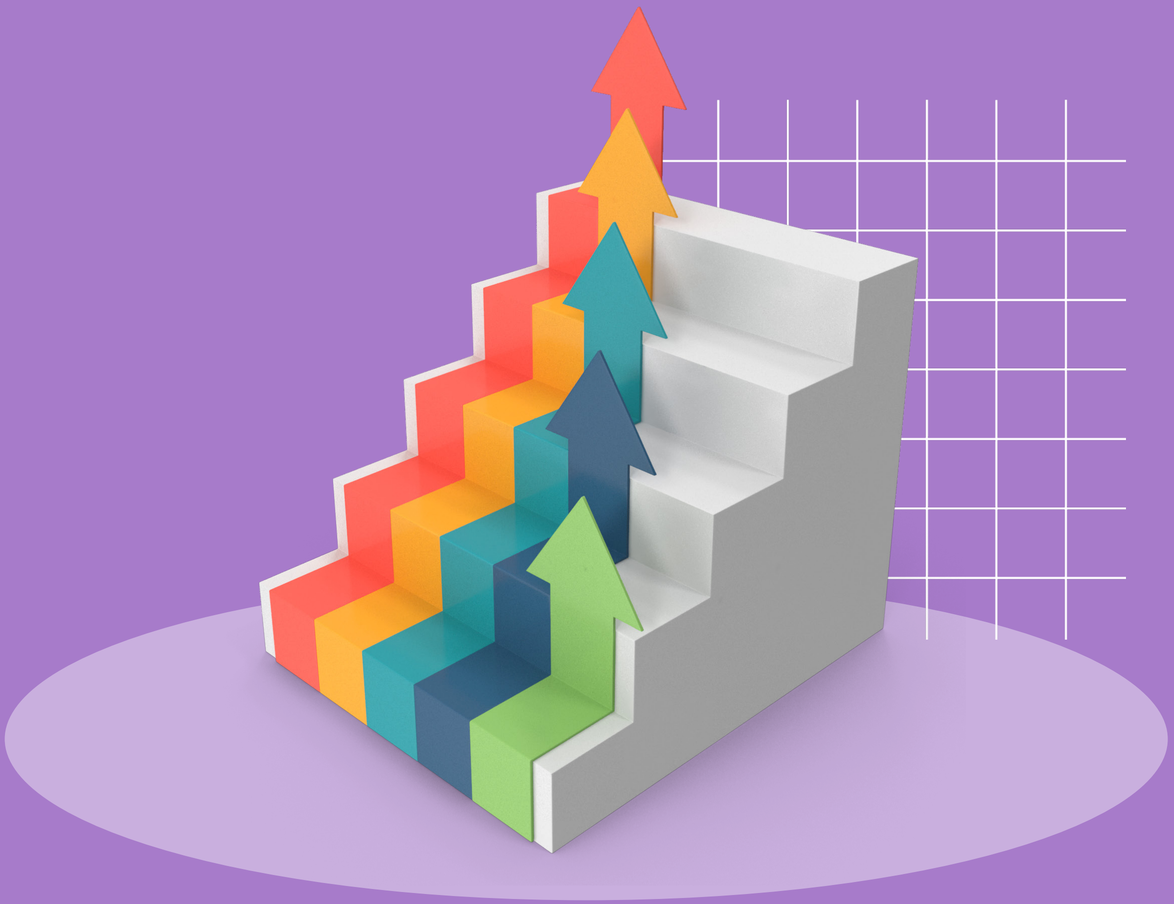
EBOOK

The Ultimate Playbook for Scaling AI

Tailored Guides to Success For Data Team Managers, Business Stakeholders, Executives, and IT Managers



Introduction

The average payback on AI investments is around 17 months, and for beginners, it's even longer, a 2020 ESI ThoughtLab study found. However, executives also report that the year's global crisis has led to a shift in expectations — hit by the economic downturn, organizations are striving to shorten payback periods and scale AI efforts across the enterprise more efficiently.

> "We've surveyed thousands of executives about how their companies use and organize for AI and advanced analytics, and our data shows that only 8% of firms engage in core practices that support widespread adoption. Most firms have run only ad hoc pilots or are applying AI in just a single business process.
>
> **Why the slow progress? At the highest level, it's a reflection of a failure to rewire the organization."**
>
> Harvard Business Review, Building the AI-Powered Organization[1]

As companies embark on their Enterprise AI journey, they find that return on investment (ROI) does not stem from cost savings alone, but rather from strategic transformation toward an approach to advanced analytics, data science, machine learning, and AI that is both agile and democratized. When it comes to pivoting the business around data, everyone has a different — yet crucial — role to play. Missing the cooperation of just one team (and in some cases, just one person) can prevent the organization from successfully scaling AI, potentially damaging business prospects for years to come.

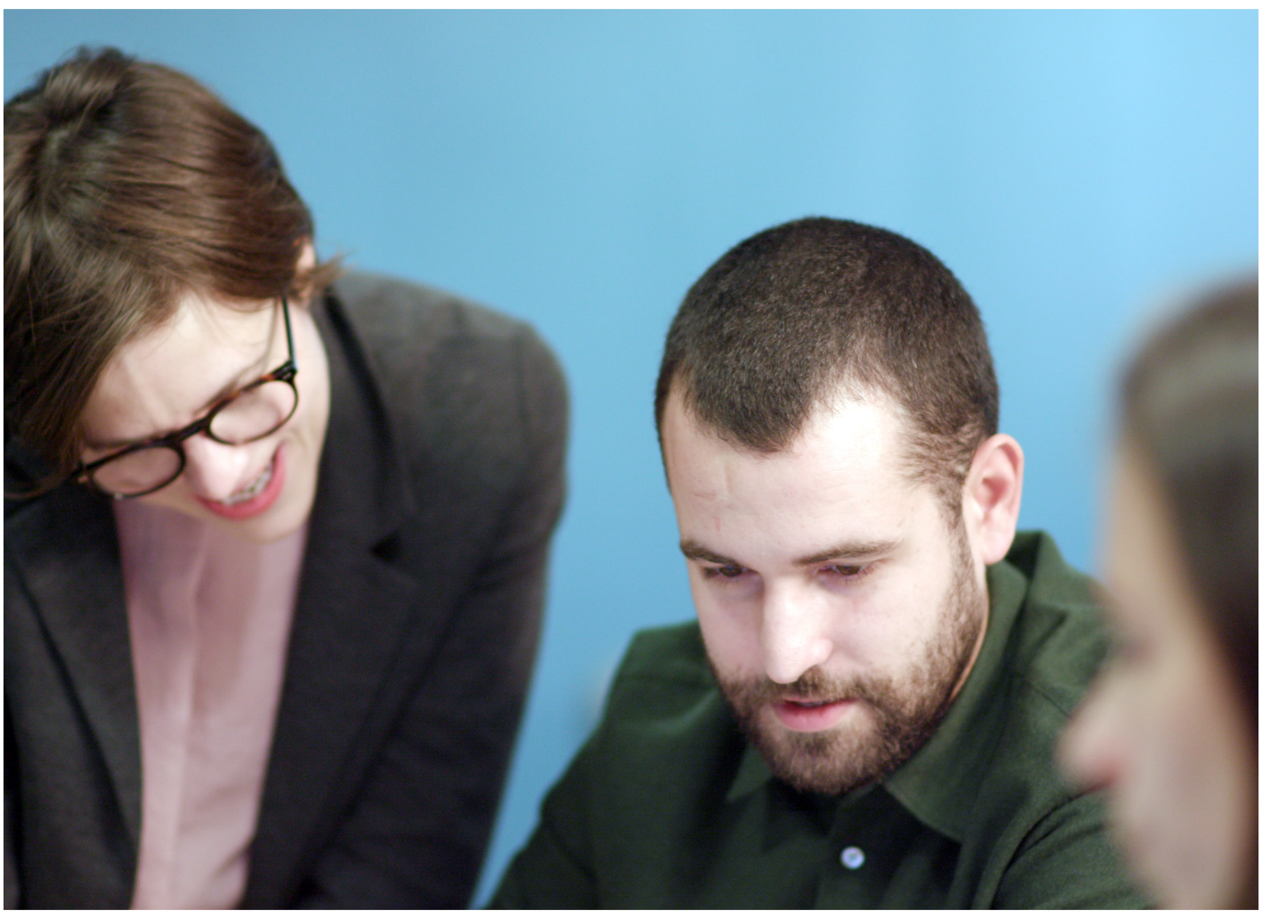
This ebook presents the part different people or roles must play in organizational transformation, providing tips, keys to success, and real-life stories about scaling AI efforts for each. The insights come from Dataiku (one of the world's leading Enterprise AI and machine learning platforms) and our years of working with organizations large and small on their data and strategic AI transformation, seeing what has brought success beyond simply technology.

[1] *https://hbr.org/2019/07/building-the-ai-powered-organization*

Data Team Managers

Data team managers play a decisive role in organizational transformation around data initiatives, as their attitude and approach often dictate the course the company takes. It's an exciting role, but it's also a big burden to bear. This section will identify:

- Some of the top challenges for data team managers that block the organization's ability to scale AI.
- The most impactful steps data team managers can take to spur organizational transformation.
- Real-life stories from data team managers who have successfully scaled AI efforts.



Top Challenges to Address

Balancing Innovation & Creativity with Productivity

By nature, data scientists want to play with new technology and work on sexy projects — this is a reality most data team managers have to face every day with their team. The problem is that if the team wants to, for example, do deep learning, they will go looking for a problem, and that problem might not be one that the business will really benefit from having solved (read: no ROI from a business perspective). On the other hand, having an unhappy team will inevitably lead to retention and turnover issues.

The solution to overcoming this challenge as a data team manager is to instill a different mindset and approach to innovation. That is, innovation doesn't always come from the techniques themselves, but from the way data scientists leverage them — for all problems, there's something new and slightly different to try. For example, data scientists might view a very classic problem (like customer churn) as boring. But if data team managers challenge people to apply new combinations of old techniques or new ways of using proven techniques to problems, innovation can take on a new meaning. This will be a win for managers' staff, who will still work on cutting-edge applications, and — critically — a win for the business, who will have solutions to their most pressing problems.

As an example, Dataiku worked with a large pharmaceutical company that was spending lots of time trying to pick the best countries for clinical trials (best meaning both where they could as well as where they optimally should conduct them). Ultimately, Dataiku helped the customer develop a recommendation system for clinical trial countries. While a recommendation engine would be considered standard and perhaps boring for data scientists in another context (say, for example, ecommerce), the team was enthusiastic about applying this technique to an unconventional use case.

Upskilling Across the Organization

Data team managers are often tasked with driving upskilling programs not just within their own teams or within a data science/AI center of excellence, but across the entire organization. This is a good thing — upskilling is critical in the move to AI democratization and scaling. For example, Accenture released a report in May estimating banks could increase revenues by 34% by investing in AI and upskilling[2], and evidence points to this benefit in other industries as well.

> 88% percent companies that see positive ROI from AI train and enable non-data-scientists to leverage AI.
>
> **- ESI ThoughtLab, Driving ROI Through AI, November 2020**

However, upskilling people across an organization is also a huge challenge. Diverse skill sets and needs mean one-size-fits-all training may not be the most efficient solution; however, the more specialized the training, the more time and effort required. In working with customers to support their upskilling programs, Dataiku has seen a few best practices to address this challenge:

Start with an assessment of skills to learn where gaps lie. The assessment should focus not just on evaluating the tech skills of business people for upskilling, but also the business skills of technical people (see Figure 1).

From there, create a training plan based on long-term upskilling goals and existing skills. For example, if the goal is for 80% of the company to be proficient in Python, the solution probably isn't holding a three-day Python training for 80% of the company. Instead, it's evaluating who already has these skills (maybe you're already at 50%) and then identifying which people will be best suited to fill in the gaps. For example, maybe those who already have SQL skills will most benefit — and the organization will have the most to gain — from upskilling.

Create tracks based on existing skills or responsibilities. For example, GE Aviation offers in-depth 100- 200- and 300-level courses to upskill and onboard end users to their self-serve data efforts in addition to a full-day executive training that is more focused to suit their needs.

[2] *https://www.technologymagazine.com/enterprise-it/accenture-banks-could-raise-revenues-34-investing-ai-and-upskilling*

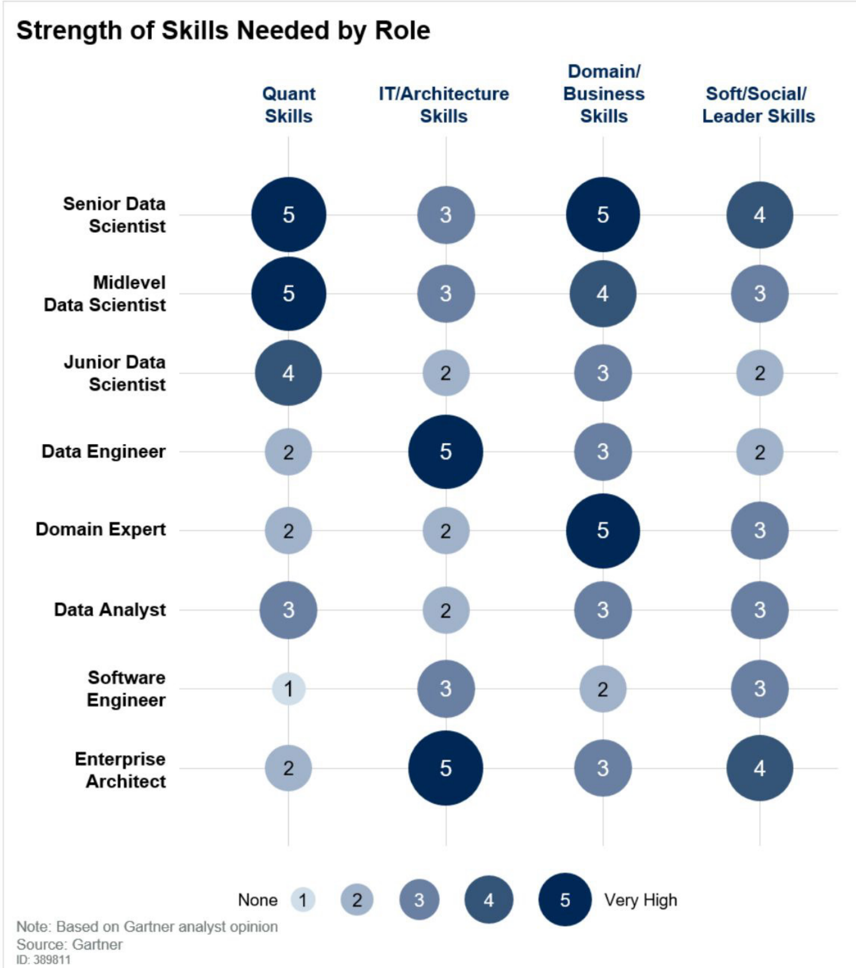


Figure 1, skills needed by role for AI initiatives. Source: Gartner Leading Upskilling Initiatives in Data Science and Machine Learning; Peter Krensky, 19 July 2019 (research available to Gartner subscribers)

The Role of Technology in Upskilling

After education, the next step is actually providing the tools that allow people in traditionally non-data roles to contribute to AI initiatives. That means tools that support easy connection to data, codeless data cleaning and exploration, AutoML, and more.

For example, Dataiku provides one simple UI for data wrangling, mining, visualization, machine learning, and deployment based on a collaborative and team-based user interface accessible to anyone on a data team, from data scientist to beginner analyst to business user.

How to Kickstart Organizational Transformation Around AI

Make a Point to Put Different People — and Data — Together

One of the fastest ways to scale AI initiatives is to create successes in all different parts and at all different levels of the business. That's often easier said than done, especially as "classic" AI use cases (think fraud, customer churn, etc.) have become more tablestakes than differentiators, not having the impact they might have five yeras ago.

At Dataiku, we've seen organizations find success when data team managers take the initiative to open lines of communication, finding an organized way to put people — and data — together that can highlight potential unique issues that fall in the middle of teams' responsibilities and are thus not being solved by anyone.

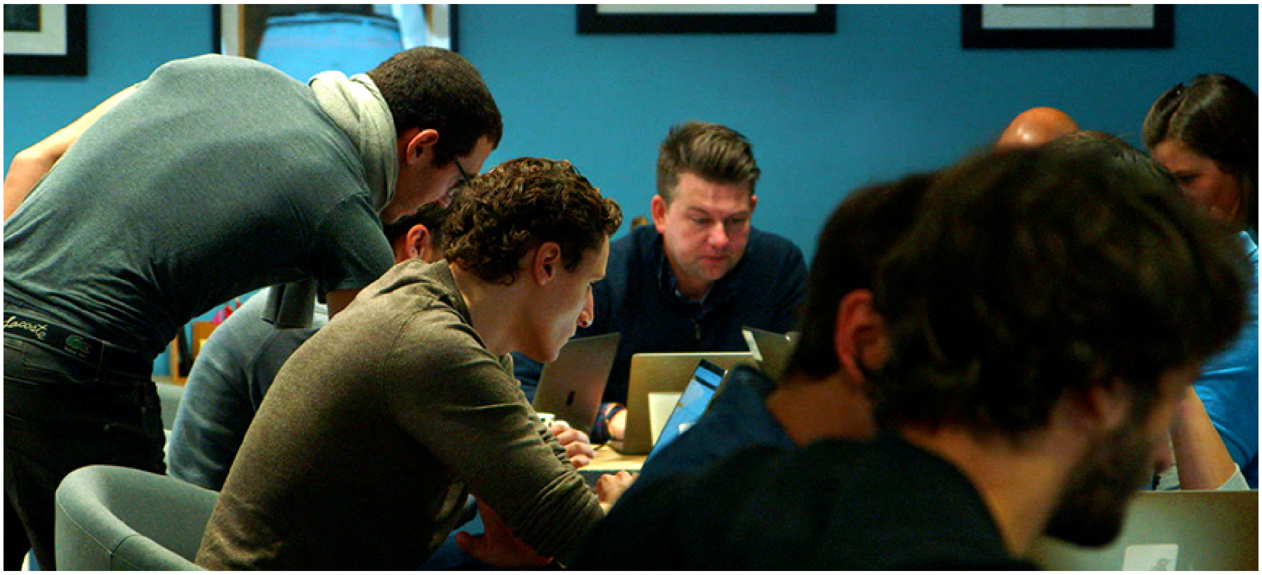
For example, Dataiku was working with a truck manufacturer that wanted to do a proof of concept (POC) on a relatively classic (but nonetheless challenging) use case: leveraging internet of things (IoT) sensors for advanced predictive maintenance. The project was extremely ambitious, and the ultimate end goal was to develop a system that would trigger alarms for drivers when maintenance was critical. In working on the project, the team put together data from different teams across the organization that they had never blended before, including classic sources like truck movement but also less obvious data such as warranty information.

Ultimately, in putting these disparate data sources together, they found some oddities: in particular, that there were some trucks that were supposed to be out of service, but instead, they were traveling around. In investigating further with business teams, they uncovered cases of warranty fraud — in other words, people were sending parts for repair for trucks under warranty, but actually using those parts in other trucks not under warranty. Ultimately, the POC was a success because the teams uncovered a problem that was even more important to solve than their initial business need.

This isn't to say that data team managers should always dig around aimlessly looking for problems to solve — it's still important to address real business concerns to get value from AI initiatives. However, building a culture that can find new, less obvious problems by bringing different people together is invaluable. As organizations across the globe become more mature in their AI efforts, it's these kinds of initiatives that will make the difference. To start, it can be as simple as getting people in the same room (like finance and marketing) and see what kinds of business problems they share and what solutions the data team can help facilitate.

Continue to Reorganize Around Specialization

AI at scale inevitably means more granularity and specialization in data roles, not less. By comparison, think about a mature finance, marketing, human resources, or any other team — instead of having one person who is generalized and does it all, these teams grow to have specialized people in each function. In the human resources example, that means recruiters, staff handling only compensation or benefits, learning and development experts, etc.



It is therefore important for data team managers to understand, encourage, and continue to organize around these nuances and specialities to allow the organization to scale AI efforts. In practice, that means:

Not entertaining ideas of “real” vs. “fake” data scientists and instead promoting everything along the scale, from the self-taught data scientist to the PhD, focusing instead on skills and how they can best be leveraged to move the organization forward.

Making distinctions between different roles on the team and allowing people to focus on what they’re best at. For example, a data scientist who is a strong coder and interested in production environments might morph into specializing as a machine learning engineer. An analyst with particularly strong communication skills across the business might become a data storyteller.

Emphasizing the value of so-called “soft skills” in data science — like the ability to communicate across teams or to explain technical concepts simply — by creating specific roles around these important facets of AI scaling.

Getting more specific when hiring, looking for specialist roles that will make the data team more efficient and effective or filling in for gaps in existing staff’s specialities.

Of course, increased specialization can mean more complexity in some ways, but that’s where data science, machine learning, and AI platforms like Dataiku come in. With everyone — no matter what their focus area — using the same overarching tool to do their work, technology can help fill the gaps between people to ensure process efficiency.

Stories

- **Hear from Chris Kakkanatt, Data Science Senior Director at Pfizer** on how he helped scale AI across the organization.
- **Hear from Somesh Saxena, Data Team Manager at GE Aviation** on their path to scaling AI through self-service analytics.

Dataiku for Scaling AI

Tom Spencer
Head of Customer Data Science @ Aviva

"If you have the wrong tools in place, you can fly solo — you can get away with inefficiencies and hide your mess a bit. Dataiku changed our team atmosphere and culture for the better through sharing capabilities."

Data Engineers Manager in the Manufacturing Industry[3]

“[Dataiku is a] must-have for anyone that wants to work in a big data environment and wants to scale. Amazing integrations with key tecs.”

Get Business & Tech To Work Together

Dataiku encourages collaboration best practices around AI initiatives where business and tech work better together. It creates a single place for discussion and data projects in the team — a win for governance — that synchronizes with other collaboration platforms such as Slack, Atlassian Confluence, or Microsoft Teams.

Capitalize & Reuse for AI at Scale

Beyond its ability to build AI pipelines, Dataiku helps document and centralize data initiatives. It provides capabilities to track, comment, and describe the various artifacts created by the data team so that they can be reused across the organization.

[3] https://www.gartner.com/reviews/market/data-science-machine-learning-platforms/vendor/dataiku/product/dataiku-dss/reviews?pageNum=5&sort=-review_date

Because to see returns on investment (ROI) in AI projects at scale, it's not enough to take on exponentially more use cases — companies must find ways to decrease both the marginal costs and incremental maintenance costs of Enterprise AI.

Automate Analytics for a Self-Serve Business

Quickly automate business processes with Dataiku via seamless scheduling processes. By empowering the business to automate repetitive analytics tasks (without having to rely on data science or IT teams), people can focus their attention on more impactful — and potentially fruitful — data projects.

Create Thousands of Analytics Apps

Dataiku facilitates the rapid creation of interactive visual front-ends for analytics outcomes. Packaged as autonomous apps, these front-ends can help leverage data and machine learning directly in operations.

Trace Your Data End-to-End

Dataiku provides transparency around how each dataset is used and what data models or insights are based on. This ultimately facilitates the implementation of robust data governance processes and eases data privacy regulation compliance efforts, e.g., the EU General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA).

A complete audit trail with Dataiku allows for detailed tracking of each modification made to data or analytics projects.

[2] *https://www.technologymagazine.com/enterprise-it/accenture-banks-could-raise-revenues-34-investing-ai-and-upskilling*

Closing Thoughts

All the factors discussed in this section that lead to success in scaling AI across an organization — retaining good data scientists while also satisfying business needs, spearheading an upskilling program, breaking down silos — are, coincidentally, also what lead to burgeoning careers for data team managers. In the next few years, companies will be looking for people with proven track records in each of these areas to take on leadership roles in data.

In other words, for data team managers, empowering the right people, building the right systems, and investing in the right technology will not only bring success at the organizational level, but at the personal level in terms of career growth opportunity. Implementing the strategies outlined in this guide can (and have for the people presented in the stories from Pfizer and GE Aviation) mean increased responsibility, support, budget, and more from the business that allow for continued growth into the next wave of AI.

Business Stakeholders

When business stakeholders aren't intimately involved with the scaling of AI, there is a serious risk of those on the technical side solving problems or providing solutions that don't actually serve the larger business at all. Subject matter experts and business stakeholders therefore need to come to the table ready and willing to participate, seeing themselves as the catalyst for turning AI efforts into business impact. This section will identify:

Some of the top challenges business stakeholders must address that can hinder the organization's ability to scale AI, including real-life stories and practical advice for how to solve them.
The most impactful steps business stakeholders can take to spur organizational transformation.

Top Challenges to Address

Data Quality

There is perhaps no bigger challenge standing in the way of scaling AI across an organization than data quality. Without systematically high-quality, consistent, source-of-truth data, it will be impossible to build effective AI systems.

But wait, business stakeholders might think — why is data quality my responsibility? Isn't that something IT or someone else should take care of? Yes and no. IT certainly has a role, but with their long list of priorities and responsibilities as well as potentially all the different business units they serve, the data quality of your particular business unit or team might not be very top of mind.

While poor data quality might not necessarily hurt the business' day-to-day operations, it's nonetheless essential to prioritize to move the organization forward in AI maturity. And it's ultimately up to the business stakeholders to care about data quality and to make it work by establishing processes and working closely with partners in IT to address it.

Dataiku for Data Governance

While no technology or tool is a magic bullet for addressing data quality, Dataiku provides transparency around how each dataset is used and what data models or insights are based on. This ultimately facilitates the implementation of robust data governance processes and eases data privacy regulation compliance efforts, e.g., the EU General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). A complete audit trail with Dataiku allows for detailed tracking of each modification made to data or analytics projects.

Changing the Way People Think and Work Day-to-Day

One enigmatic but important challenge for business stakeholders to address on the path to scaling AI across an enterprise — whether managing a team or an individual contributor — is changing the thinking around AI itself. Just like any other technology, such as the phone or the fax machine, AI needs to be fundamentally integrated with business processes, not something pasted on top.

For example, business stakeholders may ask themselves:

- Where am I (or where is my team) currently using gut feeling to make decisions? What is the decision-making process, and how can AI help not replace those intuitions, but support them?
- How can I partner with data experts to leverage AI for the automation of repetitive, low-value tasks (things like triage, labeling, etc.)?
- Where am I (or where is my team) currently using rule-based processes, and could those be improved with more advanced technology like machine learning?

Scaling AI is not just about technology — it's about evolving the culture and fabric of the organization, fundamentally changing the way people work. Part of this change falls on business stakeholders embracing a culture of working with data from the bottom up.

One way to address this challenge is to levelset on what AI really is. Many business stakeholders think of processes that make up AI — like data science and machine learning — as exact sciences. However, it's important to understand that in reality, it's all very experimental. AI is about trying, failing, and incremental improvement.

From that perspective, business stakeholders can — and should — embrace the following:

- A certain tolerance for failure, or at least understanding that AI solutions may not work as expected the first time. Scaling AI is a back-and-forth process between business and data teams, not a one-and-done solution.
- Executing projects in stages so that a first solution comes quickly and the subsequent steps involve iteration and improvements for the fastest possible time-to-value for the business.
- Refraining from asking about how accurate or "good" the AI system will be upfront. Keep in mind that data scientists and others on the technical side might not even know if the model they've created will work until they try it in a production environment on live data.
- Not every idea will work, so keep coming with ideas and problems even in the case of failure. Just because one AI project wasn't successful doesn't mean AI won't work for any business problem at hand.

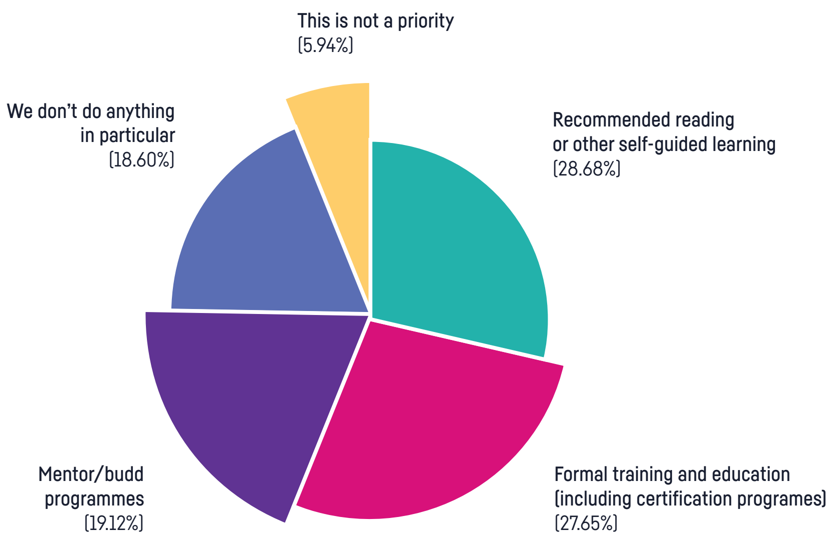
If business stakeholders can come to the table with a productive mindset and a willingness to work with data and tech teams as a partner, looking for incremental improvements and focusing on real business problems, the organization will be on the right track to scaling AI efforts.

Take Initiative to Learn

In order to scale AI efficiently, technical teams need to learn the intricacies of the business to build better solutions; but at the same time, business stakeholders should also take the initiative to learn more about AI technologies to more effectively collaborate across teams.

No one would deny that today, almost everyone in an organization needs to know how to use a computer — not just IT teams. This is not to say that everyone needs to be an expert or the most advanced user, depending on their role. But it's also not IT's responsibility to cover topics like "what is an email." In the very near future, AI will be such a part of business that knowing the basics will become table stakes.

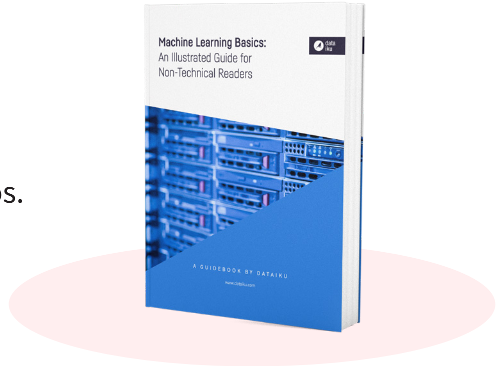
Business stakeholders who also happen to be managers can encourage and incentivize teams to learn the basic concepts of data architecture, data science, and machine learning (the foundations of AI). For individual contributors, taking on this responsibility can be a catalyst for upskilling into a new, more advanced role and an opportunity to work on AI projects that could provide a career boost.



How does your organization help its staff understand the roles data, machine learning, and/or AI play within the business?

Dataiku surveyed 350 data professionals in NYC and London during its 2019 EGG Conferences. See the full results.

Data Science, Machine Learning and AI Basics:
Get Started with the Full Set of Guidebooks and Videos.



Dataiku for Upskilling

After education, the next step is actually providing the tools that allow people in traditionally non-data roles to contribute to AI initiatives. That means tools that support easy connection to data, codeless data cleaning and exploration, AutoML, and more.

For example, Dataiku provides one simple UI for data wrangling, mining, visualization, machine learning, and deployment based on a collaborative and team-based user interface accessible to anyone on a data team, from data scientist to beginner analyst to business user.

Establish a Framework for AI Projects

Business stakeholders should come to the table with clearly defined goals, business questions, and/or key performance indicators (KPIs) that they want to achieve or address. In some cases, they might be extremely well defined (e.g., “In order to hit our numbers for the quarter, we need to reduce customer churn by 10%” or “We’re losing $N per quarter due to unscheduled maintenance, how can we better predict downtime?”) In other cases, less so (e.g., “Our service staff needs to better understand our customers to upsell them” or “How can we get people to buy more widgets?”).

In organizations with healthy processes, starting the AI lifecycle with a more well-defined business question isn’t necessarily always an imperative, or even an ideal scenario. Working with a less-defined business goal can be a good opportunity for business stakeholders to work directly with data scientists upfront to better frame the problem and brainstorm possible solutions before even beginning any data exploration or model experimentation.

Either way, business stakeholders need to establish concrete frameworks both for submitting these goals or business questions and for how to work with data teams overall. Processes around how projects are evaluated for success is also key to avoid misunderstandings between business and tech teams that can stiefel AI momentum.

As an example, GE Gas Power — a subsidiary of General Electric and a Dataiku customer — has a framework for data and business experts to work together that started out through email, which quickly became unfeasible as their AI initiative scaled. From there, they established a more formal process:

1. Business stakeholders are required to submit an analytics proposal.
2. The data team looks at the business urgency of the proposals and prioritizes them from there.
3. One a project kicks off, the team must work on translating business requirements into technical requirements, which requires close collaboration between business stakeholders and data experts.
4. From there, it's on to preliminary design where the team makes sure their expectations for features and deliverables are aligned with the business stakeholders, which means defining a minimum viable product (MVP).
5. Business stakeholders and data experts work together throughout the process, including clarifying requirements or coming back to them if they find certain data not available (or if they're lucky, that more data is available). If everything goes well, the minimum contact point is the user acceptance test (UAT).
6. After the MVP is complete and accepted by the customer, they might submit additional enhancements, and the process repeats.

Depending on how the organization is set up and what people are involved, processes might look different. But the point is that having a defined way to go from idea to AI project is pivotal in scaling efforts.

Closing Thoughts

In the near future, it is likely that there won't be such a distinct separation between data roles and business roles — business leaders and individual contributors alike will be expected to have the data competencies needed for their day-to-day work.

That's why the challenges, solutions, and strategies outlined in this section are critical not only for the benefit of the organization, but for business stakeholders' long-term career prospects. Companies will seek out and promote those in the business who are able to break down silos and incorporate data processes, leveraging technology along the way.

Executives

Three out of four C-suite executives believe that if they don't scale AI in the next five years, they risk going out of business entirely. Yet at the same time, 76% of those executives also report they struggle with how to scale.[4] This section will identify:

- Some of the top challenges executives must address to scale AI.
- The most impactful steps executives can take to spur organizational transformation.
- Real-life stories from executives who have successfully led AI scaling efforts at their organization.

[4] https://www.accenture.com/us-en/insights/artificial-intelligence/ai-investments

Top Challenges to Address

Return on Investment (ROI)

The 2020 study **Driving ROI Through AI**, which surveyed executives at 1,200 companies across 12 industries and 15 countries, illustrates that the challenges around AI ROI for executives are actually multifold:

1. Gauging the ROI on AI is more complicated than for traditional technology investments — more than 50% of executives surveyed do not even have systems in place to measure returns. It's often difficult to isolate the contribution of data alone to improvements, especially larger business outcomes (like higher profit margins, lower costs, etc.). Plus, the calculation is complicated because the value isn't all in one number — it can be spread across multiple departments and teams.
2. ROI on AI tends to be very long-term (more than a year on average). Projects can take longer than expected to implement due to unanticipated hurdles—such as not having access to the right data or difficulties in training and testing the model. Sometimes the models fail altogether, requiring that AI teams start all over
3. To generate strong ROI from AI, executives need to first lay the groundwork and put the right processes in place, which requires adequate investment. It's important to consider that costs don't just involve getting the first AI project out the door — models require maintenance over time (that's where **MLOps** comes into play) that represent real investment and costs as well.

In other words, getting ROI from AI isn't just a matter of investing in the right technology and hiring the right people. Scaling AI requires organizational change, which requires time and resources — it's not just because people have access to data that democratization simply happens.

So ROI is a challenge, but what's to be done? Many large organizations that have led successful AI transformations look for other ways to quantify success. For example, while GE Aviation has been able to quantify their efficiencies and savings via their self-service data program to the tune of millions of dollars, more importantly for them is the value the program has brought to the culture of the organization as a whole.

> "You don't always need to measure stuff to know it's successful — when talking to people, it's clear that the self-service data program makes their job easier."
>
> *— Somesh Saxena, Senior Staff Technical Product Manager at GE Aviation*

There is agreement throughout GE Aviation that the self-service data program brings enormous value, and that's not by accident. The SSD team has not only received plenty of positive feedback about the program and about Dataiku, but they're careful to share this information with the wider company to continue to feed and show the value of the program.

How to Kickstart Organizational Transformation Around AI

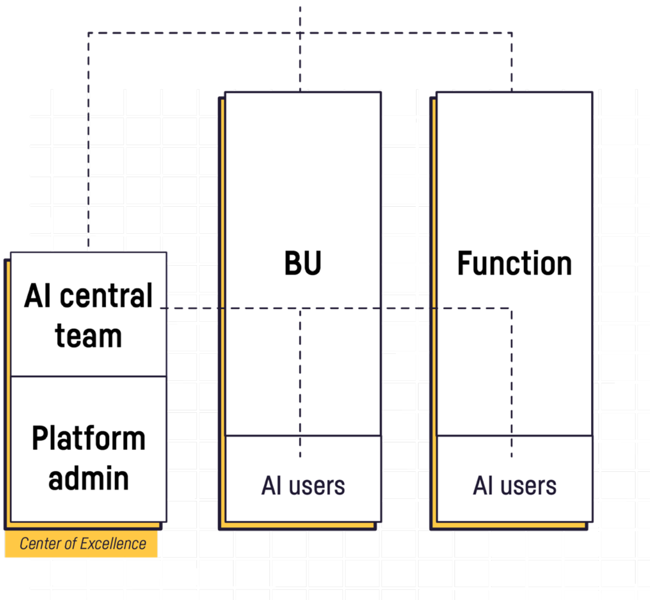
Think About Operating Models

Enabling people at all levels to use data and execute AI projects on a day-to-day basis can't happen without some serious organizational change management and process oversight. That's where operating models, including the center of excellence (CoE) come in. From an organizational perspective, CoEs provide a more unified approach to AI projects (think preventing repeated efforts across the company) and can also bring a more innovative mindset.

The three most common operating models for AI initiatives are hub and spoke, centralized, and decentralized (where the first two contain the so-called center of excellence). The latter is generally not recommended as a first organizational structure when trying to get AI initiatives off the ground; however, it is presented here because it can be a logical evolutionary step for some organizations that are farther along in their AI maturity as they scale.

Hub & Spoke

Analytics teams are scattered within the business departments with one team



- ⊕ High specialization and high productivity
- ⊕ Increased AI knowledge share and upskilling
- ⊕ Consistent practices
- ⊕ Balance between flexibility and efficiency

- ⊖ Coordination and community efforts
- ⊖ Complex budgeting

Real-World Example

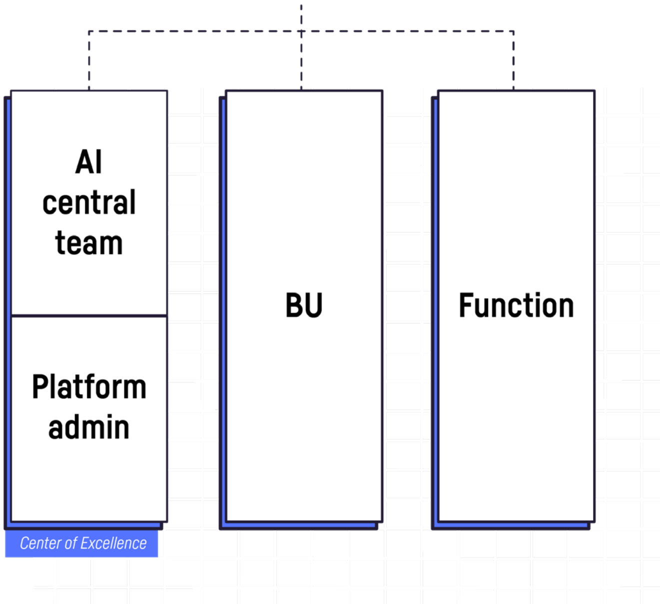
Forbes 2000 in the Manufacturing industry

- ⊕ 2,000+ Self-Service Analytics users
- ⊕ 3,000+ data products created in just three years
- ⊕ $400 million+ revenue uplift
- ⊕ Operating model being rolled-out to all the group's subsidiaries

- ⊖ Alignment can be a challenge given the number of key stakeholders
- ⊖ Complex coordination not only to set up, but to maintain
- ⊖ The size of the central team should not be underestimated

Centralized

All analytics teams are part of one single business unit



- ⊕ Control
- ⊕ Easy collaboration and reusability
- ⊕ High level of advanced AI innovation

- ⊖ Rest of the organisation is out
- ⊖ Disconnect between AI and users
- ⊖ AI innovation limited to CoE
- ⊖ Slow time to value
- ⊖ Shadow IT

Real-World Example

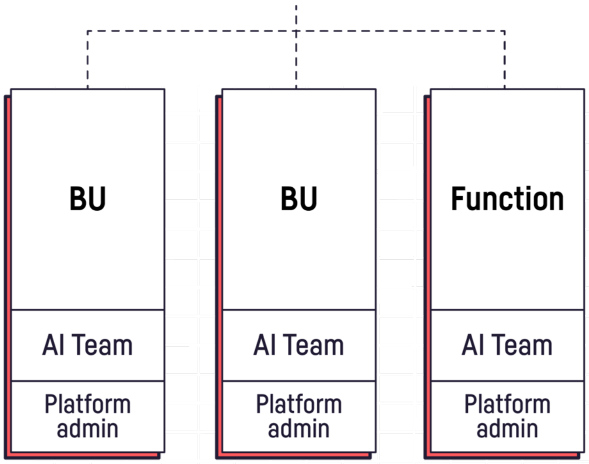
Dataiku customer – Forbes 2000 in the CPG industry

- ⊕ 0 to 350 newly hired data science and machine learning staff over four years
- ⊕ 2,500 analyses in-house a year
- ⊕ AI Assets reused hundreds of times
- ⊕ $30 million+ agency spendings saved
- ⊕ Estimated revenue increase of between $500 million and $1 billion
- ⊕ HBR article recognizing its innovative edge

- ⊖ Machine learning practitioners up to 5x removed from the business (e.g., marketers)
- ⊖ Broader organization not clear on the team's created value
- ⊖ Focus mostly on 1 single BU
- ⊖ High spread in terms of AI maturity levels within the organiz ation

Decentralized

Analytics teams are within the business departments



- ⊕ Each domain remains highly specialized
- ⊕ Strong collaboration
- ⊕ Easier to prioritize projects
- ⊕ Short time to value
- ⊕ Easy to budget

- ⊖ No sharing
- ⊖ Low reusability / capitalization
- ⊖ Lack of consistency
- ⊖ Silos
- ⊖ Higher costs

Real-World Example

Dataiku customer – Forbes 2000 organization in the banking industry

- ⊕ 3 continents and more than 10 teams involved
- ⊕ Businesses experiencing the value of data science and machine learning
- ⊕ Very relevant use cases
- ⊕ 100s of models deployed

- ⊖ Limited training capabilities and knowledge exchange between teams
- ⊖ No visibility at group level
- ⊖ Multiplication of similar tools, missing out on potential synergies

Have a Sustainable Vision

The world of Enterprise AI is evolving fast — the data science and machine learning technology revolution is exciting, with the rapid acceleration of new technologies disrupting at massive scale. Companies that are agile and able to adapt to the pace of the revolution succeed and have a leg up among their competition, while large organizations that struggle to pivot their efforts often have a hard time keeping up.

The 2020 global health crisis taught us that times of economic change tend to expose companies that aren't able to easily adapt, which is why agility and having an AI strategy that works now and in the future (whatever the future might bring) is paramount. This means organizations must be prepared to:

- Scale up and down infrastructure resources as needed.
- Support a spectrum of AI use cases depending on the company's immediate needs (from self-serve analytics to operationalized models in production).
- Easily monitor and adjust models as needed in times of volatility.
- Adapt to the ebb and flow of AI technologies — for example, look at the rise of Kubernetes and fall of Hadoop in popularity in just the last five or so years.

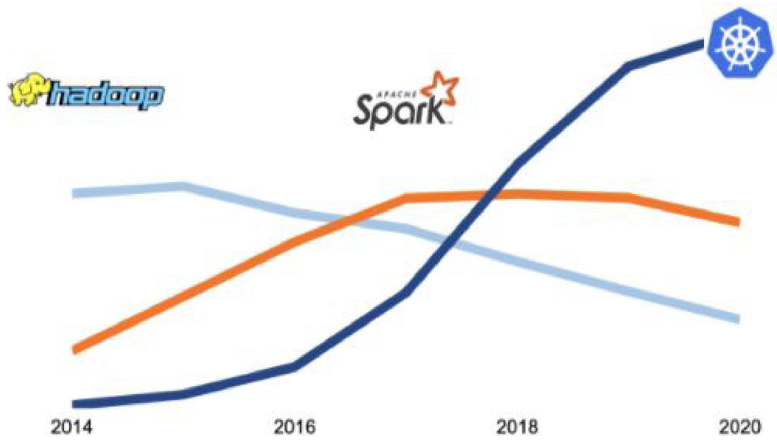


Figure 2: A look based on Google Trends at the evolution of the world of data infrastructure.

A sustainable vision of AI also means examining the costs of maintaining AI projects over time. In addition to cloud costs, there are also costs associated with continual cleaning and preparation of data, pushing to production, model maintenance, and complex technological stacks. Common sense and economics tell us not to start from scratch every time, and that is exactly the principal behind reducing these costs.

Reuse is the simple concept of avoiding rework in AI projects, from small details (like code snippets that can be shared to speed up data preparation) to the macro level (like ensuring two data scientists from different parts of the company aren't working on the same project). Capitalization in Enterprise AI takes reuse to another level — it's about sharing the cost incurred from an initial AI project (most commonly the cost of finding, cleaning, and preparing data) across other projects, resulting in many use cases for the price of one, so to speak.

Capitalization means that while tackling larger, high-priority use cases, the organization can also take on lots of other smaller use cases by reusing bits and pieces, eliminating the need to reinvent the wheel with data cleaning and prep, operationalization, monitoring and — in doing all of that — ensuring that data scientists are happy, spending their time on high-value tasks.

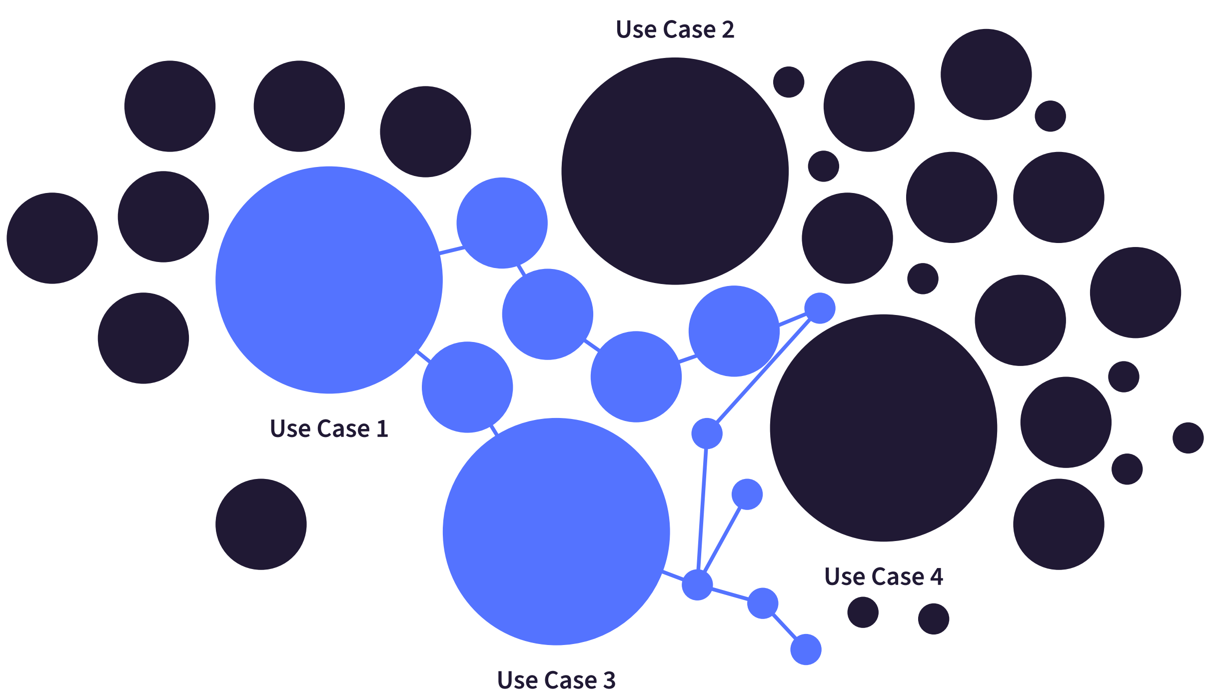


Figure 3: What capitalization and reuse across the organization can look like, leveraging parts of big, cornerstone use cases to fuel hundreds of smaller use cases with little additional marginal cost.

The Role of Technology in a Sustainable AI Vision

Data science, machine learning, and AI platforms like Dataiku are tools to enable AI at scale by providing transparency and reproducibility throughout — and across — teams:

Cost	Mitigated with Dataiku via...
Data Cleaning and Preparation	Reuse of already cleaned and prepared data across projects as well as between personas (i.e., data scientists can use data prepared by analysts).
Operationalizing and Pushing to Production	Reuse from design to production (i.e., without the need to recode models and pipelines from scratch to operationalize).
Data Scientist Hiring and Retention	Reuse of project elements across users, allowing data scientists to spend more time on higher-value tasks (which also happen to be more interesting for them, which means lower risk of brain drain) as well as ability to code in preferred languages and leverage open-source tools.
Model Maintenance	Automated scenarios, monitoring, and reuse of infrastructure across technology stacks.
Complex Technological Stacks	Freedom to reuse and adapt even across changes in technology with Dataiku as an abstraction layer, freeing people from the underlying technology.

A tool that can enable capitalization and reuse should provide at a minimum:

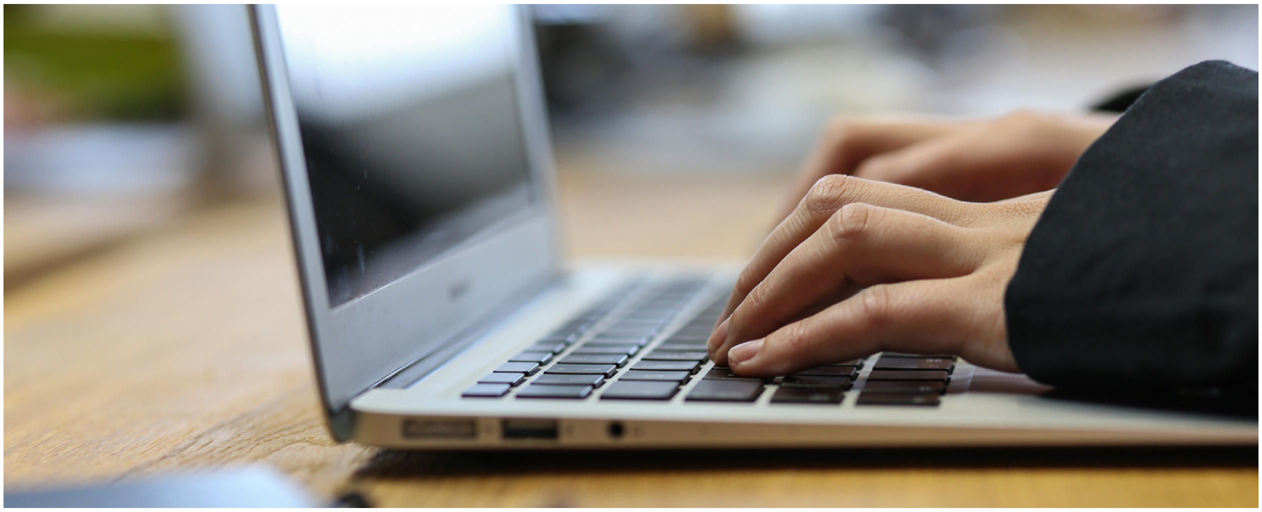
- Robust documentation so that contributors can explain what has been done in a specific project via wikis, to-do lists, versioning, activity logs, etc.
- A built-in, centralized, and structured calolog of data treatments (from data sources to data preparation, algorithms, and more) for easy consumption.
- The possibility to grab parts of data projects and input them in new projects or mix components of two different projects together.
- The possibility for advanced users to package data products as plugins to be used by others without the need to understand all the underlying complexities.
- An advanced console to monitor usage, versions, and quality to ensure easy and efficient operationalization.
- The possibility to automate scenarios using complex triggers as well as automate test production and deployment.

Dataiku provides all of these things while adding one extra layer: user interface accessible to anyone on a data team, from data scientist to beginner analyst. True inclusivity and democratization of data efforts brings capitalization and reuse to another level, as it's no longer just a question of data scientists, but of reuse by everyone across the organization.

Be a Believer

Aside from investing in people, processes, and technology that support AI at scale, one of the biggest things executives can do to aid the journey is being a true believer in AI. It sounds cheesy, but it counts — executives who invest in AI just because they feel they have to are not the ones that will come out ahead with creative or productive implementations. Those who are positive will have the potential not just to defend, but to disrupt.
Here's some food for thought about being a believer in scaling AI:

- For businesses that already historically use data for their core business (banking and insurance, travel, etc.), being a believer means thinking about embracing even more use cases beyond what's table stakes for everyone.
- For businesses that don't already use data for their core business, think "why not us?" There's big opportunities for real competitive advantages and not much to lose.
- Being a believer doesn't require an enormous leap of faith — it's not about believing that robots will take over, but the application of AI to very specific use cases that augment human abilities.
- The computer didn't become really useful or ubiquitous until the internet was invented. Even then, it took faster network speeds to get to the level of integration of computers with daily life that we have today. Right now, data is the computer — it's not useful until it's being leveraged daily by people around the organization as a part of a scaled AI strategy. By the time the equivalent of faster network speeds comes around, it will be too late to catch up.



Stories

- **Hear from Nicolas Bignell, Director of Data Science at UBS Investment Bank,** on their path to scaling through creating an AI center of excellence.
- **Hear from Ayan Bhattacharya, Managing Director of Applied AI at Deloitte Consulting,** on fueling digital transformation within the enterprise.
- **See how Mercedes-Benz, Vodafone, and Credit Suisse super-sized their data initiatives.**

Closing Thoughts

Data executives may not be involved in the day-to-day building of models or defining of business questions, but their leadership and attitude nonetheless dictate how efficient the organization can be in execution.

At Dataiku, we've seen hundreds of organizations start their Enterprise AI journey, and the most successful ones are those where transformation happens from the bottom-up and well as from the top-down. Bringing people, processes, and technology together to transform an organization simply cannot be done without support for AI initiatives from the top.

IT Managers

It goes without saying that when it comes to scaling AI, tooling, technology, and being able to go smoothly and efficiently from development or test into the real business are critical pieces, which is why IT teams play an integral role. Yet **according to a 2019 survey** of 200 IT professionals, their biggest concern remains — naturally — security. Finding the right balance between governance and democratization, security and speed, is critical for the success of AI initiatives. This section will identify:

1. Some of the top challenges IT managers must address to scale AI.
2. The most impactful steps IT managers can take to spur organizational transformation.
3. Real-life stories from IT managers who have successfully collaborated with the business to execute on AI scaling efforts at their organization.

Top Challenges to Address

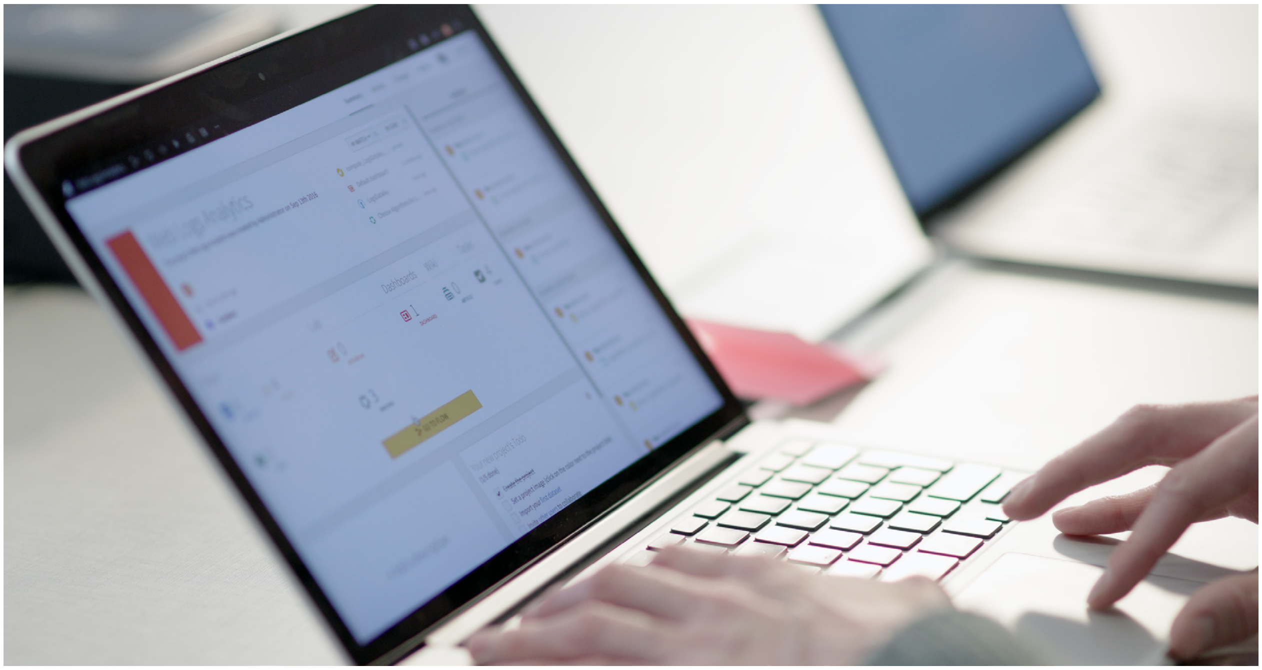
Finding the Right Balance Between Audatibility and Permission Management

Let’s be honest: the easiest way to ensure reliability and security is to make minimal changes, to err on the side of restricting rather than granting access, and to be vigilant about the adoption of new tools (especially those that promote increased access to data). In other words, the very idea of scaling AI through data democratization goes against many traditional security principals, so it’s natural that IT teams are cautious.

However, too much caution can result in the rise of shadow IT (i.e., business units developing applications, infrastructure, or processes that circumvent central IT organizations), which is a particular risk when the business wants to move fast on scaling AI efforts. It is therefore important to be a part of the conversation, taking a step back when it comes to security to determine what makes sense in the context of democratization and how IT teams can both retain an appropriate amount of control while also giving the business the freedom to work with the data they need to optimize decision making.

When it comes to democratizing AI initiatives, security has two components: auditability and permission management, and they can be viewed as two ends of the spectrum with a sliding scale in the middle. Organizations that rely entirely on auditibility give unrestricted access and then depend on those audit capabilities to ensure security. On the other end, organizations that rely entirely on permission management tightly control who has access to what at the source.

It's important to recognize that auditability is a matter of setup — it's a fixed cost when building a system. On the other hand, permissions have both a fixed cost and an ongoing cost of constant management, which requires lots of internal processes and tooling. When it comes to scaling AI, finding the appropriate balance between the two is the challenge for IT teams and their managers.



Many positions along the auditability/permission management axis make sense, but it depends on what the business is, what kind of data it has (e.g., website data is less sensitive than financial data, and jet engine data is somewhere between the two), and what tradeoffs the business and IT teams are willing to make.

In an ideal world, we would tell you based on Dataiku's hundreds of clients what this ideal balance is, and the challenge would be solved. But the reality is that to scale AI, IT managers need to conduct their own cost/benefit analysis and determine where on the sale is right for their specific scenario (this balance might even be different for different business units).

Deciding Which Technologies to Use

Innovation in IT and by IT teams is often more important to the scaling of AI than innovation in machine learning itself. That is, ultimately, most business problems won't be solved by the latest, greatest, and fanciest machine learning techniques (like deep learning or neural networks), but rather by relatively classic techniques applied in smart ways to solve business issues. However, which technologies IT chooses for data processing, storage, etc., can make a big difference in the speed at which teams can execute on AI initiatives.

That being said, it's important for IT managers to embrace innovations and new technologies, but at the same time, to be very selective about which technologies they choose. Having a deep understanding of what the business is trying to achieve versus what the trendy technologies are made for (and what they can — or can't — do) is critical.

And while making innovative choices in technology could be the key to scaling AI, IT managers also need to be reasonable. Jumping on every new technology under the sun can be confusing for those trying to develop consistent processes and systems for AI (even with an abstraction layer, e.g., AI platforms like Dataiku, on top).

Dataiku is a powerful, tech-agnostic, and expandable platform that sits on top of your tech stack, ensuring a consistent experience for end-users no matter what the underlying changes in or choices of technology. Dataiku:

- Supports all major solutions in IT and data science.
- Facilitates the progressive deployment of new technologies.
- Can be customized to build your own integrations.

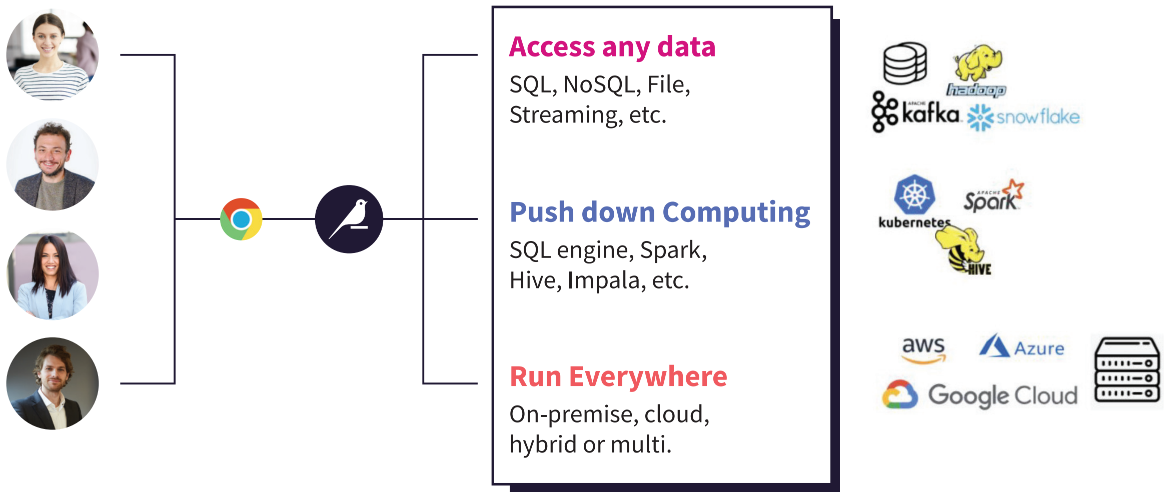


Figure 4: How Dataiku supports organizations' tech stacks, now and in the future.

Whether to Buy End-to-End or Stitch Together Best-of-Breed Tools

Today, most organizations won't consider fully building an AI platform solution. One of the biggest reasons is because of the hidden technical debt in machine learning systems identified by Google[5], which illustrates the sheer complexity of the endeavor. In other words, there is so much "glue" — so many features that are outside the core functionality of simply building a machine learning model — that building all of them from scratch to have an AI platform that truly allows for the scaling of AI efforts is prohibitively challenging.

[5] *https://papers.nips.cc/paper/2015/file/86df7dcfd896fcaf2674f757a2463eba-Paper.pdf*

Building a modern AI platform (and therefore scaling an enterprise-wide AI strategy) for most organizations today boils down to two options:

1. Buying one end-to-end platform for data science, machine learning, and AI that covers the entire lifecycle (Figure 5), from the ingestion of raw data to ETL, building models to operationalization of those models and AI systems, plus the monitoring and governance of those systems.
2. Buying best-of-breed tools for each of the steps or parts of the lifecycle and stitching together these tools to build the overall platform that is more customized for the organization and its needs.

Note that in many cases, the second option is situational, meaning it's dictated by existing investments (i.e., we already have tools for x, y, and z, what can we add to complete the stack and how can we tie it all together?) rather than driven by explicit choice in making new investments that are the best fit for the organization's needs.

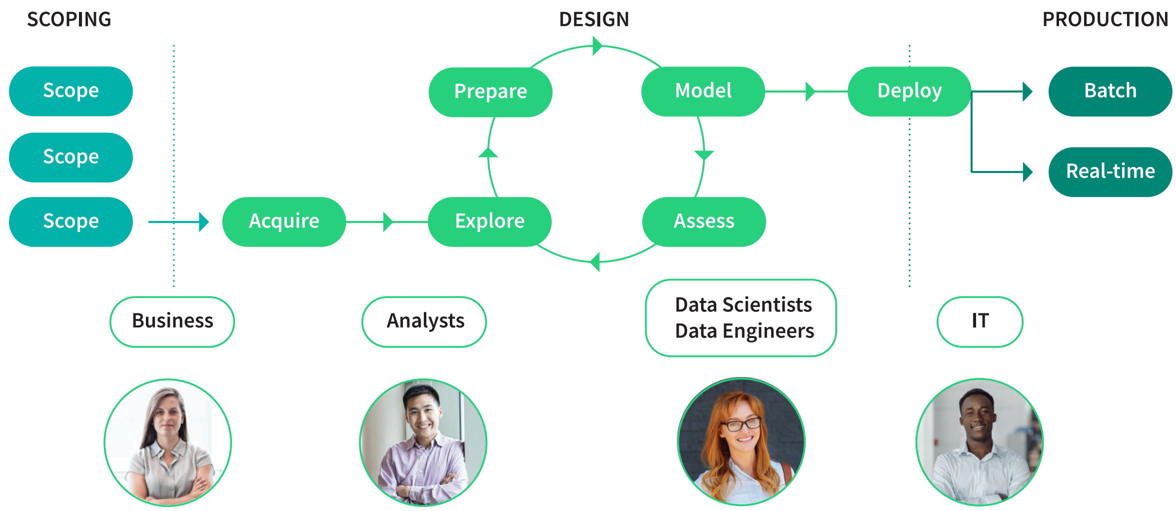


Figure 5: A representation of the data science, machine learning ,and AI project lifecycle..

Providing the very best tool for ETL, the very best for AutoML, for deploying to production, etc., will allow each team to choose the technology they want to work with, which is a tempting prospect when attempting to keep everyone happy — getting consciences across an organization is, admittedly, no easy task. However, the "glue" between these components, while not as complex as building everything from scratch, remains a huge challenge.

Besides the glue problem, there are also important components of the end-to-end lifecycle that are lost when moving from tool to tool. For example:

Data lineage is difficult to track across tools. This is problematic for all organizations across industries, as visibility and explainability in AI processes are crucial to building trust both internally and externally in these systems (and for some highly regulated industries like financial services or pharmaceuticals, it's required by law). With option two as outlined above, it will be difficult if not impossible to see at a glance which data is being used in what models, how that data is being treated, and which of those models using the data are in production vs. being used internally.

Stitching together best-of-breed tools can also **complexify the handoff between teams** (for example, between analysts and data scientists following data cleansing, or between data scientists and IT or software engineers for deployment to production). Moving projects from tool to tool means some critical information might be lost, not to mention the handoff can take longer, slowing down the entire data-to-insights process.

As a follow up to team handoffs and collaboration between data practitioners, another challenge is the **pain of managing approval chains between tools.** How can the business reduce risk by ensuring that there are checks and sign-offs when AI projects pass from one stage to the next, looking for issues with model bias, fairness, data privacy, etc.?

Option two also means **missed opportunities for automation** between steps in the lifecycle, like triggering automated actions when the underlying data of a model or AI system in production has fundamentally changed.

In the same vein, **how do teams audit and version the various artifacts between all these tools?** For instance, how does one know which version of the data pipeline in tool A matches with which model version in tool B for the whole system to work as expected?

Given the aforementioned challenges, the energy organizations put into building a modern AI platform shouldn't be spent cobbling together tools across the lifecycle, which ultimately results in losing the larger picture of the full data pipeline (not to mention adds technical debt). Instead, investing in an end-to-end platform for AI (like Dataiku) provides cost savings via reuse, the ability to focus on implementing high-impact technologies, smooth governance and monitoring, and more.

Of course, the fear that comes with investing in one end-to-end platform is that the organization becomes tied to a single vendor. This isn't a small risk and is not to be overlooked — lock in is a real consideration, as the company becomes dependent on that vendor's roadmap, decisions, and more.

To that end, it's important to invest in end-to-end technology that is open and extensible, allowing organizations to leverage existing underlying data architecture as well as invest in best-of-breed technologies in terms of storage, compute, algorithms, languages, frameworks, etc.

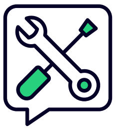
How to Kickstart Organizational Transformation Around AI

Scale Push to Production

Hands down one of the biggest steps IT teams must take to allow for organizational change around AI is scaling the ability to put entire AI projects into production. This includes not only the processes and tools for doing so, but also educating more people across the business about what pushing to production actually means so that they are educated and aware of the benefits, work involved, and risks.

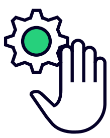
Business leaders view the rapid deployment of new systems into production as key to maximizing business value, but this is only true if deployment can be done smoothly and at low risk. Continuous integration and continuous delivery (CI/CD) concepts apply to traditional software engineering, but they apply just as well to data science, machine learning, and AI systems.

After successfully developing a model, a data scientist should push the code, metadata, and documentation to a central repository and trigger a CI/CD pipeline. An example of such pipeline could be:



- Build the model
    - Build the model artifacts
    - Send the artifacts to long term storage
    - Run basic checks (smoke tests/sanity checks)
    - Generate fairness and explainability reports

- Deploy to a test environment
    - Run tests to validate ML performance, computational performance
    - Manual validation



- Deploy to production environment
    - Deploy the model as canary
    - Fully deploy the model

Many scenarios are possible and depend on the application, the risks from which the system should be protected, and the way the organization chooses to operate. Generally speaking, an incremental approach to building a CI/CD pipeline should always be preferred — i.e., a simple or even naïve workflow on which a team can iterate on is often much better than starting with complex infrastructure from scratch.

A starting project does not have the infrastructure requirements of a tech giant, and it can be hard to know upfront which challenges deployments will face. There are common tools and best practices, but there is no one-size-fits-all CI/CD methodology. That means the best path forward is starting from a simple (but fully functional) CI/CD workflow and introducing additional or more sophisticated steps along the way as quality or scaling challenges appear.

—> Explore CI/CD in Dataiku

Keep Architecture Simple

When it comes to architecture for supporting AI systems, things can get complicated relatively quickly. Sometimes, it's simply the result of legacy systems — in enterprises with complex organizational structure and lots of history, it's impossible to start from scratch, and things might already be messy before getting started. Of course, developing a best-in-class AI platform wouldn't be so difficult if one was starting with a blank slate!

But oftentimes teams add complexity because they want to work with certain technologies or try certain things, even if the business needs don't require something so intricate, and this is problematic. Overly complicated systems can seriously hinder Enterprise AI efforts as it becomes increasingly more difficult to implement additional tools and exponentially harder to maintain over time.

Today, many companies in the United States aren't building the systems themselves but rather working with consultants for support. Regardless, the message is the same: keep it simple in terms of architecture so that even as technologies ebb and flow, switching between them is seamless both for the IT team and for the businesses they are supporting.

Build Sound MLOps Strategies

It's one thing to smoothly deploy the first versions of models, but what about the next ones? How do people in the organization make decisions to upgrade models, and who is responsible for it?

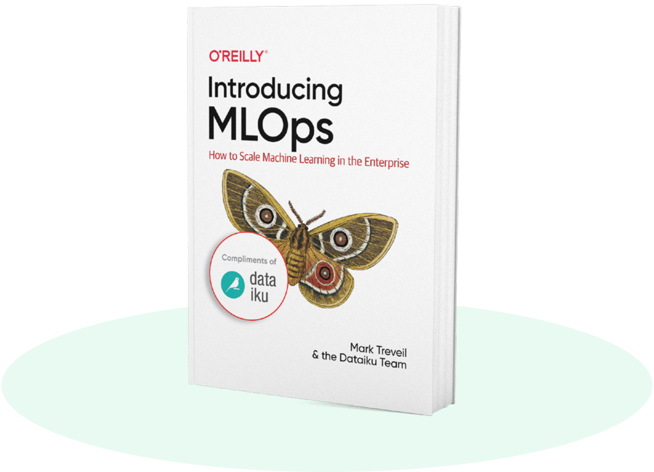
While MLOps often gets lumped in with data or AI governance, the two are not the same. While governance (practices and processes ensuring the management of data assets within an organization) is largely owned by IT managers and their teams, almost everyone in the organization — including IT teams — has a role to play in MLOps (the standardization and streamlining of machine learning lifecycle management).

MLOps isn't just important because it helps mitigate the risk of machine learning models in production (though that is one good reason to develop MLOps systems), but it is also an essential component to massively deploying machine learning efforts and in turn benefiting from the corresponding economies of scale. Going from one or a handful of models in production to tens, hundreds, or thousands that have a positive business impact will require MLOps discipline.
Good MLOps practices will help teams at a minimum:

- Keep track of versioning, especially with experiments in the design phase.
- Understand if retrained models are better than the previous versions (and promoting models to production that are performing better).
- Ensure (at defined periods — daily, monthly, etc.) that model performance is not degrading in production.

For IT managers and their teams, MLOps needs to be integrated into the larger DevOps strategy of the enterprise, bridging the gap between traditional CI/CD and modern machine learning. That means systems that are fundamentally complementary and that allow DevOps teams to automate tests for machine learning just as they can automate tests for traditional software.

—> Go further on scaling AI with robustMLOps processes in O'Reilly **Introducing MLOps**



Stories

- **AI in 2020: The Architecture and the Infrastructure**
- **Data Warehousing in the Cloud (Snowflake)**
- **Johnson & Johnson: Getting from 1 to 100 Users With Dataiku**
- **Architecting Scalable Self-Service Analytics Platforms with Shailesh Doshi, Pivotal Software**

Dataiku: A Must-Have Data and Analytics Collaborative Tool

> "Dataiku put the data manipulation and data science work in the hands of the business, helping the data and analytics democratization in the organization."
>
> *— IT Lead in the Healthcare Industry.*[6]

Storage and Compute Agnostic

Dataiku can run on-premise or in the cloud — with supported instances on Amazon Web Services (AWS), Google Cloud Platform (GCP), and Microsoft Azure — integrating with storage and various computational layers for each cloud.

Any IDE, Git-Enabled

Dataiku provides an integrated development environment for Python, R, Julia, and Scala from which you can transparently access data sources without having to manage connectivity issues. Leverage Dataiku

- In a "Notebook" style (with Jupyter Notebook).
- In a "Visual Flow" style (by creating a flow of computation represented graphically in the tool).
- By connecting your own IDE (SublimeText, Visual Studio) to the platform.

All developments can be managed in Git.

Spark & K8S Clusters: Fully Managed (at Scale)

Dataiku can either leverage existing Spark and Kubernetes clusters or create and manage its own clusters (leveraging cloud platforms).

[6] *https://www.gartner.com/reviews/market/data-science-machine-learning-platforms/vendor/dataiku/product/dataiku-dss/review/view/3445445*

Powerful Extensions via Dataiku Plugins

Dataiku Plugins enable developers to take control and expand any part of the platform by building powerful extensions to out-of-the-box functionality using Python or Java. Dataiku plugins can help connect to new data sources, provide and encapsulate a new algorithm visually for non-coders, integrate an IT process within Dataiku, and much more. Dataiku can be further extended via APIs, and it integrates with Jira and Jenkins.

Build a Robust Data Architecture, End-to-End

Dataiku architecture is built around a pattern that systematizes the push down of computation into existing technologies, and it provides all the building blocks to enable data architects to build their own robust data architecture:

- Data validators to protect the architecture against changes in underlying data sources.
- Robust deployment with auto-scale, versioning, and rollback for both batch data pipelines and real-time model scoring.
- A smart data reconstruction engine for efficient incremental data recomputation.

Automate and Monitor With APIs

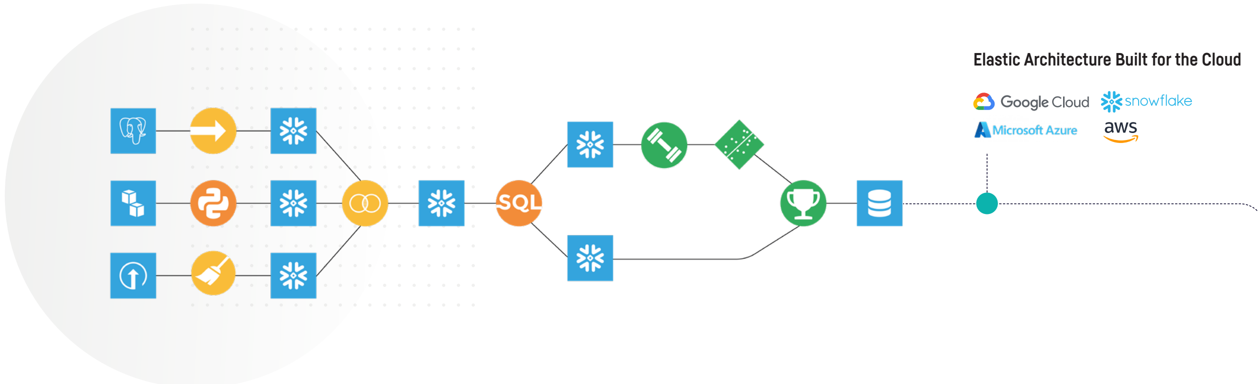
Dataiku provides an extensive API for platform setup, administration, and deployment (including automating the deployment of the full solution or new services). Administration extensions let you integrate Dataiku within your existing monitoring IT stack.

Closing Thoughts

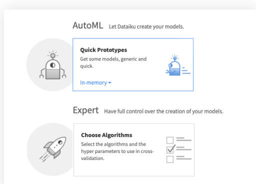
IT managers who are able to bring different profiles together — including data scientists but also analysts and business stakeholders — over the common goal of scaling AI initiatives will be in high demand worldwide as the need for organizational change around AI increases.

Those that manage to address the challenges as well as enact some of the changes brought up here will not only spark real change in the business, but they will also create career opportunities and demand for themselves as agents of change in the quest to scale AI.

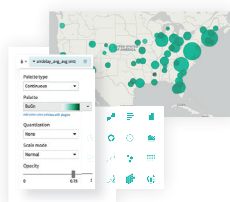
Everyday AI,
Extraordinary People



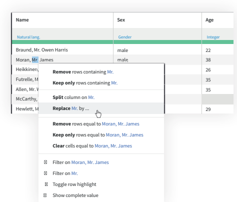
Machine Learning



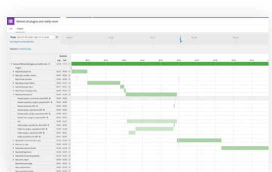
Visualization



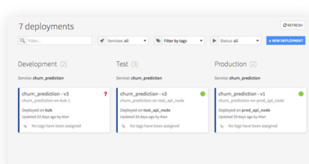
Data Preparation



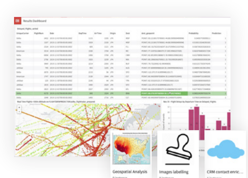
DataOps



Governance & MLOps



Applications



45,000+
ACTIVE USERS

450+
CUSTOMERS

Dataiku is the platform for Everyday AI, systemizing the use of data for exceptional business results. Organizations that use Dataiku elevate their people (whether technical and working in code or on the business side and low- or no-code) to extraordinary, arming them with the ability to make better day-to-day decisions with data.

EBOOK