

АРХИПОВА Е. С.
РОЛЬ ОБРАБОТКИ «БОЛЬШИХ ДАННЫХ»
В УПРАВЛЕНИИ СОВРЕМЕННЫМ ПРЕДПРИЯТИЕМ

Аннотация. В статье рассмотрены причины появления технологии «больших данных». Проанализированы преимущества применения данной технологии на предприятиях независимо от их отраслевой принадлежности. Представлены основные программные продукты, решающие проблемы обработки «больших данных».

Ключевые слова: большие данные, информационные технологии, управление предприятием, анализ данных.

ARKHIPOVA E. S.
THE ROLE OF BIG DATA PROCESSING
IN MODERN ENTERPRISE MANAGEMENT

Abstract. The article discusses the reasons for the emergence of Big Data technologies. The author analyzes the benefits of using Big Data for the company regardless of their industry sector. The main software products that solve the problems of Big Data processing are considered.

Keywords: big data, information technology, enterprise management, data analysis.

В условиях глобальной информатизации общества субъекты экономики создают и накапливают огромный объем данных различного рода. Производственные предприятия, финансовые рынки, банковский сектор непрерывно производят и обрабатывают потоки технической, конъюнктурной, статистической информации, способной оказывать влияние на принятие решений. Общеизвестно, что на сегодняшний день информация является активом любого хозяйствующего субъекта, от качества управления которым напрямую зависит эффективность его деятельности.

Благодаря развитию информационных и коммуникационных технологий растут возможности обработки больших объемов данных, что кардинально изменяет бизнес-процессы и бизнес-среду, например, взаимодействие с контрагентами и клиентами зачастую полностью переведено в дистанционную сферу. Однако, для субъекта экономики одним из важнейших условий является разработка программных и технических средств, способных не только с высокой скоростью обрабатывать постоянно увеличивающиеся информационные потоки, но делать более дешевыми процедуры их сбора, обработки, хранения и передачи. Бесспорно, технологии «Big Data» способны удовлетворить указанные требования современной экономики.

Итак, термин «большие данные» сам по себе вызывает множество споров. Для многих специалистов – это не более, чем тренд в информационных технологиях современности. И с этим трудно не согласиться, ведь первое упоминание о нем в журнале Nature датировано лишь 2008 годом. Для других – это полноценная, хотя и не теоретическая, а практическая категория, определяющая не только размер массива данных, который превосходит возможности обычных баз данных по занесению, хранению, управлению и анализу, но и неструктурированную информацию, перед обработкой и анализом которой бессильны традиционные алгоритмы [6]. Конкретно определение «больших данных» можно сформулировать как «технологии и архитектуры нового поколения для экономичного извлечения ценности из разноформатных данных большого объема путем их быстрого захвата, обработки и анализа» [9].

Для уточнения представленной формулировки понятие «большие данные» чаще всего характеризуют через систему свойств «пяти V»: Volume, Variety, Velocity, Veracity и Validity, т.е. объем, разнообразие, скорость, достоверность и обоснованность. Иногда специалисты выделяют шестое свойство – Value – ценность (с точки зрения важности экономической целесообразности обработки данных в соответствующих условиях) [8].

Свойство «Объем» подразумевает оперирование информацией в объеме не менее 100 Тбайт. Объем накапливаемой информации удваивается в среднем за 1,2 года [3]. По прогнозам к 2020 объем цифровой информации увеличится до 40 зеттабайт, что, в свою очередь, потребует приращения количества необходимых для ее хранения серверов в 10 раз [2].

Свойство «Разнообразие» означает, что данные аккумулируются из одного или нескольких источников и, возможно, в разных форматах. Источником Big Data являются не только корпоративные базы данных и массив социального интернета, но и весь поток информации от разнообразных датчиков, измерительных устройств, сенсорных сетей. Источники больших данных качественно дополняют традиционные источники, но ни в коем случае не заменяют их [4].

Свойство «Скорость» интерпретируется двояко. Во-первых, это скорость получения и обработки данных в режиме, близком к реальному времени, недоступном для традиционных подходов к анализу информации, а, во-вторых, это скорость накопления данных, подлежащих анализу.

Свойство «Достоверность» устанавливает порог «чистоты» данных, используемых для анализа, т.е. скорость изменения данных и их разнообразие не гарантируют их абсолютную истинность для конечного потребителя. В зависимости от сферы применения, верификация данных может иметь важное значение либо быть простой опцией.

И, конечно же, «Обоснованность», т.е. применимость для определенной сферы деятельности.

Обработка больших данных кардинально отличается от обычного процесса бизнес-аналитики, в котором для получения результата достаточно произвести арифметические операции с известными значениями показателей. Сама по себе технология «больших данных» позволяет преобразовать информацию в знания. Интерпретация исходных данных в искомый результат осуществляется посредством поэтапного моделирования, включающего в себя формулирование гипотезы, построение модели для ее проверки и переход к новой гипотезе. Указанная последовательность действий осуществляется посредством преобразования визуальных значений, составления интерактивных запросов на основе знаний, либо разработки адаптивных алгоритмов «машинного обучения».

Внедрение на предприятии технологии «больших данных» позволит не только совершенствовать процесс управления, но и обеспечить бесперебойность производственных процессов, а также более эффективно развивать ключевые бизнес-направления. Например, увеличить производительность, наладить обратную связь с потребителями или повысить уровень безопасности. Какие же преимущества для предприятий, независимо от их отраслевой принадлежности, открывают «большие данные»? К ним относят:

- минимизацию продолжительности незапланированных простоев производства. Как показало исследование, проведенное Honeywell совместно с KRC Research, незапланированные простои производства, вызванные поломкой оборудования, являются основной угрозой неполучения прибыли [5]. Осуществление непрерывного контроля технического состояния оборудования и технологических процессов позволит предприятиям различных сфер деятельности прогнозировать и предотвращать указанные угрозы. Посредством данных с сенсоров и датчиков, установленных на оборудовании, специалисты получают оперативную информацию о состоянии производственного процесса, на основании которой осуществляется планирование диагностики оборудования в целях выявления возможных неисправностей, в результате чего улучшается эффективность использования оборудования и производственных активов в целом, снижаются затраты на техническое обслуживание, повышается производительность;

- улучшение контроля качества продукции, дизайна и удобства использования продукта, повышение лояльности и уровня удовлетворенности клиентов на основе данных о поведении потребителей;

- увеличение точности при прогнозировании объемов закупок и формировании логистических цепочек поставок сырья и материалов, снижение убытков от хранения

избыточных запасов, убытков от задержек в производстве на основе данных о продажах за предыдущие периоды;

- снижение себестоимости продукции за счет оптимизации производственной цепочки;

- оптимизацию бизнес-процессов и значительное увеличение их эффективности за счет сокращения времени подготовки отчетов и обработки данных, автоматизации учета продукции, определения ее себестоимости и цены, т.е. обеспечения пользователей наиболее актуальной информацией.

На сегодняшний день основными потребителями технологии «большие данные» в России являются банки, телекоммуникационные компании и топливно-энергетический комплекс. На втором месте – компании из транспортной отрасли, торговли, а также крупные промышленные предприятия. Но, по прогнозам аналитиков, уже к 2020 году как минимум 40% всех новых данных в мире будет генерироваться именно промышленными датчиками/сенсорами [10].

Примерами успешной апробации «больших данных» в российской промышленности являются Антипинский нефтеперерабатывающий завод в Тюменской области, внедривший инновационную распределенную систему управления для автоматизации производственных процессов, а также Ильский нефтеперерабатывающий завод в Краснодарском крае, использующий систему, обеспечивающую безопасное ведение технологических процессов

В рамках исследования, проведенного компанией MPI Group, респондентами которого являлись руководители 350 крупных предприятий, 76% опрошенных выразили намерение уже в ближайшее время расширить использование технологии «больших данных» [1]. В настоящее время на рынке представлен отдельный класс продуктов, обеспечивающих решение проблемы обработки «больших данных». Рассмотрим основные из них [7; 11]:

- Electron Data Solutions – система, предназначенная для обеспечения других приложений необходимыми данными в нужное время. С ее помощью производится нормализация, изучение, распространение данных для совершения сделок;

- Electron Community Solutions – система, поддерживающая эффективный и динамичный торговый процесс на международных площадках, для общения с партнерами в формате «тет-а-тет», а также для контрибуции и интеграции рыночных данных;

- MapReduce – модель распределенной обработки данных на компьютерных кластерах, разработанная компанией Google, включающая в себя две функции: map-функцию, разделяющую данные на несколько групп, которые затем обрабатываются параллельно, и reduce-функцию, объединяющую результаты расчетов в варианты ответов;

- HADOOP – программное обеспечение для разработки и выполнения различных распределенных программ, представляющая собой мощный инструмент для обработки данных. Распределенная файловая система, позволяющая хранить информацию практически неограниченного объема для параллельной обработки больших объемов информации на множестве узлов стандартного оборудования;

- MPP (массово-параллельная архитектура), аналогично технологии MapReduce, преобразует данные, распределяя их по множеству узлов, которые выполняют обработку распределенных данных одновременно, но, благодаря более тесной интеграции с Hadoop, данные могут быть обработаны в одном запросе. На основании полученных данных собирается общий результат;

- NoSQL – технология, обеспечивающая встроенные возможности горизонтального масштабирования и параллельной обработки, быстрого получения первых результатов, при этом структура данных создается максимально простой. Используется для сбора и хранения информации в социальных сетях и проектах, решающих задачи с «тяжелыми» запросами или бизнес-аналитике.

На основе указанных платформ разрабатывается программное обеспечение для решения различных управленческих задач, функционирующее либо в режиме реального времени и потоковых данных с принятием решения на основе выбора действий, предписанных той или иной ситуацией, либо по запросу на основе анализа и прогнозирования определенных событий.

В заключение следует сказать, что «большие данные» предлагают предприятию неограниченные возможности в плане принятия более обоснованных решений, оптимизации и повышения эффективности деятельности, а также создания новых продуктов и услуг, т.е. буквально создают предприятие «реального времени». Однако, решая вопрос о внедрении технологии, не следует забывать об обеспечении в ее рамках информационной безопасности, энергетической эффективности, правильном толковании информации в условиях ее манипулирования, неопределенности и неполноты.

ЛИТЕРАТУРА

1. Big Data и блокчейн – прорыв в области анализа данных [Электронный ресурс]. – Режим доступа: <https://forklog.com/big-data-i-blokchejn-proryv-v-oblasti-analiza-dannyh/> (дата обращения 16.01.2019).
2. Cisco прогнозирует почти четырехкратный рост облачного трафика к 2020 году [Электронный ресурс]. – Режим доступа: https://www.cisco.com/c/ru_ru/about/press/press-releases/2016/11-14.html (дата обращения 14.01.2019).

3. Rob Petersen, 38 big facts on Big Data every business leader should know [Электронный ресурс]. – Режим доступа: <http://barnraisersllc.com/2012/12/38-big-facts-big-datacompanies/> (дата обращения 14.01.2019).
4. Френкс Б. Укрощение больших данных. – М.: Манн, Иванов и Фербер, 2014. – 352 с.
5. Влияние больших данных на промышленную автоматизацию [Электронный ресурс]. – Режим доступа: <https://www.itbestsellers.ru/companies-analytics/detail.php?ID=34844> (дата обращения 16.01.2019).
6. Денисова О. Ю., Мухутдинов Э. А. Большие данные – это не только размер данных // Вестник Казанского технологического университета. – 2015. – Т. 18, № 4. – С. 226–230.
7. Клеменков П. А., Кузнецов С. Д. Большие данные: современные подходы к хранению и обработке // Труды ИСП РАН. – 2012. – № 23. – С. 143–156.
8. Медведева М. Б., Басакина И., Мкртчян Д. Big data. Финансовые реалии [Электронный ресурс] // ТДР. – 2016. – № 1. – Режим доступа: <http://cyberleninka.ru/article/n/bigdata-finansovye-realii> (дата обращения: 19.01.2019).
9. Найдич А. Большие данные: насколько они большие? [Электронный ресурс]. – Режим доступа: <http://compress.ru/article.aspx?id=23469>. (дата обращения 14.01.2019)
10. Савчук И. Big Data – технология, рождающая новый тип бизнеса [Электронный ресурс]. – Режим доступа: <http://bit.samag.ru/archive/article/1352>. (дата обращения 16.01.2019).
11. Соколянский В. В., Пашков Б. С. Технологии BIG DATA и их инсталляции в экономические исследования // Вопросы экономических наук. – М., 2015. – Т. 74, № 4. – С. 167–169.