



The Bloor Group

**PUT IT ON AUTOMATIC:
CLOUD DATA WAREHOUSING HAS ARRIVED**

by Eric Kavanagh

Location, location, location! The long-standing mantra for brick-and-mortar business models speaks volumes about the value of locale. Secure the ideal physical location for your operation, and you will all but assure your long-term success.

My, how times have changed... sort of. These days, there's a new epicenter of activity, but it's not really in one physical location, but many. It's called the cloud, and it's a pervasive phenomenon in business as in life.

The other widespread reality in today's world? The power of data. Thanks in large part to the vast amounts of information created by the Web, companies large and small have recognized the intrinsic value of data. The key is to identify, organize and analyze all relevant data.

That's the purpose of data warehousing, a long-standing discipline that often separates winners from losers in big business. Since at least the 1990s, data warehouse practices have helped transform the nature of business strategy, empowering decision-makers with critical insights.

Historically, only the largest, most well financed enterprises could reasonably afford a data warehouse. The vast majority of companies just didn't have the margin to purchase the software, let alone finance the consulting or internal costs of building and maintaining one.

That was then. This is now. As cloud computing has matured over the past 30 years, dramatic changes have bent the cost curve down in dramatic fashion. Powerhouses like Amazon have driven computing costs lower and lower, fundamentally reshaping what's possible online.

At the nexus of these two powerful forces, we find a third component that completes the trifecta, and empowers data-driven companies to reinvent their future: automation. As any serious programmer will tell you, automation is the heart and soul of every application ever written. When applied properly, it shortens time-to-value from months or weeks, sometimes all the way down to days or even minutes.

When automation is leveraged in the realm of cloud data warehousing, some amazing things can happen. Most notably, businesses can get right to the task of analyzing relevant data, instead of waiting weeks or months for critical data sets to be added. This has long been a serious hurdle with data warehousing: projects take too long to deliver, are too expensive, and ultimately become politicized due to the complexity and resource-intensive nature of the work.

One of the original pioneers of data warehouse automation is WhereScape, a company founded in New Zealand back in 2002. WhereScape recognized the potential value of data warehousing, and realized there would be a need to automate the otherwise mundane tasks associated with building such solutions. Nobody likes tedious. That's a universal norm. By focusing on automating the data

integration component of data warehousing, WhereScape helped create what has become a whole market segment, one that expedites and amplifies the value of data infrastructure.

There are other key benefits to data warehouse automation besides time-to-value, which can typically be expedited by an order of magnitude. The first is quality control. When people are tasked with doing repetitive processes ad nauseum, the propensity for making mistakes increases. Very few consultants enjoy doing mappings for extracting, transforming and loading data into a warehouse (ETL).

This is a serious issue for a variety of reasons, including the law of error propagation. One mistake in an ETL mapping can wreak havoc on downstream systems. All it takes is for a single field to be erroneously mapped, or incorrectly transformed, and the data which an analyst uses for decision-making gets thrown off, sometimes dramatically. This can lead to faulty decisions, thus undermining the entire warehousing effort.

By automating ETL mappings from source to target, companies can seriously mitigate issues such as error propagation. What's more, by using a automation software like WhereScape, they can provide visibility into the otherwise complex, even spaghetti-like matrices that constitute warehouse integration processes. This facilitates troubleshooting, especially when rules are put in place to automatically identify potentially problematic ETL scripts. Over time, it also greatly reduces ETL maintenance costs, which can spike at any time, causing significant financial pain.

And then there's documentation. Often the most overlooked and underappreciated aspect of programming, documentation is critical for ensuring the integrity of data management processes. But once again, by relying on programmers to assiduously document their every move in coding, companies are basically asking for trouble. A developer would need to have an obsessive-compulsive disorder to do so efficiently. So, why not automate that process, too?

This component is especially true for any organization subject to compliance regulations like Sarbanes Oxley, Dodd-Frank, the Affordable Care Act, HIPAA or other such mandates. And as the Global Data Protection Regulation enacted by the European Union comes online, there will be another far-reaching set of rules that must be respected in regards to data lineage and use.

The automation of these key processes therefore increases chances of success significantly, especially when compared to environments in which hand-coding is the alternate method. It pays to appreciate that software is coded in languages, and such languages are not always easily understood by developers who didn't write the code. By automating processes, organizations can avoid painful situations when coders move on to other jobs, or retire.

The 'There' There

To paraphrase the great Gertrude Stein, there really is a 'there there' with cloud computing. Just follow the giants: IBM, Oracle, Microsoft, SAP and Google have all joined Amazon's Jeff Bezos in carving out their place in the virtual sky of high-performance computing. The center of gravity is now cloud.

There are many significant benefits of cloud computing, as compared to its predecessor, the company-specific, on-premises data center. In short, businesses can offload the responsibility, cost, strain and headaches of purchasing hardware, then buying and installing software, and finally handling maintenance. No more need to amortize those assets; they're just rented!

And there are very palpable, tactical benefits of cloud computing for data warehousing. Specifically, data that lives in the cloud already -- in Amazon's S3, or Microsoft Azure, or wherever -- is very easy to load and provision, as compared to traditional on-prem data which takes longer to ETL into a warehouse.

That capability was showcased in a recent Deep Dive webcast featuring Jason Laws, vice president of product management for WhereScape. Laws did a live demo of his company's automation software being used in conjunction with Snowflake's cloud-based data warehouse solution, and the results were undeniable. Cloud-based data can be loaded much faster than on-prem data, in fact by orders of magnitude.

The fact that cloud-based data can be so quickly and easily loaded into a cloud data warehouse is tremendous news for the following reason: the preponderance of data being created today is cloud native, and the percentage is actually increasing as compared to on-prem data. This is especially true in the data-centric world of sales and marketing automation.

According to the Chief Marketing Technologist Blog, there are now nearly 5,000 software companies playing in the MarTech space. What's more, the pace of growth in that sector is stunning. In 2011, the estimate was 150 companies, up to 350 in 2012, then 1,000 in 2014, 2,000 in 2015, and 3,500 in 2016.

From a data integration and management perspective, that sounds like a Gordian knot. After all, even though there are some general standards for data, such as common field names like first name, last name, email address and so forth; the bottom line is that each and every one of those vendors has a proprietary data model. The intrinsic level of complexity is utterly massive.

And yet the potential value of weaving together the big picture is downright tantalizing. As any marketer knows, additional sets of data can be invaluable for analysis, whether for customers, prospects, companies or other entities. The challenge is to integrate disparate data sets around key dimensions, in order to glean insights about what they might buy or promote.

Absent the power of automation, stitching together even a handful of data sets is a serious challenge. Historically, organizations would create direct connections between systems using enterprise application integration (EAI) software. Most of those solutions were bespoke, and tended to be rather brittle, and otherwise limited in their capacity. If one system changed on either side of the pipe, repairs had to be done.

A much more dynamic solution is now evolving in the cloud, namely the use of application program interfaces, or APIs. The design and use of APIs now dominates the discipline of data integration online, and with good reason. APIs tend to be significantly more flexible than their EAI counterparts, and as their power increases, companies can effectively weave together richer and more useful pictures of whatever matters most.

Not Your Father's Data Warehouse

There is always a correlation between software markets and their corresponding client and prospect appetites. There are hype cycles, to be sure, when a thousand flowers will bloom, with only a fraction surviving; but there's never a complete disconnect between what's being built, and what's needed. There is often a lag, however, simply because software takes time to code, and because market leaders can be difficult to dethrone.

This explains why the cloud data warehouse market took so long to materialize. The major first-mover was Amazon with its RedShift product, which market research company HG Data estimates has more than 2,500 clients. Upstart competitor Snowflake recently touted their 400th client, which represents a remarkably fast rise for a company that was founded in 2012.

And although there have been a range of advances in big data analytics since 2010, there's no reason to believe that the data warehouse market has anywhere to go but up. In fact, research firm Market Analysis estimates that by 2022, the overall market will crest \$20 billion. That would represent greater than 8% growth, year over year, from today's market.

Market Analysis notes that one key trend which will drive additional adoption is the advent of cloud-based data warehousing. Snowflake's early success is indicative of this trend, as well as the rapid rise of RedShift. And the use cases go far beyond marketing:

- Major healthcare organizations are increasingly looking to online patient portals to streamline processes and improve the administration of care;
- Financial services organizations are also leaning heavily on web-based interfaces to interact with clients and prospects, and provision services;
- Transportation companies are investing heavily in Internet of Things (IoT) technologies to better track, manage and maintain their far-flung assets;
- Oil and gas concerns are also investing in IoT technologies to optimize the

operation of expensive machinery in remote locations;

- Educational organizations are increasingly moving to cloud-based solutions for providing the spectrum of learning modules, from pre-kindergarten all the way through to professional training;
- Governments are recognizing the remarkable power of web-based systems for communicating with constituents, tracking programs, and delivering critical services;
- The list goes on and on...

All of this speaks to the tremendous value of data warehouse automation. Based upon the expanding use cases for data infrastructure, and the cost reductions afforded by cloud-based computation, it's no wonder that WhereScape and Snowflake have partnered to expedite the design, development, deployment and operation of cloud-based data warehouses.

Due to market trends, coupled with the vision and demonstrated capabilities of these two companies to execute, it's fair to say that this collaboration will prove to be effective. This is especially true because of how WhereScape has integrated its new automation solution into the Snowflake platform. Specifically, WhereScape automation for Snowflake automates the development and operations workflows based on native Snowflake functions, wizards and best practices.

In doing so, WhereScape provides Snowflake customers with the means to fast-track the design, development and deployment of their cloud data warehouse initiatives. This was showcased in that Deep Dive webcast, which demonstrated how a process that once took at least days, can now be accomplished inside of 15 minutes. By collapsing cycle times to that degree, these two companies have opened the door to a new phase in the evolution of data warehousing, one that fast-tracks delivery, limits project risk, boosts team productivity, and supports both development and operations for their clients.

Peter Nilsson, Chief Technical Officer of Aptus Health, a Snowflake customer said it best, when he went on record with this glowing review of the value WhereScape automation for Snowflake will provide for their organization:

"Our health and life sciences clients rely on us to execute flawless marketing campaigns and deliver customer insights from across our many digital platforms. To that end, we're leveraging WhereScape's ability to ease our design and implementation effort as we develop our fully integrated warehouse on Snowflake. WhereScape's tools will save us thousands of engineering hours, promote implementation accuracy, and accelerate our time to delivery. And since WhereScape natively supports Snowflake, we can use the generated code without having to tailor it to our target environment -- letting us focus on generating the insights that will drive our business forward."