Planning for affordable growth in the data warehouse with IBM Balanced Warehouse D5100

Contents

Business intelligence (BI) is no longer a luxury—it has become fundamental to the success and growth of businesses worldwide. The volume of data an average business collects and stores doubles every year, which means that some of the world's leading organizations will have data warehouses measured in petabytes (thousands of terabytes) within the next several years.

To meet the challenges associated with using rapidly expanding data stores to generate BI, companies need flexible data warehousing solutions that can grow with the organization. They need the flexibility to add data while managing costs without ripping-and-replacing the data warehouse. They also need the ability to consolidate datamarts and silos of information or migrate from a current datamart or warehouse that does not scale or meet the needs of the business. They must be able to cost-effectively store and manage growing amounts of data without wasting valuable time configuring and installing disparate hardware and software components. In addition, the total solution should be thoroughly tested and preconfigured based on best practices—not a basic reference configuration or piecemeal solution.

IBM is well positioned to help companies meet these challenges. For decades, IBM has been enabling companies to make better-informed decisions by helping them generate, manage and extend their enterprise data warehouses.

Through IBM® Balanced Warehouse™ offerings, IBM can help organizations plan for affordable growth in their data warehouses.

IBM Balanced Warehouse provides a comprehensive, optimized platform for data warehousing

IBM Balanced Warehouse is a real-time, ready-to-go data warehousing solution composed of pretested, scalable and fully integrated system components including IBM DB2® Warehouse software, servers, storage and support. The components are precision balanced for optimal performance.

The comprehensive IBM data warehousing solution portfolio offers a variety of options built from readily available and stable hardware and software components such as DB2 Warehouse, which empowers companies to control their information in real time. By integrating data warehousing with business analytics, the IBM platform allows companies to define their central business concepts and the data required to support those concepts from an enterprise-wide perspective. In addition, organizations can pull data from source systems that traditional BI and data warehousing solutions cannot access. This capability can make it easier for IT organizations to support business requirements for actionable information—not just raw data, but data with intelligence behind it to help people take action and make decisions.

IBM Balanced Warehouse solutions are rigorously tested and tuned for performance and reliability. Before a Balanced Warehouse is released, IBM laboratories configure and stress test the systems for maximum reliability and security as well as fine-tune them to meet the highest levels of query performance.

IBM Balanced Warehouse platform configurations are based on IBM best practices from successful client implementations. The platform is not a reference architecture; instead, it is supported as a total solution accessed through a single contact point. It is also not an appliance—in a typical appliance model, once the data outgrows the platform or IT decision makers decide that the company should grow in a different direction (for example, by adding higher levels of concurrency), the company must start over with its data warehousing infrastructure—which costs valuable time and money.

The IBM Balanced Warehouse model D5100 is designed for growing midsize and large organizations that manage raw data amounts in the 1TB-and-up range. The D5100 is particularly well suited for data warehouses that will grow significantly over time.

The latest addition to the IBM Balanced Warehouse D-Class family, the D5100 integrates optional high availability features that enable servicing mission-critical situations at an affordable price. With well-defined levels of modular

scalability, the D5100 provides a cost-effective warehousing solution that has the room and ability to grow as business requirements expand. It is aimed at two types of data warehouse environments: businesses that are consolidating datamarts and making an investment in data warehousing on Linux®, and growing business units in large enterprises.

IBM Balanced Warehouse provides multiple entry points for warehousing and a clear growth path

IBM Balanced Warehouse offerings range from instant reporting solutions handling gigabytes of raw data to multiterabyte, real-time, mission-critical enterprise data warehouses. C-Class platforms provide small and midsize businesses with a simplified, complete data warehousing solution designed for rapid deployment. The D-Class, including the D5100, is designed to support advanced departmental datamarts and smaller data warehouses for growing midsize and large organizations. The E-Class provides outstanding performance for large-scale enterprise data warehouses.

In addition to the assortment of IBM Balanced Warehouse offerings, IBM offers new industry model solutions that are built directly into the platform and integrated with third-party application components. These models can help companies address key business initiatives in the banking, retail, telecommunications and government sectors.

Clear growth paths are a major benefit of the IBM Balanced Warehouse platform. The fully integrated solution supports efficient growth by allowing companies to start with what they require right now and plug in additional capacity as needed: Organizations can scale to multiple terabytes of information without compromising price/performance. The platform also provides an efficient prescriptive, fully validated and tested, pre-integrated design that enables companies to better manage costs with a more predictable total cost of ownership (TCO).

IBM Balanced Warehouse D5100: Designed for growth

The transparent, modular architecture of the D5100 allows companies to define when they grow and how they grow, which enables simple and logical growth based on usage demands. Built on midrange IBM Systems hardware, this warehouse was specifically designed to incorporate automated failovers to help ensure that critical information remains available when the organization needs it most and provide affordable integrated system availability. When problems do arise, intelligent call-home features provide proactive, automated system servicing.

The platform also incorporates the latest AMD dual-core Opteron processors, which help lower power consumption without compromising performance. In addition, enterprise-type tools and features are available on midrange standard hardware, providing enhanced processing power at less cost and lower energy usage.

Five flexible modules enable companies to determine their own growth paths
The D5100 features prebalanced building blocks and a transparent modular architecture that give companies the freedom and flexibility to decide their own warehousing growth path. Based on a strict building block methodology that uses Balanced Configuration Units (BCUs), the D5100 encapsulates the basic BCU and adds layers of system functionality to create modules. The modules are specifically designed to represent a simpler, more transparent architecture—uncommon in today's warehousing appliance market.

Foundation modules serve as the keystone for all IBM Balanced Warehouse offerings, along with a minimum of one data module. The foundation module helps ensure that all D5100 systems start with proper management and administration of the Balanced Warehouse.
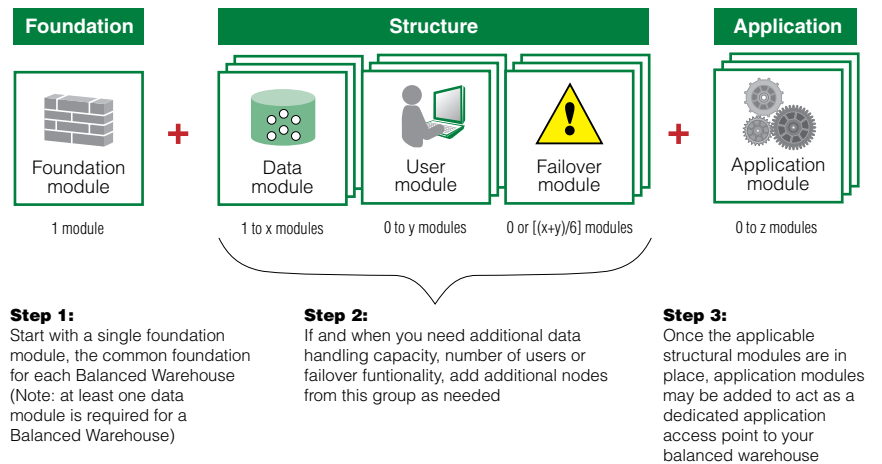
Data modules allow companies to handle more raw data by enabling the Balanced Warehouse to scale out in size, giving its users more capacity for growing stores of data. Each module provides roughly 1TB in raw data space. If an extra boost is required for expanding the 1TB-per-module limit or if faster access is required, an additional high-capacity option is available.

User modules are designed to allow businesses to grow the number of active users in the warehouse. In situations where greater numbers of users are required—both in number and with higher levels of concurrent access—the user module can provide more access. As more users in an organization start leveraging the integrated functionality of the warehouse, companies can add additional user modules as needed to facilitate user and workload growth.

Failover modules can be used to provide high availability. The failover module helps ensure that an organization's information is available regardless of potential server failures. If a server goes down, a Balanced Warehouse can perform an automated failover that mobilizes a failover module to replace the server in question. Recommended coverage, based on the total number of current data modules (x) and user modules (y), is a total of (x+y)/6 failover modules.

Application modules are designed to act as portals for applications running on the warehouse. Once the foundation and structure of the Balanced Warehouse is in place, application modules may be added to act as dedicated access points for reporting and analytic applications. These modules offer higher limits of performance and concurrency for applications and users.

Figure 1: IBM Balanced Warehouse modules create a transparent architecture

Foundation	Structure			Application
Foundation module	Data module	User module	Failover module	Application module
1 module	1 to x modules	0 to y modules	0 or [(x+y)/6] modules	0 to z modules

Step 1:
Start with a single foundation module, the common foundation for each Balanced Warehouse (Note: at least one data module is required for a Balanced Warehouse)

Step 2:
If and when you need additional data handling capacity, number of users or failover funtionality, add additional nodes from this group as needed

Step 3:
Once the applicable structural modules are in place, application modules may be added to act as a dedicated application access point to your balanced warehouse

Building a modular data warehouse with the D5100 requires five steps:

1. **Build the foundation.** *Using just a foundation module and a data module, companies can start a data warehouse with less than 1TB of raw data.*
2. **Grow the data.** *Data modules can be added to the foundation to expand capacity and accommodate the growth of data.*
3. **Add user capacity.** *At some point, additional users may strain the current warehouse; user modules can be added as needed to provide support.*
4. **Add failover capability.** *If performance is critical, the use of a failover module allows performance levels to be maintained because the failed module is completely replaced instead of shared. For a proper automated failover, a company would need $(x+y)/6$ failover modules, where x represents the number of data modules and y represents the number of user modules; for example, if there were five data modules and one user module, the company would require $(5+1)/6 = 1$ failover module.*

5. *Add an application access point. If additional support is required from the application end, an optional application module can be added to handle more demanding programs.*

IBM Balanced Warehouse in the real world

At Convergence CT—a global company that links patient data at healthcare institutions with pharmaceutical and biotech companies' data needs for clinical trial planning, recruitment processes and marketplace research—the IBM Balanced Warehouse plays a significant role.

"We've found DB2 Warehouse to be perfect for consolidating information about how cancer and diabetes patients respond to different treatments in healthcare facilities around the world," says Lambert Onuma, CEO at Convergence CT. "The IBM Balanced Warehouse provides us with a simplified, yet flexible and optimized solution for delivering analyses of that data to help save lives."

For more information

To learn more about IBM Balanced Warehouse and the D5100 model, please contact your IBM representative or IBM Business Partner or visit **ibm.com**/bi

IBM. ®