THE DEVELOPER'S CONFERENCE

Trilha – Big Data

Passo a Passo para implementar DataOps em projetos de Big Data

Eduardo Hahn

DataLakers
The Big Data Company

THE DEVELOPER'S CONFERENCE

- + 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

DataLakers
The Big Data Company

Começo.....

THE DEVELOPER'S CONFERENCE

- 2014....

IBM Big Data & Analytics Hub

Blogs

3 reasons why DataOps is essential for big data success

JUNE 19, 2014

by Lenny Liebmann
Contributing Editor, InformationWeek
Follow me on LinkedIn, Twitter

Developers once wrote application code and just "threw it over the wall" to IT operations, which then had to ensure that those applications performed well in the production environment. This was always a less-than-optimal approach, but it became untenable as the business began to depend more and more on lots of fresh code getting rolled out into production quickly and with a high degree confidence. So IT organizations are now embracing a set of best practices known as *DevOps* that improve coordination between development and operations.

"A ciência de dados é uma disciplina excepcionalmente importante hoje em dia. Mas essa ciência só é útil na medida em que pode ser executada de forma eficiente e confiável. E para que isso aconteça, você precisa de DataOps."

DataLakers
The Big Data Company

Começo.....

THE DEVELOPER'S CONFERENCE

- 2015...

From DevOps to DataOps, By Andy Palmer

Posted on Thursday, May 7th, 2015 at 1:55 PM.
Written by Andy Palmer

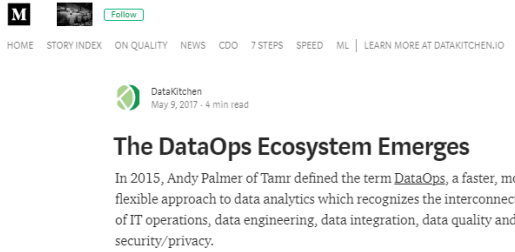
Why It's Time to Embrace "DataOps" as a New Discipline

Over the past 10 years, many of us in technology companies have experienced the emergence of "DevOps." This new set of practices and tools has improved the velocity, quality, predictability and scale of software engineering and deployment. Starting at the large internet companies, the trend towards DevOps

DataLakers
The Big Data Company

Acelerou.....

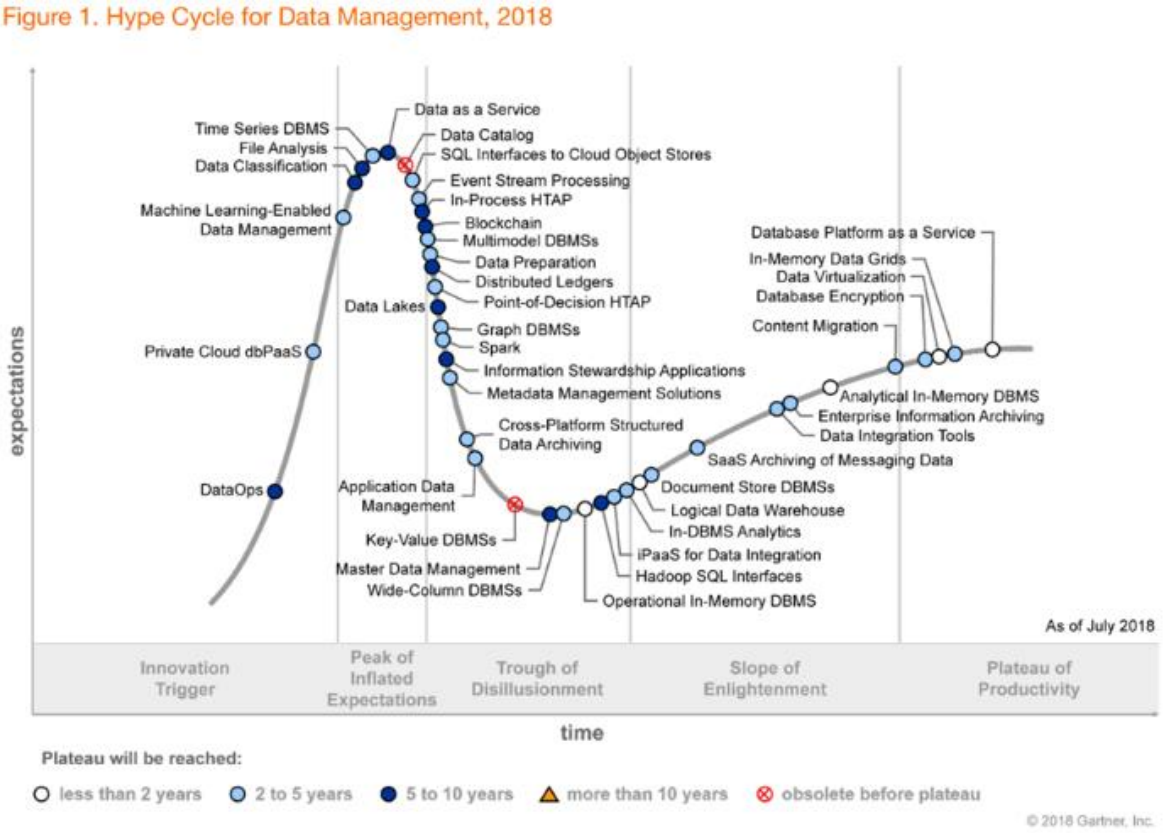
- 2017....



Agora sim.....

THE DEVELOPER'S CONFERENCE

- 2018...



DataLakers
The Big Data Company

Agora sim.....

THE DEVELOPER'S CONFERENCE

DataOps is 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.

Gartner

DataLakers
The Big Data Company

DataOps Manifesto

THE DEVELOPER'S CONFERENCE

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/

DataLakers
The Big Data Company

DataOps Manifesto

THE DEVELOPER'S CONFERENCE

- **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**

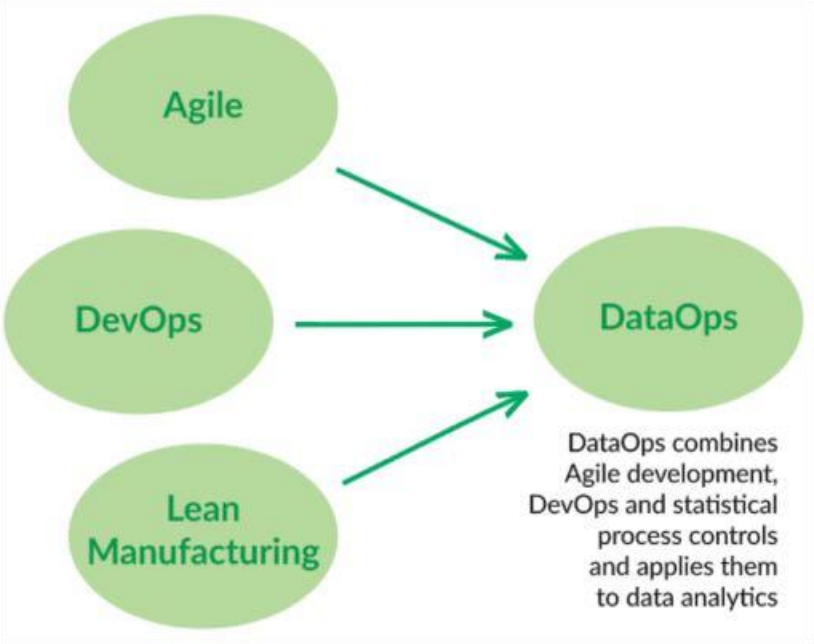
DataLakers
The Big Data Company

Genesis of DataOps

THE DEVELOPER'S CONFERENCE

People,
Process,
Organization

Technical
Environment



DataLakers
The Big Data Company

4 “As” de DataOps

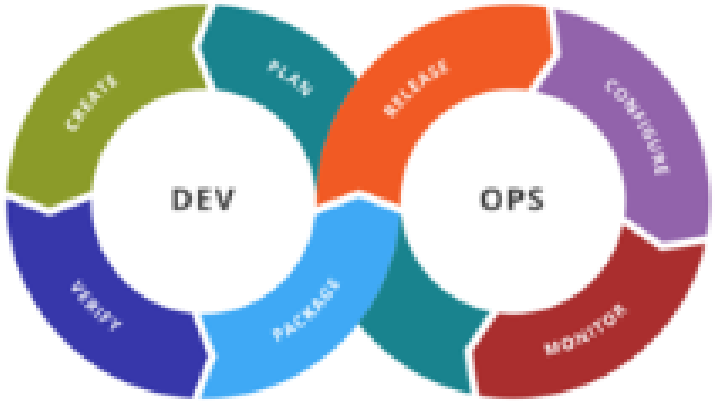
THE DEVELOPER'S CONFERENCE

- Automatize e monitorar pipelines
- Automatizar implantações
- Automatizar e monitorar a qualidade
- Automatizar sandbox

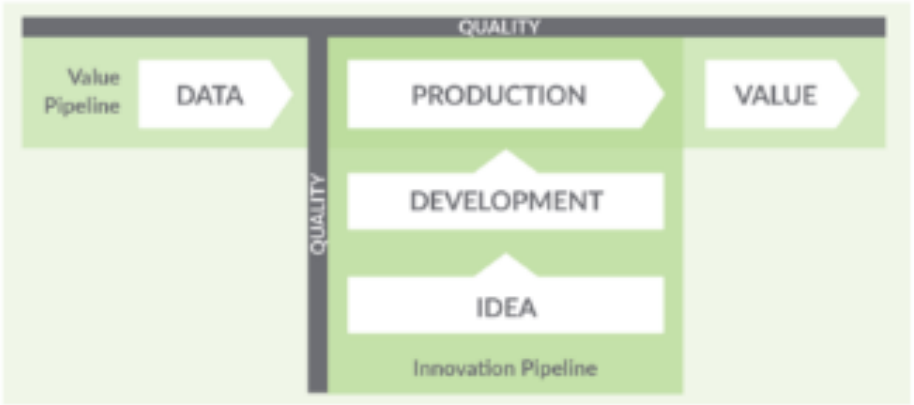
DataLakers
The Big Data Company

DataOps is NOT Just DevOps for Data

THE DEVELOPER'S CONFERENCE



DevOps



DataOps

DataLakers
The Big Data Company

DataOps is NOT Just DevOps for Data

THE DEVELOPER'S CONFERENCE

Different People and Expectations

DevOps Users & Tools

Software Engineers, comfortable with coding and complexity of multiple languages, tools, and hardware/software.

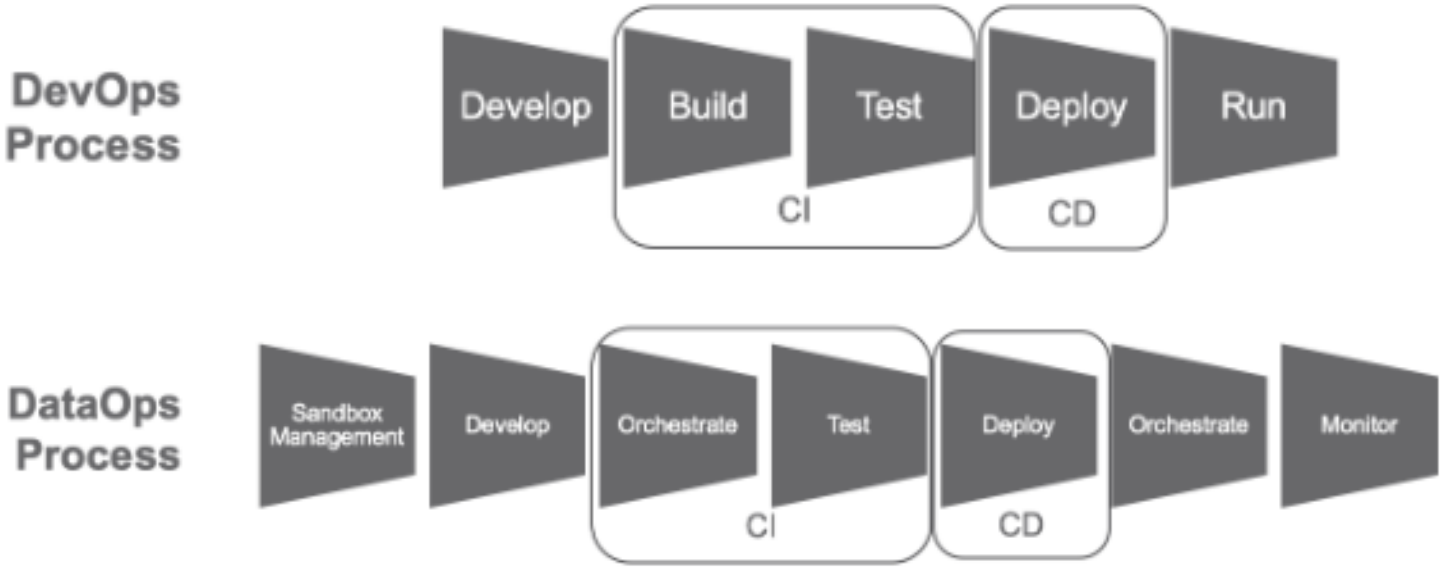
DataOps User & Tools

Data Scientists, Engineers, and Analysts who want to just analyze data and build models.

DataLakers
The Big Data Company

DataOps is NOT Just DevOps for Data

THE DEVELOPER'S CONFERENCE



DataLakers
The Big Data Company

DataOps is NOT Just DevOps for Data

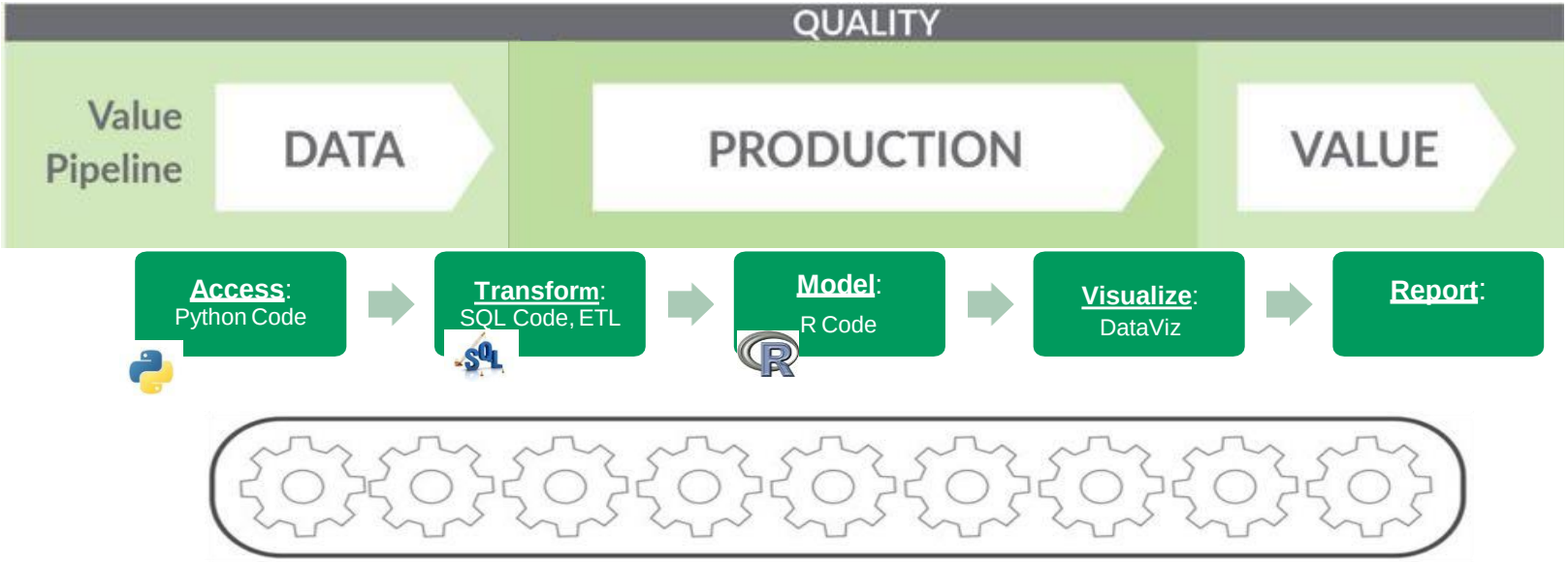
THE DEVELOPER'S CONFERENCE

DataLakers
The Big Data Company

DataOps is NOT Just DevOps for Data

THE DEVELOPER'S CONFERENCE

Automatize e monitorar pipelines

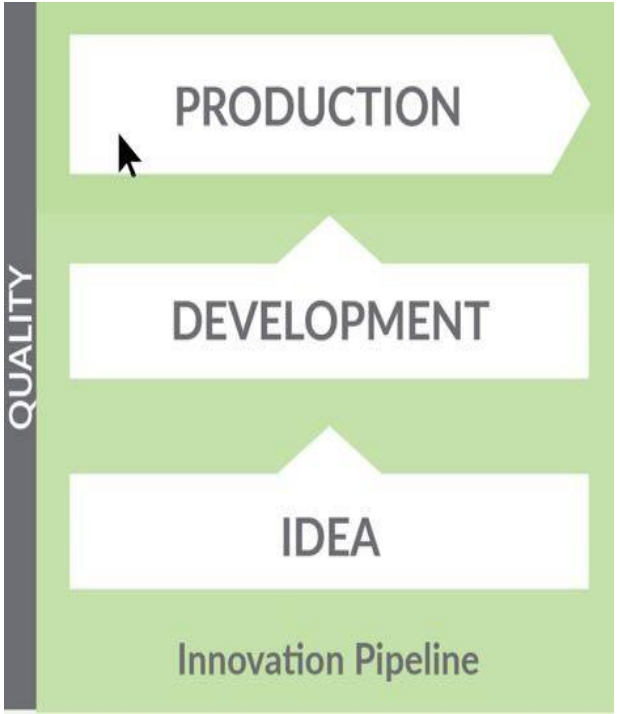


DataLakers
The Big Data Company

DataOps is NOT Just DevOps for Data

THE DEVELOPER'S CONFERENCE

Automatizar implantações



Business
Customer

Products &
Systems

Diverse Team

Data
Engineers

Data
Scientists

Data
Analysts

Diverse Tools

DataLakers
The Big Data Company

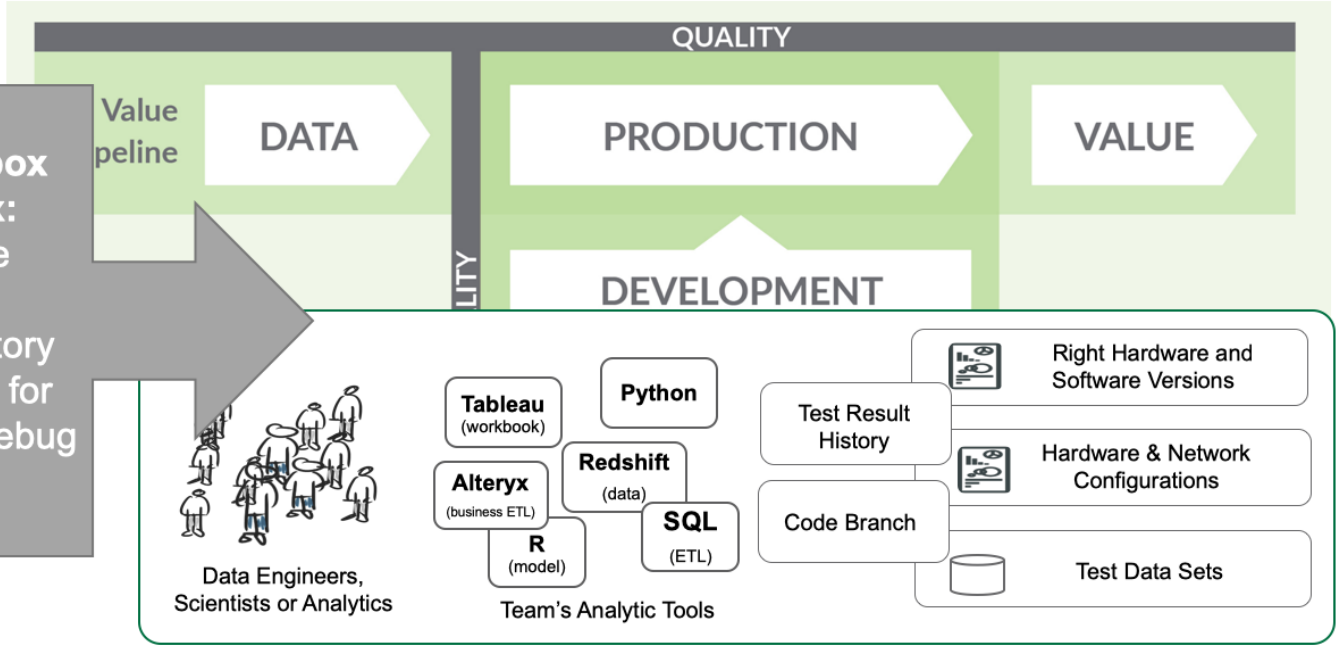
DataOps is NOT Just DevOps for Data

THE DEVELOPER'S CONFERENCE

Automatizar sandbox

Development Sandbox Creation is Complex:

Hard to create the right set of data, tools, people, history and configuration for a fast build test debug cycle

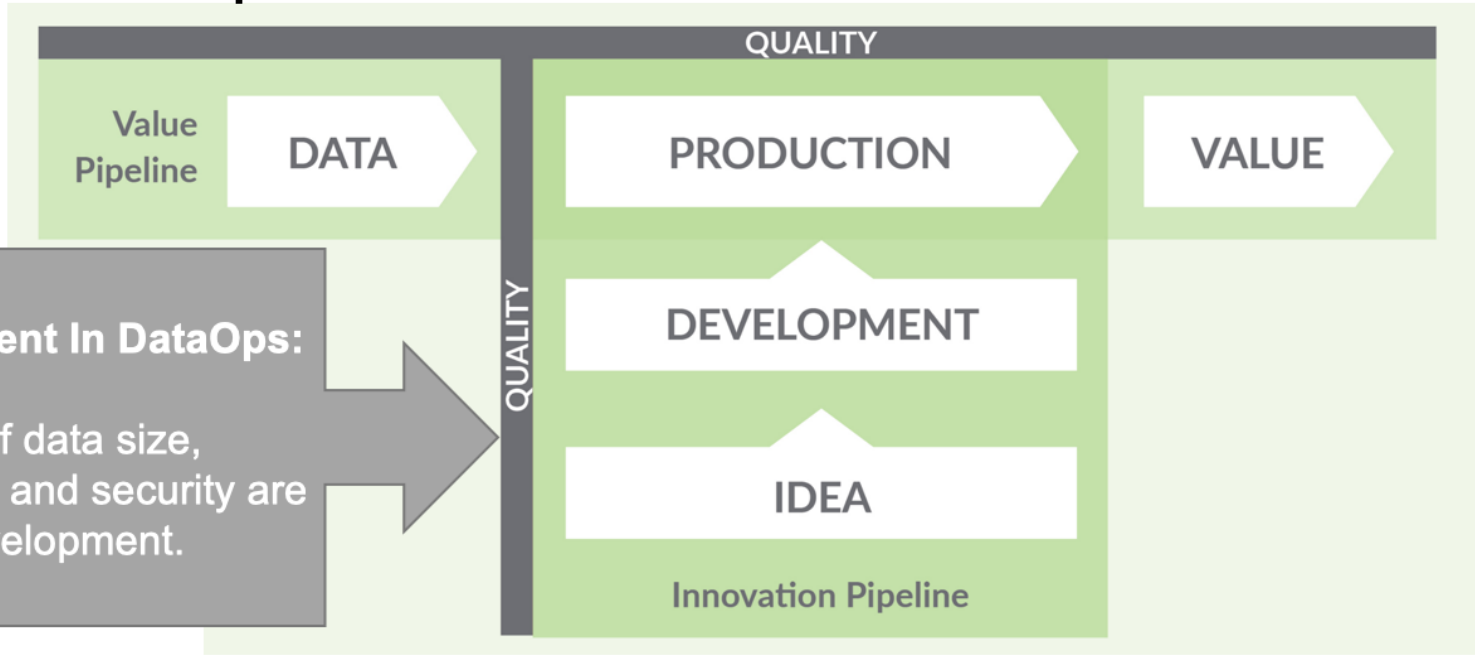


DataLakers
The Big Data Company

DataOps is NOT Just DevOps for Data

THE DEVELOPER'S CONFERENCE

Automatizar e monitorar a qualidade



Test Data Management In DataOps:

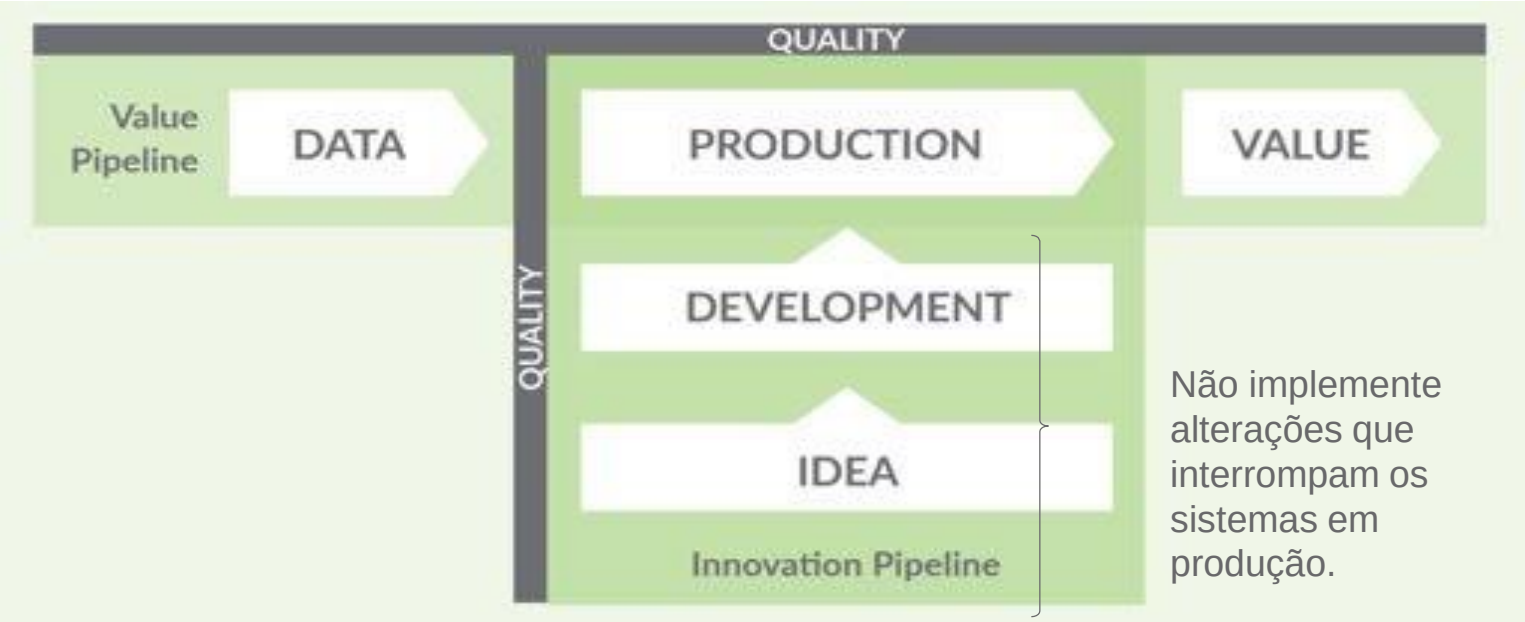
Consideration of data size, complexity, sampling and security are required in development.

DataLakers
The Big Data Company

DataOps is NOT Just DevOps for Data

THE DEVELOPER'S CONFERENCE

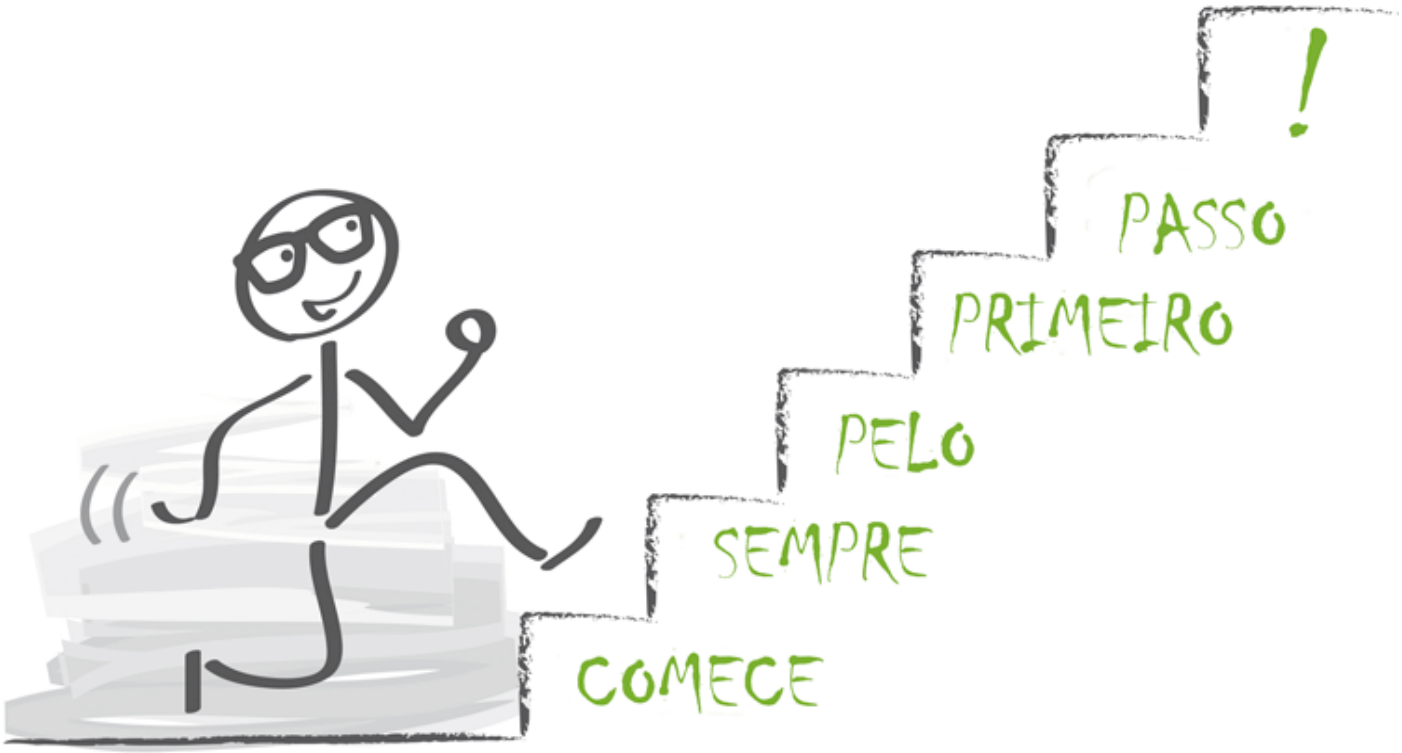
Não permita que dados de baixa qualidade cheguem aos usuários no **Value Pipeline**



DataLakers
The Big Data Company

Passos para implementar DataOps

THE DEVELOPER'S CONFERENCE



DataLakers
The Big Data Company

Passos para implementar DataOps

THE DEVELOPER'S CONFERENCE

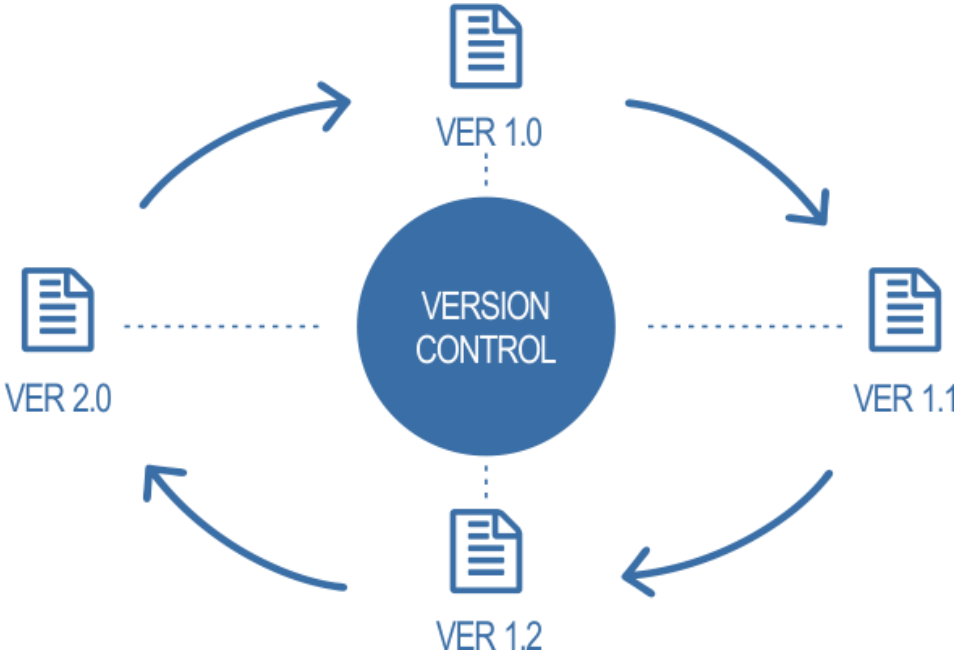
- Passo 1: incremente Data Test e Logic Test

DataLakers
The Big Data Company

Passos para implementar DataOps

THE DEVELOPER'S CONFERENCE

- Passo 2: use controle de versão (Branch&Merge).

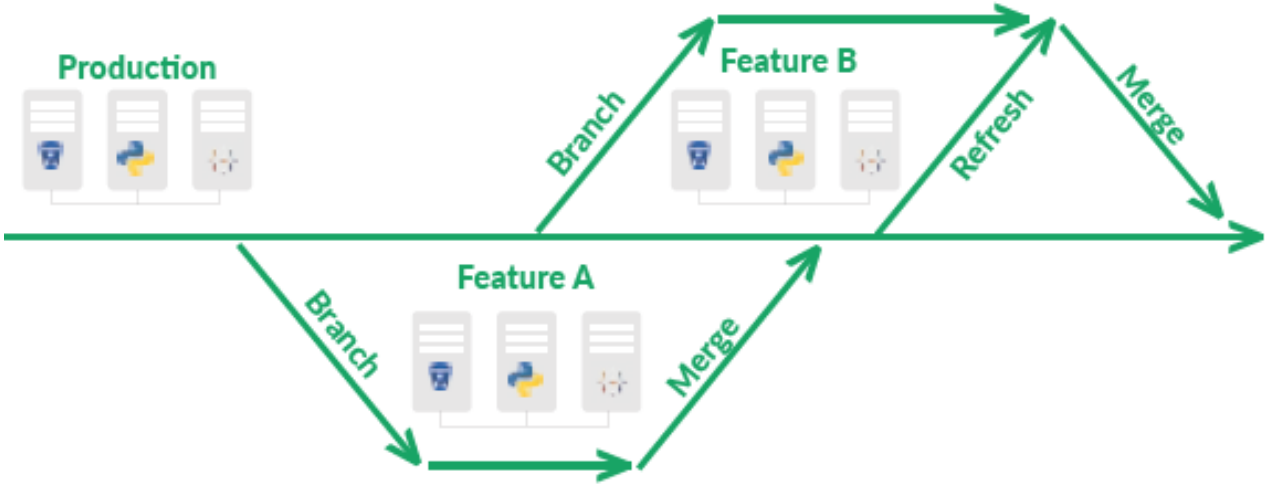


DataLakers
The Big Data Company

Passos para implementar DataOps

THE DEVELOPER'S CONFERENCE

- Passo 3: Use múltiplos ambientes



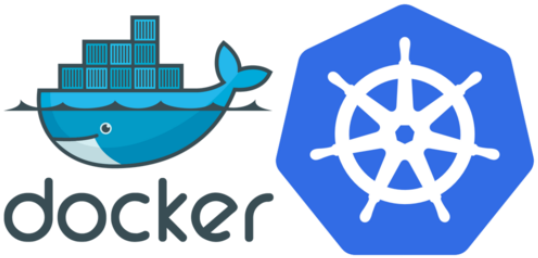
DataLakers
The Big Data Company

Passos para implementar DataOps

THE DEVELOPER'S CONFERENCE

Passo 4: Reuso e containers

Outro método de aumento de produtividade para equipes é a capacidade de reutilizar e conter o código. Cada etapa intermediária no pipeline de análise de dados recebe a saída de um estágio anterior e fornece entrada para o próximo estágio. É mais fácil para os outros membros da equipe reutilizar componentes menores, se eles puderem ser segmentados ou conteinerização. Melhor caminho é usar Docker

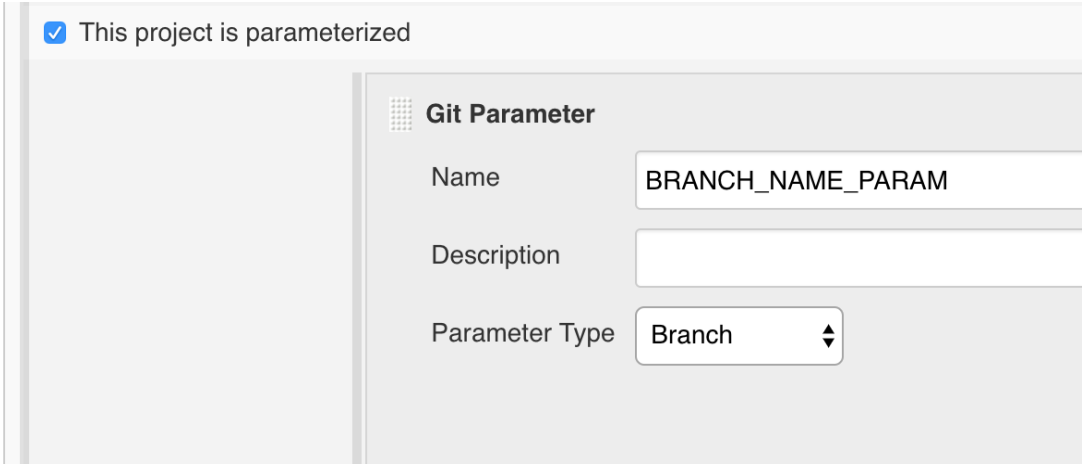


DataLakers
The Big Data Company

Passos para implementar DataOps

THE DEVELOPER'S CONFERENCE

- Passo 5: Parametrize seus processos



DataLakers
The Big Data Company

Passos para implementar DataOps

THE DEVELOPER'S CONFERENCE

Passo 6: Use Simple Unique Storage

- Data Lake: os dados são movidos de diferentes silos de dados para um repositório comum, é muito mais fácil para uma equipe de análise de dados trabalhar com ele.

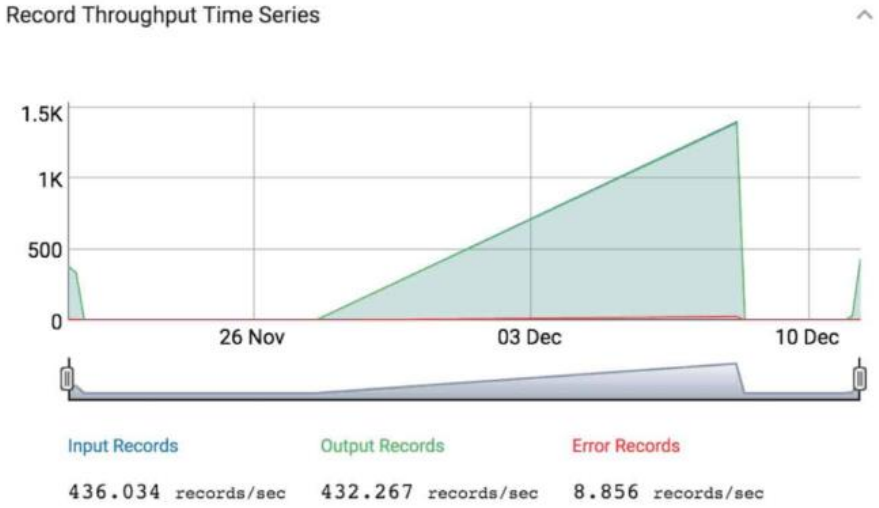
DataLakers
The Big Data Company

Passos para implementar DataOps

THE DEVELOPER'S CONFERENCE

Passo 7: Medir todas as etapas do pipeline

Defina métricas de ponta a ponta para sua arquitetura. Identifique pontos de melhorias e problemas de desempenho. Visualize uma arquitetura de dados online, visualizando como os sistemas evoluem.



DataLakers
The Big Data Company

DataOps Ecosystem

THE DEVELOPER'S CONFERENCE

Platform Solutions

composable

DATA KITCHEN

tamr

StreamSets

kinæsis

data / intelligence

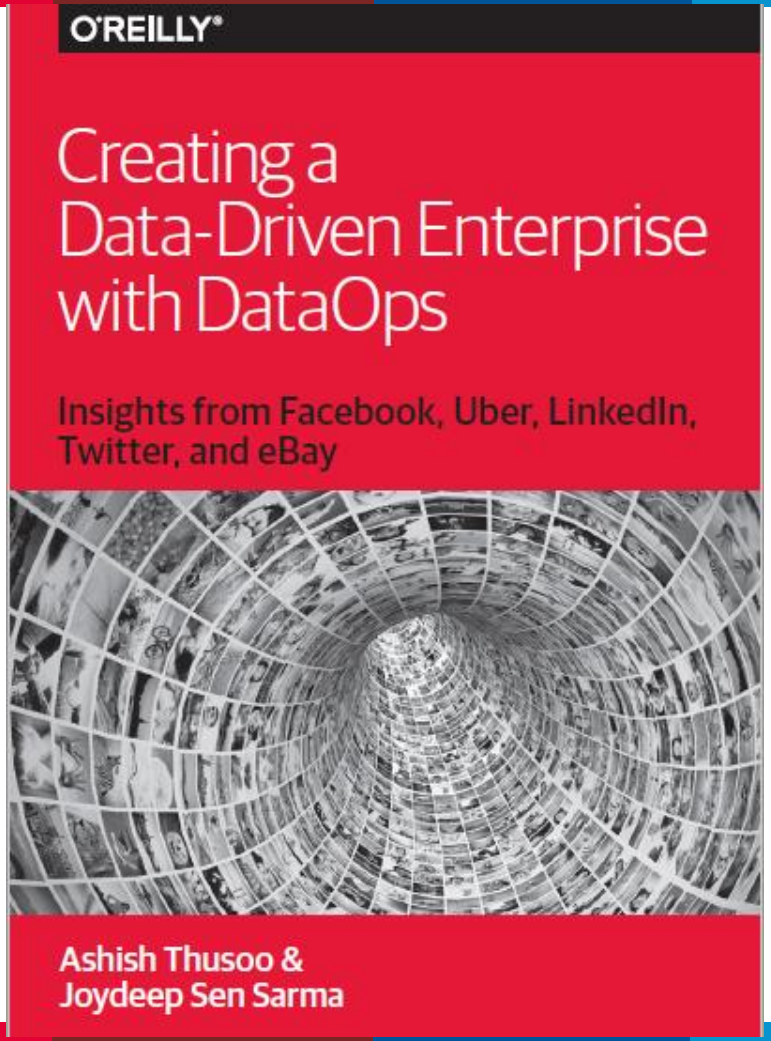
talend

ATTUNITY

DataLakers

The Big Data Company

References



THE DEVELOPER'S CONFERENCE

DataLakers
The Big Data Company

References

THE DEVELOPER'S CONFERENCE

- **DataOps Ecosystem - //**medium.com/data-ops/2017-the-year-of-dataops-b2023c17d2af
- **DataOps for Government (State of Connecticut) -** //github.com/OpenDataCT/DataOps
- **Creating a Data-Driven Enterprise with DataOps - //**www.oreilly.com/data/free/creating-a-data-driven-enterprise-with-dataops.csp
- http://dataopsmanifesto.org/
- **DataOps—It's a Secret** - //www.datasciencecentral.com/profiles/blogs/dataops-it-s-a-secret
- **The Power of DataOps** - //www.delphix.com/blog/power-dataops
- **Building a DataOps Team** - //medium.com/data-ops/building-a-dataops-team-abc375e0a6bc
- **DataOps: Industrializing Data and Analytics -** //www.eckerson.com/articles/dataops-industrializing-data-and-analytics?content=dataops-industrializing-data-and-analytics

DataLakers
The Big Data Company

Concluindo....

THE DEVELOPER'S CONFERENCE

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

DataLakers
The Big Data Company

About me

Big Family

- Eduardo Hahn
  - Founder DataLakers Tecnologia
  - Data Lover & DataOps Enthusiastic
  - eduardo.hahn@datalakers.com.br
  - @eduardohahn
  - /in/eduardohahn3

Partners

cloudera talend ATTUNITY

DataLakers
The Big Data Company

THE DEVELOPER'S CONFERENCE

Trilha – Big Data

Passo a Passo para implementar DataOps em projetos de Big Data

Eduardo Hahn

DataLakers founder & DataOps Enthusiastic

DataLakers
The Big Data Company