

Microsoft® SQL Server™ 2005

Solving Business Problems with SQL Server 2005 Analysis Services

Author: Craig Utley

Published: April 2005

Summary: This paper examines a number of common business problems and shows how SQL Server 2005 Analysis Services solves these problems in a simple and cost-effective manner. Types of business problems covered are providing a simple, integrated view of data, data mart and niche application consolidation, intelligent views of data, localization challenges, and real-time business intelligence data. These business problems typically have been challenging or impossible with other data warehousing products. Microsoft brings these new capabilities to the enterprise in a simple yet powerful package.

Copyright

This is a preliminary document and may be changed substantially prior to final commercial release of the software described herein.

The information contained in this document represents the current view of Microsoft Corporation on the issues discussed as of the date of publication. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information presented after the date of publication.

This White Paper is for informational purposes only. MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS DOCUMENT.

Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Microsoft Corporation.

Microsoft may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Microsoft, the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

© 2005 Microsoft Corporation. All rights reserved.

Microsoft, ActiveX, SharePoint, Visual Basic, Visual C#, Visual Studio, Windows, Windows Server, and Windows Server System are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

The names of actual companies and products mentioned herein may be the trademarks of their respective owners.

Table of Contents

Introduction

Too often, businesses find that their investment in first-in-class applications leads to silos of information that become difficult to integrate and analyze. This means the promise of making more informed decisions in a timely manner often goes unfulfilled. In order to effectively analyze these disparate systems, organizations have tried any number of approaches: swapping one relational reporting system for another, extracting and transforming the data and building a data warehouse, or buying new enterprise systems because the new system promises improved reporting. Each of these solutions has drawbacks, from using inadequate tools to address the problem to requiring new and costly projects to consolidate data.

The problems experienced by many businesses are nothing new. For example, many companies seek to buy software products, such as SAP, to run much of their business. Unfortunately, these products often have complex data structures that make it difficult to perform analysis effectively. A company may have customer relationship management (CRM) and supply chain management (SCM) applications that are completely separate, yet both contain information critical to making strategic business decisions. Companies have also found that in order to meet a variety of reporting and analytics needs, they have produced an array of applications and data stores in order to meet the varied and often changing needs of the analysts, information workers, and users.

In order to address these concerns, Microsoft has been involved in easing the transition to business intelligence (BI) for years. With the release of Microsoft® SQL Server™ 2005 Analysis Services, Microsoft significantly raises the bar. Organizations now have a single, consistent solution for reporting against either online transaction processing (OLTP) or online analytical processing (OLAP) data stores. SQL Server 2005 Analysis Services greatly reduces the amount of effort required to provide a consistent view of data that is integrated from an array of disparate applications and formats. Whether the goal is to integrate data from multiple applications into a consistent format, analyze data from a variety of sources, report against this data using a variety of formats and techniques, mine the data for relationships, or forecast future results with predictive analytics, SQL Server 2005 Analysis Services provides a suite of highly integrated tools to help businesses take full advantage of the spectrum of BI.

This paper will examine a number of common business problems and how SQL Server 2005 Analysis Services solves these problems in a simple and cost-effective manner. The business problems covered in this paper are:

- **Simple, Integrated View of Data.** Many businesses maintain a number of separate applications for such functions as enterprise resource planning (ERP), CRM, SCM, Human Resources, and so forth. Microsoft SQL Server 2005 Analysis Services addresses this challenge through the use of Data Source Views, a technology that allows for the creation of analytical models against the existing data structures.
- **Data Mart Consolidation.** Organizations may build data marts for each functional area and then discover that consolidating these marts into a warehouse is difficult or impossible. Microsoft SQL Server 2005 Analysis Services introduces the Unified Dimensional Model and Perspectives in order to create a single version of the truth and expose it as virtual data marts to individual departments as needed.
- **Intelligent View of Data.** Most warehousing tools today build rigid hierarchical structures to allow analysts and knowledge workers to drill up and down into data. While powerful, drilling up and down is not the only form of analysis that can be performed. Microsoft SQL Server 2005 Analysis Services allows businesses to build solutions that support flexible hierarchies, many-to-many relationships, and other dimension types that allow for more advanced analytics than are available in other solutions.
- **Localizing the Analytics Experience.** Businesses with operations that span languages must consolidate data into a single language and currency in order for analysis to be valid, but this makes it more difficult for organizations worldwide to use the solution. Thanks to a feature called Translations, Microsoft SQL Server 2005 Analysis Services lets companies build a single solution that can localize values on the fly depending on the user's location.
- **Real-Time BI.** Most business intelligence and data warehousing solutions are snapshots in time that are updated nightly or weekly. Real-time BI solutions exist but they tend to be very slow compared to snapshots. Microsoft SQL Server 2005 Analysis Services gives businesses the ability to have real time or near real time BI capabilities with snapshot-like speeds and minimal impact on production systems.

Examining the Business Problems

In order to examine the kinds of problems SQL Server 2005 Analysis Services can solve, it is best to explore real business challenges encountered by a variety of organizations today. While by no means an exhaustive list of the problems that can be addressed by SQL Server BI tools, this paper provides an overview of common problems that SQL Server 2005 Analysis Services solves in an efficient and cost-effective manner. The Microsoft focus on allowing organizations to more quickly integrate, analyze, and report against their data, regardless of where it resides, greatly lowers the total cost of ownership (TCO) while decreasing the time needed to make informed decisions using data from a wide range of applications. Table 1 provides an overview of business problems and their solutions.

Table 1: SQL Server 2005 Analysis Services features and the business problems they solve.

Business Problem	SQL Server 2005 Analysis Services Feature that Addresses this Problem
Simple, Integrated View of Data	UDM's Data Source Views
Data Mart and Niche Application Consolidation	UDM's Perspectives
Intelligent Business View of Data	UDM's Many-to-Many Dimensions, Role Playing Dimensions, Referenced Dimensions, and Fact Dimensions
Localizing the Analytics Experience	UDM's Translations and Currency Conversion Wizard
Real-Time BI	UDM's Proactive Caching

Problem 1: Simple, Integrated View of Data

Most organizations have data spread over numerous systems such as ERP, CRM, and SCM. These systems may be from separate vendors and employ different terminology, data schemas, and even database engines. However, business decision makers need to view the data from all these sources in order to make informed decisions.

Having heterogeneous data stored in disparate formats means that extraction, transformation, and loading (ETL) routines of varying complexity are required in order to consolidate the data to perform analysis. Even if an organization performs analysis against just a single source such as the CRM application, the data has to be transformed into an operational data store or data mart (usually in the form of a star schema) in order to build cubes and use advanced analytics applications.

Solution: Disconnected Semantic View

SQL Server 2005 Analysis Services new capabilities make it possible to perform advanced analysis against data in an organization's existing systems with far less effort than has been required in the past. With Microsoft UDM, or Unified Dimensional Model, organizations can create a holistic view of disparate data, including relational databases, OLTP databases, flat files, and even Web services. The UDM gives businesses the ability to model complex data structures from a variety of sources and then perform relational reporting or multidimensional analysis against the UDM. This gives businesses the benefits of relational reporting that benefits from OLAP aggregations, while at the same time providing powerful analytic capabilities against the data.

SQL Server 2005 Analysis Services can create UDM cubes without the intermediate step of building a star schema data warehouse. The UDM enables organizations to build reporting applications directly against a production system and doesn't require an intermediate data warehouse as in the past. This is because the UDM doesn't require the data to be in a star or snowflake schema, but can connect to any data source whose data is stored in a third normal form. Regardless of the approach, SQL Server 2005 Analysis Services provides the ability to design a complex schema on top of the existing schema, including calculated columns, SQL statements that mimic tables, and more. These schemas can provide the basis for building cubes. Since the cube is a separate store, it doesn't impact the performance of the underlying data source. This means that the production OLTP systems continue to run at maximum speed, while allowing the business to have access to advanced analytics capabilities and, if desired, near real-time analytics.

SQL Server 2005 Analysis Services provides this powerful capability through the use of a technology called Data Source Views. Data Source Views allow the technology staff to create complex analytical models on top of existing schemas, whether those schemas are OLTP or star schemas. This allows for the creation of complex schemas that include calculated columns, SQL queries that can be treated as tables, friendly names for items, and virtual relationships. One major advantage of Data Source Views is that they can be created by analysts who do not have Create permissions on the database. Once built, cubes can be based on these Data Source Views, allowing for the building of cubes against a richer schema while adding the ability to perform advanced analysis on production systems without adversely affecting those critical applications.

Problem 2: Data Mart and Niche Application Consolidation

Data Mart Consolidation

One of the primary strategies companies employ to develop data warehouses is to build data marts, smaller warehouses targeted to particular divisions of the company or created to solve specific problems, such as analyzing financial documents or performing market basket analysis.

Having multiple data marts presents several problems. It can be expensive to maintain a series of marts scattered across the enterprise, requiring multiple updates from the same data sources. In addition, it can be difficult to consolidate these disparate marts into a smaller number of marts or a single enterprise data warehouse.

Eliminating a number of data marts has many TCO benefits. First, work is eliminated by lowering the number of marts and reducing data replication and ETL processing. Storage space requirements are reduced by eliminating redundant data storage. Data synchronization costs are lowered as well. Data integrity risks are reduced because the need to keep multiple data marts consistent with each other is eliminated.

Solution: Single Version of the Truth

Microsoft has addressed this problem by introducing the construct called the Unified Dimensional Model (UDM). UDM is one place where businesses can consolidate an organization's business logic (dimensions, measures) and Key Performance Indicators (KPIs) as well as data to provide the "single version of corporate truth" to the consumers of BI.

SQL Server 2005 Analysis Services addresses this challenge by allowing for a single UDM to be seen differently based on the user who is looking at it; these different views are called Perspectives. These Perspectives allow a UDM to be segmented, based on the needs of various departments. This preserves the need of each department to have their own marts, while eliminating redundant data storage, ETL processes, and synchronization costs. Data integrity and consistency issues are eliminated because there is only a single physical copy of the cube.

Handling Niche Applications

Many organizations build separate applications in order to handle a variety of reporting and analysis needs. A company may build separate solutions for relational reporting, analytics, scorecards, dashboards, and so forth. These separate applications are often driven off different data stores; a relational reporting tool normally reports against a normalized OLTP structure while an analytics application typically pulls data from an OLAP cube built off a star schema.

Having separate applications using a variety of data stores leads to a number of potential issues. Data integrity is far more difficult to maintain, as there is no single version of the truth. Instead, OLTP and OLAP systems each represent a version of the truth despite the fact that OLTP systems typically have not been scrubbed during an ETL process. Separate data stores also require more storage space and tend to promote data silos, the antithesis of a single store sought after in warehousing solutions. In addition, reporting and analytics applications are often seen as completely separate applications, feeding off separate data stores and having completely different needs. Because of this, many third-party analytics applications have no support for any sort of relational reporting, and only limited abilities to produce any kind of hardcopy output of analytical data.

Solution: The UDM's Single Model

In order to address the proliferation of niche applications, Microsoft has introduced a dramatic shift in the conventional wisdom surrounding business intelligence. In the past, the separation of OLTP to OLAP was clear: companies extracted data from one or more OLTP data sources, transformed that data into a consistent and consolidated state, loaded the data into a star schema database, and then built cubes from that star schema. This process meant that analysts could examine the data in high performance cubes or report against relational data, but the structures of the data were different, the tools for analyzing or reporting were different, and even the languages used to query data were significantly different. Now these issues are eliminated thanks to the Unified Dimensional Model, or UDM.

The UDM merges relational reporting needs from OLTP sources with analysis needs from OLAP sources into a single reporting model. By using UDM, both analytics and reporting applications access the same data store. OLAP cubes are built and maintained on the fly, providing powerful analytics capabilities combined with near real-time updates to the cubes. Reporting applications receive a speed benefit from the OLAP caching performed by the UDM.

By eliminating the need for separate data models for OLTP and OLAP, a number of benefits are realized. Data integrity concerns can be focused on a single location rather than multiple data stores. Data is no longer stored in a series of data silos but stored in a central location that is used for both relational and OLAP reporting. The need for niche applications and separate data stores is reduced by allowing tools to access a common data store and look at the data either relationally or in a cube format.

Further enhancing the UDM and all reporting and analytic applications is the ability to create Key Performance Indicators (KPIs) using the KPI Framework. This allows organizations to define the key metrics of the business so that those definitions are created and stored in a single location, and all applications can access that single version of the truth. KPIs can be indicated using simple but powerful graphics such as stoplights or gauges. This allows business decision makers to quickly assess the state of the business' key drivers in a single glance.

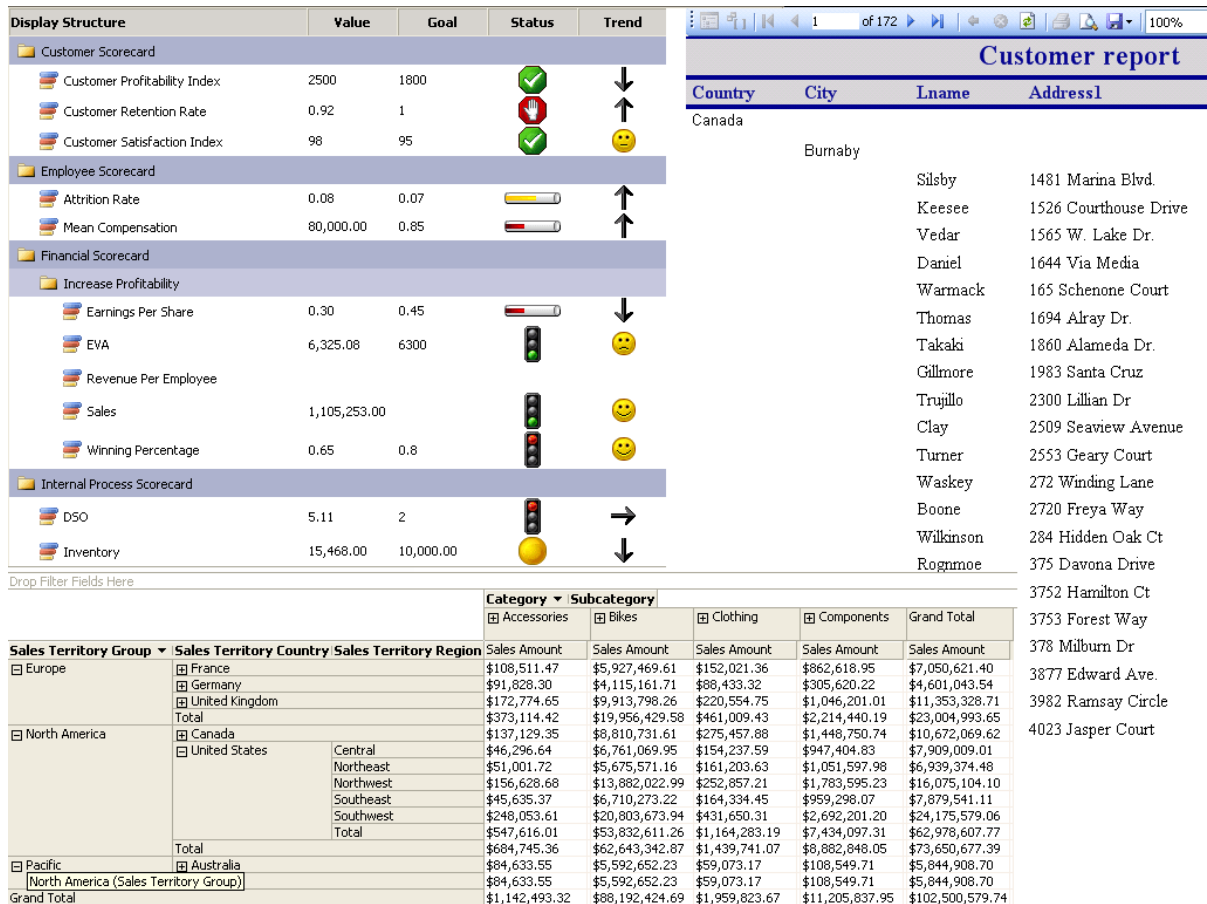


Figure 1: Mixed Reports

Problem 3: Intelligent Business View of Data

One of the most powerful features of analytics applications is the ability to drill down into data. This allows users to perform the iterative process of examining the data, forming a hypothesis, and then investigating the data further. By giving all users the ability to drill down through the data, organizations have opened up the data for any user to examine the data and answer their own questions about what underlies the numbers they are seeing. In this way, users can perform their own analysis, eliminating the need to return to Information Technology staff, repeatedly asking for new reports with lower and lower levels of detail.

While drill down is a powerful concept, it does not answer all questions. Various applications allow for some alternative analytics, but they are limited by the capabilities of the underlying data warehousing engine. These limitations make certain forms of analysis difficult.

Rather than just drilling down, users need to be able to drill across from one dimension to another. For example, when looking at a products dimension, an analyst may need to look at the customers who ordered those products or the suppliers from whom those products were purchased. Analysts may need to compare values, such as comparing order dates to ship dates, rather than focus on single values. Businesses may need to model more complex relationships than simple hierarchies, such as when an author writes more than one book and a book is written by more than one author. Data warehousing engines tend to have difficulty modeling such complex scenarios, making it nearly impossible to analyze and report against such structures.

Solution: Advanced Dimensional Modeling

SQL Server 2005 Analysis Services addresses these issues through a number of new features that open up a wealth of new analytical power. One such feature is attribute-based dimensions. OLAP products today tend to enforce a very rigid hierarchical structure on dimensions. This means that business analysts drill down from Product Group to Product Family to SKU. However, there are often a number of additional attributes to the dimension, such as size, color, weight, and so forth. These additional attributes are often tied to a particular SKU but it becomes difficult to drill down on color as a level in a dimension. Attribute-based dimensions get around this issue by focusing dimensions on attributes rather than a rigid hierarchy. All attributes can be used for drilling down and filtering, allowing any combination of hierarchies to be built on the fly.

SQL Server 2005 Analysis Services also introduces the concept of many-to-many dimensions. Traditional data warehousing solutions allow a fact to be tied to only a single dimension member, meaning that a particular fact record could be tied to one and only one product in the Product dimension. In the real world, however, items are not always so clean. A company might have customers and bank accounts in which a customer can have several accounts (checking, savings, money market, and so forth), and each account can be held by several customers (the savings account for John Doe and Jane Doe). SQL Server 2005 supports this complex relationship in which a single fact record can map to multiple dimension records, allowing for more complex and powerful analysis with easier modeling of the data.

Many data warehousing engines require cubes to be built with one dimension per type of analysis to be performed. For example, a firm might want to analyze orders by the Order Date or Ship Date. Rather than building two separate dimensions, SQL Server 2005 Analysis Services allows the warehouse designer to build a single time dimension and then use that same dimension as a generic time dimension as well as using it as an order date and ship date dimension. The ability for a single dimension to play different roles is known as a role-playing dimension and these dimensions help reduce disk storage and processing time by only storing the data once but allowing that dimension to be viewed as a number of separate dimensions by analysts and other users.

Sometimes data is stored in a fact table that is not contained in any dimension table. SQL Server 2005 Analysis Services makes it possible to pull out this data and create a one-to-one dimension with the fact table, usually referred to as a *degenerate dimension* and called a *fact dimension* in SQL Server 2005 Analysis Services. This means that it becomes simple to have a dimension that is just used to look up a particular fact record or identify all the values that make up an aggregate cell.

Problem 4: Localizing the Analytics Experience

One of the primary benefits of data warehouses is that they consolidate data into a consistent state. For example, sales in various countries are recorded in that country's local currency. However, adding British Pounds and Japanese Yen together does not make sense; instead, the BI Wizard has a feature that allows currency conversion rules to be defined directly in the UDM. Similarly, strings may be converted from one language to another, measurement units may be converted from English to metric, and so forth.

The advantage to converting data to a consistent format is obvious: comparisons across countries are invalid until the data is put in a consistent local format. This allows for analysis of data across the enterprise regardless of its country of origin. The disadvantage is that if the warehouse consolidates all data to US dollars, analysts in the Tokyo office now have a warehouse in English that may be difficult to use. The business can decide if it wants to build multiple copies of the warehouse, localized for each office, or if a single warehouse in a single language is enough.

Solution: Built-in Data and Metadata Translations and Currency Conversions

Businesses now have an additional choice thanks to SQL Server 2005 Analysis Services, which includes the ability to build the warehouse in a single language, but allow users to see localized versions of the data. This means that a single, consolidated warehouse with data formatted for US markets could be viewed by analysts and information workers in other countries in their local currency, numeric, and text formats.

There are two primary advantages to this approach. First, it eliminates the need to create a number of localized warehouses, reducing storage space and build times. Second, it allows organizations to extend analytics to the entire enterprise, as each country can now get a localized experience.

The ability to serve up localized views of a particular cube is provided by a new feature in SQL Server 2005 Analysis Services called Translations. Translations allow both the metadata and data to be translated on the fly into the local format, giving users a localized experience with a minimum of effort required compared to setting up separate physical cubes for each locale.

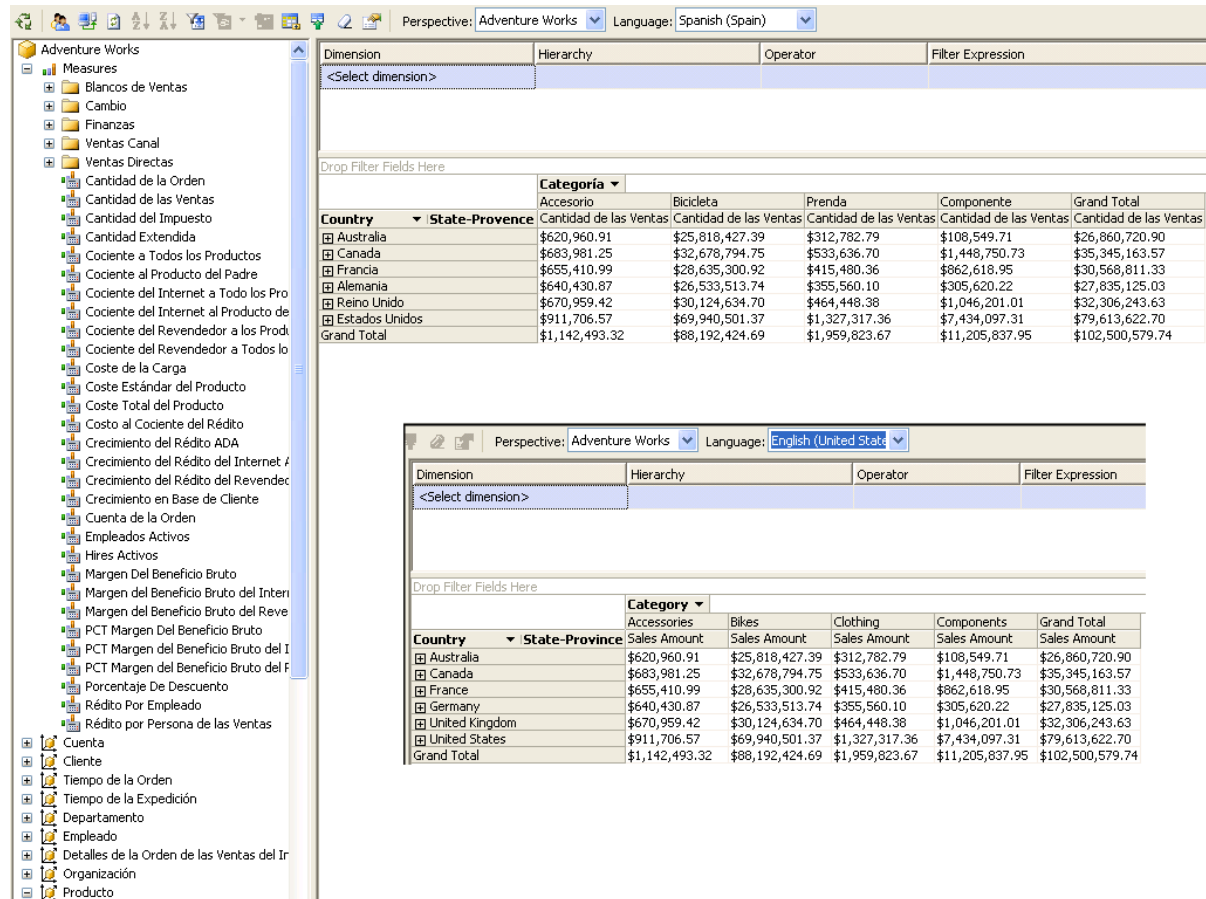


Figure 2 Multi Translation

Problem 5: Real-Time BI

Typical business intelligence solutions share a host of common attributes: the data is consolidated, consistent, read-only, and historical. The historical piece generally means that the data not only goes back in time, but it is not completely up to date. Even in a warehouse that is refreshed nightly, business decision makers are making decisions based on yesterday's information. In situations involving rapidly changing markets and conditions, yesterday's information may not be good enough.

Real-time, or near real-time, BI solutions have existed for some time, but there are typically problems with such solutions. Most near real-time solutions tend to suffer from performance issues, as they are unable to store the analytical data in the optimum format. In addition, most such solutions have a negative impact on the performance of the OLTP systems from which they pull data, as they are constantly hitting those systems to retrieve the latest data.

Solution: Right Information, Right Format, Right Time

Microsoft SQL Server 2005 Analysis Services adds the ability for enterprises to perform near real-time BI, performing analysis without waiting for nightly or weekly refreshes. In addition, the structure of the solution allows this analysis to be performed without impacting the performance of the BI solution. Finally, the performance of the OLTP systems will show little to no impact despite the near real-time BI solution.

SQL Server 2005 Analysis Services achieves this near real-time functionality through a new feature called proactive caching. Proactive caching allows cube builders to set how and when the cube is refreshed. As an example, real-time analytics can be achieved by setting the cache so that as each new transaction enters the relational database, a notification is sent to the UDM. This triggers a refresh of the cube in the background, keeping the cube in synch with the underlying relational data source. Alternatively, the cache can be set to refresh on a timed basis, such as every half hour.

One major benefit of this approach is that companies achieve real-time or near real-time analytics while maintaining the performance of a fully processed data warehouse cube. Administrators can configure the caching to control the latency and update frequency. Proactive caching allows a cube to be queried not just before it is populated but also during and after it is refreshed. If a query comes in before the cube is fully refreshed, Analysis Services 2005 will switch from a MOLAP to a ROLAP mode to answer the query. Microsoft has brought real-time BI to the mainstream while maintaining the performance formerly reserved for fully processed, historical warehouse cubes.

Summary

A number of challenges have plagued business intelligence implementations for years. For example, much of an organization's data exists in a series of silos, be it in a CRM, SCM, ERP, or other application with its own schema and format for the data. The UDM allows businesses to perform complex analytics and reporting against huge data volumes, even across heterogeneous data sources.

BI solutions tend to be historical snapshots that are out of date by the time the analysis are performed. Some BI products support near real-time analytics but often suffer from poor performance and significant overhead.

Many BI solutions also suffer from the need to consolidate far-flung data marts into a consistent warehouse. Other BI implementations suffer from a lack of flexibility in analyzing hierarchical structures or modeling complex relationships between various dimensions.

Finally, warehousing products tend to fall short when dealing with delivering data in localized languages and numeric formats, while at the same time separating the concept of reporting from analysis. This often means multiple data stores in different formats, all to be used for different purposes.

SQL Server 2005 Analysis Services addresses these problems with out-of-the-box solutions to these persistent data warehousing challenges. Data warehouses can be built directly on top of relational schemas found in CRM, SCM, and other data silos without the need to first build a star schema and with little to no impact on the production application. Analysis Services introduces technology to update OLAP cubes in near real-time with minimal impact on the source database while simultaneously speeding up relational reporting against that same source.

Analysis Services further allows for a single warehouse to be built and then presented as a series of separate data marts to departments and users throughout the organization. This allows each department to continue to analyze their subset of data without having to build and distribute a number of separate marts. Further, that single warehouse can be presented in a localized format without the need to build distinct localized versions.

Finally, the Microsoft Unified Dimensional Model combines the benefits of OLTP and OLAP data stores into a single model. OLAP cubes are built on the fly and updated in near real-time. No separate star schema is required. Reporting applications benefit from the caching used to maintain OLAP cubes which allows for faster reporting of the underlying OLTP database.

Microsoft has demonstrated an ongoing commitment to bringing incredibly powerful BI and analytical features to organizations of any size. Using these features, companies can build more powerful, flexible, and cost-effective BI solutions that can be used by a wider audience of users and tools than ever before.

Craig Utley is the Vice President of Development for KIZAN Technologies LLC, where he leads a team focused on business intelligence solutions and enterprise application design and development. He has been working with Microsoft business intelligence products since their inception and has worked on BI and data warehousing solutions for companies throughout the United States. He is an author, conference speaker, and has been chosen as a Microsoft MVP.

This paper was produced in partnership with A23 Consulting.