# Notes From the Giga Advisor: Viability of the ASP Model for Data Mining

By Lou Agosta

Published in DM Review
in May 2001

The idea of applying the ASP model to business intelligence has aroused skepticism for good reason. End-user firms are reluctant to allow anyone else to access their sensitive customer and product data. However, a new company named digiMine has assembled at least three of the four necessary components for a successful go at it. These include the talent, the technology, the operational acumen and the target market. Obviously, any such undertaking is inherently risky, and all the usual disclaimers apply. However, in conversation with the CEO, Usama Fayyad, and background checks on the overall situation, Giga was impressed with both the credentials and the delivery, as well as the results attained to date.

**The Technology.** A recent digiMine ad shows an amusing picture of the rear view of a person's shaved head with mathematical formulas on it in black marker. The caption reads: "Got data? We'll do the math." All the evidence supports the assertion. Everyone has the data. Without the data mining application to analyze it, the data is meaningless (and worthless). Likewise, the analytic application is empty without the data. digiMine incorporates Microsoft OLE/DB for Data Mining, a Microsoft interface to data mining services using the familiar SQL API. With the shipment of SQL Server 2000, Microsoft has changed the name of OLAP Services to Analytic Services. Analytic Services now contains both OLAP and data mining. In a sense, digiMine is the ultimate proof of concept. Given digiMine's price point, using an ASP can actually be a lot less risky than undertaking in-house data mining development (and the implied data warehouse) from scratch.

**The Talent.** One thing that is particularly impressive is that Usama Fayyad headed the team that developed Microsoft OLE DB Data Mining Services. The other key personnel are also impressive with Bassel Ojjeh having a proven track record in large data warehousing systems internal to Microsoft and Nick Besbeas, also a Microsoft alumnus, with extensive direct marketing experience.

**The Market.** The target for digiMine is data mining of Web log and related CRM (i.e., order entry) services. While acknow-ledging that one dot-com prospect simply vanished from the radar, Usama stated that the involvement of the bricks-and-mortar firms in e-business and their commitment to learning from the mistakes of the early pioneers is strong and growing.

**Operational Acumen.** This brings us to the issue of operational acumen. As digiMine wins additional business, they will increasingly be in the business of building infrastructure, contracting with firms such as Exodus to provide T3 lines between client and digiMine data centers. This is all good, honest data processing and, in comparison with certain kinds of market basket analysis, essentially a solved problem. However, it is not to be taken for granted. This is where the operational expertise of Bassel Ojjeh, digiMine's COO, will be tested. The good news is he does indeed bring considerable depth to the role, having built data warehouses and analytic applications for Microsoft's initiatives such as MSN, MSNBC, Expedia and CarPoint. How-ever, this is one of those cases where the provisioning of the plumbing can be just plain hard work. Costs for storage, networks, database software, etc., can mount just as quickly as the underlying data itself. Currently, digiMine estimates it has about three terabytes of data under management. This can be expected to grow rapidly as they acquire more customers, each generating 20GB of Web log data. Of course, Web logs contain a significant amount of noise data and shrink down nicely to about a tenth of their original size leaving only essential data points such as customer identity, product, "from" and "to" page, and date/ time stamp. Finally, the ability quickly and easily to reach a workable service level agreement (SLA) is on the critical path to building win-win relations between digiMine and its clients. This is a well-defined problem that can be addressed by the usual amounts of hard work and people skills.

In a useful oversimplification, digiMine is best described as an experiment in data mining using an ASP model. By implication, it is also an experiment in business intelligence and data warehousing using an ASP model because having a data warehouse of clean, consistent data is a useful target for data mining activities. To the end-user firm, the benefits are straightforward  Web logs transformed into customer identities, highlighted buying behavior, market-basket analysis and related analytics. The end-user firm pays one flat fee, installs the digiMine Data Slurper in its data center to define the logical unit of work in communicating with the digiMine Operations Center and uses a Web browser to inspect the resulting analytics. The end user doesn't need to know or care if digiMine DBAs are scurrying around like crazed weasels squeezing the underlying Microsoft SQL Server 2000 technology to perform at volume points of a terabyte and above (or even what technology lies under the covers). If there was any doubt about whether data mining via the ASP model was possible, it looks like we (and the market) are about to find out.

---

Lou Agosta is director of research in data warehousing at Giga Information Group and the author of **The Essential Guide to Data Warehousing** (Prentice Hall PTR, 2000). Agosta and his colleagues at

Giga provide objective research, advice and continuous coaching on technology for e-business. Please send comments or questions to lagosta@acm.org.