DATA+AI SUMMIT 2022

Automating Model Lifecycle Orchestration with Jenkins

ORGANIZED BY databricks

Liam Newman
Sr. Software Engineer, Verta

Conrado Miranda
CTO, Verta

Who are we?

Liam Newman

Sr. Software Engineer, Verta

- Jenkins Pipeline at Cloudbees
- Jenkins Declarative Pipeline
- CI/CD/DevOps Evangelist

Conrado Miranda

CTO, Verta

- ML platform at Twitter
- Supported 100+ DSes
- Nvidia self-driving car team

Agenda

01

Context of software lifecycle

02

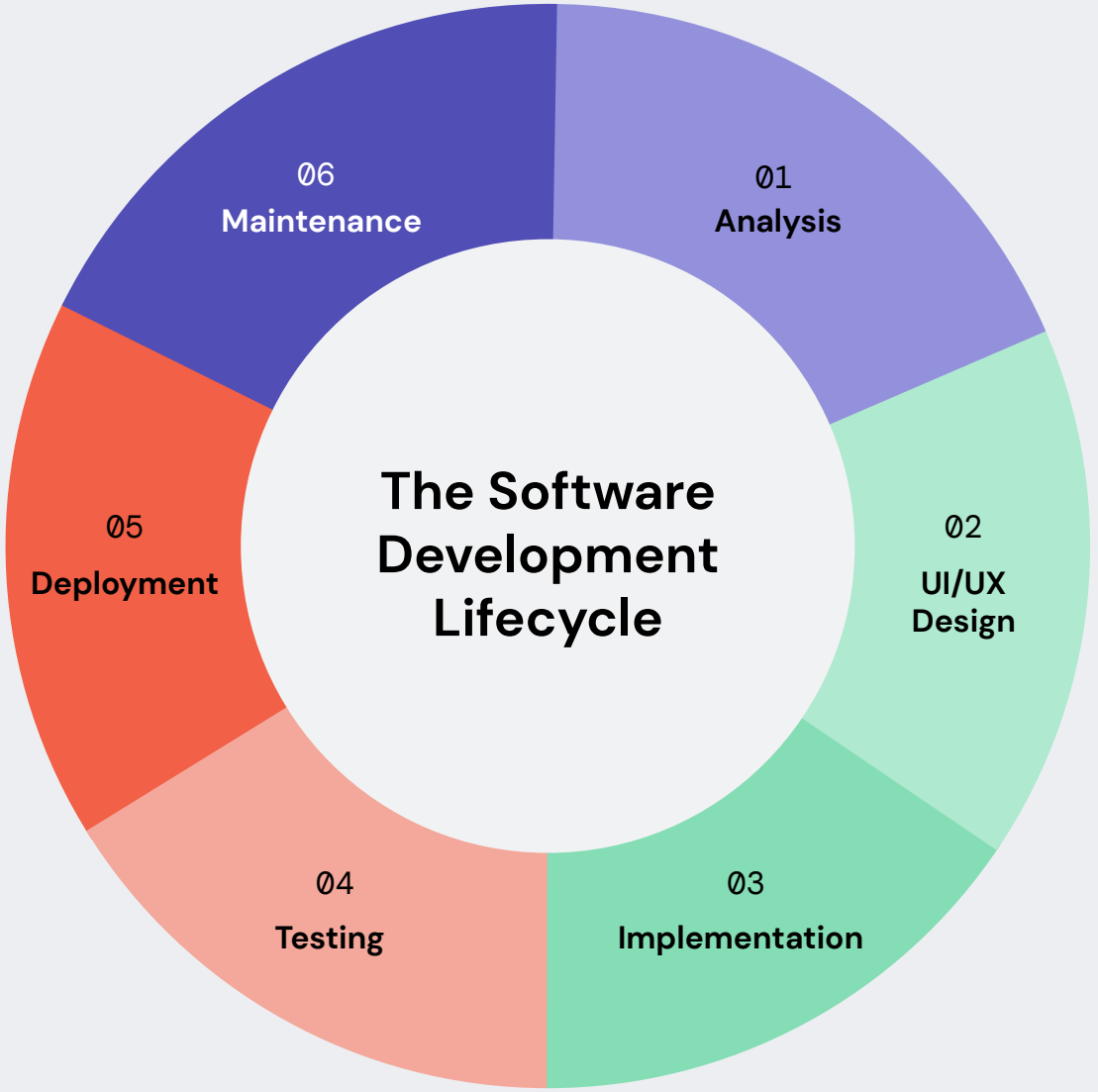
CI/CD, DevOps and Jenkins

03

Industry examples from customers

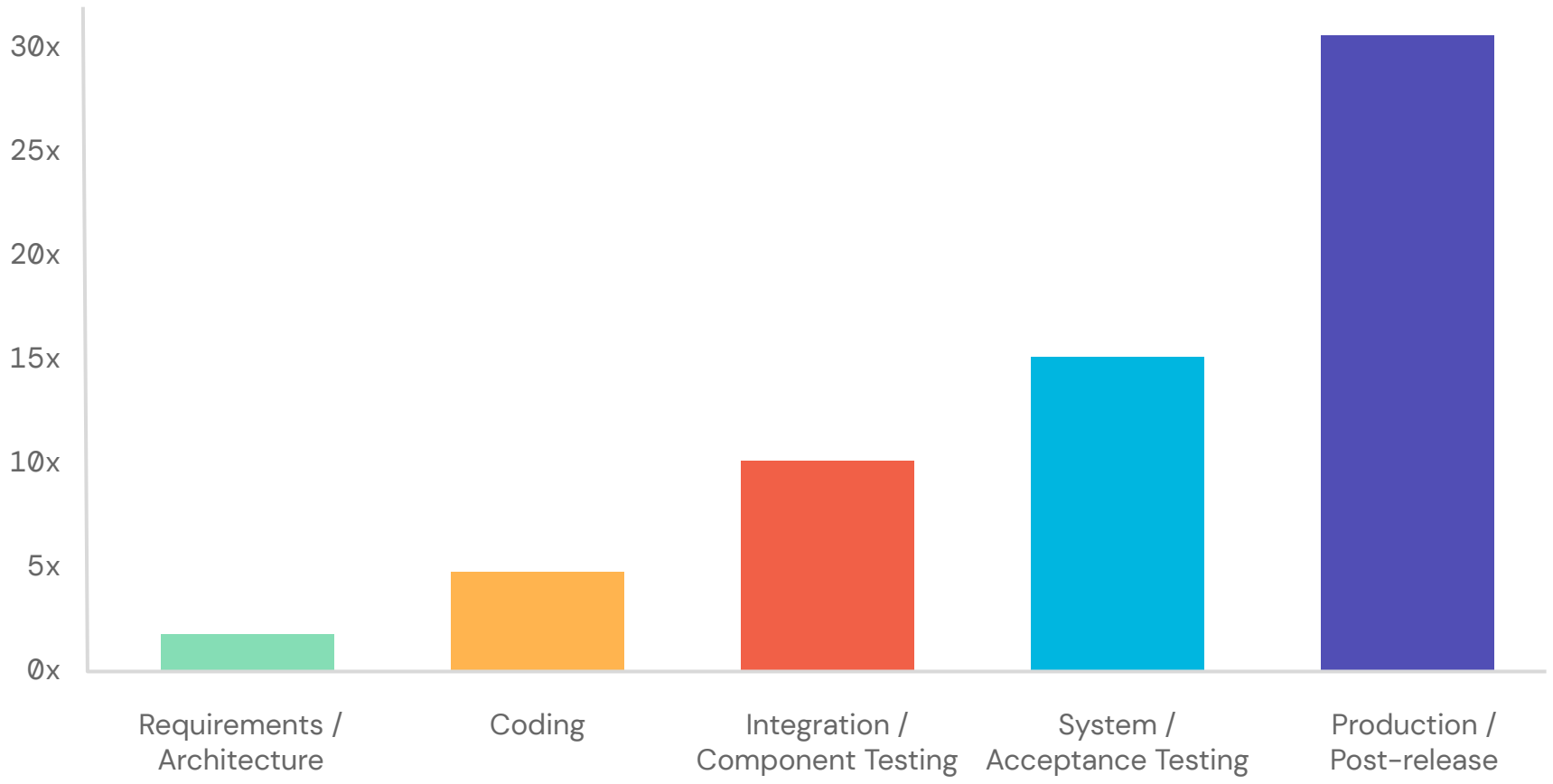
Software Lifecycle

The Software Lifecycle



Source: AgileTech - Top 6 Software Lifecycle Models & Methodologies

The Relative Cost to Fix Bugs



The cost is **30x higher** if these defects are not fixed early on.

Source: Deep Source IO - Exponential Cost to Fix Bugs

CI/CD, DevOps and Jenkins

What is CI/CD?

Continuous Integration

Automated build and test pipeline run on every change before it is merged

Continuous Delivery

CI plus automated publishing of artifacts

Continuous Deployment

Continuous Delivery plus automated deployment to production

What about “DevOps”?

Combination of tools/automation, process and culture

What is Jenkins?

- A general purpose automation tool
- Broad plugin system with tons functionality
- Coordinates actions performed on multiple machines
- Used for implementing CI/CD pipelines

Best Practices

01 Automate your processes, write them as code

02 Commit early, commit often

03 Build only once

04 Automate and streamline your tests

05 Clean your environments

06 Make CI/CD pipeline the only way to deploy to production

07 Monitor and measure your pipeline

08 Make it a team effort

DS/ML and CI/CD

01

DS/ML is still software

02

DS/ML is a component of larger software products and services

03

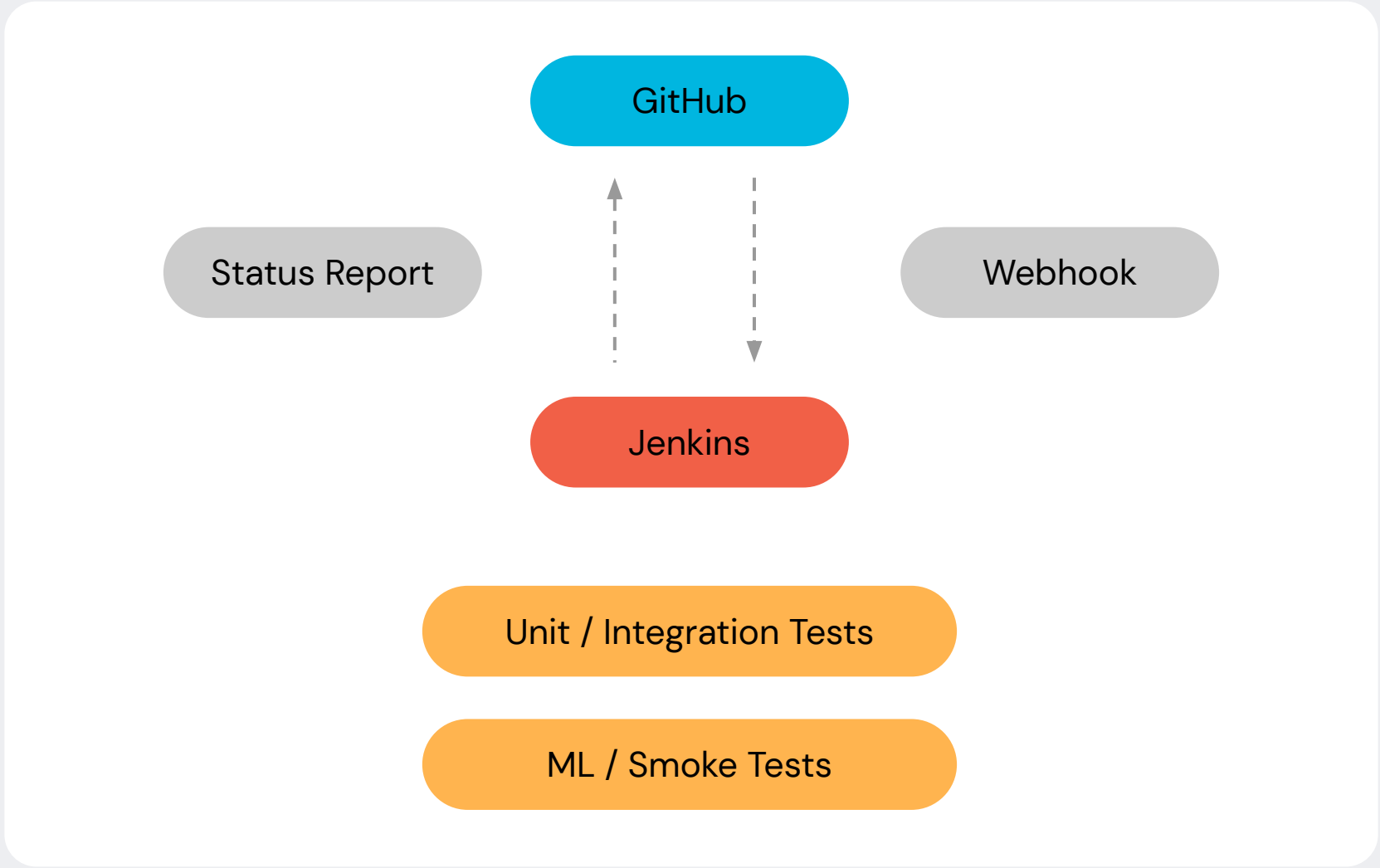
DS/ML needs to be part of software CI/CD pipeline

Why DS/ML Struggles with CI/CD

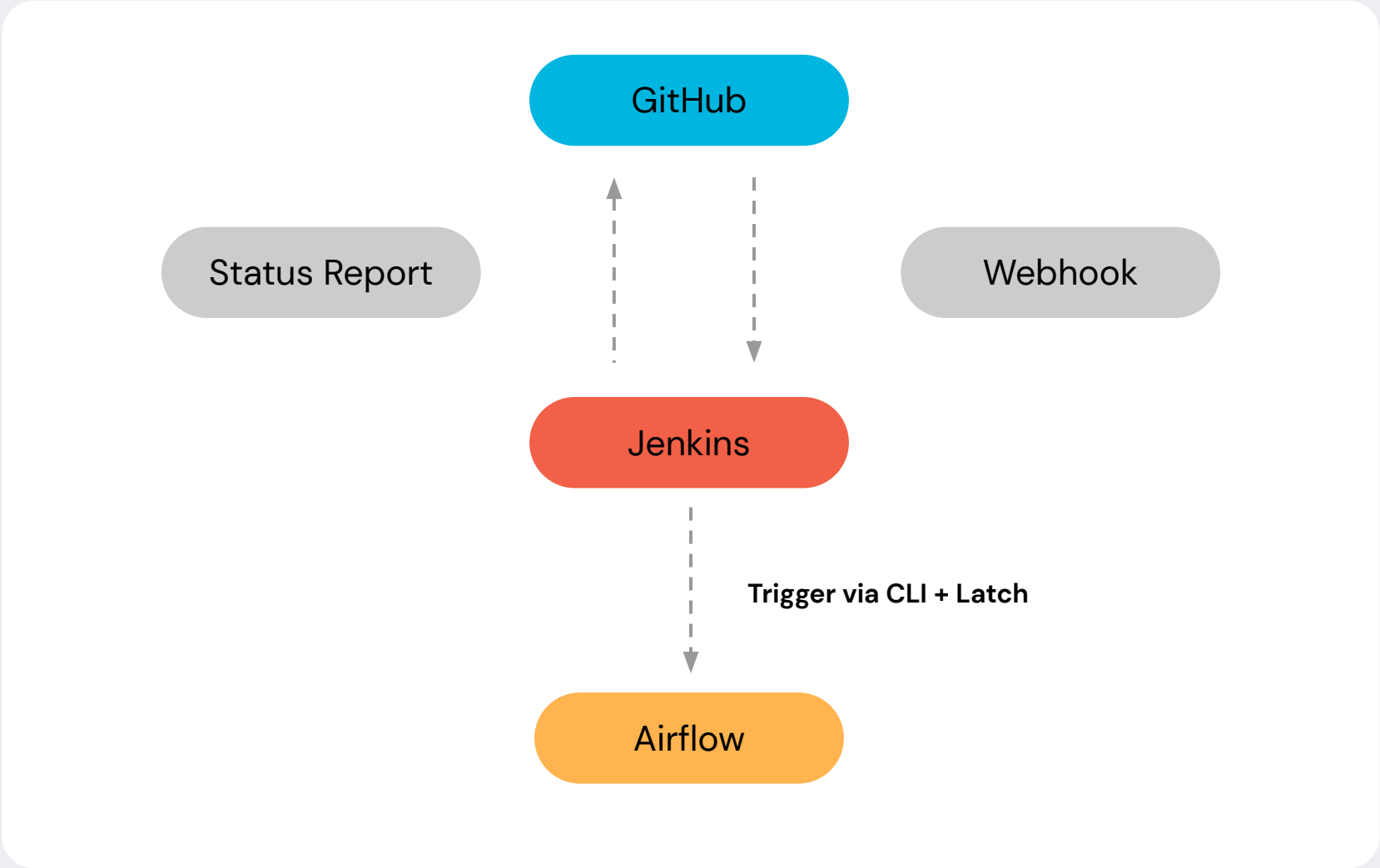
- ✕ Different standards
- ✕ Different resulting artifacts
- ✕ Missing tools and infrastructure
- ✕ Different training
- ✕ Different tools
- ✕ Missing knowledge and training

Examples

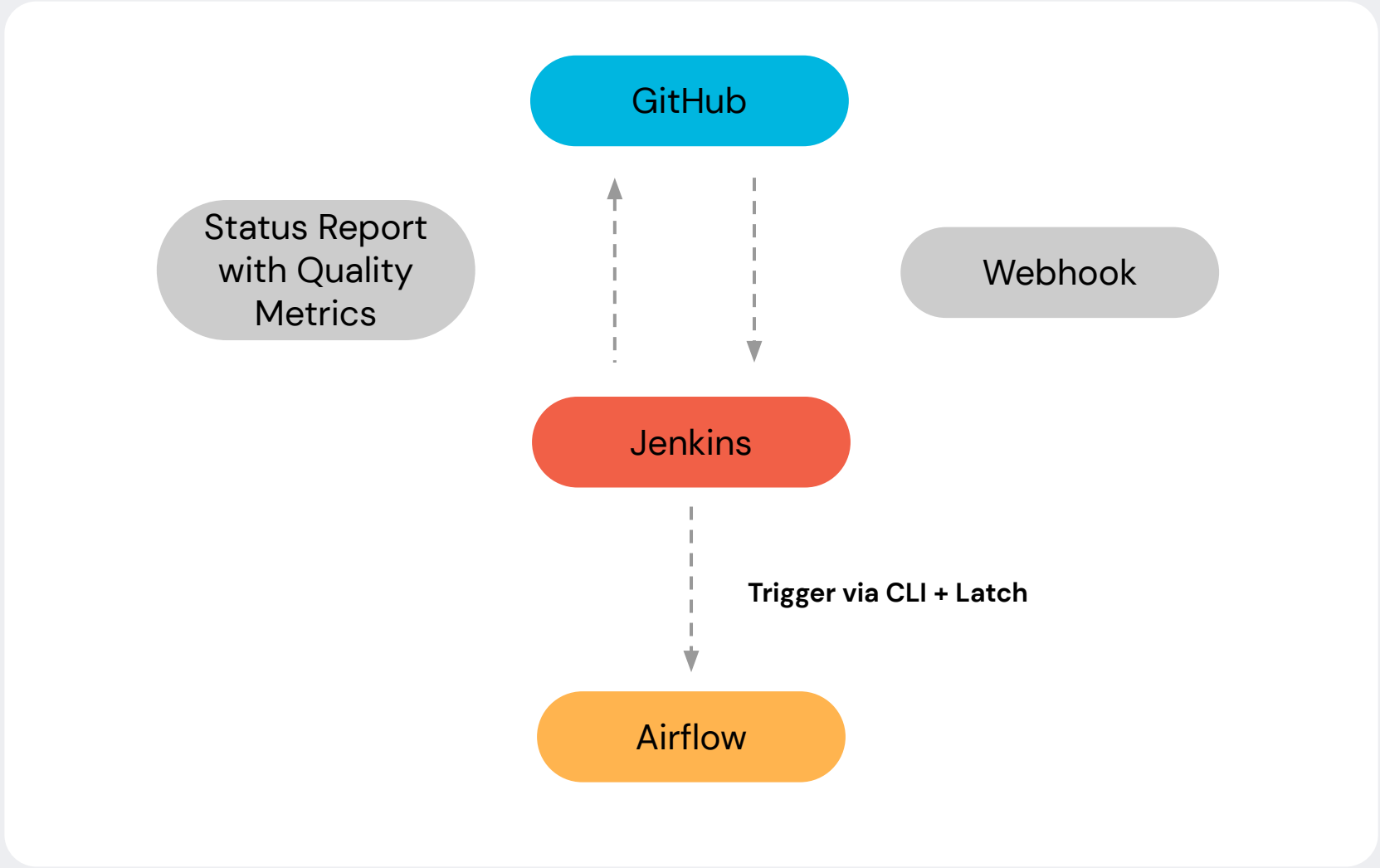
Testing ML code



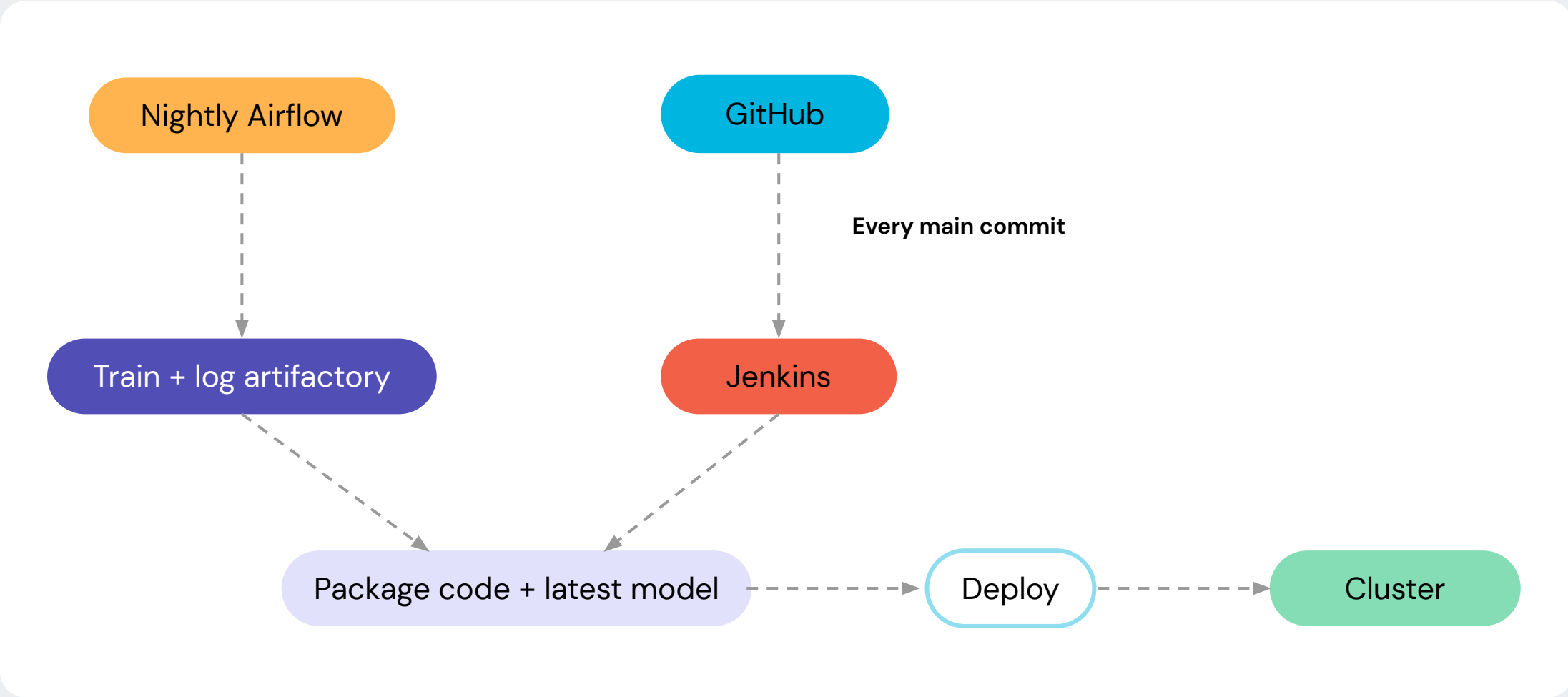
Verifying data pipeline



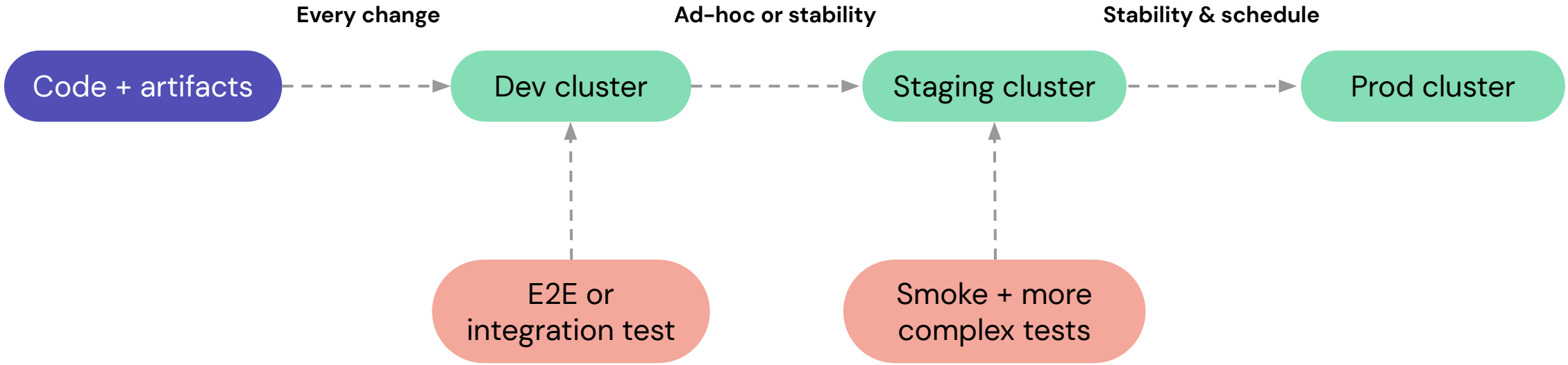
Verifying data pipeline quality



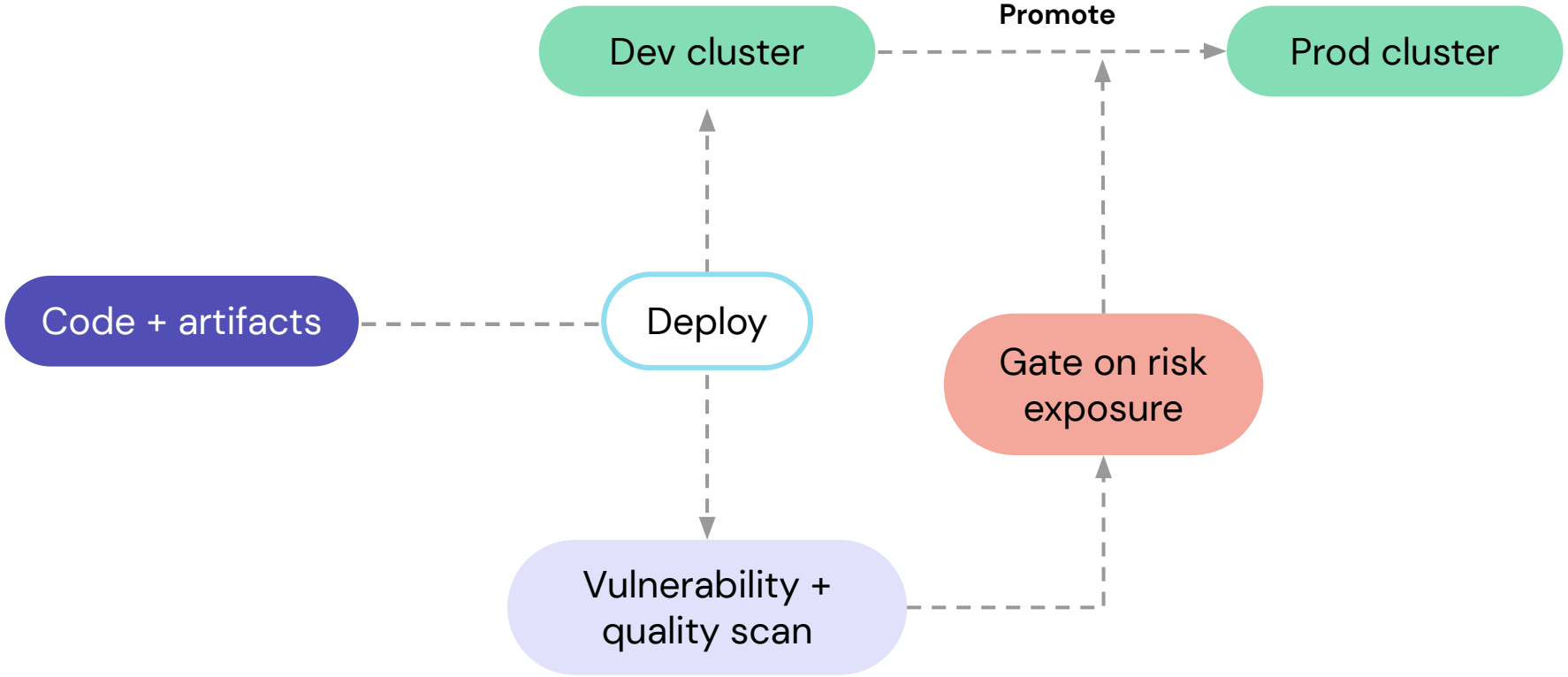
Integrating inference code and model artifacts



Model promotion: dev->staging->prod



Model promotion with compliance gates



Summary

DS/ML and CI/CD folks should work together

Devops folks want everyone at the company to do good processes.

DS/ML live in their own land. They need to become a part of the developer ecosystem at the company.

A lot of orchestration problems from DS/ML have already been solved for SW. Ask someone about it over coffee!

Bring CI/CD to your DS/ML systems

We know what we're doing! Reach out to us to talk about this.

Learn more at Verta.ai

Thanks!

Reach out anytime

liam@verta.ai

conrado@verta.ai