# Machine Learning Lifecycle - MLOps

Date: Jan 26th,2021

Subroto Mukherjee
Head of Digital Innovation & Emerging Tech Americas, Consumer Healthcare GSK
subroto.x.mukherjee@gsk.com

**Pam**
Respiratory manufacturing
North Carolina, USA

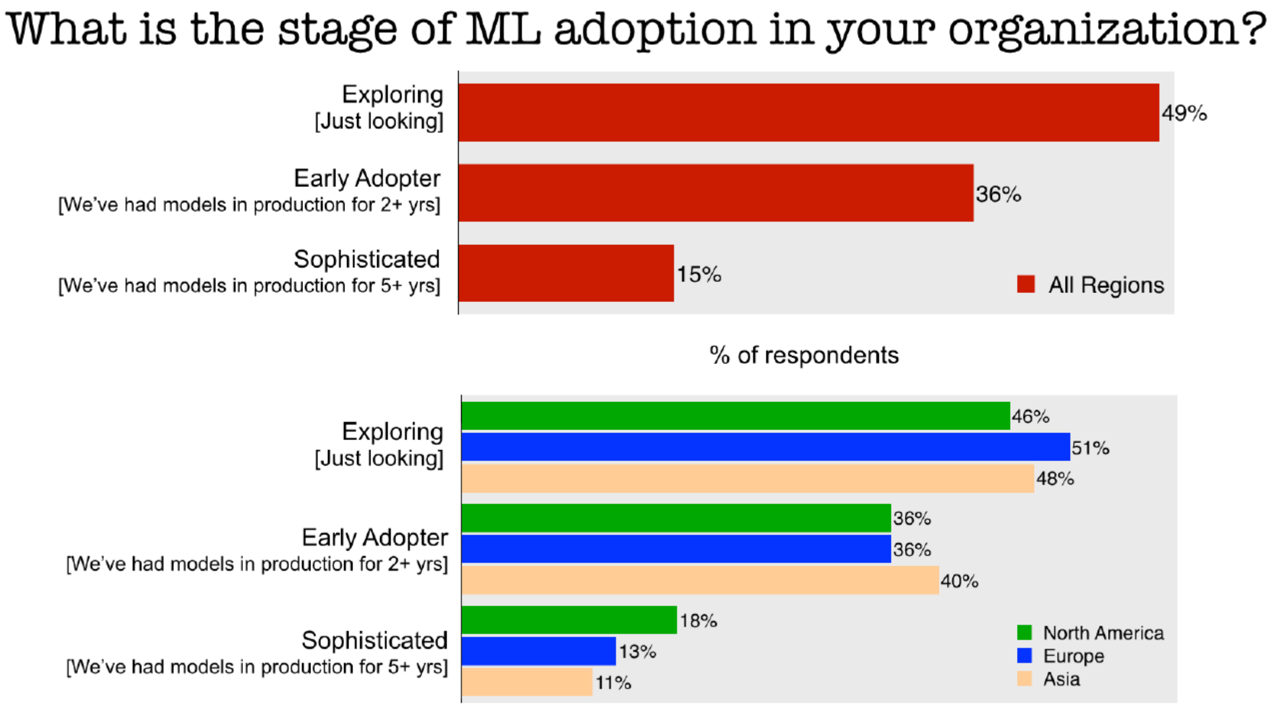
# A science-led global healthcare company with a special purpose: to help people do more, feel better, live longer.

We have 3 global businesses that research, develop and manufacture innovative pharmaceutical medicines, vaccines and **consumer healthcare products**.

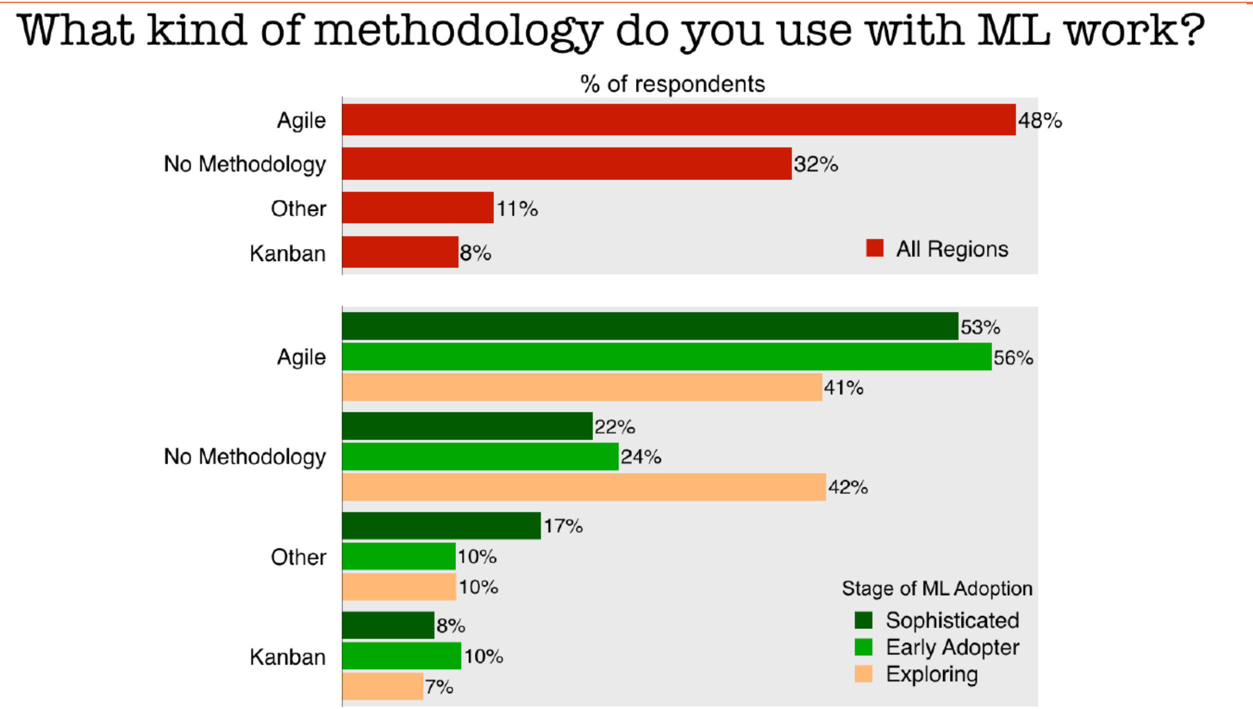
# Machine Learning(ML) Lifecycle – MLOps Outline

- Current State of Machine Learning Journey – Industry Forecast and ML Journey
- Bottlenecks and Challenges Facing Machine Learning Industry
- Different Actors , Players in the Machine Learning Journey.
- Word cloud --- MLOps, Devops, Dataops , AIOps ,Model Ops...... Lots of Ops ...
- Need For MLOps ,MLOps vs DevOps
- Use Cases - ML Use Cases
- MLOps Framework - ML Flow , Kubeflow, Air Flow
- MLOps RoadMap

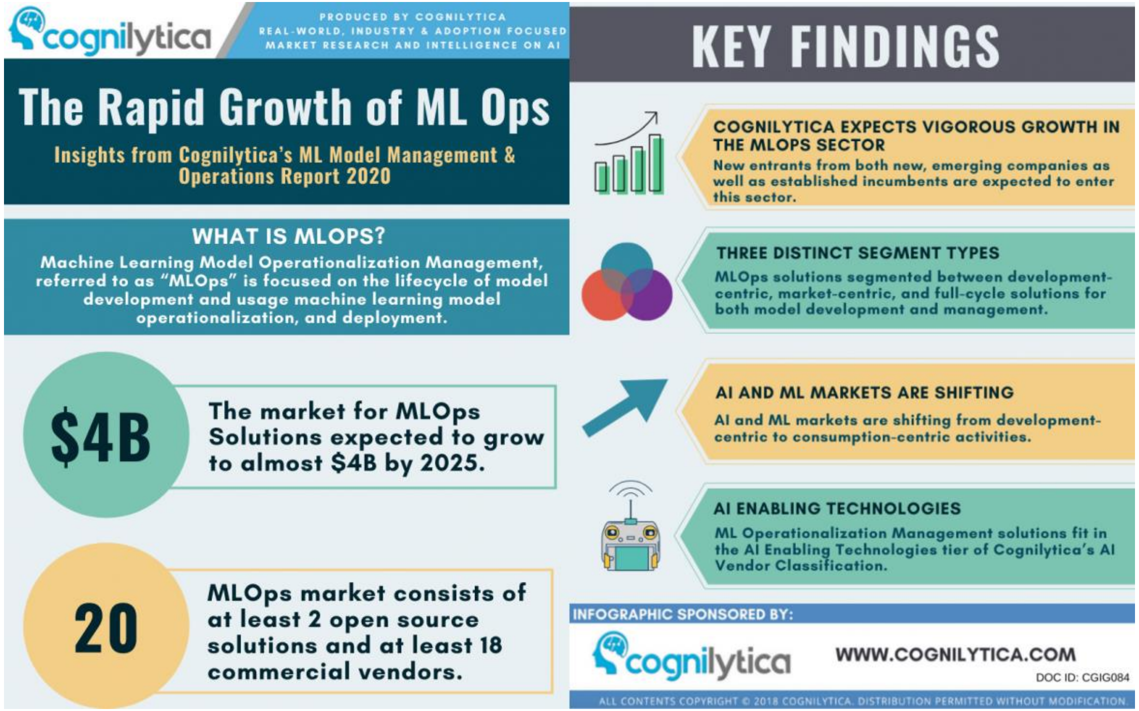
# Current State of Machine Learning Journey – Industry Forecast and ML Journey



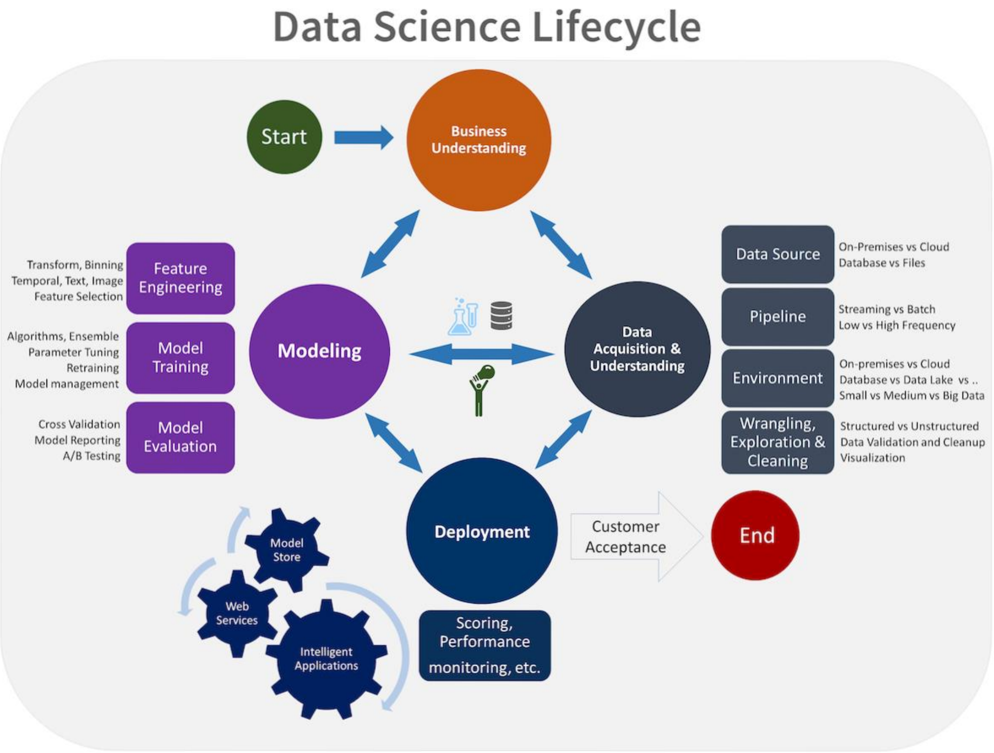
# Current State of Machine Learning Journey – Industry Forecast and ML Journey



# Current State of Machine Learning Journey – Industry Forecast and ML Journey



# Data Science Machine Learning Lifecycle




[MLOps: The end of end-to-end. Machine Intelligence is one of our main… | by Mosaic Ventures | Medium]

## Challenges Facing the Machine Learning Solutions

- The Black Box Problem
  - This inability to understand how machines are thinking plays into cultural and societal fears around artificial intelligence
  - People don’t trust machines that think, and act like them.
- Limited Talent Pool
- Data Availability and Privacy Concerns
- Complexity and Limitations

The State of Machine Learning in 2020 (sololearn.com)

## Challenges Facing the Machine Learning Solutions

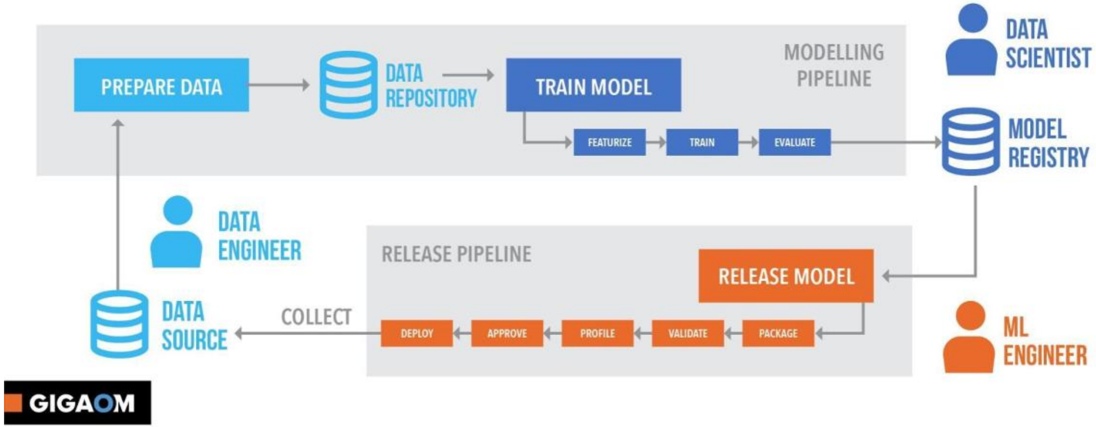
According to a Gartner Analyst-

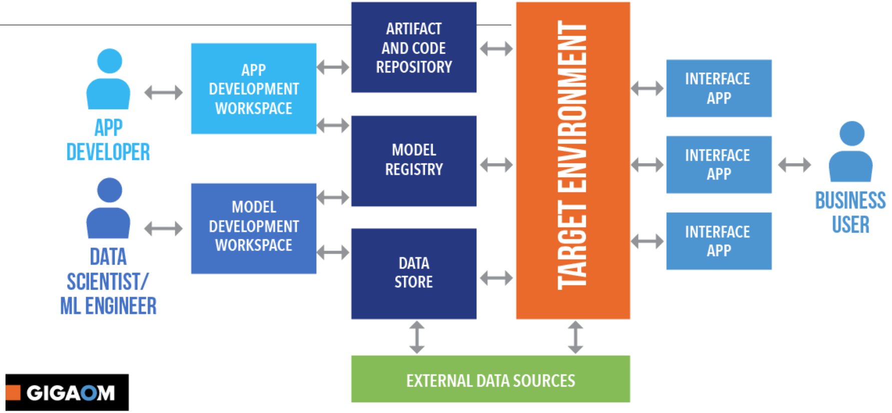
- Out of all ML Models – only 47% of those models go into production.
- 88% of AI initiatives in the corporate sector struggle to move past the test stage.

Getting Machine Learning into Production: MLOps - InformationWeek

# Introducing MLOps – Word Cloud
## MLOps, DevOps, AIOps, ModelOps, Data, Governance, Compliance

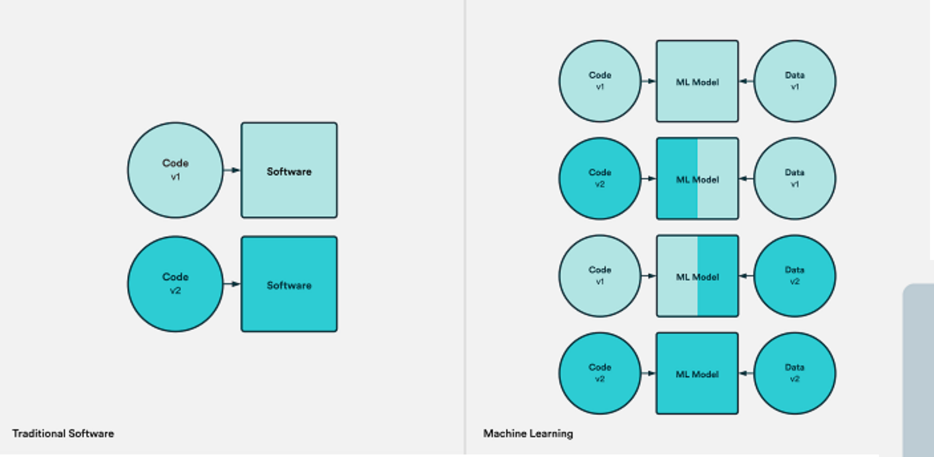
# Introducing MLOps

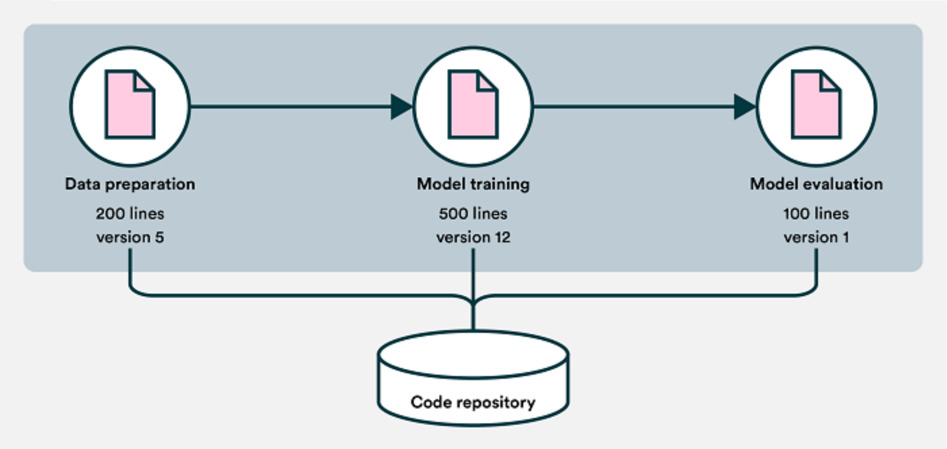



[Delivering on the Vision of MLOps (gigaom.com)]

# Introducing MLOps (Summary)

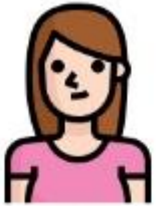



[MLOps Brings Best Practices to Developing Machine Learning - insideBIGDATA]()

# Is That All ?

## Cowboys and Ranchers Can Be Friends!



**Data Scientist**

- Quick iteration
- Frameworks they understand
- Best of breed tools
- No management headaches
- Unlimited scale


**SRE/ML Engineers**

- Reuse of tooling and platforms
- Corporate compliance
- Observability
- Uptime

## ML Model - Problems

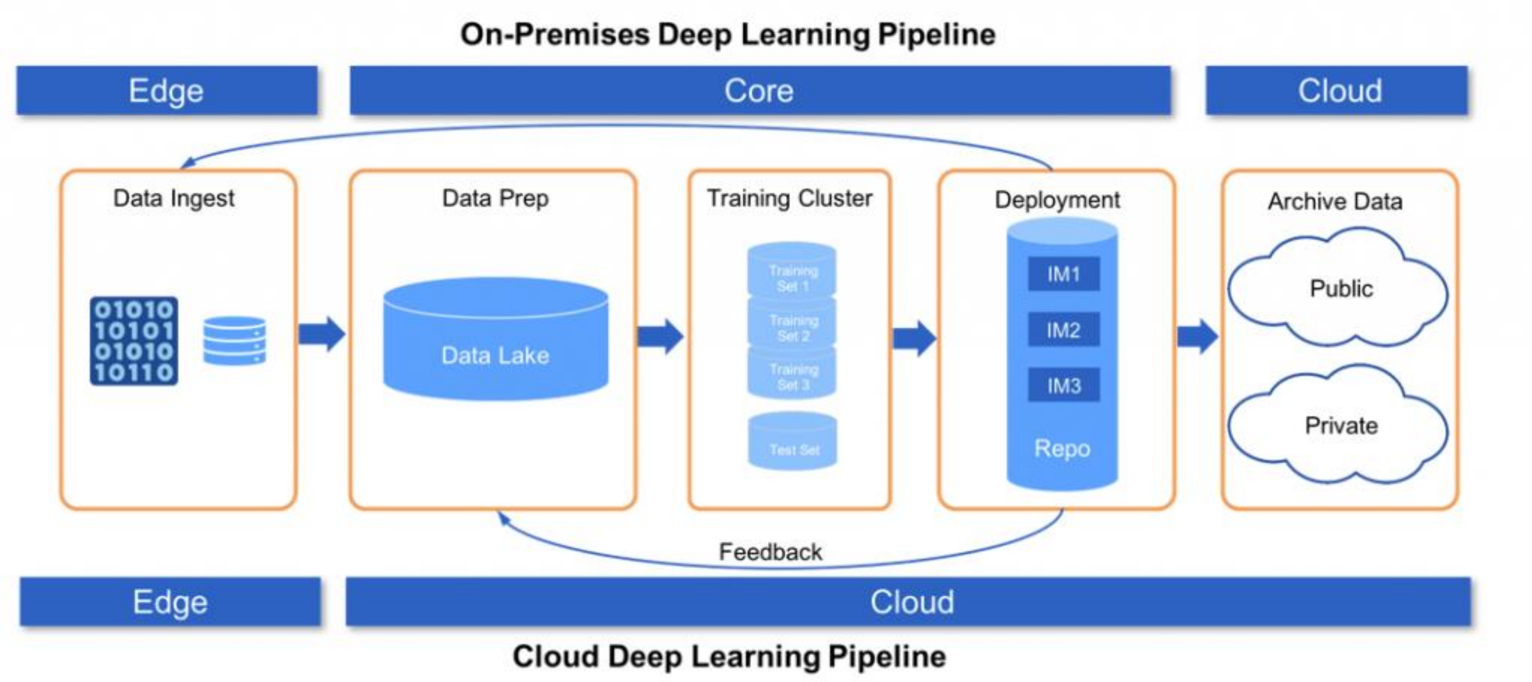
Every ML model faces a host of challenges during the four core stages of its lifecycle:

- ETL (Data pipelines)
- Algorithm training
- Inference
- Monitoring, Management, and Updates

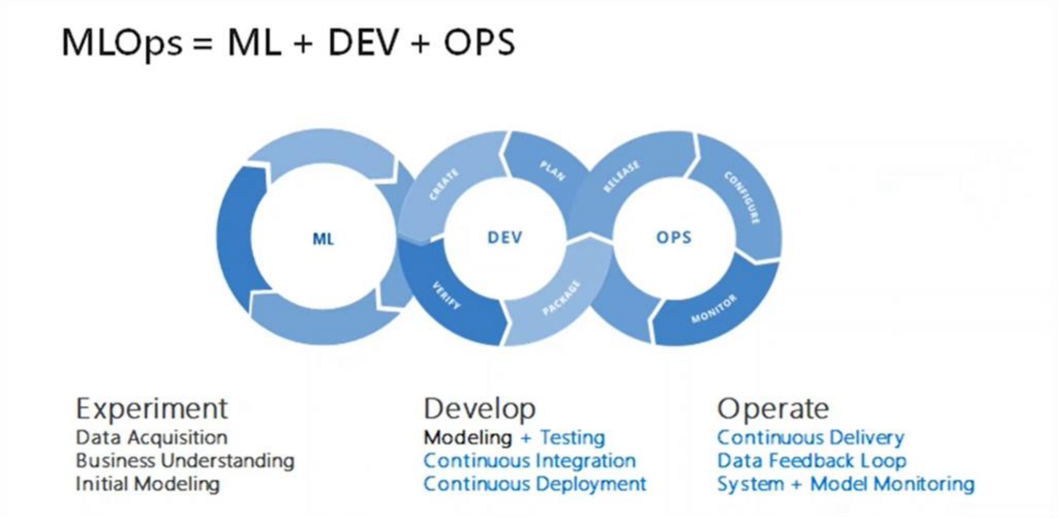
https://blogs.nvidia.com/blog/2020/09/03/what-is-mlops/

# ML Model – Data Pipelines

## Neural Network,CNN, XGBoost,Principal Component Analysis



# Devops to  MLOps Journey



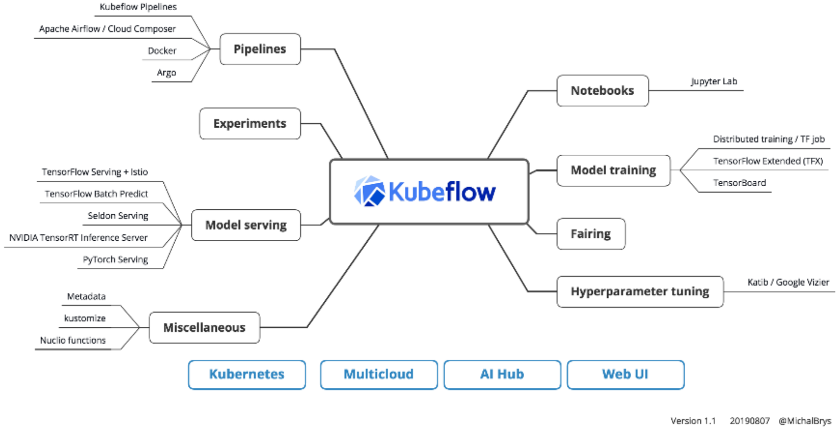

- **MLOps** is the Machine learning equivalent of DevOps. While DevOps helped optimize the production lifecycle of Big Data project
- **MLOps** seeks to solve the problems associated with the implementation of ML in production.

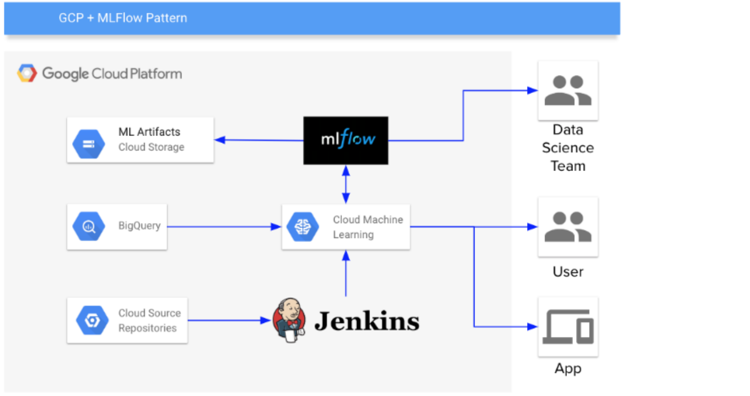
# MLOps vs Devops

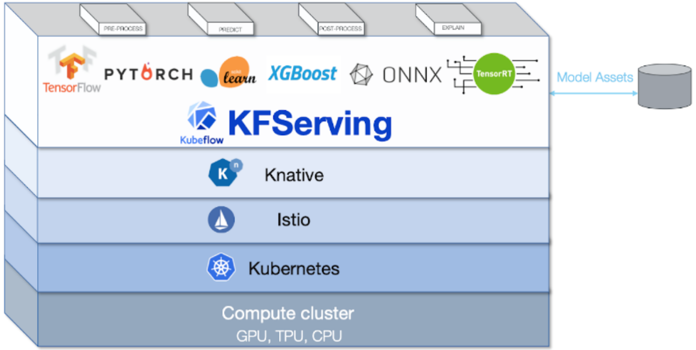
| Functionality | Devops | MLOps |
| --- | --- | --- |
| Stakeholders | Software developers(SW), Operations, QA | Data Scientist, Data Engineers, Software Developers, Operations, QA. |
| Automation(CI/CD), | Code by SW developers to Production system continuously. | 2 Streams – Data Engineering Code, Model Code are Deployed, Synced up and models are validated and deployed |
| Governance, Monitoring | Dependency Changes, Failures, Key Metrices. Ethics and Good Model not needed | Devops+ Model Performance+ Data Drift(Quality of data, distribution). Ethics and Explanability of models |
| Model Retraining | NA | Critical and required. |

Why MLOps shouldn’t be an afterthought | by Innovate & Disrupt | Medium

# MLOps Frameworks – Opensource (MLFlow , Kubeflow),Azure ML

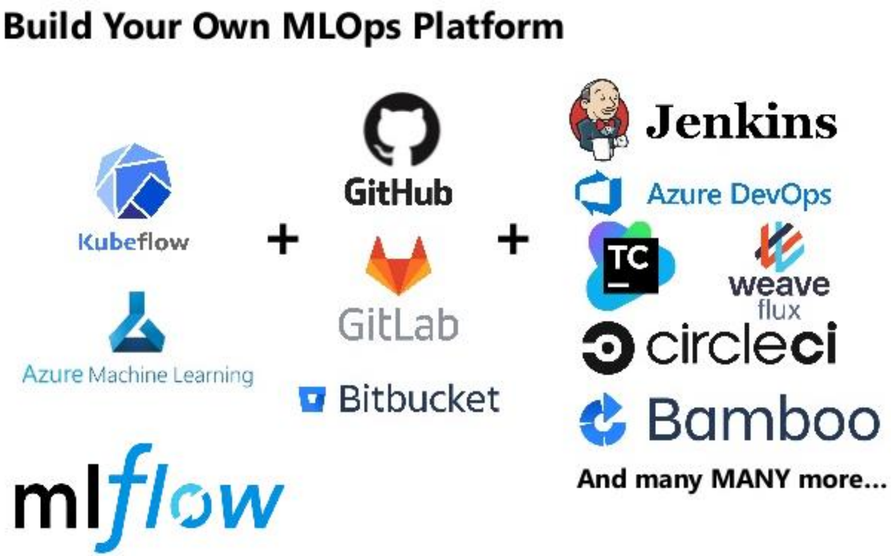



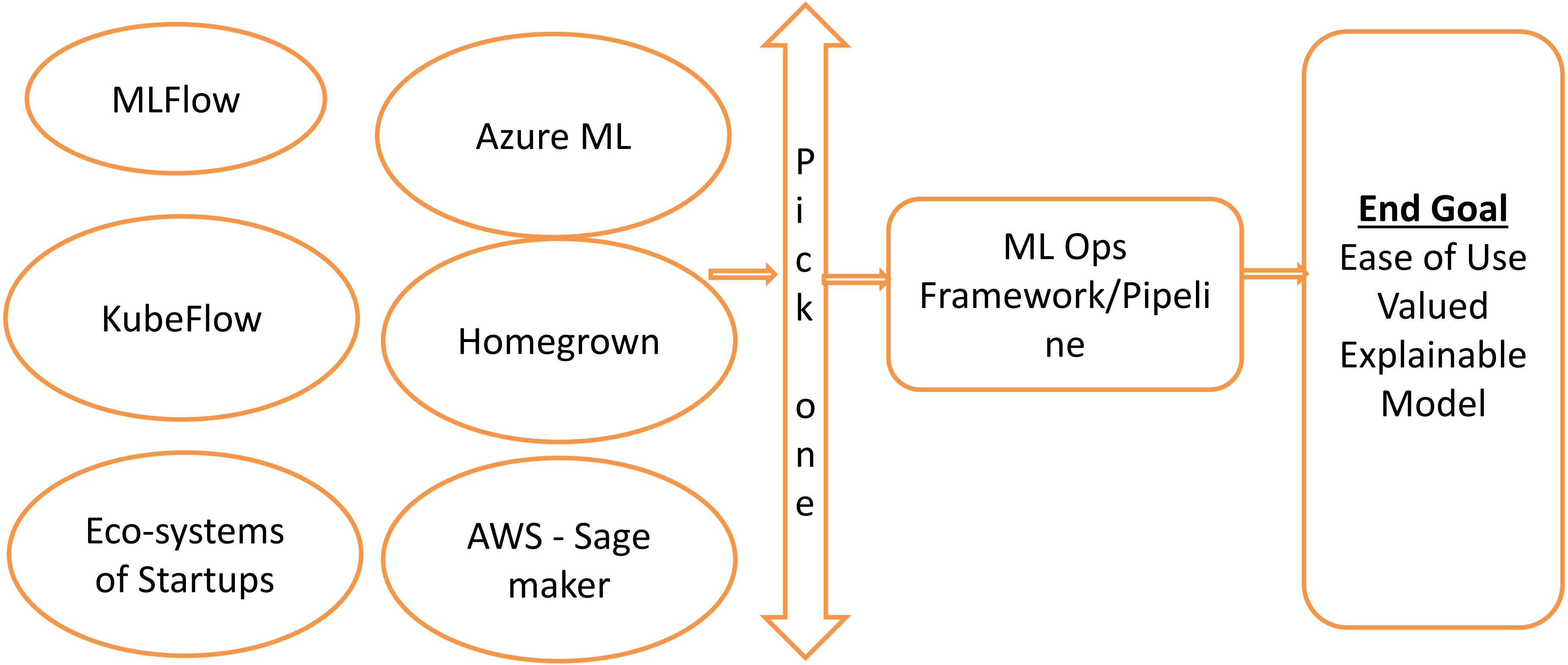




[The Cheesy Analogy of MLflow and Kubeflow | by Byron Allen | Servian | Medium]
[Kubeflow — a machine learning toolkit for Kubernetes | by Michal Brys | Medium]

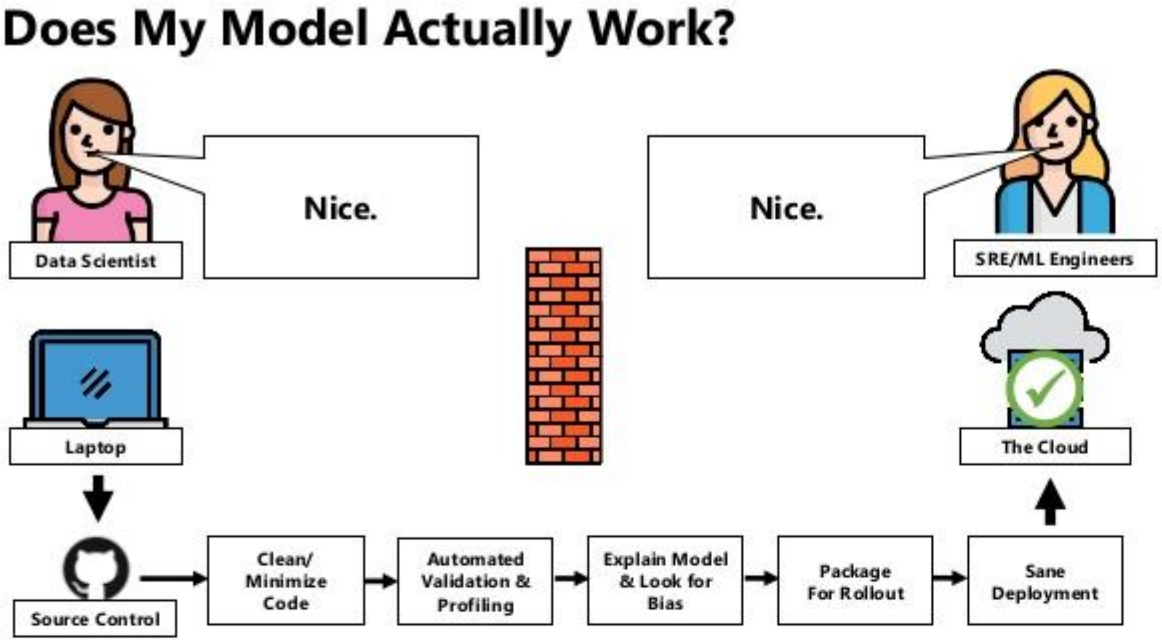
# Build your own



# MLOps Frameworks – Opensource (MLFlow , Kubeflow),Azure ML



# Model Works




https://www.slideshare.net/weaveworks/using-mlops-to-bring-ml-to-productionthe-promise-of-mlops

# ML Ops – Use Case 1 - Seasonal Prediction Value Add

**Problem Statement**

Traditional lead indicator for the Allergy season severity was the Allergy Activity Notification (AAN) which uses the combination of weather, pollen and medical allergy incidences.

**Vision**

- Imagine a world where we could predict seasonal illness and message you to be prepared and for vulnerable people to get prepared with flu or allergy medication.
- GSK working in the background with retailers for availability of OTC medications readily available .

# Machine Learning – Digital Innovation -Mission

Consumer Healthcare

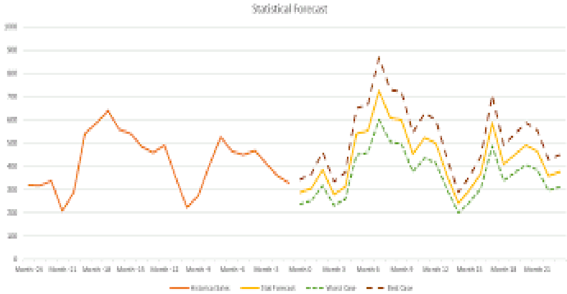
## Seasonality Forecasting Model

Primary Objectives

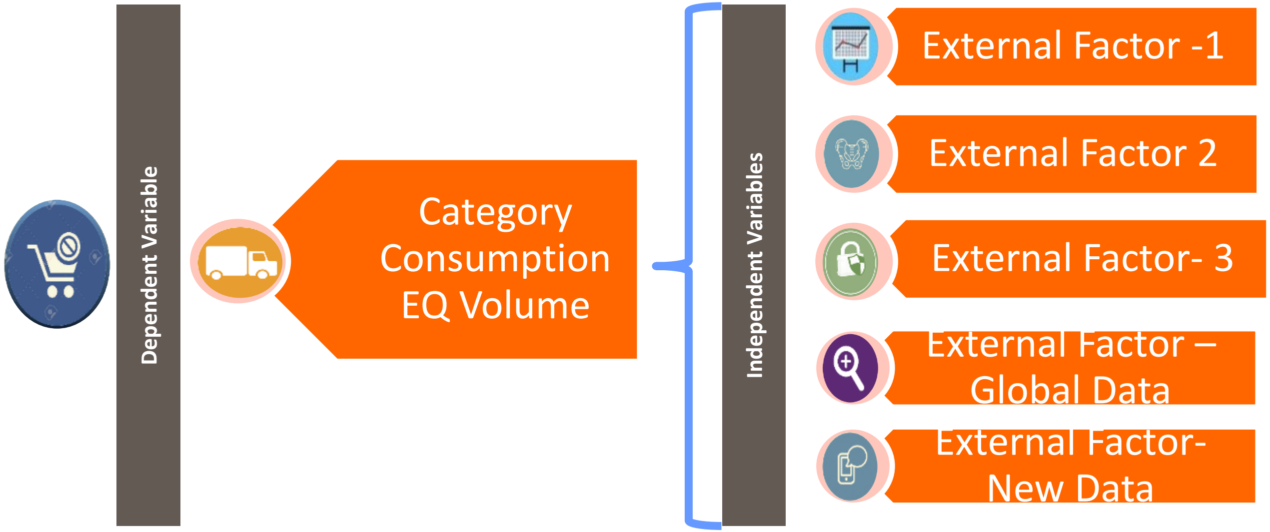
- Inform Consumers on our Brand.com
- Improve national and regional media delivery
- Enable GSK as a Category leader in seasonal category, to inform retailers of timing for seasonal activation (distribution, stock up, display and secondary support)

Secondary Objective

- Inform internal GSK functions such as Supply Chain, Finance and Leadership.



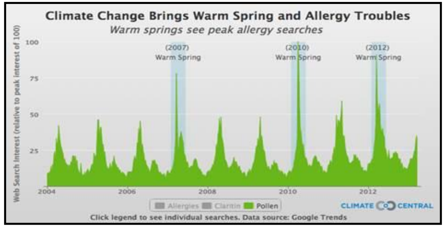
# Predictive Model Construct – General Overview

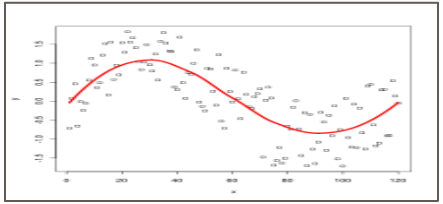


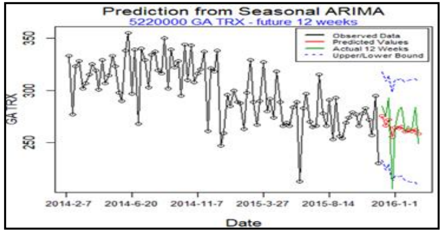

Our attempt will be to a predictive took for the Allergy Seasons.

The model will be broken down by the syndicated major U.S Regional Market Levels.

## Machine Learning: Build an allergy model and have it adjusted itself as new data comes in






### Step 1

Use Historical data to align on model form & structure. Following data sets will be considered-

- Internal Data Sets
- External Data Sets

### Step 2

The <u>supervised machine learning</u> model runs on a weekly fashion and tells where brand is at current state as well as information around when expected peak and level of peak. The model should also predict what is expected over the next few weeks.

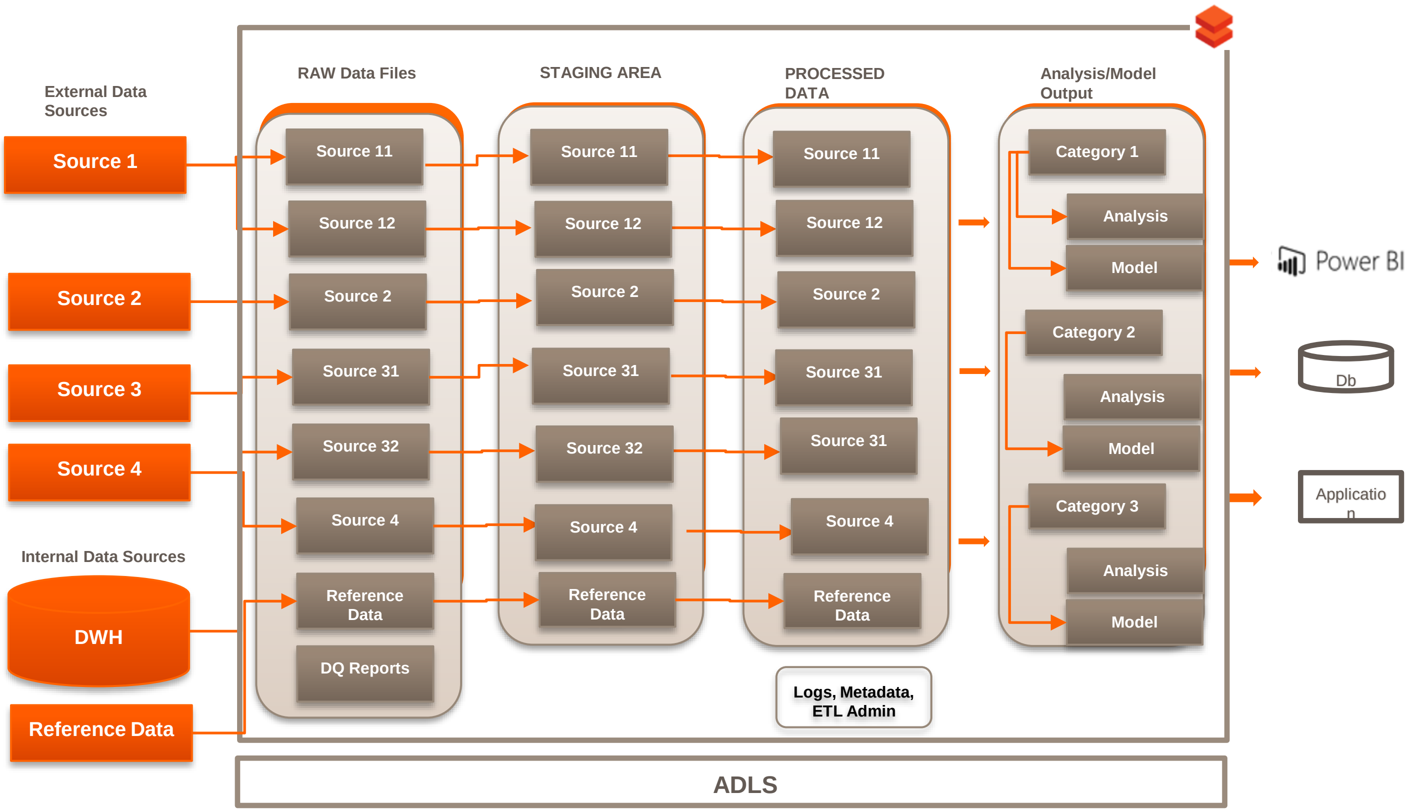
### Step 3

Once the weekly expectations are completed, the model adjusts, so that the weekly data points in step 2 become facts or inputs to the next prediction

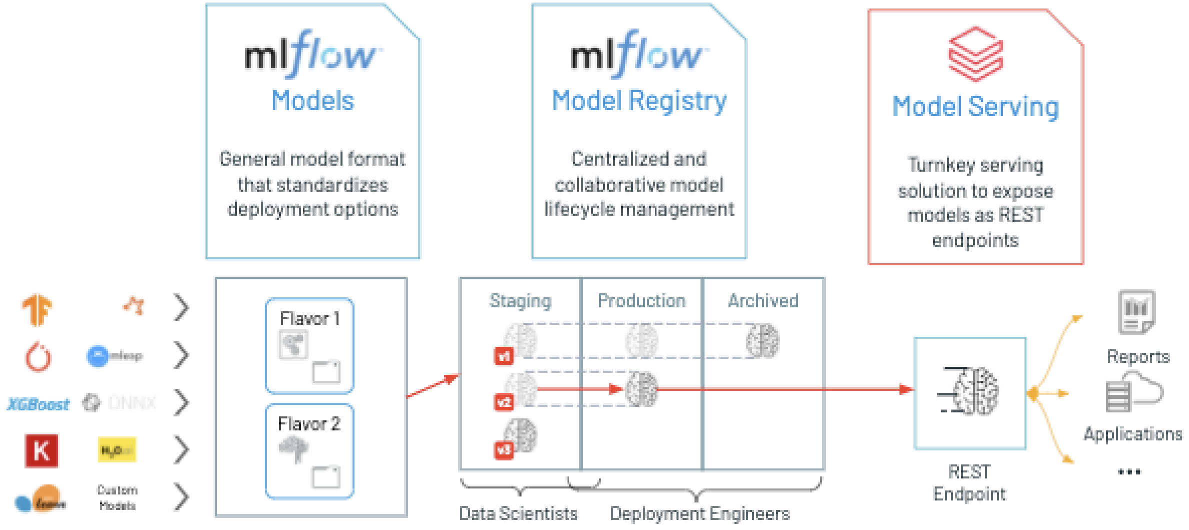
# Data Processing on Databricks

## Data Pipeline

gsk
Consumer Healthcare



# Model Works  ML FLow

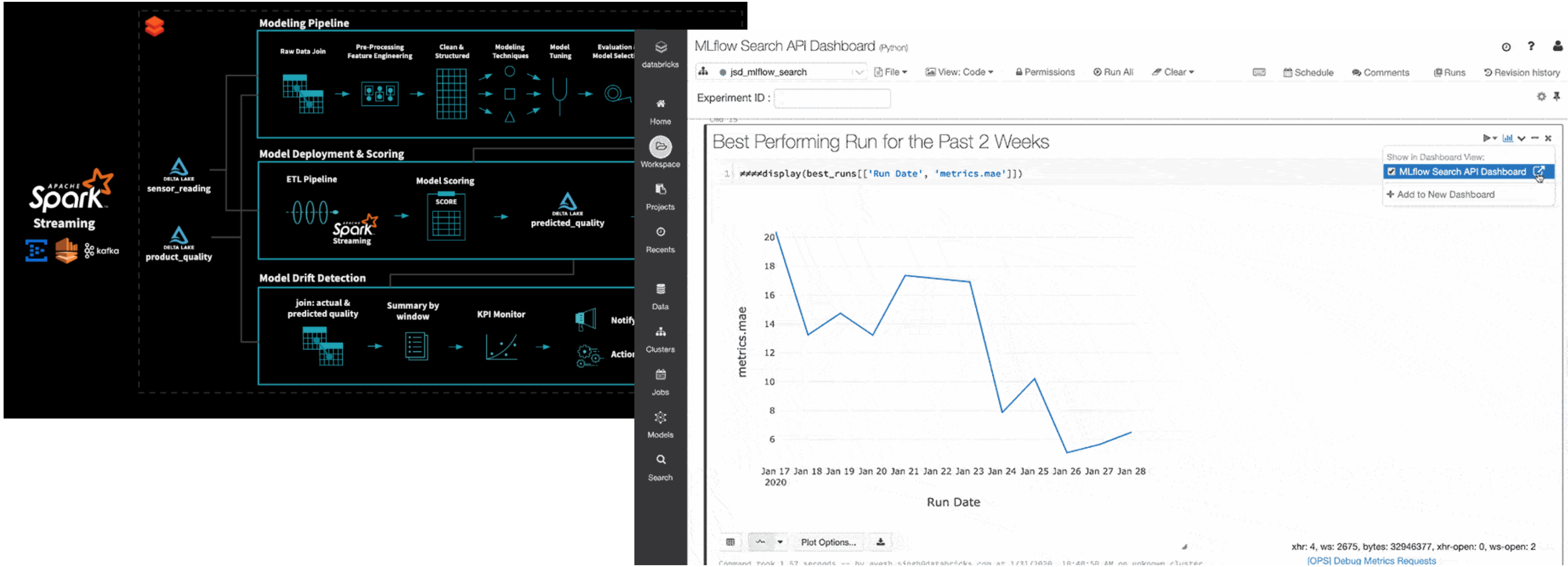



[MLflow Model Serving Intro and Overview of How It Enables Real-time Machine Learning - The Databricks Blog]

# MLOps

## Productionizing Machine Learning: From Deployment to Drift Detection – ML Flow






https://github.com/joelcthomas/modeldrift
https://databricks.com/blog/2020/02/18/how-to-display-model-metrics-in-dashboards-using-the-mlflow-search-api.html

# MLOps – Seasonal Prediction – key Takeaway

## Key Takeaway

Consumer Healthcare

- **MLOps** helped in managing the
  - Model Pipeline.
  - Model Scoring
  - Model Drift Detection

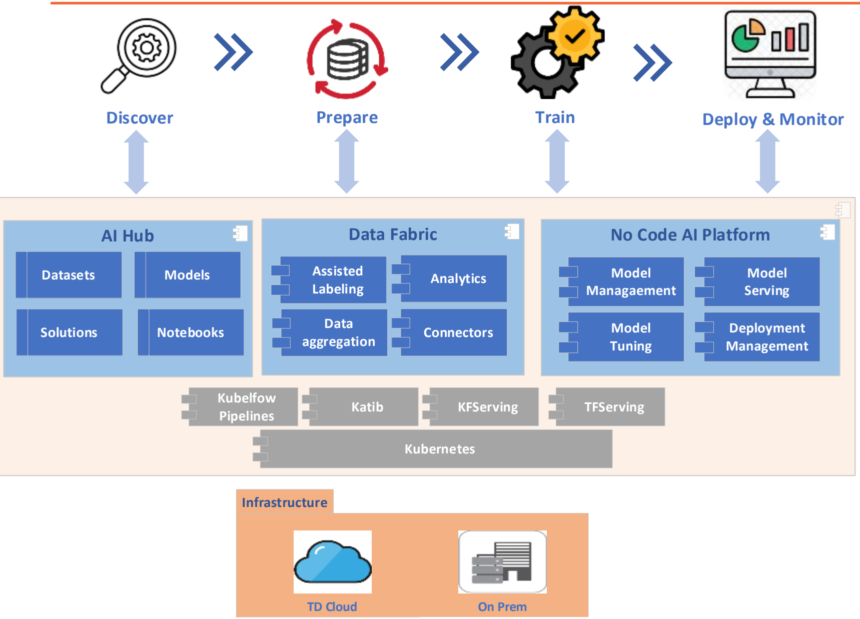
- **MLFlow quick code bytes helped with the** problems associated with the implementation of ML in production.

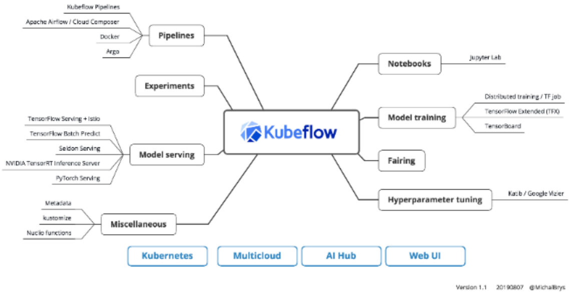
- Global Scaling is now easier.

# MLOps Use Case – 2  Using Kubeflow

- ❖ A surveillance solution needs identification of objects and activities in live video feeds.
- ❖ Operators to be able to retrain models whenever they needed to add different categories or items/activities of interest.
- ❖ Abstract all training infra/pipeline aspects from the operators because these are not technical but business folks.
- ❖ 24 hours for the new trained model to be tested and deployed.

# MLOps Use Case – 2 Using Kubeflow (ThinkDeeply.ai)




**Requirement:**

- As a small team, wanted the key resources to focus on the core job of model building and tuning and not to waste cycles on infrastructure.
- To ensure that the solutions will deploy and scale in client's infrastructur without much effort

**Tools :**

- Leveraged the Kubeflow pipelines to simplify the model training TFServi and KFServing for serving TensorFlow and Pytorch models.

**Benefits:**

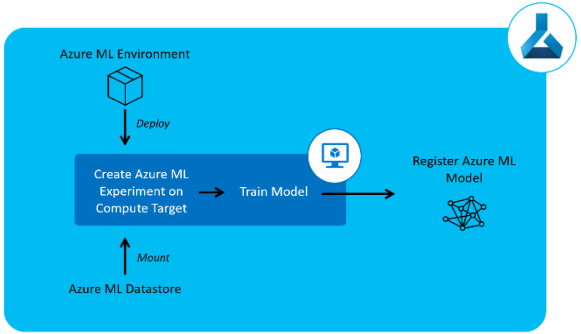
- Standardized environment simplifies task of Data Scientists and ML developers
- Significant reduction in time spent in operational and infrastructure management tasks
- Easy to deploy and scale across different environments – on-premise ar combination of Cloud

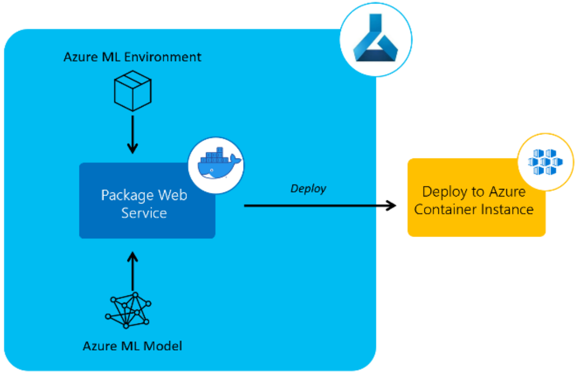
**Challenges:**

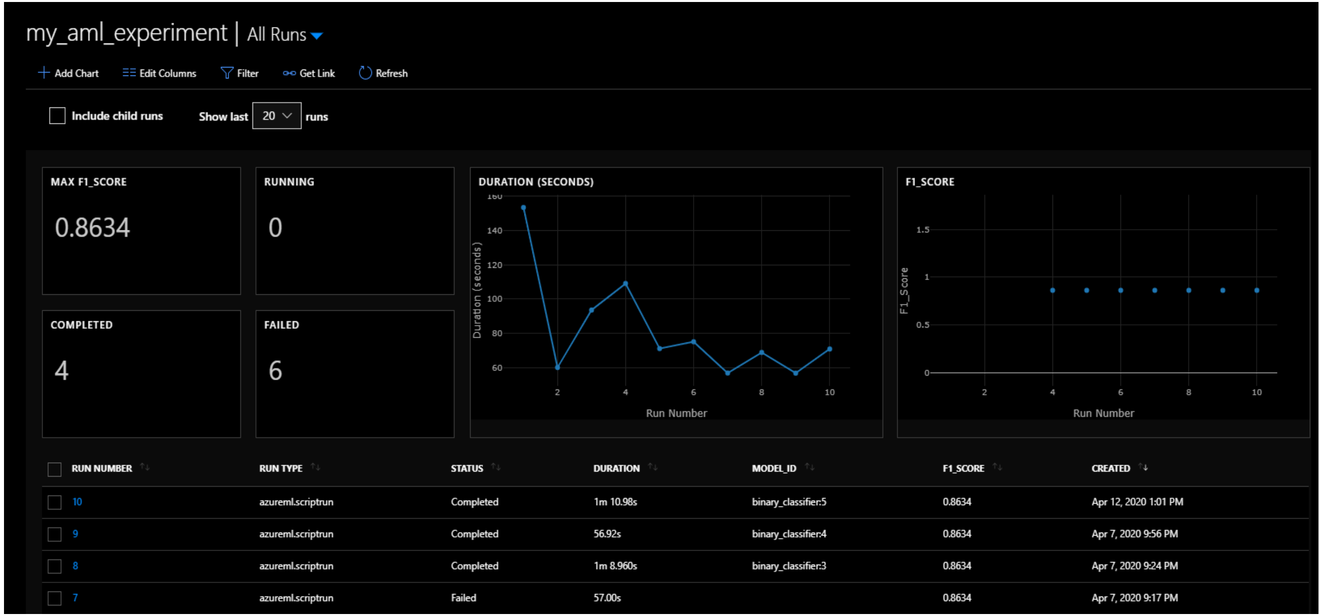
- Learning Curve
- Memory requirements

[Home | Thinkdeeply](#)
[Kubeflow — a machine learning toolkit for Kubernetes | by Michal Brys | Medium](#)

# MLOps Use Case – 3  Using Azure ML



# MLOps Benefits

## MLOps Benefits

### Automation / Observability

- Code drives **generation** and **deployments**
- Pipelines are **reproducible** and **verifiable**
- All artifacts can be **tagged** and **audited**

### Validation

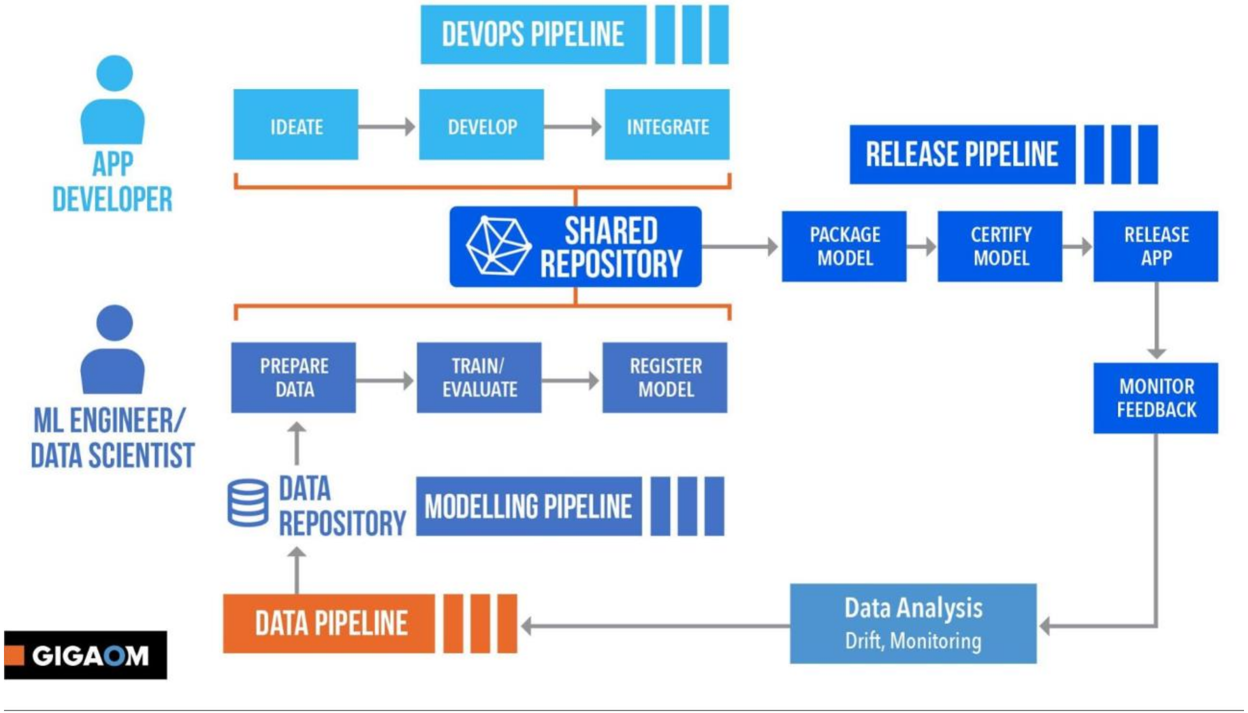
- SWE best practices for quality control
- Offline comparisons of model **quality**
- Minimize **bias** and enable **explainability**

### Reproducibility /Auditability

- Controlled rollout capabilities
- Live comparison of predicted vs. expected performance
- Results fed back to watch for drift and improve model

**== VELOCITY and SECURITY (For ML)**

# Convergence of MLOps and DevOps Teams



# MLOps Roadmap Checklist for an Organization

## Capabilities

- Process
- Technology
- People / Resources
- Culture

## Time Horizon

- Now
- Next
- Later
- Luxury

- Identify and prioritize mission-critical AI/ML workflows that can benefit from retroactive MLOps implementation.
- Create a roadmap for applying MLOps to the backlog of workflows
- Collaborate on the organization’s MLOps governance and policy enforcement strategy

# Thank you

Consumer Healthcare

– Final

Thank you

## Questions ?

## Subroto Mukherjee