Data Cube Design using SQL Server

Amna A.Hanni Al Safar
Iraqi Commission for
Computer and Informatics

Dr. Moaid A.Fadhil
Informatics Institute
for Postgraduate Studies

Abstract

When we try to extract information from a database, we need tools to help us find what's relevant and what's important and to explore different scenarios. A report, whether printed on paper or viewed on-screen, is at best a two-dimensional representation of data, a table using columns and rows. That's sufficient when we have only two factors to consider, but in the real world we need more powerful tools. On-Line Analytical Processing (OLAP) is a category of software technology that enables analysts, managers and executives to gain insight into data through fast, consistent, interactive access to a wide variety of possible views of information that has been transformed from raw data to reflect the real dimensionality of the enterprise as understood by the user. An OLAP system can be modeled by a three-level architecture that consists of a Query Client, a Data Cube engine and a Data Warehouse or Relational DB.

In this paper we introduce a Client Graphical Application Interface to Design & Create Local Cube from relational database and to connect to SQL Analysis Service OLAP Cube Server to make complex queries, switch from summary to detail data, and filter or slice data into meaningful subsets.

Key words: OLAP, Data Cube, SQL Server, Analysis Service;

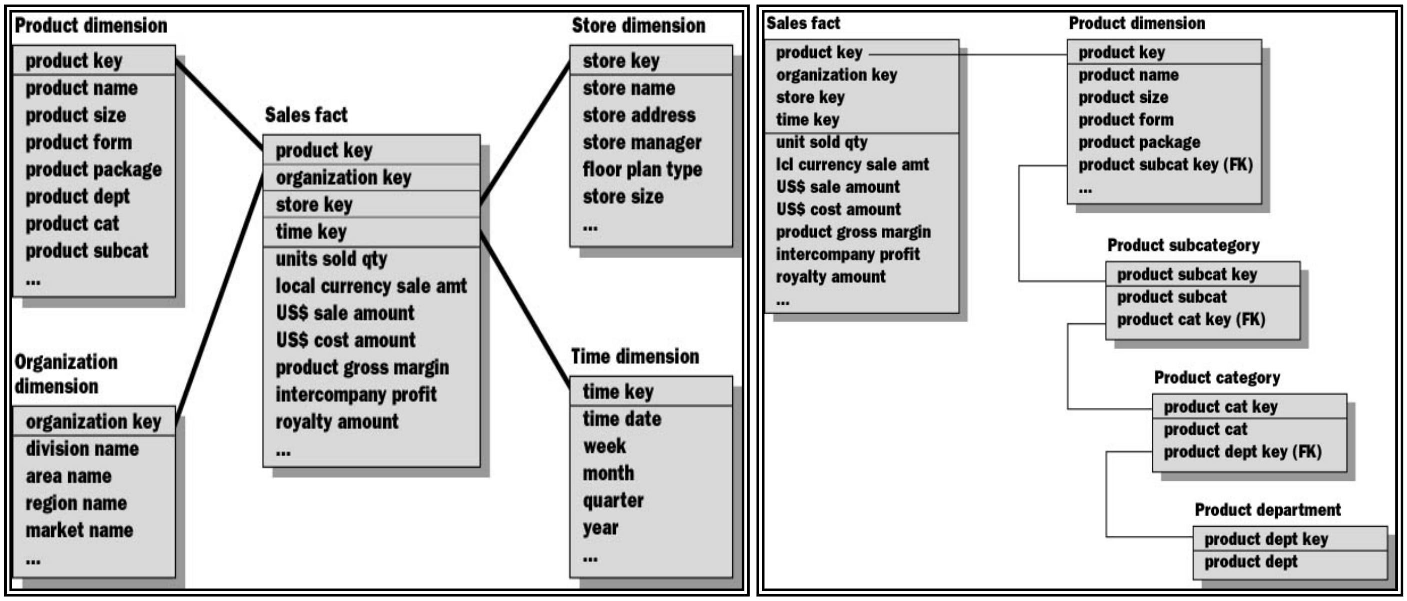
1. Introduction

On-line analytical processing (OLAP) is a term that was coined in an unpublished 1993 white paper, *Providing OLAP to User Analysts: An IT Mandate*, by E. F.Codd [1] the paper signaled a shift in the paradigm for business analysis. Instead of reviewing piles of static reports printed on green-bar paper, the OLAP analyst could explore business results interactively, dynamically adjusting the view of the data—asking questions and getting answers almost immediately. This freedom from static answers to fixed questions on a fixed schedule allows business analysts to operate more effectively and to effect improvements in business operations. Nigel Pendse, an analyst with Business Intelligence Ltd. who publishes *The OLAP Report*, provides his point of view in a Web page entitled “What Is OLAP?”, Pendse introduces a simpler model, FASMI (Fast, Analytical, Shared, Multidimensional), to characterize OLAP systems [2, 3]:

- ♣ Fast. In keeping with the spirit of the “O” in OLAP, such systems need to provide results very quickly usually in just a few seconds, and seldom in more than 20 or 30 seconds. This level of performance is key in allowing analysts to work effectively without distraction.
- ♣ Analytic. Considering the “A” in OLAP, OLAP systems generally must provide rich analytic functions appropriate to a given application, with minimal programming.
- ♣ Shared. An OLAP system is usually a shared resource. This means that there is a requirement for OLAP systems to provide appropriate security and integrity features. Ultimately, this can mean providing different access controls on each cell of a database.
- ♣ Multidimensional. Multidimensionality is the primary requirement for an OLAP system. OLAP products present their data in a multidimensional framework. Dimensions are collections of related identifiers, or attributes (product, market, and time, for example) of the data values of the system. The identifiers belonging to the collection for a particular dimension generally have some sort of structure—usually hierarchical. Sometimes there is more than one natural structure for these identifiers. The multidimensional characteristic means that an OLAP system can quickly switch among various orientations of dimensions, as well as among various subsets and structural arrangements of a dimension. Because of the multidimensional nature of OLAP systems, the users often refer to the collections of data that they implement as ***cubes***.
- ♣ Store and calculate information. OLAP systems must store and calculate information. Data for OLAP systems often come from one or more operational systems. Analytical models are applied to these data, and the results must either be stored in the system or generated at query time.

2. OLAP Data Cube basics

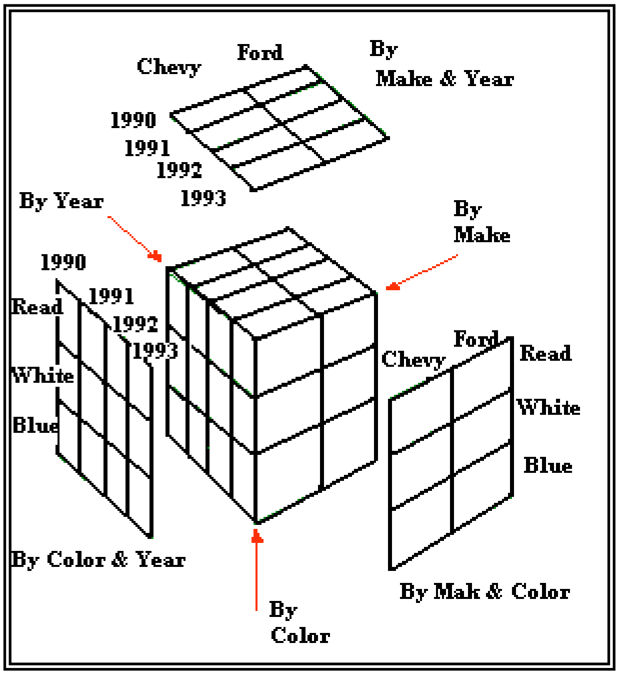
In an OLAP data model, information is viewed conceptually as ***cubes*** that consist of descriptive categories (***dimensions***) and quantitative values (***measures***) and the ***Fact table***, which relate the dimension to the measures of interest. A multidimensional data model is typically organized around a central theme, like sales, for instance. This theme is represented by a fact table. Facts are numerical measures. Examples of facts for a sales include dollars_sold(sales amount in dollars), units_sold(number of units sold) and amount_budgeted. The fact table contains the names of the facts, or measures, as well as keys to each of the related dimension tables. Such a model can exist in the form of a ***star schema*** (A relational database structure in which data is maintained in a single fact table at the center of the schema with additional dimension data stored in dimension tables. Each dimension table is directly related to the fact table by a key column), a ***snowflake schema*** (An extension of a star schema such that one or more dimensions are defined by multiple tables. In a snowflake schema, only primary dimension tables are joined to the fact table. Additional dimension tables are joined to primary dimension tables) see Fig(1). Within each dimension of an OLAP data model, data can be organized into a ***hierarchy*** that represents ***levels*** of detail on the data. For example, within the time dimension, it may have the levels years, months, and days. A particular instance of the OLAP data model would have the specific values for each level in the hierarchy [4, 5, 6].



Fig(1) Star Schema, snowflake schema

Data cubes are multidimensional structures that store the data for OLAP system. In short it’s described as a data abstraction that allows one to view aggregated data from a number of perspectives. Conceptually, the cube consists of a core or base cuboid, surrounded by a collection of sub-cubes/cuboids that represent the aggregation of the base cuboid along one or more dimensions. The dimension to be aggregated refers as the measure attribute, while the remaining dimensions are known as the feature attributes. Fig(2) gives a small, practical data cube example from the car industry. This particular data cube has three feature attributes (make, color, and year) and single measures attribute (sales, sales is computed with the sum function). By selecting cells, planes, or subcubes from the base cuboid, one can analyze sales figures at varying granularities from any such view, OLAP systems typically provide operations that allow an analyst to quickly switch to a related view. Operations typically provided include [4, 6]:

- ♣ DRILL DOWN/UP—drilling down or up is a specific analytical technique whereby the user navigates among levels of data ranging from the most summarized (up) to the most detailed (down). The drilling paths may be defined by the hierarchies within dimensions or other relationships that may be dynamic within or between dimensions.



Fig(2) car industry data cube

- ♣ SLICE AND DICE—the user-initiated process of navigating by calling for page displays interactively, through the specification of slices via rotations and drill down/up.
- ♣ Pivot—rearrange the ordering of dimensions in a result.

An aggregate is a table or structure containing pre-calculated data. Aggregations support rapid and efficient querying of a multidimensional database. Aggregations improve query response time by having the answers ready before the questions are asked. Aggregates must conform to rules derived from the hierarchical structure of all dimensions and the granularity of the base fact table [5,7].

3. Proposed system

Microsoft SQL Server 2000 ***Analysis Services*** is a middle-tier server for (OLAP). Analysis Services organizes data into cubes with pre-calculated aggregation data to provide rapid answers to complex analytical queries. To make OLAP easier to use, Microsoft SQL Server Analysis Services provides Analysis Manager; this console application provides a user interface for accessing analysis servers and their Meta data repositories. The Analysis Manager which is an administers tools to defines, modifies measures and dimensions design, build, process, and view cubes, from simple to advance Server Cubes, that is, under the Analysis Manager Tree pane.

The proposed system is a Client's tool, it helps the user to build his own cubes (Local Cube) from relational database (such as Microsoft Access or SQL) and also connect to per-built cubes by database administrator and stored in Analysis Server. After connecting to a data cube it's possible to asking question and getting answer using MDX (multi dimensional expression) then import the result to Excel sheet which allows the user to deliver nice looking graph moreover the data are exactly in the location that most business analysts appreciate in the Excel environment. No more hassling of data is needed with difficult imports from OLAP to Excel. Also it's possible after connection to send the data cube hierarchy to Word Document for farther archiving. Fig(3) shows architecture of the main functions with their activates. The algorithm to create Local cube with db provider is:

```
Input       ADODB connection to Database file
Output      Local Cube file
------------------------------------------------------------
Step 1      Create an ADODB connection to DB
Step 2      Choose the Fact Table from DB tables list
Step 3      List Fact table numerical fields (int., money, real, float…)
Step 4      Select Measures from Fact table numerical fields list
Step 5      Choose the Aggregations function for each Measure
Steps 6     Dim.(s) Choosing Loop
        6.1    Choose dim. Schema Star or Snowflake
        6.2    Choose the cube Dimension(s) from db tables
        6.3    Connect Fact table with Dim. (Set Where Relations)
        6.4    Choose Dim. Type
        6.4.1    Regular Dim.
```

6.4.1.1 List all dim Fields
6.4.1.2 Choose dimension Level from dim. Fields
6.4.2 Time Dim.
6.4.2.1 List only date type fields
6.4.2.2 Choose the Time hierarchy from prepared list
6.5 Name the dimension and accept it
6.6 Next Dim.
Step 7 Choose Cube Storage Mode
Step 8 Name and Locate cube Disk File (save File)
Step 9 Named and Create cube (Process)
Step 10 On Error goto Error handler

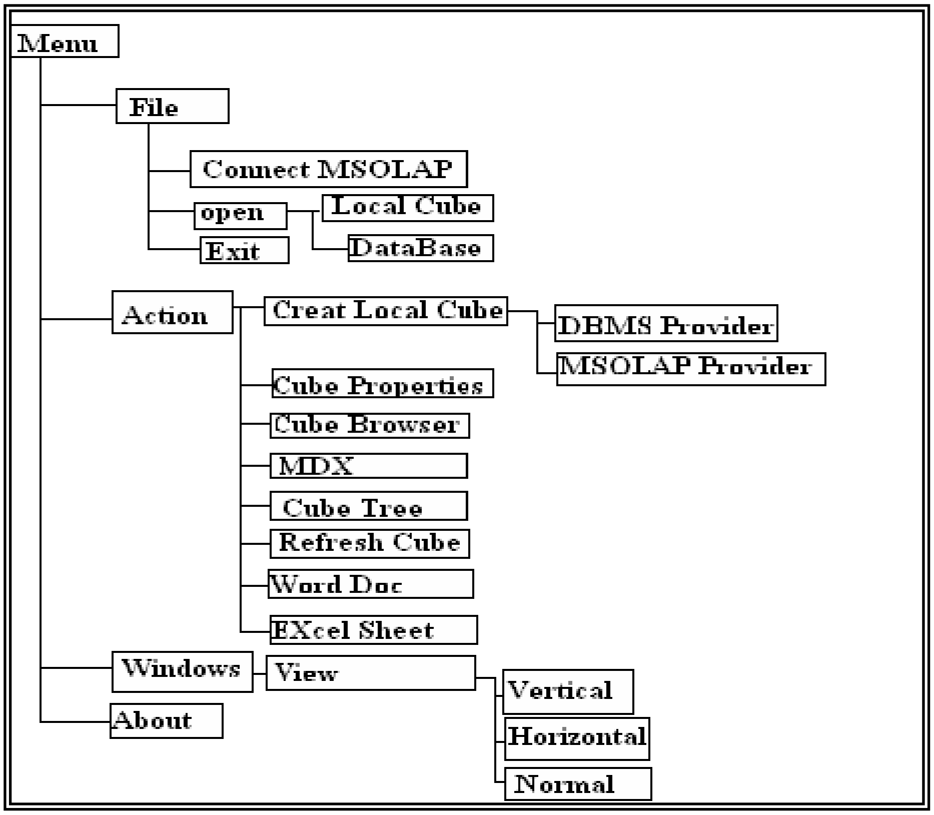


Fig (3) Main menu architecture

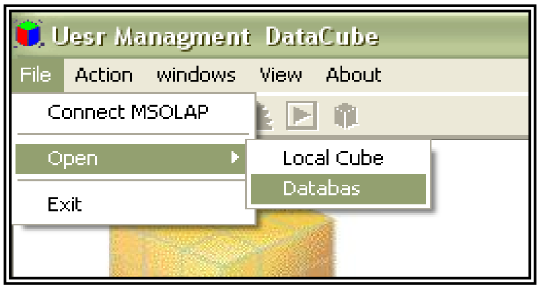
With our proposed system, the user can choose to build his own cubes from a relational DB such as SQL or from pre-built cubes by the database administrator that stored in Microsoft SQL Analysis Server, or he can connect to Analysis Server to pivoting and drilling through cube data. With self-built cubes, first the user defines the cube structure and then process the cube, processing cubes can be very time consuming. Smaller cubes can be built in a matter of minutes, but if the source data has millions of rows or the cube has many measures and dimensions with many hierarchy levels, loading and calculating the data in the cube could take hours. The size and complexity of the cubes the regular user built with our system will be limited by the power of the user personal workstation and his patience.

A local cube is stored in a single, portable file (default extension is ".cub"). Local data cubes are stored in files that are the local equivalent of Microsoft OLAP Server objects that can be stored on both server and non-server computers. End users can browse local cubes without a connection to an Analysis server; Local cubes are the only variety of cube that provides this capability.

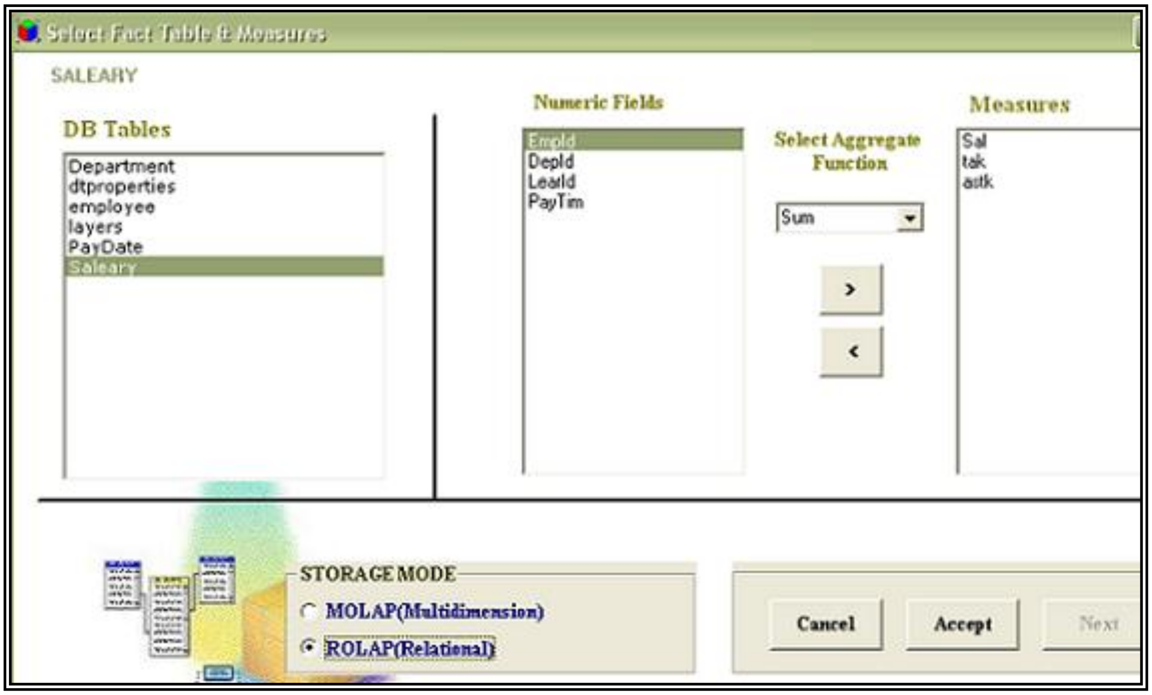
PivotTable Service is the primary interface for applications interacting with Analysis Services. It is used to build client applications that interact with multidimensional data. All client applications must communicate with the Analysis Server through a client process named ***PivotTable service***, which must reside on the client computer. PivotTable service is automatic installed on computer when installing Analysis services (server or client components) or when installing office 2000, the proposed system uses OLE DB (object link Embedding Database) for OLAP to communicate with PivotTable service. OLE DB for OLAP is a set of methods added to OLE DB to provide efficient access to multidimensional data. To assist programmers develop corporate application, OLE DB for OLAP comes with a wrapper layer called Microsoft ActiveX Data Object (multidimensional), or ADO MD, which is simpler set of instructions commonly used by tools such as Microsoft Visual Basic. The system supports Multidimensional Expression (MDX) functions as well as a full language implementation for building local cubes (also referred to as offline cubes) and querying cube data using PivotTable Service with OLE DB and (ADO). A local cube is a cube constructed on the client workstation, as opposed to a cube constructed on an Analysis server.

4. Example: Creating a Local data cube

4.1 Connecting and selecting the Provider: before creating the data cube you must connect to data provider (Relational DBMS, MSOLAOP) from the menu open your db or connect to MSOLAOP.



4.2 Begin Creating Cube : menu<Action<createLocalCube<DBMS provider, then your opened DB tables will be shows to select the Fact table from the list and then select the Measures from the Fact Table numerical fields and there function.



SALEARY

DB Tables: Department, dtproperties, employee, layers, PayDate, Saleary

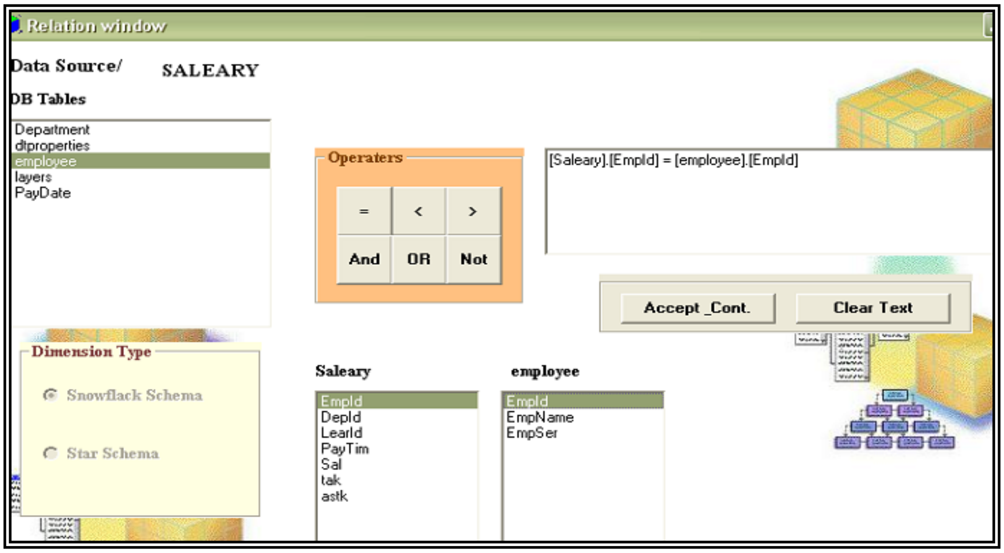
Numeric Fields: EmpId, DepId, LearId, PayTim

Select Aggregate Function: Sum

Measures: Sal, tak, astk

STORAGE MODE: ○ MOLAP(Multidimension) ● ROLAP(Relational)

4.3 Set the Relation: between the Fact Table and the dimensions tables which was select from the DB Table lists in the Relation window and set the type of dimension (star, snowflake). Then in Dimension Window select the dimension Levels hierarchy (Time, Regular) name the dimension and accept it to select another dimension or continue as needed.



SALEARY

DB Tables: Department, dtproperties, employee, layers, PayDate

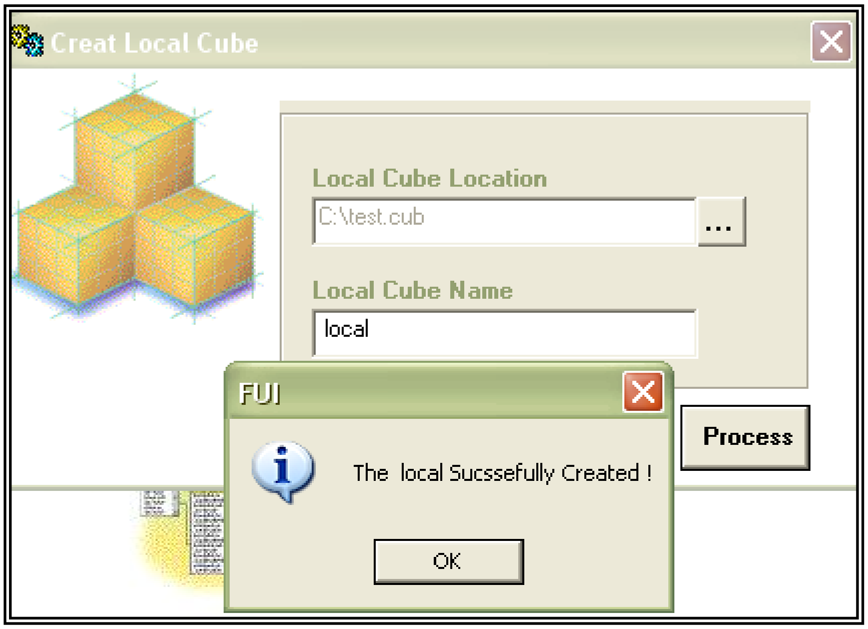
[Saleary].[EmpId] = [employee].[EmpId]

Dimension Type: ● Snowflack Schema ○ Star Schema

Saleary: EmpId, DepId, LearId, PayTim, Sal, tak, astk

employee: EmpId, EmpName, EmpSer

4.4 Process the cube: name and locate the cube in a file and begin the process to create it, you will resave a message in both success and Error.

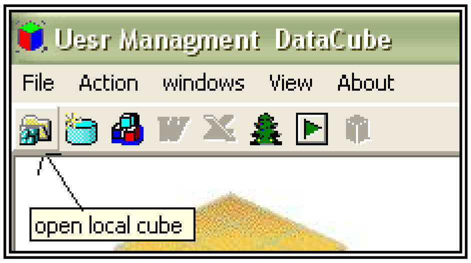


Local Cube Location: C:\test.cub

Local Cube Name: local

The local Sucssefully Created !

4.5 Open it: open the Local cube from menu< open< Localcube or from Toolbar.



5. Conclusions

1. Frees data cube from constantly creating ad hoc reports and data summaries. Because summarized data is organized along business lines, users can get the reports they need dynamically.
2. Users can take their Local cubes with them on their laptops with there own different names for the parts of the cube.
3. Local cube are available even if there isn't a good connections to the Analysis Server.
4. Using local cubes is useful for the user and Network, providing good access, organization and security for the user and improves Network traffic.

Reference

[1] E. F. Codd, Providing OLAP to User Analysts: An IT Mandate, unpublished manuscript, E. F. Codd and Associates (1993). This paper may be obtained from Hyperion Solutions Corporation at their Web site, http://www.hyperion.com.

[2] N. Pendse, “What Is OLAP?”. The OLAP Report, http://www.olapreport.com/fasmi.htm, 2003

[3] N.Colossi, W.Malloy, B.Reinwald, “Relational Extensions for OLAP” IBM Systems journal, NO 4, 2002.

[4] OLAP AND OLAP Server Definitions. www.dssresources.com/glossary/olaptrms.html, 2002

[5] “Microsoft SQL Server 7.0 Data Warehousing Training Kit”, Copyright Microsoft Corporation 2000.

[6] Faculty of Computer Science,”The Data Cube”, University Avenue Halifax NS Canada, www.cs.dal.ca, 2003-2004

[7] Hari Mailvaganam, “Designing OLAP solutions”, Data Warehousing Review copyright 2004. www.DWreview.com

تصميم مكعب بيانات باستخدام خادم لغة الاستفسار المهيكلة

د.مؤيد عبد الرزاق فاضل
معهد المعلوماتية للدراسات العليا

آمنة عبد الرزاق هاني الصفار
الهيئة العراقية للحاسبات والمعلوماتية

الخلاصة

لاستخلاص المعلومات من قاعدة البيانات نحتاج الى ادوات لتساعدنا في ايجاد المهم والمناسب منها وبعدة اساليب. فالتقارير تكون اما مطبوعة او معروضة على شاشة الحاسوب وعلى شكل صفوف واعدة كجدول ثنائي الابعاد. هذه الصيغة قد تكون كافية لعرض معلومات تقتصر على معاملين فقط الا ان العالم الحقيقي يحتاج الى ادوات اكثر قابلية.

المعالجة التحليلية الفورية هي نوع من انواع البرمجيات التي تمكن كل من المحللين والمدراء والمستخدمين من كسب بعد نظر على بياناتهم من خلال سرعة، تناغم، تفاعلية وصول الى عدد كبير من امكانية عرض المعلومات التي تم نقلها من صف من البيانات لتعكس حقيقة ابعاد العمل كما يفهمه المستخدم.

نظام (OLAP) ممكن تصميمه بثلاث مستويات من المعمارية وهي استفسارات الزبائن، مكعب البيانات ومستودع البيانات او قواعد البيانات العلائقية.

في هذه الورقة نقدم واجهة تطبيق رسومية للمستخدمين لتصميم وبناء مكعب بيانات محلي من قواعد بيانات علائقية والاتصال بخدمات التحليل الخاص بخادم لغة الاستفسار المهيكلة، لتنفيذ الاستفسارات المعقدة والابدال من العرض الاجمالي الى العرض التفصيلي للبيانات او عرض اجزاء ذات مغزى منها.