MLOps avec Databricks : incontournables et leçons apprises.

Histoires d’échecs, de succès et conseils

databricks

agileDSS

Présentateur - Vincent Fortier

Manager, Field Engineering, Eastern Canada

3.5 ans chez Databricks

3 ans chez Cloudera

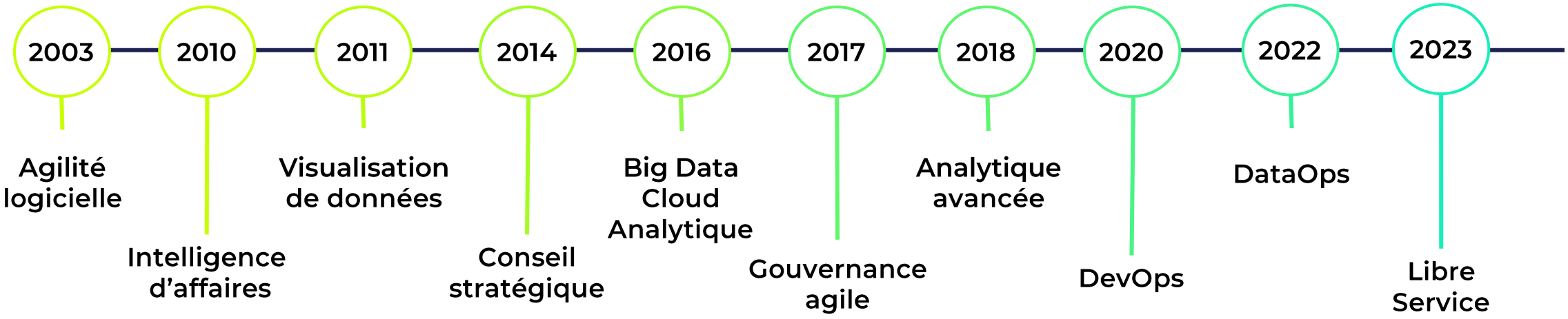
3+ années d'amélioration de ma recette de pizza

2 chiens saucisse

vincent.fortier@databricks.com

agileDSS

Fondé à Montréal en 2003, agileDSS est un collectif d'experts pluridisciplinaires qui vous accompagne dans la mise en place de votre écosystème analytique pour accélérer votre prise de décision.



databricks agileDSS

Agenda

1. Introduction
2. Causes des échecs de projets en IA
3. MLOps comme facteur de succès
4. Déployer du MLOps avec Databricks

databricks agileDSS

Introduction : chiffres clés

- En 2021, Gartner Inc. prévoyait que 85 % des implémentations d'IA échoueront
- Ils prévoyaient que cette tendance se maintiendrait en 2022

databricks agileDSS

Introduction : Définition d'un projet en IA réussi

Projet à valeur ajoutée

Projet qui a un retour sur investissement positif

Projet qui procure un avantage compétitif

databricks agileDSS

Cause des échecs de projets en IA

1. **Organisation**
2. Processus
3. Technologique

Organisation

- Engagement exécutif faible
- Manque de collaboration entre les équipes

databricks agileDSS

Cause des échecs de projets en IA

1. Organisation
2. **Processus**
3. Technologique

Processus

- Gouvernance de données non mature
- Absence de surveillance de la performance en production
- ROI non mesurable

databricks agileDSS

Cause des échecs de projets en IA

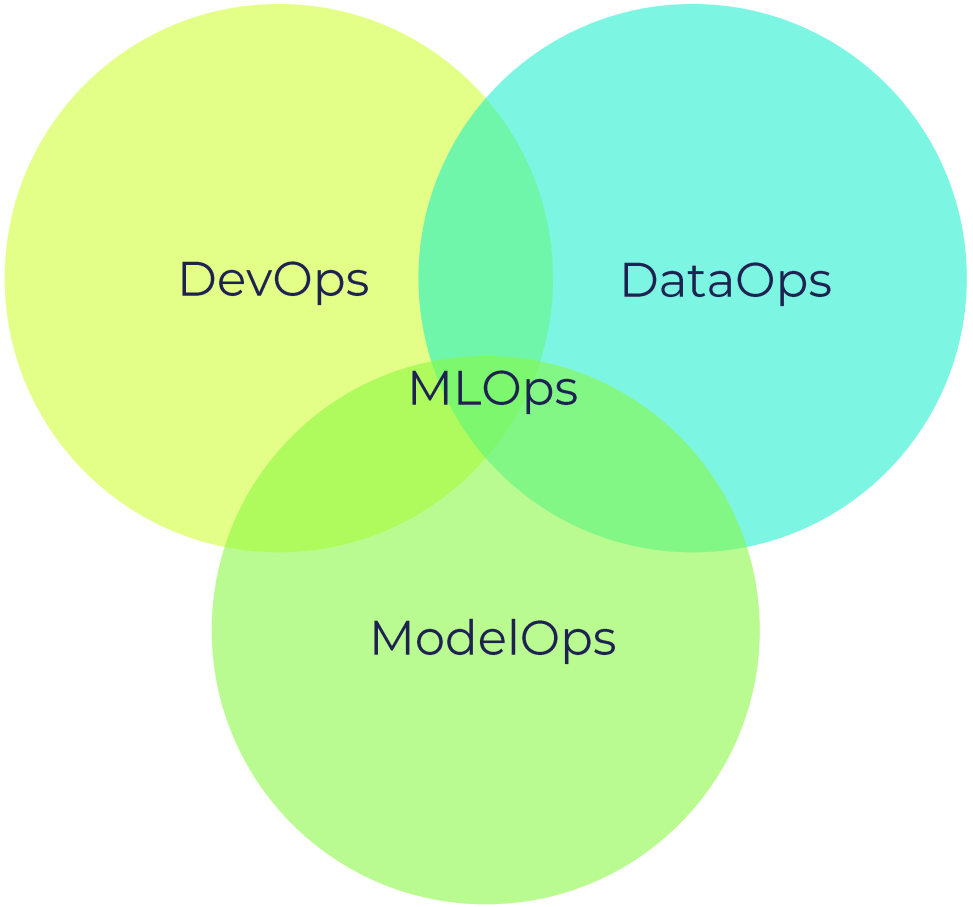
1. Organisation
2. Processus
3. **Technologique**

Technologique

- Transition du mode expérimental à la production coûteuse
- Modèles non interprétables (type boite noire)

databricks agileDSS

MLOps comme facteur de succès : définition



databricks agileDSS

MLOps comme facteur de succès : Pourquoi?

- MLOps vous aide à réduire les risques :
  - Risque de conformité : violation des régulations, de politique corporative
  - Risque technique : modèles de mauvaise qualité, infrastructure fragile
- MLOps améliore l'efficacité à long terme grâce à l'automatisation :
  - Capter des erreurs avant le déploiement en production
  - Éviter les tâches manuelles

databricks agileDSS

MLOps comme facteur de succès : Principes

P1 : Automatisation du CI/CD
P2 : Orchestration du flux de travail
P3 : Reproductibilité
P4 : Gestion des versions de données, du code, du modèle
P5 : Collaboration
P6 : Apprentissage + évaluation continus de modèles
P7 : Suivi des métadonnées de l'apprentissage auto
P8 : Surveillance continue
P9 : Boucle de rétroaction

Source : arxiv.org/abs/2205.02302

databricks agileDSS

MLOps comme facteur de succès : Intervenants

Parties prenantes

- Responsable de la valeur ajoutée de la solution IA

Ingénieur de données

- Bâti des pipelines de données

Scientifique de données

- Traduit un problème d'affaires en tâche de IA
- Bâti des modèles

Ingénieur ML

- Déploie des modèles en production

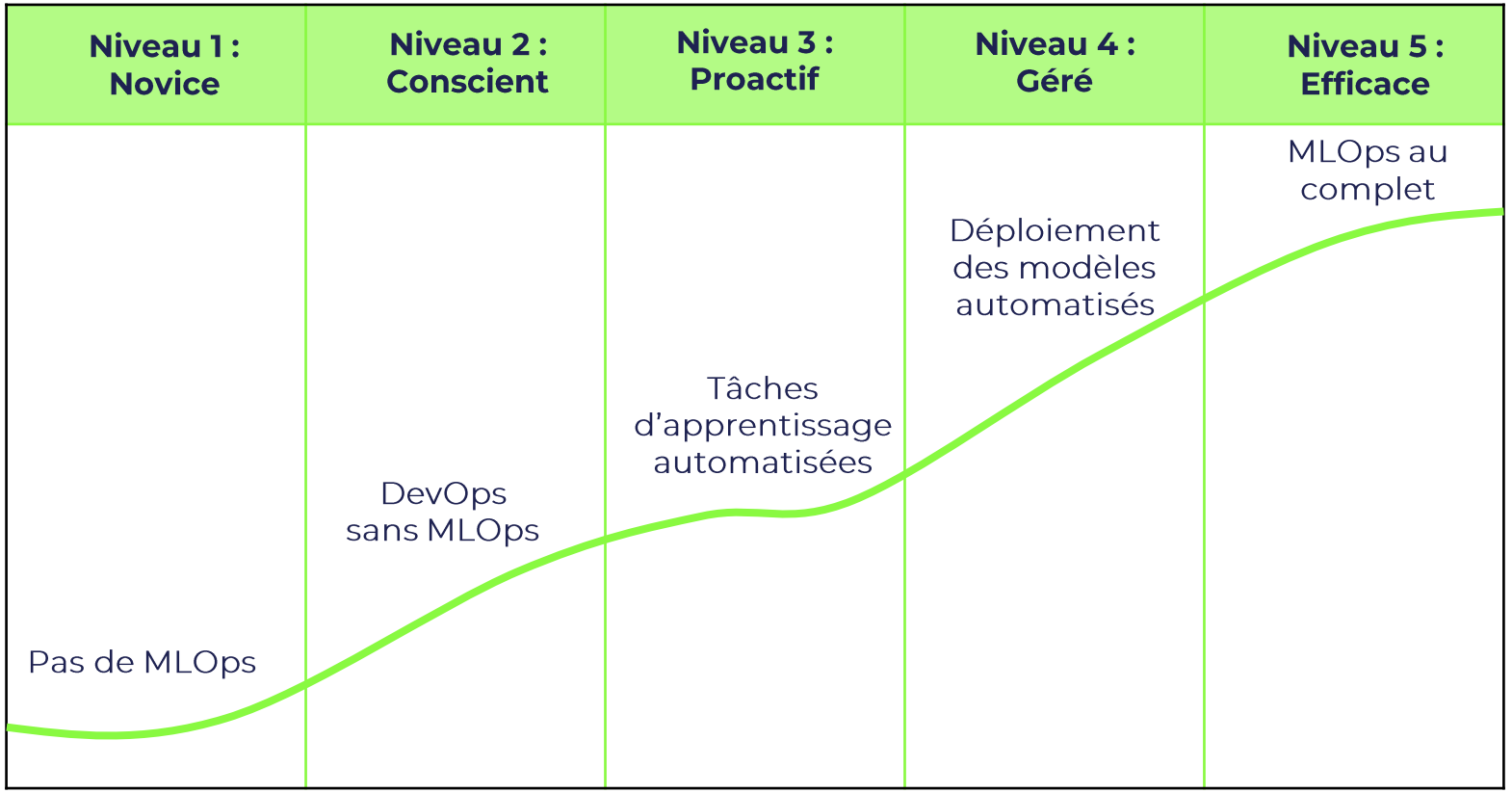
Chef gouvernance de données

- Responsable de la gouvernance des données et de la conformité

Source : arxiv.org/abs/2205.02302

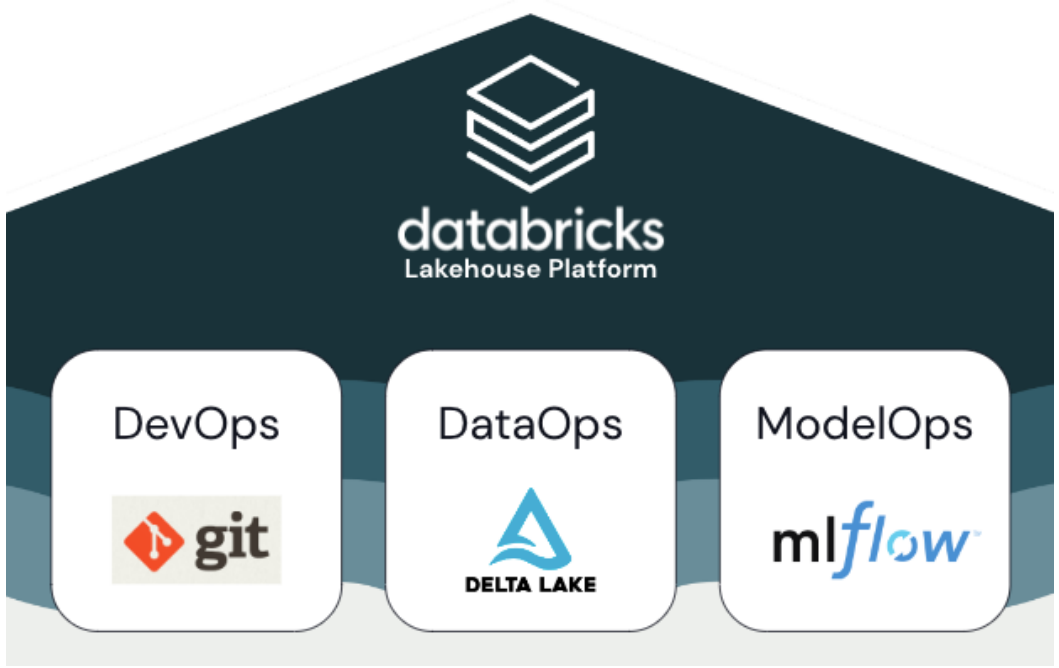
databricks agileDSS

MLOps: Niveaux de maturité



databricks agileDSS

Déployer du MLOps avec Databricks: Pourquoi ?



- Disponible sur Azure, AWS et Google Cloud
- Plateforme pour données structurées et non structurées
- Intégration facile
- Déploiement du DevOps, DataOps, ModelOps supportés
- Offre Unity catalog pour la gouvernance de données

Source: https://learn.microsoft.com/en-us/azure/databricks/_static/images/machine-learning/mlops-lakehouse.png

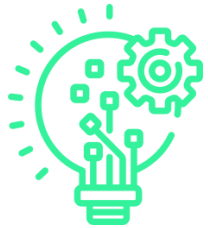
databricks agileDSS

Déployer du MLOps avec Databricks

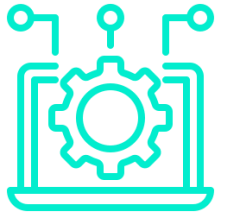
Adoptez une approche incrémentale basée sur les niveaux de maturité MLOps

Assurez-vous d'avoir l'expertise nécessaire

Favorisez les bonnes pratiques

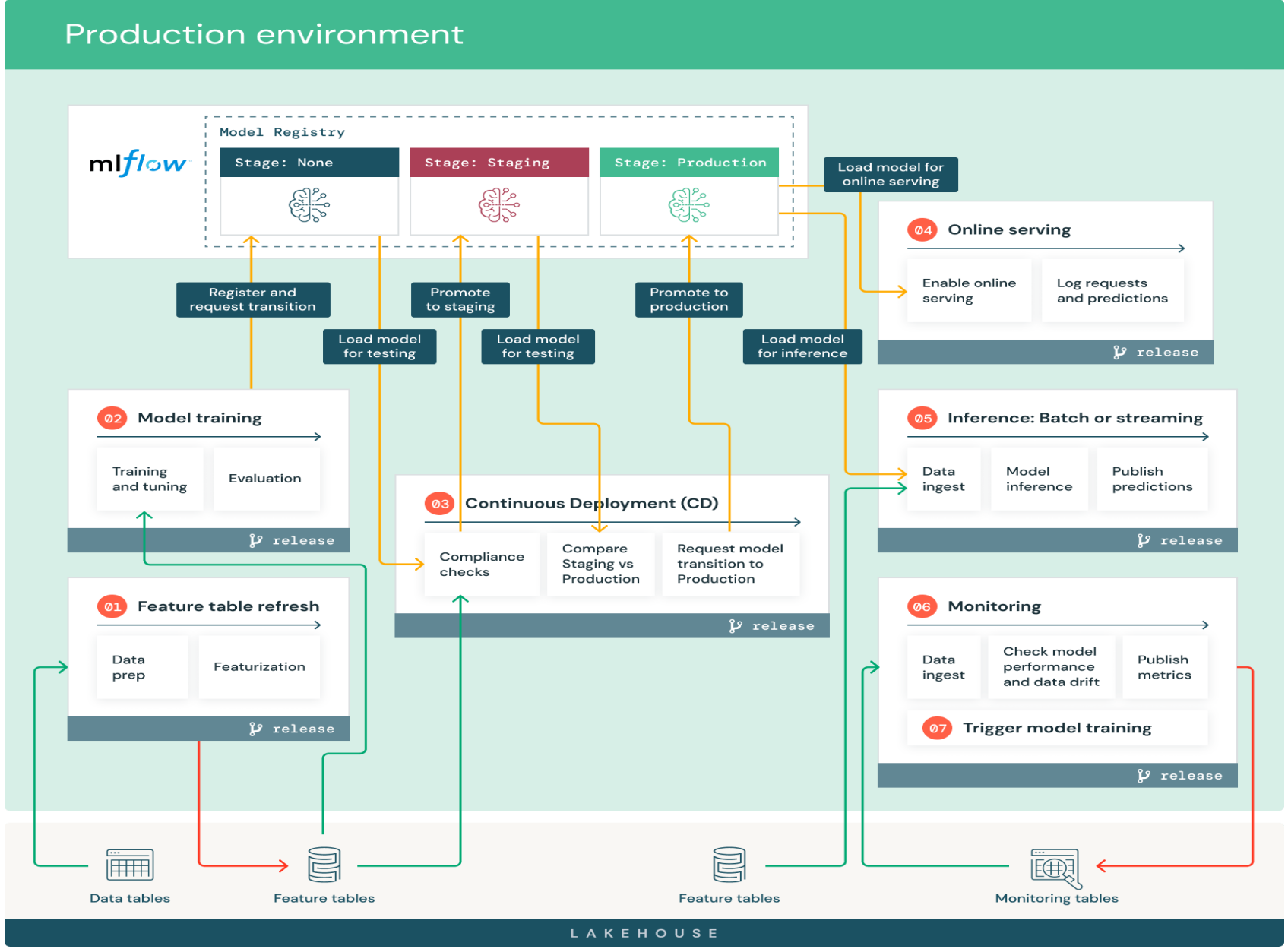


Faites des choix technologiques justifiés



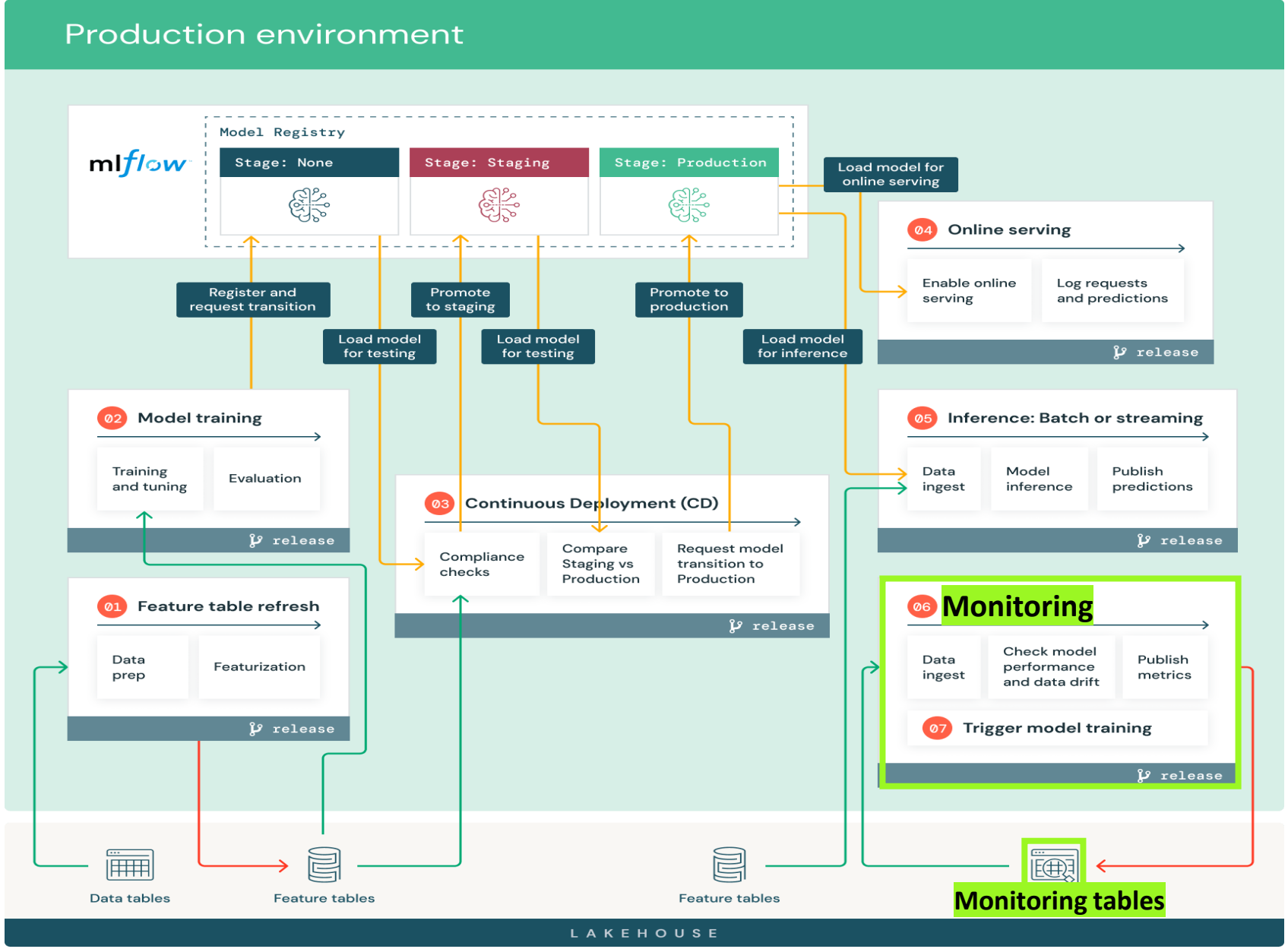
Favorisez l'approche plateforme au détriment de l'approche projet

databricks agileDSS



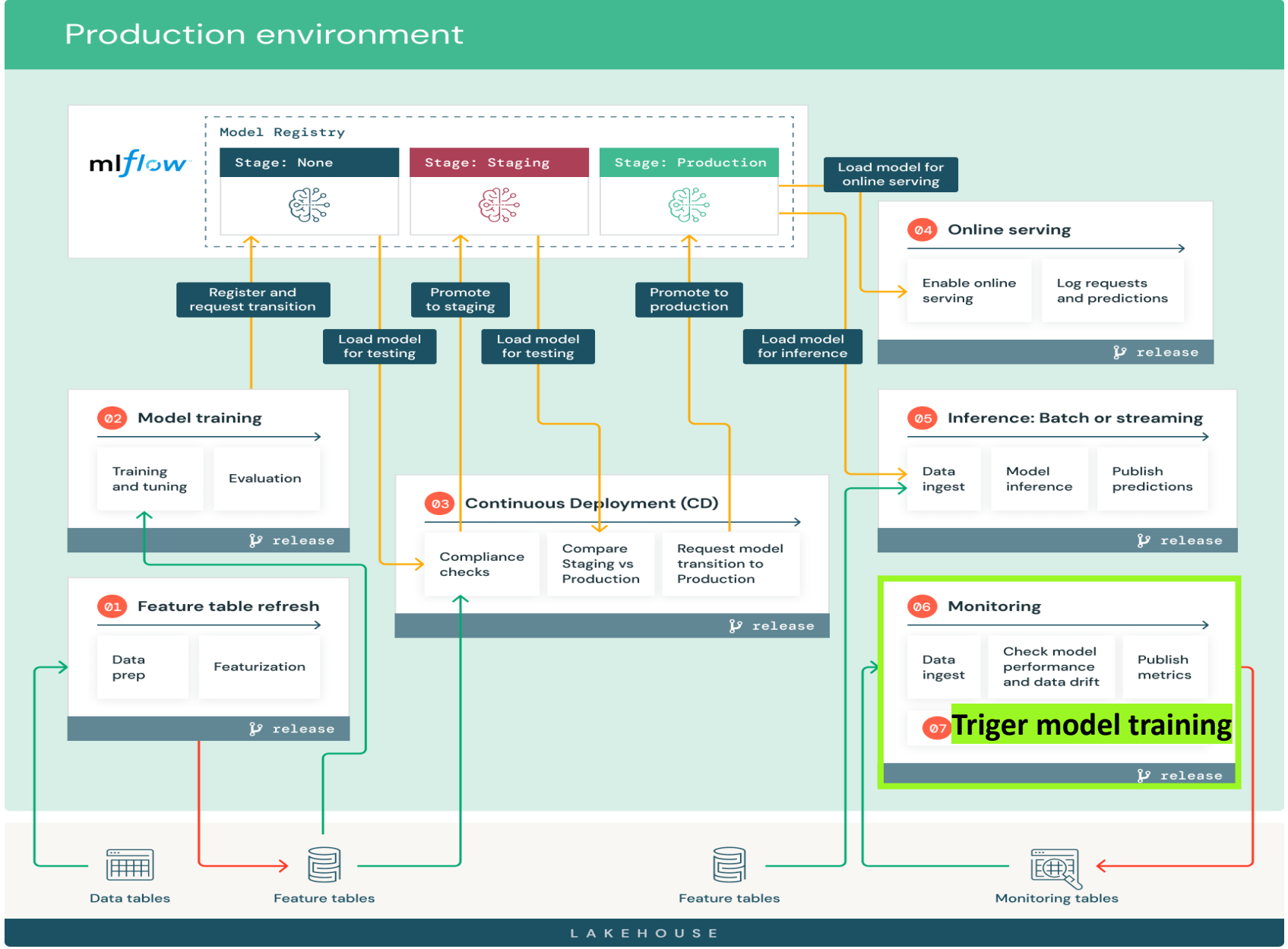
Source https://learn.microsoft.com/en-us/azure/databricks/_static/images/machine-learning/mlops-prod-diagram.png

databricks agileDSS



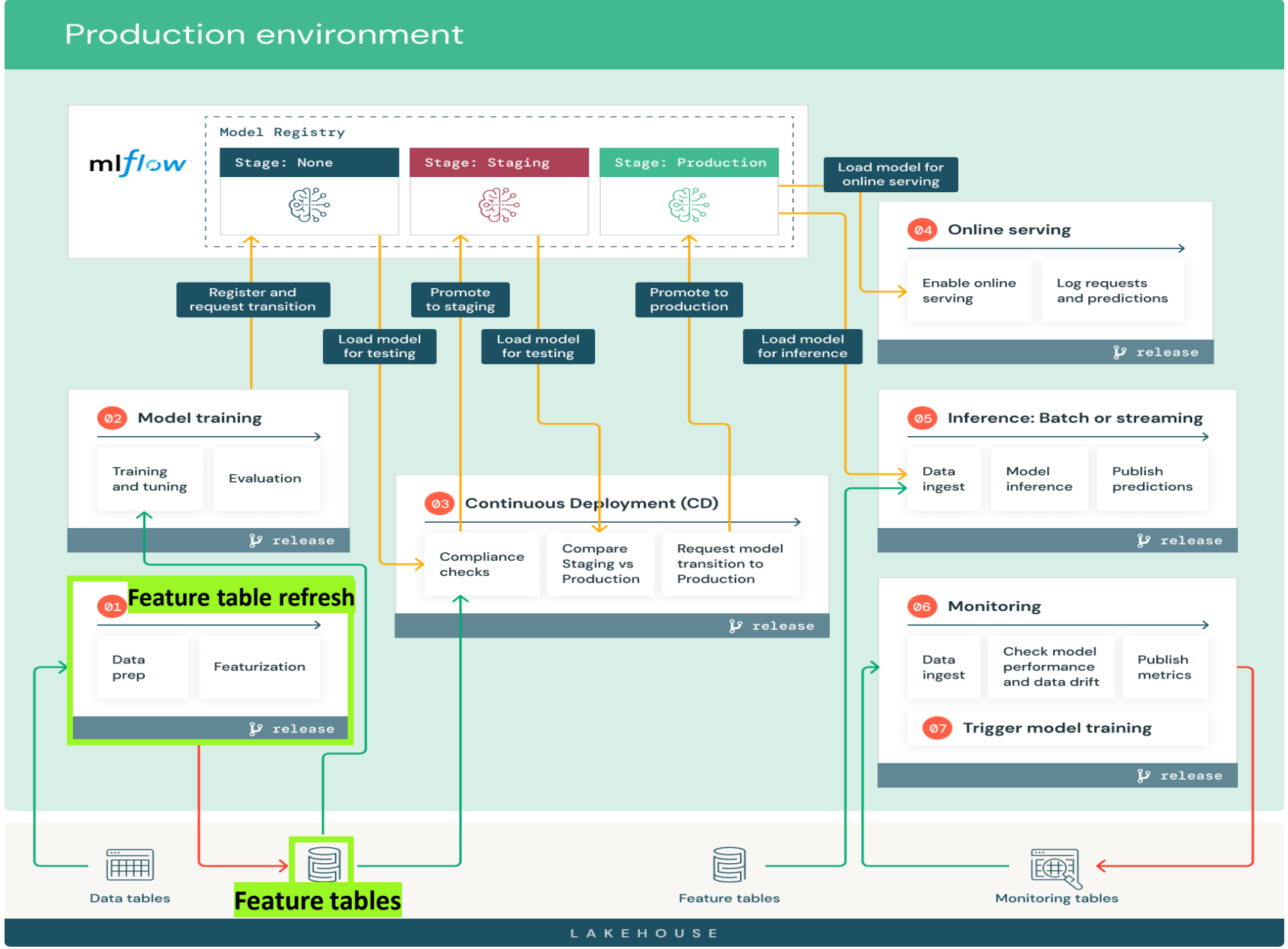
Source https://learn.microsoft.com/en-us/azure/databricks/_static/images/machine-learning/mlops-prod-diagram.png

databricks agileDSS



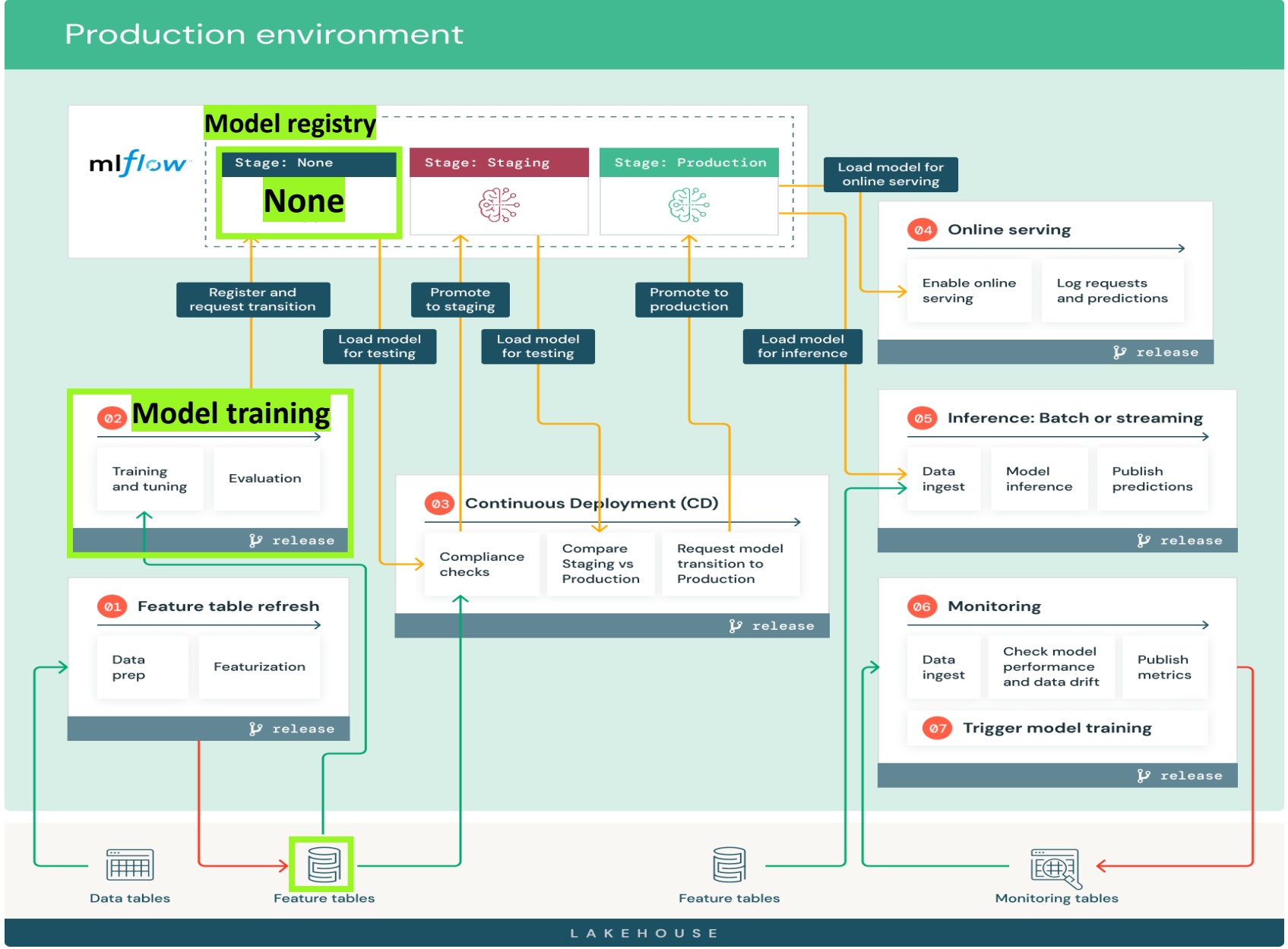
Source https://learn.microsoft.com/en-us/azure/databricks/_static/images/machine-learning/mlops-prod-diagram.png

databricks agileDSS



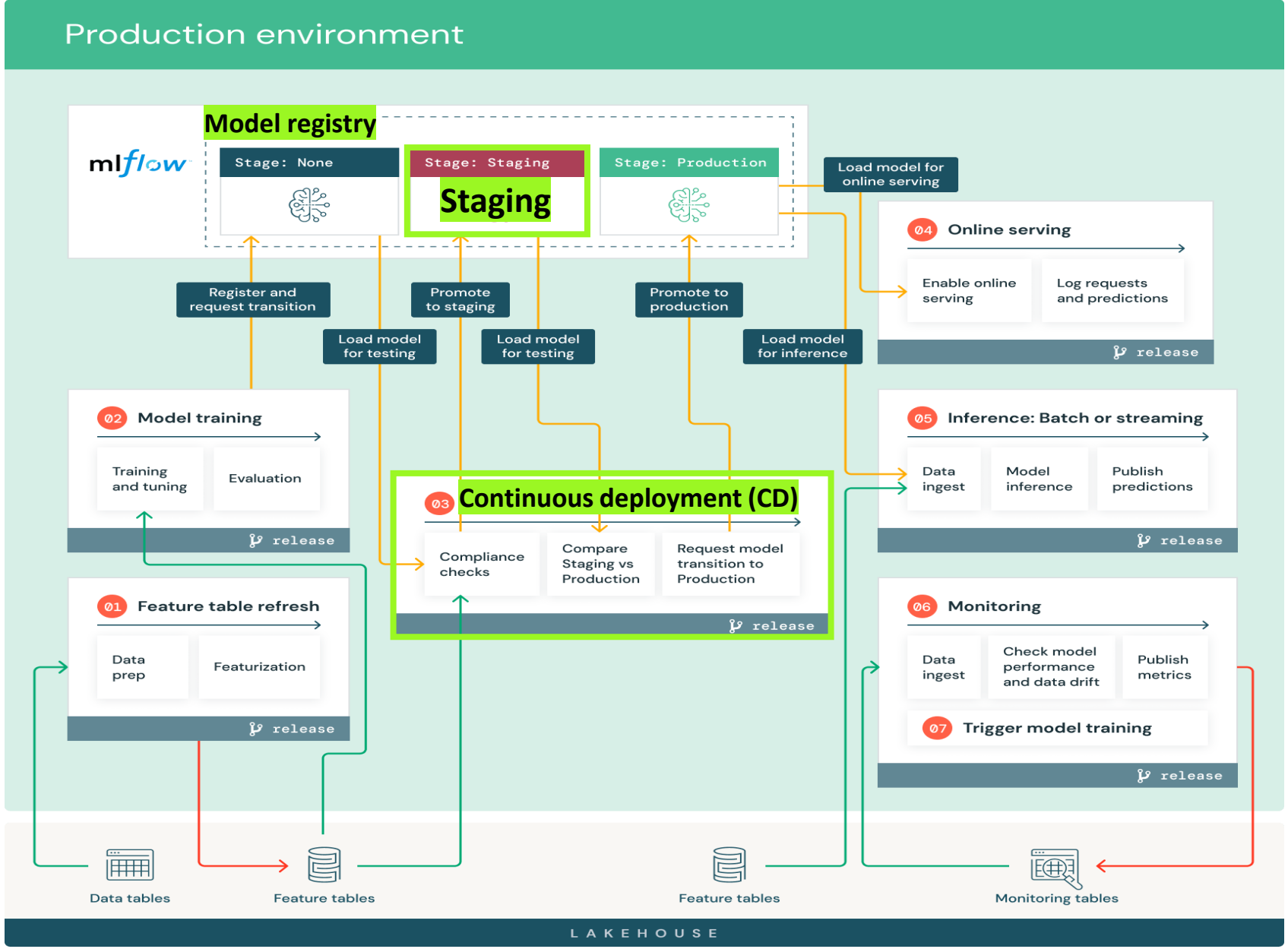
Source https://learn.microsoft.com/en-us/azure/databricks/_static/images/machine-learning/mlops-prod-diagram.png

databricks agileDSS



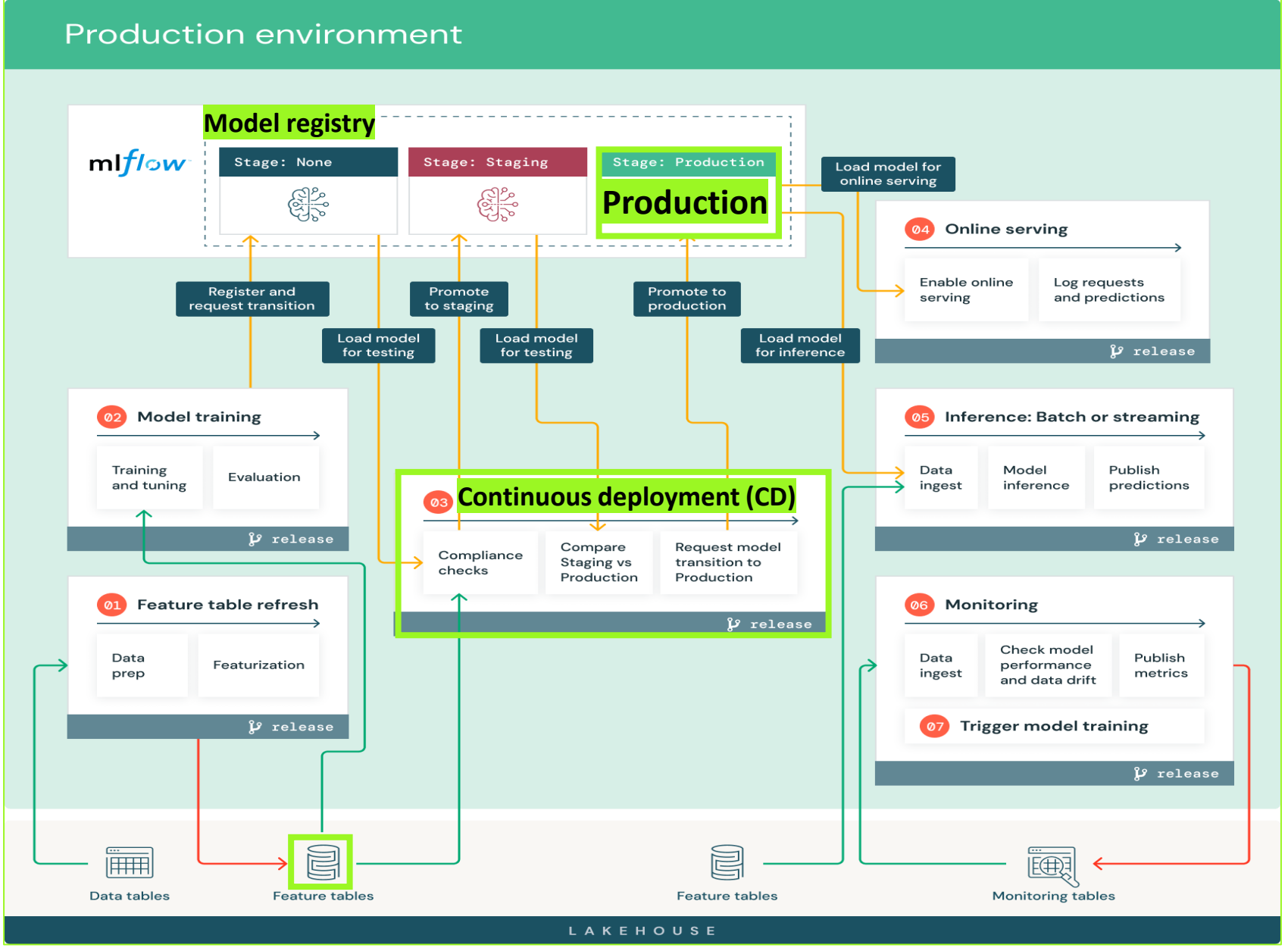
Source https://learn.microsoft.com/en-us/azure/databricks/_static/images/machine-learning/mlops-prod-diagram.png

databricks agileDSS



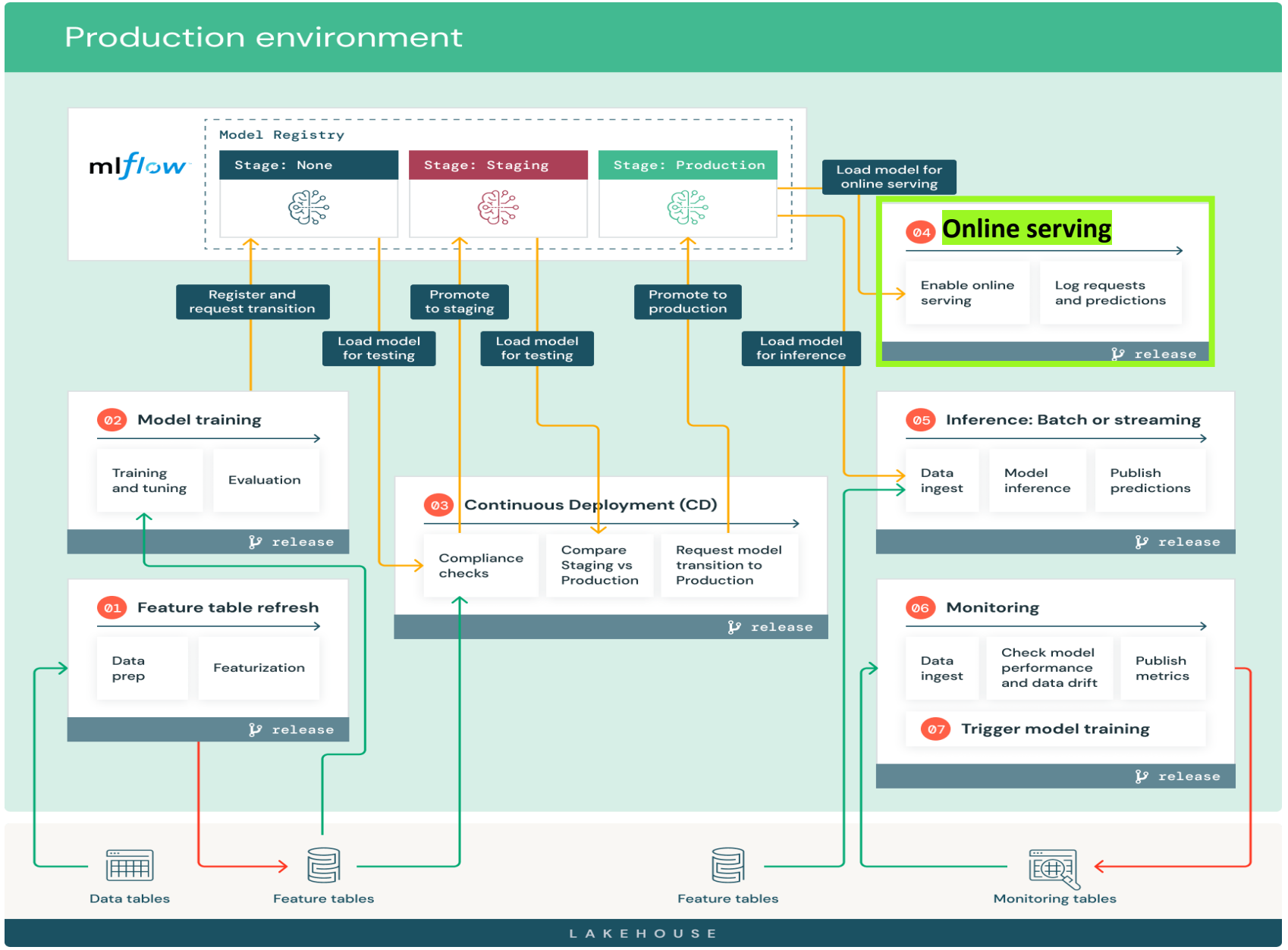
Source https://learn.microsoft.com/en-us/azure/databricks/_static/images/machine-learning/mlops-prod-diagram.png

databricks agileDSS



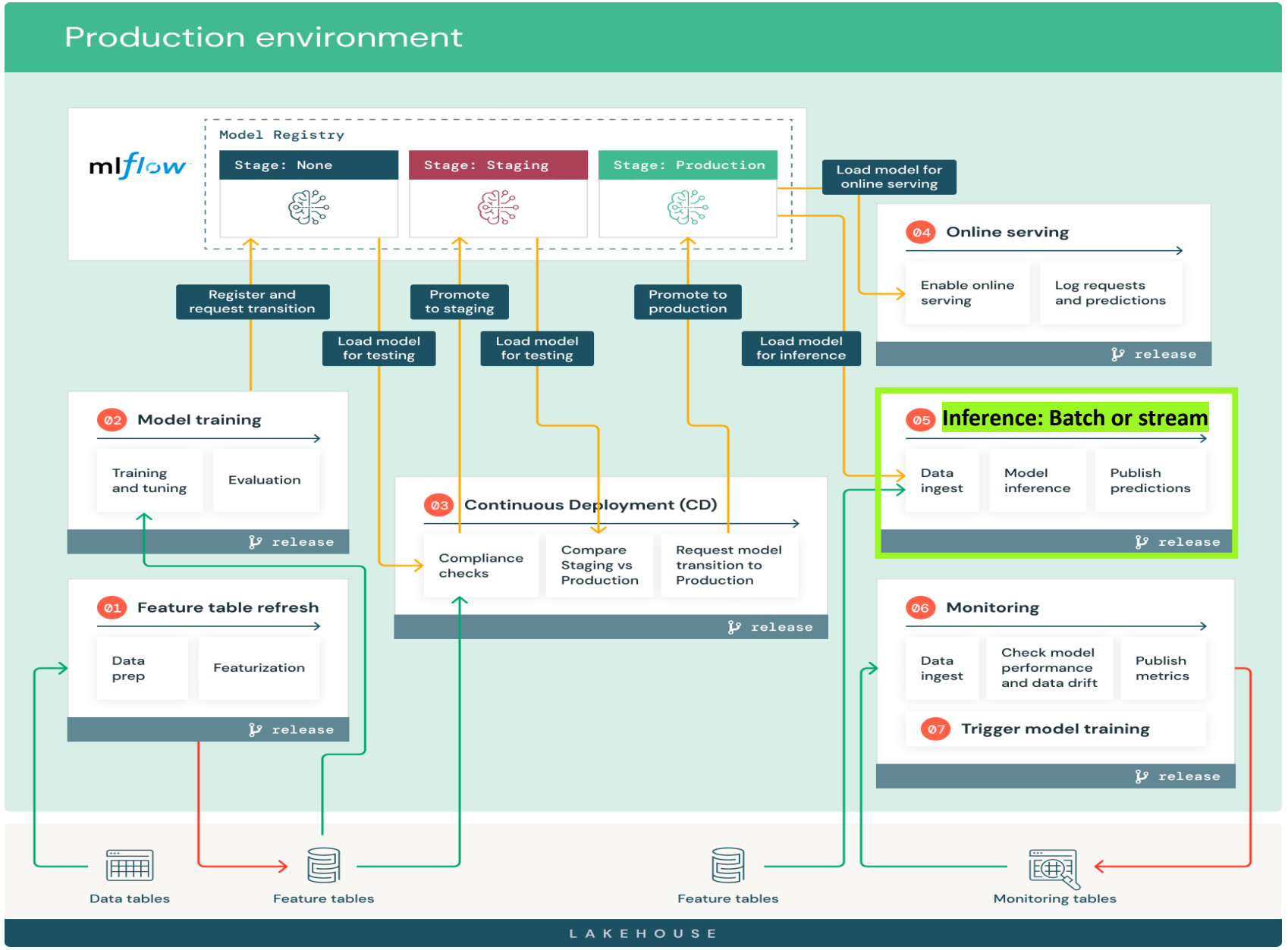
Source https://learn.microsoft.com/en-us/azure/databricks/_static/images/machine-learning/mlops-prod-diagram.png

databricks agileDSS



Source https://learn.microsoft.com/en-us/azure/databricks/_static/images/machine-learning/mlops-prod-diagram.png

databricks agileDSS



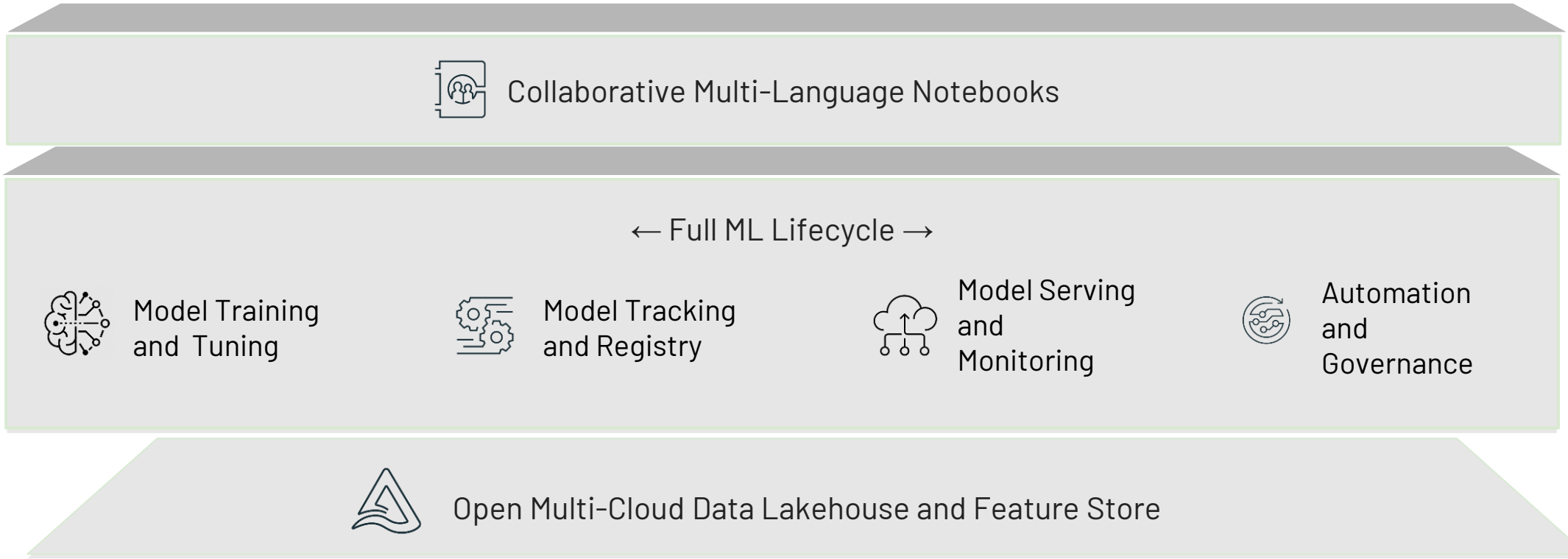
Source https://learn.microsoft.com/en-us/azure/databricks/_static/images/machine-learning/mlops-prod-diagram.png

databricks agileDSS

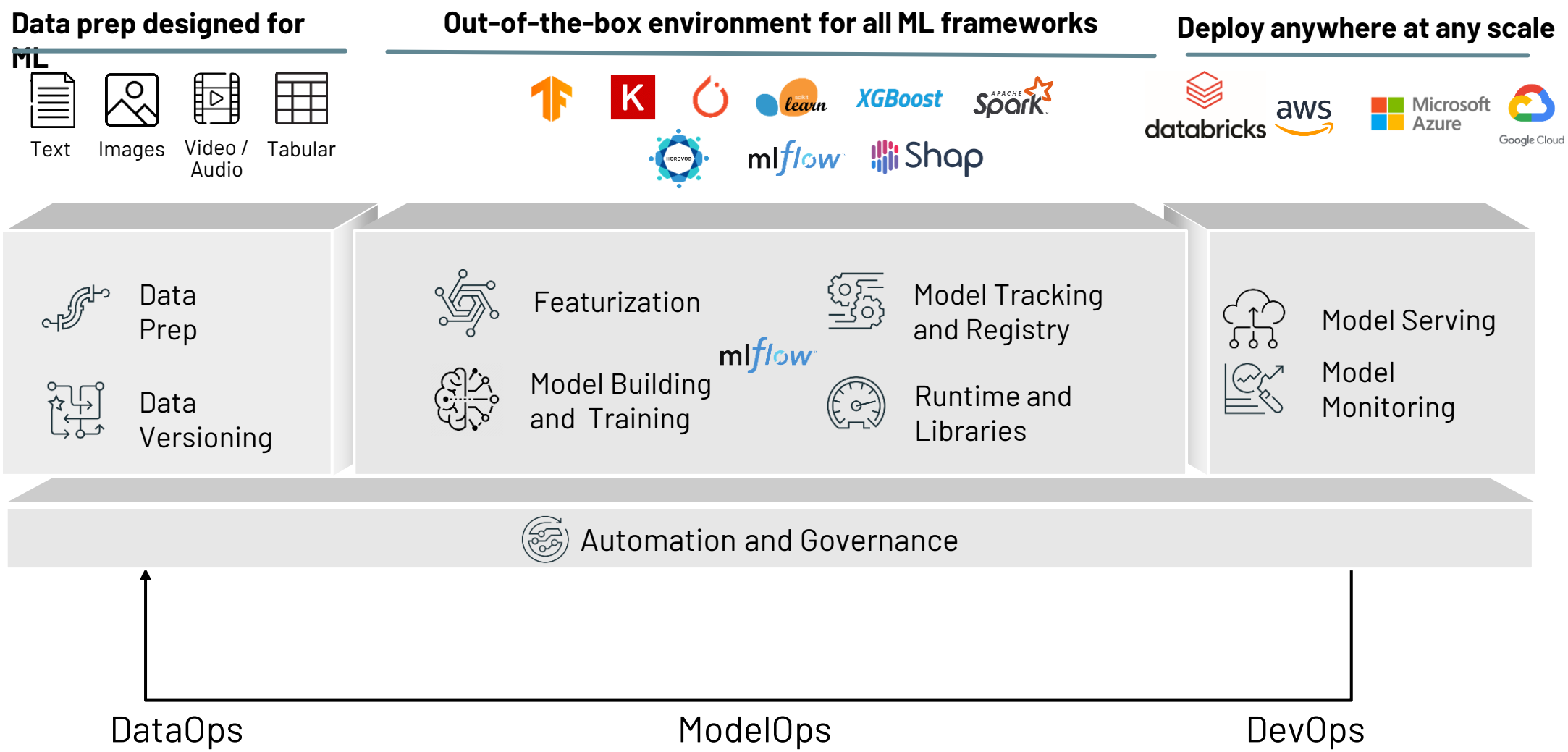
databricks

Gouvernance et MLops

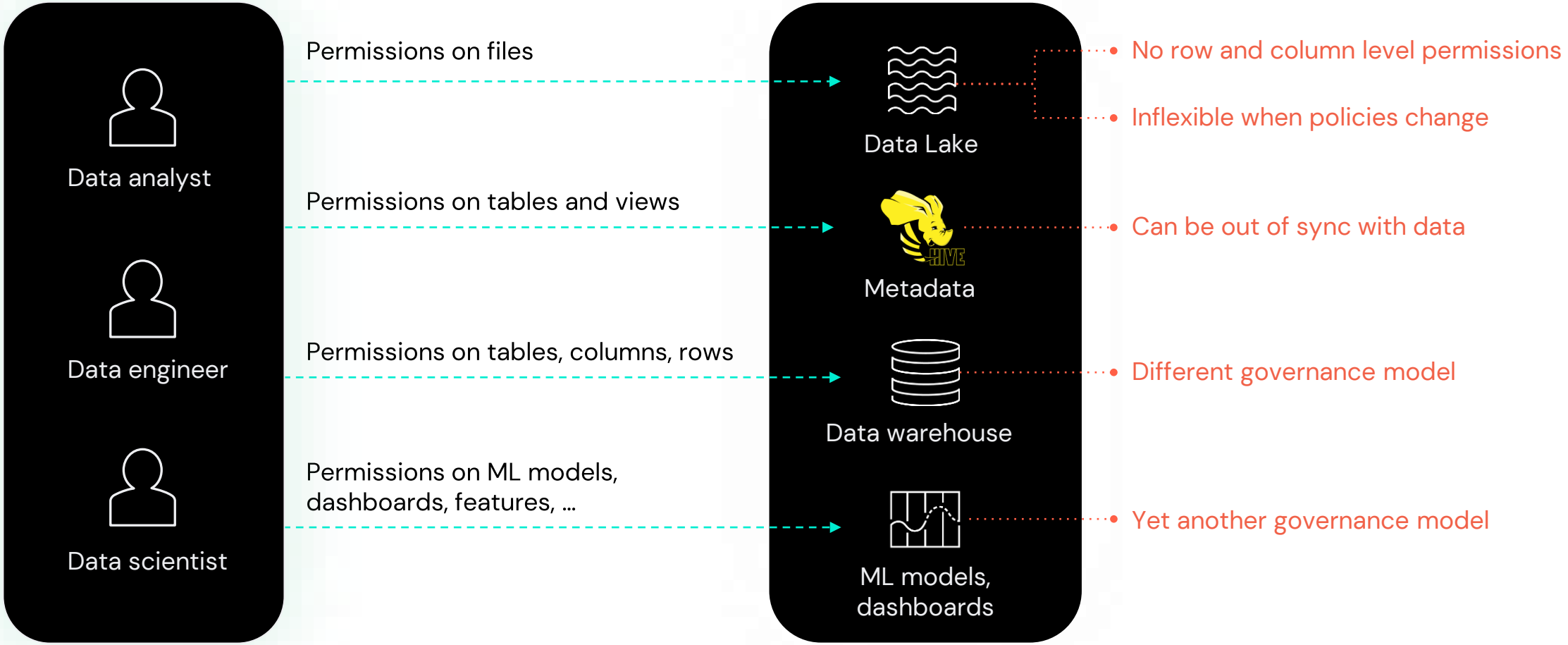
Databricks for Data Science and Machine Learning



Full ML Lifecycle: From Data Ingest to Model Deployment

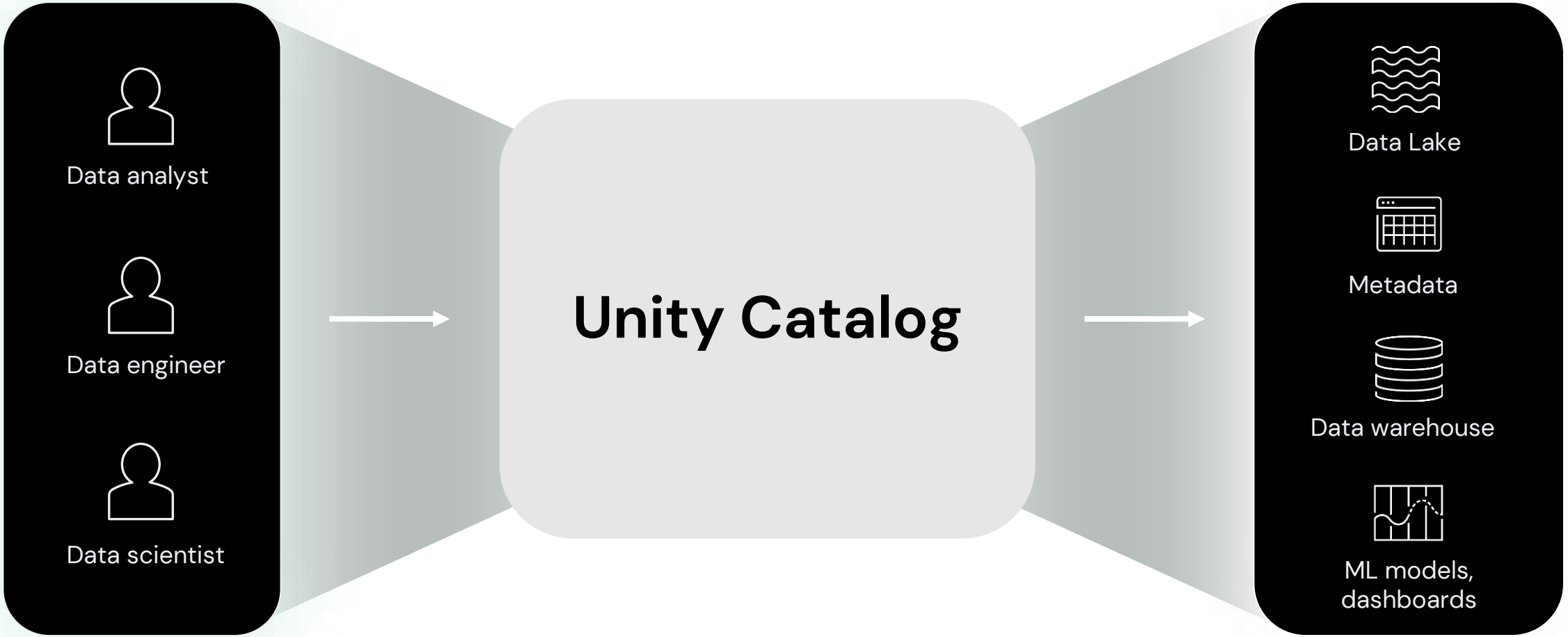


Governance for data, analytics and AI is complex



Databricks Unity Catalog

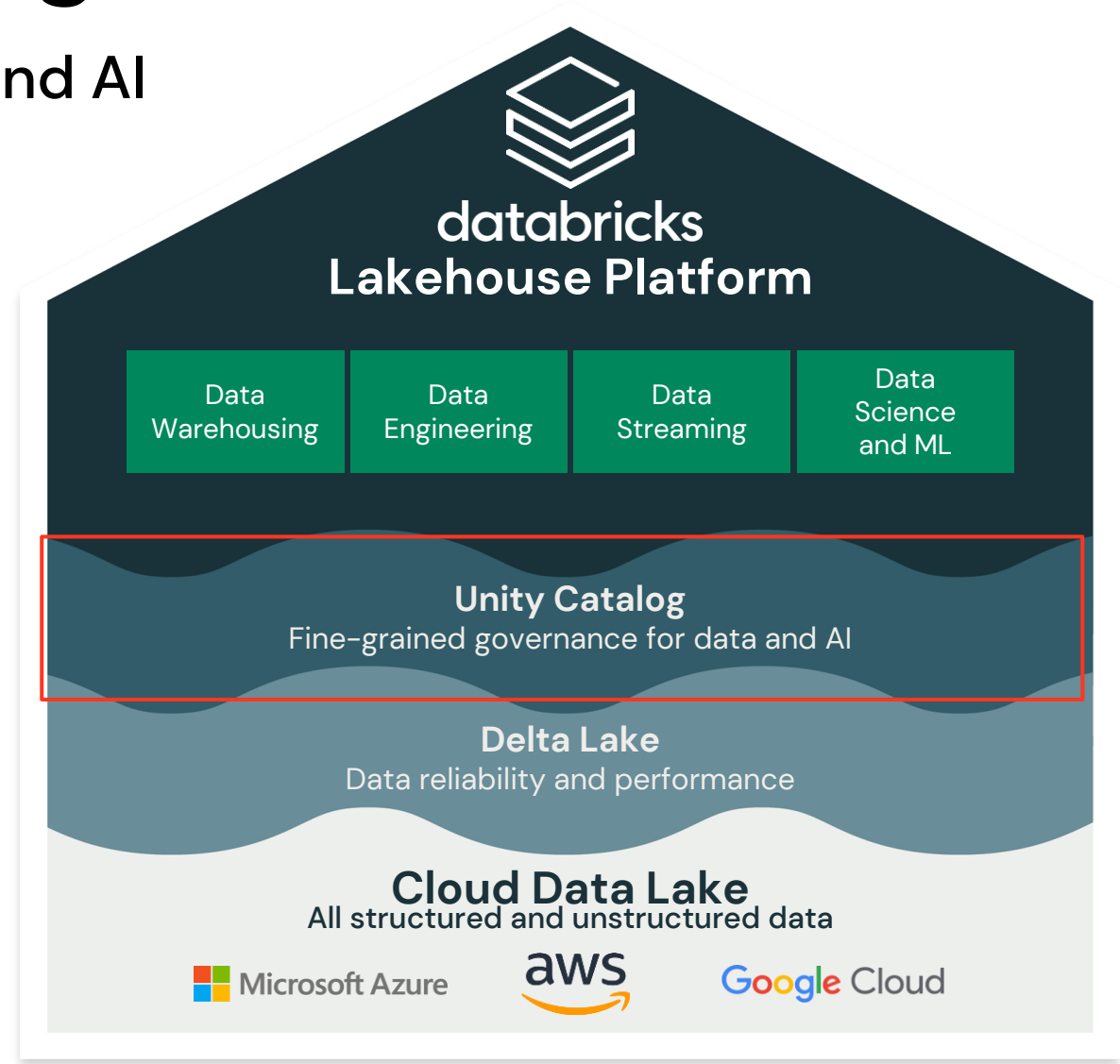
Unified governance for data, analytics and AI



Databricks Unity Catalog

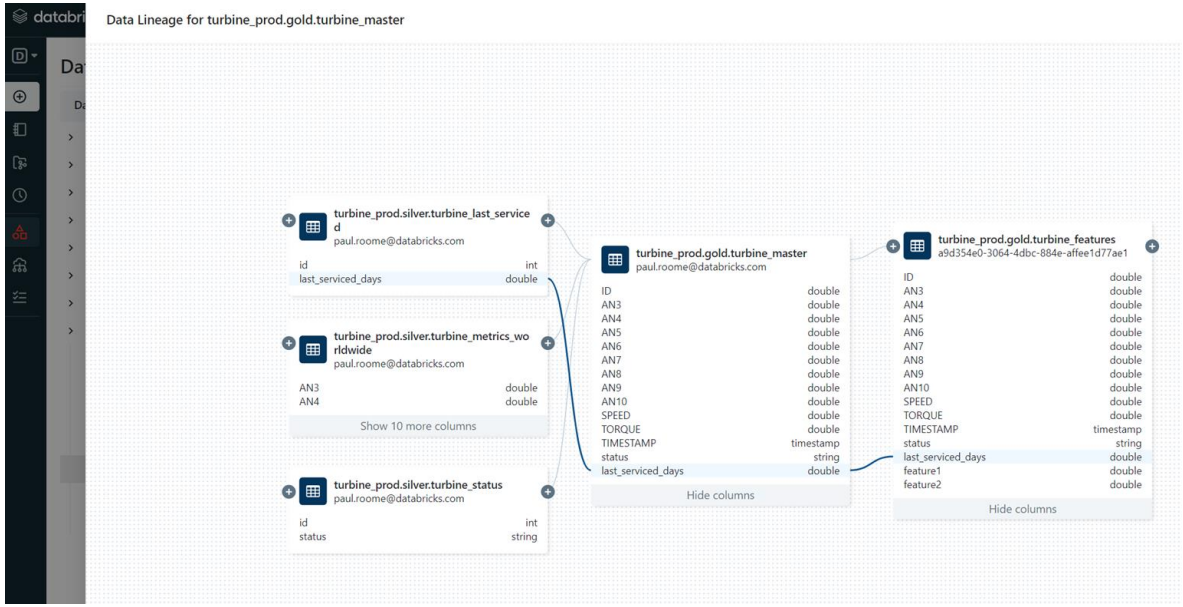
Unified governance for data, analytics and AI

- Map, secure and audit data across clouds
- Open data sharing and collaboration
- Seamless data observability and quality
- Improved productivity and discoverability
- Optimized performance and reduced cost



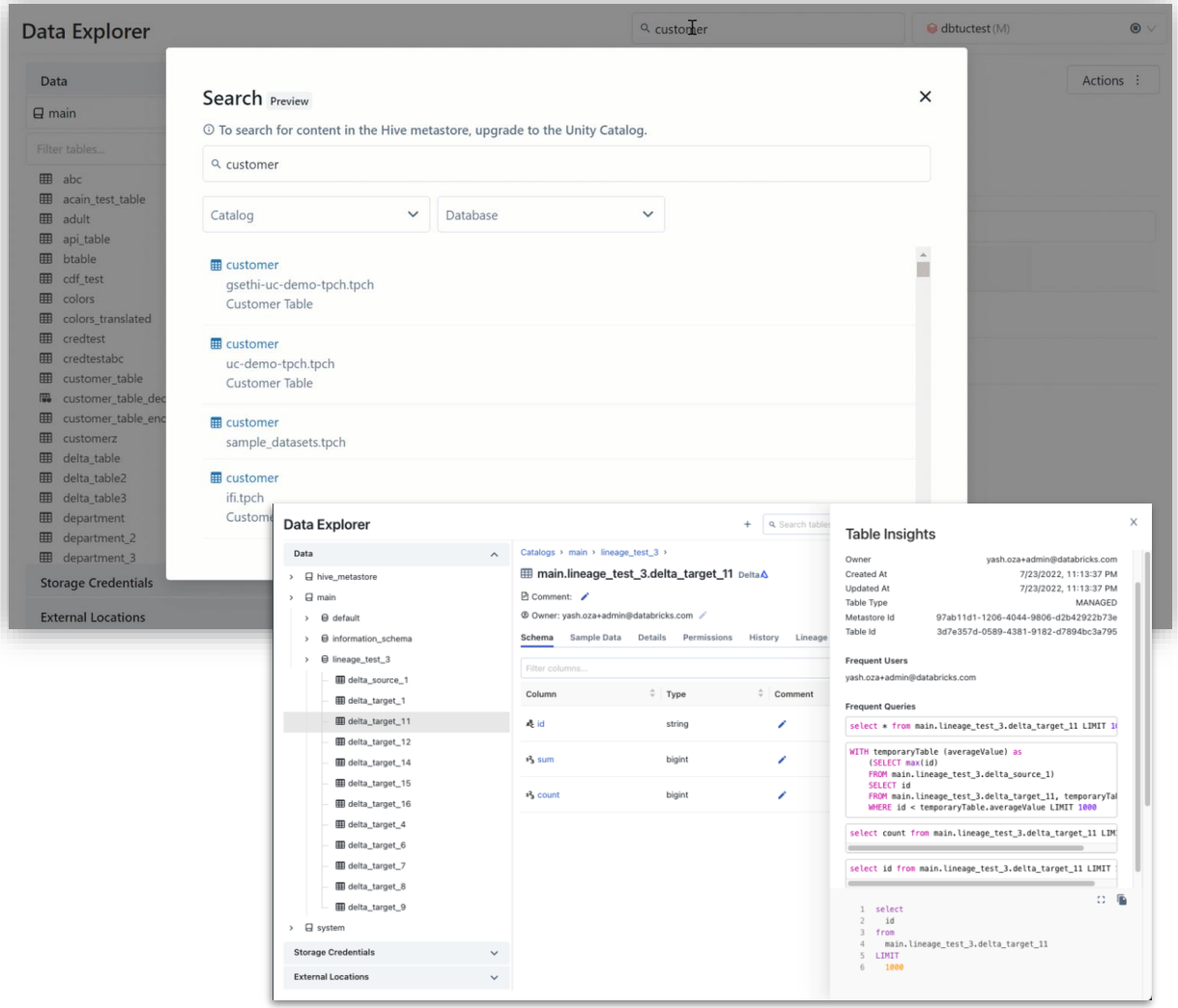
Seamless data observability and quality

- End-to-end visibility into how data flows in your organization with automated data lineage down to column level
- View lineage across tables, columns, notebooks, workflows, dashboards, models, feature stores
- Captured in real time across all workloads in SQL, Python, Scala, and R



Improved productivity and discoverability

- Quickly find, understand, and reference data from across your data estate with a unified data browsing experience
- Secure by default—leverages common permission model from Unity Catalog
- Auto-generated data insights to understand your data



Groupe utilisateurs Montréal / Québec

Joignez le groupe de discussion des utilisateurs Databricks de Montréal et Québec.

Le groupe partagera les nouvelles sur Databricks ainsi que les prochaines réunions du groupe utilisateur

Merci de votre attention

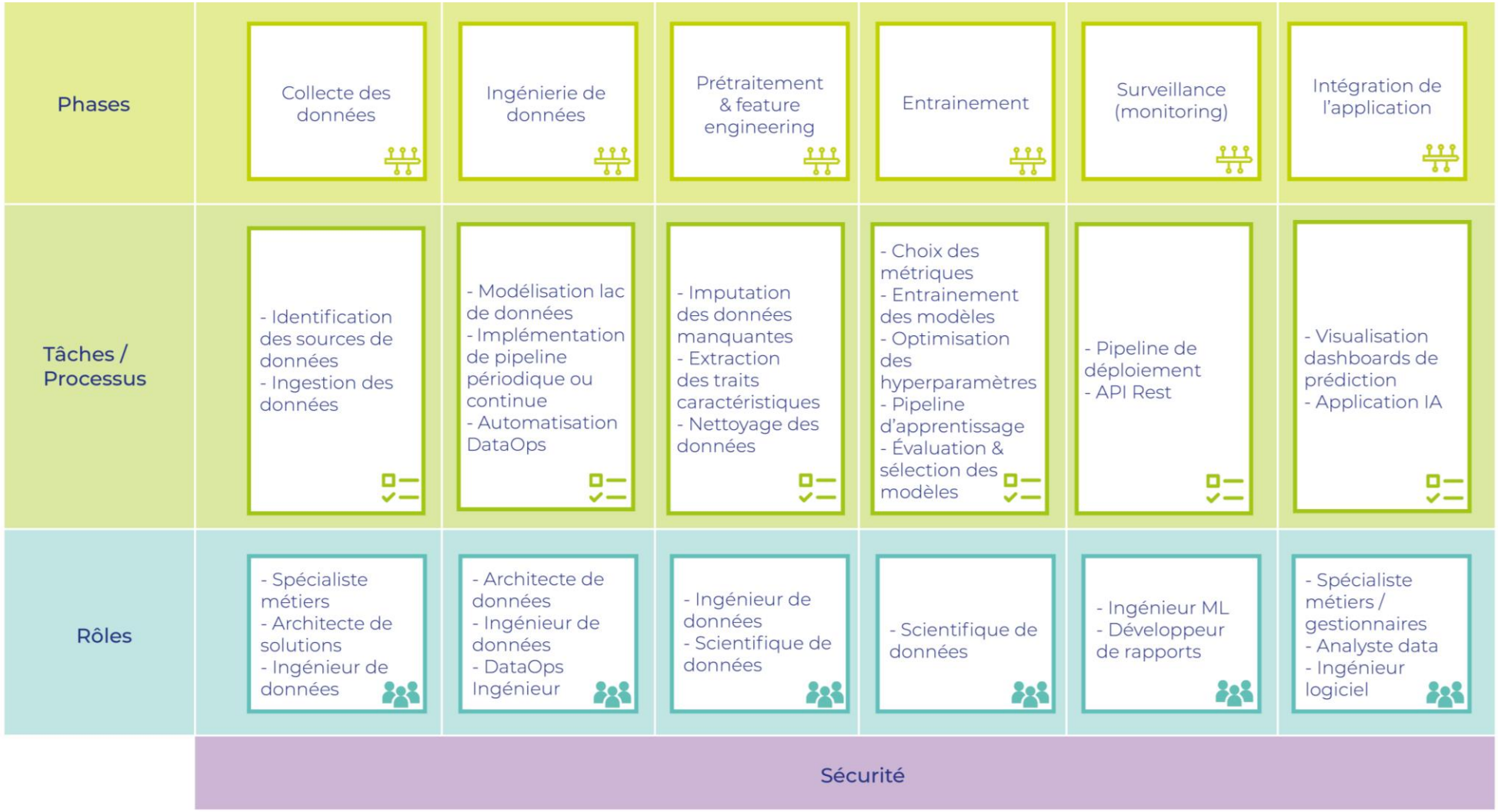
Des questions?

databricks

407 Rue McGill bureau 500, Montréal

2828 Bd Laurier Suite 700, Québec

Casse-tête ML : Vers le chemin du succès



agileDSS