



**Рис. 4 - Зависимость отношения площадей положительной и отрицательной частей первой производной линии поглощения от расхода (скорости) жидкости для децимолярного раствора  $NiSO_4$  в воде.**

Полученные результаты показывают перспективность изложенного метода для исследования массопереноса жидких сред; и по ряду показателей, таких как точность измерения малых скоростей (расходов), обеспечение измерения реверсивных потоков и простота реализации, превосходит известные методы.

#### **Список использованных источников**

1. Кремлевский П.П. Расходомеры и счетчики количества. –Л.: Машиностроение, 1975, 776 с.
2. Жерновой А.И. Латышев Г.Д. Ядерный магнитный резонанс в проточной жидкости. –М.: Атомиздат, 1964, 252 с.
3. Jacobsohn A., Wangsness R. Shapes of nuclear induction signals. – Phys. Rev., 1948, v.73, p.942–946.

УДК 002

**А.С. Омелянчук, Н.И. Белодед**

Академия управления при президенте Республики Беларусь  
Минск, Беларусь

### **ОБРАБОТКА БОЛЬШИХ ДАННЫХ В ЭПОХУ ЦИФРОВИЗАЦИИ**

*Аннотация.* Актуальность работы обусловлена тем, что с развитием технологий цифровизация стала ключевым фактором в сфере управления

городами, промышленности и сельским хозяйством. В работе рассмотрены классификация и характеристики больших данных, проведено сравнение их видов, выявлены преимущества технологий их хранения и обработки.

**A.S. Omelyanchuk, N.I. Beloded**

Academy of Public Administration under the aegis of the President  
of the Republic of Belarus  
Minsk, Belarus

## **BIG DATA PROCESSING IN THE ERA OF DIGITALIZATION**

***Abstract.** The relevance of the work is due to the fact that with the development of technology, digitalization has become a key factor in the management of cities, industry and agriculture. The work examines the classification and characteristics of big data, compares its types, and identifies the advantages of technologies for storing and processing it.*

Цифровые технологии пронизывают быт современного человека, окружают его в различных формах профессиональной деятельности и досуга, являются важной составляющей общения, познания и социализации. Чтобы идти в ногу со временем для сохранения и усиления своих позиций в мире в новых условиях, обеспечения безопасности и суверенитета, повышения конкурентоспособности экономики, достижения устойчивого роста и улучшения качества жизни населения, нам необходимо всецело содействовать развитию и применению цифровых технологий.

Целью работы является исследование современных технологий и методов хранения файлов в контексте обработки и анализа больших данных в эпоху цифровизации.

Задача работы – рассмотрение классификации и характеристик больших данных, выявление преимуществ технологий хранения и обработки данных.

**Цифровизация** – это процесс преобразования аналоговой информации и процедур в цифровой формат. В современном мире цифровизация проникает во все сферы жизни, обеспечивая быстрый доступ к данным, их удобное хранение и обработку.

Увеличение объёма данных приводит к появлению термина Big Data. **Большие данные (BigData)** - обозначение структурированных и неструктурированных данных огромных объёмов и значительного многообразия, эффективно обрабатываемых горизонтально масштабируемыми программными инструментами, появившимися в

конце 2000-х годов и альтернативных традиционным системам управления базами данных и решениям класса BusinessIntelligence.

Одна из важных проблем при работе с Big Data является классификация. Энциклопедии и имеющиеся научные труды в качестве определяющих характеристик для больших данных традиционно выделяют теорию «VVV», которая содержит в себе следующие характеристики:

1. **Объем (Volume):** Большие данные означают большие объемы информации. Обычно это многие терабайты или петабайты данных.

2. **Разнообразие (Variety):** Данные могут иметь различные форматы, включая текст, изображения, видео, аудио, структурированные и неструктурированные данные. Обработка данных разнообразных типов является одним из вызовов в анализе больших данных.

3. **Скорость (Velocity):** Данные поступают с высокой скоростью, такой как потоки данных в реальном времени, и требуют быстрой обработки и анализа.

Также, наборы данных характеризуют и по их физическому представлению. Существуют следующие категории:

1. **Структурированные данные:** это когда данные хранятся, извлекаются, или могут быть использованы в конкретном, определенном формате. Например, информация о клиенте банка может содержаться в базе данных в виде таблицы, сериализованного пакета и содержать информацию, которую возможно найти однозначно и без особых усилий (имя, возраст, номер телефона, номер счёта, состояния счёта, аресты и др.)

2. **Неструктурированные данные:** этот вид данных трудно категоризовать или структурировать. Неструктурированные данные не имеют определенной формы или общего формата, а храниться могут в виде текста, пакетов или мультимедийных файлов. Хорошим примером могут являться электронные письма, текстовые документы, презентации, видео, которые хоть и могут принадлежать конкретной категории, однако данные в них хранятся хаотично и непредсказуемое.

3. **Слабоструктурированные (полуструктурированные) данные:** представляет собой некий гибрид, смешанную категорию между структурированными и неструктурированными данными. Основное отличие заключается в том, что нельзя категоризировать, но они имеют некоторые определенные свойства (например логи, тэги), которые можно проанализировать и структурировать для их хранения.

Хранение больших данных требует специальных инфраструктурных решений. Работа с файлами и базами данных

является одной из базовых составляющих обработки и анализа больших данных. В контексте больших данных, файлы могут содержать огромные объемы информации, и эффективная обработка данных из файлов играет важную роль.

Технологии хранения файлов в контексте больших данных:

1. Системы файловых хранилищ: системы файлового хранения, такие как Network Attached Storage (NAS) и Storage Area Networks (SAN), могут быть использованы для хранения больших файлов, таких как видео, аудио, изображения и документы. Они обеспечивают высокую пропускную способность и надежность при доступе к данным.

2. Облачные файловые системы: облачные платформы, такие как Dropbox, Google Drive и Microsoft OneDrive, предоставляют возможность хранить и совместно использовать файлы в облаке. Это удобное решение для индивидуальных пользователей и малых компаний.

3. Объектные хранилища: объектные хранилища, такие как Amazon S3 и Azure Blob Storage, могут хранить большие файлы, предоставляя гибкость и масштабируемость. Они используют принцип хранения объектов и позволяют управлять метаданными.

4. Распределенные файловые системы: технологии, такие как Hadoop HDFS и GlusterFS, предоставляют распределенные файловые системы, которые могут хранить и обрабатывать большие файлы на кластерах серверов.

5. Технологии сжатия файлов: для оптимизации хранения больших файлов, можно использовать технологии сжатия, такие как ZIP, GZIP и BZIP2, которые позволяют сократить объем файлов.

Преимущества технологий хранения файлов в больших данных:

1. Эффективное управление файлами: Технологии хранения файлов обеспечивают удобное управление и доступ к файлам, что важно для организаций, которые работают с большими объемами медиа-контента.

2. Быстрый доступ и передача файлов: Облачные и распределенные файловые системы позволяют обеспечивать быстрый доступ и передачу больших файлов, что важно для совместной работы и обмена данными.

3. Масштабируемость: Объектные хранилища и распределенные файловые системы могут легко масштабироваться с ростом объема файлов.

Эффективное управление большими данными, включая хранение файлов и процессы записи, чтения и обработки, становятся все более

важными для успешных бизнес-проектов. Для записи больших объемов данных в файлы, языки программирования предоставляют стандартные библиотеки, такие как Python (модуль `open()`), Java (`FileOutputStream`) и C++ (`ofstream`). Важно оптимизировать процесс записи для уменьшения накладных расходов и повышения производительности. Для чтения больших файлов, таких как журналы серверов, логи и большие наборы данных, используются специальные алгоритмы, которые позволяют быстро извлекать информацию. Производительность при чтении больших данных может быть оптимизирована с помощью кэширования и многопоточности. Для обработки больших данных из файлов, можно использовать инструменты и технологии, такие как Apache Spark, Hadoop MapReduce и Apache Flink, которые предоставляют распределенные механизмы обработки данных. Базы данных и системы управления данными, как Apache Cassandra и MongoDB, могут использоваться для хранения и обработки структурированных данных.

Ниже представлен код, который является примером работы с файлами и выполнением операций с данными, считанными из этих файлов. Он может быть использован как одна из частей более крупной системы для обработки данных.

```
#include <iostream>
#include <fstream>
#include <string>
using namespace std;

int main() {
    string number1, number2;
    ifstream file1("big_data1.txt");
    ifstream file2("big_data2.txt");
    ofstream result("result.txt");
    getline(file1, number1);
    getline(file2, number2);

    int num1 = stoi(number1);
    int num2 = stoi(number2);

    result << "Сложение: " << num1 + num2 << endl;
    result << "Вычитание: " << num1 - num2 << endl;
    result << "Умножение: " << num1 * num2 << endl;
    result << "Деление : " << num1 / num2 << endl;
```

```
file1.close();  
file2.close();  
result.close();  
return 0;
```

Данный код представляет собой пример программы на C++, которая считывает числа из двух файлов «big\_data1.txt» и «big\_data2.txt», выполняет над ними математические операции и записывает результаты в файл «result.txt» и может быть примером использования файлового ввода-вывода в C++ для обработки данных из файлов.

В ходе работы было проведено исследование современных технологий и методов хранения файлов в контексте обработки и анализа больших данных в эпоху цифровизации. Эффективное хранение и обработка больших данных играют ключевую роль в современных информационных системах, и их изучение и развитие остаются приоритетными задачами в области информационных технологий.

### **Список использованных источников**

1. Коньков, А.Е. Цифровизация политики VS политика цифровизации / А.Е. Коньков // Вестник Санкт-Петербургского университета. Международные отношения. – 2020. – Т. 13, № 1. – С. 47-68
2. Колесников, Ю.А. Цифровизация власти через цифровизацию финансовых механизмов / Ю.А. Колесников, В.П. Городецкий // Вестник юридического факультета Южного федерального университета. – 2023. – Т. 10, № 2. – С. 118-122
3. Гордеева, Е.В. Цифровизация в образовании / Е.В. Гордеева, Ш.Г. Мурадян, А.С. Жажоян // Экономика и бизнес: теория и практика. – 2021. – № 4-1. – С. 112-115
4. Кузора, С.С. Цифровая трансформация и большие данные / С.С. Кузора, И.П. Натаров // Вестник Российского университета дружбы народов. – 2022. – Т. 9, № 2. – С. 150-161
5. Белов, В.А. Оценка временной эффективности форматов хранения больших данных в динамике роста объема данных / В.А. Белов, Е.В. Никульчев // Современные информационные технологии и ИТ-образование. – 2022. – № 4. – С. 889-895