

Analyzing Player Behavior in Video Games using Big Data

Kaung Htet Kyaw - 6729566, M.Sc. Computer Science, Assumption University

Research Paper

ABSTRACT

This paper expresses about the studies of player's behaviors in video games industry using Big Data analysis. It firstly introduces the identification and expression of issues that are emerged in video games by the player interactions and the background history of these issues. The related works or essential literature reviews associated with it are described and the suitable solutions that are associated to this are addressed. Then, analysis describes how the proposed solution can solve those difficulties in practice. The conclusion readdresses the cases of problems, summary of problem solutions and its advantages.

I. INTRODUCTION

A. Problem Statement

Nowadays, since the video game development is advancing with increasing speed, the enormous amount of data is produced by the mobile and video games and they are captured and analyzed by game developers to empower gaming analytics in order to enhance the games developed in future. [1]

Most data-driven developers in the gaming industry apply this data to find and fix bugs in software, increase revenue, and enhance the player experience. However, the sheer volume and complexity of data available can make gaming analytics difficult, particularly when trying to fully utilize data or enable strong gaming analytics use cases with outdated data analytics tools. [1]

There are six potential gaming analytics challenges. They are: [1]

1. **High Data Storage** – The quantity of data generated by a game increase exponentially with its popularity and user base, making it more difficult for game producers to collect, compile, store, manage, and analyze. [1]
2. **Cross-Platform Integration** – Game makers must gather, combine, and store gaming data from a variety of platforms, such as PCs, mobile devices (Android and iOS), and gaming consoles (eg., Microsoft Xbox, Sony

PlayStation 4/5, Nintendo Switch, etc.), in order to comply with cross-platform gaming analytics. [1]

3. **Complex Querying Needs** – The extreme diverse type of gaming data that game creators may gather presents a hurdle for gaming analytics: complex querying demands. This comprises unstructured data that deviates from any standard data model as well as organized data (stored in relational databases in tabular form). Developers may need to run SQL/relational queries on structured data, perform full-text search on unstructured data, or use machine learning (ML) methods to handle both types of data in order to serve a variety of game analytics applications. To enable gaming analytics use cases, game creators must move or replicate huge data across several tools because the majority of analytics technologies do not provide multi-model data access. This makes the game analytics software far more complicated and expensive. [1]
4. **Real-Time Analytics** – Real-time data and insights are necessary for many game analytics applications, including the following:
 - a. **Personalization:** Using real-time user interaction patterns to deliver tailored experiences, suggestions, in-game deals, or ads. [1]
 - b. **Hacker Detection** - Finding unusual occurrences or user behavior that can point to dishonest game manipulation is known as hacker detection. [1]
 - c. **Monetization:** Making use of the most recent data on user interaction and monetization to improve microtransaction offerings and pricing or increase subscription sales. [1]

A more sophisticated data infrastructure that can continually evaluate vast amounts of freshly created data is required to achieve real-time game analytics. For game producers, creating and maintaining a real-time analytics capability may be extremely difficult due to the additional expense and complexity involved. [1]

5. **Integrating External Data** – Some game creators might wish to combine qualitative data from sources outside the game with the data they gather from within the gaming environment. [1]
6. **Data Security and Privacy** – Sensitive user data, such as financial information like credit card numbers or payment accounts used for subscription or microtransaction purchases, as well as personally identifiable information like name and mailing address, are frequently gathered and stored by game developers. When it comes to using sensitive user data, this presents a problem for game creators. As part of their game analytics programs, developers wish to openly access and analyze this data; nevertheless, stricter access restrictions and other security measures are also required to comply with data security and privacy laws, prevent the exposure of sensitive data, preserve user confidence. [1]

B. Background

The game business is one that fosters creativity, strategic thinking, and fast problem-solving and it is changing quickly. To differentiate themselves from the competition, video game creation companies must establish a data-driven approach that identifies patterns and trends. [2]

Big data analytics is examining vast amounts of data gathered from video games to provide insights on player behavior in gaming business. This includes the quantity of individuals who play a game every day, the number of new users who purchases made in-app or in-game, the game's levels where players frequently become stranded in order to increase income and brand awareness. This data may then be applied as a useful tool to guide marketing, improving game design, and make smart business decisions. [2]

Big data analytics is generally used in five ways in the gaming sector:

1. Player Behavior Analysis

By analyzing player activity, including how much time people spend on playing each game, big data analytics can seek trends and patterns in their involvement with the game. This information is critical because data analysts may get valuable

insights on improving user engagement and designing games that are more sensitive to players' requirements. [2]

2. Game Customization

Predictive analytics is widely utilized in the video game business to find patterns based on historical and current player data. Game developers use these data to tailor games to individual player preferences, such as modifying difficulty levels. After all, if the player have played any game, he have probably seen that each one monitors the player's decisions and modifies the plot or difficulties to make them more accessible and customized. [2]

3. Targeted Marketing

User data analysis allows for better focused marketing initiatives. Big data analytics may be used to target certain categories of players with targeted in-app adverts and game recommendations, therefore improving the overall gaming experience and generating profitability for the company. [2]

4. Fraud Detection

Big data analytics can help prevent and identify fraud and infringement. A fair gaming environment is maintained by detecting odd activity patterns that may suggest attempts to hack. [2]

5. Server Optimization

Big data analytics can help assist improve server performance. This consists of understanding peak application hours, load distribution, and possible bottlenecks. Ensuring a seamless gaming experience increases player retention. [2]

II. LITERATURE REVIEW

A. Understanding Player Behavior Analytics in Gaming

The gaming industry has developed dramatically throughout the years. As the development progresses, competition has increased significantly. As a result, game developers are adopting complex and novel technology to increase player acquisition and retention while also standing out. One of the primary solutions that developers have implemented to better serve their gamers is behavior analytics. [3]

Developers are now leveraging their games as data collecting and analysis tools to better understand their player's interests. This has opened up new possibilities in gaming by introducing the concept of data science. [3]

As the name implies, player behavior analytics entails collecting, processing, and analyzing player data while they play the game. This information includes user preferences, conservations, and in-game behaviors, to name a few. Analyzing player activity data may provide useful insights on how to improve the gaming experience, increase player engagement, and optimize game design. [3]

If multiple players do the same action at a given level of the game, it may indicate the need to optimize that component. For instance, if players quit when they reach a certain level, it indicates the level may be "too difficult" to accomplish, prompting developers to create a simpler level. These findings assist developers in optimizing the game for a bigger audience. [3]

B. The Importance of Gaming Analytics

Gaming analytics gives developers the knowledge they need to fix problems and improve software performance, increase player acquisition and engagement, balance gameplay, minimize churn, and maximize revenue production. The following factors embraces the essence of it:

1. Player Acquisition and Onboarding

Gaming marketers may measure and monitor first user flow interactions to discover bottlenecks or drop-off spots in the player acquisition process and enhance the onboarding process for new players. [1]

2. Player Experience

Game developers may utilize user behavior and engagement data to better understand player behavior and identify which features drive the most user engagement and pleasure. From there, developers may take efforts to improve the game experience by making the proper game design decisions to please their customers. [1]

3. Player Retention

Gaming statistics can give important information about the obstacles or disappointments that players often endure before abandoning a game. Developers may use this information to start tracking churn rates, projecting player churn, and designing away experiences that have a negative influence on player retention. [1]

4. Monetization

Optimizing game monetization is one of the most important gaming analytics applications. Gaming firms may utilize data to analyze subscription sales or in-game microtransaction purchases, identify top-selling products or high-spenders, and tweak pricing and offers to enhance income. [1]

5. Game Balancing

Game developers utilize data to discover possible balance concerns in multiplayer games and apply changes to maintain their games enjoyable and fair for loyal players. [1]

6. Bug Detection

Game developers may use data to seek software defects or other unexpected behavior in their products and address them before they cause imbalanced gameplay or poor user experiences. [1]

C. Roles of Player Behavior Analytic in Gaming

The roles of player behavior analytics in gaming are as follows:

1. Enhanced Game Design

Player behavior study is critical for enhancing game design. Players offer vital comments on your game design, including parts that are difficult to navigate or don't work properly. As a developer, you may share this information with game designers to assist them improve their design. This led to the outcome of more enjoyable and interesting in gameplay. [3]

2. Tailored Gaming Experiences

Personalized experiences are essential in today's highly competitive gaming market. It's an important approach for gaining and maintaining players since it keeps them engaged in the game. [3]

During player behavior analysis, developers examine player preferences and behavior. This allows them to tailor the in-game material, incentives, and challenges to these users' desires. As a consequence, they tailor gaming experiences to specific gamers' demands. [3]

3. Predictive Analytics

Player behavior analysis can be useful in predictive analytics. Developers may predict future player demands and behavior by evaluating past data and use trends. For instance, game creators can utilize player behavior analytics to predict when a player is likely to abandon the game. This can let them assist by offering customized rewards to keep the gamer engaged. [3]

D. Types of Player Behavior Analytics

There are various kinds of behavior analytics, each aiming on specific player aspects. They are:

1. Player Segmentation

This type of player behavior analysis categorizes individuals with similar tastes and habits, allowing game creators to give gaming events and material tailored to certain player segments, resulting in increased user happiness and engagement. [3]

2. Churn Analysis

Churn analysis is a critical component of player behavior analysis. It entails evaluating gamers and identifying individuals who are likely to depart the game. To do this, developers examine players' in-game behavior to uncover probable leaving reasons. The churn study data allow developers to apply efforts to keep players. For instance, they can generate new material to keep users interested or offer personalized incentives to encourage gamers to continue playing the game. [3]

3. Monetization Analysis

Developers use player behavior research to learn how players spend money on in-game goods. They may study users' purchasing habits and optimize in-game purchases. This leads to greater gaming income. [3]

4. A/B Testing

A/B testing enables creators to compare two game versions to see which one promotes better player behavior. For instance, developers might test numerous in-game features or user interfaces to see what keeps players interested. This allows them to delete or optimize the least engaging aspects, resulting in maximum engagement. [3]

E. How Data is collected during Gaming Analytics Process

Big data analytics is employed in many aspects of the industry, including evaluating player activity, strengthening security operations and generating marketing strategies. Game creators may use this information to fine-tune their designs, and gaming firms are expanding as demand grows. [4]

1. Player Behavior Analysis

Big data is utilized to study player activity, which helps to determine player trends and preferences. Analyses look at topics like how long people play per game, trends in their style of play, and how various people's styles compare to one another. [4]

- Analysts record games involving numerous players. Game creators can enhance future game designs by monitoring players' strategies during current games. They also employ big data to examine massive quantities of interactions in order to identify wider trends among participants. [4]
- Players report problems they face while playing, and preserved messages from during games contain concerns about game design and tactics players employ to boost their performance. Technical difficulties are also recorded in log files. Analysts utilize these reports to investigate inefficiencies and flaws in games.
- Analysts examine at certain "challenges" that occur in games to see how quickly players can solve them. Gaining a better knowledge of the real difficulty level of these tasks allow game creators to improve future versions of games and make them more tailored for individual players. [4]

2. Game Customization

Analysts utilize big data to investigate player tendencies and unique playing styles. They can identify both general trends and unique subtleties by collecting enormous volumes of data from a variety of participants. As a result, game creators tailor games to specific user preferences and talents. [4]

- Aspects like as difficulty levels and the complexities of maneuvering games may be properly tuned. In this approach, game makers are producing games that are not just more difficult for players, but also more tailored to people's specific preferences. [4]
- UX/UI design is improved by analysis of user behavior and input on the overall player experience. [4]
- By acquiring a better grasp of what players desire from their gaming experience, game creators may make games more appealing to prospective consumers, thereby increasing player numbers. [4]

3. Targeted Marketing

Analyses help gaming firms design more focused marketing efforts. Companies may personalize adverts for gamers by studying their preferences and tailoring them to their specific requirements. Advertisements can display while playing the game, and they can be tailored to the smallest detail thanks to extensive analysis. [4]

AI also helps to refine marketing methods. Analysts utilize artificial intelligence algorithms to examine large amounts of data and tailor advertisements to specific players. This is part of a bigger trend that is helping to enhance marketing in a variety of industries, including e-commerce, healthcare, technology, finance, retail, and manufacturing. As a result, marketing strategies are more effective, and gaming firms are expanding at a quicker pace. [4]

F. Gaming Companies Utilizing Big Data Analytics

1. Niantic

Niantic, US software company, is best known for creating the augmented reality smartphone game Pokémon Go. It uses big

data analytics to understand player motions, enhance VR experiences, and optimize the game based on location data. [2]

2. Riot Games

Riot Games, a well-known American video game development company and esports tournament organizer, uses big data analytics for player behavior research, pattern identification, esports administration, and game continuous improvement. Riot game is well-known for creating the famous game League of Legends (LoL). [2]

3. Ubisoft

Ubisoft, a French video game publisher, uses big data analytics to evaluate player activity, and deliver more tailored in-game experiences. Some of their popular gaming series include Assassin's Creed and Far Cry. [2]

III. PROPOSED SOLUTION

The proposed solutions for addressing the challenges in gaming analytics are as follows:

1. **High Data Storage** – The scalable cloud-based storage solutions can be implemented to dynamically adjust the increasing sheer amount of data produced by popular games. Utilizing the cloud infrastructure enables for flexible storage expansion without significant upfront investments. [1] [5] [6]
2. **Cross-Platform Integration** – The unified data integration platform that is capable of aggregating data from various gaming platforms, including PCs, mobile devices, and consoles can be developed. This platform can standardize data formats and ensure seamless data flow into a centralized repository. [1] [5] [6]
3. **Complex Querying Needs** – Game creators should adopt multi-model databases and analytics tools that support various data types and querying capabilities, including SQL for structured data, full-text search for unstructured data and machine learning integrations for advanced analytics. [1] [5] [6]
4. **Real-Time Analytics** – Implement a real-time data processing pipeline that captures, processes, and analyzes

streaming data to support applications like personalization, hacker detection, and monetization strategies. [1] [5] [6]

5. **Integrating External Data** – Game developers should establish a data integration framework that incorporates external data sources, such as social media and market trends, enriching in-game data to provide comprehensive insights. [1] [5] [6]
6. **Data Security and Privacy** – Game developers and publishers should implement comprehensive data security measures, including encryption, access controls, and compliance with data protection regulations, to safeguard sensitive user information. [1] [5] [6]

IV. ARCHITECTURE DESIGN

Firstly, for the solution of high data storage, Data ingestion layer services like AWS Kinesis or Azure Event Hubs can be utilized to capture real-time gaming data. Then, the system can employ cloud storage solutions such as Amazon S3 or Azure Blob Storage for raw data, and integrate with data warehousing services like Amazon Redshift for structured data analysis. In addition, the processing layer services like Apache Spark should be used to incorporate distributed processing frameworks to handle large-scale data transformations and analysis. [6] [7]

Second, for cross-platform integration, data platform specific agents or software development kits (SDKs) can be deployed to collect and transmit to a central system. The integration layer services like middleware solutions and APIs can be applied to normalize data from different platforms, ensuring consistency and these data should be stored in the centralized data warehouses in order to allow unified analytics.

In order to design the complex queries, the developers can implement multi-model databases like Azure Cosmos DB or Amazon DynamoDB to support multiple data models. Along with it, the analytics engine platforms like Elasticsearch for full-text search and Apache Spark MLlib for machine learning should be integrated and the system should have developed an interface for allowing seamless execution of different query types, abstracting the underlying complexities from end-users. [6] [7]

In the implementation the solution of real-time analytics, the data stream ingestion tools like Apache Kafka should be utilized to ingest streaming data from game clients and employing stream processing frameworks can allow the developers to analyze data in motion. Moreover, real-time analytics dashboards like Kibana can be developed to visualize real-time metrics and insights. [6] [7]

For external data integration, developers can create external data connectors to retrieve data from external APIs and data sources. Then, these external data should be merged with internal datasets, ensuring alignment and consistency and use the enriched data to build more accurate predictive models and insights. [6] [7]

Lastly, in order to protect the privacy and data secured, the data encryption algorithms like AES-256 standards can be applied. After that, Role-based Access Control (RBAC) and Multi-Factor Authentication (MFA) can be utilized to control data access. Nonetheless, the game developers should incorporate the tools and processes to make sure the adherence to regulations like GDPR or CCPA, consisting of data anonymization and regular audits. [6] [7]

V. ANALYSIS

Addressing the difficulties of game analytics necessitates a diverse strategy that blends modern technology, smart architectural, designs and strong analytical frameworks.

Cloud-based architectures provide elasticity by allowing storage capacity to scale in response to data growth. This strategy reduces the expenses associated with overprovisioning while also ensuring excellent data availability and durability. Furthermore, incorporating distributed accessing frameworks allows for effective management of massive datasets, resulting in timely insights. [1] [5] [7]

A unified integration platform accelerates data collecting while ensuring consistency across several platforms. This unified method enables extensive analytics, giving a complete picture of player behavior and game performance across all platforms. However, data synchronization and latency concerns must be carefully considered in order to guarantee data integrity. [1] [5] [7]

The use of multi-model databases and integrated analytics engines overcomes the difficulty of various querying requirements by offering flexible data handling and processing capabilities. This method removes the need for several independent systems, simplifying the design and lowering maintenance costs. To maximize efficiency across several query types, rigorous schema design and indexing methodologies are required. [1] [5] [7]

Real-time analytics provide quick insights into player activity, allowing for dynamic customization, rapid hacker detection, and monetization modifications. While deploying such systems adds complexity and necessitates a robust infrastructure to manage high-velocity data streams, the benefits of improving player experience and operational responsiveness are significant. [1] [5] [7]

Integrating external data broadens the analytical perspective, resulting in better informed decision-making and individualization player experiences. However, issues include maintaining data quality, addressing data privacy concerns, and dealing with growing complexity in data processing pipelines. [1] [5] [7]

Prioritizing data security and privacy is critical for maintaining user confidence and complying with regulatory regulations. While installing strong security measures may add complexity and performance overhead, it is an essential investment to safeguard both users and the company from data breaches and their consequences. [1] [5] [7]

VI. CONCLUSION

The fast rise of the video game business has resulted in an explosion of user, data which presents both benefits and difficulties for game producers. This research investigated the use of Big Data to analyze player behavior and improve game design, engagement, and monetization tactics. Key difficulties, including large data storage requirements, cross-platform integration, complicated querying demands, real-time analytics, external data integration, and data security, were recognized and solved with appropriate solutions.

To address these issues, cloud-based architectures, multi-model databases, real-time data processing pipelines, and strong

security mechanisms have been proposed. Implementing these technologies enables game producers to easily handle large volumes of data, increase player engagement, detect fraudulent activity, and improve overall gaming experiences. The combination of machine learning and AI-driven analytics improves predictive modeling, customization, and adaptive gaming settings.

Despite these advances, issues persist, notably in maintaining data quality, reducing latency, assuring compliance with privacy legislation, and managing the complexity of multi-platform analytics. Future research should look into combining blockchain for safe transactions, federated learning for decentralized analytics, and AI-driven automation to improve game analytics even more.

In contrast, Big Data analytics has become a vital tool for current game development, allowing developers to build, immersive, engaging, and customized gaming experiences while improving commercial tactics. As technology advances, harnessing data-driven insights will be vital to remaining competitive in the gaming business.

VII. REFERENCES

- [D. Bunting, "6 Gaming Analytics Challenges and How to 1 Solve Them," ChaosSearch, 25 January 2024. [Online].
] Available: <https://www.chaossearch.io/blog/solving-gaming-analytics-challenges#:~:text=6%20Gaming%20Analytics%20Challenges%20and%20How%20to%20Solve,...%206%206.%20Data%20Security%20and%20Privacy%20>. [Accessed 27 February 2025].
- [B. B. D. Academy, "How is Big Data Analytics Used in 2 Gaming?," 30 December 2023. [Online]. Available:
] <https://bigblue.academy/en/big-data-analytics-in-gaming#:~:text=How%20Is%20Big%20Data%20Analytics%20Used%20in%20Gaming%3F,4%20Fraud%20detection%20...%205%20Server%20optimization%20>. [Accessed 28 February 2025].
- [iXie, "The Role of Player Behavior Analytics in Gaming," 3 iXie, 25 March 2024. [Online]. Available:
] https://www.ixiegaming.com/blog/role-of-player-behavior-analytics-in-gaming/?utm_source=chatgpt.com. [Accessed 28 February 2025].

[O. Morris, "How Big Data Analytics in Gaming Improves
4 Player Experience," DataFloq, 22 February 2024. [Online].
] Available: [https://datafloq.com/read/how-big-data-
analytics-in-gaming-improves-player-
experience/#:~:text=Big%20data%20analytics%20in%20th
e%20gaming%20industry%20involves,which%20people%
20play%2C%20purchasing%20behavior%2C%20and%20s
uccess%20rates..](https://datafloq.com/read/how-big-data-analytics-in-gaming-improves-player-experience/#:~:text=Big%20data%20analytics%20in%20the%20gaming%20industry%20involves,which%20people%20play%2C%20purchasing%20behavior%2C%20and%20success%20rates..) [Accessed 28 February 2025].

[M. R. T. Grady Andersen, "Software Architecture for Big
5 Data: Challenges and Best Practices," MoldStud, 10
] February 2024. [Online]. Available:
[https://moldstud.com/articles/p-software-architecture-for-
big-data-challenges-and-best-practices](https://moldstud.com/articles/p-software-architecture-for-big-data-challenges-and-best-practices). [Accessed 28
February 2025].

[J. Klostermann, "Big Data Solutions For The Gaming
6 Industry," Cloud Tweaks, 6 July 2016. [Online]. Available:
] [https://cloudtweaks.com/2016/07/big-data-solutions-
gaming-industry/](https://cloudtweaks.com/2016/07/big-data-solutions-gaming-industry/). [Accessed 28 February 2025].

[M. R. T. Grady Andersen, "The Impact of Big Data in
7 Technical Architecture Design," MoldStud, 24 January
] 2024. [Online]. Available: [https://moldstud.com/articles/p-
the-impact-of-big-data-in-technical-architecture-design](https://moldstud.com/articles/p-the-impact-of-big-data-in-technical-architecture-design).
[Accessed 28 February 2025].