Deploying Machine Learning models in Production
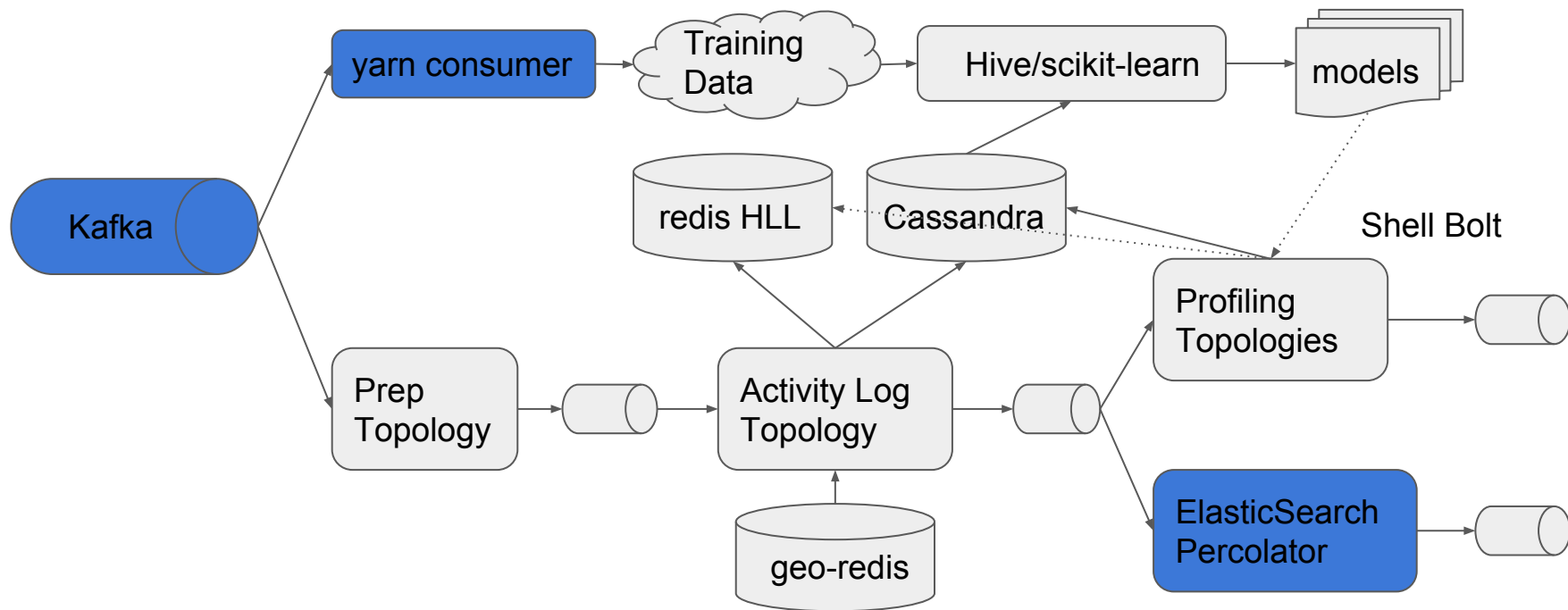
Balaji Ganesan

About me

- Engineer interested in Machine Learning, Information Retrieval, Big Data
- University of Arizona (MS), University of Madras (BE)
- Industry
  - ML - Near (adNear), Yahoo (DomainMatch), Iyottasoft (inagist.com)
  - IR - Guruji.com, Biggerboat.com (music search)
- Research
  - Collaborated with Yahoo Labs
  - Research Assistant to Dr. Suvrajeet Sen, MORE Lab, University of Arizona
  - Research Assistant to Dr. Chris Impey, Steward Observatory, University of Arizona

Blue in slides indicates external systems or work where author had no direct contribution.
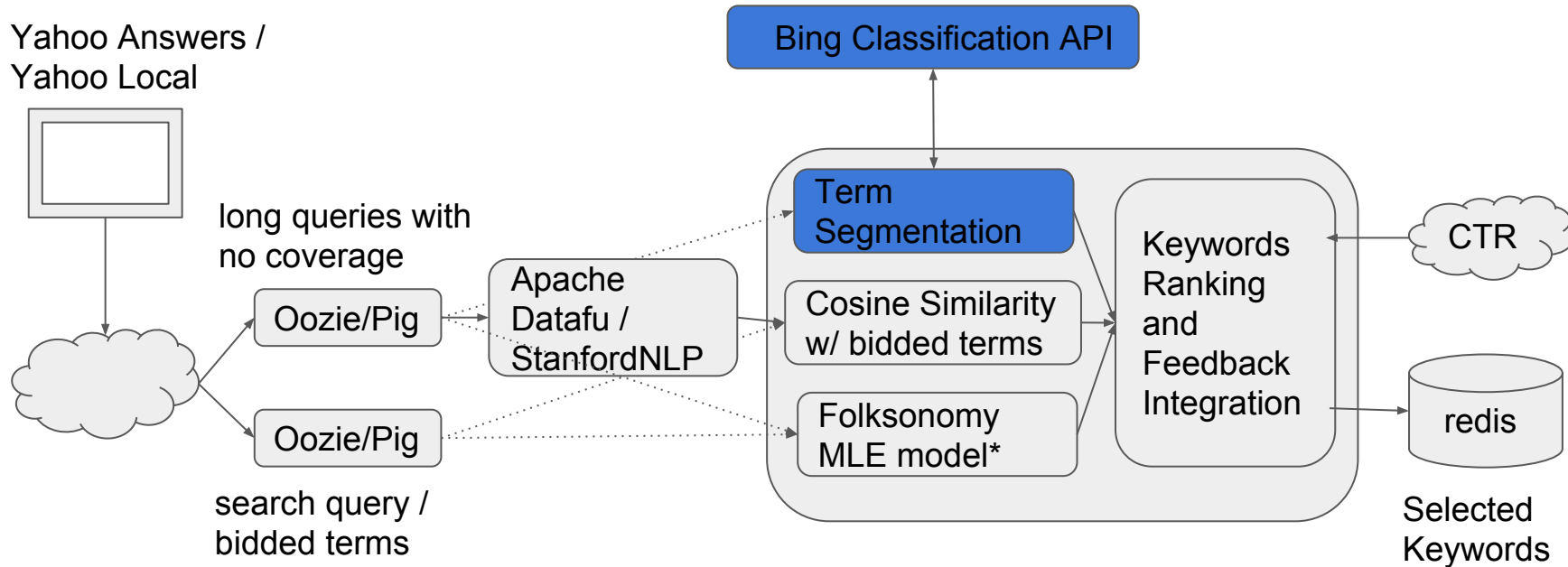
Overview

- Model Training on Single machine
  - Single or multi processes, out-of-core using Dask
- Model Training on Distributed Systems
  - IBM SystemML, Spark MLLib, Microsoft DMTK

- Model Deployment on Batch Systems
  - Pig UDFs, D2S and Ad Query Optimization systems at Yahoo DomainMatch
- Model Deployment on Streaming Systems
  - Python shell bolt on Storm, ML Pipeline at Near (adNear)
  - erlang native interface functions with ZeroMQ, ML Pipeline at Iyottasoft
- Model Deployed as a Service
  - Yahoo Content Analysis Platform, Bing Contextual Ads API

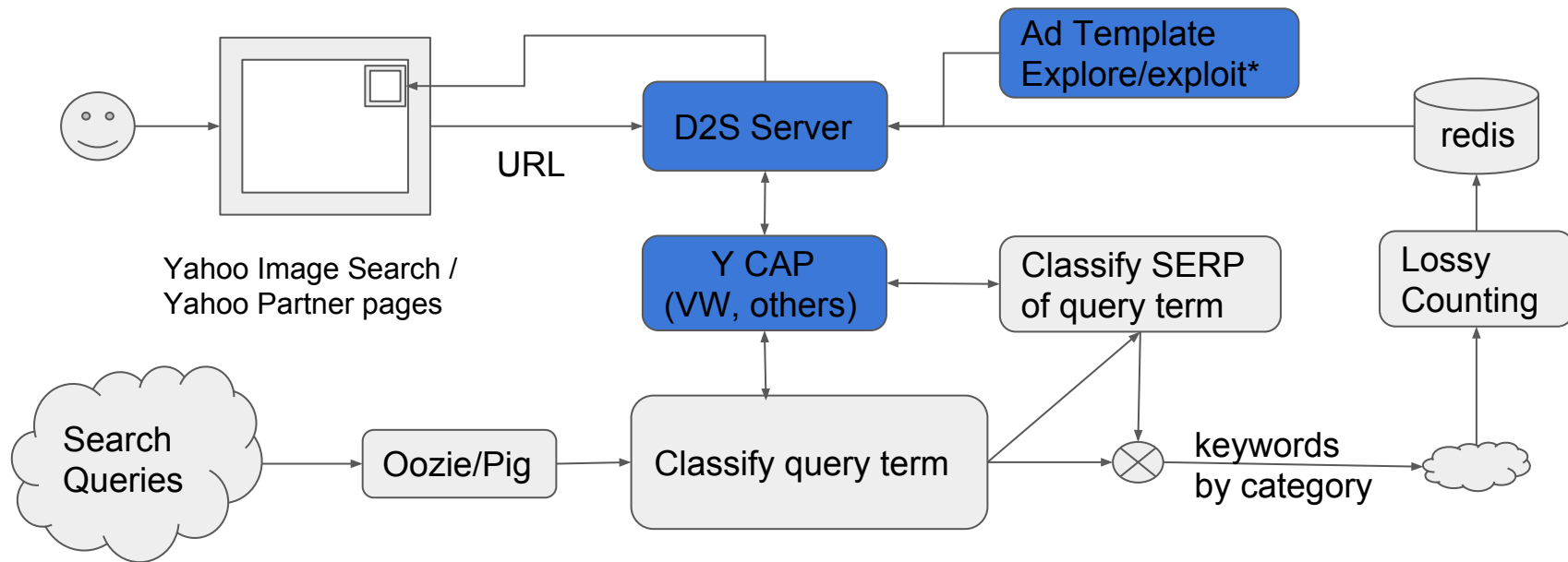# ML Pipeline at Near - model deployed on stream

Yahoo AdQO - model deployed on batch system

Yahoo Answers /
Yahoo Local

long queries with
no coverage

Oozie/Pig

Apache
Datafu /
StanfordNLP

search query /
bidded terms

Oozie/Pig

Bing Classification API

Term
Segmentation

Cosine Similarity
w/ bidded terms

Folksonomy
MLE model*

Keywords
Ranking
and
Feedback
Integration

CTR

redis

Selected
Keywords

AdQO - Ad Query Optimization

* Work done with Avihai Mejer and Dr. Dan Pelleg at Yahoo Labs Haifa; based on "Folksonomy-Based Term Extraction for Word Cloud Generation" paper by Dr. David Carmel at IBM Research Haifa Lab
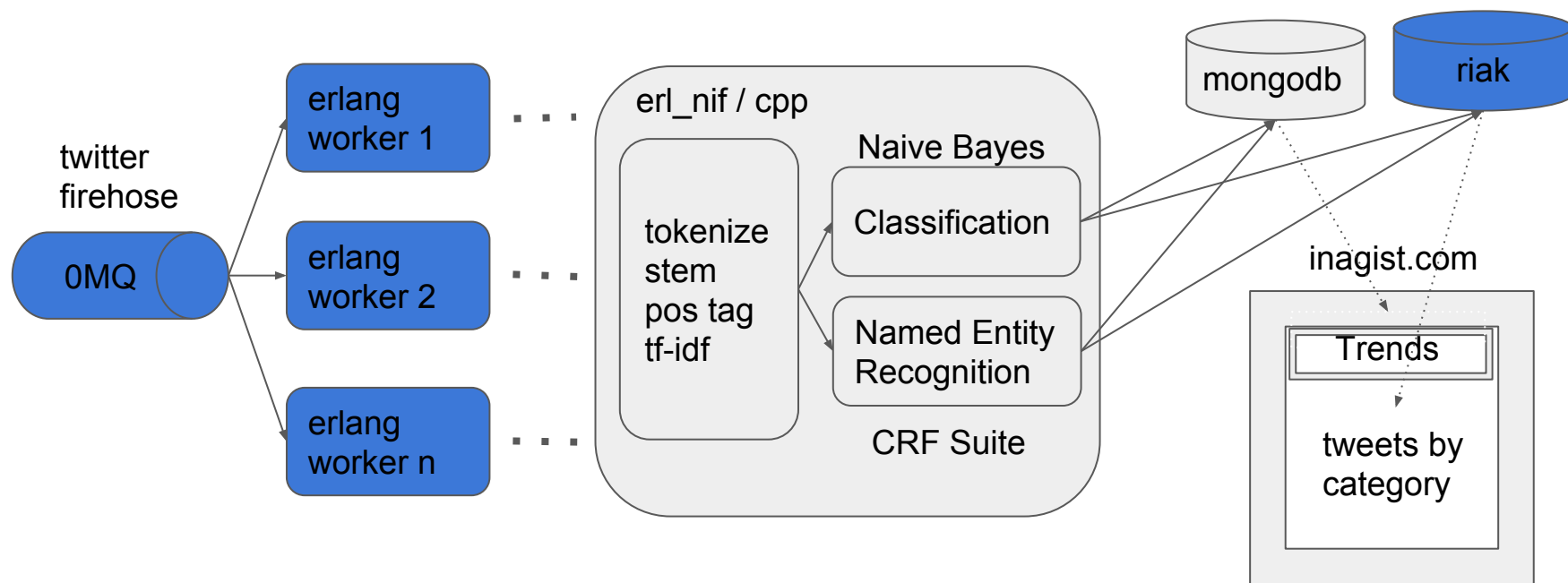
Yahoo CAP - model deployed as a service

CAP - Content Analysis Platform, VW - Vowpal Wabbit, D2S - Display to Search ad, SERP - Search Engine Result Page

* Explore/Exploit - Multi armed bandit - work done by Dr. Shivaram Kalyanakrishnan and Amit Kumar Gupta.

ML Pipeline at Iyottasoft - model deployed on stream

Thank you!