

Кузьмин Антон Вячеславович

студент

ФГБОУ ВО «Поволжский государственный
университет телекоммуникаций и информатики»

г. Самара, Самарская область

БУДУЩЕЕ ИННОВАЦИОННОЙ ТЕХНОЛОГИИ BIG DATA

Аннотация: в данной статье рассмотрена технология Больших данных (Big Data). Автором представлены признаки технологии Big Data.

Ключевые слова: технология Больших данных, Big Data, информация, информационно-коммуникационными технология.

Активное развитие на современном этапе в сфере информационных технологий различного назначения получила технология Больших данных, или Big Data. Данная категория используется как собирательное наименование совокупности подходов, инструментов и методов, предназначенных для обработки структурированных и неструктурированных данных огромного объема и значительного разнообразия, и рассматривается как один из ключевых параметров степени экономического и технологического развития системы любого уровня.

Категорию «Big Data» в научный оборот в сфере информационных технологий ввел Клифорд Линч [10, с. 18]. Непосредственно под Большими данными рассматриваются изначально не систематизированные и происходящие из самых разнообразных источников данные о развитии окружающей среды (природной и социально-экономической), генерирование которых происходит посредством реализации информационно-коммуникационных технологий. Консалтинговая компания «Форрестер» обращает внимание, что в технологии Больших данных объединяются те техники и технологии, которые позволяют извлекать смысл их данных на экстремальном пределе практичности [6, с. 47], т.е. придают целесообразность обработке и анализу данных.

В силу комплексного характера технологии при ее определении более оптимально исходить из присущих данной технологии признаков, отличающих ее от других способов обработки данных.

На начальном этапе становления Big Data на рубеже 1990–2000-х годов Дуг Лани выделил три таких признака [2, с. 115]:

– объем как наиболее характерный признак технологии, выражающийся в возможности генерировать значительное количество данных, что, во-первых, недоступно для других технологий обработки, и, во-вторых, обеспечивают огромный потенциал для исследования;

– разнообразие, отражающее возможность посредством технологии проводить анализ и обработку самых различных данных, не ограниченных какими-то отдельными параметрами. Классифицировать источники возникновения больших данных можно следующим образом: 1) любые базы данных, которые ведут государственные и негосударственные организации в рамках своей деятельности (электронные медицинские карты, дела сотрудников, организационная статистика и т. д.), 2) данные о совершаемых организациями и гражданами транзакциях и операциях, в том числе посредством применения мобильных устройств (выдача средств в банкоматах, оплата покупок и т. д.); 3) источники данных, связанные с работой сенсорных сетей (со спутников, измерительных устройств и т. д.) при условии, что информация отправляется в хранилище данных; 4) данные с регистрирующих устройств (Глобальная система определения координат, сети мобильной телефонной связи и т. д.); 5) данные о поведении пользователей, например, Интернета (запросы в поисковиках, просмотры веб-страниц и т. д.); 6) данные о выражении пользователями мнений (например, комментарии в социальных сетях) [7];

– скорость генерирования и обработки данных.

Впоследствии к этому перечню добавились еще три признака [9, с. 8–9]:

– стоимость, что стало предполагать наличие экономического эффекта в результате обработки данных;

– достоверность, что отразило возможность посредством Big Data обеспечивать качество собранных данных даже при значительности различий между ними;

– сложность, что было связано со сложностью управления данными в рамках рассматриваемой технологии, особенно в ситуациях, если источники происхождения данных очень различны. В частности, для сбора и обработки информации используется несколько групп средств: программное обеспечение, оборудование, сервисные услуги, не говоря уже о многообразии применяемых методов анализа.

Обобщая признаки (рисунок 1) можно заключить, что предназначение технологии Больших данных состоит в обработке беспрецедентно огромного объема данных, которые поступают из самых различных областей и источников в разных форматах, что требует поиска таких способов их обработки, которые бы позволили извлекать из них ценную информацию.



Рис. 1. Признаки технологии Big Data

Многообразие входящих в Big Data технологий обуславливает множество критериев классификации больших данных: по типу источников данных (интернет, медиа, машинные данные, оперативные данные), по формату контента (структурированные, полуструктурированные, неструктурированные), по форме хранения данных (документы, таблицы, графики), по способу постановки данных (чистые, нормализация, преобразование), по характеру обработки данных (потоками, в режиме реального времени) [2, с. 115–116].

Инструментарий Big Data включает в себя следующие составляющие:

– машинное обучение как комплекс подходов, алгоритмов, инструментов, которые позволяет создавать модели, способные к обучению на основе примеров – так называемое индуктивное обучение;

– извлечение информации – это задача автоматического извлечения структурированных данных из неструктурированных или слабоструктурированных машиночитаемых документов;

– интеллектуальный анализ данных (Data Mining) как совокупность методов, позволяющих найти необходимую информации в данных, находящихся в ранее неизвестных, нетривиальных, полезных и доступных для интерпретации знаниях. В основе анализа здесь лежат методы классификации, моделирования, прогнозирования, опирающиеся на такие инструменты как дерево решений, искусственные нейронные сети, генетические алгоритмы, эволюционное программирование, ассоциативная память, нечеткая логика.

Непосредственно применяемыми инструментами являются Hadoop & MapReduce, NoSQL базы данных, углубленная аналитика (статистика, предиктивная аналитика, лингвистическая обработка текстов), инструменты класса Data Discovery. Дополнительно следует говорить о развитии инструментов представления результатом обработки данных для пользователей [3].

Аналитика рынка Big Data демонстрирует постоянно возрастающую активность проникновения технологии в деятельность в разных сферах [1]. Наиболее развитие Большие данные закономерно получили в области телекоммуникаций (58%) и инжиниринге (35%). Лидерами внедрения технологии являются

зарубежные компании, прежде всего, Nasdaq, Facebook, Google, IBM, VISA, MasterCard, BankofAmerica, HSBC, AT&T, CocaCola, Starbucks и Netflix. Помимо этого, в венчурной индустрии уже сегодня Big Data воспринимаются как один из наиболее перспективных объектов для инвестирования. В целом же, значительные возможности использования рассматриваемой технологии есть в самых разнообразных сферах: энергетике, метеорологии, науке, электронной коммерции, транспорте, финансах, здравоохранении, безопасности и т. д. – везде, где требуется обрабатывать и анализировать значительные массивы данных, находить закономерности и корреляции в состояниях и процессах. Преимущества, предоставляемые технологией Больших данных, значительны. Она позволяет и детализировать сведения по конкретным запросам, и обеспечить скорость обработки постоянно меняющейся информации, и дать возможность доступа к самым разнообразным данным вне зависимости от их происхождения, и способствовать многоаспектности измерений, выстраивая такие взаимосвязи между явлениями и процессам, которые для более традиционных баз данных просто невозможны.

По сути, информация – это новый фактор производства, позволяющий производить товары и услуги с высокой долей добавленной стоимости и принимать эффективные управленческие решения, особенно в стратегической перспективе. Эксперты пришли к выводу, что использование автоматизированных технологий принятия решений позволяет увеличить производительность в среднем на 5–6% [6, с. 45].

Таким образом, повсеместный охват информационно-коммуникационными технологиями жизни и деятельности людей в разных их сферах, вызвал рост данных, которые генерируются ежедневно в значительных объемах. И это поставило проблему не просто хранения и структурирования данных, но и использование их в качестве информационного ресурса для продуцирования нового знания, принятия решений, прогнозирования социально-экономического развития. Ответом на данную потребность и стала технология Big Data, позволившая обрабатывать и анализировать огромные объемы не только структурированных, но и неструктурированных данных.

Список литературы

1. Булавин Д. Аналитический обзор рынка Больших данных / Д. Булавин, А. Чикунов, С. Будушкаева. – М.: Борд, 2015. – 51 с.
2. Жукабаева Т.К. Технология больших данных (Big Data). Основные характеристики и перспективы применения / Т.К. Жукабаева, А.Т. Кусаинова // Новости науки Казахстана. – 2016. – №1 (127). – С. 114–132.
3. Крюкова А.А. Особенности тестирования, внедрения и разработки десктопных приложений / А.А. Крюкова, В.А. Почебут // Синергия наук. – 2017. – №12. – С. 964–967.
4. Майер-Шенбергер В. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим / В. Майер-Шенбергер, К. Кукьер; пер. с англ. Инны Гайдюк. – М.: Манн, Иванов и Фербер, 2014. – 240 с.
5. Рязанцев И.С. О критериях оценки степени инновационности развития социально-экономической системы // Креативная экономика. – Т. 3. – №1. – С. 41–44.
6. Савельев А.И. Проблемы применения законодательства о персональных данных в эпоху «Больших данных» (Big Data) // Право. Журнал Высшей школы экономики. – 2015. – №1. – С. 43–66.
7. Там С.-М., Кларк, Ф. Большие данные, официальная статистика и некоторые инициативы Австралийского бюро статистики // Межгосударственный статистический комитет Содружества независимых государств [Электронный ресурс]. – Режим доступа: http://www.cisstat.com/BigData/CIS-BigData_03.pdf (дата обращения: 18.08.2017).
8. Хасаншин И.А. Анализ основных путей информационного взаимодействия электронного правительства с инновационными предприятиями // Экономика и предпринимательство. – 2016. – №7 (72). – С. 974–978.
9. Цымблер М. Big Data: несколько простых вопросов о сложном явлении // Суперкомпьютеры. – 2014. – №1 (17). – С. 8–11.
10. Черняк Л. Большие данные – новая теория и практика // Открытые системы. СУБД. – М.: Открытые системы, 2011. – №10. – С. 18–25.