

Data Migration

Introduction

Data migration is a subset of the data integration market. However, it is a very large subset. According to our research the market for data migration (including services and people as well as products) was significantly in excess of \$5bn in 2007. In our eyes that makes data migration a valid sub-market in its own right, just as ETL is a sub-market in its own right.

Historically, data migration has been accomplished using conventional data integration and data quality tools, or by means of hand coding. However, research undertaken by the Standish Group in 1999 found that more than 80% of data migration projects ran over time or budget. When Bloor Research conducted a survey on the same topic in 2007 we found that that figure had not changed. In other words, seven years of general-purpose enhancements to the tools available made no difference in the success rate of such projects.

One of the conclusions we can draw from this is that traditional tools, on their own, are not enough. We need data migration methodologies, we need experienced professionals, and we need special-purpose tools; or at least add-ons to conventional tools that have been specifically designed to support data migration.

However, the tools vendors (with some notable exceptions) have been relatively slow to recognise the importance of data migration and that it has specialised requirements over and above those that apply to data integration in general. As a result the data migration market is immature in terms of technology.

We could have included every data integration vendor in this Market Update. However, a Market Update on Data Integration Platforms is being published by Bloor Research alongside this one so we have here concentrated on those vendors that specialise in data migration, or have at least started to focus on data migration as a specialist sector.

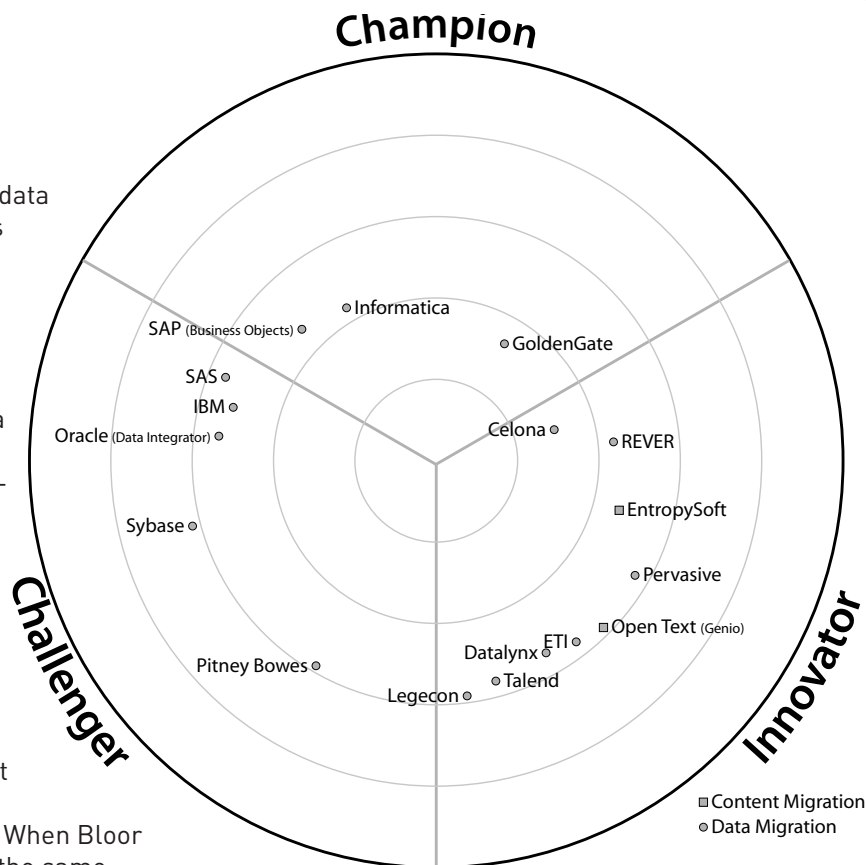


Figure 1: Bullseye assessment – the nearer the centre, the better

Key market issues

Richard Feynman, the Nobel award-winning physicist, was once asked what you need to know to be a physicist. His reply was that there were three things: “mathematics, mathematics and mathematics”. The same could be applied to data migration in that there are the three things that you need to do: “profile your data, profile your data and profile your data”. You need to know where your data is (if necessary going through a discovery exercise) and you need to understand it. If you understand your data fully (which is the role of profiling) then the remainder of any data migration project should be relatively straightforward.

Thus the first requirement of any data migration environment is that data profiling must be included either directly or via partnership. The same applies to data cleansing: there is little point in starting a new system running on poor quality data; and, in any case, if you load bad data into your new system it may break it if the rules for data differ in the new system compared to the old one (which they more or less certainly will do).

In terms of actually moving the data, traditional techniques assume that you are moving database tables (or their equivalent in a non-relational environment). However, a better approach is to migrate business entities (for example, a customer together with all of her orders). Amongst other things, this is easier to understand at a business level, enables incremental migration, facilitates parallel running and makes failback easier to implement. Unfortunately, we know of only two general-purpose vendors that explicitly support this sort of functionality today.

The other big issue when it comes to data migration is how you perform cutover to the new system. Traditionally, this was a one-time “big bang” task, typically performed over a weekend. However, this is not practical for some mission critical 24 x 7 applications, which cannot be taken down at any time. Moreover, experiences such as those of Terminal 5 at Heathrow illustrate how dangerous such an approach can be.

The alternative to big bang is known as incremental migration and, in this scenario, you effectively run off both the old system and the new one in parallel. However, in order to tell user applications where to retrieve the data, the classical approach is to insert flags into the source system(s). In other words this is an intrusive approach where one would prefer to use a method that was non-intrusive. Software that enables such a non-intrusive approach is now starting to become available and it represents a significant advance over previous offerings.

Also important in supporting what are becoming known as “zero-downtime” migrations is bi-directional synchronisation between the source and target systems. This allows the two systems to be run in parallel even after the migration is technically complete so that, for example, you might run the two systems together for a number of months until you feel completely happy about turning off the original system. This peace of mind can be further enhanced by failback facilities that automatically revert to the original system in the case of any sort of failure in the new environment.

Vendor landscape

As we have said, the major vendors have largely ignored data migration as an environment requiring special technology and facilities. The one area where there is a general exception to this rule is the recognition that it is important to have a deep understanding of the source and target

application environments. Thus leading vendors have considerable expertise in understanding SAP, Oracle, Siebel and other such applications. However, while useful from a migration standpoint this is not specific to migrations but is of general use for data integration.

The only leading data integration vendor to have specifically introduced functionality for data migration is Informatica, which announced its Data Migration Assistant earlier in 2008. It thereby has a narrow lead in this market amongst the major suppliers. Otherwise, while SAP (Business Objects Data Integrator) and Oracle (Data Integrator) are both obviously targeting their respective user communities in terms of migrations, there is a dearth of innovation from the more well-known players. Business Objects is, however, putting a major emphasis on facilitating SAP migrations and, in this regard, we expect significant new facilities to emerge from Business Objects as it works with its SAP counterparts.

There are also companies such as Datalynx and Legecon that are specialist data migration consultancies that have built their own tools but these mirror those of the traditional data integration and data quality vendors.

In order to find specially targeted data migration tools we must therefore look elsewhere, amongst smaller and more innovative companies including GoldenGate, Celona and REVER. All of these offer bi-directional synchronisation and GoldenGate and Celona both provide failback as well, though they can all be used in conventional big bang style environments. Celona and REVER also offer the ability to migrate business entities rather than just tables. However, Celona only operates in the UK and specifically targets the telecommunications sector (though it may otherwise be available via systems integrators), while REVER, which uses a model-driven approach to migrations, is, again, geographically limited (this time to Belgium and neighbouring countries and it is only available from systems integrators).

Of these three, only GoldenGate offers a significant international presence with offices in several European countries, Singapore and India in addition to the United States. It also has several major systems integrators as partners. In addition to its data migration technology per se the company also has a tool called Veridata that allows you to compare your source and target data on the fly, which is particularly useful during both testing and parallel running. Valiance Partners has

a similar product called TRUcompare that is used for testing and documenting the accuracy of migrations, especially where content has to be migrated as well as data, and where compliance requirements mean that you have to be able to prove the accuracy of your migration.

It is likely that the major vendors will start to develop the sort of facilities offered by Celona and GoldenGate in due course. However, we have only seen the very earliest signs of this and it is likely to be a while before the big boys can offer the sort of bi-directional and failback capabilities provided by these suppliers.

Also noteworthy are EntropySoft and Open Text's Genio because both of these offer content migration. That is, they understand the metadata that content management vendors wrap around the documents that they store, thus enabling easier migration (and federation) between these sources. Genio is otherwise a conventional ETL tool whereas EntropySoft (which is based in France) specialises specifically in the content market where, from a technical point of view, it is the clear market leader. The product would not be suitable, however, for more general data migrations.

Note that we are here reporting on general-purpose data migration products and not those that are limited to a particular target environment, of which the most notable are Microsoft SQL Server Integration Services (SSIS) and Oracle Warehouse Builder. While it will not be applicable any wider than it is now, SSIS will be significantly enhanced with SQL Server 2008, which will include data profiling capabilities. Oracle Data Integrator, which is based on its Sunopsis acquisition plus integration with Trillium's data quality tools, is included in this Update.

Summary and conclusions

If you have a data migration project then your first choice is whether to do this in-house or whether it might be better to contract with a specialist data migration consultancy or systems house. Some of these will be prepared to offer you a fixed price contract with a fixed time for delivery though there will be the proviso that you have discovered all of the relevant data sources and that the consultancy has thoroughly profiled all of these (which illustrates the point about profiling: that a fixed price can be offered once profiling is complete). This may be an attractive option.

Alternatively, if you opt to do the data migration yourself, and assuming that you are going to use appropriate tools rather than risk hand-coding, then you have a choice between one of the traditional ETL platforms and one of the newer vendors offering more innovative solutions. With the exception of GoldenGate this may limit your choice (for geographical reasons) if you want to opt for a specialised data migration solution.

That said, if you need to have bi-directional synchronisation, failback and/or zero-downtime migration then you really have little choice but to go to one of GoldenGate, Celona and (if failback is not required) REVER or their partners. If big bang will suffice or you don't mind an intrusive approach to incremental migration then you will have a wider choice of providers.

*Philip Howard
Data Migration
June 2008*



Suite 4, Town Hall,
86 Watling Street East
TOWCESTER,
Northamptonshire,
NN12 6BS, United Kingdom

Tel: +44 (0)870 345 9911
Fax: +44 (0)870 345 9922

Web: www.bloor-research.com
email: info@bloor-research.com