

## **Syncsort brings sort, acceleration and community twist to Hadoop integration**

**Analyst: Krishna Roy**

**Syncsort** is contributing an open source sort plug-in to the Apache Hadoop community in a bid to help promote ease of use and wider adoption of the open source data-processing framework. The company is also serving up a release of its flagship data-integration-acceleration product designed specifically for Hadoop, which is due later this year. DMExpress Hadoop Edition is designed to enable data to be extracted from and moved into Hadoop and will provide the ability to create integration jobs for MapReduce via a graphical UI. The upcoming offering also draws on DMExpress' ability to bolster performance when it comes to integration jobs.

### **The 451 take**

DMExpress Hadoop Edition was developed due to customer demand, so we see the low-hanging fruit for the new offering within Syncsort's existing installed base. That said, we also think that the free sort plug-in that the company is contributing to the Apache Hadoop community could act as a seeding strategy for its new offering, and help raise Syncsort's profile in the integration sector, which is still fairly low for a company that has been in data integration since the 1980s.

The sort plug-in that Syncsort is contributing to the Hadoop community is designed to enable third-party sort engines to be plugged into the data-processing framework. Although the plug-in will reportedly work with any sort system utility or homegrown sort code, Syncsort has its own sort engine and has built a business in bolstering sort operations with its wares. We think that many Hadoop developers will therefore want to use Syncsort's offering, which hooks into the open source sort plug-in via the Hadoop accelerator in DMExpress Hadoop Edition.

The Hadoop accelerator essentially aims to bolster the performance of MapReduce jobs by accelerating sort operations and will also reportedly invoke high-performance compression as needed to reduce storage requirements. DMExpress Hadoop Edition also moves transformations to its underlying engine for performance gains, sports a graphical UI for creating jobs and running them in MapReduce, and is designed to move data both in and out of Hadoop's file system.

DMEExpress Hadoop Edition will be available in the third or fourth quarter of 2011. Syncsort says it is not only designed to accelerate Hadoop deployments by reducing the elapsed processing time for jobs, but also brings about a number of other benefits including ease of use (because it is self-tuning) and a reduction in CPU and memory resources.

## Competition

Syncsort is not the only vendor to have embraced Hadoop from an integration perspective. Commercial open source players **Pentaho** and **Talend** were quick to realize an opportunity in supporting integration with the open source data-processing framework and served up offerings designed to move data in and out of Hadoop and create MapReduce jobs via a graphical UI, for example. **IBM** is also in the Hadoop integration fray, as is **Informatica**. Informatica is one of the reported players that Syncsort continues to encounter in DMEExpress deals – **Ab Initio Software** is the other.

We anticipate that other data management software providers such as **Oracle**, **Microsoft**, **SAP BusinessObjects**, **SAS Institute/DataFlux** and **expressor software** will enter the Hadoop integration fray due to the burgeoning popularity of the data-processing framework for 'big data.' We also see high-level parallels between **Pervasive Software's** focus on Hadoop and Syncsort's from a performance-boost standpoint. Pervasive's DataRush for Apache Hadoop aims to improve Hadoop's performance by drawing on its DataRush parallel dataflow engine to fully utilize every multicore processor in a Hadoop cluster.

Reproduced by permission of The 451 Group; copyright 2010-11. This report was originally published within The 451 Group's Market Insight Service.

For additional information on The 451 Group or to apply for trial access, go to: [www.the451group.com](http://www.the451group.com)