

Microsoft Corp. SQL Server 2000 Analysis Services

Summary

Microsoft SQL Server 2000 Analysis Services is a middle-tier server for online analytical processing (OLAP) and data mining. Analysis Services and the PivotTable Service provide the capability to design, create, and manage cubes and data-mining models from data warehouses and to provide client access to OLAP data and data-mining data. The Analysis server manages the data; PivotTable Service works with the server to provide client access to the data. With Microsoft introducing this product in late 1998 as SQL Server 7.0 OLAP Services, one might conjecture that Microsoft is too late. Are they? How well does this product stack up against Gartner criteria for products of this type, and is Analysis Services strong enough to challenge entrenched competitors?

--By Alan H. Tiedrich

Note:

Microsoft SQL Server 2000 Analysis Services is compatible with SQL Server version 7.0 OLAP Services. Metadata for cubes that were created in SQL Server 7.0 OLAP Services is upgraded automatically and transparently during the upgrade process; existing cubes do not need to be reprocessed. The existing structures for cubes, roles, shared dimensions, and so on do not need to be changed. Analysis Manager is backward-compatible with SQL Server 7.0 OLAP Services. It is capable of administering both OLAP servers (the server that ships with SQL Server 7.0 OLAP Services) and Analysis servers (the server that ships with SQL Server 2000 Analysis Services) concurrently.

Corporate Headquarters

Microsoft Corp.
One Microsoft Way
Redmond, WA 98052-6399, U.S.A.
Tel: +1 425 882 8080
Fax: +1 425 936 7329
Internet: www.microsoft.com

Overview

The purpose of Microsoft SQL Server 2000 Analysis Services is to provide rapid analytical access to data warehouse data. To accomplish this purpose, Analysis Services creates multidimensional cubes from data in the data warehouse fact and dimension tables. Numerical measures are also summarized into preaggregated values during cube construction. The design of the data warehouse structure can affect how easily these cubes can be designed and constructed. Cubes are stored in multidimensional structures that are

designed for rapid query response, combining preaggregated information with raw fact data to respond to a wide variety of queries. Cubes can contain data summarized, copied, or read directly from the data warehouse. OLAP cubes, dimensions, and partitions are processed to incorporate new or changed data from the data warehouse (OLAP data must be updated after data warehouse data is changed). The method of processing an OLAP object depends on the object and type of change made to the data warehouse, such as data addition, data change, or structural change. Realtime OLAP is an Analysis Services feature that uses realtime cubes to automatically synchronize cube data with changes in the underlying relational database.

The Multidimensional Expressions (MDX) language is used to manipulate multidimensional information in Microsoft SQL Server 2000 Analysis Services. MDX is defined in the OLAP extensions in OLE DB. Similar to SQL in many respects, MDX syntax provides for the retrieval and manipulation of multidimensional data, such as the data stored in cubes on the Analysis server. Analysis Services supports MDX functions in the definitions of calculated members, as well as a full language implementation for building local cubes and querying cube data using the PivotTable Service with OLE DB and Microsoft ActiveX Data Objects (ADO). Additionally, MDX supports the creation and registration of user-defined functions. User-defined functions can be created to operate on multidimensional data and accept arguments and return values in the MDX syntax.

Overview

Product Type	Business Intelligence Platform--Microsoft SQL Server 2000 Analysis Services is a middle-tier server for online analytical processing (OLAP) and data mining.
Product and Version	SQL Server 2000 Analysis Services.
Release Date	September 2000
User Interface	Microsoft Windows 95, 98, NT 4.0, 2000.
Editions	Standard Edition, Personal Edition, Enterprise Edition, Developer Edition.
Server Platforms	Microsoft Windows 2000 and Windows NT 4.0.
Databases Supported	<ul style="list-style-type: none"> • OLE DB drivers for Microsoft SQL Server, Oracle 7.3 and 8i, IBM DB2, JET, Exchange, and others provided by third-party software vendors. • OLE DB for ODBC driver ships with the product--provides access to other databases that are compliant with this standard.
Base Price (US\$)	<ul style="list-style-type: none"> • SQL Server 2000 is priced either per client license or per processor. • Microsoft SQL Server 2000 Standard Edition English North America CD 1 Processor License--4,999 or 10 Clients--2,249, 5 Clients--1,489.
Installed Base	Not available.
Target Markets	Enterprises seeking a business intelligence (BI) platform that runs on Microsoft Windows NT for building BI applications and doing light data mining.

Operating Requirements

Hardware/Software	Requirements
Computer	Intel or compatible (Pentium 133MHz or higher, Pentium PRO, Pentium II, or Pentium III).
Main Memory (RAM)	32MB minimum (64MB recommended).
Disk Drive	CD-ROM drive.
Hard Disk Space (1)	50MB-90MB (130MB for all components, including common files and samples), 12MB for the client only.
Operating System	<ul style="list-style-type: none"> • Microsoft Windows 2000 Server or Microsoft Windows NT Server 4.0 with Service Pack 5 or later (2). • For client components on client computers only, the following systems also qualify: <ul style="list-style-type: none"> --Windows 2000 Professional --Windows NT Workstation 4.0 with Service Pack 5 --Windows 98 --Windows 95 + DCOM95 --Windows 95 OSR2 + DCOM95
Network Software	Windows 2000, Windows NT 4.0, Windows 98, or Windows 95 built-in network software, and TCP/IP (included with Windows).
Online Product Documentation Viewer	Microsoft Internet Explorer version 5.0 or later. (3) Windows NT 4.0 Service Pack 5 or later must be installed before installing Internet Explorer version 5.0.

(1) Setup installs a number of components that can be shared by other applications and may already exist on the computer.

(2) Analysis Services should not be installed on a domain controller; this installation configuration is not supported.

(3) Internet Explorer is required for Microsoft Management Console (MMC) and HTML Help. A minimal installation is sufficient, and Internet Explorer does not need to be the default browser. Internet Explorer is not required for the client-only installation.

Features/Functions--Microsoft SQL Server 2000 Analysis Services

Description	The Analysis Services system includes a server that manages multidimensional cubes of data for analysis and provides rapid client access to cube information. Analysis Services organizes data from a data warehouse or other data source into cubes with precalculated aggregation data to provide rapid answers to complex analytical queries. Analysis Services also supports creating data-mining models from both multidimensional (OLAP) and relational data sources. PivotTable Service, the included OLE DB-compliant provider, is used by Microsoft Excel and applications from other vendors to retrieve data from the server and present it to the user or to create local data cubes for offline analysis.
Key Features Support	<ul style="list-style-type: none"> • Standard Edition and Personal Edition support the following Analysis Services Features: <ul style="list-style-type: none"> --Analysis Services --Custom roll-ups --Actions --Data mining --Unbalanced and ragged hierarchies --Dimension level security --Drill-through --Write-back allocations --Named sets • Enterprise Edition and Developer Edition support all of the above plus: <ul style="list-style-type: none"> --User-defined OLAP partitions --Partition wizard --Linked OLAP cubes --ROLAP dimension support --HTTP Internet support --Calculated cells --Write-back to dimensions --Very Large Dimension support --Realtime OLAP --Distributed partitioned cubes

Analysis Services Architecture

Ease of Use	To make online analytical processing (OLAP) and data-mining technology easier to use, Microsoft SQL Server 2000 Analysis Services provides wizards, editors, tools, and information within Analysis Manager. This console application provides a user interface for accessing Analysis servers and their metadata repositories.
Wizards, Editors, Tools	<ul style="list-style-type: none"> • Cube Wizard • Cube Editor • Dimension Wizard • Dimension Editor • Incremental Update Wizard • Partition Wizard • Storage Design Wizard • Usage Analysis Wizard • Usage-Based Optimization Wizard • Calculated Cells Wizard • Action Wizard • Virtual Cube Wizard • Virtual Cube Editor • Mining Model Wizard • Mining Model Editors • Data Views • OLE DB Data Source Locator Integration • Role Managers

Data Model

Data Storage Options	<ul style="list-style-type: none"> • Multidimensional OLAP (MOLAP)--The underlying data for a cube is stored along with aggregation data in a high-performance multidimensional structure. • Relational OLAP (ROLAP)--The underlying data for a cube is stored along with the aggregation data in a relational database. • Hybrid OLAP (HOLAP)--The underlying data for a cube is stored in a relational database and the aggregation data is stored in a high-performance multidimensional structure.
MOLAP	<ul style="list-style-type: none"> • This option is used to store the data for the cube or partition in a multidimensional structure. The aggregations will be stored with the multidimensional data. • Multidimensional OLAP (MOLAP) storage provides the potential for the most rapid query response times, depending only on the percentage and design of the cube's aggregations. In general, MOLAP is more appropriate for cubes with frequent use and the necessity for rapid query response.
ROLAP	<ul style="list-style-type: none"> • This option is used to keep the data for the cube or partition in the existing relational data store. Aggregations designed for relational OLAP (ROLAP) will also be stored in the relational database, rather than in a multidimensional structure. • ROLAP cubes are also used to deliver realtime OLAP. • ROLAP query response is generally slower than that available with MOLAP or HOLAP. A typical use of ROLAP is for large datasets that are infrequently queried, such as less recent historical data.
HOLAP	<ul style="list-style-type: none"> • This option is used to keep the data for the cube or partition in the existing relational data store and to keep aggregations in a multidimensional structure. • For queries that access summary data, hybrid OLAP (HOLAP) is equivalent to MOLAP. Queries that access base data, such as a drill-down to a single fact, must retrieve data from the relational database and will not be as fast as if the base data were stored in the MOLAP structure. Cubes stored as HOLAP are smaller than equivalent MOLAP cubes and respond faster than ROLAP cubes for queries involving summary data. HOLAP storage is generally suitable for cubes that require rapid query response for summaries based on a large amount of base data.
Partitioned Cube Storage	<ul style="list-style-type: none"> • A cube can be partitioned into separate physical sections. Each partition can be stored in a different mode, in a different physical location, and with a level of aggregations appropriate to the data in the partition. This enables fine-tuning of the performance and data management characteristics of the system. • Partition Merging--a cube's multiple partitions can be combined back into a single physical partition.
Write-Enabled Cubes	A cube can be enabled for write access by multiple simultaneous users.

Hierarchies Supported	<ul style="list-style-type: none"> • Dimensions with balanced or unbalanced hierarchies • Ragged hierarchies
Dimensions	<ul style="list-style-type: none"> • Parent-Child Dimensions--create a dimension based on two dimension table columns that together define parent-child relationships between rows in the dimension table. Parent-child dimensions support balanced, unbalanced, and ragged hierarchies. • Write-Enabled Dimensions--write access by multiple simultaneous users. Users can manipulate the dimension data to see the immediate effect on the cube data.
Virtual Cubes	Cubes can be joined into virtual cubes to access data in the combined cubes without necessitating the construction of a new cube.
Calculated Members	Define new measures and dimension members--create calculated measures and calculated dimension members by combining Multidimensional Expressions (MDX), mathematical formulas, and user-defined functions.
Custom Unary Operators	Custom unary operators use simple math operators, called <i>unary operators</i> , stored in a column to determine how the value of a level member affects the value of the parent.
Custom Roll-Up Formulas and Custom Member Formulas	MDX expressions that determine cube cell values associated with members. Custom roll-up formulas apply to all members of a level, whereas custom member formulas apply to individual level members.
Calculated Cells	Similar to custom member formulas, calculated cells are MDX statements that determine cube cell values associated with a specified group of cells. Calculated cells apply only to specified cells in a cube, whereas custom member formulas must apply to all of the cells for a given member.
Member Properties	Can define properties for dimension members and use data for these properties within a cube.
Virtual Dimensions	A virtual dimension can be created from member properties or levels of another dimension.

Performance and Scalability

Customized Aggregation Options	Using the Storage Design Wizard, the trade-off between system performance and the disk space allocated to storing aggregations can be optimized. (Analysis Services does not require that all aggregations be precalculated, instead resolving queries from the most efficient level of aggregation in the cube. As a result, Analysis Services cubes minimize the impact of the data explosion syndrome inherent in OLAP technology.)
Usage-Based Optimization	The performance of a cube can be tuned to provide quick response to the queries most often executed by directing the Usage-Based Optimization Wizard to design aggregations appropriate to those queries while maintaining reasonable storage requirements.
Data Compression and Storage Optimization	In MOLAP and multidimensional parts of HOLAP structures, storage is not used for empty cells, and a sophisticated data compression algorithm is applied to data that is stored.
Distributed Calculation	PivotTable Service incorporates functionality from the server so that calculations can often be performed on the client instead of the server. Because this distributes the computational load between the server and the client, it increases the capacity of the server, reduces network traffic, and improves performance for the clients.
Partitions	By spreading a cube over multiple files by dividing it into partitions, Analysis Services can then retrieve data in parallel to answer queries. Partitioning enables the management of storage strategy, increases scale with multiple servers, and increases performance.
Linked Cubes	A cube can be stored on a single server and referenced as a linked cube on other servers using NT Integrated Security, a Security Support Provider Interface (SSPI) provider such as Kerberos, or the http and https authentication. Users connected to any of these servers can then access the cube. This approach avoids the more costly alternative of storing and maintaining copies of a cube on multiple servers.
Distributed Partitioned Cubes	Creating distributed partitioned cubes by using remote partitions enables management of a storage strategy by storing a cube's data across multiple servers.
Incremental Updates	A cube can be updated by processing only the data that has been added rather than the entire cube; OLAP cubes can be incrementally updated while they are in use.
LAN, WAN, Internet, and Mobile Scenarios	<ul style="list-style-type: none"> • Intelligent cache management integrates the Analysis server with the PivotTable Service client, minimizing traffic over LAN and WAN connections. PivotTable Service contains an efficient multidimensional calculation engine to further minimize network traffic and to enable analysis of local multidimensional data when the client is not connected to the server. • Microsoft ActiveX controls, Active Server Pages (ASP) scripting, and ActiveX Data Objects (ADO) APIs provide a variety of solutions for querying OLAP data over the Web. • To support mobile scenarios, the user can create local cubes, which can be used without a connection to an Analysis server. Depending on their storage mode, some local cubes can be used without a connection to an Analysis server and without a connection to the cube's data source.

HTTP Authentication of Connections	HTTP or secure HTTP (HTTPS) authentication can be used in conjunction with Microsoft Internet Information Services (IIS) to establish connections to an Analysis server.
Client Support for Windows 95 and Windows 98	PivotTable Service runs on Microsoft Windows 95 and Windows 98, supporting client applications available for these platforms as well as for Microsoft Windows NT 4.0 and Windows 2000.

Integration with Other Components and Programs

Integrated Management Console	Analysis Manager, a graphical administration tool, is a snap-in to Microsoft Management Console (MMC). It provides a common framework and user interface for defining, accessing, and managing Analysis servers and databases.
Integrated Security	Cube and data-mining model access is based on Microsoft Windows NT 4.0 or Windows 2000 security.
OLE DB and ODBC Data Sources	A variety of OLE DB and ODBC data sources can be used, such as Oracle versions 7.3 and 8.0. Multiple sources can be used simultaneously.
Data Transformation Services	Using the Data Transformation Services (DTS) portion of SQL Server Enterprise Manager, packages that process cubes and data-mining models can be created, and prediction queries based on mining models can be run. These packages can be scheduled to execute automatically.
English Query	English Query allows end users to pose questions in English to query relational databases instead of forming a query with an SQL statement. With the integration of Analysis Services and English Query, client applications can also produce answers to users' questions in either a multidimensional display, as well as in traditional SQL rows. When an English Query model has been enabled for Analysis Services, entities and relationships can be associated with OLAP cube information. The English Query Model Editor includes an explicit representation of OLAP cube dimensions, levels, properties, measures, and facts. English Query generates Multidimensional Expressions (MDX) queries. After creating a project using the OLAP Project wizard, users can add cubes to the project; modify cubes, dimensions, properties, and levels; and add entities, relationships, and dictionary entries. The English Query Model Editor appears within the Microsoft Visual Studio version 6.0 development environment. From there, users can choose one of the English Query project wizards, the SQL Project Wizard or the OLAP Project Wizard, to automatically create an English Query project and model. After the basic model is created, it can be refined, tested, and compiled into an English Query application (*.eqd) and then deployed (for example, to the Web).
Meta Data Services	<ul style="list-style-type: none"> • SQL Server 2000 Meta Data Services is used to store the Analysis Services metadata--the information used to define cubes, data-mining models, and other objects on the server--in a relational database. • For each Analysis server, Microsoft SQL Server 2000 Analysis Services creates a repository called the Analysis Services repository to store metadata for the objects of the Analysis server (cubes, dimensions, and so on). By default, this repository is a Microsoft Access database on the server computer where Analysis Services is installed, but this repository can be migrated to a SQL Server database on the same or another server computer.
Excel and Visual Basic Functions	<ul style="list-style-type: none"> • Many functions in the Microsoft Excel worksheet library can be included in Multidimensional Expressions (MDX), which is automatically registered if installed on the computer with Analysis Services. • Many functions in the Microsoft Visual Basic for Applications Expression Services library--which is included with Analysis Services and automatically registered--can be included also.
Server-Side Cache	User queries, metadata, and data are stored in the Analysis server cache, making it possible to answer new queries by calculating answers from cached data rather than retrieving data from the disk.
Client-Side Cache	<ul style="list-style-type: none"> • Client applications connect to the Analysis server through the client-based PivotTable Service component. Because PivotTable Service receives metadata with data from the server in response to a query, it can often use data in the client cache to calculate the answer to subsequent queries without sending a new query to the server. • PivotTable Service shares much of the same functionality as the server, enabling it to bring the server's multidimensional calculation engine, caching features, and query management directly to the client computer. This client/server data management model optimizes performance and minimizes network traffic.

Supported APIs and Open Architecture

OLE DB	<ul style="list-style-type: none"> • Analysis Services is designed to meet the OLAP-specific requirements of the OLE DB 2.0 and later specification. • Analysis Services is also designed to meet the requirements of the OLE DB for Data Mining specification, which addresses data mining-specific provisions.
ADO	Analysis Services is compatible with Microsoft ActiveX Data Objects (ADO) and its extension for multidimensional objects, ADO (Multidimensional) (ADO MD).
User-Defined Functions	The list of built-in functions can be extended by creating libraries of functions using Component Object Model (COM) automation languages, such as Microsoft Visual Basic or Microsoft Visual C++. These libraries can be registered and their functions used in calculated member definitions and other expressions written in Multidimensional Expressions (MDX). This architecture enables the addition of customized analysis tools.
Decision Support Objects	The server object model, Decision Support Objects (DSO), can be used to create applications that define and manage cubes, data-mining models, and other objects. This object model can be used to extend the functionality of Analysis Manager or to automate the ongoing maintenance of the system.
Add-in Support	Analysis Services Add-in Manager interface can be used to create applications that extend the functionality of the Analysis Manager user interface. The Analysis Services Add-in Manager interface and DSO can be used to create custom extensions, dialog boxes, wizards, and other applications that integrate with Analysis Manager.

Server and Client Architecture

Server Architecture	<p>Analysis Services provides server capabilities to create and manage OLAP cubes and datamining models and to provide this data to clients through PivotTable Service.</p> <p>Server operations include:</p> <ul style="list-style-type: none"> • Creating and processing cubes from relational databases, usually in data warehouses. • Storing cube data in multidimensional structures, in relational databases, or in combinations of both. • Creating data-mining models from cubes or from relational databases, usually in data warehouses. • Storing data for data-mining models in multidimensional structures, relational databases, or in Predictive Model Markup Language (PMML), which is a standardized XML format.
Client Architecture/PivotTable Service	PivotTable Service, included as part of Analysis Services, is the primary interface for applications interacting with Microsoft SQL Server 2000 Analysis Services in order to accomplish such tasks as connecting to a cube or data-mining model, querying a cube or data-mining model, and retrieving schema information. It is used to build client applications that interact with multidimensional data. PivotTable Service also provides methods for online and offline data-mining analysis of multidimensional data and relational data.

Object Architecture

Analysis Server	<p>The following section summarizes the objects used to administer Microsoft SQL Server 2000 Analysis Services.</p> <p>The server component of Analysis Services, designed specifically to create and maintain multidimensional data structures and to provide multidimensional data in response to client queries.</p> <ul style="list-style-type: none"> • Each Analysis server has a repository called the Analysis Services repository. This repository stores the metadata (that is, definitions) of the objects defined on the Analysis server. • By default, this repository is a Microsoft Access database on the server computer where Analysis Services is installed. This repository can be migrated to a SQL Server database on the same or another server computer. • Each Analysis server has an associated Data folder, which stores multidimensional structures for the objects defined on the Analysis server. These structures are contained in files that are created when the objects are processed. These structures are referenced to resolve queries sent to the Analysis server. Some of these structures contain aggregations.
Databases	Databases serve as containers for related data sources, cubes, dimensions, data-mining models, and the objects they share.

Data Sources	Data sources store the specification of the information necessary to access source data for an object such as a cube. Although the term <i>data source</i> is sometimes used to refer to the source data itself, in this topic it refers to the data source object, which is used by Microsoft SQL Server 2000 Analysis Services to establish connections to the source data. A data source specifies an OLE DB provider and settings for the other properties in the connection string used to access the source data.
Dimensions	Dimensions serve as a structural attribute of a cube. A dimension is an organized hierarchy of categories (levels) that describe data in the fact table. These categories describe similar sets of members upon which the user wants to base an analysis.
Levels and Members	Levels are elements of a dimension hierarchy. Levels describe the hierarchy from the highest (most summarized) level to the lowest (most detailed) level of data.
Measures	Measures are, in a cube, a set of values that are based on a column in the cube's fact table and are usually numeric. In a cube, measures are the central values that are analyzed.
Cubes	Cubes contain a set of data that is usually constructed from a subset of a data warehouse and is organized and summarized into a multidimensional structure defined by a set of dimensions and measures.
Partitions	Partitions are the storage containers for data and aggregations of a cube.
Aggregations	Aggregations are defined as a table or structure containing precalculated data for a cube.
Roles	Roles contain a set of Microsoft Windows NT 4.0 or Windows 2000 user accounts and groups with the same access to Analysis Services data.
Commands	Commands are used to hold an administrator-defined command that is automatically executed when a client accesses a database, cube, or role. Commands include calculated members, named sets, and actions.
Member Properties	Member properties contain information about the members of a dimension level in addition to that contained in the dimension.
Data-Mining Models	Data-mining models contain a virtual structure that represents the grouping and predictive analysis of relational or online analytical processing (OLAP) data.
Data-Mining Columns	Data-mining columns contain a structure that is used to define the content of a data-mining model. A column can contain data or nested columns.

Security and Authentication

Description	<p>Access to data managed by Microsoft SQL Server 2000 Analysis Services can be restricted:</p> <ul style="list-style-type: none"> • Limits to which administrators are permitted to access Analysis Services data through Analysis Manager and perform administrative functions. • Limits for end users who access data on the Analysis server through client applications: <ul style="list-style-type: none"> --Specify which end users can access data and the types of operations they can perform. --Control end-user access at various levels of Analysis Services data, including the cube, dimension, and cube cell. • Administrator security is controlled using the Microsoft Windows NT 4.0 or Windows 2000 group named OLAP Administrators. • End-user security is controlled using: <ul style="list-style-type: none"> --Authentication during connection to the Analysis server. --Database, cube, and mining model roles defined in Analysis Manager.
End-User Security	End-user security is concerned with users' access to data on the Analysis server through client applications such as Microsoft Excel. It affects the ability of users to connect to the Analysis server, which data they can access, and whether they have read or read/write access.
Database, Cube, and Mining Model Roles	<p>A role (also called a security role) defines a set of Microsoft Windows NT 4.0 or Windows 2000 user accounts and groups with the same access to Microsoft SQL Server 2000 Analysis Services data. Roles are used to implement end-user security by controlling access to data on the Analysis server by users connected with client applications.</p> <p>Analysis Services includes three types of roles:</p> <ul style="list-style-type: none"> • Database role--can be assigned to multiple cubes or mining models in the database, thereby granting users of the role access to these cubes or mining models. • Cube role--applies to a single cube. • Mining model role--applies to a single mining model.
Levels of End-User Security	End-user security can be enforced at several levels of detail. These levels are (from least detailed to most detailed): server (Analysis Server), database, cube/mining model, dimension member, cell.

Processing Cubes

Description	<p>When a cube is processed, the aggregations designed for the cube are calculated, and the cube is loaded with the calculated aggregations and data. Processing a cube involves reading the dimension tables to populate the levels with members from the actual data, reading the fact table, calculating specified aggregations, and storing the results in the cube. After a cube has been processed, users can query it.</p>
Cube Processing Options	<p>Each of the following three processing options is appropriate in different circumstances:</p> <ul style="list-style-type: none">• Full Process--used to perform a complete load of the cube. All dimension and fact table data is read, and all specified aggregations are calculated.• Incremental update--appropriate when new data is to be added to a cube, but existing data has not changed and the cube structure remains the same.• Refresh data--causes a cube's data to be cleared and reloaded and its aggregations recalculated. This option is appropriate when the underlying data in the data warehouse has changed, but the cube's structure remains the same. <p>In addition to these three mutually exclusive options, a fourth option--incrementally update the dimensions of this cube--can be selected in conjunction with any of these options. This option allows incrementally updating of the cube's dimensions as part of the cube processing and is appropriate when rows have been added to any of the cube's dimension tables since the cube or dimension was last processed.</p>
Realtime OLAP Cube Updating	<p>For fact table data that changes often, it is impractical to use a regular cube, which must be reprocessed to recreate its aggregations each time the data supporting the dimensions or partitions of a cube changes, because this can lock the cube for long periods of time. Realtime OLAP enables ROLAP dimensions and partitions to automatically refresh themselves when data in their underlying dimension or fact tables changes. When working in concert with SQL Server 2000 as the relational data source, Analysis Services is notified about updates to dimension or fact tables associated with specific ROLAP dimensions or partitions enabled for realtime updates as those changes happen. If the Analysis server finds that a change to a dimension or fact table has occurred, it can respond to the notifications simply by flushing the Analysis server cache. (Simply adding dimension members or facts to a ROLAP cube, without changing its structure, requires no reprocessing—that is why it can perform reasonably well.)</p>
Incremental Updates	<p>Incremental updates support keeping the contents of a cube current without requiring that the cube be reprocessed in full when new data is added. An incremental update involves creating a temporary partition, filling it with updated source data, processing the temporary partition, and then merging it into another partition in the cube.</p> <p>Data to be added to a cube can come from the original fact table or from a separate fact table with a structure identical to the original.</p>
Merging Partitions	<p>To ensure that partitions can be merged, certain restrictions on merging partitions must be considered when the partitions are created. Partitions can be merged only if they meet all the criteria listed here:</p> <ul style="list-style-type: none">• They are in the same cube.• They have the same structure (the default situation).• They are stored in the same mode (MOLAP, HOLAP, or ROLAP).• They contain identical aggregation designs.• Remote partitions can be merged only with other remote partitions that are defined with the same remote Analysis server.

MDX Functions in Analysis Services

Description	Microsoft SQL Server 2000 Analysis Services provides for the use of functions in Multidimensional Expressions (MDX) syntax. MDX is a syntax used for defining multidimensional objects and querying and manipulating multidimensional data. Functions can be used in any valid MDX statement and are often used in queries, calculated members, and custom roll-up definitions. There are three types of functions in MDX (Basic/Advanced, Registered Functions, and User-Defined).
Basic/Advanced	<ul style="list-style-type: none"> • Basic MDX commands query multidimensional objects, such as cubes, to extract simple data and return multidimensional datasets. • Advanced MDX provides additional functionality to create named sets and calculated members and write information back to dimensions and cells. • MDX Function Groups--array, dimension, hierarchy, level, logical, member, numeric, set, string, tuple. • Numeric Functions--aggregate, average, correlation, count, covariance, regression, maximum, media, minimum, predict (data mining), rank, standard deviation, sum, variance. • Custom Roll-ups--in addition to the standard roll-up (that is, aggregate) functions Sum, Min, Max, Count, and more sophisticated custom roll-up functions can be defined for any given member in the CREATE CUBE statement.
Registered Function Libraries	<ul style="list-style-type: none"> • Microsoft SQL Server 2000 Analysis Services supports many functions in the Microsoft Visual Basic for Applications Expression Services library. This library is included with Analysis Services and automatically registered. • Microsoft SQL Server 2000 Analysis Services supports many functions in the Microsoft Excel worksheet library, which is automatically registered if installed on the computer with Analysis Services.
User-Defined Functions with MDX Syntax	User-defined functions can accept arguments and return values in the MDX syntax. User-defined functions can be created using Component Object Model (COM) automation languages such as Microsoft Visual Basic or Microsoft Visual C++. A user-defined function can be developed using any tool capable of generating Microsoft ActiveX libraries.

Data Mining

Data Mining Key Features	<p>Data analysis and prediction capabilities.</p> <ul style="list-style-type: none"> • The Analysis Services algorithms can train data-mining models with data from any relational data source that supports OLE DB access, as well as from multidimensional cubes created with Analysis Services. • Extensibility allows Analysis Services to be used with third-party tools such as mining model viewer components, providing flexibility and enhancement. • OLE DB support has been enhanced with the OLE DB for Data Mining specification to provide integration of third-party data-mining algorithms. • Data mining is integrated into PivotTable Service following the same pattern that was used in Analysis Services. In general, a data-mining model is treated like a cube by PivotTable Service: users can create local data-mining models, retrieve information from server cubes, and so on.
Data-Mining Model	<p>A data-mining model is a virtual structure that represents the grouping and predictive analysis of relational or multidimensional data. In many aspects, the structure of a data-mining model resembles the structure of a database table. However, while a database table represents a collection of records, or a record set, a data-mining model represents an interpretation of records as rules and patterns composed of statistical information referred to as cases. The structure of the data-mining model represents the case set that defines the data-mining model, while the data stored represents the rules and patterns learned from processing case data.</p> <p>Cases and case sets: for example, a collection of related records for a single customer is referred to as a case, and the same collection of related records for a group of customers is referred to as a case set. The case set is simply a way of viewing the physical data. Because of the innately hierarchical nature of such information, the data-mining model stores the representation of a case set as a collection of data-mining columns.</p>

Data-Mining Model Structure	<ul style="list-style-type: none"> • The structure of a data-mining model is defined primarily by a set of data-mining columns and a data-mining algorithm. The data-mining model content, created by the training process, is stored as data-mining model nodes. • The data-mining algorithm uses the data-mining column definitions to generate a predictive model by running the algorithm on training data submitted to the data-mining model. The data-mining model then stores the results obtained from analyzing the training data. Even though large amounts of training data may be inserted into a data-mining model, the training data itself is not stored. Only the analysis information gained by processing that data and the distinct column values used as part of the analysis are stored as data-mining model content.
Training a Data-Mining Model	To determine the relative importance of each attribute in a data-mining model, the model goes through a process known as mining model training. During training, data is supplied to the model for analysis. The data-mining algorithms used by the model then examine the training dataset in a variety of ways to test it so that it can draw some conclusions about classification and prediction of the data.
Data-Mining Columns	Data-mining columns are used to define the inputs and outputs used by a data-mining model. The data-mining column also provides a standard structure against which familiar SQL syntax, such as INSERT for training data and SELECT for predictive analysis, can be used.
Data-Mining Model Nodes	When a data-mining model is built and trained, the resulting data-mining model content is stored as data-mining model nodes. A node stores the attributes, description, probabilities, and distribution information for the model element it represents, as well as any cardinality information the node may possess in relation to other nodes.
Data-Mining Algorithms Supported	<p>Decision tree--a decision tree is a form of classification shown in a tree structure in which a node in the tree structure represents each question used to further classify data.</p> <p>Clustering--the classification of data into groups based on specific criteria.</p> <p>Two data-mining algorithms are included with Analysis Services: Microsoft Decision Trees and Microsoft Clustering. The decision trees algorithm is based on the notion of classification. The clustering algorithm uses an expectation-maximization method to group records into clusters (or segments) that exhibit some similar, predictable characteristic.</p> <p>Third-party data-mining algorithm providers can be added using OLE DB for Data Mining.</p> <p>Microsoft SQL Server 2000 Analysis Services supplies a number of functions that retrieve and manipulate statistical information from a data-mining model, including BottomCount, BottomPercent, BottomSum, Cluster, ClusterDistance, ClusterProbability, Predict, PredictAdjustedProbability, PredictHistogram, PredictProbability, PredictStdev, PredictSupport, PredictVariance, RangeMax, RangeMid, RangeMin, Sub-SELECT, TopCount, TopPercent, TopSum.</p>
Data-Mining Functions	

PivotTable Services

Description	<ul style="list-style-type: none"> • PivotTable Service is an OLE DB provider for multidimensional data and data-mining operations. This means that it provides OLE DB functionality for applications that need access to multidimensional data and data-mining services. By providing support for a subset of SQL and MDX, PivotTable Service enables applications to retrieve tabular and multidimensional data. The data can be displayed, included in a local cube, analyzed using sophisticated data-mining algorithms, or updated. • Offers connectivity to the multidimensional data managed by Analysis Services, other OLE DB-compliant providers, and to non-OLAP relational data sources. • The execution speed of these complex operations makes it possible to perform sophisticated analyses on the client computer itself. This, in turn, allows remote client applications to function independent of a high-speed network, intranet connection, or physical presence at a geographical location. • PivotTable Service also supports data definition language (DDL) in the connection string of the client application so that offline clients can create and modify local cubes at runtime and define temporary multidimensional objects for use in analysis.
-------------	--

Key Features

- As a stand-alone provider, PivotTable Service provides client applications with the ability to create local cube files and mining models from relational and multidimensional sources. Client applications can connect to a local cube and execute queries using MDX without interacting with the full-scale Analysis server.
- PivotTable Service can be used in a variety of development environments. Both Microsoft Visual Basic and Visual C++ developers can use either the Microsoft ActiveX Data Objects (Multidimensional) (ADO MD) object library or the OLE DB for OLAP Component Object Model (COM) interfaces to create client applications.
- PivotTable Service communicates with the Analysis server and provides interfaces that client applications can use to access OLAP data and data-mining data on the server. Client applications connect to PivotTable Service using OLE DB interfaces for C++ or the Microsoft ActiveX Data Objects (ADO) object model for Component Object Model (COM) automation languages such as Microsoft Visual Basic.
- PivotTable Service can also create local cube files that contain data from a cube on the server or from OLE DB relational databases. Local cubes can be stored as multidimensional cube files on the client computer. Local cubes can be used offline with PivotTable Service for portable analysis. That is, to query local cubes, a connection to the Analysis server is not required; connection to the local cubes' data sources is not required if the local cubes have a multidimensional OLAP (MOLAP) storage mode.
- PivotTable Service can also create local data-mining model files that contain models processed from cubes or tables on the server or from OLE DB relational databases. Local data-mining models can be stored on the client computer. Local data-mining models can be used offline with PivotTable Service for portable analysis. That is, to use local data-mining models, a connection to the Analysis server is not required.

Developing Client Applications

- PivotTable Service is the primary method of communication between a client application and a multidimensional data source or data-mining model, such as Microsoft SQL Server 2000 Analysis Services. It is used by applications provided by Microsoft (such as the Microsoft Excel PivotTable and PivotChart features) and by applications provided by third-party vendors. Programmers who want to develop custom client applications for Analysis Services must use PivotTable Service.
- PivotTable Service client applications can retrieve, display, and manipulate data from multidimensional sources, such as OLAP cubes, or from data-mining models.
- PivotTable Service client applications can also create cube files and data-mining models on the local computer and populate them with data derived from an OLE DB tabular provider such as SQL Server or an OLAP cube.
- After a cube or mining model is created, client applications can use PivotTable Service to browse and analyze the data contained in it.
- In the case of OLAP cubes, client applications can conduct what-if analyses using write-back and cell allocations.
- In the case of local multidimensional OLAP (MOLAP) cubes, such analyses can be conducted even if the client application is not connected to the original data source. In this case, PivotTable Service takes the place of the Analysis server by providing many of the functions of the server on the local computer.
- The data source for creating a local cube file or mining model can be any OLE DB data provider, such as SQL Server or Analysis Services.

Data-Mining Support

PivotTable Service supports data mining by providing support services that are very similar to the services it provides for online analytical processing (OLAP). For example, PivotTable Service can create and maintain local data-mining models just as it can create and maintain local cubes. To create a data-mining model on an Analysis server, the Decision Support Objects (DSO) must be used.

Client Cache

PivotTable Service maintains a local cache on the client computer. When PivotTable Service executes a query, the cache is used to store the data locally.

Transactions in Analysis
Services

PivotTable Service supports transaction management for allocations and write-backs to cubes on the Analysis server. An allocation or a write-back to a cube changes a cached copy of that cube in order to analyze the effects of the change. This transaction process enables users to perform what-if analysis.

Client Operations

- Retrieving Schema Information--Microsoft ActiveX Data Objects (Multidimensional) (ADO MD), ADO, or OLE DB can be used to retrieve schema row sets using PivotTable Service.
- Retrieving Data--there are two methods for retrieving data with Microsoft Visual Basic using PivotTable Service: use the Microsoft ActiveX Data Objects (Multidimensional) (ADO MD) Cellset object or the ADO DB Command and Recordset objects. OLE DB can also be used to retrieve data from a cube or data-mining model.
- Updating Information in a Cube--PivotTable Service supports a number of methods of updating the contents of both server cubes and local cubes. Transaction support for write-back operations is provided by the transaction methods of the Microsoft ActiveX Data Objects (ADO) Connection object.

Analysis Manager

A console application that provides a user interface for accessing Analysis servers and their metadata repositories.

Use Analysis Manager to:

- Administer Analysis servers. Multiple users can administer an Analysis server using Analysis Manager. Locking is applied only to the objects being edited and their dependent objects.
- Create databases and specify data sources.
- Build and process cubes.
- Create and process data-mining models.
- Specify storage options and optimize query performance.
- Manage security.
- Browse data sources, shared dimensions, security roles, and other objects.
- Find links for third-party client applications, support resources, Help updates, and product news.
- Obtain information about the complete SQL Server 2000 product.

This step-by-step Analysis Manager tutorial guides the user through building, deploying, managing, and enhancing an OLAP cube. This tutorial also provides instruction on how to create a data-mining model.

Data Transformation Services

- Data Transformation Services (DTS) in Microsoft SQL Server 2000 can be used to process cubes, data-mining models, and other objects and to create prediction tasks based on mining models. For processing activities, a DTS task called the Analysis Services Processing task is provided, and for mining model predictions, the Data Mining Prediction Query task is provided.
 - The Analysis Services Processing task in a Data Transformation Services (DTS) package is used to perform processing of one or more cubes or other objects defined in Microsoft SQL Server 2000 Analysis Services. Use DTS to extract, transform, and consolidate data from disparate sources into single or multiple destinations.
-

Programming Analysis Services Applications

Description

Analysis Services provides support for creating and integrating custom applications that enhance an OLAP and/or data-mining installation.

An object model, Decision Support Objects (DSO), provides support for the Analysis Manager user interface and for custom applications that manage OLAP metadata and control the server. An API enables your applications to extend and interact with the user interface. PivotTable Service provides access to OLAP data from the server and the ability to create local cubes.

Applications That Can Be Created

- Manage the Analysis server and create and maintain OLAP and data-mining objects such as cubes, dimensions, security roles, and data-mining models.
- Extend the user interface by adding new objects to the object tree pane and by adding and responding to new menu choices.
- Connect to the Analysis server, query data in cubes, and create local cubes.
- Combine any or all of these functions.

Analysis Services Component Tools

Analysis Services provides three component tools that can be used with custom applications: Decision Support Objects (DSO), Add-ins Interface and Objects, and PivotTable Service.

Add-ins Interface and Objects	Applications can be created that interact with and enhance the Analysis Services user interface. Analysis Manager can call various routines in the application in response to user activity in the user interface. The Analysis Manager user interface is implemented as an add-in and is called by the Analysis Services Add-in Manager in the same way that the custom add-in will be called. Add-ins can use other Analysis Services component tools to enhance, augment, and automate an Analysis Services installation. Multiple custom add-ins can be registered and operating at the same time.
Decision Support Objects	Analysis Services provides a wide range of OLAP and data-mining functionality. To access such a rich, extensible, wide-ranging set of features in a simple, straightforward fashion, the Decision Support Objects (DSO) library supplies a hierarchical object model for use with any development environment that can support Component Object Model (COM) objects and interfaces, such as Microsoft Visual C++, Microsoft Visual Basic, and Microsoft Visual Basic Scripting Edition.
The Decision Support Objects (DSO) Library	The Decision Support Objects (DSO) library of Analysis Services provides a robust set of Component Object Model (COM) objects and interfaces that can be used to create applications that can programmatically administer Analysis Services objects. This includes managing Analysis Services objects, such as servers, databases, data sources, dimensions, cubes, mining models, and roles. It also includes administering security, process cubes and mining models, and so on.
DSO Applications	<ul style="list-style-type: none"> • A custom application can use DSO in combination with other component tools of Analysis Services to enhance, augment, and automate an Analysis Services installation. • Custom applications can use DSO to control and automate functionality on the server. Applications can also create and maintain OLAP objects such as cubes, dimensions, and roles.
Developing DSO Applications	<ul style="list-style-type: none"> • DSO exposes the object model for the Analysis server. The DSO object model consists of interfaces, objects, collections, methods, and properties. • The user can develop DSO applications in Microsoft Visual Basic and other languages that support the Component Object Model (COM). The DSO object model was developed in Visual Basic and is easiest to use with that language.

Analysis

The Microsoft Corp. SQL Server 2000 Analysis Services (Analysis Services) product, part of SQL Server 2000, provides On-Line Analytical Processing (OLAP) capabilities. Microsoft Analysis Services is one of the leading products in Gartner's Business Intelligence (BI) Platforms category. Products in this category met certain criteria (See the "*Business Intelligence Platforms*" sidebar), which enable them to support IT professionals in building and deploying, as well as managing the execution of BI applications. Analysis Services is a back-end product, which is used by, but transparent to, end users. It provides server capabilities to create and manage OLAP cubes and data-mining models and to provide this data to clients through the PivotTable Service. Another method of querying Analysis Services databases is by using English Query. Analysis Services includes numerous wizards and editors that simplify using the system for DBAs and other IT professionals.

Analysis Services supports end users who need to analyze multidimensional data (data stored in "cubes," a term used for multidimensional data stores) and/or do "light" data mining. Analysis Services provides server-based capabilities to create and manage multidimensional data structures (i.e., OLAP cubes) and data-mining models and to provide multidimensional data to clients in response to the client queries through the PivotTable Service. These server-based precalculated cubes contain aggregated data (and sometimes detailed data) that can be accessed by many simultaneous users, with good response times (according to Microsoft, AT&T has 11,000 users accessing 90 cubes). Each Analysis Server has a metadata repository--the Analysis Services repository, which by default is stored in a Microsoft Access database on the server computer where Analysis Services is installed--and a Data folder that stores multidimensional structures or cubes. Optionally, the Analysis Services repository can be migrated to a SQL Server database on the same or another server computer. Data can be accessed from relational (ROLAP) and multidimensional (MOLAP), as well as hybrid relational and multidimensional (HOLAP) data stores.

Analysis Services also provides two data-mining algorithms, which can be applied to data from relational and multidimensional data stores to create data-mining models.

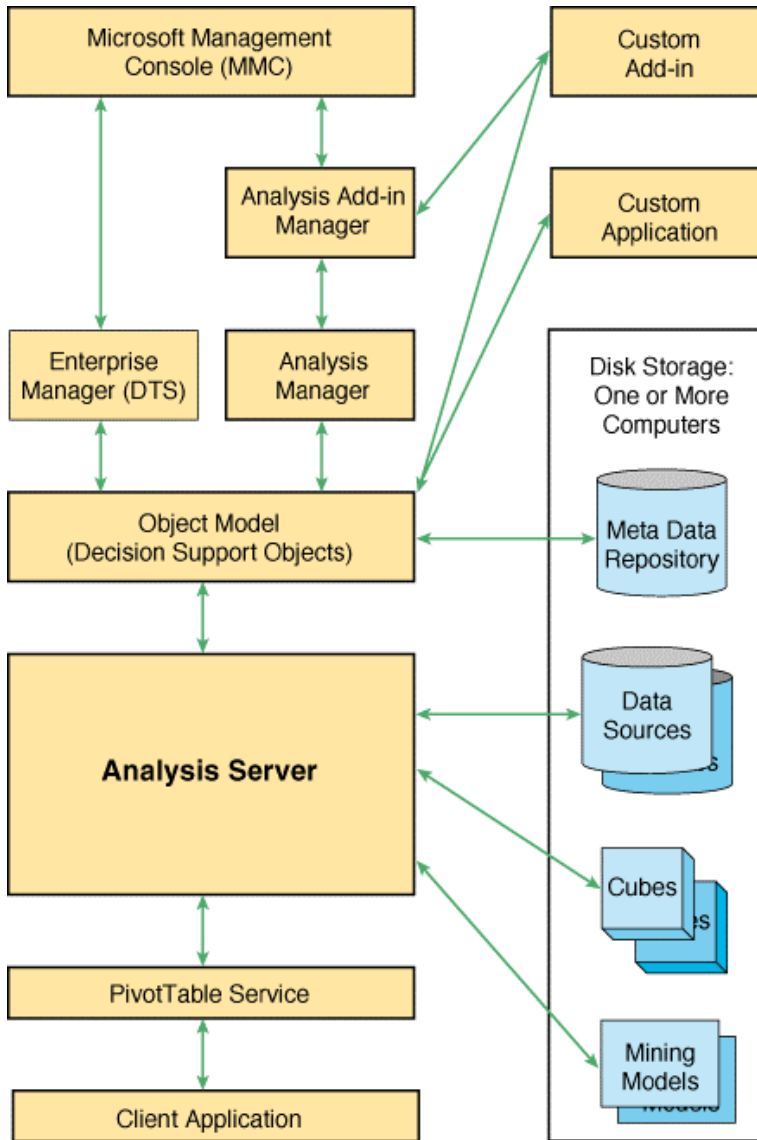


Figure
Microsoft SQL Server 2000 Analysis Services Architecture
Source: Microsoft Corp.

Although its identification as SQL Server 2000 Analysis Services seems to imply that it is integrated with the SQL Server 2000 RDBMS, this is not the case. It really is a stand-alone product that is a member of the SQL Server 2000 product set; however, a customer is required to license SQL Server 2000 in order to get Analysis Services. In fact it could operate independent of the SQL Server RDBMS. Why has Microsoft packaged it this way? First of all, SQL Server 2000 RDBMS customers will have Analysis Services in their possession and so may end up using it (rather than a competitive product). This is the same theory behind packaging productivity applications together as Office 2000. As a corollary, Microsoft wants to make BI analysis accessible to users throughout an enterprise, in contrast with competitive products, which have been targeted at environments having small numbers of professional business analysts. Also,

because of favorable pricing of SQL Server versus competitors' RDBMSs, the package may be appealing to prospective customers, provided they are using Windows NT for their database server.

SQL Server 2000 editions include the Enterprise Edition, the Standard Edition, the Personal Edition, the Developer Edition, and the Evaluation Edition. There also is a Windows CE Edition. Along with the SQL Server engine and Analysis Services, English Query, Meta Data Services (extends and renames the former repository component known as Microsoft Repository), and Data Transformation Services (DTS) are also included on the Microsoft SQL Server 2000 CD-ROM. The SQL Server RDBMS and Analysis Services each can be used mutually exclusive of the other. DTS, integrated into Microsoft SQL Server 2000, can extract, transform, and consolidate data from disparate sources into single or multiple destinations. It can be used to process cubes, data-mining models, and other objects and to create prediction tasks based on mining models. English Query allows end users to pose questions in English (it is based on a semantic model, not a keyword model) instead of forming a query with an SQL statement to query relational or multidimensional databases. English Query generates MDX or SQL queries, depending on the data source. The English Query Model Editor includes an explicit representation of OLAP cube dimensions, levels, properties, measures, and facts. Integrated with Analysis Services, English Query can support client applications that can produce answers to users' questions in either a multidimensional display, as well as in traditional SQL rows.

Analysis Services Data Models

Analysis Services supports cube partitioning, which can result in improved performance, as well as the flexibility of storing updates in a separate partition. Each partition can be stored in a different mode or a combination of OLAP modes, in a different physical location, and can have varying levels of aggregation appropriate to the data in each partition. Multiple partitions can be merged into a single physical cube using the Analysis Manager if they have the same storage mode and aggregation design. By allowing different levels of aggregation in each partition, Analysis Services provides the ability to tune data management and performance to the application characteristics. Multiple cubes can be joined into virtual cubes (no storage requirement) to access data in the combined cubes without necessitating the construction of a new cube and while still maintaining the optimal design for each cube.

Dimensions can be balanced and unbalanced, and ragged hierarchies are supported. Dimensions can be write-enabled, enabling multiple simultaneous users to change and manipulate dimension data, with users able to see the effect on cube data. It is also possible to create calculated measures and calculated dimension members by combining MDX, mathematical formulas, and user-defined functions. Properties can be defined for dimension members, and data can be assigned to these properties and queried. A virtual dimension can be created from member properties or levels of another dimension. MDX expressions that determine cube cell values associated with members can be calculated by custom roll-up formulas that apply to all members of a level, and custom member formulas that apply to individual level members.

Analysis Services offers several cube-processing options, including full processing to perform a complete cube load, incremental update of new data, and refreshing all of the data in the cube. In conjunction with any of these options, it is possible to concurrently update the cube's dimensions. Analysis Services also provides a realtime OLAP feature that uses realtime cubes to automatically synchronize cube data with live updates to the underlying relational database. Realtime cubes can be used for applications that need to monitor and analyze live data and are intended to extend OLAP capabilities rather than replace traditional cube designs and applications. Realtime OLAP enables ROLAP dimensions and partitions to automatically refresh themselves when data in their underlying dimension or fact tables changes. When working in concert with SQL Server 2000 as the relational data source, Analysis Services is notified about updates to dimension or fact tables associated with specific ROLAP dimensions or partitions enabled for realtime updates as those changes happen. If the Analysis Server finds that a change to a dimension or fact table has occurred, it can respond to the notifications by flushing the Analysis server cache.

Although Microsoft was late in coming to the OLAP market, it has delivered a very comprehensive offering in Analysis Services. It is one of the few products of its type that offer all three types of OLAP storage--MOLAP, ROLAP, and HOLAP (See the "*MOLAP, ROLAP, and HOLAP Data Stores*" sidebar). This means that it can be applied to a broad spectrum of applications, from financial

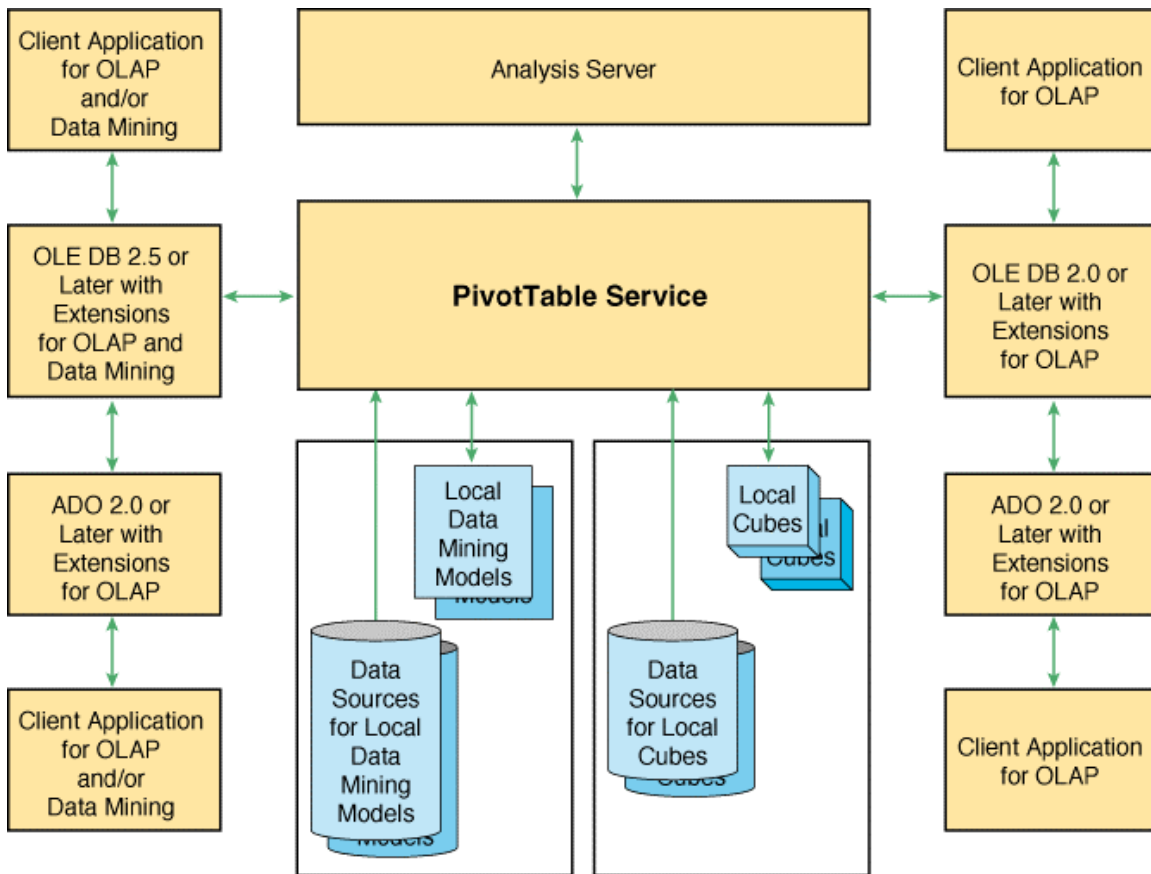
through customer relationship management and other sales and marketing applications. To satisfy the requirements for applications like budgeting and forecasting, Analysis Services provides write-back support so that data in a cube can be updated by multiple simultaneous users. Such write-back occurs to a physically separated cube partition; the write-back data is automatically applied when the cube data is read.

Analysis Services is open to the degree that any Microsoft product will allow and according to its own "open" standards. Analysis Services is designed to meet the OLAP-specific requirements of the OLE DB 2.0 and later specification. This makes it accessible by any front-end BI tool or application, like AlphaBlox 3 applications, BusinessObjects, Cognos EBI Platform, Knosys Proclarity, Brio.Enterprise, and other products from a vibrant third-party market.

End-User Access--PivotTable Service

PivotTable Service, included as part of Analysis Services, is the primary method of a client application for querying a multidimensional data source or data-mining model, such as Microsoft SQL Server 2000 Analysis Services multidimensional cubes and data-mining models. As an OLE DB provider for multidimensional data and data-mining operations, PivotTable Service provides OLE DB functionality to access multidimensional data and data-mining services. PivotTable Service client applications can create cube files and data-mining models on the local computer and populate them with data derived from an OLE DB tabular provider such as SQL Server or an OLAP cube. By providing support for a subset of SQL and MDX, PivotTable Service enables applications to retrieve tabular and multidimensional data. The data can be displayed, included in a local cube; analyzed using sophisticated data-mining algorithms; or updated. The execution speed of these complex operations makes it possible to perform sophisticated analyses on the client computer itself. PivotTable Service also supports data definition language (DDL) in the connection string of the client application so that offline clients can create and modify local cubes at runtime and define temporary multidimensional objects for use in analysis.

In stand-alone mode, PivotTable provides client applications with the ability to create cubes and data-mining models from relational or multidimensional data sources--cubes or tables on the server or from OLE DB relational databases--which can then be queried by users in local mode. Microsoft Excel and third-party applications use the PivotTable Service to retrieve data from the server for analysis by the user or to create local data cubes for offline analysis. Developers can use the PivotTable Service in creating applications that connect to the local cube and execute queries using MDXs. PivotTable Service can be used in a variety of development environments, including both Microsoft Visual Basic and Visual C++.



Figure

Client Architecture using PivotTable Service

Source: Microsoft Corp.

Calculation and Analytic Functions

Analysis Services provides substantial calculation and analytic functionality via the MDX language, which is used to retrieve and manipulate multidimensional information. Analysis Services can use three types of MDX functions: Basic/Advanced, Registered, and User-Defined (See "Features/Functions--Microsoft SQL Server 2000 Analysis Services" table). It is possible to add libraries of user-defined functions and to create calculated measures and calculated dimension members by combining user-defined functions, mathematical formulas, and MDXs. Custom roll-up formulas and custom member formulas, MDXs which determine cube cell values associated with members, are supported. Calculated cells, which are MDX statements that determine values of *specified* cells in a cube, are similar to custom member formulas, which apply to *all* cells for a specific member. Many functions in the Microsoft Excel worksheet library and in the Microsoft Visual Basic for Applications Expression Services library can be included in MDX expressions.

Data-Mining Functionality

Two data-mining algorithms are included with Analysis Services: Microsoft Decision Trees and Microsoft Clustering. The Analysis Services algorithms can train data-mining models with data from any relational data source that supports OLE DB access, as well as from multidimensional cubes created with Analysis Services. Microsoft SQL Server 2000 Analysis Services supplies a number of functions that retrieve and manipulate statistical information from a data-mining model. Analysis Services can be used with third-

party tools such as mining model viewer components. OLE DB support has been enhanced with the OLE DB for Data Mining specification to allow for integration with third-party algorithms. Data mining is integrated into the PivotTable Service following the same pattern that was used in Analysis Services. In general, a data-mining model is treated like a cube by PivotTable Service: users can create local data-mining models, retrieve information from server cubes, and so on.

The user interface for browsing the trees and clusters that are output from the data-mining models is basic, allowing the developer to take a quick look at the output of the mining model development-build process, but is not intended for analysts and end users. Microsoft expects to see a variety of tools developed by third parties; some of its Data Warehouse partners are already working to add data-mining features into their toolsets.

Analysis Manager

A console application that provides a user interface for accessing Analysis servers and their metadata repositories, Analysis Manager is used to administer Analysis servers; create databases and specify data sources; build and process cubes; create and process data-mining models; specify storage options and optimize query performance; manage security; browse data sources, shared dimensions, security roles, and other objects; find links for third-party client applications; and support resources, help updates, and product news.

Programmability of Functionality

Analysis Services provides three component tools that can be used with custom applications: Decision Support Objects (DSO), Add-ins Interface and Objects, and PivotTable Service. Using the Analysis Services Add-ins Interface and Objects, applications can be created that interact with and enhance the Analysis Services user interface. The Analysis Services Decision Support Objects (DSO) object model can be used to extend the functionality of Analysis Manager or to automate the ongoing maintenance of the system; custom applications can use DSO to control and automate functionality on the server. The DSO library provides a robust set of Component Object Model (COM) objects and interfaces that can be used to create applications that can programmatically administer the environment, which includes managing Analysis Services objects, such as servers, databases, data sources, dimensions, cubes, mining models, and roles. It also includes administering security, processing cubes and data-mining models, and so on. Applications can also create and maintain OLAP objects such as cubes, dimensions, and roles. PivotTable Service, an OLE DB provider that supports the optional OLE DB for OLAP extensions, is the primary interface between client applications and Analysis Services, functioning as a connection interface with Analysis Services cache management functionality to support client application access to OLAP data.

Pricing

SQL Server 2000 is priced either per client license or per processor. For large deployments, the per-processor license is expected to be most cost-effective. The price includes a license for Analysis Services, RDBMS, Data Transformation Services, Meta Data Services, and English Query. In a production environment where Analysis Services is installed on its own processor (unless databases and cubes are tiny) and the RDBMS on another, it would be necessary to pay for a license for each processor. The prices listed below do not include volume or other discounts.

—

Microsoft Corp. SQL Server 2000 Analysis Services: Price List

Product	Price (US\$)
Microsoft SQL Server 2000 Developer Edition English North America CD	499
Microsoft SQL Server 2000 Enterprise Edition English North America CD 1 Processor License --25 Clients	19,999 11,099
Microsoft SQL Server 2000 Standard Edition English North America CD 1 Processor License --10 Clients --5 Clients	4,999 2,249 1,489

Updated pricing can be found at www.microsoft.com/sql/productinfo/pricing.htm.

GSA Pricing

Yes.

Competitors

The primary BI Platform competitors for Microsoft SQL Server Analysis Services are Hyperion Essbase and Oracle Corp. Express. While precise market share and position statistics are unavailable, Hyperion Essbase has the largest share of these three competitors, probably roughly equal to the other two combined. Although Microsoft just entered this market at the end of 1998, they are conceivably running neck-and-neck with Oracle by now. Other RDBMS vendors have begun to build OLAP functionality into their DBMSs and could prove to be a source of additional competition down the road. IBM has already added OLAP functionality through an agreement with Hyperion, under which they have implemented Essbase functionality in DB2. Other competitors include Seagate Software Crystal Info and Holos, Gentia Software Gentia, Applix TM1, MicroStrategy MicroStrategy 7, and The SAS Institute SAS System and SAS/MDDB Server.

On the surface, it would seem that the greatest differentiators for Microsoft Analysis Services versus Hyperion Essbase and Oracle Express are that these two products support Unix environments, and Analysis Services does not. Analysis Services provides a true hybrid OLAP (HOLAP) environment, whereas Essbase does not, and Oracle has a limited hybrid functionality. There is another major and more important difference: Microsoft Analysis Services was designed with a number of scalability- and performance-enhancing features, including cube data compression (unlike competitors' cubes, which explode the source data into much larger OLAP datasets, Analysis Services compresses the data into smaller OLAP cubes), partial aggregation with on-the-fly aggregation, and partitioning, to name a few of the most significant. While competitors have retrofitted their products with features like partial and on-the-fly aggregation and partitioning, this could not be done as efficiently as if these features were built into the base products. Analysis Services was engineered to use these features to optimal advantage and has achieved remarkable data compression and performance not attainable by its competitors (according to Microsoft). Oracle Express users can use the Oracle RDBMS (Oracle 8i) as a ROLAP data store; however, some Express features are disabled in the relational Express version. The DB2 OLAP Server supports all Essbase (MOLAP) functionality identically to native Essbase.

In a limited, informal survey of BI users conducted in mid-1999, there were an adequate number of responses from Hyperion Essbase and Oracle Express users to perform a head-to-head comparison of these two BI platforms. (The survey results were qualitative in nature, i.e., intended only to provide insight into user attitudes. The results may not be statistically valid due to the relatively small number of respondents expressing an opinion. Microsoft Analysis Services--then designated as SQL Server 7.0 OLAP Services and a new entrant into the market--was not a part of this survey.) Some of the findings of this survey are presented here, where it can be conjectured as to how these products might compare to Microsoft Analysis Services if a similar survey were done, including Analysis Services; other findings of the survey deemed not relevant have not been presented and discussed here.

In the survey, Hyperion Essbase had relatively low ratings for hybrid OLAP support; Oracle Express was only slightly stronger. Microsoft Analysis services probably would have outdone both of these

products (had it been included in the survey), because it has strong hybrid OLAP support. Two of Hyperion Essbase's greatest strengths were in robust analytic functionality--but only slightly stronger than Oracle Express--and performance. We would anticipate that Essbase would show up the same way versus Analysis Services, due to Essbase's extensive functionality as a very mature product. Although Hyperion Essbase also was relatively strong in drill-through to details, Oracle Express was judged by survey respondents as slightly stronger. With strong drill-through support in MOLAP, ROLAP and HOLAP environments, Analysis Services probably would have at least equaled these two products. According to the survey, one of Oracle Express' greatest strengths versus Hyperion Essbase was full support for Unix, although Essbase does support some Unix platforms. Based on Microsoft Analysis support being restricted to Windows NT, it probably would have lagged Express and Essbase in this criterion.

Seagate Software Crystal Info contains a subset of Seagate Holos, an OLAP server that supports both relational and multidimensional OLAP databases. Seagate Holos OLAP technology is sound, and Seagate Info is a good product offering, but Seagate has been unable to mount a successful challenge to the top products. This is evidenced by its strategy--to gain market recognition--of giving away Seagate Analysis, a stand-alone query, reporting, and OLAP analysis tool, along with 50 free seats of Seagate Info. Gentia Software, like other secondary OLAP vendors such as Applix (TM1), which have good multidimensional OLAP technology, somehow has not been capable of achieving market success and has migrated into being a BI applications vendor. MicroStrategy offers relational OLAP solutions and is not as direct a competitor as other multidimensional OLAP products. SAS comprehensive BI solutions require a large commitment by a customer to its overall technology, so SAS may not be a direct competitor for Analysis Services.

Strengths

Performance and Scalability

Microsoft Analysis Services was designed with a number of scalability- and performance-enhancing features, including cube data compression, partial aggregation with on-the-fly aggregation, and partitioning, to name a few of the most significant. Spreading a cube over multiple servers by dividing it into partitions, Analysis Services can then retrieve data in parallel to answer queries, as well as distribute the processing load. OLAP cubes can be incrementally updated, even while they are in use. Having been thus engineered from the outset, Analysis Services can use these features to optimal advantage and has achieved remarkable data compression and performance not attainable by their competitors (according to Microsoft). Analysis Services also offers a number of tools that can tune system performance, including the Storage Design Wizard and the Usage-Based Optimization Wizard.

PivotTable Service shares much of the server functionality. This enables it to bring its multidimensional calculation engine, caching features, and query management directly to the client computer so that calculations can often be performed on the client instead of the server, thus distributing the computational load between the server and the client and reducing network traffic. Intelligent cache management can minimize network traffic. User queries, metadata, and data are stored in the Analysis server cache, making it possible for PivotTable Service to answer new queries by calculating answers from cached data rather than retrieving data from the disk.

Ease of Use for Administrators

Analysis Services includes a number of wizards and editors that simplify using the system for DBAs and other IT professionals.

Programmability of Functionality

Analysis Services provides three component tools that can be used with custom applications: Decision Support Objects (DSO), Add-ins Interface and Objects, and PivotTable Service. Using the Analysis Services Add-ins Interface and Objects, applications can be created that interact with and enhance the Analysis Services user interface. The DSO object model can be used to extend the functionality of Analysis Manager or to automate the ongoing maintenance of the system. Custom applications can use DSO to

control and automate functionality on the server. These applications can manage Analysis Services objects, such as servers, databases, data sources, dimensions, cubes, mining models, and roles; administer security; process cubes and data-mining models; and so on. Applications can also create and maintain OLAP objects such as cubes, dimensions, and roles.

PivotTable Service is the primary interface between client applications and Analysis Services, functioning as a connection interface with Analysis Services cache management functionality to support client application access to OLAP data. Applications also can be created that interact with and enhance the Analysis Services user interface. Analysis Manager can call various routines in the application in response to user activity in the user interface.

Partitioned, Virtual, and Linked Cubes

Analysis Services supports cube partitioning, which can result in improved performance, particularly for parallel querying, as well as the flexibility of storing updates in a separate partition. By allowing different levels of aggregation and/or modes of OLAP storage in each partition, Analysis Services provides the ability to tune data management and performance to the application characteristics. Multiple partitions can be merged into a single physical cube. Multiple cubes can be joined into virtual cubes (no storage requirement) without necessitating the construction of a new cube and while still maintaining the optimal design for each cube. A cube can be stored on a single server and referenced as a linked cube on other servers, avoiding the more costly alternative of storing and maintaining copies on multiple servers.

Security Model

Analysis Services provides strong security for both administrators and end users. Administrator security is controlled using the Microsoft Windows NT 4.0 or Windows 2000 group named OLAP Administrators. End-user security specifies which end users can access data and the types of operations they can perform, including whether they have read or read/write access. End-user security can be enforced at several levels of detail. These levels are (from least detailed to most detailed) server (Analysis Server), database, cube/mining model, and dimension member, cell. End-user security is controlled using authentication during connection to the Analysis server, and database, cube, and mining model roles defined in Analysis Manager.

Flexible Data Storage Options--ROLAP, MOLAP, and HOLAP

Microsoft SQL Server 2000 Analysis Services is one of the few products of its type that offer all three types of OLAP storage--ROLAP, MOLAP, AND HOLAP--providing the ultimate flexibility of choosing the most appropriate storage for each application. Each type of data storage offers storage and/or performance advantages for particular types of applications. See the "*MOLAP, ROLAP, and HOLAP Data Stores*" sidebar and the "*Features/Functions--Microsoft SQL Server 2000 Analysis Services*" table.

Data Mining

Two data-mining algorithms are included with Analysis Services: Microsoft Decision Trees and Microsoft Clustering. These are suitable for "light" data mining. For "serious" data mining, users will probably need to resort to special data-mining software. The Analysis Services algorithms can train data-mining models with data from any relational data source that supports OLE DB access, as well as from multidimensional cubes created with Analysis Services. Microsoft SQL Server 2000 Analysis Services supplies a number of functions that retrieve and manipulate statistical information from a data-mining model. Analysis Services can be used with third-party tools such as mining model viewer components, but does not provide its own end-user interface. OLE DB support has been enhanced with the OLE DB for Data Mining specification. Data mining is integrated into PivotTable Service.

Limitations

Unix Support

Microsoft SQL Server Analysis Services and RDBMS do not run on Unix platforms.

Support for Other Databases

Microsoft provides OLE DB drivers for Oracle 7.3 and 8i, and IBM DB2. Otherwise, customers must connect to other databases via OLE DB for ODBC, unless they obtain a direct OLE DB driver from a third party. There is no support for accessing other OLAP Servers, like Hyperion Essbase or Oracle Express.

Data Mining

While it is a good feature and nice to have some data-mining capability included, Analysis Services data mining only provides two basic models, satisfactory for "light" data mining. Users will need to resort to other special-purpose software to do "serious" data mining. An end-user interface is not provided, so users will need to use third-party software.

Insight

Microsoft SQL Server 2000 Analysis Services is a robust BI Platform that runs on Microsoft Windows NT platforms only. Despite being closed in this sense, it is open in the sense that it can access other relational databases through OLE DB support and allows front-end tools to access it through OLE DB for OLAP. While Analysis Services is part of the SQL Server 2000 product, Analysis Services functions independently of the SQL Server RDBMS. The bundling is a marketing decision, but this may be fortuitous for users of the SQL Server DBMS if they wish to develop BI applications. In addition to providing analytic functionality that comes with the product, Analysis Services also has data-mining capabilities, albeit simple ones, coming with two data-mining models: a decision tree model and a clustering model. Analysis Services has been engineered to be a scalable and high-performance product. In addition to being designed for partial data aggregation and OLAP data compression, it has extensive support for parallel processing and load distribution through cube partitioning, parallel querying against partitioned cubes, as well as wizards to help determine how to trade off storage considerations for performance. Analysis Services supports MOLAP, ROLAP, and HOLAP and write-back, which is essential for budgeting and forecasting applications, and even has a realtime cube update option. It is an impressive product, with extensive functionality that can be further enhanced and automated by custom development. Despite Hyperion Essbase having the largest market share, there is no player in this market that is dominant. Analysis Services stacks up very well against Gartner's criteria for BI Platforms and is capable of standing up successfully against its major competition in the Windows NT market.

Business Intelligence Platforms

Gartner categorizes BI tools into two categories: Enterprise Business Intelligence Suites and Business Intelligence (BI) Platforms (other Gartner Research Notes describe the characteristics of each). A BI platform offers a complete set of tools for the creation, deployment, support, and maintenance of BI applications. These are data-rich applications with custom end-user interfaces organized around specific business problems with targeted analyses and models. Gartner has defined the characteristics of a BI Platform, and these include a modular, distributed architecture that supports relevant standards like XML and OLE DB for OLAP and provides total Web deployment. It must be open and extensible so that third parties are encouraged to and can easily add functionality. The vendor must provide a strong third-party

support program. This will assist the vendor in building a substantial user base, which is essential to the success of BI Platform vendors. Of course, the product must have comprehensive BI functionality.

MOLAP, ROLAP, and HOLAP Data Stores

OLAP provides a multidimensional presentation of data warehouse data, creating cubes that organize and summarize data for efficient analytical querying. In ROLAP (relational) and MOLAP (multidimensional) data stores, the detailed data and the cubes of aggregate data are stored in the native stores. In HOLAP (hybrid OLAP), the detailed data is stored in a relational store, while the aggregates are stored in a multidimensional store. This provides the capacity of relational stores for large detailed transaction databases without duplication in a multidimensional data store, along with the high performance of a multidimensional data store. MOLAP storage provides excellent performance and data compression. It provides the potential for the most rapid query response times, appropriate for cubes with frequent use and the necessity for rapid query response. ROLAP query response is generally slower than that available with MOLAP or HOLAP, but is a preferred solution for large datasets--typically detailed sales transactions or the like that are infrequently queried or less recent historical data. ROLAP storage enables enterprises to take advantage of their investment in relational technology and enterprise data management tools. HOLAP storage offers the benefits of MOLAP for aggregations without necessitating duplication of the underlying detail data. HOLAP is ideal when the predominant type of queries will be against aggregates of large databases and rapid query response is required and also where there is also a requirement to occasionally drill down into the source data. Cubes stored as HOLAP are smaller than the equivalent MOLAP cubes and provide faster queries than ROLAP cubes.

Entire contents © 2000 Datapro Information Services, a subsidiary of Gartner Group, Inc. All rights reserved. Reproduction of this publication in any form without prior written permission is forbidden. The information contained herein has been obtained from sources believed to be reliable. Datapro disclaims all warranties as to the accuracy, completeness or adequacy of such information. Datapro shall have no liability for errors, omissions or inadequacies in the information contained herein or for interpretations thereof. The reader assumes sole responsibility for the selection of these materials to achieve its intended results. The opinions expressed herein are subject to change without notice.