



БОЛЬШИЕ ДАННЫЕ (BIG DATA) КАК ГЛАВНЫЙ РЕСУРС ЦИФРОВОЙ ЭКОНОМИКИ

М.Т.Бутабоев

Фергана, Узбекистан,
Ферганский Политехнический институт, Д.Э.Н. профессор

MAQOLA HAQIDA

АННОТАЦИЯ

Qabul qilindi: 24-iyun 2023-yil
Tasdiqlandi: 26-iyun 2023-yil
Jurnal soni: 7
Maqola raqami: 2
DOI: <https://doi.org/10.54613/ku.v7i7.768>

В данной статье мы рассматриваем стремительно развивающуюся отрасль цифровой экономики – большие данные, которые в современном мире позволяют правительствам, бизнесу и социальным структурам превращать тонны гетерогенной информации в структурированное знание. Обсуждается опыт развития науки о больших данных в лидирующих странах мира и на их примере рассматривается возможность применения больших данных в Узбекистане.

KALIT SO'ZLAR/ Ключевые слова/
keywords

цифровая экономика, цифровизация, Big Data, полу структурированные и неструктурированные данные, блокчейн, поставщик инфраструктуры, датамайнер, системный интегратор, разработчик готовых сервисов, облака.

Введение: Эпоха информационной революции затронула все сферы бизнеса и каждого человека на земле. Огромные потоки данных и информации создают больше экономической ценности, чем мировая торговля товарами. Бизнес идеи, построенные на данных, рожают новые возможности для экономического роста и решения проблем общества.

Каждый этап развития характеризовался определенным ресурсом, так основным ресурсом в эпоху сельскохозяйственной цивилизации была земля, а в эпоху индустриальной цивилизации – сырье, топливо.

В эпоху цифровой экономики основным ресурсом стали данные и профессиональная компетенция человека. Клайв Хамби, британский математик и архитектор карточки Clubcard компании Tesco одним из первых сравнил данные с новой нефтью. Нефть ценна, но в нерафинированном виде ее нельзя использовать. Как нефть нужно превратить в газ, пластик, химикаты и т.д., чтобы создать ценное вещество, способствующее прибыльной деятельности; так и данные нужно разложить на части, проанализировать, чтобы они имели ценность". [3]

Один из авторитетных онлайн-ресурсов Wired также сравнил информацию и ее анализ с нефтью. В XVIII веке не все понимали потенциальную прибыльность добычи нефти, однако те, кто понял, сумели сделать на этом состоянии. То же нас ждет и с информацией. [4]

Генеральный секретарь КНР Си Цзиньпин во время своего визита в Китайскую академию наук в 2013 году отметил: «Большие данные – это «бесплатный» ресурс в индустриальном обществе. Тот, кто контролирует данные, получит преимущество».

Сущность Big Data

Чтобы оставаться конкурентоспособными, компаниям необходимо извлекать информацию из огромных и разнообразных массивов данных. Тем компаниям, которые научились анализировать и использовать большие данные, обещано светлое будущее.

Исследования цифровой экономики начинаются с больших данных (Big data). Клиффорд Линч в своей статье «Как ваши данные растут?» [5] отмечал, что происходит взрывной рост мировых объемов информации и бизнесу необходимо воспользоваться возможностями, которые предоставляют большие массивы данных. Это значит, что необходимо постоянно собирать, изучать, упорядочивать, трансформировать, хранить данные с целью выявления полезных сведений, выводов, выдвижения свежих идей, обоснования новых решений.

Большие данные – это сочетание опробованных и вновь внедряемых технологий, которое позволяет компаниям извлекать из имеющихся данных информацию для

использования их в деятельности. Большие данные можно определить, как способность управлять большими объемами разнородных данных со скоростью, достаточной для анализа таких данных в реальном времени и своевременного реагирования. [6][7]

Источниками больших данных являются социальные сети, события, связанные с действиями пользователей в различных приложениях, потоки транзакций банковских платежей, данные из мира «Интернета вещей». Все это многообразие данных можно разделить на структурированные, полуструктурированные и неструктурированные данные. 20% от общего объема данных составляют структурированные данные, которые имеют определенную длину и формат. Это числа, даты или комбинации слов и чисел, например, как имя и адрес клиента. Неструктурированные данные составляют 80% от общего объема данных и не имеют определенного формата. К ним можно отнести записи в соцсетях, фото, наполнение веб-сайтов, спутниковые снимки, записи с камер наблюдения и видеорегистраторов.

Необходимость хранить и обрабатывать постоянно возрастающие объемы больших данных с целью извлечения из них полезной информации привела к появлению новых технологий хранения и методов аналитики больших данных. Такие инновационные технологии как Map Reduce, Hadoop и BigTable послужили толчком к появлению нового поколения систем управления данными и направлены на быстрое обнаружение новых знаний в данных и определение действий, продуктов и услуг, которые можно извлечь из добытых знаний.

Совокупность методов обнаружения в данных неочевидных, объективных, практически полезных и доступных для понимания человеком закономерностей, для принятия решений - это глубокий анализ (Data Mining).

Технология Text Mining охватывает новые методы для выполнения семантического анализа текстов, информационного поиска и управления, в том числе исследования тональности. Может использоваться для анализа больших и сверхбольших массивов неструктурированной информации.

Технология Call Mining объединяет в себя распознавание речи, ее анализ и Data Mining. Ее цель – упрощение поиска в аудио-архивах, содержащих записи переговоров между операторами и клиентами.

Технология Web Mining охватывает методы, которые способны на основе данных сайта обнаружить новые, ранее неизвестные знания и которые в дальнейшем можно будет использовать на практике. Технология Web Mining применяет технологию Data Mining для анализа неструктурированной,

неоднородной, распределенной и значительной по объему информации, содержащейся в Web-узлах.

Согласно характеристикам, которые компания Meta Group определила для больших данных, они имеют следующие особенности:

- объем (volume) - от 150 Гб в сутки;
- скорость (velocity) - непрерывное создание большого объема данных предъявляет высокие требования к скорости их обработки;
- разнообразие (variety) - современные технические средства позволяют анализировать любую структурированную и неструктурированную информацию;
- достоверность (veracity) - достоверность данных и результатов их анализа;
- изменчивость (variability) - не всегда потоки данных бывают стабильными, изменчивый поток данных сложнее поддается анализу;

- ценность (value) - ценность и значимость информации. Большие данные могут быть простыми (посты в соцсетях) или сложными для восприятия и анализа (банковские транзакции).

Отчеты об исследовании рынка показали, что объем данных растет с невероятной скоростью. По данным компании IBM к 2003 году мир накопил 5 эксабайтов данных, к 2008 году этот объем возрос до 0,18 зеттабайта, в 2013 году показатель составил 4,4 зеттабайтов. По данным Statista, объем возрос до 64,2 зеттабайт в 2020 году и 79 зеттабайт в 2021 году, до 2025 года прогнозируется рост более чем до 180 зеттабайт. В 2035 году он будет расти в геометрической прогрессии до более чем 2100 зеттабайт.



Рис. 1 Динамика роста объема данных.

Достижение объема созданных и синтезированных данных в 2020 году до нового максимума было вызвано увеличением спроса в связи с пандемией COVID-19. Кризис, вызванный распространением нового коронавируса, активизировал переход от работы в режиме офлайн к постоянному онлайн, т.е. все больше людей работали и учились из дома и чаще пользовались домашними развлечениями.

Эффективное и грамотное использование Big Data дает огромные преимущества. Обработка больших данных позволяет сделать анализ предсказательным и предложить бизнесу рекомендации на будущее. Сегодня Big Data используется в государственном управлении, промышленности, медицине, торговле, сельском хозяйстве, рынке недвижимости, спорте, туризме, транспорте, строительстве. Анализ больших данных в государственном управлении помогает принимать решения, касающиеся занятости населения, экономического регулирования, обеспечения безопасности. В медицине обработка собранных массивов данных позволяет выявлять эпидемии, оперативно принимать решения в экстренных случаях, ставить точные диагнозы, находить новые лекарства. В промышленности внедрять систему предиктивного производства для прогнозирования более точного спроса на продукцию. В сельском хозяйстве, например, используя метод точного земледелия обрабатывать информацию, собранную с сенсоров специальных метеостанций, которые стоят в полях, строить высокоточный почасовой прогноз погоды. Девелоперы, работающие на рынке недвижимости с помощью технологий Big Data анализируют потоки информации и предоставляют клиентам наиболее интересные для них варианты. Таким образом покупатель может познакомиться с понравившимся ему домом уже без продавца. В спорте обработка больших данных позволяет отбирать самых перспективных спортсменов и разрабатывать эффективные стратегии как для спортсменов, так и для каждого противника. В бизнесе Big Data предоставляет такие преимущества как упрощение процесса

планирования, увеличение скорости запуска и востребованности новых проектов, оптимизация интеграции в цепи поставок, рост целевой аудитории, повышение качества клиентского сервиса, что в свою очередь приводит к лояльности текущих клиентов.

BARC, сообщает, что организации пожинают плоды использования Больших Данных - 69% вероятность принятия лучших стратегических решений, 54% вероятность улучшения контроля операционных процессов, 52% для лучшего понимания потребителей, а также 47% для эффективного снижения затрат. Организации, которые пользуются преимуществами Больших Данных, сообщают о среднем увеличении доходов на 8% при сокращении расходов на 10%.

Структура Больших данных.

Система Big Data включает в себя поставщиков инфраструктуры, датамайнеров, системных интеграторов, потребителей и разработчиков готовых сервисов. Рассмотрим каждый элемент данной системы.

Поставщики инфраструктуры. Одной из проблем, связанных с большими данными, является их хранение. Поставщики инфраструктуры решают задачи хранения и преобразования данных. Существует три основных способа хранения данных: традиционный, в публичных облаках и в частных облаках.

При традиционном («хранение у себя») способе информация записывается на диски, ленты или локальные хранилища. При этом способе данные всегда рядом, поэтому это привычно и спокойнее, к локальным носителям можно легко и быстро подключиться и по цене это обычно приемлемо. Но они ненадежны, так как диски и серверы из-за физического износа могут выйти из строя, к ним нельзя получить доступ, если ты находишься вдалеке, и возможности носителей ограничены (нужно покупать новые носители и где-то их размещать).

Хранение данных за определенную плату, зависящую от объема данных и сопутствующих услуг, предоставляют публичные облака. В пятерку стран, предоставляющих

облачные технологии входят США, Китай, Великобритания, Германия и Япония (рис.2). Одной из первых компаний, которая стала зарабатывать совершенно новым образом, предоставляя свои неиспользуемые вычислительные ресурсы была Amazon. Сегодня крупнейшими облачными провайдерами являются Amazon Web Services (AWS), Microsoft Azure, Google Cloud, IBM Bluemix, Oracle. Лидером на мировом

рынке хранения больших данных является компания Amazon. Китайские компании быстро догоняют зарубежных конкурентов в гонке облачных услуг. Лидерами на рынке облачных услуг Китая являются Alibaba (Alibaba Cloud), Tencent (Tencent Cloud), Huawei (Huawei Cloud) Baidu (Baidu Cloud).



Рис.2 Доходы стран лидеров, предоставляющих облачные технологии. [8]

Если говорить о преимуществах публичных облаков, можно выделить следующие параметры:

- удобство (компаниям нет необходимости в покупке и обслуживании физических серверов, что максимально упрощает базовые сценарии работы);
- относительная безопасность (большинство провайдеров предоставляют защиту данных собственными алгоритмами шифрования);
- приемлемые цены (цены колебались на уровне 5-10 центов за гигабайт в месяц, а в настоящее время идет тенденция к их снижению);
- относительная надежность (в случае утери есть возможность реплицирования данных);

- новые возможности в будущем.

К недостаткам публичных облаков можно отнести психологический фактор (ощущение того, что к вашим данным имеет доступ еще кто-то), скорость доступа (скорость доступа в интернет измеряется мегабайтами в секунду, что в десятки раз медленнее доступа в локальные хранилища), цены (облачное хранение дороже локального).

Частные облака – это хранилища, которые являются частью инфраструктуры компаний и используются только ее сотрудниками. Эти хранилища используются корпоративным сектором.

В настоящее время скандинавские компании лидируют по внедрению корпоративных облачных услуг в Европе (рис.3)

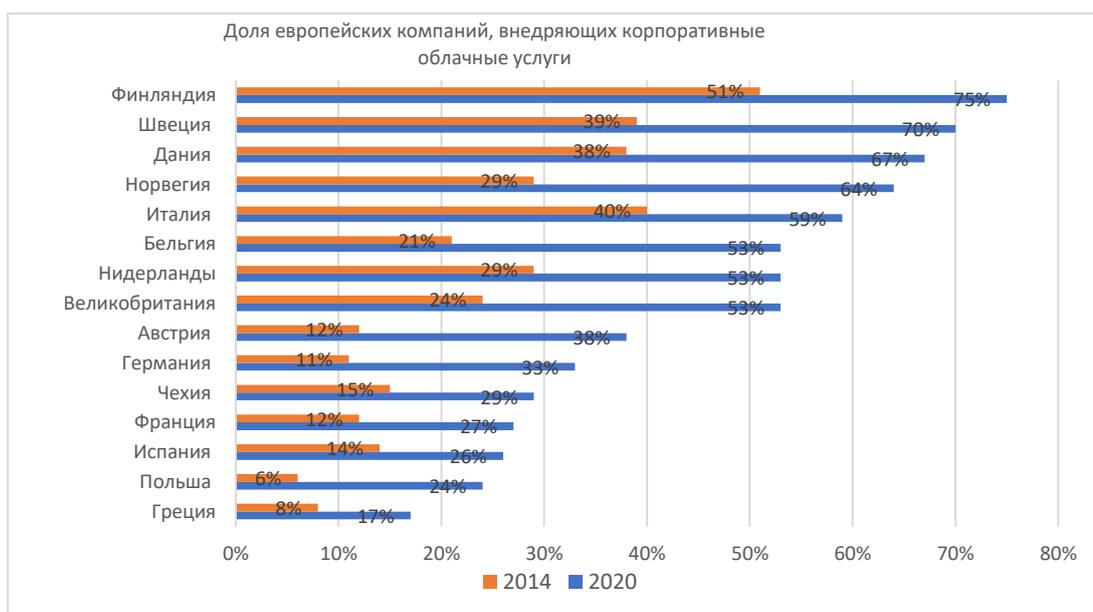


Рис. 3. Доля компаний, внедряющих корпоративные облачные услуги в отдельных европейских странах [8].

Системные интеграторы – это компании или физические лица, которые предлагают решения для автоматизации бизнес-процессов. На основе разработанного проекта они связывают автоматизированные системы друг с другом, что обеспечивает эффективную работу команды или всей организации в целом.

Согласно глоссарию компании [Gartner](#) работа системного интегратора заключается в наведении порядка в системах, поставляемых разрозненными поставщиками. В частности, компании — системные интеграторы занимаются: внедрением, планированием, координацией, составлением графиков, тестированием, улучшением, обслуживанием информационных систем.

Датамайнеры – специалисты, занимающиеся поиском, интеллектуальным и глубинным анализом данных. На основе анализа можно получить представление о зависимости данных, а на основе этих связей выдвинуть предположения для построения нейросетевых моделей. Например, анализ поведения покупателей магазинов бытовой техники может показать, что холодильники люди предпочитают покупать в кредит, покупая компьютер клиенты приобретают и принтеры. Такие зависимости в дальнейшем помогают принять более эффективные управленческие решения и увеличить уровень продаж.

Разработчики готовых сервисов разрабатывают готовые решения на основе доступа к большим данным. Например, приложение Value Arpr предлагает предпринимателям готовую к использованию экономически эффективную платформу для создания приложений и веб-решений, которая обеспечит цифровую трансформацию их бизнеса. Например, платформа предлагает разработку приложений для различных видов бизнеса, таких как доставка на заказ (еды, продуктов, лекарств), бытовые услуги (ремонтные работы, косметические процедуры на дому, мойка машин, выгул собак), онлайн консультации (врача, диетолога, преподавателей, фитнес тренеров), онлайн магазины (ювелирные изделия, одежда, книги, подарки), которая разработана с учетом требований различных бизнес-вертикалей. Предпринимателю предоставляется высоконадежное решение для привлечения клиентов, максимизации прибыли и опережения конкурентов в выбранной сфере.

Перспективы развития Big Data в Узбекистане

Вклад цифровой экономики в валовой внутренний продукт США составляет 10,9%, в Китае - 21%, в Индии - 5,5%. В Узбекистане этот показатель не превышает 2 процентов.

2020 год в Узбекистане был объявлен Годом развития науки, просвещения и цифровой экономики. В целях формирования и развития цифровой экономики нашего государства в 2020 году была утверждена стратегия «Цифровой Узбекистан - 2030». Стратегия предусматривает проведение прорывных мероприятий по таким приоритетным направлениям как цифровая инфраструктура, электронное правительство, цифровая экономика, национальный рынок цифровых технологий, подготовка специалистов в сфере информационных технологий. На 2020-2022 годы намечены мероприятия, направленные на повышение уровня охвата подключения населенных пунктов к сети Интернет, прокладка 20 тысяч километров оптоволоконных связей, внедрение более 400 информационных систем для социально-экономического развития регионов, внедрение свыше 280 информационных систем для автоматизации процессов управления, реализация программы

Использованная литература

1. Ш.Мирзиёв. Янги Ўзбекистон стратегияси// Т: «Ўқитувчи», – 2021, – 657 с.
2. Бутабоев М., Мулайдинов Ф., Захидов Г., Саттарова Х. Ракамли иқтисодиёт// Т: «Инновацион ривожланиш нашриёт матбаа уйи», - 2021, - 598 с.
3. [Michael Haupt](#). “Data is the New Oil” – A Ludicrous Proposition. [электронный ресурс] - <https://medium.com/project-2030/data-is-the-new-oil-a-ludicrous-proposition-1d91bba4f294>
4. Денис Ламехов. Почему для ритейлеров информация – это новая нефть. - [электронный ресурс] - <https://www.be-in.ru/news/35044-data-is-the-new-oil/>.

«Один миллион программистов», в рамках которой запланировано обучение 500 тысяч юношей и девушек основам компьютерного программирования, реализация проекта «Цифровое ведомство».

Внедрение системы «Электронного правительства», через которую налажено оказание 178 наименований услуг через единый портал интерактивных государственных услуг обеспечивает открытость, прозрачность, экономит время и снижает расходы населения.

В рамках стратегии «Цифровой Узбекистан – 2030» были определены 13 образцовых городов и районов, в которых на основе разработанных «дорожных карт» в течение трех лет будут проводиться работы по цифровой трансформации. В течение 2020-2021 годов в соответствии с «Дорожными картами» будут реализованы 172 проекта по расширению цифровой инфраструктуры, 54 проекта по внедрению цифровых технологий в государственное управление, 190 проектов в социальной сфере, 45 – в жилищно-коммунальном хозяйстве, 51 - в сельском хозяйстве, 17 в промышленности, 58 проектов направлены на обучение руководителей и сотрудников.

Государственный налоговый комитет республики был одним из первых, запустивших крупный проект, предусматривающий усовершенствование действующего комплекса информационных систем налоговых органов с учетом реинжиниринга бизнес-процессов налогового администрирования. Проект стал основой для автоматизированной системы контроля за уплатой НДС, целью которого стал мониторинг уплаты НДС в режиме реального времени.

В стратегии развития нового Узбекистана, предложенной Президентом Республики Ш.Мирзиёвым одним из ключевых моментов развития нового Узбекистана является развитие цифровой экономики в таких сферах как связь и телекоммуникации, информационная безопасность, образование, здравоохранение, финансовое и банковское дело, строительство, сельское и водное хозяйство, энергетика, транспорт, машиностроение и др. Узбекистан уже упускает финансово-экономическую выгоду в 5-7% по отношению к ВВП страны ежегодно. Внедрение таких технологий как блокчейн, бигдата, облачных вычислений, искусственного интеллекта приведет к росту ВВП на 30% и росту реальных доходов населения до 45 млн.сум в год к 2030 году, долгосрочной перспективе данные показатели могут вырасти до 50% к ВВП и 50 млн.сум в год в 2040-2050 годах.

Заключение: Big Data, стремительно развивающиеся отрасль цифровой экономики в ответ на реактивное увеличение объёма разнородной информации в мире, превратились в один из самых мощных инструментов прогнозирования, регулирования и реагирования в глобальном масштабе. Эффективное управление большими данными становится крайне ценным в получении прибыли больших информационных компаний и стран, уделяющим этому особое внимание. Зачатки для использования больших данных в Узбекистане уже созданы, совершенствование же этой технологии может стать сильнейшим толчком для развития цифровой экономики в государстве, которая в свою очередь имеет потенциал увеличить ВВП страны до десяти процентных пункта.

5. Lynch C. How do your data grow? //Nature. – 2008. – Т. 455. – №. 7209. – С. 28-29.
6. Соколова А. Мир bigdata в 8 терминах / (Электронный ресурс) Режим доступа: <https://rb.ru/howto/big-data-in-8-terms/>;
7. Соколова А. Как устроен рынок bigdata в России/ (Электронный ресурс) Режим доступа: <https://rb.ru/howto/big-data-in-russia/>
8. Digital economy compass 2021 [Электронный ресурс]. <https://www.statista.com/study/105653/digital-economy-compass>