

Министерство образования Республики Беларусь
Учреждение образования
«Белорусский государственный университет
информатики и радиоэлектроники»

УДК 004.42:311.3

На правах рукописи

ГРИЦКЕВИЧ
Дмитрий Сергеевич

**ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ АДАПТИВНЫХ СИСТЕМ
СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ИНФОРМАЦИИ**

АВТОРЕФЕРАТ

диссертации на соискание степени
магистра экономических наук

по специальности 1-25 80 08 – Математические и инструментальные
методы экономики

Минск 2017

Работа выполнена на кафедре экономической информатики учреждения образования «Белорусский государственный университет информатики и радиоэлектроники»

Научный руководитель: **АЛЕКСЕЕВ Виктор Федорович**,
кандидат технических наук, доцент кафедры проектирования информационно-компьютерных систем учреждения образования «Белорусский государственный университет информатики и радиоэлектроники»

Рецензент: **ПОЛУБОК Владислав Анатольевич**,
кандидат технических наук, доцент, заведующий кафедрой микропроцессорных систем и сетей института информационных технологий учреждения образования «Белорусский государственный университет информатики и радиоэлектроники»

Защита диссертации состоится «26» января 2017 г. года в 10⁰⁰ часов на заседании Государственной экзаменационной комиссии по защите магистерских диссертаций в учреждении образования «Белорусский государственный университет информатики и радиоэлектроники» по адресу: 220013, Минск, ул. П.Бровки, 6, копр. 1, ауд. 415, тел. 293-20-80, e-mail: kafei@bsuir.by

С диссертацией можно ознакомиться в библиотеке учреждения образования «Белорусский государственный университет информатики и радиоэлектроники».

СОГЛАСОВАНО:
научный руководитель
канд.техн.наук, доцент

В.Ф. Алексеев

ВВЕДЕНИЕ

В современных условиях большую роль приобретает статистика как наука и в частности ее составляющая часть – статистический анализ данных, позволяющий описать количественные закономерности жизни общества в неразрывной связи с их качественным содержанием. Выполняя задачи, наиболее приближенные к реальной жизни, анализ данных способствует ознакомлению со сложившейся ситуацией и обеспечивает приспособление к изменяющимся условиям.

Принятие управленческих, производственных, финансовых, кадровых и других решений основывается на всестороннем анализе, что позволяет делать более обоснованные выводы и прогнозы, давать оценки вероятности их выполнения и предотвращать принятие необоснованных и непродуманных решений. Этим объясняется важность проведения статистического анализа данных, позволяющего привести ясность в изучаемое явление или процесс.

Потребность в средствах анализа данных в различных областях деятельности, особенно в экономике, очень велика. Это послужило причиной развития рынка компьютерных программ для статистической обработки информации. За последние 20 лет активное развитие получили компьютерные программы, позволяющие проводить обработку больших объемов данных с целью выявления закономерностей, сравнения вероятных альтернатив выбора, построения прогнозов развития событий, обнаружения связей между явлениями и процессами и прочее. Существующие программы постоянно совершенствуются в части ускорения работы с данными, улучшения представления результатов анализа данных, повышения удобства интерфейса, совершенствования справочной системы, увеличения числа встроенных в программу процедур и методов, средств обработки данных, взаимодействия с другими программными продуктами и т.д.

На сегодняшний день существует около тысячи компьютерных программ для проведения анализа данных. Рынок программного обеспечения обработки и анализа информации характеризуется высокой конкуренцией, нередки случаи консолидации и поглощений компаний-разработчиков.

Разнообразие статистических пакетов обусловлено многоплановостью задач обработки данных с применением различных типов процедур и методов анализа для поиска ответов на вопросы из различных областей человеческой деятельности. Их используют всегда, когда необходимо получить и обосновать какие-либо суждения о группе объектов или субъектов с некоторой внутренней неоднородностью.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования

Появление огромного количества информации стало причиной возникновения ряда проблем, связанных с изучением и прогнозированием тех или иных явлений или процессов. Эта тенденция привела к необходимости создания больших хранилищ данных, а также систем обработки информации, способных работать с многочисленными потоками данных. Актуальность разработки программного обеспечения подобного рода объясняется возможностью представления результатов манипулирования большими объемами данных в сжатом или, другими словами, усредненном виде. Представленная таким образом информация позволяет дать описательную характеристику изучаемого явления или процесса и сделать прогнозы о дальнейшем развитии событий.

Однако большинство компьютерных программ для проведения статистической обработки информации на сегодняшний день являются дорогостоящими продуктами и предъявляют высокие требования к системе компьютера. Для того, чтобы решить эти проблемы, необходимо оптимизировать процесс манипулирования большими объемами данными в системах статистической обработки информации.

Необходимость разработки подобного рода системы, представляющей собой подключаемый программный модуль для манипулирования данными в условиях умеренного потребления ресурсов компьютера, делает представленную тему диссертации актуальной.

Степень разработанности проблемы

Исследование путей оптимизации процессов манипулирования данными в системах статистической обработки информации с целью увеличения производительности представлено в работах Б. Хоффмана, Б. Холла, К. Шетти и других авторов. Также следует отметить специалистов в области увеличения производительности и масштабирования систем статистической обработки информации, таких как Т. Брайен, А. Падманапхан, Б. Ольшевский, К. Гао, Г. Сербаринов и др.

Авторами работ, связанных с решением оптимизационных задач в области разработки высоконагруженных программных продуктов, являются признанные специалисты А. Шипилев, Р. Черемин, В. Ситников, С. Куксенко, В. Озеров, Н. Липский, В. Иванов, А. Федоров, Ф. Симонис, Ч. Наттер, С. Рупперт, М. Хеклер и др.

Вместе с тем необходимо отметить, что недостаточно полно проработаны вопросы, связанные с применением комплексного подхода к оптимизационным процессам манипулирования данными, с учетом специфики выявленных недостатков систем статистической обработки информации. Предложенное исследование направлено на устранение неполноты проработанности данной области.

Цель и задачи исследования

Целью диссертации является разработка подключаемого программного модуля для систем статистической обработки информации, основанного на принципах технологии обработки данных *OLAP*, учитывающего специфику выявленных недостатков статистических пакетов и позволяющего манипулировать большими объемами данных в условиях умеренного потребления ресурсов компьютера.

Для выполнения поставленной цели в работе были сформулированы **следующие основные задачи:**

1. Исследовать современные средства проведения статистической обработки информации и выявить недостатки при манипулировании большими объемами данных.

2. Разработать адаптивный подключаемый программный модуль для манипулирования большими объемами данных в пакетах статистической обработки информации в условиях умеренного потребления ресурсов компьютера.

3. Построить численные модели на основе данных, полученных с помощью разработанного программного продукта, проверить их на адекватность и провести прогнозирование тенденций изменения изученных экономических показателей на 2017 год.

Область исследования

Содержание диссертации соответствует образовательному стандарту высшего образования второй ступени (магистратуры) специальности 1-25 80 08 «Математические и инструментальные методы экономики».

Теоретическая и методологическая основа исследования

В основу диссертации легли результаты известных исследований отечественных, российских и зарубежных специалистов в области применения пакетов статистической обработки информации и исследования путей оптимизации процессов манипулирования данными в статистических пакетах с целью увеличения производительности системы.

Информационная база исследования сформирована на основе литературы, открытой информации, сведений из электронных ресурсов, а также материалов научных конференций и семинаров.

Научная новизна

Научная новизна и значимость полученных результатов работы заключается в разработке комплексного решения для работы с большими объемами данных, в основу которого положена специальная реализация технологии манипулирования многомерными данными *OLAP*, учитывающая преимущества и недостатки своих предшественников, тем самым представляя собой новую разновидность подобных систем.

Теоретическая значимость работы заключается в предложенном комплексном подходе организации работы с данными для систем статистической обработки информации, который объединяет в себе основные принципы манипулирования большими объемами данных и преимущества современных технологий для разработки систем подобного рода.

Практическая значимость диссертации состоит в разработанном программном продукте, представляющем из себя подключаемый модуль для систем статистической обработки информации и удовлетворяющий всем заявленным требованиям.

Основные положения, выносимые на защиту

1. Классификация и сравнительный анализ современных систем проведения статистической обработки информации и реализаций технологии манипулирования данными *OLAP*, позволяющие учесть преимущества и недостатки рассмотренных систем и моделей при разработке программного модуля; обоснование используемых технологий.

2. Разработанный программный продукт, предназначенный для манипулирования большими объемами данных в системах статистической обработки информации, учитывающий специфику работы с данными в условиях умеренного потребления ресурсов компьютера.

3. Построенные численные модели на основе данных, полученных с помощью разработанного программного продукта, оценка моделей на адекватность и прогноз тенденций изменения изученных экономических показателей на 2017 год.

Апробация диссертации и информация об использовании ее результатов

Результаты исследования были представлены на следующих конференциях: «Инновационные процессы в научной среде» (международная научно-практическая конференция, 8 декабря 2016 г., г.Новосибирск, Россия); «Материалы и методы инновационных исследований и разработок» (международная научно-практической конференция, 3 декабря 2016 г., г.Челябинск, Россия); «Проблемы современных интеграционных процессов и пути их решения» (международная научно-практической конференция, 13 декабря 2016 г., г.Омск, Россия).

Отдельные положения диссертации могут быть использованы при преподавании дисциплин, связанных с разработкой и оптимизацией программных продуктов.

Публикации

Изложенные в диссертации основные положения и выводы опубликованы в 6 печатных работах. Все статьи размещены в сборниках материалов международных научно-практических конференциях. Общий объем публикаций по теме диссертации составляет 18 страниц.

Структура и объем работы

Диссертация состоит из введения, общей характеристики работы, трех глав с краткими выводами по каждой главе, заключения, библиографического списка и приложений.

В первой главе приведен обзор современных средств проведения статистической обработки информации в экономике. На основе сравнительного анализа выявлены основные недостатки представленных программных продуктов. Описаны основные принципы технологии *OLAP*, проведена классификация существующих моделей и их сравнительный анализ. Сделан обзор и обоснование технологий, использованных для разработки программного модуля манипулирования большими объемами данных в системах статистической обработки информации.

Во второй главе приводится описание разработанного адаптивного программного модуля манипулирования большими объемами данных в системах статистической обработки информации. Приведена структура и модель взаимодействия компонентов разработанного программного продукта. Описаны основные принципы функционирования системы. Показаны детали организации работы с памятью, структура и персистентность хранимых данных. Рассмотрен процесс подготовки и индексирования данных. Основное отличие предлагаемой решения состоит в том, что в основу разработанного программного продукта положена специальная реализация технологии манипулирования многомерными данными, которая учитывает преимущества и недостатки альтернативных вариантов.

В третьей главе описано построение численных моделей на основе данных, полученных с помощью разработанного программного продукта. Рассмотрена модель множественной регрессии, которая отражает зависимость доходов компаний по производству косметических средств от затрат на рекламу и затрат на освоение рынка. В качестве изучаемых данных взята выборка по кварталам за 2009 – 2016 годы. Приведена оценка статистической значимости модели и коэффициентов уравнения, а также проверка на автокорреляцию и гетероскедастичность. Сделано прогнозирование тенденций изменения изученных экономических показателей на 2017 год.

В приложении представлены диаграммы представления системы, публикации автора и слайды презентации.

Общий объем диссертационной работы составляет 112 страниц. Из них 59 страниц основного текста. Работа содержит 52 иллюстрации, 16 таблиц, библиографический список из 49 наименований, список собственных публикаций соискателя из 6 наименований, 3 приложения.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** рассмотрено современное состояние проблемы, связанной с появлением огромного количества различной информации, которая в необработанном виде не несет никакой смысловой нагрузки. В качестве решения данной проблемы рассмотрено программное обеспечение в виде хранилищ данных и систем статистической обработки информации, позволяющие представить результаты манипулирования большими объемами данных в сжатом или другими словами усредненном виде, что в свою очередь позволяет дать описательную характеристику изучаемого явления или процесса и сделать прогнозы о дальнейшем развитии событий.

Также в данном разделе определены основные направления исследований, дается обоснование актуальности темы диссертационной работы.

В **общей характеристике работы** сформулированы цель и задачи данного научного исследования, показана степень разработанности проблемы, представлены положения, выносимые на защиту, приведены сведения о личном вкладе соискателя, апробации результатов диссертации, публикации, а также, структура и объем диссертации.

В **первой главе** рассматриваются задачи и области применения анализа данных. Выделены 3 основные группы статистических пакетов по признаку функциональности: универсальные, профессиональные и специализированные пакеты. Рассмотрены наиболее распространенные пакеты анализа данных: *SPSS*, *SAS*, *STATA*, *STATISTICA*, *STATGRAPHICS*, *Minitab*, *EViews*. Определен наиболее популярный среди пользователей пакет согласно статистике *Google Scholar* о количестве опубликованных научных статей за 2016 год для каждого пакета, представленной на рисунке 1.

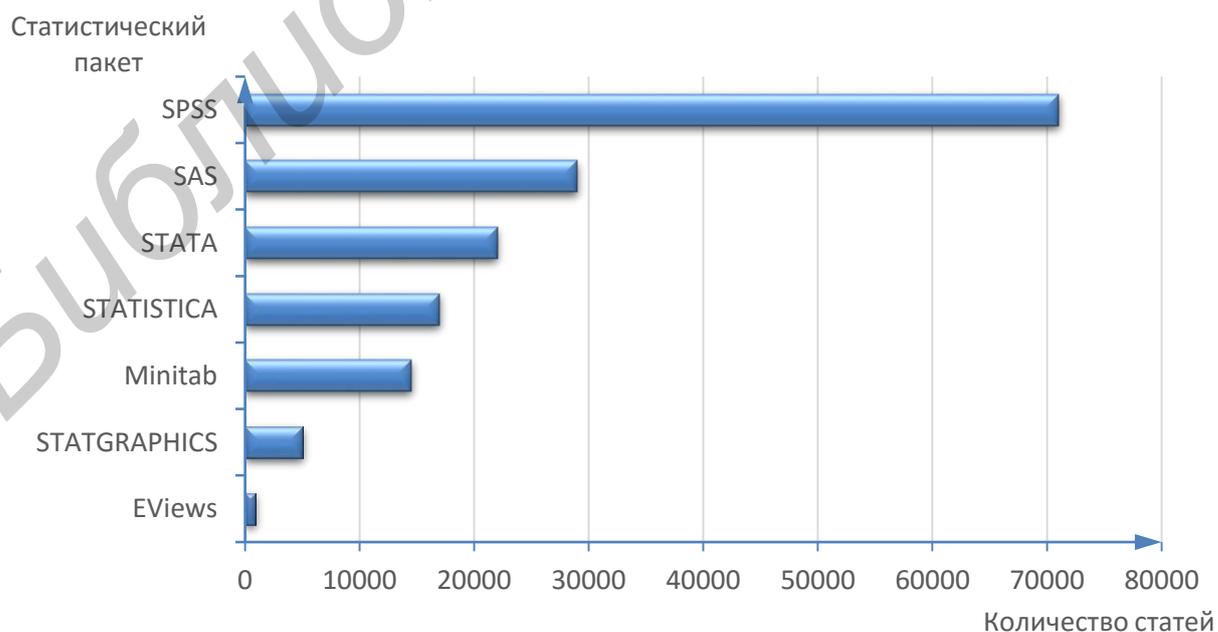


Рисунок 1 – Диаграмма, отражающая количество опубликованных научных статей за 2016 год по каждому из статистических пакетов

Определены основные параметры, влияющие на выбор того или иного пакета для работы: соответствие программного обеспечения характеру решаемых задач, объем обрабатываемых данных, производительность обработки данных, удобство в эксплуатации и наличие подробной сопровождающей документации.

Приведена сравнительная таблица, отражающая преимущества и недостатки каждого из рассматриваемых статистических пакетов. На основе таблицы сделано заключение, что одними из ключевых недостатков рассматриваемых программных продуктов являются высокая цена и высокие требования к системе компьютера.

Приведены результаты тестирования пакета *SPSS* на производительность при вычислении основных показателей описательной статистики и корреляционной матрицы, а также при проведении сортировки.

На основе результатов проведенного сравнительного анализа рассмотренных статистических пакетов, главными недостатками которых являются высокая стоимость, высокие требования, предъявляемые к системе компьютера, или несовместимость с некоторыми операционными системами, сделан вывод о необходимости разработки подключаемого модуля, который позволит осуществлять манипулирование большими объемами информации при умеренном потреблении ресурсов компьютера за счет использования технологии обработки многомерных данных *OLAP*, что значительно снизит предъявляемые требования к системе компьютера.

Рассмотрены основные принципы технологии *OLAP*, которые использовались при разработке программного продукта, и приведена классификация существующих моделей и их сравнительный анализ с целью выявления преимуществ и недостатков. Принято решение о комбинировании нескольких моделей *OLAP* при разработке программного продукта с целью достижения заявленных требований.

Обосновано использование высокоуровневого языка разработки *Java* для создания систем подобного рода. Проведен сравнительный анализ с языком программирования *C/C++*, который является более классическим подходом при разработке высоконагруженных систем. Определены основные преимущества языка *Java* по сравнению с языком *C/C++* такие как: *JIT*-компиляция, скорость разработки, сторонние вспомогательные библиотеки для работы с примитивами.

Рассмотрена технология *Apache Thrift* как способ реализации распределенного взаимодействия между элементами системы и обоснован ее выбор на основе проведенного сравнительного анализа данной технологии с альтернативными вариантами: *RMI*, *SOAP*, *Google Protocol Buffers (protobufs)*, *REST*.

Рассмотрен фреймворк *Disruptor* для обработки транзакций отличающийся высокой пропускной способностью и очень низкой латентностью.

Обосновано использование пула соединений с базой данных *HikariCP*. Приведены диаграммы, отражающие сравнительный анализ данной технологии с ее альтернативными вариантами: *commons-dbcp2*, *Tomcat*, *Vibur*, *c3p0*.

Во **второй** главе приведено описание разработанного программного модуля. Представлены основные требования, предъявляемые к системе. Рассмотрена структура компонентов модуля и модель их взаимодействия, представленная на рисунке 2.

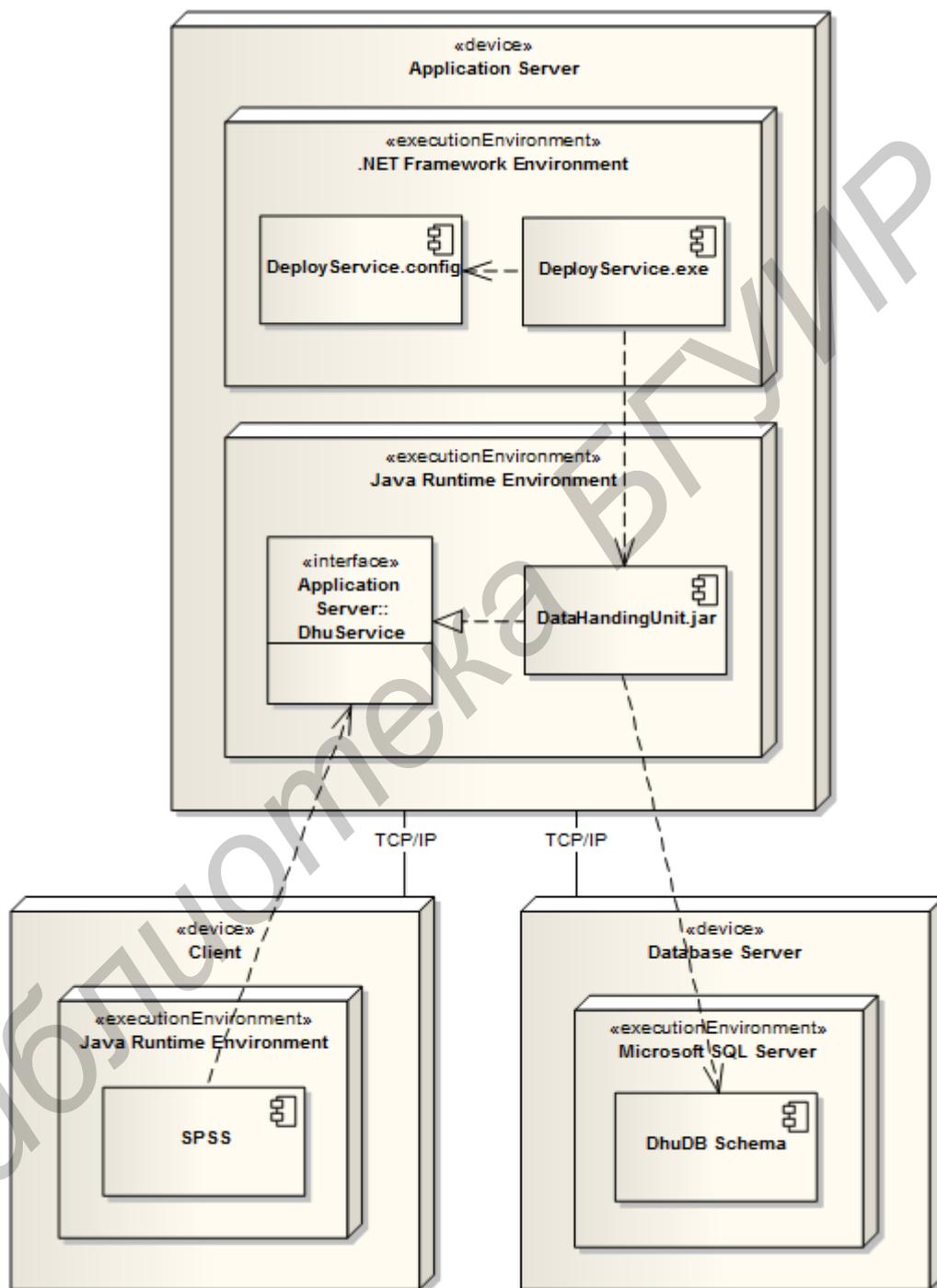


Рисунок 2 – Диаграмма развертывания системы

Описана работа с памятью *RAM* и структура хранения данных. Приведено обоснование подхода хранения данных в памяти процесса, избегая дисковых операций ввода-вывода, а также сетевого взаимодействия.

Проведен тест производительности наиболее часто используемых ме-

ханизмов организации хранения данных в различных регионах памяти, предоставляемых *JDK*. Результаты теста отражены в таблице 1.

Таблица 1 – Сравнение производительности хранения данных в памяти

Структура	Режим	Число запусков	Среднее значение	Средняя ошибка	Единицы измерения
<i>LinkedList</i>	<i>avgt</i>	5	21418.739	26207.055	<i>ms/op</i>
<i>ArrayList</i>	<i>avgt</i>	5	4315.133	5770.468	<i>ms/op</i>
<i>ByteBuffer</i>	<i>avgt</i>	5	1783.773	223.998	<i>ms/op</i>
<i>MappedByteBuffer</i>	<i>avgt</i>	5	1152,827	1203.361	<i>ms/op</i>
<i>DirectByteBuffer</i>	<i>avgt</i>	5	770.764	214.911	<i>ms/op</i>
<i>Unsafe</i>	<i>avgt</i>	5	473.691	25.812	<i>ms/op</i>
<i>Empty</i>	<i>avgt</i>	5	17.574	0.653	<i>ms/op</i>

Рассмотрены структуры хранимых ключей и значений, описана иерархия классов доступа в память. Рассмотрена реализация процесса извлечения, обработки и загрузки данных с помощью многопоточного параллельного фреймворка *Disruptor*. Описан процесс индексирования хранимых данных.

В третьей главе на основе полученных с помощью разработанного программного модуля статистических данных построена модель множественной регрессии, которая отражает зависимость дохода компаний по производству косметических средств для ухода за волосами от затрат компаний на рекламу и затрат компаний на освоение рынка. В качестве изучаемых данных была взята выборка по кварталам за 2009 – 2016 годы в пределах Германии. Результаты множественной регрессии в численном виде представлены в таблице 2.

Таблица 2 – Результаты множественной регрессии в численном виде

	Коэффициенты	Стандартная ошибка	t-статистика	P-значение	Нижние 95%	Верхние 95%
У-пересечение	-32684,4	10391,76	-3,14522	0,00382	-53938	-11430,8
Переменная X_1	2,120233	0,112294	18,88105	7,769E-18	1,89067	2,3499
Переменная X_2	1003,961	144,7055	6,937963	1,262E-07	708,005	1299,92

Модель множественной регрессии с учетом рассчитанных коэффициентов имеет вид: $Y = -32684,4 + 2,120233 * X_1 + 1003,961 * X_2$.

Построенная модель является статистически значимой в целом. Коэффициенты регрессии являются статистически значимыми. Установлено наличие явления автокорреляции случайных отклонений модели на основе критерия Бреуша-Годфри. Установлено отсутствие явления гетероскедастичности. Проведено прогнозирование тенденций изменения изучаемых экономических показателей на 2017 год.

В приложениях приведены диаграммы, отражающие варианты представления системы, опубликованные работы в рамках данного научного исследования, а также графический материал в виде слайдов.

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

1. Проведена классификация и сравнительный анализ существующих систем статистической обработки информации, на основе которого выявлены следующие недостатки: высокая стоимость, высокие требования, предъявляемые к системе компьютера, или несовместимость с некоторыми операционными системами. Проведен обзор и сравнительный анализ существующих реализаций технологии манипулирования данными *OLAP*, позволивший учесть преимущества и недостатки рассмотренных моделей при разработке программного продукта, в основу которого положена специальная реализация технологии *OLAP*, представляющая собой новую разновидность подобных систем. Выбраны и обоснованы технологии, которые были использованы в разработке программного продукта манипулирования большими объемами данных в системах статистической обработки информации: технология реализации распределенного взаимодействия *Apache Thrift*, фреймворк *Disruptor*, быстрый пул соединений с базой данных *HikariCP*. В качестве языка разработки был выбран высокоуровневый язык *Java*.

2. С учетом предложенных технологий разработан программный продукт, предназначенный для манипулирования большими объемами данных в системах статистической обработки информации, которая учитывает специфику работы с данными в условиях умеренного потребления ресурсов компьютера. Рассмотрена структура и модель взаимодействия компонентов разрабатываемого программного модуля. Представлены диаграммы компонентов и развертывания системы. Рассмотрены особенности и детали реализации системы, учитывающие преимущества выбранных технологий, а также нестандартные решения, позволяющие добиться необходимой производительности модуля с учетом заявленных требований. Описаны основные принципы функционирования системы. Представлены диаграммы последовательностей создания и активации снэпшота, диаграмма состояния снэпшота

3. Построены численные модели на основе данных, полученных с помощью разработанного программного продукта. Для изучения были использованы статистические данные Германии как экономически высокоразвитой страны. Проведена оценка статистической значимости модели и коэффициентов уравнения регрессии, выполнена проверка на автокорреляцию и гетероскедастичность. Спрогнозированы тенденции изменения изученных экономических показателей на 2017 год.

СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ

Статьи в сборниках научных трудов

1. Грицкевич, Д.С. Особенности использования языка Java для разработки высоконагруженных систем / Д.С. Грицкевич // Инновационные процессы в научной среде: сборник статей Международной научно-практической конференции (8 декабря 2016 г., г.Новосибирск). В 4 ч. Ч.3. – Уфа: МЦИИ ОМЕГА САЙНС, 2016. – С. 31–33.

2. Грицкевич, Д.С. Сравнение технологий реализации распределенного взаимодействия между элементами системы / Д.С. Грицкевич // Материалы и методы инновационных исследований и разработок: сборник статей Международной научно-практической конференции (3 декабря 2016 г., г.Челябинск). В 3 ч. Ч.2. – Уфа: МЦИИ ОМЕГА САЙНС, 2016. – С. 20–22.

3. Грицкевич, Д.С. Фреймворк Disruptor как способ увеличения производительности обработки событий / Д.С. Грицкевич // Материалы и методы инновационных исследований и разработок: сборник статей Международной научно-практической конференции (3 декабря 2016 г., г.Челябинск). В 3 ч. Ч.2. – Уфа: МЦИИ ОМЕГА САЙНС, 2016. – С. 18–20.

4. Грицкевич, Д.С. Сравнение производительностей различных способов организации работы с памятью в Java / Д.С. Грицкевич // Инновационные процессы в научной среде: сборник статей Международной научно-практической конференции (8 декабря 2016 г., г.Новосибирск). В 4 ч. Ч.3. – Уфа: МЦИИ ОМЕГА САЙНС, 2016. – С. 33–35.

5. Грицкевич, Д.С. Реализация процесса ETL с помощью фреймворка Disruptor / Д.С. Грицкевич // Проблемы современных интеграционных процессов и пути их решения: сборник статей Международной научно-практической конференции (13 декабря 2016 г., г.Омск). В 2 ч. Ч.2. – Уфа: МЦИИ ОМЕГА САЙНС, 2016. – С. 60–62.

6. Грицкевич, Д.С. Особенности реализации высокопроизводительного кэша в Java / Д.С. Грицкевич // Проблемы современных интеграционных процессов и пути их решения: сборник статей Международной научно-практической конференции (13 декабря 2016 г., г.Омск). В 2 ч. Ч.2. – Уфа: МЦИИ ОМЕГА САЙНС, 2016. – С. 62–65.

РЭЗІЮМЭ
Грыцкевіч Дзмітрый Сяргеевіч
Лікавае мадэляванне адаптыўных сістэм статыстычнай апрацоўкі
інфармацыі

Ключавыя словы: статыстычная апрацоўка інфармацыі.

Мэта працы: распрацоўка падключаемага праграмнага модуля для сістэм статыстычнай апрацоўкі інфармацыі, заснаванага на прынцыпах тэхналогіі апрацоўкі даных OLAP, які ўлічвае спецыфіку выяўленых недахопаў статыстычных пакетаў і дазваляе маніпуляваць вялікімі аб'ёмамі даных ва ўмовах умеранага спажывання рэсурсаў кампутара.

Атрыманыя вынікі і іх навізна: зроблены параўнальны аналіз сучасных сродкаў правядзення статыстычнай апрацоўкі інфармацыі. На аснове выяўленых недахопаў прынята рашэнне аб распрацоўцы праграмнага модуля, які дазваляе маніпуляваць вялікімі аб'ёмамі даных у сістэмах статыстычнай апрацоўкі інфармацыі пры ўмераным спажыванні рэсурсаў кампутара. Распрацаваны адаптыўны праграмны модуль маніпулявання вялікімі аб'ёмамі даных у сістэмах аналізу даных. На аснове атрыманых з дапамогай распрацаванага праграмнага модуля даных пабудавана лікавая мадэль і праверана на адэкватнасць; спрагназаваны тэндэнцыі змены вывучаных эканамічных паказчыкаў на 2017 год.

Вобласць ужывання: статыстычны аналіз даных у эканоміцы.

РЕЗЮМЕ

Грицкевич Дмитрий Сергеевич

Численное моделирование адаптивных систем статистической обработки информации

Ключевые слова: статистическая обработка информации.

Цель работы: разработка подключаемого программного модуля для систем статистической обработки информации, основанного на принципах технологии обработки данных *OLAP*, учитывающего специфику выявленных недостатков статистических пакетов и позволяющего манипулировать большими объемами данных в условиях умеренного потребления ресурсов компьютера.

Полученные результаты и их новизна: выполнен сравнительный анализ современных средств проведения статистической обработки информации. На основе выявленных недостатков принято решение о разработке программного модуля, позволяющего манипулировать большими объемами данных в системах статистической обработки информации при умеренном потреблении ресурсов компьютера. Разработан адаптивный программный модуль манипулирования большими объемами данных в системах анализа данных. На основе полученных с помощью разработанного программного модуля данных построена численная модель и проверена на адекватность; спрогнозированы тенденции изменения изученных экономических показателей на 2017 год.

Область применения: статистический анализ данных в экономике.

SUMMARY
Gritskevich Dmitry Sergeevich
Numerical simulation of adaptive systems and statistical information processing

Keywords: statistical processing of information.

The object of study: to develop a plug-in software module for statistical processing of information, based on the principles of processing data OLAP-specific identified deficiencies of statistical packages and allow manipulation of large amounts of data in conditions of moderate consumption of computer resources.

The results and novelty: a comparative analysis of modern means of conducting statistical processing. Based on the identified shortcomings made the decision to develop a software module that allows to manipulate large volumes of data in systems of statistical data processing with moderate consumption of computer resources. Developed an adaptive software module manipulating large volumes of data in systems of data processing. On the basis of the obtained using the developed software module of the data a numerical model and tested for adequacy; forecasted trends in the studied economic indicators in 2017.

Sphere of application: statistical analysis of data in the economy.