

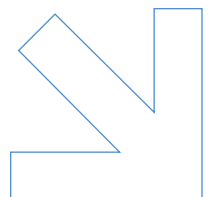
A New Perspective on Data Warehouse Architectures



baseline

C O N S U L T I N G

www.baseline-consulting.com





Evan Levy is an industry recognized information technology expert, speaker, and business consultant. As partner and co-founder of Baseline Consulting, he has developed enterprise information systems at Boeing, Verizon, and Charles Schwab. He has been published in a wide array of industry magazines, is a faculty member of The Data Warehousing Institute, and a featured speaker at Marcus Evans, DCI, the CRM Association, and DAMA International conferences and other industry events.

A New Perspective on Data Warehouse Architectures

BY EVAN LEVY

The story goes that early in the 20th century, the president of the U.S. Patent Office recommended shutting the office down. The justification was that “Everything that can be invented has already been invented.” Now, the story may be more urban myth than fact, but based on the volume of patent applications since then, the lesson is clear: innovation is alive and well and continuing to flourish.

This adage is true of data warehousing if it’s true of anything. Data warehousing has gone through its own evolution, beginning with large centralized platforms, migrating to client-server, moving toward semi-dependent servers—with a backlash somewhere in the middle—to the co-existence of a range of technology platforms and vendor software products. Every time someone predicts its demise, data warehousing solves an even bigger, more complex set of problems.

Irrespective of its architecture or scale, the objective of data warehousing to help business people make faster and more accurate decisions through the access and analysis of detailed data, has remained a central tenet of its value.

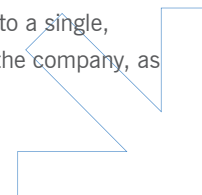
It just so happens that, while companies continue to need and leverage business analytics, data warehouse technologies remain resource and maintenance-intensive. More than ever before, data warehouse architects and development teams are being asked by management to rationalize costs and even research alternatives. As it turns out, management loves the new knowledge being gleaned via business intelligence. It just doesn’t particularly appreciate what goes into deploying business intelligence to the enterprise.

But data warehouse architecture is approaching its next milestone. The emergence of important new appliance technologies promises new efficiencies and economies of scale that couldn’t be envisioned five years ago when Unix servers were all the rage. Before we discuss the promise, let’s review the current state-of-the-art.

CLASSIC DATA WAREHOUSE ARCHITECTURES: ENTERPRISE VERSUS FEDERATED

Amdahl. Teradata. Oracle. IBM. These vendors are familiar to anyone who was involved in migrating data from operational systems and onto separate platforms back in the early 1980s. The objective of these efforts was to alleviate the processing and expensive CPU usage of the mainframes that were running our businesses by replicating data onto separate platforms for the purposes of analysis. The capability was called decision support—which became widely adopted as business users and executives alike could run reports on-demand, without relying on programmers to generate them—and the platform was the data warehouse.

The concept of an enterprise data warehouse architecture was and is to acquire and load data onto a single, centralized technology platform to avail a “single version of the truth” to different departments across the company, as shown in Figure 1.



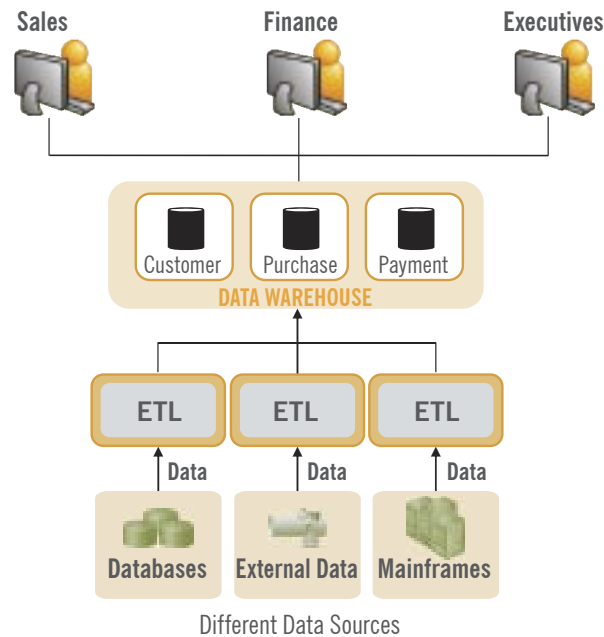


Figure 1: An Enterprise Data Warehouse

In the enterprise data warehouse environment, data is gathered from disparate operational sources both within and outside of the company, cleaned up through a process called ETL (Extraction, Transformation, and Loading), and then deployed to end-users with business intelligence software tools. The advantages of the enterprise data warehouse approach are that it:

- Establishes data integrity through the centralization of business rules and data cleansing processes
- Reduces duplication of data across systems
- Reduces staff resource costs, since fewer systems mean fewer system administrators and DBAs
- Benefits from data loading economies of scale
- Creates the acknowledged system of record for business analysis across the company
- Is inherently designed to be flexible, it can serve many masters and evolve with business requirements

As an alternative to the enterprise data warehouse, the federated data warehouse architecture looks different and offers a different set of benefits. By its nature, the federated data warehouse is comprised of one or more smaller platforms, known as data marts, which serve more focused purposes, as shown in Figure 2.

The advantages of the federated environment are:

- It allows organizations to maintain autonomous business content.
- Two or more small systems are typically less expensive than one big one, thus hardware is usually less expensive.

- The smaller scope of individual data marts mean they're generally simpler and faster to implement.
- Benefits from data loading economies of scale.
- Because they are organizationally aligned, funding and development approval is usually more straightforward than for enterprise systems.

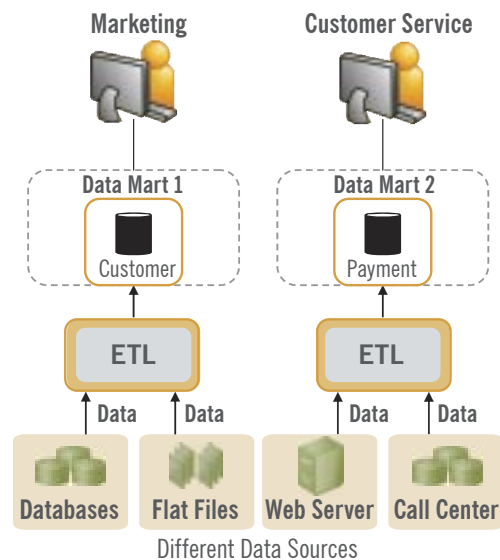


Figure 2: A Federated Data Warehouse

Both enterprise and federated data warehouse architectures have served an important function in their respective organizations: they've helped deploy business data to business people who can leverage the resulting information to make fast and differentiated business decisions. Despite this value—which some companies can peg at millions of dollars of ROI—there are typical challenges with both architectures that companies faced at the beginning and are still grappling with as their data warehouses continue to grow:

- **The care and feeding of the equipment.** Hardware versions, operating system versions, tracking and monitoring storage, file system, operating system and database backups, and other seemingly routine maintenance tasks all combine to add work, time, and expense.
- **Working with disparate vendors.** The larger and more complex the data warehouse environment, the greater the number of products and the more rampant the conversations about maintenance contracts.
- **Ongoing support and skills.** The ongoing compatibility of various software products, their versions, and operating systems is only the tip of the iceberg. When problems arise, a company needs a broad base of skills, including operating system, storage products, and DBMS products, to hire, train, and nurture. (The alternative of outsourcing these skills brings its own time and training challenges.)

Although the “classic” data warehouse architectures I've described here have been wildly successful, new architectural alternatives have emerged that can not only alleviate some of the processing and cost challenges, but can result in even faster delivery of business value.

TO THE RESCUE: DATA WAREHOUSE APPLIANCE TECHNOLOGY EMERGES

One of the most exciting new breakthroughs in data warehousing is the data warehouse appliance (e.g., DATAlegro, Netezza). A data warehouse appliance is packaged hardware and software that pre-integrates all components (server, software, OS, RDBMS, storage), providing a veritable “out of the box” data warehouse solution. Similar in concept to network security or network storage appliances, a data warehouse appliance provides data warehouse management functionality to an IT infrastructure.

Data warehouse appliance vendors make a compelling argument for their products being part of an overall data warehouse infrastructure. Many companies are facing aging equipment that, while a long way from being obsolete, is ever more expensive to maintain and upgrade. In the worst case, a data warehouse appliance can provide the processing and storage necessary to delay a costly upgrade. Or, it can represent the eventual target data warehouse, as more data is integrated onto the appliance and migrated from the data warehouse. For an aging enterprise data warehouse, this can be an elegant alternative, as shown in Figure 3.

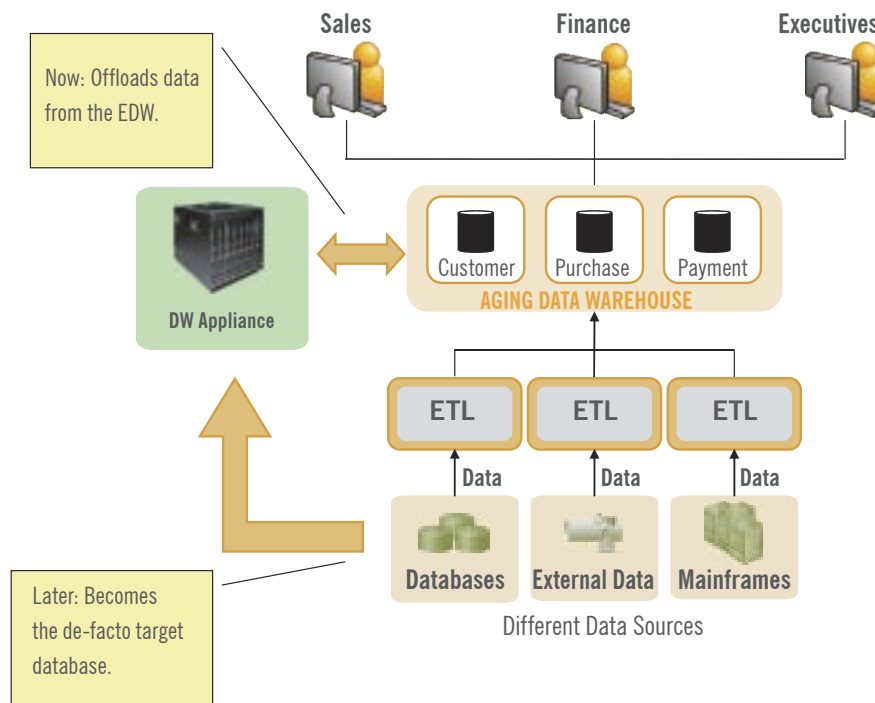


Figure 3: Augmenting and Offloading an EDW

This approach allows the company to migrate data onto the data warehouse appliance at its own pace, and, if needed, even sunset the legacy data warehouse if desired. And it mitigates the need for complex data migration work that would be required if adding a standard server.

There are similar opportunities to augment the federated data warehouse environment. For instance, a company that relies heavily on partner channel for its sales may have a small direct sales organization. Nevertheless, sales management needs information about compensation, territory assignments, and profitability, but doesn't have the resources or budget to build its own data mart.

impacting much of the staffing and infrastructure layers—including system and data administrators, and hardware and software configuration—necessary to deploy a data mart. The administrative functions become much simpler, offering the use of GUI-based wizards instead of the laundry list of configuration files and command line utilities. Using a data warehouse appliance is, comparatively speaking, plug and play.

Moreover, as sales' reporting needs grow—say they'd like to begin to monitor customer trouble tickets—they can leverage the data warehouse appliance to acquire additional operational data from the call center system, as shown in Figure 4.

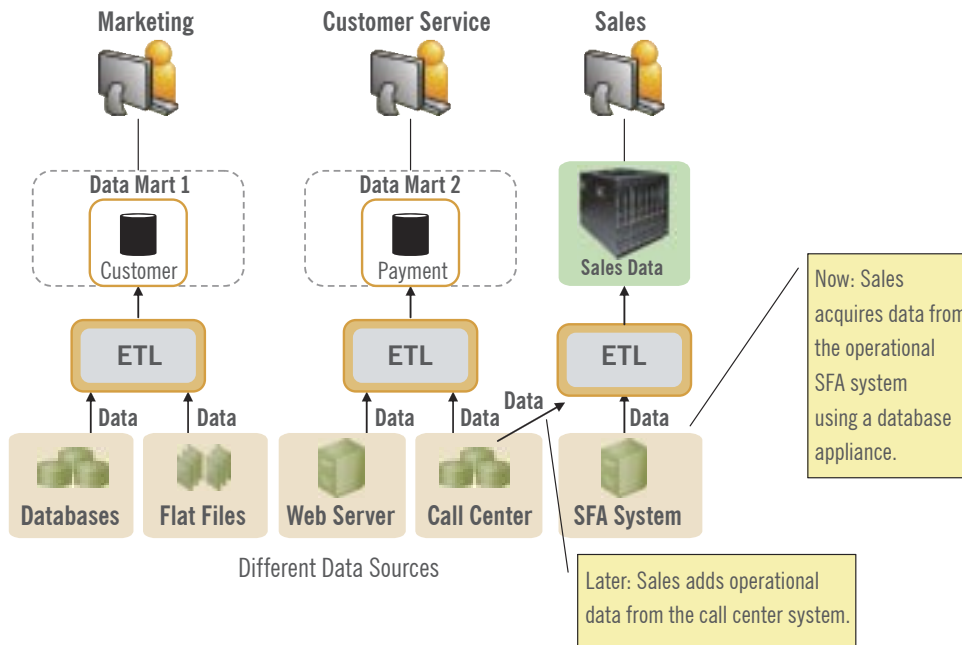


Figure 4: Standard Server versus Database Appliance

Data warehouse appliances offer the opportunity to add computing power in a straightforward and cost effective way. Eventually, this technology can scale to replace upgrades or entire platforms.

FUTURE-PROOFING YOUR DATA WAREHOUSE PLATFORM

The transition from gigabyte to terabyte-class data warehouses is occurring now, and it's changing both the way business users leverage information and the way IT manages it. As the volumes and complexity of data continue to grow, the concept of “shrink wrapped” becomes ever more elusive, but alluring. One could argue that the popularity of MS Access isn't due to the product's utility, but its ease to set up.

The issue is that companies require resources to do the acquisition, design, integration, installation, and deployment. It all takes time. The benefit of a data warehouse appliance is that it is up and running quickly with limited configuration, setup, or administration. Negligible setup, negligible cost. And that's a captivating message for both business and IT.

*Solve your data integration, business analytics,
and data warehousing problems
with Baseline Consulting.*

**Call 800-747-3709 or
visit www.baseline-consulting.com**



Thank you for reading this complimentary paper provided by DATAlegro.
For more information on this topic or DATAlegro products, please contact us at
(949) 330-7690, or visit www.datallegro.com.

Baseline Consulting Group, Inc.
15030 Ventura Boulevard Suite 19-707
Sherman Oaks, CA 91403

818.906.7638 Tel
818.907.6012 Fax
www.baseline-consulting.com