Higher Performance and Lower Cost Solutions with IBM InfoSphere Warehouse V10

IBM Redbooks Solution Guide

Access to timely, accurate information is critical for enterprises that are striving to better serve their customers, beat the competition, and foster innovation. IBM® Information Management provides a comprehensive data warehouse solution (Figure 1) so that organizations can centrally, accurately, and securely analyze and deliver information as part of their operational and strategic business applications.
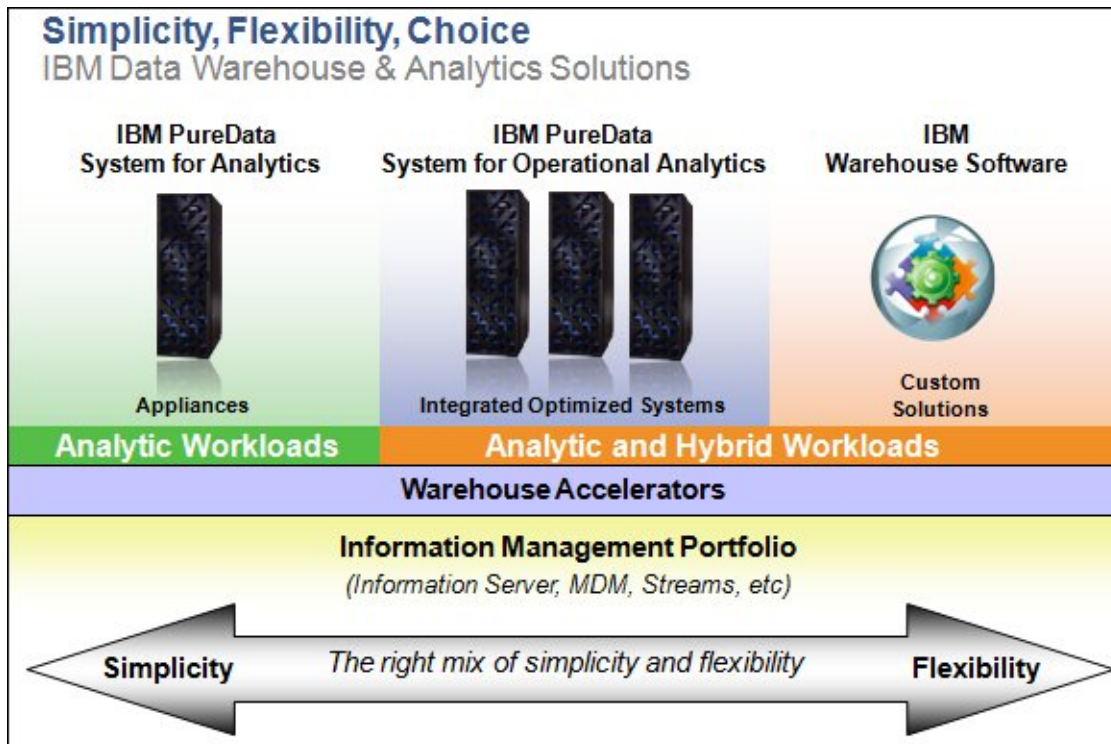

Figure 1. A comprehensive data warehouse solution

IBM InfoSphere® Warehouse V10 provides a powerful range of capabilities that go beyond the capabilities of traditional warehouses. This comprehensive platform integrates the strength of the IBM DB2® database with a dynamic data warehousing infrastructure that can handle traditional business intelligence (BI) workloads and more operational business requirements. In addition, InfoSphere Warehouse Advanced Enterprise Edition delivers an enhanced set of database performance, management, and design tools. These tools assist companies in maintaining and increasing value from their warehouses and by helping to reduce the total cost of maintaining these complex environments.

Did you know?

The volume and variety of digital information (structured and unstructured) are exploding as our planet becomes more instrumented, interconnected, and intelligent. With social media alone, we are talking about terabytes of new data. The key to success is the ability for you to gain insight from that data and leverage it for business opportunities. With IBM InfoSphere Warehouse, you have that ability.

Business value

Through advanced data warehousing technology, IBM helps organizations extract insight from virtually any type of data. IBM helps to deliver the right information at the right time and in the right context so that business leaders can make the right decisions quickly. IBM advanced warehousing solutions integrate data warehousing and business analytics to help define an organization's central business concepts and the data that is required to support those concepts. These solutions allow organizations to capture data changes from various enterprise and source systems that traditional BI and data warehousing solutions were unable to access in the past.

As a result, IT organizations can better support business requirements for actionable information. This information is not just raw data but data that is backed by intelligence that can help people to take action and to make sound business decisions.

InfoSphere Warehouse V10, which is based on DB2 10, includes a new set of advanced capabilities to enable real-time operational analytics that empowers organizations to make active, timely, and informed decisions as business events occur. InfoSphere Warehouse V10 offers the following benefits:

- Faster, accurate decision making and turnaround times

    o   Business intelligence because data is continuously fed into the warehouse
    o   Business intelligence and analytics tools for decision makers and specialized analysts

- Improved cost efficiencies

    o   Advanced storage technology
    o   Advanced recovery solutions that help enable online recovery of lost data

- High performance

    o   Star schema optimization delivery for quicker response times, delivering three times the performance on BI workloads
    o   High availability operational access that is concurrent with analytics

- Increased team productivity

    o   Built-in time travel query that enables faster historical and trend analytical queries
    o   Row and column access controls to support multiple tenant operational warehouses
    o   Basic bi-temporal support that improves developer and database administrator (DBA) productivity

- Access to and analysis of a broad array of information

    o   Unstructured information in call center notes, emails, and blogs
    o   Structured information in databases, spreadsheets, and other data sources

Solution overview

Businesses must address challenges and work to achieve on-demand access to insight. Among these challenges are bottlenecks in capturing and loading operational data that slow the ability for businesses to react in a timely manner. Also, performance challenges result from the additional resources and planning that are involved in handling heavy workloads and complex queries for analytics processing.

In addressing the challenges, businesses can target smaller customer segments and communicate with them about their individual needs and wants, while driving new market opportunities within the current business landscape. They can identify and capitalize on even the smallest trends, attaining competitive advantages that are normally realized only by more flexible and dynamic smaller businesses. They can detect small behavior patterns that can have a significant influence and effect on the business in terms of revenue, expenses, and growth. Most important, businesses can build competitive strategies around data-driven insights and, in the end, generate impressive business results.

InfoSphere Warehouse is powered by the DB2 for Linux, UNIX, and Windows data server. With its massively scalable, shared-nothing architecture, DB2 provides high performance for mixed-workload query processing of relational and basic XML data. Such advanced features as database and table partitioning, compression, multidimensional clustering (MDC), materialized query tables (MQT), and OLAP capabilities make DB2 a powerful engine for operational warehousing (Figure 2).
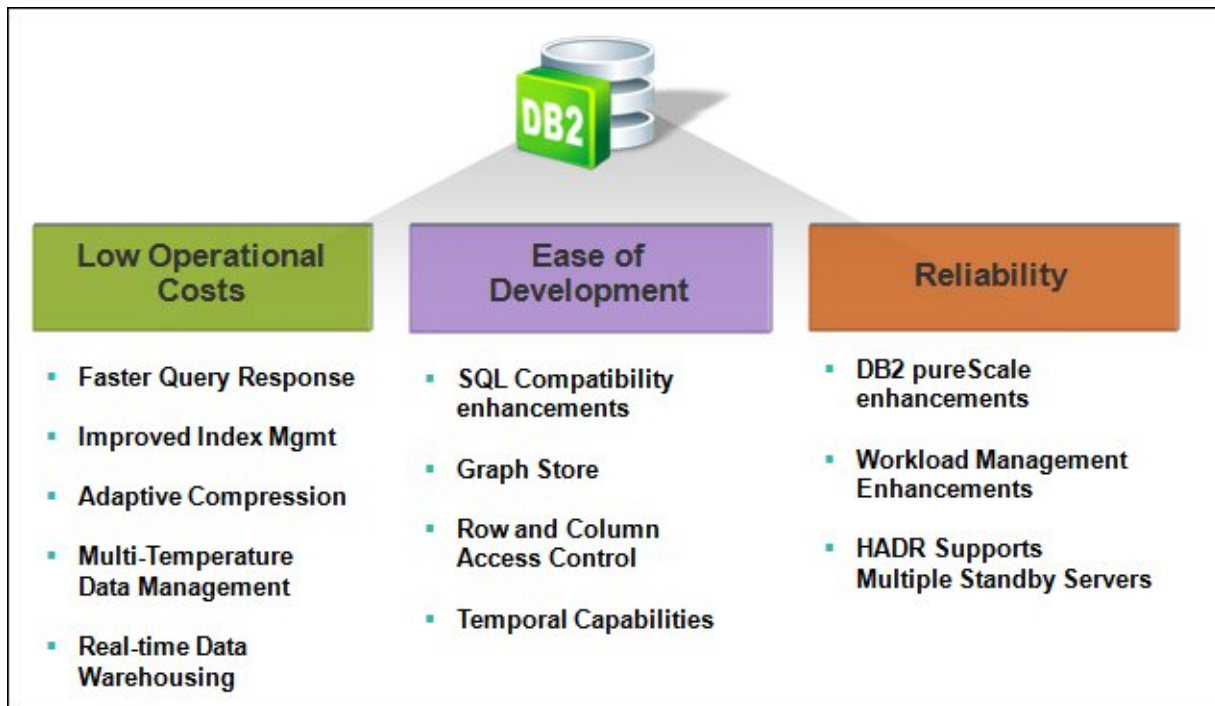

Figure 2. Building on the pillars of DB2

InfoSphere Warehouse provides advanced capabilities for database partitioning, so that IT users have multiple ways to distribute data across servers for large-scale parallelism and linear scalability. The shared-nothing architecture of DB2 helps ensure that performance will not degrade as the warehouse grows. Also, because InfoSphere Warehouse can physically cluster data on multiple dimensions, order data by value range, and limit I/O to relevant data partitions, it helps reduce the work that is needed to resolve many queries.

InfoSphere Warehouse transparently splits the database across multiple partitions and uses the horsepower of multiple servers to satisfy requests for large amounts of information. SQL statements are automatically decomposed into subrequests that are run in parallel across each database partition. Results of the subrequests are joined to provide final results.

IBM InfoSphere Warehouse includes the following rich features and functions:

- *Table partitioning* offers easy roll-in and roll-out of table data, flexible index placement, and efficient query processing. Table partitioning enhances the flexibility of table-level administration by allowing administrative tasks to be performed on individual data partitions. These tasks include detaching and reattaching a data partition, backing up and restoring individual data partitions, and reorganizing individual indexes. Time-consuming maintenance operations can be streamlined by breaking them down into a series of smaller operations. For example, backup operations can work data partition by data partition when the data partitions are placed into separate table spaces.

- By using *Continuous Data Ingest*, you can transparently load data from external sources into InfoSphere Warehouse databases without downtime and perform real-time business analysis and decision making.

- *Time Travel Query* is integrated into DB2 10 and InfoSphere Warehouse for easier and faster time-based (historical trend-based) analytics applications. The addition of zigzag join helps to significantly reduce the time for complex multidimensional business queries. Enhanced query joins and optimizer enhancements help to increase query performance of other analytic queries and to reduce the need for more indexes.

- *Adaptive compression* can also help reduce storage costs and improve performance, especially for large I/O-bound warehouse applications and query workloads. Data row compression contributes to storage space savings and helps to reduce disk access time. At the same time, the stored pages are compressed, which further enhances the compression on disk. Also, because data is compressed, more rows can be cached in the buffer pool of the database to improve query response time, and DBAs no longer need to perform REORG operations as frequently.

- New *Row-and-Column Access Control* provides easy and flexible rule and role definitions to manage and control data accesses that help to enhance security and simplify application development. These security features provide a robust and flexible set of rules and access controls to manage and help secure data accesses that help reduce security risks.

- *Multidimensional clustering* provides a flexible method to continuously and automatically cluster table data in multiple dimensions. This type of clustering reduces the amount of I/O that is required. In addition, it helps reduce the need for database maintenance activities such as reorganization.

- *InfoSphere Warehouse workload management* capabilities enable real-time delivery of business insights without compromising performance. With traditional servers, the strain of mixed workloads can inhibit the delivery of information to a broad set of users and applications. With the advanced workload management that is provided by InfoSphere Warehouse, DBAs can establish and enforce service levels for users. They can prioritize queries from different users and applications and then control the number of underlying resources that are dedicated to those processes.

- *InfoSphere Replication Server* technology is included in all editions of InfoSphere Warehouse. Organizations that are looking to provide active/active availability can use bidirectional Q replication between a pair of source and target DB2 for Linux, UNIX, and Windows data servers.

- *Embedded analytics capabilities* deliver a set of sophisticated, yet easy-to-use tools within the data warehouse. These tools provide valuable business intelligence to many users. The Cubing Services for OLAP feature enables multidimensional data analysis without extracting data from the warehouse. InfoSphere Warehouse includes basic support for the Microsoft PivotTable Service, enabling ad hoc analyses or delivery of standard spreadsheet reporting, all while working within the Microsoft Excel application. In addition, Cubing Services cubes are first-class data providers to the IBM Cognos® platform. The entire suite of Cognos clients and applications can use these powerful warehouse-based data cubes. InfoSphere Warehouse provides embedded data mining, modeling, and scoring capabilities. With these capabilities, business users can work with current data and deliver analytics in real time, helping them to quickly discover revenue opportunities.

- By using *IBM Cognos Business Intelligence*, business users can evaluate a rich set of BI capabilities without incurring up-front costs. Business users can easily access data from their data warehouse. With help from reporting and analysis features, they can deliver relevant information how, when, and where it is needed. By using the web-based user interface, enterprise-class service-oriented architecture (SOA) foundation, and the ability to access any data sources, business users can easily develop and deploy reports on the data assets within the warehouse. Combined with the Warehouse Packs (available in the Advanced Editions), Cognos Business Intelligence provides a quick way to deploy warehouse reporting and to gain rapid value and insights from data.

- *IBM InfoSphere Optim Database Administrator* helps organizations to manage databases and database changes without disruption, streamlining change-in-place and database migration scenarios. Built-in analysis and migration features help prevent application outages by ensuring that all related objects are migrated. They also support outstanding performance by ensuring that indexes are updated and facilitate availability by ensuring that privileges are migrated. InfoSphere Warehouse also includes *InfoSphere Optim Performance Manager*, which provides performance monitoring and management that can be used immediately to help improve quality of service and prevent impacts to business operations. Its intuitive, web-based user interface provides use-anywhere monitoring, alerting, and diagnosis of potential performance bottlenecks.

InfoSphere Warehouse provides a set of tools that help simplify data warehouse and analytics development and deployment. With these interfaces, users can design the warehouse and populate data structures. They can also perform analytics and manage data mining and multidimensional cubing through common interfaces.

- Design Studio provides a graphical user interface (GUI) so that architects can design, model, reverse engineer, and validate physical database schemas. Design Studio is based on IBM InfoSphere Data Architect software and can import and export models from various sources, including CA ERwin. By using the SQL Warehousing tool, DBAs can prepare and populate the data warehouse structures that are required for data mining, multidimensional analytics, and embedded analytics. Data flows, control flows, and transformations can be built by using Design Studio and deployed within the warehouse.

- IBM InfoSphere Optim™ Development Studio software helps increase development efficiency for Java data access and facilitates cross-system development and migration. It supports development for DB2, Oracle database, and IBM Informix® software. Its SQL outline feature facilitates developer and DBA collaboration by quickly isolating all the SQL for review and enables impact analysis by correlating SQL with source code, database objects, and ALTER requests.

Solution architecture

InfoSphere Warehouse Advanced Edition brings together all the components that are required for a successful, cost-efficient data warehouse solution. The components range from the development tools that are needed to create your extract, transform, and load (ETL) operations, OLAP, and data mining, to the BI tools that are used in understanding your market. InfoSphere Warehouse Advanced Edition also offers the tools that are needed to manage your backup strategy, drive consistency across your business, and bring out the best performance of your data warehouse and the applications that are connected to it.

At the center of the InfoSphere Warehouse lies the DB2 relational database engine, which provides a repository for the user data and the infrastructure to support the many functional operations that are performed on the data. Together with the InfoSphere Warehouse application that is hosted in an IBM WebSphere® Application Server, these elements combine to form the runtime component of an InfoSphere Warehouse solution. Several client products, Data Studio, Design Studio, and web browser provide the development tools and administration components that are required to support these runtime elements. Figure 3 shows the complete functional component architecture of InfoSphere Warehouse.
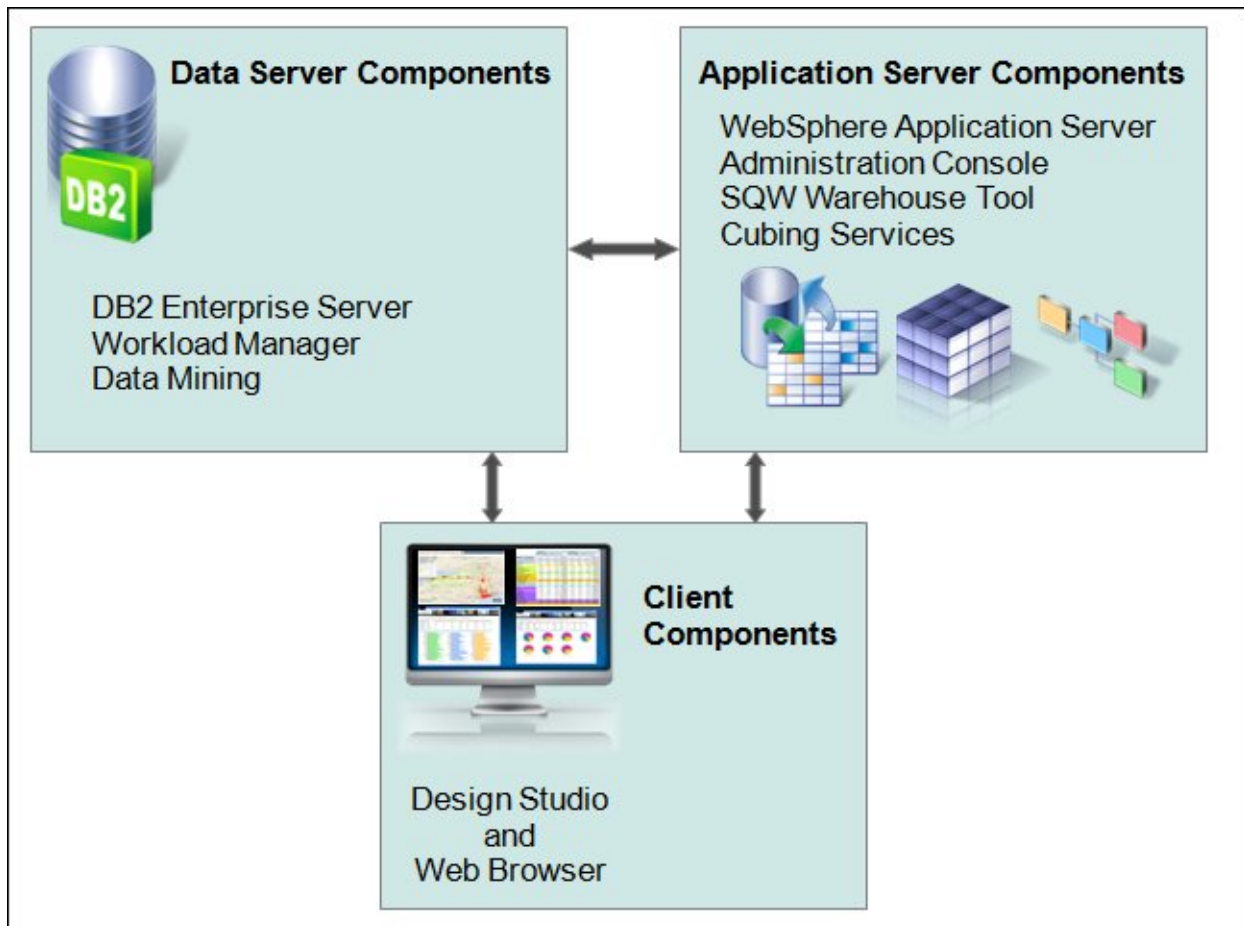

Figure 3. Functional component architecture of InfoSphere Warehouse

The instance of the DB2 relational database at the core of the data warehouse can be configured as a single or multiple partitioned database that is installed on a single hardware server or numerous hardware servers. This flexibility of DB2 results in an unlimited power that is available to the main repository of your warehouse data, which is often called the *execution database*.

In addition to the main data repository and execution database, the same DB2 instance hosts two more databases. These much smaller databases contain the metadata that is required by the InfoSphere Warehouse runtime and Cognos BI server applications.

The InfoSphere data warehouse application server component consists of enterprise Java applications that perform several important functions, including the following functions, within the complete solution:

- The administration console for the InfoSphere Warehouse solution
- The ability to store, run, and manage the ETL processes
- Store and manage the cubing services
- Store and manage data mining services

The InfoSphere Warehouse administration console has a web-based interface that allows any browser to be used in the configuration and management of the functional elements of the runtime environment. Thus, a single browser can be used to handle all production, test, and development environments.

SQL Warehousing applications perform the ETL operations on the data that is in the execution database. These operations consist of Control Flows and Data Flows that were created by using Design Studio, which is the SQL Warehousing development tool. From within Design Studio, SQL Warehousing data and control flows can also be tested and debugged against real databases and then grouped into a warehouse SQL Warehousing application. These SQL Warehousing applications are then deployed through the administration console into the SQL Warehousing runtime element.

An OLAP cubing services server, which for simplicity is called a *cube server*, is the runtime element of the cubing process. This cube server is an independent Java process that hosts the various cubes, receives incoming connection and query requests, processes the requests, constructs the result sets, and returns them to the calling application. This Java process runs independently of the installed WebSphere Application Server, but is required to reside on the same physical server, so that it can be managed by the administration console application.

A cube can be implemented within a cubing server by using the Design Studio development tool. When a cube model is successfully implemented, it can then be deployed to the InfoSphere Warehouse server.

Usage scenarios

The components that make up a core InfoSphere Warehouse implementation can be divided into four Installation categories:

- The *Data Server component* covers the main DB2 platform, which is supported on IBM AIX®, HP-UX, Solaris, various Linux implementations, and Windows.

- The *Application Server component* covers the IBM WebSphere Application Server, which is a part of the warehouse product set.

- The *Clients component* covers all of the command line and GUI-based platforms that might normally be installed on a user's personnel computer or notebook.

- The *Documentation component* covers the online and PDF versions of the product set documentation.

The InfoSphere Warehouse components that cover these categories can be installed onto a hardware platform in a range of topologies. The InfoSphere Warehouse architecture has the following common topologies (Figure 4):

●   A *one-tier architecture* is often used in development, test, and education environments. All the major components, including the clients, are installed on a single hardware platform.

●   A *two-tier architecture* is also primarily used in development and test environments. However, with an appropriate server and storage, this topology can be used for a smaller warehouse implementation.

●   In a *three-tier architecture*, the client components, the DB2 components, and the WebSphere Application Server components are installed on a separate hardware system. You use this topology on a production system.
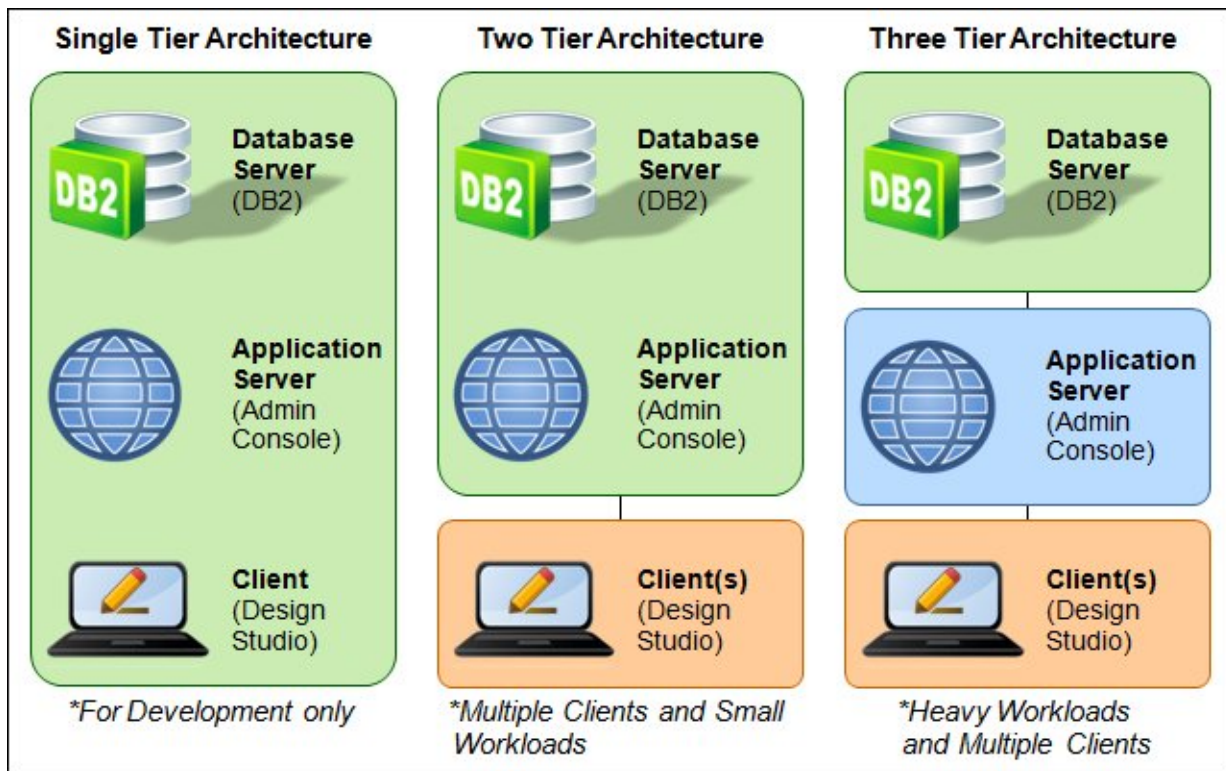

Figure 4. Three common topologies of the InfoSphere Warehouse architecture

Figure 5 demonstrates a three-tier implementation on multiple physical hardware platforms, where the database server has an administration database node and multiple data nodes.
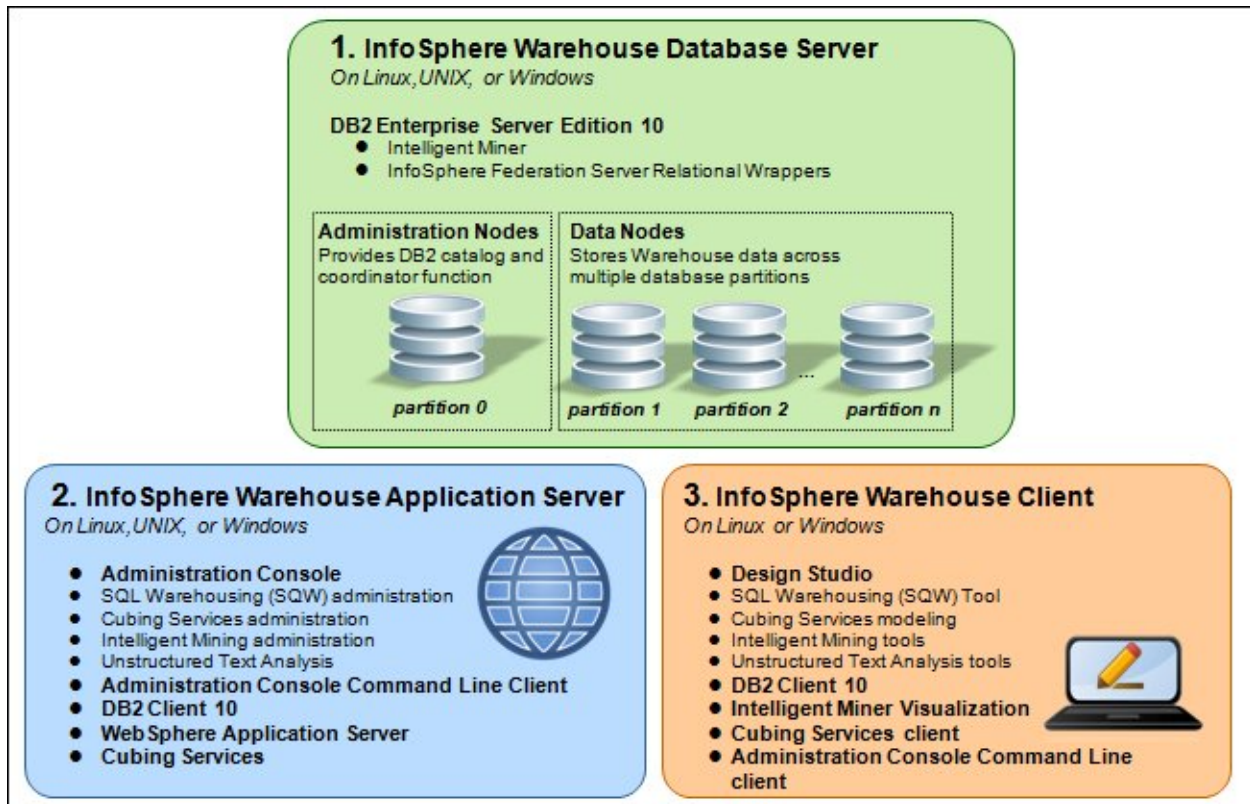

Figure 5. Three-tier solution on multiple tier physical servers

Supported platforms

InfoSphere Warehouse is a suite of products that combine the strength of DB2 Enterprise Edition with a data warehousing infrastructure from IBM. InfoSphere Warehouse has a component-based architecture that consists of a data server component group, an application server component group, and a client component group. In a typical production environment, you install each of these component groups on different computers to create a complete warehousing solution.

The system requirements for InfoSphere Warehouse can be occasionally updated. To obtain the most current information, see the InfoSphere Warehouse product page at:
http://www.ibm.com/software/data/infosphere/warehouse/sysreqs.html

Ordering information

InfoSphere Warehouse offerings range from editions for enterprise-class data warehousing to speciality editions:
- InfoSphere Warehouse V10.1 Advanced Enterprise Edition
- InfoSphere Warehouse V10.1 Enterprise Edition
- InfoSphere Warehouse V10.1 Advanced Departmental Edition
- InfoSphere Warehouse V10.1 Departmental Edition
- InfoSphere Warehouse V10.1 Developer Edition

For ordering information, contact your IBM representative or an IBM Business Partner. See also the IBM InfoSphere Warehouse V10.1 Sales Manual at:
http://ibm.co/XkAEgz

Related information

For more information, see the following documents:

- *Solving Operational Business Intelligence with InfoSphere Warehouse Advanced Edition*, SG24-8031
  http://www.redbooks.ibm.com/abstracts/sg248031.html

- *InfoSphere Warehouse: A Robust Infrastructure for Business Intelligence*, SG24-7813
  http://www.redbooks.ibm.com/abstracts/sg247813.html

- IBM InfoSphere Warehouse Information Center
  http://bit.ly/SC2IWU

- Workload Management (WLM) Tutorial
  http://ibm.co/RimG9Z

- Best Practices Workload Management
  http://www.ibm.com/developerworks/data/bestpractices/workloadmanagement

- IBM InfoSphere Warehouse V10.1 Sales Manual
  http://ibm.co/XkAEgz

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service. IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you. This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you. Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurement may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

This document was created or updated on November 8, 2012.

Send us your comments in one of the following ways:
- Use the online **Contact us** review form found at:
  `ibm.com`/redbooks
- Send your comments in an e-mail to:
  redbook@us.ibm.com
- Mail your comments to:
  IBM Corporation, International Technical Support Organization
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400 U.S.A.

This document is available online at `http://www.ibm.com/redbooks/abstracts/tips0932.html` .

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at www.ibm.com/legal/copytrade.shtml

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

AIX®
Cognos®
DB2®
IBM®
Informix®
InfoSphere®
Optim™
Redbooks (logo)®
WebSphere®

The following terms are trademarks of other companies:

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Java, and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Other company, product, or service names may be trademarks or service marks of others.