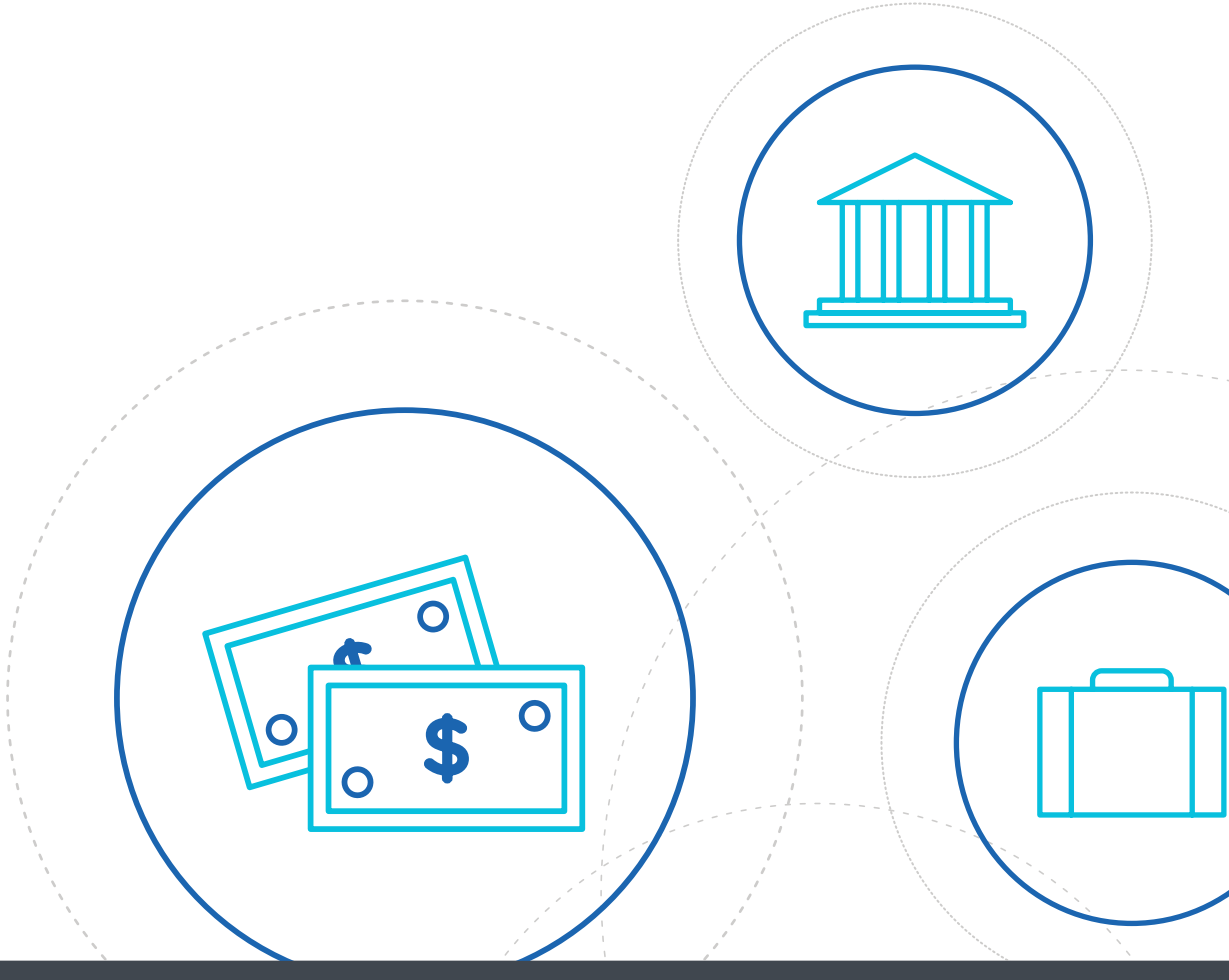
TRIFACTA

575 Market St, 11th Floor
San Francisco, CA 94105

www.trifacta.com
844.332.2821

WHITEPAPER

The Opportunity for Data Wrangling in Financial Services and Insurance



Disruption with Big Data

> "The big data revolution and the ever-increasing connected world have paved the way for disruptive change and opportunity in the financial services and insurance industries."

Financial Services and Insurance companies have built their success on recognizing the importance of data. Every interaction a customer may have with an institution can produce actionable insights that have potential business value associated with it. Those involved in commercial banking, capital markets, asset management and/or insurance know that success requires weighing millions of data points in the context of constantly changing macro environments.

The big data revolution and the ever-increasing connected world have paved the way for disruptive change and opportunity in the financial services and insurance industries. Firms now have access to never-before-seen volumes and sources of data in real or near real-time. Whether it is payment geolocation, mobile phone application usage, connected home devices, telematics providing greater specificity for insurers to tailor risk models and premiums, or real-time sentiment analysis providing leading indicators of capital market behavior, big data can provide immense payoffs.

Furthermore, dynamic and stringent regulatory requirements and existing data management can cost banks billions annually in both staffing and IT. Big data can provide a means to turn the cost of regulation into profit. Staying ahead of the competition is often a product of efficient forecasting models driven by the requirement for high quality data. Those who can predict and manage risk while identifying future trends reap sizable rewards. Harnessing the opportunity of big data allows firms to introduce novel and value-adding data sources to operating models. Ultimately, this presents a new value proposition around business innovation, risk mitigation, advanced personalization and enterprise-wide operational efficiencies.

Digital Revolution Needs Data Wrangling for Efficient Information Use

Every credit card swipe, loan payment, car accident, or equity trade leaves a trail of data that financial services or insurance companies track as a part of their internal operations. Given the number of customers these companies serve and the sheer volume of transactions and events they monitor, this data is rich with potential insights.

The financial services and insurance industries face an inflection point as the digitization of everything, along with the advent of big data initiatives, vastly expand the amount and types of data that they can quantitatively analyze. Traditionally, forecasting models have been comprised of highly structured data points such as ticker-tape readings, transaction records or actuarial tables. But there is a growing opportunity to augment these models with numerous new data sources such as connected home feeds or machine logs; along with previously untapped unstructured data sources such as social media streams, corporate emails, regulatory filings, or news headlines.

> "The opportunity big data platforms like Hadoop offer is the promise of a common environment to land, combine and analyze data of any shape and size."

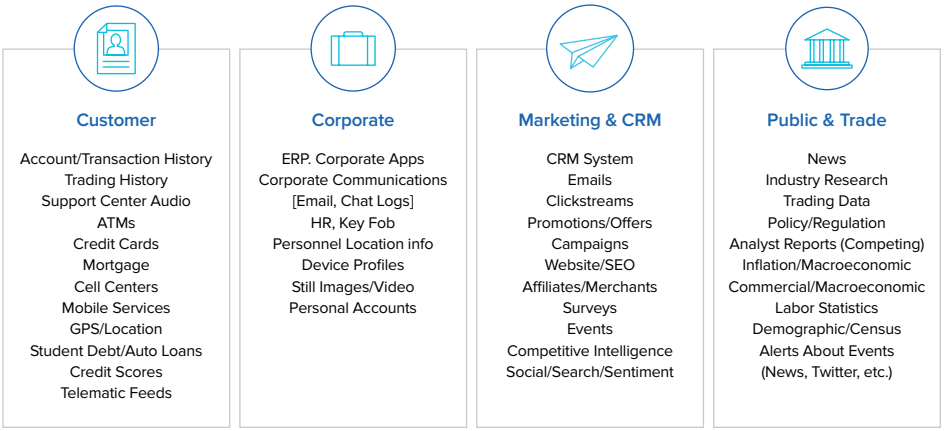


FIGURE 1: KEY FINANCIAL SERVICES AND INSURANCE INDUSTRY DATA SOURCES

To stay on top of their business, financial services and insurance firms may also listen to leading indicators of market performance or geopolitical risk via social and news sentiment analysis. Further information that was traditionally too difficult to parse and analyze, such as corporate email and chat room logs, may be excellent indicators of compliance and regulatory risk profiles. The analysis of new and existing data types requires a combination of structured, semi-structured and unstructured data sources, meaning they have to be wrangled extensively before the actual analysis can take place.

The opportunity big data platforms like Hadoop offer is the promise of a common environment to land, combine and analyze data of any shape and size. However, these projects often fail because of the challenges in defining the analysis workflow required for new exploratory analytics initiatives, and in providing tooling that empowers a diverse set of users to collaboratively work with diverse big data sources.

Common Challenges in Leveraging Data in the Financial and Insurance Industries

For all of the opportunity big data offers, there are many data management hurdles unique to the financial services and insurance industries. In addition to the overwhelming number and variety of data sources and data silos analysts must access and parse, firms must contend with many other data generating events such as working through complicated mergers and acquisitions, and dealing with both security and efficiency. Common challenges these firms face include:

Complex Regulatory and Compliance Requirements

Complexity of regulatory and compliance environments resulting from Dodd-Frank, Basel III, OEDC anti-bribery and other national and international legislative and judicial actions have placed significant pressures and complications on reporting. These firms must expend vast resources on bringing together disparate databases across multiple geographies and business logic conventions.

> "We wanted to build an in-house capability that was scalable and reusable. ... In terms of a measure of success, we achieved exactly what we wanted to do. It was easier than we thought, which was great."
>
> STEPHEN GEORGESON
> Analytics Manager
> Royal Bank of Scotland

Systems Set Up for Processing Speed Rather than Functionality

Given the outsized monetary rewards that financial services firms face to be first, best, fastest or most affordable, the industry has given rise to extensive homegrown and inflexible data systems. These systems are designed for specific tasks such as high-speed algorithm trading or highly compute-intensive risk modeling. These systems were not constructed necessarily to communicate with the broader corporate data environment. As such, there are flexibility, security and compliance constraints to uniting all of the sources necessary for data initiatives.

Inconsistent Data Structures

Firms must reconcile their corporate ambitions with the challenges of uniting data systems with incompatible interfaces and inconsistent architectures. For instance, defining assets under management may vary based on international conventions and differing regulatory environments dictate varying levels of capital constraints. Further, bringing new data sources such as telematics or unstructured sources into analysis often requires new data infrastructure and tooling. Even highly structured data such as ticker transactions occur in such high volumes that they can only be processed in modern systems and databases. Trying to combine this with unstructured data stored in data lakes and Hadoop environments requires new methods to bring the data to a coherent and usable state. Additionally, the rise of data-interchange formats such as JSON present challenges as organizations must clean, parse and unite data from new, legacy and hard to access sources in order to present a consistent view of external realities.

Burden on IT Teams

In many organizations, the IT organization is responsible for managing and disseminating data resources. As the volume and variety of data increases by terabytes of data from new sources, IT teams struggle to meet the needs of the data scientists and business users they support.

Skills Gap: Sustained IT Focus on Core Applications

Performing exploratory analytics on large, varied data sets requires a different set of skills, technologies and techniques not commonly found in existing business and data analytics teams. Finding and hiring the right mix of statistics, programming and Finance or Insurance domain expertise is difficult. Once resources with the right capabilities are on board, even then, their skills are often wasted as they spend a majority of their time in low-level cleaning tasks or waiting to get access to the data they need. Some estimate that data cleaning and prep tasks constitute 50–80% of the time and cost[1] in data warehousing and analytics projects.

Trifacta in Action

> "In the matter of a couple of hours, we managed to replicate what took me weeks to do without Trifacta."
>
> GUY NICHOLSON
> Senior Analytics Manager
> Royal Bank of Scotland

Empowering Finance and Insurance Organizations to Leverage Diverse Data to Drive Innovation

Trifacta was founded to provide an agile, broadly accessible and highly productive software platform for people who work with data. By removing friction stemming from a reliance on IT to access data and the length of time spent "wrangling" frustrating, complex data sets, Trifacta accelerates the process of making data usable.

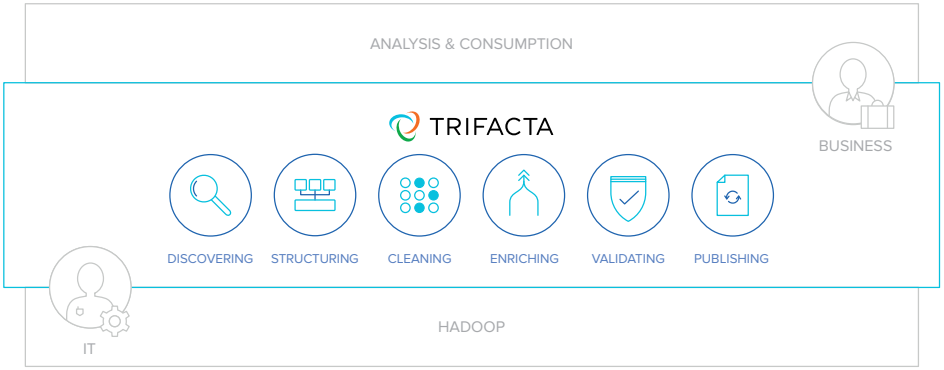


FIGURE 2: BRIDGE BETWEEN RAW DATA AND IT ANALYTICS

Trifacta is focused on empowering every organization to radically accelerate the process of preparing data of all shapes and sizes for analysis by providing an entirely new approach for how analysts access, transform and blend diverse sources of data.

Wrangling is this preparation process of converting diverse data from their raw formats into a structured and consumable format for business intelligence or statistical modeling tools. Trifacta offers an experience to non-technical users in a seamless and self-service way. Trifacta automatically **discovers** the data, **structures** it in a familiar grid interface, identifies potential invalid data and suggests the best ways to **clean** and transform the data. Trifacta learns from the user interaction, providing immediate feedback to the interactions, to better guide the user in **enriching** and **validating** the data at scale so it can be **published** with confidence to the next stage of the analytical process.

Trifacta sits between the data storage layer, being standard systems or a Hadoop platform and the visualization or machine learning applications used downstream in the process.

With Trifacta, firms can leverage their diverse data footprint through a series of simplified interactions that have proven to save organizations substantial amounts of time and resources traditionally allocated to data preparation.

Use Case Spotlight:
Data Wrangling in Fraud Detection

Important Financial Services and Insurance Industry Use Cases Powered by Data Wrangling:

- Fraud Detection and Security
- Governance, Risk and Compliance
- Intelligent Customer Insight and Service for Customer 360 Programs
- Analysis to Derive Prescriptive and Predictive Value
- New Product Development
- Optimizing Pricing with External Data
- Operational Efficiency
- Investment Management

We bridge IT and business groups by providing a common platform where the contents of diverse data are discovered and analytic requirements are defined and executed at scale.

The opportunity cost of inadequate data regulation has been highlighted recently in the financial services sector with high-profile fines ranging from the millions to billions of dollars, tarnished brand reputations and other legal action. Trifacta offers a unique solution to analyze fraud patterns across vast quantities of diverse data. In the following example, investigating groups were interested in joining different data sets within and outside the company in order to identify where fraud-related events may be occurring. Data sets to be wrangled included: customer and employee activity data, including emails and call detail records (in case of legal investigation); news and trading data sets including tickers, article topics and dates; and corporate data sets including corporate and individual relationship data within the company. Using Trifacta, analysts were able to identify and discover the structure and content of the data they previously did not have access to, and they were able to clean the data for downstream analytics and reporting.

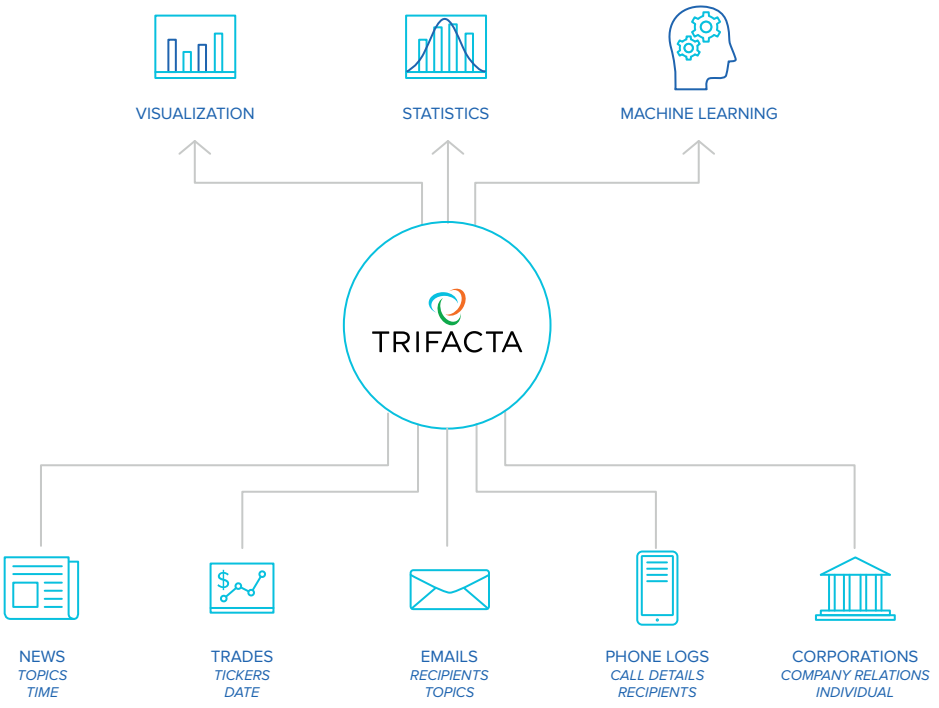


FIGURE 3: WRANGLING DIVERSE DATA SOURCES FOR FRAUD DETECTION ANALYSIS

Trifacta can accelerate the process of data analysis and significantly improve the productivity of the people in organizations that work with diverse big data sources.

The combination of big data and Trifacta enables better and faster outcomes for these important initiatives to deliver results for the organization in both innovation and cost control.

Creating Actionable Data Through Data Wrangling

Trifacta presents Financial Services and Insurance analytics teams with a new approach to dealing with the challenges of working with the scale, complexity and diversity of data today. Trifacta's approach to data wrangling utilizes the latest techniques in machine learning, data visualization and human-computer interaction to allow IT teams, data scientists, data analysts and business analysts to become more productive wrangling data themselves—allowing them to build and manage data products and transformation scripts more effectively and on-demand.

Benefits of Trifacta:

- **Empower the people who understand the data best:** By providing a breakthrough user experience, Trifacta enables business users to access any form of data and prepare it in a consumable layout for their analysis. This can be done by themselves in a self-service approach, breaking the dependencies from IT or complex hand coding languages.
- **Accelerate time to value:** productivity is drastically enhanced by an order of magnitude. Instead of frustrating cycles between IT and business to deliver and refine the data, the data analyst can explore and validate the data with immediate feedback for the insight he's looking for.
- **Lower business risk with more accurate data:** by allowing the end user to clean and validate data, the entire process is made more accurate and trustworthy. With correct information and trust in ones data, users lower the business risk in making decisions or implementing efficient data products.
- **Unlock innovation using a wider variety of data:** less time cleaning data means more time analyzing data. Trifacta liberates users from the cumbersome process of preparing data so they can focus on higher value analytics problems. With a simplified data preparation process, users can augment the number of data points to unleash business innovation, take competitive lead, optimize operational efficiency and reduce the cost of processes with a noticeable impact to the company's bottom line.

A Disruptive New Approach to Preparing Data

Experience a new way for working with diverse data—empowering analysts to interact with data in ways they never thought possible.

Interactive Exploration: Trifacta presents the user with automated visual representations of the data based upon the inferred data type of each attribute of the data. These profiles require no specification by the user and Trifacta automatically presents each data type in the most compelling visual representation—geographic elements are presented as maps; time-oriented elements are presented according the common hierarchies such as day, month, year; etc. Every profile is completely interactive—allowing the user to simply select certain elements of the profile to prompt transformation suggestions.

About Trifacta

Trifacta, the leading data wrangling solution for exploratory analytics, significantly enhances the value of an enterprise's big data by enabling users to easily transform and enrich raw, complex data into clean and structured formats for analysis. Leveraging decades of innovative work in human-computer interaction, scalable data management and machine learning, Trifacta's unique technology creates a partnership between user and machine, with each side learning from the other and becoming smarter with experience. Trifacta is backed by Accel Partners, Greylock Partners, and Ignition Partners.

For Additional Questions, Contact Trifacta

www.trifacta.com
844.332.2821

Experience the Power of Data Wrangling Today

www.trifacta.com/start-wrangling/

Predictive Transformation: Upon pulling up a data set within Trifacta, users are presented with a visual representation of the data set they are working with. These visual representations are interactive—enabling the user to click, drag or select over the specific elements or attributes of the data they'd like to manipulate. Every interaction within Trifacta leads to a prediction—the system evaluates the data you're working with and the specific interaction applied against the data to then recommend a ranked list of suggested transformation for the user to evaluate or even edit, depending upon what they're trying to do.

As users browse through the different suggested transformations presented to them, the system will also present a preview of how each transformation, when applied to the data, will impact the data itself. This iterative feedback loop is always occurring throughout the use of Trifacta—constantly taking inputs from the data and the user to intelligently recommend ways to manipulate the data and giving the user the ability to validate their work with previews of each transform.

Intelligent Execution: Every transformation step defined by the user in the application is logged in Trifacta's domain specific language called Wrangle, allowing the application to take the finished script the user is defining in Trifacta and compile that down into the appropriate execution framework, based upon the scale of the data the user is working with, and the type of transformation. Depending upon the data, Trifacta can compile down to Pig, Spark and Trifacta's own execution engine for jobs that can run on a single machine. This is all done behind the scenes—abstracting the user from the underlying execution framework.

Collaborative Data Governance: Although the core focus of Trifacta is enabling the people who know the data best to be able to access and transform it themselves, we recognize that organizations require having centralized processes for determining who has access to data, how metadata and lineage are tracked, how transformation jobs are operationalized and how data sets and transformation scripts are shared with other users. Instead of creating a completely separate governance framework in Trifacta, we have built support for the existing enterprise standard frameworks on Hadoop for security, user authentication, access controls, job scheduling and so forth. This enables Trifacta customers to simply implement existing governance policies in Hadoop instead of creating a new, entirely separate governance framework for Trifacta.

Sources

[1] New York Times, For Big-Data Scientists, 'Janitor Work' Is Key Hurdle to Insights, August 2014