Oracle Business Analytics

Business Intelligence

Enterprise Search

Analytics

Martin Butler
August 2013

Please note: This is an independently produced review of Oracle technologies. Please do not distribute or reproduce this document without permission (one hard copy may be produced per downloaded document).

Verdict

Overall, Oracle Business Analytics technologies compete well with best-of-breed options. Oracle Endeca Information Discovery was the best-of-breed option as far as we were concerned prior to Oracle acquiring Endeca. This is unique technology and worthy of serious consideration as a solution to the enterprise search problem.

Oracle's BI products represent a mature, and very large suite of solutions. However Oracle has recently added significant data visualization capability, and it compares well with newer market entrants in this space. For organizations committed to Oracle this is the obvious solution to BI needs.

Analytics, and specifically data mining and predictive analytics have been treated very differently by Oracle. Most analytics tools need data extracted from various data sources, and the resulting models stand as separate entities outside the production environment (although this can be addressed to some extent by PMML). Oracle Advanced Analytics puts the data mining algorithms into the database, along with resulting models. There are many advantages associated with this approach – not least the fact that the models are stored in a place where they can be used. Oracle has also made a significant commitment to using open source R in addition to its own data mining tools. This is a very good option for committed Oracle users, and allows them to take advantage of existing investments. It really wouldn't make a lot of sense for such users to look elsewhere.

In addition to the data and software architectures deployed by Oracle the company offers specialized hardware for analytics, BI and enterprise search purposes. It also offers connectors to Hadoop based big data sources. The configurability of these options will be of interest to many large organizations.

Introduction

It should be self-evident that there is a fundamental shift taking place in the way organizations use information technology. The era of process automation, which was characterized by transaction based systems, is being complemented by a new generation of information oriented applications – business intelligence, enterprise search and analytics.

While in reality the dividing lines between these three application types are not well defined (and will probably become less well defined) we can summarize them as follows:

Business Intelligence is primarily concerned with the processing of data so that it becomes a useful aid in decision making, performance monitoring, trend analysis and many other activities where information is needed to address business uncertainties. The input data is usually structured and most often derived from transactional data.

Enterprise Search typically addresses both structured and unstructured data, and is a facility for retrieving knowledge rich information. The sources of such information typically includes email, documents, transactional data and possibly external resources such as web sites.

Analytics primarily concerns itself with the discovery of patterns in data which might be useful to improve the efficiency and efficacy of business operations. Classic examples include more accurate targeting of customers, fraud detection, prediction of machine failure, and even swings in sentiment before they happen. Analytics address structured and unstructured data, the latter through text mining. Big data is primarily concerned with the provision of a supporting infrastructure so that very large amounts of data can be analysed at speed, and data mining techniques provide the analytical tools.

While this may seem like a disparate set of technologies and requirements, it is not. Data is fundamental to transactional and information oriented applications, and a unified data architecture is essential if complexity and inefficiency are to be avoided. And the effects of complexity should not be underestimated, it is the single most common cause of chronic information systems related problems.

These issues introduce an age-old problem. Is it better to single source everything, so that the buck stops fairly and squarely with one supplier, or is it better to go for best-of-breed? Both have advantages and disadvantages. Best-of-breed may introduce functionality that cannot be gained any other way, but it also introduces complexity. A single supplier simplifies matters considerably, but then there is the problem of lock-in. In the end it's a matter of choosing your poison, and experience shows that many large organizations prefer to deal with a single supplier for core functionality, and to add satellite applications if needed.

The rest of this document assumes the organization has an investment in Oracle applications and infrastructure, and is considering implementing information oriented systems. Emphasis is on the data mining and predictive analytics products, since this is a new domain for many organizations.

Oracle's technologies are described, with comments on the quality of the offerings where appropriate.

Oracle Advanced Analytics

Oracle has taken a different approach to analytics. The central philosophy can be summarised as: Instead of taking data to the analytics algorithms, take the algorithms to the data. As a result the analytics capability is embedded in the database. Well Oracle did start life as a database company and this is definitely a different way of thinking about the problem of analytics – specifically data mining, text mining and predictive analytics.

Being different is one thing, being better is another. The approach adopted by Oracle does have some very significant benefits. The most important is the fact that predictive models sit in the very place where they can be used – in the database. Any application that calls data from a database can just as easily call a predictive model, or some other analytics task.

> Instead of taking data to the analytics algorithms, take the algorithms to the data. As a result the analytics capability is embedded in the database.

At the present time data mining is seen as a fairly esoteric activity. Certainly the analysts who use this technology needs specialized knowledge. But as with all things, what seems exotic sooner or later becomes commonplace, and the use of predictive models will, probably within five years, become a commonplace activity. Clearly Oracle is already well positioned for this by integrating the algorithms and models into the database.

Oracle Advanced Analytics (OAA) provides two primary mechanisms to support analytical activity:

1. Oracle Data Mining is SQL based, and the actual modelling environment, Oracle Data Miner, comes as an extension to Oracle SQL Developer. It supports most of the usual data mining algorithms, and support vector machines feature quite strongly.
2. Oracle R Enterprise (ORE) extends database functionality with a library of R functions and makes database tables and views available to the R environment as native R objects. Several notable features include parallelized neural networks, the scoring of database tables, support for time series analysis, and the persistence and management of R objects in the database. Open source R packages can also be incorporated into the environment.

Analytics Algorithms
The in-database algorithms available from SQL and R include:

Problem	Algorithms
Classification	Logistic Regression Decision Trees Naïve Bayes Support Vector Machine Neural Networks
Regression	Multiple Regression Support Vector Machine Stepwise Linear Model
Anomaly Detection	One Class SVM
Attribute Importance	Minimum Description Length
Association Rules	Apriori
Clustering	Hierarchical K-Means Hierarchical O-Cluster
Feature Extraction	Nonnegative Matrix Factorization

Time series are newly supported in ORE 1.3. This facilitates aggregation and moving window analysis of time series data, with date arithmetic and integration with R time series packages such as fts, tseries, xts, zoo, forecast and arima.

Oracle Data Miner

It looks like yet another graphical, drag and drop model building environment – and indeed it is. What makes this tool particularly easy to use is the way that data, models, workflows and connections are handled, and presented to the user. It's highly productive, and as all analysts know, the organization of data and models is key to model building. This is not a linear process and an organized environment is critical.
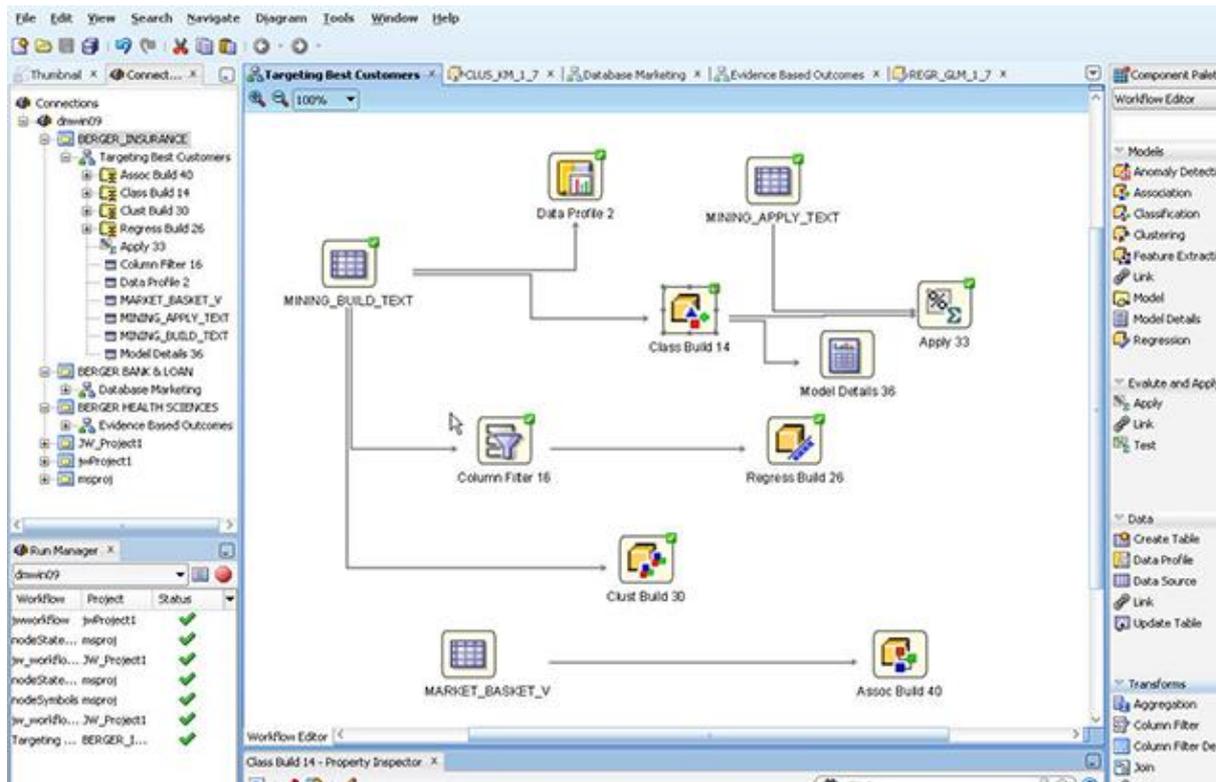
Oracle R Connector for Hadoop

The Oracle R Connector for Hadoop provides access to a Hadoop cluster from R. It can be used on the Oracle Big Data Applicance or on non-Oracle Hadoop clusters. This is part of the Oracle Big Data Connectors suite.

Oracle and R

It sounds like a contradiction in terms - an oxymoron. The bespoke with open source. Despite this, Oracle is clearly trying very hard to accommodate R in all its aspects. New releases of Oracle R Enterprise (ORE) focus on greater compatibility with, and exploitation of R's capabilities. The details can be found at Oracle's web site, and we may produce a longer paper on this topic.

Oracle Business Intelligence

Business Intelligence typically addresses the routine reporting needs in an organization, although the reporting formats may be tabular, graphical or based on charts. Oracle provides a large suite of solutions in this space and will satisfy even the most demanding of needs.

Oracle BI Foundation Suite encompasses enterprise reporting, dashboards, ad-hoc analysis, multi-dimensional OLAP, scorecards and predictive analytics in one integrated platform. This is a very large suite of products, and Oracle provides ample documentation. So there is little point drilling down into details here, and we will focus on architectural issues.

Visualizations

Data visualization is hot right now and Oracle has responded by adding numerous new visualization modes in its BI toolkit. These include 'recommended visualization' for specific data and intent, performance tiles to give prominence to a metric or set of metrics, waterfall diagrams often used in pricing analysis, map views and many, many others. It easily competes with the new generation of visualization tools, but is of course just a single feature of a much broader capability.

Oracle Exalytics

Oracle Exalytics is a combination of hardware and software that facilitates very high levels of analytics performance. This is an in-memory machine with built-in analytics capability that addresses BI, predictive analytics, data mining and Enterprise Performance Management needs. The Oracle BI suite takes advantage of the Exalytics speed, and it is claimed at least that use of Oracle Exalytics can reduce cost of ownership – but we have to take Oracle's word for that.

Oracle Essbase

Oracle Essbase provides an OLAP solution and can be implemented on the Exalytics platform for performance improvements.

Oracle Exadata

The Oracle Exadata Database Machine is a large scale, scalable data warehousing and OLTP database platform. Exalytics complements Exadata by providing analytics capability.

These various database platforms clearly need to be configured for individual organization requirements. But the diversity of function provides a rich set of configurations, and Oracle is one of only two or three supplier who can match this flexibility.

Other components of the BI suite include BI Publisher. This provides a web based platform for authoring, managing and delivering interactive reports, dashboards and other forms of formatted information.

Oracle BI Suite Standard is aimed at small to medium size organizations, and Scorecards and Strategy Management support execution of strategy by cascading goals down to individual workers and departments.

All-in-all this is a comprehensive BI environment with many, many configuration options. The improved visualization capability keeps Oracle competitive with newer technologies, but within the context of a broader architecture.

Oracle Endeca Information Discovery

Oracle Endeca Information Discovery (OEID) addresses the information search problem. It can draw upon email, social media, content management systems, transactional data, documents, and external websites – virtually all information sources.

Prior to Oracle's acquisition of Endeca, we rated it most highly of all the enterprise search technologies we looked at in 2009 (the report is still available on our web site). It is exceptional technology, and we commented at the time that this was a good acquisition for Oracle – and so it seems.

OEID indexes all the data it is pointed at regardless of structure. This means diverse information sources can be tied together by using these indexes – the name of a customer might be found in transactional data, in memos, emails, social data and so on. OEID facilitates a unified picture.

> This quite unique technology positions itself in the top three search technologies in our opinion.

At the heart of OEID is the **Oracle Endeca Server**. This is a hybrid search/analytical database, and was the distinguishing factor in our 2009 analysis, enabling very high levels of performance. The lengthy waits that can be a feature of enterprise search are much less likely. Data is stored in column format, and extensive caching means these highly compressed indexes can be scanned with great speed.

The user interface is provided by **Studio**. This is a drag and drop interactive environment that enables high productivity and supports complex queries.

The **Integration Suite** provides the connectors to a bewilderingly large number of data sources – both structured and unstructured.
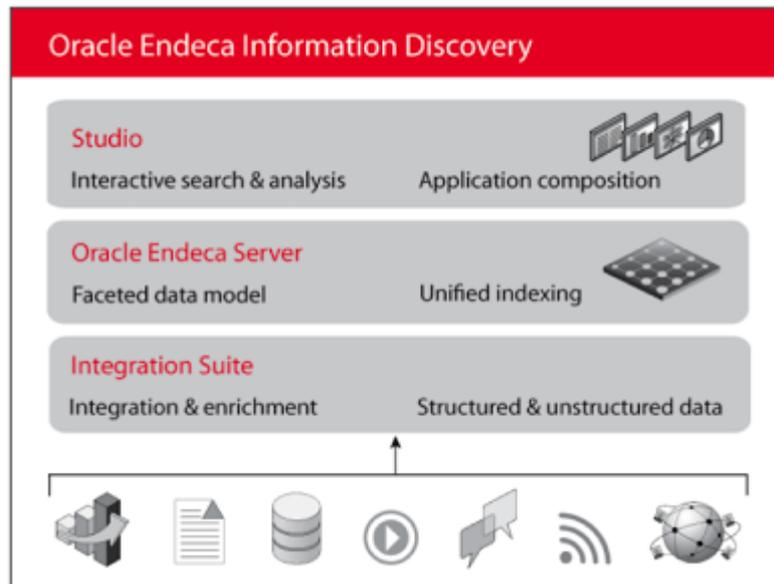
The performance of OEID is further enhanced by multicore parallel query, extensive caching, per-attribute column storage, and multiple compression techniques.

This quite unique technology positions itself in the top three search technologies in our opinion, and the innovations it incorporates make it an ideal solution for many enterprise search needs.

Further details on the OEID platform can be found in this technical overview.