

# Twelve Ways TIBCO Data Virtualization Turbocharges Data Science and Analytics

## **There is opportunity in data**

---

Data fuels digital transformation. The opportunity is massive as reflected by IDC Research, which estimates that by 2023 over 103 ZB (1ZB = 1 trillion gigabytes) of new data will be created.<sup>1</sup>

However, with opportunity comes challenge. IDC also reports that data scientists and business analysts spend as much as 91 percent of their working week on data-related activities such as searching, preparation, and governing. This means less than 10 percent goes toward the data science and analytics work that derives the insights that in turn inform digital transformation efforts.

## **The growing data science and analytics data bottleneck**

---

Data science and analytics run against data—providing that data is more difficult than ever.

Do any of these sound familiar?

- Your company's demand for analytic data is exploding.
- Your analytic user types and capabilities are more diverse.
- Your analytic tools are more varied.
- Your data is more widely distributed across traditional, big data, Internet of Things, and cloud repositories.

---

<sup>1</sup> IDC Worldwide Global DataSphere Forecast 2019-2023

- Your traditional data integration via warehousing and ETL is not keeping pace.
- As a result, your analytics data bottleneck keeps growing.

The TIBCO Data Virtualization platform helps you turbocharge your data science and analytic applications so you can deliver more insights, drive better business outcomes, and outperform your competition, far faster than ever before.

This paper outlines data science and analytic data requirements and describes how TIBCO Data Virtualization users can successfully address these challenges. A customer example illustrates how one forward-thinking company is using this powerful combination to its advantage.

## What data science and analytics users want

---

Data science and analytics solutions enhance your competitive edge by enabling you to get meaningful intelligence and insights from data and predict what is going to happen next. By predicting and acting on the discovered opportunities, you can innovate and optimize your business, deliver amazing customer experiences, and win in your markets.

However, if you ask your data scientists, business analysts, BI application developers, and others about their data requirements, you may find them frustrated by all the time they spend chasing data, rather than analyzing it—which is after all why you hired them in the first place.

When asked to describe their data nirvana, their needs are clear. They include:

- **Any and All Data** – Any dataset I might need, from anywhere and everywhere.
- **Anytime Data** – Datasets should be up-to-the-minute and available whenever I need them.
- **Accurate Data** – Datasets should be trustworthy, with lineage back to their source.
- **Easy Data** – Datasets should be easy to find, understand, access, combine, and use.
- **Agile Data** – Whenever a new, unique dataset is required, help me quickly meet this need.

## TIBCO Data Virtualization to the rescue

---

The TIBCO Data Virtualization platform is purpose-built to help your data engineering teams meet these requirements. With it, your data scientists, business analysts, BI developers, and other users, can take advantage of analytic datasets built, tested, secured, and governed by your data engineering teams. Delivered at a fraction of ETL time, cost, and rigidity, your data engineers can use it to overcome the analytics data bottleneck that is constraining your data science and analytics success.

Here are 12 ways that TIBCO's best-in-class data virtualization capabilities help you turbocharge your data science and analytics.

- 1 Data Abstraction Makes Data Easier to Understand and Use** – The TIBCO Data Virtualization platform includes powerful data abstraction capabilities that simplify your complex data, transforming it from native structure and syntax into common business semantics that are easier for your data scientists and analysts to understand and analyze. Focused on business problem solving, your data scientists and analysts will particularly appreciate this business-friendly view of your data.
- 2 Business Directory Makes Data Easier to Find** – Your data scientists and analysts can easily search, categorize, and consume IT-curated datasets using TIBCO Data Virtualization Business Directory. By encouraging analytic dataset sharing and reuse, you can accelerate projects and reduce data engineering workloads. Further, by adding the TIBCO Cloud Metadata platform, you can provide a complete, one-stop-shop data catalog solution spanning virtualized and other data assets.
- 3 Data Federation Provides the Complete Picture** – The TIBCO Data Virtualization federated query engine virtually integrates your analytic datasets. This provides your data scientists and analysts with the complete picture without the time, cost, and rigidity of physical data consolidation. Further your data scientist will value the opportunity to combine raw data in a data lake with IT-curated data from across your organization.
- 4 Standard Data Delivery Interfaces Support Heterogeneous Analytics Solutions** – The TIBCO Data Virtualization platform supports all the most popular data science and analytic tools and interfaces with ADO.Net, Hadoop, JDBC, ODBC, REST, SOAP, and more. Sharing datasets across all your data science and analytic tools helps maximize your return on TIBCO Data Virtualization investments.
- 5 Adapters Broaden Data Access** – Nearly 300 TIBCO Data Virtualization adapters provide a wide range of data source connectivity for databases, files, big data, cloud sources, packaged applications, IoT devices, and more. Beyond schema-to-schema only connectivity, TIBCO Data Virtualization adapters work hand in hand with data source optimizers to ensure accurate queries and the highest possible performance. With more data than ever to work with, data scientists and analysts will uncover new insights.
- 6 Caching Supports Analytic Sandboxing** – When your data scientists and analysts require specialized analytics sandboxes and data lakes, TIBCO Data Virtualization caching can load needed datasets into over a dozen distinct analytic databases including Greenplum, HSQLDB, DB2, SQL Server, Netezza, Oracle, SAP HANA, Sybase, Teradata, Vertica, and more.

- 7 Integrated Security Controls Data Science and Analytic Data** – The TIBCO Data Virtualization platform provides myriad fine-grained security capabilities including authentication, authorization, encryption, and masking. You can also leverage your existing security investments (LDAP, Active Directory, and Kerberos); standards (SAM, NTLM, TLS, and WSS); and policies. Your data scientists and analysts will access sanctioned data only, no more, no less.
- 8 Quality Ensures Trusted Data** – The TIBCO Data Virtualization platform provides a range of profiling, standardization, conformance, enrichment, augmentation, and validation capabilities to give your data scientists and analysts data they can trust. And when master data is an important part of your analysis, you can use the TIBCO EBX platform to further improve that trust.
- 9 Data Governance Maximizes Control** – TIBCO Data Virtualization lineage, where-used, logging, and more, provide complete visibility, traceability, and control over your analytic datasets. Your data scientists and analysts working on mission-critical financial reporting, compliance, and other projects, will especially value this extra level of governance.
- 10 Query Optimization Ensures Timely Business Insight** – TIBCO Data Virtualization classic and MPP engines, advanced query optimizers, and algorithms—the fastest in the industry—provide up-to-the-minute datasets your data scientists and analysts require. And with the TIBCO Data Virtualization platform, your team no longer needs to wait for the nightly data warehouse refresh.
- 11 Studio Accelerates New Dataset Development and Analytics Time-to-Solution** – Your data engineers use TIBCO Data Virtualization Studio to quickly model new datasets, build transformations, optimize queries, manage resources, and more. Easy to learn and use, the Studio graphical modeling environment, along with powerful code editors, speed new dataset development and thus your time-to-insight. And if these datasets require further preparation, your data scientists and analysts can use their own data preparation tools to quickly make these value-add tweaks.
- 12 Scalability Supports Multiple Lines of Business, Hundreds of Projects, and Thousands of Analytics Users** – The TIBCO Data Virtualization platform is enterprise-grade and proven in some of the world's most complex, high-volume data science and analytics environments. It can support nearly all your data science and analytic tool workloads from project to enterprise.

## Better together in the energy industry

---

Data engineering teams at hundreds of companies use the TIBCO Data Virtualization platform to turbocharge their data science and analytic efforts today. Here is an example from one of the world's 10 largest energy companies.

### Situation and requirements

Energy companies run on data and analytics. In the upstream operations at this firm, thousands of sensors continuously capture pressures, vibrations, temperatures, and hundreds of other metrics from hundreds of offshore oil rigs throughout the Gulf of Mexico. Numerous additional systems such as SAP plant maintenance add to their data deluge.

Providing this critical data to production, safety, financial, and compliance analytics applications was a huge challenge made more complex by the need to support a diverse range of engineers, managers, data scientists, and business analysts, each requiring unique, and often ad hoc slices of the data.

### TIBCO's solution

Using the TIBCO Data Virtualization platform, the company built an enterprise-wide data virtualization environment with over 600 standard business canonicals. These easy to find and use datasets simplify complex data and make it available to a wide range of data science and analytic tools while supporting over 100,000 queries per day. And with one virtual place to go for data, analytics users stopped replicating data in costly uncontrolled, rogue data marts.

TIBCO Data Virtualization classic and MPP engines, high-performance algorithms, and optimizers enable data scientists and analysts to query large datasets with as many as two million rows in a single query. Prior to the TIBCO Data Virtualization platform, the firm was often forced to constrain queries even for making important decisions on subsets of the data. To further improve query performance, TIBCO Data Virtualization caching preloaded frequently used datasets into a Netezza analytic appliance.

With its TIBCO Data Virtualization platform, the company was able to accelerate new data science and analytics projects. Reusing existing datasets across analytics efforts saved significant development time. And whenever a brand new dataset was required, it's data engineers could quickly provide it.

## Results achieved

Using the TIBCO Data Virtualization platform, the company gained a number of benefits including:

- Reduced offshore downtime by 1%, yielding significant revenue increases.
- Accelerated development of new data science and analytic applications, improving decisions, while reducing development costs.
- Enabled engineers, managers, data scientists, and business analysts to improve analytic development productivity.
- Reduced rogue data mart growth rates and costs.

## Getting started

If your data scientists and analysts are challenged by a complex data landscape, the TIBCO Data Virtualization platform can help you turbocharge your efforts so you can deliver more insights, drive better business outcomes, and outperform your competition, far faster than ever before.

Learn more at <https://www.tibco.com/products/tibco-data-virtualization>.



**Global Headquarters**  
3307 Hillview Avenue  
Palo Alto, CA 94304  
+1 650-846-1000 TEL  
+1 800-420-8450  
+1 650-846-1005 FAX  
[www.tibco.com](http://www.tibco.com)

TIBCO Software Inc. unlocks the potential of real-time data for making faster, smarter decisions. Our Connected Intelligence platform seamlessly connects any application or data source; intelligently unifies data for greater access, trust, and control; and confidently predicts outcomes in real time and at scale. Learn how solutions to our customers' most critical business challenges are made possible by TIBCO at [www.tibco.com](http://www.tibco.com).

©2018-2020, TIBCO Software Inc. All rights reserved. TIBCO and the TIBCO logo are trademarks or registered trademarks of TIBCO Software Inc. or its subsidiaries in the United States and/or other countries. All other product and company names and marks in this document are the property of their respective owners and mentioned for identification purposes only.

25 Jun 2020