Extraction Transformation and Loading (ETL) Tool Selection Process

1 Summary

During the month of September the Data Warehouse group gathered and analyzed the requirements for the Extraction Transformation and Loading software. Key participants were identified and Purchasing Office was contacted to establish a selection process.

The selection committee

The following individuals participated in the selection process: David Goldschmidt, Business Systems Analyst; Ora Fish, Project Manager; Kim Morgan, Manager Database Administration, Keith Martens, Purchasing Agent; Raj Kahlon, Data Warehouse Architect; Kate Owens, Business Systems Analyst.

The group also consulted with Dennis Milkiewicz, Manager Desktop Support and John Fuchs, Systems Architect, DBA.

Request for Proposal

On October 3, 2001, the RFP was send to the following companies: Ascential Software, Cognos Corporation, Hummingbird Corporation, Informatica Corporation, Oracle Corporation. All companies were asked to reply by October 17, 2001. The presentation dates were set for the week of October 22, 2001. *Please refer to the ETL Tool RFP for more information*

Vendor Presentations

The following three companies replied to our RFP: Ascential Software, Cognos Corporation, Informatica Corporation. All three were invited to demonstrate their products as following: Cognos Corporation on October 24; Informatica Corporation on October 25; Ascential Software on November 07.

The selection process and recommendations

During the evaluation process, it became clear that Cognos Corporation could not compete in the ETL tool space based on the relative youth and limitations of their ETL tools. Cognos was unable to show support for version or revision control, security beyond that provided by the underlying database(s), or interoperability with non-Cognos products. The selection committee recommended to avoid Cognos as an ETL vendor.

The ETL tools presented by Ascential Software and Informatica Corporation were comparable in numerous ways, though the Selection Committee's recommendation was to select Informatica Corporation as our ETL vendor. Informatica seems to be more mature and stable as a company; and they seem to provide a more mature and stable product. Informatica's solution seems simpler to use and still with a large amount of flexibility and power. While Informatica Corporation uses a pricing model that charges for each additional source and target, Ascential Software seems to cost more overall. Given that we primarily have Oracle, Microsoft Access, and flat files as sources and Oracle as a target, the Informatica cost will be fairly fixed.

References

Once DataStage from Ascential Software and PowerCenter from Informatica Corporation were identified as key potential ETL solutions, the selection committee has contacted several companies and Universities. In overall all references seemed to be quite satisfied with vendor's ability to respond and address their needs. All seems to be satisfied with the software performance and the ability to cater to their needs.

2 Product Evaluation Notes

Throughout the following sections, each of the vendors and their ETL products are evaluated, focusing on primary differences between such products.

2.1 Company/Product Philosophy

- **Ascential Software**, formed in July 2001, is focused on improving, developing, and perfecting their ETL and "back-end" tools; they do not have current plans of entering the BI tool market.
- **Cognos Corporation**, founded in 1969, seems to prefer that all components of the enterprise data warehouse are Cognos products; in other words, their ETL tools easily integrate with Cognos BI tools, etc., but have difficulty integrating with other vendor products.

  Even if we end up choosing Cognos as part of our BI tool solution, their ETL tool does not measure up to the other competitors. Further, being tied to a single vendor due to technical limitations increases the risk of failure.
- **Informatica Corporation**, founded in 1993, serves a large and impressive portion of the market. Informatica produces both ETL and BI tools, though they also support open APIs for metadata exchange.

2.2 Design/Development Environment

Products of **Ascential Software** and **Informatica Corporation** provide a number of design and development GUIs that have substantial built-in functionality. Ascential's tool seems more flexible than Informatica's based on Ascential's proprietary BASIC-like programming language; however, flexibility is attained through coding that may easily become a nightmare to maintain.

Both tools support reusability via a shared folders mechanism; both tools also promote top-down design via a graphical icon-based design palette.

Ascential's product seems slightly more comprehensive and powerful than that of Informatica's; however, Ascential's product consists of a development environment that is spread between at least 4 separate desktop applications, each requiring a separate login. This seems a bit daunting and perhaps somewhat cumbersome, though every new tool requires training, patience, etc.

2.3 Version Control

- **Ascential Software** provides a separate Version Control (VC) application that is used to label every imaginable data warehouse component with a version; the separate VC application aids in the migration from development to test to production. Note that Ascential has admitted that this is not optimal and is in the process of redesigning their VC component.
- **Cognos Corporation** does not support version control.
- **Informatica Corporation** addresses version control through the Repository Manager that allows archiving any object contained in a folder into subsequent version each time the development landmark is reached. All versions can be reverted to at any point of time. Promoting transformation from development to production is achieved through copying objects or folders of objects within a repository or across repositories.

2.4 Security

Both **Ascential Software** and **Informatica Corporation** provide products that rely on underlying database security.

Informatica's product supports a user/group model, though the granularity of permissions goes no lower than the folder level; Ascential's product is similar, though each of their separate applications seems to expect a separate login and password, which seems not only frustrating for a user, but difficult to administer and manage.

2.5 Performance and Scalability

As to be expected, all vendors claim tunable performance and scalability. All of the presented ETL tools seem architecturally flexible enough to support performance tuning, etc.

2.6 Adherence to Metadata Exchange Standards

ETL products of **Ascential Software** and **Informatica Corporation** both claim interoperability with many BI and other tools. The mechanism used to achieve such integration, as well as the philosophy behind each approach greatly differs:

- **Ascential Software** provides the MetaStage product to manage metadata. Supported in part by this product, Ascential's philosophy is to keep up with key BI tool vendors and systems such that Ascential's products properly interoperate with each system. This does not imply a partnership with each vendor and therefore could imply that the relationship between Ascential's product and BI tools is not a sound relationship. With the onus entirely on Ascential to maintain integration functionality between numerous vendors, a debilitating risk for failure seems to exist. It raises a potential question as to how Ascential is able to keep up with multiple vendor activities, especially if no explicit corporate partnership exists. This question seems to be further supported by one of the contacted references, Karen Romano's version compatibility issues.
- **Informatica Corporation** provides an open API to its metadata, allowing corporate partners to pull metadata from its API in support of BI and other tools. The responsibility of maintaining the API belongs to Informatica, while the task of pulling metadata from the API is the responsibility of the BI tool vendors.

2.7 Product Cost

While **Informatica Corporation** uses a pricing model that charges for each additional source and target, **Ascential Software** seems to cost more overall. Given that we primarily have Oracle, Microsoft Access, and flat files as sources, and Oracle as a target, the Informatica cost will be fairly fixed.

3 Reference Checks

3.1 Ascential Software

After participating in Ascential Software's presentation, we have contacted the following references:

- **Natalie Vincent, Syracuse University**
- **Karen Romano, Carrier Corporation**
- **Victoria Shaffer, Virginia Tech**

3.2 Informatica

The following references were contacted:

- **Kevin LaVallee , Capital District Physicians Health Plan, Albany**

- **Tim Porzio, Data Warehouse Project Manager, Sodexho USA, Buffalo**
- **Joe Sullivan, Project Director, University of Minnesota, Minneapolis -**

3.3 Summary of the variety of technical Forums:

Informatica(Power Center) and Ascential(Data Satge) usually cater to all the essentials features that a typical ETL tool should have, it depends upon the environment which tools fits best in terms of source and targets , price and support.

4 ETL Tool Recommendation

During the evaluation process, it became clear that Cognos Corporation could not compete in the ETL tool space based on the relative youth and limitations of their ETL tools. Cognos was unable to show support for version or revision control, security beyond that provided by the underlying database(s), or interoperability with non-Cognos products. The Selection Committee recommendation was to avoid Cognos as an ETL vendor.

The ETL tools presented by Ascential Software and Informatica Corporation are comparable in numerous ways, though our recommendation is to select **Informatica Corporation** as our ETL vendor. Informatica seems to be more mature and stable as a company; and they seem to provide a more mature and stable product. Informatica's solution seems simpler to use and still with a large amount of flexibility and power.