High Level Design (HLD)

NBA Data Analysis

Revision Number: 1.0
Last date of revision: 20/07/2024
Nandini Verma

Document Version Control

Date Issued	Version	Description	Author
20 July 2024	1.0	First Version of Complete HLD	Nandini Verma

Contents

Abstract

Each year, thousands of players from all around the world submit their name to the National Basketball Association (NBA), in hopes that their talent warrants an invitation to the upcoming draft. The draft a process that enables teams to select the best talent to add to their roster. One of the major obstacles for teams in this process is trying to weed out the bad players, so to speak, from the good. Traditional selection methods have no real empirical backing and currently possess no full-proof measure of ensuring that the players chosen will perform well in the league. The purpose of this research paper is to develop models that can assist scouts in selecting the optimal players and prevent the selection of players who are more prone to "bust" or exhibit poor performance. In this study, I have illustrate how pre-draft characteristics such as an individual's height, hand size, weight, wingspan, and other attributes serve as a predictor for their performance.

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
- List and describe the non-functional attributes like:
    - Security
    - Reliability
    - Maintainability
    - Portability
    - Reusability
    - Application compatibility
    - Resource utilisation
    - 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

NBA data analysis is the analysis based on the player's past performance which will be/can be used by the investor's, head coach, team management to select the righteous and in form player's so that the player can be used at their full potential and it will be beneficial for the team as well as the player itself. In this project, player's past performance will be given explanatory variables that cover many aspects of players abilities and skills.

The objective of the project is to perform data visualisation techniques to understand the insight of the data. This project aims apply various Business Intelligence tools such as 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, Python, Power BI are used to build the whole framework.

3 Design Details

3.1 Functional Architecture

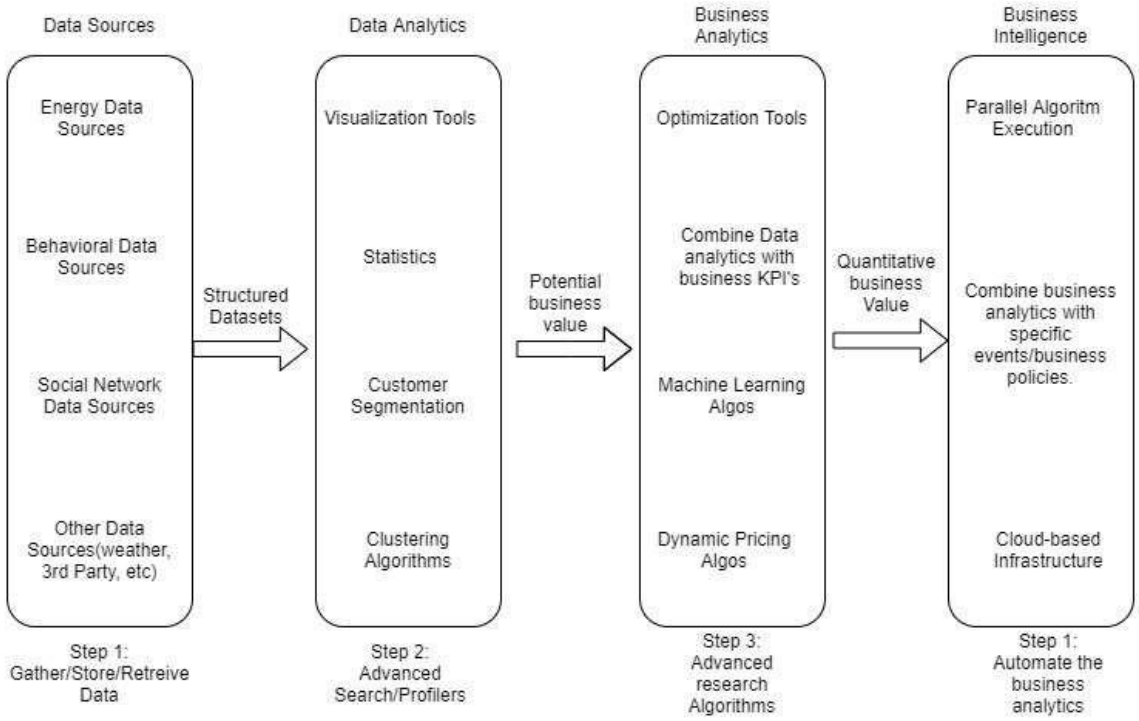
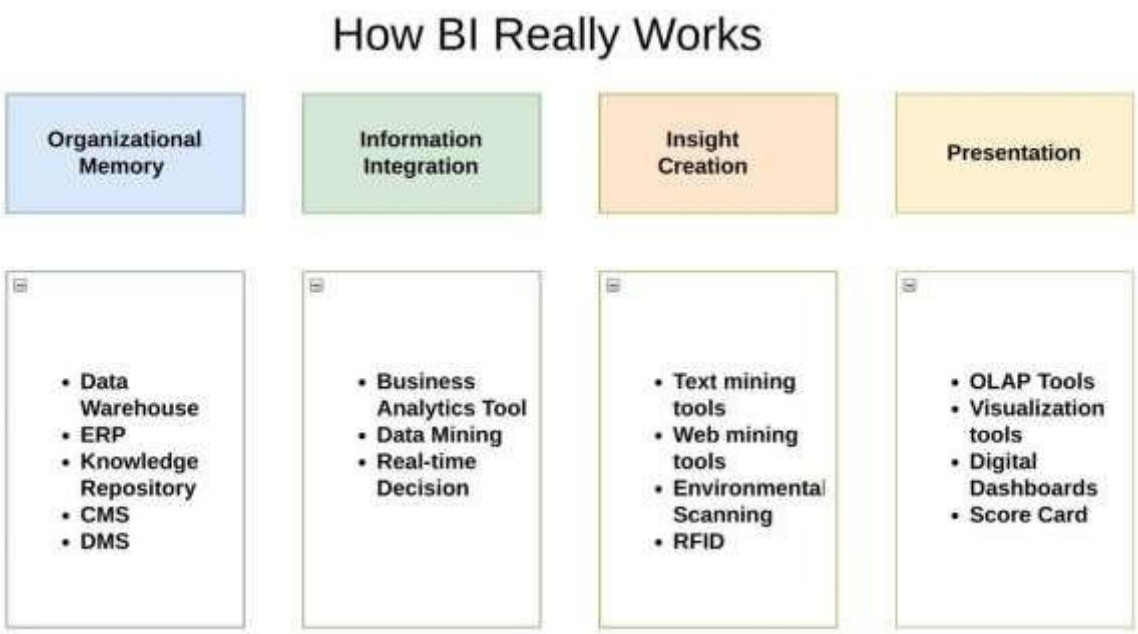


Figure 1: Functional Architecture of Business Intelligence



3.2 Optimisation

Your data strategy drives performance

- Minimise the number of fields
- Minimise the number of records
- Optimise extracts to speed up future queries by materialising 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. Computer's 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).

Optimise and materialise 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 Power BI'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>Date Time>String

4 KPIs

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

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)

Key indicators displaying a summary of the Player and its relationship with different metrics

1. Best Player of the year.
2. Overall Best Player.
3. Tallest Player.
4. Player with most basket
5. Top 5 best players

6. Influence of BMI on player

7. Player with highest agility and sprint

5 Deployment

Prioritising data and analytics couldn't come at a better time. Your company, no matter what size, is already collecting data and most likely analysing 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 organisations have shifted their focus to enabling

self-service by deploying and operating Power BI at scale, as well as organising, orchestrating, and unifying disparate sources of data for business users and experts alike to author and consume content.
Power BI priorities choice in flexibility to fit, rather than dictate, your enterprise architecture. Power BI Server and Power BI 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 Power BI to match your requirements. Below is a comparison of the three types:

TYPE PROS CONS

Power BI Server - On Premises

- Full control of hardware and software
- Infrastructure and data remain behind your firewall
- Need dedicated administrators to manage hardware and software
- Additional infrastructure needed to access off-network (mobile, external)

Power BI Server - Public Cloud (IaaS)

- Full control of software on managed hardware
- Puts infrastructure in same place as data (for migration to cloud)
- Flexibility to spin up/down hardware as needed
- Need dedicated administrators to manage software
- Additional infrastructure needed to access off-network (mobile, external)

Power BI Online (SaaS)

- Fully hosted solution (hardware, software upgrades)
- Fast to deploy
- Easy for external audience to access
- Single-site in multi-tenant environment
- Cubes are not supported
- No guest account access

Depending on your organisational roles and responsibilities, Power BI Server should be installed by a systems administrator and the designated Power BI Server Administrator in coordination with the appropriate IT roles. For Power BI Online, you will integrate with your existing technology and configure the site settings. The Data & Analytics Survey, completed by business teams, identifies and priorities data use cases, audience size, and users. You will use the information collected in both surveys to plan your deployment strategy, including sizing, installation, and configuration of your Power BI Server or integration and configuration of Power BI Online. In addition to installing Power BI Server or configuring Power BI Online, administrators will also need to plan for the client software installation of Power BI Prep Builder, Power BI Desktop, Power BI Mobile, and Power BI Bridge for Power BI Online where applicable.