



ПРИМЕНЕНИЕ ТЕХНОЛОГИИ BIG DATA В СФЕРЕ ЖЕЛЕЗНОДОРОЖНОГО СООБЩЕНИЯ

Рабинович Александр Евгеньевич

к.э.н., доцент кафедры «Прикладная информатика» Московский Политехнический университет
Россия, г. Москва

Август Алина Витальевна

магистр направления «Системная аналитика больших данных» Московский Политехнический университет
Россия, г. Москва

Аннотация

В статье представлена краткая характеристика применения информационной технологии Big Data в контексте работы сферы железнодорожного сообщения с определением общего функционала самой технологии. В частности, авторы привели примеры реализации Big Data в сегменте железнодорожного сообщения в таких странах, как Турция, Великобритания, Нидерланды, Швеция и Россия.

Ключевые слова: железнодорожное сообщение, Big Data, информационные технологии, поезда, сбор и анализ данных.

APPLICATION OF BIG DATA TECHNOLOGY IN THE FIELD OF RAILWAY COMMUNICATION

Alexander E. Rabinovich

Candidate of Economical Sciences, Associate Professor of the «Applied Informatics» Department at the Moscow polytechnic University
Moscow, Russia

Alina V. Avgust

Master of Science in Big Data Systems Analytics Moscow Polytechnic University
Russia, Moscow

ABSTRACT

The article presents a brief description of the application of Big Data information technology in the context of the work of the railway communication sector with the definition of the general functionality of the technology itself. In particular, the author cited examples of the implementation of Big Data in the railway segment in countries such as Turkey, the UK, the Netherlands, Sweden and Russia.

Keywords: railway connection, Big Data, information technology, the train, data collection, data analysis.

Износ основных фондов – одна из болезненных проблем железнодорожного транспорта. По состоянию на 2011 год более 15% железнодорожных путей имеют просроченный срок службы, более 50% контактной сети находится в эксплуатации свыше 40 лет, причем допустимые величины износа были превышены преимущественно в постсоветское время. [3] Четверть железных дорог центральной России работают в режиме, который превышает оптимальный уровень нагрузки. Развитие современных систем технического обслуживания, анализа за состоянием средств технического оснащения и влияния на нее внешней среды – необходимое условие повышения эффективности работы железных дорог. В настоящее время тратится огромное количество времени на техническое обслуживание поездов, но эффективность работы ремонтных бригад, а вместе с тем и качество ремонта остаются на довольно низком уровне, что можно объяснить отсутствием современных методов сбора, обработки (структурирования) и хранения данных о состоянии технического оснащения железных дорог, а также высокой долей ручного труда.

В 2015 году мировой рынок продуктов и услуг для работы с Big Data составил \$33,3 млрд. Такая цифра приводится на основании исследований американского агентства Wikibon. По их прогнозу, к 2026 году объем индустрии больших данных возрастет до \$85 млрд. Каждый год рынок прибавляет, в среднем, по 17%. В мировом масштабе российский рынок больших данных ничтожно мал. Однако, по сравнению с глобальным, он растет достаточно быстро – почти на 40% в год. По некоторым данным, на конец 2015 года он составлял \$500 млн., по отношению к 2014 году – \$340 млн.[6]

Известно, что большие данные существовали задолго до появления самого термина, так как поисковые системы и соцсети изначально строились на технологии обработки Big Data. По информации из открытых источников технология больших данных активно применяется в банковской сфере (Сбербанк, Газпромбанк, ВТБ24, Альфа-банк и др.), телеком-операторами (МТС, МегаФон и др.), а также крупными ритейлерами (МВидео, Ozon, Азбука вкуса и др.) и онлайн-рекламой.

В госсекторе, где технологии Big Data могут вызвать взрывное увеличение эффективности, они используются крайне слабо. По словам экспертов, среди госструктур технологии больших данных находятся на вооружении у Федеральной налоговой службы, аналитического центра правительства России, Пенсионного фонда, правительства Москвы, Фонда обязательного медицинского страхования, Федеральной службы безопасности, Следственного комитета и Службы внешней разведки.

В настоящее время течение развития мира характеризуют разные процессы, связанные как с геополитикой, так и с полномасштабным внедрением информационных технологий, что способствует усилению цифровизации множества сфер хозяйственной жизни человека. Железнодорожные перевозки, наряду с другими видами транспорта, играют большую роль в экономическом сегменте многих государств, одновременно также эффективно используя множество информационных технологий в своей работе. Опираясь на статистические данные, можно сказать, что в 2018 году на внедрение информационных технологий в мировом соотношении функционировании ж/д было потрачено 159,3 млрд. фунтов стерлингов, что на 9% выше, чем в 2017 году [7], что указывает на положительную тенденцию вовлечения этой индустрии в информационный глобальный процесс. И одной из самых инновационных технологий в сфере железнодорожного сообщения является Big Data, что обуславливает актуальность выбранной темы.

Задачей данной работы является выявление слабых сторон деятельности железных дорог и возможные пути повышения эффективности работоспособности железнодорожного полотна и поездов, путем применения технологий Big Data.

Big Data – это информационно-программное поле, в котором рассматриваются способы анализа, систематического извлечения информации из других наборов данных или обработки их иным способом, которые слишком велики или сложны, чтобы их можно было решить с помощью традиционного программного обеспечения. Важность данной технологии Big Data зависит не от того, сколько в общем контексте имеется данных, а от того, что с ними можно сделать. К примеру, говоря о превалирующих достоинствах Big Data, можно сказать, что они позволяют сократить расходы (время, финансы и прочие ресурсы), разработать новый продукт, оптимизировать предложения, а также принять разумное решение, основываясь на конъюнктуре большого числа параметров, связанных с той или иной ситуацией [5]. Железные дороги это та область, в которой производительность напрямую зависит от способности извлекать информацию из больших объемов сложных данных и уметь ее анализировать. Однако часто бывает сложно получить необходимую информацию для анализа, преградой может становиться, к примеру, законодательное регулирование. Также проблемой является неготовность большинства людей использовать технологии Big Data, так как они привыкли принимать решения, исходя из собственного опыта и руководствуясь своими собственными ощущениями, а не основываясь на данных.

Сам концепт Big Data в плоскости сегмента железнодорожного сообщения стал применяться сравнительно недавно, а вся инициатива началась со стран Европы и США. Исследователь A. Thaduri в своей статье привел данные, согласно которым Big Data уже были эффективно опробованы в качестве инструмента моделирования и прогнозирования ситуации на ж/д транспорте в разных городах и странах мира. К примеру, в Стамбуле эта технология была применена в плане начала работы автоматизированной системы сбора тарифов и ценообразования в контексте работы ж/д линий по всей стране, и, в частности, для линии BRT-us Rapid Transit [5]; в Лондоне подход Big Data был применен в качестве средства нахождения наиболее лучших путей эффективности работы для мультимодальной транспортной сети в городе, учитывая, что в Лондоне имеется надземное и подземное метро, а также ж/д составы, использующие те же линии, что и надземное метро. В Нидерландах в городе Утрехт Big Data были применены для обработки ж/д трафика, чтобы можно было в режиме реального времени видеть какую-либо поломку и направлять рабочую команду туда как можно быстрее для оптимизации технического обслуживания железной дороги. Еще одним положительным примером использования Big Data в контексте функционирования европейских железных дорог является Швеция, где менеджеры в сегменте ж/д инфраструктуры работают с данными, структурированными по времени, скорости, трафику и др. Именно Big Data позволяют обрабатывать источники всех данных, чтобы уменьшить барьеры в отношении их управления. Они применяют систему алгоритмов обнаружений аномалий, чтобы вовремя предугадать на основе имеющихся данных возможную нештатную ситуацию [5]. В целом, в странах Евросоюза с октября 2016 года работает программа, призванная повсеместно внедрить Big Data в сегмент ж/д сообщения, апеллируя к тому, что Big Data, как технология, послужит для улучшения качества обслуживания, а также повышения безопасности железнодорожного сообщения [6].

Тем не менее, и в России в отрасли ж/д сообщения имплементация Big Data уже инициирована. К примеру, ОАО «РЖД» – крупнейшая компания в анализируемой сфере – одной из первых начала процесс цифровой трансформации. ОАО «РЖД» работает

вместе с компанией «Сименс», помогающей собирать данные. Так, сотрудники используют аналитические модели, которые на базе полученных технических параметров позволяют реализовать концепцию предиктивного технического обслуживания, а также позволяют прогнозировать отказы критически важных узлов подвижного состава [1]. Помимо этого, компания использует Big Data для общего повышения безопасности собственной инфраструктуры, что в будущем увеличит эффективность функционирования всей сферы железнодорожного сообщения, которая подконтрольна конкретно ОАО «РЖД» [2].

Однако проблема поломок и износа железнодорожного полотна остается насущным вопросом. Вследствие наличия большого объема ранее не использованных для анализа статистических данных, не берется во внимание та или иная информация о состоянии рельс, о факторах окружающей среды, влияющих на состояние путей. Как итог – сходы и простои составов, нарушение графика движения грузовых и пассажирских поездов.

Так, по данным годового отчета о результатах деятельности ОАО «РЖД» за 2011 год, было выявлено 25[3] случаев изломов литых деталей тележек грузовых вагонов, четыре из которых привели к крушениям и авариям грузовых поездов, 169 – к сходам подвижного состава в поездах.

Также, по данным информационного агентства «РЖД Партнер.ру», за последние 5 лет ущерб от транспортных происшествий на сети РЖД составил 1 млрд 187 млн рублей, причем 60% транспортных происшествий произошли по вине сторонних организаций, отвечающих за ремонт и эксплуатацию ПС[4].

В связи с непрерывным развитием систем учета, объемы данных, накапливаемых крупными производственными компаниями, растут в геометрической прогрессии. Согласно отчету InternationalDataCorporation (IDC) общемировой объем данных, накопленный к 2011 году вырос в 9 раз по отношению к 2006 году и составил 1.8 Збайт информации[9]. Этот показатель и в настоящее время продолжает неуклонно расти и удваиваться каждые 2 года. Такой рост связан с возрастанием технологического уровня и постоянным внедрением инноваций, что влечет к появлению различных сенсоров и датчиков для считывания первичной информации. Регистрируемая информация, кроме того, меняется на качественном уровне. К примеру, в железнодорожном транспорте регистрируется огромное количество информации датчиками GPS. На данный момент пристальное внимание уделяется внедрению и применению учета различных показателей эффективности работы транспортной отрасли, такие как производственные, финансово-экономические, кадровые, а также в сфере безопасности работы железнодорожного транспорта. Впоследствии, возникает потребность в хранении, обработке и анализе многомерных массивов первичных данных. Часто существующие средства анализа данных не позволяют предоставить результаты обработки в адекватной форме и в срок, что нарушает принятие решений в оперативной перспективе.

Существуют следующие общепринятые этапы обработки больших данных: сбор данных, извлечение информации, интеграция и агрегация данных, анализ и моделирование, интерпретация [10]. Из-за того, что объемы первичной информации, в некоторых случаях, может достигать таких масштабов, которые ставят под сомнение возможность хранения массива данных в таком объеме, возникает вопрос очистки, фильтрации и сжатия данных. Достаточно большая часть данных может быть подвергнута компрессии без потери ключевой информации. К примеру, некоторые датчики могут коррелировать между собой на одном участке. Возникает проблема селекции фильтров данных, позволяющих осуществлять компрессию данных без потери полезной

информации, которая является ценной с точки зрения информатизации поддержки конкретного решения. Также, некоторые данные могут содержать ошибки и быть неполными, что имеет место быть из-за погрешности работы некоторых датчиков, человеческого фактора или других причин.

Ни для кого не секрет, что, при обработке больших данных, приходится сталкиваться с рядом таких проблем как: разнородность, непоследовательность и неполнота. При консолидации данных, собранных на разных географических участках железных дорог существует проблема агрегации данных различных форматов и баз данных, которые могут являться неполными и непоследовательными. Большая часть данных, собираемых на железной дороге, является разнородной. В результате чего, возникает потребность привести данные к агрегированному состоянию для дальнейшего анализа. Часто анализ операционных данных имеет ограниченное время, что, при использовании функциональных процессов обработки, анализа и визуализации данных в реальном времени, говорит о существенном повышении эффективности принимаемых решений.

По мнению Томсона П. существует многочисленный ряд возможностей для повышения показателей производительности, надежности, скорости и безопасности работы железнодорожного транспорта [11]. Обработка неструктурированных данных открывает большие возможности для оптимизации производства, повышения рентабельности бизнеса и качества услуг, появления средств для выявления «узких мест», максимальных нагрузок, предупреждения внештатных ситуаций, оперативной поддержке принятия решений.

Для обработки огромного объема разнородных реального времени больших железнодорожных данных требуется связать их не только семантически с возможностью их репозиторного хранения, но и построить систему цифровых словарей, для того чтобы точно во время иметь возможность обслуживать собственно железную дорогу, а также осуществлять ее информационное взаимодействие с другими бизнес-процессами на железной дороге.

Сегодня существует потребность в разработке и имплементации новых и новых алгоритмов анализа данных таких масштабов и объемов. Массивы этих данных можно охарактеризовать такими свойствами, как: большой объем, динамичность, разнородность, значимость и достоверность. Исходные данные могут быть структурированы и не структурированы, возможно, структурированы частично или по отдельным свойствам. Вовремя осуществленный анализ данных с технической, экономической и исследовательской точек зрения поможет осуществить поддержку принятия решений на новом уровне.

Таким образом, можно сделать вывод, что под информационной технологией Big Data понимается комбинация структурированных, полуструктурированных и неструктурированных данных, собранных организациями, которые можно «добывать» для получения информации и использовать в проектах машинного обучения, прогнозном моделировании и других аналитических приложениях. Big Data в настоящее время используются в разных сегментах экономической сферы многих государств, и одним из аспектов применения является внедрение Big Data в железнодорожном сообщении. В целом, эта технология позволяет значительно облегчить сбор большого количества данных, оптимизировать их анализ, разработать сценарии превентивного характера для возможных технических неисправностей, связанными с подвижными составами, а также повысить общий параметр безопасности железнодорожного сообщения.

Адекватное использование информации имеет жизненно важное значение для железных дорог. Основываясь на информации о сети, можно принять правильные решения, чтобы свести к минимуму затраты на эксплуатацию, техническое обслуживание и реализовывать максимальную производительность сетей инфраструктуры. Передача данных из одной организации в другую, от одного процесса к другому происходит постоянно. В связи с чем, есть потребность в новых методах для обработки и обмена информацией. Такие методы должны быть как на уровне данных, так и на уровне процессов.

Список литературы

1. Как на транспорте большие данные превратились в ценный актив [Электронный ресурс], - Режим доступа. - URL:https://www.gudok.ru/science_education/?ID=1463049
2. РЖД планируют использовать Big Data для повышения безопасности [Электронный ресурс], - Режим доступа. - URL:<https://ria.ru/20161108/1480916680.html>
3. Годовой отчет о результатах деятельности ОАО «РЖД» за 2011 год [Электронный ресурс], - Режим доступа. - URL: https://annrep.rzd.ru/reports/public/ru?STRUCTURE_ID=4259
4. ЖД транспорт. Новости [Электронный ресурс], - Режим доступа. -URL: <https://www.rzd-partner.ru/zhd-transport/news/za-poslednie-5-let-ushcherb-ot-transportnykh-proisshestviy-na-seti-rzhd-sostavil-1-mlrd-187-mln-rub/>
5. A. Thaduri. Railway Assets: A Potential Domain for Big Data Analytics // Procedia Computer Science. Vol. 53, 2015, pp. 457-467.
6. Big Data [Электронный ресурс], - Режим доступа. -URL: https://www.sas.com/en_us/insights/big-data/what-is-big-data.html
7. Big Data in railways. EU [Электронный ресурс], - Режим доступа. - URL:https://www.era.europa.eu/sites/default/files/activities/docs/cor_big_data_en.pdf
8. Rail Industry Statistic [Электронный ресурс], - Режим доступа. - URL: <https://www.statista.com/topics/1088/rail-industry/>
9. Gantz J., Reinsel D. Extrcting value from chaos/ IDC iView. - Июнь 2011 С.1-12
10. Jagadish, H.V. et al Big Data and Its Technical Challenges/ Communications of the ACM. - 2014. №7, С. 86-94.
11. Thomas, P. The Role of Big Data in Railroading / Railway Age. - Август 2014. С.44.

References

1. How in transport big data turned into a valuable asset [Electronic resource], - Access mode: https://www.gudok.ru/science_education/?ID=1463049
2. RZD plans to use big data to increase security [Electronic resource], - Access mode: <https://ria.ru/20161108/1480916680.html>
3. Annual report on the results of the activities of RZD for 2011 [Electronic resource],- Access mode: https://annrep.rzd.ru/reports/public/ru?STRUCTURE_ID=4259
4. ZD transport. News [Electronic resource], - Access mode: <https://www.rzd-partner.ru/zhd-transport/news/za-poslednie-5-let-ushcherb-ot-transportnykh-proisshestviy-na-seti-rzhd-sostavil-1-mlrd-187-mln-rub/>
5. A. Thaduri. Railway Assets: A Potential Domain for Big Data Analytics // Procedia Computer Science. Vol. 53, 2015, P. 457-467.

6. Big Data [Electronic resource], - Access mode: https://www.sas.com/en_us/insights/big-data/what-is-big-data.html
7. Big Data in railways. EU [Electronic resource], - Access mode: https://www.era.europa.eu/sites/default/files/activities/docs/cor_big_data_en.pdf
8. Rail Industry Statistic [Electronic resource], - Access mode: <https://www.statista.com/topics/1088/rail-industry/>
9. Gantz J., Reinsel D. Extrzcting value from chaos/ IDC iView. - June 2011 P.1-12
10. Jagadish, H.V. et al Big Data and Its Technical Challenges/ Communications of the ACM. - 2014. №7, P. 86-94.
11. Thomas, P. The Role of Big Data in Railroading / Railway Age. - August 2014. P.44