EB-7244 March 2017

Getting Your Analytics Platform Up To Speed

Claudia Imhoff, Ph.D., Founder of Intelligent Solutions, Inc.

Boulder BI Brain Trust (BBBT)

Sachin Grover, Agile Coach, Teradata Corporation

Sponsored by

TERADATA.

Table of Contents

The Evolution of Analytics

By Claudia Imhoff

Analytics have become every CEO's darling these days. This is evidenced by the fact that their creation and usage rank as number 1 or 2 for importance in most surveys of the C-suite. The demand for BI and analytics—such as dashboards, scorecards, key performance indicators, and customer analytics—is constantly growing, as is the need for new, diverse, and complex sources of data to support them. The challenge is that it's increasingly difficult for data warehouse implementers to keep up with these demands, especially considering the always-changing requirements of a growing, thriving business.

The term, analytics, is not generic; there are different types of analytics needed in all enterprises. Figure 1 shows the typical evolution of these types of analytics in a modern, data-driven organization. Most data warehouse implementations start with the relatively simple *descriptive analytics* like report generation, multi-dimensional, and comparative analytics. These answer the question, "What happened or is happening?" These are historical, backward-looking results that give business users a good idea of where they have been.

Unfortunately, many companies often stop there. They may not realize or see the need for logical questions that should follow after the "what" question. The first logical follow-on is, "Why is this happening?" or *diagnostic analytics*. The ability to determine the causes and effects on business performance is mandatory if the organization wants to truly understand its marketplace, operational performance, and competitive environment.

The next question is, "Will this continue?" or *predictive analytics*. For the most part, we are all creatures of habit so using history to predict the future is a pretty safe bet. Unless something happens to change the trajectory of a trend (i.e., internal or external events or actions), the predictive analyses will give business leaders a good idea of where their company will be in 6 months to a year.

And the ultimate question is, "What should I do to either maintain the trajectory predicted or change it to a more beneficial one?" or *prescriptive analytics.* These are the analytics that result in *actions.* They consist of "what if" analyses, optimizations, and guided decision making used throughout the business.

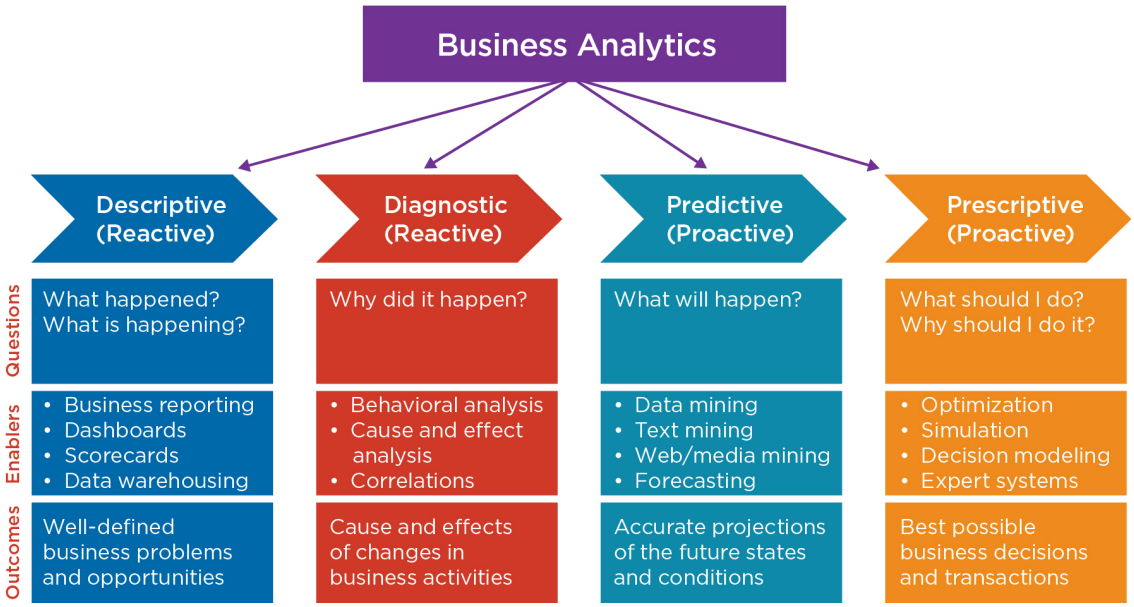

Figure 1: The Evolution of Analytics

These more advanced analytics provide the most business value to the enterprise, but are also the source of much of the complexity in terms of data requirements and technological support. They require rapid changes to existing analytical capabilities caused by evolving business situations, and demand much shorter development timeframes for new analytics. We consistently hear from business users that it takes too long to develop the analytics and supporting infrastructure, it costs too much, and changing anything after deployment is just too hard.

Fortunately, vendors and implementers alike heard these concerns and responded. Today, we are able to overcome these challenges through a combination of technological advances:

- A new analytics topology, called the Extended Data Warehouse architecture
- An updated methodology that supports agile development
- The creation of a stable analytics environment using data warehouse automation

Let's look at these in more detail.

Agile Analytics Creation

Business requirements must drive the creation of analytics; they should never be created in a vacuum. To ensure this, there must be a significant amount of collaboration between IT resources and business users to ensure these analytics are both data–driven AND business-driven. The following technological advances greatly increase the likelihood of success in responding quickly and accurately to business needs:

Extended Data Warehouse Architecture (XDW)

Figure 2 is a depiction of an updated logical architecture for analytics. The XDW shows that the traditional Enterprise Data Warehouse (EDW) is still a mandatory part of any analytics environment, especially for production analytics. However, it is not the only source of analytics. For exploratory or experimental analytics (data science), an Investigative Computing Platform must be created. And for real time operational analyses, embedded BI services can be called, and/or a real-time analytics engine may be invoked (streaming analytics).

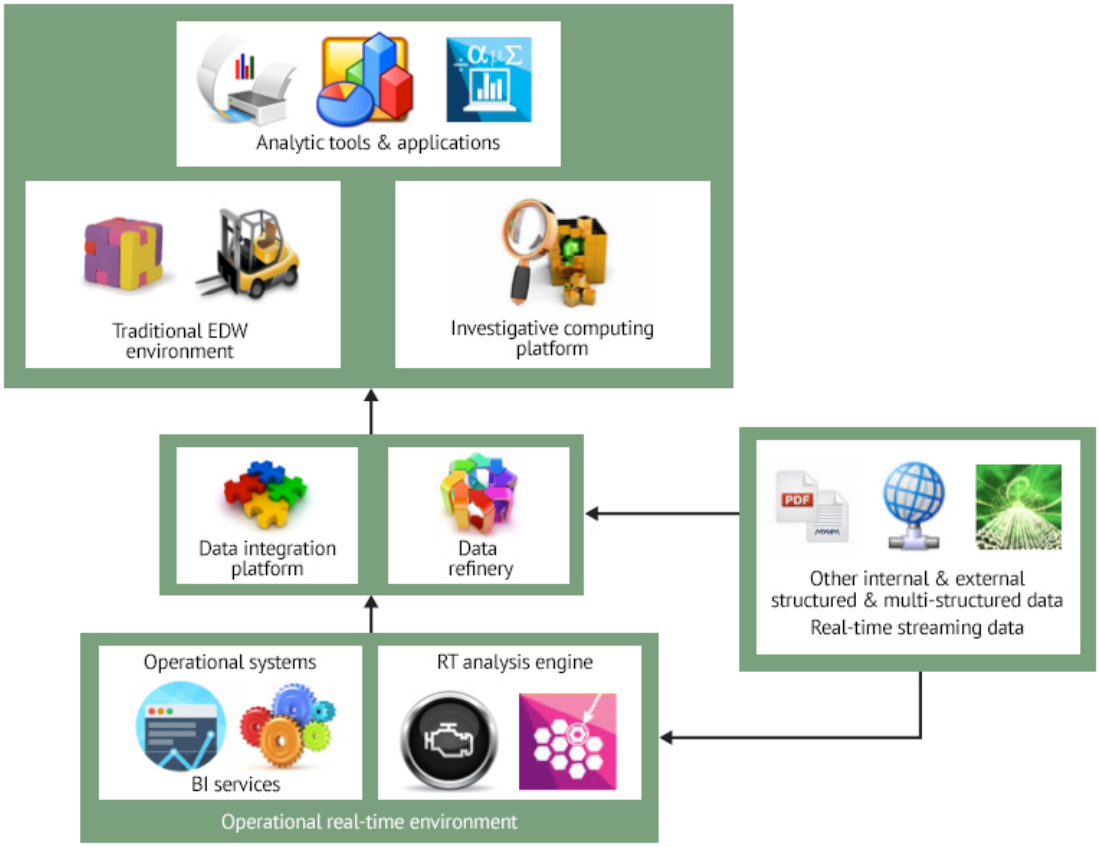

Figure 2: The Extended Data Warehouse Architecture

This should be the roadmap used for all current and future analytics projects to ensure the creation of a proper environment for each of the different analytics needed. The benefits of its usage include a reduction in redundancy, chaos and wasted efforts, and the support for agility and responsiveness to changes.

An Agile Methodology

Business requirements change along with ever-changing business situations. Therefore, analytics environments must remain flexible and responsive to accommodate rapidly changing needs. While there will be short-term business needs requiring fast adaptation of plans, the implementation team must ensure long-term business flexibility and efficiency. This means that IT and business users must collaborate to meet all these expectations—and provide quick responses. Agile development methodology suits these environments to a tee.

Creation of a Stable Analytics Environment

Traditional EDW methodologies require multiple technologies at different steps. These include at least four:

1. Logical modeling
2. Extract, Transform, and Load (ETL)
3. Physical data modeling
4. Schema development technologies

In addition, specialized skill sets are needed with each tool.

The bottom line for traditional methodologies is that the "disintegration" of these sets of tools causes higher costs, significantly more complexity, and longer development times. Agile EDW methodology uses data warehouse automation technology that supports the Agile Methodology with speeds that meet the needs of an iterative process of feedback and resolution. These technologies also reduce the number of tools to one, removing the need for multiple specialized skills. By reducing the number of tools and skills needed, it also reduces costs, time, and complexity. Data warehouse automation delivers reusable components very quickly, consolidates the technical metadata into a single repository, and makes enhancements and changes very easy to perform—improving flexibility, sustainability, and maintainability.

Connect Data-driven Analytics to Business Outcomes

While business requirements must drive the creation of all analytic results, in turn the results must drive business outcomes. IT professionals skilled in understanding data, not just coding, should support the creativity of business users—rather than cause frustration for them. The combined efforts of these two groups can then be used to connect business strategy and ROI to analytics usage. This analytical environment can be used to test business strategies more accurately, and measure the effectiveness of new strategies.

For example, if the business strategy is to reduce customer churn and increase customer profitability, the analytics team must produce customer analytics for usage by all customer-facing entities in the enterprise. Some of the analytics would include customer lifetime value, segmentation identifier, behavioral analyses, next-best product offers, and predicted response rates to campaigns.

The mantra should be: "If you can't measure it, you can't manage it". The enterprise must be able to quickly identify if something is not working, enabling the business to respond and change rapidly.

For this to work, the following technological underpinnings are needed:

1. **Productivity tools for faster time to deliver.** Data warehouse automation technology will be needed to replace the bits and pieces of technology you currently use. Here are a few suggestions for selecting this technology (a short list):

    a. Data warehouse automation technology should programmatically generate target schemas, and perform profiling and analysis of source data before ETL or ELT process

    b. It should have broad support for most database engines, and be able to create multiple data models (e.g., relational, star schema, data vault)

    c. It should produce technical metadata and other supporting documents with ease

2. **Trusted, high performance, and flexible forms of data storage** Depending on the type of analytics being created, a variety of technologies will be used. In selecting the data storage technologies, look for these functions:

    a. For the EDW, bulletproof relational technologies have worked well for a long time; however, new technologies can also enhance its performance.

    b. For the Investigative Computing Platform, Hadoop, in-memory computing, columnar storage, data compression, or appliances have all shown to work well for complex queries and "unusual" sets of data.

    c. Operational Analytics will require that services be called from the EDW (a service-oriented architecture is recommended).

    d. Finally, for the streaming analytics engines, you should look into Apache Spark, event stream processing, or complex event processing technologies. For other specialized analytical needs, you may also want to look into graph databases.

3. **Appropriate analytical techniques and technologies** Today, there is no single technology that supports all forms of analytical needs. Let's break up the technologies into the four analytical forms mentioned in Figure 1.

    a. For descriptive analytics, you can use reporting & multi-dimensional technologies, dashboards, and/or scorecards—all usually produced from the EDW. Diagnostic analytics start out using reporting & multi-dimensional technology.

    b. For more sophisticated analyses, you'll need behavioral analysis, cause and effect capability, and the ability to determine correlations between objects.

    c. For predictive analytics, you will need software that can perform forward-looking analytics, statistical and machine learning, pattern recognition, text and data mining, and forecasting.

    d. The last type of analytics support is for prescriptive analytics. The technology needed is for what-if analysis, probability and uncertainty modeling, guided decision-making, and expert systems. Also recommended is data visualization and storytelling support in your analytical products.

Summary

Today's enterprises must have sophisticated, advanced analytics to compete and flourish in a complex, changing marketplace. The underpinnings of this environment must be based on a solid, proven architecture that serves as a roadmap for all projects.

With a thoughtful selection of appropriate technologies, an eye on strategic initiatives, and a successful collaboration between the business and IT, an enterprise can expect to create an agile, reliable, and highly performing analytics environment. The rapid production of new analytics, flexibility to change analytics as situations unfold, and the incorporation of a myriad of source data can be accomplished in a minimum amount of time—with much less cost—and the flexibility to handle any changes.

From the Sponsors: Mind the Gap

Let's look at how Teradata can enable agility by starting with a framework that can be leveraged at the individual team level, and be scaled at the portfolio level where multiple teams are at play. To support speed and agility, Teradata has created an approach that is based on the principles and practices of Lean, Lean Startup, Agile development, DevOps, Systems thinking, and Kanban. We applied these principles and practices to analytics. It is important to understand that while Agile, Lean, and DevOps principles are always a sound foundation, they are applied differently for analytics from software development.

In the Lean startup method we build Minimum Viable Product (MVP) and test the market to see what kind of response we receive. If we get good response, then we put it into production. The engineering sprint label shown in Figure 3 is where we build a prototype or MVP.

To remove uncertainty, there is a strong need to create that prototype, or MVP, in the shortest time possible. Many projects fail because there is so much uncertainty in the projects—due to inaccurate requirements gathered at the beginning of a project. In traditional methodology, if the requirements are wrong up front they will snowball into wrong design, development, test, and so on. Wrong requirements generally occur because business users don't know what they want. There is a relevant statement from the book LEAN ENTERPRISE:

"Users are often unable to articulate exactly what they need, yet they often seem insistent about what they don't want … once they see it."

Business users can't tell you what they want until they see a prototype. The Teradata framework addresses this issue directly. It uses engineering sprints by applying the principles and practices of Scrum.

During a continuous series of sprints we create a prototype for users, performing a demo for them at the end of each sprint. While the prototype is being created, we are validating what we are building every day of the sprint, directly with the business users. After 4 weeks (or less) we demo the working solution in a way that business users can touch and play with the solution.

For example, suppose we were building a Business Intelligence (BI) solution. We would decompose analytic use cases and land the data in advance of the sprint. The dimensional models would then be built as prototypes directly from the landed data, concurrently building the BI solution on top. The main purpose is to get the data into the hands of business users.

The "middle layers" are bypassed during the engineering sprint, using a process called data-driven design to get into the data immediately. The prototype produces requirements that are far better than the big design up front because the business users are shown a working solution using production data and their tools. Once the prototypes are validated by business users at the sprint review, the engineering team immediately moves to new analytic use cases in the next sprint. The working solution package from the first sprint is queued up for a delivery team to build using Kanban.

Kanban is a method that focuses on the end-to-end flow. Unit testing, SIT, UAT, and deployment are part of an integrated end-to-end cycle, unlike separate phases in many implementations. If, within your environment, you have test sprints—or just deployment sprints—this is a sign that agile principles are not being fully applied.

The output of engineering sprints are working solutions, while the output of delivery cycles are product increments that are ready to be deployed to production, depending on frequency of releases. DevOps practices are used to automate the delivery pipeline that is within the build cycle.

Using this approach we can provide early delivery (within 4 weeks) and continuous delivery (because the work is divided into small chunks that provides business value, and is delivered into production through a predictable cadence).

It is important to note that automation tools provide a very big benefit, especially during the engineering sprints. Tools that can help design and develop seamlessly—and do it faster—play a big role in practicing Agile.

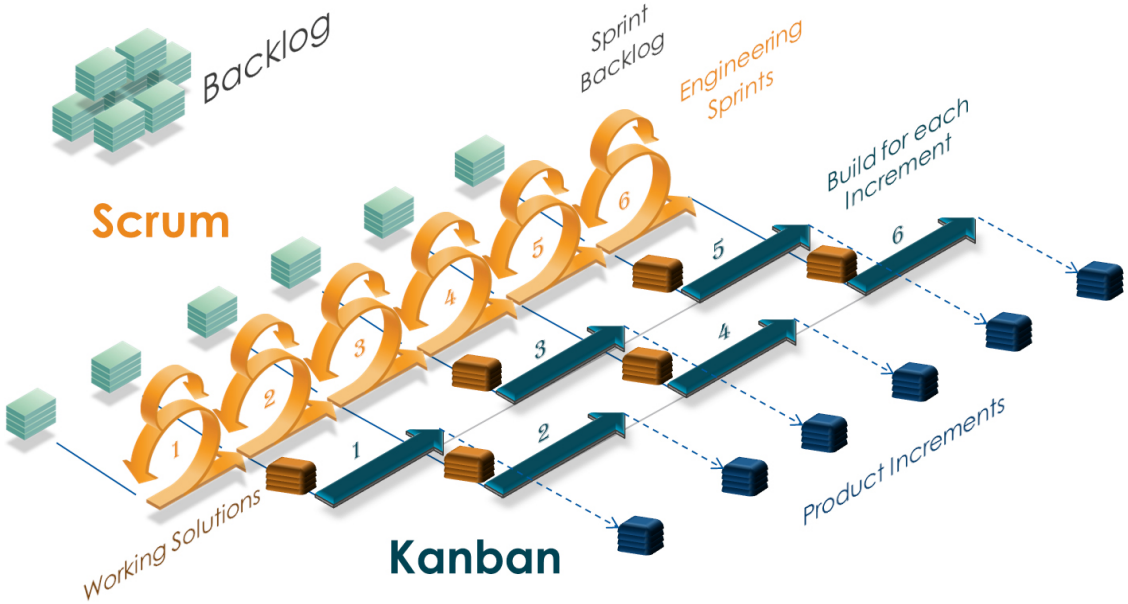

Figure 3. MVP engineering sprint label

TESCO Customer Case Study

Tesco, the second largest retailer in the world, used WhereScape's data warehouse automation software to reengineer its Teradata enterprise data warehouse—creating an integrated reporting application for global promotions in just six months, smashing the previous 3-year development forecast.

The Challenge

In response to new business challenges and opportunities, Tesco needed a global view of its data to coordinate management information. However, a review of their management information systems discovered more than 100 sandboxes where data was dumped, resulting in 20 terabytes of duplicate data.

The Solution

In place of a traditional waterfall development approach to renovate its data warehouse, Tesco used an agile development approach—along with data warehouse automation from WhereScape—while consolidating data on a Teradata EDW for a single source of the truth. Using WhereScape automation software, Tesco put a working solution in front of business users every four weeks to deliver business value continually throughout the development process.

The Benefit

WhereScape enabled quick turnarounds for trying and testing alternatives. The Tesco and Teradata Professional Service teams used WhereScape software to rapidly model a concept, generate DDL code, and test the physical structure to determine its workability, while documenting the entire process.

03.17 EB-7244