Databricks Demo: Transfer Learning with MLflow, the #1 Tool to Learn Today

Feifei Wang

Senior Data Scientist @ Databricks

feifei.wang@databricks.com

databricks

Feifei Wang

Ph.D in Applied Math & Computer Science @ Iowa State University

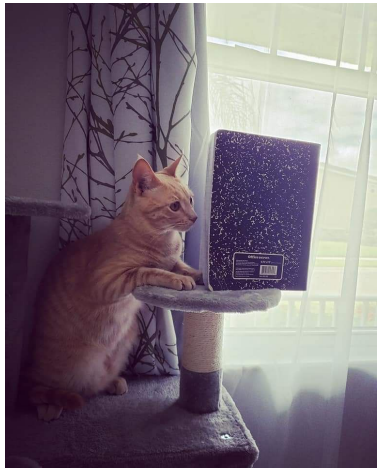
Former Senior Decision Scientist @ The Walt Disney Company

Current Senior Data Scientist @ databricks

Feifei Wang

Hobbies: Ballet and Piano

Pet: Simba



databricks

Databricks

Unified Data Analytics Platform

Collaborative Workspace

Data Quality + Consistency

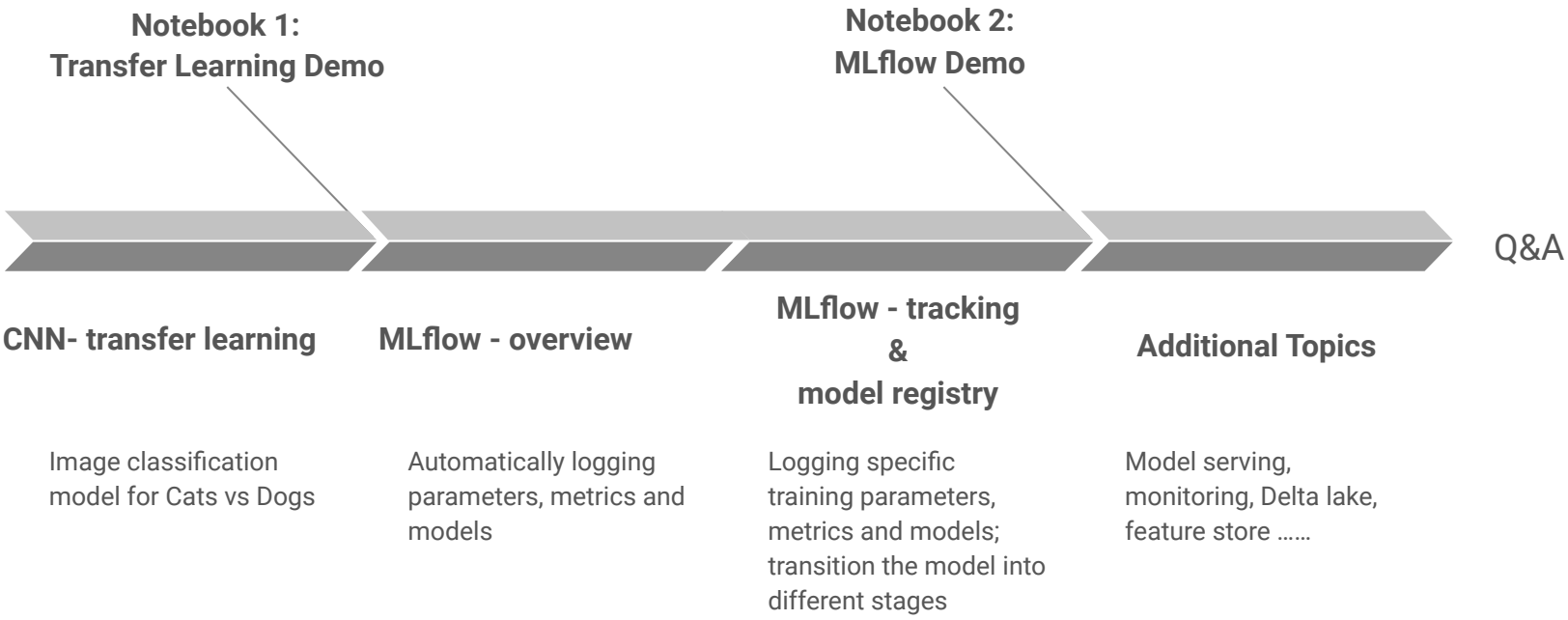
Reproduce + Productionize Models

databricks

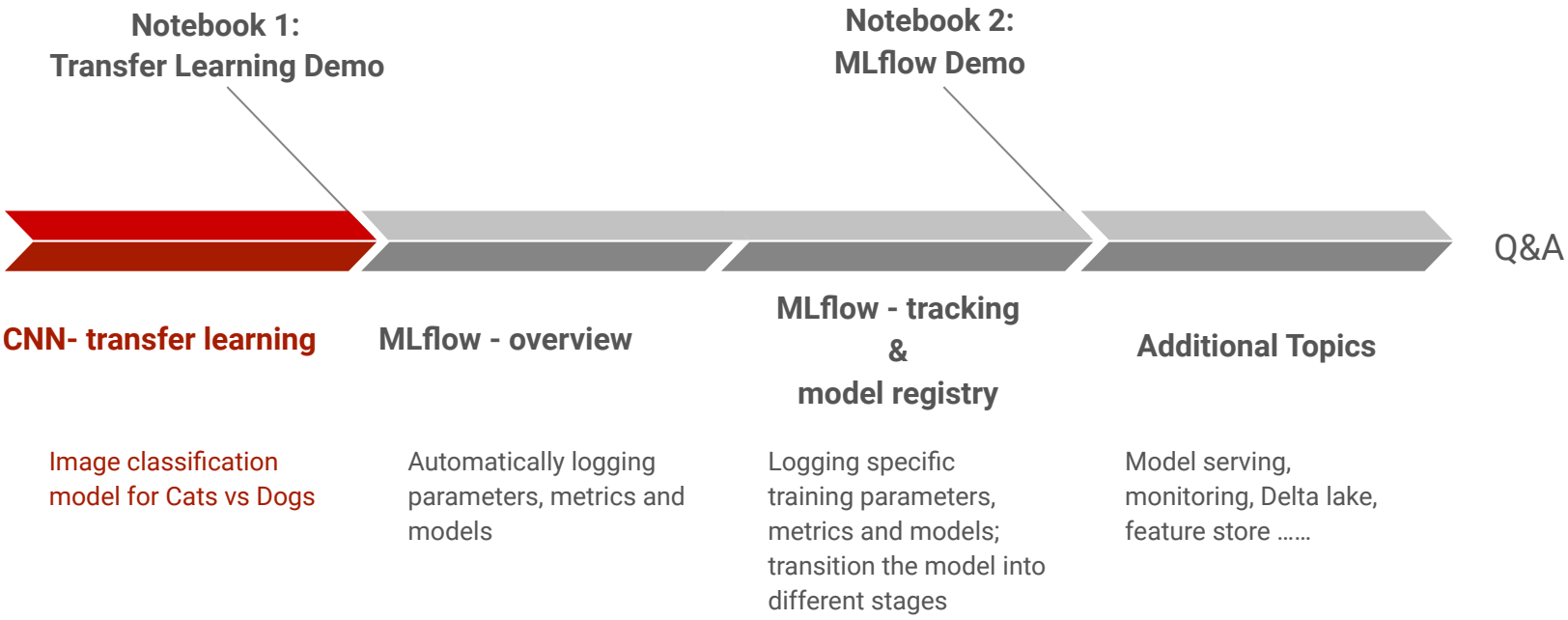
DELTA LAKE

mlflow

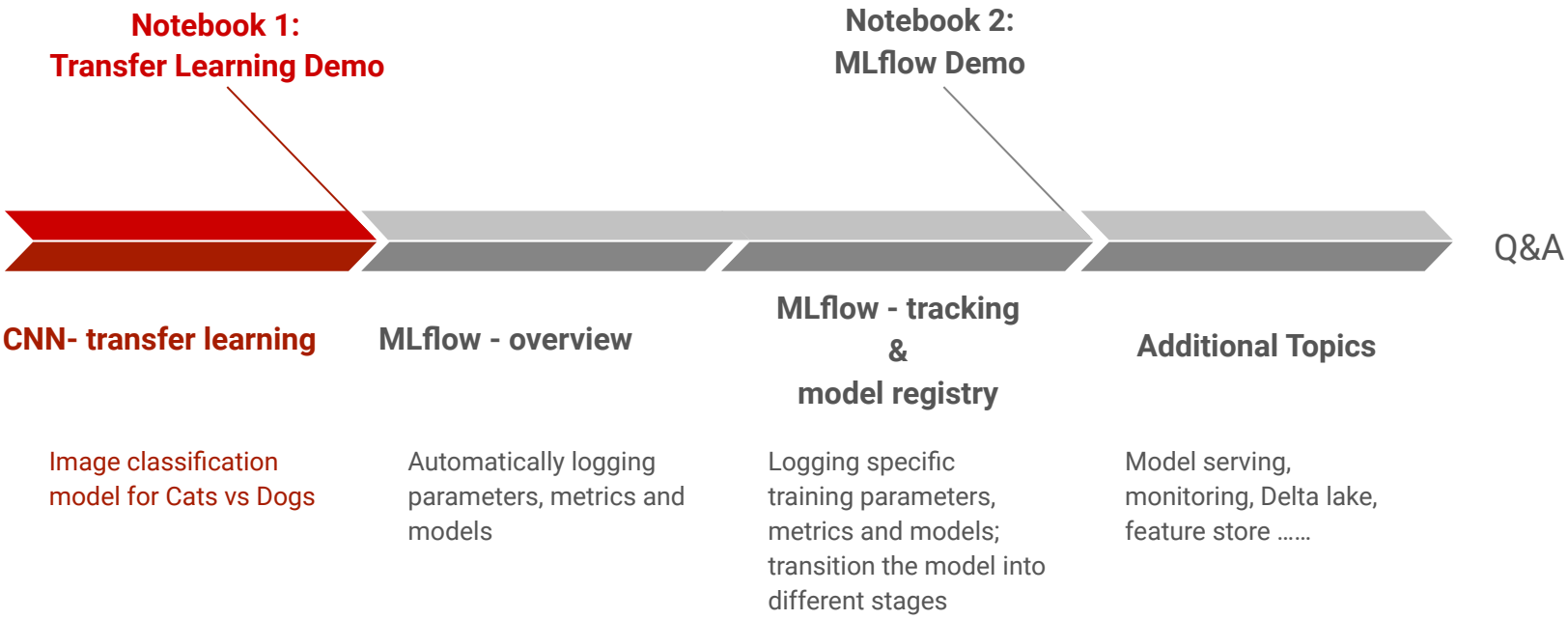
Outline


Notebook 1:
Transfer Learning Demo
Notebook 2:
MLflow Demo
Q&A
CNN- transfer learning
MLflow - overview
MLflow - tracking
&
model registry
Additional Topics
Image classification model for Cats vs Dogs
Automatically logging parameters, metrics and models
Logging specific training parameters, metrics and models; transition the model into different stages
Model serving, monitoring, Delta lake, feature store ……

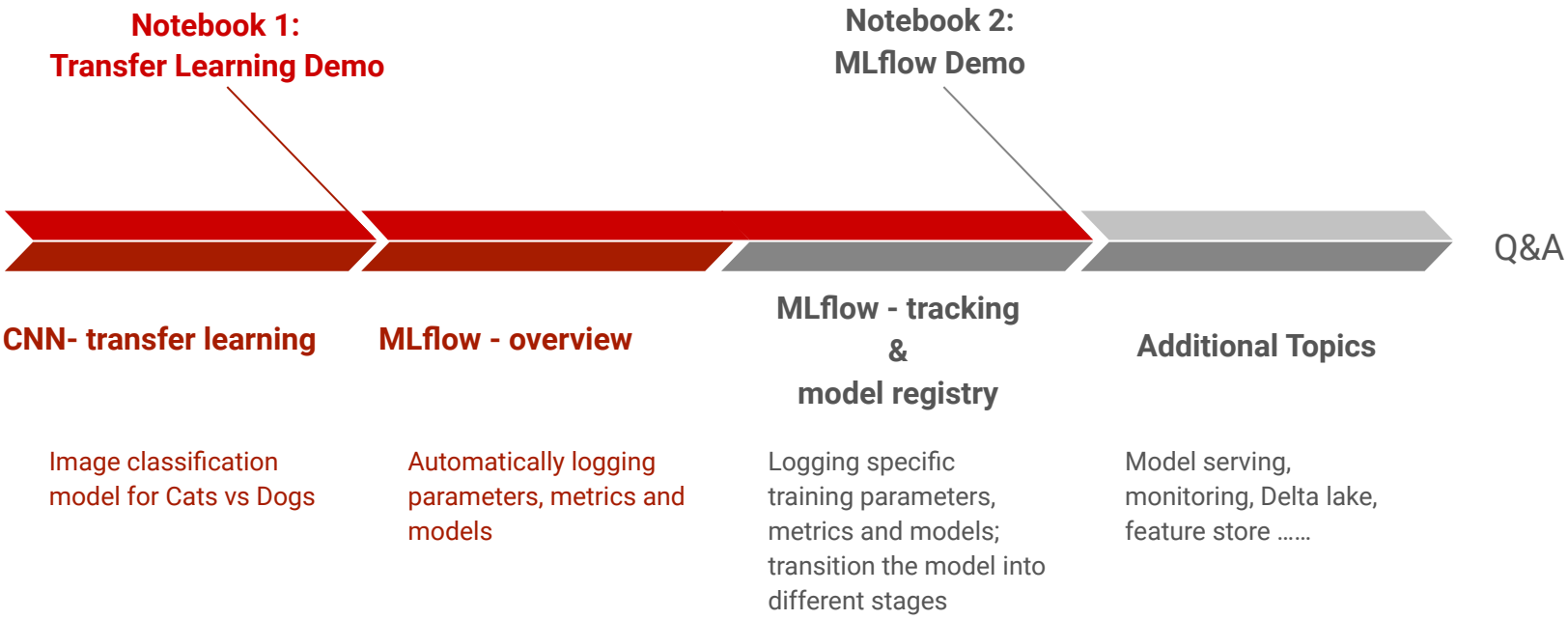
Outline


Notebook 1:
Transfer Learning Demo
Notebook 2:
MLflow Demo
Q&A
CNN- transfer learning
MLflow - overview
MLflow - tracking
&
model registry
Additional Topics
Image classification model for Cats vs Dogs
Automatically logging parameters, metrics and models
Logging specific training parameters, metrics and models; transition the model into different stages
Model serving, monitoring, Delta lake, feature store ……

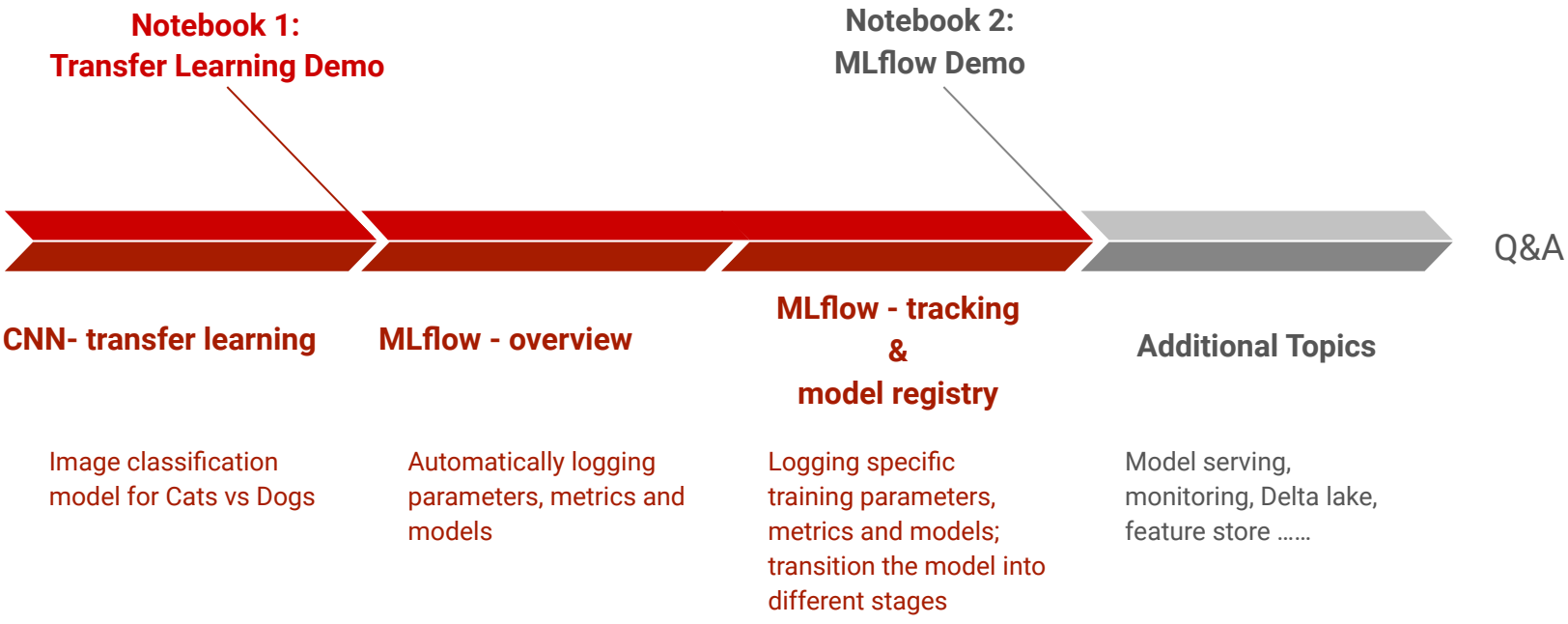
Outline


Notebook 1:
Transfer Learning Demo
Notebook 2:
MLflow Demo
Q&A
CNN- transfer learning
MLflow - overview
MLflow - tracking
&
model registry
Additional Topics
Image classification model for Cats vs Dogs
Automatically logging parameters, metrics and models
Logging specific training parameters, metrics and models; transition the model into different stages
Model serving, monitoring, Delta lake, feature store ……

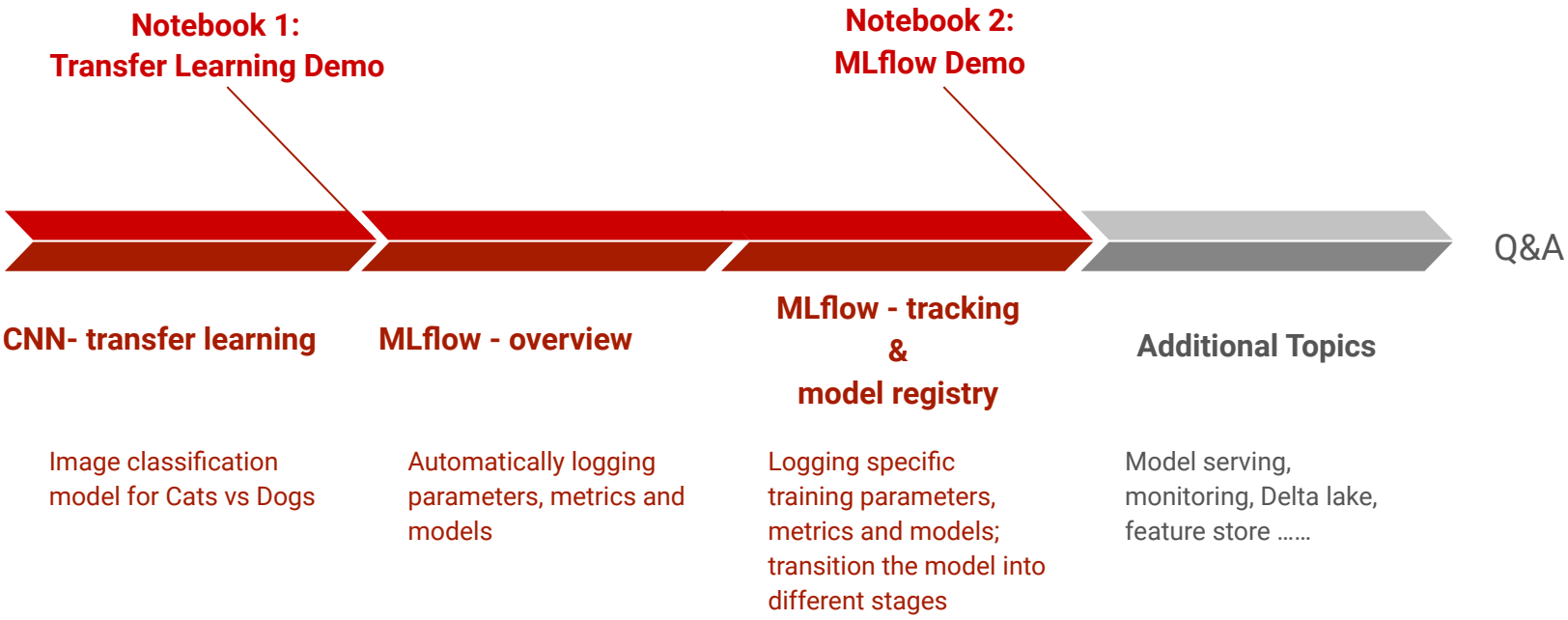
Outline


Notebook 1:
Transfer Learning Demo
Notebook 2:
MLflow Demo
Q&A
CNN- transfer learning
MLflow - overview
MLflow - tracking
&
model registry
Additional Topics
Image classification model for Cats vs Dogs
Automatically logging parameters, metrics and models
Logging specific training parameters, metrics and models; transition the model into different stages
Model serving, monitoring, Delta lake, feature store ……

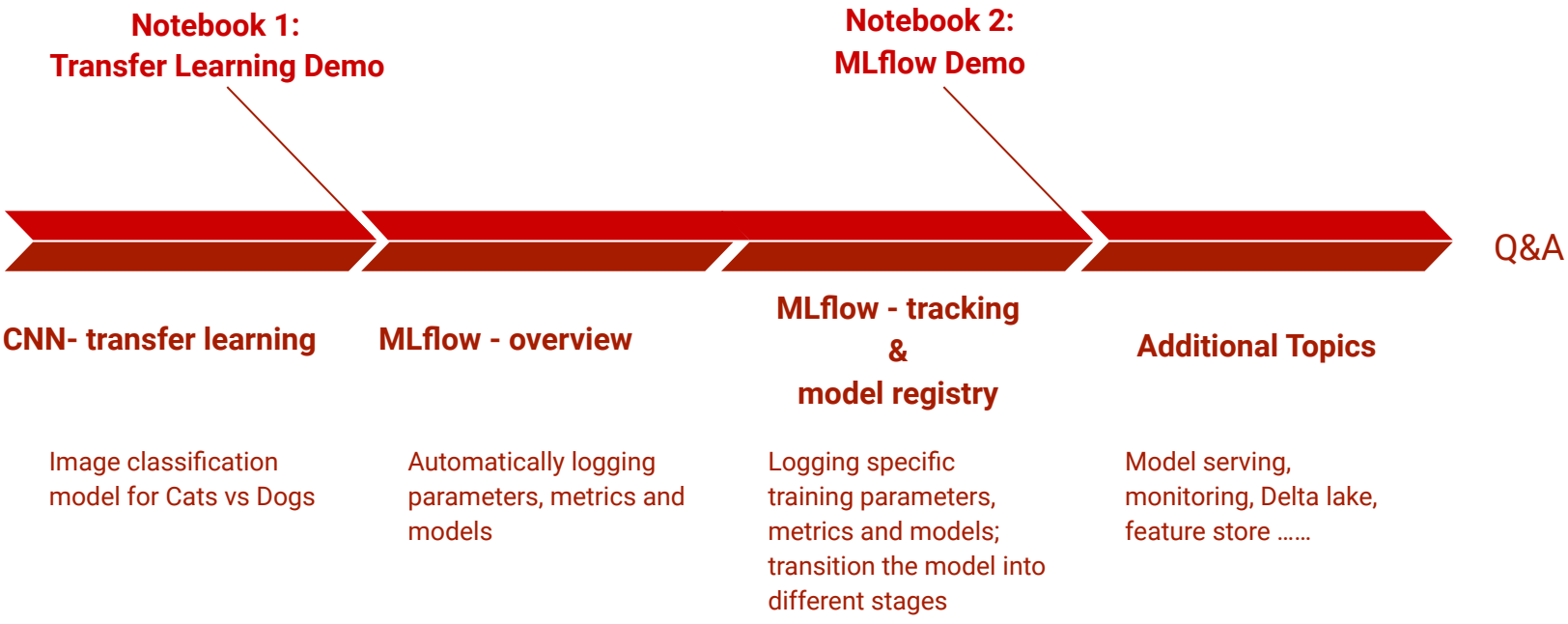
Outline


Notebook 1:
Transfer Learning Demo
Notebook 2:
MLflow Demo
Q&A
CNN- transfer learning
MLflow - overview
MLflow - tracking
&
model registry
Additional Topics
Image classification model for Cats vs Dogs
Automatically logging parameters, metrics and models
Logging specific training parameters, metrics and models; transition the model into different stages
Model serving, monitoring, Delta lake, feature store ……

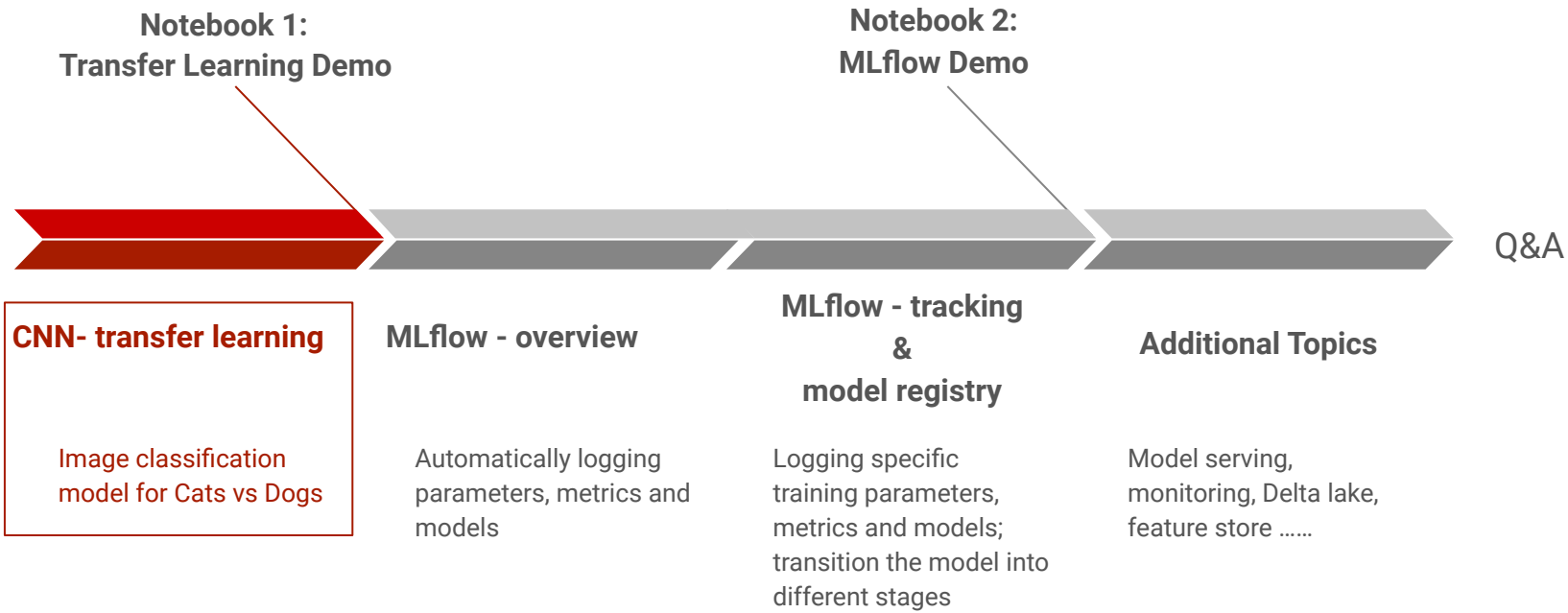
Outline


Notebook 1:
Transfer Learning Demo
Notebook 2:
MLflow Demo
Q&A
CNN- transfer learning
MLflow - overview
MLflow - tracking
&
model registry
Additional Topics
Image classification model for Cats vs Dogs
Automatically logging parameters, metrics and models
Logging specific training parameters, metrics and models; transition the model into different stages
Model serving, monitoring, Delta lake, feature store ……

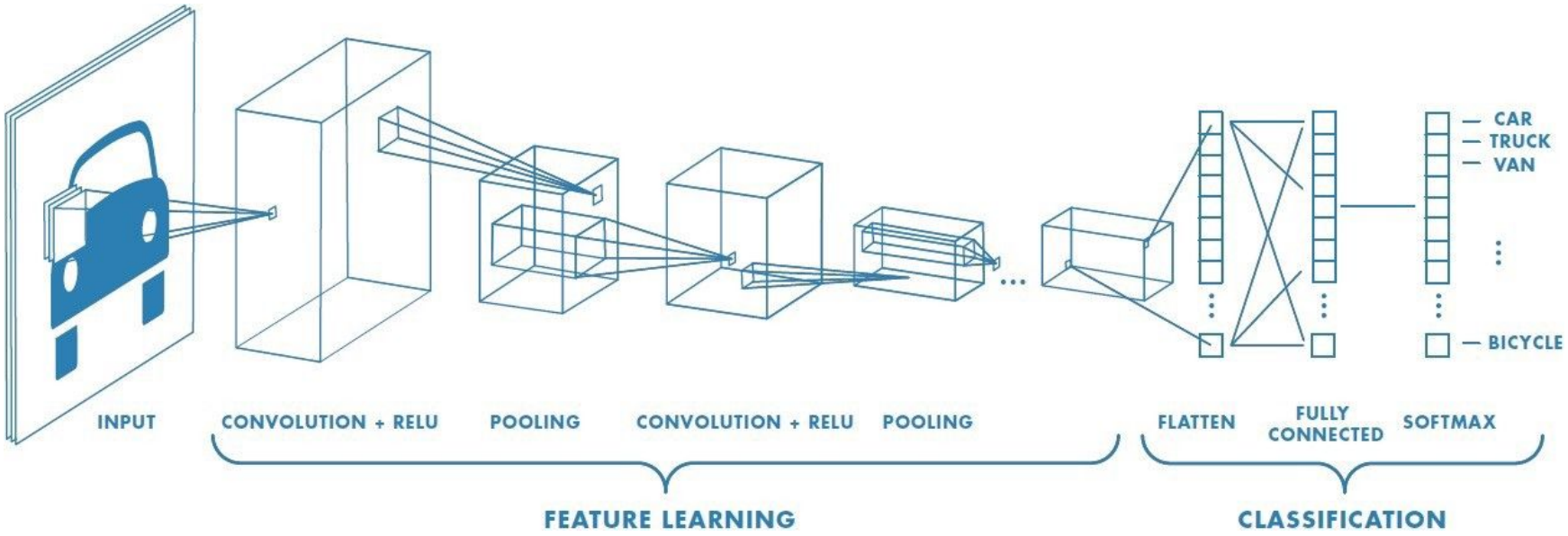
Outline


Notebook 1:
Transfer Learning Demo
Notebook 2:
MLflow Demo
Q&A
CNN- transfer learning
MLflow - overview
MLflow - tracking
&
model registry
Additional Topics
Image classification model for Cats vs Dogs
Automatically logging parameters, metrics and models
Logging specific training parameters, metrics and models; transition the model into different stages
Model serving, monitoring, Delta lake, feature store ……

Outline


Notebook 1:
Transfer Learning Demo
Notebook 2:
MLflow Demo
Q&A
CNN- transfer learning
Image classification model for Cats vs Dogs
MLflow - overview
Automatically logging parameters, metrics and models
MLflow - tracking & model registry
Logging specific training parameters, metrics and models; transition the model into different stages
Additional Topics
Model serving, monitoring, Delta lake, feature store ……

Convolutional Neural Networks



databricks

Image source

Filters and Feature Maps

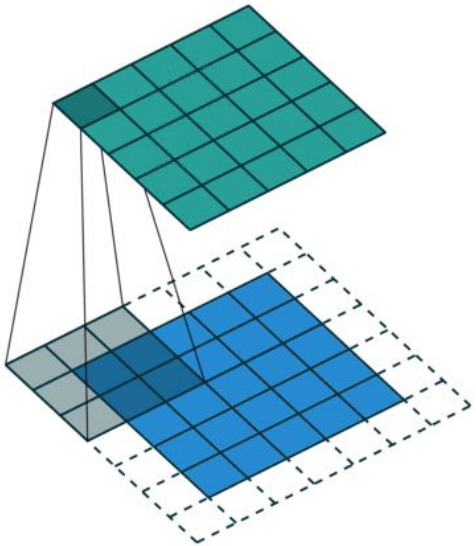
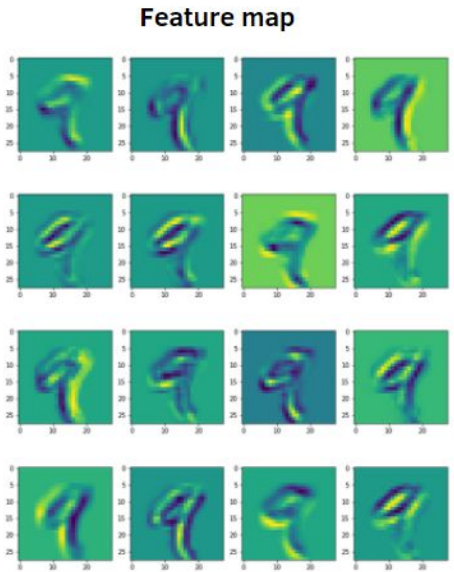


Image source



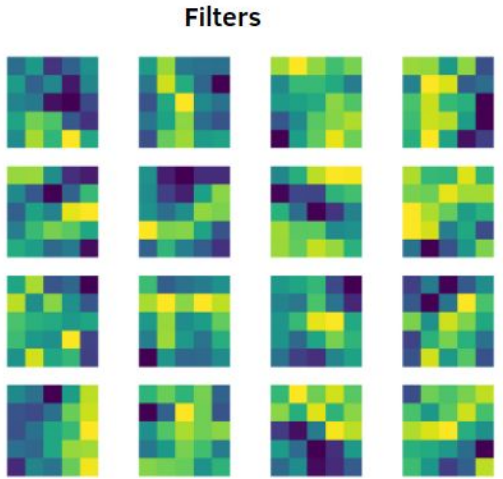
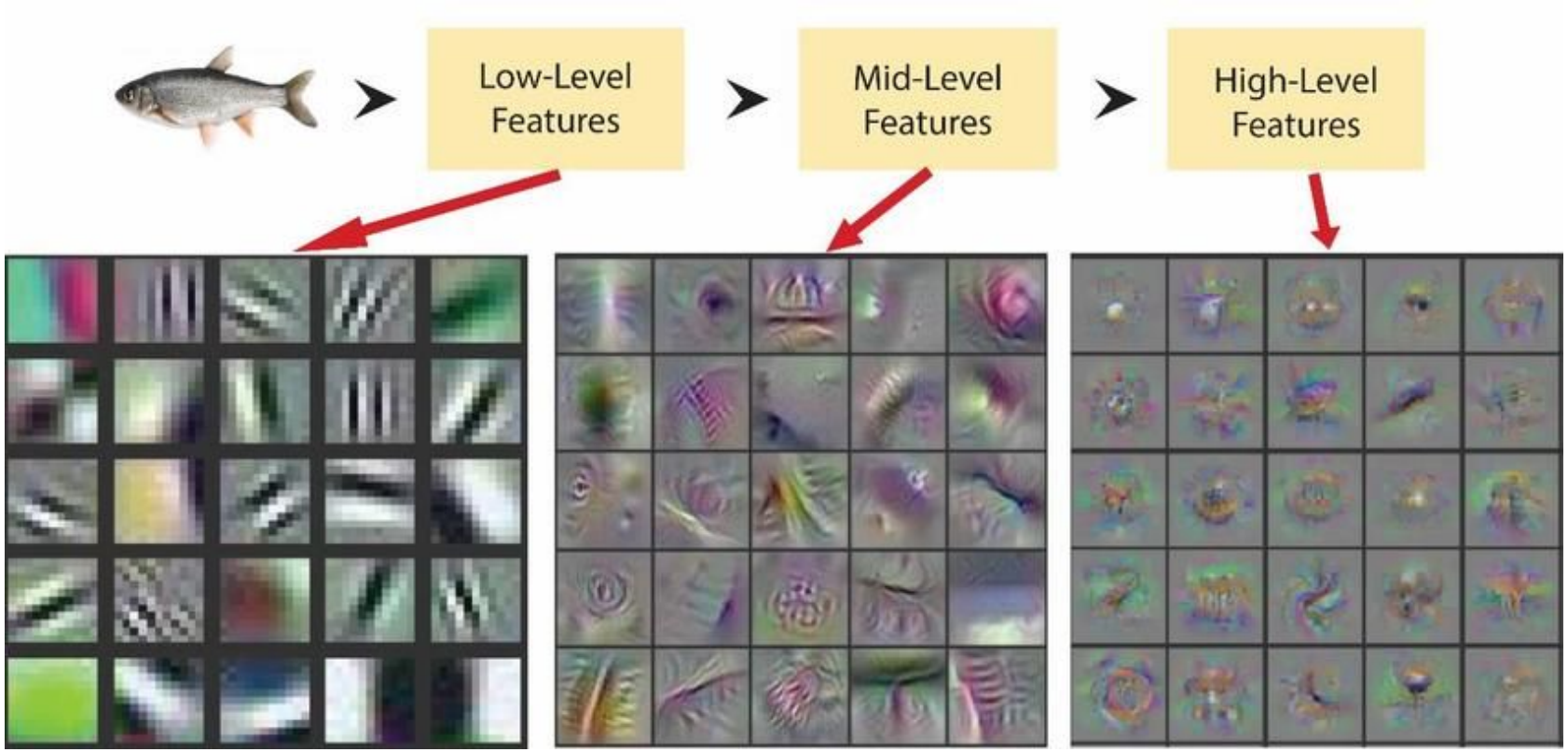


Image source

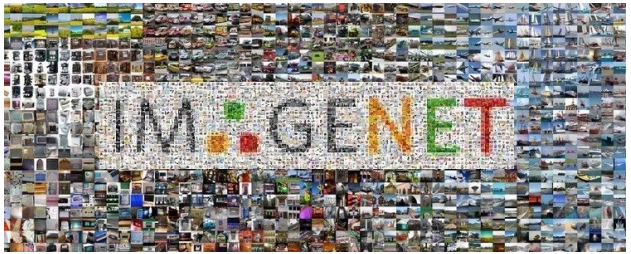
Deeper Layers → More Complex Features

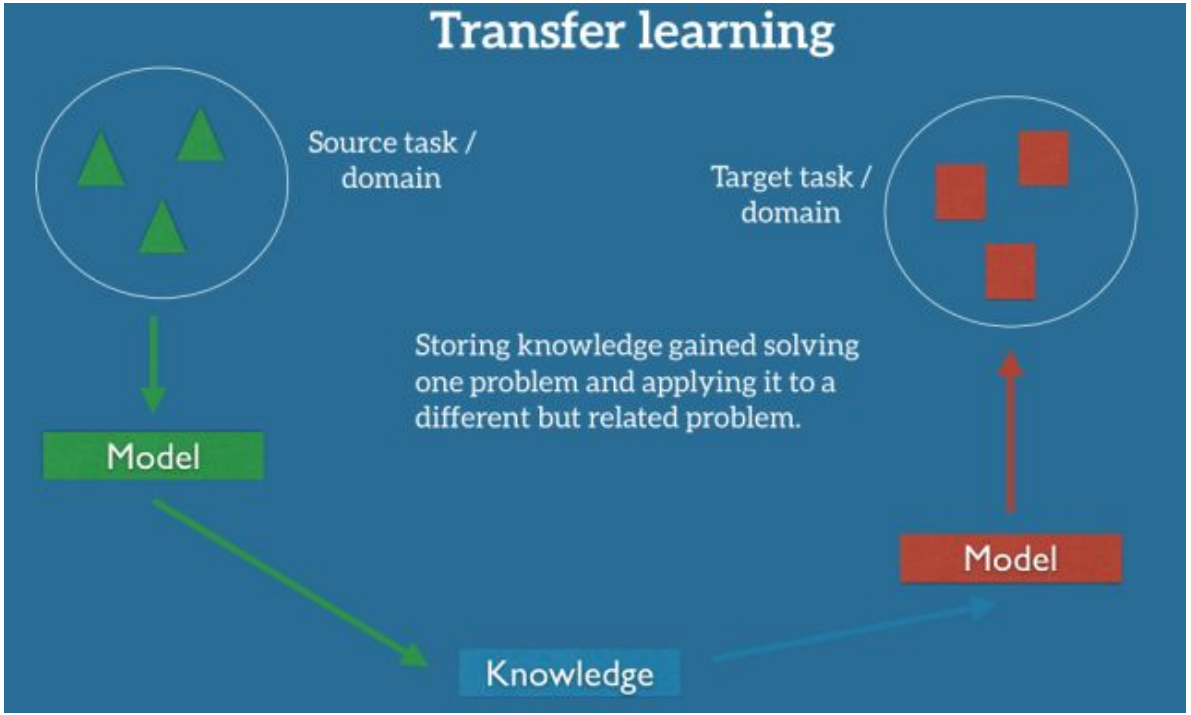


databricks

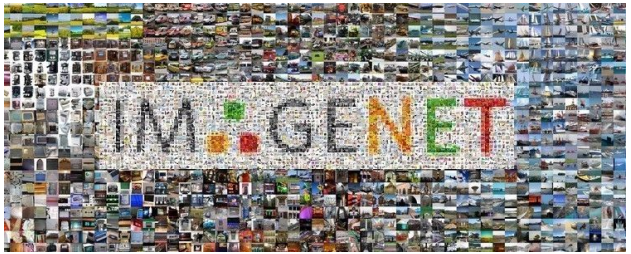
Image source

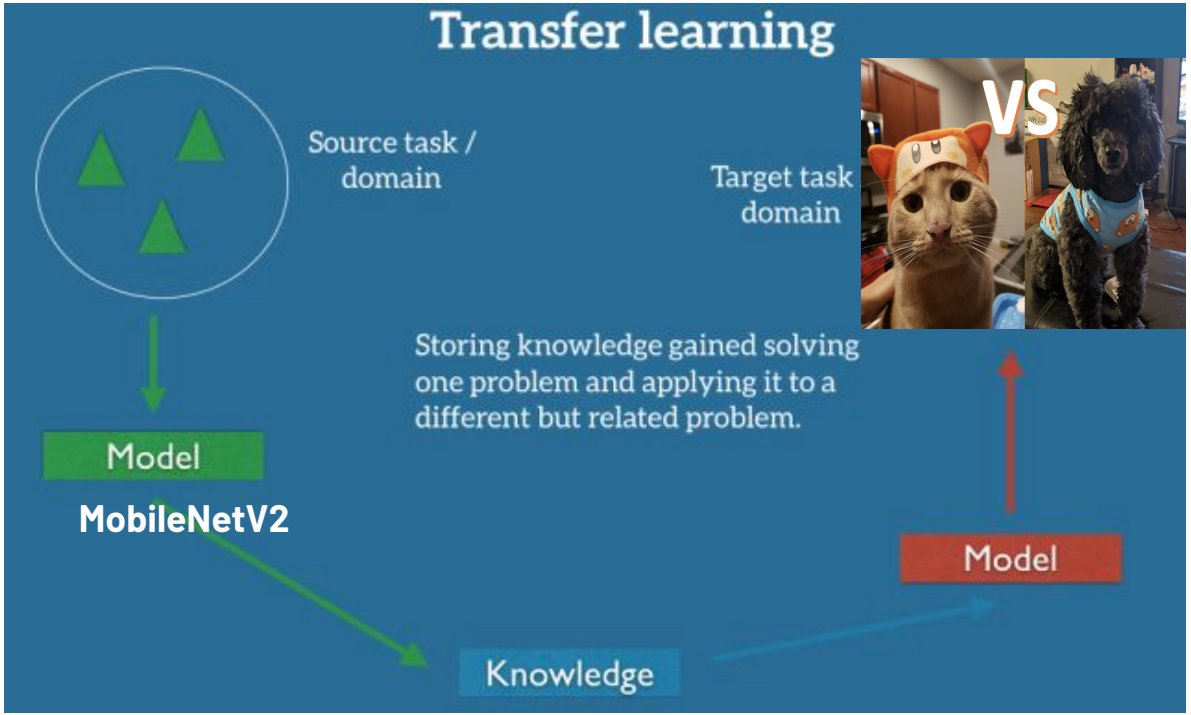
Transfer Learning





Transfer Learning





Feature Extraction

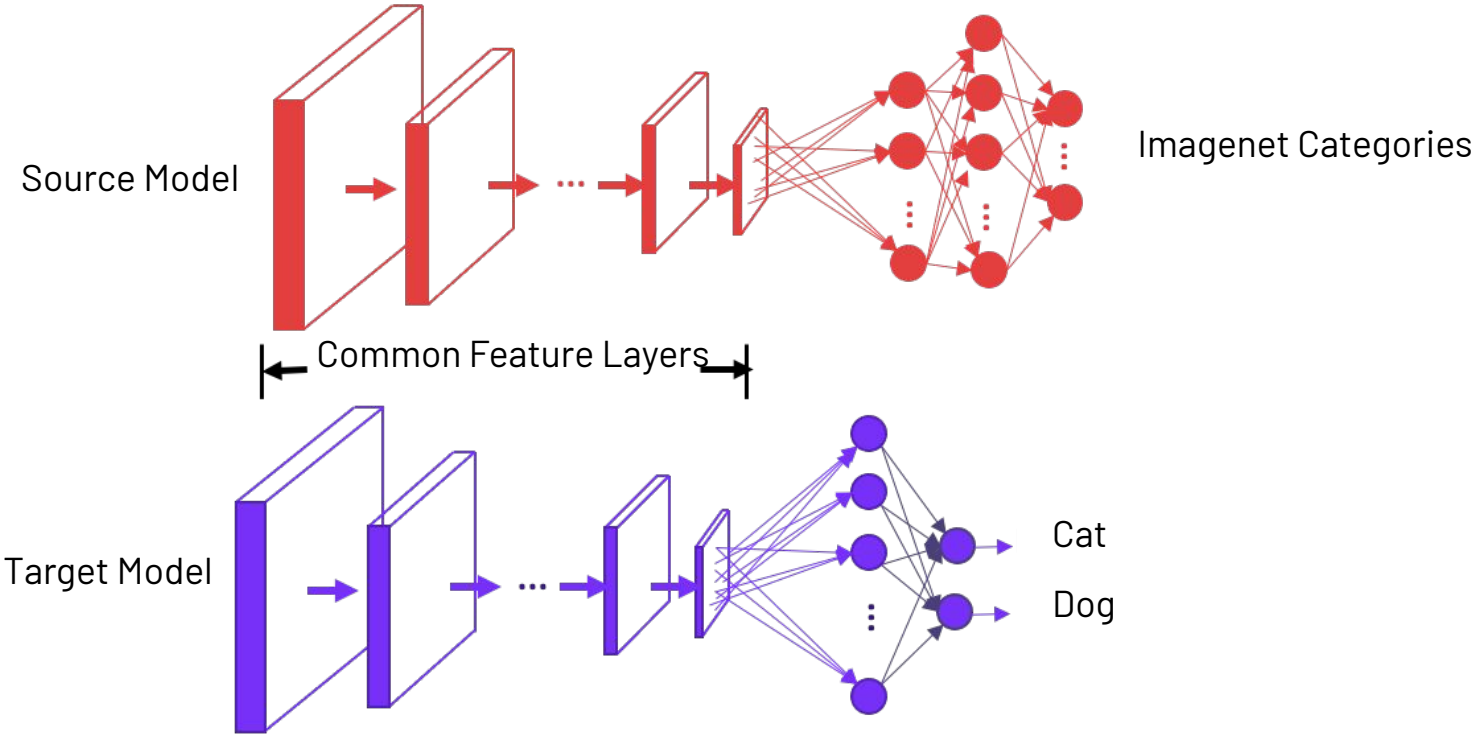
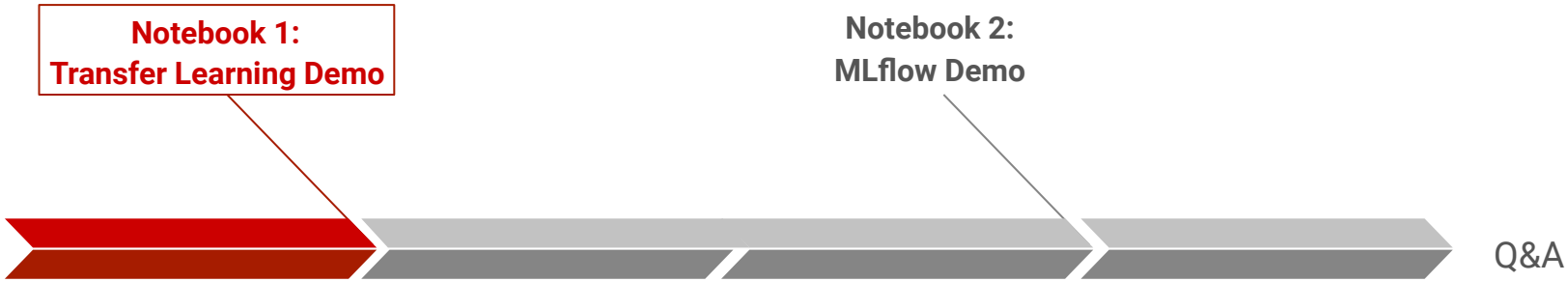


Image source

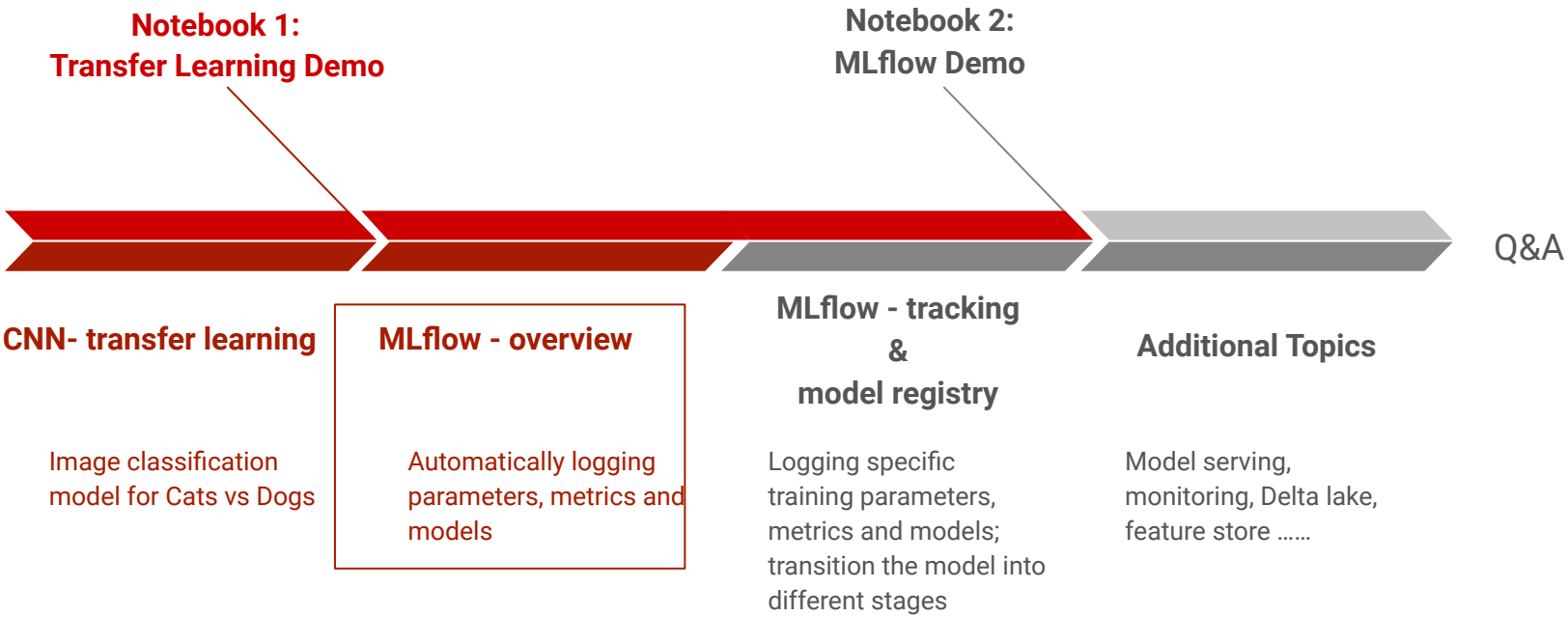
Outline


Notebook 1:
Transfer Learning Demo
Notebook 2:
MLflow Demo
Q&A
CNN- transfer learning
MLflow - overview
MLflow - tracking
&
model registry
Additional Topics
Image classification model for Cats vs Dogs
Automatically logging parameters, metrics and models
Logging specific training parameters, metrics and models; transition the model into different stages
Model serving, monitoring, Delta lake, feature store ……

Notebook Demo 1: Transfer Learning

github.com/feifeiwww

Outline


Notebook 1:
Transfer Learning Demo
Notebook 2:
MLflow Demo
Q&A
CNN- transfer learning
Image classification model for Cats vs Dogs
MLflow - overview
Automatically logging parameters, metrics and models
MLflow - tracking & model registry
Logging specific training parameters, metrics and models; transition the model into different stages
Additional Topics
Model serving, monitoring, Delta lake, feature store ……

MLflow is the most successful MLOps project

Sci-kit Learn: **34M** / month

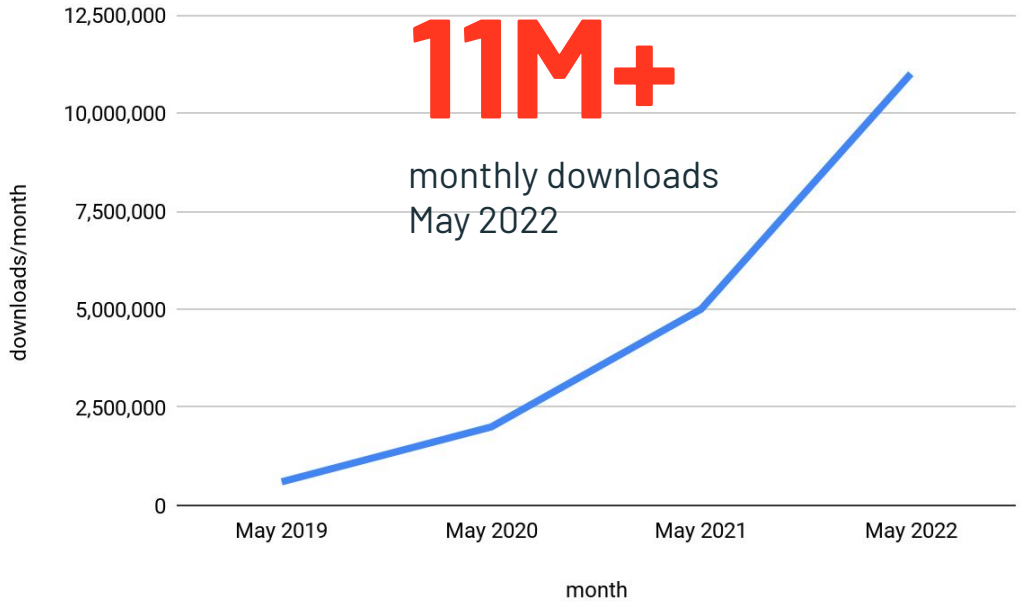
TensorFlow: **19M** / month

MLflow: **11M** / month

PyTorch: **10M** / month

...

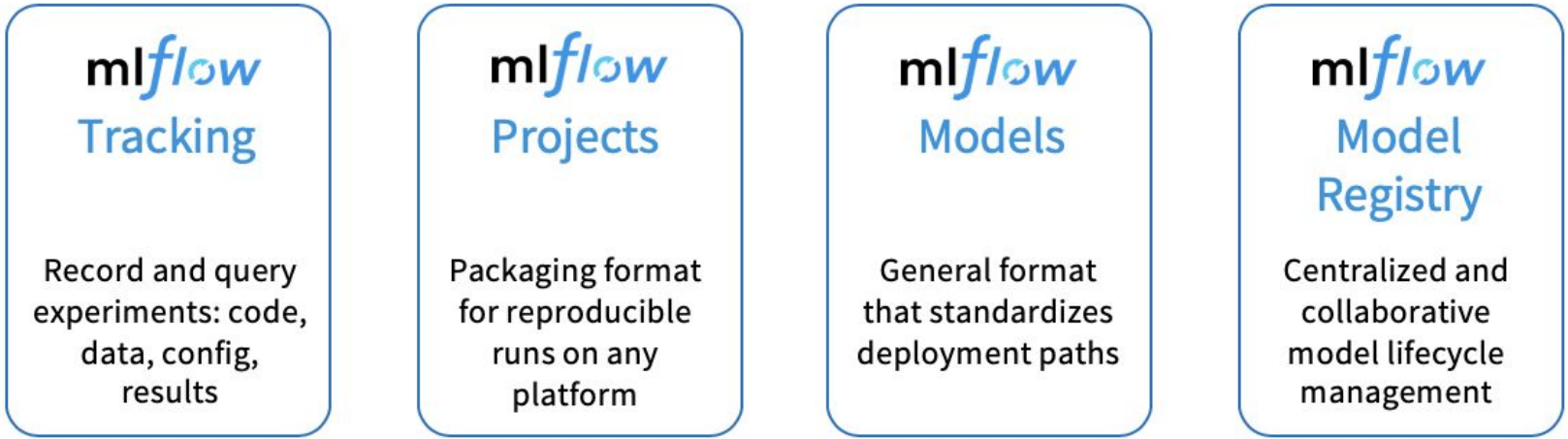
TFX: **319K** / month



MLflow

An open source platform for the machine learning lifecycle

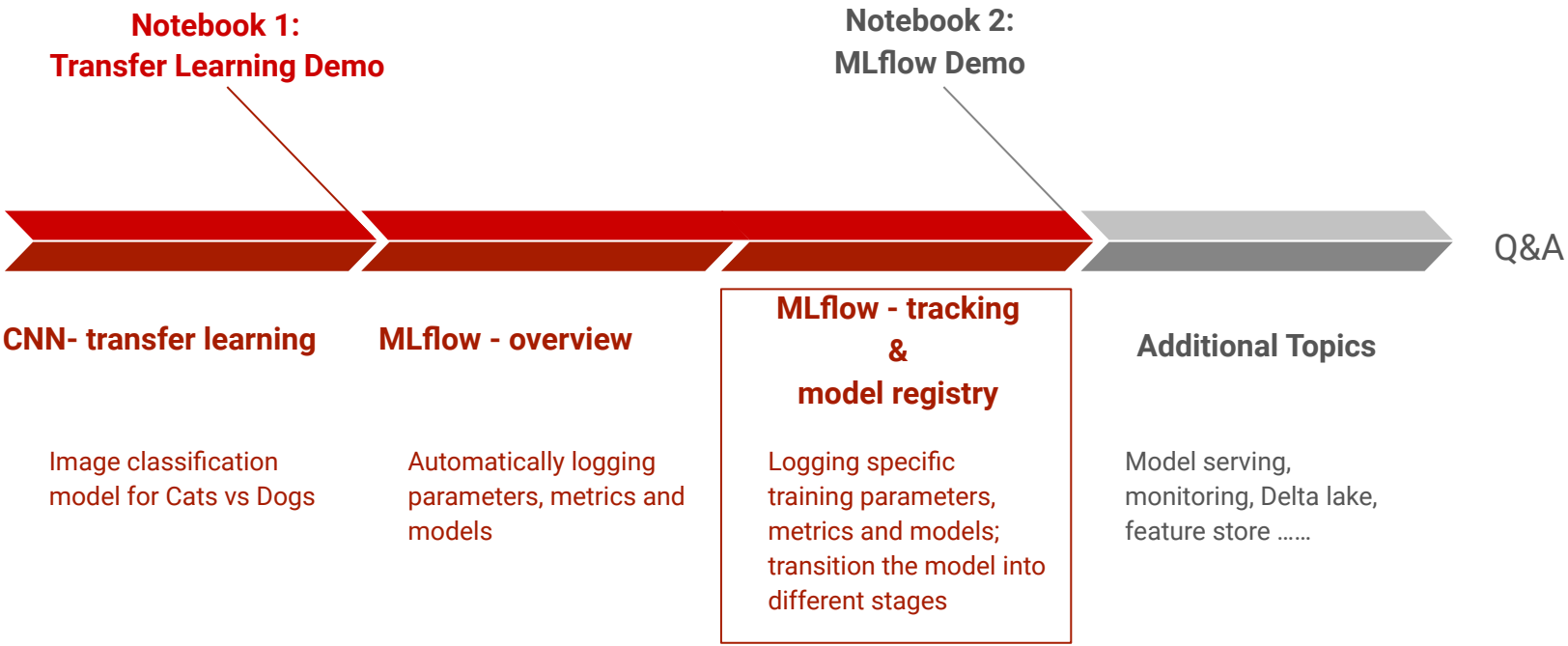
pip install mlflow



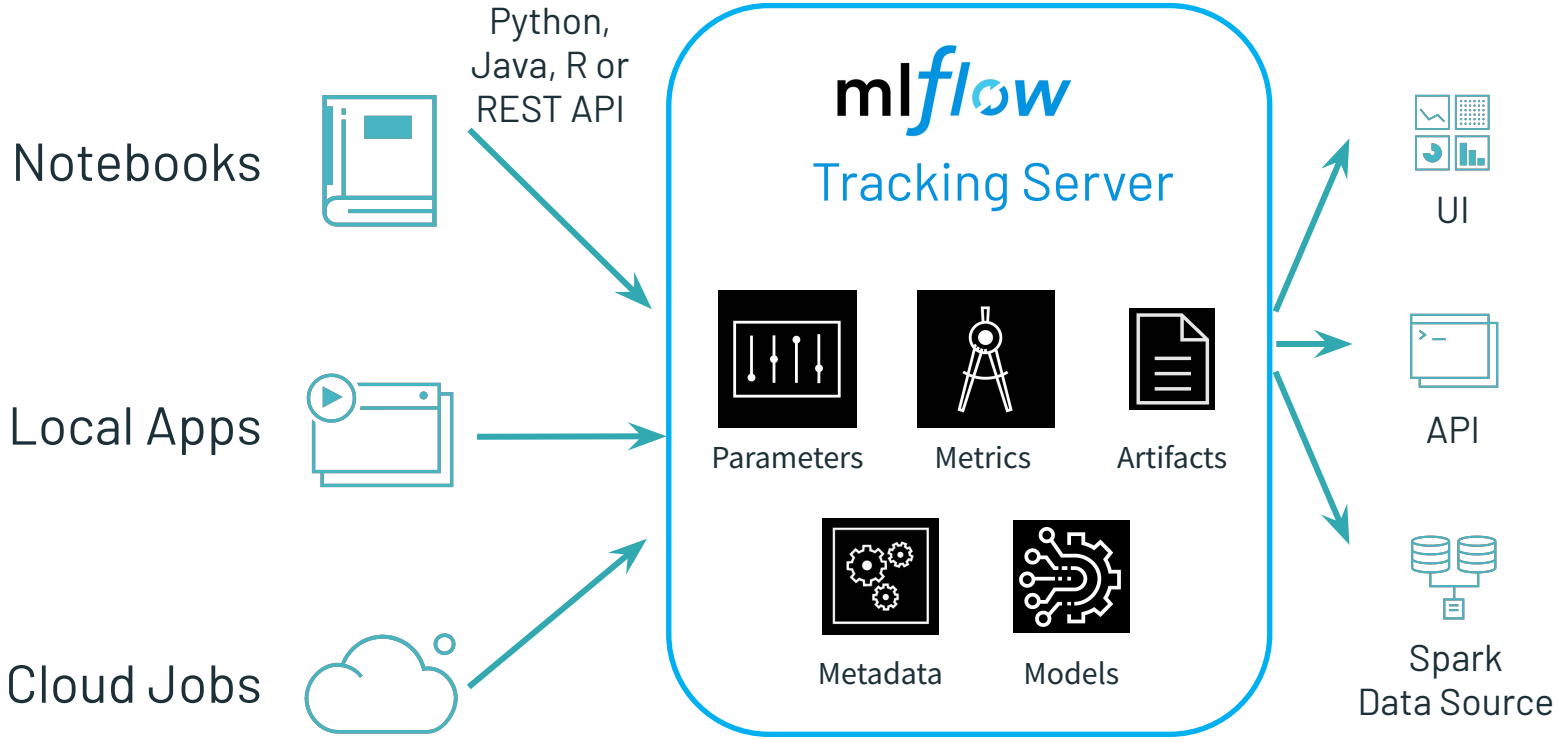
Notebook Demo 1: Transfer Learning

github.com/feifeiwww

Outline


Notebook 1:
Transfer Learning Demo
Notebook 2:
MLflow Demo
Q&A
CNN- transfer learning
MLflow - overview
MLflow - tracking
&
model registry
Additional Topics
Image classification model for Cats vs Dogs
Automatically logging parameters, metrics and models
Logging specific training parameters, metrics and models; transition the model into different stages
Model serving, monitoring, Delta lake, feature store ……

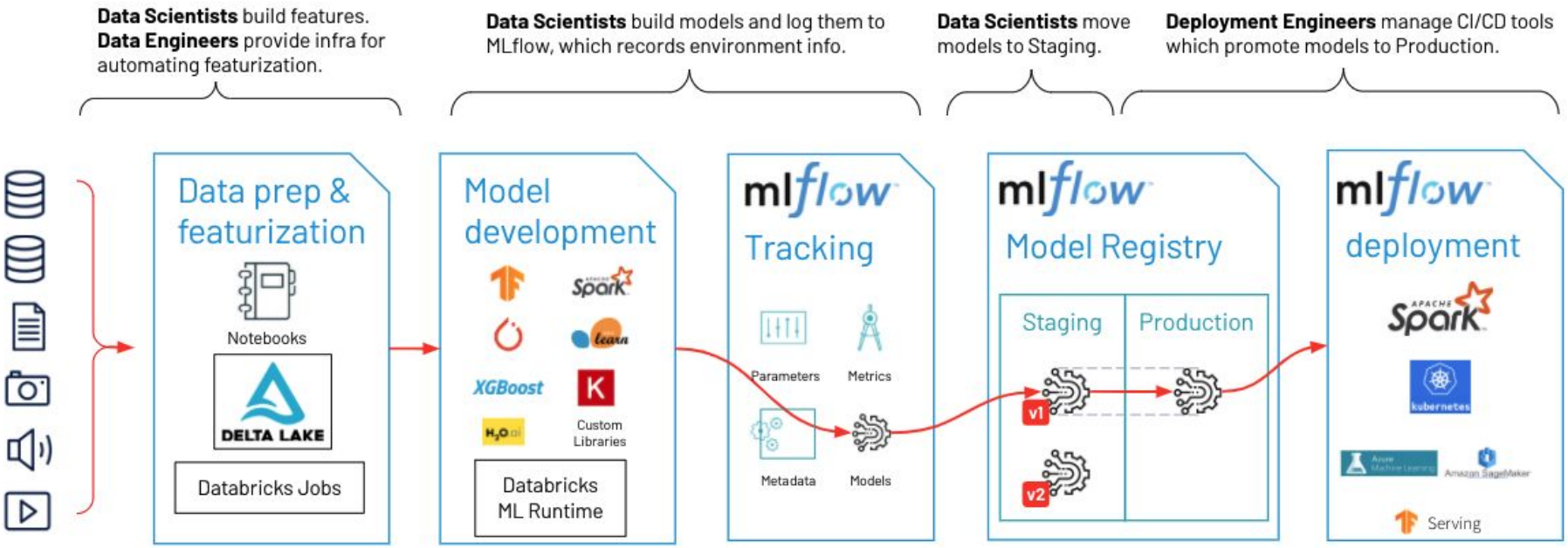
MLflow Tracking



MLflow Model Registry

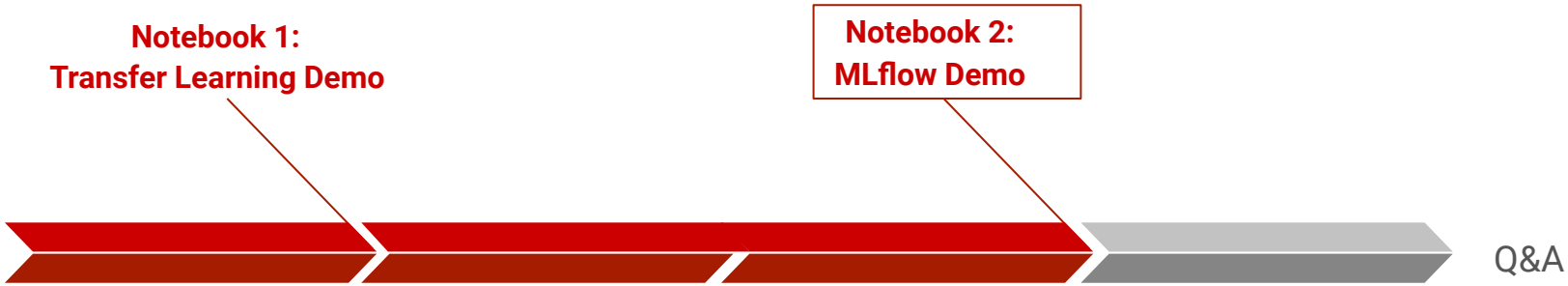
- **Central Repository**: Unique named registered models for discovery across data teams
- **Model Registry Workflow**: Provides UI and API for registry operations
- **Model Versioning**: Allow multiple versions of model in different stages
- **Model Stages**: Allow stage transition: none, staging, production, or archived
- **CI/CD Integration**: Easily load a specific version for testing and inspection
- **Model Lineage**: Provides model description, lineage and activities

The Full ML Lifecycle



databricks

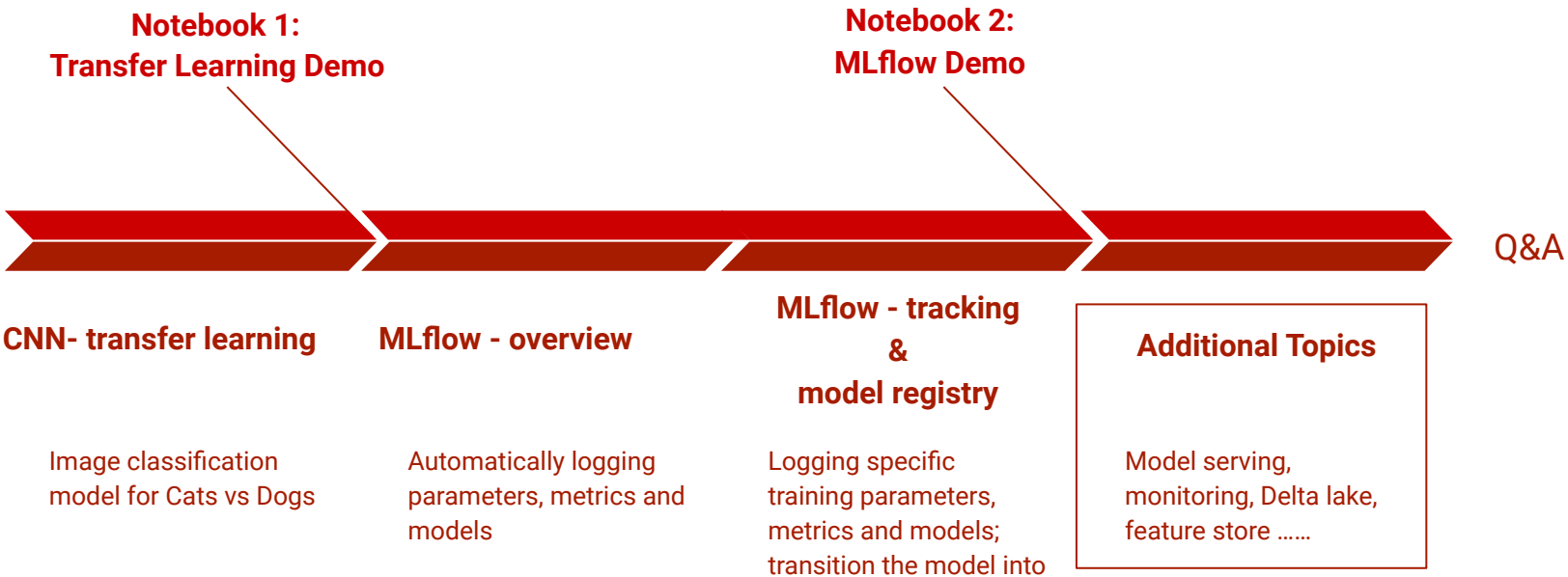
Outline


Notebook 1:
Transfer Learning Demo
Notebook 2:
MLflow Demo
Q&A
CNN- transfer learning
MLflow - overview
MLflow - tracking
&
model registry
Additional Topics
Image classification model for Cats vs Dogs
Automatically logging parameters, metrics and models
Logging specific training parameters, metrics and models; transition the model into different stages
Model serving, monitoring, Delta lake, feature store ……

Notebook Demo 2: MLflow

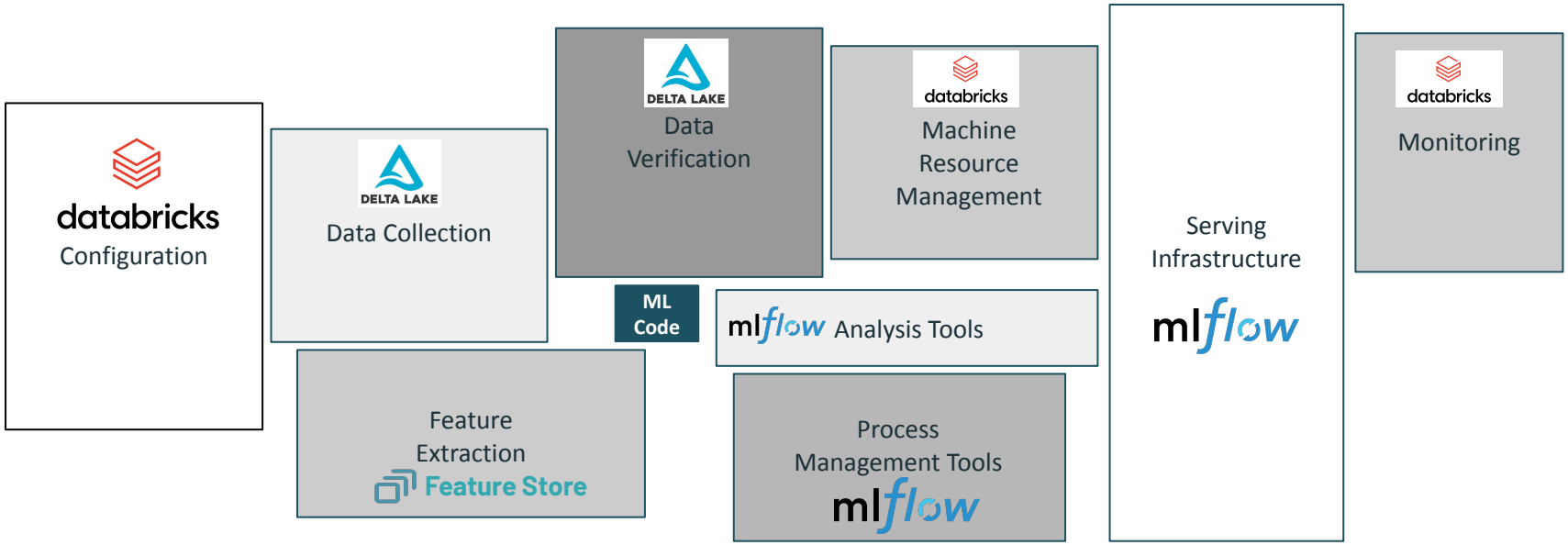
databricks

Outline


Notebook 1:
Transfer Learning Demo
Notebook 2:
MLflow Demo
Q&A
CNN- transfer learning
Image classification model for Cats vs Dogs
MLflow - overview
Automatically logging parameters, metrics and models
MLflow - tracking & model registry
Logging specific training parameters, metrics and models; transition the model into different stages
Additional Topics
Model serving, monitoring, Delta lake, feature store ……

Hardest Part of ML isn't ML

"Hidden Technical Debt in Machine Learning Systems," Google NIPS 2015



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.

References

- **CNN transfer learning tensorflow example:**
  https://www.tensorflow.org/tutorials/images/transfer_learning
- **Sign up for Databricks community edition:**
  https://docs.databricks.com/getting-started/community-edition.html
- **Databricks community edition login:**
  https://community.cloud.databricks.com/login.html
- **Get started with Databricks as a data scientist:**
  https://docs.databricks.com/getting-started/quick-start.html
- **Image classification on ImageNet state of the art:**
  https://paperswithcode.com/sota/image-classification-on-imagenet
- **Data lakehouse:**
  https://databricks.com/blog/2021/04/26/reproduce-anything-machine-learning-meets-data-lakehouse.html
- **My Github repo:**
  https://github.com/feifeiwww/20220726_Databricks_Demo_Transfer_Learning_with_MLflow

Special thanks to:
James Weaver
Brooke Wenig
Conor Murphy

Thank you! Questions?

feifei.wang@databricks.com
github.com/feifeiwww