

3D Data Management: Controlling Data Volume, Velocity, and Variety. Current business conditions and mediums are pushing traditional data management principles to their limits, giving rise to novel, more formalized approaches.

META Trend: During 2001/02, leading enterprises will increasingly use a centralized data warehouse to define a common business vocabulary that improves internal and external collaboration. Through 2003/04, data quality and integration woes will be tempered by data profiling technologies (for generating metadata, consolidated schemas, and integration logic) and information logistics agents. By 2005/06, data, document, and knowledge management will coalesce, driven by schema-agnostic indexing strategies and portal maturity.

The effect of the e-commerce surge, a rise in merger/acquisition activity, increased collaboration, and the drive for harnessing information as a competitive catalyst is driving enterprises to higher levels of consciousness about how data is managed at its most basic level. In 2001/02, historical, integrated databases (e.g., data warehouses, operational data stores, data marts), will be leveraged not only for intended analytical purposes, but increasingly for intra-enterprise consistency/coordination. By 2003/04, these structures (including their associated metadata) will be on par with application portfolios, organization charts, and procedure manuals for defining a business to its employees and affiliates. Data records, data structures, and definitions commonly accepted throughout an enterprise reduce fiefdoms pulling against each other due to differences in the way each perceives where the enterprise has been, is presently, and is headed. Readily accessible current/historical records of transactions, affiliates (partners, employees, customers, suppliers), and business processes (or rules), along with definitional and navigational metadata (see ADS Delta 896, 7 Aug 2000) enable employees to paddle in the same direction. Conversely, application-specific data stores (e.g., accounts receivable versus order status), geographic-specific data stores (e.g., North American sales vs. International sales), offer conflicting or insular views of the enterprise that, while important for feeding transactional systems, provide no “single version of the truth,” giving rise to inconsistency in the way enterprise factions function. While enterprises struggle to consolidate systems and collapse redundant databases to enable greater opera-

tional, analytical, and collaborative consistencies, changing economic conditions have made this job more difficult. E-commerce, in particular, has exploded data management challenges along three dimensions: volumes, velocity, and variety. In 2001/02, IT organizations must compile various approaches to have at their disposal for dealing with each.

Data Volume. E-commerce channels increase the depth/breadth of data available about a transaction (or any point of interaction). The lower cost of e-channels enables an enterprise to offer its goods or services to more individuals or trading partners, and up to 10x the quantity of data about an individual transaction may be collected—thereby increasing the overall volume of data to be managed. Furthermore, as enterprises come to see information as a tangible asset, they become reluctant to discard it. Typically, increases in data volume are handled by purchasing additional online storage. However, as data volume increases, the relative value of each data point decreases proportionately — resulting in a poor financial justification for merely incrementing online storage. Viable alternates/supplements to hanging new disk include:

- Implementing tiered storage systems (see SIS Delta 860, 19 Apr 2000) that cost-effectively balance levels of data utility with data availability using various media

Business Impact

Attention to data management, particularly in a climate of e-commerce, and greater need for collaboration can enable enterprises to achieve greater returns on their information assets.

- Limiting data collected to that which will be leveraged by current or imminent business processes
- Limiting certain analytic structures to a percentage of statistically valid sample data
- Profiling data sources to identify and subsequently eliminate redundancies
- Monitoring data usage to determine “cold spots” of unused data that can be eliminated or offloaded to tape (e.g., Ambeo, BEZ Systems, Teleran)
- Outsourcing data management altogether (e.g., EDS, IBM)

Data Velocity. E-commerce has also increased point-of-interaction (POI) speed and, consequently, the pace data used to support interactions and generated by interactions. As POI performance is increasingly perceived as a competitive differentiator (e.g., Web site response, inventory availability analysis, transaction execution, order tracking update, product/service delivery), so too is an organization’s ability to manage data velocity. Recognizing that data velocity management is much more than a physical bandwidth and protocol issue, enterprises are implementing architectural solutions such as:

- Operational data stores (ODSs) that periodically extract, integrate, and reorganize production data for operational inquiry or tactical analysis
- Caches that provide instant access to transaction data while buffering back-end systems from additional load and performance degradation (unlike ODSs, caches are updated according to adaptive business rules and have schemas that mimic the back-end source)
- Point-to-point data routing between databases and apps (e.g., D2K, DataMirror) that circumvents high-latency hub/spoke models that are more appropriate for strategic analysis
- Designing architectures that balance data latency with app data requirements and decision cycles, without assuming the entire information supply chain must be near real time

Data Variety. Through 2003/04, no greater barrier to effective data management will exist than the variety of incompatible data formats, non-aligned data structures, and inconsistent data semantics. By this time, inter-

change/translation mechanisms will be built into most DBMSs. But until then, app portfolio sprawl (particularly when based on a “strategy” of autonomous software implementations due to e-commerce solution immaturity), increased partnerships, and M&A activity intensifies data variety challenges. Attempts to resolve data variety issues must be approached as an ongoing endeavor encompassing the following techniques:

- Data profiling (e.g., Data Mentors, Metagenix) to discover hidden relationships and resolve inconsistencies across multiple data sources (see ADS Delta 898, 23 Aug 2000)
- XML-based data format “universal translators” that import data into standard XML documents for export into another data format (e.g., infoShark, XML Solutions)
- Enterprise application integration (EAI) predefined adapters (e.g., NEON, Tibco, Mercator) for acquiring and delivering data between known apps via message queues, or EAI development kits for building custom adapters
- Data access middleware (e.g., Information Builders’ EDA/SQL, SAS Access, OLE DB, ODBC) for direct connectivity between applications and databases
- Distributed query management software (e.g., Enth, InfoRay, Metagon) that adds a data routing and integration intelligence layer above “dumb” data access middleware
- Metadata management solutions (i.e., repositories and schema standards) to capture and make available definitional metadata that can help provide contextual consistency to enterprise data
- Advanced indexing techniques for relating (if not physically integrating) data of various incompatible types (e.g., multimedia, documents, structured data, business rules)

As with any sufficiently fashionable technology, users should expect data management marketplace ebb/flow to yield solutions that consolidate multiple techniques/solutions that are increasingly app/environment specific (see Figure 1 in Addendum). In selecting a technique or technology, enterprises should first perform an information audit assessing the status of their information supply chain to identify and prioritize particular data management issues.

Bottom Line

In 2001/02, IT organizations must look beyond traditional direct brute-force physical approaches to data management. Through 2003/04, practices for resolving e-commerce accelerated data volume, velocity, and variety issues will become more formalized/diverse. Increasingly, these techniques involve tradeoffs and architectural solutions that involve/impact application portfolios and business strategy decisions.

Figure 1 — Data Management Solutions

▲ **Volume**

- ▶ Tiered storage/hub and spoke
- ▶ Selective data retention
- ▶ Statistical sampling
- ▶ Redundancy elimination
- ▶ Offload “cold” data
- ▶ Outsourcing

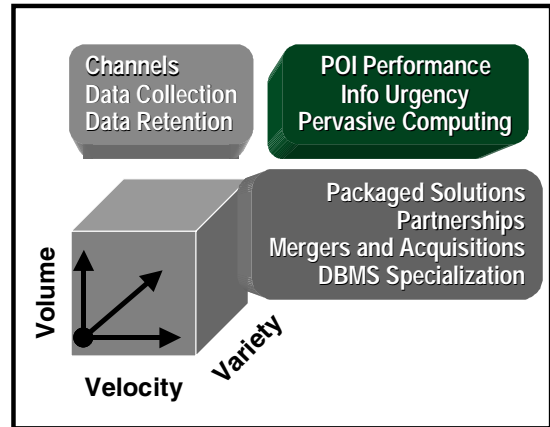
▲ **Velocity**

- ▶ Operational data stores
- ▶ Data caches
- ▶ Point-to-point data routing
- ▶ Balance data latency with decision cycles

▲ **Variety**

- ▶ Inconsistency resolution
- ▶ XML-based “universal” translation
- ▶ Application-aware EAI adapters
- ▶ Data access middleware and ETLM
- ▶ Distributed query management
- ▶ Metadata management

● **E-Business-Driven Information Explosion Factors**



Extending data management options enables greater returns on information assets

Source: META Group

META Delta
