GE Aviation

EBOOK

GE Aviation: From Data Silos to Self-Service

A Deep Dive into the Processes, People, and Technology that Enabled GE's Data Revolution

An Ebook by Dataiku in Collaboration with GE

www.dataiku.com

Introduction

Any company that wants to make any impactful change today - 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.

According to Forbes,[1] in 2018, humans and their systems produce around 2.5 quintillion bytes of data a day (by the way, a quintillion is 1018). Most of this data lies in the hands of companies, and the ones that are able to make radical business change today are those able to harness massive amounts of data and turn it into insights at scale.

This is easier said than done - transformation at this level doesn't simply mean slapping data on top of existing processes; it involves fundamental organizational change, weaving data into the very fabric of the company. To date, despite the hype around artificial intelligence (AI) in the media, very few businesses have managed to execute on incorporating the fundamental machine learning (ML) processes that enable these data insights at scale, much less automating them to enable AI services.

This white paper tells the story of GE Aviation, a company that bucks this trend and that has, at a large scale, been able to empower the organization - not just at a high-level, but down to the individual level - to use data for day-to-day processes. Specifically, it will cover:

- *How GE Aviation developed a self-service data program using Dataiku and a suite of tools that unlocks employees' ability to use data to get insights quickly.*
- *The technological stack and organizational setup at GE Aviation that enable these systems.*
- *The lifecycle of a data product at GE Aviation.*
- *How GE Aviation handles data governance and data education (including suggestions for employee onboarding material for a self-service data system).*
- *The return on investment (ROI) GE Aviation has seen from their data initiatives.*

[1] *https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#652aff2860ba*

GE Aviation

Fast Facts

Subsidiary of

General Electric

Headquarters

Evendale, Ohio
(United States)

Industry

Aerospace

Total Employees

40,0000

Users

1,841 users* of the self-service analytics system

**includes true users, i.e., excludes IT team members or other administrative functions*

Data Products

2,000+ created since March 2017

Datasets

130 published in the last year

Projects

218 in Dataiku automation

(plus 450 more automated functions in Starfish)

Key Contributors

Self-Service Analytics at GE Aviation

Somesh Saxena

Product Owner of Dataiku and Alation at GE Aviation[2]

Somesh Saxena is the Product Owner of Dataiku and Alation at General Electric Aviation.He manages a team of full-stack data engineers and helps lead the Self-Service Data Program. Somesh supports a community of over 1,400 self-service developers building digital products to make data-driven decisions.

Somesh has trained over 700 employees through the Digital Data Analyst training, which teaches digital tools, data science, and process excellence. Somesh is front and center of the digital cultural transformation at General Electric Aviation. He began his career with General Electric's Digital Technology Leadership Program exploring different areas of the business. He led projects for the company's customer portal; did full-stack web development in Cyber Security; and data ingestion, engineering, and visualization in the data analytics space.

Somesh is a Certified Scrum Product Owner from the Scrum Alliance. Somesh holds a degree in Business Administration with a concentration in Information Systems from the University of Cincinnati.

[2] *https://www.linkedin.com/in/someshsaxena*

Jon Tudor

Senior Manager of Self-Service Data Engineering and Analytics at GE Aviation[3]

Jon Tudor is the Senior Manager of Self-Service Data Engineering and Analytics at General Electric Aviation. He founded the Self-Service Data Program in 2016 and now leads the team, implementing six innovative products that enable over 1,500 users to create their own data and analytics solutions. Joining GE Aviation in 2009 as an intern, he completed the Information Technology Leadership Program in 2014, and has since held roles spanning data ingestion, big data architecture, cloud application automation, and self-service data and analytics.

Jon graduated with a BS in Business, majoring in Management Information Systems, from Miami University in 2012 with Magna Cum Laude and Honors. He's always happy to connect with others interested in digital transformation through data and analytics, so feel free to reach out on LinkedIn!

[3] *https://www.linkedin.com/in/jonathan-tudor/*

Self-Service Analytics at GE Aviation:

So Much More than BI

Sometimes when people think of self-service analytics, they often still think of old-school business intelligence (BI), which is often extremely limited only to historical data. On top of that, it can generally only be used to create rather static dashboards that ultimately don't provide much utility to the business as a whole. In fact, just a few years back (circa 2015)[4], industry leaders and analysts were still hailing BI platforms as the end-all-be-all of data-driven transformation.

But today, self-service can be (and is) so much more than that, and that's especially true at GE Aviation. **In a general sense, self-service is the system by which line-of-business professionals or analysts can access and work with data to generate insights (predictive or not) as well as data visualization with little direct support from data scientists, IT, or a larger data team (though the self-service platform itself should be supported by these personas).**

> *"More than 87 percent of organizations are classified as having low business intelligence (BI) and analytics maturity, according to a survey by Gartner, Inc. This creates a big obstacle for organizations wanting to increase the value of their data assets and exploit emerging analytics technologies such as machine learning."*
>
> *Gartner Press Release, Gartner Data Shows 87 Percent of Organizations Have Low BI and Analytics Maturity, December - 2018,*[5]

[4] *https://www.forbes.com/sites/louiscolumbus/2015/02/25/key-take-aways-from-gartners-2015-magic-quadrant-for-business-intelligence-and-analytics-platforms/#3a9bf89559aa*
[5] *https://www.gartner.com/en/newsroom/press-releases/2018-12-06-gartner-data-shows-87-percent-of-organizations-have-low-bi-and-analytics-maturity*

And indeed, GE Aviation has implemented their own version of a self-service system that serves their specific needs and requirements and that allow them to use real-time data at scale to make better and faster decisions throughout the organization:

- Engineering is using data from these tools to redesign parts and build jet engines more efficiently.
- Supply chain is using it to get better data insights into their shop floors and streamline supply chain processes.
- Finance is using it to understand key metrics such as cost, cash, etc.
- The commercial group (by leveraging data scientists) is using these tools to transform engine sensor data from customers and build analytics services for them.
- The data initiative at GE Aviation is called Self-Service Data (SSD), but in fact encompasses both self-service in the traditional sense as well as an element of **operationalization** (that is, the process of converting data insights into actual large-scale business and operational impact) for both business lines and IT users.

The SSD at GE Aviation is, in a nutshell:

- The ability for everyone (with proper access rights) to discover and use data, prepare that data, and create a data product, including developing predictive models within Dataiku.
- The ability for data product creators to share their work with other colleagues.
- The ability for data product creators to deploy data pipelines in production using macros developed in Dataiku.

Ultimately, the teams at GE Aviation have developed the SSD in such a way that anything users want to put in production, they can - provided it passes a set of checks and balances to ensure it meets database administration and data governance standards. They were able to do this through the way they chose and configured technology and their Dataiku instance, but equally important, how they built organizations around the initiative for support.

But it wasn't always this way.

The History of Self-Service Data at GE & The Digital League

Someone on the business side is responsible for giving numbers to leadership. One day, they need to deliver something specific to the leadership team, so they go to IT and ask for a report or a dataset. The IT team gives them an extract of data in Excel.

business user does his or her best with that Excel extract, comparing that data to similar data from past reports or to others' data extracts to try to figure out if it's the right data and if it's accurate. He or she then puts together a report in PowerPoint for the leadership team. Rinse and repeat about once a month.

Sound familiar?

Until 2015, this was the process of using data at GE Aviation as well. Following this process, they moved on to building reports (either weekly or monthly) in Spotfire, with the deliverable to leadership being screenshots of Spotfire - still delivered in PowerPoint.

But they still had trouble scaling these efforts not just because of a technological barrier, but also because of:

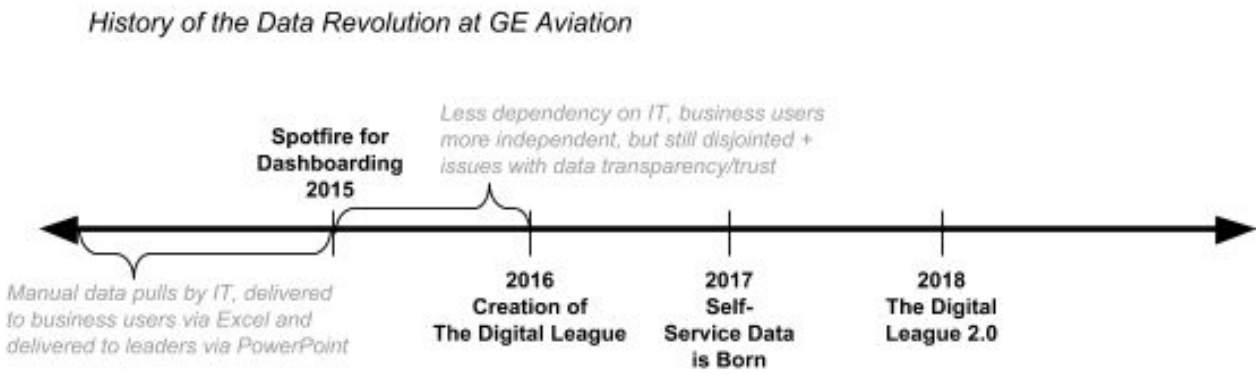
- Lack of trust and transparency: With limited visibility into how the data was being processed or where it was coming from, leadership would question the data (or the logic behind it). And there was no easy answer to these concerns - when people questioned data, there was no real source of truth. This undermined all of the efforts put forth by the business teams.
- Silos: Not only was data siloed, but individual business users also each had their own dashboards (there was no shared, central repository).
- Repetitive Efforts: Because of the aforementioned silos, by nature, lots of efforts ended up getting repeated over and over again. Business users didn't have any visibility into projects and analysis that had already been done, which caused a lot of time to be lost.
- No central vision: Ultimately, the company didn't have one overarching idea or standard for using data, and with a lack of vision, no one individual business user, much less an entire line of business, could really move forward and execute properly.

> *"The Digital League completely transformed the culture of data at GE Aviation."*
>
> *-Somesh Saxe*

GE Aviation's data revolution began in 2016 when they built The Digital League, a cross-functional team (made up of leaders from supply chain, finance, IT, and engineering lines of business) that came together under one central vision and strategy and, importantly, in one physical location. Their goal was to create data-driven products, and from the time of their creation on, anything data-related went through this team.

The Digital League worked to spark the digital data revolution at GE Aviation because it:

- **Broke down silos:** The team established itself as the principal place that controlled all data initiatives, and this helped ensure a central vision and necessary control over data processes. When getting started, the idea and practicalities of having one team owning the initiative was critical.
- **Emphasized communication and collaboration:** The Digital League was a cultural revolution - normally, these lines of business (supply chain, finance, IT, and engineering) sat physically in different buildings, communicating almost entirely through email. Having them together meant that they actually fundamentally worked together, each intimately understanding the role of the others through regular demo days and Agile methodology where all representatives were present at daily standups.
- **Drove infrastructure decisions:** In addition to driving data culture, The Digital League also drove data infrastructure decisions (specifically building a data lake) that set the stage for future innovations. Having one common repository and the data all in one place made a huge difference into changing GE Aviation's data initiatives into what they are today.



The Birth of SSD:

Technological & Organizational Setup

Following the successful launch of The Digital League and its growth over the year after founding, the SSD initiative was born in late 2016 out of the need to provide scale to data initiatives beyond the Digital League. That is, GE Aviation needed to start democratizing: bringing data into the hands of everyone at the organization in addition to providing the central vision and infrastructure. Out of this need, the SSD was born. And today, 90 percent of the users of the SSD are outside of The Digital League.

How did GE Aviation get there?

Self-service initiatives often fail in large enterprises for a variety of reasons (ongoing issues with data access, insufficient tooling - or tooling that doesn't meet users' daily needs, lack of data accuracy or data confidence, data security problems, a complete lack of connection between self-service and operationalization, etc.). All of these issues boil down to a larger problem: self-service gets treated as a one-time project that gets launched, then forgotten.At GE Aviation, this is far from the reality - they view their self-service initiative as an ongoing one that requires support and continuous improvement to find continual success. Therefore, there are two teams at GE Aviation that support their robust self-service initiative:

1 The Self-Service Data Team, who are responsible for:

- User enablement - training and best practice documentation (more on this later).
- Tool administration (including Dataiku), instance sizing, and usage monitoring.
- New process development and identification of opportunities for process automation.
- Ensuring the smooth deployment of data products.

The team supports over 1,800 users today. They are a centralized group based at the company's office in Cincinnati, but they work to support users around the globe. This team ensures that nothing is blocking people from using SSD, whether that be an initial knowledge gap or technical issues along the way. That means teaching users to do things for themselves instead of simply acting as a more traditional help team who takes tickets and solves issues on behalf of the users.

Importantly, they also ensure that the initiative keeps its momentum - that is, that it doesn't become outdated or stale and continues to evolve with the needs of the business and the users - by introducing new automations and process improvements to reduce repetitive work across the board.

For example, the team recently turned the rather arduous process of triggering data product deployment (which involved manually switching the product's environment, making sure any scenarios were turned on and running manual checks on those scenarios, etc.) into an automated process so that macros do all of these checks behind the scenes and users can simply automate their data product with a click of a button in Dataiku.

This allows users a more instant satisfaction of deployment to production, plus the Self-Service Data Team can spend more time on other priorities, like supporting any issues in production and improving education around the tools and processes.

2 The Database Admin Team, who are responsible for:

- Ensuring that data products going into production follow basic data governance policies, including naming conventions and data access rules.
- Checking that any data used in deployed self-service projects are being used appropriately.
- Along with the Self Service Data Team, helping with user support in case of any failures in deployed projects (e.g., data logic changes, etc.).

It's worth noting that neither of these two teams is responsible for evaluating the self service users' projects from the perspective of business need or business use case. That's because in the case of GE Aviation, those users are the business experts - they know best about what their needs are.

The job of the Self-Service Data and Database Admin Teams is simply to enable them to do what they need to with the data and ensure there are no technical roadblocks to using that data.

From a technical standpoint, the structure that enables these two teams to function and to support the SSD and its users consists of:

- Greenplum and Hortonworks/HIVE (for database management).
- Dataiku (for designing and deploying data products - users access, clean, and manipulate data through Dataiku).
- Alation (for data cataloging and search).
- Daasboard (an in-house tool built for monitoring the state and status of all data products).
- Spotfire (visualization of final data product for end business users).
- Starfish (an in-house tool built to enable automation of database functions and ingestion of data to the data lake).
- Published Data (reusable business domain data to act as a starting place for data exploration).

“SSD at GE Aviation was born out of a conversation in a conference room. The idea was that you would never be able to hire enough data professionals to meet the data demands of the business, so instead, why not turn the business into data professionals. Taking that premise we started to define what self-service meant for us and how it would work. ”

-Jon Tudor

The Two Prongs of Data Enablement

Today, the SSD and The Digital League both exist in parallel at GE Aviation - that is, the development of a self-serve solution did not eliminate the need for The Digital League. In fact, The Digital League continues to be a vital piece of the overall digital and data strategy at GE Aviation, managing larger initiatives that touch multiple parts of the business as sort of a top-down strategy that rounds out the SSD's bottom-up approach.

> *"The Digital League is the cultural piece [of the data-driven approach] – people are impressed that we have that in a 127-year-old company. It takes a lot to innovate, and The Digital League is enabling that; they are the next step in the innovation of manufacturing."*
>
> *-Somesh Saxena*

The Digital League, for example, might have five key metrics or objectives they are focused on in a given year. There might be other metrics or data initiatives that are also important, and the SSD will work on enabling and supporting those in other teams and departments around the company. The support provided ranges from more full-service support (help with data engineering, guidance on projects, etc.) to simple tools access that the SSD will provide.

And sometimes it works in the opposite direction as well - the SSD might support a data project for a particular team that becomes a larger priority for the company, in which case it would be taken over by The Digital League.

The Data Product & Its Lifecycle at GE Aviation

At GE Aviation, a data product generally consists of a dashboard that is consumed by users in the business units where data is coming in and updated in batches - weekly or daily (as opposed to a dashboard that looks at or shows historical data). These so-called dashboards - which are far from the well-known, static, BI-sense of the word - provide automated information flows powered by Dataiku on critical business functions (see the next section for specific use cases).

But how does an employee in a line of business department at GE Aviation go from idea to raw data to business-impacting data flows?

1 Datasets are readily available to those with appropriate access and are tagged by type for simple cataloging:

- Consumption dataset: The output of deployed data product that is consumed in a dashboard.
- Published dataset: A set of tables published and maintained by the self-service team that are domain-specific and applicable to multiple use cases. These data sets are created by collaborating with data stewards from the business domain - they belong to, and are considered the source of truth for, that domain. They offer a starting point for most self-service users instead of them recreating the same business logic by working with raw data.
- Base: Base data that is replicated from source systems throughout the organization and outside through Extract Load Transform (ELT) to enable full exploration of the data.
- Transform: Interim datasets that are created as part of a series of data engineering steps
- Sandbox: Exploratory datasets that are in development.

2 Any self-service data user can use available datasets to create projects in a design environment in order to explore that data and test various data pipelines. However, there are no scenarios running on the design environment.

- That means if the user wants to have the data for that project automatically updated daily, weekly, etc., they will need to go through necessary compliance and performance checks to have the data product deployed, as described in the previous section.
- Data products in the design environment can, however, be shared with other users.

3 Once the user decides to promote from the design to production environment, they can do so at the click of the button, which triggers a series of macros that conduct.

- Checks to ensure the data product is fit to be in production.
- Checks include verifying naming conventions, checking the schemas being used, ensuring the proper data distribution, verifying the size of the datasets being used and implementing partitioning if needed, plus memory and CPU consumption of the data product.
- Whenever a data product fails a check, the user receives a clear and descriptive error message with instructions on how they can adjust the data product to fix the error and pass the mandated checks.
- Note that these checks used to be done manually by the Database Admin and Self-Service Data Team, and the process involved a lot of back-and-forth between that team and the users that slowed down the process. Since the Self-Service Analytics Team automated this task with macros, the system has become even more self-service in the true sense of the word.

4 When the data product passes all checks, it is pushed to the production environment automatically.

- Of the 1,900 (and growing) data projects that have been built using the GE self-serve data system, 202 (10 percent) are in production.

Importantly, the GE SSD program not only has a smooth and largely automated process to remove friction from going from a design to production environment with data products - they have a similar process when something breaks to ensure that data products in production don't become obsolete.

This is critical given that data systems (and the underlying data itself) are constantly changing - a self-service data system that doesn't have a way to handle these changes can quickly become problematic.

At GE Aviation, if something breaks down and affects a data product in production, the system kicks off the following automated process:

1. An email gets sent to the Self-Service Data Team as well as the data product owner.
2. The system automatically tries to rerun the scenario(s).
3. If the automatic rerun fails, the data project owner is notified automatically to fix the problem.
    - The data project owner can use GE's homegrown monitoring tool, Daasboard, for self-service troubleshooting or to understand what the problem might be. This system gives the ingestion status of input datasets (failures versus successes) as well as information on all third-party platforms (Spotfire and Dataiku) plus server health.
    - If, following an error and consulting Daasboard, project owners are not able to fix the problem on their own, they work with the SSD Team to get to the bottom of the issue and get the project back up and running.

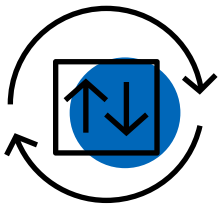
By far the number one reason that self-serve data efforts fall flat in other organizations is because of a lack of trust in the data. This could be because it's not updated regularly enough (ideally, it should be in real time), or because data and data formats change, but the company hasn't devoted ongoing resources to self-serve platform maintenance. This feedback system and way of addressing errors and issues allows business users to place even more confidence and trust in the GE SSD program.

> *"It's incredibly important to create and communicate common definitions and processes so that you and your user base are speaking the same language and to ensure that you use automation to enforce those definitions and processes. Otherwise you'll have chaos instead of outcomes."*
>
> *-Jon Tudor*

Use Cases for Self-Serve Data at GE Aviation

Many lines of business leverage the SSD program at GE Aviation to build their own data products for a range of different end goals, but here are three of the largest self-service users:

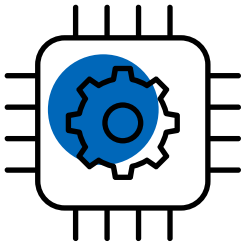
Supply Chain



Individuals who are responsible for the production or repair of hardware at GE Aviation are concerned with monitoring operations and improving the accuracy and speed of their decision making. Thus, operations-focused individuals want to know how to access their data quickly and automate trivial information-gathering processes, which is how they leverage the SSD program.

Modeling remains important for operations-focused individuals, but more often than not, they must prioritize “firefighting.” Also, given the nature of their work, data they need is often more difficult to obtain than other lines of business, which is also why they heavily leverage the SSD program.

Engineering



Engineers are involved with high-level analytical work and are not encumbered by the daily delivery demands that exist in supply chain. In addition, they already usually have access to the data they need.

However, the appeal of the SSD program in this department is that the datasets they deal with are often too large or too complex to be analyzed in Excel or Minitab. In addition, engineers are interested in automating information gathering and analysis and have a strong use case for the modeling techniques currently taught on one of the days of the advanced DDA: Digital Data Analyst Training (see The Role of Education for more on training).

Finance

Individuals in Finance spend a great deal of time and effort producing complex reports in Excel. The data for these reports exists both in the Aviation Data Lake and in the Finance Data Lake, so finance individuals need the ability to duplicate their Excel macros in Spotfire, build financial models/Monte Carlo simulations, and automate information flows from both data lakes.

The Role of Education

Many organizations who embark on a journey to provide self-service data to business users or to democratize data across the company make the mistake of thinking intuitive tools can reduce (or worse, replace) the need for proper training and continuing education around the initiative.

The SSD program at GE Aviation has taken quite the opposite approach, which has proved incredibly successful in both getting high adoption among business users as well as sustaining use of the tools and program over time. From onboarding new users to ensuring that existing users have the tools and resources they need to support themselves, the GE self-serve philosophy is heavily focused on education as a means to enablement.

Onboarding

Users of the GE SSD program can have a wide range of skills and responsibilities (the Use Cases section covers just a few); thus, job titles of those users can vary from business analyst to mechanical engineer to product quality manager to brilliant factory leader for a specific site, and more. But regardless of role and background, all GE SSD users go through the same training, DDA: Digital Data Analyst Training. As of the end of 2018, more than 1,000 employees have gone through the robust training program.

The courses currently available include:

- DDA 100: Largely self-guided, this introductory program allows potential SSD program users to watch videos to get an understanding of basic tools and programs available to them.
- DDA 200: An intensive, week-long course that teaches digital data tools, data science, and process excellence. The course includes sessions like: introduction to the data lake, data lake architecture, how to write SQL, tool deep dives (e.g., Dataiku) including how-to demos and best practices, hands on lab sessions for data wrangling (join, filter, stack, etc.), stats 101, Six Sigma, and more.
- DDA 300: A quarter-long program going more in depth on many of the topics also covered in DDA 200, especially data science topics.
- DDA for Executives: A full day training that is higher-level and focuses on the value proposition of the SSD program and how/why teams and individuals partake.

Since January 2018, DDA 200 has had 1,050 graduates

Ongoing

In addition to formal onboarding training (DDA), the GE SSD Team is also very concerned with - and spends a lot of time on - improving materials surrounding ongoing education and training. For example, they are careful that for each and every error that the system could produce that they have documentation surrounding the error and how it can be fixed.

For example, one of the automatic checks that happens when a user wants to deploy a specific data product to production is for table size. The Database Admin Team needs to make sure that tables deployed in production have no impact on the larger environment, which means any table over 200 Gb must be partitioned - that is, the table must be split based on different columns, creating subsets of the table.

But end users don't necessarily know or understand how table size might impact the system overall (much less how to partition a table), so an error message at this time about the table being too big isn't very helpful. Instead, the team devotes the time and resources to ensure that errors are specific and explain what exactly the problem is, linking to additional help or how-to if necessary. Taking it a step further, they have built macros within Dataiku to solve these problems; so there is a partitioning macro, and users can partition their table with the click of a button. This ongoing initiative is what really allows the GE SSD program to scale.

Gameification

Another important aspect to education and one of the reasons for the early success of the GE SSD programs is embedding gamification into the program. The SSD team developed a low-cost program to have individuals contribute to data quality and ultimately get early buy-in for the SSD program that increased user engagement and interest as well as prevented the initiative from falling flat due to lack of internal users or support.

> *"There's only so many times you can tell a data steward - hey, go in … and tag your stuff. Add documentation. So beyond that, the answer is gamification."*
>
> *-Somesh Saxena*

The Data Duel launched in summer 2017, and it was relatively simple: the SSD team started monitoring and tracking all user actions within Alation and Dataiku, and they rolled out a point system such that each time someone tagged a dataset, created new documentation, added a new recipe, started a new project, created a new dataset, etc., that person would receive a certain number of points.

More points unlocked the possibility to pass levels and get exclusive laptop stickers. The SSD team took the competition to the next level by adding a leaderboard so people could see others' accumulated points. The interest and involvement with the SSD program due to this gamification was undoubtedly a huge piece of the program's overall success.

> *"Gamification truly helped us enhance the experience within [our] tools, making users a lot more engaged and the environment a lot more fun to work with."*
>
> *-Somesh Saxena*

Later, the gamification program expanded even further with the Metadata Duel - a one-off challenge to encourage people to add all the metadata and documentation they could to Alation. Within the month of the contest, the SSD progam gained 21,000 titles and 1,200 descriptions for datasets with the participation of more than 70 subject matter experts.

The Role of Data Governance

With today's increasing concerns around data privacy, the role of data governance in the realm of self-serve data 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 self-serve component of a data-powered strategy before it ever gets off the ground - the key is balance.

> *"I frequently tell people. The hardest part of my job is balancing enablement and governance. It's like running a playground. You want to give everyone the ability to play and explore but keep them from hurting themselves and others."*
>
> *-Jon Tudor*

The team at GE Aviation uses Alation for data product governance. It is here where all data products are documented with a description as well as project owner and workflow of approvers. This is an important element to the GE SSD program strategy because it makes it possible from a more high-level perspective to understand how data is being used and where.

In addition, self-service follows the same data security model as the data lake environment; that is, anyone that has access to data in the lake can use that data in the tools provided by the SSD team (like Dataiku). All the tools are integrated with the data lake in a shared ecosystem.

Self-Serve ROI

In a study[6] sponsored by Teradata conducted by Forbes and McKinsey, large enterprises reported that data efforts improved company growth by just 1 to 3 percent on average. But still, in this study, only 37 percent of respondents could quantify the business case for big data analytics, while 47 percent could not and 9 percent reported "no clear vision." And yet another study by BCG[7] estimated 20 to 30 percent EBITDA gains for data-driven companies. So clearly this is something that today, most companies struggle with and that analysts themselves struggle to quantify as a whole.

So while most companies (and analysts) struggle with calculating big data ROI, including what exactly they should be measuring to get there, larger data still suggests that investing in data science (including self-serve data initiatives) and machine learning is well worth it.

At GE Aviation, the focus specifically surrounding the SSD program is in streamlining - making processes smoother and business operations overall more efficient. While they have been able to quantify their efficiencies and savings via the SSD 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 [SSD] makes their job easier."*
>
> *-Somesh Saxena*

There is agreement throughout GE Aviation that the SSD 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.

[7] *https://www.datanami.com/2015/09/08/wheres-the-roi-in-big-data/*
[8] *https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation.aspx*

Top 5 Tips from GE Aviation for Setting Up a Self-Service Data Initiative

From Somesh Saxena, Product Owner of Dataiku and Alation at General Electric Aviation:

1 **Understand the needs of the business** before building a self-service data program; regardless of the size of the company, the most important thing to solidify before trying to get a self-serve data initiative off the ground is buy-in from the business. That doesn't just mean getting business lines to sign-off, but really working with them from the beginning to understand their needs deeply and having them test tools and processes to determine what the best solution is. Only the stakeholders in different lines of business will be able to properly evaluate the functionality and ease-of-use of any self-serve data tools based on the skills of their teams. But beyond this, getting their input from the start means that come launch, those business stakeholders will already be invested and can help get the program off the ground.

> *"You have to focus on customers and get their buy-in and support. Focused wins from grassroots efforts in the business will provide the momentum to grow and get executive buy-in."*
>
> *-Jon Tudor*

2 **Education is critical:** Rolling out a self-serve data initiative involves more than just buying a tool and setting people loose. Without proper training and ramp-up, the initiative won't be successful. Certainly, having a team that's devoted to this training and education piece can be incredibly helpful - but it doesn't have to be an entire dedicated team (which isn't always feasible based on the size of the company or other resource constraints). But at the very least, someone running the initiative must ensure that education is a priority.

3 **Support is also critical:** Similarly, ongoing support for self-serve data initiatives is just as - or possibly more - important than initial education. And it can't just be a queueing system, but personal attention that allows a self-serve data user to not only get past a blocking issue or question, but learn more about why or how the problem came up and how it can best be avoided in the future. At some point, person-to-person contact and collaboration is critical: no amount of thorough documentation can make up for a hands-on experience. What's more, positive support experiences can only bolster the support of the program.

> *Whether it's demos, training sessions, whatever it takes - you have to be involved. That's the only thing that's going to drive that cultural transformation."*
>
> *-Somesh Saxena*

4 **Try for a top-down and bottom-up approach:** Grassroots efforts can be very effective. So can executive initiatives. So the best bet? Come at it from both ends. Individual contributors have to love the self-serve data solution and want to use it, but executives must be on board and pushing for it as well. That's the best way to ensure long-term success.

> *"One of the smartest things we did was to bring our customers, the business, in on day one to make the decisions. The 40 non-IT employees who decided with us what tools we would go forward with are our biggest champions and ambassadors in the business today. "*
>
> *-Jon Tudor*

5 **Branding the program is important:** It's critical to advertise self-serve data program success not just for a pat on the back, but to show that people are empowered and that people's jobs are made easier and more enjoyable. That will plant the seed for more and more people to want to use it, and the initiative can spread incredibly quickly. Similarly, branding a self-serve program as a process that allows for data to be used in an efficient and effective way has the opportunity to change a business from a cultural standpoint - especially in domains where people tend to question data, reports, or dashboards and don't understand how they're being produced.

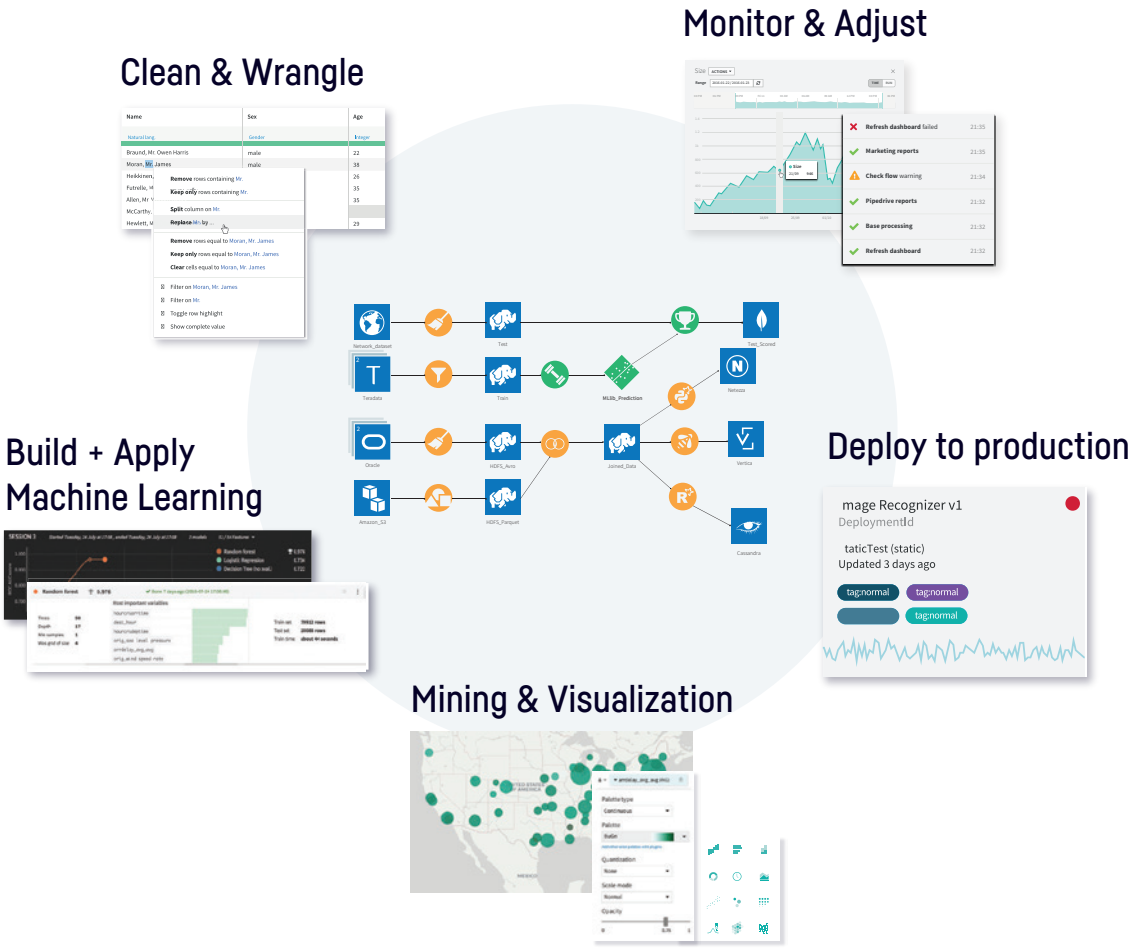
Conclusion

GE Aviation was successful in their self-serve data initiative partially because they brought in technology, but perhaps more importantly because they didn't ignore the fact that they needed to change people as well as processes in order to achieve their goals.

For further reading on scaling up data operations and democratizing data across a large organization, Dataiku recommends

- **Best Practices for a Successful AI Center of Excellence**
- **Guidebook: Staffing the AI Enterprise**
- **Enterprise AI for Business Preparedness**

Your Path to Enterprise AI



400+
CUSTOMERS

40,000+
ACTIVE USERS*
*data scientists, analysts, engineers, & more

Dataiku is one of the world's leading AI and machine learning platforms, supporting agility in organizations' data efforts via collaborative, elastic, and responsible AI, all at enterprise scale. Hundreds of companies use Dataiku to underpin their essential business operations and ensure they stay relevant in a changing world.

EBOOK

www.dataiku.com