

# Constructing Operational Big-Data Models

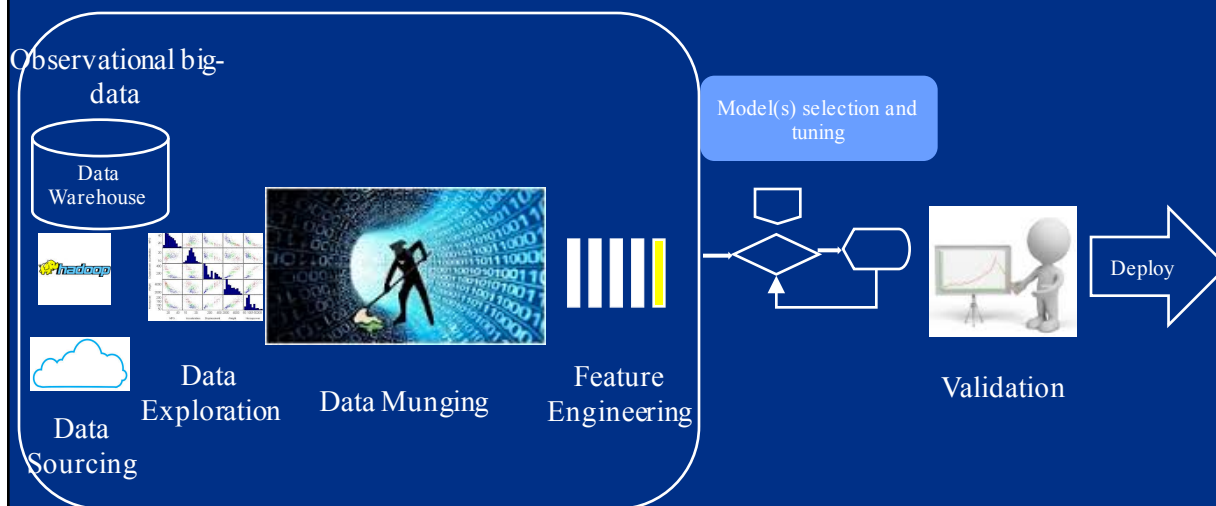
Nachum Shacham, PhD  
Director, Data Science  
PayPal

5/18/2016



© 2015 PayPal Inc. All rights reserved.

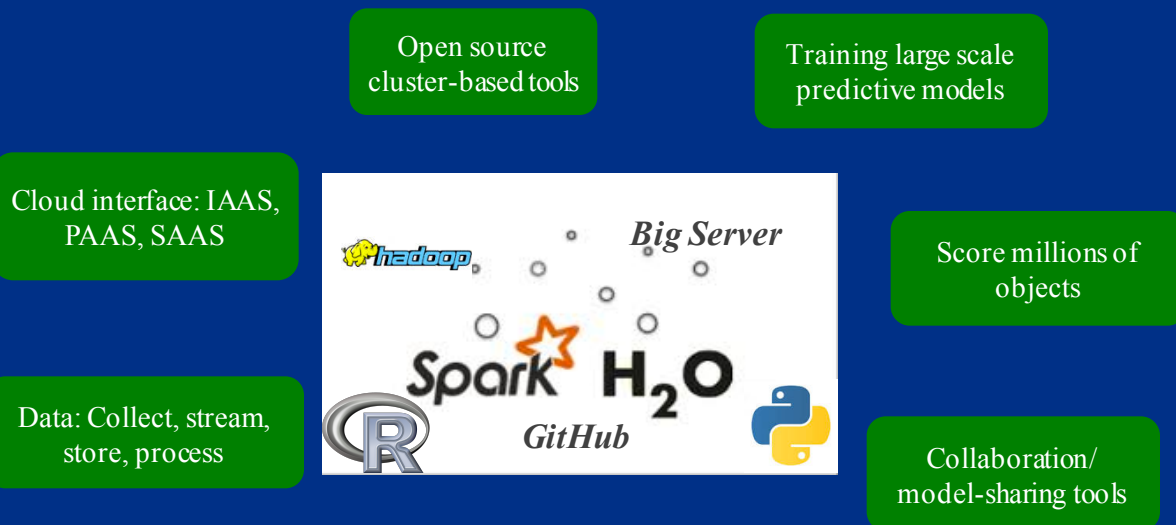
## Predictive Modeling With Big Data



PayPal

2

## Tools & Infrastructure for Data and Modeling



## Modeling: Predicting Individual Users

- Big predictive models are complex
  - Trained on historical data to score future cases
  - Hundreds/thousands of features
  - Millions of cases
- Machine-learning algorithms have been upgraded to operate efficiently on big data
  - Neural nets
  - Tree-based ensembles (Random Forest, GBM)
  - Clustering
- Do the model results work?

## How Good Is Your Model

- Performance Metrics:  $R^2$ , Error Rate(s)
  - Measure via cross validation
  - Based on historical data
- Cost of errors: False Positive, False Negative
  - Set by the users of the model
- Pre-production evaluation requires trust
- A/B testing for selecting alternatives



## Gain a Trust in the Model

- Model results: interpretable or black box
  - Can score reasoning be articulated
- Identify champion(s) who believe in the model
- Incentivize adoption of the model
- Avoid impression of positional bias
- Integrate gradually to prove results on a gradually increasing scale
- Repeatedly test and adjust
- Measure the benefits of actions based on the model

# THANK YOU



7