High Level Design (HLD)

FIFA World Cup

Revision Number: 1.0
Last date of revision: 01/06/2021

Anupam Saha

Document Version Control

Date Issued	Version	Description	Author
5th May 2023.	1.0	First Version of Complete HLD	Anupam Saha

Contents

Abstract

The FIFA World Cup is one of the most anticipated and prestigious international sporting events, captivating billions of fans worldwide. With its rich history and immense popularity, the tournament serves as a valuable source of data for comprehensive analysis. This abstract presents an overview of a dataset comprising diverse information about the FIFA World Cup, covering multiple aspects including teams, players, matches, and outcomes.

The dataset contains detailed records from various editions of the tournament, spanning decades of football history. It includes data on participating teams, their performance metrics, player attributes, match statistics, and final outcomes. The information encompasses a wide range of variables such as goals scored, assists, yellow and red cards, possession, shots on target, and more.

This study aims to leverage the FIFA World Cup dataset to gain insights into the dynamics of the competition, team performances, player trends, and strategic patterns across different editions. By utilizing data analysis techniques and statistical methods, we seek to uncover patterns, correlations, and significant factors that contribute to the success or failure of teams in the tournament. Furthermore, the dataset allows for the exploration of historical trends, such as changes in playing styles, evolution of tactics, and the impact of rule modifications on the game.

The findings derived from this analysis can offer valuable insights to football enthusiasts, coaches, analysts, and researchers. The comprehensive examination of team and player performance can aid in strategic decision-making, talent identification, and performance evaluation. Moreover, the analysis may uncover intriguing patterns and highlight the significance of various factors that influence the outcome of matches, potentially providing valuable knowledge for the development of future football strategies.

In conclusion, this abstract introduces a comprehensive dataset encompassing the FIFA World Cup, inviting researchers to explore the dataset's extensive information. The analysis of this dataset has the potential to yield valuable insights into the intricacies of the global football competition, fostering a deeper understanding of team and player performances and contributing to the broader knowledge base of the beautiful game.The championship has been awarded every four years since the inaugural tournament in 1930, except in 1942 and 1946 when it was not held because of the Second World War. The current champion is Germany, which won its fourth title at the 2014 tournament in Brazil.

1 Introduction

1.1 Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

- Present all of the design aspects and define them in detail
- Describe the user interface being implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- Include design features and the architecture of the project

5 High Level Design (HLD)

- List and describe the non-functional attributes like:
    - Security

- Reliability
- Maintainability
- Portability
- Reusability
- Application compatibility
- Resource utilization
- Serviceability

1.2 Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

2 General Description

2.1 Product Perspective & Problem Statement

The project provides the following perspectives that need to be addressed.

1. Tournament Winners and Runners-up: A FIFA World Cup dataset allows you to identify the countries that have won the tournament and finished as runners-up. You can analyze the frequency of their success, identify dominant periods in World Cup history, and track any patterns in the distribution of titles among different nations.

2. Most Successful Teams: By examining the number of World Cup titles won by each country, you can identify the most successful teams in the history of the tournament. This provides insights into the powerhouses of international football and their historical dominance.

3. Team Performance by Stage: By examining team performances in each stage of the tournament, such as the group stage, knockout rounds, and the final, you can identify teams that have consistently performed well or struggled in specific stages. This provides insights into teams' ability to handle different levels of competition.

4. Goalscoring Trends: World Cup datasets enable analysis of goalscoring trends over time. You can identify patterns in the average number of goals per match, changes in goal-scoring techniques, and assess whether the tournament has witnessed an increase or decrease in goalscoring rates.

5. Goal Distribution: By examining the distribution of goals across different stages of the tournament, you can identify whether there are specific stages where more goals are scored, such as the group stage or knockout rounds. This provides insights into the varying levels of intensity and competitiveness throughout the tournament.

This project aims apply various Business Intelligence tools such as Tableau or Power BI to get a visual understanding of the data.

2.2 Tools used

Business Intelligence tools and libraries works such as Numpy, Pandas, Excel, R, Tableau, Power BI are used to build the whole framework.

3 Design Details

3.1 Functional Architecture

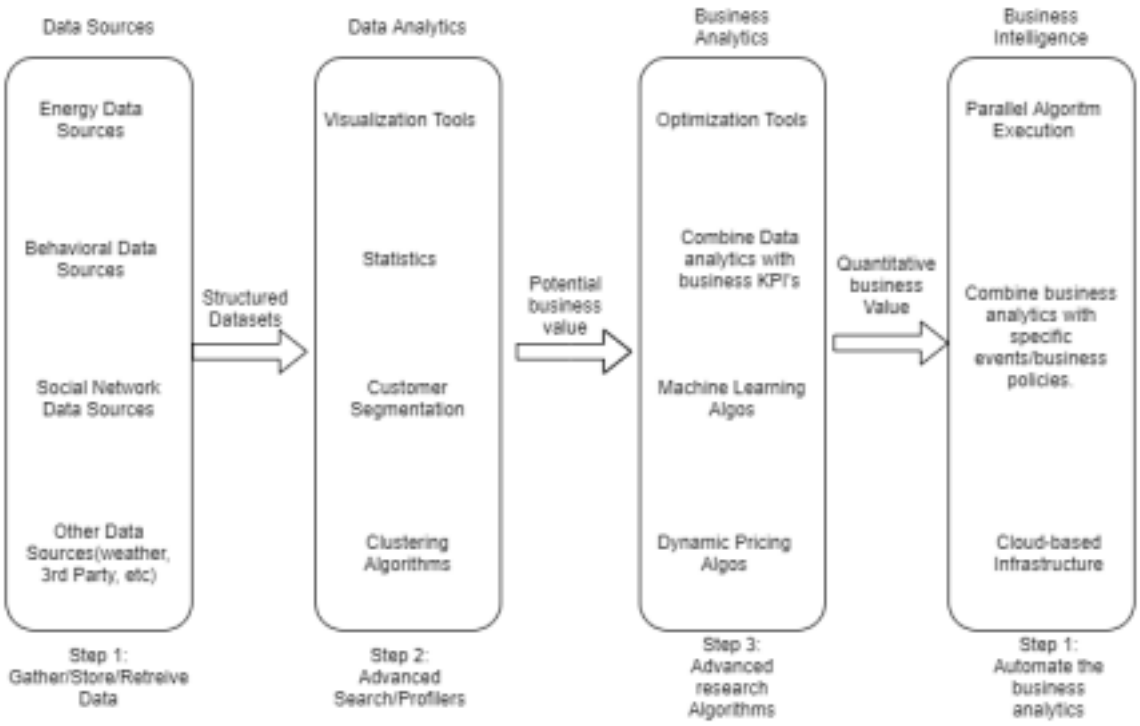
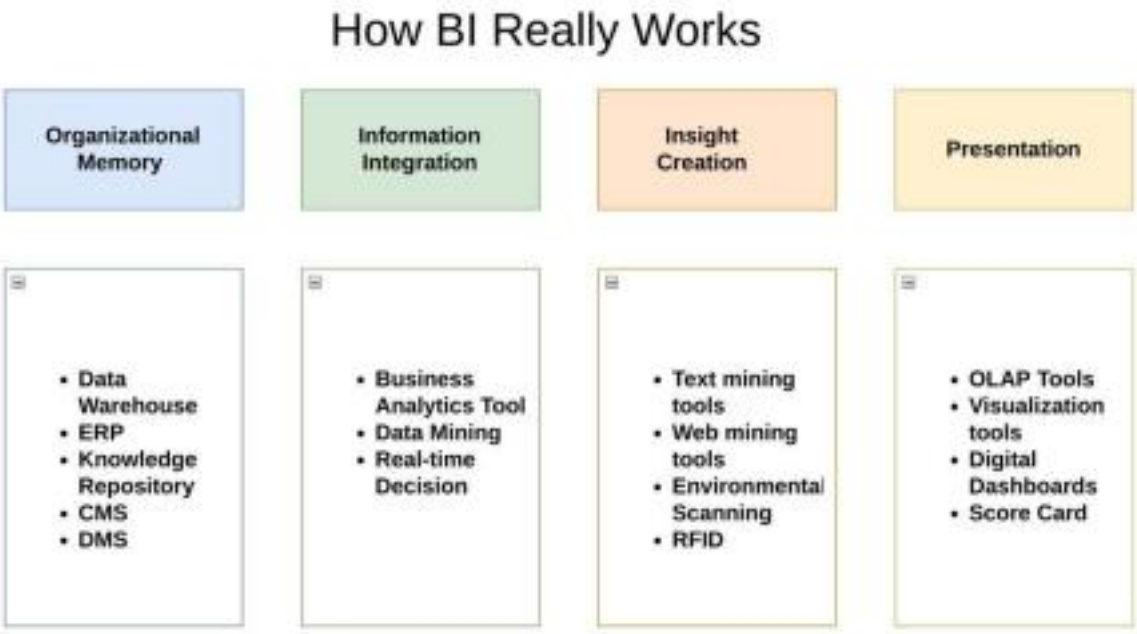


Figure 1: Functional Architecture of Business Intelligence



3.2 Optimization

Your data strategy drives performance

- Minimize the number of fields
- Minimize the number of records
- Optimize extracts to speed up future queries by materializing calculations, removing columns and the use of accelerated views

Reduce the marks (data points) in your view

- Practice guided analytics. There's no need to fit everything you plan to show in a single view. Compile related views and connect them with action filters to travel from overview to highly-granular views at the speed of thought.
- Remove unneeded dimensions from the detail shelf.
- Explore. Try displaying your data in different types of views.

Limit your filters by number and type

- Reduce the number of filters in use. Excessive filters on a view will create a more complex query, which takes longer to return results. Double-check your filters and remove any that aren't necessary.
- Use an include filter. Exclude filters load the entire domain of a dimension, while include filters do not. An include filter runs much faster than an exclude filter, especially for dimensions with many members.
- Use a continuous date filter. Continuous date filters (relative and range-of-date filters) can take advantage of the indexing properties in your database and are faster than discrete date filters.

- Use Boolean or numeric filters. Computers process integers and Booleans (t/f) much faster than strings.
- Use parameters and action filters. These reduce the query load (and work across data sources).

Optimize and materialize your calculations

- Perform calculations in the database
- Reduce the number of nested calculations.
- Reduce the granularity of LOD or table calculations in the view. The more granular the calculation, the longer it takes.
  - LODs - Look at the number of unique dimension members in the calculation.
    - Table Calculations - the more marks in the view, the longer it will take to calculate.
- Where possible, use MIN or MAX instead of AVG. AVG requires more processing than MIN or MAX. Often rows will be duplicated and display the same result with MIN, MAX, or AVG
  - Make groups with calculations. Like include filters, calculated groups load only named members of the domain, whereas Tableau's group function loads the entire domain.
- Use Booleans or numeric calculations instead of string calculations. Computers can process integers and Booleans (t/f) much faster than strings.
  Boolean>Int>Float>Date>DateTime>String

4 KPIs

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators for the insights.

As and when, the system starts to capture the historical/periodic data for a user, the dashboards will be included to display charts over time with progress on various indicators or factors

4.1 KPIs (Key Performance Indicators)

KPIs provide a framework for evaluating team and player performances, strategic effectiveness, and overall success in the FIFA World Cup. They assist in benchmarking and comparing teams, identifying areas for improvement, and tracking performance trends across editions of the tournament.

- Most Number of World Cup Winning Title
- Number of Goal Per Country
- Attendance, Number of Teams, Goals, and Matches per Cup
- Goals Per Team Per World Cup
- Matches With Highest Number Of Attendance
- Stadium with Highest Average Attendance
- Which countries had won the cup ?
- Number of goal per country
- Match outcome by home and away teams

5 Deployment

Prioritizing data and analytics couldn't come at a better time. Your company, no matter what size, is already collecting data and most likely analyzing just a portion of it to solve business problems, gain competitive advantages, and drive enterprise transformation. With the

explosive growth of enterprise data, database technologies, and the high demand for analytical skills, today's most effective IT organizations have shifted their focus to enabling self-service by deploying and operating Tableau at scale, as well as organizing, orchestrating, and unifying disparate sources of data for business users and experts alike to author and consume content.

Tableau prioritizes choice in flexibility to fit, rather than dictate, your enterprise architecture. Tableau Server and Tableau Online leverage your existing technology investments and integrate into your IT infrastructure to provide a self-service, modern analytics platform for your users. With on-premises, cloud, and hosted options, there is a version of Tableau to match your requirements. Below is a comparison of the three types:

\

Pros of Power BI:

- User-Friendly Interface: Power BI offers a user-friendly interface that makes it easy for both technical and non-technical users to create interactive visualizations and reports. The drag-and-drop functionality and intuitive design tools simplify the process of data analysis and presentation.
- Data Integration: Power BI allows integration with a wide range of data sources, including databases, spreadsheets, cloud services, and online platforms. This enables users to consolidate and combine data from multiple sources, creating a unified view for analysis.
- Interactive Visualizations: Power BI provides a rich set of visualization options, including charts, graphs, maps, and tables. Users can customize and interact with these visualizations, allowing for in-depth exploration of data and facilitating better understanding and insights.

- ·Real-Time Data Analysis: Power BI supports real-time data streaming and analysis, enabling users to monitor key metrics and make data-driven decisions in real-time. This is particularly useful for businesses that require up-to-date information for timely actions and responses.

Cons:

- Steep Learning Curve: While Power BI offers a user-friendly interface, mastering the full capabilities of the tool may require some learning and training. Users with limited technical expertise may find it challenging to explore advanced features and functionalities.
- Licensing Costs: Power BI offers both free and paid versions, with the latter providing more advanced features and capabilities. The paid licenses can be costly for small businesses or individuals on a limited budget, potentially limiting access to certain features.
- Data Processing Limitations: Power BI has certain limitations when it comes to processing large datasets or handling complex data transformations. Users may encounter performance issues or restrictions when working with extensive data volumes or complex data structures.