WHITE PAPER

A Systems Approach to Data Warehousing: The Critical Role of Using Unified Storage in Business Intelligence and Analytics Solutions

Sponsored by: NetApp

Dan Vesset   Richard L. Villars
July 2010

IDC OPINION

The use of pervasive business intelligence (BI) and analytics is a critical competitive necessity for organizations of all sizes and across industries and geographic regions. The key building block in the BI and analytics solutions deployed for decision support and automation is the data warehouse (DW).

Data warehouses, the repositories and engines for data management and analytics, play a critical role in organizations' efforts to "cash in" on their data assets. Today, the amount of data generated by operational systems and Web sources, as well the number of business end users with access to data warehouses, continues to explode. This growth often threatens to overwhelm DW software and the underlying hardware platforms (servers and storage) in terms of performance, effective asset use, and data insecurity. These threats put organizations at competitive and regulatory risk.

As organizations increase their use of fact-based decision-making processes, they must boost technology investments to support scalable and flexible data collection, transformation, loading, management, analysis, reporting, and sharing. A key component of a sound data warehouse investment strategy, therefore, is the deployment of a unified storage solution that delivers scalable performance and reliability as DW requirements grow. The solution must also deliver advanced data management and data replication capabilities that boost utilization of DW assets while protecting organizations from the risks of data loss or data misuse.

This IDC white paper examines the link between more pervasive BI and analytics and organizational competitiveness, identifies requirements that more pervasive BI and analytics pose for data warehouse platforms, and examines how companies can use unified storage solutions like those from NetApp to meet higher expectations for performance, reliability, security, and efficiency in their data warehouse environments.

THE RISE OF THE ANALYTIC ENTERPRISE

At a time of great economic volatility, there has been a significant decrease in the ability of managers to rely effectively on experience or intuition to make decisions. The old cause-and-effect mental models are becoming less relevant, while the demand to respond faster, with greater insight into ongoing internal and external organizational events based on facts rather than experience, is increasing.

Competitive Differentiation

Most decision makers rely on a mix of experience and analytics when making decisions. However, IDC research shows that:

☑ Managers at the most competitive organizations in any given industry are two times more reliant on analytics (rather than intuition) than managers at their least competitive peers.

☑ The most competitive organizations in an industry indicate two times more frequently that output of BI solutions is very influential on actions taken by their employees.

☑ The least competitive organizations in an industry view sharing of data between business units as a loss of control 2.5 times more frequently than their most competitive peers.

☑ There is a clear difference between the level of competitiveness of an organization and the percentage of power users. The median percentage of power users at organizations that ranked themselves as least competitive in their industry was only 9.5%, compared with a 20% median for those that ranked themselves as most competitive.

These are just a few examples that highlight the importance of BI and analytics solutions in differentiating between the most and the least competitive organizations within the industry.

Business Intelligence and Analytics Solution

BI and analytics solutions are composed of technology and processes and people engaged in supporting or automating decision-making processes. As the amount and the types of data and content that organizations collect are exploding, the complexity of deploying and supporting BI and analytics solutions is increasing. While it is often easy to collect this avalanche of information, categorizing it, storing it, and properly using it remain hurdles to improving the decision-making processes.

This data explosion is exacerbated by the fact that different types of decisions are being made by all employees of an organization. Some are strategic decisions made by executives, but the vast majority are operational and tactical decisions made by managers, analysts, and operational and customer-facing employees.

Today, however, much of the technology to support fact-based decision making remains focused on the siloed information access and analysis needs of a select group of decision makers.

IDC refers to the technology needed to support and automate these decision-making needs as business analytics technology. The business analytics technology portfolio, as depicted in Figure 1, is composed of two broad categories of software and hardware technology:
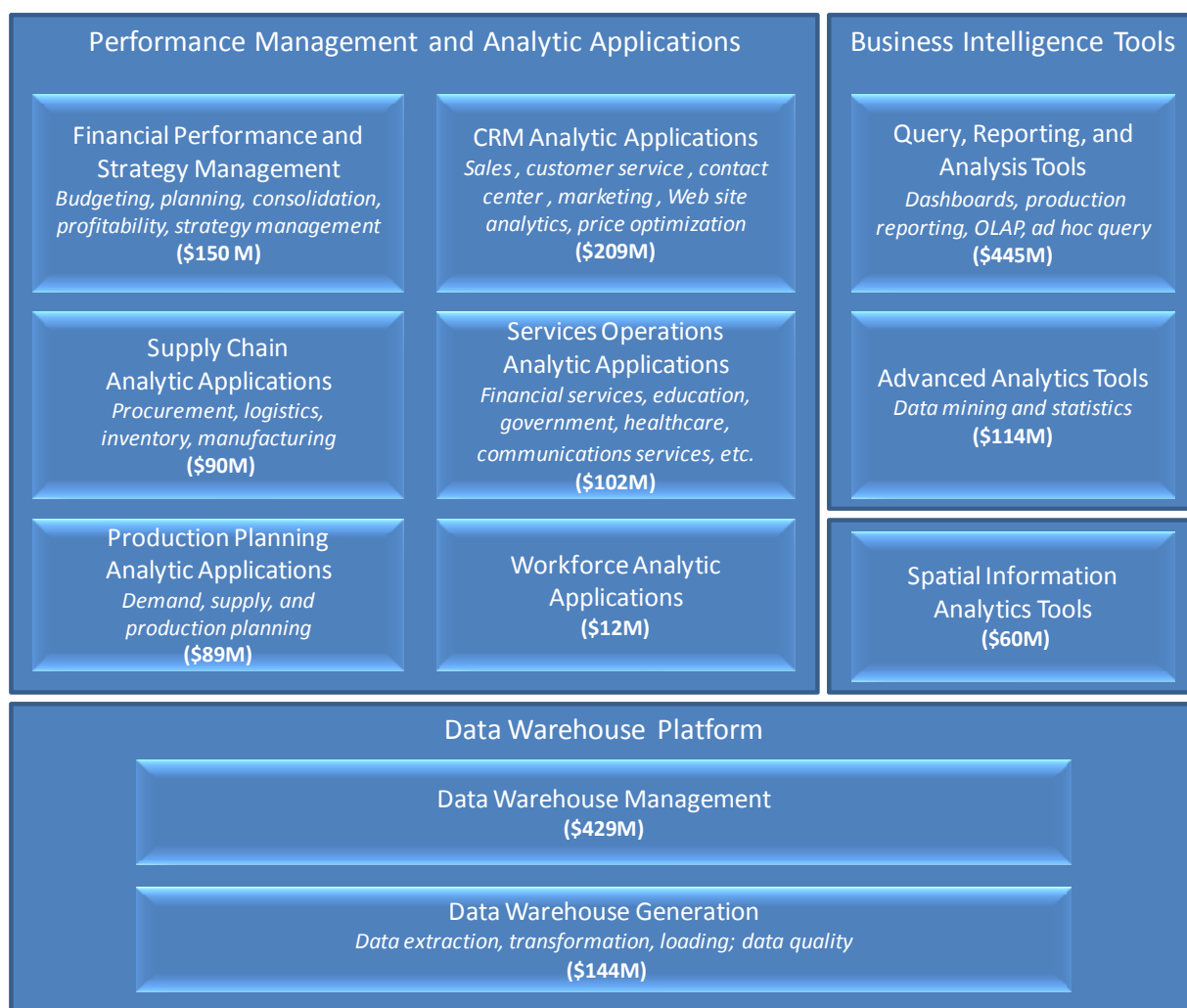
#224039   ©2010 IDC

⊡ Data warehouse platform

⊡ End user–facing BI tools and performance management and analytic applications

The latter category is composed of various horizontal and business process–specific tools and applications combining query, reporting, analysis, and data mining functionality.

FIGURE 1

IDC's Business Analytics Technology Taxonomy, 2010

Performance Management and Analytic Applications		Business Intelligence Tools
Financial Performance and Strategy Management *Budgeting, planning, consolidation, profitability, strategy management* **($150 M)**	**CRM Analytic Applications** *Sales , customer service , contact center , marketing , Web site analytics, price optimization* **($209M)**	**Query, Reporting, and Analysis Tools** *Dashboards, production reporting, OLAP, ad hoc query* **($445M)**
Supply Chain Analytic Applications *Procurement, logistics, inventory, manufacturing* **($90M)**	**Services Operations Analytic Applications** *Financial services, education, government, healthcare, communications services, etc.* **($102M)**	**Advanced Analytics Tools** *Data mining and statistics* **($114M)**
Production Planning Analytic Applications *Demand, supply, and production planning* **($89M)**	**Workforce Analytic Applications** **($12M)**	**Spatial Information Analytics Tools** **($60M)**

Data Warehouse Platform

Data Warehouse Management **($429M)**

Data Warehouse Generation *Data extraction, transformation, loading; data quality* **($144M)**

Source: IDC, 2010

In response to changing competitive pressures, a growing number of organizations are moving toward having more pervasive BI by turning to evidence-based decision making supported by the business analytics technology that enables decision makers to have the best possible intelligence about customers, operations, and the market.

Enabling pervasive BI means much more than simply distributing reports or spreadsheet to all end users. Today, the vast majority of resources continue to go into the development, deployment, and maintenance of the DW platform, which serves as the foundation for the rest of the BI and analytics solution.

THE DATA WAREHOUSE PLATFORM

The DW platform forms the foundation of BI and analytics solution. IDC defines the DW platform as being composed of two segments (refer back to Figure 1): DW generation and DW management.

☐ **Data warehouse generation.** These tools include software used in the design, cleansing, transformation, loading, and administration of the data warehouse.

☐ **Data warehouse management.** These tools include database management system (DBMS) software used to manage and process data in the data warehouse.

In each case, the solution includes the corresponding storage and server hardware. Often this hardware is tightly linked to the DW platform and is not part of the overall IT infrastructure (e.g., the corporate SAN).

Data Warehouse Market Size and Growth

According to the latest IDC market research, the DW management segment of the market reached $9.5 billion in 2009 in combined software and hardware revenue. Figure 2 depicts the DW management software market growth over the past two decades and shows the hardware portion for the past three years.

Over this period, growth came in both the small and medium-sized business (SMB) and the large enterprise segments of the market and included various DW architectures, such as consolidated central enterprise data warehouses and subject-specific data marts, as well as various form factors, such as physical and virtual appliances.

In the short term, IDC expects continued strong growth to come from projects that deploy highly focused DWs or data marts, many of which will be deployed as either physical or virtual appliances. During this time, DW technology that can be purchased and deployed in smaller increments than large, enterprisewide projects has a better chance of being approved and deployed by IT departments or systems integrators. This short-term trend, however, does not diminish (in fact, it will exacerbate) the value and need for a long-term, more scalable enterprisewide DW strategy.

Although the vast majority of the DW platform solutions are provided today as on-premises, licensed technology, some of the supply of DW platform technology will continue to migrate to offerings based on software as a service (SaaS) and eventually as cloud-based offerings that are based on clustered grids, dynamic provisioning of hardware, and subscription and pay-per-use pricing models.

FIGURE 2

Data Warehouse Market Size ($B) and Demand Trends
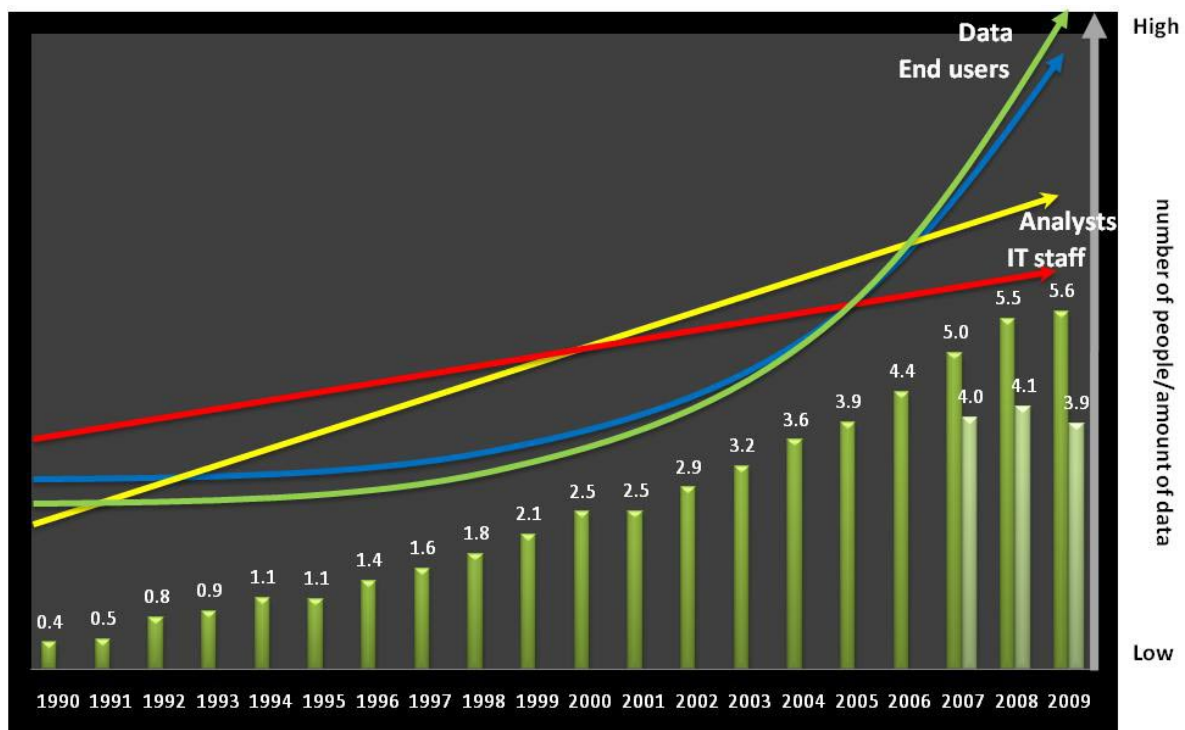
Source: IDC, 2010

Data Warehouse Development Trends

Over the foreseeable future, the DW market will be driven by the following key trends:

⊡ **Growing data volumes.** Availability of data from existing and new sources continues to expand. Organizations are looking to retain and analyze more granular data from existing operational and transactional systems. Others are looking to new sources of data — many from the Web, including data generated by consumers browsing and social networking activity. New sources and new levels of granularity of data in the communications, retail, financial services, manufacturing, utilities, government, and healthcare industries will drive the demand for DW solutions. Already about half of organizations surveyed by IDC indicate that their largest DW is at least 500GB, and within that segment, 17% have DWs larger than 10TB. But 100TB and larger DWs are no longer an anomaly.

⊡ **Growing number of business end users.** The number of business end users with direct access to DWs or indirect access to BI content based on analytics derived from DWs also continues to grow. These users are divided into two broad categories: information consumers and power users. Information consumers require self-service information access through parameterized dashboards or reports that are populated with the data from DWs. Power users are analysts who perform ad hoc and advanced analytics on large amounts of data. These are specialists who place high demands on DWs for both current and historical data.

As shown in Figure 2, the number of dedicated analysts or power users whose role it is to support business end users is not growing at the same rate as the demand for business end users. Nor is the number of IT staff tasked with supporting the DW technology growing at the same rate. The resulting supply-demand gap is therefore expanding, and this will drive continued growth in the automation of performance management and administration of the DW solution stack (including storage and software components).

These trends highlight the new requirements for performance, availability, and scalability of DWs.

Changing Requirements for Data Warehouse Platforms

Data warehouses are no longer "nice-to-have" technology used by a few quantitative analysts. Market research points to the growing criticality of data warehouses and broader business analytics solutions they support as a key factor in providing competitive differentiation enabled by better decision making. The following DW requirements are not new, but they have taken on a growing importance:

☐ **Performance.** Much of the discussion about DW performance circles around query performance as experienced by end users. This is indeed a key metric that enables rapid access to large amounts of data or complex predictive and descriptive analytics. However, just as important is the performance of the data integration process that includes tapes for data extraction, transformation, and loading. The need to consolidate data from multiple sources and update data warehouses at both batch and real-time intervals within shrinking time windows is an important indicator of more pervasive BI and analytics. In a 2009 study, IDC found that 33% of responding organizations already monitor or track information in real time to support decision making, and another 24% plan to do so in the near term.

☐ **Scalability.** Current market trends require data warehouses to scale to support terabytes of data and hundreds and sometimes thousands of concurrent users. The primary trend, today, is for scale-out architecture based on massively parallel processing solutions.

☐ **Availability.** Over just the past three years, the percentage of organizations indicating that their business analytics solutions can't be out of service for more than an hour without having material negative impact on operations has doubled. An average DW still doesn't have the same availability requirements as point-of-sales (POS) systems or email servers, but there are a growing number of use cases where data warehouses drive everything from call center operations and investment trading desks to retail pricing engines and environmental monitoring systems.

☐ **Disaster recovery.** Related to availability requirements are disaster recovery requirements. Whether driven by operational, financial, or data governance requirements, features and functionality for disaster recovery are a key part of overall business analytics solutions.

#224039   ©2010 IDC

☐ **Security.** The market trends also point to a greater need for more granular security management. The competing demands of more self-service access to broader data sets and more stringent privacy regulations need to be addressed through data governance policies. However, once established, these policies need to be manifested in technology, which will enable ongoing management, review, and adjustment of such policies in response to internal and external policy and regulation changes.

In a survey of 336 organizations conducted in 2009, IDC found that less than half of organizations interviewed have above average satisfaction with most of these DW solution capabilities (see Table 1).

TABLE 1

Satisfaction with BI and Analytics Solution (% of Respondents)

Q. On a scale of 1–5, how would you rate the level of satisfaction with the following BI-related initiatives or solution characteristics at your organizations over the past 12 months?

	1 Very Dissatisfied	2	3	4	5 Very Satisfied
Data update frequency	8.7	21.1	31.7	24.2	14.3
Availability (i.e., uptime)	6.9	11.3	24.5	35.2	22.0
Scalability (ability to handle current data volumes and number of users)	8.2	19.0	25.3	31.6	15.8
Manageability (effort to maintain)	11.0	24.0	36.4	20.1	8.4

n = 336

Source: IDC, 2009

The remainder of this white paper discusses how organizations can leverage unified storage and data management solutions from NetApp, a leading storage solutions provider, to better address the fast-evolving requirements of their BI and analytics solutions and especially their DW platforms.

BUILDING MORE SCALABLE AND FLEXIBLE BI SOLUTIONS WITH NETAPP STORAGE

NetApp is a leading worldwide supplier of storage and information management solutions for enterprises. It delivers a broad portfolio of systems, software, and support services designed to provide organizations with a scalable, flexible, and efficient foundation for storing, accessing, and protecting valuable information assets.

At the core of NetApp's solutions is a family of unified, networked storage systems that scale from compact solutions designed to meet the requirements of SMBs or remote offices to large, high-performance systems designed to meet the intense storage workloads of the largest enterprises. NetApp's solutions support multiple connection options (e.g., Fibre Channel, iSCSI, NFS, and CIFS) so that organizations can leverage a common storage asset while having flexibility in selecting the optimal connection option for different applications (e.g., BI).

From the largest to smallest systems, administrators need to learn and use only one set of tools for storage provisioning/management, data protection, remote data replication, and data retention. Table 2 provides a list of important storage use cases and a partial list of NetApp's capabilities that enhance enterprises' storage processes in support of BI and analytics solutions.

TABLE 2

NetApp Unified Storage Architectural Components and Their Value in Support of Data Warehousing

Use Case	NetApp Component(s)	Customer Benefit(s)
Data Warehouse Scalability through consolidation of information on shared storage	Replace isolated direct-attached storage pools associated with individual data marts with consolidated SAN-attached storage systems; using FlexClone Snapshot copies for space-efficient replicas of test/development	Highly scalable with modular expansion that reduces cost of incremental expansion; larger aggregate capacity including support for multiple storage tiers to reduce cost of storage
Data Warehouse Availability and Fast Recovery through advanced data replication and management	Multiple FlexClones for many point-in-time (PIT) replicas with little impact on performance; multiple copies allow frequent checkpoints	Choice of PITs upon which to perform additional data manipulation; minimal data loss; any size database; Snapshot copies verified for integrity and then stored; instantaneous access to backup copies
Data Warehousing Performance and Data Integration	Scalable SAN/NAS storage systems with flash cache for high IOP requirements; Data ONTAP operating environment and FlexVol for automated and flexible provisioning	Support for high-performance I/O requirements through intelligent use of solid state disks (SSDs); quick time to value for data integration with on-demand storage provisioning
Data Warehousing Security	Protection Manager for centralized automation of data replication and data movement; DataFort and Lifetime Key Management for heterogeneous data encryption; SNAPLock for ensuring compliance with data retention policies	Centralized control/audit of data replication to protect against inappropriate data copy; encryption of data at rest to protect against data theft; ensure retention of data sets to comply with regulatory requirements
Data Warehousing Backup and Recovery	Replicated copies using SnapVault; SnapManager for Microsoft SQL Server; SnapManager for Oracle	Application-aware, offsite separation to protect from regional disasters or system failures; improved compute and storage efficiency through the use of "follow the sun" workload migration; optional backup to tape for long-term archiving

Source: IDC, 2010

One example of how companies can leverage NetApp's FlexClone capabilities is dealing with the proliferation of data marts to meet many different groups' data warehousing needs. In any BI and analytics system, organizations invariably need copies of some or all of their data for individual data marts, development and test, and other use cases. FlexClone enables the easy creation and management of multiple, instant virtual copies of data with almost no additional storage overhead (compare

that with the 2x storage requirement of making a full copy plus the hours it takes for a full copy to complete). Because only changes that are made to the clone are stored, the additional storage requirement is minimal.

This ability is particularly important in development and test environments that require copies of the DW for updating ETL processes or designing new analytic applications. Because part or all of the data in the DW can be instantly cloned without requiring additional storage space at the time of creation, IT teams can create and allocate many individual, writable copies of data. Activities that once had to be performed sequentially because of storage limitations can now be done in parallel, increasing the speed of testing and deploying new applications.

This example highlights how NetApp's portfolio of integrated features focuses on boosting efficiency in terms of asset utilization and operations. This focus is especially critical for fast-growing environments such as DWs, where information creation and reuse are both rapid and unpredictable.

Without this focus on efficiency, the storage growth and management challenges associated with DW platforms can be highly disruptive to application development and storage operations in the datacenter and ultimately to business decision makers who rely on these applications.

NetApp Solutions for Partner Business Applications: Automation and Optimization

The final element in NetApp's storage solutions is recognition that advanced services must be optimized to address the unique requirements of critical business applications such as data warehouses. NetApp delivers targeted solutions that simplify and/or automate configuration and storage management tasks associated with specific applications. Critical applications addressed directly by NetApp's products include:

☑ SnapManager for Microsoft's Exchange, SharePoint and SQL servers

☑ SnapManager for Virtual Infrastructure (VMware, Citrix, and Microsoft)

☑ SnapManager for SAP

☑ SnapManager for Oracle

Opportunities and Challenges for NetApp

The criticality of solutions provided by NetApp presents the company with an opportunity to capitalize on the accelerating demand in the broader BI and analytics market. The trends outlined in this white paper, including the growth in data being collected, stored, managed, and analyzed, are encouraging IT groups to evaluate solutions that provide the appropriate level of performance, scalability, availability, security, and manageability features and functionality to address the demands for fact-based decision support processes.

At the same time, this opportunity is not without its challenges. In a competitive market, differentiation for NetApp must come from a combination of technical capabilities and effective customer and partner support practices. Among the challenges that NetApp and other storage suppliers continually face in the DW and BI environments are growing and evolving concerns about data security and privacy. Delivering the right levels of performance, reliability, and availability is no longer enough. NetApp must work with its customers to make sure that the data used in DW and BI environments is well protected from accidental or intentional disclosure and misuse. This will require more coordination of encryption and access control functions and policies with leading DW platform software suppliers.

ESSENTIAL GUIDANCE

As BI and analytics usage reaches critical mass in organizations, the need to address management, scalability, reliability, compliance, and cost containment capabilities becomes a critical part of the data warehouse investment decision process. Making the right architectural storage decision can have a profound effect on addressing these requirements. Storage solutions with scalable performance and advanced data management capabilities will play a key role in enabling data warehouse solutions that scale to enterprise-ready deployments.

NetApp — with its broad portfolio of unified storage systems; storage and information software solutions like SnapManager for Oracle, Microsoft SQL Server, and SAP; and wide range of virtual server environments — provides an enterprise-class foundation for scaling, managing, and protecting data warehouse deployments. It enables an advanced set of storage optimization, data protection and recovery, data manageability, and risk mitigation capabilities. NetApp's comprehensive set of features and options is also essential for reliably and cost-effectively growing data warehouse environments in the current budget-constrained economic climate.