7  62 3.69 3.190 20.00 1  0    4    2
## 9 22.8   4 140.8  95 3.92 3.150 22.90 1  0    4    2
## 10 19.2  6 167.6 123 3.92 3.440 18.30 1  0    4    4
## 11 17.8  6 167.6 123 3.92 3.440 18.90 1  0    4    4
## 12 16.4  8 275.8 180 3.07 4.070 17.40 0  0    3    3
## 13 17.3  8 275.8 180 3.07 3.730 17.60 0  0    3    3
## 14 15.2  8 275.8 180 3.07 3.780 18.00 0  0    3    3
## 15 10.4  8 472.0 205 2.93 5.250 17.98 0  0    3    4
## 16 10.4  8 460.0 215 3.00 5.424 17.82 0  0    3    4
## 17 14.7  8 440.0 230 3.23 5.345 17.42 0  0    3    4
## 18 32.4   4  78.7  66 4.08 2.200 19.47 1  1    4    1
## 19 30.4   4  75.7  52 4.93 1.615 18.52 1  1    4    2
## 20 33.9   4  71.1  65 4.22 1.835 19.90 1  1    4    1
## 21 21.5   4 120.1  97 3.70 2.465 20.01 1  0    3    1
## 22 15.5   8 318.0 150 2.76 3.520 16.87 0  0    3    2
```

# DBI 명령으로 table 지우기

`dbRemoveTable()` 은 테이블을 이름으로 지정하여 지우는 함수

```
dbRemoveTable(conn, "mtcars")
```

```
dbListTables(conn)
```

```
## character(0)
```

# 실습

1. DBI 패키지로 연결, 쓰기, 읽기를 수행해 봅니다.
  1. `dbConnect()`함수와 `SQLite()`함수를 이용해서 "./ForderForclass2" 폴더에 "db\_[자기이름영어].sqlite" 파일 이름으로 "conn" 연결 정보를 만드세요.
  2. "`dbListTables(conn)`"으로 테이블이 아무것도 없는 것을 확인하세요.
  3. "./data/recomen/chennel.csv"을 "chen" 이름의 R 객체로 불러오세요.
  4. "`dbWriteTable()`" 함수를 이용해 "chen" 데이터 객체를 DB에 "dbchen"이라는 테이블 이름으로 저장해 보세요.
  5. "`dbReadTable()`" 함수로 "dbchen"을 불러와서 "chen" 데이터 객체와 내용이 같은지 확인하세요.  
\* `identical()` 함수는 두 개의 객체가 같은지를 T/F로 결과를 주는 함수입니다.
  6. "./ForderForclass2" 폴더에 "db\_[자기이름영어].sqlite" 이 "chen" 테이블을 저장한 상태를 commit 해주세요.
  7. github에서 PR을 진행해주세요.
  8. 같은 과정을 "./data/recomen/customer.csv" 파일로 진행해 주세요. 중간에 명명은 "cust", "dbcust" 입니다.  
\* 인코딩 문제가 발생할 수 있습니다.

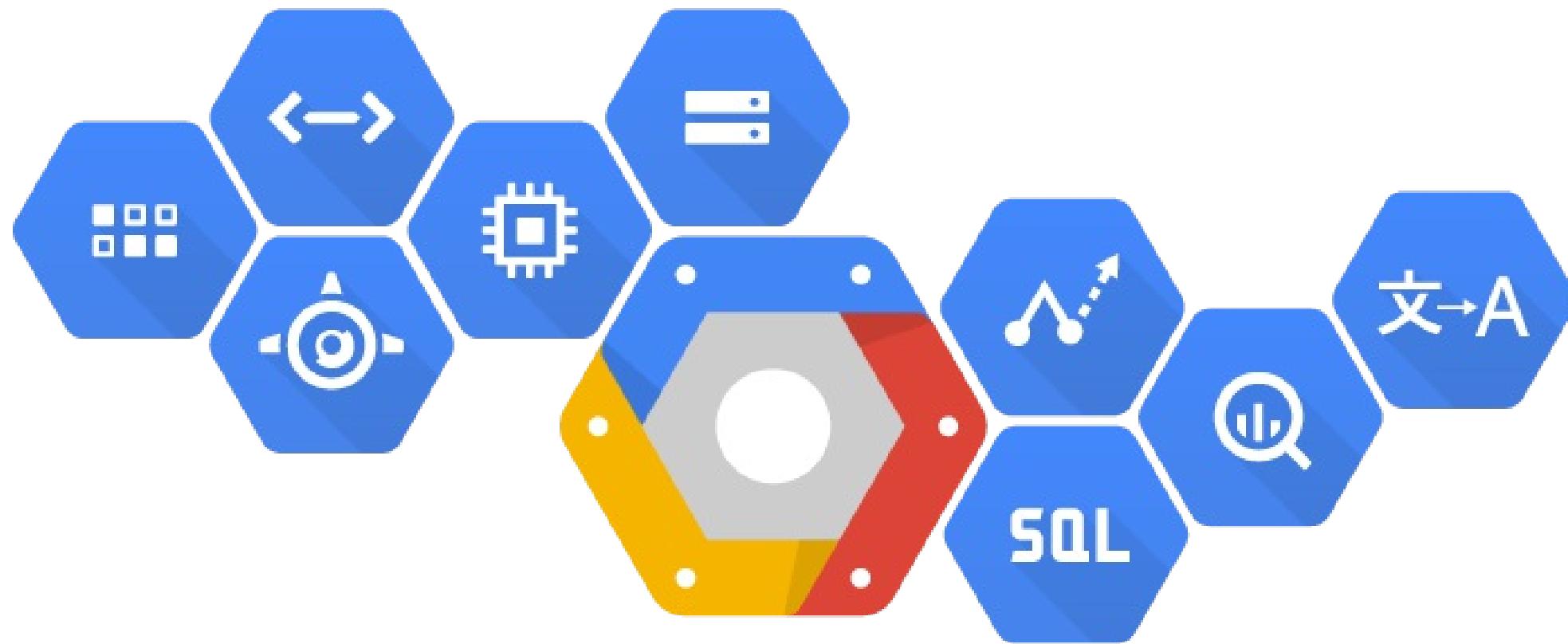
# 클라우드 서비스

# 클라우드 서비스 소개

클라우딩 컴퓨팅은 사용자의 환경 밖에서 서비스로서 제공된 확장 가능한 컴퓨팅 자원을 사용한 양에 따라 비용을 지불하고 사용하는 것

출처: P. Changanti, 가상 인프라용 클라우드 서비스, Part 1: IaaS(Infrastructure as a Service) 및 Eucalyptus

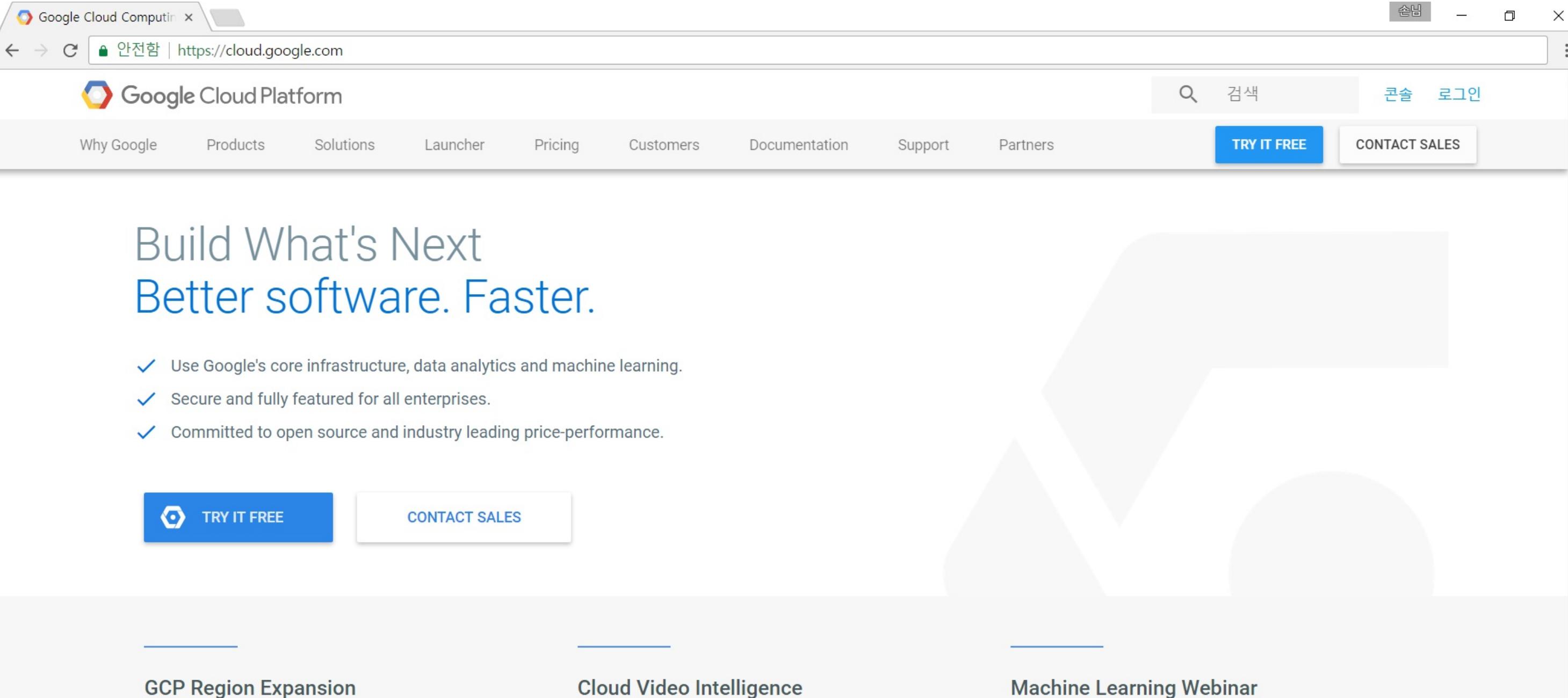
# 구글 클라우드



Google Cloud Platform

# 구글 클라우드 시작하기

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



The image shows the Google Cloud Platform homepage. At the top, there's a navigation bar with links for 'Why Google', 'Products', 'Solutions', 'Launcher', 'Pricing', 'Customers', 'Documentation', 'Support', and 'Partners'. On the right side of the header are buttons for 'TRY IT FREE' (in blue), 'CONTACT SALES' (in white), and user account options ('Cloud', 'Login'). The main headline reads 'Build What's Next' followed by 'Better software. Faster.' Below this, there's a bulleted list of benefits: '✓ Use Google's core infrastructure, data analytics and machine learning.', '✓ Secure and fully featured for all enterprises.', and '✓ Committed to open source and industry leading price-performance.' At the bottom, there are three cards: 'GCP Region Expansion', 'Cloud Video Intelligence', and 'Machine Learning Webinar'. Each card has a brief description and a link.

Google Cloud Computin x 솔림 - X

← → C 안전함 | https://cloud.google.com : :

Google Cloud Platform

Why Google Products Solutions Launcher Pricing Customers Documentation Support Partners TRY IT FREE CONTACT SALES

Cloud Platform

Build What's Next

Better software. Faster.

- ✓ Use Google's core infrastructure, data analytics and machine learning.
- ✓ Secure and fully featured for all enterprises.
- ✓ Committed to open source and industry leading price-performance.

TRY IT FREE CONTACT SALES

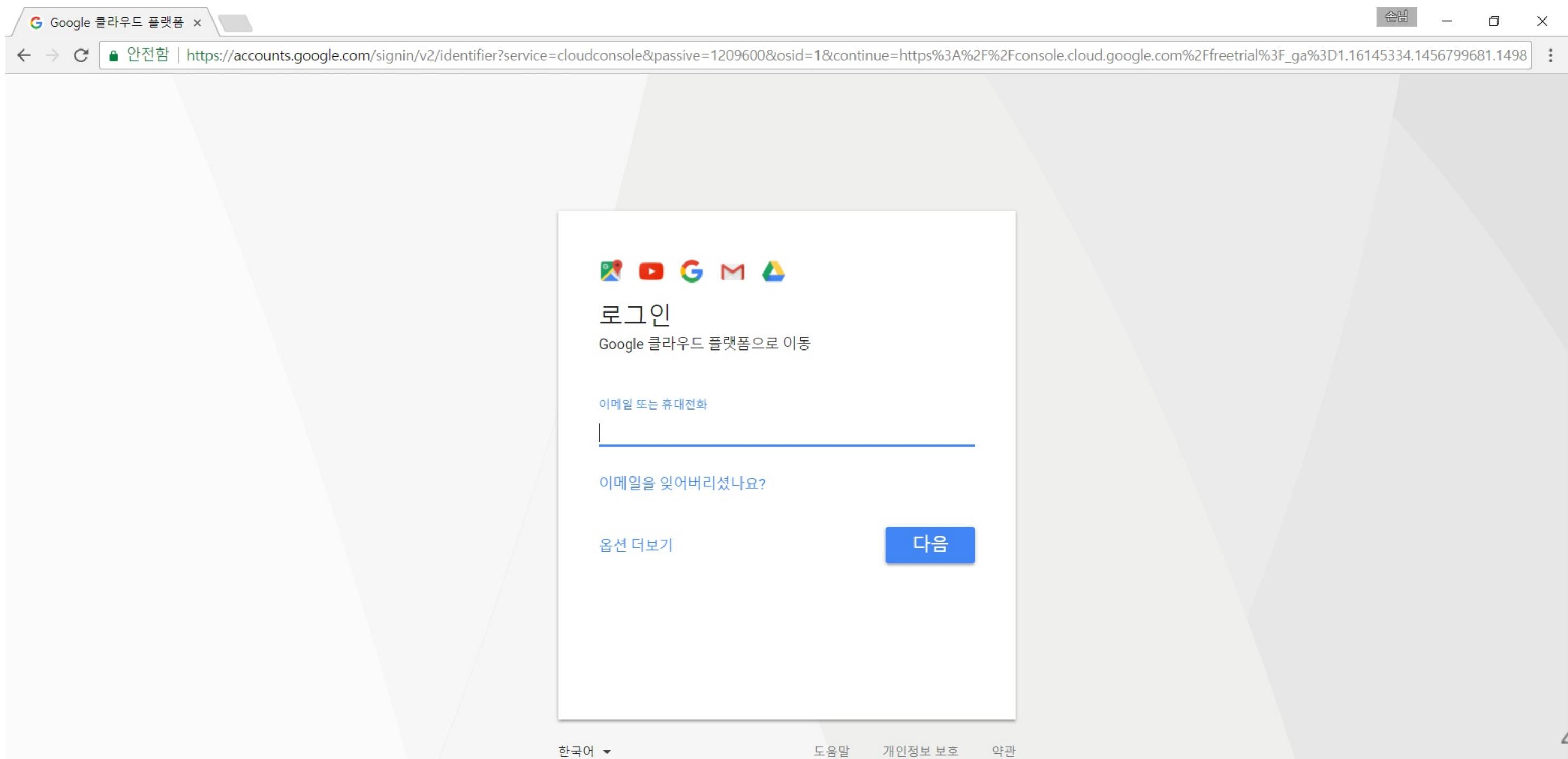
GCP Region Expansion

Cloud Video Intelligence

Machine Learning Webinar

Join us live as Google and MIT discuss the value of ML and

# 구글 클라우드 로그인



The image shows a screenshot of a web browser window titled "Google 클라우드 플랫폼" (Google Cloud Platform). The URL in the address bar is [https://accounts.google.com/signin/v2/identifier?service=cloudconsole&passive=1209600&osid=1&continue=https%3A%2F%2Fconsole.cloud.google.com%2Ffreetrial%3F\\_ga%3D1.16145334.1456799681.1498](https://accounts.google.com/signin/v2/identifier?service=cloudconsole&passive=1209600&osid=1&continue=https%3A%2F%2Fconsole.cloud.google.com%2Ffreetrial%3F_ga%3D1.16145334.1456799681.1498). The main content is a Google sign-in page for the Cloud Platform. It features icons for Google services like Google Maps, YouTube, Google, Gmail, and Google Drive. The text "로그인" (Login) is prominently displayed, along with the subtitle "Google 클라우드 플랫폼으로 이동". Below this is a text input field labeled "이메일 또는 휴대전화" (Email or phone number) with a placeholder underscore. A link "이메일을 잊어버리셨나요?" (Forgot your email?) is provided. At the bottom left is a "옵션 더보기" (More options) link, and at the bottom right is a blue "다음" (Next) button. The footer of the page includes links for "한국어" (Korean), "도움말" (Help), "개인정보 보호" (Privacy), and "약관" (Terms).

# 구글 클라우드 설문

The screenshot shows a web browser window for the Google Cloud Platform. The URL in the address bar is [https://console.cloud.google.com/freetrial?\\_ga=1.16145334.1456799681.1498616668&pli=1&page=0](https://console.cloud.google.com/freetrial?_ga=1.16145334.1456799681.1498616668&pli=1&page=0). The page title is "Google Cloud Platform".

**Cloud Platform 무료로 사용해 보기**

Google

국가  
내안민국

수락  
기능 공지사항, 성능 제안사항, 의견 설문조사 및 특별 할인에 대한 새로운 소식을 이메일로 받고 싶습니다.  
 예       아니요

Google Cloud Platform 무료 평가판 서비스 약관을 읽었으며 이에 동의합니다.  
계속 진행하려면 체크박스를 선택하세요.  
 예       아니요

동의 및 계속하기

모든 Cloud Platform 제품에 액세스  
앱, 웹사이트, 서비스를 구축하고 실행하는 데 필요한 Firebase, Google Maps API 등의 모든 기능을 이용할 수 있습니다.

무료 크레딧 300달러  
가입하여 Google Cloud Platform에서 이후 12개월 동안 사용할 수 있는 300달러의 혜택을 받아 보세요.

무료 평가판 종료 후 자동 청구되지 않음  
신용카드를 요청하는 이유는 자동 가입을 방지하기 위해서입니다. 유료 계정으로 직접 업그레이드하지 않는 한 요금이 청구되지 않습니다.

개인정보처리방침

46 / 99

# 구글 클라우드 설문 선택

The screenshot shows a web browser window for the Google Cloud Platform free trial. The URL in the address bar is [https://console.cloud.google.com/freetrial?\\_ga=1.16145334.1456799681.1498616668&pli=1&page=0](https://console.cloud.google.com/freetrial?_ga=1.16145334.1456799681.1498616668&pli=1&page=0). The page title is "Google Cloud Platform".

**Cloud Platform 무료로 사용해 보기**

Google

국가  
내안민국

수락  
기능 공지사항, 성능 제안사항, 의견 설문조사 및 특별 할인에 대한 새로운 소식을 이메일로 받고 싶습니다.

예  아니요

Google Cloud Platform 무료 평가판 서비스 약관을 읽었으며 이에 동의합니다.  
계속 진행하려면 체크박스를 선택하세요.

예  아니요

동의 및 계속하기

모든 Cloud Platform 제품에 액세스  
앱, 웹사이트, 서비스를 구축하고 실행하는 데 필요한 Firebase, Google Maps API 등의 모든 기능을 이용할 수 있습니다.

무료 크레딧 300달러  
가입하여 Google Cloud Platform에서 이후 12개월 동안 사용할 수 있는 300달러의 혜택을 받아 보세요.

무료 평가판 종료 후 자동 청구되지 않음  
신용카드를 요청하는 이유는 자동 가입을 방지하기 위해서입니다. 유료 계정으로 직접 업그레이드하지 않는 한 요금이 청구되지 않습니다.

개인정보처리방침

47 / 99

# 구글 클라우드 카드 등록

The screenshot shows a web browser window for the Google Cloud Platform (GCP) free trial. The URL in the address bar is [https://console.cloud.google.com/freetrial?\\_ga=1.16145334.1456799681.1498616668&pli=1&page=1](https://console.cloud.google.com/freetrial?_ga=1.16145334.1456799681.1498616668&pli=1&page=1). The page title is "Google Cloud Platform".

The main content area has a blue header bar with the text "Cloud Platform 무료로 사용해 보기" and the "Google" logo.

The form fields are as follows:

- 고객 정보**:
  - 계정 유형:  사업자
  - 이름 및 주소:
    - 도/시: [Input field]
    - 구/군: [Input field]
    - 주소:
      - 주소 입력란 1: [Input field]

**모든 Cloud Platform 제품에 액세스**  
앱, 웹사이트, 서비스를 구축하고 실행하는 데 필요한 Firebase, Google Maps API 등의 모든 기능을 이용할 수 있습니다.

**무료 크레딧 300달러**  
가입하여 Google Cloud Platform에서 이후 12개월 동안 사용할 수 있는 300달러의 혜택을 받아 보세요.

**무료 평가판 종료 후 자동 청구되지 않음**  
신용카드를 요청하는 이유는 자동 가입을 방지하기 위해서입니다. 유료 계정으로 직접 업그레이드하지 않는 한 요금이 청구되지 않습니다.

# 구글 클라우드 프로젝트 생성

The screenshot shows the Google Cloud Platform dashboard for the project 'konlper'. The left sidebar lists various services: Compute Engine, Datastore, Storage, SQL (selected), Spanner, STACKDRIVER, Monitoring, Debug, Metrics, Log, and Error Reporting. The main content area has tabs for 대시보드 (Dashboard) and 활동 (Activities). The Dashboard section includes:

- 프로젝트 정보**: konlper, 프로젝트 ID: konlper-168808, #636441884378. Link: [프로젝트 설정으로 이동](#).
- Compute Engine**: CPU usage chart from 6월 28일 오후 12:30 to 6월 28일 오후 1:06. CPU usage: 3.083. Link: [Compute Engine 대시보드로 이동](#).
- SQL**: 사용된 저장소(바이트) 1G, 768M. Link: [오류 보고 설정 방법 알아보기](#).
- Google Cloud Platform 상태**: 모든 서비스 정상. Link: [Cloud 상태 대시보드로 이동](#).
- 결제**: \$0.00. Link: [청구 세부정보 보기](#).
- 오류 보고**: 오류가 감지되지 않았습니다. 오류 보고를 설정하셨나요? Link: [오류 보고 설정 방법 알아보기](#).

# R MySQL 연결

DBI로 MySQL을 연결하는 패키지와 사용

```
library(RMySQL)
con <- dbConnect(MySQL(),
                  user = user,
                  password = pw,
                  host = host,
                  dbname = "recom")
dbListTables(conn = con)
dbwriteTable(conn = con, name = 'tran', value = "./data/recomen/tran.csv")
dbReadTable(conn = con, name = "tran")
```

# MySQL 인스턴스 생성

The screenshot shows the Google Cloud Platform interface for managing MySQL instances. The top navigation bar includes the project name 'SQL 개요 - konlper', a search bar, and various status indicators. The main menu bar has 'Google Cloud Platform' and 'konlper' dropdowns, along with a search icon and user profile. On the left, there's a sidebar with 'SQL' and '인스턴스' buttons. The main content area is titled 'Cloud SQL' and 'Cloud SQL 인스턴스'. It contains a descriptive paragraph about Cloud SQL instances and two buttons at the bottom: '인스턴스 생성' (Instance Creation) and '자세히 알아보기' (Learn More). The URL in the browser address bar is <https://console.cloud.google.com/sql/instances?project=konlper-168808>.

# 엔진선택

The screenshot shows a web browser window for the Google Cloud Platform SQL instance creation process. The URL in the address bar is <https://console.cloud.google.com/sql/choose-instance-engine?project=konlper-168808>. The page title is "SQL - konlper". The main content area is titled "데이터베이스 엔진 선택" (Database Engine Selection). It offers two options: "MySQL" (selected) and "PostgreSQL" (beta). Both options have a small description below them: "버전 5.6 또는 5.7" for MySQL and "버전 9.6" for PostgreSQL. A blue "다음" (Next) button is at the bottom left.

SQL - konlper

안전함 | https://console.cloud.google.com/sql/choose-instance-engine?project=konlper-168808

Google Cloud Platform konlper

SQL

데이터베이스 엔진 선택

MySQL  
버전 5.6 또는 5.7

PostgreSQL **베타**  
버전 9.6

다음

# MySQL 2세대 선택

The screenshot shows the Google Cloud Platform SQL interface. The title bar indicates the project is 'konlper' and the URL is 'https://console.cloud.google.com/sql/pre-create?project=konlper-168808'. The main navigation bar includes 'Google Cloud Platform', 'konlper', a search bar, and various status icons.

In the top left, there's a 'SQL' icon and a link to 'MySQL 인스턴스 유형 선택'. Below this, a message states: 'Cloud SQL MySQL 인스턴스에는 두 가지 유형이 있습니다. 자세히 알아보기'.

The interface displays two options:

- MySQL 2세대(권장)**: This section is highlighted with a grey background. It contains the following text:

저렴한 비용으로 높은 성능과 큰 저장용량이 제공됩니다.

  - 1세대 처리량의 최대 7배 및 저장용량 크기 20배
  - 대부분의 사용 사례에서 1세대보다 비용 낮음
  - 고가용성 장애 조치를 추가하고 복제본을 읽을 수 있는 옵션 제공
  - 백업 기간 및 유지관리 기간 구성 가능
  - MySQL 5.6 및 5.7만 지원됩니다.

**2세대 선택**
- MySQL 1세대(이전)**: This section has a white background. It contains the following text:

기본 성능과 저장용량 크기를 제공하는 이전 버전의 Cloud SQL입니다.  
MySQL 5.7을 지원하지 않습니다.

**1세대 선택**

# 인스턴스 설정

The screenshot shows the Google Cloud Platform SQL creation interface. The URL in the browser is <https://console.cloud.google.com/sql/create-instance-mysql?project=konlper-168808>. The page title is "MySQL 2세대 인스턴스 만들기".

**인스턴스 ID**  
이후에 변경할 수 없습니다. 소문자, 숫자, 하이픈을 사용할 수 있으며 문자로 시작해야 합니다.  
[입력 필드]

**데이터베이스 버전**  
MySQL 5.7

**위치**  
성능을 향상하려면 필요한 서비스와 가까운 위치에 데이터를 보관하세요.

지역	영역
us-central1	자동 선택

**머신 유형**  
성능을 향상하려면 가장 큰 표를 보관할 수 있는 메모리가 제공되는 머신 유형을 선택하세요.

db-n1-standard-1	vCPU	메모리
1	3.75GB	변경

**네트워크 처리량(MB/초)** 250/2,000

# 접속 허용 IP 설정

The screenshot shows the Google Cloud Platform SQL interface for creating a MySQL instance. The top navigation bar includes tabs for 'SQL - konlper' and 'konlper'. The main title is 'MySQL 2세대 인스턴스 만들기'. The interface displays several configuration sections:

- 유지관리 기간**: A dropdown menu set to '시스템 자동 선택 기간'.
- 유지관리 시점**: A dropdown menu set to '자동 선택'.
- 루트 비밀번호**: A section for setting the root password, with a note: '루트 사용자의 비밀번호를 설정하세요. 자세히 알아보기'. It includes a password input field, a '생성' (Create) button with a key icon, and a checkbox for '비밀번호 없음' (No password).
- Cloud SQL 플래그**: A section with a '+ 항목 추가' (Add item) button.
- 승인된 네트워크**: A section for adding IPv4 addresses to the instance's network. It includes a note: '아래에 IPv4 주소를 추가하여 인스턴스에 연결할 네트워크를 승인하세요. 아래에 추가한 주소를 통해서만 네트워크가 승인됩니다.' and a '+ 네트워크 추가' (Add network) button.

# 접속 허용 IP 설정

The screenshot shows the Google Cloud Platform SQL interface for creating a MySQL instance. The top navigation bar includes the project name 'konlper' and the URL 'https://console.cloud.google.com/sql/create-instance-mysql?project=konlper-168808'. The main page displays the 'MySQL 2세대 인스턴스 만들기' (Create MySQL 2nd Gen Instance) section. A note indicates that no password is required. Below this, there is a 'Cloud SQL 플래그' (Cloud SQL Flags) section with a '+ 항목 추가' (Add item) button. The central part of the screen is a modal window titled '새 네트워크' (New Network) for configuring network access. The modal contains fields for '이름 (선택사항)' (Name (optional)) set to '없음' (None), '네트워크' (Network) with a 'CIDR 표기' (CIDR notation) field containing '예: 199.27.25.0/24', and two buttons '완료' (Done) and '취소' (Cancel). At the bottom of the modal is a '+ 네트워크 추가' (Add network) button.

# 현재 IP 확인하기

A screenshot of a web browser window showing search results for "myip" on Google. The browser has two tabs open: "SQL - konlper" and "myip - Google 검색". The "myip - Google 검색" tab is active, displaying search results in Korean. The results include links to various websites like WhatIsMyIP.com, whatismyipaddress.com, iplocation.net, and expressvpn.com, each offering tools to check current IP address information.

SQL - konlper x myip - Google 검색 x

← → C 🔒 안전함 | https://www.google.co.kr/search?q=myip&oq=myip&aq=chrome..69i57.1526j0j4&sourceid=chrome&ie=UTF-8

Google myip

전체 동영상 뉴스 이미지 지도 더보기 설정 도구

검색결과 약 20,800,000개 (0.31초)

**What Is My IP Address - IP Address Tools and Info - WhatIsMyIP.com ®**  
<https://www.whatismyip.com/> 이 페이지 번역하기  
We provide IP address tools that allow users to perform an Internet Speed Test, IP address lookup, proxy detection, IP Whois Lookup, and more.

**My IP Information**  
My IP Information tool shows your ip address, city, state, country ...  
[whatismyip.com 검색결과 더보기 »](#)

**IP Address Lookup**  
The IP Address Lookup tool includes the following IP ...

**What Is My IP Address? IP Address Tools and More**  
[whatismyipaddress.com/](https://whatismyipaddress.com/) 이 페이지 번역하기  
IP address lookup, location, proxy detection, email tracing, IP hiding tips, blacklist check, speed test, and forums. Find, get, and show my IP address.

**What is my IP address? - IP Location**  
<https://www.iplocation.net/find-ip-address> 이 페이지 번역하기  
This webpage displays public IP address of your computer or router assigned by your ISP.

**What's My IP Address? | Online Privacy and Security Tool - ExpressVPN**  
<https://www.expressvpn.com/what-is-my-ip> 이 페이지 번역하기  
IP address lookup. What do others know about your location? See the true IP address of your VPN or proxy server. Learn to hide your IP address in 2 minutes.

**What's My IP Address? Networking Tools & More**  
[whatismyip.com/](#) 이 페이지 번역하기

# whatismyip

<https://www.whatismyip.com/>

The screenshot shows a web browser window with the URL <https://www.whatismyip.com/> in the address bar. The page displays the following information:

- WhatIsMyIP.com logo** and **THE IP ADDRESS EXPERTS** tagline.
- Navigation menu:** IP Tools, How To, Resources.
- User account links:** Log In, Create Account.
- Main content area:** Your IP Address Is: **112.217.214.251**. Below it, detailed location information is listed:
  - City: Seoul
  - State: Seoul-teukbyeolsi
  - Country: Korea, Republic of
  - ISP: LG Dacom Corporation
- Advertisement:** A Grammarly Instant Grammar Checker box with a red "Try Now" button.
- Footer links:** My IP Information, What My IP Says About Me, Proxy Check.
- Social media and stats:** Facebook icon with 42.9K likes.

# ip 입력

The screenshot shows the Google Cloud Platform SQL interface for creating a MySQL instance. The browser tabs indicate the user is on the 'SQL - konlper' and 'What Is My IP Address' pages. The main content area is titled 'MySQL 2세대 인스턴스 만들기' (Create MySQL 2nd Gen Instance). A note says '비밀번호 없음' (No password). Below it, a 'Cloud SQL 플래그' section has a '+ 항목 추가' button. The '네트워크' (Network) tab is active, showing a form to add a new network configuration. The form fields include '이름 (선택사항)' (Name (optional)) set to '없음' (None), '네트워크' (Network) with 'CIDR 표기' (CIDR notation) checked, and the IP address '112.217.214.251'. There are '완료' (Done) and '취소' (Cancel) buttons at the bottom, and a '+ 네트워크 추가' (Add Network) button at the bottom right.

# root 계정 비밀번호 생성

The screenshot shows the Google Cloud Platform SQL creation interface for creating a MySQL 2nd Generation instance. The current step is 'MySQL 2세대 인스턴스 만들기' (Create MySQL 2nd Gen instance).

**유지관리 기간** (Maintenance window):  
시스템 자동 선택 기간 (System automatic selection period): dropdown menu.

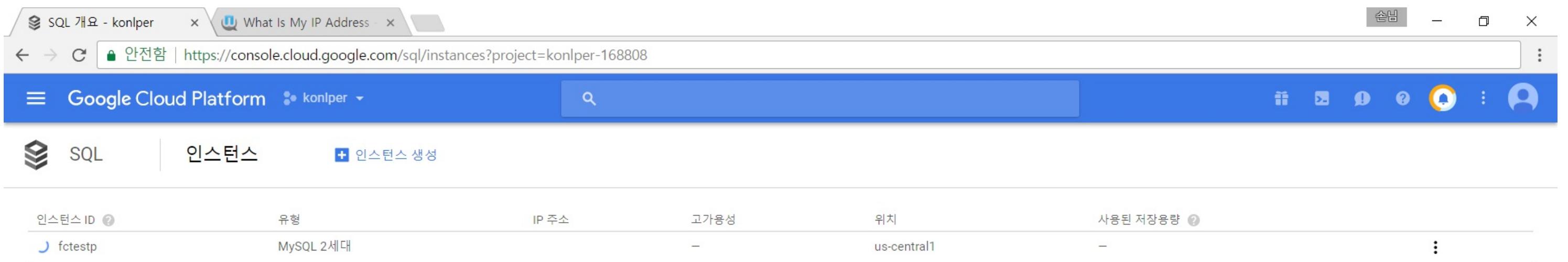
**유지관리 시점** (Maintenance time):  
자동 선택 (Automatic selection): dropdown menu.

**루트 비밀번호** (Root password):  
루트 사용자의 비밀번호를 설정하세요. 자세히 알아보기 (Set the root user's password. Learn more):  
 (password input field) and **생성** (Generate) button.  
 비밀번호 없음 (No password): checked checkbox.

**Cloud SQL 플래그** (Cloud SQL flags):  
**+ 항목 추가** (Add item) button.

**승인된 네트워크** (Approved network):  
아래에 IPv4 주소를 추가하여 인스턴스에 연결할 네트워크를 승인하세요. 아래에 주가 한 주소를 통해서만 네트워크가 승인됩니다.  
112.217.214.251 (IP address input field) and **저장/삭제** (Save/Delete) button.  
**+ 네트워크 추가** (Add network) button.

# SQL 인스턴스 생성중



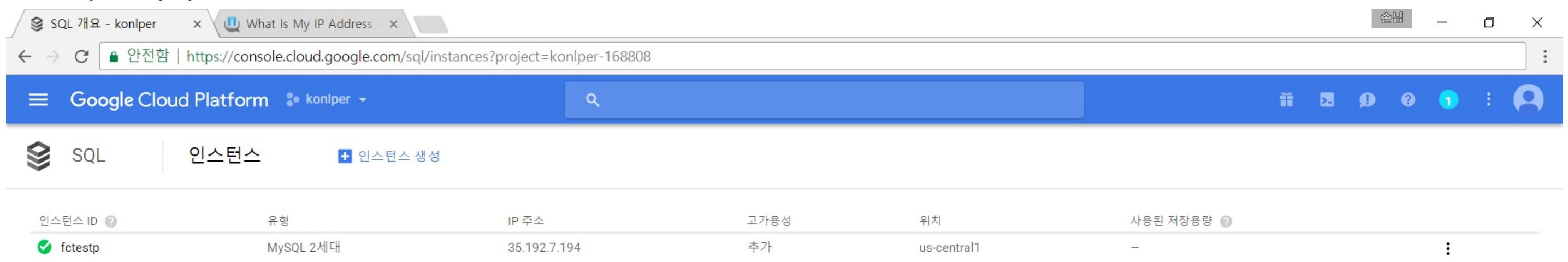
The screenshot shows the Google Cloud Platform interface for managing SQL instances. The top navigation bar includes tabs for 'SQL 개요 - konlper' and 'What Is My IP Address'. The main title bar displays a secure connection to 'https://console.cloud.google.com/sql/instances?project=konlper-168808'. The header also features the 'Google Cloud Platform' logo, a search bar, and various navigation icons.

The main content area is titled 'SQL' and shows a list of instances. A blue button labeled '+ 인스턴스 생성' (Create Instance) is visible. The table lists one instance:

인스턴스 ID	유형	IP 주소	고가용성	위치	사용된 저장용량
fctestp	MySQL 2세대	-	-	us-central1	-

# RMySQL로 연결하기

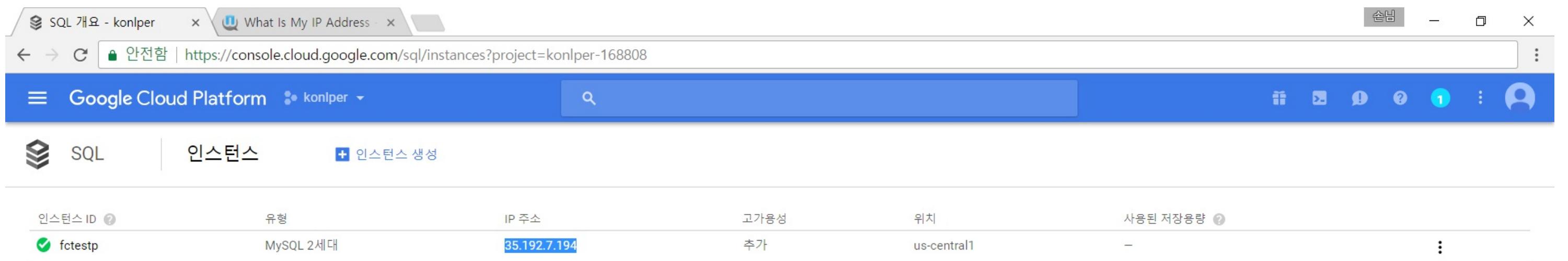
- host 주소 파악



The screenshot shows the Google Cloud Platform interface for managing SQL instances. The top navigation bar includes tabs for 'SQL 가요 - konlper' and 'What Is My IP Address'. The URL in the address bar is <https://console.cloud.google.com/sql/instances?project=konlper-168808>. The main header says 'Google Cloud Platform' and 'konlper'. Below the header, there are tabs for 'SQL' and '인스턴스' (Instances), with '인스턴스' being the active tab. A button '+ 인스턴스 생성' (Create Instance) is visible. The main content area displays a table of instances:

인스턴스 ID	유형	IP 주소	고가용성	위치	사용된 저장용량
fctestp	MySQL 2세대	35.192.7.194	추가	us-central1	-

# RMySQL로 연결하기



The screenshot shows the Google Cloud Platform interface for managing SQL instances. The top navigation bar includes tabs for 'SQL 개요' and 'What Is My IP Address'. The main title is '안전함 | https://console.cloud.google.com/sql/instances?project=konlper-168808'. The header features the 'Google Cloud Platform' logo, a search bar, and various navigation icons. Below the header, there are two main sections: 'SQL' and '인스턴스'. The '인스턴스' section contains a button for '인스턴스 생성'. A table lists the existing instance 'fctestp'. The table columns are: 인스턴스 ID, 유형, IP 주소, 고가용성, 위치, and 사용된 저장용량. The 'IP 주소' column shows '35.192.7.194'.

인스턴스 ID	유형	IP 주소	고가용성	위치	사용된 저장용량
fctestp	MySQL 2세대	35.192.7.194	추가	us-central1	-

# 데이터베이스 만들기

The screenshot shows the RStudio interface with the following components:

- File Explorer:** Shows the project structure under "dabrp\_classnote2".
- Code Editor:** Displays an R script with code for database management, including package installation, connection setup, and table creation.
- Console:** Shows the R command-line interface output, including the R version information and the start of the R license notice.

**Code Editor Content:**

```
1 if (!require(devtools)) install.packages("devtools")
2 if (!require(DBI)) devtools::install_github("rstats-db/DBI")
3 if (!require(RSQLite)) devtools::install_github("rstats-db/RSQLite")
4 if (!require(RMySQL)) devtools::install_github("rstats-db/RMySQL")
5
6
7
8 # user<- "root"
9 # pw<- "XXXXXXXXXXXXXX"
10 # host<- 'XXX.XXX.XXX.XXX'
11
12 # save(user,pw,host,file = "./gsql.RData")
13
14 load("./gsql.RData")
15
16 library(RMySQL)
17 con <- dbConnect(MySQL(),
18                   user = user,
19                   password = pw,
20                   host = host,
21                   dbname = "test")
22 dbListTables(conn = con)
23 dbWriteTable(conn = con, name = 'Test', value = as.data.frame(iris))
24 dbReadTable(conn = con, name = "Test")
25
26 dbListTables(conn = con)
```

**Console Output:**

```
R version 3.4.0 (2017-04-21) -- "You Stupid Darkness"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
```

# 데이터베이스 만들기

The screenshot shows the Google Cloud Platform SQL overview page for the 'fctestp' instance. The instance details are as follows:

- Instance Name:** fctestp
- Type:** MySQL 2세대
- Region:** us-central1
- Status:** 켜고(On)
- Storage Type:** 마스터(Master)
- Storage Size:** 1G
- Last Activity:** 6월 28일 오후 12:45
- Current Usage:** 1.14G

The page also includes sections for logs and connection information.

# 데이터베이스 만들기

The screenshot shows the Google Cloud Platform SQL Instances page. The URL in the browser is <https://console.cloud.google.com/sql/instances/fctestp/databases?project=konlper-168808>. The page title is "fctestp 데이터베이스". The main content area displays the "fctestp" instance details, including the database named "fctestp" (MySQL 2세대). The "데이터베이스" tab is selected, showing the "MySQL 데이터베이스" section. A button labeled "데이터베이스 만들기" (Create Database) is highlighted with a red box. Below it, a table lists existing databases:

이름	문자 세트	대조	유형
information_schema	utf8	utf8_general_ci	시스템
mysql	utf8	utf8_general_ci	시스템
performance_schema	utf8	utf8_general_ci	시스템

# 데이터베이스 만들기

The screenshot shows the Google Cloud Platform SQL interface for creating a new database. The main navigation bar at the top includes tabs for 'fctestp 데이터베이스' and 'What Is My IP Address'. The URL in the address bar is <https://console.cloud.google.com/sql/instances/fctestp/databases?project=konlper-168808>. The left sidebar shows the project 'konlper' and the instance 'fctestp' located in 'us-central1'. The main content area displays the 'Instances 세부정보' (Instance Details) page for 'fctestp', which is a MySQL 2世代 instance. Below this, a modal dialog box titled '데이터베이스 만들기' (Create Database) is open. It contains fields for '데이터베이스 이름' (Database Name), which is currently empty, and 'MySQL 스키마 개체 이름 규칙' (MySQL Schema Object Name Rule). It also includes dropdowns for '문자 세트' (Character Set) set to 'utf8' and '대조 (선택사항)' (Collation) set to '기본 대조' (Default Collation). At the bottom of the dialog are '취소' (Cancel) and '만들기' (Create) buttons.

# 데이터베이스 만들기

The screenshot shows the Google Cloud Platform SQL interface for creating a new database. The main navigation bar at the top includes tabs for 'fctestp 데이터베이스' and 'What Is My IP Address'. The URL in the address bar is <https://console.cloud.google.com/sql/instances/fctestp/databases?project=konlper-168808>. The left sidebar shows the project 'konlper' and the instance 'fctestp' located in 'us-central1'. The main content area displays the 'Instances Subinfo' page for 'fctestp', which is a MySQL 2nd Gen instance. A modal window titled '데이터베이스 만들기' (Create Database) is open in the foreground. It contains fields for '데이터베이스 이름' (Database Name), which is set to 'test', '문자 세트' (Character Set), which is 'utf8', and '대조 (선택사항)' (Collation), which is '기본 대조' (Default). Below the modal are buttons for '취소' (Cancel) and '만들기' (Create).

# 데이터베이스 만들기

The screenshot shows the Google Cloud Platform SQL interface. On the left, there's a sidebar with 'SQL' and a list of databases: 'fctestp' (master, us-central1). The main area shows the 'Instances 세부정보' (Instance Details) page for 'fctestp'. A modal window titled '데이터베이스 만들기' (Create Database) is open. In the modal, the '데이터베이스 이름' (Database Name) field contains 'test', the '문자 세트' (Character Set) is set to 'utf8', and the '대조 (선택사항)' (Collation) is set to '기본 대조' (Default Collation). Below the modal, there are buttons for '취소' (Cancel), '만들기' (Create), and '...'. The URL in the browser bar is <https://console.cloud.google.com/sql/instances/fctestp/databases?project=konlper-168808>.

# host와 dbname 설정

The screenshot shows the Google Cloud Platform SQL Instances page. The URL in the browser is <https://console.cloud.google.com/sql/instances/fctestp/databases?project=konlper-168808>. The page title is "fctestp 데이터베이스". The main content area displays the "fctestp" instance details, including the database named "fctestp" (MySQL 2세대). The "데이터베이스" tab is selected, showing the following table:

이름	문자 세트	대조	유형
information_schema	utf8	utf8_general_ci	시스템
mysql	utf8	utf8_general_ci	시스템
performance_schema	utf8	utf8_general_ci	시스템
test	utf8	utf8_general_ci	사용자

# 실습 진행

The screenshot shows the RStudio interface with the following components:

- Top Bar:** Shows the project path (~/project/dabrp\_classnote2 - master) and the R logo.
- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Code Editor:** Displays an R script with code for MySQL database management. The code includes requirements for devtools, DBI, RSQLite, and RMySQL packages, and connects to a MySQL database named 'test'.
- Console:** Shows the R version information and the standard GNU General Public License notice.
- Environment Tab:** Shows the Git status for the project 'dabrp\_classnote2'. A file named 'notegen/class2.Rmd' is listed under 'Staged' changes.
- File Browser:** Shows the project structure with files like .gitignore, codeForClass2.R, dabrp\_classnote2.Rproj, docs, folderForClass1, Materials, notegen, README.md, and recomen.

# R bigquery 연결

`query_exec`가 첫 실행이면 브라우저에서 권한 확인을 합니다.

```
library(bigrquery)
project <- "konlper-168808"
sql <- "SELECT * FROM [konlper-168808:recom.channel] LIMIT 5"
query_exec(sql, project = project)
```

# bigquery 인스턴스 생성

The screenshot shows the Google Cloud Platform (GCP) SQL Instances page. The URL in the browser is <https://console.cloud.google.com/sql/instances/fctestp/databases?project=konlper-168808>. The page title is "fctestp 데이터베이스 - k".

The main content area displays the "Instances 세부정보" (Instance Details) section for the "fctestp" instance. The instance type is "MySQL 2세대". Below this, there are tabs for "주요" (Main), "액세스 제어" (Access Control), "데이터베이스" (Database), "백업" (Backup), "복제본" (Replica), and "작업" (Operations). The "데이터베이스" tab is selected.

In the "MySQL 데이터베이스" section, there is a button labeled "데이터베이스 만들기" (Create Database) which is highlighted with a yellow box. Below this button is a table listing existing databases:

이름	문자 세트	대조	유형
information_schema	utf8	utf8_general_ci	시스템
mysql	utf8	utf8_general_ci	시스템
performance_schema	utf8	utf8_general_ci	시스템
recom	utf8	utf8_general_ci	사용자
test	utf8	utf8_general_ci	사용자

At the bottom right of the table, there are two small trash can icons, likely for deleting databases.

The left sidebar shows the GCP navigation menu with sections like Home, Compute Engine, MySQL 2세대, Container Registry, Docker Registry, Endpoints, Big Data, and BigQuery.

The bottom right corner of the screen shows the page number "73 / 99".

# bigquery 첫 화면

The screenshot shows the Google BigQuery web interface. At the top, there are two tabs: "fctestp 데이터베이스" and "Google BigQuery". The "Google BigQuery" tab is active, indicated by a blue search icon and the URL <https://bigquery.cloud.google.com/welcome/konlper-168808>. The page title is "Google BigQuery". On the left, there's a sidebar with a red "COMPOSE QUERY" button, "Query History", "Job History", a filter input field containing "konlper", and a dropdown menu. Below these, it says "No datasets found in this project. Please create a dataset or select a new project from the menu above." A section titled "Public Datasets" lists several public datasets: "bigquery-public-data:hacker\_news", "bigquery-public-data:noaa\_gsod", "bigquery-public-data:samples", "bigquery-public-data:usa\_names", "gdelt-bq:hathitrustbooks", "gdelt-bq:internetarchivebooks", "lookerdata:cdc", "nyc-tlc:green", and "nyc-tlc:yellow". The main content area features a "Welcome to BigQuery!" heading and a paragraph about the service's capabilities. It also provides instructions for getting started with links to "BigQuery Quickstart guide", "cost control options in BigQuery", and the "BigQuery Web UI guide".

# dataset 만들기

The screenshot shows the Google BigQuery web interface. At the top, there are two tabs: 'fctestp 데이터베이스 - k' and 'Google BigQuery'. The URL in the address bar is <https://bigquery.cloud.google.com/welcome/konlper-168808>. The main content area is titled 'Welcome to BigQuery!'. It explains that BigQuery is a web service for interactive analysis of massive datasets. Below this, there's a list of options to get started, including 'Read our [BigQuery Quickstart guide](#)' and 'Run a query against our sample data by clicking "Compose Query"'. A dropdown menu is open over these options, showing 'Create new dataset', 'Switch to project', and 'Refresh'. On the left sidebar, there's a red button labeled 'COMPOSE QUERY'. Below it are links for 'Query History' and 'Job History'. A search bar is labeled 'Filter by ID or label' with the text 'konlper' entered. Below the search bar, it says 'No datasets found in this project.' and 'Please create a dataset or select a new project from the menu above.' Under the 'Public Datasets' section, there's a list of available datasets: 'bigquery-public-data:hacker\_news', 'bigquery-public-data:noaa\_gsod', 'bigquery-public-data:samples', 'bigquery-public-data:usa\_names', 'gdelt-bq:hathitrustbooks', 'gdelt-bq:internetarchivebooks', 'lookerdata:cdc', 'nyc-tlc:green', and 'nyc-tlc:yellow'.

fctestp 데이터베이스 - k x Google BigQuery x

← → C 🔒 안전함 | https://bigquery.cloud.google.com/welcome/konlper-168808

Google BigQuery

COMPOSE QUERY

Query History

Job History

Filter by ID or label ?

konlper

No datasets found in this project.

Please create a dataset or select a new project from the menu above.

Public Datasets

- ▶ bigquery-public-data:hacker\_news
- ▶ bigquery-public-data:noaa\_gsod
- ▶ bigquery-public-data:samples
- ▶ bigquery-public-data:usa\_names
- ▶ gdelt-bq:hathitrustbooks
- ▶ gdelt-bq:internetarchivebooks
- ▶ lookerdata:cdc
- ▶ nyc-tlc:green
- ▶ nyc-tlc:yellow

Welcome to BigQuery!

Google BigQuery is a web service that lets you do interactive analysis of massive datasets—up to billions of rows. Scalable and easy to use, BigQuery lets developers and businesses tap into powerful data analytics on demand.

To get started, try one of the following options:

- Read our [BigQuery Quickstart guide](#)
- Run a query against our sample data by clicking "Compose Query"
- Create a new dataset and load some of your own data into a table using the  menu on the left
- Learn about [BigQuery](#)
- For more information, see the [BigQuery documentation](#)

Create Dataset

Dataset ID: recom

Data location: (unspecified)

Data expiration: In 60 days.

OK Cancel



# table 생성

Google BigQuery

Compose Query

Query History  
Job History  
Filter by ID or label  
konlper  
recom

Public Datasets  
bigrquery-public-data:hacker\_news  
bigrquery-public-data:noaa\_gsod  
bigrquery-public-data:samples  
bigrquery-public-data:usa\_names  
gdelt-bq:hathitrustbooks  
gdelt-bq:internetarchivebooks  
lookerdata:cdc  
nyc-tlc:green  
nyc-tlc:yellow

## Create Table

Source Data  Create from source  Create empty table

Repeat job Select Previous Job  
Location File upload Choose file No file chosen  
File format CSV

Destination Table  
Table name recom . Destination table name  
Table type Native table

Schema  Automatically detect

Name	Type	Mode
	STRING	NULLABLE

Add Field Edit as Text

Options  
Field delimiter  Comma  Tab  Pipe  Other  
Header rows to skip 0  
Number of errors allowed 0  
Allow quoted newlines

77 / 99

# 큰 데이터는 storage를 통해 업로드

The screenshot shows the 'Create Table' page in the Google BigQuery web interface. The URL in the browser is <https://bigquery.cloud.google.com/createtable/konlper-168808:recom>. The left sidebar shows 'Query History' and 'Job History', and a dropdown for 'konlper' which includes 'recom'. The 'Public Datasets' section lists several datasets like 'bigquery-public-data:hacker\_news' and 'nyc-tlc:green'. The main form is titled 'Create Table' and has the following settings:

- Source Data:** 'Create from source' is selected.
- Repeat job:** 'Select Previous Job' is chosen.
- Location:** 'File upload' is selected, and a file named 'tran.csv (1484069730 bytes)' is chosen.
- File format:** 'CSV' is selected.
- Destination Table:** 'Table name' is set to 'recom' and 'Destination table name' is also 'recom'. 'Table type' is 'Native table'.
- Schema:** 'Automatically detect' is checked.
- Name:** An empty input field.
- Type:** 'STRING'.
- Mode:** 'NULLABLE'.
- Add Field:** A button to add more fields.
- Options:** 'Field delimiter' is 'Comma', 'Header rows to skip' is '0', and 'Number of errors allowed' is '0'.



# 버킷 만들기

The screenshot shows the Google Cloud Platform Storage bucket creation interface. The browser title bar reads "버킷 생성 - konlper" and "Google BigQuery". The URL in the address bar is "https://console.cloud.google.com/storage/create-bucket?project=konlper-168808". The page header includes the Google Cloud Platform logo, project name "konlper", a search bar, and various navigation icons.

**Storage**    [← 버킷 만들기](#)

**브라우저**    이름  [C](#)

**전송**

**설정**

**기본 저장소 클래스** [?](#) [가격 책정 자세히 알아보기](#)

**Multi-Regional**  
동영상을 스트리밍하고 인기 웹 콘텐츠를 호스팅하는 데 사용하세요.  
전 세계에서 자주 액세스하는 데이터에 적합합니다.

**Regional**  
데이터를 저장하고 데이터 분석을 실행하는 데 사용하세요.  
특정 지역에서 자주 액세스하는 데이터에 적합합니다.

**Nearline**  
자주 액세스하지 않는 문서를 보관하는 데 사용하세요.  
액세스 빈도가 한 달에 한 번 미만인 데이터에 적합합니다.

**Coldline**  
거의 액세스하지 않는 문서를 보관하는 데 사용하세요.  
액세스 빈도가 1년에 한 번 미만인 데이터에 적합합니다.

**Multi-Regional 위치**  
선택한 위치 내의 지역 2곳 이상에서 중복됩니다.

[▼](#)

[라벨 지정](#)

[만들기](#) [취소](#)

버킷 생성 - konlper x Google BigQuery x

안전함 | https://console.cloud.google.com/storage/create-bucket?project=konlper-168808

Google Cloud Platform konlper :

Storage ← 버킷 만들기

브라우저 이름 ?  
Cloud Storage 전체에서 고유해야 합니다. 개인정보 보호: 버킷 이름에 민감한 정보를 포함해서는 안 됩니다. 다른 사용자가 사용하려고 시도한 버킷 이름과 일치할 경우 이 사용자에게 버킷 이름이 노출될 수 있습니다.  
yarling-stratum-4912

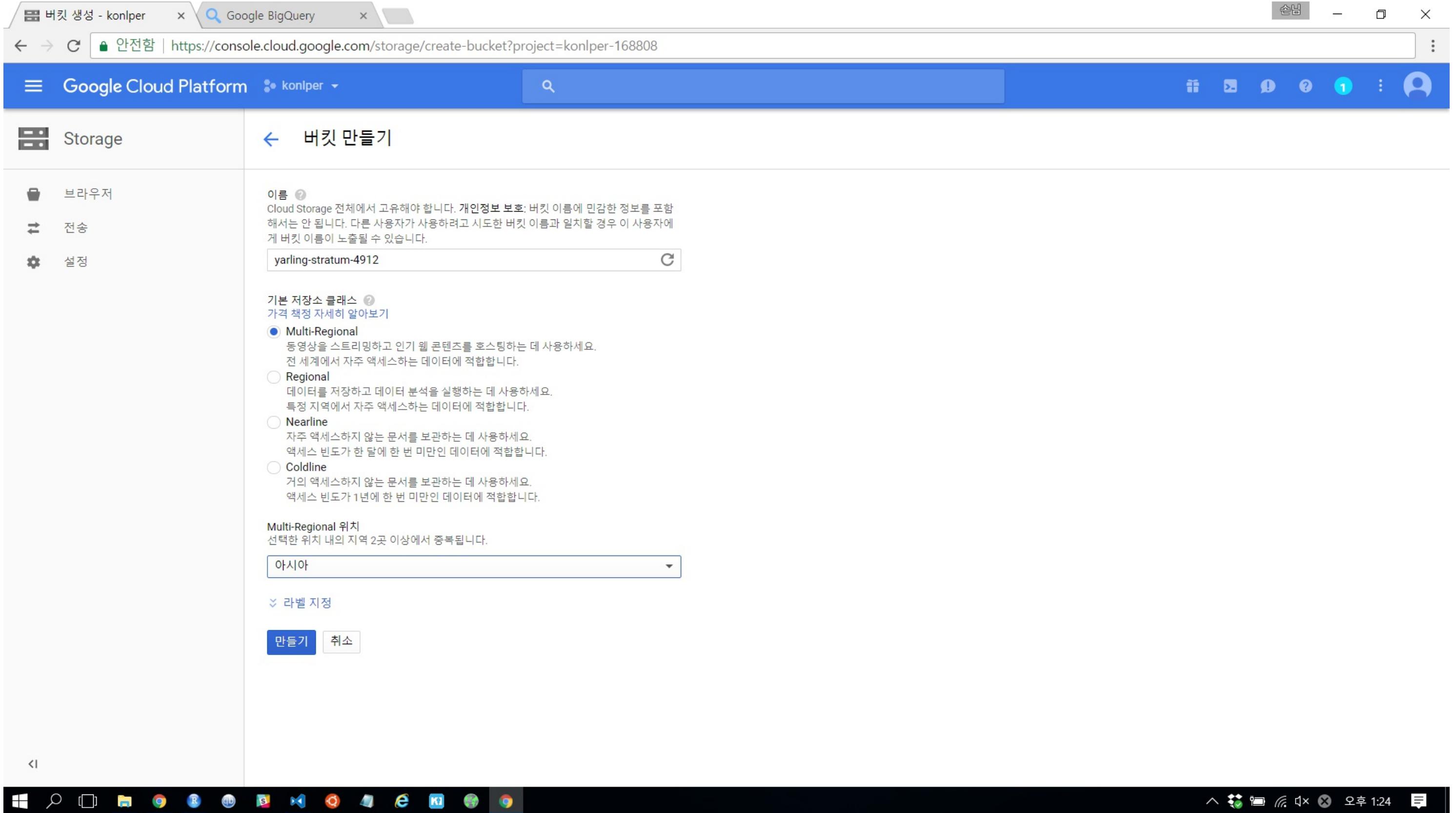
전송 설정

기본 저장소 클래스 ?  
가격 책정 자세히 알아보기  
 Multi-Regional  
동영상을 스트리밍하고 인기 웹 콘텐츠를 호스팅하는 데 사용하세요.  
전 세계에서 자주 액세스하는 데이터에 적합합니다.  
 Regional  
데이터를 저장하고 데이터 분석을 실행하는 데 사용하세요.  
특정 지역에서 자주 액세스하는 데이터에 적합합니다.  
 Nearline  
자주 액세스하지 않는 문서를 보관하는 데 사용하세요.  
액세스 빈도가 한 달에 한 번 미만인 데이터에 적합합니다.  
 Coldline  
거의 액세스하지 않는 문서를 보관하는 데 사용하세요.  
액세스 빈도가 1년에 한 번 미만인 데이터에 적합합니다.

Multi-Regional 위치  
선택한 위치 내의 지역 2곳 이상에서 중복됩니다.  
아시아

▼ 라벨 지정

만들기 취소



# 버킷 확인

The screenshot shows the Google Cloud Platform Storage browser interface. The left sidebar has a 'Storage' section with three items: '브라우저' (selected), '전송', and '설정'. The main area displays a table with one row of data:

버킷	이름	기본 저장소 클래스	위치	라벨	⋮
브라우저	<input type="checkbox"/> yarling-stratum-4912	Multi-Regional	ASIA		

At the top, there are tabs for 'yarling... 버킷 - konlper', 'Google BigQuery', and '브라우저 - konlper'. The address bar shows '안전함 | https://console.cloud.google.com/storage/browser?project=konlper-168808'. The top navigation bar includes icons for Help, Notifications, and User profile.

# 폴더 업로드

The screenshot shows a web browser window with two tabs: 'yarling... 버킷 - konlper' and 'Google BigQuery'. The active tab is 'yarling... 버킷 - konlper', which displays the Google Cloud Platform Storage browser interface. The URL in the address bar is <https://console.cloud.google.com/storage/browser/yarling-stratum-4912/?project=konlper-168808>. The page title is '안전함 | Google Cloud Platform'.

The main content area shows a bucket named 'yarling-stratum-4912'. A message at the top states '이 버킷에 개체가 없습니다.' (No objects in this bucket). On the left, a sidebar menu has '브라우저' selected. At the top right, there are buttons for '파일 업로드' (File upload), '폴더 업로드' (Folder upload), '폴더 생성' (Create folder), '새로고침' (Refresh), '공개적으로 공유하기' (Share publicly), and '삭제' (Delete).

# 업로드 확인

The screenshot shows the Google Cloud Platform Storage browser interface. The URL in the address bar is <https://console.cloud.google.com/storage/browser/yarling-stratum-4912/?project=konlper-168808>. The left sidebar has 'Storage' selected, with '브라우저' (Browser) highlighted. The main area shows a single folder named 'recomen/'. The table below provides details about this folder:

이름	크기	유형	저장소 클래스	최종 수정 시간	공개적으로 공유하기
recomen/	-	폴더	-	-	

# 업로드 확인

The screenshot shows the Google Cloud Storage browser interface. The left sidebar has a 'Storage' section with three items: '브라우저' (selected), '전송', and '설정'. The main area shows a list of files in the 'recomen' bucket of the 'yarling-stratum-4912' project. The files are:

이름	크기	유형	저장소 클래스	최종 수정 시간	공개적으로 공유하기
chennel.csv	184.7KB	application/vnd.ms-excel	Multi-Regional	17. 6. 28. 오후 1:24	<input type="checkbox"/>
competitor.csv	522.51KB	application/vnd.ms-excel	Multi-Regional	17. 6. 28. 오후 1:24	<input type="checkbox"/>
customer.csv	454.13KB	application/vnd.ms-excel	Multi-Regional	17. 6. 28. 오후 1:24	<input type="checkbox"/>
item.csv	192.9KB	application/vnd.ms-excel	Multi-Regional	17. 6. 28. 오후 1:24	<input type="checkbox"/>
membership.csv	177.42KB	application/vnd.ms-excel	Multi-Regional	17. 6. 28. 오후 1:24	<input type="checkbox"/>

# gs:// 경로확인

The screenshot shows the Google BigQuery web interface with the title "Create Table".

**Source Data:** The "Create from source" radio button is selected. Under "Repeat job", "Select Previous Job" is chosen. The "Location" is set to "Google Cloud Storage" with the path "gs://yarling-stratum-4912/recomen/chennel.csv". The "File format" is set to "CSV".

**Destination Table:** The "Table name" is "recom . chennel". The "Table type" is "Native table".

**Schema:** The "Automatically detect" checkbox is checked, with a note below stating "Schema will be automatically generated."

**Options:** The "Number of errors allowed" is set to 0. The "Write preference" is "Write if empty".

**Buttons:** At the bottom left is a blue "Create Table" button.

# 작업 결과 확인

The screenshot shows the Google BigQuery web interface. The top navigation bar has tabs for 'yarling... 버킷 - konlper' and 'Google BigQuery'. The main title bar says '안전함 | https://bigquery.cloud.google.com/jobs/konlper-168808'. The interface includes a 'COMPOSE QUERY' button, a 'Recent Jobs' section, and a 'Job History' sidebar.

**Job History** (selected)

**Recent Jobs**

**Filter jobs**

**Load** gs://yarling-stratum-4912/recomen/chennel.csv to konlper-168808:recom.chennel **Repeat Load Job** 1:34PM

<b>Job ID</b>	konlper-168808:bquijob_6e05e78e_15cecfb07fd
<b>Creation Time</b>	Jun 28, 2017, 1:34:54 PM
<b>Start Time</b>	Jun 28, 2017, 1:34:55 PM
<b>End Time</b>	Jun 28, 2017, 1:34:59 PM
<b>Destination Table</b>	<a href="#">konlper-168808:recom.chennel</a>
<b>Write Preference</b>	Write if empty
<b>Source Format</b>	CSV
<b>Source URI</b>	gs://yarling-stratum-4912/recomen/chennel.csv ( <a href="#">Open in GCS</a> )
<b>Autodetect Schema</b>	true

**Cancel Job**

**Load** gs://yarling-stratum-4912/recom/chennel.csv to konlper-168808:recom.chennel **Repeat Load Job** 1:34PM

**Filter by ID or label**

**konlper**

- recom
- chennel

**Public Datasets**

- bigquery-public-data:hacker\_news
- bigquery-public-data:noaa\_gsod
- bigquery-public-data:samples
- bigquery-public-data:usa\_names
- gdelt-bq:hathitrustbooks
- gdelt-bq:internetarchivebooks
- lookerdata:cdc
- nyc-tlc:green
- nyc-tlc:yellow

# query 실행화면

The screenshot shows the Google BigQuery web interface. On the left, there's a sidebar with 'COMPOSE QUERY' and sections for 'Query History', 'Job History', and 'Filter by ID or label'. Below these are dropdown menus for 'konlper' (selected), 'recom' (expanded), and 'chennel'. A 'Public Datasets' section lists various datasets like 'bigquery-public-data:hacker\_news' and 'bigquery-public-data:usa\_names'. The main area is titled 'New Query' and contains a code editor with the following SQL query:

```
1 | SELECT * FROM [konlper-168808:recom.chennel] LIMIT 100
```

Below the code editor, a green validation message says 'Valid: This query will process 259 KB when run.' There are buttons for 'RUN QUERY' (red), 'Save Query', 'Save View', 'Format Query', and 'Show Options'. A checkmark icon is next to the 'RUN QUERY' button. At the bottom, tabs for 'Results', 'Explanation', and 'Job Information' are visible. The 'Job Information' tab is active, displaying the error message 'Query Failed' and the error details: 'Error: Encountered " "SELECT" " at line 1, column 1. Was expecting: <EOF>' and 'Job ID: konlper-168808:bquijob\_3ada8983\_15cecfbea1b'. A note at the bottom says 'Note: You can also find errors in your queries before running them. Click the ! below the query composition box to enable real-time validation.'

# 비용 발생 경고

The screenshot shows the Google BigQuery web interface. On the left, there's a sidebar with 'COMPOSE QUERY' and sections for 'Query History', 'Job History', and 'Filter by ID or label'. Below these are dropdown menus for 'konlper' (selected), 'recom' (expanded), and 'chennel'. Under 'Public Datasets', several options like 'bigquery-public-data:hacker\_news' are listed. The main area shows a 'New Query' window with the following SQL code:

```
1 SELECT * FROM [konlper-168808:recom.chennel] LIMIT 100
```

Below the code, a message says 'Valid: This query will process 259'. There are 'RUN QUERY' and 'Save Query' buttons, and tabs for 'Results', 'Explanation', and 'Job Inform'. A modal dialog titled 'Confirm query' is open, containing the following text:

With this query, you will be billed for all the data in the table (even if your query contains a LIMIT clause). If you're using the free tier, this query still counts against your free quota.

You can use table preview instead to see records for free and without affecting quotas.

Don't show this again

At the bottom of the modal are 'Run query', 'Go to table preview', and 'Cancel' buttons. A note at the bottom of the main page says: 'Note: You can also find errors in your queries before running them. Click the ! below the query composition box to enable real-time validation.'

# query 결과 확인

The screenshot shows the Google BigQuery results page. The URL in the browser is [https://bigquery.cloud.google.com/results/konlper-168808:bquijob\\_1f8d0e04\\_15cefc9fa5](https://bigquery.cloud.google.com/results/konlper-168808:bquijob_1f8d0e04_15cefc9fa5). The main content area displays a query result table with the following data:

Row	cusID	chennel	useCnt
1	14	A_MOBILE/APP	1
2	74	A_MOBILE/APP	1
3	241	A_MOBILE/APP	1
4	304	A_MOBILE/APP	1
5	326	A_MOBILE/APP	1
6	377	A_MOBILE/APP	1
7	448	A_MOBILE/APP	1
8	518	A_MOBILE/APP	1
9	549	A_MOBILE/APP	1

# 공개 데이터셋

The screenshot shows a web browser window with four tabs open:

- yarling... 버킷 - konlper
- Google BigQuery
- yarling... 버킷
- Google Cloud Platform
- Google BigQuery Public

The main content area is the "Documentation" page for "Google BigQuery Public Datasets". The URL in the address bar is <https://cloud.google.com/bigquery/public-data/>.

The page includes a sidebar with links to various resources, a main content area with a heading and a list of datasets, and a right sidebar with a list of public datasets.

**Left Sidebar:**

- All Resources
  - Pricing and Quotas
  - Release Notes
  - Support
  - Public Datasets
    - 개요
    - 1000 Cannabis Genomes Project
    - Bay Area Bike Share Trips Data
    - Chicago Crime Data
    - Chicago Taxi Trips
    - EPA Historical Air Quality Data
    - GDELT Books Corpus
    - GitHub Data
    - Hacker News
    - Healthcare Common Procedure Coding System (HCPCS) Level II
    - IRS 990 Data
    - Major League Baseball
    - Medicare
    - NHTSA Traffic Fatality Data
    - NOAA GHCN Weather
    - NOAA GSOD Weather
    - NOAA ICOADS
    - NYC 311 Service Requests

**Main Content Area:**

BigQuery > Documentation

## Google BigQuery Public Datasets

A public dataset is any dataset that is stored in BigQuery and made available to the general public. This page lists a special group of public datasets that Google BigQuery hosts for you to access and integrate into your applications. Google pays for the storage of these data sets and provides public access to the data via BigQuery. You pay only for the queries that you perform on the data (the first 1 TB per month is free, subject to [query pricing details](#)).

**Public datasets hosted by BigQuery**

- 1000 Cannabis Genomes Project**  
Genomic open dataset of approximately 850 strains of Cannabis via the Open Cannabis Project.
- Bay Area Bike Share Trips**  
This data includes all Bay Area Bike Share trips from August 2013 to the present, and is updated daily.
- Chicago Crime Data**

**Right Sidebar:**

목차

- Public datasets hosted by BigQuery
  - 1000 Cannabis Genomes Project
  - Bay Area Bike Share Trips
  - Chicago Crime Data
  - Chicago Taxi Trips
  - EPA Historical Air Quality Data
  - GDELT Book Corpus
  - GitHub Data
  - Hacker News
  - Healthcare Common Procedure Coding System (HCPCS) Level II
  - IRS Form 990 Data
  - Major League Baseball Data
  - Medicare Data
  - NHTSA Traffic Fatality Data
  - NOAA GHCN
  - NOAA GSOD

# wikipedia dataset

The screenshot shows the Google BigQuery web interface. On the left, there's a sidebar with navigation links like 'COMPOSE QUERY', 'Query History', 'Job History', and a 'konlper' dropdown. Below that is a 'Public Datasets' section with links to datasets such as 'bigquery-public-data:hacker\_news', 'bigquery-public-data:noaa\_gsod', 'bigquery-public-data:samples', and 'wikipedia'. The 'wikipedia' link is highlighted with a red vertical bar.

In the main area, there's a 'New Query' section with a single query listed:

```
1 | SELECT * FROM [konlper-168808:recom.chennel] LIMIT 100
```

Below the query is a 'RUN QUERY' button. To the right of the query, there are tabs for 'Query Editor' and 'UDF Editor', with 'SQL' selected.

A modal dialog titled 'Create Dataset' is open in the center. It contains fields for 'Dataset ID' (set to 'test'), 'Data location' (set to '(unspecified)'), and 'Data expiration' (set to 'In 5 days'). There are 'OK' and 'Cancel' buttons at the bottom of the dialog.

At the bottom of the page, there's a table showing schema details for the 'wikipedia' dataset:

title	id	language	wp_namespace
	INTEGER	NULLABLE	REQUIRED
	STRING	REQUIRED	REQUIRED

Descriptions for the columns include:

- 'id': A unique ID for the article that was revised. These correspond to the order in which articles were created, except for the first several thousand IDs, which are issued in alphabetical order.
- 'language': Empty in the current dataset.
- 'wp\_namespace': Wikipedia segments its pages into namespaces (e.g. "Talk", "User", etc.).
- 'wp\_namespace' (description): MEDIA = 202; // =-2 in WP XML, but these values must be >0  
SPECIAL = 201; // =-1 in WP XML, but these values must be >0  
MAIN = 0;  
TALK = 1;

On the far right, there are buttons for 'Query Table', 'Copy Table', 'Export Table', and 'Delete Table'. A green checkmark icon is visible near the bottom right of the interface.

# dataset 복사

The screenshot shows the Google BigQuery web interface. On the left, there is a sidebar with navigation links like 'COMPOSE QUERY', 'Query History', 'Job History', and a 'konlper' project section containing 'recom', 'chennel', and 'test' datasets. Below this is a 'Public Datasets' section listing 'bigquery-public-data:hacker\_ne...', 'bigquery-public-data:noaa\_gsod', and 'bigquery-public-data:samples'. The 'samples' dataset is currently selected, as indicated by the red border around its name.

In the main area, a 'New Query' window is open with the following SQL code:

```
1 | SELECT * FROM [konlper-168808:recom.chennel] LIMIT 100
```

A green validation message at the bottom of the query editor says: "Valid: This query will process 259 KB when run."

Below the query editor, there are buttons for 'RUN QUERY', 'Save Query', 'Save View', 'Format Query', and 'Show Options'. A status message indicates: "Query complete (2.1s elapsed, 259 KB processed)" with a checkmark icon.

On the right, the 'Table Details: wikipedia' section is displayed. It shows three columns: 'Schema', 'Details', and 'Preview'. The 'Schema' tab is active, showing the table structure:

title	STRING	REQUIRED	The title of the page, as displayed on the page (not in the URL). Always starts with a capital letter and may begin with a namespace (e.g. "Talk:", "User:", "User Talk:", ... )
id	INTEGER	NULLABLE	A unique ID for the article that was revised. These correspond to the order in which articles were created, except for the first several thousand IDs, which are issued in alphabetical order.
language	STRING	REQUIRED	Empty in the current dataset.

Below the schema, there are options to 'Copy table', 'Export table', and 'Delete table'. The 'Copy table' option is highlighted with a red border.

At the bottom right of the interface, there is a page number: "93 / 99".

yarling... 버킷 - konlper x Google BigQuery x yarling... 버킷 x Google Cloud Platform x Google BigQuery Public x NOAA Global Surface S... x

안전함 | https://bigquery.cloud.google.com/table/bigquery-public-data:samples.wikipedia

# Google BigQuery

COMPOSE QUERY

New Query ?

1 SELECT \* FROM [konlper-168808:recom.chennel] LIMIT 100

Query Editor UDF Editor X

SQL

Query History

Job History

Filter by ID or label ?

konlper

recom

chennel

test

Public Datasets

bigquery-public-data:hacker\_news

bigquery-public-data:noaa\_gsod

bigquery-public-data:samples

github\_nested

github\_timeline

gsod

nativity

shakespeare

trigrams

wikipedia

bigquery-public-data:usa\_names

gdelt-bq:hathitrustbooks

gdelt-bq:internetarchivebooks

lookerdata:cdr

nyc-tlc:green

Valid: This query is valid.

RUN QUERY

Copy Table

Destination project: konlper (konlper-168808)

Destination dataset: test

Destination table: wikipedia\_copy

OK Cancel

Table Data

Schema

title

id: INTEGER NULLABLE A unique ID for the article that was revised. These correspond to the order in which articles were created, except for the first several thousand IDs, which are issued in alphabetical order.

language: STRING REQUIRED Empty in the current dataset.

wp\_namespace: INTEGER REQUIRED Wikipedia segments its pages into namespaces (e.g. "Talk", "User", etc.)  
MEDIA = 202; // =-2 in WP XML, but these values must be >0  
SPECIAL = 201; // =-1 in WP XML, but these values must be >0  
MAIN = 0;  
TALK = 1;  
USER = 2;  
NSER\_TALK = 2;

is\_redirect: BOOLEAN NULLABLE Versions later than ca. 200908 may have a redirection marker in the XML.

Query Table Copy Table Export Table Delete Table



# 데이터 구조 보기

The screenshot shows the Google BigQuery web interface. The top navigation bar includes tabs for 'yarling... 버킷 - konlper' (selected), 'Google BigQuery' (highlighted with a blue search icon), 'yarling... 버킷', 'Google Cloud Platform' (highlighted with a red play button icon), 'Google BigQuery Public', and 'NOAA Global Surface'. The address bar shows the URL [https://bigquery.cloud.google.com/table/konlper-168808:test.wikipedia\\_copy](https://bigquery.cloud.google.com/table/konlper-168808:test.wikipedia_copy). The main area is titled 'New Query' with a question mark icon. On the left, there's a sidebar with 'COMPOSE QUERY' and sections for 'Query History', 'Job History', 'Filter by ID or label' (with a dropdown showing 'konlper' expanded to 'recom', 'chennel', 'test', and 'wikipedia\_copy'), and 'Public Datasets' (listing 'bigquery-public-data:hacker\_ne...', 'bigquery-public-data:noaa\_gsod', 'bigquery-public-data:samples' (expanded to 'github\_nested', 'github\_timeline', 'gsod', 'natality', 'shakespeare', 'trigrams', 'wikipedia'), 'bigquery-public-data:usa\_names', and 'gdelt\_bg:bohitrustbooks'). The main query editor has a SQL tab selected, showing the query: '1 | SELECT \* FROM [konlper-168808:test.wikipedia\_copy] LIMIT 100'. Below the query, a green box says 'Valid: This query will process 35.7 GB when run.' At the bottom are buttons for 'RUN QUERY' (red), 'Save Query', 'Save View', 'Format Query', and 'Show Options'. A checkmark icon is on the right. The bottom section displays the schema for the 'wikipedia\_copy' table:

Field	Type	Nullable	Description
contributor_id	INTEGER	NULLABLE	Typically, either (_id and _username) or _ip will be set. A (very) small fraction of edits have neither _ip or (_id and _username). They show up on Wikipedia as "(Username or IP removed)".
contributor_username	STRING	NULLABLE	Typically, either (_id and _username) or _ip will be set. A (very) small fraction of edits have neither _ip or (_id and _username). They show up on Wikipedia as "(Username or IP removed)".
timestamp	INTEGER	REQUIRED	In Unix time, seconds since epoch.
is_minor	BOOLEAN	NULLABLE	Corresponds to the "Minor Edit" checkbox on Wikipedia's edit page.
is_bot	BOOLEAN	NULLABLE	A special flag that some of Wikipedia's more active bots voluntarily set.
reversion_id	INTEGER	NULLABLE	If this edit is a reversion to a previous edit, this field records the revision_id that was reverted to. If the same article text occurred multiple times, then this will point to the earliest revision. Only revisions with greater than fifty characters are considered for this field. This is to avoid labeling multiple blankings as reversions.
comment	STRING	NULLABLE	Optional user-supplied description of the edit. Section edits are, by default, prefixed with /* Section Name */.

# 큰 데이터 query 진행

```
select title,sum(num_characters) as num_characters
from [konlper-168808:test.wikipedia_copy]
where regexp_match(title,'[ss]eoul')
group by title
order by num_characters desc;
```

# query 결과

The screenshot shows the Google BigQuery interface with a query results page. The query is:

```
1 select title,sum(num_characters) as num_characters
2 from [konlper-168808:test.wikipedia_copy]
3 where regexp_match(title,'[Ss]eoul')
4 group by title
5 order by num_characters desc;
```

The results table shows the top 10 Korean titles and their character counts:

Row	title	num_characters
1	Seoul	104656429
2	User talk:JohnnySeoul	78435528
3	FC Seoul	44944412
4	Seoul National University	18422979
5	Talk:Seoul/Archive2	9263684
6	Seoul Subway Line 1	9074261
7	Seoul Foreign School	7576768
8	Seoul Metropolitan Subway	5946441
9	Seoul International School	4911143

# 과제

1. RSQLite와 DBI를 활용해서 `nycflights13` 데이터를 db Table로 만들고, 제출해 주세요.
  1. '`nycflights13`' 패키지를 설치하고 5개 데이터를 확인하세요.
  2. `dbConnect` 명령으로 SQLite 파일을 `sql_[이름].db` 으로 생성하세요.
  3. 5개의 데이터를 각각의 이름으로 `table`을 생성하세요.
  4. 1)~3)의 과정을 모두 코드로 남기고 `run_[이름].R`로 저장하세요.
  5. `run[이름].R`과 `sql[이름].db` 두 개의 파일을 `class2assignment` 폴더에 저장하세요.
  6. github에서 pull request로 제출해 주세요.
  7. `recomen` 폴더에 있는 데이터 6개도 같은 과정을 진행하고, 제출은 하지 마세요.
  8. `.gitignore`를 이용하면 push의 범위에서 제외할 수 있습니다.

# 과제

1. recomen 폴더에 있는 데이터 6개를 bigquery에 업로드하고 질의를 실행해 보세요.
  1. "bigrquery" 패키지의 기능으로 작은 용량 5개의 데이터를 업로드해 주세요.
  2. Storage 서비스를 이용해서 `tran.csv` 파일을 bigquery에 테이블로 생성하세요.
  3. `query_exec` 함수와 "select \* from [tran] limit 10" 을 실행하고 결과를 받으세요.
    - [tran]은 각자 해당하는 이름으로 변경하셔야 합니다.
4. 1)~3)의 과정을 `big_[이름].R`로 저장하세요.
5. `big_[이름].R` 파일을 class2assignment 폴더에 저장하세요.
6. github에서 pull request로 제출해 주세요.