



# OUT OF THE SANDBOX, INTO THE DATA LAB

SELF-SERVICE BUSINESS INTELLIGENCE  
BECOMES A REALITY

TERADATA®

## EXECUTIVE SUMMARY

Real innovation—the kind that turns heads and attracts the attention of customers and investors—is about more than just having the best products or services. It requires knowledge of everything from market trends to individual customer behavior, as well as the ability to act on that knowledge faster and better than the competition.

One way to get that knowledge is with an external analytics sandbox, also known as a data mart—a separate environment that gives analysts the freedom to safely explore and experiment with data without impacting production environments.

While sandboxes within data marts may be perceived to be lower cost and faster to implement, they do have limitations. Business analysts must first provision the sandbox, which might require the acquisition of hardware or software. Then, every time they need data from the data warehouse, they must make a formal request to IT staff. This taxes IT resources and bogs down the process, inhibiting analytical agility. It's also costly, risky, and resource intensive.

A better solution is a data lab, which is a self-service analytics sandbox that exists within a production data warehouse. Business users enjoy quick and easy access to required data through automated processes and user-friendly tools. And IT staff is assured of proper governance without the strain on resources created by sandboxes and data marts.

The result is that users are more likely to explore and experiment because insights are delivered faster and with greater accuracy. In short, data labs encourage innovation and business agility, thanks to built-in controls and tools that make the process faster, safer, and more accessible.

## BARRIERS TO INNOVATION

The pressure is on for companies to move faster, to be more agile, to anticipate customer requirements, and to react quickly—sometimes before the customers themselves know what they need.

However, businesses can only move as fast as they can innovate. If it takes twice as long as your competitor to envision, develop, and roll out a new product or service, eventually you'll be left behind. Work harder, faster, and smarter, and you'll take the lead.

Unfortunately, today's businesses face several obstacles that can slow down innovation and insight. One such obstacle is massive data volume. Everything from transaction records to email to social media posts is valuable, and it all contains insights into how the business is doing, what it needs to do next, and how customers are reacting to those efforts. The right kind of analytics tools can help manage huge amounts of data, but to really take advantage of such tools, organizations need a way to quickly manipulate, test, and retest the data in combination with existing enterprise data.

Furthermore, businesses are dealing with new data types—such as Web log, geospatial, and clickstream data—that carry enormous potential for insight, so much so that it can be very difficult to know where to begin. What questions should you ask? What don't you know that you need to know to take your business to the next level? What processes are required to catalog, mine, refine, and combine that data with existing enterprise data?

There's also a gap between business and IT departments. Generally, business analysts don't have the database skills to access the data themselves. Even if they did, they often don't have permission. From an IT standpoint, there must be clear processes in place to control and protect the data. Restricting access is one logical way to do this, even if it places a greater demand on IT personnel.

## DATA MARTS AND SANDBOXES

Data marts used to be an answer to some of these challenges. A data mart is a subset of the data warehouse that houses data related to a specific subject, department, or project. Approved users can access this data to run analyses and make business decisions.

The problem? Data marts create silos within the business. Users must make decisions without the benefit of the 360-degree view provided by the Integrated Data Warehouse (IDW). Data marts also consume system resources and can be costly to maintain.

A better, more cost-effective alternative is an in-database analytics sandbox. A sandbox is an analytics area that is independently managed within the data warehouse. IT staff can load it with test data and allow approved users to experiment on that data with minimal risk to the data warehouse. Analysts can explore enterprise data, combine it with untested data from external data sources, and experiment with the data to explore new business ideas before putting them into production.

Despite the obvious benefits, this too is an imperfect solution with these drawbacks:

- Slow—It can take weeks or months to submit a request for a sandbox, gain approval from IT personnel, and wait for the space to be partitioned and loaded.
- Dependent on IT staff—Business users must have IT permission and assistance to access other data within the data warehouse, load new or different datasets, or expand the size or time frame of the sandbox. This manual process can strain IT resources and delay analysis.
- Resource intensive—System resources are at risk with sandboxes. Because it takes so long to set up a new sandbox, users tend to add on to existing sandboxes without removing data from old projects. As they grow unchecked, sandboxes can quickly eat up space and processing capacity, which increases demand on the rest of the system.
- Ungoverned—Most sandboxes are not subjected to the same level of governance as other business intelligence (BI) systems, making it relatively easy for IT personnel to lose control over who has access to the sandbox, how the data is used, and how big the sandbox becomes. If left unchecked, production applications can gain access to the sandbox and utilize untested, compromised data.
- Dated—The data in an analytics sandbox is merely a snapshot, frozen in time. Unless users regularly update the data, they'll end up testing against data that may no longer be accurate or relevant.

Still, business analysts have a job to do, so they make do with what's available. The question is, at what cost? Slow, complicated processes can compromise decision making as analysis proceeds on untested, uncleaned, and out-of-date data. Users might turn to other solutions, such as data marts, simply to avoid getting bogged down in the sandbox process. Worse yet, they may avoid experimentation altogether because it's just too difficult to execute.

But the real impact to the business comes in terms of missed opportunities. If the analytics environment doesn't foster creativity, experimentation, and innovation, the business won't see the fruits of those labors, such as improved profit margins, increased response rates, higher customer satisfaction, lower churn, and so on.

## DATA LABS: A NEW APPROACH

Thankfully, there is a solution that combines the benefits of an external or internal analytical sandbox with the resources of an IDW. A data lab is a self-service workspace built within the data warehouse. Like a sandbox, a data lab allows for experimentation, exploratory data analysis, and rapid prototyping in a safe environment.

But unlike a sandbox, a data lab is created by software that enforces governance, in terms of expiration dates and size limitations. That same software assigns user roles that govern data access and permissions. These safeguards protect against unwanted access and uncontrolled growth. In addition, automation enables self-service access to required data, freeing IT staff to focus on other priorities and allowing business analysts to explore the data at will.

### REAL RESULTS: LESS COST, GREATER PERFORMANCE

Data labs are ideal solutions for businesses of all sizes and in every industry. Consider these examples:

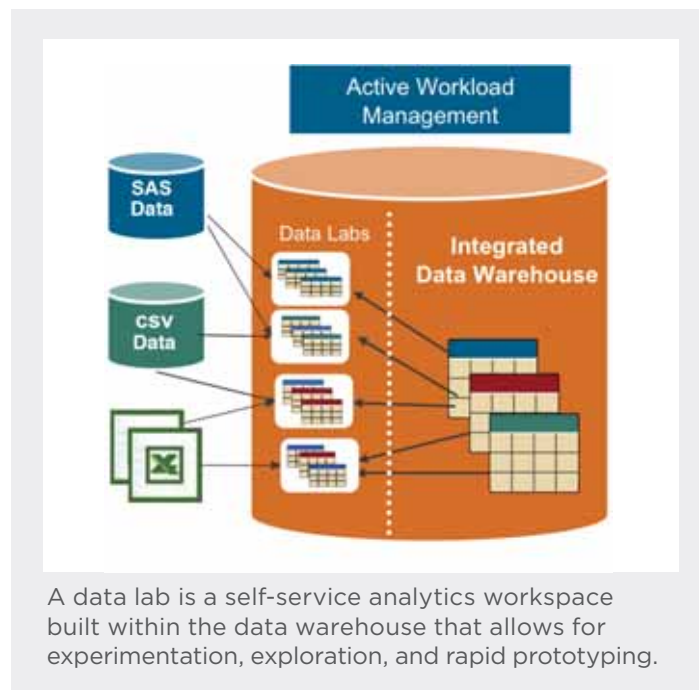
- A leading online retailer replaced existing \$500,000 analytics appliances with 90-day analytics data labs that were provided free to all organizations within the company. The data labs use excess capacity within the existing data warehouse to handle the increased load. Users can access all of the organization's data, not just what's in an individual data lab. As a result, this organization has improved analytics, reduced costs, maximized data warehouse resource utilization to 99 percent, and avoided silos.
- A major financial services provider dedicated up to 25 percent of its IDW to data labs supporting agile development, prototyping, temporary data, and personal development environments. In-database processing improved query performance while minimizing data movement and maximizing IT resources.
- A specialty retailer turned to data labs and improved its promotional filter performance by 98 percent. Target email lists that previously took three days now generate in one hour. In addition, the organization has tripled its analytics output.

Organizations that are considering moving toward a data lab environment should look for a solution that offers

- ~ Self-service access—Business analysts should be able to set up data labs and access enterprise data as needed. Critical areas for self-service include setting up the initial data lab, requesting more space or more time, and adding or deleting users for sharing and collaboration.
- ~ Governance—Data labs should be governed according to the same rules that govern the data warehouse. IT personnel should be able to limit the size of a data lab; who can gain access to a data lab; and how long a data lab will be active. Other governance concerns are data security, data quality, and data management.
- ~ Access to external data—Data within the data lab should be easily accessible so users can load data and work with a wide range of file types, including Apache™ Hadoop®, geospatial, XML, and temporal data. This gives users greater flexibility in the types of analysis they can do without IT intervention.
- ~ Access to tools—After a data lab is populated with data, users should be able to access a variety of visualization, analytics, and BI tools for exploration and deep analysis.
- ~ Workload management—A data lab environment should easily integrate into the existing workload management capabilities that are available within the data warehouse. This ensures that system resources are optimized according to established business and IT priorities.
- ~ Integration with the data warehouse—A data lab should feel built-in, that is, it should understand and follow existing data warehouse parameters, use existing database objects, and adhere to existing regulations.

## TERADATA DATA LAB

To address the need for faster, safer, and more flexible self-service BI and analytics capabilities, Teradata has introduced Teradata® Data Lab. It provides IT professionals and business analysts the on-demand database space and system resources required for short-term projects, such as ad-hoc analysis, data mining, proof-of-concept, application evaluation, and quality assurance testing. It also can enable the quick assimilation of untested and uncleaned data into a separate non-production portion of the data warehouse, where it can be analyzed along with insight-rich enterprise data.



Teradata Data Lab consists of two portlets. The Data Lab portlet is intended for both power users and business users. It is the interface through which end users can self-provision a data lab, add users to a data lab, manage a data lab's day-to-day activities, make requests to IT personnel, and so on. In addition, this portlet allows users to easily browse through the data lab hierarchy to view and manage data labs and tables.

The Lab Group Setup portlet is designed for database administrators (DBAs). It allows IT staff to create and manage data lab groups, including establishing the data lab hierarchy; who has access to the data labs; how much space can be allocated to any single data lab within the group; which tasks are automated; which tasks require IT approval; and any other criteria that govern data labs and lab groups. This portlet also allows the DBA to set automated email notifications for low data space concerns, lab expirations, as well as all request and approval notifications.

Both portlets are equipped with wizards that guide users through the process of setting up and administering a data lab. Default settings ensure that each data lab inherits the security, workload management, and other controls established within the rest of the database or agreed upon by IT staff and business users. The user-friendly interface and automated features make it quick and easy to create and govern data labs.

## CUSTOMER SNAPSHOT: 2,500 PERCENT REDUCTION IN CYCLE TIMES

A global healthcare company turned to Teradata Data Lab to address a number of pain points. Data understanding and preparation consumed a great deal of time, core processes were difficult to scale, and the current environment limited collaboration and data reuse. In addition, analytics processes were largely manual and required a great deal of rework whenever parameters changed. The company's analytics technology platform was also outdated.

This organization chose Teradata Data Lab because it allows analysts to have direct access to corporate data, thus empowering them to test hypotheses as needed. The process of setting up a workspace is fast, which makes it possible for analysts to shift their focus and priorities more quickly than in a traditional IT environment. Finally, the solution supports the rapid exploration of hypotheses with minimal IT support, making it an efficient alternative to other solutions.

The short-term results have been dramatic. Under the new system, return on investment (ROI) production cycle times improved by more than 2,500 percent—from 130-plus hours to less than five hours per business unit. The business analyst community has been much more productive with both tactical and strategic work. There has also been a reduction in data redundancy and less dependency on IT staff for low value-add work.

## TERADATA STUDIO

Teradata Studio™ supports the various functions within Teradata Data Lab. Its point-and-click interface allows users to easily load files, such as Microsoft® Excel and .csv files, as tables in the data lab. Teradata Studio™ also makes it possible to copy select columns from any Teradata table if the user has access to the table. This drag-and-drop capability streamlines the process of adding data to a data lab.

Smart Loader for Hadoop is a Teradata Studio™ feature that allows users to easily drag and drop any files from Hadoop to the data lab (or vice versa). There is no need for expertise in MapReduce or knowledge of the Hadoop distributed file system (HDFS). The tool automatically determines where the files reside.

Teradata Studio™ also has descriptive statistics through the Teradata Profiler plug-in. With these profiling capabilities, analysts can visualize the statistical characteristics of the data through values analysis, statistical analysis, frequency, and histograms.

Together, Teradata Data Lab and Teradata Studio™ automate and support the self-service BI process.

## CONCLUSION

Issues related to data access, resource allocation, and governance can impede analytics, making it difficult for companies to get the insights they need to stay ahead of the competition. Data marts and analytics sandboxes address many of these issues, but they can be costly to maintain and slow to implement.

Data labs bypass these barriers while using the power and resources of the data warehouse. Automation and governance simplify the analytics process and protect valuable data assets. This type of environment enables the kind of exploration and experimentation that drives innovation because it allows users to quickly and easily prove the success of their ideas, or to quickly fail and move on.

To learn more about the new Teradata Data Lab solution, visit Teradata Developer Exchange, where you can view Teradata Data Lab Live, a series of demos that illustrate the use of the Teradata Data Lab portlets:

<http://developer.teradata.com/viewpoint/articles/teradata-data-lab-live>.



10000 Innovation Drive Dayton, OH 45342 [teradata.com](http://teradata.com)

**TERADATA**®

Teradata Studio is a trademark, and Teradata and the Teradata logo are registered trademarks of Teradata Corporation and/or its affiliates in the U.S. and worldwide. Microsoft is a registered trademark of Microsoft Corporation. Apache is a trademark, and Hadoop is a registered trademark of Apache Software Foundation. Teradata continually improves products as new technologies and components become available. Teradata, therefore, reserves the right to change specifications without prior notice. All features, functions, and operations described herein may not be marketed in all parts of the world. Consult your Teradata representative or [Teradata.com](http://Teradata.com) for more information.