Virtual Prototyping: Bridging the Business/IT Gap

Jeff Reagan

Abstract
Many companies that embark on business intelligence (BI) initiatives experience serious setbacks to project budgets and timelines. Recent research reveals that “efforts to eliminate time and cost overruns in business intelligence and data warehousing projects are mostly unsuccessful.” In fact, these issues are so common that 62 percent of those surveyed factor delays and cost overruns into the budgets for data warehouse projects (Havenstein, 2007).

One of the many factors that contribute to the high costs and missed deadlines of BI projects is a lack of alignment between the business and information technology (IT). Prototyping is one way to bridge the gap between business and IT and regain control over BI projects. Prototypes provide a means to validate business requirements, design decisions, and data quality issues before the final product or service is delivered. Prototypes can also be controversial: their associated time and cost are perceived to be prohibitive, and the risks may outweigh the prototype’s benefits.

In this article, we explore the use of a virtual prototyping warehouse (VPW), created using data federation technology, to achieve the same goals as a physical prototype but at a fraction of the cost and time.

Introduction
Business intelligence is not driven by IT alone. Many business initiatives require access to data across many sources to drive profits and reduce costs. Among the common drivers are:

- Management: planning, budgeting, and forecasting
- Marketing: marketing analysis, customer segmentation, direct marketing
• Operations: risk reduction, quality improvement, asset reduction

Each of these business drivers requires specific information reported in very specific ways to gain the promised returns. Unfortunately, capturing requirements and consolidating data is not an easy task. Despite the challenges, these strategic projects are a must if an organization is to gain a competitive advantage. Although the challenges come in many shapes and sizes, a few stand out because they are both common and difficult to overcome:

• Communication gaps between business and IT (referred to as the business/IT gap)
• Changing requirements during development due to long project duration
• Difficulty integrating technology from multiple vendors
• Complications from dealing with multiple, disparate data sources
• Unsatisfactory data quality

Of these obstacles, the business/IT gap has the greatest negative impact, both because the problem runs so deep and because it affects BI projects at so many points within the development cycle. The virtual prototyping warehouse (VPW) development environment provides a powerful way to overcome this issue by enabling:

• More accurate determination of functional requirements through input from business early in the project
• Access by business users to integrated data throughout development
• Early identification of data quality and standardization issues
• A faster return on investment

Formally defined, a physical prototype is an original or model upon which the final project is based. In a data-sharing project, a prototype can be used to validate design decisions, test the alignment of design to business needs, and ensure that data quality is satisfactory. However, some people view prototypes as counterproductive because of the additional cost and impact on project deadlines from a prototype’s development. Many people simply do not want to do the work twice, fearing that the prototype will fail and the effort and resources will be wasted.

A virtual prototype’s goals and use are similar to those of a physical prototype. However, the two prototyping methods vary greatly in their implementation. Whereas physical prototypes require new physical data structures, data movement, and data access methods, virtual prototypes do not move data; they use existing data structures, and they access data via federation technologies. These differences help to overcome many—if not all—of the challenges physical prototypes typically present.

The Great Divide: Business versus IT

Business users communicate daily with IT about needs, business initiatives, and IT solutions. Unfortunately, what business says and IT hears are often different. This lack of understanding causes a gap that is often too wide to overcome, putting projects (particularly BI initiatives) at risk.

Some communication issues stem from misconceptions about what motivates the other party. Business users often perceive IT staff as inflexible and wanting technology solely for technology’s sake. They believe IT doesn’t understand the business and doesn’t interact well with others. On the other hand, IT perceives that business people refuse to learn anything about IT but act like they are knowledgeable and resist innovation. IT professionals believe business people are money-driven and value image over substance.

This lack of understanding and frequent miscommunication becomes detrimental during the project’s requirements-gathering stage. Because requirements are often lost in translation, IT can go down the wrong
development path. If business users are not actively involved in the development process, those errors can go undetected until the project is completed. Once they are uncovered, IT may be forced to start over to correct them.

Bridging the Business/IT Gap: Virtual Prototyping
A virtual prototype can be a highly successful way to bridge the business/IT gap, helping business and IT to successfully complete BI initiatives.

Business users can be involved earlier in the development of the solution, ensuring that what IT builds is what the business needs. It also allows all stakeholders to identify and avoid costly mistakes before they are made.

A VPW can overcome the common fears about prototypes. It uses federated technology and avoids the cost and effort of physical development. Hardware and software costs are significantly lower, and resource usage is dramatically lower because business requirements are more readily captured and vetted by users.

A VPW differs from a physical prototype in several ways:
- The VPW does not permanently store data; it leaves data at the source and accesses it only as needed to fulfill a user request
- It is not intended for full-scale production use and thus is not concerned with performance, resulting in lower hardware costs
- It can be created in a fraction of the time it takes to create a physical prototype, making it cost-effective
- The main objective of a VPW is business validation; it helps business users visualize the end result in a tangible way so they can actively participate in testing throughout the development cycle

Benefits
The VPW helps mitigate risk and bridge the business/IT gap. Business users can be involved earlier in the development of the solution, ensuring that what IT builds is what the business needs. It also allows all stakeholders to identify and avoid costly mistakes before they are made.

From a project-management perspective, the VPW helps reduce the risk of missed deadlines by providing more precise information on which to base accurate and achievable project timelines. Furthermore, it provides a method to quickly uncover and find workarounds for project showstoppers (e.g., data quality issues).

After an initial short exercise to create the prototype, the VPW aids project approval by allowing key stakeholders (including executives) to see the value of the target data model or repository up front and to determine whether it meets their requirements. Because the VPW is tangible, users can clearly see the capabilities they will receive, and decision makers can visualize the potential benefits. With this invaluable example, buy-in is much more easily achieved and continues throughout the project lifecycle.

In addition, this approach allows IT to fine-tune requirements to make sure they deliver what business users want. To further facilitate organizational support, the VPW can be tweaked to provide specific information to appeal to different stakeholders. For example, one view can be prototyped to convey a targeted message to C-level executives and a separate view can be generated from sales data to show value to regional sales managers.

A VPW that is developed at the beginning of a BI project will yield benefits throughout the entire development cycle. A VPW complements most data management tools, accelerates the implementation of the BI solution, and helps improve the overall performance of the end product.
Figure 1 demonstrates a typical BI development process that utilizes many of the more popular categories of data management tools. It describes the benefits that can be achieved at each phase.

Thanks to their ease of development, VPWs can be used after project implementation to develop and manage an ongoing change management strategy to keep up with changing business requirements.

**Potential Pitfalls**

VPWs have many benefits, but they are not without risk. For example, what if users like the VPW so much that they begin to question the need for a production data warehouse? To prevent this, set expectations early in the project, emphasizing that the intention is not to use the VPW in a production environment but rather to validate business requirements.

There are several reasons why a VPW should not be used in place of the production data warehouse. First, owners of operational applications will likely not want to compete against VPW users for their computing resources on an ongoing basis. Second, the query performance of a VPW will never match that of a physical data warehouse, which uses optimized indexing. Finally, a VPW offers no true data persistence and therefore no mechanism to handle slowly changing dimensions.

Another concern arises when the VPW produces different information from the existing BI, data warehouse, data mart, or OLAP applications. This can be viewed as either an opportunity or a risk, as it may introduce doubt about previously trusted information sources. One way to address this scenario is to proactively create an environment that promotes a “culture of opportunity” where the users are enabled and emboldened to help clean up “dirty data.” By giving users a more active role in data stewardship and data ownership, tactical and strategic company metrics will more closely match actual results, and better decisions can be made.

Another potential pitfall of a VPW is the effect it may have on existing applications that access the...
same data sources. Because virtual prototyping allows developers and users to directly access operational data sources, resources may become constrained, negatively affecting primary users of the data. This risk can be reduced by allowing the federated queries to run only during non-peak hours or by carefully monitoring their resource consumption.

A fourth potential pitfall of using a VPW is that it may contribute to “scope creep.” As users access data they’ve never seen before, they will want to see and explore even more. To counter this situation—which many would not consider detrimental to the project—put controls in place to keep developers and business users focused on the project requirements. Virtual prototyping should be viewed as both a tool for exploration and as a means to quickly validate existing requirements—with emphasis on the latter for project-related work.

**Improving Business Value**

The direct result of a VPW is ease of communication, but a more powerful effect is improved business value. The intangible benefits described thus far ultimately lead to real dollars saved through greater support and involvement from executives, improved requirements gathering, properly directed development efforts, less rework, and faster implementation.

“Time to value” is a driving component of improved business value. If it takes two years to roll out an enterprise data warehouse, or six months for a line-of-business data mart, the effective business value is reduced by the opportunities that are missed during these long development cycles. The traditional “waterfall” approach to project management often leads to such lengthy development cycles (Royce, 1970).

Figure 2 shows how the traditional implementation process addresses user testing and changes once the development cycle is almost complete. This static, “waterfall” approach to project management manifests the shortcomings outlined earlier.

Conversely, an implementation that utilizes a VPW involves users for testing and validation early in the development cycle, and keeps them involved throughout (see Figure 3). This project management model can be iterated as often as necessary to achieve the desired results, and helps to build a better relationship between the business group and IT.

As a result, choosing to create a VPW before rolling out a static production version reduces the time, effort, cost, and risk related to BI initiatives or data-sharing implementations. As shown in Figure 4, the VPW provides ROI to the business community sooner. In
addition, it maximizes the value of many BI or related initiatives, including:

- **Business performance management (BPM):** At the top of the business intelligence food chain, a BPM application relies on solid metrics and requirements to drive the business effectively. The underlying data structures needed to populate a BPM application can be derived more quickly and accurately using a VPW, with quicker buy-in and ROI to executive management.

- **Enterprise data warehousing (EDW):** The high demands on reporting for business intelligence require integrated data from many disparate data sources. A VPW provides a mechanism early in the development process so business stakeholders can identify what data should be included in and excluded from the final data warehouse design.

- **Data marts:** As with enterprise data warehousing initiatives, line-of-business data marts require discovery and integration from multiple sources and typically have a shorter time frame for implementation. These factors make a strong case for virtual prototyping.

- **Enterprise architecture:** A VPW enables an organization to quickly document what information exists within baseline systems, helping it to better structure its enterprise MDM initiatives. Organizations of all types and sizes want to gain a better understanding of what data they have and where it exists, and store that information in a master repository or an operational data store (ODS). Virtual prototyping facilitates a quicker turnaround.

**Using VPW**

The following scenarios illustrate how a VPW can help bridge the business/IT gap and cement a positive relationship between the two groups, resulting in more successful projects.

**Business-Case Justification**

When IT project prioritization and funding is a challenge, simplifying the approval process for IT projects is a must. Early in the development cycle or as a pilot, the VPW can be used to justify the data consolidation projects. Furthermore, the VPW provides a way to more accurately predict costs and enables the business to apply appropriate budgeting.

**Reverse-Engineer Production Systems Requirements**

With a VPW, IT organizations can quickly build multiple views of the enterprise data to help them better understand
business requirements. With samples of a finished data warehouse, IT can work closely with business users to build use-case scenarios that describe how the business would use each configuration. The business can define what data they want and what type of performance they expect, giving IT clear instructions on how to tweak the final product.

For example, a customer service representative may need real-time access to all customer data to address inbound support calls, so a full-scale data warehouse would be necessary. Business and IT can work together to define the criteria, identifying scales of timeliness and breadth of information to generate the right solution for each business scenario.

Prototype for Production System
Because a VPW creates a single but flexible view of the data, it is ideal for testing and verification at both the business and IT levels. This functionality enables potential modifications to be identified, modeled, reviewed, and approved more quickly and at less cost than other methods.

When building a data warehouse or data mart, testing typically does not begin until a solution is nearly complete. Developers may have to start over to correct errors, adding both time and cost to the project. A VPW enables users to create a replica of the target system in a fraction of the time and cost it takes to create the production system.

A VPW enables testing sooner and at various levels of the production system (such as sizing, tuning, security, validation of mapping relationships, and identification of data quality or standardization issues) and verifies the data to be included in production system.

Working Example: Virtual Prototyping Warehouse
The first step in designing and building a data warehouse is requirements gathering, including the identification of key data sources and the warehouse’s key performance indicators (KPIs). Automated mapping and data-profiling tools can help identify key data sources, but some domain knowledge is necessary to pare all of the enterprise’s data sources into a manageable, useful list that corresponds to the initial business requirements for the data warehouse or data mart.

Once the data sources have been identified, KPIs should be outlined. KPIs are quantifiable, agreed-upon measurements that reflect the critical success factors of the data warehouse initiative. In a data warehouse, KPIs are typically numeric values or facts that users can slice and dice by various dimensions. Typical dimensions include time, geography, customers, products, and employees.
The goal of a VPW is to validate the data sources for KPIs and help determine the data sources for the needed dimensions and associated drill-down hierarchies.

A common data warehouse KPI is the time dimension. Time dimensions are standard in their granularity (year, quarter, month, week, day, and so on), but other dimensions will likely be unique to every business. Thus, the “discovery” of these dimensions is fundamental to understanding the KPIs that drive the business. As a best practice, we suggest breaking the task of discovering these operational dimensions into manageable chunks, one dimension at a time. Rather than prototyping an entire star schema (the most common data warehouse schema) with all of the final dimensions, start small and prototype one dimension at a time, including the measures that form the basis of your KPIs for each dimension.

A common dimension that can span many data sources is the customer dimension. Customer name and region might come from one data source, customer sales measures from a second, and customer type from a third. The data would be related by primary and foreign keys.

Virtual prototyping allows the data warehouse designer to issue federated queries across the enterprise and return all values of interest from all of the candidate dimension data sources. This will allow a consistent set of dimension values to be agreed upon by all parties. This is also needed to create consistent hierarchies within each dimension. Failure to do so may lead to invalid aggregate values and inconsistent rollups when fact tables are designed, built, populated, and queried.

Once the requirements-gathering phase is complete, the following basic steps should be considered when implementing a prototype:

1. **Create ODBC/JDBC connections**: ODBC or JDBC connections to each data source enable subsequent discovery of the conformed dimensions (and the associated KPIs). This can be a manual or
automated process, depending on the availability of mapping tools.

2. Profile and discover: After the data source connections have been established, either use an automated data mapping tool (with data federation capabilities or a bridge to a data federation technology) or manually discover and use a data federation technology. The automated mapping and discovery tool will retrieve data-source schemas, their relationships, and basic data profiling data. A data federation tool will typically retrieve only data source schema information.

3. Identify tables of interest: Once the data source catalogs and schemas have been retrieved from the candidate data sources, identify tables of interest. Much of the heavy lifting can be handled by automated mapping and discovery tools, but some domain knowledge is required to finalize relationships. If only the data federation technology is available, the process can be manual and time-consuming.

4. Determine relationships and mappings: A key component of the discovery process is identifying the primary and foreign keys (or other join criteria) for the tables of interest. An automated data mapping tool will discover most of these intrinsic relationships, or the relationships can be manually inferred by writing SQL queries.

5. Build federated queries: Once an understanding of the data sources has been gleaned, the VPW designer can design and build the federated queries needed to further explore the candidate data sources for dimensional data and fact table measures (KPIs).

6. Evaluate the results: Evaluating the query results includes looking for consistency of data values; understanding the definition of the data; and evaluating the frequency distribution and data latency (e.g., date/time stamps). Get the most up-to-date and consistent data to determine the proper attributes and their sources for use in the conformed dimensions.

7. Create and publish federated views: Select the chosen dimension attributes and combine them in a series of saved, federated SQL queries. These federated views must include the appropriate fact table measure columns (KPIs) and be grouped in the SQL statement by dimension. Check that the results are reasonable and that the KPIs are accurate.

8. User validation: Once the VPW has been validated by IT staff, release it to end users. Typically, all that is required for access to the VPW is an ODBC/JDBC connection to the data federation server and a BI analytic tool. Because it is not optimized for performance, a VPW will not offer the same query response or permanent data storage as a normal data warehouse or data mart, but the VPW will provide a much quicker mechanism of validation for the business community, without the overhead and cost of building a complete data warehouse or data mart.

Although this is only an example, locating the correct data is a common goal for virtual prototyping. This can be accomplished only by accessing the relevant data sources using a single federated query. Users can view the results and judge whether the desired information has been captured.

Virtual Prototyping: Key Features
To realize the full potential of the VPW, embed it in the corporate culture as a best practice, and make it part of the project management methodology. In addition to enhancements to existing processes, consider the following key technical features and capabilities when
selecting data federation tools or building a virtual prototyping framework:

- **Data federation capabilities with automatic determination of data relationships across data source tables and columns:** Data federation is the key enabling technology for virtual prototyping. The technology should provide the technical basis to query heterogeneous data sources concurrently and return the result set as a single view of data, with intelligent data source tagging. The technology should also automatically determine data relationships across data source tables and columns.

- **An intuitive, user-friendly interface with instantaneous “what-if” analysis capability:** This allows rapid data visualization and the quick building of virtual views that include or exclude specific data sources or elements.

- **Compatibility with common reporting and analytic tools:** The data federation engine must support ODBC or JDBC access (preferably both) so that views created and published through the data federation interface can be treated as a single data source in most applications.

- **The option to share metadata with other data-sharing products:** One of the benefits of virtual prototyping is its ability to capture new metadata and business definitions for data elements. The ability to share this information with other data-sharing initiatives via metadata repositories helps leverage and show value to the virtual prototyping environment.

- **Automatic configuration and provision of ETL processes with other products:** For virtual prototyping of OLTP or operational systems, the automatic configuration and provisioning of ETL sessions, processes, and pipelines with other products allows
organizations to leverage any relationships discovered during the prototyping.

- **A small footprint, eliminating the need for large capital hardware expenditures**: The ability to install the data federation engine on an existing server or hardware platform is one of its benefits. Queries to the data federation server are then issued from client workstations, requiring only an ODBC or JDBC connection.

One of the benefits of virtual prototyping is its ability to capture new metadata and business definitions for data elements.

In addition to these key features, the organization must build a technical architecture to support a virtual prototyping environment (see Figure 6). Such an architecture should provide access to all of the necessary data sources. Standard query tools should be used to access the federated views, and the architecture should support interfaces to third-party tools such as ASG Rochade, CA-ERWIN, and Informatica. Once the necessary metadata has been collected, and the dimensional and fact data have been validated, you can build physical models and deploy BI applications more quickly than with typical prototyping methods that use waterfall project management techniques.

**Conclusion**

The implementation best practices and business cases described here demonstrate how building a VPW can benefit organizations in any development phase. Through cost reduction, accelerated time frames, and greater involvement from business stakeholders, BI projects have a better chance of succeeding with virtual prototyping.

In November 2005, analyst firm Gartner agreed that using a prototype (which it defines as “an iterative or spiral approach, but allows for adaptation of requirements based on customer feedback from prototype-assisted design reviews”) yields a more than 10 percent improvement in development productivity compared to the more traditional “waterfall” project management method (Duggan, Hotle, Light, and Solon, 2005).

Lowering the total cost of delivery and enabling business users with better data (knowledge) faster and sooner (access) and with consistent results (reliability) should be the goal for all business intelligence and data integration projects. Virtual prototyping is one approach that helps make this happen.

**References**

