

УДК 65.011.56

**Тхай Куанг Хюи**

магистрант  
Санкт-Петербургский политехнический  
университет Петра Великого  
Санкт-Петербург, Россия  
huyken0015@gmail.com

**Thai Quang Huy**

Post-graduate student  
Peter the Great St. Petersburg  
Polytechnic University  
Saint Petersburg, Russia

---

## **ОБЗОР АНАЛИЗА BIG DATA В ОТРАСЛИ РОЗНИЧНОЙ ТОРГОВЛИ**

---

---

## **OVERVIEW OF BIG DATA ANALYSIS IN THE RETAIL INDUSTRY**

---

### **Аннотация**

В настоящее время отрасль розничной торговли стала неотъемлемой частью нашей повседневной жизни. Анализ Big Data в этой отрасли становится актуальным трендом, принося множество преимуществ для предприятий, способствуя улучшению производительности бизнеса, увеличению объема продаж и более глубокому пониманию поведения покупателей. Однако, он может столкнуться с различными трудностями, особенно когда инфраструктура, обеспечивающая его, работает неоптимально, что может привести к недоступности важной информации или ее задержке. Статья исследует обзор анализа Big Data и предлагает подходящую модель для анализа в розничной сфере, что способствует улучшению эффективности бизнеса и может быть использовано в качестве справочного материала для будущих исследований в области Big Data.

### **Ключевые слова:**

большие данные, розничная торговля, искусственный интеллект

### **Abstract**

Currently, the retail industry has become an integral part of our daily lives. Big data analysis in this industry is becoming a trend, bringing many benefits to businesses, improving business productivity, increasing sales and understanding of consumer behavior. However, it can face various challenges, especially when the infrastructure that provides it is not working optimally, which may result in unavailability or delay of important information. This article explores an overview of big data analysis and proposes a suitable model for analysis in the retail industry, contributing to the enhancement of business efficiency and serving as a reference for future research in the field of Big Data.

### **Keywords:**

big data, retail, artificial intelligence, Amazon, Spark

---

## **Введение**

Big Data (Большие данные) изменили способ, которым розничные компании принимают решения в бизнесе, требуя сочетания технологии и новой ценности. Термин "Big Data" впервые появился в 2008 году и был использован редактором журнала Nature Клиффордом Линчем [1]. С тех пор Big Data обещают иметь все более широкое применение в будущем. Говоря о Big Data, мы имеем в виду не только данные, но и эффективное использование аналитических технологий. Многие розничные компании проводят анализ Big Data в реальном времени и достигают ценных результатов, стимулируя прибыль и принимая более эффективные бизнес-решения.

Однако успешная реализация анализа больших данных сильно зависит от мощной и масштабируемой ИТ-инфраструктуры.

### **Понятие и характеристики Big Data**

Gartner дал следующее определение [2]: "Big Data – это информационный ресурс большого объема, высокой скорости и/или высокой многообразности, требующий творческого и эффективного подхода к обработке данных для улучшения понимания, принятия решений и автоматизации процессов". Проще говоря, это касается больших и сложных наборов данных, которые выходят за рамки возможностей традиционных методов и инструментов анализа данных. Big Data, как правило, может быть разделены на два основных типа:

1. Структурированные данные: представляют собой организованные информации, легко поддающиеся анализу и обработке, такие как данные о продажах, ценах, товарных запасах, клиентах, поставщиках и транзакциях.

2. Неструктурированные данные: Эти данные включают сложные информации, такие как комментарии в социальных сетях, изображения товаров, видеоролики с магазинов и другие формы данных. Большая часть Big Data существует в этой форме.

Исходя из определения Gartner, Big Data характеризуются тремя основными характеристиками, известными как "3V":

- 1) Объем (Volume);
- 2) Скорость (Velocity);
- 3) Разнообразие (Variety).

Кроме вышеперечисленных характеристик, современные аналитики выделяют две дополнительные важные характеристики:

- 4) Достоверность (Veracity);
- 5) Ценность (Value).

### **Преимущества анализа Big Data в розничной торговле**

Вот конкретные ценности, которые принести ритейлерам:

Во-первых, поддержка принятия более точных решений. Принятие решений руководителями в значительной степени зависит от доступной информации. Чем более точной и полной является предоставленная информация, тем более эффективным становится принятое решение. Проведение анализа данных, касающихся предпочтений, поведения при покупках, информации о конкурентах и состоянии рынка, способствует глубокому пониманию характеристик клиентов,

потребностей рынка, а также позволяет внести соответствующие коррективы с целью оптимизации процесса принятия решений.

Во-вторых, прогноз спроса. Анализ Big Data помогает прогнозировать спрос, что позволяет компании предсказать, какие товары будут пользоваться большим спросом, и в какой момент клиенты будут делать небольшие покупки, увеличивая возможности в согласовании и доставке товаров вовремя.

В-третьих, поддержка управления цепочкой поставок: Суть розничной торговли заключается в том, что клиенты заказывают малое количество товаров и хотят получить их как можно быстрее. Если это не удовлетворяется, существует риск потери выручки. Поэтому анализ Big Data помогает компании избегать излишнего запаса товаров на складе и снижать расходы на хранение благодаря разумному распределению товаров.

В-четвертых, мониторинг производительности магазинов в реальном времени. Крупные розничные компании обычно имеют множество магазинов, распределенных по разным географическим зонам. В условиях постоянных изменений на рынке, менеджерам необходимо быстро принимать решения на основе конкретных бизнес-показателей, таких как текущий объем продаж, количество продуктов каждого типа, доход от каждого магазина за определенный период времени и многие другие данные [3].

### **Кейс использования больших данных в розничной торговле**

Amazon, одна из ведущих мировых онлайн-розничных компаний. В России, если что-то покупают онлайн и это не отечественного производства, то, скорее всего, это товар с Amazon.

Amazon обычно собирает около одного эксабайта истории покупок от своих потребителей на своем веб-сайте. Они в основном получают данные из трех основных источников: данных, предоставленных пользователями в процессе взаимодействия с услугами; автоматически обнаруживаемых данных, включающих информацию о продуктах, местоположении и других параметрах; и данных, собираемых от третьих лиц, таких как кредитная информация, история покупок и возвратов. Основной целью этого процесса является анализ тенденций покупок, интересов потребителей и различных аспектов их поведения [5].

Вот несколько способов применения Big Data в бизнесе Amazon

Персонализированная система рекомендаций. Компания использовала инструмент под названием "Comprehensive Collaborative Filtering Engine" (CFE),

настраивающий персонализированные рекомендации. Он проводит анализ покупательской модели клиента, учитывая транзакции, списки желаемого, содержимое корзины и информацию о продуктах, которые часто искомых клиентами. Эти данные используются для предоставления клиентам рекомендаций при следующей покупке. Например, если клиент добавляет DVD в корзину, система может предложить фильмы, которые часто покупают другие клиенты. Таким образом, Amazon использует силу рекомендаций для стимулирования клиентов к дополнительным покупкам, превышающим их первоначальные намерения, с целью улучшения опыта покупок и увеличения объема трат. Этот инструмент приносит около 35% годового объема продаж компании [5].

Рекомендации от Kindle (Recommendations from Kindle Highlighting). После покупки Goodreads в 2013 году Amazon интегрировала социальную сеть в учетные записи почти 25 миллионов пользователей в функции Kindle. В результате читатели Kindle могут делать заметки, выделять ключевые слова и обмениваться ими с другими, используя их как средство обмена мнениями и обсуждения книг. Amazon регулярно рецензирует ключевые слова в заметках Kindle клиентов, для выявления областей, которые интересуют пользователей, и отправляет им рекомендации по другим электронным книгам [5].

Заказ одним кликом (One-Click Ordering). Поскольку Big Data показывает, что клиенты предпочтут совершать покупки в других местах, если товар не будет доставлен быстро, Amazon создала инструмент "One-Click Ordering". Этот инструмент получил авторское право и автоматически позволяет клиентам использовать его при размещении первого заказа, сохраняя данные о месте доставки и методе оплаты. Когда выбрано "One-Click Ordering", предоставляется временное окно в течение 30 минут, в течение которого субъект принимает решение о подтверждении покупки товара [5]. В случае отсутствия изменения решения в указанный период времени товар автоматически оплачивается и направляется по заранее предоставленному адресу.

Предпродажная модель доставки (Anticipatory Shipping Model). В рамках модели применяется Big Data для прогнозирования продуктов, которые клиент, вероятно, приобретет и потребует в определенное время и месте. На основе этих данных товары направляются в местные центры распределения или склады заранее, чтобы быть готовыми к отправке непосредственно после размещения заказа. Amazon использует

методы анализа прогнозов для увеличения объема продаж и прибыли, совмещая их с возможностью сокращения времени доставки и общих затрат.

Оптимизация цепочки поставок. С целью быстрого выполнения заказов Amazon установил партнерства с производителями и осуществляет мониторинг их складов. Конкретно Amazon применяет систему Big Data для выбора складов, прилегающих к продавцам или конечным клиентам, с целью сокращения расходов на доставку в пределах от 10% до 40%.

Оптимизация цен. Big Data используется для постоянного обеспечения привлекательных цен в Amazon с целью привлечения большего числа клиентов и годового увеличения средней прибыли на 25%. Установка цен осуществляется на основе множества факторов, включая действия клиентов на веб-сайте, стратегии конкурентов, наличие товаров, предпочтения клиентов, историю заказов, ожидаемую прибыль и другие параметры. Регулярное обновление и анализ данных Big Data позволяют Amazon корректировать цены на обычные товары каждые 10 минут [5]. Такой подход позволяет компании предоставлять скидки на востребованные товары и одновременно увеличивать прибыль от менее популярных продуктов. Например, стоимость романа в категории бестселлеров New York Times может быть на 25% ниже розничной цены, в то время как роман, не входящий в эту категорию, может стоить на 10% дороже, чем предложение конкурентов.

### **Применение искусственного интеллекта в розничных организациях**

Развитие искусственного интеллекта (ИИ) в розничной торговле стали важной тенденцией в последние годы, поскольку эти технологии все более активно применяются в ведущих розничных предприятиях. Предприятия, не внедряющие ИИ, будут вынуждены уменьшать прибыль для поддержания конкурентоспособности по цене, в то время как применяющие компании смогут наслаждаться значительными финансовыми выгодами. Основной причиной этого является высокая стоимость внедрения ИИ в бизнес.

В ритейле решения, основанные на использовании ИИ с Big Data, включают в себя:

1. Чат-боты. Чат-боты, базирующиеся на ИИ, не только предоставляют клиентскую поддержку, но и совершенствуют качество услуг, оптимизируя процессы поиска и предоставляя информацию о новинках. Более 80% мировых брендов

выражают намерение внедрить чат-боты в перспективе. Иллюстративные примеры успешного использования чат-ботов включают в себя Tommy Hilfiger и Burberry [6].

2. Корректировка цен. Решение эффективно осуществлено с использованием системы ИИ. Эта система оптимизирует стоимость своих товаров, анализируя различные стратегии ценообразования и используя обширные данные о продуктах, акциях и продажах.

3. Отслеживание удовлетворенности клиентов. Искусственный интеллект может обнаруживать настроение клиентов в процессе покупок. Walmart внедрил систему распознавания лиц для мониторинга этих настроений. Если клиент чувствует дискомфорт, сотрудники магазина могут вмешаться.

4. Создание разнообразного опыта покупок. Искусственный интеллект и большие данные предоставляют клиентам возможность тестировать продукты в различных точках контакта. Alibaba использует технологию "Fashion AI" для предоставления персонализированных рекомендаций из более чем 500 000 образцов модных данных. Alibaba также использует "волшебные зеркала", позволяющие клиентам примерять различные виды одежды и проводить оплату, сканируя QR-код на экране.

5. Развитие роботов. Роботы на основе ИИ стали символом инноваций в розничной торговле, принимая участие в процессах от поставок и цепочки поставок до операций в магазинах, маркетинге, продажах, а также в офисных и служебных функциях. Скорость и точность становятся ключевыми преимуществами, делая интеллектуальных роботов решающим фактором в центрах распределения. Большинство крупных розничных компаний по всему миру внедрили роботов в своих складах [6].

### **Технологии анализа Big Data**

Сегодня существует множество технологий Big Data, разрабатываемых исследователями по всему миру, включая Россию. Помимо крупных розничных компаний с профессиональными отделами, использующими технологии, такие как Hadoop, Apache Spark, малые и средние компании обычно используют инструменты, такие как Power BI, Tableau и Excel. В этой статье фокус будет сосредоточен на представлении технологии Apache Spark, которой широко пользуются многие физические лица и организации.

Apache Spark, технология быстрой обработки больших объемов данных, с мощной поддержкой для языков программирования Java, Scala и Python. Кроме того, она предоставляет разнообразные библиотеки для обработки данных в режиме реального времени (стриминг), машинного обучения и анализа изображений. Быстрая скорость обработки данных в Apache Spark достигается благодаря параллельным вычислениям на разных компьютерах, которые происходят полностью в оперативной памяти компьютера. Основным преимуществом Apache Spark является способность обрабатывать данные в реальном времени, получая данные из разных источников и обрабатывая их непосредственно по мере их поступления. Когда речь идет о технологии анализа Big Data, обычно Hadoop первым приходит на ум у программистов и ученых в области компьютерных наук. Тем не менее, Apache Spark, несмотря на свой более поздний появление, получает более высокие оценки и выгодно выделяется благодаря 100 раз более быстрой скорости обработки в памяти и 10 раз более быстрой на диске [4]. С такими впечатляющими преимуществами применение технологии Spark принесет значительную пользу предприятию.

### Модель обработки и анализа Big Data

На основе розничной формы и теории Big Data, модель обработки и анализа больших данных построена, описанную на рисунке 1.

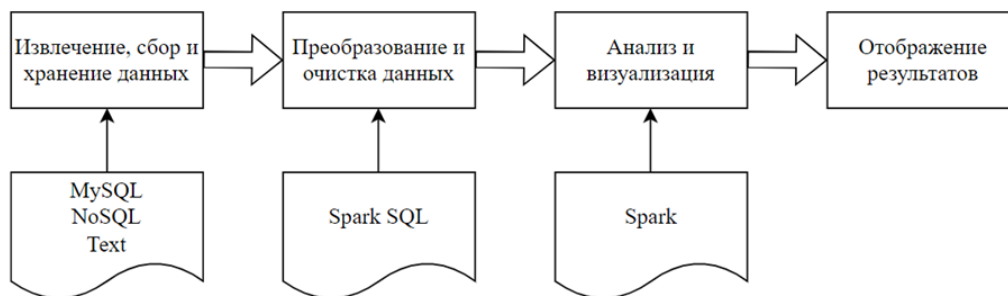


Рисунок 1 - Модель обработки и анализа больших данных на основе Apache Spark

Этап 1: Извлечение и сбор данных. В традиционных подходах данные обычно существуют в структурированной форме и хранятся в статических хранилищах данных для периодического запроса. Однако в случае больших данных исходные источники данных более гибкие, и данные часто представляют в неструктурированной форме. Ритейлеры имеют различные источники данных, и эти типы данных собираются и хранятся в форматах SQL, NoSQL или текстовых данных.

Этап 2: Преобразование и очистка данных. На этом этапе данные часто содержат избыточные компоненты или находятся в нестандартизированной форме, поэтому они должны быть преобразованы и очищены. Затем данные сохраняются в Spark SQL (компоненте Spark).

Этап 3: Анализ и визуализация. На этом этапе данные анализируются с помощью Spark. Результаты анализа могут быть непосредственно применены к распределенной системе управления данными Mllib (компоненту Spark). Это обычно будет базой данных SQL или NoSQL. Затем анализ исследуемых данных применяется к этой базе данных.

Этап 4: Отображение результатов. На этом этапе предприятие получает результаты обработки и использует их в различных задачах, таких как ценообразование на продукты, оптимизация запасов, разработка рекламных кампаний и принятие решений.

### **Вывод**

Анализ Big Data играет ключевую роль в розничной отрасли, обеспечивая более глубокое понимание клиентов с целью разработки персонализированных маркетинговых стратегий, прогнозирования потребительских запросов, оптимизации поставок и повышения эффективности бизнеса. Согласно Global Ecommerce, быстрое развитие технологий и электронной коммерции открыло глобальный рынок розничных продаж, увеличив долю онлайн-розницы до 18,1%, в то время как традиционная розница продолжает занимать 81,9% глобального рынка. Это создает большие возможности для розничных компаний внедрить технологию Big Data в свой бизнес. Для этого им нужно инвестировать в современную систему управления данными с мощной ИТ-инфраструктурой, чтобы заменить устаревшие технологии. Apache Spark считается важной технологией, о которой стоит узнать подробнее. Предложенная в статье модель может быть полезным ресурсом для предприятий в этой отрасли.

### **Список использованных источников**

1. Что такое Big Data и почему их называют «новой нефтью» [Электронный ресурс]. URL: <https://trends.rbc.ru/trends/innovation/5d6c020b9a7947a740fea65c> (дата обращения: 25.01.2023).
2. Big Data // Сайт Garner [Электронный ресурс]. URL: <https://www.gartner.com/en/information-technology/glossary/big-data>



3. Сулейкин А.С. Возможности применения технологии Big Data в крупном ритейле. European research. 2015. № 10 (11). С. 77-79.
4. Mudasir A.W., Suraiya J. Big Data: Issues, Challenges, and Techniques in Business Intelligence. Big Data Analytics. 2018. № 12 (105). С. 613-628. DOI: 10.1007/978-981-10-6620-7\_59.
5. How does Amazon use Big Data? 2023 Amazon Big Data insights [Электронный ресурс]. URL: <https://www.invisibly.com/learn-blog/how-amazon-uses-big-data/> (дата обращения: 20.12.2021).
6. AI in Retail: Uses of Artificial Intelligence in Retail Business [Электронный ресурс]. URL: <https://www.hyena.ai/what-artificial-intelligence-can-do-for-retail-industry/> (дата обращения: 19.10.2022).