

Министерство образования Республики Беларусь

Учреждение образования  
«Витебский государственный университет имени П.М. Машерова»  
(ВГУ имени П.М. Машерова)

УДК 004.62, 004.41

№ гос. регистрации 20181861

Инв. № \_\_\_\_\_



ПОДПИСАЮ

Директор по научной работе

*И. М. Прищеп* И. М. Прищеп

*20* февраля 2019г.

## ОТЧЕТ

о научно-исследовательской работе

Провести научно-исследовательские работы по разработке концепции и архитектуры единого корпоративного хранилища данных (ЕКХД), разработать программные средства, реализующие созданные модели данных для ЕКХД, представляющие собой техническое нововведение в рамках существующей у заказчика корпоративной системы

согласно договору с компанией с ограниченной ответственностью  
«АндерсанЛаб» № 10/09 от 10.09.2018 г.

(заключительный)

УСТАНОВА АДУКАЦЫІ  
«Віцебскі дзяржаўны ўніверсітэт  
імя П.М.Машэрава»  
НАВУКОВАЯ БІБЛІЯТЭКА

Научный руководитель  
зав. каф. прикладного системного  
программирования, к. ф.-м. н.

*С. А. Ермоchenko* 04.02.19

С. А. Ермоchenko

Нормоконтролер

*Т. В. Харкевич* 04.02.19

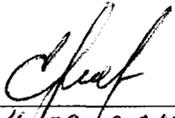
Т. В. Харкевич

Витебск 2019

Н-530

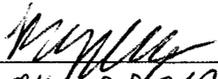
## СПИСОК ИСПОЛНИТЕЛЕЙ

Научный руководитель темы,  
кандидат физ.-мат. наук

  
04.02.2019

С. А. Ермоченко  
(введение, раздел 1,  
заключение)

Исполнители темы:  
магистрант

  
04.02.2019

А. А. Кухарев  
(разделы 1-3)

## РЕФЕРАТ

Отчет 20 с., 1 ч., 3 рис., 2 листинга, 9 источников.

### БАЗА ДАННЫХ, BIG DATA, MAP-REDUCE, REST API, ШАРДИНГ, РЕПЛИЦИРОВАНИЕ

Объект исследования – единое корпоративное хранилище данных (ЕКХД).

Предмет исследования – подходы к построению хранилища, критерии эффективности функционирования хранилища и его оптимизация.

Цель работы – проектирование концепции и архитектуры единого корпоративного хранилища данных (ЕКХД) и научное обоснование их эффективности для оптимизации подходов к построению хранилища.

Методы исследования: описательно-аналитический, проективный, метод объектно-ориентированного анализа, методы нормализации реляционных данных.

В процессе работы проводился анализ подходов к построению единого корпоративного хранилища данных, изучались критерии эффективности функционирования такого хранилища. На основании проведённого анализа проектировалась концепция и архитектура единого корпоративного хранилища, и оптимизация эффективности его работы.

Результатом работы стала разработанное хранилище, позволяющее развернуть в нём пилотный объём сущностей, соответствующих предметной области заказчика.

Практическая значимость работы обусловлена производственной необходимостью заказчика. Разработанная концепция хранилища позволит структурировать информацию, обрабатываемую в информационной системе заказчика, и унифицировать алгоритмы её обработки.

## СОДЕРЖАНИЕ

Введение .....	5
Основная часть.....	7
1 Разработка концепции и архитектуры хранилища данных .....	7
2 Проектирование верхнеуровневой модели данных .....	12
3 Проработка детализированного описания объектов, необходимых для разворачивания пилотного объёма сущностей и показателей.....	17
Заключение.....	19
Список использованных источников .....	20

## ВВЕДЕНИЕ

В рамках заключённого с компанией с ограниченной ответственностью «АндерсанЛаб» договора на выполнение научно-исследовательских, опытно-конструкторских и технологических работ, была проанализирована имеющаяся у заказчика корпоративная система отчётности и выявлен ряд её недостатков.

Во-первых, это разрозненность хранящихся в системе данных. Информация не хранится в одном хранилище, а рассредоточена в разных базах данных и частях имеющейся системы. С одной стороны такой подход позволяет независимо хранить данные, касающиеся различных несвязанных между собой бизнес-процессов, что повышает отказоустойчивость и надёжность хранения информации. Но с другой стороны затруднён централизованный анализ этих данных [1].

Во-вторых, хранящиеся данные недостаточно структурированы. Многие документы хранятся не в виде записей в базе данных, а в виде отдельных электронных таблиц Microsoft Excel, текстовых документов Microsoft Word и др. Такая ситуация приводит к снижению производительности при анализе имеющихся данных [2].

Таким образом, для улучшения имеющейся системы необходимо разработать некую платформу по интеграции и консолидации имеющихся данных, обеспечив унифицированный интерфейс доступа к этим данным для их последующего анализа.

Целью настоящего исследования является проектирование концепции и архитектуры единого корпоративного хранилища данных (ЕКХД) и научное обоснование их эффективности для оптимизации подходов к построению хранилища.

Для достижения поставленной цели необходимо решить ряд задач:

1. Проанализировать текущие и перспективные потребности заказчика с учётом современных тенденций в области анализа данных и ведущих мировых практик в этой сфере.

2. Сформировать концепцию и архитектуру ЕКХД на основании выявленных потребностей.

3. Спроектировать и разработать модель данных ЕКХД на основе выработанной концепции и архитектуры. Оценить адекватность этой модели потребностям заказчика.

4. Детализировать построенную модель данных и развернуть в разработанном хранилище пилотный объем сущностей и показателей бизнес-процессов, демонстрирующий возможности разработанной ЕКХД по хранению и анализу данных.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Милёхина, О. В. Информационные системы: теоретические предпосылки к построению / О. В. Милёхина, Е. Я. Захарова, В. А. Титова. – 2-е издание. – Новосибирск: издательство НГТУ, 2014. – 283 с.
2. Бритков, В. Б. Методы анализа больших объёмов слабоструктурированной информации / В. Б. Бритков, А. В. Булычев // Информационные технологии и вычислительные системы. – Москва: издательство Федеральный исследовательский центр «Информатика и управление» Российской академии наук, 2010. – №1. – с. 36-44.
3. Ульман, Дж. Д. Реляционные базы данных / Дж. Д. Ульман, Дж. Уидом. – Москва: Лори, 2014. – 384 с.
4. Марц, Н. Большие данные. Принципы и практика построения масштабируемых систем обработки данных в реальном времени / Н. Марц, Дж. Уоррен. – Москва: Вильямс, 2017. – 368 с.
5. Foreman, J. W. Data Smart: Using Data Science to Transform Information into Insight. – Willey, 2013. – 409 p.
6. Tan, P.-N. Introduction to Data Mining / P.-N. Tan, M. Steinbach, V. Kumar. – Pearson, 2005. – 792 p.
7. Уорсли, Дж. PostgreSQL. Для профессионалов / Дж. Уорсли, Дж. Дрейк. – Санкт-Петербург: Питер, 2003. – 496 с.
8. Core J2EE Patterns – Data Access Object [Electronic resource] / Oracle Technology Network. Java. – Mode of access: <https://www.oracle.com/technetwork/java/dataaccessobject-138824.html>. – Date of access: 02.02.2019
9. Heller, M. REST and CRUD: the Impedance Mismatch [Electronic resource] / InfoWorld. – Mode of access: <https://www.infoworld.com/article/2640739/application-development/rest-and-crud--the-impedance-mismatch.html>. – Date of access: 02.02.2019