



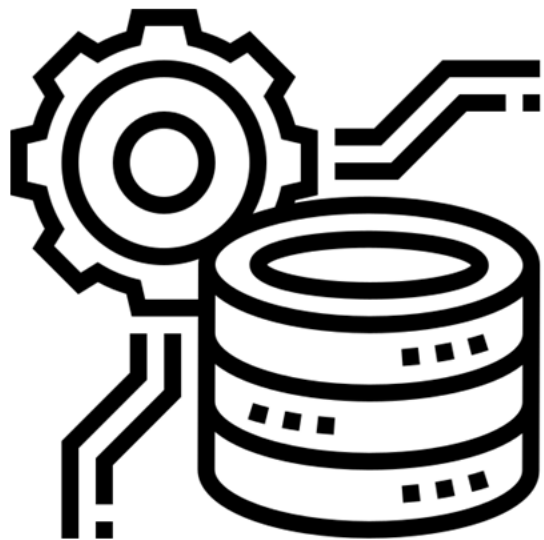
# About Our Team



**“실제 회사같은 팀을 만들어 보자”**

지향점과 방향성이 다른 사람들끼리 모여 프로젝트의 기반을 함께 준비하고,  
그 기반을 바탕으로 각자의 목표점을 잡아 분야를 확장하는 것

# Our Goal



## ETL Process

Extract Transform Load

프로세스의 End to End에 대한 경험

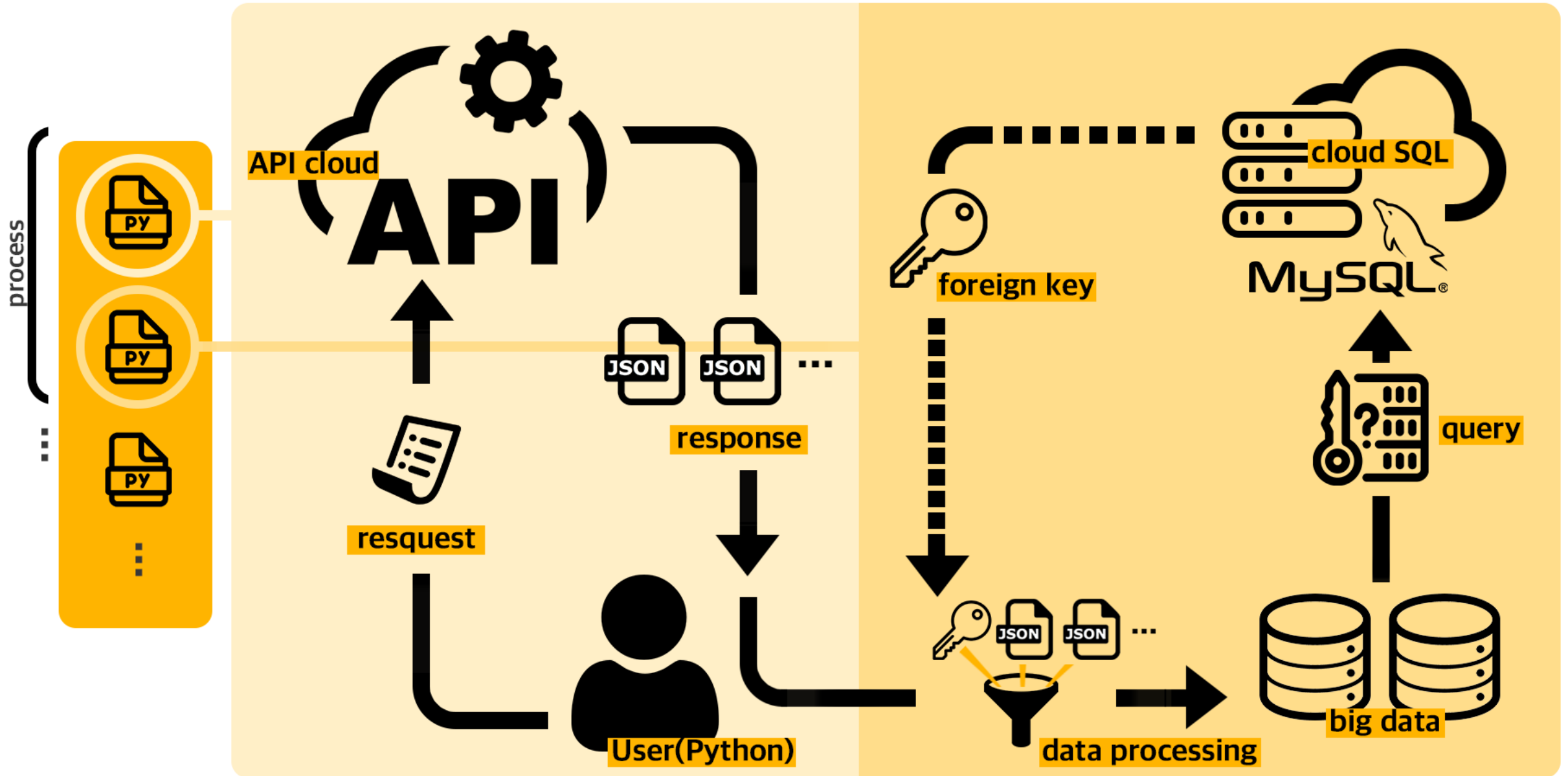


## Infrastructure

Limitation Overcome


워크스페이스가 가지는 한계의 극복

# Database Pipeline Diagram



# Process 1

데이터 수집을 위한 플랫폼 선정



Kaggle

Google Dataset Search

World Bank Open Data

CI Machine Learning Repository

Quandl

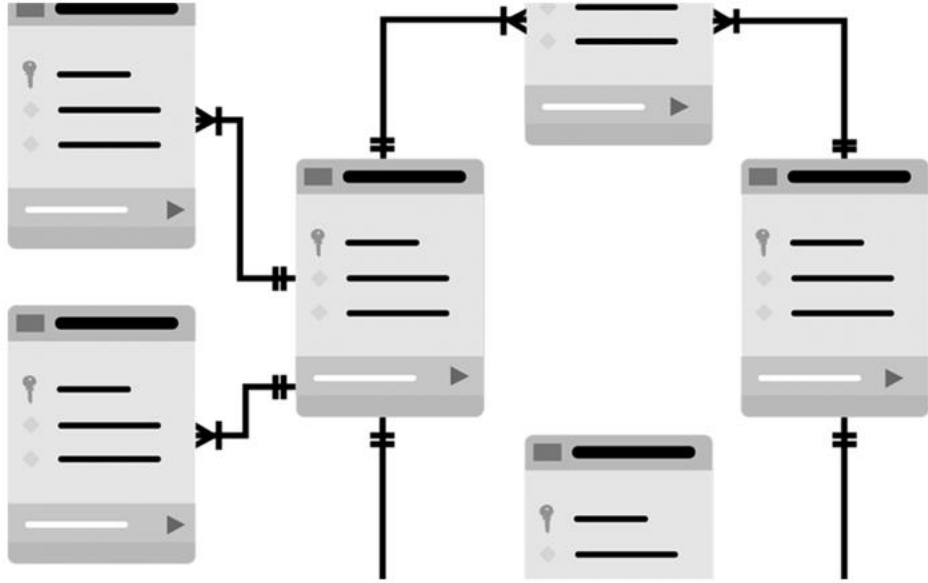
Amazon Web Services

**“실시간 데이터를 서빙해주는 곳이 없을까?”**



# Process 1

데이터 수집을 위한 플랫폼 선정

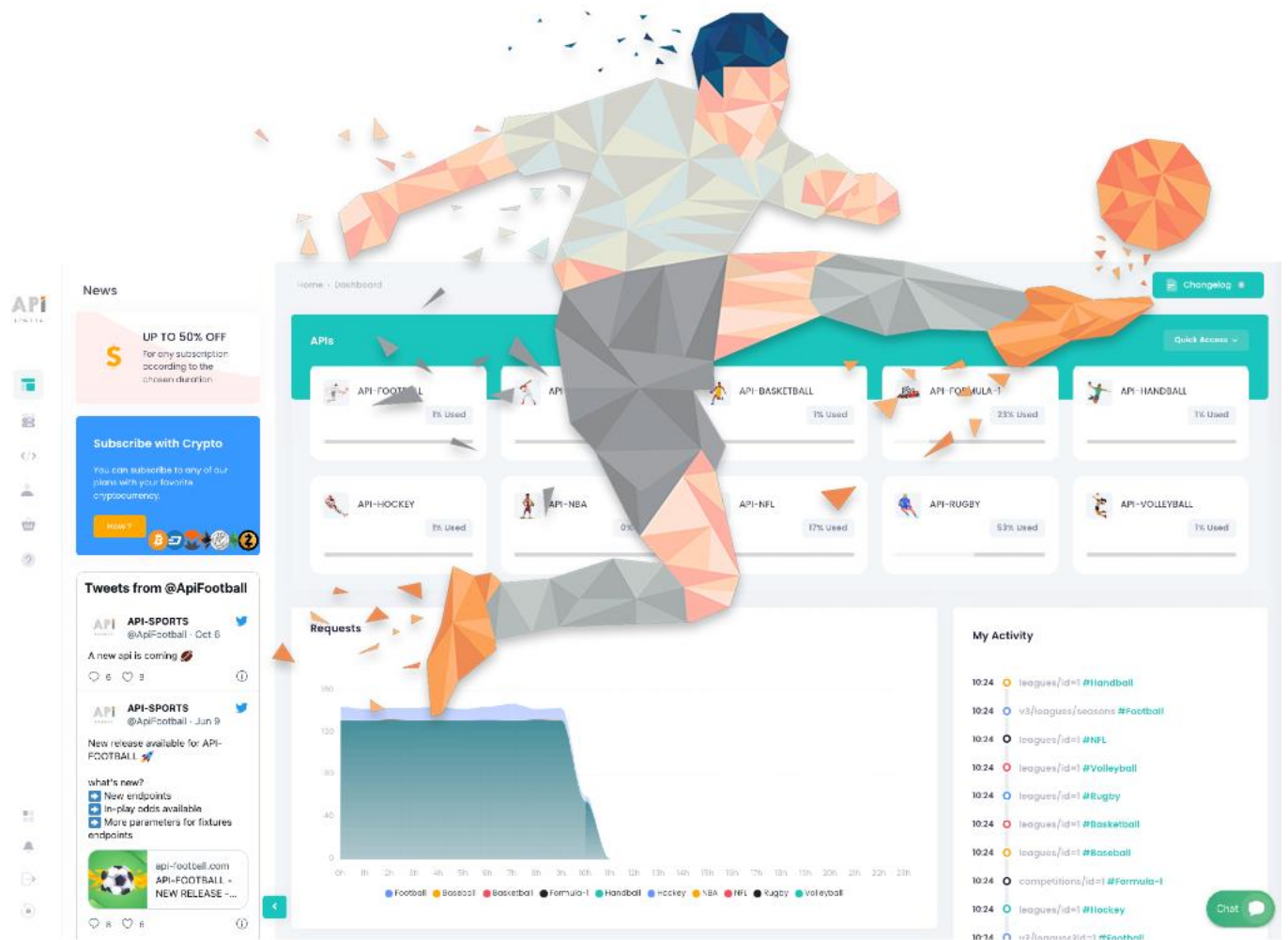


## API 서비스

JSON 파일을 통한 데이터 수집

### API 서비스 플랫폼 | “API-SPORTS”

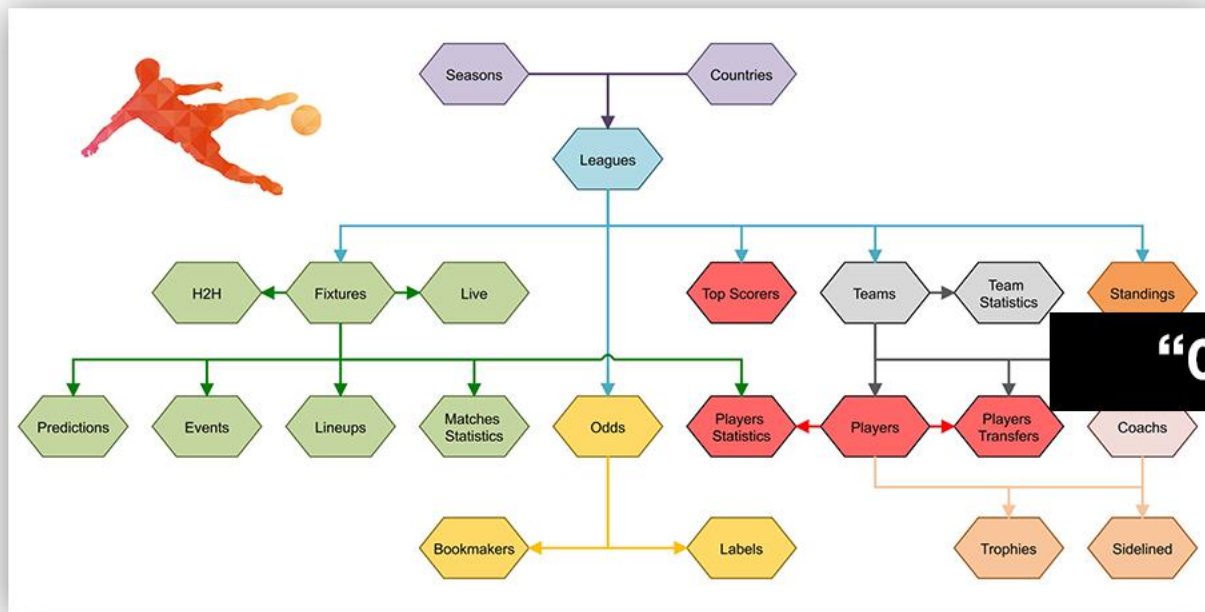
API 데이터의 가공 및 데이터베이스로의 적재를 ‘프로세스화’ 함으로서  
독자적인 실시간 서비스 시스템 개발을 계획





# Process 2

데이터 적재를 위한 데이터베이스 구조도 설계



“어떻게 데이터를 관리할 수 있을까?”



# Process 2

데이터 적재를 위한 데이터베이스 구조도 설계

| 행정구역        |              |             |          |
|-------------|--------------|-------------|----------|
| Nation_Info |              |             |          |
| 국가 ID       | nation_id    | integer(10) | NOT NULL |
| 국가 이름       | country_id   | int         | NOT NULL |
| 행정구역 도      | nation_state | varchar(30) | NOT NULL |
| 행정구역 도시     | nation_city  | varchar(30) | NOT NULL |

| 포메이션 정보        |                |            |          |
|----------------|----------------|------------|----------|
| Formation_Info |                |            |          |
| 포메이션 ID        | formation_id   | integer(1) | NOT NULL |
| 포메이션 이름        | formation_name | varchar    | NOT NULL |

| 국가정보         |              |             |          |
|--------------|--------------|-------------|----------|
| Country_Info |              |             |          |
| 국가 ID        | country_id   | int         | NOT NULL |
| 국가 이름        | country_name | VARCHAR(30) | NOT NULL |

| 라운드정보      |           |                  |          |
|------------|-----------|------------------|----------|
| Round_Info |           |                  |          |
| 라운드 ID     | round_id  | integer(20)      | NOT NULL |
| 팀 ID       | team_id   | integer(10)      | NOT NULL |
| 홈          | home_team | varchar(30)      | NOT NULL |
| 아웨이팀       | away_team | varchar(30)      | NOT NULL |
| 메치결과       | result    | 데이터 값에 따른 분할도 고려 | NOT NULL |
| 패스수        | passes    | integer(4)       | NOT NULL |
| 슈팅수        | shots     | integer(3)       | NOT NULL |

| API 팀 ID         |             |             |          |
|------------------|-------------|-------------|----------|
| hidden_league_id |             |             |          |
| 팀 ID             | team_id     | int         | NOT NULL |
| 팀 이름             | team_name   | varchar(30) | NOT NULL |
| API 팀 ID         | api_team_id | int         | NOT NULL |

| API ID 선수 ID     |               |             |          |
|------------------|---------------|-------------|----------|
| hidden_player_id |               |             |          |
| 선수 ID            | player_id     | int         | NOT NULL |
| 선수 이름            | player_name   | varchar(50) | NOT NULL |
| API 선수 ID        | api_player_id | int         | NOT NULL |

| 시즌      |                |            |          |
|---------|----------------|------------|----------|
| Seasons |                |            |          |
| 시즌 ID   | season_id      | integer(2) | NOT NULL |
| 시즌년도    | year_of_season | integer(4) | NOT NULL |

| 리그정보        |                    |             |          |
|-------------|--------------------|-------------|----------|
| League_Info |                    |             |          |
| 리그 ID       | league_id          | integer(3)  | NOT NULL |
| 시즌 ID       | season_id          | integer(10) | NOT NULL |
| 국가 ID       | country_id         | int         | NOT NULL |
| 리그 이름       | league_name        | varchar(20) | NOT NULL |
| 리그 창단일      | league_created     | integer(4)  | NOT NULL |
| 리그별 소속 팀수   | league_team_count  | integer(3)  | NOT NULL |
| 리그별라운드수     | league_round_count | integer(3)  | NOT NULL |

| 팀정보       |                 |              |          |
|-----------|-----------------|--------------|----------|
| Team_Info |                 |              |          |
| 팀 ID      | team_id         | integer(10)  | NOT NULL |
| 리그 ID     | league_id       | integer(3)   | NOT NULL |
| 시즌 ID     | season_id       | integer(10)  | NOT NULL |
| 포메이션 ID   | formation_id    | integer(1)   | NOT NULL |
| 팀 이름      | team_name       | varchar(30)  | NOT NULL |
| 구장 이름     | stadium_name    | varchar(50)  | NOT NULL |
| 구장 주소     | stadium_address | varchar(100) | NOT NULL |
| 창단일       | team_est_date   | integer(4)   | NOT NULL |
| 구단주       | owner           | varchar(50)  | NOT NULL |
| 감독        | head_coach      | varchar(50)  | NOT NULL |

| 개인 스탯          |              |             |          |
|----------------|--------------|-------------|----------|
| Personal Stats |              |             |          |
| 개인 스탯 ID       | stats_id     | integer(30) | NOT NULL |
| 시즌 ID          | season_id    | integer(10) | NOT NULL |
| 선수 ID          | player_id    | integer(20) | NOT NULL |
| 라운드 ID         | round_id     | integer(20) | NOT NULL |
| 골              | goals_scored | integer(2)  | NOT NULL |
| 어시스트           | assists_made | integer(2)  | NOT NULL |
| 퇴장여부           | outs         |             | NOT NULL |

| 우승팀    |           |             |          |
|--------|-----------|-------------|----------|
| Winner |           |             |          |
| 우승 ID  | winner_id | integer(5)  | NOT NULL |
| 시즌 ID  | season_id | integer(10) | NOT NULL |

| 선수 정보       |                |              |          |
|-------------|----------------|--------------|----------|
| Player_Info |                |              |          |
| 선수 ID       | player_id      | integer(20)  | NOT NULL |
| 시즌 ID       | season_id      | integer(10)  | NOT NULL |
| 팀 ID        | team_id        | integer(10)  | NOT NULL |
| 이름          | first_name     | varchar(20)  | NOT NULL |
| 성           | last_name      | varchar(20)  | NOT NULL |
| 등번호         | uniform_number | int(2)       | NOT NULL |
| 생년월일        | date_of_birth  | varchar(20)  | NOT NULL |
| 주르쓰는발       | main_feet      | varchar(2)   | NOT NULL |
| 키           | height         | int(3)       | NOT NULL |
| 몸무게         | weight         | int(3)       | NOT NULL |
| 포지션         | position       | varchar(20)  | NOT NULL |
| 선수 국적       | player_nation  | varchar(20)  | NOT NULL |
| 선수 사진       | player_picture | varchar(100) | NOT NULL |

| 실시간 순위 정보 |          |             |          |
|-----------|----------|-------------|----------|
| Rank_Info |          |             |          |
| 랭크 ID     | rank_id  | int         | NOT NULL |
| 팀 ID      | team_id2 | integer(10) | NOT NULL |
| 순위정보      | standing | int         | NOT NULL |

## 데이터베이스 구조도 설계

순차식 구성을 통한 트리 구조를 갖춘 데이터베이스 설계

상위 테이블 정보를 참조하는 하강식 테이블 구조 사용

당근마켓 등의 기업 데이터베이스 구조도를 참고하여 데이터테이블 구조도 자체 설계



# Process 2

데이터 적재를 위한 데이터베이스 구조도 설계

| 행정구역    |              | Nation_Info |          |  |
|---------|--------------|-------------|----------|--|
| 국가 ID   | nation_id    | integer(10) | NOT NULL |  |
| 국가 ID   | country_id2  | int         | NOT NULL |  |
| 행정구역 도  | nation_state | varchar(30) | NOT NULL |  |
| 행정구역 도시 | nation_city  | varchar(30) | NOT NULL |  |

| 포메이션 정보 |                | Formation_Info |          |  |
|---------|----------------|----------------|----------|--|
| 포메이션 ID | formation_id   | integer(1)     | NOT NULL |  |
| 포메이션 이름 | formation_name | varchar        | NOT NULL |  |

| API 리그 ID |               | hidden_league_id |          |  |
|-----------|---------------|------------------|----------|--|
| 리그 ID     | league_id     | int              | NOT NULL |  |
| 리그 이름     | league_name   | varchar(30)      | NOT NULL |  |
| API 리그 ID | api_league_id | int              | NOT NULL |  |

| 국가정보  |              | Country_info |          |  |
|-------|--------------|--------------|----------|--|
| 국가 ID | country_id   | int          | NOT NULL |  |
| 국가 이름 | country_name | VARCHAR(30)  | NOT NULL |  |

| 라운드 정보 |           | Round_Info       |          |  |
|--------|-----------|------------------|----------|--|
| 라운드 ID | round_id  | integer(20)      | NOT NULL |  |
| 팀 ID   | team_id   | integer(10)      | NOT NULL |  |
| 홈      | home_team | varchar(30)      | NOT NULL |  |
| 아웨이팀   | away_team | varchar(30)      | NOT NULL |  |
| 매치결과   | result    | 데이터 값에 따른 분할도 고려 | NOT NULL |  |
| 패스수    | passes    | integer(4)       | NOT NULL |  |
| 슈팅수    | shots     | integer(3)       | NOT NULL |  |

| API 팀 ID |             | hidden_team_id |          |  |
|----------|-------------|----------------|----------|--|
| 팀 ID     | team_id     | int            | NOT NULL |  |
| 팀 이름     | team_name   | varchar(30)    | NOT NULL |  |
| API 팀 ID | api_team_id | int            | NOT NULL |  |

“테이블을 꼭 분할해서 다루어야 할까?”

데이터트리를 통해 얻을 수 있는 장점들

**데이터의 유연성** : 관리 및 복잡한 연산의 처리

**유지 보수의 편의** : 대량의 데이터를 적은 리소스로 관리

**데이터베이스의 확장성** : 서버의 부하를 분산

**데이터의 보안성** : 접근 권한의 설정 및 제한

| 우승팀   |              | Winner      |          |  |
|-------|--------------|-------------|----------|--|
| 우승 ID | winner_id    | integer(5)  | NOT NULL |  |
| 팀 ID  | team_id      | integer(10) | NOT NULL |  |
| 우승 연도 | winning_year | 아직 가용연한     | NOT NULL |  |

| 선수 정보 |                | Player_Info  |          |  |
|-------|----------------|--------------|----------|--|
| 시즌 ID | season_id      | integer(10)  | NOT NULL |  |
| 팀 ID  | team_id        | integer(10)  | NOT NULL |  |
| 이름    | first_name     | varchar(20)  | NOT NULL |  |
| 생년월일  | date_of_birth  | varchar(10)  | NOT NULL |  |
| 주포스닝  | main_feet      | varchar(2)   | NOT NULL |  |
| 키     | height         | int(3)       | NOT NULL |  |
| 무게    | weight         | int(3)       | NOT NULL |  |
| 포지션   | position       | varchar(20)  | NOT NULL |  |
| 선수 국가 | player_nation  | varchar(20)  | NOT NULL |  |
| 선수 사진 | player_picture | varchar(100) | NOT NULL |  |

| 개인 스터    |              | Personal_Stats |          |  |
|----------|--------------|----------------|----------|--|
| 개인 스터 ID | stats_id     | integer(30)    | NOT NULL |  |
| 시즌 ID    | season_id    | integer(10)    | NOT NULL |  |
| 선수 ID    | player_id    | integer(20)    | NOT NULL |  |
| 라운드 ID   | round_id     | integer(20)    | NOT NULL |  |
| 골        | goals_scored | integer(2)     | NOT NULL |  |
| 어시스트     | assists_made | integer(2)     | NOT NULL |  |
| 퇴장여부     | outs         |                | NOT NULL |  |

| 실시간 순위 정보 |          | Rank_Info   |          |  |
|-----------|----------|-------------|----------|--|
| 랭크 ID     | rank_id  | int         | NOT NULL |  |
| 팀 ID      | team_id2 | integer(10) | NOT NULL |  |
| 순위정보      | standing | int         | NOT NULL |  |

| API 선수 ID |               | hidden_player_id |          |  |
|-----------|---------------|------------------|----------|--|
| 선수 ID     | player_id     | int              | NOT NULL |  |
| 선수 이름     | player_name   | varchar(50)      | NOT NULL |  |
| API 선수 ID | api_player_id | int              | NOT NULL |  |

# Process 2

데이터 적재를 위한 데이터베이스 구조도 설계

Anaconda Prompt (anaconda3) - mysql -h 34.64.214.96 -u root -p

([HOON]) C:\Users\WUSER>mysql -h 34.64.214.96 -u root -p

Enter password: \*\*\*\*\*

Welcome to the MySQL monitor. Commands end with ; or \g

Your MySQL connection id is 8118

Server version: 8.0.26-google (Google)

Copyright (c) 2000, 2023, Oracle and/or its affiliates.

Oracle is a registered trademark of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.

Type 'help;' or '\h' for help. Type '\w' to clear the current input statement.

mysql> \_

mysql> use scout

Database changed

mysql> show tables;

Tables\_in\_scout

country\_info

formation\_info

hidden\_league\_id

hidden\_player\_id

hidden\_team\_id

league\_info

nation\_info

personal\_stats

player\_info

rank\_info

round\_info

seasons

team\_info

winner

14 rows in set (0.01 sec)

mysql> \_

Anaconda Prompt (anaconda3) - mysql -h 34.64.214.96 -u root -p

mysql> select \* from team\_info;

| team_id | season_id | league_id | formation_id | team_name         | stadium_name       | team_est_date | owner | he |
|---------|-----------|-----------|--------------|-------------------|--------------------|---------------|-------|----|
| 1       | 8         | 1         | 5            | Manchester United | Old Trafford       | 1878          |       |    |
| 2       | 8         | 1         | 4            | Newcastle         | St. James' Park    | 1892          |       |    |
| 3       | 8         | 1         | 5            | Bournemouth       | Vitality Stadium   | 1899          |       |    |
| 4       | 8         | 1         | 5            | Fulham            | Craven Cottage     | 1879          |       |    |
| 5       | 8         | 1         | 5            | Wolves            | Molineux Stadium   | 1877          |       |    |
| 6       | 8         | 1         | 4            | Liverpool         | Anfield            | 1892          |       |    |
| 7       | 8         | 1         | 5            | Southampton       | St. Mary's Stadium | 1892          |       |    |
| 8       | 8         | 1         | 4            | Everton           | Goodison Park      | 1878          |       |    |
| 9       | 8         | 1         | 4            | Everton           | Goodison Park      | 1878          |       |    |

## 터미널 환경에서의 제어

mysql 커맨드를 통한 데이터베이스 접속, 탐색 및 관리

MySQL의 Query를 통한 데이터테이블 생성

터미널 인터페이스에 익숙해지고, 외부 프로그램에 대한 의존성 낮추기

# Process 3

MySQL 서버 환경 구축

On-Fremis

Workspace

“환경적 여건의 한계를 극복할 수 있을까?”



3.3.7



3.8.6



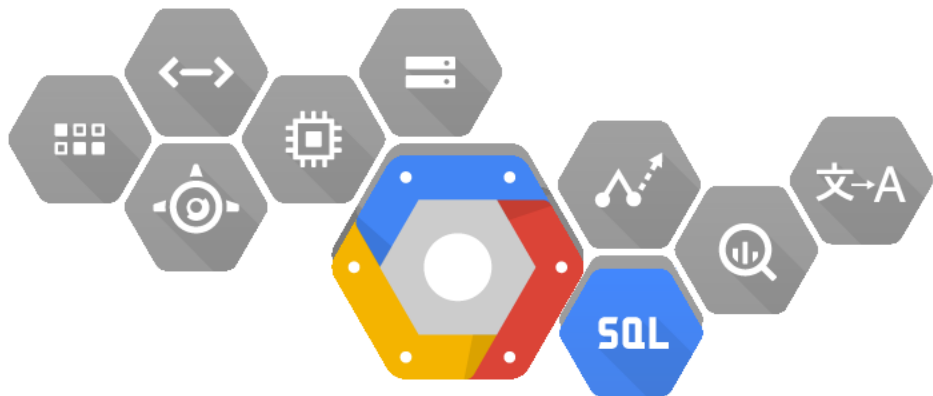
3.10.0



3.5.10

# Process 3

MySQL 서버 환경 구축



## MySQL 기반 클라우드 서비스

클라우드 서비스를 통한 데이터베이스 관리

**기업 표준으로 사용되는 GCP 환경 채택**

자체적인 서버 컴퓨터 관리 없이 클라우드를 통한 데이터 관리 목표

+ 추후 Function 기능을 통한 데이터의 실시간 업데이트 고려



SQL

기본 인스턴스

개요

쿼리 통계

연결

사용자

데이터베이스

백업

복제본

작업

데이터베이스

모든 인스턴스 > scouter

✓ scouter

MySQL 8.0

+ 데이터베이스 만들기

| 이름 ↑               | 대조              | 문자 집합 | 유형  |
|--------------------|-----------------|-------|-----|
| information_schema | utf8_general_ci | utf8  | 시스템 |

mysql

performance

scout

sys

test

1시간 6시간 1일 7일 30일 커스텀 ▼



→ 쿼리 통계로 이동하여 쿼리 및 성능에 대한 자세한 정보 보기

# Process 4

## 가공을 위한 데이터 분석

```
PS C:\Users\Playdata\Desktop> c:: cd 'c:\Users\Playdata\Desktop'; & 'C:\ProgramData\anaconda3\python.exe' 'c:\Users\Playdata\.vscode\extensions\ms-python.python-2023.6.0\pythonFiles\lib\p
ython\debugpy\adapter\..\..\debugpy\launcher' '53808' '--' 'c:\Users\Playdata\Desktop\scouter2.py'
{'get': 'players/squads', 'parameters': {'team': '42'}, 'errors': [], 'results': 1, 'paging': {'current': 1, 'total': 1}, 'response': [{'team': {'id': 42, 'name': 'Arsenal', 'logo': 'https
://media-2.api-sports.io/football/teams/42.png'}, 'players': [{'id': 20355, 'name': 'A. Ramsdale', 'age': 25, 'number': 1, 'position': 'Goalkeeper', 'photo': 'https://media-1.api-sports.io
/football/players/20355.png'}, {'id': 50999, 'name': 'M. Turner', 'age': 29, 'number': 30, 'position': 'Goalkeeper', 'photo': 'https://media-1.api-sports.io/football/players/50999.png'}, {
'id': 169295, 'name': 'K. Hein', 'age': 21, 'number': 31, 'position': 'Goalkeeper', 'photo': 'https://media-3.api-sports.io/football/players/169295.png'}, {'id': 138843, 'name': 'J. Hillso
n', 'age': 22, 'number': None, 'position': 'Goalkeeper', 'photo': 'https://media-1.api-sports.io/football/players/138843.png'}, {'id': 1117, 'name': 'K. Tierney', 'age': 26, 'number': 3, '
position': 'Defender', 'photo': 'https://media-1.api-sports.io/football/players/1117.png'}, {'id': 19959, 'name': 'B. White', 'age': 26, 'number': 4, 'position': 'Defender', 'photo': 'http
s://media-1.api-sports.io/football/players/19959.png'}, {'id': 22224, 'name': 'Gabriel Magalhães', 'age': 26, 'number': 6, 'position': 'Defender', 'photo': 'https://media-1.api-sports.io/f
ootball/players/22224.png'}, {'id': 22090, 'name': 'W. Saliba', 'age': 22, 'number': 12, 'position': 'Defender', 'photo': 'https://media-2.api-sports.io/football/players/22090.png'}, {'id'
: 61431, 'name': 'J. Kiwior', 'age': 23, 'number': 15, 'position': 'Defender', 'photo': 'https://media-1.api-sports.io/football/players/61431.png'}, {'id': 1440, 'name': 'R. Holding', 'age': 28, 'numbe
r': 16, 'position': 'Defender', 'photo': 'https://media-2.api-sports.io/football/players/1440.png'}, {'id': 2597, 'name': 'T. Tomiyasu', 'age': 25, 'number': 18, 'position': 'Defender', 'photo': 'https
://media-1.api-sports.io/football/players/2597.png'}, {'id': 284439, 'name': 'L. Sousa', 'age': 18, 'number': 96, 'position': 'Defender', 'photo': 'https://media-2.api-sports.io/football/players/284439
.png'}, {'id': 309505, 'name': 'R. Walters', 'age': 19, 'number': 97, 'position': 'Defender', 'photo': 'https://media-2.api-sports.io/football/players/309505.png'}, {'id': 49, 'name': 'T. Partey', 'age
': 30, 'number': 5, 'position': 'Midfielder', 'photo': 'https://media-2.api-sports.io/football/players/49.png'}, {'id': 1460, 'name': 'B. Saka', 'age': 22, 'number': 7, 'position': 'Midfielder', 'photo
': 'https://media-3.api-sports.io/football/players/1460.png'}, {'id': 37127, 'name': 'M. Ødegaard', 'age': 25, 'number': 8, 'position': 'Midfielder', 'photo': 'https://media-3.api-sports.io/football/pl
ayers/37127.png'}, {'id': 1161, 'name': 'E. Smith Rowe', 'age': 23, 'number': 10, 'position': 'Midfielder', 'photo': 'https://media-3.api-sports.io/football/players/1161.png'}, {'id': 2289, 'name': 'Jo
rginho', 'age': 32, 'number': 20, 'position': 'Midfielder', 'photo': 'https://media-3.api-sports.io/football/players/2289.png'}, {'id': 41725, 'name': 'Fábio Vieira', 'age': 23, 'number': 21, 'position
': 'Midfielder', 'photo': 'https://media-2.api-sports.io/football/players/41725.png'}, {'id': 1452, 'name': 'Mohamed Elneny', 'age': 31, 'number': 25, 'position': 'Midfielder', 'photo': 'https://media-
1.api-sports.io/football/players/1452.png'}, {'id': 1464, 'name': 'G. Xhaka', 'age': 31, 'number': 34, 'position': 'Midfielder', 'photo': 'https://media-2.api-sports.io/football/players/1464.png'}, {'i
d': 641, 'name': 'O. Zinchenko', 'age': 27, 'number': 35, 'position': 'Midfielder', 'photo': 'https://m
number': 72, 'position': 'Midfielder', 'photo': 'https://media-1.api-sports.io/football/players/167658.png'}, {'id': 309501, 'name': 'A. Cozier-Dub
photo': 'https://media-1.api-sports.io/football/players/313236.png'}, {'id': 309501, 'name': 'A. Cozier-Dub
football/players/309501.png'}, {'id': 278075, 'name': 'C. Ćirjan', 'age': 21, 'number': None, 'position
540, 'name': 'Mauro Bandeira', 'age': 19, 'number': None, 'position': 'Midfielder', 'photo': 'https://media-1.api-sports.io/football/players/284540.png'}, {'id': 309507, 'name': 'B. Ibrahim', 'age': 19
, 'number': None, 'position': 'Midfielder', 'photo': 'https://media-2.api-sports.io/football/pl
photo': 'https://media-3.api-sports.io/football/players/643.png'}, {'id': 127769, 'name': 'Gabr
football/players/127769.png'}, {'id': 1468, 'name': 'E. Nketiah', 'age': 24, 'number': 14, 'position': 'Attacker', 'photo': 'https://media-2.api-sports.io/football/players/1468.png'}, {'id': 727, 'name': 'B. Nelson', 'age': 24, 'number': 24, 'posit
ion': 'Attacker', 'photo': 'https://media-1.api-sports.io/football/players/727.png'}, {'id': 284062, 'name
pi-sports.io/football/players/284062.png'}, {'id': 284571, 'name': 'C. Sagoe', 'age': 19, 'number': None,
}}
PS C:\Users\Playdata\Desktop>
```

“이 많은 데이터들을 어떻게 가공하지?”

{ “Key” : [ { “key” : “value”, “key” : [ { “Key” : ... }, { .. } ], ... }, { ... } ], ... }


딕셔너리 안의 리스트 안의 딕셔너리 안의 리스트 안의 딕셔너리 형태

# Process 4

가공을 위한 데이터 분석




- 단일 request에 들어오는 데이터의 길이가 지나치게 길다
- 비슷한 유형의 데이터를 요청해도 response들의 구조가 각기 다르다





**Son Heung-Min**  
**Attacker** : Tottenham Hotspur FC / Korea Republic  
**DOB** : 1992년 7월 08일 (Chuncheon, Korea Republic)

|         |         |       |
|---------|---------|-------|
| Height: | Weight: | Foot: |
| 184cm   | 77kg    | both  |



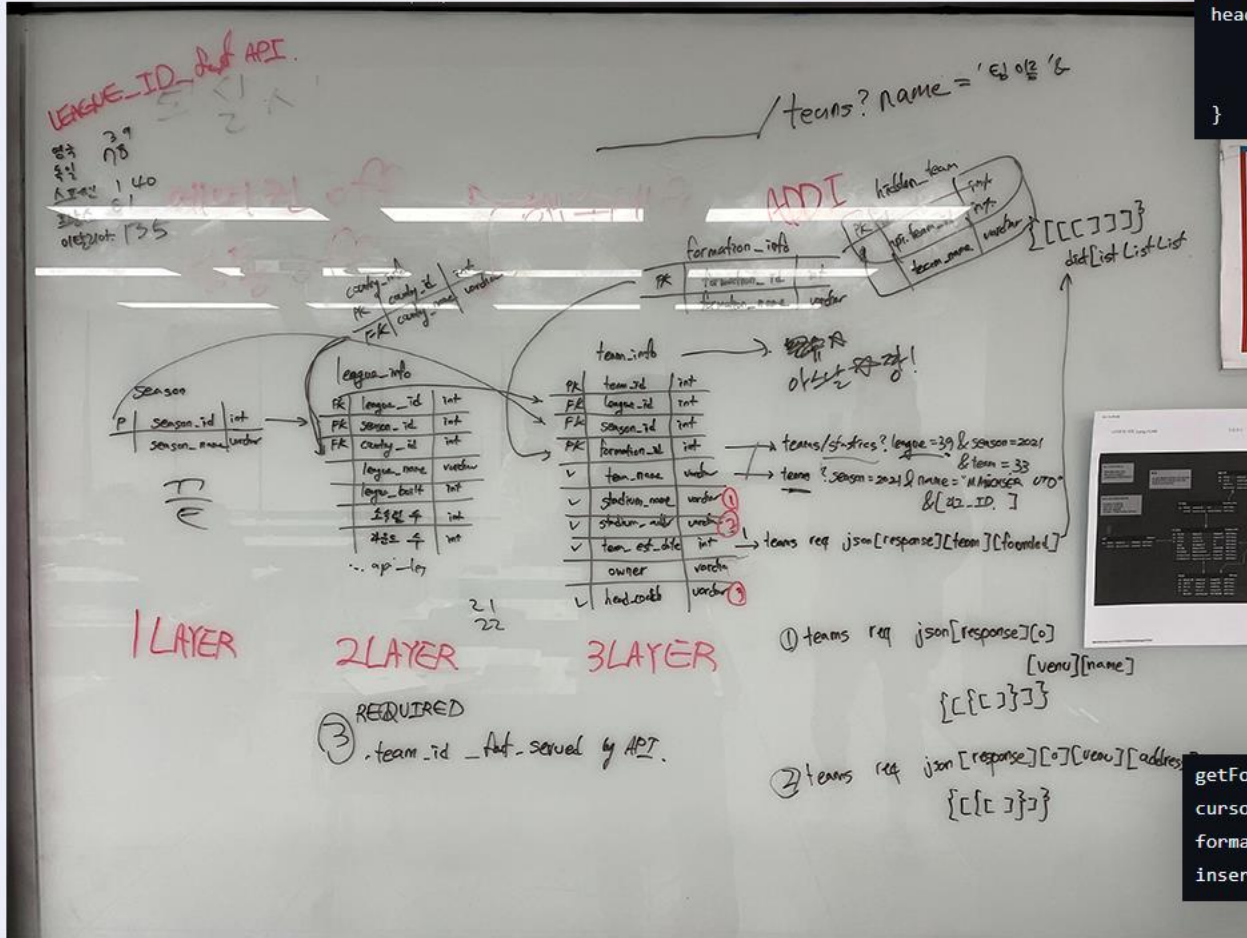


“이 많은 데이터들을 어떻게 가공하지?”



# Process 4

가공을 위한 데이터 분석



#PL 스크립트에서 필요한 상수값 + 확장성은 hidden에서 끌어올수 있음 좋음

```
tmpLeagueName = "Premier League"
```

```
tmpSeasonYear = 2022
```

```
tmpLeagueID = 39
```

```
getTeamUrl = "https://v3.football.api-sports.io/teams?season=%d&league=%d" %(tmpSeasonYear, tmpLeagueID)
```

```
headers={
```

```
    'x-rapidapi-host': "v3.football.api-sports.io",
```

```
    'x-rapidapi-key': "a68636f8f2c18511179c56f15e95080c"
```

```
}
```

request용 URI 구조 파악

```
#json으로 으로 pl 데이터 다 받아와서 indexing 후 적재 list에 이중 리스트 구현
for i in range(len(tmpTeamData)):
```

```
    #insert 리스트 내부에 인덱싱될 data labeling
```

```
    tmpList = []
```

```
    tmpList.append(tmpTeamData[i]['team']['name'])
```

```
    tmpList.append(tmpTeamData[i]['team']['id'])
```

```
    tmpList.append(tmpTeamData[i]['team']['founded'])
```

```
    tmpList.append(tmpTeamData[i]['venue']['name'])
```

```
    tmpList.append(tmpTeamData[i]['venue']['address'])
```

```
    insertTeamInfoList.append(tmpList)
```

API 데이터 구조 파악

```
getFormationSQL = 'select formation_id from formation_info where formation_name = "%s"' %convertFormation
```

```
cursor.execute(getFormationSQL)
```

```
formationRawData = cursor.fetchall()
```

```
insertFormationId = formationRawData[0][0]
```

데이터베이스에 적재된 기존 데이터들과의 관계 파악

# Process 5

## MySQL 데이터 적재

“오류들을 제어할 수 있는 방법이 없을까?”



MySQL에서 Primary Key 값은 레코드를 추가할 때 자동으로 할당됩니다. 이를 'auto-increment'라고도 부릅니다. 이 값은 테이블이 생성될 때 정의되며, 레코드가 삽입될 때마다 1씩 증가하며 자동으로 할당됩니다. 따라서, Python에서 MySQL 서버로 쿼리문을 전송할 때, 오류가 발생하여 스크립트 실행이 중단되더라도, Primary Key 값은 이미 할당된 상태이므로 해당 레코드는 테이블에 저장됩니다.

그러나, 중간에 오류가 발생하여 쿼리문이 실행되지 않았다면, 해당 레코드는 부적절한 값으로 저장될 수 있습니다. 이를 방지하기 위해, 쿼리문 실행 전에 Primary Key 값을 지정하는 쿼리문을 작성해야 합니다.



MySQL 서버에서 데이터베이스가 30초마다 실시간 업데이트 되고 있고, 파이썬에서 1분마다 쿼리문을 실행하여 가장 마지막 데이터를 가져오도록 한다면, 파이썬에는 최근 1분 동안 업데이트된 데이터만 가져올 수 있습니다.

반복문 내에서 쿼리문을 실행할 때마다, MySQL 서버와의 연결이 새롭게 이루어지지 않습니다. 이미 연결된 상태에서 쿼리를 실행하며, 이전 실행에서 반환된 결과와는 다른 정보를 가져올 수 있습니다. 그러나, 이전 결과를 제외한 최근 1분 동안 업데이트된 데이터를 가져오기 위해서는 해당 시간 동안에 업데이트된 데이터를 기준으로 쿼리문을 작성해야 합니다.

즉, 1분마다 가장 마지막 데이터를 가져오기 위해서는, 해당 데이터에 대한 타임스탬프 컬럼이 있어야 하며, 이를 활용하여 1분 이내에 업데이트된 데이터를 검색하는 쿼리문을 작성해야 합니다. 이렇게 작성한 쿼리문을 반복문 내에서 실행하면, 최근 1분 동안 업데이트된 데이터를 가져올 수 있습니다.



# Process 5

## MySQL 데이터 적재

```
for l in range(len(SeasonYearList)):
    if k % 2 == 1:
        print('wait for 120s')
        time.sleep(120)
        print('have waited 120s. 다시 일해라!')
    else :
        pass

url = "https://v3.football.api-sports.io/standings?season=%d&league=%d" %(SeasonYearList[l], ApiLeagueIdList[k])
resp = re.get(url=url, headers=headers).json()
tmpData = resp['response'][0]['league']['standings'][0]
# api에서 받은 값을 담은 빈 리스트 생성 -> 이중 리스트가 될 것임 -> ex) [['Manchester United', 1],['Arsenal',2],,,,]
tmpRankData = []
# 최종 insert할 데이터를 담은 빈 리스트 생성 -> 이중 리스트가 될 것임 -> ex) [[1,1],[2,2],,,,] <- [[팀 ID,순위],[팀 ID,순위],,,,]
insertRankData = []

# api에서 필요한 값 tmpRankData에 넣기
for m in range(len(tmpData)):
    # tmpRankData가 이중리스트가 되어야 하기 때문에 한 사이클 마다 tmpRankData에 append할 빈 리스트 생성
    tmpRankData = []

    tmpRankData.append(tmpData[m]['team']['name'])
    tmpRankData.append(tmpData[m]['rank'])

# tmpRankData에 append
insertRankData.append(tmpRankData)
```

```
# db에서 시즌년도 값 쪽 받아오기
getSeasonYearSql = 'select season_year from seasons'
cursor.execute(getSeasonYearSql)
tmpSeasonYear = cursor.fetchall()
# getSeasonYear 쿼리문으로 받아온 값 인덱싱해서 SeasonYearList에 append
for i in range(len(tmpSeasonYear)-1):
    SeasonYearList.append(tmpSeasonYear[i][0])
print(SeasonYearList)
```

```
# 각 리그 api_league_id를 받을 빈 리스트 생성
ApiLeagueIdList = []
# db에서 api_league_id 값 받아오기
getApiLeagueId = 'select api_league_id from hidden_league_id'
cursor.execute(getApiLeagueId)
tmpApiLeagueId = cursor.fetchall()
# getApiLeagueId 쿼리문으로 받아온 값 인덱싱해서 ApiLeagueIdList에 append
for j in range(len(tmpApiLeagueId)):
    ApiLeagueIdList.append(tmpApiLeagueId[j][0])
print(ApiLeagueIdList)
```

## 오류제어 중심의 .py 파일 빌드

annotations, modulations, reuses and studies

## 프로덕션 환경을 염두한 .py 파일 손코딩

각 기능별로 코드들을 정렬하여 오류를 빠르게 점검하고 추후 객체화 할 수 있도록 구성 / 오류 발견 또는 예상 구간을 주석으로 상세 표기

# What We Did

```
mysql> select * from seasons;
```

| season_id | season_year |
|-----------|-------------|
| 1         | 2015        |
| 2         | 2016        |
| 3         | 2017        |
| 4         | 2018        |
| 5         | 2019        |
| 6         | 2020        |
| 7         | 2021        |
| 8         | 2022        |

8 rows in set (0.01 sec)

## | season

- season id
- season year

```
mysql> select * from league_info;
```

| league_id | season_id | nation_id | league_name    | league_created | league_team_count | league_round_count |
|-----------|-----------|-----------|----------------|----------------|-------------------|--------------------|
| 1         | 0         | 1         | Premier League | 1888           | 20                | 38                 |
| 2         | 1         | 1         | Premier League | 1888           | 20                | 38                 |
| 3         | 2         | 1         | Premier League | 1888           | 20                | 38                 |
| 4         | 3         | 1         | Premier League | 1888           | 20                | 38                 |
| 5         | 4         | 1         | Premier League | 1888           | 20                | 38                 |
| 6         | 5         | 1         | Premier League | 1888           | 20                | 38                 |
| 7         | 6         | 1         | Premier League | 1888           | 20                | 38                 |
| 8         | 7         | 1         | Premier League | 1888           | 20                | 38                 |
| 9         | 1         | 3         | La Liga        | 1929           | 20                | 38                 |
| 10        | 2         | 3         | La Liga        | 1929           | 20                | 38                 |
| 11        | 3         | 3         | La Liga        | 1929           | 20                | 38                 |
| 12        | 4         | 3         | La Liga        | 1929           | 20                | 38                 |
| 13        | 5         | 3         | La Liga        | 1929           | 20                | 38                 |
| 14        | 6         | 3         | La Liga        | 1929           | 20                | 38                 |
| 15        | 7         | 3         | La Liga        | 1929           | 20                | 38                 |
| 16        | 8         | 3         | La Liga        | 1929           | 20                | 38                 |
| 17        | 1         | 2         | Bundesliga     | 1897           | 18                | 34                 |
| 18        | 2         | 2         | Bundesliga     | 1897           | 18                | 34                 |
| 19        | 3         | 2         | Bundesliga     | 1897           | 18                | 34                 |
| 20        | 4         | 2         | Bundesliga     | 1897           | 18                | 34                 |
| 21        | 5         | 2         | Bundesliga     | 1897           | 18                | 34                 |
| 22        | 6         | 2         | Bundesliga     | 1897           | 18                | 34                 |
| 23        | 7         | 2         | Bundesliga     | 1897           | 18                | 34                 |
| 24        | 8         | 2         | Bundesliga     | 1897           | 18                | 34                 |
| 25        | 1         | 4         | Ligue 1        | 1932           | 20                | 38                 |
| 26        | 2         | 4         | Ligue 1        | 1932           | 20                | 38                 |
| 27        | 3         | 4         | Ligue 1        | 1932           | 20                | 38                 |
| 28        | 4         | 4         | Ligue 1        | 1932           | 20                | 38                 |
| 29        | 5         | 4         | Ligue 1        | 1932           | 20                | 38                 |
| 30        | 6         | 4         | Ligue 1        | 1932           | 20                | 38                 |
| 31        | 7         | 4         | Ligue 1        | 1932           | 20                | 38                 |
| 32        | 8         | 4         | Ligue 1        | 1932           | 20                | 38                 |
| 33        | 1         | 5         | Serie A        | 1948           | 20                | 38                 |
| 34        | 2         | 5         | Serie A        | 1948           | 20                | 38                 |
| 35        | 3         | 5         | Serie A        | 1948           | 20                | 38                 |
| 36        | 4         | 5         | Serie A        | 1948           | 20                | 38                 |
| 37        | 5         | 5         | Serie A        | 1948           | 20                | 38                 |
| 38        | 6         | 5         | Serie A        | 1948           | 20                | 38                 |
| 39        | 7         | 5         | Serie A        | 1948           | 20                | 38                 |
| 40        | 8         | 5         | Serie A        | 1948           | 20                | 38                 |

40 rows in set (0.01 sec)

## | league

- league id
- season id(FK)
- nation id(FK)
- league name
- league created
- league team count
- league round count

```
mysql> select * from country_info;
```

| country_id | country_name |
|------------|--------------|
| 1          | England      |
| 2          | Germany      |
| 3          | Spain        |
| 4          | France       |
| 5          | Italy        |

5 rows in set (0.05 sec)

## | country

- country id
- country name

```
mysql> select * from nation_info;
```

| nation_id | country_id | nation_state  | nation_city     |
|-----------|------------|---------------|-----------------|
| 1         | 1          | North         | Northumberland  |
| 2         | 1          | North         | Newcastle area  |
| 3         | 1          | North         | Durham          |
| 4         | 1          | North         | Cumbria         |
| 5         | 1          | North         | Teesside        |
| 6         | 1          | Yorkshire     | North Yorkshire |
| 7         | 1          | Yorkshire     | West Yorkshire  |
| 8         | 1          | Lancashire    | Lancashire      |
| 9         | 1          | Lancashire    | Merseyside      |
| 10        | 1          | Manchester    | Manchester      |
| 11        | 1          | Humber        | Humber area     |
| 12        | 1          | Humber        | Lincolnshire    |
| 13        | 1          | Humber        | South Yorkshire |
| 14        | 1          | East Midlands | Derbyshire      |
| 15        | 1          | East Midlands | Notts           |
| 16        | 1          | East Midlands | Leicestershire  |
| 17        | 1          | East Midlands | Warwicks        |
| 18        | 1          | Servern       | Cheshire        |
| 19        | 1          | Servern       | Staffs          |
| 20        | 1          | Servern       | Shropshire      |
| 21        | 1          | Servern       | H&W             |
| 22        | 1          | Wales         | Gwynedd         |
| 23        | 1          | Wales         | Ceredigion      |
| 24        | 1          | Wales         | Powys           |
| 25        | 1          | Wales         | Dyfed           |
| 26        | 1          | Wales         | N. Glam         |
| 27        | 1          | Wales         | Mid Glam        |
| 28        | 1          | Wales         | S. Glam         |
| 29        | 1          | Wales         | Gwent           |
| 30        | 1          | East Anglia   | Norfolk         |

30 rows in set (0.09 sec)

```
mysql> select * from team_info;
```

| team_id | season_id | league_id | formation_id | team_name          | stadium_name                           | stadium_address                       | team_est_date | owner | head_coach |
|---------|-----------|-----------|--------------|--------------------|--|---------------------------------------|---------------|-------|------------|
| 1       | 0         | 1         | 5            | Manchester United  | Old Trafford                           | Sir Matt Busby Way                    | 1878          |       |            |
| 2       | 0         | 1         | 4            | Newcastle          | St. James' Park                        | St. James' Park; Street               | 1892          |       |            |
| 3       | 0         | 1         | 5            | Bournemouth        | Vitality Stadium                       | Dean Court, Kings Park                | 1899          |       |            |
| 4       | 0         | 1         | 5            | Fulham             | Craven Cottage                         | Stevenage Road                        | 1879          |       |            |
| 5       | 0         | 1         | 5            | Wolves             | Molineux Stadium                       | Waterloo Road                         | 1877          |       |            |
| 6       | 0         | 1         | 4            | Liverpool          | Anfield                                | Anfield Road                          | 1892          |       |            |
| 7       | 0         | 1         | 5            | Southampton        | St. Mary's Stadium                     | Britannia Road                        | 1885          |       |            |
| 8       | 0         | 1         | 4            | Arsenal            | Emirates Stadium                       | Queensland Road                       | 1886          |       |            |
| 9       | 0         | 1         | 4            | Everton            | Goodison Park                          | Goodison Road                         | 1878          |       |            |
| 10      | 0         | 1         | 7            | Leicester          | King Power Stadium                     | Filbert Way                           | 1884          |       |            |
| 11      | 0         | 1         | 8            | Tottenham          | Tottenham Hotspur Stadium              | Bill Nicholson Way, 748 High Road     | 1882          |       |            |
| 12      | 0         | 1         | 5            | West Ham           | London Stadium                         | Marshgate Lane, Stratford             | 1895          |       |            |
| 13      | 0         | 1         | 5            | Chelsea            | Stamford Bridge                        | Fulham Road                           | 1905          |       |            |
| 14      | 0         | 1         | 4            | Manchester City    | Etihad Stadium                         | Rousley Street                        | 1880          |       |            |
| 15      | 0         | 1         | 5            | Brighton           | The American Express Community Stadium | Village Way                           | 1901          |       |            |
| 16      | 0         | 1         | 5            | Crystal Palace     | Selhurst Park                          | Holmesdale Road                       | 1905          |       |            |
| 17      | 0         | 1         | 4            | Brentford          | Gtech Community Stadium                | 166 Lionel Rd N, Brentford            | 1889          |       |            |
| 18      | 0         | 1         | 5            | Leeds              | Elland Road                            | Elland Road                           | 1919          |       |            |
| 19      | 0         | 1         | 4            | Notttingham Forest | The City Ground                        | Pavilion Road                         | 1865          |       |            |
| 20      | 0         | 1         | 3            | Aston Villa        | Villa Park                             | Trinity Road                          | 1874          |       |            |
| 21      | 0         | 9         | 4            | Barcelona          | Spotify Camp Nou                       | Carrer d'Aragó; Arístides Maillol     | 1899          |       |            |
| 22      | 0         | 9         | 3            | Atletico Madrid    | Estadio Cívitas Metropolitano          | Risas                                 | 1903          |       |            |
| 23      | 0         | 9         | 3            | Athletic Club      | San Mam's Barria                       | Rafael Moreno Pitxitri Kalea          | 1898          |       |            |
| 24      | 0         | 9         | 4            | Valencia           | Estadio de Mestalla                    | Avenida de Suecia                     | 1919          |       |            |
| 25      | 0         | 9         | 3            | Villarreal         | Estadio de la Cerámica                 | Plaza Labrador                        | 1923          |       |            |
| 26      | 0         | 9         | 5            | Sevilla            | Estadio Ramón Sánchez Pizjuán          | Avenida de Eduardo Dato               | 1890          |       |            |
| 27      | 0         | 9         | 9            | Abanca-Balaídos    | Abanca-Balaídos                        | Avenida de Balaídos                   | 1923          |       |            |
| 28      | 0         | 9         | 5            | Espanyol           | RCDE Stadium                           | Avenida Baix Llobregat 100            | 1900          |       |            |
| 29      | 0         | 9         | 4            | Real Madrid        | Estadio Santiago Bernabéu              | Avenida de Concha Espina 1, Chamartín | 1902          |       |            |
| 30      | 0         | 9         | 5            | Real Betis         | Estadio Benito Villamarín              | Avenida de Heliópolis                 | 1907          |       |            |
| 31      | 0         | 9         | 6            | Getafe             | Coliseum Alfonso Pérez                 | Avenida de Teresa de Calcuta          | 1983          |       |            |

406 rows in set (0.03 sec)

## | team

- team id
- team name
- season id(FK)
- stadium name
- league id(FK)
- stadium address
- formation id(FK)
- team estimate date

```
mysql> select * from formation_info;
```

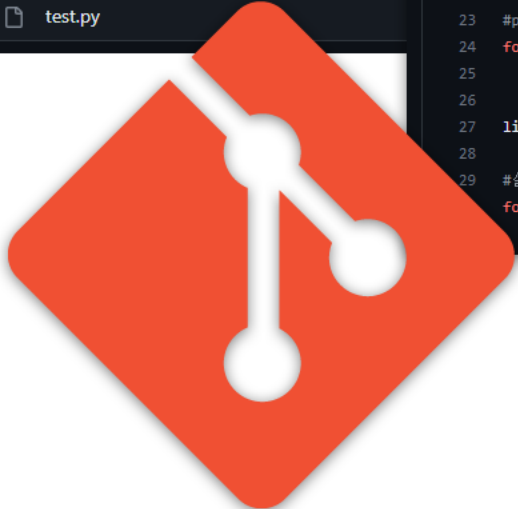
| formation_id | formation_name |
|--------------|----------------|
| 1            | 343            |
| 2            | 352            |
| 3            | 442            |
| 4            | 433            |
| 5            | 4231           |
| 6            | 532            |
| 7            | 4141           |
| 8            | 3421           |
| 9            | 432            |
| 10           | 541            |
| 11           | 3511           |
| 12           | 3412           |
| 13           | 4312           |
| 14           | 453            |
| 15           | 4222           |
| 16           | 3142           |
| 17           | 4411           |
| 18           | 4221           |

18 rows in set (0.07 sec)

## | formation

- formation id
- formation name

# What We Did



The image displays a development environment with three panels showing code and file structure.

**Left Panel (File Explorer):** Shows a project structure with folders like `.idea`, `Hidden_info/LOAD`, `League_info`, `Player_info`, `Rank_info`, `Team_info`, and files like `.DS_Store`, `README.md`, `example.json`, `ppt_diagram.pptx`, and `test.py`.

**Middle Panel (Code Editor):** Shows Python code for connecting to a MySQL database and fetching team information. The code includes imports for `requests` and `mysql.connector`, database connection details, and SQL queries to fetch team IDs and player information.

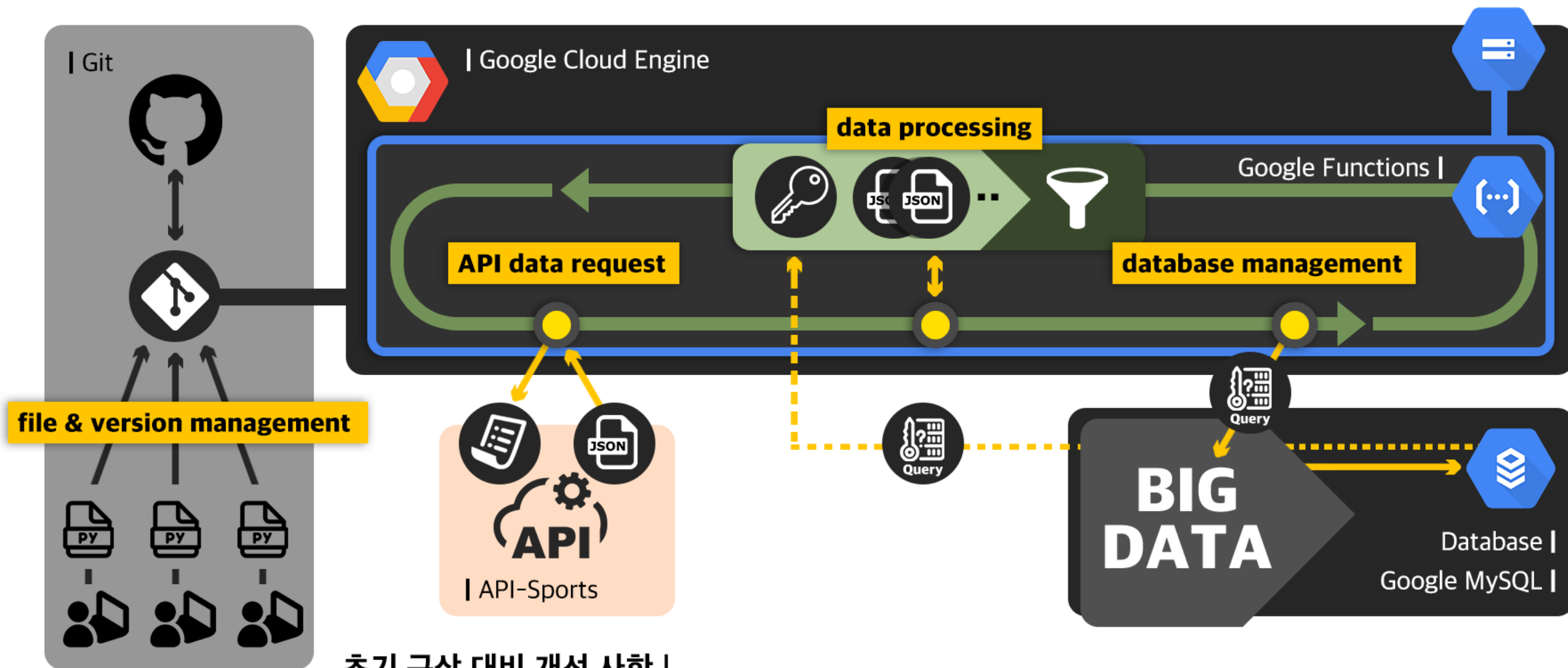
```
1 import requests as re
2 import mysql.connector
3
4
5 tmpTeamId = []
6 tmpLeagueId = 39
7 tmpSeasonYear = 2022
8
9 headers = {
10     'x-rapidapi-host': "v3.football.api-sports.io",
11     'x-rapidapi-key': "a86d420d0d8840c8e722e16cf9742f7b"
12 }
13
14 conn = mysql.connector.connect(user='root', password='tmzkdnxj1', host='34.64.214.96', database='scout')
15 cursor = conn.cursor()
16
17 getTeamApiIdSql = 'SELECT DISTINCT hidden_team_id.api_team_id FROM team_info'
18 cursor.execute(getTeamApiIdSql)
19 tmpTeamRaw = cursor.fetchall()
20
21 #####
22 #####
23 #p1 팀 id 리스트화
24 for i in range(len(tmpTeamRaw)):
25     tmpTeamId.append(tmpTeamRaw[i][0])
26
27 lili = tmpTeamId[:3]
28
29 #실 연산 시작
30 for i in range(len(lili)):
31     getBaseInfoUri = "https://v3.football.api-sports.io/players/squads?team=%s"
```

**Right Panel (Code Editor):** Shows Python code for making HTTP requests to the football API. It includes headers for the API key and host, and a GET request to fetch league information.

```
call_league.py → League_info/LOAD/Bundes_getTeamInfo.py
@@ -4,10 +4,10 @@
4     conn = mysql.connector.connect(user='root', password='tmzkdnxj1', host='34.64.214.96', database='scout',
5     port='3306')
6     cursor = conn.cursor()
7 - url = "https://v3.football.api-sports.io/leagues?name=Premier%20League&country=England&season=2022"
8 + url = "https://v3.football.api-sports.io/leagues?name=Serie%20A&country=Italy&season=2022"
9     headers = {
10     'x-rapidapi-host': "v3.football.api-sports.io",
11     'x-rapidapi-key': "a86d420d0d8840c8e722e16cf9742f7b"
12 + 'x-rapidapi-key': "a68636f8f2c18511179c56f15e95080c"
13     }
14     data = requests.request("GET", url, headers=headers).json()
15
League_info/LOAD/LA_getTeamInfo.py
... @@ -0,0 +1,56 @@
1 + import requests
2 + import mysql.connector
3 +
4 + conn = mysql.connector.connect(user='root', password='tmzkdnxj1', host='34.64.214.96', database='scout',
5 + port='3306')
6 + cursor = conn.cursor()
7 +
8 + url = "https://v3.football.api-sports.io/leagues?name=Serie%20A&country=Italy&season=2022"
9 + headers = {
10 +     'x-rapidapi-host': "v3.football.api-sports.io",
11 +     'x-rapidapi-key': "a68636f8f2c18511179c56f15e95080c"
12 + }
13 + data = requests.request("GET", url, headers=headers).json()
14 +
15 + # response 안에 실제 데이터가 있어서 그거만 빼주기
16 + reveil = data['response']
17
```



# Fixed Database Pipeline Diagram



## 초기 구상 대비 개선 사항 |

# Git / Github를 통한 파일 및 버전 관리

# Google Cloud Platform을 통한 데이터베이스 서버 관리 / 데이터 갱신 자동화



# About Presentation

