

СИСТЕМНЫЙ АНАЛИЗ БОЛЬШИХ ДАННЫХ ДЛЯ НАУК О ЗЕМЛЕ

© 2023 г. А. Д. Гвишиани^{a,b,*}, В. Я. Панченко^{c,**}, И. М. Никитина^{a,***}

^aГеофизический центр РАН, Москва, Россия

^bИнститут физики Земли им. О.Ю. Шмидта РАН, Москва, Россия

^cРоссийская академия наук, Москва, Россия

*E-mail: a.gvishiani@gcras.ru

**E-mail: viapanchenko@pran.ru

***E-mail: i.nikitina@gcras.ru

Поступила в редакцию 18.05.2023 г.

После доработки 20.05.2023 г.

Принята к публикации 25.05.2023 г.

В настоящее время большие данные — одно из наиболее обсуждаемых сегодня явлений в сфере информационных технологий. Что такое большие данные? Как они создаются? Как устроено программное обеспечение больших данных? Как теория и практика больших данных используются и могут быть использованы применительно к наукам о Земле? Как проблематика больших данных развивается в рамках Российской академии наук? Эти вопросы обсуждаются в статье, подготовленной на основе научного сообщения, представленного на заседании Президиума РАН 28 марта 2023 г.

Ключевые слова: большие данные, системный анализ, науки о Земле.

DOI: 10.31857/S0869587323060087, **EDN:** YCSNEI

Большие данные (Бод) — одно из наиболее обсуждаемых сегодня явлений в сфере информационных технологий, в том числе в их приложении к научным исследованиям, бизнесу, информационной сфере во всех её аспектах. Этот термин часто используется достаточно свободно, если не сказать волюнтаристски. Нередко те, кто говорят о Бод, концентрируют внимание скорее на интуитивном понимании термина, называя данные большими, когда их просто много. На самом деле Бод — это формализованная система понятий, имеющая очерченную область применения, снабжённая оригинальным программным обес-

печением и лежащей в его основе математикой. Последняя является частью системного анализа.

Масштабируемость системы. На разных этапах развития любой информационной системы возникает проблема превышения расчётной нагрузки. Эта проблема может быть обусловлена увеличением объёмов данных, числа пользователей, количества параметров, скорости обработки и другими подобными причинами. Соответственно, важным базовым понятием системы является *масштабируемость*.

Информационная система называется масштабируемой, если она: во-первых, способна на-



ГВИШИАНИ Алексей Джерменович — академик РАН, научный руководитель ГЦ РАН. ПАНЧЕНКО Владислав Яковлевич — академик РАН, вице-президент РАН. НИКИТИНА Изабелла Михайловна — кандидат технических наук, старший научный сотрудник ГЦ РАН.



Рис. 1. Основные и дополнительные критерии, характеризующие БД

рашивать свои вычислительные и системные ресурсы без структурных изменений; во-вторых, может увеличивать свою производительность пропорционально дополнительным ресурсам.

Разделяют вертикальную и горизонтальную масштабируемость. *Вертикально масштабируемая система* увеличивает свою производительность за счёт усиления каждого компонента системы, без изменения программ. Иными словами, производительность увеличивается за счёт усиления нескольких или всех компьютеров системы или мощности сервера, например, благодаря расширению его оперативной памяти.

Горизонтально масштабируемая система усиливает свою производительность путём добавления к системе новых компонентов. При этом для полноценного использования добавленных ресурсов может потребоваться модификация программ. Иначе говоря, горизонтально масштабируемая система — это система, позволяющая добавлять как новые компьютеры, так и новые серверы. Добавление может происходить физически или виртуально.

Критерии больших данных. Термин “большие данные” появился сравнительно недавно, в 2008 г., по аналогии с терминами “большая нефть”, “большие деньги”, “большая вода” и т.д. Его предложил редактор журнала “Nature” Клиффорд Линч, в статье “Big data: how do your data grow?”, опубликованной в спецвыпуске, посвящённом взрывному росту мировых объёмов информации [1].

В 2010 г. БД впервые характеризуются свойствами, известными как критерий 3V: volume (объём), velocity (скорость), variety (разнообразие) (рис. 1). БД должны обладать большими значениями всех 3V. Это и стало определением понятия БД.

Возникает естественный вопрос: какие значения следует считать большими? Имея в виду относительность ответа, можно утверждать, что это значения каждого из 3V, существенно превышающие стандартные пороги для соответствующих выборок. Следует уточнить, что в данном контексте скорость понимается в широком смысле, как скорость сбора, передачи, архивирования, научного анализа и роста разнообразия информации. Последнее особенно важно, потому что огромные объёмы совершенно одинаковых данных не есть БД.

Позже в некоторых случаях к 3V стали добавлять три дополнительных критерия: value (ценность) определяется значимостью входящей и обработанной информации для целей исследования БД; validity (уместность) характеризует имеющиеся шансы наличия связи исходных и обработанных данных с объектом исследований; variability (вариативность) показывает способность развёртывания исходных и обработанных данных в сторону усиления их разнообразия.

Эти параметры возникли в ответ на лавинообразный рост объёма данных, когда актуальными стали вопросы, могут ли полученные данные принести пользу и насколько им можно доверять [2]. В результате появились особые типы БД, которые, наряду с большими значениями основных 3V, обладают большими значениями одного или нескольких дополнительных V. Подобные 4V, 5V, 6V БД де-факто являются собственными множествами некоторых 3V БД.

Создание больших данных. Создание БД — это процесс, разворачивающийся во времени. У него есть начало, но нет конца. Цифровые электронные данные генерируются и увеличиваются в объёме как снежный ком. Нельзя принять моментальное решение о создании БД. Можно начать собирать данные и не останавливаться. Тогда рано или поздно мы создадим БД. Далее эти данные будут согласованно расти в направлении увеличения каждого из трёх основных V.

Классическим и достаточно очевидным примером БД являются операции в Интернете. Анализ происходящего в сети в течение 2023 г., выполненный компанией LOCALiQ [3], показал, что в среднем за одну минуту уникальные пользователи:

- 3.67 млн раз просматривали видео на YouTube;
- 5.9 млн совершили поисковые запросы в Google;
- 231 млн отправили e-mail сообщение.

Процесс создания БД следует за процессом увеличения количества данных. Так, по результатам исследования компании IDC [4] за последние 15 лет объём мировых данных вырос в 100 раз — с 1 ЗБт до 100 ЗБт (единицы измерения количества

информации в байтах приведены в табл. 1). При этом более половины данных формируется на устройствах конечных пользователей. Бод создаются из первоначального ядра обычных данных путём развёртывания во времени одновременно каждого из трёх V. Процесс имеет начало, но не имеет конца. Единожды начав создавать Бод, невозможно остановиться, поскольку сама система управления ими базируется как на вновь поступающих потоках информации, так и на неуничтожаемых архивах. Факт постоянного роста больших данных служит базой для неразрывно связанных с ними методов анализа. Важным свойством здесь является неуничтожимость Бод, обеспеченная оригинальными ПО.

Новая парадигма научных исследований. Буквально через несколько лет после появления термина “большие данные” стало ясно, что их изучение приводит к новой парадигме научных исследований. Цель научного поиска смещается с ответа на вопрос “почему?” на ответ на вопросы “что именно происходит?” и “как действовать?”. С этой целью во главу угла ставится распознавание корреляций и трендов в гигантских массивах информации. Как показывают реализованные коммерческие и научные приложения, расчёт трендов стал эффективным инструментом внедрения научных результатов в реальный сектор экономики, укорачивающим путь от их анализа *a posteriori* к прогнозированию.

Преобразование количества накопленной информации в качество решений называют феноменом Бод. Программы для игры в шахматы (например, программа “Каисса”, ставшая в 1974 г. первым чемпионом мира по шахматам среди компьютерных программ), построенные на тех же принципах, что и алгоритмы 60-х годов XX в., сегодня играют на уровне гроссмейстеров и даже превышая его. Это достижение определяется сегодняшней возможностью доступа к гигантскому множеству данных истории шахматных партий при неизменных правилах игры.

Феномен Бод ярко проявляется в прогрессе продуктов искусственного интеллекта, например, Google Translate. Система использует не строгие своды грамматических правил и исключений языка, а статистическую вероятность употребления того или иного слова с учётом контекста, рассматривая язык как беспорядочную систему данных. Простые алгоритмы, имеющие доступ к огромному объёму данных, имеют лучшие результаты, чем сложные алгоритмы, обучаемые на узких выборках [6]. Качественный скачок определяется именно огромным объёмом и разнообразием данных, которые поступают и обрабатываются очень быстро, и лишь отчасти — совершенствованием алгоритмов.

Таблица 1. Единицы измерения количества информации в байтах [5]

Название	Обозначение	Степень
Байт	Б	10^0
Килобайт	Кбайт	10^3
Мегабайт	Мбайт	10^6
Гигабайт	Гбайт	10^9
Терабайт	Тбайт	10^{12}
Петабайт	Пбайт	10^{15}
Эксабайт	Эбайт	10^{18}
Зеттабайт	Збайт	10^{21}
Йоттабайт	Йбайт	10^{24}

Модель MapReduce. Для работы с Бод не годятся привычные подходы и методы хранения и обработки информации, например СУБД. Большие данные нужно обрабатывать распределённо. Для этого необходим некоторый универсальный общий подход, реализуя который, можно обрабатывать и безопасно хранить большие объёмы в том числе и неструктурированных данных.

Соответствующая модель распределённых вычислений MapReduce была впервые предложена сотрудниками корпорации “Google” в 2004 г. [7]. В её основе лежат две процедуры функционального программирования: *map* (отображение), в которой применяется нужная функция к каждому элементу списка, и *reduce* (ограничение), объединяющая результаты работы этапа *map*. В процессе вычислений множество входных пар ключ/значение преобразуется в множество выходных пар ключ/значение. Иными словами, MapReduce — это модель параллельной обработки Бод на компьютерных кластерах за счёт соответствующего разделения задачи на независимые части. Схема работы MapReduce для наглядности представлена на классическом примере подсчёта количества слов в тексте (рис. 2).

Почти двадцать лет назад на базе модели MapReduce появилось программное обеспечение обработки Бод, альтернативное СУБД — без требований жёсткой иерархии и однородности данных. Это ПО может обрабатывать как структурированные, так и неструктурированные данные, например тексты.

Одной из основных программных реализаций модели MapReduce стала программная платформа HADOOP. Это открытая система, однако чтобы адаптировать её к собственным данным, необходимо создать некий дополнительный программный аппарат для актуализации сборки.

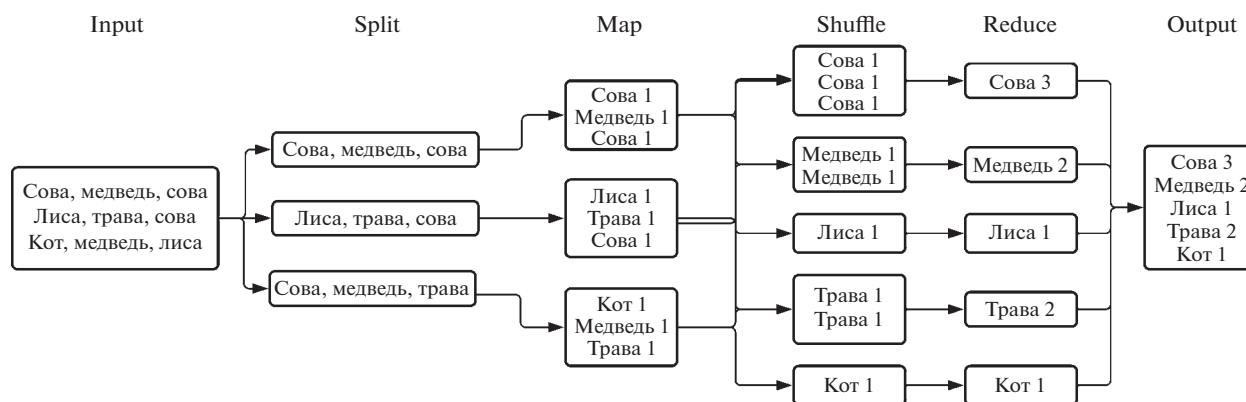


Рис. 2. Схема работы MapReduce

Сегодня такие готовые решения HADOOP применительно к обрабатываемым потокам информации находятся под санкциями. Создание аналогичных отечественных решений – важная задача российских учёных, инженеров и коммерческих компаний.

Большие данные в науках о Земле. В работах [8, 9] изучено и обосновано отнесение к большим данным информационных потоков, которые создаются в рамках следующих наук о Земле:

- метеорология – метеорологические данные сегодня составляют петабайты (10^{15} Б) информации;
- изучение Земли из космоса – данные дистанционного зондирования Земли (ДЗЗ) к 2025 г. достигнут объёма 300 экзабайт (10^{18} Б);
- экология – информация международной сети экологических наблюдений станций SMEAR составляет сотни терабайт (10^{12} Б);
- сейсмология – данные глобального сейсмического мониторинга, включая временные ряды сейсморегистрации, измеряются в петабайтах (10^{15} Б);
- геофизика – объём информации геофизической разведки и поиска полезных ископаемых достигает 100 петабайт (10^{15} Б) в сутки;
- горные науки – данные горно-добывающих и перерабатывающих комплексов составляют 130 терабайт (10^{12} Б) в год.

Следует иметь в виду, что в приведённый здесь список включены лишь те науки о Земле, которые гарантированно создают Бод. Целый ряд других дисциплин со временем тоже станет и уже становится источником больших данных. Это происходит как благодаря распространению систем наблюдений в региональных и глобальных масштабах, так и в силу появления принципиально новых типов наблюдательного оборудования. Дальнейшее распространение высокоскоростного Интернета по планете обеспечивает возмож-

ность оперативной обработки и постоянного сохранения Бод наук о Земле на базе системы HADOOP и её аналогов. А значит, потоки данных будут неуклонно увеличиваться, особенно при развёртывании международного сотрудничества. Поэтому представленный выше список Бод в науках о Земле открыт для расширения.

Региональные Бод на примере Арктической зоны Российской Федерации (АЗРФ). Перспективным представляется создание и дальнейшее развитие мультидисциплинарных региональных Бод, в том числе применительно к АЗРФ.

Источниками данных здесь служат:

1. Информация мониторинга и изучения АЗРФ, создаваемая институтами РАН, Северным (Арктическим) федеральным университетом им. М.В. Ломоносова и другими университетами, научно-исследовательскими центрами МЧС и ГК “Росатом”.
2. Центральные исследовательские проекты в Арктике:
 - система спутников дистанционного зондирования Земли “Арктика”;
 - ледовая самодвижущаяся платформа “Северный полюс”;
 - арктические станции “Снежинка”;
 - транспортный коридор “Северный морской путь”.
3. Проекты РНФ по изучению и мониторингу Арктики.

Проект “Системный анализ динамики геофизических процессов в Российской Арктике и их воздействие на развитие и функционирование инфраструктуры железнодорожного транспорта” с 2021 г. реализует Геофизический центр РАН (финансируется РНФ, проект 21-77-30010), АО “НИИАС” (РЖД) создаёт прототип Бод для АЗРФ.

Информация, генерируемая в проекте, обладает большой степенью разнообразия. Бод гене-

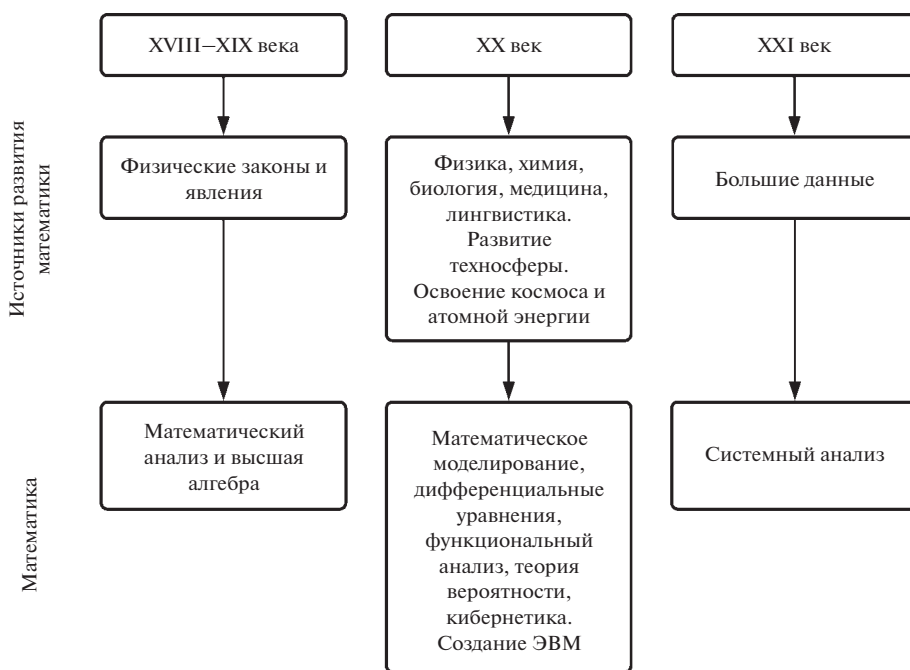


Рис. 3. Математика XX–XXI вв.

пируются по следующим направлениям: изучение землетрясений, сейсмической опасности и магнитного поля Арктики, гравитационные исследования литосферы АЗРФ, метеорологические и климатические исследования и системная интеграция полученных результатов в геоинформационную систему.

Системный анализ в математике XXI столетия.

Источником развития математики служили и продолжают служить естественные науки. В XVIII–XIX вв. двигателем математики была физика (рис. 3), в то время выдающиеся физики часто оказывались и математиками. В их числе следует упомянуть прежде всего великого английского учёного Исаака Ньютона. Открытые им фундаментальные законы физики стали основой разработанного им же дифференциального и интегрального исчисления. Золотыми буквами в историю физики, астрономии и одновременно математики вписаны имена немецкого учёного Карла Гаусса и французского учёного Пьера-Симона Лапласа.

В XX в. развитие математики стимулировали и другие естественные науки. Важную роль сыграло внедрение математического моделирования в целый ряд дисциплин — физику, астрономию, химию, биологию, лингвистику и др. Драйверами развития математики явились, кроме того, крупные инженерно-технические проекты, такие как освоение космоса и покорение атомной энергии. Тогда же были созданы компьютеры (ЭВМ) и по-

лучила свой математический аппарат кибернетика.

В XXI в. теоретическая математика получила беспрецедентное развитие. Выдающиеся результаты получены в теории дифференциальных уравнений, функциональном анализе, теории вероятностей и теории управления — математических дисциплинах, прямо представляющих математические модели как проблемы вычислений. Свой вклад в прогресс математики внесли и большие данные, которые, как уже было сказано, возникают при глубоком изучении любой научной дисциплины.

История Бод — драйвера новой части системного анализа — разворачивается на наших глазах — на протяжении последних 15 лет. Эта математическая дисциплина находится в стадии своего формирования. Тем не менее уже сегодня прослеживаются её характерные особенности, ожидающие строгой математической формализации в будущем.

Методы обработки Бод ориентированы как на структурированную, так и неструктурированную информацию. Например, матричные числовые данные и раскрывающие их неструктурированные тексты обрабатываются с помощью программного обеспечения, реализующего концепцию Map Reduce, как одно целое. Так изучается множество, состоящее из объектов нового типа.

Новый смысл приобретает и понятие высоких порогов для различных V, характеризующих Бод. Пороги возрастают вместе с Бод, являясь, как и

Таблица 2. Возможные направления исследований ИАСА и РАН. Актуальное и потенциальное сотрудничество между программами ИАСА и тематическими отделениями РАН

Программы ИАСА	Отделения РАН
Математический и прикладной системный анализ	Отделение математических наук РАН (ОМН РАН) Отделение нанотехнологий и информационных технологий РАН (ОНИТ РАН) Отделение наук о Земле РАН (ОНЗ РАН)
Биоразнообразие и природные ресурсы	Отделение биологических наук РАН (ОБН РАН) Отделение сельскохозяйственных наук РАН (ОСХН РАН) Отделение химии и наук о материалах РАН (ОХНМ РАН) Отделение физиологических наук РАН (ОФ РАН)
Экономика и смежные дисциплины	Отделение общественных наук РАН Отделение глобальных проблем и международных отношений РАН (ОГПМО РАН) Отделение историко-филологических наук РАН (ОИФН РАН) Отделение наук о Земле РАН (ОНЗ РАН) Отделение математических наук РАН (ОМН РАН) Отделение нанотехнологий и информационных технологий РАН (ОНИТ РАН)
Энергия, климат и окружающая среда	Отделение наук о Земле РАН (ОНЗ РАН) Отделение физических наук РАН (ОФН РАН) Отделение энергетики, машиностроения, механики и процессов управления РАН (ОЭММПУ РАН)
Население и общество	Отделение общественных наук РАН (ООН РАН) Отделение физиологических наук РАН (ОФ РАН) Отделение медицинских наук РАН (ОМедН РАН) Отделение математических наук РАН (ОМН РАН) Отделение глобальных проблем и международных отношений РАН (ОГПМО РАН)
Обсуждение программы Бод ИАСА	Планирование проекта Бод РАН

сами информационные потоки, возрастающими функциями времени. Изучение возможного явного вида этих функций — одна из многих задач, которые практика Бод ставит перед их будущей математической теорией. Даже простое формирование относительно полного списка подобных задач могло бы стать предметом самостоятельного математического исследования.

Международный институт прикладного системного анализа (International Institute of Applied System Analysis, IIASA). Цель работы института — создавать и совершенствовать методы системного анализа и применять их для обоснования прикладных управленческих решений. Приоритетные направления исследований — сокращение антропогенного воздействия на климат Земли, повышение устойчивости природных и социально-экономических систем и достижение целей устойчивого развития человечества.

IIASA основан в 1972 г., он расположен в Австрии, в г. Лаксенбург в здании Лаксенбургского дворца. Учредителями института выступили СССР, США и Австрия. В настоящее время членами IIASA являются 22 страны. Всего в IIASA постоянно работают около 300 учёных из разных стран — официальных членов института, включая порядка 30 специалистов из России. Согласно до-

говорённости 1972 г. с австрийским правительством, годовая аренда рабочего помещения IIASA составляет всего один евро. Членство РФ в IIASA осуществляет Российская академия наук. Совместную работу РАН и IIASA координирует Комитет по системному анализу РАН под председательством вице-президента РАН академика В.Я. Панченко. Сотрудничество РАН с IIASA непрерывно продолжается с 1972 г.

Актуальные и потенциальные направления развития постоянных программ IIASA в их сопоставлении с тематикой исследований отделений РАН приведены в таблице 2. Каждая из программ IIASA — междисциплинарная и многоцелевая, что отвечает актуальным целям и задачам РАН. Действующими и потенциальными партнёрами любой из программ института могут быть сразу несколько тематических отделений РАН.

Проект создания больших данных в IIASA в настоящее время находится в стадии обсуждения, как и проект разработки плана создания Бод РАН, цели которого, в предварительной форме, были представлены авторами данной статьи на заседании Президиума РАН 28 марта 2023 г. Предложение получило поддержку Президиума, и разработка плана создания Бод РАН началась. С учётом санкционной политики Запада расши-

рение сотрудничества РАН и IASA по тематике больших данных представляется взаимно полезным и перспективным.

Большие данные в реальном секторе экономики. Использование опыта для создания Бод РАН. Сейчас уже можно с уверенностью говорить об эффективности применения Бод в реальном секторе экономики как за границей, так и в нашей стране. По данным исследования [10], каждая третья крупная российская компания не менее трёх лет использует в своей работе технологии Бод. Среди основных направлений, в которых реализуются соответствующие проекты, можно назвать: маркетинг, розничную торговлю, банковский и страховой сектор, машиностроение, предприятия ТЭК, медицинскую и фармацевтическую отрасль. По оценке Ассоциации больших данных (Россия), к 2024 г. ожидается десятикратный рост отечественного рынка в сравнении с 2022 г. [11]. Таким образом, рынок Бод в нашей стране уже сформирован. Это направление начинает развиваться и в классических отраслях, имеющих дело, например, с разведкой, поиском и добычей углеводородов и твёрдых полезных ископаемых, горно-перерабатывающим производством, управлением и развитием железнодорожного транспорта, электроэнергетикой.

Работу с Бод затрудняет ряд узких мест (рис. 4). Самая большая трудность — нехватка кадров для ведения таких проектов. Компании вынуждены привлекать внешних специалистов, что связано с определёнными рисками, а также обучать собственных сотрудников, на что требуется время. С этой проблемой сталкиваются научно-исследовательский и университетский секторы, что обусловлено как недостаточным выпуском IT-специалистов, так и прямой зависимостью зарплаты научного персонала от публикационной активности. Нехватка квалифицированных кадров, помимо прочего, затрудняет подбор подходящей под задачи компании архитектуры и инструментов Бод.

Следует сказать и о неполном понимании ценности работы с Бод внутри организации. Интеграция результатов таких проектов в операционную деятельность компаний есть их конечная цель. Однако зачастую сотрудники не готовы принимать решения, опираясь на выводы, сделанные искусственным интеллектом на основе Бод. Аргументом здесь является недостаток понимания всей широты развития ситуации, что обусловлено восприятием Бод как ограниченного среза — выборки, а не постоянно возрастающего потока информации, меняющей представления об окружающем мире.

Сегодня проекты в сфере Бод разрабатывают несколько институтов РАН. Так, Институт проблем комплексного освоения недр им. академика



Рис. 4. Основные затруднения при работе с Бод (по данным [10])

Н.В. Мельникова РАН под руководством академика РАН В.Н. Захарова реализует проект РНФ “Решение фундаментальной проблемы сбора, хранения и анализа больших данных для развития методов прогнозной аналитики при управлении горно-техническими системами”. Пермский федеральный исследовательский центр УрО РАН под руководством академика РАН А.А. Баряха изучает возможность использования технологий Бод для повышения безопасности ведения горных работ. В Геофизическом центре РАН под руководством академика РАН А.Д. Гвишиани совместно с Институтом физики Земли им. О.Ю. Шмидта РАН ведутся фундаментальные исследования в рамках государственного задания по созданию дата-центра наук о Земле. Заинтересованность в работе с большими данными во взаимодействии с Академией наук и её институтами проявляют такие крупные представители реального сектора экономики, как ПАО “Уралкалий”, АО “НИИАС” (РЖД), компания “Металлинвест”, АК “АЛРОСА” (ПАО), ПАО “Газпром”, “Национальная компьютерная корпорация” и др.

В настоящее время Российская академия наук является наиболее крупной и важной экспертной организацией нашей страны, осуществляющей фундаментальную научную и научно-техническую экспертизу. Экспертиза РАН охватывает программы, проекты и отчёты российских вузов и научно-исследовательских институтов РАН, других министерств и ведомств, реального сектора экономики. Для эффективной работы с такими объёмами информации, которую нужно собирать, хранить и оперативно анализировать, необходимо создать в РАН современный дата-центр больших данных. Такой центр обеспечит академию постоянный доступ к материалам заявок, проектов и отчётов, что позволит выявлять тенденции исследований и оценивать их полноту. Кураторами направлений Бод по соответствующим областям науки могли бы стать научные со-

веты РАН. Например, Научный совет РАН по изучению Арктики и Антарктики уже осуществляет методологическую поддержку и экспертное сопровождение крупнейших арктических проектов.

Большие данные можно рассматривать как переход от простого накопления информации к её эффективному использованию в качестве интеллектуального капитала. Этот феномен становится особенно важным в условиях растущей сложности и неопределённости современного мира, когда принятие верных решений может зависеть от многих факторов и требовать гигантского объёма информации. В этом контексте развитие тематики Бод может стать ключевой задачей как для организаций реального сектора экономики, так и для научно-исследовательских институтов и университетов, повышая их конкурентоспособность и эффективность.

ИСТОЧНИКИ ФИНАНСИРОВАНИЯ

Работа выполнена в рамках государственного задания № 075-01349-23-00 Геофизического центра РАН, утверждённого Минобрнауки России.

ЛИТЕРАТУРА

1. *Lynch C.* Big data: How do your data grow? // *Nature*. 2008. № 455. Р. 28–29.
2. *David R., Gantz J., Rydning J.* The Digitization of the World // *From Edge to Core An IDC White Paper*. #US44413318. <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf> (дата обращения 14.02.2022).
3. What Happens in an Internet Minute in 2023: 90 Fascinating Online Stats. <https://localiq.com/blog/what-happens-in-an-internet-minute/> (дата обращения 11.05.2023).
4. Data Age 2025: The Evolution of Data to Life-Critical. Don't Focus on Big Data; Focus on the Data That's Big. IDC White Paper. Doc# US44413318 I November 2017. <https://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf> (дата обращения 11.05.2023).
5. ГОСТ 8.417-2002 Государственная система обеспечения единства измерений. Единицы величин. 2013.
6. *Майер-Шенбергер В., Кукьер К.* Большие данные. Революция, которая изменит то, как мы живём, работаем и мыслим. М.: Манн, Иванов и Фербер, 2014.
7. *Dean J., Ghemawat S.* MapReduce: Simplified Data Processing on Large Clusters // *OSDI'04: Sixth Symposium on Operating System Design and Implementation*. San Francisco, CA, 2004. Р. 137–150.
8. *Гвишиани А.Д., Добровольский М.Н., Дзеранов Б.В., Дзедоев Б.А.* Большие данные в геофизике и других науках о Земле // *Физика Земли*. 2022. № 1. С. 3–34. <https://doi.org/10.31857/S0002333722010033>
9. *Захаров В.Н., Гвишиани А.Д., Вайсберг Л.А., Дзеранов Б.В.* Большие данные и устойчивое функционирование горнотехнических систем // *Горный журнал*. 2021. № 11. С. 45–52. <https://doi.org/10.17580/gzh.2021.11.06>
10. Совместное исследование VK CLOUD и ARENA-DATA: Технологии работы с Big Data: готовность к использованию и основные барьеры. 2022. Москва. © VK Cloud
11. *Паньков В.* Государство может простимулировать предиктивные подходы в бизнесе // Тематическое приложение к ежедневной газете РБК. 15 апреля 2021. № 52 (3341).