

## **Data Warehousing Tool Evaluation – ETL Focused**

Nestor Rodriguez  
University of Texas-Pan American  
1201 W. University Drive, Edinburg, TX  
(956) 665-UTPA  
nrodriguez21@broncs.utpa.edu

Kent Lawson  
University of Texas-Pan American  
1201 W. University Drive, Edinburg, TX  
(956) 665-UTPA  
kelawson@broncs.utpa.edu

Eddie Molina  
University of Texas-Pan American  
1201 W. University Drive, Edinburg, TX  
(956) 665-UTPA  
ermolina89@yahoo.com

Jeremiah Gutierrez  
University of Texas-Pan American  
1201 W. University Drive, Edinburg, TX  
(956) 665-UTPA  
jgutierrez10@broncs.utpa.edu

## **Abstract**

This paper evaluates three different ETL tools, one of them is open source and the other two are proprietary. The purpose of the evaluation is to be able to make an informed recommendation as to which piece of software we would recommend based on each tool's strengths and weaknesses. The evaluation is based on four criteria that are relevant to any business in the process of deciding on what tool to implement. Using the evaluation, we compare the three tools and provide our recommendations.

## **Introduction**

The importance of being able to turn raw data into useful information has become a necessity for many emerging and established companies. In the business world we often hear the statement that it's not all about the mighty dollar. However, a common goal is to gain that competitive advantage over competitors. Since the process can be a costly and complicated, companies are looking for the most user-friendly and cost effective methods possible. These two criteria are often not related in the software industry. The goal is not only to be able to utilize Business Intelligence, but also to do it effectively. "Business intelligence, or BI, is an umbrella term that refers to a variety of software applications used to analyze an organization's raw data. BI as a discipline is made up of several related activities, including data mining, online analytical processing, querying and reporting (Mulcahy, 2011)." The main goal of Business Intelligence is to have correct and accurate information in order to make effective decisions.

Business Intelligence gives companies the ability to transform data into usable information, thus taking seemingly useless data and turning it into valuable information. According to Brian

Larson (2008) “Effective decisions are choices that move an organization closer to an agreed-on set of goals in a timely manner.”

Business Intelligence is a tool that can be utilized to help inform decision makers in this process. We will be looking in particular at ETL i.e. Extract, Transform, and Load process. This is a process that takes data from other sources and transforms that data to a common type, then loads that data to the data mart. We have focused on this process because the ability to take several forms of data that may be in different formats and merge them all into one source that is coherent is quite powerful. Our criteria are: Cost, Reusability, Ease of use, and Functionality. The tools that we have decided to evaluate are SAS Data Integration Studio, Clover ETL, and IBM Information Server (Datastage). Our research was limited by not having the tools at hand to personally evaluate. Our research was focused around journals, company websites, and research papers.

### **Overview of ETL Process**

When looking at ETL as a Data Warehousing tool ETL offers Extraction, Transformation, and Loading. ETL tools extract data from different sources and then clean the data and make it uniform in the transformation process. The finished product is then loaded into the data mart. ETL is a valuable process because companies have data coming from many different sources and formats. The data is merged into the data mart giving the decision makers the power to look through the data in a way that’s not possible when in their separate locations. This will increase

the ability of decision makers to be able to sift through data and make informed business decisions.

### **Evaluation Criteria**

When deciding on our evaluation criteria we decided to look at Cost, Reusability, Ease of use, and Functionality. The reason we believe these criteria are important is when looked at collectively it would give the decision maker a great foundation to make the best informed decision. These four criteria cover key success factors that relate to the effectiveness of the tool. We will be comparing all three tools against the four criteria, with the goal of being able to make a solid recommendation as to which tool would be best.

### **SAS Data Integration Studio**

**Ease of use:** SAS is known for having some of the best Business Intelligence tools in the market when compared to similar tools from Microsoft, Congos and TerraData among others ( Ta'a, Abu Bakar, and Saleh 4). This reputation has been built on the ease of use that this tool provides. SAS Data Integration Studio is worked by a point-and-click graphic user interface or GUI with customizable windows. This allows the user to manage data integration from beginning to end without unpleasant lines of code. This tool is very visual and allows for end-to-end design of processes, reducing time in development. The drag-and-drop features save users from learning programming codes or rules associated with older technology, wizards help users in many tasks such as creating target structures, building and executing ETL process flows and import and export metadata among others. The GUI is full of color and provides an attractive

view of the work being created, managed or being worked on ("SAS® Data Integration Server" 4-6).

**Cost:** Although the cost for such a tool could not be found during the time provided for this paper, the benefits outrank the cost to implement this kind of ETL tool. For this research paper a SAS online agent was contacted for pricing information and upon consistent pricing questions, the agent said that no information can be given out on the price of this tool. The online agent also mentioned that no pricing information can be found online due to “the capabilities of SAS are ...broad; from simple statistics to complex solutions across the entire enterprise. Because of this we are not able to provide pricing. (Web Chat)”

**Functionality:** “SAS software has the ability to get at just about any data source in the world with a single point of IT administration. From relational database systems to legacy mainframe systems to flat file and proprietary applications, SAS can either access them natively (native access engines) or through any number of methods including ODBC/JDBC or open standards supported with SAS® Integration Technologies (Grasse, and Nelson 4 2006).” SAS Data Integration Studio has many capabilities for companies looking to build business intelligence with the data currently held within or data that has not been gathered yet. Among the capabilities that this tool provides are “enterprise data access and processing across systems and platforms; integrated data quality, which is critical to producing accurate, consistent information; and an interactive, visual data integration development environment that enables collaboration (SAS The Power to Know 2011)...”

**Reusability:** One great feature that SAS provides is immediate reuse of metadata that was captured and documented throughout the transformation and data integration processes ("SAS®

Data Integration Server" 1-2). The collaboration provided by this ETL tool gives it easy reusability. By automating processes such as change data capture or messaging queues, SAS can make organizations more resourceful. An advantage of having a tool like SAS Data Integration Studio is that since organizations pay for this software, the principle of reusability is supported. On the other hand, many of the open source ETL tools don't have this feature (Murray 2011). This feature can allow others within an organization to reuse any work done by others and share the information to reduce maintenance and development time.

### **Oracle Warehouse Builder (OWB)**

**Ease of Use:** "OWB gives the user a highly graphical and easy-to-use interface, allowing users to model the ETL process by dragging and dropping objects on a mapping canvas. " (Singapore Stocks and Shares 2011) OWB is based off a set of GUIs for better and easier solutions of data integration. The user creates a set of objects that are then stored as metadata in a repository known as a "workspace", which is hosted off an Oracle database. The primary interface is the Design Center; from here a user can import source objects, and create ETL processes such as mappings. These mappings are how you define the flow of data from sources to targets. OWB then generates the code necessary to implement the ETL logic. Essentially, this software seems rather user-friendly, which is a major advantage for those already familiar with software similar to OWB and cuts out the time for training users.

**Cost:** It was discovered in this criterion that OWBs functionality is divided up into five feature groups, which in reality means – more licensing costs. The groups are listed as Basic ETL,

Enterprise ETL, and Application Adapters for OWB, OWB Data Quality and Data Watch and Repair. Each group requires a different license than the other, which means pricing can get out of hand quickly. For example, the Basic ETL (this is limited to building simple data warehouses) functions are available with the Oracle Database license, while Enterprise ETL (allows advanced enterprise functionality) requires the Oracle Data Integrator Enterprise Edition license. Although, for companies who already develop with Oracle don't need to pay extra licensing costs. Essentially, this makes OWB a "license free" ETL tool for companies that previously have a license for Oracle Enterprise Edition and that work with an Oracle partner. (Ribaud 2005)

**Functionality:** "Warehouse Builder's feature set facilitates highly productive development for data warehousing and BI solutions on Oracle..." (Oracle 2011) Some of these said features include but are not limited to: pluggable mappings for reusable ETL logic, Oracle Data Integrator knowledge module technology for extensible code generation, Extensibility, scripting and automation features, Built-in change management/versioning, Multiple configuration management for development, testing, production life cycle, Import Metadata (source non-Oracle and Flat Files), Target Oracle (Relational and Dimensional) and Flat Files, Debugging. Another great feature that was added to a recent version of OWB is the ability to design relational and OLAP database structures. Given this feature now makes it easier to store common data onto an Oracle Database repository, after which users can use BI applications and decision support tools to yield information in the right format. (InvestCenter 2011) Overall, since Oracle is a leader in Database and Data Warehousing, it's no surprise that they would invest heavily into making an ETL with great capabilities.

**Reusability:** A primary function of an ETL is to essentially transform data. OWB provides many methods of custom transformation of source data. These custom transformations include procedures, functions, and packages. A user is also allowed to reuse PL/SQL and to write custom PL/SQL transformations, which can then be used in OWB mappings. The major advantage to this is that a majority of relational database management systems support SQL. Meaning, that all programs written in SQL can be moved from one database to another with hardly any modifications. “OWB is also an extensible data integration and data quality solutions platform. OWB can be extended to manage metadata specific to any application, and can integrate with new data source and target types, and implement support for new data access mechanisms and platforms, enforce your organization’s best practices, and foster the reuse of components across solutions.” (AQIES 2011)

### **Clover ETL Community**

**Ease of use:** The Clover ETL products are designed by the Javelin group, which focuses on the developing software for the purpose of managing complex data solutions for small, medium or large businesses. The CloverETL Community Edition includes a free designer which makes it easy to use and looks similar to many full priced designers. The current version of CloverETL Community comes with a Graphic User Interface (GUI), which was not available with the Community edition previously. In the past, the Community Edition used a command line style prompt to create and design data management projects which made it difficult to use and not a popular tool for designers to use. The new GUI mirrors the design options and structure that many of the paid license versions of the CloverETL products as well as competitor versions.



**Cost:** The Community Edition is based on Clover's Open Source transformation engine and is available at no cost to consumers. The Community Edition, although free, does not include many of the extra features that the other packages in the CloverETL family. CloverETL also offers pay versions of their products which include Desktop, Enterprise Standard, Enterprise Corporate, and Cluster. The package that we are evaluating is CloverETL Community Edition which is Clover's open source free package. The Desktop version, according to [www.cloveretl.com](http://www.cloveretl.com), is listed at \$3995 with the version with more features prices not listed. According to [www.cloveretl.com](http://www.cloveretl.com), the features available in the Community Edition are the CloverETL Designer and the ability to run graphs on a PC/Mac/Linux desktop which are far fewer than the pay versions.

**Functionality:** CloverETL Community is Java-based and has been successfully deployed on the following Operating System platforms: Linux both 32 & 64 bit), Windows (both 32 & 64 bit), HP-UX, AIX, AS/400 (IBM System I), Solaris, and Mac OS X. The Community edition contains connectors for the following data sources: text file delimited, fix-length and combined, XML, XLS, RDBMS through JDBC, WebServices through REST/SOAP protocols, JMS, LDAP, dBase/FoxBase/FoxPro, bulk-loaders for Oracle, DB2, MS SQL, Informix, MySQL and PostgreSQL, and QuickBase. Since the Community Edition does not include many of the features that the paid versions do, the functionality of the Community Edition is very limited.

**Reusability:** The Community Edition allows for the access of transformation components that allows for common data transformation tasks such as reformatting, filtering and sorting data. Components for aggregating, merging or duplicating data are also available for use with the Community Edition. Community Edition allows for reading and writing from MYSQL, PostgreSQL, SQLite, MS SQL, Sybase, Oracle and Dervy databases. The Community Edition

also works with CSV and other flat file data files, as well as Excell worksheets, XML file and dBase files.

### **Comparative Analysis**

Upon evaluating three different ETL tools, a comparative analysis was constructed to determine each tool's strengths and weaknesses. A one (1) in the table below means that the criteria is a positive attribute for the ETL tool and a zero (0) in the table represents a negative attribute.

These scores are assigned on what we think was positive or negative according to the research constructed using previous works in the same field. We compared one tool to the other two tools and made an informative decision to distinguish if an attribute was appealing or not. The tool with the highest attribute total is the tool that would be recommend to an organization that plans to implement this type of business intelligence software. In the table below the aggregate totals are displayed leading SAS Data Integration Studio and Oracle as a tie. Clover ETL Community being the freeware version of the other suites made by Clover fell short in two out of the four criteria.

**Comparative Analysis Table**

|                             | Cost | Functionality | Reusability | Ease of Use | Total |
|-----------------------------|------|---------------|-------------|-------------|-------|
| SAS Data Integration Studio | 0    | 1             | 1           | 1           | 3     |
| Oracle (OWB)                | 0    | 1             | 1           | 1           | 3     |
| Clover ETL                  | 1    | 0             | 0           | 1           | 2     |

## **Conclusion**

Following our research, the two tools that scored highest would be the ETL tool recommended by our team. The organization's budget will be the deciding factor for the decision makers. Our recommendations are based on this comparative analysis; this recommendation is done by using only the research obtained from previous research papers and not from actual hands-on tool evaluation. For these reasons, there could be a different set of opinions once the tool is implemented.

## Work's Cited

"AQIES Technologies -Blog » Blog Archive » Oracle Warehouse Builder Interview Questions." *AQIES Technologies -Blog*. Web. 10 Oct. 2011. <<http://techblog.aqies.com/2011/03/30/oracle-warehouse-builder-interview-questions/>>.

"Defining Custom Transformations." *Oracle® Warehouse Builder Data Modeling, ETL, and Data Quality Guide 11g Release 2 (11.2)*. Web. 12 Oct. 2011. <[http://download.oracle.com/docs/cd/E11882\\_01/owb.112/e10935/custom\\_transforms.htm](http://download.oracle.com/docs/cd/E11882_01/owb.112/e10935/custom_transforms.htm)>.

"ETL Tools for Data Warehouses ~ ORACLE DATAWAREHOUSE." *ORACLE DATAWAREHOUSE*. Web. 8 Oct. 2011. <<http://oracle-datawarehousing.blogspot.com/2011/02/etl-tools-for-data-warehouses.html>>.

Grasse, Danny, and Greg Nelson. "Base SAS® vs. SAS® Data Integration Studio: Understanding ETL and the SAS Tools Used to Support It." 31st Annual SAS Users Group International. (2006): 4. Print.

Mulcahy, Ryan . "Business Intelligence Definition and Solutions." CIO.com. N.p., n.d. Web. 10 Oct 2011. <[http://www.cio.com/article/40296/Business\\_Intelligence\\_Definition\\_and\\_Solutions](http://www.cio.com/article/40296/Business_Intelligence_Definition_and_Solutions)>

Murray, John. "Sharing experiences with ETL tools." civiCRM. Creative Commons Attribution-Share Alike 3.0, 06062011. Web. 6 Oct 2011. <<http://civicrm.org/blogs/joe-murray/sharing-experiences-etl-tools>>.

"Oracle Announces General Availability of Oracle(R) Warehouse Builder 10g Release 2 | InvestCenter." *Stock Market | InvestCenter*. Web. 9 Oct. 2011. <[http://markets.financialcontent.com/stocks/news/read/129933/Oracle\\_Announces\\_General\\_Availability\\_of\\_Oracle](http://markets.financialcontent.com/stocks/news/read/129933/Oracle_Announces_General_Availability_of_Oracle)>.

"Oracle Warehouse Builder." *ETL Software Tools*. Web. 12 Oct. 2011. <<http://www.etltools.net/oracle-warehouse-builder.html>>.

"Oracle Warehouse Builder - Frequently Asked Questions." *Singapore Stocks and Shares*. Web. 12 Oct. 2011. <<http://sgstocks.tripod.com/oracle-007.htm>>.

Ribaud, Denis. "Oracle Warehouse Builder (OWB): the Alternative ETL Tool." *IRIX Software Engineering AG - Newsletter* Mar. 2005. Print.

"SAS Data Integration Chat." Interview. 10102011. Web.  
<[https://server.iad.liveperson.net/hc/42179880/?&cmd=file&file=chatFrame&site=42179880&channel=web&d=1318280426350&referrer=\(engage\)http://www.sas.com/software/data-management/data-integration/index.html&t=1&sessionkey=H7846093423733038655-645085452607327201K6237038](https://server.iad.liveperson.net/hc/42179880/?&cmd=file&file=chatFrame&site=42179880&channel=web&d=1318280426350&referrer=(engage)http://www.sas.com/software/data-management/data-integration/index.html&t=1&sessionkey=H7846093423733038655-645085452607327201K6237038)>.

"SAS® Data Integration Server." A complete solution designed to meet your data integration needs. 2009. Web. 6 Oct 2011. <<http://www.sas.com/resources/factsheet/sas-data-integration-server-factsheet.pdf>>.

"SAS Data Integration Studio." SAS The Power to Know. SAS, 2011. Web. 6 Oct 2011. <<http://www.sas.com/software/data-management/data-integration/index.html>>.

Ta'a, Azman, Muhamad Shahbani Abu Bakar, and Abdul Razak Saleh. "ACADEMIC BUSINESS INTELLIGENCE SYSTEM DEVELOPMENT USING SAS® TOOLS." SAS Global Forum. 508 (2008): 4. Web. 10 Oct. 2011. <<http://www2.sas.com/proceedings/forum2008/058-2008.pdf>>.

"Warehouse Builder Feature Groups and Licensing." *Oracle | Hardware and Software, Engineered to Work Together*. Web. 8 Oct. 2011. <<http://www.oracle.com/technetwork/developer-tools/warehouse/overview/licensing/index.html>>.