

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

УДК 004.42

Вахович  
Иван Олегович

БИЗНЕС-АНАЛИЗ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ СРЕДСТВ MICROSOFT

Автореферат на соискание академической степени  
магистра технических наук

по специальности 1-40 80 04 – Математическое моделирование, численные методы и комплексы программ

Научный руководитель  
Волорова Н.А.  
к.т.н., доцент

Минск 2015

## **ВВЕДЕНИЕ**

К настоящему времени во многих организациях накоплены значительные объемы данных, на основе которых имеется возможность решения разнообразных аналитических и управленческих задач. Проблемы хранения и обработки аналитической информации становятся все более актуальными и привлекают внимание специалистов и фирм, работающих в области информационных технологий, что привело к формированию полноценного рынка технологий бизнес-анализа.

В идеале работа аналитиков и руководителей различных уровней должна быть организована так, чтобы они могли иметь доступ ко всей интересующей их информации и пользоваться удобными и простыми средствами представления и работы с этой информацией. Именно на достижение этих целей и направлены информационные технологии, объединяющиеся под общим названием хранилищ данных и бизнес-анализа.

## **ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ**

Тема диссертации является актуальной, так как системы бизнес-анализа данных используются в настоящее время многими компаниями для анализа и стали для многих компаний стратегической основой и важнейшим фактором успешной деятельности.

Цель данного исследования состоит в исследовании существующих систем анализа данных, изучении средств, предоставляемых компанией Microsoft в этой области и создании системы для анализа производительности сотрудников компании.

В диссертации был сделан обзор популярных систем бизнес-анализа данных, выявлены их достоинства и недостатки. Были подробно изучены средства компании Microsoft. Также были сформулированы принципы разработки систем бизнес-анализа данных и основные функциональные возможности данных систем.

## **КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ**

В соответствии с определением Gartner, бизнес-анализ (BI, Business Intelligence) - это категория приложений и технологий для сбора, хранения, анализа и публикации данных, позволяющая корпоративным пользователям принимать лучшие решения. В русскоязычной терминологии подобные системы называются также системами поддержки принятия решений.

Сбор и хранение информации, а также решение задач информационно-поискового запроса эффективно реализуются средствами систем управления базами данных (СУБД). В OLTP (Online Analytical Processing) - подсистемах реализуется транзакционная обработка данных. Непосредственно OLTP-системы не подходят для полноценного анализа информации в силу противоречивости требований, предъявляемых к OLTP-системам и СППР.

Для предоставления необходимой для принятия решений информации обычно приходится собирать данные из нескольких транзакционных баз дан-

ных различной структуры и содержания. Основная проблема при этом состоит в несогласованности и противоречивости этих баз-источников, отсутствии единого логического взгляда на корпоративные данные.

Поэтому для объединения в одной системе OLTP и СППР для реализации подсистемы хранения используются концепция хранилищ данных (ХД). В основе концепции хранилищ данных лежит идея разделения данных, используемых для оперативной обработки и для решения задач анализа, что позволяет оптимизировать структуры хранения. Хранилище данных позволяет интегрировать ранее разъединенные детализированные данные, содержащиеся в исторических архивах, накапливаемых в традиционных OLTP-системах, поступающих из внешних источников, в единую базу данных, осуществляя их предварительное согласование и, возможно, агрегацию.

Подсистема анализа может быть построена на основе:

- подсистемы информационно-поискового анализа на базе реляционных СУБД и статических запросов с использованием языка SQL

- подсистемы оперативного анализа. Для реализации таких подсистем применяется технология оперативной аналитической обработки данных OLAP, использующая концепцию многомерного представления данных;

- подсистемы интеллектуального анализа, реализующие методы и алгоритмы Data Mining.

Технология хранилищ данных предназначена для хранения и анализа больших объемов данных с целью дальнейшего обнаружения в них скрытых закономерностей и, наряду с технологией Data Mining, входит в понятие "предсказательная аналитика". Data Mining, в свою очередь, изучает процесс нахождения новых, действительных и потенциально полезных знаний в базах данных. Хранилище данных - предметно-ориентированный, интегрированный, редко меняющийся, поддерживающий хронологию набор данных, организованный для целей поддержки принятия решений. Предметная ориентация означает, что хранилище данных интегрирует информацию, отражающую различные точки зрения на предметную область. Интеграция предполагает, что данные, хранящиеся в хранилище данных, приводятся к единому формату.

Основная проблематика при создании хранилище данных заключается в следующем:

- интеграция разнородных данных. Данные в хранилище данных поступают из разнородных OLTP-систем, которые физически могут быть расположены на различных узлах сети.

- эффективное хранение и обработка больших объемов данных. Построение хранилище данных предполагает накопление данных за значительные периоды времени, что ведет к постоянному росту объемов дисковой памяти, а также росту объема оперативной памяти, требующейся для обработки этих данных. При возрастании объемов данных этот рост нелинеен;

- организация многоуровневых справочников метаданных. Конечным пользователям СППР необходимы метаданные, описывающие структуру хранящихся в хранилище данных, а также инструменты их визуализации;

- обеспечение информационной безопасности хранилище данных.

На данный момент в мире наиболее популярны следующие системы анализа данных:

- IBM;
- Microsoft
- Oracle
- SAP

### **Обзор средств Microsoft для работы с BI**

Microsoft BI представляет собой набор продуктов (SQL Server, Office, SharePoint), позволяющих организациям принимать решения на основании достоверной информации, полученной из внутренних и внешних источников данных. Рассмотрим подробнее компоненты Microsoft BI.

В решении Microsoft BI пользователи могут использовать для анализа данных в хранилище и аналитических кубах следующие инструменты.

SQL Server Reporting Services (SSRS) – это серверное решение и набор программ, предназначенных для создания и отображения отчетов с произвольным точным оформлением. Созданные отчеты можно разместить либо на специальном сайте Reporting Services, либо на портале SharePoint.

В SSRS присутствуют средства для работы с регламентированными сложными отчетами строго заданной формы, а также с интерактивными отчетами.

В качестве инструментов разработки отчетов можно использовать:

Power, Report Builder, Report Designer в Business Intelligence Development Studio (Visual Studio).

Sharepoint. Функции SharePoint 2013 для бизнес-аналитики можно разделить на 3 группы: инструменты исследования данных позволяют объединять данные из различных источников, создавать интерактивные отчеты и исследовать данные.

Возможности визуализации предоставляют различные варианты отображения и совместного анализа данных в пределах организации. Реализуются через функции Power View, PerformancePoint Services, Visio Services.

Excel Services позволяет пользователям просматривать и взаимодействовать с книгами Excel (в том числе с аналитическими моделями PowerPivot), опубликованными на портале SharePoint. Непосредственно с портала (через веб-браузер) можно модифицировать таблицы (PivotTable) и диаграммы (PivotChart) на базе аналитической модели BISM, а также выполнять операции детализации данных (Drill Down To) и запрашивать предлагаемые способы просмотра данных (Data Exploration). Также поддерживаются расчетные меры и расчетные элементы, созданные в Excel. Новый элемент визуализации «шкала времени» (TimeLine) также поддерживается. Т.е. Excel Services поддерживает перечисленные функции так же, как и Excel 2013.

Application BI Servers. Администраторы могут настроить выделенные серверы SQL Server Analysis Services для поддержки дополнительных аналитических возможностей Excel Services.

PerformancePoint Services. В SharePoint 2013 PerformancePoint Services продолжает оставаться средством для создания произвольных информационных панелей с возможностью отображения карт показателей и отчетов (интерактивных, со строгим оформлением, диаграмм и таблиц из Excel, схем из Visio и т.д.). Этот компонент позволит создать нестандартные панели, которые невозможно реализовать с помощью Power View.

Visio Services. Компонент Visio Services – это сервисное приложение, позволяющее пользователям просматривать на портале рисунки Visio, в том числе и подсоединенные к источникам данных (отражающие значения данных на схемах). В SharePoint 2013 Visio Services появился параметр управления кэшем и соответствующие правила управления этим параметром, обновлены команды PowerShell, добавлена возможность комментирования на схемах Visio (\*.vsdx), что является важным дополнением для возможностей командной работы на портале.

### **Создание системы анализа данных компании**

В практической части своей диссертации я бы хотел провести анализ данных компании X. Компания X – это IT компания, которая занимается аутсорсингом. В компании X ведется учет рабочего времени. Каждый сотрудник обязан отчитываться в определенной системе учета времени о часах, которые потратил на определенные задачи каждый день. Данные о рабочем времени сотрудников хранятся в базе.

Для анализа данных нам нужно создать еще одну базу, так называемый data warehouse, данные в которую будут переливаться из основной базы. База для анализа будет денормализованной, иметь структуру звездочки.

Итак, приступим к созданию еще одной базы, назовем ее CompanyDW. В базе нам нужны таблицы измерений и таблицы фактов. Для наших отчетов нам нужно знать о том, сколько сотрудники времени работали над определенным проектом, и нужно знать сколько компания получает выгоды от работы сотрудника над определенным проектом. Таким образом у нас получается две таблицы с фактами.

В качестве фильтров нам понадобятся сами работники, отделы, заказчики, проекты и время. Это значит что у нас получится 5 таблиц с измерениями. Таблица Tasks не фигурирует в нашей новой базе, так как она становится не нужной. ETL процесс, который будет заполнять базу для анализа, будет итерироваться по таблице задач и писать уже агрегируемую информацию в новую базу в таблицу фактов о зарепорченном времени сотрудниками.

Для импорта данных из первой базы во вторую используется ETL процесс средствами Integration Services Microsoft. С помощью этого средства можно задать правила копирования из источника данных в приемник данных. В нашем случае приемником служит CompanyDW, источником – Company. Получив данные в базе CompanyDW мы можем приступить к созданию куба и после отчетов.

Итого, в базе данных для анализа у нас в таблице пользователей хранится **54** записи, в таблице заказчиков – **11** записей, в таблице проектов – **44** записи, в

таблице времени – **1000** записей, в таблице отделов – **4** записи. Для данных измерений в таблице выручки по пользователям и проектам хранится **2376** записей, а в таблице отчетов по времени – **146316** записей.

Пора перейти к созданию куба. Для этого нам нужно создать проект в Business Intelligence Studio. Первым делом в проект нужно добавить источник данных. В качестве источника данных у нас будет база CompanyDW. Добавив источник нужно добавить представления, из которых будут браться измерения и факты. В нашем случае, так как мы используем уже готовые к анализу данные, нам не нужно преобразовывать данные

Чтобы создать куб в Business Intelligence Studio нужно заполнить необходимые данные в окне создание куба. Нужно выбрать созданное нами представление источника данных и нужные нам таблицы из представления.

Прямо из Business Intelligence Studio мы можем его опубликовать и обработать. Далее этот куб будет использоваться для построения отчетов, которые будут публиковаться в системе Sharepoint.

После публикации куб доступен на аналитическом сервере MS Sql. Я могу использовать его, например в Excel, или в Sql Management Studio.

### **Отчет о производительности сотрудников**

Приступим к созданию отчетов. Отчет о производительности сотрудников – это отчет, который показывает количество отработанных часов, сгруппированные по периодам и фильтрацией по отделам, клиентам и проектам.

Для разработки отчета опять же будем использовать Business Intelligence Studio. Для создания отчета нужно создать новый проект отчетов в студии. В создавшемся проекте первым делом нужно добавить источник данных. В качестве источника мы выбираем наш куб.

После создания источника данных нужно создать датасет. Датасет можно создавать как встроенный в отчет, так и общий. Встроенный в отчет датасеты хранятся в теле самого отчета. Есть определенные плюсы такого подхода, например, для некоторого типа отчетов нельзя по-другому, но это отдельная тема. Предыдущее утверждение так же актуально и для источника данных. В нашем примере мы будем использовать встроенные источник данных и датасет.

Самый быстрый способ создать отчет – это запустить визард создания отчетов в студии.

В визарде первым делом заполняем источник данных, указывая наш куб в качестве.

Далее визард предложит создать запрос для датасета. Здесь проще воспользоваться дизайнером. Нам нужно добавить измерения в качестве параметров и данные. В качестве параметров добавим название отдела, клиента и проект. Как данные выведем год, месяц, число и количество внесенного времени.

Далее выбираем тип отображения данных, а после выбираем какие поля отображать. Все эти данные можно позже поменять, отредактировав отчет.

В итоге мы получаем нужный нам отчет. Вверху в отчете видно фильтры по отделам, клиентам и проектам. В самой таблице видим сгруппированные по периодам данные о зарегистрированном времени работы.

## **График доходов компании**

Перейдем к созданию следующего отчета. Отчет будет показывать график доходов компании за определенный период. Доход компании рассчитывается по формуле  $Доход = Сумма([количество\ часов, \ отработанных\ сотрудником] * [доход\ компании\ за\ час\ работы\ сотрудника\ исходя\ из\ дохода\ от\ его\ работы\ для\ компании])$ . Отчет будет содержать фильтры по годам, компаниям и проектам. Итак, приступим к созданию отчета. Снова воспользуемся визардом. Выбираем нужные измерения и факты в дизайнера и получаем начальный вид нашего отчета.

Далее нам нужно вместо таблицы, которая создалась в отчете по умолчанию, вставить график производительности. Все нужные данные у нас есть, так что просто удаляем таблицу и вставляем компонент Chart.

Далее нужно настроить данные для отображения в чарте. В качестве категорий выбираем даты, в качестве значений – сумму произведения дохода на часы пользователя. В легенду выводим компании. И получается нужный нам отчет.

## **Интеграция с Microsoft Sharepoint**

Платформа Sharepoint имеет широкие возможности для интеграции других средств бизнес-анализа от Microsoft. Для отображения наших отчетов нужно, чтобы были выполнены следующие действия:

SQL Server Reporting Services должен быть установлен в режиме интеграции SharePoint

Настроить репортинг сервис на странице настроек в панели администрирования Sharepoint

Фича Report Server Integration должна быть активирована на сайте в Sharepoint

Список с типом данных Отчет должен быть добавлен на сайт Sharepoint

После выполнения этих действий можно добавить вебчасть “Server Reporting Services Webpart” на требуемую страницу и отобразить требуемый отчет.

## **ЗАКЛЮЧЕНИЕ**

### **Основные результаты диссертации**

1. Была изучена теоретическая часть по предмету бизнес-анализ данных.
2. В ходе работы над диссертацией были изучены средства Microsoft для бизнес-анализа данных. Microsoft предоставляет весьма полный и широкий выбор как клиентских, так и серверных средств хранения данных, их калкуляций и анализа.
3. Был предложен метод хранения данных об отработанном времени сотрудниками компании. Была предложена схема движения данных об отработке сотрудниками. Для хранения данных используется MS SQL Server.

4. Было создано информационное хранилище, которое заполняется данными из главного хранилища данных компании. Информационное хранилище используется для создание куба для анализа данных. В качестве сервера для информационного хранилища используется MS SQL Server
5. Был спроектирован и создан куб, для анализа данных о производительности работников компании и доходов компании. Для создания куба использовалась Microsoft SQL Server Business Intelligence Development Studio.
6. Были созданы отчеты о производительности сотрудников и доходах компании. Отчеты созданы средствами Microsoft SQL Server Business Intelligence Development Studio. Они портируются в систему Microsoft Sharepoint и доступны для использования руководящим сотрудникам компании.

### **Рекомендации по практическому использованию**

1. Полученные результаты формируют теоретическую и практическую часть для разработки системы анализа данных и принятия решений в компании. Они могут быть использованы как для создания такого рода системы, так и для модернизации существующей системы
2. Результаты работы могут быть использованы как стартовый материал для изучения технологии бизнес-анализа данных.

### **Список публикаций соискателя**

- 1.А. Вахович И.О. Бизнес-анализ данных с использованием средств Microsoft.  
// Компьютерные системы и сети: материалы 49-й научной конференции аспирантов, магистрантов и студентов. Минск, 6-10 мая 2013г. / БГУИР, главн.ред.: В.А.Прытков. - Минск 2013. С.111