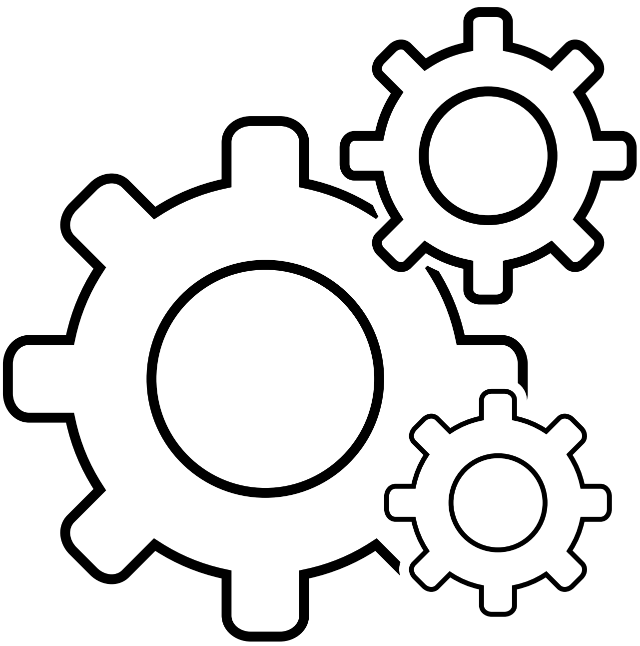
Technical Overview of Huddle's Player Props

Unified market feed, player projections model, Monte Carlo-based simulations, and real-time processing pipelines



INTRODUCTION

The **HuddleOS Odds Feed** comprises a consolidated markets feed encompassing various categories:
1. **Cores and Derivatives**: Standard markets such as moneylines, spreads, and totals, along with their derivatives.
2. **Micros**: Markets characterized by a brief betting period, typically associated with a single play in the game.
3. **Player Props**: Markets linked to player statistics during the course of the game.

While this paper predominantly concentrates on Player Props, it is important to note that the approach to market making remains strikingly similar across all market categories.

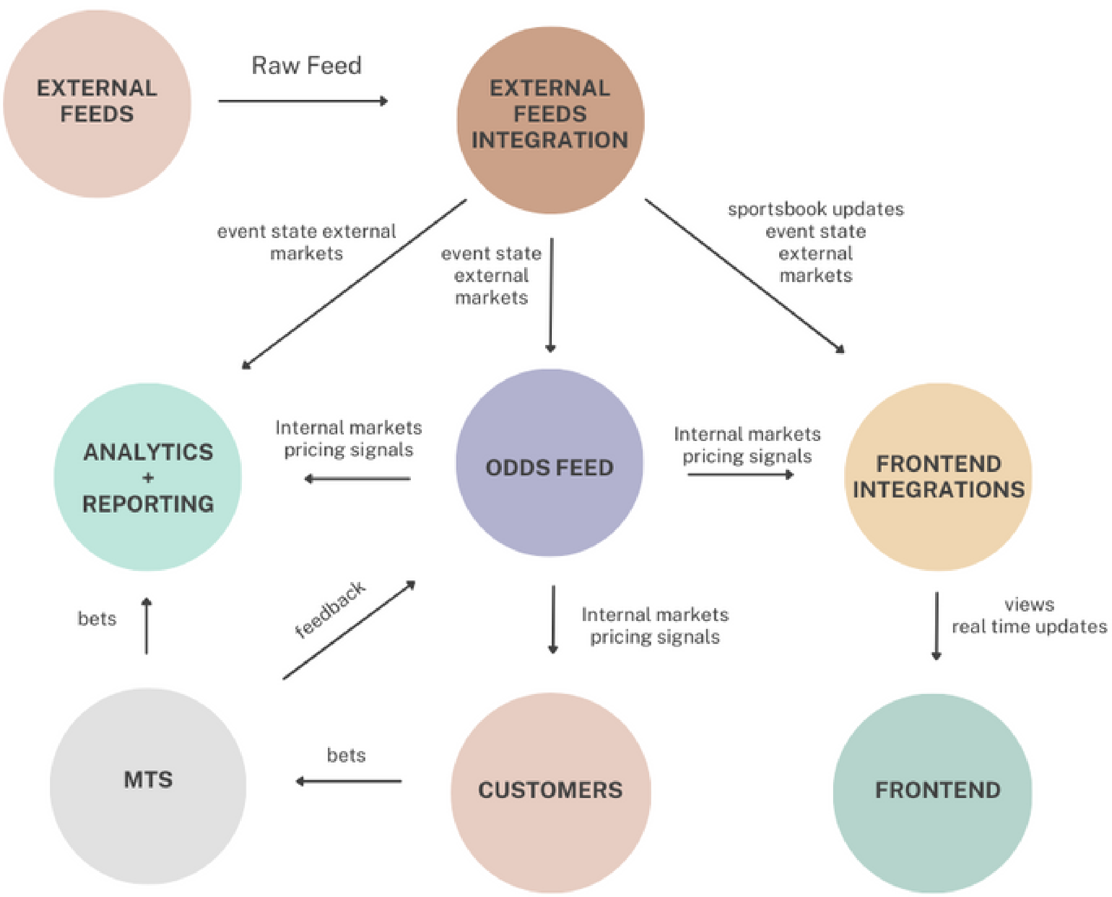


Figure 1 HuddleOS Components

PROJECTING PLAYER STATS

In order to set up **Player Props** in pre-match, it's necessary to collect **historical play-by-play data** from reliable sources. That data then goes through a standard **"extract-transform-load" (ETL) pipeline** which consists of:

1. Downloading the data from an API exposed by data providers
2. Filtering out "falsy" records
3. Transforming the input to unified schema
4. Loading the data into a data lake (a centralized repository designed to store, process, and secure large amounts of structured, semistructured, and unstructured data.)

Once the data is organized in a **structured format**, additional sets of pipelines come into play, taking on the responsibility of executing the projections algorithm. This algorithm then generates **player inputs** essential for the operation of the **pricing algorithm.**

Data Lifecycle

To compute player projections, HuddleOS leverages data obtained from play-by-play data providers via their APIs. The primary big data storage utilized by HuddleOS is ***Snowflake***, chosen for its numerous advantages. *Snowflake* is designed for cloud **scalability**, allowing seamless adjustment of resources based on data processing requirements without the need for infrastructure management. Its architecture separates computing and storage, ensuring **swift query performance**, even with extensive datasets. *Snowflake's* support for **high concurrency** enables multiple users to query and manipulate data concurrently without compromising performance.

Prior to feeding the data into projection models, a preprocessing phase is essential to prepare the data in the appropriate format. HuddleOS employs ***Apache Airflow***, hosted on ***Kubernetes***, to execute ETL pipelines. These pipelines extract data from the source, undergo necessary transformations, and subsequently store the processed data at the designated destination. This meticulous process ensures that the data is appropriately formatted and ready for use in projection models.

Apache Airflow is an open-source platform used for orchestrating complex data workflows, allowing users to define, schedule, and monitor tasks as **Directed Acyclic Graphs (DAGs)**. It provides a scalable and extensible framework for automating data processing pipelines with features for dependency management, monitoring, and scheduling.

Deploying Apache Airflow on Kubernetes enables several benefits. Kubernetes allows for easy **scaling** of Airflow components, such as workers and schedulers, based on workload demands, ensuring that the infrastructure can handle varying levels of data processing tasks efficiently. It also provides built-in f**ault tolerance** mechanisms such as auto-replication and self-healing, ensuring that Airflow remains resilient even in the face of infrastructure failures. Kubernetes **dynamically allocates resources** to Airflow components based on their requirements, optimizing resource utilization and minimizing waste. Kubernetes also allows for **automation** of deployment, scaling, and management of Airflow infrastructure, streamlining operations and reducing manual intervention in managing the infrastructure.

Figure 2 depicts the comprehensive lifecycle of data, commencing from its ingestion from the source. Following the initial transformation, the data finds its repository in Snowflake. Subsequently, a crucial step involves preparing the data to align with the model's requirements. The data must adhere to standards of cleanliness, reliability, and the appropriate format to be effectively utilized by the Player Projections Model, which leverages this input data for the computation of player projections.

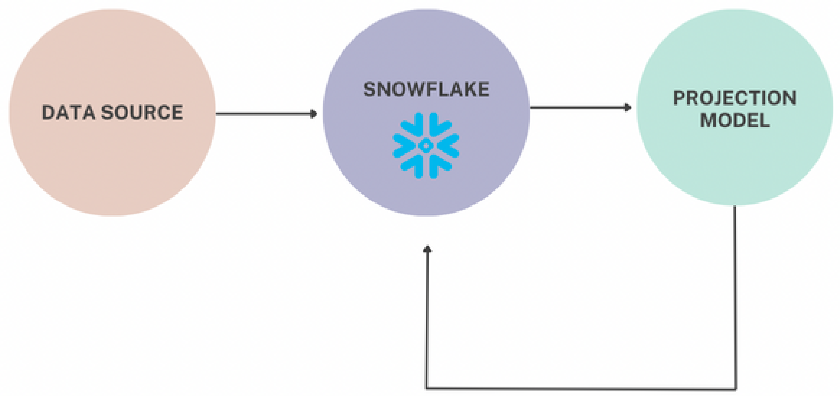


Figure 2 HuddleOS Projections Pipeline

The resultant output from the model is then once again stored in Snowflake, providing a centralized repository. From this juncture, the data becomes versatile, capable of being projected in various ways as dictated by the requirements, utilizing SQL views. Notably, all of HuddleOS's data pipelines are scripted as Airflow Directed Acyclic Graphs (DAGs), facilitating seamless scheduling and monitoring.

Player projections model

In the realm of data analysis and processing, a meticulous approach is adhered to by Huddle. Commencing with data loading, progressing through preprocessing, and culminating in the utilization and integration of machine learning processes, adherence to rigorous and established mathematical methods is paramount. This principle extends to the algorithm responsible for computing player performances in forthcoming games. Calculations are grounded in meticulously verified historical data housed in Snowflake. For each sport, Huddle's quant team meticulously crafts a specialized player projection model, tailoring the methodology to align with the distinctive dynamics and requisites of each game. This bespoke approach ensures the generation of the most accurate and pertinent predictions.

The models extensively delve into the analysis of player types, scrutinizing variables such as positional occupation, team affiliation, and the opponents faced. This scrutiny aims to capture the nuanced impacts these factors exert on performance. Multiple machine learning approaches, including regressions, clustering, and time series models, are deployed in this comprehensive process. To optimize the model's efficacy, Huddle carefully divides data into training and validation sets. Iterative testing is then conducted to ascertain the model's validity in measuring performance under various conditions. This rigorous testing and refinement process not only facilitates effective learning from the data but also enhances the model's accuracy and reliability.

CALCULATING ODDS

The collaboration between Huddle's quant and engineering teams plays a pivotal role in integrating the simulation algorithm into the real-time processing pipeline, resulting in dynamic player props markets. Leveraging their respective strengths, the quant team, specializing in statistical modeling, closely collaborates with the engineering team overseeing the real-time processing pipeline. This joint effort ensures the efficient incorporation of the simulation algorithm, enabling the swift and precise generation of player props markets. Huddle's commitment extends beyond player props, encompassing other market categories previously mentioned. By harmonizing quantitative insights with engineering capabilities, Huddle delivers a comprehensive suite of **real-time, data-driven markets** that dynamically capture the evolving dynamics of sports events across various categories.

Simulation model

At the core of Huddle's product lies a suite of **Monte Carlo-based models** meticulously developed for each sport under analysis, underscoring Huddle's dedication to precision and adaptability in sports forecasting. Monte Carlo simulations were selected for their inherent strengths, which encompass robustness in handling intricate, variable-rich environments and the capacity to furnish detailed probabilistic forecasts. These models distinguish themselves by excelling in simulating a broad spectrum of outcomes, incorporating randomness to facilitate a thorough assessment of potential scenarios and their respective probabilities.

By seamlessly integrating the Monte Carlo simulation outcomes of sports games with Huddle's player projections algorithm, a notable augmentation is achieved in Huddle's ability to forecast a diverse array of player-specific probabilities. For instance, in simulating an NBA game using player projections, we not only obtain aggregate statistics for the entire team but also intricate details for each individual player, encompassing metrics such as rebounds, two-pointers, and three-pointers. As a result, Huddle's models furnish a more comprehensive and detailed panorama of potential game scenarios, providing stakeholders with profound insights into each player's likelihood of attaining various performance milestones within the game's context.

Real Time Processing

Within the HuddleOS, multiple real-time processing pipelines form its foundation. Without delving too deeply into specifics, these pipelines are tasked with ingesting, transforming, and processing real-time updates from the pitch, encompassing play-by-play data with detailed player involvement information. The output from this initial pipeline is then utilized to fuel another real-time processing pipeline known as the odds feed pipeline. This secondary pipeline executes simulations using the aggregated play-by-play data, standard game parameters, and aforementioned player inputs to derive odds for various markets, including player props. Concurrently, the simulation results power the **Single Game Parlay (SGP)** product, which can be seamlessly combined with Huddle's complete market offering. This unified pipeline spans across all sports, facilitating the rapid integration of each new sport by Huddle's engineering and quant teams. This swift integration is crucial for meeting business KPIs.

In order to deliver the best odds possible, **Huddle's trading team fine-tunes player inputs**, monitors play-by-play data in real-time through various trading tools, and intervenes when necessary with ad-hoc suspensions and activations, odds adjustments, and potential re-settlement in case any of the data providers send faulty data.

CONCLUSION

In conclusion, this paper has provided an in-depth exploration of Huddle's comprehensive approach to sports forecasting and odds calculation. Key elements such as the unified market feed, player projections model, Monte Carlo-based simulations, and real-time processing pipelines were discussed, emphasizing Huddle's commitment to precision, adaptability, and real-time responsiveness.

Looking ahead, several focal points for future consideration and enhancement emerge:

1. Continuous Integration of New Sports: Huddle's agile approach to integrating new sports into its unified pipeline is critical for business KPIs. Future efforts should maintain and potentially expedite this integration process to keep pace with evolving market demands.

2. Enhancements in Player Projections: As sports dynamics evolve, ongoing refinement of player projection models, encompassing variables like player types, team dynamics, and opponent analysis, will be vital. Leveraging advanced machine learning approaches and expanding the scope of player-specific metrics could further enhance accuracy.

3. Optimizing Simulation Models: The integration of Monte Carlo simulations has proven valuable. Future considerations may involve fine-tuning these models and exploring additional simulation methodologies to ensure even more nuanced and accurate predictions.

4. Real-time Processing Innovations: Continuous exploration of advancements in real-time processing pipelines, especially in terms of speed, scalability, and efficiency, can contribute to more responsive and dynamic market offerings.

5. Market Expansion and Product Synergies: Building upon the success of the Single Game Parlay (SGP) product, exploring opportunities for further market expansion and synergies within the product offerings could open avenues for growth and diversification.

6. Advanced Trading Tools: Continued investment in tools that empower the trading team to fine-tune inputs and promptly respond to real-time data changes will be crucial. This includes exploring advancements in automation and machine learning for dynamic decision-making.

By focusing on these aspects, Huddle can fortify its position at the forefront of sports forecasting, delivering not only the best odds but also a cutting-edge and dynamic market experience for its customers.