



**Integrating Ingres
in the Information System:
An Open Source Approach**

WHITE PAPER

Table of Contents

Ingres, a Business Open Source Database that needs Integration	3
Scenario 1: Data Migration.....	4
Scenario 2: e-Business Application Deployment.....	7
Scenario 3: ETL for Data Warehousing.....	10
Open Source Model and Advanced Capabilities Provide for Optimized Cost of Ownership	13
Conclusion & Downloads	14

Ingres, a Business Open Source Database that needs Integration

Open source database solutions that provide enterprise-level functionality, services and support are revolutionizing today's marketplace. Companies no longer need to pay excessive licensing fees for the right to use a highly scalable, highly available, and high-performance database platform. Open source allows companies to deploy flexible, less complex solutions that are not bloated with excess functionality allowing database engineers and administrators to be more productive and able to focus on more mission-critical projects. An open source platform gives enterprise customers and developers the freedom to choose the right database solution for their specific business need.

Ingres 2006 provides enterprise customers with this level of functionality and performance on an open source platform with world-class support and services. By leveraging both the affordability and flexibility of open source and the functionality and support of enterprise solutions, Ingres 2006 provides a customizable open source database solution that enterprises can trust with mission-critical data and business applications.

Today, many organizations are deploying Ingres for extremely varied uses - from e-Business to Customer Relationship Management, from Data Warehousing to Accounting, from Technical Databases to Enterprise Resource Planning. These systems are likely to be deployed in a heterogeneous environment made up of both traditional and Open Source solutions, and where internally hosted packages coexist with systems hosted under the Software as a Service (SaaS) model.

As they add new systems, very few organizations can afford the luxury of re-designing their entire information system. Any new technology deployment in an enterprise information system will thus require the newly deployed system to interoperate, in one form or another, with a number of other applications or databases. This interoperability, which is essen-

tially based on the exchange of data between systems (termed data integration), guarantees the consistency of data in the overall information system.

Consider, for example, an organization that owns a commercial ERP, and deploys an e-Business portal based on Ingres 2006. This business also implements a new SaaS CRM application such as SugarCRM and a data warehouse that runs on Ingres 2006. Such an IT environment would not be at all unusual. Data would need to be replicated between the CRM and the ERP and collected from the e-Business Portal. In addition, data from all systems in the environment would need to be loaded into the data warehouse.

This white paper describes a number of real-life scenarios for Ingres interoperability with applications and other databases and explains how Talend Open Studio, a powerful, open source data integration solution, addresses various data integration requirements related to Ingres deployments.

Scenario 1: Data Migration

This first scenario examines a data migration example. When deploying a new business application, few organizations have the luxury of starting from a clean slate. In most cases, they need to recover data from legacy systems, enrich it with additional information, and transform the data because their legacy and new data schemas do not match.

Let's have a closer look at this actual data migration challenge.

"We needed to migrate data from our legacy home-grown CRM system running on DB2 UDB to the Salesforce.com CRM, deployed in Software as a Service mode.

We used an Ingres 2006 database as a staging area to prepare all the data before the final load of the online CRM. Data transformations (to match the Salesforce.com data structures) occurred prior to loading data in the staging area. We also en-

riched this data with user information (extracted from an internal LDAP directory) and customer address geocoding, collected from a third party provider via a set of Web Services.

And because we knew that the quality of data in the legacy CRM was sometimes poor, with many duplicate records, we channeled all data through various data quality filters. We finally loaded the enriched and cleansed data in Salesforce.com.”

– Mark S., CRM Project Manager at a regional retail chain

To accomplish this migration, Mark S. and his team were able to leverage a number of advanced components and connectors of Talend Open Studio:



The tDB2Input component (DB2 connector), which provides native connectivity to DB2. This connector includes advanced metadata introspection capabilities that enabled the project staff to quickly understand the structure of the DB2 CRM database - even though all documentation for this legacy system had long ago been lost.



The tFileInputLDIF component (LDAP connector), used to retrieve user information from the company's LDAP directory and use it as an input to the mappings.



The tWebServiceInput component (Web Services connector), which enabled the connection to a provider of geocoding services - passing the customer's address as parameters and returning the geocoded address.



The tMap component, which allowed job developers to easily map data from DB2 to MySQL, add enrichments from LDAP and the geocoding Web Services, and graphically design the data transformations from the DB2 schema to the Salesforce.com data model.



The tFuzzyMatch component, which provides advanced data matching capabilities to identify duplicate records - even if the data isn't a 100% match.



The tIngresOutput component (Ingres connector) which provides native connectivity to Ingres in which the Salesforce.com data model had been duplicated (thanks to Talend Open Studio’s advanced metadata management capabilities).

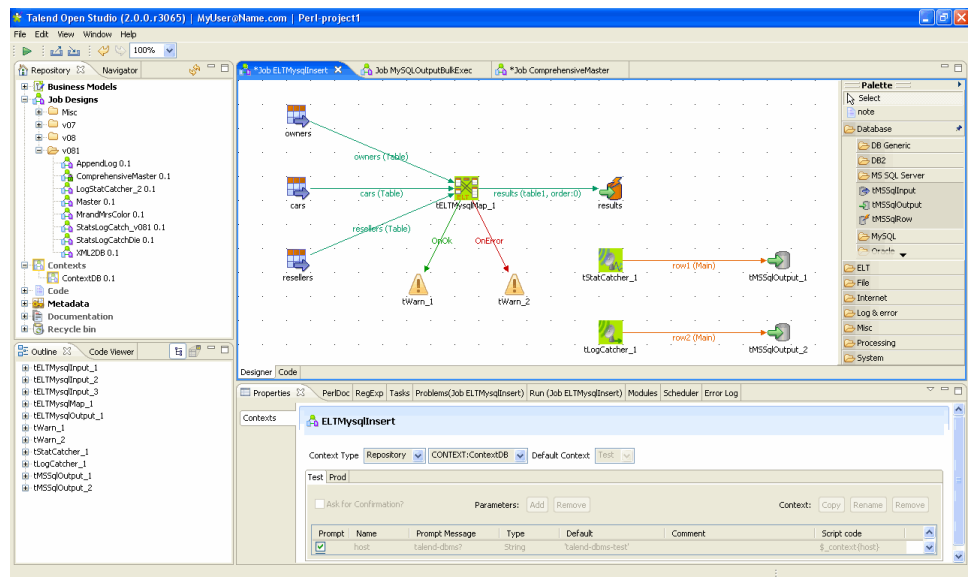


The tSalesforceOutput component (Salesforce.com business connector), which connects natively to Salesforce.com, leveraging its native Web Services APIs and SOQL language.

Mark S. concludes:

“Talend Open Studio’s intuitive drag-and-drop interface dramatically increased our productivity. We had initially estimated 35 days of work for performing this migration with custom development. Talend Open Studio allowed us to migrate our CRM in just 4 days.

And because Talend Open Studio relies on industry standard languages, our ramp-up on the product was extremely fast, and we were able to leverage existing expertise.”



Talend Open Studio’s intuitive Job Designer interface

Scenario 2: e-Business Application Deployment

This scenario studies the deployment of a new online e-Business application. This application initially needs to be loaded with reference data, typically data that pre-dates the application, and is stored in other systems, databases or even files. Then data of the online application needs to be synchronized with other systems - inside or outside the firewall.

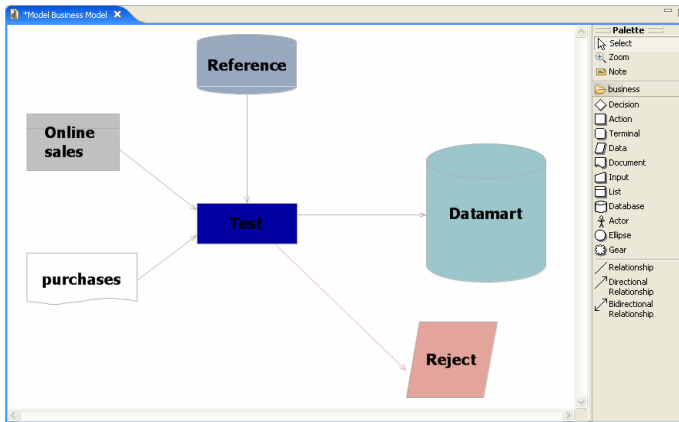
Let's have a look at this online application deployment challenge.

"The major challenge we faced when deploying our new Customer Rewards Portal was related to data integration. Initially, we needed to load existing account information, login data, and reward product information, as well as a number of reference data sets, in the underlying Ingres 2006 database. All this data existed, but was stored in a variety of formats - including SQL Server databases, flat files, and Excel spreadsheets.

Now that the Portal application is up and running, orders for reward merchandise need to be provided hourly to the warehouse's shipping application. (This is done by transferring an XML file via FTP.) Shipping bills need to be created with the courier company we use (via their Web Services). New reward points acquired by customers should be reflected in real time in the portal. (They come through a legacy messaging system.) And of course we also need to load our Ingres 2006 data warehouse with customer redemption data for analytical needs."

– Suzan K., Vice President of Customer Reward Programs at a retail bank

Because Susan K. is a business user - and not an IT staff member - her first challenge was to be able to describe the business needs of the Customer Rewards Portal. She was able to leverage the Business Modeler of Talend Open Studio to provide data integration developers with a high-level, non-technical view of the required integration processes. And because the Business Modeler stores all project-related documentation, it also enabled Susan to follow the progress of the data integration project.



Talend Open Studio's Business Modeler interface

After Susan K. had designed her requirements, IT team members were able to load the initial data sets into the portal, leveraging the following Talend Open Studio components:



The tSQLServerInput component (MS SQL Server connector), to retrieve customer account information.



The tFileInputDelimited component (File connector), to retrieve product information from a delimited flat file.



The tFileInputExcel component (Excel connector), to retrieve reference data from Excel spreadsheets.



The tFileInputRegex component (File connector), to extract reference data contained in unstructured data files.



The tMap component, to perform the transformations of data between the various sources and the portal's database.



The tIngresOutput component (Ingres connector), providing native connectivity to Ingres.

For subsequent synchronization of the portal with the other systems, specialized Talend Open Studio components were also used:



The tFileOutputXML (XML connector) component, to create the XML output for the shipping application.



The tFTP component to load this XML file on the warehouse's FTP server.



The tWebServiceInput component (Web Services connector), which invokes the courier's Web Services, passing the shipment information and getting the tracking number in return.



The tPerlFlex component, to retrieve data in real time from the legacy messaging system used by the bank. Using this component, an engineer was able to leverage existing code to communicate with this system. Further down the road, bank engineers will formalize this development and turn it into a Talend Open Studio connector, which they will contribute to the community.

And finally, for the nightly extracts from the portal to the data warehouse, the following Talend Open Studio components were used:



The tIngresInput component (Ingres connector), for extracts from the Ingres portal database.



The tMap component, to perform the transformations of data between the portal's database and the data warehouse.



The tIngresOutput component (Ingres connector), providing native connectivity to Ingres.

These complex data integration processes would have been almost impossible to implement using custom scripting. Susan K.'s team originally had envisioned using a proprietary ETL solution already in use by the bank for enterprise data warehousing. This approach, however, would have required purchasing additional tool licenses. Because of the variety of connectors involved and the number of servers for which this proprietary tool would have had to be deployed, the licensing costs would

have been very high. Furthermore, outside of the enterprise data warehousing group the bank had no expertise using the tool, and thus expensive and complex training would have been required to develop the requisite proficiency.

Susan K. reflected on the development process.

“Compared to the cost of the proprietary integration software the bank uses, the total cost of ownership of Talend Open Studio was lower by at least a factor of 20. The smooth handling of real-time data flows and native connectivity to all our sources and targets were real contributors to the significantly lower TCO. Also, the fact that Talend Open Studio leverages existing IT infrastructure greatly facilitated its deployment and created significant hardware savings.

Our initial worries about technical support have proven groundless. Unlike proprietary software, open source solutions benefit from a very helpful community, and Talend’s technical support teams have also been very effective in helping us.”

Scenario 3: ETL for Data Warehousing

This scenario examines the integration processes, which are used to load a data warehouse for analytics and are usually called ETL (Extract, Transform, Load). The data warehouse typically consolidates data from a variety of sources, such as business applications, databases, files, and external data feeds.

Beyond the data warehouse, and depending on the analytics needs, other structures such as data marts and multi-dimensional analytical structures (OLAP “cubes”) may also need to be loaded.

A data warehousing executive describes the scenario.

“Our organization is running a number of business systems, including a hosted SugarCRM (Software as a Service), as well as SQL Server and Ingres databases, which store data for a variety of back-end systems. The bulk of the data we need to process

and load (customer calls data), however, comes in positional flat files. The enterprise data warehouse we are loading runs on a DATAlegro data warehouse appliance, based on Ingres 2006. Subsequent extracts of data from DATAlegro are then performed to load a few data marts, based on MySQL and Ingres 2006.”

– Karim B., Senior Director of Data Warehousing, virtual telecommunications operator

A number of specialized components and connectors in Talend Open Studio were leveraged by this telecom operator to build their data integration scenarios.

For the extraction of data from the various sources and its transformations (aggregation, lookups, etc.), and the loading of the Netezza system, the following components are used:



The tSugarCRMInput component (SugarCRM business connector), which provides native connectivity to SugarCRM, leveraging its native Web Services APIs.



The tFileInputPositional component (File connector), to retrieve call data from the very large file.



The tIngresInput component (Ingres connector), to retrieve reference information contained in Ingres.



The tMSSqlInput component (MS SQL Server connector), to retrieve reference information contained in Microsoft SQL Server.



The tMap component, to easily map data from the positional file and SugarCRM, enrich it with Ingres and SQL Server data, and perform the transformations to the data warehouse data model.



The tIngresOutput component (Ingres connector), providing connectivity to the DATAlegro server based on Ingres 2006 for loading the data warehouse.

For the extraction of data from DATAlegro and the subsequent loading of the MySQL and Ingres 2006 data marts, the following Talend Open Studio connectors are used:



The tIngresInput component (database connector), for extracts from DATAlegro.



The tMap component, to perform the transformations of data between the data warehouse and the data marts.



The tIngresOutput component (Ingres connector), to load the data in the Ingres data marts.



The tMySQLOutput component (MySQL connector), to load the data in the MySQL data marts.

For this user of Talend Open Studio, the DATAlegro data warehouse was a new deployment, and the company did not have any legacy data integration processes to deal with. The Ingres and MySQL data marts already existed, however, and it was important to continue to load them. This demonstrates another benefit of Talend Open Studio - its native connectivity to Ingres and MySQL.

Karim B. confirms the reasons for his product choice.

“After balancing the pluses and minuses of several data integration solutions, we chose Talend Open Studio because it was able to process large data files for our calls with high performance. Compared to several engine-based tools we also evaluated, the processing of these files were 4 to 7 times faster. The native connectivity to our databases and applications was also a significant advantage.”

Open Source Model and Advanced Capabilities Provide for Optimized Cost of Ownership

The real-life examples included in this paper describe how three organizations used Talend Open Studio to address different needs: migration, operational data integration and synchronization, and data warehouse loading (ETL). There are many other cases that reflect the experience of Talend's users; some are simpler than the ones we have presented here, while others describe much more complex challenges and how they were met.

One of the key reasons for Talend Open Studio's pervasiveness in data integration projects is its open source approach, which allows it to outperform solutions based on the traditional proprietary model. It does so by supplying an open, innovative and scalable software solution that provides the flexibility to meet the data integration needs of all types of organizations.

The source code of Talend Open Studio is published under the GPL license, enabling the developer community to improve the product with enhancements that can benefit other users. Talend does not solely rely on the community for product enhancements, however; the company's own R&D team drives the product roadmap and development cycles of future versions and is committed to fixing problems as they are detected. Talend supports open standards to ensure the robustness of all its products and services. For example, Talend Open Studio is written entirely in Java, runs in Eclipse and relies on standard languages code generation: Java, Perl and SQL. This enables project teams to leverage existing expertise in these technologies.

Furthermore, Talend Open Studio's openness makes it easy for users and integrators to extend the capabilities of the product - for example, by developing and adding specific connectors.

From the deployment standpoint, Talend Open Studio processes are executed on a grid of commodity servers, leveraging existing IT

infrastructure, processing data closer to its source and facilitating the deployment.

Key benefits of Talend Open Studio include:

- A business view of the integration processes.
- Maximize productivity and ease of use through an advanced drag-and-drop interface.
- High performance and robustness with industry standard languages generation, grid deployment and support of both ETL and ELT architectures.
- Increased versatility through connectivity to all business applications, databases, files, Web Services, etc.
- Consistency and reusability of developments, as well as maintenance facilitated by the project repository.
- Ability to analyze errors and trace bottlenecks in the integration processes.

To get more information on the key features and benefits of Talend Open Studio, please visit <http://www.talend.com/products-data-integration/talend-open-studio.html>.

Conclusion

The scenarios described in this paper demonstrate how Talend Open Studio's versatility helps various applications and databases interoperate. Such interoperability is crucial to the efficient operation of any organization, ensuring the consistency of data across systems and a smooth flow of the systems.

Easy to learn, Talend Open Studio dramatically improves developer productivity. To help users architect their most complex data integration

scenarios and leverage the product to its full extent, Talend's expert consultants and partners provide a range of services.

Try Talend Open Studio and Ingres 2006 and see for yourself how these open source solutions' advanced capabilities dramatically improve both the productivity of data integration and the robustness of business applications. Products are available for free download at:

- <http://www.talend.com/download.php>
- <http://www.ingres.com/downloads/prod-cert-download.php>