BMMsoft Raises the Bar on Data Warehousing: Integrating Structured and Unstructured Data in the E-Mail, Document Management, Multi-Media, and Transactional Data (EDMT®) Server

The Market

What happens when an irresistible force meets an immovable object? We are about to find out. The irresistible force of BI, eDiscovery, compliance, fraud detection, governance, risk management, and other analytic and regulatory mandates is heading straight toward the immovable rock of year-to-year 10% reductions in information technology budgets.

The convergence of the markets for structured and unstructured data has been heralded many times. This white paper will not do so. That is because they will *not* converge. The continued separation of the markets for data warehousing, email archiving (eDiscovery), and document management will be a fact of life. However, looking into the crystal ball, what will happen is that an increasing overlap already underway will disrupt incumbents across these diverse markets.

The data warehousing database software market has been sized at about $5.2 billion by IDC.[1] The document management software market is estimated at nearly $3 billion.[2] While email archiving is relatively new and growing rapidly due to near federal regulations, it has now reached the $1 billion "take off" point. In short, at nearly $10 billion total, a product that addressed requirements across all three of these markets with a reasonable prospect of response from even one third of the enterprises, would have an outside boundary of over $3 billion. This is a substantial market under any interpretation.[3]

In the meantime, the exiting markets for these three classes of products is fragmented into silos of the traditional data warehousing vendors, email archiving, and document management, the latter sometimes including compliance and governance software. The first are well known in the market – extending from such stalwarts as HP, IBM, Oracle, Microsoft, SAP, to data warehousing appliances and column-oriented databases – and will not be repeated here (though one new developments will be noted below). Document management systems include IBM FileNet Business Process Manager (www.ibm.com), EMC Documentum (www.emc.com), OpenText LiveLink ECM (www.opentext.com), Autonomy Cardiff Liquid Office (www.cardiff.com). Strictly speaking, risk management is considered a separate market from document management. Risk management and compliance offerings include Aventis (www.aventis.com), BWise (www.bWise.com), Cura (www.curasoftware.com), Protiviti (www.protiviti.com), Compliance 360 (www.compliance360.com) and IBM, which has at least two offerings

---

[1] Cited in Information Week January 2007 - http://h20223.www2.hp.com/NonStopComputing/downloads/iwk7198-final.pdf The data warehousing tools part of the market is estimated at some $9.6 billion in the same article.

[2] http://www.researchandmarkets.com/research/c3917a/document_managemen

[3] For further background see "Enterprise Data Mashup: Data Integration for Structured, Unstructured, and Web," based on "Gartner Cool Vendors in Enterprise Data Management and Integration," Ted Friedman, David Newman, Donald Feinberg, Whit Andrews," March 31, 2007; and see also "IDC Executive Brief Enabling Better Decisions through Unified Access to Information" *IDC #34836.*

one based on Lotus Notes and one based on FileNet. This list is partial and could easily be expanded with many best of breed offerings. The result? Fragmentation. Diversity, though not in a positive sense. Many offerings instead of a comprehensive approach to unified access and unified analysis.

Five years from now data will be as heterogeneous as ever and the uses of data even more so, but individual products – single instance products, not solutions – will characterize a transformed market for database management that traverses the boundaries between email archiving, document management, and data warehousing with agility that is only dreamt about in today's world. Video clips are now common on social networking sites such as Facebook and YouTube. Corporate sponsorship of such opportunities for viral marketing is becoming more common. The requirement to track and manage product brands and images will necessitate the archiving of such material, so multi-media (image/video/audio) are being added to the mix.

This future is being driven and realized by the imperative for business transparency, risk management and compliance, and growing regulatory requirements layered on top of existing business intelligence and document management requirements. Still, document management is distinct from workflow. If an enterprise needs workflow, then it will continue to require a special purpose document management system. Workflow was invented by FileNet in 1985, acquired by IBM in 2006, and continues to lead the pack where detailed step-by-step process engineering is required. Elaborate rules-engines for enterprise decision management are different than compliance. If an enterprise requires a rule-engine for compliance and governance, then it will need a special purpose compliance, risk management, and governance system. Such solutions would be over-kill for those enterprises that require email archiving for eDiscovery, document management for first order compliance, and cross references to transactional data in the data warehouse. While the future is uncertain, one of the vendors to watch is BMMsoft.

Innovation Happens

BMMsoft has put together a product delivering functionality across these three previously unrelated silos – data warehousing, eDiscovery (e-mail), and document management - and able to be purchased as a EDMT®Server – a single part number from BMMsoft (EDMT stands for "E-Mail, Documents, Media, Transactions"). The database "under the hood" is Sybase IQ, a column-oriented data store with a proven track record and several large objectively audited benchmarks. The latest of these weighs in at 1000 terabytes – a petabyte – and was audited by Francois Raab, the same professional who audits the TPC.org benchmarks.[4] The business need is real. Based on demonstrations and conversations with Paul Krneta, Founder, and Bud Michael, CEO, so is the product.

The three keys to connect and make intelligible the data from the three different sources are:

[4] http://www.sybase.com/guinness.

1) extreme scalability to handle the data volumes – this is where a column-oriented database would come in handy since the storage compaction is intrinsic and prior to the additional compression that could be applied;
2) parallel, high performance ETL functionality to load all the data; and finally
3) search capabilities that enable high performance inquiries against the data.
Such unified access to diverse data types, intelligently connected by metadata, is also sometimes described as a "data mashup."

A part of the challenge that a start up - and up start - such as BMMsoft will face is building credibility. Of course, client success stories will contribute to that and some are coming on stream. But in the case of BMMsoft there is another consideration. The initial claims that it can play in three markets and across some ten diverse functions - let's be candid about it – initially strains credibility. However, the reason the product is for real is that metadata is an underestimated and underdeveloped opportunity. Innovations in metadata that make possible many applications that require cross referencing emails, documents, and the transactional data. For example, fraud detection, threat identification, enhanced customer relations – all require navigating across the different data types. Metadata makes that possible. That is not an easy problem to solve; and BMMsoft has demonstrated significant progress with it. Second, the column-oriented database is intrinsically skinny in terms of data storage in comparison with the standard relational database, which continues to be challenged by database obesity. As data warehouses scale up, the cost of storage technology becomes a disproportionably large part of the price of the entire system. Note that for the column-oriented approach proportional cost savings come into view and are realized. Third, this also has significant performance implications, since if there is less data – in terms of volume points – to manage, then it is faster to do so. So when all the reasons are considered, the claims are quite modest, or at least in line with common sense. The wonder is that no one thought of it sooner.

When you think about it for a minute, there is every reason that an underlying database should be capable of storing a variety of different data types and doing so intelligently. The latter intelligence is the "secret sauce" that differentiates BMMsoft. The relationships between the different types of data are built as the data is being loaded by BMMsoft using patent pending software technology. The column-orientation of the underlying data store – Sybase IQ – intrinsically condenses the amount of space required to persist the information, yielding up to an order of magnitude – more typically a factor of two or three – in storage savings, even prior to the application of formal compression algorithms. This fights database obesity across all segments – email, document, media, transactional (structured) data warehousing information. This means that the application that lives off of the underlying data is able to take advantage of performance improvements since less data is being stored and more being fetched with every data retrieval. For those enterprises with a commitment to installed Oracle or MySQL infrastructure, BMMsoft provides investment protection. The EDMT® Server runs also on Oracle and MySQL and can be easily ported to any other relational Database.

Thus, BMMsoft is a triple threat and is able to function as a standalone product addressing data warehousing, email archiving, and document management requirements as separate silos. But just as importantly, for those enterprises that need or want to compete with advanced applications in fraud detection, security threat assessment, customer data mining beyond structured data, BMMsoft offers the infrastructure and application to do so. For example, the ability to perform cross-analysis between securities traded on the stock market and those companies named in email and voice mail (remember multimedia handling) will immediately provide a short list for follow up detection on on-going insider trading or other fraudulent scheme. While hindsight is 20-20, a similar method of identifying emerging patterns through cross-analysis would have been be useful in surfacing the 8 billion dollar Societe General fraud, Madoff's nonexistent options plays at the basis of the pyramid, the Georgia Tech shooter, and relevant chatter that shows up prior to terrorist attacks. Going forward, this technology is distinct in that it can be deployed on a small, medium, or large scale to highlight emerging hot spots that require attention.

One may object – but won't the competition be able to reverse engineer the functionality and provide something similar using different methods? Of course, eventually every innovation will be competitively attacked by some more-or-less effective "work around." Read the prospectus – new start ups and existing software laboratories at HP, IBM, etc. will eventually produce innovations that challenge the contender. However, that could require three to five years. Then there is the matter of bringing it to market. IBM provides an example, based on publicly available news reports. IBM went out and purchased FileNet for about $5 billion dollars. FileNet is a great company, which virtually invented workflow, and if one requires advanced workflow capabilities, it is hands down a good choice. However, it does not do data warehousing or email archiving. As a subsidiary of IBM which delivers substantial revenue to the "mother ship," the executives in charge will set a high bar on any IBM innovations which combine email archiving and structured data warehousing with document management. In short, IBM is faced with the classic innovator's dilemma.[5] The price points that interest it – both internally and externally – are further up on the curve than the deals that BMMsoft will be able to complete. Once the up start (in this case, BMMsoft) gets a toe hold in the market, it has a good chance of marching up market, displacing the installed, legacy solutions as it goes. This happened before in the client server revolution when IBM mainframe deals at the several million dollar price point were undercut by a copy of PowerBuilder and a copy of Sybase, albeit a different version of the database. So, yes, the competition will eventually get it right. But if it strikes early and often, BMMsoft will be able to exploit first mover advantages and build an installed base that will be challenged only with great difficulty.

Business Value Lives

Given the economic dynamics of Q1 2009, many end-user firms are in reaction. They are reacting to reduced consumer confidence and consumer spending. However, regulations around the archiving of email – also called e-Discovery – will drive the installation of

[5] Clayton Christensen, *The Innovator's Dilemma*. Cambridge, MA: Harvard Business School Press, 1998.

email archiving systems in good or bad economic times. Why? Because federal regulations around email archiving and adverse decisions in federal courts in the USA due to lack of timely email retrieval are spreading fear into the hearts of CIOs and CEOs. This is not a white paper on updates to Federal Rules of Civil Procedure, but the message is not just for boy scouts anymore: be prepared.[6]

In addition, what some pundits have described as a populist backlash at the business executive excesses of the first decade of the new millennium, whether real or imagined, will require new levels of transparency and oversight in corporate operations and finances, all implying more IT systems in support of email, document management, and data warehousing. The trend will be to try to "store the universe," whether wisely or not, against future inquiries for fraud, security data mining, or even customer service. Finally, even those firms "in reaction" are looking for rational ways to reduce costs. Ongoing initiatives and projects such as data mart consolidation, which have proven cost reduction value, will continue to drive the installation of large, centralized data warehouses, document management systems, and archives.

This is occurring simultaneously with the proliferation of data marts on column-oriented databases and data warehousing appliances, the latter being main stream. The performance issues faced by the standard relational database, the dominant design in managing structured data in the corporation, will remain a dirty little secret only as long as no viable alternative is available. Do not complain unless you can do something about it. But the proliferation of alternative approaches is a direct consequence of the requirement for faster answers to local, individual queries and shorter time to value for BI/DW applications. One new development - special purposes backend accelerators from Oracle, SAP, and the underlying blade vendors such as HP, IBM, Sun, and Dell is also getting traction in spite of economic headwinds. This is a mixed blessing since such solutions would not be needed if the underlying data stores were delivering the performance required by the business and promised in the initial service level agreements.

In this environment, applications are required to show tangible results. Tangible results extend from "saving money" through "making money" with an emphasis on automating business processes that were previously performed manually as well as working smarter. Document management with a view towards compliance and risk reduction are areas where manual processes still predominate in many firms. If a manual process can be automated, then it often promises an order of magnitude return on investment. Even if the business process can only be partially automated, because the business judgment of experts and executives is still the critical path, the process still offers substantial, integral factor improvements in productivity (and value added).

[6] A nice short tutorial with examples is available at http://www.inboxer.com/index.shtml [site checked on 03/03/2009].

As noted above, the consolidation of multiple data marts onto a single platform is able to reduce the costs of coordinating multiple instances of hardware and software. If the application can also be consolidated, then even more costs can be squeezed out. Now consider the possibility of consolidating across silos such as data warehousing, email archiving, and document management that were previously considered to be unconditionally separated. If a single database image could be used to manage all of the above, new possibilities open up for more efficient utilization of database administration (DBA) staff, maintenance, and scheduling. In that sense, BMMsoft is riding the wave of new, previously undefined possibilities.

If hardware power were unlimited, storage infinite, and bandwidth ever expanding, then there would only be one data store – in the world. Since hardware, storage, bandwidth are limited in the real world, the professional in the information technology (IT) departments have job security integrating all the products and solutions required to operate the information factory needed to sustain and advance business processes. However, every once in awhile, the convergence of Moore's law, falling storage prices, advances in broadband networking are combined with software innovations in data integration, indexing, and metadata, resulting in new possibilities. The new possibility of storing email, documents, and data warehousing data in a single intelligently integrated instance is one such innovation. This will lead to a breakthrough in applications such as eDiscovery, fraud detection, customer relations, and related business intelligence operations.

No client in his right mind would choose to buy three separate products when one product will do the job. The historical development of IT has resulted in silos across email archiving, document management, and data warehousing. Those installations that have lagged behind, for whatever reason, have the opportunity to leapfrog their competitors in solving three out of three problems with a single product. Others may address two out of three. One out of three is not a bad ratio if it solves a recalcitrant problem at a competitive price while adding business value. In all cases the approach is to implement a single database in an appliance-like reference architecture that stores all three types of data, cross references the relationships between them at load time, and, using metadata "under the hood," delivering answers to business questions of the analyst's own formulation at a single interface. Data integration is performed automatically at load time, in effect, by building indexes and metadata as an intrinsic part of the process. Thus, the data integration and coordination cost is zero. The approach yields a total cost of ownership (TCO) that is less than 12% - or over 8.9 times the value of – separate, siloed products to address the same set of requirements.

Most innovative technologies start at the margin and below the dominant design of the market leaders (i.e., standard relational databases) and through superior implementation, marketing, delivery, execution, work their way towards the center. Make no mistake - because of innovations in metadata, ETL, and front-end search the approach of BMMsoft has the potential to execute a pincers movement on data warehousing, document management, e-mail archiving and eDiscovery, executing a "squeeze play" on the

existing market in a positive sense to deliver superior value to customers at increased functionality and reduced prices**.**

3 year TCO for F1000 Enterprise ($000)

Components Required	Purchase Cost	Training & Set-up	3 year Maintenance	Integration Cost	Total
Email archive	$ 300	$ 20	$ 150	$ 50	$ 520
Email compliance/retention	$ 100	$ 20	$ 50	$ 50	$ 220
Email backup	$ 100	$ 20	$ 50	$ 50	$ 220
Email Standby server	$ 100	$ 20	$ 50	$ 20	$ 190
File archive/retention/compliance	$ 500	$ 20	$ 250	$ 50	$ 820
File backup	$ 200	$ 20	$ 100	$ 50	$ 370
Text search	$ 200	$ 20	$ 100	$ 50	$ 370
eDiscovery/Litigation support	$ 1,000	$ 50	$ 500	$ 50	$ 1,600
Data warehouse	$ 1,000	$ 100	$ 500	$ 50	$ 1,650
DW & DB archive/compliance	$ 500	$ 20	$ 250	$ 50	$ 820
Data migration tools	$ 300	$ 50	$ 150	$ -	$ 500
Hardware infrastructure	$ 500	$ -	$ 50	$ -	$ 550
Total for "Operational System"	$ 4,800	$ 360	$ 2,200	$ 470	$ 7,830
Disaster Recovery Site (passive)	$ 500	$ 100	$ 300	$ 50	$ 950
Point-product TCO	**$ 5,300**	**$ 460**	**$ 2,500**	**$ 520**	**$ 8,780**
EDMT® Server Software	$ 480	$ 50	$ 240	$ -	$ 770
Hardware infrastructure	$ 100	$ -	$ 10	$ -	$ 110
Total EDMT® Server Solution	**$ 580**	**$ 50**	**$ 250**	**$ -**	**$ 880**

Full Disclose: Lou Agosta, Ph.D., is an independent industry analyst who also serves as a member of the BMMsoft Advisory Board.