

IBM's flexible, preconfigured, integrated combinations of hardware and software challenge the conventional wisdom on data warehousing.

InfoSphere Warehouse

Breaking Barriers

I've been on a mission for the last three years to show small, midsize, and large businesses how data warehousing and business intelligence (BI) can drive bottom-line profits and make them more competitive—and that they *can* afford it.

I've encountered many obstacles, but IBM's significant investments in the last two years are making it easier for companies to derive extreme value from data warehouse solutions. The output of IBM's recent efforts, the IBM InfoSphere Warehouse portfolio, is beginning to break down the perceived and real data warehousing and business intelligence barriers. The products in this portfolio allow users to integrate and manage a centralized data repository for many sources, including Informix, DB2, mainframe, and a score of other non-IBM database systems.

In this article, I'll delve into InfoSphere Balanced Warehouse and InfoSphere Warehouse, which enable companies to centrally design, manage, and deploy data warehouses through a single Web-based console.

There are, of course, many aspects to the BI world, including data integration, data cleansing, reporting, OLAP, master data management, and more; but to get there, you first need an infrastructure to store and integrate the data.

Data Warehouse Barriers

Every day, I attempt to dispel these data warehouse concerns:

- It requires scores of consultants and millions of dollars.
- It's only viable for massive corporations.
- Small IT teams shouldn't consider it.

These perceptions prevent many companies from pursuing true data warehousing. But using consolidated solutions from a single vendor (or a limited number of vendors) can help mitigate all these factors.

One misconception arises from the term "data warehouse" itself, which many associate with large enterprises. I've seen companies use that term to describe a one-table database. Other companies call a multi-terabyte repository a data mart. You can call a historical database any of the buzzwords (data mart, data warehouse, silo, reporting database), but they all mean the same thing: a solution that helps companies understand their profitability, analyze historical buying patterns, cross-sell products to their existing customers, understand their competition, or create simple management reports and dashboards.

Flexible Packaging and Licensing Options

My data warehouse customers are looking for affordability and solutions in which the components are well integrated, relatively easy to manage, and comprehensive. IBM InfoSphere Warehouse (which evolved from DB2 Data Warehouse Edition) provides a number of options, from software only to a complete bundled solution that includes hardware, operating system, software, and (optionally) services delivered by IBM or a certified business part-

by Ron Flannery

ner. Both options are completely scalable, providing great flexibility for a range of customers, from those with less intensive requirements (such as smaller companies, information silos and departmental data marts) all the way up to massive enterprise warehouses with intensive usage.

InfoSphere Warehouse software is available in a number of versions (see Table 1). With a software-only approach, customers are responsible for using in-house servers or for purchasing new servers and for obtaining any in-house or external consultants to implement the solutions. The software can be installed on new or in-house servers that meet InfoSphere Warehouse software's operating system requirements.

IBM InfoSphere Balanced Warehouse, on the other hand, includes a complete, pre-certified and pre-installed solution made up of operating system, hardware, InfoSphere Warehouse software, and services. This option may be advantageous for customers with time constraints or minimal staff, allowing them to get online quickly with a solution built from components that are guaranteed to work together.

InfoSphere Warehouse licensing is available in a variety of options that offer customers the flexibility to start with a limited amount of memory and CPUs, then upgrade as their data warehouse requirements grow. And grow they will. I've observed that many customers aren't aware of the value a data warehouse can return until they actually start using it. Once the initial value is realized, users and management typically clamor for more.

The features included in even the Starter Edition provide tremendous value. Some of these features (CubeViews and the Database Partitioning Feature, for example) were previously sold as separate DB2 add-on products.

The software licensing is available in the following editions:

- **Starter.** For small and midsize businesses, typically first-time data warehouse customers, and 10 to 60 users
- **Intermediate.** For small and midsize businesses and departmental customers with larger data warehouses that scale from 50 to 200 users
- **Base.** For departmental data marts, basic reporting data warehouses and business warehouses of unlimited size
- **Enterprise.** For warehousing solutions requiring advanced business insight and optimization through analytics.

Most warehouse solutions—at least in their initial implementations—can be sufficiently served by the Starter, Intermediate, and Base editions. Table 1 shows the specific edition features.

Single Management Console

Much of InfoSphere Warehouse's value proposition derives from the single management console, Design Studio. This Eclipse-based console can be accessed by a Web browser and provides complete management of the warehouse, including design, development, project and asset management,

job control, and data diagrams (see Figure 1, page 38).

The upper-left pane provides navigation through the various stages of the warehouse. Other items include schema management (lower left), load job sequence and control (middle), data workloads (right), and details on the current asset (lower right).

Administration Tools

InfoSphere Warehouse includes a number of tools to administer and deploy data warehouse assets. For example, Database Explorer provides connection management. Control Center (part of standard DB2 installations) provides a powerful database administration tool. (Although Design Studio includes some administrative functions, Control Center is more comprehensive.) Configuration Assistant helps perform various connection-related tasks. In addition, the WebSphere-based administration console enables administration of OLAP, data mining, SQL Warehouse jobs (to refresh your warehouses), plus process control and troubleshooting.

Modeling and Design Tools

InfoSphere Warehouse Design Studio includes a number of powerful tools, including a library of plug-ins from IBM's Eclipse-based Rational Data Architect (RDA). These plug-ins provide physical data modeling and other functions for Design Studio. The Design Studio also offers reverse engineering of existing schemas, database navigation, life-cycle management, compare-and-sync of data models, physical design, and rules validation.

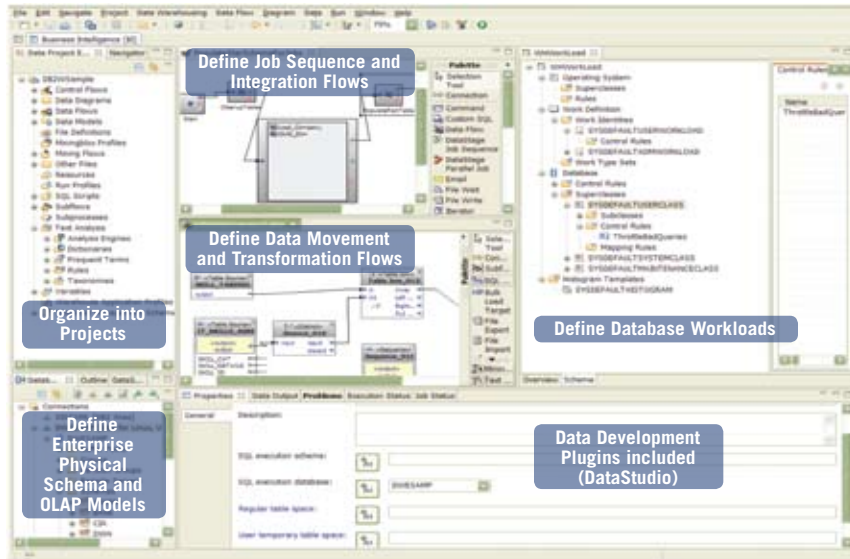
Data Movement and Transformation

One of the major barriers to creating a data warehouse is the ability to actually *populate* it. Many companies use manu-

TABLE 1. InfoSphere Warehouse software licensing options.

	Starter Edition	Intermediate Edition	Base Edition	Enterprise Edition
Administration and control tools	✓	✓	✓	✓
Modeling and design tools	✓	✓	✓	✓
Data movement and transformation	✓	✓	✓	✓
Database partitioning	✓	✓	✓	✓
Cubing Services	✓	✓	✓	✓
Workload management				✓
Deep compression				✓
Data mining and visualization				✓
Unstructured analytics				✓
In-line analytics				✓

FIGURE 1. The Eclipse-based Design Studio provides a single place to manage an entire warehouse project and all related assets via a Web browser.



al, time-consuming, and tedious processes to populate the warehouse (including FTP and manual queries). SQL Warehouse Tool provides a lightweight ETL tool (like the one in IBM WebSphere DataStage) without many of the bells and whistles. SQL Warehouse Tool is actually considered an ELT (extract, load, and transform) package; transformations and cleansing are done via other processes after the data is loaded into the warehouse.

SQL Warehouse Tool provides the ability to set up jobs to load data from a variety of data sources including DB2 (Linux, Unix, Windows, System z, and System i), Informix, Oracle, SQL Server, flat files, ODBC, and a number of other sources. Jobs can be set up to run (typically) nightly, repopulating or creating new tables in the warehouse. Because the SQL Warehouse Tool's job flows are WebSphere-based, it provides access to powerful logging, debugging, job control, and other features, as well as the ability to notify staff of the results of each job. The result is a data warehouse populated and ready to use with minimal IT effort. (The center part of the screenshot in Figure 1 shows SQL Warehouse Tool jobs being managed.)

Database Partitioning

DB2's Database Partitioning Feature (DPF) provides a way to control warehouse storage and performance. DPF, previously an add-on to DB2 Enterprise Edition, allows you to define a way to segment data across storage on the same or different servers. This capability can greatly improve query performance by allowing queries to run in parallel across multiple partitions. In addition, DPF can be used to manage the archiving and storage of historical data typically used in a data warehouse.

Cubing Services

Cubing Services, a relatively new add-on to InfoSphere Warehouse, allows you to create multidimensional cubes that comply with MDX (the de facto standard for multidimensional query) within memory on the same database server. Creating the cube on the same server saves network bandwidth and eliminates the need for a second server. In many cases, in-memory cubes result in greatly improved performance. IBM Alphablox, Cognos, or any MDX-compliant product can be used to report on the data in the cube; alternatively, these cubes can be accessed direct-

ly via the MDX query language without going through a front-end tool.

There are advantages to using the built-in cubing functions of many OLAP products like Cognos; however, Cubing Services may provide value for many environments. In addition, Cubing Services now includes CubeViews, which uses materialized query tables to design (via advisors if needed), manage, deploy, and update data for warehouse summary tables. Figure 2 shows Cubing Services serving data to a variety of front-end tools.

IBM InfoSphere Balanced Warehouse

IBM InfoSphere Balanced Warehouse is a set of preconfigured solutions, certified by IBM in a combination of hardware, software, operating system, and services. Each solution contains *all of* the core InfoSphere Warehouse features I've described; only the D- and E-class contain the Enterprise-Edition features (workload management, deep compression, data mining and virtualization, unstructured analytics, and in-line analytics).

The Balanced Warehouse extends the value of the software by providing pre-installed, ready to use hardware and software solutions, certified to work together, at a fixed price. The offering removes more barriers associated with implementing a new data warehouse solution; customers won't have to try to find hardware that can scale to the needs of the warehouse while being compatible with the InfoSphere Warehouse software. In addition, these systems are preconfigured by IBM or a business partner, making the warehouse ready to run out-of-the-box.

Of course, preconfigured options aren't the best match for every company. Many customers want to control their hardware and operating system choices or integrate them into an existing environment. In such cases, customers can purchase the proper combination of InfoSphere Warehouse software.

Operating systems vary depending on the option chosen. For example, cost-effective C-1000 and C-3000 run only on Linux and Windows. These solutions also reduce hardware and licensing costs for their Intel-based systems. (For example, consider the cost of deploying a Microsoft solution with more hardware and licensing costs.) Even if your company isn't a Linux shop, for example, the C-1000 or C-3000 are still viable options because of their pre-installed and GUI-based administration features.

Hardware in C-, D-, and E-classes is all IBM. Intel-based solutions use IBM System x (formerly known as xSeries) or BladeCenter components; other solutions use IBM System p (formerly pSeries) using the POWER6 chipset.

License Classes

The InfoSphere Balanced Warehouse solutions come in three classes:

- C-class solutions are sold by certified IBM business partners and can be configured with the features shown in Table 1 (except those in Enterprise Edition); the only difference is related to the CPU and memory limitations. These solutions are typically provided as completely installed and ready to implement.

- D-class solutions provide the features in Enterprise Edition, including data mining (in-line and text), higher storage limitations, and more reporting tools. These are typically targeted at

BI includes data integration, data cleansing, reporting, OLAP, and master data management; but to get there, you need an infrastructure to store and integrate the data.

midsize to large customers, advanced departmental data marts, and small to midsize warehouses.

- E-class solutions are designed to provide unlimited growth for large companies with warehouses that grow dynamically and are key to a company's operations.

Within each class there are configurable options that depend on warehouse size and usage.

The C-class solutions are a great starting point for companies that aren't sure how far their warehousing commitment will go; those solutions provide all the basic tools needed to create and manage a warehouse, and can be upgraded to D- and E-class as warehouse requirements grow.

Simplified Data Warehouses for All Sizes

IBM has made major financial and development efforts to address the pain points that prevent companies from implementing BI

and data warehouse solutions. InfoSphere Warehouse and InfoSphere Balanced Warehouse provide a suite of integrated products, managed from one place, that can grow as your company grows.

Investing in one of the solutions I described in this article is a great starting point, but it doesn't give you a complete BI solution. Data warehousing and BI require buy-in from business users, (sometimes) intensive requirements gathering, warehouse design, subject matter expertise, reporting tools, and more. But why not start with a solution that gets you that much closer? ■

Ron Flannery

is president and founder of One Point Solutions, an IBM Business Partner focused entirely on selling and servicing Informix, DB2 UDB, and IBM BI/data warehouse solutions. A former member of the International Informix Users Group board of directors, Flannery is the lead author of the popular *Informix Handbook* (Prentice Hall, 2000).

RESOURCES

InfoSphere Warehouse

ibm.com/software/data/infosphere/warehouse

IBM InfoSphere Balanced Warehouse

ibm.com/software/data/infosphere/balanced-warehouse

InfoSphere Warehouse Enterprise Edition

ibm.com/software/data/infosphere/warehouse/enterprise.html

FIGURE 2. Front-end tools supported by Cubing Services.

