

Data Warehousing Final Project

Team members: Bikash Ghimirey, Rahul Awale, Ramesh Kumar, Abdul Rehman Aban

Phase 1

Data Source Selection & Extraction

API Overview

- API Name: TheSportsDB
- Base URL: <https://www.thesportsdb.com/api/v1/json/3/>
- Primary Use: Retrieve sports data, including teams, matches, players, and stadium details for various leagues and sports.
- Authentication: Requires an API key.
 - Public Key: 3 (used in this project)

Rationale for Choosing This API

We chose TheSportsDB API because:

- It offers structured, JSON-based data.
- Data is relevant for sports analytics, including match results, teams, venues, and media.
- Covers multiple sports, leagues, and languages.
- It's publicly accessible and doesn't require OAuth

Endpoints Used in the Project

1. Teams by League

URL:

https://www.thesportsdb.com/api/v1/json/3/search_all_teams.php?l=English%20Premier%20League

Purpose: To get all teams in the English Premier League.

Fields Used: idTeam, strTeam, strStadium, intStadiumCapacity, strLocation, idVenue, etc.

2. Stadium Information

URL: <https://www.thesportsdb.com/api/v1/json/3/lookupvenue.php?id=<idVenue>>

Purpose: To get stadium details like construction cost

Fields Used: strCost

3. Last 5 Matches per Team

URL: <https://www.thesportsdb.com/api/v1/json/3/eventslast.php?id=<idTeam>>

Purpose: To collect the last 5 completed matches for each team.

Fields Used:

strEvent, strHomeTeam, strAwayTeam, intHomeScore, intAwayScore, dateEvent, strTime, strVenue, strStatus

Sample JSON Structure

Match Object (from /eventslast.php)

```
json
{
  "strEvent": "Arsenal vs Chelsea",
  "strLeague": "English Premier League",
  "strHomeTeam": "Arsenal",
  "strAwayTeam": "Chelsea",
  "intHomeScore": "1",
  "intAwayScore": "0",
  "dateEvent": "2025-03-16",
  "strTime": "13:30:00",
  "strVenue": "Emirates Stadium",
  "strStatus": "Match Finished"
}
```