opentext™

Best practices for using OpenText Magellan Integration Center

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

opentext™

Contents

opentext™

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, 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 for customer relationships, business performance and supplier capabilities. And, this analysis is starting to take place in realtime, as businesses operate on 24/7 requirements. The first step in this analysis process is data and content integration—accessing and consolidating 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 OpenText™ Magellan™ Integration Center as an enterprise data integration tool with a focus on its technical capabilities, key features and best practices for configuration.

opentext™

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. This is information that needs to be interpreted, integrated and made more meaningful to corporate end users.

Transform information for corporate intelligence

OpenText 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. Magellan 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, Magellan 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, Magellan Integration Center serves as the central hub for 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

Magellan 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 OpenText™ ECM Suite.

OpenText ECM Suite stores and manages all types of corporate content, from documents, such as Microsoft® Word, Microsoft® Excel® and PDF files, to structured data, including database information. OpenText ECM Suite stores this content in the form of XML files, which can be represented as a web page in a user-friendly format using style-sheets called XSLTs.

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 Magellan Integration Center can decommission legacy data by moving it to OpenText ECM Suite from outdated systems.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

opentext™

Magellan Integration Center

Magellan Integration Center has the ability to access, transform, enrich, cleanse and direct all types of information across the spectrum of enterprise systems and applications. It is a universal integration solution capable of carrying out all data exchange operations regardless of format, syntax, source or target, from XML support and mainframe connectivity to relational databases and multidimensional online analytical processing.

Magellan Integration Center extends an organization's existing investment in technology and human resources by seamlessly integrating the corporate IT infrastructure. It 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 Magellan Integration Center Repository, rather than locking users into a single vendor or proprietary solution.

Adding to this capability is a powerful component of Magellan 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 Magellan Integration Center:

- Integrate content and metadata from various repositories and web-based resources

- Archive data from any application

- Automatically apply lifecycle management rules to archived content, including records management classifications, retention rules and record series identifiers

- Full data availability for knowledge workers and eDiscovery activities

- Full record extraction from legacy systems, transported to an 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 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

Magellan Integration Center is an essential part of OpenText™ Magellan™, the AI data analytics platform.

Magellan Integration Center architecture

Magellan 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, Magellan Integration Center optimizes data management processes, avoids bottlenecks and reduces network traffic by leveraging the local database capabilities.
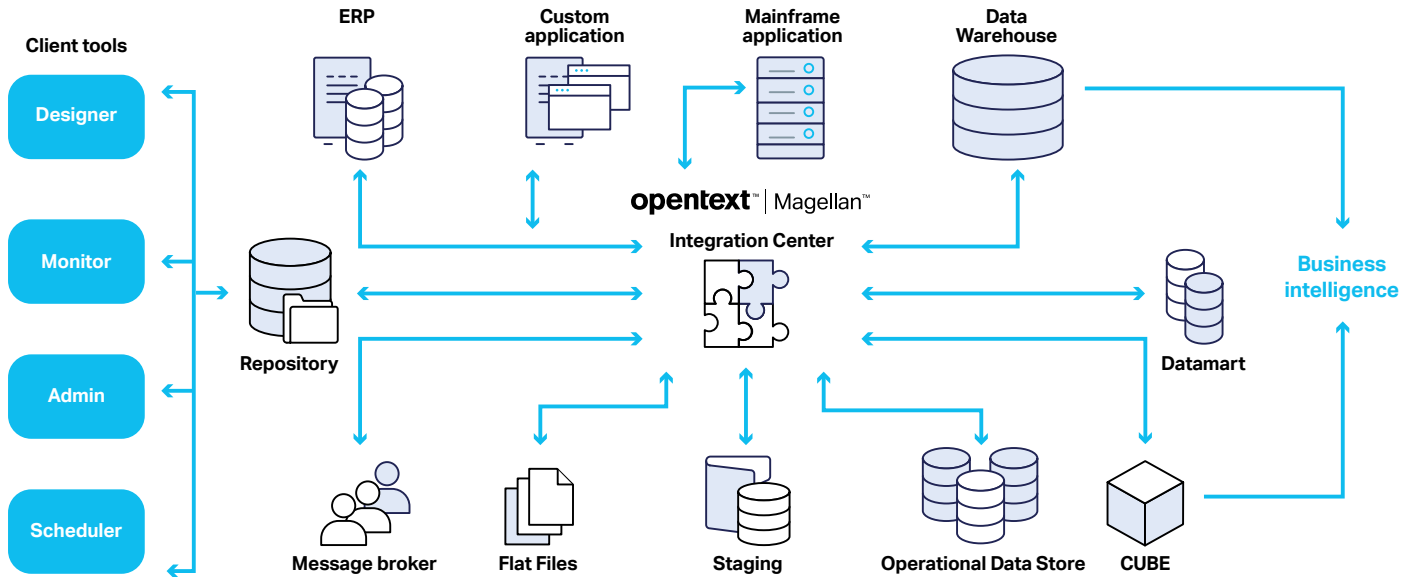
OpenText Magellan Integration Center's architecture features an extensible, component-based hub-and-spoke design

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, Magellan 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.

This structural architecture has enabled OpenText to develop a procedural approach to data transformation and exchange processing. It gives users unlimited capabilities to transform and process all types of data, whether 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 reusable transformation code to any degree of complexity.

Magellan 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

Users can install Magellan Integration Center on Microsoft® Windows®, and various UNIX® and Linux® platforms and its repository can reside on a variety of RDBMSs, including IBM DB2, Informix, Microsoft® SQL Server, and MySQL, PostgreSQL, SAP HANA, Sybase. and Oracle.

For simple ETL-type scenarios in which users want to extract structured data from a source RDBMS, transform it and then load it into a target RDBMS system, Magellan Integration Center can be installed on a server between the source and target systems.

For high-volume, time-critical projects, users can install Magellan 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. Magellan Integration Center processes can be assigned to multiple engines in this case. For additional performance gains, users can install Magellan Integration Center on the same server as the target RDBMS database to avoid network latency.

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

Magellan Integration Center components

Magellan Integration Center offers an integrated set of components that allow organizations to design, deploy and maintain data transformation and exchange processes. Magellan Integration Center's main components include the Engine, Repository, Designer, Debugger, Scheduler, Administration Tools and MetaLinks. It also includes OpenText™ Magellan™ Integration Service with its ECM Connectors, an extensive set of web services and command line APIs and utilities.

| Administration Console | Designer | Real-Time Administrator | Scheduler | Magellan Integration Center Repository | Magellan Integration Center Engine |
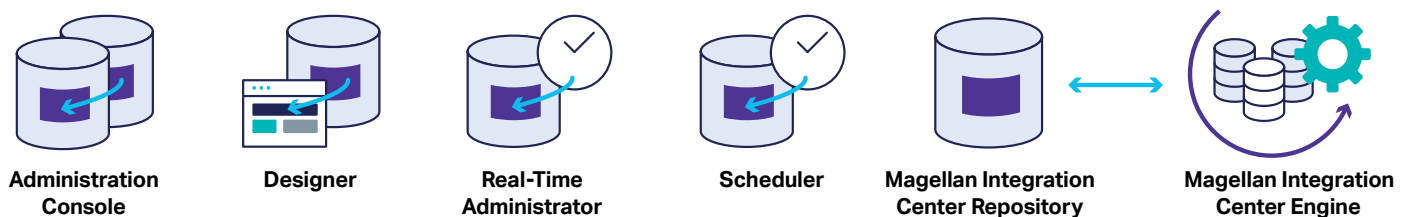
Magellan Integration Center Engine

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

opentext™

Magellan Integration Center Repository

The Magellan 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:

• Microsoft® Azure® SQL Database

• IBM DB2 UDB

• IBM DB2 UDB (CLI)

• Informix

• Microsoft SQL Server

• MySQL Community Server

• MySQL Enterprise Server

• Oracle

• PostgresSQL

• SAP HANA

• Sybase

• Sybase SQL Anywhere

For more details on versions, please refer to the Supported Connectivities section of the Installation and Administration Guide.

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. Magellan 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.

Magellan Integration Center Designer

Magellan 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 Magellan Integration Center Repository and are completely reusable from project to project. Magellan 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 Magellan Integration Center extends the functionality of the designer to support multiple designer sessions via Citrix and Remote Desktop, further enhancing the multi-user experience and improving user efficiency and productivity across disparate locales.
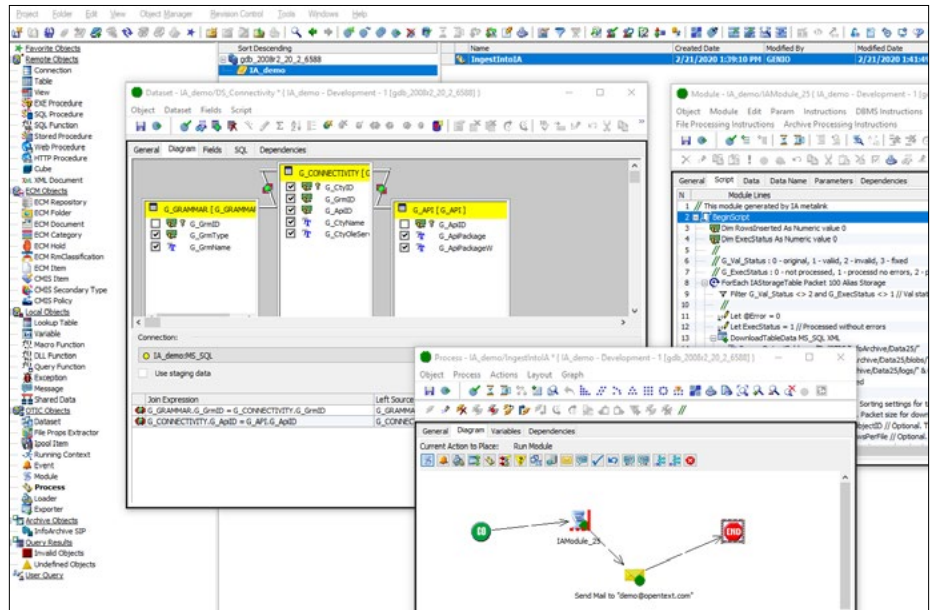
Designer is the developers' tool that is used to design and create data mappings, extracts, transformations and exchange processes

Magellan Integration Center debugger

Magellan Integration Center debugger provides several tools for controlling module or process execution steps. For example, users can set breakpoints to pause the execution of a module or process at a specific line. Users can also use the step in and step over toolbar buttons to navigate sequentially from one line to another and control when individual lines are executed.

Users can view the exceptions or console messages generated and which values are assigned after each execution step. Values can be assigned to objects to view the results of executing a module or process using those values.

Magellan Integration Center Scheduler

Magellan Integration Center Scheduler provides the ability to program process execution, on either a calendar or event basis. Scheduler also provides monitoring of process executions, full history and audit trail reporting, as well as detailed analysis of log files. Scheduler can work alongside external schedulers, such as 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

Magellan Integration Center includes the following powerful administration tools: Administration Console, Realtime Administrator and Execution/Log Viewer.

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

- Creating, initializing and connecting to repositories

- Importing and exporting repositories

- Configuring Magellan 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

Realtime Administrator is a realtime communication management application that provides an overview of Magellan Integration Center Services, as well as administration and/or execution threads running on all host machines defined for a particular repository.

It can be used to:

- View the properties of host machines defined in the active repository

- View the status of Magellan Integration Center services for host machines

- View the status of administration and execution threads

- View and stop process executions

Execution Viewer provides users with realtime monitoring of process or module executions. It lets users view or interrupt the progress of any running execution and is launched each time users manually execute a process or module in Designer or Scheduler, or view a process execution in Realtime Administrator. Users 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 Magellan Integration Center 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 or Scheduler user's guides.

Log Viewer lets users view process or module execution logs. Users can load and view the contents of the following types of execution logs:

- Process or module execution log (.xml) files on a 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 Applications

Magellan Integration Center Web Applications facilitate the execution of processes defined in Magellan Integration Center Designer. It lets users execute processes on-demand, at scheduled times or under certain conditions. It also provides history and audit trail reporting and detailed analysis of log files. Processes, events and their corresponding projects are stored in the current repository and can be accessed provided that the administrator granted the user the appropriate permissions. For information about these permissions, contact an administrator or see the installation and administration guide.

Users can also view information about particular process or module executions. Execution Viewer lets users view the details of an execution while it is in progress, whereas Log Viewer enables users to view the details of an execution after it is complete through the execution logs. From Magellan Integration Center Web Applications, users can launch OpenText Magellan Integration Center web tools.

Magellan Integration Center Web Applications offers a web interface for Realtime Administrator, Scheduler and Execution Viewer functionalities. At the same time, RESTful Web Applications enable third-party application integration with Magellan Integration Center.

opentext™

Magellan Integration Center Connectors

Magellan Integration Center includes the following groups of connector types to ensure connectivity to any source and target systems:

- RDBMS
- OLAP/multidimensional
- Text
- Tag
- iPool
- ECM Connectors
- OpenText™ InfoArchive™
- Magellan Integration Service-based ECM Connectors
- Web services and API-based

Connector types	
RDBMS	Provides ODBC-based and native access to information. These are connections to databases that store data in the form of tables, for example IBM DB2, Informix, Microsoft SQL Server, Oracle and Mainframe (as long as the customer has the proper client software).
OLAP/multidimensional	Connections to databases that store data in the form of cubes. For example, Essbase and Express.
Text	Connections to text files, including delimited, fixed length, CSV and XML located in directories. For example, AS400, JFile, JFilePro and standard flat files.
Tag	Connections to files formatted using tags. For example, XML files, web and web (HTTP).
iPool	Connections to iPools created for OpenText™ Content Server data flows. For example, document conversion service (DCS) iPool.
ECM Connectors	Connections to ECM repositories and items. For example, to support CMIS enabled systems,Content Server (Enterprise Server) and Microsoft® SharePoint®.
InfoArchive	Provide connections to InfoArchive. For example, to archive unstructured and RDBMS data.
Magellan Integration Service-based ECM Connectors	Connections leveraging Magellan Integration Service-based ECM Connectors. For example, Content Server, OpenText™ Documentum™, OpenText™ Media Management, OpenText™ Archive Center, SharePoint, FileNet P8, etc.
Web services and API-based	Magellan Integration Center enables web services, a collection of functionalities (or web methods) on a web services server, Web Method. It is a procedure that is executed when an object receives a message. Clients can use simple object access protocol (SOAP) or RESTful with JSPON payload to invoke web methods and the web services description language (WSDL) for describing web services.

For more information on how to invoke a web method on a web services server, please refer to the creating web procedures chapter in the Magellan Integration Center Designer user's guide.

These objects provide a generic way for Magellan Integration Center to integrate with generic RESTful web services. API-based connectivities include FTP, MQSeries or Kafka messaging, email or shell script.

opentext™

Magellan Integration Service with ECM Connectors

Magellan Integration Service provides a unified access interface to various content management systems based on the industry-wide CMIS 1.1 standard. The number of supported systems are defined by the available ECM Connectors. There are currently more than a dozen ECM Connectors that are implemented using the Magellan Integration Service SPI (service provider interface) framework. For more information about ECM connector versions, please refer to the Connector documentation.

Each ECM connector supports a specific ECM repository type and can be deployed as part of a Magellan Integration Service installation. Once deployed, it can be configured to access one or more ECM repositories of that type.

Supported Connectors:

- CMIS

- Dropbox

- Google Drive

- Gmail

- IBM FileNet P8

- Microsoft Exchange

- Microsoft Windows files

- Archive Center

- Content Server (CWS, REST, LAPI)

- OpenText™ Media Manager

- SharePoint

- Box

- Documentum

- OpenText™ eDOCS

Please contact OpenText Support for the latest updates on supported Connectors.

Magellan Integration Center key features

ECM support—All new Connectors rely on the CMIS model

The Magellan Integration Center 16.7 release benefits from a major shift in ECM support. In previous releases, Magellan Integration Center used its own ECM drivers and applied its own ECM model to ECM module instructions. In version 16.7, it relies extensively on OpenText Magellan Integration Service 16.7 and all new Connectors rely on the CMIS model.

As a result, Magellan Integration Center 16.7 supports new ECM (CMIS) types, provides new module instructions and relies on the joint work of Magellan Integration Center, Magellan Integration Service and SPI Connectors.

This latest version of Magellan Integration Center builds on OpenText's ongoing commitment to support organizations' information management strategy.

opentext™

InfoArchive–Native support for archiving projects

Users can process data from RDBMS objects into InfoArchive, as well as archive documents and associated metadata from a variety of ECM systems (SIP packages).

Containerization

Magellan Integration Center Engine and Web Applications support Kubernetes as a container orchestration platform for deploying the Docker images packaged in the latest release. The current Kubernetes supported version is 1.13. For users employing "kubectl" to control their Kubernetes cluster, client version 1.13, 1.14 or 1.15 can be used. Please see the official Kubernetes documentation here for more details and the latest information.

ECM objects—Native support and graphical interface for information integration projects

The ECM object support 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 (including documents, records, folders and their relevant metadata).

ECM Service Metalink: Import utility for OpenText ECM Suite objects

The Metalink allows users to connect to any ECM service and then search, browse the ECM repository and import any underlying structures for use in the MIC integration project. The new MetaLink leverages the system's APIs 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 Magellan 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 to simplify the actions a user can perform against the source or target ECM system. For example, create, update, delete, modify, add version and many other actions may now be leveraged in Magellan Integration Center projects, combining powerful existing features and transformation capabilities with the new functions relevant to the content-centric world.

Environment neutral

Magellan Integration Center is completely platform and database neutral. These features allow 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 and look up tables and custom functions for each new transformation project. At execution time, the Magellan Integration Center Engine reloads metadata-driven processes and generates the appropriate code for the target environment.

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

opentext™

Extraction

Magellan 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, Magellan Integration Center has a variety of tools to manage complex structures, such as hierarchical data dumps from mainframes or EDI files. Magellan 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

Magellan Integration Center offers multiple strategies to perform incremental extraction. Simple approaches, such as 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.

The solution can also capture data changes in most applications through its web services connectivity. By accessing the application layer of operational systems through web services, Magellan 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 and recently edited documents.

Middleware and standards support

Magellan 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, Magellan Integration Center can use FTP protocol (with or without SSL), MAPI, RSH or any external application to get access to data or push data to the target environment.

By leveraging these capabilities, Magellan Integration Center provides access to virtually any legacy system, as well as ERP, CRM and SCM systems, documents and records repositories and custom applications.

Transformation functions

Magellan 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.

It offers approximately 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 Magellan 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, Magellan Integration Center users can create macros to describe business rules. For example, a "discount" function can be calculated from a given sales amount and used everywhere across Magellan Integration Center transformations, being processed inside the solution's engine or on a remote database.

For a full description of all available functions, see the Designer user's guide PDF or Designer help.

opentext™

Support for stored procedures and SQL functions

Magellan 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 Magellan 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 Magellan 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 the remote database's 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 Magellan 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 Magellan Integration Center transformations, thereby preserving legacy investments.

Also, Magellan Integration Center can call any web service, executable, external batch or shell script for specialized transformation needs.

National language and Unicode support

Magellan Integration Center delivers comprehensive national language and Unicode support.

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

Magellan 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

Magellan 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 Magellan 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 Magellan 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 Magellan Integration Center manipulates transparently. Filters are transformed in WHERE or HAVING clauses and sorting becomes ORDER BY. Magellan Integration Center recognizes each SQL grammar, adapting itself to the source or target DBMS.

Transaction and nested transaction support

Magellan 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.

Magellan 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

Magellan 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 Magellan Integration Center.

Metadata can be defined using Designer, either by manually editing it or by importing it. Metadata can be imported from the database schema, using sample files or Magellan Integration Center MetaLinks. By using third-party solutions, such as Meta Integration, it is 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

Magellan Integration Center has multiple loading strategies, including 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.

Magellan 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. Magellan 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. Magellan 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

Magellan Integration Center provides the ability to track differences between an object definition stored in the 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, Magellan Integration Center users can automatically detect and import changes into the repository. Every change made to an object, whether located in a remote database or in Magellan Integration Center, is also stored and available for documentation purposes. This means that Magellan 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.

opentext™

Dynamic impact analysis

Magellan Integration Center provides a unique, one-of-a-kind dynamic impact analysis solution. The consistency and status of objects are automatically maintained in realtime within the design environment. The Designer provides two impact analysis modes: immediate and deferred.

In the immediate impact analysis mode, Magellan 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, Magellan Integration Center changes its status to undefined object. Objects with this status can be verified later.

This impact analysis is triggered either by changes made by the developer within the Magellan 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. Magellan Integration Center's impact analysis eliminates the need for developers to spend time manually tracking down dependencies whenever a change is made. It 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

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

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

Magellan 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.

It 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. This can include 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 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 Magellan Integration Center Repository contains all metadata used by data integration processes. This metadata is made available to users through Magellan Integration Center tools, either by querying the solution's open database repository or though XML datagrams.

Flexible scheduling

Magellan Integration Center includes a complete scheduling facility, making it possible to schedule process execution at a fixed or recurring time (daily, weekly or 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, Magellan Integration Center developers can build as complex scheduling rules as necessary.

The Scheduler is not always required, as there is support for external applications setting Magellan Integration Center variables, launching processes and receiving the result of the process. This makes it very easy to implement use of system management applications, such as 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

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

Error handling

Magellan Integration Center, through its graphical procedural language, also delivers exceptional error handling capabilities. It 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. 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 Magellan Integration Center's procedural language, which provides users with an easy and comprehensive mechanism for error

opentext™

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 Magellan Integration Center Repository. It 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 Magellan 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. Magellan Integration Center can also package the logs into different formats, such as text and XML. These can be sent as SMS messages to mobile phones or via email.

Optimizing performance and throughput

OpenText undertakes continuous testing to assess the performance of Magellan Integration Center. Multiple tests are conducted as part of the assessment, with an increasing number of Content Server application deployment instances and Magellan Integration Center engines and threads to determine the maximum document upload/download rate for the respective environment.

Disclaimer

The tests and results described in this document apply only to the OpenText configuration described herein. For testing, certification of other configurations or more information, please contact OpenText Support or the Engineering team for more information.

All tests described in this document were run on equipment located in the OpenText Performance Laboratory and were performed by the OpenText Performance Engineering Group. Note that using a configuration similar to that described in this document, or any other certified configuration, does not guarantee the results documented herein.

For any OpenText production deployment, OpenText recommends a rigorous performance evaluation of the specific environment and applications to ensure that there are no configuration or custom development bottlenecks present that hinder overall performance.

opentext™

Below are the major findings from a Magellan Integration Center assessment.

For the upload metadata and document scenario:

Content Server instances	Threads per Content Server instance	Magellan Integration Center Engines	Total Magellan Integration Center threads	Achieved documents per second
1	48	1	56	100
2	48	2	112	171
3	48	2	198	237

For the download metadata and document scenario:

Content Server instances	Threads per Content Server instance	Magellan Integration Center Engines	Total Magellan Integration Center threads	Achieved documents per second
1	48	1	56	118
2	48	2	112	165
3	48	2	198	200

For the download metadata only scenario:

Content Server instances	Threads per Content Server instance	Magellan Integration Center Engines	Total Magellan Integration Center threads	Achieved documents per second
1	48	1	56	170
2	48	2	112	350
3	48	2	198	515

Users can help optimize the solution's performance even before creating a Magellan Integration Center process by cleansing the data of .csv, .txt and xml files before loading the clean data into a RDMS system. This drastically improves read/write speed for projects that involve thousands of records.

Aside from data cleansing, Magellan 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 Magellan Integration Center Engine that can share the processing load in parallel. Multiple engines allow users 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, users can change the Magellan 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

Magellan 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, Magellan Integration Center provides the key features required to implement complex failover strategy. When triggering a process execution, users are able to define the list of the Magellan 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.

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

Performance measurement

Magellan 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 each target system is available in these logs. As a result, data integration process administrators can easily spot any potential performance problems.

Magellan 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, Magellan Integration Center delivers multiple ways of optimizing process executions, including offloading part of, or the entire process execution, on the source or target system.

For more information, see the Magellan Integration Center unique processing methodology section.

Magellan Integration Center provides multiple reading and writing strategies (single, packet and bulk) that enable users to optimize data movements based on the particular topology of their 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 Magellan Integration Center Engine. This is very simple to achieve with Magellan 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 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

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

opentext™

The solution 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. Magellan 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

Magellan 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 Magellan Integration Center offers.

### Transformation performed exclusively by the engine

Magellan 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

- The required transformations cannot be performed natively by the source/target databases

- 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

| Extract	Transform	Load
RDBMS	Magellan Integration Center Engine	RDBMS

Data access method one
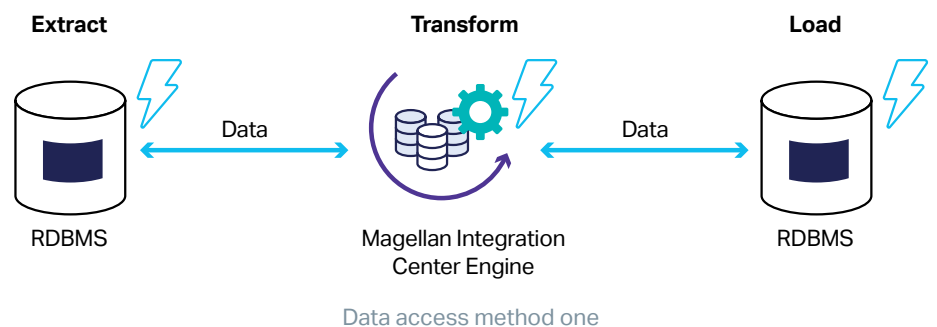
Transformation performed partially by the engine and remote databases

In this model, Magellan 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.

Transform/extract	Transform	Load/transform
RDBMS	Magellan Integration Center Engine	RDBMS

Data access method two
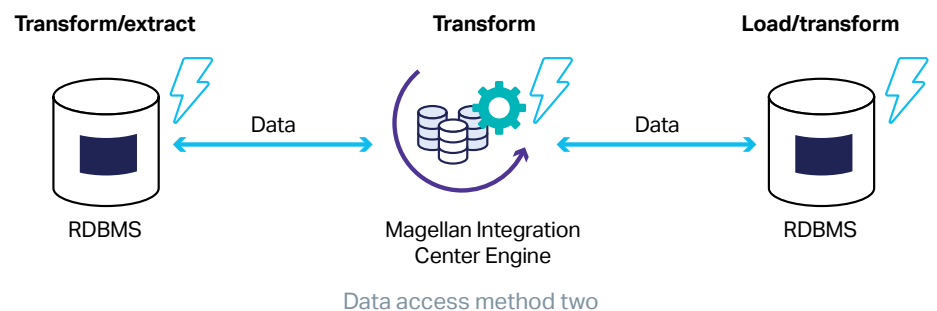
Transformations performed by the engine and remote databases

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 transport the data through a communication layer (the network). In this context, Magellan 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.

Extract	Transform	Transform
Magellan Integration Center Engine	RDBMS 1	RDBMS 1

SQL — Extract/load
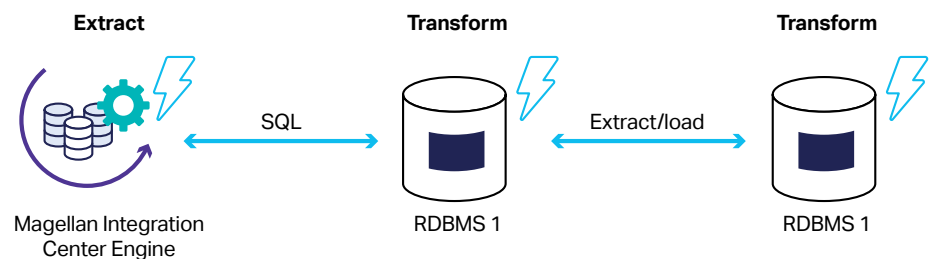
Data access method three

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, Magellan 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, Magellan 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 OpenText: A trusted link between people and information

OpenText connects people, data and applications in mission-critical environments with an award-winning suite of solutions. For more than 20 years, OpenText 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.

For more information on sales and support, please contact us.

About OpenText

OpenText, The Information Company, enables organizations to gain insight through market leading information management solutions, on-premises or in the cloud. For more information about OpenText (NASDAQ: OTEX, TSX: OTEX) visit: opentext.com.

Connect with us:

- OpenText CEO Mark Barrenechea's blog
- Twitter | LinkedIn

opentext.com/contact