

УДК 004.82

ТЕНДЕНЦИИ РАЗВИТИЯ ТЕХНОЛОГИЙ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ И ИНСТРУМЕНТАРИЯ ХРАНЕНИЯ РАЗНОФОРМАТНЫХ ДАННЫХ И АНАЛИТИКИ

М.Р. Биктимиров¹, А.М. Елизаров², А.Ю. Щербаков³

¹ *Всероссийский институт научной и технической информации РАН, г. Москва*

² *Институт математики и механики им. Н.И. Лобачевского Казанского (Приволжского) федерального университета*

³ *НИУ «Высшая школа экономики»; Федеральный исследовательский центр «Информатика и управление» РАН, г. Москва*

¹ marat@ras.ru, ² amelizarov@gmail.com, ³ x509@ras.ru

Аннотация

Статья посвящена анализу тенденций развития технологий обработки Больших Данных и инструментария хранения разноформатных данных и аналитики, который проведен в рамках работ по программе фундаментальных исследований Отделения математических наук РАН «Алгебраические и комбинаторные методы математической кибернетики и информационные системы нового поколения», а также гранта РФФИ № 14-07-00783 «Способы хранения и обработки большого объема научно-справочных данных на современных аппаратных платформах».

Ключевые слова: Большие данные, анализ, информация, программное обеспечение, распределенные вычисления, системы хранения, облачные технологии.

КЛЮЧЕВЫЕ ХАРАКТЕРИСТИКИ БОЛЬШИХ ДАННЫХ

Как известно (см., например, [1]), «Большие Данные» (Big Data) сегодня являются одним из ключевых драйверов (стимулов) развития информационно-коммуникационных технологий (ИКТ). Это направление развития ИКТ, относи-

тельно новое для России, получило широкое распространение в западных странах. Связано это с тем, что в эпоху информационных технологий, особенно после бума социальных сетей, стало накапливаться значительное и все возрастающее количество информации, связанной с каждым пользователем интернета, что в конечном счете дало развитие направлению Больших Данных.

Термин «Большие Данные» вызывает множество споров. Многие полагают, что этот термин характеризует лишь объем накопленной информации, но не стоит забывать о технической стороне: названное направление развития ИКТ включает в себя технологии хранения и вычисления, а также сервисные услуги. Следует отметить, что к названной сфере относится обработка именно большого объема информации, который затруднительно обрабатывать традиционными способами. Например, международная исследовательская компания Forrester (см. [2]) определяет это понятие как технологию в области аппаратного и программного обеспечения, которая объединяет, организует, управляет и анализирует данные, характеризующиеся «четырьмя V»: объемом (Volume), разнообразием (Variety), изменчивостью (Variability) и скоростью (Velocity):

- Volume – это очень большой объем информации, накопленный в базах данных, его трудоемко обрабатывать и хранить традиционными средствами СУБД; поэтому востребованы новые подходы и усовершенствованные инструменты обработки этих данных.

- Variety – это разнообразие (многообразие) форматов данных (главный критерий Больших Данных): большие массивы данных, поступающие из разных источников в различных форматах, разной степени структурированности – табличные данные в СУБД, иерархические данные, текстовые документы, видео, изображения, аудиофайлы и т. д.; поэтому востребована возможность одновременной обработки структурированной и неструктурированной разноформатной информации. Главное отличие структурированной информации – в том, что она может быть классифицирована. Примером может служить информация о клиентских транзакциях. Неструктурированная информация включает в себя видео-, аудиофайлы, свободный текст, информацию, поступающую из социальных сетей. На сегодняшний день 80% появляющейся информации входит в груп-

пу неструктурированной. Такая информация нуждается в комплексном анализе, чтобы сделать ее полезной для дальнейшей обработки;

- **Variability** – изменчивость информации: например, таковой является информация, непрерывно поступающая с датчиков некоторых устройств или из интернета и имеющая важное значение для анализа, прогнозирования и принятия решений;

- **Velocity** – скорость накопления и обработки данных; данный признак указывает как на увеличивающуюся скорость накопления данных (90% информации было собрано за последние 2 года), так и на скорость их обработки; в последнее время в ряде задач стали более востребованы технологии обработки данных в реальном времени.

К перечисленным свойствам Больших Данных сегодня добавляют:

- **Veracity** – достоверность данных: все большую значимость пользователи стали придавать достоверности имеющихся данных. Так, у интернет-компаний есть проблема по разделению действий, проводимых на сайте компании роботом и человеком, что приводит в конечном счете к затруднению анализа данных;

- **Value** – ценность накопленной информации: Большие Данные должны быть полезны компаниям и приносить им определенную выгоду, например, должны помогать в усовершенствовании бизнес-процессов, составлении отчетности или оптимизации расходов.

Таким образом, при соблюдении указанных выше условий накопленные объемы данных можно относить к числу больших.

СФЕРЫ ПРИМЕНЕНИЯ БОЛЬШИХ ДАННЫХ

Сфера использования технологий Больших Данных весьма обширна (см. обзор, размещенный нами в [3]). Так, с их помощью можно узнать о предпочтениях клиентов, эффективности маркетинговых кампаний или провести анализ рисков. Ниже представлены некоторые результаты опроса IBM Institute о направлениях использования Больших Данных в компаниях (см. [4, 5]).

Большинство компаний использует Большие Данные в сфере клиентского сервиса, второе по популярности направление – операционная эффективность, в

сфере управления рисками Большие Данные на текущий момент времени менее распространены.

Следует отметить, что Большие Данные являются одной из самых быстро-растущих сфер информационных технологий: согласно статистике (например, [6]), общий объем получаемых и хранимых данных удваивается каждые 1,2 года. За период с 2012 по 2014 годы количество данных, ежемесячно передаваемых мобильными сетями, выросло на 81%. По оценкам Cisco [7], в 2014 году объем мобильного трафика составил 2,5 эксабайта (единица измерения количества информации, равная 10^{18} стандартным байтам) в месяц, а уже в 2019 году он будет равен 24,3 эксабайтам. Таким образом, Большие Данные – это уже устоявшаяся сфера технологий, несмотря на относительно молодой возраст, получившая распространение во многих сферах бизнеса и играющая немаловажную роль в развитии компаний.

ТЕХНОЛОГИИ БОЛЬШИХ ДАННЫХ

Технологии, используемые для сбора и обработки Больших Данных, можно разделить на три группы: программное обеспечение (ПО), технологическое оборудование и сервисные услуги.

К наиболее распространенным инструментам обработки Больших Данных из группы ПО относятся (см. также [8]):

SQL – язык структурированных запросов, позволяющий работать с базами данных. С его помощью можно создавать и модифицировать данные, а управлением массивом данных занимается соответствующая система управления базами данных.

NoSQL (термин расшифровывается как Not Only SQL (не только SQL)) [9] включает ряд подходов, направленных на реализацию базы данных, имеющих отличия от моделей, используемых в традиционных, реляционных СУБД. Поэтому их удобно использовать при постоянно меняющейся структуре данных, в частности, для сбора и хранения информации в социальных сетях.

MapReduce [10] – модель распределения вычислений. Применяется для параллельных вычислений с использованием очень больших наборов данных (петабайты и более). В программном интерфейсе на обработку программе пе-

редаются не данные, а, наоборот, программа – данным. Таким образом, запрос представляет собой отдельную программу. Принцип работы заключается в последовательной обработке данных двумя методами – Map и Reduce. Map выбирает предварительные данные, а Reduce агрегирует их.

Hadoop – используется для реализации поисковых и контекстных механизмов высоконагруженных сайтов – Facebook, eBay, Amazon и др. Отличительной особенностью является то, что система защищена от выхода из строя любого из узлов кластера, так как каждый блок имеет, как минимум, одну копию данных на другом узле.

SAP HANA [11] – это высокопроизводительная NewSQL-платформа для хранения и обработки данных, которая обеспечивает высокую скорость обработки запросов. Сочетание в этой платформе технологий OLAP [12] и OTLP [13] создает унифицированный ракурс данных, полученных из систем обработки транзакций, систем анализа, принятия решений и планирования. Еще одним отличительным признаком является то, что SAP HANA упрощает системный ландшафт, уменьшая затраты на поддержку аналитических систем.

К *технологическому оборудованию*, как известно, относят серверы и инфраструктурное оборудование. Серверы включают в себя хранилища данных. К инфраструктурному оборудованию относят средства ускорения платформ, источники бесперебойного питания, комплекты серверных консолей и др.

Сервисные услуги включают в себя услуги по построению архитектуры системы базы данных, обустройству и оптимизации инфраструктуры, а также обеспечению безопасности хранения данных.

Программное обеспечение, оборудование, а также сервисные услуги вместе образуют комплексные платформы для хранения и анализа данных. Такие компании, как Microsoft, HP, EMC, предлагают услуги по разработке, развертыванию решений Big Data и управления ими.

ПРИМЕНЕНИЕ БОЛЬШИХ ДАННЫХ В РАЗЛИЧНЫХ ОТРАСЛЯХ

Большие Данные получили широкое распространение во многих отраслях бизнеса – их используют в здравоохранении, телекоммуникациях, торговле, логистике, финансовых компаниях, а также в государственном управлении. Ниже представлено несколько примеров применения Больших Данных в некоторых из этих отраслей.

Розничная торговля. В базах данных розничных магазинов может быть накоплено множество информации о клиентах, системе управления запасами и поставками товарной продукции. Эта информация может быть полезна во всех сферах деятельности магазинов. Так, например, с помощью накопленной информации можно управлять поставками товара, его хранением и продажей, а также прогнозировать спрос и поставки товара. Система обработки и анализа данных может решить и другие проблемы ритейлера, например, оптимизировать затраты или подготавливать отчетность.

Финансовые услуги. Большие Данные дают возможность проанализировать кредитоспособность заемщика, также они полезны для кредитного скоринга и андеррайтинга. Внедрение технологий Больших Данных позволяет сократить время рассмотрения кредитных заявок, а также проанализировать операции конкретного клиента и предложить банковские услуги, подходящие именно ему.

Телеком. В телекоммуникационной отрасли широкое распространение Большие Данные получили у сотовых операторов. Операторы сотовой связи наравне с финансовыми организациями имеют одни из самых объемных баз данных, что позволяет им проводить наиболее глубокий анализ накопленной информации. Главной целью анализа данных являются удержание существующих клиентов и привлечение новых. Для этого компании проводят сегментацию клиентов, анализируют их трафики, определяют социальную принадлежность абонентов. Помимо использования Больших Данных в маркетинговых целях, эти технологии применяются для предотвращения мошеннических финансовых операций.

Горнодобывающая и нефтяная промышленности. Большие Данные используются как при добыче полезных ископаемых, так и при их переработке и сбыте. На основании поступившей информации предприятия могут делать выводы об эффективности разработки месторождения, отслеживать график капитального ремонта и состояния оборудования, прогнозировать спрос на продукцию и цены.

По данным опроса Tech Pro Research [15], наибольшее распространение Большие Данные получили в телекоммуникационной отрасли, а также в инжиниринге, ИТ, в финансовых и государственных предприятиях. По результатам данного опроса, менее популярны Большие Данные в образовании и здравоохранении.

ПРИМЕРЫ ИСПОЛЬЗОВАНИЯ БОЛЬШИХ ДАННЫХ В КОМПАНИЯХ

На сегодняшний день Большие Данные активно внедряются в зарубежных компаниях. Такие компании, как Nasdaq, Facebook, Google, IBM, VISA, Master Card, Bank of America, HSBC, AT&T, Coca Cola, Starbucks и Netflix, уже активно используют ресурсы Больших Данных.

Сферы применения обработанной информации разнообразны и варьируются в зависимости от отрасли и задач, которые необходимо решить. Далее будут представлены примеры применения технологий Больших Данных на практике.

Компания HSBC [16] использует технологии Больших Данных для противодействия мошенническим операциям с пластиковыми картами. С помощью этих технологий компания увеличила эффективность службы безопасности в 3 раза, распознавание мошеннических инцидентов – в 10 раз. Экономический эффект от внедрения таких технологий превысил 10 млн. долл. США.

Система Антифрод VISA [16] позволяет в автоматическом режиме вычислить операции мошеннического характера, в настоящее время она позволяет предотвратить мошеннические платежи на сумму до 2 млрд. долл. США ежегодно.

Суперкомпьютер Watson компании IBM анализирует в реальном времени поток данных по денежным транзакциям. По данным IBM [16], Watson на 15% увеличил количество выявленных мошеннических операций, на 50% сократил

ложные срабатывания системы и на 60% увеличил сумму денежных средств, защищенных от транзакций такого характера.

Компания Procter & Gamble (P&G) с помощью Больших Данных проектирует новые продукты и составляет глобальные маркетинговые акции [16]. P&G создала специализированные офисы Business Spheres, где можно просматривать информацию в реальном времени. Таким образом, у менеджмента компании появилась возможность мгновенно проверять гипотезы и проводить эксперименты. В P&G считают, что Большие Данные существенно помогают в прогнозировании деятельности компании.

Ритейлер офисных принадлежностей OfficeMax с помощью технологий Больших Данных анализируют поведение клиентов [17]. Анализ Больших Данных позволил увеличить B2B выручку на 13%, уменьшить затраты на 400 тыс. долларов США в год.

По мнению компании Caterpillar [16], ее дистрибьюторы ежегодно упускают от 9 до 18 млрд. долл. США прибыли только из-за того, что не внедряют технологии обработки Больших Данных. Последние позволили бы клиентам более эффективно управлять парком автомобилей за счет анализа информации, поступающей с датчиков, установленных на них. На сегодняшний день уже есть возможность анализировать состояние ключевых узлов, их степени износа, управлять затратами на топливо и техническое обслуживание.

Компания Luxottica group является производителем спортивных очков таких марок, как Ray-Ban, Persol и Oakley. Она применяет технологии Больших Данных для анализа поведения потенциальных клиентов и «умного» смс-маркетинга. В результате Luxottica group выделила более 100 миллионов наиболее ценных клиентов и повысила эффективность маркетинговой кампании на 10% (см. [16]).

С помощью Yandex Data Factory разработчики игры World of Tanks анализируют поведение игроков. Технологии Больших Данных позволили проанализировать поведение 100 тысяч игроков World of Tanks с использованием более 100 параметров (информация о покупках, играх, опыт и др.). В результате анализа был получен прогноз оттока пользователей. Данная информация позволяет уменьшить уход пользователей и работать с участниками игры адресно. Разра-

ботанная модель оказалась на 20–30% эффективнее стандартных инструментов анализа игровой индустрии (см. [18]).

Министерство труда Германии использует Большие Данные в работе, связанной с анализом поступающих заявок на выдачу пособий по безработице (см. [16]). Так, проведенный анализ информации показал, что 20% пособий выплачивалось незаслуженно. В результате министерство сократило расходы на 10 млрд. евро.

Детская больница Торонто [19] внедрила проект Project Artemis. Это информационная система, которая в реальном времени собирает и анализирует данные о младенцах и непрерывно отслеживает более 1250 показателей состояния каждого ребенка, что позволяет прогнозировать появление нестабильного состояния и вовремя начинать профилактику заболеваний у детей.

ТЕКУЩЕЕ СОСТОЯНИЕ МИРОВОГО РЫНКА

В 2014 г. Большие Данные, по мнению Data Collective [20], стали одними из приоритетных направлений инвестирования в сфере венчурной индустрии. Согласно данным информационного портала Компьютерра [20], связано это с тем, что разработки в данном направлении начали приносить значительные результаты пользователям. За 2014 год количество компаний с реализованными проектами в сфере управления Большими Данными увеличилось на 125%, объем рынка по сравнению с 2013 годом вырос на 45%.

Большую часть выручки рынка Больших Данных, по мнению Wikibon [21], в 2014 году составили сервисные услуги, их доля была равно 40% в общем объеме выручки. Согласно данным Wikibon, в 2014 году 36% выручки принесли приложения и аналитика Больших Данных, 17% – вычислительное оборудование и 15% – технологии хранения данных. Меньше всего выручки было сгенерировано NoSQL-технологиями, инфраструктурным оборудованием и обеспечением сетью компаний (корпоративные сети).

Наибольшей популярностью пользуются такие технологии Больших Данных, как in-memory платформы компаний SAP, HANA, Oracle и др. Результаты опроса T-Systems показали (см., например, [22]), что такие технологии выбрали 30% опрошенных компаний. Вторыми по популярности стали NoSQL-платформы (18% пользователей), также компании использовали аналитические платформы

компаний Splunk и Dell (их выбрало 15% компаний). Наименее полезными для решения проблем Больших Данных, по результатам опроса, оказались продукты Hadoop/MapReduce.

По данным опроса Accenture (см. [23]), в более чем 50% компаний, использующих технологии Больших Данных, затраты на них составляют от 21% до 30%. Согласно анализу Accenture, 76% компаний считают, что такие расходы будут увеличиваться, а 24% компаний не изменят своего бюджета на технологии Больших Данных. Это говорит о том, что в последних из упомянутых компаний технологии Больших Данных уже стали устойчивым направлением использования ИКТ, ставшим неотъемлемой составляющей их развития.

Результаты опроса Economist Intelligence Unit survey [23] подтверждают положительный эффект от внедрения Больших Данных: 46% компаний заявили, что с помощью этих технологий они улучшили клиентский сервис более, чем на 10%; 33% компаний оптимизировали запасы и улучшили продуктивность основных активов; 32% компаний улучшили процессы планирования.

БОЛЬШИЕ ДАННЫЕ В РАЗНЫХ СТРАНАХ МИРА

На сегодняшний день технологии Больших Данных чаще всего внедряются в компаниях США, но и другие страны мира начали проявлять интерес к этим технологиям. В 2014 году, по данным IDC [24], на страны Европы, Ближнего Востока, Азии (за исключением Японии) и Африки пришлось 45% рынка ПО, услуг и оборудования в сфере Больших Данных. Также, согласно опросу CIO [24], компании из стран Азиатско-Тихоокеанского региона быстрыми темпами осваивают новые решения в области анализа Больших Данных, безопасного хранения и облачных технологий. Латинская Америка находится на втором месте по количеству инвестиций в развитие технологий Больших Данных, опережая страны Европы и США. Ниже даны описание и прогнозы развития рынка Больших Данных нескольких стран с использованием этих данных [24].

Китай. Объем информации Китая составляет 909 эксабайт, что равно 10% общего объема информации в мире, к 2020 году объем информации достигнет 8060 эксабайт, увеличится и доля информации в общемировой статистике, через

5 лет она будет равна 18%. Потенциальный рост Больших Данных Китая имеет одну из самых быстрорастущих динамик.

Бразилия по итогам 2014 года накопила информации на 212 эксабайт, что составляет 3% от общемирового объема. К 2020 году объем информации вырастет до 1600 эксабайт, что составит 4% информации всего мира.

Индия. По данным EMC [25], объем накопленных данных Индии по итогам 2014 года составляет 326 эксабайт, что составляет 5% от общего объема информации. К 2020 году объем информации вырастет до 2800 эксабайт, что составит 6% информации всего мира.

Япония. Объем накопленных данных Японии по итогам 2014 года составляет 495 эксабайт (8% от общемирового объема информации). К 2020 году объем информации вырастет до 2200 эксабайт, но доля рынка Японии уменьшится и составит 5% от общего объема информации всего мира. Таким образом, объем рынка Японии уменьшится более, чем на 30%.

Германия. По данным EMC [25], объем накопленных данных в Германии по итогам 2014 года составляет 230 эксабайт (4% от общемирового объема информации). К 2020 году объем информации вырастет до 1100 эксабайт и составит 2% от общемирового. На рынке Германии большую долю выручки, по прогнозам Experton Group [26], будет генерировать сегмент сервисных услуг, доля которых в 2015 году составила 54%, а в 2019 году увеличится до 59%; доли программного обеспечения и оборудования, наоборот, уменьшатся.

В целом объем рынка вырастет с 1,345 млрд евро в 2015 году до 3,198 млрд. евро в 2019 году, средний темп роста составит 24%. Таким образом, на основании аналитики CIO и EMC можно сделать вывод, что развивающиеся страны мира в ближайшие годы станут рынками активного развития технологий Больших Данных.

ОСНОВНЫЕ ТЕНДЕНЦИИ РЫНКА

По мнению IDG Enterprise [27], в 2015 расходы компаний на сферу Больших Данных составили в среднем 7,4 млн. долл. США на компанию (у крупных компаний это примерно 13,8 млн. долл. США, у малых и средних – 1,6 млн. долл. США). Больше всего средств инвестировано в такие области, как анализ и визуализация данных и их сбор. Согласно имеющейся информации, инвестиции в

2015 году были использованы на улучшение качества данных, совершенствование планирования и прогнозирования, а также увеличение скорости обработки данных. Компаниями финансового сектора, по данным Bain Company's Insights Analysis [24], были произведены значительные инвестиции (в 2015 году потрачено около 6,4 млрд долл. США на технологии Больших Данных, средний темп роста инвестиций составит 22% до 2020 года; интернет-компании потратили порядка 2,8 млрд. долл. США, средний темп роста увеличения затрат на Большие Данные составит 26%). При проведении опроса Economist Intelligence Unit survey были выявлены приоритетные направления развития Больших Данных в 2014 году и на ближайшие 3 года (см. также [24]).

ВЫВОДЫ

По прогнозам IDC [25], тенденции развития рынка Больших Данных выглядят следующим образом:

- в следующие 5 лет затраты на облачные решения в сфере технологий Больших Данных будут расти в 3 раза быстрее, чем затраты на локальные решения; станут востребованными гибридные платформы для хранения данных;
- рост приложений с использованием сложной и прогнозной аналитики, включая машинное обучение, ускорится, рынок таких приложений будет расти на 65% быстрее, чем приложения, не использующие прогнозную аналитику;
- медиа-аналитика станет ключевым драйвером роста рынка технологий Больших Данных;
- ускорится тенденция внедрения решений для анализа постоянного потока информации, которая применима для интернета вещей;
- к 2018 году 50% пользователей будут взаимодействовать с сервисами, основанными на технологиях когнитивных вычислений.

СПИСОК ЛИТЕРАТУРЫ

1. *Kord Davis with Doug Patterson. Ethics of Big data: Balancing Risk and Innovation.* O'Reilly Media, Inc., 2012. 82 p. ISBN: 978-1-4493-1179-7.

2. Hopkins B., Evelson B. Expand your digital horizon with Big Data. 2011. 16 p. http://www.asterdata.com/newsletter-images/30-04-2012/resources/forrester_expand_your_digital_horiz.pdf

3. Обзор рынка BigData. <http://habrahabr.ru/company/moex/blog/256747/>

4. IBM: Бизнес выбирает когнитивные решения. <https://www.crn.ru/news/detail.php?ID=109059>

5. Будзко В.И. BigData. Новый вызов. <http://www.myshared.ru/slide/535343/>, 18 марта 2016 г.

6. «Мировой объем данных увеличивается более чем в два раза каждые два года, большие объемы данных открывают новые возможности и изменяют роль ИТ», <tweet> <http://russia.emc.com/about/news/press/2011/20110628-01.htm>

7. Cisco: в период с 2012 по 2017 гг. объем мобильного трафика вырастет в 13 раз Пресс-релиз. http://www.cisco.com/c/ru_ru/about/press/press-releases/2013/02-020713a.html

8. Big Data от А до Я. Часть 1: Принципы работы с большими данными, парадигма MapReduce. <https://habrahabr.ru/company/dca/blog/267361/>

9. NoSQL: <http://nosql-database.org/>

10. Apache Hadoop. MapReduce Tutorial. https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html

11. Apache Hadoop. <http://hadoop.apache.org/>

12. SAP HANA. <https://hana.sap.com/abouthana.html>

13. Cios K.J. Data Mining: A Knowledge Discovery Approach. Springer 2007, 606 p. ISBN 978-0-387-33333-5.

14. Wakelin P., Day S., Read S., McKenna F. CICS transaction gateway V3.1: the WebSphere connector for CICS. SG24-6133-00, IBM, 2001. 269 p.

15. <http://www.techproresearch.com/>, <http://22century.ru/popular-science-publications/gartner-preliminary-results/attachment/bigdata>

16. Медетов А.А. Термин Big Data и способы его применения // Молодой ученый. 2016. № 11. С. 207–210.

17. Эра Big Data. Как ритейлеры используют данные о своих клиентах. <http://retail-community.com.ua/news/tehnologii/3677-era-big-data-kak-riteyleryi-ispolzuyut-dannye-o-svoih-klientah>

18. Предсказание оттока игроков из World of Tanks от Yandex Data Factory. Лекция для Малого ШАДа. <https://habrahabr.ru/company/yandex/blog/281777/>

19. *Сергей Немалевич*. Гадание на больших данных. <https://republic.ru/biz/1152340/>

20. *Андрей Васильков*. Data Collective: «большие данные» — приоритетное направление инвестиций. <http://www.computerra.ru/97457/data-collective-relies-on-big-data/>

21. Wikibon Big Data Capital Markets Day 2014, Jeff Kelly. <http://www.slideshare.net/jeffkelly568/wikibon-big-data-capital-markets-day-2014>

22. Небольшие выводы о больших данных, Генеральный директор T-Systems в России. Алексей Тоскин 19 августа 2014. <http://www.computerra.ru/105278/nebolshie-vyivodyi-o-bolshih-dannyih/>

23. 10 лучших продуктов на рынке Big Data в 2015 году. <https://www.crn.ru/news/detail.php?ID=107118>

24. *Измалкова С.А., Головина Т.А.* Использование глобальных технологий «big data» в управлении экономическими системами //Изв. Тульского гос. ун-та. Экономические и юридические науки. 2015. № 4-1. С. 151–158.

25. Большие данные. Большие возможности. <http://www.emc.com/ru-ru/big-data/index.htm>

26. Тренды мирового e-commerce рынка в 2015–2016 годах. https://habrahabr.ru/company/SECL_GROUP/blog/268455/

27. *Семенов Ю.А.* Обзор по материалам ведущих фирм, работающих в сфере сетевой безопасности. <http://book.itep.ru/10/2015.htm>

Благодарности

Работа выполнена при финансовой поддержке РФФИ (проект № 14-07-00783).

TECHNOLOGY TRENDS HANDLING OF BIG DATA AND TOOLS STORAGE OF MULTIFORMAT DATA AND ANALYTICS

M.R. Biktimirov¹, A.M. Elizarov², A.Yu. Scherbakov³

¹ Russian Institute for Scientific and Technical Information of Russian Academy of Sciences (VINITI RAS), Moscow

² N.I. Lobachevskii Institute of Mathematics and Mechanics. Kazan (Volga Region) Federal University, Kazan

³ National Research University "Higher School of Economics"; Federal Research Centre "Information and Management" of the Russian Academy of Sciences, Moscow

Abstract

This article analyzes the development trends of processing Big Data tools and multi-format data storage and analysis. This analysis was carried out as part of our program of basic research of the Department of Mathematical Sciences, Russian Academy of Sciences "Algebraic and Combinatorial Methods of Mathematical Cybernetics and information systems of the new generation", as well as RFBR grant number 14-07-00783 "Way to store and process a large volume of scientific and reference data modern hardware platforms".

Keywords: *Big Data, storage systems, analysis, information, software, grid computing, cloud computing*

REFERENCES

1. Kord Davis with Doug Patterson. Ethics of Big data: Balancing Risk and Innovation. O'Reilly Media, Inc., 2012. 82 p. ISBN: 978-1-4493-1179-7.
2. Hopkins B., Evelson B. Expand your digital horizon with Big Data. 2011. 16 p. http://www.asterdata.com/newsletter-images/30-04-2012/resources/forrester_expand_your_digital_horiz.pdf
3. Obzor rynka BigData. <http://habrahabr.ru/company/moex/blog/256747/>
4. IBM: Biznes vybiraet kognitivnye reshenija. <https://www.crn.ru/news/detail.php?ID=109059>

5. *Budzko V.I.* BigData. Novyj vyzov. <http://www.myshared.ru/slide/535343/>
6. «Mirovoj ob'em dannyh uvelichivaetsja bolee chem v dva raza kazhdye dva goda, bol'shie ob'emy dannyh otkryvajut novye vozmozhnosti i izmenjajut rol' IT», <tweet> <http://russia.emc.com/about/news/press/2011/20110628-01.htm>
7. Cisco: v period s 2012 po 2017 gg. Ob'em mobil'nogo trafika vyrastet v 13 raz Press-reliz. http://www.cisco.com/c/ru_ru/about/press/press-releases/2013/02-020713a.html
8. Big Data ot A do Ja. Chast' 1: Principy raboty s bol'shimi dannymi, paradigma MapReduce. <https://habrahabr.ru/company/dca/blog/267361/>
9. NoSQL: <http://nosql-database.org/>
10. Apache Hadoop. MapReduce Tutorial. https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html
11. Apache Hadoop. <http://hadoop.apache.org/>
12. SAP HANA. <https://hana.sap.com/abouthana.html>
13. *Cios K.J.* Data Mining: A Knowledge Discovery Approach. Springer 2007, 606 p. ISBN 978-0-387-33333-5.
14. *Wakelin P., Day S., Read S., McKenna F.* CICS transaction gateway V3.1: the WebSphere connector for CICS. SG24-6133-00, IBM, 2001. 269 p.
15. <http://www.techproresearch.com/>, <http://22century.ru/popular-science-publications/gartner-preliminary-results/attachment/bigdata>
16. *Medetov A.A.* Termin Big Data i sposoby ego primenenija // Molodoj uchenyj. 2016. No 11. S. 207–210.
17. Jera Big Data. Kak ritejlery ispol'zujut dannye o svoih klientah. <http://retail-community.com.ua/news/tehnologii/3677-era-big-data-kak-riteyleryi-ispolzuyut-dannye-o-svoih-klientah>
18. Predskazanie ottoka igrokov iz World of Tanks ot Yandex Data Factory. Lekcija dlja Malogo ShADa. <https://habrahabr.ru/company/yandex/blog/281777/>
19. *Sergej Nemalovich.* Gadanie na bol'shih dannyh. <https://republic.ru/biz/1152340/>
20. *Andrej Vasil'kov.* Data Collective: «bol'shie dannye» – prioritnoe napravlenie investicij. <http://www.computerra.ru/97457/data-collective-relies-on-big-data/>

21. Wikibon Big Data Capital Markets Day 2014, Jeff Kelly. <http://www.slideshare.net/jeffkelly568/wikibon-big-data-capital-markets-day-2014>

22. Nebol'shie vyvody o bol'shikh dannyh, General'nyj direktor T-Systems v Ros-sii. Aleksej Toskin, 19 avgusta 2014. <http://www.computerra.ru/105278/nebolshie-vyvodyi-o-bolshih-dannyih/>

23. 10 luchshih produktov na rynke Big Data v 2015 godu. <https://www.crn.ru/news/detail.php?ID=107118>

24. *Izmalkova S.A., Golovina T.A.* Ispol'zovanie global'nyh tehnologij «big data» v upravlenii jekonomicheskimi sistemami //Izv. Tul'skogo gosudar-stvennogo universi-teta. Jekonomicheskie i juridicheskie nauki. 2015. No 4-1. S. 151–158.

25. Bol'shie danye. Bol'shie vozmozhnosti. <http://www.emc.com/ru-ru/big-data/index.htm>

26. Trendy mirovogo e-commerce rynka v 2015–2016 godah. https://habr-habr.ru/company/SECL_GROUP/blog/268455/

27. *Semenov Ju.A.* Obzor po materialam vedushhikh firm, rabotajushhikh v sfere setевой bezopasnosti (ITJeF-MFTI) <http://book.itep.ru/10/2015.htm>

СВЕДЕНИЯ ОБ АВТОРАХ



БИКТИМИРОВ Марат Рамилевич – кандидат технических наук, ВРИО директора Всероссийского института научной и технической информации Российской академии наук (ВИНИТИ РАН), г. Москва

Marat Ramilevich BIKTIMIROV – Candidate of Engineering Sciences, Acting Director of Russian Institute for Scientific and Technical Information (VINITI RAS), Moscow, Russia.

Current scientific interests: system analysis, data mining, cloud computing, information security.

e-mail: marat@ras.ru



ЕЛИЗАРОВ Александр Михайлович – доктор физико-математических наук, профессор, заслуженный деятель науки Республики Татарстан, заведующий кафедрой дифференциальных уравнений Института математики и механики им. Н.И. Лобачевского Казанского (Приволжского) федерального университета.

Alexander Mikhailovich ELIZAROV – Doctor of Physics and Mathematics, Professor, Honoured Worker of Science of the Republic of Tatarstan, Head of Chair “Differential Equations” of the Lobachevskii Institute of Mathematics and Mechanics, Kazan Federal University, Kazan, Russia.

Current scientific interests: data mining, recommender systems, cloud computing, knowledge extraction technologies.

email: amelizarov@gmail.com



ЩЕРБАКОВ Андрей Юрьевич – доктор технических наук, профессор НИУ «Высшая школа экономики», главный научный сотрудник Федерального исследовательского центра «Информатика и управление» РАН, г. Москва

SHCHERBAKOV Andrey Yurievich – Doctor of technical Science, Professor of National Research University "Higher school of Economics", chief researcher, Federal Research Center "Information and Management" RAS, Moscow, Russia

Current scientific interests: information security, cryptography, data mining, cloud computing.

e-mail: x509@ras.ru

Материал поступил в редакцию 4 октября 2016 года