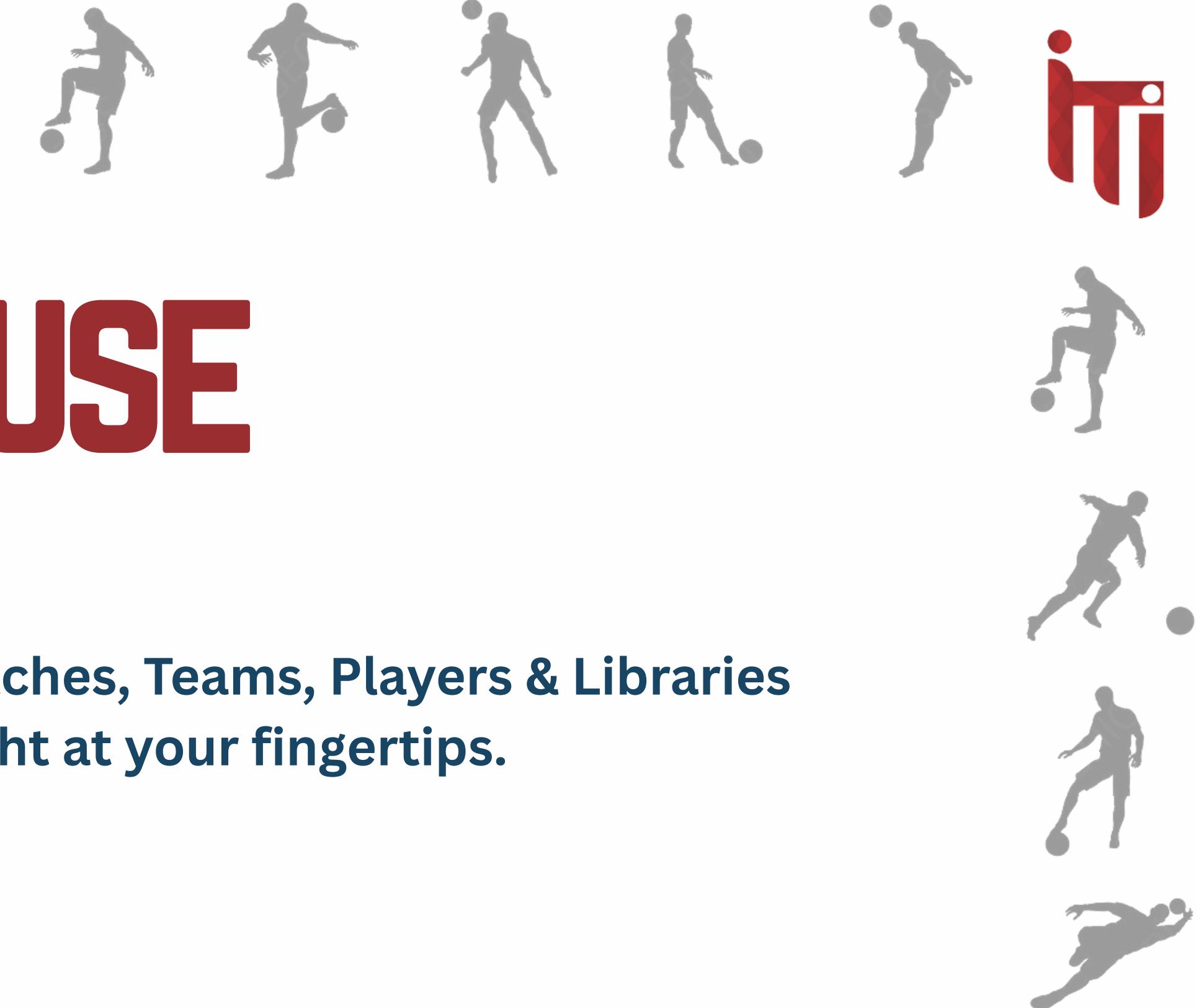




KICK HOUSE

Historical Matches, Teams, Players & Libraries
right at your fingertips.





TEAM MEMBERS





OVERVIEW

Kick House is a real-time football analytics platform designed to deliver instant live scores, detailed match statistics, and rich insights from leagues worldwide.

It provides fans, analysts, and sports enthusiasts with up-to-the-second data, helping them stay informed and engaged, while offering a smooth and interactive user experience





PROJECT AIM

Football isn't just a game it's passion, rivalry, and unforgettable moments, **Kick House** transforms the way fans live the sport by delivering real-time scores, instant match insights.

Whether it's the roar of a last-minute goal or the tension of a penalty shootout, kick house keeps you in the heart of the action, anytime, anywhere.





AUDIENCE

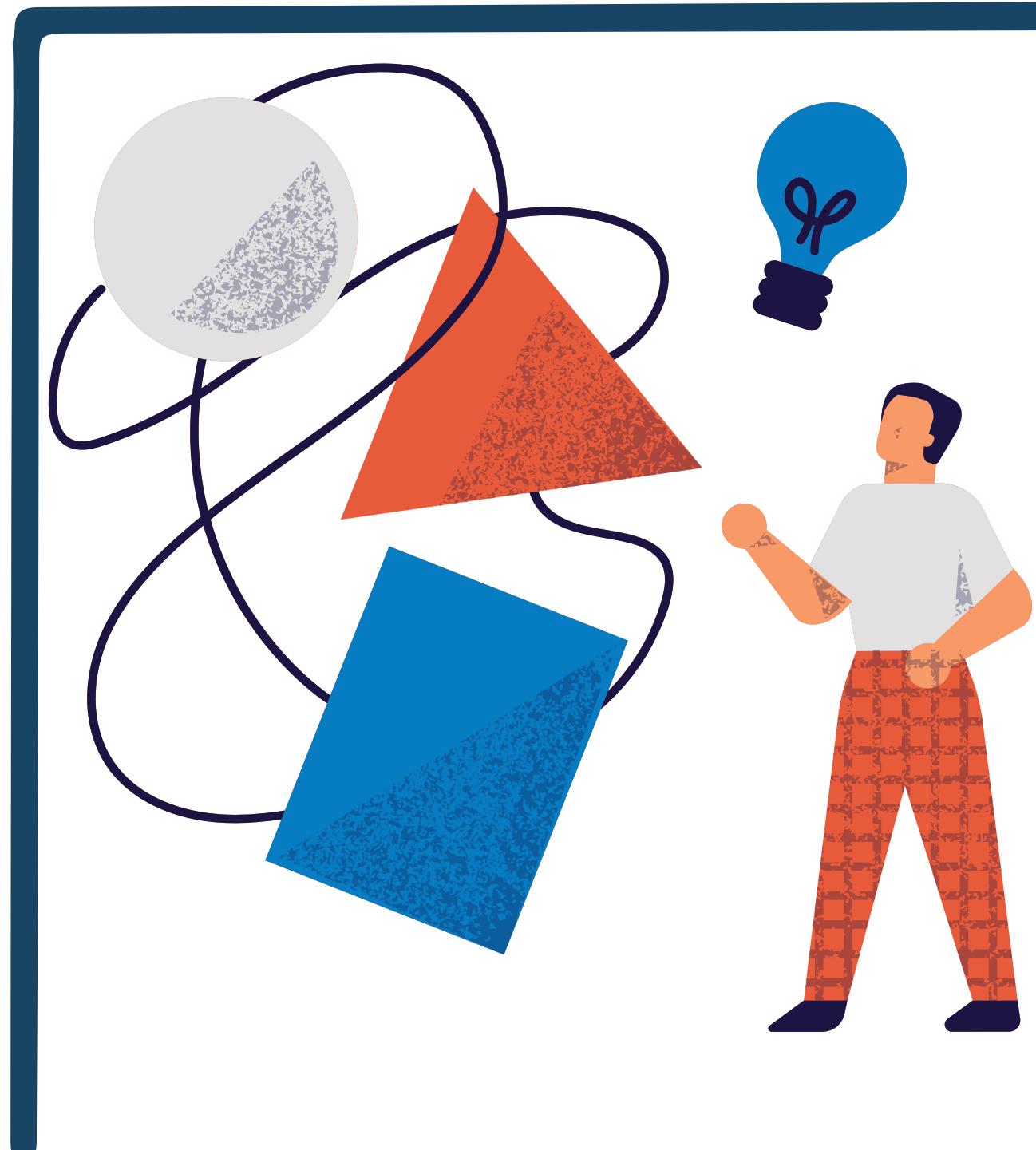
“Hard football fans from all over the world”





PROBLEMS

- Massive variety of big data (live scores, player statistics, historical records, media content).
- Much of the data is unstructured and hard to process.
- Difficult to integrate and analyze in real time.
- We accepted the problem that the free API did not work with us.



SOLUTION



AGENDA

- 1. Overview**
- 2. Project Aim**
- 3. Audience**
- 4. Problems**
- 5. Solution**
- 6. Work Flow**
- 7. Valid data flow**
- 8. Invalid data flow**
- 9. Data modeling**
- 10. Orchestration**
- 11. Branding**
- 12. Dashboards**
- 13. Conclusion**



1



Build a real-time data pipeline to collect and process data from multiple sources.

2



Use stream processing to handle live sports updates instantly.

3



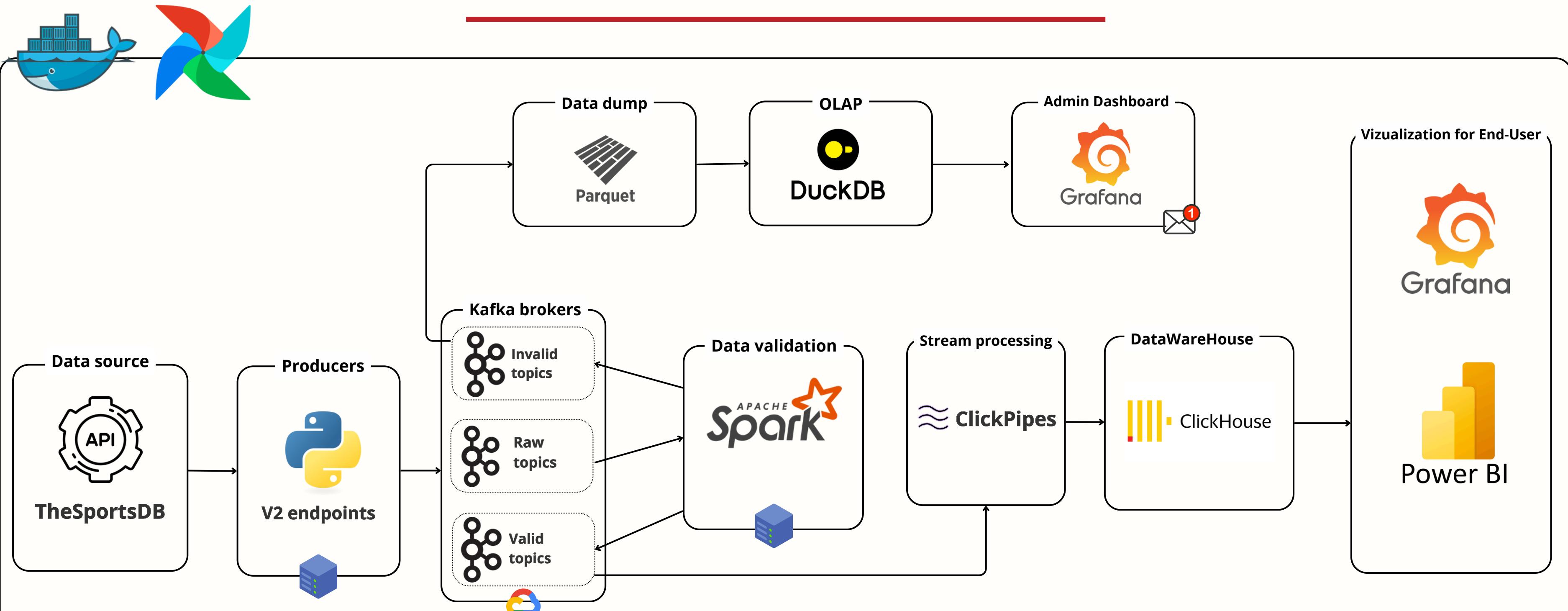
Store data in a centralized, scalable database for easy access.

4



Implement data cleaning and transformation to standardize formats.

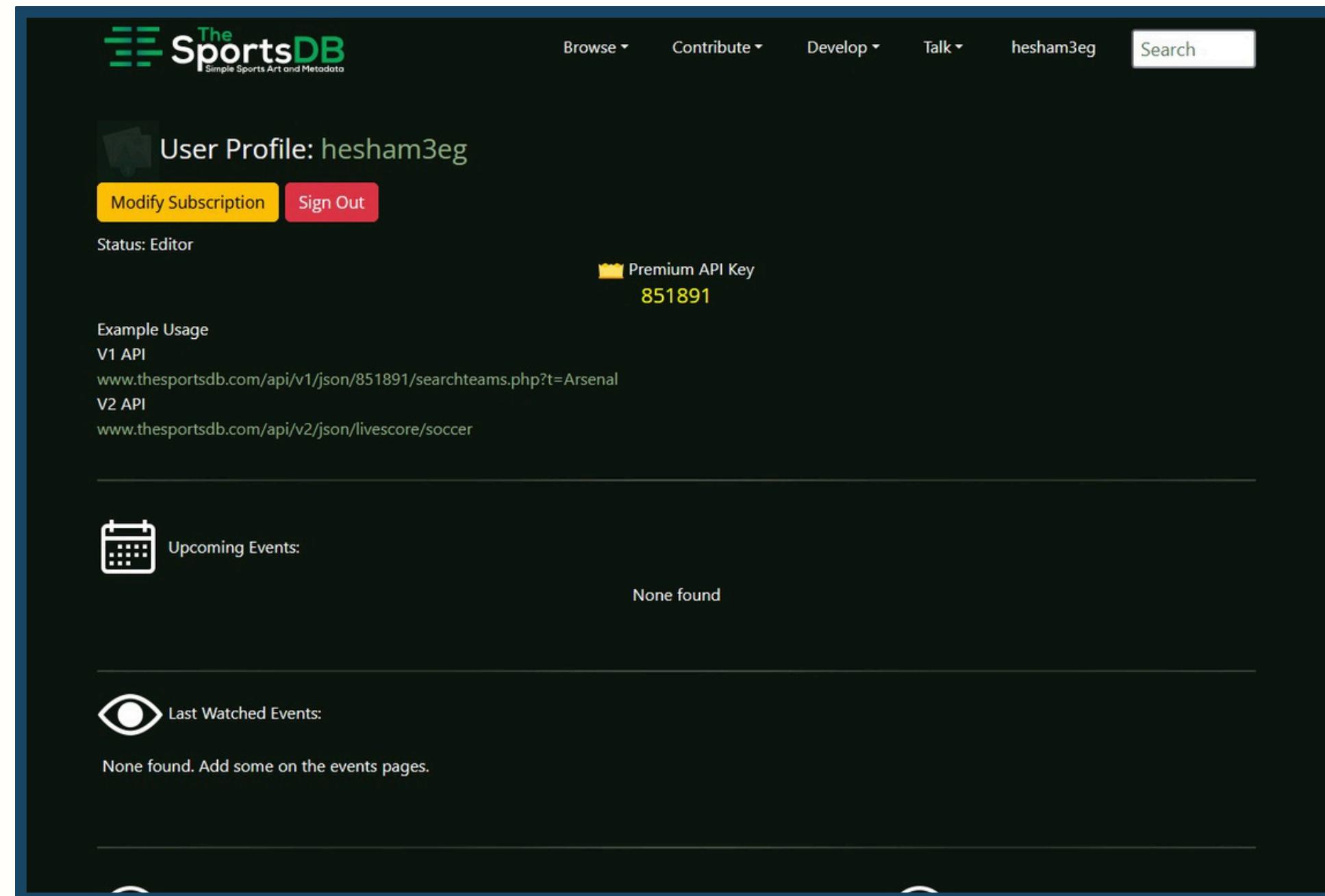
WORK FLOW





DATA SOURCE

This API allows you to search for sports data, analytics, schedules, social media links and fan artwork.



TheSportsDB User Profile: hesham3eg

Status: Editor

Premium API Key
851891

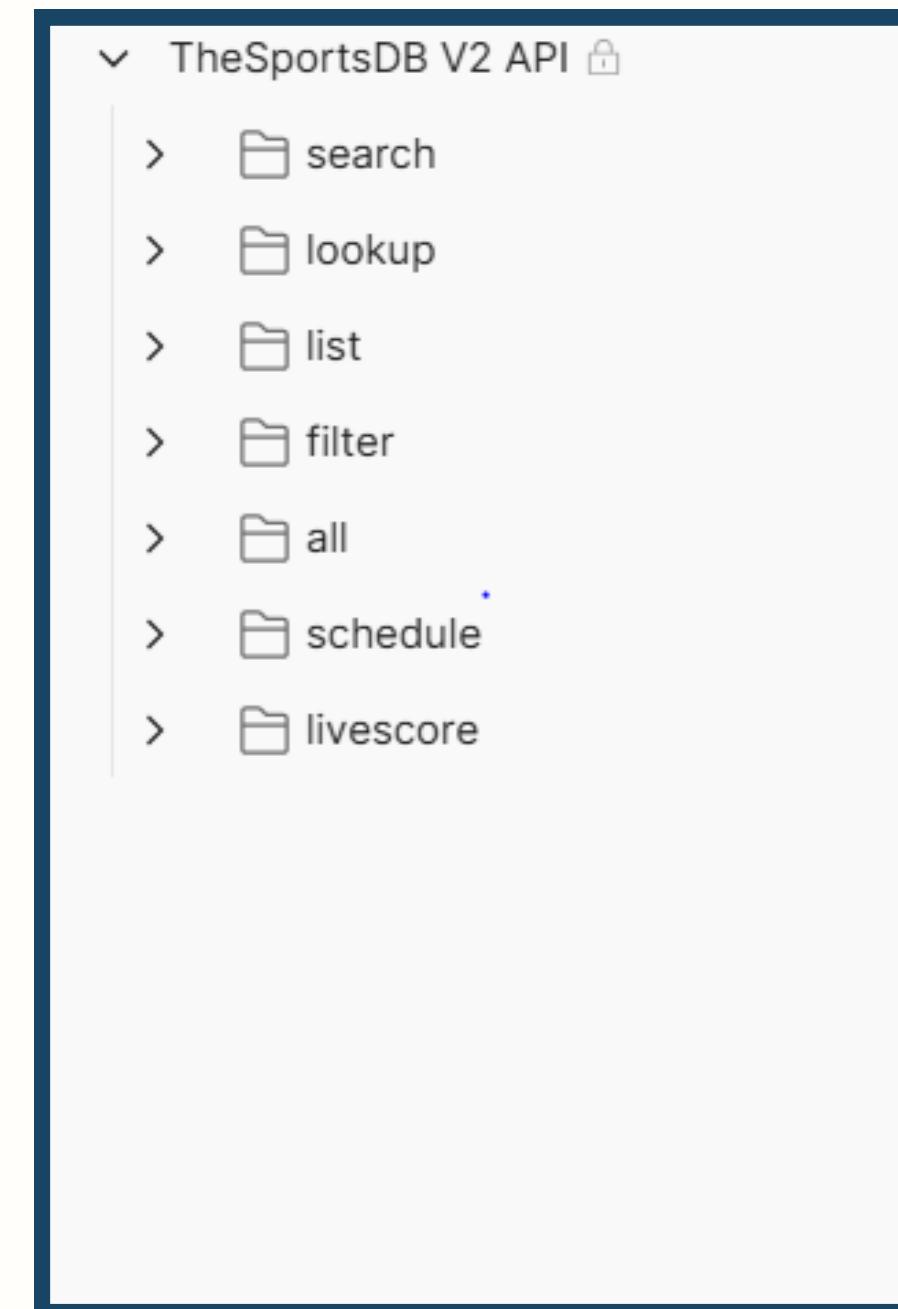
Example Usage

V1 API
www.thesportsdb.com/api/v1/json/851891/searchteams.php?t=Arsenal

V2 API
www.thesportsdb.com/api/v2/json/livescore/soccer

Upcoming Events:
None found

Last Watched Events:
None found. Add some on the events pages.



TheSportsDB V2 API

- > search
- > lookup
- > list
- > filter
- > all
- > schedule
- > livescore



Python | PRODUCERS

- Developed Kafka producers using a Python script to publish data to the topics

The screenshot shows a file explorer window with a dark theme. At the top level, there is a folder named ".venv". Below it is a "config" folder containing a ".env" file. The main directory is named "producers", which contains several Python files: __pycache__, broadcast_producer.py, common.py, create_topics.py, dynamic_once_livescore.py, event_details_producer.py, event_lookup_producer.py, event_producer.py, league_producer.py, livescore_producer.py, player_producer.py, schedule_producer.py, team_producer.py, and venue_producer.py.

```
> .venv
config
  .env
producers
  __pycache__
  broadcast_producer.py
  common.py
  create_topics.py
  dynamic_once_livescore.py
  event_details_producer.py
  event_lookup_producer.py
  event_producer.py
  league_producer.py
  livescore_producer.py
  player_producer.py
  schedule_producer.py
  team_producer.py
  venue_producer.py
```

VALID DATA FLOW





KAFKA TOPICS

Designed and implemented a Kafka topic architecture with:

- Raw topics
- Valid topics
- Invalid topics

The image displays two side-by-side screenshots of a Kafka topic management interface, likely from a tool like Confluent Cloud or a custom application. Both screens have a dark header bar with a search bar labeled "Search by Topic Name". Below the header are buttons for "Delete selected topics" and "Copy selected to...". The main area is a list of topics, each preceded by a small square icon.

Left Screenshot (Raw Topics):

Topic Name
rejected.soccer.broadcast
rejected.soccer.event
rejected.soccer.event.highlights
rejected.soccer.event.lineup
rejected.soccer.event.stats

Right Screenshot (Valid Topics):

Topic Name
soccer.venue
validated.soccer.broadcast
validated.soccer.event
validated.soccer.event.highlights
validated.soccer.event.lineup
validated.soccer.event.stats
validated.soccer.event.timeline



DATA VALIDATION

- Ensured that all ID fields contain valid (**non-null**) values.
- Filtered the dataset to include only records where sport = '**soccer**'.
- Added the **ingested_at** timestamp during data ingestion for data warehouse requirements.





STREAM PROCESSING

ClickPipes is responsible for automatically ingesting data from Kafka into ClickHouse without the need for custom ETL code.

In your case, it:

- Consumes data from Kafka topics
- Loads it into ClickHouse staging tables.

Name	Type	Status	Data Size	Table data	Records	Created on
raw_event_lookup_stream	Apache Kafka	Running	7.85 KB	View table Preview	23	August 10, 2025
raw_venue	Apache Kafka	Running	43 KB	View table Preview	97	August 9, 2025
raw_team	Apache Kafka	Running	72.88 KB	View table Preview	155	August 9, 2025
raw_schedule	Apache Kafka	Running	54.84 KB	View table Preview	1,382	August 9, 2025
raw_player	Apache Kafka	Running	476.37 KB	View table Preview	2,670	August 9, 2025
raw_livescore	Apache Kafka	Running	29.01 KB	View table Preview	1,243	August 9, 2025
raw_league	Apache Kafka	Running	16.19 KB	View table Preview	18	August 9, 2025
raw_event_timeline	Apache Kafka	Running	55.43 KB	View table Preview	1,596	August 9, 2025
raw_event_stats	Apache Kafka	Running	9.08 KB	View table Preview	846	August 9, 2025
raw_event_lineup	Apache Kafka	Running	35.63 KB	View table Preview	1,266	August 9, 2025
raw_event_highlights	Apache Kafka	Running	48.84 KB	View table Preview	372	August 9, 2025
raw_event_stream	Apache Kafka	Running	584.14 KB	View table Preview	9,654	August 9, 2025
raw_channel_stream	Apache Kafka	Running	16.03 KB	View table Preview	150	August 9, 2025

DATA MODELING





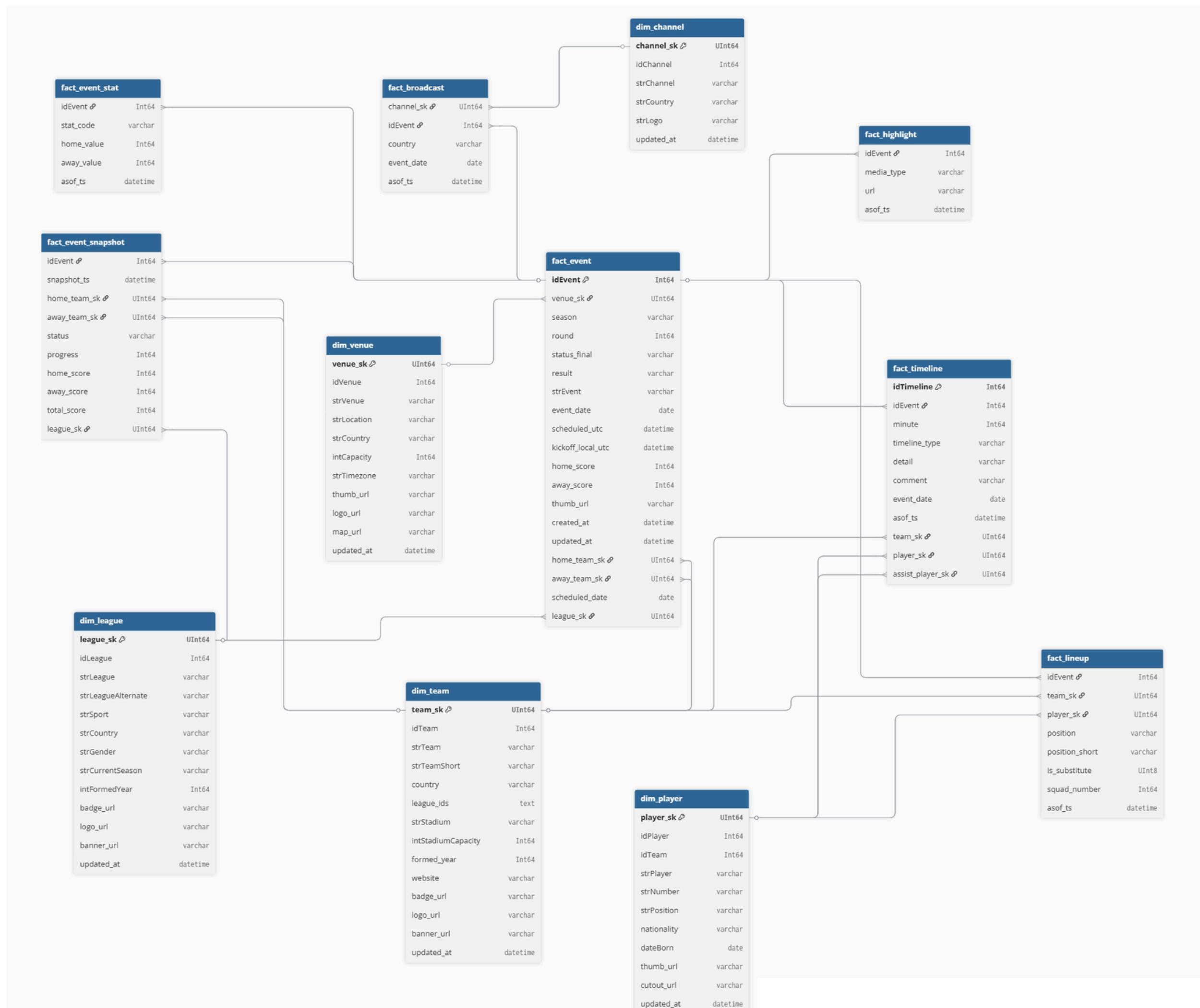
STAR CONSTELLATION SCHEMA

followed the Kimball data model approach, which is data warehouse design methodology, by creating schemas based on **data marts**.

There are two main types of data processing:

- real-time
- batch data





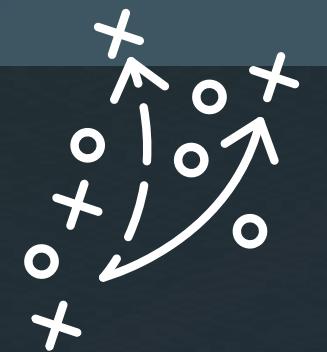


DATA WAREHOUSE

- Designed and implemented a scalable data warehouse in ClickHouse, structured into 5 dimension tables and 6 fact tables, optimizing query performance and supporting complex analytical workflows.

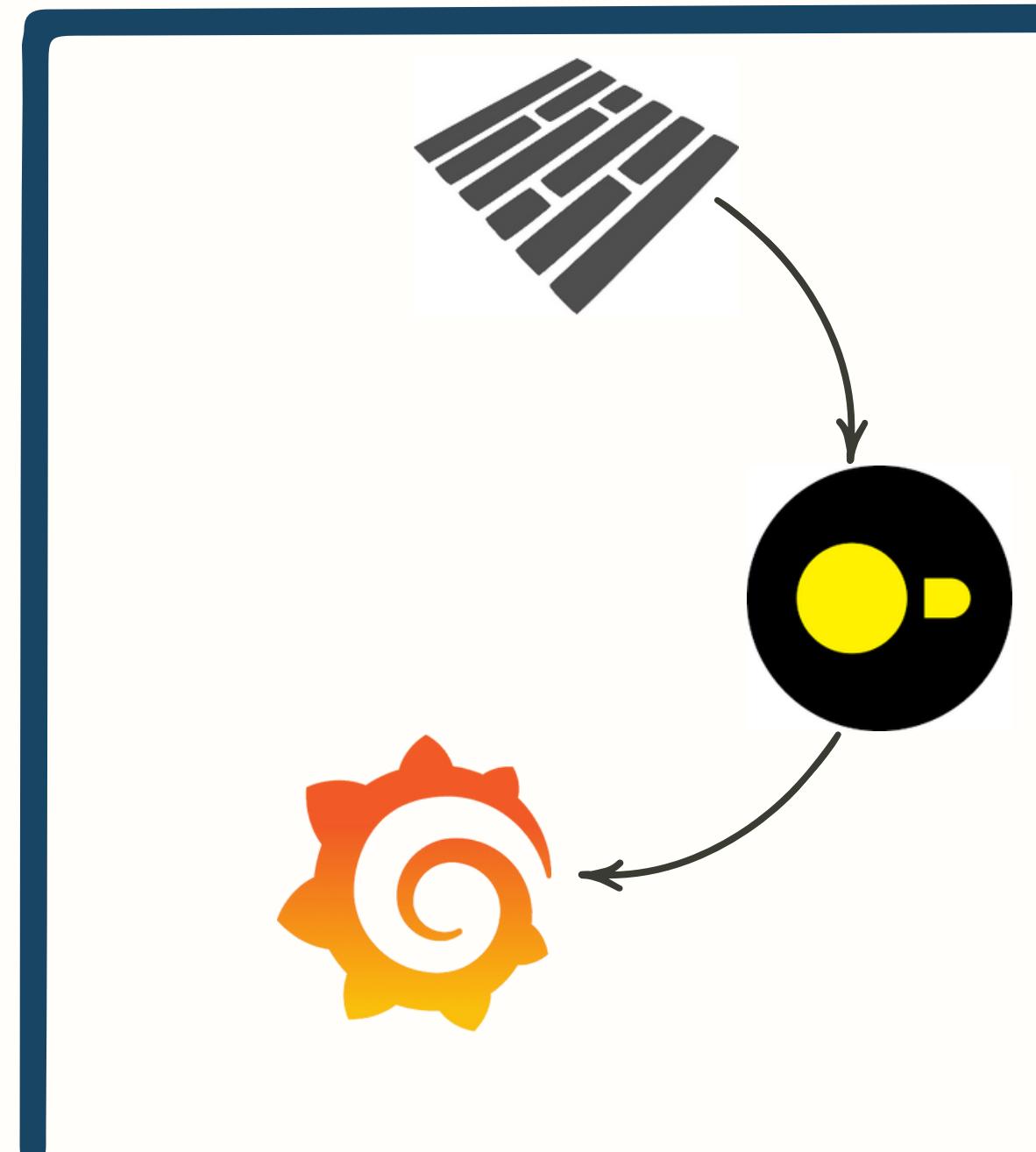
Tables (12)	
dim_channel	RMT
dim_league	RMT
dim_player	RMT
dim_team	RMT
dim_venue	RMT
fact_broadcast	MT
fact_event	RMT
fact_event_snapshot	MT
fact_event_stat	MT
fact_highlight	MT
fact_lineup	MT
fact_timeline	MT

INVALID DATA FLOW



DATA DUMP

We used DuckDB to directly read and query large Parquet files without the need for data loading into a traditional database. This allowed us to run complex OLAP queries efficiently and connect the results to Grafana for real-time visualization and analytic.



ORCHESTRATION WITH APACHE AIRFLOW

Central scheduler to manage all producers & data flows:

- Scheduling: Different frequencies for historical & live data.
- Retries: 3 attempts before sending failure email.
- Alerts: Daily dashboard + email with invalid data analysis.
- Monitoring: Using DuckDB + Grafana to get insights on rejected messages





AIRFLOW DAGS OVERVIEW

Total: 11 DAGs

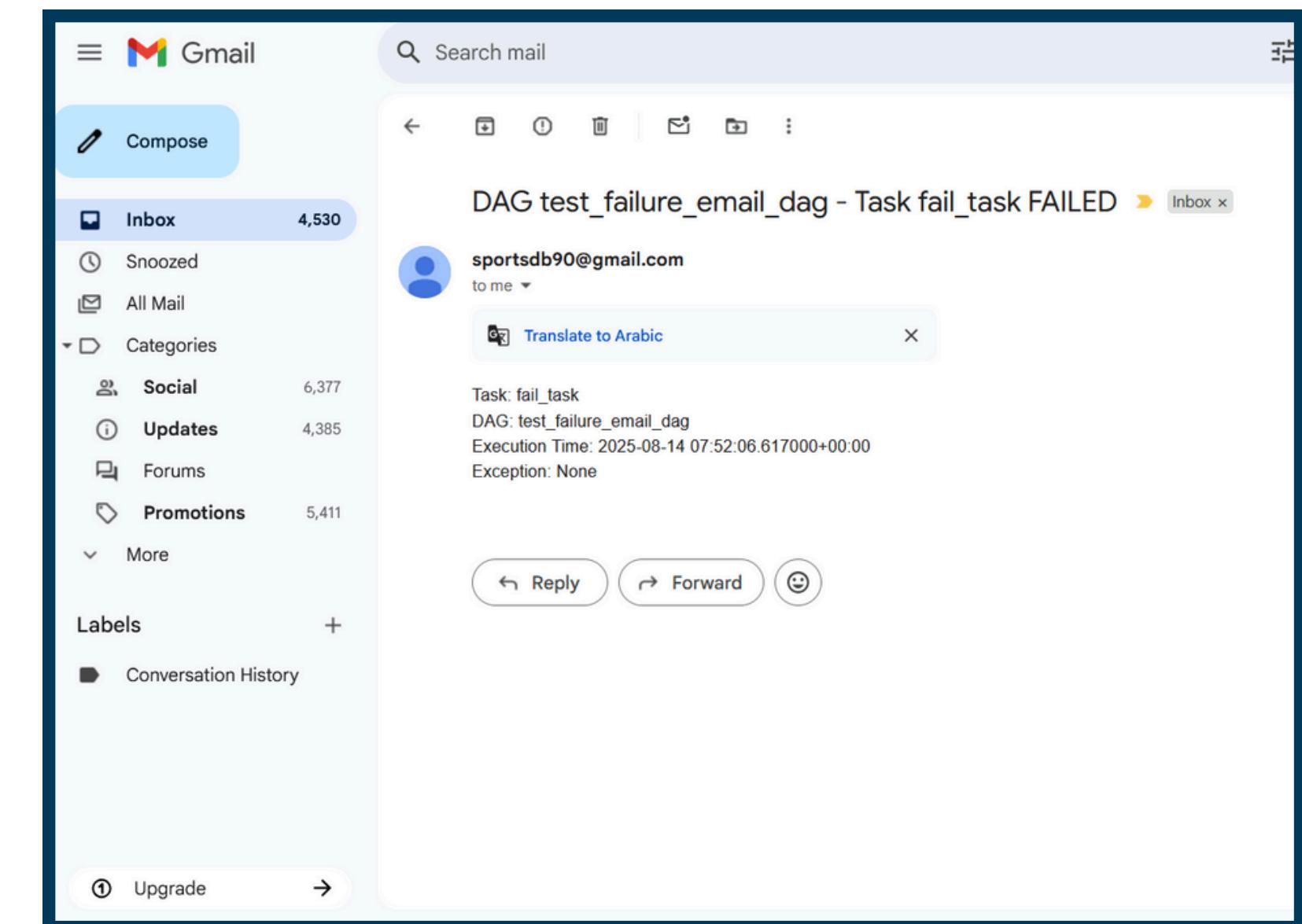
- Historical Data (daily at midnight)
- Live Data (every 2 minutes)
- Invalid Data (daily at 6 AM)

DAG	Schedule	Latest Run	Next Run
event_lookup_dag	0:10:00	2025-08-13, 23:49:44	2025-08-13, 23:59:44
event_proucer_daily_dag	0 0 * * *	2025-08-13, 23:05:52	2025-08-14, 00:00:00
event_stats_daily_dag	0 0 * * *	2025-08-13, 23:03:22	2025-08-14, 00:00:00
league_proucer_daily_dag	0 0 * * *	2025-08-13, 23:00:01	2025-08-14, 00:00:00
live_score_dag	0:02:00	2025-08-13, 23:53:22	2025-08-13, 23:55:22



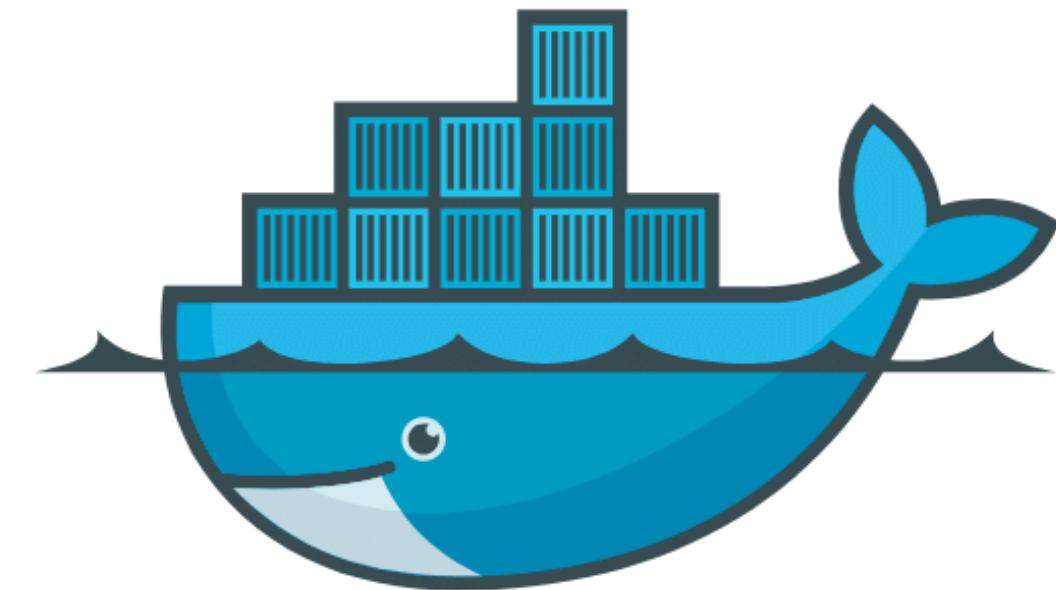
ALERTING OF DAGS

He implemented a module that triggers an automated email alert in case of any failure in our producers, ensuring that issues are promptly reported to the development.



| DOCKER

- We used Docker to deploy virtualized containers across multiple machines, ensuring high availability and avoiding a single point of failure.
- By distributing services like Kafka, Spark, Grafana, and Airflow across different hosts, the system can continue operating even if one machine goes down.



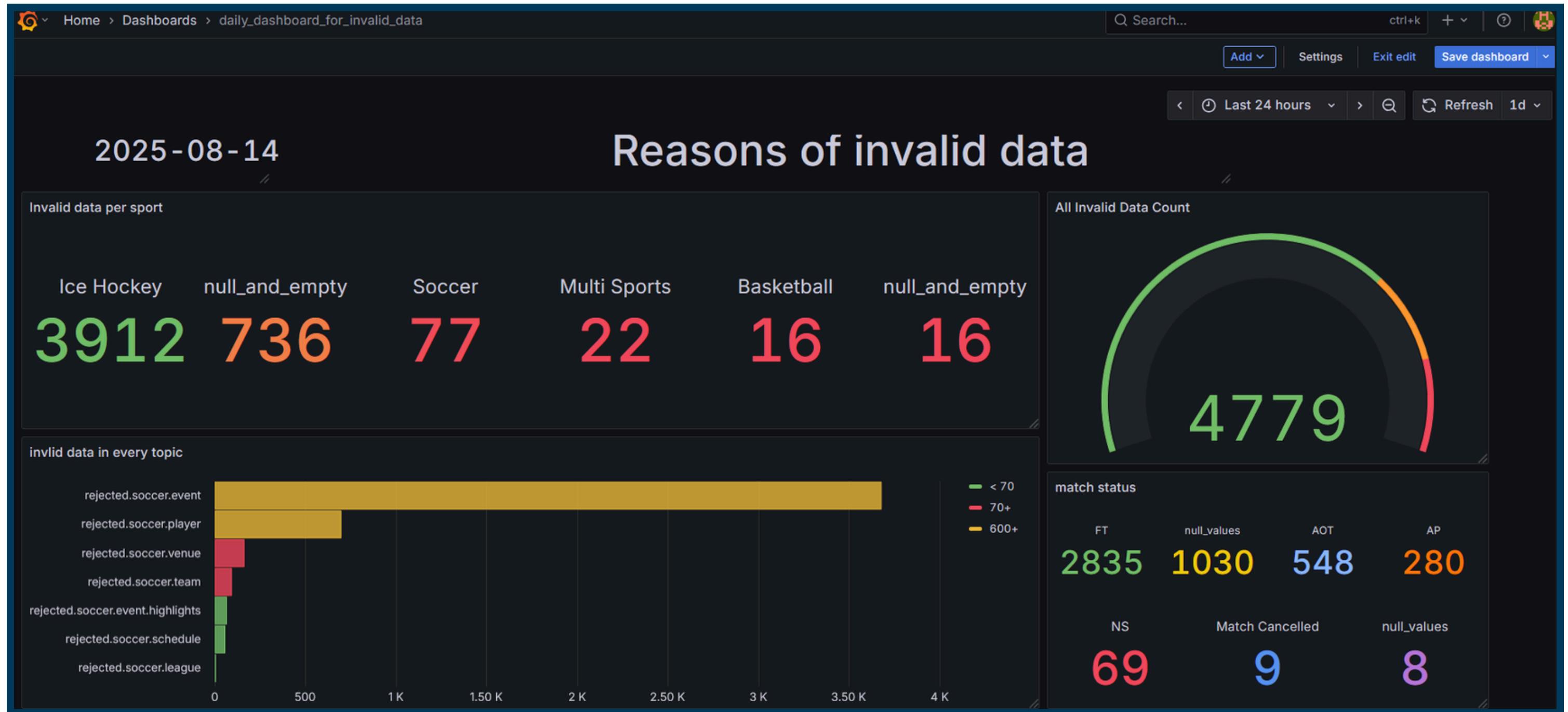


ADMIN DASHBOARD

We developed an admin dashboard that provides detailed insights into invalid data, including the specific records, error types, and timestamps.



ADMIN DASHBOARD





MONITORING OF GRAFANA

It is responsible for daily monitoring on Grafana, automatically sending email updates summarizing any changes or alerts detected on the dashboard.

The screenshot shows an email from Grafana to the user 'sportsdb90@gmail.com'. The subject of the email is '[FIRING:1] DatasourceError invalid data (Invalid data per sport)'. The email body contains a summary of the alert, labels, annotations, and links to silence the alert, view the dashboard, or view the panel. The Grafana logo is visible in the top right corner of the email interface.

[FIRING:1] DatasourceError invalid data (Invalid data per sport) [Inbox](#)

Grafana <sportsdb90@gmail.com>
to me

Grafana

invalid data > DatasourceError

1 firing instances

Firing **DatasourceError** **View alert**

Summary
There is a new invalid data . Take a look

Labels
alertname
grafana_folder
rulename
DatasourceError
invalid data
Invalid data per sport

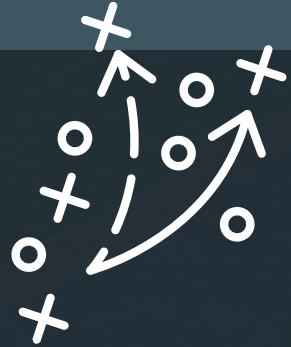
Annotations
Err invalid format of evaluation results for the alert definition : frame cannot uniquely be
or identified by its labels: has duplicate results with labels {sport=}

Silence **View dashboard** **View panel**

Observed 33s before this notification was delivered, at 2025-08-14 03:43:40 +0000 UTC

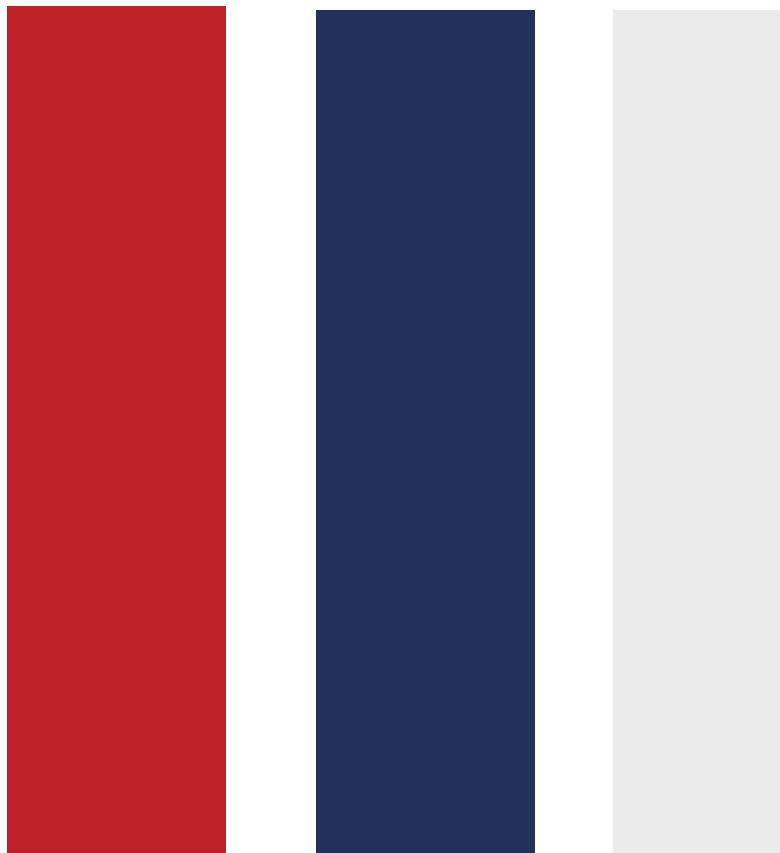
© 2025 Grafana Labs. Sent by [Grafana v12.1.1](#).

DASHBOARD





PROJECT BRANDING



With each Dashboard we made sure to choose the most what makes the project pop more, and ensure captivating the user's attention to the correct workflow





PROJECT BRANDING



inspired by ancient Egyptian outlines and
Ancient Greek Olympics art to tie the old with a
new modern twist





BACKGROUND COLOURS



With each Dashboard we made sure to choose the most what makes the project pop more, and ensure captivating the user's attention to the correct workflow





DEMO – LIVE MATCHES

KICK HOUSE

league	Home_Team	score	Away_Team	status	last_update	activity
Spanish La Liga	Girona	0 - 0	Rayo Vallecano	1H	2025-08-15 21:20:22	ongoing
Australia Tasmania NPL	Glenorchy Knights	1 - 1	Kingborough Lions United	2H	2025-08-15 16:10:22	finished
Australia South Australia NPL	Para Hills Knights	4 - 4	North Eastern MetroStars	2H	2025-08-15 16:00:22	finished

Updates

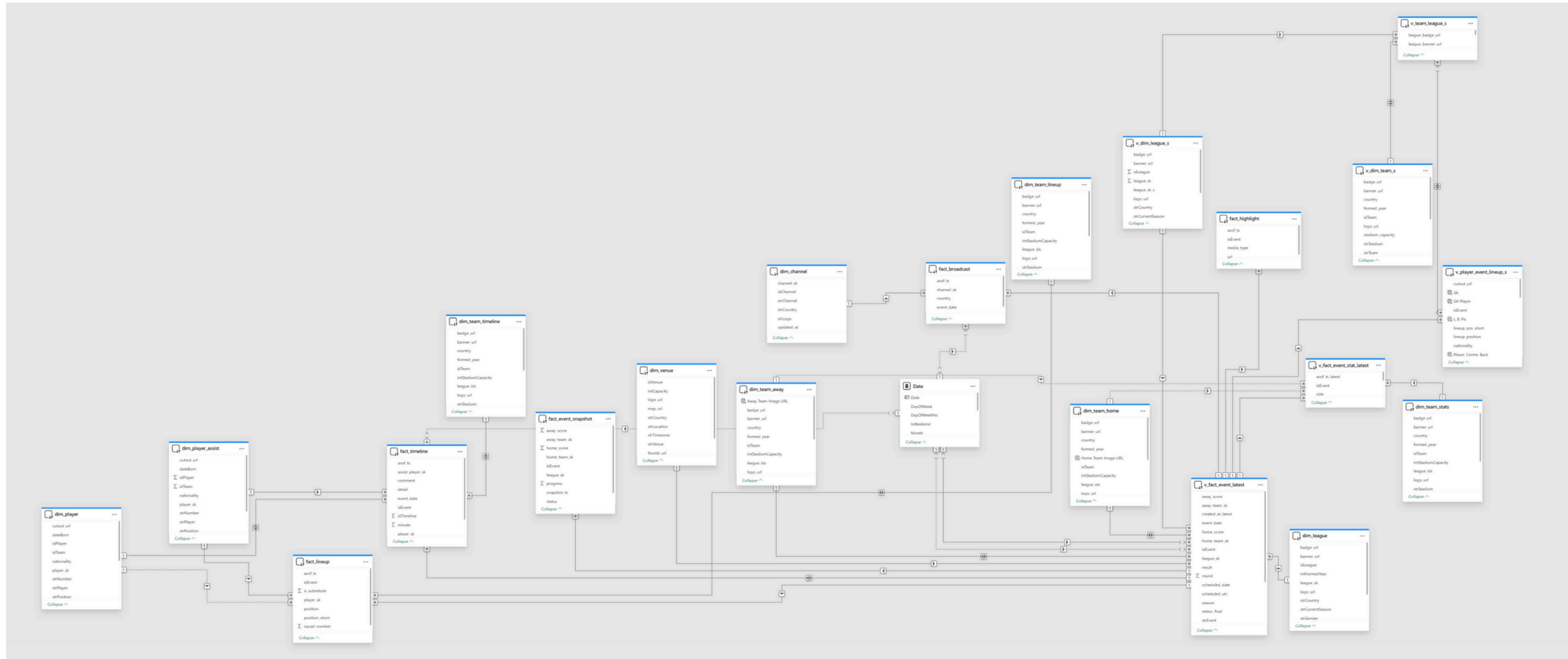
- > Ferencváros vs Noah
- > Malmo FF vs Iberia 1999
- > Al Qadsiah vs Al-Orobah

Live Goals by League

League	matches	goals	gpm
alia South Australia NPL	10	10	1.0
Club Friendlies	2	1	0.5
Brazilian Serie B	1	1	1.0



OUR POWER BI SCHEMA





DEMO – SOCCER ANALYTICS

The screenshot shows a Microsoft Power BI report titled "KickHouse". The report features a large soccer ball graphic on the left and a grid of player silhouettes on the right. The title "KICK HOUSE" is prominently displayed in red. Below the title, a subtitle reads: "Historical Matches, Teams, Players & Libraries right at your fingertips." At the bottom, there are five navigation tabs: "Schedule", "Past Matches" (which is selected), "Player", "Team", and "Line-up". The report is presented in a clean, modern interface with a white background and a dark sidebar.

KickHouse • Last saved: Today at 10:18 PM

Search

File Home Insert Modeling View Optimize Help

Sensitivity Publish Prep data for AI Copilot ...

Share

KICK HOUSE

Historical Matches, Teams, Players & Libraries
right at your fingertips.

Schedule / Past Matches / Player / Team / Line-up

Cover Schedule Past Matches Player Team Lineup +

Storage Mode: Mixed 116%



CONCLUSION

Successfully built a real-time sports analytics pipeline with reliable ingestion, processing, and visualization, leveraged ClickHouse, ClickPipes, Kafka, Spark, and DuckDB for scalable and efficient data handling, delivered curated data models enabling advanced analysis in Power BI and Grafana.





THANK YOU!

