

# Make sure your ODS isn't odious

Deciding when operational data stores should be merged, removed or left alone. *by Dan Graham*

The operational data store (ODS) has been a standard architectural feature since the 1990s when the book *Building the Operational Data Store*, by Bill Inmon, Claudia Imhoff and Greg Battas, first gave it a name. Historically, this architectural construct has served us well in bridging the technology gap between online transaction processing (OLTP) systems and enterprise data warehousing. Now, as enterprise data warehouse (EDW) technology evolves, the arrival of active workload management dramatically alters the technical foundations of the ODS. This is not to say

the ODS is dead; some ODSs will be consolidated into the active data warehouse, some will stay as they are and, it is hoped, some will just go away.

## Many faces

In the mid-1990s, six distinct ODS types were defined by Inmon, Imhoff and Battas. The “canonical” definition that was later adopted by Gartner says an ODS is:

- > Subject-oriented (like a data warehouse)
- > Made up of integrated data (standard, consistent data formats)

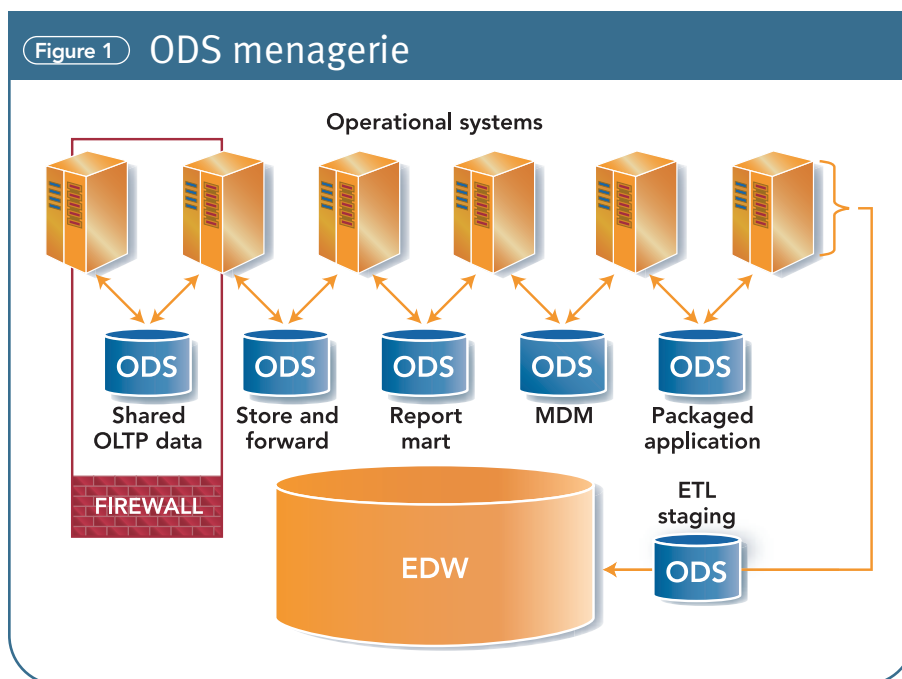
- > Volatile (changes as often as the source system)
- > Current (low-latency data capture; no historical detail)

When limited in scope to customer or product data, the canonical ODS is similar to today's master data management (MDM).

Yet, as one humorist said, “An ODS is any data extracted from production systems that's not in the EDW.” A quick Google search finds that the following workloads are often called an ODS: (See figure 1, left.)

- > Operational reporting (i.e., data marts in disguise)
- > Extract, transform and load (ETL) staging areas
- > Embedded operational analytics (i.e., SAP-BW)
- > Queuing service to business partners over the firewall
- > Store and forward transactions
- > A report mart when an EDW is politically forbidden
- > Low-latency repository for master records exchange (i.e., MDM precursors)

Yet, in reality, only the last definition fits the canonical definition of an ODS. Even so, the most common repositories named “ODS” are actually data marts in disguise used for operational reporting. These ODS report marts are needed because most operational application reports only include data contained from the oper-



Some organizations employ a jumbled collection of ODSs that support the varying workload environments.

ational application—they lack the integrated data from other applications. Additionally, OLTP performance would suffer if complex reports were run against the operational database. This is the classic mixed workload problem. Furthermore, up-to-the-minute operational reports cannot be supplied by EDWs that load data only once each night. Hence, an “operational report mart” is born on its own server and called an ODS.

### A surplus of report marts

Technology limits and years of revolving-door developers have left many IT organizations with a glut of ODS report marts. The marts were quick solutions at the time. While most ODSs started out as small, simple deployments, demanding users got more subject areas, historical data and summary tables added. Data movement, data replication and data scope requirements grew accordingly. While it was never planned this way, some sites have 10 to 30 ODSs—only a few of which

fit the original goals of an ODS. (See figure 2, below.)

But some ODS deployments match the canonical definition. This MDM precursor serves as the “system of record” to multiple applications. These kinds of ODSs also proliferated; in some cases, one ODS was created per consumer touchpoint (i.e., channel) or major application. Notice that, often, these master records also exist inside the EDW.

All of this leaves the DBA and CIO with a growing problem: Duplicate ETL jobs refreshing duplicate data in multiple ODSs. Each ODS consumes a server, storage, software licenses, backups, labor, etc. Many IT sites report spending more on ODS maintenance, support and licenses than on their EDW. Sound familiar? The similarities with data mart proliferation are not a coincidence. The same forces are at work but from slightly different angles. Like data mart consolidation, ODS consolidation into an EDW will reduce labor and costs.

### Innovation required

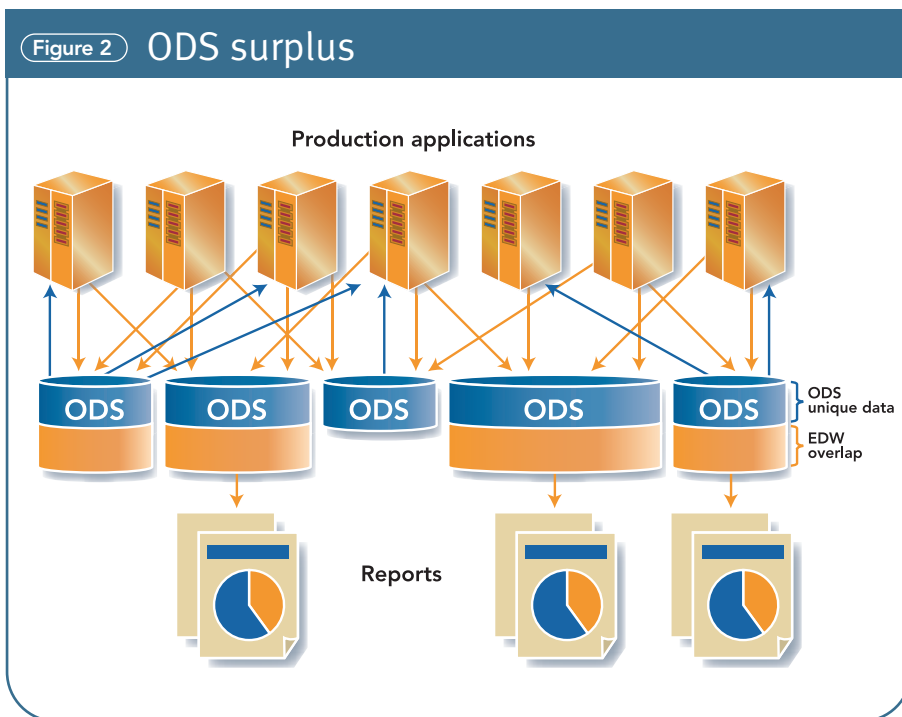
Unlike data mart consolidation, ODS consolidation could not begin until technical innovations arrived. First is the need for low-latency data loading to provide fresh data in the EDW for operational reports. It was not until continuous loading technology, such as TPump, connected the message queuing bus to the data loaders that fresh low-latency data arrived in the EDW. With the modern enterprise service bus (ESB) and TPump enhancements, this capability is easier than ever to deploy. Many DBAs add replication from GoldenGate and the mini-batch data load technique to get an arsenal of choices for fresh data loading.

In the Teradata 2006 Data Warehouse Survey, nearly 40% of respondents said they have some form of “active load” in operation. A few are loading data continuously, many more with latency in minutes or hours. But users must be cautious that near real-time data loading does not cause poor performance for concurrent operational reports or queries.

Similarly, to consolidate the MDM-style ODS, tactical queries must run at sub-second speeds. This mixed workload of data loading, ad hoc queries, tactical queries and reports needs to meet multiple service-level goals—requiring another innovation.

Teradata Priority Scheduler will ensure certain tasks have a “fast path” through the system so that high-priority service-level goals are met. At runtime, Teradata Active System Management can analyze each database request before and during execution, assigning it to workload groups with associated priorities. Thus, when CPU is allocated, Teradata Active System Management can favor real-time queries first, reporting second and data loading third.

In the survey mentioned earlier, more than 50% of Teradata customers said they have deployed either Teradata Priority



An oversupply of ODSs within an organization can result in data management overhead.

Scheduler or Teradata Active System Management. So, with these two key technologies in place—active load and active workload management—ODS consolidation is now possible.

### Consolidating ODS into the active data warehouse

Just because something is called an ODS does not make it a consolidation candidate; it's the actual workload that counts. Most organizations will benefit by first consolidating ODS report marts and MDM precursors. In contrast, it is best to avoid consolidating an ODS that receives high-volume OLTP updates. (See figure 3, below.)

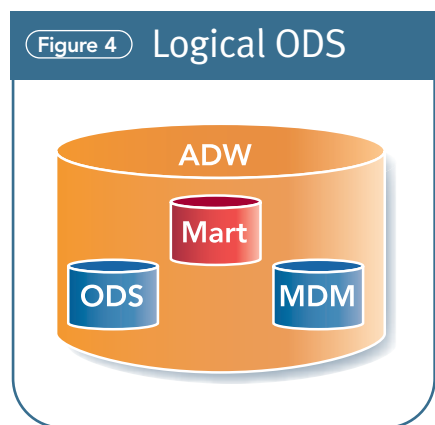
Start by developing a plan to consolidate several ODS instances over a 12- to 18-month period. ODS consolidation targets should be selected based on:

- > Amount of data to migrate
- > Workload redundancy
- > Return on investment (ROI)
- > Project complexity

Consolidating an ODS into the EDW requires the following steps:

- > Implement workload management (i.e., Teradata Active System Management)
- > Merge the ODS data model into the EDW
- > Implement active loads using TPump, replication and mini-batch
- > Tune ODS queries in the EDW using tactical query best practices
- > Run parallel production environments to test the logical ODS

Implementing Teradata Active System Management is the recommended first step because it enhances the existing EDW and gives you time to optimize Teradata Active System Management while working on subsequent tasks. Presumably, if the ODS and EDW both have normalized data models, merging them should not be difficult. If the Teradata EDW is reasonably mature, the ODS report mart data might already be in the EDW and, if so, only the appli-



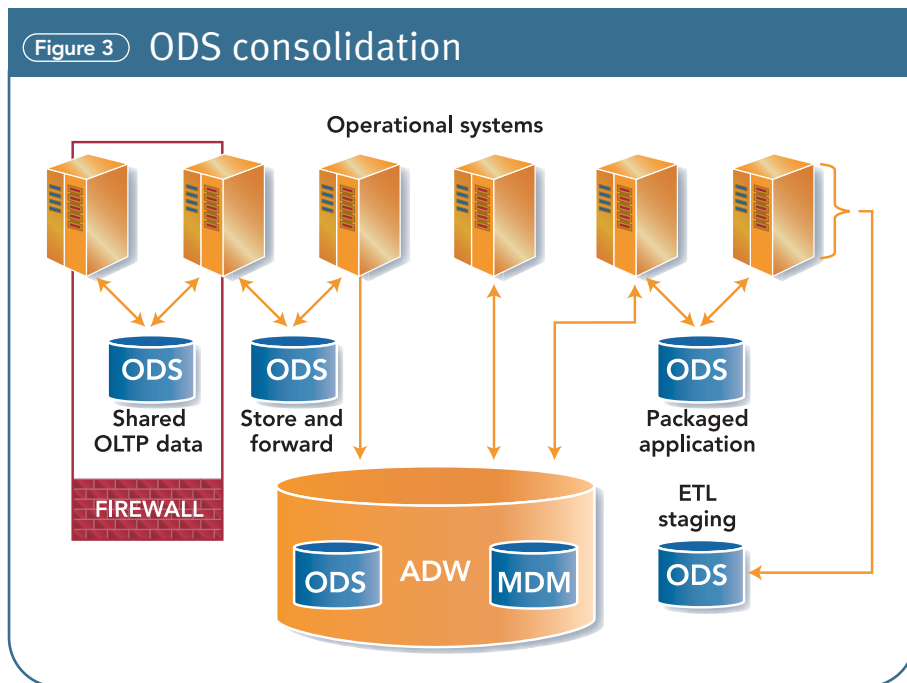
Within Teradata, an ODS report mart can be designated as a “logical ODS.”

cation accessing the ODS will need some migration effort.

While tuning tactical queries for the MDM precursor workload, pay special attention to techniques that localize the work, such as row-level locking and single-AMP access.

Not every ODS is odious, onerous or odorous; some are vital. ODS report marts are excellent candidates for consolidation into the most cost-effective platform in your architecture—the EDW. As in data mart consolidation, Teradata recommends moving ODS instances into the EDW where they end up being a “logical ODS.” (See figure 4, above.) This should be transparent to users and applications with no difference in service levels. What changes is a reduction in costs and complexity. The result, of course, is an *active* data warehouse—active loading with active workload management to meet service level agreements. That will make your ODS melodious! **T**

*Dan Graham joined Teradata in 1989. He later moved to IBM where he became a strategy executive for its Global Business Intelligence Solutions. Now, with more than 30 years in IT, Graham has returned to Teradata to lead its Active Data Warehouse Technical Marketing.*



The greatest benefit will be realized by consolidating ODS report marts and MDM precursors first into an active data warehouse.