Лекция 1

Тема «Архитектура Business Intelligence Tools»

Business Intelligence Is Data Analysis

Having designed a data warehouse, the next step is to understand and make business decisions from your data warehouse. Business intelligence is nothing more than analyzing your data and making actionable decisions. An example of business analytics is shown through the analysis of results from a product placed on sale at a discounted price, as commonly seen in any retail store. If a product is put on sale for a special discounted price, there is an expected outcome: increased sales volume. This is often the case, but whether or not it worked in the company ' s favor isn't obvious. That is where business analytics come into play. We can use SSAS 2008 to find out if the net effect of the special sale was to sell more product units. Suppose you are selling organic honey from genetically unaltered bees; you put the 8 - ounce jars on special — two for one — and leave the 10 - and 12 - ounce jars at regular price. At the end of the special you can calculate the *lift* provided by the special sale — the difference in total sales between a week of sales with no special versus a week of sales with the special. How is it you could sell more 8 - ounce jars on special that week, yet realize no lift? It ' s simple — the customers stopped buying your 10 - and 12 - ounce jars in favor of the two - for - one deal; and you didn ' t attract enough new business to cover the difference for a net increase in sales.

You can surface that information using SSAS 2008 by creating a Sales cube that has three dimensions: Product, Promotion, and Time. For the sake of simplicity, assume you have only three product sizes for the organic honey (8 - ounce, 10 - ounce, and 12 - ounce) and two promotion states (" no promotion " and a " two - for - one promotion for the 8 - ounce jars "). Further, assume the Time dimension contains different levels for Year, Month, Week, and Day. The cube itself contains two measures, " count of products sold " and the " sales amount. " By analyzing the sales results each week across the three product sizes you could easily find out that there was an increase in the count of 8 - ounce jars of honey sold, but perhaps the total sales across all sizes did not increase due to the promotion. By slicing on the Promotion dimension you would be able to confirm that there was a promotion during the week that caused an increase in the number of 8 - ounce jars sold. When looking at the comparison of total sales for that week (promotion week) to the earlier (non - promotion) weeks, lift or lack of lift is seen quite clearly. Business analytics are often easier described than implemented, however.

Microsoft Business Intelligence Capabilities

Different types of organizations face different challenges. Whether you work in a large or a small company, business intelligence (BI) is critical to provide the business insight you need to help everyone in every department of your organization succeed. To help you address specific BI needs, you typically need to perform various operations or tasks on your data. Figure 1 - 8 provides you with a list of various tasks typically performed for business intelligence in an organization and how SQL Server 2008 helps in various parts of business intelligence. You can have a single tool helping you with multiple BI tasks or multiple tools being used for each BI task. Your organization may only be utilizing some tasks for your BI needs. Now let ' s look at each operation in detail and how Microsoft SQL Server 2008 products help you in performing these operations.

Integrate
Integration
Services
DQP RDBMS
Store Model Explore Visualize Deliver
SQL Server
MD Store UDM
Analysis Services
SMDL Report Builder
Report Designer
Reporting Services
Report
Server
Figure 1-8

Integrating Data

Typically, organizations have data available from different backend systems. In order to build a data warehouse, you typically integrate all the data into a staging database. SQL Server Integration Services

(SSIS) helps you in integrating data from backend systems to a single system. SSIS helps you in extracting data, cleaning the data, and then loading it to a single system. If you have multiple SQL Server
relational databases, you can integrate the data for your data warehouse using distributed queries.

Storing Data

Your organization ' s data grows over time. Hence you do need to store the data for efficient access. You can store the data in multiple ways, from simple text files to an efficient database management system. SQL Server 2008 provides you with the ability to store your data in a relational database engine or the multidimensional database engine (SQL Server Analysis Services).

The Model

Once your organization ' s data has been stored, you need to create a model to analyze the data. You can create models on the data stored in an Analysis Services database or the relational database system. Databases created in Analysis Services conform to the Unified Dimensional Model (UDM). You learn more about UDM later in this chapter and throughout this book. In order to analyze the data from your relational database system, SQL Server Reporting Services provides you with a way to model the data using the Semantic Model Definition Language (SMDL). SMDL then helps you to analyze and report the data to satisfy your business needs.

Exploring Data

Once you have a model and the underlying data, you need to explore the data to interpret and get the intelligence from the data that will help you to meet your business needs. SQL Server Reporting Services helps you to explore the data from your models via two ways: ad - hoc analysis using Report Builder and through a structured format using Report Designer. Report Builder and Report Designer help you to easily explore the data using the models without the need of learning the query language to query your database engines. You learn more about Report Builder and Report Designer and how to explore data from Analysis Services in Chapter 20 .

Visualizing

Once you explore the data, you typically build reports that can be delivered to end users who can interpret the data and make intelligent business decisions to enhance your organization. Report Designer helps you to visualize the data as efficient reports that can then be deployed on to your Reporting Services server.

Deliver

Once you build your report on top of the data, you need a way for users to retrieve the reports easily. The Reporting Services server helps users to view the reports with appropriate authentication. In addition, Reporting Services allows you to deliver the reports at needed intervals to the appropriate users in your organization.

Microsoft SQL Server 2008 provides a platform to perform various business intelligence tasks to access, integrate, and manage data for your organization and help in building an efficient data warehouse. In addition, SQL Server 2008 offers a robust, scalable, and enterprise - ready data warehouse platform. With Microsoft SQL Server 2008, you can bring together and manage all your data assets to help ensure that the critical information you put in decision - makers ' hands is high - quality, complete, and timely, which can help them make better decisions. In addition to SQL Server 2008, which forms the core of the business intelligence platform, Microsoft offers additional products that form a fully integrated set of BI technologies to help make building, managing, and using BI for your organization less complicated and more economical. The result is that you and your organization can have the advantage of a complete set of BI capabilities. Figure 1 - 9 shows Microsoft business intelligence products. You can see that SQL Server
Analysis Services is the core business intelligence platform from Microsoft.

The majority of the consumers use Microsoft Office Excel as a core BI client for their organization. Due to this, there is very tight integration between SQL Server Analysis Services and Excel 2007 so you can analyze the data from your multidimensional databases effectively via pivot tables in Excel. You can use Excel 2007 to retrieve and view calculations that define your organization ' s performance from SQL Server Analysis Services such as Key Performance Indicators. This helps end users to easily interpret and understand how your organization is performing and make appropriate decisions.

In addition to Excel 2007, SharePoint Server 2007 and Performance Point 2007 form the suite of products from Microsoft that help in business intelligence for your organization.

Performance Point 2007 helps in analysis, forecasting, input from multiple people and departments, and the combination of multiple related reports. It offers an integrated performance management application that delivers a robust infrastructure to support your business planning. Built on the Microsoft BI platform, Office Performance Point Server 2007 can help your people continuously interact and contribute throughout the process of business planning, budgeting, and forecasting. With Office Performance Point Server 2007, you can manage consolidation and provide monitoring tools such as

scorecards and analysis tools that can help your organization track its changing performance — all through the familiar and easy - to - use Microsoft Office system environment.

Microsoft Office SharePoint Server 2007 offers an integrated suite of server capabilities that can help organizations connect people, processes, and information. With Office SharePoint Server 2007, decision - makers can easily access all their BI information, including scorecards, reports, and Office Excel spreadsheets. Office SharePoint Server 2007 also offers collaboration and powerful built - in search and content management features. When you deliver Microsoft BI through Office SharePoint Server 2007, you have one central location from which you can provide business intelligence capabilities to every employee and quickly connect your people to the information they need.

Scorecards, Analytics, Planning
(PerformancePoint Server 2007)

Microsoft Buisiness Intelligence
An end-to-end integrated offering

Collaboration and Content
(Office SharePoint Server 2007)

End-user Analysis
(Excel 2007)

Integration
Integration Services

Analysis
Analysis Services

SQL Server 2008
RDBMS

Reporting
Reporting Services

Performance
Management
Applications

BI
Platform

Figure 1-9

SQL Server Analysis Services 2008

SQL Server 2008 is the Microsoft business intelligence platform. Analysis Services 2008 is the multidimensional database engine. In addition to Analysis Services, SQL Server 2008 contains other services such as Integration Services (tools and engine to Extract, Transform, and Load) and Reporting Services, among other things. Integration Services, Analysis Services, and Reporting Services together form the core of the business intelligence platform with SQL Server as the backend. Analysis Services not only provides you with the ability to build dimensions and cubes for data analysis but also supports several data mining algorithms, which can provide business insight into your data that are not intuitive. Next you learn about the overall architecture of Analysis Services 2008 followed by the concept of the Unified Dimensional Model (UDM), which helps you to have a unified view of your entire data warehouse.

SSAS 2008 is a scalable, reliable, and secure enterprise class multidimensional database server. The architecture of Analysis Services allows it to provide scalability in terms of scale - out and scale - up features and in terms of very large database capabilities. Several instances of Analysis Services can be integrated together to provide an efficient scale - out solution. Similarly, Analysis Services is also 64 - bit enabled and scales - up on a large - scale system. On the other hand, the service has been architected with efficient algorithms to handle large dimensions and cubes on a single instance. Analysis Services provides a rich set of tools for creating multidimensional databases, efficient and easy manageability, as well as profiling capabilities.

The *Business Intelligence Development Studio* (BIDS) integrated within Visual Studio 2008 is the development tool shipped with SQL Server 2008 used for creating and updating cubes, dimensions, and Data Mining models. The *SQL Server Management Studio* (SSMS) provides an integrated environment for managing SQL Server, Analysis Services, Integration Services, and Reporting Services. SQL Server Profiler in the SQL Server 2008 release supports profiling SSAS 2008, which helps in analyzing the types of commands and queries sent from different users or clients to SSAS 2008. You learn more about BIDS and SSMS in Chapter 2 with the help of a tutorial. You learn about profiling an instance of SSAS 2008 using SQL Server Profiler in Chapter 15 . In addition to the above - mentioned tools, SSAS 2008 provides two more tools: the Migration Wizard and the Deployment Wizard. The Migration Wizard helps in migrating SQL Server 2000 Analysis Services databases to SQL Server 2008 Analysis Services. The Deployment Wizard helps in deploying the database files created using BIDS to SSAS 2008.

The SSMS provides efficient, enterprise - class manageability features for Analysis Services. Key aspects of an enterprise class service are availability and reliability. SSAS 2008 supports fail - over clustering on

Windows clusters through an easy setup scheme, and fail - over clustering certainly helps provide high availability. In addition, SSAS 2008 has the capability of efficiently recovering from failures. You can set up fine - grain security so that you can provide administrative access to an entire service or administrative access to specific databases, process permissions to specific databases, and read - only access to metadata and data. In addition to this, certain features are turned off by default so that the service is protected from hacker attacks.

Analysis Services 2008 natively supports XML for Analysis (XMLA) specification defined by the XMLA Advisory Council. What this means is that the communication interface to Analysis Services from a client is XML. This facilitates ease of interoperability between different clients and Analysis Services. The architecture of Analysis Services 2008 includes various modes of communication to the service as shown in Figure 1 - 10 . Analysis Services 2008 provides three main client connectivity components to communicate to the server. The Analysis Management Objects (AMO) is a new object model that helps you manage Analysis Services and the databases resident on it. The OLE DB 10.0 is the client connectivity component used to interact with Analysis Services instances for queries that conform to the OLE DB standard. The ADOMD.Net is .NET object model support for querying data from Analysis Services. In addition to the three main client connectivity components, two other components are provided by Analysis Services 2008. They are DSO 10.0 (Decision Support Object) and HTTP connectivity through a data pump. DSO 8.0 is the extension of the management object of Analysis Server 2000 so that legacy applications can interact with migrated Analysis Server 2000 databases on Analysis Server 2005. The data pump is a component that is set up with *IIS* (Internet Information System) to provide connection to Analysis Services 2008 over *HTTP* (Hypertext Transfer Protocol).

Even though XMLA helps in interoperability between different clients to Analysis Server, it comes with a cost on performance. If the responses from the server are large, transmission of XML data across the wire may take a long time depending on the type of network connection. Typically slow wide area networks might suffer from performance due to large XML responses. To combat this, SSAS 2008 supports the options for compression and binary XML so that the XML responses from the server could be reduced. These are optional features supported by SSAS 2008 that can be enabled or disabled on the server.

Analysis Services 2008 stores metadata information of databases in the form of XML. Analysis Services provides you with the option of storing the data or aggregated data efficiently in an optimized multidimensional format on an Analysis Services instance or storing them in the relational database as a relational format. Based on where the data and/or aggregated fact data is stored, you can classify the storage types as MOLAP (Multidimensional OLAP), ROLAP (Relational OLAP), or HOLAP (Hybrid OLAP).

MOLAP is the storage mode in which the data and aggregated data are both stored in proprietary format on the Analysis Services instance. This is the default and recommended storage mode for Analysis Services databases because you get better query performance as compared to the other storage types. The key advantages of this storage mode is fast data retrieval while analyzing sections of data and therefore provides good query performance and the ability to handle complex calculations. Two potential disadvantages of MOLAP mode are storage needed for large databases and the inability to see new data entering your data warehouse.

ROLAP is the storage mode in which the data is left in the relational database. Aggregated or summary data is also stored in the relational database. Queries against the Analysis Services are appropriately changed to queries to the relational database to retrieve the right section of data requested. The key advantage of this mode is that the ability to handle large cubes is limited by the relational backend only. The most important disadvantage of the ROLAP storage mode is slow query performance. You will encounter slower query performance in ROLAP mode due to the fact that each query to the Analysis Services is translated into one or more queries to the relational backend.

The HOLAP storage mode combines the best of MOLAP and ROLAP modes. The data in the relational database is not touched while the aggregated or summary data is stored on the Analysis Services instance in an optimized format. If the queries to Analysis Services request aggregated data, they are retrieved from the summary data stored on the Analysis Services instance and they would be faster than data being retrieved from the relational backend. If the queries request detailed data, appropriate queries are sent to the relational backend and these queries can take a long time based on the relational backend. If you choose the data and/or aggregated data to be stored in the optimized MOLAP format, you get better query performance than the ROLAP format, where data is being retrieved from the relational database. The MOLAP format helps Analysis Services to retrieve the data efficiently and thereby improves the query performance.

Based on your requirements and maintainability costs you need to choose the storage mode that is appropriate for your business. SSAS 2008 supports all three storage modes.

The Unified Dimensional Model

Central to the architecture is the concept of the Unified Dimensional Model (UDM) which, by the way, is unique to this release of the product. UDM, as the name suggests, provides you with a way to encapsulate access to multiple heterogeneous data sources into a single model. In fact, with the UDM, you will be buffered from the difficulties previously presented by multiple data sources. Those difficulties were often associated with cross - data - source calculations and queries — so, do not be daunted by projects with lots of disparate data sources. The UDM can handle it! The UDM itself is more than a multiple data - source cube on steroids; it actually defines the relational schema upon which your cubes and dimensions are built. Think of the UDM as providing you with the best of the OLAP and relational worlds. UDM provides you with the rich metadata needed for analyzing and exploring data along with the functionality like the complex calculations and aggregations of the OLAP world. It supports complex schemas, and is capable of supporting ad - hoc queries that are needed for reporting in the relational world. Unlike the traditional OLAP world that allows you to define a single fact table within a cube, the UDM allows you to have multiple fact tables. The UDM is your friend and helps you have a single model that will support all your business needs. Figure 1 - 11 shows a UDM within SQL Server Analysis Services 2008 that retrieves data from heterogeneous data sources and serves various types of clients.

Key elements of the UDM are as follows:

Heterogeneous data access support: UDM helps you to integrate and encapsulate data from heterogeneous data sources. It helps you combine various schemas into a single unified model that gives end users the capability of sending queries to a single model.

Real - time data access with high performance: The UDM provides end users with real - time data access. The UDM creates a MOLAP cache of the underlying data. Whenever there are changes in the underlying relational database, a new MOLAP cache is built. When users query the model, it provides the results from the MOLAP cache. During the time the cache is being built, results are retrieved from the relational database. UDM helps in providing real - time data access with the speed of an OLAP database due to the MOLAP cache. This feature is called proactive caching. You learn more about proactive caching in Chapter 21 .

Rich metadata, ease of use for exploration, and navigation of data: UDM provides a consolidated view of the underlying data sources with the richness of metadata provided by the OLAP world. Due to rich metadata supported by OLAP, end users are able to exploit this metadata to navigate and explore data in support of making business decisions. UDM also provides you with the ability to view specific sections of the unified model based on your business analysis needs.

Rich analytics support: In addition to the rich metadata support, the UDM provides you with the ability to specify complex calculations to be applied to the underlying data; in this way you can embed business logic. You can specify the complex calculations by a script - based calculation model using the language called MDX (Multi Dimensional eXpressions). UDM provides rich analytics such as Key Performance Indicators and Actions that help in understanding your business with ease and automatically take appropriate actions based on changes in data.

Model for Reporting and Analysis: The UDM provides the best functionality for relating to both relational and OLAP worlds. UDM provides you with the capability of not only querying the aggregated data that are typically used for analysis, but also has the ability to provide for detailed reporting up to the transaction level across multiple heterogeneous data sources.

Another handy aspect of using the UDM is the storage of foreign language translations for both data and metadata. This is handled seamlessly by the UDM such that a connecting user gets the metadata and data of interest customized to his or her locale. Of course, somebody has to enter those translations into the UDM in the first place; it is not actually a foreign language translation system.

Summary

Reading this chapter may have felt like the linguistic equivalent of drinking from a fire hose; it is good you hung in there because now you have a foundation from which to build as you work through the rest of the book. Now you know data warehousing is all about structuring data for decision support. The data is consumed by the business analyst and business decision - maker and can be analyzed through OLAP and Data Mining techniques.

OLAP is a multidimensional database format that is a world apart in form and function when compared to an OLTP relational database system. You saw how OLAP uses a structure called a cube, which in turn relies on fact tables (which are populated with data called facts) and dimension tables. These dimension tables can be configured around one or more fact tables to create a star schema. If a dimension table is deconstructed to point to a chain of sub - dimension tables, the schema is called a snowflake schema. By choosing SQL Server 2008 you have chosen a business intelligence platform with great features with

reliability, availability, and scalability. The SQL Server business intelligence platform is the fastest growing with highest market share product in the market. The rest of this book illustrates the power of SQL Server Analysis Services 2008, which is the core part of the BI platform from Microsoft.

In the unlikely event that you didn ' t read the Introduction, mention was made that you should read at least the first three chapters serially before attempting to tackle the rest of the book. So, please do not skip Chapter 2 , an introduction to Analysis Services, and Chapter 3 , an introduction to the technology behind the most famous acronym in business analytics, MDX.