



INTERNATIONAL
INSTITUTE FOR
ANALYTICS™

RESEARCH BRIEF
RESEARCH & ADVISORY NETWORK



COMPLIMENTARY
EDITION

Operationalizing Analytics with Modern Analytics Platforms

BILL FRANKS

APRIL 2019

Sponsored by

alteryx | The Thrill
of Solving



THE BIG IDEAS

- Simple, interactive, visual interfaces to orchestrate analytics processing
- Empower experts to code in popular languages
- Enable less sophisticated users to skip coding with pre-packaged functionality
- Collaboration tools to share more effectively and modernize protocols
- Integrated governance and process management helps lower risk and add transparency
- Push button deployment with no re-coding simplifies and enables the rapid operationalizing of results
- Maximize impact of analytics through easy integration with applications and business processes
- Capitalize on the full range of platform capabilities and not pigeon hole as only handling certain functionality

Introduction

The analytics landscape has matured, broadened, and totally transformed in recent years. This transformation has been driven by a wide range of factors including big data, the cloud, the increase in processing capability, the decrease in storage and processing costs, and the near-ubiquitous availability of algorithms. As a result, **an organization must also totally transform the tools and processes it uses for analytics if it is to recognize its analytics potential.**

Most organizations have embraced new data, new data platforms, and new processing environments. However, **upgrades to analytics software have lagged behind.** Many organizations still use legacy analytics software, many of which now include updated functionality to handle the new data and platform environments. In some cases, this may be enough. And, large organizations will likely make use of legacy analytics packages for years to come. However, **it is critical that organizations also begin**

to take advantage of modern analytics platforms that have been built from the ground up to take advantage of today's full range of capabilities.

This paper will explain what makes these modern platforms different from those of the past and how those differences add value in today's business environment.

- **Moving beyond purely code-based environments** and toward workflows that are easy-to-use, visually appealing, and intuitive
- **Expanding analytic creators and consumers**
- **Enabling collaboration** in ways never before present within analytics packages
- **Easing the deployment effort** to operationalize and maintain analytics processes
- **Simplifying the integration of analytics** with applications and other business processes both simpler and safer



The purpose of this paper isn't to recommend any specific platform or technology. Rather, it is to make the case that organizations need to look at the new breed of platform now available in the marketplace. Once the decision to explore the space is made, then the comparison of the various offers in the marketplace can take place to identify which best meets the needs of a given organization.

Perhaps the biggest driver of this paper is the fact that **many organizations that own an analytics platform don't use it to its potential. Often, whatever purpose drives the initial purchase is where the platform gets pigeon-holed.** As a result, some companies view modern analytics platforms as data preparation tools, some as statistical modeling tools, some as geospatial tools, and some as an orchestration tool. In all cases, the organizations would be better off if they understood the full scope of the platform they own and made use of the platform's full capabilities.

What Drives the Need for Modern Analytics Platforms?

Change happens. Virtually every aspect of how business is done has been transformed over the past few decades. It has been barely 30 years since computers became truly ubiquitous in the workplace. The Internet has been a substantive factor for less than 20 years. Big data is less than 10 years old. The cloud has had a serious impact for less than 5 years. In short, it is hard to find anyone making use of the same tools in the same way as they did in the past to do their job today.

Given these huge, change-driving trends, it is reasonable to expect the methods of generating and consuming analytics to change substantially as well. However, **many organizations are still heavily dependent upon software and protocols that carry over many aspects of what was needed in the past without accounting for all of what is needed today.**

Organizations are now trying to implement more analytics processes and at a faster pace than ever before. This puts a strain on the analytics team, which must be as efficient and effective as possible. At the same time, traditional deployment processes are long, painful, and expensive. In fact, one of the biggest barriers to realizing value is the inability to effectively deploy analytics processes. **Combined, these issues severely dampen the impact of analytics for organizations today. Changes are required!**

It is hard to fully understand what is possible today without experiencing it firsthand. Think about all of the tools that smartphones have replaced. Smartphones have, for most people, completely replaced cameras, Rolodexes, traditional telephones, and more. On top of that, smartphones have provided new, now indispensable functionality that people didn't realize they needed until they saw it. Examples of this include GPS functionality, mobile browsing, and highly functional mobile applications. Instead of a myriad of best of breed products, consumers use a single, integrated smartphone to do it all.

Smartphones consolidated, integrated, and improved upon a range of traditional tools within a single platform. Even better, they also incorporated new, never-before-seen functionality that drove their value even higher. This is analogous to what modern analytics platforms have done. **Modern analytics platforms have not only consolidated and integrated most of what analytics organizations made use of in the past, but they've incorporated additional and highly valuable functionality that goes beyond the past.** Just like a smartphone, these platforms enable an organization to both replace best of breed approaches, but also expand capabilities at the same time.

It is important to keep in mind that a consolidated toolset can have its challenges too. Any given component may not be as strong as a best of breed



option. It is necessary to consider the tradeoffs between the benefits of an integrated tool and the potentially superior functionality of best of breed options.

Any organization still stuck in a classic, code-based analytics environment will fall hopelessly behind. Similarly, the old, expensive, and risk-laden approach of having an IT organization recode a working process into a different language and environment as part of placing it into production must go. **Today, it is possible to address the entire spectrum of analytics from data acquisition, to analysis and modeling, to deployment and operationalization with one platform.** This lowers cost and risk while simplifying and streamlining efforts. It is a win for the business team, the analytics and data science organization, and IT.

The next sections will dive into some of the specifics of why modern platforms are so powerful and why every organization should consider one.

Ease of Use

Even for those who love to code, programming environments can be difficult to use. For those who are not technically adept, such environments are virtually impossible to navigate. Both experts and novices can benefit from tools designed to be intuitive and visual.

Today's modern analytics platforms enable users to lay out an end-to-end analytics process in a visual palette that allows anyone to quickly see what is happening, in what order, and what comes next. Instead of having to read a separate set of documentation or opening each piece of code to see what it pulls in and pushes out, users can see the process laid out in front of them visually.

Understanding the flow of a complex process is difficult complex coding language describe the flow. However, it is quite simple to comprehend when laid out in a workflow palette.

Once a workflow is laid out, it is also easier to use it again and again. Repeatability has always been important, and it is even easier to build repeatable processes when they are intuitively laid out for others to see and integrate into their own processes.

Ease of use also implies intuitive user interfaces that make finding what is needed easy and logical. Ideally, this will be a browser-based interface so that people can access their analytics processes from anywhere. The easier a tool is to use, the more people will make use of it. Given the need for broad deployment of analytics across an organization, it is important to enable more people than just the experts to get involved in the process of creating and operationalizing analytics.

Of course, there is a balance between enabling more users and keeping them safe. Making it easy to create a process and execute it does not mean that it is easy to ensure that the process is accurate and appropriate. Expert users such as data scientists may have virtually unfettered access to data and algorithms. Front line employees may only be able to execute pre-approved processes with tight bounds on the parameters that can be changed. In any case, it will be critically important to have processes in place to ensure that the platform is not used inappropriately.

Finally, a breadth of built-in "widgets" that perform specific, common tasks adds much to ease of use. Simply dropping an icon onto the palette and setting the parameters appropriately is fast. **Having such widgets means that moderately skilled users can create fairly complex processes safely.** It is also easy for an expert to validate what the process is doing by examining the visual workflow and widget settings. What's more, additional widgets can be added by users in order to expand functionality. So, an organization's capabilities will grow over time.



Flexibility

One of the nicest benefits of the new breed of analytics platform is the level of flexibility they provide. The platforms don't attempt to embed all functionality directly, as they focus on enabling access to all required functionality in one way or another. This means lots of formal integration with other tools to enable users to interact with their organization's entire analytics ecosystem. Some examples to illustrate this will be provided later in the brief.

One type of functionality that will make both experts and novices alike happy is the ability to use either pre-built widgets or to embed code you've written yourself directly within the workflow. **It is a "both / and" not an "either/or" with respect to coding and pre-built templates.** This is important because it allows people who love to code to continue to do so while still providing them all of the benefits of the visual flow and other features discussed in this brief. At the same time, those who don't want to code (or can't) can stick to the widgets.

There are inevitable cases where something complex requires a bit of custom code. Being able to simply drop that code into the visual analytics flow quickly and transparently blends the best of both worlds together. If problems occur with a raw code widget, of course, it will be necessary to dig into the code. But, at least it will be clear where in the process the problem is and exactly what code is causing the problem.

Not only is custom coding supported, but there are usually multiple languages supported as well. **This lets users make use of whatever their preferred language is.** While also allowing the mixing and matching of different languages within the same analytics workflow. Of course, to use different languages, it is necessary to have each piece of code handle specific steps in the process. Rather than choosing only one of the common languages such as

SQL, R, or Python users can choose the best for each situation. This is distinctly different from traditional platforms where often only one language was supported. At the same time, introducing more languages does introduce more complexity. So, it is best to stick to the minimum actually required for a given project. Having a team's approach as standardized as possible is best.

The level of integration with other tools and platforms is also impressive compared to times past. In general, it has become much easier to integrate different toolsets as much industry attention has been given to simplifying and documenting protocols to do so.

Today's analytics platforms come out of the box with direct integration with common programming languages and common toolsets. Languages like SQL, R, and Python are a snap to use. Similarly, tying into common visualization tools, data platforms, or reporting tools comes out of the box.

To the extent that it is necessary to interact with a tool not officially integrated, the capability to define a connection is present as long as the tool follows standard industry protocols. Providers of analytics platforms are continually increasing the portfolio of out of the box integrations. These integrations cover both on-premise and cloud-based deployments.

The analytics platform integrates with different tools and data sources and unifies the process of data input to analytic output into a single, visual workflow.

Collaboration & Sharing

Perhaps the area where modern analytics platforms stand out most compared to historical standards is in the way they build in robust sharing and collaboration functionality. It used to be that most analytics professionals would write custom code and save it to their hard drive. Possibly, the code files were saved to



a shared network drive where others could theoretically see it. However, the reality was that nobody ever looked at anyone else's code. People would do what they needed themselves. In a pinch, someone would ask a peer if they had code that could be cut and pasted into their work.

Today's analytic solutions keep all code and metadata in a central location that is easily accessible to those with permission to access it. The repository also tracks changes and versions over time. In addition to the raw code and process settings, the platforms allow users to embed detailed documentation at any point. Documentation can include anything from business definitions, to data caveats, to commentary on why a given approach was utilized.

The platforms also enable social functionality so that one user can post a question about a specific process and the owner and/or community that supports the process will receive the question and be able to respond. Being able to interact with others quickly and within the context of the question allows for improved productivity, less risk of errors, and a higher level of confidence within the community that analytics processes are sound.

Of course, having the functionality available is only the starting point. Without a change to the team's culture and processes, no benefit will be realized. It takes work to document properly, to answer questions, and to develop a shared sense of community and ownership. It is very easy to slip into historical habits with new analytic solutions. The platform is an enabler, but not an automatic solution, with regard to addressing these issues.

While a platform might enable end-to-end functionality from ETL to deployment, not all users will be comfortable navigating each of those steps. Many processes will require input at various times from any number of people serving roles such as:

- Line of business owners who help define the business problem and establish the goals and purpose of a project
- Data stewards who own the data itself and ensure that it is accurate and that it is being used properly
- Data engineers and data analysts to prepare any data required for an analysis
- Data scientists and analytics professionals who will drive the analytics components
- IT and DevOps experts who will manage the deployed process

Each person can add his or her input to the overall process within the same shared workflow so that all can see the progress. At the same time, everyone will handle only the pieces of the flow that their expertise dictates. **As long as each person receives the input expected and produces the output expected, it is not necessary for everyone to understand the full details of each other's work.** In the end, however, the full process will be transparent and available to review.

Once a process is complete, it can be shared and made available to any other person or process that might need to access it. **Either the detailed workflow can be made available so that it can be examined and edited, or the workflow can be packaged such that it can only be provided the proper inputs and executed as-is.** Thus, any type of user or application can benefit from the same process at the level that makes sense for the situation. This level of flexibility enables broad deployment and the scaling of usage.

Governance and Process Management

While the ability to share workflows is usually thought of as an enabler of collaboration, it also enables better governance. The more people and processes that are making use of the exact same



workflow, the easier it is to ensure consistency, maintain quality, and lower risk.

When everyone is building their own processes, it is easy for divergence to arise. Multiple processes may purport to provide the same information, yet the answers each process provides are actually different. This can be due to differing definitions, the usage of different raw data sources, among other causes. Regardless of the cause, it is not a good thing to have a question inconsistently answered. Especially when operationalizing results, it is critical that results can be trusted.

By sharing workflows, it is easier to track the lineage of any output data or results and the logic used to generate them. Having the processes handled by a shared server also avoids the problem of “it worked on my computer” where local setting differences can cause the same process that works on one computer to run into problems on another.

It is important to understand that the sharing functionality of modern analytics platforms goes beyond the simple sharing of code. Rather an entire process can be wrapped into a callable process that other users or applications can access easily. It is also possible to wrap up just a portion of a process and make it available. **This enables an ever-growing library of widgets that have been tested, documented, and made available to speed development of new processes.** The key here is to ensure any shared processes are in fact tested, documented, and ready for prime time. Without proper governance, an analytics platform can lead to a lot of sloppy and inaccurate processes being easily distributed.

Another aspect of modern analytics platforms that greatly aid governance is the use of code management. **Utilizing version control to track and log process changes allows transparency in how a process evolved and also makes it easy to roll back**

changes that are deemed negative after they are tested.

The platforms also integrate notebook-style functionality, either through Jupyter Notebooks integration or providing similar functionality. This provides power users who do code with access to all of the benefits of an integrated notebook approach. The span and benefits of notebooks is outside the scope of this brief.

Streamlined Deployment

The deployment and operationalization of an analytics process has always been an area where analytics has fallen short in large organizations. Many analytics processes are never deployed in an official manner, but rather maintained and executed on a regular basis by the creator of the process in a pseudo-production mode. Those processes that do get deployed often require substantial modification and re-coding in order to enable roll out on platforms that are different from those where the process was created. This introduces huge costs, more room for error, and a timeline that is often unacceptable.

By enabling the sharing of processes, modern analytics platforms substantially remove many of these historical hurdles. **With the push of a button, a new process can be registered and made available on-premises or in the cloud.** Role-based permissions can also be configured to govern who and what can access the process. The benefits of push-button deployment are hard to overstate in terms of the amount of risk, complexity, and cost that are removed from an enterprise analytics initiative.

One of the biggest benefits of an analytics platform is the ability to deploy and operationalize a process using the exact same language(s) it was created with. No more re-coding or re-engineering a process. It can be deployed exactly as built. This is a breakthrough that was made achievable with the



advancement of the underlying processing platforms. These platforms have become much more flexible and contain broader integrations with other technologies than in the past. Due to the complexity, problems can be expected from time to time, but the overall effort to deploy should be greatly reduced and there will be more consistency in the way the processes are deployed.

Security and transparency are both aided when only one version of a workflow must be tested and maintained instead of the “original development version” and the “converted production version”. When an analytics user is able to confirm that a process has been tested and proven accurate, the process will be ready to deploy without any of the classic overhead that would then follow.

One huge implication of this direct deployment of analytics processes is that it changes who should be responsible for what components of the operationalized production process. Historically, the analytics team would hand off a process to the IT team. Once the IT team oversaw the conversion effort to deploy the process, IT would then own the entire production process, including the code and analytics logic it contains.

Today, the analytics organization can (and should) maintain ownership of the underlying code and logic since that organization created it and understands it best. The IT team should continue to maintain ownership of scheduling, resource allocation, and process monitoring. If a problem occurs, IT then alerts the analytics organization so that it can diagnose and fix the problem.

Simplified Integration

Deployment of analytics today means more than just having a batch process that periodically runs and spits out data. Today, multiple other users and business processes may need to call upon a process in an on-

demand fashion. This is at the heart of operationalized analytics. Today’s platforms make use of modern application program interface (API) protocols that make it much easier for one process to call upon another. Representational State Transfer (**REST**) APIs **make calling a complex function as easy as requesting a web page, even borrowing the syntax of web navigation.**

By making complex analytics processes available with a simple API call, the ease of integration is greatly enhanced. Embedding complex analytics into business processes becomes easier than ever from a syntax point of view. More important, **developers of downstream processes don’t need to understand the guts of the process, but only what inputs are needed and what outputs will be provided.** The barriers to pervasive and operationalized analytics are tremendously reduced!

As discussed previously, the analytics organization will own the logic within the process while the IT team will own the environment, the allocation and management of containers, and the general resource management of the systems during execution. This keeps everyone working within their areas of strength and increases the scale possible within the organization in terms of the number and depth of analytics processes that can be deployed and operationalized. As also mentioned in other sections, making a process easy to call does not mean it can’t be misused. It will be critical that the time is taken to document the process and its intended uses, as well as to audit its usage over time. There will be a time commitment from needed people to ensure ongoing success.

Conclusions

There are a number of important takeaways from this brief.

First, **every organization should take the time to understand and assess the functionality of analytics platforms currently available.** The new breed of modern analytics platform outlined in this brief provides a lot of new and exciting functionality.

Second, these modern solutions have functionality spanning the entire analytics process from data acquisition to deployment. Historically, a variety of platforms and tools had to get involved in order to deploy and operationalize analytics processes. **Today, a single platform can either eliminate the need for the different tools and platforms or, at minimum, manage the process of orchestrating the use of them.** Organizations should ensure they don't pigeon hole a platform into a single niche.

Third, having a single solution that can **deploy processes as-is decreases the cost of deployment, lowers the complexity of the process, and lowers risk.** Being able to more quickly deploy complex analytics processes also increases the usage of the processes and the value that they drive.

Next, the flexibility of modern analytics platforms provides tremendous value. Being able to intermix different languages, make use of visual user interfaces as well as coding interfaces, and to provide access to a wider audience that goes beyond deep technical experts all enable an organization to be more effective with its analytics initiatives. **Such capability also future-proofs an organization by being able to adapt and integrate new tools, techniques, and platforms that will certainly arrive in the future.**

Next, the visual workflows contained within the platforms **makes understanding what is happening, when it is happening, and any upstream and downstream dependencies much, much easier than in the past.** Just as visualization tools assist in the presentation and comprehension of analytics results, so visualizing a process assists in the understanding and comprehension of that process.

Next, the integrated collaboration and governance capabilities provide functionality and value that is largely not in use today. **Modern analytics solutions provide a jumpstart toward becoming a more cohesive and integrated organization.**

Finally, **the ability to easily deploy and operationalize a process as-is while also making that process available via simple function calls is a game changer.** By making deployment of, and access to, an analytics process very simple, it is possible to deploy more analytics, to a broader set of applications and business processes, and with tighter integration.

Modern analytics platforms are disrupting the analytics marketplace. As discussed in the introduction, they are replacing and integrating multiple classical tools and approaches with one that covers it all. Beyond that, the platforms provide functionality that many organizations haven't yet brought to their analytics but should.

In this sense, the parallel to how a smartphone replaced calendars, contact systems, and more is quite appropriate. **Few people have failed to embrace the changes that smartphones can bring to their lives. Similarly, few organizations should be failing to embrace the changes that modern analytics platforms can bring to their analytics.**



ABOUT THE AUTHOR

BILL FRANKS

Bill Franks is Chief Analytics Officer for The International Institute For Analytics (IIA), where he provides perspective on trends in the analytics & big data space and helps clients understand how IIA can support their efforts to improve analytics performance. Franks is also the author of the books *Taming The Big Data Tidal Wave* and *The Analytics Revolution*. He is a sought after speaker and frequent blogger who has been ranked a [top 10 global big data influencer](#). His work, including several years as Chief Analytics Officer for Teradata (NYSE: TDC), has spanned clients in a variety of industries for companies ranging in size from Fortune 100 companies to small non-profit organizations. You can learn more at <http://www.bill-franks.com>.

About the Research Sponsor:



Alteryx is the only quick-to-implement, code-free and code-friendly, self-service data analytics platform that allows data scientists and line of business analysts alike to break the barriers to insight, so everyone can experience the thrill of getting the answer faster. The Alteryx platform makes it easy to securely access, cleanse, and prepare all the relevant data; blend it all together to prepare the perfect analytic data set; then perform sophisticated analytics, and output easily consumable analytics that drive deeper insights and improved decision making and business performance. [Learn more](#).

IIANALYTICS.COM

Copyright © 2019 International Institute for Analytics. Proprietary to subscribers. IIA research is intended for IIA members only and should not be distributed without permission from IIA. All inquiries should be directed to membership@iianalytics.com.