Logotipo

Descrição gerada automaticamente

UNIVERSITÁ DI PISA

ENGEGNERIA DELL’INFORMAZIONE

**PC Performance Evaluation on games**

**Project Report**

**Bruno Augusto Casu Pereira de Sousa**

**Master’s degree Computer Engineering – Computer Systems and Networks**

**883II 22/23 Large-Scale and Multi-Structured Databases**

Pisa, 2023

**Contents**

[1. Introduction 3](#_Toc144168205)

[2. Preliminary evaluation and datasets 4](#_Toc144168206)

[3. Design 6](#_Toc144168207)

[3.1 Actors and Requirements 7](#_Toc144168208)

[3.2 Use Case diagram 9](#_Toc144168209)

[3.3 Class Diagram 11](#_Toc144168210)

[4. Data model and DBMS architecture 11](#_Toc144168211)

[4.1 Document DB 11](#_Toc144168212)

[4.2 Key-Value DB 11](#_Toc144168213)

[4.3 Replicas and Sharding 11](#_Toc144168214)

[5. Implementation 11](#_Toc144168215)

# **1. Introduction**

The application proposed in this project is a platform for PC systems performance evaluation for games, allowing users to submit its personal computer system parts, and receive and evaluation if its system can run a specific game. In general, video games for the PC platform provide a set of components, such as the CPU and GPU models as reference for the minimum requirements to execute the game without performance problems. However, for a regular consumer, comparing PC parts without a specific metric is not an easy task, as the number of available CPUs and GPUs in the market is very large, and doing and objective evaluation if certain hardware is more powerful than other is not simple. Not only that, but the requirements vary a lot from one specific game to another, as different graphics engines and other features may influence the use of computational resources, making the comparison and evaluation of different system configurations not trivial. The objective of the service provided is then comparing benchmarks and performance test results for the components involved (User system and the game Requirements) to obtain an objective comparison between the expected performance, thus resulting in the evaluation if a users’ system can run the game based on the metrics evaluated.

In the case of not meeting the requirements, the developed application will suggest some hardware upgrades for the user, based on relevant metrics, including a price range and desired performance. As the components specifications and prices are stored in the platform database system, an additional functionality will be provided for the users, which will allow them to check benchmarks and compare the PC components performance and value. Not only the PC games system requirements will be managed, but to provide a better user experience, the presented platform will also provide some reviews for the games, in case a User just want to browser and check popular games and obtain information if its system can run those games. An additional feature will also allow a User to post a simple review of one or more games during its session in the platform, complementing exiting reviews and metrics that can be seen on the available games listed. This type of system evaluation system combined with the PC games overview is, therefore, useful as often when a new PC game is released many consumers asks questions like “Can I Run It?” or “What do I need to Run It?”. Finally, the manager will also be able to obtain statistics and additional information about which games the Users are searching, as a way to better identify trending games.

To handle the operations, a document, and a key-value database management systems (DBMS) were used, respectively, MongoDB and Redis, and the application modules were designed using Python, with the available libraries for those DBMS. With that the PC games and components information will be stored and managed in the document database (DB) and the user session and searches will be managed by the key-value DB, providing a fast and reliable platform.

# **2. Preliminary evaluation and datasets**

The first step to develop the proposed application was to obtain the datasets that are used on the platform, as well as to analyze its main characteristics in terms of data structures and organization. The evaluation is also fundamental to assess the feasibility of the proposed system as the database will be constructed based on the items, values and other information found in the researched datasets. As the primary source for PC games information, Steam is the largest gaming platform in the world, and it gather a huge collection of available games for purchase while also providing community support for discussions and games evaluations (<https://store.steampowered.com/about/>). Steam then contains most of the data necessary for the evaluation proposed, as its large game inventory along with the review data are made available.

For the platform proposed, three main sections are necessary to be implemented in the platform database system, so that it is possible to provide the proper information for the end users:

1. The PC game minimum requirements, defined by a CPU and GPU reference, along with addition descriptions (storage size or memory requirements). For example, the game “Forza Horizon 5”, found on Steam, has a system requirement of an Intel i5-4460 as the CPU, and the NVidia GTX 970 as the GPU.
2. A list of reviews for recent PC games, which must contain a brief description of the User experience with the game, and if that user recommends the game. For example, a review for the game "Dead by Daylight" was posted on the date 2017-11-22 recommending the game. This review also contained the text: “Fun to play with friends”, as a description.
3. The description and benchmark performance of the main PC components relevant for executing a software like a video game. This includes for which platform the component is designed (Desktop, Laptop, etc.), test results showing a metric for the component performance, and the retail price. For example, the Graphics card from NVidia model GeForce GTX 1650 for the Desktop platform has a price of 209 USD with a G3Dmark result (benchmark test) of 7807 points.

For the first two items on the list, the data can be extracted from the Steam platform, as for every game listed, the website provides the system requirements and the User reviews. To obtain then a compilation of the data from Steam some datasets were found in the Kaggle platform (<https://www.kaggle.com/>). A number of datasets were found, containing a significant amount of information. However, since the focus of this platform is not to manage and provide a complex platform for games reviews, a constrained and organized dataset was selected, containing over 260.000 reviews for a selection of games on Steam. In addition to this, a dataset was also found containing in a very organized way all the system requirements for the Steam games, where for every item, a CPU and GPU model were referenced, allowing the proposed evaluation platform to be developed.

Aside from the PC games general information and reviews, to obtain the information in the third item of the list, a consistent dataset must be used, and it should provide some known performance metrics for the PC parts, to compare them with the game system requirements. two additional datasets were found also using the Kaggle platform, and they contain lists for many PC parts (CPU and GPU) along with its retail price and benchmarks results. These datasets were fundamental for the project development as the proposed application must have a way to evaluate if a submitted hardware component can meet the minimum requirements to execute the game. The datasets then used for the development of the platform are listed below:

**PC video games requirements**

Source: [*https://www.kaggle.com/datasets/baraazaid/pc-video-game-requirements*](https://www.kaggle.com/datasets/baraazaid/pc-video-game-requirements)

CSV File size: 10MB

Number of documents: 80k

**Steam review dataset**

Source: [*https://www.kaggle.com/datasets/luthfim/steam-reviews-dataset*](https://www.kaggle.com/datasets/luthfim/steam-reviews-dataset)

CSV File size: 120MB

Number of documents: 260k

**CPU benchmark v4**

Source: *https://www.kaggle.com/datasets/alanjo/cpu-benchmarks*

CSV File size: 300kB

**GPU benchmark v7**

Source: <https://www.kaggle.com/datasets/alanjo/gpu-benchmarks>

CSV 120 kB

CPU+GPU Number of documents: 6k

To organize these datasets in the MongoDB collections, some modifications were done adapting and removing some data fields that were unnecessary. Also, the formatting of some of the fields had to be corrected, as it was making it difficult to quickly find components based on the system requirements, as on the dataset used, the name of the component suppressed the manufacturer name, which caused some discrepancy when querying the components dataset (as in that dataset, the name of the manufacturer was kept). To correct and improve the query description, a custom identifier (UUID) was included to better identify the components included in the DB. With all the needed modifications the final collections were then imported in MongoDB, converting the CSV files in the JSON document format. Figure 1 contains a diagram illustrating the process done for the dataset management:

Diagrama

Descrição gerada automaticamente

Figure 1 – Diagram of the management of datasets for the platform developed.

# **3. Design**

Staring the design of the platform, it is fundamental to understand the functional and non-functional requirements needed to provide the service. Also, the actors that are involved in the system must also be highlighted, with its respective functions inside the platform operation. Following this analysis, the use cases will be defined using the appropriated diagram, along with the Class diagram, providing an overview of all the modules that will be implemented to handle the system.

Another relevant mention to the overall design of this platform is the evaluation and characterization of the operations and access in the DBMS, considering the CAP theorem (Consistency, Availability, Partition tolerance). In the theory, DBs are then classified depending on the predominant types of operation, as well as the performance indicators that must be implemented in the system, following the functional and non-functional requirements. For the PC games evaluation platform, most of the access in the system are Read operation, as the application works more as a consulting system, where Users go to obtain needed information about their computer systems and its performance in video games.

With that, consistency is not the key performance parameter to be enforced in the platform, as eventually, the access to the requested data will be consistent, even though at some point a User may see a deprecated version of some game or review information due to the distributed DB. For most part, updates should not be recurrent in the collections, except for the comments and review collection, where then there is more dynamic and fast insertion of new data. It may occur that new comments were written in some portion of the database while a read request was being processed, and the replicas were not updated yet. In that case the Eventual Consistency paradigm would be used, and this issue should not detriment extensively the user experience as, eventually, the User will retrieve the most recent version of the PC game of interest with the proper reviews updated.

Analyzing the other two pillars of the CAP theorem with the platform proposed system, those performance indicators then have significant more relevancy, as low availability may in fact be detrimental to the user experience, as they will not be able to quickly get the information they were looking for. Also, partitioning the large number of reviews, in a way that data loss is minimized in case of failure is also a relevant parameter that must be present in the PC games evaluation platform. Therefore, the system proposed is more oriented to the AP side of the CAP theorem diagram.

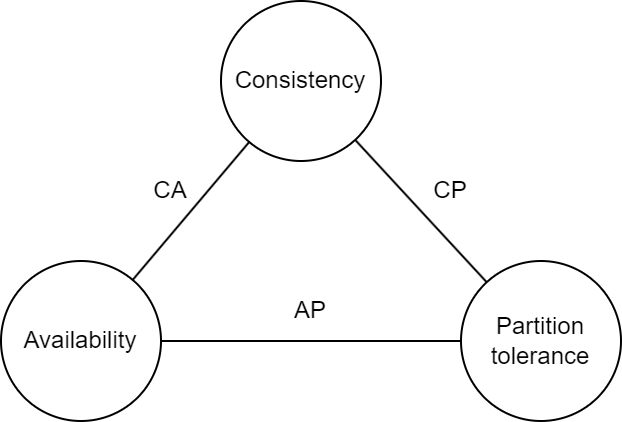


Figure 2 – CAP Theorem diagram.

## 3.1 Actors and Requirements

Moving to the requirement for the system, as previously presented, the overall objective of the platform is to provide information about PC games and its computational requirements, allowing users to check if its system is able to run a PC game. A fundamental consideration on this platform, is that as the service aims to provide evaluation and hardware information, meaning that there is not a significant importance in maintaining User personal information. In that way the design of the platform considers the User as an anonymous actor, and the information provided by this entity is only kept as a Session information (like tracking cookies in a browser search, for example). In that way, there is no registration in the platform, and any user can access the database resources, and submit its system for a quick evaluation. The same is true for the reviews management, as the goal of the platform is not to serve as a reference for games reviews, meaning that a simpler structure was built to integrate the reviews and games statistics on the DBMS. Thought they are not considered as the main objective of the platform, providing additional information about the PC games and trending games can be appreciated as most of the target users for the proposed system also have this as a common interest.

Two actors are then defined in the PC games and hardware evaluation platform, **Users** and **Managers**. The functional and non-functional requirements are listed below:

**Functional Requirements**

Users:

1. View most reviewed PC games.
2. View best reviewed PC games.
3. View latest reviews of a selected PC game.
4. View the System Requirements for a selected PC game.
5. View a selected Component benchmarks and price.
6. View the Components with the highest value metric.
7. Submit its own System Configuration.
8. Submit a simple review of a PC game.
9. Update its submitted System Configuration.
10. Obtain a System Evaluation: verify if its own System Configuration meets the System Requirements for a selected PC game, in terms of CPU and GPU.
11. Obtain a suggestion for a System Upgrade, based on a selected PC game.

Managers:

1. Create/update/delete a PC game information.
2. Create/update/delete a component information.
3. Delete a PC game review.
4. View statistics for Components, based on the categories.
5. View statistics for PC game reviews.
6. View all PC game reviews.

**Non-Functional Requirements**

1. The platform should provide a simplified way for users to verify if it submitted system is compatible with a particular PC game.
2. The access to the database should have low latency, improving user experience when evaluating its system.
3. The platform should be reliable, avoiding data loss in the case of hardware failure.
4. The access to the reviews and the evaluation should have good availability for improved user experience.

## 3.2 Use Case diagram

The Use Case diagram the illustrates the overall actions performed by the actors in the system. The diagram has been divided in four sections, to better illustrate the operation in each domain, and the correlation between the actions performed.

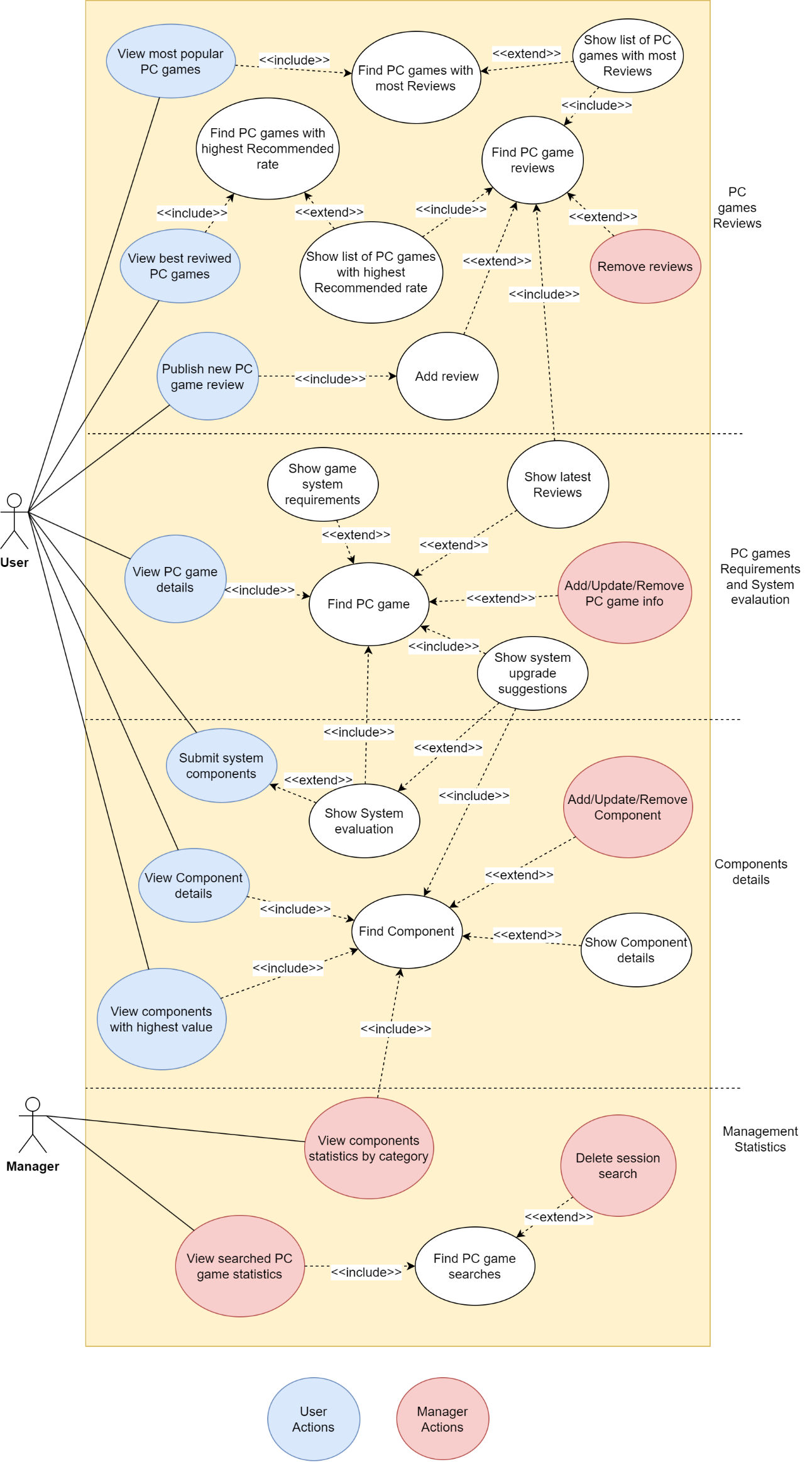


Figure 3 – Platform Use Case diagram.

## 3.3 Class Diagram

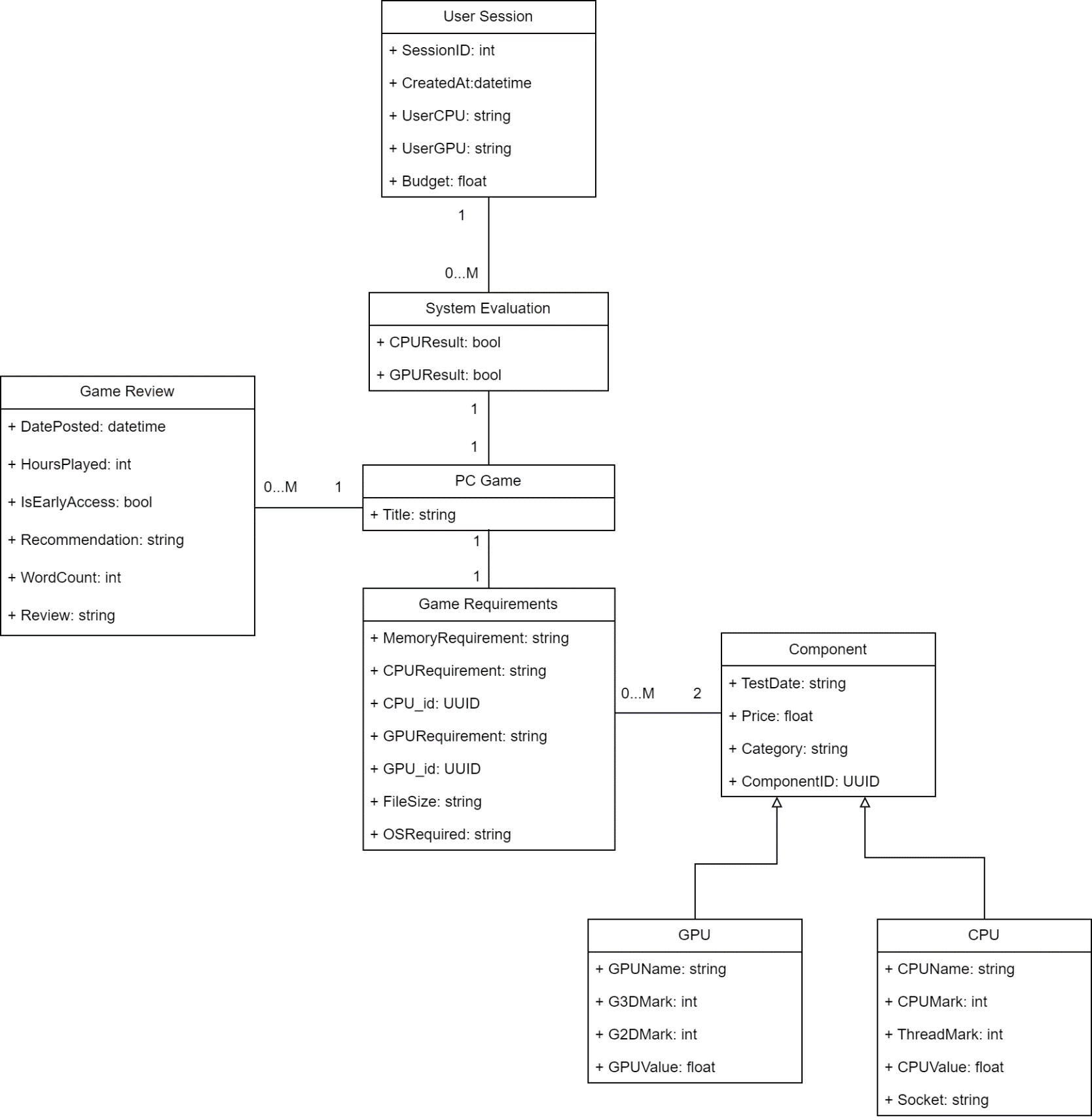


Figure 4 – Class diagram.

A brief description of the relation between entities is:

* A User (Session) have a One-to-Many relation with a system evaluation. Users obtain a system evaluation when submitting their information and searching for a PC game.
* A system evaluation has a One-to-One relation with a PC game. In this application Users then can search for multiple games in one session, thus having multiple evaluations, but each evaluation must be tied to one PC game.
* A PC game has a One-to-One relation with a Game requirement, and have a One-to-Many relation with a Game review, since one game may have multiple reviews.
* A Game requirement have a Many-to-Two relation with a Component, since for every game exactly two components (One CPU and One GPU) are present in its requirements, and those components may be included in multiple system requirements.
* A Component may be either a CPU or a GPU, having specific fields that differentiate between the two hardware.

When implementing the classes in Python, the structure of the classes was adapted to better organize the operations. A complete description of the classes and attributes created is shown:

|  |  |  |
| --- | --- | --- |
| System Requirements Class | | |
| **Attribute** | **Type** | **Description** |
| \_id | string | MongoDB id |
| title | string | Title of the PC game (unique) |
| memory | string | Memory requirement |
| file\_size | string | Required disk space |
| os\_required | string | Operating system required |
| cpu | string | CPU model required |
| gpu | string | GPU model required |
| cpu\_id | string | UUID of the CPU component |
| gpu\_id | string | UUID of the GPU component |

|  |  |  |
| --- | --- | --- |
| Game Review Class | | |
| **Attribute** | **Type** | **Description** |
| \_id | string | MongoDB id |
| title | string | Title of the PC game (unique) |
| is\_early\_access\_review | bool | Review posted in the early access period (true of false) |
| recommendation | string | If the user recommends the game or not |
| hour\_played | int | Number of hours that the user played the game |
| word\_count | int | Number of words in the review text |
| review | string | User review text |

|  |  |  |
| --- | --- | --- |
| Component Class | | |
| **Attribute** | **Type** | **Description** |
| \_id | string | MongoDB id |
| component\_id | string | Custom identifier for components |
| category | string | Component category (Desktop, Laptop, Mobile, Workstation or Server) |
| price | float | Retail price in USD ($) |
| test\_date | int | Year when the benchmark test was done |
| CPU Class | | |
| name | string | CPU model name |
| cpu\_mark | int | CPUMark benchmark score |
| thread\_mark | int | ThreadMark benchmark score |
| cpu\_value | float | Metric that consists in dividing the CPUMark score by the price (references how much performance per dollar the component has) |
| socket | string | Socket name for the CPU model |
| cores | int | Number of CPU cores |
| GPU Class | | |
| name | string | GPU model name |
| g3d\_mark | int | G3DMark benchmark score |
| g2d\_mark | int | G2DMark benchmark score |
| gpu\_value | float | Metric that consists in dividing the G3DMark score by the price |

|  |  |  |
| --- | --- | --- |
| User Session Class | | |
| **Attribute** | **Type** | **Description** |
| session\_id | string | MongoDB id |
| created\_at | datetime | Title of the PC game (unique) |
| user\_cpu | string | Submitted CPU model |
| user\_gpu | string | Submitted GPU model |
| budget | float | Maximum value the user is willing to spend on a system upgrade (in USD) |

|  |  |  |
| --- | --- | --- |
| Evaluation Class | | |
| **Attribute** | **Type** | **Description** |
| session\_id | string | MongoDB id |
| title | datetime | Title of the PC game (unique) |
| user\_cpu\_result | bool | Submitted CPU model |
| user\_gpu\_result | bool | Submitted GPU model |

# **4. Data model and DBMS architecture**

For the application proposed two types of DB systems will be used, a Document DB and a Key-Value DB. The document DB allows the information on the components, games, and reviews to be stored in a readable and consistent way, where the data structures are then stored in JSON format. The NoSQL structure that this type of DB manages data in a flexible and schema-less format, providing an intuitive and developer-friendly approach to data management. They are also optimized for high-performance read and write operations, thanks to their efficient schema design and support for distributed data storage.

Considering that the proposed platform should provide high availability, as discussed in the previous sessions mentioning the CAP theorem, a Key-Value DB was selected to handle the information of the users during a session. In this model, each piece of data (the value) is associated with a unique identifier (the key), optimizing data retrieval as values are directly accessed using their associated keys. Key-value databases are known for their simplicity and efficiency in handling basic data storage and retrieval tasks, which is the main motivation for selecting this type of data management system for the session information.

## 4.1 Document DB (MongoDB)

For the documents DB organization three collections were designed, based on the datasets found, they are:

1. ***components***

Interface gráfica do usuário, Texto

Descrição gerada automaticamente

With over 6.000 components imported from the dataset, the documents for the CPUs and GPUs have the following structure, respectively:

{

"\_id": {

"$oid": "64b32f37c7c0483e58ed4834"

},

"cpuName": "AMD EPYC 7763",

"price": 7299.99,

"cpuMark": 88338,

"cpuValue": 12.1,

"threadMark": 2635,

"cores": 64,

"testDate": 2021,

"socket": "SP3",

"category": "Server",

"componentID": "2f35cddf-10a7-48eb-afa5-d2716d4fa113"

}

{

"\_id": {

"$oid": "64b32f65c7c0483e58ed5745"

},

"gpuName": "GeForce GTX 1080 Ti",

"G3Dmark": 18284,

"G2Dmark": 934,

"price": 604.15,

"gpuValue": 30.26,

"testDate": 2017,

"category": "Desktop",

"componentID": "dc07dd3b-125a-4b19-9f7c-9417329d58e8"

}

1. ***game\_reviews***

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

The game reviews were extracted from the Steam database, and over 260.000 reviews were included, covering over 40 different games. The general document structure for this collection is as the example:

{

"\_id": {

"$oid": "64b481baa2900f566a40140a"

},

"date\_posted": {

"$date": "2017-09-12T00:00:00.000Z"

},

"hour\_played": 474,

"is\_early\_access\_review": false,

"recommendation": "Recommended",

"title": "Dead by Daylight",

"word\_count": 56,

"new\_review": "The is addictive and fun to play Licensed characters and constant updates make it even betterThe biggest downside is that it has become toxic because of school kids who are bullied at school and who take out their frustrations in the There are teams who deliberately demote their ranks to troll low rank killers out of "

}

1. ***system\_requirements***

Texto, Carta

Descrição gerada automaticamente

{

"\_id": {

"$oid": "64b475f4a2900f566a3ed8da"

},

"Memory": " 4 GB",

"GPU": "Intel HD 4000",

"CPU": "Intel Core2 Duo E8400 or Athlon 200GE",

"File Size": "23 GB",

"OS": "Windows 7 64-bit",

"title": "Valorant",

"CPU\_id": "f56d95f9-c561-4880-b414-394354cf6332",

"GPU\_id": "c70e31fd-687c-4901-9968-3b9d3c574e9a"

}

One consideration on the system requirement’s structure is that, by observing the read patterns to obtain the PC game information, it was noticed that embedding the latest reviews on the document for the requirements would greatly improve the speed of which the data is retrieve and displayed for the user, as accessing the reviews collection were consuming some time. In this way, the latest 20 reviews were embedded in the system\_requirement documents, allowing for a much faster data acquisition when a user searches for a PC game and would like to check the latest reviews. Though this process does speed the data display (making higher availability for the user), since the database will be distributed, it might occur that at some point the reviews shown to a user are not consistent, as the replicas are still being updated. However, this should not disrupt the user experience, as the priority of this application is to provide good availability, relying on eventual consistency. The document structure for the reviewed games should be then:

{

"\_id": {

"$oid": "64b475f4a2900f566a3ed8e7"

},

"Memory": " 8 GB",

"GPU": "GeForce GTX 460",

"CPU": "Intel Core i3-4170",

"File Size": "25 GB",

"OS": " 64-bit Operating Systems (Windows 7,Windows 8.1)",

"title": "Dead by Daylight",

"CPU\_id": "2deb053d-3aa7-4254-baf4-8e0bdb4365f2",

"GPU\_id": "3ef6ac00-98f1-48b5-b82b-a9cace4c60dc"

"reviews":

{

"\_id": {

"$oid": "64b481baa2900f566a401408"

},

"date\_posted": {

"$date": "2017-11-22T00:00:00.000Z"

},

"hour\_played": 654,

"is\_early\_access\_review": false,

"recommendation": "Recommended",

"title": "Dead by Daylight",

"word\_count": 5,

"new\_review": "Fun to play with friends"

},

{

"\_id": {

"$oid": "64b481baa2900f566a40140d"

},

"date\_posted": {

"$date": "2018-09-29T00:00:00.000Z"

},

"hour\_played": 12,

"is\_early\_access\_review": false,

"recommendation": "Recommended",

"title": "Dead by Daylight",

"word\_count": 1,

"new\_review": "good"

} ...

}

## 4.2 Key-Value DB (Redis)

## 4.3 Replicas and Sharding

# **5. Implementation**

To deploy the developed monitoring system in a production environment some additional features must be present, such to detect fault in the sensor and actuators, as well as to track the number of packs that are lost. A secure protocol can also be added to the frames sent in the WSN to prevent intrusions and attacks. With some improvements the monitoring system can be adapted and provide a better quality of service for applications similar to the use-case defined in the project.