



# A teoria à prática na implementação de DataOps e MLOps

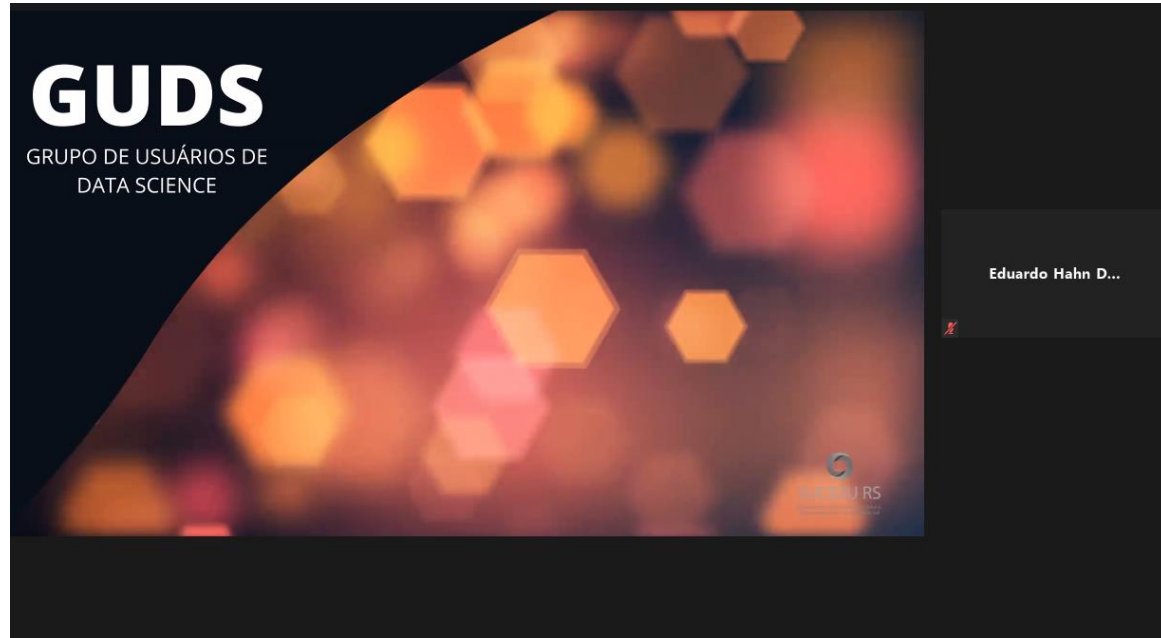
Eduardo Hahn

DataLakers founder & DataOps Enthusiastic

- + BuzzWord
- “**DataOps** is an automated, process-oriented methodology, used by analytic and data teams, to improve the quality and reduce the cycle time of data analytics.” Wikipedia
- “**DataOps** is about more than speed and quality. With a culture of continuous improvement, organizations can deliver data analytics solutions more efficiently, releasing valuable team members for more valuable activities, such as building innovative new products.” Eckerson Group

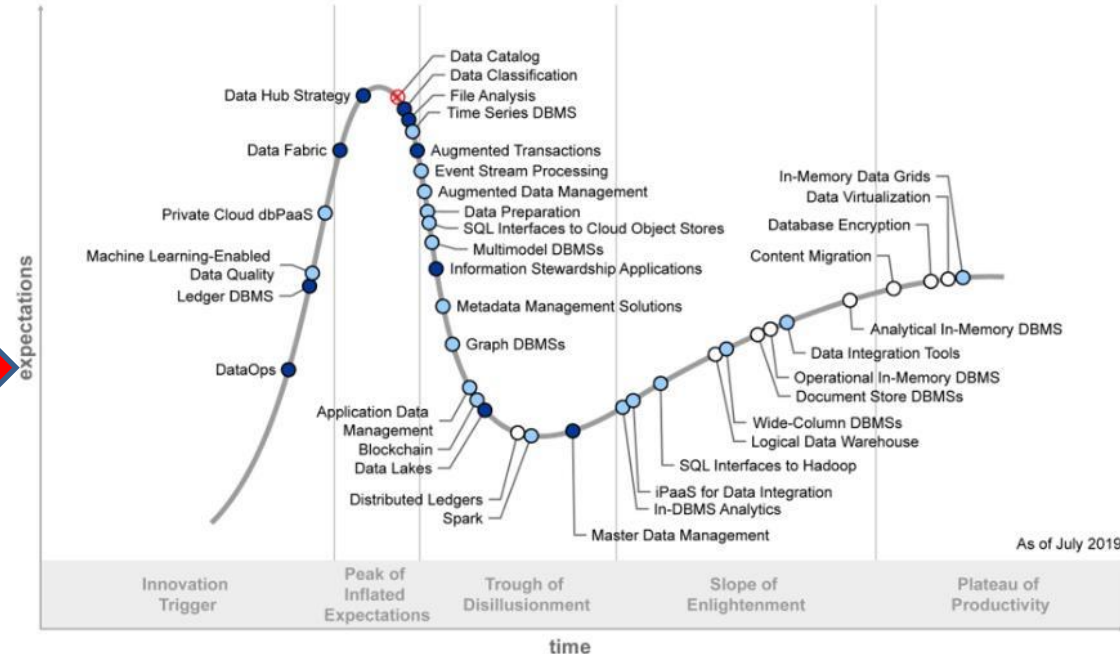
# Onde já ouviu falar sobre DataOps e MLOps?

- GUDS
- Medium
- Meetup
- 1ª vez
- Outro lugar???



# Hype Cycle

## Hype Cycle for Data Management, 2019



As of July 2019

Plateau will be reached:

○ less than 2 years   ● 2 to 5 years   ● 5 to 10 years   ▲ more than 10 years   ✕ obsolete before plateau

Source: Gartner  
ID: 369950

# DataOps Manifesto

## The DataOps Manifesto

Through firsthand experience working with data across organizations, tools, and industries we have uncovered a better way to develop and deliver analytics that we call DataOps.

**Whether referred to as data science, data engineering, data management, big data, business intelligence, or the like, through our work we have come to value in analytics:**

- Individuals and interactions over processes and tools
- Working analytics over comprehensive documentation
- Customer collaboration over contract negotiation
- Experimentation, iteration, and feedback over extensive upfront design
- Cross-functional ownership of operations over siloed responsibilities

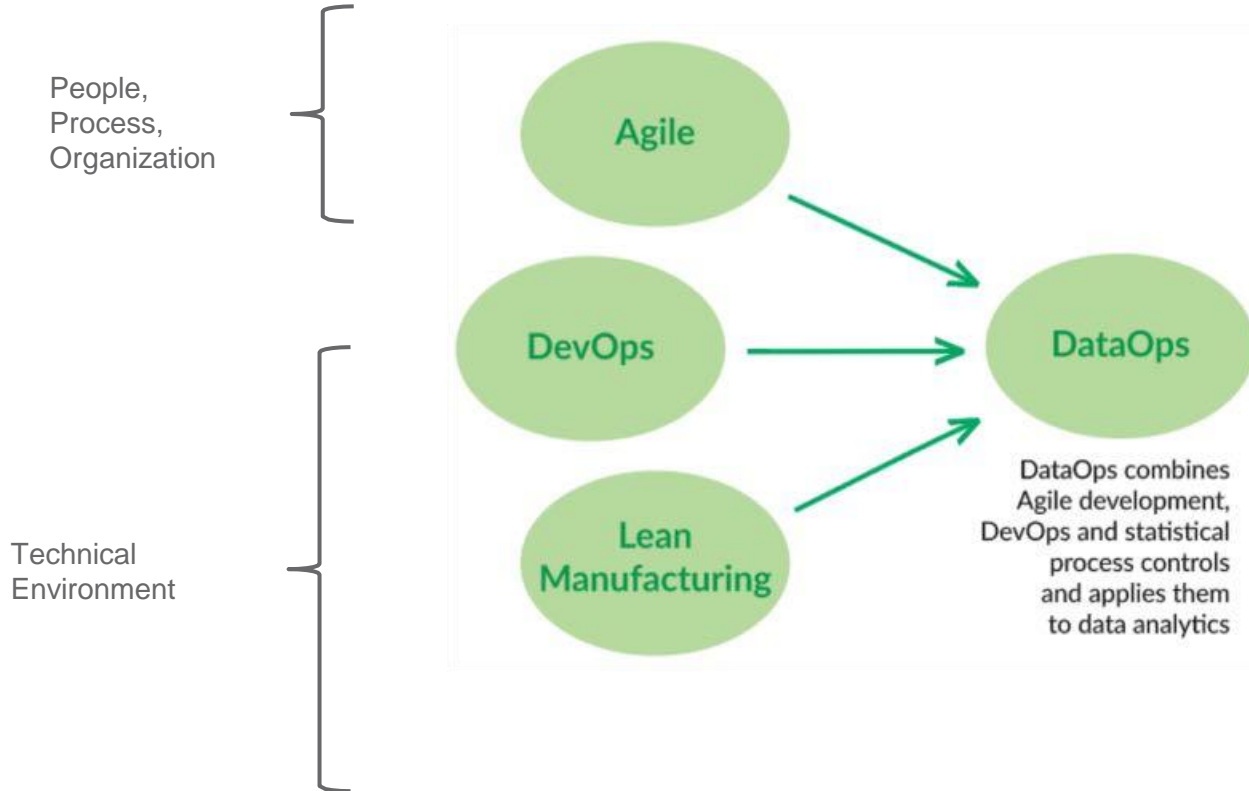
[dataopsmanifesto.org/](https://dataopsmanifesto.org/)

# DataOps Manifesto

## ➤ Princípios do DataOps

- 1. Satisfaça continuamente o seu cliente
- 2. Valor do trabalho analítico
- 3. Abrace a mudança
- 4. É um esporte em equipe
- 5. Interações diárias
- 6. Auto-organização
- 7. Reduza o heroísmo
- 8. Reflita
- 9. Os códigos
- 10. Orquestração
- 11. Faça tudo ser reproduzível
- 12. Ambientes descartáveis
- 13. Simplicidade
- 14. Análise de dados é manufatura
- 15. A qualidade é primordial
- 16. Monitorar a qualidade e o desempenho
- 17. Reutilizar
- 18. Melhorar os tempos dos ciclos

# Genesis of DataOps



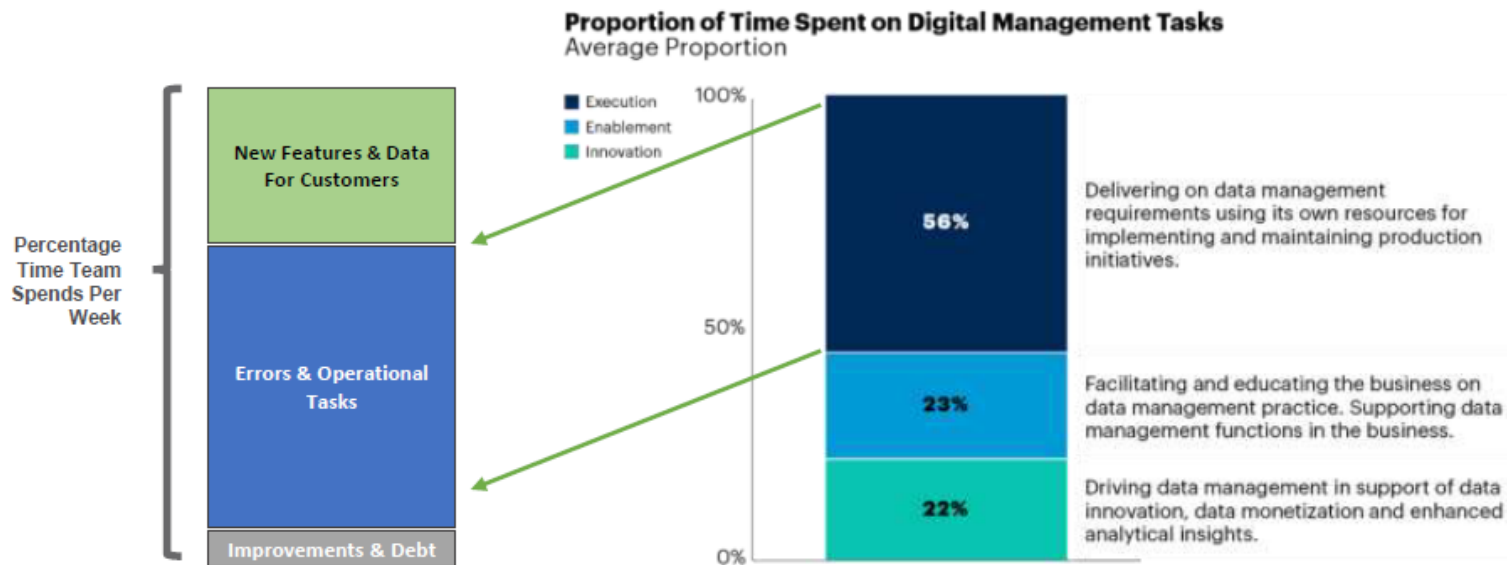
# Desafios de Data Analytics

1. Objetivos de negócio movem-se rápido
2. Dados vivem sem silos
3. Formatos de dados não otimizados
4. Dados errados
5. Dados ruins Relatórios ruins
6. Gestão do Pipeline de dados nunca termina
7. Processo manual esgotado
8. A armadilha da "esperança e heroísmo"



# Desafios de Data Analytics

Pesquisa Gartner - apenas 22% do tempo em novas iniciativas, 56% em execução operacional.



# DataOps benefícios

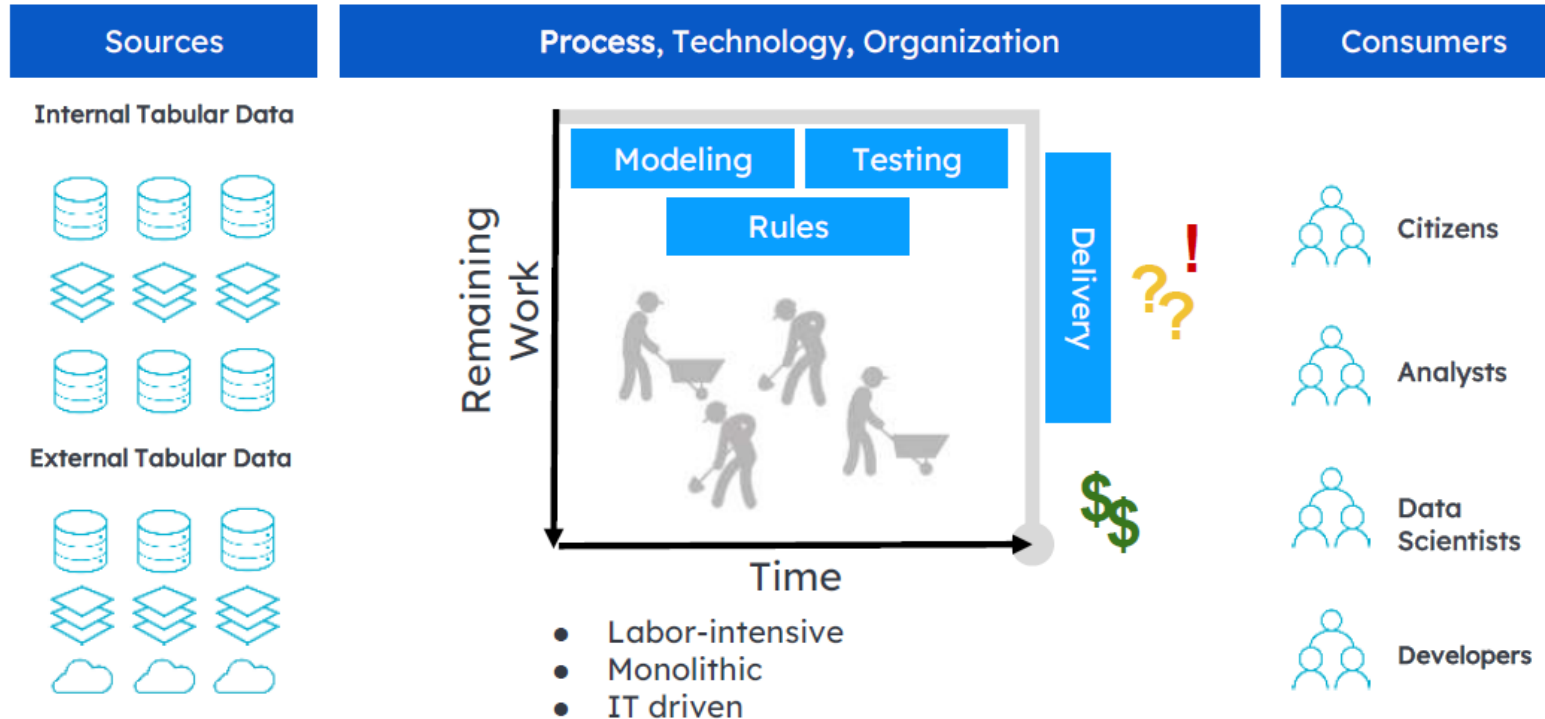


- Um dos principais benefícios de DataOps é entregar aos times, ferramentas que auxiliem no controle e gestão do processo de pipeline de dados

# DataOps benefícios

- Reduzir o tempo para análise de dados (insight)
- Melhorar a qualidade de dados
- Reduzir o custo marginal de perguntar para o negócio a próxima pergunta
- Aprimorar a capacidade de integração e comunicação do time
- Promover maior eficiência do time através de processos ágeis, reuso e deploy

# DataOps Framework



# DataOps Framework

## Sources

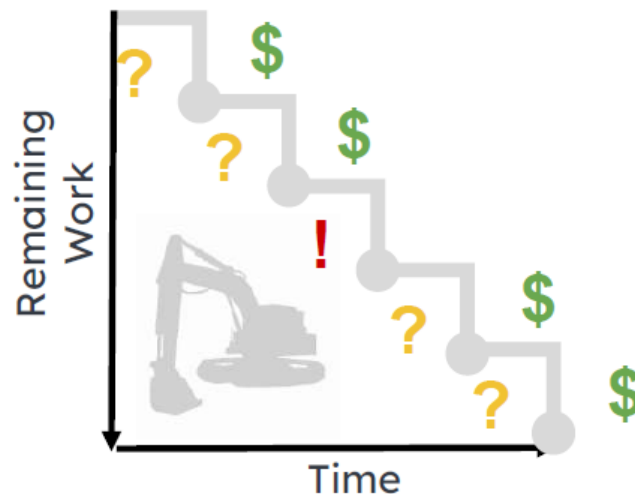
### Internal Tabular Data



### External Tabular Data



## Process, Technology, Organization



- Automated
- Incremental
- Collaborative

## Consumers



Citizens



Analysts

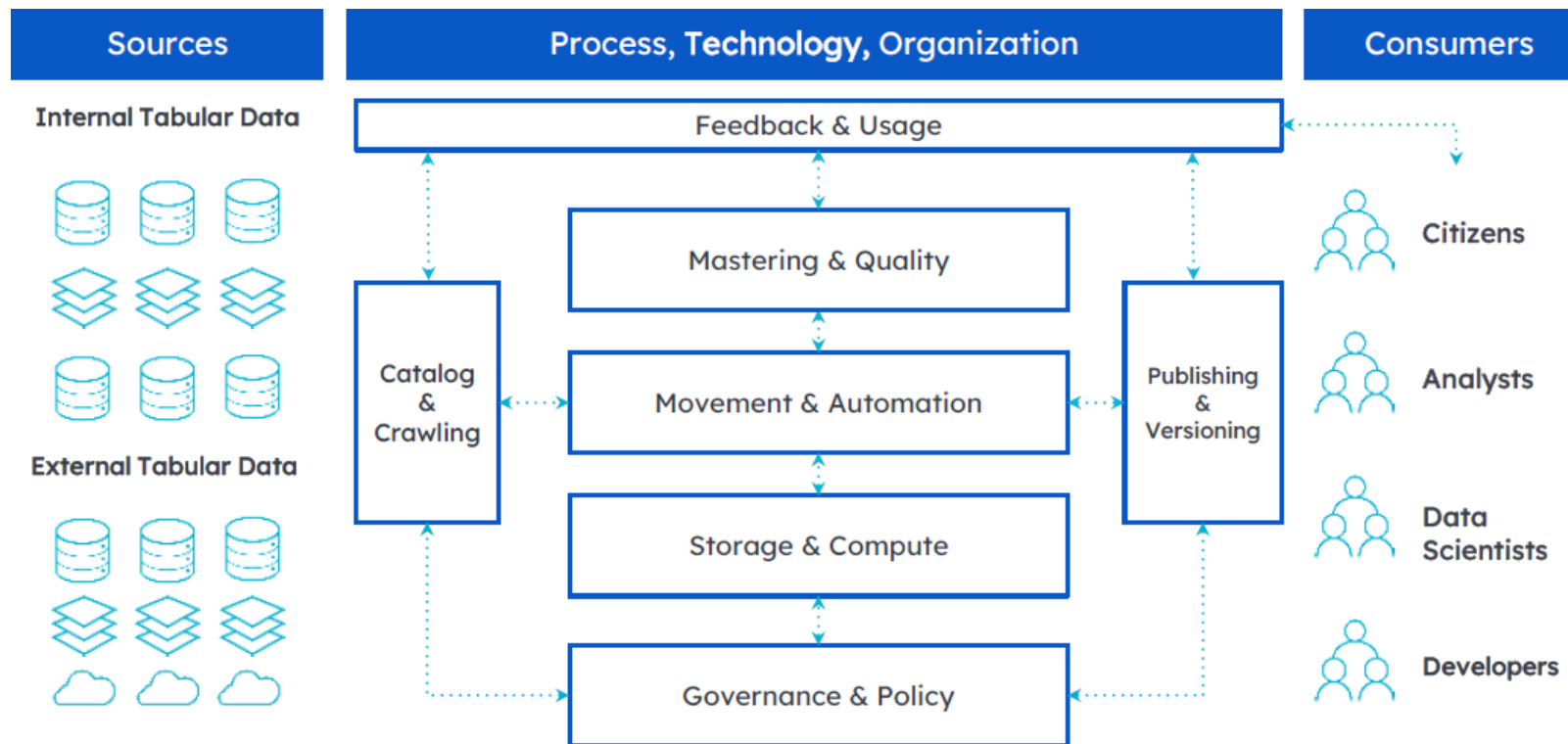


Data Scientists



Developers

# DataOps Framework



# DataOps Framework

- Processo
  - Agile - modelo de entrega incremental
- Tecnologia
  - Arquitetura - seleção de ferramentas que compõem a cadeia de ingestão de dados
  - Infraestrutura - seleção de plataforma para suportar arquitetura
- Organização
  - Funções - divisão do trabalho entre equipes de habilidades mistas
  - Estrutura - modelo de trabalho para projetos em equipes técnicas e de negócios

# DataOps Framework

## Getting Started - Processos

- Ágil é a chave
  - Se ainda não estiver lá, escolha um modelo que funcione (Scrum, SAFe)
- Avaliar projetos para entrega de produtos de dados
  - Pontuação na disponibilidade de dados x valor da solução de um problema
- Definir projeto de alto valor e rico em dados que exigirá uma solução complexa
  - Implementar o processo para garantir a funcionalidade de ponta a ponta



# DataOps Framework

## Getting Started – Tecnologia

- **Identifique o caminho para uma arquitetura de serviço moderna e modular**
  - Criar projeto para a próxima geração de plataforma de gerenciamento de dados
  - Revisitar a estratégia de migração na nuvem (cloud first)
- **Separar processos monolíticos**
  - Encapsular componentes em APIs, expor como serviços
- **Comece a construir com novas tecnologias**
  - Escolha um subconjunto de ferramentas para prova de conceitos para substituir a tecnologia antiga, se existir

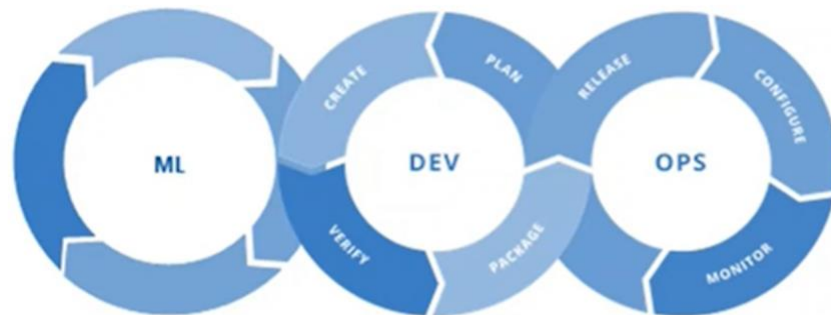
# DataOps Framework

## Getting Started – Organização

- Equipe atual
  - Identificar as principais funções existentes
  - Encontre os melhores candidatos para novas funções
- Criar equipe multifuncional
  - Data consumers
  - Data Engineer
  - Data Analyst
  - Data Architect
- Escolha o seu modelo operacional
  - Iniciar com serviços compartilhados para o primeiro projeto
- Garanta o alinhamento executivo
  - CDO ou equivalente

# MLOPS

## MACHINE LEARNING OPS



**Experiment**  
Data Acquisition  
Business Understanding  
Initial Modeling

**Develop**  
Modeling + Testing  
Continuous Integration  
Continuous Deployment

**Operate**  
Continuous Delivery  
Data Feedback Loop  
System + Model Monitoring

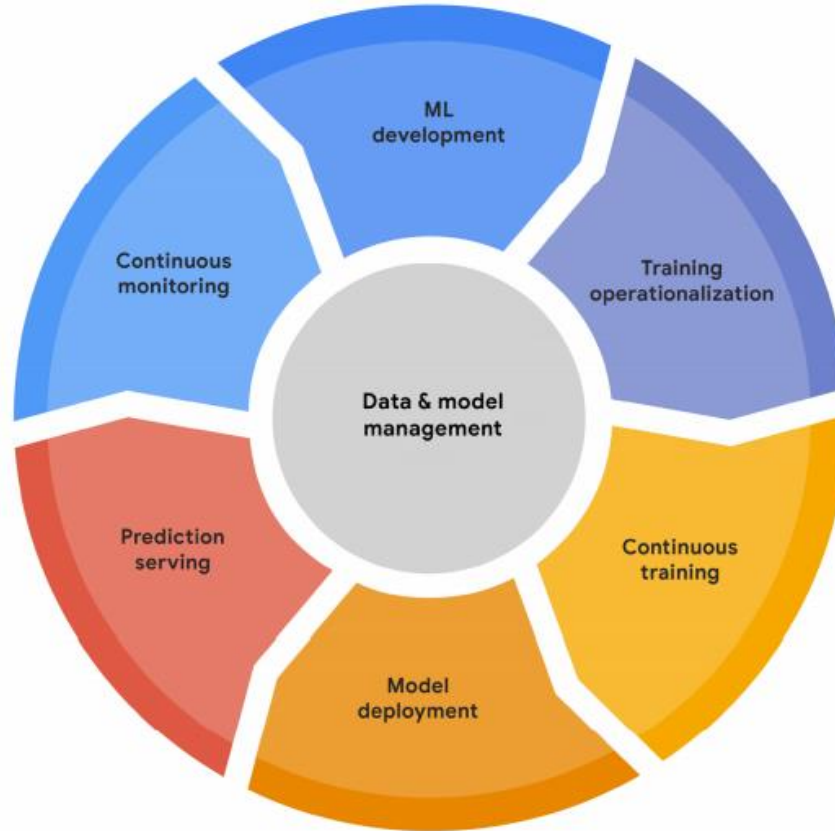
*Fonte: Azure ML*

# MLOps, Machine Learning Ops

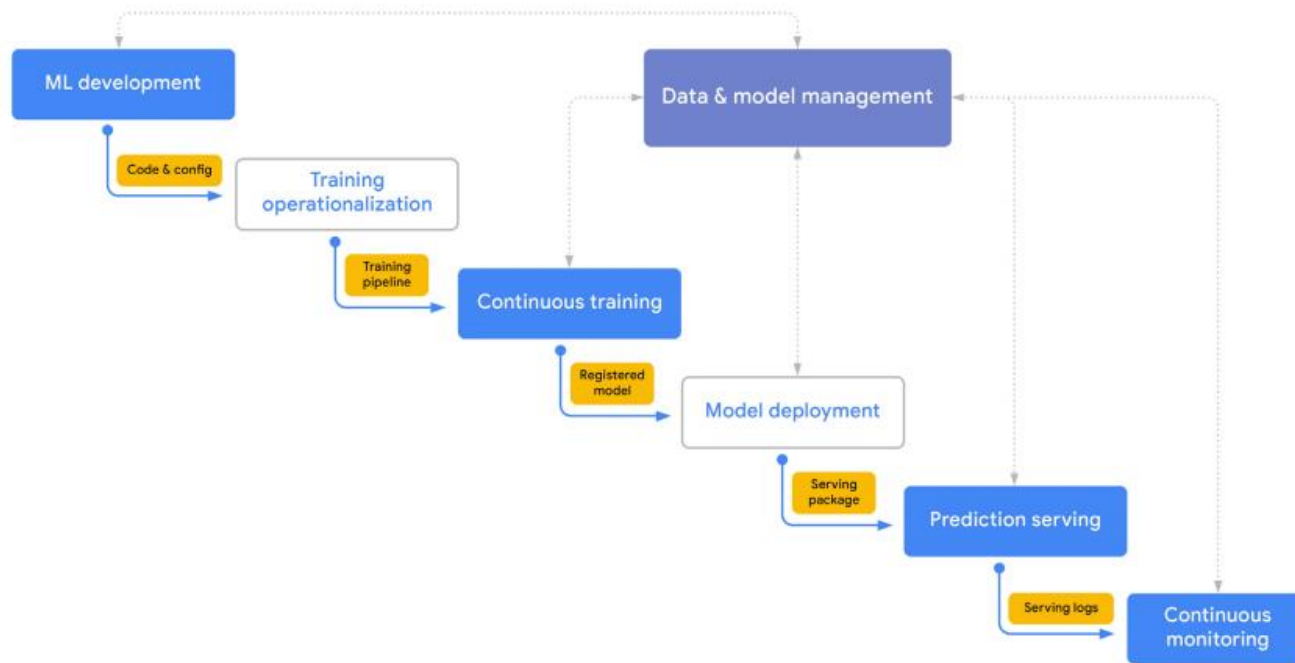
- MLOps surgiu em 2015 pela primeira vez com a crescente discussão dos cientistas de dados em como melhorar seus processos de treinamento de modelos, o qual exige constantes execuções do modelo, com a preparação de datasets, configuração de serviços e coleta de resultados.

Sculley ,Google, 2015 "*Hidden Technical Debt in Machine Learning Systems*"

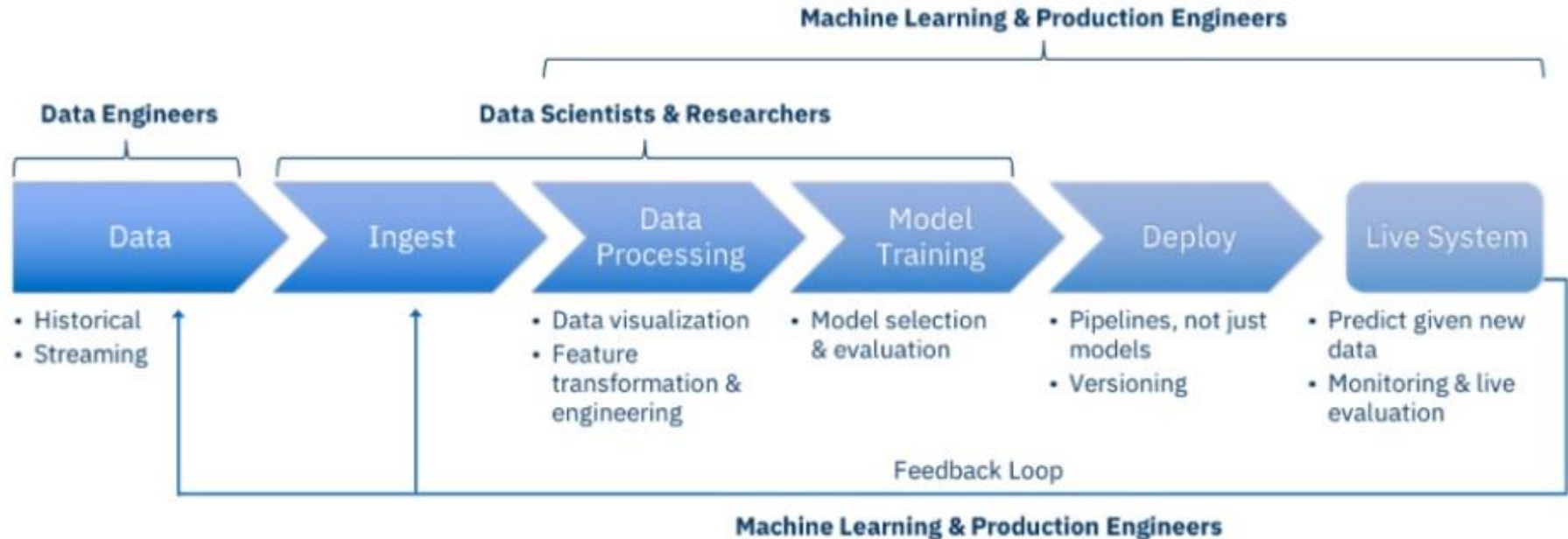
# MLOps, Machine Learning Ops



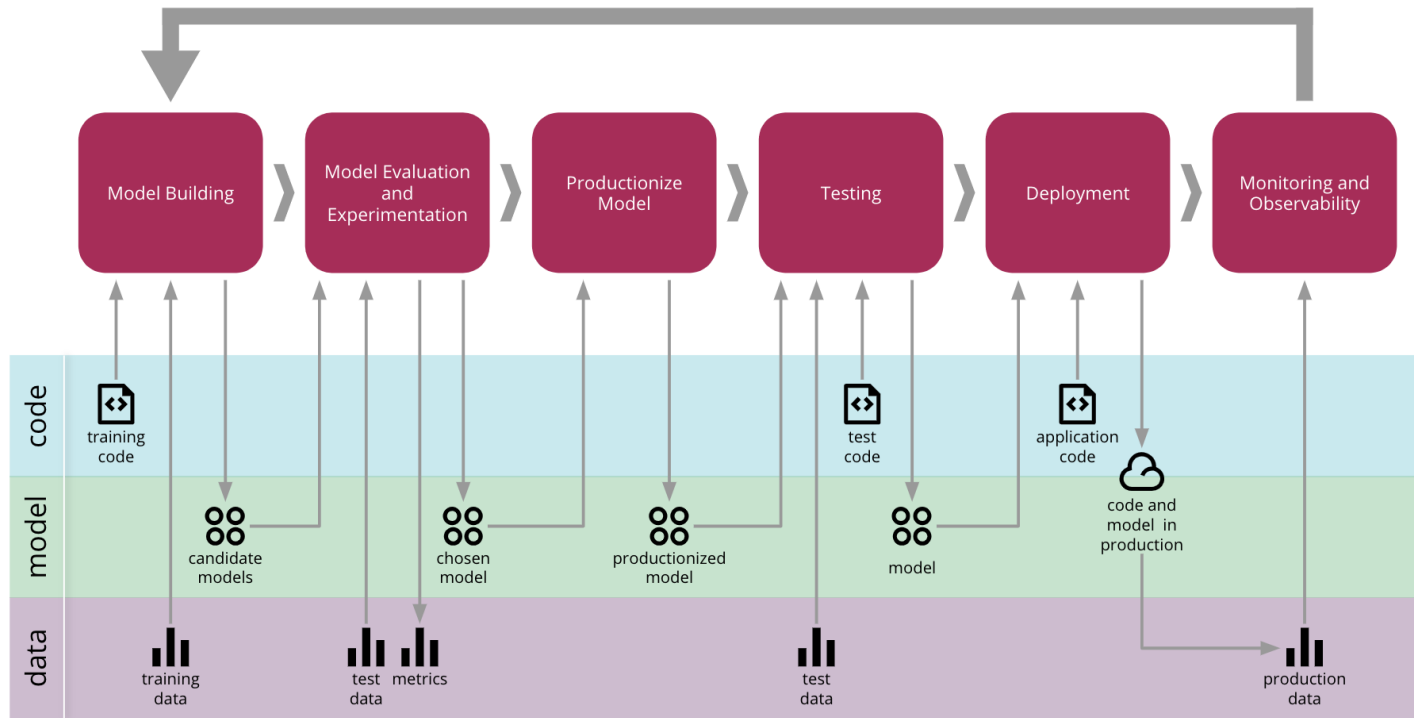
# MLOps: End-to-end workflow



# MLOps, Machine Learning Ops



# MLOps, Machine Learning Ops



Fonte: Martin Fowler, 2019, Continuous Delivery for Machine Learning end-to-end process



# MLOps, Machine Learning Ops

Principais objetivos com adoção do MLOps :

- Reduzir o tempo e dificuldade de publicar modelos em produção
- Melhorar a comunicação entre as equipes
- Auxiliar no gerenciamento do processo, rastreamento de falhas e controle de versões
- Implementar um ciclo de vida no projeto de ML
- Padronização do processo de ML, facilitando melhor controle do processo de implantação e mudança.



**DataOps Team**

# DataOps Team



# DataOps Team

	Role	Goals	Tools
Consumers	<b>Citizen</b>	Use data to make business decisions	Viz, CRM, Excel, PowerPoint, Word, Web Search
	<b>Analyst</b>	Deliver insights to the business, typically through dashboards and reports	Viz, Excel, SSDP, Web Search
	<b>Scientist</b>	Deliver insights to the business, typically through models and algorithms	R, Python, SAS, SSDP
	<b>Developer</b>	Build applications which leverage corporate data	Python, Java, JS, SQL, REST
Preparers	<b>Engineer</b>	Deliver and manage data pipelines	ETL, SQL
	<b>Curator</b>	Ensure consumers have the data they need, in the form they need it	MDM, Catalog
	<b>Steward</b>	Create policies and drive governance	MDM, Catalog, Governance
Suppliers	<b>Source Owner</b>	Define and manage purpose, processes (data creation, consumption) & users (i.e., access) of the data source	EDW, SQL, ERWin, LDAP, SAP

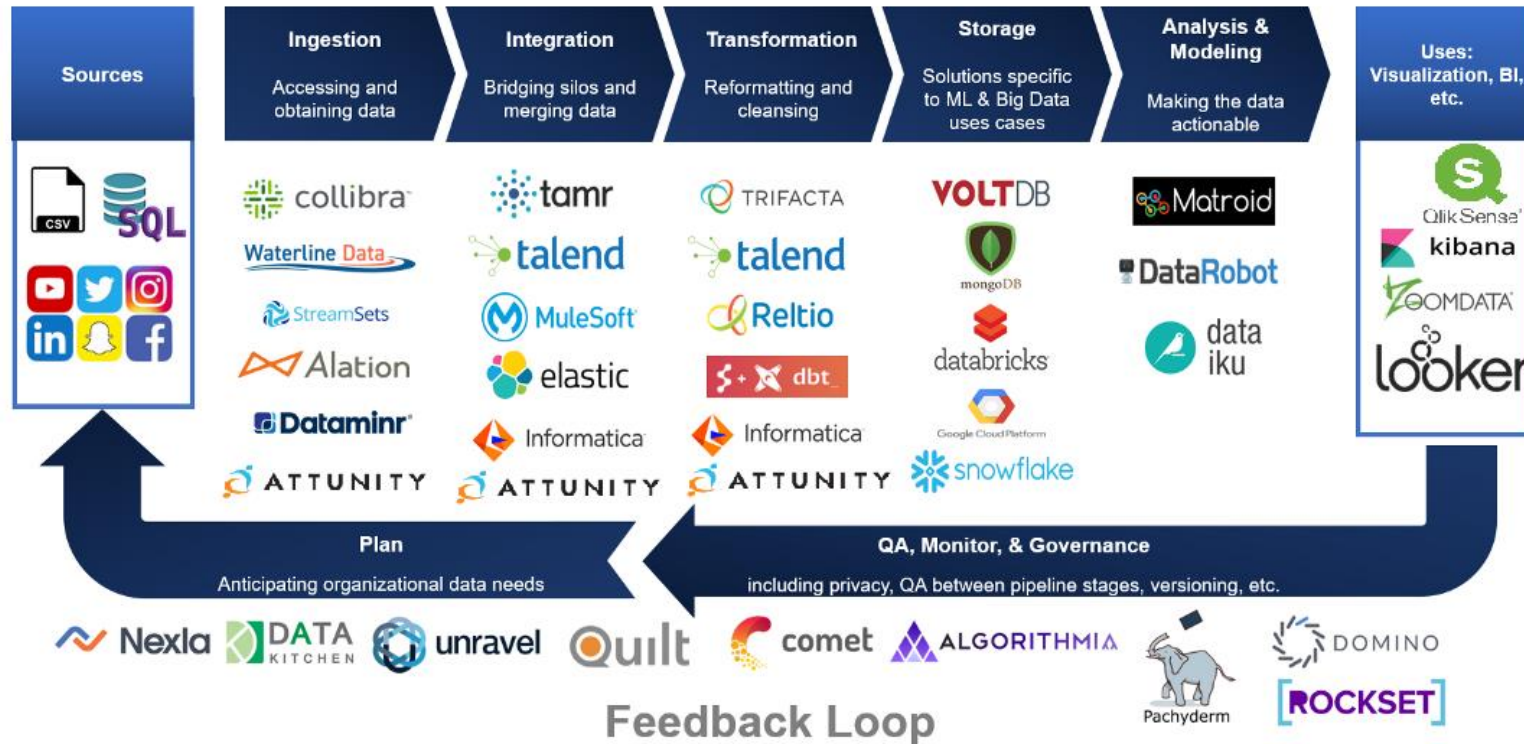


# DATAOPS ECOSYSTEM

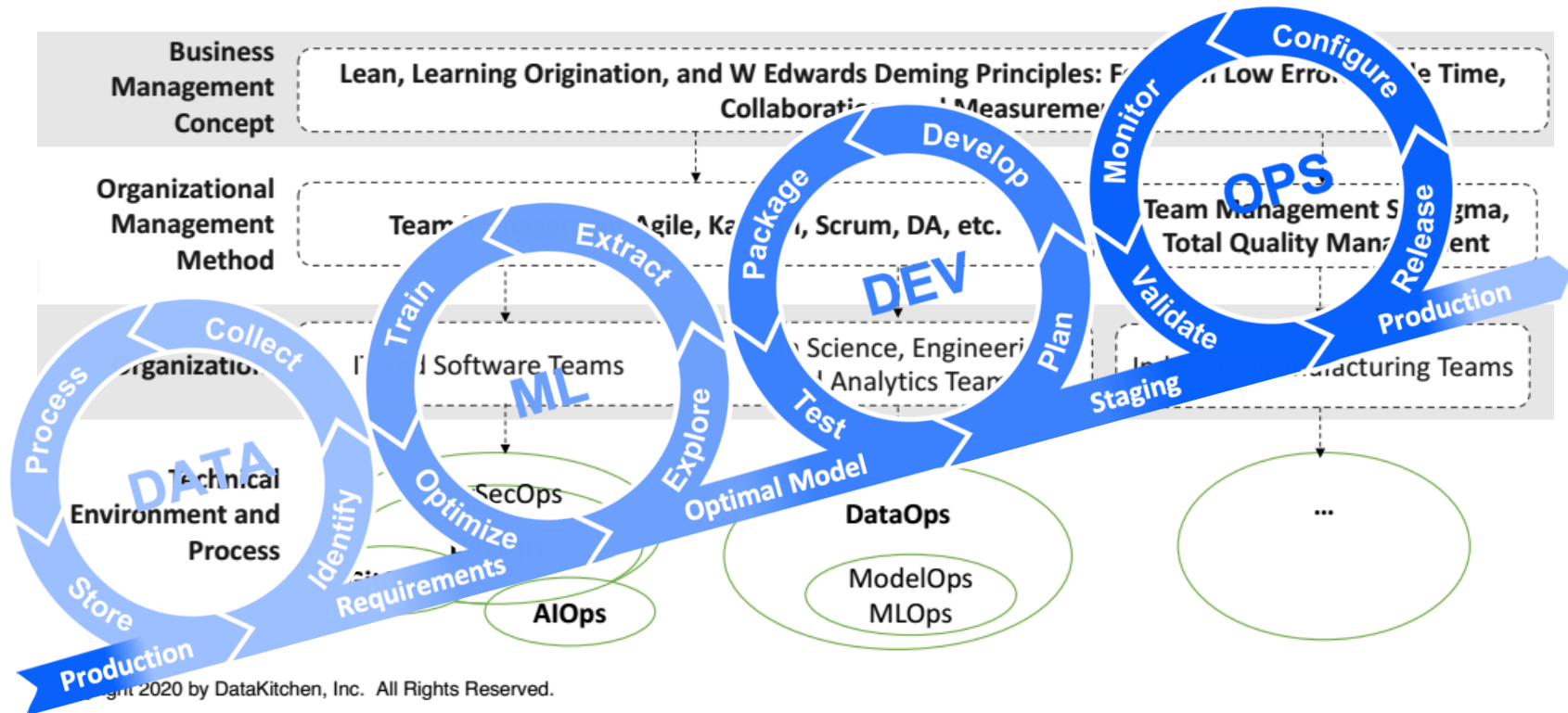


# DataOps Ecosystem

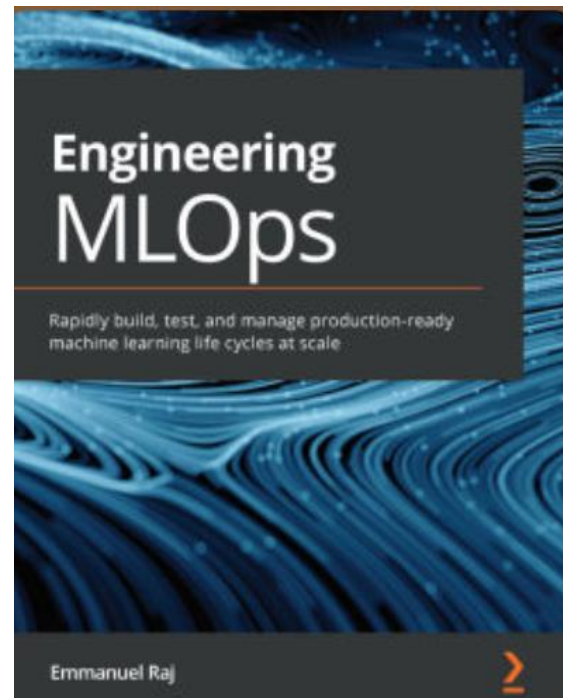
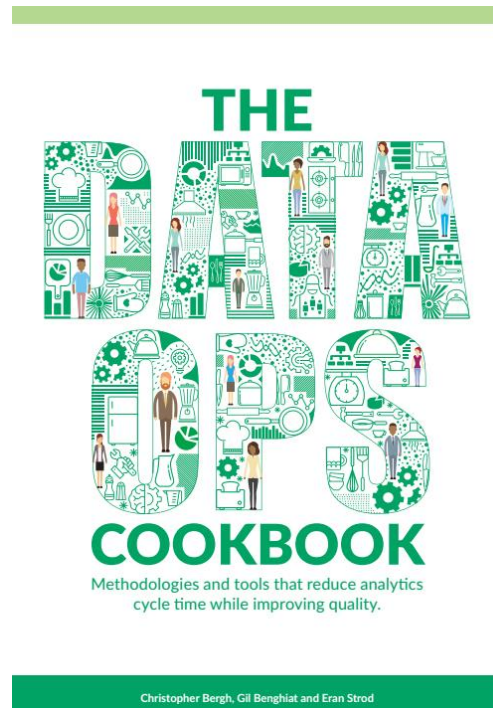
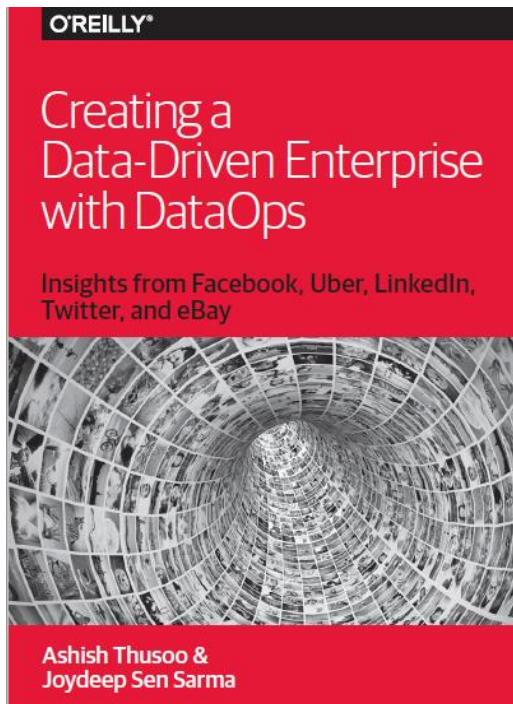
## Data Delivery Pipeline, from Sources to Uses



# DevOps vs DataOps (e outros\*Opses)



# Mais conteúdo





# References

- **DataOps Ecosystem** - [//medium.com/data-ops/2017-the-year-of-dataops-b2023c17d2af](https://medium.com/data-ops/2017-the-year-of-dataops-b2023c17d2af)
- **DataOps Manifesto** - [//dataopsmanifesto.org/](https://dataopsmanifesto.org/)
- **Continuous Delivery for Machine Learning** - [//martinfowler.com/articles/cd4ml.html](https://martinfowler.com/articles/cd4ml.html)
- **DataOps—It's a Secret** - [//www.datasciencecentral.com/profiles/blogs/dataops-it-s-a-secret](https://www.datasciencecentral.com/profiles/blogs/dataops-it-s-a-secret)
- **The Power of DataOps** - [//www.delphix.com/blog/power-dataops](https://www.delphix.com/blog/power-dataops)
- **Building a DataOps Team** - [//medium.com/data-ops/building-a-dataops-team-abc375e0a6bc](https://medium.com/data-ops/building-a-dataops-team-abc375e0a6bc)
- **Guia para profissionais de MLOps** - [//cloud.google.com/resources/mlops-whitepaper](https://cloud.google.com/resources/mlops-whitepaper)
- **MLOps.community** - [mlops-community.slack.com](https://mlops-community.slack.com)
- **ML Ops Best Practices on Google Cloud** - [//www.youtube.com/watch?v=20h\\_RTHetZI](https://www.youtube.com/watch?v=20h_RTHetZI)

# Concluindo....

- As empresas que desejam implementar DataOps/MLOps devem concentrar seus esforços em três áreas:
  - Cultura
  - Processos
  - Tecnologia

# About me

## ➤ Eduardo Hahn

- Founder DataLakers Tecnologia
- Data Lover & DataOps Enthusiastic
- [eduardo.hahn@datalakers.com.br](mailto:eduardo.hahn@datalakers.com.br)
- [@eduardohahn](#)
- [/in/eduardohahn3](#)



Partners

**CLOUDERA**

**denodo**



Google Cloud Platform



# A teoria à prática na implementação de DataOps e MLOps

Eduardo Hahn

DataLakers founder & DataOps Enthusiastic