

Best Practices for Executing New Analytics Initiatives

Introduction

The world runs on data. We value every medical record, consumer purchase, and stock trend to illuminate a path forward in our decision-making, narrowing our margin of error with each new insight. Recently, our dependency on data—and the rate at which we generate it—has only accelerated. IBM recently estimated that the world is generating 2.5 quintillion bytes of data each day, so much that 90% of the world's data had been created in the past two years. Today, leveraging this wealth of data is irreversibly intertwined with profitability.

But set aside the value we derive from everyday, fundamental data sources. What value do organizations miss out on by not digging deeper, exploring new data sets, and spearheading exploratory initiatives? As we know, the explosion of big data isn't just by depth, but also by breadth. Today, there are a variety of new data types available, of which many organizations have already taken advantage. According to Aberdeen Group and Ventana Research, 71% of organizations analyze more than six data sources and 23% use more than 20. Forward-thinking companies understand that they need a diverse data portfolio for cutting-edge initiatives and to gain competitive advantages.

Though organizations often first consider the risk in undertaking an exploratory project, there's the flip side of that risk, too—if you choose not to place a bet on big data, and your competition does, what will that mean for you business? Take companies in the top third of their industry, for example, who saw a 5% productivity increase and a 6% profitability gain from data-driven decision making, the Harvard Business Review reported.

Whether you're already planning a new analytics initiative, skeptical, or simply curious, this paper will help demystify the process of getting started.

Defining your Initiative

New analytics initiatives are a product of the research you put in up front. When considering a new initiative, think carefully about what you're trying to accomplish and what data sources you might need to leverage. Identify current pain-points on your team or areas where your analysis efforts could modernize or improve. Then, gut check against industry trends and standards—what are others doing to drive growth and how can you compete? All of these questions can inform how your initiative will take shape.

Leveraging Non-Traditional Data Sources

Nearly every organization has utilized a basic collection of data on customers, sales, finance, or marketing to manage their business over the years. More recently, the advent of the internet and new digital technologies has led to an ever-growing diversity of new data sources from which businesses must derive value.

Of these non-traditional data sources, social media data is a fairly common example. With an estimated 2.307 billion active global users on social networks and growing, social media data has become a critical component of enhanced marketing, customer service, and product development initiatives. Internet of Things (IoT) data is another great example of a non-traditional data source gaining serious traction across industries—Internet-connected devices, such as Nest thermometers or Tesla cars, are generating huge volumes of data and, as more and more devices enter the market, this figure will only grow. Cisco predicts that the amount of devices connected to the Internet will rise to 50 billion by 2020.

The tricky part comes in decoding these complex files, which typically arrive in a semi-structured format such as JSON, and joining it with other, more structured data sets, such as Excel files. This is true of other unstructured

data as well—text files, video logs, and web chats, for example, which are increasingly important, but difficult to work with using traditional technologies. For organizations embarking upon new analytics projects utilizing non-traditional sources, understanding how to successfully leverage these new data sets is critical.

New Analytic Initiatives in Financial Services

A new analytics initiative gaining traction in the financial services industry is Customer 360°, or gaining a complete view of customers' shifting needs, preferences, and behaviors to personalize service and predict churn.

For example, in order to improve customer service, The Royal Bank of Scotland set a lofty goal: analyzing unstructured customer web chat conversations, of which they receive upwards of 200,000 per month. Previously, the small team at The Royal Bank of Scotland simply couldn't keep pace with the scale of incoming data—on average, they could only analyze and categorize 200 web chats per month and were unable to make informed decisions.

This was highly complex data, yet after launching this initiative with the help of Trifacta, The Royal Bank of Scotland is well on its way to analyzing 100% of web chat data—a huge milestone from its previous 0.1%—and have dramatically increased visibility into customer activity. The team at RBS even reported that leveraging Trifacta to more effectively prepare and analyze this data led to a multi-million dollar business benefit within the first year of deployment.

New Analytics Initiatives in Healthcare

As healthcare data becomes increasingly digitalized, pharmaceutical companies and healthcare providers are presented with opportunities to work with new combinations of data to drive advanced research and more personalized care.

Sanofi, a pharmaceutical company headquartered in France, took advantage of disparate cancer research data and launched Project Data Sphere, an initiative that aggregated the collective body of cancer into a consolidated data hub for analysis. Doing so brought together researchers and clinicians to leverage new sources and combinations of data that allowed them to explore better cancer treatment options.

For Sanofi, the key to successfully launching this initiative was empowering non-technical researchers and clinicians—those who knew the data best—with a visual summary of the data's contents, as well as the ability to manipulate it themselves. Trifacta allowed Sanofi to accelerate data curation; today, the team only spends hours on preparation tasks they had previously expected to take months.

Short-Term Goals, Long-Term Gains

No matter the type of data you choose, it's important that your goal, such as improved fraud detection or more personalized patient care, has wiggle room for you to consider new alternatives and test out ideas. Launching a new analytics initiative should drive results. But it is also exploratory in nature; don't expect results overnight. Your goal should plan for the long-term by thinking strategically about where your organization will be, what forces are shaping your industry, and what it'll take to get there.

Defining the Stakeholders

Arguably, the most important part of a new analytics initiative is the people behind it. Identifying who will champion this initiative forward is intrinsically tied to its success, which means it's critical to get the right people involved early on. Below, we've identified some of the key players that you should involve when planning a new analytics initiative.

Business Lead

The business lead is the person dedicated to this initiative's results. They understand the potential impact it can have on the business and what that translates to resources or dollars. At the same time, they're also thinking long-term, understanding that building upon the success of one initiative could provide a beneficial domino effect for their team, function, or the entire company. They are hugely important, primarily because they have the most at stake to gain from a successful initiative, and should be the first identified person in any new analytics initiative.

Technology Lead

The technology lead is the person (or team) responsible for implementing, maintaining, and, in some cases, training others on the right technology to execute the initiative. They must examine existing technologies, ensure compliance and security, as well as plan for a technology solution and process that will last well into the future. The technology lead architects the end-to-end data platform and applications required for the initiative and understands how various users will access the data they need. They are instrumental in establishing a rock-solid foundation for the initiative and a valuable resource in maintaining its success.

Data Analysts & Scientists

Data analysts and scientists are in the trenches of data wrangling, applying best practices to achieve their desired output for analysis. Ideally, those performing the data wrangling also have the necessary business context to do so—not only is this more efficient, but it allows users to discover unexpected trends in their data early on. This is especially important when exploring new data sources when the value of these datasets isn't always obvious at the outset.

While data wrangling is often considered “janitorial” work, the process involves multiple steps that can have a huge impact on downstream analysis. Data analysts and scientists must consider each transformation and its impact on their datasets in order to generate a successful output. These six steps include:



Discovering: An umbrella term for the entire process; in it, analysts learn what is in the data and what might be the best approach for productive analytic explorations



Structuring: Necessary because data comes in all shapes and sizes; analysts need to format and standardize into a consolidated dataset



Cleaning: Involves taking out data that might distort the analysis, such as a null or inconsistent value



Enriching: Allows analysts to ask questions about other data that might be useful in the analysis, or new data that they can derive from existing data



Validating: Surfaces data quality and consistency issues, or verifies that they have been properly addressed by applied transformations



Publishing: Delivers the output of data wrangling efforts for downstream project needs such as visualization or statistics

Downstream Consumers

After the necessary transformations have been completed and the appropriate output generated, there are several downstream applications for analytic or visualization use. During this phase, it's common for project leaders or executives to get involved to understand how the data impacts their business objectives. While analysts can spot trends early on in the data wrangling process, this is when they're able to build more robust insights and showcase those insights to others across the organization. However, and this is key, the quality of analysis for downstream consumers is directly correlated with the data wrangling efforts put in up front.

Defining the Technology

The core technology elements required for a successful initiative are often labeled the "big data stack." These three elements have different variations, but are essentially responsible for one of the three core phases in your initiative.

Data Storage & Processing Platform

Any new analysis initiative relies on the underlying technology infrastructure where the data will be stored and processed. Given the scale and diversity of data that makes up new analytics initiatives, Hadoop has gained tremendous traction as the best data storage and processing platform for executing these projects. With Hadoop, organizations have the agility to store any data set in its native format, regardless of the data's format or volume. The Hadoop platform also supports a variety of different processing engines including Spark and MapReduce that can support data of all sizes.

Data Wrangling Solution

Given that the data making up these initiatives is typically not well-understood or well-structured ahead of time, it's critical that your team implements a data wrangling solution that enables analysts to intuitively explore the data and manipulate it into the appropriate format for analysis. This will allow the business individuals with most context for the data to interact with it themselves, instead of having to send rigid requirements to a technical counterpart and rely on them to prepare the data. An effective data wrangling solution will ensure that business users spend less time focused on the preparation and cleaning of data and more time focused on generating value from these non-traditional sources.

Analysis & Visualization Applications

Once your data has been effectively prepared, it will need to be more broadly consumed by the organization through the use of various downstream applications. Depending upon the specific initiative, your organization may need to leverage one or multiple tools to make best use of the data. These can include applications for data visualization, statistics, machine learning, or a custom application developed internally.

Mapping People to Technology

Assigning the right people to technology isn't always cut and dry. In many cases, you'll have multiple people working with data, which, with the right technology, instills a collaborative environment and more accurate results. For most organizations, collaborative data wrangling and analysis is a welcome change from the siloed efforts that legacy tools encourage.

To start, take stock of existing resources, considering both technical and business knowledge, in order to assess who is the intended user or administrator of each aspect of the technology platform. Given that modern data technologies allow for increased flexibility and autonomy among business units—today, users can accomplish more with less skills given the right tools.

Roles, Responsibilities & Technology

Data Architects: responsible for the selection, integration and implementation of the different technologies that make up the end-to-end analysis platform

Platform Administrators: responsible for managing the data storage and processing infrastructure

Data Engineers: utilizes open source and vendor-driven technology to land data onto the platform and automate processes

Data Analysts & Scientists: leverage data wrangling and analysis solutions to drive new sources of value for one or multiple business departments

Getting Started

After understanding both the right technologies to adopt and the necessary resources needed, you're well on your way to getting started. Researching the best vendors, implementing those technologies into the organization, and training your team will all take time, but the investment will prove worthwhile in the long run. Ensure that you have the appropriate stakeholders involved as you start this process and set realistic goals for your timeline.

Researching Vendors

Vendor research can leave most of us in a daze—every vendor's value proposition can look the same, so what's the difference? To help decipher the best from the rest, it's important to look under the hood as well as holistically. As you bring in each vendor, make sure representatives from each group identified above are in the room.

Implementation

Implementing any new technology can be tricky, but the technical nature of big data can be particularly complicated. In most cases, it's worth the extra cost to buy internal services from your vendor of choice, or hire an outside consulting team for additional help. This ensures that everything is set up correctly without any hiccups down the line.

Training

Training is another time-intensive process, but well worth the up-front cost. It's important to ensure that your team is comfortable with your new technology for high adoption rates and a successful initiative. Again, leverage any services provided by your vendor of choice for the most efficient and effective training, both at the onset and as

follow-up questions ensue. Encourage questions and work with real data—your team will become increasingly comfortable as they become accustomed to repeatable steps.

Launching the Initiative

Once you've completed implementation and training, you might feel eager to get started—and rightfully so. However, it's important that as you actually put your initiative into motion, you don't bite off more than you can chew. Specifically, keep these two tips in mind:

Start with a single question or hypothesis

Don't rush to tackle every new question you've stockpiled or leverage every available data source; rather, prioritize your efforts and start with one. Not only will this better set your team up for success, but will also help demonstrate success to the rest of the organization.

Set small goals and build

Set realistic goals, timelines, and expectations for your team. Don't promise immediate results—help executives understand that this is a long-term gain, and improvements in efficiency will come, but not necessarily immediately.

Next Steps

For most organizations, taking those first steps is the hardest part. Steven Totman, Cloudera's financial services industry expert, put it best when he said: "When it comes to challenges with big data, I think first thing is just where to start. There are so many different potential use-cases—you can use it for risk, you can use it for fraud, you can use it for customer 360°—and we actually see customers struggle with that. We usually ask them, 'What data have you not had access to? What would you like to have that you've never seen before?'" Then, Totman says, he's able to quickly show his customers the data they've been missing by using Trifacta's visual interface. And they get to work.

In your own organization, ask those questions, leverage the right technology, and don't be scared off by your "lack" of technical resources. Launching a new initiative is exciting, especially when you're focused and committed. And, it's even more exciting when that hard work pays off.

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.