Golf Analytics Platform

Thomas Kulch

June 2025

1 Project Kickoff

Project Goals
The primary objective of this project is to design and implement a comprehensive Real-Time Golf Analytics Platform that utilizes big data technologies and cloud computing to demonstrate enterprise-scale data management capabilities. This platform will simulate the data infrastructure required to manage multiple golf courses while providing real-time analytics, simple predictive modeling, and operational optimization.
Project Scope
•Multi-course golf chain simulation.
•Real-time data streams including live scoring, weather integration, and booking systems
•Historical data analysis covering years of simulated golf operations.
•Simple machine learning models for pricing optimization, demand forecasting, and player analytics.
•Cloud-based data lake architecture with bronze/silver/gold data layers.
•Interactive dashboards for course management and player insights. (Potentially out of scope)

Project Deliverables
Cloud Infrastructure: Fully deployed AWS architecture with data lake, processing clusters, and streaming services
Data Pipeline: End-to-end ETL/streaming pipeline processing millions of records
Machine Learning Models: Trained models for pricing optimization and player analytics
Real-Time Dashboard: Interactive analytics platform for course operations
Technical Documentation: Architecture diagrams, code documentation, and deployment guides
Final Report: Comprehensive analysis of system performance, scalability, and business insights
Presentation: Technical presentation demonstrating system capabilities and findings

Datasets
This will be a mixture of generated historical golf data and "real time" data for live scoring updates.

2 Team Discussions

My skills include Python programming, algorithms and data structures, pandas, basic sklearn, advanced SQL with database administration experience and ETL with SSIS. Knowledge of product/project instrastructure will help with this project as well. I will be using Python, PySpark/Apache and hopefully AWS for this project.// My skill gaps are in AWS, I have basic skills as I host a discord bot on AWS and I work with Azure servers at work. I have no knowledge of PySpark/Apache and hope to learn through lectures/reading/researching and trial and error through this project.

3 Skills and Tools Assessment

•SQL
•Python/Pandas
•PySpark
•AWS
•Streamlit for website

External resources will be code documentation, youtube, 5110 class content, LLM's for guidance

4 Submission for This Iteration

My initial challenge is visualizing what this project will look like as I have never done something like this before. I will rely on research and lecture content to help plan this out.

I originally wanted this to be a database project with some data preprocessing, database management and visualizations, but I have done projects like this and I want to branch out my skills and do something new to put on my resume.

I have attached my Excel tracker and will use that to discipline myself and set goals as well as track progress.