BI, Built to Order, On-demand

Automating data warehouse delivery

January 2015

A White Paper by

Dr. Barry Devlin, 9sight Consulting
barry@9sight.com

A quiet evolution is underway. While media attention focuses on big data, the ongoing work of delivering business value through daily decision making support still continues through business intelligence and data warehousing efforts. The quiet but conspicuous evolution has been in the ease with which these projects proceed. The evolution has been data warehouse automation.

This paper explores three key ways in which data warehouse automation changes the design, development and ongoing maintenance of data warehouses and marts. First, we examine how automation addresses the old conundrum of delivering consistent, quality data in the timeframe demanded by modern business needs. Second, we explore how streamlining the overall process provides a single repository of metadata and integrated tooling to speed and simplify development. Third, we observe how business and IT can truly collaborate in delivering business solutions.

Using examples from three WhereScape clients, a large bank, a smaller credit union and a small, for-profit childcare organization, this paper offers a clear business and IT rationale for adopting data warehouse automation and shows its benefits across very different business types and sizes.

Contents

Sponsored by:
us.wherescape.com

WhereScape®

Data-driven business has become a modern mantra of consultants and vendors in data management. The vision of Data Lakes filling commodity server farms, running on free, open source software, promises new opportunities to predict customer behavior, perfect production, and overwhelm the competition. Whether from social media or the Internet of Things, a wave of data, exabytes high, is breaking on business shores, raising business expectations of data, and its deliverers, to unprecedented levels. The excitement is palpable.

And yet, for students of the history of data management, alarm bells are ringing. Haven't we been here before, and struggled to deliver?

Strangely enough, this paper is not about big data and its delivery. While the big data wave has rolled forward, a quiet revolution has been taking place in the old world of data warehousing. New approaches to addressing old problems of defining, delivering and maintaining data warehouses and marts have emerged. New tools and techniques have begun to deliver data more quickly and accurately to satisfy business intelligence (BI) needs. Agile approaches simplify and speed maintenance of the data resource as business needs change. We call it *data warehouse automation*, and for those businesses adopting it, the gains are significant.

Most of us are aware of business' unending frustration with IT's roadblocks to data. But, contrary to another current myth, the answer is *not* free for all, self-service access by users to every piece of data through some sexy visualization tool. That comes after the data is consolidated, blended, cleansed and certified for use. The creation of a high quality data resource has always been what data warehousing has been about. And that's what data warehouse automation is about too—but faster, better and more flexible than traditional ETL (extract, transform and load) tools.

Data warehouse automation is the use of an integrated set of tools and techniques that automate the design, delivery and maintenance of data warehouses and marts.

With data warehouse automation , we can move from IT's old need—or necessity—to control everything to empowering both business and IT people to each do what they do best. Business defines what data is needed and how it should be analyzed iteratively, with IT capturing the business needs and applying quality and production values in flight.

But, does any of this matter in the new world of big data? Absolutely. Much of the long-term value of big data emerges only when this is combined with business data that has been cleansed and consolidated through traditional data warehousing approaches. And, for big data itself, how can you hope to create high-quality big data if you cannot create and maintain high-quality traditional, "small" data?

The swinging pendulum— data quality vs. time-to-value

The pendulum swings again from data consistency and quality to speed of delivery and time to value of business decisions. And then back again. It's an old story… back and forth we go. Data warehouses major on consistency and quality of extensive data, garnered from a variety of sources and integrated according to a common enterprise model. It's often a slow and painful process. Data marts, on the other hand, focus on satisfying the known needs of a set of business users for a relatively limited set of data as quickly as possible. The problem arises when you have twenty different but overlapping data marts offering competing answers to the same question.

The history of data warehousing is littered with the corpses of enterprise data warehouse projects that simply took too long for the business. They were overtaken by a plague of data marts that started well, but were overcome by the complexities of multiple feeds and inconsistent answers. Which in turn, led to another data warehouse project designed to integrate the data marts. In such situations, IT becomes the (often reluctant) guardian of quality, the promiser of consistent data. It takes time and effort using traditional development approaches. Worse still, as business needs change at an ever-increasing tempo, maintenance becomes an unwinnable race against time.

Business, exasperated with IT, builds stand-alone solutions. Speedy delivery is possible—at least with smallish data marts and in the first iteration. Users are happy. And then they encounter quality and consistency issues. Maintainability also becomes a nightmare.

Of course, the real requirement is to have *both* consistency and speed of delivery, quality and minimum time to value. Data warehouse automation addresses both needs by integrating the design and delivery of the data model, database structure, and the population process in one place—whether for a warehouse or mart. All the design and population metadata is stored together in a single repository, allowing development to flow smoothly and iteratively from user requirements, through database design, to creation of population routines. By integrating all the steps of the design and development process, consistent and quality data can be delivered quickly to the business for immediate review and early acceptance. Time to value is minimized and the business is delighted.

Automating the full warehouse / mart design and development cycle allows the needs for data consistency and time to value to be correctly and equally balanced.

Compare this to the traditional approach. Modeling, database design and development of population routines required multiple, disconnected iterations involving business users, modelers, database administrators and ETL programmers at different times, each using different and unconnected tools. These gaps and tool transitions slowed the process and gave rise to design errors and inconsistencies. A slow, sequential process spread over weeks or months left IT exhausted and business frustrated.

Now, the common environment and shared metadata repository offered by data warehouse automation overcomes these problems. What was separate steps by different parties becomes a collaborative team effort between business and IT.

Maintenance and ongoing development also benefit directly from the existence of such a shared repository, where a current, ongoing and reliable description of the data and processes of the entire mart or warehouse is available for inspection and use at all times. Updates and changes of content and

processes are easier to do and carry less risk of unintended consequences when undertaken with automation tools based on the metadata repository.

When the repository is stored in a standard relational format, the metadata is directly accessible through SQL, eliminating lock-in to the automation tool and easing migration between platforms. Metadata created to support a data mart can be used as the basis for a data warehouse, even on a different database platform, and *vice versa*. This is particularly useful in the evolutionary situations between marts and warehouses described above.

Example: Union Bank gets compliant with WhereScape

MUFG Union Bank, headquartered in San Francisco and operating in seven US states, is a wholly-owned subsidiary of multinational Bank of Tokyo-Mitsubishi UFG, with a strong focus on meeting rapidly changing customer needs. In the past, that generally meant business units building their own stand-alone data marts and similar shadow IT systems, an approach that was cited a number of years back for data quality problems by the bank's risk and compliance department. The result has been a move towards more centrally-managed data marts (based on SQL Server) and a plan to migrate these and some Oracle-based data warehouses to a Greenplum-based enterprise data warehouse (EDW) over time[1].

All these migrations are being supported by WhereScape's data warehouse automation tool, RED. Using this approach, deployments take less than a third of the time they would have taken using traditional tools, and changes can be made in minutes if necessary. Migrations of data marts to the EDW are simplified by the ability to port metadata from one platform to the other. Perhaps most interestingly, the balance between business and IT in development teams has shifted dramatically from heavily IT-oriented to much more business-focused, driving high-speed, agile iterations in delivery. Previously, it took three developers to work through the requirements of one business analyst at Union Bank. Now, one developer can support three business analysts.

> *WhereScape enabled us to design, develop, document and deploy a production-ready solution in 8 weeks. Using traditional data warehouse development methods would have taken us 6-8 months.*
>
> *Ryan Fenner, VP, Data Solutions Architect, Union Bank*

Easy for IT means easy on the business

Business intelligence begins with business and ends with intelligence. Obviously! So why is so much focus put on IT? The short answer is that between that business need and the desired intelligence, there's a lot of intensive preparation work to do. Work best done by people with a particular knowledge and skill set around data: IT. We've already seen the danger to data quality if business takes over this preparation entirely. But can we make this preparation faster and easier for IT? And can we bridge that initial gap between business need and data preparation?

For more than two decades, data warehouse consultants and experts have advocated that data warehouse builders automate this preparation. They should avoid hand-coding or scripting ETL programs. Eliminate documentation by coders (who don't do it anyway). Rather, adopt rules-driven engines with a drag-and-drop design process and that run (often in parallel) on dedicated servers. And for nearly two decades, many developers have been ignoring that advice. Why? Because hand-coded procedures already existed or were needed for some very specialized transformations, reducing the attractiveness of the ETL tools and impacting the business case for another server in the warehouse environment. Because the ETL tools introduced yet another design environment and metadata store into the project, moving IT further from the business users.

But there's one more point. ETL tools automate only one step—the creation of the population procedure—of the full data warehouse delivery process. While this is valuable, it automates only one part of the whole process. The developers are left to manually link this step to data modeling and database design. And because the whole process is so iterative—the warehouse or mart design will often change as a result of discoveries made when building the ETL, and *vice versa*—automating only one step does not give as much benefit as might be assumed. The real strength of data warehouse automation is that it addresses the whole cycle—modeling, database design and management, development, operations, logging, documentation, and so on.

Data warehouse automation redefines traditional ETL by expanding the scope of automation to the full development and maintenance process.

Data warehouse automation also simplifies the data preparation task through one more radical move: swapping the letter order from ETL to ELT. Doing transformation after load—in the warehouse or mart environment—has far-reaching consequences, making life a lot easier for IT and getting business much more intimately involved. Doing transformation in the warehouse means it can be written in SQL, optimized with the underlying relational database function, and make use of all the strengths of the database server. There's no additional ETL server license, one less data transfer to do.

The traditional design and development process comprises at least three separate phases:

1. Business requirements and warehouse/mart modeling, involving the business users
2. Population design and warehouse/mart optimization, excluding the business users
3. Warehouse/mart (as proposed/delivered) review by business users

In the first phase, the work is usually done in design tools that offer no access to actual data. This initial disconnect with the reality of existing data may lead to beautiful theoretical models that are difficult or even impossible to populate, as is discovered in phase 2, when IT goes away and dives into the source databases. In the third phase, which often occurs some considerable time after phase 1, users

are presented with something which is often significantly different to what they saw and agreed to earlier. Immediately, the business begins to doubt the ability of IT to deliver what they need and question the eventual relevance of the final result. One or more further iterations through the three phases are usually needed, increasing business doubts and discomfort.

Using a data warehouse automation approach, the above process from data sources to business needs is dramatically simplified. It is reduced to a single modeling and design phase, involving business users, data modelers and population experts, who work with the actual source data and explore what can realistically be delivered. The data design and processing needed are captured in metadata during the process and the required SQL is auto-generated on the spot. Any iterations required—and, of course, there will be some—can be done immediately or later, based on the one, common environment. The users see immediate results and IT gets instant feedback. Furthermore, there's no code for IT to design, write, test and (not) document. The generated metadata and SQL procedures are the output. The result is a process that involves multiple iterations of prototyping; failing fast and redoing, converging to a final "prototype" that is actually the fully production-ready final product. Business users are intimately involved at all stages of the process so business buy-in is assured. Users can refine requirements right up to the moment of production deployment.

Example: Delta Community Credit Union empowers IT with WhereScape

Delta Community Credit Union is Georgia's largest credit union, with assets exceeding $4.7 billion and more than 310,000 members. Delta Community provides its members with a typical full service array of financial products including checking and savings accounts, credit card services, insurance, investment and financial programs.

To improve responsiveness to the financial needs of its fast-growing membership, Delta Community developed a Teradata enterprise data warehouse to consolidate and analyze its loan and member information. The company selected WhereScape to rapidly create Teradata native objects, including tables, models, views and indexes, as well as set-based procedures to load the data warehouse.

Delta Community Credit Union's first project with WhereScape focused on membership analytics and was successfully deployed in three months. Projects specific to deposit portfolio performance and business lending were then completed in the following six months.

These speedy results contrasted with Delta Community's prior data warehouse implementation attempt which had simply failed to deliver. A review of that effort revealed cycles of lengthy requirements analysis by the business that produced extensive documentation, followed by a separate design phase and significant levels of rework by IT, that often delivered no usable or timely results. Business was understandably frustrated and skeptical that this second attempt would do better. However, the agile approach offered by WhereScape, where requirements could be quickly prototyped and delivered, enabled the incoming project leader to build a new sense of partnership between the business and IT. And from the IT point of view, the self-documenting features of WhereScape proved a great boon to a small team with a limited budget. The outcome was a satisfied business community, whose ever-evolving needs were being met speedily and successfully, providing significant opportunity for business growth.

> *WhereScape's integration and best practices empowered us to take on a large implementation that quickly delivered value to our stakeholders. The agile, iterative development approach is ideal for us, because it helps us quickly adapt as our business needs dictate.*
>
> *Su Rayburn, AVP, Information Management & Analytics, Delta Community Credit Union*

Business and IT—hand in hand

The old world where business and IT live on different planets is so last millennium! Sandal-wearing IT in the data center vs. pinstriped business suits is rapidly moving to a new paradigm: the biz-tech ecosystem[2]. This phrase recognizes the new reality that most, if not all, new business advances and competitive advantage today spring from information technology. To succeed, business must take full advantage of IT knowledge and skills. And IT must step up to becoming knowledgeable and fully involved in the business too. This symbiosis of business and IT is still emerging, but its clearest need has long been in the area of business intelligence.

In traditional IT projects, the starting point is requirements gathering, a procedure where IT hunts down reluctant business people, sits them in a conference room and feeds them pizza until a requirements document is produced. This approach is particularly ineffective in BI. Business needs are often loosely described, simply because the users cannot know how they really want to play with the data until they have the data to play with. Fearful that they may miss out on some important information, and expecting that future changes may be difficult or costly, users tend to ask for *"all the data, now"*. Such requests lead only to IT frustration and business disappointment; they cannot be satisfied.

Data warehouse automation starts from a different space. Based on the agile manifesto[3] and approaches arising from it, data warehouse automation brings users and IT together from the beginning to work as a team. Initial, known business needs—however sketchy—direct the first investigations of what is possible. Hands-on prototyping of data mart deliverables starts early and continues throughout the process. Users can play with the prototype and discover if it begins to meet their needs. And having played with the data, they may find that their needs are actually different to what they first thought.

Building on agile development principles, data warehouse automation engages business and IT in a fully collaborative development process.

Users see the challenge of getting the right data from the underlying sources and can even contribute their knowledge of what the right sources might be. Formal requirements documents are eschewed in favor of working models, where the data definitions and population steps are recorded directly in the metadata of the warehouse or mart. No more throwing documents back and forth over the fence. With ongoing delivery of small wins, the cost of change is reduced. Business and IT are free to fail, empowered to innovate.

Data warehouse automation moves the relationship between IT and business from competitive or contentious to highly cooperative. Collaborative design, development and ongoing evolution and maintenance is enabled when business are involved in the full process. IT works at the pace that the business wants; business is available when IT needs direction.

Example: Learning Care Group educates business and IT together

Learning Care Group operates more than 900 corporate and franchise schools across the country and around the world with the capacity to serve more than 100,000 children between the ages of six weeks and 12 years.

Built through a series of mergers and acquisitions, the five brands within Learning Care Group have long maintained different systems, different ways of managing data as well as different business defi-

nitions and metrics, making it difficult to measure and manage company's progress at various levels. In addition, the company had no data warehouse for reporting and analytics, opting instead to capture data on a weekly basis and provide snapshots of the data to business users.

> *WhereScape was a key ingredient in enabling LCG to deliver an information delivery platform with a limited budget and short time to execute that will continue adding and maintaining the agility required to serve the company's needs for business intelligence.*
>
> *Ahmad Malik, VP Technology, Learning Care Group*

Learning Care Group wanted to rapidly deploy an information delivery platform consisting of a Microsoft SQL Server data warehouse and BI solution to help to streamline business processes and understand the data across its brands, as well as plan for future data integration / application consolidation projects. The company also wanted to simplify information availability across all levels of the organization, with limited project budget and staffing. Conventional ETL solutions to feed the data warehouse were deemed too expensive and technically difficult for existing staff, who were relatively new to the world of data integration and data warehousing. WhereScape was selected to streamline the warehouse development process and to engage business and IT staff in a collaborative environment that enabled rapid delivery and expansion of an integrated data warehouse environment.

By taking a well-managed, highly productive, iterative approach to development, facilitated by data warehouse automation tools, Learning Care Group could focus resources where needed—on real, analytic and reporting work by and for the business.

Conclusions

Although big data and, increasingly, the Internet of Things have been making all the headlines and Hadoop has become the darling of the analysts, the need for good old business intelligence has not gone away. In fact, it can be argued that the requirement has become even stronger as businesses move ever faster, attempt to reinvent themselves to compete, and comply with increasing levels of regulation. Businesses large and small need to deliver BI data warehouses and marts more quickly and more efficiently—simply to survive.

The traditional approach to design and development of data warehouses and marts is a relic of the earliest days of the industry in the 1980s. It may have worked for large organizations with highly skilled and extensive IT departments. And in reality, even for them, it presented challenges. In today's fast moving business and slimmed down IT, the traditional development approach is often no longer fit for purpose.

Data warehouse automation has been operating in stealth mode for more than a decade now. The products have matured to the stage where they are appropriate from the smallest to the largest implementations. The benefits accrue over three distinct areas. First, the approach provides a good bal-

ance between data quality and time to value by eliminating unnecessary gaps and hand-overs in the integrated design and delivery process. This benefit extends from initial concept through ongoing maintenance and upgrade of the warehouse or mart. Second, by simplifying and streamlining the design and development tasks of IT, business sees immediate benefits from data warehouse/mart projects: earlier delivery of better deliverables with lower business involvement costs. Third, a new collaborative relationship emerges between business and IT that drives more innovative and successful future BI developments.

Data warehouse automation is central to delivering both high quality decision making information and extreme agility in reacting to ever-changing market needs.

As the scope and variety of data used by business expands, and further emphasis is placed on data-driven decision making, it becomes ever more imperative that the core business information in business intelligence systems is of both high quality and extreme agility. Data warehouse automation is central to delivering on both of these aims.

Dr. Barry Devlin is among the foremost authorities on business insight and one of the founders of data warehousing, having published the first architectural paper on the topic in 1988. With over 30 years of IT experience, including 20 years with IBM as a Distinguished Engineer, he is a widely respected analyst, consultant, lecturer and author of the seminal book, "Data Warehouse—from Architecture to Implementation" and numerous White Papers. His new book, "Business unIntelligence—Insight and Innovation Beyond Analytics and Big Data" (http://bit.ly/BunI-Technics) was published in October 2013.

Barry is founder and principal of 9sight Consulting. He specializes in the human, organizational and IT implications of deep business insight solutions that combine operational, informational and collaborative environments. A regular tweeter, @BarryDevlin, and contributor to ITKnowledgeExchange and TDWI, Barry is based in Cape Town, South Africa and operates worldwide.

[1] Stedman, C, *"Former Shadow IT Worker Helps Bring Analytics Data into the Light"*, May 2014, http://us.wherescape.com/searchdatamanagement-former-shadow-worker-helps-bring-analytics-data-light/

[2] Devlin, Barry, *"Business unIntelligence—Insight and Innovation Beyond Analytics and Big Data"*, (2013), Technics Publications LLC, NJ, http://bit.ly/BunI_Book

[3] Principles behind the Agile Manifesto, http://agilemanifesto.org/principles.html