



Рисунок 2 - Архитектура MERN приложения

Алгоритм выдачи рекомендации будет основан на коллаборативной фильтрации [4]. Он позволит прогнозировать предпочтения пользователя, сравнивая его интересы с интересами других пользователей. Благодаря этому в рекомендательном разделе предлагаются те элементы, которыми интересовались другие пользователи, а текущий нет.

Библиографический список

1. JustWatch - The Streaming Guide // JustWatch URL: <https://www.justwatch.com/> (дата обращения: 01.11.2022).
2. Кинориум // Кинориум URL: <https://ru.kinorium.com/> (дата обращения: 01.11.2022).
3. Все, что нужно знать о стеке MERN // machinelearningmastery.ru URL: <https://machinelearningmastery.ru/everything-you-need-to-know-about-the-mern-stack-43d27ddd480a/> (дата обращения: 02.11.2022)
4. Как работают рекомендательные системы. Лекция в Яндексе // Хабр URL: <https://habr.com/ru/company/yandex/blog/241455/> (дата обращения: 02.11.2022).

БОЛЬШИЕ МАССИВЫ ДАННЫХ В БАНКОВСКОЙ СФЕРЕ: ПРОБЛЕМЫ ХРАНЕНИЯ И ОБРАБОТКИ

Н.В. Павлович

Научный руководитель – Тонкович И.Н., канд. хим. наук, доцент

**Белорусский государственный университет
информатики и радиоэлектроники**

Банковская индустрия генерирует из внутренних и внешних источников большие массивы данных. Ценность этих данных в значительной степени зависит от того, как они обрабатываются и хранятся.

Проблемы, с которыми приходится сталкиваться при обработке и хранении больших массивов данных в этой сфере, могут повлечь за собой серьезные последствия для бизнеса. Речь, в первую очередь, идет о репутационном риске банка, под которым понимают риск, связанный с негативным восприятием банка со стороны клиентов, контрагентов, акционеров.

Установлено, что в процессе хранения данных, а точнее физической составляющей этого процесса, могут возникнуть следующие проблемы:

1 Существуют внешние факторы, влияющие на работоспособность хранилища данных, исходя из которых и возникает проблема надёжности и отказоустойчивости. Хранилище данных, как система, должно выполнять свои функции в полном объеме не взирая на факторы, оказываемые внешней средой. В хранилище должно быть предусмотрено полное или частичное резервирование всех компонент, а также система мониторинга и оповещения о возможных и существующих проблемах.

2 Проблемы с доступностью, как правило, возникают из-за каких-либо внутренних сбоев, однако хранилище должно обеспечивать доступность данных в любой момент времени. Такое состояние достигается путем зеркалирования сегментов, на которых хранятся пользовательские данные. В случае отказа одного из сегментов, должно происходить автоматическое переключение на резервный сегмент, зеркало упавшего сегмента.

3 В связи с постоянным накоплением данных объем хранимой информации непрерывно возрастает, что приводит к проблемам с производительностью. В зависимости от объемов и типов хранимой информации, необходимо поддерживать количество накопителей и процессорных подсистем на должном уровне.

4 Нарастание производительных мощностей влечет за собой проблему масштабируемости. С ростом объемов хранимой информации необходимо иметь возможность наращивать число жёстких дисков, объёма кэш-памяти, проводить аппаратную модернизацию и расширение функционала. Все перечисленное должно производиться без значительного переконфигурирования и потерь функциональности, что позволяет гибко подходить к проектированию инфраструктуры.

Следующий исследованный аспект – выявление проблем, связанных с процессом обработки данных.

Проблема проектирования потоков данных. Обработка данных начинается с проектирование логической модели. При загрузке данных из систем-источников в хранилище данные подвергаются ряду последовательных преобразований. Любую цепочку преобразований можно разбить на несколько этапов, результатом выполнения каждого из которых является обновление постоянной таблицы в хранилище. Каждому этапу соответствует слой хранилища данных, объединяющий в себе множество таблиц, обновляемых на соответствующем этапе преобразований. Обобщенная экосистема хранилища данных включает в себя: слой загрузки данных (физическое перемещение данных и систем-источников в экосистему хранилища); слой детальных данных (интеграция данных из нескольких систем-источников, приведение в соответствие с концептуальной моделью хранилища, построение версионности [1]); слой enterprise витрин (вычисление произвольных аналитических показателей, востребованных широким кругом пользователей) и слой справочников и правил преобразования (правила получения одних атрибутов на основе других, соответствие исходных ключей и ключей хранилища). Основной проблемой при построении модели является проектирование потоков данных. Необходимо придерживаться последовательности в цепочках

преобразований, выстраивая потоки данных таким образом, чтобы пройти каждый слой.

Проблема обновления хранилища. Обновление хранилища данных предполагает под собой выполнение огромного количества SQL-запросов. Запросы, возвращающие одинаковый набор данных, могут отличаться по времени работы, что говорит плохой оптимальности. Для установления причины долгой работы необходимо изучить план запроса, представляющий собой последовательность операций, которые выполняет БД для получения требуемого в запросе результата. После установления причины необходимо реализовать решение найденной проблемы. Практическое применение оптимизации позволяет в значительной степени повысить производительность распределенной базы данных, что приведет к снижению потребности в приобретении дополнительных производительных мощностей [2].

Проблема разрастания хранилища. Хранилище данных непрерывно накапливает данные, что приводит к увеличению занимаемого места. Для решения используются механизмы инкрементальной загрузки, чтобы уменьшить объем данных в потоке, что в свою очередь снижает потерю производительности, и механизмы регламентных работ, которые удаляют или переносят в другие системы хранения данные.

Библиографический список

1. Павлович, Н.В. Data vault: преимущества и недостатки / Павлович Н.В. // Новые информационные технологии в научных исследованиях: материалы XXVI Всероссийской научно-технической конференции студентов, молодых ученых и специалистов, Рязань, 2021 г. / Рязанский государственный радиотехнический университет имени В.Ф. Уткина. – Рязань, 2021. – С. 35–36.
2. Павлович, Н.В. Оптимизация запросов в Greenplum / Н.В. Павлович // Электронные системы и технологии [Электронный ресурс]: сборник материалов 58-й научной конференции аспирантов, магистрантов и студентов БГУИР, Минск, 18-22 апреля 2022 г. / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: Д.В. Лихаческий [и др.]. – Минск, 2022. – С. 53–56.

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ ПРИМЕНЕНИЯ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПОИСКА ДОСУГА ЧЕЛОВЕКУ

А.А. Петунин

Научный руководитель – Цуканова Н. И. доцент каф. ВПМ
**Рязанский государственный радиотехнический университет
 имени В.Ф. Уткина**

В настоящее время, в рамках актуального процесса цифровизации всех сфер деятельности в нашей стране, актуальными становятся исследования в области применения новых информационных технологий в различных отраслях нашей жизни.