

Федеральное государственное образовательное бюджетное
учреждение высшего образования
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ
РОССИЙСКОЙ ФЕДЕРАЦИИ»

Кафедра «Бизнес-информатика»

Учебное электронное издание на диске

Т.Е. Точилкина, А.А. Громова

ХРАНИЛИЩА ДАННЫХ И СРЕДСТВА БИЗНЕС-АНАЛИТИКИ

УЧЕБНОЕ ПОСОБИЕ

Москва • 2017

УДК 004.9(075.8)
ББК 65с
Т64

Рецензенты:

кандидат экономических наук, доцент Н.Ф. Алтухова
(Финансовый университет)

кандидат экономических наук Д.В. Турчановский
(АО «Зарубежнефть»)

Точилкина Т.Е., Громова А.А.

Т64 Хранилища данных и средства бизнес-аналитики: учебное пособие / Т.Е. Точилкина, А.А. Громова – М.: Финансовый университет, 2017. – 161 с.

ISBN 978-5-7942-1387-4

В учебном пособии представлен базовый материал по технологии хранилищ данных, современным тенденциям в области бизнес-аналитики; приведены описания практикумов по разработке хранилищ данных и бизнес-анализу; сформулированы требования к результатам самостоятельного выполнения заданий студентами; даны тесты для самоподготовки студентов.

Предназначено для формирования у студентов практических навыков разработки хранилищ данных и средств бизнес-аналитики с помощью современных программных продуктов.

Для студентов бакалавриата, обучающихся по направлениям «Прикладная информатика», «Бизнес-информатика», «Экономика», «Менеджмент».

УДК 004.9(075.8)
ББК 65с

ISBN 978-5-7942-1387-4

© Т.Е. Точилкина, А.А. Громова, 2017
© Финансовый университет, 2017

Federal State – Funded Educational Institution
of Higher Education

“FINANCIAL UNIVERSITY UNDER THE GOVERNMENT
OF THE RUSSIAN FEDERATION”

The Department of Business Informatics

Educational electronic edition on CD

Tatiana E. Tochilkina, Alla A. Gromova

DATA WAREHOUSE AND BUSINESS ANALITICS TOOLS

MANUAL

Moscow • 2017

UDC 004.9(075.8)

Reviewers:

Candidate of Economic Sciences, Associate Professor N.F. Altukhova
(Financial University)

Candidate of Economic Sciences D.V. Turchanovsky
(JSC “Zarubezhneft”)

Tatiana E. Tochilkina, Alla A. Gromova

Data Warehouse and Business Analytics tools: manual /
Tatiana E. Tochilkina, Alla A. Gromova – M.: Financial University,
2017. – 161 p.

ISBN 978-5-7942-1387-4

The manual contains information about the data warehousing and analytics trends in the business. The paper presents workshops on developing data warehouses and using BI to analyze; independent work; tests.

The purpose – formation of practical skills development of data storage and use of business intelligence tools.

The manual can be recommended to bachelors of “Applied informatics”, “Business informatics”, “Economics”, “Management”.

UDC 004.9(075.8)

ISBN 978-5-7942-1387-4

© Tatiana E. Tochilkina,
Alla A. Gromova, 2017
© Financial university, 2017

ОГЛАВЛЕНИЕ

Введение	8
Глава 1. Технология хранилищ данных	11
1.1. От транзакционных систем к системам аналитическим ...	11
1.2. Характеристики хранилищ данных	15
1.3. Модели данных хранилищ данных	17
1.4. Сценарий функционирования хранилища данных	25
1.5. Контрольные вопросы	28
1.6. Задания для самостоятельного выполнения	28
Глава 2. Современные тенденции в области бизнес-аналитики	31
Глава 3. Проектирование хранилищ данных с ERwin Data Modeler	38
3.1. Основы разработки хранилищ данных с ERwin Data Modeler	38
3.2. Настройка программной среды для практикума	39
3.3. Практикум по разработке хранилищ данных с ERwin Data Modeler	39
3.4. Контрольные вопросы	57
3.5. Задание для самостоятельного выполнения	57
Глава 4. Разработка хранилищ данных с MS SQL Server	59
4.1. Основные компоненты MS SQL Server для практикума ...	59
4.2. Настройка программной среды для практикума	60
4.3. Практикум по разработке хранилищ данных с MS SQL Server	60
4.4. Контрольные вопросы	81
4.5. Задания для самостоятельного выполнения	81
Глава 5. Использование надстроек MS Excel для бизнес-аналитики	83
5.1. Основы использования надстроек MS Excel для анализа	83
5.2. Настройка программной среды для практикума	84
5.3. Практикум по бизнес-анализу с надстройками MS Excel	84
5.4. Контрольные вопросы	92
5.5. Задание для самостоятельного выполнения	93

Глава 6. Анализ данных с помощью Microsoft Power BI	94
6.1. Подключение к хранилищу данных	94
6.2. Основы использования надстроек по формированию отчетов в Power BI	96
6.3. Основы использования облачного сервиса powerbi.com ...	98
6.4. Практикум по формированию отчетов в Power BI	102
6.6. Контрольные вопросы	105
6.6. Задание для самостоятельного выполнения	105
Глава 7. Анализ данных с помощью системы класса Data Discovery – Tableau	106
7.1. Подключение к хранилищу данных и внешним источникам	106
7.2. Основы использования надстроек по редактированию и агрегации данных в Tableau	107
7.3. Практикум по формированию отчетов в Tableau	112
7.4. Контрольные вопросы	122
7.5. Задание для самостоятельного выполнения	122
Глава 8. Формирование аналитической отчетности с помощью генератора отчетов SAP Crystal Reports	123
8.1. Создание отчета в SAP Crystal Reports	123
8.2. Основы расчетов в SAP Crystal Reports	135
8.3. Практикум по формированию шаблона отчета в SAP Crystal Reports с помощью мастера стандартных отчетов	138
8.4. Контрольные вопросы	143
8.5. Задание для самостоятельного выполнения	143
Тесты для самоподготовки	144
Ключи к тестам	150
Словарь	151
Заключение	158
Литература	159

CONTENTS

Introduction	8
Chapter 1. Data Warehouse	11
Chapter 2. Modern trends in business intelligence	31
Chapter 3. Developing Data warehouse design with ERwin Data Modeler	38
Chapter 4. Developing and deploying Data warehouse with MS SQL Server	59
Chapter 5. Applying MS EXCEL ADD-INS for BI	83
Chapter 6. Business analysis with Microsoft Power BI	94
Chapter 7. Business analysis with Tableau	106
Chapter 8. Creation of analytical reporting with SAP BO Crystal Reports	123
Tests for self-study	144
Keys to tests.....	150
Glossary	151
Conclusion.....	158
Bibliography.....	159

ВВЕДЕНИЕ

В современных условиях ведения бизнеса компаниям недостаточно уметь накапливать информацию о собственной операционной деятельности. Чтобы добиться действенных конкурентных преимуществ, компании вынуждены заниматься оперативным и глубинным анализом и прогнозированием, основанном на данных, хранящихся в разных информационных системах в разных форматах не только в самом копании, но и вне ее. Технологическим решением этой проблемы является использование аналитических систем, сочетающих технологии хранилищ данных, оперативного анализа данных, интеллектуального анализа данных и современных технологий визуализации.

Аналитики компании Gartner Group считают, что к 2018 г. большинство бизнес-пользователей организаций будут в своей повседневной деятельности использовать средства бизнес-аналитики; более половины крупных организаций по всему миру будут использовать данные средства в качестве конкурентного преимущества. К 2020 г. средства бизнес-аналитики привлекут на 40% больше новых инвестиций в организации.

Учебное пособие предназначено для студентов бакалавриата, обучающихся по направлениям «Прикладная информатика» (дисциплина «Хранилища данных и средства бизнес-аналитики»), «Бизнес-информатика» (дисциплина «Бизнес-аналитика для управления организацией»), «Экономика» (профили «Корпоративные финансы» и «Страхование»), дисциплина «Информационные технологии бизнес-аналитики»), «Менеджмент» (профиль «Антикризисное управление», дисциплина «Информационные технологии бизнес-аналитики»). Оно состоит из 8 глав:

- в главе 1 представлен базовый материал по технологии хранилищ данных;
- глава 2 представляет основные современные тенденции в области бизнес-аналитики;
- в главе 3 даны основы проектирования хранилищ данных с ERwin Data Modeler, даны рекомендации по организации программной среды для практикума; дано пошаговое описание практикума по проектированию витрины данных с ERwin Data Modeler; сформулированы требования к результатам самостоятельного выполнения задания по проектированию витрины данных;
- в главе 4 представлены основы разработки хранилищ данных с MS SQL Server, даны рекомендации по организации программной среды для практикума; дано пошаговое описание практикума по разработке хранилища данных с MS SQL Server; сформулированы требования к результатам самостоятельного выполнения задания;
- в главе 5 представлены основы использования надстроек MS Excel для бизнес-анализа, даны рекомендации по организации программной среды для практикума; дано пошаговое описание практикума по анализу таблиц с помощью надстроек MS Excel данных с MS SQL Server; сформулированы требования к результатам самостоятельного выполнения задания;
- главы с 6 по 8 содержат описания примеров использования средств бизнес-аналитики для анализа информации и поддержки принятия с помощью современных технологий.
- в главе 6 представлены основы использования BI-решения компании Microsoft Power BI, описаны возможности использования как настольного, так и облачного приложений, описываются основы формирования аналитических отчетов, информационных панелей и их публикации в сети интернет, дано пошаговое описание практикума по анализу данных с помощью системы Microsoft Power BI.
- в главе 7 описываются принципы анализа информации с помощью решения компании Tableau Software. Tableau – инновационная система бизнес-аналитики, которое позволяет каждому анализировать любые типы данных и создавать блестящую отчетность за минуты. Глава содержит описание возможностей по подключению и настройке данных, формированию аналитических отчетов, информационных панелей и историй (интерактивных ситуационных табло); дано поша-

говое описание практикума по анализу данных с помощью системы Tableau.

- глава 8 содержит описание возможностей самого популярного генератора отчетов SAP BO Crystal Reports. Crystal Reports – программа или библиотека, позволяющая представить информацию в удобочитаемом структурированном виде. Другими словами сделать из данных информацию (документ, отчет), которую можно распечатать или сохранить в различных электронных форматах. Глава иллюстрирует возможности системы в области формирования аналитических отчетов; дано пошаговое описание практикума по формированию шаблонов отчетов.

В конце учебного пособия приводятся тесты для самоподготовки с ключами для правильных ответов, словарь и список литературы.

Главы 1 (п. 1.1–1.3, 1.5, 1.6), 3, 4 и 5 разработаны Т.Е. Точилкиной, главы 1 (п. 1.4), 2, 6–8 разработаны А.А. Громовой.

Материалы, представленные в учебном пособии, прошли апробацию в процессе обучения студентов-бакалавров в рамках аудиторных лекционных, практических и самостоятельных занятий в 2015/2016 учебном году.

ТЕХНОЛОГИЯ ХРАНИЛИЩ ДАННЫХ

1.1. ОТ ТРАНЗАКЦИОННЫХ СИСТЕМ К СИСТЕМАМ АНАЛИТИЧЕСКИМ

Для автоматизации операционных задач (учет платежей в бюджет, учет расходов бюджета, учет клиентов, учет договоров, учет заказов, учет взаиморасчетов, учет запасов и пр.), которые решаются сотрудниками «нижнего» звена финансовых учреждений, производственных, консалтинговых компаний и других организаций, традиционно используются учетные системы, называемые также **OLTP-системами** или **транзакционными** системами (см. словарь) [4, 8, 9, 12].

Некоторые из OLTP-систем являются комплексными и состоят из ряда модулей. Например, системы управления ресурсами предприятий (ERP-системы) имеют модульную структуру, объединяя различные службы предприятия в единый управленческий контур.

OLTP-система осуществляет учет и хранение первичной информации о работе организации, обрабатывая огромное количество транзакций, производя «горы» данных, связанных с операционной деятельностью.

Чтобы обеспечить качественное автоматизированное выполнение операционных задач организации база данных OLTP-системы должна удовлетворять ряду требований. Как правило выделяют следующие требования. База данных OLTP-системы должна быть

- оптимизирована на выполнение максимального количества транзакций за короткие промежутки времени. При этом показателем эффективности является количество транзакций, выполняемых за секунду;
- иметь нормализованную реляционную структуру;
- удовлетворять жестким требованиям ссылочной целостности данных.

В нормализованной реляционной базе данных можно хранить факты без избыточного дублирования. Однако вследствие нормализации база данных OLTP-системы представляет собой сотни, тысячи, десятки тысяч связанных таблиц; тысячи, десятки и сотни тысяч колонок (рис. 1.1). При этом одних лишь наименований таблиц и колонок как правило недостаточно для понимания содержащейся в них информации. Поэтому собранная в базе данных OLTP-системы информация мало пригодна для анализа без участия специалистов-программистов. В связи с этим для конечных пользователей OLTP-система предоставляет набор встроенных отчетов.

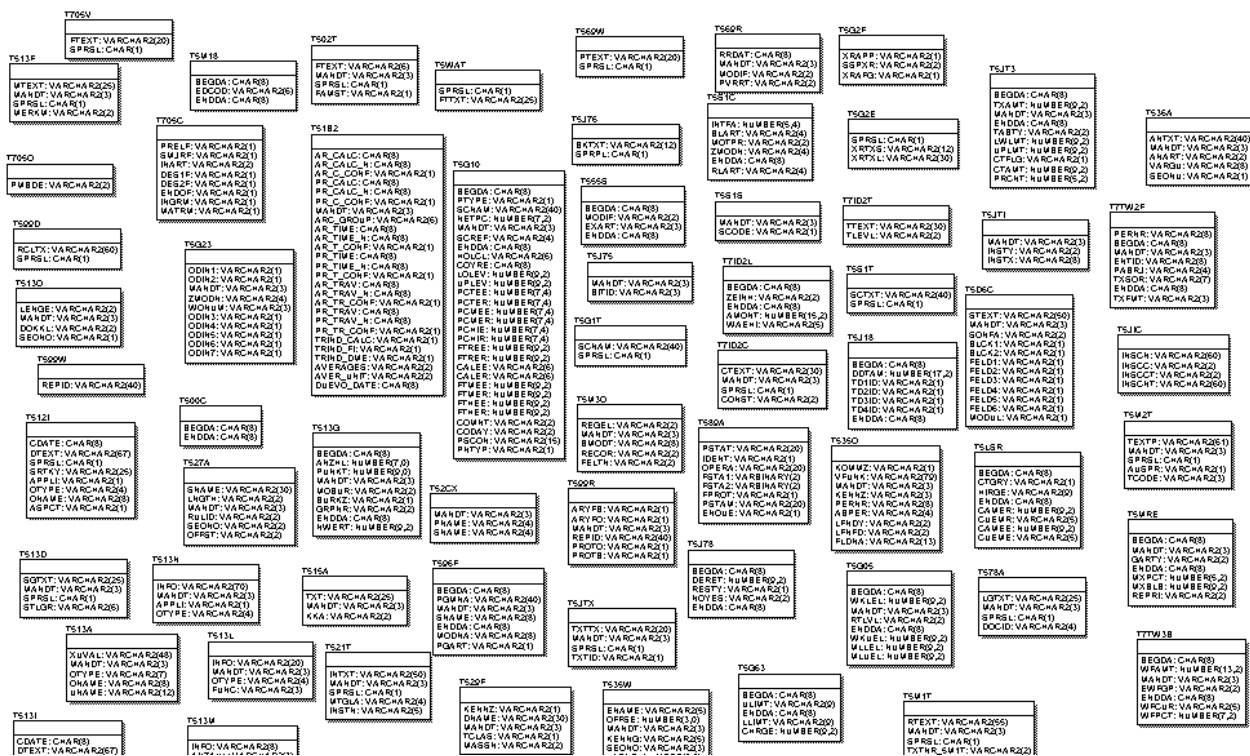


Рисунок 1.1 – Фрагмент структуры базы данных OLTP-системы

Получение агрегированной информации (которая как правило и требуется лицам, принимающим решения на уровне всей организации) из базы данных OLTP-системы часто требует выполнения операций соединения по многим таблицам, содержащим большое число записей. Такие запросы могут выполняться часами, перегружая OLTP-систему, нарушая тем самым нормальную работу подразделений организации.

Кроме того, для планирования и оптимизации ресурсов руководителям нужно знать, как на загрузку предприятия влияют сезонные

и годовые тренды. Например, можно сравнивать продажи в течение первого квартала этого года с продажами в течение первого квартала предшествующих лет или попытаться оценить влияние новой компании маркетинга, проходящей в течение определенных периодов, рассматривая продажи в течение тех же самых периодов. Однако OLTP-системы не предназначены для хранения, анализа информации за длительный период времени. Данные в большинстве OLTP-систем архивируются сразу после того, как они становятся неактивными. Например, заказ может стать неактивным после того, как он выполнен; банковский счет может стать неактивным после того, как он был закрыт.

Таким образом характеристики OLTP-системы не позволяют использовать их для оперативного анализа непосредственно лицами, принимающими решения.

В современной организации одновременно действуют несколько (иногда несколько десятков) OLTP-систем (рис. 1.2).

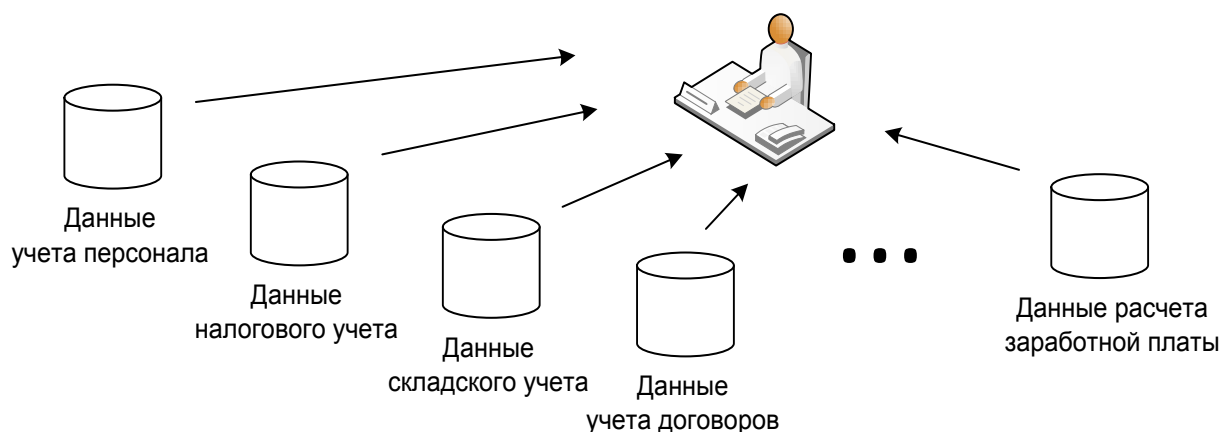


Рисунок 1.2 – Пример многообразия OLTP-систем в организации

Для предоставления необходимой для принятия решений на уровне всей организации информации приходится анализировать данные из всех OLTP-систем. Более того в силу глобализации экономики, увеличения конкуренции организациям приходится учитывать также многочисленные внешние факторы, использовать кроме внутренних еще и внешние источники данных, например, данные от поставщиков и партнеров, данные по законодательству, данные из социальных сетей и т.д.

Основная проблема многообразия источников состоит в несогласованности и противоречивости содержащихся в них данных, в отсутствии единого логического взгляда на корпоративные данные.

Итак, успешная деятельность организации невозможна без принятия обоснованных управленческих решений. Такие решения могут быть построены на основе всестороннего анализа результатов выполнения бизнес-процессов в самой организации, а также многочисленных внешних факторов. Время принятия решений в современных условиях сокращается. Роль информационных технологий, поддерживающих процессы бизнес-анализа и принятия управленческих решений, возрастает.

Следствием этого стало активное развитие особого класса информационных систем – **информационно-аналитических систем** (см. словарь), ориентированных на оперативную аналитическую обработку данных, извлекаемых из множества источников данных как внутри, так и вне организации, предназначенных для помощи управляющему персоналу организации в принятии обоснованных своевременных решений.

Информационно-аналитическая система базируется на нескольких информационных технологиях. Как правило информационно-аналитическая система сочетает технологию хранилищ данных, технологию оперативного анализа данных, технологию интеллектуального анализа данных и современные технологии визуализации (рис. 1.3).

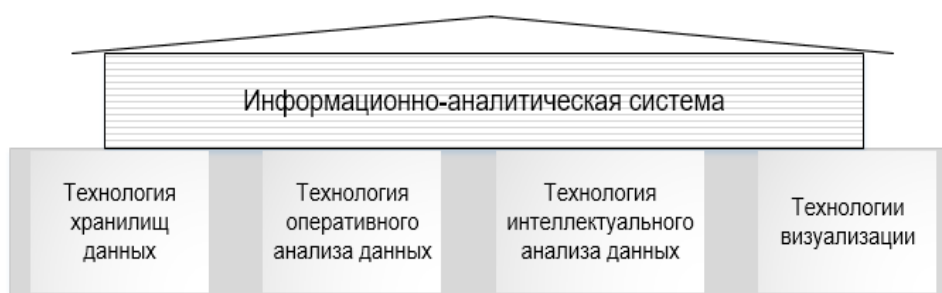


Рисунок 1.3 – Базовые технологии информационно-аналитической системы

Упрощенно структуру информационно-аналитической системы можно представить следующим образом (рис. 1.4). В центре информационно-аналитической системы обычно находится хранилище данных, в котором собираются очищенные, консолидированные, агреги-

рованные данные, собранные из OLTP-систем организации, ее партнеров и других источников. Данные из OLTP-систем копируются в хранилище данных таким образом, чтобы в процессе анализа не использовались ресурсы транзакционных систем и не нарушалась стабильность их работы. Данные в хранилище, как правило, доступны только для чтения. Данные исследуются с помощью инструментов бизнес-аналитики. Результаты анализа данных представляются лицу, ответственному за принятие решений, с помощью инструментов визуализации.

Более подробно архитектура информационно-аналитической системы рассмотрена во второй главе пособия.

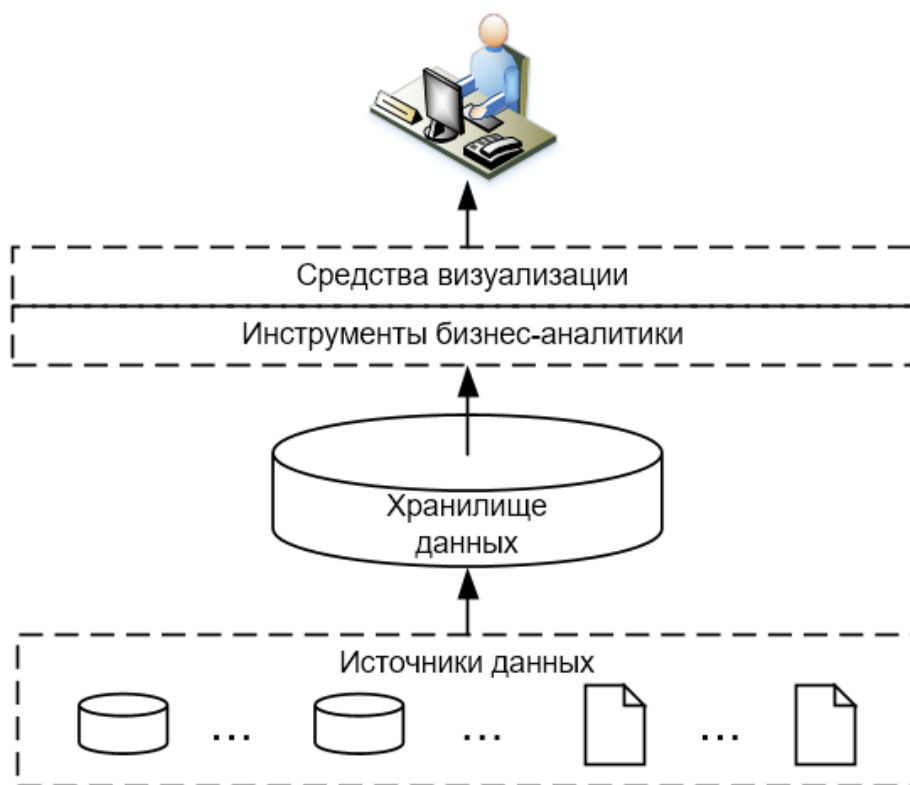


Рисунок 1.4 – Укрупненная структура аналитической системы

1.2. ХАРАКТЕРИСТИКИ ХРАНИЛИЩ ДАННЫХ

Согласно классическому определению Б. Инмона, **хранилище данных** представляет предметно-ориентированный, интегрированный, привязанный ко времени и неизменяемый набор данных, предназначенный для поддержки принятия решений.

Перечисленные характеристики принципиально отличают хранилище данных от базы данных OLTP-системы. Рассмотрим эти характеристики подробнее.

Предметная ориентированность предполагает, что данные в хранилище данных объединяются в категории и хранятся в соответствии с областями, которые они описывают (продажи, бюджетные поступления и т.п.).

Реализуется предметная ориентированность через использование особых схем организации данных (см. п. 1.3). Это позволяет упростить создание аналитических запросов, а также увеличить скорость их выполнения.

Предметная ориентированность данных в хранилище данных отличается от организации данных в базе данных OLTP-систем, группирующихся в соответствии с бизнес-операциями (выписка счетов, отгрузка товара и т.п.), оптимизированных на интенсивное оперативное выполнение операций вставки/обновления/удаления небольших порций данных; имеющих, как правило, нормализованную реляционную структуру баз данных; удовлетворяющих жестким требованиям ссылочной целостности данных.

Интегрированность означает, что данные для анализа не берутся напрямую из источников, в частности, баз данных OLTP-систем предприятия. Исходные данные извлекаются, проверяются, очищаются, унифицируются и т.п. (см. ETL в словаре), чтобы удовлетворять требованиям аналитика, исследующего не отдельную бизнес-функцию, а деятельность всего предприятия. Процесс ETL позволяет повысить качество данных аналитической системы и скорость выполнения аналитических запросов к хранилищу данных.

Привязка ко времени подразумевает, что данные в хранилище всегда жестко «привязаны» к определенному периоду времени. Данные, выбранные из баз данных OLTP-систем и других источников, накапливаются в хранилище в виде исторических слоев, каждый из которых относится к конкретному периоду времени. Это позволяет анализировать тенденции в развитии бизнеса.

Неизменяемость означает, что, попав в определенный исторический слой хранилища, данные уже никогда не будут изменены. Стабильность данных способствует достоверности результатов аналитических запросов.

Это также отличает хранилища данных от баз данных OLTP-систем. В OLTP-системах данные все время меняются, и один и тот же аналитический запрос, выполненный дважды с интервалом в 10 минут, может дать разные результаты, приводя к несогласованным управленческим решениям.

Таким образом, хранилище можно охарактеризовать как особую базу данных, содержащую данные

- масштаба организации;
- собранные из различных источников за длительный период времени, очищенные и консолидированные;
- привязанные ко времени;
- структурированные в целях упрощения выполнения аналитических запросов.

1.3. МОДЕЛИ ДАННЫХ ХРАНИЛИЩ ДАННЫХ

Модель данных хранилища может быть представлена на трех уровнях [9, 10]

- концептуальной модели данных;
- логической модели данных;
- физической модели данных.

Концептуальная модель является отражением предметной области, в рамках которой планируется построение хранилища данных, а также спектра аналитических задач, для решения которых разрабатывается хранилище.

Логическая модель расширяет концептуальную, уточняя состав таблиц и взаимосвязи между ними, добавляя колонки, а также определения для таблиц и колонок.

Физическая модель описывает реализацию объектов логической модели на уровне объектов конкретной базы данных.

Концептуальная модель данных хранилища

В основе концептуальной модели хранилища лежит **многомерная модель данных**. Основными понятиями многомерной модели данных являются: **куб** (гиперкуб) данных, **факты**, **показатели** (меры), **измерения**, **иерархии**, **агрегаты**, **срез** (см. словарь).

Поскольку многомерная модель данных предназначена для предоставления VI-информации, все показатели, измерения, иерархии должны иметь названия, доступные для понимания лицам, принимающим решения в организации. Кроме того, модель предусматривает возможность сопровождать названия объектов развернутыми описаниями. Это позволяет ответственному за принятие решений, удостовериться, что он работает с требуемой аналитической информацией.

На рис. 1.5 приведен пример упрощенного гиперкуба, разработанного для анализа поступлений в бюджет Российской Федерации (значения показателя намеренно изменены). «Поступления в бюджет» – единственный показатель данного гиперкуба. Три измерения гиперкуба: «Время», «Доходы бюджета», «ОКАТО» – позволяют анализировать факты поступлений в бюджет в контексте

- определенного времени (конкретный месяц определенного года, конкретный квартал определенного года);
- вида дохода бюджета (НДС, налог на прибыль, НДФЛ, налоговые доходы);
- административно-территориальной принадлежности корреспондентов (Артемьевский район, Егорьевский район, Истринский район).

В ячейках гиперкуба «хранятся» фактические значения показателя (факты). Факты могут быть **аддитивными**, **полуаддитивными** и **неаддитивными** (см. словарь). В приведенном на рис. 1.5. примере представлены аддитивные факты, так как для показателя «Поступления в бюджет» может быть выполнено суммирование по всем измерениям куба данных.

В гиперкубе на рис. 1.5 отображены как детализированные, так и агрегированные значения показателя «Поступления в бюджет». Детализированные значения – это факты поступлений в бюджет по конкретному виду дохода бюджета за конкретный месяц от корреспондентов определенного района. Примером детализированного значения может служить факт поступлений в бюджет по налогу на прибыль в феврале 2013 г. от корреспондентов Артемьевского района – 25 млн руб. Примером агрегированного значения в рассматриваемом гиперкубе могут быть все налоговые поступления в бюджет в феврале 2013 г. от корреспондентов Артемьевского района – 75 млн руб.

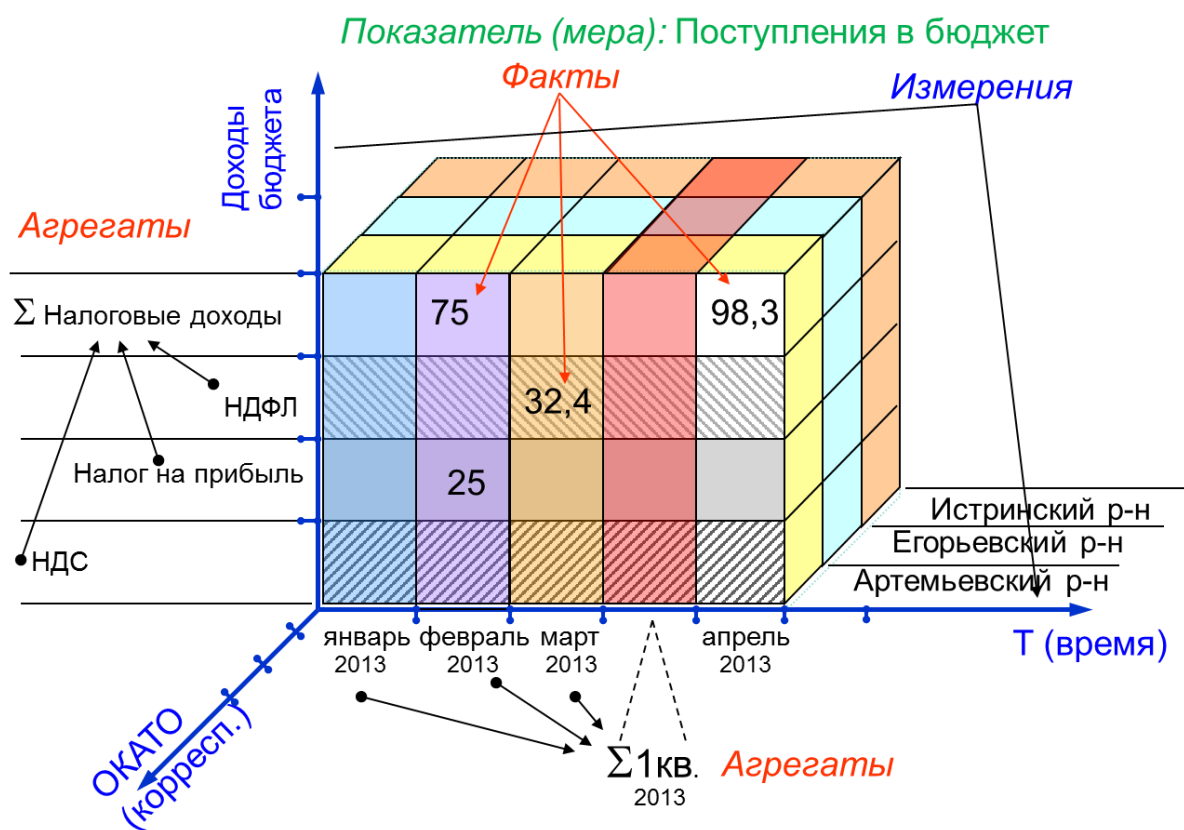


Рисунок 1.5 – Пример куба данных для анализа поступлений в бюджет

При выполнении аналитического запроса к гиперкубу данных на каждом измерении фиксируется одно или несколько значений. Результатом выполнения аналитического запроса к гиперкубу является срез – куб меньшей размерности при фиксированных значениях для каждого измерения.

Например, при выполнении запроса «*Каковы были поступления в бюджет по налогу на прибыль от корреспондентов Артемьевского района за февраль 2013 г.?*» на каждом измерении гиперкуба будет зафиксировано по одному значению, и результатом выполнения этого запроса станет срез гиперкуба, состоящий из одной ячейки.

При выполнении запроса «*Каковы были поступления в бюджет по налогу на прибыль и НДС от корреспондентов Артемьевского и Егорьевского районов за январь и февраль 2013 г.?*» на каждом измерении гиперкуба будет зафиксировано по два значения, и результатом выполнения этого запроса станет срез гиперкуба, состоящий из восьми ячеек (рис. 1.6).

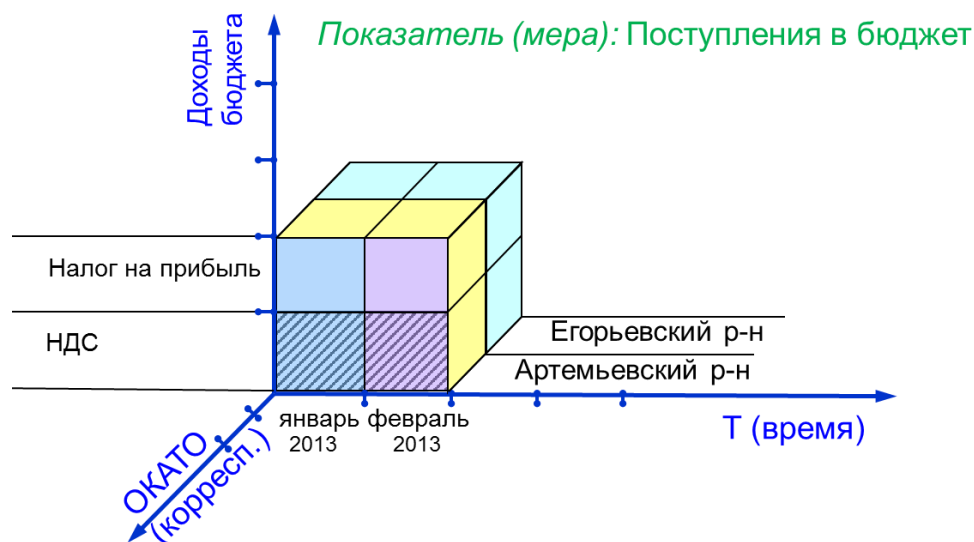


Рисунок 1.6 –Пример среза куба данных

Многомерная модель позволяет делать срезы куба данных и поворачивать их нужной гранью любым удобным аналитику образом.

Таким образом, используя концептуальную модель хранилища, аналитик и/или лицо, принимающее решения, могут оперативно самостоятельно сформировать запрос в соответствии с текущей аналитической задачей.

Логическая модель данных хранилища

Концептуальная модель хранилища (гиперкуб) на логическом уровне представляется с помощью **схемы «Звезда»** или ее варианта - **схемы «Снежинка»** (см. словарь).

Схема «Звезда» имеет одну **таблицу фактов** и несколько **таблиц измерений** (см. словарь). Таблицы измерений являются денормализованными.

Схема «Снежинка» имеет одну таблицу фактов и несколько нормализованных таблиц измерений.

Одна схема «Звезда», как и гиперкуб, охватывает отдельную сферу бизнеса предприятия. Так, схема «Звезда» на рис. 1.7 охватывает сферу продаж. В соответствующей концептуальной модели «Сумма» – представляет единственный показатель гиперкуба, а «Клиент», «Продавец», «Время», «Товар» – его четыре измерения.

Схема «Звезда» обеспечивает высокую скорость выполнения запроса посредством **денормализации** данных (см. словарь).

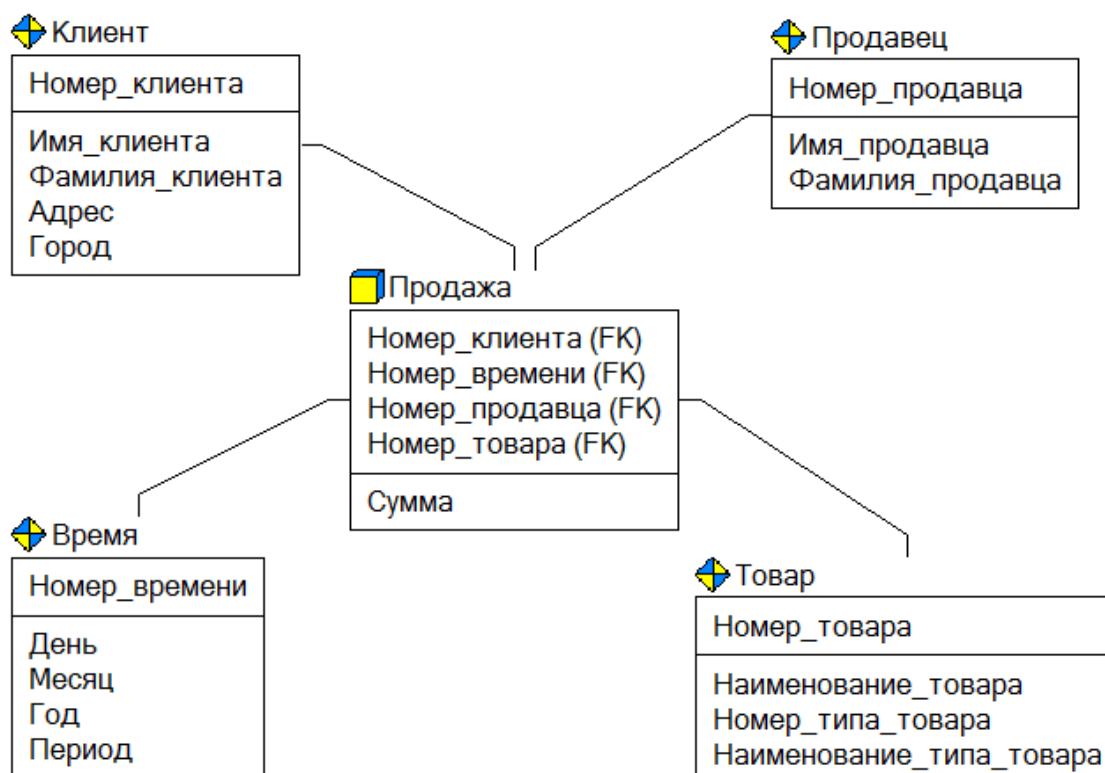


Рисунок 1.7 – Пример схемы «Звезда» для анализа продаж

Невозможно создать универсальную денормализованную структуру данных, обеспечивающую высокую производительность при выполнении любого аналитического запроса. Поэтому схема «Звезда» строится так, чтобы обеспечить наивысшую производительность при выполнении одного самого важного запроса, либо для группы похожих запросов.

Так, хранилище, разработанное на основе схемы на рис. 1.7, позволит оперативной ответить на вопросы, связанные продажами в компании.

Все факты, необходимые для ответа на запросы, помещаются в таблицу фактов. Так, на рис. 1.7 в таблице фактов «Продажа» все факты относительно суммы продаж размещаются в колонке «Сумма».

Таблица фактов может состоять из миллионов строк и содержать агрегированные или детализированные данные, которые могут помочь ответить на требуемые вопросы. Она содержит данные, которые хранились бы во многих таблицах нормализованных реляционных баз данных.

Для анализа полезны таблицы фактов, содержащие детальные данные. Например, предпочтительнее взять за основу факты продажи

отдельного товара отдельному клиенту, а не суммы продаж для разных стран. Агрегированные данные будут вычислены.

Таблица фактов является центральной таблицей в схеме «Звезда».

Таблицу фактов окружают меньшие таблицы, называемые таблицами размерности или измерений. Таблицы измерений содержат неизменяемые либо редко изменяемые данные типа справочника. В рассмотренном примере на рис. 1.7 предполагается, что записи в таблицах «Клиент», «Продавец», «Время», «Товар» не изменяются или изменяются редко.

Таблицы измерений соединены с таблицей фактов в виде звезды радиальными связями. В этих связях таблицы размерности являются родительскими, таблица факта – дочерней. Таблица фактов и таблицы размерности связаны идентифицирующими связями, при этом первичные ключи таблицы размерности мигрируют в таблицу фактов в качестве внешних ключей, как показано на рис. 1.7.

Таблица фактов, как правило, содержит уникальный составной ключ, объединяющий первичные ключи таблиц измерений (см. рис. 1.7). Чаще всего это целочисленные значения либо значения типа «дата/время».

В схеме «Звезда» измерения денормализованы. Так, на схеме на рис. 1.7 в таблице измерения «Время» неключевые столбцы «День», «Месяц», «Год», «Период» функционально зависимы; в таблице измерения «Товар» функциональная зависимость прослеживается между неключевыми столбцами «Наименование товара» и «Наименование типа товара». Вследствие этого, с одной стороны, в хранилище наблюдается дублирование данных, а с другой – ускорение времени выполнения запросов и упрощение восприятия схемы. Поэтому в долгосрочных рекомендуется использовать именно схему «Звезда».

В краткосрочных пилотных проектах для сокращения времени разработки хранилища иногда используют схему «Снежинка» с ее нормализованными таблицами измерений, так как она ближе по структуре к источникам – нормализованным базам данных OLTP-систем.

На рис. 1.8. представлен схема «Снежинка», на основе которой планируется выполнить пилотный проект по разработке хранилища данных для анализа продаж компании. В данной схеме таблицы «Клиент», «Продавец», «Время», «Товар» были заимствованы из баз данных OLTP-систем, и остались нормализованными (см. дополнительные таблицы измерений «Регион», «Период», «Тип товара»). Это

позволит сократить сроки проекта. Однако скорость выполнения аналитических запросов к такому хранилищу будет ниже, чем к хранилищу, построенному на основе схемы «Звезда» на рис. 1.7.

Дополнительные таблицы измерений в схеме «Снежинка», обычно соответствуют верхним уровням **иерархии измерения** (см. словарь) и находятся в соотношении «один ко многим» с главной таблицей измерений, соответствующей нижнему уровню иерархии. Эти дополнительные таблицы иногда называют **консольными** (см. словарь). На схеме, представленной на рис. 1.8, таблицы «Регион», «Период», «Тип товара» являются консольными.

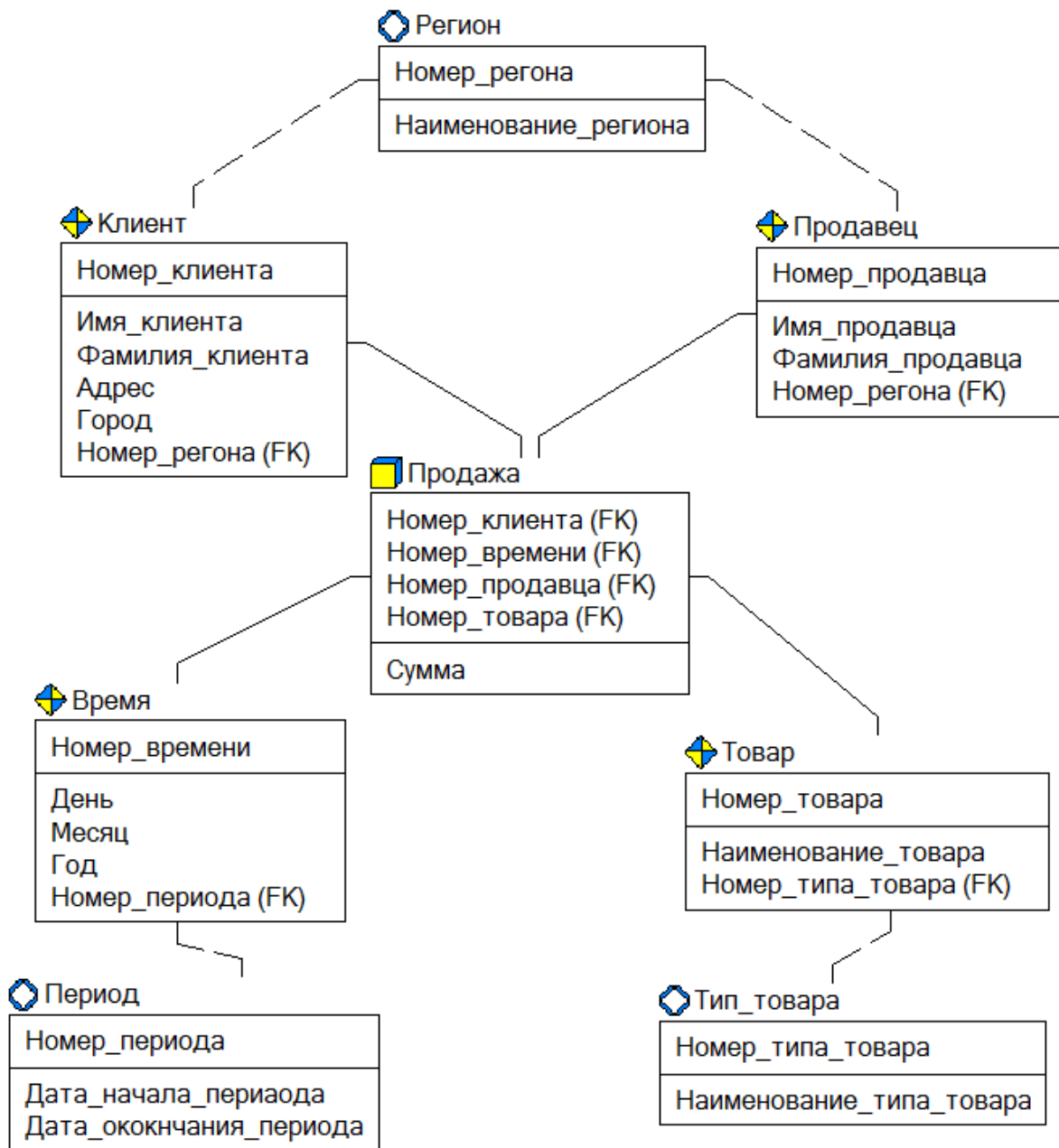


Рисунок 1.8 – Пример схемы «Снежинка» для анализа продаж

Консольные таблицы могут быть связаны только с таблицами измерений, причем консольная таблица в этой связи родительская, а таблица измерений – дочерняя. Связь может быть идентифицирующей или неидентифицирующей.

Консольная таблица не может быть связана с таблицей факта. Она используется для нормализации данных в таблицах измерений.

Бывают ситуации, когда в схеме «Звезда» денормализованная таблица измерений оказывается слишком большой (имеется в виду количество колонок), при этом к части колонок запросы выполняются чаще, чем к другим колонкам.

Например, в результате анализа статистики по видам запросов к хранилищу было обнаружено, что количество запросов о продажах с детализацией по наименованию товара в десять раз меньше, чем количество запросов о продажах по типам товаров. В таких случаях изменяют схему «Звезда»: таблицу измерения разбивают на две отдельные таблицы, связав их неидентифицирующей связью, как показано на рис. 1.9.

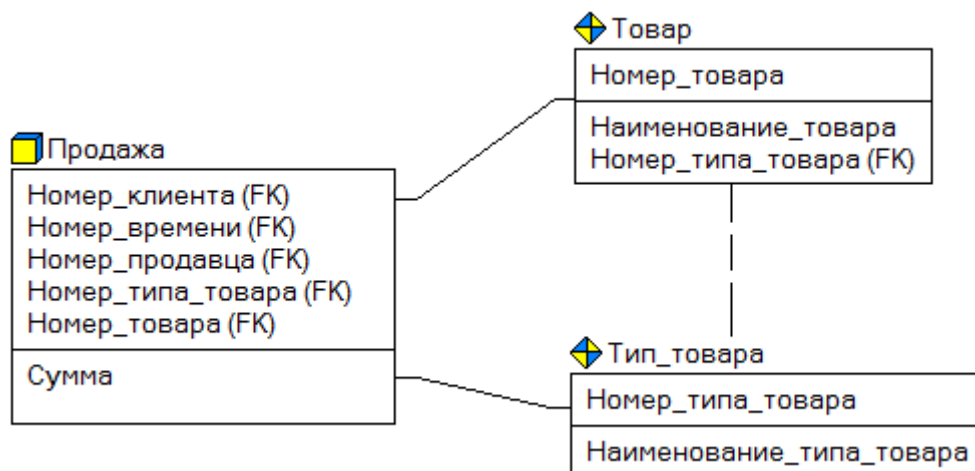


Рисунок 1.9 – Таблица измерения разбита на две связанные таблицы

Хранилище данных предприятия содержит данные из разных сфер деятельности и состоит из набора связанных схем «Звезда» или «Снежинка». Согласование разных схем производится через общие таблицы измерений.

Физическая модель данных хранилища

Физическая модель может быть представлена с помощью схем «Звезда» или «Снежинка», дополненными описанием реализации

объектов логической модели на уровне объектов конкретной базы данных.

Так, на рис. 1.10 представлена физическая модель хранилища, соответствующая логической модели на рис. 1.7, отображающая ряд дополнительных характеристик объектов базы данных SQL Server 2012, а именно: тип данных колонок, а также возможность отсутствия значений колонок (опция Null/Not Null).

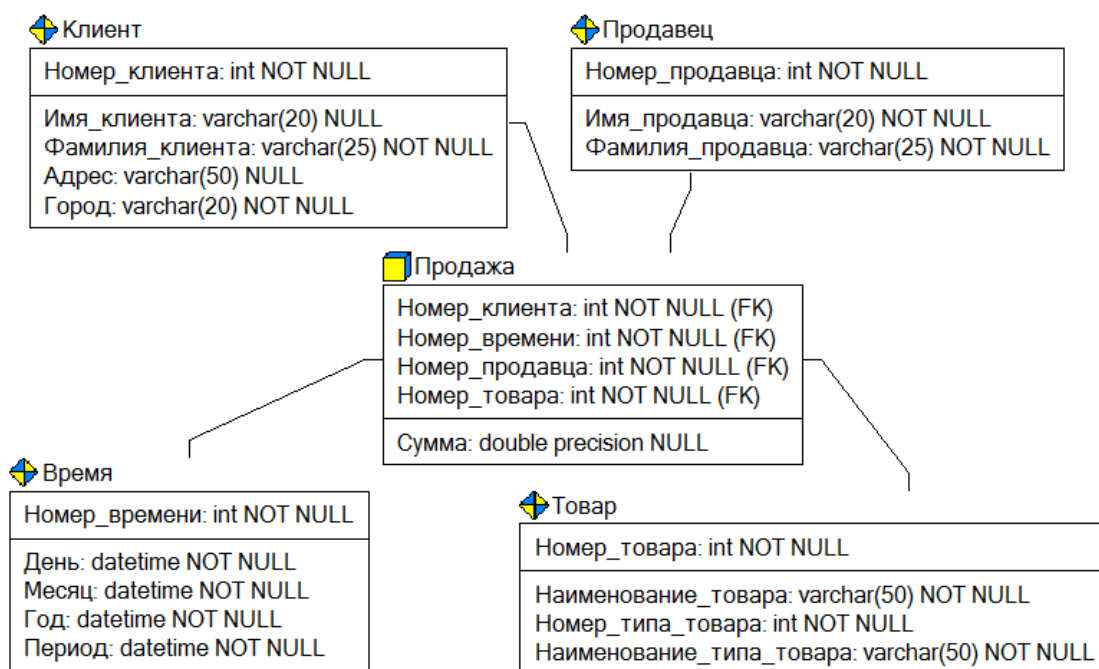


Рисунок 1.10 – Пример физической модели хранилища

1.4. СЦЕНАРИЙ ФУНКЦИОНИРОВАНИЯ ХРАНИЛИЩА ДАННЫХ

По заданному регламенту в ХД собираются данные из различных источников – баз данных систем оперативной обработки. В хранилище поддерживается хронология: наравне с текущими хранятся исторические данные с указанием времени, к которому они относятся. В результате необходимые доступные данные об объекте управления собираются в одном месте, приводятся к единому формату, согласовываются и, в ряде случаев, агрегируются до минимально требуемого уровня обобщения.

В общем виде, сценарий функционирования хранилища данных выглядит следующим образом (рис. 1.11):

1. Извлечение, преобразование и загрузка данных.
2. Хранение данных.
3. Анализ данных (рабочие места пользователей).

Извлечение, преобразование и загрузка данных

Данные поступают из различных внутренних транзакционных систем, от подчиненных структур, от внешних организаций в соответствии с установленным регламентом, формами и макетами отчетности. Вся эта информация проверяется, согласуется, преобразуется и помещается в хранилище и витрины данных. После этого пользователи с помощью специализированных инструментальных средств получают необходимую им информацию для построения различных табличных и графических представлений, прогнозирования, моделирования и выполнения других аналитических задач.

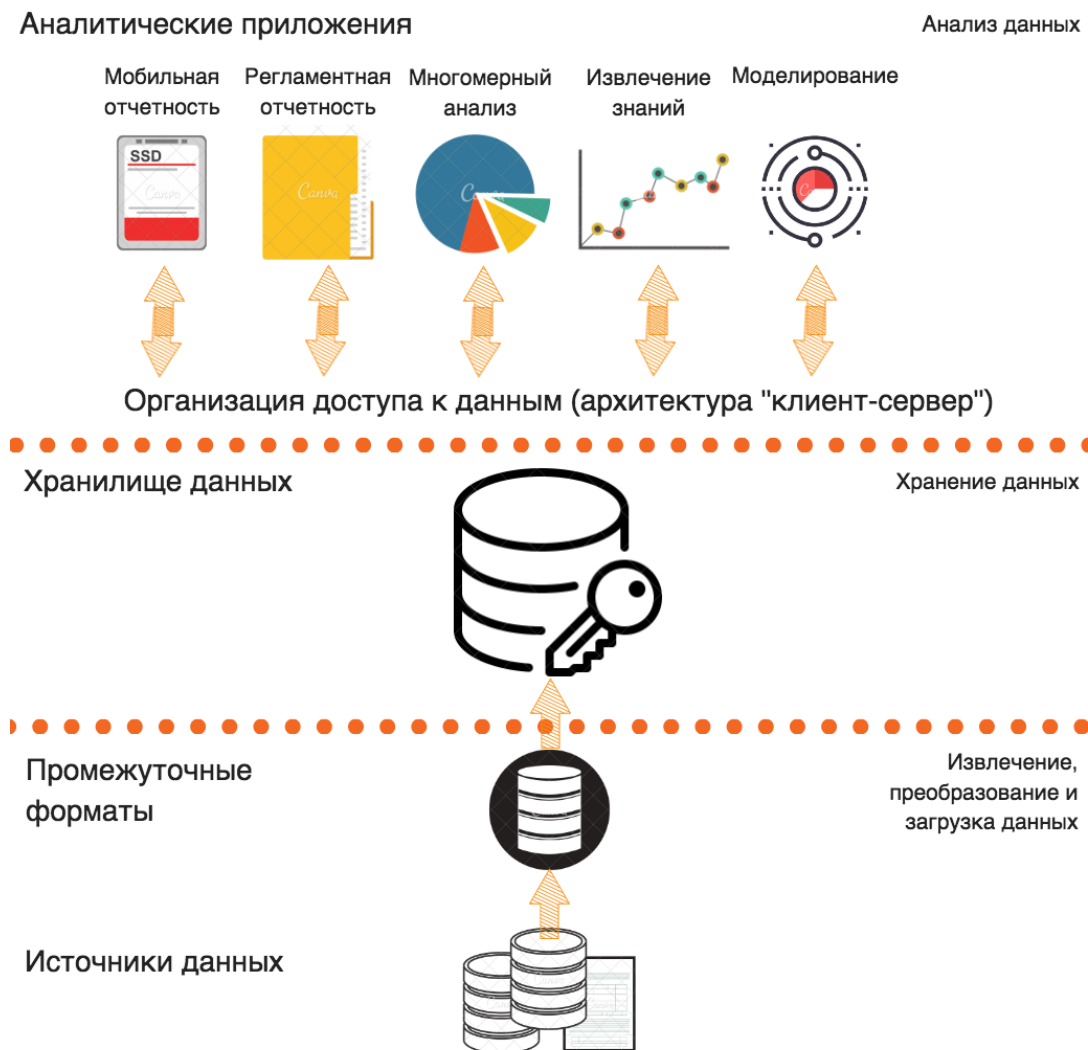


Рисунок 1.11 – Сценарий функционирования хранилища данных

В ходе наполнения хранилища информация подвергается много-ступенчатой обработке. На первом этапе производится буферизация сведений, поступающих из разных источников информации, что позволяет стабилизировать процесс загрузки и выполнять сверку при-сланных данных. Первую буферную зону называют шлюзом. На сле-дующем этапе данные преобразуются и перемещаются из одной бу-ферной зоны в другую. Здесь устраняется структурная неоднород-ность данных (унифицируются длины полей, преобразуются форматы и пр.), проверяется их целостность, служебным полям хранилища присваиваются соответствующие значения, решаются задачи инте-грации данных. На этом этапе производится также очистка информа-ции. Сведения, не прошедшие процедуры преобразования, очистки или проверки, попадают в специальную зону, где их в полуавтомати-ческом режиме можно исправить, дополнить, после чего заново за-грузить в хранилище.

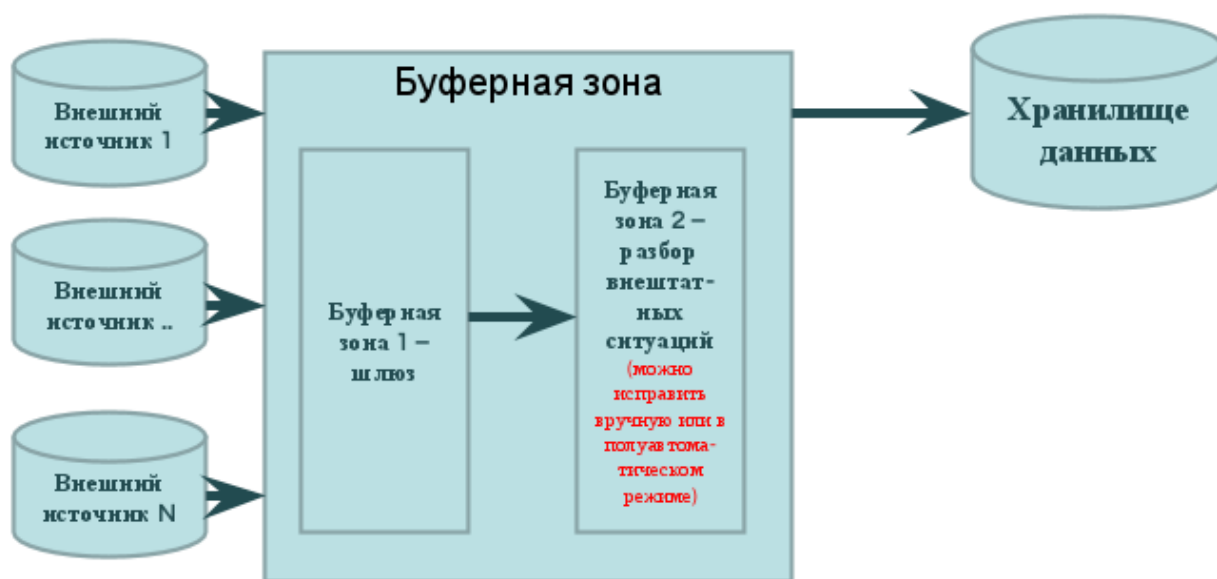


Рисунок 1.12 – Загрузка данных в хранилища данных

Из источников данных информация перемещается на основе не-которого регламента в централизованное хранилище (более того, от-дельные таблицы, могут обновляться по разному расписанию: еже-дневно, еженедельно, ежемесячно, либо не обновляться вовсе). Часто, необходимые для анализа данные не хранятся в окончательном виде ни в одной из транзакционных систем. Эти данные либо можно полу-

чить из исходных баз данных путем специальных преобразований, вычислений и агрегирования, либо они могут быть загружены в хранилище в ручном или полуавтоматическом режиме.

Таким образом, процесс настройки хранилища данных подразумевает выполнение следующих шагов:

1. Создание модели хранилища данных с помощью специализированного CASE-средства.
2. Импорт описания метаданных.
3. Настройка внешних источников.
4. Загрузка данных.
5. Хранение данных.
6. Предоставление данных для анализа.

1.5. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Каковы основные причины появления технологии хранилищ данных?
2. Каковы основные характеристики хранилища данных?
3. Каковы основные понятия гиперкуба данных?
4. Каковы основные характеристики моделей данных «Звезда» и «Снежинка»?
5. В чем заключаются достоинства и недостатки введения иерархической структуры измерений гиперкуба?

1.6. ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОГО ВЫПОЛНЕНИЯ

Задание 1

Имеется модель данных витрины для аналитической системы (см. рис. 1.13).

Известно, что эта модель отвечает требованиям бизнеса. Для указанной модели:

- дайте характеристику представленной модели данных;
- назовите направление бизнеса, для анализа которого разработана витрина;
- приведите примеры запросов, которые можно сформулировать к представленной витрине (не менее 5 запросов);

– приведите примеры структур аналитических отчетов, которые могут быть сгенерированы аналитической системой для лиц, принимающих решения (не менее двух структур);

Результат оформить как документ Word. В названии документа указать номер группы и фамилию. Документ сохранить в личной папке.

Задание 2

Для модели на рис. 1.14 выполнить те же пункты, что в задании 1.

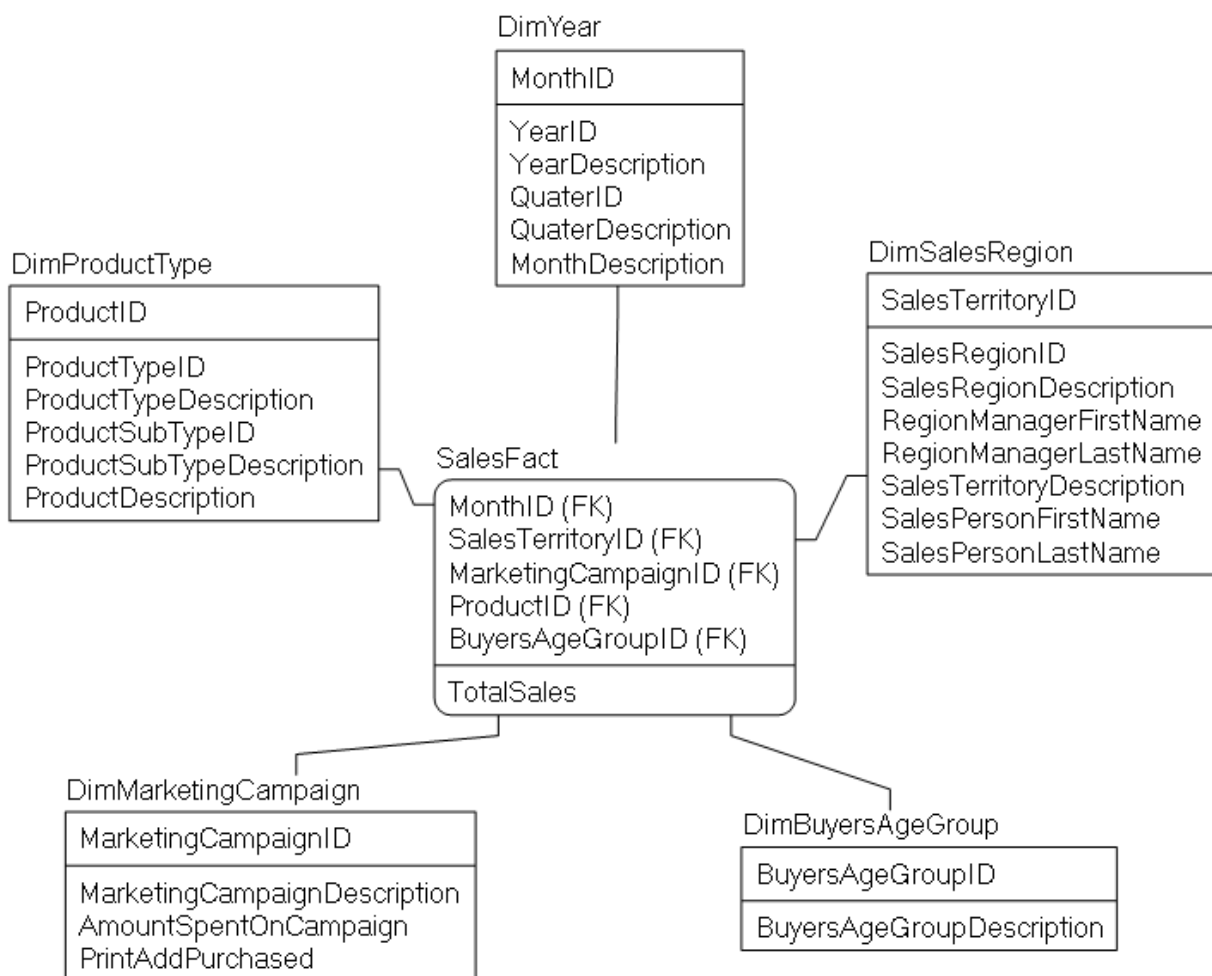


Рисунок 1.13 – Модель витрины данных для задания 1

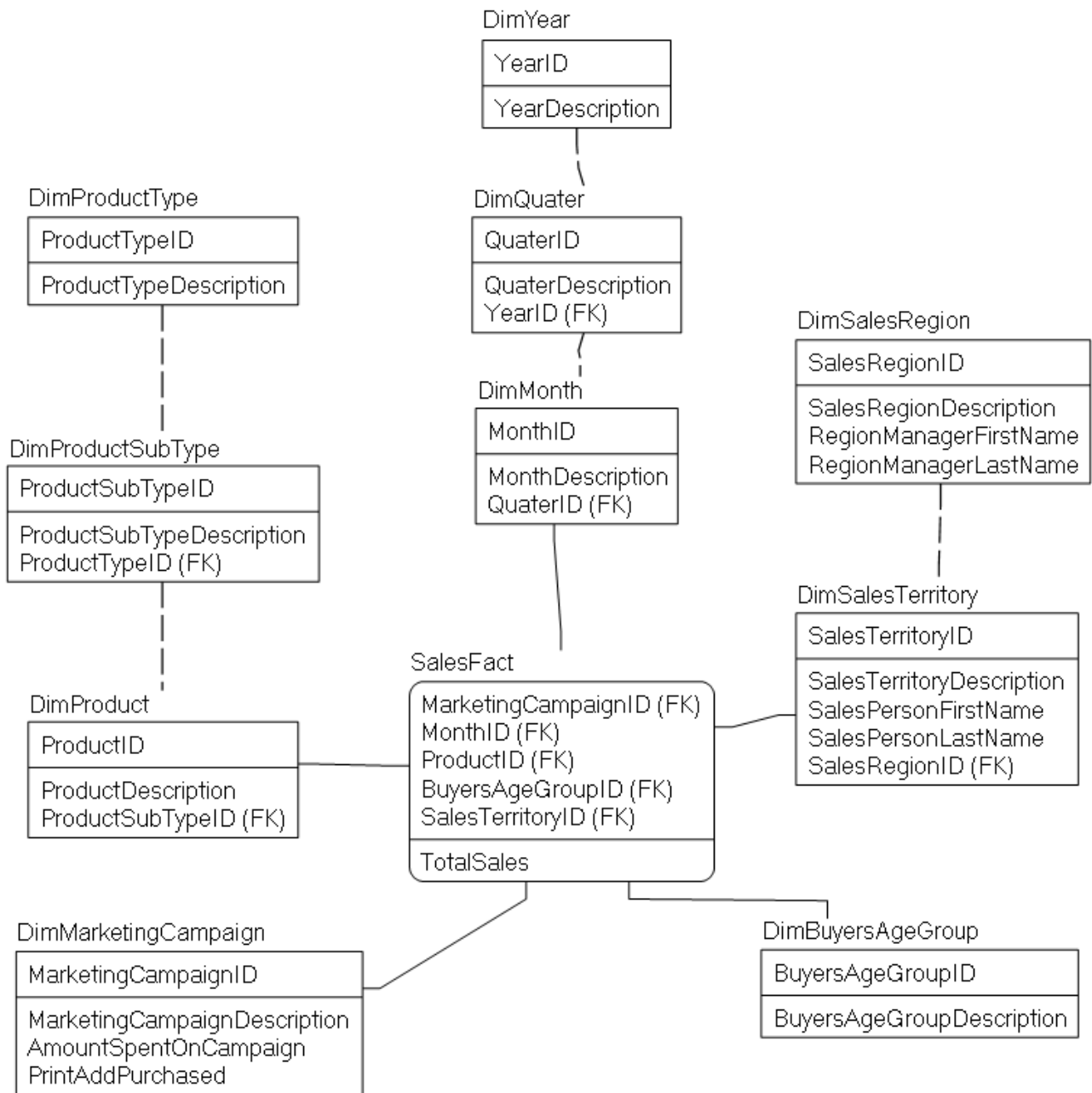


Рисунок 1.14 – Модель витрины данных для задания 2

СОВРЕМЕННЫЕ ТЕНДЕНЦИИ В ОБЛАСТИ БИЗНЕС-АНАЛИТИКИ

До недавнего времени подход к классификации технологий анализа данных выглядел следующим образом:

- 1) оперативный анализ данных;
- 2) интеллектуальный анализ данных.

Под оперативным анализом данных понимались, в первую очередь, технологии OLAP (On-line Analytical Procassing), а также техники drill-down, ad-hoc и т.п.

OLAP – это совокупность концепций, принципов и требований, лежащих в основе программных продуктов, облегчающих аналитикам доступ к данным [5].

Данная концепция была предложена Эдгаром Кодом в 1993 году. Принципы OLAP были описаны в статье «Обеспечение OLAP для пользователей – аналитиков» (Providing OLAP to User-Analysts: An IT Mandate). Всего было описано 12 правил, которым должна следовать технология, именуемая OLAP. В 1995 г. к ним было добавлено еще 6, кроме того правила были разбиты на четыре группы:

– В-основные правила:

- Многомерное концептуальное представление данных (Оригинальное правило 1).
- Интуитивное манипулирование данными (Оригинальное правило 10).
- Доступность: OLAP как посредник (Оригинальное правило 3).
- Пакетное извлечение против интерпретации (Новое).
- Модели анализа OLAP (Новое).

- Архитектура «клиент-сервер» (Оригинальное правило 5).
- Прозрачность (Оригинальное правило 2).
- Многопользовательская поддержка (Оригинальное правило 8).
- S-специальные особенности:
 - Обработка ненормализованных данных (Новое).
 - Сохранение результатов OLAP: хранение их отдельно от исходных данных (Новое).
 - Исключение отсутствующих значений (Новое).
 - Обработка отсутствующих значений (Новое).
- R-особенности представления отчетов:
 - Гибкость формирования отчетов (Оригинальное правило 11).
 - Стандартная производительность отчетов (Оригинальное правило 4).
 - Автоматическая настройка физического уровня (Замена оригинального правила 7).
- D-управление измерениями:
 - Универсальность измерений (Оригинальное правило 6).
 - Неограниченное число измерений и уровней агрегации (Оригинальное правило 12).
 - Неограниченные операции между размерностями (Оригинальное правило 9).

Drill-down (детализация данных) – это специфическая аналитическая техника, которая позволяет пользователю перемещаться по различным уровням данных, от агрегированных до детализированных. Пути углубления иерархией измерений или другими связями. Техника используется в OLAP-кубах.

Ad-hoc – запросы используются в OLAP-кубах при формировании нерегламентированных отчетов. Это единоразовые запросы, используемые для решения конкретной проблемы.

Под интеллектуальным анализом данных подразумевались технологии Data Mining.

Термин Data Mining получил свое название от двух понятий: поиска ценной информации в большой базе данных (data) и добычи горной руды (mining). Оба процесса требуют или просеивания огромного количества сырого материала, или разумного исследования и

поиска искомых ценностей. Понятие Data Mining, появившееся в 1978 г., приобрело высокую коммерческую популярность лишь в 2000-х, после появления технологий BigData. Так как лишь на большом количестве информации можно построить качественные прогнозы и расчеты.

Все методы Data Mining можно поделить на два подхода: статистические методы и кибернетические.

Статистические методы представляют собой четыре взаимосвязанных раздела:

- предварительный анализ природы статистических данных (проверка гипотез стационарности, нормальности, независимости, однородности, оценка вида функции распределения, ее параметров и т.п.);
- выявление связей и закономерностей (линейный и нелинейный регрессионный анализ, корреляционный анализ и др.);
- многомерный статистический анализ (линейный и нелинейный дискриминантный анализ, кластерный анализ, компонентный анализ, факторный анализ и др.);
- динамические модели и прогноз на основе временных рядов.

К статистическим методам Data Mining относятся:

- Дескриптивный анализ и описание исходных данных.
- Анализ связей (корреляционный и регрессионный анализ, факторный анализ, дисперсионный анализ).
- Многомерный статистический анализ (компонентный анализ, дискриминантный анализ, многомерный регрессионный анализ, канонические корреляции и др.).
- Анализ временных рядов (динамические модели и прогнозирование).

Второе направление Data Mining – это множество подходов, объединенных идеями компьютерной математики и использования теории искусственного интеллекта. К этой группе относятся такие методы:

- искусственные нейронные сети (распознавание, кластеризация, прогноз);
- эволюционное программирование (в том числе алгоритмы метода группового учета аргументов);

- генетические алгоритмы (оптимизация);
- ассоциативная память (поиск аналогов, прототипов);
- нечеткая логика;
- деревья решений;
- системы обработки экспертных знаний.

Также, часто под бизнес-аналитикой подразумевают понятие Business Intelligence (BI). BI же характеризуют в трех аспектах:

- 1) Процесс превращения данных в информацию и знания о бизнесе для поддержки деятельности лица принимающего решения.
- 2) Информационные технологии, позволяющие превратить данные в информацию и знания.
- 3) Непосредственно, сами знания о бизнесе, добытые в результате углубленного анализа данных.

Если рассматривать технологическую часть понятия BI, то до последнего времени преобладало следующее понимание архитектуры аналитических систем (рис. 2.1):

1) Источники данных, куда могут входить любые мастер-системы, где формируются необходимые для анализа данные (ERP, CRM и др.), информация из открытых систем и даже бумажные документы (которые могут быть предварительно отсканированы и распознаны или же осуществлен ручной ввод информации).

2) ETL (Extract, Transform, Load) – процедуры, обеспечивающие выгрузку, трансформация, очистку и загрузку данных в хранилище данных.

3) Централизованное хранилище данных и возможные витрины.

4) Над хранилищем данных могут быть настроены различные средства бизнес-аналитики, которые используют для анализа данные хранилища:

- a. Системы формирования и мониторинга KPI (Key Performance Indicators – ключевые показатели деятельности организации) .
- b. Системы прогнозирования.
- c. Имитационное моделирование.
- d. OLAP – технологии.
- e. Системы формирования регламентной отчетности.
- f. Data Mining и др.

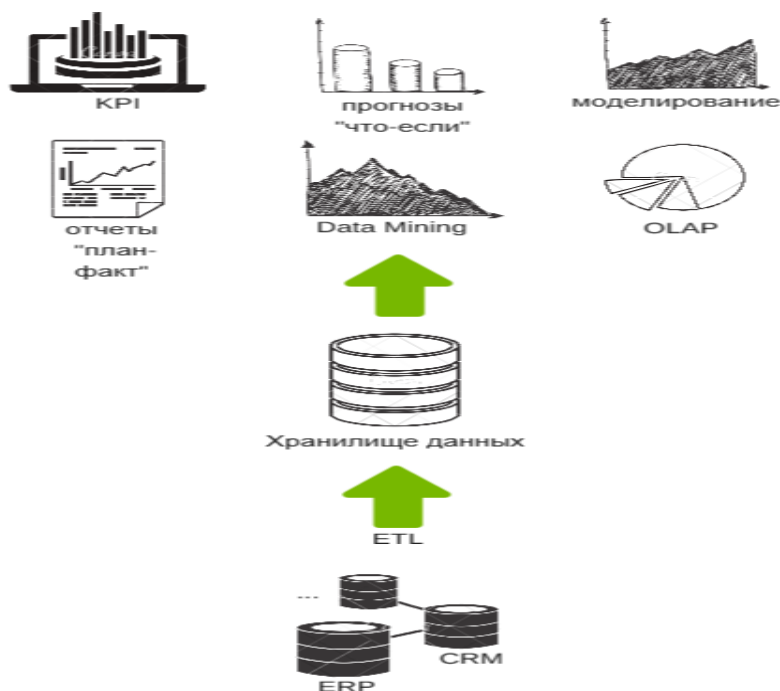


Рисунок 2-1 – Архитектура информационно-аналитической системы

В течении последний нескольких лет баланс сил с сфере бизнес-анализа сдвигался от ИТ-подразделений к бизнесу. В последние годы стала формироваться тенденция перехода от аналитических платформ для формирования отчетности до платформ самообслуживания. Современные системы бизнес-ориентированы и легко-масштабируемы под ландшафт поставщика [1].

Этот значительный сдвиг резко ускорился в последние годы, и, наконец, достиг переломной точки, которая требует нового взгляда на ВІ и аналитику.

Сегодня под системами бизнес-аналитики понимают информационные системы, позволяющие любым бизнес-пользователям пользоваться интуитивно-понятными средствами анализа для быстрого получения результата.

В перечне систем бизнес-аналитики появилось больше систем ориентированных на неподготовленных бизнес-пользователей: системы самостоятельной визуализации (Self-Service), системы Data Discovery, предикативная (прогнозная) аналитика, BigData.

Системы самостоятельной визуализации помогают бизнес-пользователю самостоятельно найти ответ на свой вопрос, представив информацию в понятном для восприятия виде, например, с помощью инфографики.



Рисунок 2.2 – BI-системы и аналитические платформы

Data Discovery – рассматривается Gartner как альтернативный подход к созданию аналитических решений по сравнению с традиционными промышленными системами бизнес-анализа. Обычно, решения типа Data Discovery предлагают пользователям интерактивный графический пользовательский интерфейс, базирующийся на архитектуре in-memory, что соответствует запросу бизнеса на простые и быстрые в работе BI-системы.

Что касается больших данных, то BigData – это зонтичный термин, подразумевающий серию подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объемов и значительного многообразия для получения воспринимаемых человеком результатов, эффективных в условиях непрерывного прироста, распределения по многочисленным узлам вычислительной сети, сформировавшихся в конце 2000-х гг., альтернативных традиционным системам управления базами данных и решениям класса Business Intelligence (Компания Gartner Group предпочитает термин Hadoop – ориентированные системы).

По результатам исследований Gartner Group к 2018 г., большинство бизнес-пользователей и аналитиков в организациях, будут иметь

доступ к «средствам самообслуживания» для подготовки данных для анализа в рамках перехода к развертыванию современных BI-платформ. Большинство автономных предложений подготовки данных самообслуживания будут реализованы в рамках одной платформы либо были интегрированы в качестве особенностей существующих аналитических платформ. Smart-приложения, приложения по управлению знаниями, BigData, системы визуализации и Data Discovery будут сходить в одной форме обнаружения данных следующего поколения, которые будут включать в себя подготовку данных самообслуживания и генерацию на естественном языке.

Однако, это изменение в фокусе BI в Magic Quadrant не должно интерпретироваться организациями в качестве рекомендации немедленно заменить все существующие технологии отчетности, построенные на основе единого хранилища данных, на системы Data Discovery.

Сдвиг на рынке BI дает возможность для появления инновационных подходов к BI, которые могут быть реализованы, как крупными игроками, так и новичками, открывается возможность для различных стартапов.

Переполненный рынок с большим количеством новых участников рынка, быстрое развитие и постоянные инновации создают сложную среду для поставщиков. Тем не менее, эти рыночные условия также создают возможность для покупателей для инвестиций в новые продукты, которые лучше подходят для работы в современных условиях, чем их предшественники.

ПРОЕКТИРОВАНИЕ ХРАНИЛИЩ ДАННЫХ С ERWIN DATA MODELER

3.1. ОСНОВЫ РАЗРАБОТКИ ХРАНИЛИЩ ДАННЫХ С ERWIN DATA MODELER

CA ERwin Data Modeler – один из инструментов моделирования данных, проектирования баз данных и хранилищ данных. CA ERwin Data Modeler Community Edition (CE) – бесплатная усеченная версия CA ERwin Data Modeler, которая специально разработана для исследования основных возможностей инструмента. В том числе для проведения практических занятий со студентами. Версия Community Edition поддерживает основные функции моделирования данных, проектирования баз данных и хранилищ данных, включая прямой и обратный инжиниринг, полное сравнение. Основное ограничение версии состоит в количестве объектов модели – не более 25.

Для разработки модели данных хранилища данных CA ERwin Data Modeler поддерживает на физическом уровне специализированную нотацию Dimensional Modeling – нотацию многомерного моделирования.

На основе разработанной модели хранилища можно сгенерировать объекты каталога базы данных на серверах различных платформ. Инструмент поддерживает возможность обратного инжиниринга каталога базы данных в модель физического уровня, а также возможность отслеживания изменений в проекте с помощью функционала Complete Compare (полное сравнение) [10, 11].

CA ERwin Data Modeler позволяет документировать разные источники для хранилища данных, включая реляционные и нереляционные источники. Инструмент позволяет документировать правила манипуляции данными. На основе документированной в модели информации встроенный построитель шаблонов ERwin Report Designer может сгенерировать отчеты различных форматах: HTML, TXT, PDF.

3.2. НАСТРОЙКА ПРОГРАММНОЙ СРЕДЫ ДЛЯ ПРАКТИКУМА

Для аудиторного выполнения практикума можно использовать на выбор одну из редакций приложения: CA ERwin Data Modeler 7.i или CA ERwin Data Modeler Community Edition 9.i.

Работа с приложением CA ERwin Data Modeler в аудитории может выполняться в одном из двух режимов:

1) с локально установленным на студенческий компьютер приложением;

2) с приложением, опубликованном в облаке Microsoft Azure в гибридной коллекции RemoteApp Финансового университета.

Для выполнения прямой генерации каталога хранилища данных из модели, а также обратной генерации и операций полного сравнения, потребуется подключение к базе данных в MS SQL Server. Параметры подключения преподаватель сообщает на занятии.

3.3. ПРАКТИКУМ ПО РАЗРАБОТКЕ ХРАНИЛИЩ ДАННЫХ С ERWIN DATA MODELER

Цель практикума

Закрепить теоретические знания и сформировать практические навыки проектирования хранилища данных с ERwin Data Modeler.

План практикума

Практикум включает следующие этапы:

1. Изучение бизнес-кейса, включающего краткое описание компании, характеристику системы автоматизации производства, требований руководства компании к аналитической системе.

2. Проектирование элементов витрины данных для аналитической системы согласно условиям и требованиям бизнес-кейса.

3. Создание проекта витрины данных в среде ERwin Data Modeler.

4. Генерацию каталога витрины данных в базу данных на MS SQL Server.

5. Сопровождение проекта витрины данных с ERwin Data Modeler, включая обратную генерацию каталога и полное сравнение каталога и модели витрины данных.

Порядок выполнения практикума

1. Изучите описание бизнес-кейса.

БИЗНЕС-КЕЙС

Компания MaxMin занимается производством и продажей небольших статуэток, раскрашиваемых вручную. Компания имеет несколько производственных направлений. Миниатюры изготавливаются из гипса, олова или алюминия. MaxMin распространяет свои товары по трем каналам. Компания содержит пять собственных розничных магазинов. Помимо этого компания владеет сайтом MaxMin.com, на котором ведется online-торговля, и осуществляет оптовые поставки сторонним дистрибьюторам.

Заместитель руководителя компании MaxMin по производству хочет иметь возможность анализировать статистику, предоставляемую системой автоматизации производства. Для анализа ему нужны не печатные отчеты, а интерактивный аналитический инструмент.

В настоящее время система автоматизации производства сохраняет все данные в текстовые файлы, а не в базу данных. Поэтому руководством поставлена задача спроектировать и построить витрину данных, которая станет хранилищем информации по производству.

В ходе производства изделий система автоматизации производства управляет всеми станками компании Max Min. Каждый станок (машина) реализует полный цикл производства изделий, включая

- заполнение формы сырьем (гипсом, оловом или алюминием);
- затвердевание материала;
- удаление изделия из формы после затвердевания;
- при необходимости автоматизированная раскраска изделий (оловянные фигурки не раскрашиваются);
- сушку после покраски (при необходимости).

Покраска и сушка могут производиться за несколько этапов в зависимости от сложности изделия. По мере готовности изделия проходят проверку, выполняемую оператором станка.

Оператор станка регистрируется в системе. В ходе этого процесса оператор сообщает системе автоматизации производства тип производимых изделий и объем загруженного в машину сырья. Оператор также делает в системе запись при отбраковке изделий.

В ходе беседы с заместителем руководителя по производству были выявлены следующие требования к данным, необходимые ему для эффективного принятия решений:

- число принятых изделий по объему сырья, видам изделий, машинам и дням;
- время формовки и затвердевания по видам изделий, машинам и дням;
- время покраски и сушки по типам краски, видам изделий, машинам и дням;
- сворачивание по подтипам изделий, которые сворачиваются по типам;
- сворачивание по типам машин, которые сворачиваются по материалам (гипс, олово или алюминий);
- сворачивание машин по фабрикам, которые сворачиваются по странам;
- сворачивание дней по месяцам, месяцев — по кварталам, кварталов — по годам;
- возможность фильтрации информации по производителю и дате покупки машины.

Анализ файла-экспорта из системы автоматизации производства показал, что для каждого вида производимых изделий есть отдельная строка, в которой присутствует следующая информация:

- тип изделия;
- объем сырья;
- номер машины;
- личный номер оператора;
- время и дата начала производства (когда серия начата);
- время и дата окончания производства (когда серия закончена);
- флаг отбраковки

2. Создайте документ MS Word, сохраните описания элементов витрины данных для аналитической системы в форме таблиц (табл. 3.1 – табл. 3.3).

Таблицы являются результатом анализа предметной области, консультаций с заместителем руководителя по производству.

Таблица 3.1 – Характеристика фактов для витрины данных

<i>№</i>	<i>Название колонки</i>	<i>Описание факта</i>
1	AcceptedProducts	Количество принятых изделий
2	RejectedProducts	Количество отбракованных изделий
3	ElapsedTimeForManufacture	Время производства в минутах

Таблица 3.2 – Характеристика измерений для витрины данных

<i>№</i>	<i>Название измерения</i>	<i>Содержание</i>	<i>Родительское измерение</i>
1	DimCountry	Страна	–
2	DimPlant	Завод	DimCountry
3	DimMachine	Машина	DimPlant, DimMachineType
4	DimMachineType	Тип машины	–
5	DimProduct	Изделие	DimProductSubType
6	DimProductSubType	Подтип изделия	DimProductType
7	DimProductType	Тип изделия	–
8	DimBatch	Партия	–

Таблица 3.3 – Характеристика колонок для модели витрины данных

<i>№</i>	<i>Имя колонки</i>	<i>Имя таблицы</i>	<i>FK</i>	<i>PK</i>	<i>Тип</i>
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>
1	AcceptedProducts	FactManufactoring	–	–	int
2	RejectedProducts	FactManufactoring	–	–	int
3	ElapsedTimeForManufacture	FactManufactoring	–	–	decimal(6,2)
4	MachineNumber	FactManufactoring	+	+	int
5	ProductCode	FactManufactoring	+	+	int
6	BatchNumber	FactManufactoring	+	+	int
7	DateOfManufacture	FactManufactoring	–	+	datetime
8	CountryCode	DimCountry	–	+	int
9	CountryName	DimCountry	–	–	varchar(30)
10	PlantNumber	DimPlant	–	+	int
11	PlantName	DimPlant	–	–	varchar(30)
12	CountryCode	DimPlant	+	–	int
13	MachineType	DimMachineType	–	+	int
14	MachineTypeDescription	DimMachineType	–	–	varchar(200)
15	MachineNumber	DimMachine	–	+	int
16	MachineName	DimMachine	–	–	varchar(50)
17	Manufacturer	DimMachine	–	–	varchar(50)
18	DateOfPurchase	DimMachine	–	–	datetime
19	MachineType	DimMachine	+	–	int
20	PlantNumber	DimMachine	+	–	int
21	ProductCode	DimProduct	–	+	int
22	ProductName	DimProduct	–	–	varchar(50)
23	ProductSubTypeCode	DimProduct	+	–	int
24	BatchNumber	DimBatch	–	+	int
25	BatchName	DimBatch	–	–	varchar(50)
26	ProductSubTypeCode	DimProductSubType	–	+	int

1	2	3	4	5	6
27	ProductSubTypeName	DimProductSubType	–	–	varchar(50)
28	ProductTypeCode	DimProductSubType	+	–	int
29	ProductTypeCode	DimProductType	–	+	int
30	ProductTypeName	DimProductType	–	–	varchar(50)

3. Запустите CA ERwin Data Modeler.

4. Создайте новую модель данных. Для этого в меню File выберите команду New. В открывшемся диалоговом окне установите тип новой модели «Physical» или «Logical/Physical», затем выберите в качестве целевого сервера SQL Server версии 2012/2014 как показано на рис. 3.1. Закройте диалоговое окно.

5. Настройте свойства модели данных. Для этого в меню Model выберите команду Model Properties. В открывшемся диалоговом окне в закладке General установите имя новой модели «Витрина данных Производства»; выберите нотацию моделирования IDEF1x для логического уровня модели, нотацию Data Warehousing для физического уровня; в секции Modeling Features установите опции Is Dimensional и Data Movement как показано на рис. 3.2. Не закрывайте диалоговое окно.

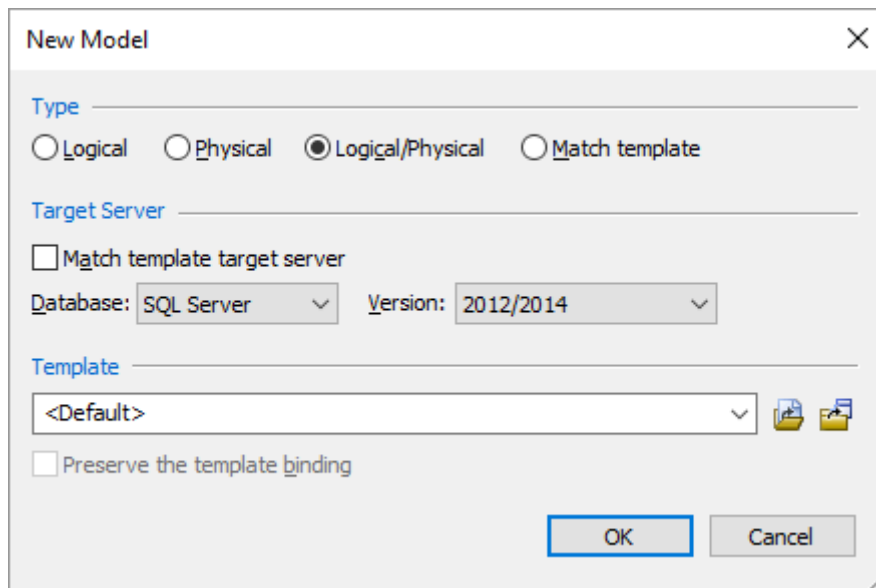


Рисунок 3.1 – Диалоговое окно для создания новой модели

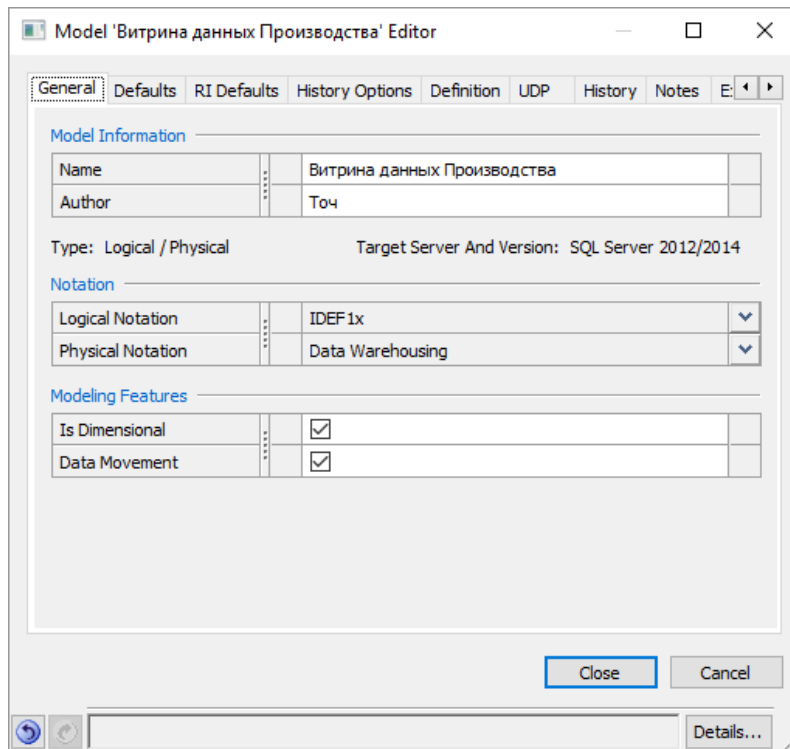


Рисунок 3.2. – Диалоговое окно для задания основных свойств модели

Переключитесь в закладку Defaults. В разделе Default Theme выберите в качестве темы по умолчанию Classic Theme как показано на рис. 3.3. Не закрывайте диалоговое окно.

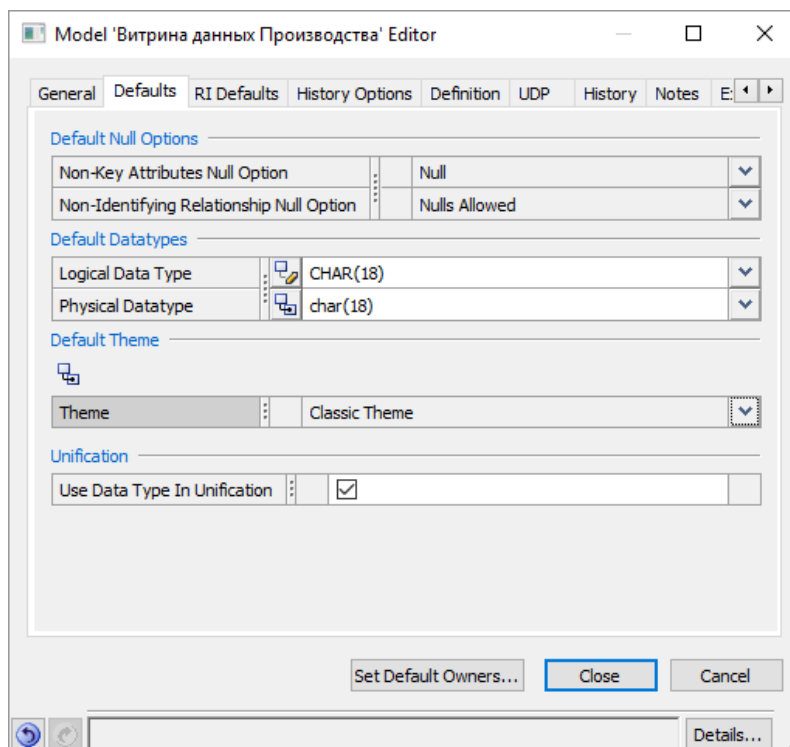


Рисунок 3.3 – Диалоговое окно для задания свойств модели по умолчанию

Переключитесь в закладку Definition. Введите описание модели данных: «Это модель данных Витрины данных Производства». Закройте диалоговое окно.

6. Настройте свойства текущей диаграммы. Для этого в меню Diagram выберите команду Diagrams. В открывшемся диалоговом окне в разделе Name измените имя диаграммы на «DM_Витрина данных Производства», выберите режим просмотра диаграммы Dimensional как показано на рис. 3.4. Затем переключитесь в закладку Display; отмените опцию отображения теней Display Shadows, установите режим рисования линий связи Diagonal как показано на рис. 3.4. Не закрывайте диалоговое окно.

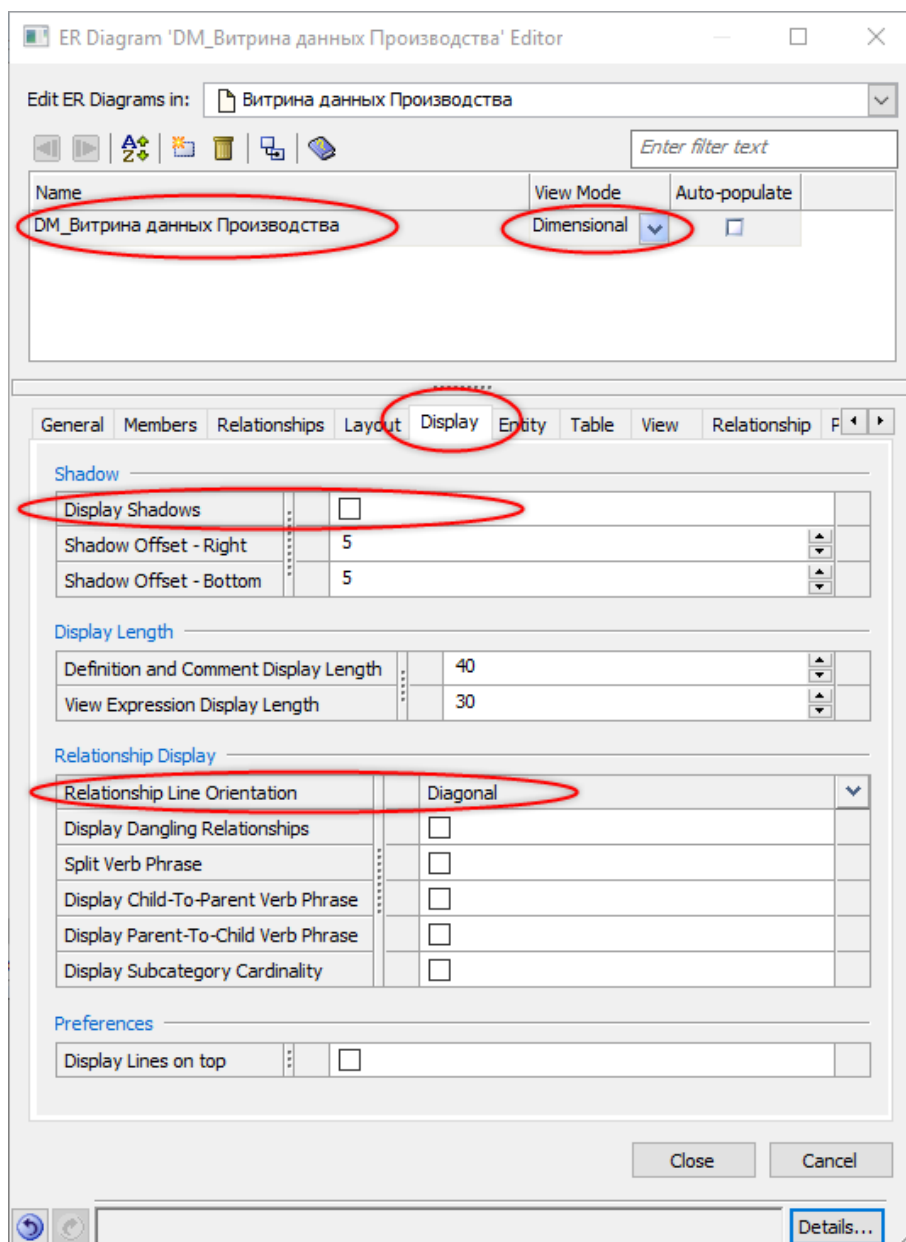


Рисунок 3.4 – Диалоговое окно для задания свойств диаграммы

Переключитесь в закладку Table. Установите опцию выделения внешних ключей Display Physical Foreign Key (FK) Designator как показано на рис. 3.5. Закройте диалоговое окно.

7. Настройте уровень отображения объектов диаграммы. Для этого в меню View выберите команду Display Level. В появившемся списке выберите режим отображения Column (рис. 3.6). Теперь для каждой таблицы будут отображаться: название, область первичных ключей, область остальных колонок (рис. 3.7).

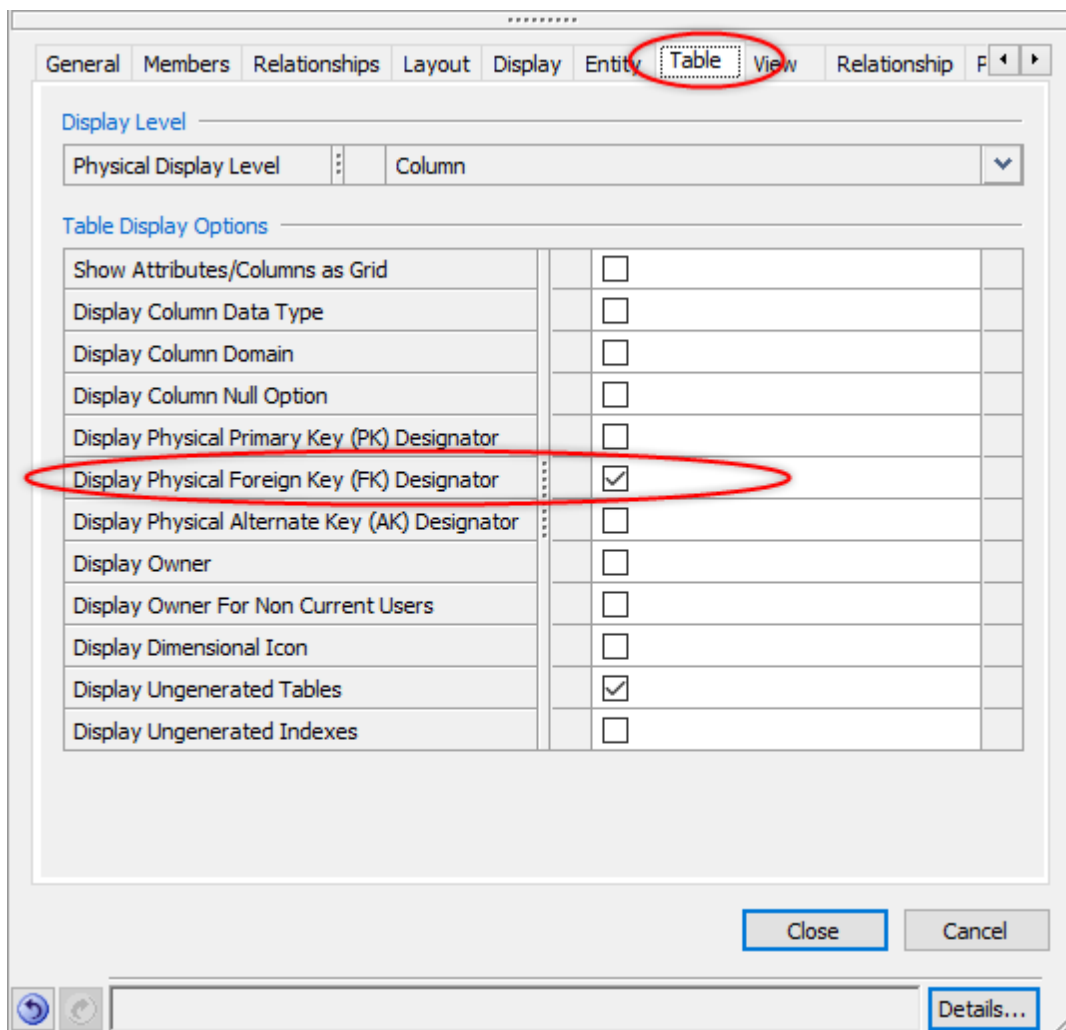


Рисунок 3.5 – Фрагмент диалогового окна для задания свойств диаграммы

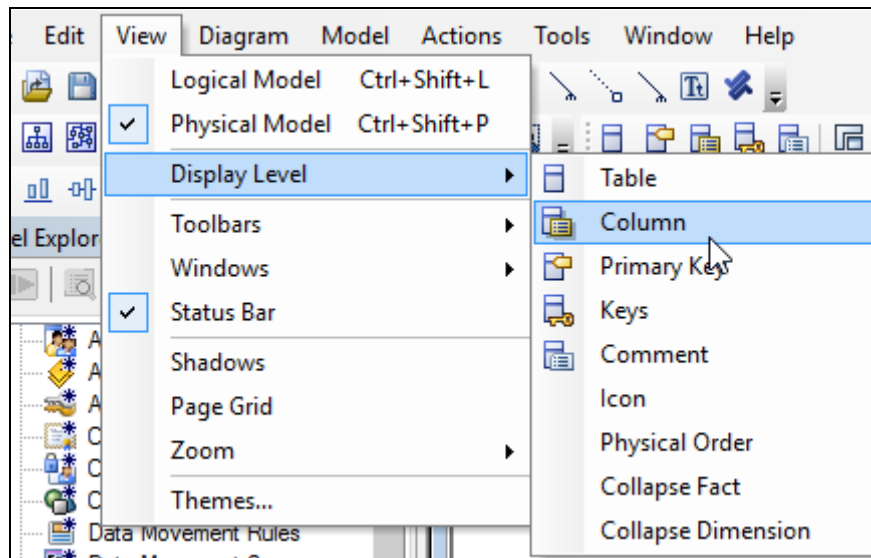


Рисунок 3.6. – Настройка уровня отображения объектов диаграммы

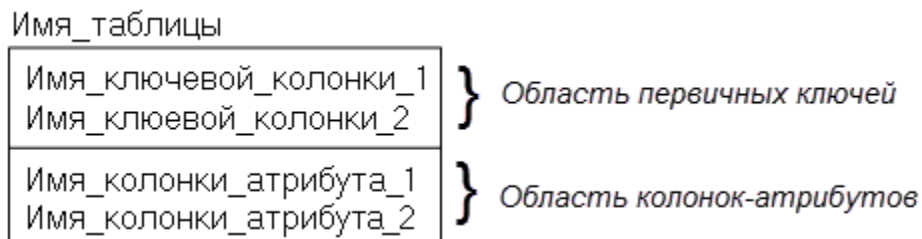


Рисунок 3.7 – Отображение таблицы в режиме Column

8. Поместите на диаграмму новую таблицу. Введите имя таблицы FactManufactoring. В область атрибутов добавьте три колонки согласно табл. 3.1: AcceptedProducts, RejectedProducts, ElapsedTimeForManufacture (рис. 3.8).

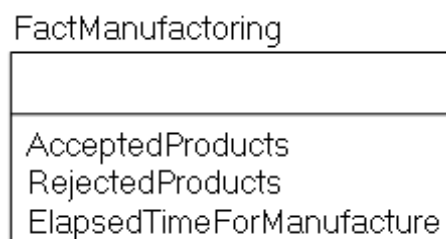


Рисунок 3.8 – Заготовка таблицы фактов

9. Добавьте на диаграмму еще восемь таблиц-измерений согласно табл. 3.2: DimCountry, DimPlant, DimMachine, DimMachineType, DimProduct, DimProductSubType, DimProductType, DimBatch.

10. Добавьте для каждой из таблиц измерений ключевые колонки и атрибуты как показано на рис. 3.9 согласно табл. 3.3. Для отбора ключевых колонок используйте столбец РК (первичный ключ) таблицы 3.3. Обратите внимание, что связи между таблицами на схеме отсутствуют. Кроме этого в таблицах измерений и таблицах фактов меньше колонок, чем представлено в табл. 3.3.

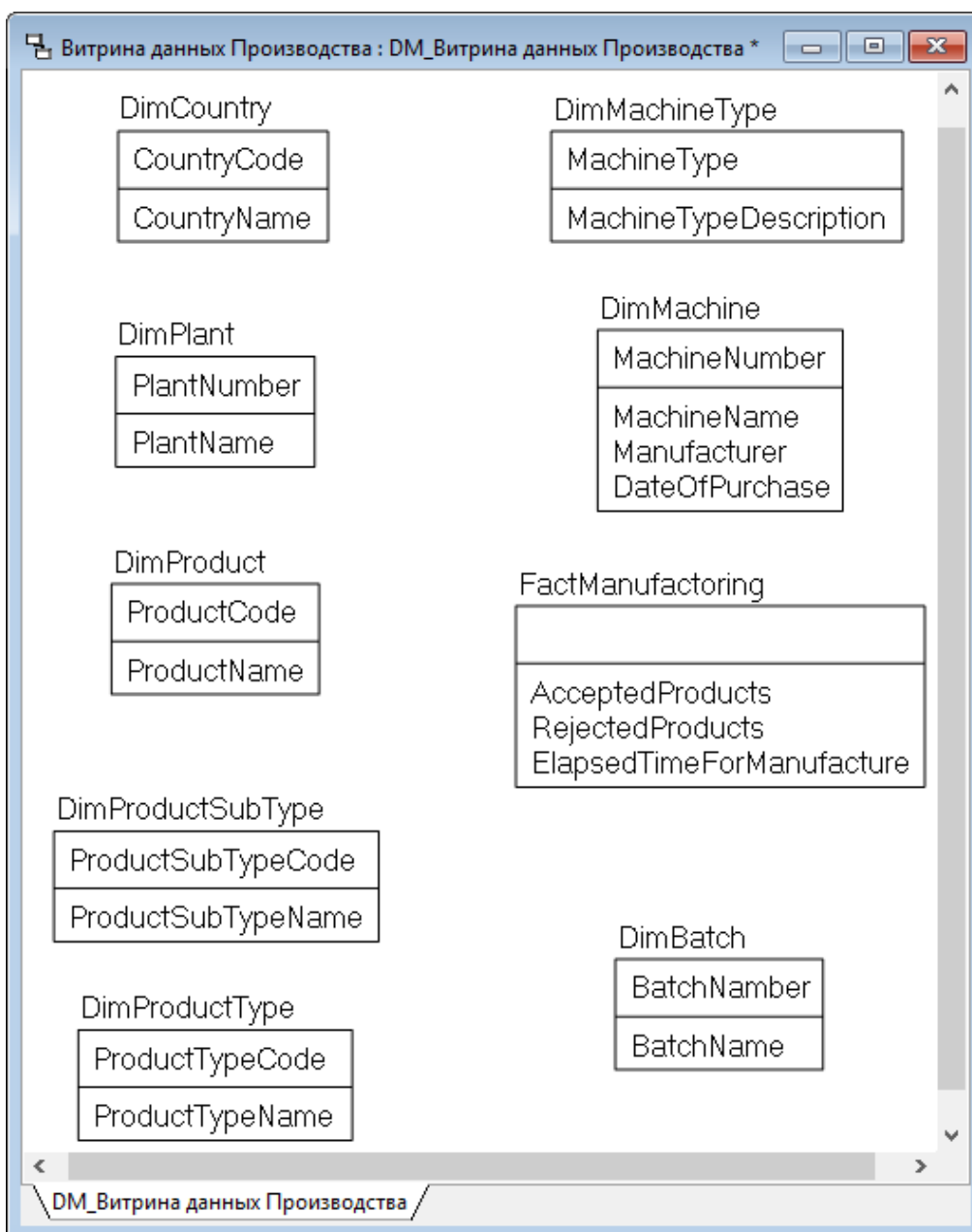


Рисунок 3.9 – Заготовка модели витрины данных

11. Добавьте в таблицу фактов FactManufacturing ключевую колонку DateOfManufacture.

12. Добавьте линии связи между таблицами согласно табл. 3.4. В результате модель примет вид как показано на рис. 3.10. Обратите внимание, что по идентифицирующей связи первичные ключи таблиц-измерений автоматически мигрировали в ключевую область таблицы-фактов, а по не идентифицирующей связи первичные ключи одних таблиц-измерений автоматически мигрировали в неключевую область дочерних таблиц-измерений.

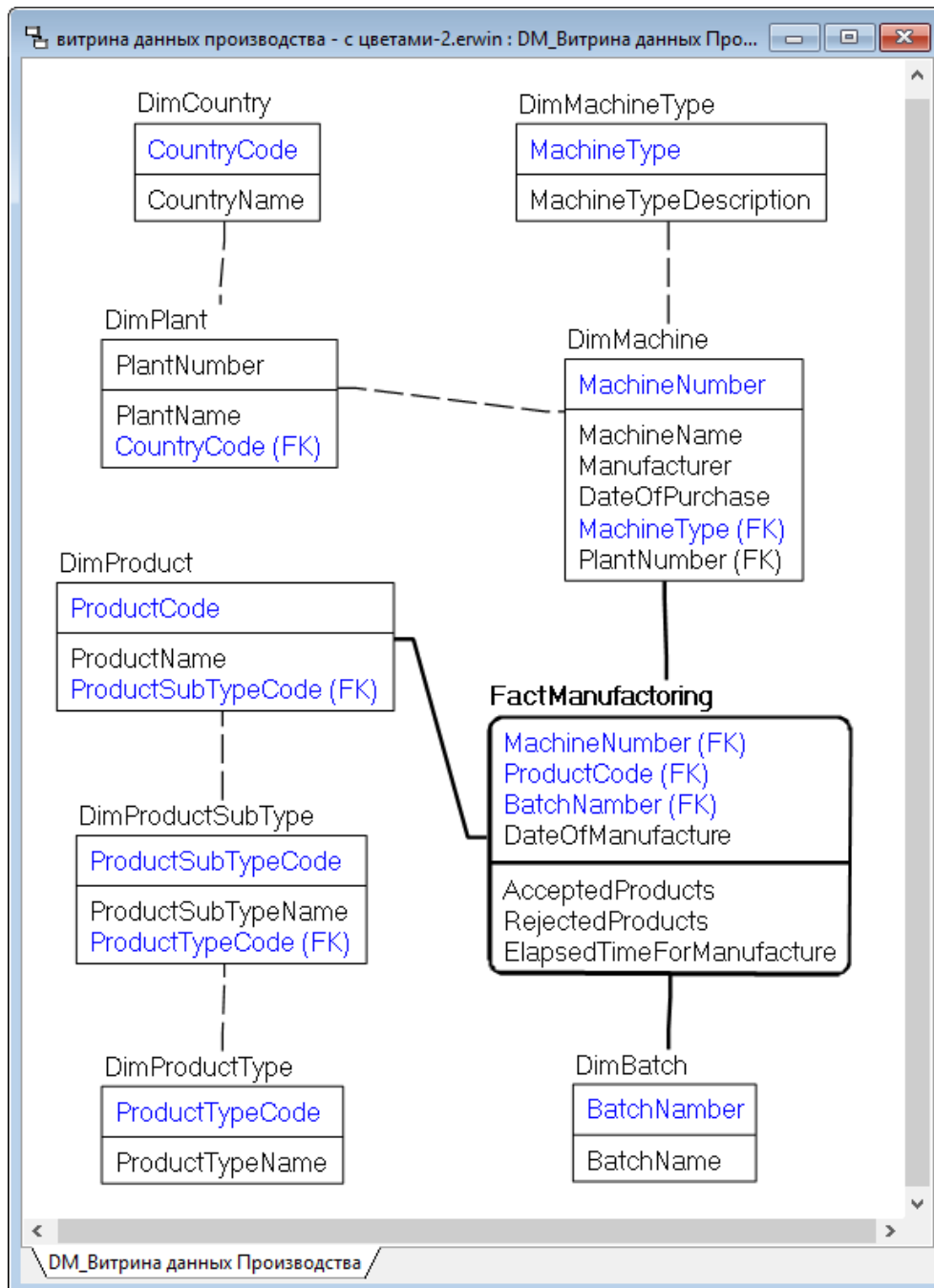


Рисунок 3.10 – Модель витрины данных Производства со связями

Таблица 3.4 – Характеристика отношений между таблицами

№	Таблица-родитель	Таблица-потомок	Тип связи
1	DimMachine	FactManufactoring	Идентифицирующая
2	DimBatch		
3	DimProduct		
4	DimCountry	DimPlant	Неидентифицирующая
5	DimPlant	DimMachine	Неидентифицирующая
6	DimMachineType	DimMachine	Неидентифицирующая
7	DimProductType	DimProductSubType	Неидентифицирующая
8	DimProductSubType	DimProduct	Неидентифицирующая

13. Скорректируйте настройки отображения диаграммы модели. Для этого в меню Diagram выберите команду Diagrams. В появившемся диалоговом окне выберите из списка диаграмму «DM_Витрина данных Производства»; переключитесь в закладку Table; затем в разделе Table Display Options установите опцию Display Column Data Type. Закройте диалоговое окно. В результате проведенных настроек на диаграмме витрины данных появилась информация о типах колонок.

14. Скорректируйте типы колонок. Для корректировки свойств колонок используйте навигатор модели Model Explorer как показано на рис. 3.11. В открывшемся диалоговом окне в выпадающем списке Table выберите таблицу, затем для этой таблицы выберите колонку, свойства которой требуется редактировать; скорректируйте тип данных колонки (рис. 3.12). Исправьте типы колонок для всех таблиц модели согласно табл. 3.3.

Полученная в результате модель витрины данных Производства с типами колонок представлена на рис. 1.3. Обратите внимание, что типы колонок – внешних ключей (содержат метку FK – Foreign Key) корректировать нельзя! Эти колонки меняются автоматически, после того, как вы изменили тип соответствующего первичного ключа в таблице на другом конце связи.

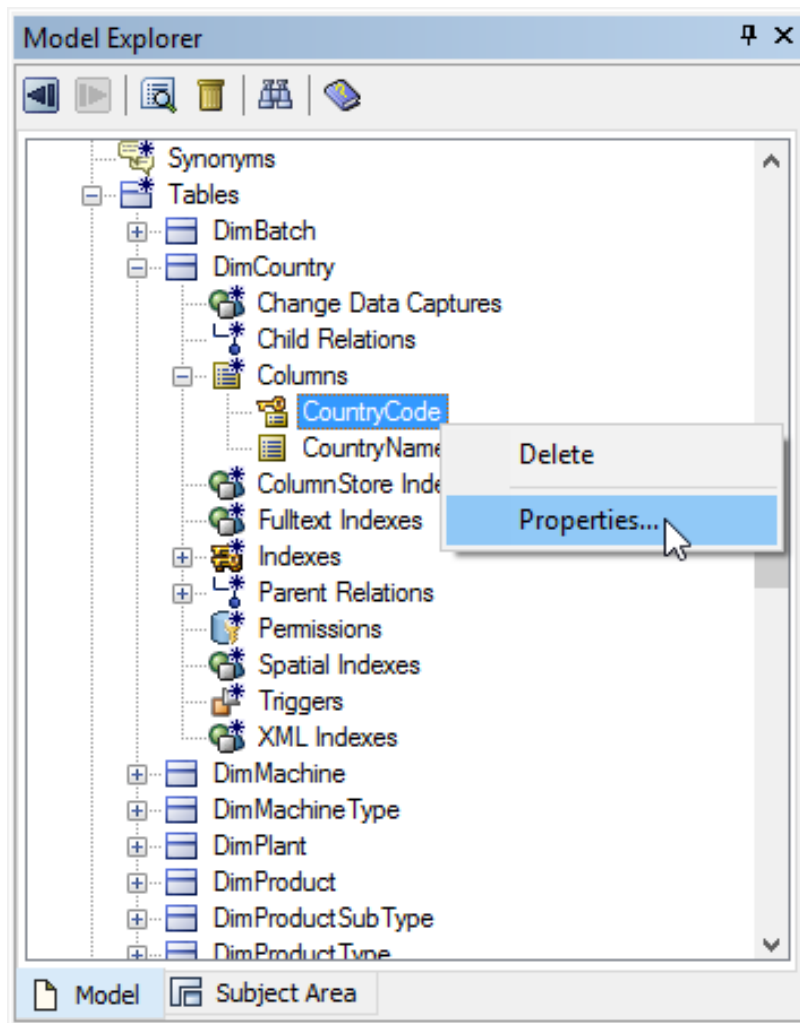


Рисунок 3.11 – Переход к редактированию свойств колонок в навигаторе модели

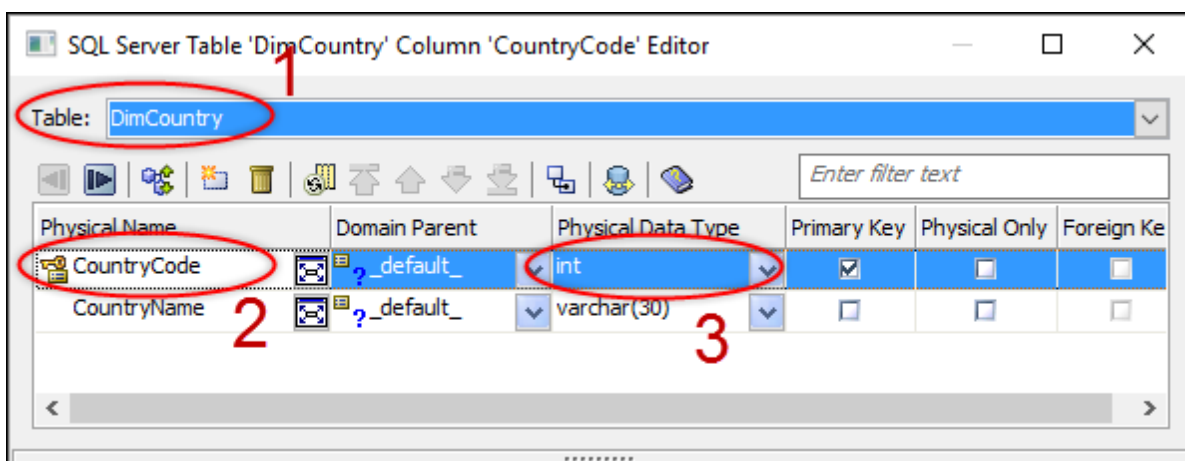


Рисунок 3.12 – Фрагмент диалогового окна редактирования свойств колонок

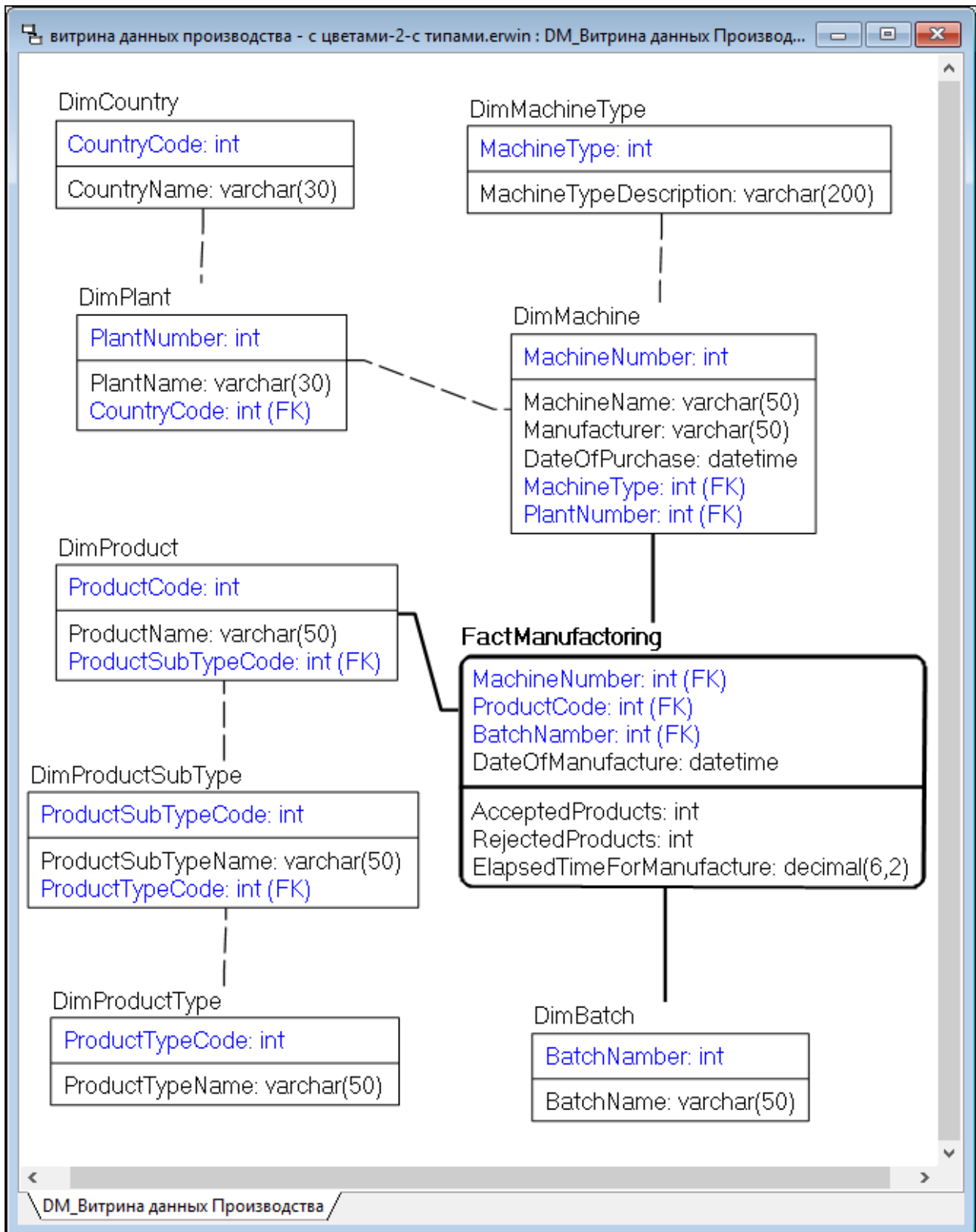


Рисунок 3.13 – Модель витрины данных Производства с типами колонок

15. Создайте шаблон отчета по разработанной модели витрины данных с помощью ERwin Report Designer. Для этого в меню Tools выберите команду Report Designer. В появившемся диалоговом окне выберите команду File\New Report. В открывшемся диалоговом окне Report Editor перейдите на закладку Report Design. В разделе Select Report Subject выберите Table. В правой части диалогового окна отобразится список возможных полей объекта таблица. Выберите для формирования шаблона отчета следующие поля: @TableName, Column Physical Name, Physical Data Type, Null Option. Не закрывайте диалоговое окно (рис. 3.14).

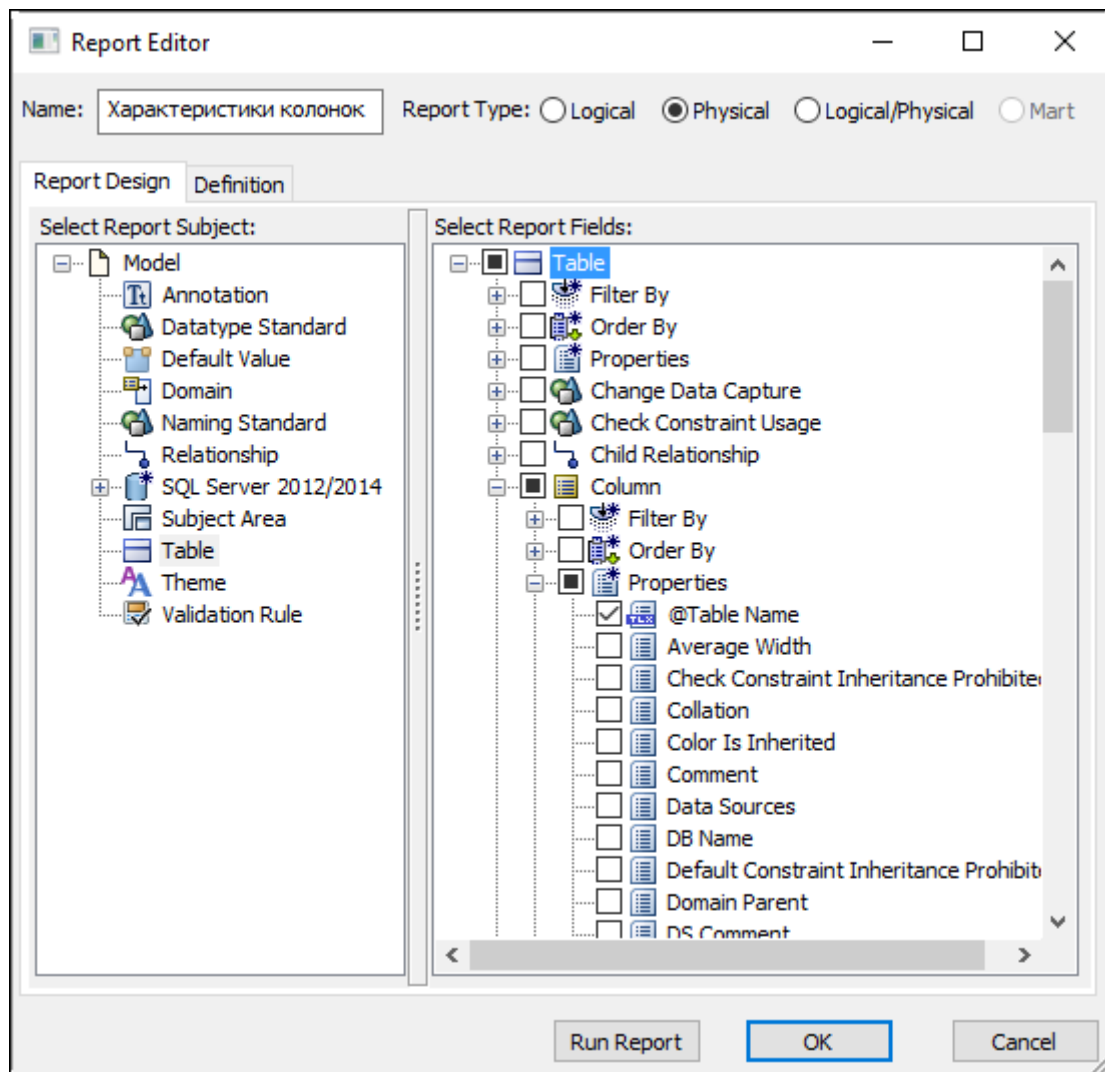
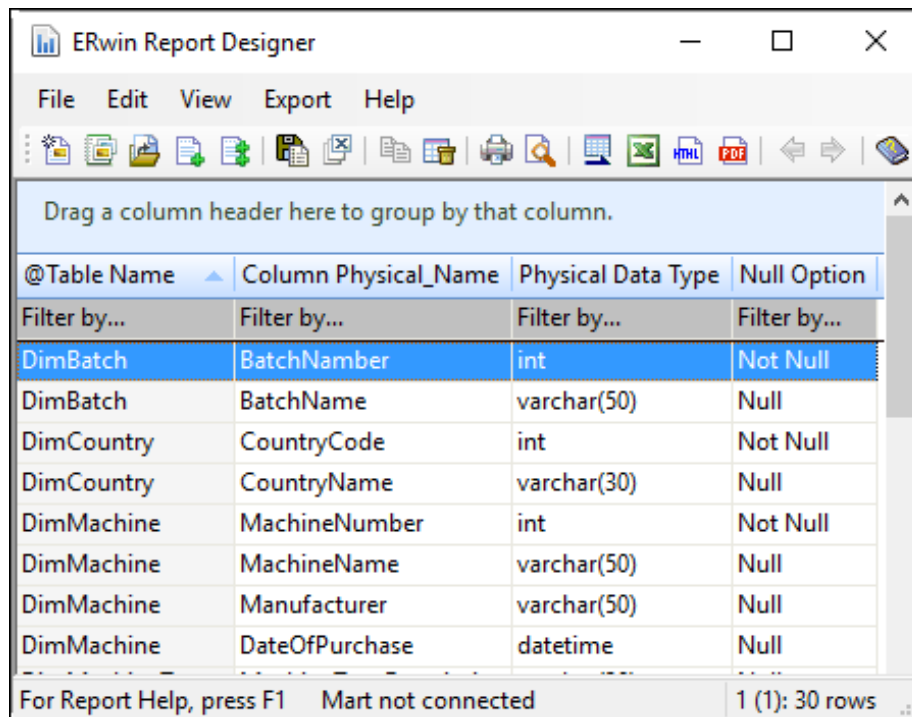


Рисунок 3.14 – Фрагмент шаблона отчета в ERwin Report Designer

16. Сгенерируйте отчет по модели витрины данных Производства на основе разработанного шаблона. Для этого в диалоговом окне Report Editor нажмите кнопку Run Report (рис. 3.14). Фрагмент сгенерированного отчета представлен на рис. 3.15.



@Table Name	Column Physical_Name	Physical Data Type	Null Option
Filter by...	Filter by...	Filter by...	Filter by...
DimBatch	BatchNumber	int	Not Null
DimBatch	BatchName	varchar(50)	Null
DimCountry	CountryCode	int	Not Null
DimCountry	CountryName	varchar(30)	Null
DimMachine	MachineNumber	int	Not Null
DimMachine	MachineName	varchar(50)	Null
DimMachine	Manufacturer	varchar(50)	Null
DimMachine	DateOfPurchase	datetime	Null

Рисунок 3.15 – Фрагмент сгенерированного отчета в ERwin Report Designer

17. Ознакомьтесь с возможностью создания спецификаций источников данных для каждой колонки. Для этого в меню Model выберите пункт Data Source. В открывшемся диалоге можно создать и специфицировать источник данных (рис. 3.16).

18. Ознакомьтесь с возможностью создания правил манипулирования данными и с возможностью назначения созданных правил для конкретных таблиц (меню Model, пункт Data Movement Rules) (рис. 3.17).

19. Подключитесь к вашей базе данных на указанном преподавателем экземпляре сервера SM SQL Server (рис. 3.18).

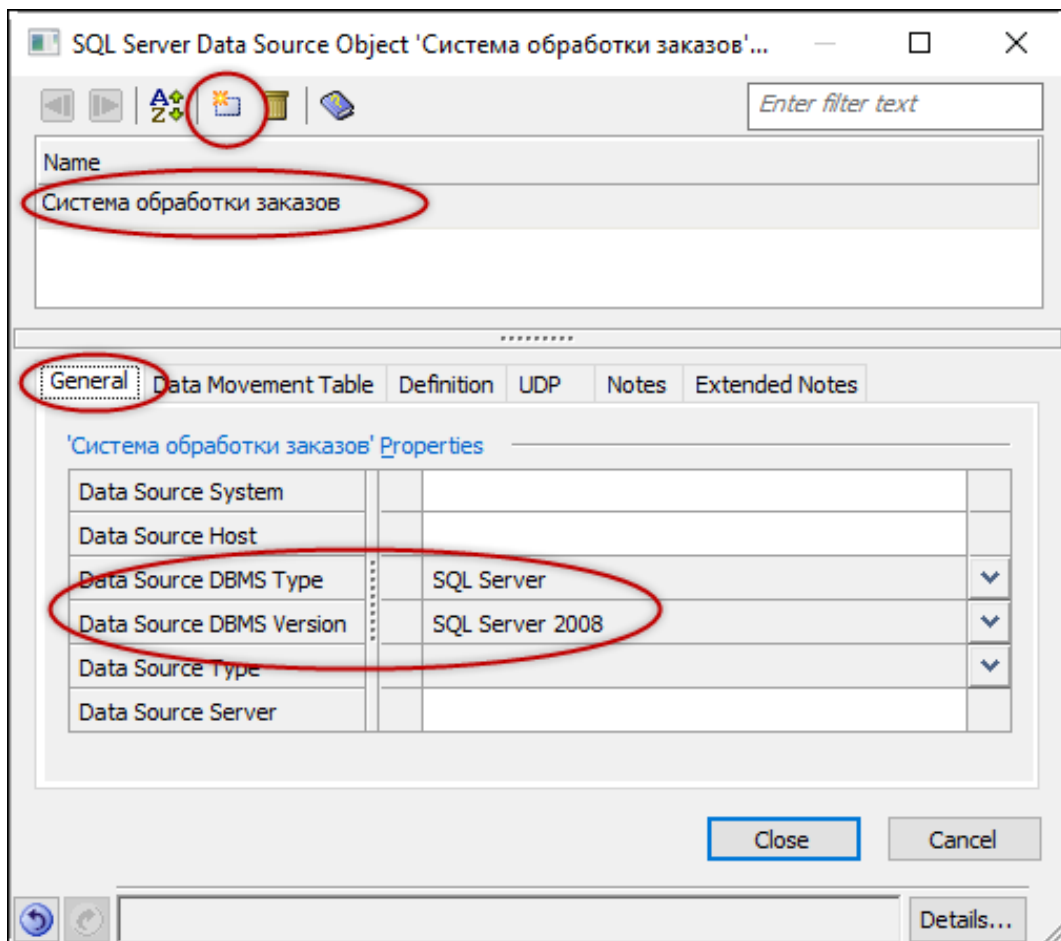


Рисунок 3.16 – Диалоговое окно для спецификации источника данных

20. Выполните прямую генерацию каталога витрины данных в вашу базу данных на MS SQL Server (меню Actions\Forward Engineering). Запустите MS SQL Server Management Studio; проверьте список сгенерированных таблиц.

21. Выполните обратную генерацию каталога витрины данных из MS SQL Server в модель физического уровня в CA ERwin Data Modeler (меню Actions\Reverse Engineering). Сгенерированную модель сохраните в личную папку.

22. Выполните полное сравнение (Complete Compare) каталога витрины данных из MS SQL Server и разработанной вами модели в CA ERwin Data Modeler физического уровня, предварительно добавив новую таблицу в вашу модель.

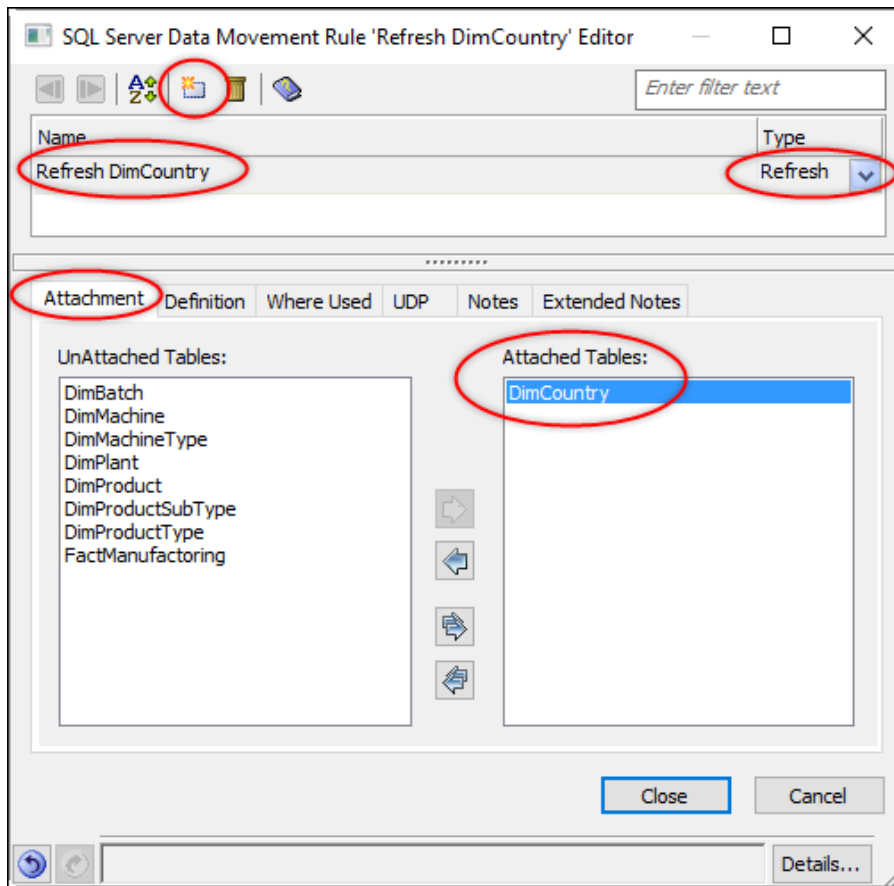


Рисунок 3.17 – Диалоговое окно для создания правил манипулирования данными

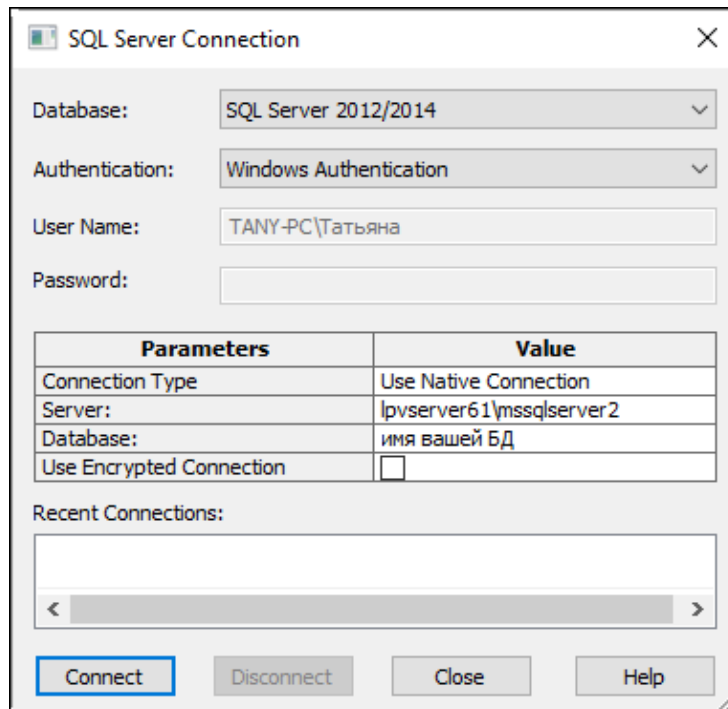


Рисунок 3.18 – Диалоговое окно для подключения к базе данных

3.4. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Каковы основные функциональные возможности СА ERwin Data Modeler в области проектирования и поддержки хранилищ данных?
2. В чем отличие редакции Community Edition?
3. К какому типу относится разработанная в рамках практикума схема витрины данных: звезда или снежинка?
4. Каковы недостатки разработанной в рамках практикума схемы витрины данных?
5. Каким образом в СА ERwin Data Modeler документируются источники данных для колонок?
6. С помощью каких инструментов СА ERwin Data Modeler можно поддерживать в актуальном состоянии модель хранилища данных?

3.4. ЗАДАНИЕ ДЛЯ САМОСТОЯТЕЛЬНОГО ВЫПОЛНЕНИЯ

Для внеаудиторного самостоятельного выполнения практического задания можно использовать приложение СА ERwin Data Modeler, опубликованное в облаке Microsoft Azure RemoteApp в гибридной коллекции Финансового университета. Для использования опубликованного в коллекции приложения требуется выполнить инструкции, представленные по адресу <http://www.fa.ru/dep/dit/projects/Pages/AzureRemoteAPP.aspx>. Другой вариант самостоятельной работы предполагает установку на домашний компьютер приложения СА ERwin Data Modeler Community Edition. Для установки приложения требуется заполнить анкету по адресу <https://www.ca.com/us/register/forms/ca-erwin-data-modeler-community-edition-evaluation-software.aspx?nu=y>. По указанному адресу электронной почты придет письмо от компании СА, в котором будет указан адрес для скачивания приложения.

В результате самостоятельного выполнения задания должен быть создан отчет, содержащий следующие разделы.

1. Описание бизнес-кейса, включая краткую характеристику деятельности компании; описание проблем бизнеса, требующих разработки аналитической системы на основе хранилища данных; характеристику действующих в компании OLTP-систем.

1. Проект витрины данных для аналитической системы, включая описание предметной области для витрины данных; формулировку основного вопроса, для ответа на который разрабатывается аналитическая система; список вариантов вопросов, которые могут сформулировать лица, принимающие решения, работая с аналитической системой (не менее 5 вариантов вопросов); примеры структур аналитических отчетов, которые могут быть сгенерированы аналитической системой для лиц, принимающих решения (не менее двух структур); список фактов для таблицы фактов с определенными для каждого факта источником данных; список измерений с ключами и атрибутами с определенными для каждой колонки источником данных; разработать схему витрины данных в СА ERwin Data Modeler, используя нотацию Dimensional Modeling.

2. Шаблон отчета по разработанной модели витрины данных, созданный с помощью ERwin Report Designer, а также сгенерированный на основе этого шаблона отчет по модели.

3. Прямую генерацию каталога витрины данных на MS SQL Server, включая скриншоты фрагмента сгенерированного каталога и соответствующей диаграммы из MS SQL Server Management Studio.

4. Обратную генерацию каталога витрины данных из MS SQL Server в модель физического уровня в СА ERwin Data Modeler, включая скриншот модели данных.

5. Полное сравнение (Complete Compare) каталога витрины данных из MS SQL Server и модели физического уровня в СА ERwin Data Modeler (после внесения изменений в одну из сравниваемых сторон), включая скриншот сгенерированной ERwin таблицы результатов сравнения.

Отчет оформить как документ Word. В названии документа указать номер группы и фамилию. Документ сохранить в личной папке. К отчету приложить файлы всех разработанных моделей, шаблонов, отчетов.

РАЗРАБОТКА ХРАНИЛИЩ ДАнных С MS SQL SERVER

4.1. ОСНОВНЫЕ КОМПОНЕНТЫ MS SQL SERVER ДЛЯ ПРАКТИКУМА

В рамках практикума будут использоваться службы Analysis Services и Integration Services [2, 5, 8].

Службы MS SQL Server Analysis Services (SSAS) позволяют анализировать большие объемы данных. С их помощью можно проектировать, создавать и управлять многомерными структурами, которые содержат как детальные, так и агрегированные данные из нескольких источников данных.

Для управления OLAP-кубами и работы с ними используется среда SQL Server Management Studio. Для создания новых кубов используется среда MS SQL Server Data Tools. Службы SSAS устанавливаются в процессе установки SQL Server вместе с другими компонентами SQL Server, или SSAS можно установить в виде отдельного компонента.

Службы MS SQL Server Integration Services (SSIS) представляют собой платформу для построения высокопроизводительных решений интеграции данных и решений потока операций, включая операции извлечения, преобразования и загрузки (Extract, Transform, Load – ETL) для хранилищ данных.

Службы SSIS содержат графические инструменты и мастера для построения и отладки пакетов; задачи для выполнения функций потока операций, таких как:

- выполнение инструкций SQL и работа с сообщениями электронной почты;
- источники данных и назначения для извлечения и загрузки данных;

- преобразования для очистки, статистической обработки, слияния и копирования данных;
- службу управления, службу SSIS для администрирования пакетов служб SSIS;
- API-интерфейсы для программирования объектной модели служб SSIS.

К типичным случаям применения пакетов служб SSIS совместно с SSAS относят:

- слияние данных из разнородных хранилищ данных;
- заполнение хранилищ данных и витрин данных;
- очистку и стандартизацию данных.

4.2. НАСТРОЙКА ПРОГРАММНОЙ СРЕДЫ ДЛЯ ПРАКТИКУМА

Для аудиторного выполнения практического задания потребуются инструменты MS SQL Server Management Studio 2012 (или выше) и MS SQL Server Data Tools 2012 (или выше). Работать с этими приложениями в аудитории можно в одном из двух режимов:

- 1) с локально установленными на студенческий компьютер приложениями;
- 2) с приложениями, опубликованном в облаке Microsoft Azure в гибридной

Параметры подключения к базе данных аналитических служб MS SQL Server преподаватель сообщает на занятии.

4.3. ПРАКТИКУМ ПО РАЗРАБОТКЕ ХРАНИЛИЩ ДАНЫХ С MS SQL SERVER

Цель практикума

Закрепить теоретические знания и сформировать практические навыки разработки хранилища данных с MS SQL Server.

План практикума

Практикум включает следующие этапы:

1. Создание проекта аналитических служб на основе шаблона аналитических служб MS SQL Server.

2. Развертывание созданного проекта на указанном преподавателем экземпляре служб Analysis Services, в результате чего в экземпляре служб Analysis Services будут автоматически сгенерированы объекты, определенные в проекте; будет выполнена обработка объектов в экземпляре служб Analysis Services (куба и его измерений), в результате которой данные из базовых источников данных будут автоматически скопированы в объекты куба.

3. Анализ данных куба.

Порядок выполнения практикума

1. Запустите MS SQL Server Data Tools. Откроется диалоговое окно MS Visual Studio.

2. Создайте новый проект, выбрав команду New Project. В открывшемся диалоговом окне (рис. 4.1) выберите шаблон аналитических служб Analysis Services Multidimensional and Data Mining Project, введите имя нового проекта «Проект ХД», введите имя решения, включающего проект: «Решение ХД», определите месторасположение нового проекта. Нажмите кнопку ОК.

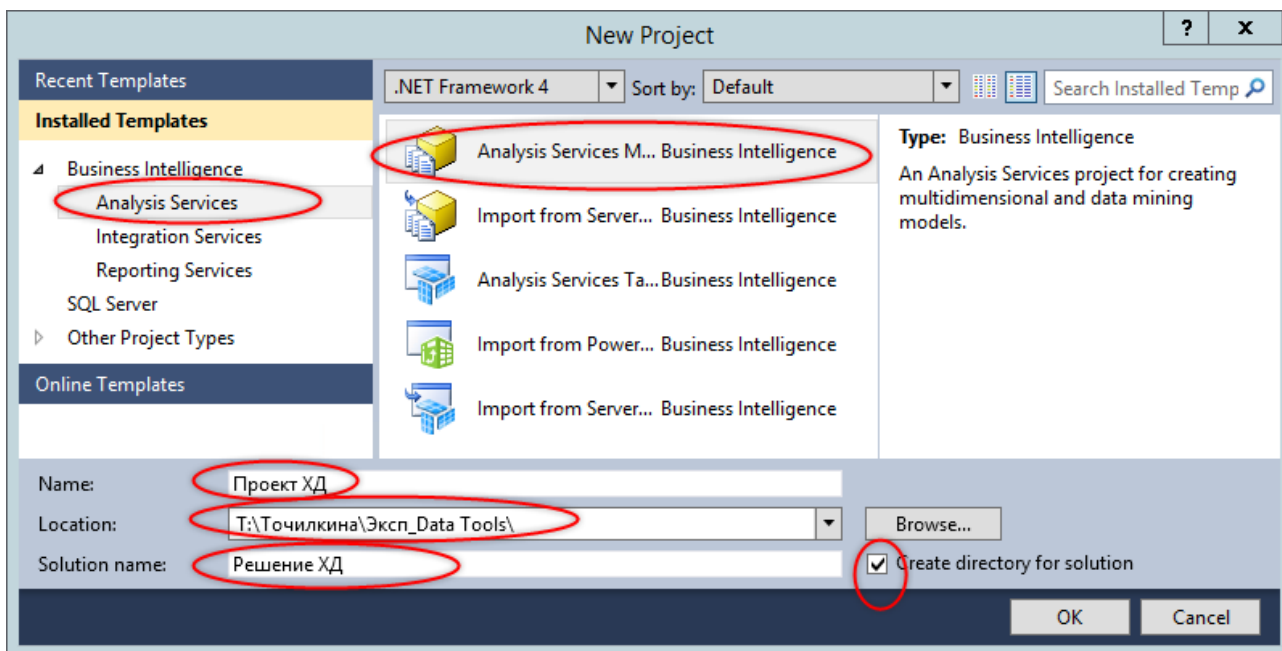


Рисунок 4.1 – Диалоговое окно создания проекта на основе шаблона

3. Определите источник данных. Для этого в обозревателе решения «Проект ХД» щелкните правой кнопкой мыши на элемент «Data Sources». В появившемся контекстном меню выберите пункт New Data Source. Откроется Мастер источников данных Data Source Wizard.

Нажмите кнопку Next. В открывшемся диалоге выберите опцию «Create a data source based on an existing or new connection» (рис. 4.2). В данном практикуме определим источник данных на основе нового соединения. Нажмите кнопку New.

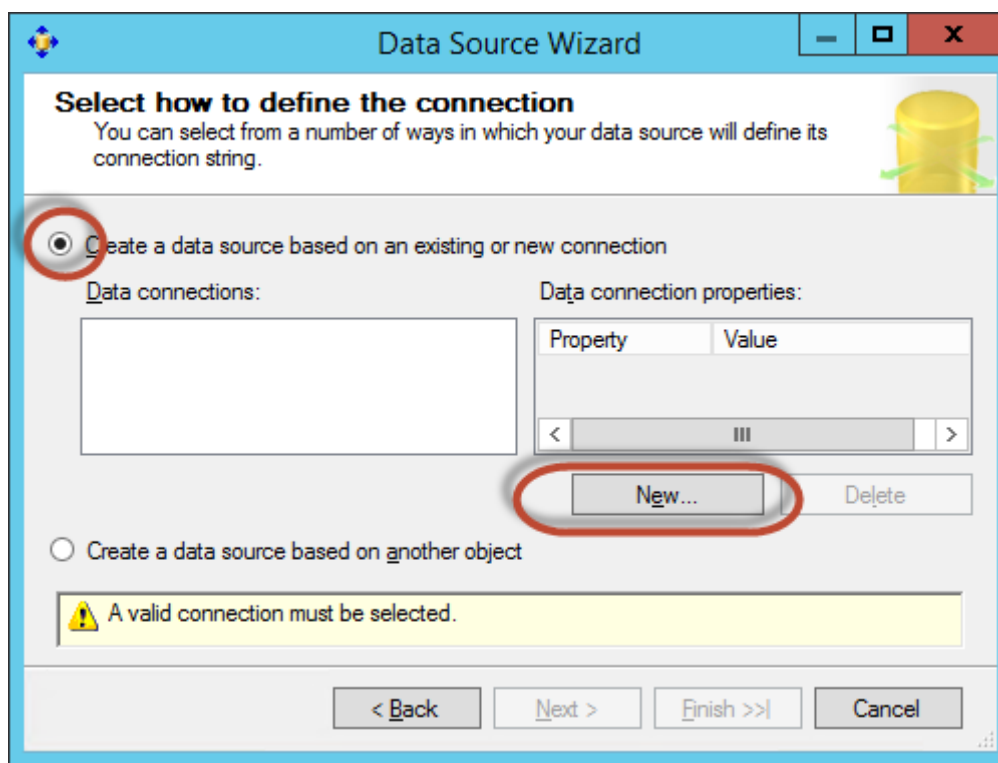


Рисунок 4.2 – Выбор метода подключения к источнику данных

В появившемся диалоговом окне Connection Manager (Диспетчер соединений) определите свойства соединения для источника данных как показано на рис. 4.3. Имя сервера: «lpvserver61\mssqlserver2», название базы данных «AdventureWorksDW2008». Протестируйте корректность соединения с источником данных, нажав на кнопку Test Connection. После успешного завершения проверки нажмите кнопку ОК.

В открывшемся диалоге нажмите кнопку Next (рис. 4.4). В новом диалоге выберите опцию «Use a specific user name and password», введите имя и пароль пользователя Windows для доступа к базе данных (рис. 4.5), нажмите кнопку Next.

В диалоге Complete the Wizard (Завершение работы мастера) введите имя источника данных: Adventure Works DW. Нажмите кнопку Finish (рис. 4.6).

В проекте «Проект ХД» появился новый источник данных (рис. 4.7).

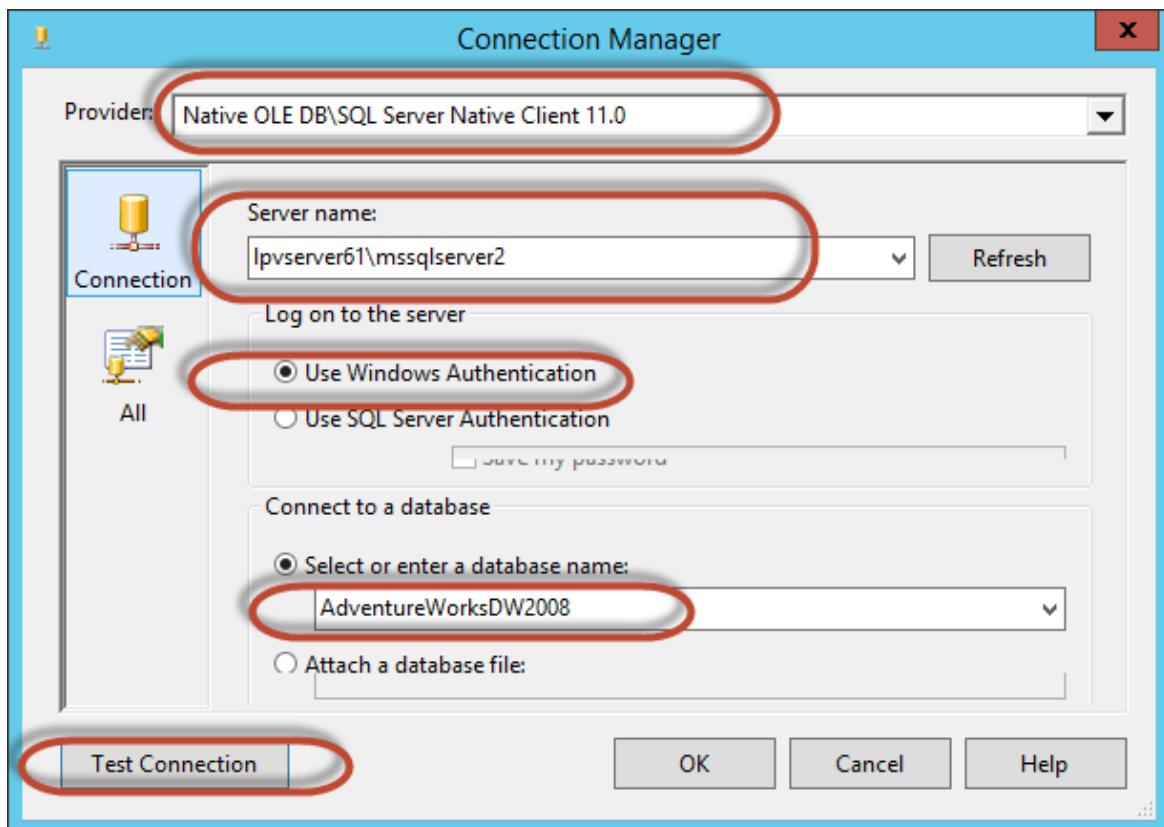


Рисунок 4.3 – Определение свойств соединения для источника данных

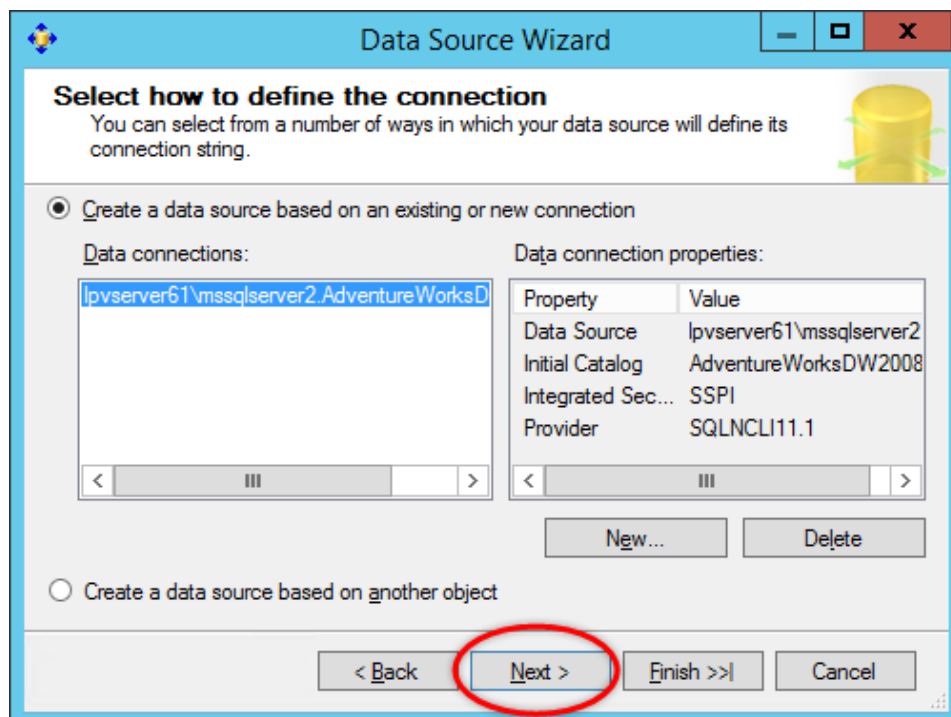


Рисунок 4.4 – Свойства подключения

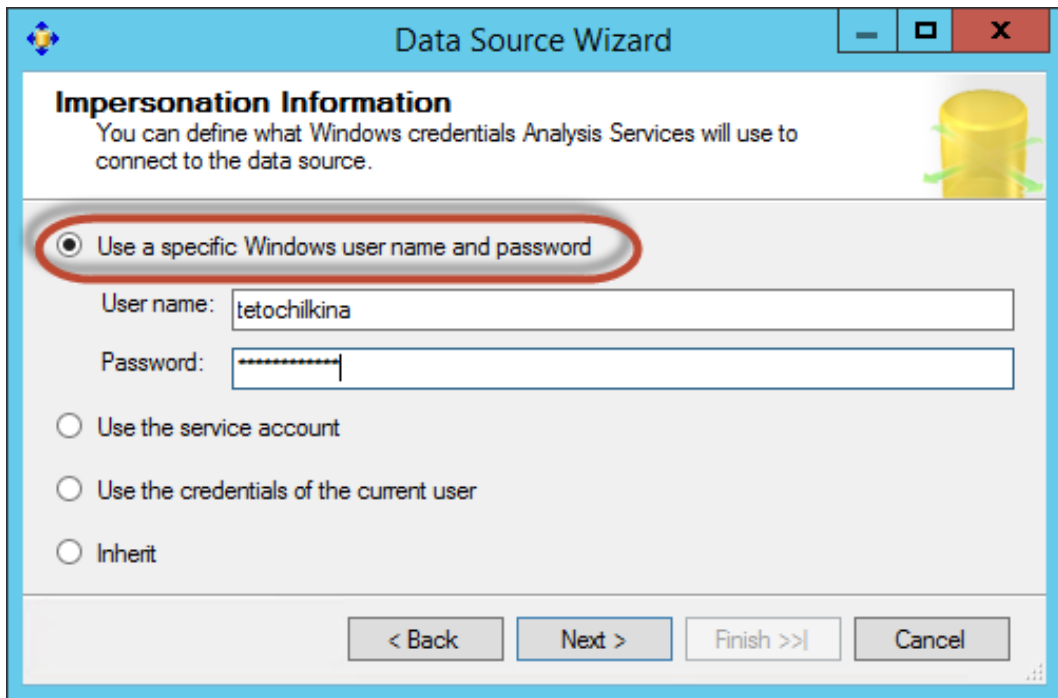


Рисунок 4.5 – Информация об олицетворении

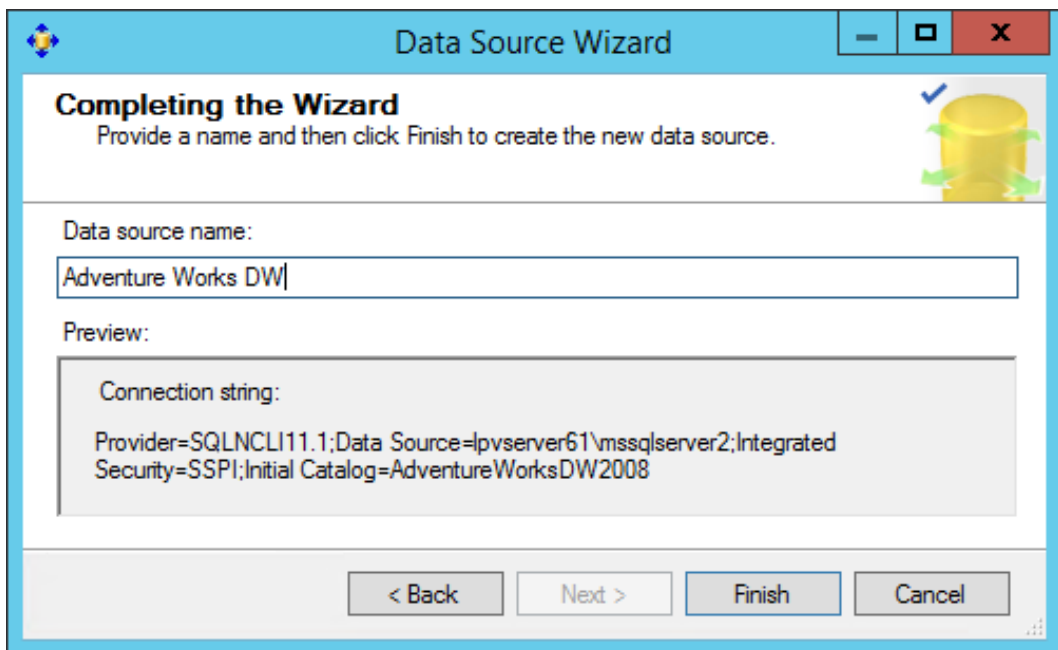


Рисунок 4.6 – Завершение создания источника данных

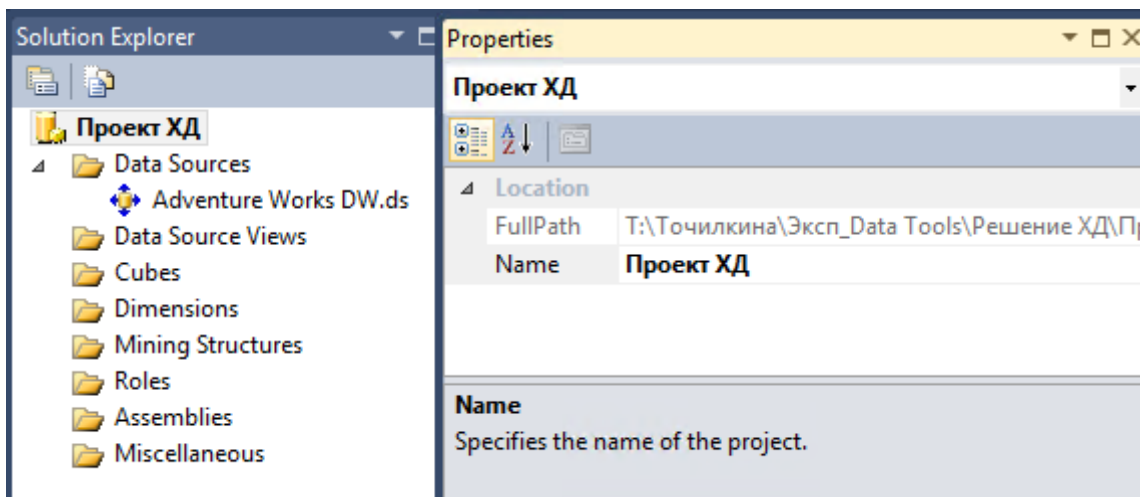


Рисунок 4.7 – Новый источник данных в проекте

4. Определите новое представление источника данных. Представление источника данных является отдельным представлением метаданных из указанных таблиц и представлений, определяемых источником данных для проекта. Хранение метаданных в представлении источника данных позволяет работать с метаданными в процессе разработки, не устанавливая соединений с базовыми источниками данных.

В обозревателе решения в проекте «Проект ХД» щелкните правой кнопкой мыши на элемент Data Source View. В появившемся контекстном меню выберите пункт New Data Source View. Откроется Мастер представления источников данных Data Source View Wizard. Нажмите кнопку Next.

Будет открыта страница Select a Data Source (Выбор источника данных). В разделе Relational Data Source (Источник реляционных данных) выберите источник данных «Adventure Works DW». Нажмите кнопку Next.

На открывшейся странице Select Tables and Views (Выбор таблиц и представлений) в области Available objects (доступные объекты) выберите 5 таблиц из базы данных Adventure Works DW: DimCustomer (dbo), DimDate (dbo), DimGeography (dbo), DimProduct (dbo), FactInternetSales (dbo).

Переместите эти таблицы в область Included objects (Включенные объекты) как показано на рис. 4.8. Нажмите кнопку Next.

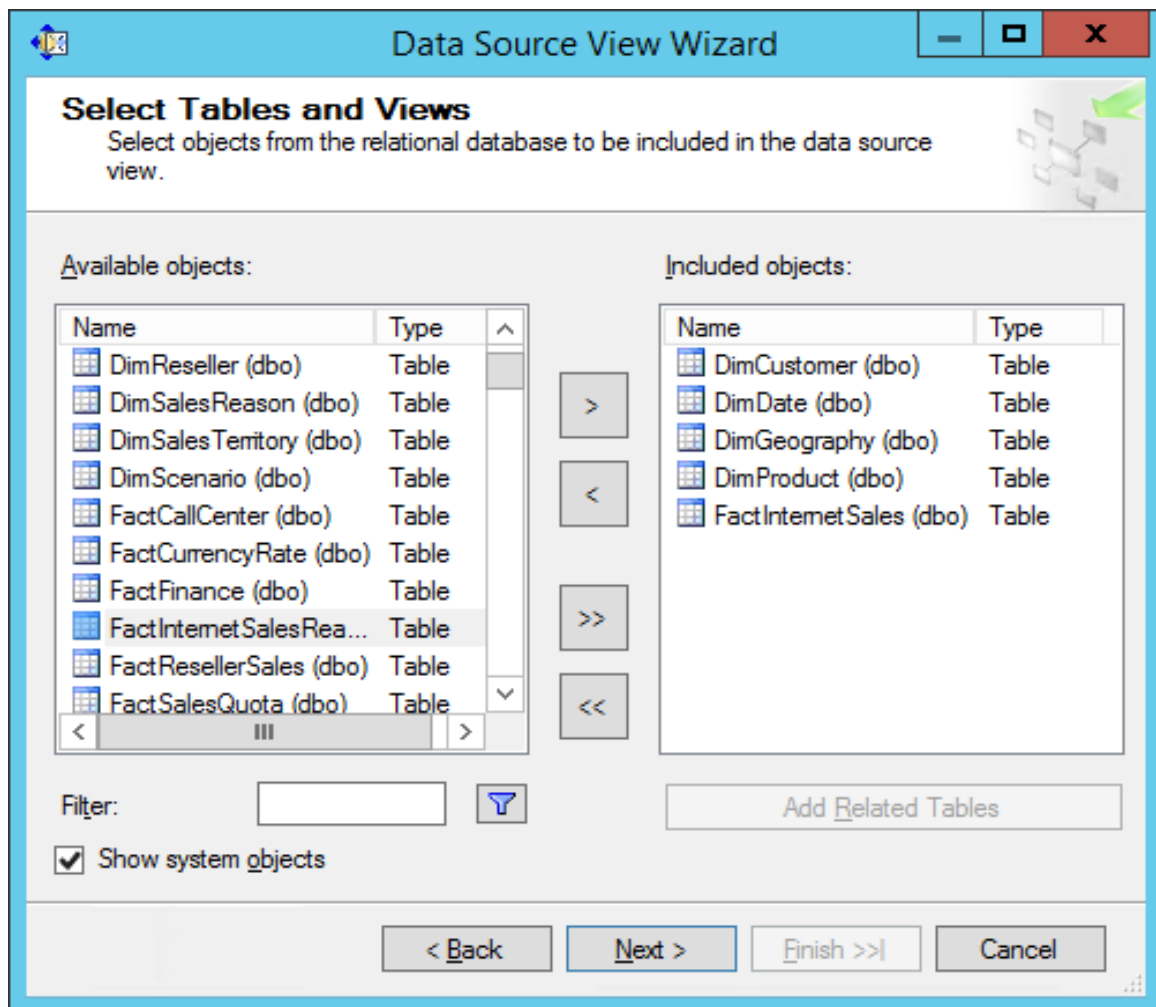


Рисунок 4.8 – Выбор таблиц для представления источника данных

На завершающей странице Мастера введите имя нового представления источника данных «Представление AdvWorksDW», нажмите кнопку Finish (рис. 4.9).

Информацию о созданном представлении источника данных удобно просматривать и редактировать в конструкторе представлений (рис. 4.10).

5. Измените имена таблиц по умолчанию. Для этого в конструкторе представлений в области диаграммы выберите требуемую таблицу и в окне Properties введите значение свойства Friendly Name согласно данным из таблицы 4.1. В результате изменения имен таблиц представление источника данных примет вид как показано на рис. 4.11.

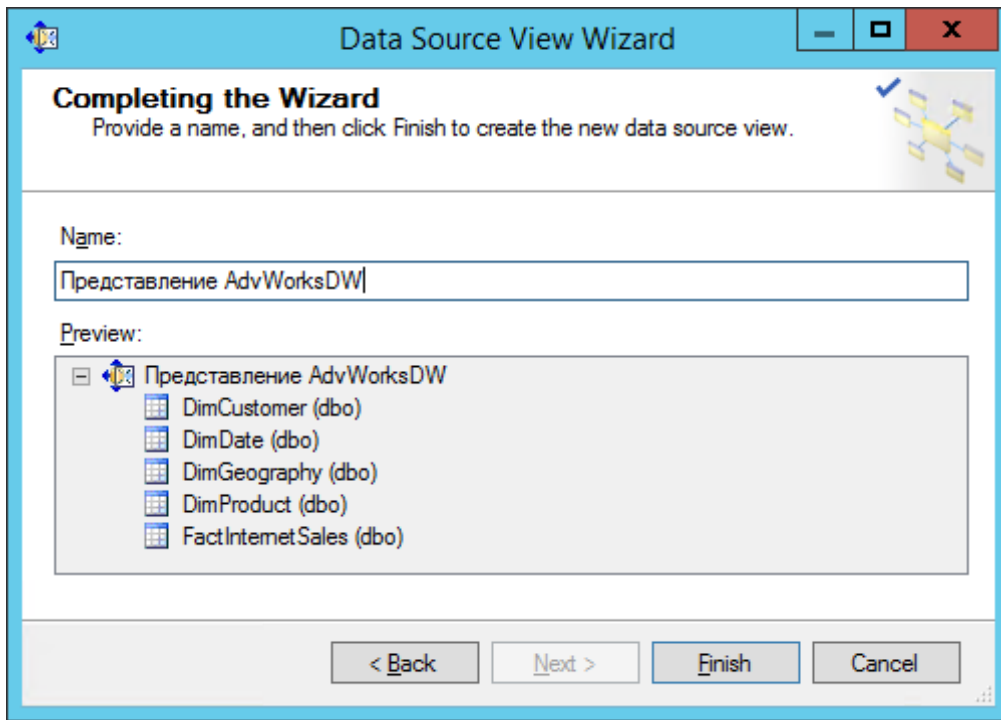


Рисунок 4.9 – Выбор таблиц для представления источника данных

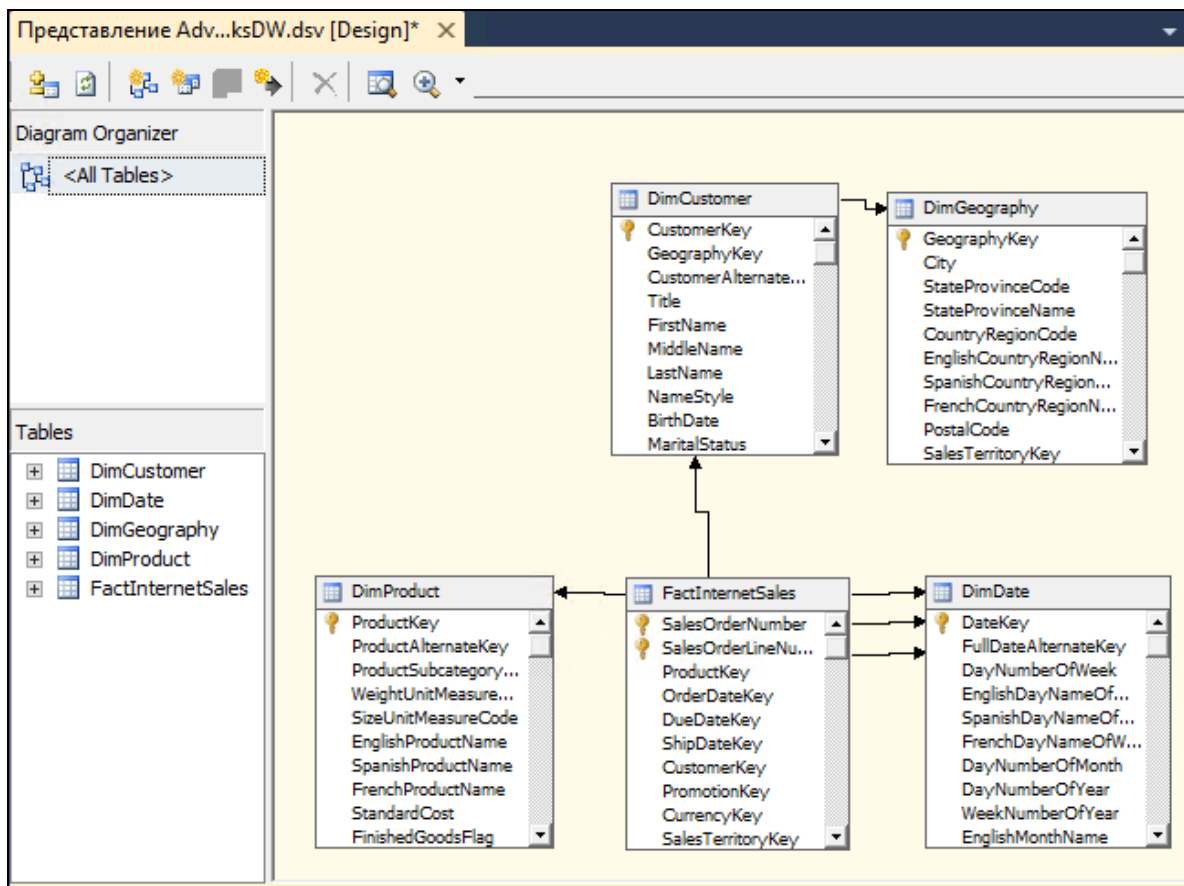


Рисунок 4.10 – Представление источника данных в конструкторе представлений

Таблица 4.1 – Удобные названия таблиц для представления источника данных

№	Название таблицы	Удобные названия
1	DimCustomer	Клиент
2	DimDate	Дата
3	DimGeography	География
4	DimProduct	Продукт
5	FactInternetSales	Продажи через Интернет

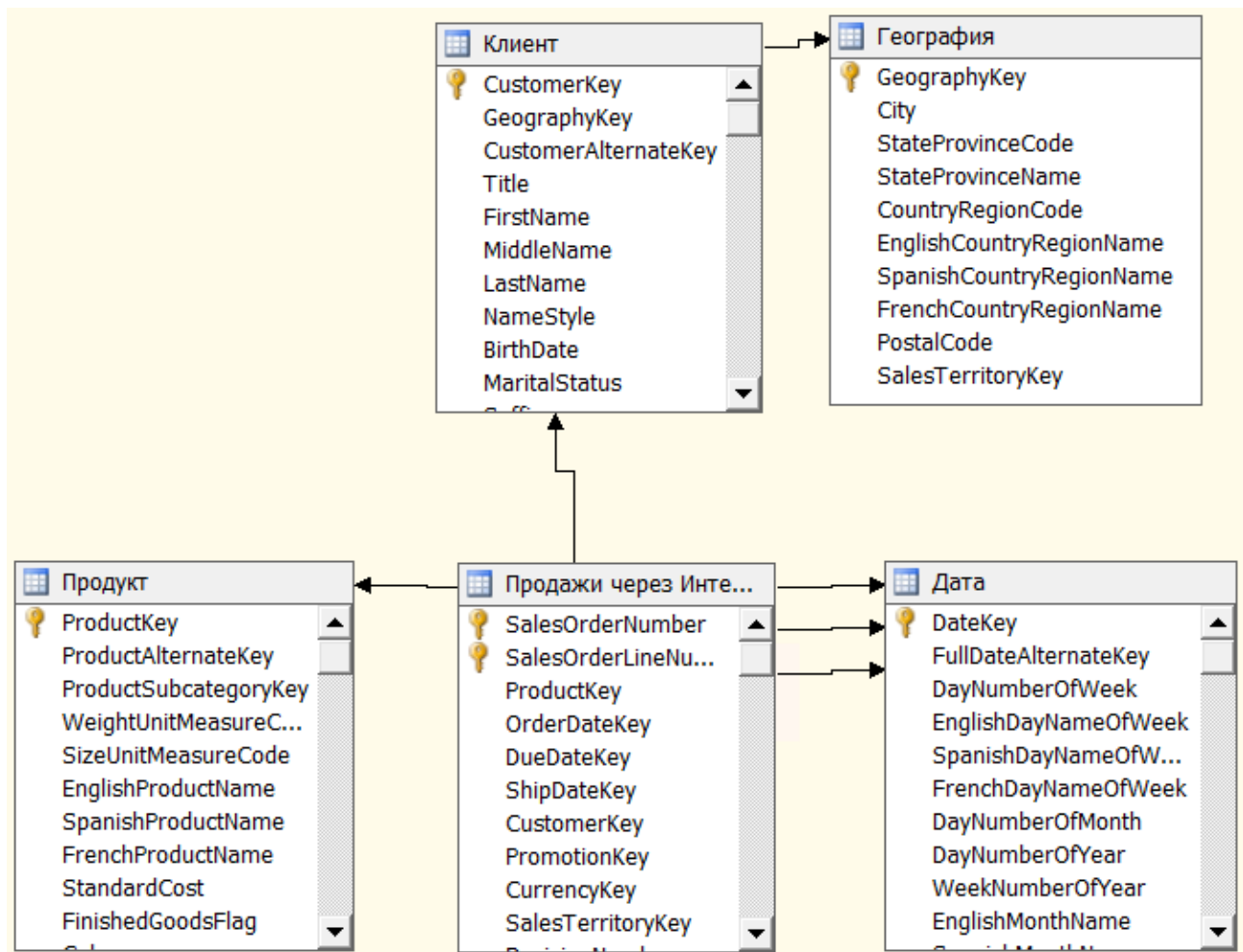


Рисунок 4.11 – Измененные имена таблиц представление источника данных

6. Создайте измерение Date. Для этого в обозревателе решений в проекте «Проект ХД» щелкните правой кнопкой мыши по разделу Dimensions, выберите команду New Dimension. В появившемся окне мастера нажмите кнопку Next. В диалоге Select Creation Method выберите Use an existing table (Использовать существующую таблицу). Нажмите кнопку Next (рис. 4.12).

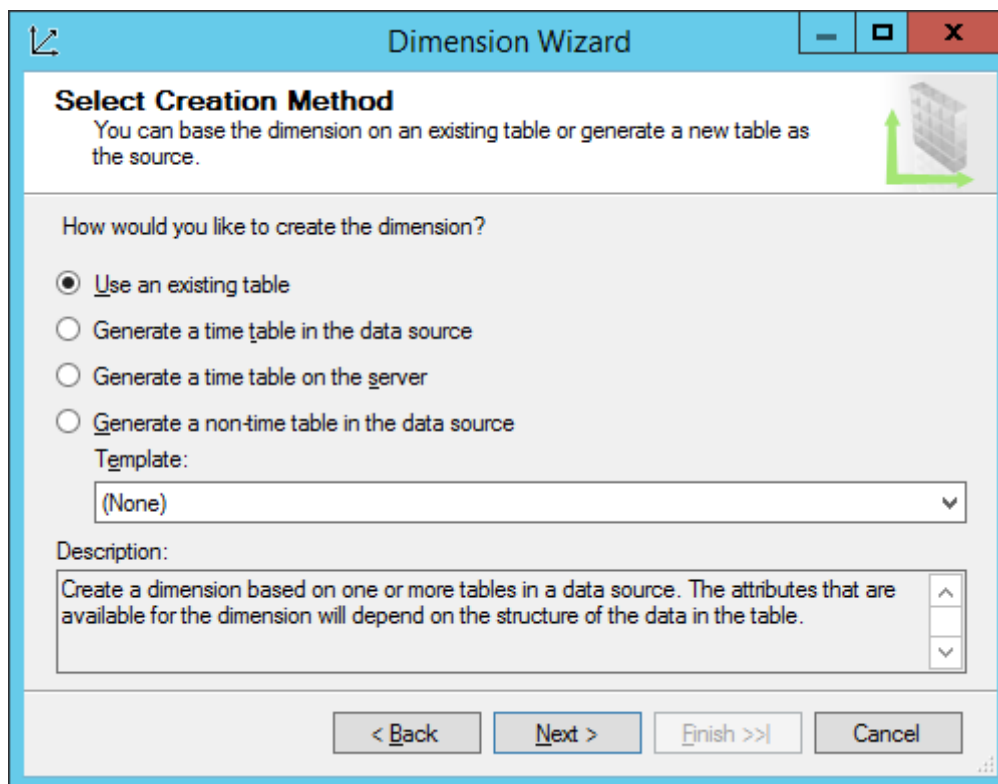


Рисунок 4.12 – Выбор метода создания измерения

В диалоге Specify Source Information (Определение исходных сведений) убедитесь, что выбрано представление источника данных «Представление AdvWorksDW». В списке Main table выберите таблицу «Дата». Нажмите кнопку Next (рис. 4.13).

В появившемся диалоге Select Dimension Attributes (Выбор атрибутов измерения) установите флажки для атрибутов: Date Key, Full Date Alternate Key, English Month Name, Calendar Quarter, Calendar Year, Calendar Semester.

Измените тип перечисленных атрибутов в соответствии с табл. 4.2 как показано на рис. 4.14. Нажмите кнопку Next.

В окне Completing the Wizard (Завершение работы мастера) отображается измерение «Дата» и его атрибуты (рис. 4.15). Нажмите кнопку Finish.

В меню File выберите команду Save All.

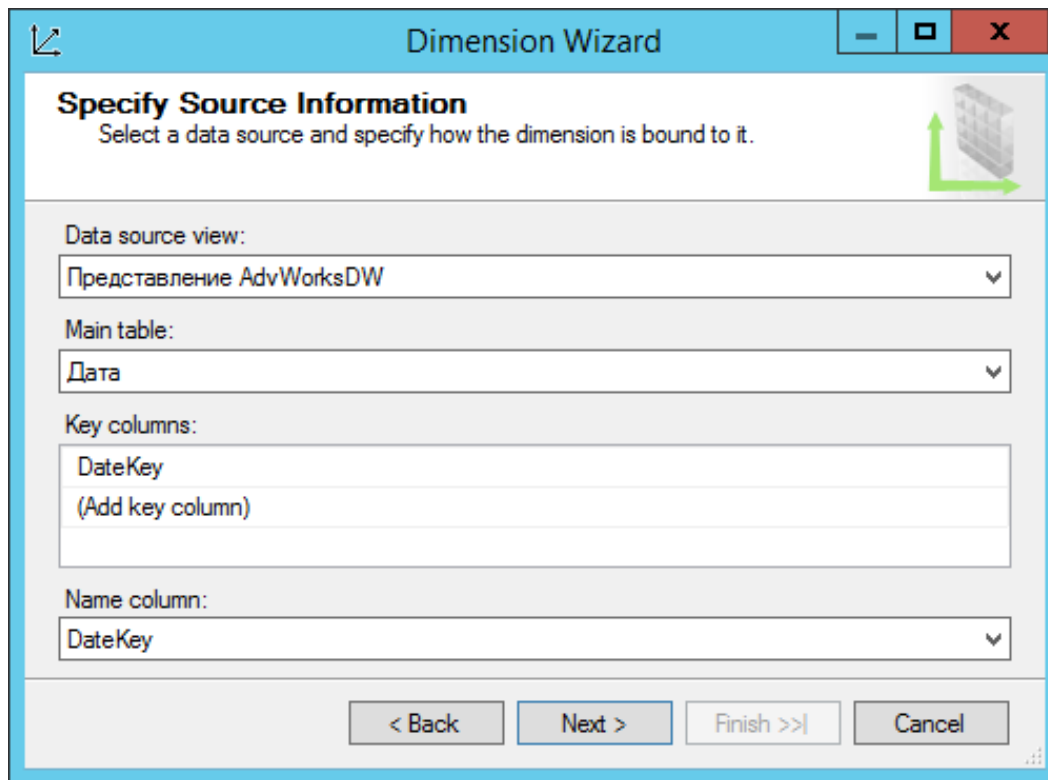


Рисунок 4.13 – Определение исходной информации об измерении

Таблица 4.2 – Характеристика типов атрибутов

<i>№</i>	<i>Название атрибута</i>	<i>Новый тип атрибута</i>
1	Date Key	Regular
2	Full Date Alternate Key	Date
3	English Month Name	Month
4	Calendar Quarter	Quarter
5	Calendar Year	Year
6	Calendar Semester	Half Year

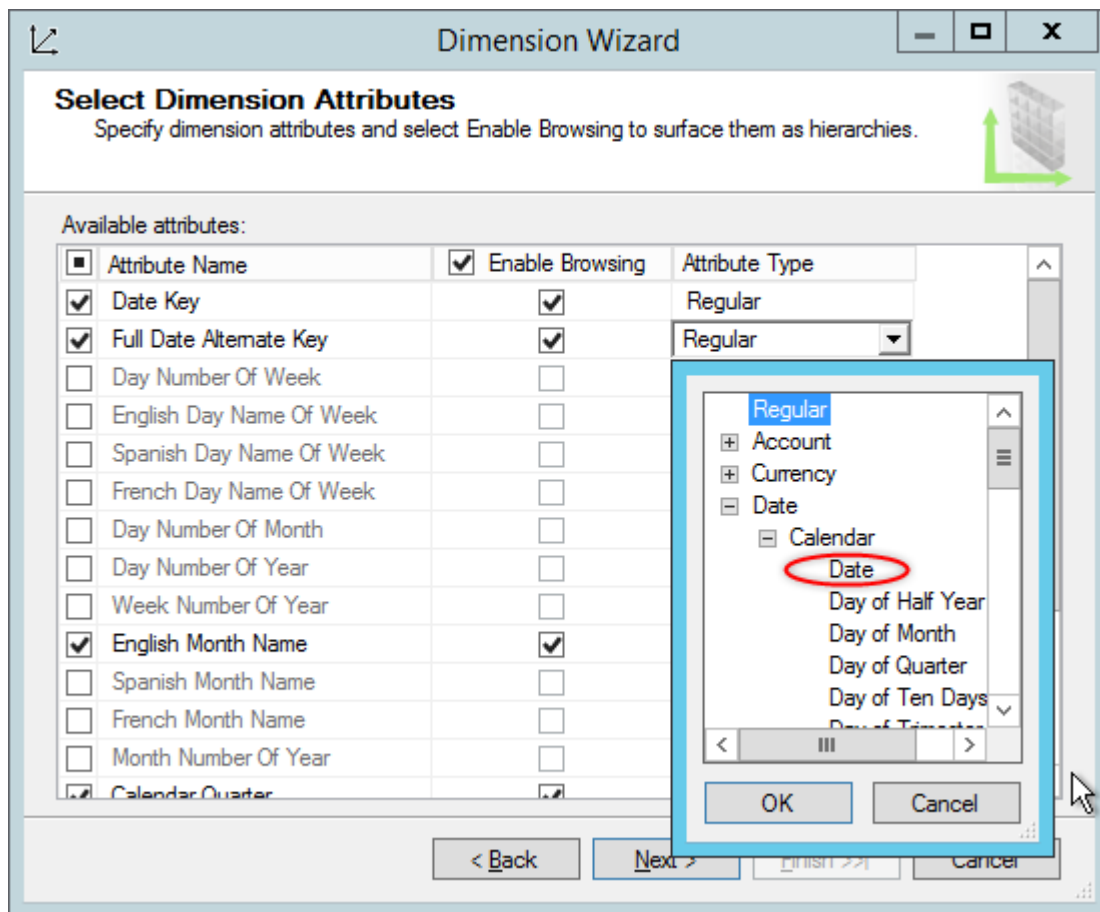


Рисунок 4.14 - Выбор типа атрибута

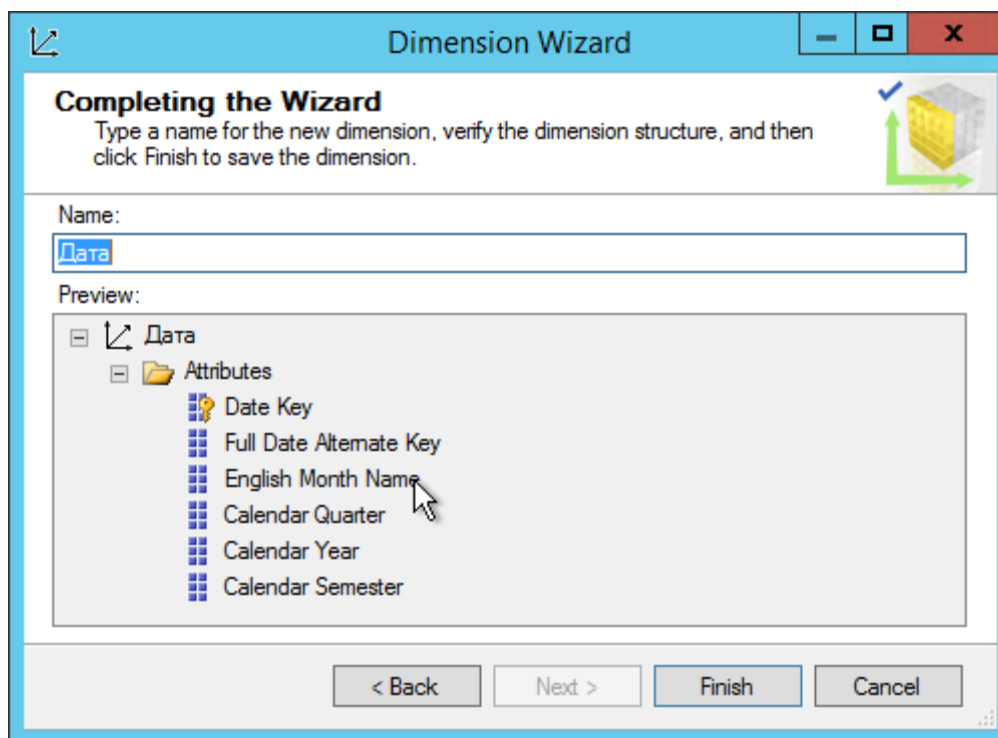


Рисунок 4.15 - Измерение «Дата» и его атрибуты

7. Создайте куб с помощью мастера кубов. Для этого в обозревателе решений в проекте «Проект ХД» щелкните правой кнопкой мыши по разделу Cubes, выберите команду New Cube. В появившемся окне мастера нажмите кнопку Next. В диалоге Select Creation Method выберите Use an existing table. Нажмите кнопку Next.

В диалоге Select Measure Group Tables (Выбор таблиц групп мер) убедитесь, что выбрано представление источника данных «Представление AdvWorksDW». Нажмите кнопку Suggest, чтобы мастер кубов предложил таблицы для создания групп мер (фактов). В результате анализа Мастер предложил в качестве таблицы группы мер (фактов) «Продажи через Интернет» (рис. 4.16).

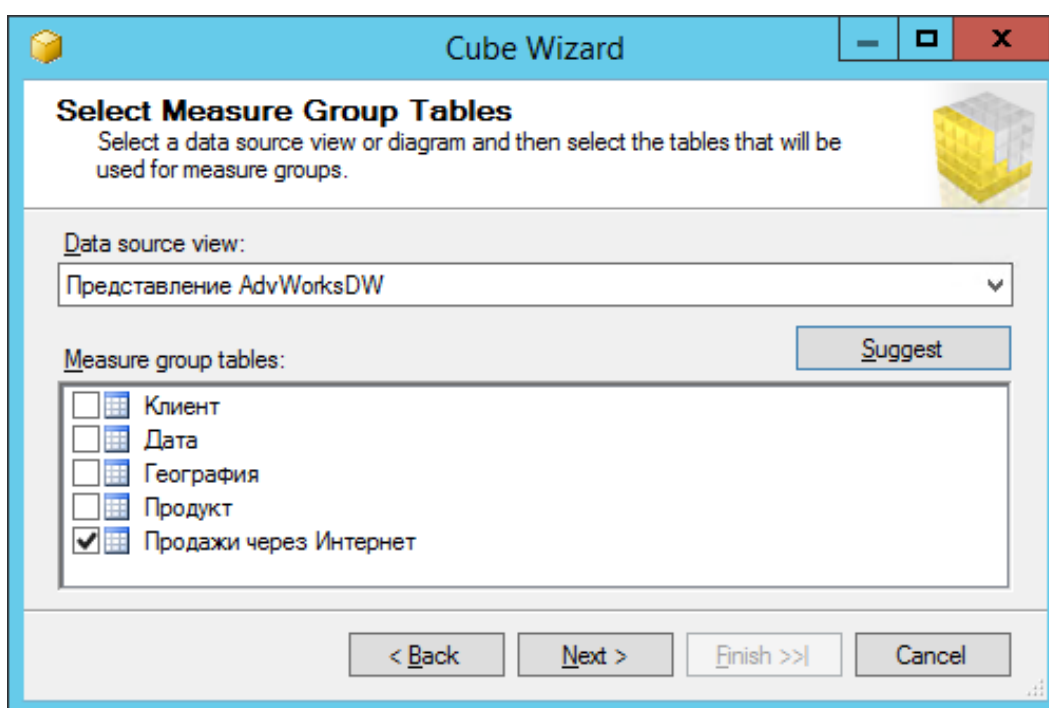


Рисунок 4.16 – Выбор таблицы группы мер (таблицы фактов)

Нажмите кнопку Next. На странице Select Measure (Выбор мер) просмотрите выбранные меры в группе мер Продажи через Интернет и снимите флажки для следующих мер: Promotion Key, Currency Key, Sales Territory Key, Revision Number. Нажмите кнопку Next (рис. 4.17).

В диалоге Select Existing Diminutions (Выбор существующих измерений) выберите ранее созданное измерение «Дата» (рис. 4.18) и нажмите кнопку Next.

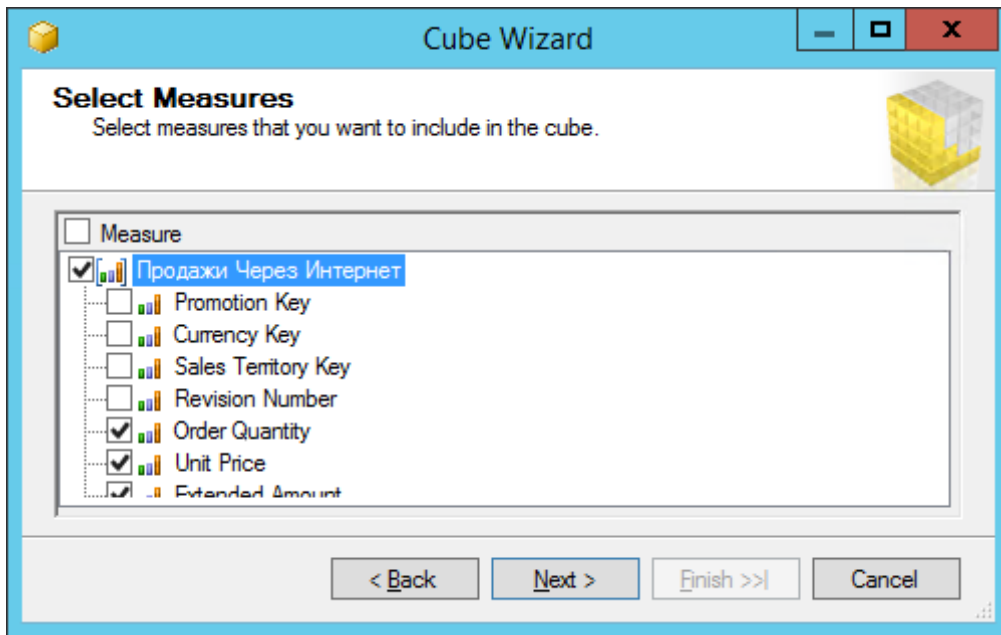


Рисунок 4.17 – Выбор мер для включения в куб

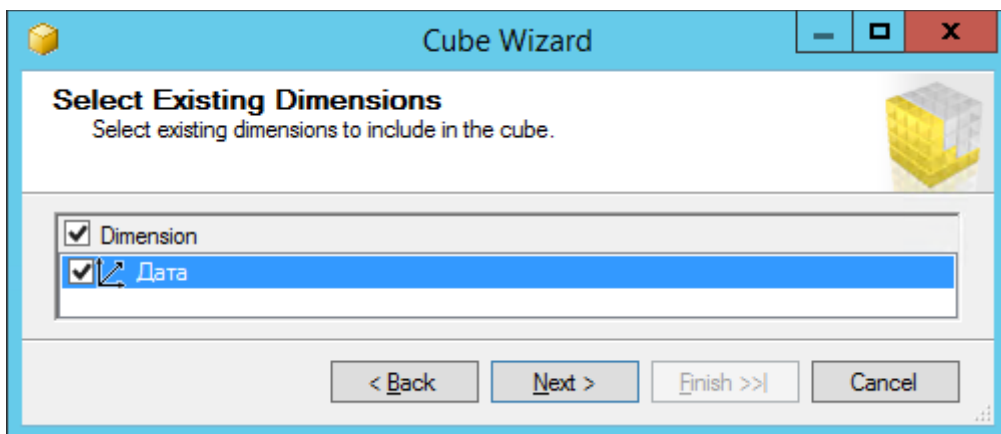


Рисунок 4.18 – Выбор существующего измерения «Дата»

В диалоге Select New Dimensions (Выбор новых измерений) выберите новые измерения: «Клиент», «География» и «Продукт». Убедитесь, что флажок «Продажи через Интернет» снят (рис. 4.19). Нажмите кнопку Next.

В окне Completing the Wizard измените имя куба на «Куб учебный». В области просмотра будет отображена группа мер «Продажи через Интернет» и входящие в нее меры (рис. 4.20). Нажмите кнопку Finish.

В конструкторе кубов будет отображен куб «Куб учебный» (рис. 4.21).

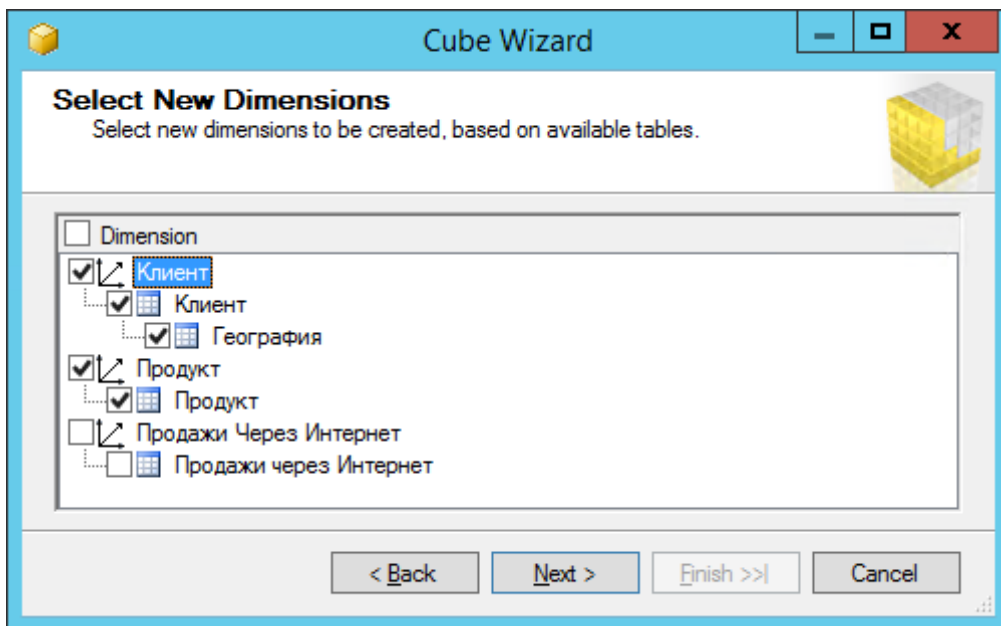


Рисунок 4.19 – Выбор новых измерений для создания

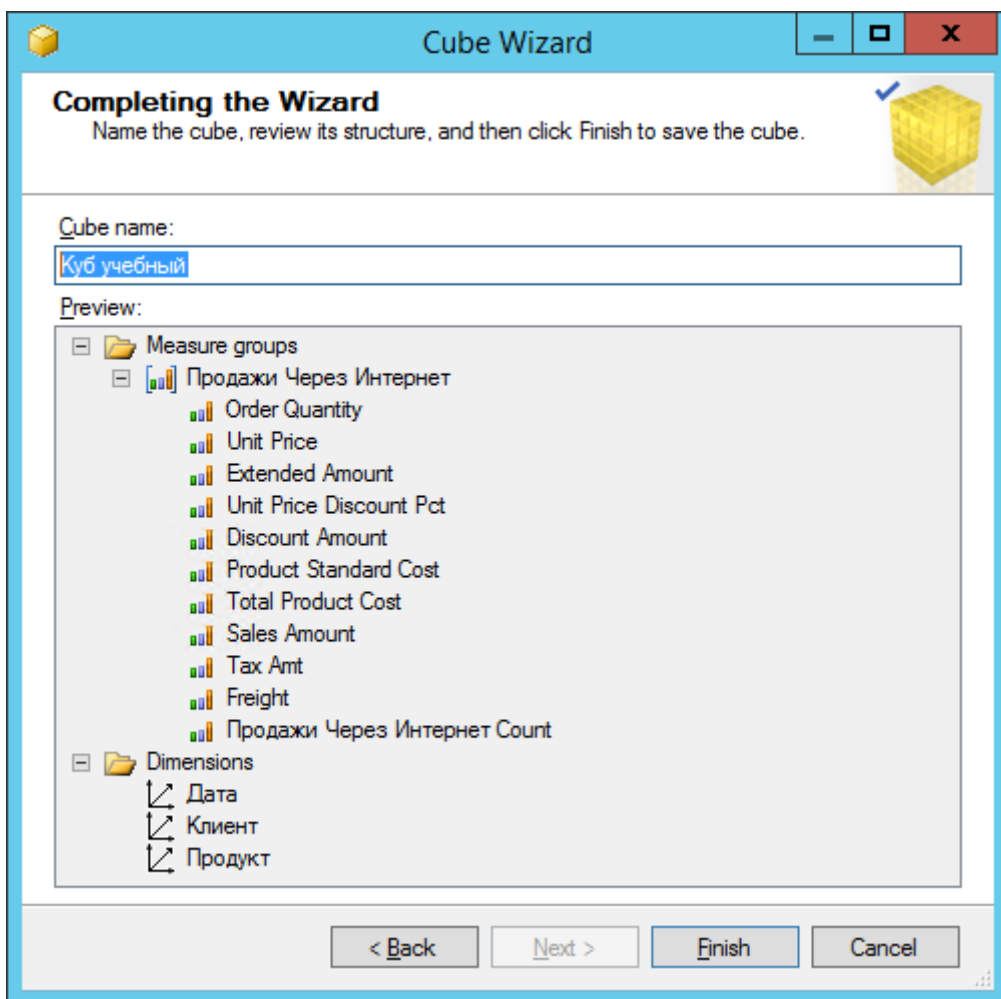


Рисунок 4.20 – Куб и его структура

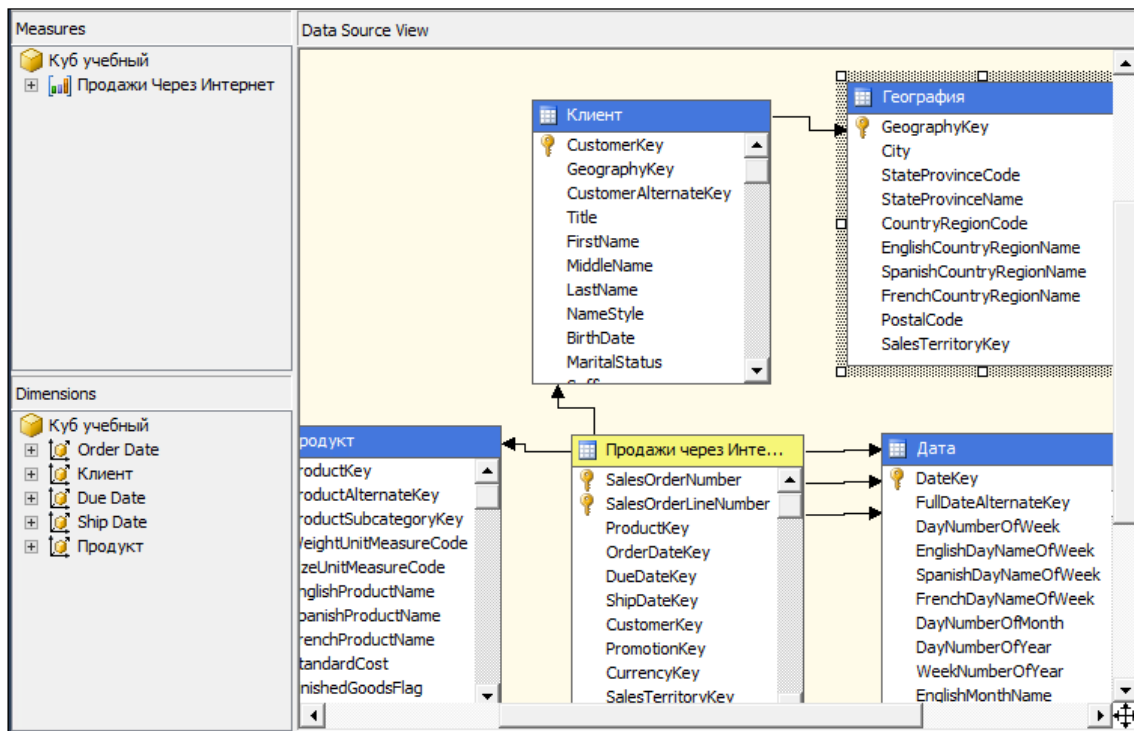


Рисунок 4.21 – Куб «Куб учебный» в конструкторе кубов

В меню File выберите команду Save All.

8. Добавьте атрибуты к измерениям «Клиент» и «Продукт» с помощью конструктора измерений. Для этого дважды щелкните измерение «Клиент» в узле Dimensions обозревателя решений (рис. 4.22).

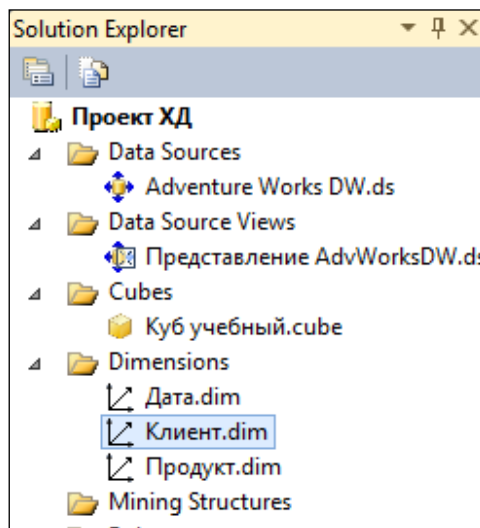


Рисунок 4.22 – Измерение «Клиент» в узле Dimensions обозревателя решений

В области Attributes обратите внимание на атрибуты «Customer Key» и «Geography Key», которые были созданы мастером кубов (рис. 4.23).

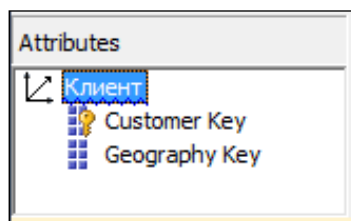


Рисунок 4.23 – Измерение «Клиент» в конструкторе измерений

В конструкторе измерений перетащите следующие столбцы из таблицы «Клиент» в области Data Source View в область Attributes: BirthDate, MaritalStatus, Gender, EmailAddress, YearlyIncome, TotalChildren, NumberChildrenAtHome, EnglishEducation, EnglishOccupation, HouseOwnerFlag, NumberCarsOwned, Phone, DateFirstPurchase, CommuteDistance. Аналогично перетащите следующие столбцы из таблицы «География» в области Data Source View в область Attributes: City, StateProvinceName, EnglishCountryRegionName, PostalCode (рис. 4.24).

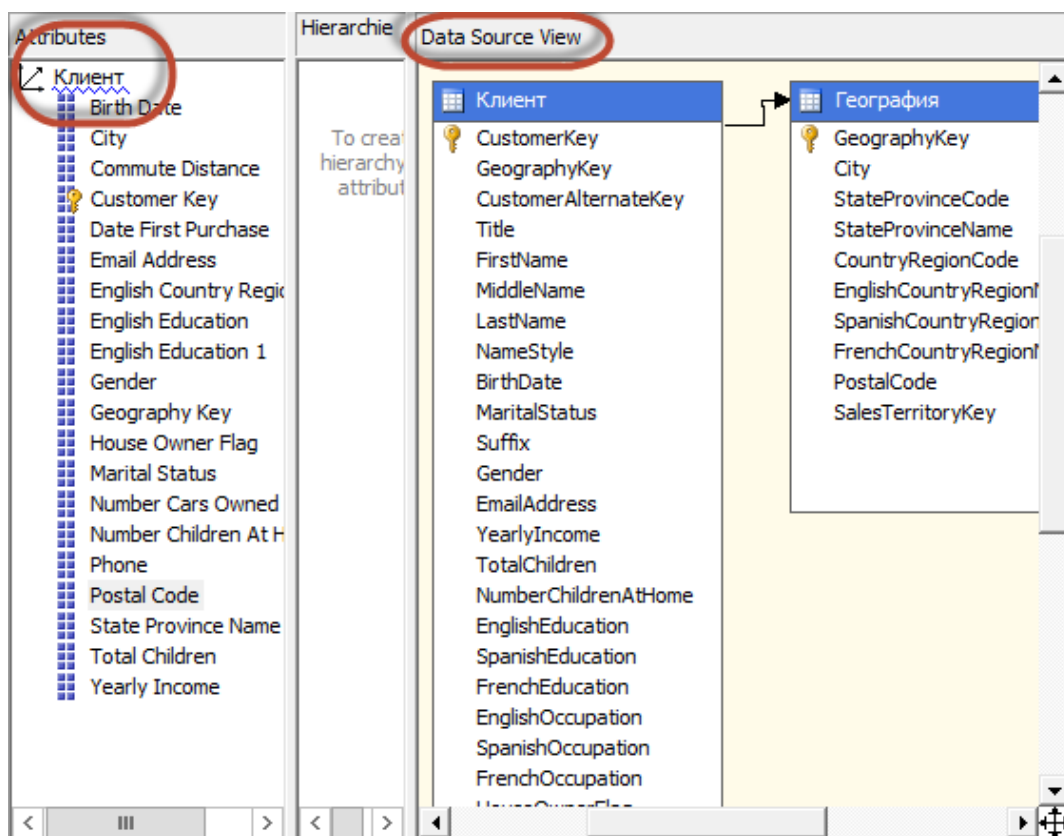


Рисунок 4.24 – Измерение «Клиент» с добавленными атрибутами

Аналогично измерению «Клиент» добавьте атрибуты для измерения «Продукт»: StandardCost, Color, SafetyStockLevel, ReorderPoint,

ListPrice, Size, SizeRange, Weight, DaysToManufacture, ProductLine, DealerPrice, Class, Style, ModelName, StartDate, EndDate, Status.

В меню File выберите команду Save All.

9. Просмотрите результаты изменения куба в конструкторе кубов. Для этого в обозревателе решений в узле Cubes дважды щелкните куб «Куб учебный».

Обратите внимание, что в кубе «Куб учебный» доступно пять измерений (рис. 4.25), хотя на уровне базы данных было создано только три измерения, что отражено в обозревателе решений (рис. 4.22).

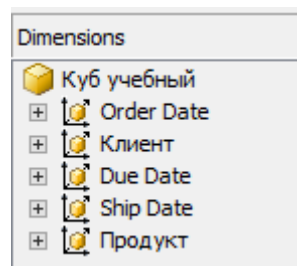


Рисунок 4.25 – Измерения в конструкторе кубов

Куб имеет больше измерений, чем база данных, потому что измерение «Дата» из базы данных служит основой для трех отдельных измерений куба «Куб учебный», связанных с датами. Эти измерения основаны на разных связанных с датами фактах из таблицы фактов. Эти измерения даты называются также ролевыми измерениями. Три измерения даты куба дают пользователям возможность разделить куб по трем отдельным фактам, связанным с каждой продажей: дате заказа товара (Order Date), сроку выполнения заказа (Due Date) и дате отгрузки заказа (Ship Date). Повторно используя одно измерение базы данных для нескольких измерений куба, службы Analysis Services упрощают управление измерениями, используют меньше места на диске и уменьшают общее время обработки данных.

10. Переключитесь в конструктор кубов и изучите содержание закладок. Обратите внимание, что куб нельзя просмотреть в закладке Browser, поскольку он еще не был развернут на экземпляре служб Analysis Services. На данном этапе куб в проекте «Проект ХД» представляет собой лишь определение куба. Во время развертывания и обработки куба создаются объекты в экземпляре служб Analysis Services, которые затем заполняются данными из базовых источников данных.

11. В обозревателе решений в узле Cubes щелкните правой кнопкой мыши по кубу «Куб учебный», затем в контекстном меню выберите команду View Code (рис. 4.26).

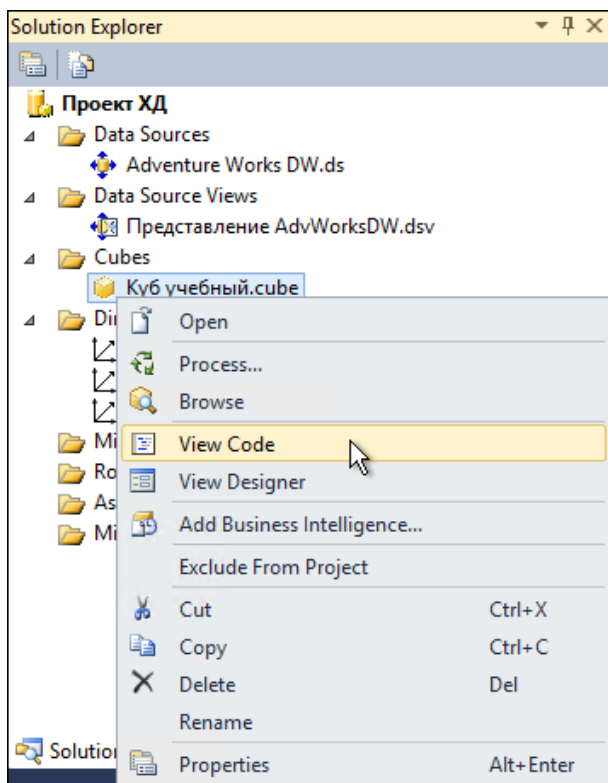


Рисунок 4.26 – Контекстное меню куба «Куб учебный» в обозревателе решений

На вкладке «Куб учебный.cube [XML]» отображается XML-код куба «Куб учебный» (рис. 4.27). Этот код используется для создания куба в экземпляре служб Analysis Services в процессе развертывания. Закройте вкладку XML-кода.



Рисунок 4.27 – Фрагмент XML-кода куба «Куб учебный»

12. Разверните проект на экземпляре аналитических служб. Чтобы просмотреть куб и данные измерения для объектов куба «Куб учебный» проекта «Проект ХД», требуется развернуть проект на указанном преподавателем экземпляре служб Analysis Services. В процессе развертывания проекта служб Analysis Services в экземпляре служб Analysis Services создаются объекты, определенные в проекте. В процессе обработки объектов в экземпляре служб Analysis Services производится автоматическое копирование данных из базовых источников данных в объекты куба.

Посмотрите свойства развертывания проекта. Для этого щелкните правой мыши по проекту «Проект ХД» в обозревателе решений и выберите пункт Properties. Для конфигурации Deployment введите имя целевого сервера согласно структуре <ИмяСервера>\<ИмяЭкземпляра>: «lpvserver61\mssqlserver2». Нажмите кнопку ОК (рис. 4.28).

В обозревателе решений щелкните правой кнопкой мыши проект «Проект ХД» и выберите пункт Deploy. Ход выполнения развертывания отображается в двух окнах: Output (рис. 4.29) и Deployment Progress – Проект ХД (рис. 4.30).

Развертывание куба «Куб учебный» для экземпляра Analysis Services и его обработка успешно завершены.

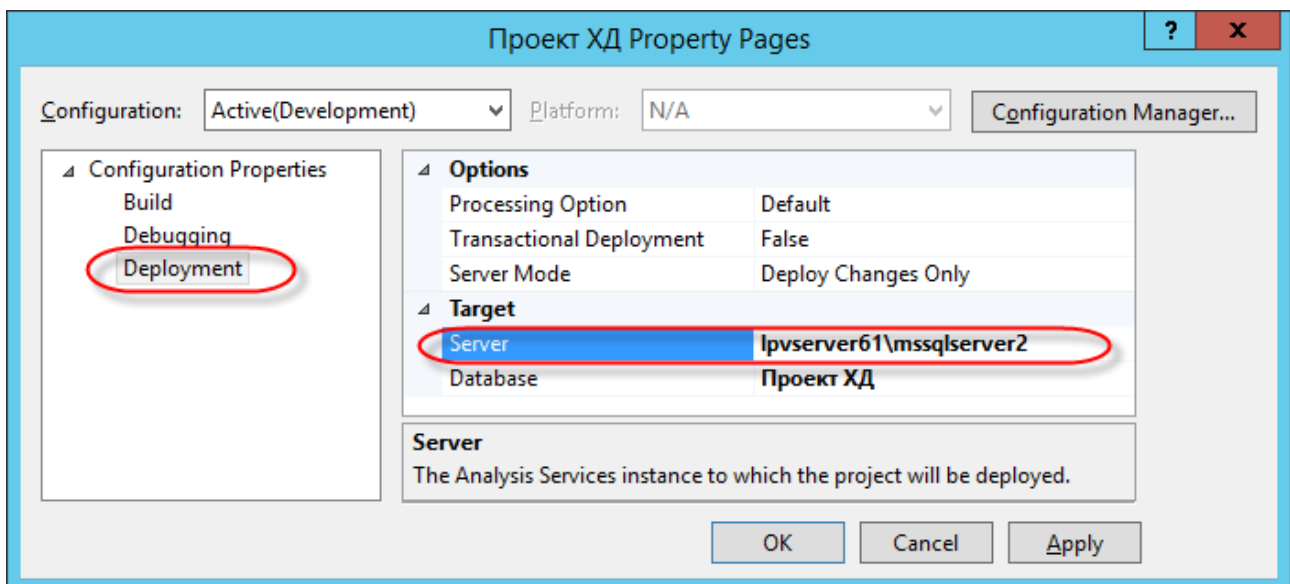


Рисунок 4.28 – Свойства конфигурирования развертывания проекта

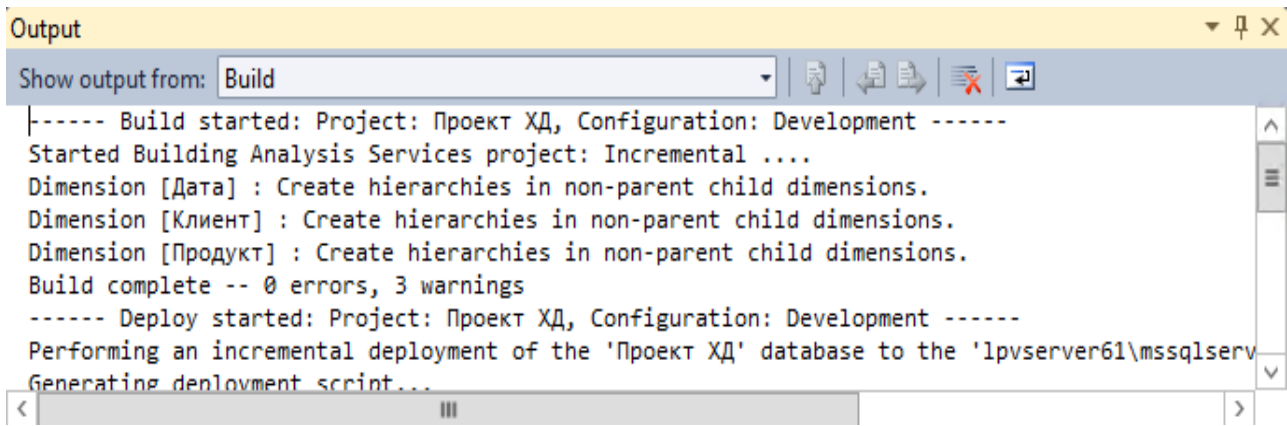


Рисунок 4.29 – Окно Output

13. Просмотрите куб. После развертывания данные куба отображаются на вкладке Browser конструктора кубов, а данные измерений отображаются на вкладке Browser конструктора измерений. Теперь куб можно анализировать средствами MS SQL Server или с помощью других инструментов бизнес-аналитики.

14. Выполните экспорт куба MS Excel (вкладка Browser, команда Analyze in Excel) и исследуйте его с помощью инструментария сводных таблиц MS Excel.

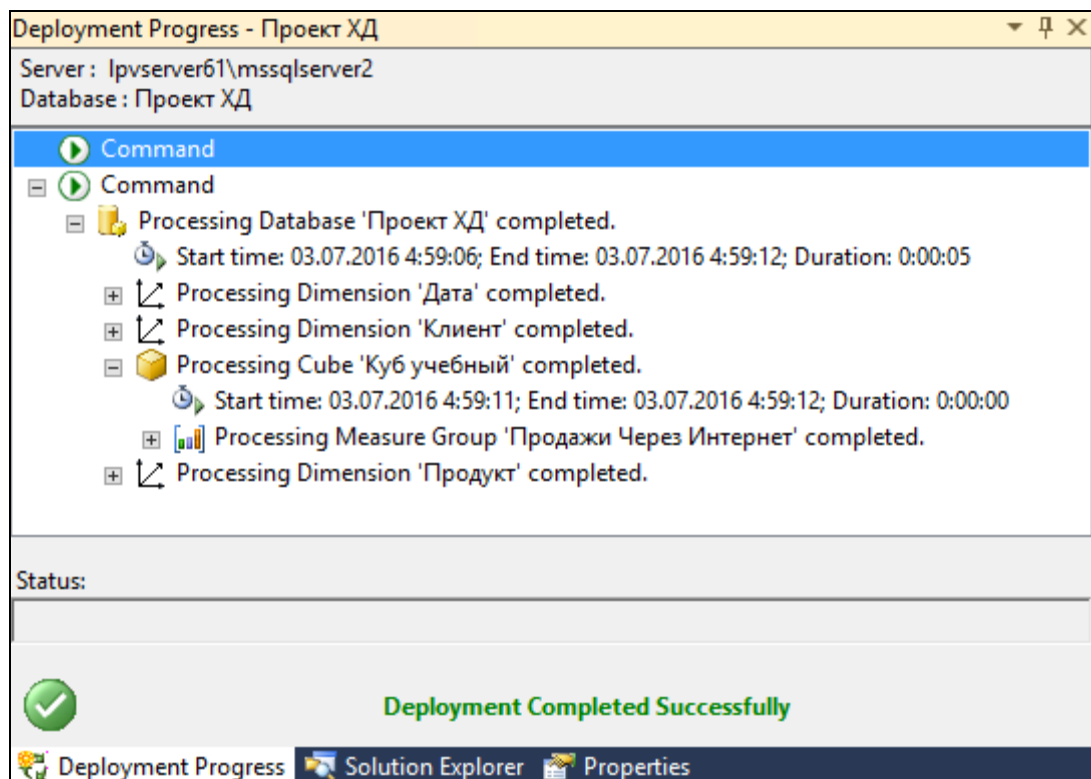


Рисунок 4.30 – Окно Deployment Progress – Проект ХД

4.4. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Каково назначение служб MS SQL Server Analysis Services?
2. Какой компонент SQL Server используется для операций ETL?
3. Какой инструмент SQL Server используется для управления OLAP-кубами и работы с ними?
4. Какой инструмент SQL Server используется для создания новых OLAP-кубов?
5. В чем отличие проекта OLAP-куба (проекта аналитических служб) от развернутого проекта OLAP-куба?
6. Каким образом можно проводить анализ данных из хранилища данных, разработанного в среде SQL Server?
7. Как экспортировать данные куба в файл MS Excel?

4.5. ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОГО ВЫПОЛНЕНИЯ

Задание 1

Для выполнения задания 1 можно использовать приложение MS SQL Server Management Studio 2012 (или выше), опубликованное в облаке Microsoft Azure RemoteApp в гибридной коллекции Финансового университета.

Альтернативой может стать установка MS SQL Server 2012 (или выше) на домашний компьютер. В этом случае для скачивания программного обеспечения рекомендуется воспользоваться услугами Центра академического программного обеспечения Microsoft DreamSpark Premium Финансового университета, используя ссылку <https://e5.onthehub.com/WebStore/Welcome.aspx?vsro=8&ws=80324a2d-f658-e211-b607-f04da23e67f6>.

В рамках самостоятельной работы разработка витрины данных осуществляется для бизнес-кейса, представленного в п. 3.3. Требуется создать объекты каталога витрины данных на экземпляре сервера MS SQL Server средствами MS SQL Server Management Studio.

В результате выполнения задания должен быть создан отчет, содержащий описание бизнес-кейса (см. п. 3.3), скриншот содержимого каталога вашей базы данных в MS SQL Server Management Studio, демонстрирующий состав таблиц каталога, а также скриншот диаграммы витрины данных.

В отчете также представить таблицу сравнения возможностей CA ERwin Data Modeler и MS SQL Server Management Studio в области моделирования витрины данных, основываясь на собственном опыте выполнения заданий из п. 3.3 и 4.3.

Отчет оформить как документ Word. В названии документа указать номер группы и фамилию. Документ сохранить в личной папке.

Задание 2

Для выполнения задания 2 можно использовать приложение MS SQL Server Data Tools 2012 (или выше), опубликованное в облаке Microsoft Azure RemoteApp в гибридной коллекции Финансового университета. Альтернативой может стать установка MS SQL Server 2012 (или выше) на домашний компьютер.

Для выполнения задания требуется:

- изучить структуру базы данных AdventureWorksDW;
- сформулировать бизнес-кейс по разработке хранилища данных, предполагая, что единственным источником данных станет AdventureWorksDW;
- разработать проект аналитических служб MS SQL Server на основе шаблона Analysis Services Multidimensional and Data Mining Project, включая определение источника данных; создание представления источников данных; создание измерений; создание куба;
- развернуть проект на указанном преподавателем экземпляре служб MS SQL Server Analysis Services;
- провести анализ данных разработанного хранилища с помощью любого инструмента бизнес-аналитики.

По результатам выполнения задания сформировать отчет. В названии отчета указать номер группы и фамилию. Документ сохранить в личной папке.

ИСПОЛЬЗОВАНИЕ НАДСТРОЕК MS EXCEL ДЛЯ БИЗНЕС-АНАЛИТИКИ

5.1. ОСНОВЫ ИСПОЛЬЗОВАНИЯ НАДСТРОЕК MS EXCEL ДЛЯ АНАЛИЗА

Интеллектуальный анализ данных выполняется аналитическими службами MS SQL Server, которые поддерживают множество различных алгоритмов анализа данных. Таблицы MS Excel выступают в роли интерфейса между аналитиком и MS SQL Server. Поэтому для работы с надстройками Excel требуется подключение к аналитическим службам MS SQL Server.

Можно назвать шесть основных этапов типового процесса глубинного анализа данных, продвигаемого компанией Microsoft:

- 1) формулировка проблемы;
- 2) подготовка исторических данных;
- 3) исследование и оценка исторических данных;
- 4) построение модели анализа данных;
- 5) исследование и оценка модели анализа данных;
- 6) развертывание и корректировка модели анализа.

Инструменты надстроек Excel Analyze и Excel Data Mining автоматизируют этот процесс. Надстройка Excel Analyze включает следующие инструменты анализа: Анализ ключевых факторов влияния; Поиск категорий; Заполнение по примеру; Прогноз; Выделение исключений; Анализ сценария (Анализ сценария поиска решения и Анализ гипотетических вариантов); Расчет прогноза; Анализ покупательского поведения.

5.2. НАСТРОЙКА ПРОГРАММНОЙ СРЕДЫ ДЛЯ ПРАКТИКУМА

Для аудиторного выполнения практического задания потребуются инструмент MS Excel 2010 (или выше) с подключенными надстройками анализа таблиц и MS SQL Server Data Tools 2012 (или выше). Работать с приложением MS Excel в аудитории можно в одном из двух режимов:

1) с локально установленным на студенческий компьютер приложением;

2) с приложением, опубликованном в облаке Microsoft Azure в гибридной коллекции RemoteApp Финансового университета.

Кроме этого потребуется файл DMAddins_SampleData1.xlsx, содержащий подготовленные данные для проведения интеллектуального анализа данных.

Параметры подключения к экземпляру MS SQL Server преподаватель сообщает на занятии.

5.3. ПРАКТИКУМ ПО БИЗНЕС-АНАЛИЗУ С НАДСТРОЙКАМИ MS EXCEL

Цель практикума

Закрепить теоретические знания и сформировать практические навыки бизнес-анализа с помощью надстроек интеллектуального анализа данных MS Excel.

План практикума

Практикум включает следующие этапы:

1. Изучение содержания книги Excel с образцами данных для анализа.

2. Подключение к аналитическим службам MS SQL Server для активации инструментов интеллектуального анализа данных MS Excel.

3. Изучение возможностей следующих инструментов интеллектуального анализа данных:

- анализ ключевых факторов влияния;
- поиск категорий;
- заполнение по примеру;
- прогноз;
- выделение исключений.

Порядок выполнения практикума

1. Скопируйте книгу DMAddins_SampleData1.xlsx, сохраните копию файла под именем DMAddins_SampleData1-*Фамилия*.xlsx. Откройте скопированную книгу. На странице Introduction дано описание и назначение всех страниц книги.

2. Переключитесь на страницу Table Analysis Tools Sample. Рабочая область на этой странице отформатирована как таблица. Благодаря этому в группе закладок «Работа с таблицами» появилась закладка Analyze. На ленте меню Analyze представлены инструменты интеллектуального анализа данных (рис. 5.1)

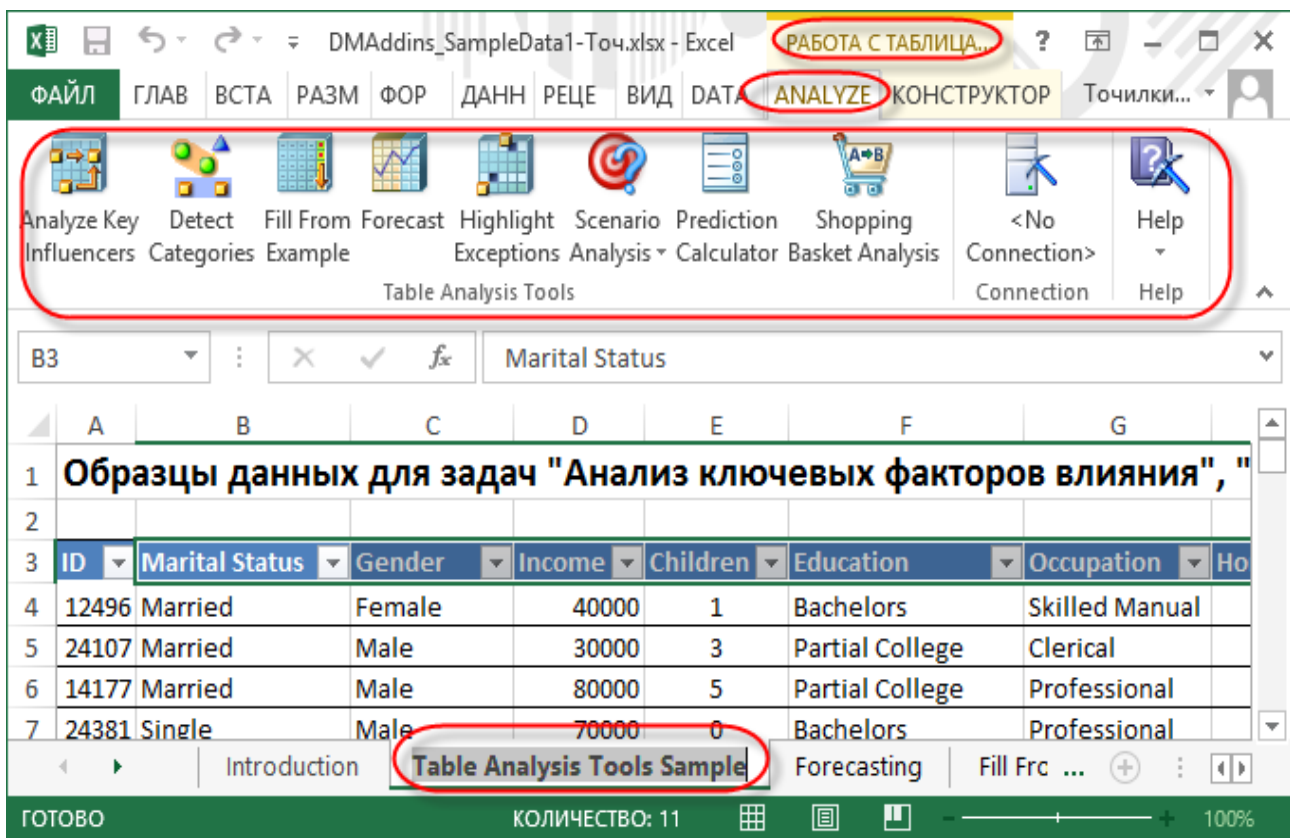


Рисунок 5.1 – Лента меню Analyze

3. Выполните подключение к экземпляру MS SQL Server. Для этого на ленте меню Analyze выберите пункт Connection. В открывшемся диалоговом окне нажмите кнопку New. В новом диалоге укажите параметры подключения к экземпляру сервера «\srvserver61\mssqlserver2» (рис. 5.2). Выполните тестовое подключение, затем нажмите кнопку ОК.

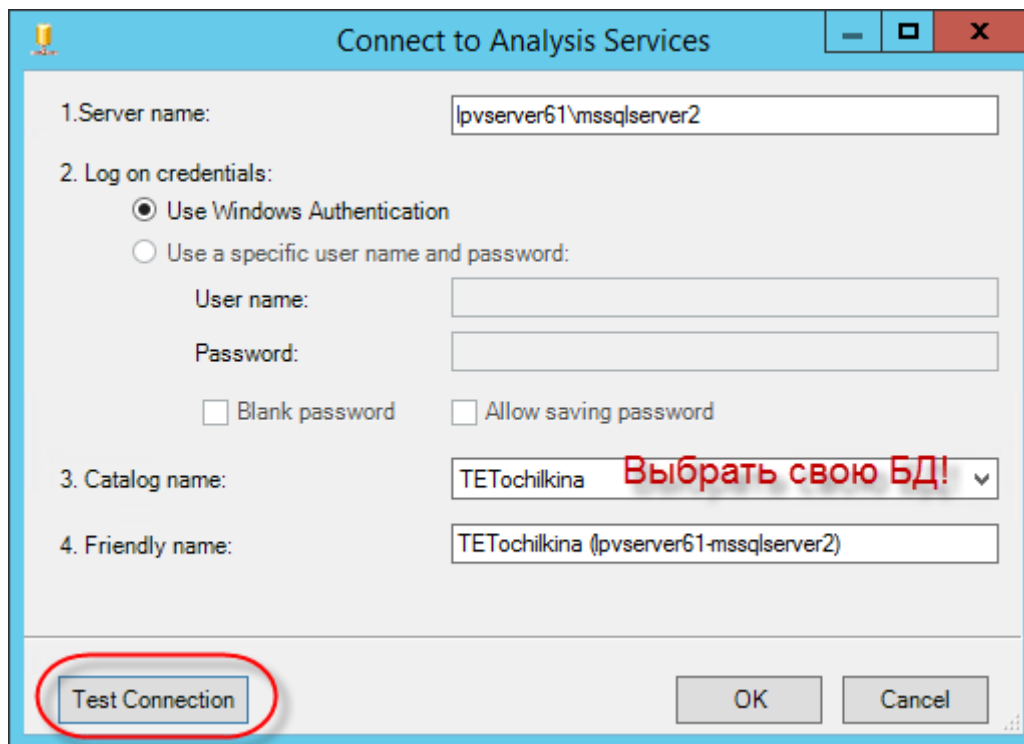


Рисунок 5.2 – Параметры подключения к серверу

4. Выявите влияние демографических данных клиентов на их решение о покупке велосипеда. Для этого на странице Table Analysis Tools Sample щелкните по пункту меню Analyze Key Influencers (Анализ ключевых факторов влияния). В появившемся диалоговом окне мастера выберите анализируемый столбец Purchased Bike (куплен велосипед).

Щелкните ссылку Choose columns to be used for analysis (Выбор столбцов для анализа). В диалоговом окне Advanced Columns Selection (Расширенный выбор столбцов) выберите столбцы, влияющие, по вашему мнению, на продажи велосипедов: Marital Status, Cars, Region, Gender. Упрощение анализа осуществляется для повышения его скорости и точности. Нажмите кнопку Run, потом Add Report. В результате будет создана новая страница в книге Excel, содержащая отчет о ключевых факторах влияния на покупку велосипеда (рис. 5.3).

В отчете содержатся факторы, связанные с каждым из результатов (Yes-велосипед был куплен, No – не куплен), и вероятности отношений в графической форме. Результаты упорядочены в порядке убывания влияния для каждого из возможных значений в столбце Purchased Bike.

	A	B	C	D
1	Key Influencers Report for 'Purchased Bike'			
2				
3	Key Influencers and their impact over the values of 'Purchased Bike'			
4	Filter by 'Column' or 'Favors' to see how various columns influence 'Purchased Bike'			
5	Column	Value	Favors	Relative Impact
6	Cars	2	No	
7	Marital Status	Married	No	
8	Region	North America	No	
9	Cars	0	Yes	
10	Marital Status	Single	Yes	
11	Cars	1	Yes	
12	Region	Pacific	Yes	
13				
14	Discrimination between factors leading to 'No' and 'Yes'			
15	Filter by 'Column' to see how different values favor 'No' or 'Yes'			
16	Column	Value	Favors No	Favors Yes
17	Cars	2		
18	Cars	0		
19	Marital Status	Married		
20	Marital Status	Single		
21	Cars	1		
22	Region	Pacific		
23	Region	North America		

Рисунок 5.3 – Отчет о ключевых факторах влияния

Обратите внимание, что колонка Gender (Пол) отсутствует в отчете, значит, согласно проведенному исследованию, она не оказывает заметного влияния на приобретение велосипеда.

Согласно отчету наиболее значимым фактором, заставляющим НЕ покупать велосипед, является наличие двух автомобилей. Наиболее значимым фактором, заставляющим покупать велосипед, является отсутствие автомобилей.

4. Проведите кластеризацию клиентов. Для этого сначала сделайте копию страницы Table Analysis Tools Sample. Перейдите на созданную копию страницы. Запустите инструмент Detect Categories.

В появившемся диалоговом окне мастера можно отменить использование в анализе некоторых колонок, содержащих уникальные значения, такие как имена людей и идентификаторы строк, т.к. они не будут полезны для анализа. Обратите внимание, что выбор столбца ID отменяется автоматически. Остальные столбцы останутся выбранными.

Можно указать максимальное количество создаваемых категорий. По умолчанию инструмент автоматически определяет число категорий. Оставьте значение по умолчанию.

Установите опцию Append a Category column to the original Excel table, чтобы созданная колонка категорий была автоматически добавлена к исходной.

Нажмите кнопку Run. Инструмент создаст новую страницу Categories Report. Сгенерированный отчет по категориям клиентов содержит две таблицы (рис. 5.4) и диаграмму Category Profiles (рис. 5.5).

	A	B	C	D
1	7 categories were detected			
2				
3	To rename a category, edit the 'Category Name' below.			
4	('Category Name' changes are visible in the 'Category' column of the source Excel table)			
5	Category Name	Row Count		
6	Category 1	189		
7	Category 2	141		
8	Category 3	158		
9	Category 4	149		
10	Category 5	126		
11	Category 6	129		
12	Category 7	108		
13				
14				
15	Category Characteristics			
16	Filter the table by 'Category' to see the characteristics of different categories.			
17	Category	Column	Value	Relative Importance
18	Category 1	Income	Very Low:< 39050	
19	Category 1	Region	Europe	
20	Category 1	Occupation	Manual	
21	Category 1	Occupation	Clerical	
22	Category 1	Commute Distance	0-1 Miles	
23	Category 1	Cars	0	
24	Category 1	Children	2	
25	Category 1	Children	1	
26	Category 1	Education	Partial High School	
27	Category 1	Education	High School	
28	Category 1	Children	3	

Рисунок 5.4 – Таблицы отчета о найденных категориях клиентов

Изучите отчет по категориям (рис. 5.4) для выявления закономерностей в группах клиентов.

В первой таблице отчета представлены найденные категории с указанием временных имен и количеством строк исходных данных, попадающих в соответствующую категорию.

Вторая таблица представляет характеристику категорий клиентов. Рекомендуется выполнять фильтрацию записей этой таблицы по категории. Колонка *Relative importance* графически показывает, насколько важна пара «атрибут-значение» в качестве отличительного фактора данной категории.

Чем длиннее полоска, тем в большей степени этот атрибут является репрезентативным для данной категории.

В нашем примере видно, что группа клиентов (Category 1) сформирована главным образом из клиентов из Европы с низким уровнем дохода и небольшим расстоянием от дома до работы. Можно дать этой категории более информативное имя. Новое имя категории сразу отображается в столбце назначения категорий на листе исходных данных.

Диаграмма *Category Profiles* (рис. 5.5) позволяет в интерактивном режиме подробно просматривать обнаруженные закономерности. Содержит элемент управления «Сводная таблица», позволяющий интерактивно фильтровать поля и изменять их расположение.

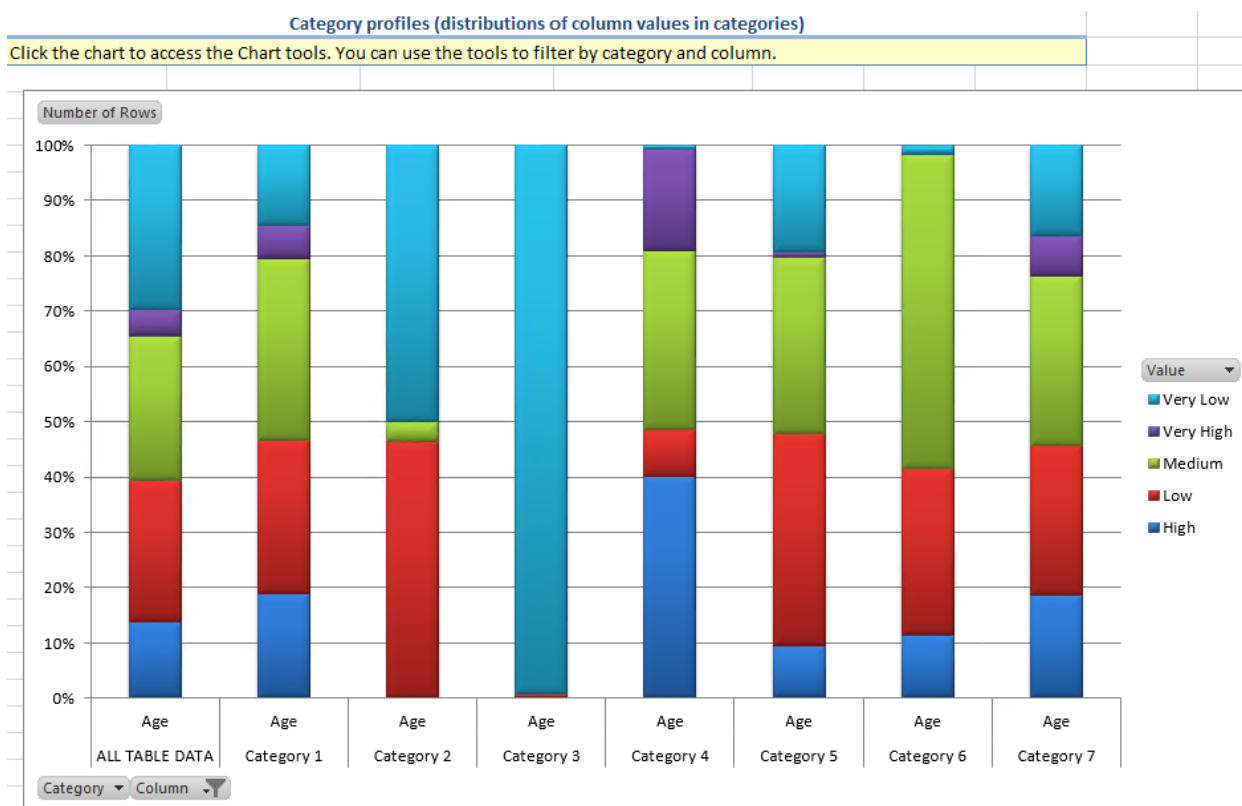


Рисунок 5.5 – Диаграмма «Профили категорий» отчета по категориям клиентов

5. Выполните автозаполнение отсутствующих значений в колонке на основе выявления закономерностей в данных таблицы. Для этого перейдите на страницу Fill From Example. Обратите внимание, что в колонке High Value Customer отсутствует большая часть значений.

Запустите инструмент Fill From Example.

В появившемся диалоговом окне мастера можно настроить параметры анализа. Нажмите кнопку Run. В книге Excel на странице Fill From Example в колонке High Value Customer будут добавлены отсутствующие значения. Кроме этого будет создана новая страница High Value Customer Patterns с отчетом о найденных в данных закономерностях (рис. 5.6).

Pattern Report for 'High Value Customer'			
Key Influencers and their impact over the values of 'High Value Customer'			
Filter by 'Column' or 'Favors' to see how various columns influence 'High Value Customer'			
Column	Value	Favors	Relative Impact
Region	Pacific	No	
Commute Distance	5-10 Miles	No	
Gender	Female	No	
Education	Partial High School	No	
Education	Bachelors	No	
Commute Distance	1-2 Miles	No	
Occupation	Professional	No	
Commute Distance	2-5 Miles	Yes	
Children	5	Yes	
Region	Europe	Yes	
Home Owner	No	Yes	
Education	Partial College	Yes	
Children	3	Yes	
Cars	2	Yes	

Рисунок 5.6 – Отчет о найденных закономерностях в данных

В отчете представлены созданные на основе анализа закономерностей правила, которые были использованы для заполнения отсутствующих значений. Описание каждого значимого правила содержит имя и значение столбца, а также относительное влияние на прогноз.

6. Составьте прогноз продаж в Европе и Северной Америке. Для этого перейдите на страницу Forecasting. Запустите инструмент Forecasting. В появившемся диалоговом окне мастера можно указать число прогнозируемых периодов – 6. Нажмите кнопку Run.

В результате работы инструмента в исходную таблицу будут добавлены строки со спрогнозированными значениями. Кроме того на новой странице книги будет создан отчет о прогнозе. Отчет содержит график, который отображает тренды за предысторию и спрогнозированные значения. Значения предыстории выводятся сплошной линией, а прогнозы – пунктирной (рис. 5.7).

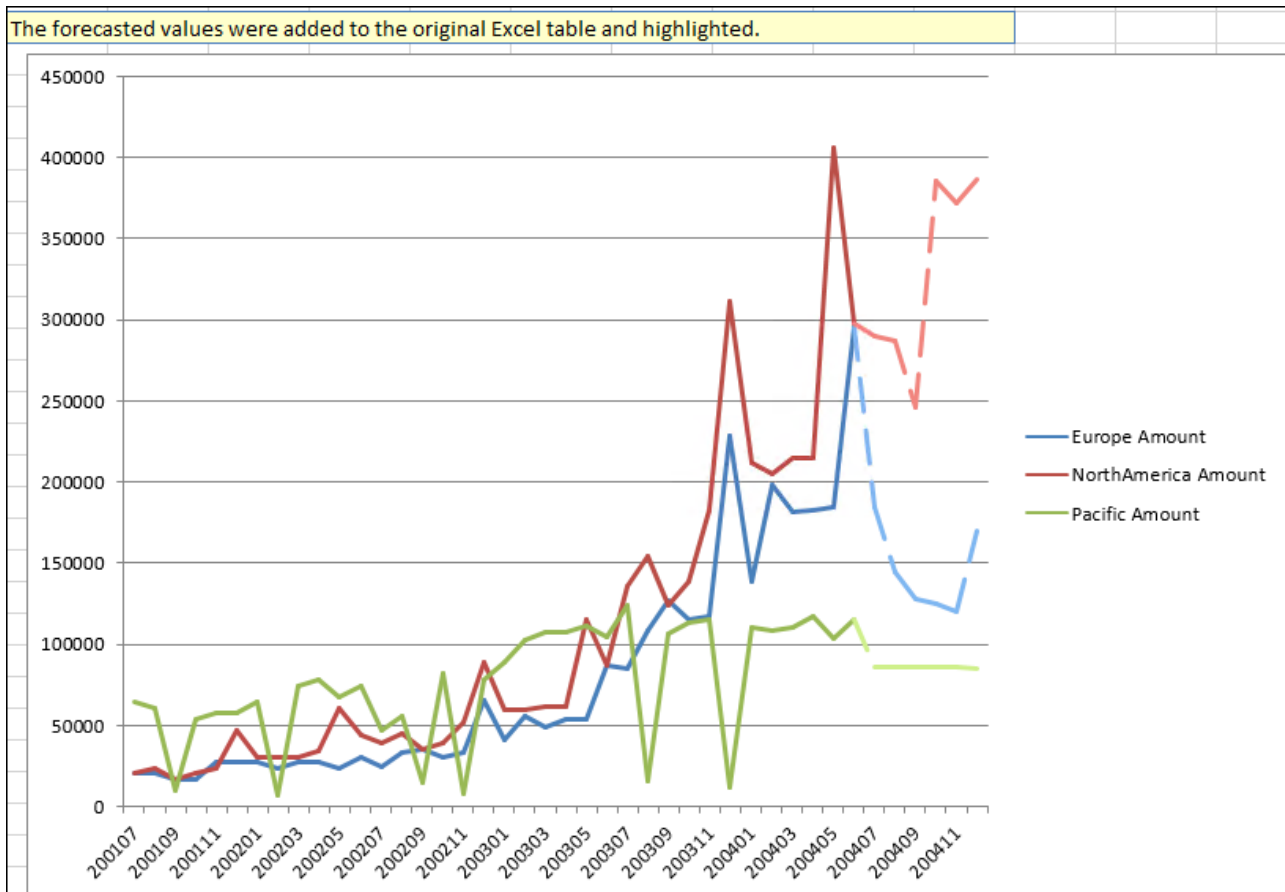


Рисунок 5.7 – Прогноз продаж

7. Найдите аномалии в данных. Для этого перейдите на страницу Table Analysis Tools Sample. Запустите инструмент Highlight Exceptions (выделение исключений). Сгенерированный отчеты отображается на новой странице Table2 Outliers (рис. 5.8).

В полученном отчете отмечены 34 выброса (аномальных значения), которые находились выше порога исключений (75). Большинство аномальных значений находится в столбцах Children и Cars. Переключитесь на страницу с исходными данными, найдите выделенные строки и проанализируйте аномальные значения. Проверьте, не является ли аномальное значение ошибкой ввода данных. Темная

подсветка строки означает, что строка требует внимания. Светлая подсветка ячейки означает, что значение в этой ячейке рассматривается как подозрительное.

Highlight Exceptions Report for Table2	
The outlier cells are highlighted in the original table.	
Exception threshold (more or fewer exceptions)	75
Column	Outliers
Marital Status	0
Gender	0
Income	4
Children	9
Education	1
Occupation	3
Home Owner	0
Cars	7
Commute Distance	4
Region	0
Age	6
Purchased Bike	0
Итого	34

Рисунок 5.8 – Отчет о найденных аномалиях в данных

5.4. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Как подготовить рабочую область страницы книги Excel, чтобы отобразилась лента меню Analyze?
2. Какие инструменты интеллектуального анализа данных предоставляет надстройка Analyze?
3. Где осуществляется интеллектуальный анализ данных при использовании надстройки Analyze?
4. В какой форме надстройка Excel Analyze сохраняет отчеты по результатам анализа таблиц?

5.5. ЗАДАНИЕ ДЛЯ САМОСТОЯТЕЛЬНОГО ВЫПОЛНЕНИЯ

Для внеаудиторного самостоятельного выполнения практического задания можно использовать приложение MS Excel 2010 (или выше) с установленными надстройками интеллектуального анализа данных, опубликованное в облаке Microsoft Azure RemoteApp в гибридной коллекции Финансового университета.

Альтернативой может стать установка MS Excel 2010 (или выше) на домашний компьютер. В этом случае для скачивания программного обеспечения рекомендуется воспользоваться услугами Центра академического программного обеспечения Microsoft DreamSpark Premium Финансового университета, используя ссылку <https://e5.onthehub.com/WebStore/Welcome.aspx?vsro=8&ws=80324a2d-f658-e211-b607-f04da23e67f6>. После установки MS Excel следует установить надстройки интеллектуального анализа данных.

В результате самостоятельного выполнения задания должна быть создана книга Excel, содержащая отчеты по итогам интеллектуального анализа данных с помощью *всех* инструментов надстройки Analyze.

Рекомендуется использовать исходные данные из практикума (см. п. 5.3), изменяя постановку задачи анализа. Например, в самостоятельной работе при отработке навыков работы с инструментом «Анализ ключевых факторов влияния» можно проанализировать влияние региона и количества детей у клиента на его решение купить велосипед.

Условия проведения и результаты каждого вида анализа сохранить в одном документе Word. В названии документа указать номер группы и фамилию. Документ сохранить в личной папке. К отчету приложить файл Excel.

АНАЛИЗ ДАННЫХ С ПОМОЩЬЮ MICROSOFT POWER BI

6.1. ПОДКЛЮЧЕНИЕ К ХРАНИЛИЩУ ДАННЫХ

Инструмент Power BI доступен в качестве отдельного приложения Power BI Desktop или на портале powerbi.com.

Для подключения источника данных необходимо нажать иконку «Получить данные» и выбрать тип подключаемого источника (рис. 6.1).

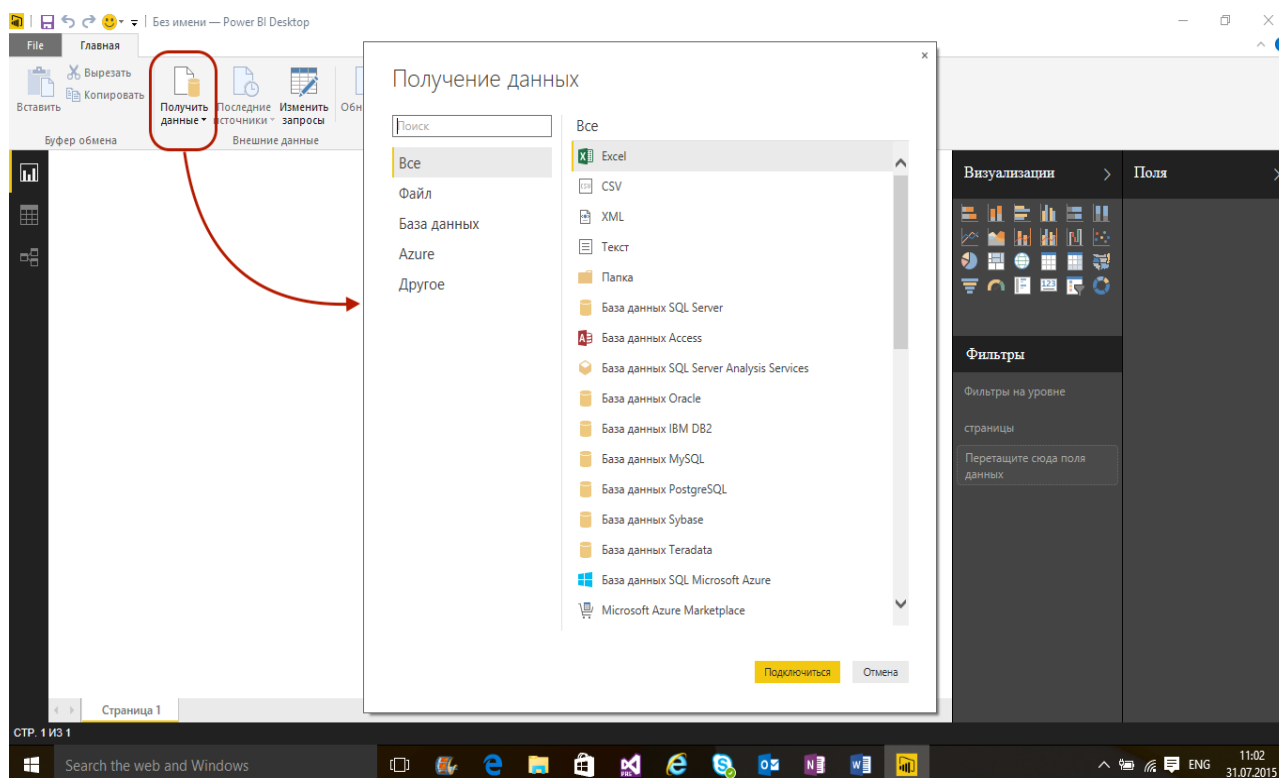


Рисунок 6.1 – Подключение источника данных

Power BI поддерживает множество форматов, таких как csv, xlsx, возможность загрузки данных из веба, SQL сервера, SharePoint, Dynamics CRM, GitHub, Oracle, Google Analytics и многие другие (рис. 6.2).

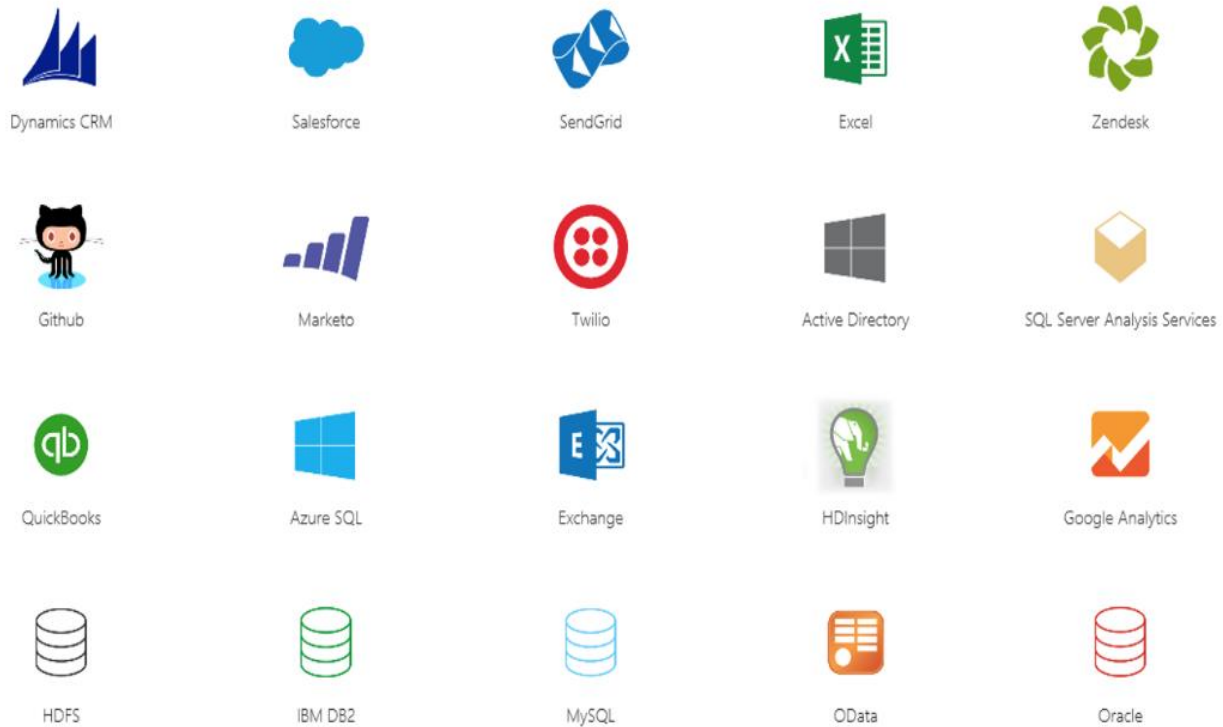


Рисунок 6.2 – Форматы данных, с которыми работает Power BI

В программе Power BI Desktop есть два режима – режим редактирования данных и режим создания отчетов. В режиме редактирования можно работать со своими данными – добавлять новые столбцы, объединять существующие, сортировать данные, менять их формат и т.д.

Во втором режиме, режиме создания отчетов, можно визуализировать данные, создавать красивые диаграммы и графики для их наилучшего представления, наносить данные на карту.

6.2. ОСНОВЫ ИСПОЛЬЗОВАНИЯ НАДСТРОЕК ПО ФОРМИРОВАНИЮ ОТЧЕТОВ В POWER BI

Главное окно системы представлено на рис. 6.3, 6.4.

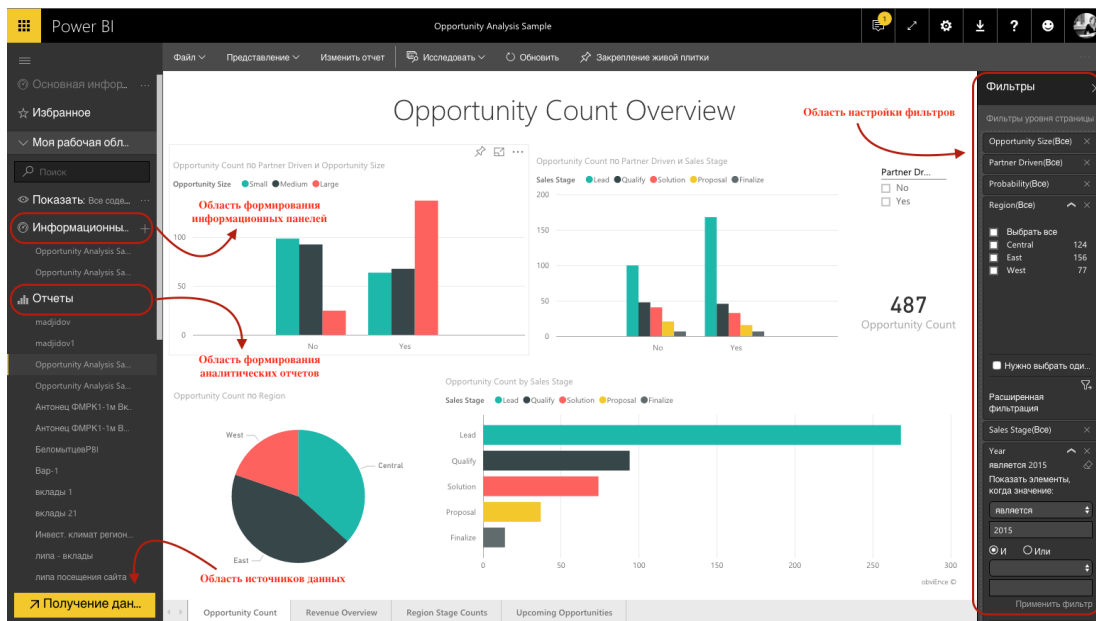


Рисунок 6.3 – Главное окно Power BI

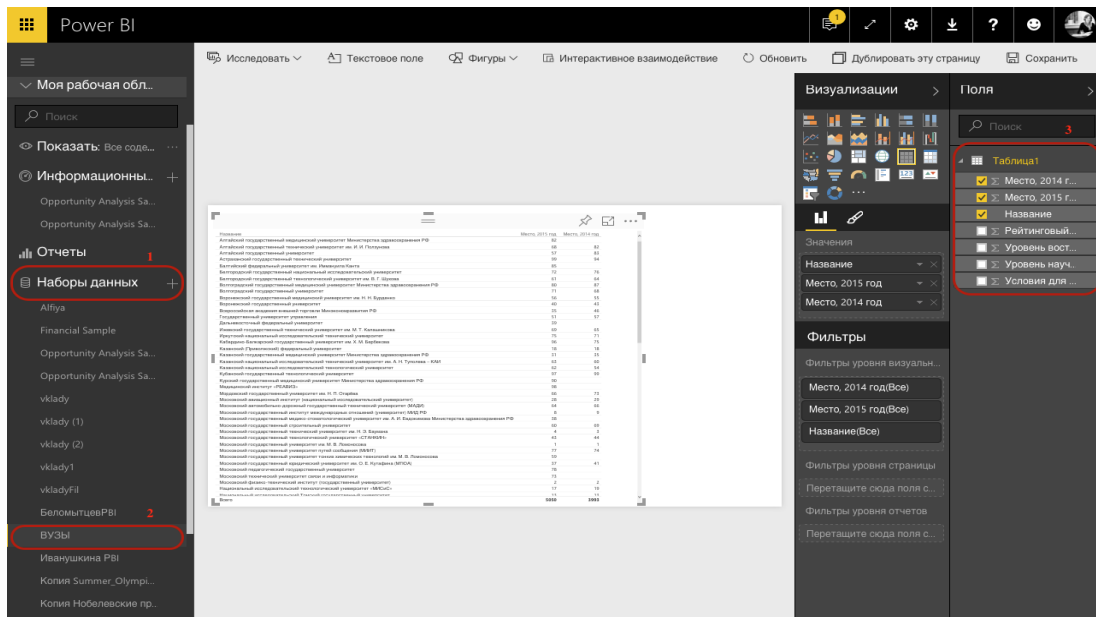


Рисунок 6.4 – Формирование аналитического отчета

Для формирования отчета необходимо в левой части главного окна выбрать набор данных, на основе которого необходимо разработать отчет (рис. 6.4).

На следующем этапе выбрать в правой части окна поля таблиц, которые необходимо отобразить в отчете (рис. 6.5).

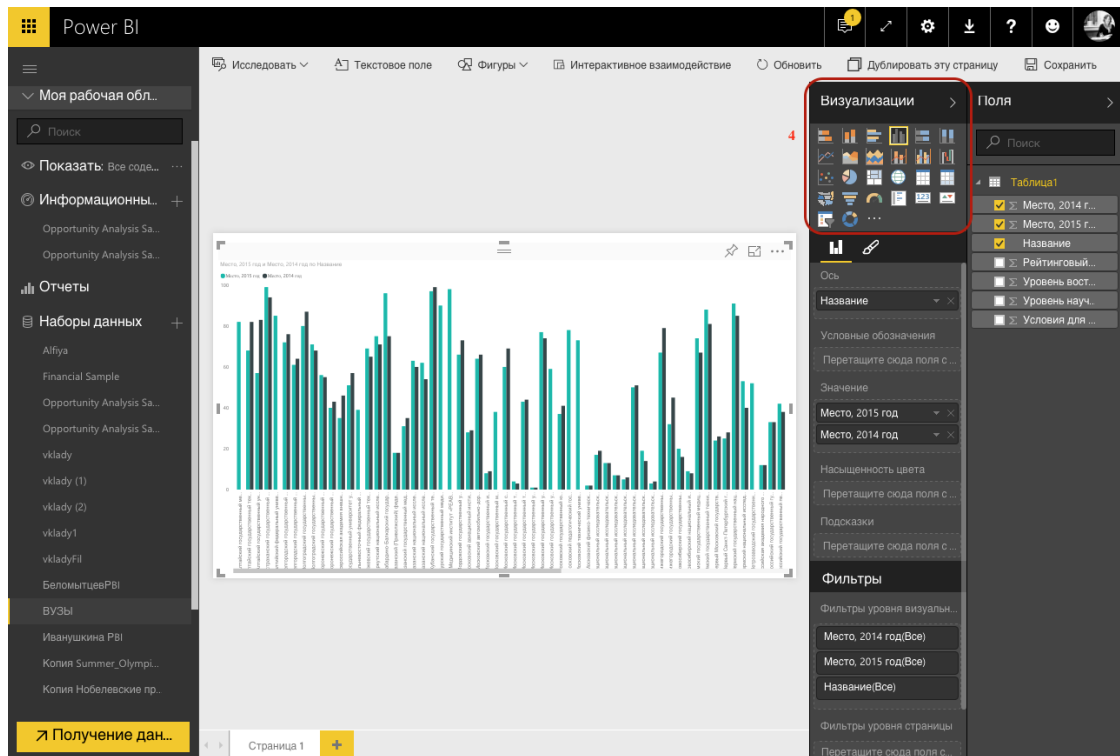


Рисунок 6.5 – Формирование аналитического отчета

Далее необходимо выбрать тип метафоры для визуализации данных и настроить фильтры.

В одном отчете может быть несколько визуализаций по разным полям одного набора данных.

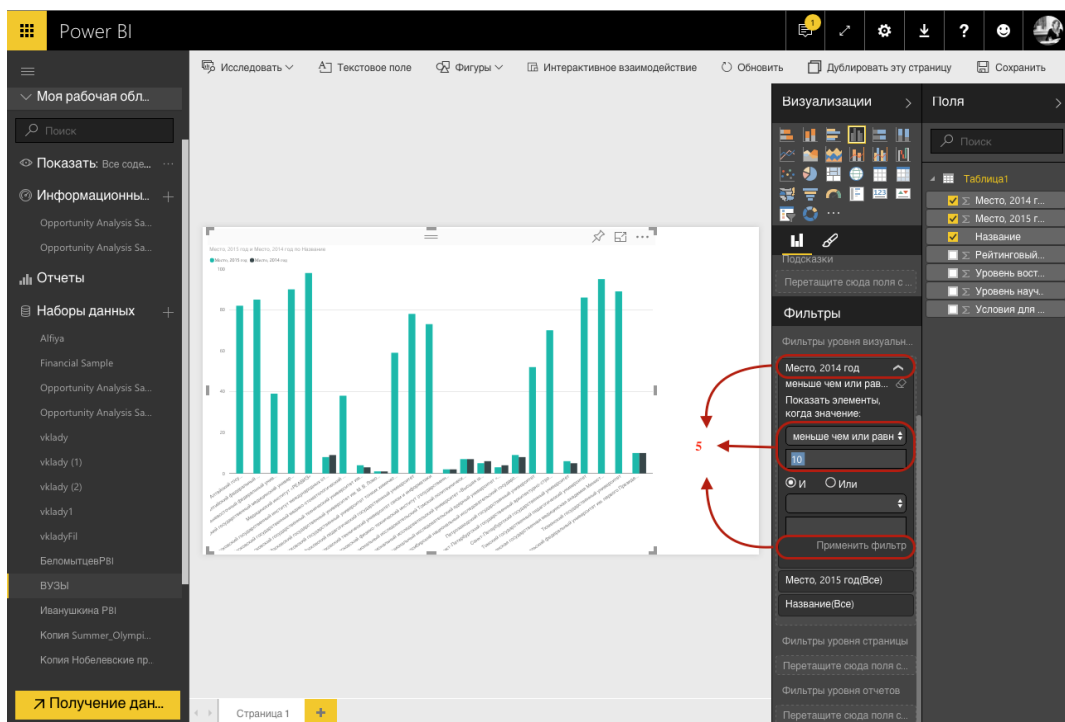


Рисунок 6.6 – Формирование аналитического отчета

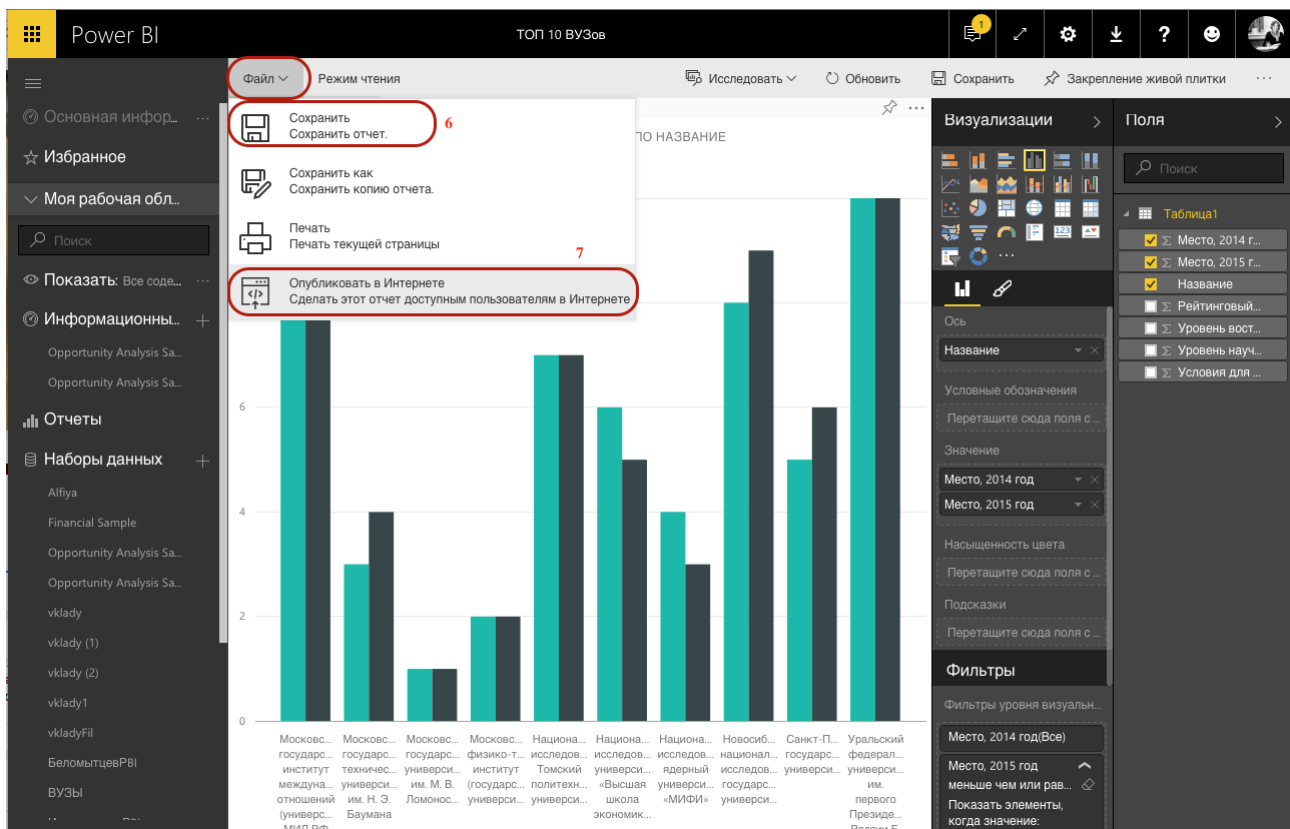


Рисунок 6.7 – Формирование аналитического отчета

Для того, чтобы сохранить отчет, необходимо выбрать пункт меню «Файл», «Сохранить. Сохранить отчет» (рис. 6.7).

6.3. ОСНОВЫ ИСПОЛЬЗОВАНИЯ ОБЛАЧНОГО СЕРВИСА POWERBI.COM

Для формирования информационной панели мониторинга – «Файл», «Опубликовать в интернете». В открывшихся окнах выбрать «Создать код внедрения» и «Опубликовать» для публикации на портале Power BI. По большей части, портал Power BI выполняет все те же функции, что и Power BI Desktop, за исключением отсутствия возможности редактирования данных. Информационная основа портала – панели мониторинга (рис. 6.8, 6.9).

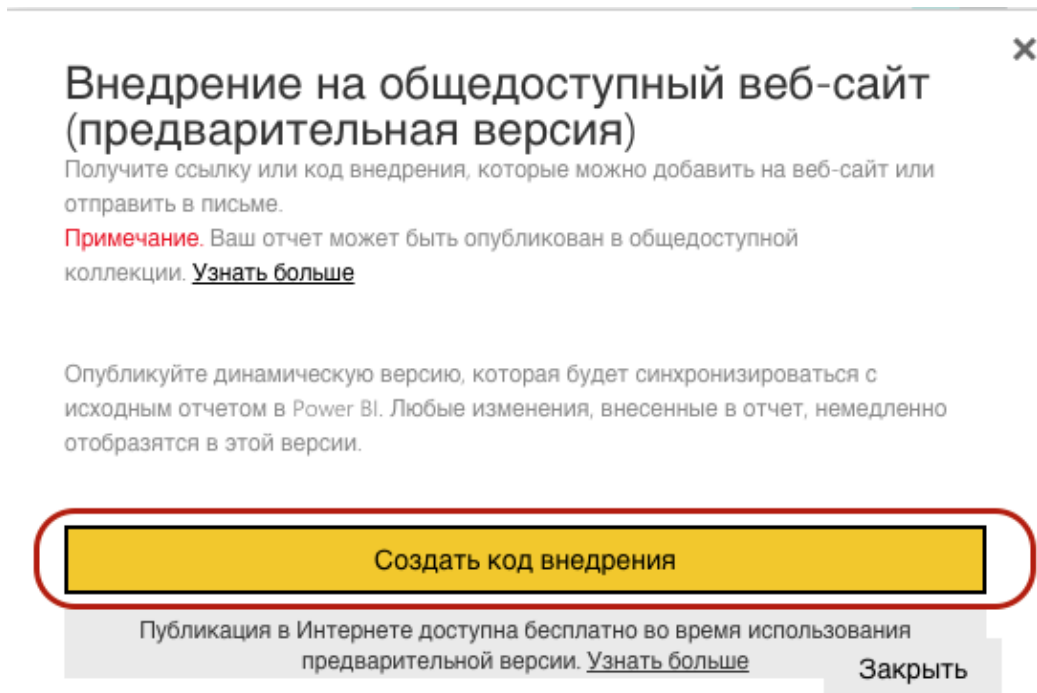


Рисунок 6.8 – Создание кода внедрения

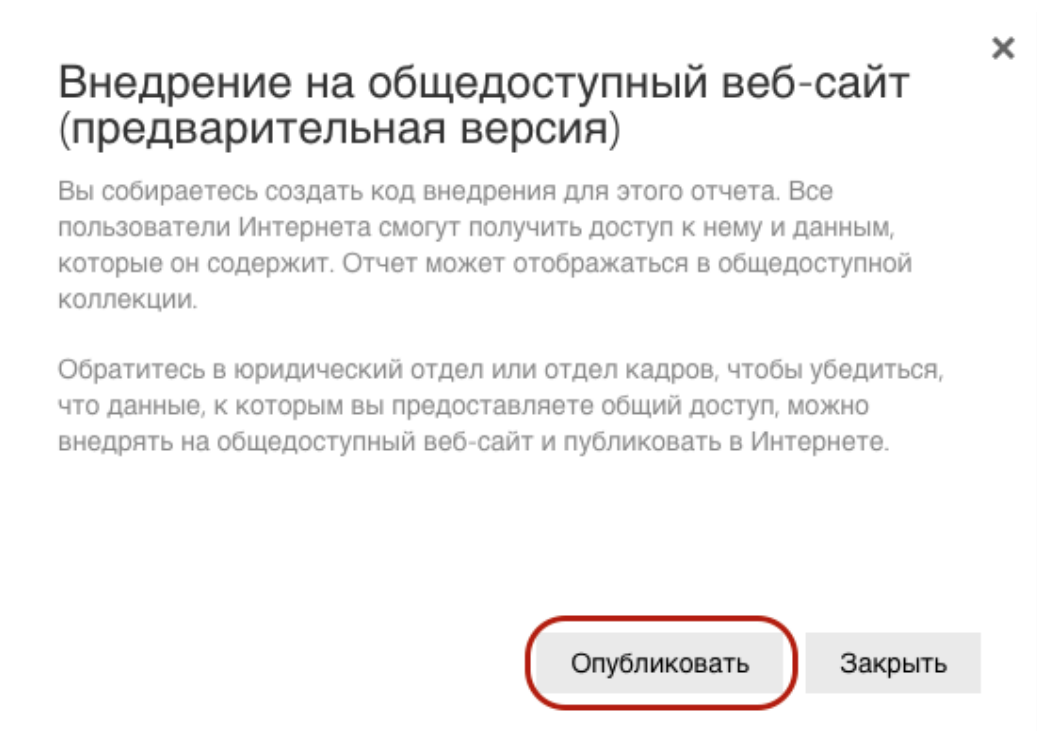


Рисунок 1 Создание кода внедрения

Сохранить полученный код.

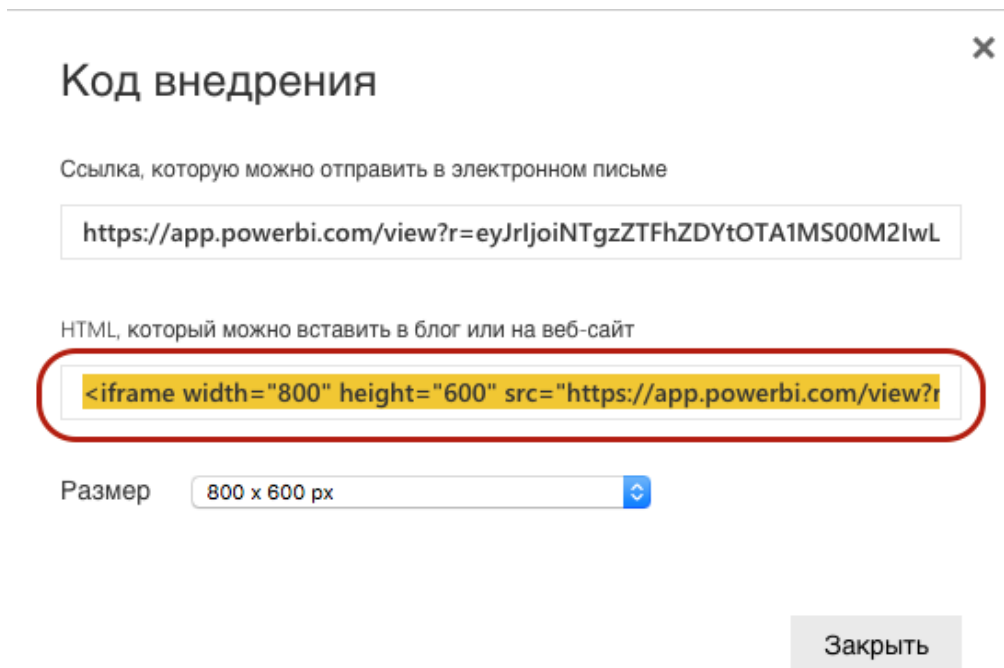


Рисунок 6.9 – Создание кода внедрения

Для создания информационной панели мониторинга необходимо нажать плюс в области информационных панелей и ввести наименование панели (рис. 6.10).

На следующем этапе необходимо нажать кнопку «Добавить плитку».

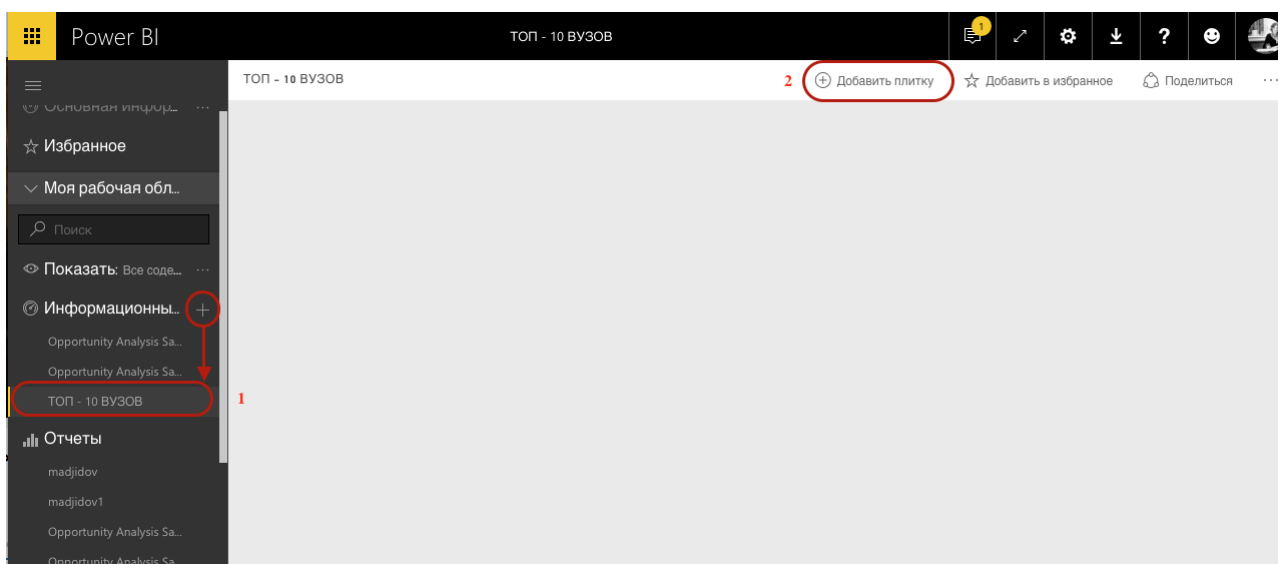


Рисунок 6.10 – Создание информационной панели мониторинга

Для добавления информации отчета на панель мониторинга необходимо выбрать кнопку «ВЕБ-содержимое». Также для формирование информативной панели можно добавить изображения, свободный текст, видео (рис. 6.11).

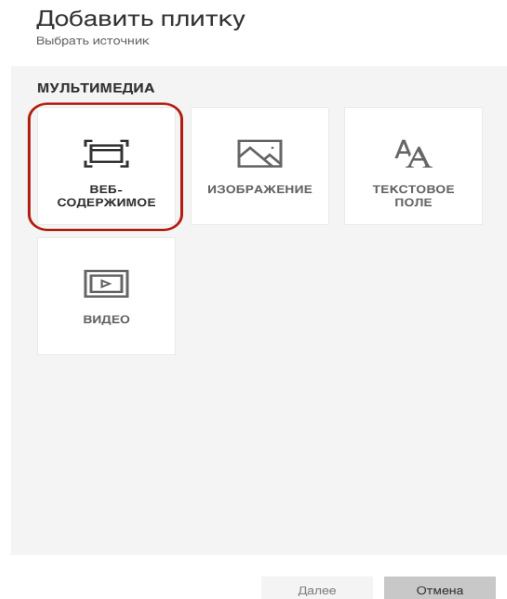


Рисунок 6.11 – Создание информационной панели мониторинга

Вставить сохраненный код с область «код внедрения» (рис. 6.12).

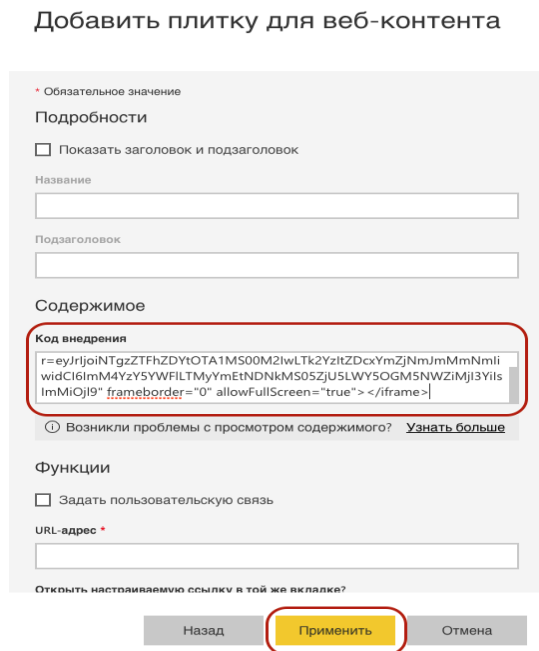


Рисунок 6.12 – Создание информационной панели мониторинга

Можно закреплять на панели мониторинга данные из различных отчетов и отслеживать их изменение в режиме реального времени. Также, на портале есть возможность поделиться отчетом со всеми, кто работает с этими данными (рис. 6.13).



Рисунок 6.13 – Полученная информационная панель мониторинга

6.4. ПРАКТИКУМ ПО ФОРМИРОВАНИЮ ОТЧЕТОВ В POWER BI

Цель практикума:

Закрепить теоретические знания и сформировать практические навыки анализа данных с помощью системы Microsoft Power BI.

План практикума:

Практикум включает в себя следующие этапы:

1. Изучение бизнес-кейса, включающего в себя описание ситуации.
2. Формирование отчетов (визуализаций), с помощью которых можно ответить на поставленные в бизнес-кейсе вопросы.
3. Загрузка источников данных.
4. Публикация полученных отчетов.

Порядок выполнения практикума

1. Изучите описание бизнес-кейса

БИЗНЕС-КЕЙС

ДАНА ИНФОРМАЦИЯ ПО УСЛОВИЯМ ВКЛАДОВ ВО ВСЕХ БАНКАХ РОССИЙСКОЙ ФЕДЕРАЦИИ:

- ПРОЦЕНТНАЯ СТАВКА;
- ВОЗМОЖНОСТЬ ПОПОЛНЕНИЯ ВКЛАДА;
- МИНИМАЛЬНАЯ СУММА ДЛЯ ОТКРЫТИЯ ВКЛАДА;
- МИНИМАЛЬНАЯ СУММА ПОПОЛНЕНИЯ.

ТАКЖЕ, ДАНА ИНФОРМАЦИЯ ПО САМИМ БАНКАМ:

- НАИМЕНОВАНИЕ БАНКА;
- МЕСТО БАНКА В РЕЙТИНГЕ БАНКОВ ПО ЧИСТЫМ АКТИВАМ.

НЕОБХОДИМО ВЫБРАТЬ САМЫЕ ВЫГОДНЫЕ ВКЛАДЫ В САМЫХ НАДЕЖНЫХ БАНКАХ С МИНИМАЛЬНОЙ ПЕРВОНАЧАЛЬНОЙ СУММОЙ.

ПОЛУЧИТЬ ДАННЫЕ О ВКЛАДАХ ИЗ ИМЕЮЩЕГОСЯ ФАЙЛА EXCEL (VKLADY.XLSX) ИЛИ НА САЙТЕ [HTTP://WWW.RBC.RU/RATINGS/](http://www.rbc.ru/ratings/).

2. Загрузите данные о вкладах в банках.

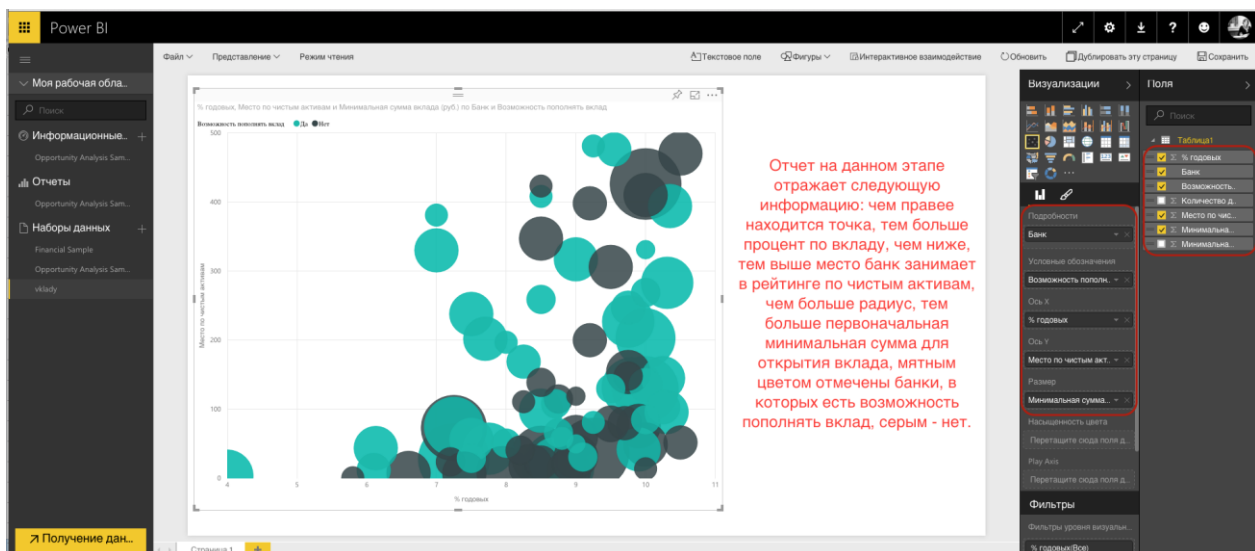


Рисунок 6.14 – Формирование аналитических отчетов

3. Задайте условия для первой визуализации: выбрать пузырьковую диаграмму (рис. 6.14).
4. Настройте фильтры, выбрав, % по вкладу больше 3, банк входит в топ-10 (место по чистым активам меньше или равно 10) (рис. 6.15).

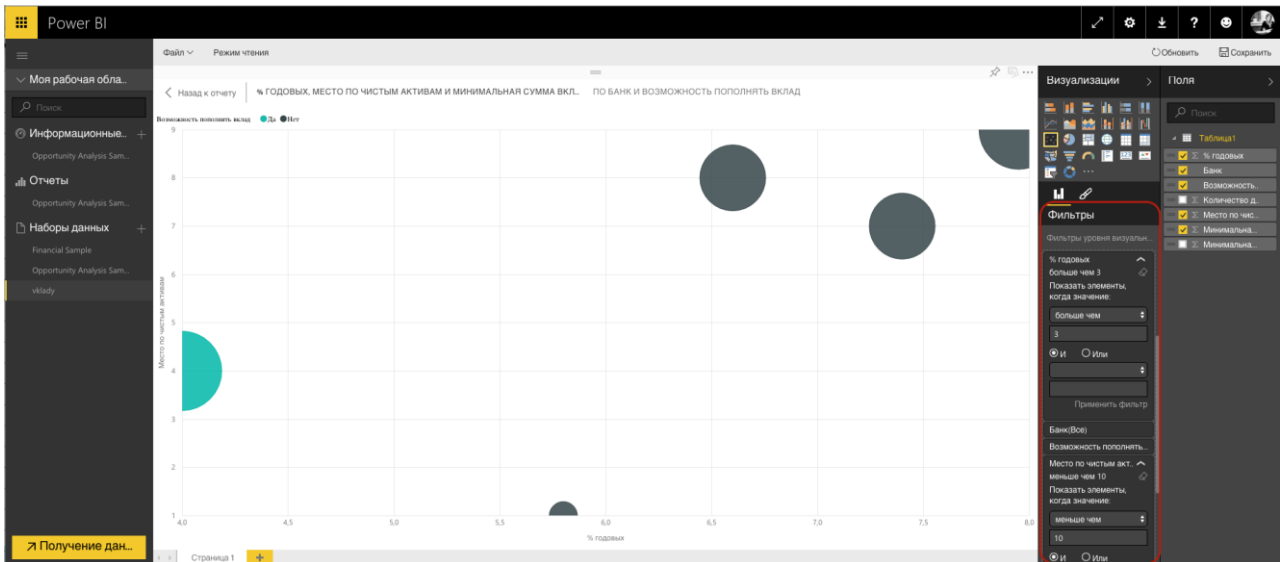


Рисунок 6.15 – Настройка фильтров

5. Добавьте другие визуализации по своему желанию.
6. Сформируйте информационную панель мониторинга (рис. 6.15).
7. Опубликуйте полученные отчеты (рис. 6.9)

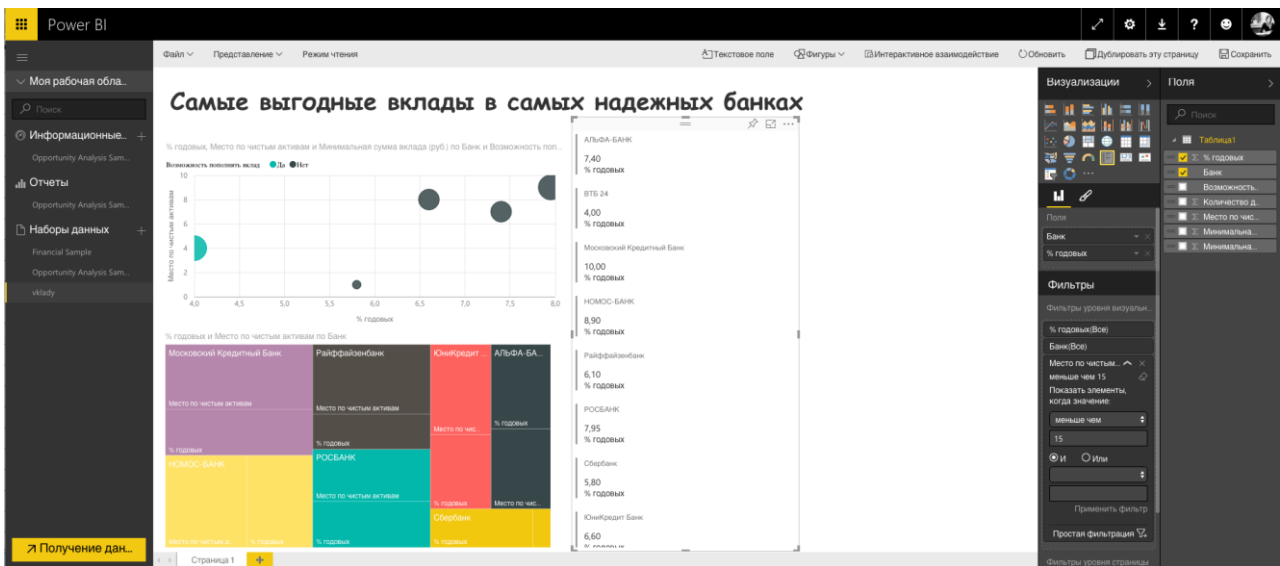


Рисунок 6.15 – Формирование информационной панели мониторинга

6.6. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Сколько режимов работы с данными поддерживает Power BI?
2. Существует ли возможность манипуляции с данными в портальной версии www.powerbi.com?
3. Существует ли возможность редактирования данных в портальной версии www.powerbi.com?
4. Каким образом можно с помощью power bi в режиме реального времени отслеживать динамику показателей?
5. Каким образом с помощью системы power bi можно анализировать информацию из открытых источников.

6.7. ЗАДАНИЕ ДЛЯ САМОСТОЯТЕЛЬНОГО ВЫПОЛНЕНИЯ

Для выполнения задания необходимо скачать систему Power BI Desktop по адресу <https://powerbi.microsoft.com/ru-ru/get-started/> либо воспользоваться системой установленной в аудиториях Финуниверситета. Также, можно использовать портальную версию системы, находящуюся по адресу www.powerbi.com. Для входа в систему можно использовать логин и пароль Microsoft, который автоматически формируются для всех студентов Финуниверситета:

- логин: номер_студенческого@edu.fa.ru;
- пароль: такой же как пароль для входа в компьютерных классах Финуниверситета.

Для закрепления полученных знаний и навыков в рамках самостоятельной работы каждому студенту необходимо реализовать минимум 5 аналитических отчетов и минимум одну информационную панель мониторинга на основе хранилища данных созданного ранее (пункты 3-4 данного пособия).

Требования к результатам выполнения задания:

- 1) Описание предметной области.
- 2) Формулировка основной цели и задач рассматриваемой организации.
- 3) Скриншоты отчетов и информационных панелей мониторинга, иллюстрирующих решение сформулированных задач.
- 4) Кроме того, полученные отчеты и панели мониторинга должны быть загружены на портал powerbi.com и выслана ссылка для проверки проекта.

АНАЛИЗ ДАННЫХ С ПОМОЩЬЮ СИСТЕМЫ КЛАССА DATA DISCOVERY – TABLEAU

7.1. ПОДКЛЮЧЕНИЕ К ХРАНИЛИЩУ ДАННЫХ И ВНЕШНИМ ИСТОЧНИКАМ

Для подключения источника данных необходимо выбрать в левой части в области «Подключение» (Connect) выбрать тип источника. В открывшемся окне выбрать файл-источник (рис. 7.1).

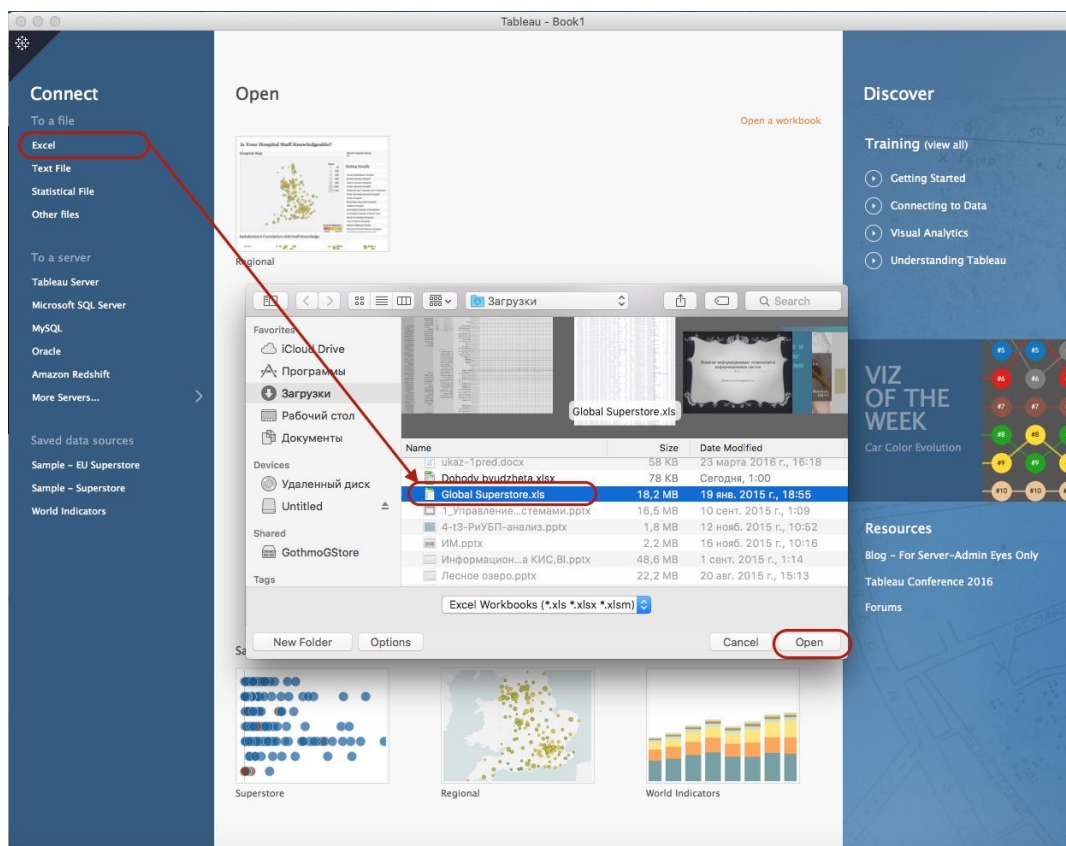


Рисунок 7.1 – Подключение источника данных

Если в виде источника, например, выбрана база данных, то поочередно необходимо выбрать таблицы с необходимыми для анализа данными. Если выбран файл Excel, поочередно выбрать нужные листы.

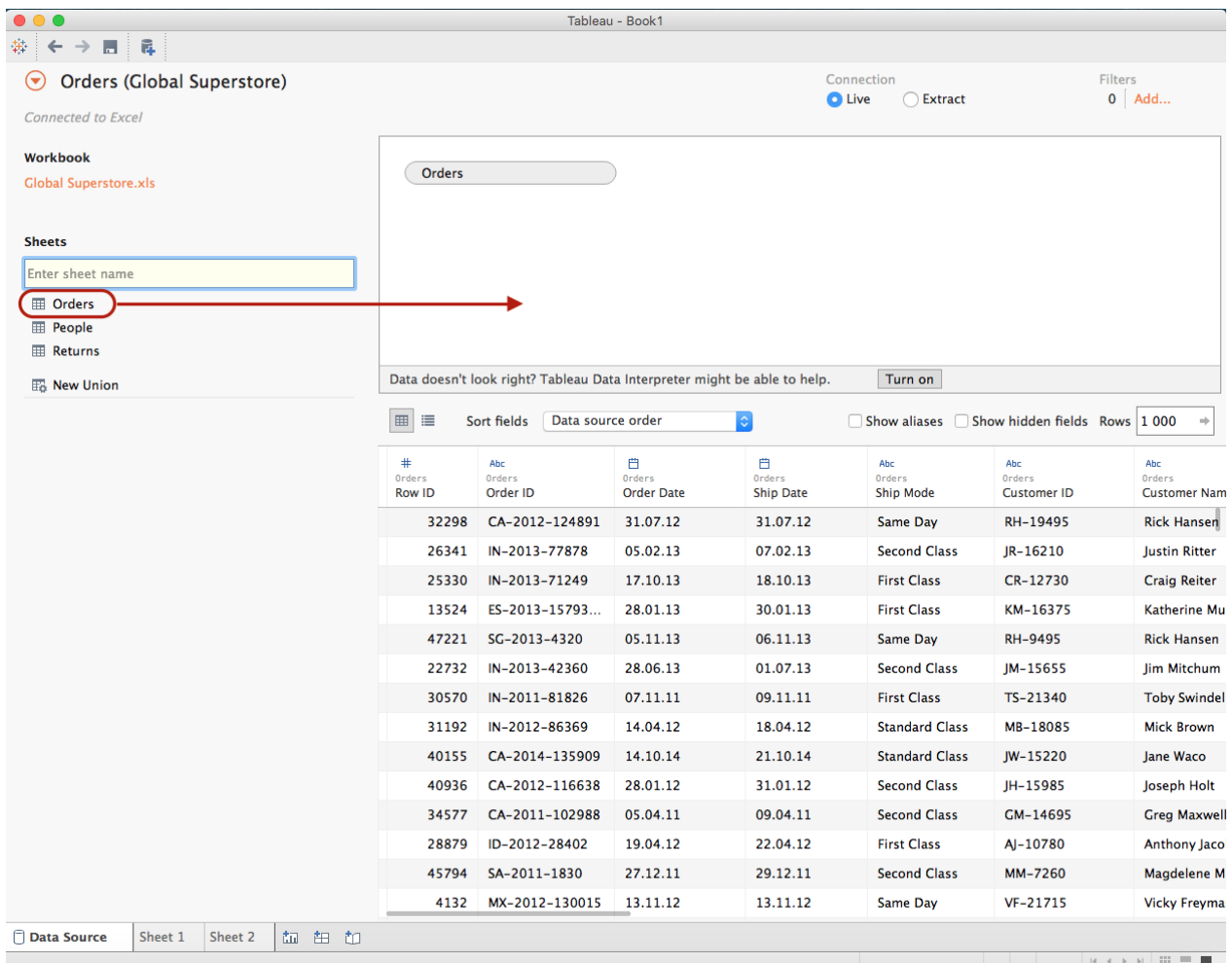


Рисунок 7.2 – Определение таблиц данных, на основе которых будет формироваться отчет

7.2. ОСНОВЫ ИСПОЛЬЗОВАНИЯ НАДСТРОЕК ПО РЕДАКТИРОВАНИЮ И АГРЕГАЦИИ ДАННЫХ В TABLEAU

Если для удобства построения отчетов необходимо сменить имена полей, нужно в контекстном меню выбранного поля нажать «Переименовать» (Rename) и ввести новое имя (рис. 7.3).

Orders Row ID	Orders Order ID	Orders Order Date	Orders Ship Date	Orders Ship Mode	Orders Customer ID
3229	IN-2011-134801	07.11.11	31.07.12	Same Day	RH-19495
2634	IN-2012-86369	14.04.12	07.02.13	Second Class	JR-16210
2533	CA-2014-135909	14.10.14	18.10.13	First Class	CR-12730
1352	CA-2012-116638	28.01.12	30.01.13	First Class	KM-16375
4722	CA-2011-102988	05.04.11	06.11.13	Same Day	RH-9495
2273	ID-2012-28402	19.04.12	01.07.13	Second Class	JM-15655
30570	IN-2011-81826	07.11.11	09.11.11	First Class	TS-21340
31192	IN-2012-86369	14.04.12	18.04.12	Standard Class	MB-18085
40155	CA-2014-135909	14.10.14	21.10.14	Standard Class	JW-15220
40936	CA-2012-116638	28.01.12	31.01.12	Second Class	JH-15985
34577	CA-2011-102988	05.04.11	09.04.11	Second Class	GM-14695
28879	ID-2012-28402	19.04.12	22.04.12	First Class	AJ-10780
45794	SA-2011-1830	27.12.11	29.12.11	Second Class	MM-7260
4132	MX-2012-130015	13.11.12	13.11.12	Same Day	VF-21715

Рисунок 7.3 – Переименование полей

Для изменения формата поля в контекстном меню выбрать новый формат (рис. 7.4).

Orders Order ID	Orders Order Date	Orders Ship Date	Orders Ship Mode	Orders Customer ID	Orders Customer Name	
891	31.07.12	31.07.12	Same Day	RH-19495	Rick Hansen	
8	05.02.13	07.02.13	Second Class	JR-16210	Justin Ritter	
9	17.10.13	18.10.13	First Class	CR-12730	Craig Reiter	
13...	28.01.13	30.01.13	First Class	KM-16375	Katherine Murray	
13...	05.11.13	06.11.13	Same Day	RH-9495	Rick Hansen	
22732	IN-2013-42360	28.06.13	01.07.13	Second Class	JM-15655	Jim Mitchum
30570	IN-2011-81826	07.11.11	09.11.11	First Class	TS-21340	Toby Swindell
31192	IN-2012-86369	14.04.12	18.04.12	Standard Class	MB-18085	Mick Brown
40155	CA-2014-135909	14.10.14	21.10.14	Standard Class	JW-15220	Jane Waco
40936	CA-2012-116638	28.01.12	31.01.12	Second Class	JH-15985	Joseph Holt
34577	CA-2011-102988	05.04.11	09.04.11	Second Class	GM-14695	Greg Maxwell
28879	ID-2012-28402	19.04.12	22.04.12	First Class	AJ-10780	Anthony Jacobs
45794	SA-2011-1830	27.12.11	29.12.11	Second Class	MM-7260	Magdelene Mors
4132	MX-2012-130015	13.11.12	13.11.12	Same Day	VF-21715	Vicky Freymann

Рисунок 7.4 – Изменение формата поля

При неверном импорте данных из источника и необходимости разбить полученное поле на несколько полей, выбрать в контекстном меню «Расщепление» (Split) (рис. 7.5).

#	Orders ID	Orders Ship Mode	Orders Customer ID	Orders Customer Name
32298	CA-2012-12489	Same Day	RH-19495	Rick Hansen
26341	IN-2013-77878	Second Class	JR-16210	Justin Ritter
25330	IN-2013-71249	First Class	CR-12730	Craig Reiter
13524	ES-2013-15793	First Class	KM-16375	Katherine Murray
47221	SG-2013-4320	Same Day	RH-9495	Rick Hansen
22732	IN-2013-42360	Second Class	JM-15655	Jim Mitchum
30570	IN-2011-81826	First Class	TS-21340	Toby Swindell
31192	IN-2012-86369	Standard Class	MB-18085	Mick Brown
40155	CA-2014-135909	Standard Class	JW-15220	Jane Waco
40936	CA-2012-116638	Second Class	JH-15985	Joseph Holt
34577	CA-2011-102988	Second Class	GM-14695	Greg Maxwell
28879	ID-2012-28402	First Class	AJ-10780	Anthony Jacobs
45794	SA-2011-1830	Second Class	MM-7260	Magdelene Morse
4132	MX-2012-130015	Same Day	VF-21715	Vicky Freymann

Рисунок 7.5 – Разбиение поля

Для добавления дополнительных полей из источника данных, в левой области источников выбрать необходимую таблицу или лист. В открывшемся окне выбрать правило объединения данных и поля, по которым будет формироваться связь (рис. 7.6).

Orders Ship Date	Orders Ship Mode	Orders Customer ID	Orders Customer Name	Orders Segment	Orders City	Orders State
31.07.12	Same Day	RH-19495	Rick Hansen	Consumer	New York C...	New York
07.02.13	Second Class	JR-16210	Justin Ritter	Corporate	Wollongong	New South Wa...
18.10.13	First Class	CR-12730	Craig Reiter	Consumer	Brisbane	Queensland
30.01.13	First Class	KM-16375	Katherine Murray	Home Office	Berlin	Berlin
06.11.13	Same Day	RH-9495	Rick Hansen	Consumer	Dakar	Dakar
01.07.13	Second Class	JM-15655	Jim Mitchum	Corporate	Sydney	New South Wa...
09.11.11	First Class	TS-21340	Toby Swindell	Consumer	Porirua	Wellington
18.04.12	Standard Class	MB-18085	Mick Brown	Consumer	Hamilton	Waikato
21.10.14	Standard Class	JW-15220	Jane Waco	Corporate	Sacramento	California
31.01.12	Second Class	JH-15985	Joseph Holt	Consumer	Concord	North Carolina
09.04.11	Second Class	GM-14695	Greg Maxwell	Corporate	Alexandria	Virginia
22.04.12	First Class	AJ-10780	Anthony Jacobs	Corporate	Kabul	Kabul
29.12.11	Second Class	MM-7260	Magdelene Morse	Consumer	Jizan	Jizan
13.11.12	Same Day	VF-21715	Vicky Freymann	Home Office	Toledo	Parana

Рисунок 7.6 – Добавление данных к источнику

Для формирования нового отчета необходимо в нижней части окна нажать кнопку создания нового отчета. В результате откроется окно формирования отчета.

Левая часть окна содержит область измерений и область фактов. При неверном автоматическом импорте и необходимости поменять измерение на факт, нужно перенести атрибут из области измерений в область фактов (при необходимости изменить факт на измерение, операция производится аналогично). В области параметров можно создать расчетный показатель. Для этого в контекстном меню нужно выбрать «Создать новый расчетный показатель» (Create calculated field) (рис. 7.7).

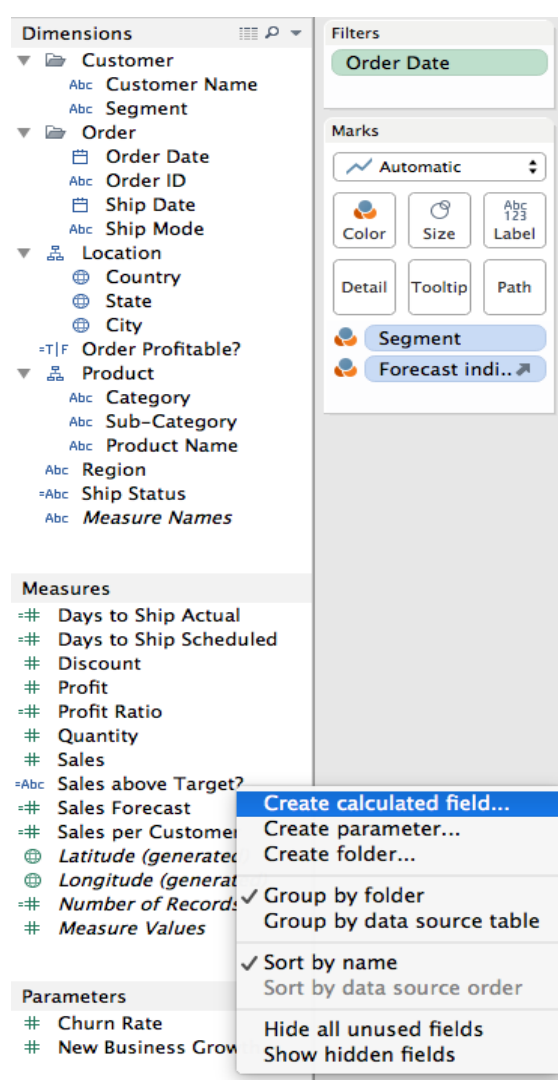


Рисунок 7.7 – Создание расчетного показателя

Для настройки фильтров по какому-либо из показателей, необходимо этот показатель перенести мышкой в область «Фильтры» (Filters).

Область «Метки» (Marks) позволяет отображать дополнительные свойства анализируемой информации за счет цвета, размера, начертания. Например, для раскрашивания данных в разные цвета в зависимости от какого-либо показателя или измерения, необходимо этот показатель перенести в зону «Цвет» (Color) области «Метки» (Marks). Аналогично, можно показать дополнительные свойства информации за счет размера объекта визуализации перенеся мышкой (рис. 7.8).

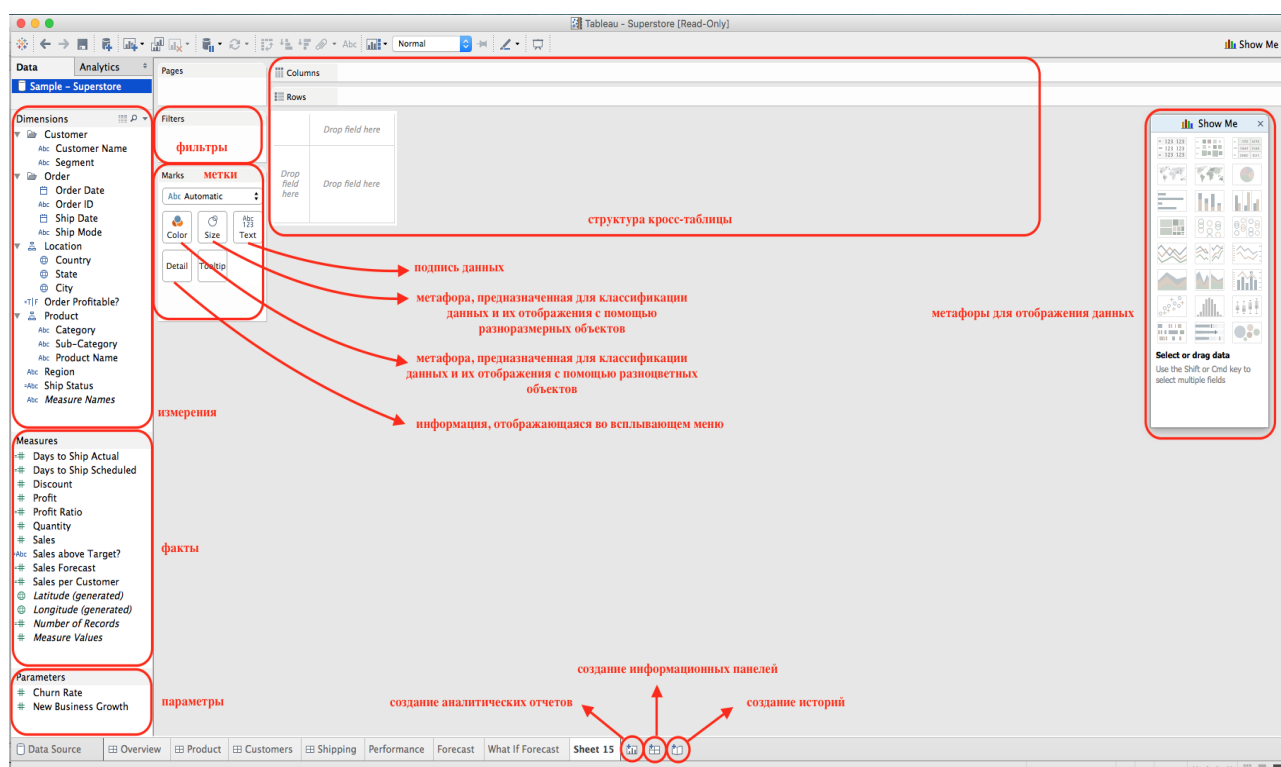


Рисунок 7.8 – Окно формирования отчетов

Область кросс-таблицы предназначена для определения, какие данные откладываются по строкам, какие по столбцам. Метафоры необходимы для визуализации данных в виде графиков, гистограмм, бабл-чатов, карт и т.д.

7.3. ПРАКТИКУМ ПО ФОРМИРОВАНИЮ ОТЧЕТОВ В TABLEAU

Цель практикума

Закрепить теоретические знания и сформировать практические навыки анализа данных с помощью системы Tableau.

План практикума

Практикум включает в себя следующие этапы:

1. Изучение бизнес-кейса, включающего в себя описание ситуации.
2. Загрузка источников данных.
3. Формирование аналитических отчетов.
4. Формирование информационных панелей.
5. Формирование историй.
6. Публикация полученных ответов.

Порядок выполнения практикума

1. Изучение бизнес-кейса

БИЗНЕС-КЕЙС

Торговая компания, занимающаяся продажами техники по всему миру. Имеются следующие данные:

- номер заказа;
- дата заказа;
- дата и условия доставки;
- личные данные о покупателе;
- данные о месте продажи: рынки сбыта (экономические пространства, объединяющие страны, где осуществляются предложения продажи и услуг), части света, страны, регионы, города.
- данные о продуктах: категория, суб-категория, наименование товара;
- данные о продажах: выручка, прибыль, количество проданных товаров, скидки.
- данные о возвратах.

Необходимо проанализировать следующую информацию:

- Объем продаж в разрезе дней. Выявить дни с самым большим объемом продаж.
- Объем продаж в разрезе рынков сбыта.
- Прибыль в разрезе продуктов.

2. Для анализа объемов продаж необходимо в области измерений (dimensions) выбрать поле «Дата заказа» (Order Date) и поместить его в колонки (columns). В строки (Rows) перенести «Объемы продаж» (Sales) из области фактов (Measures) (рис. 7.9).

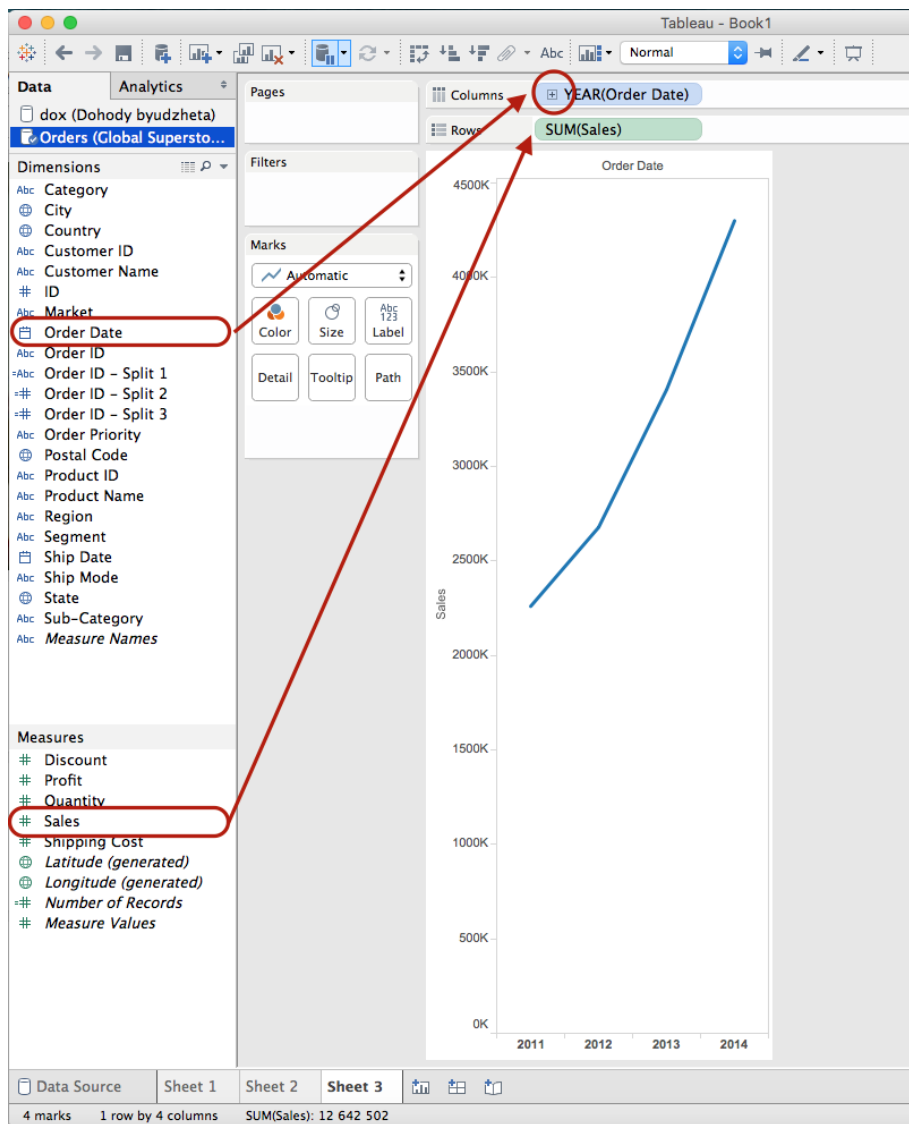


Рисунок 7.9 – Создание аналитического отчета

Для детализации данных по временным периодам необходимо нажать на знак + у поля «Год» (Year.Order Date). Детализировать до получения необходимого временного интервала.

Для отображения данных о продажах по годам разными цветами, где каждому году соответствует свой цвет, необходимо перенести поле «Год» (YEAR(Order Date)) в область «Метки» (Marks), зону «Цвет» (Color) (Рис.7.10).

Если по получившемуся показателю «День» (DAY(Order_Date)) нажать правой кнопкой мыши, откроется контекстное меню, где также можно выбрать детализацию временного показателя. Если детализацию выбрать в верхней части меню, то произойдет группировка по выбранному временному интервалу (например, если выбрать «Месяц» (Month), то январь 2011 просуммируется с январем

2012, 2013 и т.д. годов и так по всем месяцам). Если же детализацию выбрать в нижней части контекстного меню, то будут отображены все данные без группировки в зависимости от заданного интервала.

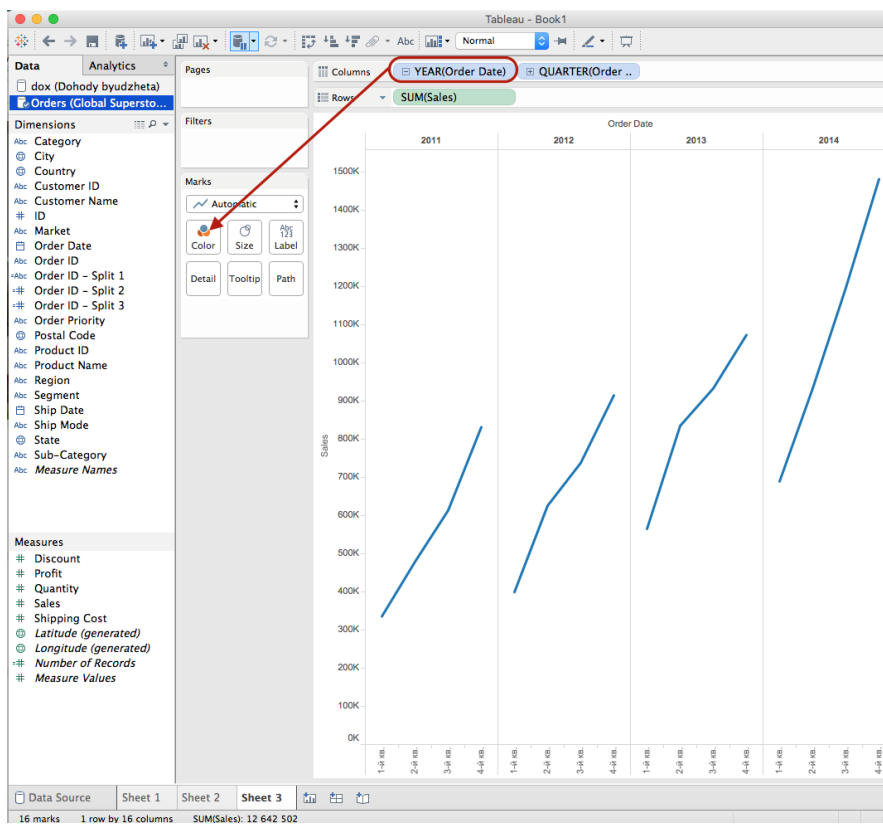


Рисунок 7.10 – Создание аналитического отчета

Для анализа объема продаж на каждый день, необходимо выбрать «День» (Day) в нижней части контекстного меню (рис. 7.11).

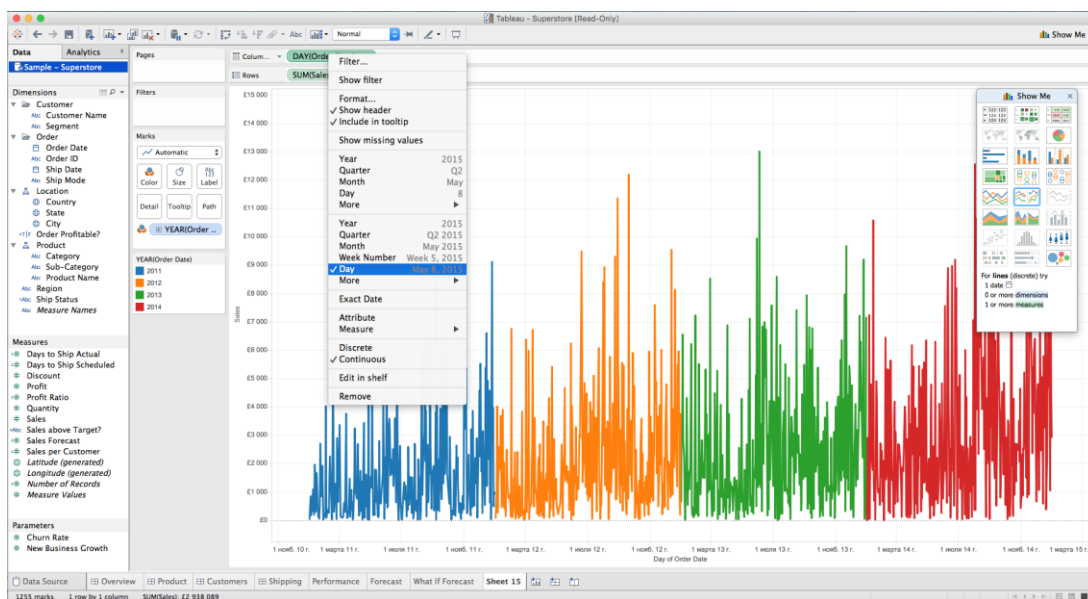


Рисунок 7.11 – Создание аналитического отчета

1) Для анализа объема продаж по рынкам сбыта необходимо добавить измерения «Категория» (Category) и «Суб-категория» (Sub-category) в колонки области кросс-таблицы. Измерение «Рынки» (Market) и факт «Объем продаж» (SUM(Sales)) в строки области кросс-таблицы. Измерение «Рынки» (Market) появится в области измерений после добавление к источнику данных листа «Возвращения» (Returns) – пункт 7.2.

Выбрать информацию по следующим рынкам сбыта:

- Азиатско-тихоокеанский регион и Япония – APAC;
- Европейский союз – EU;
- Латинская Америка – LATAM;
- США – US.

Для фильтрации по рынкам сбыта перенести в область фильтров (Filters) измерение «Рынки» (Market). В открывшемся окне выбрать необходимые рынки сбыта.

Для ввода дополнительного параметра, показывающего прибыль, необходимо перетащить мышкой факт «Прибыль» (Profit) в зону «Цвет» (Color) области «Метки» (Marks). Выбрать визуализацию – гистограмма (рис. 7.12).

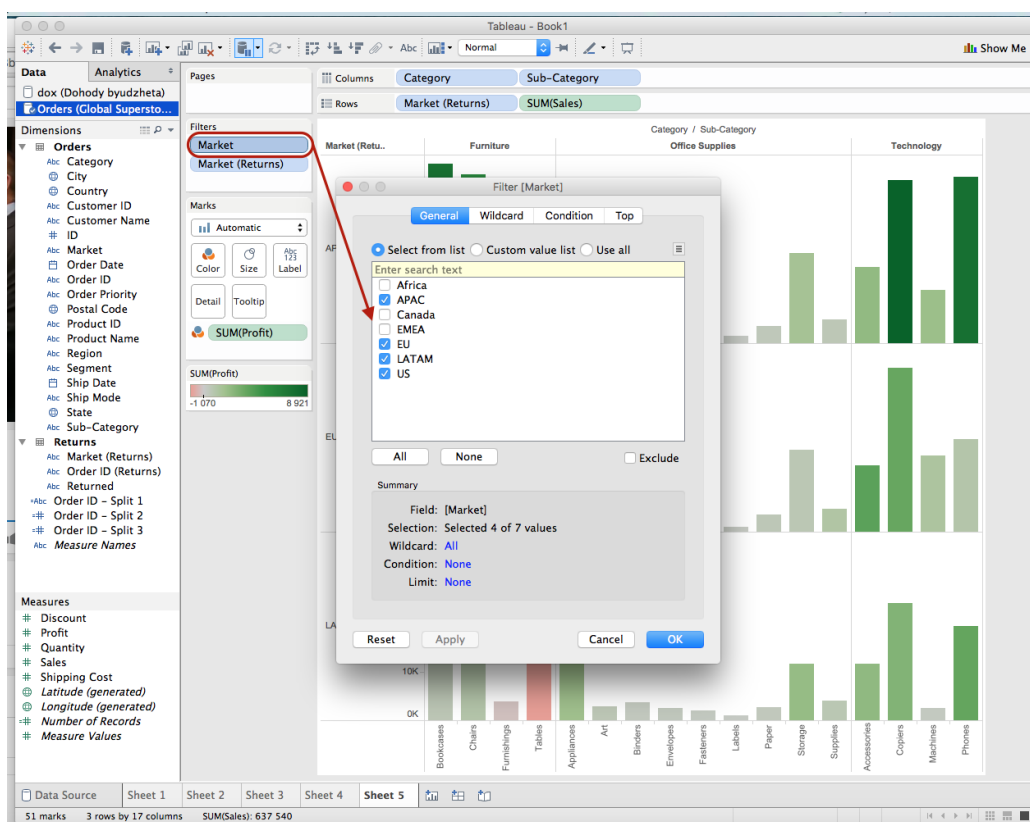


Рисунок 7.12 – Создание аналитического отчета

Для исключения нулевых данных выбрать правой кнопкой мыши столбец NULL и нажать «Исключить» (Exclude) (Рис.7.13).

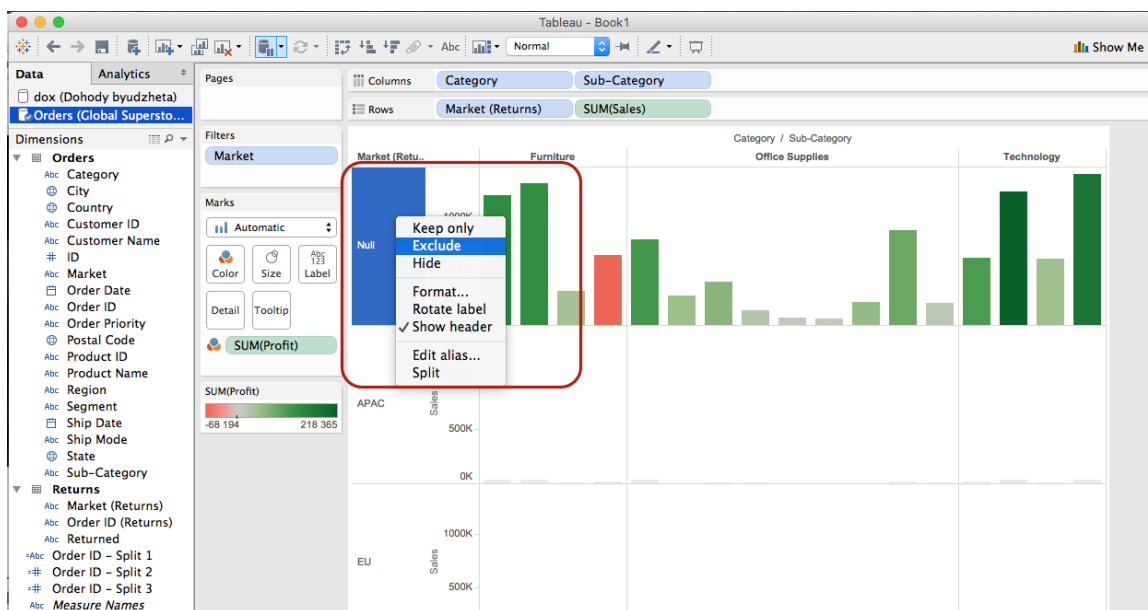


Рисунок 7.13 – Создание аналитического отчета

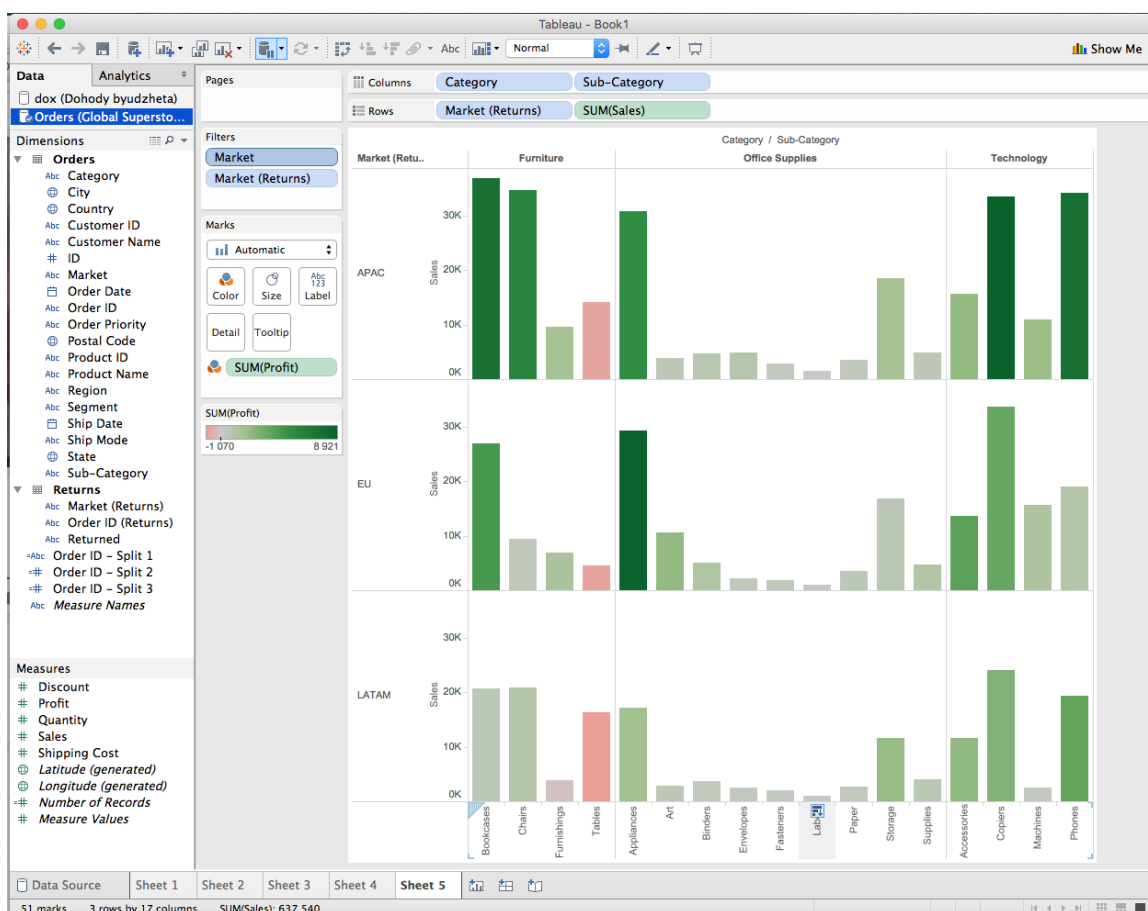


Рисунок 7.14 – Создание аналитического отчета

2) Для анализа прибыли в область кросс-таблицы добавить измерения «Категория» (Category) и «Суб-категория» (Sub-category) в строки и факт «Прибыль» (SUM(Profit)).

Для создания иерархии «Продукт» (Product) перетащить мышкой измерение «Суб-категория» (Sub-category) в «Категория» (Category). В открывшемся окне ввести наименование иерархии «Продукт» (Product) (продукт – это категория и входящие в нее суб-категории) (Рис. 7.15–7.19).

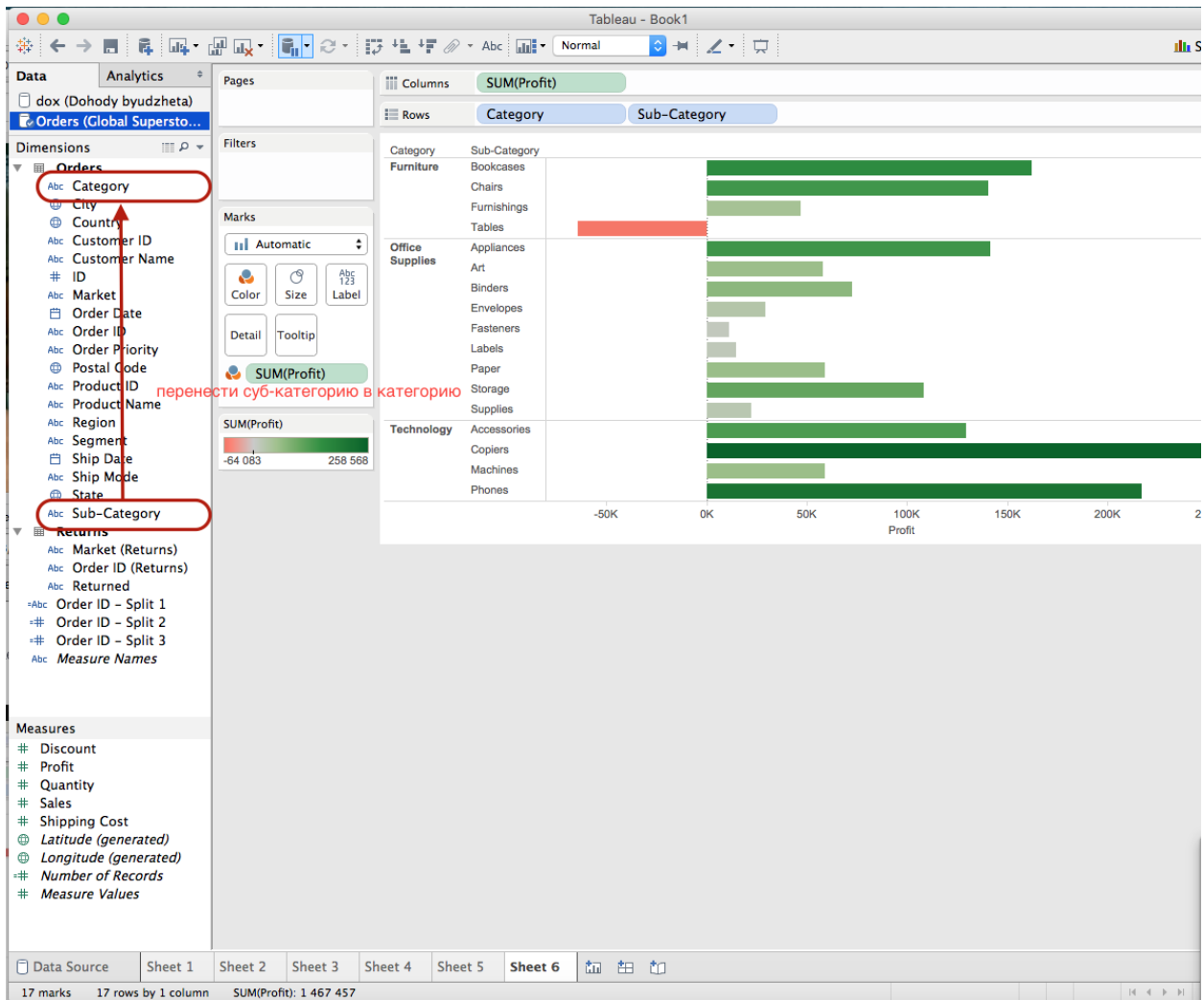


Рисунок 7.15 – Создание иерархии

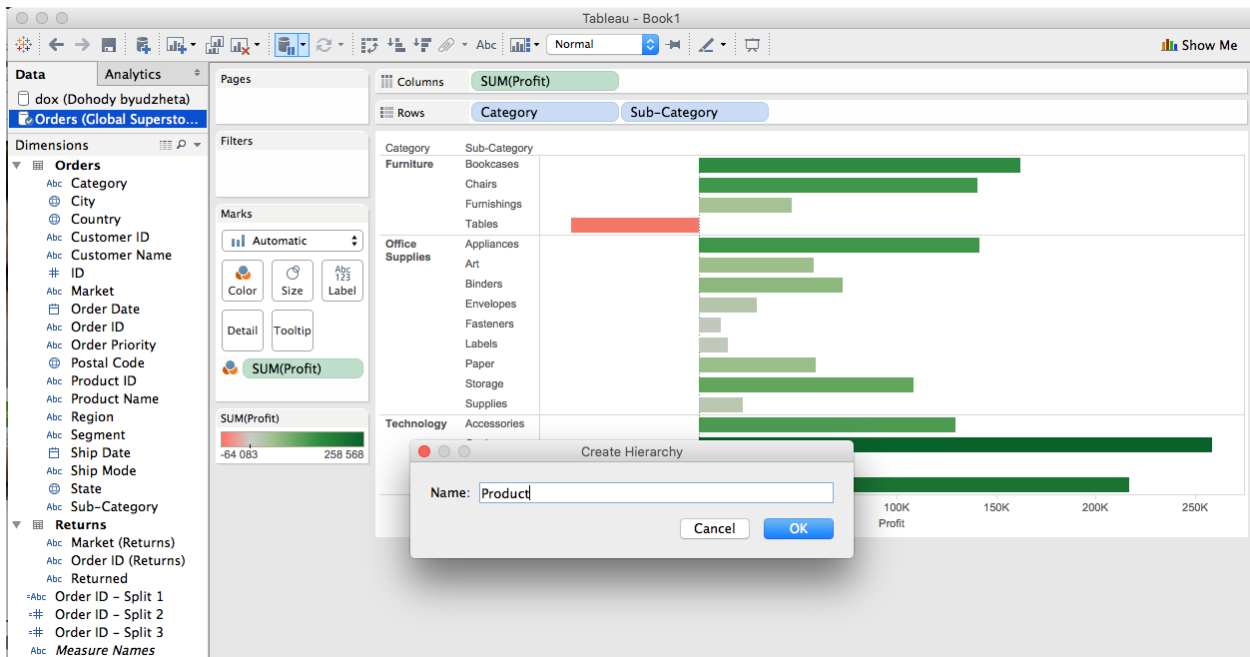


Рис. 7.16 – Создание иерархии

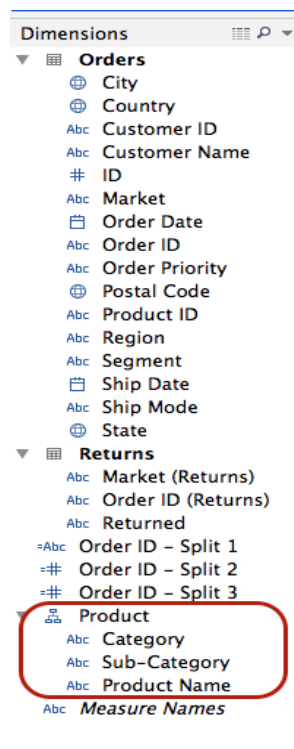


Рисунок 7.17 – Иерархия показателей

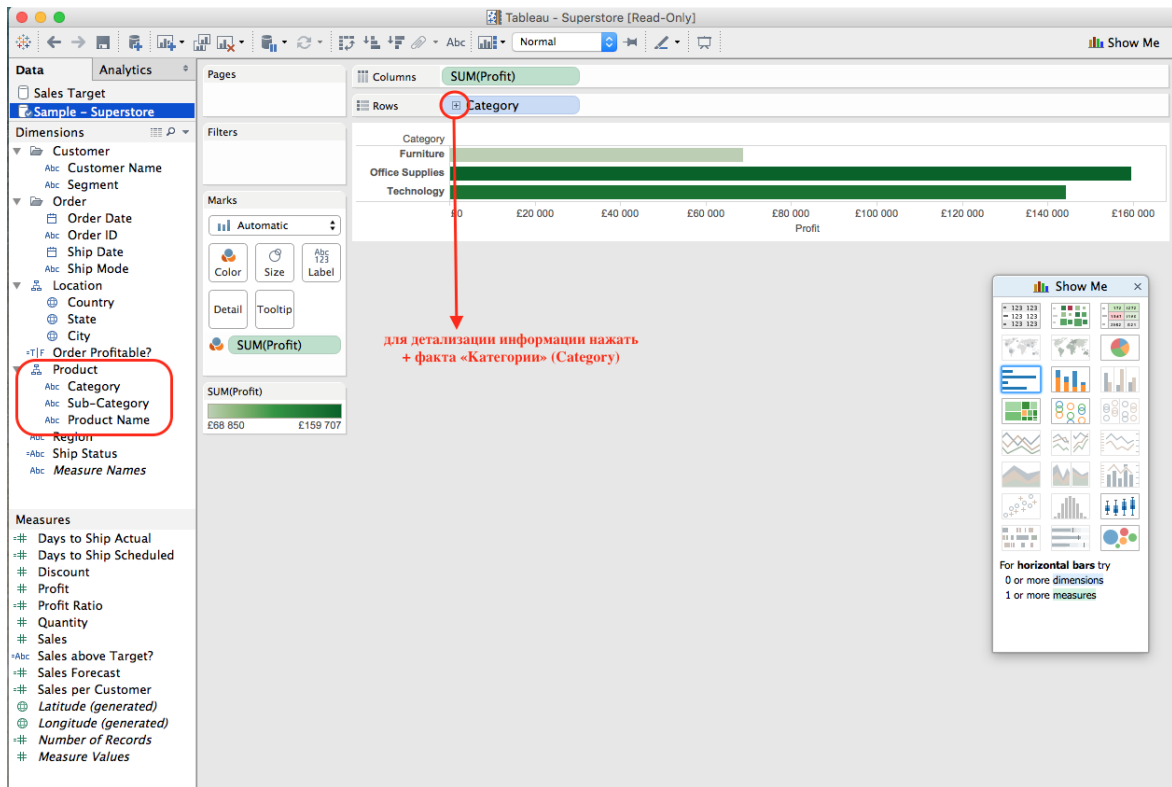


Рисунок 7.18 – Иерархия показателей

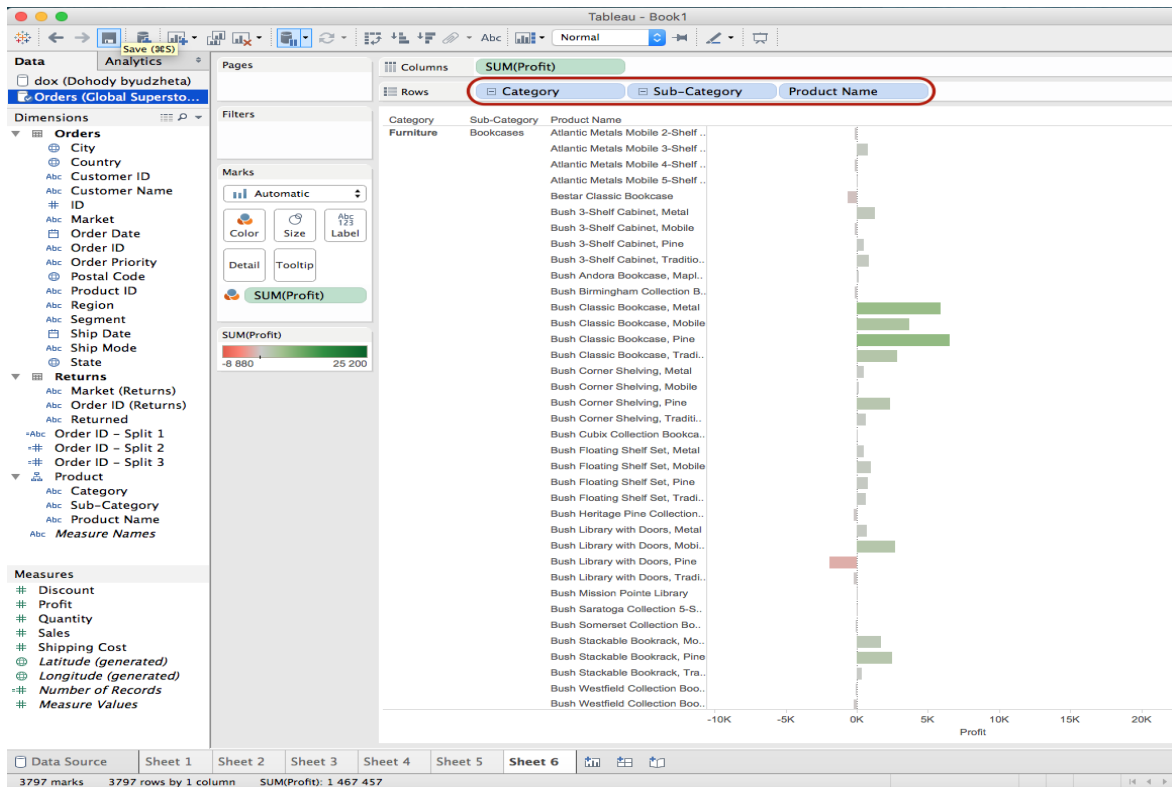



Рисунок 7.19 – Создание аналитического отчета

Сформировать еще несколько отчетов на основе данных в продажах.

2. Для формирования информационной панели мониторинга необходимо в нижней части окна добавить новую информационную панель (dashboard), нажав кнопку создания информационных панелей . В открывшемся окне, в левой части перечислены все сформированные отчеты. Необходимо перенести отчеты в область формирования информационной панели. Для более наглядной визуализации можно добавить картинки, интернет-источники, текст. Можно настроить расположение, высоту и ширину каждого окна (рис. 7.20).

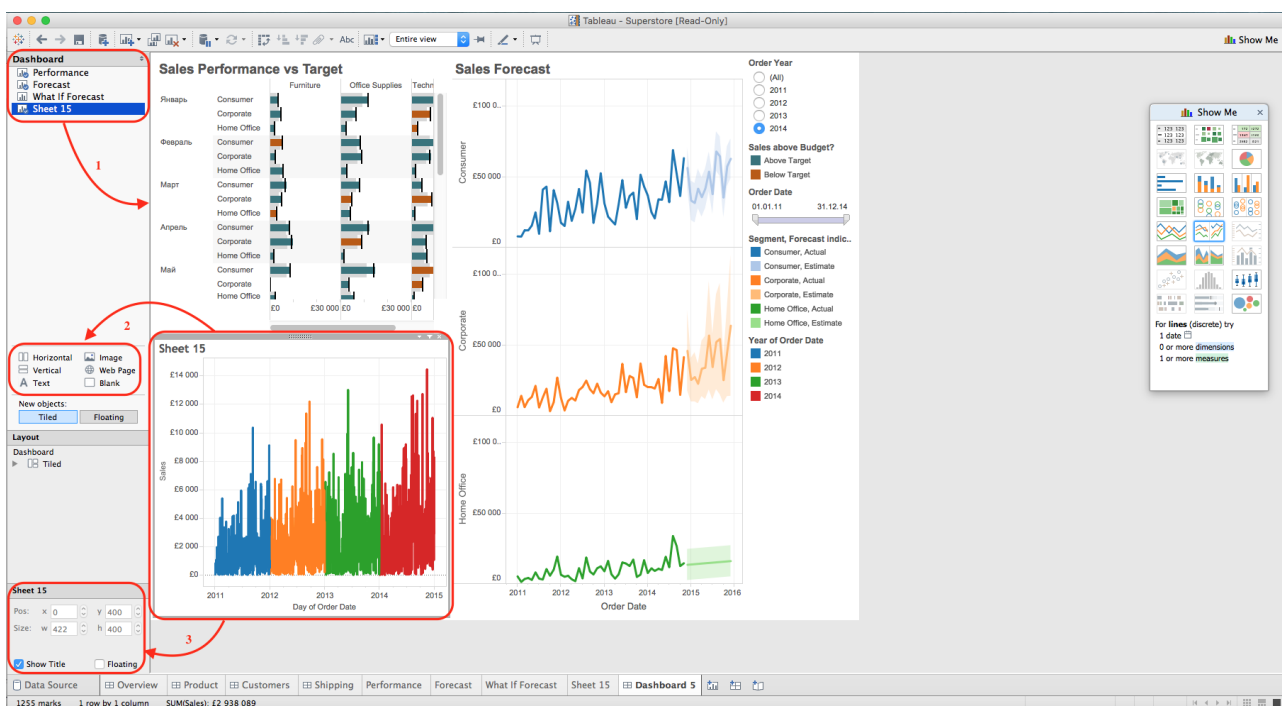



Рисунок 7.20 – Создание информационной панели мониторинга

3. Для формирования набора информационных панелей мониторинга необходимо создать новую историю. Для этого в нижней части окна нажать кнопку создания историй . В левой части открывшегося окна будут перечислены все сделанные ранее информационные панели. Перетаскив их в область формирования историй, можно сделать историю – набор панелей мониторинга (рис. 7.20).

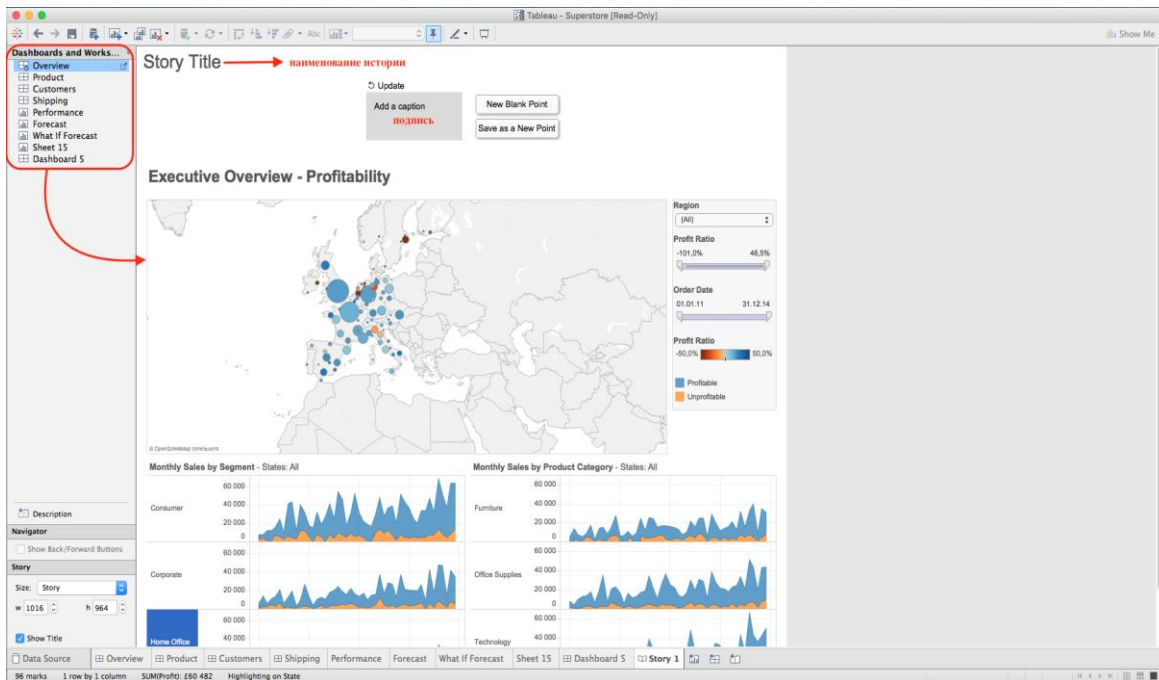


Рисунок 7.20 – Создание истории

4. Для публикации результатов на портале публичных данных, необходимо выбрать пункт меню «Сервер» (Server), Tableau Public, Save to Tableau Public. Далее необходимо будет авторизоваться или зарегистрироваться на портале Tableau Public (рис. 7.21).

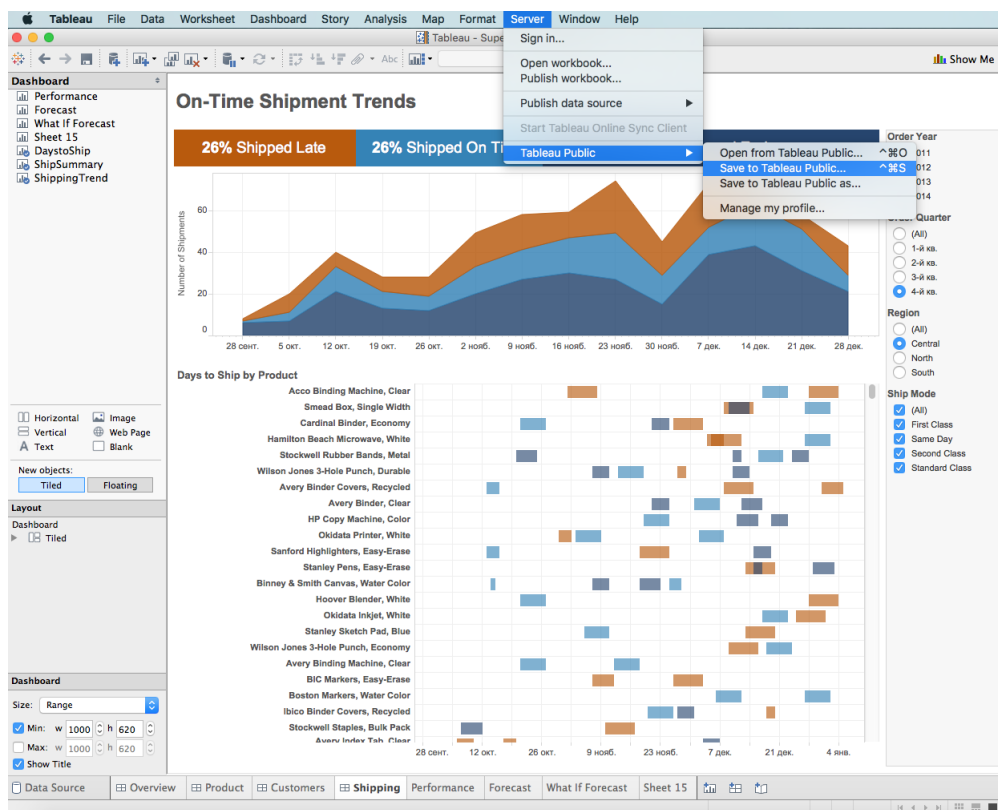


Рисунок 7.21 – Публикация отчетов

7.4. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Чем отличается информационная панель мониторинга, сформированная в Tableau от истории?
2. Каким образом в Tableau можно добавить расчетные показатели?
3. Каким образом в Tableau можно формировать иерархии показателей?
4. Каким образом можно отобразить в аналитическом отчете Tableau дополнительные характеристики?
5. Существует ли возможность редактирования данных в портальной версии <https://public.tableau.com>?

7.5. ЗАДАНИЕ ДЛЯ САМОСТОЯТЕЛЬНОГО ВЫПОЛНЕНИЯ

Для выполнения задания необходимо скачать систему Tableau Desktop по адресу <http://www.tableau.com/products/desktop>. Для получения лицензионного ключа необходимо зарегистрироваться в академической программе Tableau по адресу <http://www.tableau.com/academic>. Также можно воспользоваться системой установленной в аудиториях Финуниверситета.

Для закрепления полученных знаний и навыков в рамках самостоятельной работы каждому студенту необходимо реализовать минимум 5 аналитических отчетов, минимум две информационные панели мониторинга и минимум одну историю на основе хранилища данных созданного ранее (пункты 3-4 данного пособия).

Требования к результатам выполнения задания:

- 1) Описание предметной области.
- 2) Формулировка основной цели и задач рассматриваемой организации.
- 3) Скриншоты отчетов и информационных панелей мониторинга, иллюстрирующих решение сформулированных задач.
- 4) Кроме того, полученные отчеты и панели мониторинга должны быть загружены на портал public.tableau.com и выслана ссылка для проверки проекта.

ФОРМИРОВАНИЕ АНАЛИТИЧЕСКОЙ ОТЧЕТНОСТИ С ПОМОЩЬЮ ГЕНЕРАТОРА ОТЧЕТОВ SAP CRYSTAL REPORTS

8.1. СОЗДАНИЕ ОТЧЕТА В SAP CRYSTAL REPORTS

Crystal Reports содержит средства, позволяющие создавать отчеты пользователям, не являющимся специалистами в области информационных технологий. Наиболее просто создать отчет с помощью мастера отчетов **Report Creation Wizard**. Мастер отчетов содержит несколько окон, в которых пользователю предлагается создать отчет в несколько шагов [7].

1. Выбрать тип отчета.
2. Выбрать источник данных для отчета.
3. Связать таблицы источника данных.
4. Создать поля отчета.
5. Сгруппировать данные.
6. Создать суммирующие поля для каждой группы.
7. Сортировать данные отчета.
8. Создать диаграммы на основании данных.
9. Установить правила выборки данных.
10. Выбрать шаблон форматирования полей отчета.

Следовательно, мастер отчетов позволяет создавать сложные отчеты, которые объединяют данные из разных таблиц баз данных, фильтруют данные по заданным критериям, группируют и сортируют их.

Для создания нового отчета необходимо выполнить команду меню **File | New** или щелкнуть на кнопке создания нового отчета на странице **Start Page**.

Первый шаг – выбор типа нового отчета. Можно выбрать один из четырех типов:

- стандартный – Standard;
- кросс-таблица - Cross-Tab;
- почтовая этикетка – Mail Label;
- OLAP.

После выбора типа создаваемого отчета щелчка на кнопке ОК в диалоговом окне **Crystal Reports Gallery** мастер **Standard Report Creation Wizard** (рис. 8.1).

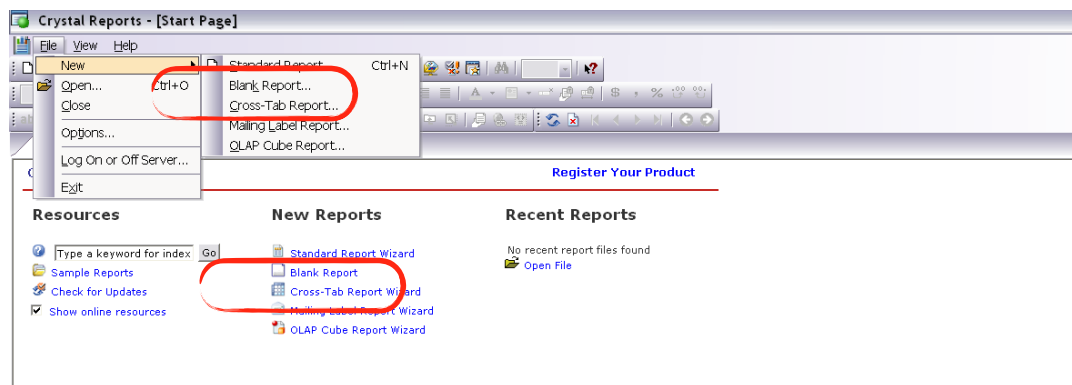


Рисунок 8.1 – Окно Start Page

Окно мастера **Data** содержит два древовидных списка. В левом списке **Available Data Sources** перечислены источники данных, которые могут быть использованы для создания отчета. Доступ к базе данных осуществлен с помощью ODBC или драйверов прямого доступа. Правый список **Selected Tables** содержит выбранные таблицы источников данных (рис. 8.2).

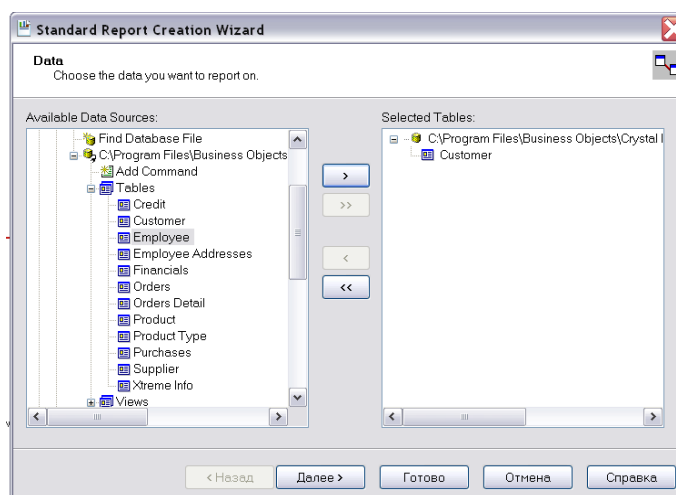


Рисунок 8.2 – Окно мастера Standard Report Creation Wizard (Data)

Следующее окно – окно мастера **Standart Report Creation Wizard (Link)**. Если в качестве источника данных было выбрано более одной таблицы, в окне **Link** показываются таблицы отчета в графическом виде – таблица изображается в виде прямоугольника, внутри которого содержится список полей таблицы. Индексированные поля помечены цветными маркерами. По умолчанию таблицы автоматически связываются по полям с одинаковыми названиями. Связи таблиц можно удалять, создавать и менять их свойства. Подробнее связывание таблиц будет рассмотрено. Заметим, что если в качестве источника данных выбрана настольная база данных (например, формата DBF или, как в рассматриваемом случае, база MS Access), то связь между таблицами возможна только по индексированным полям. Для ODBC-источников данных такого ограничения нет (рис. 8.3).

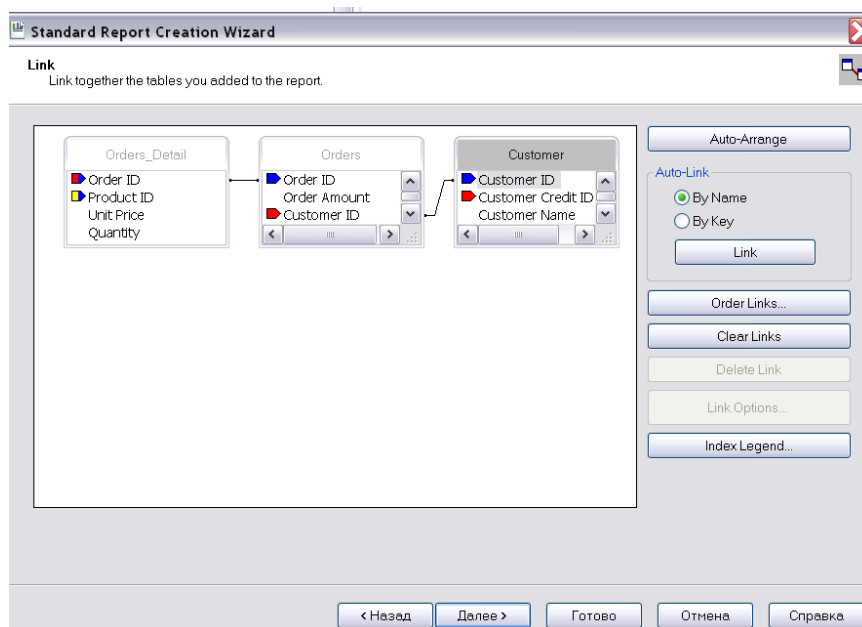


Рисунок 8.3 – Окно мастера Standart Report Creation Wizard (Link)

Третье окно мастера Standard Report Creation Wizard (Fields) содержит два древовидных списка (рис. 8.4). Левый список Available Fields содержит поля таблиц, на основе которых могут быть созданы объекты отчета – поля (Fields) базы данных. Правый список Fields to Display содержит поля, которые будут включены в отчет. Наименование поля в правом списке состоит из двух частей, разделенных точкой. Слева от точки показывается наименование таблицы, справа – наименование поля, например, Customer.Last Year's Sales.

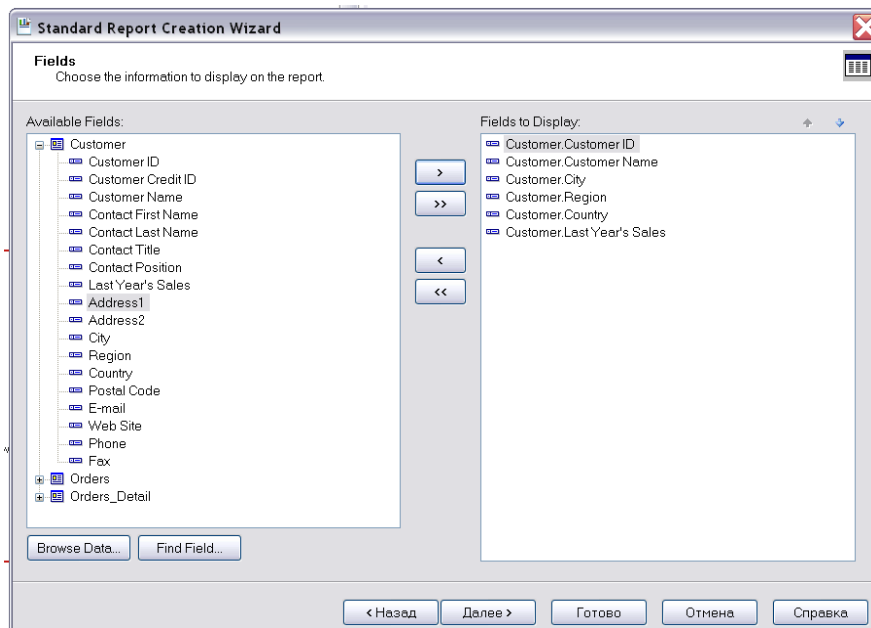


Рисунок 8.4 – Окно мастера Standard Report Creation Wizard (Fields)

Окно Grouping позволяет сгруппировать данные по какому-либо полю, причем, сортировка групп может быть установлена по возрастанию значения поля (числового, строкового или даты), по убыванию или в специальном порядке.

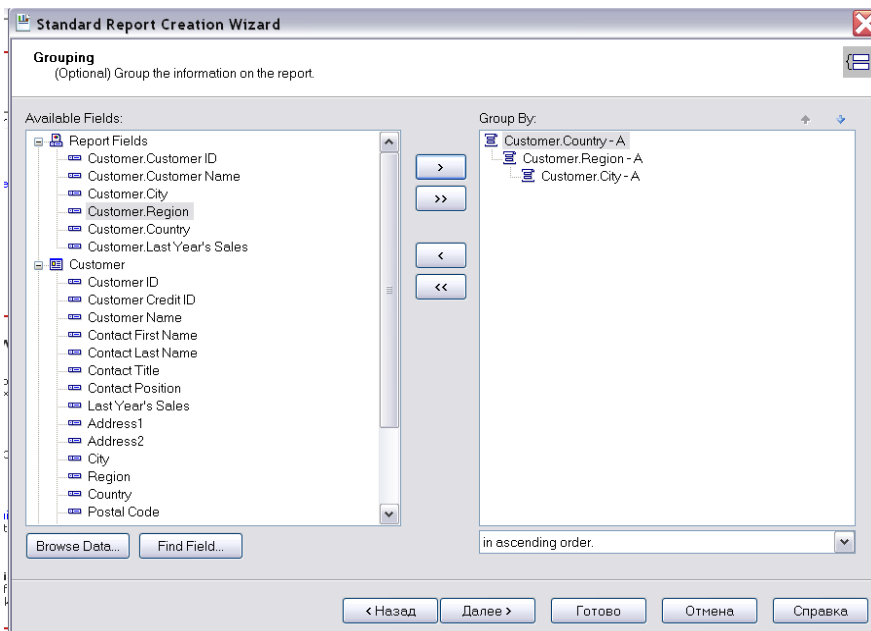


Рисунок 8.5 – Окно Grouping

Левый список окна Grouping – Available Fields содержит поля отчета и поля таблиц, на основе которых могут быть созданы группы отчета (рис. 8.5). Правый список Group By содержит группы отчета. Для создания в отчете новой группы необходимо в левом списке переместить указатель на строку с наименованием поля отчета, по которому будет проведена группировка, и щелкнуть кнопкой (в рассматриваемом примере данные отчета будут сгруппированы сначала по странам, внутри каждой страны по регионам и, наконец, внутри каждого региона по городам).

Можно также вычислять суммарные значения числовых полей в каждой группе. Например, если в отчет включено поле **Customer.Last Year's Sales** (продажи клиенту в прошлом году) и в отчете созданы группы по странам, регионам и городам, можно вычислить объемы продаж в каждом городе, регионе и стране. Вычисляться могут не только сумма, но и среднее значение, количество клиентов и другие агрегатные значения. Для вычисления агрегатных значений служит окно **Summaries**.

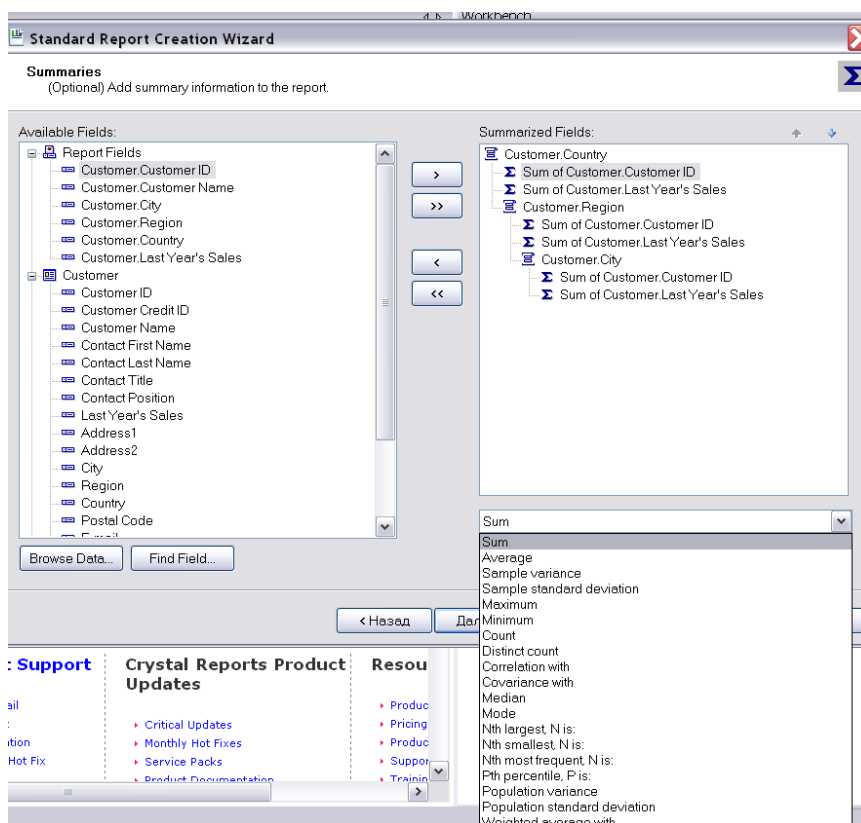


Рисунок 8.6 – Окно Summaries

Левый список окна **Summaries** (рис. 8.6) содержит поля отчета и поля таблиц, на основе которых в отчете может быть проведено агрегирование данных. Правый список **Summarized Fields** содержит древовидный список отобранных для вычисления агрегатных значений полей. В правой нижней части окна расположен раскрывающийся список агрегатных функций (суммирование, вычисление среднего, максимального и минимального значения, подсчет количества записей и т.д.). В примере будет вычисляться количество клиентов в каждой стране и сумма продаж в каждой стране, регионе и городе.

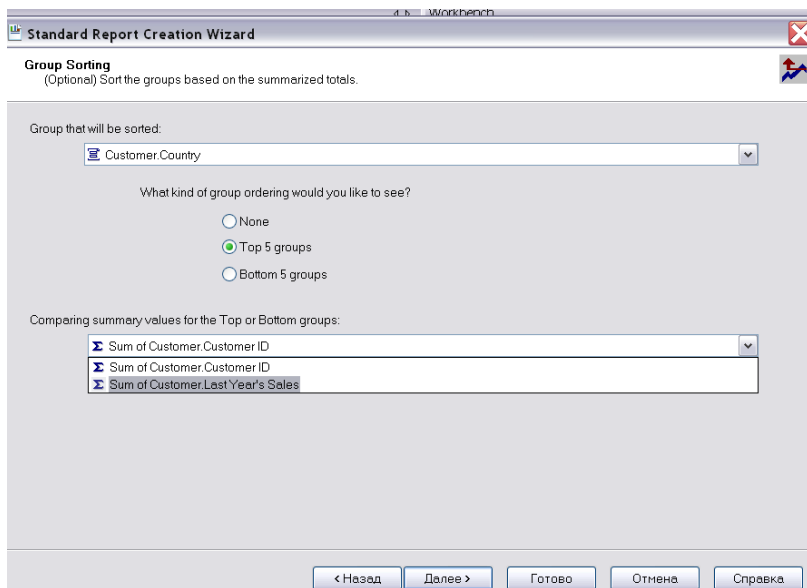


Рисунок 8.7 – Окно Group Sorting мастера Standard Report Creation Wizard

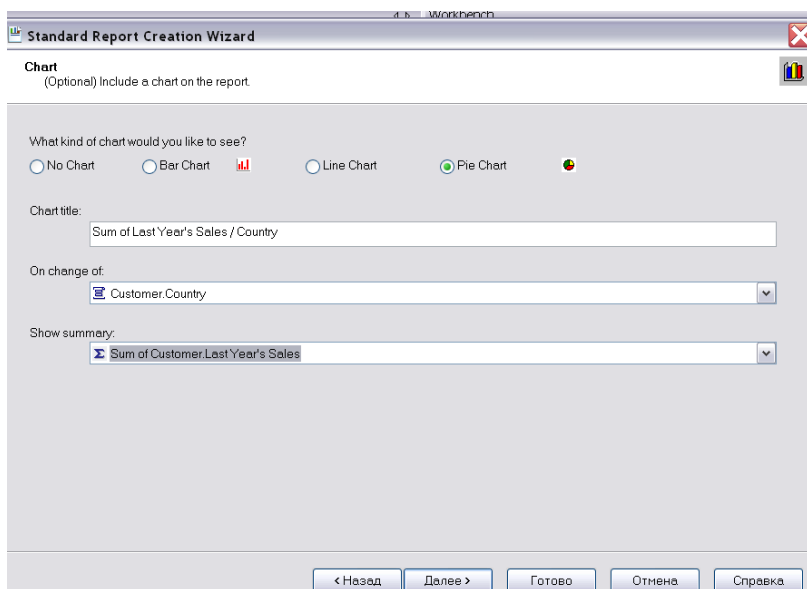


Рисунок 8.8 – Окно Chart мастера Standard Report Creation Wizard

Окно **Group Sorting** мастера **Standard Report Creation Wizard** (рис. 8.7) позволяет более эффективно обрабатывать сгруппированные данные. В верхней части окна расположен раскрывающийся список, содержащий группы, уже созданные в отчете с помощью окна **Grouping**.

Радиокнопки в средней части окна позволяют выбрать тип группировки, а раскрывающийся список в нижней части – выбрать критерий отбора групп. (Так, в примере на выбор специальной группировки **Top 5 groups** и значения суммы **Sum of Customer.Last Year's Sales** в качестве критерия позволит оставить в отчете только те страны, которые занимают пять первых мест по объему продаж).

Окно **Chart** предназначено для включения в отчет диаграмм трех типов – диаграммы в виде столбцов (**Bar**), линий (**Line**) и секторные диаграммы (**Pie**).

Диаграммы могут быть созданы только в отчете, содержащем группировку данных и вычисление агрегированных значений для этих групп.

Окно **Chart** содержит поле **Chart title** для внесения заголовка диаграммы, списки **On change of** и **Show summary**, в которых следует выбрать группу и агрегированные данные для диаграммы.

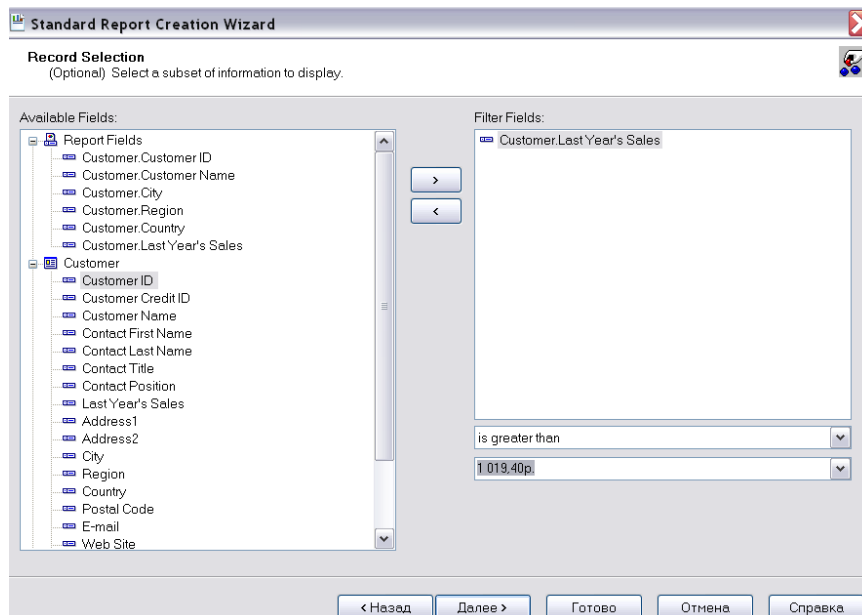


Рисунок 8.9 – Окно Record Selection мастера Standard Report Creation Wizard

Окно **Record Selection** (рис. 8.9) служит для создания правил отбора данных в отчет и содержит два списка. Левый древовидный список **Available Fields** содержит все доступные поля таблиц базы дан-

ных. Правый список **Filter Fields** содержит поля, на основе которых создаются правила отбора данных в отчет.

Последнее окно мастера **Standard Report Creation Wizard - Template** служит для выбора шаблона форматирования отчета. Список в окне **Template** содержит десять шаблонов. В примере выбран шаблон **Confidential Underlay** (рис. 8.10).

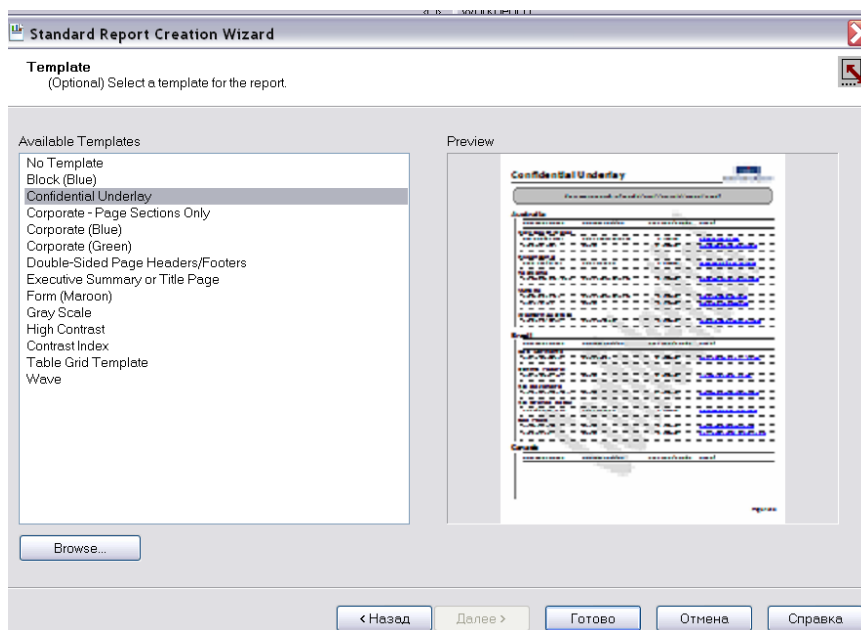


Рисунок 8.10 – Окно Template мастера Standard Report Creation Wizard

После щелчка на кнопке **Finish** (Готово) можно просмотреть созданный отчет на вкладке **Preview** главного окна Crystal Reports (рис. 8.11).

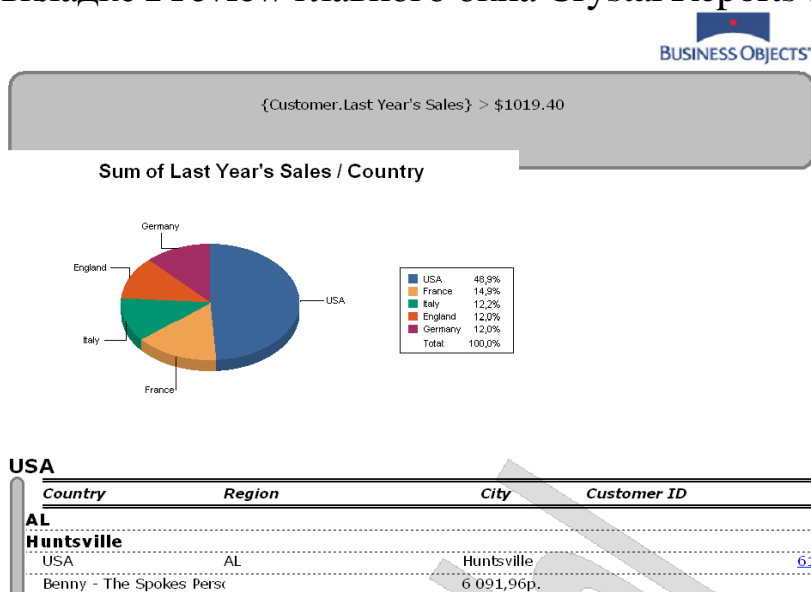


Рисунок 8.11 – Просмотр отчета, созданного с помощью мастера Standard Report Creation Wizard

Для корректировки представления отчета в верхнем меню выбрать пункт «Вид», подпункт «Проект» (рис. 8.12).

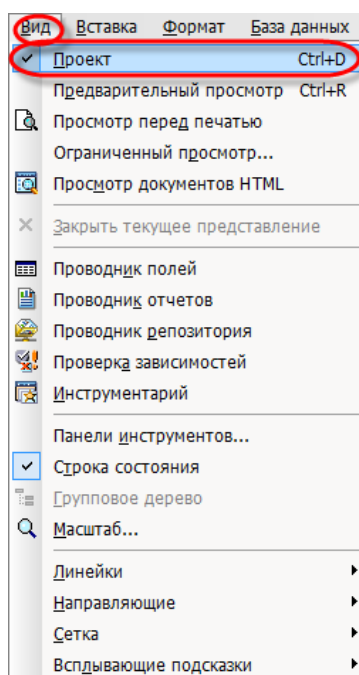


Рисунок 8.12 – Настройка отчета вручную

Отредактировать структуру отчета. Для этого необходимо переключиться на закладку «Проект» (рис. 8.13).

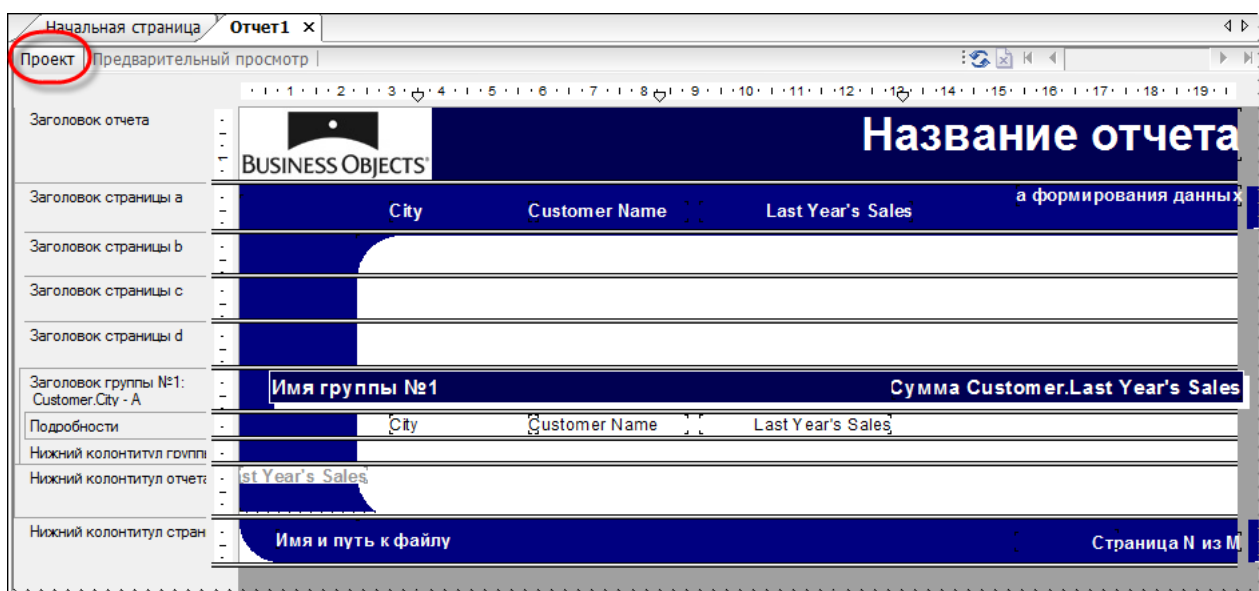


Рисунок 8.13 – Настройка отчета вручную

Для просмотра отчета, необходимо переключиться на вкладку «Предварительный отчет» (рис. 8.14).

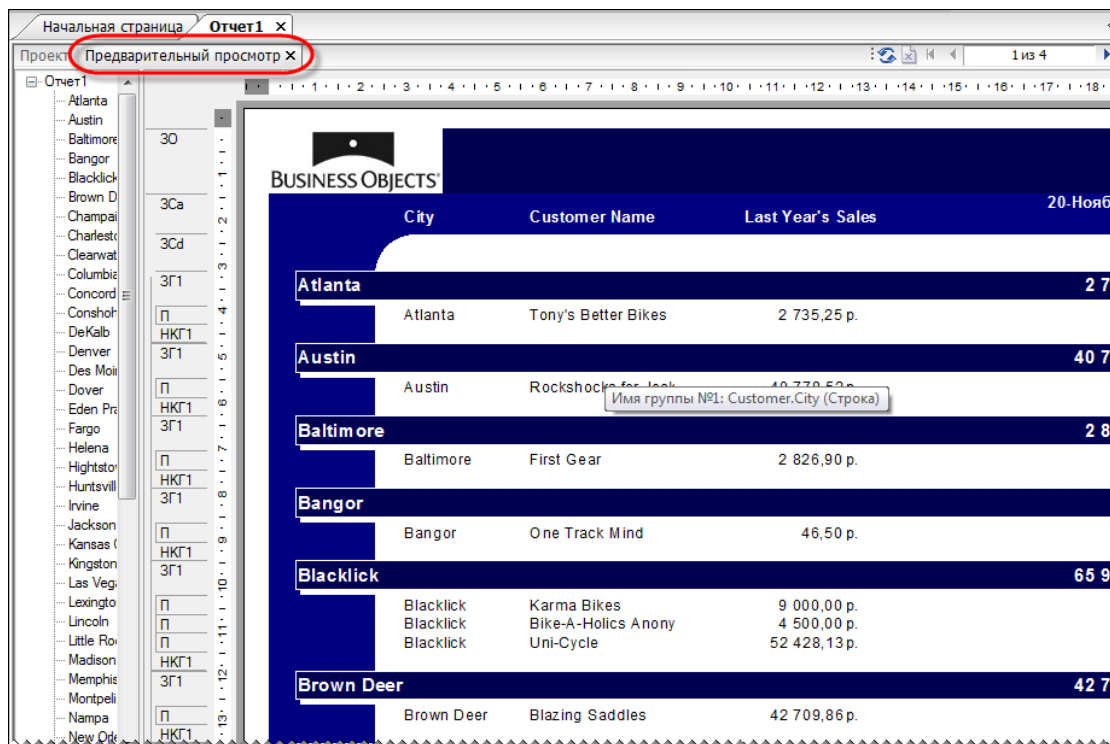


Рисунок 8.14 – Настройка отчета вручную

После начала создания отчета на вкладке «Проект» по умолчанию формируется пять секций.

Таблица 8.1 – Секции закладки «Проект»

Секция	Место вставки	Состав секции
Report Header (заголовок отчета)	Один раз в начале отчета	Титульная страница, логотип, вводная информация, сведения, которые нужно вставить один раз в начале отчета
Page Header (заголовок страницы)	Вверху каждой страницы	Заголовки полей, дата/время создания или печати отчета, заголовок отчета
Details (подробности)	При каждом чтении/новой записи из базы данных	Поля базы данных, или формулы, которые выводятся для каждой записи
Report Footer (колонтитул отчета)	Один раз в конце отчета	Общие итоги, другие сведения, применимые к отчету в целом
Page Footer (колонтитул страницы)	Внизу каждой страницы	Номер страницы, имя отчета

Диаграммы и объекты **Cross-Tab** применяются к отчету в целом, и могут быть вставлены либо в заголовок, либо в колонтитул отчета.

Высоту каждой секции можно менять, для этого необходимо подвести указатель мыши к краю секции так, чтобы указатель приобрел вид двунаправленной стрелки. После этого нужно нажать левую кнопку мыши, и изменить высоту секции.

Если какая-либо из созданных по умолчанию секций не нужна в отчете, ее можно скрыть. Для этого нужно щелкнуть правой кнопкой мыши в области заголовка секции (слева на экране) и в раскрывшемся контекстном меню выбрать команду «Скрыть». В окне конструктора выбранная секция после этого будет отображаться покрытой косой штриховкой, а на печать выводиться не будет. Снять скрытие секции можно также из контекстного меню.

Секция **Details** в дизайнера отчета – самая узкая, но именно в ней, как правило, отображается основная информация из полей базы данных.

Для этого необходимо в окне **Field Explorer**, на вкладке **Database** выделить требуемое поле, и не отпуская левую кнопку мыши, перетащить его в секцию **Details** конструктора, разместить его там, и после этого отпустить левую кнопку мыши. По умолчанию **Crystal Reports** для помещенного в секцию **Details** поля автоматически в области **Page Heading** создает заголовок, который, опять же по умолчанию, совпадает с именем поля. Если поле поместить в любую другую секцию, заголовок не создается. Если потом перетащить это поле в секцию **Details**, заголовок все равно уже создан не будет (рис. 8.15).

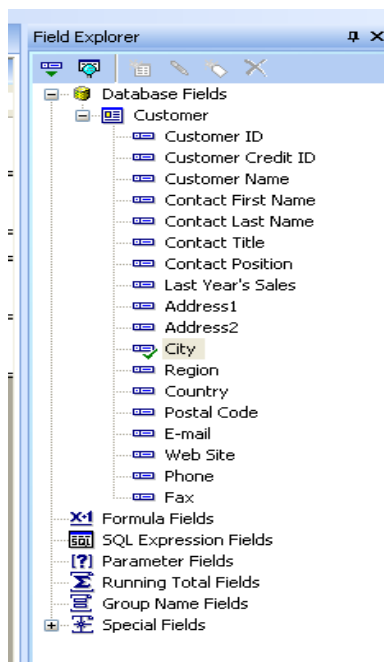


Рисунок 8.15 – Окно Field Explorer

Кроме полей базы данных в отчет практически всегда нужно добавить текстовую информацию: заголовки, комментарии, примечания. Для вставки текстового поля в отчет необходимо в меню **Insert** выбрать пункт **Text Object** (рис. 8.16).

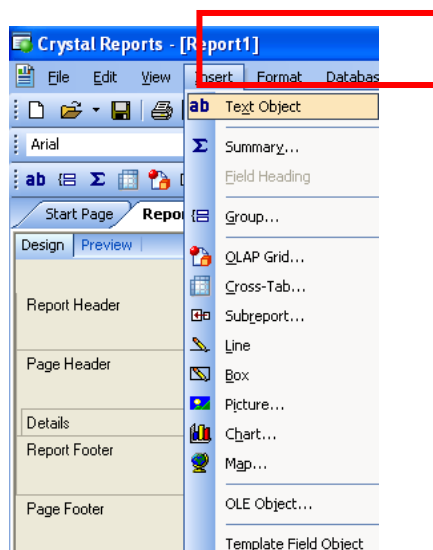


Рисунок 8.16 – Вставка текстового поля с помощью меню

Помимо текстовых полей, в отчет могут быть включены специальные поля содержащие дополнительную информацию, такую как номер страницы номер записи, дата отчета и т.д. Для вставки специального поля необходимо в списке полей диалогового окна **Field Explorer** выбрать раздел **Special Field** (рис. 8.17).

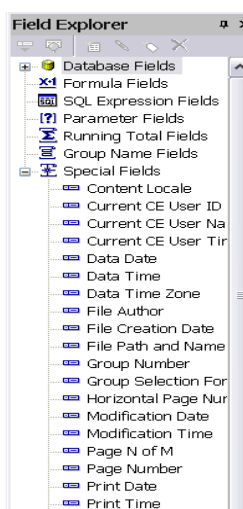


Рисунок 8.17 – Раздел Special Field

Группа полей **Special Field** содержит следующие специальные поля.

- Page Number – номер страницы.
- Total Page Count – общее количество страниц в отчете.
- Page N of M – показывает номер текущей страницы из общего количества страниц в отчете.
- Report Title – заголовок отчета
- Report Comments - комментарий к отчету.
- File Path and Name – имя файла отчета и путь к нему.
- File Author – имя автора отчета.
- File Creation Date – дата создания шаблона отчета.
- Print Date – дата печати отчета.
- Print Time – время печати отчета.
- Data Date – дата последнего обновления данных.
- Data Time – время последнего обновления данных.
- Modification Date – дата последнего изменения шаблона отчета.
- Modification Time – время последнего изменения шаблона отчета.
- Record Number – номер записи отчета.
- Group Number – номер группы отчета.
- Record Selection Formula – поле отображает в режиме просмотра формулу, использующуюся для отбора записей.
- Group Selection Formula – поле отображает в режиме просмотра формулу, использующуюся для отбора записей (фильтрация по агрегатным значениям).

8.2. ОСНОВЫ РАСЧЕТОВ В SAP CRYSTAL REPORTS

Если в отчет необходимо вставить расчетное поле, необходимо создать формулу. Кроме вычислений, формулы **Crystal Reports** имеют еще одно важное назначение – с их помощью можно производить форматирование в зависимости от контекста, например, за счет раскрашивания данных в разные цвета.

Окно **Fields Explorer** имеет пункт **Formula Fields** (рис. 8.18).

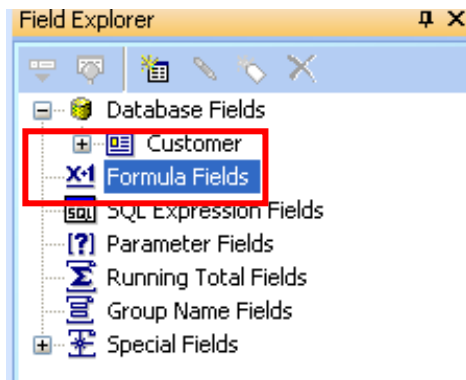


Рисунок 8.18 – Пункт Formula Fields

Пока в отчете нет ни одной формулы, вкладка пуста, для создания новой формулы следует нажать кнопку **New** (рис. 8.19). Откроется окно для ввода названия формулы:

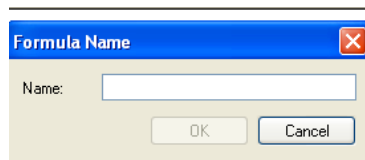


Рисунок 8.19 – Создание новой формулы

Появляется следующее окно, в котором можно вписать формулу вручную:

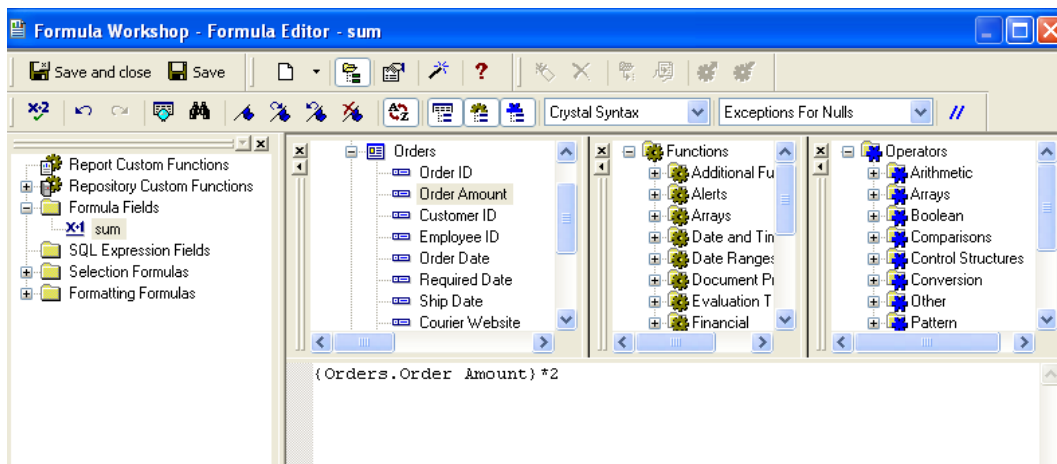


Рисунок 8.20 – Окно редактора формул

Чтобы сохранить формулу, следует закрыть редактор формул. При закрытии система предложит сохранить формулу (рис. 8.20).

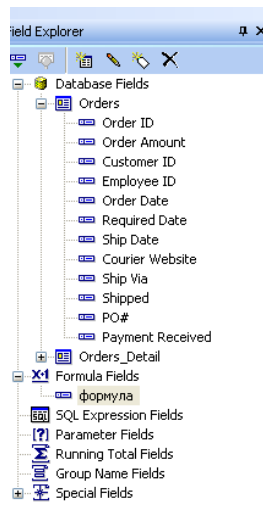


Рисунок 8.21 – Просмотр формулы

Теперь поле с формулой можно вставить в отчет методом перетаскивания.

Синтаксис формул:

1. Текст. При использовании текста в формулах он должен быть заключен в двойные или одинарные кавычки. В одной формуле можно одновременно использовать оба вида кавычек, но нужно соблюдать парность: Пример: “Текст”, ‘Текст’

2. Числа. Числа должны содержать в качестве десятичного разделителя точку. Не допускаются пробелы внутри числа, разделители тысяч и знаки валюты.

3. Поля данных. При использовании в формуле поле должно быть заключено в фигурные скобки {}, с указанием имени таблицы и, через точку, имени поля: Пример: {Orders.Costs}

4. Поля формул и параметрические поля. Имена этих полей заключаются в фигурные скобки. Перед именем поля формулы автоматически устанавливается знак @, перед именем параметрического поля – знак ?. Пример: {@SumAccount}, {?DocDate}

5. Операции. При использовании операций в формулах в нужном месте просто вводится знак операции. До и после знака операции можно вводить пробел, но он не обязателен.

6. Функции. При использовании функций вводится имя функции и необходимые параметры в круглых скобках. При вводе параметров необходимо соблюдать синтаксис в зависимости от типа параметра: поля заключать в фигурные скобки, текст – в одинарные или двойные кавычки и т. д. Пример: Sum({Pays.Qv})

7. Скобки.

{ } – в фигурные скобки заключаются имена полей;

[] – в квадратные скобки заключаются индексы массивов;

() – в круглые скобки заключаются параметры, а также они позволяют группировать операции и определять порядок их выполнения.

8. Переменные. В Crystal Reports переменные определяются с помощью ключевого слова, которое начинается с названия типа данных и суффикса Var.

Пример:

NumberVar x ; – переменная x числового типа

CurrencyVar CCost ; – переменная CCost денежного типа

Значения присваиваются переменным с помощью оператора присваивания «:=»:

X := 24 ;

Присвоить значение переменной можно непосредственно при объявлении:

StringVar LName := "Иванов" ;

Объявление переменной должно заканчиваться точкой с запятой. В других случаях точка с запятой используется для разделения нескольких операторов, если они написаны в одну строку.

9. Комментарии. Комментарии в формулах начинаются со знака «//» (двойной слэш) и действуют до конца строки. В одной строке допускается совместное написание выражения и комментария, но комментарий должен быть справа.

8.3. ПРАКТИКУМ ПО ФОРМИРОВАНИЮ ШАБЛОНА ОТЧЕТА В SAP CRYSTAL REPORTS С ПОМОЩЬЮ МАСТЕРА СТАНДАРТНЫХ ОТЧЕТОВ

Цель практикума

Закрепить теоретические знания и сформировать практические навыки анализа данных с помощью генератора отчетов SAP BO Crystal Reports.

План практикума

1. Изучение описания бизнес-кейса.
2. Подключение источника данных.
3. Формирование шаблона отчета.

Порядок выполнения практикума

1. Изучите описание бизнес-кейса.

БИЗНЕС-КЕЙС

Торговая компания, занимающаяся продажами техники по всему миру. Имеются следующие данные:

- номер заказа;
 - дата заказа;
 - дата и условия доставки;
 - личные данные о покупателе;
 - данные о месте продажи: регионы, города.
 - данные о продуктах: номер, наименование товара;
 - данные о продажах: выручка, количество проданных товаров.
- Необходимо проанализировать следующую информацию:
- необходимо проанализировать последние покупки клиентов проживающих в США.

2. Для создания отчета выполнить команду меню Файл – Создать – Стандартный отчет. Либо выбрать «Мастер стандартных отчетов» (рис. 8.22).

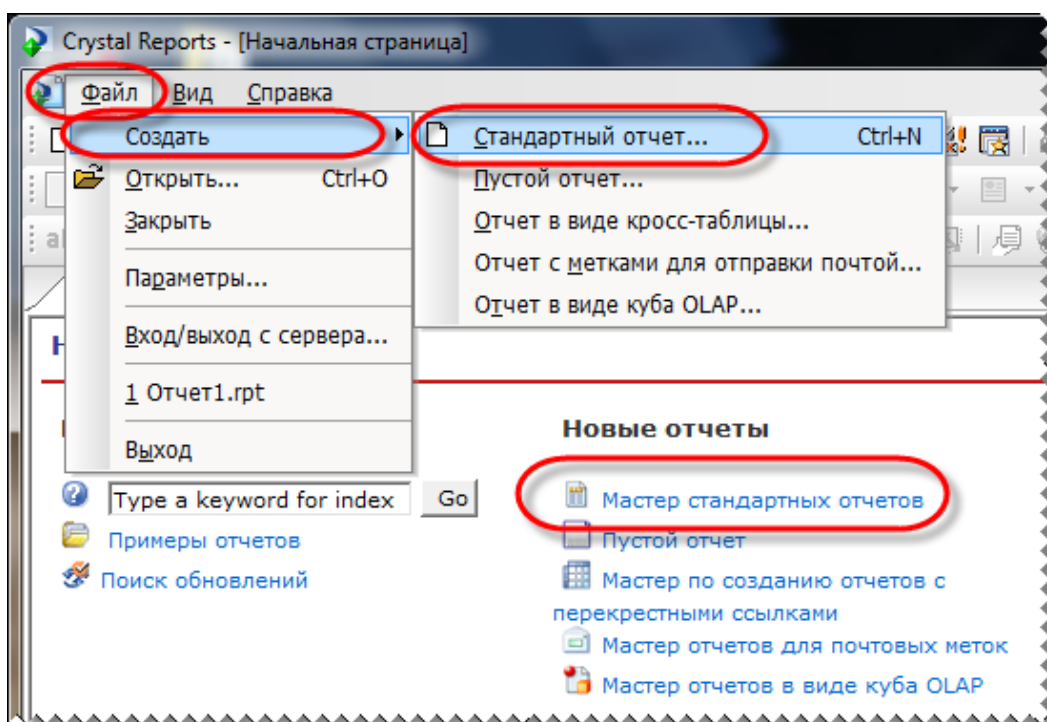


Рисунок 8.22 – Создание отчета

В окне «Данные» мастера создания отчетов в качестве источника данных выбрать Создать новое соединение – Access/Excel. Подключить файл xtreme (рис. 8.23).

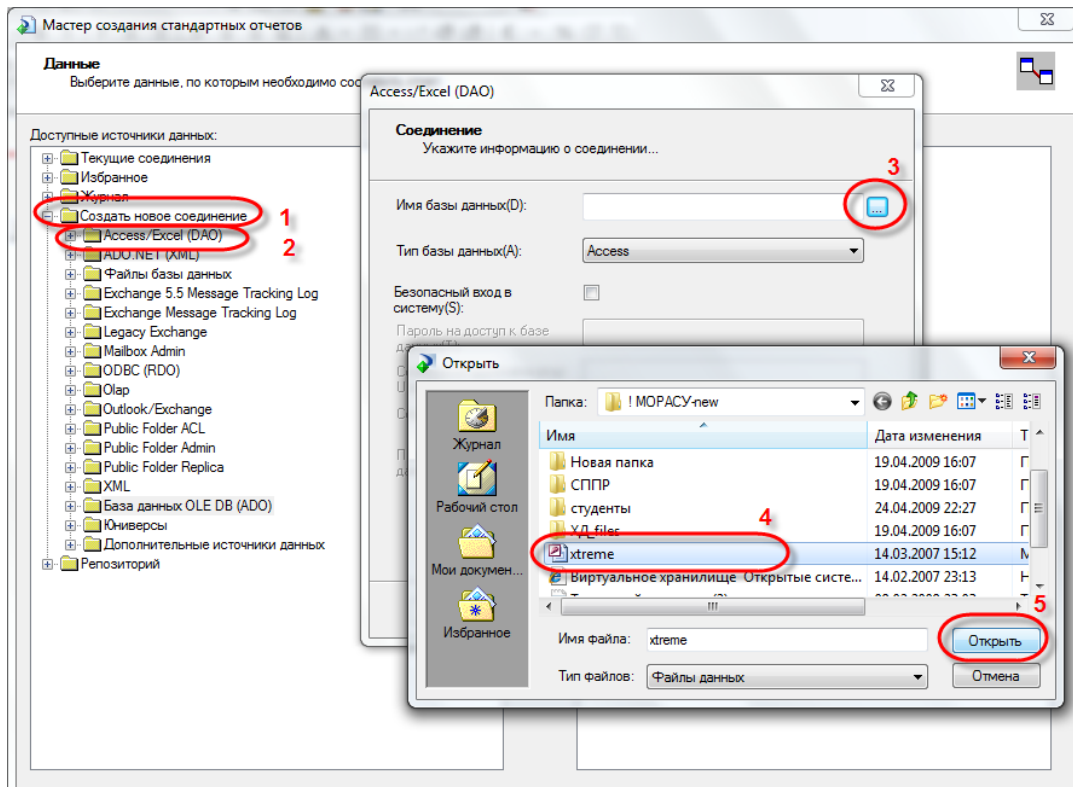


Рисунок 8.23 – Создание соединения с источником

После включения базы данных xtreme в список источников выберите в разделе Tables таблицу Customer и щелкните на кнопке >. Таблица Customer включается в правый список Selected Tables. Щелкните на кнопке «Далее» (рис. 8.24).

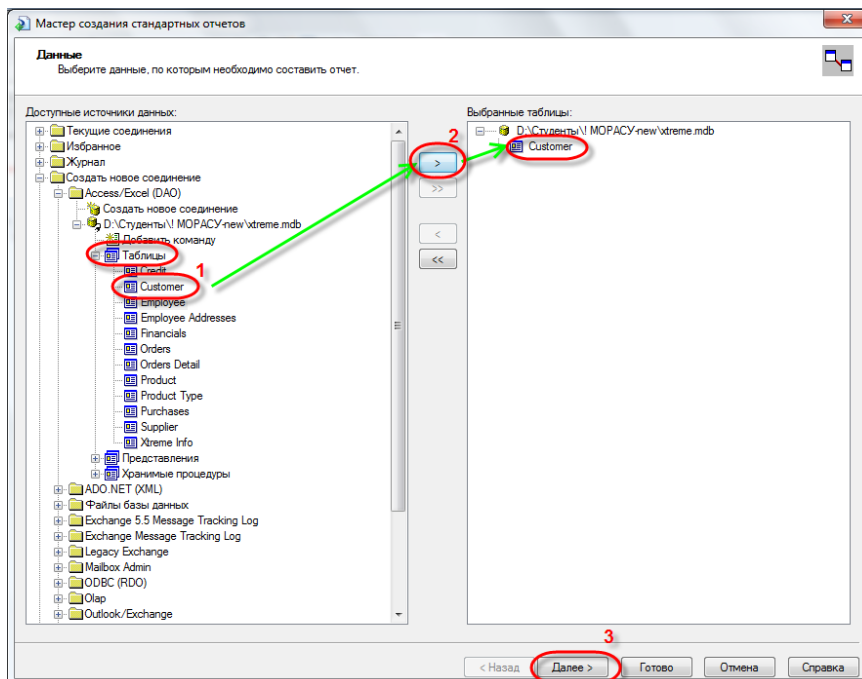


Рисунок 8.24 – Выбор данные для отчета

В окне «Поля» включите в отчет поля Customer Name и Last Year's Sales. Щелкните на кнопке «Далее» (рис. 8.25).

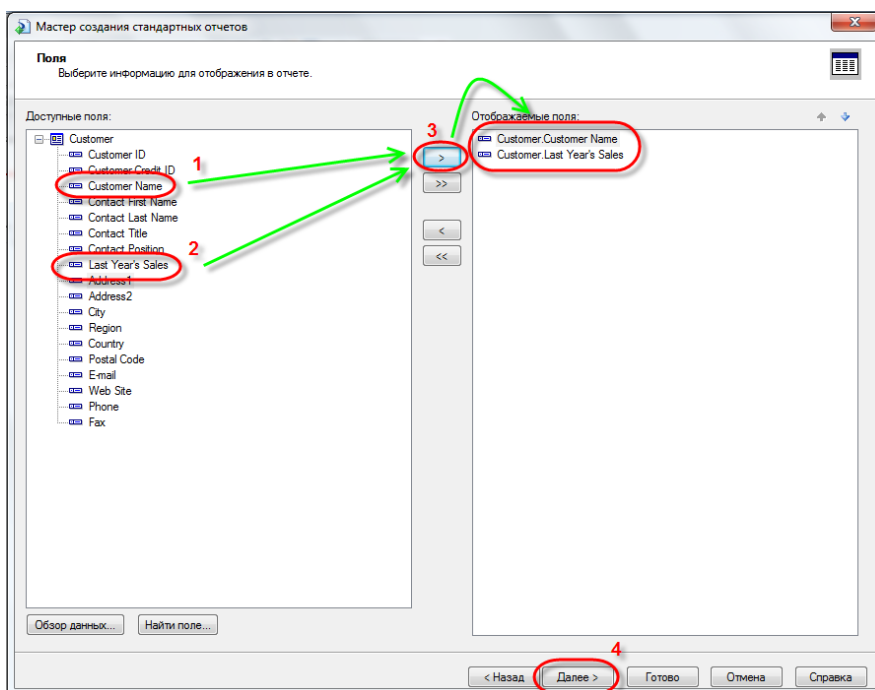


Рисунок 8.25 – Выбрать поля для отображения

В окне «Группировка» включите в отчет группировку по полю City (рис. 8.26).

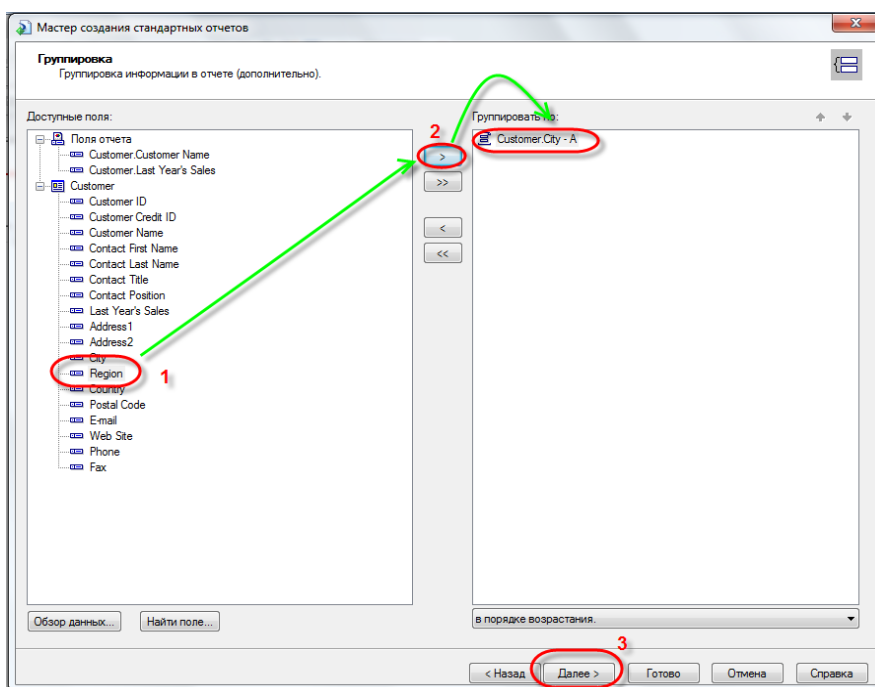


Рисунок 8.26 – Настройка группировки

Щелкните несколько раз на кнопке «Далее» до тех пор, пока не появится окно «Выбор записей». В окне «Выбор записей» задайте условие отбора Country is equal to USA (рис. 8.27).

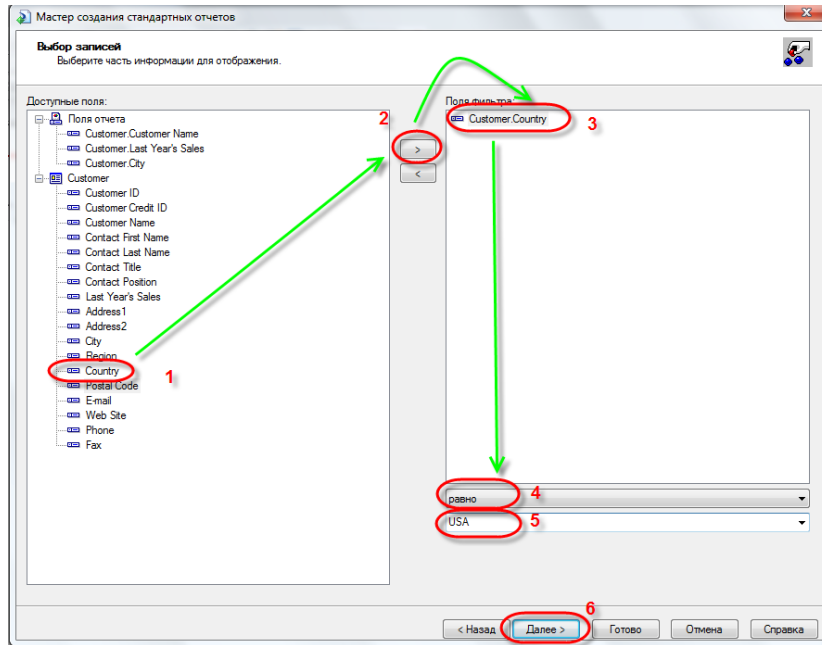


Рисунок 8.27 – Настройка фильтров

Щелкните на кнопке «Готово».
Выберите шаблон (рис. 8.28).

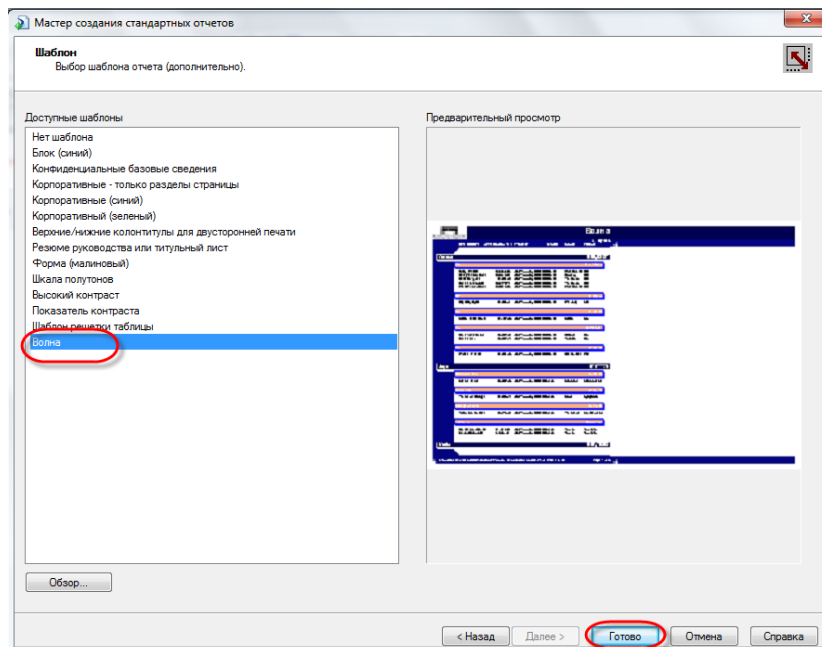


Рисунок 8.28 – Выбор шаблона отчета

Сохраните отчет.

8.4. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. В какую секцию отчета необходимо добавить поля для их повторения в виде шапки отчета на каждой странице?
2. Каким образом в SAP Crystal Reports можно создавать вычисляемые поля?
3. Каким образом в SAP Crystal Reports можно добавить дополнительные поля?
4. Каким образом в SAP Crystal Reports можно добавить свободный текст?
5. Каким образом можно предварительно просмотреть формируемый вручную отчет?

8.5. ЗАДАНИЕ ДЛЯ САМОСТОЯТЕЛЬНОГО ВЫПОЛНЕНИЯ

Для выполнения заданий можно воспользоваться системой установленной в аудиториях Финуниверситета.

Для закрепления полученных знаний и навыков в рамках самостоятельной работы каждому студенту необходимо реализовать минимум 3 аналитических отчета на основе хранилища данных созданного ранее (пункты 3-4 данного пособия).

Требования к результатам выполнения задания:

- 1) Описание предметной области.
- 2) Формулировка основной цели и задач рассматриваемой организации.
- 3) Скриншоты отчетов, иллюстрирующих решение сформулированных задач.
- 4) Кроме того, полученные отчеты должны быть сданы преподавателю в виде сохраненного проекта SAP Crystal Reports.

ТЕСТЫ ДЛЯ САМОПОДГОТОВКИ

- 1.** Таблица фактов содержит
 - a) множество метаданных модели
 - b) данные, являющиеся предметом анализа
 - c) схему «звезда»
 - d) список трансформаций данных, необходимых для процесса ETL
 - e) категории, по которым производится анализ данных
- 2.** Таблица измерений содержит
 - a) множество метаданных модели
 - b) данные, являющиеся предметом анализа
 - c) схему «снежинка»
 - d) категории, по которым можно анализировать данные
 - e) список трансформаций данных, необходимых для процесса ETL
- 3.** Консольная таблица характеризуется следующими утверждениями
 - a) Является обязательной частью схемы «звезда»
 - b) Содержит данные, являющиеся предметом анализа
 - c) Является частью схемы «снежинка»
 - d) Повышает производительность аналитических запросов
 - e) Не может быть связана непосредственно с таблицей фактов
 - f) Увеличивает степень нормализации модели данных
- 4.** Требования к транзакционным системам включают
 - a) Быструю вставку и обновление
 - b) Быстрое объединение и выборки (joins and selects)
 - c) Прямой доступ к текущей информации
 - d) Максимум ссылочной целостности
 - e) Доступ к агрегированной исторической информации
- 5.** Требования к аналитическим системам включают
 - a) Быструю вставку и обновление
 - b) Быстрое объединение и выборки (joins and selects)
 - c) Прямой доступ к текущей информации
 - d) Максимум ссылочной целостности
 - e) Доступ к агрегированной исторической информации

- 6.** Схема «звезда» характеризуется следующими утверждениями:
- a) Включает одну таблицу фактов
 - b) Может включать несколько таблиц фактов
 - c) Включает только одну таблицу измерений
 - d) Может включать несколько таблиц измерений
 - e) Все сведения об иерархиях измерения хранятся в одной таблице измерения
 - f) Сведения об иерархиях измерения хранятся в нескольких связанных таблицах
- 7.** Схема «снежинка» характеризуется следующими утверждениями:
- a) Включает одну таблицу фактов
 - b) Может включать несколько таблиц фактов
 - c) Включает только одну таблицу измерений
 - d) Может включать несколько таблиц измерений
 - e) Все сведения об иерархиях измерения хранятся в одной таблице измерения
 - f) Сведения об иерархиях измерения хранятся в нескольких связанных таблицах
- 8.** Для таблиц, связанных идентифицирующей связью в модели ERwin Data Modeler, справедливы утверждения:
- a) Первичный ключ родительской таблицы не мигрирует по связи в дочернюю таблицу;
 - b) Первичный ключ родительской таблицы автоматически мигрирует в дочернюю таблицу в область ключевых колонок;
 - c) Первичный ключ родительской таблицы автоматически мигрирует в дочернюю таблицу в область атрибутов;
 - d) Идентификация объектов дочерней таблицы невозможна без объекта родительской таблицы.
- 9.** Для таблиц, связанных неидентифицирующей связью в модели ERwin Data Modeler, справедливы утверждения:
- a) Первичный ключ родительской таблицы не мигрирует по связи в дочернюю таблицу;
 - b) Первичный ключ родительской таблицы автоматически мигрирует в дочернюю таблицу в область ключевых колонок;
 - c) Первичный ключ родительской таблицы автоматически мигрирует в дочернюю таблицу в область атрибутов;
 - d) Идентификация объектов дочерней таблицы невозможна без объекта родительской таблицы.

- 10.** Для связи между таблицей-измерением и таблицей-фактов справедлив утверждения:
- a) Таблицы должны быть связаны идентифицирующей связью;
 - b) Таблицы могут быть связаны связью многие-ко-многим;
 - c) Таблицы могут быть связаны неидентифицирующей связью;
 - d) Таблица-измерение должна находиться на родительском конце связи;
- 11.** Для связи между таблицей-измерением и консольной таблицей справедливы утверждения:
- a) Таблицы должны быть связаны идентифицирующей связью;
 - b) Таблицы могут быть связаны связью многие-ко-многим;
 - c) Таблицы могут быть связаны неидентифицирующей связью;
 - d) Таблица-измерение должна находиться на родительском конце связи;
- 12.** Для прямой генерации каталога витрины данных в СА ERwin Data Modeler требуется:
- a) Переключиться на логический уровень модели данных;
 - b) Переключиться на физический уровень модели данных;
 - c) Выполнить подключение к базе данных;
 - d) Заменить все неидентифицирующие связи идентифицирующими.
- 13.** Поддержка модели хранилища данных в актуальном состоянии в СА ERwin Data Modeler осуществляется с помощью инструмента:
- a) Forward Engineering;
 - b) Reverse Engineering;
 - c) Complete Compare;
 - d) Report Designer.
- 14.** Компонент MS SQL Server, который используется для операций ETL:
- a) Database Engine;
 - b) Analysis Services;
 - c) Integration Services;
 - d) Reporting Service;
- 15.** Инструмент MS SQL Server, который используется для создания новых OLAP-кубов:
- a) SS Data Tools;
 - b) SS Management Studio;
 - c) Report Designer;
 - d) Database Engine.

- 16.** Использование надстроек Excel для анализа таблиц требует:
- a) Использование данных из хранилища данных;
 - b) Использование данных на страницах Excel, форматированных произвольным способом;
 - c) Использование данных на страницах Excel, форматированных как таблица;
 - d) Подключения к аналитическим службам MS SQL Server;
 - e) Автономной работы Excel.
- 17.** Под оперативным анализом данных подразумеваются следующие технологии:
- a) OLAP;
 - b) Data Mining;
 - c) Имитационное моделирование.
- 18.** Под интеллектуальным анализом данных подразумеваются следующие технологии:
- a) OLAP;
 - b) Data Mining;
 - c) Имитационное моделирование.
- 19.** Особенностью систем Data Discovery являются:
- a) Использование технологий in-memory;
 - b) Простота формирования отчетов;
 - c) Расширенные возможности использования методов глубокой аналитики;
 - d) Возможности формирования регламентных отчетов.
- 20.** Для того, чтобы выбрать определенные данные в Power BI необходимо:
- a) Написать SQL-запрос для выделения нужных данных;
 - b) Интерактивно перенести поля для фильтрации в область «Фильтры»;
 - c) В комбобоксе области «Фильтры» выбрать поля и прописать условия для отбора данных;
 - d) Настроить фильтры в системе невозможно.
- 21.** Для того, чтобы выбрать определенные данные в SAP Crystal Reports необходимо:
- a) Написать SQL-запрос для выделения нужных данных;
 - b) Настроить фильтры при формировании отчета в окне «Фильтры»;

- c) Интерактивно перенести поля для фильтрации в область «Фильтры»;
- d) Настроить фильтры в области «Проекты».
- e) Настроить фильтры в системе невозможно.

22. Для публикации отчетов, сформированных в системе Power BI необходимо:

- a) Зарегистрироваться на powerbi.com и загрузить отчет через портал;
- b) Выбрать пункт меню: файл-сохранить как-сохранить как публикацию.
- c) Выбрать пункт меню: файл-опубликовать в интернете-создать код публикации.

23. В системе Power BI существуют следующие возможности управления данными:

- a) Можно формировать хранилища данных по схеме «звезда».
- b) Можно переименовывать поля и менять их формат.
- c) Можно изменять структуру источника данных.
- d) Можно менять наименование и формат поля, разделять одно поле на несколько.
- e) Нет возможности настраивать источники данных.

24. В системе Tableau существуют следующие возможности управления данными:

- a) Можно формировать хранилища данных по схеме «звезда».
- b) Можно переименовывать поля и менять их формат.
- c) Можно изменять структуру источника данных.
- d) Можно менять наименование и формат поля, разделять одно поле на несколько.
- e) Нет возможности настраивать источники данных.

25. В системе Tableau существуют следующие возможности управления данными:

- a) Можно загружать несколько источников, но использовать для формирования отчетов только один.
- b) Можно загружать и одновременно использовать несколько источников.
- c) Можно загружать и использовать только один источник данных.

- 26.** В системе Power BI существуют следующие возможности управления данными:
- a) Можно загружать несколько источников, но использовать для формирования отчетов только один.
 - b) Можно загружать и одновременно использовать несколько источников.
 - c) Можно загружать и использовать только один источник данных.
- 27.** В системе SAP Crystal Reports существуют следующие возможности управления данными:
- a) Можно загружать несколько источников, но использовать для формирования отчетов только один.
 - b) Можно загружать и одновременно использовать несколько источников.
 - c) Можно загружать и использовать только один источник данных.
- 28.** Для формирования наименования отчета в SAP Crystal Reports необходимо вставить данные в блок:
- a) Report Footer
 - b) Page Footer
 - c) Page Header
 - d) Report Header
 - e) Details
- 29.** Для добавления всех данных из определенного поля базы данных в отчет необходимо переместить поле в блок:
- a) Report Footer
 - b) Page Footer
 - c) Page Header
 - d) Report Header
 - e) Details
- 30.** Для добавления нумерации страниц в отчет необходимо добавить данные в блок:
- a) Report Footer
 - b) Page Footer
 - c) Page Header
 - d) Report Header
 - e) Details

КЛЮЧИ К ТЕСТАМ

Таблица 9.1 – Правильные ответы к заданиям теста

<i>№ теста</i>	1	2	3	4	5	6	7	8	9	10
<i>Правильный ответ</i>	b	d	c, e, f	a, c, d	b, e	a, d, e	a, d, f	b, d	c	a, d

<i>№ теста</i>	11	12	13	14	15	16	17	18	19	20
<i>Правильный ответ</i>	c	b, c	c	c	a	c, d	a	b	a, b	b, c

<i>№ теста</i>	21	22	23	24	25	26	27	28	29	30
<i>Правильный ответ</i>	b, d	c	e	d	b	b	a	d	e	b

СЛОВАРЬ

BAВOK. Business Analysis Body of Knowledge – профессиональный стандарт/свод знаний по бизнес-анализу, разрабатываемая международным институтом бизнес-анализа (International Institute of Business Analysis, ИВА).

Data Discovery. Альтернативный подход к созданию аналитических решений по сравнению с традиционными промышленными системами бизнес-анализа. Решения типа data discovery предлагают пользователям интерактивный графический пользовательский интерфейс, базирующийся на архитектуре in-memory, что соответствует запросу бизнеса на простые и быстрые в работе BI-системы.

ETL. Extract, Transform, Load («извлечение, преобразование, загрузка»). Процесс управления данными, который включает в себя:

- извлечение данных из внешних источников;
- трансформацию и очистка извлеченных данных, чтобы они соответствовали потребностям бизнес-модели;
- загрузку преобразованных данных в хранилище данных.

HOLAP. Hybrid OLAP. Технологии OLAP, основанные на использовании для поддержки многомерного представления данных средств как реляционных, так и многомерных баз данных.

MOLAP. Multi-Level OLAP. Технологии OLAP, где источниками данных служат многомерные базы данных.

OLAP. On-Line Analytical Processing. Технологии интерактивной аналитической обработки данных для поддержки принятия решений, ориентированные главным образом на нерегламентированные интерактивные запросы. OLAP имеет дело, как правило, с историческими данными и часто используется в хранилищах данных.

OLTP-система. OLTP – On-line Transaction Processing System – система обработки транзакций в реальном времени) осуществляет учет и хранение первичной информации о работе организации, наце-

лена на поддержание ее повседневной деятельности. Классическим примером OLTP-системы является «1С-Бухгалтерия». В каждой организации одновременно действуют несколько (иногда несколько десятков) OLTP-систем.

ROLAP. Relational OLAP. Технологии OLAP, использующие представление куба данных, которое поддерживается над реляционной базой данных.

Агрегат (Aggregate). Скалярное значение, вычисляемое по некоторому множеству детализированных записей.

Агрегирование данных (Data Aggregation). Операция над кубом данных, которая строит куб меньшей размерности путем замены совокупностей значений элементов одного или нескольких измерений значениями элементов более высоких уровней и формирования в качестве содержимого ячеек результирующего куба соответствующих агрегированных значений показателей.

Атрибут измерения (Attribute). Дополнительный элемент информации, относящийся к измерению и не являющийся при этом уникальным идентификатором или описанием этого измерения.

Бизнес-анализ. Набор задач и методов используемых в работе в процессе взаимодействия с заинтересованными лицами для того, чтобы понять структуру, политики и действия организации, и рекомендовать решения, которые позволят организации достичь своих целей.

Бизнес-аналитика (Business Intelligence). Предоставление точной, полезной информации соответствующим лицам, ответственным за принятие решений, в требуемых временных рамках для поддержания эффективности принятия решения.

Бизнес-правило. Политика, предписание, стандарт, правило или математическая формула, определяющая или ограничивающая некоторые стороны бизнес-процессов.

Бизнес-требования. Объем информации, который в совокупности описывает потребность, которая инициирует один или больше проектов, призванных предоставить решение и получить требуемый конечный бизнес-результат. Бизнес-требования включают бизнес-возможности, бизнес-цели, метрики успеха, концепция и границы и ограничения.

Бизнес-цель. Финансовая или нефинансовая выгода, которую организация ожидает получить в результате реализации проекта или другой инициативы.

Витрина данных (Data Mart). Хранилище данных, создаваемое для поддержки принятия решений в интересах какого-либо подразделения компании или для обеспечения каких-либо конкретных аспектов ее деятельности. Источником данных для витрины данных может быть общее хранилище данных компании или оно создается и функционирует независимо. Объем данных в витрине данных и его потребности в вычислительных ресурсах обычно существенно ограничены по сравнению с общим хранилищем данных.

Вращение измерений (Dimension Rotation). Операция, обеспечивающая изменение порядка измерений куба данных при его визуализации. Позволяет представить его пользователю в другом ракурсе.

Генератор отчетов. Программа или библиотека, позволяющая представить информацию в удобочитаемом структурированном виде. Другими словами сделать из данных информацию (документ, отчет), которую можно распечатать или сохранить в различных электронных форматах. Могут использоваться как в составе программ, программных систем и комплексов, так и самостоятельно (для анализа имеющихся данных безотносительно формирующих их систем).

Гиперкуб (Data Hypercube). См. Куб данных.

Гранулярность (Granularity). Представляет степень детализации фактов в кубе данных. Определяется количеством измерений, связанных с таблицей фактов, и иерархиями измерений.

Дезагрегирование данных (Data Disaggregation). Операция над кубом данных, которая строит новый куб большей размерности путем замены значений некоторых элементов измерений совокупностями значений элементов более низких уровней иерархии и декомпозиции соответствующих значений исходных показателей. Дезагрегирование данных позволяет, например, перейти от рассмотрения годовых данных по некоторым категориям продуктов к квартальным данным.

Денормализация (Denormalization). Процесс, противоположный нормализации. Процесс преобразования отношений базы данных к виду, допускающему избыточность данных, ориентированных в первую очередь на высокую производительность при выполнении аналитических запросов. Основное назначение денормализации состоит в ограничении межтабличных соединений в запросах.

Иерархия измерений (Hierarchy). Структура, состоящая из связанных между собой измерений, организованных в два или более уровня. Измерение, находящееся на более высоком уровне иерархии содержит одно или несколько измерений из нижележащего уровня иерархии. Иерархии определяют пути агрегирования/деагрегирования данных.

Измерение (Dimension). Способ ранжирования, используемый для разделения агрегированных фактов на составляющие их части.

Информационно-аналитическая система. Особый класс информационных систем, предназначенных для аналитической обработки данных. Они хранят как единое целое информацию, извлекаемую из множества учетных баз данных организации и других источников.

Консольная таблица (Outrigger table). Таблица в базе данных со схемой типа звезды или типа снежинки. В базе данных со схемой типа звезда консольные таблицы могут использоваться для хранения дополнительных атрибутов членов измерения. В базе данных со схемой типа снежинки консольные таблицы необходимы для хранения информации для построения иерархии измерения (отношений родитель-потомок между членами измерения).

Концептуальная модель данных (Conceptual data model). Является отражением предметной области, в рамках которой планируется построение хранилища данных, а также спектра аналитических задач, для решения которых разрабатывается хранилище.

Куб данных (Data Cube). Основная структура данных в многомерных моделях данных. Имеет несколько независимых измерений – систему координат представляемого пространства данных. Комбинации значений координат по всем измерениям определяют точки куба, называемые ячейками. С ячейками ассоциируются значения переменных, называемых показателями и имеющих, как правило, числовые типы.

Логическая модель данных (Logical data model). Аналитическая модель, изображающая логическую структуру данных, независимую от физической структуры данных или механизмов хранения данных.

Многомерная модель данных (Multi-Dimensional Data Model). Модель данных, оперирующая многомерными представлениями данных в виде кубов данных. Операционные возможности многомерных моделей данных включают, в частности, операции агрегирования и дезагрегирования данных, построения проекций куба, вращения измерений куба и др.

Модель данных (Data model). Абстрактное, самодостаточное, логическое определение объектов, операторов и прочих элементов, в совокупности составляющих абстрактную машину доступа к данным, с которой взаимодействует пользователь. Эти объекты позволяют моделировать структуру данных, а операторы – поведение данных.

Модель (Model). Упрощенное представление реальности, разработанное с целью донести информацию до определенной аудитории для обеспечения анализа, взаимодействия и понимания.

Нормализация (Normalization). Процесс преобразования отношений базы данных к виду, отвечающему нормальным формам. Цель нормализации: исключить избыточное дублирование данных, которое является причиной аномалий, возникших при добавлении, редактировании и удалении строк таблицы.

Показатель (Measure). Переменная, значениями которой являются факты, ассоциируемые с ячейками куба данных.

Проекция куба данных (Data Cube Projection). Операция над кубом данных, которая строит куб меньшей размерности при фиксированных значениях для каких-либо измерений. Этот термин обозначает также результат указанной операции.

Развертка (Drill down). Операция перемещения вниз по уровням иерархии измерения гиперкуба. См. также Дезагрегирование данных.

Свертка (drill up). Операция перемещения вверх по уровням иерархии измерения гиперкуба. См. также Агрегирование данных.

Срез куба (Data Cube Slice). См. Проекция куба данных.

Ссылочная целостность (Referential integrity). Необходимое качество реляционной базы данных, заключающееся в отсутствии в любом ее отношении внешних ключей, ссылающихся на несуществующие кортежи.

Схема «Звезда» (Star Schema). Схема реляционной базы данных, служащая для поддержки многомерного представления содержащихся в ней данных. Описываемая база данных включает таблицу фактов и ряд связанных с нею при помощи внешних ключей таблиц измерений. Таблицы измерений могут иметь составные первичные ключи и являются денормализованными. Благодаря этому упрощается восприятие структуры данных пользователем и формулировка запросов, уменьшается количество операций соединения таблиц при обработке запросов. Однако в связи с избыточностью данных возрастает требуемый для их хранения объем памяти.

Схема «Снежинка» (Snowflake Schema). Разновидность схемы типа звезды, предусматривающая нормализацию таблиц измерений. Первичные ключи в них состоят из единственного атрибута (соответствуют единственному элементу измерения). Это позволяет минимизировать избыточность данных и более эффективно выполнять запросы, связанные со структурой значений измерений.

Таблица измерений (Dimension Table). Таблица в базе данных со схемой типа звезды или типа снежинки. Каждая таблица измерений с помощью внешнего ключа связана с таблицей фактов. В схеме типа звезды таблица измерений денормализована. Напротив, в схеме типа снежинки используются нормализованные таблицы измерений, декомпозированные по уровням иерархии элементов измерений. Каждая такая таблица соответствует единственному элементу измерения.

Таблица фактов (Fact Table). Таблица в базе данных со схемой типа звезды или типа снежинки. Каждая ее строка соответствует некоторой ячейке куба данных. Она содержит набор фактов и по одному значению внешнего ключа для каждой таблицы измерений.

Транзакция. Логически целостная операция по обработке данных, обеспечиваемая последовательностью взаимно обусловлен-

ных (логически связанных) простых операций с данными. В базе данных транзакция предполагает цепочку логически связанных изменений данных.

Факт аддитивный (Additive Fact). Факты, допускающие агрегирование относительно любого измерения куба данных.

Факт неаддитивный (Nonadditive Fact). Факты, которые не могут агрегироваться ни по какому измерению куба данных.

Факт полуаддитивный (Semiadditive Fact). Факты, которые допускают агрегирование относительно одних измерений и не допускают относительно других.

Факт (Fact). Значение показателя, соответствующее какой-либо ячейке куба данных.

Физическая модель данных (Physical data model). Представляет реализацию объектов логической модели на уровне объектов конкретной СУБД.

Хранилище данных (Data Warehouse). Централизованный склад данных предприятия, содержащий полученные из разных источников консолидированные, очищенные ретроспективные данные.

Элемент измерения (Dimension Element). Уровень в иерархии значений координат некоторого измерения куба данных. Например, для измерения времени может быть задана иерархия «год – квартал – месяц». В этом случае данному измерению соответствует три элемента измерения, каждый из которых характеризуется своим номером уровня в иерархии.

Ячейка (Cell). Структурный элемент куба данных, соответствующий набору значений по всем измерениям. С ячейками куба ассоциируются значения показателей.

ЗАКЛЮЧЕНИЕ

В учебном пособии представлен базовый теоретический материал по технологии хранилищ данных, современным тенденциям в области бизнес-аналитики, необходимый для выполнения студентами практических заданий.

Кроме теоретического материала пособие содержит описание шести практикумов: от проектирования хранилища данных до применения средств бизнес-аналитики. Для каждого практикума даны рекомендации по организации программной среды; определены цель и план практикума, дано пошаговое описание практикума. Для закрепления полученных в ходе практикума знаний и навыков студентам предложены задания для самостоятельного выполнения и тесты для самоподготовки.

В результате изучения пособия, выполнения всех практикумов и самостоятельных заданий студенты закрепляют теоретические знания об организации и проектировании хранилищ данных, технологиях оперативного анализа данных и интеллектуального анализа данных; приобретают практические навыки разработки хранилищ данных и средств бизнес-аналитики с помощью современных инструментальных средств.

ЛИТЕРАТУРА

1. A Guide to the Business Analysis Body of Knowledge (BABOK Guide v3.0) [Электронный ресурс] // International Institute of Business Analysis, ИБА 2016: [сайт]. [2016]. URL: <https://www.iiba.org/babok-guide.aspx> (дата обращения: 10.10.2016).
2. Analysis Services Tutorials (SSAS) [Электронный ресурс] // MSDN. MS SQL Server 2016: [сайт]. [2016]. URL: <https://msdn.microsoft.com/en-us/library/hh231701.aspx> (дата обращения: 05.06.2016).
3. Codd, E.F.; Codd S.B. and Salley C.T. Providing OLAP to User-Analysts: An IT Mandate (1993).
4. *Дейт К. Дж.* Введение в системы баз данных. – 8-е изд. – М.: «Вильямс», 2016. – 1328 с.
5. Многомерное моделирование. Учебник по Adventure Works [Электронный ресурс] // MSDN. MS SQL Server 2014: [сайт]. [2016]. URL: [https://msdn.microsoft.com/ru-ru/library/ms170208\(v=sql.120\).aspx](https://msdn.microsoft.com/ru-ru/library/ms170208(v=sql.120).aspx) (дата обращения: 05.06.2016).
6. *Кондратьев В.В.* (ред.), Конструктор регулярного менеджмента, М.: Инфра-М, 2015. – 256 с.
7. *Маклаков С.В., Матвеев Д.В.* Анализ данных. Генератор отчетов Crystal Reports СПб.БХВ-Петербург, 2003. – 496 с. ISBN 5-94157-230-1.
8. *Полубояров В.В.* Использование MS SQL Server 2008 Analysis Services для построения хранилищ данных. – М.: Национальный Открытый Университет «ИНТУИТ», 2010. – 487 с.
9. *Сарка Д.* Microsoft SQL Server 2012. Реализация хранилищ данных: учебный курс Microsoft-М.: Изд-во «Русская редакция», 2014. – 816 с.

10. *Точилкина Т.Е.* Многоуровневые модели данных предприятия и их поддержка в СА ERwin Data Modeler // Вестник развития науки и образования. – 2009. – № 2. – С. 25–28.
11. *Точилкина, Т.Е.* Принципы создания информационных систем и моделирования бизнес-процессов с использованием пакета программ ALLFusion Modeling Suite: учебное пособие. Ч.II. Моделирование данных и проектирование баз данных с ERwin Data Modeler / Т.Е. Точилкина. – М.: изд. АБиК М-ва финансов РФ, 2010. – 221 с.
12. *Туманов В.Е.* Проектирование хранилищ данных для приложений систем деловой осведомленности (Business Intelligence Systems) 2-е изд., испр. – М.: Национальный Открытый Университет «ИНТУИТ», 2016. – 958 с.

Учебное электронное издание на диске

**Татьяна Евгеньевна Точилкина
Алла Александровна Громова**

**ХРАНИЛИЩА ДАННЫХ И
СРЕДСТВА БИЗНЕС-АНАЛИТИКИ**

Учебное пособие

Публикуется в авторской редакции

Техническое редактирование
и компьютерная верстка *Л.Б. Галкиной*

Подписано в печать 17.02.2017.

Гарнитура Times.

Объем 8,5 МБ. Тираж 9 экз. Заказ № 118.

Финансовый университет
Москва, 125993 (ГСП-3), Ленинградский просп., 49