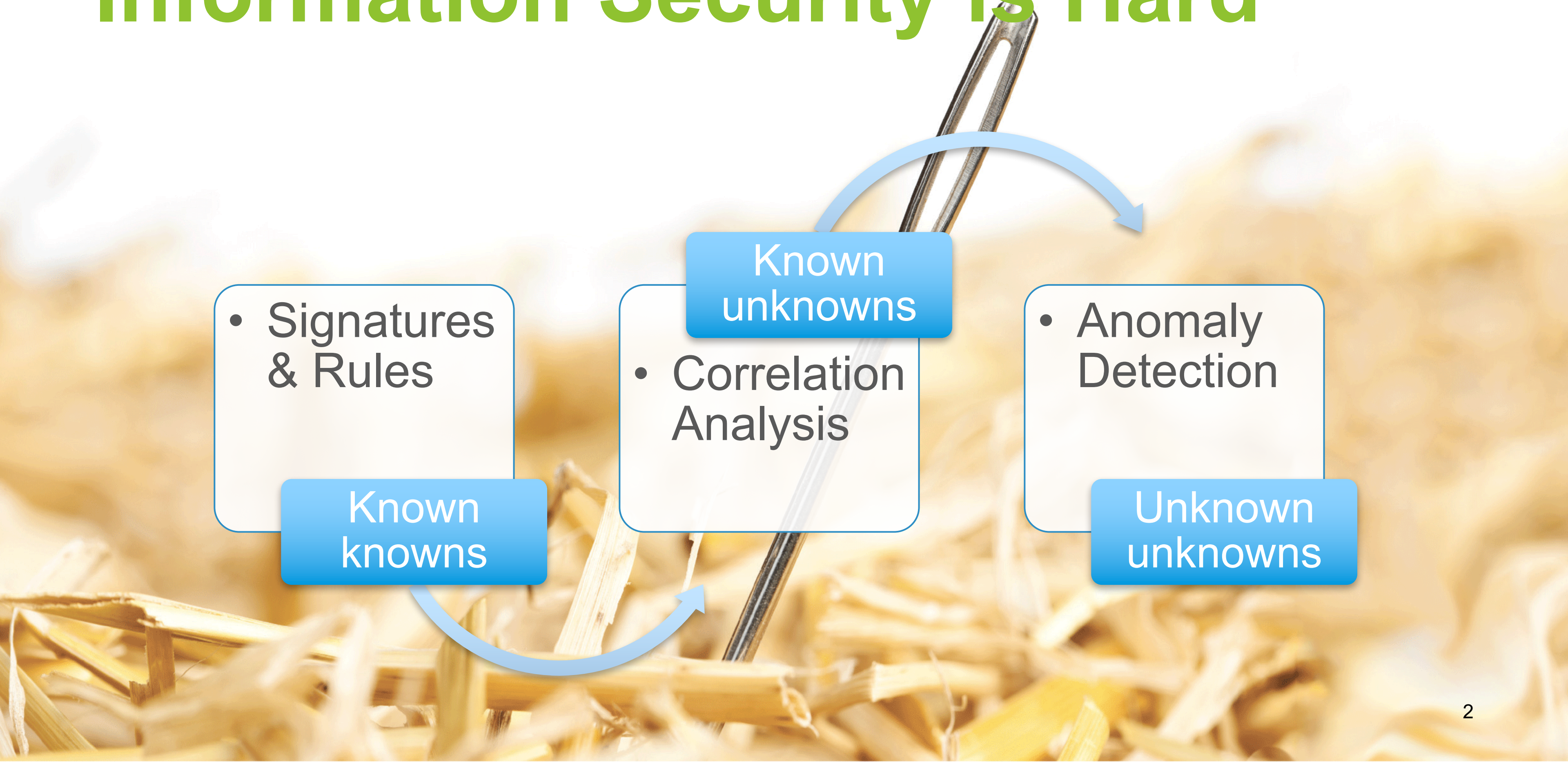
SPARK SUMMIT 2017

Needle in the Haystack

User Behavior Anomaly Detection for Platform Security

Wei Deng, Ping Yan

Information Security is Hard



- Signatures & Rules

Known knowns

- Correlation Analysis

Known unknowns

- Anomaly Detection

Unknown unknowns

User In-app Behavior

WHO

WHEN

WHERE

HOW

WHAT

Entities

client IP

timestamps

user agents

...

Derived

hour of day

day of week

geo country

...

Actions

logins

UI page views

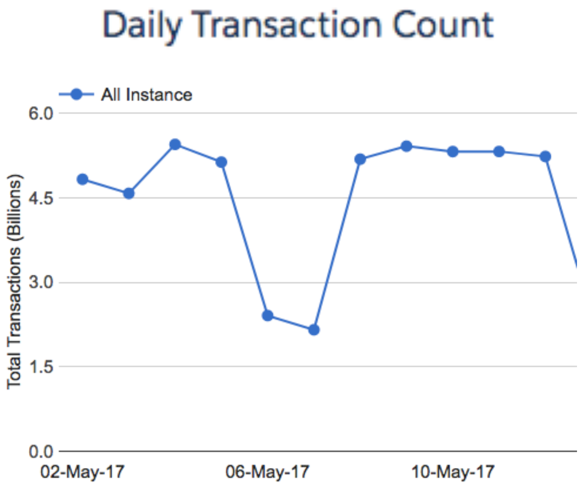
API calls

password reset

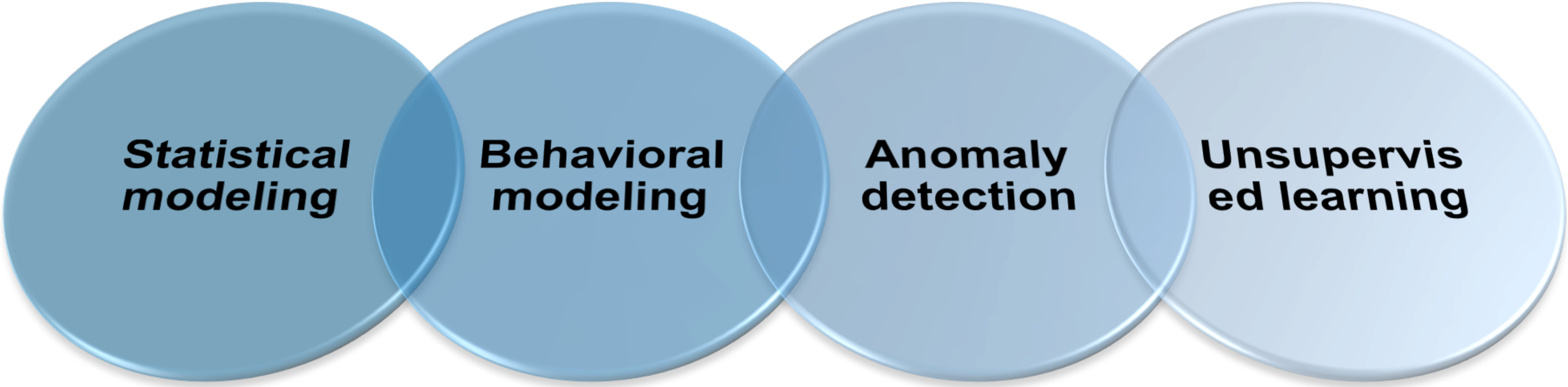
...

Challenges

- Size of data, speed to response
- Variability of threats
- Little to none ground truth
- Feature selection
- User behavior novelty



A Behavior Anomaly Detection Approach



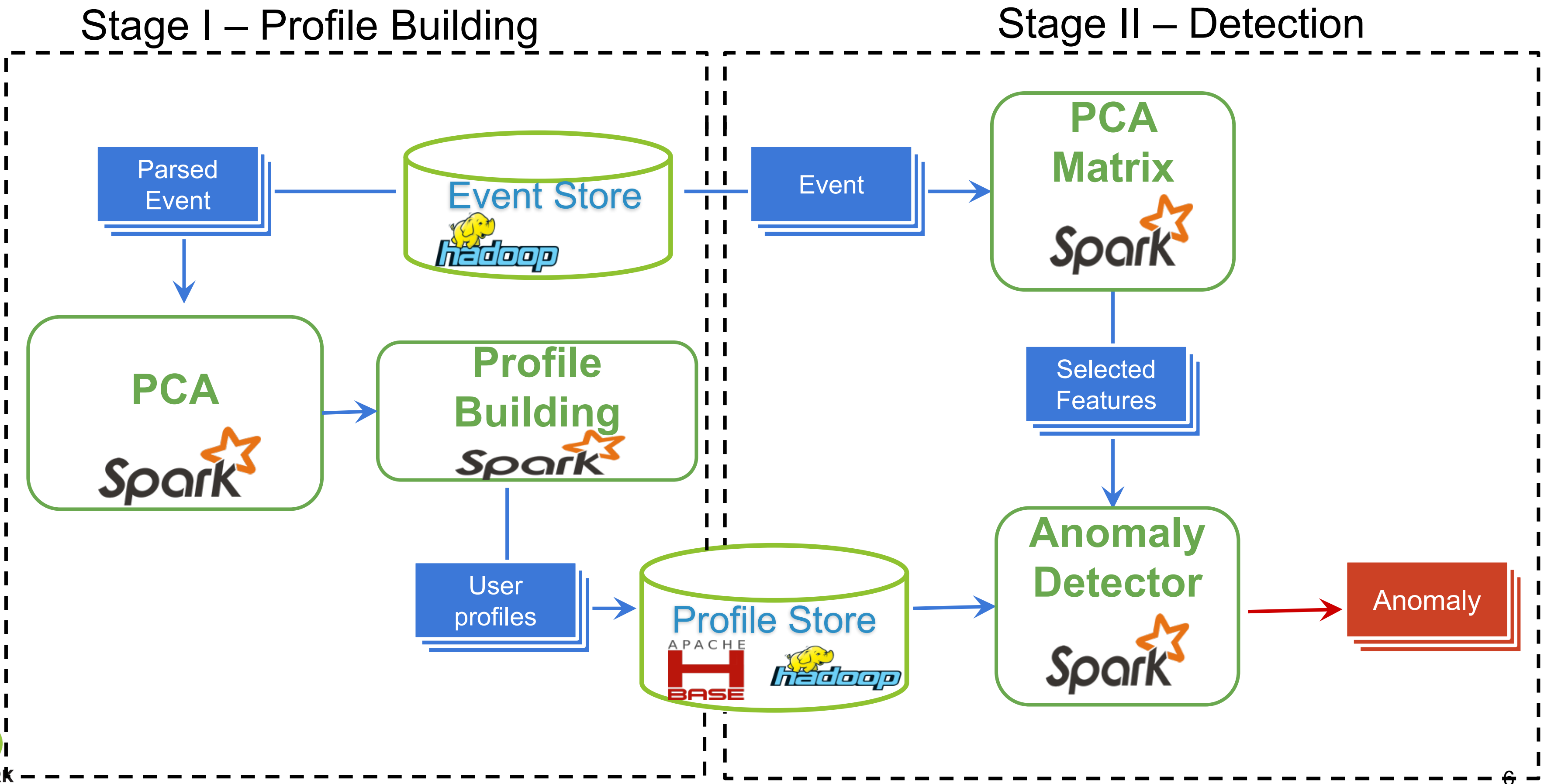
Statistical modeling

Behavioral modeling

Anomaly detection

Unsupervised learning

Anomaly Detection Engine



Stage I – Profile Building

Stage II – Detection

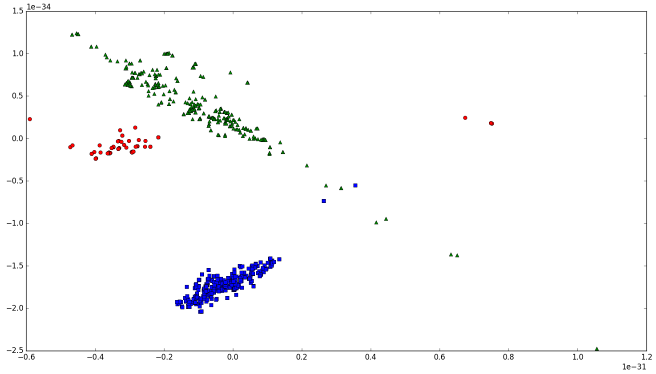
PCA

- High variance variables

  Representative of original data set

- Low variance variables

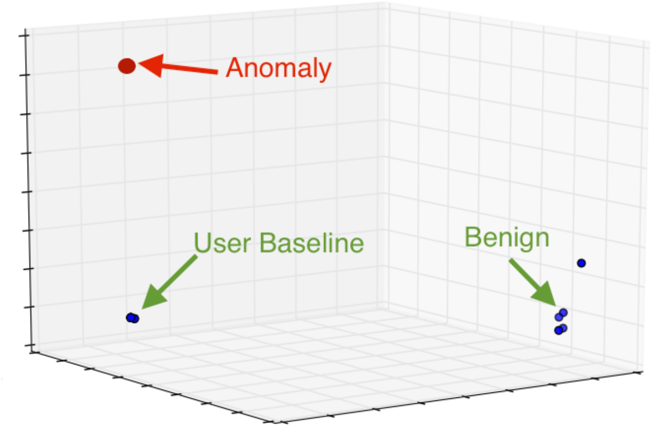
  Representative of users’ stable behavior



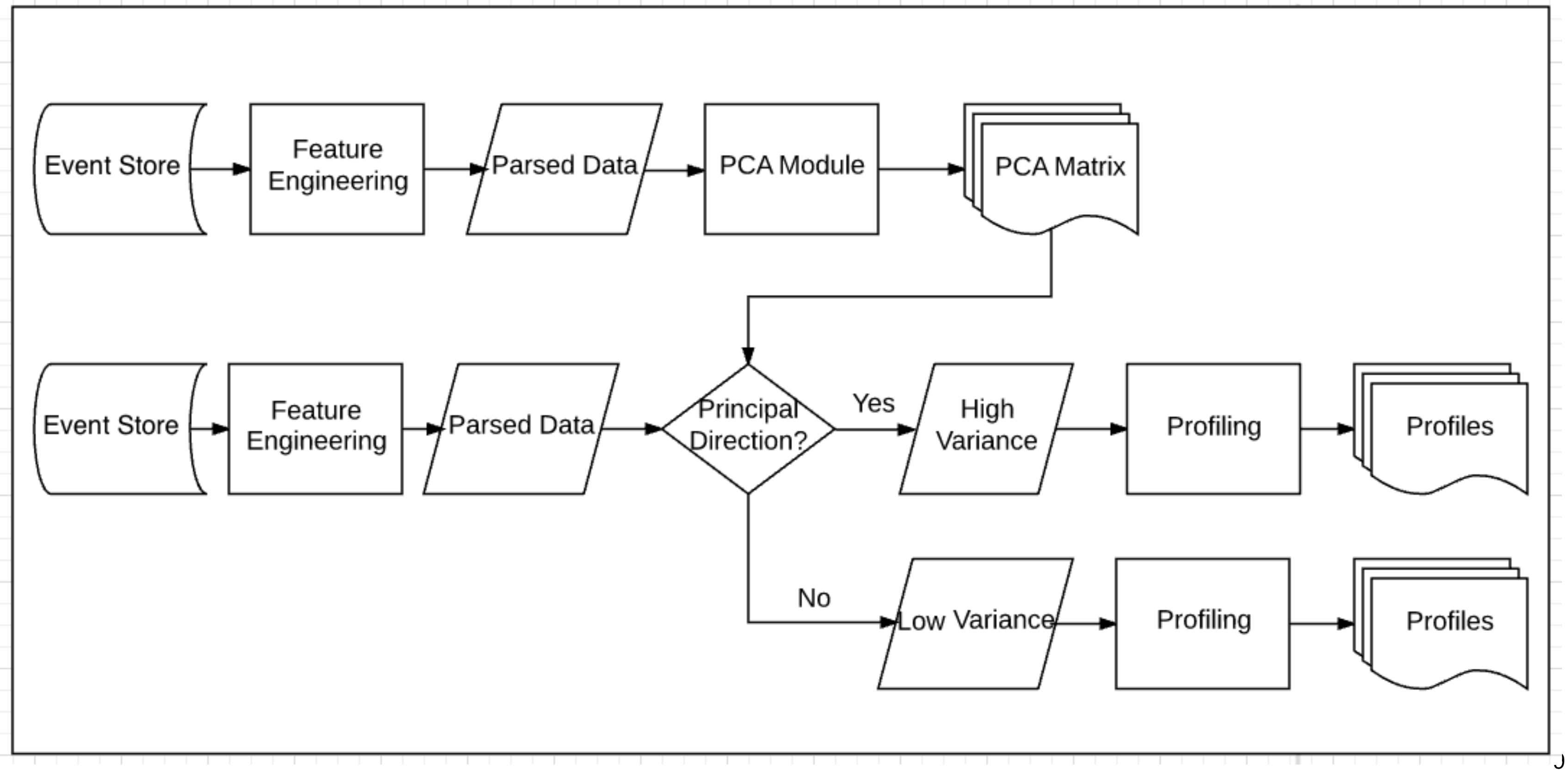
Profile Building

3 statistics to summarize each user's historical distribution

- User's behavior baseline
- Variance of user's behavior
- Legitimate non-typical behavior



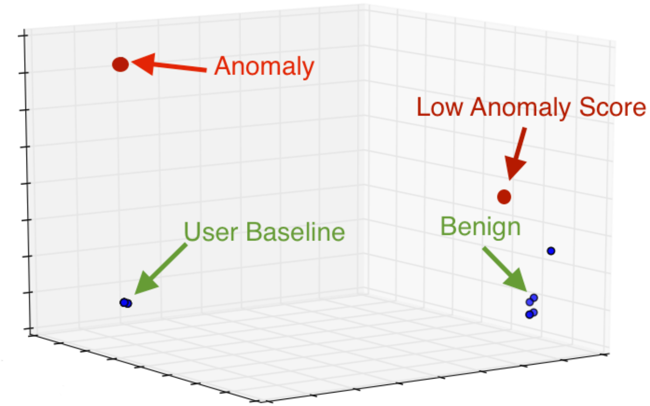
Profile Building



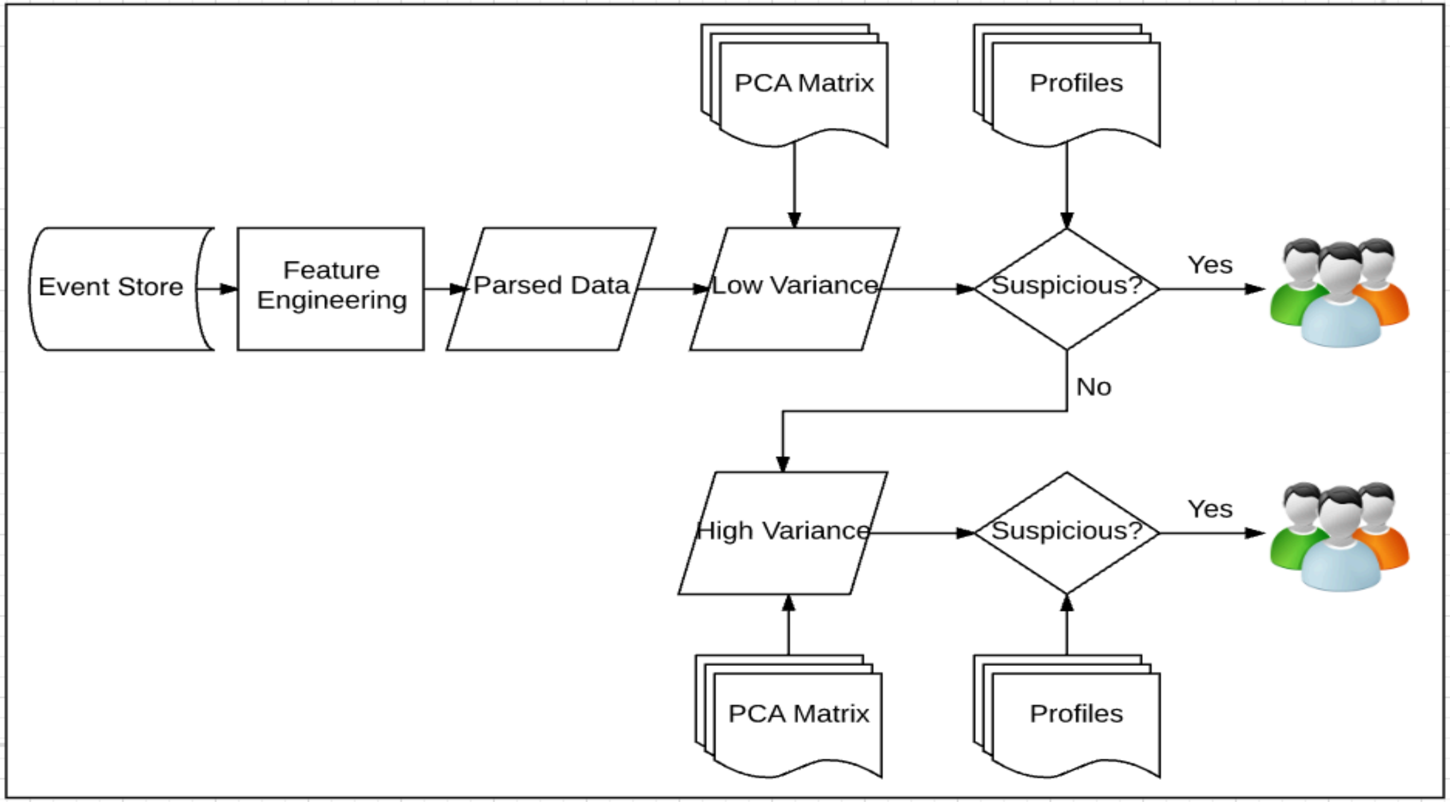
SPARK SUMMIT 2017

Detection

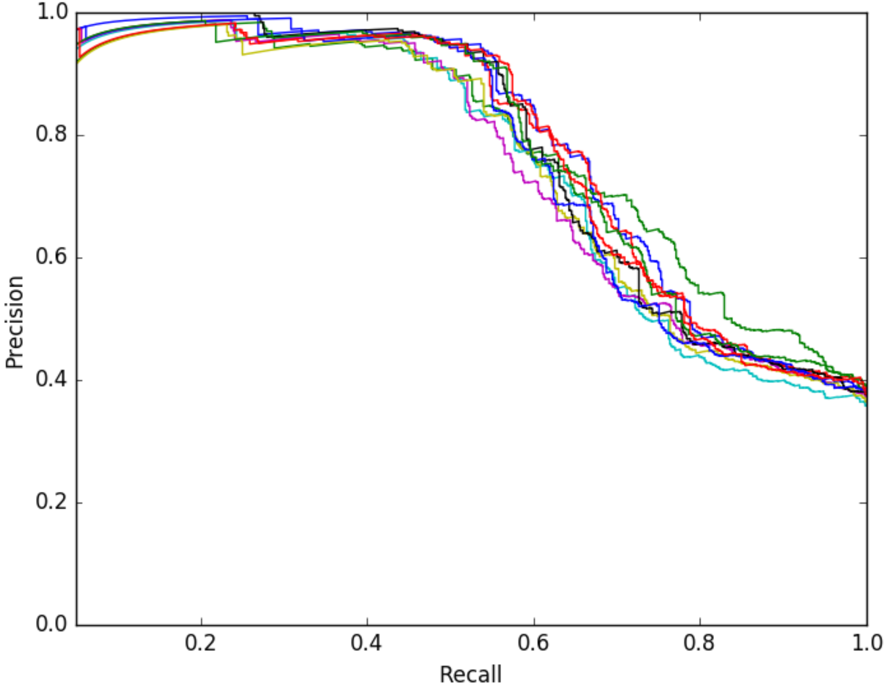
- Deviation from user's baseline
- Re-scale deviation score with a correction factor



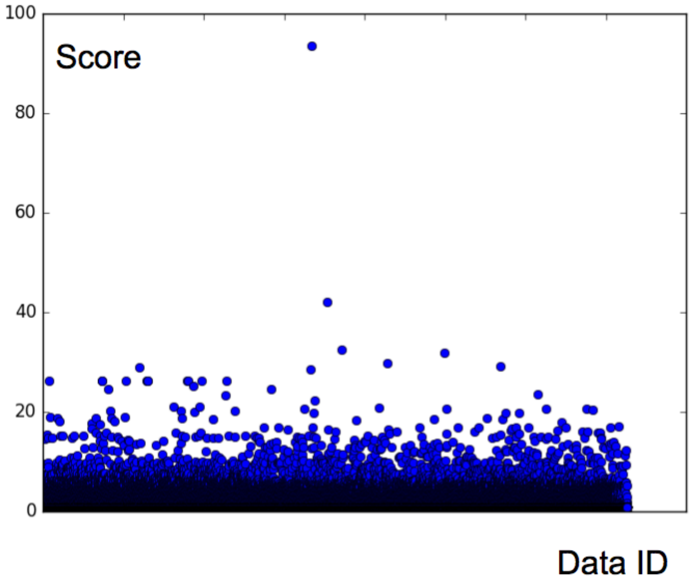
Detection

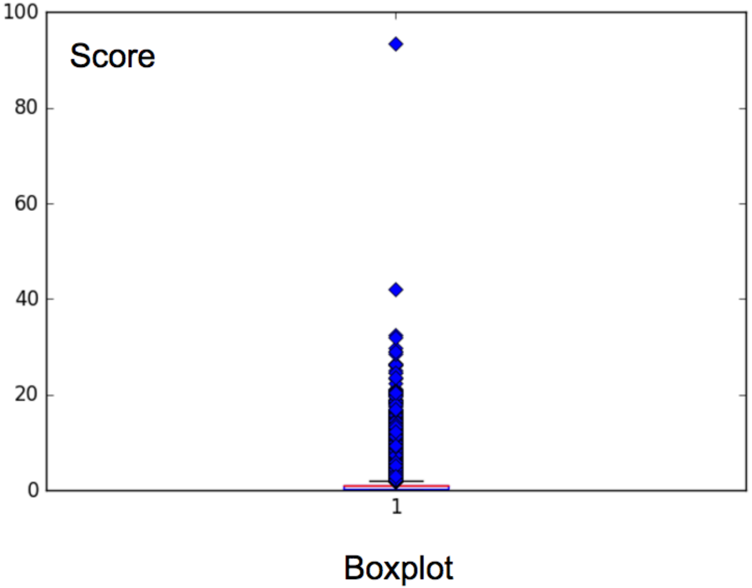

PCA Matrix
Profiles
Event Store
Feature Engineering
Parsed Data
Low Variance
Suspicious?
Yes
No
High Variance
Suspicious?
Yes
PCA Matrix
Profiles

Evaluation with Synthetic Data



Deployment





SPARK SUMMIT 2017

Thank You.

Wei Deng
wdeng@salesforce.com

Ping Yan
pyan@salesforce.com
🐦 @pingpingya