

ПРОЕКТИРОВАНИЕ АНАЛИТИЧЕСКОГО СЕРВИСА С ИНТЕЛЛЕКТУАЛЬНОЙ ДОБЫЧЕЙ ДАННЫХ И ПРЕДОСТАВЛЕНИЕМ ПРОГНОЗОВ

Усиков А. В.

Кафедра программного обеспечения интеллектуальных и компьютерных систем, Учреждение образования «Гродненский государственный университет имени Янки Купалы»

Гродно, Республика Беларусь

E-mail: andrew.usikov@gmail.com

В статье изложены основные принципы построения аналитического сервиса, связанного с интеллектуальной добычей данных и предоставлением прогнозов. Выделены требования к сервису и указаны преимущества многоуровневой архитектуры сервиса. Кратко описаны составляющие компоненты сервиса, их взаимодействие и организация.

ВВЕДЕНИЕ

Современные условия ведения бизнеса, характеризующиеся возрастающей жесткой конкуренцией и нестабильностью экономических условий, предъявляют повышенные требования к оперативности и качеству принимаемых решений на всех уровнях управления предприятием или организацией. При этом объем информации, которую необходимо учитывать для формирования оптимальных обоснованных решений, неуклонно растет.

Это приводит к ситуации, когда становится невозможно эффективно управлять компанией без использования современных средств информационного обеспечения.

Одной из основных проблем в мире большого объема данных является избыточность ненужных данных для конкретного пользователя. В качестве решения такого рода проблемы можно использовать OLAP-средства.

С помощью интеллектуального анализа данных из данных, накопленных организацией, можно обнаружить ранее неизвестные, нетривиальные, практически полезные и доступные интерпретации знаний, которые необходимы для принятия решений организации.

I. КРАТКАЯ ХАРАКТЕРИСТИКА ВОЗМОЖНОСТЕЙ СЕРВИСА

Предлагаемый сервис будет обрабатывать с накопленные клиентские данные. Однако эти данные могут иметь различную структуру, поэтому само хранение данных тоже представляет собой достаточно сложную задачу.

Изначально для организации хранения данных большого объема необходимо хранилище, откуда и будет происходить выборка, а, в дальнейшем, анализ данных. Благодаря аналитической отчетности данные из хранилища представляются в виде, удобном для дальнейшего анализа. Наиболее удобным инструментом для получения аналитической отчетности являются OLAP-кубы. Технология OLAP дает возможность в реальном времени генерировать описательные и

сравнительные сводки данных и получать ответы на различные другие аналитические запросы [1].

Оперативные данные собираются из различных источников, очищаются, интегрируются и складываются в реляционное хранилище. При этом они уже доступны для анализа при помощи различных средств построения отчетов. Затем данные подготавливаются для OLAP-анализа. Они могут быть загружены в специальную базу данных OLAP или оставлены в реляционном хранилище.

Наличие самообучающегося модуля для прогнозирования и модуля интеллектуального анализа данных дает преимущество и интерес клиентов к данной функциональности. Обеспечение доступа к подобному сервису в любое время также является положительным аспектом. Каждый клиент может взаимодействовать с сервисом и, таким образом, следить за отображенной аналитической информацией.

II. ОСНОВНЫЕ ТРЕБОВАНИЯ К СЕРВИСУ

Основным требованием к функциональности разрабатываемого сервиса является наличие следующих возможностей для обработки больших объемов данных: аналитические сводки данных, интеллектуальный поиск данных (алгоритмы добычи данных), использование нейронных сетей для прогнозирования, принятия решений и выявления тенденций, корреляций, типовых образцов и исключений.

Данные как пользовательские, предназначенные для анализа, так и данные бизнес-логики, необходимо где-то сохранять. В силу этого следует выделить отдельный компонент хранилища данных, который будет отвечать за централизованное хранение данных, оперативный доступ, манипуляция над ними и обеспечение сохранности.

Сервис должен предоставлять также удаленный доступ к проанализированным и результирующим данным, поэтому необходимо предоставлять публичную конечную точку пользова-

телям для доступа в любое время к сервису и получать проанализированные данные.

Кроме этого, клиенту необходима среда для возможности загрузки снимков данных, их доработки, конфигурирования необходимых настроек и операций, слежения за статусом выполнения определенных операций, быстрого просмотра и проверки полученных данных для их дальнейшего использования [2].

III. ОПИСАНИЕ МНОГОУРОВНЕВОЙ АРХИТЕКТУРЫ СЕРВИСА

Многоуровневая архитектура сервиса обеспечивает группировку связанной функциональности приложения в разных слоях, выстраиваемых вертикально. Функциональность каждого слоя объединена общей ролью или ответственностью. Слои слабо связаны и между ними осуществляется явный обмен данными. Правильное разделение приложения на слои помогает поддерживать строгое разделение функциональности, что в свою очередь, обеспечивает гибкость, а также удобство и простоту обслуживания.

Функциональные области приложения разделяются на многослойные группы (уровни). Сервис состоит из шести взаимодействующих друг с другом слоев:

- уровень хранения данных;
- уровень надстроек;
- уровень доступа к данным;
- уровень бизнес-логики;
- уровень сервисов;
- уровень клиентов.

На уровне хранения данных расположен MS SQL Server 2012 и RavenDB Server. Данный слой ответственен за безопасное хранение информации, проведения манипуляций над ней и сохранение целостности данных.

Уровень надстроек включает в себя компоненты, которые содержат в себе дополнительную функциональность при работе с данными из хранилища. Надстройка над MS SQL Server в виде Microsoft Analysis Services включает в себя набор средств для работы с OLAP и интеллектуальным анализом данных. Надстройка Versioning Bundle над RavenDb будет создавать снимки для каждого документа на все изменения, вносимые в него, или когда он удаляется.

Уровень доступа к данным содержит в себе реализацию и контракты, позволяющие уровню бизнес-логики обращаться к данным не заботясь о специфике формирования запросов и различных ограничений, поэтому для потребителей этого слоя доступ к данным выглядит просто и прозрачно.

Уровень бизнес-логики содержит в себе набор компонентов, ответственных за отдельные

виды функциональности. При этом основная логика работы с сервисом, хранение настроек, манипуляции с ними, алгоритмы интеллектуального анализа и алгоритмы по созданию нейронных сетей.

Уровень сервисов имеет в себе реализацию всех сервисов, которые инкапсулируют в себе взаимодействия компонент с уровня бизнес-логики.

На уровне клиентов расположены приложения третьих сторон и приложение для управления настройками, просмотра данных и режимами работы анализа.

Каждый слой агрегирует ответственности и абстракции уровня, расположенного непосредственно под ним. При строгом разделении на слои компоненты одного слоя могут взаимодействовать только с компонентами того же слоя или компонентами слоя, расположенного прямо под данным слоем. Более свободное разделение на слои позволяет компонентам слоя взаимодействовать с компонентами того же и всех нижележащих слоев.

Благодаря такому типу архитектуры достигается слабое связывание между слоями, т.к. связи между ними основываются на абстракциях и событиях. Каждый слой инкапсулирует в себе все детали и типы данных, необходимые для функционирования слоя. Разделение функциональности между слоями достаточно четкая. Верхние слои, такие, как, например, слой клиентов, посылают команды нижним слоям, таким, как слой сервисов, и могут реагировать на события, возникающие в этом слое, обеспечивая возможность передачи данных между слоями вверх и вниз. Причиной этого является то, что каждый слой может обращаться только к нижележащему слою и к нему же может обратиться только вышележащий слой.

Выводы

Спроектированный сервис будет полезен всем организациям, которые заинтересованы в получении аналитических сводок и другой ранее неизвестной информации по своим накопленным данным. Например, организации могут проводить анализ воздействия рекламы, сегментацию клиентов, поиск признаков прибыльных клиентов, анализ предпочтений товаров, прогнозирование объемов продаж и многое другое.

1. Кудрявцев, Ю. А. OLAP технологии: обзор решаемых задач и исследований / Ю. А. Кудрявцев // Бизнес-информатика. – 2008. – № 1. – С. 66–70.
2. Усиков, А. В. Общие подходы создания и организации универсальной и гибкой архитектуры клиентской части веб-систем, специализирующихся на сборе и анализе различной информации / А. В. Усиков, В. А. Ломакин // Наука–2012: сб. науч. ст. В 2 ч. Ч. 2. Гродно: ГрГУ. – 2012. – С. 116–118.