

## БОЛЬШИЕ ДАННЫЕ (BIG DATA) КАК ГЛАВНЫЙ РЕСУРС ЦИФРОВОЙ ЭКОНОМИКИ

**Бутабоев Мухаммаджон Гуйчиевич**

Ферганский Политехнический институт, профессор кафедры “Экономика”, д.э.н.

**Исманов Иброхим Набиевич**

Ферганский политехнический институт, зав. кафедрой “Бухгалтерский учет и аудит”, д.э.н., профессор

[i.ismanov@ferpi.uz](mailto:i.ismanov@ferpi.uz)

**Нишанова Икболой**

Ферганский политехнический институт, магистрант

### *Аннотация*

*концепция цифровой экономики стала новой концепцией для всего мира, и этот период продолжает оказывать влияние на все сферы бизнеса. В этом исследовании говорится, что ценность информации, создает большую элегантность, чем себестоимость товара. Автор высказывает свое мнение о роли больших баз данных в экономике различных стран, в том числе Узбекистана.*

### *Ключевые слова*

*информационной революции, эпоха цифровой экономики, большие данные, анализ больших данных, объёмов информации, хранение данных.*

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.

Чтобы оставаться конкурентоспособными, компаниям необходимо извлекать информацию из огромных и разнообразных массивов данных. Тем компаниям, которые научились анализировать и использовать большие данные, обеспечено 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 объёмов 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, полу структурированные 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.

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.

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 оффлайн 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.

Система Big Data включает в себя поставщиков инфраструктуры, датамайнеров, системных интеграторов, потребителей и разработчиков готовых сервисов.

*Поставщики инфраструктуры.* Одной из проблем, связанных с большими данными, является их хранение. Поставщики инфраструктуры решают задачи хранения и предобработки данных [2]. 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.

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 [3]. Лидером на мировом рынке хранения больших данных является компания Amazon [4]. 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) [10].



**Рисунок 1. Доходы стран лидеров, предоставляющих облачные технологии [8]**

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

*Системные интеграторы* – это компании или физические лица, которые предлагают решения для автоматизации бизнес-процессов.

*Датамайнеры* – специалисты, занимающиеся поиском, интеллектуальным и глубинным анализом данных.

*Разработчики готовых сервисов* разрабатывают готовые решения на основе доступа к большим данным.

#### **Типы больших данных.**

В многообразии данных можно выделить следующие:

- ✓ структурированные данные;
- ✓ полу структурированные данные;
- ✓ неструктурированные данные.

Под структурированными данными понимаются данные, которые имеют определенную длину и формат. Примерами таких данных являются числа, даты, а также комбинации слов и чисел (например, имя адрес клиента и.т.д.)

Сегодня (2022г) по мнению экспертов, структурированные данные составляют около 20 % от общего объема данных. Структурированные данные хранятся в базе данных, к которой можно делать запросы с помощью специального языка программирования, например SQL (язык структурированных запросов). Ныне 80% данные составляет не структурированные которые не имеют определенного формата (записи в соц. сетях, фото наполнение веб-сайтов.)

Не структурированная информация включает в себя видео, аудио файлы, свободный текст, информацию, поступающую из социальных сетей.

Данная информация нуждается в специальном анализе, чтобы сделать ее полезной для дальнейшей обработки.

#### **Источники данных делятся на две категории:**

- данные, генерируемые машинами;

Это данные, создаваемые компьютерами без вмешательства человека например спутниковые снимки, записи с камер наблюдения, видео регистров

- данные, генерируемые человеком,

Это данные, которые создаются человеком при взаимодействии с компьютером. Так данные о посещении веб-сайтов, игровые данные, мобильные данные, данные социальных сетей.

При необходимости хранить большие объемы данных и осуществлять к ним доступ, возможны два варианта:

- 1) Приобрести более мощный компьютер с большим количеством процессоров, оперативной памяти, дискового пространства и т.д.
- 2) Приобрести сотни и даже тысячи маломощных компьютеров, объединенных в кластер. Здесь можно приложить для офлайн-обработки или онлайн-обработки по запросу.

Например платформы для больших данных компании Oracle.

Сбор обработки хранения анализ передачи данные.

Например Сбербанк (Россия) занимает от 40 % до 90 % большие данные финансовые услуги в регионе. Банк анализирует данные по 130 миллион частных и 2 миллиона корпоративных клиентов.

«Большие данные» Сбербанка впервые становится доступны всем желающим.

Информация о доходах и расходах, сбережений граждан, текущие тенденции о розничной торговле, недвижимости, финансах и другие.

По мнению американских исследователей Э. Бриньольфссона и Э.Макафи, люди пока еще намного превосходят машины в навыках трех видов: творчество уровня; эмоции, общение, забота, воспитание, лидерство и т. д.; сноровка и мобильность (в частности аккуратное перемещение в пространстве).

Одновременно многоуровневое сетевое взаимодействие компаний, а также оборудования и различной техники порождает проблему разработки стандартов такого взаимодействия (стандарты обмена данными, их хранения, идентификации и пр.), выводя на новый уровень сложности проблему цифрового доверия и безопасности (кибербезопасности).

Решая проблему безопасности онлайн-платежей, компания Alibaba создала сервис Alipay, который предоставляет услугу депонирования и систему оценки продавцов, заложив тем самым основы прозрачности и доверия, что ускорило распространение интернет торговли в Китае.

Далее широкое распространение различных датчиков и устройств M2M в условиях цифровизации всех отраслей экономики порождает проблему мониторинга и оценки эффективности работы этих датчиков и устройств, поддержки их работоспособности.

Но поскольку датчики устанавливаются на сложных продуктах (сооружения, оборудование, бытовая техника и пр.), то важным фактором конкурентоспособности бизнеса может стать его умение обеспечивать целостное управление сложными продуктами на протяжении всего цикла его жизни – от разработки, производства, продажи и установки до утилизации (модель PLM).

Видимо, это далеко не полный перечень последствий и проблем, связанных с цифровизацией экономики и компаний.

Информационный взрыв произошел уже давно, однако значительно позже появились технологические возможности хранения и обработки возрастающего информационного потока (уже говорят об информационном потоке) структурированных, а главное – неструктурированных данных. Лишь в 2008 г. была сформулирована проблема извлечения знаний, доступных для восприятия человеком, из больших массивов цифровой информации, которая получила название «большие данные» (Big Data). Big Data – это многоплановое и многоуровневое понятие, основными характеристиками которого являются объем данных, скорость их генерации и доступности, разнообразие источников и типов данных, форм их хранения. Также важной характеристикой является неструктурированность или плохая структурированность генерируемых данных (например, данных мобильной связи, мессенджеров и других интернет-сервисов), которые нужно интерпретировать в понятных для пользователя терминах.

По сути, проблемы Big Data связаны с генерацией, доставкой, хранением, анализом и интерпретацией данных и являются обобщением решений класса Business Intelligence, включая поиск знаний в базах данных (Data Mining), технологии машинного обучения и другие технологии, прогнозную аналитику и пр.

В рамках этого тренда с 2013 г. В вузовских образовательных программах появилось направление Data Science.

В настоящее время многие исследователи говорят о том, что все данные стали большими, поэтому Big Data рассматривается как инструмент работы с данными, алгоритмы извлечения знаний из этих данных, которые меняют многие принципы управления и работы с потребителями, способствуют созданию новых цепочек формирования стоимости. Сейчас проблема перемещается от Big Data в сторону Fast Data – обработки потока данных в реальном масштабе времени, а также Smart Data – интеллектуальных данных, которые помогают понять суть или смысл проблемы, явления, сообщения на основе семантического анализа.

Также важной проблемой является монетизация аналитики этих данных и развитие технологий управления на базе этой аналитики. По мере развития цифровых технологий мир бизнеса будет вынужден действовать быстрее и разумнее, повышая скорость анализа данных и темпы действий.

В сфере аналитики данных выделяют несколько последовательных стадий;

- 1) Descriptive analytics – дескриптивная, описательная аналитика;
- 2) Descriptive analytics -аналитика, связанная с распознаванием образов, определением того что представляет собой объект анализа или воздействия;
- 3) Predictive analytics - предиктивная, предсказательная, прогнозная аналитика. Как следует из названия, Predictive analytics ориентирована на формирование прогноза данных или событий будущего на основе использования методов математической статистики, анализа данных, теории игр, моделирования (выдвигается гипотеза и проверяется, что произойдет, если...?) и других методов обработки накопленных данных.

Такая аналитика широко распространена в сфере страхования, в банках (оценка платежеспособности заемщиков), на фондовом рынке и в других областях. Подробно Predictive analytics рассматривается в книге Эрика Сигеля «Просчитать будущее (Predictive analytics)»;

4) Predictive analytics – прескриптивная, предписывающая аналитика. Она позволяет предупреждать о каких-то явлениях, выдавать рекомендации к действию, предписания. Как правило, Predictive analytics основана на статистике прошлого и информации, получаемой в режиме реального времени. В рамках этого подхода строятся модели будущего и в зависимости от полученных результатов вырабатываются решения, которые принимаются без участия человека (например, делается рассылка предложений потребителям), либо эти предписывающие решения передаются человеку (решения по ремонту оборудования и т.п.).

Классическим примером предписывающей аналитики, который приводится во многих книгах, стала ситуация с юной покупательницей американского розничного гиганта Target, который по хаотичным покупкам школьницы вычислил, что она беременна, и послал ей спецпредложение одежды для будущих мам, о чем в феврале 2012 г. Написало издание New York Times Magazine в статье «Как компании изучают ваши секреты».

Считается, что примером прескриптивной аналитики в действии служат разделы «Рекомендуемые товары» на сайте Amazon или «Возможно вы их знаете» на Facebook. Этот инструмент позволяет собрать информацию и подготовить клиенту персонализированное предложение, что несомненно, может стать конкурентным преимуществом.

В промышленной сфере программное решение для аналитики технического обслуживания от американской компании Aspen Tech предупреждает о потенциальных сбоях на самых ранних стадиях эксплуатации оборудования. Программа не только информирует о том, когда произойдет сбой, но также указывает причины его возникновения, предоставляя рекомендации по тому, что можно сделать, чтобы его избежать или смягчить возможные последствия.

Новая форма – прескриптивное (предписывающее) обслуживание оборудования позволяет прогнозировать будущие отказы и предупреждать их, предписывая корректирующие действия.

### **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» [1]. 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 оптико-волоконных связей, внедрение более 400 информационных систем для социально-экономического развития регионов, внедрение свыше 280 информационных систем для автоматизации процессов управления, реализация программы «Один миллион программистов», в рамках которой запланировано обучение 500 тысяч юношей и девушек основам компьютерного программирования, реализация проекта «Цифровое ведомство».

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 [9]. 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.

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.

### **Список использованной литературы**

1. Ш.Мирзиёв. Янги Ўзбекистон стратегияси// Т: «Ўқитувчи», - 2021, - 657 с.
2. Академия Big Data. Введение в аналитику больших массивов данных. М.: 2021, 400 с.
3. Бутабоев М., Мулайдинов Ф., Захидов Г., Саттарова Х. Рақамли иқтисодиёт// Т: «Инновацион ривожланиш нашриёт матбаа уйи», - 2021, - 598 с.
4. Голиков М. Развитие больших данных в России. М.: 2022, 113 с.
5. 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>
6. Денис Ламехов. Почему для ритейлеров информация – это новая нефть.- [электронный ресурс]-<https://www.be-in.ru/news/35044-data-is-the-new-oil>
7. Lynch C. How do your data grow? //Nature. – 2008. – Т. 455. – №. 7209. – с. 28-29.
8. Соколова А. Мир bigdata в 8 терминах / (Электронный ресурс) Режим доступа: <https://rb.ru/howto/big-data-in-8-terms/>;

9. Соколова А. Как устроен рынок bigdata в России/ (Электронный ресурс) Режим доступа: <https://rb.ru/howto/big-data-in-russia>

10. Kudbiev, D., Tursunova, D., & Qudbiyev, N. (2022). Buxgalteriya Hisobini Avtomatlashtirish Masalalari. Journal of Integrated Education and Research, 1(2), 107-112.

11. Nabievich, I. I., & Jaloliddinovich, M. E. (2022). Accounting of fixed assets according to ifrs-tasks and solutions for the organization of accounting by component parts. ACADEMICIA: An International Multidisciplinary Research Journal, 12(8), 53-60.

12. Digital economy compass 2021 [Электронный ресурс]. <https://www.statista.com/study/105653/digital-economy-compass/>

