

УДК 004.032.24

*Багдасарьян А.Д. магистрант кафедры «Сервис Транспортных Систем»,
Набережночелнинский институт ФГАОУ ВО «Казанский федеральный
университет», e-mail: arsenalkingt10@gmail.com,*

*Бахарева В.С. «Набережночелнинский институт (филиал) ФГАОУ ВО КФУ»,
Набережные Челны, e-mail: bahareva115@gmail.com*

СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ БИЗНЕС-РЕШЕНИЙ НА БАЗЕ ОБЛАЧНЫХ ТЕХНОЛОГИЙ

Аннотация: в статье анализируются технологии построения систем поддержки принятия решений, применяемые в управлении крупными экономическими системами, такими как предприятия, бизнес-проекты разного уровня реализации, государственные программы, социальные инициативы и проекты. Рассматриваются модели реализации распределенных аналитических вычислений в облачной инфраструктуре Amazon Web Services, использующие распределенную обработку данных Map Reduce, технологии хранения и обработки данных Hadoop, Mongo DB, Apache Kylin.

Ключевые слова: системы поддержки принятия решений, распределенные вычисления.

Введение

На сегодняшний день возрастающие требования к объемам обрабатываемых данных, оперативности анализа и минимизации капитальных затрат приводят к необходимости использования арендованных информационных вычислительных систем, одним из общепринятых наиболее эффективных вариантов использования вычислительных ресурсов является приобретение облачного сервиса.

Целью исследования является произвести анализ существующих сервисных моделей облачных технологии и выделить основные характеристики облачных технологий в рамках решения прикладной задачи разработки системы поддержки принятия решений. Проанализировать разработанные облачные системы поддержки принятия стратегических решений к переходу на облачные технологии.

Ни для кого не секрет, что современные IT-технологии имеют значительные конкурентные преимущества. Необходимо рассматривать их как динамичный и гибкий, но при этом эффективный для управления ресурсами. В связи с этим стремительное развитие IT-технологий возрастает. Переход к «облачным технологиям» является значимым моментом. В последние годы наряду с удобством и гибкостью облачных платформ использования информационных технологий существенно влияют на развитие данного направления, тем не менее возможности его развития еще не исчерпаны.

Облачные технологии

Облачные технологии – это то, что заставляет задуматься о будущем IT, способствующие воспринимать его систему деятельной. В данный период перевод IT-инфраструктуры любой компании в «облако», позволяет создать платформу для работы с «облачными технологиями».

Модель облачных технологий рассматривает возникновения ресурсов для повсеместного сетевого доступа по требованию к пулу разделяемых конфигурируемых вычислительных ресурсов и систем хранения данных, которые оперативно выделяются и освобождаются без необходимости непосредственного управления со стороны пользователя.

По способу владения облачные структуры делятся на типы: частные, публичные и гибридные.

Публичные облака доступны большой промышленной группе или широкой общественности; они принадлежат и поддерживаются предприятием, которое продает облачные IT-сервисы. Под термином «облако» обычно понимается именно открытое публичное облако. Сторонний поставщик при помощи Web-приложений динамически предоставляет ресурсы совместного использования через Интернет и выставляет счета от их использования.

Закрытые (частные) облака располагаются за сетевым экраном компании и ей же управляются. Это сервисы, которые создаются и управляются внутри предприятия. Ответственность за поддержку такого облака несет предприятие.

В случае необходимости компании могут разворачивать инфраструктуру на частном облаке. Такая модель используется организациями, нуждающимися в большей степени контроля над своими ресурсами и обеспечения безопасного использования своих данных. Такие облака создаются только для одной организации, обеспечивается контроль над данными, безопасность и качество предоставляемых услуг. На входе в облако устанавливается брандмауэр компании. Доступ к облаку и ресурсам предоставляется только сотрудникам и клиентам.

Гибридные облака – это совмещение закрытого и открытого облака, в котором используются сервисы, расположенные как в закрытом, так и в открытом пространстве. За управление такими сервисами ответственность распределяется между предприятием и провайдером открытого облака. При использовании гибридного облака предприятия определяют цели и требования к облачным сервисам, выбирая более подходящий вариант.

Говоря об архитектуре промышленных распределенных систем поддержки принятия решений, размещаемых на облачных серверах, то чаще всего она является трехзвенной.

Первое звено — это клиентское приложение, с помощью которого пользователь работает с системой. Как правило, в качестве такого приложения выступает обычный интернет-браузер или мобильное приложение, что не требует установки специального ресурсотребовательного ПО. Вторым звеном является так называемый «Front-end» сервер, занимающийся обработкой огромного потока заявок и ставящий их в очередь к третьему звену архитектуры — «Back-end» серверу. Он хранит данные промышленной системы и реализует ее логику.

Система поддержки принятия бизнес-решений на базе облачного хранения.

Проблематика системы поддержки принятия решений (СППР) является актуальной темой для исследований во всем мире, как в силу недостаточной разработанности, так и по причине перспектив, открывающихся при ее использовании. Новые технологии и модели бизнеса появляются постоянно, порождая все большее количество информации, что ставит проблему ее агрегации и анализа [1].

В настоящее время во многих организациях драйвером бизнес-процессов являются электронные таблицы, которые распространяются по электронной почте. Для успешной обработки больших объемов данных создаются сводные таблицы. Время на создание одного отчета в подобной вычислительной среде составляет несколько часов, и требует участия квалифицированного специалиста, или целого отдела. К тому же, подобная система исключает интерактивность, наличие которой является одним из предикатов системы бизнес-аналитики [2].

Несмотря на явные преимущества использования системы бизнес-аналитики, очень немногие Российские компании применяют ее. Причиной тому является, в том числе, отсутствие в организациях инфраструктуры, необходимой для развертывания систем обработки больших объемов данных, отвечающих требованиям отказоустойчивости и безопасности данных, а, также, отсутствие ресурсов для создания и поддержания подобной системы.

В подобной ситуации, закономерным шагом является использование готовой инфраструктуры, предоставляемой в виде сервиса, так называемые «облачные технологии» [3]. Эти системы подразделяются на основные компоненты, отвечающие за

- сбор исходных данных из различных источников,
- преобразование данных (приведение к стандартному виду, денормализация),

- группировку данных,
- инструменты многомерной аналитической обработки данных,
- создания отчетов и уведомлений.

Необходимо отметить, что графическая визуализация отчетов и интерфейс пользователя играют значительную роль в практике применения той или иной системы бизнес-аналитики, помогая лучше понять информацию, полученную из данных [4]. Ключевым компонентом таких систем является компонент, отвечающий за расчетную часть аналитики, и с которым связана система доступа и хранения собранных данных.

Как отмечают авторы [5, 6], применение принципа Map Reduce дает возможность обрабатывать неструктурированные данные объемом более одного терабайта, являющееся неоспоримым преимуществом по сравнению с реляционной системой управления базой данных (СУБД). Лидером рынка реализации возможности параллельного выполнения запроса на многих машинах является Apache Hadoop. Стоит отметить документо-ориентированную базу данных Mongo DB, реализующую Map Reduce. Учитывая особенности реализации, Mongo DB обеспечивает запись данных с большой скоростью, при этом не гарантирует консистентность записанных данных. Таким образом, Mongo DB больше подходит для хранения и обработки информации, потери которой в небольшом количестве не являются критичными (например, он-лайн опросы, данные сети датчиков), и не подходит для обработки финансовых данных.

Система управления и обработки данных Apache Hadoop не является БД в привычном понимании – это комплекс программных систем, состоящий из множества компонентов и позволяющий хранение данных в произвольном формате. Ее использование позволяет проводить распределенную обработку данных в сети, состоящей из множества вычислительных узлов. Стоит отметить, что настройка окружения Apache Hadoop может представлять определенную сложность. Для упрощения ее внедрения рекомендуется применять готовую к использованию среду,

предоставляемому как сервис. В частности компания Amazon предлагает продукт Elastic Map Reduce (EMR), настроенный как сервис Apache Hadoop. Недостатком такого решения является то, что доступ к инфраструктуре ограничен, невозможно обновлять версию программного обеспечения (ПО), в результате чего EMR может оказаться несовместим со всеми возможными расширениями, созданными для Apache Hadoop. Преимуществом является низкая цена.

Таким образом, самым подходящим для аналитики больших объемов данных является Apache Hadoop, остается ответить на вопрос подходит ли реализация Map Reduce для OLAP вычислений? В основе OLAP лежит принцип организации реляционных таблиц данных под названием схема «звезды», однако Map Reduce обрабатывает не реляционные данные и без помощи дополнительных программных компонентов реализовать OLAP невозможно. Существуют компоненты, части экосистемы Hadoop, специально разработанные для аналитических вычислений (рис.) – это Apache Hive – компонент для преобразования и агрегирования данных. Данные в Hive хранятся в виде таблиц, а для доступа к данным используется язык структурированных запросов Hive QL, т. е. эта СУБД с ограничениями на основе Hadoop. С помощью Hive можно построить и агрегировать данные в таблицы, которые используются для построения «OLAP-кубов», многомерных массивов данных, где факты распределены по нескольким «измерениям».

Недостатком Hive является низкая скорость выполнения запроса (несколько минут) по сравнению с традиционными СУБД из-за трансляции языка Hive QL в Map Reduce и последующего выполнения в среде распределенных вычислений. Достоинством является возможность обработки намного большего объема данных, чем СУБД.

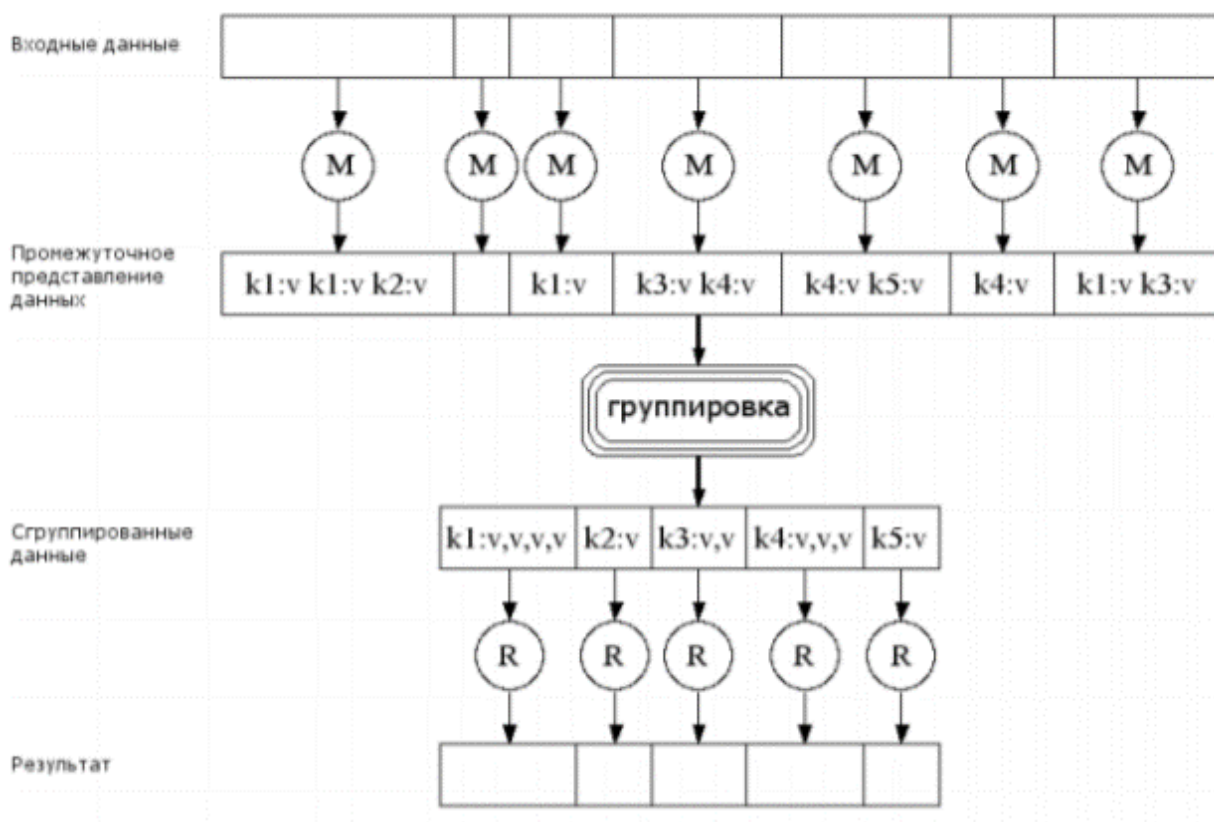


Рисунок 1 – Этапы выполнения вычислений в Hadoop Map Reduce

Проект Apache Kylin позволяет ускорить обработку запроса (секунды) и построить «OLAP-куб» по таблицам из HIVE на основе схемы «звезда», сохранить полученный массив в промежуточное хранилище, чтобы в дальнейшем использовать язык структурированных запросов ANSI SQL для выборки данных. Пользователю предоставляется удобный графический интерфейс для создания взаимосвязей таблиц и создания моделей «кубов». Также Kylin позволяет разграничить доступ к тому или иному массиву данных, создавать уведомления о событиях для пользователей. Недостатком является отсутствие сведений об интеграции Kylin с облачной инфраструктурой Amazon.

Осуществление процесса поддержки принятия решений вычислительно трудоемкая задача, для оперативного решения которой можно использовать модели облачных сервисов. В отличие от других классов интернет ресурсов облачные модели, вне зависимости от сервисной модели и модели

развертывания, должны удовлетворять основным положениям, прописанным в NIST Definition of Cloud Computing. Уровень обеспечения провайдером пользователей сервисов полностью регламентируется SLA соглашением и может быть пересмотрен по инициативе пользователя. Главные особенности облачных сервисов, положительно влияющих для реализации конечной цели поддержки принятия решения, являются: единство пула ресурсов, эластичность, масштабируемость и оплата только за фактически использованные ресурсы.

Выводы

Таким образом, задачу аналитики больших объемов данных нельзя назвать тривиальной, для создания мощного инструмента анализа необходимо совместить несколько систем, подняв уровень абстракции представления данных в несколько раз, что само по себе не может не представлять опасности получения некорректных результатов из-за ошибки программного кода того или иного компонента или неправильной его конфигурации.

Очевидно, что вероятность того, что ошибка не будет обнаружена долгое время, выше в случае использования частной инфраструктуры, нежели в случае использования облачной инфраструктуры. Использование «облачных технологий» для информационной СППР на основе обработки большого объема данных является наиболее целесообразным.

Список использованных источников

1. Тиханычев О. В. О некоторых проблемах предметной области поддержки принятия решений // Программные продукты и системы. 2016. № 3. С. 24–28.
2. Бизнес-аналитика и большие данные в России 2016 [Электронный ресурс]. URL: http://www.cnews.ru/reviews/bi_bigdata_2016. (дата обращения: 20.05.2021).

3. Батура Т. В., Мурзин Ф. А., Семич Д. Ф. Облачные технологии: основные модели, приложения, концепции и тенденции развития // Программные продукты и системы. 2014. № 3 (107). С. 64–72.
 4. Karacapilidis N. Mastering Data-Intensive Collaboration and Decision Making: Research and Practical Applications in the Dicode Project. Springer. 2014. 226 p.
 5. Дергачева А. А. Анализ данных на основе платформы SQL-Map Reduce // Научно-технический вестник информационных технологий, механики и оптики. 2014. № 1 (89). С. 66–71.
 6. Sun J., Wei X., Jian M., Jiasen S. Leverage RAF to find domain experts on research social network services: A big data analytics methodology with MapReduce framework // International Journal of Production Economics, 2015. N 165. pp. 185–193
-

Bagdasaryan A.D. master student of the department "Service of Transport Systems", Naberezhnye Chelny Institute, Kazan Federal University, e-mail: arsenalkingt10@gmail.com

Bakhareva V.S. "Naberezhnye Chelny Institute (branch) FSAEI HE KFU", Naberezhnye Chelny, e-mail: bahareva115@gmail.com

CLOUD-BASED BUSINESS DECISION SUPPORT SYSTEMS

Abstract: The article analyzes the technologies of building decision support systems used in the management of large economic systems, such as enterprises, business projects of different levels of implementation, government programs, social initiatives and projects. We consider models for implementing distributed analytical computing in the Amazon Web Services cloud infrastructure that use Map Reduce distributed data processing, Hadoop, Mongo DB, and Apache Kylin data storage and processing technologies.

Keywords: decision support systems, distributed computing.