
AWS
re:Invent

WPS301

Designing AI/ML applications for impact

Eric Greene

Lead AI/ML Solutions Architect,
Public Sector Asia-Pacific and Japan

AWS

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The AWS Well-Architected Framework helps you understand the pros and cons of decisions you make while building systems on AWS

Operational excellence

Security

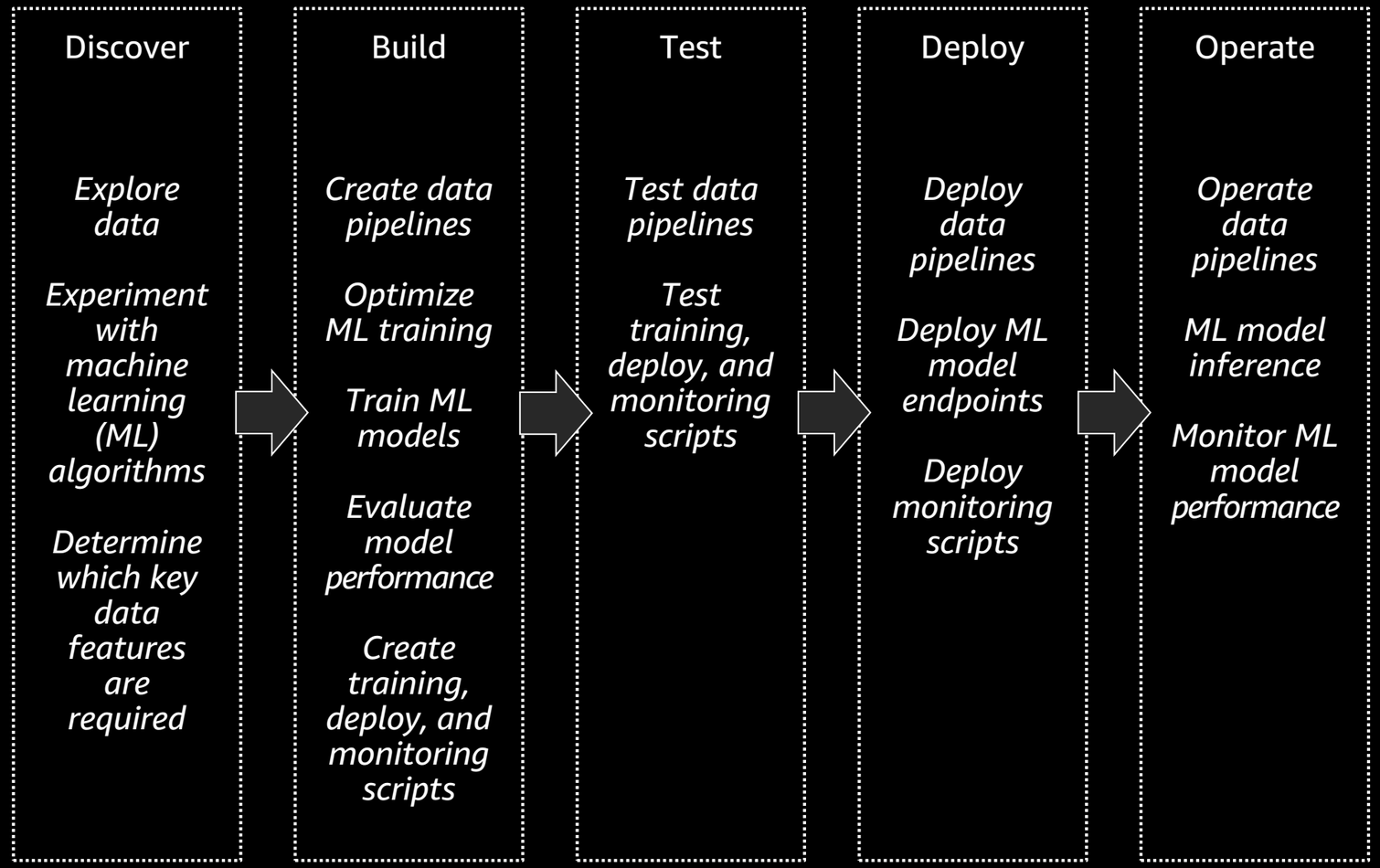
Reliability

Performance efficiency

Cost optimization

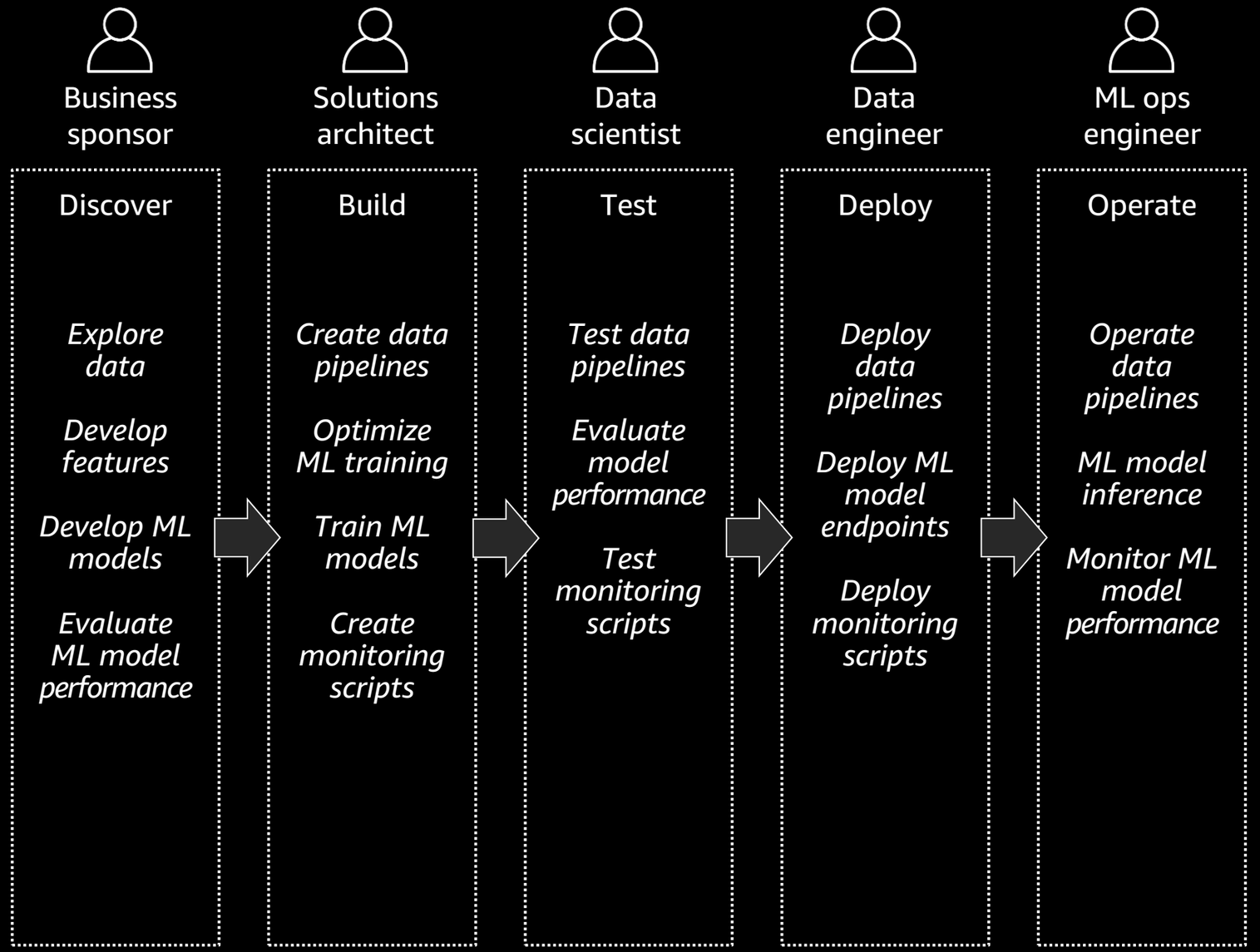
Operational excellence

Run, monitor, and gain insights into systems to deliver business value and continually improve supporting processes and procedures



Operational excellence

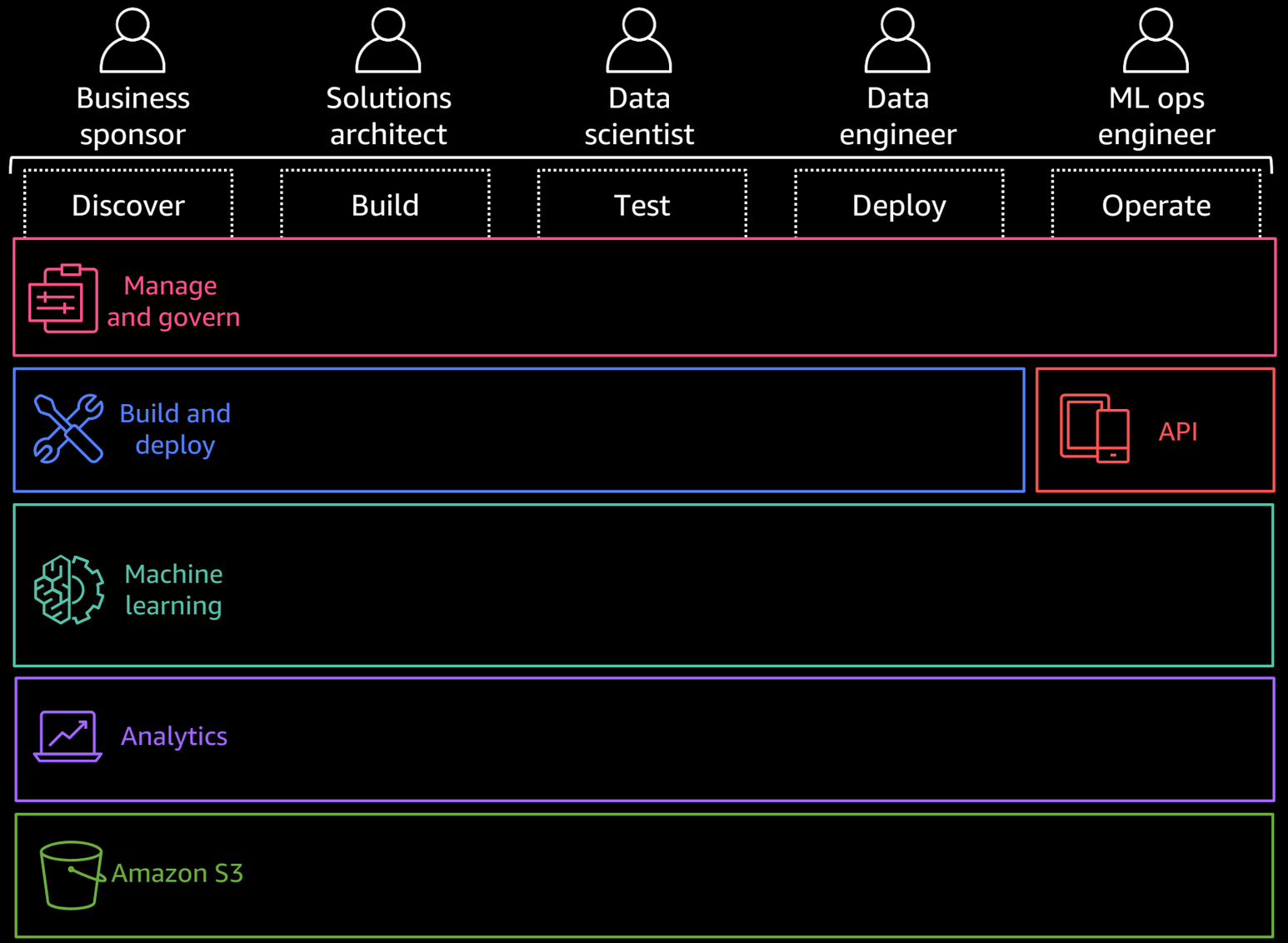
Establish cross-functional teams



Operational excellence

Establish cross-functional teams

Identify the end-to-end architecture and operational model early

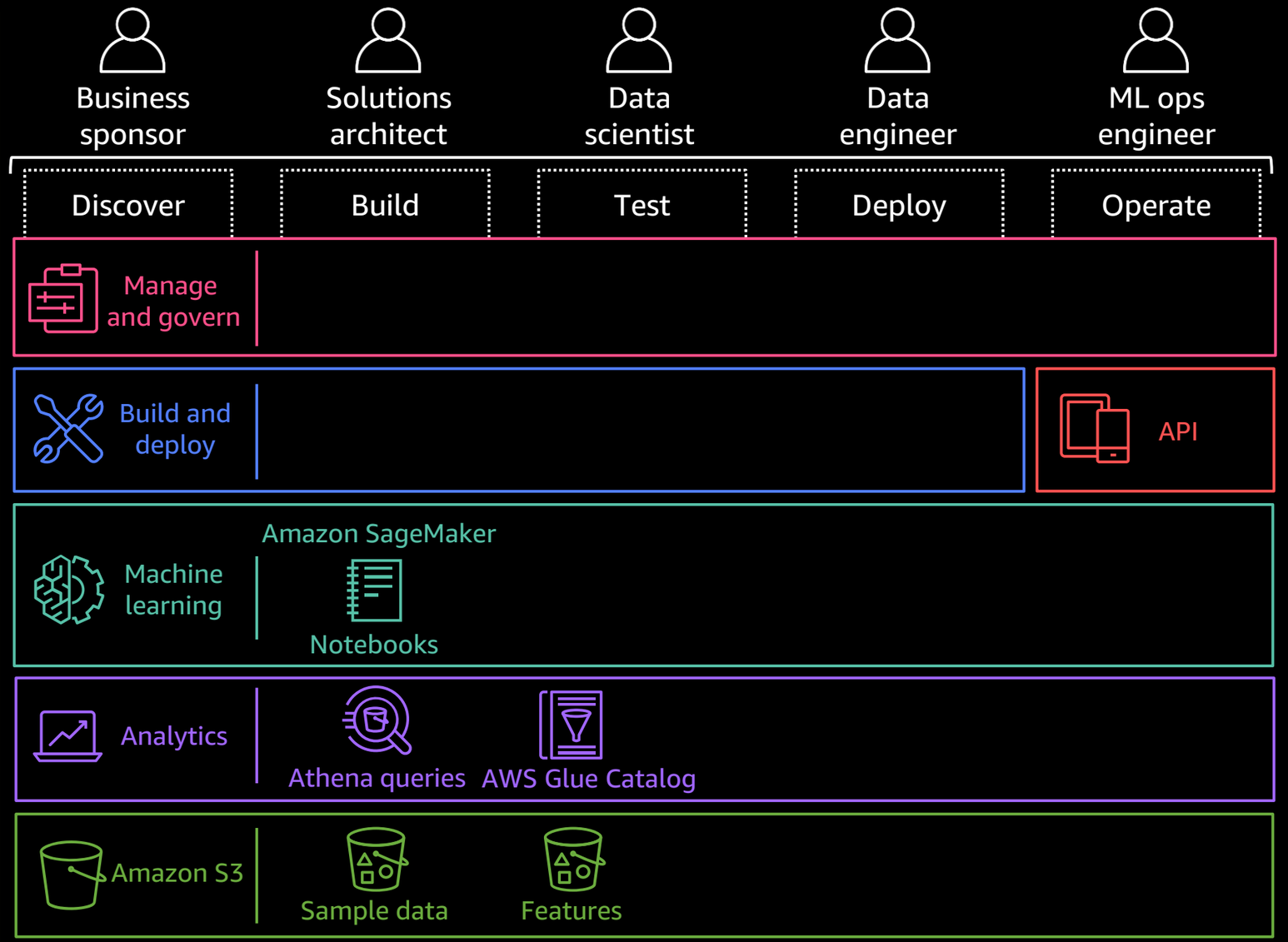


Operational excellence

Establish cross-functional teams

Identify the end-to-end architecture and operational model early

Document ML discovery activities and findings



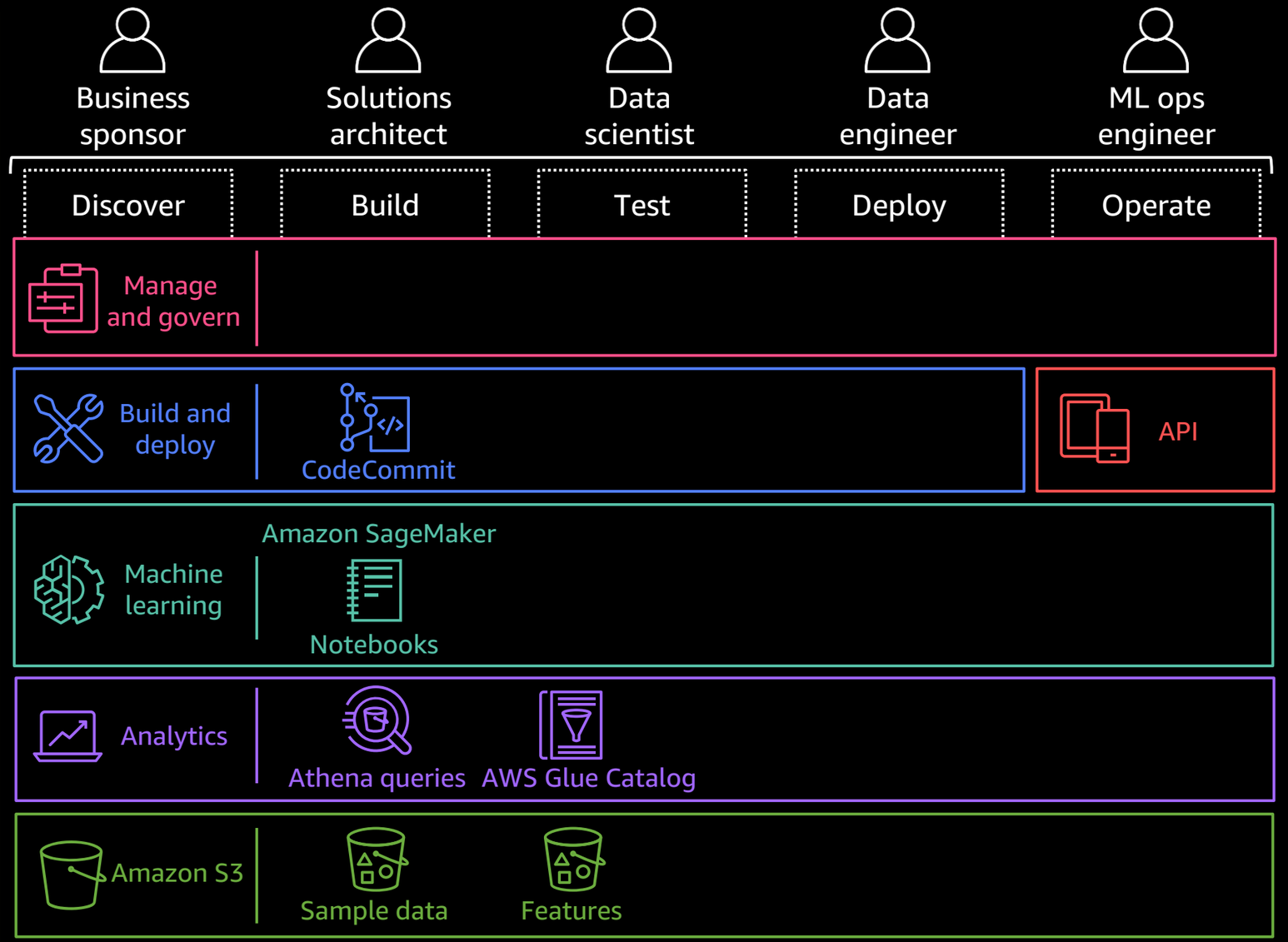
Operational excellence

Establish cross-functional teams

Identify the end-to-end architecture and operational model early

Document ML discovery activities and findings

Version ML inputs and artefacts



Operational excellence

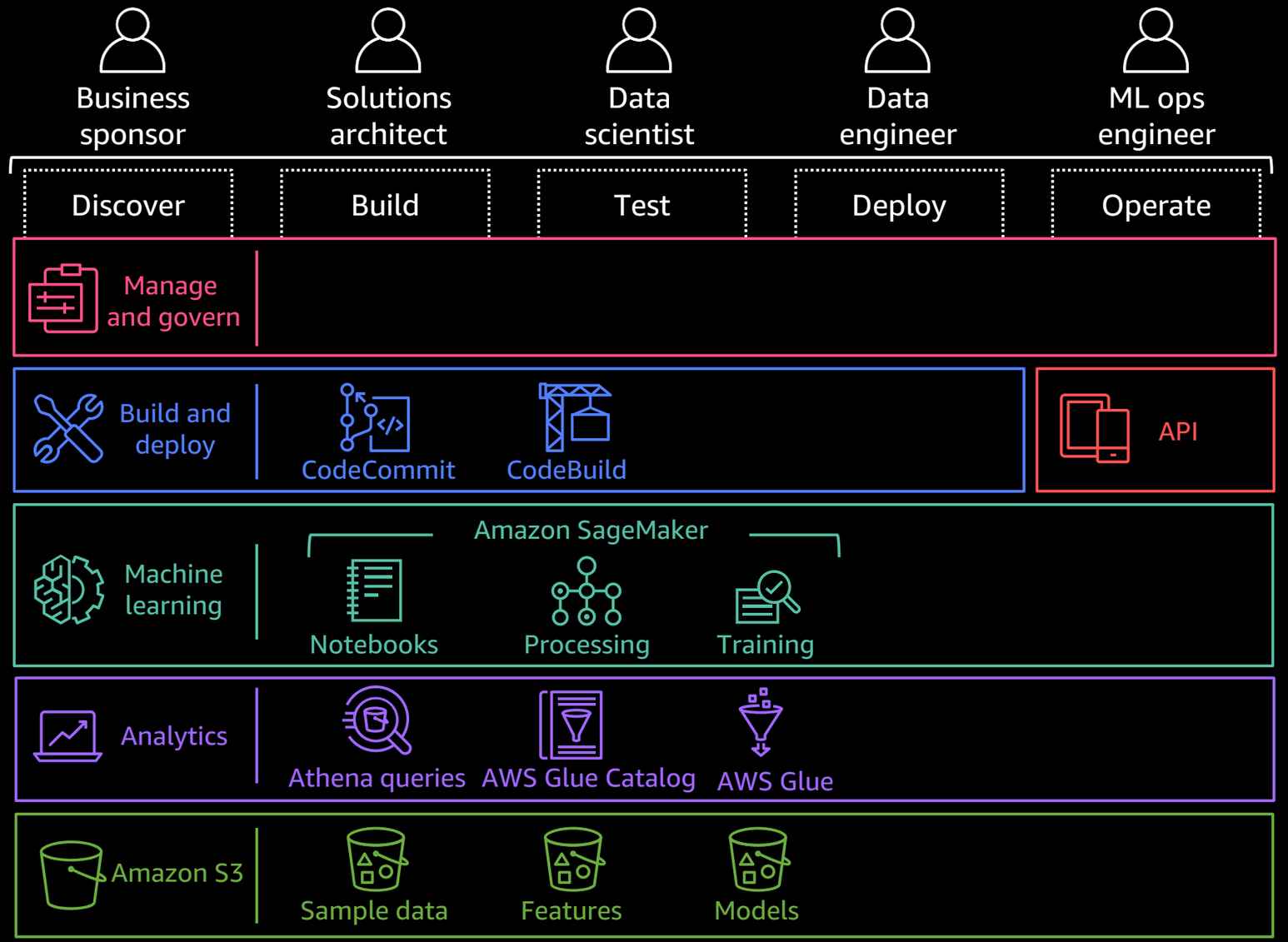
Establish cross-functional teams

Identify the end-to-end architecture and operational model early

Document ML discovery activities and findings

Version ML inputs and artefacts

Establish a model retraining strategy



Operational excellence

Establish cross-functional teams

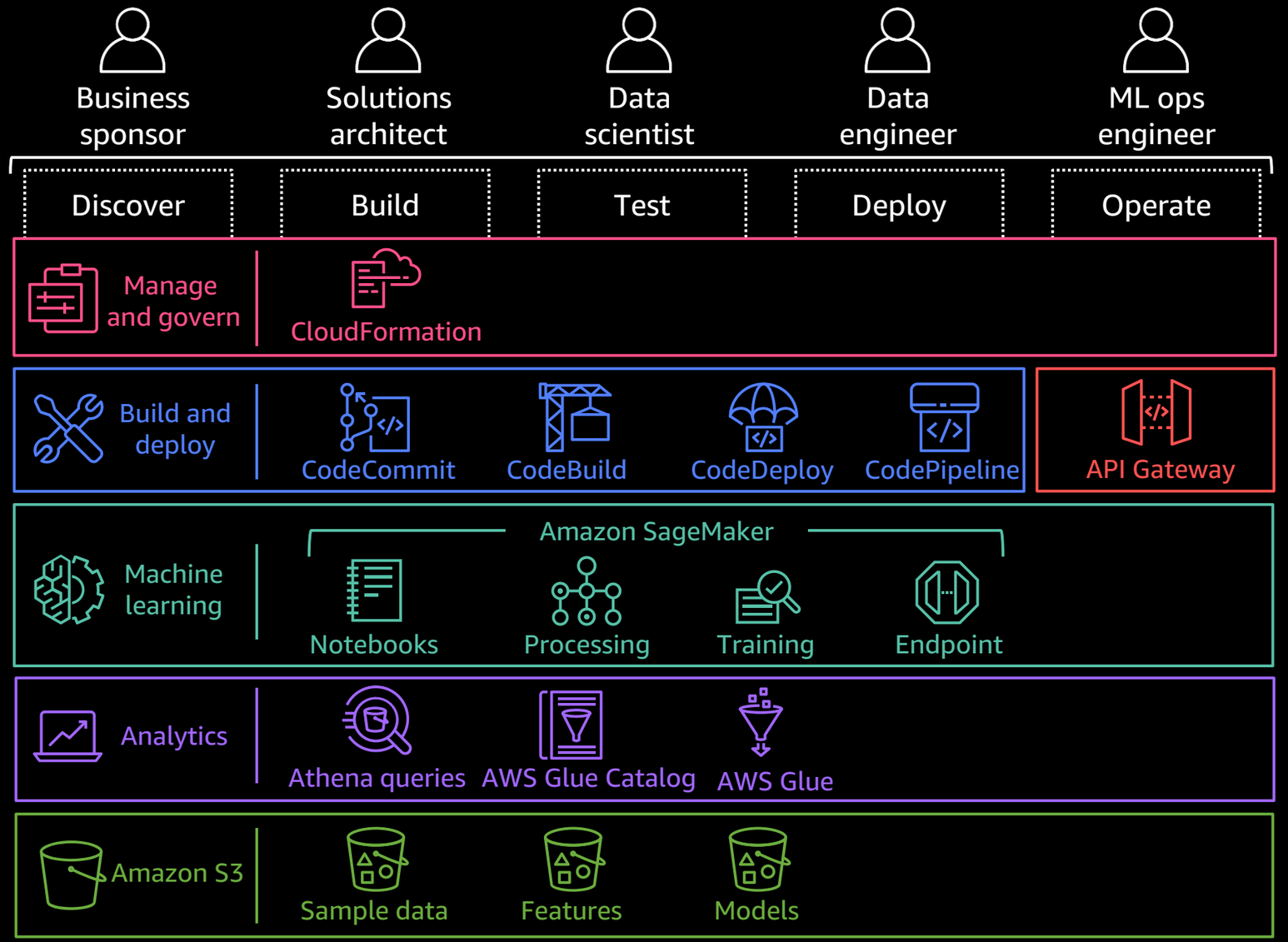
Identify the end-to-end architecture and operational model early

Document ML discovery activities and findings

Version ML inputs and artefacts

Establish a model retraining strategy

Automate ML deployment pipelines



Operational excellence

Establish cross-functional teams

Identify the end-to-end architecture and operational model early

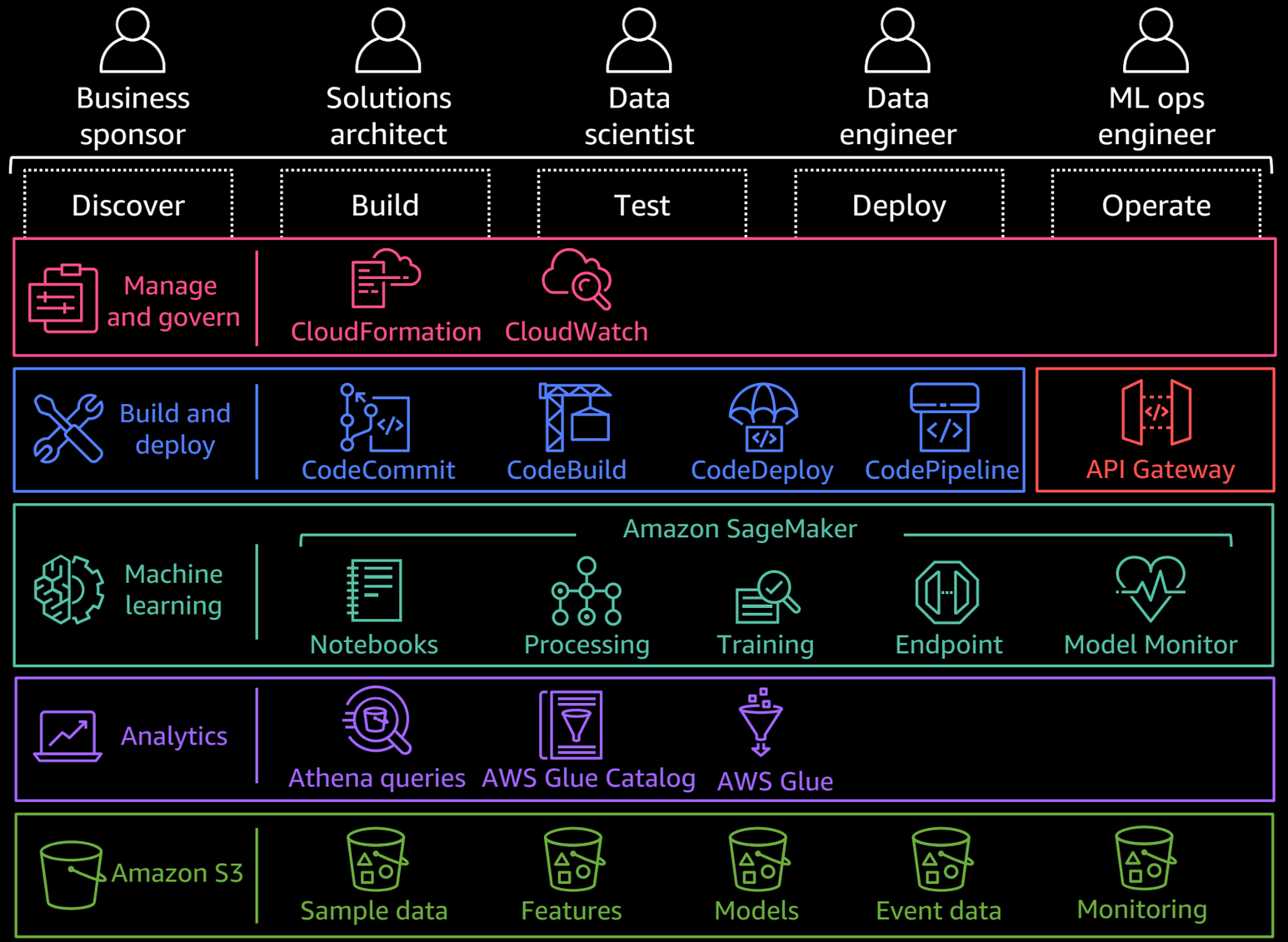
Document ML discovery activities and findings

Version ML inputs and artefacts

Establish a model retraining strategy

Automate ML deployment pipelines

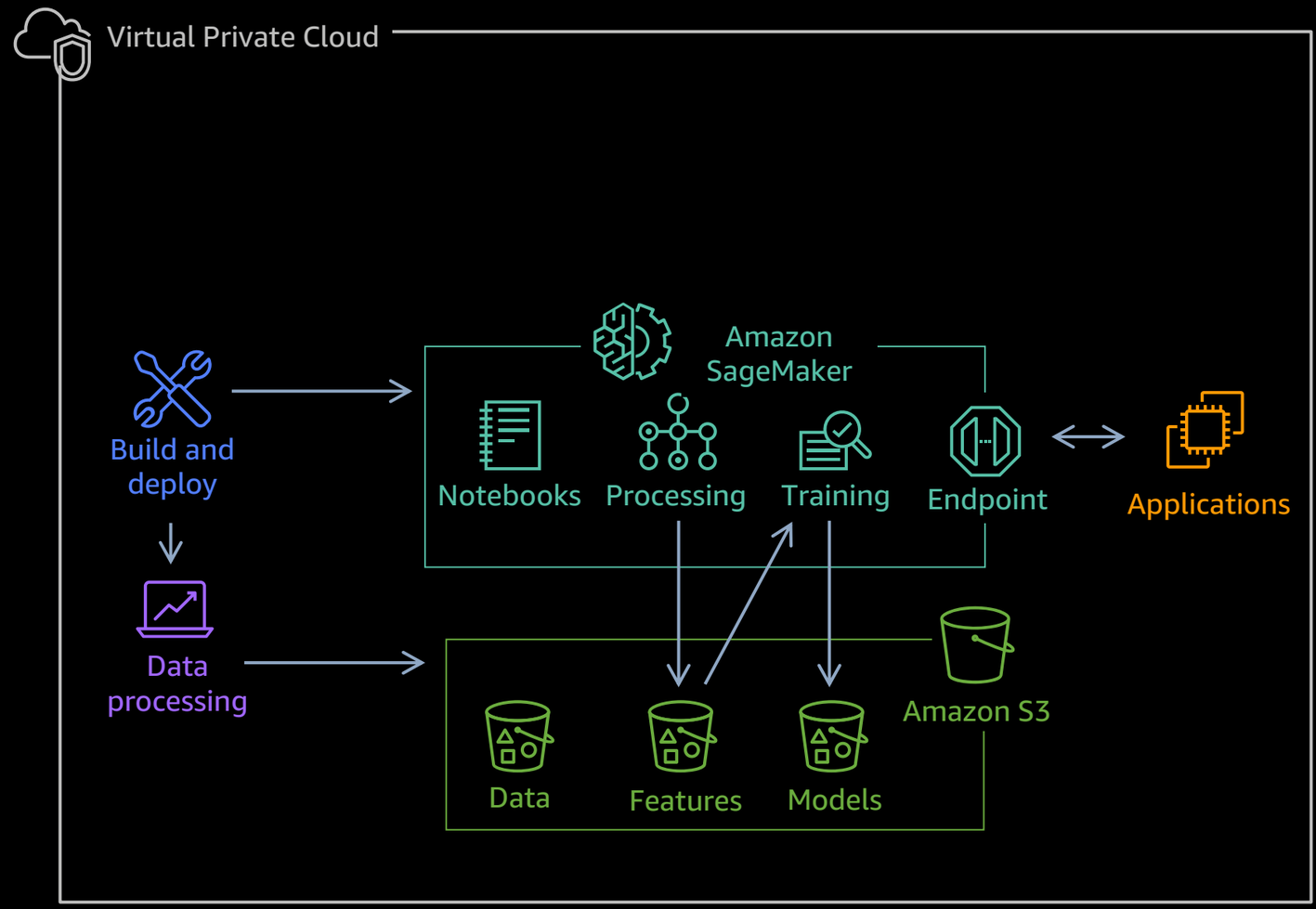
Continuously monitor and measure ML workloads



Operational excellence

Security

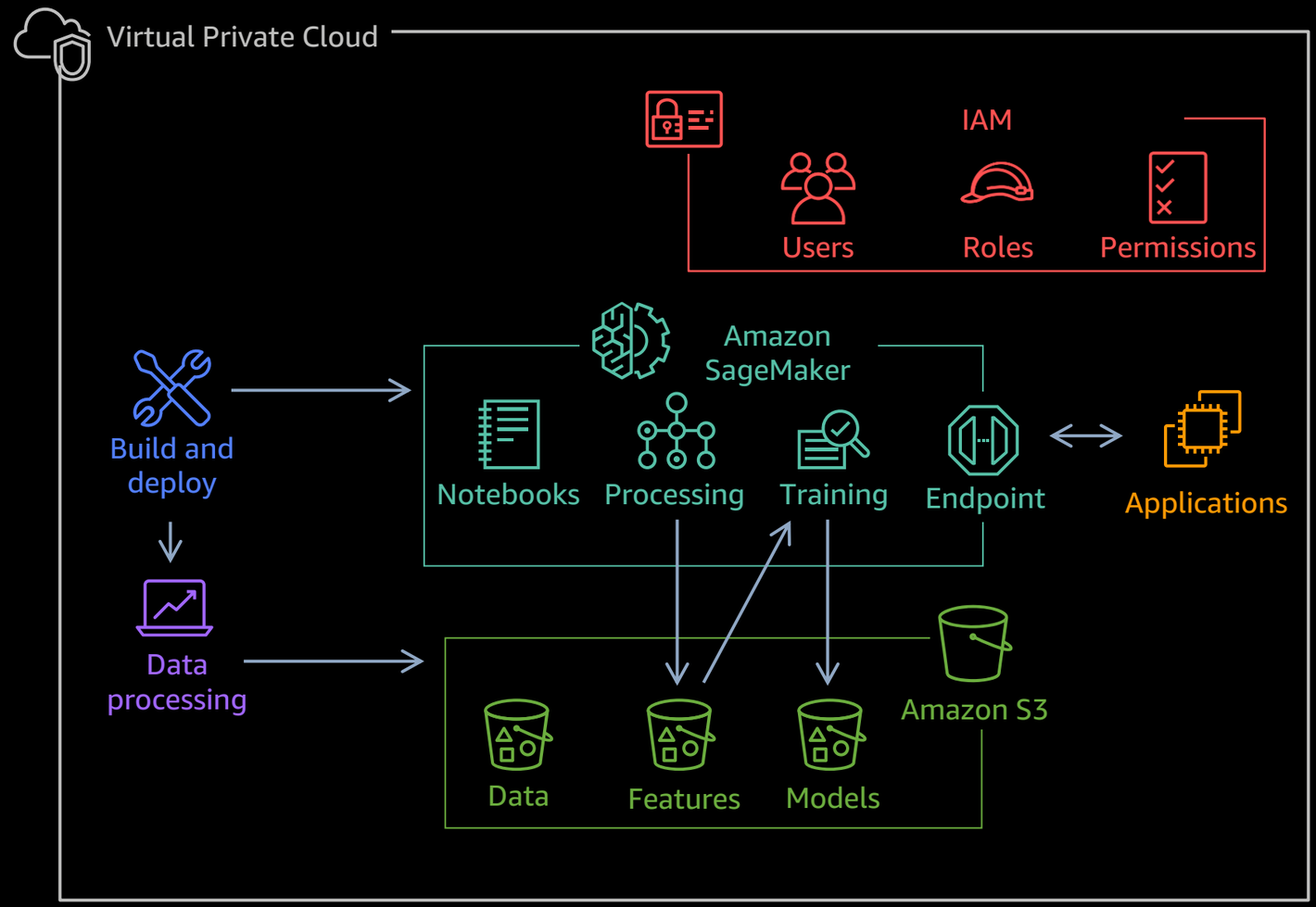
Protect information, systems, and assets while delivering business value through risk assessments and mitigation strategies



Operational excellence

Security

Restrict access to ML systems

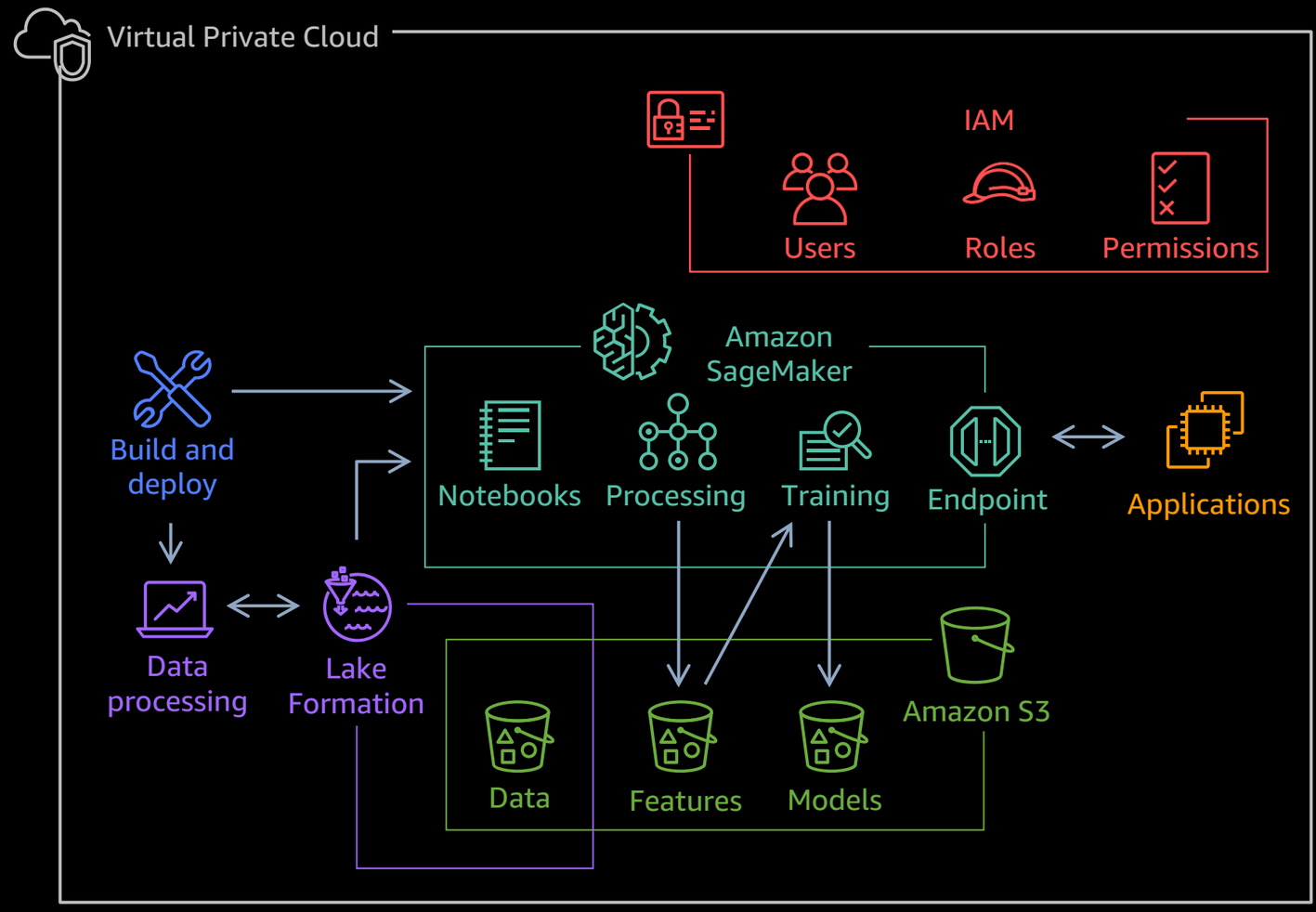


Operational excellence

Security

Restrict access to ML systems

Provide data governance



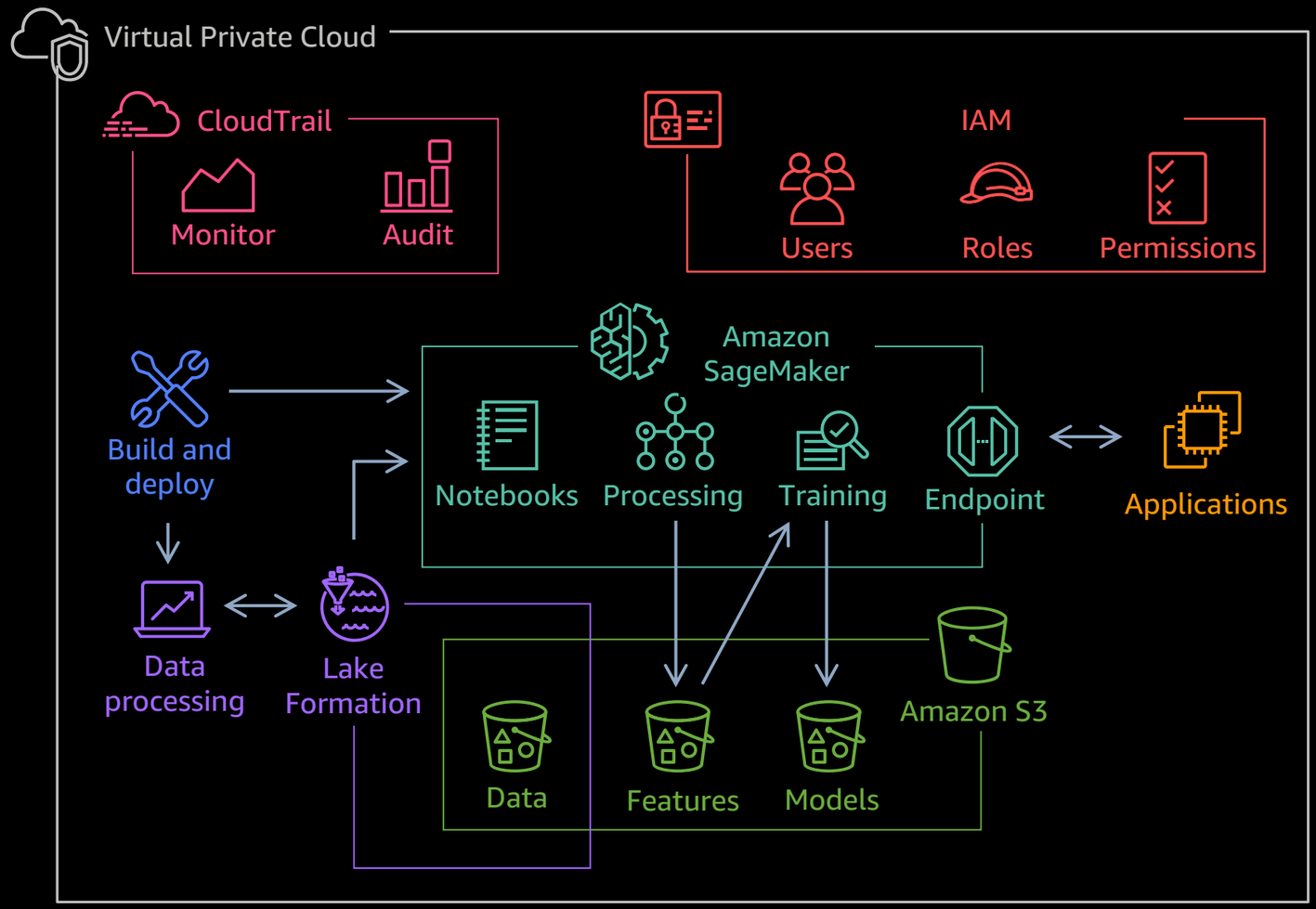
Operational excellence

Security

Restrict access to ML systems

Provide data governance

Enforce data lineage



Operational excellence

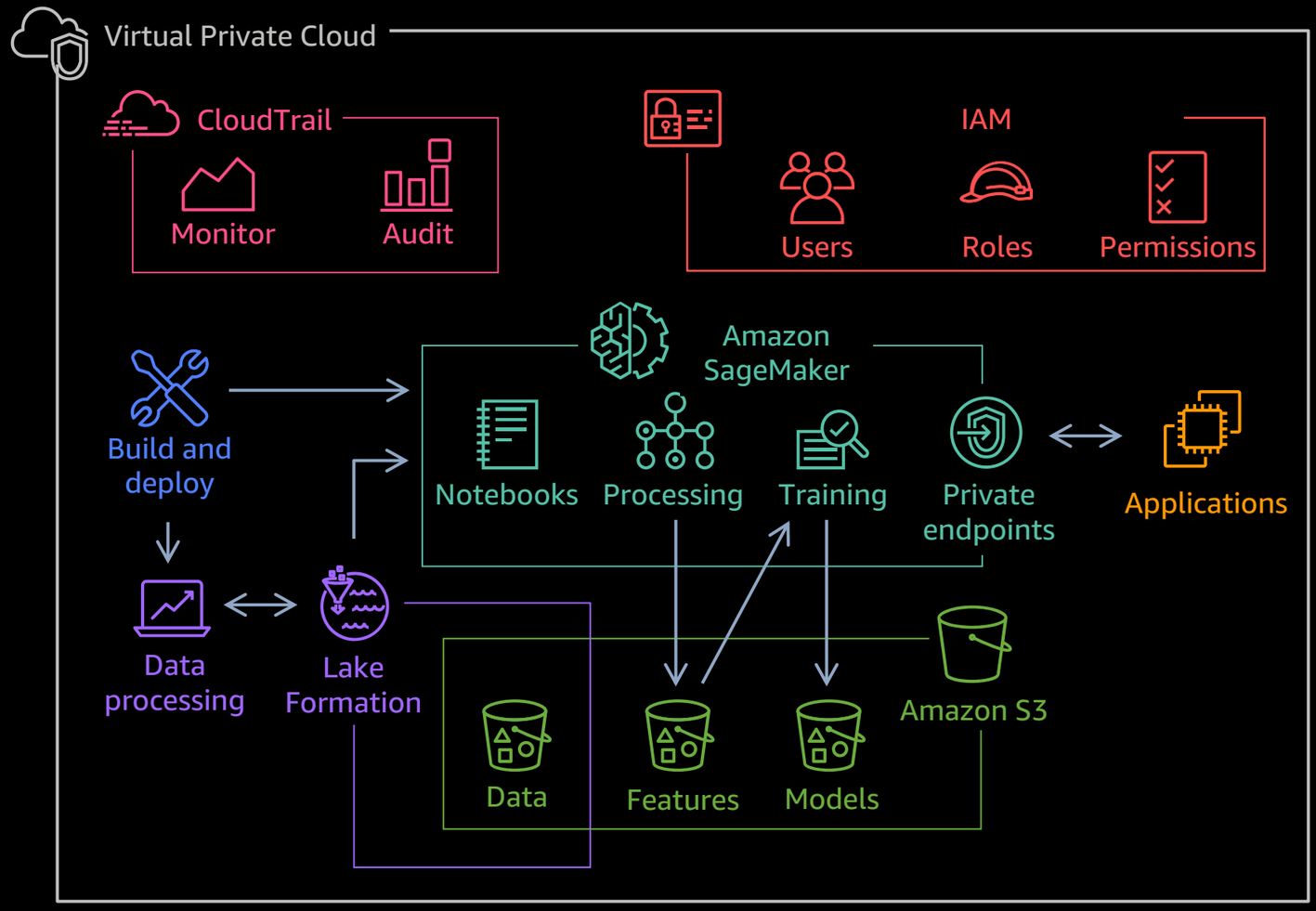
Security

Restrict access to ML systems

Provide data governance

Enforce data lineage

Meet regulatory compliance

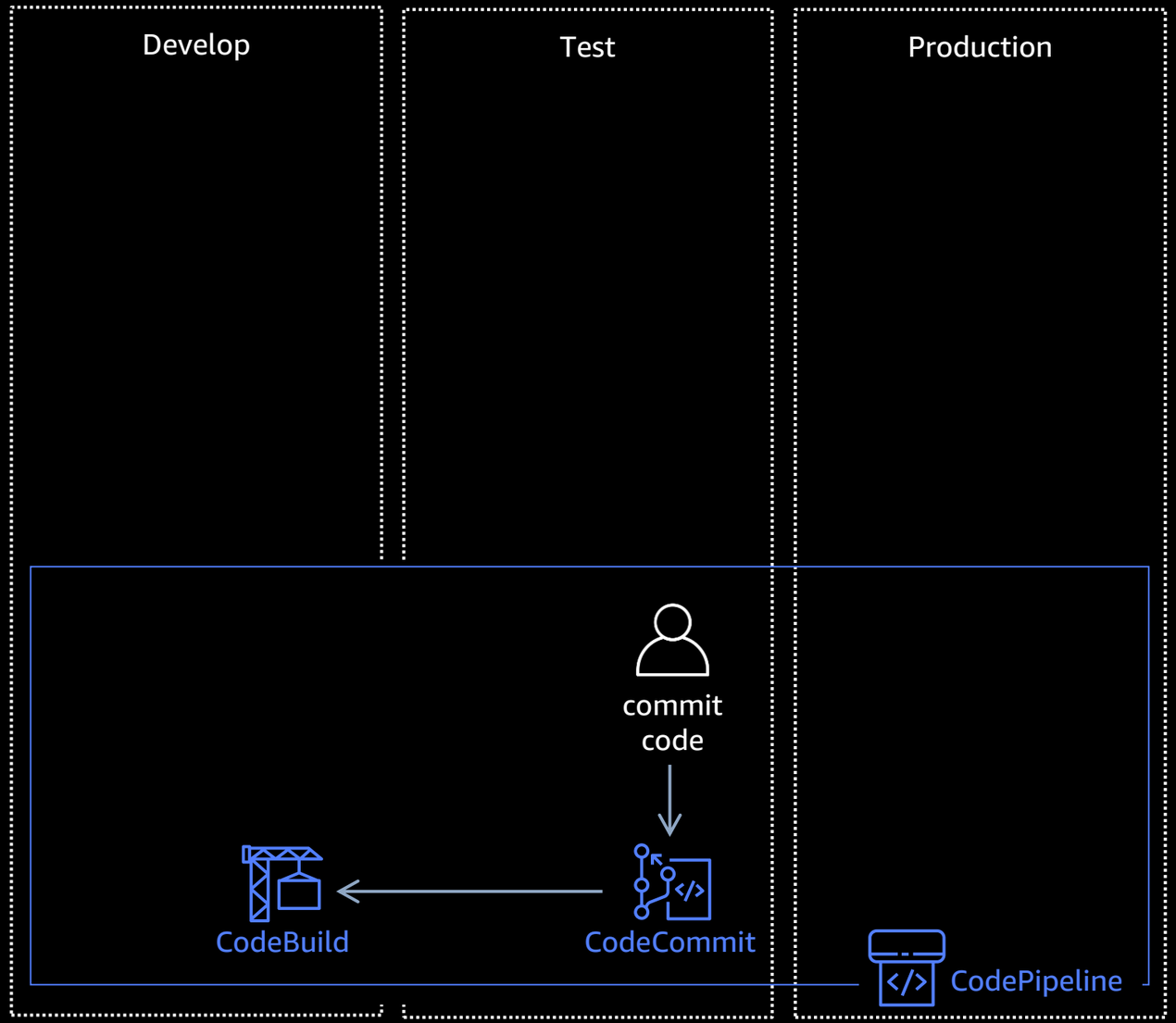


Operational excellence

Security

Reliability

Recover from infrastructure or service disruptions, dynamically acquire computing resources to meet demand, and mitigate disruptions such as misconfigurations or transient network issues

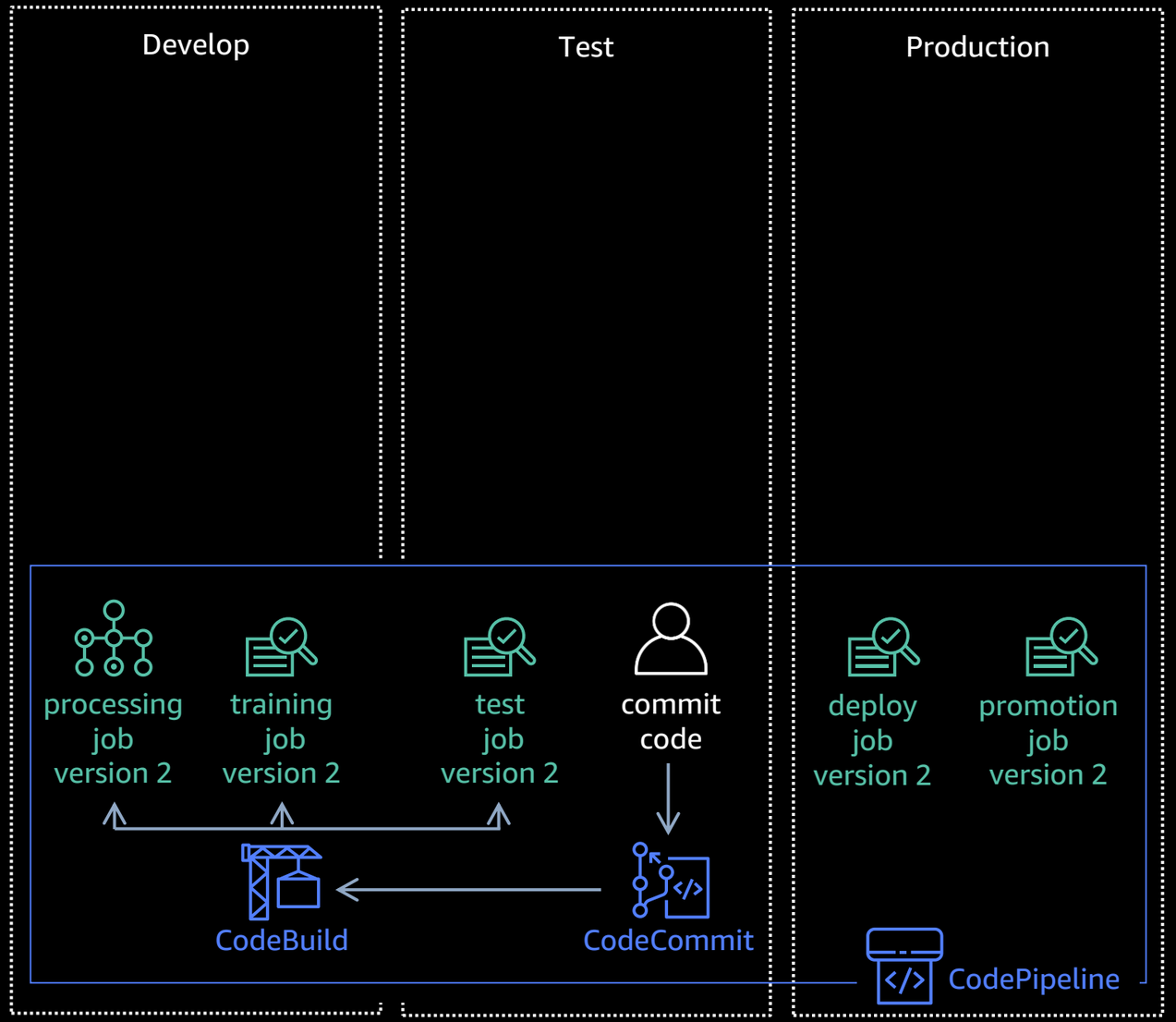


Operational excellence

Security

Reliability

Manage changes to model inputs through automation



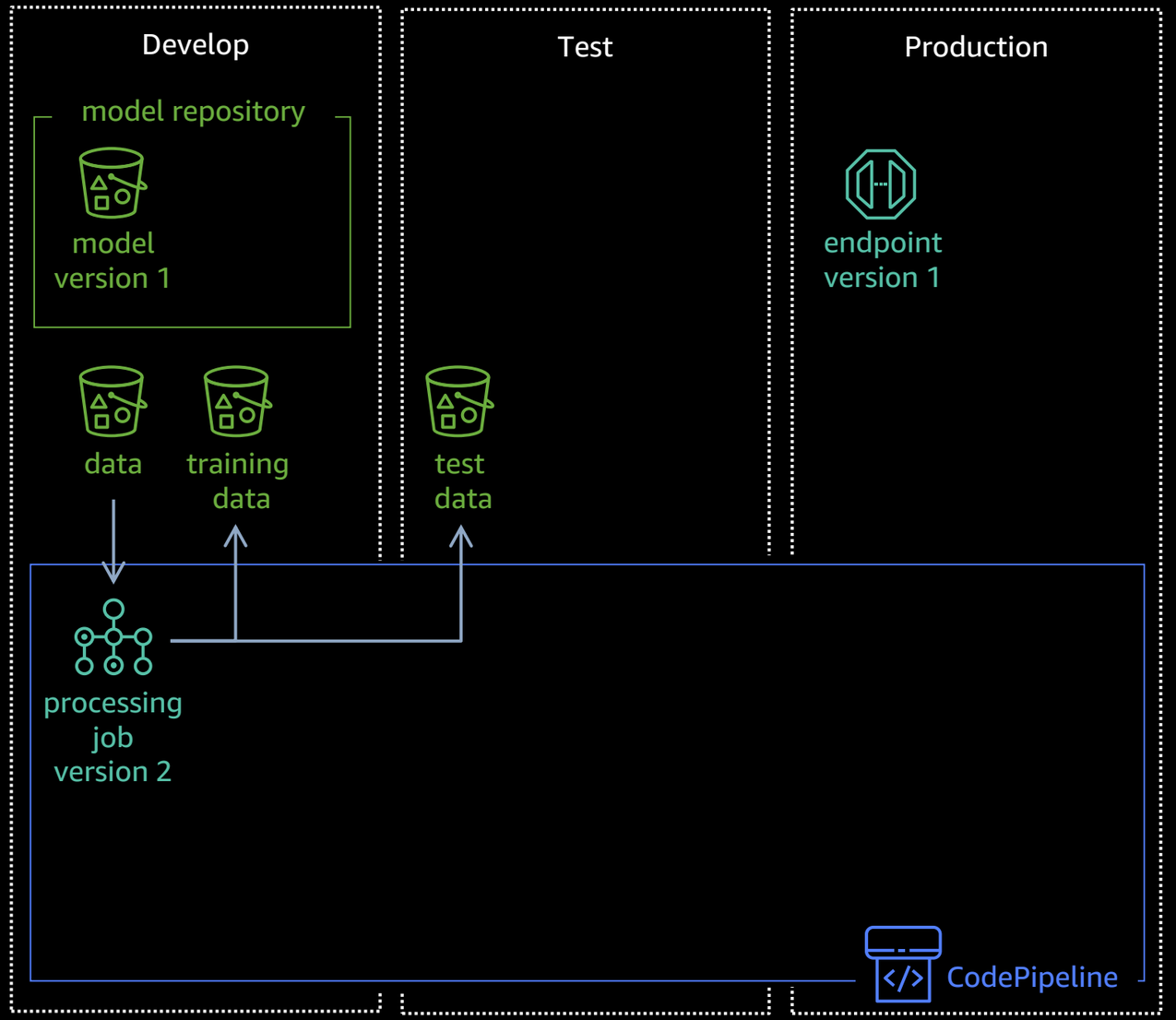
Operational excellence

Security

Reliability

Manage changes to model inputs through automation

Train once and deploy across environments



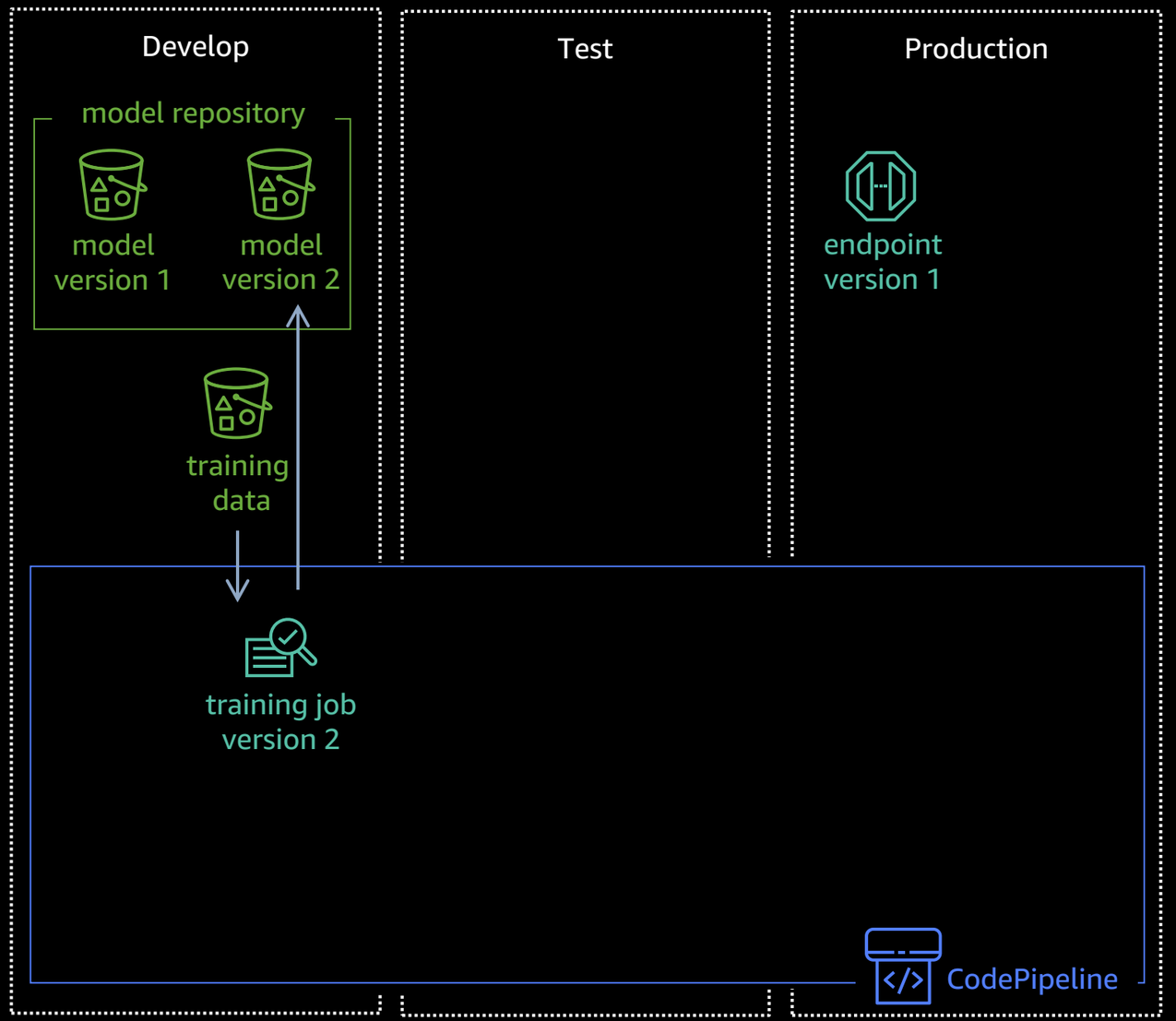
Operational excellence

Security

Reliability

Manage changes to model inputs through automation

Train once and deploy across environments



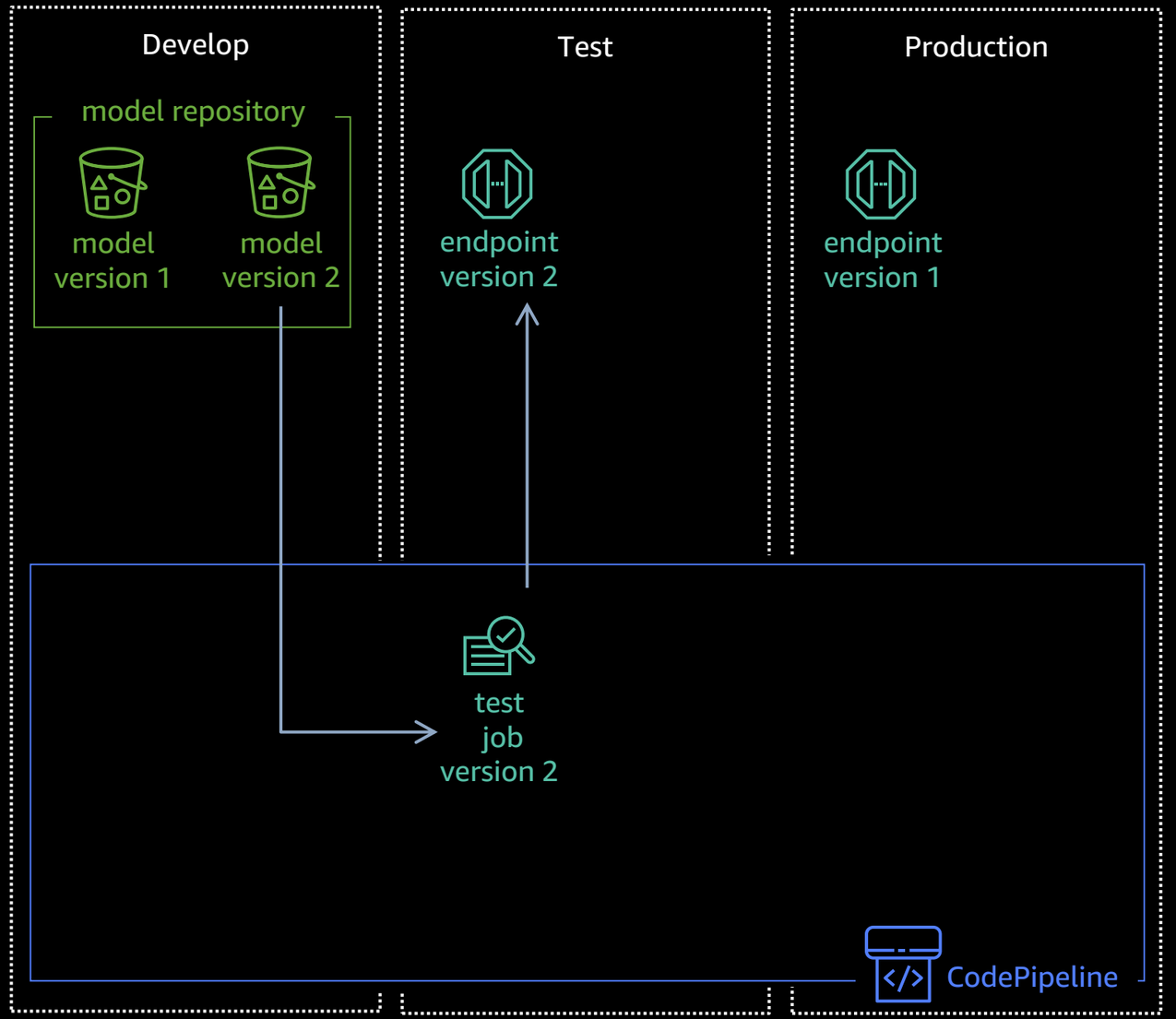
Operational excellence

Security

Reliability

Manage changes to model inputs through automation

Train once and deploy across environments



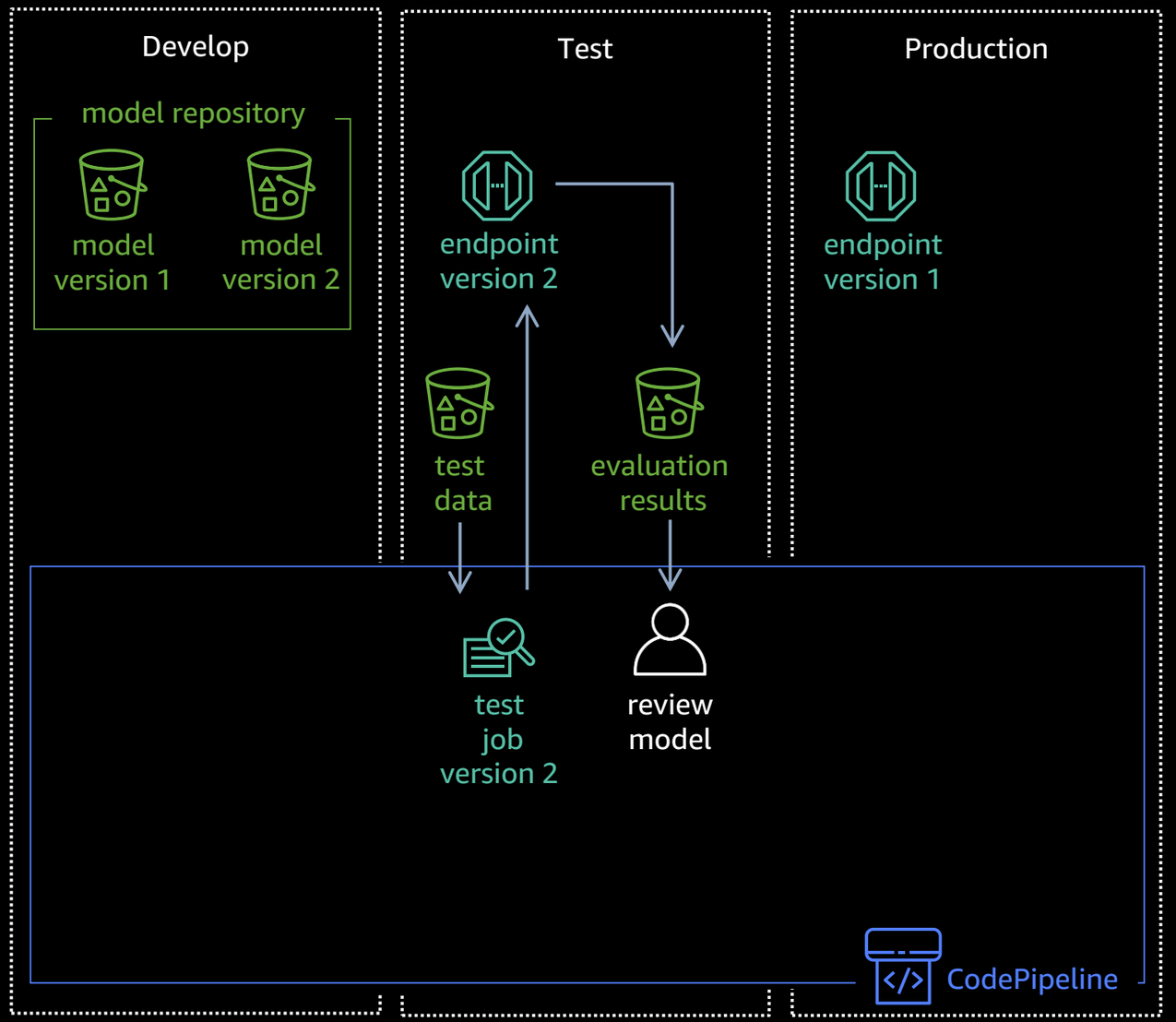
Operational excellence

Security

Reliability

Manage changes to model inputs through automation

Train once and deploy across environments



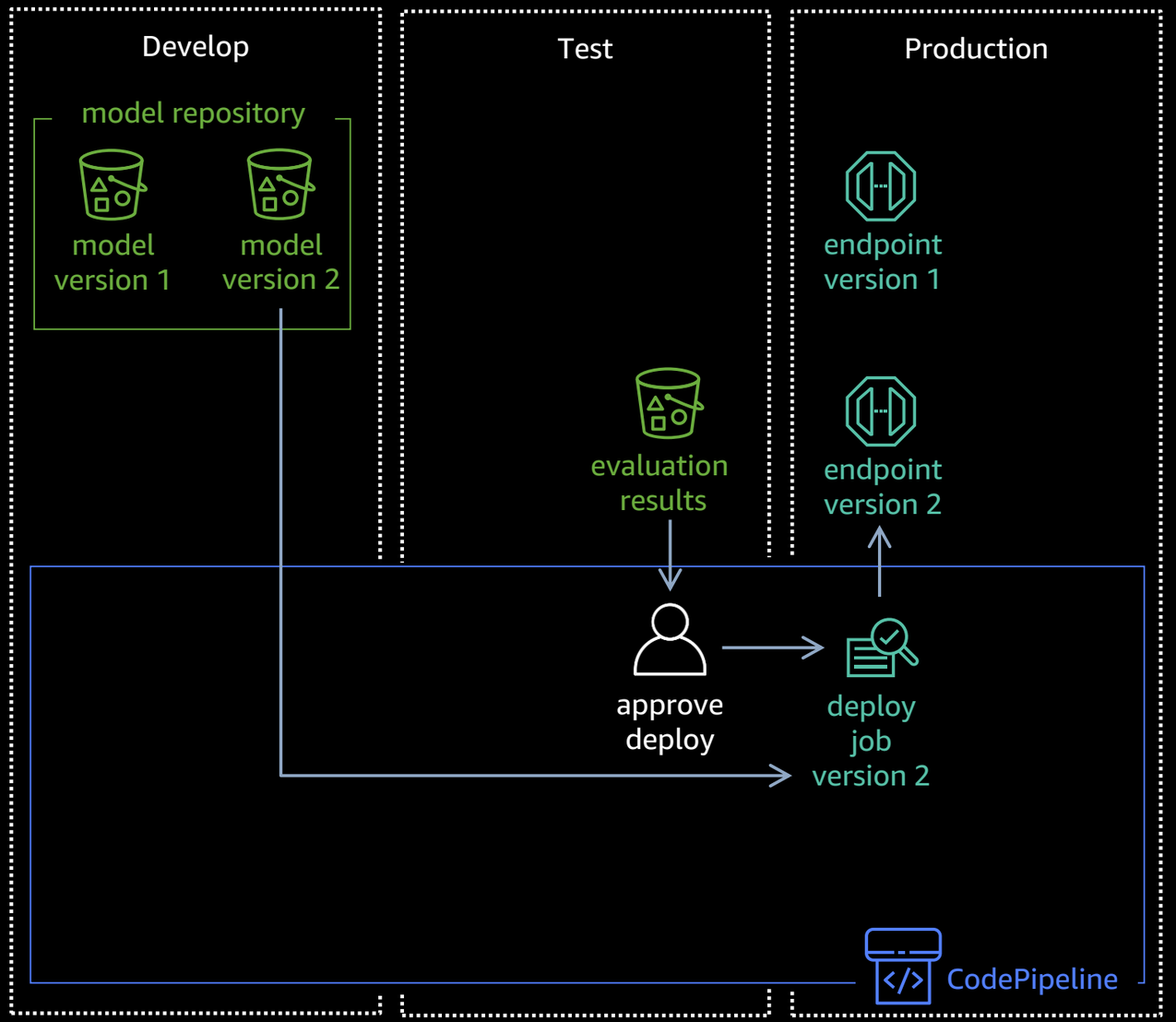
Operational excellence

Security

Reliability

Manage changes to model inputs through automation

Train once and deploy across environments



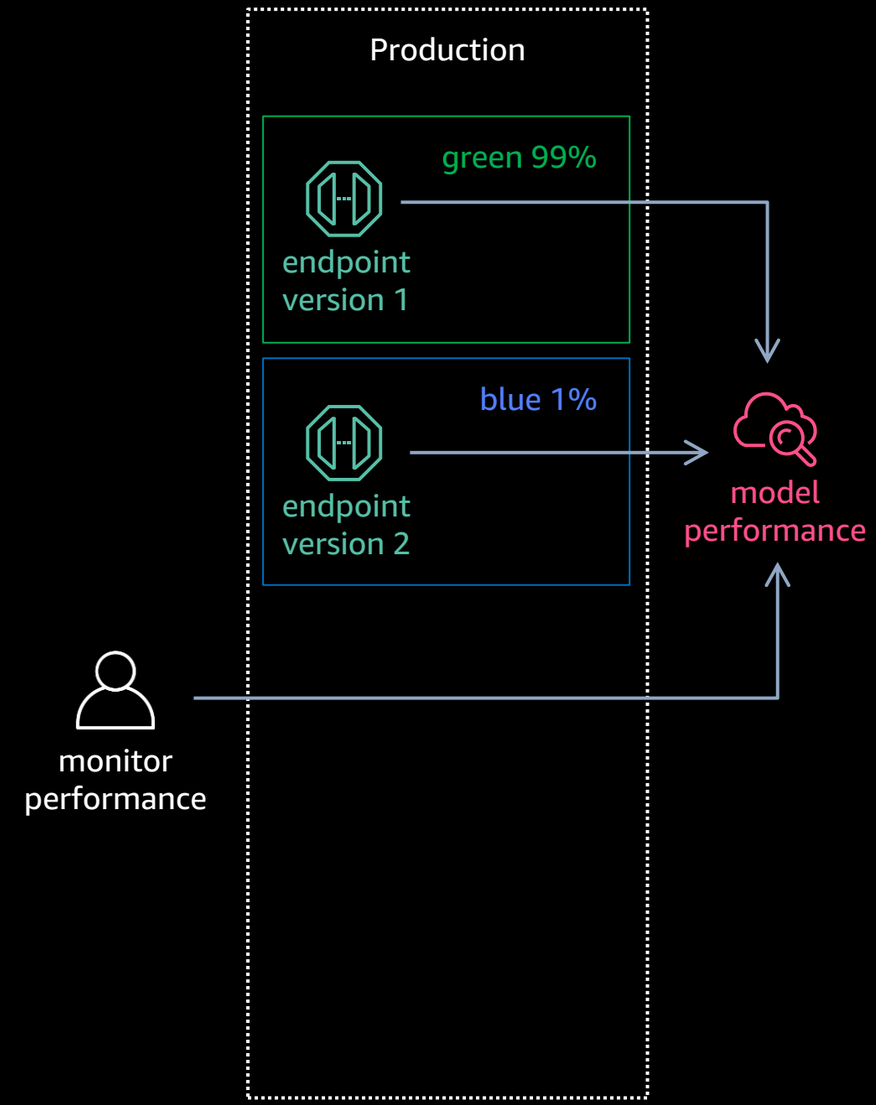
Operational excellence

Security

Reliability

Manage changes to model inputs through automation

Train once and deploy across environments



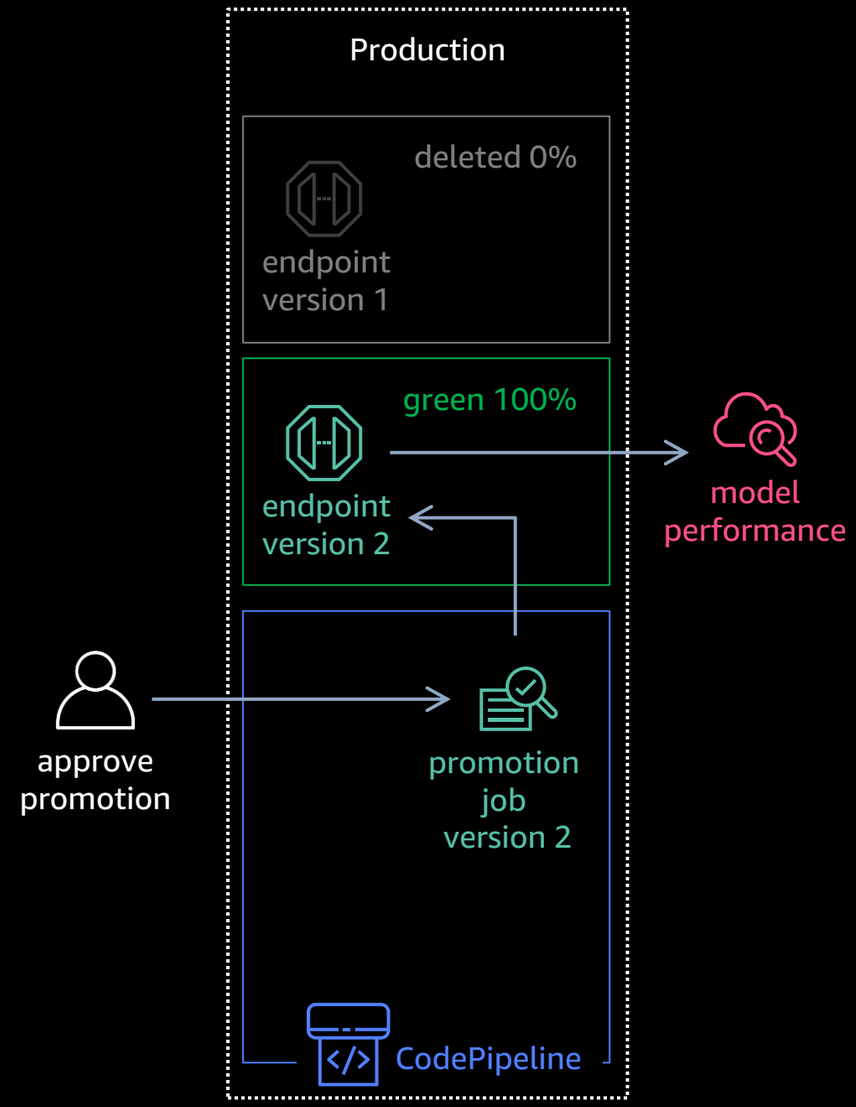
Operational excellence

Security

Reliability

Manage changes to model inputs through automation

Train once and deploy across environments



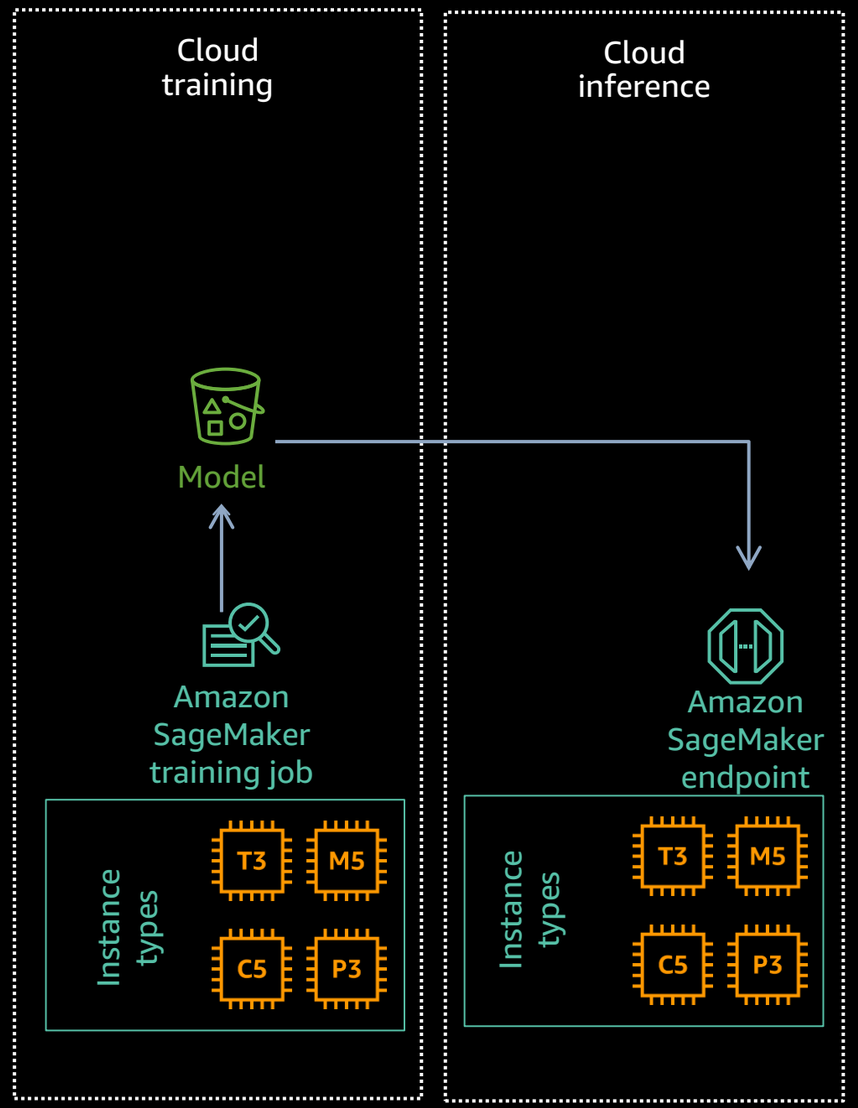
Operational excellence

Security

Reliability

Performance efficiency

Efficient use of computing resources to meet requirements and maintain efficiency as demand changes and technologies evolve



Operational excellence

Security

Reliability

Performance efficiency

Optimize compute for your ML workload



Operational excellence

Security

Reliability

Performance efficiency

Optimize compute for your ML workload



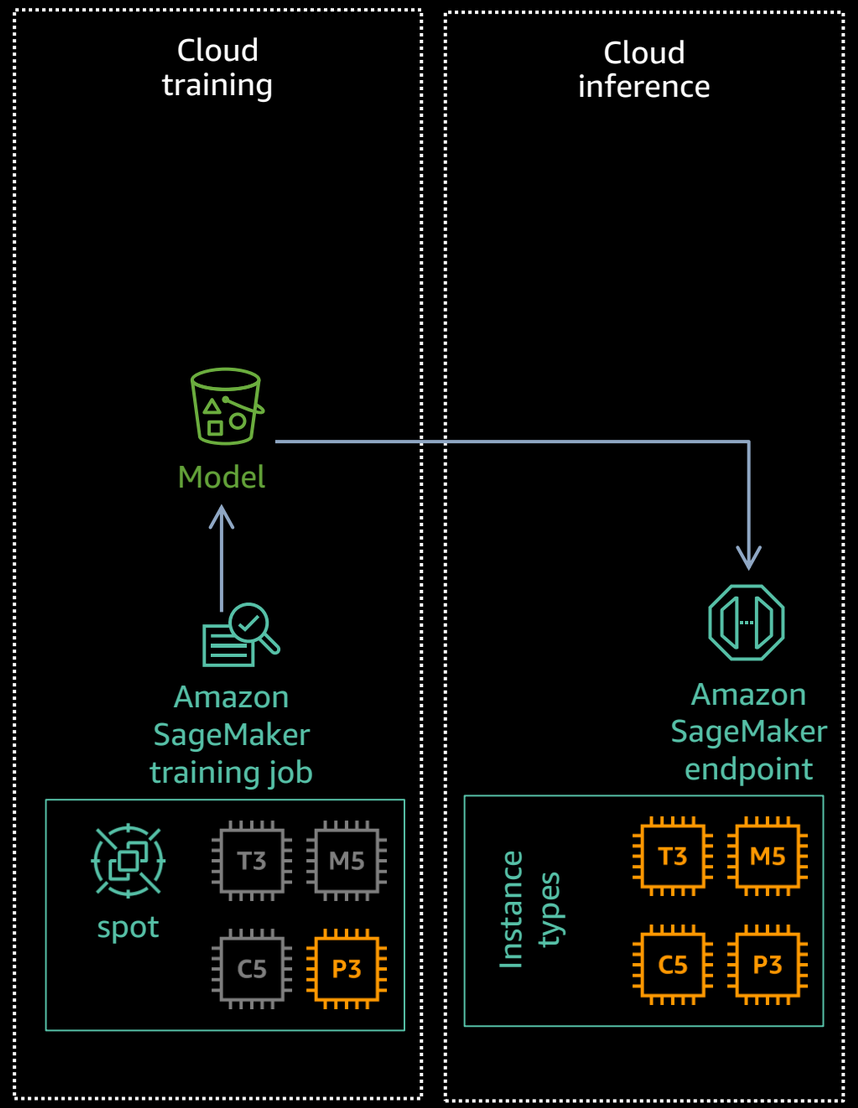
Operational excellence

Security

Reliability

Performance efficiency

Optimize compute for your ML workload



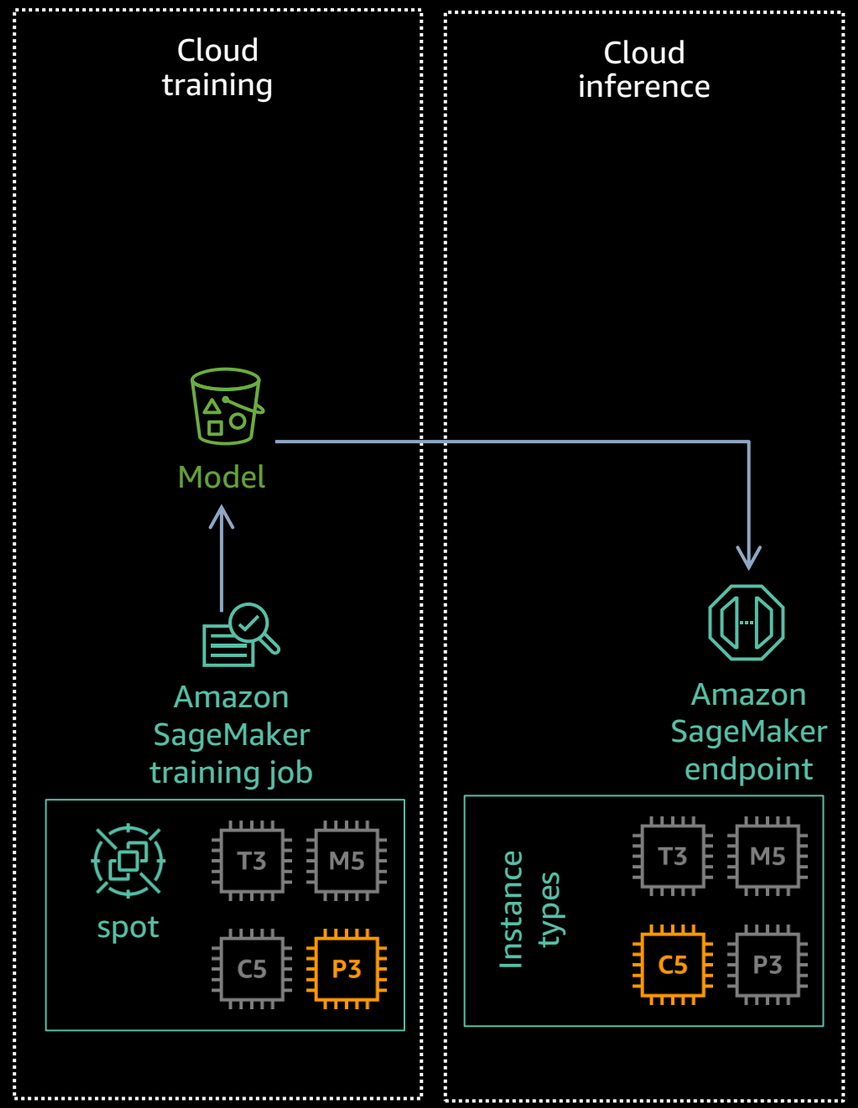
Operational excellence

Security

Reliability

Performance efficiency

Optimize compute for your ML workload



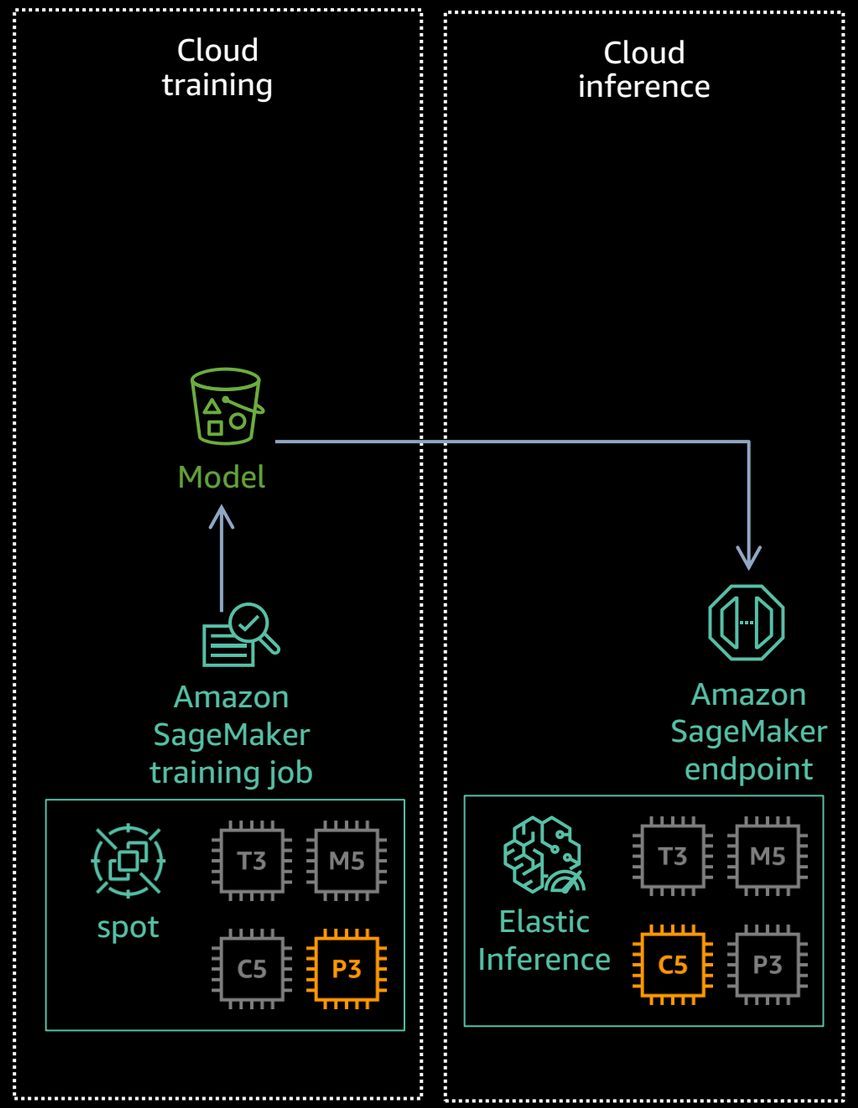
Operational excellence

Security

Reliability

Performance efficiency

Optimize compute for your ML workload



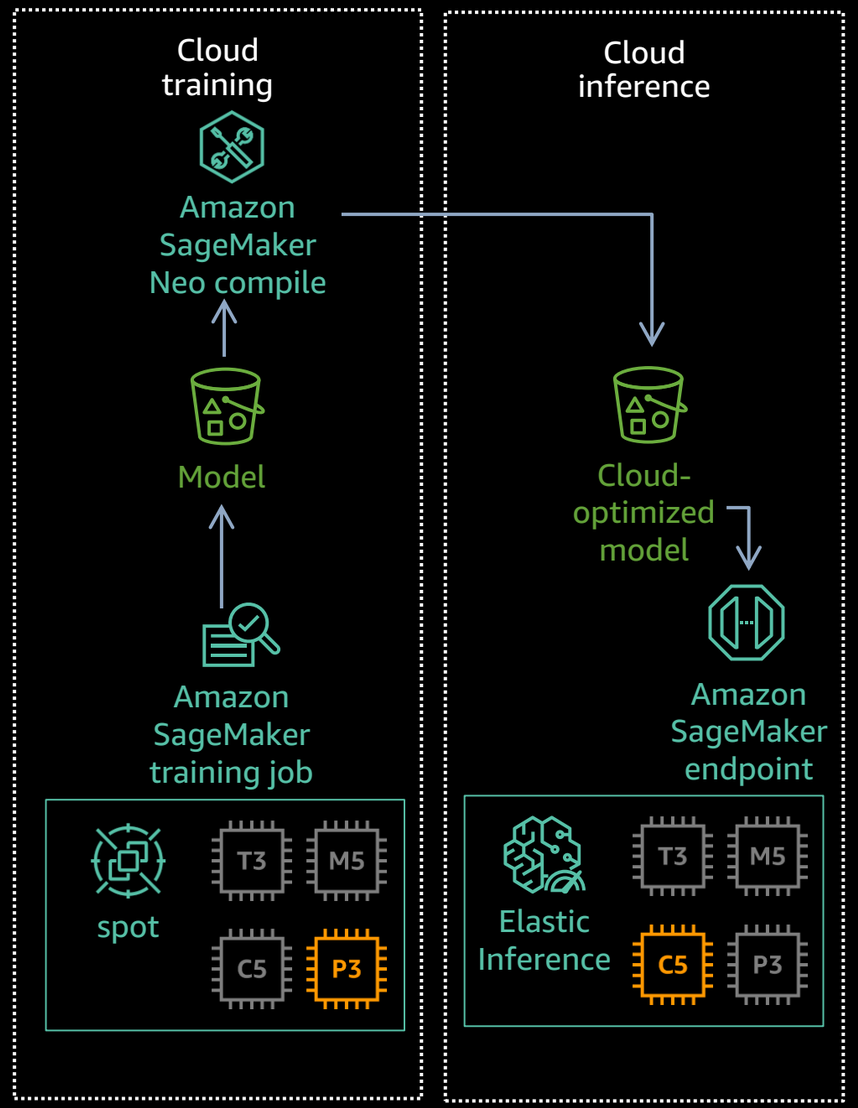
Operational excellence

Security

Reliability

Performance efficiency

Optimize compute for your ML workload



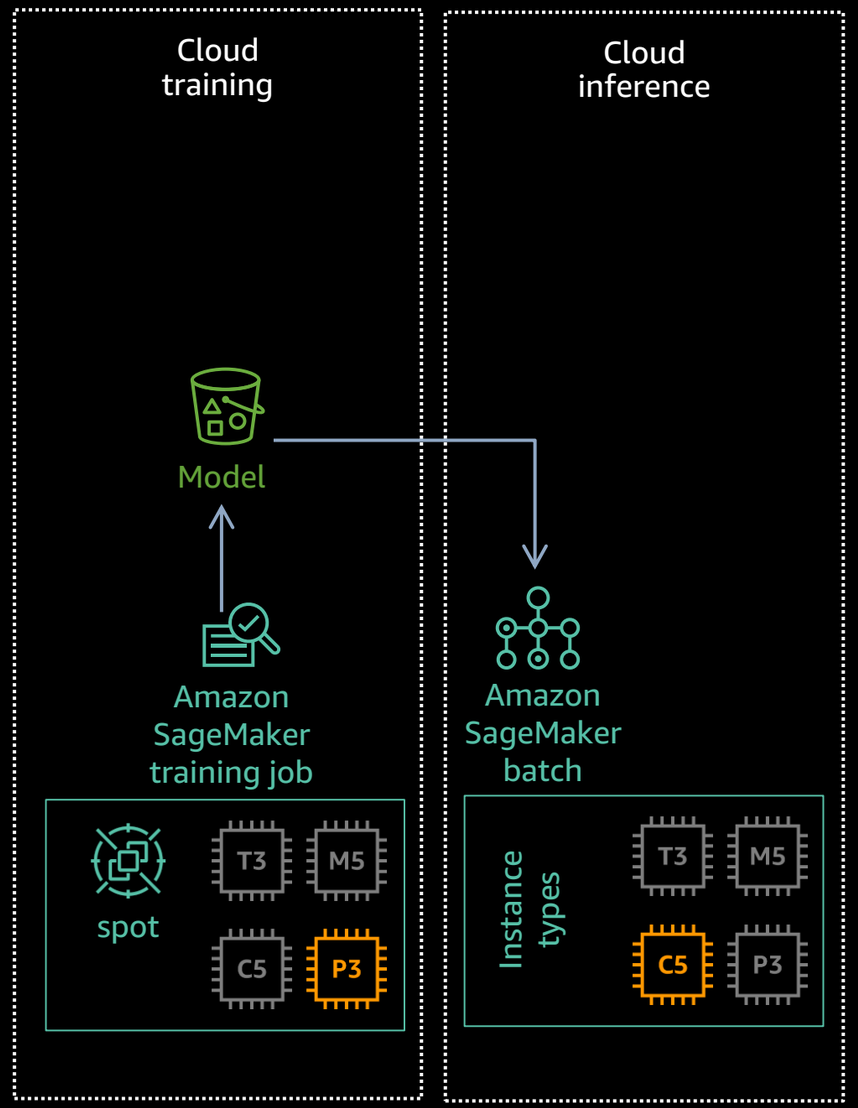
Operational excellence

Security

Reliability

Performance efficiency

Optimize compute for your ML workload



Operational excellence

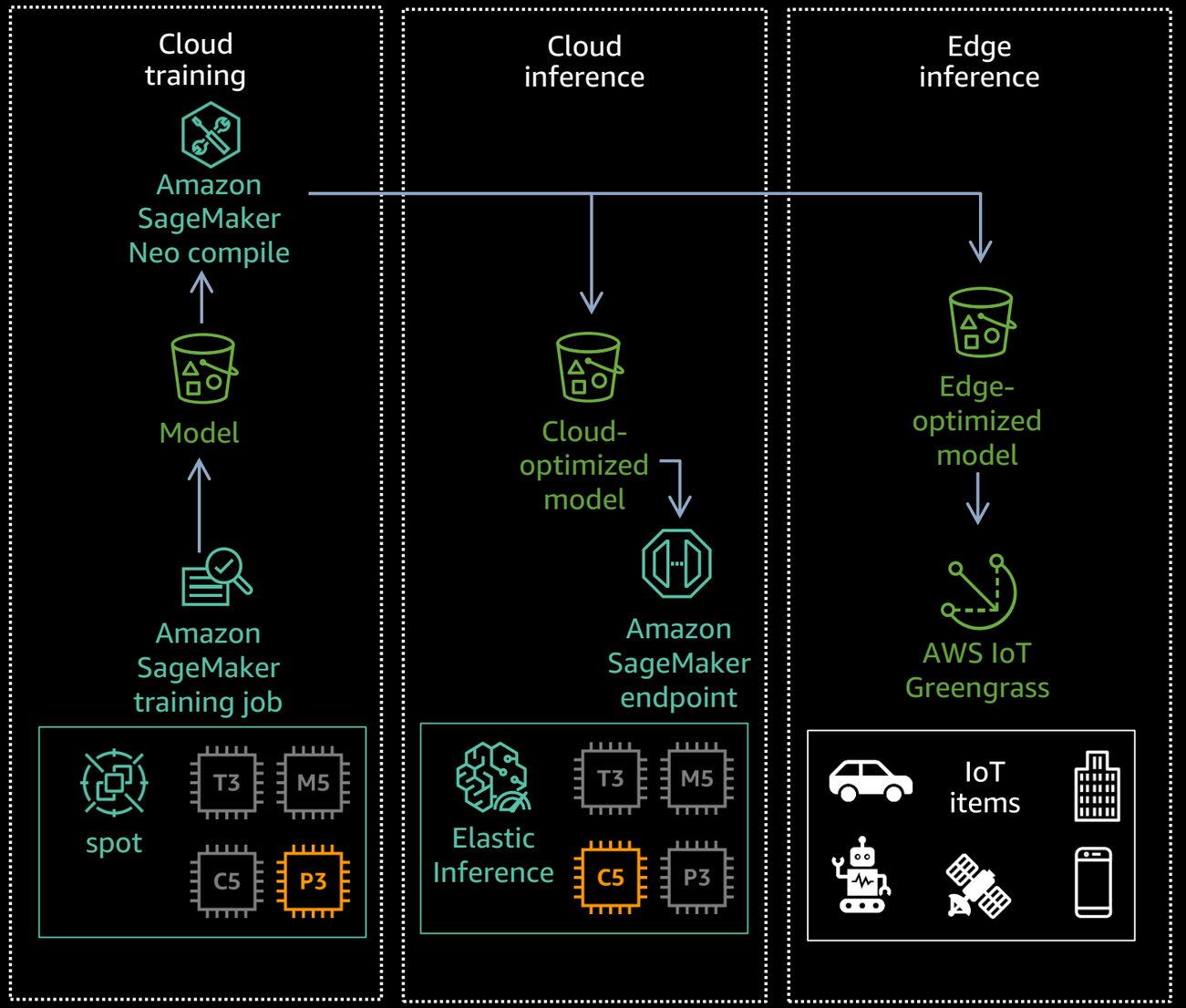
Security

Reliability

Performance efficiency

Optimize compute for your ML workload

Define latency and network bandwidth performance requirements for your models



Operational excellence

Security

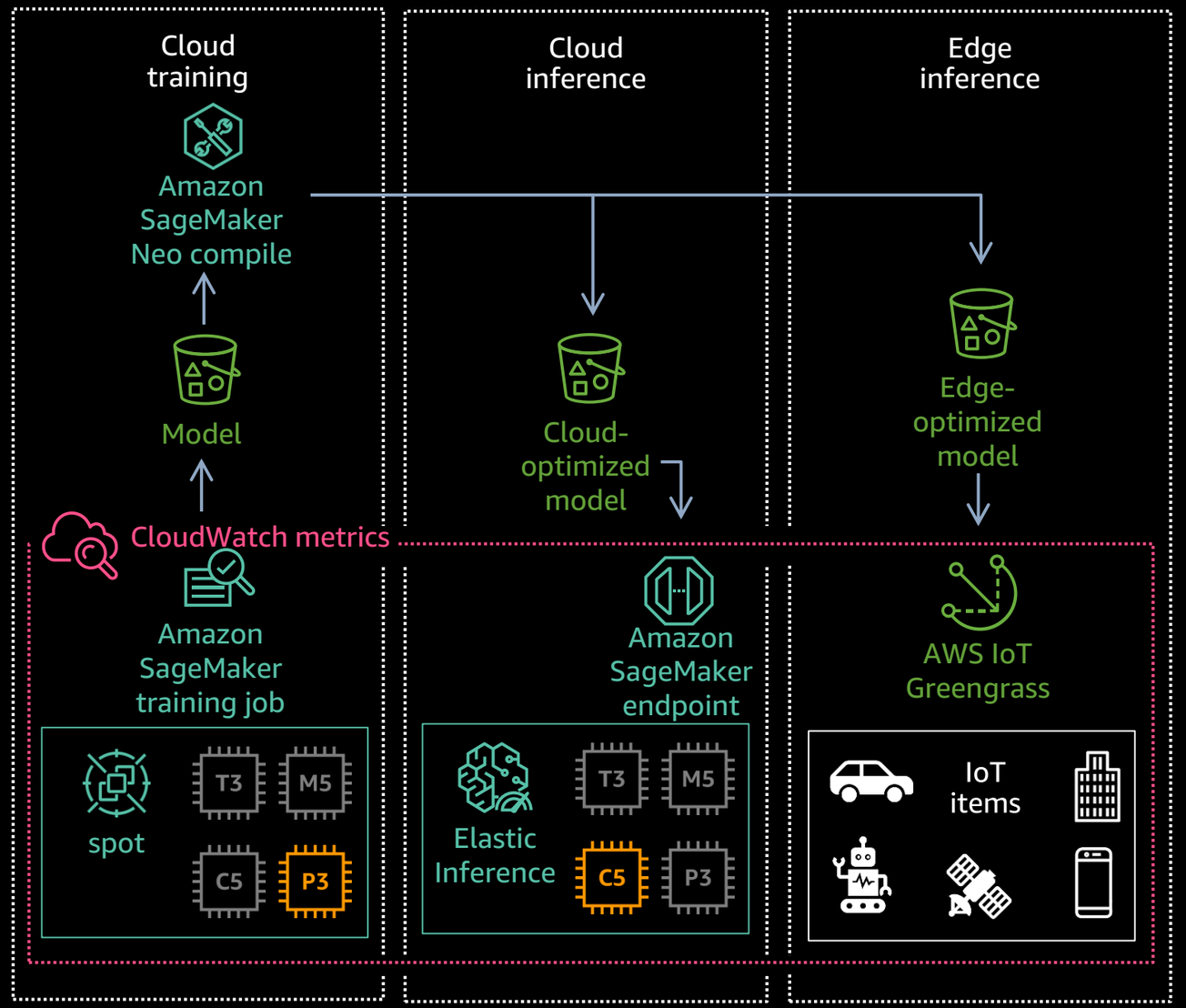
Reliability

Performance efficiency

Optimize compute for your ML workload

Define latency and network bandwidth performance requirements for your models

Continuously monitor and measure system performance



Operational excellence

Security

Reliability

Performance efficiency

Cost optimization

Practices that enable building and operating cost-aware systems that achieve business outcomes, minimize costs, and maximize return on investment



Operational excellence

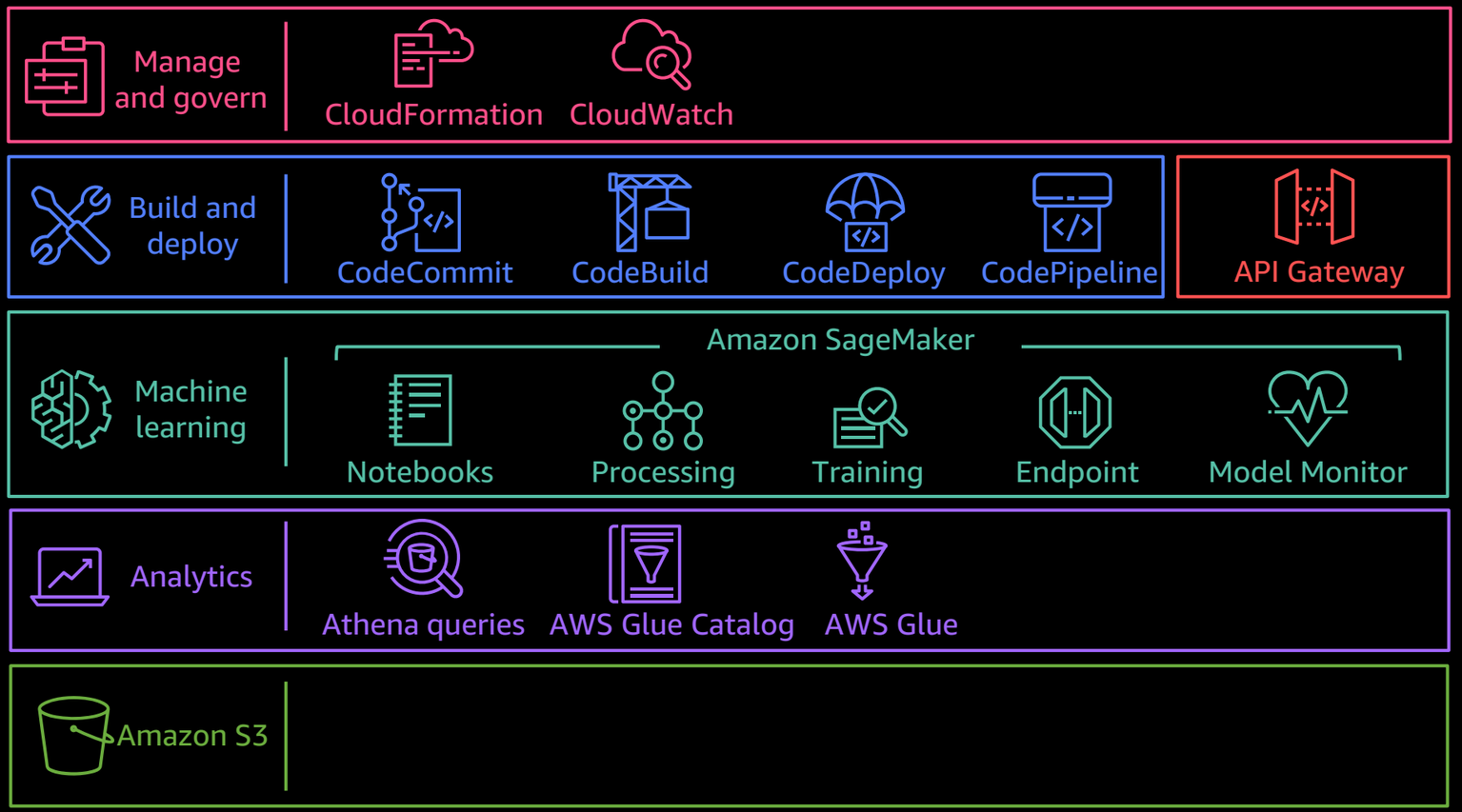
Security

Reliability

Performance efficiency

Cost optimization

Use managed services to reduce cost of ownership



Operational excellence

Security

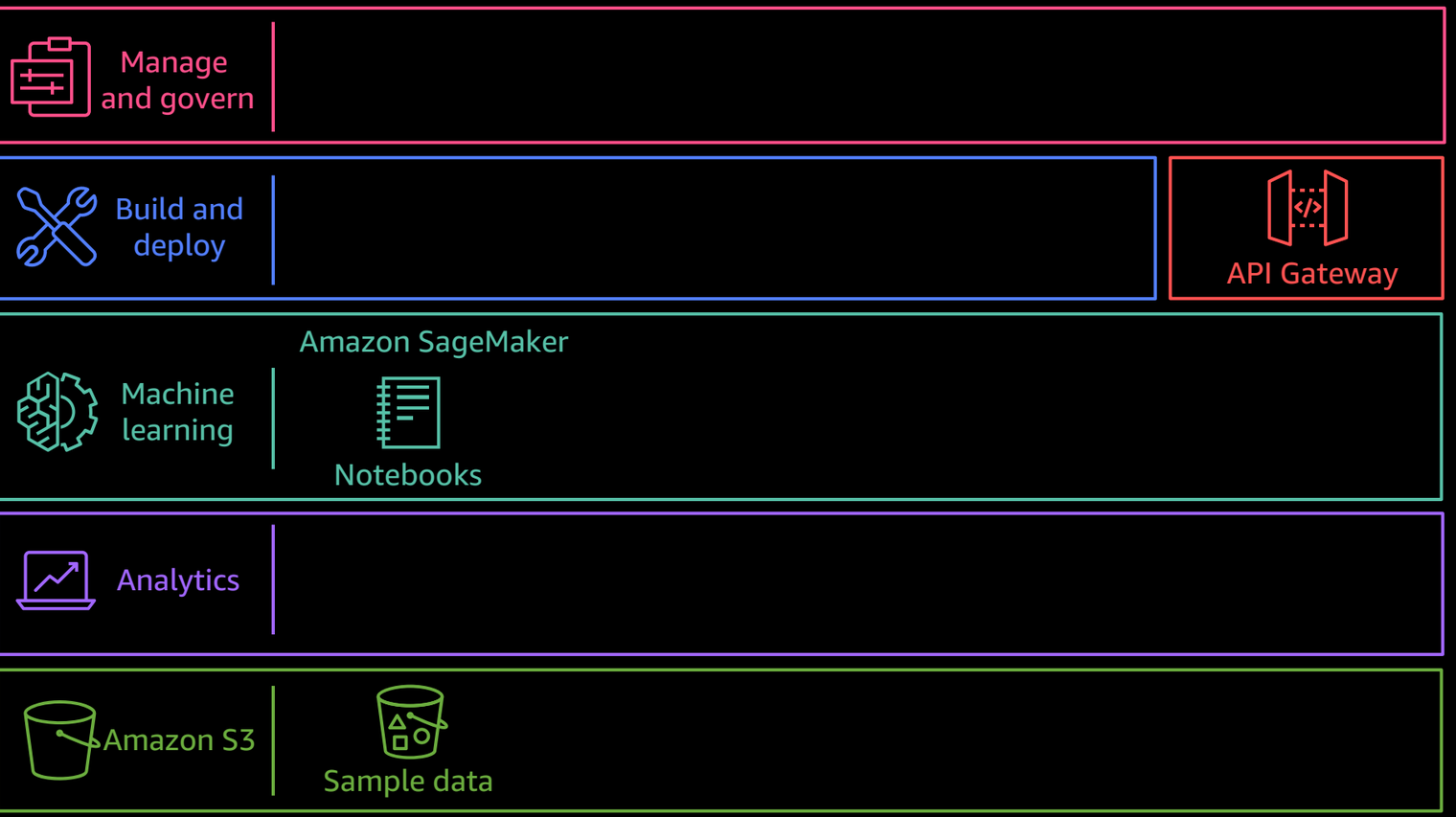
Reliability

Performance efficiency

Cost optimization

Use managed services to reduce cost of ownership

Experiment with small datasets



Operational excellence

Security

Reliability

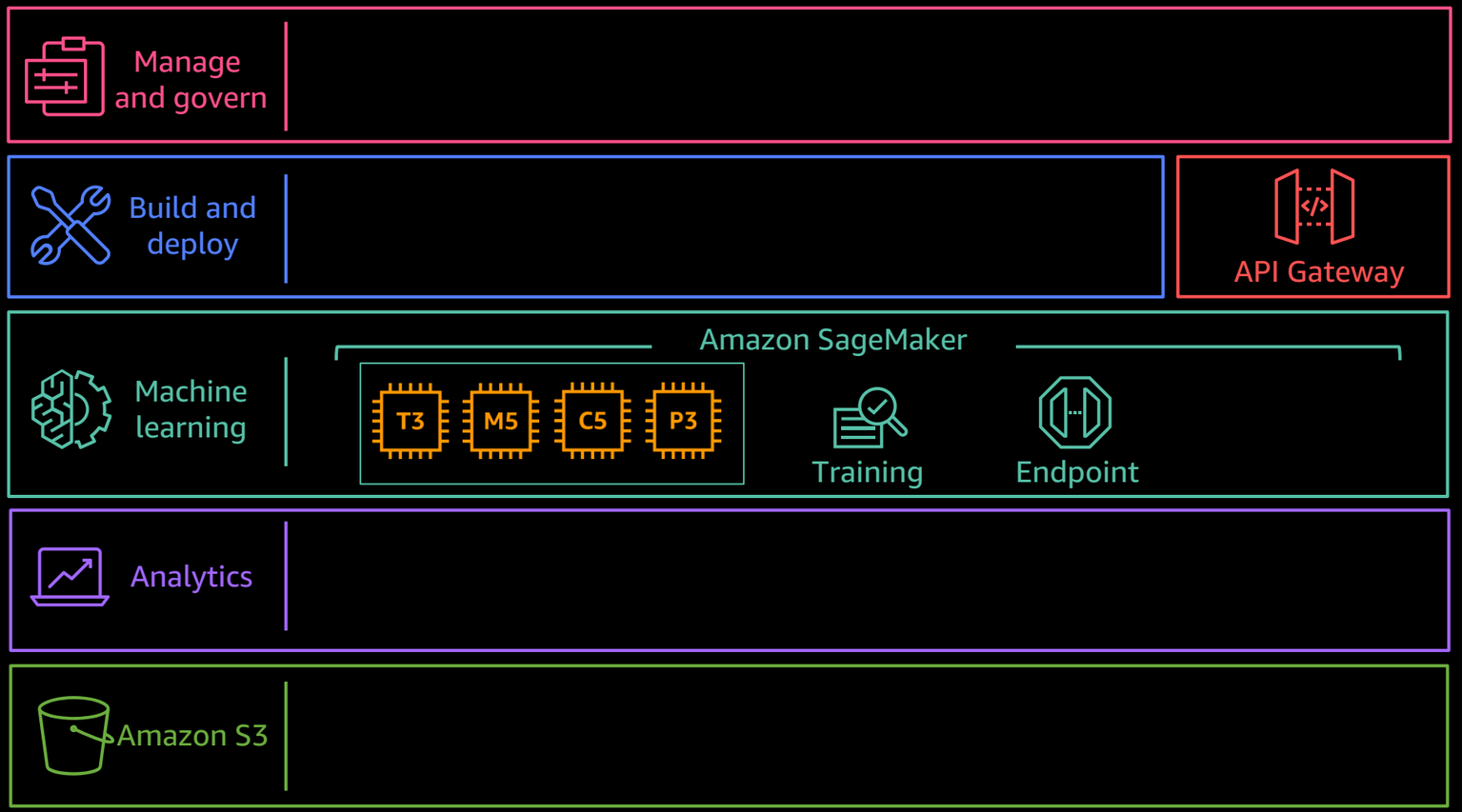
Performance efficiency

Cost optimization

Use managed services to reduce cost of ownership

Experiment with small datasets

Rightsize training and model hosting instances



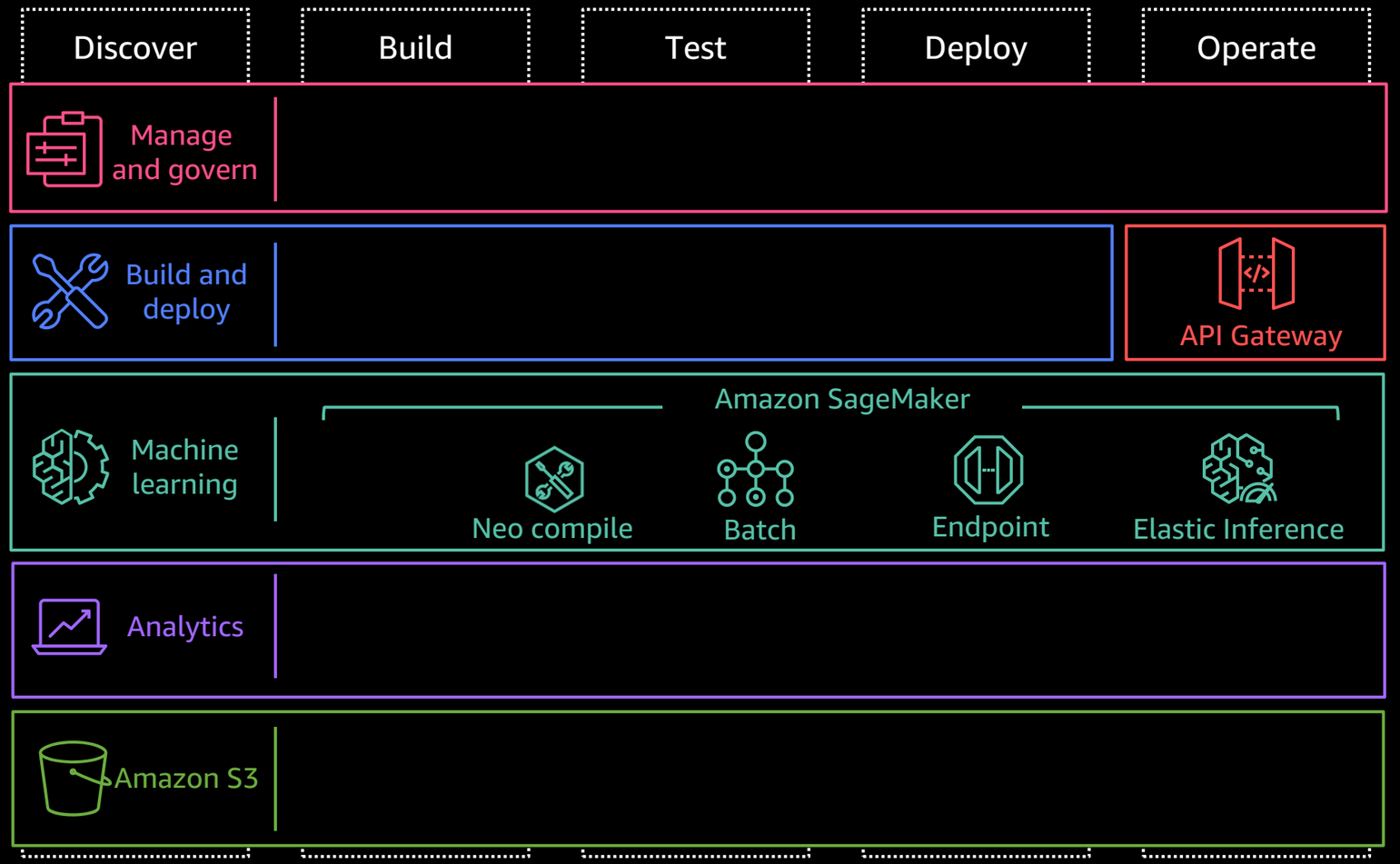
Operational excellence | Security | Reliability | Performance efficiency | Cost optimization

Use managed services to reduce cost of ownership

Experiment with small datasets

Rightsize training and model hosting instances

Account for inference architecture based on consumption patterns



Operational excellence

Security

Reliability

Performance efficiency

Cost optimization

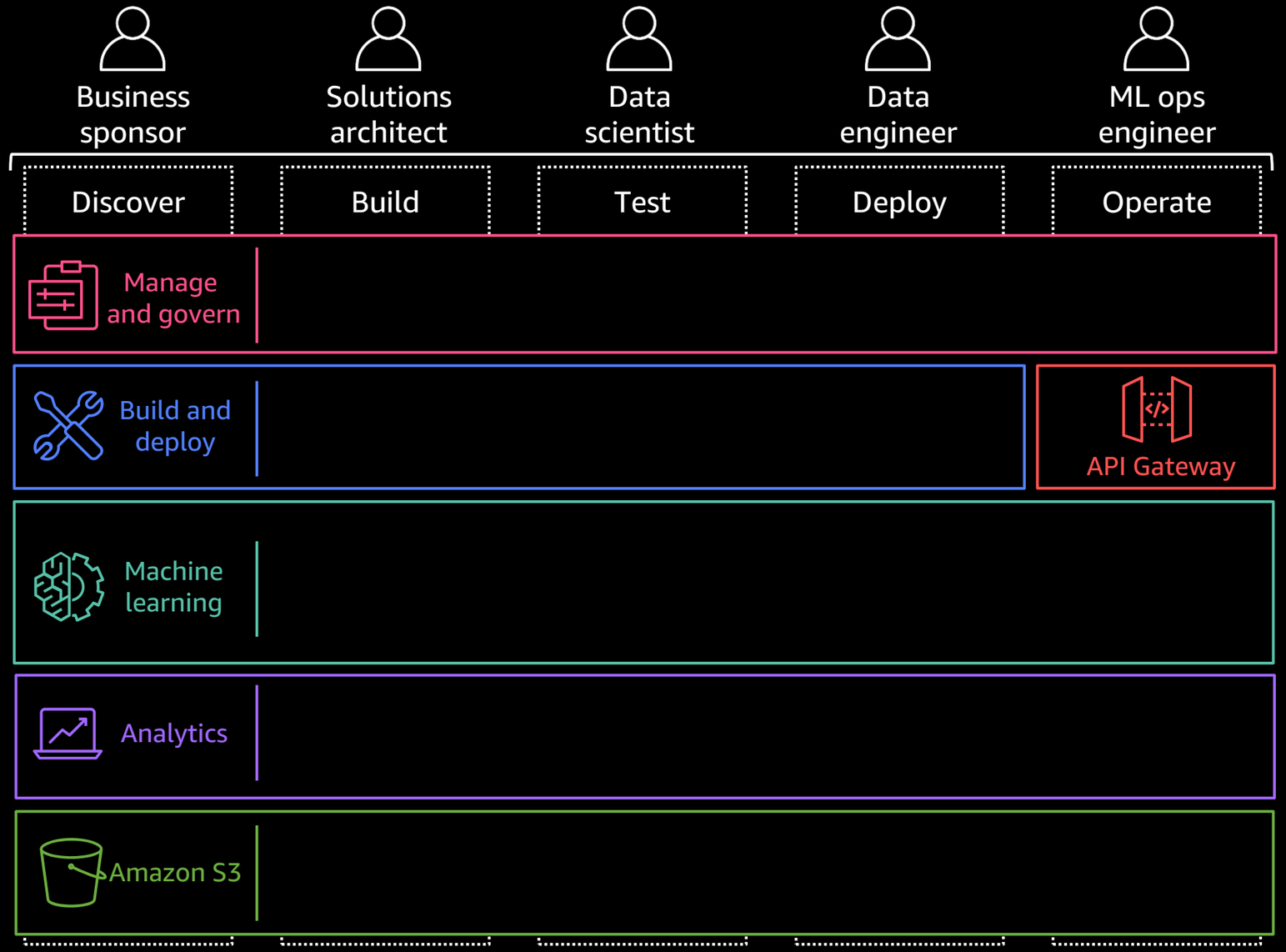
Use managed services to reduce cost of ownership

Experiment with small datasets

Rightsize training and model hosting instances

Account for inference architecture based on consumption patterns

Define overall ROI and opportunity cost



More information

The AWS Well-Architected Framework
aws.amazon.com/architecture/well-architected

Amazon SageMaker
aws.amazon.com/sagemaker

Thank you!

Eric Greene

www.linkedin.com/in/greene/

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Please complete the session survey

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.