

از سلسله وبینارهای

گروه تخصصی معماری سازمانی انجمن انفورماتیک ایران

**{مقدمه ای بر MLOps و معماری
سازمانی مبتنی بر هوش مصنوعی
{Introduction to MLOps and AI
Based Enterprise architecture}**

{رضا گنجی}

{مدیر تولید نرم افزار}

{عنوان وبینار}
{webinar title}

آنچه خواهیم گفت ...

معرفی

مقدمه ای بر DevOps

چرایی MLOps

معرفی MLOps

ساختار سازمانی مبتنی بر MLOps

معرفی ابزارها

معرفی MIFlow



{عنوان وبینار}
{webinar title}

مراجع :

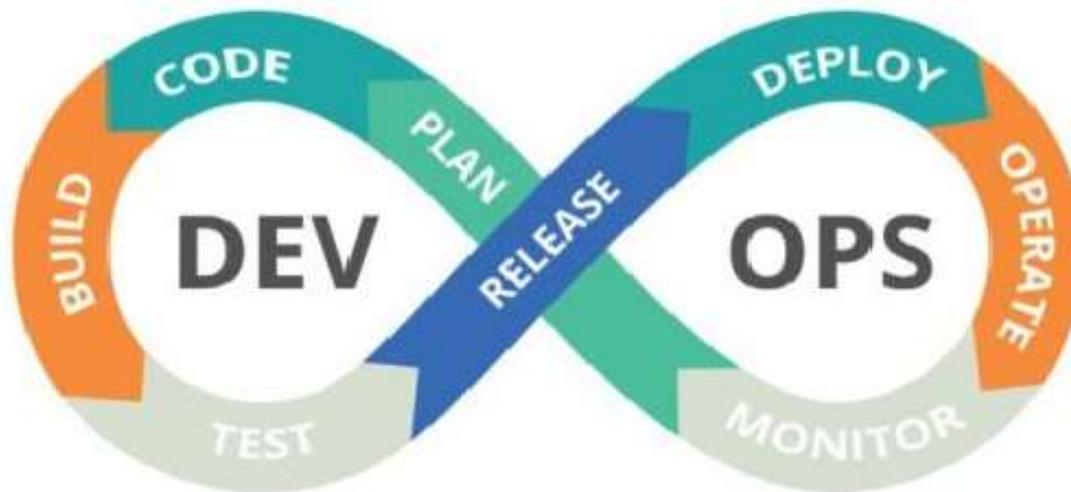
- Practical MLOps by Noah Gift, Alfredo Deza
- RedHat MLOps base model

{عنوان ويبينار}
{webinar title}

- عدم نگهداری تاریخچه الگوریتم ها و دلیل انتخاب مدل ها
- عدم وجود مانیتورینگ
- عدم وجود چرخه CI/CD
- عدم چابکی
- عدم نگهداری اجزای مختلف
- عدم امکان مقایسه سریع الگوریتم های موجود

مقدمه ای بر DevOps

- تضمین کیفیت - عملکردی، غیرعملکردی و یکپارچگی داده‌ها
- امنیت و کنترل - انطباق با استانداردهای صنعت
- عملیات - انطباق در دسترس بودن بر اساس قراردادهای سطح خدمات
- چابکی



{عنوان وبینار}
{webinar title}

Wikipedia :

MLOps or ML Ops is a paradigm that aims to deploy and maintain machine learning models in production reliably and efficiently. The word is a compound of "machine learning" and the continuous development practice of DevOps in the software field. Machine learning models are tested and developed in isolated experimental systems. When an algorithm is ready to be launched, MLOps is practiced between Data Scientists, DevOps, and Machine Learning engineers to transition the algorithm to production systems.

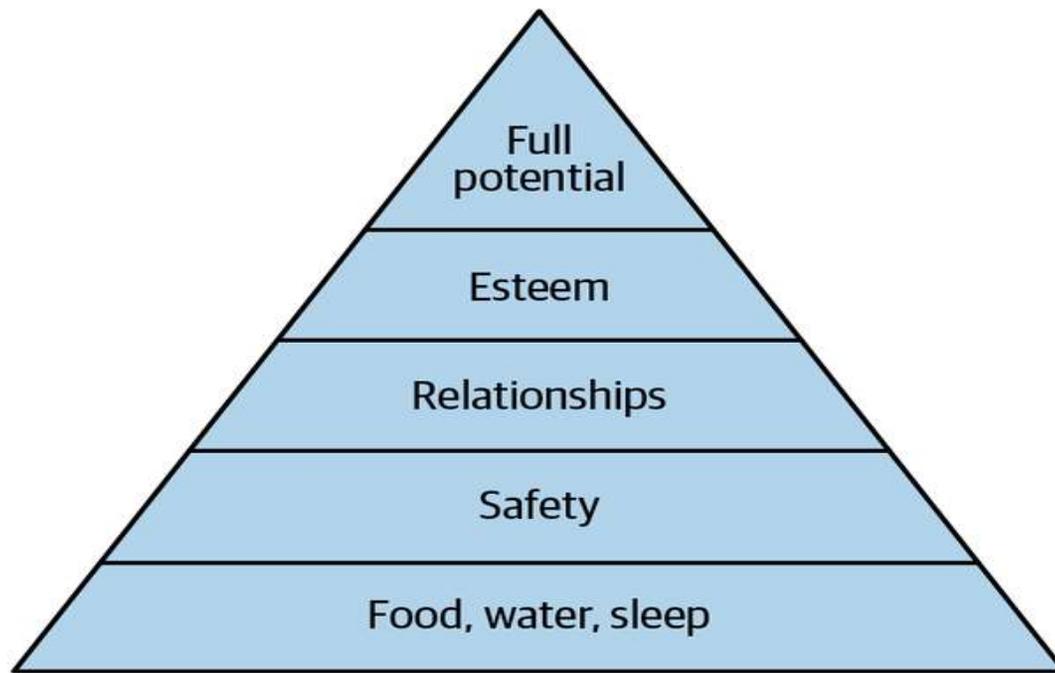
چرا MLOps ؟

- سرعت بیشتر ارائه محصولات (Faster Time To market)
- بهره وری بیشتر : استاندارد کردن و هزینه های عملیاتی کمتر
- تشخیص ریسک : نظارت بیشتر , پیش بینی ریسک , امنیت و ...

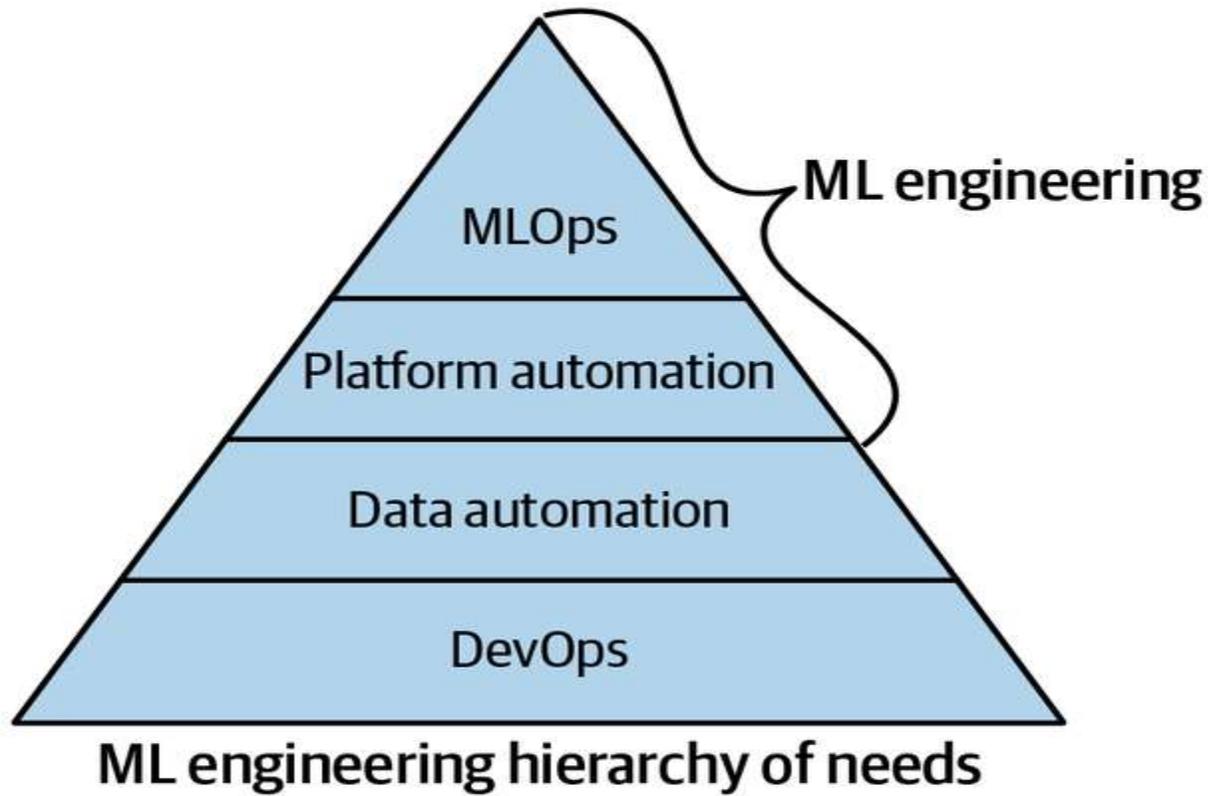
شبهات DevOps و MLOps

- Continues Integration
- Continues Delivery
- Microservices
- Infrastructure As a code
- Monitoring And Instrumentation
- Effective Technical Communication
- Effective Technical Project Management

{عنوان وینار}
{webinar title}



Maslow's hierarchy of needs



Data Automation یا Data Ops

Data Ops پایه اصلی MLOps هست و هدف ان ادغام DevOps و متدولوژی های چابک و تولید ناب هست برای رسیدن به دادهایی در راستای اهداف سازمان

{عنوان وبینار}
{webinar title}

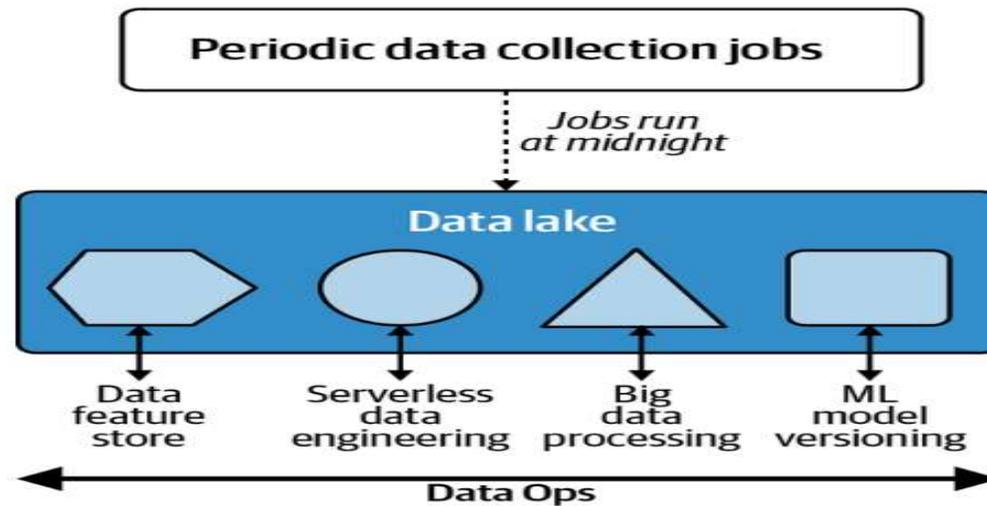
- همسو کردن داده ها با کسب و کار
- فرآیندهای چابک برای مدیریت داده ها و توسعه تجزیه و تحلیل فرآیندهای
- DevOps برای بهینه سازی کد، ساخت محصول و تحویل

پارادایم های اصلی Data Ops

- Cross-functional communication
- Agile mindset
- Integrated data pipeline
- Data-driven culture
- Continuous feedback

{عنوان وبینار}
{webinar title}

- Periodic collection of data and running of jobs
- Processing streaming data
- Serverless and event-driven data
- Big data jobs
- Data and model versioning for ML engineering tasks

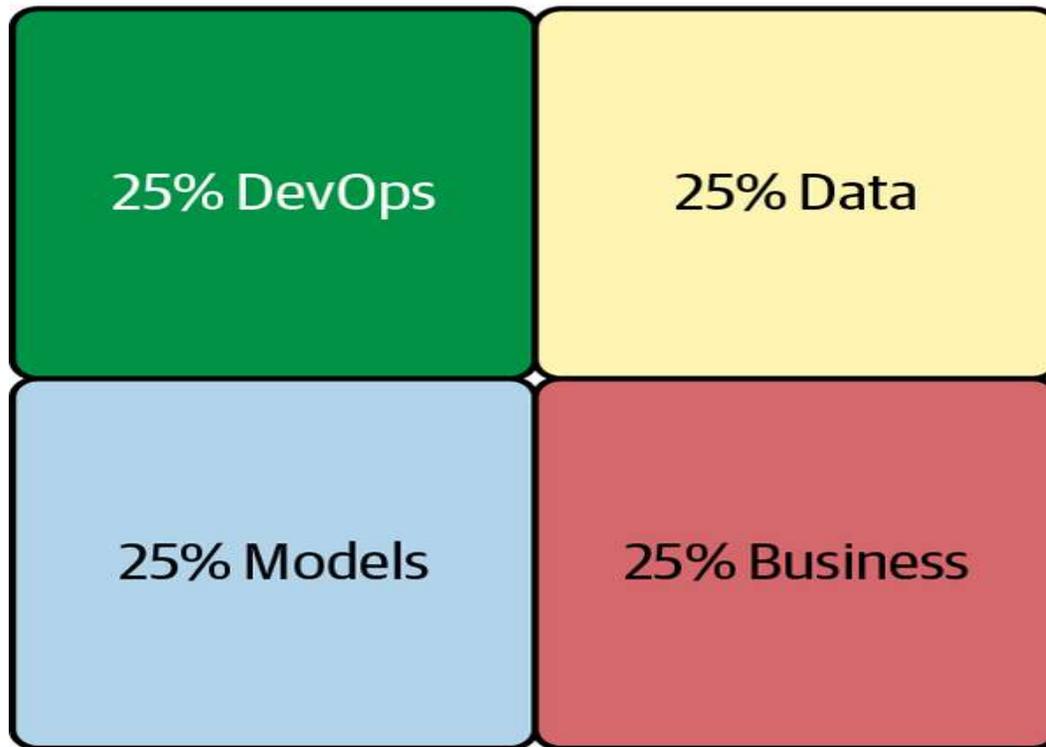


Platform Automation

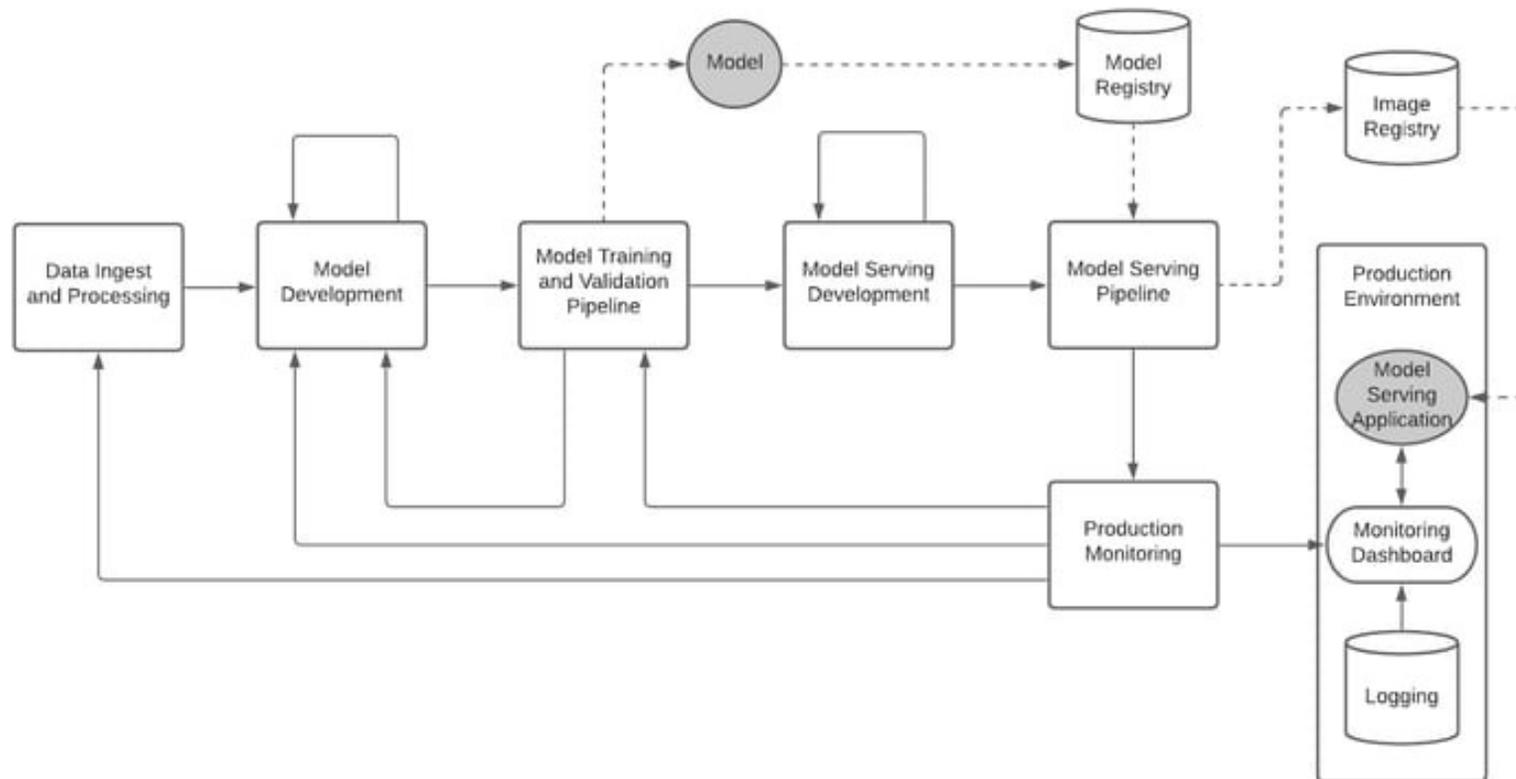


{عنوان ويبينار}
{webinar title}

MLOPS



معماری مرجع MLOps

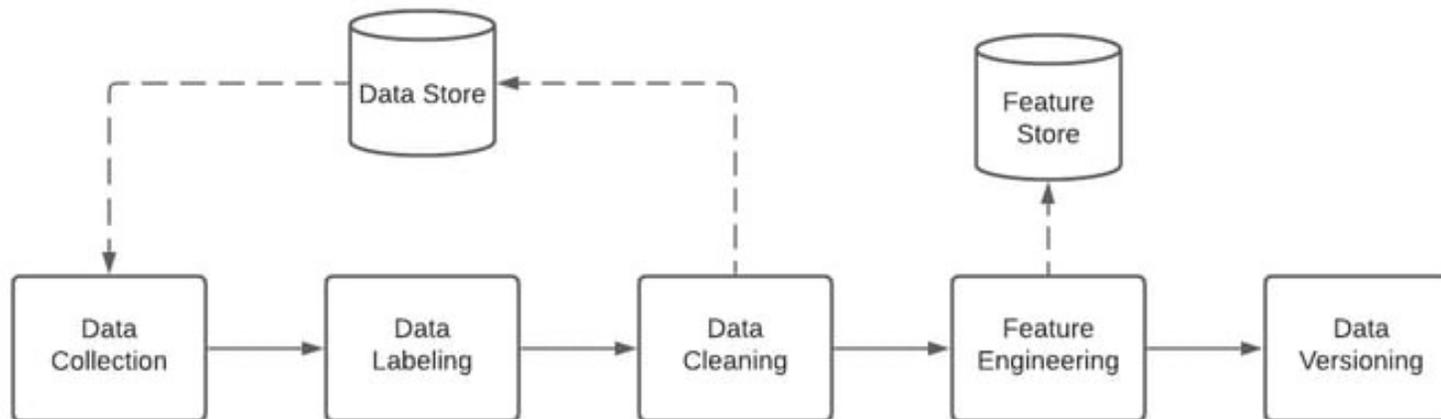


{عنوان وبینار}
{webinar title}

جمع اوری و پردازش داده ها (Data Ingest and Processing)

مهندس داده ها :

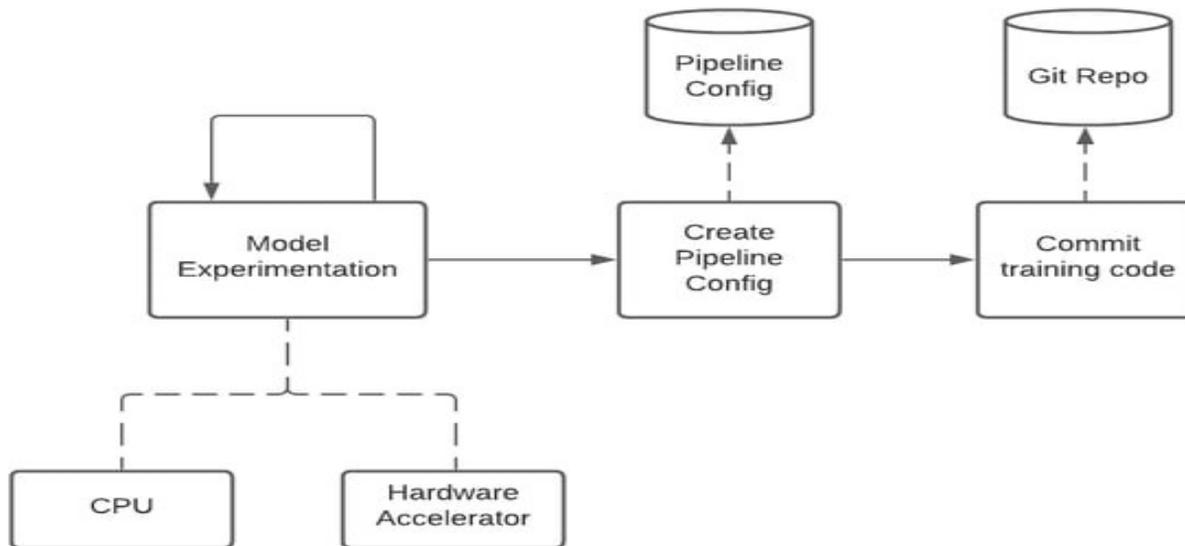
- جمع اوری داده ها از دیتابیس یا data lake (SQL database or HDFS/S3) یا از چند منبع مختلف
- برچسب گذاری داده ها
- پاکسازی داده ها (Normalization or Deduplication)
- مهندسی بردار ویژگی ها



توسعه مدل (Model Development)

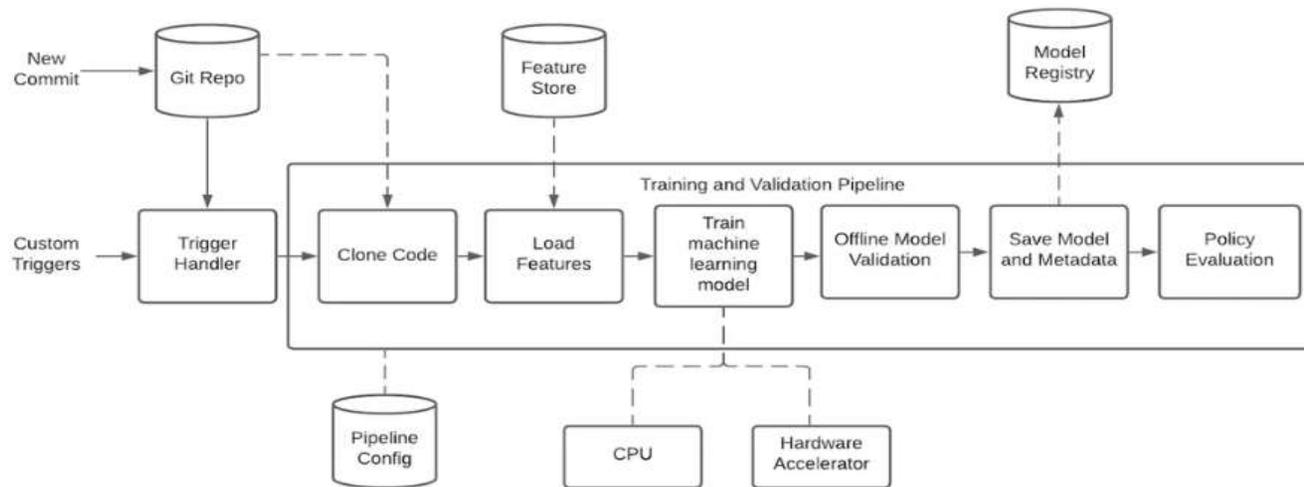
:Data scientist

- نوشتن کد برای ساختن مدل یادگیری ماشین
- تولید مدل های مختلف با استفاده از داده ها
- مقایسه مدل ها با ابزار یا ماتریس سردرگمی (Confusing matrix)
- ارزیابی توان سخت افزار و پردازنده برای الگوریتم ها
- تنظیمات خط تولید



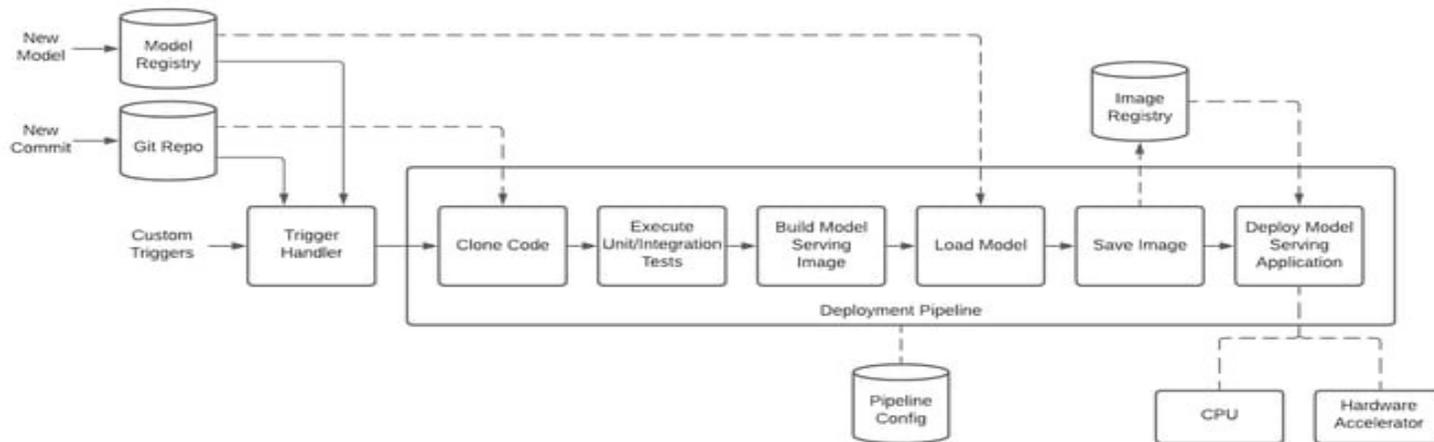
یادگیری و اعتبار سنجی (Model Training and Validation)

- فرایند اتوماتیک و تکرار شونده
- فعال شدن در زمان تغییر کد ها , رفع باگ , تغییر کتابخانه های نرم افزاری



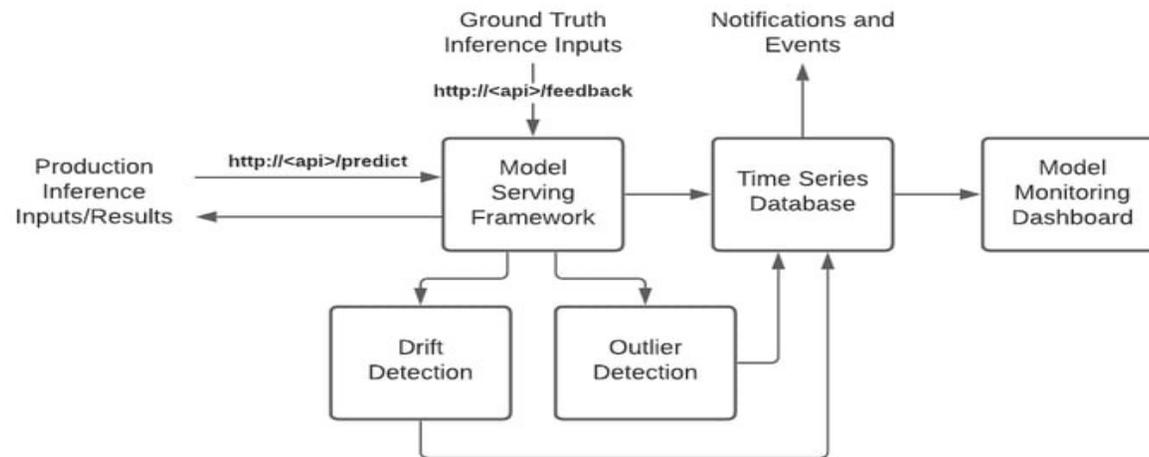
Model Serving

- معماری مبتنی بر میکروسرویس
- خروجی مبتنی بر داکر
- انتخاب بهترین مدل



مانیتورینگ محصول (Production Monitoring)

- نگهداری رفتار سیستم
- Drift detection
- Outlier detection
- سیستم هشدار



پروفایل ریسک در استقرار

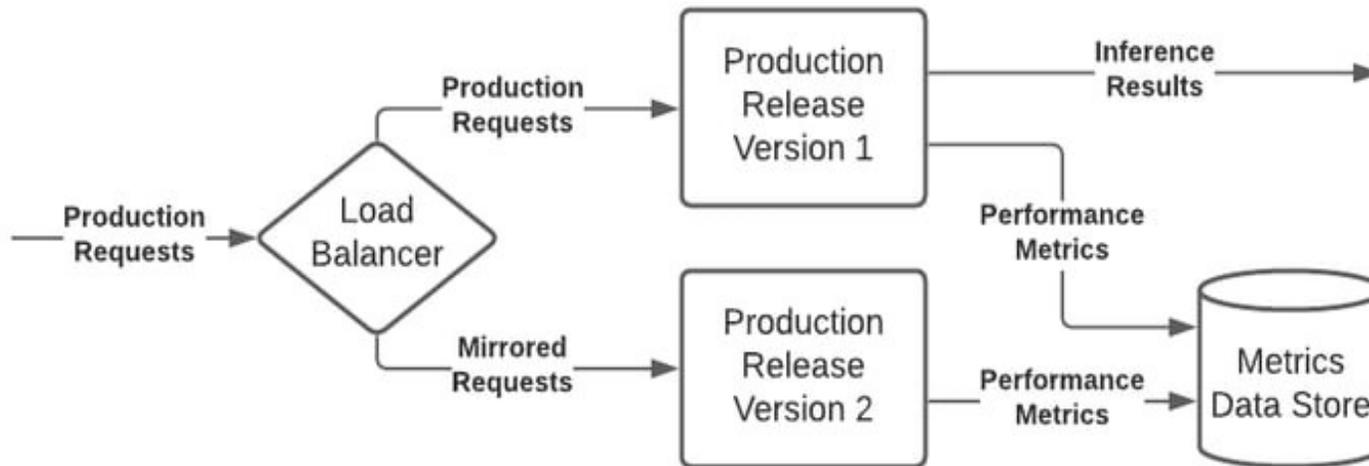
سطح ریسک	شرح	نوع تغییر
کم	تغییرات کد کم در سطح رفع باگ	Model Serving Microservice - Bug Fixes
کم	تغییرات کتابخانه های نرم افزاری	Model Serving Microservice - Library updates
کم	تغییرات base image در کانتاینر	Model Serving Microservice - Base Image Update
متوسط	اضافه کردن ویژگی های جدید مثل متریک های جدید	Model Serving Microservice - Feature Enhancements
متوسط	تغییرات API مورد استفاده سایر برنامه های کاربردی	Model Serving Framework - API Change
کم	تغییرات جزئی در داده ها	Model - Minor Training Dataset Change
متوسط	تغییرات مهم در داده ها مانند نوسازی داده ها , تغییر برجسب ها و یا تغییر بخش مهمی از داده ها	Model - Major Training Dataset Change
متوسط	تغییرات در هایپر پارامتر ها مثل نرخ یادگیری , دقت مدل و over/under fit	Model - Hyperparameter Tuning
زیاد	تغییرات مدل ها مثل سوئیچ از الگوریتم Gradient random forest به الگوریتم boosting و یا تغییر لایه های نرم افزاری	Model - Architecture or Model Type Change

{عنوان وبینار}
{webinar title}

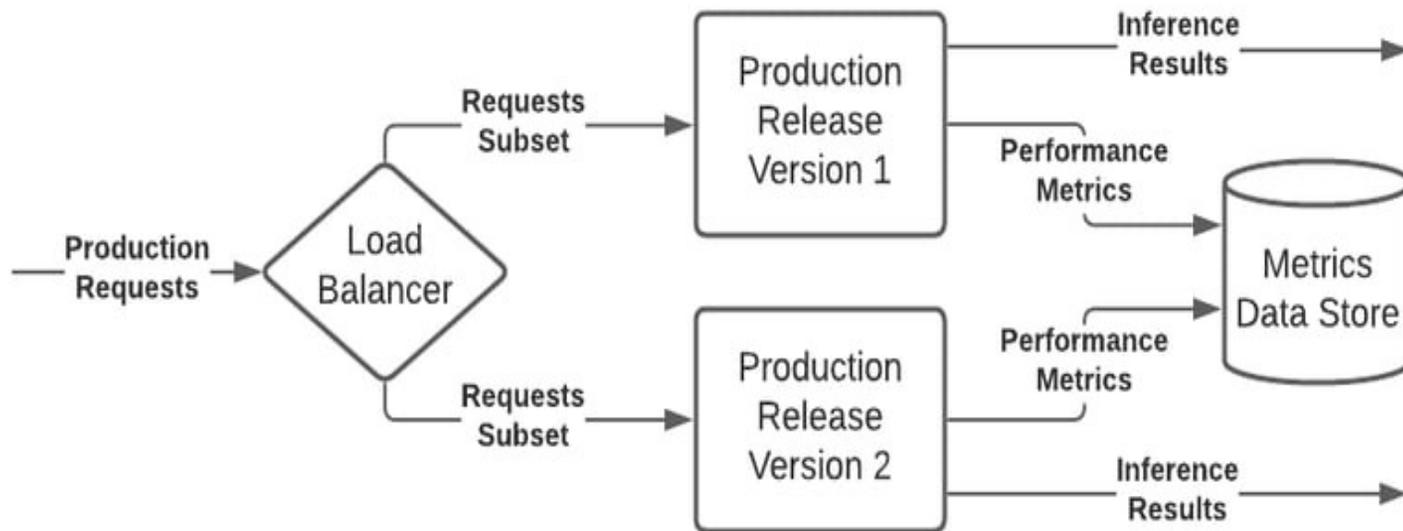
- Shadow Testing •
- A/B Test Pattern •
- Canary Test Pattern •

Shadow Testing

- ریسک گریزترین حالت تست
- تاخیر در اجرا



- انتخاب سرور ها به صورت تصادفی
- تاخیر در اجرا

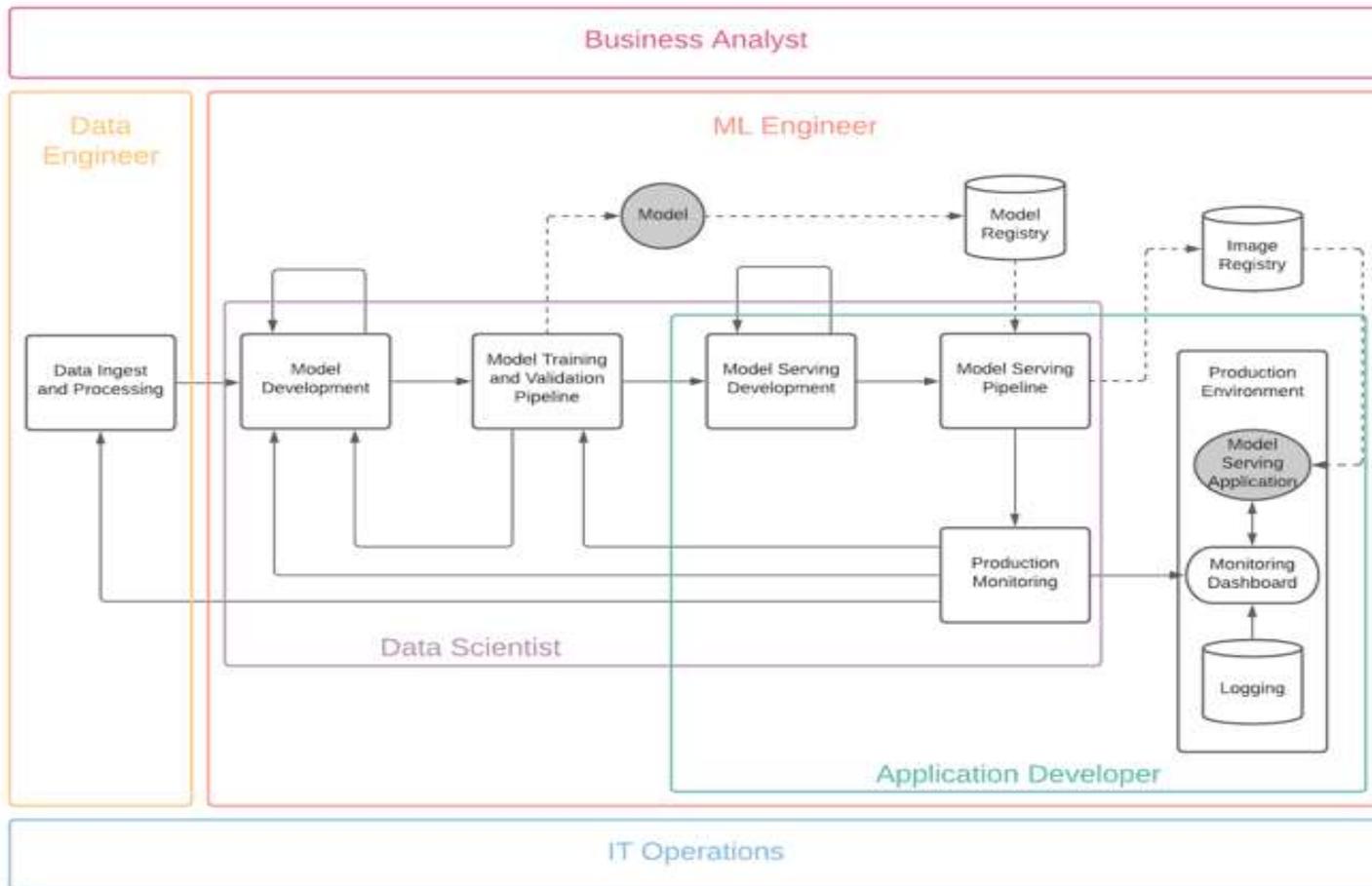


Canary Testing



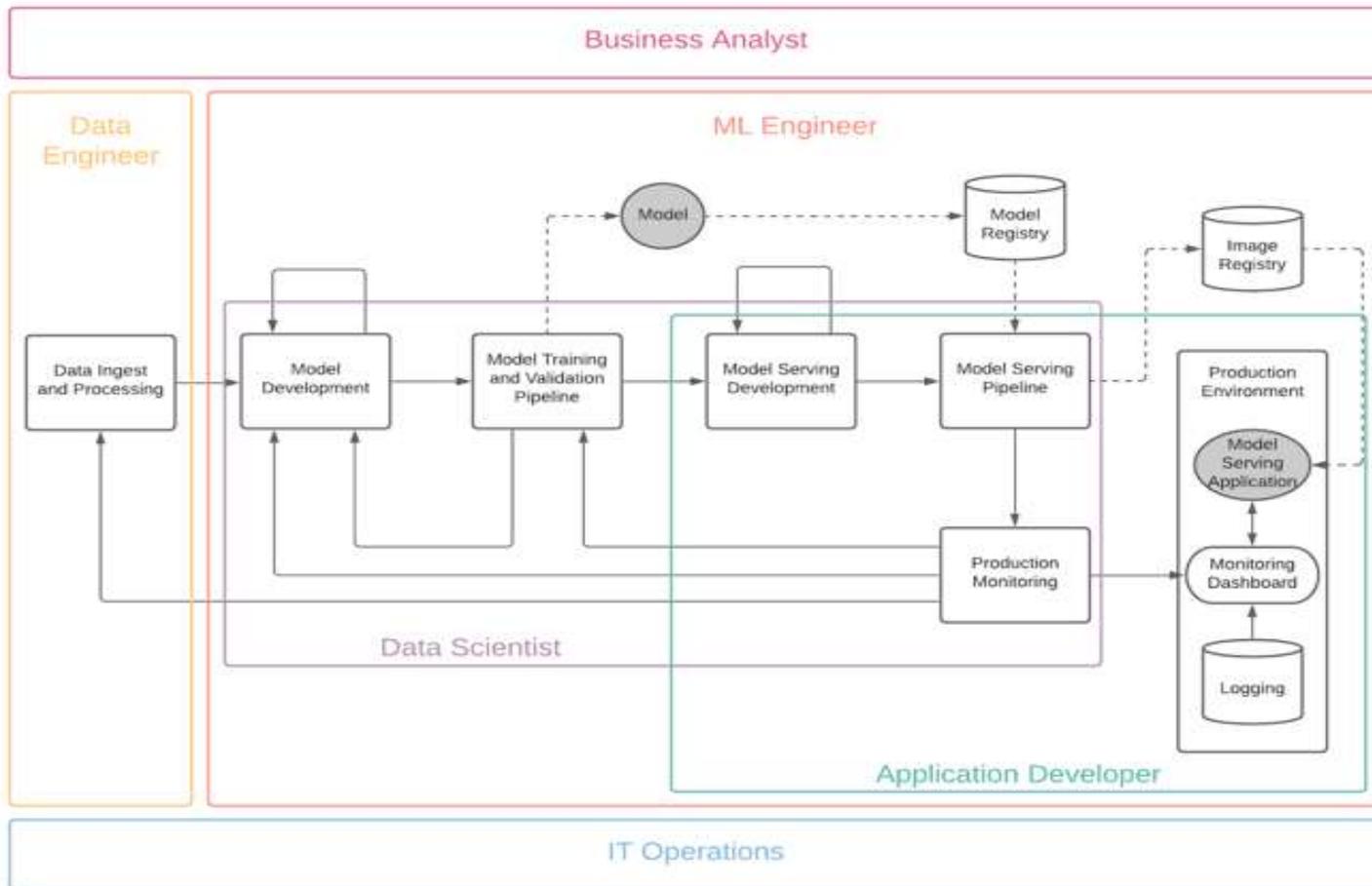
{عنوان وینار}
{webinar title}

ساختار سازمانی مبتنی بر ML/Ops



{عنوان وبینار}
{webinar title}

ساختار سازمانی مبتنی بر ML/Ops



{عنوان وبینار}
{webinar title}

Business Analyst

- تعریف یوزکیس ها و نیازمندی های بیزینسی
- پیگیری ROI
- اطمینان از پوشش نیازمندی بعد از پیاده سازی



{عنوان وبینار}
{webinar title}

Data Engineer

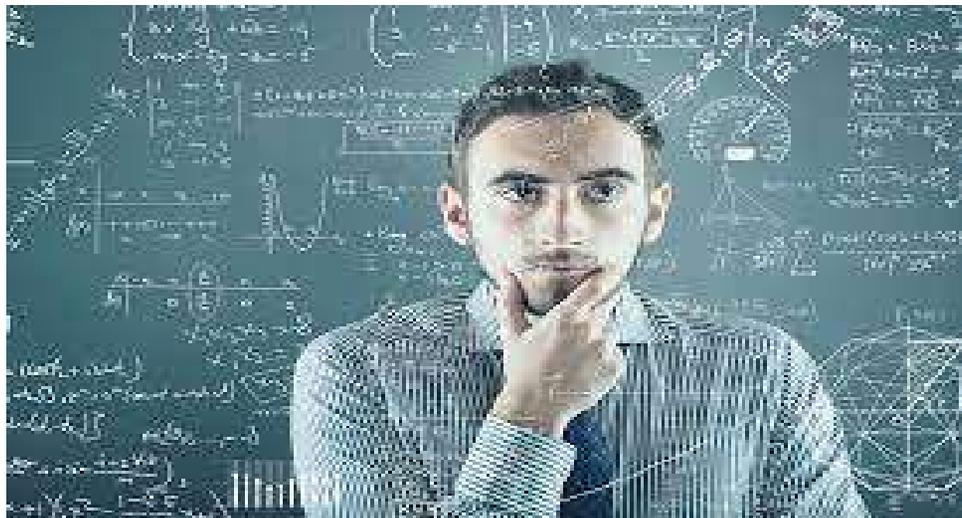
- دریافت داده خام و آماده سازی (clean,prepare,organize)
- مجوز دریافت و در صورت نیاز تغییر داده های حساس و قابل فهم
- ارائه داده در فرمت استاندارد سازمان
- ایجاد ساختار های توسعه و معماری و ابزاری برای داده ها



{عنوان وبینار}
{webinar title}

Data Scientist

- فهم معنایی داده ها
- فرایند تکرار پذیر برای انتخاب الگوریتم مناسب
- ایجاد مدل نهایی یادگیری ماشینی



{عنوان وبینار}
{webinar title}

Application Developer

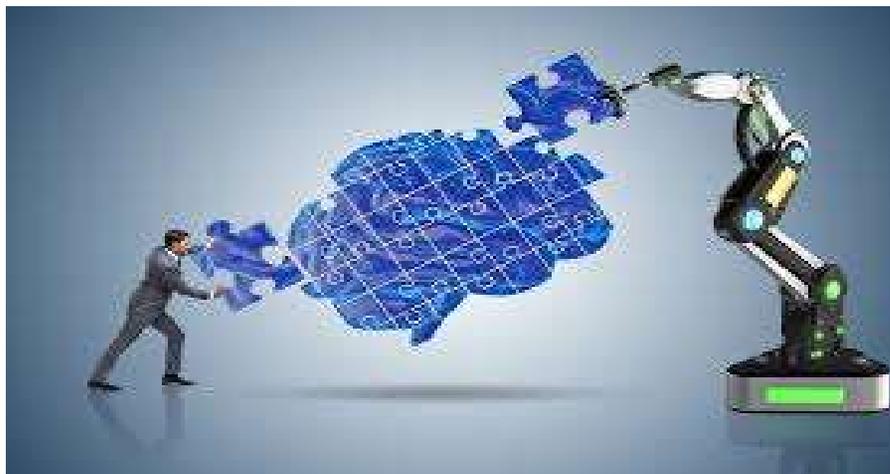
- ایجاد میکروسرویس برای مدل یادگیری
- تمرکز روی تحویل نهایی محصول



{عنوان وبینار}
{webinar title}

ML Engineer

- نظارت و کنترل و ساختن خط تولید مبتنی بر ML



IT Operations

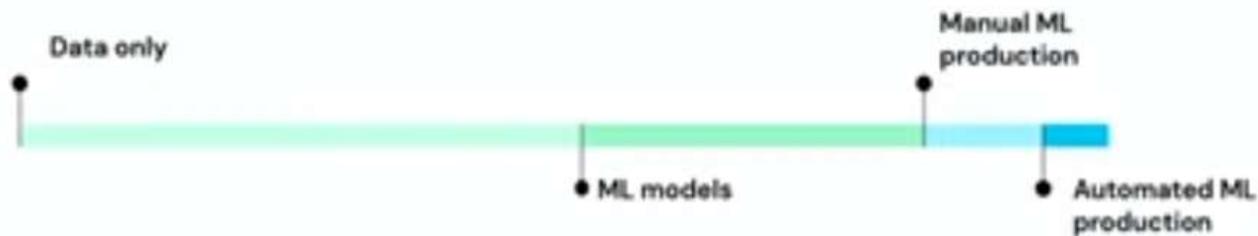
- نصب سیستم ها
- مانیتورینگ سیستم ها
- نگهداری از سیستم ها
- تصمیم برای توپولوژی استقرار



{عنوان وبینار}
{webinar title}

ML maturity levels

Almost every company wants to be a Data+AI company, but the majority are new to machine learning (ML).



ابزارهای مطرح این حوزه

MLOps tools	strengths	weakness
Kubeflow	Seamless integration with Kubernetes	Requires Kubernetes expertise
MLflow	Easy setup and usage	May lack some advanced features
TensorBoard	Powerful visualization for TensorFlow models	Focused on visualization; may lack other MLOps features
DVC	Strong data versioning and pipeline orchestration	Limited MLOps-specific features
Apache Airflow	Excellent for complex workflow orchestration	Additional components needed for full MLOps support
TFX	Scalable model training and serving with TensorFlow	Tightly integrated with TensorFlow
Metaflow	Easy to use and collaborate on data science projects	May lack extensive enterprise-grade features

{عنوان وبینار}
{webinar title}

ابزارهای مطرح این حوزه

MLOps tools	strengths	weakness
Kubeflow	Seamless integration with Kubernetes	Requires Kubernetes expertise
MLflow	Easy setup and usage	May lack some advanced features
TensorBoard	Powerful visualization for TensorFlow models	Focused on visualization; may lack other MLOps features
DVC	Strong data versioning and pipeline orchestration	Limited MLOps-specific features
Apache Airflow	Excellent for complex workflow orchestration	Additional components needed for full MLOps support
TFX	Scalable model training and serving with TensorFlow	Tightly integrated with TensorFlow
Metaflow	Easy to use and collaborate on data science projects	May lack extensive enterprise-grade features

{عنوان وبینار}
{webinar title}

MLflow

- پروژه متن باز
- نگهداری تجارب (experiments) الگوریتم های مختلف روز داده ها به منظور مقایسه ان ها
- نگهداری کد های ML به منظور بسته بندی و ارسال نسخه و همینطور استفاده مجدد
- مدیریت و دیپلوی مدل های مختلف ML
- مدیریت چرخه تولید و عملیات در یادگیری ماشین

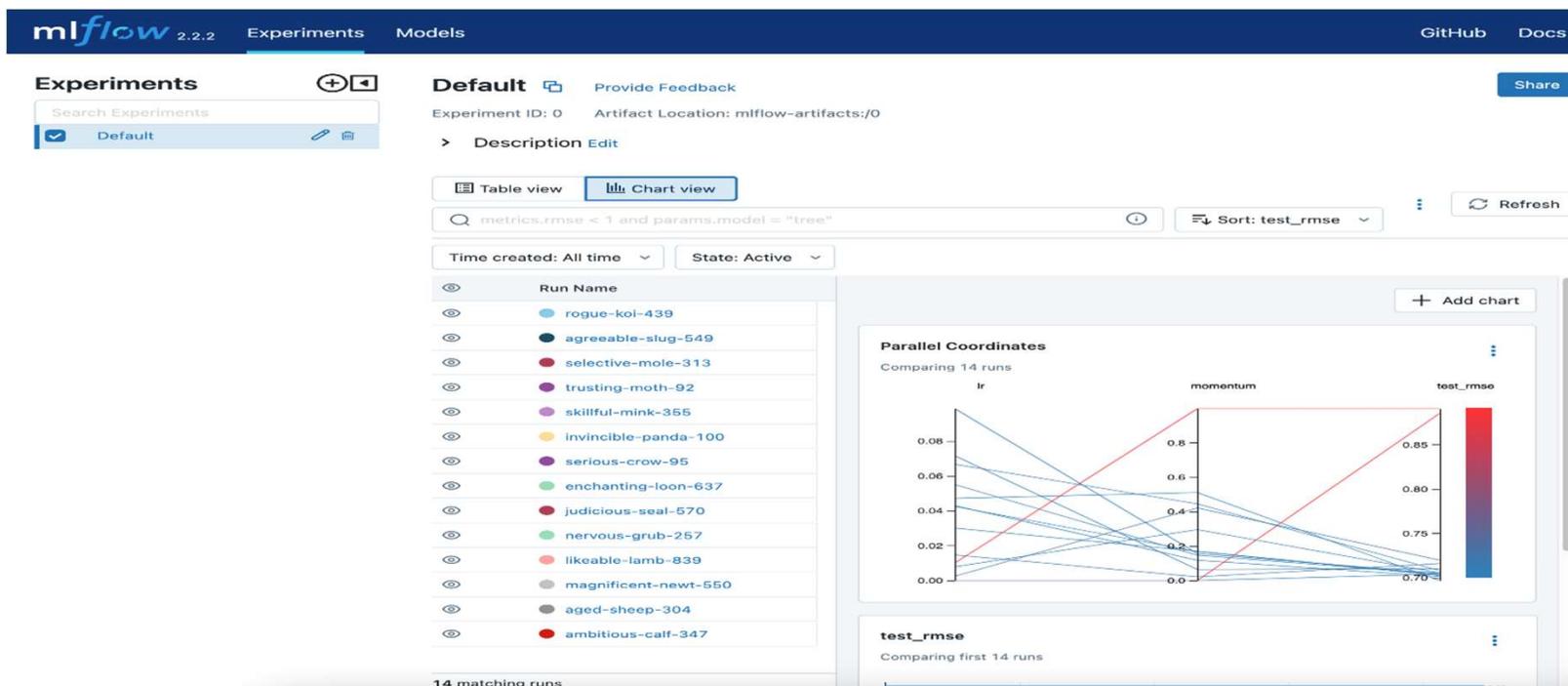
معماری سازمانی در MLflow



{عنوان وبینار}
{webinar title}

MIFlow Tracking

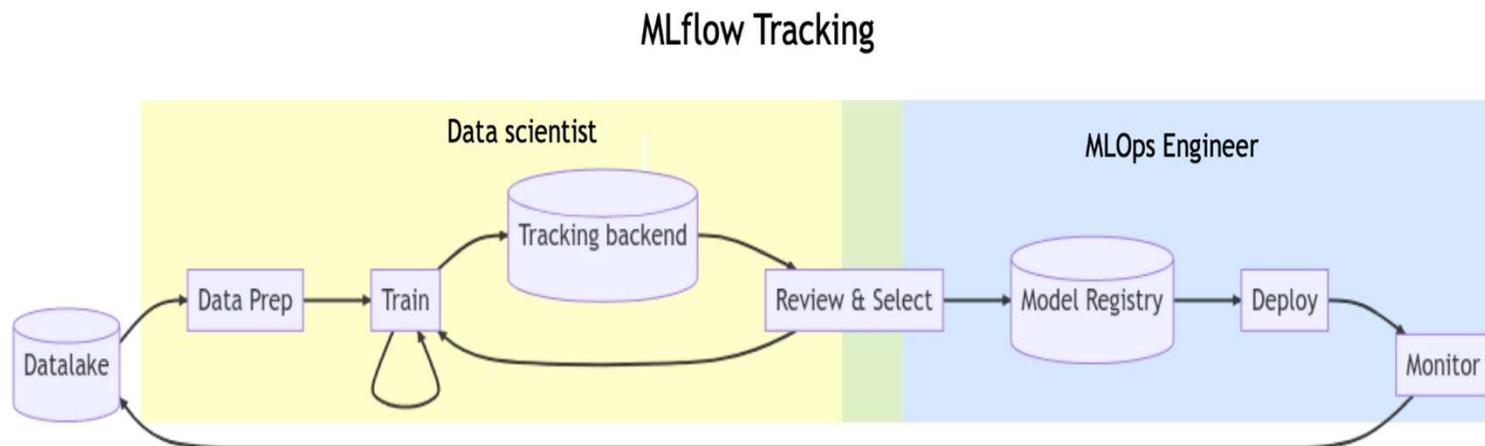
- لاگ و ثبت پارمترها
- نگهداری ورژن کد ها
- ثبت متریک ها
- داشبرد مدیریتی برای مشاهده وقایع



{عنوان وبینار}
{webinar title}

{file_name}

MLflow tracking



{عنوان وینار}
{webinar title}

MLFlow Models

- مجموعه ای از ابزار ها برای یادگیری و استقرار مدل
- استفاده از ابزار هایی مثل داکر برای استقرار مدل
- مقایسه پارامتر های مدل برای انتخاب مدل بهینه

The screenshot displays the MLflow web interface for a model run named 'ambitious-calf-347'. The interface includes a navigation bar with 'Experiments' and 'Models' tabs. The main content area shows the run details, including the Run ID, Date, Source, Duration, and Parent Run. A table lists the metrics for the run, showing 'test_loss' as 0.697, 'train_loss' as 0.479, and 'val_loss' as 0.5. The 'Artifacts' section is expanded, showing a tree view of the model artifacts, including 'data', 'MLmodel', 'conda.yaml', 'python_env.yaml', and 'requirements.txt'. The 'MLflow Model' section provides instructions on how to use the model, including a 'Model schema' table and 'Make Predictions' code snippets for Spark and Pandas DataFrames.

Name	Value
test_loss	0.697
train_loss	0.479
val_loss	0.5

Name	Type
No schema. See MLflow docs for how to include input and output schemas with your model.	

```
import mlflow
from pyspark.sql.functions import struct, col
logged_model = 'runs:/bc19d004f1e24ae28e34f5ac24e96578/model'

# Load model as a Spark UDF. Override result_type if the model does not return double values.
loaded_model = mlflow.pyfunc.spark_udf(spark, logged_model, result_type='double')

# Predict on a Spark DataFrame.
df.withColumn('predictions', loaded_model(struct(*map(col, df.columns))))

# Predict on a Pandas DataFrame:
import mlflow
logged_model = 'runs:/bc19d004f1e24ae28e34f5ac24e96578/model'

# Load model as a PyFuncModel.
loaded_model = mlflow.pyfunc.load_model(logged_model)

# Predict on a Pandas DataFrame.
import pandas as pd
```

{عنوان وبینار}
{webinar title}

به گروه تخصصی معماری سازمانی انجمن انفورماتیک ایران بپیوندید...



www.isi-ea.ir

{عنوان وبینار}
{webinar title}