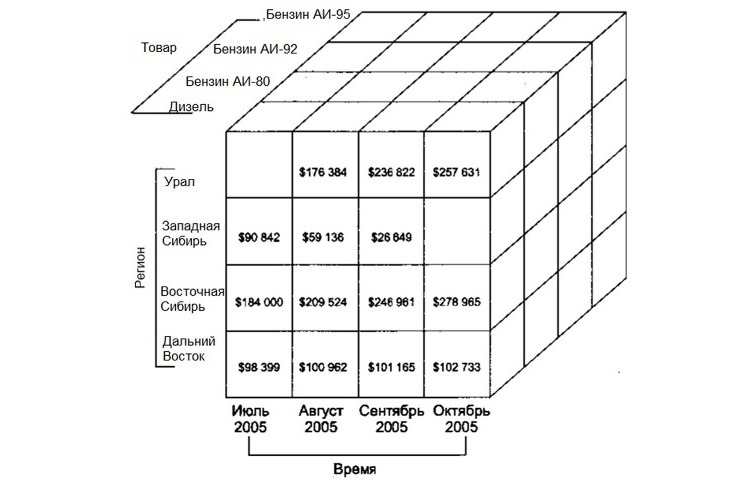
* OLAP-кубы

В рамках данной работы будут рассмотрены следующие вопросы:

* + Что представляют собой OLAP-кубы?
  + Что такое меры, измерения, иерархии?
  + Какие виды операций можно выполнять над OLAP-кубами?

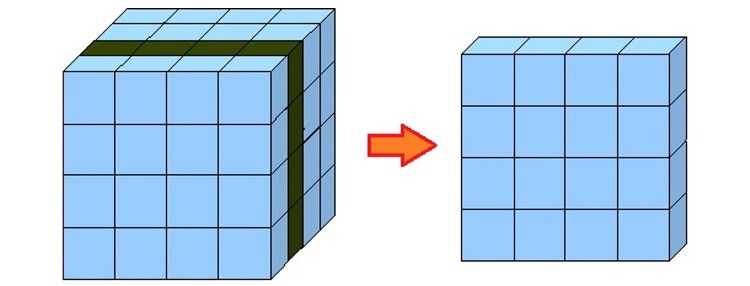
**Понятие OLAP-куба**  
  
Главный постулат OLAP – многомерность в представлении данных. В терминологии OLAP для описания многомерного дискретного пространства данных используется понятие куба, или гиперкуба.  
  
*Куб* представляет собой многомерную структуру данных, из которой пользователь-аналитик может запрашивать информацию. Кубы создаются из фактов и измерений.  
  
*Факты* – это данные об объектах и событиях в компании, которые будут подлежать анализу. Факты одного типа образуют меры (measures). Мера есть тип значения в ячейке куба.  
  
*Измерения* – это элементы данных, по которым производится анализ фактов. Коллекция таких элементов формирует атрибут измерения (например, дни недели могут образовать атрибут измерения "время"). В задачах бизнес-анализа коммерческих предприятий в качестве измерений часто выступают такие категории, как "время", "продажи", "товары", "клиенты", "сотрудники", "географическое местоположение". Измерения чаще всего являются иерархическими структурами, представляющими собой логические категории, по которым пользователь может анализировать фактические данные. Каждая иерархия может иметь один или несколько уровней. Так иерархия измерения "географическое местоположение" может включать уровни: "страна – область – город". В иерархии времени можно выделить, например, такую последовательность уровней: Измерение может иметь несколько иерархий (при этом каждая иерархия одного измерения должна иметь один и тот же ключевой атрибут таблицы измерений).  
  
Куб может содержать фактические данные из одной или нескольких таблиц фактов и чаще всего содержит несколько измерений. Любой конкретный куб обычно имеет конкретный направленный предмет анализа.  
  
На рисунке 1 показан пример куба, предназначенного для анализа продаж продуктов нефтепереработки некоторой компанией по регионам. Данный куб имеет три измерения (время, товар и регион) и одну меру (объем продаж, выраженный в денежном эквиваленте). Значения мер хранятся в соответствующих ячейках (cell) куба. Каждая ячейка уникально идентифицируется набором членов каждого из измерений, называемого кортежем. Например, ячейка, расположенная в нижнем левом углу куба (содержит значение 98399),задаетсякортежем[Июль2005,ДальнийВосток,Дизель].Здесьзначение98399 показывают объем продаж (в денежном выражении) дизеля на Дальнем Востоке за июль 2005 года.  
  
Стоит обратите также внимание на то, что некоторые ячейки не содержат никаких значений: эти ячейки пусты, потому что в таблице фактов не содержится данных для них.

* 

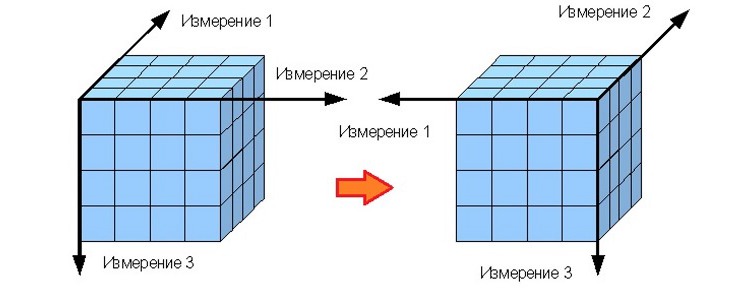
**Рис. 1.** Куб с информацией о продажах нефтепродуктов в различных регионах

* Конечной целью создания подобных кубов является минимизация времени обработки запросов, извлекающих требуемую информацию из фактических данных. Для реализации этой задачи кубы обычно содержат предварительно вычисленные итоговые данные, называемые *агрегациями*(aggregations). Т.е. куб охватывает пространство данных большее, чем фактическое – в нем существуют логические, вычисляемые точки. Вычислять значения точек в логическом пространстве на основе фактических значений позволяют функции агрегирования. Наиболее простыми функциями агрегирования являются SUM, MAX, MIN, COUNT. Так, например, используя функцию MAX, для приведенного в примере куба можно выявить, когда произошел пик продаж дизеля на Дальнем Востоке и т.д.  
    
  Еще одной специфической чертой многомерных кубов является сложность определения точки начала координат. Например, как задать точку 0 для измерения "Товар" или "Регионы"? Решением этой проблемы является внедрение специального атрибута, объединяющего все элементы измерения. Этот атрибут (создается автоматически) содержит всего один элемент – All ("Все"). Для простых функций агрегирования, например, суммы, элемент All эквивалентен сумме значений всех элементов фактического пространства данного измерения.  
    
  Важной концепцией многомерной модели данных является подпространство, или подкуб (sub cube). Подкуб представляет собой часть полного пространства куба в виде некоторой многомерной фигуры внутри куба. Так как многомерное пространство куба дискретно и ограничено, подкуб также дискретен и ограничен.  
    
  **Операции над OLAP-кубами**  
    
  Над OLAP-кубом могут выполняться следующие операции:
  + срез;
  + вращение;
  + консолидация;
  + детализация.

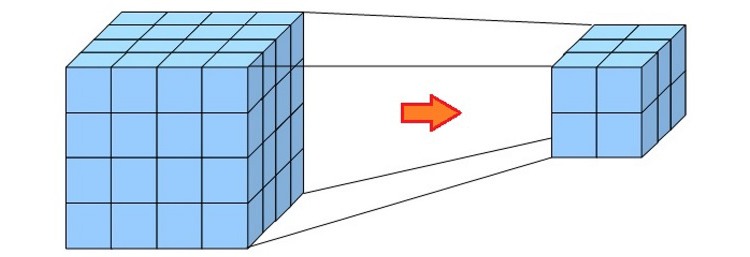
*Срез* (рисунок 2) является частным случаем подкуба. Это процедура формирования подмножество многомерного массива данных, соответствующее единственному значению одного или нескольких элементов измерений, не входящих в это подмножество. Например, чтобы узнать, как продвигались продажи нефтепродуктов во времени только в определенном регионе, а именно на Урале, то необходимо зафиксировать измерение "Товары" на элементе "Урал" и извлечь из куба соответствующее подмножество (подкуб).

* 

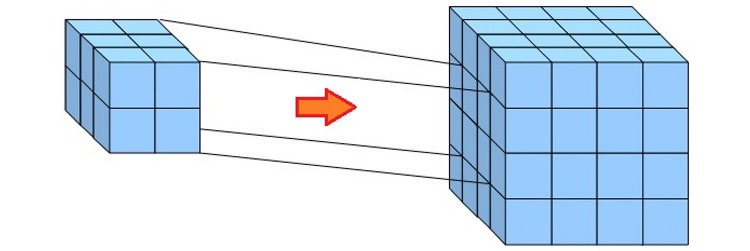
**Рис. 2.** Срез OLAP-куба

* *Вращение* (рисунок 3) – операция изменения расположения измерений, представленных в отчете или на отображаемой странице. Например, операция вращения может заключаться в перестановке местами строк и столбцов таблицы. Кроме того, вращением куба данных является перемещение внетабличных измерений на место измерений, представленных на отображаемой странице, и наоборот.
* 

**Рис. 3.** Вращение OLAP-куба

* *Консолидация* (рисунок 4) – операция перехода от детального представления данных к агрегированному. Например, переход к просмотру данных о продажах не по месяцам, а по годам.
* 

**Рис. 4.** Консолидация данных OLAP-куба

* *Детализация* (рисунок 5) – обратная консолидации операция, которая определяет переход от агрегированного представления данных к детальному.
* 

**Рис. 5.** Детализация данных OLAP-куба