

# MLOps – Machine Learning Operations

**Machine Learning Architects Basel**

Bassem Ben Hamed

December 2022





# Plan

- **C'est quoi MLOps ?**
- **Principes du MLOps**
- **DevOps vs MLOps**
- **Niveaux de maturité**
- **Données et cas d'utilisation**



# C'est quoi MLOps ?

MLOps (Machine Learning Operations) est un ensemble d'outils et de pratiques visant à assurer une intégration transparente entre le cycle de vie d'un projet ML et le processus d'exploitation global. MLOps a pour but d'atteindre principalement 3 objectifs :

01

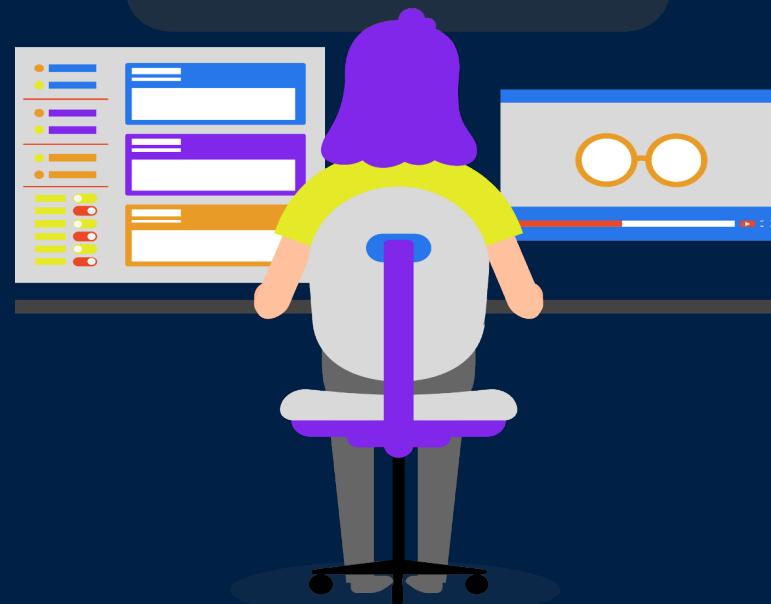
Accélérer  
l'expérimentation et le  
développement de  
modèles

02

Déploiement plus rapide des  
modèles mis à jour dans la  
production

03

Assurance  
qualité



# C'est quoi MLOps?

## Phases du MLOps

Le processus MLOps comprend principalement 3 phases : "Conception de l'application alimentée par le ML", "Expérimentation et développement du ML" et "ML pour la production". Ces trois phases sont interconnectées et s'influencent mutuellement.



- Ingénierie des exigences
- Hiérarchisation des cas d'utilisation du ML
- Vérification de la disponibilité des données
- Ingénierie des données
- Ingénierie des modèles ML
- Test et validation
- Déploiement de modèles
- Pipelines CI/CD
- Surveillance et déclenchement

# Principes du MLOps

## Automation



Automate the workflow steps without any manual intervention

## Continuous X



Includes Continuous Integration, Deployment, Testing & Monitoring

## Versioning



Tracks ML models and data sets with version control systems

## Experiment Tracking



Multiple experiments on model training are parallelly executed

## Testing



Tests for features & data, model development & infrastructure

## Monitoring



Monitor to assure that the ML model performs as expected

## Reproducibility



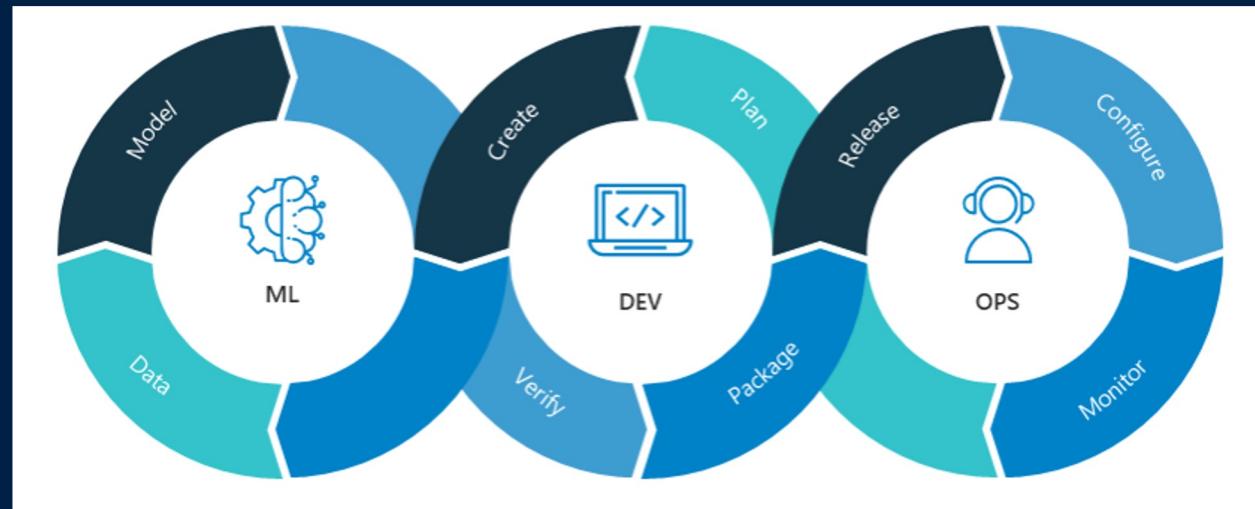
Should produce identical results given the same input

# DevOps VS MLOps

Devops ne peut être défini comme un concept unique ; c'est une combinaison de tout ce qui crée le processus des release du logiciel tout en assurant :

- la rapidité
- une qualité élevée
- une meilleure communication et collaboration entre les développeurs et les responsables des opérations.

DevOps est issu du génie logiciel et concerne le développement et les opérations de production de logiciels à grande échelle. D'autre part, MLOps est dédié aux systèmes ML.



# DevOps VS MLOps

## Comment fonctionne MLOps ?

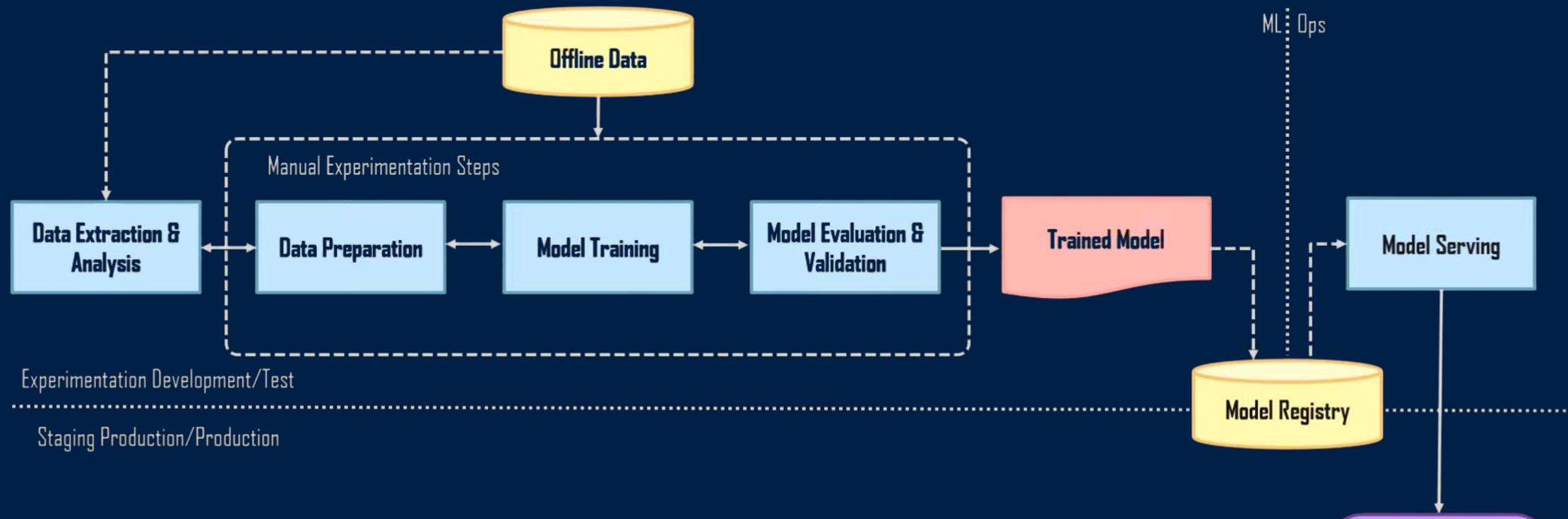
Le processus de livraison d'un modèle ML à une production comprend principalement les étapes suivantes :



# Niveaux de maturité MLOps

## Niveau 0 : Processus manuel

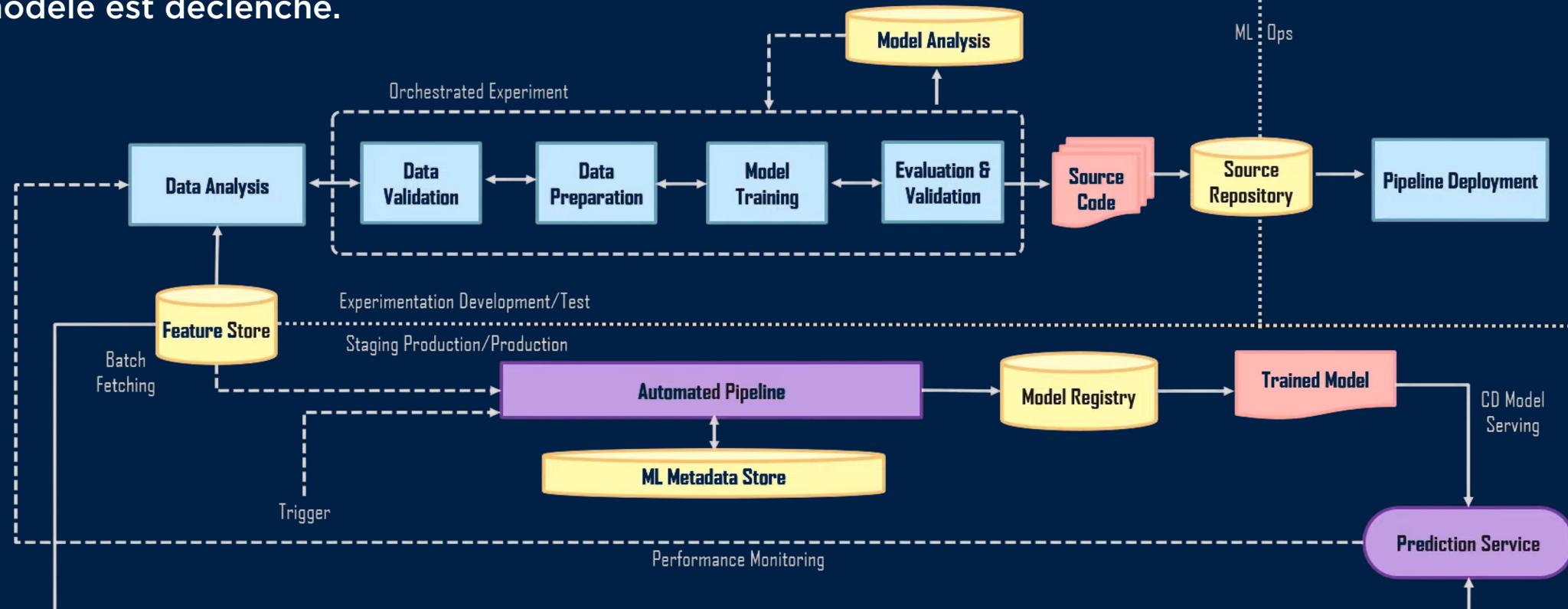
Il s'agit du processus typique de la science des données. Chaque étape de chaque pipeline, comme la préparation et la validation des données, la formation et le test des modèles, est exécutée manuellement. mise en scène de la production



# Niveaux de maturité MLOps

## Niveau 1 : Automatisation du pipeline ML

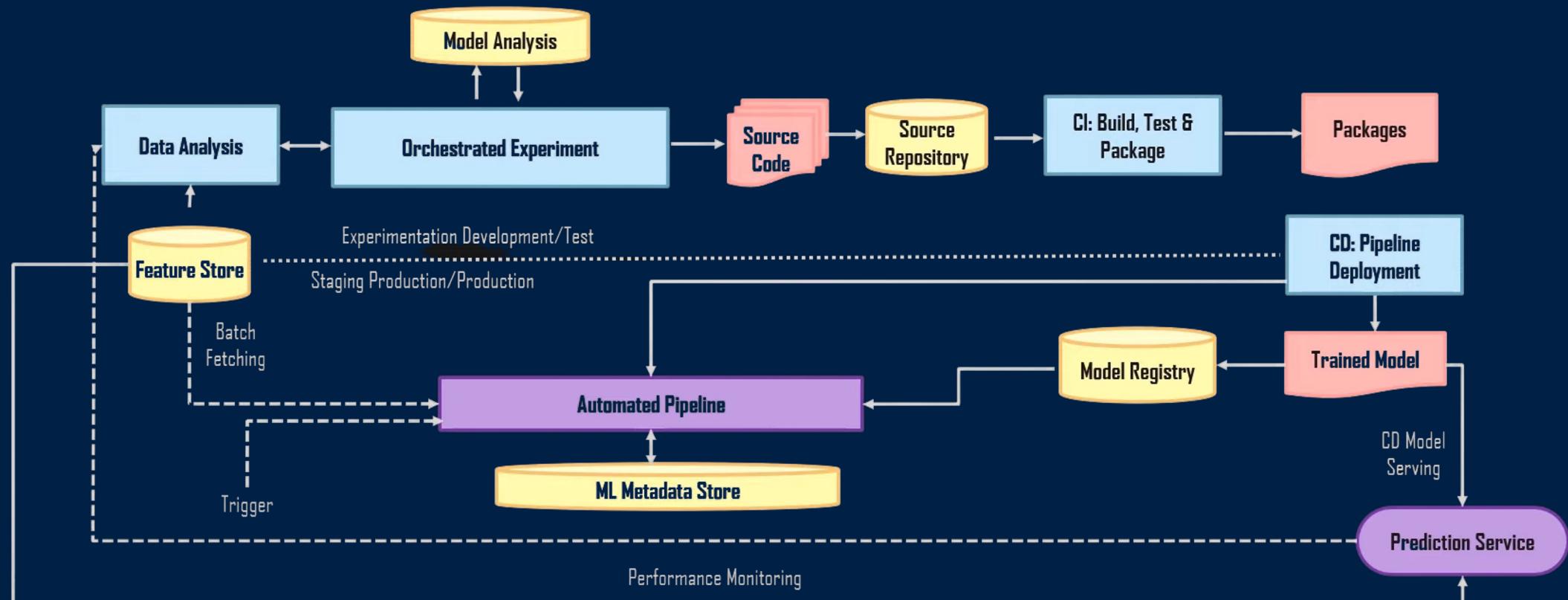
Ce niveau comprend l'exécution automatique de l'entraînement du modèle. On introduit ici la formation continue (CT) du modèle. Dès que de nouvelles données sont disponibles, le processus de réentraînement du modèle est déclenché.



# Niveaux de maturité MLOps

## Niveau 2 : Automatisation du pipeline CI/CD

On introduit un système CI/CD pour effectuer des déploiements rapides et fiables de modèles ML en production. La principale différence par rapport à l'étape précédente est qu'on construit, teste et déploie maintenant automatiquement les composants des données, du modèle ML et du pipeline de formation ML.



# Données et cas d'utilisation

## Qu'est-ce que la fraude ?

**La fraude bancaire est l'utilisation de moyens potentiellement illégaux pour obtenir de l'argent, des actifs ou d'autres biens appartenant ou détenus par une institution financière.**



# Données et cas d'utilisation

## Pourquoi utiliser ML pour la détection des fraudes ?

L'idée derrière l'utilisation de la ML est que les transactions frauduleuses présentent certains modèles qui les différencient des transactions authentiques. Les algorithmes de ML reconnaissent ces modèles et sont capables de différencier les fraudeurs des clients légitimes.



# Données et cas d'utilisation

## Jeu de données pour la détection des fraudes dans les transactions par carte de crédit

Le jeu de données qu'on va utiliser est hébergé sur la plateforme Kaggle.

Il s'agit d'un ensemble de données de transactions par carte de crédit simulées contenant des transactions légitimes et frauduleuses pour la période du 1er janvier 2019 au 31 décembre 2020. Il couvre les cartes de crédit de 100K clients effectuant des transactions avec un pool de 800 marchands.

### Credit Card Transactions Fraud Detection Dataset

Simulated Credit Card Transactions generated using Sparkov

