

**Practical Open Source Data Integration
Case Studies & Implementation Examples**

CASE STUDIES

Table of Contents

ETAI unifies its data production architecture	4
The City of Brantford manages its municipal infrastructure	6
Eurofins chooses a universal integration solution	8
UMIT fights cancer with data integration, mining and analysis	10
MediaLab mitigates risks and accelerates data integration	12
Pokolm automates manufacturing processes	14
Eassyssur strengthens its data integration processes	16

Practical Open Source Data Integration Case Studies & Implementation Examples

Over the past few years, open source has established itself as a key component of the overall data integration market. Many organizations have adopted this model for their data integration projects. These organizations span all industries, all continents, and all company sizes. More importantly, their projects range from ETL for data warehousing or business intelligence to operational data integration, data migration, data synchronization, etc.

This document presents a few selected case studies, illustrating real-life implementations of open source data integration and its associated benefits.



ETAI

unifies its data production architecture

Talend Open Studio automates the integration of heterogeneous and disparate data and increases significantly the quality of production tasks

A data specialist

ETAI's advanced expertise and media mix apply to three industries: automotive, manufacturing and retail. The company offers diverse information solutions to these industries: databases, software, conventions, trade shows and magazines. With 400 employees, ETAI generates of 60 million euros (\$80 million) in revenue.

ETAI's business consists of producing and selling data. For example, for the automotive industry, ETAI collects raw data from suppliers (vehicles and parts manufacturers), consolidates and reconciles this data, and sells applications based on the thus-created technical database. These products are provided as either DVD-ROM or internet portal subscription to automotive dealers, repair shops, and other interested parties.

Complex data architecture

To build automotive industry applications, ETAI collects information related to vehicles (over 50,000), spare parts, parts equivalence, suitability of parts for vehicles, repair durations and techniques, costs, etc.

"This is a vast domain, and there is no standard for data," explained Philippe Bobo, Director of Software and Information System at ETAI. "We are getting disparate data repository formats, but also many paper documents which information needs to be keyed in. We have decided to keep data entry applications in the most efficient formats, to increase data entry productivity and reduce the cost of updates."

As a result, each time a new version of a database product is released (over a hundred times a year), all these repositories need to be processed: data consolidation, reconciliation and cleansing. "This process is very time consuming, both in time and resources, and always presents data quality challenges," recalled Philippe Bobo. "We used to run numerous programs in Access, Java, Python, PL/SQL, etc. Also, many steps of the process are manual and require domain expertise."

Conversely, in ETAI's other businesses (directories, magazines, trade shows, etc.), it made sense to consolidate non-technical repositories, to leverage data models that are similar between verticals.

"We looked for a data integration solution that could not only automate the assembly processes for the automotive technical databases, but also merge all data management systems for ETAI's other businesses," explained Philippe Bobo. "It was critical for us to concentrate our efforts around a single tool that would meet all our needs. Down

the road, even our internal systems will be using the same data integration environment.”

Advanced market research

“Our tool research was very methodical, and started by analyzing the company’s integration processes (both manual and automated) with the help of consultants from Apsia,” recalled Philippe Bobo. “Subsequently, several products that would conform to our primary criteria – especially the applicability to all of our businesses, and the ability to process the high data volumes of the automotive databases – were selected for a pilot project.”

Various subjects were presented to the vendors of solutions that were compatible with ETAI’s budget constraints. “Compared to proprietary solutions, Talend Open Studio offers greater flexibility,” confirmed Philippe Bobo. “Not a single step of the prototype was blocking for the tool. This is partly due to its use of industry standard languages for which component libraries are readily available, and which accelerates the ramp up of the teams.”

Talend Open Studio was also the fastest for executing a complex process of the prototype, processing 2 million records 30 to 50% faster than the other contenders and than the existing custom program.

Beyond the complexity of consolidating and reconciling data, the heterogeneity of data sources also created an important challenge. ETAI’s data sources include MySQL, DB2/400, Access, SQL Server, Oracle, Excel, XML, etc. Manufacturers’ data usually comes in complex flat files. The technical database is hosted on Oracle but is deployed on MySQL. And the company’s backend systems run on AS/400. A broad connectivity palette was hence a strong requirement.

Choosing the industry’s first open source data integration solution

After each vendor had built their prototype, ETAI compared the results and weighted each criteria based on the risk it presented for the success of the projects. Beyond the performance and feature set of each tool, the robustness of the vendor, its geographical location and its ability to assist the ETAI teams in the long run were critical factors. Talend’s solution, which best met ETAI’s criteria, was selected.

“The price of the solution was also a factor,” explained Philippe Bobo. “The traditional licensing model would have slowed down the deployment of the solution on our seven sites, for budgetary reasons. Talend’s Open Source model does not cancel all costs but it alleviates them significantly, especially in the deployment phase. It also makes them easier to predict.”

The first project that ETAI is building with Talend Open Studio simulates daily the consolidation of all data entry repositories. This simulation enables to immediately detect and fix data discrepancies. “Thanks to these daily tests of data

Key facts

100 releases/year of the database product
50,000 information points related to vehicles
2,000,000 parts equivalence

Industrialization of assembling processes for technical automobile databases
Consolidation, reconciliation and cleansing of technical data
Consolidation of non technical repositories

MySQL, DB2/400, Access, SQL Server and Oracle databases, Excel and XML files

quality, we can now maintain control over an important cause of errors. When we launch the assembly chain for the shippable database, we know it will run without error and we won’t need to edit the consolidated data.”

To get the most out of Talend Open Studio, ETAI elected to use Talend’s services offerings. IT teams first attended advanced training sessions. ETAI also subscribed to a Gold technical support contract with guaranteed response times. And finally an expert consultant from Talend visits ETAI on a regular basis to provide assistance to the development teams.

“The support Talend provides is excellent, corroborating the importance we attributed to this factor,” indicated Philippe Bobo. “The advanced expertise and the professionalism of the consultants we have worked with are an important asset for the success of our projects.”

“When we launch the assembly chain for the shippable database, we know it will run without error and we won’t need to edit the consolidated data. Talend’s Open Source model does not cancel all costs but it alleviates them significantly, especially in the deployment phase.”

Philippe Bobo
Software and IS Director



The City of Brantford manages its municipal infrastructure

Talend Open Studio leverages Web Services to provide real-time data integration between systems that manage infrastructure, work orders, utilities, etc.

A vibrant community life in a strategic location

Situated on the picturesque Grand River, the City of Brantford features a vibrant community life and a strategic location in the heart of Southern Ontario. It is just a short drive from Toronto, Hamilton, London, Niagara Falls and Buffalo, New York.

Acknowledged internationally as “The Telephone City”, Brantford marks with pride that dramatic moment in 1874 when Alexander Graham Bell invented the telephone; launching one of the world’s greatest communication industries. Three years later, Brantford was incorporated as a City with 10,000 individuals. Today, the City of Brantford continues to grow with a population of 90,192 (source: Canada 2006 Census).

Proud to be the hometown of “the Great One”, Wayne Gretzky, Brantford is a city where amateur and minor sports are an integral part of community life. Wayne’s father, Walter Gretzky, still lives here in the city he’s always called home.

A new Enterprise Asset Management system

The I.T. Services Department of the City of Brantford is responsible for the maintenance, integration and availability of the off-the-shelf “best of breed” systems and internally developed custom applications implemented by the City to service its citizens. The City has deployed the Avantis Enterprise Asset Management software solution, to manage asset and work order information. As part of this deployment, I.T. Services needed to integrate to Avantis data contained in disparate systems operated by the City to manage thousands of linear assets within the water distribution, sanitary and storm sewer networks. Most of this data is contained in a custom system, MIDS (Municipal Infrastructure Data Standard) which along with Avantis uses Oracle as the back-end DBMS.

“We had multiple data integration needs related to the Avantis deployment,” explained Scott Hall, Manager of Corporate Information for the City of Brantford. “Initially, we needed to load the system with data related to our infrastructure. Then, once the system was up and running, we needed to maintain the data in sync between the reference applications and Avantis – ensure that any updates to the source data would be propagated to Avantis.”

Historically, data integration processes used by the City of Brantford were developed as Java programs, and executed by the Windows Scheduler. However, maintenance of these processes was complex and costly. “The MIDS to Avantis Interface project was a good opportunity for us to study data integration solutions,” explained Muzammil Rajpurkar, Database Administrator with the City of Brantford. “We looked at several GUI-based data integration products, focusing specifically on their robustness and the return on investment we could get out of them.”

Robust and versatile data integration

After discarding several process-oriented integration solutions, I.T. Services settled on the data integration approach, more versatile and better suited for performing both batch loads and real-time data synchronization. “We selected Talend Open Studio for several reasons,” said Muzammil Rajpurkar. “The breadth of connectors provided by the solution allows us to connect to all our sources and targets. It is also a very robust solution, which guarantees that the synchronization processes are running reliably and that our systems are kept in sync, all the time, without any data loss. And then, the open source model limited the need for an upfront investment, an important factor in a fiscally responsible organization, financed by taxpayers’ money.”

Of special importance for I.T. Services is the native support for Web Services, built into Talend Open Studio. “The Talend processes extract data from the Oracle source systems in real-time, every time a record is created or updated,” detailed Muzammil Rajpurkar. “This data is converted to an XML format, and it is transformed and enriched through XSLT transformations to match the format required by Avantis. It is then inserted or updated in Avantis through their Web Services API. Finally, the return codes from the Web Services are analyzed to confirm the success of the operation. All these operations are performed automatically by Talend.”

In addition to the Avantis data integration project, I.T. Services have initiated other projects using Talend Open Studio. One of them consists of extracting data from the Daffron utilities management software that manages electrical and water supply in the City of Brantford. “The Daffron system is based on DB2 on the AS/400 platform, and it was important for us to extract the water meter location and contact information on the utilities networks from this system, so that they are also loaded in the Avantis platform in order to generate work orders,” clarified Muzammil Rajpurkar. “Talend Open Studio, with its native DB2/400 connectors, allowed us to develop this process very quickly.” Beyond these projects, Talend Open Studio has become the standard integration platform for all data migration and synchronization processes deployed by the City of Brantford.

“Beyond the obvious cost savings, the open source model also brings a tremendous advantage: the community that supports it,” concluded Muzammil Rajpurkar. “We have received excellent support from other users and from the Talend team via the forums and bugtracker. Open source is clearly the model of the future for software development, and Talend as a vendor has fully embraced it.”

Key facts

Management of municipal infrastructure: water distribution, sanitary and storm sewer networks
Work orders integration

Oracle, DB2/400, XML, Web Services, custom APIs
Packaged and custom applications: Avantis, Daffron, Municipal Infrastructure Assets Management

Real-time data synchronization
Data migration

“Beyond the obvious cost savings, the open source model also brings a tremendous advantage: the community that supports it. We have received excellent support from other users and from the Talend team via the forums and bugtracker.”

Muzammil Rajpurkar
Database Administrator

Eurofins

chooses a universal integration solution

Talend Integration Suite becomes the integration tool of the entire Eurofins Group, for local as well as enterprise-wide projects.

One of the first world providers of bio analysis

Founded in Nantes in 1987, Eurofins is a French and German company, figuring among the first providers of bio analysis in the world. It is specialized in analytical services for the environment as well as pharmaceutical and food industries.

The group has more than 7000 employees, it generated revenue of 460 million euros in 2007, and owns a network of over 100 laboratories in 29 countries and a portfolio of about 25,000 analytical methods to validate the traceability, the authenticity, the origin, the safety, the identity and the purity of biological substances and numerous products. The Group is permanently investing in order to provide to its customers the best possible level of quality and safety, in all the markets here they are present.

A multi-domain growth

Founded by 4 persons in 1987, Eurofins mostly bases its growth on mergers and acquisitions of laboratories with specific expertise. The challenge of integration lies in the Group's ability to share data and to manage the collaboration between laboratories specialized in various types of analyses. "We have decided to leverage the advantage of proximity: each laboratory represents an entry point for the customer and should be able to offer the range of services provided by the other laboratories of the Groups," explains Pierre-Eric Menuet, EAI Project Leader at Eurofins.

This strategy of proximity has, of course, important consequences on the management of the information systems. "Until 2005, in order to streamline processes, each new laboratory which joined Eurofins was equipped with the same information system. But we realized that this strategy could be problematic for employees as well as customers. The management has thus decided to permit diversity among production systems, by replacing only some components and by developing interfaces to integrate other applications with the HQ's information system," explains Eric Bello, Coordinator of Systems Integration at Eurofins.

Various integration needs

This change of strategy has generated tremendous integration needs. Indeed, Eurofins had to juggle with multi-cultural, multi-domain and multi-format issues. For instance, the firm had to take into account integration needs in real-time (ongoing collaboration between the information systems) and batch (load of the online portal and visualization by clients of their analyses results).

To meet those various integration needs, Eurofins had initially decided to resort to a proprietary solution. "After

Key facts

Ongoing integration of information systems
Loading of the online portal
Visualization of analysis results by customers

Replacement of Oracle Data Integrator

Speed of development, vast range of connectors,
speed of execution
Support reactivity, community dynamism

Speed of development, vast range of connectors and speed of execution

The generalization of Talend Integration Suite is in progress today, and several profits have already been highlighted by the Eurofins' integration team.

"First, even if we haven't yet reached our cruising speed, we have from the beginning benefited from the user-friendliness of the tool: this has clearly sped up the developments. We are able today to work much faster than in the past," comments Pierre-Eric Menuet. "Secondly, Talend support is very reactive, the community is very dynamic and the product evolves quickly in order to integrate the new components we need."

"Besides, the tool offers a vast range of connectors which provide time savings. Thanks to our expertise in terms of databases, we improved the MS SQL Server connectors and we have shared these improvements with the community. Those have also been included in the tool. Again, we wouldn't have been able to work so fast without the Open Source approach of Talend," concludes Eric Bello. "Finally, thanks to the component library and the shared repository, we were able to re-use some rules already defined in other projects and to gain productivity. Talend offering answers our integration goals: gain in efficiency on each integration project by using past experience."

Eurofins highlights the speed of execution of Talend Integration Suite which offers higher performances in real-time environment than Sunopsis. For all these reasons, the usage of Talend will be generalized in the entire Group for local as well as enterprise-wide projects (operational, CRM or ERP).

"Talend Integration Suite is becoming the official integration tool of the whole Eurofins group, an all-purpose solution that allows both to feed project information at the group level but also to meet many daily needs of flow automation and data integration."

Pierre-Eric Menuet
EAI Project Leader

analyzing what was offered on the market, we had chosen Sunopsis. Unfortunately, this decision occurred just before this technology was acquired by Oracle," says Pierre-Eric Menuet. "We have quickly been pushed to abandon this solution for two reasons: first, our growth model turned out to be incompatible with the new pricing model of Oracle, based on a cost per target CPU. Secondly, we realized that Oracle Data Integrator didn't quite meet our 'real-time' integration needs and we had to do some specific developments – which didn't fit our wish for standardization and would have made maintenance more complex."

"At the same time, we wanted to be able to have a clear inventory of the data flows which go through our information system, in order to better control and homogenize these flows, even though all integration projects are different," adds Eric Bello. "Our goal is both to preserve the autonomy of all the laboratories, which have their own IT departments, and to offer consistent global services (network, directory, etc.)."

Talend Open Studio to start and Talend Integration Suite to industrialize the integration process

Because he had worked for a services firm before joining Eurofins, Pierre-Eric Menuet knew the Talend Open Studio solution. "We ran some tests on limited data sets and then we used the tool in a pilot project," explains Pierre-Eric Menuet. "Several observations have helped us reach a decision. For instance, the Talend solution happened to be very use and to learn. Moreover, the Open Source approach, and the fact that the code is open, help us manage the tool. We can easily follow and participate to the evolution of the tool, even give our opinion and contribute to its development. This visibility on the tool's roadmap comes along with transparent relationships with the software vendor. Besides, two days of training helped us become completely proficient."

Today, two laboratories have been integrated thanks to Talend's solution and 7 or 8 projects are planned for the coming year.

In a subsequent step, Eurofins has decided to go into high gear and has implemented Talend Integration Suite: "We needed a collaborative layer because many developers often work together on the same projects," describes Eric Bello. "We also wanted to develop a shared and centralized repository in order to inventory our data flows and to run impact analysis, which are used for the management of change. In view of this enterprise development, Talend Integration Suite brings additional services. This solution is becoming the official integration tool of the whole Eurofins group, an all-purpose solution that allows both to feed project information at the group level but also to meet many daily needs of flow automation and data integration."

UMIT

fights cancer with data integration, mining and analysis

Talend Open Studio helps UMIT do data processing and statistical analysis

A University specialized in cancer treatment

The University for Health Sciences, Medical Informatics and Technology (UMIT), based in Hall (Austria) is a key participant to the IMGuS project. A life science data warehouse system to enable systems biology in prostate cancer is managed by UMIT/biomed. In coordination with five other research groups located in Germany and Austria, UMIT manages the technical infrastructure and the data warehouse part of the project.

The IMGuS project

Prostate cancer is the most frequent tumor type in male individuals and the second most frequent cause of male death. The IMGuS project aims at the application of high throughput data processing to identify molecular signatures allowing the stratification of patients who are susceptible to curative treatment of prostate cancer and who need treatment. Therefore, patient samples that are already available at the University Clinic of Urology, University of Innsbruck, are used. The established technology platforms of the different partners will be utilized in order to generate complementary genomic, proteomic, and metabolomic data using samples from healthy controls, low risk and high risk prostate cancer patients. The results for both groups will be analyzed using statistical and data mining methods in order to discover molecular signatures for novel therapy and prediction approaches. The generated data are integrated and persistently stored in a clinical data warehouse, which is developed and managed by the Institute of Biomedical Engineering at UMIT.

Data processing is key to cancer research

"A large part of cancer research today consists of data processing and statistical analysis," explains Dr. Bernhard Tilg, Professor and Board Member at UMIT Institute of Biomedical Engineering. "The goal of these projects is to identify molecular signatures associated with certain types of tumors, so that efficient and non-intrusive diagnostic mechanisms can be designed. Some cancer treatments have high success rates, when the disease is diagnosed in time, but the key problem remains the diagnostic."

"We use data integration to combine several different data sources to perform advanced analysis and statistics on the

Key facts

Multi-stage loading of a prostate cancer research data warehouse

Data providers spanning five universities and research centers, plus medical publications, legacy systems, reference medical databases (genome cartography, etc.)

Various formats: CSV files, high resolution images, RDBMS, XML data, Web Services etc. PostgreSQL target

whole set," clarifies Dr. Bernhard Pfeifer, Associate Professor at UMIT Institute of Biomedical Engineering. "And because of the amount of data the high throughput sources create, an automated approach is mandatory. We looked at a number of data integration solutions, both proprietary and open source, and settled on Talend's solutions because of their flexibility, openness, and high performance."

Indeed, it is critical for the project that the chosen data integration solution not only work with all data sources, but also be able to integrate specific data processing approaches – for example since various medical devices deliver data in different formats, pre-processing of this data is required. Talend's open architecture allowed UMIT to develop specific components to access and process this data.

The PostgreSQL-based LINDA data warehouse, which is the basis for the statistical analysis of the IMGuS project data, is loaded in two stages. The first stage, dubbed Electronic Data Capture or EDC, centralizes data from all the different sources: patient samples, reference medical data, genome cartography, etc. "The complexity of the Electronic Data Capture stage is very high," explains Bernhard Pfeifer. "Not only are the data providers very diverse – five different universities and research centers – but the formats vary vastly: very large CSV files, high resolution images, RDBMS, XML data, etc."

Administrative data is also loaded at this stage: patient demographics, information about the biological source a certain sample comes from (tissue, serum...), or information on the data source the information is stored in.

The second loading stage reconciles, transforms, cleanses and enriches the data contained in the EDC and loads the LINDA data warehouse. "At this stage, we need to bring in reference data from external providers – medical publications, legacy systems, reference medical databases. "Talend's native support of Web Services and XML brings tremendous value to the project," explains Bernhard Pfeifer. "It allows very easily to parse and to cross reference external sources of data, reducing greatly the time it would otherwise take to enrich the data warehouse."

The frequent refresh of the data warehouse, performed every night, ensures that researchers can use ad-hoc query and data mining tools, and apply advanced statistical models, to extract data relevant for their research.

"UMIT/biomed relies entirely on Talend's solutions for all data integration needs," concludes Bernhard Pfeifer. "We have high hopes that the IMGuS project will contribute to the reduction of prostate cancer mortality rates, and data integration is a critical part of this project. Talend is helping us save lives!"

“UMIT/biomed relies entirely on Talend's solutions for all data integration needs. We have high hopes that the IMGuS project will contribute to the reduction of prostate cancer mortality rates, and data integration is a critical part of this project. Talend is helping us save lives!”

Dr. Bernhard Pfeifer
Associate Professor



MediaLab

mitigates risks and accelerates data integration

Talend Open Studio allows integrating into SAP data from nine sources in six different formats.

A “Research Aggregator”, made in New Zealand

MediaLab is an independent research and development specialist that leverages a unique partnership model: the company connects commercial companies and researchers to develop new technologies that meet specific and complex business needs. After identifying such a need, MediaLab partners with commercial organizations, universities and governments agencies, and deploys project teams to develop innovative solutions.

Mostly involved in the Information & Communications Technology industry, MediaLab is a fast growing New Zealand-based company. Created in 2005, it manages a range of research projects and develops its own products, especially for life cycle management (optimization and modeling) and resource management. With a staff of 16 (up from 5 in 2006), MediaLab partners with companies such as Telecom New Zealand, Alcatel Lucent, OpenCloud, Kordia and CityLink. In addition, the company offers an extended range of professional services, varying from the optimization of technology investments and operations research to consultancy and network testing.

Nine data sources

As part of a project that was recently implemented for a major telecommunications operator in the Asia-Pacific region, MediaLab was looking for a solution that could handle complex integration challenges, in a very short timeframe.

“Our team of statisticians uses an internally-developed modeling software to compute data and provide information for management decision support. The results of this processing needed to be integrated in the SAP system of our client. This modeling software is linked to 9 data sources and outputs needs to reach two different stakeholders every month, one of them being SAP,” explains Florent Mara, Technology Analyst at MediaLab. “We had developed internally a data integration solution based on Perl scripts, but the development was too long and wasn’t reliable enough for this highly strategic type of project. In brief, our solution did not offer sufficient stability, functionality and scalability. We started to look for a robust data integration solution that would better meet our needs.”

The timeframe to develop this project was rather short: “We only had three months to develop our integration solution – actually, more like 5 to 6 weeks when accounting for tests and deployment. Custom scripts development could not possibly have worked, both in terms of timing and in terms of features,” continues Florent Mara. “We thus looked for a solution that would be capable of reusing the scripts we had developed previously, and managing all the complex parameters we handle. We also needed an important level of flexibility, to be able to meet changing requirements during the development and testing phases. Talend Open Studio was not only meeting these needs, but we could also extend it with our own components.”

Robustness and performance

Before the execution of the statistical computation, MediaLab needs to manipulate data from these 9 sources, in 6 different formats: convert format, cleanse the data, load it in a database, and then launch the integration jobs and calculations. The resulting data is then integrated into SAP via an XML format, after another format conversion.

“We needed the ability to build custom connectors, in order to smoothly combine the integration processes with our statistical software and produce the required SAP output. After evaluating several integration solutions available on the market, we realized that Talend Open Studio was the only tool that would allow us to reuse existing scripts and develop our own components. In addition, it was bringing us the major advantages of open source: openness, customization, and no licensing costs,” explains Florent Mara. “The other solutions we envisioned did not meet our criteria: they were too expensive, too inflexible, or too closed.”

Time saving and risk mitigation

MediaLab is currently in the process of testing their developments, and will be installing them soon into their production environment. Yet, the first benefits of the chosen solution can already be felt, especially with regards to time savings: “Even though the integration of all systems isn’t completed yet, Talend Open Studio helped us decrease by about half the monthly development costs, which translated into significant savings for our customer,” rejoices Florent Mara. “In addition, because processes developed with Talend Open Studio are more robust than the ones we used to build internally, we are mitigating the risks linked to the project, to project changes and to its very tight timeframes – an important factor in the telecommunications industry.”

“Given our initial successes and our fast rampup with the solution, we have now incorporated it to another aspect of our project,” pursues Florent Mara. “Thanks to its automation capabilities, Talend Open Studio now manages and runs part of our statistical calculations, which further increases the reliability and time savings.”

Talend Open Studio offers MediaLab greater maintenance flexibility: “We can easily add or modify data sources and create quickly new transformation and integration scripts. Talend Open Studio is becoming MediaLab’s integration solution: we have started to use it for other projects, such as Network performance measurement,” concludes Florent Mara. “With the time saving and the increased reliability brought by Talend’s tools, we are reducing our development costs and reinforcing our credibility with our clients and prospects. We are currently initiating important projects with major global players in the telecommunications field, in which Talend’s solutions might play a strategic role.”

Key facts

9 different sources, 6 different formats Integration into SAP

Format conversion, data cleansing, database loading
Statistical calculations

Short project timeframe
Need for flexibility and extensibility

“We needed the ability to build custom connectors. (...) We realized that Talend Open Studio was the only tool that would allow us to reuse existing scripts and develop our own components. In addition, it was bringing us the major advantages of open source: openness, customization, and no licensing costs.”

Florent Marat
Technology Analyst

Pokolm

automates manufacturing processes

Talend Open Studio improves data quality, sharply increasing productivity and customer satisfaction

A leader in precision tooling

Founded in 1994, Pokolm Frästechnik GmbH is a leader in precision tooling (milling cutter bodies, inserts and arbor, solid carbide tooling) for mould making, die shops, machine building, model making and tool making. Pokolm's clients include automotive, aeronautics and space companies, and mechanical construction. Markets served by Pokolm include Europe, the USA, Russia, Australia, China, etc.

Half a million references

Precision tooling is a highly complex domain, with a high number of permutations (shape, size, precision, material, carbide grade, type of coating, type of mount, etc.). In order to decrease delivery times and to ensure the accuracy of each order, it was imperative for Pokolm to process smoothly all the ordering, manufacturing and inventory data. This problem had become even more acute with the deployment of an online store.

"Our customers and distributors expect a fast turnaround time – a single missing milling cutter can bring a production shop to a halt while they wait for the part to be delivered," explained Maik Böttcher, IT manager at Pokolm. "However, there are over 500,000 different combinations for the tooling we provide, and we cannot maintain an exhaustive inventory of all parts. We needed to streamline the processing of order data, and existing inventory in order to speed up our manufacturing and delivery process."

Data processed to automate the manufacturing process come a variety of sources: the online store, orders placed off-line by Pokolm's sales representatives and network of distributors, the Sage KHK production shop management system, a Crossbase catalog system, MS SQL Server and Oracle databases, and several custom applications developed in-house.

Open source data integration

"The complexity of the data sources and formats, and the need for fast data processing, drove us toward automating as much of the processing as we could," said Maik Böttcher. "The price tags of the traditional data integration solutions were way too high for a company like ours, so we decided

to look at open source solutions. We ended up running tests with both Talend Open Studio and Pentaho Data Integrator (ex-Kettle) but quickly found out that Talend Open Studio offers much broader capabilities and a far greater ease of use.”

Without a data integration solution, most of the developments would have been performed manually, with very little possibility of reusing and sharing code between the different parts of the system. “Talend Open Studio has enabled us to centralize and streamline the development of our data integration processes,” explains Maik Böttcher. “The tool’s visual presentation of sources, targets and integration jobs makes it very easy to understand the flow of data and the business rules that apply. The productivity of our development team has been increased at least by a factor of ten, and the resulting processes are far richer than what we would have been able to achieve with manual developments.”

Pokolm’s developers have quickly become involved members of the Talend community. “The collaborative spirit among Talend users is excellent, and the forum on talendforge.org is a great example of this collaborative work,” says Maik Böttcher. “Not only is this forum already very rich with information, but many community members are actively involved in providing answers and help to newbie users – who in turn quickly become contributors. We have also used the Wiki and the BugTracker, and the documentation provided by Talend is of excellent quality.”

A sharp increase in data quality

Pokolm has been able to automate in a record time the flowing of data between the business applications and the manufacturing systems, and the benefits are already obvious for all parties involved. “Not only are orders processed in real-time, which makes the planning of the manufacturing jobs a lot easier to manage, but there is also a tremendous increase in the quality of the data – the machine operators no longer need to go hunting for missing information that used to cause huge delays and losses of productivity,” clarifies Maik Böttcher. “We have also reduced the number of products that were returned because of non-compliance with the order to virtually zero.”

In less than three months, Pokolm’s development team has developed and deployed 67 new data integration processes. Every morning, the management of the company is receiving a report generated by Talend that shows the status of the manufacturing orders. And Pokolm is already finding other ways to leverage their expertise with the tool: mailing list management, direct marketing, trade show leads management, etc.

Key facts

Processing of ordering, manufacturing and inventory data for precision tooling
500,000 possible inventory combination

Oracle, MS SQL Server, Sage KHK, Crossbase, custom applications
Complex data sources and formats

Ten-fold increase in productivity
Real-time processing of orders
Data and product quality increase
67 new data integration processes developed in 3 months
Active community involvement

“Since we started to implement Talend’s solutions, we have seen tangible business benefits: sharp increase in data quality leading to higher productivity and higher customer satisfaction; as well as technical benefits: increased productivity of the development team and facilitated teamwork. Talend Open Studio has become our standard for all of Pokolm’s data integration needs,” concludes Maik Böttcher.

“*Since we started to implement Talend’s solutions, we have seen tangible business benefits: sharp increase in data quality leading to higher productivity and higher customer satisfaction; as well as technical benefits: increased productivity of the development team and facilitated teamwork. Talend Open Studio has become our standard for all of Pokolm’s data integration needs.*”

Maik Böttcher
IT Manager



Easyssur

strengthens its data integration processes

Talend Open Studio allows to streamline all data extraction and integration processes.

A Web-based insurance broker

A pioneer of online insurance sales, Easyssur is part of April Group, a leading insurance broker in France, widely recognized for its business performance, its quality of service and its innovative products.

Established in 1991, Easyssur builds targeted insurance offerings through several commercial brands:

- www.redcoxx.fr (for all customers)
- www.bnrasurances.fr
(for customers with bad driving records)
- www.alloassurances.com
(independent workers)

By using Web, telephone or video-conference technologies to handle customer relationships, Easyssur reinvents both proximity and personalized service.

A modern company that leverages new technologies, Easyssur has reduced its structural costs, for the greatest benefits of its customers: its call center is located in France to leverage trained personnel, qualified to manage French contracts and to protect the interests of clients when they submit a claim; management of contracts is available 24x7 through Web sites; each client has access to a personal advisor.

Data extraction and integration needs

In order to better analyze its business and to improve its predictions, Easyssur has developed a business intelligence platform: "We extract data from several applications – our back office systems: an ERP for brokers and our CRM/Rate Calculator, both of them using MS SQL Server databases; Excel spreadsheets; our Sage accounting software – in order to integrate and consolidate them in our Data Warehouse and to perform calculations, comparisons and predictions," explains Arnaud Fournier, co-Director at Easyssur in charge of Internet Development. "We can get a detailed snapshot of our activity – new business, contract cancellations, claims, etc. – as well as of our accounting and finance data, in order to better analyze it."

In parallel, Easyssur needs to provide monthly to its partner insurance companies information about the contracts: contact info, levels of coverage, claims, renewals, etc. "Easyssur's main advantage is that we can provide economical services, without decreasing the coverage. The company uses cost tools that help us to reduce costs. In every aspect of our business – customer relationship, claims, relationship with insurance companies – we try to be creative and to capitalize on new technologies to offer our clients the best service, at the lowest cost," explains Clément Chaumette, IT analyst at Easyssur. "We were looking for a reliable and economical way to perform our data extractions. Because we were using MS SQL Server 2000 databases, we first looked at what Microsoft had to offer. But a limited functional coverage, and a lack of ease of use, drove us to look at Open Source solutions. Among them, two products appeared interesting: Pentaho Data Integrator (ex-Kettle) and Talend Open Studio. The first one presented a significant inconvenience: the scripting language used for custom data

manipulations was JavaScript, a pretty limited language originally created for the Web and not as efficient as Java or Perl for processing massive data volumes. Conversely, we had significant Java expertise – Talend with Java jobs was the ideal solution for us.”

Clément Chaumette adds: “Another determining factor in favor of Talend Open Studio was its debug mode, and the ability to see transparently the Java code generated. This has helped us to greatly reduce development cycles. Furthermore, Talend’s solutions prove their maturity with frequent updates, always increasing feature set, an active community, and a growing ecosystem of partners. The ergonomics of the solution is a key advantage (the tMap component for joining several sources is a real killer!). Beyond these advantages, Talend Open Studio offer economical advantages: no license cost and quick results obtained after only a short training.”

User friendly, stable and flexible

“Our project started in January 2007 and a prototype was ready the following month. A lot of work had to be done around data semantics and the design of the data warehouse – the 80/20 rule for BI applied: 80% for the work on data and loading processes, and 20% for analysis. We have now a pretty strong database to integrate all of the enterprise’s data,” declared Clément Chaumette. “We never could have created manually the thousands of lines of code that Talend Open Studio generated automatically. This allowed us to spend more time on data than on ‘plumbing’ and save us lots of valuable time.”

According to Clément Chaumette, Talend Open Studio brings Easyssur three key advantages: “The first advantage is maintainability. The solution is easy to understand and to master, the environment is intuitive and its appropriation requires minimum training. Secondly, Talend Open Studio is very stable, which is a must have for a production environment. And finally, Talend Open Studio offers maximum flexibility: our company is growing very quickly, and we need to be able to meet immediately new requirements and thus to be able to modify quickly our integration processes.”

Easyssur also uses Talend Open Studio for implementing MDM (Master Data Management) in order to control the quality and consistency of data between several applications. “Some information is sometimes entered multiple times in our different systems (back office and CRM for example),” explains Clément Chaumette. “With automated processes, we can track duplicates and anomalies. Thanks to Talend, we make our data more reliable – continuously.”

After close to a year of using Talend Open Studio, Easyssur is looking at several directions for expanding its use. First, they are looking at increasing the number of information sources handled by Talend Open Studio, including integration of data from the call center. They are also studying the possibility to leverage it for recouping data from their

Key facts

Data Warehouse loading
Master Data Management
Integration between operational systems
(call center, Web site, CRM, etc.)
Data quality

MS SQL Server database
Excel files
Sage accounting system
Active Directory

incidents manager (a modified version of the Mantis open source solution). Another possible application would be to automatically update the company’s Active Directory.

“At some point, our CRM will communicate in real-time with our back office. Today, we are manually transferring some data entered by our customers on the Web site; soon this process will be automated, which will save us lots of time. Originally a tool we were using only for a few specific processes, Talend Open Studio has grown on us to become our single integration platform, our ‘Swiss Army knife’ for data manipulation. When a new data-related need arises, we immediately look for the way to implement it using Talend...” concludes Clément Chaumette.

“Originally a tool we were using only for a few specific processes, Talend Open Studio has grown on us to become our single integration platform (...). This allows us to spend more time on data than on ‘plumbing’ and save us lots of valuable time.”

Clément Chaumette
IT Analyst

For more information on Talend's open source data integration solutions: <http://www.talend.com>
Contact information for your region: <http://www.talend.com/contact>

© 2008 Talend Inc. All rights reserved.

Java is a registered trademark of Sun Microsystems, Linux is a registered trademark of Linus Torvalds. All other brand names or products referenced herein are acknowledged to be trademarks or registered trademarks of their respective owners.