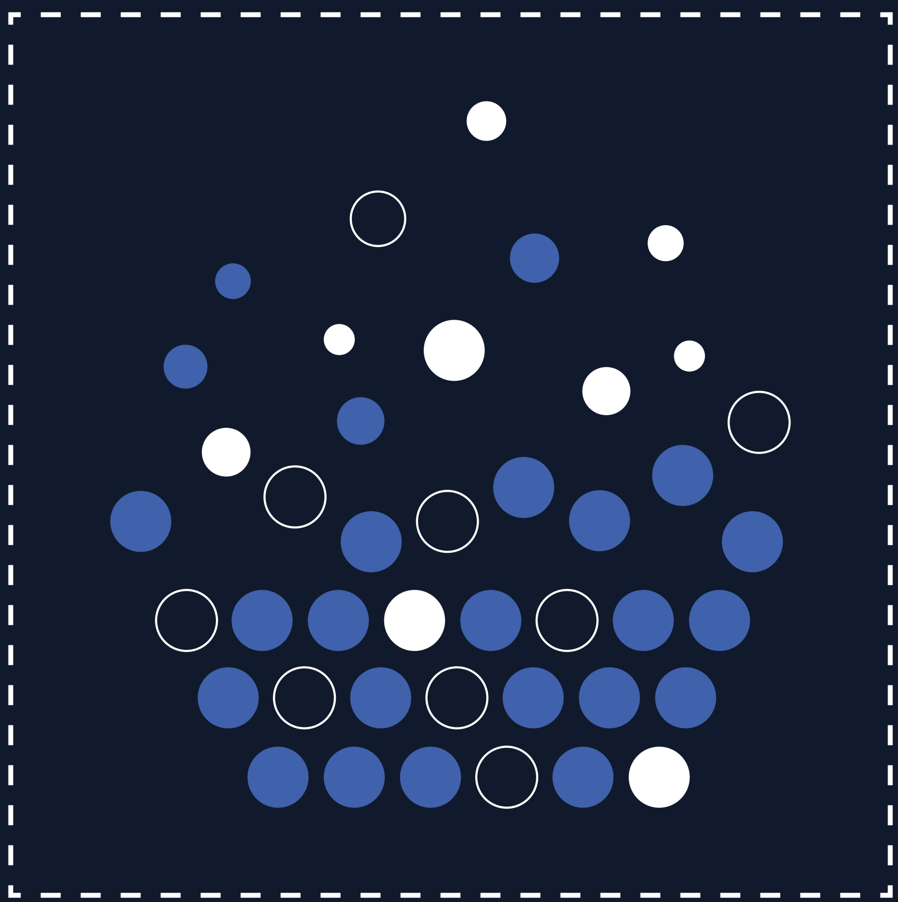
data
iku

ENABLING AI SERVICES THROUGH OPERATIONALIZATION + SELF-SERVICE ANALYTICS

And Why a Data-Powered Company Needs Both



WHITE PAPER

www.dataiku.com


Ross 128
Lalande 21185
10 light years
Wolf 359
GJ725
ε Eridani
UV Ceti
Lacaille 9352
ε Indi
τ Ceti
YZ Ceti

Introduction

You are the CXO of a company that serves as a sales platform for thousands of different clients. You and your management team have identified a list of processes that could be improved via better use of your data and advanced analytics. For instance, in order to help your clients increasesales (and thus increase the stickiness of the platform), you decide that the development team should surface a custom recommendation module offering three product recommendations per client.

Requirements include that:

- The recommendations be available in real time and
- The recommendation engine can be updated without causing platform downtime.

At the same time, there are internal requests from different teams (like sales and marketing) who want to make data-driven decisions (for example, key trending products with positive reviews from social networks or how transformation rates are influenced by the historical browsing behavior of a visitor), but their old dashboards are static and don't address their needs. Even though the data exists internally, they can't get insights for themselves because they don't have direct, regular, monitored access to data that can help them do their jobs.

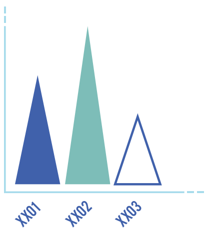
Which need should be prioritized? And how do you even begin to tackle these projects with an approach that will be sustainable and reproducible for other projects and requests down the road?

THE ANSWERS TO THESE QUESTIONS ARE NOT EASY (ANYONE - OR ANY PRODUCT - THAT CLAIMS TO MAKE IT EASY IS EITHER LYING OR GROSSLY OVERSIMPLIFYING). BUT IN ORDER TO BECOME A TRULY DATA-POWERED COMPANY, IT'S NECESSARY TO SET UP THE TOOLS, PROCESSES AND THE ORGANIZATION TO MAKE THE DEVELOPMENT AND USE OF DATA AND MODELS PERVASIVE.

INDEED, THE KEY TO DELIVERING BUSINESS VALUE FROM DATA LIES IN FOUR AREAS:

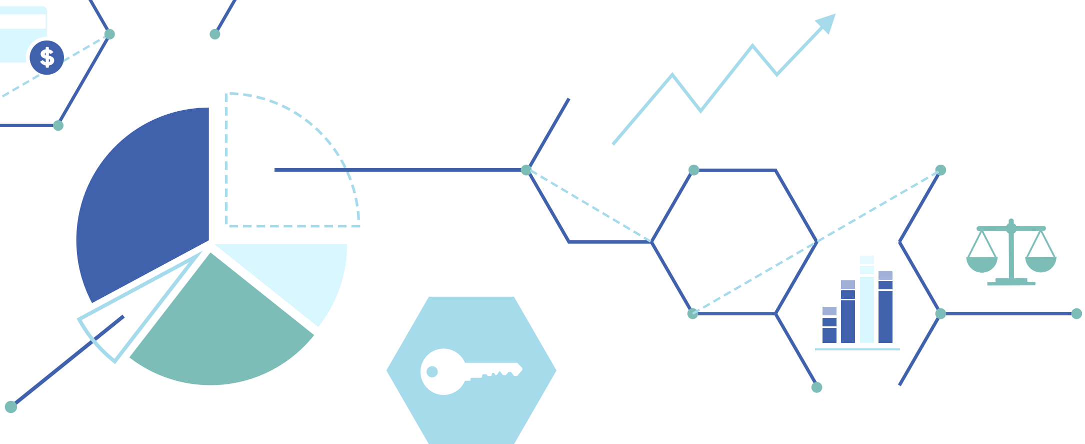
1 Aligning advanced analytics projects with actual business value (and getting those projects out of the lab and into a production environment where there is real, business impact) - also known as operationalization (o16n).

2 Supporting the need for quick answers to ad-hoc questions to augment daily, individual decisions through self-service analytics (SSA).



3 Getting beyond the limits of small data to scale these efforts from o16n as well as SSA in order to become a data-powered company (i.e, one where the development as well as use of data and models is pervasive).

4 Align o16n and SSA to benefit from the virtuous cycle of o16n pre-packaging data, results, and processes to improve SSA, and SSA both freeing up key resources and supporting identification of key value creation opportunities for o16n.



So in the previous scenario concerning the operationalized model as well as the user-driven data requests, when it comes to the question of what to prioritize, the answer is: Both.

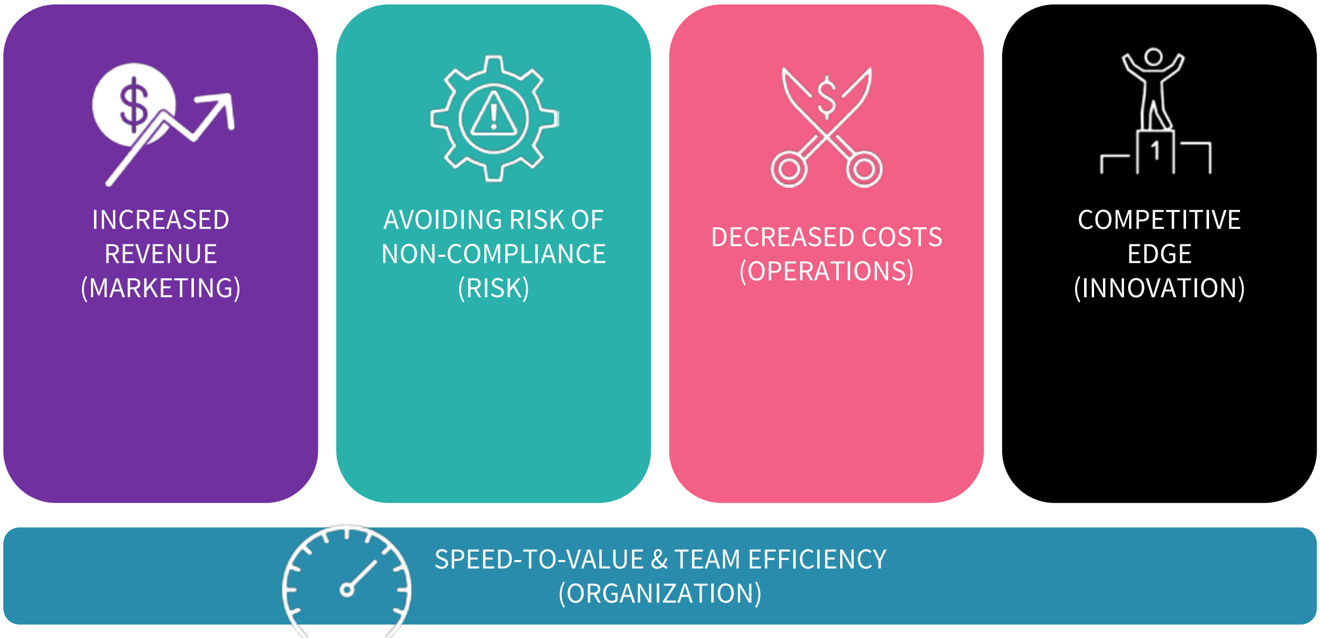
This white paper will explore why this method is effective, what each component really means, and how to get there.

Above all, it will answer the question:

How can I build a holistic data management architecture to support current and future business intelligence, advanced analytics, machine learning, and AI initiatives?

Why It All Matters

Any company that wants to make any impactful change - whether that's decreasing costs or risks, increasing revenue, creating innovative new products, or making employees and the organization more efficient overall - has the opportunity to do so using today's not-so-secret weapon: data. More specifically, the massive amounts of data available can be used to gain insights at scale vis-à-vis processes like reuse and automation.



Transformation at this level doesn't simply mean slapping data on top of existing processes; it involves fundamental organizational change, weaving data into the fabric of the company. And there are several components to success in execution, to be sure: like top-down support of the initiative from executives, fundamental company structure(s) and staffing, or available tools and architecture.

Self-Service Analytics (SSA)

sɛlf sərvəs ænəlɪtɪks | (n)

The system by which line-of-business professionals or analysts can access and work with data to generate insights - predictive or not - and data visualization with little direct support from data scientists, IT, or larger data team (though the SSA platform itself should be supported by these personas).

Operationalization (o16n)

apəreʃənələzeʃən | (n)

operationalize (v)

The process of converting data insights into actual large-scale business and operational impact. This means bridging the huge gap between the exploratory work of designing machine learning models and the industrial effort (not to mention precision) required for deployment within actual production systems and processes. The process includes, but is not limited to: testing, IT performance tuning, setting up a data monitoring strategy, and monitoring operations.

For example, a recommendation engine on a website, a fraud detection system for customers, or a real-time churn prediction model that is at the heart of a company's operations cannot just be APIs exposed from a data scientist's notebook - they require full operationalization after their initial design.

Ultimately, the shift to a data-powered organization requires two things:

1 Operationalization (o16n), enabling top-down changes through large, scalable, impactful data initiatives, combined with...

2 Self-service analytics (SSA), which rounds out the transformation with a bottom-up approach by putting increasingly sophisticated data analysis in the hands of the many.

In fact, it is the interplay and balance between the two that makes a successful data-powered company that executes on all initiatives to its fullest potential.

Self-Service Analytics to Operationalization:

The Interplay

It's not uncommon for organizations to implement only SSA and stop there. This is perhaps the result of several years back (circa 2015)[1] where industry leaders and analysts thought that business intelligence (BI) platforms were the be-all and end-all of data-driven transformation. On the other side, there are also some organizations that implement only a system for o16n, thinking that it overpowers the need for SSA. However, the most successful enterprises do both.

SSA is useful because it ensures wide coverage of an organization's data initiative. It is more bottom-up, and it puts data in the hands of everyone at the organization, empowering them to use it to drive day-to-day decisions. While organizations that rely only on SSA are agile and can answer quick, small-scale questions, they may lack the larger picture; that is, a focus on bigger business questions and problems that have opportunity to make the large-scale changes and projects driven by machine learning or AI.

The Path to the Data-Powered Organization

Becoming a data-driven business involves fundamental organizational change via self-service analytics and operationalization, which should coexist and drive each other forward.

Provide Self-Service Analytics

VALUE **Quick answers** to ad-hoc questions

SCALE **For smaller-scale** or one-off business questions; immediate insights

FOR/BY WHOM **Used regularly by business** team(s), but built and maintained behind the scenes by data and/or IT team

DRIVEN BY **Intimate business knowledge** *(i.e., of the specific problems or roadblocks the business is facing and how data can help solve them)*

IMPACT **Frees resources to work on more high-gain opportunities** *(IT/data teams don't constantly have to pull data or build one-off reports and dashboards)*

Execute on Model Operationalization

Keras TensorFlow

VALUE **Robust machine** learning solutions driving business value

SCALE For **larger** and more complex business problems; longer-term execution and vision

FOR/BY WHOM For **larger** and more complex business problems; longer-term execution and vision

DRIVEN BY **Technical knowledge** *(i.e., of how data systems and projects work and can be optimized and automated)*

hadoop R python

IMPACT **Frees resources to work on more high-gain opportunities** *(self-designed data products providing business value can easily be operationalized to prevent repetitive work on the business side)*

Foundation

Data science, machine learning, and AI (whether driven by self-service analytics or operationalization) cannot happen in a central team that exists in a vacuum without deep understanding of the business.

data iku

> ***“Self-service analytics and BI users will produce more analysis than data scientists will by 2019.”***

Gartner Press Release, January - 2018 [2]

By contrast, o16n is what ensures depth. It is more top-down - it’s all about answering large, strategic business questions and needs to make sweeping change through large-scale data projects. Operationalization can also mean the automation of projects previously addressed with SSA, moving them away from ad-hoc status to save time and resources when that makes business sense.

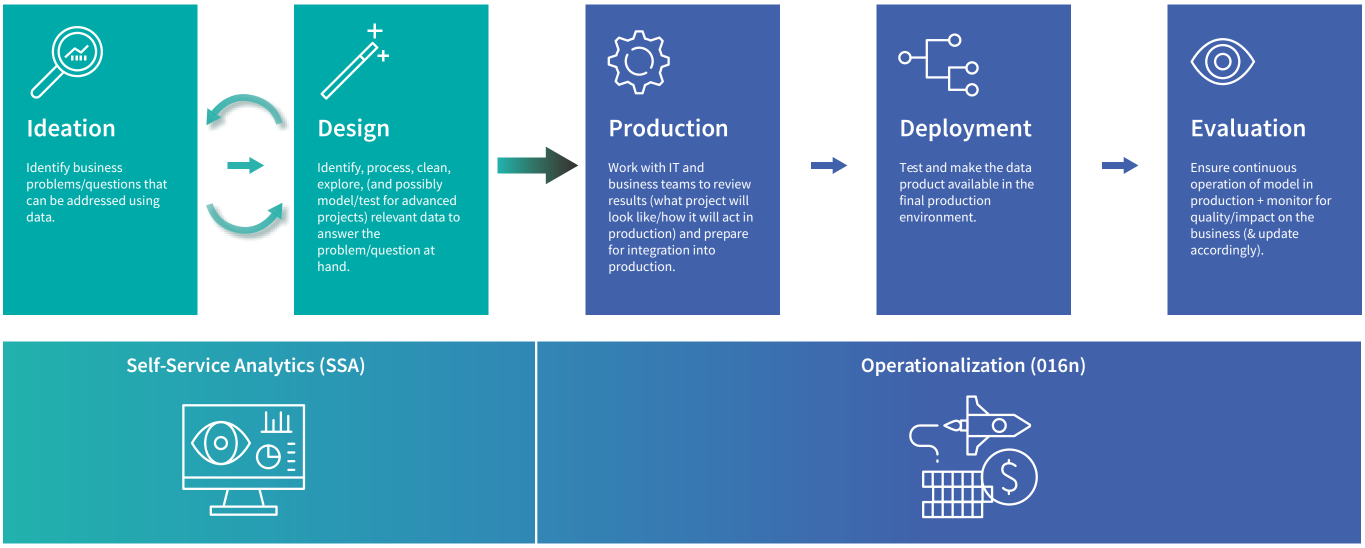
> ***“Architectures that support deploying diverse analytics into production drive user adoption. Automated deployment architectures that provision analytics into operations or embed them into business processes are emerging as a top priority for technical professionals.”***

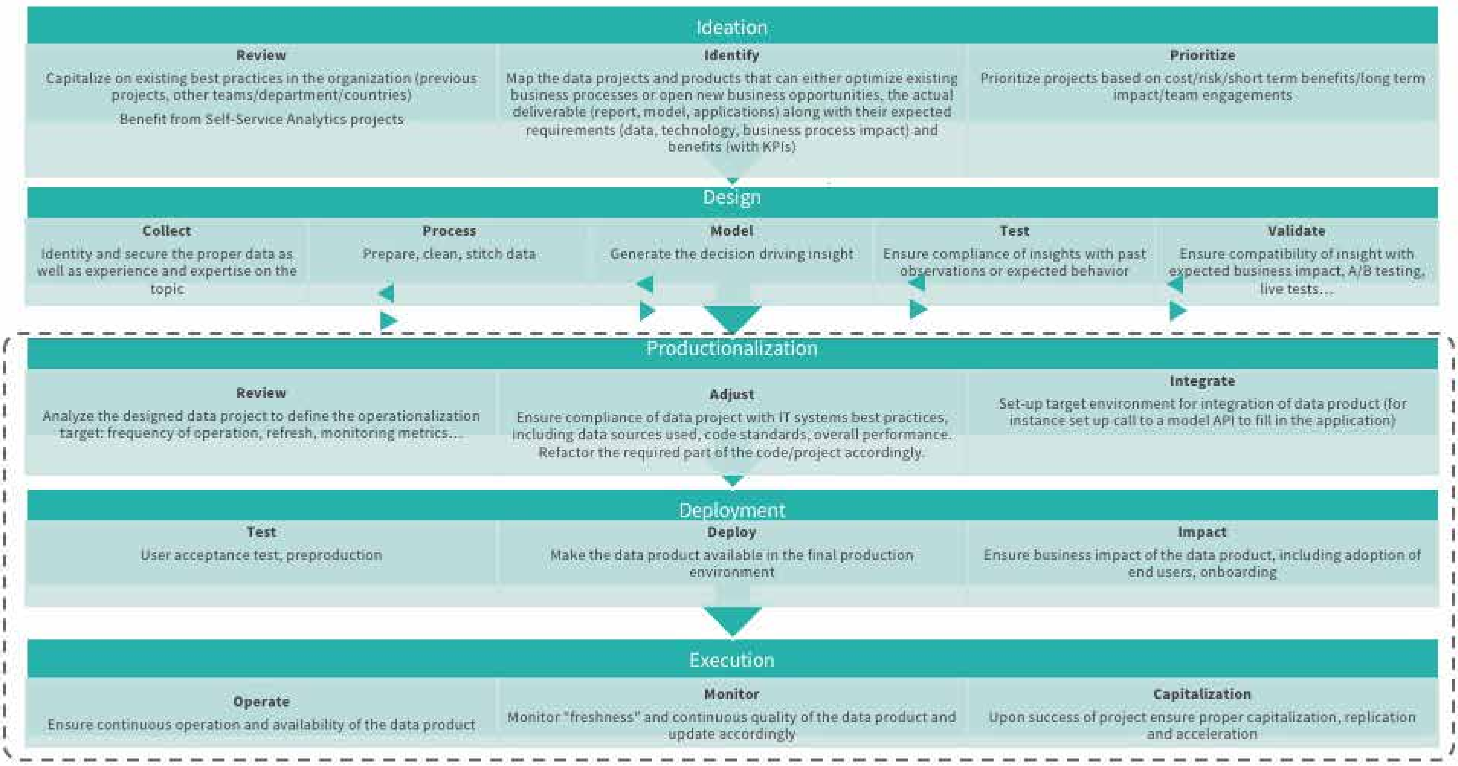
Gartner, Solution Path for Implementing a Comprehensive Architecture for Data and Analytics Strategies, June - 2018

Despite their seeming opposition, SSA and o16n actually drive each other forward and work best when using a common platform for data governance and communication, built on a philosophy of breaking down silos to work with data across teams throughout the enterprise. Many organizations mistakenly see the two as a progression; that is, that one must start with SSA and later progress to o16n. Or worse, that SSA is sufficient to become a data-driven enterprise.

In fact, SSA and o16n can (and should) coexist, each scaled at an appropriate rate with the size of teams, the company, and its initiatives. But incorporating both SSA and o16n into a larger data strategy can be challenging because the two have such different uses, goals, drivers, and audiences. The next sections will focus on what it takes to bring each to life in an organization.

From Self-Service Analytics to Operationalization:
Data Project Stages & Management





Self-Service Analytics:
Avoid the Pitfalls

In order to successfully execute on SSA in the enterprise, it's first valuable to understand why efforts to implement self-service analytics often fail. SSA can miss expectations in terms of its utility (ultimately becoming an expensive project that provides little value) when:

1. **Data access is an ongoing problem.** Often, data exists and has been centralized in a data lake, but the vast majority of individuals cannot easily access it because they lack the technical skill or the proper access rights. On top of this, SSA often requires opportunistic use of data from a local database or even a spreadsheet, which can be challenging to cross-reference or use along with the centralized data. Implementing a solution that doesn't rectify this issue can render a SSA system useless.

2. **SSA tools and processes are not tailored to the users' daily needs.** When SSA is not built to suit its users, those users will often be faced with a system that requires them to go through repetitive or complex steps to perform the analysis suited to their own job and responsibilities. This can equate to users who don't support the use of the system and who also might use it less and less if it's not convenient for their workflows, slowing (or halting) adoption.

3. **There is no confidence in the data.** By far the number one reason that SSA efforts fall flat is because of a lack of trust in the data. This could be because it's not updated regularly enough (ideally, it should be in real time), or because data and data formats change, but the company hasn't devoted ongoing resources to SSA platform maintenance. Or it might be because business users have access to datasets with no context on what exactly the data is - that is, where it comes from or what it means.

   The problem quickly balloons when managers or executives don't trust the results of projects done using SSA because of data quality or context issues. If SSA systems are not properly maintained and monitored, ensuring the data that business teams are using is validated, contextualized, and in the proper format, it won't be of any use to the enterprise. It's important not to confuse self-service with self-sufficiency - SSA cannot be built once and then forgotten.

4. **There is confidence in the data ... but there shouldn't be.** Perhaps worse than no data confidence is a situation when SSA users have confidence in the data given to them, use it to create projects and deliver insights, but in fact, the data is not correct. It's critical that IT teams validate (and continue to monitor) data being delivered through an SSA platform to ensure that it is accurate. The onus falls also on the SSA platform users to ensure they understand the data they are using and ask questions about any doubts in quality or accuracy.

5. **Data security suffers.** Partially related to data confidence are issues of data security. Often, SSA solutions don't provide a centralized (virtual) workspace that allows for control over who can access what datasets and that prevents datasets from being downloaded and manipulated on employees' local machines. This matters because ultimately, SSA projects can be shut down if they don't have a solution for controlled, monitored, access.

6. **SSA is too centered around small data.** SSA solutions that are built for small data are doomed to failure. Business users around the company who want to use SSA should have the ability to easily work with large datasets in meaningful ways, perhaps even by applying machine learning models or doing predictive analytics projects. Too often, SSA is limited to simple dashboarding without the ability to combine datasets to extract more meaningful insights.

7. **SSA is completely disconnected from o16n.** As previously mentioned, SSA is a good first step to becoming a data-powered enterprise, but it can only be a minor success without the second piece of the puzzle (more on that later). Initiatives not connected to o16n - or without a longer-term plan of introducing o16n - will fail to make a massive impact in the organization because they cannot scale. Being able to automate and deploy solutions that started out in the realm of SSA where it makes sense means more time for meaningful analysis.

Self-Service Analytics:

Keys for Execution

In order to be successful in self-service analytics, enterprises need to implement a system that aims to smooth the path (i.e., reduce delay, false-starts, and roadblocks) between data and insights, shortening the time to go from business question to data-supported answer.

Trust is a key component to successful SSA execution in a few different ways:

- The enterprise needs to trust employees' ability to use data in a self-service context (see section: **The Role of Data Governance**).
- People working on SSA need to trust the data that they're working with. That means there has to be someone continually responsible for its quality, making sure it's regularly updated, formatted, and being used appropriately (see section: **Organizational Structure around SSA and O16N** for more detail).
- Managers and executives need to trust the insights delivered from SSA projects. In theory, if they trust the employees' use of the data and trust the data itself, this should be a given. But in 2016,[3] a study of more than 2,000 data and analytics decision makers in 10 countries by KMPG revealed that only 38 percent of respondents have a high level of confidence in their customer insights, and only one third trust the analytics they generate from their business operations. So clearly, this point is still worth mentioning.

In addition to establishing trust, on a tactical level, SSA execution requires a setup that allows a wide audience that may or may not have specific knowledge (e.g., coding or other technical ability) to work easily with big data in various formats. And the wide audience also needs to be able to present their results easily as well, whether that's with a specific predictive output or more visual representations (e.g., dashboards).

This can be achieved through a platform that sits on top of data storage architecture and that combines data access, cleaning, exploration, analysis, and visualization, all accessible through a visual interface that means non-technical users can work with huge datasets on data projects easily.

All of these factors point to the need to create a single self-service experience for all teams at the company whenever possible (note that this may not be possible for extremely large enterprises with data teams of hundreds of people that report to specific business units).

But in general, some centralization is beneficial in terms of maintenance costs - no one wants to have to maintain several SSA systems. It's also beneficial in terms of cutting the time from question to answer. If everyone is speaking the same language and using a common platform, the layer of trust is easier to achieve, and presenting results becomes smoother throughout the organization.


N
1 567 Km
15 721 Km
Tóquio
Roma
Cartagena das Indias
Washington DC
4 266 Km
9 913 Km
Goa
7 053 Km
Rio de Janeiro
Moscovo
5 044 Km
Macau
12 123 Km
2 013 Km
Madrid
Londres
2 542 Km
Melbourne
2 810 Km
Pequim
10 539 Km

Operationalization:

Avoid the Pitfalls

Like SSA, it's helpful before diving in to the execution of o16n to learn from the mistakes of others. Generally, operationalization efforts fail when:

1. **The process is too slow.** The reality around most data projects is that they don't bring real value to the business until they're in a production environment. Therefore, if this process isn't happening quickly enough - both in terms of total overall start-to-finish time-to-insights as well as the ability to rapidly iterate once something's in production - o16n efforts will fall flat. In addition to having strong communication between teams responsible for o16n, proper tools that allow for quick, painless incorporation of machine learning models in production are the key to a scalable process. Speed is also of the essence in that feedback from models in production should be delivering timely results to those who need it. For example, if the data team is working with the marketing team to operationalize churn prediction and prevention emails, the marketing team should have immediate insight into whether the churn prevention emails sent to predicted churners are actually working, or if they should re-evaluate the message or the targeted audience.

2. **Lines of business are not involved in the process.** Operationalization happening in a vacuum without any input from business teams is doomed to failure, as projects tend to get delivered that don't address real needs, or do so superficially. Even when data team members are embedded in the lines of business, they need to work closely together every step of the way to ensure alignment of project goals with business realities. Similarly, o16n should interplay with SSA in the sense that projects started with SSA might be good candidates for operationalization. When the teams don't work together, these opportunities can be easily missed.

3. **There is a lack of alignment between data science, engineering, and/or traditional IT.** Data projects are beasts of their own when it comes to integrating into a production environment. On the data science side, they need to provide reliable, business-validated predictions or outcomes not only at first, but also over time. On the data engineering side, they leverage advanced data processing systems that need to be tuned and monitored. On the traditional IT side, they need to abide by all of the requirements of enterprise-ready, robust IT systems. A lack of coverage or alignment between these three dimensions quickly leads - at best - to the inability to get data innovation up and running, and at worst to actual in-production failures.

4. **There is a lack of follow-up and iteration.** The point of rapid operationalization is to get models out of a sandbox and into production quickly in order to evaluate their impact on real business processes. If models are operationalized and then forgotten, they could - over time - have adverse effects on the business. Instead, constant monitoring, tweaking, and follow-up is the key.

Operationalization:

Keys for Execution

Logistically speaking, operationalization is often more difficult for enterprises to execute on (especially compared to SSA) because it requires coordination, collaboration, and change not just at the organizational level, but often at the system architecture and deployment/IT levels as well.

But it's truly the final and most important step of the process, as data projects are rendered incomplete without being operationalized - that is, incorporated into the fabric of the business to see monetary results that have real impact. So what are the steps to smooth execution?

The first, and perhaps most overlooked, step is simply learning from, listening to, and working with business teams to ensure that the solution that will be operationalized is a viable one. While this sounds like a no-brainer, it's often overlooked, resulting in lots of lost time and effort from data teams in building a solution that doesn't actually provide any business value in the end.

Data teams work hard to ensure that their projects meet business needs, but the reality is that they often lack the context or business knowledge necessary to build the most optimal solution to a specific problem.

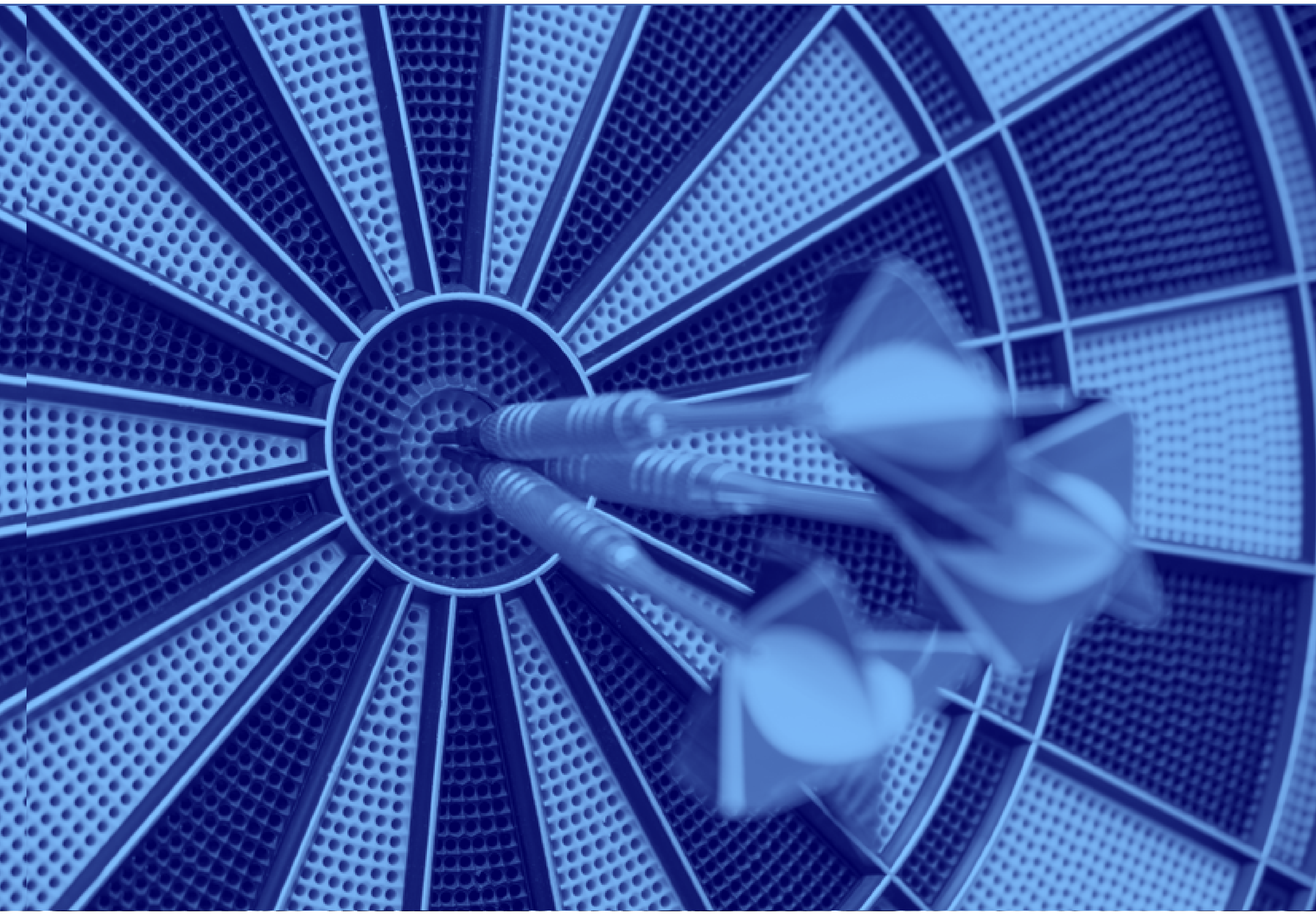
For example, say that the data science team for a website offering personalized loans is tasked with creating a more sophisticated fraud detection model. If they set out to create this on their own without speaking to the team currently handling fraudulent claims, they might create a model that is perfectly reasonable from a technical perspective, but doesn't answer the original business problem - like, for example, the fact that too many people are being caught as potential fraud and going to manual review, overwhelming the operations teams. With this contextual knowledge, the data team can tune the model to better balance it to the reality on the ground.

From a methodology standpoint, o16n requires a consistent, efficient process for integration, testing, deployment, and then measuring impact and monitoring performance (followed by, of course, making necessary modifications and integrating, testing, deploying, etc., those modifications). Inconsistent packaging and release can lead to a subtle degradation of a model's performance between development and production.

Traditionally, it's the data engineering or IT team that is responsible for refactoring of the data product to match target IT ecosystems requirements (including performance and security). However, this handoff between data team and IT or data engineering teams can be significantly eased when the two are working with the same tools and are aligned on project goals - so again, communication (even between technical teams) is key.

Following release, it is critical to implement an efficient strategy for the retraining and updating of models. Implementing a retrain-in-production methodology is a key to o16n success; without it, retraining a model becomes an actual deploy-to-production task, with the result requiring significant manpower and a loss of agility.

Additionally, a successful o16n strategy involves functional monitoring, which is used to convey the model's performance to the business sponsors, owners, or stakeholders. This provides an opportunity to demonstrate the end-results of the model in production for evaluation. And going hand-in-hand with functional monitoring is a having a viable rollback strategy in case something goes wrong. For more detail on successful o16n strategy, read Data Science Operationalization: Finding the Common Ground in 10 Steps [4] .



Optimal Organizational Structure Around SSA and O16N

It should be clear by now that self-service analytics and operationalization are not simply one-time projects, but an investments that involves the devotion of plenty of resources (both technical and personnel). So how can companies structure their organizations in a way that supports these initiatives?

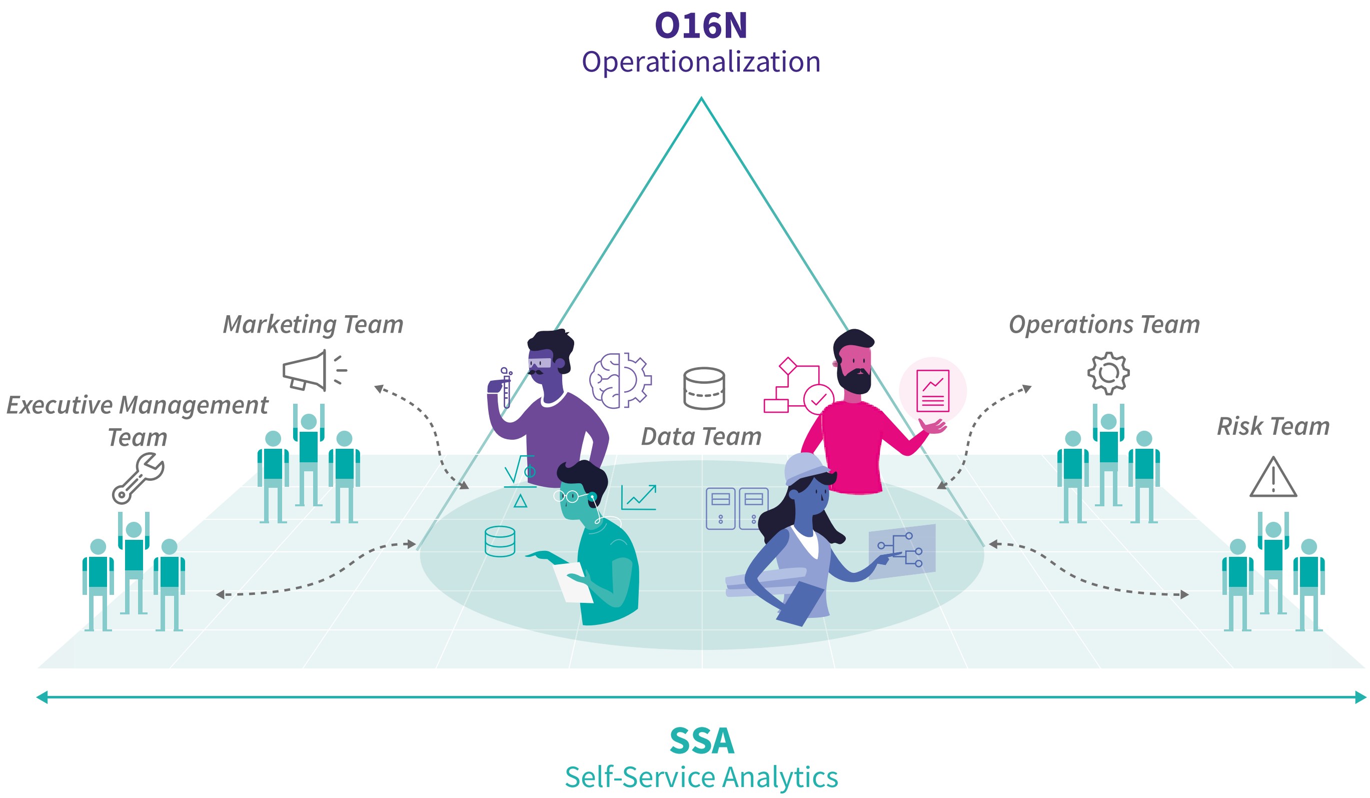
The most efficient model for the deployment of effective SSA and o16n systems is rooted in the fact that data science needs are intertwined with business needs and can arise from any line of business or department. That is, data initiatives - whether driven by SSA or o16n - cannot happen in a central team that exists in a vacuum without a deep understanding of the business (or connections to those who have that knowledge). For example, it would be impossible for a data team to execute successfully on a customer churn prediction and prevention project without input from teams like marketing and/or sales.

Instead, SSA and o16n feed off of each other and thrive together in organizations that establish a central data team as a center of excellence, a sort of internal consultant that can be deployed to activate data efforts across the company through a combination of SSA and o16n. The is the model that Pfizer,[5] Daimler [6], and more leading companies are increasingly turning to in order to fully transform into the era of data science, machine learning, and - increasingly - AI.

In practice, this kind of organizational model means establishing:

- ☐ A platform for data access, discovery/exploration, visualization/dashboarding, as well as for machine learning modeling and deployment into production that can be the basis of both a thriving SSA and o16n environment. This platform should ensure that everyone across the organization (regardless of his or her technical skill set) is working with data that can be trusted and that they can produce desired outcomes - whether that means dashboards, a predictive model, etc.
- ☐ A centralized - and, importantly, not siloed - data team or organization that maintains said platform, ensuring that all data is accessible, accurate, and generally usable in a SSA context. This team would also be responsible for larger deployment and o16n efforts based on SSA projects or other data projects executed along with business units.
- ☐ A means of collaboration and communication between business units and data team(s) so that any questions arising from data projects (whether via SSA or o16n) can be easily addressed in context - especially questions surrounding where data comes from, what it means, and how it can be accurately used in projects. This means of collaboration should also ensure that any larger data project produced with SSA is validated by the data team to ensure.
- ☐ Feedback loops connecting operationalized data projects to business objectives and ensure they continue to meet those objectives (and can be easily adjusted if not).

The data-team-as-internal-consultant model might look something like this, with the SSA component spanning horizontally and the o16n component arising vertically out of those results and relationships with different teams and lines of business:



Note that in this model, all types of roles from across the organization are involved with different stages of data projects. In this model, data scientists - while they may be centralized - are not the only people to provide value or insights. Instead, they act as specialists (or perhaps enablers and catalysts). Meanwhile, business users, analysts, data engineers/IT, and more provide the day-to-day fuel of ideas, data preparation work, and more that allow the data-powered organization to run continuously without depending on a specific person, role, or team to keep moving.

	Business owners	Data Stewards	Data analysts	Data scientists	Data engineers	Applicative IT environment owners/devops
Ideation	Ideation led through specific workshops involving the broader community: business stakeholders from data driven and non data driven business areas, variety of data stakeholders, executives. Specific workshops formats help support ideation and prioritization process					
Design	Anticipate and prepare business process impact Review process and results of data product, validate and enrich with business know-how	Support identification of relevant data and data cleansing/processing effort, ensure compliance with data usage guidelines	Design data product with the support of the wider data experts	Support data analyst with advanced know-how and tools packaging		
Productionalization		Ensure proper availability and readiness of data required in production environment, identify new data sources bringing business value.	Document project	Define monitoring metrics, warning strategies and scenarios	Refactoring of the data product to match target IT ecosystems requirements (including performance, security)	Prepare integration within target IT ecosystem environment
Deployment	Prepare the business/operation teams to the use/changes of the new data project	Provide data sources homogenous to				Test in various pre production environments and deploy in production environment
Execution	Leverage new data product and monitor/validate KPI		Monitor performance of data project over time and follow up on models updating when required, or kick off redesign if needed	Package plugins/macros		Monitor continued execution

The Role of Data Governance

With today's increasing concerns around data privacy, the role of data governance in the realm of SSA and o16n efforts is an important one that is worth touching on. And it goes without saying that democratizing the use of data across an enterprise through these techniques should never come at the expense of good data governance policies.

However, data governance policies that are too strict will kill any efforts to implement the important SSA component of a data-powered strategy before it ever gets off the ground. The key, as with the interplay between SSA and o16n strategies, is balance.

That is, building a solid (yet flexible) strategy that allows lines of business access to the data they need while also restricting any access they have no business need to access. And, of course, maintaining a workable feedback strategy that allows users to gain the access they don't have (but need) with appropriate amounts of friction that won't kill data projects because of overbearing data access problems. This balance is best struck in a centralized environment, where roles and rights can be managed and updated easily as obligations and priorities evolve over time.

The Role of Time

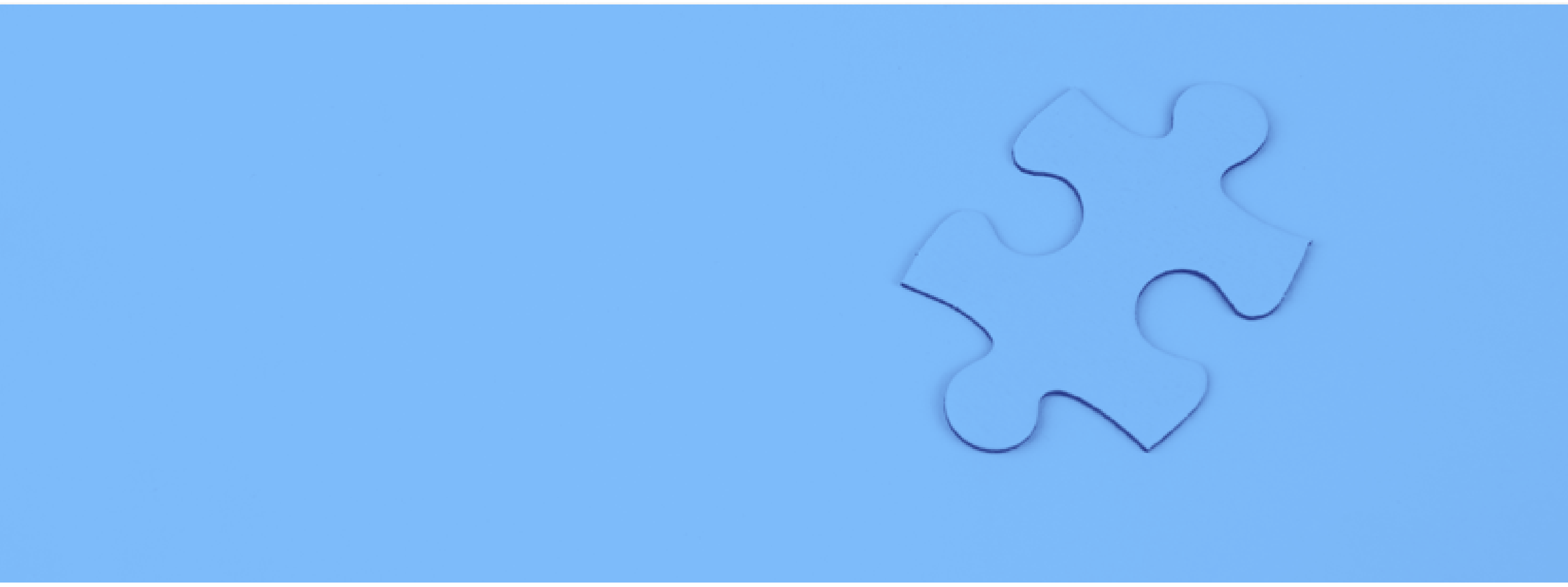
With all of the pressure to move into the era of machine learning or even deep learning and AI - whether from investors, competition, executives, or board members - organizations sometimes move too quickly in these areas. They rush to invest in the latest technologies or to hire the best data scientists without proper consideration for how these parts will fit into the whole.

As previously mentioned, the biggest challenge to the change is organizational, so considering the parts of a whole is exceedingly important (yet often overlooked). Therefore, the best results to instituting this organizational change often come with time and with gradual change that eventually have radical impact. By contrast, radical changes often ultimately end up being too abrupt, not to harsh, or not well thought out, and they fail to have that same impact.

That means when it comes to implementing SSA and o16n, it's critical to balance smaller, shorter-term wins with a more long-term plan for bigger change. This approach can save more time and money in the end as staff can deal with the progress and adapt to changes.

Specifically, playing the long-game
when it comes to organizational change brings:

1. Time to hire the right staff that can bring the right combination of skills and attitude toward a data culture.
2. A more holistic approach to technology, which is ever-changing as the latest algorithms, data storage systems, etc., continue to evolve. Taking a more holistic approach for the long-term allows companies to make better decisions that don't pigeonhole them into one specific technology, language, or resource.
3. More solid governance practices, considering how decisions about data access and use will affect the organization long-term and building a governance program to match (and, not to mention, to reduce liability).



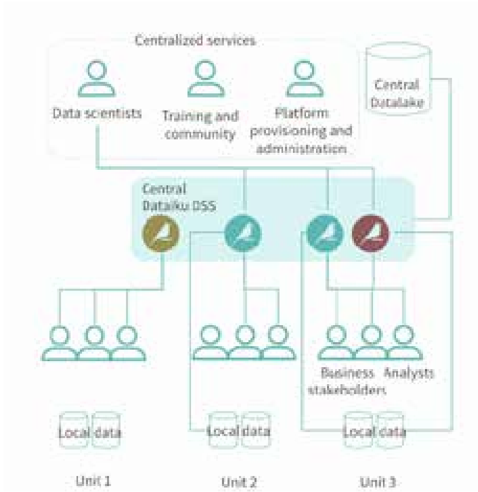
Case Studies

Dataiku has been a part of the transformation of enterprises worldwide into data-powered businesses through the establishment of SSA and o16n processes. Here are three common ways that large enterprises structure their SSA and/or o16n efforts:

Self-Service Analytics with Data Lake

- Training and community are handled centrally.
- Ideas are generated at the unit level, or suggested centrally based on successful projects in other units.
- Projects are developed at the unit level using centralized tools, opportunistic centralized expertise, mixing central data lake data and opportunistic local data.
- Deployment and productivizations are handled centrally and deployed in decentralized business processes.
- One of the benefits of this approach is that as new data is being used and prepared by the SSA users, the central operations will be able to identify which new data sources are the most relevant to ingest into the data lake for easier and broader consumption.

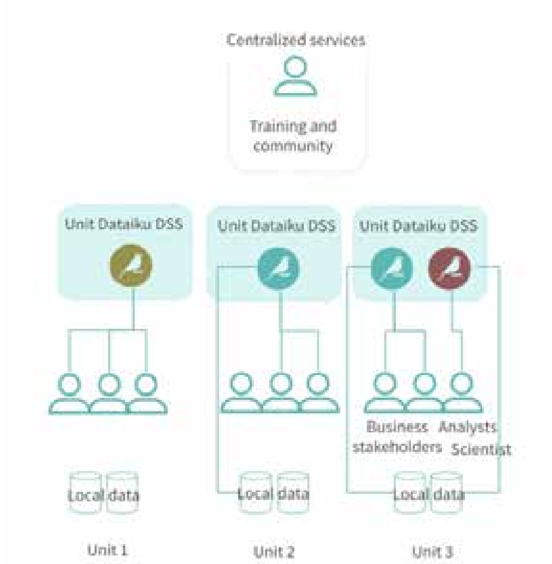
Component	Centralized	Decentralized
Platform	X	
Data	X	X
Operationalization	X	
Training	X	
Standard usage		X
Advanced usage	X	
Community	X	
Ideation	(replication)	X
Impact		X



Disseminated Operations

- Training and community are handled centrally.
- Platform, data, and execution are handled separately in each unit.

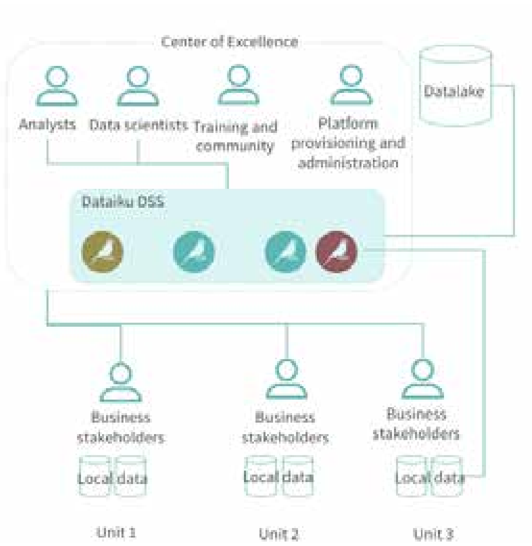
Component	Centralized	Decentralized
Platform		X
Data		X
Operationalization		X
Training	X	
Standard usage		X
Advanced usage		X
Community	X	
Ideation		X
Impact		X



Center of Excellence (Recommended)

- Ideas are generated at the unit level or suggested centrally based on successful projects in other units.
- Projects are developed at the central level, mixing central data lake data and opportunistic local data.
- Deployment and operationalization are handled centrally.
- Data and projects are available to the wider organization to facilitate reuse.

Component	Centralized	Decentralized
Platform	X	
Data	X	X
Operationalization	X	
Training	(X)	
Standard usage	X	
Advanced usage	X	
Community	X	
Ideation	(replication)	X
Impact		X



CONCLUSION

Operationalization and self-service analytics are critical pieces to the transformation of a data-powered enterprise, including the movement toward AI. However, that doesn't mean that the two systems in and of themselves are a final destination; rather, o16n and SSA are processes requiring continuous improvement and evolution.

Businesses will continue to collect new data (potentially even from new data sources) from which to draw insights. Models always have room for improvement, and of course, shifts in the organization can require a shift in o16n and SSA systems to align. All of these evolutions will mean refining - and perhaps enhancing - these existing systems.

We hope that this white paper has prepared you to take the initial steps. If you're just starting your path to o16n and SSA, know that even by considering and understanding how the two work together, you are well ahead of the adoption curve in this space. To understand at a deeper level where your organization stands on the path to SSA and o16n, we've also prepared a quiz - [7] take it here to see your results.

Endnotes

1 www.forbes.com/sites/louiscolumbus/2015/02/25/key-take-aways-from-gartners-2015-magic-quadrant-for-business-intelligence-and-analytics-platforms/#434510b559aa
2 www.gartner.com/en/newsroom/press-releases/2018-01-25-gartner-says-self-service-analytics-and-bi-users-will-produce-more-analysis-than-data-scientists-will-by-2019
3 www.assets.kpmg.com/content/dam/kpmg/xx/pdf/2016/10/building-trust-in-analytics.pdf
4 https://pages.dataiku.com/data-science-operationalization
5 https://blogs.wsj.com/cio/2018/05/09/seeking-insights-into-rare-diseases-pfizer-scales-ai-analytics-platform/
6 https://blogs.wsj.com/cio/2018/05/09/seeking-insights-into-rare-diseases-pfizer-scales-ai-analytics-platform/
7 https://dataiku.typeform.com/to/IlKsts

data iku

Your Path to Enterprise AI

Dataiku is the centralized data platform that moves businesses along their data journey from analytics at scale to enterprise AI. Data-powered businesses use Dataiku to power self-service analytics while also ensuring the operationalization of machine learning models in production.

20,000+

ACTIVE-USERS

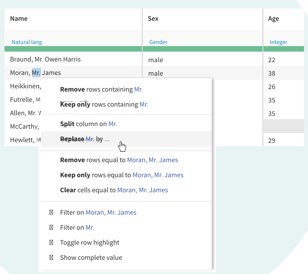
*data scientists, analysts, engineers, & more

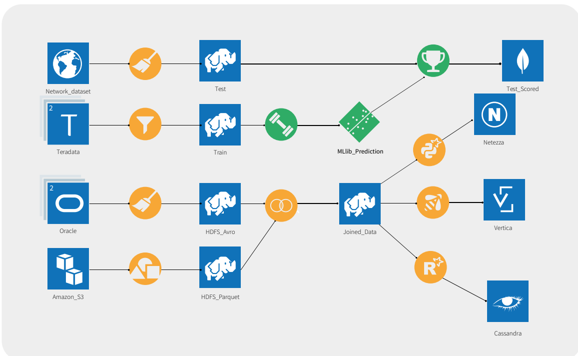
200+

CUSTOMERS

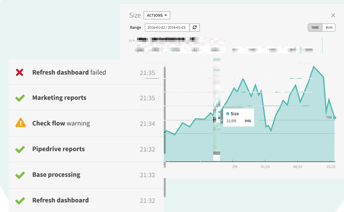
Santander
paloalto NETWORKS
PREMERA BLUE CROSS
FOX NETWORKS GROUP
Unilever
NXP
SEPHORA
KUKA
BNP PARIBAS

1. Clean & Wrangle

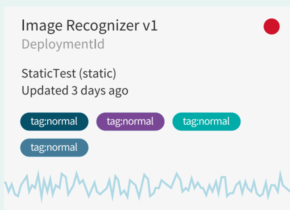




5. Monitor & Adjust

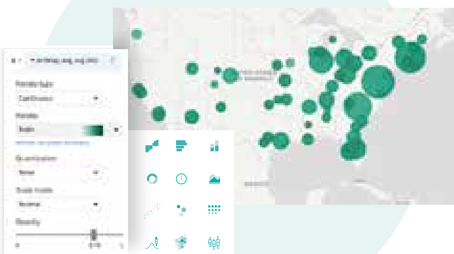


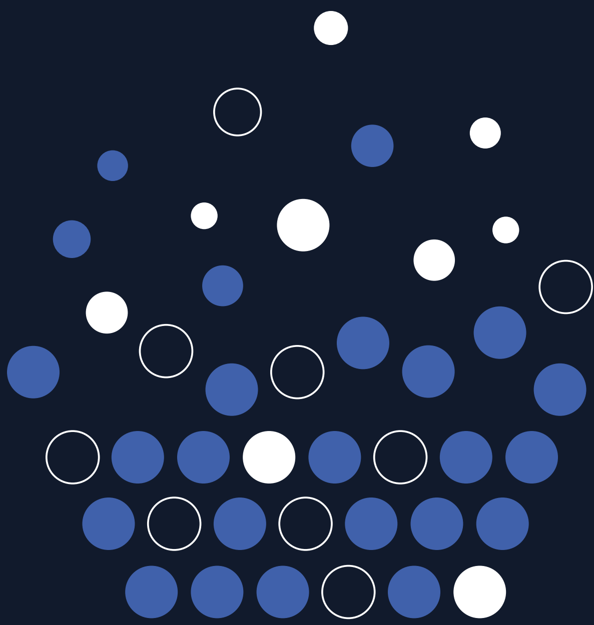
4. Deploy to production



2. Build + Apply Machine Learning

3. Mining & Visualization





WHITE PAPER