Technical Report

Microsoft Business Intelligence on NetApp Storage

John S Parker, NetApp
March 2010 | TR-3829

MICROSOFT BUSINESS INTELLIGENCE ON NETAPP STORAGE

This technical report will discuss the basics of deploying a business intelligence solution with a Microsoft® business intelligence stack using SQL Server® as the base and Analysis Services as the key aggregate and data analysis tool. This report discusses the use of Snapshot™ copies and FlexClone® volumes in providing scalable and efficient storage solutions.

1   INTRODUCTION

Today's economic climate is showing the need for businesses to start looking at all aspects of their business. A key tool that is being used is a renewed emphasis on developing better means for understanding the business and developing trends for the future. The way that business is doing that is through better business intelligence (BI) operations.

Microsoft has a suite of tools that have been developed expressly for the purpose of developing a robust business intelligence environment. NetApp has a strong solution to help business get the most out of its BI investment. NetApp offers a solution for backup and recovery of the Microsoft SQL Server data warehouse and Microsoft Analysis Services databases. NetApp® storage solutions create a robust development and test infrastructure to reduce storage space for both Microsoft SQL Server and Analysis Services. NetApp provides the ability to scale the Analysis Services environment to meet the growing needs of reporting while keeping storage growth to a minimum and increasing the efficiency of the current storage infrastructure.

1.1   PURPOSE AND SCOPE

The purpose of this technical report is to give guidance on how to get the most out of your Microsoft business intelligence environment. This technical report will discuss in general terms the integration points for NetApp in a Microsoft BI solution; the ability for Microsoft to take advantage of NetApp backup and recovery; and use of FlexClone in the environment to create an easy-to-manage, highly efficient, and sustainable environment. This technical report does not discuss performance, which will be discussed in subsequent technical reports.

Note: This architecture is designed to show how NetApp integrates with the Microsoft Business Intelligence services; this is not the only configuration possible. The combination of Microsoft products and NetApp functionality can be applied to many different combinations of hardware and software tools.

2   SYSTEM ARCHITECTURE

The key components of this business intelligence architecture contain a data warehouse, an online transaction processing (OLTP) database, an Analysis Service cube, and an extract transform and load (ETL) process (Figure 1). In addition to the traditional areas of the ETL process this architecture also discusses the use of a FlexClone for population of the development and test environment as well as the use of a FlexClone to create multiple share Analysis Services databases and cubes.

Stepping through the process that was used in the technical report starts out with a production OLTP database. The scenario for the production OLTP system requires that the ETL process be pushed off to another system to ensure the OLTP processing SLA is met. The use of a FlexClone allow a point in time capture of the database that can be mounted on a separate system for processing of the database loading it into the operational data store (ODS) and into the data warehouse.

The creation of a fully populated data warehouse and ODS development and test environment is managed through using a FlexClone of the production data warehouse and ODS. This clone allows for the entire dataset to be available without the storage cost.
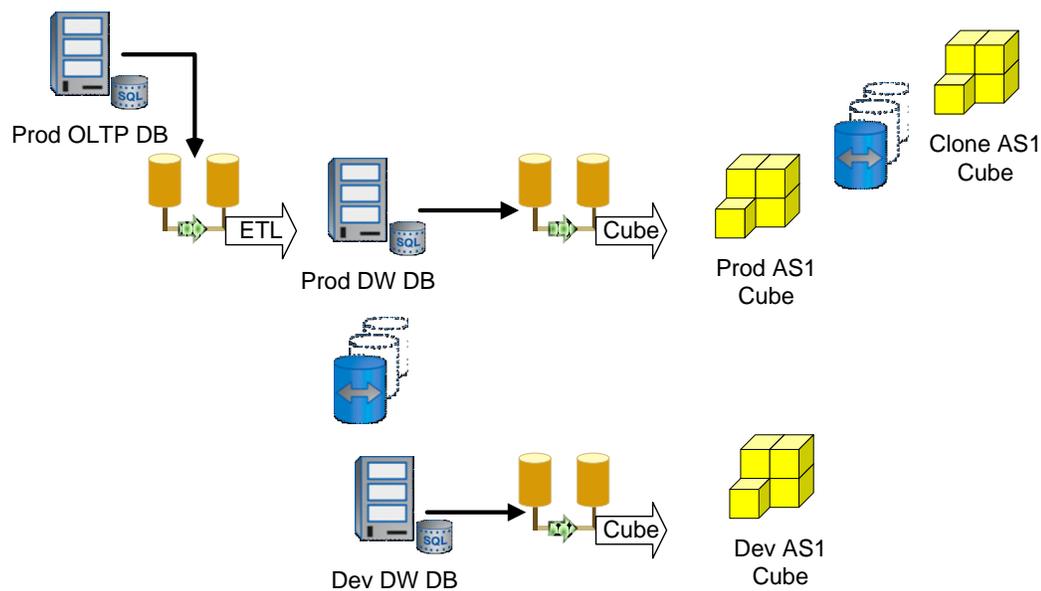
Figure 1) Business intelligence overview.

Taking a deeper look at the individual components of the entire system, the individual server break for this test includes four physical hosts ranging from single-CPU dual-core systems with 16GB of RAM to two-CPU dual-core systems with 32GB of RAM. See Table 1 for an additional description of the physical hardware.

Table 1) Physical host descriptions.

Host Name	Role	CPU Cores	Host Memory
SEA-SQL1	OLTP	4	16GB
SEA-SQL2	Analysis Services	4	16GB
SEA-SQL3	Data Warehouse	4	32GB
SEA-ETL1	ETL	2	16GB

2.1   DATA WAREHOUSE LAYOUT AND CONFIGURATION

The SQL Server instance is a Microsoft SQL Server 2008 instance with the memory capped at 30GB and four processor cores. The data warehouse configuration is two databases: DW1 and DW2. These two databases have different schema and processes to load. This is established this way to show different methods for loading a data warehouse and the different types of reporting and decision making that takes place with the two different databases.

DW1 is a 1.7TB database (Figure 2) that is based on the OLTP system. There have been no transformations of the data from the OLTP system. The schemas between DW1 and the OLTP database are identical. This represents a large normalized operational data store designed for historical reporting. This database has 13 data files and one log file broken into four LUNs, separated into four volumes, two aggregates, across a cluster pair of controllers.
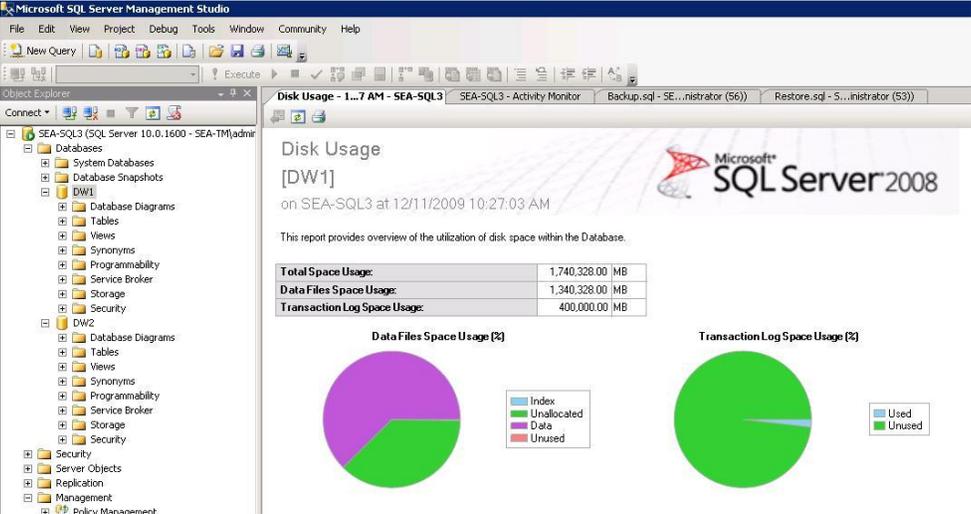
Figure 2) View of DW1.
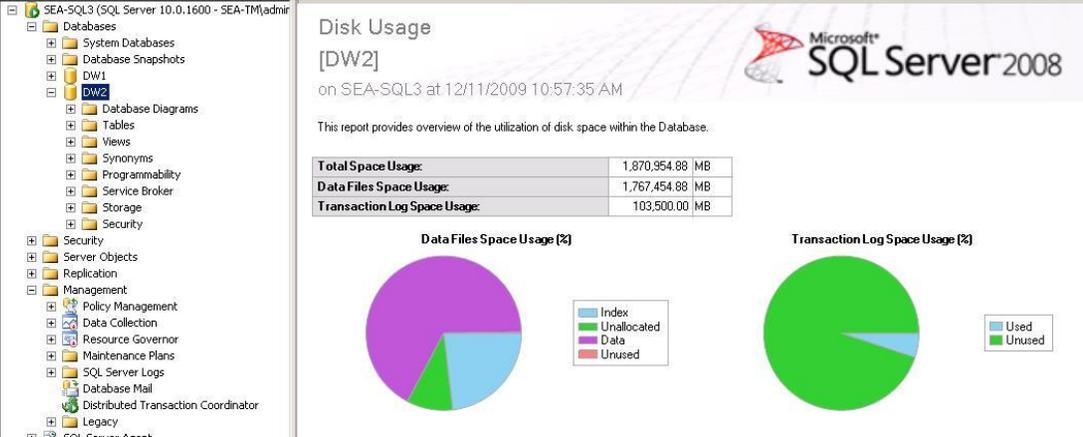
Figure 3) View of DW2.

The database DW2 is 1.8TB in size (Figure 3). The data has been transformed into aggregate-level data with both star and snowflake tables for large query performance. This database is populated by an ETL process that transforms the data from the OLTP system and places the data in a dimensional data warehouse schema. The schema is layer out so that long-term reporting trend analysis and data mining are easier to develop new reports to meet the needs of the data consumers. The database is broken into 19 data files and four log files spread across four LUNs, four volumes, and two aggregates across two

clustered controllers (Table 2). This allows for throughput and redundancy to make sure that the data is available.

Table 2) LUN layout.

Hostname	Controller Name	Aggregate	Volume	LUN	Size in GB	Total Size/GB
SEA-SQL1	C53050-4	SQLAGGR6	SQLTempdb	SQL1Tempdb	50	
	C53050-4	SQLAGGR6	SQLSysdb	SQL1Sysdb	20	
	C53050-1	SQLAGGR3	SQLData5	SQL1Data1	400	
	C53050-1	SQLAGGR3	Sys_Info1	SQL1SysInfo1	100	570
SEA-SQL2	C53050-2	SQLAGGR4	SQLData6	SQL2Data1	700	
	C53050-2	SQLAGGR4	Sys_Info2	SQL2Sys_info	100	800
SEA-SQL3	C53050-4	SQLAGGR6	SQLTempdb	SQL3Tempdb	100	
	C53050-4	SQLAGGR6	SQLSysdb	SQL3Sysdb	20	
	C63050-1	SQLAGGR1	SQLDATA1	SQL3Data1	1,000	
	C63050-1	SQLAGGR1	SQLDATA1	SQL3Data2	1,000	
	C63050-1	SQLAGGR1	SQLDATA2	SQL3Data3	1,000	
	C63050-1	SQLAGGR1	SQLDATA2	SQL3Data4	1,000	
	C63050-2	SQLAGGR2	SQLDATA3	SQL3Data5	1,000	
	C63050-2	SQLAGGR2	SQLDATA3	SQL3Data6	1,000	
	C63050-2	SQLAGGR2	SQLDATA4	SQL3Data7	1,000	
	C63050-2	SQLAGGR2	SQLDATA4	SQL3Data8	1,000	
	C53050-3	SQLAGGR5	Sys_Info3	SQL3Snap_Info	1,000	9,120
SEA-ETL1	C53050-3	SQLAGGR5	SQLDATA7	ETL1Data	400	
	C53050-4	SQLAGGR6	SQLSysdb	ETL1Sysdb	20	
	C53050-4	SQLAGGR6	SQLTempdb	ETL1Tempdb	50	470

2.2 ANALYSIS SERVICES CONFIGURATION

The Analysis Services Instance has two Analysis Services databases that correspond to the operational data store and the data warehouse databases. The AS databases each have data cubes that have been built to gather data and build aggregate data for fast reporting and analysis. Each cube is placed on a NetApp LUN to take advantage of the NetApp Snapshot technology for backup and cloning.

The Analysis Services instance in this setup is using a single LUN for each AS database (Figure 4). The view in Figure 4 shows the LUN established through SnapDrive® and how it is mounted using mount points to manage the space and directories.
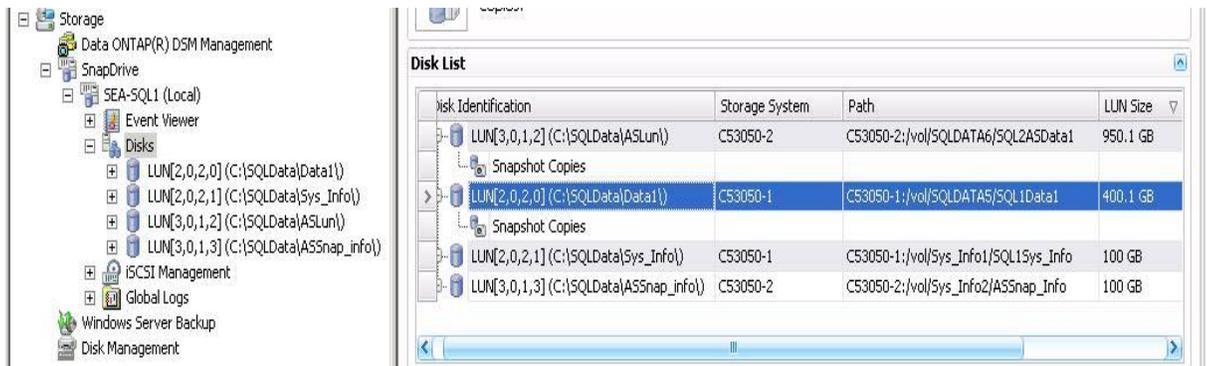
Figure 4) SnapDrive view of LUNs.

The cube is currently estimated at approximately 430GB. The database properties (Figure 5) show the estimated size and location of the data.

3 NETAPP BUSINESS INTELLIGENCE

The data warehouse is a key component of any business intelligence activity. The management of the data warehouse becomes critical as the data volume grows within the data warehouse. Larger databases in them create problems that smaller OLTP databases do not have. Making sure of recovery time objectives (RTOs), query performance on large data sets, managing indexes to meet the query needs of reports, keeping ETL jobs working to make sure that the data maintains relevancy, and increasing backup times as the volume of data grows are just a few of the problems that come with managing a data warehouse. This technical report will demonstrate how NetApp helps to give administrators and stewards the tools they need to meet the SLAs for their environment.

This section will cover the entire solution, discussing key tools to helping make sure that the data warehouse can be developed and managed on a NetApp storage platform, extract transform and load using FlexClone, Analysis Services guide to backup and recovery, cloning AS cubes, and integrating with Analysis Services scalable shared databases (SSDBs).

3.1 DATA WAREHOUSE MANAGEMENT

NetApp brings many different tools to help manage a data warehouse. Looking at each of the items described above will show solutions to each of these problems with managing a large data warehouse on Microsoft SQL Server.
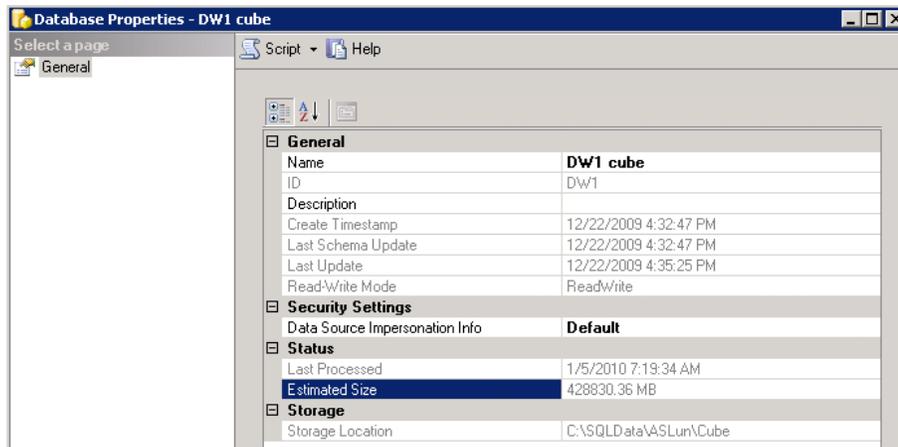
Figure 5) Analysis Services cube properties.

DEVELOPMENT AND TESTING OF INDICIES

One challenge of large SQL Server databases is developing and managing indexes. The creation of an index on a large table can at time be very intrusive in the database. This affect of creating large indexes requires more thought and development put into the index before deployment to production. Large queries

require effective indexes in order to perform correctly and return data in a reasonable time frame. NetApp's solution to helping make sure that you have quality indexes is through an improved development and test process. The key to this is having an appropriate data set in the development or test environments that matches production.

The traditional method of doing this is with a complete copy of the database. The challenge there is now you suddenly have to spend more for storage. NetApp FlexClone volumes allow you to have full access to the data in a read/write Snapshot copy, which takes up small amount of space for the block pointers and any additional changes to the database. This database can be on hosted on a separate SQL Server instance, thus moving the processing over to another location. This method reduces the effect on the production environment and allows database developers access to a full set of data. The development process does not have to deviate but is enhanced through the using a full set of data. This process allows database developers and operational database administrators to better time how long the new maintenance for the indexes will take as well as help to find issues with the data before being moved to production.
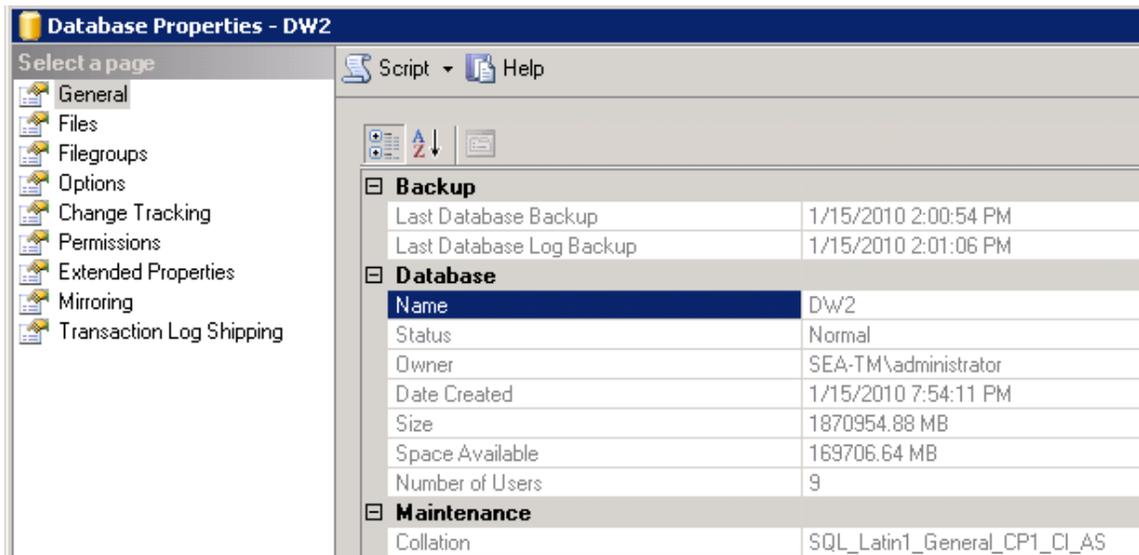
Figure 6) DW2 data warehouse DB properties.

The process for creating clones can be done through SnapManager® for SQL Server (SMSQL) or through Power Shell scripts that create the clone and attach it to a specific SQL Server instance. The clone can reside on either the host for the parent instance or on a separate SQL Server instance. For more information about creating clones with SMSQL, see the best practice guide for SMSQL 5.0 at http://media.netapp.com/documents/tr-3768.pdf. Comparing the size of the database (Figure 6) to the clone database (Figure 7), the clone database matches the original database in size and content. The actual space that the clone takes up on the storage platform is only the size of the original Snapshot copy plus any changes. This makes sure of storage efficiency that both are actual used space and development data sets that match the production environment.
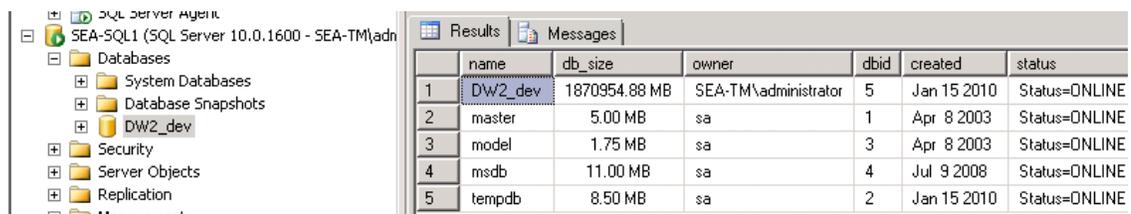
Figure 7) Development database FlexClone volume.

SPECIALIZED REPORTING CLONES

Specialized reporting can create unique changes in the larger data warehouse environments. The queries required to support specialized reports at times require specialized table statistics or tailored query plans to be effective. In a traditional data warehouse environment specialized queries are often compromised due to the requirements of higher priority queries. The compromise can lead to a combination of duplicate tables and even databases to satisfy the needs of certain reports. Year end and quarterly reports are prime examples of specialized reporting that can require different statistics for query performance.

The NetApp solution to this issue is to create FlexClone volumes that allow the customization of a database to handle these specific requirements (Figure 8). The ability to make a read/write clone of the data warehouse allows the administrator to create specific queries for specialized reports; these specialized queries can have different table statistics to allow each report to run in the required time.

Figure 8) Database and FlexClone volumes DB sizes.

The key to a successful business intelligence environment is the ability to get consistent and fast reports from the reporting environment. The reporting environment for this technical report consists of creating FlexClone volumes of the data warehouse to enable specific statistics for specialized queries. These specialized databases allow for deviation of statistics for specific tables and enable the use of the same code base for stored procedures to execute different query plans to successfully build reports.

3.2   EXTRACT TRANSFORM AND LOAD PROCESSING

The processing of the Analysis Services cube and the loading of the data warehouse are broken into stages that break up the process into manageable components. The cube is processed by a job that runs on a schedule managed from the Analysis Services host. In this sample scenario the cube is processed once a day. After the cube has been processed for the day, then it is cloned and the clones are attached to separate Analysis Services instances. Cloning of the cube allows for query scaling.

The data warehouse is loaded on a daily basis using an Integration Service job that gathers data from clones of the OLTP database. By using clones of the database processing resources can be moved off to separate hardware without impacting the production environment. This job takes place before the Analysis Services cube is processed.

Overview of steps (Figure 9):

1. Create a FlexClone volume of the OLTP database. This can be done through the System Manager user interface or through the controller command line interface.

2. Attach the OLTP clone to the processing SQL Server instance.

3. Run the ETL Integration Services job extracting new data from the OLTP clone and populating the data warehouse.

4. Process the Analysis Services cubes through a job on the AS host 1.

5. Detach the Analysis Services cube from AS host 1.

6. Create clone Snapshot copy of the Analysis Services directory.

7. Reattach the Analysis Services cube on AS host 1.

8. Attach clone to AS host 2-5.

9. Back up the data warehouse.

Extract Transform and Load
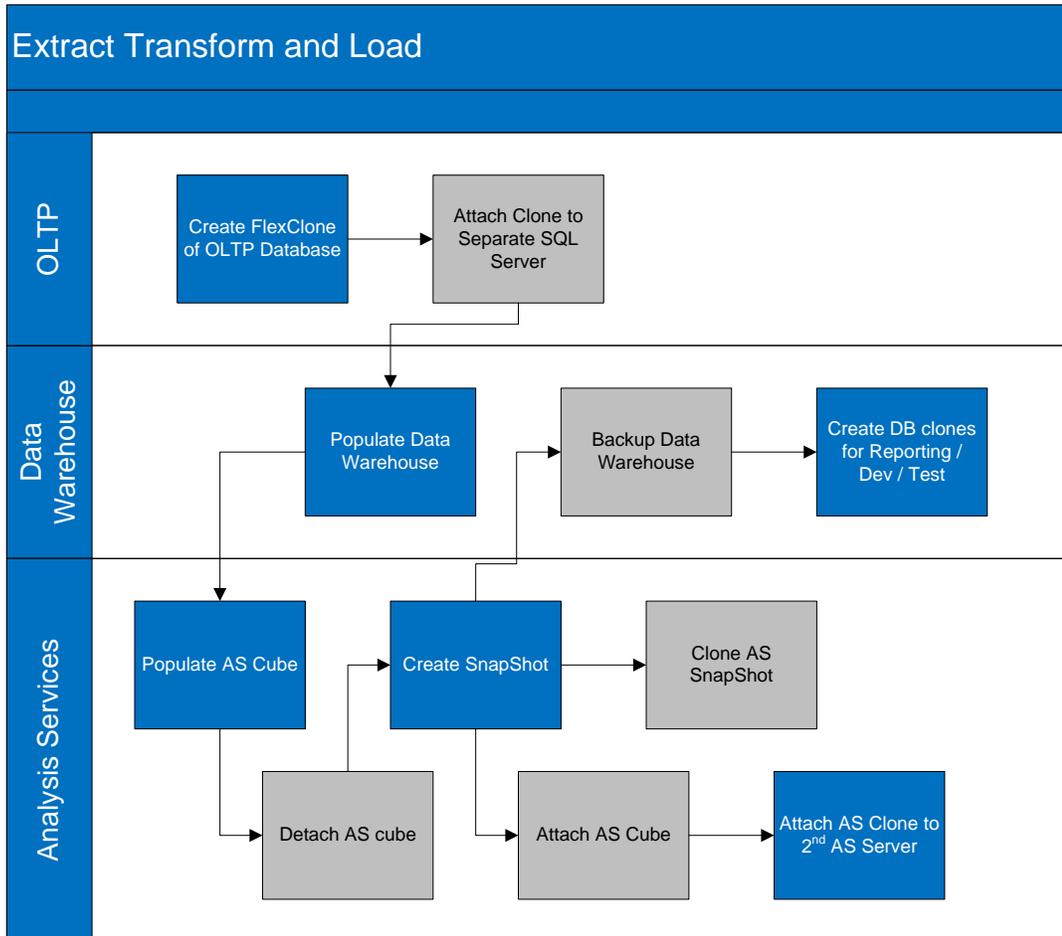
Figure 9) ETL process flow.

3.3   ANALYSIS SERVICES BACKUP AND RECOVERY

The current method for backup and recovery of Analysis Services databases is a copy of the database and its components to a separate area. The copy of the database requires additional disk space and time to create the copy. The NetApp approach to backup and recovery is to create a Snapshot copy of the volume that the data is currently residing on. This Snapshot copy gives you a point in time that you can restore the database. The key for this recovery is to make sure that the location of the data is in the volume that you are creating the Snapshot copy.

Snapshot copies allow for faster backup and restore times. It is important to make sure that you have the backup take place after each change to the cube. Scheduling the Snapshot copy after the Analysis Services cube is processed will make sure of reliability of your backup.

The recovery process requires that you make sure that all Analysis Services processes are off so that the restore of the Snapshot copy will be complete. After restoring the Snapshot copy, then restart the Analysis Services process to resume database operations.

3.4   ANALYSIS SERVICES CUBE CLONING AND SCALABLE SHARED DATABASE INTEGRATION

Analysis Service 2008 introduced the ability to mount read-only databases on separate Analysis Service hosts, making it easier to create a reporting farm. This feature is Scalable Shared Databases (SSDBs), which allows for the creation of multiple read-only databases to help make sure the reporting needs for heavy reporting environments are met. The method for creating these databases comes down to creating a copy of the database to a shared location where multiple hosts can access the database. This copy can be time consuming and expensive due to the actual database space requirements. NetApp can help reduce both the amount of time required to copy the data but also by using FlexClone to reduce the amount of storage required to create these SSDBs in the reporting farm.
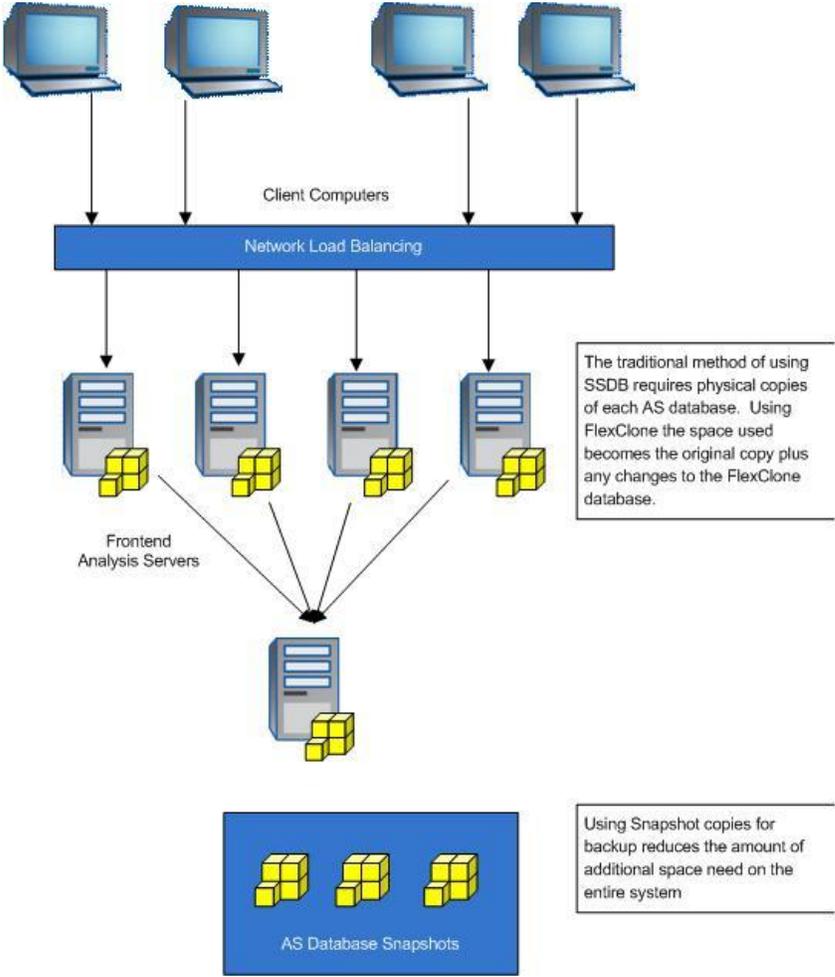
Figure 10) Scalable Shared Database Architecture

The cloning process can work either against previous Snapshot copies of the LUN that has the Analysis Services cube. The Snapshot copy can then be transported to additional Analysis Services instances on separate hosts. Once the cloned LUN has been attached to the new host, attach the AS cube to the new instance. The cloned cube can either operate in read-only mode or read/write. The decision to use read-only or read/write is based on the needs of the reports taking place on the second AS instance.

When mounting the new database with SQL Server Management Studio it is important to remember the need to have the correct mode if you need to have the database in read-only mode. The FlexClone process can create a new version of the database that is either read-only or writable (see Figure 11). The combination of using the SSDB and FlexClone allows fast deployment of reporting databases within the reporting farm. There is also a reduction in the number of physical copies that require duplicate storage space within the environment. This allows scaling quickly and easily within the farm.
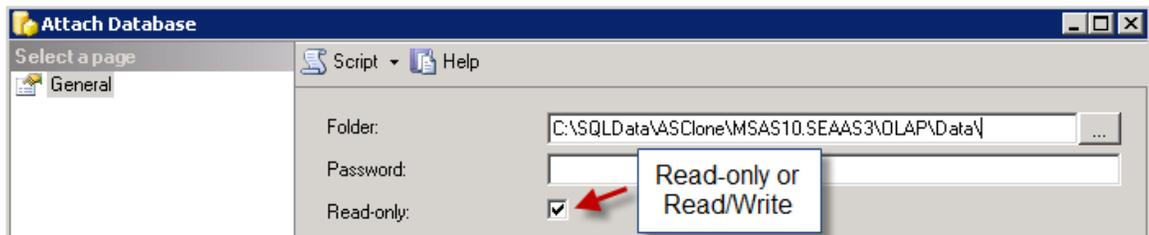
Figure 11) AS attach DB GUI dialog.

4 SUMMARY

The growing needs of today's business intelligence systems and restrictions on storage budgets require an infrastructure that is both flexible and efficient to meet the requirements of the business. The ability to reduce the required amount of storage for the databases, while increasing the size and flexibility of the development and test environments helps increase the efficiency of the underlying storage. NetApp clones allow database queries to be specialized without requiring multiple physical copies of the data. This use of FlexClone for the database increases efficiency without increasing the amount of storage infrastructure required.

The backup and recovery of an Analysis Services database and cubes with NetApp Snapshot is a fast and efficient process. The backup of an Analysis Services database using NetApp Snapshot demonstrates the flexibility of NetApp in maintaining uptime and recovery times for the business intelligence environment. Then integrating the FlexClone volume on the Analysis Services database creates a fast method for deploying Scalable Share Databases across multiple Analysis Services servers.

NetApp storage offers a solution to the expanding business intelligence environment through the use of Snapshot copies and FlexClone volumes. Snapshot copies help to provide backup and recovery for both Analysis Services and SQL Server data warehouses as well as providing the flexible development and test environments for both database platforms.

5 REFERENCES

TR-3373: "Data ONTAP 7G: The Ideal Platform for Database Applications"
http://media.netapp.com/documents/tr-3373.pdf

TR-3410: "Data ONTAP 7G: FlexVol and FlexClone for Microsoft SQL Server"
http://media.netapp.com/documents/tr-3410.pdf

TR-3650: "Microsoft SQL Server 2008: Decision Support Workloads 1TB Data Warehouse on IA64"
www.netapp.com/us/library/technical-reports/tr-3650.html

TR-3696: "Microsoft SQL Server 2005 Relational Engine: Storage Fundamentals for NetApp Storage Systems
www.netapp.com/us/library/technical-reports/tr-3696.html

TR-3768: "SMSQL Best Practice Guide"
http://media.netapp.com/documents/tr-3768.pdf

WP-7078: "Development and Test Solution for Microsoft SQL Server"
http://media.netapp.com/documents/wp-7078.pdf