

# From Model(s) to Credit Scoring System

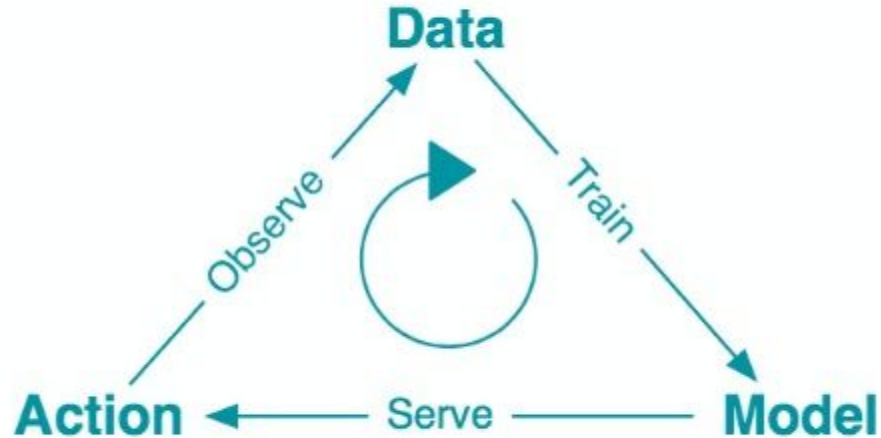
Akezyt Janedittakarn

Midas AI

[akezyt@midas.ai](mailto:akezyt@midas.ai)

<https://www.midas.ai>

# Machine Learning Life Cycle



Credit: Operationalizing Machine Learning Using GPU Accelerated, In-Database Analytics

# Outline

- Data Pipeline
- Model Training
- Model Deployment
- Measurement
  - Real traffic
  - Feedback loop

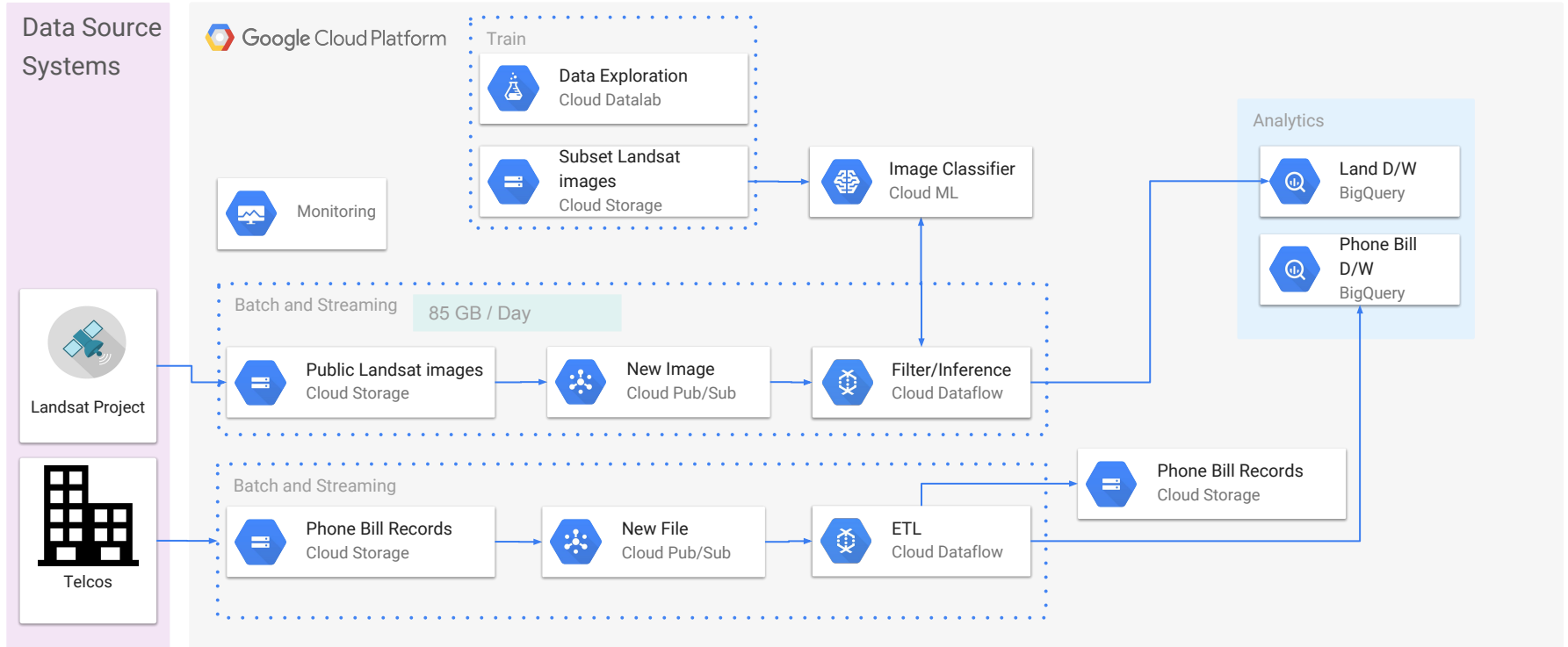
# Data Pipeline

- Most data need cleansing and transformation before ML training
- Multiple stages and continuously happening -> Pipeline
- Usually end result is data warehouse
- Key challenges:
  - Batch and stream
  - Infrastructure management
  - Unstructured data
- Sometimes you need to build ML to generate features for other ML

# Data Sources - Credit Scoring

- Phone bill payment record
  - Already structured data
  - ETL into data warehouse such as BigQuery
- 
- Landsat satellite images
  - Used for creating history of flood/drought in the past
  - Unstructured and need to be processed by image classifier
  - Existing 15 M images
  - 85 GB new images every day
- 
- Other sources such as demographics or weather can be used too

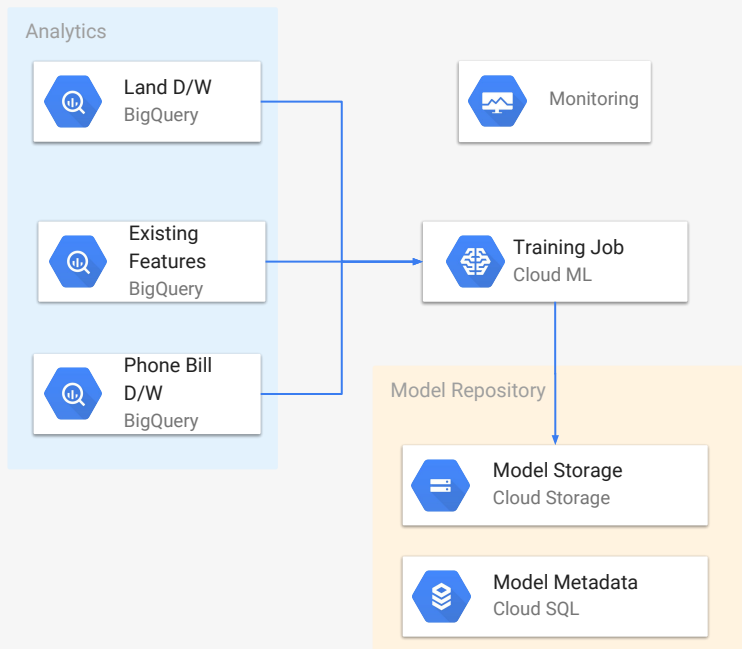
# Data Pipeline Architecture for Credit Scoring



# Train and Keep Track of Model Metadata

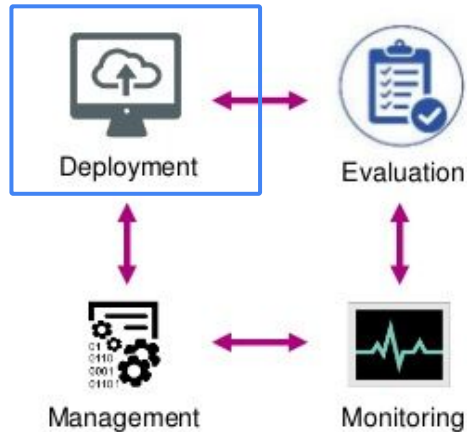
- Locally train or train on cloud
- Measure offline metric
- Record training parameter, data used, and metric in model repository
- In addition to model metric we need to consider:
  - Speed of model training and inference
  - Infrastructure constraints to train model with real dataset
  - Incremental training possible?

# Model Training





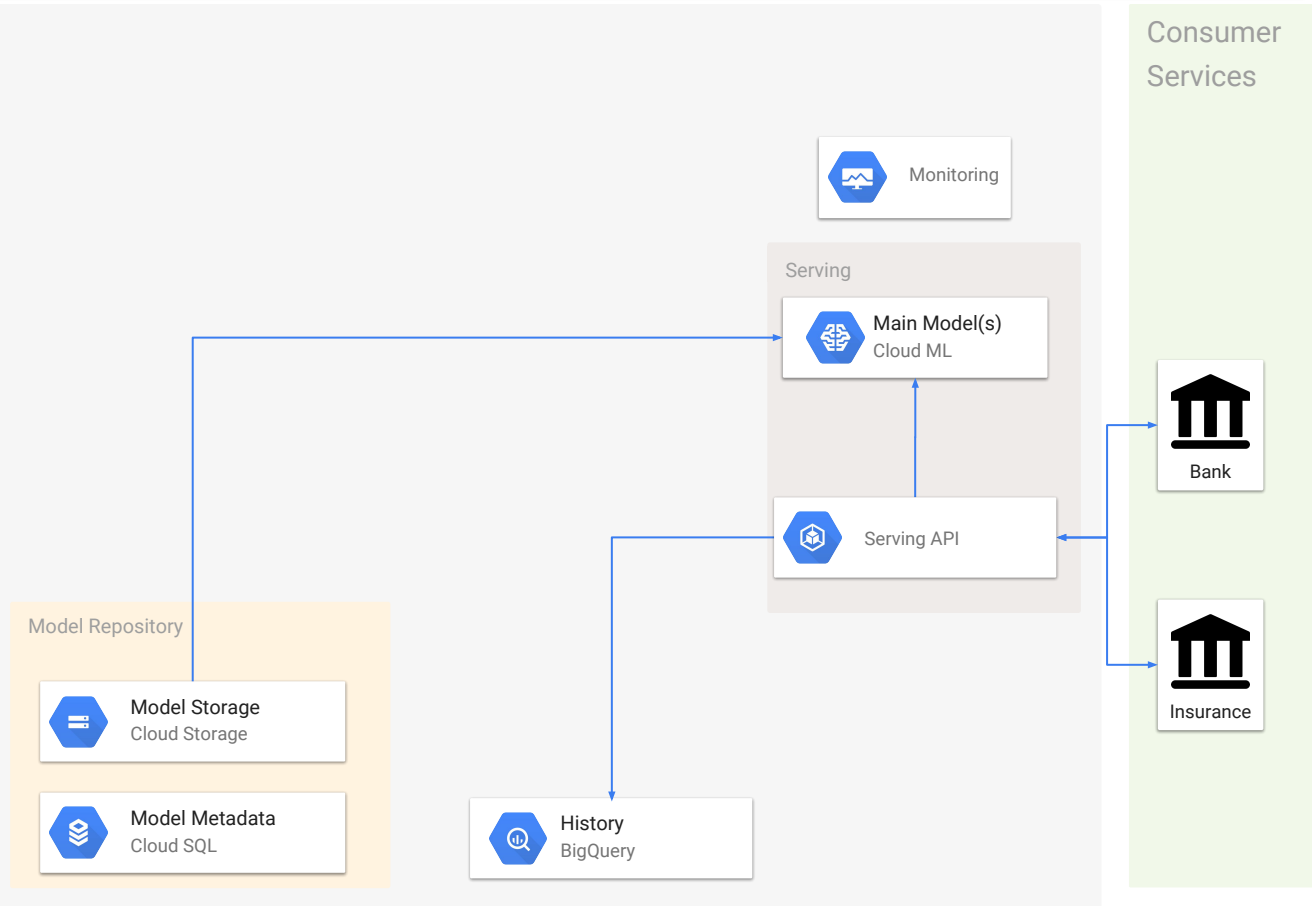
# Model Deployment



- Use model to give result to user
- Typically REST API
- Implementation
  - Custom web service
  - TensorflowServing
  - Cloud ML Engine
- **Serving API on top**

Credit - Production and Beyond: Deploying and Managing Machine Learning Models

# Deployment



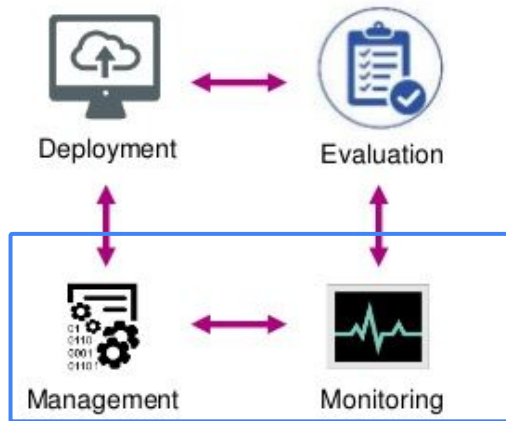
# Model Evaluation



- Business Metrics != Model Metrics
- Keep track which model/version served which requests
- Compare between multiple models

Credit - Production and Beyond: Deploying and Managing Machine Learning Models

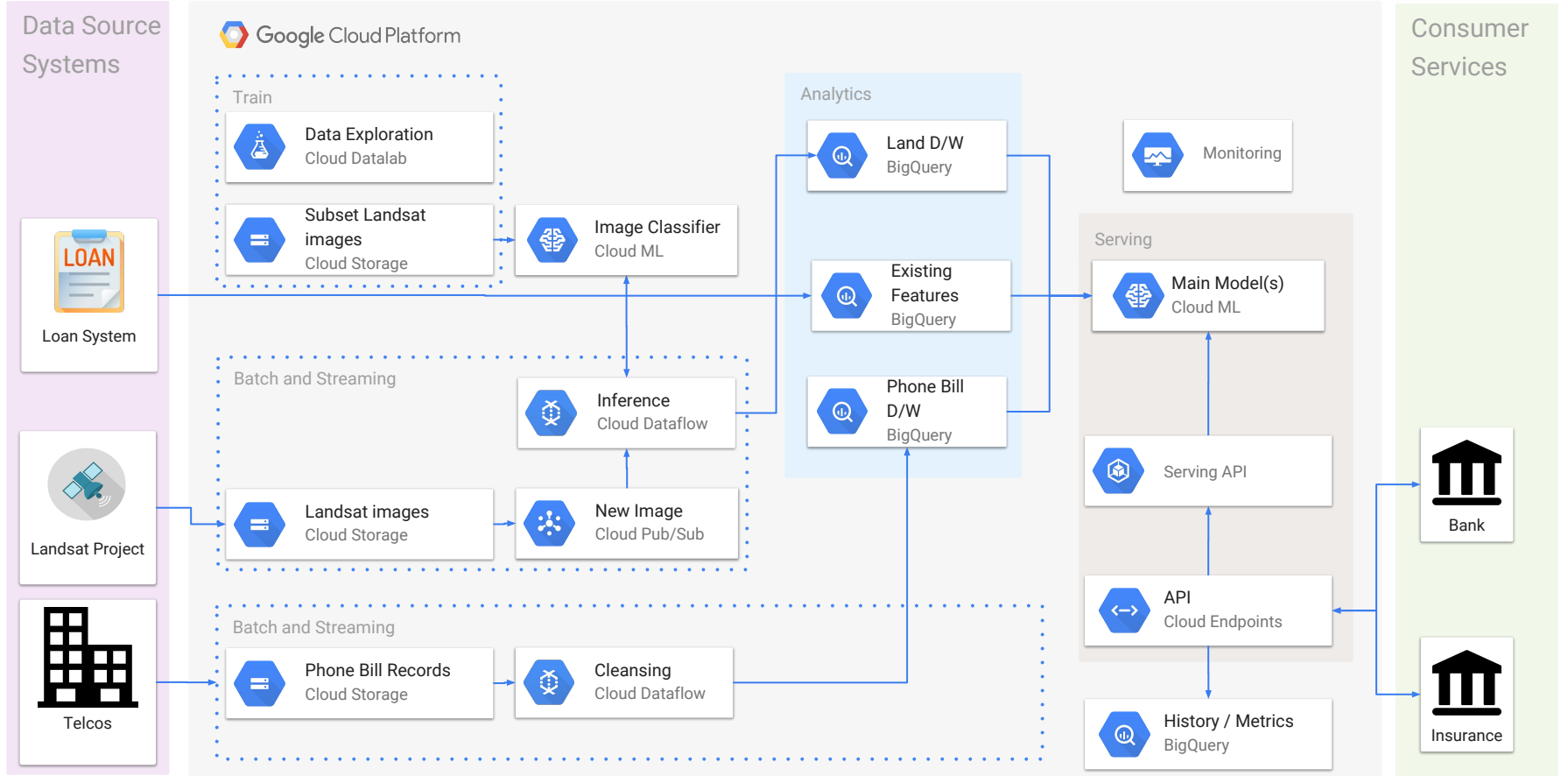
# Model Monitoring & Management



- Monitor metrics overtime
- React to feedback
  - Update model?
- Basic health of model and overall system

Putting Everything Together

# Architecture: Credit Scoring with Alternative Data



# Summary

- Model is central part of Machine-Learning Driven system
- But it takes solid system to deliver value to customer
- A lot of practices from system engineering can be borrowed to combine with unique characteristics of ML to deliver highest impact to business

# References

- Production and Beyond: Deploying and Managing Machine Learning Models - <https://www.slideshare.net/turi-inc/model-management>
- Machine Learning In Production - <https://www.slideshare.net/samkiller/machine-learning-in-production-60296902>
- Production ML Systems - <https://developers.google.com/machine-learning/crash-course/production-ml-systems>
- Data Science for Startups - <https://towardsdatascience.com/data-science-for-startups-data-pipelines-786f6746a59a>



Thank You