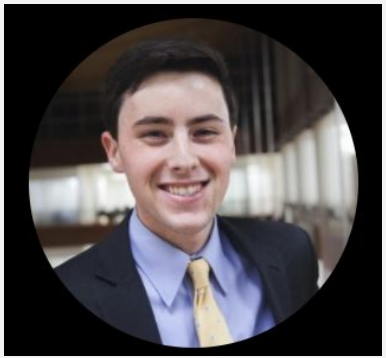
Unified ML Monitoring Hands-on Workshop

Presented by

Max Fisher, *Solution Architect*

Amy Wang, *Senior Solution Architect*

Meet the Presenters



Max Fisher
Solution Architect
max.fisher@databricks.com

Amy Wang
Solution Architect
amy.wang@databricks.com

Agenda

- Introductions, 5m
- Databricks Overview & Lakehouse, 5m
- Unified ML Monitoring Overview, 15m
- Hands-on lab, 55m
- Q&A, 10m

The Data and AI Company

Lakehouse

One simple platform to unify all of your data, analytics, and AI workloads

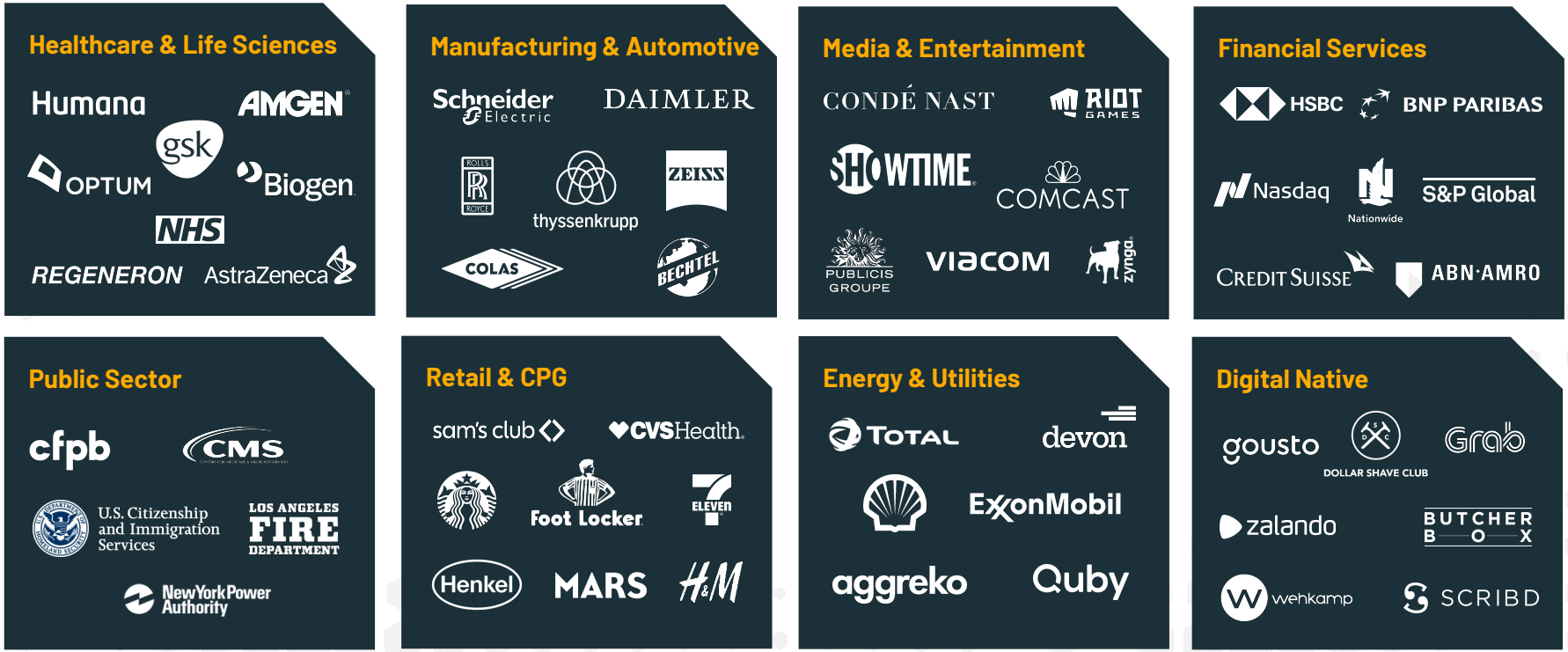
CUSTOMERS

5000+

Across the globe

ORIGINAL CREATORS

We're working with enterprises in every industry



databricks

What is the “Lakehouse” Paradigm?

databricks

Data Warehouses

Pros

- Great for Business Intelligence (BI) applications

Cons

- Limited support for Machine Learning (ML) workloads
- Proprietary systems with only a SQL interface

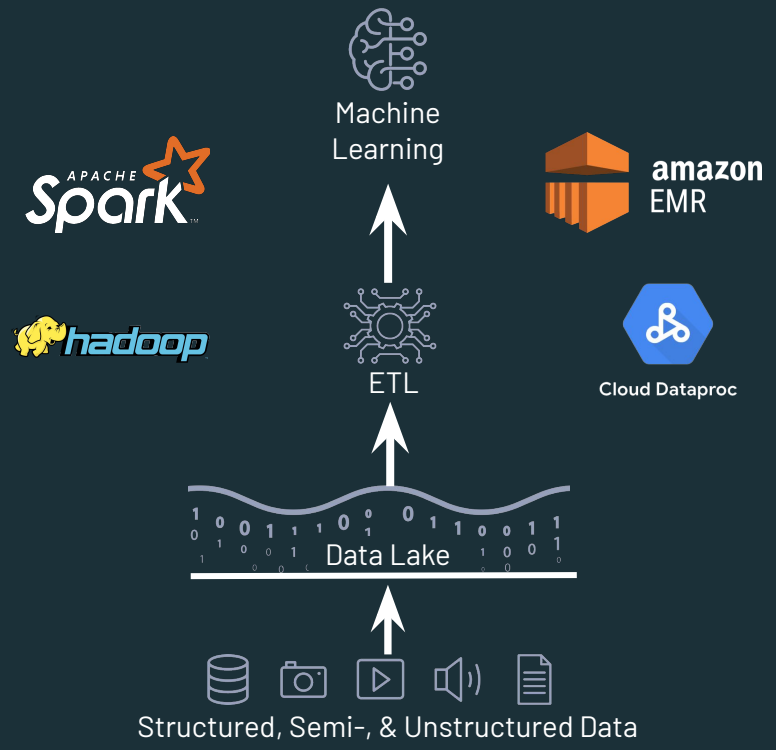
Data Lakes

Pros

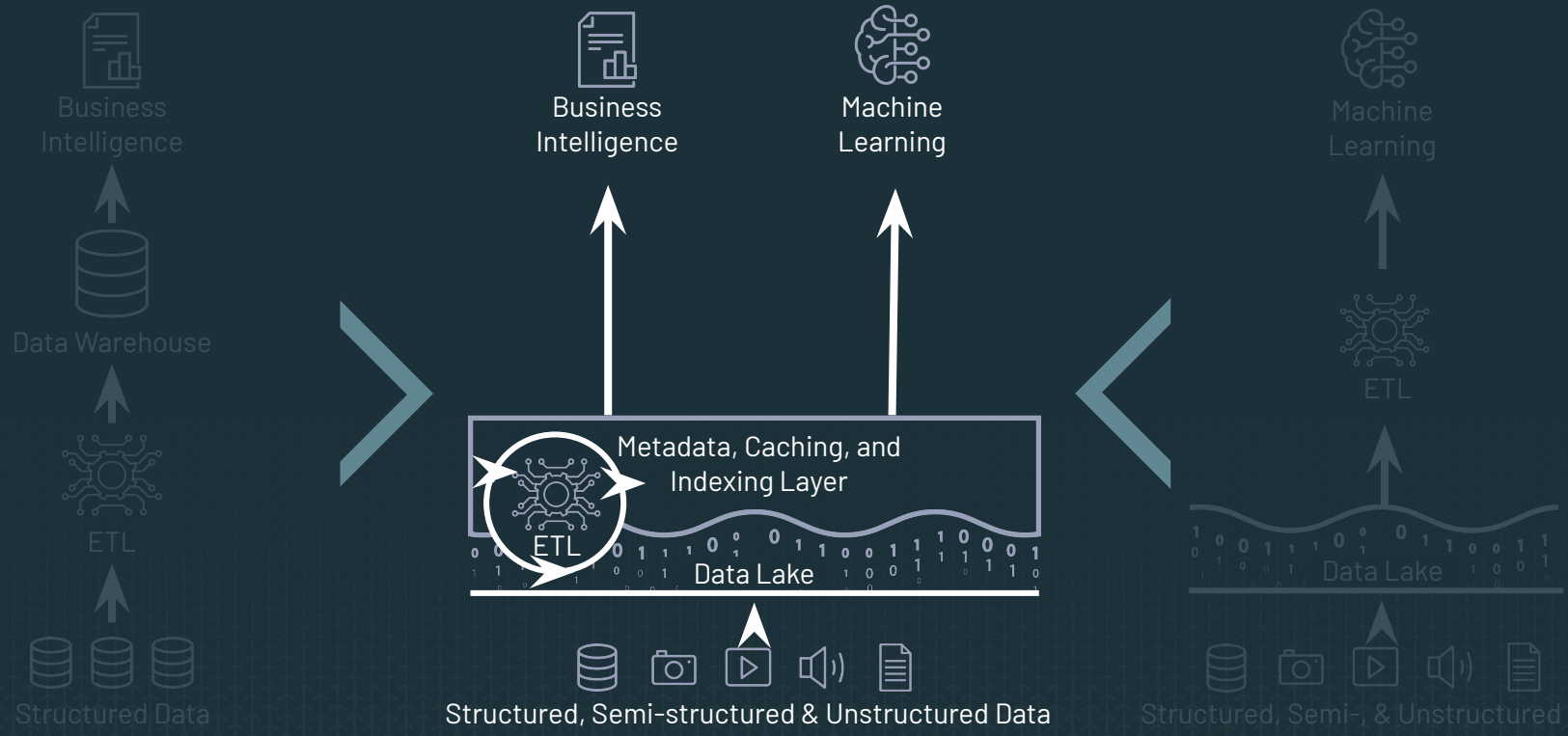
- Supports ML
- Open formats and big ecosystem

Cons

- Poor support for BI
- Complex data quality problems

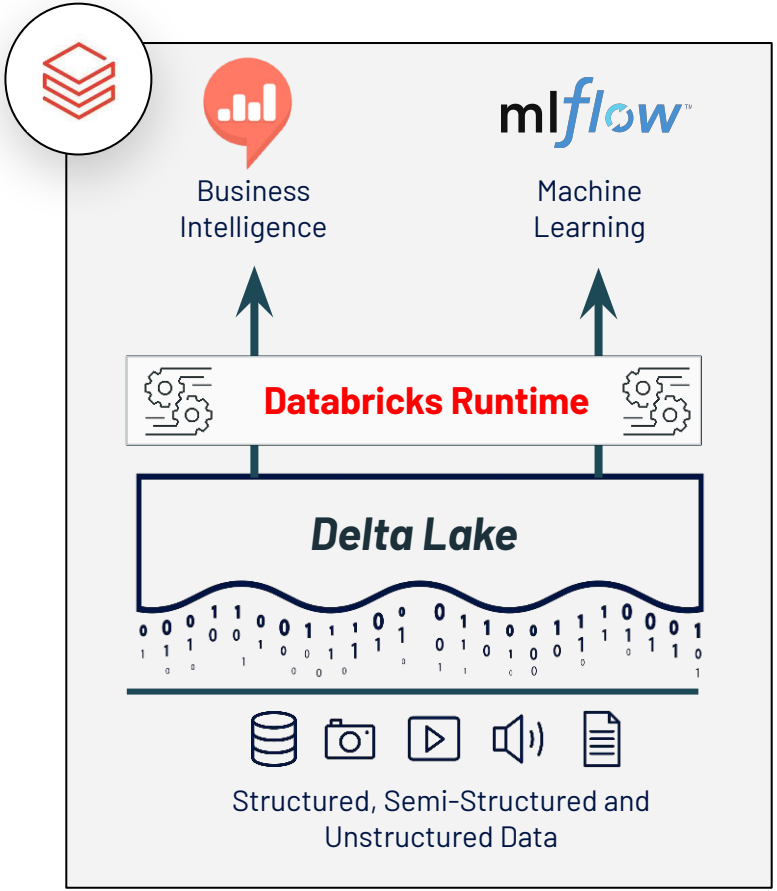


New Way Forward: lakehouse



databricks

The Databricks Lakehouse Platform



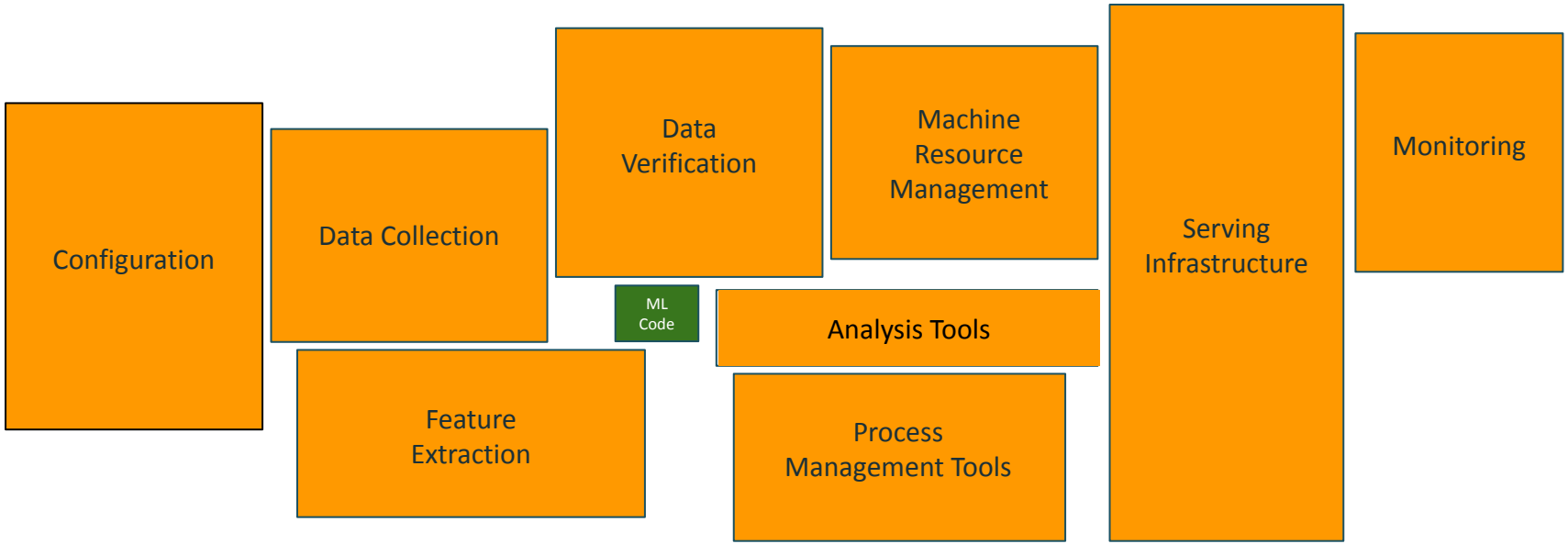
Introduction to the Lab

databricks

Quick Poll: What is the biggest challenge you face with your Machine Learning Projects?

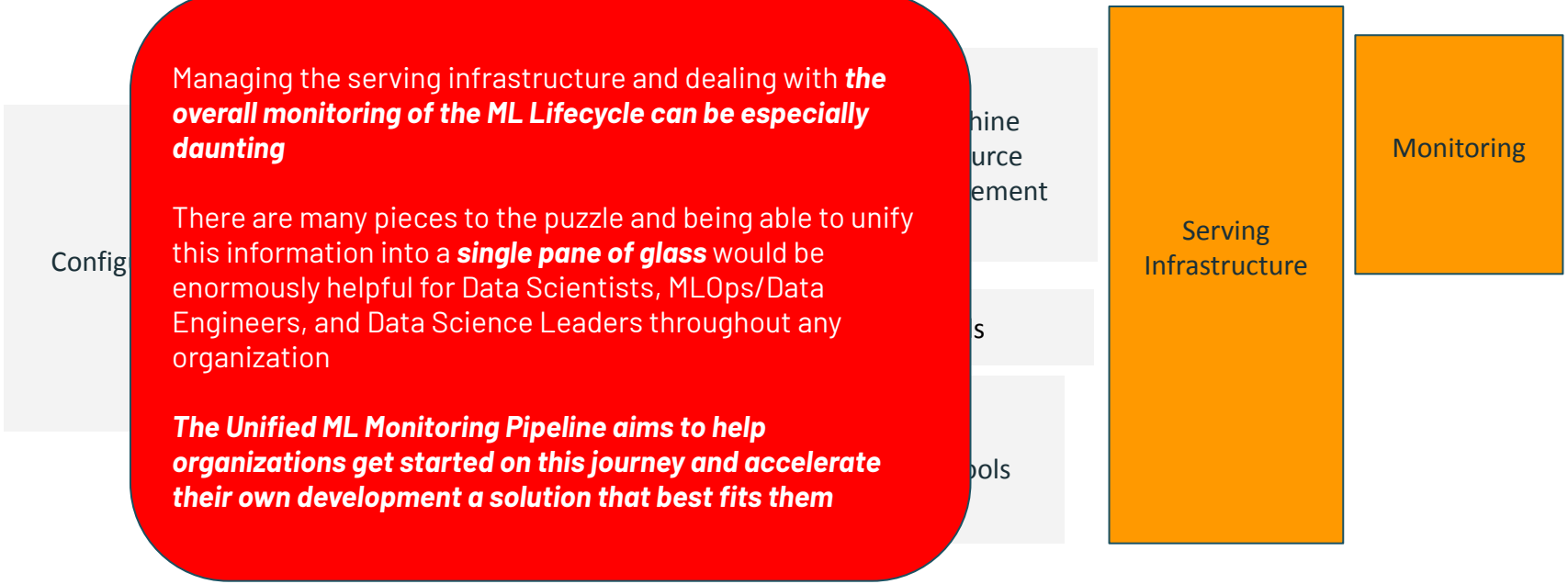
- Figuring out the right algorithm to use to train your model?
- Tracking all the data associated with training and testing your model?
- Operationalizing your finished models?
- Governing the lifecycle of your machine learning models?

Hardest Part of ML isn't ML, it's Data

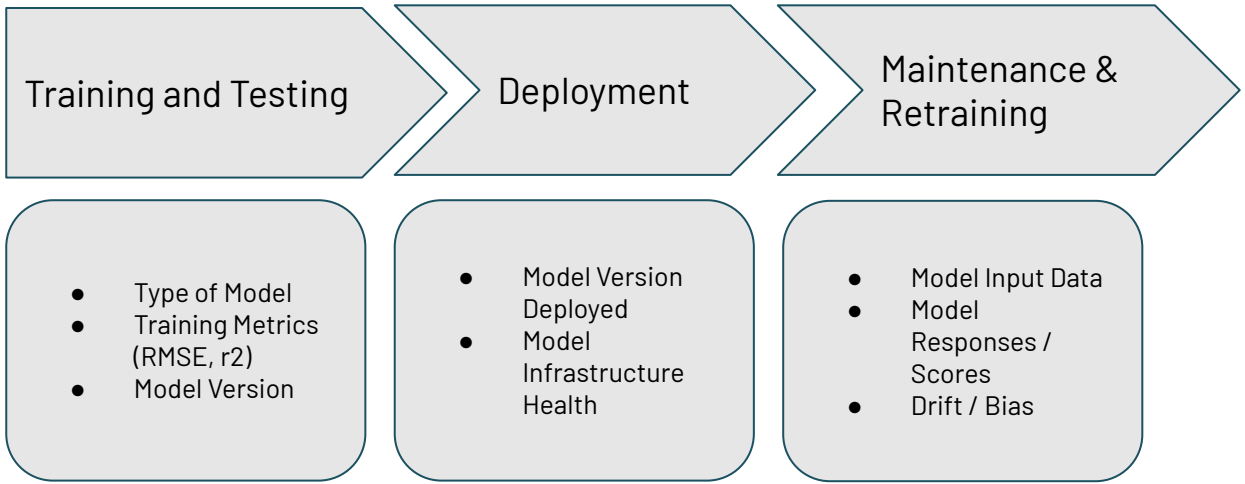


Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small green box in the middle. The required surrounding infrastructure is vast and complex.

Monitoring Can Be A Complex Part of this Problem



What do we need to track from the ML Lifecycle?



Where can we find all this information?

The Current Picture Across Clouds

databricks

Azure Machine Learning

Amazon SageMaker

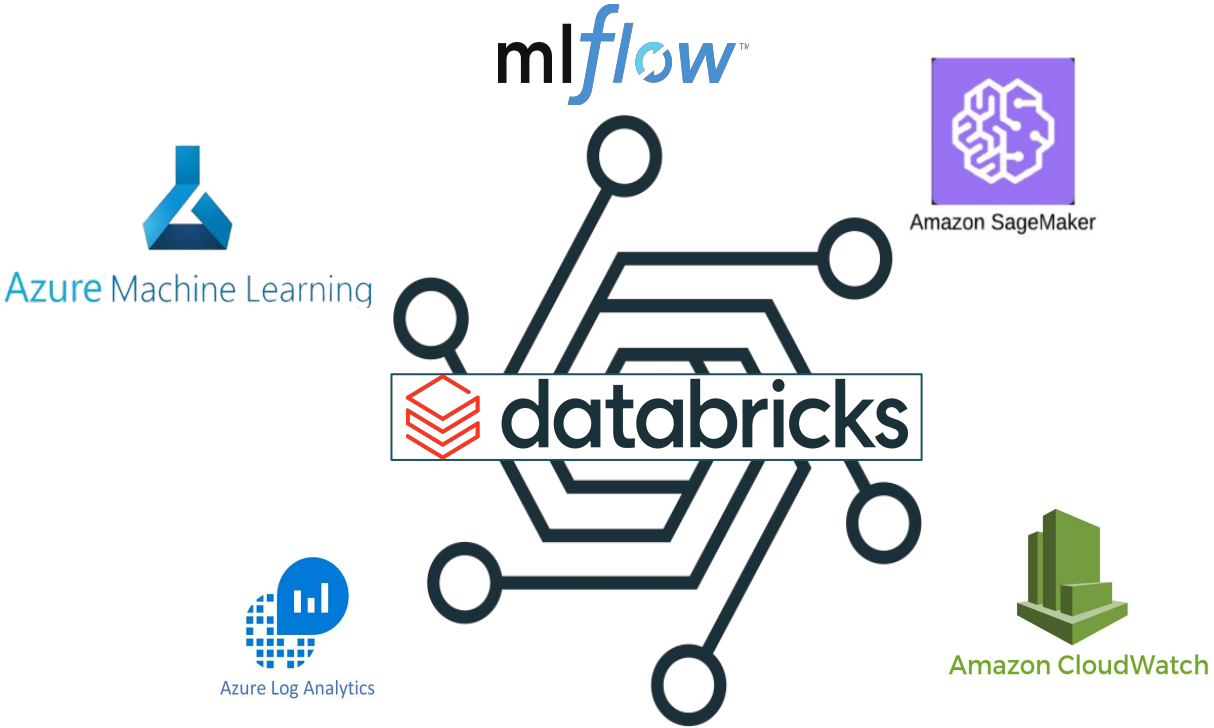
mlflow™

Azure Log Analytics

Amazon CloudWatch

As customers become multi-cloud themselves, it makes this problem even more self-evident with an expanded amount of services and tools to monitor.

The Solution

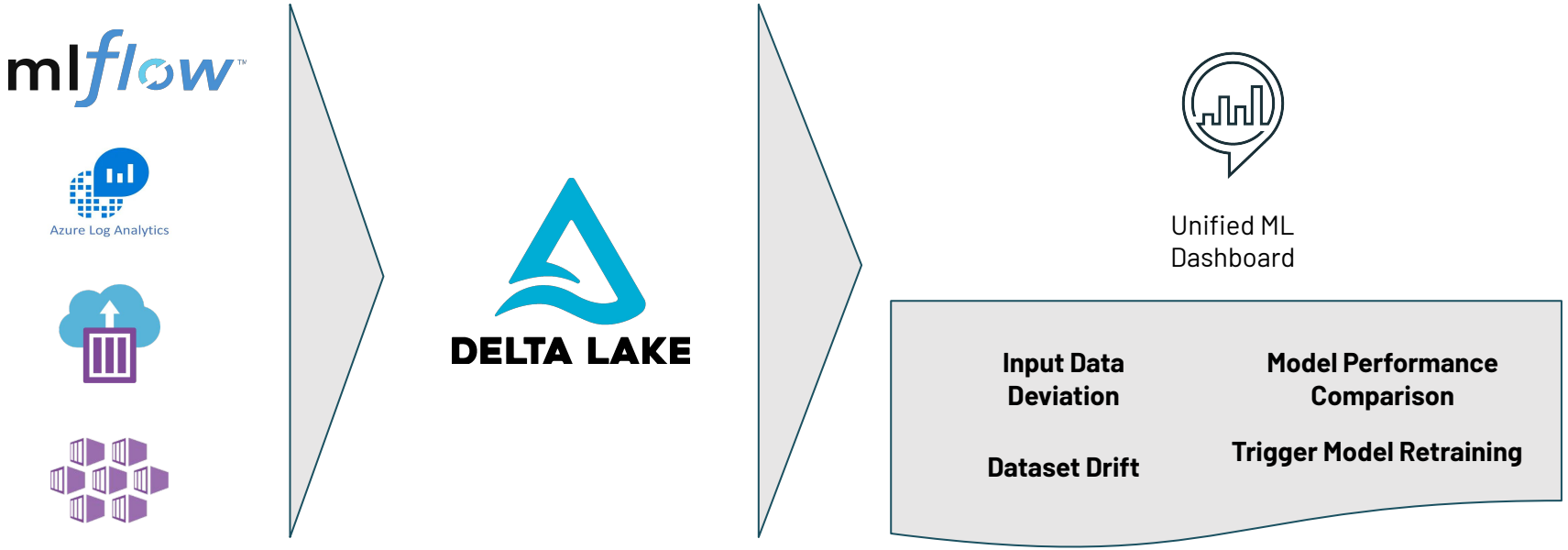


Unite MLFlow data with monitoring and log information from external services like AzureML or SageMaker

Leverage SQL Analytics to visualize and set alerts on metrics like endpoint traffic or model drift

Databricks is uniquely positioned to be the Unified Monitoring Platform for Machine Learning

What Solutions Can Be Powered by this Data?



Being able to unify all this data in Databricks allows Data Science teams to be able to better understand the performance of their models, better respond to changes in the data, and ultimately deliver more value

Why is capturing this information important?

83% CEOs say AI is a strategic priority

MIT Sloan Management Review

$3.9T Business value created by AI in 2022

Gartner

85% Of big data projects fail

Gartner

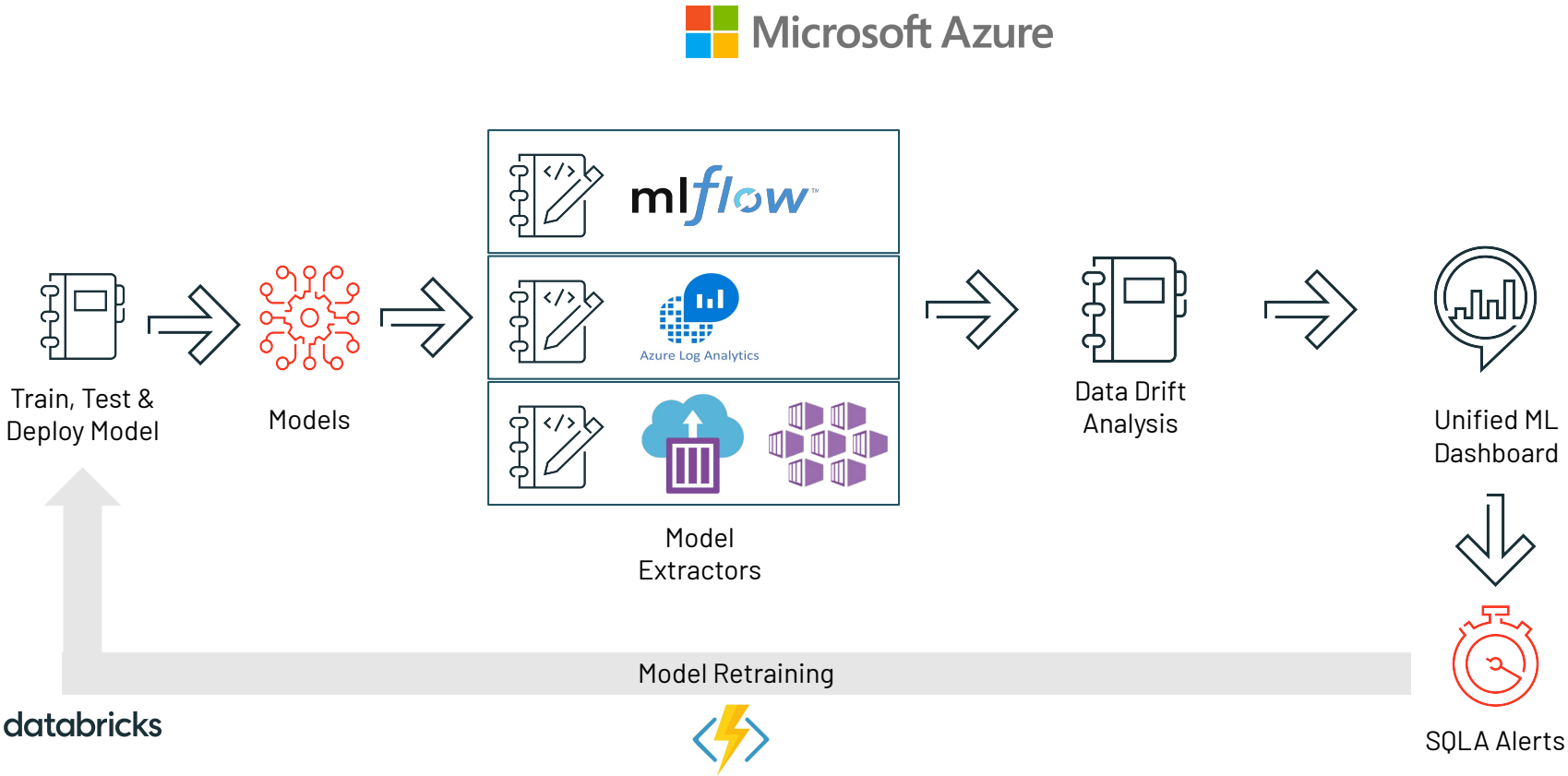
87% Of data science projects never make it into production

VB

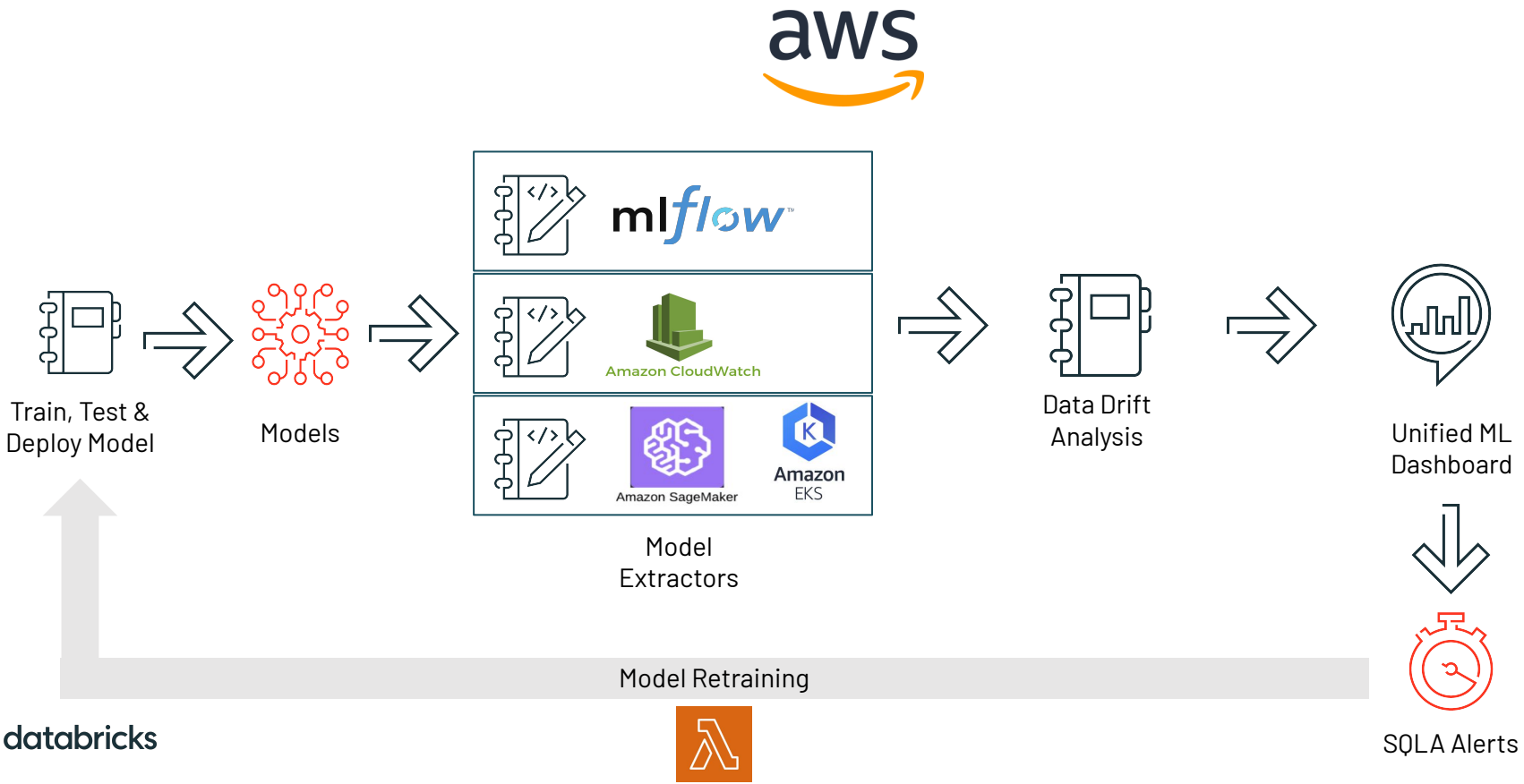
How Can I Do This in Databricks?

databricks

Architecture of the Solution



Architecture of the Solution



Hands-on Lab

Max will be leading the lab

Amy is available in the chat to answer questions or provide assistance

Content Links

- GitHub Repository for all Lab Materials:
  - mpfis/unified-ml-monitoring-on-databricks (github.com)
- PDF of this Presentation
- Lab guide

Similar Workshops

- Introduction to Delta and Lakehouse
  - Description
  - First and third Tuesday every month, sign-up link
- What's new in Databricks?
  - Description
  - First Wednesday every month, sign-up link

Q&A and Post-workshop Discussion

Thank you for joining us. We're here to answer any questions you have about the content, Databricks, or anything else. We'll send out a copy of this presentation after we wrap up.

We'll send out a survey after the workshop. Let us know what you liked and didn't like so that we can continue to improve.

If you're not currently in contact with your Databricks account team and would like to be, let the presenters know and we can assist.

Contact info for presenters: max.fisher@databricks.com, amy.wang@databricks.com

We're Hiring Solution Architects

If you or someone in your network is interested,
we have East and West coast positions available.

Email ericka.styles@databricks.com with your resume or LinkedIn.
You can ask your presenters for more information.

databricks.com/company/careers

Looking for applicants with backgrounds in any of these skills:
data engineering, machine learning, **or** analytics.

Also hiring in Europe and Asia.