**BC/NW 2017 № 1 (30):14.2**

**ПРИМЕНЕНИЕ СИСТЕМ БИЗНЕС АНАЛИТИКИ ДЛЯ ОБРАБОТКИ ПОКАЗАТЕЛЕЙ ВУЗА**

Тумкина А.И.

В век современных технологий любая крупная или небольшая, частная или государственная компания, использует в своей деятельности информационные системы. Это говорит о том, что большинство предприятий уже имеет определенный объем накопленных данных, который представляет собой немалую ценность.

У крупного предприятия и ВУЗа много общего: сопоставимы финансовые показатели, численность персонала, филиальная сеть и т.п. Выпускаемые вузом специалисты — очень трудоемкая «продукция», производство которой невозможно без эффективного управления учебным заведением. Сегодня консолидируется научный и образовательный потенциал учебных и научно-исследовательских учреждений, возникают объединенные вузы, комплексы и ассоциации университетов, а потому роль информационной инфраструктуры становится все более важной. Система отечественного высшего образования стоит на пороге интеграции отечественных образовательных стандартов с международными, а значит, необходимо управление качеством образовательного процесса.

Для того чтобы существующие хранилища данных способствовали принятию управленческих решений, информация должна быть представлена аналитику в нужной форме. Иначе говоря, специалист должен иметь развитые инструменты доступа к данным хранилища и их обработки.

Однако большинство информационно-аналитических систем являются статистическими. Системы такого типа содержат в себе определенный набор запросов, поэтому каждый новый запрос должен быть формально описан, закодирован программистом и затем выполнен. с вновь появившимися. Такой процесс занимает много времени и лишает систему гибкости. Кроме того, отчёты, построенные в этих системах громоздки, неструктурированны и неинформативны. [1]

Динамические системы поддержки принятия решений, напротив, ориентированы на обработку разовых запросов аналитиков. аналитиков с системами такого типа заключается в интерактивной последовательности формирования запросов и изучения их результатов.

Таким образом, динамические системы могут быть задействованы не только в области оперативной аналитической обработки данных, но и в других базовых сферах, в которых поддержка принятия управленческих решений осуществляется на основе накопленных данных.[1]

Основным недостатком реляционной модели с точки зрения множественности измерений является невозможность объединения, просмотра и анализа данных.

Более быструю реакцию на запросы обеспечивают многомерные СУБД за счёт организации данных в виде упорядоченных многомерных массивов. Массивы представляют собой гиперкуб, в котором все хранимые ячейки имеют одинаковую мерность, и витрины данных, являющиеся предметно – ориентированным подмножеством хранилища данных. Системы, используемые для работы с многомерной моделью данных, называются OLAP-системами. [2]

OLAP-система (On-Line Analytical Processing, аналитическая обработка) – это технология комплексного многомерного анализа данных, заключающаяся в подготовке суммарной информации на основе крупных массивов данных.

В состав OLAP-системы входит множество компонентов (Рисунок 1). Высший уровень системы определен источником данных, многомерной базой данных, посредством которой реализован механизм построения отчётов, OLAP-сервер, клиент. Клиент-серверный принцип построения системы предоставляет удалённый, многопользовательский доступ к серверу многомерной базы данных. [2]

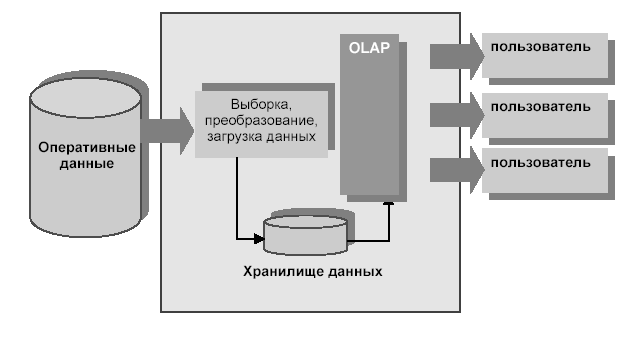


Рисунок 1 Структура OLAP – системы

Разработка OLAP-системы подразумевает выполнение следующего алгоритма [2]:

* 1. Определение и создание в хранилище данных таблицы фактов, измерений, иерархий и мер проектируемого куба;
  2. Разработка сценария преобразования данных из хранилища – источника в хранилище OLAP;
  3. Построение и публикация OLAP-куба на сервере.

Таблица фактов и таблицы измерений определяются на основании схемы источника данных. (Рисунок 2).

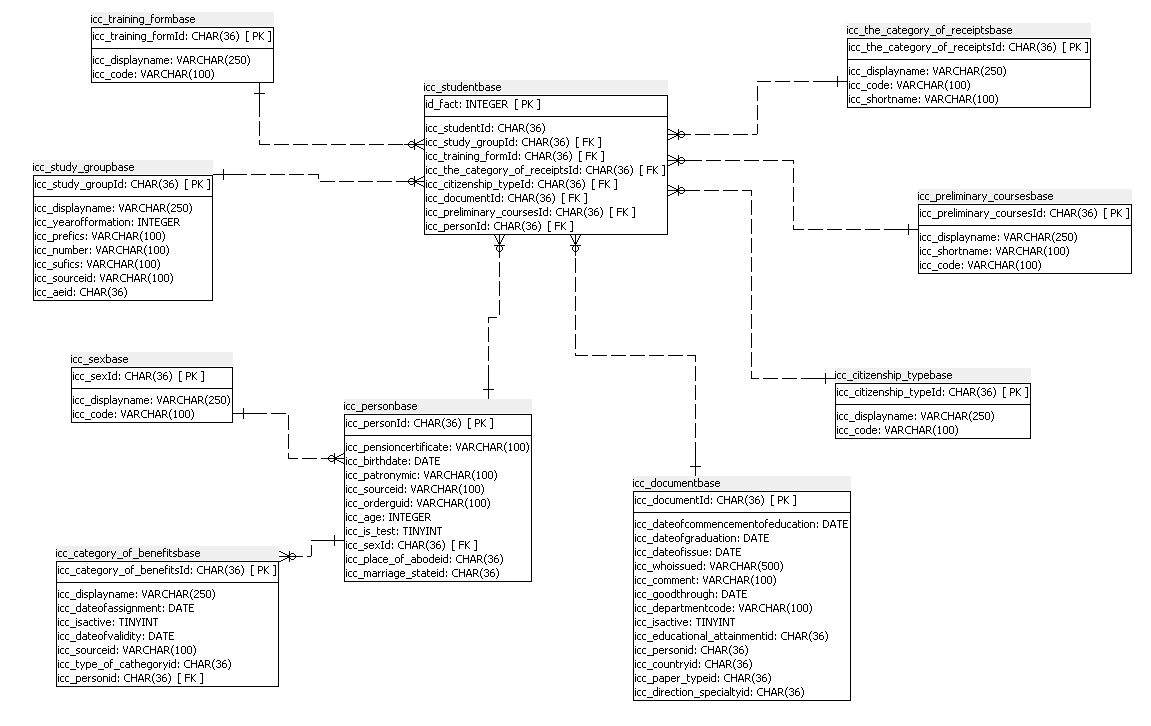


Рисунок 2 Схема данных многомерного куба

Таблицы не содержат каких-либо сведений. Для их наполнения необходимо сформировать «подключение-источник», из которого информация будет забираться, и передаваться в «подключение-цель».

После настройки источников составляется схема передачи данных, в которой используется два основных типа документа: трансформация (transformation) и работа (job).[1]

* трансформация (transformation) используется для описания потоков данных, полученных при извлечении, преобразовании и выгрузке данных.

Каждая трансформация, создаваемая в проекте, состоит из последовательных шагов: чтение данных, их сортировка, соединение, запись данных (рисунок 3).

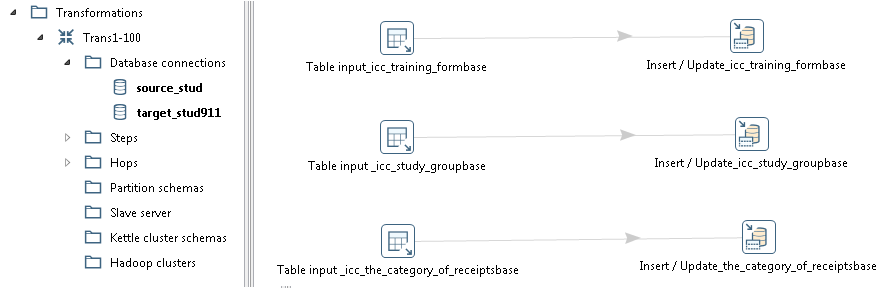


Рисунок 3 Потоки данных

* работа (job) применяется для координации таких действий с данными, как определение их потока и управление последовательным запуском трансформаций.

Применение функциональной декомпозиции помогает выделить в работе элементы, из которых она состоит. Такие элементы называются записями (рисунок 4).

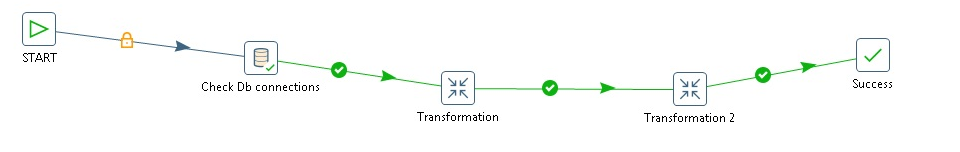


Рисунок 4 Последовательный запуск трансформации

Многомерный куб создаётся при помощи графического интерфейса продукта Pentaho Mondrian Workbench. Визуальное создание схемы куба выполняется в режиме ROLAP схемы. XML - модели работают с многомерными кубами, используя существующую таблицу фактов и таблицы измерений.

К создаваемому кубу привязывается таблица фактов. Далее задаются измерения куба – это параметры, по которым будет производиться срез данных (рисунок 5).

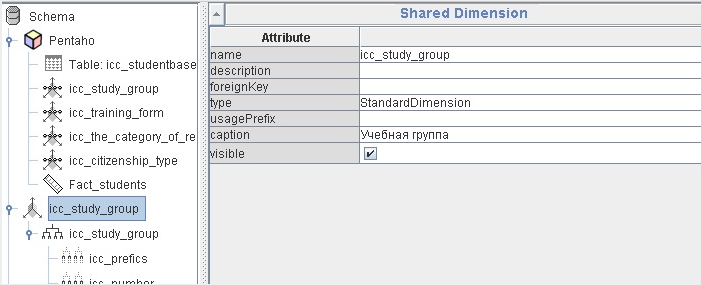


Рисунок 5 Создание измерения

В каждом измерении создаётся иерархия, а затем в ней выделяется один или несколько уровней. Для этого таблицу фактов связывается с таблицей того измерения, в котором создаётся уровень.

Инструментом сбора статистики в OLAP-системе является мера. Каждая создаваемая мера поддерживает функцию агрегирования, которая производит вычисление обобщающего значения для выбранных строк таблицы [2]. По результатам запроса можно получить информацию о том, какое количество студентов – граждан РФ, иностранного государства, а также иностранцев, приравненных в правах к гражданам РФ, было принято в учебное заведение в течение нескольких лет на каждую, из имеющихся в ВУЗе, форм обучения по конкурсу, без экзаменов или на платной основе.

По завершении настройки многомерный куб публикуется на сервере, где конечному пользователю предлагается выбрать интерфейс для выполнения запросов.

Результатом выполнения запроса является таблица, в полях которой доступны сведения, отвечающие заданным пользователем параметрам.

****

Рисунок 6 Табличный режим отображения отчёта

Возможности системы позволяют представить данные в виде диаграмм и гистограмм (рисунок 7), а также просмотреть статистику по выборке: минимальный и максимальный показатель, среднее значение, суммарное значение, значение, разброс (рисунок 8).

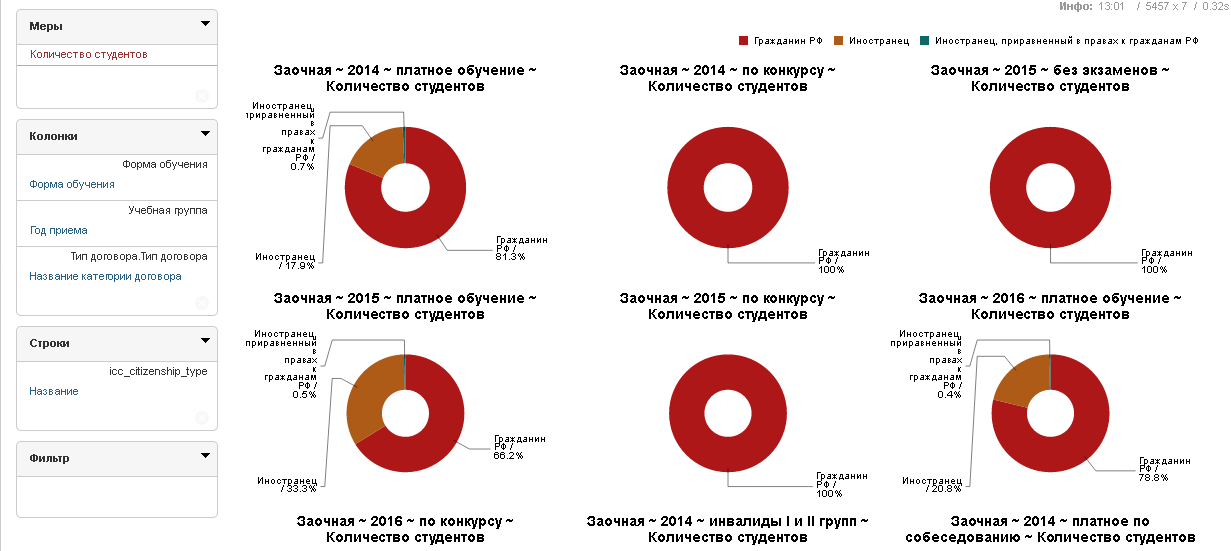


Рисунок 7 Круговая диаграмма



Рисунок 8 Основная статистика по выборке

В заключении хочется отметить, что OLAP-система является современным инструментом анализа и планирования деятельности крупной организации. Благодаря структурированию информации анализ выполняется оперативно, а отчёты формируются в различных срезах и с произвольной глубиной детализации.

Используемый в данном проекте OLAP-клиент Saiku имеет интуитивно понятный интерфейс, что позволяет рядовому пользователю работать с запросами без привлечения программистов.

**Литература**

1. «Методы и модели анализа данных: OLAP и Data Mining». Учебное пособие. А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. Издательство БХВ-Петербург, 2004г.

2. «Введение в OLAP-технологии Microsoft». Наталия Елманова, Алексей Фёдоров. Издательство Диалог-МИФИ, 2002г.