

REAL PRACTICES: PERFORMANCE SCALING MICROSOFT SQL SERVER 2008 ANALYSIS SERVICES AT MICROSOFT adCENTER

Abstract

This white paper explains how Microsoft® adCenter implemented a Microsoft SQL Server® 2008 Analysis Services Scalable Shared Database on EMC® Symmetrix VMAX™ storage. Leveraging TimeFinder® clones and Enterprise Flash Drives with the read-only feature of SQL Server 2008 Analysis Services allowed adCenter to dramatically scale out OLAP while maintaining SLAs and decreasing system outages.

December 2010



Copyright © 2010 EMC Corporation. All Rights Reserved.

EMC believes the information in this publication is accurate of its publication date. The information is subject to change without notice.

The information in this publication is provided “as is”. EMC Corporation makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com.

Part Number h8129

Table of Contents

| | |
|---|-----------|
| Executive summary | 4 |
| Product solution..... | 4 |
| Key results..... | 4 |
| Audience..... | 5 |
| Statement of intent..... | 5 |
| Introduction | 5 |
| Microsoft adCenter..... | 5 |
| Key components..... | 6 |
| Windows NLB | 6 |
| Microsoft SQL Server 2008 Analysis Services | 6 |
| EMC Symmetrix VMAX | 7 |
| EMC Symmetrix TimeFinder | 7 |
| Scalable Shared Database deployment | 7 |
| Legacy system | 7 |
| Infrastructure | 8 |
| Cube design | 10 |
| Security..... | 13 |
| Management | 13 |
| Performance | 19 |
| Server performance | 19 |
| Analysis Services data structure | 20 |
| Storage performance | 21 |
| Conclusion | 23 |
| References | 24 |

Executive summary

Great information technology departments leverage business intelligence to provide a competitive advantage. Some of the world's largest enterprise users run Microsoft® SQL Server® 2008 Analysis Services to provide critical data analytics. Top-end Analysis Services systems deliver massive processor power and memory capacity for the most complex, multi-terabyte data warehouses to handle business-critical workloads with record-setting performance, as well as high reliability, availability, and serviceability.

A single server running Analysis Services cannot support an unlimited number of concurrent users or queries. Although an optimized cube design and efficient multidimensional expressions (MDX) can help to maximize performance, concurrent queries depleting server resources eventually require a scale-out approach.

In a storage area network (SAN) environment, it is possible to leverage the read-only database feature of SQL Server 2008 Analysis Services to separate data processing from reporting functions. Implementing a read-only reporting volume enables multiple query servers to access the same database concurrently, which can help to save terabytes of storage space without complicated configurations. The shared database volume only has to be able to sustain the combined input/output (I/O) workloads that the parallel query servers might generate. Microsoft calls this technology Scalable Shared Database.

The EMC® Symmetrix VMAX™ array utilizes TimeFinder® clones to facilitate rapid data copy and performance isolation with minimal impact to processing and reporting servers. Implementing Enterprise Flash Drives (EFDs) within the Symmetrix® VMAX array has created a new ultra-high-performance storage tier that removes previous performance limitations imposed by rotating disk drives. The significantly lower latency of EFDs dramatically improves the performance of latency-sensitive applications. For real-world workloads, EFDs feature high sustained performance with very low response times.

Product solution

The solution illustrates how the EMC Symmetrix VMAX SAN array, TimeFinder clones, and Enterprise Flash Drives integrate with Microsoft SQL Server 2008 Analysis Services to provide a highly performant scale-out environment for enterprise business intelligence and analytics.

Key results

Microsoft adCenter engineering and the Microsoft SQL Customer Advisory team (SQL CAT) partnered with EMC to design the custom solution detailed in this case study. adCenter deployed a custom scale-out solution with Microsoft SQL Server 2008 Analysis Services deployed on EMC Symmetrix VMAX arrays. The systems utilize EFDs to dramatically increase performance.

Microsoft SQL CAT has created a program called Real Practices that publishes examples of SQL Server running in real-world environments. This Real Practices case

study describes how Microsoft adCenter takes advantage of groundbreaking performance and scalability enhancements and innovative design tools available from Microsoft and EMC.

Note: For security reasons, the sample names of forests, domains, servers, databases, and other internal resources used in this paper do not represent real resource names used within Microsoft and are for illustration purposes only.

Audience

This white paper assumes that readers are IT professionals and technical decision-makers, already familiar with the concepts of SANs, Windows Server®, Active Directory® directory service, and SQL Server. Specifically, knowledge about SQL Server 2008 Analysis Services, SQL Server Integration Services, and Business Intelligence Development Studio, and a high-level understanding of Analysis Services optimization techniques for cube processing and query performance is helpful. Detailed product information is available in the [SQL Server 2008 TechCenter](#) and available in the [Analysis Services 2008 Performance Guide](#).

Statement of intent

The purpose of this paper is to share Microsoft and EMC knowledge, experiences, and recommendations related to the architecture and design of SQL Server 2008 Analysis Services for high performance and scalability in a large-scale OLAP environment. The solution discussed here is intended to serve as a template to assist in creating a custom solution. Deploying a custom solution will require adaptation of the information discussed in this paper to fit specific enterprise requirements.

Introduction

Microsoft adCenter

Microsoft adCenter is the paid search advertising system that provides ads for the Microsoft Bing™ search engine. adCenter provides a complete range of solutions for online advertising campaigns on Bing and offers customers a wide range of intelligence and analytics tools to target, reach, and engage the right audience with the right ad at the right time. The adCenter system uses a bid-based rendering system to display pay per click (PPC) ads based on user searches. Because adCenter uses PPC, advertisers only pay when an ad is clicked. adCenter utilizes advanced technology to deliver relevant advertising to Bing users. Advertisers can produce highly targeted ads based on user location and demographic information as illustrated in Figure 1.

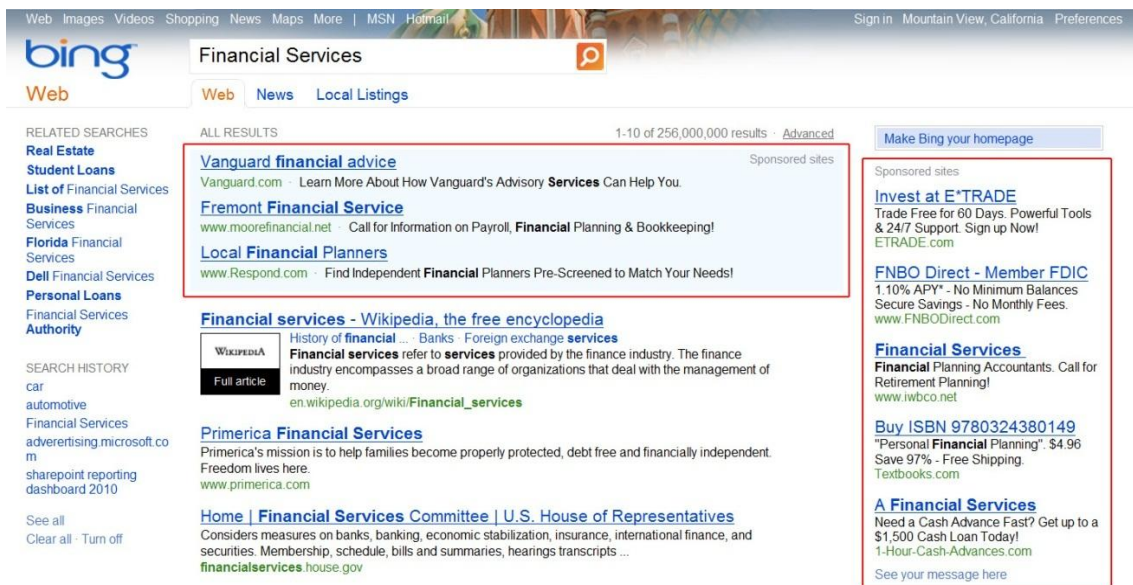


Figure 1. Paid search ad placement

With Microsoft adCenter on Bing organizations are able to:

- Easily import Google and Yahoo! campaigns to Bing using adCenter.
- Efficiently deliver business marketing.
- Control costs, paying only when someone clicks on an ad.
- Easily set up, monitor, and make changes to a search campaign with adCenter.
- Target geographically and demographically using business intelligence tools.

Search advertising with Bing puts ads in front of consumers who are actively seeking information on specific products, with unique profiling features to increase relevance and deliver the clicks that count.

Key components

Windows NLB

Network Load Balancing (NLB) for Windows Server 2008 allows the distribution of TCP/IP requests to multiple systems in order to optimize resource utilization, decrease computing time, and ensure system availability.

Microsoft SQL Server 2008 Analysis Services

SQL Server Analysis Services is the online analytical processing (OLAP) system from Microsoft. Analysis Services delivers rapid answers to advanced business intelligence questions. Integration with Microsoft Office and an open architecture allow a diverse user base to rapidly access deep analytical data. Analysis Services utilizes multidimensional OLAP (MOLAP) and an advanced caching engine to optimize data access.

EMC Symmetrix VMAX

EMC Symmetrix VMAX provides high-end storage for the virtual data center. Symmetrix VMAX scales up to 2 PB of usable protected capacity and consolidates more workloads with a much smaller footprint than alternative arrays.

EMC Symmetrix TimeFinder

TimeFinder is a family of products that enable volume-based replication within a single Symmetrix system. Data is copied from Symmetrix devices using array-based resources without using host CPU or I/O. The source Symmetrix devices remain online for regular I/O operations while the copies are created. The TimeFinder family has three separate and distinct software products, TimeFinder/Mirror, TimeFinder/Clone, and TimeFinder/Snap.

Scalable Shared Database deployment

Legacy system

adCenter relies on Microsoft technology, including Windows Server and SQL Server, to deliver exceptional advertiser value and allow advertisers to efficiently market products and services. The adCenter intelligence solutions are driven by a multifaceted OLAP environment based on SQL Server 2008 Analysis Services. This large-scale environment includes numerous data marts that consolidate information from various sources through SQL Server 2008 Integration Services. The adCenter data marts, based on SQL Server 2008 Enterprise Edition, maintain and stage up to 2 TB of data in relational databases, equivalent to one month of Web activities at a current accumulation rate of approximately 60 to 70 GB per day. The OLAP cubes built on top of these relational stores hold 13 months of data and are up to 3 TB in size.

The Microsoft adCenter engineering teams are experts who specialize in designing, implementing, and operating its OLAP infrastructure and solutions. The OLAP Development Team is responsible for designing all enterprise OLAP-based data marts and key analytics services within adCenter. The engineering team is responsible for implementing the business data models and analytics solutions. The operations team deploys and runs the data services, server hardware, and storage systems. For details on how adCenter works with SSAS, see the [Accelerating Microsoft adCenter with Microsoft SQL Server 2008 Analysis Services white paper](#).

The adCenter engineering and operations teams previously ran SQL Server 2008 Analysis Services on standalone servers. Each server would perform both processing and reporting operations. Soon after building the initial standalone systems, adCenter engineering became victims of previous successes. The Analysis Services OLAP cube systems were so popular with adCenter business data consumers that the cubes were soon overwhelmed.

Bilal Obeidat, lead developer, stated, “As a part of our cube scale/performance roadmap we have completed moving a majority of queries onto coarser grained cubes

which perform faster and meet a majority of user query requirements. As a side effect, the finer grained cubes also perform better by offloading many queries. As a result there is a need to find a solution where the emphasis is more on user scalability and less on database scalability.”

Objectives of this initiative include:

- Improve user scalability by maintaining current cube query response SLAs while supporting up to 4x the number of queries of the same pattern and mix.
- Increase cube uptime significantly.
- Eliminate negative query performance impact to end users during daily cube processing. Other than the scheduled cube farm downtime each day, there will be no impact to user query performance during cube load/processing.

adCenter engineering utilized Microsoft Excel[®], SQL Server Management Studio, and a customized UI built using Microsoft Silverlight[®] as client systems.

When asked why adCenter uses SQL Server Analysis Services, Bilal stated, “The customer has more flexibility than using canned reports. So cubes does support ad-hoc without the cost of building a mechanism to create canned reports. They can use Excel to slice and dice without waiting on creating and releasing these reports.”

Infrastructure

The adCenter Scalable Shared Database system utilizes one factory server and three reporting servers. Each server is an HP BL460 server with 32 GB of RAM installed, running eight processors. The servers are running the 64-bit edition of Windows Server 2008 Enterprise. SQL Server 2008 Analysis Services is installed on the factory and three reporting servers.

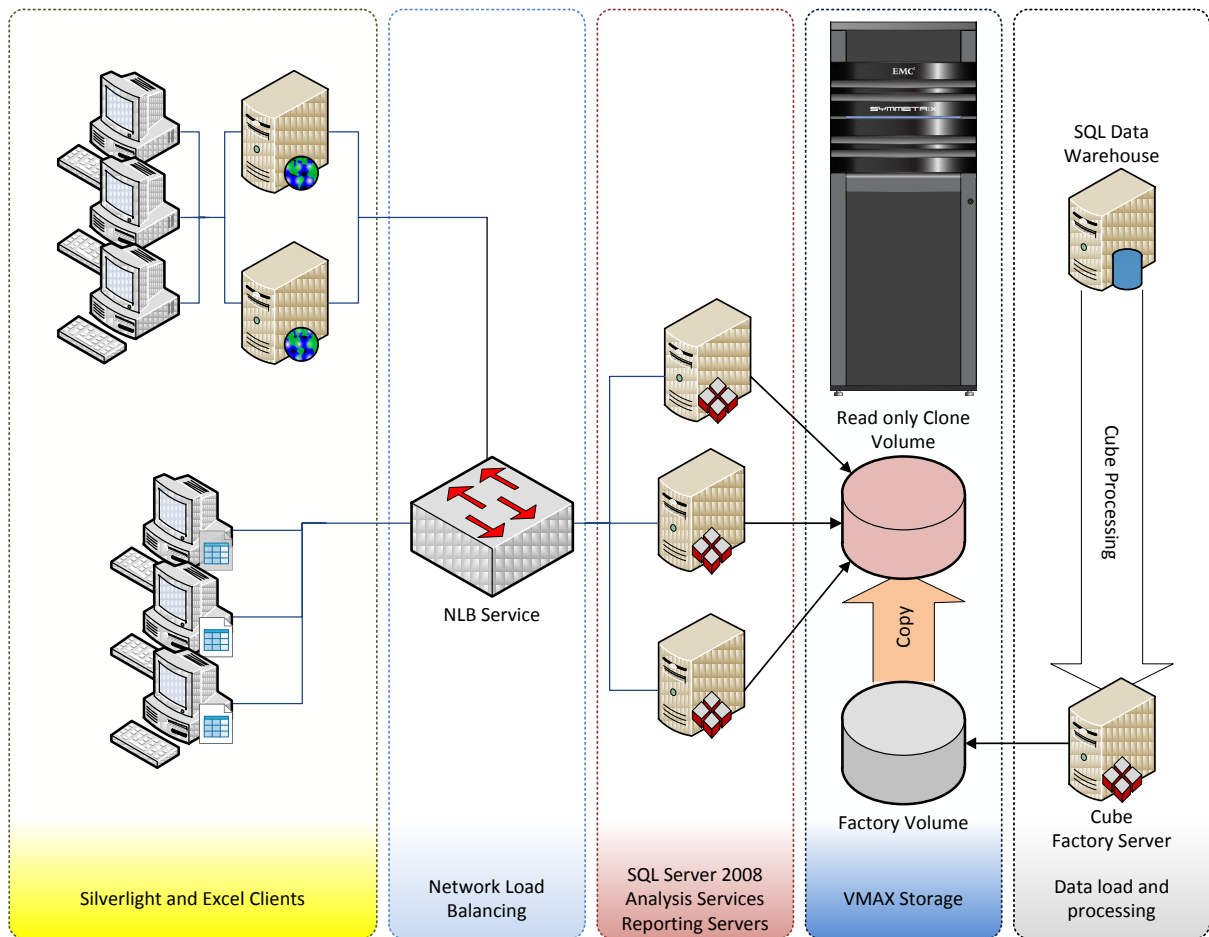


Figure 2. adCenter cube implementation

The reporting servers use [Microsoft NLB](#) for traffic balancing between servers. NLB is used because it allows easy network configuration, especially when multiple network interfaces are involved. NLB has some limitations; all network traffic is directed to all nodes, thus limiting overall bandwidth. One server running NLB has the same network bandwidth as three nodes using NLB. If more network bandwidth is needed, then a faster network technology will be needed; that is, 10 gigabit Ethernet. Fortunately the adCenter deployment was not hampered by network bandwidth.

Note: Hardware network load-balancing devices are designed to handle large-scale web traffic. One characteristic of web traffic is that it has a very short connection time. In order to protect the network from denial-of-service attacks hardware load balancers are configured with an extremely short connection time-out value. SQL Server and SQL Server Analysis Services will usually have a long connection time that requires the hardware load balancer to be configured with this longer connection time-out value. Unfortunately this can leave the device open to attack. When using a hardware load balancer for a Scalable Shared Database check with the load-balancer vendor for proper supported configuration instructions.

Each server is connected to an EMC Symmetrix VMAX array via dual host bus adapters. EMC PowerPath® software balances load across the multiple independent pathways that connect the host to storage, and ensures redundancy in the event of failure. The factory server utilizes a 3 TB factory volume hosted across 80 400 GB 10,000 rpm Fibre Channel disks in a mirrored and striped configuration (RAID 1 + 0). Every 24 hours the factory volume replicates changes to a 3 TB reporting volume made up of nine 400 GB EFDs configured in a distributed parity stripe configuration (RAID 5).

Cube design

adCenter engineering utilizes Analysis Services cubes to provide the data for business reporting. Multiple data marts feed the OLAP processing and reporting system. Analysis Services provides a key performance indicator (KPI) services framework that is leveraged to provide rich consolidated business data.

Data feeds based in SQL Server Integration Services (SSIS) are utilized to process and validate data on an hourly and daily basis. Every hour, the SSIS package loads new facts into a relational staging database. Separate SSIS packages load dimension data from other sources. Data is further processed by daily extract, transform, load (ETL) operations.

SQL Server databases are used to maintain and stage relational facts and dimensions. Each OLAP cube has its own set of facts and dimension databases to process against in order to minimize the need for cross-database or cross-server querying. Currently, adCenter maintains one month of activity data in a set of facts databases while maintaining the complete set of dimension data.

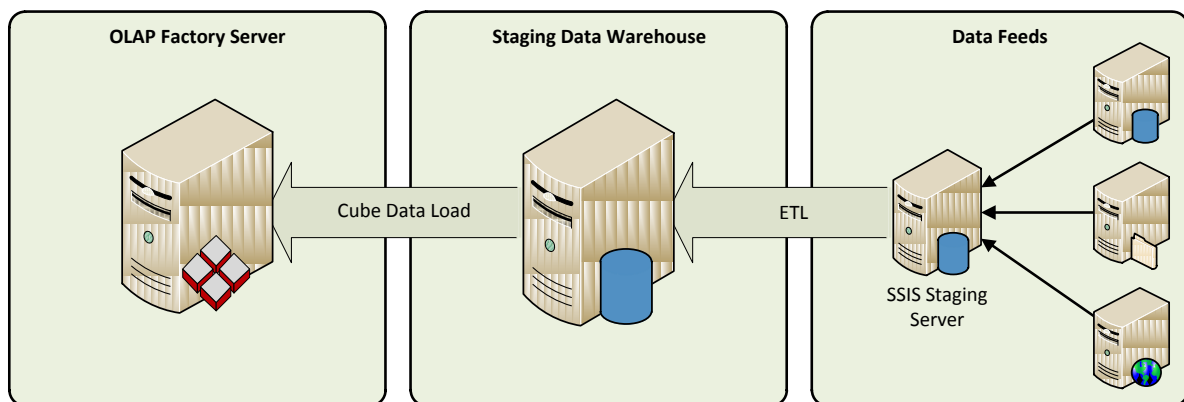


Figure 3. Daily data load

The OLAP servers maintain the Analysis Services databases and cubes. The OLAP reporting servers handle the user queries. The factory server processes cube data and performs maintenance tasks. At the end of the ETL processing cycle an XML for Analysis (XMLA) script is run on the factory server to add new dimension and facts data to the cubes. Every quarter a *ProcessUpdate* operation is used to refresh dimension data.

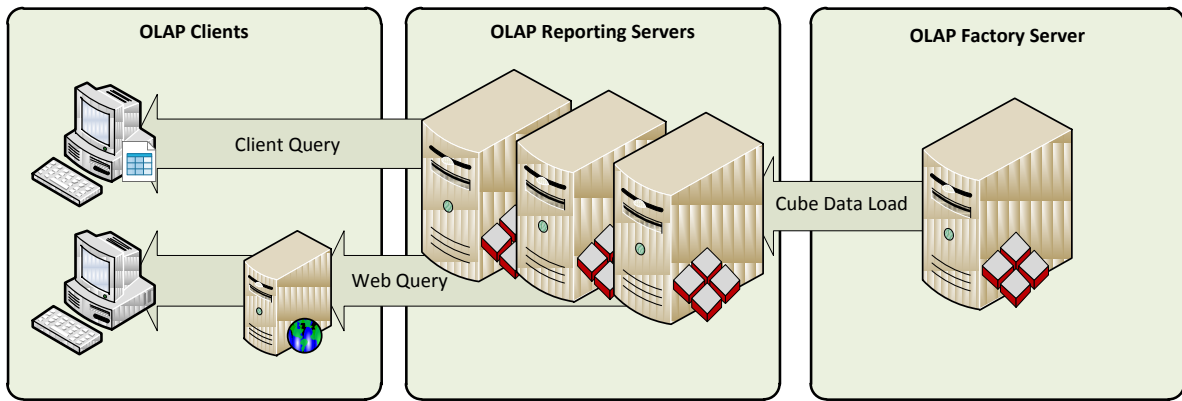


Figure 4. Data query

The adCenter hypercubes are designed using a star schema (Figure 5; table and column names have purposely been obfuscated). The single Fact_KPI table stores advertising usage metrics that are further described by the 11 dimensional tables. The fact table is partitioned by day, thus allowing the multicore servers to process the fact table in parallel, decreasing processing time.

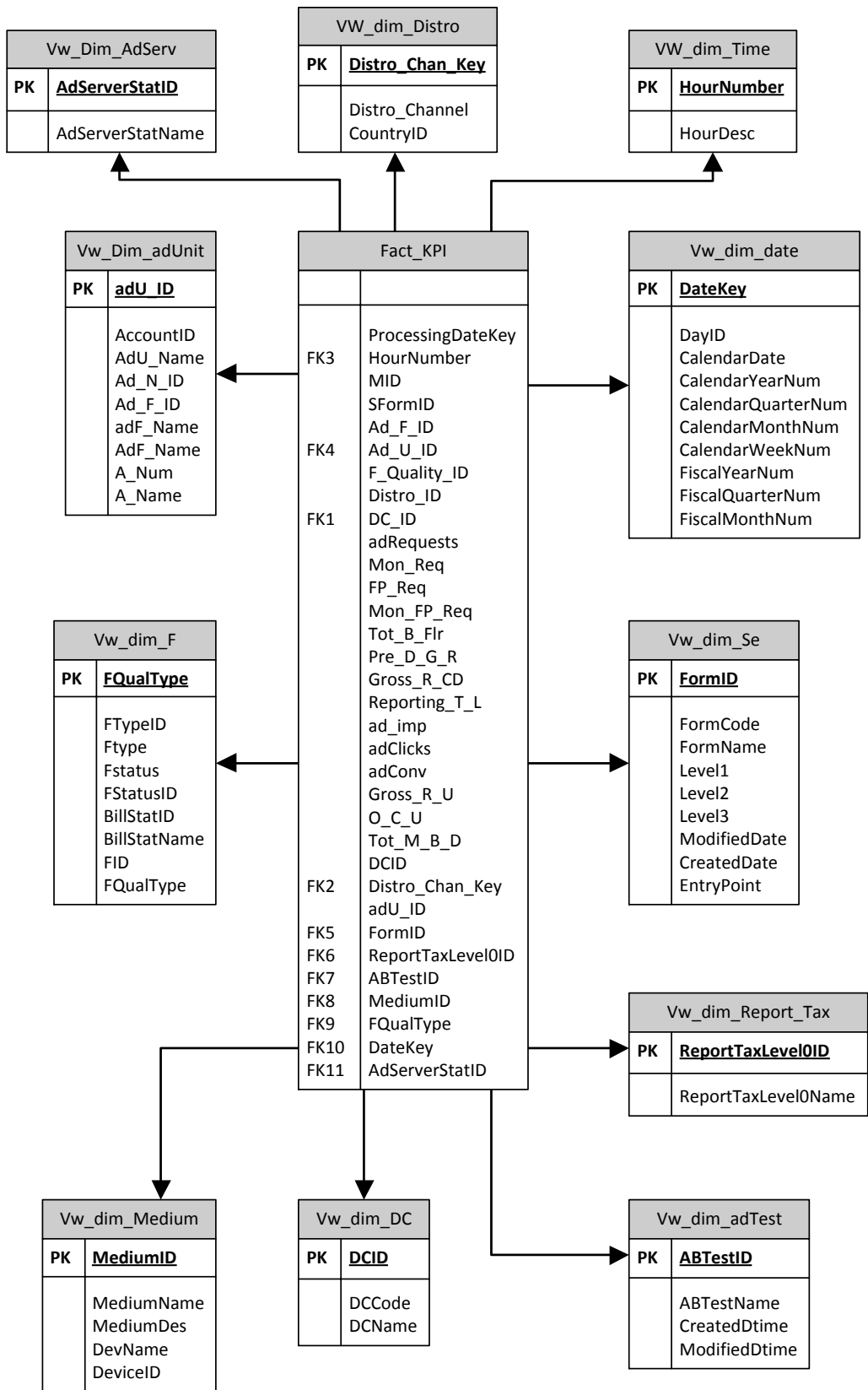


Figure 5. OLAP schema

Security

adCenter must comply with rigorous security requirements, including the Sarbanes-Oxley Act and Payment Card Industry Data Security Standard, and is required to secure all user Personally Identifiable Information (PII). The adCenter engineering and operations teams provide OLAP data to many clients that use Microsoft Office and Silverlight to access data. Analysis Services and SQL Server 2008 offer security feature enhancements that help provide effective management of security configuration, strong authentication and access control, and powerful encryption and key management capabilities.

SQL Server 2008 Analysis Services provides deployment and configuration options that allow an administrator to tailor each application to help meet company security requirements. SQL Server 2008 uses policy-based management to help manage and detect noncompliance with security policies for data across the enterprise.

Policy-based management helps ensure compliance with configuration policies for servers, databases, and database objects across the enterprise and helps reduce exposure to security threats by using the new Surface Area facet to control active services and features. Administrators using policy-based management can use SQL Server Management Studio to create policies to manage entities on the server, such as the instance of SQL Server, databases, or other SQL Server objects.

In addition to these security measures sensitive data can be encrypted. Encryption helps to protect sensitive data through built-in cryptographic capabilities and support for enterprise key management solutions. SQL Server 2008 has built-in cryptography hierarchy to create asymmetric keys, symmetric keys, and certificates.

Management

Running a Scalable Shared Database system in a production environment presents many challenges. It all starts with the Analysis Services factory node, which runs much like any other Analysis Services server except that it exclusively performs processing and maintenance tasks. The data refresh of the multidimensional cube on the factory server involves the data load tasks (loading data from the relational data sources) and Process Cube as described in Figure 6. The processing cube not only creates the cube fact data but also a set of aggregations that significantly improve query performance.

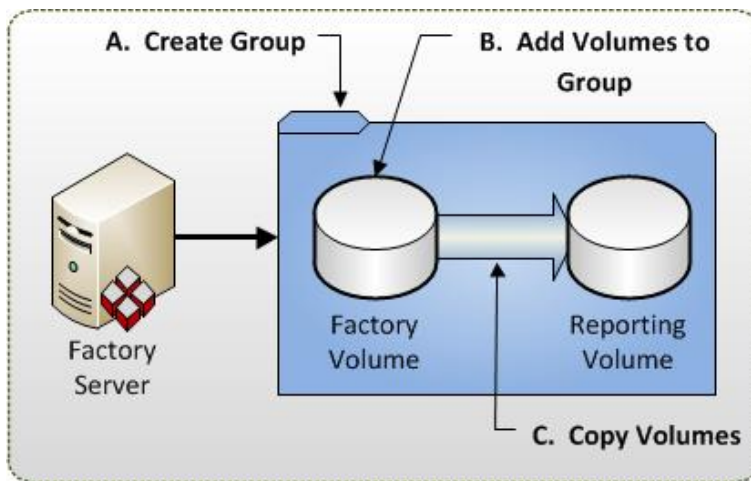


Figure 6. Clone group

The next steps are to detach the cube, fracture the clone, and sync the clone disks (as noted in Figure 7) to make the processed data available for the three query Analysis Services servers.

The underlying disks to hold and sync this data are TimeFinder clones (to facilitate rapid data copy and performance isolation), which are controlled using EMC Solutions Enabler for VMAX. Solutions Enabler is managed using standalone executable files. The clones are controlled by a device group that is created and managed by the factory server. The device group is set up using the SYMDG.EXE command (Step A in Figure 6).

```
symdmg create <group_name> -type regular
```

SYMLD.EXE is used to add the factory and clone volume devices to the control group (Step B in Figure 6).

```
symld -g <group_name> add dev <device> -sid <array>
```

The factory and clone volumes are initialized using the SYMCLONE.EXE command. The create argument is used only when the clone is first synced (Step C in Figure 6).

```
symclone -g <group_name> create DEV001 -precopy -differential  
sym ld DEV002
```

While the clone volume is synchronizing the reporting servers will not have access. The initial setup is a manual operation. The copy operation can be monitored using the query command.

```
symclone -g <group_name> query DEV001
```

Microsoft SQL Server 2008 Analysis Services directly interfaces with the Windows NTFS file system. Because a database exists both in server memory and on array storage, it is necessary to commit changes to disk prior to finishing a replication session. To ensure consistency the adCenter operations team detaches the cube from the processing server (for more information on cube detach, please refer to <http://msdn.microsoft.com/en-us/library/cc280661.aspx>) (Step D in Figure 7). Data

integrity is further ensured using a file system flush. SYMNTCTL.EXE allows Windows® hosts to manipulate storage with a single executable command (Step E in Figure 7).

```
symntctl flush -drive <drive letter>
```

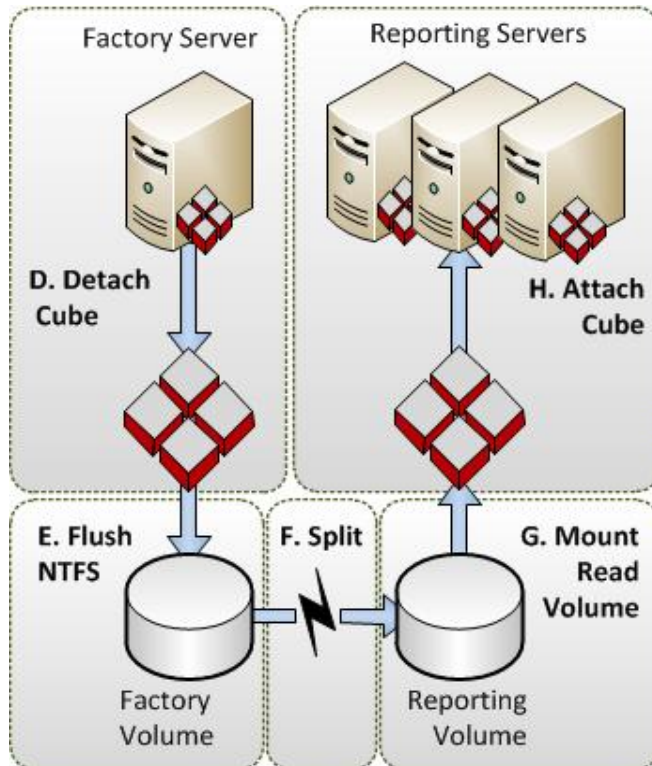


Figure 7. Clone setup

After the factory volume is in a consistent state, that is, all data has been flushed to disk, the SYMCLONE.EXE activate command is used to split the clone. The activate command will put the reporting volume into a read write state that is visible on the reporting servers (Step F in Figure 7).

```
symclone -g <group_name> activate DEV001 sym ld DEV002 -  
noprompt
```

Manually bring the volume online using Windows Disk Management. The same drive letter or mount point is used on all reporting servers and the factory server. SYMNTCTL commands the device using its array serial number and device ID. This is done because each time a server is rebooted the Windows disk number can change. Using the device ID ensures that only the defined volume is manipulated. This is especially useful when automating the cloning process (Step G in Figure 7).

```
symntctl.exe mount -sid <array> -symDev <device> -drive <drive letter>
```

Note: Windows Server 2008 has implemented a safety feature for SAN storage devices. When initially presented, or occasionally after a reboot, a SAN volume may appear as “offline.” In this state SYMNTCTL.EXE will not be able to manipulate the device. When automating the data copy operations this can prove problematic. Either use DISKPART.EXE to set the SAN POLICY to ONLINEALL. For example, use DISKPART> SAN POLICY=OnlineAll or programmatically check the volume’s state. Visit Microsoft TechNet for more information on the VDS SAN policy.

The reporting volume is now set to read only. Having a shared read only volume allows all reporting servers to cache the contents of the reporting volume without compromising the data integrity of the SQL Server shared nothing architecture.

```
symntctl.exe flag -sid <array> -symdev <device> -set READONLY
```

Each subsequent reporting server only needs to execute the mount command; as the volume is mounted the servers will note the read only flag.

```
symntctl.exe mount -sid <array> -symDev <device> -drive <drive letter>
```

Every night the processed cube is distributed to the reporting servers (Figure 8). Due to the size of the data set, standard network file transfers can become problematic. The VMAX array, utilizing TimeFinder, allows rapid data transfers. Ensuring the reporting volume has consistent data, the refresh cycle is started by detaching the fully processed factory cube.

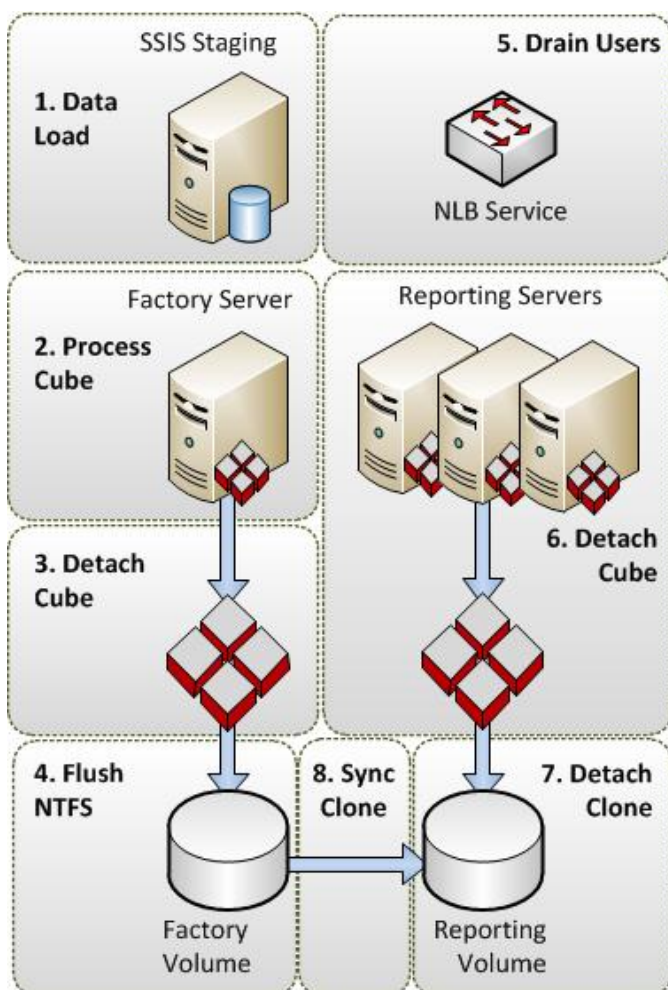


Figure 8. Daily data refresh

The nightly data refresh is very similar to the initial environmental setup. After the data has been loaded and processed the factory cube is detached. The factory volume is then flushed (Step 4 in Figure 8).

```
symntctl flush -drive <drive letter>
```

The reporting servers utilize the user draining features of NLB to terminate existing connections (Step 5 in Figure 8). Once connections have been terminated each server executes a cube detach operation (Step 6 in Figure 8). If the cube is not properly detached from the reporting servers then the mount and attach operation may fail (a reboot may be necessary to recover).

SYMNTCTL.EXE is used to detach the clone volume from each reporting server (Step 7 in Figure 8).

```
symntctl.exe unmount -sid <array> -symdev <device>
```

The clone volume is now synced with the factory master (Step 8 in Figure 8). This sync operation will replicate all content from the factory volume to the reporting volume. The VMAX array keeps a bitmap table of all changes, allowing an optimal

differential copy operation. A TimeFinder clone has many different copy modes. To ensure optimal performance of both the factory and reporting servers the pre-copy option is used. adCenter engineering set an SLA that guarantees maximum performance 23 hours a day, and allows one hour for the clone re-sync operation. Allowing the cloned reporting volume to copy all changes before being presented to the reporting servers ensures that the processing operations will not interfere with reporting.

When the recreate command is issued, the clone volume is marked as not ready. This will prevent the reporting servers from utilizing the device for the duration of the sync operation, and it ensures that the reporting servers are not erroneously reading old data.

```
symclone -g <group_name> recreate DEV001 sym 1d DEV002 -noprompt
```

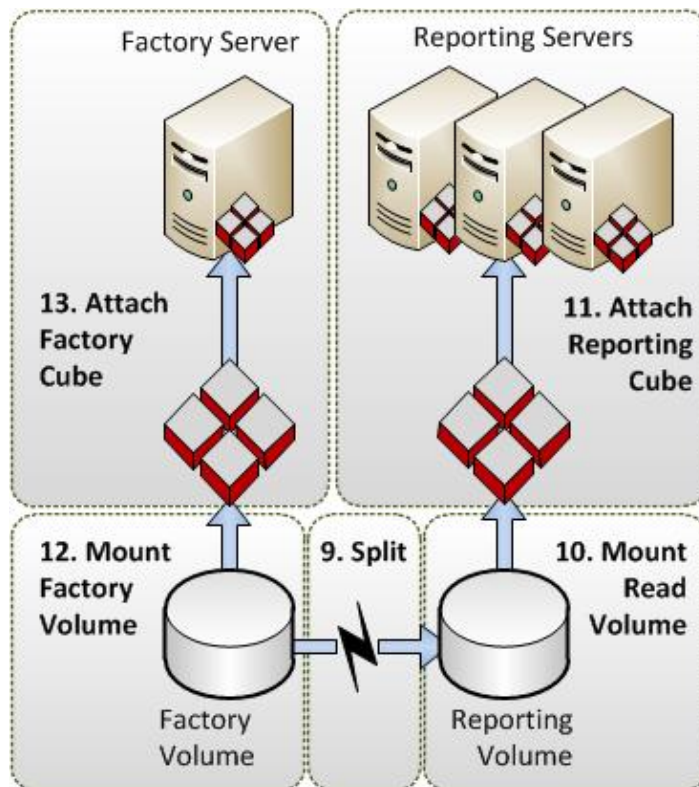


Figure 9. Daily clone mount

To facilitate automation, the SYMCLONE verify command can be used. The verify command will return a script-friendly message, rather than the full status returned by the query command.

```
symclone -g <group_name> verify DEV001 -copied
```

Once the sync operation is complete the device is fractured (this operation is also called a clone split) using the SYMCLONE activate command. This will split the devices and enable the reporting servers to view the reporting volume (Step 9 in Figure 9).

```
symclone -g <group_name> activate DEV001 sym 1d DEV002 -  
noprompt
```

Mounting the volume on the three reporting servers is exactly the same as the initial setup. On the first reporting server mount the volume and set it to read only (Step 10 in Figure 9).

```
symntctl.exe mount -sid <array> -symDev <device> -drive <drive  
letter>
```

```
symntctl.exe flag -sid <array> -symdev <device> -set READONLY
```

The remaining reporting servers will now mount the volume. The read only flag needs to only be set once (Step 11 in Figure 9).

```
symntctl.exe mount -sid <array> -symDev <device> -drive <drive  
letter>
```

After the reporting servers have successfully mounted the cube, NLB is set to again allow client connections. adCenter application development has automated this cloning operation. Depending on the amount of data generated by each daily cube processing, the sync time can vary, usually taking no more than a few minutes. As soon as the clone is fractured the factory server mounts the factory cube and can begin data load and processing (Steps 12 and 13 in Figure 9).

Performance

Server performance

SQL Server Analysis Services Scalable Shared Database has two distinct pieces: factory processing and reporting. Data processing is handled by a single server. It is critical for the factory server to have extremely robust processing capability, a fast network connection, and ability to process I/O operations rapidly.

Processing data aggregations puts a tremendous load on server processors and memory, and requires fast data transfers. When choosing a factory server, fast memory and I/O buses are the most important parts of the server. Having many processors in the server with a slow data transfer system will result in the processors waiting on data, thus slowing the processing cycle.

The reporting servers represent a scale-out solution to the problem of performance. Since servers can be added as necessary, the blocking factor is the performance of the shared storage. As long as the absolute performance of the storage array is not exceeded, the performance of the reporting servers will be linear. Two servers are twice as fast as a single server; three servers are three times as fast.

Analysis Services data structure

| | |
|-------------------------------|----------|
| 1.fact.data | 9,766 KB |
| 1.Customer.Customer Id.fact | 2,113 KB |
| 1.Customer.Postal Code.fact | 2,064 KB |
| 1.Promotion.Promotion Id.fact | 872 KB |
| 1.Promotion.Cost.fact | 810 KB |
| 1.Promotion.End Date.fact | 251 KB |
| 1.Time By Day.Time Id.fact | 191 KB |
| 1.Promotion.Start Date.fact | 170 KB |
| 1.Customer.City.fact | 52 KB |
| 1.Time By Day.The Day.fact | 48 KB |
| 1.Time By Day.The Month.fact | 7 KB |

Figure 10. File layout

Figure 10 provides an example of the many smaller sized files that make up the Analysis Services data partition folder. In each partition folder, the largest file(s) typically is the *.fact.data file, which contains the detail data defined by the measure group partition metadata that was transformed and compressed from the source data during Analysis Services processing. To improve query performance, Analysis Services also creates a set of *.agg files that contain aggregations of the underlying data and *.map files that contain the internal indexing structure. All of these files also have associated *.hdr or header files. Before querying the data or aggregations, the Analysis Services storage engine will first query the header files to determine what information exists.

There are many other files like the dimension *.store files. The purpose of calling out specific files is to observe the disk I/O pattern associated with the measure group partition folder when a query is executed. When the query is executed, the process will first query the header files to determine if the desired data exists for each of the affected partition groups (that is, in each of the file folders). If there is an aggregation that can answer the query, the result will be obtained from the associated aggregation file.

If the answer cannot be found within an aggregation, then the required data is obtained from *.fact.data and pushed to the Analysis Services formula engine (into memory). More simply put, for each partition folder required to answer the query, there is a process that will jump from header to aggregation to data (and combinations in between) to obtain the required result set and/or data. From a disk perspective, a single query that encompasses many partitions and/or many aggregations (common in enterprise environments) will result in random I/O as opposed to serial I/O interactions with disk.

Storage performance

When processing data, Analysis Services must read data from header, fact, and dimension files. This data can be distributed across many data files that are located across the storage system. As noted in [Analysis Services data structure](#), Analysis Services directly interfaces with the Windows NTFS file system. Generally speaking, NTFS generates an I/O request based on the size of the file. If the file is extremely large, then the I/O request will be broken up into smaller requests.

Accessing all of these files generates a lot of random I/O. Each query can spin up many separate processes, decreasing processing time and driving up random I/O. More complex queries can generate even more random I/O. Why is random I/O important? A hard disk is both an electronic and a mechanical device. The disk has two basic performance modes:

- Sequential access
- Random access

Sequential access offers the best performance. A 10,000 rpm Fibre Channel disk drive will perform roughly 1,400 sequential I/Os per second (IOPS). Depending on the I/O size the limiting factor is the maximum bus speed of the disk. When accessed randomly, this same 10,000 rpm disk can only perform between 100 and 120 IOPS. Each I/O access that requires head movement will increase the latency of the I/O request. This head movement time is measured by drive manufacturers by a metric called average seek time.

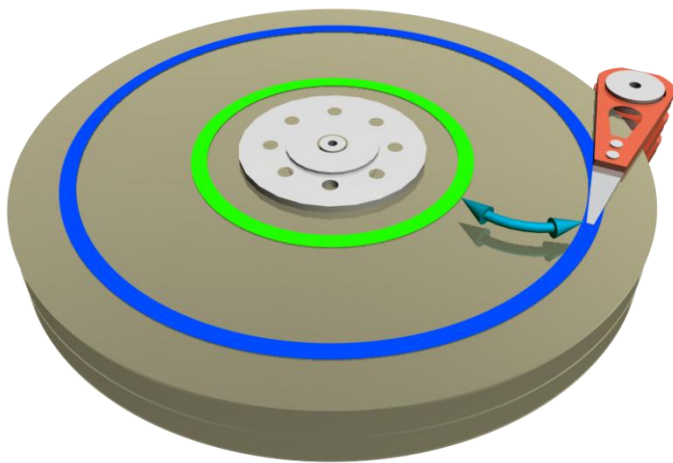


Figure 11. Head seeking

Excessive head seeking occurs (Figure 11) when many sequential I/O requests are sent to the same disk. The disk read/write head is forced to move to many regions of the disk. Drive manufacturers measure this poor performance with a metric called maximum seek time.

When Analysis Services performs more random I/O than the configuration allows, the administrator can either add more disks — if one disk produces 100 IOPS then two

will produce 200 IOPS — or move to EFDs. Adding many reporting servers to the mix, with each reporting server accessing the same physical disks, can negatively affect overall performance. One possible solution to this problem is to add enough physical disks to the storage volume to handle the maximum peak random workload. Another solution is to utilize EFDs. Remember that performance is not infinite; eventually any system will develop bottlenecks!

adCenter uses a hybrid approach. The factory server utilizes standard hard disk drives while the reporting servers are connected to shared EFDs. EFDs allow all of the reporting servers to simultaneously access the reporting data without negatively affecting performance.

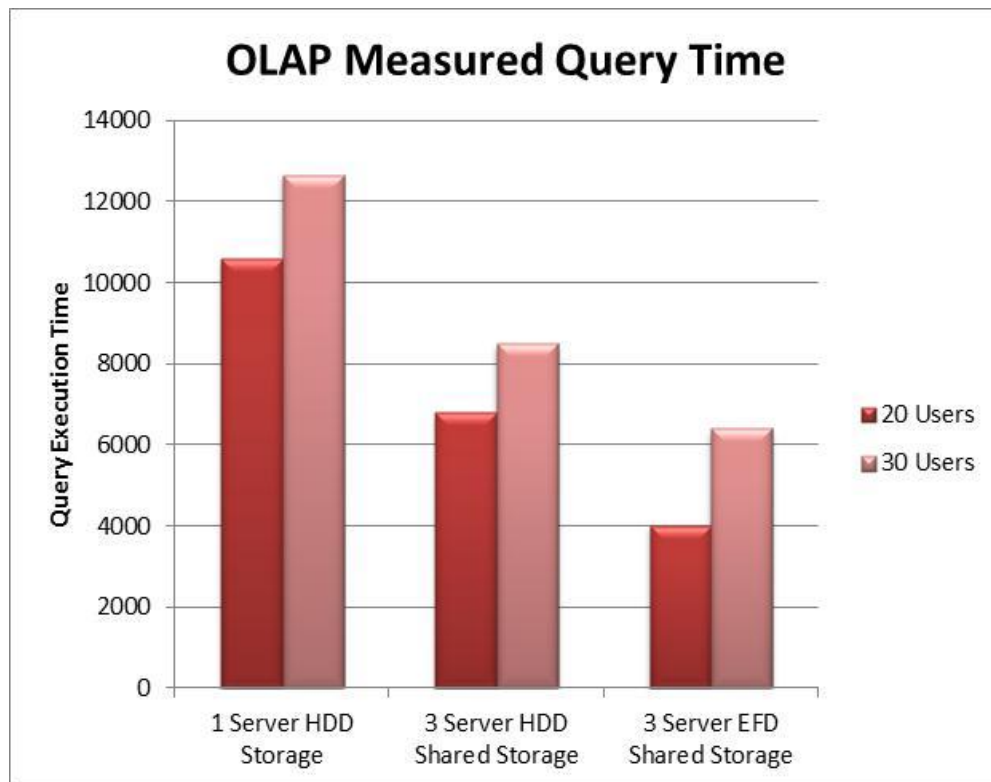


Figure 12. EFD performance

The EFDs have the added benefit of lowering the cost of ownership of the array. Assuming a single EFD can perform more than 4,000 IOPS, EFDs allow adCenter to replace 40 10,000 rpm disk drives with just one EFD. This savings is significant. Not only is adCenter saving power and space in the data center, but also the lifetime operational savings can allow the capital expenditure to be recovered in a short time.

The adCenter OLAP engineering team created a testing environment that measured the execution time of standard user queries. The contrast between EFDs and traditional drives is highlighted in the figure. The standard 20-user test queries executed 41 percent faster on EFDs than rotating disks. The 30-user test was 25 percent faster (owing to the saturation of the reporting server network).

The Scalable Shared Database allowed the reporting servers to drastically improve query load performance. Taking the baseline of a single reporting server, adding a second reporting server doubled the query workload and adding a third server tripled the number of queries executed in a given test.

Conclusion

Combining the normally compartmentalized technology disciplines of servers, storage, networking, and OLAP technology can seem daunting. This custom solution has succeeded precisely because of cross-functional, and indeed, cross-corporate cooperation. Taken together these parts have been combined to create a whole solution that is much greater than the sum of its collective parts.

The stated business requirements were:

- Improve user scalability by maintaining current cube query response SLAs while supporting up to 4x the number of queries of the same pattern and mix.
- Increase cube uptime significantly.
- Eliminate negative query performance impact to end users during daily cube processing. Other than the scheduled cube farm downtime each day, there will be no impact to user query performance during cube load/processing.

The existing excellent OLAP cube design was extended using NLB and read-only reporting cubes. These technologies allowed the number of users to scale logarithmically, while the cube query times were sustained. The Symmetrix VMAX array underpinned NLB and the read-only reporting servers. Adding EFDs to the solution allowed the query times to be further reduced.

Cube uptime was significantly improved by implementing rapid refresh with TimeFinder clones. Using array-based data replication allowed the VMAX array to offload copy overhead from OLAP servers.

Eliminating negative query performance was accomplished by separating the data processing and reporting functions. Now, data is loaded, processed, and aggregated without affecting reporting. The only time the reporting cubes are taken offline is during the nightly data refresh.

Finally, implementing EFDs in the Symmetrix VMAX array not only increased performance and decreased latency; it also allowed the adCenter team to reduce the total cost of ownership (TCO) of the entire solution.

References

- Microsoft SQL Server page on Microsoft.com:
<http://www.Microsoft.Com/SQLServer>
- “Deploying a Scalable Shared Database”:
<http://msdn.microsoft.com/en-us/library/ms345584.aspx>
- Analysis Services page on Microsoft.com:
<http://www.microsoft.com/sqlserver/2008/en/us/analysis-services.aspx>
- EMC Solutions for SQL Server and Business Intelligence page on EMC.com:
<http://www.emc.com/solutions/application-environment/microsoft/solutions-for-sql-server-business-intelligence.htm>
- *EMC Tiered Storage for Microsoft SQL Server 2008 Enabled by EMC Symmetrix VMAX with FAST* EMC white paper:
<http://www.emc.com/collateral/hardware/white-papers/h6708-tiered-storage-ms-sql-server-2008-symmetrix-vmax-fast-wp.pdf>
- Microsoft SQL Customer Advisory Team:
<http://www.SQLCAT.com>
- *Accelerating Microsoft adCenter with Microsoft SQL Server 2008 Analysis Services* Microsoft white paper:
<http://sqlcat.com/whitepapers/archive/2009/09/19/accelerating-microsoft-adcenter-with-microsoft-sql-server-2008-analysis-services.aspx>
- *Analysis Services 2008 Performance Guide* :
<http://sqlcat.com/whitepapers/archive/2009/02/15/the-analysis-services-2008-performance-guide.aspx>
- *Analysis Services Distinct Count Optimization Using Solid State Devices* case study:
<http://sqlcat.com/technicalnotes/archive/2010/09/20/analysis-services-distinct-count-optimization-using-solid-state-devices.aspx>
- *Analysis Services Distinct Count Optimization* white paper:
<http://sqlcat.com/whitepapers/archive/2008/04/17/analysis-services-distinct-count-optimization.aspx>
- *Professional Microsoft SQL Server Analysis Services 2008 with MDX*:
<http://www.wrox.com/WileyCDA/WroxTitle/Professional-Microsoft-SQL-Server-Analysis-Services-2008-with-MDX.productCd-0470247983.html>
- *Scale-Out Querying for Analysis Services with Read-Only Databases* :

<http://sqlcat.com/whitepapers/archive/2010/06/08/scale-out-querying-for-analysis-services-with-read-only-databases.aspx>

- Microsoft Windows valid disk SAN policy:
[http://msdn.microsoft.com/en-us/library/bb525577\(VS.85\).aspx](http://msdn.microsoft.com/en-us/library/bb525577(VS.85).aspx)
- Denny Lee's Blog:
<http://dennygee.com/>
- Michael Anderson's Blog:
<http://sqlvelocity.typepad.com/>

Authors

Michael Anderson, Denny Lee, Dennis Dewitt

Contributors

Bilal Obeidat, John Montgomery

Technical Reviewers

John Sirmon, Chuck Heinzelman, Steve Howard, Txomin Barturen, Ken Taylor, Mike Mcghee, Justin Kennedy, Thomas Kejser, and Nicholas Dritsas