Homework Assignment 3

IT 532, Fall 2018

Due Date – Oct 17

1. Question 3.2, 3.4 and 3.6 (page 85)

3.2 Using an example of an application domain that you are familiar with,  
describe the various components of the multidimensional model, that  
is, facts, measures, dimensions, and hierarchies.

Aswer: Take insurance business process as example, the E-R model can be composed of customer,policy,claim and employee. One customer may have many policies, and they may use one policy to claim many times, the employee will deal the claim through documents and processes.

When we want to transfer this ER model to multidimensional model. Firstly, let us think about business process and user requirement, what is the main purpose of the insurance data warhouse? Obviously, we want to know the claim information. So we can separate the claim amount and claim paid as out fact table.Secondly, what are the dimensions we need to use to analyze the data? In this case, we can separate claim OLTP information to become a new dimension table, and let customer, policy, employee become other dimension tables. Finally, how about other useful dimensions, especially we want some high level analysis ,we can consider to create two hierarchy, one for calendar, another for place, calendar is composed of year,quarter,month,date. On the other hand, place is made of country,state, city.

Sum up the main point, our multidimensional model will be as below:

|  |  |
| --- | --- |
| Facts | * claim-information (includes claim amount,paid amount as measures) |
| Dimensions | * claim * customer * policy * employee |
| Hierarchy | * Calendar(year,quarter,month,date) * Place(country,state, city) |

3.4 Discuss the role of measure aggregation in a data warehouse. How can  
measures be characterized?

Answer: Measures are the content of fact tables, which are the ‘fact’ we want to know through traversing the different level of dimensions. How can we get the measures, we need to aggregate the number from OLTP.

Measures can be categoried in 3 aspects. The first one is by if the measure is meaninful using some aggregation function. By using this categorization method, measures can be categorized into additive,semi-additive and nonadditive. The second one is by if the aggregation function is distributed, for example, sum,count,max,min are distributable, but distinct count is not. The third one is by if the aggregation is derivable, for example, average can be get from sum/count , but median and mode can’t derived from other function.

3.6 Describe the various OLAP operations using the example you defined  
in Question 3.2.

Answer:

|  |  |
| --- | --- |
| original cube | place 🡪state ; calendar🡪quarter |
| find the total claim amount by country | ROLLUP(claim-information, place→ country, sum(claim\_amount)) |
| drill down the claim amount in calendar by month | DRILLDOWN(claim-information,calendar→ month) |
| see data focus on customer dimension , and sort by customer\_name | SORT(claim-information, customer, customer\_name ASC) |
| change the point of view to employee | PIVOT(claim-information, employee🡪x,customer🡪y,policy🡪Z,claim🡪X1,place🡪X2,calendar🡪X3 |
| find the total claim amount for specific country | ROLLUP(claim-information, place→ country, sum(claim\_amount))  SLICE(claim-information, location, country= US) |
| see data in specific criteria | dice(claim-information, location.country=’US’ AND calendar.year=’2018’) |
| Compare the claim amount between 2018 and 2017 | ROLLUP(claim-information, calendar→ year, sum(claim\_amount))  SLICE(claim-information, calendar, year= 2018)  Rename(claim-information, claim-information 🡪claim-information-2018,claim-amount🡪 claim-amount-2018)  ROLLUP(claim-information, calendar→ year, sum(claim\_amount))  SLICE(claim-information, calendar, year= 2017)  Rename(claim-information, claim-information 🡪claim-information-2017,claim-amount🡪 claim-amount-2017)  DRILLACROSS(claim-information-2018, claim-information-2017) |
| Sum all claim amout | SUM(claim-information, claim-amout) BY Time |

1. Download and import data to SQL server(database: AdventureWorksDW2012). Create an Analysis Services multidimensional project using SQL Server Data Tools, create a Data Source View, imported the necessary tables into it,  create necessary dimension tables and fact table, deploy this project CUBE. Use Excel to calculate sales amount of black product ordered in quarter 1 and quarter 2,  the sales amount only includes two cities in California: Berkeley, Beverly Hills

Answer:

|  |  |  |  |
| --- | --- | --- | --- |
| **Sales Amount** | **Column Labels** |  |  |
| **Row Labels** | **1** | **2** | **Grand Total** |
| **Berkeley** | **10397.4182** | **41233.4471** | **51630.8653** |
| Black | 10397.4182 | 41233.4471 | 51630.8653 |
| **Beverly Hills** | **8999.72** | **13276.9964** | **22276.7164** |
| Black | 8999.72 | 13276.9964 | 22276.7164 |
| **Grand Total** | **19397.1382** | **54510.4435** | **73907.5817** |