

SAS sheds light on 'big data' strategy, aka in-memory analytics, Hadoop data management

Analyst: Krishna Roy

30 Jul, 2013

SAS's in-memory strategy – underpinned by its SAS LASR Analytic Server – rests on the belief that in-memory processing should not be restricted to certain analytic purviews, such as data visualization, data exploration and modeling building, but applied to a whole gamut of tasks, including data preparation, model validation and deployment. As a result, the advanced analytic behemoth is bringing the performance boost germane to its LASR Analytic Server to a range of new and existing offerings for high-performance computing and 'big data' needs. SAS has also integrated the offering within Hadoop in order to provide in-memory analytics on data stored in the open source data-processing framework. That said, the company also views data management as important as analytics. This is reflected in its big-data management strategy for Hadoop, which spans data transformation, metadata management and acceleration of a range of tasks.

The 451 Take

SAS's continued mantle as the largest privately held software vendor with \$2.8m in revenue in 2012, combined with its ongoing dominance in advanced analytics, makes it a potent force. Moreover, the company has assembled a high-performance and big-data analysis strategy, which is well conceived. However, the products to support it are still a work in progress, although the company is picking up the pace of product launches. Furthermore, we'd like to see SAS expand into other emerging sources of big data, aside from Hadoop, and bring similar capabilities to machine-generated information, which we see as the next big-data opportunity.

Context

SAS moved into in-memory analytics a couple of years ago with its LASR Analytic Server. LASR has subsequently become a shining star within the company's portfolio – albeit one that has its light refracted through a lens other than its own. Why? LASR is not sold as a stand-alone product, but applied to others in the data management, advanced analytics and visual analysis domains, given it was engineered with these focuses in mind. With that end game, SAS's much-vaunted SAS Visual Analytics offering, which was introduced in 2012, was one of the early beneficiaries of LASR.

When it comes to Hadoop, SAS continues to tread several paths. It draws on the Hadoop ecosystem, including a strategic alliance with Cloudera announced in May, to bring the benefits of the open source data-processing framework – including its MapReduce programming libraries, which SAS deems complementary to its own eponymous programming language – to its installed base. SAS also uses Hadoop for storage in LASR deployments and as an acceleration platform for data management and analytics.

Products

So what exactly is the SAS LASR Analytic Server? It is an in-memory computational engine, designed specifically for the computing demands of advanced analytics. It is a 'read-mostly,' stateless, in-memory server providing secure, multi-user concurrent access to data for diverse analytic use cases including exploration, model building and decision-making. LASR can run on a single stand-alone server, in a distributed mode on Hadoop or alongside MPP relational databases like Teradata, Greenplum or Oracle. LASR can read/write data from the popular Hadoop formats like compressed sequence files, as well as SAS dataset format. SAS recommends Cloudera's Hadoop distribution. However, it is also working with Pivotal and a number of other Hadoop distributors.

LASR is increasingly being applied to an array of vertical and horizontal offerings since SAS hatched a strategy to extend it across its product portfolio to meet high-performance and big-data needs associated with deriving insights from masses of data and handling complex questions quickly in seconds. For example, SAS now has LASR-driven offerings for specific business scenarios, including marketing optimization, risk management, fraud detection and anti-money laundering, which dovetails with its strategy to provide advanced analytics to solve business problems rather than just serve up the mere algorithms themselves, although it provides those, too, for certain use cases. The company also has a slew of other LASR-based applications on the way later this year and in early 2014 as part of a mission to wrap high-performance advanced analytic techniques into readily

understandable offerings to support business-oriented analysis scenarios. Furthermore, LASR itself is a continued focus for development. A recommendation engine, for example, is in the works for collaboration, filtering and recommendation-style analysis.

So far, one of the most common uses cases for in-memory analysis has been visualization and regression-based analysis by dint of its integration with SAS Visual Analytics. That said, in most cases, end users wouldn't be aware of the regression-based approach, which would be explained in plain English via a type of interface guide for laypeople introduced last December. Visual Analytics has also undergone a series of enhancements since it first hit the streets in 2012. For example, it is now available in the SAS cloud as one of the company's recent forays into business-oriented predictive analytics as a service. Visual Analytics also now supports six forecasting algorithms, including Holt-Winters method to account for seasonality and, furthermore, has recently been furnished with Decision Tree-based analysis, with a continued emphasis on insulating end users from the underlying complexity of the analytic technique in question. SAS has also replaced the data-preparation interface within Visual Analytics with SAS Visual Data Builder and SAS Visual Analytics Administrator in a bid to improve preparation and admin capabilities. Visual Data Builder – which was introduced in December and includes a user-friendly query builder and the ability to read and write to any SAS source – is designed to make data acquisition into a self-service process capable of being carried out by business analysts, which are the core audience for Visual Analytics. Forecasting is also a prevalent use case for Visual Analytics.

Furthermore, SAS now has six other domain-specific in-memory analytics offerings that draw on LASR and are part of its high-performance computing arsenal. These products are all, in essence, about bringing in-memory processing to analytic model development – with a domain twist – to make them more applicable and understandable. The offerings in this product stable are SAS High-Performance Statistics, SAS High-Performance Data Mining, SAS High-Performance Text Mining, SAS High-Performance Optimization, SAS High-Performance Forecasting and SAS High-Performance Econometrics.

The main premise behind these offerings is to allow customers to develop predictive models using big data, including hundreds or thousands of variables, perform frequent modeling iterations and use complex analytical techniques to get answers to difficult problems. Besides predictive modeling, each of the six high-performance analytics products also includes a common set of procedures for data exploration, data transformation (binning, missing value imputation), dimension reduction and variable selection. The high-performance procedures take advantage of parallel processing and multi-threading in MPP distributed architectures (such as Hadoop, Cloudera, Teradata, Pivotal and Oracle), as well as on single-machine SMP architectures.

Hadoop continues to be a central focus for the company's big-data and HPC ambitions. SAS is, for example, continuing to bring to market accelerators, which started coming online earlier in 2013. For example, SAS Scoring Accelerator aims to make Hadoop into a high-performance model scoring and deployment environment for publishing scored models crafted in its longtime flagship advanced analytics offerings, SAS Enterprise Miner and SAS/STAT statistical analysis software. The SAS Data Quality Accelerator, which is scheduled for availability later in the year, is another notable offering in this portfolio. It is designed to draw on Hadoop's suitability for tasks associated with data quality such as profiling data.

SAS also provides SQL-like transformations for Hadoop's Hive and Hive Server2. The latter takes advantage of Kerberos for security and addresses some of the long-standing concurrency issues related with Hive. Customers can also create a job in SAS Data Integration Studio, including the Hive transformation, and mix SAS with Hive data. SAS also provides MapReduce transformations contending that MapReduce breaks down when it comes to performing iterative tasks, which are better handled by the in-memory approach supplied by LASR.

Competition

SAS tends to get compared with IBM most frequently, particularly since Big Blue acquired SAS's long-standing rival, SPSS, in 2009. Certainly, when it comes to high-performance computing and analytics and data management applied to the Hadoop big-data environment, IBM can also bring a plethora of products to the table that we think can match SAS. That said, SAS's overall approach differs from IBM's in that it focuses mainly on advanced analytics, bringing data management products for Hadoop and other environments and high-performance computing to bear as a facilitator to better analysis.

We also think Oracle and SAP provide competition to SAS on a number of fronts, although SAP HANA and Oracle's Exadata platform are regarded by SAS as complementary rather than directly competitive. Furthermore, it is also worth noting that SAS's support for IBM database environments makes it a partner rather than a competitor in some situations, too.

We would also like to draw attention to some of the main differences between SAS LASR Analytic Server and SAS Visual Analytics and other offerings employed for visual analysis and discovery, with which these offerings are often falsely compared. QlikTech International and Tableau Software, while ongoing champions of in-memory visual analysis, don't have the advanced analytic capabilities supplied by LASR, which is able to do simulation, what-if analysis and handle a number of different types of analytic techniques, including regression, clustering and Decision-Tree-based

analysis. That said, Tableau is looking to add more advanced analytic smarts, and we can't rule out QlikTech following a similar path.

When it comes to Hadoop-based analysis, there are a number of pure plays focused on this market, including Datameer, Karmasphere and Platfora. Interestingly, the latter also takes an in-memory approach to Hadoop analytics. Informatica, Global IDs, Talend, Actian's Pervasive Software business, Pentaho, Cirro and Syncsort provide a second layer of competition when it comes to data management capabilities for Hadoop.

Actian/Pervasive is also in the big-data advanced analytic fray, which is also a significant and growing business for Pentaho. Informatica provides parsing capabilities as well as support for Hadoop as a source and target, while Syncsort brings a sort and acceleration twist to the open source data-processing framework, which it regards as the operating system for big data. Having introduced the ability to move data in and out of the Hadoop open source processing framework as an early big-data player, Global IDs is now looking to better support Hadoop as an ETL environment. This is where its strategy competes against SAS, which has a similar end game in mind. Informatica, Talend, Syncsort, Pentaho and other integration players are on a similar mission.

SWOT Analysis

Strengths

SAS's strategy to build rather than buy the vast majority of its products makes for a cohesive, integrated portfolio of analysis offerings that are better able to handle big-data needs as a result of LASR. It also has a solid lineup for customers using Hadoop.

Opportunities

Expanding the company's purview beyond its traditional audience of data scientists and data analysts is a fertile avenue to explore and, furthermore, now made possible with offerings such as SAS Visual Analytics and SAS High-Performance Analytics products - although there is still more work to be done on this front.

Weaknesses

The company could do more to support other emerging big-data environments aside from Hadoop, including NoSQL databases and machine-generated information. It could also benefit from forging relationships with Hadoop distributors other than Cloudera.

Threats

IBM is the closest in being able to match SAS's capabilities for high-performance computing and Hadoop. The latter is the focus for a variety of players from a data management and analysis perspective, presenting increasing competition - particularly when the customer is not an existing SAS account.

Reproduced by permission of The 451 Group; © 2013. This report was originally published within 451 Research,Âs Market Insight Service. For additional information on 451 Research or to apply for trial access, go to: www.451research.com