SYBASE®

Overview Guide

Sybase Data Integration Suite

1.2

DOCUMENT ID: DC00583-01-0120-01

LAST REVISED: November 2007

Sybase, Inc., One Sybase Drive, Dublin, CA 94568.

Contents

About This Book

Audience

This book is intended for customers who have purchased Sybase® Data Integration (DI) Suite components:

- Search
- Data Federation
- Replication
- Real-Time Events
- ETL

How to use this book

This book consists of:

- Chapter 1, “Introducing Sybase Data Integration Suite,” describes the features and components of DI Suite.
- Chapter 2, “Data Integration Techniques,” provides various sample solutions using DI Suite components.

Related documents

This section describes the DI Suite documentation set, which you can find on the Getting Started CD and the various SyBooks™ CDs.

The DI Suite Getting Started CD includes:

- *Sybase Data Integration Suite 1.2 Release Bulletin* for your platform – contains last-minute information that was too late to be included in the books.
- *Sybase Data Integration Suite 1.2 Installation Guide* for your platform – describes installation procedures for the various components of the DI Suite.
- *Sybase Software Asset Management and User’s Guide* – describes asset management configuration concepts and tasks.
- Release bulletins, installation guides, and administration guides for these Sybase products, which are included with DI Suite:
  - Data Federation 1.2
  - EAServer 6.0.2

- Enterprise Connect™ Data Access 15.0
- Real-Time Data Services 4.5
- Replication Agent™ 15.0
- Replication Server® 15.0.1
- Sybase Search 3.5
- Sybase ETL 4.2

DI Suite includes a separate SyBooks™ CD for each DI Suite component. In addition to the documents listed below, each CD also includes the *DI Suite Overview Guide*, and the *DI Suite New Features Guide*.

- SyBooks CD for Data Federation includes:
  - Product manuals for Data Federation 1.2.
- SyBooks CD for Replication includes:
  - Product manuals for these Sybase products, which are included with the Replication component of DI Suite:
    - Replication Server 15.0.1
    - Replication Agent 15.0
    - Enterprise Connect Data Access 15.0
- SyBooks CD for Real-Time Events includes:
  - Product manuals for these Sybase products, which are included with the Real-Time Events component of DI Suite:
    - Replication Server 15.0.1
    - Replication Agent 15.0
    - Real-Time Data Services 4.5
    - EAServer 6.0.2
    - SQL Anywhere Studio 10.0.1
    - Business Activity Monitoring 6.2 Administration Guide
  - *Data Integration Common Services* online topics
- SyBooks CD for Search includes:
  - Product manuals for Sybase Search 3.5.
- SyBooks CD for ETL includes:

- Product manuals for Sybase ETL 4.2.

Other sources of information

Use the Sybase Getting Started CD, the SyBooks CD, and the Sybase Product Manuals Web site to learn more about your product:

- The Getting Started CD contains release bulletins and installation guides in PDF format, and may also contain other documents or updated information not included on the SyBooks CD. It is included with your software. To read or print documents on the Getting Started CD, you need Adobe Acrobat Reader, which you can download at no charge from the Adobe Web site using a link provided on the CD.
- The SyBooks CD contains product manuals and is included with your software. The Eclipse-based SyBooks browser allows you to access the manuals in an easy-to-use, HTML-based format.

  Some documentation may be provided in PDF format, which you can access through the PDF directory on the SyBooks CD. To read or print the PDF files, you need Adobe Acrobat Reader.

  Refer to the *SyBooks Installation Guide* on the Getting Started CD, or the *README.txt* file on the SyBooks CD for instructions on installing and starting SyBooks.
- The Sybase Product Manuals Web site is an online version of the SyBooks CD that you can access using a standard Web browser. In addition to product manuals, you will find links to EBFs/Maintenance, Technical Documents, Case Management, Solved Cases, newsgroups, and the Sybase Developer Network.

  To access the Sybase Product Manuals Web site, go to Product Manuals at http://www.sybase.com/support/manuals/.

Sybase certifications on the Web

Technical documentation at the Sybase Web site is updated frequently.

❖ **Finding the latest information on product certifications**

1. Point your Web browser to Technical Documents at http://www.sybase.com/support/techdocs/.
2. Click Certification Report.
3. In the Certification Report filter select a product, platform, and timeframe and then click Go.
4. Click a Certification Report title to display the report.

❖ **Finding the latest information on component certifications**

1 Point your Web browser to Availability and Certification Reports at http://certification.sybase.com/.

2 Either select the product family and product under Search by Base Product; or select the platform and product under Search by Platform.

3 Select Search to display the availability and certification report for the selection.

❖ **Creating a personalized view of the Sybase Web site (including support pages)**

Set up a MySybase profile. MySybase is a free service that allows you to create a personalized view of Sybase Web pages.

1 Point your Web browser to Technical Documents at http://www.sybase.com/support/techdocs/.

2 Click MySybase and create a MySybase profile.

Sybase EBFs and software maintenance

❖ **Finding the latest information on EBFs and software maintenance**

1 Point your Web browser to the Sybase Support Page at http://www.sybase.com/support.

2 Select EBFs/Maintenance. If prompted, enter your MySybase user name and password.

3 Select a product.

4 Specify a time frame and click Go. A list of EBF/Maintenance releases is displayed.

Padlock icons indicate that you do not have download authorization for certain EBF/Maintenance releases because you are not registered as a Technical Support Contact. If you have not registered, but have valid information provided by your Sybase representative or through your support contract, click Edit Roles to add the “Technical Support Contact” role to your MySybase profile.

5 Click the Info icon to display the EBF/Maintenance report, or click the product description to download the software.

Conventions

The formatting conventions used in this guide are:

Formatting example	Indicates
command names and method names	When used in descriptive text, this font indicates keywords such as: • Command names used in descriptive text • C++ and Java method or class names used in descriptive text • Java package names used in descriptive text
myCounter variable *Server.log* *myfile.txt*	Italic font indicates: • Program variables • Parts of input text that must be substituted • Directory and file names
sybase\bin	A backward slash ("\") indicates cross-platform directory information. A forward slash ("/") applies to information specific only to UNIX. Directory names appearing in text display in lowercase unless the system is case sensitive.
File | Save	Menu names and menu items are displayed in plain text. The vertical bar shows you how to navigate menu selections. For example, File | Save indicates "select Save from the File menu."
create table table created	Monospace font indicates: • Information that you enter on a command line or as program text. • Example output fragments
setup -is:tempdir <*full path to alternate temp directory*>	Brackets indicate information that must be supplied by the user.

Accessibility features

This document is available in an HTML version that is specialized for accessibility. You can navigate the HTML with an adaptive technology such as a screen reader, or view it with a screen enlarger.

Sybase Data Integration Suite documentation has been tested for compliance with U.S. government Section 508 Accessibility requirements. Documents that comply with Section 508 generally also meet non-U.S. accessibility guidelines, such as the World Wide Web Consortium (W3C) guidelines for Web sites.

Note You might need to configure your accessibility tool for optimal use. Some screen readers pronounce text based on its case; for example, they pronounce ALL UPPERCASE TEXT as initials, and MixedCase Text as words. You might find it helpful to configure your tool to announce syntax conventions. Consult the documentation for your tool.

For information about how Sybase supports accessibility, see Sybase Accessibility at http://www.sybase.com/accessibility. The Sybase Accessibility site includes links to information on Section 508 and W3C standards.

If you need help

Each Sybase installation that has purchased a support contract has one or more designated people who are authorized to contact Sybase Technical Support. If you cannot resolve a problem using the manuals or online help, please have the designated person contact Sybase Technical Support or the Sybase subsidiary in your area.

CHAPTER 1

Introducing Sybase Data Integration Suite

This chapter introduces Sybase Data Integration Suite and provides information on how to use the various components for your data integration needs.

About Data Integration Suite

Data Integration (DI) Suite combines key data integration techniques, including data federation, replication, real-time events, search, and ETL, with integrated tools for development and administration.

Key features include:

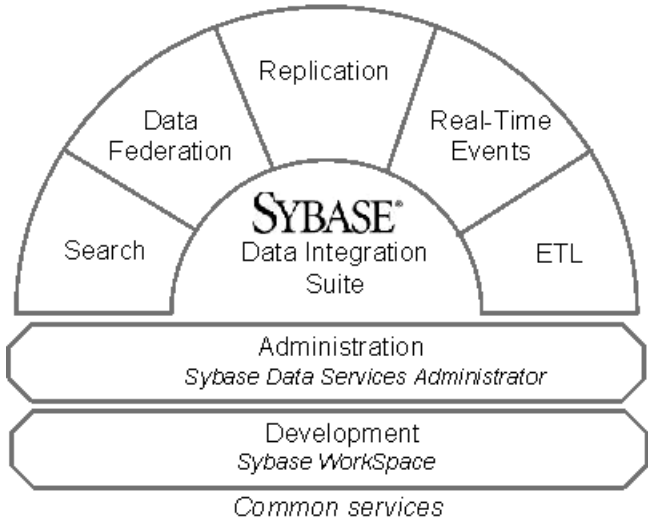
- Access to multiple, diverse data sources, and the ability to create a single, integrated view of enterprise data.
- Access to a variety of heterogeneous data sources including mainframe data sources. Replication, a component of DI Suite, ensures continuous availability of data across your enterprise by delivering real-time information to, and maintaining synchronized copies of distributed data at, multiple sites.
- Capture and propagation of real-time events in data sources to time-critical applications.
- Search and query of information in structured and unstructured data in your enterprise using natural language across numerous data formats.

- Extraction of data from multiple, heterogeneous data sources and loading of data into one or more data targets using a comprehensive set of transformation functions within a complete simulation and debugging environment.
- Development of applications in your data integration environment using an integrated toolset built on the Eclipse framework.
- Management of DI Suite components using a common system administration console.
- An installer that performs installation of all DI Suite components in your environment. The common installer enables interactive installation and silent installation using script-driven installation utility that automates the installation.
- Enterprise capabilities such as high availability, scalability and performance, reliability, security, and runtime monitoring, through the various components of the DI Suite.

Data Integration Suite architecture

Figure 1-1 illustrates the various components of Data Integration Suite.

Figure 1-1: Data Integration Suite



Suite components

DI Suite provides a modular framework that integrates various data integration techniques such as data federation, data distribution and synchronization, data events capture and propagation, data extraction, transform and load, and search. For more information, see "Data Integration Suite components" on page 4.

Integrated administration

Sybase Data Services Administrator (DSA) provides a single management console to manage various servers installed with DI Suite components. For more information, see "Sybase Data Services Administrator" on page 9.

Integrated modeling, metadata, and development

Sybase WorkSpace is the application development environment you can use to build and deliver any kind of application; from event- and data-driven applications to Web-based, composite, and mobile applications. See "Sybase WorkSpace" on page 11.

Common services

DI Suite includes a core layer of integrated services, internally used by its components. These common services include an application server, service container, messaging system, business activity monitoring system, global catalog, and security infrastructure.

For more information about common services, see the *Data Integration Common Services* online topics on your SyBooks CD for Real-Time Events component.

Data Integration Suite components

This section discusses, in further detail, each component of DI Suite:

- Search
- Data Federation
- Replication
- Real-Time Events
- ETL

It also includes more details about the DI Suite administration and development tools:

- Sybase Data Services Administrator
- Sybase WorkSpace

Search

The Search component provides advanced data services for querying, locating, and analyzing your enterprise data. It automatically processes, locates, and analyzes the most relevant information in databases, intranets and Internet, centralized repositories, libraries, file systems, network drives, and existing document management systems in your enterprise.

The Search component:

- Extracts core concepts found in unstructured data across file systems, databases, and Web
- Supports all forms of digital information such as Microsoft Word, Excel, and PowerPoint, plain text files, Adobe Acrobat PDF files, and HTML

  Note Sybase Search uses the Sybase Search Content Adapter, which is an add-on option you can purchase separately, to perform searches across proprietary document formats such as Microsoft Word and Adobe Acrobat PDF documents.

- Automatically captures and aggregates all of your unstructured data
- Enables search in natural language

Subcomponents

Sybase Search includes a central hub server, several satellite containers that include the indexing and search modules, and a Web administration server to provide the search capabilities for your environment.

Related documents

For more information about Sybase Search, view the product manuals on the SyBooks CD for Sybase Search. For more information about how you can use the Search component for data integration, see "Example 1: Improving data retrieval using concept-based searches" on page 17.

Data Federation

Data Federation streamlines integration of data from many distributed sources while providing access to integrated views of your enterprise data.

The Data Federation component:

- Enables a single, federated virtual view of your enterprise data across multiple and diverse data stores such as databases, applications, or files. A "federated" approach indicates that data is delivered from original sources, rather than from data replicas or data marts
- Provides transparent access to your most current data, regardless of its location
- Enables you to create flexible, scalable, and efficient solutions to integrate and deliver data from diverse distributed sources to key enterprise applications

Subcomponents

Sybase Data Federation includes Data Federation servers that provide the Enterprise Information Integration (EII) capabilities and tooling. Data Federation servers are installed with the Data Integration Suite installer while tooling is installed with Sybase WorkSpace.

Related documents

For more information about Data Federation, view the product manuals on the SyBooks CD for Sybase Data Federation. For more information about how you can use the Data Federation component for data integration, see "Example 2: Federating data access" on page 20.

Replication

The Replication component is the data distribution and data synchronization component of DI Suite. It replicates transactional data and synchronizes operational data across heterogeneous databases in your enterprise.

The Replication component:

- Supports movement and synchronization of transactional data in heterogeneous databases such as Sybase Adaptive Server Enterprise (ASE), Oracle, IBM DB2, and Microsoft SQL Server
- Delivers operational data across complex and broadly distributed data sources in near real-time without interrupting critical business applications
- Maintains the integrity of data at the transaction level
- Delivers data efficiently across the enterprise
- Enables bidirectional replication across distributed, heterogeneous systems

Subcomponents

Replication includes the following subcomponents for bidirectional replication and data synchronization:

- Replication Server, for distribution and synchronization of operational data
- Replication Agents, for capturing transactions and transferring them to Replication Server
- DirectConnect™, for access to a variety of LAN-based, heterogeneous data sources, as well as Mainframe data sources

Mainframe options

Sybase Replication supports replication of transactional data from a mainframe-based DB2 server. This support is available as add-on options, which you can purchase separately. The various options are:

- Sybase Replication Agent for DB2 UDB for OS/390 – supports replication of transactional data from DB2 UDB running on a mainframe system to target databases.

- Mainframe Connect™ with DirectConnect for z/OS – supports replication of data to a mainframe system. This option enables client applications and Replication Server to access data stored in mainframe database management systems (DBMSs), file systems, and applications. In addition, Mainframe Connect allows you to create custom mainframe applications that provide access to data stored on mainframe-based DB2 server.

Note Sybase Replication includes basic, simple-to-configure support for replicating to mainframe DB2 through DirectConnect. The Mainframe Connect option is an alternative that provides enhanced performance and functionality while replicating to mainframe DB2. You must purchase the Mainframe Connect option separately.

Related documents

For more information about the Replication subcomponents, view the product manuals on the SyBooks CD for Sybase Replication. For more information about how you can use the Replication component for data integration, see "Example 3: Consolidating data in real-time" on page 22.

Real-Time Events

Real-Time Events proactively captures and moves time-critical events from your heterogeneous data sources to business applications through a messaging infrastructure.

The Real-Time Events component:

- Eliminates the need for intrusive and costly polling applications
- Combines heterogeneous data movement with real-time messaging in one easily integrated, open-standards-based solution, eliminating the need for custom coding and lowering total cost of ownership
- Enables simple auditing and data management to meet security and regulatory requirements with change data capture capability
- Enables event-driven information from multiple systems to be pushed directly to the message bus for a consolidated real-time view of key operational data from multiple data sources
- Supports standard messaging infrastructures such as WebSphere MQ, BEA WebLogic JMS, TIBCO EMS, and Sybase EAServer JMS.

Subcomponents

The following are the Real-Time Events subcomponents, you can use to capture data changes and propagate these changes to standard messaging architectures:

- RepConnector – capture events nonintrusively from a database such as ASE or Oracle and deliver these events to any standard messaging infrastructure such as WebSphere MQ, BEA WebLogic JMS, and TIBCO EMS. Real-Time messaging through RepConnector is achieved using the RepConnector Server, Replication Server, and Replication Agents subcomponents.
- ASE Active Messaging – capture events from the ASE database and publish directly to any standard messaging infrastructure such as WebSphere MQ, TIBCO EMS, and Sybase EAServer JMS. ASE Active Messaging is easy to configure, and provides high performance and enhanced transactional messaging support for ASE databases.

Related documents

For more information about the various Real-Time Events subcomponents, view the product manuals on the SyBooks CD for Sybase Real-Time Events. For more information about how you can use the Real-Time Events component for data integration, see “Example 4: Managing risks in financial services” on page 24.

ETL

Sybase ETL extracts data from multiple, heterogeneous data sources and loads it into one or more data targets using a comprehensive set of transformation functions.

The ETL component:

- Enables extraction of data from heterogeneous data sources such as flat files, XML documents, Microsoft Excel, and heterogeneous databases such as Oracle, MS SQL, DB2, Sybase ASE, Sybase IQ, and MS Access.
- Supports a variety of transformation capabilities and enables you to convert, cleanse, merge, and split data streams, which you can then insert, update, or delete data in a data target. Data transformation capabilities are provided through customizable transformation components and extended JavaScript functions.
- Loads data to target databases or warehouses through native connections. In addition, bulk loading capabilities are provided for enhanced performance.

- Provides a scalable grid architecture that enables parallel transformation processing across operating system boundaries and machines.

Subcomponents

Sybase ETL includes these components:

- ETL Server – a scalable and distributed grid engine, which connects to data sources and extracts and loads data to data targets using transformation flows designed using ETL Development.
- ETL Development – graphical user interface (GUI) tools for ETL development and deployment for use with ETL Server. These tools provide a complete simulation and debugging environment to speed the development of ETL transformation flows.

Note ETL Development is available only on Windows.

Related documents

For more information about ETL, view the product manuals on the SyBooks CD for Sybase ETL. For more information about how you can use the ETL component for data integration, see “Example 5: Consolidating business data for trend analysis” on page 26.

Sybase Data Services Administrator

Sybase Data Services Administrator (DSA) is a graphical interface you can use to manage DI Suite components. It provides a visual representation of DI Suite components organized as panes (shown in Figure 1-2) that includes GUI-based server managers accessed via Sybase Central plug-ins or Web consoles.

Note Use ETL Development, rather than DSA, to manage and administer the ETL component. For more information about ETL Development, view the product manuals on the SyBooks CD for Sybase ETL.

Table 1-1 lists the various DI Suite components and their subcomponents and how they are accessed.

Figure 1-2: Data Services Administrator

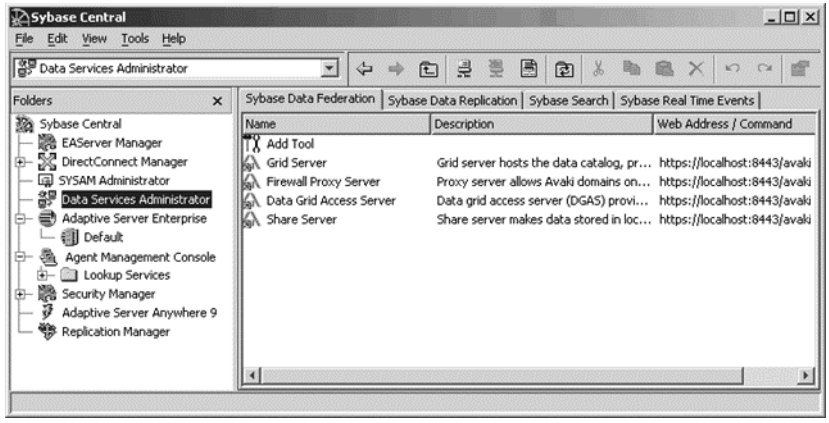


Table 1-1: Data Services Administrator plug-ins

Component	Server manager	Accessed via
Data Federation	Data Grid Access server	Web console
	Firewall Proxy server	
	Grid server	
	Share server	
Replication	Agent Management console	Sybase Central plug-in
	DirectConnect Manager	
	Replication Manager	
	SySAM administrator	
Search	Sybase Search server	Web console
Real-Time Events	EAServer Manager	Sybase Central plug-in
	Enterprise Security Manager	
	Replication Manager	

DSA comes with online help that presents information on performing specific administration tasks. You can access this help after you install any DI Suite component.

Sybase WorkSpace

Sybase WorkSpace provides development capabilities for the Data Federation, Replication, Sybase Search, and Real-Time Events components of the DI Suite.

Sybase WorkSpace is an integrated development environment that provides robust tools, enabling developers and analysts to collaborate when modeling and developing enterprise applications. Built on the Eclipse framework, Sybase WorkSpace allows you to work in a flexible, standards-based, and consistent environment as you go through modeling, development, exploration, deployment, and debugging cycles.

Note Sybase WorkSpace is available only on Windows.

Sybase WorkSpace combines modeling, data management, services assembly and orchestration, Java development, and mobilization in a single tool.

Data Federation development

Sybase WorkSpace provides the tooling for Data Federation servers installed by Data Federation. Sybase WorkSpace provides a graphical, metadata-driven modeling tool for data integration that lets you combine data from heterogeneous data sources. You can build data services, provision or search for data sources, and import or create schema models for data services.

Replication development

Sybase WorkSpace lets you manage and set up a heterogeneous replication system. It provides clients using databases in the replication system with local data access, thereby reducing load on the network and centralized computer systems. After installing and configuring Sybase Replication Server, you can connect to and manage the replication system with replication definitions, publications, articles, and subscriptions.

Search service development

Sybase WorkSpace allows you to manage Search services, in which you can summarize business process logic and data into reusable units. It lets you administer Search servers, create document stores, document categories, and document groups, test a Search query, and manage a scheduled task.

Real-Time Events development

For Real-Time Event management, RepConnector tooling facilitates the capture of real-time database events. RepConnector Manager, which is provided with Sybase WorkSpace, enables you to set up RepConnector connection profiles and connections for communication between databases and messaging systems.

Related documents

For more information about Sybase WorkSpace and which component to use for development of DI Suite components, access the online bookshelf that comes with the software. You can open this bookshelf in the embedded Web browser in the software or in an external Web browser. This bookshelf is organized into collections of topics for each component of Sybase WorkSpace.

You can also access the Sybase WorkSpace bookshelf at Sybase product manuals at http://www.sybase.com/support/manuals/.

Data Integration Suite solutions

Data Integration Suite components provide the tools to address various data integration challenges in your enterprise.

Table 1-2 provides a list of common data integration problems including problem details and the solution from DI Suite.

Table 1-2: Common data integration problems and solutions

Problem	Details	Solution
Trapped, displaced, disconnected, and inaccessible data	• Valuable data is trapped in diverse data stores or incompatible data schemas, and blocked by legacy systems. • Data among data stores is not integrated. • No real-time data is available as changes in data transactions between data stores are not distributed and synchronized.	Replication component – connects to any supported heterogeneous data store and keeps the data in your enterprise integrated and near real-time.
No single view of enterprise data	• Access to operational data impacts the operating performance of the system. • Warehoused data is not real-time. • No unified view of operational data from various data stores. This affects access to critical data for time-sensitive applications.	Data Federation component – helps you to: • Achieve single view across enterprise data. • View operational data in real time. • Create distinct views to view operational data and warehouse data.
No way to provide a unified view of enterprise data without giving complete user access to this data	• Distinct tools are used to view different data across the enterprise. • Access restrictions and stringent sharing rules restrict users from accessing certain data.	Data Federation component – allows you to copy and share data across the enterprise that has strict data share and copy rules. You can also enable single or distinct views of your enterprise data.

Problem	Details	Solution
No automated way to use real-time and historical data to improve the efficiency and quality of business processes	Business Intelligence (BI) tools access only the data warehouse and not operational data stores	Data Federation component – helps you to federate between data in data warehouse and operational systems to obtain a combined view of current and historical data.
No way to identify and prevent unauthorized use of services	• Developers write services and embed the access logic within their code. • Unable to add permissions in the code.	Data Federation component – provides the ability to create users and set access permissions for each service. This enables you to separate the business logic embedded within the code from the service access.
No single solution to deal with large volumes of unstructured information in the enterprise	• Manual tagging of documents based on the content is done to include these documents for search. This is extremely labor intensive and requires extensive tagging of documents. • Use of search engines to perform a keyword search instead of context-based search to search similar content. Keyword search displays results that match in keyword but in a different context.	Search component – enables you to search unstructured data, as well as data in the relational databases. This component provides the ability to do automatic categorization of your documents based on its content and enables search of documents containing similar text, in natural language.
No holistic view of the enterprise after acquisitions	• Each acquired business generates its own reports. • Consolidating reports is difficult and time-consuming	DI Suite provides the following choices: • Replication component – replicate all data to a central database and create aggregated views. • Data Federation component – federate access to data in various businesses.
Information across the enterprise is not synchronized or real-time and this impacts users who use applications to access data	• Applications cannot obtain events as they occur in data stores. • Central repository is updated with data changes only at intervals. Therefore, data is not current.	Real-Time Events component – changes from the source database are captured and pushed as events in a message bus. Applications that subscribe to these events apply the changes to the target database.

Problem	Details	Solution
Difficult to manage complex data extraction methods and data transformation rules	• Various legacy systems require distinct data extraction methods. Building and managing different data extraction methods is difficult and cumbersome. • Target data warehouse has complex transformation rules that increase the complexity of data load.	Sybase ETL component – provides GUI tools you can use to build and manage multiple data transformation and loads from multiple legacy systems to your target data warehouse. It helps you: • Monitor transformation flows in a comprehensive simulation and debugging environment. • Maintain data quality and enhance speed of development of transformation flows through intuitive reuse of previously built transformation processes. • Execute transformation flows in a scalable, grid architecture available on a variety of platforms.

Benefits to enterprise applications

The benefits that Data Integration Suite offers to your enterprise applications are broadly classified into these main areas:

- Operational business intelligence – synthesizes analytical information with data from your organization's operational processes to provide just-in-time information. With integrated data, you can optimize business operations. You can view current information from your operational data stores. You can also analyze historical data from your data warehouse.

  To know how you can implement a Sybase solution for your operational business intelligence enterprise application, see:

  - "Example 2: Federating data access" on page 20.
  - "Example 3: Consolidating data in real-time" on page 22.
- Risk management – combines data from multiple sources to reduce the uncertainty inherent in many key business decisions. With the up-to-the-moment information flow that DI Suite provides, financial services firms can access market, risk, and customer data to make insightful decisions.

To know how you can implement a Sybase solution for your risk management enterprise application, see “Example 4: Managing risks in financial services” on page 24.

- Competitive differentiation – competitive strategies can be revealed by combining information. You can combine business data from disparate data sources to obtain a single, unified view of your enterprise data that may provide you a competitive edge in developing innovative and new data-driven products and services.

  To know how you can implement a Sybase solution for your competitive differentiation enterprise application, see “Example 5: Consolidating business data for trend analysis” on page 26.

- Business event response – monitor information as it flows in your enterprise. You can combine this information with data from analytic systems or transactional data to quickly make informed decisions or alter processes. DI Suite provides information about changes to business conditions that help you to take necessary actions.

  To know how you can implement a Sybase solution for your business event response enterprise application, see “Example 4: Managing risks in financial services” on page 24.

- Compliance – seamlessly search through all types of information, regardless of data format and type. The results enable you to check whether these documents comply with legal standards and regulatory requirements.

  To know how you can implement a Sybase solution for your compliance enterprise application, see “Example 1: Improving data retrieval using concept-based searches” on page 17.

CHAPTER 2

Data Integration Techniques

This chapter shows, through a variety of examples, various ways you can implement Data Integration Suite components.

Example 1: Improving data retrieval using concept-based searches

A global professional recruitment company handles millions of hiring clients and job seekers across the globe, painstakingly searching vast amounts of data in different sources to match job specifications against the resumes of job seekers.

Resumes are stored in relational databases and document systems based on the formats in which they are received. Job specifications are in Microsoft Word files on the Internet while enquiries from companies and job seekers are saved in the in-house file systems. Recruitment consultants must know the format and location of information they want to search.

Incoming resumes and vacancies are manually sorted. Each piece of data is manually tagged to aid search. If errors are made in specifying metadata information, that information may be lost to any future searches. Each resume is tagged manually with keywords to locate these resumes in a future search. This system is time-consuming and gives inaccurate results. Often, the same keywords are used for very different candidates and the search yields results that do not accurately match job descriptions.

Each format and file type is searched using that tool's internal search capability. This requires the searcher to capably perform a smart search, which, if not executed properly, can yield many unwanted results. There is no capability to search in natural language.

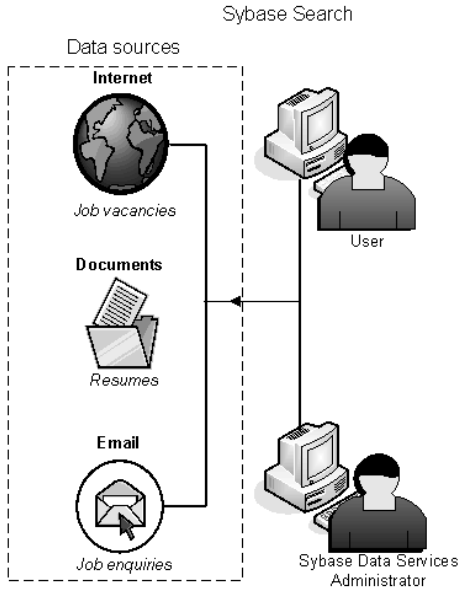
The recruitment company wants to:

- Automate matching suitable candidates to prospective jobs, thereby improving the speed at which candidates become available for selection by recruiting companies.
- Combine information from different sources to allow a single search.
- Conduct smart, fast, and accurate searches on information residing in various data stores and formats.
- Provide a multidimensional search capability to search information using a variety of criteria. For example, the ability to search by job description, skill sets, or location of any candidate.
- Improve the percentage of first-pass hits based on information provided in the job specification.

Sybase Search provides transparent access to structured and unstructured data in your organization using concept-based search capability across numerous data formats.

Note Sybase Search uses the Sybase Search Content Adapter, which is an add-on option you can purchase separately, to perform searches across proprietary document formats such as Microsoft Word and Adobe Acrobat PDF documents.

Figure 2-1: Sybase Search data flow



Data flow

Sybase Search connects to each data source, which are file systems and databases in the customer's organization. Using the content-based catalog and search tool, the Search component automatically analyzes, indexes, and categorizes data and prepares the system for users to perform category specific searches. It extracts and processes the text content from file systems, databases, and Internet where the content is unstructured.

Recruitment consultants can use Sybase Search to search for information in any of these ways:

- Query the categorized information. Perform a powerful search by including advanced search options such as additional metadata, category, and document group search features.
- Review the document summaries of resumes before querying this information. Sybase Search automatically extracts the most conceptually relevant information from larger documents and then summarizes this information based on the concepts provided by the user, providing instant, precise, and relevant information.
- Use resumes or job descriptions to query and find similar resumes.

Administration

Data Services Administrator (DSA) enables you to administer the Search component through a GUI-based server manager accessible via a Web console.

Example 2: Federating data access

Customer service is a top priority for a major commercial and retail bank. The bank strives to maintain good relationships with, and retain current customers, as well as attract new ones. However, the level of service they can currently offer their customers is seriously affected by:

- The inability to efficiently access customer information from remote data sources, such as heterogeneous databases, flat files, and Web services.
- Lack of data-level integration between the source data stores and the data warehouse, which prevents an integrated view of operational data and historical data from the central warehouse.

Customer data is stored in disparate geographical locations in heterogeneous databases such as Microsoft SQL Server, Sybase ASE, and Oracle, flat files, and Web services.

Twice a day, batch files are sent from each of these locations to a central warehouse. Some client applications can access source data directly at different locations. However, executing a query on operational data sources slows the systems.

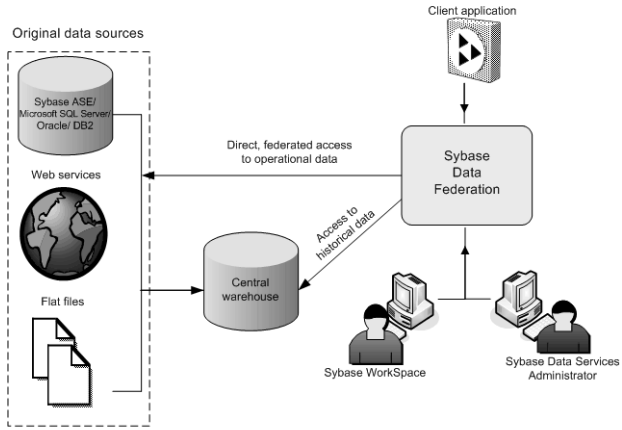
Other client applications access the central repository for historical data, which is not synchronized and real-time.

Customer service representatives require an integrated view of customer data when they interact with customers, allowing them to offer better service across product lines.

Additionally, a “single customer view” allows the bank to develop a comprehensive profile of each customer so it can offer suitable products and services. This view, which must incorporate customer data, enable customer service representatives to identify and reward loyal customers with benefits such as rapid loan approval and quick check clearing.

Sybase Data Federation, a component of the DI Suite, enables customer data to be accessed, integrated, and utilized across the entire enterprise, regardless of the source of the end user or application.

Figure 2-2: Sybase Data Federation data flow



Data flow

Data remains at the source. The Data Federation component connects directly to the operational data sources and provides a single, unified operational data view. It also connects to the central warehouse to create a historical data view. Sybase Data Federation then combines both these views to provide a "single integrated view" of both operational data and historical data to users and client applications.

Development

Sybase WorkSpace helps you to set up unified data framework, data security, and integrated data views:

- Create a unified data service layer for both read and write applications.
- Create **customized, distinct data views** to view operational and historical data. For example, you can create two distinct views, one to query the operational database for an account number and another to query the data warehouse for historical information about the account, such as number of transactions, average account balance, and any penalties or fees charged against the account over a certain period of time.
- Create a **single data view** from operational data and historical data. Two distinct data views are integrated into a single view of customer data available to anyone who needs it.

Administration

Data Services Administrator (DSA) enables you to administer the Data Federation component. DSA uses Web consoles to manage integration of data from multiple sources, Web services, and data views.

Example 3: Consolidating data in real-time

A company has expanded via mergers and acquisitions. This expansion has resulted in new data continually being added by departments and business units at various geographical locations. This data is inaccessible due to incompatible legacy systems and data schemas.

Product and customer data is stored in heterogeneous databases such as Sybase ASE, Microsoft SQL Server, Oracle, and legacy mainframe systems. Numerous in-house tools are used to extract data from heterogeneous data stores and load data into the central warehouse. Data in the central repository is not synchronized, and this lag in information yields imprecise data for analysis.

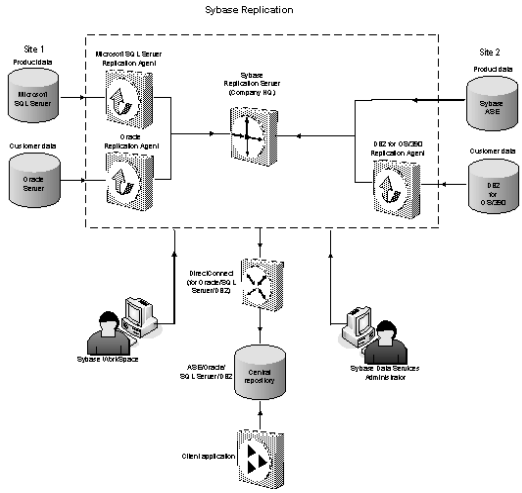
Each office sends data changes in batches, which are consolidated into the central data repository. The company processes batch updates two to four times a day, which does not provide real-time data. More frequent updates place burden on the operational system, while less frequent updates increase latency.

In-house client applications are used to access data from the central repository.

The primary requirement in this example is to gain a holistic view of the entire business and integrate data from multiple heterogeneous databases into one central repository. Information in the repository must be constantly updated and synchronized with real-time changes in various locations.

Sybase Replication provides bidirectional, heterogeneous replication and synchronization of operational data across enterprise, client/server, and mobile systems in near real-time.

Figure 2-3: Sybase Replication data flow



Data flow

At each site, the databases that contain customer and product data are identified. The central repository is initially loaded using tools such as Sybase ETL or other third-party ETL tools. This information is then kept up to date using replication, explained in detail below.

Transactions from heterogeneous databases are received through the Replication Agents connected to the central Replication Server. These replication agents read transaction log files from various databases and pass the transactions to the central Replication Server.

The central Replication Server connects to a repository that stores the consolidated information using DirectConnect. The central repository is now the data warehouse for the company. It is loaded with complete product and customer data and is in synchronization with the remote databases.

Any existing in-house or external client application can access this central repository and perform tasks on the consolidated data.

Development

Sybase WorkSpace helps you to set up the heterogeneous replication system. It provides clients using databases in the replication system with local data access, thereby reducing load on the network and centralized computer systems. After installing and configuring Replication Server, you can connect to and manage the replication system with replication definitions, publications, articles, and subscriptions.

Administration

Data Services Administrator (DSA) enables you to administer the Sybase Replication component through GUI-based server managers accessible via Sybase Central plug-ins. DSA enables you to manage and monitor all distributed components of the replication environment from a single site.

Example 4: Managing risks in financial services

A trading system needs proactive notifications of changes to customer data and securities data. For example, if the buy limit for a security is reached, information must be sent immediately to all trading systems. Currently, the securities operational database must be polled, which introduces information lag, which in turn, introduces risks to credit limit breaches. Such a system does not allow the company to take measures to evade this risk or to identify security breaches early.

At various locations, customer and securities data is stored in heterogeneous databases such as Microsoft SQL Server, Sybase ASE, and Oracle.

The current system periodically polls the database that holds credit limit thresholds resulting in performance degradation to the operational systems. The frequency of polling determines how quickly the trading company can identify whether there is a credit-limit break.

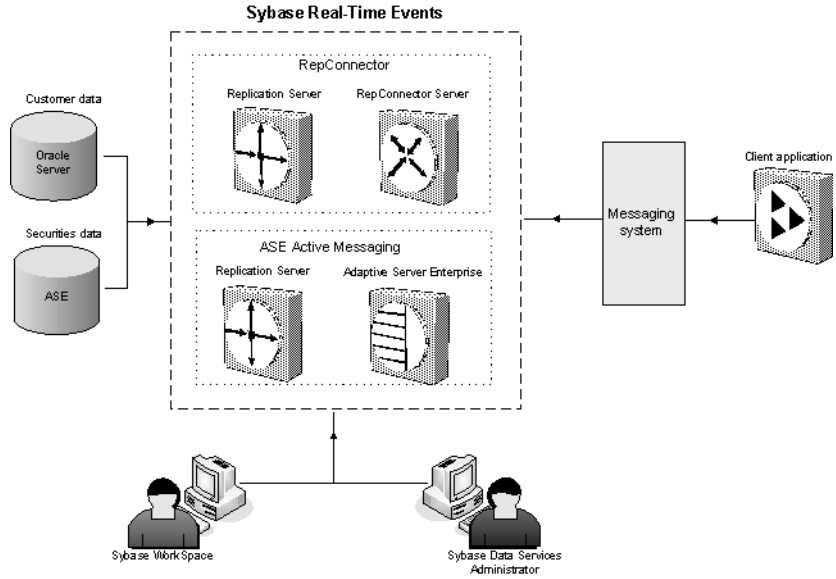
The trading company wants to automatically prevent trading once a customer's credit limit has been reached. Polling the securities operational database is not frequent enough to accomplish this goal.

Sybase Real-Time Events pushes time-critical data from heterogeneous databases to messaging architecture, eliminating the information lags created by batch updates or intermittent polling processes.

Choose either of these:

- RepConnector – enables nonintrusive capture of events from a database such as ASE or Oracle to any message bus using RepConnector.
- ASE Active Messaging – enables capture of events from an ASE database and publishes these events, via messaging services. While this option provides additional functionality and performance for ASE database, it requires an ASE installation in your environment.

Figure 2-4: Sybase Real-Time Events data flow



Data flow

When a customer places a trade request, the trading company uses Real-Time Data Services to send messages from their customer and securities databases to a message bus. The trade request is read by a client application that computes credit risk. If the credit limit is not reached, the system allows the customer's trade. If the buy limit is reached, all trading systems connected to the message bus are immediately notified to take necessary precautionary measures.

The events are delivered to external applications in real-time through a message bus such as IBM WebSphere MQ, BEA WebLogic JMS, TIBCO EMS, or Sybase EAServer JMS.

Development

Sybase WorkSpace helps you to set up and configure the messaging service for the following:

RepConnector – RepConnector Manager plug-in configures and sets up the RepConnector.

ASE Active Messaging – Sybase WorkSpace supports ASE Active Messaging using the Database Development tool to generate SQL code that sends or receives messages from the JMS provider.

Administration

Data Services Administrator (DSA) enables you to administer the Real-Time Events component through GUI-based server managers accessible via Sybase Central plug-ins.

Example 5: Consolidating business data for trend analysis

A company has expanded its business via mergers and acquisitions and now has data from many business units and acquired companies. There is no central repository that consolidates and integrates data to provide a complete view of the business. The lack of data integration has resulted in users writing and running individual reports and consolidating this information to analyze trends.

At various locations, customer and sales data is stored in relational databases, enterprise package applications, and legacy systems. Some custom reporting programs are used to extract data from relational databases and legacy systems.

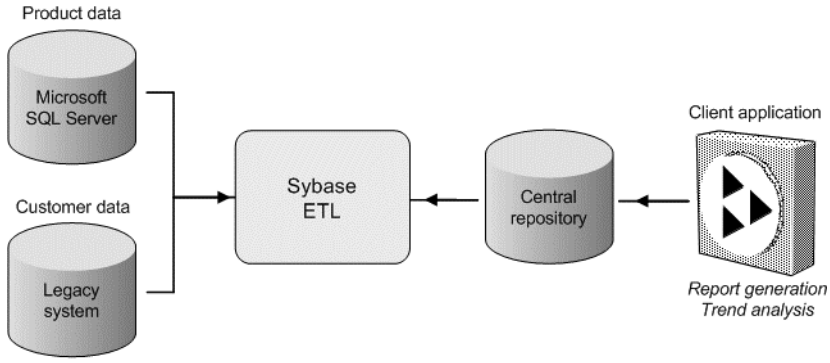
Different reports run against different systems, thereby making report generation resource intensive. Multiple reports reduce the ability to correlate data from different systems.

Data is extracted from data sources and cleaned. The cleaned data is then loaded into the warehouse using an ETL tool. However, there still exists problems such as:

- Individual data stores are of poor quality, and include duplicate entries, orphaned data, spelling variations, and data entry errors.
- Data is not consolidated well enough to support ad-hoc analysis.

The main requirement is to consolidate business data into one central repository. Then, consolidated data must be cleaned, aggregated, and restructured in the central repository. Data in the central repository must enable reporting tools to generate ad-hoc reports on the consolidated information for analyzing trends.

Figure 2-5: Sybase ETL data flow



Data flow

The data stores that contain the business data are identified. Sybase ETL extracts data from individual data stores and loads it into a central repository. The raw consolidated data in the central repository is cleaned, aggregated, and restructured and made ready for access. You can use reporting tools on this central repository to generate ad-hoc reports and trend analyses.

Administration and development

Sybase ETL Development provides GUI tools for administration and development of Sybase ETL.

Index

S

T