

# RECIPES FOR



# SUCCESS

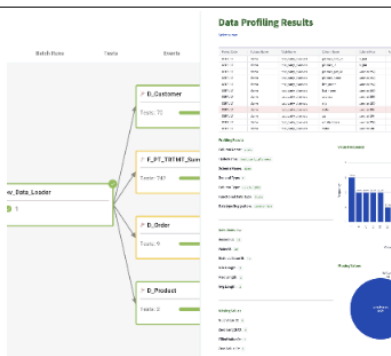
The Complete Guide to an  
Enterprise DataOps Transformation

by Chris Bergh, Eran Strod, and James Royster

# TRY DATAKITCHEN OPEN SOURCE DATA OBSERVABILITY

Data Quality Testing, Validation, Anomaly  
Detection, End-to-End Production Observability,  
and Alerting!

Install Now



# **Recipes for DataOps Success**

**The Complete Guide to an  
Enterprise DataOps Transformation**

**by Chris Bergh, Eran Strod,  
and James Royster**

Recipes for DataOps Success

*The Complete Guide to an Enterprise DataOps Transformation*

© 2021 DataKitchen, Inc. All Rights Reserved.

To order additional copies of this book:

**[info@datakitchen.io](mailto:info@datakitchen.io)**

Printed in the United States of America

Layout and cover design by John Perry and Ariel Plotkin-Gould



# Contents

<b>INTRODUCTION</b>	<b>3</b>
<b>EDUCATE</b>	<b>5</b>
Why Do DataOps . . . . .	7
<i>Better than Shake 'n Bake!</i> . . . . .	13
For Data Team Success, What You Do is Less Important Than How You Do It . . . . .	15
<i>"Chicken &amp; Rice Guys" Chicken and Rice</i> . . . . .	19
6 Steps to an Enterprise DataOps Transformation . . . . .	21
<i>Bungeoppang</i> . . . . .	31
The Business Case for DataOps . . . . .	33
<i>Red Lentil Curry / Dal</i> . . . . .	37
<b>FIND</b>	<b>39</b>
Launch Your DataOps Journey with the DataOps Maturity Model . . . . .	41
<i>Slovak Sunday Bone Broth Soup</i> . . . . .	51
Jump-Starting Your DataOps Journey . . . . .	53
<i>Peanut Butter Energy Bites</i> . . . . .	59
4 Easy Ways to Start DataOps Today . . . . .	61
<i>Chocolate Stout Cupcakes with     Irish Whiskey Filling and Baileys Frosting</i> . . . . .	67
<b>ESTABLISH</b>	<b>69</b>
Finding an Executive Sponsor for Your DataOps Initiative . . . . .	71
<i>Gil's Easy Chicken Cacciatore</i> . . . . .	75
Pitching a DataOps Project That Matters . . . . .	77
<i>Risotto alla Monzese</i> . . . . .	81
<b>DEMONSTRATE</b>	<b>83</b>
Prove Your Team's Awesomeness with DataOps Process Analytics . . . . .	85
<i>Grandma's Italian Meatballs</i> . . . . .	91
<b>ITERATE</b>	<b>93</b>
Eliminate Your Analytics Development Bottlenecks . . . . .	95
<i>Pav Bhaji</i> . . . . .	105

<b>EXPAND</b>	<b>107</b>
Do You Need a DataOps Dojo? . . . . .	109
<i>Spinach Madeline</i> . . . . .	113
DataOps Engineer Will Be the Sexiest Job in Analytics . . . . .	115
<i>Kerala Style Chicken Stew</i> . . . . .	119
Improving Teamwork in Data Analytics with DataOps . . . . .	121
<i>Spinach-Mushroom Quiche</i> . . . . .	133
Governance as Code . . . . .	135
<i>Slow Cooker Hangi Pork</i> . . . . .	145
<b>CONCLUSION</b>	<b>147</b>
Why Are there So Many -Ops Terms? . . . . .	149
<i>Mom's Keto Chocolate Peanut Butter Fat Bombs</i> . . . . .	157
A Guide to Understanding DataOps Solutions . . . . .	159
<i>Gil's Old Fashion Fudge</i> . . . . .	163
What a DataOps Platform Can Do For You . . . . .	165
<i>Chapssal Doughnuts</i> . . . . .	169
<b>DATAOPS RESOURCES</b>	<b>172</b>
<b>ABOUT THE AUTHORS</b>	<b>173</b>
<b>ADDITIONAL RECIPES</b>	<b>175</b>
<i>Chicken Breasts with Marsala Wine</i> . . . . .	177
<i>Dakgalbi</i> . . . . .	178
<i>Isaac's Special Chicken</i> . . . . .	180
<i>White Russian Tiramisu Cake</i> . . . . .	181

# Introduction

Business agility separates the leaders from the laggards. An agile business monitors the environment, quickly detects change, forms and executes plans and makes adjustments based on feedback. As data professionals, we see the foundational role that data and analytics play throughout this process. If analytics are bureaucratic and error-prone, people will naturally seek workarounds, resulting in diminished agility. Business agility depends upon analytics agility.

We are entering an era where analytics agility will be a key competitive differentiator for enterprises. Organizations bogged down by data errors and sluggish analytics team productivity will find themselves at a significant disadvantage. If companies want to be more agile, they must start with the data analytics team. Agile analytics can transform an enterprise from the inside.

When data and analytics are accurate, people learn to trust data. When a data team responds to requests immediately and on-demand, business stakeholders work more closely with the analytics team. When users and data professionals work closely together, it unlocks creativity spurred by insights that drive organizations toward new products and services, innovative marketing strategies and new markets.

DataOps is a data analytics methodology that serves as the vehicle for transformational change led by analytics. It emphasizes observability and meta-orchestration to produce error-free analytics that can be created and updated at lightning speed. DataOps is the secret sauce that can build market-leading analytics capabilities that will raise a company's business agility. We've written extensively about DataOps over the past years. If you are new to the topic, please see our first book, [“The DataOps Cookbook”](#) (over 12,000 downloads and counting) and the other resources listed in the Appendix section.

Many people ask us how to begin their DataOps journey. We used to answer that question by talking about the [“Seven Steps to Implement DataOps.”](#) Over time, we understood that some people were asking a broader question about using

DataOps to transform their enterprise. The problem statement was aiming at how a data professional can lead a DataOps initiative. How do you build support for DataOps? What is the best first project? How can you transfer DataOps from a single team to the greater enterprise? A data scientist excited about the potential benefits of DataOps may never have led an organizational change initiative.

Imagine if a person could time travel back to the 1980s and tried to evangelize Agile development. That person would face a lot of naysayers. “We’ve never done it that way.” “I don’t get how that benefits us.” “Your methods don’t align with how we allocate resources for projects.” From our perspective, we know that the Agile advocate is correct, but our intrepid time traveler would need a way to convince skeptics.

We hope this book will help you evangelize and lead a DataOps transformation at your organization. We’ve included all of the insight that we’ve gained from our own experiences coaching data analytics professionals on the best way to lead organizational change using DataOps. We hope that these materials will help you on your DataOps journey.

Our book’s title (“Recipes for DataOps Success”) refers to the orchestrated pipelines that drive DataOps. One of the DataKitchen Platform’s lesser-known features is how it helps data teams share development and operations “Recipes,” improving collaboration and promoting reuse throughout the organization. To have some fun with this metaphor, we asked our coworkers at DataKitchen (our data chefs) to share their favorite recipes with you. You’ll find these sprinkled throughout the book. Enjoy and bon appetit!



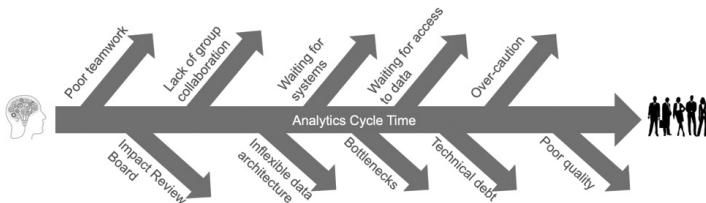
**Educate**



# Why Do DataOps

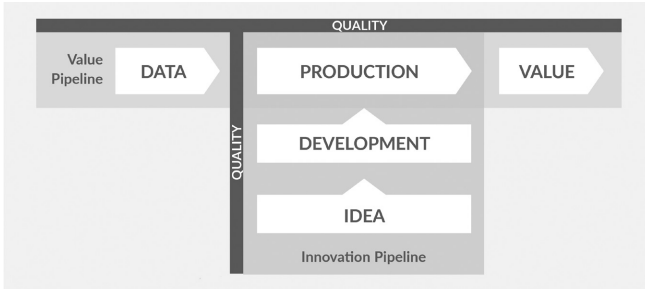
If you are frustrated with your enterprise's data analytics, you are not alone. [VentureBeat](#) reported that 87% of data science projects never make it into production. It's no surprise then that, despite soaring investments in AI and data science, the percentage of organizations that describe themselves as “data driven” has fallen from [37% to 31%](#) since 2017.

Too often, data science remains a manual process, conducted by highly trained artisans. The technology research firm [Gartner](#) asserts that 80% of AI projects resemble alchemy, run by wizards whose talents will not scale in the organization. Imagine an automobile manufacturing plant run without automation. It would suffer from inconsistent quality, long cycle times, waste, inflexibility and bottlenecks. No one in the 21st century would ever run an operations team that way. Yet, walk down the hall to your data analytics group and observe – poor quality, minor changes take months to implement, manual processes, 75% of the day is hijacked by unplanned work, and oversubscribed resources limit overall productivity. It's a classic case of good people doing their utmost to overcome the limitations inherent in poor business processes.



Data teams can learn a lot from the quality methods used in automotive and other industrial manufacturing. Methodologies like [Lean manufacturing](#) and the [Theory of Constraints](#) apply just as well to data operations and analytics development as traditional factories. Analytics is a pipeline process. Data sources enter the enterprise, are loaded into databases, undergo processing and transformation, and then feed into charts, graphs and predictive analytics. From a process perspective,

this workflow is a manufacturing operation. New analytics are much like manufacturing engineering, creating new and improved operational capabilities. As every factory manager knows, change management is a critical aspect of operations.



## FOLLOWING THE LEAD OF THE SOFTWARE INDUSTRY

If you haven’t encountered these ideas before, you may think that I am writing something revolutionary. Actually, the methods described here are widely implemented in the software industry. While a data team might require six months to release a 20-line SQL change, Amazon recently disclosed that their Amazon Web Services (AWS) team performs 50,000,000 code releases per year. If your data team had the same processes and methodologies in place as Amazon, you could ask a complex question about your customer segmentation or operations and receive an answer the same day. The number of “what-ifs?” that you could pose would increase by 50X. Imagine what that could do for creativity and business innovation in your enterprise.

## DATAOPS – APPLYING MANUFACTURING METHODS TO DATA SCIENCE

The data analytics industry today is much like the software industry of the 1990’s – producing releases at a slow pace and incurring technical debt. The good news is that the software industry discovered a path forward using classic manufacturing methodologies. Furthermore, these ideas are gaining traction in the data analytics world.

The data science industry refers to these methods under the umbrella term [DataOps](#). Just to be clear, DataOps is not a single vendor. It is not a particular tool. You do not have to throw away your existing infrastructure and start over. DataOps augments your existing operations. It is a new approach to data science which draws upon three widely-adopted methodologies that are supported by tools and software automation: [Agile](#) Software Development, [DevOps](#) and statistical process controls ([SPC](#)).

## AGILE DEVELOPMENT

One axiom in the Theory of Constraints is that small lot sizes reduce inventory, minimize waste and increase the overall system throughput of a manufacturing



operation. This insight inspired the software industry to create a methodology called Agile development. Studies show that Agile projects complete over 30% percent faster and with a 75% lower defect rate. Today, two-thirds of software organizations describe themselves as either “pure agile” or “leaning towards agile.”

Traditional project management utilizes a Waterfall sequential methodology. Projects are executed according to lengthy, complex schedules with a single deliverable at the end. There are several problems with this methodology in data science. In analytics, business conditions are constantly changing so whatever business colleagues needed several months ago has changed or is no longer of value. In other words, requirements have a shelf life. Additionally, waterfall projects are, by design, methodical (slow) and inflexible. Waterfall projects subject to rapid-fire requirements flowing in from business users never exit the planning (and replanning) phase.

In a nutshell, Agile project management delivers valuable features in short intervals and seeks immediate feedback. Large initiatives are broken into small increments and delivered iteratively. In Agile, the data science team responds faster and aligns more closely with the requirements and immediate priorities of end-users. The Agile methodology is particularly effective in environments where requirements are quickly evolving — a situation well known to data science professionals.

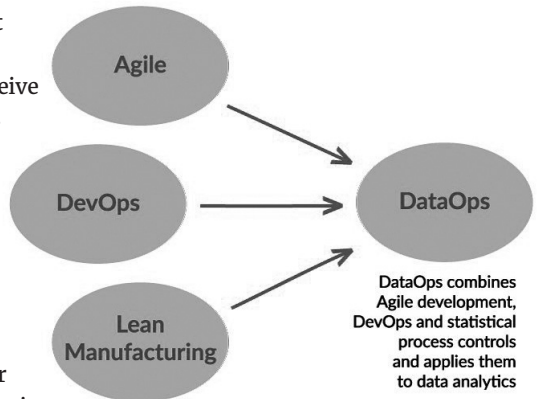
Some enterprises understand that they need to be more Agile, but organizations typically do not receive much benefit from Agile methods if quality is poor or deployment processes involve lengthy and laborious manual steps. “Agile development” alone may not make a team more “agile.”

## DEVOPS

Imagine clicking a button in order to fully test and publish new analytics into the production pipeline. That’s how Amazon and others deploy software releases in minutes or seconds. This approach to releasing software is called DevOps.

Traditionally, software organizations waited weeks or months for IT to install and configure development environments for new projects. DevOps automates this process by placing it under software control. At the push of a button (or command), DevOps spins up a virtual machine and configures it with software and data. A data scientist can be up and running on a new development project in minutes.

DevOps also automates testing. An extensive battery of tests verify and validate



that new analytics work and will operate error-free in an environment that exactly matches production. No more throwing new analytics over the wall and hoping that it doesn't break anything. When testing is complete, analytics are quickly published to users via an automated workflow. This method of publishing software is also called [continuous delivery](#) or continuous deployment, and it is a central tenet of DataOps.

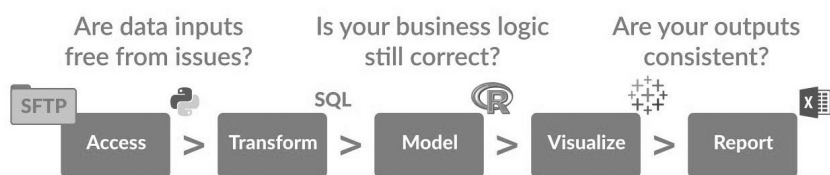
When environment creation, test and deployment are placed under software control, they can happen in seconds or minutes. This is how companies like Amazon attain such rapid cycle time.

Agile development and DevOps work hand in hand. Agile enables enterprises to quickly specify and commit to developing new features, while DevOps speeds execution, test and release of those features. Neither of these methods would be as effective without the other. Additionally, it's impossible to move quickly when a team is plagued by quality errors.

## STATISTICAL PROCESS CONTROLS

Modern enterprises have hundreds or thousands of data sources flowing into their data pipeline. The sheer quantity of data powering analytics exceeds the monitoring capacity of the typical data team. Left unchecked, errors eventually creep into data, and data errors can break or invalidate analytics. If you've ever received a report that was based upon incorrect data, you have experienced this first hand.

DataOps approaches data errors the same way that a manufacturing operation controls supplier quality, work-in-progress and finished goods. DataOps borrows a methodology, straight from lean manufacturing, called statistical process control (SPC). Tests monitor data flowing through the pipeline and verify it to be valid, complete and within statistical limits. Every stage of the data pipeline monitors inputs, outputs and business logic. Input tests can catch process drift at a data supplier or upstream processing stage. Output tests can catch incorrectly processed data before it is passed downstream. Tests ensure the integrity of the final output by verifying that work-in-progress (the results of intermediate steps in the data pipeline) matches expectations.



If an anomaly occurs at any point in the workflow or pipeline, the data team will be the first to know, through an automated alert, and they can take action. Test results can also be displayed in dashboards, making the state of the data pipeline

transparent from end to end.

## **TURNING DATA INTO VALUE**

As enterprises develop and deploy data analytics with DataOps, they can attain the same level of productivity that we see in leading software companies. Analytics will be created and deployed rapidly and statistical process controls will ensure that quality remains high. The data science team will respond to requests for new analytics with unprecedented speed and accuracy.

DataOps offers a new approach to creating and operationalizing analytics that minimizes unplanned work, reduces cycle time and improves code and data quality. It is a methodology that enables data science teams to thrive despite increasing levels of complexity required to deploy and maintain analytics in the field. Without the burden of inefficiencies and poor quality, data science teams can focus on their area of expertise; creating new models and analytics that fuel business innovation and create competitive advantage.



# Better than Shake 'n Bake!

*Contributed By Joanne Ferrari*

**Prep Time:** 15 minutes — **Cook Time:** 45 minutes — **Total Time:** 1 hour

Better than Shake 'n Bake! Easy, inexpensive & tastes better too! Common ingredients come together in this copycat Shake 'n Bake recipe that's even better than the original.

## INGREDIENTS

3 cups dried bread crumbs, ground very fine

3 Tbsp cornmeal

3 Tbsp corn starch

1 Tbsp granulated onion, (powder can work too)

1 1/2 tsp granulated garlic

1 tsp finely ground black pepper

1 1/2 tbsp fine salt

1 1/2 tsp chili powder

1 tsp ground dry thyme

1 1/2 tsp ground dry oregano

## INSTRUCTIONS

1. Mix together all of the ingredients very well. I use a food processor or mixer to make sure everything is very well blended.
1. Store in a cool place in an airtight container like a 1 quart/liter mason Jar.

## TO PREPARE THE CHICKEN

I use one whole chicken, cut in pieces and well-trimmed but about 3 lbs of any kind of chicken pieces you like will do. You can, of course, make any amount you need.

1. Preheat the oven to 375 degrees.
2. Simply wet chicken pieces with water, drain well and drop them, one at a time into a plastic bag containing some of the homemade shake 'n bake. I usually start with a half cup of the coating in the bag, which is equivalent to what is in an envelope if you bought it at the supermarket. You can always add a little extra if you need it at the end, but I find this is the best way to maximize the use you get out of a batch.
3. Shake the bag and press the coating onto the individual chicken pieces. Place the coated pieces on a parchment paper-lined baking sheet. Don't crowd the pieces, they will crisp much better if there is space between them.

## BAKING THE CHICKEN

At this point, you can drizzle a little canola oil or peanut oil over the coated pieces to maximize browning but this step is completely optional. I have an oil spritzer which is ideal for this purpose because you can spritz about 9 or 10 pieces of chicken with only about a tablespoon of oil. This is a method I often use for Oven Fried Chicken too.

Bake for about 45-55 minutes depending upon the size of the chicken pieces being used. Boneless skinless chicken breasts can be ready in as little as 25 minutes depending on size. I use my meat thermometer to ensure that the internal temperature is 175-180 degrees F to ensure they are fully cooked.

Let the chicken pieces rest for 5 to 10 minutes before serving.



# For Data Team Success, What You Do is Less Important Than How You Do It



In today's on-demand economy, the ability to derive business value from data is the secret sauce that will separate the winners from the losers. Data-driven decision-making is now more critical than ever. Analytics could mean the difference between finding the right mix of strategic moves or falling behind. In fact, [Forrester Research](#) predicted that insight-driven companies would grow seven to 10 times faster than the global GDP through 2021.

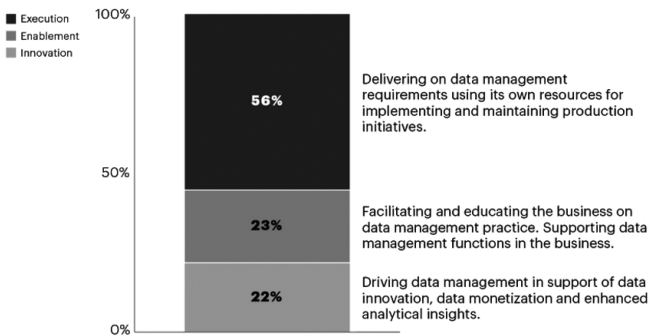
Most enterprise companies recognize the need to be data-driven, yet 60% of data projects fail to move past preliminary stages, and 87% of science projects never make it to production. More surprisingly the number of data-driven companies has actually fallen from [37% to 31%](#) since 2017, despite increased investment

## WHAT GIVES?

Becoming data-driven is hard and data teams are suffering. They are caught between the competing demands of data consumers, data providers, and supporting teams. Typically, data consumers live in an Amazon world and expect trusted, original insight on-demand. Yet data providers often send inaccurate, late, or error-prone data sets. The flawless collaboration and production required from teams in other parts of the organization often just isn't there.

Taken together, the need to manage complex toolchains and data, as well as collaborate with other organizations, roles, locations, and data centers, saps the data team's time. In fact, most data teams spend more time fixing errors and addressing operational issues than innovating and providing business value. According to Gartner, only 22% of a data team's time is spent on new initiatives and innovation. As a result, many data teams are not meeting expectations, or worse, are beaten down and disempowered.

**Proportion of Time Spent on Digital Management Tasks**  
Average Proportion



### March 2020 Gartner Survey: “Data Management Struggles to Balance Innovation and Control”

Figure 1: According to Gartner, only 22% of a data team's time is spent on new initiatives and innovation.

## FOCUS ON OPERATIONS, NOT THE NEXT FEATURE

Data teams can learn important lessons from other industries. According to management guru Dr. W. Edwards Deming, 94% of problems are “common cause variation,” and to decrease this variation you must focus on the system or process, not look for a person to blame. A relentless process focus has led to dramatic improvements in the auto industry, where lean manufacturing principles have led to dramatically higher levels of productivity and quality. Or more recently in software development, the principles of DevOps have enabled companies to perform millions of software releases each year.



---

**“We realized that the true problem, the true difficulty, and where the greatest potential is — is building the machine that makes the machine . In other words, it’s building the factory. I’m really thinking of the factory like a product.” Elon Musk**

---

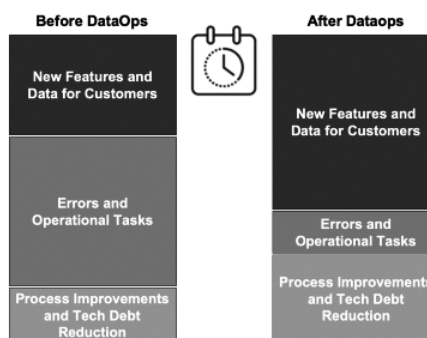
This mind shift was more recently highlighted by Elon Musk who said “we realized the true problem, the true difficulty, and where the greatest potential is — is building the machine that builds the machine. In other words, it’s building the factory. I’m really thinking of the factory like a product.” Successful data organizations are also wise to think of their data pipelines like a factory where quality and efficiency must be managed. But how can a data team shift its focus from the next big tool, technology or data feature to the people and process?

## A SOLUTION TO THE SUFFERING

In data analytics, DataOps provides the path forward. DataOps aligns the people, processes, and technologies of the data analytics organization. Supported by automation, it puts the focus on the underlying systems and managing the ‘data factory.’ Companies that follow DataOps principles spend less time worrying about the next model, algorithm, tool, visualization or even the data itself, but instead focus on how to develop, deploy, test, monitor, collaborate and measure their analytic operations.

By doing so, these companies realize multiple, simultaneous benefits.

- They experience orders of magnitude improvements in cycle time. They are able to deploy new features quickly and confidently, often improving from months/weeks to days/hours.
- They lower or even eliminate costly and embarrassing errors, enabling them to build a strong culture of trust with their data customers.
- They dramatically increase productivity. Better intra- and inter-team collaboration means less time spent on meetings and bureaucracy.



**Figure 2: DataOps reduces time spent on errors and operational tasks and increases innovation.**

All of this creates the time and space for the data team to focus on what they signed up for in the first place — creating innovative analytics and delivering business value.

Data organizations that neglect to modernize their processes, risk being left behind in an increasingly on-demand economy. DataOps enables teams to reclaim control of their data pipelines, reduce time- and soul-sucking errors and minimize the time from new ideas to the deployment of working analytics.

In any economy, but especially in challenging times, the most innovative companies will be those that can quickly adapt to rapidly evolving market conditions. The data teams that adopt DataOps and produce robust and accurate analytics more rapidly than their peers will power strategic decision-making that sustains a competitive advantage.

# “Chicken & Rice Guys” Chicken and Rice

*Contributed By Andrew Sadoway*

C&RG was a spot that many who worked at the DataKitchen Cambridge office used to frequent.

## INGREDIENTS

### marinade and chicken

- juice of 1/2 lemon
- 2 garlic cloves, minced
- 1 tsp kosher salt
- 1 tsp paprika (sweet, hot, or smoked)
- 3/4 tsp ground coriander
- 1 1/2 tsp ground cumin
- Pinch of ground cloves
- 1 tsp dried oregano
- 2 lbs boneless, skinless chicken thighs (about 6)
- 1 Tbsp olive oil

### rice

- 1 Tbsp olive oil
- 1/2 tsp ground turmeric
- 3/4 tsp ground cumin
- 2 cups basmati or another long-grain white rice
- 3 1/2 cups chicken stock
- 1 tsp kosher salt

### side salad

- 1/2 head iceberg lettuce, chopped
- 2 medium tomatoes, chopped
- 1/2 small white or red onion, chopped (optional)
- 1 small cucumber, chopped (optional)

### yogurt sauce

- 1 cup plain yogurt
- 2 Tbsp mayonnaise
- 1 Tbsp distilled white or apple cider vinegar
- 1/2 tsp granulated sugar
- 1/2 tsp kosher salt

### hot sauce

- 1 Tbsp harissa
- 1 Tbsp sriracha (or other hot sauce)
- extras (optional)
- fresh cilantro, chopped, for garnish
- 3 large pocketless pita breads, toasted, halved

## INSTRUCTIONS

1. Marinate the chicken: combine marinade and pour into a bowl. Add the chicken, coat evenly and leave in the fridge, covered, for 30 minutes
2. Cook the chicken: Once the chicken is done marinating, heat a large deep skillet with a lid on medium-high heat in a single layer. Brown them, 5–8 minutes per side. Check the internal temp with a meat thermometer: it should reach at least 165 F.
3. Boil the rice: Add 1 Tbsp olive oil to the pot, heat it, add spices and rice and toast together for 1 minute, stirring frequently. Add the stock and salt and cook.
4. Make the sauces: while the rice cooks, make the yogurt and hot sauces
5. Make the salad: chop lettuce, tomatoes, and other (optional) veggies. Combine and season with salt and pepper.
6. Chop the chicken: when the chicken is done cooking, remove it and chop into bite-sized pieces on a cutting board. Return the pieces to the pan, coating them with the oil and spices in the pan.

*Attribution: Chicken & Rice Guys*



# 6 Steps to an Enterprise DataOps Transformation



DataOps reenvision how data analytics are conceived, created, deployed, supported, maintained and monitored. It removes the barriers that previously isolated users, data scientists and data operations from each other. DataOps represents nothing less than a transformational change that permeates the data and analytics teams.

As a career data professional, you may find it fairly straightforward to wrap your mind around the tools that implement DataOps. However, leading a DataOps initiative is about more than technologies and workflows. DataOps champions are leading cultural change, which also involves overcoming skepticism.

A DataOps champion may encounter resistance to change in the form of conflicting incentives, entrenched culture and a lack of buy-in. An organization may be hierarchical and silo'ed, but data cuts across teams, locations, and data centers. Major changes in data analytics methodologies and workflows are bound to infringe upon existing norms.

We have watched organizations implement DataOps using a variety of approaches. The most successful transitions to DataOps address both technical and human factors. Successful DataOps programs follow a gradual and methodical approach that establishes a beachhead with a first project, recruits allies and builds value iteratively. We summarize our recommended process for DataOps Enterprise Transformation in the six steps below (Figure 1).

# A Six-Step Process To Bring DataOps To Your Organization

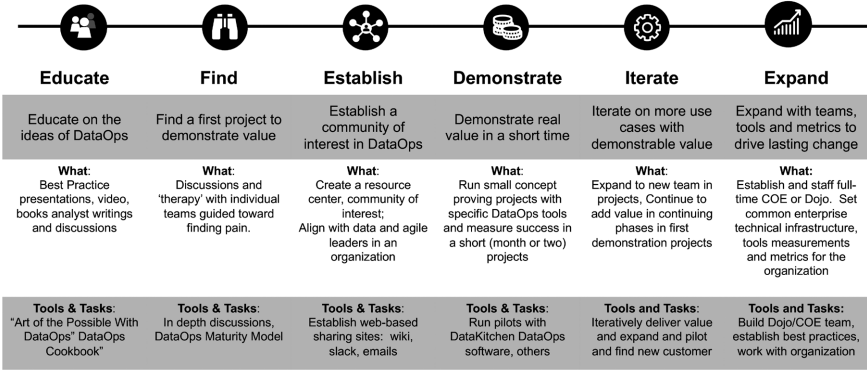


Figure 1: DataOps Enterprise Transformation can be accomplished in six steps.

## EDUCATE

DataOps introduces new methodologies, supported by tools automation, that shortens data analytics cycle time, improves collaboration, virtually eliminates errors and provides unprecedented transparency into data operations. DataOps can support your current toolchain or ease migration to new tools and technologies.

The best way to begin a transition to DataOps is by educating yourself and your team about how DataOps improves agility and quality. The team needs to learn: what is possible, what other enterprises have achieved, and what DataOps experts cite as best practice. Fortunately, there are many resources to assist you:

- [What is DataOps? Most Commonly Asked Questions](#)
- [The Seven Steps of DataOps](#)
- [The DataOps Cookbook](#)
- [DataOps YouTube Channel](#)
- [DataKitchen DataOps Webinars](#)
- [The DataOps Manifesto](#)

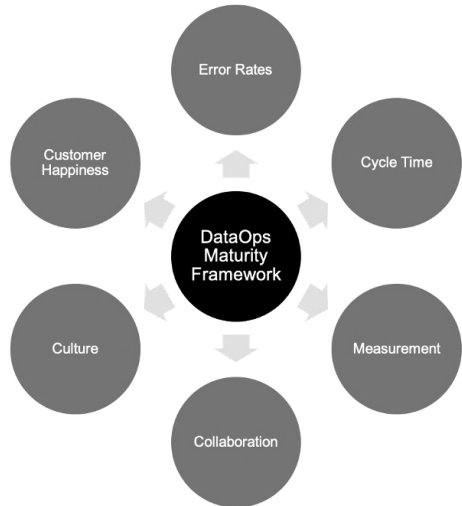


Figure 2: The six dimensions of DataOps maturity

Your investment in DataOps education should stimulate your vision of improving your organization's workflows by applying DataOps principles. Nothing is more effective at proving DataOps' potential impact than a mini-project.

## FIND

A mini or pilot project can serve as a proof of concept for potential DataOps benefits. Choose your first project in consultation with your team and, if possible, an executive sponsor. Ideally, it should demonstrate meaningful improvement in a key performance parameter. Ideally, a first project leads to a quick win. A shorter schedule is eminently preferable to an extended development effort. That's not to say that you have to get it perfect in one shot. Iterated improvements demonstrate how value builds by using Agile development. If you can't decide where to begin, our DataOps Maturity Model may be helpful.

### Measure DataOps Maturity

The [DataOps Maturity Model](#) can help organizations understand their DataOps strengths and weaknesses. Maturity models are commonly used to measure an organization's ability to improve in a particular discipline continuously. DataKitchen's DataOps Maturity Model outlines a measurement approach for building, monitoring, and deploying data and analytics according to DataOps principles. With this model, teams can understand where they are today and how to move up the curve of DataOps excellence.

DataOps employs automated orchestration to simplify complex toolchains, environments, and team collaboration, so that the data team can quickly and continuously deliver high quality, error-free insight. To implement DataOps, organizations need to prioritize improvements in the six areas shown in Figure 2.

Improve these areas by implementing core DataOps capabilities such as automated testing and monitoring, toolchain orchestration, version control, sandbox creation and management, and continuous deployment. Many DataOps capabilities relieve bottlenecks in workflow processes.

### Eliminating Bottlenecks

Most data teams are interested in DataOps because they seek to accelerate the creation and deployment of new data analytics (data, models, transformation, visualizations) without introducing errors. Reducing project cycle time or eliminating errors are both excellent starting points. Errors are a major source of unplanned work, which is a bottleneck that limits the throughput of the overall system. To minimize errors, start tracking errors and form a quality circle to explore root causes. Add tests to your data operations pipelines and continuous deployment pipelines so that your data team can address errors before they affect users.

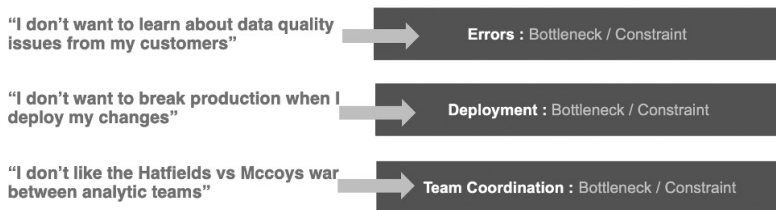


Figure 3: DataOps relieves productivity constraints such as data errors, deployment errors, and lack of team coordination.

To reduce project cycle time, study and measure the workflow processes from the inception of an analytics requirement to the delivery of published analytics. Every workflow process includes constraints and bottlenecks. Improve overall cycle time by mitigating these constraints in your development processes (Figure 3).

One approach that we recommend involves using the [Theory of Constraints](#) to alleviate your workflow bottlenecks. A bottleneck is a step in your end-to-end lifecycle process that acts as a constraint on the overall throughput of the entire system. For example, waiting six weeks for approval from an impact review board severely constrains agility. A data organization's bottlenecks often leave the following telltale signs:

- **Work in Progress (WIP)** – In a manufacturing flow, work in progress usually accumulates on an input queue feeding into a constraint. In data analytics, you may notice a growing list of requests for a scarce resource. For example, if it takes 20 weeks to provision a development system, your list of requests for them is likely to be long.
- **Expedite** – Look for areas where you are regularly diverting resources to ensure that critical analytics reach users. In data analytics, data errors are a common source of unplanned work.
- **Cycle Time** – Pay attention to the steps in your process with the longest cycle time. Naturally, if a process step is starved or blocked by a dependency, the bottleneck is the external factor. If it takes months to receive data sets from the central IT department, work with them to set-up a regular, automated feed into a locally controlled data lake.
- **Demand** – Note steps in your pipeline or process that are simply not keeping up with demand. For example, often, less time is required to create new analytics than to test and validate them in preparation for deployment. This disparity can be addressed using DevOps techniques.

Whatever your choice of projects, invest in activities that will garner support and demonstrate how DataOps produces measurable results.

The [DataKitchen DataOps Platform](#) and other [DataOps tools](#) can play a critical role in shortening the cycle time of your DataOps model project. A DataOps Platform is purpose-built to augment an existing toolchain with DataOps automation. It can help you hit the ground running.



## ESTABLISH

Many DataOps transformations start with a small number of contributors who serve as the core team. As excitement grows, you will find that your more established team will need more structure to keep everyone rowing in the same direction. Here are some ways you can support and encourage your team's growth:

- **Community-of-interest** – Find allies and cultivate a community of interest (COI) around DataOps methods and automation. One ready-made resource is the group of engineers and data scientists who understand Agile development and DevOps. These folks will understand the power of process improvement to boost productivity and quality in software development and data analytics. DataOps communities often come alive with palpable energy as DataOps benefits win over converts.
- **Executive sponsor** – A C-level sponsor can tie the project's activities into the larger organization's strategic goals. An executive can explain the value to others and provide guidance as the project team faces obstacles or grapples with trade-offs. The executive sponsor provides resources and budget as a skunkworks matures into an official project. With support from data science or engineering managers, you can gain approval for your COI to devote part or all of their time to DataOps officially.
- **DataOps strategy** – A DataOps strategy keeps everyone on the same page. As your team grows beyond its core members, a written strategy empowers everyone to contribute their creativity. If your DataOps initiative has specific initial goals, a strategy clearly communicates them to the team.
- **Shared workspace** – A shared workspace and communication channels help the team interact around tasks and builds a shared identity. Some teams have a physical space, but others are entirely virtual. Your DataOps COI may benefit from collaboration tools such as a Wiki, Slack channel or an email list. With a budget, you can establish a resource center to support your DataOps projects.
- **Build value** – DataOps is an iterative process that builds value using automation. Everyone who works on your DataOps initiative should be helping to create or enhance the value-creation machine. If that machine runs 24x7, it creates value long after the data scientists and programmers have deployed their solutions. For example, tests that ensure data quality keep creating value as new data flows through the data analytics pipelines.

The value that DataOps builds should manifest in tangible improvements. Demonstrate the benefits of DataOps, and win converts, using metrics.

## DEMONSTRATE

DataOps will deliver an unprecedented level of transparency into your operations and analytics development. DataOps automated orchestration provides an

opportunity to collect and display metrics on all of the activities related to analytics (Figure 4). Why not use DataOps analytics to shine a light on the benefits of DataOps itself?

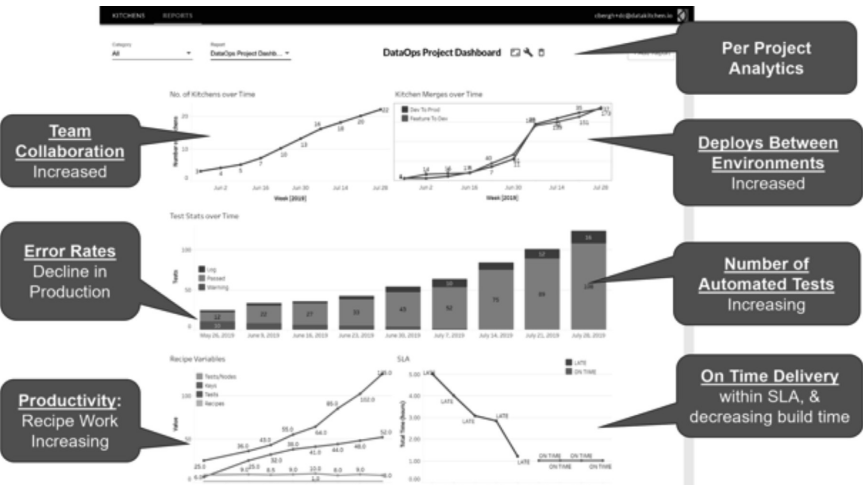


Figure 4: A typical DataOps dashboard

Figure 4 shows a typical DataOps dashboard with metrics related to team collaboration, error rates, productivity, deployments, tests, and delivery time. The metrics above might benefit from a short explanation:

- **Team Collaboration** – Measure teamwork by the creation of virtual workspaces, also known as “Kitchens.” Each Kitchen creation corresponds to a new project or sub-project in a team context.
- **Error Rates** – The graph shows production warnings at a rate of 10 per week, falling to virtually zero. This reduction in errors is the positive result of the 100+ tests that are now operating 24x7 checking data, ETL, processing results, and business logic. As the number of tests increases, the data pipeline falls under increasingly robust quality controls.
- **Productivity** – Measure team productivity by the number of tests and analytics created. The rise in “keys” (steps in data pipelines) coupled with the increase in test coverage shows a thriving development team. Also, the number of Kitchen merges at the top right shows the completion of projects or sub-projects. The “Feature to Dev” metric shows new analytics ready for release. “Dev to Prod” merges represent deployments to production (data operations).
- **On-time Delivery** – Mean deployment cycle time falls sharply, meeting the target service level agreement (SLA).

Choose your metrics to reflect your DataOps project objectives. The metric gives the entire team a goal to rally around. The number of possible DataOps metrics is as varied as the architectures that enterprises use to produce analytics. When your team focuses on a metric and iterates on it, you’ll see significant improvements in each sprint.

## **ITERATE**

In Agile Software Development, the team and its processes and tools are organized around publishing releases to the users every few weeks (or at most every few months). A development cycle is called an iteration (or a sprint). At the beginning of an iteration, the team commits to completing working and valuable changes to the code base. With iterations occurring at short intervals, the organization can continuously reassess its priorities and incorporate them into future iterations. This method allows the development team to adapt to changing requirements more easily. Each iteration adds value, so the final product is continually improved.

In an increasingly competitive marketplace, Agile methods allow companies to become more responsive to customer requirements and accelerate time to market. Agile also improves ROI by monetizing features with each iteration instead of waiting months for a big release. Unlike classic software development, agile projects build value with each iteration. In DataOps, iterations build upon each other, so value grows over time.

While initial iterations may have focused on one project, demonstration of success encourages a DataOps team to broaden its scope. Iterations can address new goals: tackling additional bottlenecks, adding new data sets, and working with new teams.

## **EXPAND**

As your DataOps initiative grows beyond the early stages, you will expand to incorporate more staff, resources, and a broader scope. One best practice incorporates DataOps into the organization chart. A sign of DataOps maturity is building a common technical infrastructure and tools for DataOps using centralized teams. It's also important to establish enterprise-wide measurements and metrics. Work with other teams throughout the organization to bring DataOps benefits to every corner of the enterprise.

## DataOps Technical Services

One approach standardizes a set of software services that support the rollout of Agile/DataOps. The DataOps Technical Services (DTS) group provides a set of central services leveraged by other groups. Examples of technologies that can be delivered ‘as a service’ include:

- Source code control repository
- Agile ticketing/Kanban tools
- Deploy to production
- Product monitoring
- Develop/execute regression testing
- Development sandboxes
- Collaboration and training portals/wikis
- Test data management and other functions provided ‘as a service’

The DTS group can also act as a services organization, offering services to other teams. Below are some examples of services that a DTS group can provide:

- Reusable deployment services that integrate, deliver and deploy end-to-end analytic pipelines to production.
- Central code repository where all data engineering/science/analytic work can be tracked, reviewed and shared.
- Central DataOps process measurement function with reports
- ‘Mission Control’ for data-production metrics and data-team development metrics to demonstrate progress on the DataOps transformation

Another important tool employed by maturing DataOps organizations helps train practitioners in DataOps methods and best practices so they can return to their team and lead local DataOps efforts.

## DataOps COE

The Center of Excellence (COE) model leverages the DataOps team to solve real-world challenges. The goal of a COE is to take a large, widespread, deep-rooted organizational problem and solve it in a smaller scope, proof-of-concept project, using an open-minded approach. The COE then attempts to leverage small wins across the larger organization at scale. A COE typically has a full-time staff that focuses on delivering value for customers in an experimentation-driven, iterative, result-oriented, customer-focused way. COE teams try to show what “good” looks like by establishing common technical standards and best practice. They also can provide education and training enterprise-wide. The COE approach is used in many enterprises, but the DevOps industry has more often standardized on Dojos as a best practice.

## DataOps Dojo

A DataOps Dojo is a place where DataOps beginners go for a short period of intense, hands-on training. In Japan, a dojo is a safe environment where someone can practice new skills, such as martial arts. Companies like [Target](#) employ the Dojo concept effectively to build lean, Agile and DevOps muscles. The Dojo offers a separate workspace where teams learn new skills while working on actual projects that deliver customer value.

Dojos provide an environment where teams gain practical experience without worrying about introducing errors into the production environment. The staff rotates in for weeks or months at a time to learn new skills by working on real-world projects. They then bring those skills ideas back to their original teams.

## CHAMPIONING DATAOPS

DataOps can serve as a positive agent of change in an otherwise slow and process-heavy organization. Remember that leading change in technical organizations is equal parts people, technology and processes. DataOps offers the potential to reinvigorate data team productivity and agility while improving quality and predictability. Our six-step program should help you introduce and establish DataOps in your data organization. In our experience, many data organizations desperately need the benefits that DataOps offers. They need people to champion a DataOps initiative. Can your organization count on you?



# Bungeoppang

*Contributed By Brandon Stephens*

Super-popular Korean street snack

## INGREDIENTS (for 6 bungeoppang)

1 cup all-purpose flour

½ teaspoon kosher salt

½ teaspoon baking soda

1 tablespoon brown or white sugar

1 cup plus 2 tablespoons water

1 tablespoon vegetable oil

- Sweet red beans (canned or homemade): for homemade, use the method from my patbingsu recipe
- Bungeoppang special pan

## INSTRUCTIONS

- Combine flour, kosher salt, baking soda, and sugar in a bowl. Add water and mix it well.
- Sieve the mixture through a strainer to get a silky batter without any lumps.
- Heat up the bungeoppang pan and turn the heat down to low.
- Open the pan and grease both the upper and lower fish molds with a light coating of vegetable oil.
- Pour the batter into one side of the fish mold until it's ½ full. Add 1 tablespoon of sweet red beans to the center, and then gently fill up the rest of the fish mold to totally cover the red beans.
- Close the mold and cook for about 3 minutes over low heat.
- Turn the pan over and let it cook another 3 minutes. Open it and turn it over again for another 30 seconds, to make the bread a little more crispy.
- Take out and serve immediately.

Attribution: Maangchi





# The Business Case for DataOps

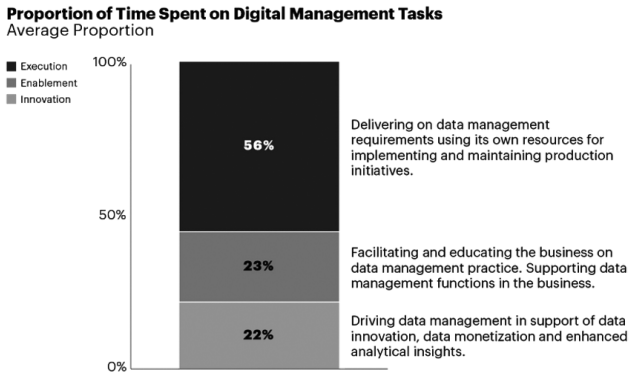
BY JAMES ROYSTER



Savvy executives maximize the value of every budgeted dollar. Decisions to invest in new tools and methods must be backed up with a strong business case. As data professionals, we know the value and impact of DataOps: streamlining analytics workflows, reducing errors, and improving data operations transparency. Being able to quantify the value and impact helps leadership understand the return on past investments and supports alignment with future enterprise DataOps transformation initiatives. Below we discuss three approaches to articulating the return on investment of DataOps.

## RESOURCE REDEPLOYMENT

In a recent Gartner survey, data professionals spent 56% of their time on operational execution and only 22% of their time on innovation that delivers value. An effective DataOps strategy can help a team invert this ratio and provide more value to the company.



### March 2020 Gartner Survey: “Data Management Struggles to Balance Innovation and Control”

Figure 1: Data professionals spend only 22% of their time on innovation.

Gartner describes the time spent on “operational execution” execution as using the data team to implement and maintain production initiatives. A big percentage of the time that data scientists spend on operational effort is consumed servicing data errors.

In teams with mature DataOps practices, including some long-time DataKitchen customers, data professionals have indeed flipped the ratio and spend much less time on nonvalue-added activities. Instead these organizations commit 20% of their time implementing automation and writing tests. As a result, they reduced the time spent on errors and manual processes to nearly zero. This allows the team to spend significantly more time focusing on high-value efforts and meaningful collaborations. Good rules of thumb are:

- If you’ll perform an operation twice in a year, then automate it.
- If it can be wrong, test it.

Implementing DataOps automation requires about 20% of a data professional’s time, but it completely eliminates data team participation in operations, saving them 56% of their time; a net savings of 36%. For a team of ten data professionals, this savings is the equivalent of adding more than 3.5 full-time employees to value added activities. These newly available resources can be redeployed to create more capacity for the company’s analytics-hungry product teams.

Another way to demonstrate the impact of DataOps on FTEs is showing the math.

	<u>DataOps Cost/Benefit Example</u>
<u>\$130,000.00</u>	<u>FTE average salary</u>
<u>\$156,000.00</u>	<u>FTE fully burdened cost</u>
<u>10</u>	<u>Team size</u>
<u>\$1,560,000</u>	<u>Team total cost</u>
<u>56%</u>	<u>Operational Execution</u>
<u>20%</u>	<u>DataOps time spend on automation and testing</u>
<u>36%</u>	<u>DataOps net time savings</u>
<u>\$561,600</u>	<u>Value of data team resources redeployed</u>

Thirty six percent of the total time of a ten-person team, based on a full-time employee (FTE) cost of \$156,000 amounts to \$561,000. This significant sum can be redeployed to higher value-add activities.

### INSOURCING THROUGH DATAOPS

Many companies overcome their staffing limitations by outsourcing critical work to third parties. When internal analytics workflows are automated, there is little advantage to outsourcing. With DataOps, the work can often be performed much less expensively through automated orchestrations that are developed and managed in house. Automation can free up both direct and indirect resources. It enables companies to redirect the utilization of their own staff and reduce the dependency on external resources. If your company spends millions on consulting fees and outside contractors, DataOps automation could make a significant contribution to the bottom line. In one real-world example, a DataKitchen customer realized a net savings of \$70 million dollars as effort transitioned fully from outside agencies to internal resources.

### COST OF SLOW DECISION-MAKING

What can you do with the resources that are freed up from DataOps automation? One approach applies these resources to business analytics that expedite and improve decision-making.

Analytics agility leads to business agility. When the data team delivers analytics rapidly and accurately, analytics do a better job supporting decision-makers. When an organization can make decisions faster and better, it is able to capture opportunities that it would have otherwise missed or misjudged. With analytics playing a central role in corporate strategy, analytics agility can be a competitive advantage.

In one example, using analytics to understand customers and markets significantly improved product launch success at one DataOps enterprise. With rapidly produced analytics, they were able to improve market segmentation to maximize revenue in the early product lifecycle, boosting lifetime product revenue.

When executives evaluate whether to invest in a DataOps initiative, they need to understand the business benefits. Improved productivity, reduced outsourcing costs, and greater business agility together build a strong business case for DataOps. It may help to start with a mini or pilot project that demonstrates DataOps benefits. Improvement of a key metric may provide the justification that you need to secure investment in a larger DataOps program.

# Red Lentil Curry / Dal

*Contributed By Rajeew Singh*

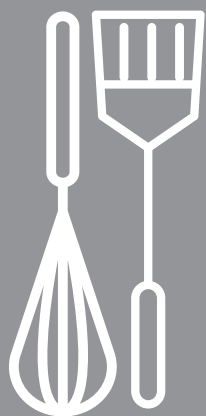
## INGREDIENTS

1.5 cup lentils  
1 tomato  
1 tsp turmeric powder  
1 tbsp ghee / clarified butter  
1 tsp cumin seeds  
1/2 cup chopped onion  
1 tsp ginger  
1 tsp garlic finely chopped  
1 tsp chili powder  
cilantro leaves  
salt

## INSTRUCTIONS

- Wash 1.5 cup lentils and soak them in water for about 2 hours.
- Add the soaked lentils, with 1 chopped tomato, 1 tsp turmeric powder and salt into an instant pot.
- Set the instant pot to Pressure Cook mode with a timer of 7 minutes and let it naturally release the steam for 5 minutes before opening the lid.
- Note, Instant pot is quicker, but the lentils can be easily cooked in a large saucepan. You can combine the lentils with about 3.5 cups of water and cook the lentils for about 25-35 minutes.
- Once the lentils are cooked, heat a pan and add 1 tbsp ghee / clarified butter. Once the ghee is hot, add 1 tsp cumin seeds, 1/2 cup chopped onion, 1 tsp ginger finely chopped, 1 tsp garlic finely chopped, 1 tsp chili powder and cook them until the onion turns translucent.
- Add the contents of this pan to the instant pot.
- Garnish with some finely chopped cilantro leaves.





**Find**





# Launch Your DataOps Journey with the DataOps Maturity Model



Most enterprise companies recognize the need to be data-driven, yet 60% of data projects fail to move past preliminary stages, and 87% of data science projects never make it to production. More surprisingly the number of data-driven companies has actually fallen from 37% to 31% since 2017, despite increased investment.

## WHY?

Becoming data-driven is hard. Data teams are caught between the competing demands of data consumers, data providers, and supporting teams. Typically, data consumers live in an “Amazon world” and expect trusted, original insight on-demand. Yet data providers often send inaccurate, late, or error-prone data sets. The flawless collaboration demanded of stakeholders often just isn’t there.

Taken together, the need to manage complex toolchains and data, as well as collaborate with other organizations, roles, locations, and data centers, saps the data team’s time. In fact, most data teams spend more time fixing errors and addressing operational issues than innovating and providing business value. According to Gartner, only 22% of a data team’s time is spent on new initiatives and innovation (Figure 1). As a result, many data teams are not meeting expectations, or worse, are beaten down and disempowered.

In data analytics, DataOps provides the path forward. Research shows that “organizations that adopt a DevOps- and DataOps-based approach are more successful in implementing end-to-end, reliable, robust, scalable and repeatable solutions,” says Gartner’s Sumit Pal. (Gartner, November 2018)

## March 2020 Gartner Survey:

“Data Management Struggles to Balance Innovation and Control”

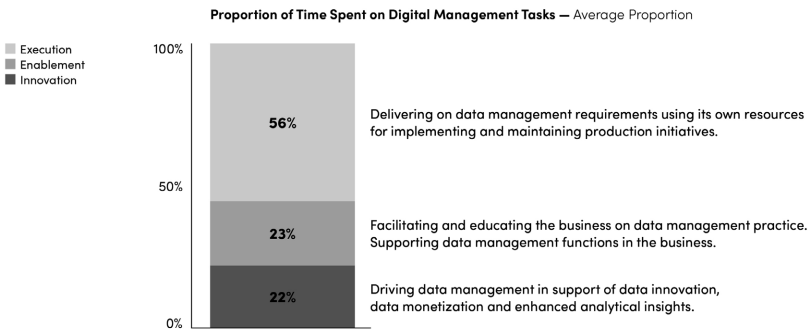


Figure 1: Only 22% of a data team's time is spent on new initiatives and innovation.

## WHAT IS DATAOPS?

DataOps is a set of technical practices, cultural norms, and architectures that enables:

- Rapid experimentation and innovation for the fastest delivery of new insights to customers
- Low error rates
- Collaboration across complex sets of people, technology, and environments
- Clear measurement and monitoring of results

DataOps draws on the principles of Agile, DevOps, and lean manufacturing to transform data processes. Supported by automation, it puts the focus on the underlying systems and managing the ‘data factory.’ Companies that implement DataOps realize multiple, simultaneous benefits. They:

- Experience orders of magnitude improvements in cycle time. They are able to deploy new analytics quickly and confidently, often delivering in hours/days instead of weeks/months.
- Lower or even eliminate costly and embarrassing errors, enabling organizations to build a strong culture of trust with their data customers.
- Dramatically increase productivity. Better intra- and inter-team collaboration means less time spent on meetings and bureaucracy and more on innovation.

## A MATURITY MODEL CAN HELP YOU GET STARTED

Because DataOps impacts your end-to-end analytic lifecycle, implementing DataOps can feel overwhelming. Even though a majority of respondents in a 2020 Seagate/ IDC survey said that DataOps was “very” or “extremely” important, only 10% have implemented DataOps fully across the enterprise. Success requires a mindset shift and most companies struggle with where to begin and how to even make modest progress towards their goals. A DataOps Maturity Model can be an incredibly useful tool to help organizations understand where they are today and how to get where they need to go.

## WHAT IS IMPORTANT IN DATAOPS?

To begin a DataOps initiative, it is first important to understand what is important (and what isn't) for DataOps success. DataOps requires a focus on the state of your data operations and processes, not the next new feature or tool. Typically, data teams can spend far too much time worrying about data types (e.g., batch, streaming, big, small, structured, unstructured), database types (e.g., hadoop, Spark, graph, NoSQL, object stores), data tools (e.g., ETL, BI, data science, data prep, catalog, etc), or specific design paradigms (e.g., lakes, warehouses, ML models, etc).

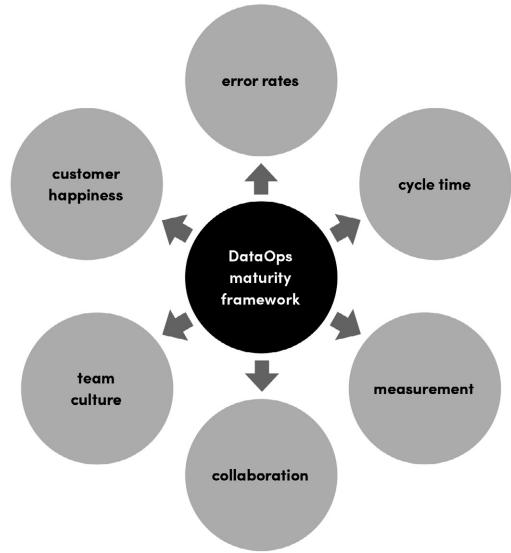
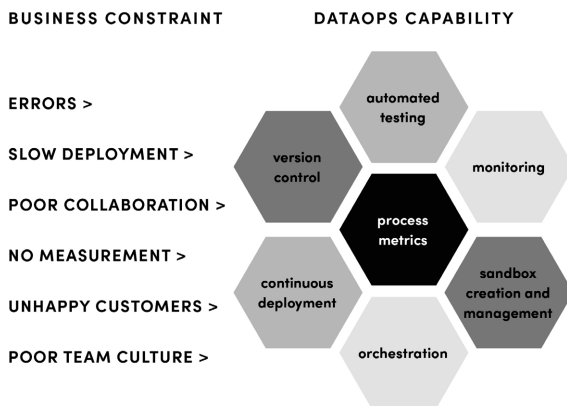


Figure 2: The 6 dimensions of DataOps Maturity

DataOps employs automated orchestration to simplify complex toolchains, environments, and team collaboration, so that the data team can quickly and continuously deliver high quality, error-free insight. To implement DataOps, organizations need to prioritize improvements in the six following areas (Figure 2).

- Error Rates
- Cycle Time
- Collaboration
- Measurement
- Team Culture
- Customer Happiness



**Figure 3: DataOps capabilities that address key business constraints affecting data analytics organizations**

Each of these areas can be improved by implementing core DataOps capabilities such as automated testing and monitoring, toolchain orchestration, version control, sandbox creation and management, and continuous deployment. Figure 3 highlights areas where DataOps capabilities can help you address key business constraints.

## A DATAOPS MATURITY MODEL

Maturity models are commonly used to measure an organization’s ability to continuously improve in a particular discipline. This document outlines a maturity model measurement approach for building, monitoring, and deploying data and analytics according to DataOps principles. With this model, teams can understand where they are today, and what needs to be done to move up the curve.

The model provides a structure for reviewing your organization’s capabilities across the six different DataOps dimensions. Results will enable you to customize strategies to get started or improve. A robust DataOps program will be optimized across all the dimensions.

## THE SIX PRIMARY DATAOPS DIMENSIONS

A DataOps Maturity Assessment asks questions across the six categories that form the dimensions of the DataOps Maturity Model. Along each dimension, progress toward maturity can be categorized as:

Level 5 Optimized: Focus on continuous improvement and change.

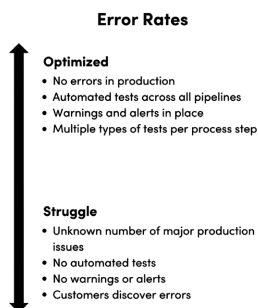
Level 4 Quantitative: Processes are measured and controlled.

Level 3 Consistent: Automated processes are being applied across the entire data analytic development lifecycle.

Level 2 Basic: Processes are documented and partly automated.

Level 1 Struggle: Processes are unrepeatable, poorly controlled, manual, and reactive.

FIGURE 4: Progress toward DataOps maturity for Error Rates.



## Cycle Time

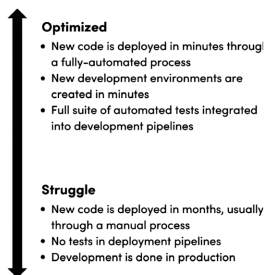


FIGURE 5: Progress toward DataOps maturity for Cycle Time.

Every company is on a journey toward achieving excellence and will have strengths and weaknesses. Just because an organization is large, does not mean that it is excellent. In fact, the flaws in a process or methodology become particularly noticeable when a team grows. Low ratings on any dimension should not be viewed as a negative, but instead as an opportunity for improvement.

## PRODUCTION ERROR RATES

Organizations that follow DataOps principles typically have less than one error per year. That is orders of magnitude better than the industry norm. In a recent DataOps survey, only 3% of the companies surveyed approached that level of quality. Eighty percent of companies surveyed reported three or more errors per month. Thirty percent of respondents reported more than 11 errors per month.

To reduce the level of errors, robust DataOps programs use [automated testing, monitoring, and orchestration in their production pipelines](#). Inspired by statistical process control, they will have tests running in production across all pipelines, sources, and tools, multiple types of tests per process step, and error alerts in place.

In contrast, teams that struggle will have no automated tests in production. This results in costly and embarrassing errors, often discovered by customers. (Figure 4)

## DEPLOYMENT CYCLE TIME

Many organizations experience lengthy cycle times for creating analytic environments or deploying new [analytics that run weeks and months](#). This is often due to manual processes with little to no automation or automated testing in place. In the worst case scenario, development work is done in the production technical environment that also hosts live data operations.

Deployment cycle time can be shortened through a strong program of testing, deployment automation, and environment management. Optimized DataOps programs can deploy new analytics and create new development environments in hours, or even minutes. In DataOps, [error-free automated deployment](#) is realized through a full suite of tests. (Figure 5)

FIGURE 6: Progress toward DataOps maturity for Collaboration.

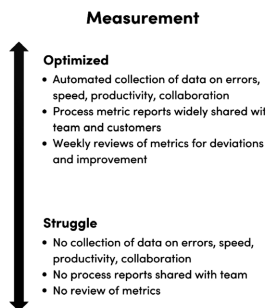
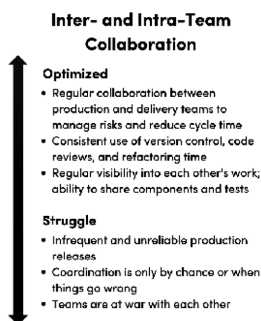


FIGURE 7: Progress toward DataOps maturity for Measurement.

## WELL-COORDINATED INTER- AND INTRA-TEAM COLLABORATION

Organizations with an optimized DataOps program have high levels of inter- and intra-team collaboration. In these organizations, production and development teams regularly collaborate to reduce risk, speed cycle time, and achieve overall greater productivity. They are able to share assets between teams and have visibility into each other's work. These teams also leverage environments, version control, and code review processes for successful collaboration.

On the contrary, organizations that struggle tend to have unreliable processes, chance meetings (and often yelling when things go wrong). Analytic and line-of-business teams are often at war with each other. (Figure 6)

## SUCCESS AND FAILURE MEASURED

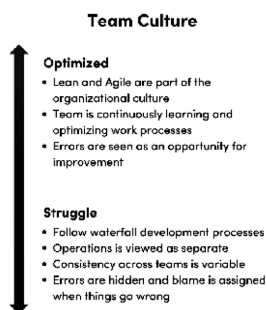
You can't improve what you don't measure, yet it is surprising how many data analytics teams don't measure their own processes. Optimized DataOps teams continuously measure success and failure through detailed process analytics on errors, deployment speed, and team productivity. Metrics are regularly shared and reviewed with the team and internal customers, with a focus on improvement. Conversely, teams that struggle don't track metrics or create reports, or worse, don't collect any data at all. (Figure 7)

## TEAM CULTURE

DataOps draws upon the principles of Agile and Lean manufacturing to transform processes that manage data on its journey toward value creation. Successful DataOps teams follow Agile principles which are a strong part of the overall company culture. These organizations are focused on continuous learning and optimization and errors are viewed as an opportunity for improvement.

On the opposite end of the spectrum are companies that follow waterfall principles. Errors go undiscovered or are hidden and blame is passed around when things go wrong. (Figure 8)

FIGURE 8: Progress toward DataOps maturity for Team Culture.



## Happy Customers

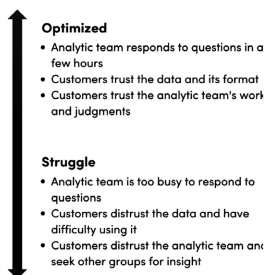


FIGURE 9: Progress toward DataOps maturity for Happy Customers.

## HAPPY CUSTOMERS

Customers won't adopt analytics they don't trust. At the end of the day, delivering trusted, timely insight is a critical measure of success. Best-in-class data teams will respond to customer requests within hours and always provide timely, useful insight. Teams that struggle are often 'too busy' to respond. Customers begin to look elsewhere for insight – the death knell for a data analytics team. (Figure 9)

## WHAT DOES GOOD LOOK LIKE?

No company will initially excel across all six dimensions. Initially, typical results will look like Figure 10. However, the results of your maturity assessment will help your organization plan a roadmap for success. The goal of a DataOps maturity assessment is to provoke a discussion across your organization, to identify areas for improvement, and to guide investment in the processes and tools that can help.

The good news is that success in one particular dimension does not have to be traded-off against others. In organizations that don't practice DataOps, this is a common practice. They often trade speed for quality (or vice versa). For example, in order to reduce fear and uncertainty over errors, a team may establish practices, like documentation, checks and balances, and lots of meetings, that lengthen their cycle time and reduce productivity. With DataOps practices in

## DataOps Maturity Model Score

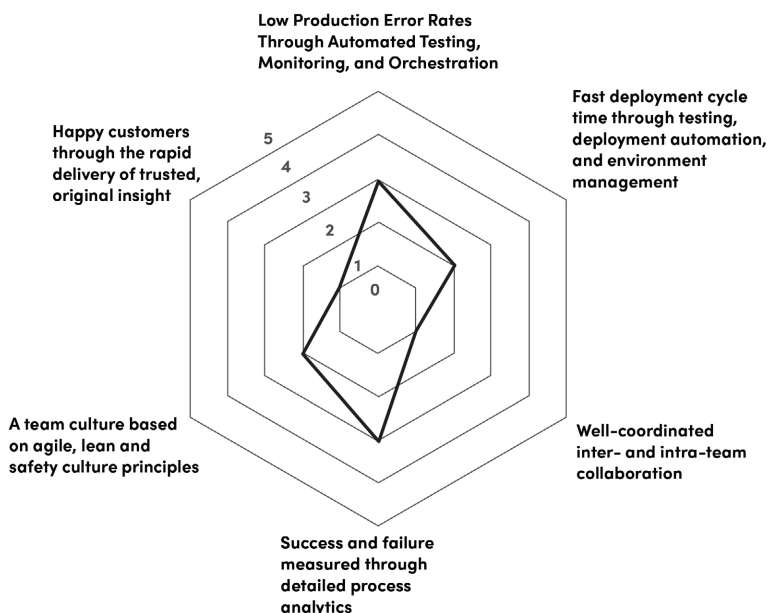


Figure 10: Typical DataOps Maturity Model results

place, you can excel in both speed and quality. Best-in-class data organizations do well across the board, leading to overall greater productivity and lower costs.

By focusing on the right areas, a data team can start to look more like Bristol Myers Squibb (formerly Celgene), a company that is now several years into their DataOps journey (Figure 11). This team initially overcame obstacles that prevented analytics responsiveness and quality. Data was organized in silos – using a variety of technologies and isolated platforms. Without the right processes and tools in place, the data engineering and analytics teams spent a majority of their time on data engineering and pipeline maintenance. This distracted them from their main mission – producing analytic insights that help the business attain its objectives.

After implementing DataOps, they now achieve excellence across all critical dimensions:

- Very, very few errors or missed SLAs
- Weekly cycle time of new changes/features/data
- Detailed process metrics
- Agile culture
- High inter- and intra-team coordination
- High customer satisfaction



## BMS DataOps Maturity Model Score

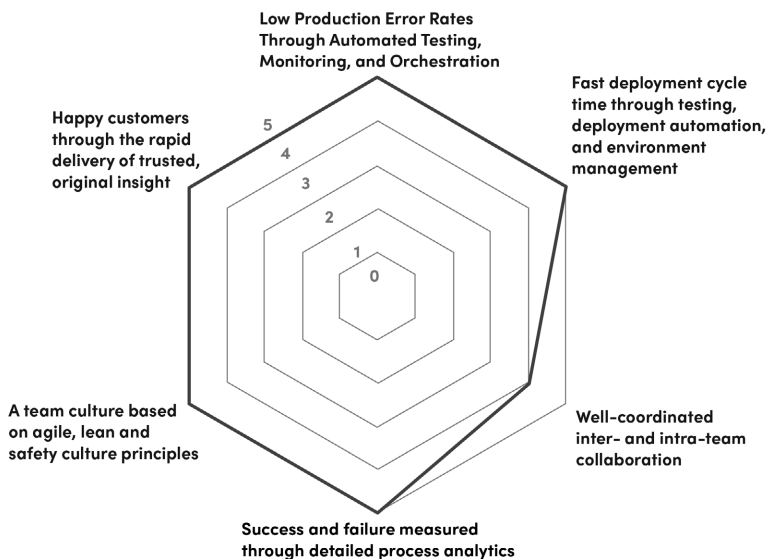


Figure 11: BMS is optimized across all dimensions of the DataOps maturity model

### THE JOURNEY TO EXCELLENCE

As most organizations come to recognize the benefits of a DataOps program, adoption is often a no-brainer. DataOps provides the foundation for analytic excellence. It streamlines the development of new analytics, shortens cycle time, and automates the data analytic pipeline, freeing the team to focus on value-add activities. It also controls the quality of the data flowing through the pipeline so users can trust their data. With DataOps in place, the team is productive, responsive, and efficient.

Because implementation of DataOps requires a mindset shift, one of the biggest challenges becomes where and how to start. The DataOps Maturity Model provides a quick objective way for organizations to assess the maturity of their DataOps initiative and breaks down the critical elements of a DataOps program into concrete, actionable areas for improvement.



# Slovak Sunday Bone Broth Soup

*Contributed By Michael Hutnyan*

## INGREDIENTS

a 3 L/quart stockpot  
bones (roughly 600-900 g/20-32 oz)  
salt  
1 onion  
chunk celery root, 1/2 kohlrabi, cabbage core, etc  
parsley top  
4-6 carrots  
2-3 parsley root  
1 small zucchini  
dried vegetable flavoring  
parsley for garnish  
Egg noodles

## INSTRUCTIONS

1. Wash bones, place in pot and fill 3/4 with cold water. Add about 1 1/2 tbsp of salt (more or less to preference) and set over low heat. The broth should never boil away, only have the occasional bubble rise to the top. If it does boil, of course, it's still tasty, but the broth will be cloudy instead of clear.
2. Add a peeled onion cut in half, the first vegetables (celery root, kohlrabi, cabbage core, etc), and parsley top (if you have one). Leave the broth on low heat for at least 3-4 hrs, if not longer.
3. About an hour before serving, add carrots and parsley root. Do not slice, although you can cut them in half lengthwise if they are bigger.
4. Make zucchini noodles and cut into 2 inch/5 cm lengths. I like to put them in a sieve and put the sieve into the broth for a few minutes to warm up and soften the noodles but not cook them.
5. Strain out the carrots and parsley root, cool for a minute, and chop. Put carrots, parsley, and zucchini noodles in a soup tureen, large mixing bowl, or another pot.
6. Ladle the hot broth through a sieve into the soup tureen, sprinkle some dried vegetable flavoring and/or salt to taste, add a handful of chopped parsley.
7. Serve piping hot over cooked egg noodles. Hot pepper can be added to individual bowls if desired.

## NOTES

- Use only raw bones. No roasting bones beforehand, no leftovers from roasted carcasses. The flavor is different.
- Use bones from any animal, preferably raised in a sustainable manner. Beef Marrow Bones preferred.
- A bit of fat (or skin) and meat on the bones adds flavor.
- Note that the vegetables are put in whole, or cut in half, don't cut them up when putting the soup together.
- If you don't have parsley root, parsnips would do as well. If you don't have either, leave it out.
- The more vegetables go in at the beginning, the sweeter the broth will be, you can choose as many or as little as you like. I save green cabbage cores or cauliflower stems in the freezer and throw them in as well.
- This recipe is for three liters/quarts

# Jump-Starting Your DataOps Journey



At DataKitchen, we are believers in delivering value. We work with our customers to find a first project that can drive real benefits and meet their critical business needs. Customers use our DataKitchen technology and their experience to address a focused business problem.

Recently we've been working with customers in various industries: transportation, telecommunication, and consumer goods to jump-start their DataOps journey. Their business users often have no concept of what it takes to design and deploy robust data analytics. The gap between expectations and execution is one of the main obstacles keeping these analytics teams from succeeding. Managers may ask for a simple change to a report or model or a new dataset. They don't expect it to take weeks or months.

These teams are trying to answer two simple questions. First, how can their team collaborate to reduce the cycle time to create and deploy new data analytics (data, models, transformation, visualizations, etc.) without introducing errors? And second, where to start this process? We've written about how to apply the ['Theory of Constraints'](#) to choosing your first DataOps win. The answer relates to finding and eliminating the bottlenecks that slow down analytics development.

What follows are examples of different types of bottlenecks, why they were selected first, and the benefits of resolving those bottlenecks with DataKitchen.

# ENABLING RAPID DEPLOYMENT TO PRODUCTION: FROM MONTHS TO DAYS

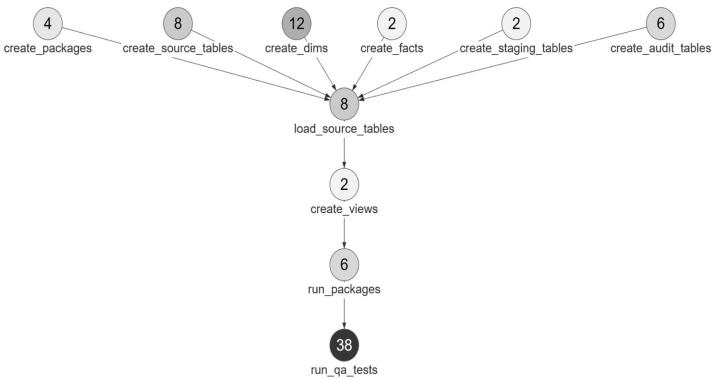
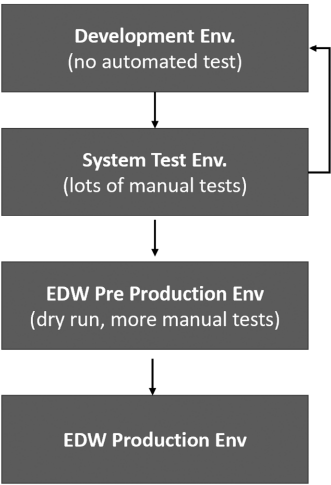
A telecom company needs to increase the rate at which new data features are delivered into production in the EDW. In one past example, it took four months to complete the development cycle from new ideas into production (i.e., to move code from development to production).

We worked to identify outcomes that will define success. Those include:

- Automate the manual tests, which alone will bring about major improvements in cycle time.
- Add new activities into the integration and test environment that will dramatically decrease the time it takes to find and fix issues, thus speeding features into production.
- Rapid addition of features and deployment into production in order to reduce the time from months to days (or faster).

Part of their challenge is that their current process involves a four-stage manual deployment from development to production (see below). This manual process introduces complexity, slowness, and errors (see diagram).

This customer has many tools, including data science, visualization and governance tools, in their analytics toolchain. They chose to focus on changes to their core data warehouse as the first bottleneck to address with DataOps. Their other teams have similar challenges, but a focused adoption by the data warehouse team was the biggest bottleneck that offered a significant short-term business benefit if addressed.



So how do you make sure that when you move something from a Dev Environment through each of the other separate environments, when the business needs it, that everything still works? The answer is automated testing. This company, like many today had very little automated testing in place. Almost all testing to prove that new code works (in this case they are using SQL-based data transformation on Oracle DB) was done by hand. The DataKitchen Recipe below describes how they created dozens of automated tests in the DataKitchen platform that prove that everything works as they moved the new SQL code from one environment to another.

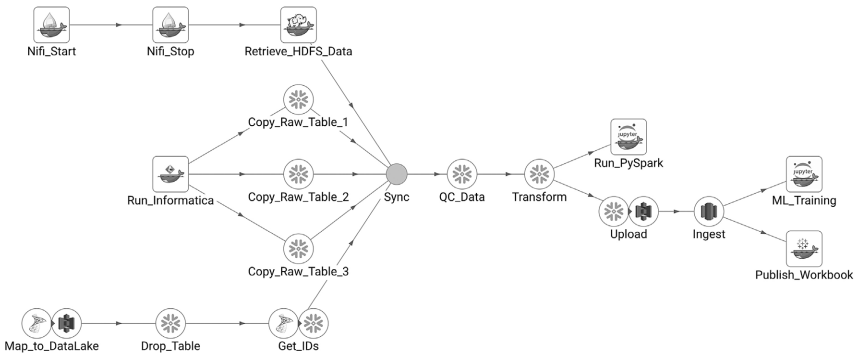


In some ways, the bottleneck is not just technical in nature. Quickly moving code from development into production can be scary – with sometimes painful and costly business implications. What if we make a mistake? Will we get yelled at by the business? Will the business make a critical (and wrong!) decision based on erroneous data? Ensuring new feature deployment success requires both a platform like DataKitchen and an effective approach to writing tests. As part of working with the customer we spend time educating them on how to write great tests as well as the core principles of DataOps.

## **REDUCING ERRORS IN A MULTI-TECHNOLOGY TOOLCHAIN THROUGH IMPROVED ORCHESTRATION AND COLLABORATION**

A transportation company has challenges, both real-time and in batch, managing the workflow orchestration of data streaming from their vehicles into actionable insight for their employees. Like many companies, they do not have just one data architecture, they have several – batch, streaming, big data, small data, on-premises, cloud, and prescriptive and predictive models – all working together. Plus, they have different teams managing the creation and the operation of these pipelines in different locations. Whew! What that means is that the disparate teams need to develop a common outcome (report, dashboard, model, etc.) together by addressing these business challenges:

- Data Pipelines = complexity – tools, platforms, teams
- Complexity = delays and risk – in value and innovation
- Slow delivery of data products (Data Science)
- Too much manual intervention/testing (Business Reporting)
- Slower delivery than desired



But to meet those challenges, they need to work on their current data operations, technologies and delivery:

- **Orchestration:** Running pipelines of various technologies at the right time
- **Iteration:** Rapidly creating, iterating and deploying data science and data engineering pipelines and their data products (reports, dashboards, models) – Full Pipeline CI/CD
- **Quality:** Detecting issues and errors in complex, multi-tool data pipelines
- **Hybrid:** Enabling on-premise and cloud approaches and an evolving technology landscape
- **Collaboration:** Enabling multiple teams to collaborate more effectively to deliver data products to end-users faster with higher quality. Also enabling transparency of all operations in a complex pipeline.

So, what bottleneck did they focus on first? Orchestration of the toolchain for low error execution and enhanced collaboration. They started with a variety of technologies that the company currently uses, including on-prem:

- Oracle
- SQL Server
- Informatica
- Apache Nifi
- Hadoop/HDFS
- PySpark Cloud
- Redshift
- S3
- Tableau online
- MLflow

They then created a single DataKitchen Recipe, which provides a framework to detect issues in streaming data across a multi-technology toolchain, runs tests that ensure the quality of streaming data, and alerts users if the condition is not met. That Recipe uses all their existing tools to transform data into



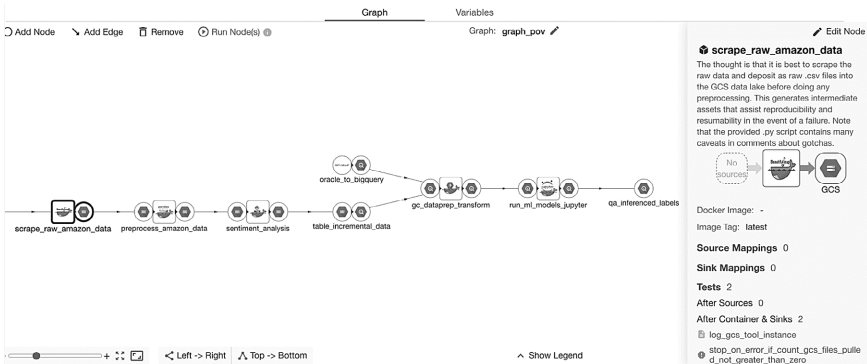
business insight. Other important considerations for the team were that the data was validated to be ‘fit for purpose’ so that the assumptions made while transforming data remain true. They also focused upon Recipe CI/CD so that once a change passes its tests across any of the technologies, it is deployed to the pipeline in less than five minutes.

## **FREED DATA SCIENTISTS’ TIME THROUGH AUTOMATION OF MACHINE LEARNING DEPLOYMENT AND PRODUCTION**

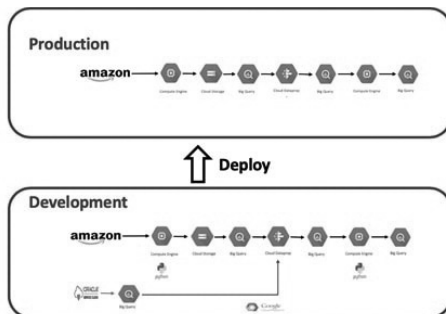
A consumer product team recently completed a machine learning (ML) use case combining web scraped data with Oracle service cloud data. It works great; the business users love the early versions of the product, but:

- The various pipeline components are run by hand in an inconsistent manner, and the orchestration is not automated
- The movement of code from sandbox to development is manual
- There are few tests to ensure data quality
- The web scraping component frequently fails, and there is insufficient notification and restart capabilities

DataKitchen and the company have identified another bottleneck that first needs to be addressed. Talented (and expensive) Data Scientists can create the first version of an idea but have no interest in running it on a day to day basis. How can that ML model and all its associated data transformation, code, and UI be put into operation following a DataOps Approach? To address this bottleneck the team created a Recipe in DataKitchen that provides the ability to orchestrate the components developed in the ML use case. That Recipe has the ability to run data tests and detect errors in the data in the production environment. It also has the ability to detect operational errors (e.g., the web scraping issues), alert users to those errors, and permit re-start of the processing of the data pipeline (Recipe). See diagram below.



## WHERE IS YOUR BOTTLENECK? WHERE CAN YOU START DOING DATAOPS TODAY?



DataOps applies lean manufacturing management methods to data analytics. One leading method, the Theory of Constraints, focuses on identifying and alleviating bottlenecks. Data analytics can apply this method to address the constraints that prevent the data analytics organization from achieving its peak levels of productivity. Bottlenecks lengthen the cycle time of developing new analytics and prevent the team from responding quickly to requests for new analytics. If these bottlenecks can be improved or eliminated, the team can move faster, developing, and deploying with a high level of quality in record time. If you have multiple bottlenecks, you can't address them all at once. As these examples have shown, there are multiple ways to provide immediate, clear value for doing DataOps!

# Peanut Butter Energy Bites

*Contributed By Eric Estabrooks*

## INGREDIENTS

- 2/3 cup creamy peanut butter
- 1/2 cup semi-sweet chocolate chips
- 1 cup old fashioned oats
- 1/2 cup ground flax seeds
- 2 tablespoons honey
- 1/2 teaspoon vanilla extract

## INSTRUCTIONS

1. Combine all 5 ingredients in a medium bowl. Stir to combine.
2. Place in the refrigerator for 15-30 minutes so they are easier to roll.
3. Roll into 12 bites and store in the fridge for up to a week. An ice cream scooper is a good measurement.



## 4 Easy Ways to Start DataOps Today



The primary source of information about [DataOps](#) is from vendors (like DataKitchen) who sell enterprise software into the fast-growing DataOps market. There are over [100 vendors](#) that would be happy to assist in your DataOps initiative. Here's something you likely won't hear from any of them (except us) — you can start your DataOps journey without buying any software.

It's important to remember that DataOps is a culture and methodology, implemented using automated augmentation of your existing tools. You are free to select one of many best-in-class free and open source tools. When we started sharing the [“Seven Steps of DataOps”](#) a few years ago, our intent was (and still is) to evangelize DataOps as a free and open methodology.

If you are a CDO or a VP, you have the power to institute broad change, but what if you are an individual contributor? What can you do? This is a common question that we hear from our conversations with data scientists, engineers and analysts. An individual contributor has assigned duties and usually no ability to approve purchases. How can one get started given these limitations?

DataOps is not an all-or-nothing proposition. There are small but impactful things that an individual contributor can do to move forward. Hopefully, with metrics in place, you can show measured improvements in productivity and quality that will win converts. As your DataOps activities reach enterprise scale, you may indeed decide that it's much easier to partner with a vendor than to build and support an end-to-end DataOps Platform from scratch. When that day arrives, we'll be here, but until then, here are some suggestions for DataOps-aligned improvements you can make with open-source tools and a little self-initiative.

## DATAOPS OBJECTIVES

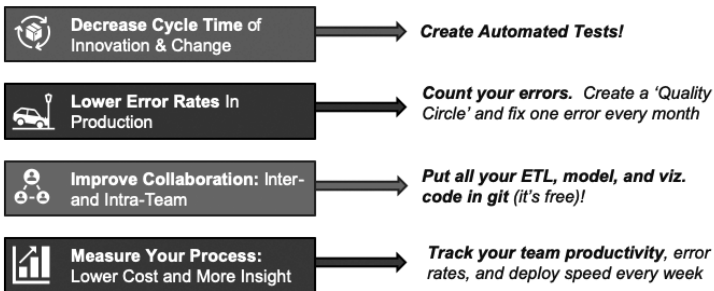
DataOps includes four key objectives:

- **Measure Your Process** — As data professionals, we advocate for the benefits of **data-driven decision making**. Yet, many are surprisingly unanalytical about the activities relating to their own work.
- **Improve Collaboration**, both **Inter-** and **Intra-**team — If the individuals in your data analytics team don't work together, it can impact analytics cycle time, data quality, governance, security and more. Perhaps more importantly, it's fun to work on a high-achieving team.
- **Lower Error Rates in Development and Operations** — Finding your errors is the first step to eliminating them.
- **Decrease the Cycle Time** of Change — Reduce the time that elapses from the conceptualization of a new idea or question to the delivery of robust analytics.

We view the steps in analytics creation and data operations as a manufacturing process. Like any complex, procedure-based workflow, the data analytics pipeline has bottlenecks. We subscribe to the **Theory of Constraints**, which advises to find and mitigate your bottlenecks to increase the throughput of your overall system.

If that's too abstract, we'll suggest four projects, one in each of the areas above, that will start the ball rolling on your DataOps initiative. These tasks illustrate how an individual contributor can start to implement DataOps on their own.

### Don't Buy Our Software ... Start Today With DataOps



4 simple projects to get started with DataOps.

## MEASURE YOUR PROCESS

**Internal analytics** could help you pinpoint areas of concern or provide a big-picture assessment of the state of the analytics team. A burn-down chart, velocity chart, or tornado report can help your team understand its bottlenecks. A data arrival report enables you to track data suppliers and quickly spot delivery issues. Test Coverage and Inventory Reports show the degree of test coverage of the data analytics pipeline. Statistical process controls allow the data analytics team to monitor streaming data and the end-to-end pipeline, ensuring that everything is operating as expected. A Net Promoter Score is a customer satisfaction metric that gauges a team's effectiveness.

	Source 1	Source 2	Source 3	Source 4	Source 5
3/13/16					
3/12/16					
3/11/16					
3/10/16					
3/9/16					
3/8/16					
3/7/16					
3/6/16					
3/5/16					
3/4/16					
3/3/16					

The data arrival report shows which data sources meet their target service levels.

When you bring these reports to the team, it will help everyone understand where time and resources are being wasted. Perhaps this will inspire a project to mitigate your worst bottleneck, leading to another project in one of the next areas.

## IMPROVE COLLABORATION

Conceptually, the data analytics pipeline is a set of stages implemented using a wide variety of tools. All of the artifacts associated with these tools (JSON, XML, scripts, ...) are just source code. Code deterministically controls the entire data analytics pipeline from end to end.

If the code that runs your data pipeline is not in source control, then it may be spread out on different systems, not revision controlled, even misplaced. You can take a big step toward establishing a controlled, repeatable data pipeline by putting all your code in a source code repository. For example, Git is a free and open-source, distributed version control system used by many software developers. With version control, your team will be better able to reuse code, work in parallel and trace bugs back to source code changes. Version control also serves as the foundation for DataOps continuous deployment, which is an excellent long-term goal.

## LOWER ERROR RATES

Maybe the test coverage report mentioned above helped you understand that your data operations pipeline needs more tests. Tests apply to [code](#) (analytics) and streaming [data](#). Tests can verify inputs, outputs and business logic at each stage of the data pipeline. Testing should also confirm that new analytics integrate seamlessly into the current production pipeline.

Below are some example tests:

- The number of customers should always be above a certain threshold value.
- The number of customers is not decreasing.
- The zip code for pharmacies has five digits.

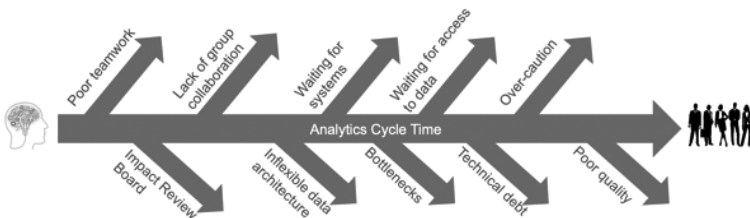
```
1 select count(*) from customers :: is above a threshold
2
3 select count(*) from customers :: is larger or equal to than last time
4
5 select count(*) from raw_specialty_pharmacy where len(zip) != 5 :: is zero
6
```

Every processing or transformation step should include tests that check inputs, outputs and evaluate results against business logic.

When you have started counting and cataloging your errors, start a quality circle, find patterns and aim to fix one error per month.

## DECREASE THE CYCLE TIME OF CHANGE

In many enterprises, lengthy cycle time is a primary reason that analytics fail to deliver on the promise of improving data-driven decision making. When the process for creating new analytics depends on manual processes, there are many opportunities for a project to go off track.



Factors that derail the development team and lengthen analytics cycle time

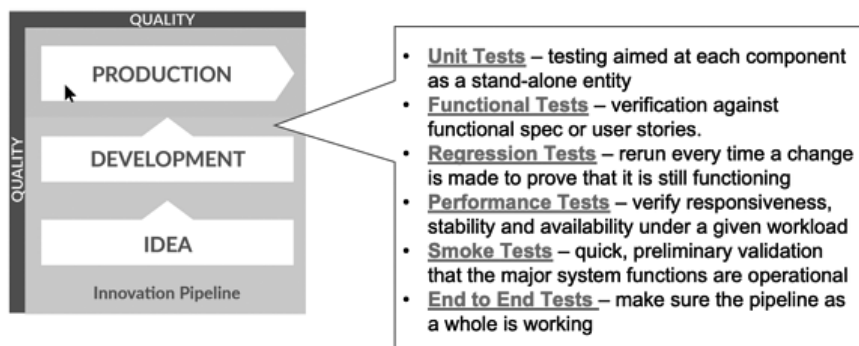
Leading software organizations deploy new and updated applications through an automated procedure that might resemble something like this:

1. Spin-up hardware and software infrastructure
2. Check source code out of source control
3. Build
4. Test
5. Deploy into production



The first step in creating an efficient, repeatable build process is to minimize any dependencies on manual intervention. Each of these steps is a whole topic unto itself, but when you are starting out, a good place to focus is on testing. Your code tests should fully validate that analytics work, can handle errors such as bad data (by stopping or sending alerts) and integrate with the existing operations pipeline.

The image below shows the many different kinds of tests that should be performed. We explain each of these types of tests in our [Guide to DataOps Tests](#).



A broad set of tests can validate that the analytics work and fit into the overall system.

Tests that validate and monitor new analytics enable you to deploy with confidence. When you have certainty, you can deploy and integrate new analytics more quickly.

## CONCLUSION

There are many small yet effective projects that you can start today that will serve your DataOps goals. Hopefully, we've given you a few ideas.



# Chocolate Stout Cupcakes with Irish Whiskey Filling and Baileys Frosting

*Contributed By Lauren Meyer*

**Prep Time** 30 mins — **Cook Time** 20 mins — **Total Time** 50 mins

Servings: 1 dozen cupcakes

## INGREDIENTS

1/2 cup Guinness (or stout of your choice)  
1/2 cup (1 stick) unsalted butter, room temperature  
1/2 cup cocoa powder  
1 cup all-purpose flour  
1 cup granulated sugar  
1 tsp baking soda  
1/4 tsp salt  
1 large egg  
1/3 cup sour cream  
1 batch Irish Whiskey Filling (recipe below)  
1 batch Baileys Frosting (recipe below)

## IRISH WHISKEY FILLING

4 oz. bittersweet chocolate, finely chopped  
1/3 cup heavy cream  
1 Tbsp unsalted butter, room temperature  
3 Tbsp Baileys Irish Cream  
1 tsp Irish whiskey

## BAILEYS FROSTING

2 cups confectioners' sugar  
1/2 cup (1 stick) unsalted butter, room temperature  
4 Tbsp Baileys Irish Cream

## INSTRUCTIONS

### Stout Cupcakes

1. Preheat oven to 350 degrees and line a 12-cavity cupcake tin with papers.
2. Bring stout and butter to a simmer in a large, heavy saucepan over medium heat. Add cocoa powder to the saucepan and whisk the mixture until it's smooth. Remove saucepan from heat.
3. In a separate medium bowl, whisk together the flour, sugar, baking soda, and salt.
4. In the bowl of a stand mixer or in a separate large bowl with a hand mixer (or whisk) beat together egg and sour cream, until combined.
5. Add the chocolate stout mixture to the egg mixture and beat until just combined.
6. Add the dry mixture to the wet mixture and mix until just combined, taking care not to over-mix.
7. Divide batter among cupcake liners, filling them about  $\frac{3}{4}$  of the way.
8. Bake for about 17-20 minutes, until a toothpick stuck into the center of a cupcake comes out clean.
9. Let the cupcakes cool in the pan for a few minutes and then take them out to cool completely on a wire rack. Once cupcakes are cooled completely, core out a small section from the middle using either a knife or a cupcake corer.
10. Spoon Irish whiskey filling into centers of cupcakes. Frost cupcakes with Bailey's frosting. I used a Wilton 1A pastry tip for mine.

### Irish Whiskey Filling

1. Place the chocolate in a heatproof bowl.
2. In a small saucepan, bring cream just to a boil (keep a close eye on it and remove from heat right when it starts boiling). Pour cream over chocolate in bowl and let sit for 1 minute. Then, stir until chocolate is completely melted and smooth.
3. Add butter, Baileys, and Irish whiskey, and stir to combine.

### Baileys Frosting

1. In the bowl of a mixer or in a large bowl with a hand mixer, mix butter on medium speed until it's nice and fluffy. Add confectioners' sugar one cup at a time and beat until well-combined.
2. Add the Baileys and beat until combined. If frosting is too thin, add more confectioners' sugar a couple tablespoons at a time.

### Notes

Recipe adapted from Serious Eats and Smitten Kitchen.

For an easy way to core cupcakes, see the link for a [cupcake corer](#).

Attribution: [We Are Not Martha](#), Author: Sues



**Establish**



# Finding an Executive Sponsor for Your DataOps Initiative

BY JAMES ROYSTER



DataOps revolutionizes how data analytics work gets done. Like many other “big ideas,” it sometimes faces resistance from within the organization. For most organizations, data is a means to an end. The organization’s primary focus is on its mission, whether that is a product or a service. As data professionals, we communicate the value of data-driven insights. Although many of our colleagues appreciate the value of insight, they generally pay little attention to the process of uncovering that insight unless there is an issue or error.

If you are launching a DataOps initiative, executive sponsorship can give you air cover while building DataOps capabilities on the ground. A C-level sponsor can tie the project’s activities into the larger organization’s strategic goals. An executive can explain the value to others and provide guidance as the project team faces obstacles or grapples with trade-offs. The executive sponsor provides resources and budget as a skunkworks matures into an official project. To pitch a transformational concept like DataOps to an executive, put yourself in his or her shoes.

## CONNECTING DATAOPS TO BUSINESS OUTCOMES

Executives rarely have the opportunity to passively reflect on the past. Every quarter comes with a new goal, and the success or failure of initiatives impacts the company's short and long-term success. If someone comes along with an idea that can improve business outcomes — an executive will be “all ears.” Articulating how DataOps can contribute to the success of a key initiative will speak to an executive's priorities.

Translate DataOps' impact into benefits that your executive understands and cares about. DataOps offers ways to slash analytics development cycle time, streamline workflows, and virtually eliminate errors in data operations. These capabilities help business leaders rapidly capitalize on opportunities and gain insight into the marketplace, often well before the competition.

An executive is always on the lookout for ways to grow revenue and maximize resources. Circumstances present the business with an endless stream of opportunities to make investments that spur growth or implement efficiencies. Companies can't jump on every opportunity. They have to select the best of the bunch based on return-on-investment (ROI), risk assessment, or another preferred metric.

A high-performance organization maximizes its ability to select and leverage opportunities. Data is the modern business decision apparatus (just ask Google, Target, Amazon, or Facebook). If people leverage their data more effectively and rapidly, and with fewer errors, they can pursue opportunities more quickly and efficiently. DataOps improves business agility, which itself sustains a competitive advantage.

## THE OPPORTUNITY COST OF INEFFICIENT ANALYTICS

DataOps also improves the efficiencies of data analytics workflows. The data team spends less time on manual processes, such as data prep, integration, documentation, execution of data operations and recovery from errors, and more time on new models and analytics that create value. Living with less efficient data analytics workflows has an opportunity cost. Without DataOps automation, the enterprise pursues fewer opportunities or the wrong opportunities.

Linking the overt benefits of DataOps to business impact is key to earning executive support. If you pitch DataOps only as a way to make data analytics more efficient, an executive will likely not see the full value. Executives have a tremendous responsibility to the organization and its employees, so they must carefully choose where they place their energies. Your role is to articulate how DataOps can impact objectives in the business domain. Connect the dots for how DataOps helps the organization more effectively achieve its mission.



## INSIGHTS THAT TRANSFORM THE ENTERPRISE

When pitching your DataOps project to a potential executive sponsor, it may help to discuss it relative to a broader strategy that you outline, for example, in a slide presentation. You may be starting with a single project, but DataOps can help improve metrics that reflect teamwork, productivity, quality, and more. DataOps is a transformational concept that revolutionizes how data science and analytics work gets done. Ultimately, the impact of DataOps extends beyond just the data team. It promotes collaboration across the entire enterprise and, through analytics, helps people discover creative insights that stimulate growth.



# Gil's Easy Chicken Cacciatore

*Contributed By Gil Benghiat*

If you can cut things into pieces, you can make this easy recipe.

## INGREDIENTS

- 6-8 boneless, skinless chicken thighs (about 2-4 pounds) – or an equivalent type of chicken
- 3 peppers (red, yellow, and orange for some color)
- 1 large onion
- 28 oz can of crushed tomatoes
- 6 oz can of tomato paste
- 1 package of sliced mushrooms (about 2 cups) - optional
- 2 teaspoons turmeric

## INSTRUCTIONS

- Cut chicken, peppers, and onion into pieces
- Combine and mix (with a DataKitchen spoon) all ingredients in a pot
- Until the chicken is cooked and the vegetables are soft and tender simmer covered on the stove for 45-60 minutes or bake in an oven-proof pot (e.g. a Dutch oven) at 350 degrees for 45-60 minutes.
- While cooking mix with a DataKitchen spoon every 20 minutes.
- Serve over your favorite pasta.



# Pitching a DataOps Project That Matters



DataOps addresses a broad set of use cases because it applies workflow process automation to the end-to-end data analytics lifecycle. DataOps reduces errors, shortens cycle time, eliminates unplanned work, increases innovation, improves teamwork, and more. Each of these improvements can be measured and iterated upon.

These benefits are hugely important for data professionals, but if you made a pitch like this to a typical executive, you probably wouldn't generate much enthusiasm. Your data consumers are focused on business objectives. They need to grow sales, pursue new business opportunities, or reduce costs. They have very little understanding of what it means to create development environments in a day versus several weeks. How does that help them "evaluate a new M&A opportunity by Friday?"

If you pitch DataOps in terms of its technical benefits, an executive or co-worker might not understand its full potential value. Instead, explain how agile and

error-free analytics serves the organization's mission. What would it mean to monetize data more effectively than competitors? Data is the modern business decision apparatus (just ask Google, Target, Amazon, or Facebook). DataOps enables companies to rapidly assess and pursue opportunities, avoiding strategic mistakes, and shrinking time-to-market. What would it mean for a company to lead its industry in savvy and business agility? When discussing a DataOps initiative with an executive or colleague, focus on his/her top business objective and find a project related to it. Impactful DataOps projects are those that help colleagues and executives pursue their objectives. Below we suggest some additional unconventional approaches to finding high-visibility DataOps projects.

## **FIND UNHAPPY ANALYTICS USERS**

A strained relationship between the data team and users can point to a potential DataOps pilot project. A data team with unhappy users is ripe for transformational change. You may instinctively wish to turn away from grumbling users. You should be thankful for them. The more vocal and unhappy the customers are, the bigger the opportunity to turn the situation around and bring high-impact improvements to the broadest possible group. A large community of dissatisfied customers is also likely to be a higher priority for managers and executives. Ask your unhappy customers or colleagues what concerns them most about the data analytics team. User discontent may be expressed in feelings and observations. User surveys can organize and quantify user anecdotes into actionable priorities. The list of possible issues is long, but you might hear feedback that includes:

- Data science/engineering/analytic teams do not deliver the insight that the business customers need.
- The data team takes too long to deliver analytics.
- Users mistrust the data itself or the team working on the data.
- Stakeholders have hired consultants or shadow teams to do data work.

## **BE GRATEFUL FOR NEGATIVE FEEDBACK**

Negative feedback often stems from deep, underlying issues. The data team may not deliver relevant analytics because business users and data analysts are isolated from each other. Users may mistrust data and analytics because of errors. When business units hire their own data analysts, it's a sign that they are underserved. They may feel like the data organization is not addressing their priorities.

User feedback may feel concrete to users, but as a data professional, you will have to translate these requirements into metrics. For example, users may not trust the data. That may seem abstract and not directly actionable. Try measuring your errors per week. If you can show users that you are lowering that number, you can build trust. A test coverage dashboard can illustrate progress in quality controls. Demonstrating your success with data can help gradually win over detractors. What other problems have eroded trust? You may need to look for more than one contributing factor.

In many organizations, analytics follows a complex path from raw data to processed analytics that create value. Your data crosses organizational boundaries, data centers, teams, and organizations. Errors can creep in anywhere along this path. What are the historical drivers of issues/errors? Which teams own each part of the process? A lack of responsiveness sometimes squanders trust. Measure how fast teams can respond to errors and requests.

Another common user complaint is that data analytics teams take too long to deliver requested features. The length of time required to deliver analytics can be expressed in a metric called cycle time. Benchmark how fast you can deploy new ideas or requests into production. To reduce cycle time, examine the data science/engineering/analytic development process. For example, how long does it take to create a development environment? How up-to-date are development environments? How well-governed are development environments?

## **CREATING A FEEDBACK LOOP OF TRUST**

As DataOps improves trust in data and data team responsiveness, business users will naturally begin to work more closely with the data team. As the data team becomes more agile, interaction with users increases in importance. DataOps focuses on delivering value to customers in short, frequent iterations. The value that business users receive after interacting with the data team reinforces the value of working together. DataOps enterprises frequently observe greater and more frequent communication and collaboration between users and the data team. The positive feedback loop of collaboration and value creation encourages users and data professionals to invest in working closely together. In the end, the quality of collaboration that DataOps fosters becomes the engine that takes an organization to new heights.





# Risotto alla Monzese

*Contributed By Gianluca Paris*

Serves 4

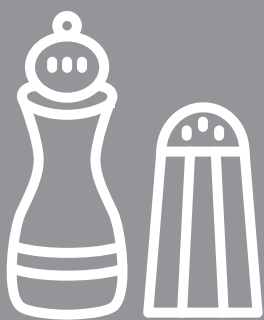
## INGREDIENTS

320g of big grain rice  
1 red onion  
1 sausage  
100g of butter  
40g grated Grana Padano cheese  
A bag of saffron  
A pot of vegetable broth  
A spoon and a half of olive oil  
Salt

## INSTRUCTIONS

1. Chop the red onion. Put a spoon and a half of olive oil into a pot, wait for it to be hot and fry the red onion.
2. In the meantime boil the broth in another pot and keep it hot for the whole recipe time as you will need it.
3. Cut the sausage, when the onion changes color, put the sausage into the pot and fry it, then add the rice.
4. Once the rice becomes transparent, add broth until the content of the pot gets covered. At the same time add the bag of saffron.
5. When the broth gets absorbed completely, add another ladle of broth and continue this way for 18–20 minutes (add broth only when is absorbed!). After 18–20 minutes the rice will be cooked. Now turn off the stove (mandatory!) and add salt, butter (from the fridge) and cheese.
6. Stir until the butter melts completely and enjoy!





**Demonstrate**



# Prove Your Team's Awesomeness with DataOps Process Analytics

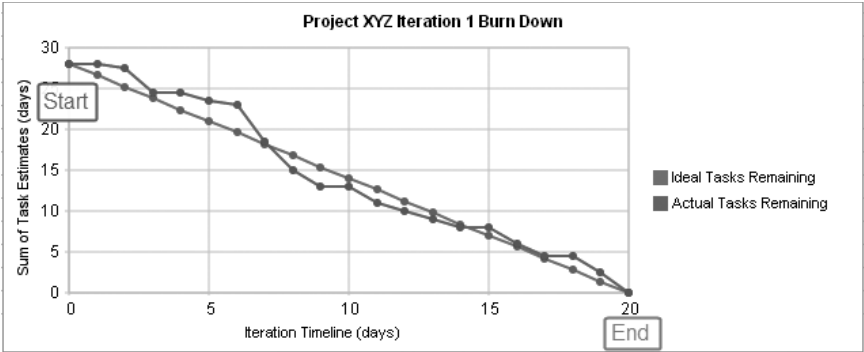


Do you deserve a promotion? You may think to yourself that your work is exceptional. Could you prove it?

As a Chief Data Officer (CDO) or Chief Analytics Officer (CAO), you serve as an advocate for the benefits of [data-driven decision making](#). Yet, many CDO's are surprisingly unanalytical about the activities relating to their own department. Why not use DataOps analytics to shine a light on yourself?

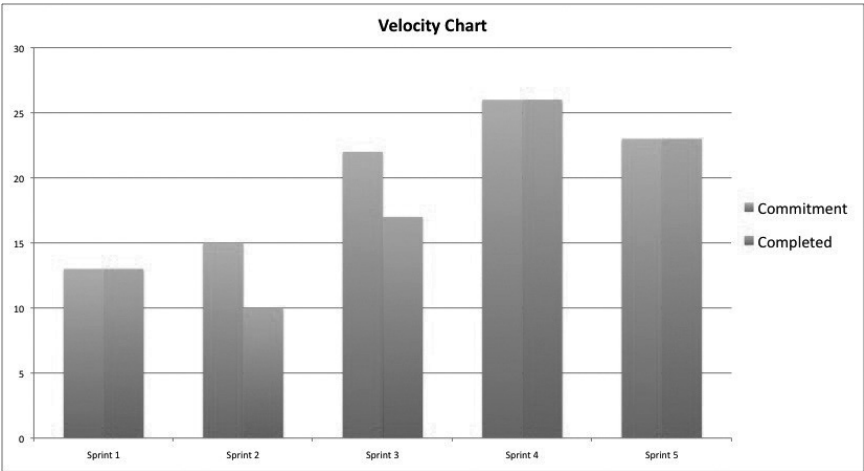
Internal analytics could help you pinpoint areas of concern or provide a big-picture assessment of the state of the [analytics team](#). We call this set of analytics the CDO Dashboard. If you are as good as you think you are, the CDO Dashboard will show how simply [awesome](#) you are at what you do. You might find it helpful to share this information with your [boss](#) when discussing the data analytics department and your plans to take it to the next level. Below are some reports that you might consider including in your CDO dashboard:

**BURN DOWN CHART**



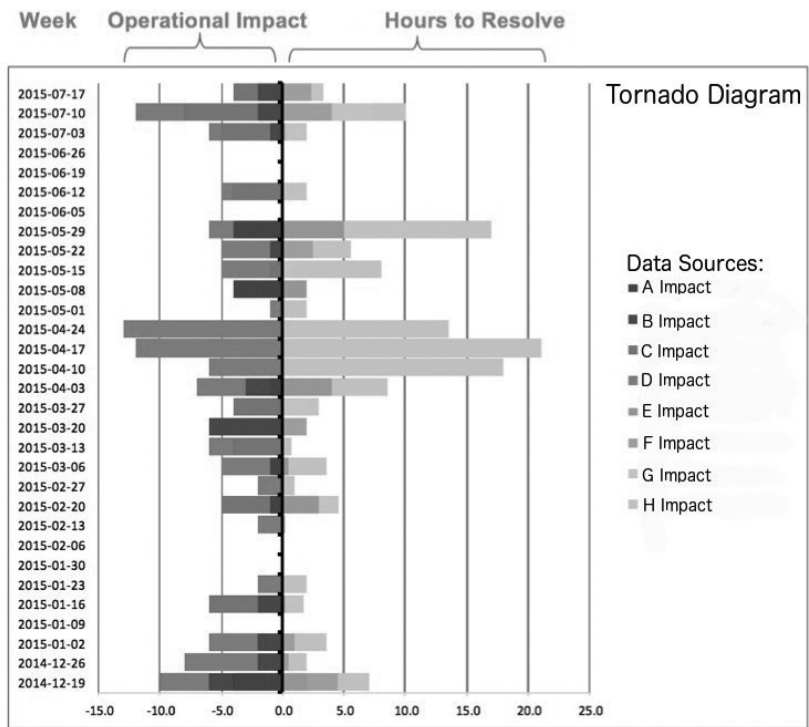
The burn down chart graphically represents the completion of backlog tasks over time. It shows whether a team is on schedule and sheds light on the productivity achieved in each development iteration. It can also show a team’s accuracy in forecasting its own schedule.

**VELOCITY CHART**



The velocity chart shows the amount of work completed during each sprint — it displays how much work the team is doing week in and week out. This chart can illustrate how improved processes and indirect investments (training, tools, process improvements, ...) increase velocity over time.

# TORNADO REPORT



The Tornado Report is a stacked bar chart that displays a weekly representation of the operational impact of production issues and the time required to resolve them. The Tornado Report provides an easy way to see how issues impacted projects and development resources.

DATA ARRIVAL REPORT

	Source 1	Source 2	Source 3	Source 4	Source 5
3/13/16					
3/12/16					
3/11/16					
3/10/16					
3/9/16					
3/8/16					
3/7/16					
3/6/16					
3/5/16					
3/4/16					
3/3/16					
	Key:		missing		
			late		
			on time		

A large organization might receive hundreds of data sets from suppliers and each one could represent dozens of files. All of the data has to arrive error-free in order to, for example, build the critical Friday afternoon report. The Data Arrival report tracks how vendors perform relative to their respective service level agreements (SLA).

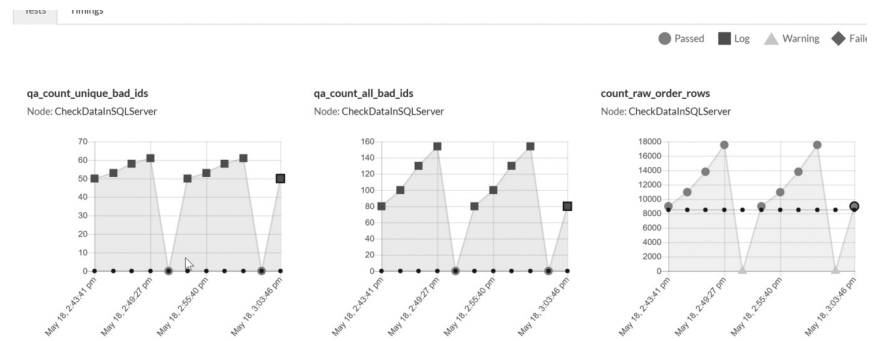
The Data Arrival report enables you to track data suppliers and quickly spot delivery issues. Any partner that causes repeated delays can be targeted for coaching and management. The Tornado Report mentioned above can help quantify how much time is spent managing these issues in order to articulate impact. These numbers are quite useful when coaching a peer organization or vendor to improve its quality.

TEST COVERAGE AND INVENTORY

The Test Coverage and Inventory Reports show the degree of test coverage of the data analytics pipeline. It shows the percent of tables and data covered by tests and how test coverage improves over time. The report can also provide details on each test. In a DataOps enterprise, results from tests run on the production pipeline are linked to real-time alerts. If a process fails with an error, the analytics team can troubleshoot the problem by examining test coverage before or after the point of interest.

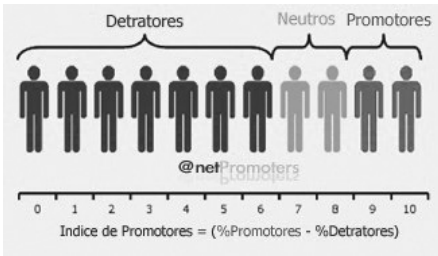


# STATISTICAL PROCESS CONTROLS



The data analytics pipeline is a complex process with steps often too numerous to be monitored manually. [Statistical Process Control](#) (SPC) tests inputs, outputs and business logic at each stage of the pipeline. It allows the data analytics team to monitor the pipeline end-to-end from a big-picture perspective, ensuring that everything is operating as expected.

## NET PROMOTER SCORE



A Net Promoter Score is a customer satisfaction metric that gauges a team's effectiveness. For a data team, this is often a survey of internal users who are served by analytics. The Net Promoter Score can show that the data analytics team is effective at meeting the needs of its internal customer constituency or that satisfaction is improving.

## CONCLUSION

One of the main goals of analytics is to improve decision-making. The CDO DataOps Dashboard puts information at the fingertips of executives, so they have a complete picture of what is happening in the data analytics domain. When it's time to review performance, the CDO DataOps Dashboard can help you show others that the analytics department is a well-oiled machine. Now, about that promotion...



# Grandma's Italian Meatballs

*Contributed By Mark Sampson*

## INGREDIENTS

- 1 lb. Ground Beef
- 1 clove of garlic minced (I like garlic so I use a large clove or more)
- 2 Eggs
- ½ cup of dry parsley flakes
- 1 cup Italian style flavored bread crumbs
- 2 slices of crust removed bread soaked in water (wring out good before adding)
- ½ cup of formaggio (Romano Cheese, the good stuff)

Very little oil: less than a tablespoon (I usually drizzle quickly over mixture)

Salt & Pepper (*I usually shake both to cover the ingredients above*)

## Instructions

Preheat oven to 400 degrees.

Put all ingredients in a big bowl. Mix/Knead well. I always make a marble size tasting ball that I cook in the microwave for about 20-30 seconds – rotating halfway through. I sometimes find I need to add more salt.

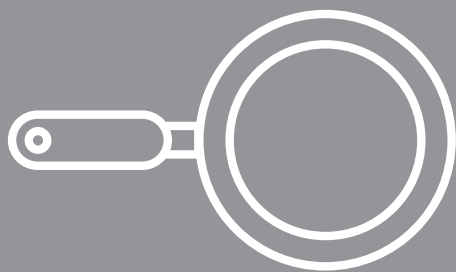
Spray a cookie sheet with olive oil spray (or wipe on olive oil). Roll into ping pong or golf ball-sized spheres.

Bake for 20 minutes (I do 11 minutes, then flip over and bake another 9 minutes). Eat them while they are hot (by themselves or butter a piece of scali bread and put a warm ball in there) or place them in your tomato sauce.

Bonus material: using just plain tomato sauce or crushed tomatoes, use these meatballs to add joyful flavoring by simmering for hours.

Each pound makes about 18-20 golf ball size meatballs (I usually make 3 lbs at a time)





**Iterate**



# Eliminate Your Analytics Development Bottlenecks



## APPLYING THE THEORY OF CONSTRAINTS TO DATA ANALYTICS

Business users often have no concept of what it takes to design and deploy robust data analytics. The gap between expectations and execution is one of the main obstacles holding the analytics team back from delighting its users. Managers may ask for a simple change to a report. They don't expect it to take weeks or months.

Analytics teams need to move faster, but cutting corners invites problems in quality and governance. How can you reduce cycle time to create and deploy new data analytics (data, models, transformation, visualizations, etc.) without introducing errors? The answer relates to finding and eliminating the bottlenecks that slow down analytics development.



Figure 1: The creation of analytics in a large data organization requires the contribution of many groups.

## YOUR DEPLOYMENT PIPELINE

Analytics development in a large data organization typically involves the contribution of several groups. Figure 1 shows how multiple teams work together to produce analytics for the internal or external customer.

Tasks in development organizations are often tracked using Kanban boards, tickets or project tracking tools. Figure 2 is a Kanban board, representing a project, with a yellow sticky note for each task. As tasks progress through milestones, they move from left to right until they reach the “Done” column.

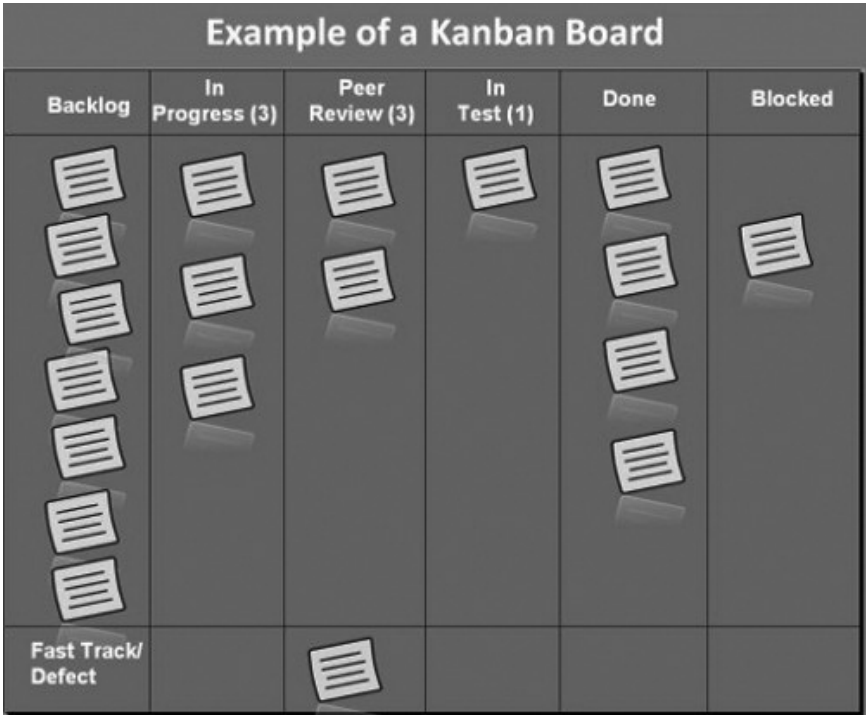


Figure 2: Example Kanban Board

Each of the groups shown in figure 1 tracks their own projects. Figure 3 shows the data analytics groups again, but each with their own Kanban boards to track the progress of work items. To serve the end goal of creating analytics for users, the data teams are desperately trying to move work items from the backlog (left column) to the done column at the right, and then pass it off to the next group in line.

Data professionals are smart and talented. They work hard. Why does it take so long to move work tickets to the right? Why does the system become overloaded with so many unfinished work items forcing the team to waste cycles context switching?

To address these questions, we need to think about the creation and deployment of analytics like a manufacturing process. The collective workflows of all of the data teams are a linked sequence of steps, not unlike what you would see in a manufacturing operation. When we conceptualize the development of new analytics in this way, it offers the possibility of applying manufacturing management tools that uncover and implement process improvements.



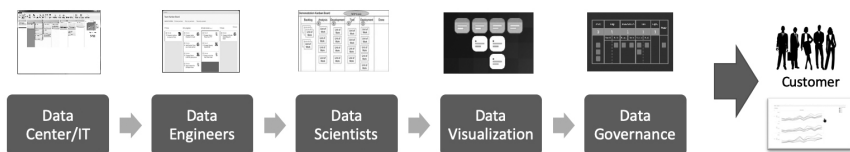


Figure 3: The development pipeline with Kanban boards

## THE THEORY OF CONSTRAINTS

One of the most influential methodologies for ongoing improvement in manufacturing operations is the [Theory of Constraints](#) (ToC), introduced by Dr. Eliyahu Goldratt in a business novel called “The Goal,” in 1984. The book chronicles the adventures of the fictional plant manager Alex Rogo who has 90 days to turn around his failing production facility. The plant can’t seem to ship anything on time, even after installing robots and investing in other improvements dictated by conventional wisdom. As the story progresses, our hero learns why none of his improvements have made any difference.

## THE BOTTLENECK

The plant’s complex manufacturing process, with its long sequence of interdependent stages, was throughput limited by one particular operation — a certain machine with limited capacity. This machine was the “constraint” or bottleneck. The Theory of Constraints views every process as a series of linked activities, one of which acts as a constraint on the overall throughput of the entire system. The constraint could be a human resource, a process, or a tool/technology.

In “The Goal,” Alex learned that “an improvement at any point in the system, not at the constraint, is an illusion.” An improvement made at a stage that feeds work to the bottleneck just increases the queue of work waiting for the bottleneck. Improvements after the bottleneck will always remain starved. Every loss of productivity at the bottleneck is a loss in the throughput of the entire system. Losses in productivity in any other step in the process don’t matter as long as that step still produces faster than the bottleneck.

Even though Alex’s robots improved efficiency at one stage of his manufacturing process, they didn’t alleviate the true system constraint. When Alex’s team focused improvement efforts on raising the throughput of the bottleneck, they were finally able to increase the throughput of the overall manufacturing process. True, some of their metrics looked worse (the robot station efficiency declined), but they were able to reduce cycle time, ship product on time and make a lot more money for the company. That is, after all, the real “goal” of a manufacturing facility.

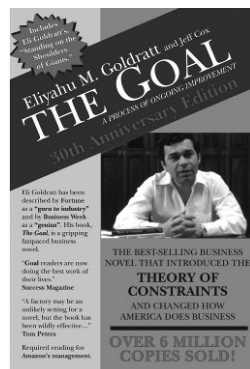


Figure 4

## FINDING YOUR BOTTLENECK

To improve the speed (and minimize the cycle time) of analytics development, you need to find and alleviate the bottleneck. This bottleneck is what is holding back your people from producing analytics at a peak level of performance. The bottleneck can often be identified using these simple indications:

- **Work in Progress (WIP)** — In a manufacturing flow, work-in-progress usually accumulates before a constraint. In data analytics, you may notice a growing list of requests for a scarce resource. For example, if it takes 40 weeks to [provision a development system](#), your list of requests for them is likely to be long.
- **Expedite** — Look for areas where you are regularly being asked to divert resources to ensure that critical analytics reach users. In data analytics, [data errors](#) are a common source of unplanned work.
- **Cycle Time** — Pay attention to the steps in your process with the longest cycle time. For example, some organizations take 6 months to shepherd 20 lines of SQL through the [impact review board](#). Naturally, if a step is starved or blocked by a dependency, the bottleneck is the external factor.
- **Demand** — Note steps in your pipeline or process that are simply not keeping up with demand. For example, often less time is required to create new analytics than to test and validate them in preparation for deployment.

## EXAMPLE BOTTLENECKS IN DATA ANALYTICS

You may notice a common theme in each of the example bottlenecks above. A bottleneck is especially problematic because it prevents people on the analytics team (analysts, scientists, engineers, ...) from fulfilling their primary function — creating new analytics. Bottlenecks distract them from high priority work. Bottlenecks redirect their energy to non-value add activities. Bottlenecks prevent them from implementing new ideas quickly.

When managers talk to data analysts, scientists and engineers, they can quickly discover the issues that slow them down. Figure 5 shows some common constraints. For example, data errors in analytics cause unplanned work that upsets a carefully crafted Kanban board. Work-in-progress (WIP) is placed on hold and key personnel context switch to address the high-severity outages. Data errors cause the Kanban boards to be flooded with new tasks which can overwhelm the system. Formerly high priority tasks are put on hold, and management is burdened, having to manage the complexity of many more work items. Data errors also affect the culture of the organization. After a series of interruptions from data errors, the team becomes accustomed to moving more slowly and cautiously. From a Theory of Constraints perspective, data errors severely impact the overall throughput of the data organization.

A related problem, also shown in figure 5, occurs when deployment of new analytics breaks something unexpectedly. Unsuccessful deployments can be another cause of unplanned work which can lead to excessive caution, and burdensome manual operations and testing.

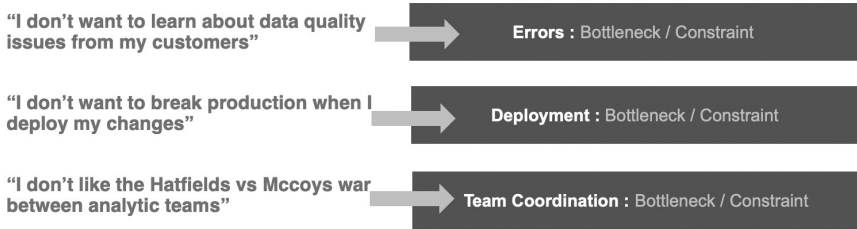


Figure 5: Translating problems to constraints

Another common constraint is [team coordination](#). The teams may all be furiously rowing the boat, but perhaps not in the same direction. In a large organization, each team's work is usually dependent on each other. The result can be a serialized pipeline. Tasks could be parallelized if the teams collaborated better. New analytics wouldn't break existing data operations with proper coordination between and among teams.

A wide variety of constraints potentially slow down analytics development cycle time. In development organizations, there are sometimes multiple constraints in effect. There is also variation in the way that constraints impact different projects. The following are some potential rate-limiting bottlenecks to rapidly deploying analytics:

- [Dependency on IT](#) to make schema changes or to integrate new data sets
- [Impact Review Board](#)
- Provisioning of development systems and [environments](#)
- [Long test cycles](#)
- Data [errors](#) causing unplanned work
- [Manual orchestration](#)
- Fear of breaking existing analytics
- Lack of [teamwork](#) among data engineers, scientists, analysts, and users
- [Long project cycles — deferred value](#)

When you have identified a bottleneck, the Theory of Constraints offers a methodology called the Process Of On-Going Improvement (POOGI) to address it. If you have many active bottlenecks that all need to be addressed, it may be more effective to focus on them one at a time. Below, we will suggest a method that we have found particularly effective in prioritizing projects.

## ALLEVIATING THE BOTTLENECK

Once identified, the Theory of Constraints recommends a five-step methodology to address the constraint:

1. **Identify the constraint**
2. **Exploit the constraint** — Make improvements to the throughput of the constraint using existing resources

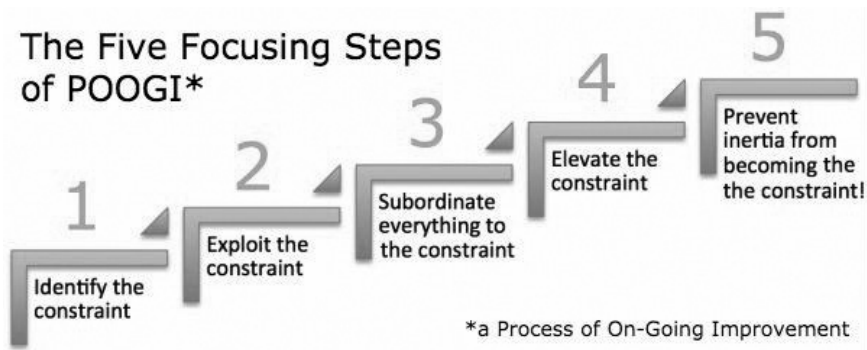


Figure 6: Source: Theory of Constraints Institute, Process of On-Going Improvement (POOGI)

**3. Subordinate everything to the constraint** — Review all activities and make sure that they benefit (or do not negatively impact) the constraint. Remember, any loss in productivity at the constraint is a loss in throughput for the entire system.

**4. Elevate the constraint** — If after steps 2–3, the constraint remains in the same place, consider what other steps, such as investing resources, will help alleviate this step as a bottleneck

**5. Prevent inertia from becoming a constraint by returning to step 1.**

## THE THEORY OF CONSTRAINTS APPLIED TO IT

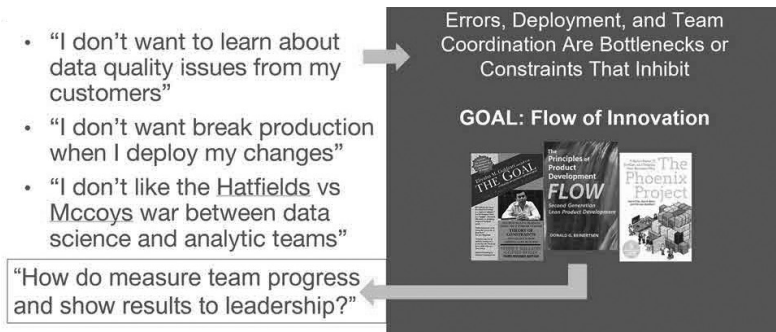


Figure 7: Errors, deployment and team coordination are bottlenecks that inhibit the flow of analytics innovation

A leading book on DevOps, called “The Phoenix Project,” was explained by author Gene Kim to be essentially an adaptation of “The Goal” to IT operations. To alleviate their bottleneck, the team in the book implements [Agile](#) development (small lot sizes) and [DevOps](#) (automation). One important bottleneck was a bright programmer named Brent who was needed for every system enhancement and was constantly being pulled into unplanned work. When the team got better at relieving and managing their constraints, the output of the whole department dramatically improved.

## PRIORITIZING DATAOPS PROJECTS BASED ON DESIRED OUTCOMES

If you have identified multiple bottlenecks in your development process, it may be difficult to decide which one to tackle first. [DataOps](#) is a methodology that applies [Agile](#), [DevOps](#) and [lean](#) manufacturing to data analytics. That's a lot of ground to cover. One way to approach this question is to think like a product or services company.

The data organization creates analytics for its consumers (users, colleagues, business units, managers, ...). Think of analytics as your product and data consumers as your customers. Like any product or service organization, perhaps you should simply ask your customers what they want?

The problem is that customers don't actually know what products or services they want. What customer would have asked for Velcro or Post-It notes or Twitter? Many data professionals can relate to the experience of working diligently to deliver what customers say they want only to receive a lukewarm response.

There is much debate about how to listen to the voice of the customer ([Dorothy Leonard, Harvard Business School](#), The Limitations of Listening). Customer preferences are reliable when you ask them to make selections within a familiar product category. If you venture outside of the customer's experience, you tend to encounter two blocks. People fixate on the way that products are normally used, preventing them from thinking outside the box. Second, customers have seemingly contradictory needs. Your data analytics customers want analytics to be error-free, which requires a lot of testing, but they dislike waiting for lengthy QA activities to complete. Data professionals might feel like they are in a no-win situation.

Management consultant [Anthony Ulwick](#) contends ([Harvard Business Review](#))

that you should not expect your customers to recommend solutions to their problems. They aren't expert enough for that. Instead, ask about desired *outcomes*. What do they want analytics to do for them? The customers might say that they want changes to analytics to be completed very fast so they can play with ideas. They won't tell you to implement automated orchestration or a data warehouse which can both contribute to that outcome.

The outcome-based methodology for gathering customer input breaks down into five steps.

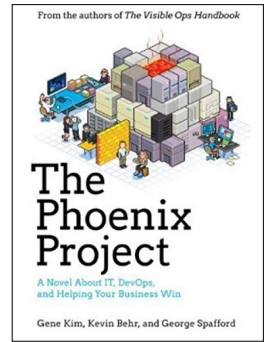


Figure 8



Figure 9: Many data professionals can relate to the experience of working diligently to deliver what customers say they want only to receive a lukewarm response.

### Step 1 — Plan outcome-based customers interviews

Deconstruct, step by step, the underlying processes behind your delivery of data analytics. It may make sense to interview users like data analysts who leverage data to create analytics for business colleagues.

### Step 2 — Conduct Interviews

Pay attention to desired outcomes not recommended solutions. Translate solutions to outcomes by asking what benefit the suggested feature/solution provides. Participants should consider every aspect of the process or activity they go through when creating or consuming analytics. A good way to phrase desired outcomes is in terms of the type (minimize, increase) and quantity (time, number, frequency) of improvement required. Experts in this method report that 75% of the customers' desired feedback is usually captured in the first two-hour session.

### Step 3 — Organize the Data

Collect a master list of outcomes, removing duplicates and categorize outcomes into groups that correspond to each step in the process

### Step 4 — Rate the outcomes

Conduct a quantitative survey to determine the importance of each desired outcome and the degree to which the outcome is satisfied by the current solution. Ask customers to rate, on a scale of 1–10, the importance of each desired outcome (Importance) and the degree to which it is currently satisfied (Satisfaction). These factors are input into the [opportunity algorithm](#) below which helps rate outcomes based on potential.

The opportunity algorithm makes use of a simple mathematical formula to estimate the potential opportunity associated with a particular outcome:

**Opportunity = Importance + (Importance - Satisfaction)**

*Note that if Satisfaction is greater than Importance, then the term (Importance - Satisfaction) is zero not negative.*

When you are done, you should have produced something like the below example.

### Step 5 — Guide Innovation

Table 1 reveals which outcomes are important to users and deprecates those outcomes that are already well served by the existing analytics development process. The outcomes which are both important and unsatisfied will rise to the top of the priority list. This data can be used as a guide to prioritize process improvements in the data analytics development pipeline and process.

Desired Outcome	Importance	Satisfaction	Opportunity
<b>Minimize data errors</b>	9.5	3.2	15.8
<b>Release new analytics (iterations) weekly</b>	8.3	4.2	12.4
<b>Provision development environments within 2 days</b>	9.5	7.5	11.5
<b>Test new analytics and approve deployment within 8 hours</b>	9.1	8.4	9.8
<b>Minimize time of impact review of new analytics</b>	5.1	1.0	9.2
<b>Minimize time to make schema changes</b>	7.7	6.6	8.8

**Table 1: Desired outcomes ranked by opportunity strength**

## THE PATH FORWARD FOR DATAOPS

DataOps applies manufacturing management methods to data analytics. One leading method, the Theory of Constraints, focuses on identifying and alleviating bottlenecks. Data analytics can apply this method to address the constraints that prevent the data analytics organization from achieving its peak levels of productivity. Bottlenecks lengthen the cycle time of developing new analytics and prevent the team from responding quickly to requests for new analytics. If these bottlenecks can be improved or eliminated, the team can move faster, developing and deploying with a high level of quality in record time.

If you have multiple bottlenecks, you can't address them all at once. The opportunity algorithm enables the data organization to prioritize process improvements that produce outcomes that are recognized as valued by users. It avoids the requirement for users to understand the technology, tools, and processes behind the data analytics pipeline. For DataOps proponents, it can provide a clear path forward for analytics projects that are both important and appreciated by users.





# Pav Bhaji

*Contributed By Anuja Waikar*

Pav Bhaji is a famous street food enjoyed in Mumbai

This recipe serves 4

## INGREDIENTS

2 large potatoes, diced  
3/4 cup peas  
1 cup cauliflower florets  
1/2 cup green bell pepper, finely chopped  
1/2 cup carrots, diced  
1 medium onion, finely chopped  
1 green chili, finely chopped (optional)  
2 medium tomatoes, finely chopped  
2 teaspoon ginger-garlic paste (if you don't have paste – finely grate ginger and garlic instead)  
1 tablespoon oil  
2 tablespoons butter (the more the better!)  
Pinch of turmeric (optional)  
1 teaspoon red chili powder (vary as per spice level you want)  
3 tablespoon pav bhaji masala powder (available in any Indian store. Best brand I used: Everest )  
Salt  
Dinner rolls and butter to toast it

## GARNISH

Sprinkle little lemon  
Chopped onions  
Coriander leaves  
Butter

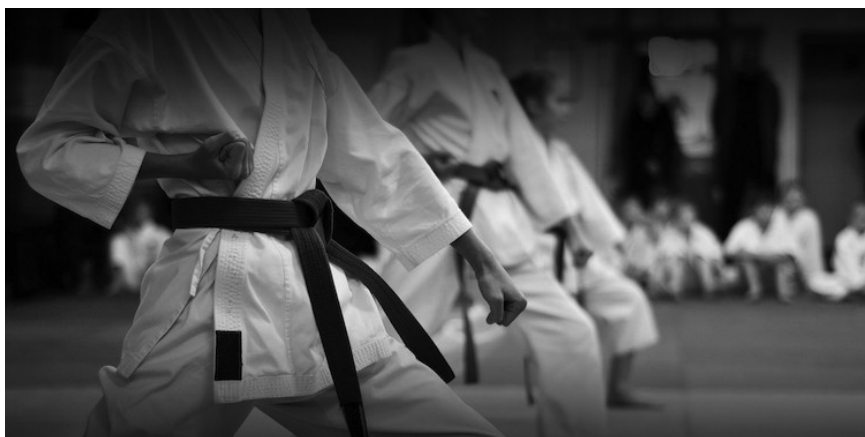




**Expand**



# Do You Need a DataOps Dojo?



As DataOps activity takes root within an enterprise, managers face the question of whether to build centralized or decentralized DataOps capabilities. Centralizing analytics brings it under control but granting analysts free reign is necessary to foster innovation and stay competitive. The beauty of DataOps is that you don't have to [choose between centralization and freedom](#). You can choose to do one or the other — or both. Below we'll discuss some standard DataOps technical services that could be developed and supported by a centralized team. We'll also discuss building DataOps expertise around the data organization, in a decentralized fashion, using DataOps centers of excellence (COE) or DataOps Dojos.

## DATAOPS TECHNICAL SERVICES

A centralized team can promote DataOps adoption by building a common technical infrastructure and tools to be leveraged by other groups. Centralizing analytics helps the organization standardize enterprise-wide measurements and metrics.

For example, some teams may recognize services revenue in the quarter booked, and others may amortize the revenue over the contract period. With a standard metric supported by a centralized technical team, the organization maintains consistency in analytics.

A centralized team can publish a set of software services that support the rollout of Agile/DataOps. The DataOps Technical Services (DTS) group provides a set of central services leveraged by other groups. DTS services bring the benefits of DataOps to groups that aren't ready to implement DataOps themselves. Examples of technologies that can be delivered 'as a service' include:

- Source code control repository
- Agile ticketing/Kanban tools
- Deploy to production
- Product monitoring
- Develop/execute regression testing
- Development sandboxes
- Collaboration and training portals/wikis
- Test data management and other functions provided 'as a service'

The DTS group can also act as a services organization, offering services to other teams. Below are some examples of services that a DTS group can provide:

- Reusable deployment services that integrate, deliver, and deploy end-to-end analytic pipelines to production
- Central code repository where all data engineering/science/analytics work can be tracked, reviewed and shared
- Central DataOps process measurement function with reports
- 'Mission Control' for data production metrics and data team development metrics to demonstrate progress on the DataOps transformation

DTS creates robust DataOps services and capabilities, but if an organization wishes to seed DataOps practices throughout the organization, it should plan methods to transfer DataOps solutions and "know-how" to data scientists and engineers in the periphery of the organization.

## **DATAOPS CENTER OF EXCELLENCE**

The center of excellence (COE) model leverages the DataOps team to solve real-world challenges. The goal of a COE is to take a large, widespread, deep-rooted organizational problem and solve it in a smaller scope, proof-of-concept project, using an open-minded approach. The COE then attempts to leverage small wins across the larger organization at scale. A COE typically has a full-time staff that focuses on delivering value for customers in an experimentation-driven, iterative, result-oriented, customer-focused way. COE teams try to show what "good" looks like by establishing common technical standards and best practices. They also can provide education and training enterprise-wide. The COE approach is used in many enterprises, but the DevOps industry has more often standardized on Dojos as a best practice.

## **DATAOPS DOJO**

Demand for skilled DataOps engineers is skyrocketing, and like DevOps engineers, they are hard to find and harder to hire. Enterprises moving towards DataOps transformation may find it worthwhile to build DataOps expertise organically in each team within the data organization.

A DataOps Dojo is a place where DataOps beginners go for a short period of intense, hands-on training. In Japan, a dojo is a safe environment where someone can practice new skills, such as martial arts. Companies like Delta Airlines and John Deere employ the Dojo concept effectively to build lean, Agile, and DevOps muscles. The Dojo offers a separate workspace where teams learn new skills while working on actual projects that deliver customer value.

Dojos provide an environment where teams gain practical experience without worrying about introducing errors into the production environment. The staff rotates in for weeks or months at a time to learn new skills by working on real-world projects. They then bring those skills ideas back to their original teams.

## **DATAOPS TRANSFORMATION**

Each of the approaches described above can deliver DataOps benefits to the enterprise. Nevertheless, it can be challenging to grow DataOps expertise in-house without the benefit of mentorship. DataKitchen offers DataOps Transformation Advisory Services that address DataOps methodologies, strategy, tools automation, and cultural change.





# Spinach Madeline

*Contributed By Jessica Dias de Oliveira*

## INGREDIENTS

2 packages frozen chopped spinach

4 tablespoons butter

2 tablespoons all-purpose flour

2 tablespoons chopped onions

1/2 cup evaporated milk

1/2 cup spinach/vegetable liquid

1/2 teaspoon black pepper

3/4 teaspoon celery salt

3/4 teaspoon garlic salt

Salt to taste

6-ounce roll of jalapeno cheese (or substitute Velveeta with 2 minced jalapenos), cut into small pieces

1 teaspoon Worcestershire sauce

Cayenne to taste

Buttered bread crumbs (optional)

## INSTRUCTIONS

1. Cook the spinach according to package directions. Drain and reserve the liquid from the pot for the butter-flour roux in the next step.
2. Melt the butter in a saucepan over low heat. Add the flour, stirring constantly until blended and smooth, but not brown. Add the onions and cook until soft but not brown. Add the milk and one-half cup of the reserved liquid from the spinach pot. Stir constantly to avoid any lumps. Cook, stirring, until smooth and thick. Add the seasonings and cheese and stir until the cheese is completely melted.
3. Pour into a casserole dish and top with buttered bread crumbs (optional).
4. Bake in a preheated 350-degree oven until bubbly, about 30 minutes.
5. Serve warm as a dip or side. Makes about 8 servings.

*Attribution: Spinach Madeline is from River Road Recipes, first published in 1959 by the Junior League of Baton Rouge. From nola.com.*



# DataOps Engineer Will Be the Sexiest Job in Analytics



Years ago, prior to the advent of [Agile development](#), a friend of mine worked as a release engineer. His job was to ensure a seamless build and release process for the software development team. He designed and developed builds, scripts, installation procedures and managed the [version control](#) and issue tracking systems. He played a mean mandolin at company parties too.

The role of release engineer was (and still is) critical to completing a successful software release and deployment, but as these things go, my friend was valued less than the software developers who worked beside him. The thinking went something like this — developers could make or break schedules and that directly contributed to the bottom line. Release engineers, on the other hand, were never noticed, unless something went wrong. As you might guess, in those days the job of release engineer was compensated less generously than development engineer. Often, the best people vied for positions in development where compensation was better.

## RISING FORTUNES

Today, the fortunes of release engineers have risen sharply. In companies that are implementing DevOps there is no more important person than the release engineer. The job title has been renamed DevOps engineer and it is one of the most highly compensated positions in the field of software engineering. According to salary surveys, experienced DevOps engineers make six-figure salaries. DevOps specialists are so hard to find that firms are hiring people without college degrees, if they have the right experience.

Whereas a release engineer used to work off in a corner tying up loose ends, the DevOps engineer is a high-visibility role coordinating the development, test, IT and operations functions. If a DevOps engineer is successful, the wall between development and operations melts away and the dev team becomes more agile, efficient and responsive to the market. This has a huge impact on the organization's culture and ability to innovate. With so much at stake, it makes sense to get the best person possible to fulfill the DevOps engineer role, and compensate them accordingly. When DevOps came along, the release engineer went from fulfilling a secondary supporting role to occupying the most sought after position in the department. Many release engineers have successfully rebranded themselves as DevOps engineers and significantly upgraded their careers.

## DATAOPS FOR DATA ANALYTICS

A similar change, called [DataOps](#), is transforming the roles on the data analytics team. DataOps is a better way to develop and deliver analytics. It applies [Agile development](#), [DevOps](#) and [lean manufacturing](#) principles to data analytics producing a transformation in data-driven decision making.

Data engineers, data analysts, data scientists — these are all important roles, but they will be valued even more under DataOps. Too often, data analytics professionals are trapped into relying upon non-scalable methods: [heroism](#), [hope or caution](#). DataOps offers a way out of this no-win situation.

The capabilities unlocked by DataOps impacts everyone that uses data analytics — all the way to the top levels of the organization. DataOps breaks down the barriers between data analytics and operations. It makes data more easily accessible to users by redesigning the data analytics pipeline to be more flexible and responsive. It will completely change what people think of as possible in data analytics.

In many organizations, the DataOps engineer will be a separate role. In others, it will be a shared function. In any case, the opportunity to have a high-visibility impact on the organization will make DataOps engineering one of the most desirable and highly compensated functions. Like the release engineer whose career was transformed by DevOps, DataOps will boost the fortunes of data analytics professionals. DataOps will offer select members of the analytics team a chance to reposition their roles in a way that significantly advances their career. If you are looking for an opportunity for growth as a DBA, ETL Engineer, BI Analyst, or another role look into DataOps as the next step.

And watch out Data Scientist, the real [sexiest job of the 21st century](#) is DataOps Engineer.



# Kerala Style Chicken Stew

*Contributed By Shruthy Vakkil*

## INGREDIENTS

3 Table Spoon Coconut oil. {{Variation override - You can use vegetable oil too}}

2 Green cardamom

3 Clove

1-Inch Cinnamon

3-4 Black peppercorn

1 Bay leaf

2 Cups Onion # preferably julienne cut

1 Teaspoon Ginger # Finely chopped

1 Teaspoon Garlic # Finely chopped

2 green chili

10-12 Curry leaves

1.5 LB Chicken

.25 Cup Potato # Small cubes

.25 Cup Carrot # Small cubes

2 cup coconut milk

Salt to taste

## INSTRUCTIONS

1. In a pan, heat oil. Once the oil is hot, add cardamom, cloves, cinnamon, peppercorn, and bay leaf.
2. Saute for a few seconds. (Don't let it burn)
3. Now add onion and saute till they turn translucent.
4. Add ginger and garlic and fry until the raw smell is gone.
5. Add green chilies and curry leaves and fry for a minute.
6. Now add chicken and cook for 2 mins
7. Add 1 cup coconut milk and a little salt.
8. Cover and cook for 10-15 minutes.
9. Add potato and carrot and cook until chicken and vegetables are done.
10. Add the remaining 1 cup coconut milk and cook for another 5 minutes.
11. Pour a little (1 teaspoon) coconut oil on top





# Improving Teamwork in Data Analytics with DataOps



When enterprises invite us in to talk to them about DataOps, we generally encounter dedicated and competent people struggling with conflicting goals/priorities, weak process design, insufficient resources, clashing mindsets, and differing views of reality. Inadequate workflow processes prevent them from doing their best work. The team lacks the structural and contextual support necessary to enable successful teamwork.

Imagine that a Vice President of Marketing makes an urgent request to the data analytics team: “I need new data on profitability ASAP.” At many organizations the process for creating and deploying these new analytics would go something like this:

1. The new requirement falls outside the scope of the development “plan of record” for the analytics team. Changing the plan requires departmental meetings and the approval of a new budget and schedule. Meetings ensue.
2. Padma, a Data Engineer, requests access to new data. The request goes on the IT backlog. IT grants access after several weeks.

3. Padma writes a functional specification and submits the proposed change to the [Impact Review Board](#) (IRB), which meets monthly. A key-person is on vacation, so the proposed feature waits another month.
4. Padma begins implementation. The change that she is making is similar to another recently developed report. Not knowing that, she writes the new analytics from scratch. The test environment does not match “production.” so her testing misses some corner cases.
5. Testing on the target environment begins. High-severity errors pull Eric, a Production Engineer, into an “all-hands-on-deck” situation, putting testing temporarily on hold.
6. Once the fires are extinguished, Eric returns to testing on the target and uncovers some issues in the analytics. Eric feeds error reports back to Padma. She can’t easily reproduce the issues because the code doesn’t fail in the “dev” environment. She spends significant effort replicating the errors so she can address them. The cycle is repeated a few times until the analytics are debugged.
7. Analytics are finally ready for deployment. Production schedules the update. The next deployment window available is in three weeks.
8. After several months have elapsed (total cycle time), the VP of Marketing receives the new analytics, wondering why it took so long. This information could have boosted sales for the current quarter if it had been delivered when she had initially asked.

Every organization faces unique challenges, but the issues above are ubiquitous. The situation we described is not meeting anyone’s needs. Data engineers went to school to learn how to create analytic insights. They didn’t expect that it would take six months to deploy twenty lines of SQL. The process is a complete hassle for IT. They have to worry about governance and access control and their backlog is entirely unmanageable. Users are frustrated because they wait far too long for new analytics. We could go on and on. No one here is enjoying themselves.

The frustration sometimes expresses itself as conflict and stress. From the outside, it looks like a teamwork problem. No one gets along. People are rowing the boat in different directions. If managers want to blame someone, they will point at the team leader.

At this point, a manager might try beer, donuts and trust exercises (hopefully not in that order) to solve the “teamwork issues” in the group. Another common mistake is to coach the group to work more [slowly and carefully](#). This thinking stems from the fallacy that you have to choose between quality and cycle time. In reality, you can have both.

We recommend a process-oriented solution that addresses everyone’s goals and priorities, coordinates tasks, provisions resources, and creates a shared reality. DataOps can turn a band of squabbling data professionals into a high-performance team.

## DATAOPS IMPROVES TEAMWORK

DataOps shortens the [cycle time](#) and improves the quality of data analytics. Data teams that do not use DataOps may try to reduce the number of errors by being more [cautious and careful](#). In other words, slowing down. DataOps helps organizations improve data quality while going faster. This might seem impossible until you learn more about how DataOps approaches analytics development and deployment.

DataOps is a set of methodologies supported by tools and automation. To say it in one breath; think [Agile](#) development, [DevOps](#) and [Lean](#) manufacturing (i.e., statistical process controls) applied to data analytics. DataOps comprehends that enterprises live in a multi-language, multi-tool, heterogeneous environment with complex workflows. To implement DataOps, extend your existing environment to align with DataOps principles. You can implement DataOps by yourself in [seven steps](#), or you can adopt a DataOps Platform. Here, we'll describe how a DataOps Platform works and illustrate it with an example of a real-life analytics development project.

### DATAOPS JOB #1:

#### ABSTRACTING, SEPARATING, AND ALIGNING RELEASE ENVIRONMENTS

Enterprises that collocate development and production on the same system face a number of issues. Analytics developers sometimes make changes that create side effects or break analytics. Development can also be processor-intensive, impacting production performance and query response time.

DataOps provides production and development with dedicated system [environments](#). Some enterprises take this step but fail to align these environments. Development uses cloud platforms while production uses on-prem. Development uses clean data while production uses real-world data. The list of opportunities for misalignment are endless. DataOps requires that system environments be aligned. In other words, as close as possible to identical. The more similar, the easier it will be to migrate code and replicate errors. Some divergence is necessary. For example, data given to developers may have to be sampled or masked for practical or governance reasons.

Figure 1 below shows a simplified production environment. The system transfers files securely using SFTP. It stores files in S3 and utilizes a Redshift cluster. It also uses Docker containers and runs some Python. Production alerts are forwarded to a Slack channel in real-time. Note that we chose an example based on Amazon Web Services, but we could have selected any other tools. Our example applies whether the technology is Azure, GCP, on-prem or anything else.

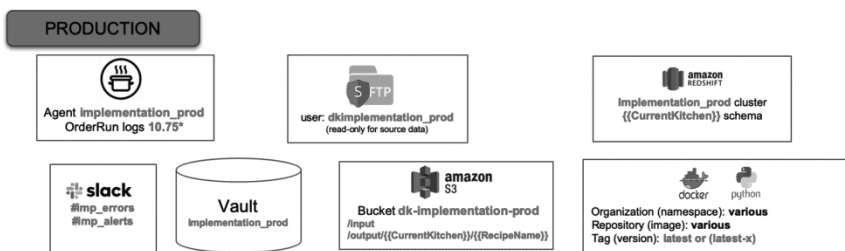


Figure 1: Simplified production technical environment

DataOps segments production and development into separate release [environments](#) — see Figure 2. In our parlance, a release environment includes a set of hardware resources, a software toolchain, data, and a security Vault which stores encrypted, sensitive access control information like usernames and passwords for tools. Our production engineer, Eric, manages the production release environment. Production has dedicated hardware and software resources so Eric can control performance, quality, governance and manage change. The production release environment is secure — the developers do not have access to it.

The development team receives its own separate but equivalent release environment, managed by the third important member of our team; Chris, a [DataOps Engineer](#). Chris also implements the infrastructure that abstracts the release environments so that analytics move easily between dev and production. We'll describe this further down below. Any existing team member, with DataOps skills, can perform the DataOps engineering function, but in our simplified case study, adding a person will better illustrate how the roles fit together.

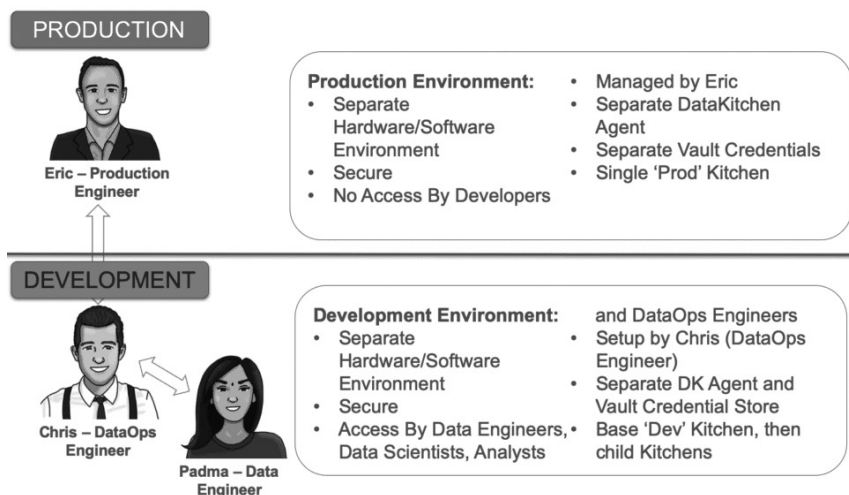


Figure 2: Production and development maintain separate but equivalent environments. The production engineer manages the production release environment and the DataOps engineer manages the development release environment.

Chris creates a development release environment that matches the production release environment. This alignment reduces issues when migrating analytics from development to production. Per Figure 2, the development environment has an associated security Vault, just like the production environment. When a developer logs into a development workspace, the security Vault provides credentials for the tools in the development release environment. When the code seamlessly moves to production, the production Vault supplies credentials for the production release environment. Figure 3 below illustrates the separate but equivalent production and development release environments. If you aren't familiar with "environments," think of these as discrete software and hardware systems with equivalent configuration, tools, and data.

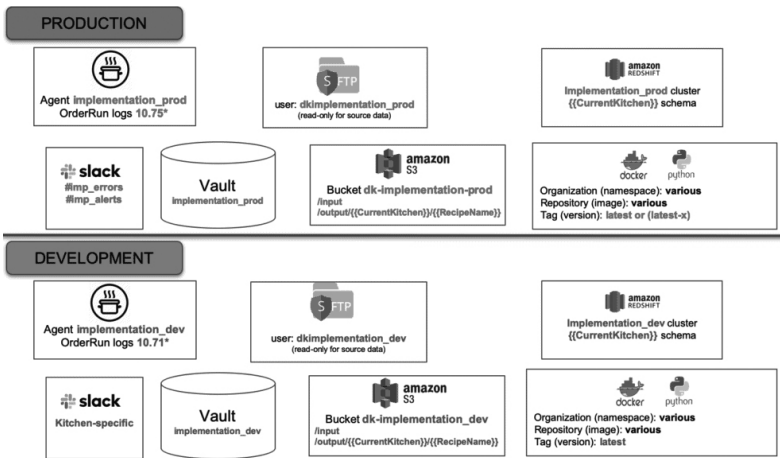


Figure 3: DataOps segments the production and development workspaces into separate but equivalent release environments.

Chris uses DataOps to create and implement the processes that enable successful teamwork. This activity puts him right at the nexus between data analytics development and operations. Chris is one of the most **important and respected** members of the data team. He creates the mechanisms that enable work to flow seamlessly from development to production. Chris makes sure that environments are aligned and that everyone has the hardware, software, data, network and other resources that they need. He also makes available software components, created by team members, to promote **reuse** — a considerable multiplier of productivity. In our simple example, Chris manages the tasks that comprise the pre-release process. Padma appreciates having Chris on the team because now she has everything that she needs to create analytics efficiently on a self-service basis. Eric is happy because DataOps has streamlined deployment, and expanded testing has raised both data and analytics quality. Additionally, there is much greater visibility into the artifacts and logs related to analytics, whether in development, pre-release or in production. It's clear that Chris is a key player in implementing DataOps. Let's dive deeper into how it really works.

## A DATAOPS “KITCHEN”:

### A RELEASE ENVIRONMENT, WORKSPACE, AND PIPELINE BRANCH

Our development team in Figure 2 consists of Chris and Padma. In a real-world enterprise, there could be dozens or hundreds of developers. DataOps helps everyone work as a team by minimizing the amount of rekeying required so that analytics move seamlessly from developer to developer and into production. DataOps also organizes activities so that tasks remain coordinated and team members stay aligned. The foundation of these synchronized activities is a virtual workspace called a “Kitchen.”

A Kitchen is a development workspace with everything that an analytics developer requires. It contains hardware, software, tools, code (with [version control](#)) and data. A Kitchen [points](#) to a release environment which gives it access to all of the resources associated with that environment. A Kitchen also enforces workflow and coordinates tasks.

The processing pipelines for analytics consist of a series of steps that operate on data and produce a result. We use the term “Pipeline” to encompass all of these tasks. A DataOps Pipeline encapsulates all the complexity of these sequences, performs the orchestration work, and tests the results. The idea is that any analytic tool that is invocable under software control can be [orchestrated](#) by a DataOps Pipeline. Kitchens enable team members to access, modify and execute workflow Pipelines. A simple Pipeline is shown in Figure 4.

Pipelines, and the components that comprise them, are made visible within a Kitchen. This encourages the reuse of previously developed analytics or services. Code reuse can be a significant factor in reducing cycle time.

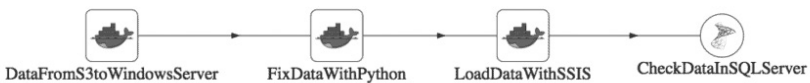
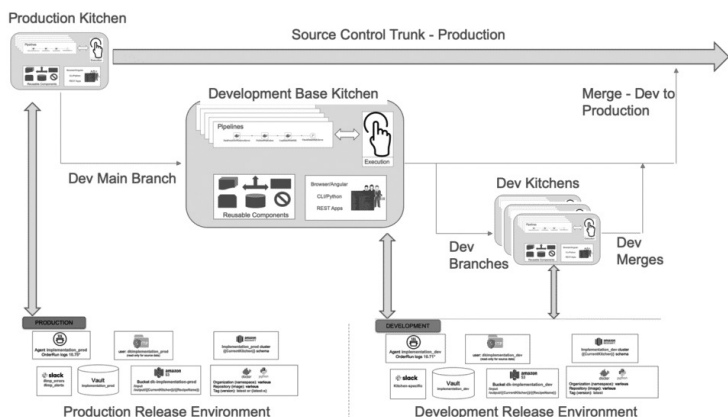


Figure 4: A simple DataOps pipeline is represented by a directed acyclic graph (DAG). Each node in the graph is a sequence of orchestrated operations.

Kitchens also tightly couple to [version control](#). When the development team wants to start work on a new feature, they instantiate a new child Kitchen which creates a corresponding Git [branch](#). When the feature is complete, the Kitchen is merged back into its parent Kitchen, initiating a Git merge. The Kitchen hierarchy aligns with the source control branch tree. Figure 5 shows how Kitchen creation/deletion corresponds to a version control branch and [merge](#).



**Figure 5: Kitchens point to a release environment. They represent source control branches and merges, and also serve as development, test, and release workspaces.**

Kitchens may be persistent or temporary; they may be private or shared, depending on the needs of a project. Access to a Kitchen is limited to a designated set of users or “Kitchen staff.” The Vault in a release environment supplies a Kitchen with the set of usernames and passwords needed to access the environment toolchain.

DataOps empowers an enterprise to provide people access to data, eliminating gatekeepers. As mentioned above, developers access test data from within a Kitchen. In another example, a Pipeline could extract data from a data lake and create a data mart or flat file that serves Alteryx, Tableau and Excel users in the business units. DataOps promotes and enables data democratization, providing everyone access to the data relevant to their job. When “self-service” replaces “gatekeepers,” more work gets done in parallel and analytics development cycle-time accelerates significantly.

Figure 6 shows a Kitchen hierarchy. The base Kitchen is “demo\_production,” which points to the production release environment described earlier. This Kitchen is Eric’s workspace, and it enables him to coordinate his interactions with the development team. There is only one Kitchen corresponding to Eric’s production release environment. No iterative work takes place in production. Instead, think of “demo\_production” as a manufacturing flow where assembly lines run on a tight schedule.

Chris’ workspace is a Kitchen called “demo\_dev.” The “demo\_dev” Kitchen is the baseline development workspace, and it **points** to the development release environment introduced above, at the bottom of Figure 2. In our example, Chris’ Kitchen serves as a pre-release staging area where merges from numerous child development Kitchens consolidate and integrate before being deployed to production. With release environments aligned, Kitchens don’t have to do anything different or special for merges across release environments versus merges within a release environment.

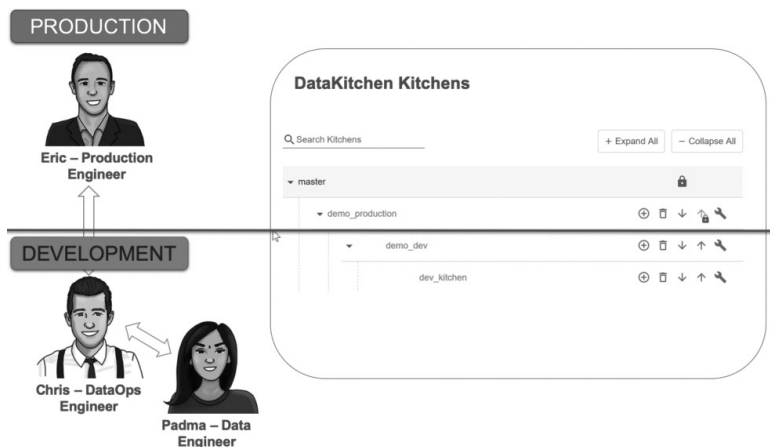


Figure 6: Eric, Chris, and Padma each have personal Kitchens, organized in a hierarchy that aligns with their workflow.

Every developer needs a workspace so they may work productively without impacting or being impacted by others. A Kitchen can be persistent, like a personal workspace, or temporary, tied to a specific project. Once Kitchen creation is set-up, team members create workspaces as needed. This “self-service” aspect of DataOps eliminates the time that developers used to wait for systems, data, or approvals. DataOps empowers developers to hit the ground running. In Figure 6, Padma has created the Kitchen “dev\_kitchen.” Padma’s Kitchen can leverage Pipelines and other services created by the dev team.

## DATAOPS SEGREGATES USER ACTIVITY

With multiple developers sharing a release environment, the DataOps Platform segregates developer activity. For example, all of the developer Kitchens share the Redshift cluster shown in Figure 2. Note the notation “`{{CurrentKitchen}}`” associated with Redshift in Figure 2. Each developer has a Redshift schema within the cluster identified by their Kitchen name. For example, an access by Padma would target a schema identified by her unique Kitchen name “dev\_kitchen.” The DataOps Platform uses Kitchen names and other identifiers to segregate user activity within a shared release environment. Segregation helps keep everyone’s work isolated while sharing development resources.

## A DATAOPS PROCESS

Now let’s look at how to use a DataOps Platform to develop and deliver analytics with minimal cycle time and unsurpassed quality. We’ll walk through an example of how DataOps helps team members work together to deploy analytics into production.

Think back to the earlier request by the VP of Marketing for “new analytics.” DataOps coordinates this multi-step, multi-person and multi-environment workflow and manages it from inception to deployment.



## Step 1 — Starting From a Ticket

The Agile Sprint meeting commits to the new feature for the VP of Marketing in the upcoming iteration. The project manager creates a JIRA ticket.

## Step 2 — Creation of the Development Kitchen

In a few minutes, Padma creates a development Kitchen for herself and gets to work. Chris has automated the creation of Kitchens to provide developers with the test data, resources, and Git branch that they need. Padma's Kitchen is called "dev\_Kitchen" (see Figure 6). If Padma takes a technical risk that doesn't work out, she can abandon this Kitchen and start over with a new one. That effectively deletes the first Git branch and starts again with a new one.

## Step 3 — Implementation

Padma's Kitchen provides her with pipelines that serve as a significant head start on the new profitability analytics. Padma receives the test data (de-identified) she needs as part of Kitchen creation and configures toolchain access (SFTP, S3, Redshift, ...) for her Kitchen. Padma implements the new analytics by modifying an existing Pipeline. She adds additional [tests](#) to the existing suite, checking that incoming data is clean and valid. She writes tests for each stage of ETL/processing to ensure that the analytics are working from end to end. The tests verify her work and will also run as part of the production flow. Her new pipelines include orchestration of the data and analytics as well as all tests. The tests direct messages and alerts to her Kitchen-specific Slack channel. With the extensive testing, Padma knows that her work will migrate seamlessly into production with minimal effort on Eric's part. Now that release environments have been aligned, she's confident that her analytics work in the target environment.

Before she hands off her code for pre-production staging, Padma first has to [merge](#) down from "demo\_dev" Kitchen so that she can integrate any relevant changes her coworkers have made since her [branch](#). She reruns all her tests to ensure a clean merge. If there is a conflict in the code merge, the DataOps Platform will pop-up a three panel UI to enable further investigation and resolution. When Padma is ready, she updates and reassigns the JIRA ticket. If the data team were larger, the new analytics could be handed off from person to person, in a line, with each person adding their piece or performing their step in the process.

## Step 4 — Pre-Release

In our simple example, Chris serves as the pre-release engineer. With a few clicks, Chris merges Padma's Kitchen "dev\_Kitchen" back into the main development Kitchen "demo\_dev," initiating a Git merge. After the merge, the Pipelines that Padma updated are visible in Chris' Kitchen. If Chris is hands-on, he can review Padma's work, check artifacts, rerun her [tests](#), or even add a few tests of his own, providing one last step of QA or governance. Chris creates a schedule that, once enabled, will automatically run the new Pipeline every Monday at 6 am. When Chris is satisfied, he updates and reassigns the JIRA ticket, letting Eric know that the feature is ready for deployment.

## Step 5 — Production Deployment

Eric easily merges the main development Kitchen “demo\_dev” into the production Kitchen, “demo\_production,” corresponding to a Git merge. Eric can now see the new Pipelines that Padma created. He inspects the test logs and reruns the new analytics and tests to be 100% sure. The release environments match so the new Pipelines work perfectly. He’s also happy to see tests verifying the input data using DataOps statistical process control. [Tests](#) will detect erroneous data, before it enters the production pipeline. When he’s ready, Eric enables the schedule that Chris created, integrating the new analytics into the operations pipeline. DataOps redirects any Slack messages generated by the new analytics to the production Slack channels.

## Step 6 — Customer Sees Results

The VP of Marketing sees the new customer segmentation and she’s delighted. She then has an epiphany. If she could see this new data combined with a report that Padma delivered last week, it could open up a whole new approach to marketing — something that she is sure the competitors haven’t discovered. She calls the analytics team and...back to Step 1.

## DATAOPS BENEFITS

As our short example demonstrated, the DataOps Teamwork Process delivers these benefits:

- **Ease movement between team members with many tools and environments** — Kitchens align the production and development environment(s) and abstract the machine, tools, security and networking resources underlying analytics. Analytics easily migrate from one team member to another or from dev to production. Kitchens also bind changes to source control.
- **Collaborate and coordinate work** — DataOps provides teams with the compelling direction, strong structure, supportive context and shared mindset that are necessary for effective teamwork.
- **Automate work and reduce errors** — Automated orchestration reduces process variability and errors resulting from manual steps. Input, output and business logic tests at each stage of the workflow ensure that analytics are working correctly, and that data is within statistical limits. DataOps runs tests both in development and production, continuously monitoring quality. Warnings and errors are forwarded to the right person/channel for follow up.
- **Maintain security** — Kitchens are secured with access control. Kitchens then access a release environment toolchain using a security Vault which stores unique usernames/passwords.
- **Leverage best practices and re-use** — Kitchens include Pipelines and other reusable components which data engineers can leverage when developing new features.
- **Self-service** — Data professionals can move forward without waiting for resources or committee approval.

- **Data democratization** — Data can be made available to more people, even users outside the data team, who bring contextual knowledge and domain expertise to data analytics initiatives. “Self-service” replaces “gatekeepers” and everyone can have access to the data that they need.
- **Transparency** — Pipeline status and statistics are available in messages, reports and dashboards.

## SMOOTH TEAMWORK WITH DATAOPS

DataOps addressed several technical and process-oriented bottlenecks that previously delayed the creation of new analytics for months. Their processes can improve further, but they are now an order of magnitude faster and more reliable. At the next staff meeting, the mood of the team is considerably improved:

*Manager:* Good morning, everyone. I’m pleased to report that the VP of Marketing called the CDO thanking him for a great job on the analytics last week.

*Padma* (Data Engineer): Fortunately, I was able to leverage a Pipeline developed a few months ago by the MDM team. We were even able to reuse most of their tests.

*Chris* (DataOps Engineer): Once I set-up Kitchen creation, Padma was able to start being productive immediately. With matching release environments, we quickly migrated the new analytics from dev to production.

*Eric* (Production Engineer): The tests are showing that all data remains within statistical limits. The dashboard indicators are all green.

DataOps helps our band of frustrated and squabbling data professionals achieve a much higher level of overall team productivity by establishing processes and providing resources that support teamwork. With DataOps, two key performance parameters improve dramatically — the development cycle time of new analytics and quality of data and analytics code. We’ve seen it happen time and time again.

What’s even more exciting is the business impact of DataOps. When users request new analytics and receive them in a timely fashion, it initiates new ideas and uncharted areas of exploration. This tight feedback loop can help analytics achieve its true aim, stimulating creative solutions to an enterprise’s greatest challenges. Now that’s teamwork!







# Governance as Code



Data teams using inefficient, manual processes often find themselves working frantically to keep up with the endless stream of analytics updates and the exponential growth of data. If the organization also expects busy data scientists and analysts to implement data governance, the work may be treated as an afterthought, if not forgotten altogether. Enterprises using manual procedures need to carefully rethink their approach to governance.

With DataOps automation, governance can execute continuously as part of development, deployment, operations and monitoring workflows. Governance automation is called DataGovOps, and it is a part of the DataOps movement. Instead of starting with a typical wordy definition of data governance, let's look at some examples of the problems that governance attempts to solve:

1. The VP calls a quarterly meeting with the global sales force to review the forecast for each territory. Some salespeople display only direct product sales – others commingle products, services and non-recurring engineering. Some team members include verbal commitments, whereas others report only bookings. Without a single definition of “sales,” it's hard to obtain an accurate picture of what's happening.
2. Data resides in different locations and under the control of different groups within the enterprise. It's hard to track and manage the organization's data assets. It's difficult to even know where to look.
3. Some users export sensitive customer data to their laptop in order to work remotely using self-service tools. Some of this regulated data falls under [GDPR](#), [GLBA](#) or California's [CCPA](#).
4. The journey from raw data to finished charts and graphs spans groups, data centers and organizations. The data pipeline follows a complex execution path with numerous tools and platforms involved. When there is an issue to fix, who owns each part of the data analytics pipeline?
5. Data is notoriously incomplete and full of errors. How can/should it be cleaned? Is it fit for a given use? How is data quality assured?

Often data governance initiatives attempt to address these issues with meetings, checklists, sign-offs and nagging. This type of governance is a tax upon data analyst productivity. DataGovOps offers a new approach to governance by building automated governance into development and operations using DataOps tools and methods. “Governance-as-code” actively incorporates governance into data team workflows. With DataGovOps automation, governance is no longer a forgotten afterthought that is deferred until other more important work is complete.

DATA GOVERNANCE

In her book, “[Disrupting Data Governance: A Call to Action](#),” data governance expert Laura Madsen envisions a more agile model for data governance by redirecting the focus of governance towards value creation through promoting the usage of data (figure 1). Instead of focusing on how to limit users, governance should be concerned with promoting the safe and controlled use of data at scale. Data governance is then more about active enablement than rule enforcement. In other words, can we design data quality, management and protection workflows in such a way that they empower, not limit, data usage? This can be done if we take a DataOps approach to governance.

	Percentage of Importance
Increase Usage of Data	40%
Data Quality	25%
Data Management (i.e. lineage)	25%
Data Protection	10%

Figure 1: Data governance should emphasize quality, management, protection and most importantly, increasing usage. Source: Laura Madsen

DATAOPS AND GOVERNANCE

In the past couple of years, there has been a [tremendous proliferation of acronyms with the “Ops” suffix](#). This was started in the software space by DevOps – the merger of development (Dev) and IT operations (Ops). Since then, people have been creating new Ops terms at a pretty rapid pace. It’s important to remember that these methods have roots in foundational business management methodologies.

To understand the historical roots of Ops terms, we have to go back to manufacturing quality methods like Lean manufacturing and the writings of quality pioneer W. Edwards Deming. These methodologies were applied in industries across the globe and, more recently, introduced into the software domain under the guise of methods you may find familiar.

For example, [Agile](#) development is an application of the [Theory of Constraints](#) (TOC) to software development. The TOC observed that it was possible to lower manufacturing latency, reduce errors and raise overall system throughput in manufacturing assembly lines using small lot sizes. Agile brings these same benefits to software development by utilizing short development iterations.



DevOps is an application of Lean manufacturing to application development and operations. DevOps automation eliminates waste, reduces errors and minimizes the cycle time of application development and deployment. DevOps has been instrumental in helping software teams become more agile.

Data analytics differs from traditional software development in significant ways. DevOps by itself is insufficient to improve agility in data organizations because data analytics includes both a code and data factory. Whereas quality is generally code dependent in traditional software development, quality is both code and data dependent in data analytics. To design robust, repeatable data pipelines, analytics organizations must turn to automated orchestration, tests and statistical process control (hearkening back to W. Edwards Deming, Figure 2).

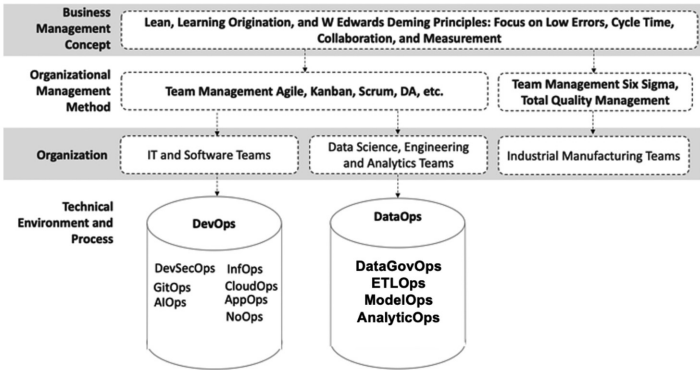


Figure 2: DataGovOps grew out of the DataOps movement in order to apply automation to data governance.

When these various methodologies are backed by a technical platform and applied to data analytics, it's called DataOps. DataOps automation can enable a data organization to be more agile. It reduces cycle time and virtually eliminates data errors, which distract data professionals from their highest priority task – creating new analytics that add value for the enterprise.

## DATAGOVOPS

All of the new Ops terms (Figure 2) are simply an effort to run organizations in a more iterative way. Enterprises seek to build automated systems to run those iterations more efficiently. In data governance, this comes down to finding the right balance between centralized control and decentralized freedom. When governance is enforced through manual processes, policies and enforcement interfere with freedom and creativity. With DataOps automation, control and creativity can coexist. DataGovOps uniquely addresses the DataOps needs of data governance teams who strive to implement robust governance without creating innovation-killing bureaucracy. If you are a governance professional, DataGovOps will not put you out of a job. Instead, you'll focus on managing change in governance policies and implementing the automated systems that enforce, measure, and report governance. In other words, governance-as-code.

# THE ROLE OF DATAGOVOPS IN DATA GOVERNANCE

Data governance can keep people quite busy managing the various aspects of governance across the enterprise:

- Business glossary – Defines terms to maintain consistency throughout the organization. A glossary builds trust in analytics and avoids misunderstandings that impede decision-making.
- Data catalog – A metadata management tool that companies use to inventory and organize the data within their systems. Typical benefits include improvements to data discovery, governance, and access.
- Data lineage – Consider data’s journey from source to ETL tool to data science tool to business tool. Data lineage tells the story of data traversing the system in human terms.
- Data quality – Evaluated through a data quality assessment that determines if data is fit for use.
- Data security – Protecting digital data from the unwanted destructive actions of unauthorized users
- Defined roles and responsibilities – Holding people accountable for adhering to governance and policies

Governance is, first and foremost, concerned with policies and compliance. Some governance initiatives are somewhat akin to policing traffic by handing out speeding tickets. Focusing on violations positions governance in conflict with analytics development. Data governance advocates can get much farther with positive incentives and enablement rather than punishments.

DataGovOps looks to turn all of the inefficient, time-consuming and error-

DATA GOVERNANCE FOCUS		DATA GOVOPS FOCUS
1. Business Glossary & Data Catalog	➔	1. Business Glossary & Data Catalog as Code
2. Data Lineage	➔	2. Process Lineage
3. Data Quality Definitions	➔	3. Automated Data Testing
4. Data Security	➔	4. Self-Service Sandbox and Test Data Management
5. Defined Roles and Responsibilities	➔	5. Agility in Defined Roles and Responsibilities

Figure 3: Focus of data governance and DataGovOps

prone manual processes associated with governance into code or scripts. DataGovOps reimagines governance workflows as repeatable, verifiable automated orchestrations. Figure 3 shows how DataGovOps strengthens the pillars of governance: business glossary and data catalogs, data lineage, data quality, data security, and governance roles and responsibilities.

# Automate Change through Governance as Code

Figure 4 represents a deployment of new analytics from a development environment to a production environment. Imagine you have an existing system that does some ETL, visualization, and data science work. Let’s say you want to add a new data table, join it to another fact table, and update a model and report. The table is new data, and it should also be added to the data catalog. DataGovOps views governance as code or configuration. The orchestration that deploys the new data, new schema, model changes, and updated visualizations also deploys updates to the data catalog. The orchestrations that implement continuous deployment include DataGovOps governance updates into the change management process. All changes are deployed together. Nothing is forgotten or heaped upon an already-busy data analyst as extra work. DataGovOps deploys the changes in the catalog as a unit with the ETL code, models, visualizations, and reports.

Automating governance ensures that it happens in a timely fashion. With manual governance processes, there is always a danger that high-priority tasks will force the data team to defer catalog updates – and occasionally drop the ball. If data catalogs are a deployable unit, updates are more likely to get done, and everyone directly participates in governance via DataGovOps orchestrations.

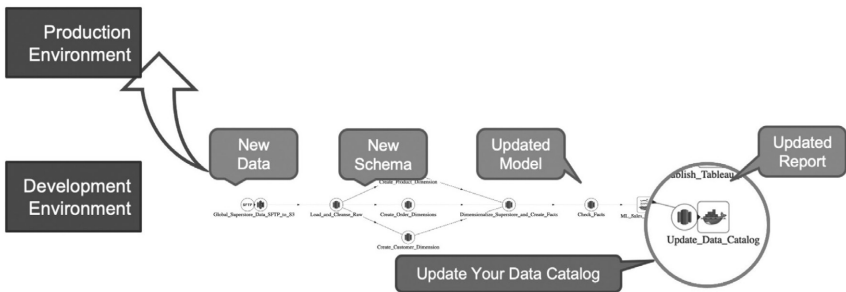


Figure 4: The orchestrations that implement continuous deployment incorporate DataGovOps updates into the change management process.

## DataGovOps Focuses on Process Lineage, Not Just Data Lineage

Data analytics is a profession where your errors get plastered on billboards. When a chart is missing or a report looks wrong, you may find out about it when the VP calls asking questions. Data lineage helps you get those answers.

Figure 5 depicts a data pipeline that ingests data from sftp, builds facts and dimensions, forecasts sales, visualizes data and updates a data catalog. Many data organizations use a mix of tools across numerous locations and data centers. They may use [hybrid cloud](#) with some [centralized data teams and decentralized development](#) using self-service tools. Data lineage helps the data team keep track of this end-to-end process. Which team owns which steps in the process? Which tools are used? Who made changes and when?

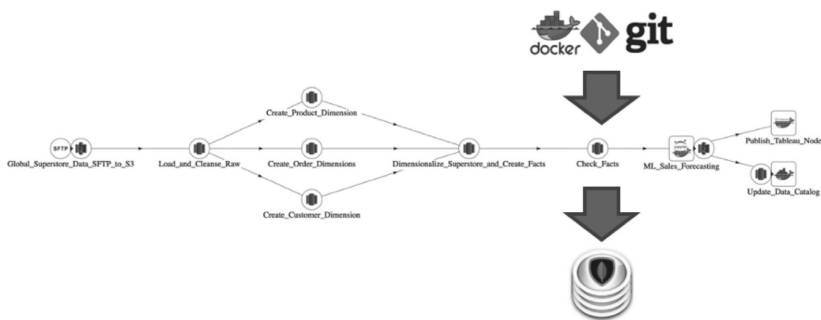


Figure 5: All artifacts that relate to data pipelines are stored in version control so that you have as complete a picture of your data journey as possible.

DataGovOps records and organizes all of the metadata related to data – including the code that acts on the data. Test results, timing data, data quality assessments and all other artifacts generated by execution of the data pipeline document the lineage of data. All metadata is stored in version control so that you have as complete a picture of your data journey as possible. DataGovOps documents the exact process lineage of every tool and step that happened along the data’s journey to value.

## DATAGOVOPS AUTOMATES TESTING AND DATA QUALITY

Manual governance programs evaluate whether data is fit for purpose by performing a data quality assessment. A labor-intensive assessment can only be performed periodically, so at best, it provides a snapshot of data quality at a particular time. DataGovOps takes a more dynamic and comprehensive view of quality. DataGovOps performs continuous testing on data at each stage of the analytics pipeline. Real-time error alerts pinpoint exactly where a problem was detected. Quality assessment is performed as an automated orchestration, so you always have an updated status of data quality. Additionally, DataGovOps performs statistical process control, location balance, historical balance, business logic and other tests, so your data lineage is packed with artifacts that document the data lifecycle. (Figure 6)

If your users see an error in charts, graphs or models, they won’t care whether the error originated with data or the transformations that operate on that data. DataGovOps tests the code that operates on data so that ETL operations and models are validated during deployment and monitored in production.

All of this testing reduces errors to virtually zero, eliminating the stress and embarrassment of having to explain mistakes. When analytics are correct, data is trusted, and the data team has more time for the fun and innovative work that they love doing.

## DATAGOVOPS ENABLES SELF-SERVICE ANALYTICS

A lot of organizations have begun to rely heavily on self-service analytics. From the CDO’s perspective, self-service analytics spur innovation, but can be difficult to manage. Data flowing into uncontrolled workspaces complicates security and

# Automated Testing and Monitoring

Lower Your Error Rates and Embarrassment!

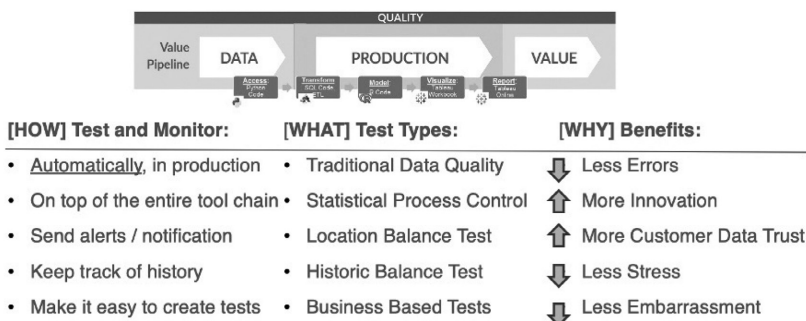


Figure 6: DataGovOps engages in automated testing of data and code to improve analytics quality.

governance. Without visibility into decentralized development, the organization loses track of its data sources and data catalog, and can't standardize metrics. The lack of cohesion makes collaboration more difficult, adds latency to workflows, creates infrastructure silos, and complicates analytics management and deployment. It's hard to keep the trains running on time amid the creative chaos of self-service analytics.

## Self-Service Sandboxes

DataGovOps relies upon self-service sandboxes to improve development and governance agility simultaneously. If manual governance is like handing out speeding tickets, then self-service sandboxes are like purpose-built race tracks. The track enforces where you can go and what you can do, and are built specifically to enable you to go really fast.

A self-service sandbox is an environment that includes everything a data analyst or data scientist needs in order to create analytics. For example:

- Complete toolchain
- Standardized, reusable, analytics components
- Security vault providing access to tools
- Prepackaged datasets – clean, accurate, privacy and security aware
- Role-based access control for a project team
- Integration with workflow management
- Orchestrated path to production – continuous deployment
- DataKitchen Kitchen – a workspace that integrates tools, services and workflows
- Governance – tracking user activity with respect to policies

Self-service environments are created on-demand with built-in background processes that monitor governance. If a user violates policies by adding a table

to a database or exporting sensitive data from the sandbox environment, an automated alert can be forwarded to the appropriate data governance team member. The code and logs associated with development are stored in source control, providing a thorough audit trail.

Note that the self-service sandbox includes test data. Access to test data is a significant pain point for many enterprises. It sometimes takes several months to obtain clean, accurate, and privacy-aware test data that has passed security checks. Once set-up, a self-service environment provides test data on demand. The self-service sandbox enables data teams to deploy faster and lower their error rate. This capability empowers them to iterate more quickly and find solutions to business challenges. The provision of test data on demand is called Test Data Management.

**Test Data Management**

In data science and analytics, test data management (TDM) is the process of managing the data necessary for fulfilling the needs of automated tests, with zero human intervention (or as little as possible).

That means that the TDM solution is responsible for creating the required test data, according to the requirements of the tests. It should also ensure that the data is of the highest possible quality. Poor quality test data is worse than having no data at all since it will generate results that can't be trusted. Another important requirement for test data is fidelity. Test data should resemble, as closely as possible, the real data found in the production servers.

Finally, the TDM process must also guarantee the security and privacy of test data. It's no use to have high-quality data that is as realistic as possible but lack secure, privacy-aware data for testing.

**DATAGOVOPS IS MISSION CONTROL FOR YOUR DATA**

In space flight, a “mission control” center manages a flight from launch until landing, providing stakeholders with complete situational awareness. To properly govern data, you similarly need to know what's happening at a glance – with an ability to quickly drill down into the details. DataGovOps serves as mission control for your data and data pipelines. It provides a single-pane-of-glass view of data and operations, enabling the data team to quickly locate and diagnose problems (Figures 7, 8, & 9).

DAILY BUILD SUMMARY					
Type	Variations/Recipes	Build Status	Last Status	Schedule	Next Build
Analytic	scrape-data	COMPLETE - ON-TIME	Orchestration COMPLETED	AT 08:00 AM	Tomorrow 08:00 AM
Analytic	pov	COMPLETE - ON-TIME	Orchestration COMPLETED	AT 09:15 AM MONDAY through FRIDAY	Tomorrow 09:15 AM
Analytic	gc_dataprep	COMPLETE - ON-TIME	Orchestration COMPLETED	AT 11:00 AM on MONDAY	Monday 11:00 AM
Analytic	generate_labels	COMPLETE - ON-TIME	Orchestration COMPLETED	AT 01:00 PM DAILY	Tomorrow 01:00 PM
Analytic	graph_amazon_data	UPCOMING	Error in Orchestration	AT 12:00 PM every TUESDAY	Tuesday 12:00 PM

Figure 7: DataGovOps mission control view: Daily Build Summary

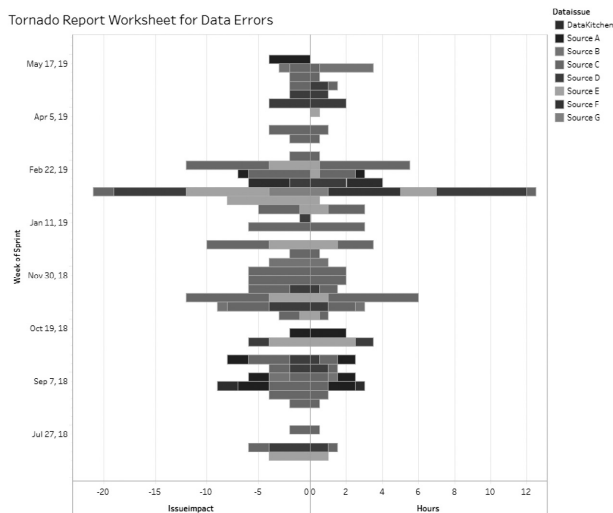


Figure 8: DataGovOps mission control view: the Tornado Report displays a weekly representation of the operational impact of data analytics issues and the time required to resolve them.

	Source 1	Source 2	Source 3	Source 4	Source 5
3/13/16					
3/12/16					
3/11/16					
3/10/16					
3/9/16					
3/8/16					
3/7/16					
3/6/16					
3/5/16					
3/4/16					
3/3/16					
	Key:		missing		
			late		
			on time		

Figure 9: DataGovOps mission control view: The Data Arrival report enables you to track data suppliers and quickly spot delivery issues.

## CONCLUSION

The concept of governance as a policing function that restricts development activity is out-moded and places governance at odds with freedom and innovation. DataGovOps provides a better approach that actively promotes the safe use of data with automation that improves governance while freeing data analysts and scientists from manual tasks. DataGovOps is a prime example of how DataOps can optimize the execution of workflows without burdening the team. DataGovOps transforms governance into a robust, repeatable process that executes alongside development and data operations.





# Slow Cooker Hangi Pork

*Contributed By Campbell Wu*

Hangi is a traditional New Zealand Māori method of cooking food using umu, basically, a type of oven made with heated rocks buried in a pit. Using meats like pork, beef, lamb and chicken this method is usually used on special occasions.

**Prep Time:** 15 minutes — **Cook Time:** 8 hours — **Total Time:** 8 hours 156 mins. — **Yield:** 8 1x

## INGREDIENTS

2 kg fatty pork (pork shoulder or belly) cut into large chunks  
1 cup stuffing mix  
1/2 pumpkin, cut into large chunks  
2 large sweet potatoes, cut into large chunks  
4 strips smoked bacon  
2 tsp smoked paprika  
smoked salt (or sea salt if you can't find one)  
freshly ground black pepper  
banana leaf

## INSTRUCTIONS

1. Make your stuffing according to packet instructions, form into a large ball then set it aside.
2. Lay large banana leaf on a table, arrange bacon in one layer on the bottom, place stuffing ball in the middle, then place meat, sweet potatoes and pumpkin. Season with smoked paprika, smoked salt and freshly ground black pepper.
3. Wrap the meats and vegetables with the banana leaf then secure it with another wrap of aluminum foil. Set it aside.
4. Using aluminum foil, crumple four small rolled-up aluminum foils then place them on the bottom of the slow cooker. Pour enough water to cover the balls then place wrapped meat on top.
5. Cover with damp cloth on top with the sides hanging outside the slow cooker, slow cook for 8 hours on low heat.
6. Remove from pot, unwrap then serve.

*Attribution: Ang Sarap Blog, Author: Raymund*





## Conclusion



# Why Are there So Many -Ops Terms?



It is challenging to coordinate a group of people working toward a shared goal. Work involving large teams and complex processes is even more complicated. Technology-driven companies face these challenges with the added difficulty of a sophisticated technical environment. It is no wonder then that the technology industry sometimes struggles to find coherent terminology to describe its own processes and workflows.

In the past couple of years, there has been a tremendous proliferation of acronyms with the “Ops” suffix. This was started in the software space by the merger of development (dev) and IT operations (Ops). Since then people have been creating new Ops terms at a pretty rapid pace:

[AIOps](#) — Algorithmic IT Operations synonymously titled as “Artificial Intelligence for IT Operations.” Replaces manual IT operations tools with an automated IT operations platform that collects IT data, identifies events and patterns, and reports or remediates issues — all without human intervention.

[AnalyticsOps](#) — Schedule, manage, monitor and maintain models under automation.

**AppOps** — The application developer is also the person responsible for operating the app in production; the operational side of application management, including release automation, remediation, error recovery, monitoring, maintenance.

**ChatOps** — The use of chat clients, chatbots and real-time communication tools to facilitate how software development and operation tasks are communicated and executed.

**CloudOps** — Attain zero downtime based on “continuous operations”; run cloud-based systems in such a way that there’s never the need to take part or all of an application out of service. Software must be updated and placed into production without any interruption in service.

**DataOps** — a collection of data analytics technical practices, workflows, cultural norms and architectural patterns that enable: rapid innovation and experimentation; extremely high quality and very low error rates; collaboration across complex arrays of people, technology, and environments; and clear measurement, monitoring and transparency of results. In a nutshell, DataOps applies [Agile development](#), [DevOps](#) and [Lean manufacturing](#) to data analytics (Data) development and operations (Ops).

**DataSecOps** — DevSecOps for data analytics

**DevOps** — a set of practices that combines software development (Dev) and information-technology operations (Ops) that aims to shorten the systems development life cycle and provide continuous delivery with high software quality.

**DevSecOps** — views security as a shared responsibility integrated from end to end. Emphasizes the need to build a security foundation into DevOps initiatives.

**GitOps** — use of an artifact repository that always contains declarative descriptions of the infrastructure currently desired in the production environment and an automated process to make the production environment match the described state in the repository.

**InfraOps** — the layer consisting of the management of the physical and virtual environment, which may very well be within a cloud environment. On top of this layer would be Service Operations (‘SvcOps’) and Application Operations (‘AppOps’).

**MLOps** — machine learning operations practices meant to standardize and streamline the lifecycle of machine learning in production; orchestrating the movement of machine learning models, data and outcomes between the systems.

**ModelOps** — automate the deployment, monitoring, governance and continuous improvement of data analytics models running 24/7 within the enterprise.

**NoOps** — no IT infrastructure; software and software-defined hardware provisioned dynamically.

There are probably even more Ops terms out there (honestly, got tired of googling). Naturally, people have found this confusing and have questioned whether all these acronyms are necessary. As students of management methodology and lovers of software tools, we thought we might take a stab at trying to sort this all out.

# TAYLORISM

After the industrial revolution (~1760 to 1840), manufacturing still greatly relied upon human labor. Naturally, managers looked for ways to improve efficiency. [Fred W. Taylor](#) (1856–1915) revolutionized factories with a methodology called “scientific management” or “Taylorism.” To improve plant productivity, Taylorism timed the movements of workers, eliminating wasted motion or unnecessary steps in repetitive jobs. Applying analysis to manufacturing processes produced undeniable efficiencies and naturally, provoked resentment by labor when taken to extremes. Taylorism took a top-down approach to managing manufacturing and treated people as automatons.

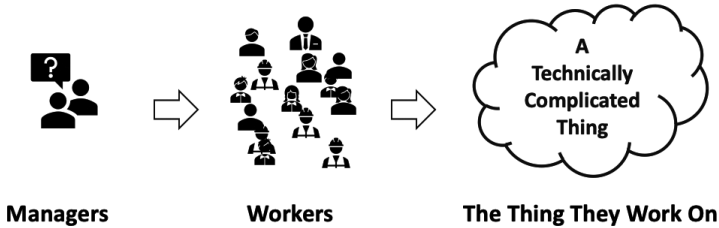


Figure 1: When trying to produce a “technically complex thing” (TCT) organizations need communication between managers and the people doing jobs.

Top-down, dictatorial control hardly works in modern manufacturing endeavors, which have grown in scale and complexity. When trying to produce a “technically complex thing” (TCT) organizations need communication between managers and the people doing jobs. A TCT could be industrial manufacturing (like ventilators), software or data analytics. These endeavors demand a culture of honesty, safety, numeracy, trust and feedback. New management methods emerged based on these requirements.

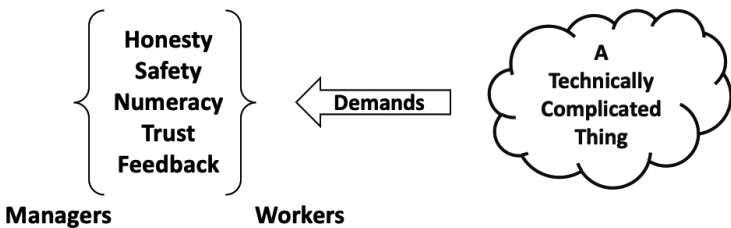


Figure 2: Production of a “technically complex thing” demands that a culture of honesty, safety, trust and feedback exist between managers and employees.

# MANAGING TECHNICALLY COMPLEX THINGS

In the era of producing TCT’s, we have many methodologies that organize human activity to deliver benefits to society and value to individuals while eliminating waste. We sometimes group all of these methodologies underneath the term “Lean manufacturing.” Lean seeks to identify waste in manufacturing processes by focusing on eliminating errors, cycle time, collaboration and measurement.

Lean is about self-reflection and seeking smarter, less wasteful dynamic solutions together.

### APPLYING LEAN TO SOFTWARE DEVELOPMENT

As the software industry emerged, companies began to understand that lean principles could be equally transformative in the context of software development. Development organizations began to apply “lean” to their software development processes.

### LEAN IN DATA SCIENCE

More recently, data analytics organizations are applying lean principles to their methodologies. Enterprises following this path find that these methods help data science/engineering/BI/governance teams produce better results more efficiently. The application of lean principles in the technology space is facilitated by progress in automating the technical environment underlying the end-to-end processes. DevOps continuous deployment played a critical role. After all, there’s little point in performing weekly sprints if it takes months to deploy a release. The Agile management process was a step forward, but insufficient. Dev teams needed the support of the technical environment to optimize management processes further.



Figure 3: In the era of TCTs, manufacturing methods like Six Sigma, Total Quality Management and the Toyota Way focus on eliminating errors, cycle time, collaboration and measurement. These methods built upon the pioneering work of W. Edwards Deming.



Figure 4: Lean manufacturing principles applied to software development took the form of Agile, Scrum, Kanban and Extreme Programming.

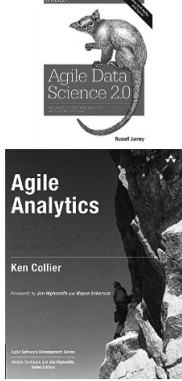


Figure 5 : Applying lean principles to data analytics



## THE TECHNICAL ENVIRONMENTS ENABLING AGILE

DevOps supplies the technical environment that enables Agile to be applied in software development and IT. With DevOps, dev teams create automated processes that deploy new features or bug fixes in minutes. The flexibility that DevOps and Agile enable have helped many companies attain a leadership position in their markets. However, DevOps and Agile together could not enable these same efficiencies in data analytics.

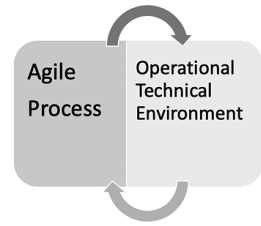


Figure 6: The benefits of Agile development can't be fully utilized without optimizing the operational technical environment.

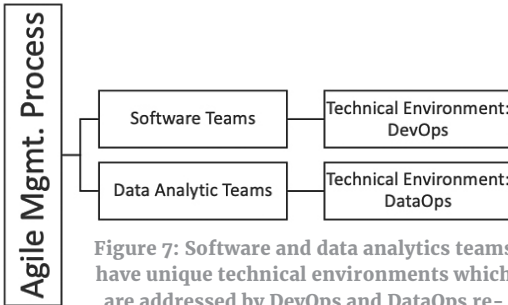
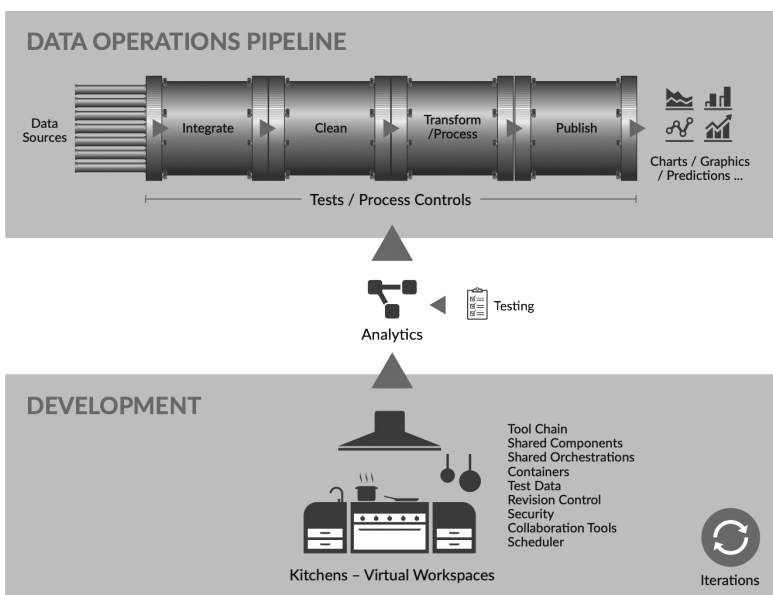


Figure 7: Software and data analytics teams have unique technical environments which are addressed by DevOps and DataOps respectively.

## DATA ANALYTICS REQUIRES MORE THAN DEVOPS

Data analytics differs from traditional software development in significant ways. For example, when a data professional spins up a virtual sandbox environment for a new development project, they need data in addition to a clone of the production technical environment. In software development, test data is usually pretty straightforward. In data analytics, there could be governance concerns. Data quality affects outcomes. There could be concerns about the age of data. A model trained on data that is three months old might work differently on data that is one day old. Also, predictive analytics can be invalidated if data doesn't all originate from the same point in time. There are many issues to consider when provisioning test data in data analytics.

Another major difference between software development and data analytics is the data factory. Streams of data continuously flow through the data analytics pipeline. Data analytics more resembles a manufacturing process than a software application. For these reasons (and more) DevOps by itself is insufficient to enable Agile in data organizations. Data analytics created DataOps; a technical environment tuned to the needs and challenges of data teams.



**Figure 8: Data analytics must orchestrate many pipelines including data operations.**

## TEAMS AND TECHNICAL ENVIRONMENTS DEFINE OPS

DevOps is the foundational technical environment for IT and software teams. DataOps is the technical environment of data analytics teams. The figure below shows how DevOps and DataOps serve as the foundation for all other Ops'. Each of the other Ops' represent branches off the DevOps and DataOps trees. Perhaps a new Ops is coined for a subgroup of a team and/or the requirement to use different methods or tools. For example, DevSecOps emphasizes security in DevOps development. DataSecOps performs the same function for DataOps.

When terms point to the same team members and the same genre of tools, the Ops terms are synonymous. For example, ModelOps, MLOps, and AnalyticOps focus on the unique problems of data scientists creating, deploying and maintaining models and AI-assisted analytics using ML and AI tools and methods. Maybe the industry doesn't need all three of these terms.

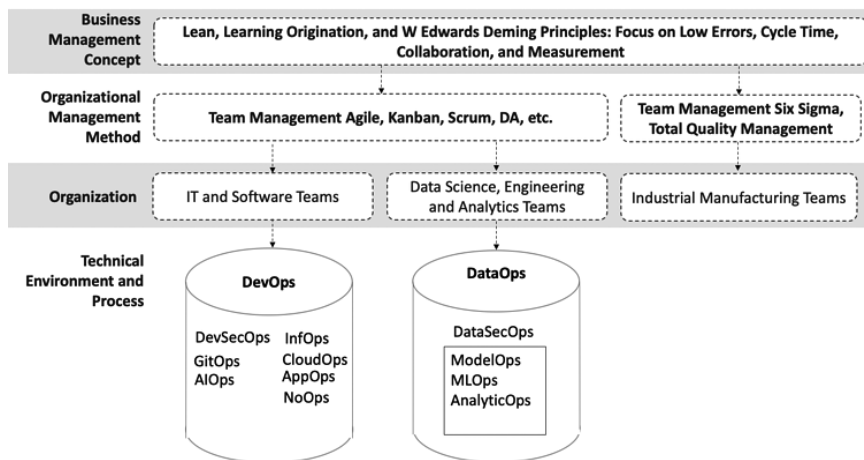


Figure 9: Ops terms can be organized by team and technical environment/process.

## STAY LEAN

Whenever a term or acronym gains momentum, marketers go to great lengths to associate their existing offerings with whatever is being hyped. Sometimes that creates a backlash that drowns out some good ideas. You may believe that you do not need a new Ops term or you may find that it helps to galvanize your target audience and increases focus on the technical environment critical to your projects. Stay focused on the goals of lean manufacturing. Anything that eliminates errors, streamlines workflow processes, improves collaboration and enhances transparency aligns with DevOps, DataOps and all the other possible Ops' that are out there.



# Mom's Keto Chocolate Peanut Butter Fat Bombs

Contributed By Nick Bracy

## INGREDIENTS

16 ounces (2 pkg) softened full fat cream cheese

6 tablespoons crunchy peanut butter (Ingredients should list just peanuts. No added sugar)

8 tablespoons Lakanto Monkfruit Sweetener with Erythritol or another granulated sugar substitute with Erythritol such as Swerve (not one with maltodextrin or sucralose such as Splenda)

4 tablespoons unsweetened cocoa powder

4 tablespoons Lily's Sugar-Free (stevia sweetened) Dark Chocolate Chips finely chopped

## INSTRUCTIONS

1. Place chocolate chips in a food processor and chop finely. Add softened cream cheese, crunchy peanut butter, sugar substitute, cocoa powder, and process until well combined. You can also chop the chips by hand with a sharp knife and mix everything together in a bowl if you prefer.
2. With a small cookie scoop or a spoon scoop about one tablespoon and place onto a parchment-lined baking sheet.
3. Freeze for 20-30 minutes to firm up. Remove and place in freezer bags to keep frozen.
4. Makes about 36 servings or you can make them larger for fewer servings.

Approximate macros for each fat bomb.

Net Carbs 1.3g

Fiber 0.8g

Total Carbs 5.1g

Protein 1.6g

Fat 6.2g

Calories 67

**Note:** I suggest you make the recipe as written the first time and then adjust it according to your taste by adding a little more or less sweetener, cocoa or peanut butter, although that may alter the macros.



# A Guide to Understanding DataOps Solutions



## BREAKING THROUGH THE NOISE

DataOps is the hot topic on every data professional's lips these days, and we expect to hear much more about DataOps in the coming years. This is not surprising given that DataOps holds true potential for enabling enterprise data teams to generate significant business value from their data. [Companies that implement DataOps](#) find that they are able to reduce cycle times from weeks (or months) to one day, virtually eliminate data errors, and dramatically improve the productivity of data engineers and analysts.

As a result, vendors that market DataOps capabilities have grown in pace with the popularity of the practice. To date, we count over 100 companies in the [DataOps ecosystem](#). However, the rush to rebrand existing products as related to DataOps has created some marketplace confusion. Because it is such a new category, both overly narrow and overly broad definitions of DataOps abound. As a result, it is easy to get overwhelmed when trying to evaluate different solutions and determine whether they will help you achieve your DataOps goals.

## SO, WHAT IS DATAOPS ANYWAY?

In short, [DataOps](#) is a set of technical practices, cultural norms, and architectures that enable:

- Rapid experimentation and innovation for the fastest delivery of new insights to customers;
- Low error rates;
- Collaboration across complex sets of people, technology, and environments;
- Clear measurement and monitoring of results.

Similarly, Gartner defines DataOps as, “a collaborative data management practice focused on improving the communication, integration, and automation of data flows between data managers and data consumers across an organization.”

Like its DevOps cousin, key elements of DataOps include increased deployment frequency, automated testing and monitoring, version control, and collaboration.

This sounds great and you are ready to get started, but the next big question is how can your organization best achieve this transformation? How can you sift through all the marketing speak and find the solutions that will truly help you?

## UNDERSTANDING DATAOPS SOLUTIONS

DataOps addresses a broad set of workflow processes, including analytics creation and your end-to-end data operations pipeline. In general, it's not a single tool you can purchase and forget. Fundamentally, any DataOps solution should improve your ability to orchestrate data pipelines, automate testing and monitoring, and speed new feature deployment – while continuing to choose the right tool for the right part of the job.

To be certain, many companies that are marketing their products as DataOps solutions play a critical role in the ecosystem. However, it is important to understand exactly what role they play. If you purchase a fancy new ETL tool, will you suddenly realize all the benefits of DataOps? Probably not.

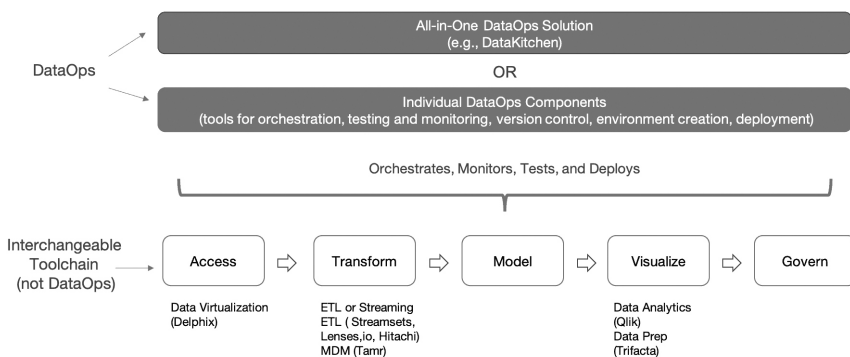
When evaluating DataOps solutions, consider the following ways that companies are marketing their capabilities.

**The Data Toolchain** – Many tools being marketed today as DataOps solutions are simply independent components of the data toolchain that collect, store, transform, visualize, and govern the data running through the pipeline. Although all of these technologies play an important role in the value pipeline, they do not ensure that each step in the data pipeline is executed and coordinated as a single, integrated, and accurate process or help people and teams better collaborate. Remember that a DataOps process automates the orchestration and testing of these tools across the pipeline. In fact, in a true DataOps environment, it does not matter which data tools you use. Your team can continue to use the ETL or analytics tools they like best or add new tools at any time. Typically, components of toolchain are being marketed as DataOps solutions in two different ways.



- **DataOps Rebranding** – One of the reasons that the concept of DataOps has become so muddled is because some companies are rebranding the actual concept of DataOps to fit with what their product does. For example, DataOps has been rebranded as ETL (e.g., Hitachi Vantara, Attunity), streaming ETL (e.g., StreamSets, Lenses.io), or data virtualization (e.g., Delphix).
- **The Halo Effect** – Because DataOps is a hot marketing term it is not surprising that many data companies are using this concept in their marketing to generate interest. The companies doing “halo effect” marketing are using the correct definition of DataOps. However, if you read closely, the message is generally that, “DataOps is great, but use our tool first.” Some examples of this type of marketing are IBM’s marketing of its Cloud Pak for Data, Trifacta for end-user data prep, and Qlik for data analytics.

## The DataOps Ecosystem



**Data Process Tools** – Data process and automation tools are being correctly marketed as important components of a DataOps solution. You’ll need some combination of these tools if you decide to [implement DataOps yourself](#). Many popular DevOps tools can also be used.

- Orchestration of end-to-end multi-tool, multi-environment pipelines can be facilitated by tools like Apache Airflow or Saagie.
- Automated Testing and Monitoring at every step in production and development pipelines is important to catch and address errors before they reach the business user. iCEDQ is a leading testing and monitoring platform.
- Environment and Deployment technologies allow teams to spin-up self-service work environments and innovate without breaking production. New features can be deployed with the push of a button. There are a host of tools built for this purpose, including well-known open-source tools such as Git (version control), Docker (containerization), and Jenkins (CI/CD).

**All-in-One DataOps Platforms** – Building a DataOps environment is challenging and requires a true organizational transformation and commitment of time and resources. Even the best-equipped organizations can encounter obstacles trying

to bring it all together. DataKitchen offers the first end-to-end platform that can serve as a foundation for your DataOps initiative. It seamlessly automates and manages workflows related to both data operations and new analytics development, using the tools you already have. In fact, the DataKitchen platform can interoperate with any of the data toolchain and process tools mentioned above. The platform fosters collaboration by providing a single view of the entire pipeline. Version control and environment management enable work to move seamlessly from person to person or team to team. The platform also provides useful metrics that show whether your DataOps initiative is adding value.

DataOps, when implemented correctly, holds exciting promise for data teams to be able to reclaim control of their data pipelines and deliver value instantly without errors. It is easy to get confused by all the marketing noise, but remember that DataOps, at its core, is a collaborative process that orchestrates data pipelines, automates testing and monitoring, and speeds new feature deployment. Whether you use an all-in-one tool like DataKitchen or build it yourself, the right combination of tools, processes, and people are critical to make DataOps a success.

# Gil's Old Fashion Fudge

*Contributed By Gil Benghiat*

This tastes as good as the fudge they sell at tourist destinations. This is a multi-hour project and is great for a rainy or snowy day. Make sure you have a candy thermometer and parchment paper before you start.

## INGREDIENTS

4 cups sugar  
½ cups powdered cocoa  
¼ teaspoon salt  
2 cups milk  
¾ cups butter  
1 ½ teaspoons vanilla

## INSTRUCTIONS

1. Line an 8- or 9-inch square pan with parchment paper.
2. Mix sugar, cocoa and salt in heavy 4-quart saucepan or larger; stir in milk. Cook over medium heat, stirring constantly until mixture comes to full rolling boil. Boil, without stirring, until mixture reaches 234°F on candy thermometer or until small amount of mixture dropped into very cold water, forms a soft ball which flattens when removed from water. This can take a while.
3. Remove from heat. Add butter and vanilla. DO NOT STIR. Cool at room temperature to 110°F (1-2 hours). Fold with wooden DataKitchen spoon until fudge thickens and just begins to lose some of its gloss (about 7 minutes).
4. Quickly spread in prepared pan; cool completely. Cut into squares.
5. Store in a tightly covered container at room temperature or in the refrigerator.

**NOTE:** For best results, do not double this recipe. The directions must be followed exactly. In the third step, beat too little and the fudge is too soft. Beat too long and it becomes hard and sugary.

*Attribution: Ingredients and instructions tweaked from [Hershey's Fudge](#)*



# What a DataOps Platform Can Do For You



Leading software companies perform millions of code releases per year. Typical data analytics organizations perform less than 10. This gap explains why most data analytics projects fail to deliver. Without the capability to move at lightning speed, data analytics can't adapt to fast-paced markets and keep up with the endless stream of requests generated by business users. Despite soaring levels of investment, the percentage of organizations that describe themselves as “data-driven” has fallen since 2017.

Software teams have faced similar challenges and found answers. The methods that yielded tremendous improvements in software development productivity can deliver similar results for data organizations. In the data industry, the process of going from 10 releases per year to millions is called [“DataOps.”](#)

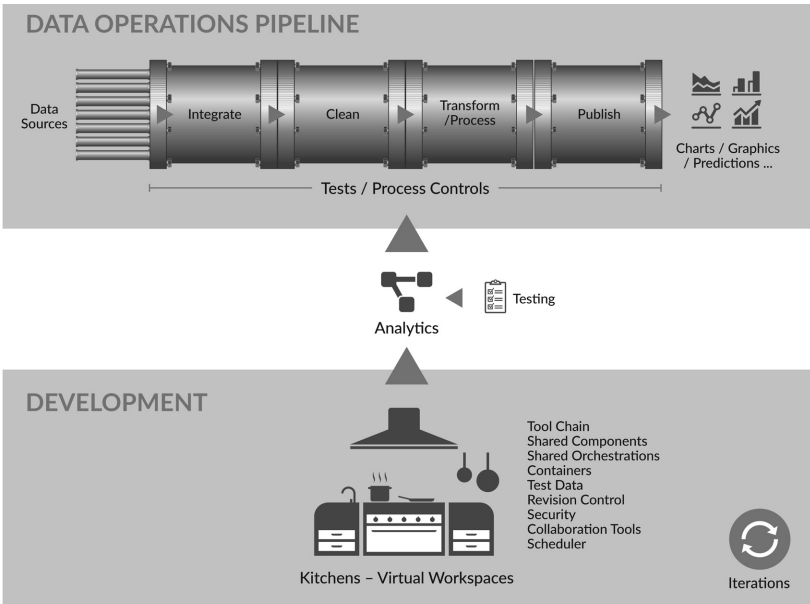
DataOps enables data organizations to accelerate the development of new analytics, deploy confidently with the push of a button, and reduce data errors to virtually zero. This represents an orders-of-magnitude decrease in analytics cycle time and improvement in quality. Sound impossible? It's already happened in

software companies, like Amazon, Facebook, Netflix and many others. If your data organization neglects to modernize its processes, it risks being left behind in an increasingly “on-demand economy.”

## ACCELERATING YOUR DATAOPS INITIATIVE

The goal of DataOps is to enable the analytics team to keep pace with user requests. Data analysts and business users can unlock enormous creativity when they work closely together. When it takes six months to release a 20-line SQL change, innovation is stymied; users get frustrated.

Applying DataOps requires a combination of new methods and automation that augment an enterprise’s existing toolchain. Some organizations build DataOps capabilities from scratch, but the fastest way to realize the benefits of DataOps is to adopt an off-the-shelf DataOps Platform. As the DataOps Platform is a relatively new product category, unlike anything else on the market, there is still a general lack of understanding about how it delivers such significant improvements in analytics productivity and quality.



## DATAOPS PLATFORM

A DataOps Platform unifies the end-to-end workflow and processes related to data analytics planning, development and operations into a single, common framework, improving overall collaboration. It incorporates your existing tools into automated orchestrations that drive analytics creation and the transformation of raw data to insights. The DataOps Platform accomplishes this goal by managing the creation, deployment and production execution of analytics. DataOps Platforms offer four fundamental capabilities:

- Spins up safe and synchronized workspaces – Using virtualization, DataOps separates and harmonizes your production and development environments. Aligning the two technical environments avoids unexpected errors during deployment. Access control secures each workspace and domain. When it's time to start a new project, data scientists spin up self-serve development sandboxes in minutes – this includes test data, validation tests, tools, a password vault, – in short, everything they need. No more waiting months for IT.
- Automates deployment – New analytics pass extensive validation tests and seamlessly move from development to production engineering and then operations, with a few clicks. Verification tests replace your impact review board, minimizing the time and effort required to deploy.
- Orchestrates, tests, and monitors the data pipeline – Data flows in from hundreds or thousands of sources and is integrated, cleaned, processed and published in analytics. As millions of data points flow through the pipeline, tests distributed throughout the data pipeline monitor work in progress and check data for anomalies. Virtually zero errors reach user analytics. When errors are found, DataOps takes appropriate action based on severity: warnings, alerts, or even suspension of a data source. Dashboards that summarize test results and activity provide unprecedented visibility into operations and development. The DataOps Platform provides quality and productivity metrics, showing the progress of your DataOps initiative.
- Fosters Collaboration – DataOps automates workflows to coordinate tasks and improve teamwork. Workspace environments provide the structure to move analytics through the development workflow, from person to person, eventually reaching production. Sandboxes feature reusable analytics components saving time and enforcing standardization. Coupled with source control, workspaces branch and merge, providing centralized control of artifacts. With a DataOps Platform, everyone has a common view of the development and operations pipelines.

## **FOCUSING ON VALUE ADD**

With an orchestrated data operations pipeline, quality controls and an automated development workflow, the DataOps Platform minimizes unplanned work. Task coordination between team members and groups leads to a more transparent and robust workflow. DataOps tests virtually eliminate data errors. The DataOps Platform enables data professionals to develop and deploy new analytics at lightning speed, delighting users and delivering insights that meaningfully impact the enterprise's goals and initiatives.





# Chapssal Doughnuts

*Contributed By Brandon Stephens*

This recipe is one of the most popular Korean snacks, chapssal doughnuts. They're a modern Korean treat combining traditional Korean rice cakes with Western-style deep-fried doughnut balls. On the outside, the dough is crispy and chewy, and on the inside, there's soft, lightly sweet red bean paste.

## INGREDIENTS

### *For the dough:*

- 1 cup glutinous rice flour (aka sweet rice flour, Mochiko powder, or chapssalgaru)
- 1 tablespoon flour
- ¼ teaspoon baking soda
- ¼ teaspoon kosher salt
- 1 tablespoon sugar
- 1 tablespoon unsalted butter, melted
- ¼ cup plus 3 tablespoons hot water

### *For sweet red bean paste (about 1 pound):*

- 1 cup dried azuki beans (aka red beans, or pat) 7 ounces or 200 grams
- ¼ cup sugar
- ¼ cup rice syrup (or corn syrup)
- ¼ teaspoon kosher salt
- 1 teaspoon vanilla extract

### *For frying and coating:*

- vegetable oil
- 2 tablespoons of white sugar





# DataOps Resources

The Agile Manifesto	<a href="http://agilemanifesto.org/">http://agilemanifesto.org/</a>
DataOps Blog	<a href="http://bit.ly/2Ef2Hto">http://bit.ly/2Ef2Hto</a>
The DataOps Manifesto	<a href="http://dataopsmanifesto.org">http://dataopsmanifesto.org</a>
DataOps Videos	<a href="http://bit.ly/2UFcKO8">http://bit.ly/2UFcKO8</a>
Scrum Guides	<a href="http://www.scrumguides.org">http://www.scrumguides.org</a>
Statistical Process Control	<a href="https://en.wikipedia.org/wiki/Statistical_process_control">https://en.wikipedia.org/wiki/Statistical_process_control</a>
W. Edwards Deming	<a href="https://en.wikipedia.org/wiki/W._Edwards_Deming">https://en.wikipedia.org/wiki/W._Edwards_Deming</a>
Wikipedia DataOps	<a href="http://bit.ly/2DnlqR1">http://bit.ly/2DnlqR1</a>
Wikipedia DevOps	<a href="https://en.wikipedia.org/wiki/DevOps">https://en.wikipedia.org/wiki/DevOps</a>
DataKitchen Website	<a href="http://datakitchen.io">datakitchen.io</a>
DataOps Maturity Model Assessment	<a href="https://datakitchen.io/dataops-maturity-model/">https://datakitchen.io/dataops-maturity-model/</a>
The DataOps Cookbook	<a href="https://datakitchen.io/the-dataops-cookbook/">https://datakitchen.io/the-dataops-cookbook/</a>

# About the Authors

**Christopher Bergh** is a Founder and Head Chef at DataKitchen where, among other activities, he is leading DataKitchen's DataOps initiative. Chris has more than 25 years of research, engineering, analytics, and executive management experience.

Previously, Chris was Regional Vice President in the Revenue Management Intelligence group in Model N. Before Model N, Chris was COO of LeapFrogRx, a descriptive and predictive analytics software and service provider. Chris led the acquisition of LeapFrogRx by Model N in January 2012. Prior to LeapFrogRx Chris was CTO and VP of Product Management of MarketSoft (now part of IBM) an innovative Enterprise Marketing Management software. Prior to that, Chris developed Microsoft Passport, the predecessor to Windows Live ID, a distributed authentication system used by 100s of Millions of users today. He was awarded a US Patent for his work on that project. Before joining Microsoft, he led the technical architecture and implementation of Firefly Passport, an early leader in Internet Personalization and Privacy. Microsoft subsequently acquired Firefly. Chris led the development of the first travel-related e-commerce web site at NetMarket. Chris began his career at the Massachusetts Institute of Technology's (MIT) Lincoln Laboratory and NASA Ames Research Center. There he created software and algorithms that provided aircraft arrival optimization assistance to Air Traffic Controllers at several major airports in the United States.

Chris served as a Peace Corps Volunteer Math Teacher in Botswana, Africa. Chris has an M.S. from Columbia University and a B.S. from the University of Wisconsin-Madison. He is an avid cyclist, hiker, reader, and father of two college age children.

**Eran Strod** is a Marketing Chef at DataKitchen where he writes white papers, case studies and contributes to the DataOps blog. He is passionate about applying process-oriented management science to data and analytics.

Eran was previously Director of Marketing for Atrenne Integrated Solutions (now Celestica) and has held product marketing and systems engineering roles at Curtiss-Wright, Black Duck Software (now Synopsys), Mercury Systems, Motorola Computer Group (now Artesyn), and Freescale Semiconductor (now NXP), where he was a contributing author to the book "Network Processor Design, Issues and Practices." Eran began his career as a software developer at CSPI working in the field of embedded computing.

Eran holds a B.A. in Computer Information Science and Psychology from the University of California at Santa Cruz (Stevenson College) and an M.B.A. from Northeastern University. He is a proud dad and enjoys hiking, travel and watching the New England Patriots.


**James Royster** is the Senior Director of Operations & Analytics at Adamas Pharmaceuticals. He has more than 20 years of experience in business analytics, pharmaceutical brand launch strategy, and project management. He and his teams have provided analytic tools to facilitate decision-making in complex situations, adapting to rapidly changing priorities. As Head of Data Strategy & Operations at Celgene leading an MS product launch, he orchestrated a team of over 100 internal and external resources operating as a single unit and was responsible for data infrastructure development and data quality.

# TRY DATAKITCHEN

## OPEN SOURCE DATA OBSERVABILITY

Data Quality Testing, Validation, Anomaly Detection, End-to-End Production Observability, and Alerting!

**Install Now**



The screenshot displays the DataKitchen web application. On the left, a data lineage diagram shows a flow from 'Data Loader' to 'E\_Customer', 'E\_Partner', 'E\_Order', and 'E\_Product' topics. On the right, the 'Data Profiling Results' section is visible, showing a table of data quality metrics for various topics. Below the table, there are charts including a bar chart and a pie chart, along with a 'DataKitchen' logo.

## **Additional Recipes**





# Chicken Breasts with Marsala Wine

*Contributed By Joanne Ferrari*

## INGREDIENTS

4 Boneless Chicken Breasts  
2 Eggs  
1 ½ cup Dry Italian Seasoned Breadcrumbs  
1/3 cup Parmesan Cheese  
1 ½ cup Dry Marsala Wine  
¼ cup Butter  
2 Tbs. Olive Oil  
8 oz. Sliced Mushrooms  
Salt & Pepper to Taste

## INSTRUCTIONS

1. Beat eggs with salt and pepper in a medium bowl. Combine breadcrumbs and cheese in a small bowl.
2. Dip chicken in beaten eggs then coat with breadcrumb mixture.
3. Press mixture onto chicken and let stand 10-15 min.
4. Melt butter with oil in a large heavy skillet.
5. When butter foams, add chicken.
6. Cook over medium heat 2-3 min. each side or until chicken has a light golden crust.
7. Add wine and mushrooms.
8. Cover and reduce heat.
9. Simmer 15-20 min. or until chicken is tender.
10. If the sauce looks too dry, add a little more wine.
11. Serve immediately over pasta.

# Dakgalbi

*Contributed By Brandon Stephens*

Spicy stir-fried chicken with vegetables

**INGREDIENTS** — *Serves 2 to 3*

## FOR CHICKEN AND MARINADE:

- 1 pound deboned chicken thigh (or drumsticks), cut into small bite-sized pieces
- 2 tablespoons milk
- 1 tablespoon soy sauce
- ¼ teaspoon ground black pepper

## FOR THE SEASONING SAUCE:

- 12 garlic cloves, minced
- 1 teaspoon peeled and minced ginger
- 2 tablespoons soy sauce
- ½ cup water
- ⅓ cup gochu-garu (Korean hot pepper flakes)
- 2 tablespoons rice syrup (or corn syrup, or sugar)
- ½ teaspoon ground black pepper

## VEGGIES AND RICE CAKES:

- 4 ounces sliced rice cake (1 cup), soaked in cold water at least 10 minutes
- 8 ounces cabbage, cored and cut into bite-sized pieces
- 4 ounces (½ of large onion), sliced
- 1 small carrot (about ⅓ cup), peeled and sliced
- 1 or 2 green chili peppers, sliced
- ¾ cup peeled sweet potato, sliced into ¼ inch thick bite-size pieces.
- 12 perilla leaves (or basil leaves), cut or torn a few times
- ½ cup water
- 1 bowl of rice (optional)
- ¼ cup chopped fermented kimchi (optional)

## INSTRUCTIONS

### MARINATE CHICKEN:

- Combine the chicken, milk, soy sauce, and ground black pepper in a bowl and mix all together with a spoon.
- Cover and set aside.

### MAKE SEASONING SAUCE:

- Combine the minced garlic, ginger, soy sauce, water, gochu-garu (Korean hot pepper flakes), rice syrup, and ground black pepper in a bowl. Mix well with a spoon and set aside.

### COOK AND SERVE:

- Spread the cabbage on the bottom of a large, heavy, and shallow pan or skillet.
- Add onion, carrot, green chili pepper, sweet potato, rice cake, and perilla leaves in that order.
- Add the chicken in the center. Pour the seasoning sauce over the chicken and spread it with a wooden spoon. Add 1/2 cup water.
- Cover and cook for 3 to 4 minutes over medium-high heat until it starts boiling. Turn down the heat to medium. Open and stir with a (DataKitchen) wooden spoon so that the pan doesn't burn and the ingredients and sauce mix evenly. Cover and cook another 13 to 15 minutes over medium heat, stirring occasionally until the chicken and sweet potato are cooked thoroughly.
- Keep the heat low during the meal. Cook, stir, eat, and talk. The pieces will be hot, so be careful! Turn off the heat when the chicken and potato are totally cooked.
- Give a bowl to each diner. They can each take some out of the pan into their bowl, and eat. When it's almost totally finished, make some fried rice by adding some rice and chopped kimchi to what's left on the grill. Stir with a wooden spoon over medium heat for a few minutes. Serve in separate bowls, or give everyone a spoon and let them eat from the pan together.

Attribution: Maangchi

# Isaac's Special Chicken

*Contributed By Gil Benghiat*

This recipe is from the 1960s where you combine processed ingredients into a fast and easy dish.

## INGREDIENTS

- 1-2 Chickens cut in quarters or pieces or parts that you like
- 1 can whole berry cranberry sauce
- 1 packet dried onion soup
- 1 bottle Catalina dressing

## INSTRUCTIONS

1. Mix ingredients in a baking dish
2. Roll chicken in it
3. Bake uncovered 1 hour at 350 degrees

# White Russian Tiramisu Cake

*Contributed By Joanne Ferrari*

## INGREDIENTS

3 cups Strong Brewed Coffee, cooled  
½ cup Kahlua, divided  
1 cup Mascarpone Cheese  
16oz Cream Cheese, softened  
½ cup White Sugar  
⅓ cup Light Brown Sugar  
48 Lady Fingers  
¼ cup of Cocoa Powder (for dusting)  
2 cups Heavy Whipping Cream  
½ cup Confectioners' Sugar  
1 TBSP Clear Vanilla

## INSTRUCTIONS

1. Place a mixing bowl in the freezer to chill. Combine 3 cups of strong, cooled, brewed coffee and ¼ cup of Kahlua in a container. Set aside.
2. Using a stand mixer combine mascarpone, cream cheese, white sugar, brown sugar, and remaining ¼ cup of Kahlua. Beat until smooth.
3. Dip a ladyfinger into the coffee-Kahlua mixture. Place the ladyfinger dipped-side down in a 13" x 9" pan. Repeat until the bottom of the pan is covered in dipped ladyfingers.
4. Spread a layer of the cheese mixture over the ladyfingers. Dust with cocoa powder. Repeat for a second layer. Set aside.
5. Remove the chilled mixing bowl from the freezer; using a stand mixer combine heavy whipping cream, vanilla and confectioners' sugar. Using the whisk attachment beat on medium speed until soft peaks form and the mixture is firm. Do not overbeat.
6. Spread whipped cream on top of the pan containing the ladyfinger/cheese mixture. Dust with cocoa powder.  
Optional: if you like, top with chocolate shavings.  
Keep refrigerated.

Attribution: Ann's Entitled Life



## What the Experts are Saying

"Recipes for DataOps" answers the all-important question: how do we get started with DataOps? Concise and well-written, this book is a wonderful primer for any data & analytics professional or business owner who finally wants to get real value from their data assets."

— Wayne Eckerson  
President, The Eckerson Group

The authors of "Recipes for DataOps" have understood what most of the industry has yet to learn - the key to data success lies not in having large data science teams or the latest machinery, components, and tools, but in establishing efficient, value-driven work processes. This book is a great step-by-step guide to unlocking that capability and achieving a DataOps culture - the data- and AI equivalent of Lean manufacturing. It is a long journey, but rewarding from the start. This book is one of the few good DataOps guides available, and I recommend it to everyone that is working with data on a daily basis - data engineers, analysts, data scientists, product owners, and data team managers. Moreover, that Maori slow-cooked pork seems delicious.

— Lars Albertsson  
Founder of Scling

Chris Bergh, Eran Strod, and James Royster have written a unique book that is the go-to guide for DataOps transformation. It covers an impressive breadth and scope of topics and explains them in a highly accessible way. If your organization is in any way struggling to deliver high-quality data analytics at speed you owe it to yourself to read this book, there is something to learn for everyone.

— Harvinder Atwal  
Author, Practical DataOps: Delivering Agile Data Science at Scale

DataOps is one of the most important innovations in the data industry in the last decade. It will transform how your organization delivers analytic capabilities, drives value, and shifts to data-supported decisions. The latest book from DataKitchen is the "how-to" manual that you need to start your DataOps transformation.

— Laura Madsen  
Author, Disrupting Data Governance

Takes the path to success with DataOps to a whole new level of understanding. There are so many actionable insights in the book.

— Jesse Anderson  
Author, Data Teams

This book is a great read and really important for any organization that wants to transform with DataOps rather than tinker around the edges. The book covers important concepts critical to the success of DataOps such as the Theory of Constraints, Process Measurement, and DataGovOps. It clarifies the full approach - business requirements first, tools second - so you are not creating more constraints before you have even started. This is a must-read for anyone open to finding better ways of working through DataOps.

— Simon Trewin  
Author, The Dataops Revolution: Delivering the Data-Driven Enterprise