

Faster Data Prep, Better Predictions

Solution Highlights

- World-class integrated data preparation and predictive modeling in a fraction of the time of traditional methods
- Better data quality inputs that result in more accurate predictive models
- 60–80 percent faster data cleaning and blending for predictive modeling
- Cloud-based solution for immediate results

Advanced Analytics Trends

In order to gain a competitive edge and improve their business, organizations need to take advantage of new technologies like automated machine learning and data preparation tools that allow them to adopt a user-friendly approach to their operations, automate processes, optimize outcomes, and deliver impactful insights. With these technologies, employees of all types and skill levels in every corner of the organization are empowered to identify business functions where analytics, artificial intelligence (AI), and machine learning deliver bottom-line value.

Traditionally, data preparation and developing and deploying machine learning models have each required teams of business analysts and data scientists months to complete. Organizations cannot afford to spend unnecessary amounts of time using traditional methods if they want to keep up with the rapidly changing competitive landscape. Trifacta and DataRobot envision a world in which data preparation and predictive modeling accelerate and optimize the creation of highly accurate models and impactful insights that increase human productivity, speed up decision-making, and deliver tangible competitive advantages.

The Advanced Analytics Data Pipeline

To realize this vision of user-focused advanced analytics, data-driven organizations need to modernize their data pipeline, starting with implementing a well-defined and iterative process to deliver the most accurate and actionable predictive insights from data.

The first phase of this process is to identify, test, and train various machine learning models to uncover the most robust predictive solution. In order to deliver the results, the next step is to operationalize the model. See Fig. 1 for more details.

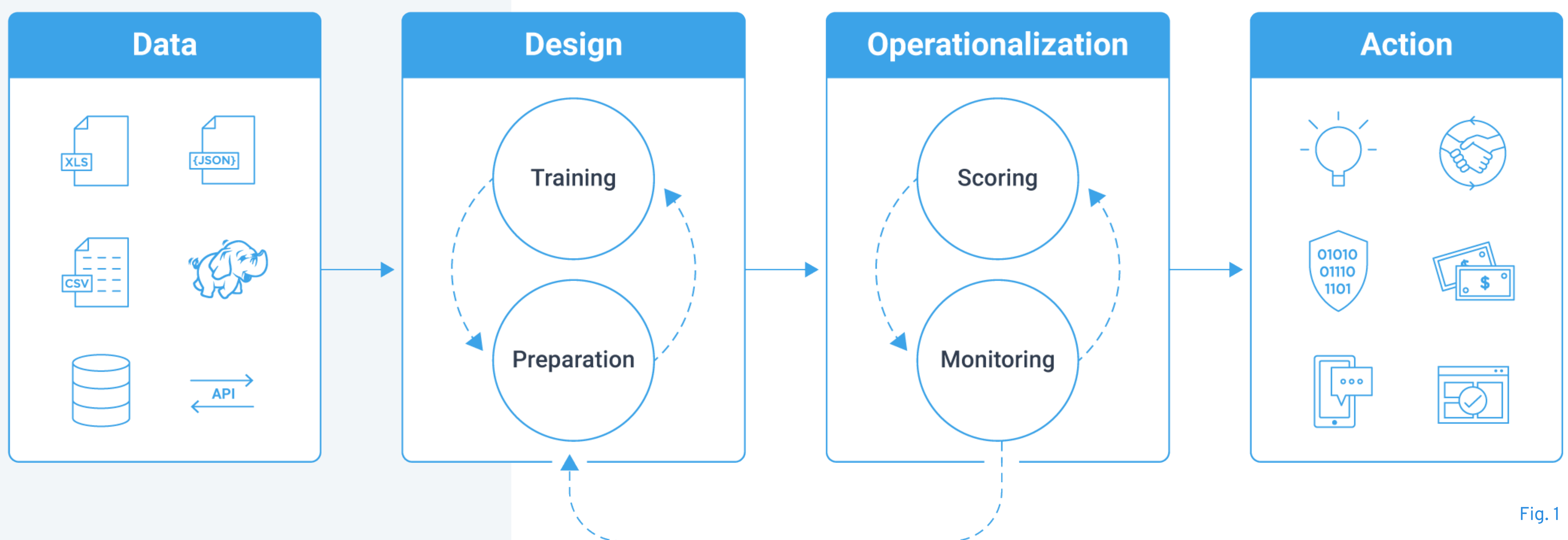
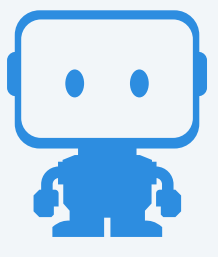


Fig. 1

Designing an advanced analytics solution consists of accessing, structuring, cleaning, combining, enriching, and deriving metrics and variables that will act as inputs for the models. Data quality has a direct impact on the accuracy of the model, but data preparation is an often-underestimated challenge. Often, data scientists and business analysts spend countless hours preparing data to get it to a clean and consistent state for predictive modeling, which can take up to 80 percent or more of a data analytics project timeline.

While data scientists can use their R or Python skills to prepare data, using programming languages is not the ideal solution because they require a lot of custom code to discover and structure the data. This may leave unidentified errors in the data, in addition to being complicated and slow. Data scientists would certainly prefer to focus on tuning models or supporting more predictive initiatives to advance their business departments rather than spending their precious time on the arduous work of cleaning data. Similarly, business analysts often have to massage data in Excel in order to get it into a usable state for machine learning. The problem with these manual methods? They are non-repeatable, non-scalable, and error-prone – in other words, they are inefficient, time-consuming, and downright painful.



Data Prep and Modeling with Trifacta and DataRobot

- Upload and connect to internal and external data sources
- Wrangle the data into a consistent and clean state
- Test and identify the best model to be implemented
- Refine the models to get a better score (Can involve re-iteration of 1, 2, 3)
- Operationalize the model

Trifacta and DataRobot empower users to upload or connect to their own disparate datasets of any size and form; clean, standardize, and combine them into a consistent state; then ingest the prepared dataset into DataRobot, which tests, trains, validates, and identifies the best model with the most robust prediction to operationalize.

A Better Approach to Data Preparation and Predictive Modeling

Imagine a world in which, instead of spending time preparing data, analysts and data scientists focus solely on building predictive models and boosting the level of insight. Trifacta and DataRobot provide a solution that enables data scientists and domain experts like business analysts to prepare data at scale and generate the most accurate predictive models in a fraction of the time of traditional approaches. See Fig. 2 for more details.

Both DataRobot and Trifacta run in the cloud on Amazon Web Services (AWS), offering a rapid approach to getting users started with advanced analytics initiatives. Additionally, the solution can scale as needed to address any data preparation and machine learning challenges.

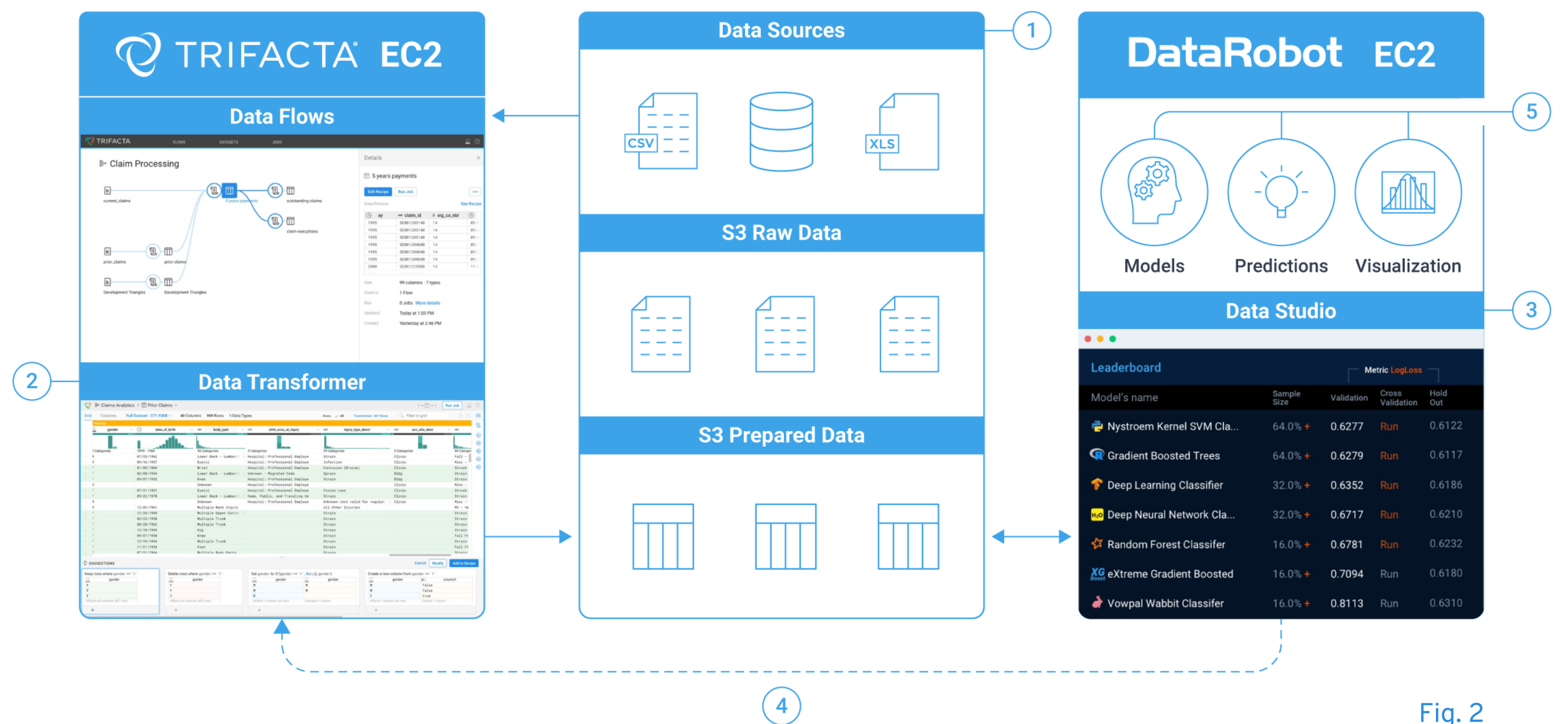


Fig. 2

Data Sources

Data Wrangling Activities

Discover

Structure

Clean

Enrich

Validate

DataRobot

Build

Validate

Tune

Deploy

Machine Learning Activities

Actionable Predictive Models & Analytics

TRIFACTA

Wrangling data is the most time-consuming and inefficient part of any data project, often taking up to 80 percent of a project's time and resources. Trifacta enables anyone to more efficiently explore and prepare diverse data by utilizing machine learning to provide a breakthrough user experience, workflow, and architecture.

- Assess data quality upfront
- Clean raw and messy data
- Combine disparate data sources
- Onboard external or third-party information of any form and volume
- Structure, enrich, and create calculations to inform predictive models and analytics solutions

DataRobot

The DataRobot automated machine learning platform captures the knowledge, experience, and best practices of the world's leading data scientists to deliver unmatched levels of automation and ease-of-use for machine learning initiatives. DataRobot enables users of all skill levels, from business people to analysts to data scientists, to build and deploy highly accurate predictive models in a fraction of the time of traditional modeling methods.

- Empower users to build and deploy predictive models
- Deliver more predictions in less time with automated machine learning
- Deploy and scale ML/AI applications
- Amplify data scientists' skills

Learn more at <https://www.datarobot.com/partner/trifacta/>

Contact Us

DataRobot
 One International Place, 5th Floor
 Boston, MA 02110
www.datarobot.com
info@datarobot.com