Data Warehouse Automation Tools

Product Categories and Positioning

By Wayne Eckerson

September 2015

About the Authors

Wayne Eckerson has been a thought leader in the business intelligence and analytics field since the early 1990s. He is a sought-after consultant, noted speaker and expert educator who thinks critically, writes clearly and presents persuasively about complex topics. Eckerson has conducted many groundbreaking research studies, chaired numerous conferences, and written two widely read books on performance dashboards and analytics.

Eckerson is founder and principal consultant of Eckerson Group, a research and consulting firm that helps business and analytics leaders use data and technology to drive better insights and actions. His firm helps companies develop strategies and roadmaps that maximize their investment in data and analytics.

About Eckerson Group

Eckerson Group is a research and consulting firm that helps business and analytics leaders use data and technology to drive better insights and actions. The firm helps companies develop strategies and roadmaps that maximize their investment in data and analytics. Its consultants and researchers each have more than 20 years of experience in the field and are uniquely qualified to help business and technical leaders succeed with business intelligence and analytics, big data management, data governance, performance management, and the internet of things.

About This Research

This market overview is part of a larger research study on data warehouse automation tools. This base report, which classifies products in an industry framework and matches them to business requirements, is accompanied by profiles of the four leading data warehouse automation products: Attunity Compose, Magnitude Kalido, TimeXtender, and WhereScape. All reports can be downloaded from www.eckerson.com.

Abstract

Data warehouse automation (DWA) tools eliminate the manual effort required to design, deploy and operate a data warehouse. By providing an integrated development environment, DWA tools enable developers and business users to collaborate around designs and iteratively create data warehouses and data marts. This turns data warehouse development from a laborious, time consuming exercise into an agile one.

There are four DWA tools in the market today. One primary distinction between the tools is the approach they take to designing data warehouses. Some follow a model-driven approach and others a data-driven approach. Another distinction is the data warehouse platforms they support. Some focus on just one platform, while others take a more ecumenical approach. Beyond these two attributes, there are many factors that customers should consider before purchasing a DWA product.

Overview

A data warehouse is the centerpiece of a data-driven organization. A data warehouse contains clean, integrated, trustworthy data that business professionals use to make decisions. Unfortunately, creating a data warehouse is not easy or cheap. It requires piecing together disparate sets of information from different systems managed by different groups across an organization.

Fortunately, modern organizations have better tools to integrate data fiefdoms and create a holistic data environment for decision making. Data Warehouse Automation (DWA) tools, which began emerging more than a decade ago, eliminate much of the manual effort and coding required to design, build, manage, operate, and document data warehouses and data marts.

DWA tools eliminate much of the manual effort and coding required to design, build, manage, operate, and document data warehouses and data marts.

Every industry has used automation to increase productivity, reduce manual effort, improve quality and consistency, and speed delivery. Henry Ford introduced the assembly to produce automobiles, and today Uber and countless other startups use the Internet and digital processing to reduce friction in commercial transactions. Thus, the time has come to introduce automation to data warehousing.

Challenges

Old habits die hard. Too many companies rely on traditional labor-intensive, hand-coded approaches to building data warehouses. These companies have invested millions of dollars in development tools (e.g. extract, transform, and load or ETL tools) and the people who write code with these tools. Also, developers often resist automation because they fear it will put them out of work.

In addition, many consultancies make money through hand coding and don't want to undermine lucrative business models with automation tools. And it's interesting that no large software vendor offers data warehouse automation tools, although some embed similar capabilities inside their platforms. The data warehouse vendors profiled in this report are small; the largest generates only $25 million in revenue. This lack of industry endorsement is another factor that has checked the growth of the DWA tool market

Many consultancies make money through hand coding and don't want to undermine their business models with automation tools.

Hefty Price Tag. DWA tools are not inexpensive—a single development platform and license costs $50,000 or more and enterprise deployments can exceed a million dollars. But the beauty of automation tools is that they provide huge return on investment since they eliminate an inordinate amount of hand coding which consumes lots of time and money. Nonetheless, the chief executive officer of one data warehouse automation vendor commented, "Companies risk a lot when they purchase a data warehouse automation tool, so it's imperative that they succeed with their initial project."

Old School. Finally, DWA tools are somewhat old school. While many business intelligence (BI) and data warehousing managers are testing Hadoop, Spark and other big data platforms to build or augment their data warehouses, the current crop of DWA vendors is largely focused on the traditional world of relational data warehousing. DWA vendors have an opportunity to bring automation into the brave new world of big data, but given their limited size and resources, most seem content for now to pursue the traditional data warehouse market, which still has an enormous need for automation.

Benefits

Despite these challenges, the time is right for DWA to become a household acronym—at least within information technology (IT) circles and even in the offices of chief financial and operations officers. Data warehouses are under attack from business managers and big data proponents as being too slow, costly, and inflexible. DWA tools reverse the equation and make traditional data warehouses agile again.

Design, Build and Operate. DWA tools address the entire data warehousing lifecycle from analysis and design to build, test, and operate. They support an agile, iterative and collaborative approach to gathering requirements and designing business solutions. The tools convert requirements and designs into metadata, which are used to generate database schema and transformations and manage workflows.

DWA tools support an agile, iterative and collaborative approach to gathering requirements and designing business solutions.

Specifically, DWA tools automatically generate data warehouse schema (e.g., star, snowflake, third normal form, or data vault), staging tables, data marts, indexes, OLAP cubes, and, in some cases, business metadata for business intelligence (BI) tools. They also schedule, manage and monitor data warehousing processes, including data loads, validations, indexing, and transformations, and they track errors, monitor performance, and validate designs. In most cases, DWA tools minimize or eliminate the need for ETL tools, simplifying and speeding data warehouse development.

Standardized Development. DWA tools provide an integrated development environment to build, design, and operate a data warehouse. As such, the tools impose a standard method and style of development that improves quality and consistency. Some DWA tools automatically convert field names to a standard naming convention. And using version control, development teams can easily implement build, test, and production environments.

Change Management. With DWA tools, administrators can inspect the impact of any change on a data warehouse environment before making modifications to the design. They then push a button that regenerates the code needed to redeploy the data warehouse and related structures. As a result, DWA tools make it easy and quick to modify data warehouses to meet new requests, turning sclerotic environments into agile ones.

DWA tools make it easy and quick to modify data warehouses to meet new requests, turning sclerotic environments into agile ones.

Roll Back. Meanwhile, the DWA tool logs all changes and tracks versions so that administrators, if they desire, can roll back the data warehouse environment to a previous state with the push of a button. Moreover, DWA tools also automatically generate the documentation that describes the structures and processes in the DW environment. The best DWA tools produce separate documentation for business users and technical administrators.

Product Categories

Creating a short list. If your company is eager to reap the benefits of DWA tools, the next step is to select a product that aligns with your requirements. Although all DWA tools automate data warehouse development and promise to create agile data warehouses, they are not all alike. The good news is that there are only four DWA vendors on the market today, so exploring their similarities and differences is not an arduous ordeal.

This report is designed to help your company sort through the range of available DWA options and create a short list of products to evaluate. It defines the major segments in the DWA market and identifies providers in each segment. It also suggests questions you should ask each vendor to ascertain whether their cloud BI service fits your business requirements.

DWA Segments. There are many attributes to evaluate when selecting DWA tools. (See appendix "Questions to Ask DWA Vendors".) Perhaps the most important is design approach, which forms the basis of our visual framework for evaluating DWA products. (See figure 1 below)

Figure 1. DWA Design Approach
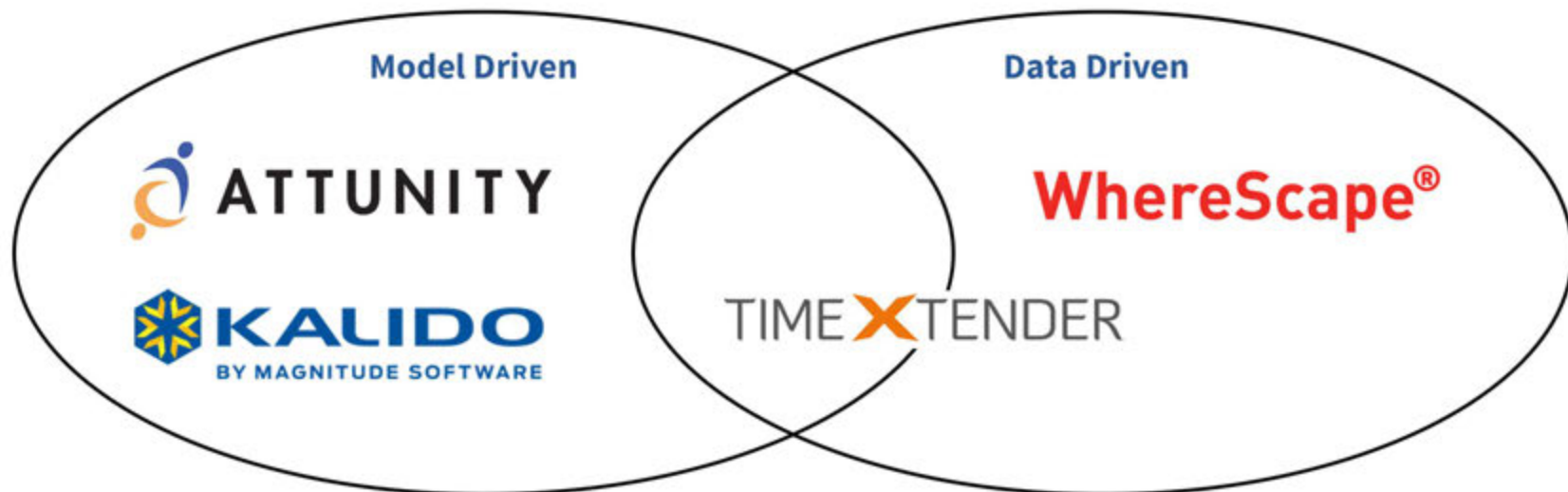
DW Design Approach

The biggest distinction among DWA tools is their approach to designing data warehouses. The differences here represent a philosophical choice about the best way to gather requirements and design data structures to support business needs.

Model-Driven Approach

Proponents of a model-driven approach—namely Magnitude and Attunity—believe that the best way to capture business requirements is to create a conceptual or logical model that visually represents how target users see the business and want to analyze it. The visual model enables business users and developers to collaborate around a vision of the desired solution in terms that business users can understand. (See figure 2.)

**Figure 2.
A Conceptual Model
from Magnitude Kalido**
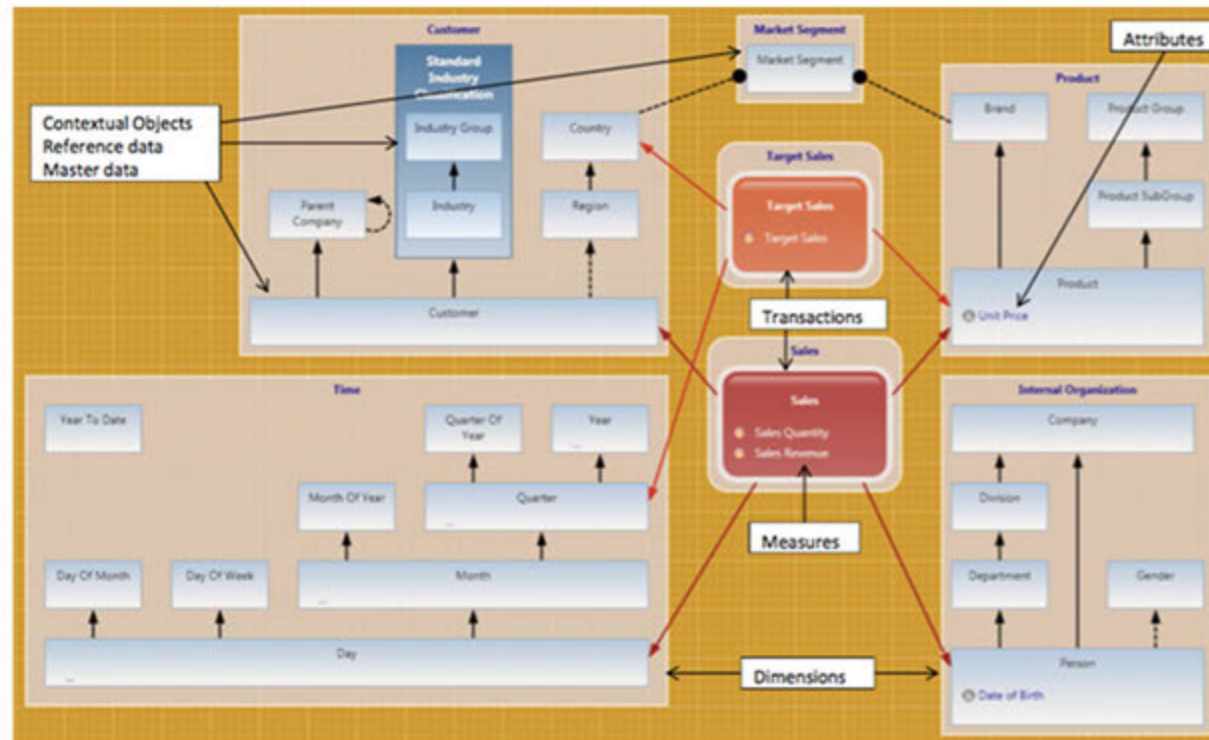
This conceptual model for a sales report shows key facts (e.g., sales transactions) and dimensions (e.g., customer, product, organization, and time) with their attributes and hierarchies. The arrows show relationships between entities.

The model contains all the required entities (e.g., facts and dimensions), their attributes and hierarchies and rules governing their relationships (e.g., linkages between entities). Once business users validate the model, developers then find appropriate data, import it into the tool, and map it to the conceptual model.

Enterprise Views. The benefit of a model-driven approach is that it clearly defines and documents business requirements upfront. Business managers or technical administrators can review the model and determine whether it conforms to business standards and aligns with other business models and solutions in the organization. Consequently, the model-driven approach helps drive toward an enterprise view of the business and avoids the creation of data silos and conflicting reports.

Grab Only What You Need. This approach also focuses the data discovery process. Rather than exploring all data for its potential to meet business requirements, a model-driven approach defines exactly what facts and dimensions a solution requires, narrowing the data exploration and extraction process. In other words, in a model-driven approach, developers are more likely to pull only the data they need, not the data they might need.

However, with a model-driven approach, it can be difficult to keep the physical data warehouse in sync with a conceptual model. For example, Magnitude Kalido separates its modeling environment from its data warehousing module, which administrators use to create load jobs, manage processes, set execution parameters, attach external code, etc. In recent releases, Kalido has moved more DW configuration functions to its modeling environment to improve synchronization.

Should We Buy a Model-Driven DWA Tool?

Yes, if you want to:
- **Use a conceptual or logical model** to generate physical data structures.
- **Define and capture rules** in a conceptual or logical model.
- **Create an enterprise data model** that spans multiple data marts.
- **Collaborate with business users** using a visual model of the BI solution.

Data-Driven Approach

Proponents of a data-driven approach, such as WhereScape, believe that it makes no sense to define requirements in the abstract without looking at actual data first. They argue that modeling creates unjustified expectations among business users who will be disappointed if data doesn't exist to support their requirements.

Data-driven proponents believe that developers and business users should collaborate around actual data, not models. With data-driven DWA tools, developers quickly generate data to show business users and get their feedback. By working iteratively, they can rapidly prototype and build a solution that meets user requirements.

Data-driven DWA tools, such as WhereScape RED, steps developers through the process of creating facts, dimensions, hierarchies, surrogate keys, and indexes, among other things. In essence, a data-driven DWA tool develops a data warehouse at the physical level and generates all the SQL, stored procedures, and scripts required to build physical tables in the target database. (See figure 3.)

In essence, a data-driven DWA tool develops a data warehouse at the physical level

Figure 3. WhereScape RED's Data-Driven Design Environment

Developers build a data warehouse by converting loaded data (right column) into various data warehousing structures (e.g., dimensions, staging tables, fact tables, aggregates, cubes, and indexes in left-hand column) using templates and wizards. Actual data appears in center panel
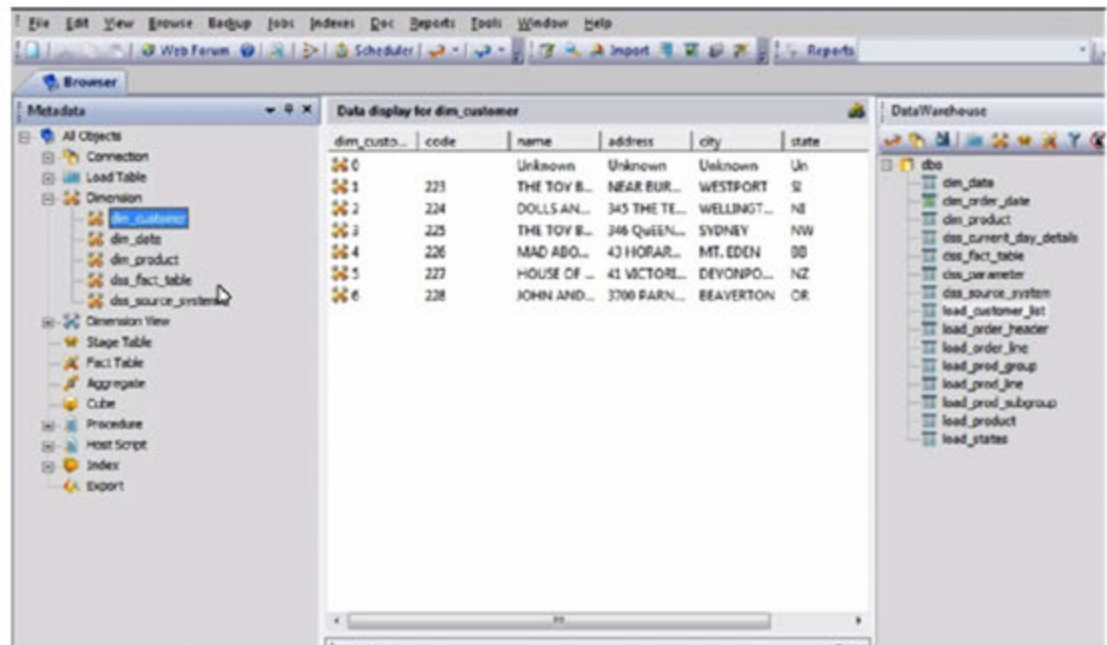
Developer Tools. Without a conceptual or logical model, the data-driven approach doesn't explicitly display rules or data relationships in a way that is easily accessible or understandable to business users. As a result, data-driven tools are primarily developer tools. The way developers and business users collaborate is by reviewing DWA data outputs, usually displayed in Excel or a BI tool.

In addition, data-driven tools are best used against well-understood sources of data, like data warehouses. More than half of WhereScape's customers, for example, use the tool to generate data marts. To support more complex data sources, WhereScape recently shipped a data profiling and model-testing tool called WhereScape 3D that helps customers better understand the nature and condition of source data prior to loading it. The tool also helps developers validate third party models using actual data prior to importing them into WhereScape RED.

Should We Buy a Data-Driven DWA Tool?

Yes, if you want to:

- **Design a data warehouse** using actual data rather than a data model.
- **Collaborate with business users** by iteratively prototyping BI solutions with actual data.
- **Create realistic expectations** for a BI solution based on data.

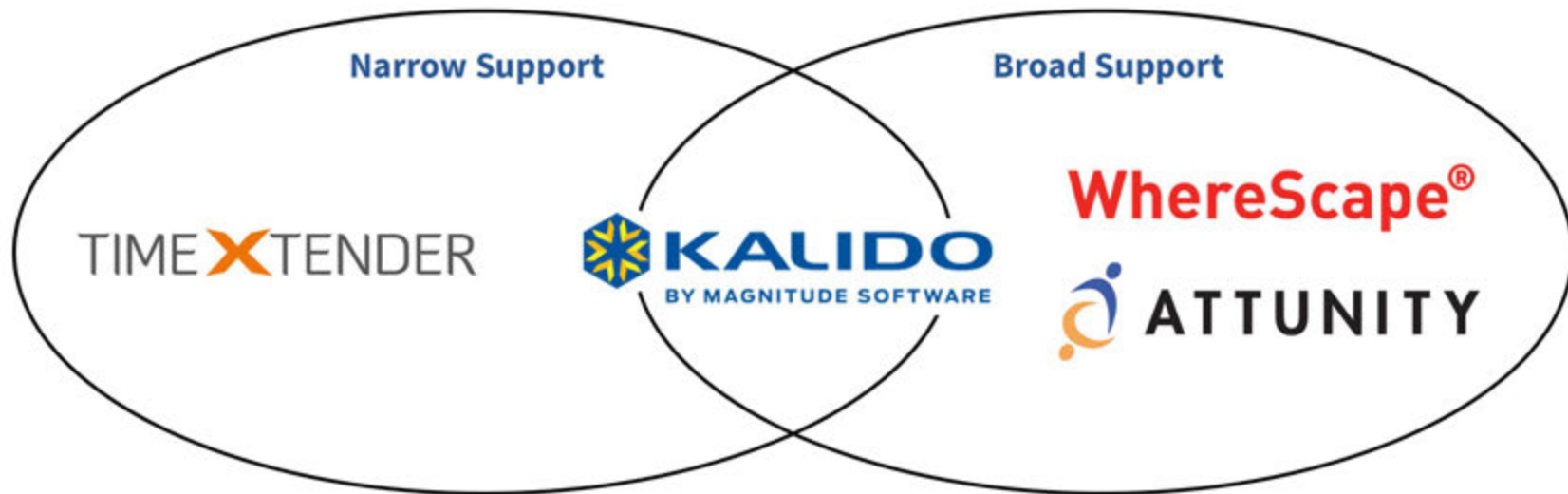## Data Warehouse Platform Support

Another significant distinction among DWA tools is the target DW platforms supported. (See figure 4.) on pg10.

It takes a lot of effort for a DWA vendor to port its code to a new database. Although DWA tools capture metadata in a generic format, they must then generate database-specific code from it. This is harder than it sounds. Each database has its own variation of SQL and supports its own language for stored procedures and user-defined functions, and applies indexes in unique ways. Moreover, database vendors continually issue new versions and releases and DWA vendors must test their code against each to ensure it generates the proper objects. So each database port represents a significant investment by a DWA vendor.

Eckerson Group

Figure 4: Platform Support

Single Platform Support

Databases. One vendor that has opted out of the multi-platform arms race is TimeXtender, which only generates Microsoft SQL Server data warehouses. But even that task is Herculean since Microsoft ships frequent upgrades. TimeXtender performs exhaustive testing of each new SQL Server release and often finds bugs that have evaded Microsoft.

TimeXtender believes it can provide broader and deeper automation by focusing on a single platform rather than many. For instance, TimeXtender automatically defines dimension types rather than present users with a dimension selection wizard. And it automatically optimizes load management jobs using built-in intelligence.

Given the size of the Microsoft SQL Server market, especially in Europe where TimeXtender is based, a single platform strategy makes sense. Besides Microsoft's partner network, TimeXtender has made inroads with Qlik's partner network, which builds many data warehouses on Microsft SQL Server.

In contrast, other vendors prefer to take a broader approach by supporting the major data warehouse database platforms. For instance, Attunity Compose supports Teradata, Oracle, SQL Server, Oracle Exadata, MySQL, and IBM Netezza, whereas WhereScape RED supports these plus Greenplum,

IBM DB2, and Microsoft Analytic Platform System. Wherescape is also testing Amazon Redshift and Microsoft Azure databases and will soon be the first DWA vendor to provide a connector to Hive and Hadoop.

Data Access. On the other side of the data warehouse are the source systems that DWA tools must access to extract and load data. Most DWA tools use file loaders, ODBC or native data access and load utilities of the target database. In other words, they ingest data in batch. But as data sources proliferate, DWA vendors may need to take a more tailored approach to connecting to source data.

Besides selling a DWA tool, Attunity also sells a raft of connectivity software, including data replication, change data capture, and cloud data delivery tools. In fact, data connectivity is Attunity's heritage. Not surprisingly, Attunity has embedded its replication software into its DWA tool (Attunity Compose) providing high-performance connectivity and change data capture from relational, legacy and NoSQL sources. This gives Attunity's DWA software a unique advantage in the marketplace.

TimeXtender also offers a unique twist to data access. Besides offering ODBC access to relational databases (as well as OLE DB and ADO), it provides adapters to various business applications, including Microsoft Dynamics, Infor, Salesforce.com, and SAP R/3. In general, the adapters grab data through the application interface, not the database. WhereScape RED also accesses Salesforce and SAP Business Warehouse.

BI Tools. Platform support is also defined by the types of front-end tools that a DWA tool supports. In general, most DWA tools focus on generating database structures, but some generate metadata files (i.e., semantic layers) for BI tools. Most of the BI tools can generate Microsoft Analysis Services cubes, but Magnitude Kalido goes a step further by shipping a component that generates BI metadata for IBM Cognos, SAP BusinessObjects, and QlikView. WhereScape RED generates data schema that Tableau can consume.

Conclusion

The time is ripe for data warehousing professionals to implement DWA tools. Automation is critical to the success of data warehousing projects. DWA tools eliminate the manual effort required to design, deploy and operate a data warehouse and enable developers and business users to iteratively create and quickly modify data warehouses and data marts. With DWA tools, developers can deliver agile data warehouses that put information into the hands of data-hungry business users.

The four DWA tools in the market today provide end-to-end automation capabilities. This report charts the key attributes that organizations should evaluate prior to selecting a DWA tool. Chief among them are design approach and platform support. But there are many more attributes that organizations should evaluate when making a buying decision. DWA tools are not cheap, but they can save organizations a bundle of money in development costs, and more importantly, salvage the reputations of IT shops bogged down with manual coding.

APPENDIX: Evaluation Criteria and Questions

This report has reviewed two key attributes of DWA tools: design approach and platform support. But there are numerous other attributes that you should evaluate prior to selecting a DWA tool to ensure it's a good fit for your organization. Below is a more comprehensive list of attributes to evaluate, along with questions to ask vendors.

1. Design Approach. As discussed above, identify whether your vendor supports a data-driven or model-driven approach. Questions to ask:
a. Can you build a conceptual or logical model in your tool?
b. Does the tool require you to load data first before designing models?

2. Platform Support. As discussed above, find out what databases, BI tools and data access tools the DWA tool supports. Questions to ask:
a. What database versions and releases does your tool support?
b. Do you offer more than ODBC connectivity to source data?
c. How do you handle file imports?
d. Do you generate BI tool metadata?

3. Metadata. All DWA tools convert designs into metadata, which drive the creation of data warehouse structures. Questions to ask:
a. Where does the tool store its metadata?
b. Can developers inspect and alter the metadata?
c. Can developers insert custom code into the metadata?
d. Can the tool track code changes and incorporate them into a version?

4. Data Access. Most DWA tools automatically generate the programs, scripts and mappings traditionally created by ETL tools. Questions to ask:
a. Does your DWA replace or augment your existing ETL tool?
b. How does your tool ingest data?
c. Does your tool optimize load performance?
d. Can your tool support incremental loads and change data capture?
e. Can your tool connect to applications as well as databases?
f. Can your tool profile source databases?
g. Does your tool support data matching and cleansing rules?

5. Architecture. Most DWA tools are Windows desktop tools and use a database and/or Web server for an application server. Questions to ask:
a. What are your tool's systems requirements?
b. What does it use for a server?
c. What components does your tool support and how do they interact?
d. What other tools are required, if any, to use your product?
e. Do you run on private or public clouds?

6. Models. DWA tools either create or import models, or both, and in some cases reverse engineer models from existing databases. Questions to ask:
a. What types of models does your tool generate? What's the preferred or default model?
b. Does your tool import Erwin or packaged models?
c. Can your tool reengineer a model from an existing database?
d. Can the model control the physical implementation of the data?
e. Does the tool apply a default or customer-driven naming convention to all fields?

7. Version Control. Version control is critical to data warehouse operations. Questions to ask:
a. Can we roll back our data warehouse loads, designs, transformations, views and hierarchies to a previous state?
b. Can we create separate instances for design, test, and production and deploy each as needed?

8. Security. Does the tool secure data so only authorized developers and users can see it? Questions to ask:
a. Does the tool support row- and column-level security?
b. Does the tool integrate with our company's security platform?

9. Team Development. Some DWA tools run on desktops, making team-based development challenging. Questions to ask:

a. Can multiple developers work on a model or project at the same time?

b. Does the tool prevent two developers from editing the same objects or does it publish potential conflicts so developers can synchronize changes manually?

c. Can developers zoom in on one part of a model or project to simplify development?

10. Change Management. The core value of DWA tools is the ease with which developers can make changes to a design. Questions to ask:

a. Does the tool check and validate changes against predefined and default rules before deploying them?

b. Does the tool display impact analysis and data lineage programs?

c. Does the tool alert developers to potential errors and conflicts when the design changes?

11. Documentation. Most DWA tools automatically generate documentation. Questions to ask:

a. Does the tool generate business and technical documentation?

b. How easy is it to share the documentation?

12. Operations. A data warehouse has many moving parts that must be coordinated. Questions to ask:

a. Does the tool come with a scheduler or work with third party schedulers?

b. Does the tool issue warnings and error messages when jobs fail to complete?

c. Can administrators configure jobs and process to optimize performance?

13. Pricing. DWA vendors offer many types of pricing. Questions to ask:

a. Do you charge by named user, target database, number of sources, volume of data?

b. What are your annual maintenance charges?

14. Support and Services. New customers require training and support to deliver a successful initial project. Questions to ask:

a. What training classes do you provide and at what cost?

b. What level of support do you provide and at what price?

c. How many partners do you have? And how skilled are they in your tool?

d. Will you design and manage our entire DW environment?

15. Vendor. Most DWA vendors are small, but some have been acquired by larger companies. Questions to ask:

a. How viable is your company? (Revenues, profitability, cash on hand)

b. How many active customers and references do you have?

Need help with your business analytics or data management and governance strategy?

Want to learn about the latest business analytics and big data tools and trends?

Check out **Eckerson Group** research and consulting services