An Overview of Big Data and NoSQL in the Video Game Industry

Cristiana COSTAN
Bucharest University of Economic Studies
Faculty of Cybernetics, Statistics and Economic Informatics
Bucharest, Romania
costancristiana21@stud.ase.ro

The connection between relational and non-relational databases and game development is covered in this article. The concept of big data and its implications in the gaming industry is also stated, the capability to store user data becoming essential for marketing and game improvements. The ideas in this article are reinforced by creating a web application to see how NoSQL databases can be implemented in game development and how they can help us extract relevant data about users. Overall, the findings of this paper are intended to guide people who want to be part of or are part of the gaming industry in the choices related to which database to utilize in future games and to provide insights on how they might obtain information about potential consumers to improve their products or services.

Keywords: *Big Data, NoSQL, Video Games, Gaming Industry*

1 Introduction

The video game market is shifting so quickly that in order to have an impact, game developers need to design products with a strong emphasis on user needs. Companies are investing more money than ever in creating games designed to achieve popularity and generate profits. As video games have become a popular form of entertainment and relaxation, more and more individuals are choosing to spend their leisure time playing them. Game producers can offer potential consumers the experiences they desire due to the evolution of data and its integration into various industries. A lot of games have to be able to store information on player accounts, characters, or even in-game items. Depending on the situation, relational databases or non-relational databases may be used to store such information. For instance, non-relational databases are preferred when the stored data changes frequently, as their scalability is crucial in managing multiple modifications made in short periods of time. In contrast, relational databases are preferred when the game information is static and does not change often. Additionally, in some cases, both types of databases can be utilized for the same product, allowing the game to include simultaneously static and dynamic components that can be stored in different formats, corresponding to each database used.

The volume of data generated by the large number of users who play video games is massive, therefore NoSQL is mostly utilized for storing such data because it is more scalable, faster, cheaper, and flexible, addressing challenges in real-time applications like online games that relational databases are unable to solve.

Big data is often used for analytics in the gaming industry. Analysis of player information assists in the development of marketing guidelines for the gaming industry and the improvement of games, whether they are still in testing or are already available on the market. Thus, gaming companies can effectively promote their game and retain as many of the users they attract.

2 Overview of Big Data

Big data refers to massive data sets with a large, complex, and varied structure [1]. These data are produced by online transactions, logs, posts, search queries,

mobile and computer applications. Their massive growth led to difficulties in storing, managing, analyzing, and visualizing data through standard database software tools.

There are three main characteristics that define big data: variety, velocity, and volume. Through variety, big data generally has three types, depending on the distinct sources it might come from structured, semi-structured, and unstructured. Structured data is defined as data that is already labeled and easily sorted, whereas unstructured data is unpredictable and challenging to analyze. Semi-structured data is in the middle, not conforming to specified fields but having tags to distinguish data items. The volume of big data is larger than terabytes and petabytes (10^{15} bytes), and its size still increases. Velocity, which is crucial for organizations when they acquire data, is the rate at which data may be created and processed.

According to Dr. Aditya Vailaya, an expert in computer science, machine learning, and statistical pattern recognition, large amounts of user information can be utilized to improve product recommendations or deliver targeted information [2]. Big data analytics has applications in merchandising, pricing, marketing, and advertising. Understanding users' interests based on previous purchases or preferences and understanding the characteristics of best-selling products are only two of the analytical uses of big data.

3 SQL vs NoSQL

Structured Query Language (SQL) is a non-procedural data access language developed in the 1970s at the IBM research laboratories [3]. Nowadays, most database management systems implement a dialect of SQL, setting SQL as the predominant language for enterprise databases in our surroundings.

The term "NoSQL" was first utilized in 1998 to refer to a relational database that did not employ SQL [4]. The main reason for the NoSQL movement was stated by Eric Evans, a blogger who contributed to popularizing the term. He declared that this alternative appeared due to the need to solve problems that relational databases could not. Nowadays, as NoSQL might include SQL, it is usually considered as "not only SQL" by developers.

NoSQL databases are preferred over SQL databases in certain situations when scalability, high availability, and low latency are needed [5]. The scale-out strategy on which NoSQL databases are often based enables a cheaper process of scaling to large data volumes and provides the ability to upgrade or change the structure of a database with no downtime [6]. NoSQL is mostly utilized to manage big data and real-time web applications. Four different forms of data storage are included in NoSQL databases: document, key-value, graph, and column. The performance of NoSQL databases varies according to their data model.

Relational databases use the ACID (Atomicity, Consistency, Isolation, Durability) model for transaction processing due to its high reliability and consistency [7]. Atomicity refers to transaction failure and commits. If it were for a transaction to fail, the database would be left in the state before the transaction started. Only when all the transactions are successful the data would be updated. Consistency represents the stable state in which the database needs to be both before and after a transaction occurs. Isolation refers to a transaction being able to start only when another transaction finishes. Durability means that a completed transaction has permanent effects in case of later system failures. Because of their different priorities, NoSQL databases use the BASE (Basic Availability, Soft state, Eventual consistency) model, derived from the CAP (Consistency, Availability, and Partition-tolerance) theorem. Maintaining a single copy of current data, having the data available for updates, and being able to

continue operating even in the presence of network partitions represent the elements of the CAP theorem. “ACID properties are to RDBMS as BASE is to NoSQL systems.” [8]. Basic availability refers to the extent to which a user can modify the data. During periods of disconnection or synchronization delays, the user can still make changes and his actions will synchronize with the database later, once connectivity is restored. Soft state represents changing the state without any input, which ensures consistency. Eventual consistency means that if an updated database element is not further updated for a significant period of time, all users will see the same updated item value.

Addressing relational database management systems (RDBMS) vs NoSQL in detail, both have their advantages and disadvantages. NoSQL is easily scalable, the database does not necessarily need administrators, is faster, cheaper, and more flexible, some of the suppliers can manage hardware failure, and is utilized for big data applications. RDBMS is easy to design, execute, maintain, and use, and information is kept in one place, being secured. When looking at drawbacks, each disadvantage of NoSQL can be solved by SQL and vice-versa. For example, SQL presents issues in high availability, which can be managed by using NoSQL.

Modern big data-driven applications require the performance and scalability offered by NoSQL. Cisco replaced Oracle RAC with Neo4J, while Netflix switched from Oracle RDBMS to Apache Cassandra. These changes improved latency, shortened query times, and decreased expenses. These are only a few of the benefits brought by the usage of NoSQL databases. NoSQL databases are considered a better fit for handling large datasets with a lot of data streaming quickly from various sources [9].

4 The Video Game Industry

Video gaming is considered to belong to the field of culture and involves recreational activities [10], allowing individuals to distance themselves from their everyday problems by redirecting their attention to new goals from within imaginary worlds.

“The development of this industry characterizes our generation: fast paced, technologically oriented, and targeted toward the young and young at heart.” [11]

The first computer game, called “Spacewar”, was developed more than 50 years ago by Steve Russell, a Massachusetts Institute of Technology (MIT) student, as a demonstration of expertise and not for profit. In a short period of time, this game gained popularity within the United States of America (USA) and entrepreneurs sought the opportunity to make a profit off of this, game development becoming a business. The starting point of this industry began from the enjoyment of Nolan Bushnell playing Spacewar during his time at university, which led to the development of “Atari” in 1972.

Nowadays, the video game industry is growing rapidly, with over 10 video game genres, including sandbox, real-time strategy (RTS), shooters, and multiplayer online battle arena (MOBA) [12]. The strong competition determines companies to constantly improve game graphics and features in order to enhance user experience. User feedback represents the key factor in game development. In 2018 the production costs of some video games ranged from $100 million to $500 million. These gigantic investments represent a risk taken by the companies, the profits depending solely on the users.

5 SQL and NoSQL in the Gaming Industry

Choosing RDBMS or NoSQL when creating a game is up to the developers nowadays. If the game is based on static elements one can utilize relational databases and whether a faster save of the data is needed NoSQL can be employed. The benefits of NoSQL can be

seen, for instance, in scoreboards and player statistics, which are dynamic components. Table 1 highlights the main advantages of utilizing NoSQL databases in the development of video games.

Table 1. Advantages of NoSQL in the Video Game Industry

Use Case	Advantages of Utilizing NoSQL
Storing User Profiles	NoSQL's flexibility and scalability enable the rapid creation and continuous updates of user profiles, allowing many users to manage their accounts efficiently.
Storing User Items	In video games where inventories contain items with attributes that change based on user modifications, NoSQL can efficiently manage a wide variety of items with diverse attributes, handling large amounts of data.
Storing Game States Which Adapt to Player Choices	NoSQL efficiently stores and manages dynamic data from player decisions that impact game progression and outcomes.
Storing Leaderboard	NoSQL is ideal for managing leaderboard data, enabling fast updates and optimized querying.
Storing Player Statistics	Player statistics are frequently generated during gameplay, and NoSQL is crucial for handling the volumes of data produced.

In 2018 Fortnite announced hitting a new peak of 3.4 million concurrent players and shared within their article the "challenges of rapidly scaling a game and its online services" [13]. Fortnite has a service named MCP which is comprised of 9 MongoDB shards, utilized for retrieving data regarding game profiles, statistics, items, matchmaking info, and more.

In the same year, Riot Games wrote in their blog about deploying League of Legends to more than 12 disparate game shards, divided by regions [14]. MySQL was employed as the database backend for storing player account information, Riot putting a high priority on ACID. Both Riot and their parent company, Tencent, widely utilize MySQL and thus have a significant number of engineers who are familiar with how it operates. Because of this and the type of data that needs to be stored, MySQL is employed to protect user accounts.

In 2022 Amazon Web Services (AWS) posted in their AWS for Games Blog the architecture employed in "New World", a massively multiplayer online role-playing game [15]. Every second, millions of states change, thus appearing their need of utilizing Amazon DynamoDB, a NoSQL database which provides high performance and reliability and is suitable for game data, which frequently lacks organization and coherence. Moreover, employing Amazon DynamoDB allowed developers to successfully expand the game from a singular test server during development to over 500 in production without making any changes to the code.

Blizzard Entertainment, a video game developer and publisher, has listed on its careers page many jobs, including in Engineering & Technology. For Overwatch, a game which in 2024 surpassed 100 million players [16], experience with SQL databases was a plus, whilst for World of Warcraft (WoW), another popular game released by Blizzard, knowledge of relational or non-relational databases represented a plus. It is possible to conclude from these optional requirements that WoW uses both SQL and NoSQL, while Overwatch uses SQL.

Thus, the decision between RDBMS and NoSQL depends on the needs of the game developers. Developers prefer RDBMS

when dealing with static data, such as game characters, fixed-attribute items, game rules, and dialogues that don't change based on player actions. In contrast, NoSQL is more popular when dealing with player statistics, multiplayer leaderboards, match scores, and game states that adapt to player choices.

6 Big Data's Impact on Gaming

The gaming industry typically utilizes big data for analytics. The term "big data analytics" refers to the processing and analysis of massive amounts of data in order to gain useful information [17]. With the aim to assist analysts in making data-informed decisions, big data analytics makes it possible to find trends, patterns, and correlations in large amounts of raw data.

Big data analytics offers those in the gaming sector the ability to analyze large data sets, consumer preferences, market trends, and other crucial business data [18]. This information aids gaming companies in improving player experiences and developing customized products and services. Furthermore, a new era of gaming has emerged as a result of the application and review of game analytics, in which data analysis is now crucial to the success of game development businesses rather than just advantageous.

Nowadays, various methods are employed in big data analytics to enhance the gaming experience and increase user engagement. Player behavior analysis, which examines in-game movements, purchase history, and playtime patterns can lead to improved engagement and higher customer loyalty. Predictive analytics utilize historical data to predict future trends and player actions, aiding in improving player satisfaction and reducing game dropouts. Personalization algorithms leverage machine learning to create game recommendations, in-game content customization, and adapt advertising messages based on player information. Real-time analytics enable immediate correction of any problems, improving the overall gaming experience by tracking and examining server performance, live player behavior, and in-game transactions. Heat maps and spatial data analysis focus on player activities from within the game environment in order to improve level design, balance game difficulty, and ensure an engaging player experience. A/B testing allows game developers to compare two versions of a game or a feature among various player segments to determine which performs better and is essential for making data-driven decisions about game design and updates, ensuring that modifications reflect user preferences. Finally, the economic analysis examines player spending patterns, the balance of the virtual economy, and pricing strategies to optimize in-game economies, ensuring they remain engaging, fair, and profitable.

In the rapidly evolving gaming industry, these analytical methods provide gaming companies valuable insights on player behavior and market trends, enabling them to provide their customers with quality products and services that suit their preferences.

Ubisoft, SEGA Europe, and Kolibri Games represent only some examples from the gaming industry in which big data analytics are employed to ensure player engagement through personalized content, targeted marketing strategies, and gameplay adjustments. SEGA, for instance, gathers more than 25,000 data events each second, such as player behavior and in-game interactions.

Another interesting topic is related to how big data analytics can be improved with the utilization of machine learning (ML). Ubisoft leverages machine learning, which evolved from years of data gathering, to analyze player behaviors and enhance game features [19]. Since the early 2010s, advancements in tools and methods have allowed them to move beyond basic analysis to applying machine learning for actionable

insights and innovation.

The involvement of machine learning in behavioral analysis is also suggested in an article that discusses the use of neural networks to observe patterns and behaviors of players in the game World of Warcraft (WoW), a popular Massively Multiplayer Online Role-Playing Game (MMORPG) [20]. The characteristics utilized were analyzed using big data and data visualization tools. This experiment was conducted with the aim of showing how player analysis in an online gaming environment works on a small scale when using neural networks. Additionally, other analysis methods from those previously mentioned in the paper have been used for various studies regarding the behavior of WoW players. These demonstrations show us that by using big data and neural networks, companies can streamline their analyses and improve their games more easily.

One type of artificial intelligence (AI) program that can detect and produce text, among other things, is called a large language model (LLM) [21]. Massive amounts of data are utilized to train LLMs, which are based on machine learning. Although there is still room for improvement in the understanding of data by the LLMs, accurate outputs can be achieved when a more detailed and precise user input request is made [22]. This study also demonstrates the possibilities offered by the use of neural networks in the gaming industry, this time analyzing data related to the behavior of players in a gaming community, the data being extracted from an online forum for online board games. It has been discovered that exploratory data analysis (EDA) can be used effectively in this context, being an analysis that provides important insights from data.

These findings underscore the importance of big data in the gaming industry and highlight how game developers can benefit greatly from the analysis that can be performed with this data. Additionally, the chapter illustrates how machine learning has improved this sector by making it easier to understand players through big data analysis.

7 Building a Game

For a deeper understanding of NoSQL in the gaming industry I created a browser game with states that adapt to player decisions, as illustrated in Figure 1. It contains choices from three distinct categories: health, sanity, and supplies. On every in-game day, each user answers three questions.

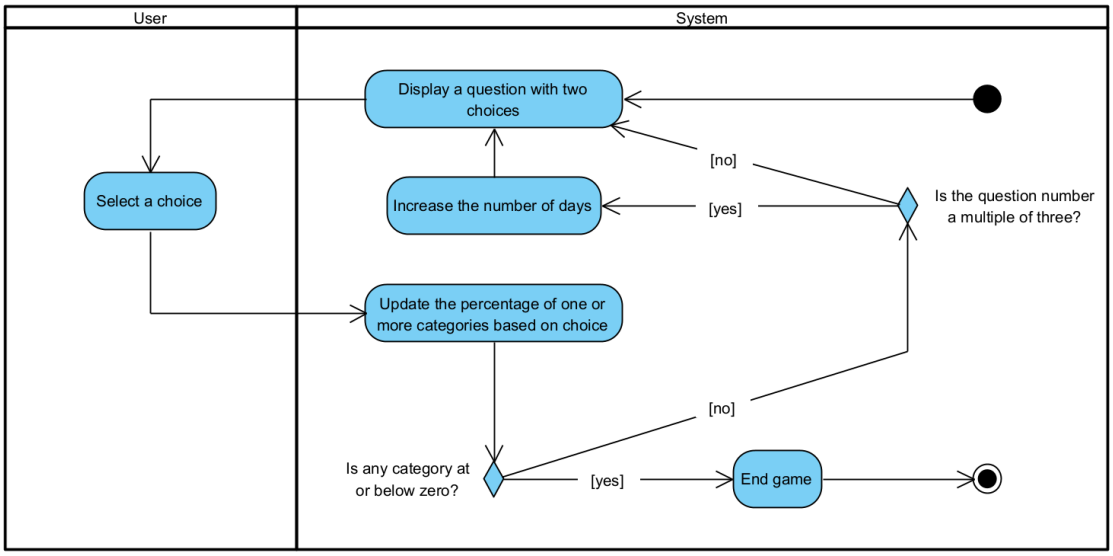

Fig. 1. Activity Diagram for Game Flow

Each category's percentage is affected by the choices made, and the game is over when one of the categories hits zero. A visual representation of the game's core features is provided in Figure 2.

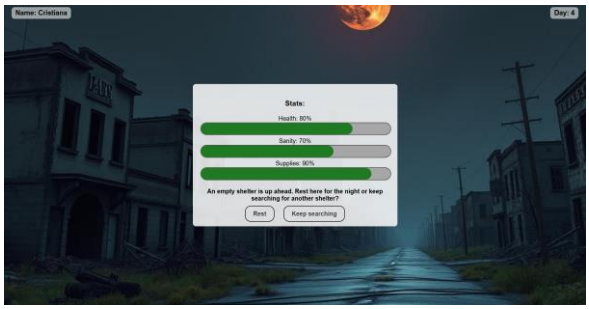

Fig. 2. Gameplay Overview

Since the game is still in its early stages, it merely keeps track of user data to generate the online leaderboard, which shows details about the players and how many days each of them survived. As depicted in Figure 3, every time the player completes the game, a POST request is made to the MongoDB database to insert the leaderboard data.

```
data = request.get_json()
days = int(data.get('days_survived'))
user = users_collection.find_one({"username": username})
high_score = int(user['high_score'])
if days < high_score:
    result = users_collection.update_one(
        {"username": username},
        {"$set": {
            "last_score": days
        }}
    )
else:
    result = users_collection.update_one(
        {"username": username},
        {"$set": {
            "last_score": days,
            "high_score": days
        }}
    )
```

Fig. 3. Python code for Updating User Score in MongoDB

The number of days survived is checked in the Python backend, entered into the database as the last score, or updated as both the last and high scores, and displayed on the final screen, as shown in Figure 4.

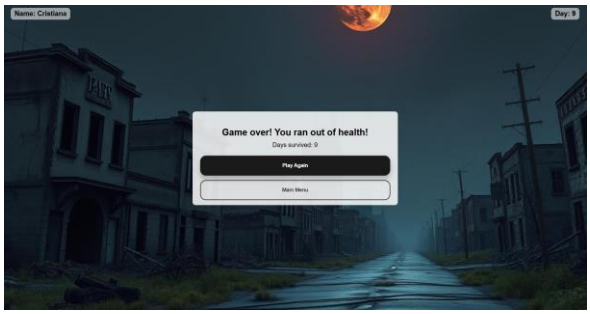

Fig. 4. Game Over Screen

All users with a score greater than zero days of surviving have their data extracted from the database and sorted in descending order, with a limit of ten users. This data is utilized to create the leaderboard. The code utilized for this process is displayed in Figure 5.

```
@app.route('/leaderboard')
def leaderboard():
    if 'username' not in session:
        return redirect(url_for('login'))
    username = session['username']
    leaderboard_info = users_collection.find({"high_score": {"$gt": 0}}).sort({"high_score": -1}).limit(10)
    return render_template('leaderboard.html', username=username, leaderboard_info=leaderboard_info)
```

Fig. 5. Python Code for Retrieving Leaderboard Data

If we were to think about the future of the game and the possibility of it becoming used by a multitude of people, the leaderboard data, as depicted in the leaderboard from Figure 6, would be extracted from a huge dataset, requiring NoSQL for its management. This will ensure continuous availability, low response times, and high-speed writes, ensuring that the database will intake data rapidly.

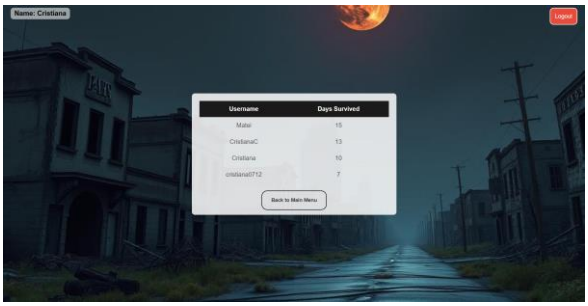

Fig. 6. Scoring Leaderboard

Taking all this into account, relational databases are recommended for developing a game that contains static features, including items with fixed basic attributes. When a game has dynamic elements, such as character traits, inventory items, states that change based on player choices, or player statistics, NoSQL databases are preferred over relational databases.

8 Conclusions and Future Work

The existence of big data has changed the gaming industry, and the emergence of NoSQL has made it easier to store this type

of data. Both relational and NoSQL databases can provide various benefits for game developers, and the choice of a database type is made depending on the needs of the respective game. NoSQL databases are preferred over relational databases in cases where games will have a huge number of players, each statistic of each player having to be stored in the database. Additionally, NoSQL makes it easier to store large amounts of data from which information can be collected, processed using pre-existing algorithms, and used to easily analyze player behavior.

In terms of data storage, relational databases and NoSQL databases are both significant for the gaming sector. While NoSQL databases are more appropriate for handling dynamic and massive volumes of player data, relational databases are better suited when utilizing static elements.

Regarding the future development of the application, potential analyses can be conducted to improve the game based on the data provided by the players, such as time played and maximum scores achieved. As a concrete example, if a player takes a long time to reach an average score, the game can be altered and made simpler, or difficulty levels can be added that users can choose at the beginning of the game.

References

[1] S. Sagiroglu and D. Sinanc, "Big data: A review," *2013 International Conference on Collaboration Technologies and Systems (CTS)*, San Diego, CA, USA, 2013, pp. 42-47.

[2] N. Salim, "What's All the Buzz Around "Big Data?": Meet Dr. Aditya Vailaya [The Good, the Bad, and the Ugly: Engineering FACTS]," *IEEE Women in Engineering Magazine*, vol. 6, no. 2, Dec 2012, pp. 24-31.

[3] D. C. Kreines, "Oracle SQL: the essential reference," *"O'Reilly Media, Inc."*, 2000.

[4] C. Strauch, U. L. Sites and W. Kriha, "NoSQL databases," *Lecture Notes, Stuttgart Media University*, vol. 20, no. 24, Feb 2011.

[5] B. Feltrin, "SQL versus NOSQL performance in Big Data analytics," *Master's thesis*, University of Zagreb, Faculty of Economics, Zagreb, 2022.

[6] MongoDB, "When to Use a NoSQL Database," mongodb.com. https://www.mongodb.com/resources/basics/databases/nosql-explained/when-to-use-nosql (accessed Dec 2024).

[7] G. C. Deepak, "A critical comparison of NOSQL databases in the context of ACID and BASE," *Culminating Projects in Information Assurance*, St Cloud State University, 2016.

[8] V. N. Gudivada, D. Rao and V. V. Raghavan, "NoSQL Systems for Big Data Management," *2014 IEEE World Congress on Services*, Anchorage, AK, USA, 2014, pp. 190-197.

[9] J. Bhogal and I. Choksi, "Handling Big Data Using NoSQL," *2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshops*, Gwangju, Korea (South), 2015, pp. 393-398.

[10] R. Baltezarevic, B. Baltezarevic and V. Baltezarevic, "The video gaming industry: From play to revenue," *International Review*, no. 3-4, 2018, pp. 71-76.

[11] P. Zackariasson and T. L. Wilson, "Paradigm shifts in the video game industry," *Competitiveness Review*, vol. 20, no. 2, Dec 2012, pp. 139-151.

[12] D. Pavlovic, "Video Game Genres: Everything You Need to Know," hp.com. https://www.hp.com/us-en/shop/tech-takes/video-game-genres (accessed Dec 2024).

[13] The Epic Team, "Postmortem of Service Outage at 3.4M CCU," fortnite.com. https://www.fortnite.com/news/postmortem-of-service-outage-at-3-4m-ccu (accessed Dec 2024).

[14] Riot Games, "Globalizing Player Accounts," technology.riotgames.com. https://technology.riotgames.com/news/

globalizing-player-accounts (accessed Dec 2024).
[15] AWS for Games Content Team and N. Walsh, "The Unique Architecture behind Amazon Games' Seamless MMO New World," aws.amazon.com. https://aws.amazon.com/blogs/gametech/the-unique-architecture-behind-amazon-games-seamless-mmo-new-world/ (accessed Dec 2024).
[16] Blizzard Entertainment, "Relive Your Overwatch Glory!," overwatch.blizzard.com. https://overwatch.blizzard.com/en-us/news/24104603/relive-your-overwatch-glory/ (accessed Dec 2024).
[17] T. Mucci and C. Stryker, "What is big data analytics?," ibm.com. https://www.ibm.com/think/topics/big-data-analytics (accessed Dec 2024).
[18] F. Hagverdiyev, "Data Governance in Gaming Industry," *Problems of Information Technology*, vol. 15, no. 1, 2024, pp. 44-51.
[19] C. Watters, "Shaping the Future of Games: How Machine Learning and Data Science Will Help Transform the Industry," news.ubisoft.com. https://news.ubisoft.com/en-us/article/5eZ3g9FHSlIkSIkp12lyVL/shaping-the-future-of-games-how-machine-learning-and-data-science-will-help-transform-the-industry (accessed Dec 2024).
[20] V. P. Barros and P. Notargiacomo, "Big data analytics in cloud gaming: Players' patterns recognition using artificial neural networks," *2016 IEEE International Conference on Big Data (Big Data)*, Washington, DC, USA, 2016, pp. 1680-1689.
[21] Cloudflare, "What is a large language model (LLM)?," cloudflare.com. https://www.cloudflare.com/learning/ai/what-is-large-language-model/ (accessed Dec 2024).
[22] V. Sharma and M. Dave, "SQL and NoSQL Databases," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 2, no. 8, August 2012.

Cristiana COSTAN graduated from the Faculty of Cybernetics, Statistics and Economic Informatics of the Academy of Economic Studies in 2024, obtaining a Bachelor's Degree in Economic Informatics. She is currently pursuing a master's degree in Databases - Business Support and will earn her degree in 2026. Her main areas of interest include data analysis, database management, and web development.