NFL Intelligence Hub - Development Log

# Project Overview

The NFL Intelligence Hub is a full-stack application designed to collect, process, analyze, and visualize NFL player statistics. It integrates data engineering, backend API development, and frontend design. The tech stack includes React, Tailwind CSS, Golang, PostgreSQL, PySpark, and optional Airflow for ETL scheduling.

# Planning Phase

- Defined project scope and objectives  
- Selected tech stack suitable for junior to intermediate college-level developers  
- Designed database schema: players, teams, rushing\_stats, etc.  
- Established Git workflow and project folder structure

# Data Engineering

- Web scraping scripts created using `requests` and `BeautifulSoup` for stats like rushing, kicking, etc.  
- Extracted and transformed data into structured pandas DataFrames  
- Cleaned and validated data (e.g., removing header rows, missing values)  
- Inserted cleaned data into PostgreSQL using SQLAlchemy  
- Resolved data mismatches, like inconsistent team abbreviations

# Database Design and Issues

- Created tables with proper primary and foreign key constraints  
- Resolved foreign key issues due to mismatched column names or missing references  
- Used `ON CONFLICT` SQL clauses to handle duplicates in `players` and `teams`  
- Inserted base team and player data before adding statistics to maintain referential integrity  
- Example bug: NULL `player\_id` in `rushing\_stats` traced back to FK mismatch

# API Development (Planned)

- Decided to use FastAPI and Uvicorn for serving data  
- Planned to start API with GET requests only  
- Will add Pydantic models if response validation becomes necessary  
- Deferred creation of SQL indexes until after basic endpoint testing

# Bug Fixes and Technical Issues

- Fixed team name abbreviation mismatches using SQL update logic  
- Identified and corrected column naming conflicts causing FK reference issues  
- Addressed insert errors due to incorrect or incomplete data values (e.g., 2TM issue)  
- Ensured proper data copying and Pandas view/assignment warnings were avoided

# Next Steps

- Build out API endpoints and validate with frontend requests  
- Construct frontend UI with team/player filters and stat charts  
- Optional: Automate ETL using Airflow or scheduled cron jobs  
- Finalize README with architecture diagram and setup instructions