

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Школа инженерного предпринимательства
Направление подготовки 27.04.05 Инноватика
Технологическое брокерство

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
Построение хранилища данных для OLAP-системы в целях поддержки принятия решений в организации

УДК 659.44:659.1:338.46

Студент

Группа	ФИО	Подпись	Дата
ЗНМ04	Тлюняева А.А.		

Руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ШИП	Юдахина О.Б.	к.э.н., доцент		

КОНСУЛЬТАНТЫ:

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ШИП	Черепанова Н.В.	к.ф.н., доцент		

Нормоконтроль

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель	Громова Т.В.			

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ШИП	Попова С.Н.	к.э.н., доцент		

**Планируемые результаты освоения ООП
27.04.05 Инноватика «Технологическое брокерство»**

Код компетенции	Наименование компетенции
Универсальные компетенции	
УК(У)-1	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий
УК(У)-2	Способен управлять проектом на всех этапах его жизненного цикла
УК(У)-3	Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели
УК(У)-4	Способен применять современные коммуникативные технологии, в том числе на иностранном (-ых) языке (-ах), для академического и профессионального взаимодействия
УК(У)-5	Способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия
УК(У)-6	Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки
Общепрофессиональные компетенции	
ОПК(У)-1	Способен решать профессиональные задачи на основе истории и философии нововведений, математических методов и моделей для управления инновациями, компьютерных технологий в инновационной сфере
Профессиональные компетенции	
ПК(У)-1	Способен выбрать (разработать) технологию осуществления (коммерциализации) результатов научного исследования (разработки)
ПК(У)-2	Способен организовать работу творческого коллектива для достижения поставленной научной цели, находить и принимать управленческие решения, оценивать качество и результативность труда, затраты и результаты деятельности научно-производственного коллектива
ПК(У)-3	Способен произвести оценку экономического потенциала инновации, затрат на реализацию научно-исследовательского проекта
ПК(У)-4	Способен найти (выбрать) оптимальные решения при создании новой наукоемкой продукции с учетом требований качества, стоимости, сроков исполнения, конкурентоспособности и экологической безопасности
ПК(У)-5	Способен разработать план и программу организации инновационной деятельности научно-производственного подразделения, осуществлять технико-экономическое обоснование инновационных проектов и программ
ПК(У)-6	Способен применять теории и методы теоретической и прикладной инноватики, систем и стратегий управления, управления качеством инновационных проектов
ПК(У)-7	Способен выбрать (или разработать) технологию осуществления научного эксперимента (исследования), оценить затраты и организовать его осуществление
ПК(У)-8	Способен выполнить анализ результатов научного эксперимента с использованием соответствующих методов и инструментов обработки
ПК(У)-9	Способен представить (опубликовать) результат научного исследования на конференции или в печатном издании, в том числе на иностранном языке

ПК(У)-10	Способен критически анализировать современные проблемы инноватики, ставить задачи и разрабатывать программу исследования, выбирать соответствующие методы решения экспериментальных и теоретических задач, интерпретировать, представлять и применять полученные результаты
ПК(У)-11	Способен руководить практической, лабораторной и научно-исследовательской работой студентов, проводить учебные занятия в соответствующей области
ПК(У)-12	Способен применять, адаптировать, совершенствовать и разрабатывать инновационные образовательные технологии
Дополнительно сформированные профессиональные компетенции университета в соответствии с анализом трудовых функций выбранных обобщенных трудовых функций профессиональных стандартов, мирового опыта и опыта организации	
ДПК(У)-1	Проводить аудит и анализ производственных процессов с целью уменьшения производственных потерь и повышения качества выпускаемого продукта
ДПК(У)-2	Разрабатывать программы коммерциализации и маркетинга инновационных проектов на основе комплексного анализа рынка

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Школа инженерного предпринимательства
Направление подготовки 27.04.05 Инноватика
Технологическое брокерство

УТВЕРЖДАЮ:
Руководитель ООП

(Подпись) (Дата) (Ф.И.О.)
Попова С.Н.

ЗАДАНИЕ

на выполнение выпускной квалификационной работы

В форме:

магистерской диссертации

(бакалаврской работы/магистерской диссертации)

Студенту:

Группа	ФИО
ЗНМ04	Тлюняевой Анастасии Артуровной

Тема работы:

Построение хранилища данных для OLAP-системы в целях поддержки принятия решений в организации	
Утверждена приказом директора (дата, номер)	

Срок сдачи студентом выполненной работы:	06.06.2022 г.
--	---------------

ТЕХНИЧЕСКОЕ ЗАДАНИЕ

Исходные данные к работе <i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.).</i>	Объектом исследования является дочернее предприятие ПАО «Газпром». Предмет исследования – процесс проектирования хранилища данных для OLAP-системы. Информационно-методическая база исследования: учебно-методические пособия; научная отечественная и зарубежная литература; журналы и периодические издания; отчетная документация предприятия; сайт предприятия, сайты систем планирования и аналитики.
--	--

<p>Перечень подлежащих исследованию, проектированию и разработке вопросов (аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание Процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</p>	<p>Цель исследования – на основе анализа информационно-аналитической системы предприятия и требований пользователей разработать алгоритм построения хранилища данных для OLAP-системы с целью формирования аналитической отчетности предприятия по дебиторской задолженности.</p> <p>Задачи исследования:</p> <ol style="list-style-type: none"> 1. Изучение теоретических особенностей хранилищ данных и OLAP технологий. 2. Анализ информационно-аналитической системы предприятия и процедуры формирования отчетов по дебиторской задолженности. 3. Выявление и анализ требований пользователей к процедуре формирования отчетов. 4. Построение и анализ алгоритма проектирования хранилища данных. 5. Построение хранилища данных для дочернего предприятия ПАО «Газпром». 6. Построение отчёта для управления дебиторской задолженностью.
<p>Перечень графического материала (с точным указанием обязательных чертежей)</p>	<p>Таблицы, рисунки</p>
<p>Консультанты по разделам выпускной квалификационной работы (с указанием разделов)</p>	
<p>Раздел</p>	<p>Консультант</p>
<p>Социальная ответственность</p>	<p>Черепанова Наталья Владимировна</p>
<p>Раздел на иностранном языке</p>	<p>Надеина Луиза Васильевна</p>
<p>Названия разделов, которые должны быть написаны на русском и иностранном языках:</p>	
<p>Глава 2 Сбор и анализ требований для проектирования хранилищ данных</p>	<p>Collection and analysis of requirements for the design of data warehouses</p>

<p>Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику</p>	<p>20.12.2021 г.</p>
---	-----------------------------

Задание выдал руководитель:

<p>Должность</p>	<p>ФИО</p>	<p>Ученая степень, звание</p>	<p>Подпись</p>	<p>Дата</p>
<p>Доцент ШИП</p>	<p>Юдахина О.Б.</p>	<p>к.э.н., доцент</p>		<p>20.12.2021</p>

Задание принял к исполнению студент:

<p>Группа</p>	<p>ФИО</p>	<p>Подпись</p>	<p>Дата</p>
<p>ЗНМ04</p>	<p>Глюняева Анастасия Артуровна</p>		<p>20.12.2021</p>

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Школа инженерного предпринимательства
Направление подготовки (специальность) 27.04.05 Инноватика
Уровень образования магистратура
Период выполнения (осенний / весенний семестр 2021/2022 учебного года)

Форма представления работы:

магистерская диссертация
(бакалаврская работа, магистерская диссертация)

**Построение хранилища данных для OLAP-системы в целях поддержки
принятия решений в организации**

**КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН
выполнения выпускной квалификационной работы**

Срок сдачи студентом выполненной работы:

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
13.03.2022	Теоретическая часть	20
24.05.2022	Практико-ориентированная часть	60
25.05.2022	Социальная ответственность	10
03.06.2022	Часть на английском языке	10

Составил преподаватель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ШИП	Юдахина О.Б.	к.э.н., доцент		

Принял студент:

ФИО	Подпись	Дата
Тлюняева Анастасия Артуровна		

СОГЛАСОВАНО:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ШИП	Попова С.Н.	к.э.н., доцент		

РЕФЕРАТ

Выпускная квалификационная работа содержит 134 страниц, 50 рисунков, 5 таблиц, 54 использованных источников, 1 приложение.

Ключевые слова: хранилища данных, базы данных, OLAP-технологии, проектирование, визуализация, оптимизация, дебиторская задолженность.

Объект исследования – дочернее предприятие ПАО «Газпром».

Предмет исследования – процесс проектирования хранилища данных для OLAP-системы.

Цель работы – на основе анализа информационно-аналитической системы предприятия и требований пользователей разработать алгоритм построения хранилища данных для OLAP-системы.

В процессе исследования проводилось изучение особенностей проектирования хранилища данных для OLAP системы.

В результате исследования был разработан алгоритм построения хранилища данных для OLAP-системы на базе платформы «Optimacros».

Основные конструктивные, технологические и технико-эксплуатационные характеристики: магистерская диссертация состоит из четырех частей: в первой рассмотрены теоретические аспекты построения хранилищ данных. Во второй осуществлён сбор и анализ требований пользователей. В третьей – представлен процесс построения хранилища для формирования отчетов по дебиторской задолженности. В четвертой части описаны принципы социальной ответственности исследуемой компании.

Степень внедрения: разработанные формы отчета ПАО «Газпром» находятся на этапе тестирования наполнения данными.

Область применения: результаты исследования могут применяться для построения хранилища любого предприятия.

Экономическая эффективность/ значимость работы заключается в повышении эффективности управления дебиторской задолженностью.

В будущем планируется расширение хранилища данных предприятия.

Определения, обозначения, сокращения, нормативные ссылки

Хранилище данных (Data Warehouse) – предметно-ориентированная информационная корпоративная база большого объема, специально разработанная и предназначенная для сбора и предварительной обработки данных, с целью предоставления информации пользователям, для статистического анализа, подготовки отчетов и анализа бизнес-процессов.

OLAP (OnLine Analytical Processing) – система аналитической обработки данных, предназначенная для подготовки отчетов, построения прогностических сценариев и выполнения статистических расчетов на базе больших информационных массивов, имеющих сложную структуру.

ETL-системы (Extraction, Transformation, Loading) – комплекс программных средств, предназначенный для извлечения данных из различных и разнотипных источников, приведения их к единому формату, оценки их качества, очистки, обогащения и загрузки в хранилище с целью дальнейшей аналитической обработки.

OLTP (Online Transaction Processing) – транзакционная система, обеспечивающая обработку транзакций в режиме реального времени.

Витрина данных (Data Mart) – подмножество хранилища данных, представляющее собой массив тематической, узконаправленной информации.

База данных (Database) – упорядоченный набор структурированной информации или данных, хранящийся в электронном виде в компьютерной системе.

SQL (Structured Query Language) – язык программирования, используемый в большинстве реляционных баз данных для запросов, обработки и определения данных, а также контроля доступа.

Система управления базами данных (СУБД) – комплекс программно-языковых средств, позволяющих создать базы данных и управлять данными.

Оглавление

Введение.....	10
1 Теоретические аспекты построения хранилищ данных.....	13
1.1 Понятие и особенности структуры хранилищ данных	13
1.2 Анализ методологических подходов построения хранилищ данных	19
1.3 Управление хранилищем данных.....	26
1.4 OLAP технология.....	33
1.5 Сравнительный анализ существующих систем планирования и аналитики	43
2 Сбор и анализ требований для проектирования хранилищ данных	48
2.1 Краткая характеристика деятельности предприятия и анализ дебиторской задолженности	48
2.2 Анализ информационной инфраструктуры предприятия для создания отчетов по дебиторской задолженности.....	57
2.3 Сбор и анализ требований пользователей.....	59
3 Проектирование хранилищ данных для формирования отчетов с целью принятия управленческих решений	64
3.1 Проектирование хранилищ данных (концептуальная, логическая и физическая модели)	64
3.2 Преобразование измерений в системе «Optimacros»	75
3.3 Построение кубов и отчетов	81
4 Социальная ответственность	98
Заключение	108
Список использованных источников	111
Приложение А Раздел ВКР, выполненный на иностранном языке	116

Введение

В настоящее время, хранилища данных являются одной из самых актуальных тем в индустрии информационных технологий. В связи с тем, что за последние годы многие системы, которые проектировались с использованием традиционных методов и приемов, недостаточно хорошо оптимизированы для выполнения запросов к данным, многие поставщики стараются не отстать от современных вызовов и предлагают различные технологические решения данной проблемы.

Именно поэтому, на сегодняшний день происходит активное развитие технологий анализа данных. Одним из этапов данного изменения является создание и эксплуатация хранилищ данных с организацией периодического выполнения пакетных работ, которые направлены на извлечение из операционных данных неких исторических выборок данных, их очистку, преобразование и последующую загрузку в хранилище. Современные средства аналитики данных позволяют выявлять скрытые модели благодаря возможностям прогнозирования, самообучения и адаптации. Эти технологии интуитивны и включают возможность визуализации данных, которая помогает мгновенно анализировать миллионы строк и столбцов.

Аналитика данных распространяется на все аспекты нашей жизни, какими бы вопросами мы не задавались. Например, о сотрудниках, финансах или о том, что нравится или не нравится потребителям, а также, что влияет на их поведение. Аналитика дает ответы на все упомянутые вопросы и помогает принимать обоснованные решения.

На сегодняшний день, большая часть работы в организациях по-прежнему выполняется людьми, но автоматизация набирает все большее распространение. Пользователь выполняет запросы, затем анализируются результаты, взаимодействуя с визуальными представлениями данных, и создаются модели для прогнозирования будущих тенденций или выводов.

Цель магистерской диссертации – на основе анализа информационно-аналитической системы предприятия и требований пользователей разработать алгоритм построения хранилища данных для OLAP-системы с целью формирования аналитической отчетности предприятия по дебиторской задолженности.

Объект исследования – дочернее предприятие ПАО «Газпром».

Предмет исследования – процесс проектирования хранилища данных для OLAP-системы.

Для достижения поставленной цели необходимо решить следующие задачи:

1) изучение теоретических особенностей хранилищ данных и OLAP технологий;

2) анализ информационно-аналитической системы предприятия и процедуры формирования отчетов по дебиторской задолженности в этой системе;

3) выявление и анализ требований пользователей к процедуре формирования отчетов по дебиторской задолженности предприятия;

4) построение алгоритма проектирования хранилища данных и проведение анализа построения;

5) построение хранилища данных для дочернего предприятия ПАО «Газпром»;

6) построение отчёта для управления дебиторской задолженностью при помощи хранилища данных и OLAP технологий.

Теоретической и методологической основой данной работы послужили труды отечественных и зарубежных исследователей в области проектирования хранилищ данных и OLAP технологий.

Информационную базу данного исследования составили материалы научно-практических конференций и семинаров, публикации в научных и периодических изданиях по теме исследования, а также официальные сайты компаний, рассматриваемых в данной работе.

Практическая значимость состоит в том, что результаты разработок и предложенные рекомендации, полученные в итоге исследования, будут использованы для повышения эффективности управления дебиторской задолженностью дочернего предприятия ПАО «Газпром».

Научная новизна работы заключается в разработке алгоритма построения хранилища данных для OLAP-системы на базе платформы «Optimacros» для формирования аналитической отчетности предприятия по дебиторской задолженности с целью поддержки принятия обоснованных управленческих решений в организациях.

1 Теоретические аспекты построения хранилищ данных

1.1 Понятие и особенности структуры хранилищ данных

В современных условиях ведения бизнеса, для того, чтобы добиться действенных конкурентных преимуществ, необходим оперативный и глубокий анализ, основывающийся на данных, хранящихся в разных информационных системах. Именно поэтому, простого умения накапливать информацию об операционной деятельности компании, в настоящее время становится недостаточно. Технологическим решением данной проблемы является использование аналитических систем, которые сочетают в себе технологии хранилищ данных, оперативного анализа данных, интеллектуального анализа данных, а также современных технологий визуализации.

Концепция хранилища данных появилась с подачи компании IBM в конце 80-х, и была обусловлена развитием регистрирующих информационных систем. Именно в тот период возникло понимание ограниченности возможностей применения регистрирующих систем для анализа данных и построения эффективных систем поддержки и принятия решений [6].

Согласно классическому определению, хранилище данных (Data Warehouse) – предметно-ориентированная информационная корпоративная база большого объема, специально разработанная и предназначенная для сбора и предварительной обработки данных, с целью предоставления информации пользователям, для статистического анализа, подготовки отчетов и анализа бизнес-процессов [7, 24].

Хранилища данных создаются для решения задач анализа, планирования и прогнозирования. Данные в хранилищах представляют собой регулярно обновляемый и пополняемый набор, который является единым для всей организации и позволяет отразить общую картину деятельности на любой

период времени. Структура данных проектируется таким образом, чтобы выполнение запросов осуществлялись с максимальной эффективностью.

Основными характеристиками хранилищ данных являются [6, 10]:

1) предметная ориентированность – данные в хранилище объединяются в категории и хранятся в соответствии с предметными областями. Данная характеристика реализуется через использование особых схем организации данных, что позволяет упростить создание аналитических запросов, а также увеличить скорость их выполнения;

2) интегрированность – данные для анализа берутся не напрямую из источников. Исходные данные извлекаются, проверяются, очищаются и унифицируются для того, чтобы удовлетворять требованиям аналитика;

3) привязка ко времени – данные всегда привязаны к определенному периоду времени и накапливаются в хранилище в виде исторических слоев, каждый из которых относится к конкретному периоду, что позволяет анализировать все основные тенденции развития организации;

4) неизменяемость – после попадания в хранилище, данные уже никогда не будут изменены. Стабильность данных способствует достоверности результатов аналитических запросов.

Кроме того, существует перечень определенных требований, предъявляемых к хранилищам данных, а именно [12, 27]:

- 1) поддержка высокой скорости получения данных из хранилища;
- 2) возможность получения и сравнения срезов данных;
- 3) наличие удобных утилит просмотра данных в хранилище;
- 4) поддержка внутренней непротиворечивости данных;
- 5) полнота и достоверность хранимых данных;
- 6) поддержка процесса пополнения данных.

Стоит отметить, что хранилища данных и базы данных не синонимичные понятия. Существует несколько основных отличий [27, 49]:

1) данные, находящиеся в хранилище, доступны исключительно для чтения;

2) базы данных подвержены регулярным изменениям в процессе работы пользователей, а хранилище данных относительно стабильно. Так, данные в хранилище, как правило, обновляются согласно расписанию (ежедневно / еженедельно). Процесс обновления представляет собой добавление новых данных за определенный период без влияния на прежнюю информацию, которая уже какое-то время находится в хранилище;

3) данные, находящиеся в хранилищах, предназначены не для повседневной работы, как в обычных базах данных, а для обеспечения принятия решений, требующих анализа, прогнозирования и планирования.

Таким образом, обычные базы данных представляют собой источники данных, которые попадают в хранилище. Кроме того, хранилище также может пополняться за счет внешних источников, например, статистических отчетов.

Источниками данных могут выступать традиционные системы сбора и регистрации данных, отдельные документы, а также наборы данных. При этом, они характеризуются: территориальным и административным размещением, степенью достоверности, частотой обновления, а также системой хранения и управления данными [28, 39].

Далее, необходимо рассмотреть структуру хранилищ данных. Типичная структура хранилищ данных денормализована, что напрямую влияет на повышение скорости выполнения запросов. Основными составляющими структуры хранилищ данных являются таблицы измерений и фактов [7, 10].

Говоря о таблицах измерений, они содержат неизменяемые или редко изменяемые данные. В каждой такой таблице содержится первичный ключ для идентификации данных, находящихся в таблице, а также один или несколько описывающих атрибутов и ссылки на другие измерения.

Говоря о таблицах фактов, можно сказать, что они могут содержать невероятно огромное количество записей, в связи с чем, хранение повторяющихся текстовых описаний не является выгодным. Намного более лучшим вариантом является – их перемещение в меньшие по объему таблицы измерений.

Таблицы фактов – это основа хранилища данных, которая содержит сведения об объектах или событиях, совокупность которых будет в дальнейшем подвергаться анализу. Кроме того, таблицы также содержат факты, дающие числовую характеристику какому-либо аспекту предметной области и внешние ключи для доступа к таблицам измерений, содержащих описание объектов [10, 19].

Для многомерного анализа пригодны таблицы фактов, содержащие как можно более подробные данные, т.е. соответствующие членам нижних уровней иерархии соответствующих измерений. Например, предпочтительнее взять за основу факты продажи товаров отдельным заказчикам, а не суммы продаж для разных регионов, т.к. последние будут вычислены OLAP-средством в процессе агрегирования.

Стоит отметить, что в таблице фактов нет никаких сведений о том, как группировать записи при вычислении агрегатных данных. Например, в ней могут быть значения ключевых атрибутов товаров или клиентов, но отсутствует информация о том, к какой категории или типу может относиться данный товар. Эти сведения, в дальнейшем используемые для построения иерархий в измерениях куба, содержатся в таблицах измерений. Как ключевые, так и некоторые не ключевые поля (например, дата) соответствуют будущим измерениям OLAP куба, а числовые поля, содержащиеся в таблице фактов, используются в дальнейшем для получения агрегатных данных [32].

Одно измерение куба может содержаться как в одной таблице (в том числе и при наличии нескольких уровней иерархии), так и в нескольких связанных таблицах, которые соответствуют различным уровням иерархии в измерении. OLAP-куб создаётся из соединения таблиц фактов и измерений с применением схемы «звезды» или схемы «снежинки». Рассмотрим более подробно каждую из них.

Говоря о схеме «звезда» (схема звёздного соединения), можно сказать, что она используется, в том случае, если информация об измерении располагается в отдельной таблице. Другими словами, она представляет собой

модель данных, в центре которой находится таблица фактов, а таблицы измерений связаны с ней при помощи ключей (рис. 1) [14, 27].

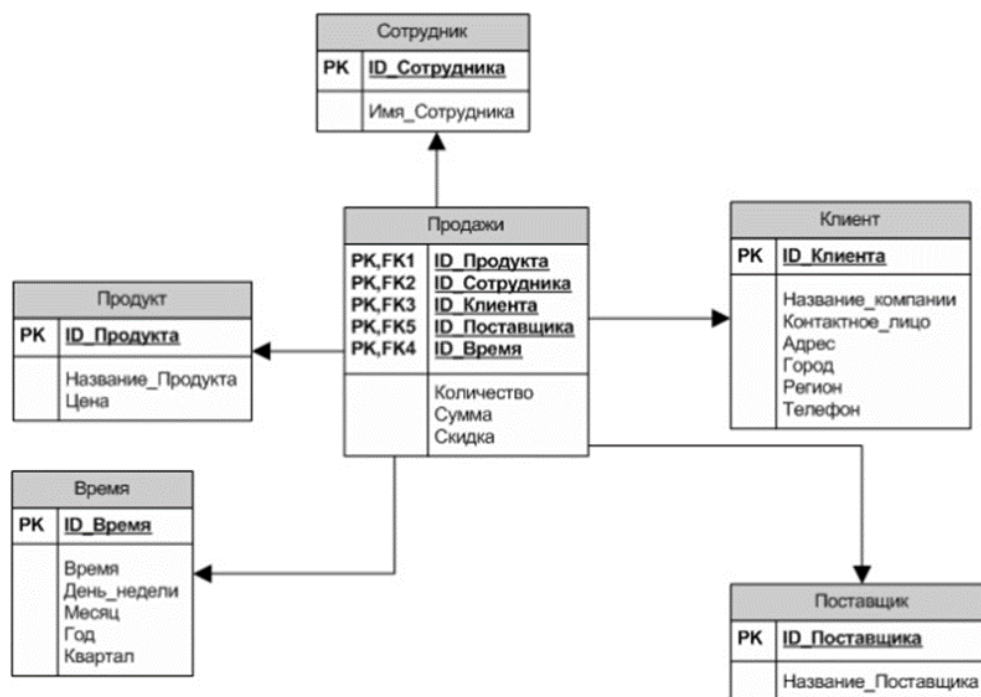


Рисунок 1 – Пример схемы «Звезда» хранилища данных [27]

Для более эффективной работы с иерархическими измерениями была разработана модификация схемы звезды, которая получила название «снежинка». Подобное название она получила за счет своей формы, в виде которой отображается логическая схема таблиц в многомерной базе данных.

Говоря о схеме «снежинка», стоит отметить, что размещение всей информации об измерении в одной таблице оказывается не всегда оправданным. Например, если продаваемые товары объединены в группы (имеет место иерархия), то тем или иным способом необходимо показать, к какой группе относится каждый товар, что может привести к многократному повторению наименований групп. Подобное не только вызовет рост избыточности, но также повысит вероятность возникновения противоречий, в том случае, если, один и тот же товар ошибочно отнесут к разным группам.

Так же, как и в схеме «звезда», схема «снежинка» (рис. 2) представлена централизованной таблицей фактов, соединенной с таблицами измерений. Отличительной особенностью является то, что в схеме «снежинка» одни

таблицы измерений могут быть связаны с другими таблицами измерений. Причем, чем больше связей между ними, тем сложнее выглядит структура схемы «снежинки» [14, 27].

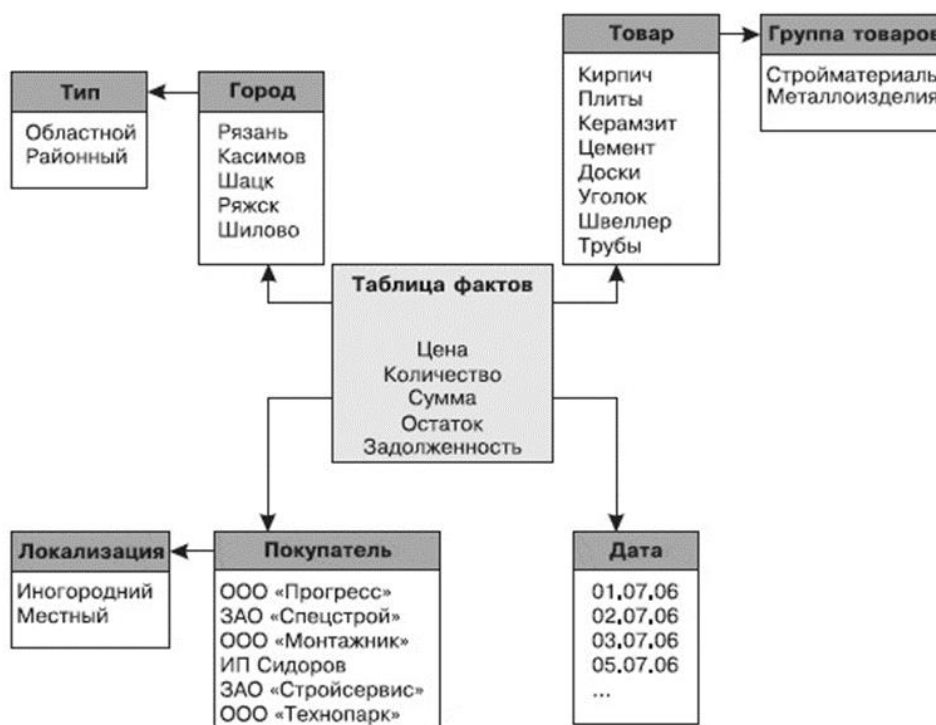


Рисунок 2 – Пример схемы «Снежинка» хранилища данных [27]

Недостатками схемы «снежинка» по сравнению со схемой «звезда» являются: более трудная для понимания и реализации структура данных и более сложная процедура добавления значений в измерения.

К преимуществам схемы «снежинка» можно отнести: лучшее соответствие представлению данных в многомерной модели, низкая вероятность появления ошибок несоответствия данных и компактность представления иерархических данных.

Так, особенность хранилищ данных состоит в организации периодического выполнения пакетных работ, которые направлены на извлечение из операционных данных исторических выборок, их очистку, преобразование и загрузку. Информация является ценным ресурсом, ценность которого зависит от времени. В результате чего, ее значимость для бизнеса напрямую зависит от своевременности и достоверности.

1.2 Анализ методологических подходов построения хранилищ данных

Прежде чем изучить существующие подходы к реализации хранилищ данных, необходимо определить основные уровни передачи данных. Информационную инфраструктуру предприятия можно представить в виде пяти иерархических уровней, каждый из которых характеризуется степенью агрегированности информации и своей ролью в процессе управления. Для графической наглядности представления информационной инфраструктуры предприятия может служить «аналитическая пирамида», разработанная компанией «Gartner» и представляющая собой иерархическую структуру, в которой различные классы информационных систем расположены на различных уровнях, а именно [4, 11, 47]:

- 1) уровень транзакционных систем;
- 2) уровень хранилищ данных;
- 3) уровень витрин данных;
- 4) уровень OLAP-систем;
- 5) уровень аналитических приложений.

Пример аналитической пирамиды, отражающей движение по данным уровням представлен на рис. 3.

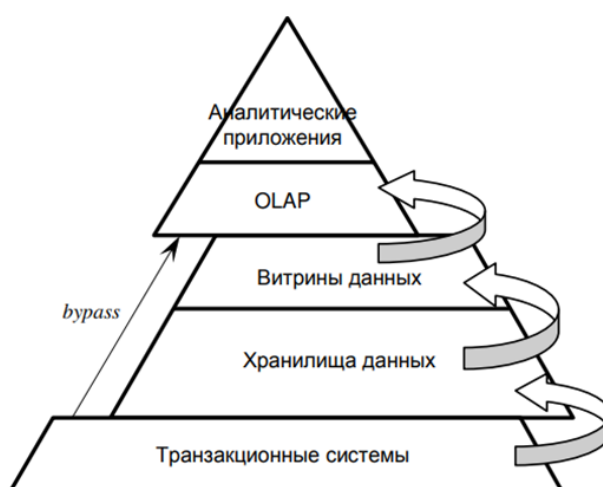


Рисунок 3 – Аналитическая пирамида [11]

Данный иерархическая структура достаточно хорошо отражает возможные варианты передачи данных. Так, данные из транзакционных систем могут поступать в аналитические приложения как последовательно, так и более коротким путем, минуя один или несколько уровней («bypass» – «прямая передача»). Способ передачи данных зависит от технических возможностей программных продуктов, а также от того, каким образом предполагается дальнейшее использование тех или иных данных [29].

Так, для того, чтобы обеспечить движение данных, необходимо осуществить проектирование хранилища данных в соответствии с выбранной архитектурой. Архитектура хранилищ данных представляет собой информационную систему, содержащую исторические и коммутативные данные из нескольких источников. Особенностью любой архитектуры является осуществление проектирования на основе пяти классов данных [10]:

- 1) данные из источников;
- 2) оперативный склад данных;
- 3) хранилища данных;
- 4) витрины данных;
- 5) метаданные (данные о данных).

Оперативный склад данных создается как промежуточный буфер между оперативными системами и хранилищем данных. Он служит аккумулятором данных, поступающих от источников, перед их загрузкой в хранилище. Данные, попадая в оперативный склад, преобразовываются в единые форматы, очищаются, объединяются и синхронизируются [29, 37].

Оперативный склад данных, как и хранилище, предметно ориентирован, но имеет одно отличие – информация в складе может изменяться со временем в соответствии с изменениями, происходящими в источниках данных.

Витрины данных (Data Mart) можно представить в виде логически или физически разделенных подмножеств хранилища данных. Иногда эти структуры хранения данных называют также «киосками данных». Обычно они строятся для обслуживания потребностей определенной группы

пользователей. Именно с витринами данных и работает большинство конечных пользователей [37].

Говоря о метаданных, стоит отметить, что это важнейший элемент хранилища данных, представляющий собой информацию о структуре, размещении и трансформации данных. Именно благодаря метаданным обеспечивается эффективное взаимодействие различных компонентов хранилища [32].

Существует два основных подхода к архитектуре хранилищ данных, а именно: корпоративная информационная фабрика CIF Билла Инмона и хранилище данных с архитектурой шины Ральфа Кимболла.

Согласно подходу Инмона, среда баз данных делится на четыре основных уровня, а именно [10]:

- оперативный уровень;
- хранилище с атомарными данными;
- ведомственный уровень;
- индивидуальный уровень.

Первый уровень содержит данные из систем обработки транзакций и предназначен для поддержки ежедневного функционирования организации. Данные из оперативных систем приводятся к единообразию и загружаются в хранилище с атомарными данными. Архитектура Билла Инмона гарантирует согласованность данных, т.к. все ведомственные данные поступают из атомарного хранилища. Причем, последние три уровня включают в себя непосредственно само хранилище данных, а четвертый уровень, как правило, носит временный характер и размещается на персональных компьютерах пользователей.

Билл Инмон определяет хранилище данных в качестве централизованного репозитория для всего предприятия, где хранилище находится в центре «Корпоративной информационной фабрики», обеспечивая логическую основу для предоставления бизнес-аналитики (рис. 4) [10].

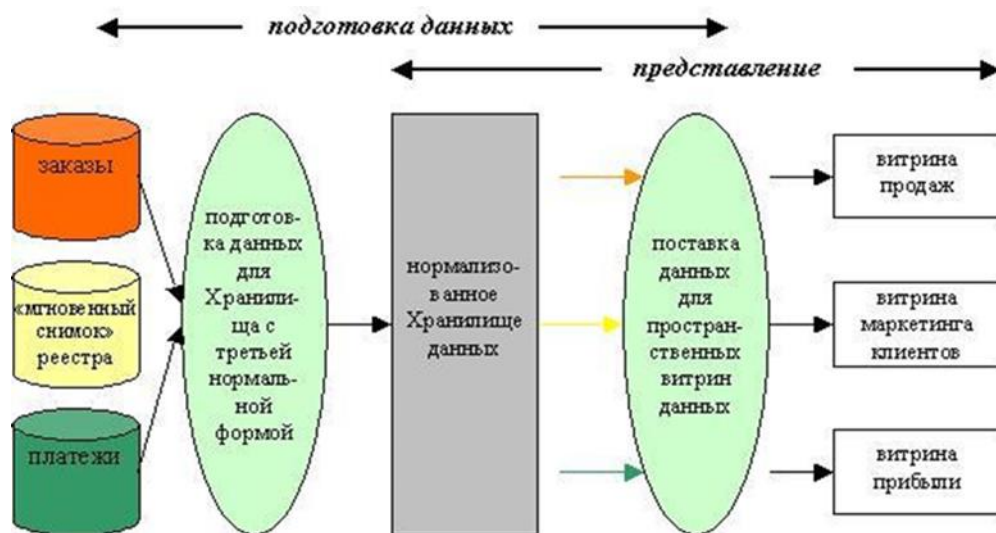


Рисунок 4 – CIF (Corporate Information Factory) Билла Инмона [10]

Преимуществами подхода CIF (Corporate Information Factory) Билла Инмона являются [3, 24]:

- целостность данных в витринах данных;
- единое хранилище данных на физическом уровне;
- быстрота создания хранилища;
- высокая степень детализации данных;
- централизованное управление деятельностью организации.

Недостатки подхода CIF (Corporate Information Factory) Билла Инмона являются [24]:

- отсутствие прямого доступа в хранилище;
- сложность в использовании системы обычными пользователями;
- затраты больших объемов ресурсов на создание и поддержание.

Далее, рассмотрим подход к проектированию хранилища данных с архитектурой шины BUS (Data Warehouse Bus) Ральфа Кимболла. В подходе Кимболла первичные данные преобразуются в информацию, пригодную для использования, на этапе подготовки данных, которая начинается со скоординированного извлечения данных из источников. Пространственная модель хранилища данных содержит ту же атомарную информацию, что и нормализованная модель согласно подходу Билла Инмона, однако

информация структурирована по-другому для того, чтобы облегчить ее дальнейшее использование, а также выполнение запросов [10, 24].

Согласно подходу Ральфа Кимболла, модель включает как атомарные данные, так и обобщающую информацию в соответствии с требованиями производительности или пространственного распределения данных. Запросы в процессе выполнения обращаются к более низкому уровню детализации без дополнительного перепрограммирования со стороны пользователей.

В отличие от подхода Билла Инмона, пространственные модели строятся для обслуживания бизнес-процессов. Например, данные о заказах, которые должны быть доступны для общекорпоративного использования, вносятся в пространственное хранилище данных один раз, в отличие от СIF-подхода, в котором их пришлось бы копировать сначала в витрины данных отделов маркетинга, а затем продаж и финансов. После появления информации о бизнес-процессах, консолидированные пространственные модели могут выдавать их характеристики [24].

Матрица корпоративного хранилища данных с архитектурой шины выявляет и усиливает связи между показателями бизнес-процессов – фактами и описательными атрибутами – измерениями (рис. 5) [10].

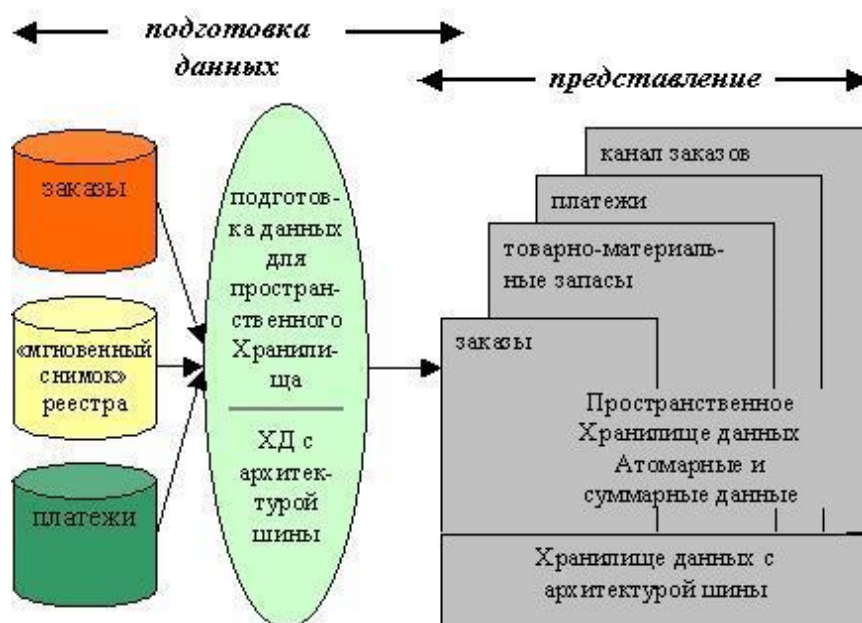


Рисунок 5 – BUS (Data Warehouse Bus) Ральфа Кимболла [10]

Таким образом, архитектура хранилищ данных согласно подходу Ральфа Кимболла основана на следующих основных принципах [10, 24]:

- использование пространственной модели организации данных с архитектурой «звезда»;
- хранилище представляет собой коллекцию витрин данных, которые могут быть пространственно разобщенными;
- в состав хранилища входят несколько витрин (атомарных данных, агрегированных данных, персональная витрина);
- использование двухуровневой архитектуры, которая включает стадию подготовки данных, недоступную для конечных пользователей, и хранилище;
- хранилище данных не является единым физическим репозиторием.

Так, оба подхода направлены на создание корпоративного хранилища данных, в результате чего, единство конечного объекта говорит о схожести требований, которым должен удовлетворять любой из названных подходов.

Первое требование – для принятия обоснованных управленческих решений, варьирующих от проблем выживания бизнеса до роста прибыли, предприятия нуждаются в программном продукте для хранения, анализа и интерпретации накапливаемых со временем данных. Для достижения поставленных целей, предприятиям необходимо в полной мере использовать возможности первичных данных, что, в свою очередь, требует соответствующих операционных систем и систем обратной связи. Именно для этого создаются хранилища данных.

Еще одним требованием является – точность и своевременность данных. Другими словами, каждый отдельный пользователь должен иметь доступ к данным при помощи понятных и простых способов построения запросов [24].

В техническом отношении оба подхода признают необходимость наличия в хранилище как атомарных, так и суммарных данных, а также доступность обеих категорий пользователям для анализа.

Говоря об отличиях, стоит отметить, что данные подходы имеют различные методологии к построению баз данных, которые составляют основу хранилища. Так, если Ральф Кимболл использует пространственную организацию баз данных с архитектурой «звезда» как на стадии подготовки, так и презентации данных, то Билл Инмон комбинирует два подхода. В его модели атомарные данные организованы в реляционные базы и находятся в нормализованном хранилище данных. Причем, суммарные данные доступны для использования через специализированные хранилища и средства (OLAP) [16, 48].

Таким образом, рассмотренные архитектуры отличаются между собой способами обращения с атомарными данными: их пространственной организацией у Кимболла и нормализованной у Инмона.

Еще одним принципиальным отличием является вопрос физической организации хранилища. Согласно подходу Инмона, хранилище данных представляется физически целостным реально существующим объектом, а в соответствии с подходом Кимболла, хранилище данных представляет собой виртуальный объект [16].

Указанные отличия являются принципиальной разницей между данными подходами. Стоит отметить, что на практике подавляющее большинство организаций предпочитают использовать подход Ральфа Кимболла, т.к. подобная архитектура хранилища данных позволяет усиливать связи между показателями бизнес-процессов и описательными атрибутами. Кроме того, пространственная организация с архитектурой «звезда» облегчает быстрый доступ к данным и требует меньше времени на выполнение запросов, упрощая работу с атомарными данными. Схема предоставляет широкие возможности для манипулирования атомарными данными и изменения их формата и способа представления по мере необходимости. Однако, вне зависимости от конечного выбора организации, оба подхода к архитектуре хранилищ данных являются весьма эффективными.

1.3 Управление хранилищем данных

Во все времена, данные представляют собой огромную ценность для компаний, которые могут работать с ними, несмотря на затраты времени, средств и сложности, которые связаны с этим процессом. Если говорить о создании высоко функциональной и надежной системы, то ключевой элемент этого процесса – регулярное и бдительное управление. В противном случае, могут появиться проблемы, которые способны существенно снизить эффективность и ценность хранилища данных для бизнеса.

Управление хранилищами данных осуществляется при помощи ETL-систем (Extraction, Transformation, Loading), которые представляют собой комплекс программных средств, предназначенный для извлечения данных из различных и разнотипных источников, приведения их к единому формату, оценки их качества, очистки, обогащения и загрузки в хранилище с целью дальнейшей аналитической обработки (рис. 6) [34, 35].

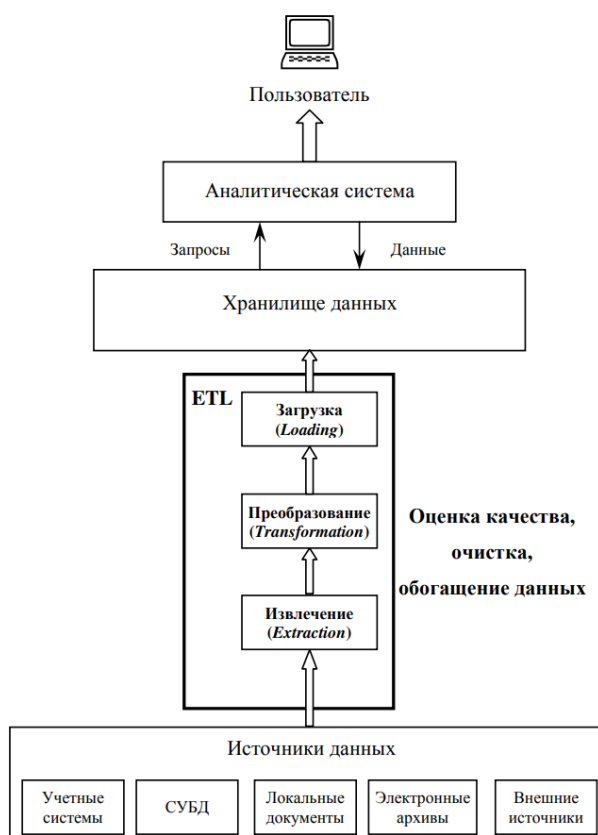


Рисунок 6 – Алгоритм процесса ETL [35]

Процесс ETL преобразует совокупность данных оперативной системы и других источников в многомерную модель хранилища данных. Движение данных к приёмнику от источника называют «потоким данных». Необходимые для работы потоки данных формируются и описываются аналитиками.

На этапе извлечения, исходные данные загружаются в промежуточную область, как правило, в виде вспомогательных таблиц, и подвергаются проверке на соответствие спецификациям и возможности последующей загрузки в хранилище данных.

Стоит отметить, что для правильного представления данных в хранилище из источников должны извлекаться не только сами данные, но и информация, описывающая их структуру, из которой будут сформированы метаданные для хранилища [40, 51].

В некоторых случаях, когда требуется извлекать данные из нескольких источников одного типа, для них создается общая таблица (рис. 7) [35, 51].

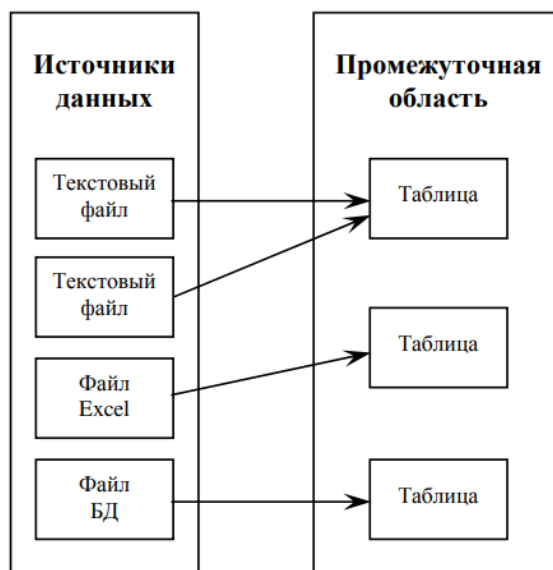


Рисунок 7 – Извлечения данных из нескольких источников [51]

Извлечение данных из источников разных типов связано с проблемами преобразования форматов данных, созданных в различных приложениях и, вероятно, использующих различную кодировку, в единый универсальный формат, который поддерживается хранилищем данных.

Исходные данные часто содержат факторы, мешающие их дальнейшей корректной обработке: пропуски, противоречия, дубликаты, аномальные значения, шумы и т.п. Поэтому в процессе ETL производится выявление таких факторов и их устранение, т.е. очистка данных.

Так как хранилища данных могут строиться на основе различных моделей, ETL процесс реализуется с учетом всех особенностей используемой модели. В процессе преобразования данные группируются и приводятся к виду, соответствующему структуре хранилища (рис. 8) [51, 52].

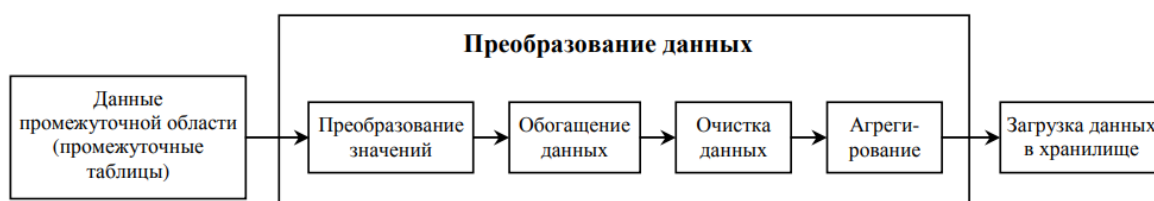


Рисунок 8 – Схема преобразования данных [51]

Преобразование значений необходимо в том случае, если коды, используемые в источниках для обозначения наименований одних и тех же объектов, их свойств, признаков, характеристик, не совпадают или не соответствуют обозначениям, принятым в хранилище. В таком случае, их необходимо преобразовать в соответствии с требованиями хранилища.

Обогащение данных включает как поиск дополнительной информации (внешнее обогащение), так и создание новой (внутреннее обогащение). Нередко данных недостаточно для последующего анализа. В данном случае производится поиск дополнительной информации. Такой процесс называется внешним обогащением данных. Внешнее обогащение часто связано со сбором информации об объектах предметной области, участвующих в бизнес процессе. Например, данные о результатах исследований предпочтений клиентов на основании проведенного анкетирования [53].

Создание новых данных (внутреннее обогащение) может понадобиться в том случае, если требуется вычисление каких-либо параметров на основе существующих. Например, система содержит поля о количестве и стоимости

проданного товара, а в таблице хранилища данных есть поле «Сумма». Тогда необходимо вычислить сумму как произведение стоимости товара на количество проданных единиц. Таким образом, будет создано новое поле, которое содержит дополнительную информацию. Новыми данными, могут выступать как экономические, финансовые, так и другие показатели, которые можно вычислены на основе имеющихся данных.

Так, основываясь на данных о продажах можно рассчитать рейтинг товаров и создать новое поле, в котором для каждого товара будет указано соответствующее рейтинговое значение по какой-либо выбранной шкале (например, по пятибалльной, десятибалльной и т.п.).

Сбор данных в процессе ETL производится из большого количества разнообразных источников, многие из которых не содержат автоматических средств поддержки целостности, непротиворечивости и корректного представления данных. В связи с этим, данные могут стать причиной неправильных результатов анализа и даже сделать невозможным применение некоторых аналитических алгоритмов и методов. Поэтому, при разработке стратегии ETL этапу очистки уделяется достаточно большое внимание [6, 25].

Очистка данных – одна из наиболее важных и сложных задач ETL-процесса. Особое внимание уделяется данным, которые содержат наиболее критичные ошибки, из-за которых они не могут быть загружены в хранилище. Такими ошибками могут быть, например, некорректные форматы данных – буква или пробел в числовом значении, неправильный разделитель и т.д.). Наиболее критичными принято считать ошибки, которые усложняют дальнейшую работы с данными и делают невозможным последующее взаимодействие [33].

Очистка также применяется для согласования полей с атрибутами хранилища данных. В процессе преобразования данных в случае необходимости сразу же производится их агрегирование.

Последним этапом ETL процесса является загрузка данных в хранилище. Загрузка заключается в переносе данных, подвергшихся

преобразованию в структуры хранилища данных из промежуточных таблиц. Первыми в хранилище загружаются данные в таблицы измерений, которые содержат описательную информацию, необходимую для таблиц фактов, а затем загружаются данные в таблицы фактов. После завершения загрузки, перед тем, как сделать данные доступными для анализа, необходимо убедиться в их корректности и достоверности. Для этого требуется выполнить дополнительные операции, которые позволят сравнивать данные как в различных разрезах, так и с источниками данных.

Подобными операциями могут быть [5, 8]:

1) общая сумма за месяц должна соответствовать ежедневной сумме по исследуемому месяцу;

2) общая выручка по всем отделениям сбыта за текущий месяц должна соответствовать сумме продаж по отдельным отделениям и т.п.

Другими словами, необходимо наличие вспомогательных итоговых ячеек, которые будут служить ориентиром для определения корректности расчетов по отдельным составляющим.

Стоит также отметить, что система управления хранилищем данных состоит, из трех подсистем, а именно: подсистемы загрузки данных, подсистемы представления данных и обработки запросов и подсистемы администрирования хранилища.

Говоря отдельно о каждой из них, подсистема загрузки данных представляет собой программное обеспечение управления оперативным складом. При помощи данного управления можно извлекать данные из источников в соответствии с определенным заранее регламентом и приводить их к единому формату. Данная подсистема отвечает за логическую согласованность, качество, а также интеграцию данных.

Каждый отдельный источник данных может потребовать разработку собственного загрузочного модуля, который должен решать два типа задач, а именно: первоначальную загрузку данных в оперативный склад и регламентное пополнение хранилища данными [11, 12].

Кроме того, данная подсистема также по регламенту извлекает детализированные данные из оперативного склада, осуществляет их агрегирование, консолидацию, трансформацию, после чего – помещает данные в хранилище и витрины данных. Именно в этой подсистеме должны быть определены все бизнес-модели консолидации данных по иерархическим измерениям и вычисления зависимых бизнес-показателей по независимым исходным данным.

Говоря о подсистеме обработки запросов и представления данных, можно сказать, что она характеризуется, как специальное программное обеспечение, которое служит для извлечения данных из хранилища, их аналитической обработки и представления конечным пользователям. Как правило, можно выделить три типа данного программного обеспечения, предназначенные для выполнения следующих задач [15, 19]:

1) регламентированная отчетность – заранее предопределенные запросы и формы представления результатов. В данном случае не требуется быстрой скорости реакции, в результате чего из соображений экономии для реализации часто используют технологию ROLAP, т.е. реляционная OLAP-системы, имеющие прямой доступ к существующим базам данных или использующие данные, которые уже выгружены в локальные таблицы.

2) Нерегламентированные запросы пользователей – способ общения с хранилищем данных, при котором каждый последующий запрос определяются результатами предыдущего. Для такого программного обеспечения, в отличие от регламентированной отчетности, требуется высокая скорость обработки запросов. Именно поэтому, оно реализуется на основе технологии MOLAP. MOLAP – множественная система, состоящая из нескольких независимых измерений, вдоль которых могут быть проанализированы определенные совокупности данных. Также, данное ПО реализуется специальными инструментами построения сложных нерегламентированных запросов с интуитивно понятным для аналитиков графическим интерфейсом.

3) Выявление скрытых закономерностей, которое реализует сложные статистические алгоритмы и алгоритмы искусственного интеллекта, предназначенные для поиска скрытых закономерностей в данных, а также представления этих закономерностей в виде моделей и многовариантного прогнозирования по ним развития ситуаций по схеме «IF THEN ELSE».

Подсистема администрирования хранилища выполняет различные задачи, связанные с поддержанием системы и обеспечением ее устойчивой и эффективной работы и расширения [28, 30]:

- 1) администрирование метаданных системы;
- 2) администрирование доступа к данным (обеспечивает сопровождение профилей пользователей, разграничение доступа к конфиденциальным данным);
- 3) администрирование хранилища данных (поддержание архитектуры хранилища, его эффективной и непрерывной работы, защита и восстановление данных после сбоев).

Таким образом, управление является одной из главных составляющих при работе с хранилищем данных. Кроме того, отражение реальной картины быстро меняющегося состояния рынка, а также адекватное реагирование и принятие обоснованных управленческих решений влияют на специфику построения хранилища данных. В результате чего, необходимо тщательно подходить к каждому этапу работы с хранилищем.

Возможность получения первых результатов в кратчайшие сроки достаточно сильно повышает вероятность положительного результата всего проекта. Именно поэтому, необходимо основываться на наиболее понятной методологии построения хранилищ данных, поэтапно реализовывая сложную систему в хорошо выверенной на практическом опыте архитектуре и осуществлять грамотное управление данными хранилища. Так, все этапы разработки и эксплуатации системы должны опираться на надежные, хорошо зарекомендовавшие себя в многочисленных проектах средства, полностью интегрированные в единую среду.

1.4 OLAP технология

В предыдущих параграфах данной работы многократно упоминались такие понятия, как «OLAP-системы» и «OLAP-средства». Термин «OLAP» неразрывно связан с понятием хранилищ данных (Data Warehouse). OLAP (OnLine Analytical Processing) – система аналитической обработки данных, предназначенная для подготовки отчетов, построения прогностических сценариев и выполнения статистических расчетов на базе больших информационных массивов, имеющих сложную структуру [7, 18].

Концепция OLAP была впервые сформулирована в 1993 году Е.Ф. Коддом. Система представляет собой совокупность концепций, принципов и требований, которые лежат в основе программных продуктов, облегчающих аналитикам доступ к данным для быстрой обработки сложных запросов к хранилищу.

Основной причиной использования OLAP для обработки запросов является – скорость. Реляционные базы данных хранят данные в отдельных таблицах, в связи с чем, сложные многотабличные запросы в них выполняются достаточно медленно. Время обработки запросов в OLAP-средствах значительно меньше и составляет примерно 0,1% от аналогичных запросов в реляционную базу данных [18, 51].

В настоящее время, превращение разрозненных данных в структурированную информацию, можно считать одной из самых актуальных задач, стоящих перед современными предприятиями. Причем, обработка данных и принятие решений должны быть настолько оперативными, насколько этого требуют интересы бизнеса. Именно для этого предназначены специальные системы аналитической обработки данных – OLAP-системы.

Основная цель OLAP-средств – обеспечивать необходимой информацией пользователей и системы поддержки принятия решений. При решении различных вопросов во время анализа предметной области требуется формирование разнообразных выборок, которые носят характер «набор

атрибутов – числовой параметр». OLAP-технология направлена на быстрое извлечение подобных данных из хранилища и ее предоставление в удобной для пользователей и программных систем виде. Данные, необходимы для анализа могут располагаться как в одном месте, так и в разных исходных таблицах, а их извлечение может производиться по достаточно сложным алгоритмам [35, 48].

Какое-либо событие, например, продажа, характеризуется множеством параметров, которые принимают как качественные, так и количественные значения. Каждое конкретное событие представляется последовательностью конкретных значений этих параметров.

Таким образом, цель OLAP-технологии состоит в анализе зависимости значений количественных параметров – фактов от значений качественных параметров – измерений (рис. 9) [32, 35].

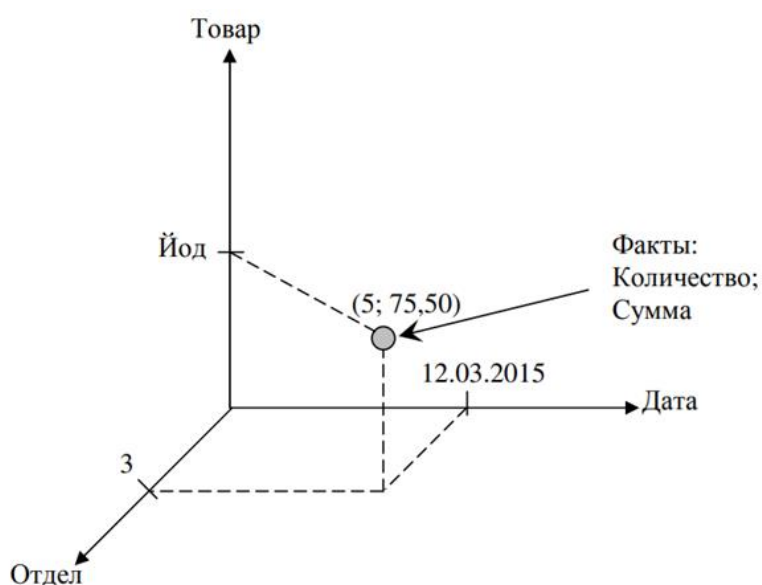


Рисунок 9 – Пример измерений и факта [35]

Задача аналитика состоит в определении стойких взаимосвязей между измерениями (зависимыми данными) и фактами (числовыми параметрами). Таким фактом может быть, например, выручка за период.

Если требуется провести анализ зависимости данного параметра от поставщиков, вида товара и временного периода, то OLAP-технология

позволяет быстро сформировать необходимую выборку, извлекая данные из хранилища, и предоставляя ее конечному пользователю. Причем, исходные данные далеко не всегда могут быть сгруппированы в соответствии с требованиями. Например, если пользователя интересуют данные за год, а в хранилище поступали данные только по итогам недели.

Подобные данные можно представить в виде реляционной таблицы, в строках которой располагаются значения атрибутов. При этом, один из атрибутов является анализируемым фактом, который предполагается зависимым от других. Однако на практике подобная форма представления данных является далеко не эффективной.

Таким образом, по данным рис. 10 можно увидеть, что даже при небольших объемах информации, отчет представленный в виде многомерной (двумерной) таблицы намного нагляднее и информативнее отчета с реляционной (построчной) формой организации [15, 17].

Реляционная модель

Товар	Дата	Объем продаж
Холодильник	12.03.2015	12
Холодильник	20.05.2015	5
Телевизор	12.03.2015	2
Телевизор	08.04.2015	18
Электропечь	08.04.2015	19

Многомерная модель

Товар	12.03.2015	08.04.2015	20.05.2015
Холодильник	12		5
Телевизор	2	18	
Электропечь		19	

Рисунок 10 – Варианты представления данных [15]

Так, если компания имеет не три вида товара, а тридцать и необходимо исследовать не три месяца, а все месяцы года, при реляционном представлении получится отчет в 360 строк. В случае же многомерного представления данных можно получить достаточно компактную таблицу, которую даже при подобном объеме данных, можно адекватно анализировать.

В многомерной модели каждое значение факта определяется набором значений измерений. В рассмотренном примере каждое значение факта «Объем продаж» определяется набором значений измерений «Товар» и «Месяц». Причем, в многомерной модели не все значения показателей должны иметь и реально имеют какие-либо значения.

Факт может быть определен как переменная или формула. Значения переменных единожды вводятся из внешнего источника, после чего хранятся в базе данных и не подвергается изменению. Значения фактов, которые имеют тип «Формула», вычисляются по заранее сформированным формулам и алгоритмам. Причем, для такого факта в базе хранится не его значения, а именно формула, по которой эти значения могут быть вычислены в дальнейшем [50].

Разумеется, значение факта может зависеть от значений любого количества измерений. Например, если в последнем примере добавить дополнительное измерение «Менеджер», то значение факта будет зависеть уже от значений трех измерений. Углубляясь в терминологию, речь будет идти уже не о двумерной таблице, а о трехмерном кубе.

Основная суть технологии OLAP заключается в проектировании многомерных кубов из рабочих данных, доступных для запросов пользователей. Здесь по одной оси откладываются значения первого атрибута, по другой оси – значения второго атрибута и т.д. Так, пользователь получает интуитивно понятную модель данных, организованную в виде многомерных кубов [43, 50].

Таким образом, по одной оси куба откладываются, например, места продаж, по другой – виды товара, по третьей – период (годы), а значениями анализируемого факта в этой многомерной (трехмерной) структуре будут являться соответствующие им объемы продаж. Размещение измерений может осуществляться в зависимости от основных требований пользователей. Графическая интерпретация представлена на рис. 11 [15, 17].

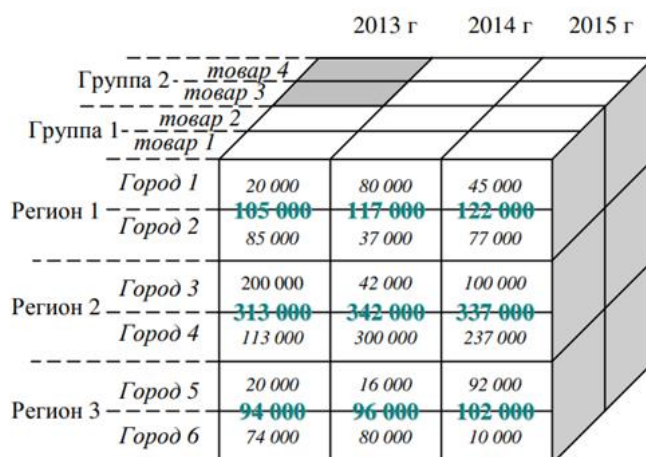


Рисунок 11 – Многомерный куб [15]

На рисунке 11 серым сегментом показано, что для «Город 1» данных о продаже «Товара 3» и «Товара 4» в 2013 году нет. Куб также можно представить в виде двумерной таблицы, в столбцах (и/или строках) которой будут находиться два или более измерений куба (рис. 12) [15, 17].

Место	товар 1			товар 2			товар 3			товар 4		
	2013 г	2014 г	2015 г	2013 г	2014 г	2015 г	2013 г	2014 г	2015 г	2013 г	2014 г	2015 г
Регион 1	105 000	117 000	122 000	40 000	68 000	90 000	30 000	75 000	78 000	130 000	115 000	80 000
Регион 2	313 000	342 000	337 000	80 000	202 000	250 000	25 000	65 000	70 000	77 000	79 000	85 000
Регион 3	94 000	96 000	102 000	120 000	45 000	80 000	46 000	50 000	55 000	125 000	115 000	120 000

Рисунок 12 – Двумерная таблица куба [15]

На рисунке 12 в ячейках таблицы указаны значения факта «Объем продаж», агрегированные по регионам и группам товаров. Основываясь на количестве измерений в кубе, можно определить его размерность. Куб OLAP может быть двумерным, трехмерным и т.д. в зависимости от количества измерений.

Элементы (значения), совокупность которых образует измерение, называются метками. Метки визуальнo отображаются как точки, откладываемые на гранях мультикуба. Например, измерение «Место» может состоять из меток Томск, Северск и т.д. В кубе могут быть указаны не все значения фактов. Также, совершенно необязательно, чтобы система с OLAP хранила данные исключительно в многомерной структуре. Ключевым

моментом является то, что именно для пользователя эти данные должны выглядеть именно так [7, 10].

Стоит также отметить, что именно за счет специальных способов хранения многомерных данных, незаполненные элементы кубов не приводят к потере памяти.

В хранилищах данных, основанных на многомерной модели, используются два основных варианта организации данных: поликубическая и гиперкубическая модель.

Системы, поддерживающие поликубическую модель (например, Oracle Express Server), предполагают, что в хранилище может быть определено множество гиперкубов с различной размерностью и с различными измерениями в качестве их осей. Например, значение факта «Количество работников», скорее всего, не зависит от измерения «Список товаров» и однозначно определяется двумя измерениями: «Год» и «Район». В поликубической модели в подобном случае может быть объявлено два различных гиперкуба: двумерный для факта «Количество работников» и трехмерный для факта «Объем продаж» [10, 31].

В случае гиперкубической модели, как правило, все факты определяются одним набором измерений. Другими словами, если «Объем продаж» определяется тремя измерениями, при описании факта «Количество работников» необходимо использовать те же три измерения, которые являются избыточными для данного показателя «Список товаров» [10].

Многомерная модель допускает применение операций: агрегация, детализация, формирование среза и вращение [18, 25]:

1) операция агрегации – для пользователя измерения «Отдел», «Город» и «Регион» являются точно такими же, как и «Год». Однако каждое из них соответствует более высокому уровню агрегации значений показателя «Объем продаж». В процессе проведения анализа пользователь осуществляет работы с разными срезами данных и переходит от детализированных данных к агрегированным, т.е. производит операцию агрегации. Например, посмотрев,

результаты продаж в 2021 году в городе «А», руководство может изъявить желание ознакомиться, как выглядит ситуация на уровне региона.

2) Операция детализации – переход к более детализированным данным. Например, изучив информацию на уровне региона, пользователь может получить более точную информацию о работе конкретного отделения.

3) Формирование среза – пользователь, как правило, не работает одновременно со всем гиперкубом данных. В большинстве случаев, извлекаются двумерные таблицы из многомерного куба с помощью сечения плоскостями. Другими словами, происходит разрезание измерения куба по интересующим его меткам, в результате чего получает двумерный срез.

Подмножество гиперкуба, которое получается при фиксации значения одного или нескольких измерений, называется срезом (рис. 13) [10, 18].

<i>Место</i>	<i>2013 г</i>	<i>2014 г</i>	<i>2015 г</i>
<i>Город 1</i>	20 000	80 000	45 000
<i>Город 2</i>	85 000	37 000	77 000
<i>Город 3</i>	200 000	42 000	100 000
<i>Город 4</i>	113 000	300 000	237 000
<i>Город 5</i>	20 000	16 000	92 000
<i>Город 6</i>	74 000	80 000	10 000

Рисунок 13 – Двухмерная таблица продаж [10]

Например, если зафиксировать значение какого-либо товара измерения «Список товаров», можно получить подмножество гиперкуба, которое содержит данные об истории продаж этого вида товара в разных местах продаж и в различные годы.

Кроме того, куб может включать в себя значения нескольких фактов. Соответственно и в срезе такого куба отображаются значения нескольких фактов (рис. 14) [10, 18].

<i>Место</i>	<i>2013 г</i>		<i>2014 г</i>		<i>2015 г</i>	
	<i>Сумма</i>	<i>Количество</i>	<i>Сумма</i>	<i>Количество</i>	<i>Сумма</i>	<i>Количество</i>
<i>Регион 1</i>	105 000	5 000	117 000	5 500	122 000	6 000
<i>Регион 2</i>	313 000	13 000	342 000	16 000	337 000	16 500
<i>Регион 3</i>	94 000	4 000	96 000	4 200	102 000	4 600

Рисунок 14 – Двухмерная таблица с двумя фактами [10]

4) Операция вращения применяется для изменения порядка представления измерений, т.е. визуализации. Как уже отмечалось ранее, операция предоставляет возможность визуализации данных в форме, наиболее комфортной для их восприятия пользователями (рис. 15) [10, 18].

<i>Год</i>	<i>Регион 1</i>	<i>Регион 2</i>	<i>Регион 3</i>
2013	105 000	313 000	94 000
2014	117 000	342 000	96 000
2015	122 000	337 000	102 000

Рисунок 15 – Транспонированная двухмерная таблица [10]

Например, если пользователь изначально вывел отчет, в котором данные о продажах по годам были перечислены в столбцах, а по регионам – в строках, он может поменять местами столбцы, строки и фильтры, решив, что такое представление будет наиболее наглядным.

Сложность, с которой можно столкнуться при применении OLAP технологий – создание запросов, выборе базовых данных и разработке схемы, в результате чего большинство современных продуктов OLAP поставляются вместе с огромным количеством предварительно настроенных запросов. Еще одна проблема состоит противоречивости в базовых данных. Для применения технологии, они должны быть не только непротиворечивыми, но и полными.

На данный момент, все больше поставщиков программных средств утверждают, что их продукция относится к категории OLAP, однако на самом деле это может означать что угодно.

Критерии OLAP были сформулированы в 1993 году Эдгаром Коддом. С течением времени они были переработаны и объединены в так называемый тест «FASMI» (Fast Analysis of Shared Multidimensional Information), который определяет основные требования к OLAP продуктам [7, 18]:

1) «Fast» (быстрый) – быстрое время доступа к аналитическим данным. Простые запросы обрабатываются в течение 1-ой секунды, а более сложные – немногим более 20-ти секунд. Подобную скорость достаточно сложно достигнуть при больших объемах данных;

2) «Analysis» (анализ) – возможность логического и статистического анализа, а также сохранения результатов в доступном для конечного пользователя виде. Важно обеспечить пользователей всеми необходимыми функциональными возможностями для проведения анализа;

3) «Shared» (разделяемый доступ) – доступ к данным должен осуществляться по многопользовательскому принципу с поддержкой соответствующих механизмов блокировок и средств авторизованного доступа;

4) «Multidimensional» (многомерность) – многомерное представление данных, включая полную поддержку для иерархий;

5) «Information» (информация) – возможность доступа к информации независимо от ее объема и места хранения. При этом, мощность OLAP-продуктов определяется в том, сколько входных данных они могут обрабатывать, а не какое количество гигабайт они могут хранить.

Говоря о сетевых OLAP-системах, можно сказать, что они построены на основе архитектуры «клиент-сервер». При данной архитектуре, сервер обеспечивает извлечение данных, необходимых для создания кубов из хранилища данных, для удобного и эффективного просмотра. Данные могут храниться либо в исходных реляционных таблицах, либо в виде многомерных структур. Существует несколько основных способов хранения [20, 32]:

1) MOLAP (Multidimensional OLAP) – исходные и агрегатные данные хранятся в виде многомерных структур. В данном случае, на этапе загрузки данных из источников выполняется расчёт агрегатов. Например, по нескольким исходным значениям вычисляется «Общий итог за период». Вычисленные данные сразу же сохраняются в специальном многомерном виде, обеспечивающем их быстрое извлечение в будущем, что позволяет манипулировать данными;

2) ROLAP (Relational OLAP) – исходные данные хранятся в реляционных таблицах, а агрегаты могут не существовать вообще или создаваться по первому запросу к системе;

3) HOLAP (Hybrid OLAP) отличается от MOLAP тем, что исходные данные остаются в тех же реляционных таблицах, в которых они изначально находились, а вычисленные агрегатные данные хранятся в виде многомерных структур.

Хранение данных по принципу MOLAP позволяет манипулировать данными, как многомерным массивом. Благодаря чему, скорость вычисления агрегатных значений одинакова для любого из измерений. Однако в подобном случае, базы данных становятся избыточными, т.к. полностью содержат исходные реляционные данные.

Преимущество ROLAP заключается в упрощении процедуры загрузки и интеграции данных в хранилище, т.к. реляционные СУБД лежат в основе построения многих систем оперативной обработки, которые обычно являются главными источниками данных для хранилища [20, 25].

Гибридная модель HOLAP является наиболее подходящей, когда имеется большой объем исходных данных и высокая степень детализации, а для анализа используются в основном обобщенные данные.

Так, в супермаркете, который ежедневно обслуживает большое количество покупателей, используется OLTP-система. При этом, в реляционной системе отображаются данные о каждой конкретной покупке. С точки зрения анализа, наиболее значимыми являются обобщенные данные (например, по группам товаров, неделям и т.п.). Именно поэтому, исходные детализированные данные подвергаются агрегированию и вычисленные агрегаты сохраняются в многомерной структуре хранилища (рис. 16) [18].

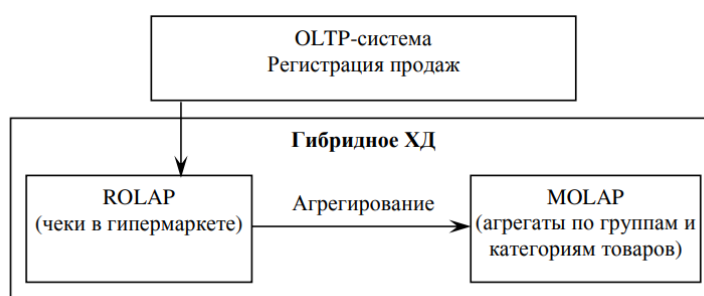


Рисунок 16 – Поток данных в системе [18]

Таким образом, на основании вышеизложенного материала, можно сказать, что OLAP представляет собой уникальную систему многомерного представления данных в режиме реального времени, которая подходит для всех видов бизнеса. Кроме того, OLAP предоставляет строительные блоки для инструментов бизнес-моделирования, инструментов интеллектуального анализа данных, инструментов отчетности о производительности, а также, позволяет пользователям делать данные куба срезов по различным измерениям и фильтрам, что является довольно удобным для анализа временных рядов. Технология является мощной системой онлайн-аналитики для визуализации, которая способна обеспечить быстрое время отклика.

1.5 Сравнительный анализ существующих систем планирования и аналитики

На сегодняшний день происходит активное развитие технологий анализа данных. Современные средства аналитики данных позволяют выявлять скрытые модели благодаря возможностям прогнозирования, самообучения.

Эти технологии интуитивны и включают возможность визуализации данных, которая помогает мгновенно анализировать миллионы строк и столбцов. Аналитика данных распространяется на все аспекты нашей жизни, какими бы вопросами мы не задавались. Например, о сотрудниках, финансах или о том, что нравится или не нравится потребителям, а также, что влияет на их поведение. Аналитика дает ответы на все упомянутые вопросы и помогает принимать обоснованные решения.

В настоящее время подавляющее большинство компаний хранит и анализирует информацию в MS Excel. Однако время не стоит на месте, а объем наработанных данных и сведений растет с невероятной скоростью, в результате чего анализ полученной информации (отчетности и т.п.) становится все более затруднительным из-за массива данных. Именно поэтому, многие

крупные компании стремятся перейти на более оптимальные системы планирования и аналитики, которые позволят выполнять обработку и анализ больших объемов данных, необходимых для эффективного принятия решений, а также осуществлять планирование и мониторинг текущей операционной деятельности компании.

Ранее выбор систем планирования и аналитики был ограничен только двумя вариантами (рис. 17) [2, 36].



Рисунок 17 – Устоявшиеся варианты систем планирования [2]

Выбирая между менее затратными и более затратными платформами, большинство компаний опираются на индивидуальные возможности и цели предприятия. Однако при обоих вариантах, разумеется, предстоит столкнуться как с преимуществами, так и с недостатками платформ.

Так, преимуществами MS Excel, Access и Power BI состоят в максимальной гибкости изменений и экономии затрат на внедрение. Недостатками являются: отсутствие единого источника данных, вероятность потери данных, отсутствие разграничения прав пользователей, а также работа с версиями, которая полностью осуществляется в ручном режиме [2, 17, 38].

Преимущества Oracle, SAP, IBM, SAS и 1C состоят в широком стандартном функционале для работы с большими данными, наличие любых

интерфейсов и коннектора для сбора и передачи данных. Недостатками являются: низкий уровень Self-Service и потребность в поддержке от ИТ-команды, долгий срок внедрения, а также большая стоимость внедрения и поддержки [23, 34, 46].

Таким образом, если изучить все стороны указанных альтернатив, становится понятно, что анализировать большой объем данных не только затруднительно, но также занимает достаточно большое количество времени, при этом не исключая риска потери данных.

Кроме того, помимо вышеупомянутых систем, также существуют такие альтернативные платформы, как Anaplan и Optimacros, в основе которых лежит моделирование по принципу многомерного представления данных (OLAP) особенности которого были рассмотрены в предыдущем параграфе. Данные системы совмещают в себе все основные функции альтернативных систем аналитики, но также практически полностью доведена до автоматизма.

Преимущества Anaplan и Optimacros состоят в гибкости и низкой стоимости развертывания, версионности и многомерности, а также высоком уровне Self-Service у ПО. Недостатками являются дополнительные затраты на ПО и консалтинг при развертывании, а также ограниченный функционал управления регламентом процесса.

Кроме того, основными преимуществами данных систем аналитики также являются [2, 33]:

- 1) целостная модель позволяет исключить риск потери данных;
- 2) совместная одновременная работа пользователей через Интернет, без необходимости установки и обновления приложений и плагинов;
- 3) контроль доступов и действий пользователей;
- 4) разграничение прав доступа на уровне справочников в разрезе пользователей;
- 5) аналитическое хранение и моделирование (OLAP) на основе многомерных кубов данных (аналог сводных таблиц);

б) возможность масштабирования (горизонтальное и вертикальное масштабирование) моделей и мастер-данных;

7) справочники, иерархии и форматы данных позволяют моделировать на основе многомерных кубов данных и строить системы корпоративного уровня;

8) консолидационное планирование (Bottom-up Planning) и каскадирование целей (Top-down Planning);

9) создание интегрированной модели за счет вовлечения в процесс планирования сотрудников разного уровня (от фронт-лайн до топ-менеджмента) из разных отделов (финансы, продажи, маркетинг, производство, HR и IT);

10) возможность автоматической интеграции с внешними источниками данных;

11) встроенный OLTP модуль (DWH/КХД) позволяет хранить любые объемы данных в SQL/noSQL форматах и строить витрины данных для BI и EPM.

Так, Anaplan и Optimacros создают продукт для аналитиков в бизнесе, которые смогут самостоятельно строить системы планирования и отслеживания эффективности. Продукт содержит в себе гибкость MS Excel и надежность оперативных систем. Единственное, что необходимо – хорошее знание системы, чтобы перенести всю информацию из баз данных. После чего, можно не беспокоиться о «растягивании данных», т.к. система автоматически применит условие за выбранный пользователем период анализа.

Таким образом, основным отличием от традиционных систем планирования является принцип Self-Service конструктора, для использования которого требуются только аналитики, а также проектирование модели при помощи OLAP технологий. Для работы необходим только браузер и доступ в корпоративный домен или Интернет.

Так, платформы Anaplan и Optimacros имеют меньшую стоимость внедрения и владения при большем уровне гибкости и Self-Service, а также

внушительный перечень преимуществ по сравнению с альтернативными системами аналитики (рис. 18) [2, 24, 44].

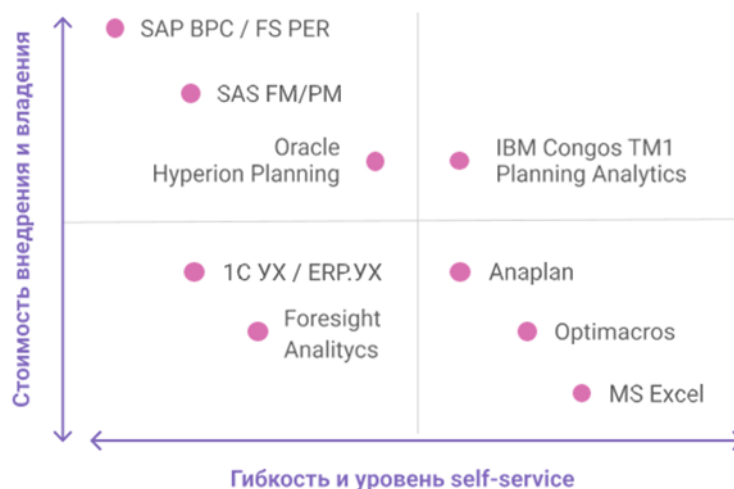


Рисунок 18 – Сравнение систем планирования и аналитики [2]

Таким образом, аналитические инструменты обеспечивают широкое преобразование бизнес-выводов благодаря автоматическому обновлению и автоматизации процессов обнаружения, очистки и публикации данных. Кроме того, подавляющее большинство компаний не только хранят, но и анализируют информацию в нескольких системах, в результате чего, становится затруднительным поиск и обнаружение необходимой информации.

Так, для оптимизации рабочего процесса и роста эффективности принимаемых решений, наилучшим решением является – проектирование хранилища данных для одной системы планирования и аналитики, которая будет являться единой для всех бизнес-процессов, выполняемых организацией. Не нужно искать отдельный файл с необходимыми сведениями, все сфокусировано в одном месте, а в случае добавления новых данных – система обновляется. При работе с большим массивом данных подобная функция является решающей.

2 Сбор и анализ требований для проектирования хранилищ данных

2.1 Краткая характеристика деятельности предприятия и анализ дебиторской задолженности

В качестве объекта исследования данной работы было выбрано дочернее предприятие ПАО «Газпром», основным видом деятельности которого является производство пара и горячей воды (тепловой энергии). Организационно правовая форма – публичное акционерное общество.

К дополнительным видам деятельности компании относятся [13]:

- 1) производство электроэнергии тепловыми электростанциями;
- 2) производство электроэнергии прочими электростанциями и промышленными блок-станциями;
- 3) передача электроэнергии и ее распределение;
- 4) деятельность по обеспечению работоспособности тепловых электростанций и электрических сетей;
- 5) производство и распределение пара и горячей воды (тепловой энергии) тепловыми электростанциями;
- 6) подготовка строительного участка;
- 7) производство общестроительных работ по строительству тепловых и прочих электростанций;
- 8) монтаж зданий и сооружений из сборных конструкций и т.п.

Компания обеспечивает централизованное отопление и горячее водоснабжение собственных источников теплоснабжения, а также других объектов тепловой генерации, за исключением небольших локальных районов теплоснабжения от изолированных ведомственных и корпоративных тепловых источников.

Компания осуществляет свою деятельность с 2004 года и оказывает услуги как юридическим, так и физическим лицам.

Предприятие является оператором самой протяженной теплоэнергетической системы в мире. В эксплуатации компании находится более 17 000 км тепловых сетей, в том числе около 8 600 км магистральных и 8 433 км разводящих. Также компания эксплуатирует свыше 10 тыс. тепловых пунктов.

В настоящее время компания эксплуатирует и обслуживает следующие теплоэнергетические объекты [13]:

- 141 источник генерации тепловой энергии, в том числе 45 районных и квартальных тепловых станций (РТС и КТС), 96 малых котельных (МК, АИТ, ПК) общей тепловой мощностью 4996,014 Гкал/ч;

- 17 034 км тепловых сетей (в однострубно́м исчислении), в том числе 8 600,06 км магистральных тепловых сетей (включая тепловые вводы) и 8 433,94 км разводящих тепловых сетей;

- 10 611 тепловых пунктов;

- 24 насосные станции.

На сегодняшний день в компании работает свыше 14 тыс. сотрудников. Высокий профессионализм и эффективная работа каждого сотрудника являются гарантом успеха и достижения поставленных целей.

Предприятие имеет большое количество потребителей, которые выступают, как в лице физических, так и юридических лиц. Потребитель заключает сделку с компанией производителем тепловой энергии, в ходе которой, компания исполняет свои обязательства, а ее контрагент оплачивает стоимость через какое-то время. Другими словами, происходит отсрочка платежа. Подобный временной разрыв приводит к образованию дебиторской задолженности потребителя перед исследуемой организацией.

В связи с тем, что подобный долг может быть закрыт, как через пару дней, так и продолжить существовать длительный период времени, грамотное управление дебиторской задолженностью является одним из ключевых моментов организации эффективного функционирования предприятия.

Кроме того, т.к. речь идет о компании, осуществляющей производство и распределение воды и тепловой энергии, к данному вопросу стоит относиться с большой серьезностью. Подобное обусловлено тем, что величина дебиторской задолженности энергетической компании с внушительным перечнем потребителей и деловых партнеров, административных округов и районов распределения, несомненно, имеет огромные размеры.

В зависимости от взаимоотношений с деловыми партнерами на предприятии, дебиторскую задолженность можно классифицировать следующим образом [13, 21]:

1) текущая ДЗ – представляет собой краткосрочную задолженность, при которой срок выплаты долга не превышает 12 месяцев;

2) долгосрочная ДЗ – представляет собой, как можно понять из названия, долгосрочную задолженность, при которой выплаты долга ожидается не ранее, чем через 12 месяцев;

3) просроченная ДЗ – представляет собой задолженность, при которой срок выплат по договору уже прошел, но денежные средства не были перечислены клиентом;

4) сомнительная ДЗ – представляет собой задолженность, при которой существует риск не взыскания, если долг не обеспечен залогом или какой-либо другой гарантией;

5) безнадежная ДЗ – представляет собой задолженность, при которой возврат денежных средств не представляется возможным по причине банкротства, ликвидации, реструктуризации организации заемщика или истечения срока исковой давности.

Таким образом, необходимо провести анализ дебиторской задолженности исследуемой организации для того, чтобы определить ситуацию и оценить существующий механизм регулирования.

Информация о дебиторской задолженности дочернего предприятия ПАО «Газпром» за период 2019 – 2021 гг. представлена в табл. 1. Единицы измерения – тыс. руб. [1, 13].

Таблица 1 – Дебиторская задолженность ПАО «Газпром» за 2019 – 2021 гг.

Наименование показателя	На 31.12.2021г.	На 31.12.2020г.	На 31.12.2019г.
Дебиторская задолженность, в т.ч.	57 267 323	45 038 343	44 996 990
Дебиторская задолженность (платежи по которой ожидаются более чем через 12 месяцев после отчетной даты), в т.ч.	9 069 570	9 053 781	6 758 206
покупатели и заказчики	957 796	245 934	31 340
авансы выданные	15 840	20 058	29 260
прочие дебиторы	8 095 934	8 787 789	6 697 606
Дебиторская задолженность (платежи по которой ожидаются в течение 12 месяцев после отчетной даты), в т.ч.	48 197 753	35 984 562	38 238 784
покупатели и заказчики	34 503 804	23 023 961	23 875 932
зadolженность участников (учредителей) по взносам в уставный капитал	–	–	–
авансы выданные	748 965	702 288	1 057 435
прочие дебиторы	12 944 984	12 258 313	13 305 417

Таким образом, по сравнению с 2019 годом размер дебиторской задолженности компании увеличился на 21%. Причем, большая часть приходится на покупателей и заказчиков (рост на 97% – долгосрочных, на 30% – краткосрочных). Кроме того, по некоторым показателям, дебиторская задолженность за анализируемый период снизилась. Например, в 2021 году платежи, которые ожидаются в течение 12 месяцев после отчетной даты по прочим дебиторам, меньше почти на 3%.

Подобное может быть обусловлено тем, что по сравнению с отчетным периодом перечень потребителей и деловых партнеров претерпел какие-либо изменения.

Каждый месяц в компанию приходят n-е количество новых потребителей и закрываются старые. Кроме того, доля какого-либо из перечисленных ранее видов задолженности также могла существенно измениться.

Так, распределение по видам дебиторской задолженности в общем объеме ДЗ компании за 2021 год представлен на рис. 19 [13].

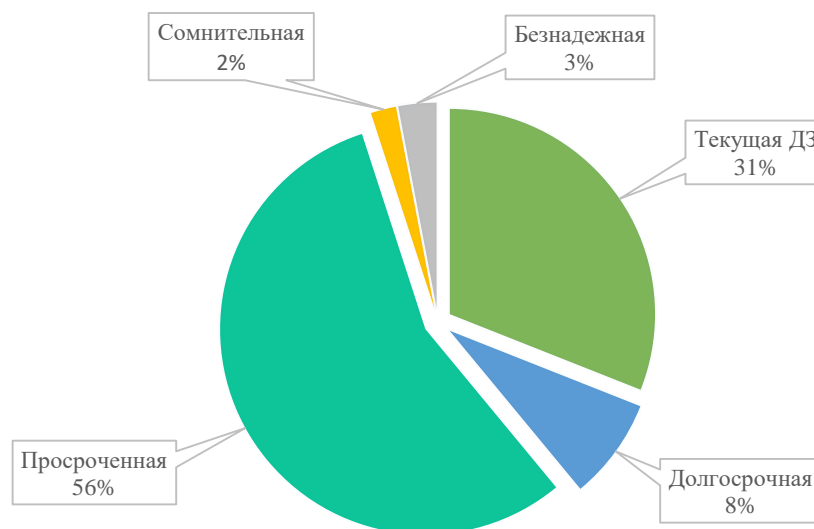


Рисунок 19 – Структура ДЗ ПАО «Газпром» за 2021г.

Таким образом, на рис. 19 можно увидеть, что большая часть дебиторской задолженности компании является просроченной, т.е. денежные средства не были перечислены в срок, оговоренный по договору. Стоит отметить, что при нормальном функционировании, большая часть задолженности относится к текущим краткосрочным долгам, что не оказывают негативного влияния на предприятие и не тормозят его развитие. Подобный итог говорит о необходимости наличия грамотного механизма управления дебиторской задолженностью на предприятии, который позволит обеспечить нормальную платежеспособность. Для этого, необходимо изучить перечень наиболее популярных мер, осуществляемых для эффективного управления ДЗ.

Как правило, большинство компаний для управления дебиторской задолженностью используют комплекс мер, в состав которых входят как предупредительные, так и взыскательные юридические мероприятия.

Существует несколько основных этапов, а именно [21, 26]:

- 1) предварительная проверка партнера до стадии заключения договора;

2) особые условия для каждого конкретного партнера (график, сроки, штрафы и т.п.);

3) оперативный контроль за исполнением обязательств;

4) взыскание (должник не собирается добровольно погашать долг).

Первые три этапа помогают предотвратить возникновение просроченных долгов. Говоря об управлении уже имеющейся дебиторской задолженностью, в исследуемой организации применяются несколько основных мероприятий по отношению к потребителям, а именно [13]:

1) обзвон потребителей;

2) встреча с потребителем в отделение сбыта;

3) информационное сообщение о задолженности;

4) направление обращения Управы / Префектуры / ДЖКХиБ;

5) направление обращения в вышестоящий отраслевой орган;

6) распространение информации через СМИ;

7) направление в МЖИ (заявление);

8) направление в МЖИ (обращение);

9) ПДК;

10) направление в МТУ Ростехнадзора (обращение);

11) направление в Прокуратуру (обращение);

12) служба корпоративной защиты

13) направление в адрес директора предприятия.

Каждое из указанных мероприятий применяется в отношении к потребителю последовательно в зависимости от рейтинга платежной дисциплины. Данный рейтинг рассчитывается исходя из периода образования долга и имеет вариацию: ДЗ свыше 3 лет, 0, 2, 3 и 4.

Причем, некоторые из перечисленных мероприятий применяются ко всем потребителям, без исключений, а некоторые только к контрагентам с наиболее высоким рейтингом (3 и 4). В подавляющем большинстве случаев, рейтинг равный четырем говорит о том, что задолженность является просроченной или безнадежной, а потребители – наиболее проблемные.

Схема последовательности применения указанных мероприятий в зависимости от ситуации представлена на рис. 20.

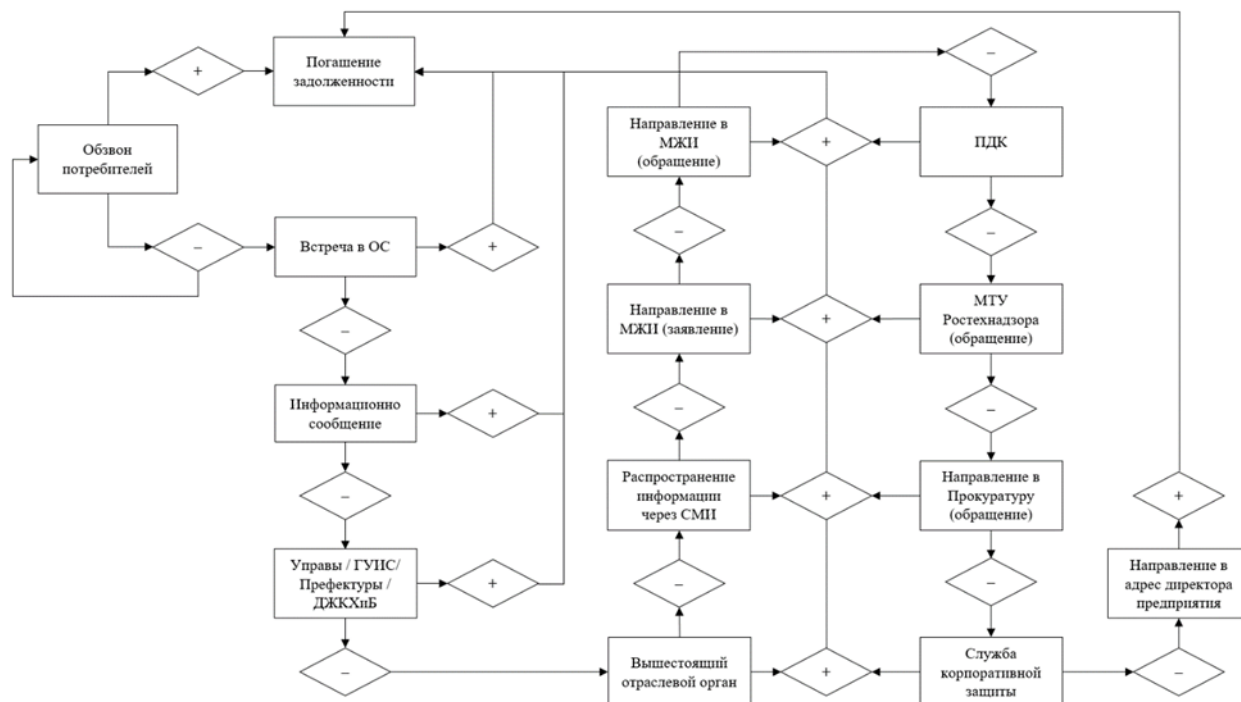


Рисунок 20 – Процесс управления ДЗ ПАО «Газпром»

Данная схема отражает последовательность выполнения каждого мероприятия в зависимости от положительного и отрицательного результата по предыдущему.

Таким образом, если срок выплаты задолженности приближается или имеется иная причина связи с клиентом – сотрудники компании осуществляют обзвон потребителей. В рамках данного мероприятия не имеет значения, какой рейтинг платежной дисциплины у обзваниваемых должников. Уведомление о наличии задолженности направляется всем клиентам компании. Далее, в случае невыполнения обязательств, происходит встреча с недобросовестными должниками в отделение сбыта, где происходит разговор с сотрудниками, отвечающими за управление дебиторской задолженности в компании по конкретному отделению и округу.

Если после первых двух мероприятий взаимодействия с потребителем задолженность не была погашена – сотрудником направляется

информационное сообщение о задолженности с целью напомнить клиенту о существовании долга. Причем, по каждому мероприятию необходимо фиксировать дату и время взаимодействия, а также оставлять небольшие комментарии, если того требует ситуация.

Первые три мероприятия являются довольно стандартными, не несущими большой «угрозы» для потребителей и являются исключительно уведомительными мерами. В том случае, если классические уведомления не возымели эффекта, в ход вступают направления обращений в Управы / ГУИС/ Префектуры / ДЖКХиБ, вышестоящий отраслевой орган, МЖИ, ПДК, МТУ Ростехнадзора, прокуратуры и т.д. В крайних случаях происходит распространение информации о задолженности потребителя через средства массовой информации.

Если ни одно из существующих на предприятии мероприятий не принесло ожидаемого результата (погашение долга), происходит направление в адрес директора предприятия. Как правило, до данной меры не доходит. Клиент выполняет свои долговые обязательства на моменте обращения в Управы / ГУИС/ Префектуры / ДЖКХиБ. Однако подобный исход характерен для потребителей с рейтингом платежной дисциплины не более двух. Для потребителей с более высокими рейтингами, не изменяющимися в лучшую сторону за период анализа, в ряде случаев приходится проходить полный список мер, вплоть до взыскания и списания.

В данном процессе задействованы не только финансисты и бухгалтеры предприятия. Для управления и автоматизации просроченной задолженности привлекаются менеджеры, юристы, а также сотрудники отдела безопасности.

Таким образом, можно сделать вывод, что в компании присутствует достаточно эффективных механизмов регулирования дебиторской задолженностью потребителей, который предполагает последовательное применение мероприятий-мер по уведомлению о наличии долга. Кроме того, сотрудники компании выделяют несколько возможных причин наличия у потребителей дебиторской задолженности, а именно:

- отсутствует полный пакет документиков для передачи в суд;
- не согласны с начислениями;
- взыскание задолженности в судебном порядке нецелесообразно (сумма долга менее 20 тыс. руб.)
- ДЗ возникла в результате уточнения / перераспределения льгот;
- планируется перерасчет;
- подготовлены и направлены документы на КПД, для предоставления рассрочки платежей.

Также, стоит отметить, что причины большого количества просроченной задолженности могут быть связаны не только с внутренней структурой управления компании, но также и с рядом внешних факторов. Так, внешними для исследуемой компании причинами могут служить [5, 21]:

1) финансовые трудности – потеря платежеспособности контрагентом, явившаяся следствием «кассового разрыва», связанного с кризисом. В настоящее время, достаточно большое количество потребителей находятся в шатком положении;

2) форс-мажор, стихийные бедствия и другие непрогнозируемые разрушительные явления, смерть или потеря дееспособности контрагента;

3) недобросовестность партнера – некоторые контрагенты заведомо не планируют выполнять условия договора. В данной сфере подобное происходит крайне редко, однако не стоит полностью исключать такую вероятность. При этом, избежать возникновения задолженности просто – качественно проверять потенциальных контрагентов до заключения договора.

Так, компания имеет внушительный перечень деловых партнеров, который не так просто контролировать. Для того, чтобы обеспечить своевременное выявление обязательств, необходимо акцентировать внимание не только на проработке комплекса мер по отношению к взысканию долга, но также структурировать и визуализировать имеющиеся данные в виде, наиболее эффективном для анализа.

2.2 Анализ информационной инфраструктуры предприятия для создания отчетов по дебиторской задолженности

В предыдущем параграфе было выявлено, что дочернее предприятие ПАО «Газпром» имеет действенный механизм управления дебиторской задолженностью. Далее, для того, чтобы определить наиболее эффективные пути для снижения величины просроченной задолженности в общем объеме задолженности контрагентов перед компанией, необходимо изучить основные инфраструктуры предприятия для формирования отчетности.

Компания ПАО «Газпром» долгие годы имела в своем распоряжении несколько основных систем планирования и аналитики, а именно: MS Excel, SAP Logon и Power BI. Процентное соотношение применения каждой из перечисленных площадок представлено на рис. 21 [13].

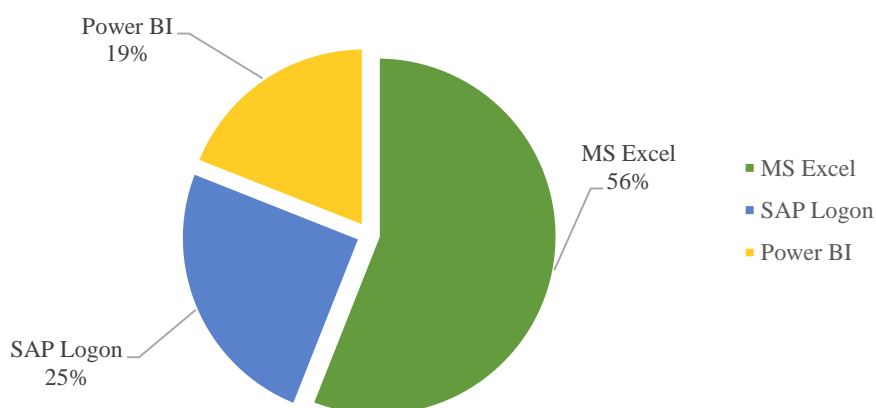


Рисунок 21 – Системы планирования и аналитики ПАО «Газпром»

Основные преимущества и недостатки указанных систем планирования и аналитики были рассмотрены в первой главе данной работы. Далее, необходимо описать основное применение каждой из них для исследуемой организации. Говоря о MS Excel, стоит отметить, что около 56% отчетов и форм предприятия формируются и анализируются именно в этой системе планирования и аналитики. Программа зарекомендовала себя, как один из наиболее гибких инструментов для моделирования отчетов на любом

предприятию. Однако, при большом объеме данных, высока вероятность того, что тяжелые файлы с большим количеством колонок, строк и листов будут загружаться достаточно длительный период времени. Кроме того, работа в MS Excel предполагает (в большей степени) ручной труд, что несомненно является не эффективным с точки зрения автоматизации и экономии времени.

Говоря о Power BI, стоит отметить, что данная площадка используется для аналитики отчетов и форм связанных с движением денежных средств, например, бюджет движения денежных средств, бюджет доходов и расходов и т.п. Площадка позволяет визуализировать и выделять наиболее важную информацию, а также делиться ей со всеми отдельными сбыта. Если говорить об управлении дебиторской задолженностью, то Power BI не используется для формирования отчета. В данном процессе принимают участие две из перечисленных ранее систем аналитики, а именно: MS Excel и SAP Logon.

Говоря о площадке SAP Logon, стоит отметить, что в связи со сложившейся ситуацией в мире, компания SAP приняла решение приостановила работу в Российской Федерации на неопределенный срок. Так как компания не стала прогнозировать будущее, еще неизвестно, уйдет ли SAP из России окончательно, или же через некоторое время возобновит свою работу. В результате чего, дочернее предприятие ПАО «Газпром» приняло решение отказаться от данной площадки из-за неопределенности, которая может препятствовать эффективному планированию и аналитики текущей деятельности.

Формирования отчёта по управлению дебиторской задолженностью происходит на основе ежедневного отчета по авансам потребителей. Алгоритм получения необходимой информации для отчета и его составление имеет следующий вид:

- 1) выгрузка ежедневного отчета по авансам потребителей в формате MS Excel из системы SAP;

- 2) перенос данных из отчета по авансам в шаблон файла MS Excel по управлению дебиторской задолженностью посредством копирования данных.

Если учесть тот факт, что компания отказывается от площадки – SAP, данный алгоритм претерпит существенные изменения.

Фактически, большая часть отчетов компании формируются и анализируются в MS Excel. По каждому отчету существуют отдельные файлы шаблоны, содержащие стандартный набор колонок и вычислений. Для того, чтобы проанализировать ситуацию за новый период, необходимо вставить данные в файл за прошлый период и, в случае необходимости, протянуть данные колонок с формулами для всех значений строк или же наоборот – удалить ненужные для периода анализа строки.

В настоящее время, копирование, вставка и протягивание данных давно отошли в сторону. На смену ручному труду сотрудников давно пришли системы планирования и аналитики, которые предполагают автоматизацию обновления данных и позволяют хранить в себе всю отчетность и аналитику компании, а не только какую-то ее часть.

Таким образом, руководством компании было принято решение внедрить на предприятие новую площадку для аналитики, которая позволит удовлетворить основным требованиям пользователей, а также даст возможность автоматизировать процесс обновления данных во всех формах отчетов компании.

2.3 Сбор и анализ требований пользователей

Прежде чем осуществлять переход на новую систему планирования, необходимо определить основные требования пользователей отчета по управлению дебиторской задолженностью компании. В ходе сбора информации посредством личного разговора с сотрудниками отделений сбыта, ответственными за процесс управления дебиторской задолженностью, было выявлено, что 91% рабочего персонала поддерживает решение

относительно автоматизации процесса формирования отчетов и форм компании, при помощи перехода на новую платформу.

Таким образом, было принято решение провести опрос, на основании которого можно было бы определить, какие моменты в действующей форме отчета устраивают сотрудников отделений, какие не устраивают, а также, что можно было бы добавить / изменить, чтобы повысить эффективность. Так, на рис. 22 представлены результаты опроса 45 сотрудников относительно положительных моментов действующей формы отчета и процесса ее формирования.

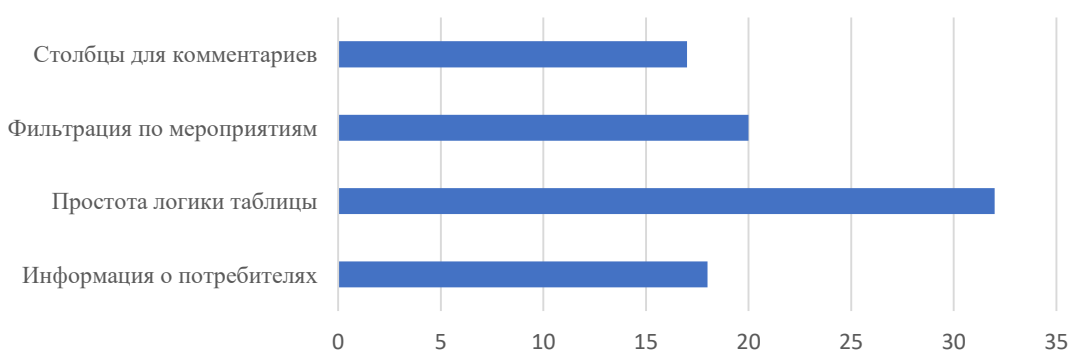


Рисунок 22 – Положительные моменты действующей формы отчета

Так, на основании рис. 22 можно сделать вывод, что важными моментами при работе с отчетом по управлению дебиторской задолженностью являются:

- наличие базовой информации о потребителях (наименование делового партнера, номер договора, название административного округа, ИНН, группа потребления, альтернативная группа и т.д.);
- возможность фильтрации по колонкам мероприятий (выбор значения, необходимого для рассмотрения и анализа);
- наличие столбцов, в которых можно оставлять комментарии по соответствующему потребителю и мероприятию. Данная информация помогает разобраться в ситуации, если ранее с потребителем работал другой сотрудник компании и отследить, как закончилось последнее взаимодействие;

– простота логики таблицы. Данный момент является одним из ключевых и наиболее значимых для пользователей отчета, т.к. все столбцы располагаются последовательно, в зависимости от возрастания серьезности мер, предъявляемых к потребителям, а также имеют простые и понятные наименования, не вызывающие трудностей при идентификации.

Далее, необходимо определить, какие моменты в работе с отчетом вызывают у сотрудников компании различного рода затруднения. Результаты опроса по данной теме представлены на рис. 23.

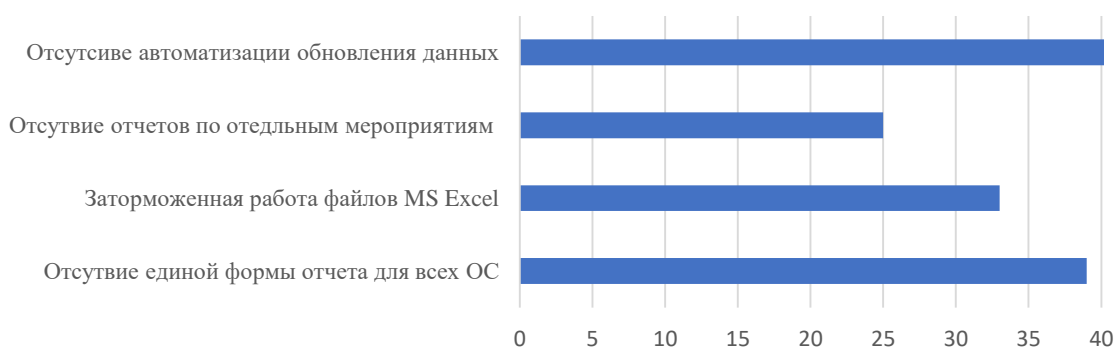


Рисунок 23 – Отрицательные моменты действующей формы отчета

По данным рис. 23 можно увидеть, что, как уже отмечалось ранее, большинство опрошенных отрицательно относятся к ручному обновлению данных и определяют данную особенность к слабой стороне формирования отчета по управлению дебиторской задолженностью. Кроме того, также было установлено, что работа в файлах MS Excel сопровождается заторможенностью системы, т.к. отчет содержит большое количество строк и столбцов, а также некоторые вспомогательные листы для расчётов, что сказывается на эффективности работы.

Также, по результатам опросов было выявлено, что на предприятии присутствует проблема, связанная с тем, что каждое отдельное отделение сбыта осуществляет формирование отчета по управлению дебиторской задолженностью в разных форматах. Так, у одного отделения могут быть добавлены какие-либо вспомогательные колонки, а у другого нет. В

результате чего, анализ отчета, содержащего данные по всем отделениям, становится затруднительным из-за противоречия форматов представления.

Помимо вышеизложенных трудностей, стоит отметить, что 63% опрошенных сотрудников отметило, что из-за большого количества колонок в таблице, становится затруднительным возвращаться к первым столбцам, перекачивая бегунок при анализе мероприятий в конце отчета. В результате чего, еще одной слабой стороной действующей формы является – отсутствие возможности анализа отдельных мероприятий.

После определения положительных и отрицательных моментов отчета, сотрудники компании предложили собственные варианты улучшения отчета, которые представлены на рис. 24.



Рисунок 24 – Предложения для действующей формы отчета

Так, часть из предложенных улучшений способна нейтрализовать некоторые отрицательные стороны, рассмотренных ранее, а часть представляет собой внедрение возможности проведения дополнительного точечного анализа, как по мероприятиям, так и по различным показателям контрагентов компании.

Таким образом, на основании результатов опроса, были сформулированы требования пользователей к отчету по управлению дебиторской задолженности в будущей системе планирования, а именно:

- 1) возможность аналитики по отдельным мероприятиям;

2) наличие колонок, содержащих всю необходимую вспомогательную информацию о потребителях;

3) наличие дополнительной колонки о рейтинге платежной дисциплины потребителя на отчетную дату;

4) возможность фильтрации по характеристикам: административный округ, отделение сбыта, муниципальный район, группа потребления, альтернативная группа, принадлежность, категория потребителя, рейтинг платежной дисциплины, итог по каждому из мероприятий;

5) возможность формирования списка изменений по: потребителям, группе потребления, альтернативной группе, принадлежности потребителей;

6) возможность формирования сводных таблиц по: рейтингу платежной дисциплины и группе потребления;

7) автоматизация обновления данных отчета;

8) возможность одновременной работы нескольких пользователей.

Таким образом, на основании сформированных требований пользователей, руководством компании было принято решение внедрить на предприятие систему «Optimacros», которая представляет собой платформу для оптимизационного и консолидационного планирования, бизнес-анализа и интеграций. Данная платформа содержит в себе все те преимущества и гибкость, которые имеют применяемые на предприятие системы аналитики. Основные преимущества системы были рассмотрены в первой главе данной работы.

Для того, чтобы внедрить данную платформу, необходимо осуществить несколько шагов проектирования хранилища данных исследуемой организации. Полученный результат позволит обеспечить взаимодействие хранилища данных и OLAP технологий, лежащих в основе системы «Optimacros», а также осуществлять обработку данных, консолидацию и их отображение на различных отчётах с помощью средств визуализации. Другими словами, технология способствует развитию инструментария для создания инфраструктуры принятия решений.

3 Проектирование хранилищ данных для формирования отчетов с целью принятия управленческих решений

3.1 Проектирование хранилищ данных (концептуальная, логическая и физическая модели)

В предыдущей главе было установлено, что для внедрения в дочернее предприятие ПАО «Газпром» системы «Optimacros» необходимо осуществить проектирование хранилища данных.

Проектирование хранилища данных представляет собой создание структуры данных, ориентированной на высокую производительность при выполнении аналитических запросов. Существенным преимуществом подобной структуры является – простота модели хранилища, которая не затрудняет ее понимание и не ухудшает эффективность выполнения запроса.

К проектированию хранилищ данных предъявляются следующие основные требования [10, 19]:

- 1) структура данных хранилища должна быть проста и понятна конечным пользователям;
- 2) статические данные, которые регулярно подвергаются модификации должны быть выделены;
- 3) требований к запросам должны быть просты;
- 4) обеспечение поддержки сложных запросов SQL, которые требуют последовательной обработки тысяч записей.

При проектировании хранилища данных зачастую строится три модели, а именно: концептуальная, логическая и физическая. Рассмотрим более подробно каждую из них.

Концептуальная модель – характеризуется определением отношения между сущностями. Как правило, подобная модель данных представляет собой диаграмму «сущность-связь», целью которой является – определение и

анализ потребностей организации. Концептуальная модель состоит из сущностей, атрибутов и связей.

Кроме того, данная модель не отражает особенностей физической реализации или программных ограничений [14, 32].

В концептуальной модели могут быть сущности двух видов:

1) сильные (независимые) – существование которых не зависит от других сущностей;

2) слабые (зависимые) – существование которых зависит от других сущностей, которые являются более сильными по отношению к ним.

Таким образом, одна и та же сущность может быть сильной по отношению к одной и слабой по отношению к другой.

Атрибутами являются характеристики сущностей и предназначены для идентификации. Значения каждого атрибута выбираются из домена – множества, включающего в себя все значения, которые могут быть присвоены этому атрибуту. Значения атрибутов могут быть как статическими, т.е. не меняющимися с течением времени, так и динамическими (например, регион, категория, объем и т.д.). Пример представлен на рис. 25 [14].

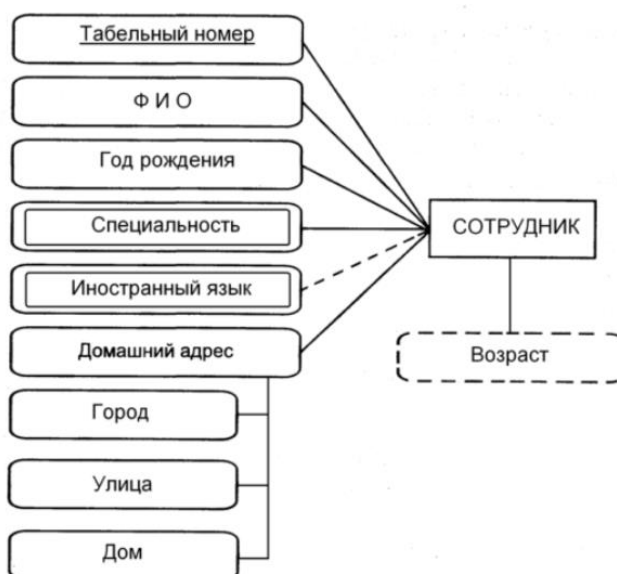


Рисунок 25 – Пример описания сущности атрибутами [14]

Для наглядности проектирования по каждой из указанных ранее моделей, рассмотрим технологию на примере кейса: дочернее предприятие

ПАО «Газпром» занимается производством тепловой энергии на территории Российской Федерации и осуществляет ее распределение в различные регионы страны. Руководством было принято решение внедрить технологию хранилища данных на предприятии для построения модели управления дебиторской задолженностью. Для данного примера можно выделить следующие характерные сущности: продукт (Product), клиент (Customer), дебиторская задолженность (Receivables), сотрудники (Employees), отделение сбыта (Sales Office), регион (Region), изменчивость спроса (Demand Variability), сезонность (Seasonality).

Таким образом, на рис. 26 представлена концептуальная модель сущностей и связей для данного примера.

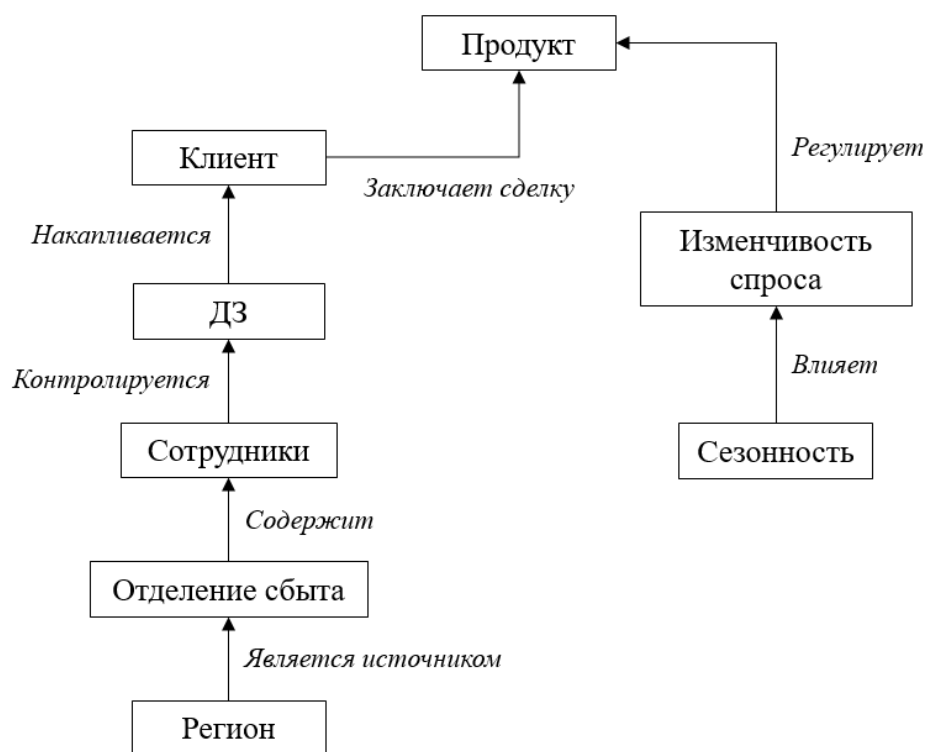


Рисунок 26 – Сущности и связи хранилища данных

Необходимо определить список атрибутов для каждой из сущностей:

1) продукт:

- код продукта;
- категория продукта;
- название продукта.

2) Клиент:

- код клиента;
- номер контокоррентного счета клиента;
- номер договора;
- категория потребителя;
- наименование делового партнера / потребителя;
- адрес потребителя.

3) Дебиторская задолженность:

- код дебиторской задолженности;
- сумма долга;
- период образования долга;
- рейтинг платежной дисциплины.

4) Сотрудники:

- код сотрудника;
- категория сотрудника;
- контактная информация.

5) Отделение сбыта:

- код отделения;
- название отделения.

6) Регион:

- код региона;
- название региона.

7) Изменчивость спроса:

- код изменчивости;
- название;
- ранг.

8) Сезонность:

- код сезона;
- характеристика.

Таким образом, на основании указанных выше атрибутов, можно построить схему ER-модели. Для этого необходимо поместить на схему атрибуты каждой из определенных ранее сущностей (рис. 27).

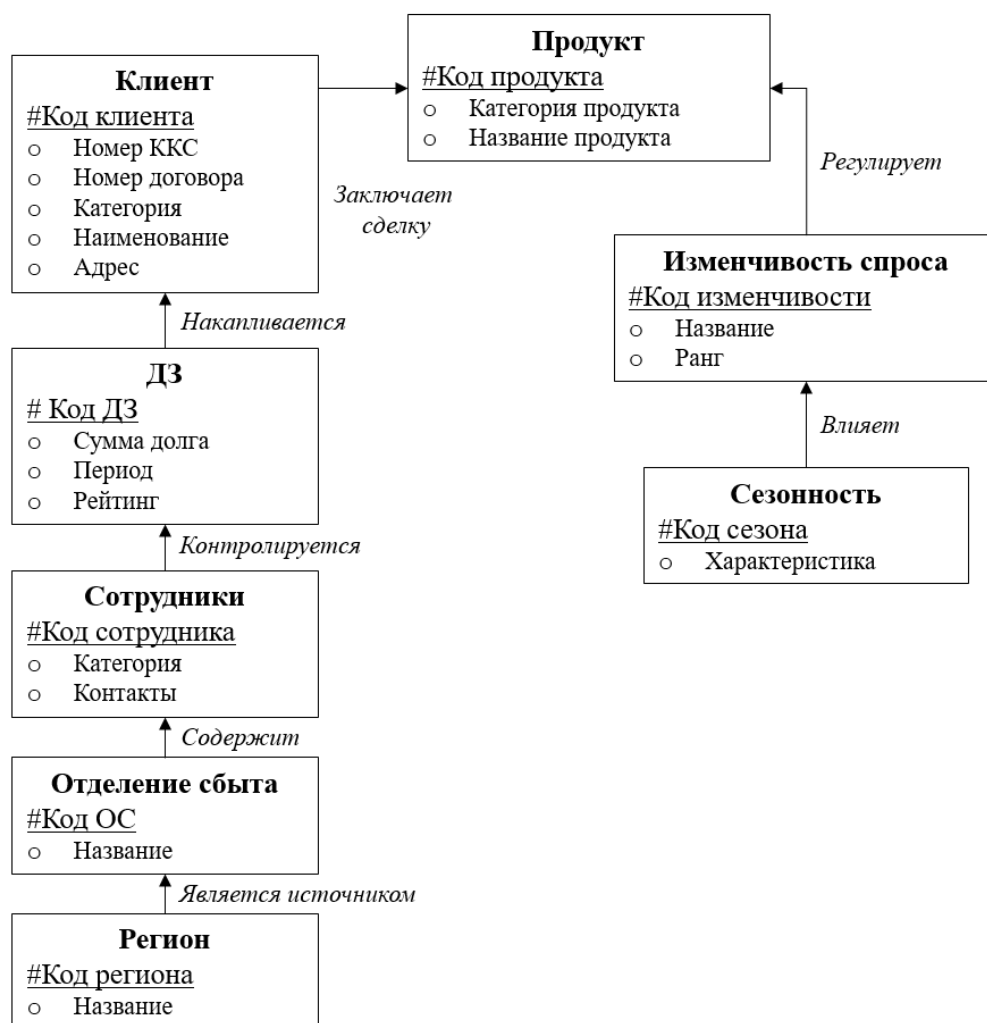


Рисунок 27 – Концептуальная модель хранилища данных

Построенная модель является моделью предметной области высокого уровня и включает в себя сущности, связи и атрибуты. Данная модель не описывает типы данных атрибутов, не учитывает нормализацию данных и не показывает ключи, по которым происходит соединение сущностей. Следующим этапом является проектирование логической модели хранилища данных.

Логическая модель напоминает концептуальную, но при этом отражает начало физического представления данных. В логической модели, в отличие

от концептуальной, в которой связи могут быть: «один ко многим» и «многие ко многим», запрещены связи «многие ко многим», а используются промежуточные сущности. Подобное обусловлено тем, что связи «многие ко многим» не согласуются с идеей реляционного хранения [37].

Логическая модель данных является начальным прототипом будущей базы данных и строится в терминах информационных единиц, но без привязки к конкретной СУБД. Предварительным средством разработки логической модели данных в настоящий момент являются различные варианты инфологических моделей – ER-диаграмма (Entity-Relationship, диаграммы сущность-связь).

Стоит отметить, что одну и ту же ER-модель можно преобразовать как в реляционную модель данных, так и в модель данных для иерархических и сетевых СУБД, или в постреляционную модель данных.

Логическая модель отражает логические типы сущностей, атрибуты данных, описывающие сущности, и связи между сущностями. Данной модели присущи определенные особенности, а именно [37, 45]:

- модель включает все сущности и отношения между ними;
- все атрибуты сущностей определены;
- для каждой сущности определен первичный ключ;
- определены внешние ключи;
- появляется нормализации.

Логическая модель данных генерируется из концептуальной модели при помощи нотация IDEF1X – метод для разработки реляционных баз данных, использующий условный синтаксис. Так, данная модель уже является реляционной и должна учитывать выбранный подход к построению хранилища данных, а именно – нормализованное хранилище данных [37].

Целью построения логической модели является получение графического представления логической структуры исследуемой предметной области. На данном уровне появляется возможность определить ключевые атрибуты,

которые станут первичными ключами. Таким образом, логическая модель, полученная после генерации, представлена на рис 28.

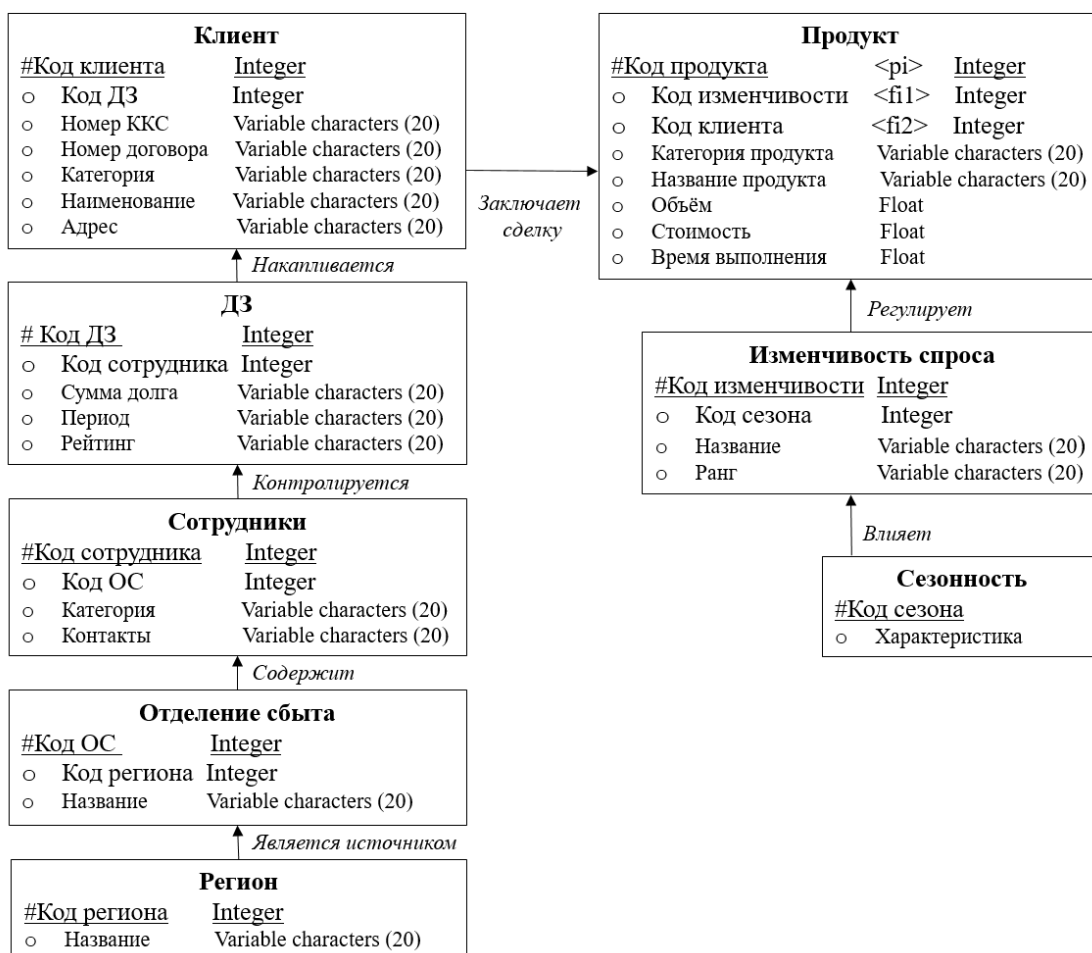


Рисунок 28 – Логическая модель хранилища данных

Так, все сущности данной модели имеют атомарные атрибуты и не образуют повторяющиеся группы, что соответствует первой нормальной форме (1NF). Кроме того, все сущности также находятся и во второй нормальной форме (2NF), т.к. они, как уже выяснилось, находятся в первой нормальной форме, и все не ключевые атрибуты сущностей полностью зависят от первого ключа. Сущности находятся также и в третьей нормальной форме, т.к. они находятся во второй нормальной форме, и никакой не ключевой атрибут не зависит от другого не ключевого атрибута. Другими словами, данная модель соответствует методологии нормализованного хранилища в третьей нормальной форме.

Следующим этапом является построение физической модели хранилища данных из логической. Физическая модель данных применяется для построения схемы базы данных, которая будет использоваться в информационной системе.

Так, решения, принятые на уровне логической модели, определяют границы, в пределах которых можно развивать физическую модель данных. В пределах этих границ можно принимать различные решения.

Таким образом, в физической модели сущности становятся таблицами, а атрибуты – столбцами этих таблиц. Связи реализуются в соответствии с используемыми в базе данных механизмами, например, механизмы обеспечения ссылочной целостности и хранимые процедуры.

Физическая модель показывает структуру таблиц и включает в себя название полей, типы данных полей, ограничения, первичные и внешние ключи и отношения между таблицами. К ключевым особенностям физической модели данных можно отнести [37, 41]:

- спецификация таблиц и полей;
- внешние ключи применяются для определения отношений между сформированными таблицами;
- возможность возникновения необходимости денормализации в соответствии с требованиями пользователей;
- возможность отличая от логической модели, связанная с ограничениями на физическом уровне;
- различность модели для различных СУБД.

Физическая модель является графическим отображением структуры хранилища данных и позволяет анализировать его таблицы (атрибуты, индексы), процедуры. Все это необходимо для последующей корректной генерации DDL-скрипта построения хранилища данных. Данный этап является последним и выполняется непосредственно перед созданием хранилища [41].

Вначале необходимо заменить сущности и атрибуты из логической модели на таблицы и колонки. В соответствии с ограничениями СУБД, таблицы и колонки должны быть на английском языке и не содержать пробелов.

Таким образом, определим таблицы с первичными и внешними ключами, поля и типы данных физической модели хранилища данных, создаваемого для исследуемой компании:

1) таблица «PRODUCT» – продукт:

- Product_ID [integer] PK – код продукта;
- Demand Variability_ID [integer] FK – код изменчивости;
- Customer_ID [integer] FK – код клиента;
- Category_Product [varchar 20] – категория продукта;
- Product_Name [varchar 20] – название продукта;
- Volume [float] – объём;
- Price [float] – стоимость;
- Lead_Time [float] – время выполнения.

2) Таблица «CUSTOMER» – клиент:

- Customer_ID [integer] PK – код клиента;
- Receivables_ID [integer] FK – код ДЗ;
- KKS_Number [varchar 20] – номер контокоррентного счета;
- Number_of_contract [varchar 20] – номер договора;
- Category_Customer [varchar 20] – категория клиента;
- Company_Name [varchar 20] – наименование компании;
- Address [varchar 20] – адрес.

3) Таблица «RESEIVABLE» – дебиторская задолженность:

- Receivables_ID [integer] PK – код ДЗ;
- Employees_ID [integer] FK – код сотрудника;
- Sum [varchar 20] – сумма долга;
- Period [varchar 20] – период образования долга;

- Rating [varchar 20] – рейтинг платежной дисциплины.
- 4) Таблица «EMPLOYEES» – сотрудники:
- Employees_ID [integer] PK – код сотрудника;
 - Office_ID [integer] FK – код отделения сбыта;
 - Employees_Category [varchar 20] – категория сотрудника;
 - Contacts [varchar 20] – контактная информация.
- 5) Таблица «SALES OFFICE» – отделение сбыта:
- Office_ID [integer] PK – код отделения;
 - Region_ID [integer] FK – код региона;
 - Office_Name [varchar 20] – название отделения.
- 6) Таблица «REGION» – регион:
- Region_ID [integer] PK – код региона;
 - Region_Name [varchar 20] – название региона.
- 7) Таблица «DEMAND VARIABILITY» – изменчивость спроса:
- Demand_Variability_ID [integer] PK – код изменчивости;
 - Season_ID [integer] FK – код сезона;
 - Demand_Variability_Name [varchar 20] – название изменчивости;
 - Rank [varchar 20] – ранг.
- 8) Таблица «SEASONALITY» – сезонность:
- Season_ID [integer] PK – код сезона;
 - Characteristic [varchar 20] – характеристика.

На основании перечисленных наименований таблиц с первичными и внешними ключами можно осуществлять построение физической модели хранилища данных в соответствии с упомянутой ранее нотацией IDEF1X, которая в дальнейшем будет использована для настройки DDL скрипта – конечного пункта проектирования.

Предпоследний этап проектирования хранилища данных – физическая модель данных представлена на рис. 29. На нем отображены все связи и сущности процесса управления дебиторской задолженностью.

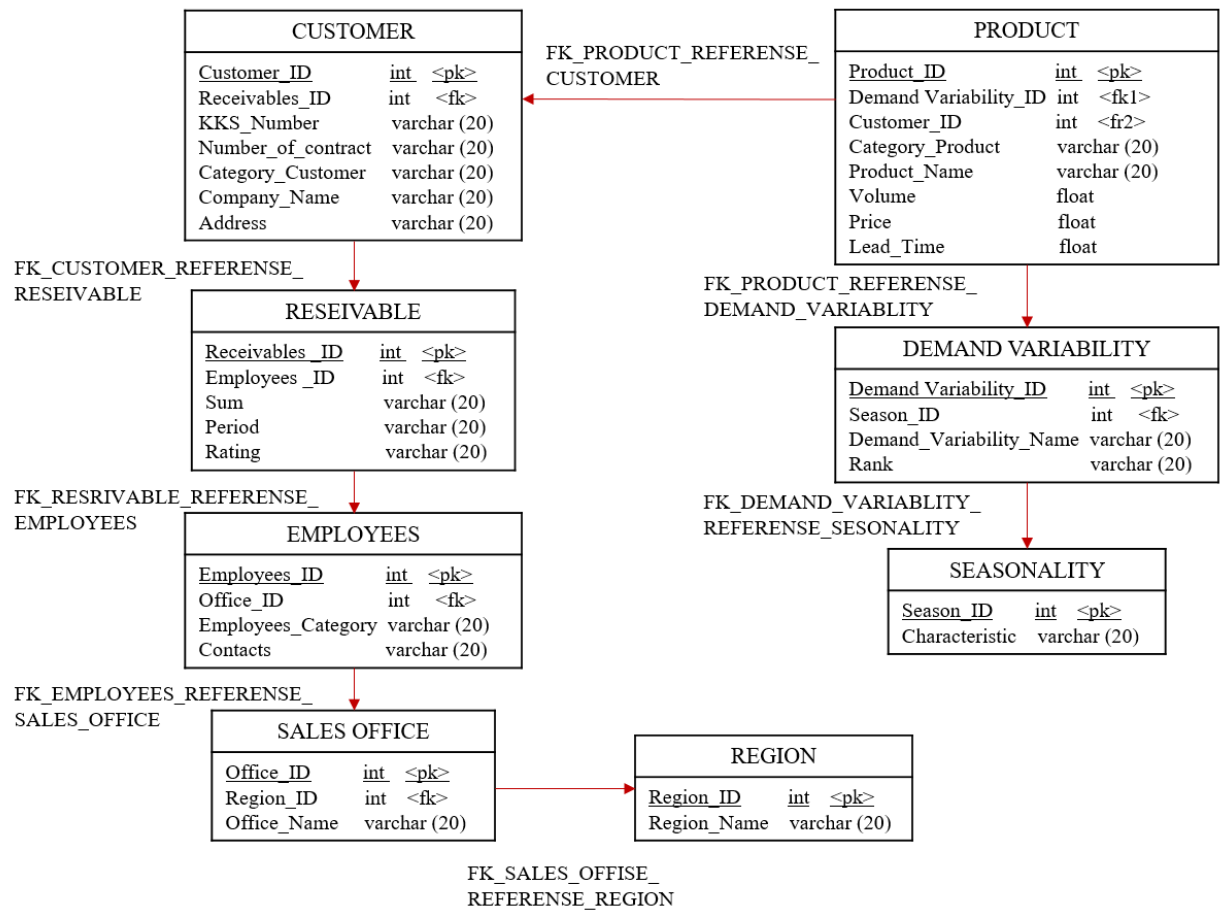


Рисунок 29 – Физическая модель хранилища данных

Таким образом, главной целью построения физической модели данных является генерация DDL-скрипта для создания базы данных. Пример небольшой части подобного скрипта представлен на рис. 30.

```

create table CUSTOMER (
    Customer_ID          int          not null,
    State_ID             int          null,
    Category_Customer    varchar(20)  null,
    Company_Customer     varchar(20)  null,
    State                varchar(20)  null,
    Location             varchar(20)  null,
    constraint PK_CUSTOMER primary key (Customer_ID)  constraint PK_CUSTOMER
primary key (Customer_ID)
)
go

```

Рисунок 30 – DDL-скрипт для создания базы данных

Такой скрипт имеет следующую структуру: вначале происходит проверка на существование таблиц и индексов. Если таблицы или индексы уже существуют, то происходит их удаление. Далее создаются таблицы, после чего определяются внешние ключи и ограничения целостности. Как правило, скрипт формируется в терминах СУБД «Microsoft SQL Server 2017» – платформа для решения критически важных задач в масштабе предприятия, обладающая высокой доступностью, повышенной производительностью и безопасностью. Когда всех действия с физической моделью данных завершены, можно запускать скрипт.

Таким образом, проектирование хранилищ данных представляет собой последовательность шагов по реализации концептуальной, логической и физической моделей. Каждая последующая модель формируется на основании данных о предыдущей, модернизируется и отражает все основные сущности, связи и атрибуты баз данных. В качестве итога проектирования выступает DDL-скрипт, который является конечным пунктом для построения хранилища данных.

3.2 Преобразование измерений в системе «Optimacros»

Прежде чем приступить к преобразованию измерений для формирования отчета по управлению дебиторской задолженностью, необходимо рассмотреть основные особенности платформы «Optimacros», в которой будет происходить дальнейшая работа с данными компании.

Система «Optimacros» – российская универсальная платформа для оптимизационного и консолидационного планирования, бизнес-анализа и интеграций. Также, систему можно охарактеризовать, как конструктор для построения систем планирования, бюджетирования, план-факт анализа, прогнозирования, бизнес-аналитики, BI, EPM и отслеживания эффективности

компаний. В основе системы лежит построение моделей посредством OLAP-систем (технология многомерного представления данных) [2].

В настоящее время подавляющее большинство компаний хранит и анализирует информацию в Excel, который уже долгие годы зарекомендовал себя, как эффективный инструмент работы с электронными таблицами. Однако время не стоит на месте, а объем наработанных данных и сведений растет с невероятной скоростью, в результате чего анализ полученной информации (отчетности и т.п.) становится все более затруднительным из-за массива данных. Именно поэтому, некоторые крупные компании приходят к решению внедрить в функционал своей организации систему «Optimacros», которая дает возможность обработки и анализа больших объемов данных, необходимых для эффективного принятия решений, а также планирования и мониторинга текущей операционной деятельности компании.

Основным отличием от традиционных систем планирования является то, что «Optimacros» – Self-Service конструктор, для использования которого требуются только аналитики (моделирование без программирования, команд разработчика, ТЗ и т.п.). Продукт содержит в себе гибкость MS Excel и надежность оперативных систем.

Данный аналитический инструмент обеспечивает широкое преобразование бизнес-выводов благодаря автоматическому обновлению и автоматизации процессов обнаружения, очистки и публикации данных. Не нужно искать отдельный файл с необходимыми сведениями, все сфокусировано в одном месте, а в случае добавления новых данных – система обновляется. При работе с большим массивом данных подобная функция является решающей.

Система «Optimacros» имеет возможность гибкой настройки измерений. Так, в связи с тем, что компания осуществляет распределение тепловой энергии по нескольким административным округам, в которых также присутствует распределение по муниципальным районам, было принято

решение создать два измерения: «Административный округ» и «Муниципальный район».

Каждому административному округу соответствует свое отделение сбыта. Причем, у более крупных округов отделение соответствует названию округа, а некоторые маленькие округа могут относиться к одному отделению с видоизмененным названием-связкой. Для таких случаев лучше всего подойдет добавление свойства в справочник-измерение с прописью соответствующей формулы-условия. Для проектирования данного отчета подобные манипуляции не требуются, поэтому указанный шаг будет пропущен.

Для того, чтобы присвоить конкретный район к конкретному округу, в качестве родительского справочника (верхнего уровня иерархии) было выбрано измерение АО. Причем, для дальнейшего анализа общей суммы по отчетам, у измерения АО также был настроен верхний уровень «Total». Вид иерархия измерения МР на примере двух административных округов представлен на рис. 31.

	Имя элемента	Справочник	Родитель
Total	Total	Административный округ (АО)	
ЦАО	ЦАО	Административный округ (АО)	Total
Арбат	Арбат	Муниципальный район (МР)	ЦАО
Басманный	Басманный	Муниципальный район (МР)	ЦАО
Замоскворечье	Замоскворечье	Муниципальный район (МР)	ЦАО
Красносельский	Красносельский	Муниципальный район (МР)	ЦАО
Мещанский	Мещанский	Муниципальный район (МР)	ЦАО
Пресненский	Пресненский	Муниципальный район (МР)	ЦАО
Таганский	Таганский	Муниципальный район (МР)	ЦАО
Тверской	Тверской	Муниципальный район (МР)	ЦАО
Хамовники	Хамовники	Муниципальный район (МР)	ЦАО
Якиманка	Якиманка	Муниципальный район (МР)	ЦАО
САО	САО	Административный округ (АО)	Total
Аэропорт	Аэропорт	Муниципальный район (МР)	САО
Беговой	Беговой	Муниципальный район (МР)	САО
Бескудниковский	Бескудниковский	Муниципальный район (МР)	САО
Войковский	Войковский	Муниципальный район (МР)	САО
Восточное Дегунино	Восточное Дегунино	Муниципальный район (МР)	САО
г.о. Химки	г.о. Химки	Муниципальный район (МР)	САО
Головинский	Головинский	Муниципальный район (МР)	САО
Дмитровский	Дмитровский	Муниципальный район (МР)	САО
Западное Дегунино	Западное Дегунино	Муниципальный район (МР)	САО
Коптево	Коптево	Муниципальный район (МР)	САО
Левобережный	Левобережный	Муниципальный район (МР)	САО
Молжаниновский	Молжаниновский	Муниципальный район (МР)	САО
Савеловский	Савеловский	Муниципальный район (МР)	САО
Сокол	Сокол	Муниципальный район (МР)	САО
Тимирязевский	Тимирязевский	Муниципальный район (МР)	САО
Ховрино	Ховрино	Муниципальный район (МР)	САО
Хорошевский	Хорошевский	Муниципальный район (МР)	САО

Рисунок 31 – Измерение «МР» в системе Optimacros

Для преобразования измерения «Отделения сбыта» также был добавлен верхний уровень «Total» для будущих расчетов суммы и возможности создания сводных таблиц. Справочник представляет собой список отделений от 1 до 12 по аналогии с измерением МР. Отличием является – наличие только одного родительского уровня иерархии – «Общая сумма».

Для преобразования измерения CUSTOMER было принято решение взять за основу номер контокоррентного счета потребителя, который представляет собой счёт, на котором фиксируются все операции банка с клиентом и учитывается взаимная задолженность банка и клиента. Также, были созданы свойства-колонки, которые содержат в себе основную информацию по потребителю: административный округ, отделение сбыта, муниципальный район, ИНН, номер договора и наименование делового партнера. В случае необходимости, можно добавить дополнительный свойства, например, категория потребителя, дата заключения договора и т.д. в зависимости от дальнейшей логики многомерного представления данных.

Для дополнения описания потребителей необходимо также добавить измерения: «Альтернативная группа» и «Группа потребления» (рис. 32).

	Имя элемента	Справочник	Родитель
Total	Total	Группа потребления (ГП)	
Управляющие организации	Управляющие организации	Группа потребления (ГП)	Total
0001 ДЕЗ	0001 ДЕЗ	Альтернативная группа (АГ)	Управляющие организации
0002 ЖСК	0002 ЖСК	Альтернативная группа (АГ)	Управляющие организации
0004 Ведомственное жилье	0004 Ведомственное жилье	Альтернативная группа (АГ)	Управляющие организации
0025 Управляющие компании по муницип. жилью	0025 Управляющие компании по муницип. жилью	Альтернативная группа (АГ)	Управляющие организации
0026 Управляющие компании по ТСЖ и ЖСК	0026 Управляющие компании по ТСЖ и ЖСК	Альтернативная группа (АГ)	Управляющие организации
0027 Прочие управляющие компании	0027 Прочие управляющие компании	Альтернативная группа (АГ)	Управляющие организации
0029 Управляющие организации по факту отпуска	0029 Управляющие организации по факту отпуска	Альтернативная группа (АГ)	Управляющие организации
0048 Юр. лицо в МКД на прямых расчетах-неакт.	0048 Юр. лицо в МКД на прямых расчетах-неакт.	Альтернативная группа (АГ)	Управляющие организации
0049 Физическое лицо частный жилой фонд	0049 Физическое лицо частный жилой фонд	Альтернативная группа (АГ)	Управляющие организации
0050 Юр. лицо (жилая квартира)	0050 Юр. лицо (жилая квартира)	Альтернативная группа (АГ)	Управляющие организации
Социальная сфера	Социальная сфера	Группа потребления (ГП)	Total
0008 Департамент культуры	0008 Департамент культуры	Альтернативная группа (АГ)	Социальная сфера
0009 Департамент образования	0009 Департамент образования	Альтернативная группа (АГ)	Социальная сфера
0010 Департамент сем. и молод. полит. Москвы	0010 Департамент сем. и молод. полит. Москвы	Альтернативная группа (АГ)	Социальная сфера
0013 Департамент здравоохранения	0013 Департамент здравоохранения	Альтернативная группа (АГ)	Социальная сфера
0015 ОУ здравоохранения	0015 ОУ здравоохранения	Альтернативная группа (АГ)	Социальная сфера
0016 ОУ образования	0016 ОУ образования	Альтернативная группа (АГ)	Социальная сфера
0017 ОУ культуры	0017 ОУ культуры	Альтернативная группа (АГ)	Социальная сфера
0019 ОУ соц. защиты населения	0019 ОУ соц. защиты населения	Альтернативная группа (АГ)	Социальная сфера
0020 МОСКОМСПОРТ	0020 МОСКОМСПОРТ	Альтернативная группа (АГ)	Социальная сфера
0021 Департамент соц. защиты населения	0021 Департамент соц. защиты населения	Альтернативная группа (АГ)	Социальная сфера
Прочие бюджетные организации	Прочие бюджетные организации	Группа потребления (ГП)	Total
0014 ГУВД и прокуратуры	0014 ГУВД и прокуратуры	Альтернативная группа (АГ)	Прочие бюджетные организации
0022 УСЗН	0022 УСЗН	Альтернативная группа (АГ)	Прочие бюджетные организации
0023 Прокуратура	0023 Прокуратура	Альтернативная группа (АГ)	Прочие бюджетные организации
0024 УСК МФ РФ	0024 УСК МФ РФ	Альтернативная группа (АГ)	Прочие бюджетные организации
0028 Прочие бюджетные организации	0028 Прочие бюджетные организации	Альтернативная группа (АГ)	Прочие бюджетные организации
0031 Министерство обороны РФ	0031 Министерство обороны РФ	Альтернативная группа (АГ)	Прочие бюджетные организации
0045 ДИГМ, МИК	0045 ДИГМ, МИК	Альтернативная группа (АГ)	Прочие бюджетные организации

Рисунок 32 – Измерение «АГ» в системе Optimacros

Альтернативная группа представляет собой четырехзначный код, который соответствует основной специфике деятельности контрагента, а

также имеет в качестве верхнего уровня – группу потребления, которая в свою очередь характеризуется сферой (например, социальная сфера). Иерархия измерения АГ, как и предыдущие измерения, имеет родительский справочник с верхним уровнем «Total».

Также, по аналогии с АГ и ГП, вводится дополнительная характеристика для потребителей и деловых партнеров, а именно – измерение «Принадлежность». Данное измерение не имеет верхних уровней и представляет собой список возможных вариантов принадлежности организации контрагента (например, к фондам, сферам, бюджетным учреждениям и т.п.). После того, как все рассмотренные ранее измерения прошли этап создания и преобразования, необходимо приступить к формированию вытекающего из них показателя.

В рамках данного и предыдущих параграфов упоминалось понятие «Категория потребителя». Данная категория представляет собой расчетное наименование контрагента, которое присваивается в зависимости от принадлежности и альтернативной группы по каждому отдельному потребителю. Данное измерение представляет собой список существующих на предприятии категорий, а именно: ГУК; ЧУК; МО, силовые ведомства; перепродавцы; социальная сфера, прочие бюджетные организации; строительные организации и прочие.

Так как в компании применяется комплекс мер-мероприятий для управления дебиторской задолженностью, необходимо также создать измерение-список действующих мероприятий. Данное измерение будет необходимо для внедрения в систему логики расчета и создания, своего рода, сводных таблиц для аналитики. Также, для детализации комментариев, необходимы измерения: «Причина наличия ДЗ» и «Причина обзвона». Данные измерения будут служить в качестве выпадающего списка отдельных кубов отчета. Причем, данная функция возможна для любого из упомянутых ранее справочников модели. Все созданные и преобразованные измерения можно

использовать в дальнейшем, как для построения мультикубов отчета и mapping, так и для настройки форматов отдельных кубов.

Для построения мультикуба управления дебиторской задолженностью не требуется создание измерения PRODUCT, EMPLOYEES, DEMAND VARIABILITY и SEASONALITY, т.к. в исходном файле структуры отчета данные колонки не фигурируют и не оказывает существенного влияния для проектирования логики алгоритма.

Измерения, созданные в системе «Optimacros» содержат уникальные Item Name, которые не могут повторяться дважды. Однако, в случае необходимости добавления идентичных названий, одним из возможных вариантов решения может случить – создания нумерованных Item Name, например, 1.1, 1.2 и т.д. Создав дополнительное свойство в формате текста и вставив нужные названия в строку, необходимо выбрать в качестве Display Name созданное свойство и таким образом получить визуально идентичные Item элементы.

Как правило, для решения одного кейса всегда существует как минимум два варианта реализации, т.е. понятие системы, как Self-Service конструктора действительно оправдано.

Возвращаясь к проектированию измерений, стоит отметить, что для создания мультикуба отчета также необходимо измерение времени. Время имеет несколько классических альтернатив, а именно: день, месяц, год. Так как вводные данные о потребителях для плана управления дебиторской задолженностью формируются на основе ежедневного отчета об авансах потребителей, в качестве измерения времени будут случить Days Subset – дни.

Таким образом, необходимые измерения были созданы и прошли необходимые этапы преобразования. Следующим этапом работы в системе является – построение мультикуба отчета с соответствующей для нужд организации логикой расчетов, визуализации данных и основным требованиям пользователей, которые были определены посредством опроса сотрудников отделений сбыта во второй главе данной работы.

3.3 Построение кубов и отчетов

В системе «Optimacros» есть возможность создания кубов с настраиваемым количеством расчётных столбцов различного формата данных. Кубы создаются на основе доступных измерений, помимо них также есть измерения времени (года, месяца, дни).

Для создания мультикуба отчета плана по управлению дебиторской задолженностью необходимо выбрать измерения, которые будут выступать в качестве строк, колонок и фильтров. В данном кейсе, строки выступают в формате справочника-измерения номера контокоррентного счета потребителя, кубы являются колонками, а измерение времени переброшены в фильтр. Подобное расположение составных частей отчета позволит в дальнейшем наиболее быстро ознакомиться с данными за другую дату.

Кубы отчета плана по управлению дебиторской задолженностью создаются на основе колонок файла MS Excel, который был предоставлен со стороны компании. После того, как все кубы были добавлены, необходимо определиться с форматами каждого из них. Данный этап является достаточно ответственным, т.к. в случае смены формата после импорта данных – куб очищается. В результате чего, данные необходимо импортировать заново. Именно поэтому, для того, чтобы избежать потери времени в будущем, необходимо определиться с форматами кубов на начальном этапе. Кроме того, наиболее эффективно использовать в качестве форматов измерения справочников, т.к. в дальнейшем при помощи существующих в мультикубе измерений можно строить различные сводные таблицы по конкретному показателю.

Так, для тех кубов, информация о которых уже находится в справочнике, необходимо установить формат соответствующего измерения. Исключением является случай, если аналитик хочет использовать данные в кубе для формирования нового справочника при помощи скрипта для автоматизации обновления данных.

Таким образом, для данного кейса, кубы импорта будут иметь форматы, указанные на рис. 33.

	Мультикуб	Формула	Формат
План по Досудебной работе			
Округ	План по Д...	'5.13 ...	Административный округ (АО)
Отделение сбыта	План по Д...	'5.13 ...	Text
Номер контокоррентного счета	План по Д...	'5.13 ...	ККС
ИНН потребителя	План по Д...	'5.13 ...	Text
Группа потребления	План по Д...	'5.13 ...	Группа потребления (ГП)
Альтернативная группа	План по Д...	'5.13 ...	Альтернативная группа (АГ)
Принадлежность	План по Д...	'5.13 ...	Принадлежность
№ договора	План по Д...	'5.13 ...	Text
Дата заключения	План по Д...	'5.13 ...	Date
Дата расторжения договора	План по Д...	'5.13 ...	Date
Потребитель	План по Д...	'5.13 ...	Text
Статус	План по Д...	'5.13 ...	Text
Муниципальный район	План по Д...	'5.13 ...	Муниципальный район (МР)
Начислено за месяц 2021 (факт с учетом всех перерасчетов)	План по Д...	'5.13 ...	Number
Начислено за месяц 2021 с учетом 1/12	План по Д...	'5.13 ...	Number
Начислено за месяц 2021 с учетом 1/12 и перерасчетов	План по Д...	'5.13 ...	Number
ДЗ начальная дата образования	План по Д...	LEFT('Пери...	Text
ДЗ конечная дата образования	План по Д...	IF ...	Text
Оплачено в отчетном месяце	План по Д...	SELECT('Он...	Number
ДЗ всего на отчетную дату	План по Д...	SELECT('Д...	Number
Период образования долга	План по Д...	SELECT('Пе...	Text
Авансы на отчетную дату	План по Д...	SELECT('Ав...	Number
Оплачено в текущем месяце	План по Д...	'5.13 ...	Number
ДЗ всего на текущую дату	План по Д...	'5.13 ...	Number
Период образования долга	План по Д...	'5.13 ...	Text
Авансы на текущую дату	План по Д...	'5.13 ...	Number
Признак расчета "с учетом 1/12"	План по Д...	'5.13 ...	Text
Наличие дополнительного соглашения ЭДО	План по Д...	'5.13 ...	Text

Рисунок 33 – Настройка форматов кубов

После настройки форматов кубов можно приступать к загрузке данных в систему. Загрузка происходит посредством экспорта и импорта, что позволяет выгрузить данный мультикуб в настроенном формате в виде файла MS Excel и в последующем, вернуть его обратно в систему с уже наполненным содержимым. После выгрузки данных, необходимо перенести сведения из файла MS Excel, предоставленного компанией, посредством обычного копирования данных. При этом, важно осуществить проверку корректности цифр и текста, а также отслеживать соответствие ККС в измерении номеру контокоррентного счета в импортируемой колонке. Наиболее простой способ – копирование столбца и его вставка в колонку измерения. Такие действия дают системе понять, что необходимо вставить данные по конкретной строке в конкретную часть измерения.

После корректировки файла, осуществляется импорт данных. Данные проходят проверку аналитика на корректность и, в случае несоответствия – устраняются ошибки.

Далее необходимо интерпретировать логику формул из файла MS Excel в систему «Optimacros», а именно, колонок: категория потребителя и рейтинг платежной дисциплины. На основе данных кубов можно визуализировать отчет в виде дашбордов по каждой конкретной группе и рейтингу. Кроме того, рейтинг платежной дисциплины, как уже упоминалось в данной работе, является неотъемлемой частью для определения перечня мероприятий по управлению дебиторской задолженностью, которые будут предъявлены к деловому партнеру.

Категория потребителя рассчитывается исходя из принадлежности и альтернативной группы. В компании присутствуют определенные условия для присвоения конкретной категории. Например, если альтернативная группа в мультикубе mapping равняется категории «X», тогда итоговой категорией будет категория «X». В расчете данного показателя можно встретить логику трех видов: принадлежность равняется категории «X»; альтернативная группа равняется категории «X»; и принадлежность, и альтернативная группа равняется категории «X». Интерпретация данной логики на систему представлена на рис. 34.

```
IF SELECT('Mapping Принадлежность'. 'mapping', 'Принадлежность': 'Принадлежность') = 'Категория X'  
THEN 'Категория X' ELSE  
  
IF SELECT('Mapping АГ'. 'Mapping', 'Альтернативная группа (АГ)': 'Альтернативная группа') = 'Категория X'  
THEN 'Категория X' ELSE  
  
IF SELECT('Mapping АГ'. 'Mapping', 'Альтернативная группа (АГ)': 'Альтернативная группа') = 'Категория X'  
AND SELECT('Mapping Принадлежность'. 'mapping', 'Принадлежность': 'Принадлежность') = 'Категория X'  
THEN 'Категория X' ELSE
```

Рисунок 34 – Логика расчета категории потребителя

Функция SELECT представляет собой аналог функции ВПР MS Excel. Данная функция возвращает в куб значение из куба X в соответствии со значениями куба Y в формате измерения Z.

Другими словами, если необходимо подтянуть данные о сумме из МК на измерении, то это можно сделать в том случае, если в МК-получатель имеет куб на данном измерении.

Для подобного сопоставления данных были созданы мультикубы mapping для принадлежности и альтернативной группы на соответствующих измерениях. Мультикуб mapping строиться без измерения времени и является уникальным для любого временного периода, нося справочный характер. Причем, важно, что куб, в котором итоговым значением является категория потребителя должен иметь формат «Категория потребителя» (т.е. формат справочника). В противном случае система не будет пропускать логическое выражения, не воспринимая его, как верное. Пример мультикуба mapping представлен на рис. 35.

	mapping
Бюджет. ГБУ "Жилищник"	ГУК
Бюджет. ГБУ "ЭВАЖД"	
Бюджет. Минобороны	МО, силовые ведомства
Бюджет. Силовые ведомства	МО, силовые ведомства
Бюджет. ТСО 2го уровня	Перепродавцы
Прочие	Социальная сфера, прочие бюджетные организации
Прочие бюджетные	Социальная сфера, прочие бюджетные организации
Прочие. Внутригрупповой оборот	Перепродавцы
Прочие. ГБУ "Жилищник"	ГУК
Прочие. ГБУ "ЭВАЖД"	ГУК
Прочие. ГК "Капитал-Инвест"	Прочие
Прочие. ГК "ПИК"	Прочие
Прочие. ГК "ПИК". Строительные организации	Строительные организации
Прочие. ГК "ЮНИ-ДОМ"	Прочие
Прочие. ГК "Юни-Сервис"	Прочие
Прочие. ГК "Юни-Сервис". Строительные организации	Строительные организации
Прочие. ГУП "ДЕЗ"	ГУК

Рисунок 35 – Мультикуб mapping для принадлежности

Следующим расчетным показателем является рейтинг платежной дисциплины потребителя. Данный показатель имеет большое значения для всего дальнейшего алгоритма, т.к. является решающим для определения перечня мероприятий, которые будут применены к контрагенту.

Рейтинг присваивается от 0 до 4 на основе периода образования долга. Также, существует рейтинг – «ДЗ свыше 3-х лет», который присваивается наиболее проблемным потребителям.

Большая часть мероприятий применяется к контрагентам с рейтингом 3 и 4. Наиболее серьезными считаются потребители, период образования долга у которых превышает год.

Для удобства мониторинга данных, в системе присутствует настройка условного форматирования, которая позволяет подсвечивать значения в зависимости от изменения в отчётном периоде. Таким образом, было принято решение создать дополнительный куб – рейтинг платежной дисциплины на отчетную дату. В качестве отчетной даты в данном случае выступает последний день предыдущего месяца. Логика условного форматирования имеет следующий вид: если ячейка подсвечивается красным – рейтинг упал, если зеленым – повысился. В данном кейсе можно проследить исключительно повышение рейтинга на новый уровень задолженности (рис. 36).

Категория потребителя	Рейтинг платежной дисциплины потребителя (1-2-3-4) на отчетную дату	Рейтинг платежной дисциплины потребителя (1-2-3-4) на текущую дату
Прочие		0
Прочие		0
Прочие	0	2
Прочие	2	2
Прочие	0	2
Прочие		0
Прочие	2	2
Прочие	4	4
Прочие	0	2
Прочие	2	3
Прочие	0	2
Прочие	4	4
Прочие		0

Рисунок 36 – Условное форматирование рейтинга ПДП

На рисунке можно увидеть, что рейтинг подсвечивается красным, но на отчетную дату ячейка является пустой. Это говорит о том, что в прошлом анализируемом периоде данный потребитель еще не являлся клиентом компании и не имел задолженность перед ней.

Далее необходимо прописать логику, согласно которой в кубах соответствующих мероприятий будет высвечиваться значение «Да», т.е. данное мероприятие по отношению к потребителю соответствующей строки необходимо применить.

Мероприятия, которые предъявляются к потребителям с рейтингом 1 и выше: обзвон потребителей; встреча с потребителем в отделение сбыта;

информационное сообщение о задолженности. Механизм взаимодействия с потребителями для данных мероприятий был прописан во второй главе данной работы.

Мероприятия, которые применяются к потребителям с рейтингом 3 и выше – перечисленные ранее мероприятия, а также: направление обращения Управы / Префектуры / ДЖКХиБ; направление обращения в вышестоящий отраслевой орган; распространение информации о задолженности потребителя через СМИ; направление в МЖИ (заявление); направление в МЖИ (обращение); ПДК; направление в МТУ Ростехнадзора (обращение); направление в Прокуратуру (обращение); служба корпоративной защиты направление в адрес директора предприятия соответствующего района ЭФ. Для данных мероприятий помимо даты и комментария, также созданы кубы: номер обращения и результат рассмотрения.

Большая часть мероприятий рассчитывается исходя из значения рейтинга по логике: если рейтинг равняется 1 или 2 или и т.д., тогда «Да», иначе пустота. Однако некоторые из них имеют свою уникальную логику расчета, которая базируется на периоде образования долга и номере договора.

Кроме того, в связи с тем, что на новую дату могут происходить определенные изменения, например, появиться новые или закрываться старые потребители, измениться принадлежность, альтернативная группа или группа потребления, необходимо создать новые представления, в которых будет отображаться список изменений по сравнению с отчетной датой.

Новые представления являются копией старых со скрытыми ненужными для данной страницы кубами и Boolean кубом, который будет отсеивать данные по ККС, не соответствующим условию. Пример подобных условий представлен на рис. 37.

```
IF SELECT('Номер контокоррентного счета', 'Days':DAY(BOMONTH() - 1)) <> ""
AND SELECT('Группа потребления', 'Days':DAY(BOMONTH() - 1)) = 'Группа потребления' AND 'Номер контокоррентного счета' <> ""
THEN 'Изменения'. 'Без изменений' ELSE IF SELECT('Номер контокоррентного счета', 'Days':DAY(BOMONTH() - 1)) <> ""
AND SELECT('Группа потребления', 'Days':DAY(BOMONTH() - 1)) <> 'Группа потребления'
AND 'Номер контокоррентного счета' <> "" THEN 'Изменения'. 'Изменение' ELSE ""
```

Рисунок 37 – Boolean условие для изменений в группе потребления

Таким образом, было создано четыре Boolean куба с условием фильтрации для каждого конкретного показателя, а также получены отдельные листы со списками изменений, с которыми можно ознакомиться при нажатии, по аналогии с листами файлов MS Excel. Сведения из этих представлений полезны для мониторинга корректности данных об альтернативной группе и группе потребления. Так, если в прошлом периоде у потребителя была другая АГ или ГП, можно узнать, является ли это истиной или ошибкой при заполнении данных о контрагентах.

После того, как все данные и логика расчетов были перенесены, необходима сверка. Для более удобной проверки и навигации в будущем необходимо создание Boolean куба, который будет выступать в качестве куба «Общего условия».

В связи с тем, что план управления ДЗ содержит большое количество колонок и строк по разным округам, отделениям и другим характеристикам, которые требуется сужать для поиска желанной информации, было принято решение создать дополнительный мультикуб со списком условий. Данный МК представляет собой список строк (кубов) в форматах, соответствующих названию кубов (рис. 38).

Список условий

<----Основные---->	
Данные по ККС	Да
Выберете АО	
Выберете ОС	
Выберете МР	
Выберете АГ	
Выберете ГП	
Выберете Принадлежность	
Выберете категорию потребителя	
Выберете рейтинг потребителя	
<----Мероприятия---->	
Обзвон потребителей	
Встреча с потребителем в ОС	
Информационное сообщение о задолженности	
Направление обращения Управы / Префектуры / ДЖКХиБ	
Направление обращения в вышестоящий отраслевой орган, (через Ф11)	
Распространение информации о задолженности потребителя через СМИ	
Направление в МЖИ (заявление)	
Направление в МЖИ (обращение)	
ГДК	
Направление в МТУ Ростехнадзора (обращение)	
Направление в Прокуратуру (обращение)	
Служба корпоративной защиты	
Направление в адрес директора предприятия соответствующего района ЭФ	
<----Изменения---->	
Группа потребления	Изменение
Альтернативная группа	Изменение
Принадлежность	Изменение
Потребитель	Новый

Рисунок 38 – Мультикуб для настройки фильтрации

Данный мультикуб имеет несколько разделителей, которые помогают классифицировать кубы по категориям, отвечающим за конкретные составные части отчета. В качестве основных кубов выступают кубы на измерениях созданных справочников и отражают вариативность информации о потребителях.

Разделитель «Мероприятия» создан для фильтрации основного отчета по значениям да/нет в кубе по конкретному мероприятию. Кубы «Изменения» настроены фиксировано, для того, чтобы пользователь мог сразу ознакомиться со списком изменений.

Таким образом, в данном мультикубе, при нажатии на каждую ячейку высвечивается выпадающий список, в котором можно выбрать нужный административный округ, муниципальный район, отделение сбыта и т.д. или «Да» по какому-то либо из мероприятий. Фильтр позволит отследить более узкие данные и провалиться в интересующую пользователя комбинацию.

При этом, для того, чтобы связать данные из МК «Список условий» с данными из отчета по управлению дебиторской задолженностью компании, необходимо создать Boolean куб с соответствующей логикой связи. Каждый аналитик по-своему проводит интеграцию, видоизменяя формулу в тех или иных местах (рис. 39).

```
'ККС' = 'Список условий'. 'Данные по ККС' AND (IF 'Список условий'. 'Выберете МР' = "" THEN TRUE ELSE 'Муниципальный район' = 'Список условий'. 'Выберете МР') AND (IF 'Список условий'. 'Выберете категорию потребителя' = "" THEN TRUE ELSE 'Категория потребителя' = 'Список условий'. 'Выберете категорию потребителя') AND (IF 'Список условий'. 'Выберете рейтинг потребителя' = "" THEN TRUE ELSE 'Рейтинг (На измерении)' = 'Список условий'. 'Выберете рейтинг потребителя') AND (IF 'Список условий'. 'Обзвон потребителей' = "" THEN TRUE ELSE 'Обзвон' = 'Список условий'. 'Обзвон потребителей') AND (IF 'Список условий'. 'Встреча с потребителем в ОС' = "" THEN TRUE ELSE 'Встреча' = 'Список условий'. 'Встреча с потребителем в ОС') AND (IF 'Список условий'. 'Информационное сообщение о задолженности' = "" THEN TRUE ELSE 'ИС' = 'Список условий'. 'Информационное сообщение о задолженности') AND (IF 'Список условий'. 'Направление обращения Управы / Префектуры / ДКХИБ' = "" THEN TRUE ELSE 'Управы / Префектуры / ДКХИБ' = 'Список условий'. 'Направление обращения Управы / Префектуры / ДКХИБ') AND (IF 'Список условий'. 'Направление обращения в вышестоящий отраслевой орган, (через Ф11)' = "" THEN TRUE ELSE 'Через Ф11' = 'Список условий'. 'Направление обращения в вышестоящий отраслевой орган, (через Ф11)') AND (IF 'Список условий'. 'Распространение информации о задолженности потребителя через СМИ' = "" THEN TRUE ELSE 'Через СМИ' = 'Список условий'. 'Распространение информации о задолженности потребителя через СМИ') AND (IF 'Список условий'. 'Направление в ИЖИ (заявление)' = "" THEN TRUE ELSE 'ИЖИ (заявление)' = 'Список условий'. 'Направление в ИЖИ (заявление)') AND (IF 'Список условий'. 'Направление в ИЖИ (обращение)' = "" THEN TRUE ELSE 'ИЖИ (обращение)' = 'Список условий'. 'Направление в ИЖИ (обращение)') AND (IF 'Список условий'. 'ПДК' = "" THEN TRUE ELSE 'ПДК' = 'Список условий'. 'ПДК') AND (IF 'Список условий'. 'Направление в ИТУ Ростехнадзора (обращение)' = "" THEN TRUE ELSE 'ИТУ Ростехнадзора (обращение)' = 'Список условий'. 'Направление в ИТУ Ростехнадзора (обращение)') AND (IF 'Список условий'. 'Направление в Прокуратуру (обращение)' = "" THEN TRUE ELSE 'В Прокуратуру (обращение)' = 'Список условий'. 'Направление в Прокуратуру (обращение)') AND (IF 'Список условий'. 'Служба корпоративной защиты' = "" THEN TRUE ELSE 'Служба КЗ' = 'Список условий'. 'Служба корпоративной защиты') AND (IF 'Список условий'. 'Направление в адрес директора предприятия соответствующего района ЭФ' = "" THEN TRUE ELSE 'Направление в адрес директора' = 'Список условий'. 'Направление в адрес директора предприятия соответствующего района ЭФ') AND (IF 'Список условий'. 'Выберете ГП' = "" THEN TRUE ELSE 'Группа потребления' = 'Список условий'. 'Выберете ГП') AND (IF 'Список условий'. 'Выберете АГ' = "" THEN TRUE ELSE 'Альтернативная группа' = 'Список условий'. 'Выберете АГ') AND (IF 'Список условий'. 'Выберете Принадлежность' = "" THEN TRUE ELSE 'Принадлежность' = 'Список условий'. 'Выберете Принадлежность')
```

Рисунок 39 – Boolean куб «TOTAL Фильтр»

В общей сложности, логика выглядит следующим образом: «Если 'Список условий'. 'Выберете X' = "" THEN TRUE ELSE 'X' = 'Список условий'. 'X'», что позволяет «не затирать» данные в случае итога первой части формулы – FALSE.

После завершения общего фильтра, мультикуб отчета по управлению дебиторской задолженностью можно считать практически завершенным. Далее, необходимо перенести представления на контактную таблицу (это необходимо, чтобы неопытные пользователи не сбили настройки введенных формул). Также в КТ добавляется представление МК списка условий. После того, как все необходимые для отчета МК находятся в одном месте, следующими этапами со стороны организации будут – тестирование и наполнение данными по новым периодам времени.

Как правило, для автоматизации наполнения данными создаются интеграционные скрипты, которые самостоятельно загружают и преобразуют файлы MS Excel в необходимый формат. Данная процедура производится в определенное в системе время. Как правило это происходит ночью, чтобы с начала рабочего дня можно было приступать к работе с отчетом. Для данного кейса интеграционный скрипт создаваться не будет, т.к. подобная автоматизация вводится на конечном этапе переноса в хранящее всех отчетов и форм компании.

Как уже отмечалось в данной работе, отчет по управлению дебиторской задолженностью формируется на основе ежедневного отчета об авансах потребителей. Кубы, представленные на рис. 33, являются классическими для большинства отчетов компании. Так, для автоматизации обновления текущего и последующих отчетов компании, из отчета об авансах, необходимо создать мультикуб данного отчета по аналогии с прописанным ранее алгоритмом и связать его с планом по управлению ДЗ при помощи прямых ссылок на кубы. Таким образом, при обновлении данных в одном мультикубе, они автоматически обновляются во всех кубах МК, в которых присутствуют ссылки.

В ходе построения мультикуба отчета, было принято решение добавить дополнительные кубы: «Банкрот» и «Реструктуризация». Данные для этих кубов берутся из отчета по авансам потребителей за месяц. Однако, данный МК построен на измерении нумерованного справочника (Num: 1,2,3...), что

затрудняет классический перенос данных при помощи функции SELECT. Для того, чтобы сопоставить данные из МК на num в МК на измерении, необходимо настроить скрипт.

Настройка скрипта осуществляется по шаблону из библиотеки скриптом для конкретной манипуляции с данными. С помощью скриптов (сценариев) можно автоматизировать часто выполняемые пользователями задачи и заменить работу некоторых функций.

Для переноса данных необходимо настроить скрипт внутри модели, такие скрипты называются обычными. Также существуют интеграционные скрипты, о которых уже упоминалось ранее. При настройке интеграционного скрипта, необходимая информация берется не из модели, а извне. Кроме того, существуют комбинированные скрипты, которые совмещают возможности обычных и интеграционных скриптов [3].

Для создания и настройки скрипта также необходим дополнительный корневой скрипт, который будет служить в качестве основы. Его также можно скопировать из библиотеки скриптов и добавить в список, без необходимости последующего запуска.

Таким образом, для настройки скрипта необходимо указать только основные составляющие переноса: мультикуб источник/приемник, куб источник/приемник, измерение времени источника/приемника. Если говорить о переносе данных, то остальные части скрипта являются стандартными. Однако в случае с автоматическим обновлением справочника, дополнительной особенностью является – указание условия отсеивания уже существующих Item Name при помощи функции FINDBYNAME.

Стоит отметить, что, после загрузки новых данных, скрипт необходимо активировать вновь. Это требуется для того, чтобы система могла считать данные за новый период, т.к. с момента предыдущего запуска, значения изменились. В качестве VIEW скрипта выступает новое представление мультикуба с пометкой «4S» в названии, в котором все измерения находятся в кубах и строках, и не содержат фильтров.

Так, скрипт для переноса данных из месячного отчета об авансах потребителей в план по управлению дебиторской задолженностью представлен на рис. 40.

```

const ENV = {
  CORE: "core_export_import",
  SHOUTER_LVL: "INFO",
  SRC: {
    TYPE: "OM_MULTICUBE",
    PARAMS: {
      NAME: "Отчёт 5.13",
      VIEW: "4S_Отчёт 5.13",
      // FORMULA_FILTER:
    }
  },
  DEST: {
    TYPE: "OM_MULTICUBE",
    PARAMS: {
      NAME: "Для Досудебной работы",
      TRANSFORM: {
        DOWNLOAD_TRANSFORM_FILE: false,
        DOWNLOAD_IMPORT_REPORT_FILE: true,
        CHARSET: "UTF-8",
        SRC_TO_DEST_COLUMN_MAP: {
          "125": "Признак банкротств",
          "134": "Реструктуризация (период)",
        },
      },
      DIMENSIONS: {
        dim1: {
          NAME: "ds. Отчёт 5.13",
          SRC_COLUMN_NAME: "ds. Отчёт 5.13",
          CAN_USE_IN_CLEANER: false,
          ON_VALUE: "AS_IS"
        },
        dim2: {
          NAME: "ККС",
          SRC_COLUMN_NAME: "ККС",
          CAN_USE_IN_CLEANER: false,
          ON_VALUE: "AS_IS"
        }
      }
    }
  },
  CUSTOM_COLUMNS: [],
  SRC_COLUMN_PREPARE_DATA_MAP: {}
},
CLEANER: {
  STATUS: false,
  FORMAT_FORMULA_MAP: {
    "No Data": null,
    "Date": `DATE("")`,
    "Entity": `""`,
    "Time Entity": `""`,
    "Version": `""`,
    "Line Item Subset": `""`,
    "Number": `0`,
    "Boolean": `FALSE`,
    "Text": `""`
  }
}
};

om.common.resultInfo()
.actionsInfo()
.makeMacrosAction(ENV.CORE)
.appendAfter()
.environmentInfo()
.set(ENV, ENV);

```

Рисунок 40 – Настройка скрипта переноса данных

Одной из особенностей системы является возможность создания автоматизированных и регулярно обновляющихся, в зависимости от поступления данных, дашбордов. Дашборд представляет собой композицию из графиков и таблиц, публикуемых из мультикубов модели [3].

Особенностью дашбордов системы состоит в том, что ввод и редактирование данных в карточках мультикубов и справочников доступны

прямо на дашборде, а все внесенные изменения сразу отображаются в соответствующих источниках.

Таким образом, для более наглядной визуализации управления дебиторской задолженностью, было принято решение построить дашборд, который будут отражать динамику изменения ситуации, а именно: аналитику по группе потребления и рейтингу платежной дисциплины потребителя. Блок дашборда, в котором содержится информация по рейтингу платежной дисциплины потребителя представлена на рис. 41.

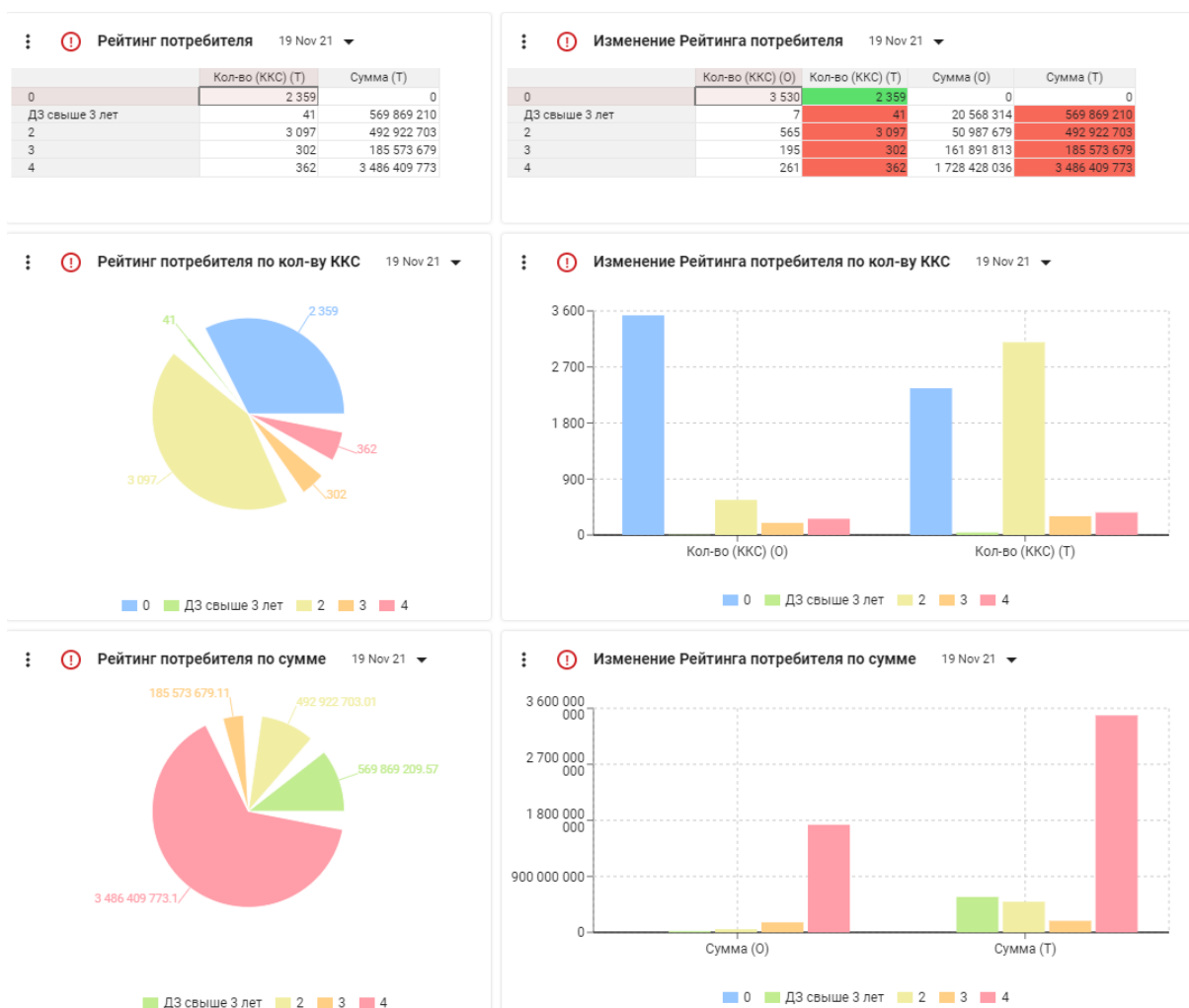


Рисунок 41 – Аналитика по рейтингу потребителя

На рисунке можно увидеть, что дашборд содержит информацию о количестве потребителей по каждому рейтингу, а также сумму дебиторской задолженности этих потребителей. Также, есть возможность отследить динамику изменения количества и суммы по сравнению с отчетной датой.

Для визуализации также были добавлены графики, построенные на основе сводных таблиц данных. Так, можно увидеть, что наибольшее количество потребителей имеет рейтинг платежной дисциплины равный двум. Данный рейтинг является довольно стандартным. При подобном рейтинге, потребители являются вполне добросовестными плательщиками. Однако, стоит обратить внимание, что по сравнению с отчетной датой, количество потребителей и сумма задолженности по данному рейтингу выросла почти в 5 и 10 раз соответственно. В данном случае, подобное изменение, очевидно, связано с разницей в количестве потребителей за текущую и отчетную даты (рис. 42).

	31 Oct 21	19 Nov 21
Кол-во (ККС)	4 558	6 161
Сумма	1 961 875 841	4 734 775 365

Рисунок 42 – Количество контрагентом и сумма ДЗ

Говоря о сумме дебиторской задолженности, можно сказать, что наибольшая сумма приходится на потребителей с рейтингом платежной дисциплины равным четырем. Подобный рейтинг свидетельствует о большой давности периода образования долга, который может являться как просроченным, так и безнадежным. Кроме того, благодаря графической визуализации данных также можно легко определить, по какому из рейтингов платежной дисциплины образовалась большая часть дебиторской задолженности за период и предпринять соответствующие меры для взыскания долга.

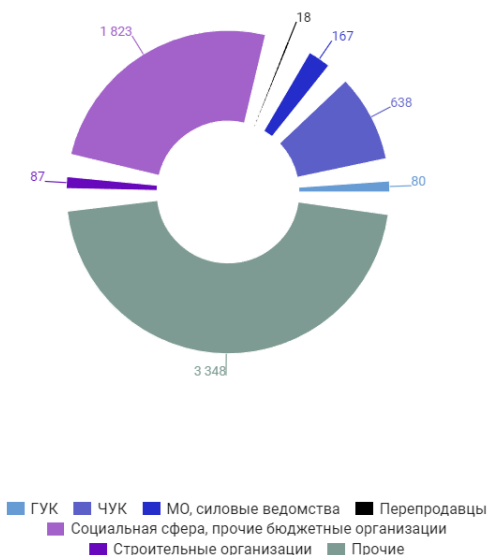
Также, хотелось бы отметить, что перечень мер, описываемый ранее, представляет собой действия исключительно по досудебной работе.

Блок дашборда, в котором содержится информация по группе потребления представлен на рис. 43. На дашборте представлена информация по группе потребления, необходимая для мониторинга количества потребителей, относящихся к конкретной группе.

⋮ ⓘ **Изменения в группах потребления** 19 Nov 21 ▾

	Кол-во ККС (O)	Кол-во (ККС) (T)	Сумма (O)	Сумма (T)
ГУК	70	80	988 597 773	1 386 170 642
ЧУК	556	638	727 287 814	2 408 550 543
МО, силовые ведомства	71	167	28 896 813	38 390 398
Перепродавцы	18	18	102 054 825	143 282 111
Социальная сфера, прочие бюджетные организации	866	1 823	29 818 847	196 061 592
Строительные организации	6	87	198 436	4 540 788
Прочие	2 971	3 348	85 021 333	557 779 292

⋮ ⓘ **Группы потребления по кол-ву ККС** 19 Nov 21 ▾



⋮ ⓘ **Группы потребления по сумме** 19 Nov 21 ▾

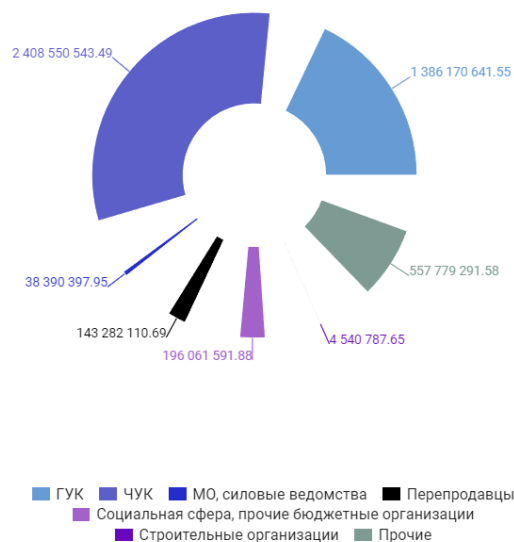


Рисунок 43 – Аналитика по группе потребления

Подобная визуализация данных является достаточно эффективной не только для оценки управления дебиторской задолженностью компании, но и для другой статистической информации. Дашборд дает возможность в считанные минуты ознакомиться с ситуацией и определить проблемные моменты, с которыми в дальнейшем необходимо работать. Кроме того, подобное представление данных о деятельности компании является достаточно информативным для руководства, а автоматическое обновление является дополнительным бонусом со стороны платформы. Благодаря многоуровневым связям между отдельными отчетами компании, можно существенно облегчить ручной труд и значительно сэкономить время.

Кроме того, проделанная работа позволяет сформулировать алгоритм построения хранилища данных для OLAP-системы (рис. 44).

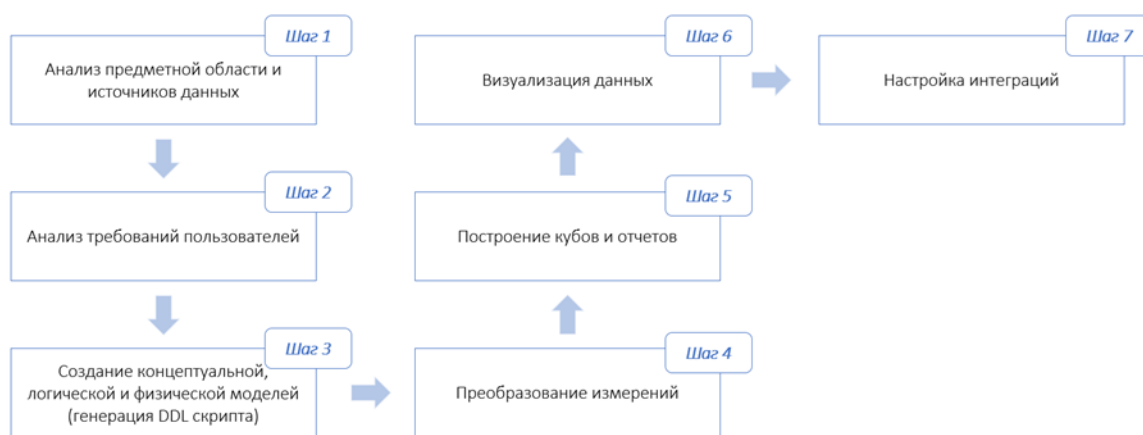


Рисунок 44 – Алгоритм построения ХД для OLAP-системы

Данный алгоритм представляет собой последовательность действий, в виде семи основных шагов проектирования, которые включают как анализ предметной области и основных источников данных, так и построение моделей для последующей настройки DDL скрипта.

Каждый из указанных этапов был проделан для исследуемой в данной работе организации, что позволило удовлетворить всем основным требованиям пользователей новых форм отчета по управлению дебиторской задолженностью, главным из которых было – автоматизация обновления данных, предполагающая снижение количества манипуляций с отчетом посредством ручного труда сотрудников.

Данный алгоритм является универсальным для любого предприятия, вне зависимости от сферы деятельности. Логика, заложенная в последовательность этапов, позволит более детально изучить все основные моменты специфики деятельности по каждому отдельному отчету, а также сконцентрировать информацию в единой информационной системе (в данном случае – системе «Optimascros»), что поспособствует более быстрому доступу пользователей к информации, а также повышению эффективности и обоснованности управленческих решений.

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
ЗНМ04	Глюняевой Анастасии Артуровной

Школа	Инженерного предпринимательства	Отделение (НОЦ)	
Уровень образования	Магистратура	Направление/специальность	27.04.05 Инноватика / Технологическое брокерство

Исходные данные к разделу «Социальная ответственность»:

<p><i>1. Описание организационных условий реализации социальной ответственности</i></p> <p>– заинтересованные стороны (стейкхолдеры) программ социальной ответственности организации, проекта, инновационной разработки, на которых они оказывают воздействие;</p> <p>– стратегические цели организации, проекта, внедрения инновации, которые нуждаются в поддержке социальных программ;</p> <p>– цели текущих программ социальной ответственности организации</p>	<p>Прямые стейкхолдеры организации: сотрудники; акционеры и инвесторы; поставщики и подрядчики; потребители.</p> <p>Косвенные стейкхолдеры организации: государственные органы; местные сообщества и общественные организации.</p> <p>Стратегическая цель ПАО «Газпром» – укрепление статуса лидера среди глобальных энергетических компаний посредством диверсификации рынков сбыта, обеспечения энергетической безопасности и устойчивого развития, роста эффективности деятельности, использования научно-технического потенциала.</p>
<p><i>2. Законодательные и нормативные документы</i></p>	<p>Конституция Российской Федерации; Трудовой кодекс Российской Федерации.</p>

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

<p><i>1. Анализ факторов внутренней социальной ответственности:</i></p> <p>– принципы корпоративной культуры исследуемой организации;</p> <p>– системы организации труда и его безопасности;</p> <p>– развитие человеческих ресурсов через обучающие программы и программы подготовки и повышения квалификации;</p> <p>– системы социальных гарантий организации;</p> <p>– оказание помощи работникам в критических ситуациях.</p>	<p>1. Безопасность труда.</p> <p>2. Стабильность заработной платы.</p> <p>3. Развитие человеческих ресурсов через обучающие программы и программы подготовки и переподготовки кадров.</p> <p>4. Оказание помощи работникам в критических ситуациях.</p>
<p><i>2. Анализ факторов внешней социальной ответственности:</i></p> <p>– содействие охране окружающей среды;</p> <p>– взаимодействие с местным сообществом и местной властью;</p> <p>– спонсорство и корпоративная благотворительность;</p> <p>– влияние разработки, проекта, инновации на стейкхолдеров</p> <p>– влияние разработки, проекта, инновации на окружающую среду, возможное содействие охране окружающей среды;</p> <p>– ответственность перед потребителями товаров и услуг (выпуск качественных товаров),</p> <p>– готовность участвовать в кризисных ситуациях и т.д.</p>	<p>1. Спонсорство и корпоративная благотворительность.</p> <p>2. Содействие охране окружающей среды.</p> <p>3. Ответственность перед потребителями товаров и услуг (выпуск качественных товаров) и т.д.</p>
<p><i>3. Правовые и организационные вопросы обеспечения социальной ответственности:</i></p> <p>– анализ правовых норм трудового законодательства;</p>	<p>1. Трудовой кодекс РУзГосударственное управление в сфере труда. Контроль и надзор за соблюдением законодательства о</p>

<p>– анализ специальных (характерные для исследуемой области деятельности) правовых и нормативных законодательных актов;</p> <p>– анализ внутренних нормативных документов и регламентов организации в области исследуемой деятельности.</p>	<p>труде. от 12.05.2001г. № 220-П – Ведомости Олий Мажлиса, 2001 г., №5, ст 9, ст 89.)</p> <p>2. Кодекс РУз об административных правонарушениях от 21.12.2009г. № ЗРУ-236 — СЗ РУ, 2009 г., № 52, ст. 551)</p> <p>Нарушение трудового законодательства и иных нормативных правовых актов, содержащих нормы трудового права.</p> <p>3. Правила внутреннего трудового распорядка.</p> <p>4. Правила по обеспечению безопасных условий и охране труда.</p> <p>5. Положение о нормированном рабочем дне</p> <p>6. Положение о порядке хранения персональных данных работника. штатное расписание.</p> <p>7. Трудовой договор сотрудников;</p> <p>8. Должностная инструкция.</p> <p>9. График отпусков;</p> <p>10. Положение о структурном подразделении и т.д.</p>
<p>Перечень графического материала:</p>	
<p>При необходимости представить эскизные графические материалы к расчётному заданию (обязательно для специалистов и магистров)</p>	<p>Таблицы</p>

<p>Дата выдачи задания для раздела по линейному графику</p>	<p>24.04.2022 г.</p>
--	----------------------

Задание выдал:

Руководитель ООП, должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ШИП	Попова С.Н.	к.э.н., доцент		24.04.2022

Консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ШИП	Черепанова Н.В.	к.ф.н., доцент		24.04.2022

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
ЗНМ04	Тлюняева Анастасия Артуровна		24.04.2022

4 Социальная ответственность

Корпоративная социальная ответственность – международная бизнес-практика, которая прочно вошла в корпоративное управление в конце XX века. На сегодняшний день, внедрение мероприятий по КСО становится неотъемлемой частью деятельности любой компании.

Термин «корпоративная социальная ответственность» достаточно часто заменяется такими понятиями, как «социальная ответственность бизнеса», «социальная инициатива предприятия», «корпоративные социальные программы», «социальное партнерство», «корпоративная филантропия», «корпоративная благотворительность» и т.д. [42].

Таким образом, социальная ответственность бизнеса (СОБ) – это ответственность организации за воздействие ее решений и деятельности на общество и окружающую среду через прозрачное и этическое поведение, которое [3, 42]:

- содействует устойчивому развитию, включая здоровье и благосостояние общества;
- учитывает ожидания заинтересованных сторон;
- соответствует применяемому законодательству и согласуется с международными нормами поведения;
- введено во всей организации.

Другими словами, под социальной ответственностью бизнеса (СОБ) понимается добровольный вклад бизнеса в развитие общества в социальной, экономической и экологической сферах.

Экологическая составляющая КСО направлена на решение проблем, которые напрямую связаны с негативным воздействием на окружающую среду. Например, загрязнение водных и земельных ресурсов, атмосферы; истощение природных ресурсов; стремительное изменение климата и т.п. Для решения указанных проблем социально ответственные компании, в рамках

своих стратегий по КСО, разрабатывают экологическую политику, направленную на снижение их негативного воздействия на окружающую среду [9].

Экономическая составляющая КСО предусматривает как увеличение прибыли для компании, так и улучшение благосостояния общества. Например, инвестиции, направленные на развитие устойчивости продукции; эффективное использование трудовых ресурсов; создание устойчивой цепочки поставок; проведение исследований в интересах развития бизнеса; эффективное использование материалов и ресурсов, применяемых для производства конечного продукта; этическое поведение по отношению как к потребителям, так и к конкурентам [9, 42].

Говоря о социальной составляющей КСО, можно сказать, что в большей степени она затрагивает вопросы этики и проявляется в ряде действий, направленных на интересы персонала и местного сообщества. Например, соблюдение прав человека (в отношении как персонала, так и населения); соблюдение правил охраны труда; профессиональное развитие кадров; развитие регионов присутствия (проведении благотворительных и волонтерских акций, повышении уровня занятости среди населения) и т.п.

Основная цель корпоративной социальной ответственности – достижение целей устойчивого развития общества, под которыми понимается удовлетворение потребностей нынешнего поколения, без создания угроз для удовлетворения потребностей будущих поколений. Успешная политика корпоративной социальной ответственности учитывает интересы и способствует повышению качества жизни основных заинтересованных сторон компаний, к которым относятся: сотрудники, акционеры, инвесторы, органы государственной власти, клиенты, деловые партнеры, профессиональные сообщества, общество и т.д. [22].

Корпоративная социальная ответственность регламентируется рядом положений и рекомендаций, например, GRI (Global Reporting Initiative); AA 1000S (AccountAbility 1000 Series); SA 8000 (Social Accountability 8000) и т.д.

Таким образом, развитие корпоративной социальной ответственности (КСО) создает более благоприятные условия для улучшения социального климата в обществе, а также способствует повышению качества жизни населения [22, 54].

Кроме того, социальную ответственность бизнеса (СОБ) можно разделить, как на внутреннюю, так и на внешнюю. К внутренней СОБ относятся [54]:

- безопасность на рабочем месте;
- стабильная, достойная оплата труда;
- гарантия социального и медицинского страхования сотрудников;
- предоставление персоналу возможности дополнительного обучения: программы подготовки, повышения квалификации и т.п.;
- оказание финансовой помощи в экстренных случаях.

К внешней СОБ относятся:

- социальные инвестиции и благотворительность;
- забота об окружающей среде;
- ответственность компании перед потребителями;
- взаимодействие с властью и местными сообществами.

Таким образом, на основании изученного теоретического материала о корпоративной социальной ответственности, необходимо провести анализ эффективности программ КСО предприятия ПАО «Газпром», который содержит несколько основных этапов, а именно:

- 1) определение стейкхолдеров организации;
- 2) определение структуры программ КСО;
- 3) определение затрат на программы КСО;
- 4) оценка эффективности и выработка рекомендаций.

Первым делом, необходимо определить стейкхолдеров данной организации. Стейкхолдеры – заинтересованные лица и институты внутри или вне компании, на которые деятельность организации оказывает как прямое,

так и косвенное влияние. Под прямыми стейкхолдерами понимаются лица, которые взаимодействуют с продуктом напрямую (например, сотрудники компании, потребители и т.д.), а косвенные – те, кто оказывают на него опосредованное влияние (например, государственные органы, общественные организации, конкуренты и т.д.) [54].

Компания ПАО «Газпром» выделяет следующие основные группы стейкхолдеров: акционеры и инвесторы, сотрудники, государственные органы, поставщики, потребители, местные сообщества, общественные организации. Структура стейкхолдеров представлена в табл. 2 [13].

Таблица 2 – Стейкхолдеры организации

Прямые стейкхолдеры	Косвенные стейкхолдеры
Сотрудники	Государственные органы
Акционеры и инвесторы	Местные сообщества и общественные организации
Поставщики и подрядчики	
Потребители	

Таким образом, на основании табл. 2 можно сделать вывод, что перечень прямых стейкхолдеров предприятия больше, чем косвенных. Каждый из перечисленных стейкхолдеров оказывает свое влияние на деятельность организации. Так, сотрудники компании осуществляют работы по производству и распределению пара и горячей воды для потребителей. Поставщики и подрядчики оказывают влияние на деятельность компании посредством поставки технического оборудования и оснащения для строительства тепловых электростанций, монтажа зданий, сооружений и т.д. Акционеры и инвесторы являются ключевыми фигурами для повышения стоимости компании и укрепления ее лидирующей позиции [13].

Говоря о косвенных стейкхолдерах, государственные органы осуществляют регулирования и стимулирование развития нефтегазового комплекса, а местные сообщества и общественные организации контролируют

промышленную и экологическую безопасность деятельности компании, что в совокупности способствует надежному, эффективному и сбалансированному обеспечению потребителей природным газом и другими видами энергоресурсов.

Далее, необходимо рассмотреть структуру программ корпоративной социальной ответственности, которая позволит составить портрет КСО компании ПАО «Газпром». Структура программ КСО исследуемой организации представлена в табл. 3 [13].

Таблица 3 – Структура программ КСО

Наименование мероприятия	Элемент	Стейкхолдеры	Сроки реализации мероприятия	Ожидаемый результат от реализации мероприятия
Добровольное медицинское страхование	Эквивалентное финансирование	Сотрудники	Ежегодно	Страхование от несчастных случаев и болезней
Развитие массового спорта и здорового образа жизни	Социально-ответственное поведение	Сотрудники	Ежегодно	Популяризация здорового образа жизни
Курсы повышения квалификации и переподготовки	Социальные инвестиции	Сотрудники	Ежегодно	Повышение квалификации сотрудников
Проведение форумов по вопросам КСО и результатов деятельности компании	Социально значимый маркетинг	Поставщики и подрядчики	Ежегодно	Улучшение имиджа предприятия
Проведение презентаций о компании и ее деятельности	Социально значимый маркетинг	Акционеры и инвесторы	Ежеквартально	Улучшение имиджа предприятия
Проведение информационных форумов и бизнес игр	Социально-ответственное поведение	Потребители (общество)	Ежегодно	Лояльные клиенты / общество

Согласно табл. 3 можно сделать вывод, что в структуре программ КСО организация имеет большое количество мероприятий для сотрудников.

Так, работникам компании предоставляется добровольное медицинское страхование, в рамках которого они получают квалифицированную медицинскую помощь высококлассных специалистов профиля. Также, компания делает большой уклон на физкультурно-оздоровительную деятельность, которая предусматривает компенсацию расходов работников и членов их семей на посещение спортивных сооружений, секций и кружков учреждений культуры и спорта. Ежегодно в компании проводятся спартакиады по различным видам спорта, в которых принимают участие лучшие спортсмены.

Также, компания предоставляет своим сотрудникам возможность прохождения курсов повышения квалификации и переподготовки, т.к. считает, что высокий профессионализм и эффективная работа каждого сотрудника являются гарантом успеха и достижения поставленных целей.

Кроме того, компания регулярно проводит различные форумы и презентации для акционеров, инвесторов, поставщиков и подрядчиков, на которых подробно рассказывает о основных результатах деятельности компании, ее целях и миссии, а также о влиянии КСО на экономическую эффективность [13].

Так, сотрудники ожидают устойчивой системы оплаты труда, социальной поддержки, безопасных условий труда и возможностей как профессионального, так и карьерного роста. Поставщики и подрядчики заинтересованы в устойчивости, надежности и платежеспособности компании, соблюдение взаимных обязательств и благоприятных условиях взаимодействия. Акционеры и инвесторы ожидают от компании возрастание капитала, получение дивидендов и участие в принятии решений. Потребители заинтересованы в качестве, безопасности и доступности продукта, реализуемого компанией.

Таким образом, можно сказать, что мероприятия корпоративной социальной ответственности, реализуемые предприятием, соответствуют ожиданиям ее внутренних стейкхолдеров.

Следующим этапом для анализа является определение затрат на программы корпоративной социальной ответственности, которые были изложены в таб. 2. Стоимость реализации будет рассчитываться исходя из планируемого периода, суммы затрат по каждому мероприятию и среднего количества сотрудников организации. Таким образом, затраты на мероприятия КСО компании ПАО «Газпром» представлены в таб. 4 [13].

Таблица 4 – Затраты на мероприятия КСО

№	Мероприятие	Единица измерения	Период	Стоимость реализации на планируемый период
1	Добровольное медицинское страхование	руб.	год	400 000 000
2	Развитие массового спорта и здорового образа жизни	руб.	год	30 000 000
3	Курсы повышения квалификации и переподготовки	руб.	год	300 000 000
4	Проведение форумов по вопросам КСО и результатов деятельности компании	руб.	год	80 000
5	Проведение презентаций о компании и ее деятельности	руб.	квартал	250 000
6	Проведение информационных форумов и бизнес игр	руб.	год	120 000
	ИТОГО:			700 480 000

На основании табл. 4 можно сделать вывод, что все указанные мероприятия имеют в качестве единиц измерения – рубли. Стоимость реализации мероприятий по добровольному медицинскому страхованию, развитию массового спорта, здорового образа жизни, а также курсов повышения квалификации рассчитана исходя из минимального

среднестатистического количества сотрудников, желающих воспользоваться данными привилегиями от предприятия за 2021 год. Так, в связи с тем, что с ежегодно количество работников компании растет, данная сумма может существенно измениться в следующем году.

Кроме того, т.к. за 2021 год компания ПАО «Газпром» получила порядка 7,5 млрд. руб. чистой прибыли, расчетные затраты на мероприятия по КСО составляют 9% от указанной суммы. Таким образом, можно сказать, что соотношение затрат на мероприятия является вполне оптимальным. При этом, эффект от указанных мероприятий значим не только для общества и стейкхолдеров, но и для самого предприятия [13].

Таким образом, на основании проведенного исследования корпоративной социальной ответственности компании ПАО «Газпром» можно сделать несколько основных выводов:

1) программы КСО предприятия соответствуют основной стратегической цели компании – укрепление статуса лидера среди глобальных энергетических компаний посредством диверсификации рынков сбыта, обеспечения энергетической безопасности и устойчивого развития, роста эффективности деятельности сотрудников, а также повышение лояльности со стороны основных деловых партнеров и потребителей компании.

2) Преобладающей является внутренняя корпоративная социальная ответственность предприятия, т.к. компания стремится обеспечить своих сотрудников возможностью самореализации и карьерного роста, организовать эффективную систему охраны труда, а также создать благоприятный рабочий климатом и системы поощрений, которые позволят замотивировать сотрудников на эффективное достижение основных стратегических целей компании.

3) Программы корпоративной социальной ответственности отвечают интересам каждого отдельного стейкхолдера, что способствует повышению лояльности со стороны каждого из них и соответствует стратегической цели компании.

4) Основными преимуществами при реализации программ корпоративной социальной ответственности для компании являются:

- профессиональное развитие кадров при помощи курсов повышения квалификации и переподготовки;
- повышение заинтересованности и вовлеченности сотрудников в достижении основных стратегических целей компании;
- эффективное использование материалов и ресурсов, применяемых для производства конечного продукта;
- создание устойчивой цепочки поставок;
- инвестиции, направленные на развитие устойчивости продукции;
- развитие регионов присутствия при помощи проведения форумов, деловых игр и т.д.

5) Затраты на реализацию программ корпоративной социальной ответственности являются вполне оптимальными, т.к. компания ПАО «Газпром» занимает первое место в рейтинге RAEX-600 (рейтинг 600 крупнейших компаний в России) в списках: 20 наиболее прибыльных компаний, а также 10 крупнейших компаний в нефтяной и нефтегазовой промышленности 2021 года. Таким образом, компания располагает достаточным бюджетом для реализации рассмотренных в работе программ.

Для совершенствования существующих программ корпоративной социальной ответственности компании было сформулировано несколько рекомендаций, а именно:

- ежегодное проведение тимбилдингов для улучшения взаимоотношений сотрудников компании и благоприятного рабочего климата в различных отделах организации;
- осуществление психологических тренингов для выявления каких-либо проблемных моментов во взаимодействии сотрудников или проблем личного характера, которые могут сказываться на качестве работы;

– проведение аудита по соблюдению социальной ответственности поставщиками и подрядчиками посредством тестирования/инструктажа на тему социальной ответственности бизнеса.

Таким образом, в настоящее время тема социальной ответственности бизнеса приобретает большое значение, поскольку в современном обществе считается безответственным заботиться только о «зарабатывании денег» и уклоняться от ответственности перед своими стейкхолдерами. Социальная ответственность бизнеса действительно является важным моментом, поскольку, в случае безответственности со стороны компаний, возникает риск не только потери клиентов, ценных сотрудников, деловых партнёров, а также потеря поддержки со стороны общества. Таким образом, одной из целей социальной ответственности компании ПАО «Газпром» становится повышения лояльности со стороны разных групп стейкхолдеров и, как следствие, финансовой результативности.

Кроме того, ведущиеся в научном мире дискуссии по корпоративной социальной ответственности способствуют появлению и развитию масштабных программ и инициатив, способных перевести целые регионы на новый уровень устойчивого развития и применения эффективных современных методов ведения социально и экологически ответственного бизнеса.

Заключение

В современном обществе вопрос относительно более эффективного хранения и анализа данных имеет большое значение. Превращение данных в структурированную информацию, описанную в экономических терминах, можно считать одной из самых актуальных задач, стоящих перед компаниями, поскольку для принятия обоснованных управленческих решений, организациям необходимо регулярно отслеживать эффективность своей деятельности. Кроме того, для любой компании, вне зависимости от сферы, часто приходится работать с данными из разных источников. В связи с чем, проектирование хранилища данных является уникальным инструментом, позволяющим оперативно получать данные, обеспечивая различные отделы организации необходимой информацией.

Целью данной работы являлось проектирование хранилища данных для OLAP системы в целях поддержки принятия решений в организации.

Для выполнения данной цели были изучены теоретические особенности хранилищ данных и OLAP технологий, а также проведен анализ построения хранилища данные на предприятии.

Таким образом, было установлено, что основным предназначением хранилищ данных является – предоставление пользователям информации для статистического анализа, а также обеспечение высокой скорости получения данных, возможность получения и сравнения срезов данных, непротиворечивость, полнота и достоверность данных.

Также, в процессе рассмотрения актуальных тем в современной индустрии информационных технологий были выявлены системы аналитической обработки данных – OLAP-системы, представляющие собой технологию комплексного многомерного представления данных. На основе теоретической информации об OLAP технологиях были построены многомерные модели, иллюстрирующие наглядность и информативность структурирования данных в многомерных таблицах. Реализация концепции

оперативной аналитической обработки информации является наиболее оптимальным путем для эффективного управления информационными процессами на предприятии.

В рамках проведенного исследования было построено хранилище данных, а также выполнено наполнение данными для применения OLAP технологий в системе планирования и аналитики – «Optimacros». Проведенное исследование позволяет сделать следующие основные теоретические и практические выводы.

Теоретические результаты:

- 1) проанализированы понятия хранилищ данных и OLAP технологий;
- 2) исследованы основные подходы к проектированию хранилищ данных (корпоративная информационная фабрика (CIF), хранилище данных с архитектурой шины (BUS));
- 3) исследованы основные шаги проектирования (концептуальный, логический, физический, DDL скрипт).

Практические результаты:

- 1) в ходе рассмотрения основных платформ для планирования и аналитики исследуемой организации, было выявлено, что большая часть отчетов и форм компании формируется и анализируется в MS Excel, что является не совсем удобным с точки зрения автоматизации обновления данных и говорит о том, что в компании подавляющее большинство действий с отчетами осуществляется посредством ручного труда сотрудников;
- 2) в ходе проектирования хранилища данных для выбранной компании были выявлены основные сущности, связи и атрибуты для процесса управления дебиторской задолженностью;
- 3) в ходе моделирования в системе «Optimacros» были созданы измерения с различной консолидацией по уровню иерархии, а также построены аналитические кубы отчета с визуализацией, необходимой для эффективного анализа дебиторской задолженности;

4) в качестве заключительного этапа работы в системе, был построен дашборд, на котором представлена визуализация статистической информации о состоянии дебиторской задолженности предприятия в виде сводных таблиц, графиков и диаграмм;

5) на основании проделанной работы по проектированию хранилища данных и преобразований измерений в системе «Optimacros» был разработан алгоритм построения хранилища данных для OLAP-системы, который является универсальным для любого предприятия, вне зависимости от сферы деятельности.

Таким образом, разработанная форма отчета позволит оперативно получать информацию о состоянии дебиторской задолженности контрагентов перед компанией, объединяя сведения из разных источников в одной реляционной базе данных. Созданные формы позволят быстро и эффективно осуществлять поиск, обновление и анализ данных. Удобный интерфейс платформы позволяет легко ориентироваться в спроектированной модели, не требуя наличия каких-либо специальных навыков программирования.

Информационные технологии в области хранилищ данных стабильно развиваются и дальнейшие исследования будут проводиться с учетом следующих основных тенденций:

- тенденция к отказу от традиционной нормализации реляционных отношений, нарушающих естественные иерархические связи между объектами и их атрибутами в конкретной предметной области;
- организация хранилищ на основе многомерного представления;
- реализация концепции оперативной аналитической обработки информации (OLAP).

Таким образом, хранилища данных выступают в качестве основы современных информационных систем. Накапливаемая в них информация представляет собой ценный материал, в результате чего эффективные методы ее хранения и обработки необходимы для получения новых знаний.

Список использованных источников

1. Бюджетный кодекс Российской Федерации [Электронный ресурс]: федер. закон от 31.07.1998 г. № 145-ФЗ (ред. от 16.04.2022) // «Консультант Плюс»: справочная правовая система. – URL: <http://www.consultant.ru> (дата обращения: 11.01.2022)
2. Optimacros [Электронный ресурс] // Официальный сайт. – URL: <https://optimacros.ru/> (дата обращения: 07.03.2022). – Текст: электронный.
3. Аникеева О.П. Управление социальной ответственностью// В мире научных открытий. Красноярск: Научно-инновационный центр. – 2018, № 3 (15). – С. 180-184.
4. Бабич, Т. Н. Планирование на предприятии: Учебник / Т. Н. Бабич, Ю. В. Вертакова. – М.: КноРус, 2018. – 799 с.
5. Баринов, В. А. Бизнес-планирование: Учебное пособие / В. А. Баринов. – М.: Форум, 2018. – 144 с.
6. Барсегян, А. А. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / М. С. Куприянов, В. В. Степаненко, И. И. Холод. – СПб.: БХВ-Петербург, 2007. – 384 с.
7. Барсегян, А. А. Методы и модели анализа данных: OLAP и Data Mining: Учебное пособие / А. А. Барсегян, М. С. Куприянов, В. В. Степаненко, И. И. Холод. – СПб.: БХВ-Петербург, 2004. – 336 с.
8. Басовский, Л. Е. Финансовый менеджмент: Учебник /Л.Е. Басовский. – М.: Риор, 2019. – 589 с.
9. Божук С.Г. Корпоративная социальная ответственность. Учебник для академического бакалавриата / С.Г. Божук. – М.: Юрайт, 2019. – 226 с.
10. Вавренюк, А. Б. Особенности проектирования и разновидности хранилищ данных / А. Б. Вавренюк. // Молодой ученый. – 2018. – № 37 (223). – С. 8-11.
11. Варфоломеева, А.О. Информационные системы предприятия: Учебное пособие / А.О. Варфоломеева. – М.: ИНФРА-М, 2017. – 283 с.

12. Волоха, А. В. Microsoft SQL Server 2005: Учебник / А. В. Волоха. – Санкт-Петербург: Питер, 2006. – 304с.
13. Газпром [Электронный ресурс] // Официальный сайт. – URL: <https://www.gazprom.ru/> (дата обращения: 12.05.2022). – Текст: электронный.
14. Дадян, Э. Г. Данные: хранение и обработка: Учебник / Э. Г. Дадян. – М.: ИНФРА-М, 2019. – 236 с.
15. Дадян, Э. Г. Современные базы данных. Основы. Часть 1: Учебное пособие / Э. Г. Дадян. – М.: ИНФРА-М, 2017. – 88 с.
16. Джабраилов, Ш. В. Сравнительный анализ методологий проектирования хранилищ данных / Ш. В. Джабраилов, Ю. А. Орлова, Ш. Ш. Камбарова. // Молодой ученый. – 2020. – № 20 (310). – С. 27-30.
17. Домрачева, А. А. Business Intelligence в экономике / А. А. Домрачева, Н. Ю. Сайбель. // Научно-методический электронный журнал Концепт. – 2017. – № 2. – С. 41-46.
18. Зыкин, С. В. Реализация OLAP-технологии на основе межмодельных преобразований данных // Информационные технологии моделирования и управления. – 2008. – № 1 (44). – С. 71-77.
19. Каримова, Н. О. Компьютерная идентификация и современные системы интеллектуальной обработки данных / Н. О. Каримова. // Молодой ученый. – 2017. – № 4 (138). – С. 22-24.
20. Каширин И. Ю. Интерактивная аналитическая обработка данных в современных OLAP-системах / И. Ю. Каширин, С. Ю. Семченков. // Журнал «Бизнес-информатика». – 2009. – № 8(2). – С. 12-19.
21. Клычова, Г. С. Дебиторская задолженность: сущность и определение / Г. С. Клычова, З. З. Хамидуллин // Бухгалтерский учет в бюджетных и некоммерческих организациях. – 2019. – № 16 (400). – С. 40-45
22. Коваленко Ю.С. Корпоративная социальная ответственность и ее влияние на деятельность фирмы // Молодой ученый. – 2019. – № 22 (260). – С. 546-548.

23. Козлов, А.Ю. Статистический анализ данных в MS Excel: Учебное пособие / А.Ю. Козлов, В.С. Мхитарян, В.Ф. Шишов. – М.: ИНФРА-М, 2017. – 320 с.
24. Коновалов, М. В. Обзор и сравнительный анализ промышленных хранилищ данных и баз данных / М. В. Коновалов. // Молодой ученый. – 2018. – № 24 (210). – С. 24-28.
25. Коровкин, С. Д. Решение проблемы комплексного оперативного анализа информации хранилищ данных / С. Д. Коровкин, И. А. Левенец, И. Д. Ратманова, В. А. Старых, Л. В. Щавелёв. // СУБД. – 1997. – № 5-6. – С. 47-51.
26. Красина, Ф.А. Финансовый менеджмент: учебное пособие / Ф.А. Красина. – Томск, 2017. – 200 с.
27. Кукарцев, В. В. Теория баз данных: учебник / В. В. Кукарцев, Р. Ю. Царев, О. А. Антамошкин. – Красноярск: Сибирский федеральный университет, 2017. – 180 с.
28. Лесковец, Ю. Анализ больших наборов данных / Ю. Лесковец, А. Раджараман. – М.: ДМК, 2016. – 498 с.
29. Лещаков, И. Н. Архитектура информационной системы предприятий / И. Н. Лещаков. // Молодой ученый. – 2021. – № 21 (155). – С. 13-16.
30. Макшанов, А. В. Системы поддержки принятия решений: учебное пособие для вузов / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. – Санкт-Петербург: Лань, 2020. – 108 с.
31. Макшанов, А. В. Технологии интеллектуального анализа данных: учебное пособие / А. В. Макшанов, А. Е. Журавлев. – Санкт-Петербург: Лань, 2022. – 212 с.
32. Мартишин, С. А. Проектирование и реализация баз данных в СУБД MySQL с использованием MySQL Workbench: Методы и средства проектирования информационных систем и техноло / С. А. Мартишин, В. Л. Симонов, М. В. Храпченко. – М.: Форум, 2018. – 61 с.
33. Мюллер, Р. Дж. Базы данных и UML. Проектирование / Р. Дж. Мюллер. – М.: ЛОРИ, 2019. – 420 с.

34. Немуров, Е. В. Актуальность внедрения ВІ систем на предприятиях в условиях современного рынка // Международный научно-технический журнал «Теория. Практика. Инновации». – 2018. – №2(26). – С. 15-22.
35. Нестеров, С. А. Основы интеллектуального анализа данных: учебное пособие / С. А. Нестеров. – Санкт-Петербург: Лань, 2020. – 40 с.
36. Орлов, П. А. Сравнительный анализ эффективности использования современных облачных хранилищ / П. А. Орлов. // Молодой ученый. – 2017. – № 20 (154). – С. 154-156.
37. Парфенов, Ю. П. Постреляционные хранилища данных: Учебное пособие / Ю. П. Парфенов. – М.: Юрайт, 2018. – 121 с.
38. Самойлова, И. А. Введение в ВІ-технологии / И. А. Самойлова. // Молодой ученый. – 2015. – № 3 (83). – С. 78-80.
39. Советов, Б. Я. Базы данных: учебник для СПО / Б. Я. Советов, В. В. Цехановский, В. Д. Чертовской. – М.: Юрайт, 2018. – 420 с.
40. Стружкин, Н. П. Базы данных: проектирование. практикум: Учебное пособие для академического бакалавриата / Н. П. Стружкин, В. В. Годин. – Люберцы: Юрайт, 2016. – 291 с.
41. Тартынский, В. А. Реализация хранилищ данных в системах поддержки принятия решений / В. А. Тартынский. // Молодой ученый. – 2009. – № 9 (9). – С. 31-34.
42. Тульчинский Г.Л. Корпоративная социальная ответственность. Технологии и оценка эффективности. Учебник и практикум / Г.Л. Тульчинский. – М.: Юрайт, 2019. – 338 с.
43. Турсынбаева, С. Ж. Особенности построения хранилищ данных в финансовых организациях / С. Ж. Турсынбаева. // Молодой ученый. – 2021. – № 16 (358). – С. 14-19.
44. Федорова, Г. Н. Информационные системы: Учебник для студ. учреждений сред. проф. образования / Г. Н. Федорова. – М.: ИЦ Академия, 2016. – 208 с.

45. Фуфаев, Д. Э. Разработка и эксплуатация автоматизированных информационных систем: учебник / Д. Э. Фуфаев. – М.: Academia, 2017. – 352 с.
46. Хаббард Дж. Автоматизированное проектирование баз данных / Дж. Хаббард. – М.: Мир, 2007. – 294 с.
47. Хансен Г. Базы данных и управление / Г. Хансен, Д. Хансен. – М.: Бином, 1999. – 699 с.
48. Харрингтон Дж. Л. Проектирование реляционных баз данных / Дж. Л. Харрингтон. – М.: Лори, 2000. – 230 с.
49. Хомоненко А. Д. Базы данных / А. Д. Хомоненко, В. М. Цыганков, М. Г. Мальцев. – СПб.: КОРОНА, 2004. – 736 с.
50. Цикритизис Д. Модели данных / Д. Цикритизис, Ф. Лоховски. – М.: Финансы и статистика, 1985. – 344 с.
51. Черняев, А. С. ETL: обзор инструментов / А. С. Черняев, М. А. Балова. // Молодой ученый. – 2019. – № 1 (239). – С. 23-26.
52. Чубукова И. А. Data Mining: учебное пособие / И. А. Чубукова. – М.: БИНОМ, 2006. – 382 с.
53. Чумакова, М. В. Современное состояние процессов планирования на предприятии / М. В. Чумакова. // Молодой ученый. – 2021. – № 20 (124). – С. 451-457.
54. Эскиндаров М.А. Корпоративная социальная ответственность. Учебник для бакалавриата / М.А. Эскиндаров. – М.: КноРус, 2016. – 38 с.

Приложение А

Раздел ВКР, выполненный на иностранном языке

Collection and analysis of requirements for the design of data warehouses

Студент:

Группа	ФИО	Подпись	Дата
ЗНМ04	Тлюняева А.А.		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ШИП	Юдахина О.Б.	к.э.н., доцент		

Консультант – лингвист ОИЯ ШБИП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИЯ	Надеина Л.В.	к.филол.н.		

2 Collection and analysis of requirements for the design of data warehouses

2.1 Description of the company's activities and analysis of accounts receivable

A subsidiary of Gazprom was chosen as the object of the study of this work. The main activity is the production of steam and hot water (heat energy). The form of legal-organization is a public joint stock company.

The company includes additional activities [13]:

- 1) the electricity generation by thermal power plants;
- 2) the electricity generation by other power plants and industrial block stations;
- 3) transmission and distribution of the electricity;
- 4) activities to providing the operability of thermal power plants and electrical networks;
- 5) production and distribution of steam and hot water (heat energy) by thermal power plants;
- 6) preparing for the building site;
- 7) works of the building thermal and other power plants;
- 8) installation of buildings, structures made by prefabricated elements and etc.

The company provides centralized heating and hot water supply of its own heat supply sources and other thermal generation facilities. Except of small local heating areas from isolated departmental and corporate thermal sources. The company has been operating since 2004. It provides services to legal and physical entities.

The company is the provider of the longest thermal power system in the world. It operates over 17,000 km of heating networks, including about 8,600 km of trunking heat and 8,433 km of distribution heat.

Currently, the company operates and maintains services to the following thermal power facilities [13]:

- 10,611 heat offices;
- 24 pumping stations;
- 141 sources of thermal energy generation, including 45 district and quarter heating stations (BHS and CH), 96 small boiler houses (SB, AHS, ПК) with a thermal power of 4996.014 Gcal/h;
- 17,034 km of heating networks (in single-pipe terms), including 8,600.06 km of trunking heat (including heat inputs) and 8,433.94 km of heat distribution networks.

At present, the company employs over 14000 persons. The company has a capability team with the necessary professional knowledge skills, and abilities to train young employees, familiarize them with advanced achievements and sharing accumulated experience to them. For the company, high professionalism and effective work are the guarantee of reaching goals and success of each employee.

The company has a large number of consumers as physical and legal entities. The consumer bargains with a thermal energy company in which the company complies their obligations and its counterparty pays the price after some time. In other words, there is a payment deferment. Such a time gap leads to the consumer's accounts receivable formation under the investigated organization.

In connection with the fact that such a house can be closed in a couple of days and continue to exist quite a long period of time, competent management of accounts receivable is one of the key moments of the organization to the effective functioning of the company.

In addition, the topic should be taken very seriously because it's a company engaged in the production and distribution of water and thermal energy. It connected with the fact that the value of the accounts receivable of an energy company with an impressive list of consumers, business partners and administrative districts distribution certainly has huge dimensions.

Depending on the relationship with business partners/consumers at the company, accounts receivable is classified as follows [13, 21]:

1) current AR- is a short-term debt in which the repayment period does not pass over 12 months;

2) non-current AR- based on the name, it is a long-term debt, in which repayment period is expected no earlier than in 12 months;

3) overdue AR is a debt in which the repayment period is under the contract has already passed, but the funds have not been transferred by the client;

4) doubtful AR- is a debt in which there is a risk of non-repayment if the debt without the deposit or some other guarantee;

5) uncollectable AR – is a debt in which the repayment of funds is not possible due to failure, liquidation, reorganization of the borrower's organization or the end of the period of limitation.

In this way, it is necessary to analyze the accounts receivable of the investigated organization. Information on the accounts receivable of «Gazprom» for the period 2019 – 2021 is presented in Table. 5.

Table 5

Accounts receivable of PJSC "Gazprom" for 2019 – 2021

Indicator description	As for 31.12.2021	As for 31.12.2020r.	As for 31.12.2019r.
Accounts receivable, including	57 267 323	45 038 343	44 996 990
Accounts receivable (for which payments are expected more than 12 months after the reporting date), including	9 069 570	9 053 781	6 758 206
Buyers and customers	957 796	245 934	31 340
Payment of an advance	15 840	20 058	29 260
Other debtors	8 095 934	8 787 789	6 697 606
Accounts receivable (for which payments are expected during 12 months after the reporting date), including	48 197 753	35 984 562	38 238 784
Buyers and customers	34 503 804	23 023 961	23 875 932
Debt of participants (founders) of contribution to the registered capital	–	–	–
Payment of an advance	748 965	702 288	1 057 435
Other debtors	12 944 984	12 258 313	13 305 417

In this way, based on this table, it can be concluded that, compared with 2019 the amount of the company's accounts receivable increased by 21%. Moreover, most of them accrue to buyers and customers (growth by 97% – long-term, by 30% - short-term). In addition, according to some indicators, accounts receivable has decreased due to analyzed period. For example, in 2021, payments that are expected during 12 months after the reporting date for other debtors are hardly 3% lower.

This may be due to the fact, that compared to the reporting period the list of consumers and business partners has changed.

Every month, the n-th number of new customers come to the company and the old ones are closed. In addition, the share of any previously listed types of debt also could be essentially changed. That way, the distribution by type of accounts receivable of the company for 2021 is completely shown in figure 45 [13].

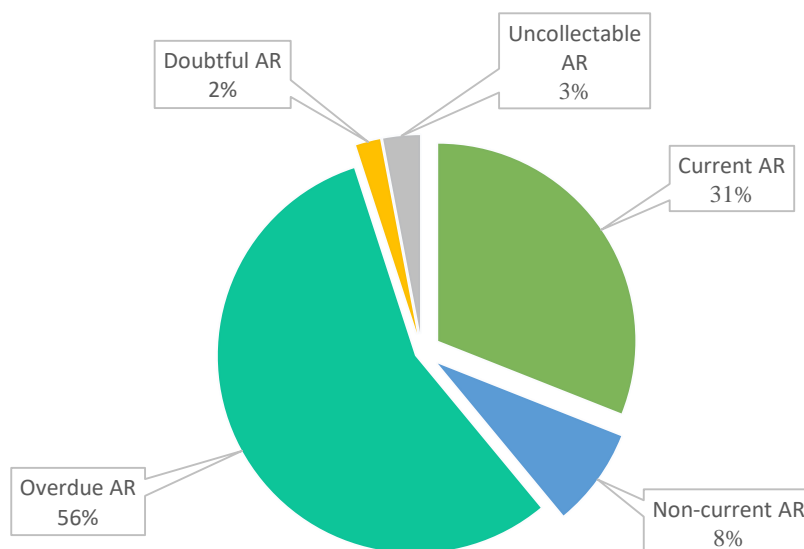


Fig. 45 The structure of the AR of «Gazprom» for 2021

So on the figure 45 can be seen that most of the company's accounts receivable is overdue, i.e., the funds have not been transferred under the time defined in the contract. It stands to mention that in appropriate functioning, most of the debt refers to current short-term debts, which don't have a negative influence to the company and don't prevent its development.

Such a result means that a competent mechanism for managing accounts receivable at the company is needed, which will provide normal paying capacity. To this result, it is necessary to study the list of the most popular measures realized for effective AR management.

As usual, most companies use a complex of measures to manage accounts receivable, which include both preventive and demanding legal measures.

There are several main stages, following [21, 26]:

- 1) previous verification of the partner before the conclusion of the contract;
- 2) special conditions for each specific partner (schedule, deadlines, penalty charges, etc.);
- 3) operating control over the complaint of obligations;
- 4) recovery (if the debtor is not going to repay the debt by own will).

The first three stages help to prevent the appearance of overdue debts. Speaking about the management of existing accounts receivable, in the investigated organization there are some several basic measures which applied in relation to consumers, namely [13]:

- 1) calling consumers;
- 2) meeting with the consumer in the sales functions;
- 3) information message about the debt;
- 4) the direction of the appeal to the Administration / Prefectures / HCS;
- 5) the direction of the appeal to higher-level agency;
- 6) distribution of information about consumer's debt through the media;
- 7) direction to the MHI (declaration);
- 8) direction to the MHI (appeal);
- 9) RCDPC;
- 10) direction to the ITM Rostechnadzor (appeal);
- 11) direction to the Prosecutor's Office (appeal);
- 12) corporate security service;
- 13) direction for the addressee the company's director of the relevant district.

Each of these measures is applied to the consumer consistently. It depends on the rating of the payment discipline. Based on the period of debt formation, this rating is calculated and has a variation: AR over 3 years, 0, 2, 3 and 4.

Moreover, some of the listed measures apply to all consumers, without exceptions, and some – only to counterparties with the highest rating (3 and 4). Predominantly, the rating of payment discipline which equal to four indicates that the debt is overdue or uncollectible, and consumers are the most problematic.

The plan of the sequence of application of these measures, which depend on situation, is shown in figure 46.

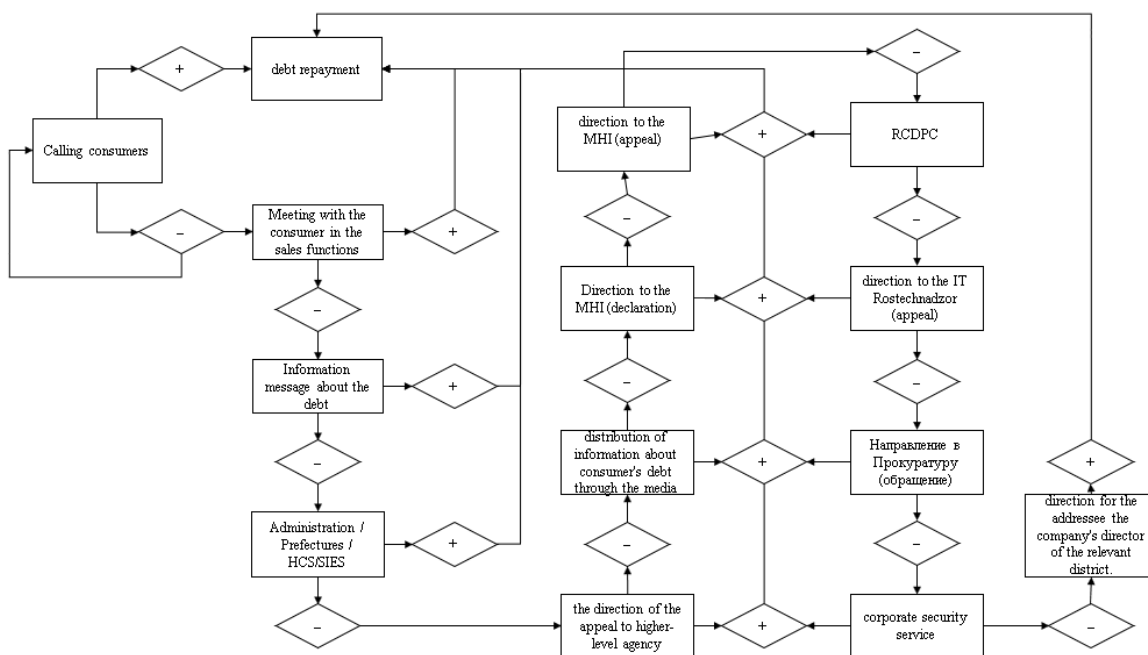


Fig. 46 The process of managing the AR of «Gazprom»

This plan reflects the sequence of completion of each measure depending on the positive and negative results of the previous one.

In this way, if the repayment period of the debt gets close or there is another reason for communication with the client, the company's employees call consumers. As for this event, it doesn't matter what the payment discipline rating of the debtors who was being called.

Notification about the debt is sent to all clients of the company. Then, in occasion of non-fulfillment of obligations, there is a meeting with unfair debtors in

the sales function, there is a meeting with employees who responsible for managing the accounts receivable in the company for a relevant department and district.

If after the first two measures of engagement with the consumer the debt is not repaid, the information message is sent by an employee to remind the client of the existence of the debt. Moreover, it is necessary to fix the date and time of engagement and also leave short comments for each event, if some situation requires it.

The first three measures are standard enough, without a big "threat" to consumers and include exclusively notification measures. Providing that, the classical notifications didn't have an effect, the directions of appeals to the Administration / SIES / Prefectures / HCS, the high-level agency, the Ministry of Internal Affairs, the MHI, RCDPC of Rostechnadzor, the Prosecutor's Office, etc., come into a play. At worst, information about the consumer's debt is distributed through the media.

If none of the existing measures at the company hasn't brought the expected result (repayment of the debt), it leads to direction for the addressee the company's director. Generally, this measure doesn't reach. The client fulfills the debt obligations at the time of applying to the Administration / SIES / Prefecture / HCS. However, such a result is typical for consumers with no more than two of payment discipline rating. For consumers with higher ratings, which don't change for the better during the analysis period, sometimes it is necessary to go through a full list of measures, up to recover and debiting funds.

This process engages not only financiers and accountants of the company. Managers, lawyers and also security officers are engaging in managing and automating overdue debts.

In this way, based on the investigation of accounts receivable management, it can be concluded that the company has a quiet effective mechanism for regulating consumer's debt. In addition, the company's employees have several possible reasons for the existence of accounts receivable from consumers, such as:

- debt recovery in court is inadvisable (amount of debt less than 20 thousand RUB);
- AR caused as a result of the clarification / redistribution of benefits;
- disagree with the charges;
- documents have been prepared and sent for efficiency, for the provision of payment by installments;
- recalculation is planned;
- absence of a complete documentation package for transfer to the court;
- recalculation is planned.

Also, it is interesting to note that the reasons for a large number of overdue debts may be connected with not only to the internal management structure of the company, but also to a number of external factors [5, 21]:

1) financial difficulties – loss of paying capacity by the counterparty as a result from a «cash gap» associated with the crisis. Currently, quite a large number of consumers are in a precarious position;

2) force majeure, natural disasters and other unpredictable destructive phenomena. This also includes the death or disablement of the counterparty;

3) partner's unconscientiousness – some of counterparties don't plan to fulfill the terms of the contract consciously. In this area, this happens incredibly rarely, but you should not completely exclude this possibility. At the same time, it is simple to prevent the appearance of debt – to check potential counterparties qualitatively before concluding a contract.

So, the company has an impressive list of consumers and business partners, which is not so easy to control. In order to provide the recognition of debt obligation for each certain consumer timely, it is necessary to pay attention not only on development the debt recovery's package measures, but also structure and visualise the available data in that form in which to be most effective for analyzing further work.

2.2 Analysis of the company's information infrastructure for the creation of accounts receivable reports

In the previous paragraph, it was revealed that the subsidiary of «Gazprom» has effective mechanism for managing accounts receivable. Then, in order to determine the most effective ways to reduce the amount of overdue debt in the total amount of counterparty's debt to the company, it is necessary to study the basic infrastructure of the company for reporting.

«Gazprom» company has several planning and analytics systems, as follows: MS Excel, SAP Logon and Power BI. The percentage of use of each of the listed systems is shown in figure 47 [13].

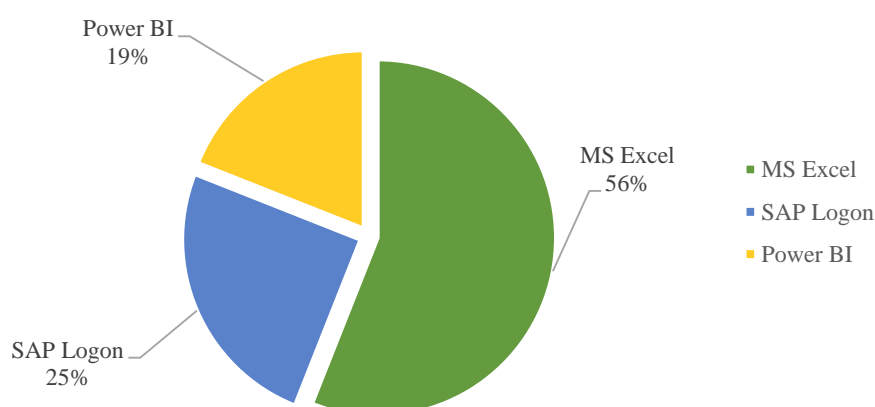


Fig. 47 Planning and analytics systems of PJSC «Gazprom»

The main advantages and disadvantages of these planning and analytics systems were described in the first chapter of this work. Then, it's necessary to describe the main application of each of them for the investigated organization.

Speaking about MS Excel, it is interesting to note that about 56% of reports and forms of the company are generated and analysed in this planning and analytics system. The program has proven itself as one of the most flexible tools for modeling reports in any company. However, working with a large amount of data, there is a high chance that heavy files with a large number of columns, rows and pages will be

loaded for a long period of time. Working in MS Excel include (mostly degree) handwork, which is undoubtedly ineffective in terms of automation and time saving.

Speaking about Power BI, it is interesting to note that this platform is used for the analysis of reports and forms related to funds flow, for example, the funds flow budget, the budget of income and expenses, etc. The platform allows you to visualize, select the most important information and share it by each sales function. Speaking about managing of accounts receivable, it's fair to say that Power BI isn't used to report generation. Two of the previously listed analytics systems (MS Excel and SAP Logon) take part in this process.

Speaking about the SAP Login platform, it's interesting to note that the company SAP decided to suspend work in the Russian Federation for the indefinite period. It was happened because of the actual situation in the world. As the SAP didn't predict the future, we don't still know whether SAP leave Russia forever or resume its work after a while. As a result, a subsidiary of PJSC «Gazprom» decided to refuse from this system because of uncertainty, which may prevent the effective planning and analysis of actual activities.

The formation of a report on the management of accounts receivable occurs on the base of a daily report on consumer advances. The algorithm of receiving the necessary information for the report and its compilation has the following form:

- 1) uploading a daily report on consumer advances in MS Excel format from the SAP system;
- 2) transferring data from the advances report to the MS Excel file template for managing accounts receivable by copying data.

To take into attention the fact that the company is currently refusing the SAP platform, this algorithm will be subjected by significant changes. In fact, most of the company's reports are generated and analyzed in MS Excel. For each report, there are separate template files containing a standard set of columns and calculations. In order to analyze the situation for the new period, employees should pass the data into the file for the previous period and stretch the column's data with formulas for all row values, or the opposite – delete unnecessary rows for

the analysis period. Currently, copying, passing and stretching data has been taken back seat. The handwork of employees has been long ago replaced by planning and analytics systems that involve automating data updates and allow to store all the company's reporting and analytics, and not just some of its part.

This way, the company's management decided to introduce a new analytics platform to the company, which will meet the basic requirements of users and also allow automate the process of updating data in all forms of the company's reports.

2.3 Collecting and analysing user's requirements

Before make proceed to a new planning system, it is necessary to determine a basic user's requirements of the company's accounts receivable management report.

So, during the collection of information through the personal conversation with employees of sales function who responsible for the process of managing accounts receivable, were revealed that 91% of the working staff support the decision to automate the process of generating reports and form of the company by switchover a new platform. This way, it was decided to carry out of a survey, on the base of which it would be possible to determine which points in the current report form are suitable the employees of the departments, which don't suitable and also what could be added / changed in order to improve efficiency. So, figure 48 shows the results of the survey of 45 employees regarding the positive aspects of the current report form and the process of its formation.

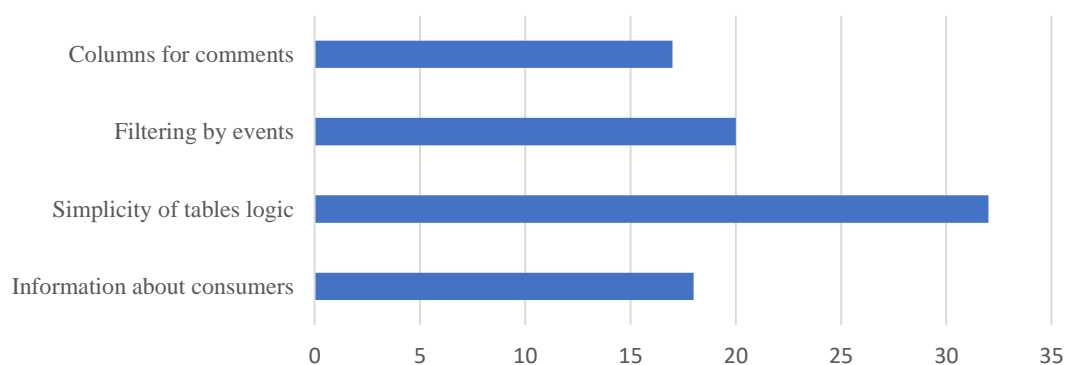


Fig. 48 Positive aspects of the current report form

So, based on figure 48, one can come to the conclusion that the important points of working with the accounts receivable management report are:

- existence of basic information about consumers (name of business partner, contract number, name of administrative district, TIN, consumer group, alternative group, etc.);
- possibility of filtering by event columns (selection of the value required for consideration and analysis);
- presence of columns, in which they can leave comments on the relevant consumer and event. This information helps to understand the situation. For example, another employee of the company worked with the consumer earlier and new employees can track when and how the last interaction with the consumer was ended;
- simplicity of the table's logic. This moment is one of the most significant for users of the report, because all columns are arranged sequentially, depending on the increasing severity of the measures imposed on consumers. Also, these tables have simple and understandable titles, they don't cause difficulties in identification.

Next, it's necessary to determine which moments in working with the report cause various kinds of difficulties for the company's employees. The results of the survey on this topic are presented in figure 49.

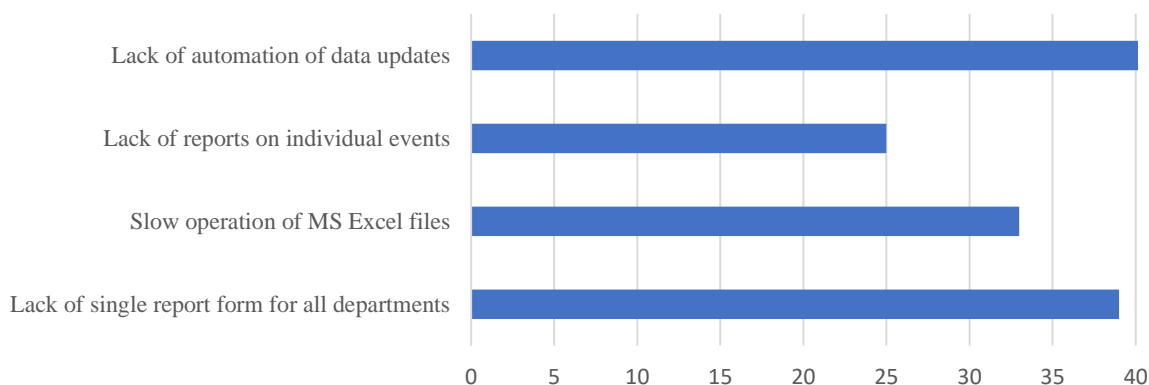


Fig. 49 Negative aspects of the current report form

In figure 49, it is interesting to note that, as already noted earlier, the most of respondents have a negative attitude to manual updating of data and identify this

feature as the weak side of the formation of the accounts receivable management report. In addition, also it was found that working in MS Excel files is accompanied by overload system, because the report contains a large number of rows, columns and some auxiliary pages for calculations, which affect to efficiency of work. Also, according to the results of surveys, it was revealed that the company has a problem which associated with the fact that each separate sales function generates a report on the management of accounts receivable in different formats. In this way, some auxiliary columns may be added in one department while the other does not. As a result, analyses the report containing data for all departments becomes difficult because of the contradiction of presentation formats.

In addition to the difficulties above, 63% of the surveyed employees noted that due to the large number of columns in the table, it becomes difficult to return to the first columns by rolling the slider when analyzing activities at the end of the report. As a result, another disadvantage of the current form is the impossibility to analyze an individual event.

After determining the positive and negative aspects of the report, the company's employees offered their own variants for improving the report, which are presented in figure 50.

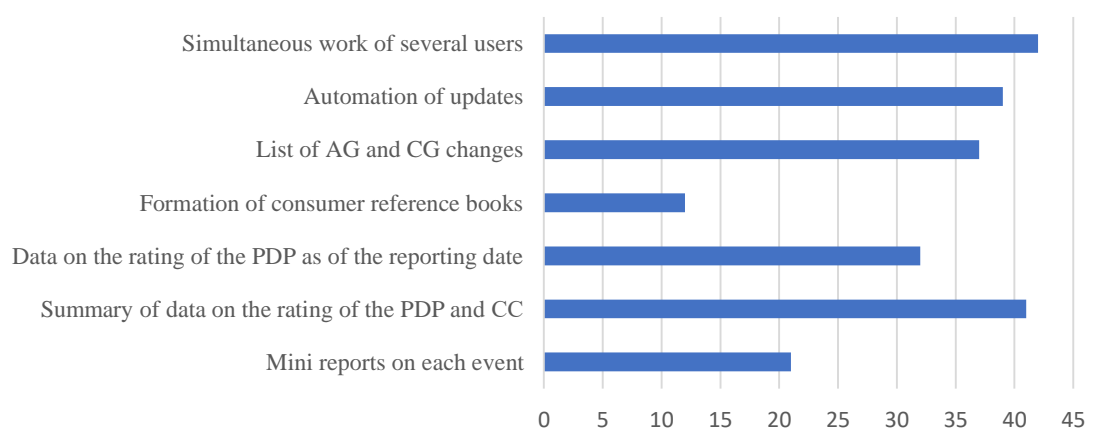


Fig. 50 Suggestions for improving the current report form

In this way, some of the proposed improvements are able to neutralize the negative aspects and some are the introduction of the possibility of point-to-point analysis for various indicators of the company's counterparties.

So, based on figure 50 and the results of the survey, the main requirements of users for the accounts receivable management report in the future system planning and analytics were formulated:

- 1) existence of columns containing all the necessary auxiliary information about consumers;
- 2) existence of an additional column on the rating of the consumer's payment discipline at the reporting date;
- 3) the possibility of filtering by characteristics: administrative district, sales function, municipal district, consumption group, alternative group, affiliation, consumer category, payment discipline rating, total for each of the events;
- 4) possibility of forming a list of changes by: consumers, consumption group, alternative group, consumer affiliation;
- 5) possibility of forming summary tables by: rating of payment discipline and consumption group;
- 6) possibility of analytics for individual events;
- 7) automating of the report's data updates;
- 8) possibility of concurrent work of the several users.

Therefore, based on the formed user's requirements, the company's management decided to introduce in company the «Optimacros» system is a platform for optimization and consolidation planning, business analysis and integrations. This platform contains all the advantages and flexibility that the analytics systems which used in the company. The main advantages of the system were described in the first chapter of this work.

In order to implement this platform, it is necessary to carry out several steps of designing the data warehouse of the investigated organizations: conceptual, logical, physical data model. The achieved result will allow for the interaction of the data warehouse and OLAP technologies underlying the «Optimacros» system, as well as data processing, consolidation and its representation on various reports using visualization tools. In other words, the technology will contribute to the development of tools for creating a decision-making infrastructure.

Data warehouse design is the creation of a denormalized data structure which is firstly focused on process of high productivity the analytical queries execution. A significant advantage of such structure is the simplicity of the warehouse model, which doesn't complicate its understanding and doesn't decline the efficiency of query execution.

There are some basic requirements which are usually presented on design of data warehouses [10, 19]:

- 1) the warehouse data structure should be simple and understandable for users;
- 2) static data that is regularly modified (daily, weekly, etc.) should be highlighted;
- 3) the query requirement should be simplified;
- 4) providing the support for the SQL complex queries, that require consistent processing of thousands or millions of records.

In designing a data warehouse, three models are often built, as follow: conceptual, logical and physical. Let's look at each of them in further detail.

The conceptual model is characterized by the definition of the position between entities. As a rule, such a data model is an "entity-position" diagram, the purpose of which is to identify, formalize and analyses the necessity of an organization in the data aspect.

The conceptual model consists of entities, attributes, positions and it doesn't reflect the physical realization or program constraints features [14, 32].

In the conceptual model can be two types of entities:

- 1) strong (independent) – that existence doesn't depend on other entities;
- 2) weak (dependent) – whose existence depends on other entities that are stronger in a relation to them.

In this way, the same entity can be strong in relation to one and weak in relation to the other one. Characteristics of entities are the attributes and it used for identification. The values of each attribute are selected from a domain – a set that includes all the values which can be assigned to this attribute. Attributes' values can be as a static, i.e., not changing over a time, as dynamic (for example, region).

The logical model is very similar to the conceptual one, however, it also reflects the beginning of the physical representation of data. In the logical model, against of the conceptual one, in which connections can be: "one to many" and "many to many", connections of the second sort are prohibited. Instead of that, the intermediate entities are used. This is due to the fact that the logical data model is relational, and the many-to-many relationships don't accept the idea of relational warehouse [37].

The logical data model is the initial prototype of the future database. It is building in terms of information units, but without reference to a definite DBMS. At the moment a preliminary resource of developing a logical data model are various variants of infological (information-logical) models – ER-diagram (Entity-Relationship, entity-position diagrams).

It is interesting to note that the same ER model can be transformed as into a relational data model, as into a data model for hierarchical and network DBMS, or a post-relational data model.

The logical model reflects logical entity types, data attributes that the entities described, and relationships between it.

This model has the following features [37, 45]:

- includes all entities and relationships between them;
- all attributes should be defined for each entity;
- an initial key should be defined for each entity;
- external keys are determined;
- occurrence of normalization.

Talk about logical data model, by the way, it generated from a conceptual model, also, logical model to design by IDEF1X.

IDEF1X is a method for developing relational databases with conditional syntax specially designed for convenient construction of a conceptual schema. This model is relational and should take into account the chosen approach to building a data warehouse [37].

The purpose of the constructing a logical model is to obtain a graphical representation of the logical structure about the subject area under study. At the logical level we can identify key attributes, which will become initial keys later. The key attributes are located at the top of the entity and they are candidates for initial keys.

The next step is to build a physical data warehouse model from a logical one, which is used for building a database schema for information system.

So, the decisions are made at the level of logical modeling, define some boundaries within, which makes it possible to develop a physical data model. The relationships which contained in the logical data model should be transformed into tables, but for each table we can additionally declare different indexes, that increase the speed of accessing data.

This way, in the physical model, the entities become tables and attributes become columns of these tables. Links are implemented in accordance with the mechanisms used in the database.

The physical model shows the structure of tables and includes field names, field data types, constraints, initial and external keys and relationships between tables.

The main features of the physical data model [37, 41]:

- specification of all tables and fields;
- external keys use to define relationships between tables;
- denormalization with the requirements of users;
- possibility of difference from the logical model associated with limitations at the physical level;
- different models for various DBMS.

The physical model is a graphical representation of the warehouse data structure and allows analyze its tables (attributes, indexes), procedures. All of this is necessary for the correct generation of the DDL script. This stage is the last and it's important to perform it before the creation of the repository [41].

Thus, the main purpose of building a physical data model is to generate a DDL script for creating a database. Such script has the following structure: it checks the tables and indexes. If tables or indexes already exist, they are deleted. Next, tables are created, after which external keys and integrity of constraints are determined. As a rule, the script is formed in terms of the Microsoft SQL Server 2017 DBMS – a platform for critical tasks solutions on a company scale, with high availability, increased performance and security. When all actions with the physical data model are completed, the script can run.

So, the design of data warehouses is a sequence of steps for the implementation of conceptual, logical and physical models. Each subsequent model is formed on the basis of data on the previous one, upgraded and reflects all the main entities, relationships and attributes of databases. The DDL script is the final point for building a data warehouse as a result of the design.

Currently, many companies store and analyze information in Excel, which has proven as an effective tool for working with spreadsheets. However, time doesn't stand still and the volume of information is growing at an incredible rate. As a result, the analysis of the information received (reporting, etc.) becomes increasingly difficult due to the array of data. Some large companies come decided to implement the «Optimacros» system into the functionality of their organization. This system helps to analyze large amounts of data for effective decision-making and planning or monitoring the current operational activities of the company.

The main difference from traditional planning systems is a Self–Service constructor, which requires only analytics (modeling without programming, developer teams, technical specifications, etc.). To work, you only need a browser and access to a corporate domain (local area network) or the Internet. The product contains the flexibility of MS Excel and the reliability of operational systems.