QUICK TAKE

September 6, 2006

Oracle Is Off To A Solid Start In Information Quality With Oracle Warehouse Builder

by J. Paul Kirby
with Keith Gile, Connie Moore, and Lucy Fossner

EXECUTIVE SUMMARY

The latest release of Oracle Warehouse Builder (OWB), OWB 10g Release 2, includes significantly more information quality capabilities than previous releases. Although OWB is primarily an extraction, transformation, and load (ETL) tool, customers can use its information quality functionality without using OWB for ETL. OWB does not offer all features of the market leaders but will be attractive to customers because of its functionality and price. Enterprise architects should include OWB on their list of information quality (IQ) offerings to evaluate as part of enterprise information quality initiatives.

TARGET AUDIENCE

Enterprise architecture professional, information management professional

ORACLE MAKES A SOLID MOVE INTO THE INFORMATION QUALITY MARKET

Oracle now occupies a place at the table among IQ software vendors. The recent release of the Oracle 10g Release 2 database is accompanied by a new version of Oracle's ETL product, OWB. Aside from improvements in the ETL area — such as support for packaged applications (like SAP R/3, Oracle's E-Business Suite, and PeopleSoft), ease of use, and data lineage for items defined in the OWB repository — OWB adds new information quality functionality. There are genuine gaps in the product's functionality — like cross-table matching and comprehensive multilanguage data — but its capabilities will satisfy many organizations' needs, at a relatively low price point, too.[1]

OWB's IQ features hit many of the right targets for information quality, including:

- **Basic IQ cleansing.** OWB offers format standardization based on patterns and custom business rules implemented in easy to understand transformation logic. OWB also offers a native interface for name and address cleansing provided by some of the market leaders in IQ (requiring a license for the third-party products).

- **Multi-algorithm matching.** OWB matches records together for deduplication and subsequent cross-table joining using both rules and tunable mathematical algorithms which create scores.

- **Ongoing monitoring.** IT can create data auditors that assess data compliance with business rules and send alerts when errors exceed configurable thresholds. Auditing results are computed rule

by rule, displayed either as the percentage of records which passed or as a Six Sigma score. Auditing can be scheduled within OWB or invoked by a third-party scheduler.

- **Integrated profiling and cleansing.** The profiling and cleansing functions are configurable within a single GUI framework and presented as a series of integrated tasks, properly reflecting the fact that profiling and cleansing are two faces of the same coin.

- **Data profiling.** OWB is the most advanced in data profiling. Along with generating information about data types, distinct values, and empty columns, OWB will infer patterns, compare information against built-in formats, assess compliance with business rules, and discover the relational integrity of columns and tables. Profiling results are first displayed in text or graphs; users can drill down to the row level to see the actual data underlying a specific result. Results can be exported in CSV and HTML format, as well as compared against historical profiling sessions.

OWB is different from the market leaders in how its IQ functionality approaches metadata. Because OWB is primarily an ETL tool, a single metadata repository is used by ETL and IQ functionality alike — creating efficiencies for organizations that want to use OWB for both. However, this will require a shift in perspective for those that only want to use OWB for IQ. Customers aren't forced to use OWB for ETL; they just have to create basic IQ metadata in an ETL context.

OWB HAS BOTH OPPORTUNITIES AND CHALLENGES WITH IQ

Oracle will need to put several more pieces in place before going head-to-head against IQ market leaders. Specific challenges include:

- **Cross-table matching.** The most difficult task for information quality software is determining which records pertain to the same entity (e.g., which records describe the same customer or which customers live in the same household). Sometimes records can be deduplicated or householded within a single table, and OWB will do this. However, complex information quality problems involve different tables pertaining to the same entity (often from multiple systems) that must be cleansed in unison. This OWB will not do, but there are two workarounds. First, you could join multiple tables into a single data stream, table, or view before matching. Or, you could match one table in isolation, and then cleanse another table in isolation using the fully cleansed data from the first.

  However, if multiple tables could be consolidated so readily, there would not be much need for matching in the first place. Comprehensive cross-table matching is central to IQ, adding value to master data management (MDM) and customer data integration (CDI) efforts, making matching a key area for needed improvement in future OWB releases.

- **Comprehensive multilanguage data.** OWB is UTF-16 compliant, but multilanguage IQ means providing for multiple languages, countries, and cultures at a very detailed level (e.g., custom address rules for each country and specific language parsers). Although OWB can support this through its integration with other information quality software, Oracle needs to offer a standalone solution to be a market leader.

- **Reference data and lookup tables.** Cleansing data means more than correcting its format; it also means correcting its content. Using patterns and business rules is powerful, but using actual reference data to correct data is better, particularly for names and addresses. In future releases of OWB, Oracle should natively provide for arbitrary types of reference data.

- **Automatic inference of metadata.** Specifying how an item should be cleansed should be better automated. Even the leaders can improve on how much metadata can be inferred automatically, but OWB needs to catch up.

Shortcomings though they may be, the fact that the list of major gaps isn't longer is great news for Oracle customers and prospects. OWB isn't a toy; its functionality base is solid.

OWB's Information Quality Functionality Is A Bargain

OWB's basic information quality functionality, along with standard ETL functionality, comes at no additional cost to a database license. This includes OWB's native name and address interface, as well as matching functionality, for example. The advanced Data Quality Option, including profiling, business rules, and auditing, costs $15,000 per CPU. Separate licenses aren't required for each data *source*, just for the data repository into which the metadata for multiple sources flows.

All in all, the basic level of functionality is a genuine bargain for not having additional cost, and the level of investment for OWB's advanced features, although tangible, is not substantial considering the amount of functionality provided.

RECOMMENDATIONS

INCLUDE ORACLE IN INFORMATION QUALITY VENDOR EVALUATIONS

Proofs of concept are the modern standard for evaluating vendors in information quality. Customer prospects need to know that performance, scalability, and raw functionality will meet their requirements before they invest. Be sure to include the following in your evaluations:

- **Name and address cleansing.** The venerable use for IQ software has been name and address cleansing, which OWB offers only through an interface to separately licensed third-party products. Determine whether your name and address cleansing needs justify the additional investment.

- **Cross-table cleansing.** OWB's multitable cleansing is strongest around sequentially cleansing each table in isolation. If you need to cleanse data *across* multiple tables at once, verify that workarounds are manageable.

- **Business rule complexity.** Pick several of your most complex rules, and verify that OWB can handle them.

- **Business user friendliness.** Compare how tech savvy your business users are to the ease of use of OWB's GUI.

W H A T   I T   M E A N S

ORACLE IS OFF TO A SOLID START IN INFORMATION QUALITY

Oracle is wise to invest in IQ; combining data integration and IQ in a single product adds value to both. Moreover, IQ is a key component of a modern information infrastructure platform. The fact that all the point players in the IQ software market are being acquired by platform vendors — Firstlogic by Business Objects, Ascential Software by IBM, and Similarity Systems by Informatica — is further evidence that Oracle's investment should continue.[2]

ENDNOTES

[1] In Q1 2006, Forrester assessed the state of the information quality software market to see how the vendors stacked up against each other. The results: Harte-Hanks Trillium Software, Firstlogic, and IBM are the market Leaders for their comprehensive, highly scalable data cleansing and profiling capabilities, including support for international customer-focused data and the ability to cleanse noncustomer data. See the January 17, 2006, Tech Choices "The Forrester Wave™: Information Quality Software, Q1 2006."

[2] Business Objects' acquisition of Firstlogic for approximately $69 million and Informatica's acquisition of Similarity Systems for approximately $48 million added information quality pieces to the acquiring vendors' data management puzzle. See the April 27, 2006, Quick Take "Business Objects And Informatica Acquire Information Quality Software Vendors."