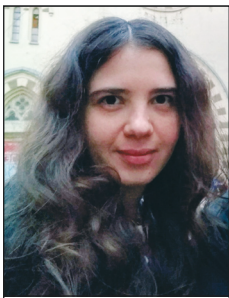


АНАЛИЗ ПРИМЕНИМОСТИ ТЕРМИНА BIG DATA К АВТОМАТИЗИРОВАННОЙ СИСТЕМЕ ОПЕРАТИВНОГО УПРАВЛЕНИЯ ПЕРЕВОЗКАМИ



Е.В. Корнеева



В.Г. Сидоренко

Рассмотрены характеристики, отличающие Big Data от традиционной работы с данными. Выполнен анализ соответствия Автоматизированной системы оперативного управления перевозками третьего поколения (АСОУП-3) приведенным характеристикам. Результаты анализа продемонстрировали, что АСОУП-3 способна достичь требований, предъявляемых к категории Big Data, а также показаны направления доработки для такого развития.

Ключевые слова: Big Data, база данных, Автоматизированная система оперативного управления перевозками (АСОУП)

Железнодорожная отрасль предполагает долгосрочное использование всех задействованных активов. Благодаря этому фактору отрасль становится потенциальным полигоном для внедрения инновационных решений, направленных на продуктивное использование баз данных и интеллектуальных систем, предполагающих использование искусственного интеллекта и машинного обучения. Одним из терминов, часто используемых при обсуждениях современного распространения информационных технологий в мире, является работа с большими данными.

Информационные технологии все больше оказываются вплетены в систему бизнес-процессов железнодорожной отрасли, делая систему эффективнее и надежнее. В настоящее время железнодорожные предприятия всего мира находятся в поисках рево-

люционно новых путей применения современных интеллектуальных систем в железнодорожных процессах [1–10], которые условно можно сгруппировать в три вида:

- процессы, связанные с техническим обслуживанием железных дорог и транспорта;
- мероприятия контроля и инспекции на железнодорожном транспорте;
- эксплуатация и управление сложными, переплетенными графиками и планами, в которых задействованы человеческий фактор и железнодорожная техника [11–13].

В каждой из рассматриваемых групп работа ведется с потоком информации, поступающей с производства и применяемой для дальнейшей работы. Примером успешной работы с большим количеством сведений о

Корнеева Екатерина Владимировна, аспирант кафедры «Управление и защита информации» Института транспортной техники и систем управления Российского университета транспорта (ИТТСУ РУТ (МИИТ)), технолог 2 категории ПКБ-ЦЦТ ОАО «РЖД». Область научных интересов: автоматизированные системы на железнодорожном транспорте, управление процессами перевозок, разработка программного обеспечения. Автор более 10 научных работ.

Сидоренко Валентина Геннадьевна, доктор технических наук, профессор кафедры «Управление и защита информации» Института транспортной техники и систем управления Российского университета транспорта (ИТТСУ РУТ (МИИТ)). Область научных интересов: теория автоматического управления, управление и планирование движения транспортных средств, оценка качества управления движением транспортных средств, планирование работы и обучение оперативных работников транспортных систем. Автор более 180 научных работ.

событиях на железных дорогах Российской Федерации является Автоматизированная система оперативного управления перевозками (АСОУП) разработки ПКТБ-ЦЦТ ОАО «РЖД». На время создания статьи актуальной версией автоматизированной системы является АСОУП третьего поколения (АСОУП-3) [14].

В ОАО «РЖД» сильна тенденция консолидации и централизации разнообразных существующих информационных потоков. Отчасти АСОУП-3 приобретает роль системы, объединяющей потоки данных систем ОАО «РЖД» [15].

Целью данного исследования является анализ применимости термина Big Data и методов, присущих Big Data, к функционирующей АСОУП-3.

Материалы и методы

Для достижения цели исследования использованы теоретические методы — анализ и сравнение.

Обсуждение

Big Data (большие данные) — это общий термин для набора данных, настолько больших или сложных, что становится трудно обрабатывать их с помощью традиционных методов управления данными [16]. Обычно для применения термина Big Data требуется соответствие характеристикам [17], называемым «три V»: объем (Volume) данных, их разнообразие (Variety) и скорость (Velocity) генерации данных [16]. Часто характеристики дополняются достовер-

ностью (Veracity) и полезностью (Value), становясь «четырьмя V» и «пятью V» [18;19]. Прочие возможные характеристики не называются определяющими. Данные характеристики не получили однозначных и устойчивых значений, указывающих на границы для работы именно с большими данными [20].

1. Объем данных. Не существует точного представления о количестве данных, считающихся Big Data. Обобщить предлагаемые значения [17] можно до диапазона от терабайт до зеттабайт.

Объемы данных в АСОУП-3 можно оценить по заявленным техническим требованиям для функционирования системы. Максимальные требования к программно-техническому комплексу разработки и тестирования ядра и компонентов АСОУП-3 приведены в табл. 1.

Максимальные требования к программно-техническому комплексу системы контроля качества и продуктивной системы АСОУП-3 приведены в табл. 2.

Общая предполагаемая численность пользователей — не менее 60000 человек.

Из приведенных данных следует, что объем данных на этапе разработки и тестирования не превышает 4 Тб, объем данных производственной базы данных — 4,5 Тб, тестовой базы данных — 13 Тб. Значения находятся на нижней границе названного ранее диапазона, поэтому допустимо принять значения подходящими категории Big Data.

Таблица 1

Максимальные требования к программно-техническому комплексу систем разработки и тестирования АСОУП-3

Подсистема (разработка и тест)	Central Pro-cessing Unit (CPU), cores	Random Access Memory (RAM), GB	Hard Disk Drive (HDD), GB
	Центральное процессорное устройство (ЦПУ), количество ядер	Оперативное запоминающее устройство (ОЗУ), Гб	Постоянное запоминающее устройство (ПЗУ), Гб
Интеграционная сервисная платформа АСОУП-3	22	80	250
Банк показателей работы АСОУП-3	12	80	150
Ядро АСОУП-3	96	1280	3000
Автоматизированный комплекс ведения нормативно-справочной информации ядра АСОУП-3	14	100	300
Подсистема мониторинга	4	30	100

Таблица 2

**Максимальные требования к программно-техническому комплексу
системы контроля качества и продуктивной системы АСОУП-3**

Компонент	Требования к промышленному полигону	Требования к тестовому полигону
БД ядра	3 физических сервера. Требования к серверу: 24 ядра (cores), 512 Гб ОЗУ (RAM), ПЗУ (HDD): 64 Гб раздел операционной системы (ОС) + 1 Тб (резерв)	3 физических сервера. Требования к серверу: 24 ядра (cores), 512 Гб ОЗУ (RAM), ПЗУ (HDD): 64 Гб раздел операционной системы (ОС) + 1 Тб (резерв)
Сервер приложений ядра	6 Логических разделов сервера (Logical Partition Access Resources (LPAR)), суммарно: 52 ядра (cores), 264 Гб ОЗУ (RAM), 500 Гб ПЗУ (HDD)	6 LPAR суммарно: 52 ядра (cores), 264 Гб ОЗУ (RAM), 500 Гб ПЗУ (HDD)
Шина	2 LPAR: 12 ядер (cores), 64 Гб ОЗУ (RAM), 64 Гб (ОС) + 300 Гб (Data)	2 LPAR: 12 ядер (cores), 64 Гб ОЗУ (RAM), 64 Гб (ОС) + 300 Гб (Data)
Банк показателей работы АСОУП-3	2 LPAR: 12 ядер (cores), 64 Гб ОЗУ (RAM), ПЗУ (HDD): 64 Гб раздел операционной системы (ОС) + 364 Гб раздел для данных (Data)	2 LPAR: 12 ядер (cores), 64 Гб ОЗУ (RAM), ПЗУ (HDD): 64 Гб раздел операционной системы (ОС) + 300 Гб раздел для данных (Data)
Автоматизированный комплекс ведения нормативно-справочной информации ядра АСОУП-3	12 LPAR суммарно: 24 ядра (cores), 100 Гб ОЗУ (RAM), 1 Тб ПЗУ (HDD)	12 LPAR суммарно: 24 ядра (cores), 100 Гб ОЗУ (RAM), 1 Тб ПЗУ (HDD)
Pentaho BI	6 LPAR суммарно: 72 ядра (cores), 384 Гб ОЗУ (RAM), 1 Тб ПЗУ (HDD)	6 LPAR суммарно: 72 ядра (cores), 384 Гб ОЗУ (RAM), 1 Тб ПЗУ (HDD)
Сетевая База АСОУП-3		2 сервера базы данных (БД): 40 ядер (cores), 2048 Гб ОЗУ (RAM), 7.75 Тб ПЗУ (HDD), 2 сервера приложений: 20 ядер (cores), 256 Гб ОЗУ (RAM), 600 Гб ПЗУ (HDD)

2. Скорость как характеристика Big Data — это скорость, с которой данные генерируются и обрабатываются [21]. Критерием Big Data считается создание большого количества данных за малое время.

В АСОУП-3 зачастую данные генерируются в режиме реального времени. Источником данных являются сообщения об операциях на железной дороге, создаваемые операторами или в автоматическом режиме на железнодорожных станциях, а также сведения, передаваемые смежными автоматизированными системами. В среднем время обработки сообщений в АСОУП-3 составляет 0,11 с.

Программное обеспечение АСОУП-3 способно обеспечивать обработку не менее 2,7 млн сообщений об операциях эксплуатационной работы в сутки, что в среднем составляло бы 31 сообщение в секунду, и предполагает сохранение производительности при десятикратном увеличении интенсивности потока сообщений. Согласно источнику [15], на текущий момент количество операций на сети железнодорожных дорог, регистрируемых АСОУП-3 за сутки, уже составляет около 10 млн, что в среднем составляет 116 операций в секунду.

При уменьшении влияния человеческого фактора путем ввода большего числа программных модулей, позволяющих автоматизировать формирование сообщений, и при оптимизации работ смежных систем — источников данных, возможно добиться генерации данных в АСОУП-3 только в режиме реального времени. Это позволило бы уменьшить время от совершения события на железнодорожном транспорте до обработки и регистрации поступающей информации, а также повысить достоверность данных, о которой речь пойдет позже.

3. Разнообразие данных. Термин Big Data предполагает обращение не только со структурированными данными, но и с разными типами неструктурированных данных без конкретных предпочтений по их многообразию [16].

Данные АСОУП-3 описывают широкий спектр процессов, осуществляемых на железных дорогах России, а также перевозочный процесс сопредельных железнодорожных администраций [15;22]. Как уже ранее отмечалось, данные генерируются как из источников самой АСОУП-3, так и из смежных систем.

Архитектура АСОУП-3 [23] предполагает разделение программного обеспечения и данных на три модели:

1. Процессная модель АСОУП-3 [24;25], включающая:

- комплекс эталонных процессов и взаимосвязей между процессами;
- комплекс исполняемых экземпляров процессов, отражающих их реализацию в связи с определенной перевозкой.

2. Обеспечивающие модели АСОУП-3 [26]:

- инфраструктурная составляющая сети железных дорог;
 - модель нормативно-справочной информации (в нее входят нормы и заданные ограничения);
 - модель показателей эксплуатационной работы;
 - экономическая модель.
3. Объектно-событийная модель АСОУП-3 [15]:
- локомотивная составляющая [27–29];
 - бригадная составляющая [27];
 - отправочно-сбытовая составляющая;
 - поездная составляющая [30];
 - вагонная составляющая;
 - контейнерная составляющая [31].

Схематическое изображение моделей АСОУП-3 представлено на рис. 1.

Система обеспечивает возможность хранения исторических данных глубиной до 10 лет.

Сведения, поступающие в обработку АСОУП-3, направляются в соответствующую модель, структурируются и распределяются. Таким образом, АСОУП-3 на текущий момент работает уже со структурированными данными, включенными в архитектуру АСОУП-3 по строго определенной структуре.

Работа с неструктурированными данными требовала бы доработки системы и разработки специализированной подсистемы в АСОУП-3.

4. Достоверность и точность данных.

Качество данных определяет точность их анализа. На достоверность влияют ошибки, искажение информации, шум и аномалии в данных [16]. Сомнения в достоверности данных могут быть вызваны тем, что внешние источники данных могут быть непроверенными, а значит, и сами данные тоже [17]. Увеличение количества сведений, получаемых путем автоматизированного съема данных, и совершенствование методов их обработки сопутствуют повышению достоверности информации [32–34].

Так как источником данных АСОУП-3 является либо сама система, либо смежные системы, то данные подвергаются перекрестной проверке и контролю, что позволяет минимизировать возможные искажения. Это позволяет считать, достоверность данных АСОУП-3 достаточно высокой.

5. Полезность данных.

Большие данные хранятся и анализируются с определенной целью, и эта цель зависит от организации и вида собираемых данных. Данные должны помогать добиться поставленных организацией цели. Полезность данных является обязательным атрибутом Big Data, хотя сложно сказать, насколько определяющим, так как любые собираемые данные должны стремиться быть полезными.



Рис. 1. Схема моделей АСОУП-3

Данные АСОУП-3 являются неотъемлемой частью процесса управления перевозками на железной дороге, они однозначно полезны в работе.

Например, статистическая отчетность АСОУП-3 предоставляет достоверное и полное представление о производственном положении холдинга ОАО «РЖД». В статистическом учете это требование выражается в сплошной и непрерывной регистрации всех фактов производственной деятельности в хронологическом порядке. Статистическая отчетность включает данные обо всех производственных операциях, осуществленных как холдингом ОАО «РЖД» в целом, так и его филиалами, дочерними и зависимыми обществами и иными подразделениями.

В то же время использование термина Big Data применимо не к любому большому количеству данных, соответствующих характеристикам «трех V» или «пяти V», а к случаям реализации передовых методов анализа данных, например, предиктивной аналитики или иных [7].

В наши дни существует множество инструментов и структур больших данных, новые технологии появляются очень быстро [35;36]. Специалисты по

обработке больших данных обязательно используют методы различных современных и актуальных технологий.

Выводы

Проведенный анализ показал, что условия «пяти V» не полностью применимы к данным АСОУП-3, но при этом вполне достижимы при некоторых доработках, расширении функций системы. Подобные модернизации позволили бы применять в отношении АСОУП-3 термин Big Data.

В вопросе использования передовых методов анализа данных АСОУП-3 предстоит обновить набор используемых методов и инструментов для решения новых современных задач.

Заключение

Подводя итог, следует отметить, что для соответствия категории Big Data предстоит провести модернизацию АСОУП-3 в части применения новых инструментов работы с данными, а также принять к работе новые для АСОУП-3 типы данных, расширив сферу работы системы.

Литература

1. Thaduri A., Galar D., Kumar U. Railway assets: A potential domain for big data analytics // *Procedia Computer Science*. – 2015. – №53. – С. 457–467. – Текст: непосредственный.
2. Ponzini J. How Big Data, Predictive Analytics And Machine Learning Changed The Railway Industry : сайт. – 2021. – URL: <https://www.forbes.com/sites/sap/2021/02/11/how-big-data-predictive-analytics-and-machine-learning-changed-the-railway-industry/?sh=15a71bf5e81f> (дата обращения: 06.12.2021). – Текст: электронный.
3. Innovative Business Planning & Traffic Routing for Kazakhstan Railways: сайт компании «SAP». – 2020. – URL: <https://www.sap.com/bin/sapdxc/inm/attachment.7378/pitch-deck.pdf> (дата обращения: 06.12.2021). – Текст: электронный.
4. Easton J., Kohli Sh. Challenges and opportunities for big data in the Digital Railway: сайт журнала «Железнодорожные технологии». – 2017. – URL: <https://www.railtechnologymagazine.com/Comment/challenges-and-opportunities-for-big-data-in-the-digital-railway> (дата обращения: 01.12.2021). – Текст: электронный.
5. van Gulijk C., Dacre M. Big data risk analysis for rail Railways: сайт журнала «Железнодорожные технологии». – 2015. – URL: <https://www.railtechnologymagazine.com/Comment/big-data-risk-analysis-for-rail> (дата обращения: 02.12.2021). – Текст: электронный.
6. Технологии Big data в транспортном планировании: сайт Общероссийской общественной организации «Российская академия транспорта». – 2020. – URL: http://rosacademtrans.ru/bigdata_transportplanning/ (дата обращения: 05.12.2021). – Текст: электронный.
7. Machine Learning at the Core of Automated Inspection and Predictive Maintenance Railways . - сайт Совета безопасности на железнодорожном транспорте и стандартов. – 2018. – URL: <https://www.rssb.co.uk/what-we-do/insights-and-news/Blogs/Machine-Learning-at-the-core-of-automated-inspection-and-predictive-maintenance> (дата обращения: 07.12.2021). – Текст: электронный.
8. BIM и инженерные формализованные онтологии на цифровой железной дороге Европы в объединении EULYNX-экономика данных / Климов А.А. [и др.] // *International Journal of Open Information Technologies*. – 2018. – Vol. 6. – № 8. – С. 38–65. – Текст: непосредственный.
9. Цифровая железная дорога – ERTMS, BIM, GIS, PLM и цифровые двойники/ Климов А.А. [и др.] // *Современные информационные технологии и ИТ-образование*. – 2017. – Vol. 13. – № 3. – С. 129–166. – Текст: непосредственный.
10. Марков, Л. Большие данные выходят на маршрут : Сайт газеты «Гудок». – 2021. – URL: <https://gudok.ru/newspaper/?ID=1588822&archive=2021.12.06> (дата обращения: 23.12.2021). – Текст: электронный.
11. Бородин, А.Ф. Свидетельство о государственной регистрации программы для ЭВМ №2016613861 Российская Федерация. Интегрированная система автоматизированной разработки технологического процесса при движении грузовых поездов по расписанию. Версия 1: № 2016611091: заявл. 12.02.2016: опубл. 11.04.2016 / А.Ф. Бородин, А.И. Сафронов, Л.Б. Сушенцева [и др.]; заявитель открытое акционерное общество «Российские железные дороги».
12. Свидетельство о государственной регистрации программы для ЭВМ №2019613065 Российская Федерация. Программное обеспечение технологии организации вагонопотоков (АСОВ). Очередь 2018 года: №2019611807: заявл. 22.02.2019: опубл. 06.03.2019 / М.А. Агеева, А.Ф. Бородин, О.Н. Кириллова [и др.]; заявитель Открытое акционерное общество «Российские железные дороги».
13. Построение архитектуры интеллектуальной системы управления городской рельсовой транспортной системой / В.М.Алексеев, Л.А.Баранов,Л.А.Кулагин, В.Г.Сидоренко. – Текст: непосредственный // *Мир транспорта*. – 2021. – Т.19, №1. – С. 18–46.
14. Кисиль, Ю.А. АСОУП – единая платформа для систем управления перевозочным процессом/ Ю.А. Кисиль, Д.Ю. Фоменков, В.Н. Давиденко. – Текст: непосредственный // *Автоматика, связь, информатика*. – 2021. – №7. – С. 10–12.
15. Урусов, А.В. 50 лет – это только начало! / А.В.Урусов. – Текст: непосредственный // *Автоматика, связь, информатика*. – 2021. – №7. – С. 2–4.
16. Cielen D., Meysman A.D.B., Ali M. *Introducing Data Science*. Manning Publications Co, 2016. – 300. – Текст: непосредственный.

17. Everts S. Information Overload // Институт Истории науки США: официальный сайт. —2016. —URL: <https://www.sciencehistory.org/distillations/magazine/information-overload> (дата обращения: 07.12.2021). —Текст: электронный.
18. Anil J. The 5 V's of big data // Компания «IBM»: официальный сайт. —2016. —URL: <https://www.ibm.com/blogs/watson-health/the-5-vs-of-big-data/> (дата обращения: 07.12.2021). —Текст: электронный.
19. Ghofrani F., He Q., Goverde R.M.P., Liu X. Recent applications of big data analytics in railway transportation systems: A survey // Transportation Research Part C. —2018. —№90. —С. 226–246. —Текст: непосредственный.
20. Davis B. The 7 pillars of Big Data // Petroleum Review. —2015. —№1. —Текст: непосредственный.
21. Al Nuaimi E., Al Neyadi H., Mohamed N. Applications of big data to smart cities // Journal of Internet Services and Applications. —2015. —№6(1). —С. 25. —Текст: непосредственный.
22. Богомолов, В.Ю. Свидетельство о государственной регистрации программы для ЭВМ №2021610-465 Российская Федерация. Программный модуль АСОУП-3 «УЧЛЮК ИСГ» заявка №2020666448. заявл. 10.12.2020; рег. 14.01.2021 / В.Ю. Богомолов, О.В. Зудова, Е.В. Корнеева [и др.]; заявитель открытое акционерное общество «Российские железные дороги».
23. Урусов, А.В. Реализация АСОУП нового поколения / А.В.Урусов. —Текст: непосредственный // Автоматика, связь, информатика. —2016. —№6. —С. 5–6.
24. Развитие общесистемной составляющей АСОУП-3 / В.Н. Давиденко, С.Н. Юсупов, В.С. Тимошенко, Д.Н.Чубчев. —Текст: непосредственный // Автоматика, связь, информатика. —2016. —№6. —С. 7–9.
25. Баврин, Г.Н. Автоматизированная система управления перевозками на DB2/ Г.Н.Баврин, В.Н. Якимец. —Текст: непосредственный // Автоматика, связь, информатика. —2011. —№7. —С. 24–26.
26. Трегубов, А.Г. Совершенствование отраслевой системы классификации и кодирования/ А.Г. Трегубов. —Текст: непосредственный // Автоматика, связь, информатика. —2011. —№7. —С. 34–35.
27. Одинцов, В.Н. Эволюция оперативных дислокационных моделей локомотивов и локомотивных бригад / В.Н. Одинцов, Т.Б. Прилипко, В.Н. Коротков. —Текст: непосредственный // Автоматика, связь, информатика. —2011. —№7. —С. 27–29.
28. Богомолов, В.Ю. Информационное сопровождение бизнес-процессов управления локомотивным комплексом / В.Ю. Богомолов. —Текст: непосредственный // Автоматика, связь, информатика. —2016. —№6. —С. 15–16.
29. Богомолов, В.Ю. Основной тренд — цифровизация локомотивного хозяйства/ В.Ю. Богомолов, В.Н. Коротков. —Текст: непосредственный // Автоматика, связь, информатика. —2021. —№7. —С. 29–32.
30. Щербакова Н.В. Проектирование базы данных перевозочного процесса: от поездной модели до полнообъектной / Н.В. Щербакова. —Текст: непосредственный // Автоматика, связь, информатика. —2011. —№7. —С. 20–21.
31. Бабкина, И.Н. Гарантируемая безопасность пассажирского движения / И.Н. Бабкина, Н.С. Пряхина. —Текст: непосредственный // Автоматика, связь, информатика. —2021. —№7. —С. 25–28.
32. Гапанович, В.А. Некоторые вопросы управления ресурсами и рисками на железнодорожном транспорте на основе состояния эксплуатационной надежности и безопасности объектов и процессов (проект УРРАН) / В.А. Гапанович, А.М. Замышляев, И.Б.Шубинский. —Текст: непосредственный // Надежность. —2011. —№1. —С. 2–8.
33. Шубинский, И.Б. Основные научные и практические результаты разработки системы УРРАН / Шубинский, И.Б., Замышляев А.М. —Текст: непосредственный // Железнодорожный транспорт. —2012. —№10. —С. 23–28.
34. Замышляев, А.М. Прикладные информационные системы управления надежностью, безопасностью, рисками и ресурсами на железнодорожном транспорте / А.М. Замышляев. —Ульяновск: Областная типография «Печатный двор», 2013. —143 с. —Текст: непосредственный.
35. Применение современных технологий программирования к автоматизации планирования движения поездов метрополитена/ В.Г. Сидоренко, А.И. Сафронов, К.М. Филипченко, М.А. Чжо. —Текст: непосредственный // Автоматика на транспорте. —2016. —Т.2, №3. —С. 331–347.
36. Сидоренко, В.Г. Архитектура многопоточного программного продукта, реализующего планирование логистических процессов / В.Г. Сидоренко, А.С. Петров. —Текст: непосредственный // Информатизация образования и науки. —2020. —№1(45). —С. 25–38.