
CITO Research
Advancing the craft of technology leadership
Machine Intelligence:
The Golden Age of Analytics
SPONSORED BY
nutonian

CONTENTS

Introduction

The exploitation of data in the business world can be viewed as a supply chain. In the past, we had a small supply of data that was carefully curated and modeled in a data warehouse. When possible, predictive analytics were painstakingly applied. In short, we had a low-resolution view of the world and applying advanced analytics was difficult and expensive.

The improved economics for storing and analyzing data have still not produced a golden age of analytics

In the modern world, much has changed. The data supply chain has a much different shape but is not yet complete. At the beginning of the supply chain we have raw materials in abundance. In addition to data warehouses, we have big data from myriad sources, and far more enterprise applications pumping out curated data. The level of signal varies widely by data source, but the means to store and analyze data have been democratized. It has never been cheaper to hunt for signals in large data sets. But we still have a big bottleneck in the supply chain. The abundance of data and the improved economics for storing and analyzing it have still not produced a golden age of analytics.

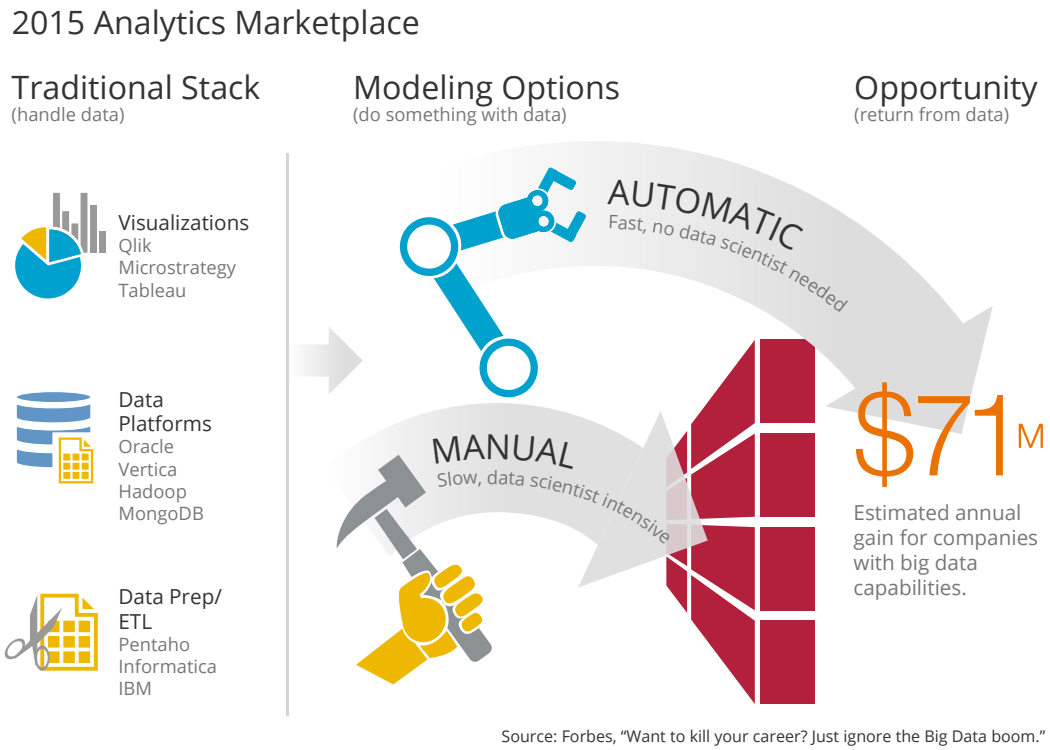
The problem is that analytics has not been sufficiently productized. To enter the golden age of analytics, we need systems that allow a normal human, not a virtuoso data scientist, to find signals, model them, apply analytics, and connect insights to business actions. We need such systems because we cannot create virtuoso data scientists at scale nor can we educate data scientists about all aspects of our businesses. The golden age of analytics requires systems that package and productize data science and advanced analytics for use by business analysts, who understand key value-creating processes and high-value questions that need answers.

At CITO Research, we are constantly on the lookout for products that empower business analysts. One very promising approach is machine intelligence. This paper explains how the application of machine intelligence can break the analytics bottleneck.

Born from the necessity of a new approach to advanced analytics, machine intelligence is an approach that automatically and independently discovers analytic models. Drawing on today's virtually unlimited compute power, machine intelligence solutions automate the modeling process while promoting interpretability and interactivity regardless of an end-user's experience with advanced analytics. Machine intelligence is a disruptive innovation that represents a complete shift in the way data scientists, analysts, and organizations approach advanced analytics. The availability of automated machine intelligence applications will enable organizations to scale their data science initiatives across levels and functional areas.

"If finding the meaning and insights in data is only a virtuoso activity, then great data analytics will be as rare as Michelin three star restaurants."

– Dan Woods, CTO and Editor, CITO Research



Source: Forbes, "Want to kill your career? Just ignore the Big Data boom."

Hypothesis Hubris

Prior to the industrial revolution, individual craftspeople made everyday goods. Today, advanced analytics are manually produced by a small group of highly specialized individuals. Despite continued innovations in data infrastructure and visualization tools, analytical modeling remains the domain of technically sophisticated users applying complex techniques and using tools developed more than 20 years ago. Decades of advances in both data management technologies and computing have made oceans of data available, yet unlocking answers is tied to specialized and labor-intensive methods.

Tools for analytical modeling such as SAS, MATLAB, and Python have not evolved far from their hypothesis-driven roots and offer a host of challenges that limit the potential user community and may ultimately distort or obscure final results. Primarily built for use by data scientists and statisticians, these tools continue to rely on slow and error-prone manual testing frameworks developed decades ago.

This testing framework requires specialized individuals to first make assumptions about the underlying data and manually manipulate the data to find how well it fits a particular assumption. The manual process of cleaning, transforming, modeling, and validating data leads to a process serial in nature and rife with opportunities to introduce biases and errors. While highly-skilled practitioners may feel confident in their abilities to generate accurate analytical models, the resulting model will only reflect the specific questions the user thought to ask. In other words, even virtuosos are ill-served by these analytical tools when contending with modern analysis needs.

The manual process of cleaning, transforming, modeling, and validating data is rife with opportunities for bias and error

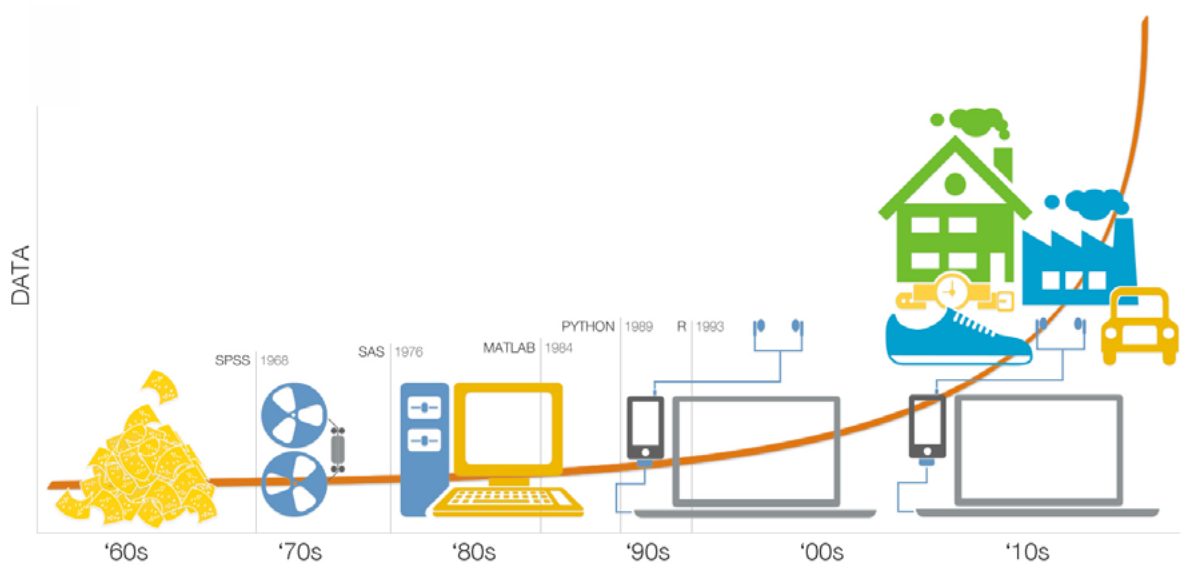
Analytical modeling has proven value in discovering and driving competitive advantage, yet businesses are constrained by the challenges of manual modeling, which include:

- **Model Composition:** Analytical models are composed of both individual variables and combinations of variables, called features. Blindly including every variable from a dataset in a model will often cause the model to "over fit" the data; additionally, there is an almost infinite number of potential features that could be comprised from those variables. A data scientist or statistician needs to ascertain the best subset of variables and features to use in a model, but the act of predetermining features for inclusion introduces bias well before the modeling process begins.

- **Algorithm Selection:** Next, a machine learning algorithm needs to be chosen. Different algorithms have different strengths, and a data scientist only has a finite amount of time to test a finite number of algorithms. Some algorithms prioritize interpretability at the expense of accuracy, while most prioritize accuracy at the expense of interpretability. The field of data science tends to favor accuracy, leading to complex analytical models that are hard to explain to non-data scientists and difficult to implement quickly. Additionally, the act of predetermining a certain algorithm structure and measuring how well the data approximates the forced structure introduces further bias in the modeling process.

- **Iteration:** Once both the features and algorithm have been selected, the data scientist then trains the algorithm to produce an analytical model from a subset of features. Algorithms can easily be "overtrained," unable to successfully be generalized on new datasets, and properly training algorithms to provide a generalizable model requires weeks to months of manual effort by a data scientist.

- **Model Selection:** After the algorithm generates a number of candidate analytical models, the data scientist must manually test and validate the results. There are a number of statistical methods for selecting a final model, including Bayes factor and Akaike criterion, but each can only be applied separately. Depending on the goal of the project, the data scientist may choose a certain way to select the winning model and omit critical information contained within the others.



Automation Acceleration

Just as technological innovations enabled manufacturing industrialization in the 18th century, new breakthroughs are enabling the industrialization of advanced analytics today. These breakthroughs are being driven by machine intelligence that takes advantage of nearly unlimited computing power. Not only can machine intelligence relieve the strain on the limited number of data scientists, but it can extend their capabilities into new areas by making advanced analytics accessible, repeatable, efficient, and cost effective.

Machine intelligence dramatically improves time-to-value at both ends of the analysis cycle: first, automation searches through billions of models every second to discover the optimal analytical models without human intervention; second, it empowers users to leverage their business expertise and rapidly iterate to find the answer that is right for each specific business opportunity, allowing them to conduct what-if analyses and see the impact their changes would have on the model. This empowers users to ask questions such as, What if the price of this commodity goes up (or down)? What if demand fluctuates?

Machine intelligence allows businesses to scale their data science initiatives into every department and incorporate critical subject matter expertise into the process. By incorporating a holistic vision for data science across previously siloed areas of data transformation, feature creation and selection, model creation, and model selection, machine intelligence now enables a number of new advantages for companies that adopt it.

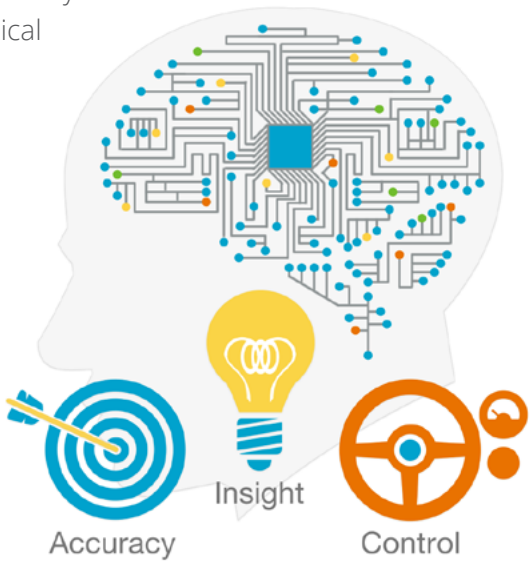
- **Automatic Input Transformation:** Bigger is not always better, and modeling across thousands of possible input variables can lead to subpar results. Instead of wasting computation on irrelevant noise, machine intelligence automatically selects the right data to use for your modeling project by creating features and selecting only the variable and feature interactions with the most explanatory power.

- **Evolutionary Nonlinear Inference:** Machine intelligence leverages genetic algorithms to automatically determine the optimal analytical model structure from the underlying data. Instead of approximating the data within a predetermined structure, machine intelligence tests hundreds of millions of models per second to distill complex data into its most concise form and structure.

- **Human Interactive Modeling:** Users are encouraged to interpret and interact with the results, encoding it with their domain-specific knowledge, which shifts the focus away from programming and statistics expertise. There will never be a substitute for human creativity and domain expertise, so interpretable results promote interactive, rapid iteration, enabling higher confidence levels in results and quicker time-to-value for an organization.

- **Multi-Objective Model Optimization:** Recognizing that not all use cases are equal, machine intelligence allows for different approaches to different situations. Rather than optimizing for a single, predefined "best" outcome, machine intelligence presents a series of solutions that reflect the tradeoff between accuracy and simplicity. This series of choices enables end users to identify the optimal solution for their specific circumstance.

Numerous vendors are touting "data science for the masses" in response to the growing market need for easy, advanced analytics. However, the majority of these tools only address specific aspects of improving analytical modeling and fall short of delivering machine intelligence. For example, visualization tools have begun adding connectivity to various data science tools while more complex analytical modeling tools are now offering drag-and-drop functionality which still require advanced coding, data science, and statistics knowledge to implement. These approaches yield only incremental improvements over existing siloed processes.



The key goal is to make advanced analytics available to a wider audience. Vendors that empower hundreds or thousands of people and provide them with the ability to find important signals in data will be the largest creators of value.

Enter Eureqa®

Nutonian, a Massachusetts-based company with deep roots in research, takes a fresh approach with its flagship product, Eureqa. Founded in 2011 by two of the world's leading data scientists, Nutonian empowers business professionals to automatically discover analytical models via sophisticated algorithms requiring no human intervention. Similarly to how applications like Excel have become standard tools for basic spreadsheet and office productivity, Eureqa's revolutionary machine intelligence capabilities provide similar ease of use for the more complex data science activities within an organization. With Eureqa, someone who has normal data analysis and manipulation skills can actually do one of the most challenging tasks in the world of data science: invent new models.

Eureqa's difference lies in its use of free form modeling. Free form modeling leverages sophisticated evolutionary algorithms to automatically find the deeper patterns underlying a dataset without human intervention. Eureqa independently transforms, hypothesizes, tests, and validates models to achieve the same type of results one would expect from a seasoned scientist. The process works because automation and scalable computation resources enable Eureqa to repeat this process hundreds of millions of times per second. In this way, a vast number of models is considered, but only those that matter rise to the top. Compared to the age in which most other analytical modeling techniques were created, the recent emergence of cloud platforms and SaaS delivery have led to plummeting costs for computation, storage and processing systems, all of which Eureqa takes full advantage of.

Despite the high-powered evolutionary algorithms used behind the scenes, Eureqa presents analytical models with transparent mathematical, visual and written explanations. By prioritizing the interpretability of the results, Eureqa makes the process of evaluating analytical models easy to understand and conducive to iteration and further exploration. Not only can business users explore results using Eureqa's intuitive interface, but its models can be easily exported to other tools, such as SAS, MATLAB, R, Python, or Tableau, and used throughout a company's analytic ecosystem. Data scientists can use Eureqa to dramatically accelerate model development, and business users can export new KPIs from Eureqa into their visualization dashboards.

Like a concierge, Eureqa acts as a trusted advisor that immediately offers understanding into the best way to spend your time but accepts direction regarding your wishes and needs. Eureqa provides intuitive methods for simulating new scenarios, addressing follow-up questions and iterating on any unclear areas, with specific and interactive functions that are accessible to any type of user. Additionally, industry-specific analyses are available for accelerating analysis life cycles on common project types, and users can build custom interactive reports for collaboration with colleagues.

Consider some of the features of Eureqa:

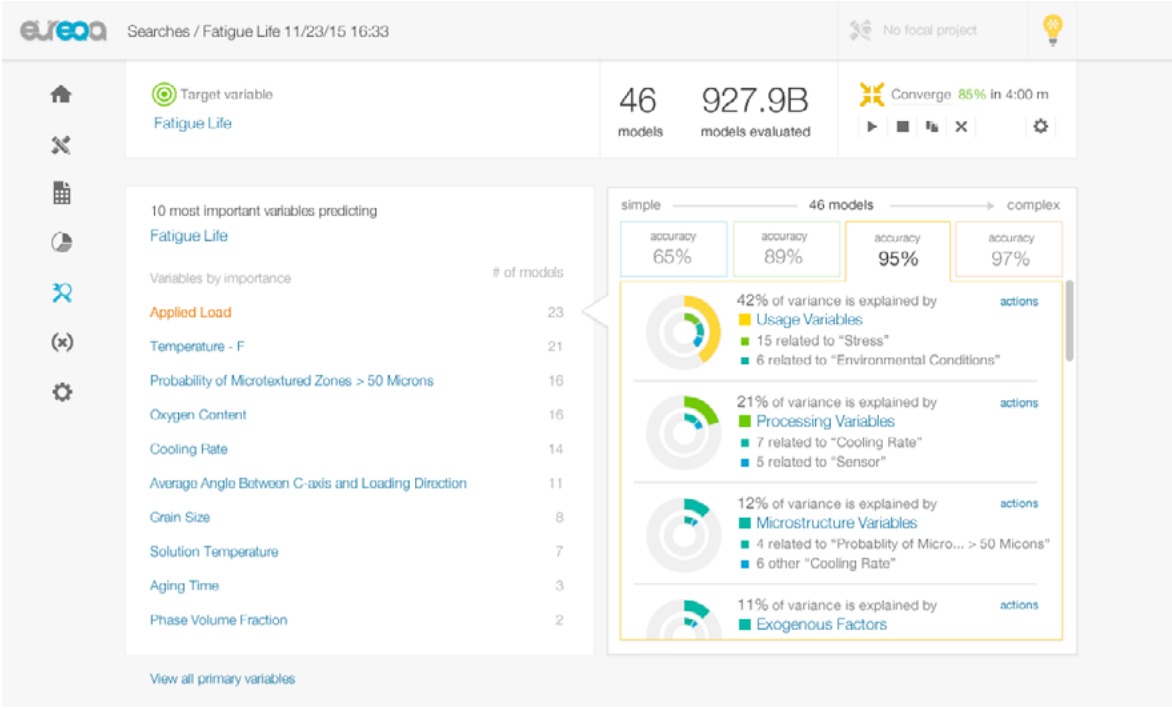
- **Automatic:** Eureqa automates best practices while searching through all potential relationships, effectively transforming raw data into accurate analytical models. This automated virtuosity enables businesses to scale data science initiatives quickly and efficiently alongside the dynamic nature of business requirements.

- **Infinite:** Eureqa is not bound by user-bias or an existing finite number of testable algorithms, allowing unique non-linear relationships and solutions to be discovered. Eureqa detects patterns that may not have been programmed or thought of before; it arrives at the right answer instead of the expected answer (see "Finding Problems in the Mix: Metal Manufacturing" later in this paper).

- **Dynamic:** Eureqa adapts results and gets smarter over time as it is given new information. As evolving market and business conditions cause changes in the underlying data and as users leverage their domain expertise, Eureqa dynamically learns and adapts accordingly.

- **Iterative:** Eureqa re-computes results in real time in response to user input, enabling effective what-if analysis as users explore various outcomes based on changes to particular variables. Incorporating critical domain expertise from business users without needing to go back to the drawing board every time leads to dramatically improved results in a shorter period of time.

In essence, Eureqa acts as a virtual data scientist, automatically building analytical models and providing the right level of abstraction for any user within an organization to consume, interact with, and iterate on the data.



Finding Problems in the Mix: Metal Manufacturing

Today, many product components are created using metal that has been reduced to powder; in this process, parts can be manufactured without milling. As a leader in the production of these powders, Rio Tinto, Fer Et Titane (RTFT) was having a problem. As sample parts were created using powder-based metals to ensure quality, batches showed inconsistencies. There seemed to be no real pattern to the problem occurrences, but one thing was certain: there was a quality issue in the making.

Thanks to Internet of Things-style sensor data, RTFT had more information than ever about the processing conditions under which the powders were being produced. But in all that data, what could explain the sporadic quality problems that were plaguing them?

Marko Litalien, Superintendent of Process Development, was certain that the data held the answer. He looked at some multivariate analysis tools and found they were only capable of handling a small portion of the available data. Litalien discovered Eureqa while reading about its success in uncovering laws of science from raw data. He talked to a colleague, Frédéric Benoit, a process engineer, about giving Eureqa a try.

As Eureqa analyzed the data, Benoit first saw something familiar: a condition he expected to be related to the problem. Then he saw Eureqa discover another relationship: one he believed was irrelevant and never would have thought to include. Further research by metallurgists on staff confirmed the relationship between the findings and their quality issues. This newfound link is still largely unknown within their industry, leading to a tremendous advantage over RTFT's competitors.

Says Benoît, "We're living in an era where we have an enormous amount of data that no one really understands. When I see an application like this that can take that data and make sense of it, I think we're going in the right direction."

Eureqa in Action

Nutonian has a large and growing market share, with more than 80,000 unique installations globally, more than 4,000 active community members, and more than 300 commercial deployments. The company provides solutions for every business, regardless of size and in-house expertise, by discovering complex and nonlinear analytical models across use cases such as forecasting, system identification, and classification in transparent and scalable ways. Here is a sampling of some of the companies using Eureqa to uncover value in big data.

Nutonian provides solutions for every business, regardless of size and in-house expertise

Maximizing Sensor Monitoring Capabilities

Extracting oil at maximum efficiency throughout every step of the process is a high priority for the US energy industry. Monitoring the quality of individual well zones is one area of particular focus. Surveillance of well performance enables engineers to proactively control equipment settings to mitigate risk and increase production. Developing cutting-edge sensor packages capable of operating in extreme conditions is a must, along with the ability to identify high-value patterns within the most chaotic data.

A Fortune 500 energy company turned to Eureqa to determine the optimal sensor configuration capable of operating within intense environmental conditions and limited power constraints while delivering critical information. Their scientists understood the physical, chemical and geological principles involved, but had difficulty translating their strong domain knowledge into predictive models capable of accurately identifying the optimal relationships across disparate datasets. Eureqa automatically built transparent models open to rapid iteration, enabling the scientists to leverage their understanding of the wells to significantly improve overall performance by applying real-time course corrections.

Optimizing Targeted Promotional Material

Sending product samples is a common and effective way for consumer packaged good manufacturers to promote new products. A major nutrition manufacturer turned to Eureqa to improve their audience targeting and send samples only to highly qualified prospects. Due to the high cost of sending samples, the marketing department needed to be precise and methodical in selecting recipients. The marketing analysts had already been building models in SPSS and SAS to select sample recipients, but models took too long to build and required purchasing numerous expensive data feeds to augment their survey response data.

The company compared results from their in-house techniques to results using Eureqa. After using Eureqa to merge thousands of internal and external data attributes and automatically derive the finite number of inputs to maximize customer targeting, their predictive accuracy improved by over 15%. Eureqa also provided features that allowed the analysts to tune models to important business criteria, allowing them to statistically take marketing objectives and the cost of sending samples into account during the targeting process and confidently anticipate response rates. Best of all, the company saw enormous improvement in turnaround time and time-to-value by leveraging Eureqa's automated modeling process.

With Eureqa, predictive accuracy improved by over 15%

Preventing Late Payments

What if sales are going well, but business to business payments are consistently late? A Fortune 500 utilities company faced this common business challenge. With a multi-billion exposure in late payments, the company continually ran into challenges and unpredictability with their average daily working capital. Determining the factors that result in late payments involved the inclusion of many data sets: product data, contract data, and supply chain data as well as geographic and demographic variations in payment timing. Internal experts at the company collected data and produced monthly KPIs, but struggled to reach a holistic understanding of the factors driving late payments.

Merging and analyzing the data with Eureqa automatically revealed a rank-ordered list of contract characteristics driving late payments, allowing the company to focus on changes with the greatest impact on the bottom line. Additionally, Eureqa was able to provide a risk probability of late payment for all transactions, accurately identifying the overwhelming majority of late payments.

Conclusion

Nutonian is one of the early examples of successful productization of advanced analytics. Companies that use Nutonian will exponentially increase their analytical capacity, leading to a golden age of analytics.

Nutonian's Eureqa combines intuition and data science at the heart of decision-making, leading to real value for business leaders and increased analytical capacity for any organization. Instead of relying on incremental improvements to legacy analysis methods, Eureqa's value lies in its ability to leverage the power of machine intelligence coupled with raw computation to expose results in a way that incorporates human creativity and domain expertise.

CITO Research believes that Eureqa represents a groundbreaking development with its ability to automatically convert raw data into understanding and sustainable competitive advantage. Organizations can quickly build automated analytical applications that scale across all levels of the business and adapt dynamically to changing business needs. Eureqa effectively industrializes and decentralizes the practice of consuming and understanding data, sparking an analytics revolution for organizations with complex challenges across manufacturing, utilities, finance, and more.

The only thing Eureqa is waiting for is your business acumen and curiosity.

Learn more about Eureqa ▸

This paper was created by CITO Research and sponsored by Nutonian

CITO Research

CITO Research is a source of news, analysis, research and knowledge for CIOs, CTOs and other IT and business professionals. CITO Research engages in a dialogue with its audience to capture technology trends that are harvested, analyzed and communicated in a sophisticated way to help practitioners solve difficult business problems.

Visit us at http://www.citoresearch.com