Collaborative Business Intelligence for Notes

A Revolution in Business Intelligence

Lotus Notes and Business Intelligence are terms that you would not expect to see together in the same sentence. Practically speaking, the Notes world and the BI world have literally been on other planets. This is unfortunate since Notes is especially adept at handling those types of rich context-full content that would especially enrich BI in the enterprise. The main reason that these worlds have not been closer is that the content in Notes has been particularly difficult to access from BI products. Until now. Sun & Son, through its Data Modeler for Notes, now provides the essential missing connection for quick, easy and powerful integration between Lotus Notes and Cognos BI. This is a revolution in itself. The Data Modeler opens Notes to a quantum leap in Business Intelligence capabilities. At the same time it makes Cognos BI more powerful, more pervasive and more indispensable to the enterprise through the integration of the huge and important Notes installed base.

Notes and Business Intelligence

Lotus Notes has been in use and growing for over 15 years. Originally designed for collaboration with shared documents, it has grown and expanded in ways that would have amazed its original developers. It is a leading email platform, a leading Web application server, a sophisticated development platform and the foundation for millions of collaborative applications, many of them mission critical to their organizations. Notes has, from its release, been intrinsically multiplatform and offers organizations a broad range of supported platforms on which to run the Domino server. Notes has a formidable and growing presence worldwide. There are currently over 125 million licensed Notes users.

Enterprise Business Intelligence is the category applied to the set of tools which organizations use to understand their data, make decisions and manage execution. Enterprise BI typically includes the following capabilities: Enterprise Reporting, Analysis (OLAP), Metrics Management (scorecarding), and Enterprise Planning. This field is also frequently referred to as Corporate Performance Management. Where in the past Enterprise BI has typically been the domain of financial officers and senior management, it is now seen as an essential set of capabilities for managers and knowledge workers throughout the organization.

Enterprise BI has traditionally been associated with the IT segments of data management and data warehousing. Since leading BI platforms originated from these disciplines, they include the capability of metadata modeling, an

important technology from the data warehousing world. These modeling tools allow the underlying data structures to be abstracted into business user-friendly data models for reporting, analysis and scorecarding. Multiple underlying data sources can be combined into a single metadata model. The use of metadata models allows users to interact with simple and meaningful data structures which hide an underling complexity of different sources and different structures.

Most organizations rely on multiple critical applications for different aspects of their operations, e.g. ERP, HR, CRM and Customer Support. Organizations have found it prohibitively expensive to 'rip & replace' many legacy applications. Instead, they are able to use Enterprise BI to tie their separate applications together and provide a comprehensive view of their organization through enterprise reports and dashboards.

Cognos is the technology leader in the Enterprise BI field. It has continued to raise and set the technology standard for Enterprise BI. Cognos' latest offering, Cognos 8 BI, includes browser-based interfaces for query, reporting, analysis and metrics management, all powered by a single BI server with integrated metadata modeling services. Over 3,500 organizations in all business segments rely on Cognos for their business intelligence needs.

Traditionally there has been very little interaction or overlap between Notes and Enterprise BI. Notes actually existed long before the general popularity of Enterprise BI, but as organizations large and small have standardized on leading Enterprise BI vendors' products, their focus has been primarily on structured data in relational databases (RDBMS) like IBM DB2 or Oracle. Whereas RDBMS are optimized for highly structured row and column data, the Notes data store is optimized instead for semi-structured document-based content. This has made Notes difficult to integrate into BI solutions. Even today, in the same organization, users of BI products like Cognos 8 BI tend to be isolated from the communities of Notes developers, administrators and users.

A significant part of the appeal of Notes over the years has been the speed and ease with which users can build deployable applications. A Notes developer can develop a prototype and then expand this to a full application quickly and without extensive training. This does not mean that Notes is not a serious development environment - it contains a wealth of powerful development tools - but Notes applications can begin simple and grow to add more sophisticated capabilities over their life cycle. Notes application development is also very cost effective. A common rule of thumb is that it can take 5 - 10 times more developer resources to develop the same application in a J2EE environment than with Notes.

For these reasons, many companies which use Notes now have hundreds or thousands of existing Notes applications. Some were built to well defined specifications and sound development practices. Many others started out as grass roots 'quick hacks' and have taken on lives of their own as they have been enhanced and expanded. As a result, many IT shops are now expected to support and maintain a bewildering array of Notes applications, some acquired from third part vendors, and many developed in house. In most cases these applications do not comply with corporate data standards.

As BI has become a greater priority for corporate IT, Notes has often come to be viewed as a large and troublesome installed base, unable to be effectively managed and effectively integrated into Enterprise Reporting and BI systems. In some organizations this has even threatened Notes' continued role as a standard for application development.

Issues with bringing Notes Data to the Enterprise

Notes has been used extensively for many types of document-based applications. Typical examples of Notes applications include: document management, tracking, workflow processing, CRM, SFA, self service Web applications, and applications which require disconnected use. Notes applications serve myriad business needs. Many Notes applications are in fact critical to the day to day business operations of their organizations.

More and more, Notes applications are also being expected to fit in to and serve broader enterprise roles. Over the years, Notes applications have become more sophisticated and more complex. Many modern Notes applications consist of multiple Notes databases and are expected to interoperate with enterprise business systems and non-Notes databases. This process further highlights some of the inherent limitations of the Notes architecture.

The basic units of data storage in Notes are the document and the database. A Notes document is roughly analogous to a record in a relational system, but unlike a relational table, there is no enforced data structure or schema. The fields in a Notes document are initially defined by the form with which it was composed, but each document can contain a different mix of fields and field values based on the specific application. The Notes database contains all of the documents and design elements together in one .NSF file.

The Notes data structure is highly optimized for semi-structured data. It easily handles rich text, attachments, embedded graphics and active objects, but, at the same time, it does not scale well to large numbers of documents in the same database. Notes works most efficiently with applications which

manage fewer that 100k documents in a single database. Document management requirements for millions of documents are typically met by RDBMS-based document management systems.

Notes uses a View (or Folder) presentation to display collections of documents. Sets of documents can be displayed based on selection formulas (Views), user actions (Folders), or search results (full text search). Notes Views can be rendered either in the Notes client or in a Web browser.

In a Notes View each document is typically displayed as one line in the view. View structures are flexible, easy to build, and highly optimized for quick real-time access to data. Views enforce the Notes security model so that users only see documents to which they have access. Designer level access to the database is required to build or modify views, so view maintenance must be done by Notes developers, not by end users.

Notes views are design elements within the Notes database (.NSF file). The corresponding view index tables are also stored in the Notes database. This allows views to be easily replicated to all instances of the database. For the same reasons, a Notes view is limited to displaying only documents stored in its own database. This makes it difficult to produce an overview interface for a multi-database Notes application.

Notes Views have some other notable limitations. They do not work well to produce formatted (print quality) reports, reports based on specified time frames, charting and analysis, drill down or roll-up. It is also difficult to distribute Notes views outside of the Notes application. Notes views do not allow easy ad hoc querying and filtering except through full text search. Notes views also cannot display the formatted content of rich text fields.

As Notes applications become more entrenched in the organization, requests come from users for more and more ways to view the data contained in the application. These requests are typically implemented as additional Notes Views. It is not unusual for legacy Notes applications to have hundreds of Notes Views. This puts a growing performance load on Domino servers. Each time a Notes document is modified, the Domino server must update all the indexes for all the views which display that document. Many of the requirements for these views could be better served through BI reporting.

There is a clear need in Notes environments for Enterprise level Reporting solutions. Notes Views serve an essential purpose, but fall short in their ability to provide more powerful and flexible reporting functions. In the Notes world, there has been a long unmet need for powerful, easy to use reporting, charting and analysis capabilities. These BI capabilities cannot be

limited to a single database, but must be easily available for reporting across multiple Notes databases or between Notes and other database types.

Previous Options & Approaches

Over the years there have been a number of partial solutions to these problems around reporting from Notes, but unfortunately each has addressed only a part of the problem.

Lotus provides a NotesSQL utility which serves as an ODBC driver and allows SQL-based access to the contents of Notes databases. This provides a basic level of accessibility to Notes content from external systems. Unfortunately, NotesSQL is a mixed bag. It also suffers from a number of significant limitations.

- NotesSQL is not stable under intensive usage and is therefore not suited to high volume or mission critical access. Part of this is due to limitations in the underlying ODBC architecture. As an ODBC driver, it is also limited to Windows machines.
- NotesSQL exposes very Notes Form, View and Folder as a separate data table. The result is often a bewilderingly confused view of the target database. Without a clear understanding of the structure of the database, most users find it difficult to understand how to structure a meaningful query.
- Many Notes elements exposed as NotesSQL tables can have horrible performance characteristics. A user has to fully understand how NotesSQL works and the details of the underlying data structure in order to achieve acceptable levels of performance for queries.

Over the years, packaged reporting solutions for Notes have been provided by Crystal Reports and other vendors. These consist mainly of reporting and charting tools which access Notes databases using NotesSQL or native connectors. These tools have best served the need for higher quality formatted reporting than provided by Notes. They have tended to be difficult to use, though, since they require a clear understanding of the structure of the Notes database to produce meaningful reports. A user had to be a Notes developer to generate reports, and then only after a lot of work. Since these tools do not include metadata modeling capabilities, the reports they produce are limited to data from one database at a time. For all of these reasons, these tools have not gained a broader acceptance, and have served instead as niche products for Notes developers. The broader market for Enterprise BI users has not been tapped.

Collaborative BI for Notes - A Broader Vision

We believe that the most effective way to bring Notes and Enterprise BI together is not by adding reporting capabilities to Notes, but by making Notes a first class player in the Enterprise BI world. Many of the reporting issues that have seemed insurmountable in the past can be effectively resolved by making Notes data readily accessible to Enterprise BI, with its metadata modeling capabilities.

Cognos 8 BI, and its predecessor Cognos ReportNet, are built upon a fully integrated data modeling capability. Since metadata modeling is a new concept for many Notes developers, we will expand upon this a bit.

Application data structures are designed primarily for efficient storage and transaction processing. In many cases, the data design strategies pursued for efficient application processing go counter to the strategies one would pursue for efficient and flexible reporting from the same data. This has been one of the major reasons for the keen interest in data warehousing solutions in recent years. For data warehousing, data is extracted from source applications, transformed into structures more suitable to reporting and analysis, and loaded into the data warehouse (ETL). While the data warehousing approach is valuable or essential in some cases, it requires additional tools, additional steps, additional data storage, and is ill suited to real time data access.

In many cases it is preferable to access the data directly in the source applications, but still provide the same kinds of transformations done with ETL through metadata modeling. Metadata modeling tools provide application interfaces which allow the data modeler (the person doing data modeling) to create a simpler, more readily understandable abstracted views of the data. These views hide the underlying complexity at the database level. The user sees a data model which is presented in a way which makes sense to a business user and which is easily applied to reporting and analysis. A single metadata model can combine data from multiple source databases. The BI services use the metadata model to resolve queries based on the user model to native queries to the individual data sources and take care of stitching together multiple query result into a single result set for the user or client application.

Cognos Framework Manager is the metadata modeling tool provided with the Cognos BI products. All of the Cognos 8 BI tools: Query Studio, Report Studio, Metrics Studio, Analysis Studio are built on top of the same metadata management layer. Users of any of these BI tools see and access the

business models of the data, not the native structures of the underlying data sources.

This highlights an important advantage of using metadata modeling techniques. By taking this approach we can separate the roles for data modeler and report designer. The data modeler must have a clear understanding of the structures of the native data sources and the kinds of business data models which will be meaningful to BI consumers. The report designers or the consumers of the data need only concern themselves with the business value of the queries, reports and analyses they produce. They do not need to understand the underlying data sources or their structures.

As we have seen, other reporting solutions that have been available for Notes have only served niche markets, primarily because they have been difficult to use. Since they lack metadata modeling capabilities, these tools require users to understand both the data structure of the databases and the reporting requirements. Business users find the underlying data structures difficult to grasp, and developers are tasked with creating reports – a job better done by business users. On top of this, it is impossible to do multi-database Notes reporting without a metadata modeling capabilities or extensive programming. By linking Enterprise BI with Notes though metadata modeling, multiple Notes databases can be joined together into a single 'data surface'. Alternatively, Notes data can be joined with data from relational databases or ERP systems in the same metadata model. By making enterprise BI available to Notes, Notes databases are no longer islands of data by themselves, but can participate as first class citizens of the Enterprise BI world.

But maybe we're getting ahead of ourselves. First we have to be able to provide access to our Notes data from our Enterprise BI platform.

Enterprise Data Access Strategies for Notes Data

Through our experience with Collaborative BI for Notes, we see three primary strategies to accessing Notes content from BI applications. There is no single best solution for all business cases. Each approach has differing strengths and weaknesses when applied to different usage models. We will discuss each strategy separately.

ODBC Access: NotesSQL is an ODBC driver that is available free from Lotus and can be set up in minutes. As mentioned previously, it does have performance and platform limitations. It is most appropriate when used against small to moderate sized databases (less than 50k documents) and under light query loads. Since ODBC is a Windows 'standard', this approach

will only work when queries are generated from a Windows machine. This means that the BI application which generates queries must be installed on a Windows machine. The Domino server, however, can be on any platform, as long as it is not running on the same machine as the BI application. NotesSQL requires Lotus Notes, Lotus Domino or Domino Offline Services (DOLS) to be installed on the same machine since it requires access to Notes runtime resources.

Native Access: Native access is provided by application drivers which execute queries within the BI application against a Notes database. There are currently two approaches to Native Access: Notes 'Classic', using native Notes protocols and NSFDB2.

Notes 'Classic': This refers to the classic Notes .NSF file structure. Native drivers are developed by BI vendors using Notes APIs. These drivers translate their own queries into Notes Remote Procedure Calls (NRPC) and then translate the results back into result sets that the calling application can use. These drivers can offer better performance and reliability than the ODBC access provided by NotesSQL. They are however expensive to develop, unique to each application, and therefore not available from all vendors. Each connector is specific to each vendor's application and must be maintained by the vendor. A poorly implemented native driver could actually be more problematic that ODBC, since there is no control over the quality of a vendor's implementation.

NSFDB2: New with Notes/Domino 7 is the option to store Notes (.NSF) data in the DB2 database. This capability is still under limited availability at the time of writing. This is an exciting development from a BI perspective. With NSFDB2, Domino 7 does not store Notes documents in a true tabular row and column layout in DB2. Instead, each document is stored in a separate record as a BLOB. This in itself does not make data access any easier, but one of the promising new NSFDB2 capabilities is DB2 Access Views (DAV). A DAV is a true native DB2 view which accesses the underlying Notes data. As Cognos and other leading BI platforms all provide native DB2 connectors, the NSFDB2 option promises fast, reliable and scalable access to Notes data stored in DB2 with ND7. A DAV must be built to support each data structure being queried. DAVs are built using Domino Designer as are other Notes design elements. DAVs are currently somewhat limited in capability, but we expect them to be enhanced in upcoming Notes releases.

Externally Staged Data: The third data access strategy is to copy the data out of the Notes database into an external relational database. This could be the best option if the Notes database is very large, other options offer poor

performance or are not available due to platform or database availability. Since data must be periodically refreshed from Notes to the relational database, this option is most viable for content that is fairly static and does not change frequently. It will clearly not support real time queries against dynamic Notes applications. By applying ETL techniques to the data staging, some metadata modeling can be done as the data is extracted.

Many of these strategies have the inherent problem that some of the data comes out of Notes as the wrong data type. The most efficient way to query Notes databases is against Notes views. Unfortunately all data elements in Notes views are formatted as strings – even numeric and date values. This is a big problem when your reports are summing numeric data, or sorting by dates. Typically this requires an addition processing step to correct data types.

In addition to the data access strategy, it is also important to understand the security implications of bringing Notes data into the BI world. Notes has a highly efficient and very effective access model which protects Notes content through multiple levels of security. This model is strictly enforced in the Notes client and by Domino based Web applications. What happens though when Notes data is extracted out of the Notes environment and into Enterprise BI applications? Cognos 8 BI also has its own security model, but the problem is that the models for Notes security and Cognos security do not directly map to one other. Where Notes security is primarily on the database and document level, Cognos security is primarily at the report level.

This brings up several important security questions. What happens when a report accesses multiple Notes databases with different Access Control Lists (ACLs)? If Notes content is sensitive, what will happen when the data moves outside of the Notes world? A report can be emailed at the click of a mouse. We mention these security issues not to offer easy solutions, but to illustrate that the security considerations of Notes Workplace BI are variable and potentially complex. For any Notes Workplace BI solution, trade-offs between performance, real time vs. background processing and secure storage of generated reports will have to be specifically addressed.

As we have seen, different BI use cases demand different strategies to access Notes data. A complete Notes Workplace BI solution will therefore support different data access strategies. Across all data access modes, the solution should provide a consistent set of metadata modeling tools. And, on top of this, it should offer world class BI capabilities: reporting, metrics, analysis and planning.

Data Modeler for Notes

Having laid out the principles and the challenges of bringing Enterprise BI and Notes together, we would like to show how we have addressed and solved these long standing problems. Sun & Son has developed of the Data Modeler for Notes for Cognos 8 BI and Cognos ReportNet. The Data Modeler brings powerful metadata modeling to Notes and optimizes query access to Notes data for Cognos BI. The Data Modeler provides the essential 'glue' between Notes and Cognos BI to provide full Notes Workplace BI capabilities.

The Data Modeler allows the organization to publish a single and consistent metadata model which hides the underlying anarchy at the database level. This metadata model makes reporting, analysis and performance management across the entire Notes application base possible and practical.

The Data Modeler is packaged as a Notes application. It provides an easy to use wizard-based interface for creating Cognos metadata models from Notes databases. The wizard interface takes the modeler through a process of progressively simplifying the complexity of the target Notes database. The result is a ready to use business user-friendly data model. Much of the 'clutter' of Notes database design elements, from a BI perspective, is focused on workflow and UI and is not important for reporting or analysis. The Data Modeler lets you select the appropriate forms and fields and apply labels and formatting options to produce data tables which are optimized for quick data access from Cognos BI and easy reporting.

The Data Modeler produces two outputs for each Notes database it processes. 1. Hidden data access views are created as query targets and are optimized for query performance. Each Notes form selected in the Data Modeler is exposed as a data table in Cognos BI. Each is represented in the Notes database as a data access view. 2. An XML definition file is created for Cognos Framework Manager. The XML file contains the complete metadata model definition and automatically maps back to the views which the Data Modeler created for data access. When the XML definition file is loaded into Framework Manager, the tables (Query Subjects) defined by the Data Modeler are ready to publish to Cognos. Data type fidelity is preserved, regardless or not whether the data access mode actually preserves data types.

If a data model needs to be changed, it is easy to simply return to the Data Modeler and rerun one of the database wizards. All previous selections are preserved, so once the changes are made, the Data Modeler automatically updates the output. This is especially valuable to Notes users who may be new to metadata modeling techniques. You can learn as you go and an

easily update the model as you better understand both the modeling process and the BI users' reporting requirements.

Once the Data Modeler metadata definition file is brought into Framework Manager you can perform a number of additional operations which further enhance the models you have created. You can set relationships between tables in the same Notes database, combine and relate models from multiple separate Notes databases, or combine and relate data from Notes databases and external data sources. Framework Manager also provides advanced data modeling tools to generate dimensional models of data, allowing drill-down and roll-up capabilities in Notes Workplace BI reports.

Once the metadata model package is published in Framework Manager, it becomes immediately available to Cognos BI users. Users can begin querying, reporting and analyzing their Notes data using the Cognos BI studios. Cognos 8 queries the data sources directly, so the Data Modeler does not in any way impede or impact data performance. In fact, the data access views are optimized for the best query performance possible.

The Data Modeler fully supports the separation of roles between data modeler and report designer. The data modeler needs to have a full understanding of the design structure of the source databases. The report designer gets to work with the business user-friendly data models created by the Data Modeler and Framework Manager and can focus on the business value of his reports.

The Data Modeler supports a number of different strategies for data access. Each of these is optimized for maximum performance through the tight matching of the data access views to the XML metadata definition file. The Data Modeler provides access models for any Notes reporting need.

NotesSQL: Data Modeler support for NotesSQL optimizes the query connection for best possible performance through the ODBC interface. Data type fidelity is preserved, and you can be accessing your data sources from Cognos in minutes.

NSFDB2: For databases running Domino 7 with DB2 you have the option in the Data Modeler to automatically create DB2 Access Views for data access. Data access through DAVs is fast and reliable, and no additional work is required. At the time of writing, ND7 NSFDB2 is still under limited availability and certain Data Modeler features are not yet supported by Lotus in NSFDB2.

Staged Data: If externally staged data is the preferred option, the Data Modeler can automate the creation of the extended data store and manage the transfer of Notes data. DB2, Oracle, SQL Server and Sybase are supported target databases. No additional effort is required. As always with the Data Modeler, data type fidelity is preserved.

You can now build BI solutions that combine all data access strategies into a single metadata model. This allows you to accommodate the varying requirements of different Notes applications individually within a larger enterprise BI environment.

The Data Modeler manages virtually all of the steps for integrating Notes with BI. At each step it applies the best practices that Sun & Son and Cognos have developed for rich and successful Business Intelligence from Notes. Some of the rich capabilities supported in the Data Modeler are:

- Subform support
- Computed column support
- Custom selection support
- Options for handling multivalue fields
- Custom field labels
- Data sorting options for optimized query execution

One word of caution: The capabilities provided by the Data Modeler are powerful, but existing Notes application structures are often messy. Some Notes naming conventions do not translate well into the BI world and metadata modeling concepts are new to most Notes developers. There is a learning curve, and the data modeling process can require multiple iterations or professional services to 'get it right'. Once the data models are set up, though, they are available to any BI needs.

The Data Modeler is available free for evaluation for 30 days from Cognos at http://www.cognos.com/solutions/data/ibm/lotus_domino.html. The fully supported production version is available from Sun & Son http://www.sunandson.com and from select business partners.

Data Modeler and Business Partners

The Data Modeler for Notes is also sold and supported by a network of IBM and Cognos business partners. These business partners understand both Notes and Cognos and can provide enablement and implementation services for Data Modeler. A current list of Data Modeler partners is available at http://www.sunandson.com. The Data Modeler can provide interesting and

lucrative opportunities for Lotus business partners interested in adding Enterprise BI to the set of services they provide to their customers. The Data Modeler can also provide a significant value add for Cognos business partners who provide services for organizations using Lotus Notes.

Sun & Son and Cognos enthusiastically support Notes ISVs interested in adding Enterprise BI capabilities to their Notes applications. Please refer to http://www.sunandson.com for more information.

Workplace BI Vision

The Data Modeler for Notes is the first installment of a strategic collaboration between IBM, Cognos and Sun & Son. We call this initiative Workplace BI. Workplace BI is the seamless integration of the Lotus collaborative capabilities of Notes, Portal and Workplace with the industry leading BI capabilities from Cognos. Traditionally BI vendors have considered their work done when a report was generated. The fact is that BI users generate reports to understand performance, identify problems or make decisions. All of these are collaborative processes. Workplace BI brings Lotus collaboration into the report itself so that the report, any report, can become an active work surface. Peoples' names become collaboration-enabled through LiveNames support, and collaborative team activities can be launched at the click of a mouse. Business Intelligence becomes in integral part of the Lotus Workplace strategy. To see examples of Workplace BI at work, take a look at the Sun & Son Active Reports for Cognos portlets at http://www.sunandson.com/workplacebi.htm.

Ravi Har Singh Khalsa
CEO & Chief Architect
Sun & Son
May 10, 2006