4 DataOps

What is DataOps?

Agile approach to designing, implementing and maintaining a distributed data architecture that will support a wide range of open source tools and frameworks in production.

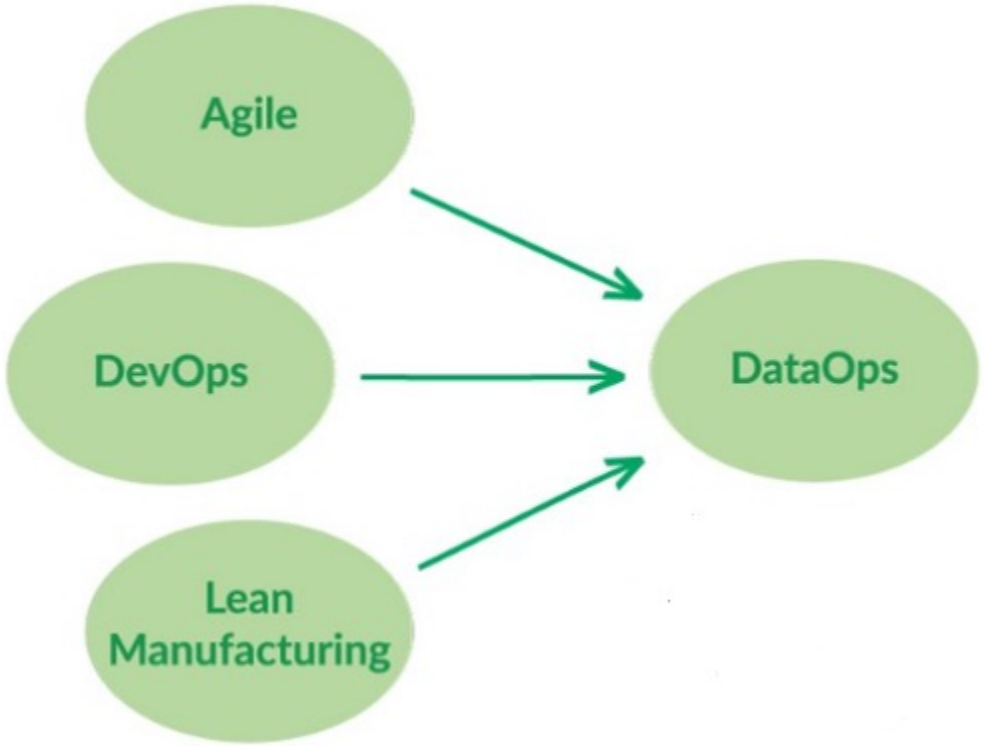
Goal:

Create business value from big data

How?

- Speed up innovation and experimentation to deliver insights from data
- Maintain high data quality and very low error rates
- Enhance collaboration across people, technology, and environments
- Enforcers clear measurement, monitoring, and transparency of results

DataOps origins



Agile

Enables organizations to respond rapidly to customer requirements and accelerate time to value.

Lean Manufacturing

Focuses on the minimization of waste within a system without sacrificing productivity.

DevOps

Accelerates the build lifecycle using automation

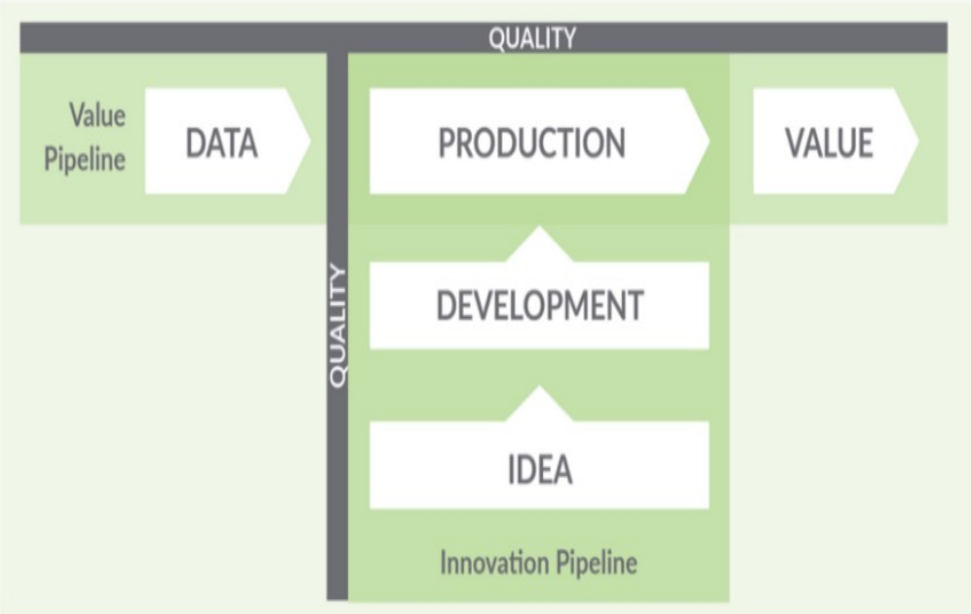
DataOps solutions

Main focus of DataOps is:

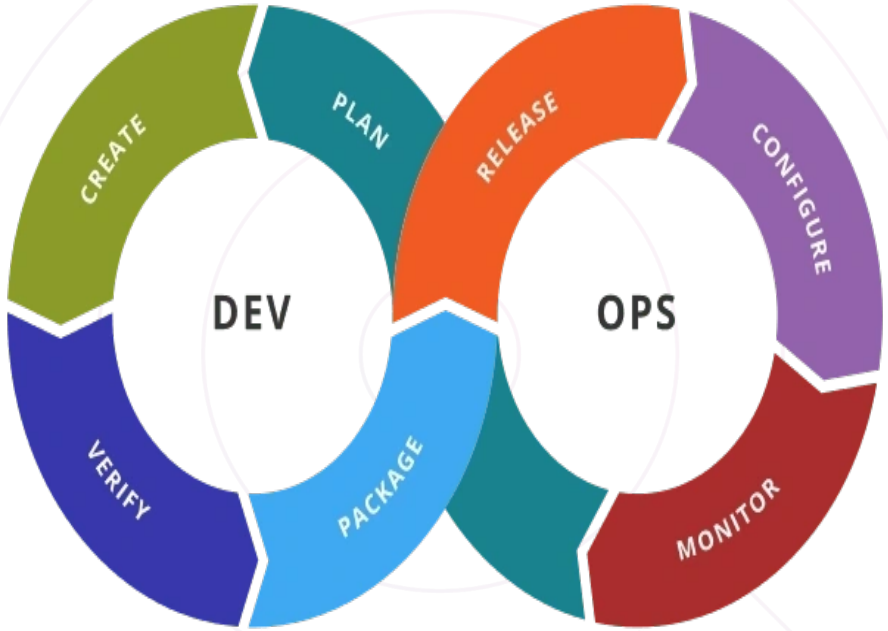
- Improve workflow in data teams
- Enhance collaboration across different data groups
- Improve access to data
- Speed up the release process
- Improve data architectures
- Alleviate process bottle necks
- Identify and reduce technical debt
- Guarantee quality at every step

Not just DevOps for data

DataOps



≠



Statistical Process Control (SPC)

DataOps users

Software Engineer

Coding in complex set of tools

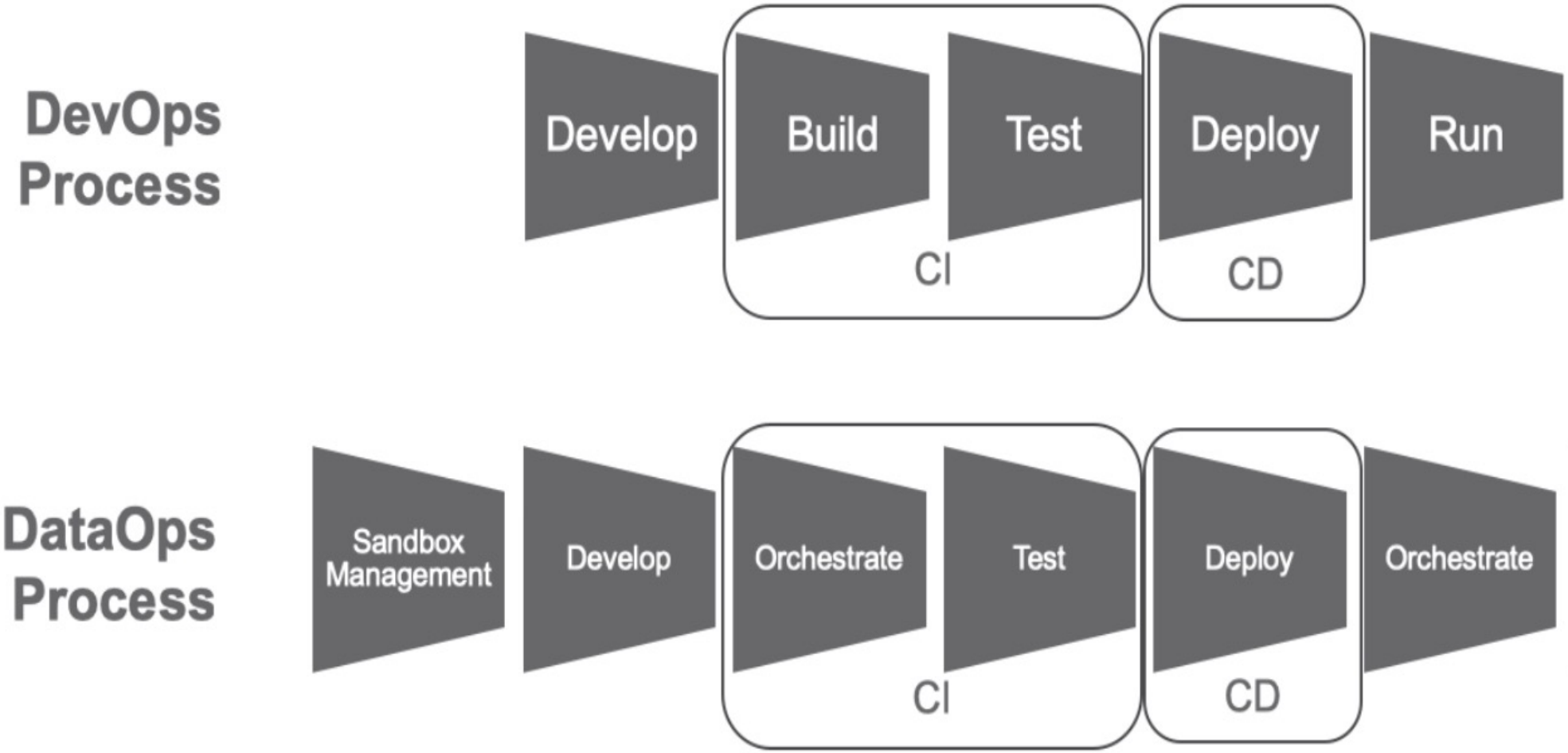
Love technology

Data Scientist and Analyst

Analyse data and build models

Technology is a means to an end

The process


DevOps Process
Develop
Build
Test
CI
Deploy
CD
Run
DataOps Process
Sandbox Management
Develop
Orchestrate
Test
CI
Deploy
CD
Orchestrate
Monitor

Importance of testing

Quality of Solution = $f(data, code)$

	Data Fixed	Data Variable
Code Fixed		Value Pipeline
Code Variable	Innovation Pipeline	

Benefits of DataOps

Enhances collaboration

- Sets collaboration parameters for cross-functional teams
- Facilitates a 360 view of execution by enforcing rigorous planning

Enforces robust solutions

- Removes human unpredictability from the equation
- Solutions are built thinking about reliability

Offers flexibility

- Well-defined processes allow adaptability
- Reduces time to move changes across systems

Incorporates the Agile mind-set

- Which comes with all benefits of the agile framework
- If you already practiced agile, easier to incorporate DataOps

Challenges of adopting DataOps

Fragmented Organizations

- DataOps helps reducing the effect of departmental silos
- Planning and collaboration across departments are key

Steep Learning curve

- Technology changes fast and upskilling is not always easy
- Training should be at the center of a mature DataOps roadmap

Choosing the right tools

- Build some buy some strategy is the most common
- When choosing a tool, think about integration and scalability

There is not one-size-fits-all solution

- None single solution for everything that you will need
- Achieving maturity requires time, investment and some research