

АВТОМАТИКА. ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА

МРНТИ 73.29.75, 73.29.81

Т.К.Жукабаева, А.Т.Кусаинова

Евразийский национальный университет им. Л.Гумилёва,
г. Астана, Казахстан

ТЕХНОЛОГИЯ БОЛЬШИХ ДАННЫХ (BIG DATA). ОСНОВНЫЕ ХАРАКТЕРИСТИКИ И ПЕРСПЕКТИВЫ ПРИМЕНЕНИЯ

Аннотация. В работе представлены ключевые понятия технологии Больших данных: основные характеристики, методы, этапы технологии, сферы применения технологии. Произведен анализ мировых достижений в данной области. Приведены примеры использования в работах зарубежных авторов. Проанализирован современный рынок использования технологии Больших данных. Выполнен сравнительный анализ с СУБД. В рамках проведенных исследований разработана функциональная блок-схема технологии Больших объемов данных.

Ключевые слова: технология Больших данных, анализ данных, структурированные данные, неструктурированные данные.



Түйіндеме. Біздің жұмысымызда Үлкен деректер технологиясының негізгі ұғымдары, кезеңдері, әдістері, сипаттамалары келтірілген. Аталған саладағы әлемдік жетістіктерге талдау жасалып, шетелдік авторлардың мысалдары ұсынылған. Қазіргі заманғы нарықта үлкен деректерің қолдану аймағына талдау жасалған. Зерттеу барысында Үлкен деректер технологиясына функционалды блок-сызбасы жасалды.

Түйінді сөздер: үлкен деректер, деректерді талдау, құрылымдалған деректер, құрылымсыз деректер.



Abstract: In our work key concepts of Technology Big Data: basic characteristics, methods, stages of technology, scope of Big Data have been presented. The analysis of the world achievements in this field has been done, examples of foreign authors have been demonstrated. The current market has

been analyzed. Comparative analysis of DBMS has been done. In the research the functional flowchart of Big Data has been developed.

Key words: Big data, data analysis, structured data, unstructured data.

Введение. На сегодняшний день одним из активно развивающихся направлений в области информационных технологий является технология Больших данных (Big Data). В последние годы Большие данные являются общепризнанным признаком экономического и технологического развития. Исследования консалтинговой компании «Gartner» прогнозируют, что технология Больших данных окажет существенное влияние на информационные технологии в производстве, здравоохранении, торговле, государственном управлении и в других отраслях, которые используют большой поток информации. В Казахстане на период до 2030 г. исследования технологии Больших данных для управления производственными и социальными процессами, наряду с обработкой неструктурированных данных, признано приоритетным направлением в области информационных технологий [1].

Общая характеристика технологии Больших данных

Понятие «технология Больших данных» в сфере информационных технологий, которое появилось недавно, ввел Клиффорд Линч [2]. Приведем самые распространенные определения. Большие данные – это:

– группа технологий и методов производительной обработки динамически растущих объемов данных в информационных технологиях [3];

– серия подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объемов и значительного многообразия для получения воспринимаемых человеком результатов, эффективных в условиях непрерывного прироста, распределения по многочисленным узлам вычислительной сети [4].

Большие данные классифицируются в зависимости от типа источника данных (Интернет или медиа-источники, сгенерированные машинные данные), по формату контента (структурированные, полуструктурированные, неструктурированные), по спо-



Рис. 1. Классификация технологии Больших данных

субу хранения данных и т. д. Для наглядности на рис. 1 представлена общая классификация технологии Больших данных [5].

Основные признаки технологии Больших данных. Впервые в 2001 г. признаки «Три V» выделил ведущий аналитик Gartner Дуг Лани в [6], а именно разнообразие, объем, скорость, которые являются основными характеристиками технологии:

разнообразие – помогает эффективно проводить анализ данных;

объем – (самое важное из свойств технологии), т. е. количество сгенерированных данных определяют значение и потенциал исследования (рис. 2);

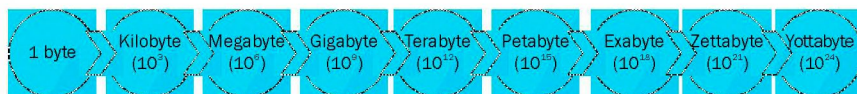


Рис. 2. Объем данных, используемых в технологии Больших данных

скорость – т. е. с какой скоростью данные генерируются и обрабатываются в соответствии с требованиями.

Зикопулоус [7] предложил добавить еще 2 признака – стоимость и достоверность (Veracity and Value), таким образом получив «5V»:

Стоимость – признак технологии, описывающий экономический эффект, который технология обеспечивает пользователям.

Достоверность – качество собранных данных, которые могут значительно различаться. Точный анализ зависит от достоверности исходных данных.

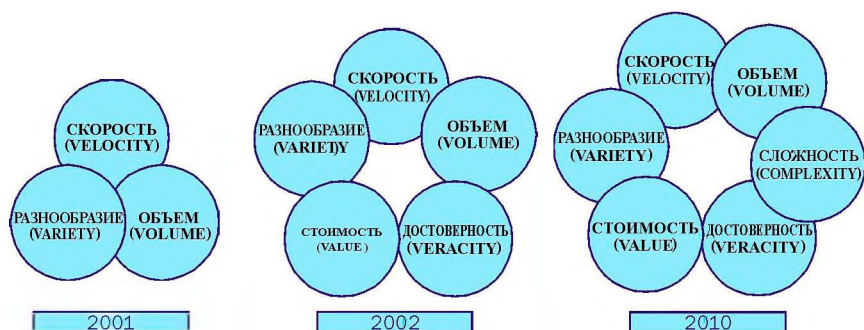


Рис. 3. Основные признаки технологии Больших данных

Позже к признакам добавили такое понятие, как «сложность», поскольку управление данными может быть очень сложным, особенно при обработке больших объемов данных, которые приходят из разных источников (рис. 3) Средства, используемые для сбора и обработки Больших данных, можно разделить на несколько групп (рис. 4): программное обеспечение, оборудование, сервисные услуги. Например, MapReduce – модель распределения вычислений, NoSQL (NotOnly SQL) – это совокупность подходов, направленных на реализацию базы данных, имеющих отличия от моделей, используемых в традиционных, реляционных СУБД. Hadoop – используется для реализации поисковых и контекстных механизмов высоконагруженных сайтов Facebook, eBay, Amazon и др.

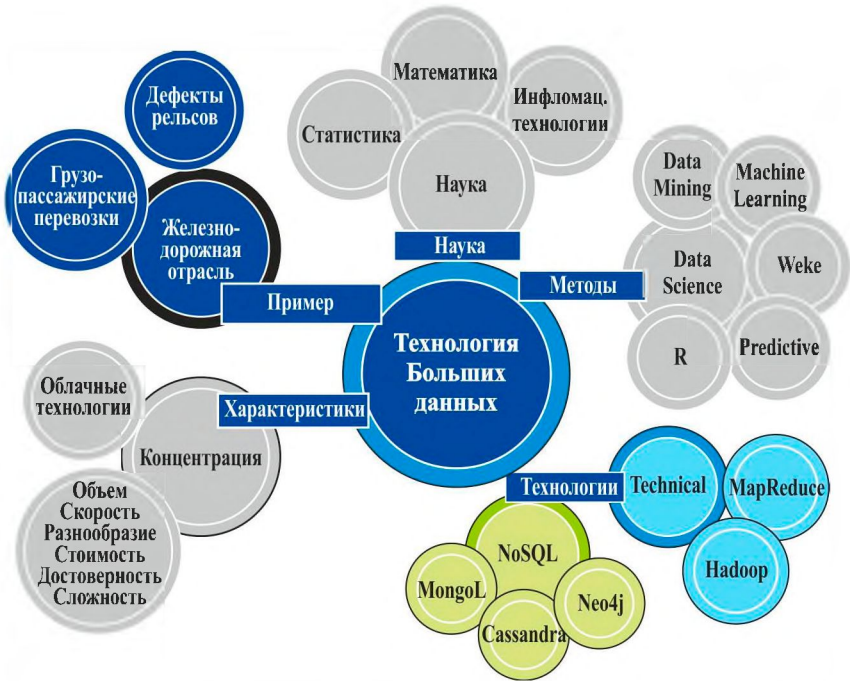


Рис. 4. Основные характеристики, технологии, методы, примеры использования технологии Больших данных

Методами анализа, применимыми к технологии Больших данных, являются методы класса DataMining: обучение ассоциативным правилам, машинное обучение, статистический анализ, R (язык программирования для статистической обработки данных и работы с графикой), Weka (Waikato Environment for Knowledge Analysis, свободное программное обеспечение для анализа данных), Predictiveanalytics (прогнозная аналитика, предиктивная аналитика). В таблице представлены методы анализа данных, применимые к технологии Больших данных, отсортированные по годам появления.

**Методы анализа данных, применимые к технологии
Больших данных**

Год	Название	Описание
2014	BDAF	Архитектура Больших данных [8]
2014	HACE	Анализ данных [9]
2014	Storm	Параллельные вычисления [10]
2014	Predictive	Прогнозная аналитика[11]
2013	MRAM	Мобильные агенты [12]
2013	CBDMASP	Статистический анализ данных [13]
2013	SODSS	Система принятий решения [14]
2013	MLPACK	Библиотека машинного обучения [15]
2013	NoSQL	Реализация БД [16]
2012	Starfish	Самообучаемая аналитическая система [17]
2012	ODT MDC	Безопасность [18]
2012	MLAS	Машинное обучение [19]
2012	PIMRU	Машинное обучение [20]
2012	R	Язык программирования для обработки данных [21]
2011	DOT	Фреймворк [22]
2011	GLADE	Многоуровневая системная архитектура [23]
2011	Hadoop	Параллельные вычисления [24]
2011	Mahout	Машинное обучение [25]
2011	Radoop	Анализ данных, машинное обучение [26]
2011	Weka	ПО для анализа данных[27]
2010	Pregel	Графический анализ данных [28]
2007	CUDA	Параллельные вычисления [29]
2004 (2007)	MapReduce	Модель распределения вычислений [30]

Более подробное описание методов анализа, применимых к технологии Больших данных, представлены в работе [31]. В рамках исследования были изучены основные характеристики, классификация технологии Больших данных, а также выделены основные технологии и методы анализа, применимые к технологии.

«Большие данные в числах»

По данным из «Аналитического обзора рынка BigData», активное развитие технологии Больших данных получили в телекоммуникационных предприятиях (58 %) и инжиниринге

(35 %) [32]. Также технология Больших данных активно внедряется в зарубежных компаниях всех отраслей. Такие компании, как Nasdaq, Facebook, Google, IBM, VISA, MasterCard, BankofAmerica, HSBC, AT&T, CocaCola, Starbucks и Netflix, уже используют ресурсы Больших данных [32].

В 2014 г. Большие данные, по мнению DataCollective, стали одними из приоритетных направлений инвестирования в сфере венчурной индустрии. Только за 2014 г. количество компаний с реализованными проектами в сфере управления большими данными увеличилось на 125 %. Объем рынка вырос на 45 % по сравнению с 2013 г. [32].

По данным сайта электронного правительства Республики Казахстан, с целью изучения и практического применения технологий Больших данных в Казахстане на базе АО «Национальные информационные технологии» создана лаборатория Больших данных. Сегодня в лаборатории уже ведутся исследовательские работы по анализу профиля пользователя республиканского портала eGov. Выявлено, что основными источниками данных являются портал электронного правительства eGov, центры обслуживания населения (ЦОН), электронные обращения граждан, СМС-сообщения, звонки в Единый контакт-центр, социальные сети. Например, в настоящее время портал электронного правительства eGov содержит около 15 ТБайт информации, которая, в свою очередь, ежемесячно пополняется на 1 ТБайт. Кроме этого, в планах лаборатории Больших данных обрабатывать не менее 3 ТБайтов данных из других открытых источников [33]. Мировые достижения «в числах» показывают, что технология Больших данных активно развивается и уже достигла положительных результатов.

Применение технологии Больших данных в производстве

К 2015 г. несмотря на небольшой срок существования технологии Больших данных, уже есть оценки эффективного использования на реальных примерах. Один из самых высоких показателей относится к энергетике. Так, по оценкам аналитиков, аналитические технологии BigData способны на 99 % повысить точ-

ность распределения мощностей генераторов [34]. Применение технологии Больших данных в различных отраслях [35, 36]:

- энергетика (влияние погоды на генерацию энергии), анализ данных от «умных» счетчиков, исследовательские инфраструктуры для эффективного использования энергии в зданиях;

- наука (Большого Адронного коллайдера, пан-Европейская инфраструктура для оценки качества при тестировании наноматериалов, антипреступная и антикоррупционная обсерватория);

- E Commerce (анализ поведения и покупательских моделей, интеграция каналов взаимодействия, моделирование поведения клиентов);

- транспорт (влияние погоды и трафика на доставку и потребление топлива);

- колл-центр (анализ расшифровок разговоров для понимания поведения клиентов);

- финансы (решения по рискам, анализ мнения клиентов, борьба с отмыванием денег);

- ИТ (анализ логов от разных транзакционных систем);

- Телеком (анализ операций и сбоях сети).

Примеры в железнодорожной индустрии

В сфере железнодорожных компаний лидирующее место занимает Union Pacific Railroad, которая оказалась в десятке лучших после внедрения технологии Больших данных. Современный пакет датчиков и обработка огромного потока данных от них методами технологии снизили частоту схода с рельсов на 75 %, или около 40 млн. дол. [37]. В работе [38] (рис. 5) представлен пример использования технологии Больших данных в железнодорожной индустрии. Авторы выделили основные этапы работы, методы, применимые при анализе Больших данных. Выделенные этапы работы: методы анализа, разработка алгоритма, теория/постановка задачи, исследование/реализация. На каждом этапе предложены свои методы решения поставленных задач, например, тензорный анализ или применение графических моделей.



Рис. 5. Пример использования технологии Больших данных в железнодорожной индустрии

Примеры в здравоохранении и биомедицинской информатики

Технология Больших данных расширяет возможности предприятий здравоохранения. Например, в больнице «скорой медицинской помощи» Colchester, принадлежащей системе NHS, на следующий год после внедрения технологии Больших данных летальность снизилась на 158 случаев [35]. Сегодня эта технология дает возможность не только собирать, хранить, анализировать всю информацию о состоянии пациентов в режиме реального времени, но и помогает медицинским работникам при-

нимать сложные решения. Также технология Больших данных активно используется в биомедицинской информатике [39]. Авторы обращают внимание, что достижения в «-omics», устройств визуализации, и других передовых биомедицинских технологий в последние годы способствовали экспоненциальному росту биомедицинских данных, и началась эра Больших данных в области здравоохранения и биомедицины. Сердечно-сосудистые заболевания являются областью, где эффективное использование аналитики Больших данных открывает огромные перспективы для оценки рисков и качественного лечения.

Подготовка кадров в сфере Больших данных

Мировые научные сообщества активно занимаются подготовкой специалистов в области Больших данных. Семь университетов Китая готовят специалистов для работы с Большими данными. В перспективе до конца 2015 г. численность молодых специалистов должна составить 40 тыс. чел. Создаются различные университетские программы (University of Washington: Certificate in Data Science, New York University: Data Science at NYU, University of Southern California (UCS): Master of Science in Data Science), он-лайн курсы обучения (Udacity, Coursera), Индустриальные конференции и выставки (Big Data Techcon, Big Data Innovation summits) [40, 43]. Так, в России активно развиваются магистерские программы по данной тематике [44, 45]: «Системы Больших данных», «Интеллектуальные системы обработки больших данных». Особенностью данных программ является ее ориентация на потребности бизнеса в новой технологии. Программа концентрирует свое внимание на деятельности предприятия как на системе с развитой информационной инфраструктурой, обеспечивающей автоматизацию решения управленческих задач. Авторами в 2015 г. в Евразийском национальном университете на кафедре «Вычислительная техника» были введены дисциплины магистратуры и докторантуры: «Интеллектуальный анализ данных», «Работа с данными и аналитика Больших данных».

На сегодняшний день технологию Больших данных активно внедряют в производство, науку и другие отрасли во всем мире. По достигнутым результатам можно сделать вывод, что внедре-

ние новых информационных технологий, в частности технологии Больших данных, будет способствовать не только уменьшению затрат на обработку данных предприятий, но и улучшению качества работы в сфере их деятельности.

Сравнение технологии Больших данных с системой управления базами данных

Традиционные архитектуры СУБД описываются следующими понятиями: дисковое хранение, транзакции, индексация, журнализация, многопоточность, блокировка. Производительность самых мощных серверов растет медленнее, чем объемы данных. Выделены некоторые недостатки традиционной архитектуры СУБД:

- оптимизация и индексы требуют денежных расходов на дополнительные операции, но при этом наблюдается также и сильное падение скорости при нагрузках;
- денормализация схемы вызывает избыточность данных;
- очередность операций создает проблему хранения и ограниченности очереди [46] (рис. 6).



Рис. 6. Компоненты традиционной архитектуры СУБД

В свою очередь, применение технологии Больших данных обеспечивает согласованность данных, невзирая на производительность, небольшие трудозатраты на развертывание больших кластеров и многое другое. На рис. 7 показано изменение традиционной архитектуры путем внедрения технологии Больших данных.

С помощью Google Trends, который является публичным web-приложением корпорации «Google», основанным на поиске Google, произведено сравнение популярности запросов «Боль-



Рис. 7. Развитие архитектуры с помощью технологии Больших данных

шие данные» и «СУБД». В итоге получен график «Динамика популярности», который показывает, как часто термины «Большие данные» и «СУБД» ищут по отношению к общему объему поисковых запросов в различных регионах мира и на различных языках. На рис. 8 представлен итоговый график популярности терминов «Большие данные» и «СУБД» за 2004-2015 гг.

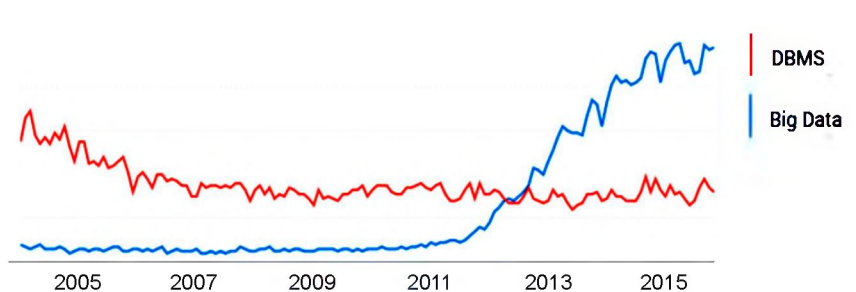


Рис. 8. Динамика популярности терминов «Большие данные» и «СУБД»

Это может свидетельствовать о том, что использование технологии Больших данных является большим шагом вперед в информационных технологиях, который позволяет оптимизировать работу пользователей, получить более качественный уровень работы.

Функциональная блок-схема технологии Больших объемов данных

В ходе изучения основных свойств, признаков и этапов технологии Больших данных была разработана функциональная блок-схема работы технологии Больших объемов данных [47-51] с учетом изученных примеров. На функциональной блок-схеме (рис. 9) детально рассмотрены этапы технологии: поиск источника Больших данных, манипулирование и хранение данных, анализ, потребление. Также следует отметить, что потребителями данных могут быть люди, бизнес-процессы, корпоративные организации, некие группы людей.



Рис. 9. Функциональная блок-схема технологии Больших объемов данных

Разработанная функциональная блок-схема технологии Больших объемов данных наглядно показывает основные этапы с учетом характеристик технологии.

Заключение. Внедрение технологии Больших данных является новым шагом в научно-техническом развитии в различных областях общества. Несмотря на то, что технология Больших данных находится сейчас на начальном этапе развития, она уже достигла высоких показателей в области информационных технологий. Данная технология активно применяется в различных сферах деятельности, в частности в железнодорожной отрасли. Представлен подробный анализ технологии Больших данных, а также основные признаки, характеристики, методы данной технологии. Приведены примеры успешного внедрения технологии в международных организациях железнодорожной индустрии, здравоохранении. Следует отметить, что с учетом результатов наших исследований изучение и внедрение технологии Больших данных в IT-индустрии Казахстана необходимо проводить на основе опыта зарубежных коллег. Перспективы дальнейшего исследования проблемы обусловлены более детальным изучением существующих алгоритмов технологии Больших данных.

Список литературы

1. Синергия форум объединяя науку, бизнес и общество. Результаты проекта «Системный анализ и прогнозирование в сфере науки и технологий» по направлению «Информационные и телекоммуникационные технологии» в Казахстане на период до 2030 г. – Алматы, 2014. – 219 с.
2. *Черняк Л.* Большие данные – новая теория и практика // Открытые системы. СУБД. – М.: Открытые системы, 2011. – № 10. – С. 18-25.
3. *Алексеев М.* Big Data – революция в области хранения и обработки данных, 2014 г. – [Электронный ресурс]: www.slideshare.net
4. *Иванов П.Д., Вампилова В.Ж.* Технологии Big Data и их применение на современном промышленном предприятии // Электронное

научно-техническое издание «Инженерный журнал: наука и инновации». – М.: МГТУ. – 2014. – № 8 (32). – 10 с.

5. Ibrahim Abaker Targio Hashem, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, Samee Ullah Khan. The rise of «big data» on cloud computing: Review and open research issues. *Information Systems* 47, 2015. – P. 98-115.

6. Laney D (2001) 3d data management: Controlling data volume, velocity and variety. Technical Report 949, META Group (now Gartner). [Электронный ресурс]: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-ManagementControlling-Data-Volume-Velocity-and-Variety.pdf>

7. Zikopoulos P, Parasuraman K, Deutsch T, Giles J, Corrigan D (2012) Harness the power of big data The IBM big data platform. McGraw Hill Professional, New York, NY. – [Электронный ресурс]: <http://books.google.com/books?id=HhSON0xOCQ0C>

8. Demchenko Y, de Laat C, Membrey P. Defining architecture components of the big data ecosystem. In: *Proceedings of the International Conference on Collaboration Technologies and Systems*, 2014. – P. 104-112.

9. Wu X, Zhu X, Wu GQ, Ding W. Data mining with big data. *IEEE Trans Knowl Data Eng.* – 2014. – № 26 (1). – P. 97-107.

10. Apache Storm, February 2, 2015. [Online]. Available: UR. – [Электронный ресурс]: <http://storm.apache.org>

11. Finlay, Steven *Predictive Analytics, Data Mining and Big Data. Myths, Misconceptions and Methods.* Basingstoke: Palgrave Macmillan. – 2014. – P. 15.

12. Essa YM, Attiya G, El Sayed A. Mobile agent based new framework for improving big data analysis. In: *Proceedings of the International Conference on Cloud Computing and Big Data*, 2013. – P. 381-386.

13. Ye F, Wang ZJ, Zhou FC, Wang YP, Zhou YC. Cloud based big data mining and analyzing services platform integrating r. In: *Proceedings of the International Conference on Advanced Cloud and Big Data*, 2013. – P. 147-151.

14. Demirkan H, Delen D. Leveraging the capabilities of service oriented decision support systems: putting analytics and big data in cloud. *Decision Support Syst.* – 2013. – P. 412-421.

15. Curtin RR, Cline JR, Slagle NP, March WB, Ram P, Mehta NA, Gray AG. MLPACK: a scalable C++ machine learning library // *J Mach Learn Res.* – 2013. – № 14. – 801 p.

16. Фаулер М, Садаладж П.Дж. NoSQL: новая методология разработки нереляционных баз данных – NoSQLDistilled. – М.: «Вильямс», 2013. – 192 с.

17. Wonner J, Grosjean J, Capobianco A, Bechmann D Starfish: a selection technique for dense virtual environments. In: Proceedings of the ACM Symposium on Virtual Reality Software and Technology, 2012. – P. 101-104.

18. Laurila JK, Gatica Perez D, Aad I, Blom J, Bornet O, Do T, Dousse O, Eberle J, Miettinen M. The mobile data challenge: big data for mobile computing research. In: Proceedings of the Mobile Data Challenge by Nokia Workshop, 2012. – P. 1-8.

19. Bu Y, Borkar VR, Carey MJ, Rosen J, Polyzotis N, Condie T, Weimer M, Ramakrishnan R. Scaling datalog for machine learning on big data, CoRR, vol. abs/1203.0160, 2012. – [Электронный ресурс]: <http://dblp.uni.trier.de/db/journals/corr/corr1203.html#abs.1203.0160>.

20. Bu Y, Borkar VR, Carey MJ, Rosen J, Polyzotis N, Condie T, Weimer M, Ramakrishnan R. Scaling datalog for machine learning on big data, CoRR, vol. abs/1203.0160, 2012. – 29 p.

21. Robert A. Muenchen (2012). The Popularity of Data Analysis Software. – [Электронный ресурс]: www.r-bloggers.com

22. Huai Y, Lee R, Zhang S, Xia CH, Zhang X. DOT: a matrix model for analyzing, optimizing and deploying software for big data analytics in distributed systems. In: Proceedings of the ACM Symposium on Cloud Computing, 2011. – P. 4-14.

23. Rusu F, Dobra A. GLADE: a scalable framework for efficient analytics. In: Proceedings of LADIS Workshop held in conjunction with VLDB, 2012. – P. 1-6.

24. Apache Hadoop, February 2, 2015. – [Электронный ресурс]: [Online]. Available: <http://hadoop.apache.org>.

25. Apache Mahout, February 2, 2015. – [Электронный ресурс]: [Online]. Available: <http://mahout.apache.org/>

26. Hasan S, Shamsuddin S, Lopes N. Soft computing methods for big data problems. In: Proceedings of the Symposium on GPU Computing and Applications, 2013. – P. 235-247.

27. Ian H. Witten; Eibe Frank; Mark A. Hall (2011). Data Mining: Practical machine learning tools and techniques, 3rd Edition. MorganKaufmann, SanFrancisco. Retrieved, 2011. – P. 1-19.

28. Malewicz G, Austern MH, Bik AJ, Dehnert JC, Horn I, Leiser N, Czajkowski G. Pregel: A system for large scale graph processing. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, 2010. – P. 135-146.

29. Nvidia, Cuda, February 2, 2015. [Online]. Available: [Электронный ресурс]: URL: http://www.nvidia.com/object/cuda_home_new.html

30. MapReduce: Simplified Data Processing on Large Clusters Jeffrey Dean and Sanjay Ghemaw, OSDI, 2004, – [Электронный ресурс]: To appearing. – 13 p.

31. Chun Wei Tsai, Chin Feng Lai, Han Chieh Chao, Athanasios V. Vasilakos. Big data analytics: a survey // Journal of Big Data. – 2015. – Т. 2. – 32 p.

32. Булавин Д., Чикунов А., Будушкаева С. Аналитический обзор рынка Больших данных. – М.: Борд, 2015. – 51 с.

33. Сайт электронного Правительства Республики Казахстан. – 2015. – [Электронный ресурс]: egov.kz

34. Обзор: Бизнес-аналитика и большие данные в России. – 2014. – [Электронный ресурс]: complex.imex.ru

35. Технологии «больших данных» в важнейших отраслях. CITO Research при поддержке QlikView, 2014 г. – [Электронный ресурс]: www.rbcgr.com – 11 p.

36. Manyika, James et al. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute, June, 2011. – 156 p.

37. Официальный сайт компании UnionPacificRailroad. [Электронный ресурс]: <http://www.up.com/>

38. Allan M Zarembski Some Examples of Big Data in Railroad Engineering // IEEE International Conference on Big Data, 2014. – 15 p.

39. Po-Yen Wu, Janani Venugopalan, Sonal Kothari, ChanchalaKaddi, Chih-Wen Cheng, John H. Phan, and May D. Wang Big Data Analytics – Biomedical and Health Informatics for Personalized Cardiovascular Disease Care. – [Электронный ресурс]: <http://lifesciences.ieee.org/publications/newsletter/april-2014/539-big-data-analytics-biomedical-and-health-informatics-for-personalized-cardiovascular-disease-care>

40. University of Washington / Tacoma Master of Science in Computer Systems with Concentration in Big Data Management. University of Washington, 2015. – [Электронный ресурс]: www.tacoma.uw.edu

41. Data Science – USC – University of Southern California, 2015. – 15 p.
42. Bellevue colleges Healthcare Data Analyst Certificate. Bellevue College Continuing Education. Washington, 2014. – [Электронный ресурс]: www.bellevuecollege.edu
43. Жуков Л. Профессия Data Scientist: конференция «Большие данные в национальной экономике». – М., 2013. – 20 с.
44. Национальный исследовательский университет Высшая школа экономики. Магистерская программа «Системы больших данных». – [Электронный ресурс]: <http://www.hse.ru/ma/bigdata/moduls>
45. Московский физико-технический институт. – [Электронный ресурс]: <https://mipt.ru/education/chairs/da/education/masters/>
46. Вовченко А. Управление разнотипными Большими данными, ВМК МГУ ИПИ РАН, 2014. – 10 с.
47. Divakar Mysore, Shrikant Khupat, Shweta Jain. Big data architecture and patterns, Part 1: Introduction to big data classification and architecture // IBM Corporation, 2013. – [Электронный ресурс]: www.ibm.com
48. Divakar Mysore, Shrikant Khupat, Shweta Jain. Big data architecture and patterns, Part 2: How to know if a big data solution is right for your organization // IBM Corporation, 2013. – 6 p. – [Электронный ресурс]: www.ibm.com
49. Divakar Mysore, Shrikant Khupat, Shweta Jain. Big data architecture and patterns Big data architecture and patterns, Part 3: Understanding the architectural layers of a big data solution// IBM Corporation, 2013. – [Электронный ресурс]: www.ibm.com
50. Divakar Mysore, Shrikant Khupat, Shweta Jain. Big data architecture and patterns, Part 4: Understanding atomic and composite patterns for big data solutions // IBM Corporation, 2013. – [Электронный ресурс]: www.ibm.com
51. Divakar Mysore, Shrikant Khupat, Shweta Jain. Big data architecture and patterns, Part 5: Apply a solution pattern to your big data problem and choose the products to implement it // IBM Corporation, 2013. – [Электронный ресурс]: www.ibm.com

Жукабаева Т.К., PhD доктор, и.о. доцента кафедры «Вычислительная техника», 8-701-726-85-44, e-mail: zhukabaeva_tk@enu.kz

Кусаинова А.Т., PhD докторант кафедры «Вычислительная техника», 8-705-505-50-67, e-mail: ainurkussainova89@gmail.com