

Разработка информационно-аналитической системы на основе многомерного хранилища данных

Пучков Е.В., Пономарева Е.И. (каф. ПМиВТ, Ростов-на-Дону)

Ключевая роль в развитии системы российского образования традиционно принадлежит классическим университетам, которые реализуют программы обучения разных уровней и проводят исследования по приоритетным направлениям. Современные университеты не уступают по сложности крупным коммерческим предприятиям, а потому их нормальное функционирование уже невозможно без хранилищ данных и построенных на их основе аналитических приложений[1].

Целью данной статьи является развитие автоматизированной технологии балльно-рейтинговой системы оценки успеваемости студентов для мониторинга и анализа качества учебного процесса в ходе проведения ежемесячной аттестации студентов и по результатам экзаменационной сессии.

Для достижения цели были поставлены следующие задачи:

- 1) спроектировать многомерное хранилище данных;
- 2) разработать процедуры извлечения, трансформации и загрузки данных в хранилище;
- 3) построить OLAP-куб, добавить вычисляемые члены;
- 4) построить модели data mining для решения прикладных аналитических задач;
- 5) создать и опубликовать отчеты на сервере отчетов.

Перечислим основные функциональные требования, предъявленные к разрабатываемой системе.

1. Создать информационную базу, позволяющую проводить анализ и формирование отчетных документов по следующим учетным элементам: студент, учебная группа, учебный курс, специальность, дисциплина, преподаватель, кафедра, институт.

2. Рассчитать академическую активность по формуле:

$$\left(\frac{C-H}{C}\right) \times 100\%,$$

где С – общее количество студентов, которые должны держать экзамен,
Н – количество неаттестованных студентов (не допуск, неявка, неудовлетворительная оценка).

3. Рассчитать абсолютную успеваемость по формуле:

$$\left(\frac{T}{C}\right) \times 100\%,$$

где Т - количество студентов, получивших оценку «отлично» и «хорошо»,
С - общее количество студентов, которые должны держать экзамен.

4. Провести анализ посещаемости занятий.

5. Рассчитать средний балл успеваемости студентов.

6. Провести анализ успеваемости по аттестациям студентов.

7. Проанализировать пересдачи экзаменов по дисциплинам и причины отчисления студентов.

8. Проанализировать изменения в составе студентов по годам набора по ряду признаков (регион, родной язык, возраст, пол, служба в армии и т.д.).

Физическая реализация системы представляет собой комбинированную двух- и трехуровневую архитектуру для Windows-приложений и Web-приложений соответственно. Управление данными происходит на базе Microsoft SQL Server 2008 Standard. Windows-приложения напрямую обращаются к серверу баз данных. Web-приложения обращаются через сервер приложений (Internet Information Services) на базе Microsoft Windows Web Server 2008.

Исходные данные для хранилища были получены из автоматизированной системы управления «Университет». Разработка информационно-аналитического приложения велась на платформе Microsoft Business Intelligence, которая включает в себя инструменты для управления данными и бизнес-анализа: SQL Server Integration Services, SQL Server 2008 Analysis Services, Службы отчетов SQL Server 2008 Reporting Services, SQL Server Database Management Studio, SQL Server Business Intelligence Development Studio, Приложение Office Excel 2007.

На первом этапе построено хранилище данных по схеме снежинка. Она избавлена от дублирования данных и является более простой в обслуживании[2]. В таблицах фактов содержатся меры, например, количество пропусков, аттестация студентов и внешние ключи таблиц измерений, чтобы была возможность просматривать информацию в различных разрезах. На основе существующих таблиц построен куб (рис.1)

Затем были установлены дополнительные связи между фактами и измерениями. При отсутствии явных соединений между таблицами создавались связи многие-ко-многим и ссылочные связи, чтобы соединить меру и измерение. Добавлены именованные наборы и вычисляемые члены, позволяющие, например, отфильтровать студентов, которые являются «круглыми» отличниками и рассчитать дополнительные измерения[3].

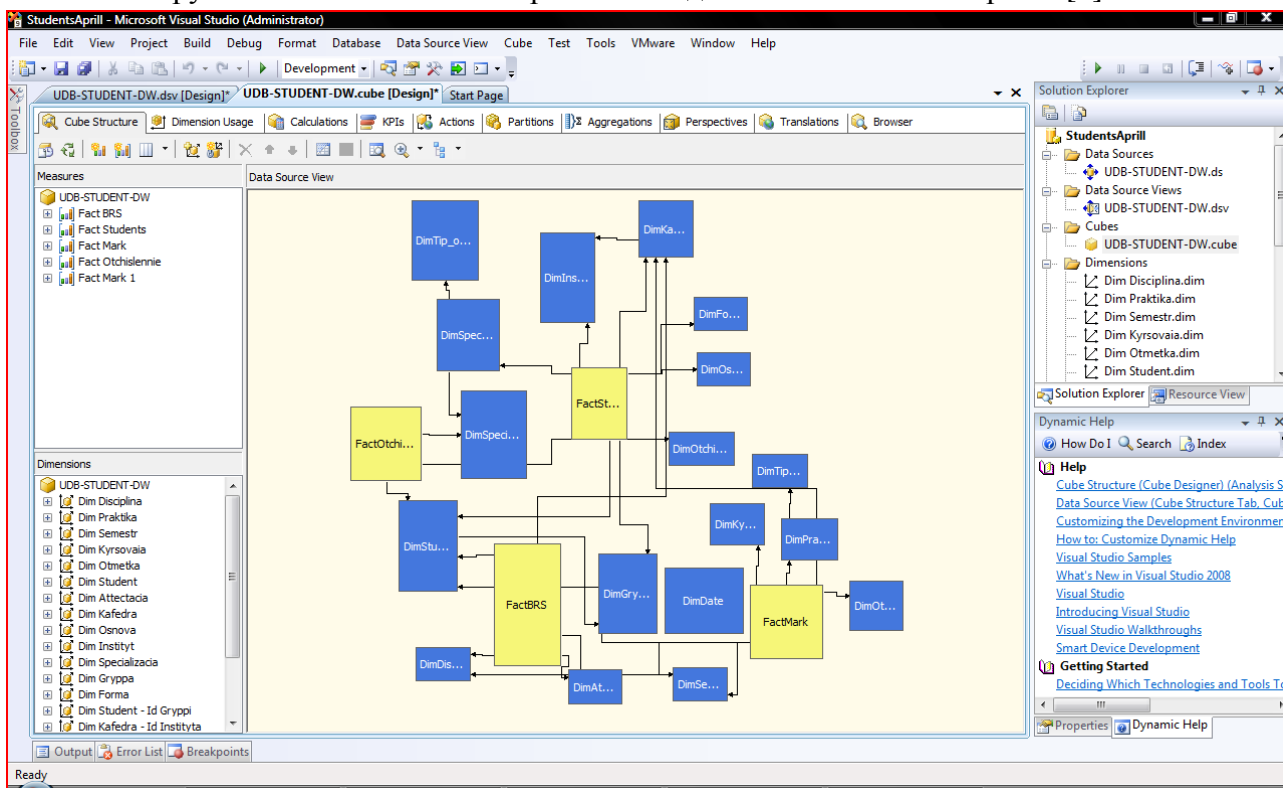


Рис.1. Структура куба в Analysis Services.

В листинге 1 создается вычисляемый член, с помощью которого рассчитывается академическая активность:

Листинг 1. Расчет академической активности.

```

CREATE MEMBER CURRENTCUBE.[Measures].[Academic] AS
    ([Measures].[Fact Mark Count]
    - ([Measures].[Fact Mark Count]
    ,[Dim Otmetka].[Id Otmetki].&[1]))
    /
    [Measures].[Fact Mark Count],

```

```

FORMAT_STRING = "Percent",
ASSOCIATED_MEASURE_GROUP = 'Fact Mark';

```

Для построения модели интеллектуального анализа данных и выявления изменений в составе студентов по ряду признаков использовался алгоритм кластеризации. Алгоритм кластеризации (Microsoft) представляет собой алгоритм сегментации, реализуемый с помощью службы SQL Server 2008 Analysis Services (SSAS). Алгоритм использует итерационные методы для группировки вариантов в наборы данных в кластерах, содержащих подобные характеристики. Модели кластеризации определяют связи в наборе данных, который невозможно логически получить с помощью случайного наблюдения[4]. В качестве входных данных использовались атрибуты измерения «Студент». В результате была построена модель кластеризации. На рис. 2 отображена диаграмма кластеров.

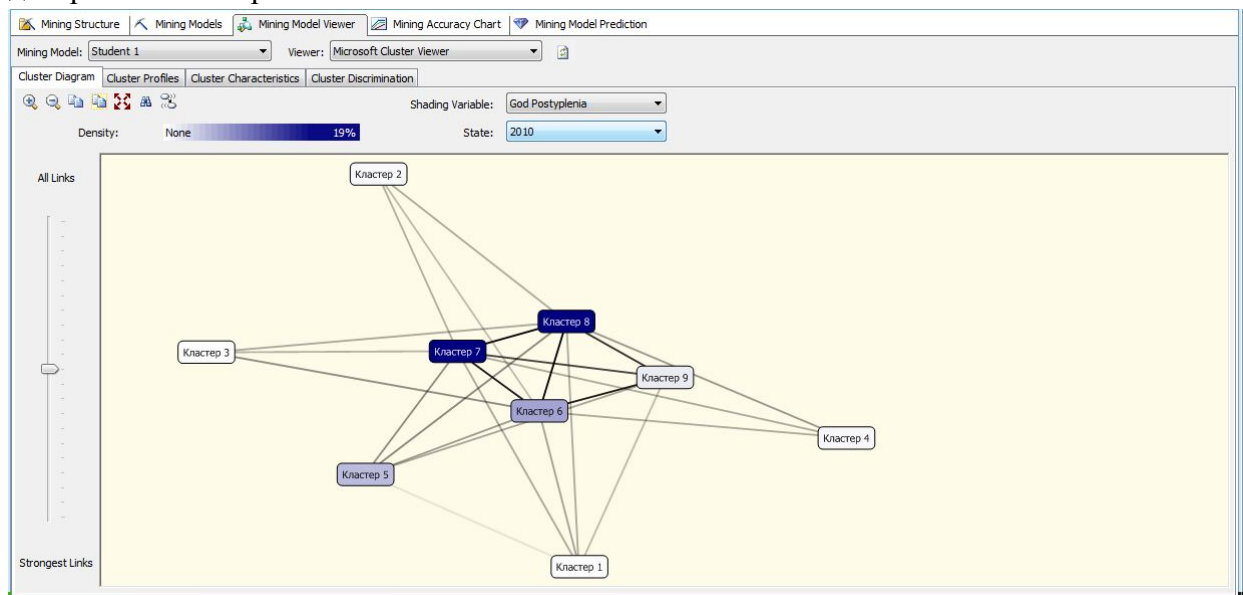


Рис. 2. Диаграмма кластеров.

Процесс заполнения хранилищ данных требует надлежащей подготовки и поддержки данных. Для этого следует использовать возможности преобразования потоков данных (DataFlow Transformations). Эти преобразования можно использовать, к примеру, для построения выборки, вычисления количества записей и объединения данных.

Для загрузки данных в хранилище создан пакет (рис.3), содержащий следующие этапы: подготовка базы данных (БД), создание скриптами T-SQL необходимых таблиц в хранилище, загрузка в созданные таблицы данных из источников(на этом этапе происходит преобразование данных и вычислений с помощью T-SQL), установка связей между таблицами в хранилище.

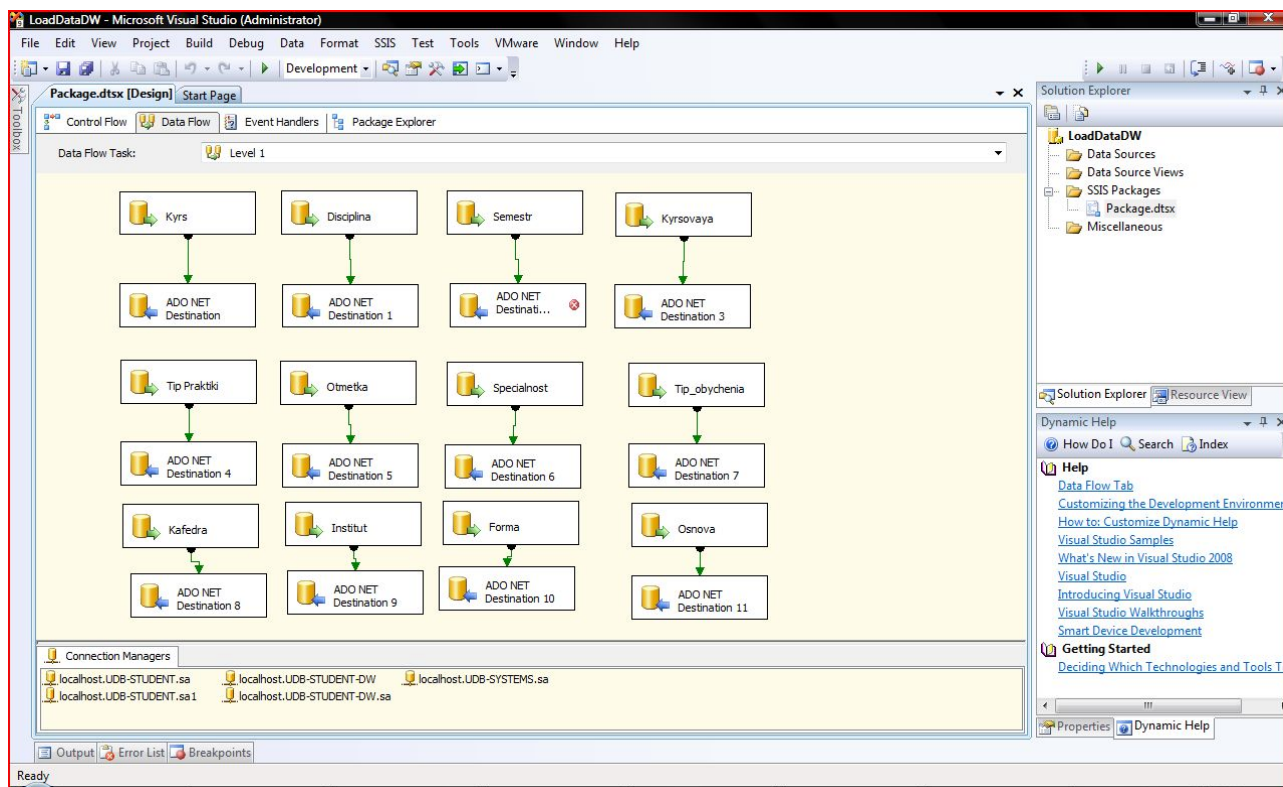


Рис.3. Загрузка данных из таблиц-источников в хранилище.

С помощью конструктора отчетов спроектированы отчеты на основе многомерной базы данных. Конструктор отчетов в Business Intelligence Development Studio используется для создания инфраструктуры и схемы отчета. Сервер отчетов (Report Server), как понятно из его названия, предназначен для передачи отчетов пользователям. Управлять сервером отчетов можно через Web-интерфейс, а также с помощью SSMS (SQL Server Management Studio).

В результате была создана информационно-аналитическая система, которая выполняет требуемый круг задач для руководителей образовательного учреждения. Разработаны хранилище данных, процедуры по загрузке данных, OLAP-куб, модели data mining, сервер отчетов.

В дальнейшем предполагается дополнить информационно-аналитическую систему такими средствами визуализации отчетов, как информационные панели (dashboards) и карты показателей (scorecards), а также доступом к отчетам через Internet.

Литература.

1. Пономарева Е.И., Мисюра В.В. Разработка системы поддержки принятия решений с использованием OLAP-технологий / сборник тезисов XII Всероссийского симпозиума по прикладной и промышленной математике (весенняя сессия). - 2011
2. Taniar, David Progressive methods in Data Warehousing and Business Intelligence: Concepts and Competitive Analytics / Monash University. – Australia. – 2009. – 390 p.
3. Сивакумар Харинатх, Мэтт Кэррол, Сетху Минакхисундарам, Роберт Зар, Денни Гуанг-Ю Ли / Microsoft SQL Server Analysis Services 2008 и MDX для профессионалов. – 2010
4. Библиотека MSDN (по-русски) [Электронный ресурс] / Microsoft. – URL: <http://msdn.microsoft.com/ru-ru/library/>