Paper 027-2012

Custom Rollup: When the Cube's Default Behavior Doesn't Do the Right Job!

Paresh Patel, Shlomish Consulting, Inc., Wauconda, IL

ABSTRACT

Aggregation tables are primarily used as a method of Cube data storage and enhanced query performance, but they can also be used to satisfy business queries by creating custom rollup. Aggregation tables are built to define data in the cube that is available to satisfy query requests without having to summarize the data on the fly. An important aspect of Cube design is cube granularity. When cubes roll up to incorrect numbers, it's simply wrong.

 '**Very wrong. In practice people like Business Managers can get surprisingly tense about this kind of error, and you end up getting aggravation about your aggregations**.'

This paper presents a practical challenge encountered when rolling up cube data with attributes that have a **many-to-many relationship** in a cube that was resolved using aggregated tables.

.

INTRODUCTION

Default Cube behavior is to aggregate from the lowest level, summing everything up to produce aggregated totals. The Cube in question's original view presented Country, Products, and Number of Distinct Clients. When the business requested removal of the Product dimension to get distinct clients for Country, it gave inflated number as the relationship between Product and clients was many-to-many (i.e. each product can have multiple clients and each client can have a multiple products). This was resolved using aggregated tables rather than building different cubes.

Many of us still equate querying and reporting in cubes as being similar to SQL querying from a database, but in cubes, these are MDX queries. We forget that cubes are multi-dimensional structures and that we can utilize different aggregated tables to present different views to the business in the same cube. HOLAP is a hybrid approach in which higher-level aggregations that are commonly accessed are stored on a server, while the more granular information is stored in the data warehouse. This technique was developed to enable larger definitions of cubes without affecting the cube build time. A cube designer can add details or dimensionality into the cube without increasing the overhead cost of the MOLAP cube.

.

AGGREGATION DESIGN SCENARIO
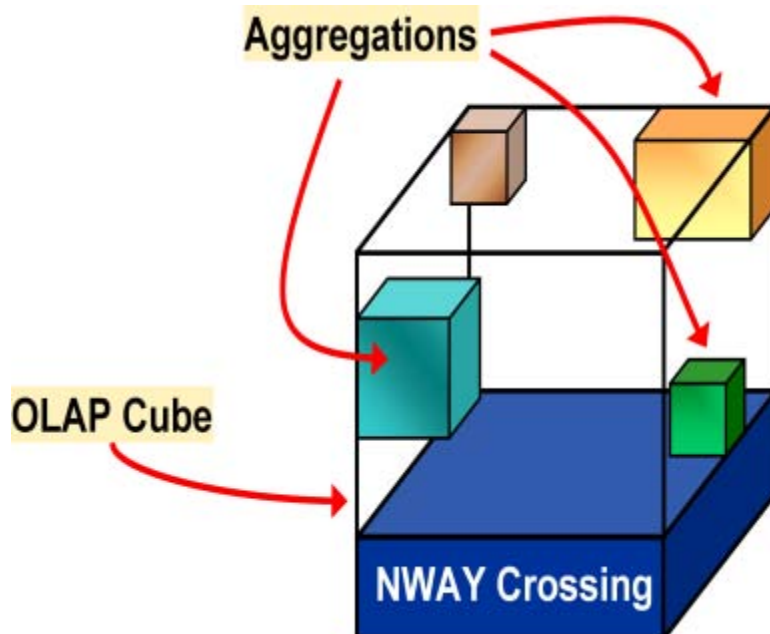
Custom Rollup: When the Cube's Default Behavior Doesn't Do the Right Job!

Figure 1. Aggregations in Cube

Before you begin to build an OLAP cube, you need to understand your data well. The end result of this activity is to build a tool to be used by decision makers and analysts to enable them to gain insights into and across what is likely a diverse set of products, geographic locations(country), or any number of other possible dimensions. Once you develop an understanding of your source data, you should be ready to determine how you would like to aggregate that data. Here is the actual scenario of the data in question. An international firm has different products offered across various countries. A client can have the multiple products and those products can be offered to multiple clients, so it is a many-to-many relationship.

```
* Given Summarized Table – Country Customer table;

data cust_cntry;
length DomicileCountry $2 ProductCode $6 ProductName $20 Customer_ID $10 Name $20 DOB 8 ;
format DOB mmddyy10.; informat DOB mmddyy10.;
infile datalines;
input Customer_ID Name &DOB DomicileCountry ProductCode ProductName;
datalines;
CX011   Bob Dole        07/04/1842   US   652941 Mutual Funds
CX012   Bruce Wayne     11/21/1940   US   652941 Mutual Funds
CZ022   Tony Blair      10/14/1969   UK   752948 Bonds
CZ089   Kevin Rudd      03/30/1978   AU   146941 Commodities
CX011   Bob Dole        07/04/1842   US   652535 Shares
CX012   Bruce Wayne     11/21/1940   US   652535 Shares
CZ022   Tony Blair      10/14/1969   UK   652941 Mutual Funds
CZ089   Kevin Rudd      03/30/1978   AU   752948 Bonds;
run;
```

CHALLENGES

By default, OLAP cubes calculate aggregations at each possible intersection of levels in one or more dimensions. However, there are situations where aggregations cannot be completed on the fly for all intersections within the cube due to business rules for some data levels. Also, the source data table isn't a fully summarized table (which fits nicely into the OLAP Studio GUIs) and derived measures are not possible due to specific business rules.

In this Scenario, when the granularity of a cube is at the country and product level, **the total number of distinct clients for country per product is 8**.

When this cube is built and then the Business User removes the product level, **we are still left with 8 distinct clients, when the true roll-up should now give 4 distinct clients**. **This is because a client can have many products, and each product may be linked to multiple clients. Here it is a many-to-many relationship between clients and products.**

Custom Rollup: When the Cube's Default Behavior Doesn't Do the Right Job!

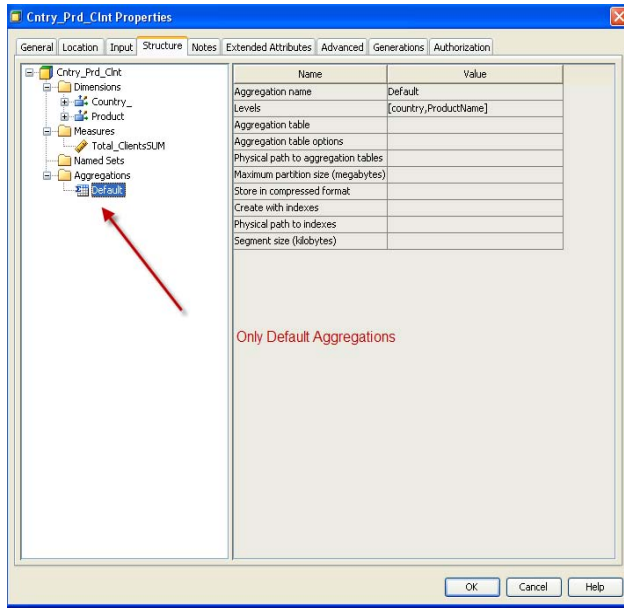## CUBE BUILT WITHOUT STORED AGGREGATION—DEFAULT BEHAVIOR

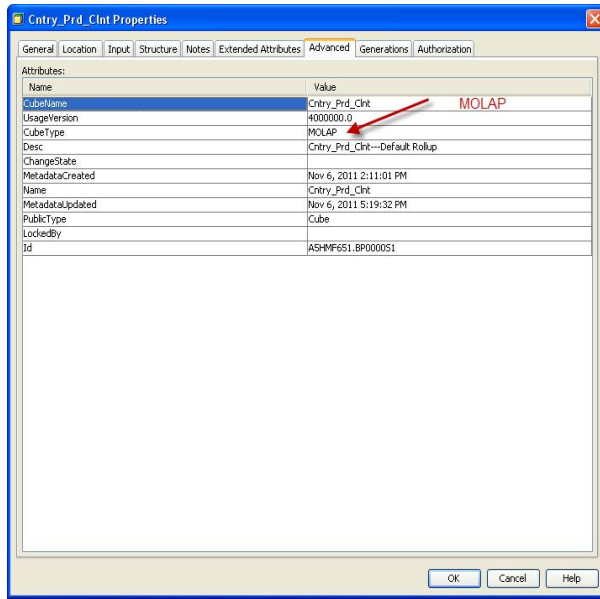**Figure 2. Initial Cube built with default Aggregation**

Figure 3. Initial Cube with MOLAP Structure

Here is a cube built with the granularity of Country and Product. It properties show that this is a MOLAP cube. With the MOLAP approach relevant, SAS ® proprietary, highly indexed aggregation tables are created and stored within the physical cube. Since the data is pre-aggregated, the response is quickly returned to the end users.

VIEWING CUBE IN ENTERPRISE GUIDE

In Enterprise Guide the OLAP Analyzer supports all of the functionality required to navigate through multidimensional data, add topic-specific business calculations, and extract information from multidimensional sources for further analysis with advanced statistical procedures or data mining.

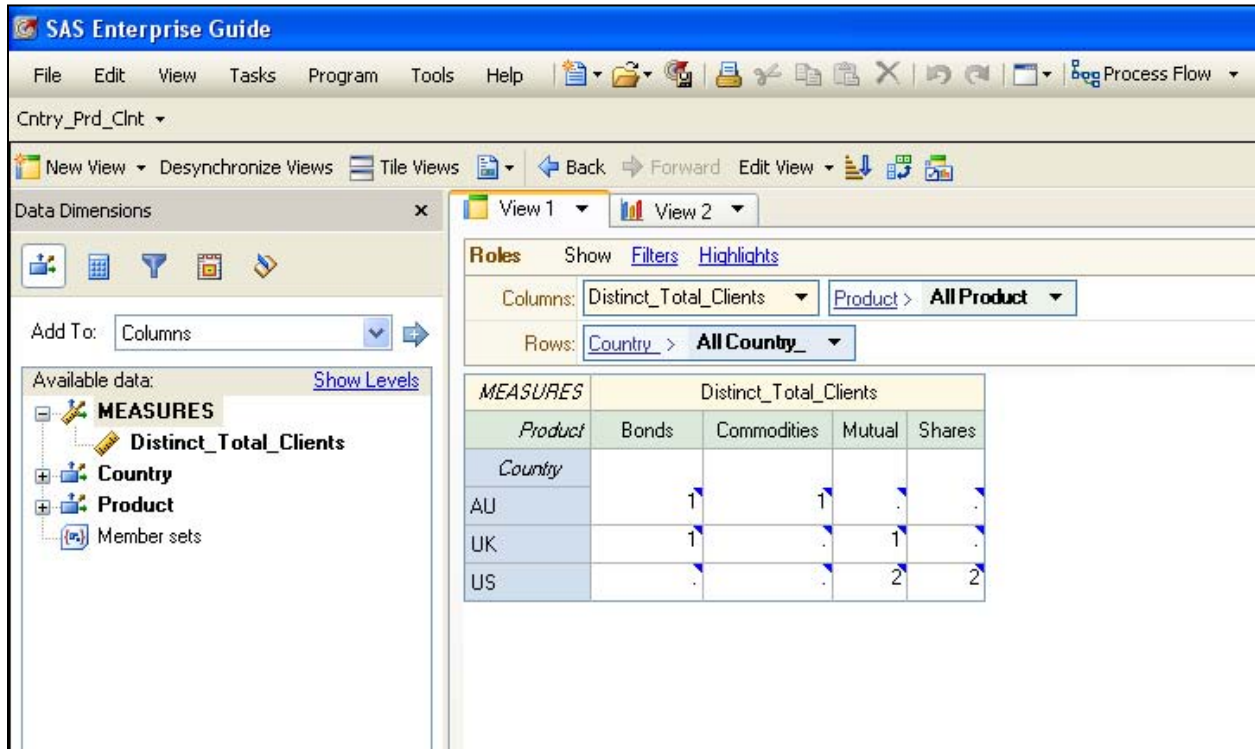**Custom Rollup: When the Cube's Default Behavior Doesn't Do the Right Job!**

Figure 4. Initial Cube with Country and Product Dimension selected in Enterprise Guide

The cube gives the correct number of clients per product per country. This is also the granularity of the data, so we do not encounter any challenges here.

DEFAULT ROLLUP BEHAVIOR

When the Product dimension is removed, the default rollup will render 8 distinct clients. But we know from the data above that it must roll up to 4 distinct clients.
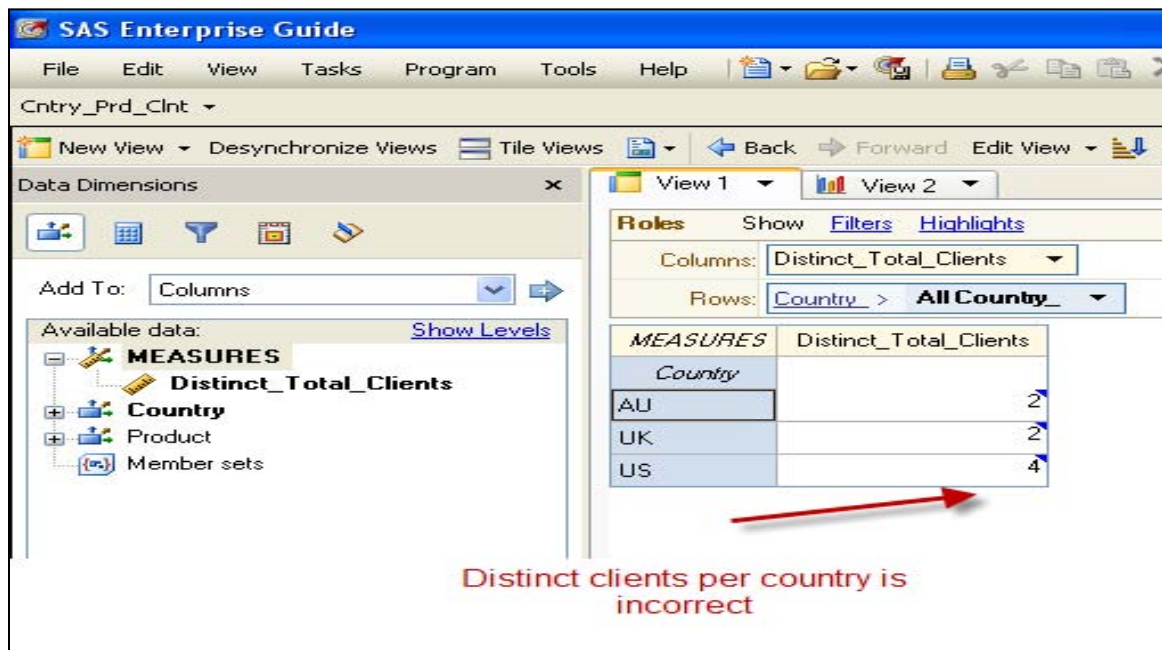

Figure 5 Distinct Clients per Country is incorrect with default roll-up

Custom Rollup: When the Cube's Default Behavior Doesn't Do the Right Job!

In SQL, when encountering a similar situation, we would simply write a query to select distinct clients per country. However, in order to accomplish the same thing in a cube when deviating from the normal behavior of a cube, we need a stored aggregation.

CUBE BUILT WITH STORED AGGREGATION

The same data as previously discussed was used for creating a different cube. The default granularity was the same: Distinct clients per country per product. **In this case, however, the scenario of a user rolling up to country level was resolved by using a stored aggregation. This stored aggregation was from an external table**.
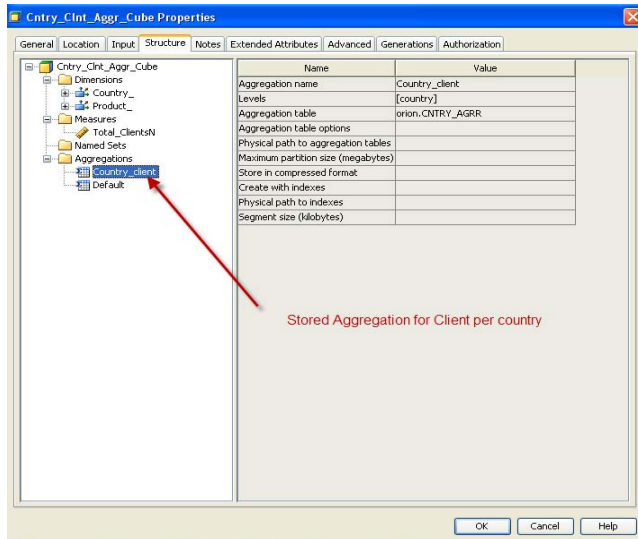

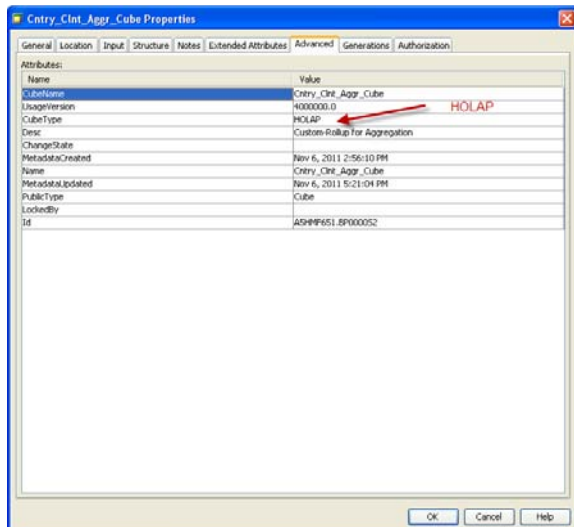**Figure 6.Cube built with additional Aggregation**

Figure 7. Initial Cube with HOLAP Structure

We can specifically assign an individual table for each Aggregation. It is important to note that the individual aggregated tables must have columns that completely match those listed in the aggregation step & this table must have at least one column defined in the MEASURE. With a HOLAP approach, a mix of the SAS ® proprietary aggregation tables and aggregated tables is used. This provides the business intelligence administrator the flexibility to establish the location of the multi-dimensional data. The location choice depends on the access frequency, administration, and processing overhead of the data.

Custom Rollup: When the Cube's Default Behavior Doesn't Do the Right Job!

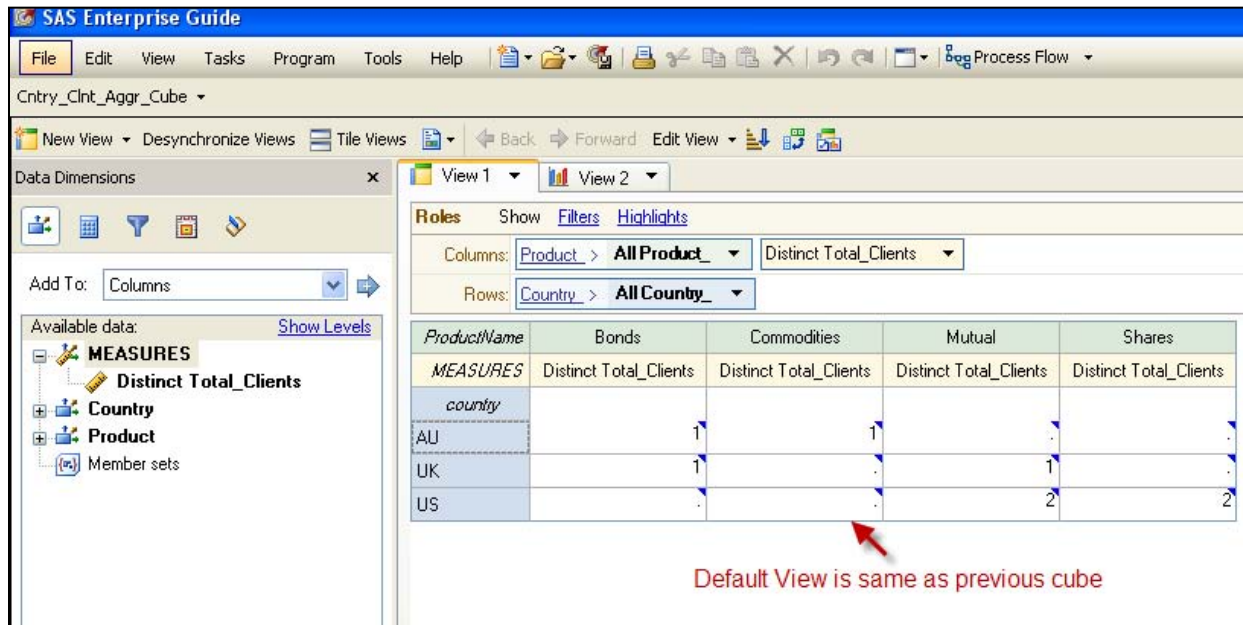## VIEWING CUBE IN ENTERPRISE GUIDE

Figure 8. Cube with Country and Product Dimension selected in Enterprise Guide

The cube gives the correct number of clients per product per country. This also is the granularity of the data so we do not encounter any challenges here.

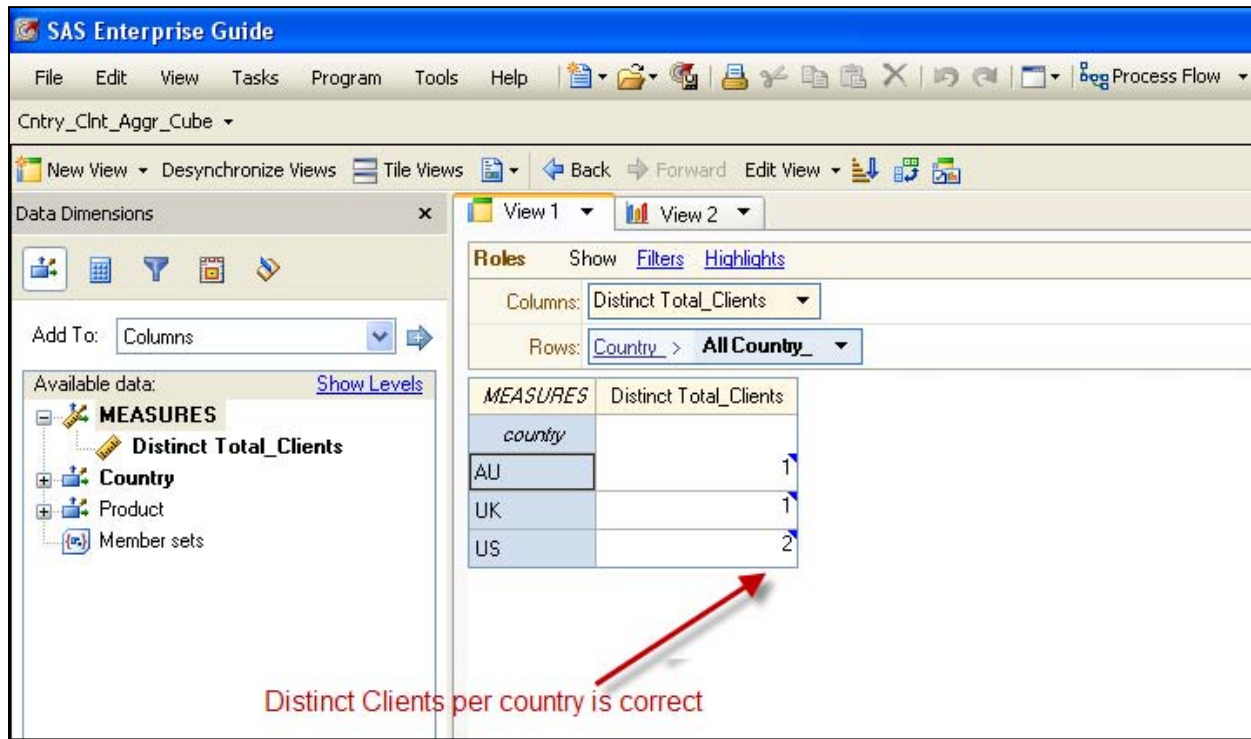## CUSTOM ROLLUPS IN CUBE USING AGGREGATION TABLE.

Figure 9. Distinct Clients per Country is correct using Custom Rollup

In this cube, when the product dimension is removed from the view, the cube no-longer does the rollup of distinct clients. Instead it uses the aggregation stored for that level which gives us the correct clients per country.

Custom Rollup: When the Cube's Default Behavior Doesn't Do the Right Job!

CONSIDERATIONS

Data Integrity

The cube input data and the aggregation tables need to be consistent with each other. It is vital that the data in aggregation tables stays in synch with the cube input data. An aggregation table always needs to be created from all input data (and not a subset of the input data).

NUNIQUE OPTION SCENARIO(NUMBER OF UNIQUE VALUES FOR LEVEL-NAME IN HIERARCHY-NAME)

This option entails directly using the features included into "SAS OLAP Cube Studio". It should be noted that NUNIQUE measures have a major effect on the performance of the cube and require a specific tuning.

If we had a data similar to

```
data cust_cntry;
length DomicileCountry $2 ProductCode $6 ProductName $20 Customer_ID $10 Name $20 DOB 8 Investments  8 ;
format DOB mmddyy10.; informat DOB mmddyy10. Investment comma12.;
infile datalines;
input Customer_ID Name &DOB DomicileCountry ProductCode ProductName Investments;
datalines;
CX011    Bob Dole         07/04/1842   US   652941 Mutual Funds      110,000
CX012    Bruce Wayne      11/21/1940   US   652941 Mutual Funds      120,000
CZ022    Tony Blair       10/14/1969   UK   752948 Bonds             210,045
CZ089    Kevin Rudd       03/30/1978   AU   146941 Commodities       325,000
CX011    Bob Dole         07/04/1842   US   652535 Shares            430,000
CX012    Bruce Wayne      11/21/1940   US   652535 Shares            445,000
CZ022    Tony Blair       10/14/1969   UK   652941 Mutual Funds      595,000
CZ089    Kevin Rudd       03/30/1978   AU   752948 Bonds           1,000,000
Run;
```

CUBE DESIGNER – SELECT MEASURES

On this page, you can view the cube's currently defined measures, add new measures, and remove existing measures. You can have a maximum of 1024 measures per cube. **Additionally available in the bottom of the same screen (same step) is the "Unique member count measures".**
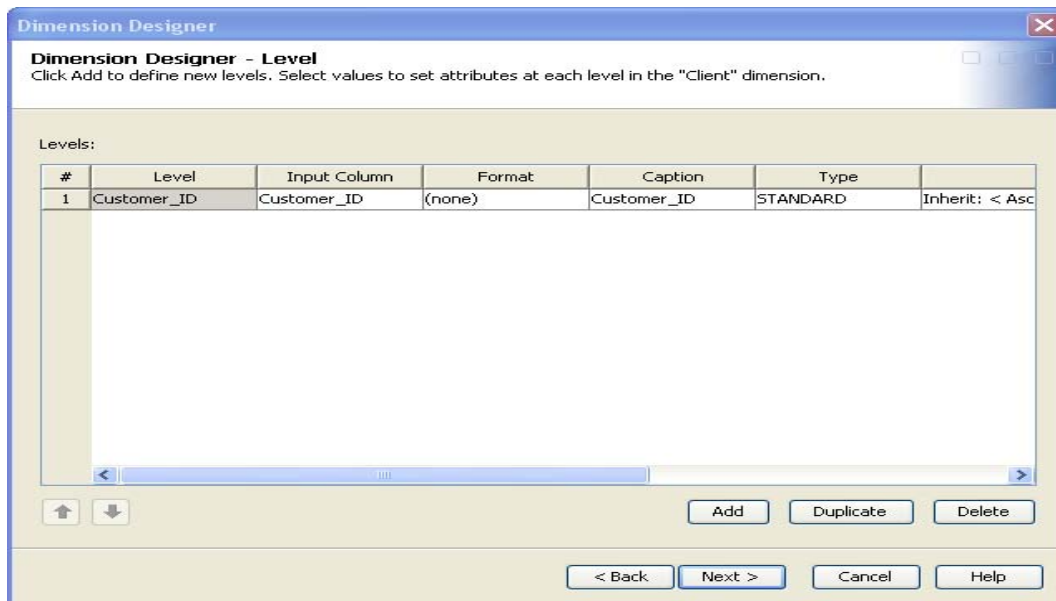

Figure 10.Customer Id is selected as a Level instead of a Measure

Custom Rollup: When the Cube's Default Behavior Doesn't Do the Right Job!

Here we can initially select customer-id as a level and then use the Cube Designer – Unique Member Count Measures.

CUBE DESIGNER – UNIQUE MEMBER COUNT MEASURE

To create the unique member count measure, select the level and hierarchy for the new measure. Selecting the level in a tree structure (dimension, hierarchy, level) automatically assigns the dimension and hierarchy to that measure. A name automatically is generated for that unique member count measure. This name consists of the level name, "NUNIQUE", and the parent hierarchy name
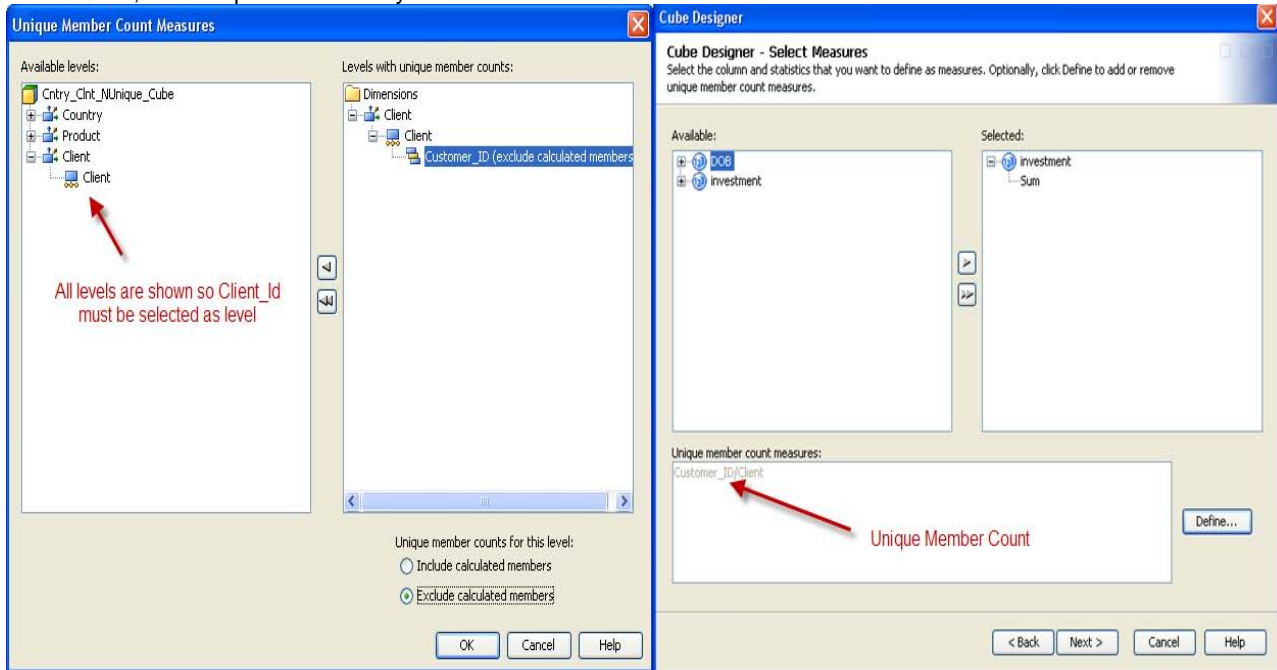

Figure11. Customer Id is selected as a Level   **Figure 12. Customer Id appears as NUNIQUE**

Custom Rollup: When the Cube's Default Behavior Doesn't Do the Right Job!

<u>**CUBE DESIGNER – UNIQUE MEMBER COUNT MEASURE**</u>
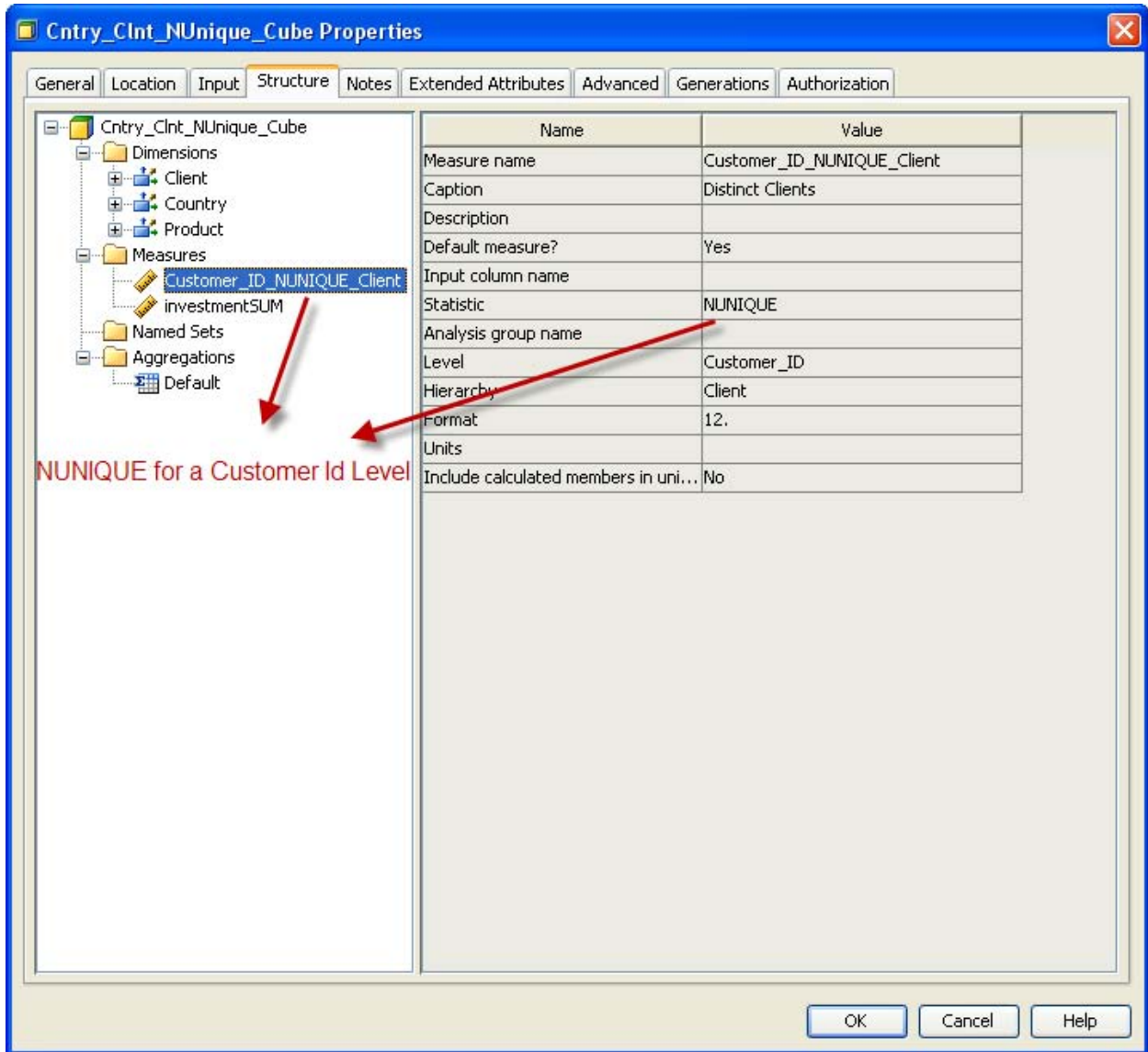

Figure 13. Cube properties showing Statistics as NUNIQUE

Custom Rollup: When the Cube's Default Behavior Doesn't Do the Right Job!

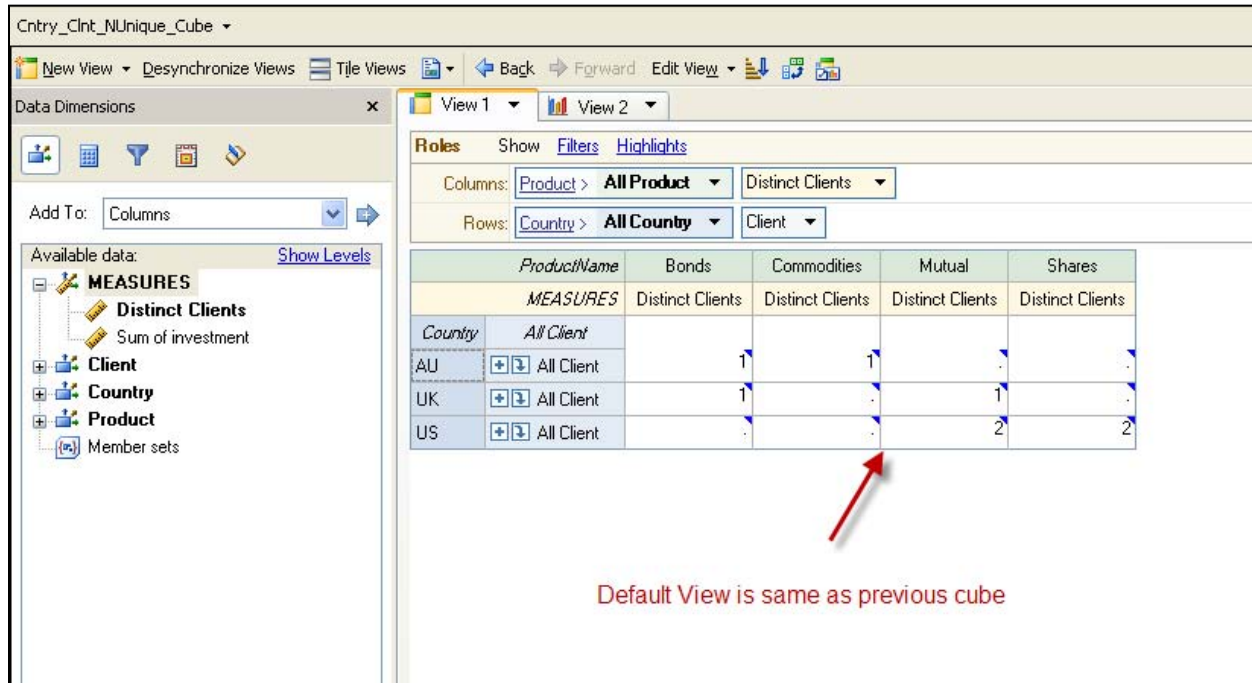<u>**VIEWING CUBE IN ENTERPRISE GUIDE**</u>

Figure 14. Cube with Country and Product Dimension selected in Enterprise Guide

The cube gives the correct number of clients per product per country. This also is the granularity of the data so we do not encounter any challenges here.
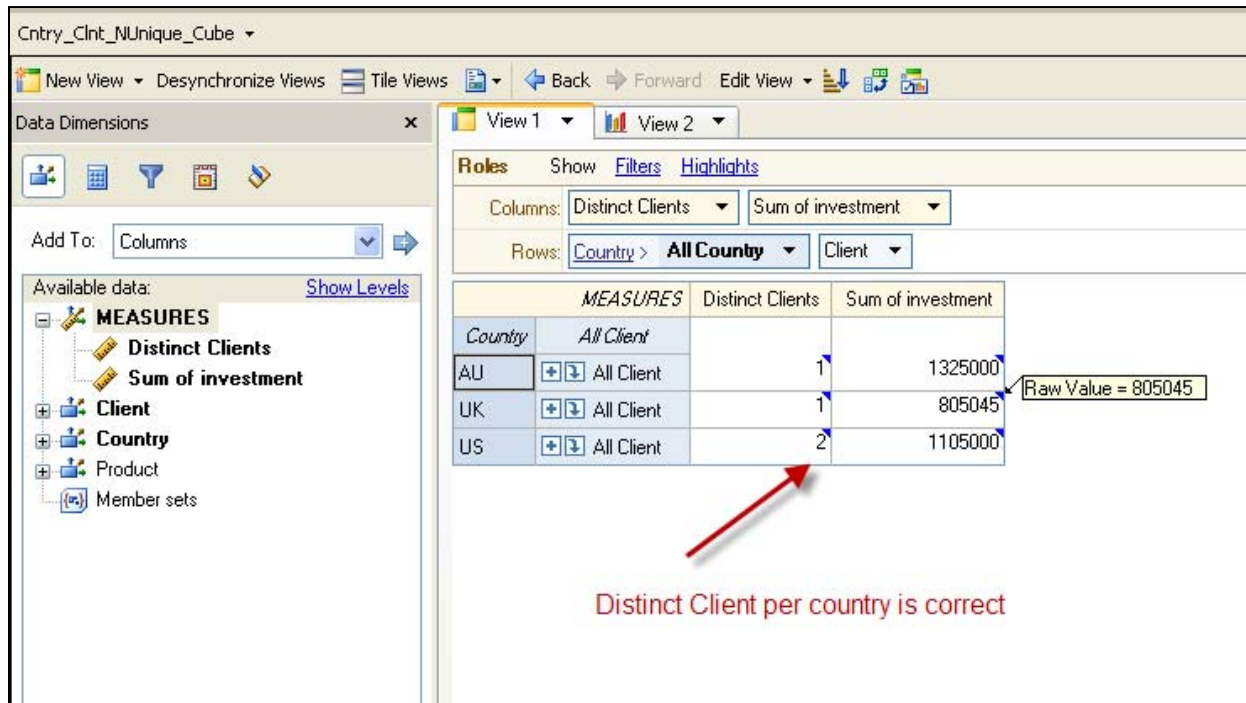

Figure 15. Cube with Country Dimension selected in Enterprise Guide

Custom Rollup: When the Cube's Default Behavior Doesn't Do the Right Job!

In this roll up, we get the correct value of distinct clients in the cube as it has NUNIQUE statistics associated with the level of customer id. For any report that has both non-NUNIQUE and NUNIQUE measures, the SAS OLAP server will make multiple passes through the aggregations in order to build the result set

LIMITS

As has been noted, NUNIQUE is the number of unique values for a given level of a given hierarchy (Number of Unique Values for level-name in hierarchy-name). "It is important to note that only one combination of level and hierarchy can be defined for a measure. After a combination has been used to create a unique member count measure, the combination cannot be used again." This means that it is not possible to define combinations of different dimensions to calculate distinct values if there are multiple dimensions within a single cube that require calculating distinct values.

Generally speaking, with a high volume of data it is possible to have a very quick response from a cube if you have generated the appropriate aggregations. However, when one of the involved measures is NUNIQUE, the appropriate aggregations may not be useful for relatively simple queries, forcing the cube to go to the fact table (and therefore to the higher volume of data) to calculate the count distinct.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Paresh Patel

Shlomish Consulting, Inc.

2299 Bluewater Dr

Wauconda, IL-60084

Cell: 847-989-1728

Email:pareshmpatel@hotmail.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.