Building a Cube in Analysis Services Step by Step and Best Practices

Rhode Island Business Intelligence Group
November 15, 2012

Sunil Kadimdiwan
Agenda

- Data Warehousing
  - Introduction
  - Star/Snowflake Schemas
- SQL Server Analysis Services
  - Overview
  - Components
- Questions?
Sunil Kadimdiwan: InfoTrove Inc.

- Working with SQL Server since 1993
- Focusing on Business Intelligence solutions
  - PowerPivot & Power View, Analysis Services, Reporting Services, Integration Services and SQL Server
- Frequent speaker: User Groups & Code Camps
- sunilk AT infotrove DOT com or 508-532-1180
- Find me in LinkedIn: Sunil Kadimdiwan
What does a cube mean?

- Excel Demo
- CrossTab report
  - a sophisticated one perhaps!
# A simple example

<table>
<thead>
<tr>
<th>Order Date</th>
<th>Product</th>
<th>City</th>
<th>Qty</th>
<th>Unit Price</th>
<th>Total Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/22/09</td>
<td>Lock Washer 4</td>
<td>Seattle</td>
<td>200</td>
<td>$3.00</td>
<td>$600.00</td>
</tr>
<tr>
<td>01/14/10</td>
<td>Touring Rim</td>
<td>Boston</td>
<td>5</td>
<td>$4000</td>
<td>$20000</td>
</tr>
</tbody>
</table>

Order Date = Date, Month/Week, Quarter, Year
Product = Product, Subcategory, Category
City = City, County, State, Region, Country, Territory
Metrics/Measures = Qty, Unit Price, Sales Total
Another example

- A line item on a bill received from a doctor
- What are the dimensions to derive?
  - Date (of treatment)
  - Doctor (maybe called "provider")
  - Patient
  - Procedure
  - Primary Diagnosis
  - Location (doctor's office/hospital)
  - Billing Organization (an organization the doctor belongs to)
  - Responsible Party (patient or patient's legal guardian)
  - Primary Payer (often an insurance plan)
  - Secondary Payer (party's spouse's insurance plan)
  - Etc.
Business Intelligence Architecture

- Monitoring & Administration
- Metadata Repository
- Data Sources
- Infotrove Data Warehouse
- Data Marts (OLAP Cubes)
- Reporting
- Scorecards & Executive Dashboards

* PeopleSoft, Oracle, SAP, CRM
Star Schema

- A **Star Schema** contains a fact table and one or more dimension tables.
  1. A Fact Table: Stores numeric fact (measures) such as Sales dollars, Costs, Unit Sales etc.
  2. Dimension Tables: They surround the Fact table, and they store descriptive information about the measures.
- The shape looks like a Star
Star Schema
Snowflake Schema
Analysis Services 2008 R2

- 4th version of Analysis Services
  - SQL Server 7.0, 2000, 2005 and 2008
- Business Intelligence Development Studio
- SQL Server Management Studio
What is running?

- **Hardware**
  - HP Pavilion dv6; Dual core, 8GB RAM
  - Windows 7 Ultimate

- **Software**
  - SQL Server 2008 R2
  - Analysis Services 2008 R2
Database environment

- Adventure Works DW 2008 R2
- Contoso Retail DW
Analysis Services Database

- An Analysis Services database includes:
  - Data Source
  - Data Source View
  - Dimension
  - Cube
  - Security Role
Data Source Connection

- Defines how Analysis Services connects to a physical data store using a managed Microsoft .NET Framework or native OLE DB provider.
- The connection string contains server name, database, security, timeout, and other connection-related information.
- Analysis Services directly supports many data sources.
  - Supported: Microsoft SQL Server databases and databases created by other products, including Oracle, DB2, and Teradata.
Connection to SQL Server

- Following providers will work:
  - OLE DB provider for SQL Server
  - SQL Native Client 10.0
  - .NET Provider/SqlClient Data Provider
    - (Avoid using .NET data sources – OLEDB is faster for processing in practices)
Data Source Views

- New feature Since AS 2005 and onwards
- A data source view is a logical representation of the data that is used by Analysis Services objects, built from the data sources already defined in the database.
- Can contain multiple Data Sources
- A data source view can contain relationships, calculated columns, and queries that are not present in the underlying data sources.
Dimension

- All dimensions are based on tables or views in a data source view.
- All dimensions are shared since AS 2005
- The structure of a dimension is largely driven by the structure of the underlying dimension table or tables.
- The simplest structure is called a star schema, which is where each dimension is based on a single dimension table that is directly linked to the fact table by a primary key - foreign key relationship.
A dimension consists of:

- Attributes that describe the entity
- User-Defined Hierarchies that organize dimension members in meaningful ways
  - such as
  
  Date -> Month -> Quarter -> Year
Attributes

- New feature since AS 2005
- Containers of dimension members
- Typically have one-many relationships between attributes in the same dimension:
  - All attributes implicitly related to the key
Measure Group

- In a cube, a measure is the set of values, usually numeric, that are based on a column in the fact table in the cube.
- A measure group contains one or more or all the measures from a single fact table. It can’t contain measures from different fact table.
Measure Group Advantages

- Measure groups provide the following advantages:
  - They can be partitioned and processed separately
  - They allow to include measures from different fact tables.
  - They are grouped by granularity: Same measure group, same granularity.
  - Security can be applied to specific measure groups
Inside a Cube

- Measures and Measure Groups
- Dimensions Relationships
- Calculations
- Actions
- Partitions
- Perspectives
Types of Dimensions

Different Dimension Relationships
- Regular Dimension Relationship
- Reference Dimension Relationship
- Fact Dimension Relationship
- Role Playing Dimension
- Parent-Child Hierarchy
Regular Dimension Relationships

- A traditional star schema design
- The Primary Key in the dimension table joins directly to Foreign Key in the fact table.
Reference Dimension Relationships

- Snowflake schema
- A Reference dimension using columns from multiple tables, or the dimension table links a dimension that is directly linked to the fact table.
Role Playing Dimension

It is used in a cube more than one time, each time for a different purpose.

- Each role-playing dimension is joined to a fact table on a different foreign key.
- Example, you might add a Time dimension to a cube three times to track the times that
  - products are ordered,
  - products are shipped,
  - Orders are due.
A parent-child hierarchy is a hierarchy in a standard dimension that contains a parent attribute. A parent attribute describes a self-join, within the same dimension table.

Example: Employee Hierarchy

- An employee is an employee who reports to his/her manager. His manager is an employee as well
- Employee Key self joins to ParentEmployeeKey
Slowly Changing Dimension

Some attribute values may change over time.

Two basic techniques:
- Type 1 change
- Type 2 change
**Slowly Changing Dimension – Type I**

A Type I change, is to simply overwrite the old value with the new one.

<table>
<thead>
<tr>
<th>key</th>
<th>salespersonid</th>
<th>salesperson</th>
<th>territory</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>405</td>
<td>tom jones</td>
<td>Northeast</td>
</tr>
<tr>
<td>102</td>
<td>410</td>
<td>paul tomlinson</td>
<td>Southwest</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>key</th>
<th>salespersonid</th>
<th>salesperson</th>
<th>territory</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>405</td>
<td>tom jones</td>
<td>Southeast</td>
</tr>
<tr>
<td>102</td>
<td>410</td>
<td>paul tomlinson</td>
<td>Southwest</td>
</tr>
</tbody>
</table>
Slowly Changing Dimension – Type 2

Create a new dimension row with the new value and a new surrogate key.

Mark the old row or timestamp as no longer in effect. The fact table will use the new surrogate key to link new fact measurements.

<table>
<thead>
<tr>
<th>key</th>
<th>salespersonid</th>
<th>salesperson</th>
<th>territory</th>
<th>startdate</th>
<th>enddate</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>405</td>
<td>tom jones</td>
<td>Northeast</td>
<td>4/24/2007</td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>410</td>
<td>paul tomlinson</td>
<td>Southwest</td>
<td>3/21/2005</td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>405</td>
<td>tom jones</td>
<td>Southeast</td>
<td>12/1/2009</td>
<td></td>
</tr>
</tbody>
</table>
A named set is a MDX expression that returns a set of dimension members.

You can define named sets and save them as part of the cube definition.

It allows you to reuse the same named set throughout the cube.

Typical example:

- Create a list Top 10 Customers based on Sales, Worst Performing Products etc.
- You can reuse in different queries.
Default Members

- Every Dimension has a Default Member
  - Usually the “All” member is the default member.
- Default Measures
  - The measures dimension also has a default measure
  - In our sample cube [Adventure Works], the default member for the cube is [Reseller Sales Amount]
Security

- Setup Roles
Advanced Topics

- KPI
- Actions
- Partitions
- Aggregations
- Perspectives
- Translations
- Browser
Questions?