РЕАЛИЗАЦИЯ КОНЦЕПЦИИ «BIG DATA» ДЛЯ ОБРАБОТКИ СТРУКТУРИРОВАННЫХ И НЕСТРУКТУРИРОВАННЫХ ДАННЫХ В СИСТЕМЕ ЭФФЕКТИВНОГО УПРАВЛЕНИЯ

## Конькова Н.А.

Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Государственный университет - учебно — научно — производственный - комплекс», г. Орел, Россия (302020, г. Орел, Наугорское шоссе, д. 29, e-mail: nataliya konkova@mail.ru)

В настоящее время во всем мире наблюдается большой интерес к технологиям класса Big Data, связанный с постоянным ростом данных, которыми приходится оперировать крупным компаниям. Накопленная информация для всех организаций является важным активом, однако обрабатывать ее и извлекать из нее пользу с каждым днем становится все сложнее и дороже.

**Ключевые слова:** Big Data, технология, большие данные, информационные системы, программное обеспечение.

## IMPLEMENTATION OF THE CONCEPT «BIG DATA» TREATMENT STRUCTURED AND UNSTRUCTURED DATA IN AN EFFECTIVE MANAGEMENT SYSTEM

## Konkova N.A.

Federal state educational institution of higher professional education «State University - teaching - research - production - a complex», Orel, Russia (302020, Orel, Naugorskoe highway, 29, e-mail: nataliya\_konkova@mail.ru)

Currently, worldwide there is a great interest in technology class Big Data, related to the constant growth of data that is necessary to operate large companies. Collects information for all organizations is an important asset, but handle it and benefit from it every day becoming more and more expensive.

The key words: Big Data, technology, big data, information systems, software.

Від Data представляет собой новый ресурс конкурентоспособности в современной экономике. Полезную информацию можно извлекать из любых источников информации, в том числе: электронной почты, социальных сетей, фотографий и т.д. Главное — научиться правильно обрабатывать и анализировать полученные данные, превращая информацию в актив и стратегический ресурс развития организации [5].

Кроме того необходимо подчеркнуть, что большие данные в нашем мире были всегда.

Крылатым этот термин стал в 2012 году, когда появились новые и очень привлекательные возможности по получению ключевой для бизнеса информации из источников с постоянно растущими объемами данных. Часто считают, что Big Data начинается с нескольких сотен терабайт. Но на самом деле здесь не стоит говорить о конкретных цифрах - огромный объем данных для одной компании может быть «мелочевкой» для другой. Необходимо понимать, что наиболее важным аспектом является не насколько велики данные, а сколько полезной информации из них можно извлечь.

Объемы структурированных и неструктурированных данных сегодня растут быстро и постоянно. В результате компании вынуждены менять подходы к их сбору, хранению и обработке. Поэтому под Big Data следует понимать системы, которые позволяют загружать, преобразовывать, обслуживать, обрабатывать данные и, самое важное, извлекать из них значимую и критически важную бизнес-информацию. Причем все это делать с приемлемой для бизнеса скоростью.

Актуальные примеры технических и ориентированных на бизнес задач применения технологий Big Data отражены на рисунке 1.

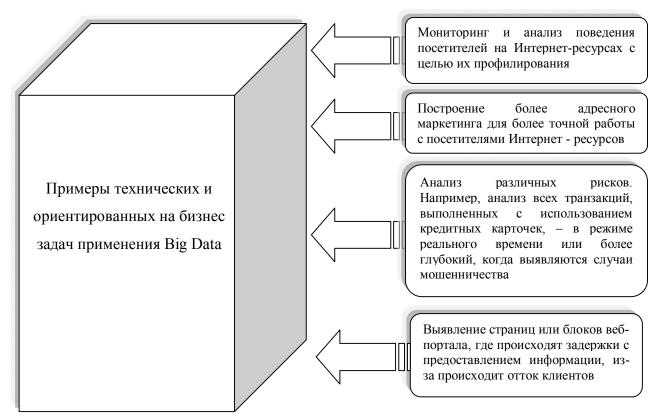


Рисунок 1 - Примеры технических и ориентированных на бизнес задач применения Big Data

В настоящее время на рынке информационных систем и программного обеспечения синонимом Big Data является экосистема Hadoop. Большая часть ее компонентов - это

открытое программное обеспечение. В экосистему входят, в частности, решения Sqoop, Flume и Pig, которые используются для переноса данных в Hadoop-кластеры или из них. МарRedice берет на себя обработку данных. Cascading и WebHDFS применяются для разработки приложений, Hive- для организации доступа к данным через SQL. В числе поставщиков не SQL-решений для исследования и выявления новой информации необходимо отметить MongoDB, Cassandra, HBase, Mahoot, Rhadoop. Кроме того, практически все основные разработчики ВІ-продуктов включают в их функциональность возможности для доступа к данным, которые постоянно содержатся в Hadoop-кластерах. Этот перечень компонентов можно расширять еще и еще, поскольку все больше компаний выходит на рынок со своими предложениями, так или иначе связанными с Hadoop.

Поверх компонентов Hadoop нередко устанавливаются коммерческие приложения, например, для интеграции с базами данных или формирования аналитической отчетности. Зачастую Hadoop используется в связке со стандартными технологиями хранения и обработки данных, а иногда дополняется такими инновационными решениямиl.

Однако большие данные — это не только Hadoop. На рынке есть достаточно много мощных, легко масштабируемых решений для организации хранилищ данных и работы с ними. Среди них можно выделить Greenplum, Netezza и Teradata, которые предлагают эффективную производительность при обработке информации благодаря использованию массивно-параллельной архитектуры (massive parallel processing, MPP).

Потребители Big Data можно разделить на 2 большие категории - это технологические компании и корпоративный рынок. В корпоративном сегменте в первую очередь это финансовые организации: банки, страховщики, инвестиционные компании. Активный спрос мы видим со стороны компаний, ориентированных на массового клиента, - розничные банки, розничная торговля, в том числе электронная, туристический сектор. Отдельный блок клиентов составляют организации, чей бизнес традиционно связан с обработкой большого объема данных. Например, поставщики бизнес- и финансовой информации. Кроме того, средства Big Data нужны крупным веб-порталам различной направленности с миллионами посетителей и огромным трафиком.

Тем не менее на большие данные не стоит смотреть исключительно «технологическим» взглядом, как на способ более инновационно организовать работу с информацией. Есть сферы деятельности, где можно использовать только средства Big Data, так как данных слишком много. Но в целом для таких технологий приоритетная задача – дать бизнесу правильное понимание ситуации на рынке, текущих трендов, потребностей, поведения и мотивации клиентов или других важных для компании вопросов [8].

Помимо значительных преимуществ применения технологий Big Data необходимо



Рисунок 2 - Ключевые проблемы внедрения технологий Big Data

По прогнозам специалистов, к 2020 году количество получаемых данных увеличится в 44 раза по сравнению с 2009 годом. С ростом количества информации растет ее уязвимость для внешних кибератак. В настоящее время безопасность данных и управление рисками являются приоритетными проблемами в области применения технологий Big Data. Трудности с защитой данных в проектах связаны не только с большими объемами информации, но и разными методами доступа к ней и с ограниченными бюджетами, выделяемыми на информационную безопасность. В последние годы перед государствами встает вопрос обеспечения собственной кибер и цифровой безопасности, но индустрия при этом развивается настолько быстрыми темпами, что правительствам приходится скорее предпринимать экстренные меры по латанию дыр в законодательстве, а не разрабатывать и принимать эффективные законы.

В России в настоящее время наиболее производительные предприятия пока не заинтересованы в применении больших данных. На Западе применение данных проектов имеет место быть, например, в нефтяных и горнодобывающих компаниях. Здесь обрабатываются большие потоки данных, которые поступают с приборов, установленных на оборудовании, трубопроводах, транспорте и др. И если какие-то параметры меняются, реакция должна последовать незамедлительно. Качество добываемого сырья могло измениться, следовательно, по правилам компании нужно везти его на другой перерабатывающий завод. Резко повысилось давление в газопроводе - необходимо быстро запустить комплекс мер, чтобы разобраться в ситуации и устранить проблему. Во всех

подобных случаях от скорости принятия решения зависит очень много. В лучшем случае – бизнес понесет чисто экономические потери, в худшем – есть риск аварий и чрезвычайных ситуаций с целым рядом тяжелых последствий. Однако обеспечить приемлемую скорость без технологий Big Data крайне сложно.

Таким образом, когда мы говорим о Big Data, мы понимаем, что это связано с тремя основными аспектами управленческой деятельности крупных компаний: большим объемом информации, ее разнообразием или необходимостью обрабатывать данные очень быстро. С другой стороны, под этим термином часто понимают совершенно конкретный набор подходов и технологий, призванных решить различные задачи бизнеса. В основе одного из таких подходов лежит система распределенных вычислений, где обработка больших объемов данных требует для себя не одну высокопроизводительную машину, а целую группу таких машин, объединенных в кластер.

## Литература

- 1. Арутюнов, Г. Перспективы развития рынка ІТ и телекома в России и за рубежом [Электронный ресурс] / Г. Арутюнов // Технологии и средства связи. Режим доступа: http://www.tssonline.ru/articles2/reviews/perspektivy-razvitiya-rynka-it-i-telekoma-v-rossii-i-za-rubezhom/.
- 2. Майданов, С. Оценка технологий Больших данных [Электронный ресурс] / С. Майданов, В. Сухомлинов // Открытые системы. Режим доступа: http://www.osp.ru/os/2013/02/13034554/.
- 3. 15Большие данные (Big Data) [Электронный ресурс] // TADVISER. Режим доступа: http://www.tadviser.ru/index.php \_(Big\_Data).
- 4. 17Bid Data подарят мировой экономике \$3 трлн [Электронный ресурс]// Interfax. Режим доступа: http://www.interfax.ru/business/txt/337925.
- 5. 19Bid Data: технологии анализа больших данных [Электронный ресур] // State Soft. Режим доступа: http://www.statsoft.ru/products/Enterprise/detail.php?ELEMENT ID=828.
- 6. Территория больших данных [Электронный ресурс]. Режим доступа: http://www.forinnovations.ru/forum/program/25/.
- 7. Технологии больших данных [Электронный ресурс]. Режим доступа: http://taskdata.com/index.php?option=com\_content&view=article&idru.
- 8. Технологии хранения и обработки больших данных [Электронный ресурс]. Режим доступа: http://www.lektorium.tv/course/?id=22932.