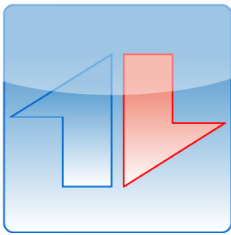


arlanis UDC

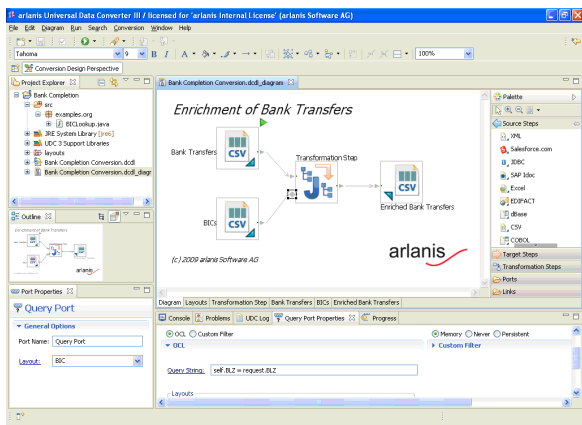
**The challenge of data integration:
Concepts and implementation**



Whitepaper, May 2009

Overview

Today, data integration is an important and essential part of any enterprise IT solution. To be competitive, it is extremely important that company data and collected information can be used in the most convenient way for the success of the enterprise. New software applications must be integrated without interruption of already running and available IT-systems.



Picture 1: arlanis UDC V3 Designer

The conversion of data is rarely a simple task: data is spread over different systems in different data formats, data has to be sorted, filtered, aggregated or split when transferred from one system to another, the content of data sources needs to be collected based on complex rules, which have to be validated during a transformation. Needless to say that data is often collected from multiple sources and must be further processed with one single application.

This whitepaper presents a first introduction and overview about the data integration topic. Inside this whitepaper; requirements, opinions and options are shown and a general solution is presented, based on software by the arlanis Software AG.

Requirements for Tools for Data Integration

Compatible to on-line running systems

Data integration will be done between running systems. Essentially this means, that it is not possible to replace or enhance existing applications while in production environments. Software for data integration must be compatible with different data sources without changing them. This is one of the most important features of modern integration tools. This feature also ensures the consistency when new applications are introduced in existing systems.

Integration in existing processes

Data integration is often carried out multiple times. For instance, periodical data transfers from a production database to a second database are necessary for analysis and reporting. Thus it must be possible to integrate data integration software in existing processes. Such integration is established between different layers in different environments. The data integration software must understand protocols like web services, JEE or native access for fulfilling customer requirements.

Interoperability with legacy systems

Converting data from and to legacy systems is a major part of the data integration story. Such systems can't be changed and mostly operate since decades. At least data must be read from such systems with the help of a data integration software. For example for the purpose of analysis or reporting. It is extremely important that the data integration tool is able to read and write various modern or older data formats.

Support of open standards and protocols

A data integration software should support standards and should be based on standards. As an example independent data transfer can be realized with XML and relational database access can be done with the help of the JDBC protocol. Beside these examples many more standards are existent in the world of information technology. With supporting those standards the data integration and data migration is not only bound to a concrete vendor data source and thus improves flexibility and use.

Support of heterogeneous platforms

With the evolving of the IT infrastructure of any company, hardware and software also become more and more heterogeneous. Thus a data integration tool should support all platforms in use and be equal in version number and functionality. An optimum approach of such software should be developed in a language that supports platform independence. Support and training costs are minimal only in heterogeneous environments.

Extensibility

Data integration software normally should come with many different adapters for the various data formats. This is one of the most important features for selecting data integration software. Although many pre-defined adaptors can make the customer's life easier, adjustments to customer systems might be needed. The integration tool must be adjustable to customer requirements. The data converting process must be changeable through scripts and a scripting language.

Performance, Stability and Scalability

The quality of data integration and data migration directly depends on performance,

stability and scalability of the used software. For best performance, conversion scripts can be verified and compiled prior to execution. Stability is important if the conversion is part of a periodical running process. The data integration software must not interrupt working in such long-running environments. Scalability with increasing workloads is an important factor. This means that any data integration solutions should also work with high volume of data.

Scripting language and the integrated development environment

A conversion consists of various mappings between input and output data. By nature such mappings range from very easy to fairly complex. The data integration software solution should support simple mappings as well as complex algorithms. Today a graphical user interface is an essential part of such a data integration software solution. For complex mappings a script language is required. This scripting language should follow a proven standard, should be debuggable and should be compilable for higher execution speed. In addition an extensible function library is required.

Expenses

Expenses for a data integration tool are definitely a decision point. But to compare only the purchase price might be a source of trouble for the future. Important facts are surrounding services of the vendor like support, training and maintenance. In integration scenarios there is often a need for customization because of the user dependent data sources.

All of them result in a special customer solution. To achieve the highest revenue all services and the tool should be provided by one single vendor or vendors business partners. In this case you have only one contact for all your needs.

Data Integration Options

Do-it-yourself solutions

Since data migration IT teams often use self-made solutions like scripts or small software projects. The main advantage is that these developers have a deep understanding about the internal data structures and data sources.

The main disadvantage is that these solutions tend to be unmanageable. There are many reasons: mainly that those projects are often not part of the core business. Other projects are more important and therefore the integration solution is out of time or control. There are unpredictable costs for improvement and maintenance of these self made solutions. In most cases new features are hard to realize and create unpredictable costs.

The advantage of a do-it-yourself solution like the knowledge of the internal structure can be saved in special integration software. In most cases there exist scripting possibilities and extension interfaces. A do-it-yourself solution is not necessarily needed.

Database or application server vendor dependent solutions

Vendor dependent solutions are often bundled with a database or application server. Those tools are mostly only usable in the context of the vendor application, i.e. the database or application server.

Because of the alignment to the vendor product the use in heterogeneous systems is often error prone. To provide solutions for heterogeneous systems or platforms is not their main focus. Thus the customer can expect perfect solutions for the vendor tools but integrating the tool in customer's grown IT landscape might be a problem.

Independent, customer-oriented solutions

Independent, customer-oriented solutions remove the disadvantages of do-it-yourself and vendor dependent solutions. Based on a robust and proven kernel these data integration solutions provide a connection to many different data sources. They also support several of the protocol standards like web services or JEE. A lot of time and money can be saved through the use of a pre-fabricated kernel system. An important characteristic of such a solution is that the customer only uses and pays what really is needed. Because of this the data integration tool is customizable for the real need.

An integrated development environment with graphical user interface is a characteristic of modern data integration tools. Within this environment conversion scripts can be created. In addition the execution and control of conversion processes will be possible.

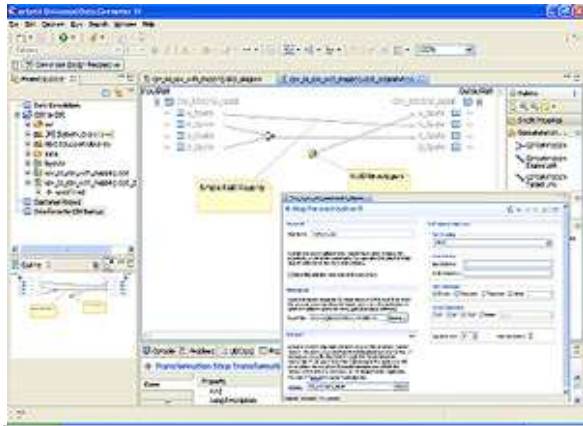
arlanis Data Integration Tools

The arlanis Software AG provides sophisticated knowledge in this specific area of data integration and data migration. This knowledge is collected in the arlanis software solutions and presented to the customer. Customized solutions can be realized with help of sales-partners and integration-partners. All products are supporting well known standards like JDBC, JMS, DIF or EDIFACT to name only a few of them. Services, seminars and trainings are complementary to the products.

arlanis UDC V3 Integration Designer

The heart of every conversion is the arlanis UDC Engine (Universal Data Converter) product which is based on a flexible and proven core. With the help of the graphical integrated development environment (picture 1 and 2) data can be easily migrated between different systems. All configuration and conversion

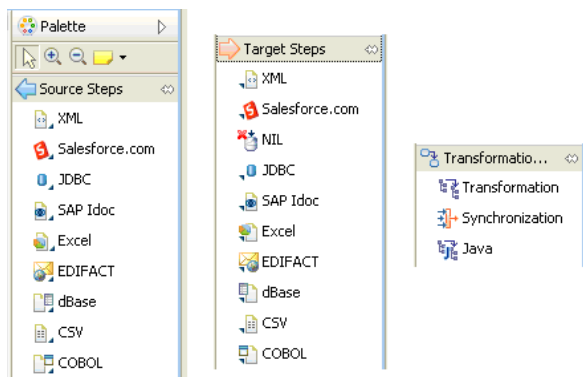
information is stored in external files and can be used for automated and unattended running.



Picture 2: Configuration of step details

The arlanis Universal Data Converter can be used to manage the task of data integration. Starting point of each conversion is the business idea of the future integration task itself. Along with the idea you can model the conversion on a higher level perspective (Picture 1). As you can see, all relevant information is shown. This means which input sources are used, which outputs are necessary or on which way the data will travel between input and output. The UDC white board shows the data transformation in overview. This drawing is fully functional (but not in deep specified yet).

Needless to say, that all conversions, presented on the white board, are printable and could be saved as an image or HTML file. This means, during the design of your conversion you will automatically receive all the artifacts which are required for your documentation.



Picture 3: A selection of the various adapters and components

Various input and output adapter are out-of-the-box provided. For instance: CSV, EDIFACT, RDBMS as well as SAP IDoc and COBOL. Those input and output adapter can be connected in different ways.

After definition of the higher level or business level of your conversion the technical data has to be assigned to each steps (see picture 2). For this task special editors are shipped within the product. On picture 2 you can see the graphical data mapping designer and, displayed below, the parameter editor for the CSV data input.

Within the Designer the conversions can be executed and tested. If everything works as desired the result is loaded into the arlanis UDC V3 Conversion Platform.

arlanis UDC V3 Integration Platform

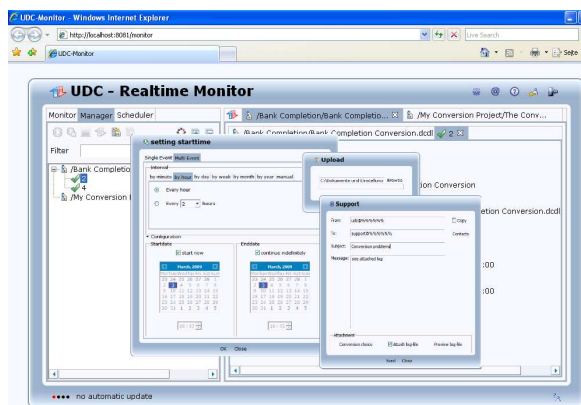
Integrated into the UDC product family is an integration platform. This means, all your conversions can be deployed to a server (with running UDC platform). With help of this platform conversions could be monitored and managed. You want to execute your conversion in periodical manner, in a specified term or once a single time? No problem, everything can be adjusted. All functionality of the UDC V3 Platform may be managed over the network. The interfaces are fully web service enabled and compatible with nearly all language standards. This means, data conversion tasks can be remote planned, started and monitored. An arlanis UDC integration job may be inserted in different kinds of business processes with help of its web service infrastructure.

Once a conversion is developed it can be easily exported from the designer and load to the platform without any additional work. The platform itself owns a web service interface for managing, viewing and scheduling integration tasks.

The arlanis UDC is the unique component between legacy architectures and „The Cloud“. Within a Platform as a Service Environment the UDC plays the role as the Integration as a Service part.

arlanis UDC V3 Monitor and Management User Interface

The monitoring and management user interface is part of the UDC product line since Version 3. With help of this, Web 2.0, compatible interface scheduling of batch processes is possible. This can happen once, at a predefined time or periodically. At each time the state of each conversion instance can be examined. Conversions can be started, stopped or monitored.



Picture 4: arlanis UDC V3 Monitor and Management User Interface

arlanis UDC V3 Runtime Engine

After exporting a conversion out of the UDC designer, there is the possibility to execute this conversion inside the platform. Sometimes it is a requirement that the conversion should only execute once, at very different times or triggered by an external scheduler.

In such cases the UDC Runtime Engine is the first choice. A conversion, described through the

Conversion Archive (CAR), is able to run inside the engine. No additional work is necessary.

arlanis Services

To achieve sustainable success an optimal use of arlanis software is needed. The arlanis software AG provides services around the products for guaranteeing this success. A comprehensive range of partners ensures high quality of these services in addition.

Consulting

If special know-how or temporary help is needed arlanis consulting is the right choice for you. This includes project planning and controlling as well as analysis and improvement of your company's data. However, consulting might be used for defining and implementing your business rules for data integration as well. For instance, if special messages must be defined in the EDIFACT standard for incoming and outgoing transfer.

Seminars

Seminars can save a lot of time and money in your projects. Taking workshops will help you to decrease initial time to understand product behaviour. In addition, technology seminars are offered for topics like SOA, architecture, web services and Java enterprise. These seminars can help to understand applications and systems from a more abstract point of view.

Custom Development

Custom development ranges from single components to complex projects. Needless to say, that the development of adaptors for special data formats is possible.

Conclusion

Data integration, migration and data transformation are a main part of every IT strategy and enterprise application development.

For example, data from different companies or departments must be integrated. Also data from different sources must be unified for analysis and reporting.

The service oriented approach offers the largest range of flexibility considering to protecting your IT-investments. The consequent splitting of the arlanis UDC product family into partial system components, allows delivering user customised solutions.

The UDC Runtime Engine and the UDC Platform can be combined to a full EAI solution. This specific solution will be enabled to develop continuously according to your enterprise requirements. To complete the UDC product family the arlanis Software AG offers the UDC Graphical Mapping- and Conversion Designer.

The arlanis Software AG provides additional services in the field of Service Oriented Architecture, Data Migration and –Conversion. This includes technology workshops and architecture- and strategy consulting. Additionally, the arlanis Software AG can support you by developing sub-projects and by the development of client-specific adaptors.

LINKS

Generell: <http://www.arlanis.com>

Produkte zur Datenintegration:

<http://www.arlanis.com/en/software-solutions.html>

Partner:

<http://www.arlanis.com/en/company/partners.html>