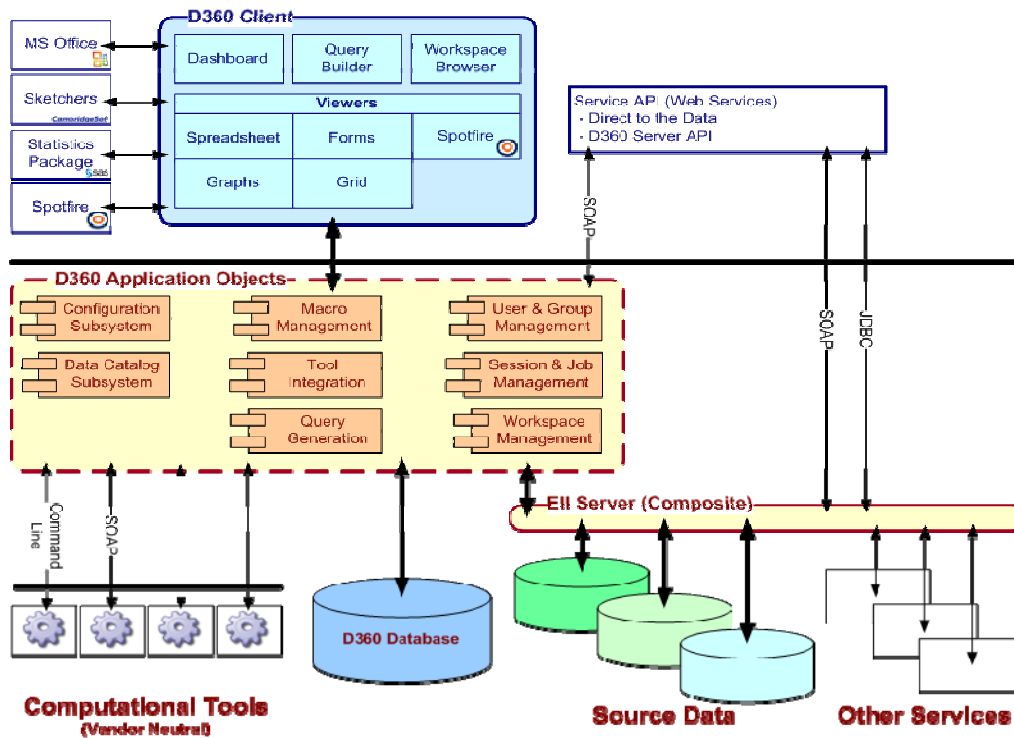


# Benchware Discovery 360° for the Technical User

## Architecture

D360 is a multi-tiered application built from a dynamic collection of interchangeable components and services, as shown in the diagram below.



The system has several major architecture components:

- **D360 Client:** A rich tool through which users interact with the functionality of the system. This is a Java application, which can be deployed as standard desktop-installed software, or through a web-based deployment mechanism (Java WebStart).
- **D360 Application Server:** The server is a collection of services that store configuration, generate queries, manage sessions and user jobs, execute data transformation logic, manage user workspaces, and interact with computational tools and other services. These are typically deployed within a Java/J2EE environment on a variety of supported platforms (Windows, Linux, Solaris).
- **EII Server:** A key component of the solution is an Enterprise Information Integration server. D360 uses an embedded solution from Composite Software. The Composite engine is a proven system, used in large organizations in a variety of industries to access diverse types of data. The EII server supports access to “federated” data, from a variety of data sources. These data sources may be underlying databases, external services and tools, files, etc. Combined with the scientific knowledge provided through D360, the solution is optimized for dealing with the typical source database systems used throughout discovery organizations.
- **D360 Database:** An Oracle database is used to store information about jobs being run, store virtual data sets while users are interacting with them, and provide storage for the workspace subsystem.
- **Tools, Source Systems and Services, and Computation Engines:** As described below, D360 is often integrated with a variety of end-user tools, source systems, customer and third-party services, and computation engines. D360 includes “out of the box” integration with the most common tools and services in each of these areas, and simple mechanisms (through configuration and pluggable components) to add integration with new services and tools.

The architecture of D360 has been designed to fulfill several high-level design goals:

- **Proven performance and scalability:** In order to support the needs of users throughout large discovery organizations, a solution such as D360 must perform well, and scale from small numbers of users up to large deployments with thousands of users. D360 has been optimized to ensure response time and query performance are as fast as possible, allowing users to focus on analyzing their data and making decisions rather than waiting for query results. In addition, real-world deployments of D360 have had utilization of greater than 10,000 queries per month, and scalability testing has been performed to significantly higher levels.
- **Modular collection of components and services:** Although D360 is a comprehensive solution that can be used “out of the box”, an important consideration was to ensure that customers could extend, replace, remove, or add modules as necessary. This includes components within the user interfaces as well as services and data sources on the back end.
- **Configuration-driven:** All aspects of the application make use of the D360 configuration system, allowing their behavior to be determined dynamically, and to be changed without the need to develop or modify any code. In most cases, configuration can be applied globally, on an individual user basis, or to groups of users.
- **Open integration options:** Users of D360 typically have a variety of tools they interact with, ranging from Microsoft Office to Spotfire to highly specialized applications. In addition, customer organizations typically have a variety of services, data sources, and computational tools they need to tie together into a comprehensive solution. D360 has been designed to provide open integration options for hooking these tools and services together as necessary.
- **Easy to setup, use and maintain:** Finally, Tripos has focused on ensuring that the power added through D360 to satisfy the design goals above is achieved while also maintaining ease of use and maintenance. Though a number of extension and configuration points are available, the system is usable after deployment simply by configuring at least one data source. Tripos has also worked to ensure ongoing maintenance and support activities are minimized, to ensure the overall cost of ownership is as low as possible.

## Configuration and Customization

### Data Sources

The most obvious and immediate need for configuration arises with access to data sources, as access to data is required to make use of functionality in the system. Data sources are connected to the EII server through a graphical configuration mechanism, which mostly involves simply providing connection information and then using “auto-discovery” capabilities to find and interpret the available data. Query generation and data display within the D360 application are controlled by the Data Catalog. The Data Catalog controls how fields are presented to users, manages important relationships between data, and provides metadata which will be used in visualization or analysis. Tripos provides administrative tools for building the data catalog.

### Data Transformation and Business Rules

Once D360 has been connected to data sources, customers frequently chose to implement business rules for transforming data as it is retrieved. This often involves operations such as pivoting result data into distinct columns, aggregating values at various levels, or automatically converting data from storage units to an alternate display format. Most of these operations can be controlled by an end user through the query user interfaces, but defaults or corporate business rules can be implemented through configuration. When more advanced transformation or specialized aggregation rules are desired, new services can be deployed into the system as pluggable components available to the user.

## **Service and Tool Integration**

A variety of mechanisms are commonly utilized to integrate external services and tools. First, customers frequently have compute engines and predictive models (in-house or third party) that they desire to connect to datasets. These tools may generate additional data to be included in datasets, or calculate data based on other information or structures in those datasets. D360 includes a framework for integration of these tools, which can invoke engines locally on the server machine (as executables) or defer to remote services. Remote services are typically deployed as SOAP web services, which are then hooked up to D360 through simple configuration. This configuration involves providing a definition for the service API, typically in a Web Services Definition Language (WSDL) file, and providing metadata about how dataset information should be passed to the service and how results should be displayed in the UI. Current customers use this to connect not only to standalone compute tools, but also to workflow tools and integrated models, including those utilizing Pipeline Pilot.

In addition to services invoked from the system, customers frequently desire to connect desktop tools within the D360 client. D360 contains standard integration with the most common desktop tools, including Microsoft Excel, PowerPoint, Spotfire, and various statistical packages. This tool integration can appear in a variety of ways. In the most common scenario, data is exported to the external tool from D360. When the external tool accepts standard data formats, this export can simply be defined through new configuration entries, including invoking the tool through command line or direct APIs. For more complex data needs, or for deeper integration of the tool within the client application, components can be implemented to D360 APIs and then “plugged in” to the client. In the most advanced cases, complex components and tools are tied deeply into menus, datasets, and functionality throughout the system, building a powerful integrated application utilizing D360 as a platform.

## **Visualization**

A number of visualizations (spreadsheets, graphs and charts, grids, forms, similarity maps, etc) are available in D360 datasets out of the box. However, customers occasionally desire to extend these visualizations, and add new ways to visualize data.

Many types of data presentation can be created through configuration of existing components. In fact, the available viewers are determined dynamically based on configuration data. In some cases, a customer may choose to add a new viewer by going beyond configuration, and creating new visualization code. This code is implemented to standard product APIs, and “plugged in” to the D360 client dynamically. In other cases, a customer may desire to integrate an existing tool as a visualization mechanism. This can be done through configuration and connections of the external tool to standard interfaces. These interfaces provide several levels of integration – ranging from simply pushing data out, to deep, two-way communication.

D360 includes an example of this with the standard integration to Spotfire. A user can choose to invoke Spotfire as a D360 viewer, which launches the external tool and provides access to data from the user’s dataset automatically. Events are then propagated in both directions between D360 and Spotfire, allowing a user to make selections and changes in either application and see the results in the other. Finally, when the user saves the D360 data set, the entire Spotfire session, including all setup, visualization, selections, and data, are persisted, and restored in the original format if the dataset is opened at a later point.

## **External Access and Sharing**

Finally, in many D360 deployments it is desirable to access data in D360 (including visualizations, queries, etc) in external tools such as SharePoint. These systems can be connected together by hooking into standard APIs within the D360 application server.

## **Other Areas**

The areas listed above provide several examples of the typical configuration and customization that customers may choose. Many other aspects of D360 are configurable and extensible, allowing customers to utilize D360 as a platform on which broader solutions are built, and to integrate D360 into other applications used throughout Discovery workflows.

