

2. Авдеева, А.А. Контроль сжигания газообразного топлива [Текст] / А.А. Авдеева. – М.: Энергия, 1971. – 256 с.
3. Коллеров, Д.К. Метрологические основы газоаналитических измерений. (Теория и практика получения градуировочных и поверочных газов и газовых смесей) [Текст] / Д.К. Коллеров. – М.: Изд-во стандартов, 1967. – 396 с.
4. Коллеров, Д.К. Газоанализаторы. Проблемы практической метрологии [Текст] / Д.К. Коллеров. – М.: Изд-во стандартов, 1980. – 176 с.
5. Заец, Е.А. Статические погрешности промышленных хроматографов [Текст] / Е.А. Заец // Измерительная техника. – 1976. – № 9. – С.83-86.
6. Коллеров, Д.К. Организация метрологического контроля за газоанализаторами [Текст] / Д.К. Коллеров // Измерительная техника. – 1970. – №10. – С.70-74.
7. ГОСТ 13320-81 Газоанализаторы промышленные автоматические. Общие технические условия [Текст]. – М.: Изд-во стандартов, 1989. – 33 с.
8. Грязина, Л.И. Проблема хранения чистых газов и поверочных газовых смесей [Текст] / Л.И. Грязина // Измерительная техника. – 1975. – № 6. – С.82-85.
9. Теплюх, З.М. Принципи побудови високоточних дросельних синтезаторів газових сумішей / З.М. Теплюх [Текст] // Вісник Національного університету “Львівська політехніка” “Автоматика, вимірювання та керування”. – 2006. – №551. – С.87-94.
10. Теплюх, З.М. Пристрої для встановлення рівності опору дроселів синтезатора газових сумішей [Текст] / З.М. Теплюх, Є.П. Пістун, І.В. Ділай // Вимірювальна техніка та метрологія / Міжвідомчий науково-технічний збірник. – 2002. – №59. – С.78-81.
11. Ділай, І.В. Дросельні синтезатори газових сумішей заданого складу [Текст] / І.В. Ділай, З.М. Теплюх // Вісник Тернопільського державного технічного університету. – 2008. – № 3. – С.165-171.
12. Ділай, І.В. Підвищення точності газового аналізу в енергетиці [Текст] / І.В. Ділай, З.М. Теплюх // Збірник праць 5-ї міжнар. науково-практичної конференції “Проблеми економії енергії”. – Львів: НВФ “Українські технології”. – 2008. – С.22-31.
13. Ділай, І.В. Побудова подільників тиску для живлення газодинамічних дросельних синтезаторів [Текст] / І.В. Ділай, З.М. Теплюх // Вісник Національного університету “Львівська політехніка” “Теплоенергетика. Інженерія довілля. Автоматизація”. – 2009. – № 659. – С.120-128.

*Пропонується побудова сховища г/м даних з використанням OLAP-технологій. Розроблена структура багатовимірного сховища даних і пакети автоматичної обробки і передачі даних з оперативної бази даних в сховищі. На основі отриманого сховища реалізований гіперкуб з агрегацією даних спостережень по трьох вимірах*

*Ключові слова: OLAP-куб, гіперкуб, багатовимірні сховища даних, гідрометеорологічна галузь*

*Предлагается построение хранилища г/м данных с использованием OLAP-технологии. Разработана структура многомерного хранилища данных и пакеты автоматической обработки и передачи данных из оперативной базы данных в хранилище. На основе полученного хранилища реализован гиперкуб с агрегацией данных наблюдений по трем измерениям*

*Ключевые слова: OLAP-куб, гиперкуб, многомерные хранилища данных, гидрометеорологическая отрасль*

УДК 004.65:551.501

# ИНТЕЛЛЕКТУАЛЬНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ЗАДАЧАХ СОВЕРШЕНСТВОВАНИЯ ГИДРО- МЕТЕОРОЛОГИЧЕСКИХ СИСТЕМ

**М. Р. Валентюк**  
Ассистент\*

Кафедра информационных систем  
Севастопольский национальный технический  
университет  
ул. Университетская, 33, г. Севастополь, Украина,  
99000

E-mail: mvalent85@gmail.com

## 1. Введение

В гидрометеорологической (г/м) отрасли накоплен большой объем данных в виде измерений различных

параметров (температура воздуха, давление, скорость ветра и другие), необходимых для создания климатических моделей и прогнозирования. Ценность г/м данных заключается именно в наборах данных (а не

в отдельно взятых записях), по которым можно выявить закономерности, тенденции изменения погодных условий, а также зависимости между отдельными параметрами и измерениями. Именно поэтому в г/м отрасли необходимы не только средства ввода, редактирования и получения небольших выборок данных, но и эффективные средства аналитической обработки для:

- выявления сезонных и других периодических колебаний, тенденций, пиков измерений;
- сравнительного анализа измерений в отдельных точках в разные моменты времени, а также анализа зависимостей между наблюдаемыми параметрами и измерениями.

В настоящее время наиболее распространенными способами аналитической обработки и анализа данных, а также составления отчетов, являются средства Excel и механизм построения запросов к базам данных. Однако пользователю необходимо владеть знаниями и опытом работы с программным продуктом Excel, а для получения информации из базы данных нужно знать структуру и язык SQL, что не входит в «базовые» знания аналитиков г/м отрасли.

В последнее время всё большее распространение приобретают OLAP (OnLine Analytical Processing)-технологии, являющиеся наиболее эффективными при анализе больших объемов данных. Кроме того, работа с OLAP-системой проста и очевидна для пользователя [1].

Возникновение OLAP-технологии, как направления интеллектуальных информационных технологий, было вызвано потребностями крупных компаний в экономической области. В современной экономике конкурентные преимущества в значительной степени определяются интеллектуальным багажом компании, то есть знаниями, накопленными за многие годы [2]. В результате знания и имеющиеся данные служат основным компонентом потенциала компании. В г/м области работы, связанные с рассмотрением и использованием технологии OLAP для анализа данных в г/м записях, на территории бывшего СНГ не проводились. Причиной этому, возможно, является отсутствие всякой конкуренции в г/м отрасли.

## 2. Постановка задачи

Подсистема хранения г/м данных должна иметь двухуровневую архитектуру. На верхнем уровне располагается обобщенная информация для аналитиков г/м отрасли, которым требуются средства анализа данных. Нижний уровень соответствует источнику данных.

Анализ данных в хранилищах базируется на технологиях интеллектуального анализа данных (ИАД) [2]. ИАД реализуется в технологиях, одной из которых является технология интерактивной аналитической обработки данных OLAP.

Таким образом, для подсистемы хранения г/м данных необходимо создание двух баз данных: оперативной, которая является источником данных, и аналитической (OLAP). При этом оперативная база проектируется для быстрого внесения данных, а OLAP-база – для быстрого построения сложных отчетов [3].

Для синхронизации оперативной и OLAP баз данных необходимо обеспечить перенос данных и назначить расписание для этого процесса. Далее в аналитической базе строятся OLAP-кубы в соответствии с требованиями пользователей – аналитиков г/м отрасли.

## 3. Создание хранилища данных для г/м отрасли

Средством удобного анализа данных является многомерное хранилище данных и многомерные кубы, построенные на его основе [4]. Источником данных для хранилища в г/м отрасли служат реляционные базы данных (БД) (рис. 1). Они называются оперативными: с ними работают приложения пользователей, в реальном времени изменяющие данные. Оперативная г/м БД «Метео» включает в себя три таблицы:

- Point\_observ – таблица, содержащая в себе информацию о точках наблюдения. О каждой точке наблюдения заносится информация либо о названии, либо о ее координатах;

- Parameter – таблица, содержащая в себе информацию о видах наблюдений, например, температура воды, уровень моря, давление и другие. Для каждого измеряемого параметра заносится его название и единица измерения;

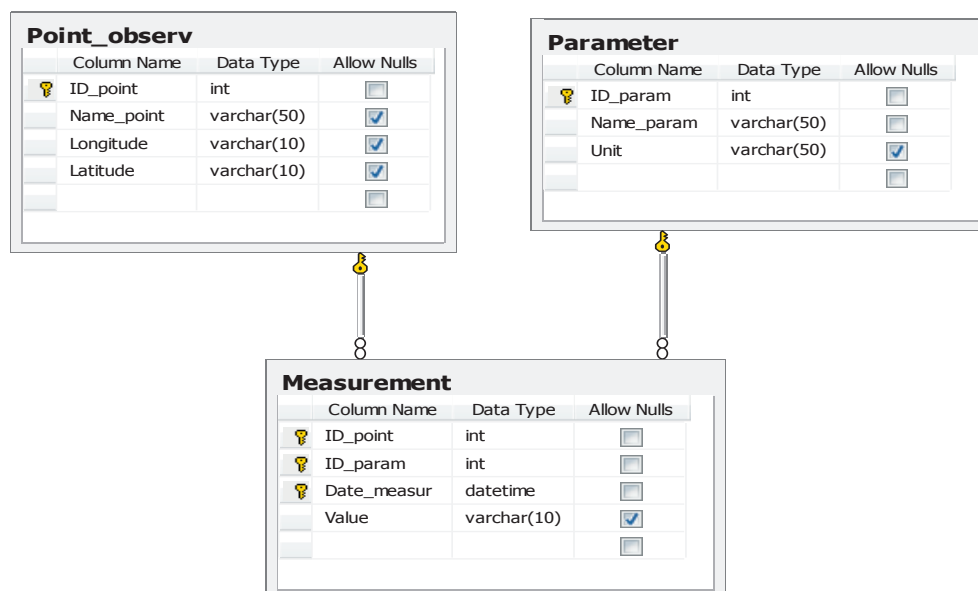


Рис. 1. Структура оперативной г/м БД «Метео»

– Measurement – таблица, содержащая информацию о наблюдениях, а именно где, когда и какой параметр получил значение, помещаемое в атрибут Value.

Оперативная БД спроектирована на минимальное время выполнения запроса на изменение данных.

Хранилище г/м данных «Meteo\_Mart» отличается от оперативной БД тем, что оно относительно стабильно – данные в нем изменяются согласно расписанию, например, ежедневно в определенное время. Процесс пополнения хранилища представляет собой добавление новых данных за определенный период без изменения уже имеющейся информации.

Структура хранилища данных существенно отличается от структуры оперативной БД тем, что она денормализована и содержит избыточность. Структура г/м хранилища приведена на рис. 2.

одной таблице, то такая структура хранилища носит название «звезда» [6].

Таблица фактов содержит сведения об объектах или событиях, совокупность которых будет в дальнейшем анализироваться. Эта таблица содержит уникальный первичный ключ, объединяющий первичные ключи таблиц измерений [5]. На основе структуры оперативной БД построена таблица фактов Measurement\_Fact, в которой первые три поля соответствуют измерениям OLAP – куба, а на основании поля Value будут получены агрегатные данные.

Использование агрегатных данных объясняется тем, что в большинстве случаев для анализа и прогнозирования используются не отдельные, а суммарные данные [7]. Для построения куба могут быть использованы различные виды агрегации – среднее (AVG), максимальное (MAX) и минимальное (MIN) значения.

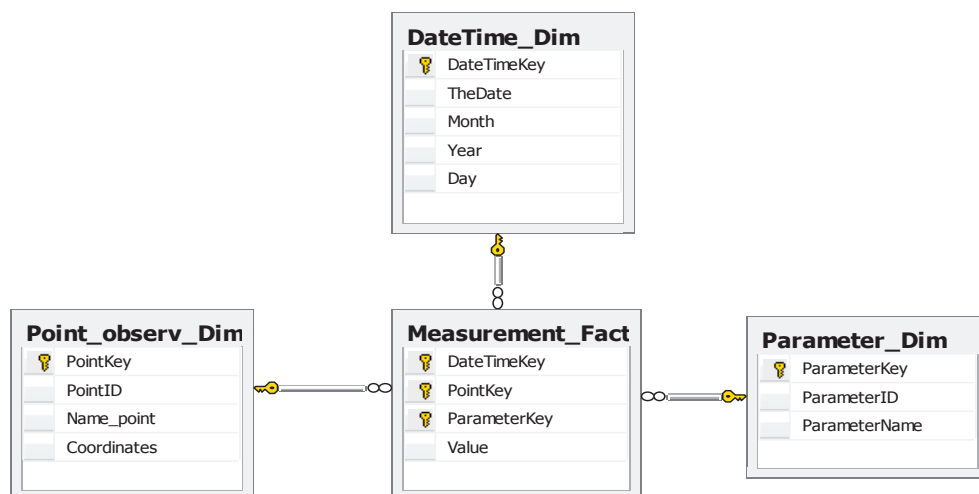


Рис. 2. Структура хранилища г/м данных Meteo\_Mart

#### 4. Заполнение хранилища данных с помощью Data Transformation Services

Для автоматического заполнения хранилища данными из оперативной БД был создан пакет службы Data Transformation Services (DTS) [8 – 9], содержащий описание последовательности всех действий, которые следует выполнить при переносе данных (рис. 3).

Основными составляющими хранилища являются таблицы измерений и фактов. Таблицы измерений содержат неизменяемые данные [5]. В приведенной структуре хранилища имеется три таблицы измерений: Date\_Time\_Dim, Point\_observ\_Dim и Parameter\_Dim. Каждая таблица измерений содержит ключевое поле с меткой Key, поле с меткой ID для идентификации записей с оперативной БД при пополнении хранилища, и одно или несколько описательных полей. Так как каждое измерение содержится в

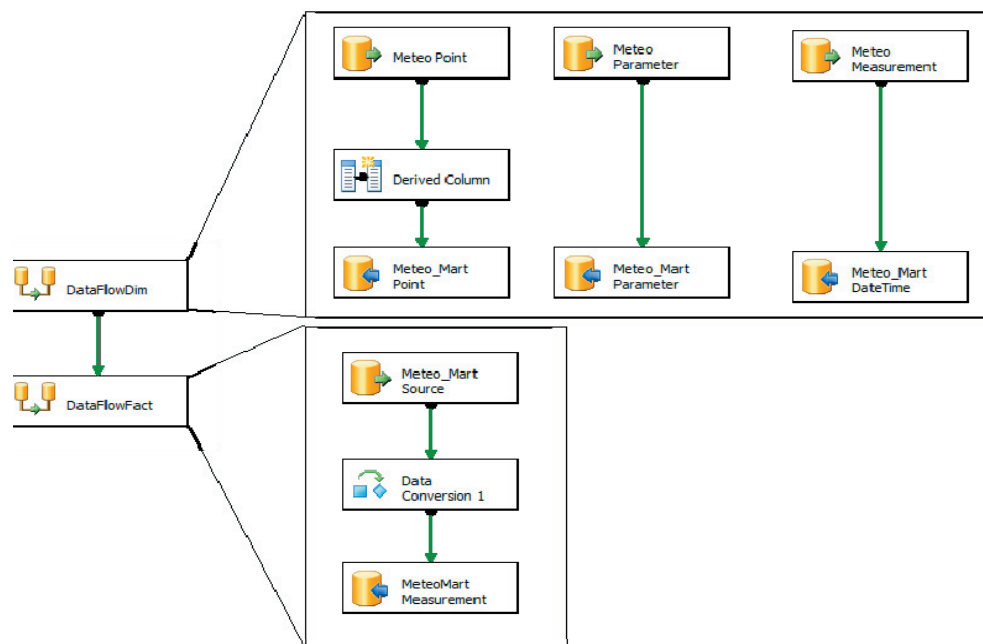


Рис. 3. Структура пакета DTS для заполнения хранилища Meteo\_Mart

Пакет, заполняющий хранилище г/м данных, состоит из двух обобщенных потоков данных: потока переноса данных в таблицы измерений и потока добавления новых данных в таблицу фактов. В первом потоке три задачи выполняются параллельно, так как таблицы измерений не связаны непосредственно друг с другом.

Такое выполнение задач существенно сокращает время пополнение данными хранилища.

Задача заполнения таблицы измерения «точка наблюдения» выполняется в три этапа:

1) выполняется выборка данных из таблицы «точка наблюдения» оперативной БД;

2) два столбца «широта» и «долгота» преобразовываются в один столбец «координаты»;

3) устанавливаются соответствия между столбцами таблицы хранилища и результатом выборки с учетом преобразования столбцов.

Заполняется таблица «точка наблюдения» хранилища данных.

Таблица измерений «Параметр» заполняется напрямую из таблицы «Параметр» оперативной БД без какой-либо предварительной обработки.

Источником данных таблицы измерений «ДатаВремя» является SQL-запрос:

```
select distinct
  Date_measur as TheDate, -- заполнение таблицы
  DateTime_Dim
  DatePart(ММ,Date_measur) as [Month],
  DatePart(YYYY,Date_measur) as [Year],
  DatePart(DD,Date_measur) as [Day]
from Measurement
```

Из рис. 3 видно, что задача заполнения данными главной таблицы фактов может быть выполнена только после того, как будут заполнены все таблицы измерений.

Исходный набор данных, преобразуемый в таблицу фактов, представляет собой результат следующего запроса:

```
select
  Meteo_Mart.dbo.DateTime_Dim.DateTimeKey,
  Meteo_Mart.dbo.Point_observ_Dim.PointKey,
  Meteo_Mart.dbo.Parameter_Dim.ParameterKey,
  Meteo.dbo.Measurement.Value
from
  Meteo.dbo.Measurement inner join Meteo_Mart.dbo.
  DateTime_Dim
  on Meteo.dbo.Measurement.Date_measur = Meteo_
  Mart.dbo.DateTime_Dim.TheDate
  inner join Meteo_Mart.dbo.Parameter_Dim
  on Meteo.dbo.Measurement.ID_param =Meteo_Mart.
  dbo.Parameter_Dim.ParameterID
  inner join Meteo_Mart.dbo.Point_observ_Dim
  on Meteo.dbo.Measurement.ID_point = Meteo_Mart.
  dbo.Point_observ_Dim.PointID
```

Так как для создания агрегатных данных поле должно быть числового типа, то после выполнения запроса происходит преобразование типов данных – преобразование исходного поля Value строкового типа в вещественный тип.

Для того, чтобы данные в хранилище соответствовали состоянию оперативной БД, было создано расписание, согласно которому созданный пакет будет выполняться автоматически каждый день в 23:00.

## 5. Создание OLAP-кубов на основании г/м хранилища данных

Созданное хранилище данных является источником для создания OLAP-кубов. Оси куба называются измерениями и содержат параметры, по которым необходимо произвести анализ [5].

В качестве одного из измерений обычно используют время.

Источником данных для формирования измерений куба служат таблицы измерений г/м хранилища. На пересечениях осей – измерений находятся данные, количественно характеризующие процесс – меры. Для вычисления мер куба используется таблица фактов г/м хранилища.

Работник г/м отрасли, анализирующий информацию, может «разрезать» куб по разным направлениям, получать сводные или, наоборот, детальные сведения и осуществлять другие манипуляции с кубом.

В качестве мер в трехмерном кубе, изображенном на рис. 4, использованы максимальные значения параметров наблюдений, а в качестве измерений – год, точка наблюдения (координаты) и параметр наблюдения.

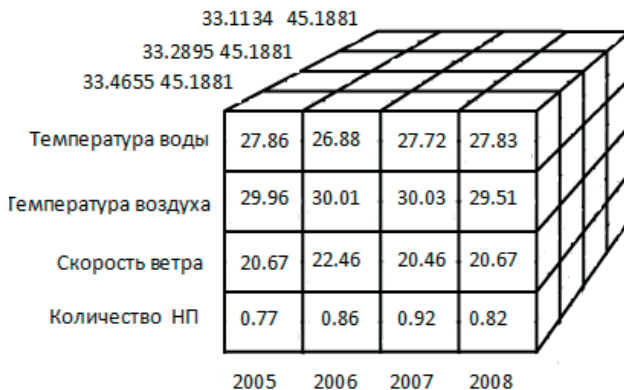


Рис. 4. Пример куба с тремя измерениями

## 6. Выводы

Задачей работников г/м отрасли, которые занимаются анализом данных и прогнозированием, является нахождение закономерностей в больших массивах данных.

Поставленная задача была решена с использованием службы MS SQL Server'a (рис. 5) и технологии OLAP. Для увеличения скорости предоставления необходимой для аналитиков информации было разработано хранилище г/м данных (компонент Database Engine), созданы пакеты автозаполнения хранилища информацией из оперативной БД (Integration Services), а также продемонстрирована возможность построения OLAP-кубов (Analysis Services).

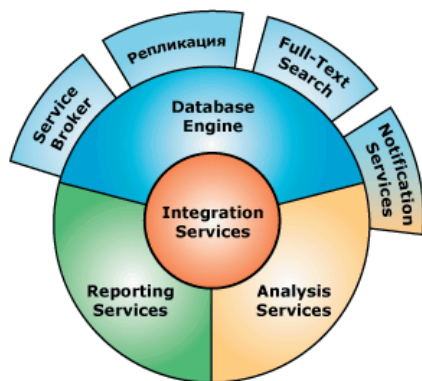


Рис. 5. Связи и взаимодействия между интерфейсами компонентов

Достоинствами применения OLAP-технологии по сравнению с другими средствами аналитики являются:

- быстрый отклик системы, чтобы аналитик «не терял мысль» (не более 5 секунд),
- полнофункциональный анализ максимально удобным способом;
- многопользовательский доступ, защита информации;
- многомерное концептуальное представление данных в виде кубов с иерархическими измерениями;
- получение информации в нужном объеме, без избыточности [10].

В дальнейшем планируется произвести анализ выборок данных, необходимых г/м работникам для анализа и прогнозирования с целью создания библиотеки многомерных OLAP кубов по различным измерениям.

### Литература

1. Некоторые соображения о роли и месте технологии OLAP [Электронный ресурс] / Сайт Olap.ru. – Режим доступа: [www/ URL: http://www.olar.ru/basic/think.asp](http://www.olar.ru/basic/think.asp) – 07.04.2012 г. – Загл. с экрана.
2. Башмаков, А. И. Интеллектуальные информационные технологии [Текст]: учеб. пособие / А. И. Башмаков, И. А. Башмаков. – М.: Изд-во МГТУ им. Н.Э.Баумана, 2005. – 304 с.
3. Creating and Manipulating Multidimensional Tables with Locational Data Using OLAP Cubes [Электронный ресурс] / Сайт Directions Magazine. – Режим доступа: [www/ URL: http://www.directionsmag.com/articles/creating-and-manipulating-multidimensional-tables-with-locational-data-usin/123565](http://www.directionsmag.com/articles/creating-and-manipulating-multidimensional-tables-with-locational-data-usin/123565) – 06.04.2012 г. – Загл. с экрана.
4. Иванова, А. Аналитическая обработка данных [Текст] / А. Иванова // Журнал БУТЕ Россия. – 2006. – №2 (90). – С. 46-49.
5. Полубояров, В. В. Использование MS SQL Server Analysis Services 2008 для построения хранилищ данных [Текст] / В. В. Полубояров. – М.: Интуит, 2010. – 487 с.
6. Введение в OLAP [Электронный ресурс] / Российский ежемесячный компьютерный журнал КомпьютерПресс. – Режим доступа: [WWW/ URL: http://www.compress.ru/article.aspx?id=10348&iid=425](http://www.compress.ru/article.aspx?id=10348&iid=425) – 10.03.2012 г. – Загл. с экрана.
7. Создаем OLAP куб [Электронный ресурс] / Сайт Хабрахабр компании «Тематические медиа». – Режим доступа: [WWW/ URL: http://habrahabr.ru/post/67731/](http://habrahabr.ru/post/67731/) – 05.04.2012 г. – Загл. с экрана.
8. Учебник по службам SSIS [Электронный ресурс] / Библиотека Microsoft Developer Network. – Режим доступа: [WWW/ URL: http://msdn.microsoft.com/ru-ru/library/ms169917%28v=sql.105%29.aspx](http://msdn.microsoft.com/ru-ru/library/ms169917%28v=sql.105%29.aspx) – 16.03.2012 г. – Загл. с экрана.
9. Create OLAP cube for SQL Server 2005 Analysis Services [Электронный ресурс] / Сайт blrf.net. – Режим доступа: [www/ URL: http://www.blrf.net/blog/74/microsoft/create-olap-cube-for-sql-server-2005-analysis-services/](http://www.blrf.net/blog/74/microsoft/create-olap-cube-for-sql-server-2005-analysis-services/) – 14.04.2012 г. – Загл. с экрана.
10. Smart Business Intelligence Solutions with Microsoft® SQL Server® 2008 [Электронный ресурс] / Сайт Safari Books Online. – Режим доступа: [www/ URL: http://my.safaribooksonline.com/book/databases/microsoft-sql-server/9780735625808](http://my.safaribooksonline.com/book/databases/microsoft-sql-server/9780735625808) – 07.04.2012 г. – Загл. с экрана.