

УДК 004.6:004.42

ИСПОЛЬЗОВАНИЕ BIG DATA В КОМПЬЮТЕРНЫХ ИГРАХ



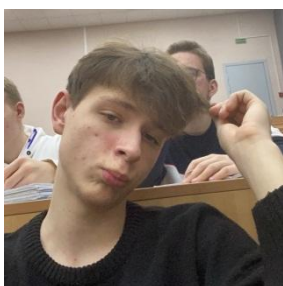
И.Д. Стаселько
Студент БГУИР



Т.Д. Позняков
Студент БГУИР



А.Ю. Сычев
Студент БГУИР



А.П. Протасов
Студент БГУИР



В.Я. Анисимов
Доцент кафедры физико-математических наук

Аннотация. Большие данные стали актуальным понятием, так как интернет сфера постоянно расширяется и прогрессирует. В данной работе анализируется использование больших данных в видеоиграх, технологии баз данных для решения задач игровой отрасли, и такие сервисы как Amazon Web Service. Нами рассматриваются примеры использования больших данных в видеоиграх и способы монетизации данных приложений. Проблемы разработки видеоигр касаются в основном игрового дизайна и количества различных характеристик

Ключевые слова: Большие данные, Supercell, Blizzard, технологии NoSQL.

Введение. Big data — это различные инструменты, подходы и методы обработки как структурированных, так и неструктурированных данных для того, чтобы их использовать для конкретных задач и целей. В современном мире Big data—социально-экономический феномен, который связан с тем, что появились новые технологические возможности для анализа огромного количества данных. Большие данные помогают разработчикам балансировать, тестировать и монетизировать свои игры. В будущем большие данные, скорее всего, будут иметь еще больше вариантов использования в видеоиграх и их разработке. Есть несколько примеров того, как игровые компании успешно внедряют решения больших данных с самого начала разработки своих продуктов, и есть некоторые компании, которые адаптировались к ним позже. Игровым компаниям не обязательно использовать большие данные, чтобы предоставлять свои услуги должным образом, в отличие от популярных социальных сетей, которые обрабатывают огромные объемы данных каждую секунду. Тем не менее, такие компании, как Supercell и Zynga, являются прекрасными примерами того, как большие данные могут помочь игровой компании добиться успеха. Петри Кяркес из Supercell принял участие в Aalto Big Data Breakfast в декабре 2015 года и в своей презентации он перечислил следующие варианты использования больших данных в своей компании:

1. Ежедневные показатели
2. Игровой баланс и показатели популяции игроков

3. А / В тестирование
4. Маркетинговая оптимизация
5. Обнаружение бота

Supercell - не единственная компания, использующая большие данные для решения тактических и стратегических задач. Рассмотрим некоторые варианты использования с примерами.

Уравнивание шансов. Балансировка является важной частью разработки как одиночной, так и многопользовательской игры, т.к. игра должна оставаться честной и захватывающей. Сложность однопользовательских игр должна увеличиваться постепенно, чтобы игроки не выходили из игры, потому что игра слишком легкая или слишком сложная. Многопользовательские игры требуют балансировки для асимметричных функций, таких как игровые персонажи с разными атрибутами и способностями. Большие данные можно использовать для балансировки как одиночной, так и многопользовательской игры, но в рамках данной работы мы рассмотрим многопользовательский аспект.

Балансировка персонажа. Blizzard Entertainment собрала более 35 миллионов игроков для игры в шутер от первого лица Overwatch [1]. Игра была запущена в мае 2016 года с 22 различными игровыми персонажами, известными как герои, было представлено четыре различных режима игры и девять разных карт. Существует четыре разные категории героев, и все герои имеют разные навыки, которые Blizzard необходимо сбалансировать [2]. Игры с таким количеством героев и способностей не могут быть полностью сбалансированы для всех режимов игры, карт и навыков игрока (от уровня начинающих до профессиональных игроков). Поэтому некоторые комбинации героев стали так называемым «мета», что означает, что эти герои и комбинации предпочтительнее других. В течение 2017 года игроки высказывали опасения, что некоторые, более мобильные герои стали слишком сильными, и так называемая «dive comp meta» (группировка некоторых персонажей с целью координации атак на противника) не позволяет полноценно играть другими героями, что делает игру предсказуемой и однозначной. Дайв Комп вкратце означает использование состав определенных мобильных героев для осуществления скоординированных атак противника. Директор игры Джефф Каплан ответил длинным постом, посвященным многим вопросам в то же время со следующими частями, решающими проблему погружения, используя статистику показать, что все хорошо. Во-первых, он упомянул шесть наиболее выбираемых героев из более привычного игрового режима Quick и только один из них, Генджи, является одним из героев «Dive Comp Meta». Во-вторых, он перечислил семь самых популярных персонажей из режима соревновательной игры, и в этом списке были как очень мобильные герои, ориентированные на «Дайв Комп», так и медленные герои, которые считаются избитыми героями «Дайв Комп» [3]. Он закончил статистику своего героя, заявив, что самая популярная героиня среди лучших игроков соревнований - Ана. Ана это герой поддержки, который не считается лучшим вариантом для погружений. Ответ Каплана был эффективным методом использования статистики, полученной с использованием инструментов сбора больших данных от Blizzard, чтобы доказать свою точку зрения и избежать поспешности только потому, что игроки чувствовали, что изменение было необходимо.

Балансировка игровой среды. В дополнение к балансировке героев, некоторые карты в играх также нуждаются в балансировке. Приведем соответствующий пример. В режиме игры под названием «Штурм» атакующая команда, состоящая из шести человек, имеет десять минут, чтобы захватить две команды защиты, также состоящих из шести человек, делающих все возможное для предотвращения захвата. На официальном игровом форуме игроки начали высказывать опасения по поводу режима игры, благоприятствующего защищающейся стороне. Когда игрок отправил вопрос о проценте побед на картах Assault в режиме быстрой игры, игровой директор Джефф Каплан ответил статистикой за период времени около двух месяцев.

Технологии баз данных для решения задач игровой отрасли. Реляционные базы данных широко применяются с момента их первого использования в 1970-х годах. Несмотря на взрывной рост баз пользователей онлайн-приложений и соответственно данных, генерируемых онлайн-системами и мобильными системами в течение этого времени, новые решения, которые лучше подходили бы для управления данными такого большого объема, не появлялись и не получали широкого распространения вплоть до недавнего времени. Вместо этого были внедрены различные методы увеличения емкости традиционных систем.

– *Шардинг*— методика разделения данных по разным серверам, которая требует знания местонахождения сервера с данными и ограничена отсутствием возможности объединения данных с разных серверов. Необходимо сопровождение схем для каждого сервера.

– *Денормализация*— еще один метод, включающий группирование и индексирование избыточных данных, зачастую приводящий к задержке и проблемам с поддержанием параллелизма в системах реляционных баз данных.

– *Распределенное кэширование*, которое предусматривает кэширование последней полученной информации в памяти, полезно, если может возникнуть потребность в этих данных. Приложение (Web-приложение, игра, социальная сеть, поисковая система и т. д.) сначала проверяет наличие необходимых данных в системе распределенного кэширования (например, memcached) вместо обращения к реляционной базе данных.

Новейшие технологии NoSQL позволяют обойти типичные недостатки реляционных баз данных и избежать вытекающей из этих недостатков необходимости применения методов расширения, описанных в предыдущем абзаце. Этим технологиям не требуются схемы и объединения, они не являются реляционными. Такие базы данных могут обрабатывать структурированные, полуструктурированные и неструктурированные данные. Дополнительные данные могут добавляться в хранилища в любое время независимо от формата и незамедлительно становятся доступными для анализа. Эти базы данных легко справляются с иерархическими вложенными структурами данных и благодаря своей гибкости поддерживают возможность автоматического распределения данных по серверам по мере увеличения и уменьшения объема информации без какого-либо влияния на производительность.

Существует множество типов баз данных NoSQL, включая документо-ориентированные хранилища, поколоночные хранилища, хранилища типа «ключ-значение», базы данных XML и базы данных на основе графов. На данный момент применяется более 150 баз данных NoSQL, и их число продолжает увеличиваться. Существует несколько популярных вариантов NoSQL с открытым исходным кодом, включая поколоночные хранилища (например, Apache CouchDB и MongoDB) и поколоночные хранилища с широкими строками (например, Apache Cassandra, Apache Hadoop и Hbase). Существуют также аналитические инструментальные средства с открытым исходным кодом, работающие поверх СУБД, например парадигма MapReduce в СУБД Hadoop. Механизм машинного обучения Apache Mahout обеспечивает классификацию, кластеризацию и совместную фильтрацию. Для исполнения статистической обработки с массовым параллелизмом непосредственно на узлах Hadoop была интегрирована система R с открытым исходным кодом. Также популярны коммерческие варианты NoSQL, которые интегрируются с Hadoop и другими инструментальными средствами с открытым исходным кодом и значительно расширяют возможности последних, добавляя средства аналитики, глубинного анализа текста, внутренней обработки приложений, функции MapReduce и возможности построения графиков.

Сервисы Amazon для обслуживания компании Supercell. Игры Supercell требуют тысячи серверов в любой момент времени. Для каждого из своих игроков Supercell нуждается в сеансе на стороне сервера—и количество серверов, которые требуются, растет с увеличением числа одновременных пользователей, играющих в игру. Вся игровая инфраструктура работает на Amazon Web Services (AWS). Supercell использует инстансы Amazon Elastic Compute Cloud (Amazon EC2), распределенные по нескольким зонам доступности, для повышения доступности и Amazon Simple Storage Service (Amazon S3) для хранения до 10 терабайт данных о событиях игр каждый день. Долгосрочное хранение, резервное копирование и архивирование осуществляется в недорогом сервисе Amazon Glacier storage. Эластичная балансировка нагрузки используется для распределения входящего веб-трафика и трафика событий между экземплярами. Supercell использует Amazon CloudFront, Amazon ElastiCache, Amazon API Gateway и AWS Lambda для доставки динамического веб-контента по всему миру, используя пограничные местоположения для улучшения задержки. Supercell использует Amazon Kinesis для доставки внутриигровых данных в режиме реального времени, обрабатывая 45 миллиардов событий в день. В то время как недавно он запустил собственные экземпляры Hadoop на Amazon EC2, теперь он использует Amazon Elastic MapReduce (Amazon EMR) для предоставления управляемой платформы Hadoop перед передачей данных в службу хранилища данных, работающую поверх экземпляров Amazon EC2. Данные Kinesis также считываются в Amazon DynamoDB. "Amazon DynamoDB хорош для таких случаев использования, как наш, где масштабируемость и скорость имеют решающее значение", - говорит Илихарью. Для мониторинга, управления и контроля своей среды Supercell использует ряд инструментов в консоли AWS. Amazon CloudWatch и AWS CloudTrail позволяют ему отслеживать текущую среду и предоставлять подробный журнал всех действий в среде разработки и производства. "AWS API—один из самых важных сервисов для нас, мы часто его используем", - говорит Илихарью. "Например, команды API позволяют нам легко управлять и загружать новые серверы и даже запускать среды полного нагрузочного тестирования для моделирования различных ситуаций и определения наших потребностей в емкости". Supercell также хорошо использует Amazon API Gateway и использует AWS Lambda для запуска кода в ответ на события и автоматического управления соответствующими вычислительными ресурсами. Илихарью говорит: "Мы используем AWS Lambda и API Gateway при создании API-сервисов, которые нуждаются в масштабируемости."

Тестирование. Существуют различные методы и инструменты для тестирования видеоигр. Как правило, тестировщики игр и агенты обеспечения качества (QA) постарайтесь убедиться, что [6]:

Игра сбалансирована;

Игра не падает ни в каких ситуациях;

Как старые, так и новые функции работают по назначению;

Большие данные предлагают помощь в тестировании игр, предоставляя возможность собирать большие выборки тестовых данных и анализировать данные в прямом эфире. "Наша главная бизнес-задача состоит в том, чтобы выяснить, что заставляет людей играть в наши игры, что делает их привлекательными. Сначала мы проявляем творческий подход, а затем используем полученные данные для подтверждения правильности принятых решений. У нас есть гипотеза, игра идет в прямом эфире, и мы проверяем нашу гипотезу. -Янне Пельтола, менеджеры данных компании Supercell на примере компании Hewlett Packard Enterprise о больших данных решения Supercell [7]. А/Б тестирование означает, что помимо живой версии игры существует еще и другая вариация игры. Эти изменения могут включать, например, изменения в пользовательском интерфейсе (Пользовательский Интерфейс) или способности в персонажах игрока. Цель состоит в том, чтобы собрать достаточно данных с обеих сторон. Актуальная версия, также известная как "контрольная" версия, и вариант версии, чтобы увидеть, если модификация приводит к желаемым результатам в варианте исполнения. "Один

интересный тест A/B Supercell провел о подключении Facebook. Тест был разработан, чтобы посмотреть, хотели ли люди, которым понравились игры, также звать их друзей, играть, и посмотреть, как эти переменные влияют на удержание игрока. С точки зрения A / B тестирования, теперь мы вытаскиваем данные в HPE Vertica. Что-то, что занимало два-три часа, в прошлом сейчас занимает четыре минуты." Blizzard имеет отдельную среду PTR (Public Test Realm), где все игроки могут протестировать следующий патч до фактического выпуска на живых серверах. PTR полезен для тестирования новых героев и карт, чтобы убедиться, что нет ничего слишком сильного в них, прежде чем они будут выпущены для миллионов игроков [8]. К нарушающим правила игры вещам относятся, например: слабые способности, слабые персонажи игроков и ошибки в картах или героях, которые ломают весь игровой сервер. Проблема с тестированием в PTR заключается в том, что оно не очень популярно по сравнению с их нынешними активными серверами и поэтому объемы собранных данных действительно небольшие по сравнению с их тестовыми серверами [9]. Согласно сообщению на форуме игрового директора - Джеффа Каплана в январе 2017 года только 0,26% игроков Overwatch вошли в PTR. В то время общее количество игроков для Overwatch было где-то между 20 и 25 миллионов так что количество игроков в области PTR было где-то между 52000 и 65000 [10]. Несмотря на то, что PTR не так популярен, как хотелось бы Blizzard, более 50000 игроков, предоставляющих данные в тестовой среде, для них все равно намного больше, чем количество сотрудников, из всей организации Activision Blizzard [11]. Также стоит отметить, что некоторые игроки не просто играют на PTR, но и сообщают о своих находках в официальный игровой форум или другие форумы Overwatch, такие как различные субреддиты Overwatch, сообщество Reddit [12].

Монетизация. Монетизация является самым важным фактором для игровой компании, потому что без хорошей системы монетизации на самом деле не имеет значения, является ли игра хитом или нет. В конце концов, деньги - главная цель игровой компании, быть прибыльной компанией без хорошо продуманной модели монетизации невозможно. Большие данные могут быть очень полезны для модели монетизации F2P, поскольку экономическая архитектура A-R-M-D, позволяет улучшать монетизацию игры постоянно, используя данные игрока в процессе. Премиум-модель не обязательно обладает этим преимуществом, потому что большинство разработок делается до начала монетизации.

Zynga была одной из первых компаний, внедривших разработку игр на основе данных в своих странах. F2P Facebook игры, такие как FarmVille и Mafia Wars. Член-учредитель Андрей Трейдер сказал в интервью The Wharton School: "грязный маленький секрет Zynga - из пяти корпоративных ценностей ни одна не является более важной, чем управление метриками. Для Zynga, это означало, что если вы не можете измерить что-то, не стройте его". Zynga внимательно следила за охватом, удержанием позиций и доходом при этом удержание является наиболее важным фактором. Их доход был получен за счет продажи внутриигровых предметов, которые хотели их клиенты. Как например, было упомянуто, что женщины покупали вещи в основном для внешности, в то время как мужчины покупали вещи, чтобы обыграть своих друзей в игре. Они использовали свои показатели, чтобы выяснить, что их игроки хотели [4]. Глядя на цифры игроков и продаж, Supercell является одним из лучших примеров использования модели F2P и больших данных для получения максимальной прибыли, не пугая неплательщиков. Supercell использует облачные сервисы Amazon, в частности Amazon MapReduce вместе с платформой Hadoop для сбора и анализа более 45 миллиардов игровых событий в день. В принципе, каждое действие, происходящее в играх Supercell, отправляется на их облако и анализируется с использованием большого набора данных, чтобы убедиться, что все находится под контролем [5].

Заключение. Индустрия видеоигр выросла больше, чем кто-либо мог себе представить в то время, когда впервые были опубликованы коммерческие видеоигры. Этого не произошло бы без помощи Интернета, хотя игры имели свою большую долю игроков еще до настоящего

Интернета и распространялись по всему миру. Видеоигры могут быть запущены на различных видах устройств, и подключение к Интернету воспринимаются как должное для многих из этих новых игр, которые появились в 21 веке. Объем данных в интернете вырос неожиданно быстро, и в конечном итоге традиционные реляционные базы данных не могли больше идти в ногу с объемом, скоростью и разнообразием данных. В то время как реляционные базы данных все еще служат своей цели для небольших томов, скорости и разнообразия данных, концепция больших данных была необходима, чтобы справиться с этой проблемой с базами данных NoSQL и фреймворками, такими как Hadoop и HPCSS. Большие данные принесли больше в таблицу разработчиков игр с точки зрения балансировки, монетизации и тестирования. Решения по балансировке и монетизации все еще могут быть приняты интуитивно, но всегда есть данные, доступные для изучения эффектов должным образом. Разработчики также могут объяснить свои действия или бездействие, показав сырую статистику своим игрокам. С надеждой, что игроки все равно будут услышаны, хотя разработчики могут просто посмотреть на данные и увидеть, если кажется, что что-то не так вместо того, чтобы читать мнения игроков на игровых форумах.

Список литературы

- [1]. Tweet by PlayOverwatch. [Online]. Available: <https://twitter.com/PlayOverwatch/status/919925924769906688> [Accessed: 3- December-2017]. Starodubtseva, M.N. Atomic force microscopic observation of peroxynitrite-induced erythrocyte cytoskeleton reorganization / M.N. Starodubtseva, T.G. Kuznetsova, S.A. Chizhik, N.I. Yegorenkov // Micron. – 2007. – Vol. 38, № 8. – P. 782-786.
- [2]. "Heroes – Overwatch". [Online]. Available: <https://playoverwatch.com/en-us/heroes/> [Accessed: 29-Nov-2017].
- [3]. "OW Update/Balancing cycle is excruciatingly slow: Part 2". [Online]. Available: <https://us.battle.net/forums/en/overwatch/topic/20757706588?page=2#post-27> [Accessed: 29-Nov-2017].
- [4]. K. Watanabe, T.Fukamachi, N.Ubayashi and Y.Kamei, Poster: Automated A/B Testing with Declarative Variability Expressions, IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW), IEEE 2017"Supercell Case Study". [Online]. Available: <https://aws.amazon.com/solutions/casestudies/supercell/> [Accessed: 29-Nov-2017].
- [5]. L. Levy and J.Novak, "Game Development – Game QA & Testing", 2009, pp. 58-69.
- [6]. "Supercell adopts HPE Vertica Analytics Platform". [Online]. Available: <https://www.vertica.com/wp-content/uploads/2017/06/Supercell-Success-Story.pdf> [Accessed: 29-Nov-2017].
- [7]. K. Watanabe, T.Fukamachi, N.Ubayashi and Y.Kamei, Poster: Automated A/B Testing with Declarative Variability Expressions, IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW), IEEE 2017.
- [8]. "Is the Sky Really Falling?". [Online]. Available: <https://us.battle.net/forums/en/overwatch/topic/20752500232?page=5#post-99> [Accessed: 29-Nov-2017].
- [9]. Tweet by PlayOverwatch. [Online]. <https://twitter.com/PlayOverwatch/status/824757676693270529> [Accessed: 29-Nov-2017].
- [10]. "Activision Blizzard on Forbes Global 2000". [Online]. Available: <https://www.forbes.com/companies/activision-blizzard/> [Accessed: 29-Nov-2017].
- [11]. "Undocumented PTR nerfs to Winston and Reinhardt". [Online]. Available: https://www.reddit.com/r/Competitiveoverwatch/comments/7e1lc4/undocumented_ptr_nerfs_to_winston_and_reinhardt/ [Accessed: 29-Nov-2017]
- [12]. "Clash of Clans" Maker Supercell Posts \$2.3B In Revenue, \$930M In Profit For 2015 As Growth Slows. [Online]. Available: <http://www.ibtimes.com/clash-clans-makersupercell-posts-23b-revenue-930m-profit-2015-growth-slows-2333237> [Accessed: 29- Nov-2017].

USING BIG DATA IN COMPUTER GAMES

I.S. Staselko
Student of BSUIR

T.D. Poznyakov
Student of BSUIR

A.P. Protasov
Student of BSUIR

V. Ya. Anissimov
*Assistant Professor, Candidate of Physico-
Mathematical Sciences*

A.Y. Sichev
Student of BSUIR

Abstract. Big data has become a relevant concept since the Internet sphere is constantly expanding and progressing. In this paper, we analyze the use of big data in video games, database technology for solving the problems of the gaming industry, and services such as Amazon Web Service. We consider examples of the use of big data in video games and ways to monetize application data. The problems of developing video games relate mainly to game design and the number of different characteristics.

Keywords: Big Data, Supercell, Blizzard, NoSQL technologies.