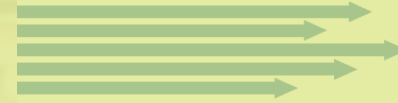




Building the Business Case
for DataOps

A man with glasses and a beard, wearing a blue shirt, is high-fiving a woman with dark hair, wearing a black and white striped shirt. They are in an office setting with a large window in the background. A green circular graphic is overlaid on the left side of the image.

What we'll cover



- Background
- DataOps ROI
 - Production ROI
 - Development ROI
 - Productivity ROI
 - Customer Examples
- Selling to Stakeholders



Background

Successful Analytics ROI = Adopted Analytics



Successful Analytics Is Hard

Data Analytic Projects



The results are in...



60% fail altogether – Gartner



30% have more than 11 errors per month – DataKitchen



87% of data science projects never get to production – Gartner



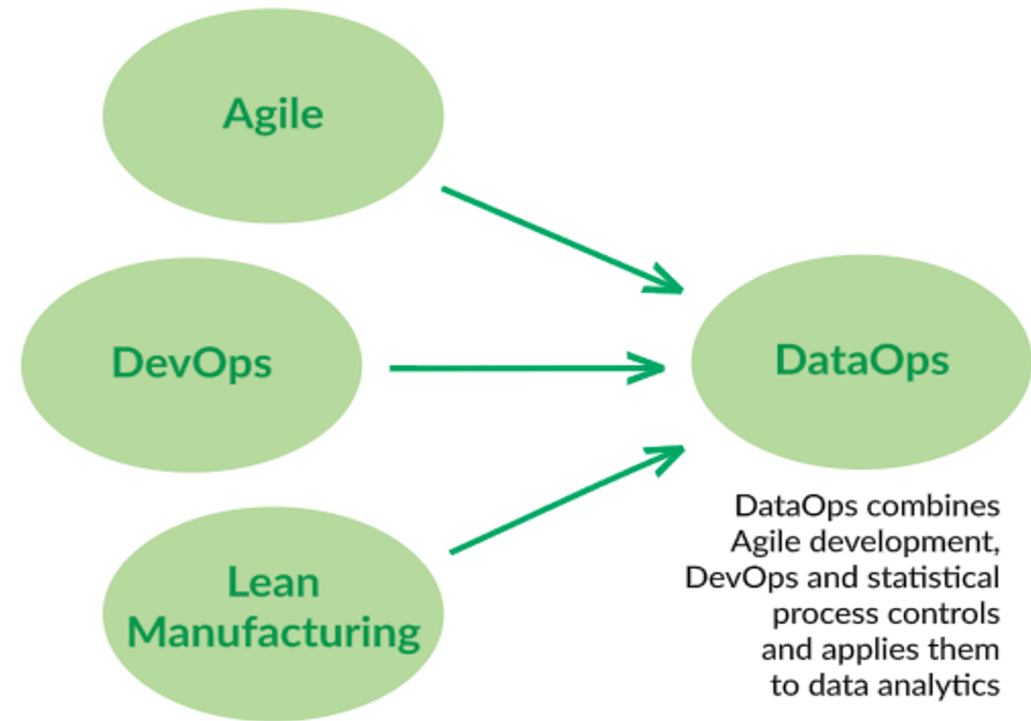
79% Teams want weekly therapy!

What's needed: DataKitchen DataOps Process Hub

Align People, Process & Technology

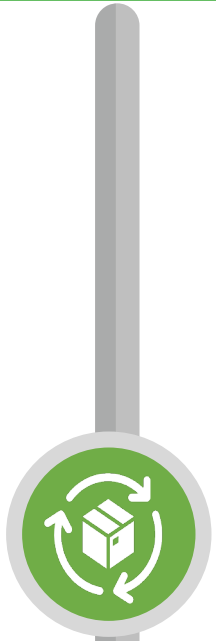
- Fewer error rates in production
- Faster (& safer) deployment cycle time
- Improve Productivity
- Measured Success

... Fast, easy, using existing tools



ROI Before DataOps

Current "Levels"



CYCLE
TIME



ERROR-FREE
DAYS



TEAM
PRODUCTIVITY



PROCESS
MEASUREMENT

Poor Results



COST &
PRODUCTIVITY



UNHAPPY
CUSTOMERS

ROI After DataOps

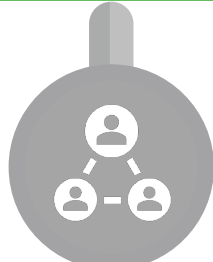
'Levels' After DataOps



CYCLE
TIME



ERROR-FREE
DAYS



TEAM
PRODUCTIVITY



PROCESS
MEASUREMENT

ROI Results



COST &
PRODUCTIVITY



UNHAPPY
CUSTOMERS



DataOps ROI

- Background

- DataOps ROI

- Deployment ROI
- Production Error ROI
- Productivity ROI
- Customer Examples

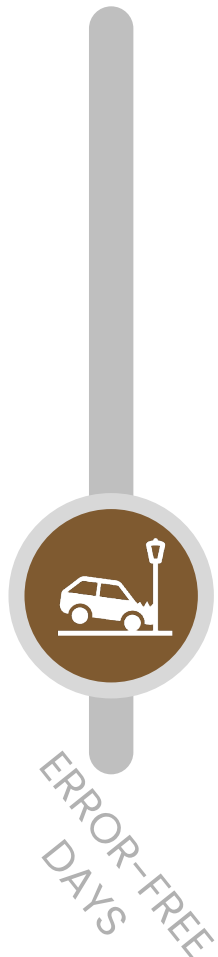
- Selling the DataOps Business Case

- Conclusion

A photograph of three office workers in a modern setting. A man in a grey shirt is seated at a desk, typing on a keyboard. A woman in a yellow sweater stands behind him, holding a tablet and looking at the screen. Another man is partially visible on the left, looking towards the others. A large, semi-transparent yellow rectangle is overlaid on the right side of the image, containing the title text. A faint circular graphic is visible on the left side of the image.

Production Error ROI

Production Data Errors Are Costly

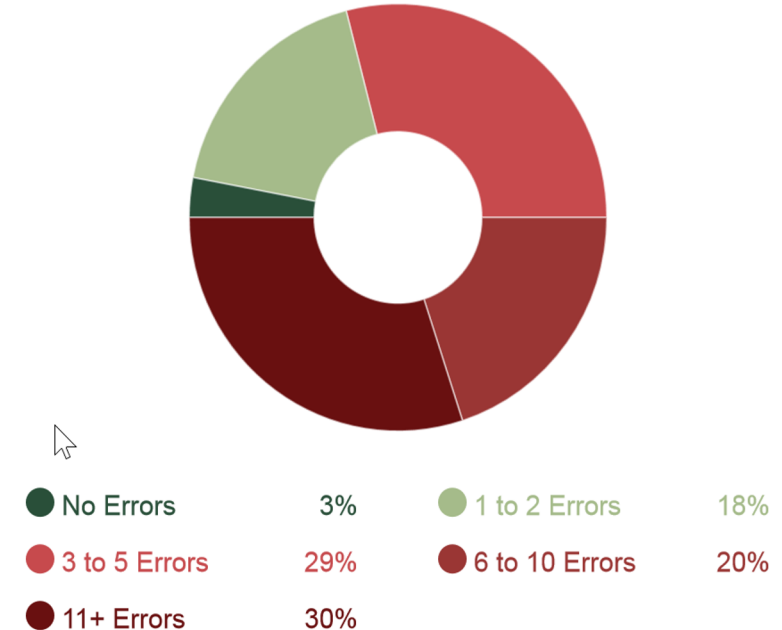


Data Errors Today

- Companies have staggering error rates
- Most companies have way too many errors per month – 79%
- Every data error is costly
- Reduces trust in the data and your team

Error-free days are all too rare!!

On average, how many errors (e.g., incorrect data, broken reports, late delivery, customer complaints) do you have each month?



– Source: Eckerson/DataKitchen



Focus on Errors & the Production Cycle

Before DataKitchen

Errors &
Operational Tasks

Process Improvements
& Tech Debt Reduction

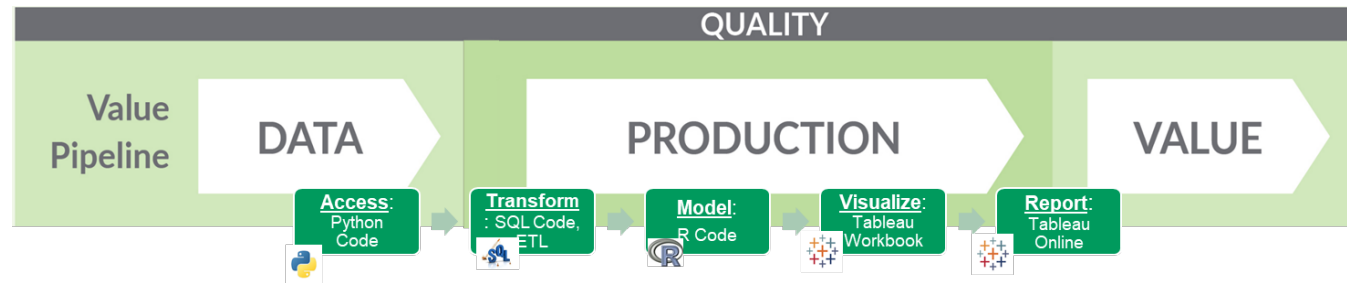
New Features &
Data for Customers



- Only 22% of time is spent on innovation; 78% on errors & manual execution. – Gartner (2020)
- 52% of Data Engineers said errors are a major source of burn out. – DataKitchen & data.world (2022)

Production Testing & Monitoring

Lower Error Rates & Embarrassment!



[HOW] Test & Monitor:

- Automatically
- On top of the tool chain
- Send alerts / notification
- Keep track of history
- Make it easy to create tests

[WHAT] Test Types:

- Traditional Data Quality
- Statistical Process Control
- Location Balance Test
- Historic Balance Test
- Business Based Tests

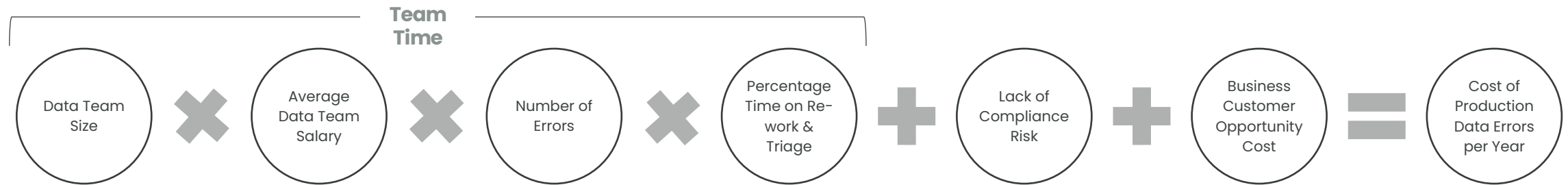
[WHY] Benefits:

- ↓ Less Errors
- ↑ More Innovation
- ↑ More Customer Data Trust
- ↓ Less Stress
- ↓ Less Embarrassment



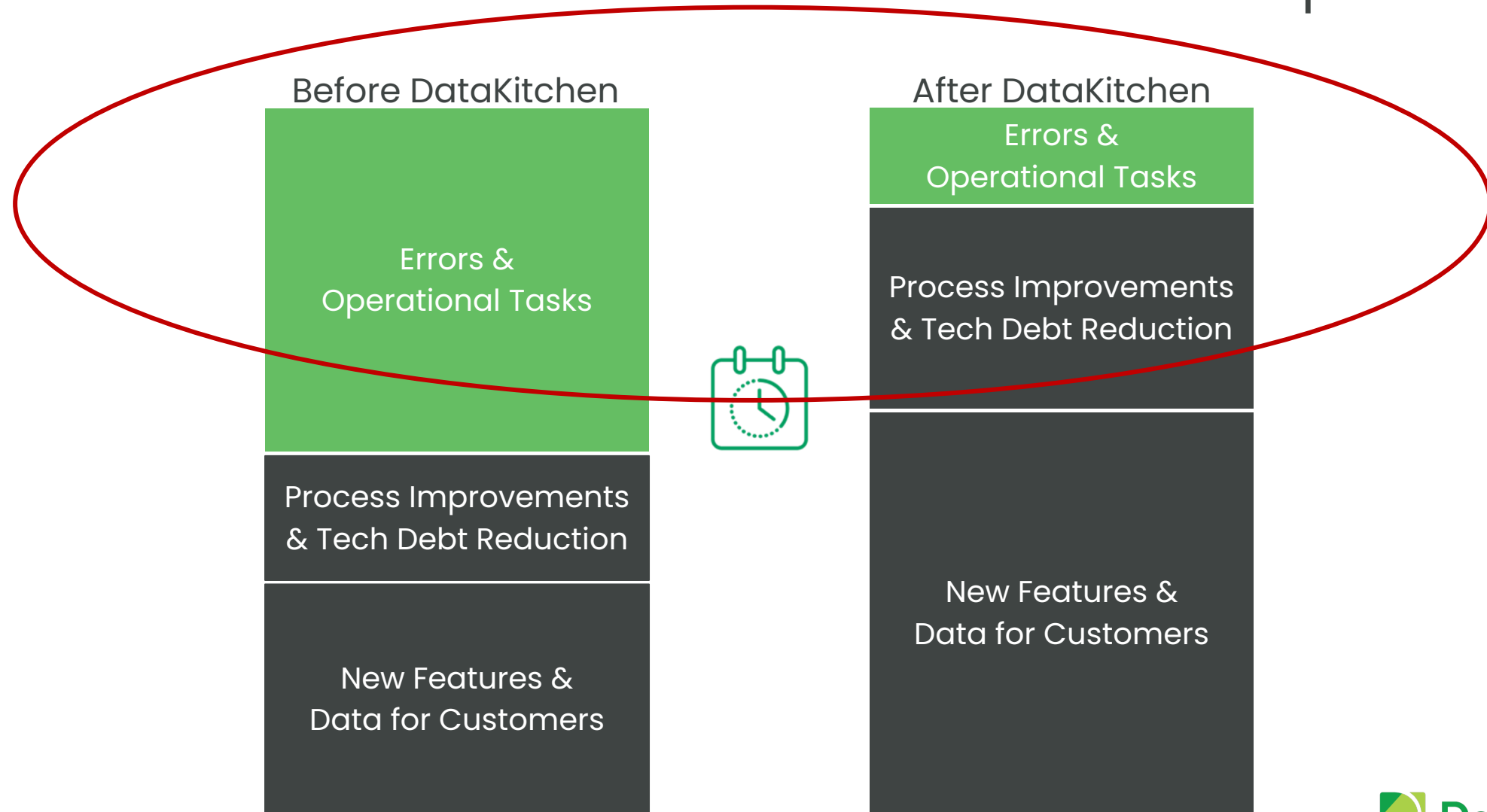
Production Data Errors Are Costly

Team Time, Compliance & Opportunity Costs

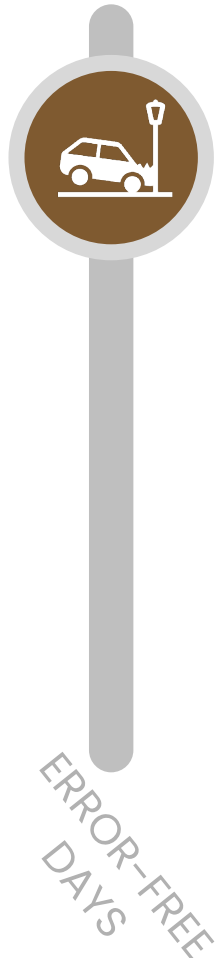


- Wrong data, or wrong reports/models are very costly.
- Business users lose trust in the data & have an opportunity cost.
- Data errors can cause compliance risk.
- Data Team spend lots of time on find/fix/triage/repair of errors – costly.

Increase Innovation Time with DataOps



Production Data Errors Reduce ROI



Problem

- With Significant Data Errors in Production
- Team Productivity Suffers
- Customer Data Trust Suffers
- Willingness to Innovate Suffers

Solution

- **Connect to your existing pipelines quickly & seamlessly** without required modification to your pipelines,
- **Test & monitor your data at-rest** that does not require extracting the data from where it is currently stored.
- **Require minimal configuration**
- **Provide rich context** that enables rapid triage & alerting
- **Prevent issues from happening in the first place !**

A man and a woman in business attire are standing in a modern office, looking at a tablet. The man is on the left, wearing a dark suit and a blue shirt. The woman is on the right, wearing a dark suit and glasses, holding the tablet. They are both smiling. A large green rounded rectangle is overlaid on the right side of the image, containing the text "Deployment ROI".

Deployment ROI

Deployment Cycle Time ROI Challenges



Challenge

- Companies take weeks or months to get new ETL, schema, models, visualizations & governance into production.
- Slow deployment or risky deployment

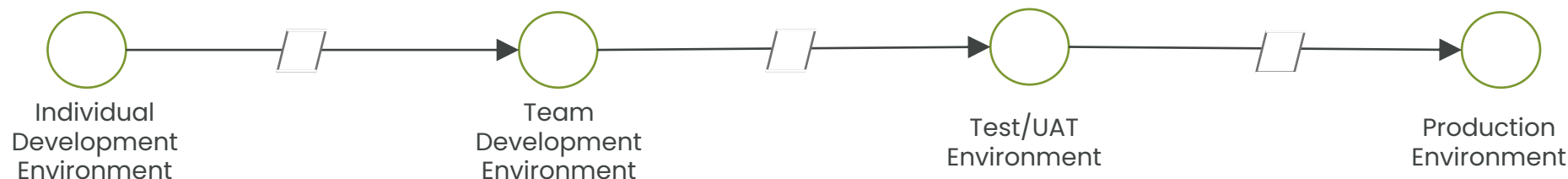
How?

- They move work manually
- Or do it with CI/CD

Need

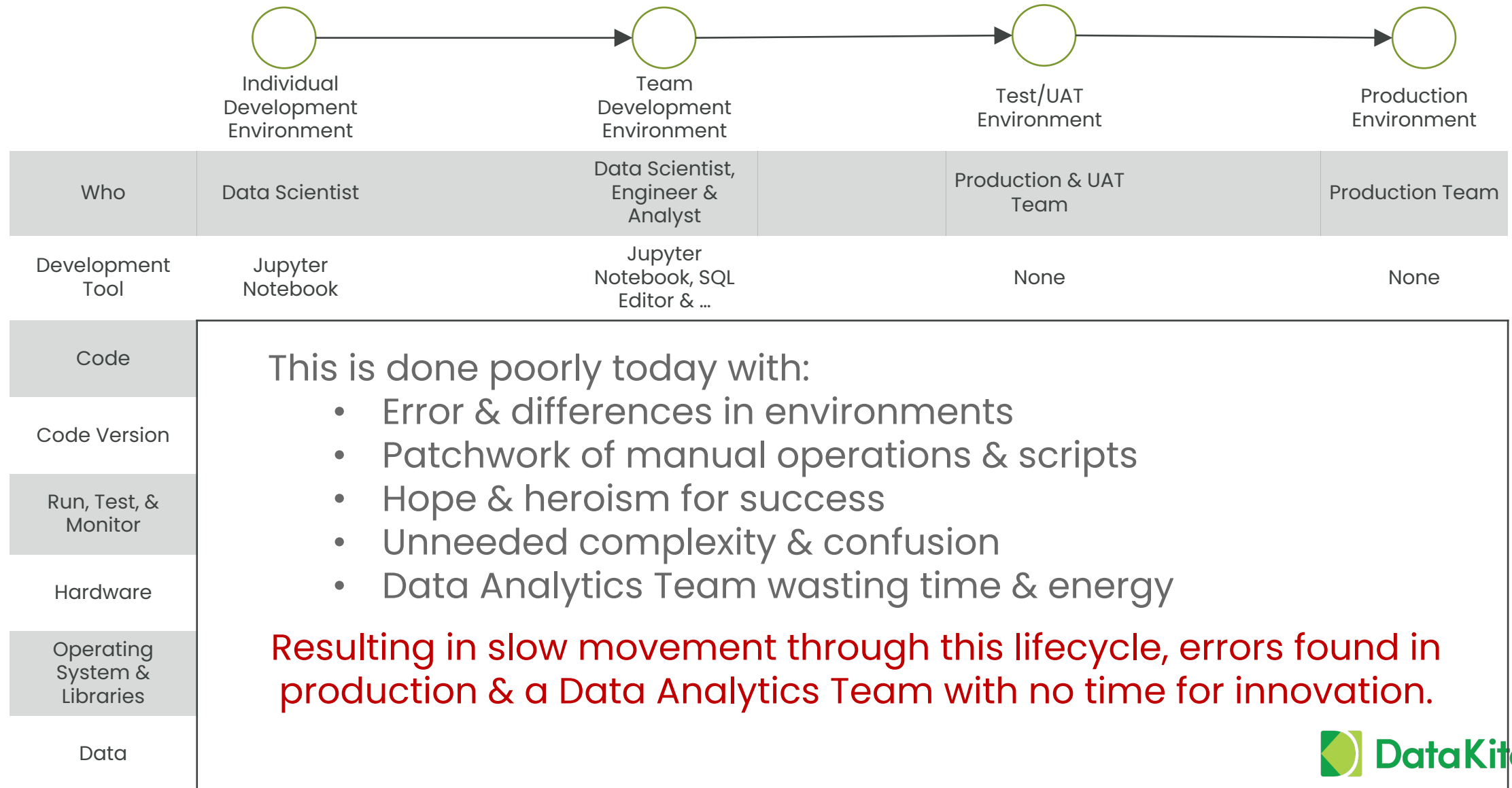
- But few have lowered the risk through a complete set of functional & regression test based on up-to-date & realistic test data
- Fast & low risk deployment

Data Development Lifecycle Complexities



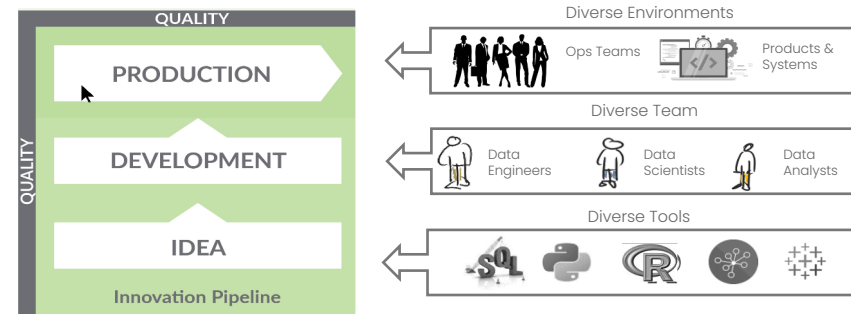
Who	Data Scientist	Data Scientist, Engineer & Analyst		Production & UAT Team	Production Team
Development Tool	Jupyter Notebook	Jupyter Notebook, SQL Editor & ...		None	None
Code	Python	Python, SQL, & others		Python, SQL, & others	Python, SQL, & others
Code Version	Version 1	Version 2 & 3		Version 3	Version 3
Run, Test, & Monitor	Ad hoc run: check unit tests	Ad hoc run: check unit, functional tests		Ad hoc run: check functional & monitor tests	Scheduled run: check monitor tests
Hardware	Laptop & test data	Development Servers & DB		Test Servers & DB	Production Servers & DB
Operating System & Libraries	Windows + libraries	Unix + Various libraries		Unix + Various libraries	Unix + Various libraries
Data	Development Data	Development Data		Copy Production Data	Production Data

Data Analytic Development Lifecycle



Best Practices in Deployment DataOps

Iterate to Innovate & Deliver Customer Value



[HOW] Automate Test,

- Automatically
- Pull pain forward
- Automate deployment
- Make it easy to create tests

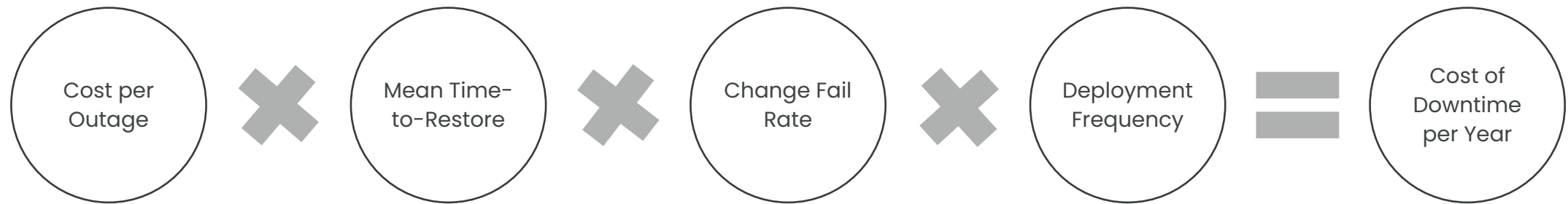
[WHAT] Dev Practices

- Code based unit test insufficient
- Test data to test your code
- Reuse production tests for regression, functional & performance tests

[WHY] Benefits

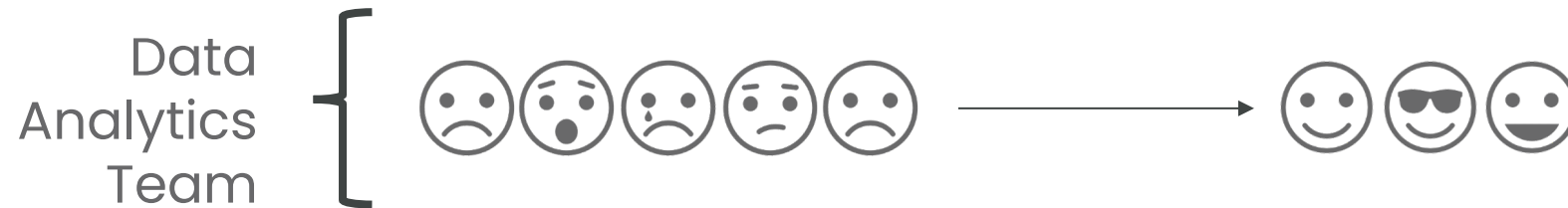
- ↓ Regression errors
- ↑ Faster deployment
- ↑ Remove individual blinders
- ↓ Lower risk of change
- ↓ Less customer eye rolls

Deployment Failure is Costly



- Deploy failures are very costly
- If they fail without an outage, you are lucky but just take your team's time
- If a deploy fails with a data error or down time, then you need to calculate the cost of not having your data system available to be used

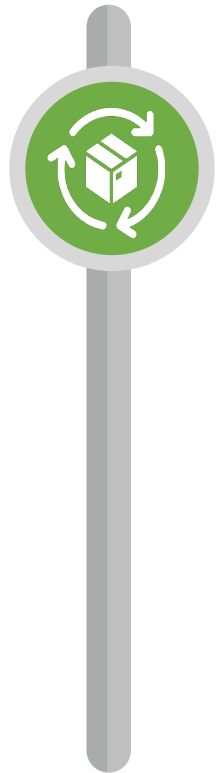
DataOps ROI: Faster, Better & Happier



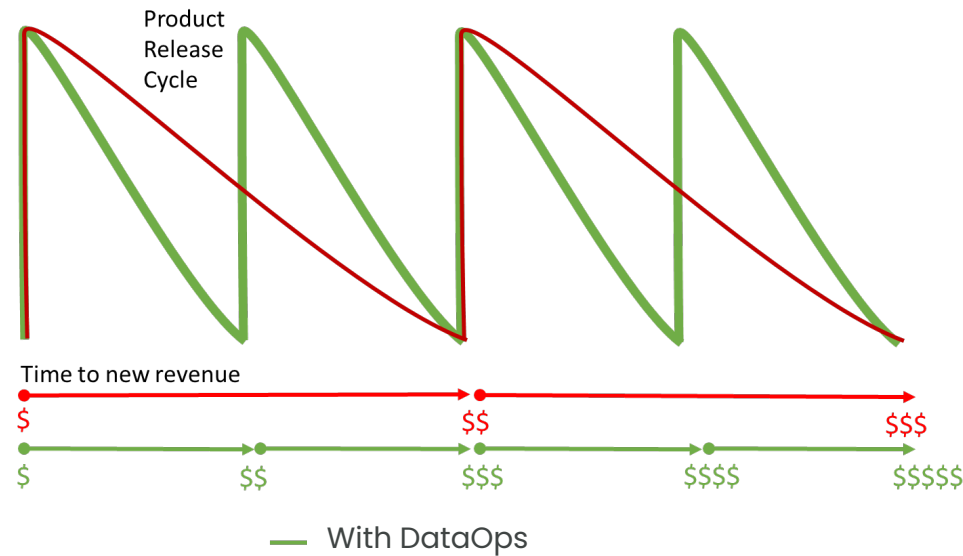
Before DataKitchen

After DataKitchen

Deployment Cycle Time ROI



CYCLE
TIME

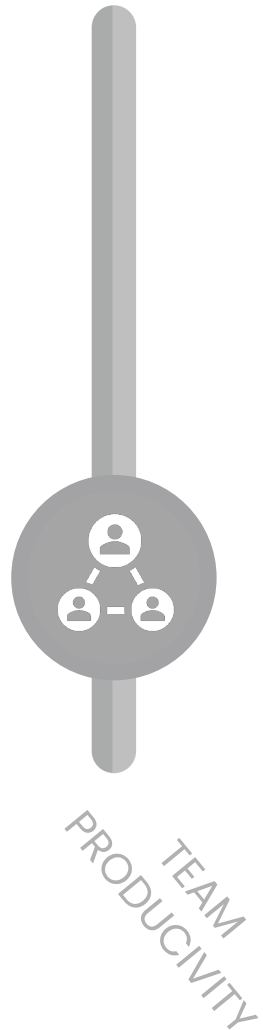


- Faster releases of analytics products
- Increased learning of your team
- Better insight
- Faster & more value for your data analytics



Productivity ROI

Data Team Productivity Challenges



Challenge

- Data Analytics Teams are not very efficient or effective
- Lots of re-work, delays, excess meetings & context switching make for a slow, frustrated workforce

How?

- They work in isolation from other team members – head in the sand
- They 'throw their work over the wall & hope it works. Then rush to fix things when they break.

Need

- Need to spend more time 'on keyboard' & doing actual work
- Have a system that allows them to make changes quickly & judge the impact (i.e., regressions) of those changes downstream

Lots of Sources of Wasted Time

Waste	Bug fixing & re-work	5%
	Incident response	5%
	Waiting/Blocked/Handoffs	2%
	Over-engineering	2%
	Extra effort for Sub-optimal solutions or workarounds	2%
	Relearning (e.g. lack of documentation or SME support)	3%
	Un-implemented work	2%
	Unnecessary meetings	2%
Unplanned Activity	Unplanned chats	2%
	Unplanned but value-add	11%
Core Management	1:1s	10%
	Team meetings	5%
	Company meetings	3%
	Other Admin tasks	10%

- From a customer survey of their data team
- Over 60% of their time is not doing value-added work.
- Look at the cost of all the overhead your team is doing!

Value – Key Question



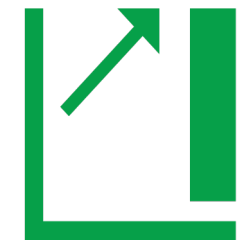
Current Data & Analytics Staffing Levels



Amount of
Insight Produced

Total amount of Data & Analytics Created
Insight (dashboards, PowerPoints, reports,
etc.)

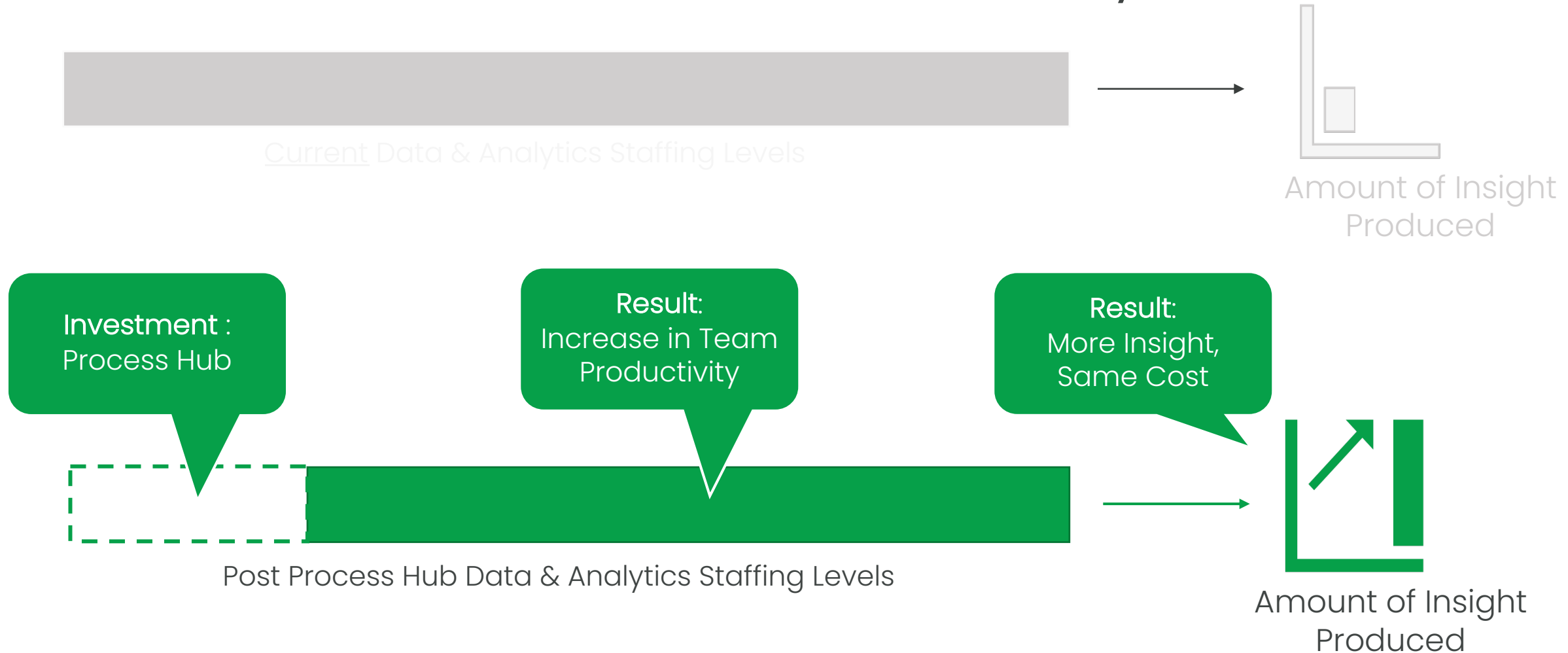
Key Question:
*How to increase the total amount of
Data & Analytics insight generated
without continually adding more staff
(& cost)?*



Amount of
Insight Produced



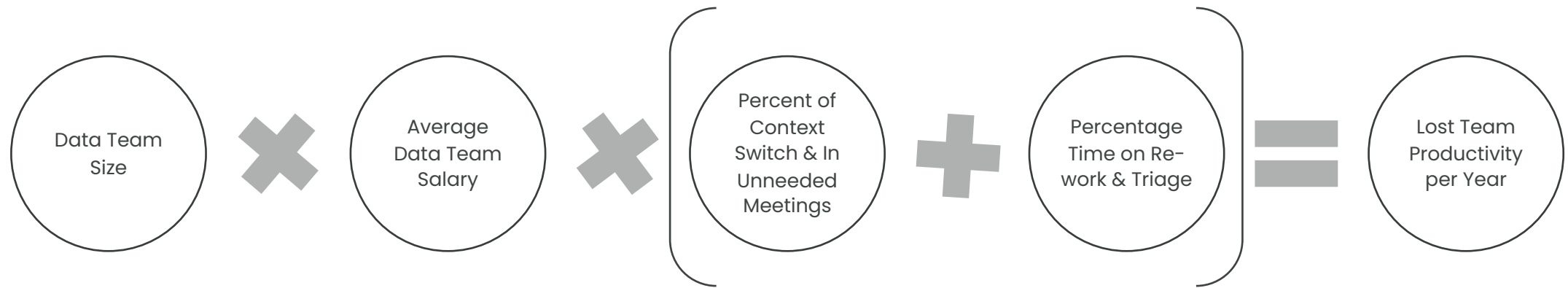
Process Hub Increases Productivity



DataKitchen ROI: Efficiency Ratio

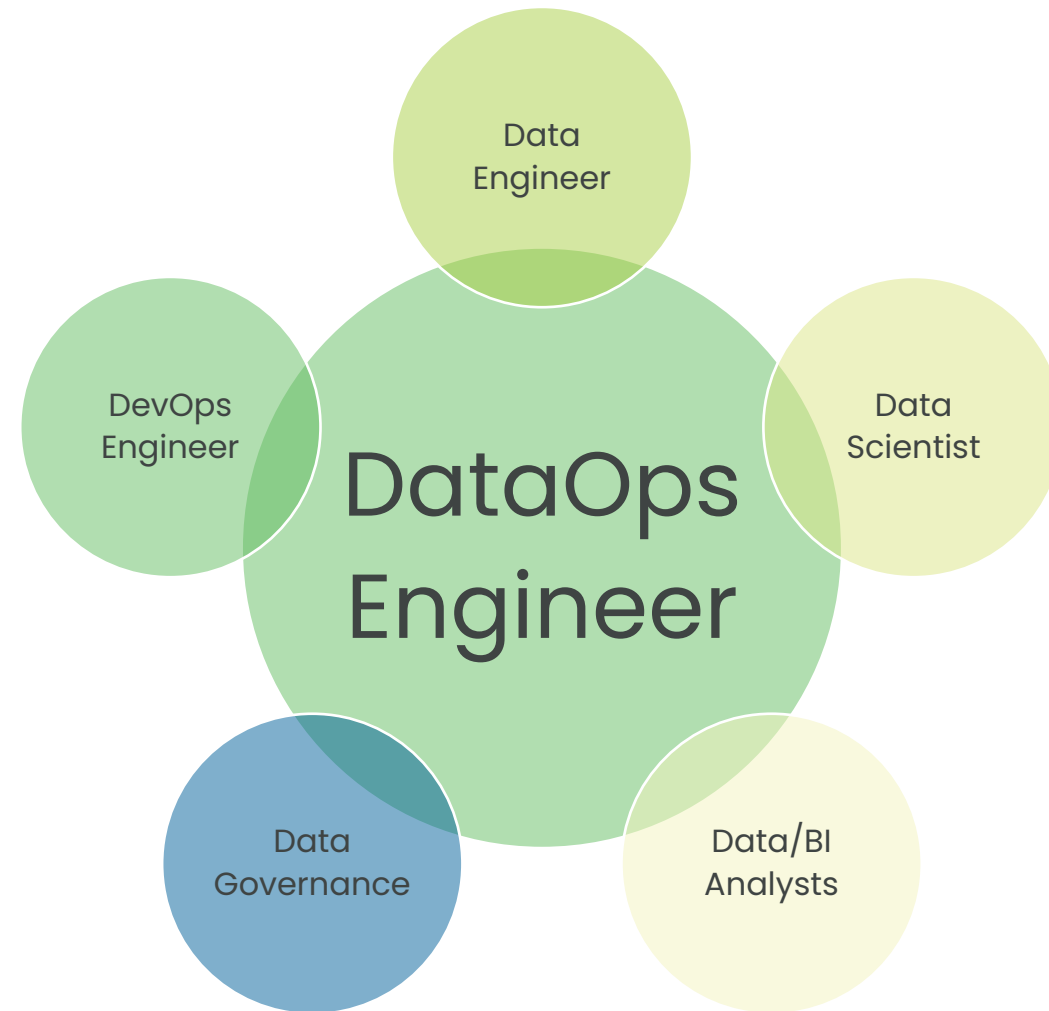
- The number of Data Engineers needed to support each Data Analyst or Data Scientist ranges from 1:1 to 12:1 (depending on your source).
 - In our experience with DataKitchen the ratio is flipped
 - 1 Data Engineer can support 12 Data Analysts.
- Therefore, you will see a **12x to 144x increase in efficiency** using DataKitchen.
- Even if you are *super* conservative & say that there is just a 20% improvement, with 10 people using the system you save the equivalent 2 people which is more than the total cost of the system

Poor Data Team Productivity Is Costly



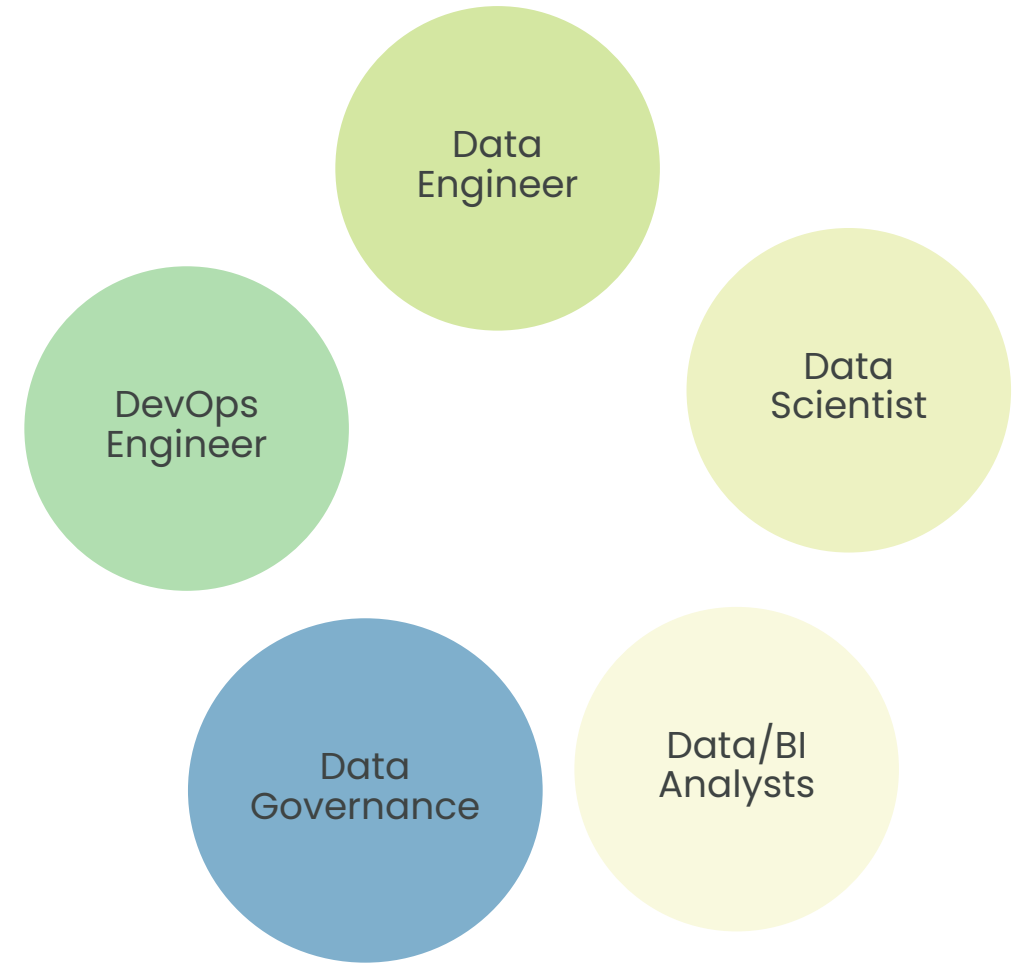
- Lack of Time Spent on Development Task is cost
- Time in Unneeded Meetings, Context Switching, Waiting & Re-work are all costly

DataOps Engineering Is About Productivity

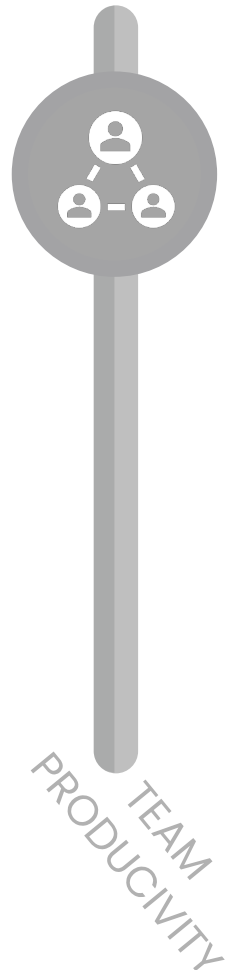


Current Challenges with Productivity

- Throw it over into prod?
- Definition of done means 'it worked for me'
- I only focus on my little part
- Someone else's problem
- Willful blindness
- Task-focused (not value-focused)
- Project-focused (not product-focused)
- Hope that it works
- Reliance on manual checks
- Lack of willingness to 'pull the pain forward'



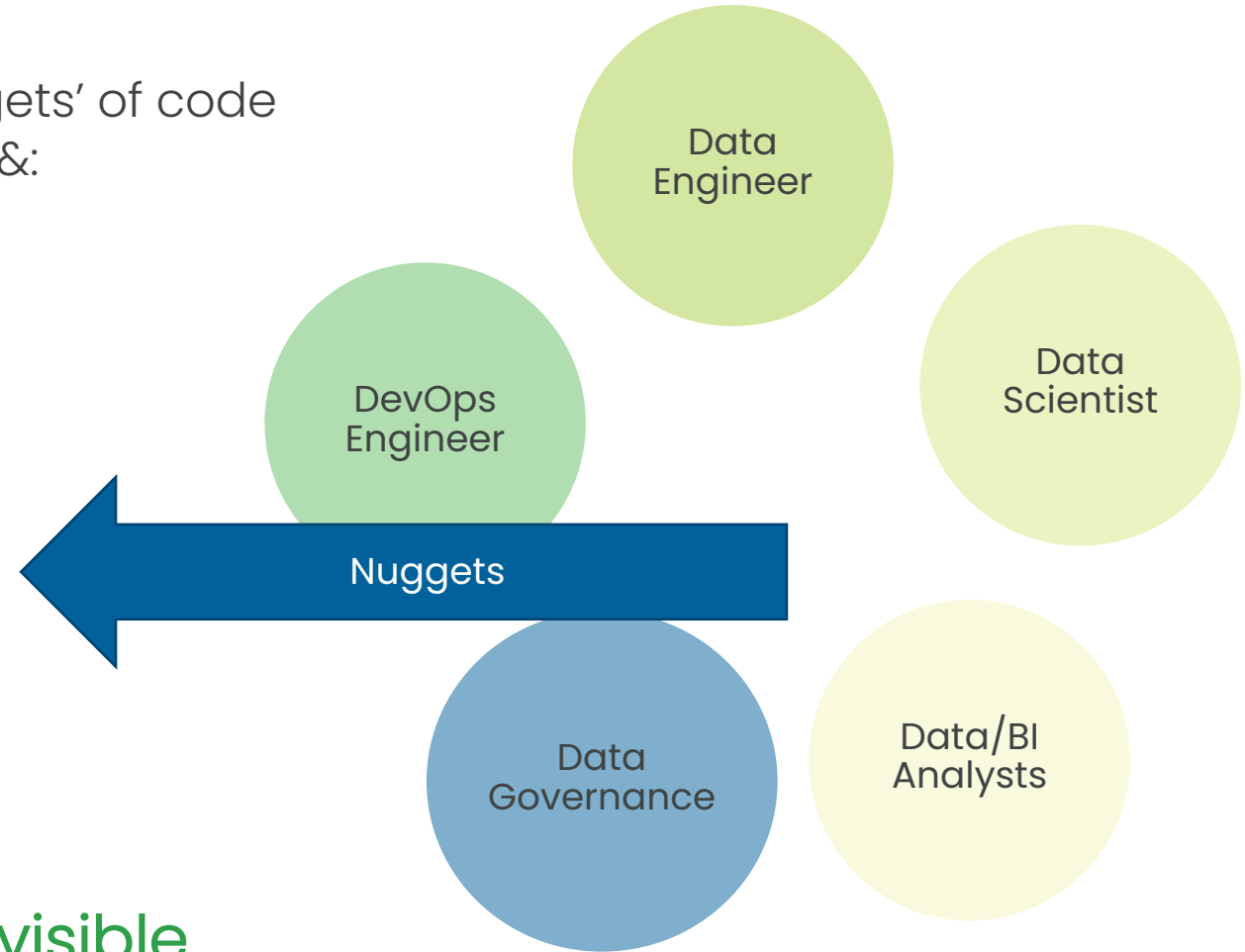
DataOps Engineering Is About Productivity Through Shared Abstraction



DataOps Engineers take 'nuggets' of code (i.e., ETL, SQL, python, xml, etc.) &:

- Put into pipelines
- Create tests
- Run the factory
- Automate deploys
- Work across people
- Measure success
- Enable DataOps self services

Make what is invisible, visible.

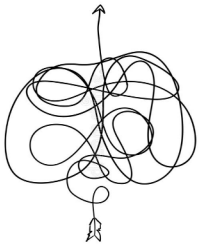
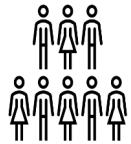


A man and a woman in business attire are standing in a modern office, looking at a tablet. The man is on the left, wearing a dark suit and a blue shirt. The woman is on the right, wearing a dark suit and glasses, holding a tablet. They are both smiling. A large green rounded rectangle is overlaid on the right side of the image, containing the text "Customer ROI Example #1".

Customer ROI Example #1

Business Analytics ROI – The Problem

Business
Analytics
Analysts



Base Data
(Data Lake/
Warehouse)

- Business Analytic Teams MUST deliver value to their customer, make rapid changes & have ongoing, updated production deliverables
- Lots of data tables from IT & diverse data cause Business Analytics (BA) Analysts to do a **huge amount of fragile, error prone & high effort data work** to get insight
- Repetitive data processes & query work **reduces the time the Business Analytics Team needs create original insight**
- Inconsistencies in how the Analytics Team applies business data process & business logic cause errors. **Business Analytics firefighting & stress are common.**
- Lack of a tailor-made data representation, up-to-date documentation, & reusable business logic cause Analysts extra time to do ad-hoc analysis & make training new Business Analytics Analysts difficult.
- New insight requires manual time. Reuse is low & costs are high.
- So, you can either continuously throw BA staff at the problem or work with IT & take many months to build something that is costly to change & out of date on arrival.

3 Ways a Process Hub Increases Productivity

DataKitchen Process Hub software plus staffing Analytic Engineering makes the Business Analytics staff much more efficient in three ways:



'The Wedge': Trusted Analyst-ready **data sets**, under Business Analytics control, with rapid changes, understandability, enables ease of answering questions with high reuse.

Benefit: lowers the cost-per-question & Business Analytics productivity rises



'The Hammer': Full **automation**, from source to value delivery enabling fully zero touch end to end ongoing insight delivery.

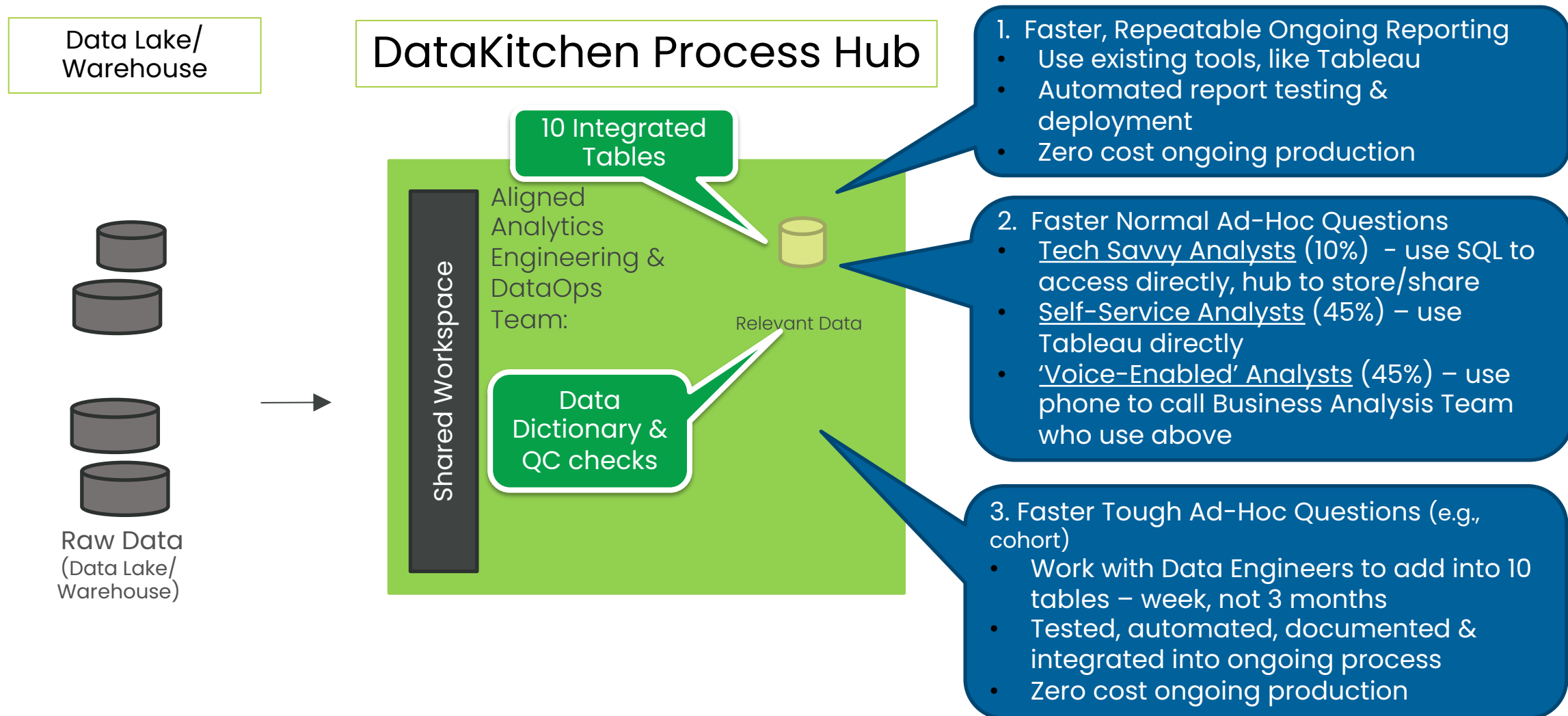
Benefit: reduces recurring work & Business Analytics get more insight generating staff



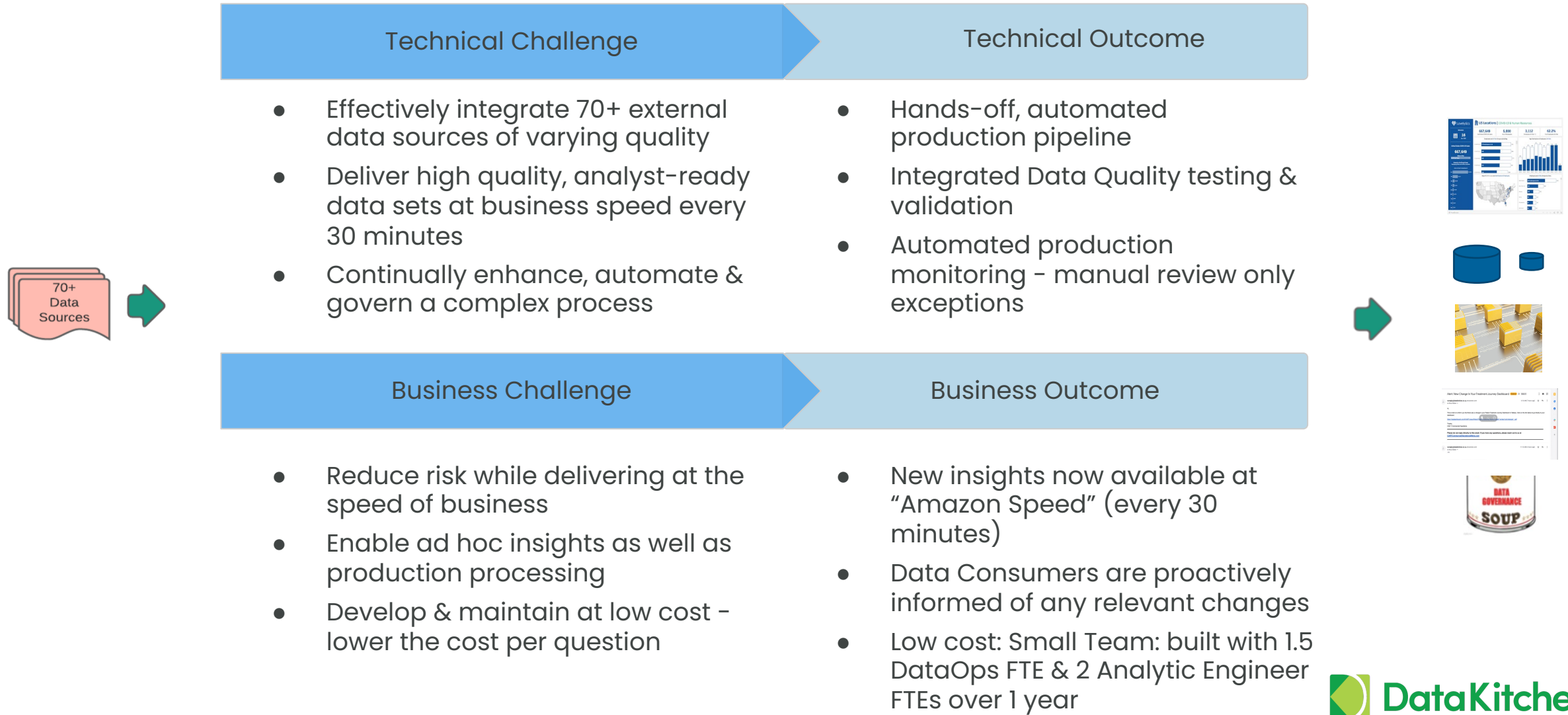
'The Store': A centralized **repository** under version control for reuse, sharing, best practices, storing of all Business Analytics intellectual property.

Benefit: lowers risks & redundancies & Business Analytics productivity rises

Benefits of the Wedge



The Hammer – Fast, Flexible & Fully Automated



The Store – Eliminates Hidden Code, Tribal Knowledge

```
--All diagnosed with new CC-486 MB dx codes
WITH all_aml_diagnosed AS
(SELECT patient_id, min(service_date) service_date
 FROM (SELECT patient_id, service_date
        FROM ptd_current.raw_celgene_onc_dx_fact a
        LEFT JOIN ptd_current.raw_celgene_onc_diagnosis_code b
              ON a.diagnosis_id = b.diagnosis_id
        WHERE upper(b.diagnosis_code) IN ('205.0',
        '205.00', '205.01', '205.02', '205.8', '205.80', '205.81',
        '205.82', '205.9', '205.90', '205.91', '205.92', 'C92.00',
        'C92.01', 'C92.02', 'C92.50', 'C92.51', 'C92.52', 'C92.60',
        'C92.61', 'C92.62', 'C92.A0', 'C92.A1', 'C92.A2', 'C92.Z0',
        'C92.Z1', 'C92.Z2'))
 UNION
 SELECT patient_id, service_date
 FROM ptd_weekly_current.raw_ptd_dx_fact
 WHERE upper(diagnosis_code) IN ('205.0',
        '205.00', '205.01', '205.02', '205.8', '205.80', '205.81',
        '205.82', '205.9', '205.90', '205.91', '205.92', 'C92.00',
        'C92.01', 'C92.02', 'C92.50', 'C92.51', 'C92.52', 'C92.60',
        'C92.61', 'C92.62', 'C92.A0', 'C92.A1', 'C92.A2', 'C92.Z0',
        'C92.Z1', 'C92.Z2'))
 )
GROUP BY 1 ),
```

Business Analytics Tribal Knowledge =
High Costs & Slow Speed to Insight

Code from a vendor's laptop

- What is it?
- Where is it used?
- Who wrote it?
- What is its history?
- Is it tested?
- How do I share it?
- How do I find it?
- Can anyone see it & change it?
- Will it be lost?
- Could I just use the resulting data?
- Are these codes up to date?

One Customer's Transformation



Bristol-Myers Squibb



Schema Changes per
Data Engineer per week

Before DataKitchen

1

With DataKitchen

12



Cycle time to publish
new visualizations

weeks/months

next day



Data Analysts supported
by **ONE** Data Engineer

0.5

12



Number of automated
tests each build

0

1,000



Number of errors
each build

frequent

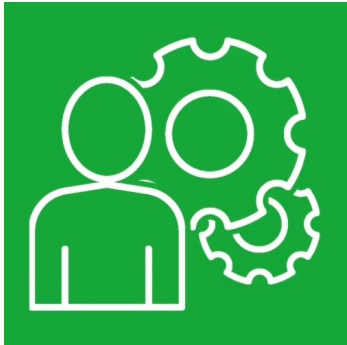
0





Customer ROI Example #2

Increase Productivity & Trust



“The DataKitchen Platform enables our team to easily bake new tests into our pipelines as we learn more over time. When we first implemented DataOps, **we reduced errors to just about one per quarter.**

We kept adding tests over time, & now **it's been several years since we've had any major glitch.**

This has **dramatically increased** the efficiency of our team & the **confidence of our end stakeholders** in the data.”

– Rajesh Gill, Former Celgene Sr Manager of Business Analytics, Forecasting & DataOps Strategy

Faster, Better & Happier

	Before DataKitchen	After DataKitchen
Deployment Latency	 weeks/months	 hours/minutes
Production Errors		
Data Analytics Team		

Celgene Example

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



Selling the Business Case to Internal Stakeholders

Just Get Started (...if you can)

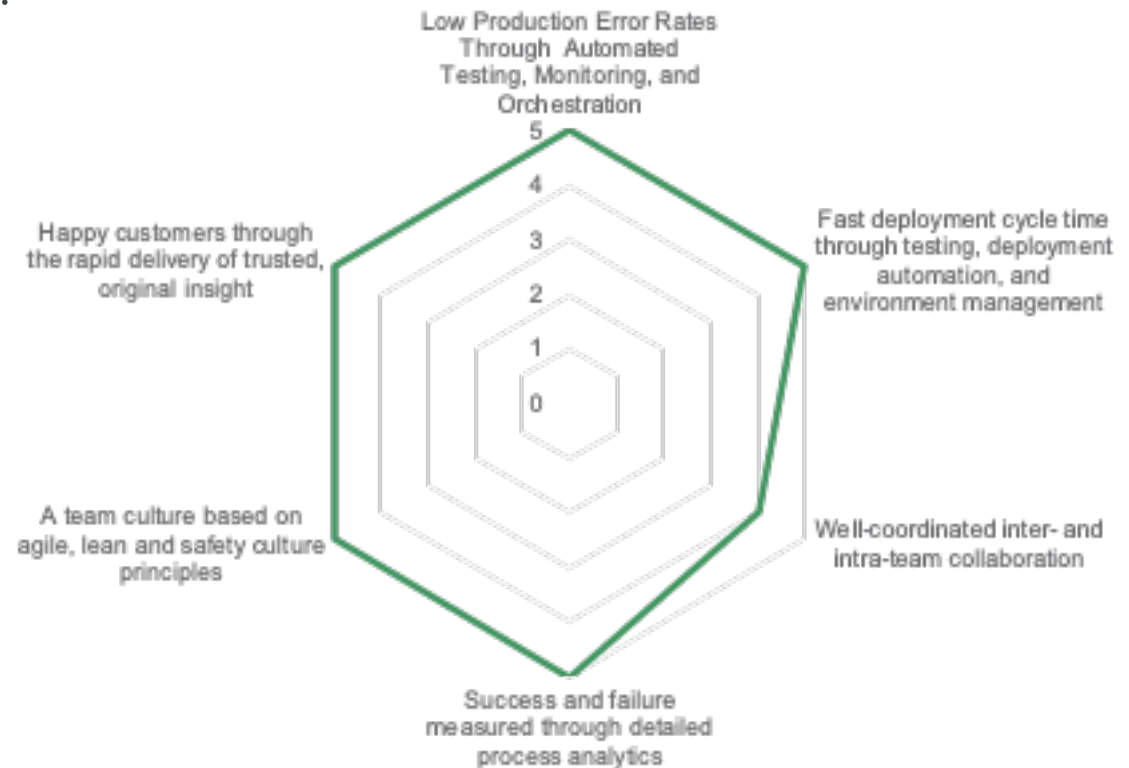
- Do not sell the details of DataOps. Just get started & show results.
- Most stakeholders don't care how the sausage is made.
- Fix low-hanging fruit for fast wins.
 - Publish metrics so all are accountable.
- Consider:
 - Do you have control over the 'data factory'?
 - Can you implement small changes without senior leadership buy-in?



Find Key Projects with Maturity Model

A DataOps Maturity Model Assessment:

- Assesses organizational readiness & provides a roadmap for DataOps success
- Identifies weaknesses across key operational areas
 - Error rates
 - Cycle time
 - Collaboration
 - Team culture
 - Measurement
 - Customer happiness

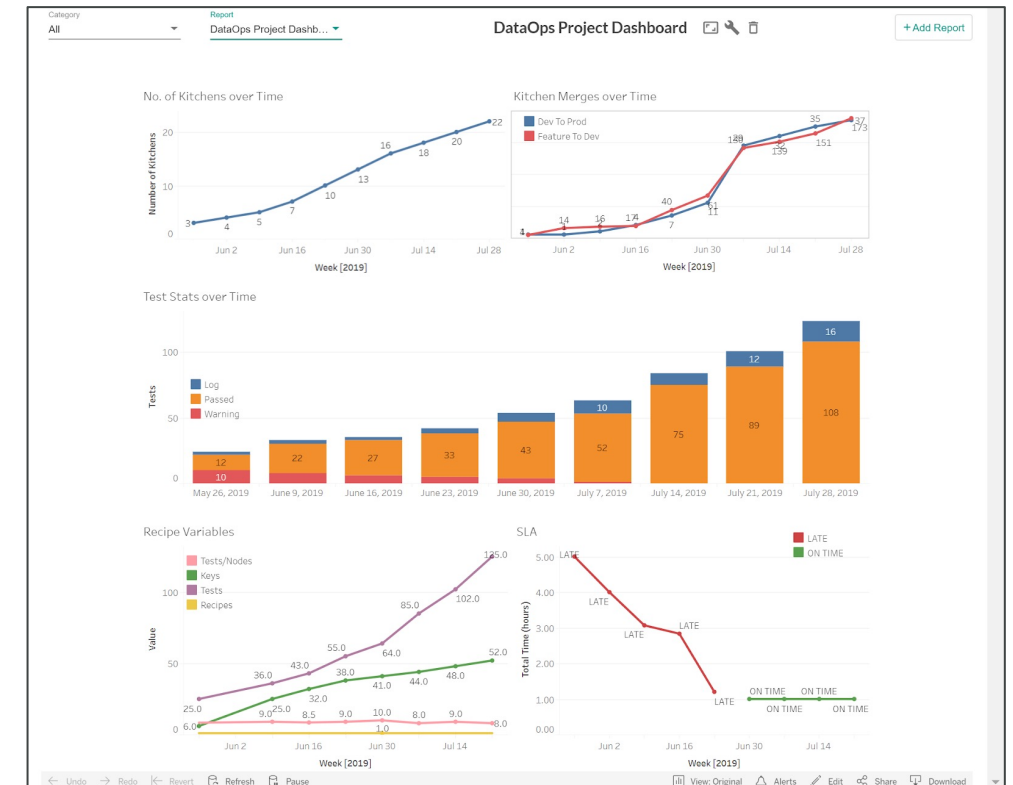


Take the DataOps [Maturity Model Assessment](#)



Establish a Baseline

- Measure your team's activities
 - Productivity of individual contributors (fixing errors, meetings, value-add)
 - Cycle time, error rates, incidents
- Time spent on waste is often shocking
- Give reporting time to stabilize
- Set target goals & track them quarterly



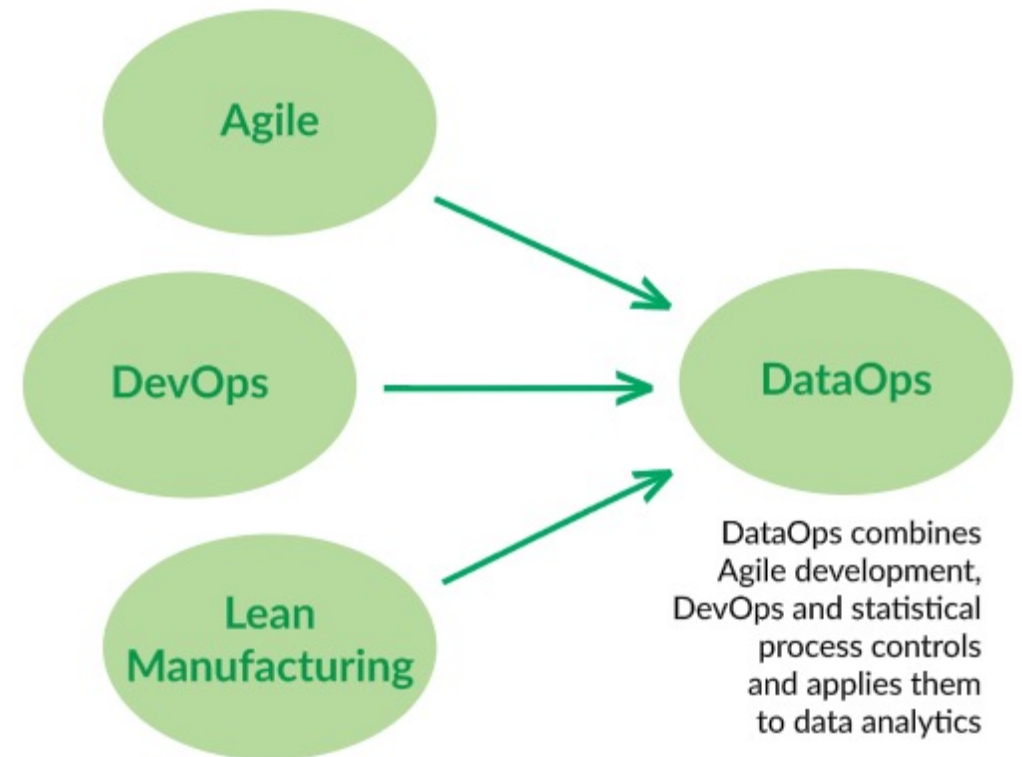
Demystify the Data & Analytics Factory

- Be prepared to give a tour to those who care
- Share what can (and will) go wrong
- Involve stakeholders in the process
- Support is easier to gain once they understand the problem, solutions & benefits



DataOps is NOT Just DevOps for Data

- DevOps can frame the conversation with IT
- Beware of oversimplifying DataOps' power:
 - DataOps is not just a CI/CD problem
 - DataOps involves complex toolchains & teams
 - Focus on the outcomes of DataOps (vs those of DevOps)



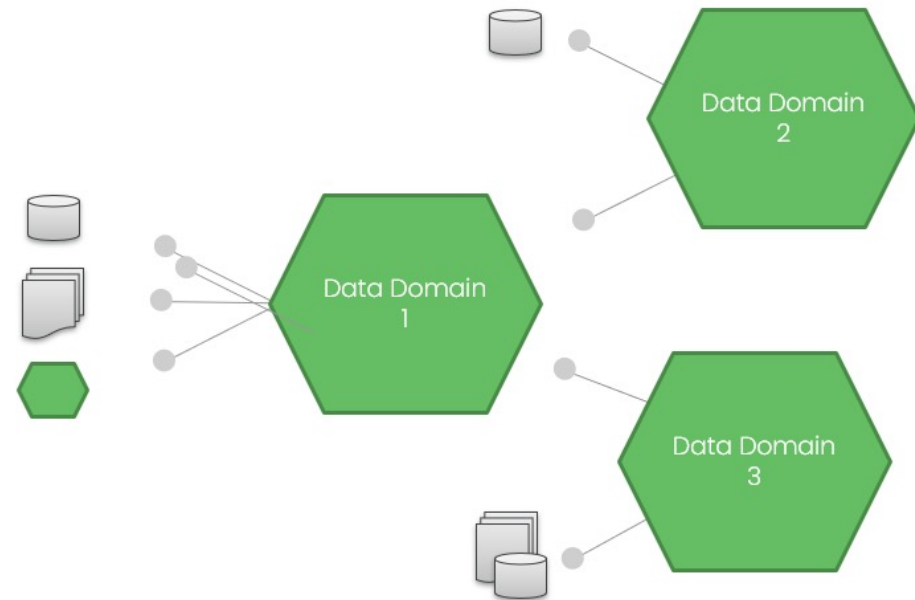
Piggyback DataOps with Other Initiatives

- Cloud Migration?
- Data Mesh?
- Data Fabric?
- Buying Snowflake?

DataOps is central to the success of these projects.

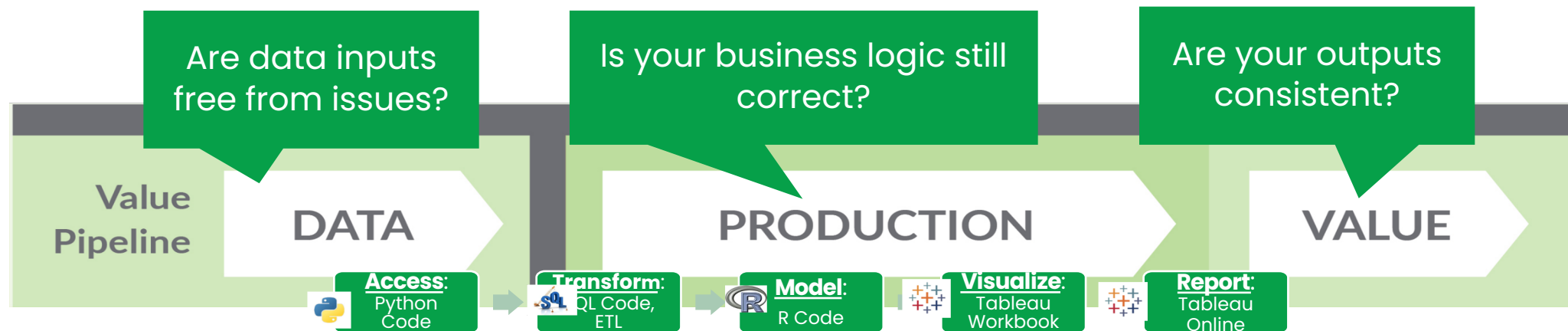
For more reading:

- [Build a Data Mesh Factory with DataOps](#)
- [Your Cloud Migration is Actually an Agility Initiative](#)
- [DataOps Enables Your Data Fabric](#)



You're Not Too Busy to Do It Right

- Start small with Production DataOps; focus on reducing errors & showing impact.
- Don't tolerate work-arounds.
- Give visibility to senior leadership. What & why things are broken.





Conclusion

After DataOps

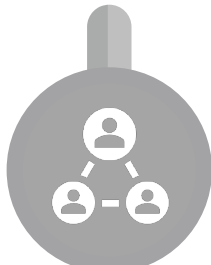
'Levels' After DataOps Transformation



CYCLE
TIME



ERROR-FREE
DAYS



TEAM (S)
PRODUCTIVITY



PROCESS
MEASUREMENT

Transformation Results

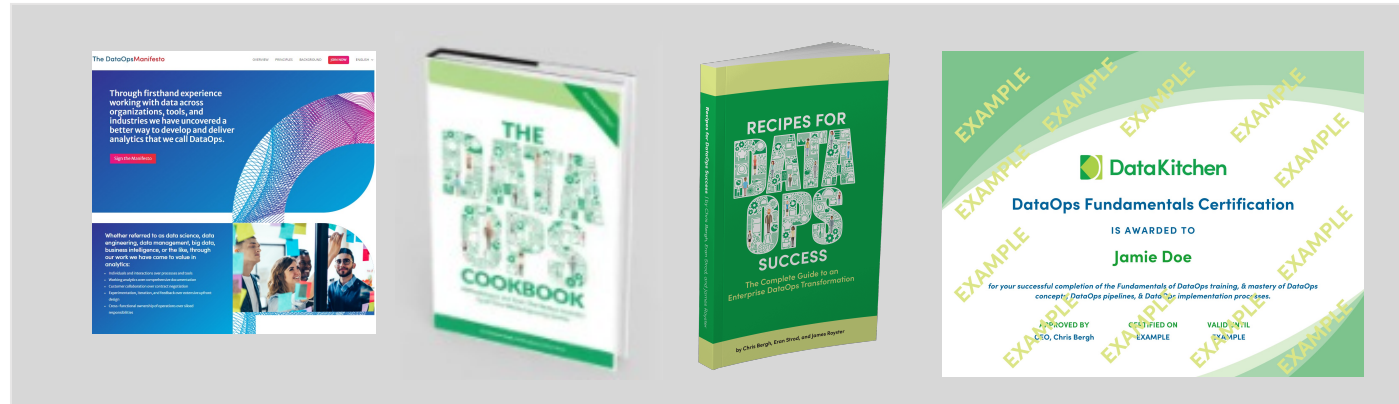


COST &
PRODUCTIVITY



UNHAPPY
CUSTOMERS

Next Steps to Learn More



- Read & Sign [The DataOps Manifesto](#)
- Download your free copies of [The DataOps Cookbook](#) & [Recipes for DataOps Success](#)
- Get Certified in [The Fundamentals of DataOps](#)