

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ**

**Федеральное государственное бюджетное образовательное учреждение
высшего образования**

**«Уральский государственный экономический университет»
(УрГЭУ)**

КУРСОВАЯ РАБОТА

по дисциплине «Формализация информации и Big Data»

Тема: история создания и использования Big Data

Институт цифровых технологий
управления и информационной
безопасности

Направление подготовки
*Математическое обеспечение и
администрирование
информационных систем*

Направленность (профиль)
*Разработка и администрирование
информационных систем*

Кафедра
*Шахматного искусства и
компьютерной математики*

Студент
Тептин Владислав Александрович
Группа
АИС-21-1

Руководитель
д.т.н., профессор
Часовских Виктор Петрович

Екатеринбург
2025 г.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	3
ОПРЕДЕЛЕНИЕ И КЛЮЧЕВЫЕ ХАРАКТЕРИСТИКИ BIG DATA.	4
ОТЛИЧИЕ BIG DATA ОТ ТРАДИЦИОННЫХ ДАННЫХ.	5
РАЗВИТИЕ ТЕХНОЛОГИЙ ОБРАБОТКИ ДАННЫХ: ОТ РАННИХ КОНЦЕПЦИИ К BIG DATA.	6
РАННИЕ КОНЦЕПЦИИ, ИСТОКИ И ИХ ЗНАЧЕНИЕ.	6
ЭВОЛЮЦИЯ И ТРАНСФОРМАЦИЯ МЕТОДОВ СБОРА ДАННЫХ.	8
ТЕХНОЛОГИЧЕСКОЕ РАЗВИТИЕ И ХРАНЕНИЕ ДАННЫХ В ЭПОХУ BIG DATA.	10
ТЕХНОЛОГИЧЕСКИЕ ДОСТИЖЕНИЯ.	10
РАЗВИТИЕ ТЕХНОЛОГИЙ ХРАНЕНИЯ ДАННЫХ.	12
ПРИМЕНЕНИЕ BIG DATA В РАЗЛИЧНЫХ СФЕРАХ.	14
БОЛЬШИЕ ДАННЫЕ В ГОСУДАРСТВЕННОМ УПРАВЛЕНИИ.	14
БОЛЬШИЕ ДАННЫЕ В ЗДРАВООХРАНЕНИИ.	16
БОЛЬШИЕ ДАННЫЕ В БИЗНЕСЕ.	17
ПЕРСПЕКТИВЫ РАЗВИТИЯ BIG DATA.	18
ЗАКЛЮЧЕНИЕ.	19
СПИСОК ЛИТЕРАТУРЫ.	20

ВВЕДЕНИЕ

В современном мире объем данных, генерируемых каждым днем, растет в геометрической прогрессии. С развитием цифровых технологий, интернета, социальных сетей облачных сервисов и интернет вещей (IoT) возникла необходимость обрабатывать и анализировать большие массивы данных. Концепция Big Data (большие данные) стала ключевым направлением в информатике, аналитике и бизнесе, позволяя находить скрытые закономерности, прогнозировать тенденции и принимать более обоснованные решения.

История появления Big Data берет свое начало еще в середине XX века, когда люди столкнулись с необходимостью обработки больших объёмов информации. Однако настоящий прорыв произошёл в начале XXI века, вместе с развитием вычислительных мощностей, распределенных систем и алгоритмов машинного обучения.

В данной курсовой работе рассматриваются история возникновения, основные этапы развития и применение технологии Big Data в реальной жизни. Основное внимание уделяется влиянию больших данных на науку, бизнес и повседневную жизнь, а также перспективам дальнейшего развития этой области.

ОПРЕДЕЛЕНИЕ И КЛЮЧЕВЫЕ ХАРАКТЕРИСТИКИ BIG DATA.

Понятие Big Data (большие данные) появилось в конце XX – начале XXI веков и обозначает огромные объемы информации, которые невозможно обрабатывать традиционными методами. Big Data охватывает структурированные и неструктурированные данные, поступающие из множества разных источников, таких как социальные сети, сенсоры, датчики, интернет вещей (IoT), базы данных, облачные сервисы и т.д.

Широкое введение термина “большие данные” связывают с Клиффордом Линчем, редактором журнала Nature, подготовившим к 3 сентября 2008 года специальный выпуск с темой “Как могут повлиять на будущее науки технологии, открывающие возможности работы с большими объемами данных?”, в котором были собраны материалы о феномене взрывного роста объёмов и многообразия обрабатываемых данных и технологических перспективах в парадигме вероятного скачка “от количества к качеству”; термин был предложен по аналогии с расхожими в деловой англоязычной среде метафорами “большая нефть”, “большая руда”.

Изначально для описания больших данных использовались три ключевые характеристики, известные как модель 3V:

1. Volume (объём) – огромные объёмы данных, измеряемые в терабайтах и петабайтах, которые невозможно обработать с помощью традиционных методов.
2. Velocity (скорость) – высокая скорость генерации, обработка и передачи данных в режиме реального времени.
3. Variety (разнообразие) – данные могут иметь разную структуру, например:
 - Структурированные (таблицы без данных);
 - Неструктурированные (тексты, изображение, аудио, видео).

ОТЛИЧИЕ BIG DATA ОТ ТРАДИЦИОННЫХ ДАННЫХ.

Основные различия между Big Data и традиционными данными представлены в таблице 1:

Критерий	Традиционные данные	Big Data
Объём	Мегабайты - гигабайты	Петабайты - зеттабайты
Источники	Базы данных, документы	Социальные сети, сенсоры, веб-лог-файлы и т.д.
Структура	Четко структурированные данные	Неструктурированные и полуструктурированные данные
Скорость обработки	Пакетная обработка	Потоковая обработка в реальном времени
Инструменты	SQL, традиционные БД	Hadoop, Spark, NoSQL, машинное обучение

Таблица 1. Различия Big Data и традиционных данных.

Традиционные данные имеют четкую структуру, ограниченные источники и обрабатываются с помощью классических реляционных баз данных. В отличие от них, Big Data характеризуется огромными объемами, высокой скоростью поступления и разнообразием данных.

РАЗВИТИЕ ТЕХНОЛОГИЙ ОБРАБОТКИ ДАННЫХ: ОТ РАННИХ КОНЦЕПЦИИ К BIG DATA.

РАННИЕ КОНЦЕПЦИИ, ИСТОКИ И ИХ ЗНАЧЕНИЕ.

История Big Data берет свое начало за долго до появления современных технологий обработки и хранения информации. С древних времен люди стремились собирать, анализировать и систематизировать данные для управления государством, ведения торговли и научных открытий. Первые попытки систематизации данных можно проследить ещё в древнем мире, например налоговые и торговые записи в древнем Египте. Хотя в те времена объем данных был относительно мал, уже тогда осознавалась необходимость их систематизации и обработки.

С появлением механических вычислительных устройств произошёл скачок в обработке данных. Среди ключевых событий можно выделить:

- Перфокарты – перфокарты впервые начали применять в 1804 году в Жаккардовых ткацких станках для управления узорами на тканях. В информатике перфокарты впервые были применены в “интеллектуальных машинах” коллежского советника С. Н. Корсакова в 1832 году, механического устройства для информационного поиска и классификации записей. Позже эта технология легла в основу первых компьютеров.
- Табуляторы и первые вычислительные машины – появление устройств, ускоряющих обработку больших массивов данных, что особенно актуально для статистики, бухгалтерии и науки.

Эти разработки стали прообразом технологии Big Data, заложив основы для автоматизации сбора и анализа информации.

XX век ознаменовался стремительным развитием вычислительной техники и методов хранения и обработки данных:

- 1950 – 1960-е годы – появление первых электронных компьютеров, используемых для обработки больших объемов информации.

- 1970-е годы – разработка реляционных данных.
- 1980 – 1990-е годы – распространение персональных компьютеров и интернета, резкий рост объемов данных.

В этот период сложились основы современных методов работы с данными. Но традиционные базы данных уже начали сталкиваться с ограниченностью обработки больших объемов данных.

Таким образом, стремление к сбору, обработке и хранению информации существовало задолго до открытия электричества, а концепты, созданные до появления первых ЭВМ, заложили фундамент для дальнейшего развития этой области. В последствии создание персональных компьютеров и интернета кратно увеличили возможности для работы с данными, но также увеличились количество информации и потребность в новых методах её обработки.

ЭВОЛЮЦИЯ И ТРАНСФОРМАЦИЯ МЕТОДОВ СБОРА ДАННЫХ.

На протяжении веков методы сбора данных развивались в соответствии с технологическими возможностями и потребностями общества. В древние времена информации фиксировали вручную в виде записей на глиняных табличках, пергаменте или папирусе. Позже, с развитием бюрократических систем и государственных институтов, сбор данных приобрел более организованный характер.

К традиционным методам сбора данных можно отнести:

- Ручной учёт – записи в книгах, переписях, налоговых реестрах.
- Перфокарты – ранняя механизированная форма обработки данных
- Анкетирование и переписи – сбор статистических данных о населении, используемый в государственном управлении.
- Реляционные базы данных – появление в 1970-х годах позволило упростить систематизацию и поиск информации.

Эти методы были эффективны в своё время, но с ростом объёмов данных, росла потребность в новых подходах к их сбору и обработке.

Конец XX – начало XXI века ознаменовались бурным развитием цифровых технологий, что привело к изменению способов сбора информации.

Основные этапы:

- 1980 – 1990-е – развитие компьютеров и появление корпоративных баз данных, автоматизация сбора данных в бизнесе.
- 2000-е годы – широкое распространение интернета и веб-технологий, появление первых веб-аналитических инструментов.
- 2010-е годы – рост социальных сетей, интернета вещей и технологий машинного обучения, которые значительно изменили методы сбора данных.

Эволюция методов сбора данных прошла путь от ручного учета до автоматизированных цифровых технологий. Развитие интернета, облачных

сервисов и искусственного интеллекта привели к созданию глобальной инфраструктуры Big Data, обеспечивающей быстрый и эффективный сбор информации из различных источников.

ТЕХНОЛОГИЧЕСКОЕ РАЗВИТИЕ И ХРАНЕНИЕ ДАННЫХ В ЭПОХУ BIG DATA.

ТЕХНОЛОГИЧЕСКИЕ ДОСТИЖЕНИЯ.

Развитие концепции Big Data стало возможным благодаря технологическим достижениям в области вычислительной техники, сетевых технологий и программного обеспечения. Современные технологии позволяют не только собирать и хранить огромные объёмы данных, но и анализировать их в реальном времени. Рассмотрим основные инновации, сыгравшие решающую роль в эволюции методов работы с большими данными:

1. Распределенные вычисления и обработка данных.

Объём данных, создаваемых ежедневно, требует распределенных методов обработки, например таких как:

- Hadoop – одна из первых платформ для распределенной обработки больших данных, использующая файловую систему HDFS;
- Apache Spark – более быстрая альтернатива Hadoop, поддерживающая обработку данных в реальном времени.
- Google MapReduce – технология, позволяющая распределять обработку больших данных на множество серверов.

2. Облачные технологии.

Появление облачных вычислений позволило компаниям отказаться от дорогостоящей инфраструктуры и использовать внешние ресурсы для хранения и обработки данных.

3. Искусственный интеллект и машинное обучение.

Big Data тесно связано с развитием искусственного интеллекта и машинного обучения, которые позволяют анализировать данные и находить закономерности.

4. Развитие сетевых технологий.

Технологии 5G, Wi-Fi, IoT обеспечивают передачу данных в реальном времени, что играет ключевую роль в таких сферах, как например:

- Умные города;
- Промышленный IoT (анализ данных с датчиков на производстве).

5. Базы данных нового поколения.

С ростом объемов данных традиционные реляционные базы данных уступили место NoSQL и NewSQL решениям, среди которых имеются:

- MongoDB – документоориентированная база данных для хранения неструктурированной информации;
- Apache Cassandra – высокопроизводительная система для распределенного хранения данных.
- Google BigQuery – облачная аналитическая база данных для быстрой обработки больших массивов данных.

Технологические достижения стали основой для эффективной обработки больших данных. Современные вычислительные мощности, облачные технологии, ИИ и IoT обеспечивают новые возможности для анализа данных, а их дальнейшее развитие приведет к ещё большей автоматизации процессов работы с информацией.

РАЗВИТИЕ ТЕХНОЛОГИЙ ХРАНЕНИЯ ДАННЫХ.

С развитием Big Data и ростом объёмов информации традиционные методы хранения данных стали недостаточны эффективными.

До появления современных технологий хранения информации использовались следующие методы:

- Магнитные ленты;
- Жесткие диски (HDD);
- Реляционные базы данных.

Однако с ростом объёмов информации и необходимостью высокой скорости обработки эти технологии стали уступать место новым решениям.

Современные подходы к хранению данных обеспечивают распределённость, отказоустойчивость и высокую скорость обработки. Среди них можно выделить:

1. NoSQL базы данных.

С появлением неструктурированных данных традиционные SQL-базы стали менее эффективны. NoSQL базы данных обеспечивают высокую масштабируемость и гибкость.

2. Облачные технологии хранения.

Облачные технологии позволяют компаниям хранить большие объёмы данных без необходимости поддерживать собственные сервера.

3. Распределенные файловые системы.

Традиционные файловые системы не справляются с масштабами Big Data, поэтому были разработаны новые распределенные решения:

- HDFS – файловая система, позволяющая хранить и обрабатывать данные в кластерах;
- Serph – объектное и блочное хранилище;

- Google File System – прототип, на основе которого построены многие современные распределенные хранилища.

4. SSD и NVMe хранилища.

Современные серверные решения используют твердотельные накопители и интерфейс NVMe, обеспечивающий высокую скорость чтения и записи данных. Это особенно важно для работы с Big Data, где важна скорость обработки информации.

5. Блокчейн и децентрализованные хранилища.

Блокчейн обеспечивает надежное хранение данных без центрального управления.

Эволюция технологий хранения данных прошла путь от простых магнитных носителей до сложных облачных и распределенных систем. Современные решения обеспечивают гибкость, скорость и надёжность, необходимые для эффективной работы с Big Data.

ПРИМЕНЕНИЕ BIG DATA В РАЗЛИЧНЫХ СФЕРАХ. БОЛЬШИЕ ДАННЫЕ В ГОСУДАРСТВЕННОМ УПРАВЛЕНИИ.

Большие данные играют важную роль в современном государственном управлении, помогая анализировать огромные массивы информации для повышения эффективности работы государственных структур. Их использование позволяет оптимизировать принятие решений, улучшать качество предоставляемых услуг и обеспечивать безопасность граждан.

Возможности Big Data позволяют:

1. Анализ данных для государственного управления.

Государственные структуры используют Big Data для повышения эффективности работы, оптимизации распределения бюджетных средств и улучшения качества услуг для населения. Основные направления применения:

- Аналитика данных для принятия решений – обработка статистической информации о населении или экономике;
- Автоматизация государственных услуг – анализ обращений граждан, оптимизация работы государственных порталов.
- Предсказательное моделирование – прогнозирование потребностей в социальной поддержке, оценка экономических рисков.

2. Применение Big Data в безопасности.

Технологии анализа данных активно используются для обеспечения национальной безопасности и правоохранительной деятельности:

- Системы видеонаблюдения и распознавания лиц;
- Прогнозирование преступности;
- Кибербезопасность.

3. Развитие умных городов.

Big Data играет ключевую роль в концепции “умных городов”, где анализ данных позволяет:

- Оптимизировать работу общественного транспорта;
- Улучшать управление энергопотреблением;
- Автоматизировать контроль за экологической обстановкой.

БОЛЬШИЕ ДАННЫЕ В ЗДРАВООХРАНЕНИИ.

Big Data в здравоохранении открывает новые возможности для диагностики, лечения и управления медицинскими ресурсами. Современные технологии позволяют анализировать большие объёмы данных о пациентах и заболеваниях, проводить персонализированное лечение и прогнозировать эпидемии.

Возможности Big Data позволяют:

1. Улучшение диагностики и лечения.

Big Data помогает врачам быстрее и точнее ставить диагнозы, анализируя медицинские данные пациентов, например:

- Автоматический анализ рентгеновских снимков, МРТ, УЗИ при помощи ИИ;
- Геномный анализ – разработка лекарств на основе анализа ДНК.

2. Оптимизация работы медицинских учреждений.

- Электронные медицинские карты;
- Анализ загруженности больниц;
- Телекоммуникационные технологии – удалённые консультации с врачами.

3. Развитие фармацевтики.

Big Data используется при разработке новых лекарств, ускоряя процессы клинических испытаний и выявляя потенциальные побочные эффекты.

БОЛЬШИЕ ДАННЫЕ В БИЗНЕСЕ

Big Data играет ключевую роль в современном бизнесе, позволяя компаниям более эффективно анализировать поведение клиентов, прогнозировать рыночные тенденции и повышать эффективность бизнес-процессов. Благодаря обработке огромных массивов данных компании могут персонализировать маркетинг, оптимизировать цепочки поставок, а также снизить финансовые риски.

1. Анализ данных для маркетинга и продаж.
 - Персонализация рекламы;
 - Анализ потребительских предпочтений.
2. Оптимизация бизнес-процессов.
 - Прогнозирование спроса;
 - Управление цепочками поставок
 - Финансовая аналитика
3. Big Data в банковской сфере.
 - Оценка платежеспособности клиентов;
 - Обнаружение мошеннических транзакций;
 - Оптимизация инвестиционных стратегий.

Big Data оказывает огромное влияние на различные сферы деятельности, повышая эффективность процессов, улучшая качество услуг и предоставляя новые возможности для анализа информации. В будущем роль больших данных будет только возрастать, способствуя цифровой трансформации различных отраслей.

ПЕРСПЕКТИВЫ РАЗВИТИЯ BIG DATA.

Big Data продолжает активно развиваться, расширяя свое влияние на различные сферы деятельности.

Одним из ключевых направлений станет более тесная связи больших данных с ИИ. Современные алгоритмы уже способны анализировать огромные объёмы информации, но в будущем их работа будет более автономной, быстрой и точной. Это приведет к созданию интеллектуальных систем, которые смогут адаптироваться к изменениям данных и принимать решения в реальном времени.

Квантовые вычисления также внесут значительный вклад в развитие Big Data, обеспечивая невероятную до этого скорость обработки информации. Их внедрение позволит решать сложные вычислительные задачи, которые сегодня требуют огромных ресурсов или вовсе невозможны.

Однако вместе с ростом возможностей Big Data увеличиваются и создаваемые угрозы. Главной проблемой остается защита персональных данных и кибербезопасность. С увеличением объёмов информации возрастает риск утечек. Это требует разработки новых методов шифрования и ужесточения законодательства в сфере цифровой безопасности.

ЗАКЛЮЧЕНИЕ.

В ходе данной курсовой работы были рассмотрены история создания и использования больших данных, их ключевые характеристики, технологические достижения, а также перспективы развития. Большие данные оказали значительное влияние на различные сферы деятельности способствуя автоматизации процессов, повышению эффективности работы организаций и принятию более точных решений.

Развитие технологий хранения и обработки данных стало ключевым фактором, позволившим эффективно управлять огромными объемами информации. С появлением облачных вычислений, ИИ и машинного обучения анализ данных стал более доступным, а алгоритмы точными и предсказуемыми. Однако, вместе с этим, возникли проблемы, связанные с защитой персональных данных и кибербезопасностью.

Таким образом, Big Data продолжает развиваться, оказывая все большее влияния на экономику, науку и повседневную жизнь. В будущем развитие технологий обработки данных сделает большие данные неотъемлемой частью цифрового мира.

СПИСОК ЛИТЕРАТУРЫ

- Маккинси Д. Big Data: как данные меняют мир. – М.: Манн, Иванов и Фербер, 2017.
- Майер-Шенбергер В., Кукье К. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. – М.: АСТ, 2014.
- Митчелл Т. Машинное обучение. – М.: Диалектика, 2021.
- Таненбаум Э. Компьютерные сети. – СПб.: Питер, 2020.
- Чен М., Мао С., Лю Ю. Big Data: A Survey. – Mobile Networks and Applications, 2014.
- Gandomi A., Haider M. Beyond the hype: Big Data concepts, methods, and analytics. – International Journal of Information Management, 2015.
- ISO/IEC 20547:2019 Information technology – Big Data reference architecture.
- European Commission. General Data Protection Regulation (GDPR), 2016.
- Gartner Report. The Future of Big Data Analytics. – 2023.