

Best Practices for Using Open Text Integration Center - A Technical Whitepaper

Open Text Connectivity Solutions Group

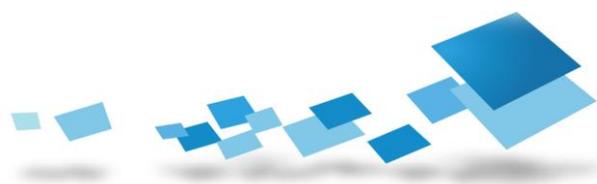
Abstract

This paper is intended for IT professionals interested in understanding and learning about OpenText Integration Center. It presents an introduction to the solution, its architecture, and key features, as well as useful best practice information and use scenarios.

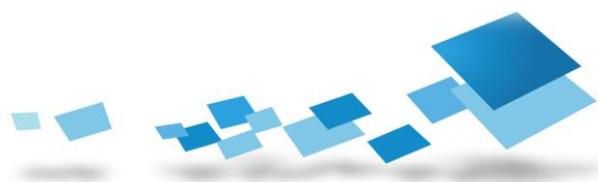


Contents

- Introduction4**
- Integration for the Enterprise5**
 - Transform Information for Corporate Intelligence.....5
 - Integration in the ECM World5
- Open Text Integration Center7**
- Open Text Integration Center Architecture8**
- Adaptable Architecture.....10**
- Integration Center Components 11**
 - The Integration Center Engine 11
 - The Integration Center Repository 11
 - Integration Center Designer 12
 - Integration Center Scheduler 13
 - Administration Tools 13
 - Web Services Publisher 14
 - Integration Center Connectors 15
 - Web Services Connectors 15
 - API Connectors 15
 - Database Connectors 16
 - Text Connectors 16
- Integration Center Key Features 18**
 - ECM Objects – Native Support and Graphical Interface for Information Integration Projects 18
 - Environment Neutral 18
 - Extraction 19
 - Incremental Extraction/Change Data Capture 19
 - Middleware and Standards Support 19
 - Transformation Functions 20
 - Data Mapping 21
 - Transaction and Nested Transaction Support 22
 - Validation against Metadata 22
 - Data Loading 22
 - Tracking Changes 23
 - Dynamic Impact Analysis 23
 - Auto-Documentation 24
 - Versioning 24
 - Metadata Management 24
 - Flexible Scheduling 25
 - Data Quality Management 25
 - Error Handling 26



- Audit and Monitoring26
- Optimizing Performance and Throughput.....27**
 - Failover Capabilities.....27
 - Performance Measurement27
 - Process Optimization and Tuning28
 - Parallelism and Process Slicing.....28
- Processing Methodology29
 - Transformation Performed Exclusively by the Engine29
 - Transformation Performed Partially by the Engine and Remote Databases29
 - Transformations Performed by the Engine and Remote Databases30
- About Open Text Connectivity: Your Trusted Link between People and Information32**

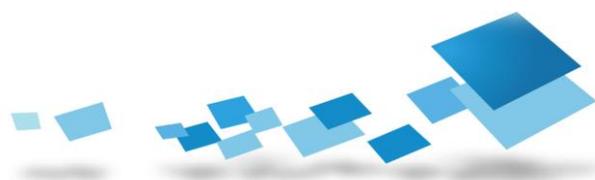


Introduction

The challenge of managing and leveraging all of the data within an enterprise grows increasingly complex. More and more applications, such as Customer Relationship Management (CRM), Enterprise Resource Planning (ERP), and Supply Chain Management (SCM), have become embedded in the enterprise's daily business and, combined with Web applications and legacy systems, they have created an elaborate and complicated IT environment. Many of these applications represent large investments by the company, yet the data contained in these systems is often isolated and not easily accessible.

In today's competitive and demanding business environment, organizations are recognizing the value of analyzing all enterprise data and content to gain a 'single version of the truth' about customer relationships, business performance, and supplier capabilities. And this analysis is starting to take place in 'real time', as businesses operate on 24/7 requirements. The first step in this analysis process is Data and Content Integration—accessing and consolidating the disparate data and systems to feed data warehouses, operational data stores and analytic applications alongside corporate content repositories and new Web 2.0 technologies—which is the basis for analysis of the entire enterprise. Moreover, to enable faster implementation of business processes, organizations need a solution that provides the ability to exchange data between all systems in their IT environment.

This paper discusses Open Text Integration Center as an enterprise data integration tool with a focus on its technical capabilities, key features, and best practices for configuration.



Integration for the Enterprise

The ability to seamlessly and transparently consolidate and interpret information from across the organization has unequivocal business value. However, in today's global economy, a proliferation of data sources—both structured and unstructured, and in various formats and software applications—has brought about a new mindset. No longer is information understood as simply data, but now includes content and other information sources; information that needs to be interpreted, integrated, and made more meaningful to corporate end users.

Transform Information for Corporate Intelligence

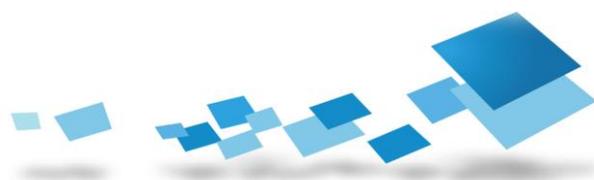
Open Text helps organizations fuse traditional data management and enterprise content management approaches into a single comprehensive information management strategy allowing organizations to realize the true value of their people, processes, and information. Open Text Integration Center supports a complete 360 degree view of the relevant information across the enterprise, ensuring that all business-critical decisions are based on current and complete information regardless of the location or format, for the purposes of business intelligence, governance, and process optimization.

Serving as the hub of the information system, Open Text Integration Center is an enabling data and content integration technology that gives organizations the ability to quickly adapt to new and changing business processes. With powerful and flexible capabilities that transform information from where it is to where it needs to be, Integration Center serves as the central hub to all data- and content-driven systems. With its unique suite of features that help to reduce the cost and duration of development cycles, it offers organizations the ability to break down information silos and cross application boundaries to enable access to data and content by providing a unified view of information throughout the information ecosystem — from leading-edge ERP and CRM systems to legacy applications.

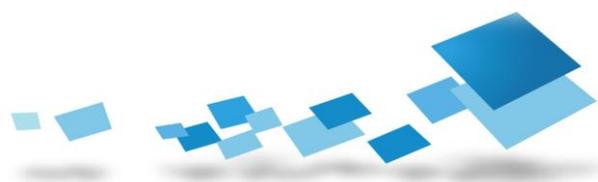
Integration in the ECM World

Open Text Integration Center is the perfect conduit for moving unstructured, semi-structured, and structured information between any source system, such as an Oracle database or file repositories, and any target system, including Open Text ECM Suite.

Open Text ECM Suite stores and manages all types of corporate content, from documents, such as MS Word, Excel, and PDF files, to structured data (database information). The Suite stores this content in the form of XML files, which can be represented in a web page in a user-friendly format using style-sheets called XSLT's.



When migrating corporate content and metadata from a competitor's document management system to the Open Text Enterprise Content Server, only Integration Center ensures that all types of content make the transition, but its uses go beyond competitor migration scenarios. Open Text Integration Center can decommission legacy data by moving them to Open Text ECM Suite from outdated systems. Open Text Enterprise Web Services provides the connection between Integration Center and Open Text Enterprise Content Server, and Content Link—a sample web services library—ensures that Integration Center is ready to connect from the start.



Open Text Integration Center

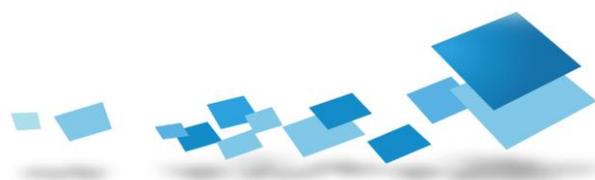
With its ability to access, transform, enrich, cleanse, and direct all types of information across the spectrum of enterprise systems and applications Open Text Integration Center provides a universal integration solution capable of carrying out all data exchange operations regardless of format, syntax, source, or target — from XML support to mainframe connectivity, relational databases to multi-dimensional on-line analytical processing.

Open Text Integration Center extends an organization's existing investment in technology and human resources by seamlessly integrating the corporate IT infrastructure. Integration Center distributes data extraction, transformation, and load processes across computing resources and is platform- and database-independent. These capabilities allow organizations to select the operating system and RDBMS of their choice to store the Integration Center Repository rather than locking users into a single vendor or proprietary solution.

Adding to this capability is a powerful component of Integration Center that uses the native capabilities of source and target relational databases to delegate transformation processes, leverage existing technology, and minimize network traffic.

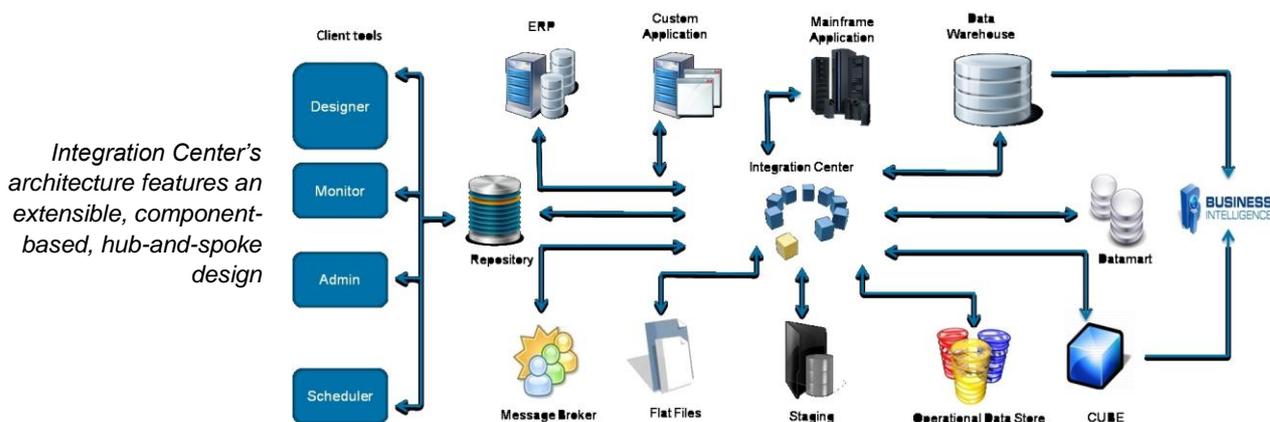
Some key features, abilities, and benefits of Integration Center are:

- Integrate content and metadata from various repositories and web-based resources
- Archive data from any application
- Automatically apply lifecycle management rules to archived content (Records Management classifications, retention rules, and Record Series Identifiers)
- Full data availability for knowledge workers and e-Discovery activities
- Full record extraction from legacy system and transport to ECM archive
- Schedule processes to run on demand, at regular intervals, or under certain conditions
- Full monitoring and audit trail for reporting
- Automated purging of archived legacy data can be done in compliance with corporate authority and applicable retention policies
- Automatic logging of all errors and anomalies
- Recover gracefully from errors and network outages with Engine failover and automatic reconnections
- Parallel processing, and support for multiple Engines on multi-core processors



Open Text Integration Center Architecture

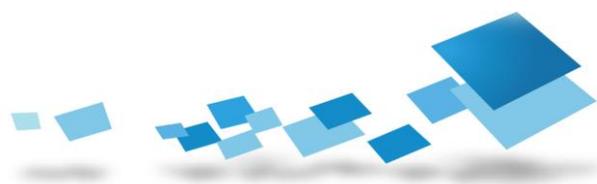
Integration Center’s architecture features an extensible, component-based, hub-and-spoke design. A centralized engine and metadata repository (the hub) exchanges data among data sources and targets (the spokes). Unlike other hub-and-spoke architecture products, Integration Center optimizes data management processes, avoid bottlenecks, and reduce network traffic by leveraging the local database capabilities.



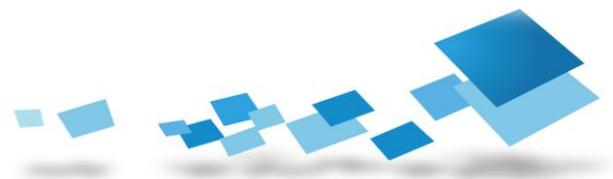
The benefit of a hub-and-spoke architecture with a centralized and open repository is that organizations can maintain full control of all data exchange processes, business rules, and metadata that make up any and all projects within the enterprise, instead of being locked into disparate closed systems. This enhances environment management and empowers knowledge workers to make better, more efficient use of business intelligence and analytical applications.

Since its initial development, Integration Center has followed an open and extensible design concept in order to provide a solid platform for future development, simplifying development of additional functionality and unifying the look and feel of different applications in the Center.

This structural architecture has enabled Open Text to develop a procedural approach to data transformation and exchange processing that gives users unlimited capabilities to transform and process all types of data, whether its traditional structured data or less structured information in content repositories, wikis, and blogs. With this approach, users are not limited to the functions provided by the tool. Instead, they are free to develop their own re-usable transformation code to any degree of complexity.



Integration Center is built on client/server architecture, and incorporates a centralized and open metadata repository. It can be implemented within a distributed deployment model, allowing multiple developers to work on projects simultaneously with complete version control and customized access privileges.



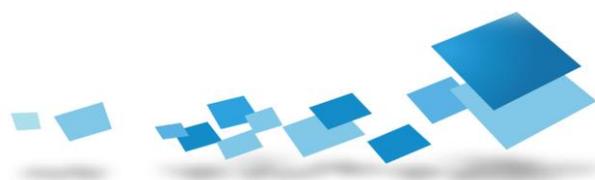
Adaptable Architecture

You can install Open Text Integration Center on Windows, UNIX and Linux platforms and its repository can reside on a variety of RDBMSs, including MS SQL and Oracle.

For simple ETL-type scenarios in which you want to extract structured data from a source RDBMS, transform it, and then load it into a target RDBMS system, simply install Open Text Integration Center on a server between the source and target systems.

For high-volume, time-critical projects, you can install Open Text Integration Center on multiple servers and define multiple engines on each box, thereby deploying multiple Integration Center Engines to share and expedite the extract, transform, and load process. Integration Center Processes can be assigned to multiple engines in this case. For additional performance gains, you can install Integration Center on the same server as the target RDBMS database to avoid network latency.

As outlined previously, Open Text Integration Centre can migrate semi-structured information, such as Microsoft Word, Excel, and PDF documents, to and from Open Text ECM Suite. When dealing with large volumes of documents, installing Integration Center on the same server as Open Text ECM Suite ensures that all web services communication happens locally, dramatically increasing the speed of document ingestion. You can use parallel processing to take advantage of multiple CPU processors on the server.



Integration Center Components

Integration Center offers an integrated set of components that allow organizations to design, deploy, and maintain data transformation and exchange processes. Integration Center's main components include the Engine, Repository, Designer, Scheduler, Administration Tools, MetaLinks, DataLinks, ECM Objects and an extensive set of Web Services (Web Services Publisher) and command-line API's and utilities.



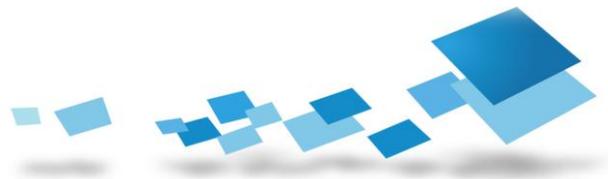
The Integration Center Engine

Integration Center uses a scalable, multi-threaded, transformation engine that brokers information from any source to any target. Integration Center architecture supports distribution and synchronization of data transformation and exchange processes over multiple engines. This is crucial as data volumes increase in size and as transformation processes increase in complexity. It allows Integration Center to leverage the power of existing distributed computing resources. The Integration Center Engine supports Windows®, UNIX (Sun® Solaris, IBM AIX, and popular Linux platforms (SUSE and Red Hat)).

The Integration Center Repository

The Integration Center Repository stores and manages all aspects of data transformation and exchange process metadata. All technical metadata (such as data structures and transformation rules), business metadata (such as business rules and data flows), and production metadata (such as programs and logs), are stored in this repository. The Repository is database-neutral and completely open. It can reside on any of the following:

- IBM DB2 UDB 6, 7, 8, and 9
- Informix 7.x, 11.5, and 2000
- Microsoft SQL Server 7.x, 2000, 2005, 2008
- MySQL Enterprise 5
- MySQL 4.1.x
- Oracle 8.0.x, 8.1.x, 9.x, 10.x, 11.x
- Sybase 11.x, 12, 12.5, 15
- Sybase SQL Anywhere 8, 9, 10



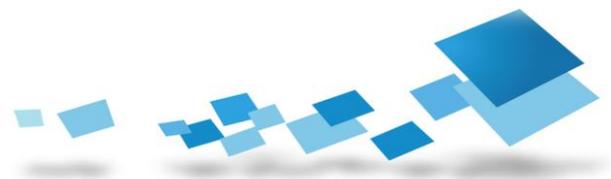
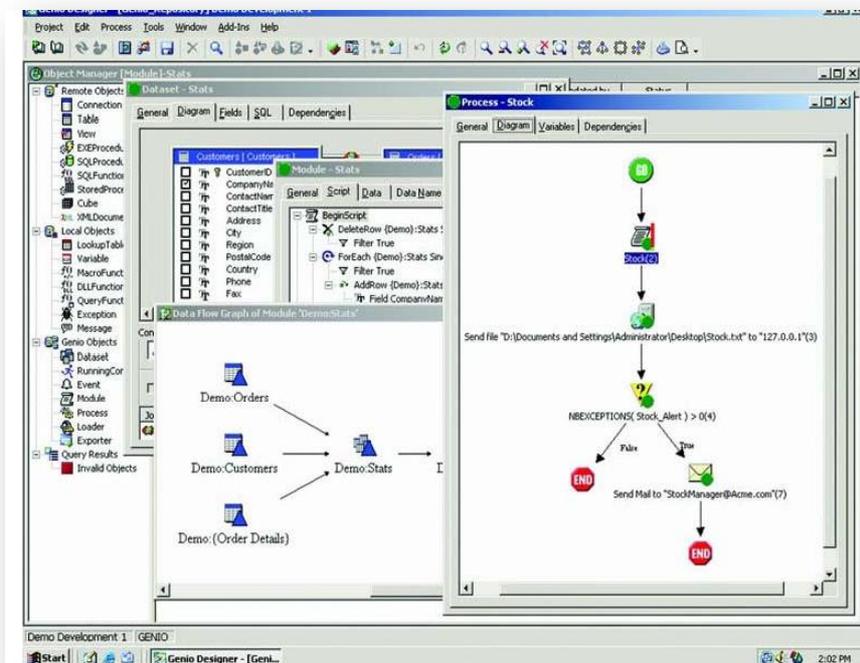
Each component of a data transformation and exchange process is created as an object and stored in this repository. Relationships between objects are automatically maintained, with a comprehensive set of dependency management features. Integration Center's dependency management capabilities provide dynamic impact analysis whenever changes to metadata are identified. Every dependent object impacted by a change (internally or externally) is automatically identified before the next data transformation and exchange process is executed. This ensures information quality and consistency, and reduces the time required to maintain data integration processes.

Integration Center Designer

Integration Center Designer is a multi-user graphical environment for designing data transformation and exchange processes. Data structures can be imported directly from source and target systems or using metadata bridges (MetaLinks). User-defined business rules, functions, and procedures created in Designer are stored as objects within the Integration Center Repository, and are completely reusable from project to project. Integration Center also incorporates a graphical interface that provides a complete and powerful graphical procedural scripting environment for designing data transformation processes of any complexity.

The latest version of Integration Center extends the functionality of the Designer to support multiple Designer sessions via Citrix and Remote Desktop – to further enhance the multi-user experience and improve users' efficiency and productivity across disparate locales.

Designer is the developer's tool that is used to design and create data mappings, extracts, transformations, and exchange processes



Integration Center Scheduler

Integration Center Scheduler provides the ability to program process execution either on a calendar basis or on an event basis. Scheduler also provides monitoring of process executions, as well as full history and audit-trail reporting. Scheduler can work alongside external schedulers like IBM Tivoli® or CA-Unicenter. The substitution process is straightforward, as it can be implemented using standard API or command line interface calls to the underlying architecture.

Administration Tools

Integration Center includes powerful administration tools: Administration Console, Real-Time Administrator, and Execution/Log Viewer.

Administration Console is the central management tool for Integration Center, which you can use to perform essential administrative tasks, including:

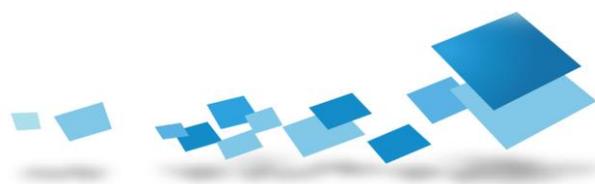
- creating, initializing, and connecting to repositories
- importing and exporting repositories
- configuring Integration Center Services
- defining hosts and configuring Loaders used for bulk transfers between source and target Tables
- defining users and their rights
- defining and connecting projects
- importing and exporting projects to and from a repository

Real-Time Administrator is a real-time communication management application that provides an overview of Integration Center Services as well as administration and/or execution threads running on all host machines defined for a particular Repository.

You can use Real-Time Administrator to:

- view the properties of host machines defined in the active repository
- view the status of Integration Center Services for host machines
- view the status of administration and execution threads
- view and stop Process executions

Execution Viewer provides you with real-time monitoring of Process or Module executions. It lets you view or interrupt the progress of any running execution, and is launched each time you manually execute a Process or Module in Designer or Scheduler, or view a Process execution in Real-Time Administrator. You can also launch it as a standalone application (GenRun.exe) from within Windows Explorer for log viewing purposes. In this case, the application is called Log Viewer.



Execution Viewer launches only when a Process or Module is executed manually. When a Process or Module is executed as a result of a scheduled program, Execution Viewer does not run. For more information on managing Process/Module executions, see Designer, Scheduler, or Real-Time Administrator Help.

Log Viewer lets you view Process or Module execution logs. You can load and view the contents of the following types of execution logs:

- Process or Module execution log (.xml) files on your local machine. To view these files, launch Log Viewer as a standalone application (GenRun.exe) from Windows Explorer.
- Process logs listed in the Logs or History view in Scheduler. To view these logs, launch Log Viewer by double-clicking any Process log in Scheduler.

Web Services Publisher

Open Text Integration Center Web Services Publisher extends Integration Center into a full-featured web services publishing agent. Web Services Publisher exposes Integration Center installations to remote and local clients, allowing them to use Integration Center without installing any of the Client or Server applications locally.

With a fully documented set of web methods (standard APIs), remote users can make calls to Integration Center, using web services to perform such functions as logging in, setting global variables, running processes, and viewing logs generated by Execution Viewer.

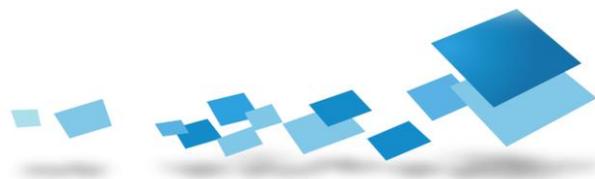
Some key features and benefits of Web Services Publisher are:

- No installation required for end users
- Execute Processes remotely from a custom user interface
- Pass variable values to a Process at execution

The Integration Center Repository must be installed on one of the following RDBMS for use with Web Services Publisher:

- MySQL
- Microsoft SQL
- Oracle
- DB2
- Informix
- Sybase

Web Services Publisher uses JDBC drivers to connect to these RDBMS.



Integration Center Connectors

Open Text Integration Center includes five groups of connector types to ensure connectivity to any source and target systems:

- Web Services connectors
 - ECM Metalink
- API Connectors
 - Mainframe Intelligent Connectors
 - Integration Center MetaLinks
 - Integration with FTP, Messaging Queues
- Database Adapters (ODBC or Native DB drivers)
- Text/XML Connectors

Web Services Connectors

Integration Center can interface with thousands of systems, by leveraging its Web Services connectivity. This universal and comprehensive Web Services support allows Integration Center to interface with all internal and external Web Services-compliant applications. It also allows Integration Center to participate in Service Oriented Architectures (SOA), and integrate other applications' functions in Data Integration processes.

ECM Metalinks

Integration Center comes with connectors called Metalinks for ECM content repositories such as the Open Text ECM Suite, Filenet, and Documentum to name a few. ECM Metalinks specialize in the integration of external information into the Open Text ECM and Archive services platform.

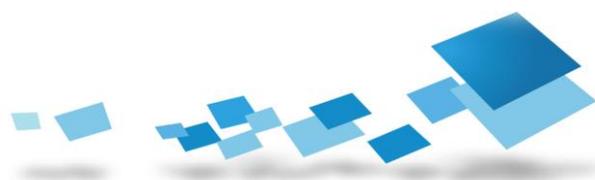
API Connectors

Integration Center includes several types of intelligent API connectors, which enable it to connect to applications/systems with very complex database schemas or lacking web services connectivity.

They are 'pluggable' metadata bridges embedded in Designer that enable the importation of data structures from ECM Repositories, CASE tools, ERP systems, XML Schema, or Web Services Description Language (WSDL) documents.

Integration Center is certified by SAP for both CA-ALE and BW-STA interfaces.

The solution also includes pre-built connectors for middleware and resources such as: FTP, MQ Series, Lotus Notes, MS Exchange and HTTP/HTTPS, as well as a framework for building connectors against additional APIs.



Mainframe Intelligent Connectors

Integration Center includes a set of Mainframe Connectors, which consist of two tiers: a dedicated piece of data access middleware (installed on the host), and an ODBC driver for the specific legacy system. This structure allows Integration Center to extract information from various systems on the mainframe, including: VSAM, IMS/DB, Adabas, Image, Allbase, Eloquence, KSAM, and FDGen files.

Integration Center MetaLinks

The Integration Center MetaLinks are 'pluggable' metadata bridges embedded in Designer that enable the importation of data structures from ECM Repositories, CASE tools, ERP systems, XML Schema or Web Services Description Language documents.

Database Connectors

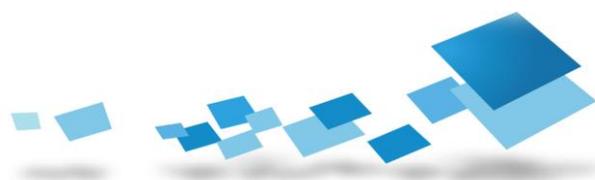
Integration Center Database Connectors connect to most relational databases, including Oracle, Microsoft SQL Server, Sybase, IBM DB2, Teradata, Essbase, and others.

Native population of multi-dimensional databases, such as Essbase, lets users directly create all hierarchies or members, set all necessary attributes, and load or refresh cubes. Through native access, users do not require an additional staging area or complex, multi-layer tools from multiple vendors. There are two advantages to this approach — namely much better performance due to the elimination of any staging area, and maintenance of programmatic control of multi-dimensional cubes within the transformation logic.

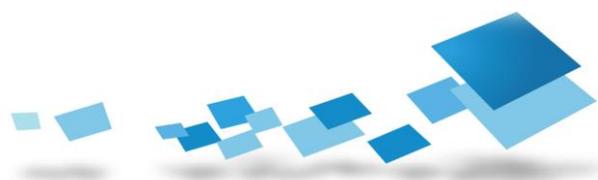
The complete list of available connectivities is included in the *Installation and Administration Guide*.

Text Connectors

Integration Center also natively accesses files such as Fixed Length Files (including mainframe flat files), Delimited Files (CSV), or XML files, and allows processing of any complex files (such as EDI, IDoc, or WebLogs).



Organizations can access information stored in SAP applications, combine it with data from other sources, and then share it with other systems throughout the enterprise. Integration Center delivers native connectivity to extract SAP data, supports bi-directional data interchange with SAP through the SAP IDoc format, and populates SAP BW with data from other systems.



Integration Center Key Features

ECM Objects – Native Support and Graphical Interface for Information Integration Projects

The new ECM Object support introduced in the latest release of Integration Center, makes this platform uniquely and truly capable of understanding and integrating all forms of corporate information: structured (such as database records and views) and unstructured (documents, records, folders and their relevant metadata).

ECM Service Metalink: Import Utility for Open Text ECM Suite objects

The new Metalink allows users to connect to any ECM Service then search, browse the ECM repository and import any underlying structures for use in the OTIC Integration Project. The new MetaLink leverages the system's API's to correctly import and render the object's properties and metadata based on the ECM connection parameters and built-in ECM grammar.

Metadata Mapping Made Easy

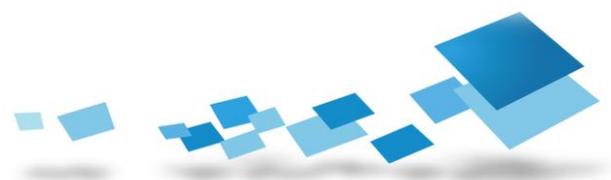
The Integration Center ECM Objects in the Designer interface enable the user to define and map source system information to target document models, easily define metadata value expressions and map XML and database values to various custom metadata fields.

ECM Instructions

Specialized ECM instructions and system functions have been built into the platform in order to simplify the actions a user can perform against the source or target ECM system. For example: Create, Update, Delete, Modify, AddVersion and many other actions may now be leveraged in Integration Center projects, combining existing powerful OTIC features and transformation capabilities with the new functions relevant in the Content-centric world.

Environment Neutral

Integration Center is completely platform and database neutral. These features allow Integration Center users to develop generic business rules without binding them to any specific environment. Objects created in Designer are stored in the centralized metadata Repository. This centralized development model eliminates the need to re-code business rules, lookup tables, and custom functions for each new transformation project. At execution time, the Integration Center Engine reloads metadata-driven processes and generates the appropriate code for the target environment.



Currently, Integration Center supports natively, either in 32 bit mode or in 64 bit mode, six main platforms (Windows, Sun Solaris, IBM AIX, SUSE and Red Hat Linux).

Extraction

Integration Center extracts data from the source databases using native SQL grammar, making it possible to optimize the use of source database power and minimize network traffic. By accessing only the source rows that are pertinent to the transformation work, the Engine avoids loading all the data into a staging area.

When working with text sources, Integration Center has a variety of tools to manage complex structures like hierarchical data dumps from mainframes or EDI files. Integration Center function remains the same, regardless of whether the source or target is a text file or database table.

Incremental Extraction/Change Data Capture

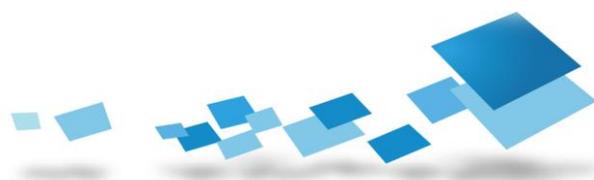
Integration Center offers multiple strategies to perform incremental extraction. Simple approaches, like selection limits based on time stamps, use of database log tables (for example, Oracle Snapshot), or use of database triggers to capture changes, can easily be implemented to enable incremental extraction. These techniques are environment-neutral and can be implemented without any additional software investments.

Integration Center can also capture data changes in most applications through its Web Services connectivity. By accessing the application layer of operational systems through Web Services, Integration Center can get access to any business transactions that occurred in these systems during a certain period of time (such as new purchase orders, updated product records, recently edited documents).

Middleware and Standards Support

Integration Center supports all middleware, providing ODBC, Web Services, HTTP or command line interfaces. It also natively supports message-oriented middleware (MOM), such as IBM WebSphere MQ. Also, Integration Center can use FTP protocol (with or without SSL), MAPI, RSH, or any external application to get access to data or to push data to the target environment.

By leveraging these capabilities, Integration Center provides access to virtually any legacy system, ERP, CRM, SCM systems, Documents and Records Repositories and custom applications.



Transformation Functions

Integration Center has a complete set of transformation functions, making it as capable as a programming language, but providing a graphical and optimized user interface to make the design of transformation routines more productive.

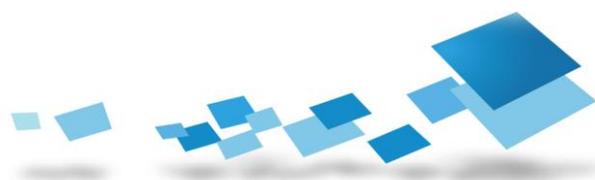
Integration Center offers roughly 150 generic functions that can be used to build complex expressions or custom functions. These functions cover the entire spectrum of string, dates, number or Boolean manipulation. Complex clauses—such as IF, THEN, ELSE or CASE—can also be written in expressions. These functions can be processed inside the Integration Center Engine on Windows, UNIX or Linux, but they can also be automatically translated in native SQL functions in order to execute them on the database engine side.

Using these standard functions, Integration Center users can create their own macros to describe business rules. For example, a function 'Discount' can be calculated from a given sales amount and used everywhere across Integration Center transformations, being processed inside the Integration Center Engine or on a remote database.

For a full description of all available functions, see the *Designer User's Guide* PDF or Designer Help.

Support for Stored Procedures and SQL Functions

Integration Center can also invoke stored procedures or any piece of SQL code that can be executed on databases. These SQL scripts can be declared in the Integration Center Repository to guarantee the reusability of existing code defined within relational databases, either source or target. These stored procedures and SQL Functions can be used to retrieve data, either to extend Integration Center's transformation feature set or simply to trigger external processing on the database side. This also enables better distribution of processing by allowing the use of remote databases transformation functions within data integration processes. For example, Oracle sequences can be reused this way.



Support for External Functions

To extend the processing capabilities of Integration Center, it is also possible to use any legacy function in a DLL written in C++ or any other language. These external functions are declared once in Designer, and can be used seamlessly in all Integration Center transformations, thereby preserving legacy investments.

Also, Integration Center can call any Web Service, executable, external batch or shell script for specialized transformation needs.

National Language and Unicode Support

Integration Center delivers comprehensive National Language Support and Unicode support.

It allows simultaneous connections to multiple systems encoded in different character sets and exchanges data between these systems. Integration Center supports most single-byte, double-byte and multi-byte character sets as well as Unicode. Whenever possible, Integration Center can convert data from one character set to another and simultaneously manipulate strings encoded in different code pages.

Integration Center's user interface also fully supports Unicode. It allows manipulation of metadata encoded in different character sets, and delivers support for international development teams.

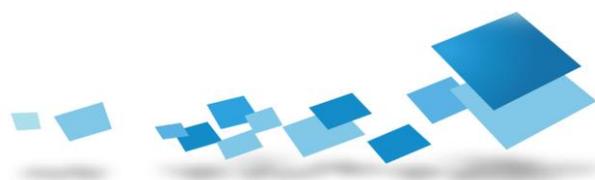
Data Mapping

Integration Center provides different ways to define mapping. Whenever possible, the tools can automatically detect mapping based on field names, field order, or any custom algorithm. Also, simple graphical mapping from the source to target can be manually defined using drag and drop functionality, and more complex mapping can be done using the Integration Center's graphical procedural language.

Aggregating, Filtering, Sorting, and Creating Joins

When multiple sources, heterogeneous or otherwise, are required, users are able to define datasets (logical views on the information system) to de-normalize, join, aggregate, sort, and distinguish data from the various source systems.

These datasets can combine multiple objects from heterogeneous systems, and can also be used in other datasets.



These operations are defined graphically inside Integration Center, with no need to write SQL code. Nevertheless, they cover the entire functional spectrum of relational database features. Users can define regular joins, external left or right joins, full outer joins, calculated joins, or recursive joins involving the same table or view several times, through aliases that Integration Center manipulates transparently. Filters are transformed in WHERE or HAVING clauses, and sorting becomes ORDER BY. Integration Center recognizes each SQL grammar, adapting itself to the source or target DBMS.

Transaction and Nested Transaction Support

Integration Center uses the traditional transactional mechanisms of relational databases, COMMIT/ROLLBACK, on one or more databases. Transactions can be distributed on several systems. They can occur at any moment during the execution of the interface, for each single treatment, or for each important step inside a single treatment for each functional unit. For example, it is possible to validate a transaction for each client change when processing a single table containing invoices. The COMMIT/ROLLBACK mechanisms can be triggered automatically or conditionally, based on data test results on previous interface executions. This is typically used to add failover capabilities to existing interfaces.

Integration Center can do conditional COMMIT/ROLLBACK for any module within a process. This means that Designer can create a complex condition for loading and data quality checking, and decide whether to commit or rollback the load.

Validation against Metadata

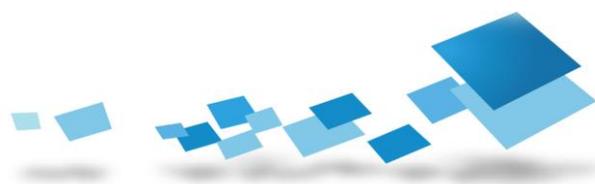
Integration Center is completely metadata-driven and always validates the business rules and changes against the metadata and operational sources and targets. While it is possible to write logically incorrect statements, their physical grammar is automatically tested and validated by Integration Center.

Metadata can be defined using Designer, either by manually editing it, or by importing it. Metadata can be imported from database schema, using sample files or using Integration Center MetaLinks. By using third-party solutions, such as Meta Integration, it's also possible to implement bi-directional metadata exchange with complementary software, such as BI tools (BO, Cognos), metadata repositories or industry standards (CWM).

Data Loading

Integration Center has multiple loading strategies—single row, packet, and bulk. In certain cases, the loading can be done by the source database directly, when the developer has decided to bypass the engine altogether.

For more details, refer to the topic, 'Integration Center's Unique Processing Methodology'.



Integration Center's design environment does not impose a pre-defined methodology to implement data-loading processes. It has been designed to be highly productive and generic to support different needs. Integration Center provides a comprehensive and flexible solution that supports both a full data refresh strategy as well as data updates when needed.

Delete, Insert, and Update strategies are all supported natively. Integration Center also provides high-level, user-friendly commands such as SmartInsert and SmartUpdate, designed to simplify row additions or updates in tables (database Merge).

Tracking Changes

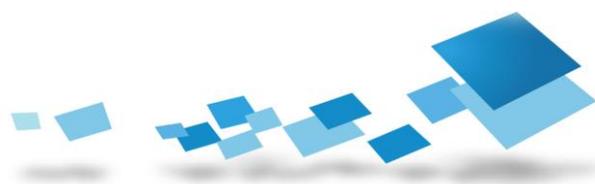
Integration Center provides the ability to track differences between an object definition stored in the Integration Center Repository and the state of the same object as it exists in the remote system (physical object in the source or target). By utilizing the 'Track Changes' wizard, Integration Center users can automatically detect and import changes into the Integration Center Repository. Every change made to an object, whether it's located in a remote database or in Integration Center, is also stored and available for documentation purposes. This means that Integration Center is always consistent with data structures as they exist on remote sources and targets, ensuring data accuracy and consistency in every data transformation and exchange process.

Dynamic Impact Analysis

Integration Center provides a unique, one-of-a-kind, Dynamic Impact Analysis solution. The consistency and status of objects are automatically maintained in real-time within the design environment. Integration Center Designer provides two impact analysis modes— immediate and deferred.

In the immediate impact analysis mode, Integration Center immediately checks the impact of a change on an object and on all dependent objects. If the change impacts the integrity of an object, its status is automatically changed to 'Invalid Object'.

When an object is modified in the deferred impact analysis mode, Integration Center changes its status to 'Undefined Object'. Objects with this status can be verified at a later date.



This impact analysis is triggered either by changes made by the developer within the Integration Center environment or by the 'Track Changes' feature. For example, if a source data structure changes, the 'Track Changes' feature detects it, and the impact analysis identifies the effect of the change on all related interfaces. Integration Center's impact analysis eliminates the need for developers to spend time manually tracking down dependencies whenever a change is made. Integration Center provides a persistent list of invalid and undefined objects, allowing developers to know the exact state of their metadata, and the immediate consequence of making a change to it. This dynamic impact analysis helps developers fix impacted objects by providing a thorough description of required changes, and even auto-correction mechanisms. This decreases the length of the maintenance cycle and increases developer productivity.

Auto-Documentation

Integration Center Designer automatically manages the documentation of projects, including dependencies between objects, modification history and comments.

Integration Center users can automatically print or generate HTML documentation, dependency graphs, or dataflow graphs at any time. This significantly reduces documentation efforts, and ensures the accuracy of project documentation.

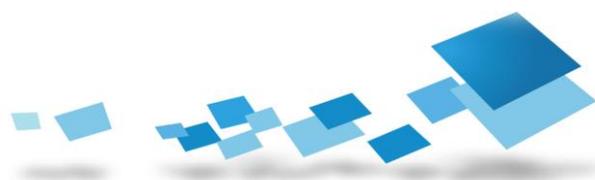
Versioning

Integration Center is built on client/server architecture and leverages an open metadata repository. It can be implemented in a centralized or distributed deployment model, allowing multiple developers to work on projects simultaneously with complete version control and customized access privileges.

Integration Center natively supports version management and status management (for example, Development, Test, and Production). All versions of data integration projects are independent from each other and can be used in parallel. All objects in these projects have timestamps for creation and modification as well as user information and comments. The history of data structure modifications is also maintained automatically by the tools.

Metadata Management

With the vast amount of information that organizations currently have at their disposal, there is an ever-increasing need to collect, manage and reuse that information. Organizations want to know what information they possess, its location, its origin, and its size. This 'data about data' is called metadata. It can describe any characteristic of the data—such as the content, its structure, its quality, or any attributes related to its processing or changes.



Quite simply, metadata is an important catalog of information from any number of sources—data warehouses, data exchange tools, business intelligence tools, ERP CRM, SCM, business process modeling, workflow, data quality tools, ECM systems, or any other application dealing with data.

Metadata secures the lineage of data, enabling knowledge workers to gain access to business rules, and to understand where the data came from and how it has been handled to date. This makes the time they spend on query and analysis activity more productive.

Metadata management provides critical access for both business users and technical users working with the data. Depending on the type of user, metadata can serve either as a blueprint to the inner technical workings of the warehouse, or as a roadmap to assist in navigating the warehouse and locating useful information.

Metadata also delivers valuable help to organizations when it comes to their compliance with regulatory rules.

The Integration Center Repository contains all metadata used by data integration processes. This metadata is made available to users through Integration Center tools—either by querying Integration Center’s open database repository or through XML datagrams.

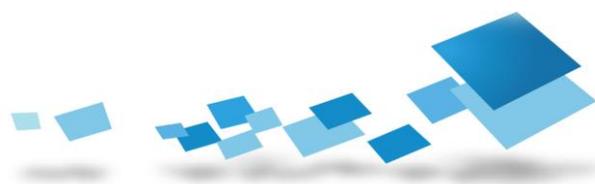
Flexible Scheduling

Integration Center includes a complete scheduling facility, making it possible to schedule process execution at a fixed or recurring time, periodically (daily, weekly, monthly), triggered by outside events, or from the polling service (file based events). Data Integration processes can also be triggered by external events or Message Oriented Middleware (MOM) such as IBM WebSphere MQ. Combining these functions, Integration Center developers can build as complex scheduling rules as necessary.

Integration Center Scheduler is not always required, as there is support for external applications setting Integration Center variables, launching processes and receiving the result of the process. This makes it very easy to implement use of system management applications like IBM Tivoli or CA Unicenter. The substitution process is straightforward, and it can be implemented on UNIX, Linux, or Windows, using standard API calls or command line utilities.

Data Quality Management

Integration Center delivers data cleansing capabilities through its own functions or through partner products. Using Integration Center built-in functions (such as String functions or Soundex), and Integration Center procedural language, it’s possible to implement a basic cleansing process. By leveraging third-party products, such as Harte-Hanks Trillium Software or SPAD DQM, it’s possible to implement a more complex cleansing process involving address cleansing, pattern matching, and more.

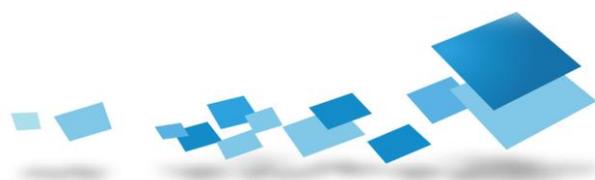


Error Handling

Integration Center, through its graphical procedural language, also delivers exceptional error handling capabilities. Integration Center automatically reports all errors and anomalies in its log. Technical exceptions (such as Datatype issues, or constraint violation), are automatically handled by the tool, while other exceptions types, such as business-rules-driven exceptions, can be handled through user-defined exceptions. Integration Center users can then implement various exception handling strategies and decide if the execution should be stopped after a certain number of exceptions, or if incriminated data should be output into rejection files. All of this can be done using Integration Center's procedural language, which provides users with an easy and comprehensive mechanism for error handling. It is possible to define any logical test and implement virtually any type of processing according to organization business rules.

Audit and Monitoring

All process execution information is logged into the Integration Center Repository. Integration Center logs everything from the SQL statements used to all timing information and anomalies — either system or designer defined exceptions. This log information can be accessed from the Scheduler client, or by using reporting tools to query Integration Center's open repository. It is also possible to use email as a delivery vehicle to inform DBAs or system administrators of processing results. Integration Center can also package the logs into different formats (such as Text, XML). These can be sent as SMS messages to mobile phones or via email.



Optimizing Performance and Throughput

You can help optimize Open Text Integration Center performance even before you create an Integration Center process by cleansing the data of .csv, .txt and xml files before you load the clean data into a RDMS system. This drastically improves read/write speed for projects that involve thousands of records.

Aside from data cleansing, Integration Center offers numerous ways to optimize the performance for ETL and document migration projects. The most obvious method is to set up multiple instances of Integration Center Engine that can share the processing load in parallel. Multiple engines allow you to group records or documents into batches and assign each batch to a different engine for processing.

Another method for improving performance for projects that involve hundreds of thousands of records or documents is to turn off the writing of status and exceptions to the Console.

For operations involving writing to RDBMS tables, you can change the Open Text Integration Center project to write using Packet Mode. Single Mode is slower but provides better error tracking. Packet Mode is faster than Single Mode but its error tracking is less precise.

Failover Capabilities

Integration Center does not impose any methodology on failover functionality. The open architecture enables the developer to use any preferred technology for failover systems—including “power-off” restarts or complex rules on continuity.

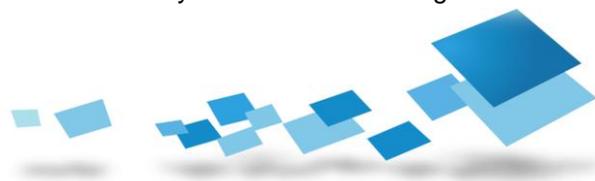
To permit such implementations, Integration Center provides the key features required to implement complex failover strategy. When triggering a process execution, Integration Center users are able to define the list of the Integration Center Engines and timeout for each one.

If the process execution fails, Scheduler can automatically trigger a fail process that will implement the desired failover strategy or restart the same process. Within each process, users can define restart points and therefore automate process restarting.

Open Text training and best practices routinely teach these different approaches, and can help Integration Center developers find the best approach for each specific situation.

Performance Measurement

Integration Center has a unique performance meter inside its logs. All the different tasks are timed, including the module coherence tests and SQL statement performance, as well as the load processes. Also the volume of impacted data on every single target system is readably available in these logs.



As a result, the Data Integration process administrators can easily spot any potential performance problems.

Integration Center can automatically email this report to the developer or the system administrator after each execution, as well as keep it in the Repository, making it possible to analyze the performance.

It is also possible to use performance measurement tools to detect and isolate networks, machines, or any other potential bottlenecks.

Process Optimization and Tuning

Once again, Integration Center openness delivers multiple ways of optimizing process executions. Integration Center offers ways of offloading part of, or the entire process execution, on the source or the target system.

For more information, see the 'Integration Center Unique Processing Methodology' section.

Integration Center provides multiple reading and writing strategies (single, packet, and bulk) that enable you to optimize data movements based on the particular topology of your information system.

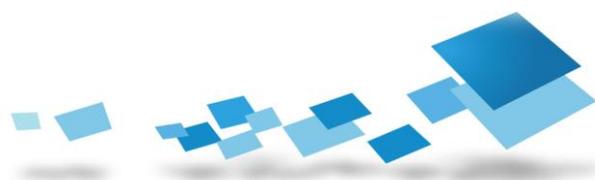
When loading needs to be optimized, indexes can be dropped and restored on the target so that the database engine can accept rows at the speed they are sent from the Integration Center Engine. This is very simple to achieve with Integration Center's SQL procedures and SQL functions.

When required, SQL statements can be tuned using the database SQL analyzer or database hints that ensure maximum utilization of resources. The tuning of the target is usually based on the performance of the database and the type of end-user tools used, and how they are used. Typically, the database environment provides tools to assist tuning by showing statistics on the use of disk, indexes and processor time for queries.

Parallelism and Process Slicing

Integration Center has native multi-threading capabilities that enable processing tasks to be performed in parallel—it communicates through global variables and events between the processes. Increased performance can be achieved by splitting the load across multiple CPUs or across multiple physical servers, depending on where the bottleneck is located.

Integration Center has a simple facility to perform source slicing based on the RDBMS sources. Again this architecture is open and does not impose a foreign methodology on the developer. Integration Center gives users an easy way of defining multiple execution contexts that will handle a subdivision of the entire process with its 'Running Context'.



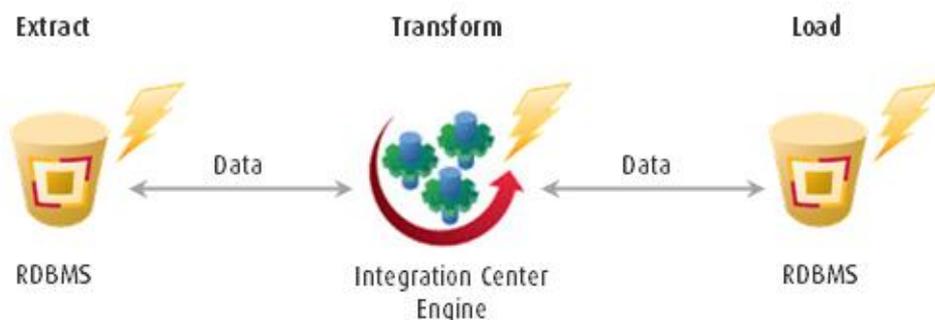
Processing Methodology

Integration Center offers a unique methodology that distributes transformation workload by offloading certain tasks to idle database engines during off-peak hours to maximize efficiency and system performance. The following graphics and descriptions depict the three modes of transformation processing that Integration Center offers:

Transformation Performed Exclusively by the Engine

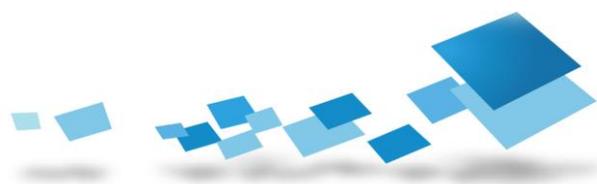
Integration Center extracts data from any source database, transforms it using the Engine, and then loads it into the target database tables. This model is suitable whenever data sources and targets are heterogeneous or whenever the required transformations cannot be performed natively by the source/target databases, or when operational requirements dictate that processing be undertaken away from the source (such as when dealing with OLTP systems where getting the source RDBMS system to undertake the transformations would impose an unacceptable performance hit).

Data Access Method One

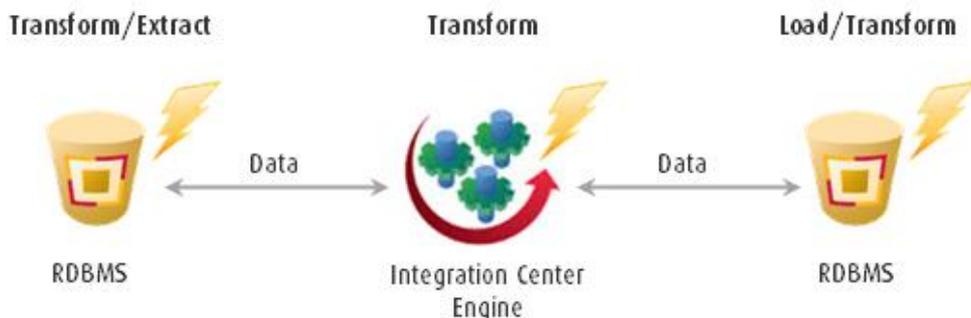


Transformation Performed Partially by the Engine and Remote Databases

In this model, Integration Center takes advantage of relational database features by offloading some of the transformation directly to a data source or target. Transformed data is extracted from the source database, brokered through the Engine – where additional transformation may occur – and loaded into target database tables, where, again, further transformation may be conducted. In this case, aggregations and consolidations are typically processed at the source, reducing the network bandwidth required to transport data.



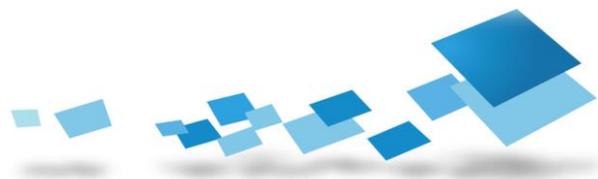
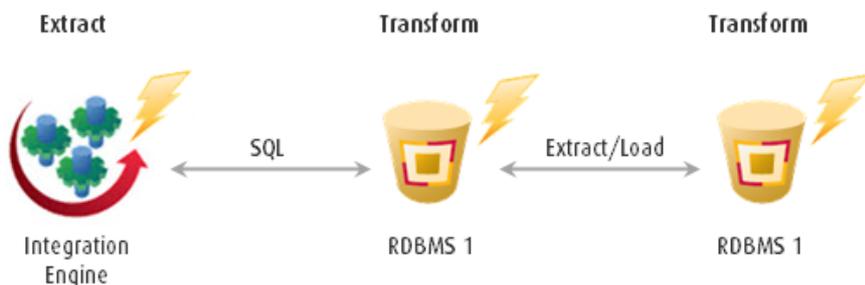
Data Access Method Two



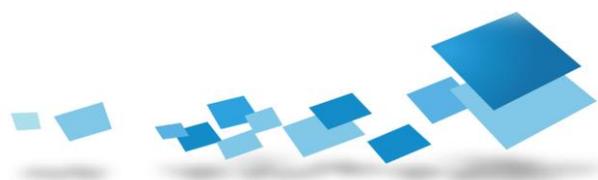
Transformations Performed by the Engine and Remote Databases

With the third model, the source and target are on either part of the same logical server, or visible to each other (using database link). In this situation, it is not necessary for the data to leave the server or to transport the data through a communication layer (the Network). In this context, Integration Center can operate as a dynamic code generator, and only send SQL orders that have been adapted to the relational database. The RDBMS manages the extractions, transformations, and insertions (or updates). As a result of this architecture, this model outperforms the previous two. This is because no network bandwidth is used and it fully exploits the processing capabilities of the database platform, including Massive Parallel Processing (MPP) architectures.

Data Access Method Three



Each data access mode is accessible through a common user interface and data integration process. These various modes are defined using the same graphical metaphor and programming methodology. By maximizing user control over data flow, Integration Center data access architecture enables users to significantly improve the performance of their data exchange processes. Being able to select, manage, and summarize only relevant data, and control the platform on which work is executed, vastly improves performance. Regardless of which data access model is chosen, Integration Center impact analysis capabilities are maintained, ensuring that if changes are made to any element of the data exchange process, administrators are notified prior to the next scheduled execution.



About Open Text Connectivity: Your Trusted Link between People and Information

Open Text Connectivity connects people, data and applications in mission-critical environments with an award-winning suite of solutions. For over 20 years Open Text Connectivity has continued to combine the best of both worlds: the strength of one of the largest software companies and the spirit of a customer-focused business.

Sales: 1 877.359.4866
Integration-Experts@opentext.com

Support: connsupport@opentext.com
1 905 762 6400

www.opentext.com