

## How come Syncsort is so fast and what does that mean?

Published: 1st March 2011

**S**yncsort has always had a reputation for high performance but historically it paid no attention to marketing and not much to sales so it was only hearsay that suggested it was a product worth looking at. That has changed recently with the company coming under new management, so I've been taking a look under the covers to see what is different about DMExpress, which is the company's data integration product, and how come it holds the ETL record (by a big margin) for loading data into a warehouse.



There are a number of important points. The first is that the technology was originally developed to run on mainframes back in the 1970s. At that time you needed to be parsimonious with your use of resources and you had to squeeze every bit of performance out of whatever little memory you could access. That frugality has been carried through into today's product (which runs on all leading platforms). For example, by default, DMExpress uses 15% of available memory on whatever platform it is running. Other tools in this space typically take all available memory. Similarly, the product directly connects to the disk drives on the source and target rather than going through the operating system, thus cutting out any overhead associated with that process.

Secondly, and the biggest thing that makes Syncsort unique in the data integration space, is that the product is built around an optimiser in much the same way that databases have an optimiser. Of course, this only makes sense if you have lots of different ways of achieving the same

results. Most ETL and data integration platforms don't have more than a few different algorithms for performing joins and sorting, for example, so it is arguable that they wouldn't get much better performance if they did have an optimiser, because their choices are so limited. Syncsort, on the other hand, has some 30 different sort algorithms and a similarly large number of join and other algorithms. The optimiser then creates a transformation plan in the same way that a database optimiser creates a query plan. Moreover, this optimiser is dynamic so that it monitors data movement as it is happening and, if it finds that the current algorithms being used are not optimal, then it can dynamically change the transformation plan.

I could go on but suffice it to say that DMExpress is extremely efficient and, for bulk loading at least, probably the fastest product available on the market. But efficiency, (which means better use of resources and therefore reduced costs) and performance aren't everything, even if they are a lot.

If we go back to the discussion about the optimiser, the other big advantage with having an optimiser (and the algorithms to support it) is that the engine is effectively self-tuning. This means that you do not constantly have to tune your ETL processes, which in turn means that developers need to spend less time on maintaining existing processes and can spend more time on servicing business requests for more functionality and capability.

Finally, I should say something about Syncsort's positioning. You might think that it is a direct competitor to Informatica or IBM and in some cases that may be true. However, it can also be treated as complementary to those products. Many companies have significant investments in IBM or Informatica as a platform, using them for B2B purposes for example, or

employing their data quality and profiling tools. Syncsort is not in those markets but it does have the necessary metadata support to allow it to act as a data movement engine in conjunction with those environments whereby you can continue to design your transformations within Informatica, say, but then use DMExpress to actually move the data. To support this Syncsort is positioning DMExpress as a Data Integration Accelerator. It certainly is.

**Philip Howard**

*Research Director - Data Management  
Bloor Research*



© Bloor Research 2011