

*А. О. Утепбергенов*

*Магистрант Санкт-Петербургского государственного  
университета,*

*Кафедра «Управления персоналом»*

## **БОЛЬШИЕ ДАННЫЕ В УПРАВЛЕНИИ ЧЕЛОВЕЧЕСКИМИ РЕСУРСАМИ**

**Абстракт:** Проведен анализ перспектив развития BIG DATA в управлении персоналом, основанных на технологиях анализа больших данных. Показана возможность применения технологий Big Data для повышения производительности управления персоналом. Сформулированы рекомендации службам управления человеческими ресурсами по использованию технологий Big Data в управлении персоналом инновационного развития.

**Ключевые слова:** технологии Big Data, инновации, платформы, сфера применения, риски и перспективы применения больших данных.

**A. O. Utepbergenov**

*Saint-Petersburg State University, Department of Human Resource  
Management*

## **BIG DATA IN HUMAN RESOURCE MANAGEMENT**

**Abstract:** An analysis of the prospects of development of BIG DATA in human resource management, based on big data analysis technologies. Shows the possibility of using Big Data technologies to improve human resource management productivity. Formulated recommendations for human resource management services for the use of Big Data technologies in the management of innovative personnel development.

*Key words: Big Data technologies, innovation, platforms, scope, risks and prospects of application of big data.*

Большие данные стали ценным ресурсом для компаний, это означает новый подход и методологию для управления данными о сотрудниках и многочисленные возможности для управления персоналом. Сфера управления персоналом на основе анализа больших данных несет новые возможности для более эффективного найма, управления талантами, сохранения и оптимизации персоналов. В эпохе цифровизации организациям необходимо применять современные технологии, чтобы оставаться успешными.

Теоретики HRM рассматривали совокупности практик управления персоналом либо по отдельности, либо как системы (Combs, Liu, Hall, & Ketchen, 2006), разделили на двух раздел. Первый «HR-исследования и практика» объединяет статьи, в которых в общих чертах обсуждается использование BD в HRM. Второй включают высокоэффективные системы практики работы (Huselid, 1995), в которых были применены подходы BD. Это отбор и найм; оценка и развитие; информация, обучение и знания; и стратегия, эффективность и результативность.

Структурированные данные имеют заранее определенный формат. Полу структурированные или слабоструктурированные данные – это данные, зачастую собранные из различных источников. Структура данных документирована, но в зависимости от источника данных конкретный формат представления информации может быть разным[1].

Поскольку термин англоязычный, логично начать обзор с авторитетных англоязычных словарей и ресурсов. В июне 2013 г. Оксфордский английский словарь добавил определение термина «big data» [2], которое можно перевести так:

«Данные очень большого размера, как правило, в том смысле, что представляют серьезные трудности в материально-техническом обеспечении по манипуляциям и управлению ими; направление вычислений с использованием такого типа данных» [3].

Что интересно, сайт [oxforddictionaries.com](http://oxforddictionaries.com), который фокусируется на текущих, актуальных значениях слов и практическом употреблении, дает следующее определение «big data»: «Очень большие массивы данных, которые могут быть проанализированы с помощью компьютеров, чтобы выявить закономерности, тенденции и взаимосвязи, в особенности в отношении поведения людей и их взаимодействий» [4]. Технологический термин в данной интерпретации тесно связан с социально-гуманитарной компонентой, т. е. по отношению к первоначальному определению наблюдается эволюция понятия в сторону его гуманизации, а содержание понятия меняется от «серьезных трудностей» к практическим возможностям анализа человеческой деятельности.

В Российском сегменте интернета популярна версия, что термин «big data» ввел в научный оборот редактор научного журнала «Nature» Клиффорд Линч в специальном номере от 3 сентября 2008 г., посвященном драматическому росту объемов информации [5]. В него он отнес любые массивы неоднородных данных свыше 150 Гб в сутки [6]. Эта версия появилась и воспроизводится в статьях (в том числе научных) [7], [8] и бизнес-презентациях с подачи издания «Открытые системы. СУБД» с 2011 г [9]. Распространению информации также могло послужить включение этой версии в русскоязычную статью Википедии о «больших данных» [10]. Однако даже самый поверхностный анализ отсылает к более ранним референциям. Так, раз уж упомянута русская версия Википедии, ее англоязычная версия отсылает к исследователю Джону Машей,

специалисту в области информатики (John Mashey, computerscientist), который популяризовал термин в 1990-х гг[11].

Эксперты по цифровым технологиям и экономике сходятся во мнении, что в ближайшие десятилетия на рынок труда существенно повлияет цифровизация. По оценкам Глобального института McKinsey, в мире к 2036 году будет автоматизировано до 50% рабочих процессов. Это приведет к значительному высвобождению персонала, сокращению количества рабочих мест, требующих средней квалификации, и увеличению разницы в уровнях оплаты труда [12].

Изначально в совокупность подходов и технологий включались средства массово-параллельной обработки неопределённо структурированных данных, такие как СУБД NoSQL, алгоритмы MapReduce и средства проекта Hadoop.

MapReduce — модель распределённых параллельных вычислений в компьютерных кластерах, представленная компанией Google. Согласно этой модели приложение разделяется на большое количество одинаковых элементарных заданий, выполняемых на узлах кластера и затем естественным образом сводимых в конечный результат.

NoSQL (от англ. NotOnly SQL, не только SQL) — общий термин для различных нереляционных баз данных и хранилищ, не обозначает какую-либо одну конкретную технологию или продукт.

Hadoop — свободно распространяемый набор утилит, библиотек и фреймворк для разработки и выполнения распределённых программ, работающих на кластерах из сотен и тысяч узлов. Считается одной из основополагающих технологий больших данных.

Обычно BIG DATA описываются при помощи следующих характеристик[13]:

1. Объем (Volume) – количество сгенерированных и хранящихся данных. Размер данных определяет значимость и потенциал данных, а также то, могут ли они быть рассмотрены как Большие данные.

2. Разнообразие (Variety) – тип данных. Большие данные могут состоять из текста, изображений, аудио, видео.

3. Скорость (Velocity) – скорость. Здесь подразумевается скорость, с которой данные генерируются и обрабатываются.

4. Изменчивость (Variability) – противоречивость наборов данных может препятствовать их обработке и управлению ими.

5. Достоверность (Veracity) – качество данных напрямую влияет на точность проведения анализа данных.

#### Методы анализа Big Data:

– Кластерный анализ – статистический метод классификации объектов, – который приводит к разделению разнообразных групп на более мелкие группы подобных объектов.

– Краудсорсинг – метод сбора, категоризация и обогащение данных силами широкого круга лиц, привлечённых на основании публичной оферты, без вступления в трудовые отношения, обычно посредством использования сетевых медиа.

– Машинное обучение (machine learning). Класс методов искусственного интеллекта, характерной чертой которых является не прямое решение задачи, а обучение в процессе применения решений множества сходных задач.

– Математическая модель, построенная по принципу организации и функционирования биологических нейронных сетей – сетей нервных клеток живого организма.

– Сетевой анализ (networkanalysis). Набор методов, используемых для описания и анализа отношений между дискретными узлами в графе или сети.

– Прогнозная аналитика (predictive analytics). Класс методов анализа данных, концентрирующийся на прогнозировании будущего поведения объектов и субъектов с целью принятия оптимальных решений.

– Имитационное моделирование. (simulation modeling) – метод исследования, при котором изучаемая система заменяется моделью, с достаточной точностью описывающей реальную систему, с которой проводятся эксперименты, с целью получения информации об этой системе.

#### Платформы Больших данных

– Hadoop <http://hadoop.apache.org/> Предоставляет интерфейс к Java, свободно распространяется под лицензиями ApacheLicense 2.0 и GNU GPL пакет программного обеспечения, состоящий из управляющего модуля HadoopCommon, распределенной файловой системы HDFS, планировщика заданий YARN и вычислительной платформы HadoopMapReduce. Развивается с 2005 года.

– Spark <https://spark.apache.org/> Предоставляет интерфейсы к Scala, Java, Python и R, распространяется под лицензией ApacheLicense 2.0. Вычислительная платформа, развивающаяся с 2014 года.

– Elasticsearch <https://www.elastic.co/products/elasticsearch> Совместно с системой сбора Logstash и платформой аналитики Kibana составляют интегрированную систему сбора, хранения, поиска и аналитики данных.

– HortonworksDataPlatform (HDP) <https://hortonworks.com/products/data-platforms/hdp/> Платформа управления данными, включающая HDFS, Hadoop, HBase, HCatalog, Pig, Hive, Oozie, Zookeeper, Ambari, WebHDFS, TalentOS, Sqoop, Flume, и Mahout [14].

Всю собираемую Big Data информацию можно классифицировать в зависимости от источников, из которых она была получена[15]:

1. Операционные данные. Это данные о клиентах, поставщиках, партнерах и сотрудниках, доступные в процессе онлайн-обработки транзакций и/или полученные из онлайн-базы данных аналитической обработки.

2. Темные данные. Информация, которая не хранится или не собирается организациями специально, а формируется случайно в процессе ведения бизнеса или взаимодействия с сетевыми сервисами и остается в Интернет-архивах.

3. Коммерческие данные. До появления возможностей технологии Big Data в разных отраслях промышленности существовали агрегаторы коммерчески ценной информации.

4. Официальные данные. Информация, распространяемая государственными органами, открытые публичные реестры, опубликованные нормативные акты, является наиболее достоверной и структурированной.

5. Информация из социальных сетей и сервисов. Вовлеченность бизнеса и частных лиц в функционал крупных социальных сетей (Facebook, ВКонтакте, LinkedIn, Twitter, Instagram и др.) создала еще один источник данных о спросе, тенденциях в определенных сегментах рыночных отношений, новых и перспективных продуктах, услугах и компаниях[16].

Большие данные, без сомнения, имеют преимущества по сравнению с традиционным управлением персоналом, однако они имеют и свои минусы, выражающиеся в следующих рисках:

- Большие данные неоднородны, поэтому их сложно обрабатывать для статистических выводов. Чем больше требуется параметров для прогнозирования, тем больше ошибок накапливается при анализе;
- Хранение и обработка Big Data связаны с повышенной уязвимостью для кибератак и всевозможных утечек. Яркий пример — скандалы с профилями Facebook;
- Несвоевременность использования данных (информация часто не выходит за рамки отдельных подразделений предприятия);
- Проблемы конфиденциальности информации о сотрудниках и клиентах;
- Повышенный риск утечки информации вследствие использования большого количества устройств хранения данных;
- Нехватка HR-менеджеров, обладающих достаточными навыками анализа и использования специального программного обеспечения.

Плюсы и перспективы:

- Динамическая схема: Как упоминалось выше, эта СУБД позволяет гибко работать со схемой данных без необходимости изменять сами данные;
- Масштабируемость: BDA горизонтально масштабируема, что позволяет легко уменьшить нагрузку на сервера при больших объёмах данных;
- Собственно, Big Data — это комплекс, состоящий из аппаратного и программного обеспечения системы, позволяющий осуществлять сбор, обработку, хранение и отображение информации о состоянии объектов в реальном масштабе времени.



– Контактные элементы удобны для быстрого «схватывания» Big Data ситуации с уже введенной частью исходных данных и для понимания того, как наиболее быстро и эффективно ввести оставшиеся данные [17].

Технологии Big Data – это и вызов, и возможность. Полнота управления данными, возможности многоканальной интеграции данных и возможности анализа данных станут способностью предприятий к достижению устойчивого развития. Понимание того, как большие данные увеличивают конкурентное преимущество компании и трансформирует бизнес-модели, является незаменимым процессом для реализации ценности больших данных. Технологии больших данных – это множество способов хранения, обработки, анализа информации, которую в связи с отсутствием четкой структуры, огромной скорости и объема пополнения, невозможно обработать классическими приемами статистического анализа.

На мой взгляд, преимущество технологий, основанных на Big Data, состоит в разумном сочетании объективного контроля за деятельностью сотрудников и постоянного поиска оптимальной мотивации для них на основе полученных данных.

### *Литература*

1. ГОСТ Р7.05-2008 Библиографическая ссылка. Общие требования и правила составления. Федеральное агентство по техническому регулированию и метрологии. Национальный стандарт РФ. ГОСТР 7.0.5-2008. М., Стандартинформ, 2008.

2. URL: <http://blog.oxforddictionaries.com/2013/06/oed-june-2013-update/> (дата обращения: 02.02.2022) - Tweet geekery and epic crowdsourcing: an Oxford English Dictionary update.

3. <https://www.tandfonline.com/doi/full/10.1080/09585192.2019.1674357> (дата обращения: 04.12.2021)

4. Как Big Data помогает в подборе персонала: технологии для хедхантеров [Электронный ресурс] <https://aiconference.ru/ru/article/kak-big-data-pomogaet-v-podbore-personala-tehnologii-dlyahedhanterov-93666>. (дата обращения: 20.11.2021)
5. Lynch C. Big data: how do your data grow? // Nature. 2008. Vol. 455. № 7209. P. 28–29. (дата обращения: 17.03.2022)
6. URL: <https://www.calltouch.ru/glossary/big-data/> (дата обращения: 17.11.2022) – Calltouch. Big data
7. Чехарин Е.Е. Большие данные: большие проблемы // Психологическая наука и образование. 2016. № 3 (21). URL: <http://cyberleninka.ru/article/n/bolshie-dannye-bolshie-problemy> (дата обращения: 25.01.2022).
8. Толстова Ю.Н. Социология и компьютерные технологии // Социологические исследования. 2015. № 8. С. 3–14.
9. Черняк Л. Большие Данные – новая теория и практика // Открытые системы. СУБД. 2011. № 10. С. 18–25. URL: <https://www.osp.ru/os/2011/10/13010990/> (дата обращения: 18.12.2021).
10. URL: [https://ru.wikipedia.org/wiki/Большие\\_данные](https://ru.wikipedia.org/wiki/Большие_данные) (дата обращения: 25.12.2021) - Википедия. Большие данные
11. URL: [https://en.wikipedia.org/wiki/Big\\_data#cite\\_note-13](https://en.wikipedia.org/wiki/Big_data#cite_note-13) (дата обращения: 05.02.2022) - Wikipedia. Big data
12. Цифровая экономика и ее основные характеристики. [Электронный ресурс]. Режим доступа: <http://www.mckinsey.com/global-locations/europe-andmiddleeast/russia/ru/> (дата обращения: 17.02.2022).

13. Армстронг М., Тейлор С. Практика управления человеческими ресурсами. - 14 изд. — СПб.: Питер, 2018. — 785 с.
14. Гобарева, Я.Л., Ширнин Г.В. Большие данные в банковской сфере // Валютный контроль. Валютное регулирование. — 2014. - №8. — С. 58-63. (дата обращения: 11.11.2021)
15. *Lawrence J.* The role of big data in employee performance development, 2013.[Электронный ресурс] / HRZone . — URL: <http://www.hrzone.com/perform/people/the-role-of-big-data-in-employee-performance-development>. (Дата обращения 19.06.2021)
16. Алан Моррисон и др. Большие Данные: как извлечь из них информацию // Технологический прогноз. Ежеквартальный журнал, российское издание. — 2010. — № 3. URL: <https://www.pwc.ru/ru/services/technology.html> (дата обращения: 30.07.2019).
17. Леснова Е. М., Пестов И. Е. Разработка метода обнаружения и коррекции ошибок для распределенной информационной сети на основе больших данных // Материалы XVI Санкт- Петербургской международной конференции «Региональная информатика «РИ-2018» (Санкт- Петербург, 24–26 октября 2018 г.) СПб., 2018. С. 570–571. (дата обращения: 23.03.2022)