

Big Data Developments in Transaction Analytics

Scott M Zoldi, PhD
Vice President
FICO

Credit Scoring and Credit Control XIII
August 28-30, 2013

This presentation is provided for the recipient only and cannot be reproduced or shared without Fair Isaac Corporation's express consent.
© 2013 Fair Isaac Corporation.

FICO™

Agenda



- » **Transaction Analytics**
Streaming Transaction Profiles
- » **Deeper Understanding**
Behavior Sorted Lists
- » **Beyond a Single View**
Connecting the Holistic Customer
- » **Self-Calibrating Analytics**
Essential In-Stream Self-learning

Transaction Analytics

Streaming Transaction Profiles

Transaction Fraud Detection Long Term Streaming Analytic Problem



1993 Falcon® introduced

- » Real-time Predictive Analytics Decisions
- » Transaction profiles for recursive variable updates
- » Neural Networks for complex rapid score computation
- » Fraud scores operationalized in authorization systems
- » <10s of millisecond response times



2013

- » 65% of Payment cards transactions processed by Falcon®
- » More than a petabyte of transactions processed in real-time annually



Boiling Down Transaction Histories

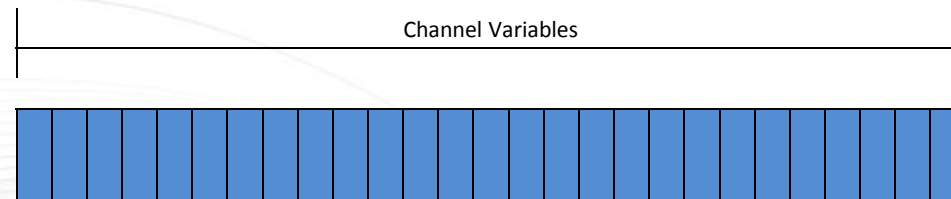


Cardholder

Transaction history

| |
|--------------------------------|
| June 3, \$28.00, Amazon, ... |
| June 4, \$33.51, Grocery, ... |
| June 4, \$100.00, ATM, ... |
| June 5, \$1.00, Preauth, ... |
| June 5, \$53.88, Gas, ... |
| June 8, \$28.00, Amazon, ... |
| June 9, \$60.00, ATM, ... |
| June 9, \$9.50, Fast food, ... |
| ... |
| ... |
| ... |
| ... |
| July 30, \$51.18, Grocery, ... |
| July 31, \$55.11, Gas, ... |
| Aug 1, \$28.16, Amazon, ... |
| Aug 2, \$80.00, ATM, ... |
| Aug 3, \$8.75, Fast food, ... |
| ... |
| Now |

Cardholder Transaction Profile



- » In many applications it is impractical to retrieve the entire transaction history for making predictions
- » Transaction Profiles contain recursive variables to summarize the relevant predictors

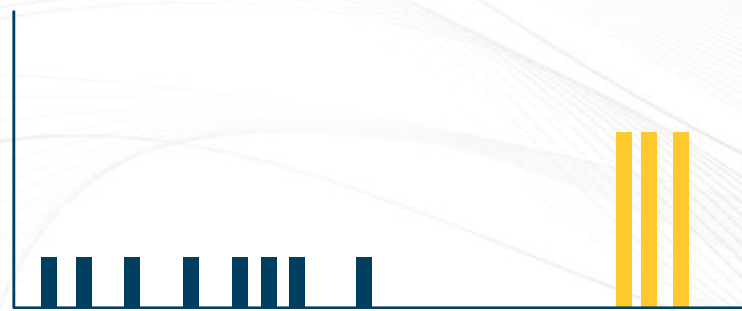
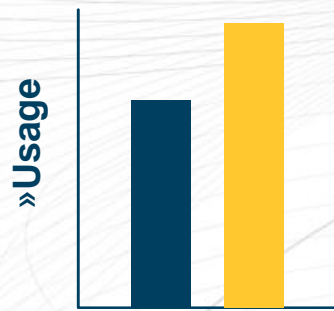
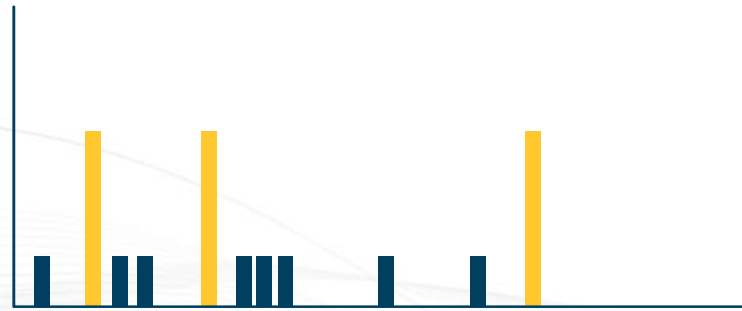
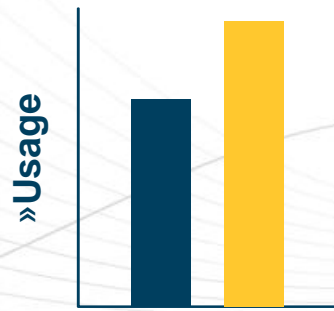
Capturing Patterns Using Analytics

» Summarized data

» Transaction data

■ » In Country

■ » Cross Border



» Month-End

» Time

Two Very Different Behaviors Look The Same With Summarized Data

Transaction Profiling Technology



» Transaction Profiles

- » Adjust based on a customer's real-time activity
- » Individualized to each customer
- » Optimized recursive mathematical variables

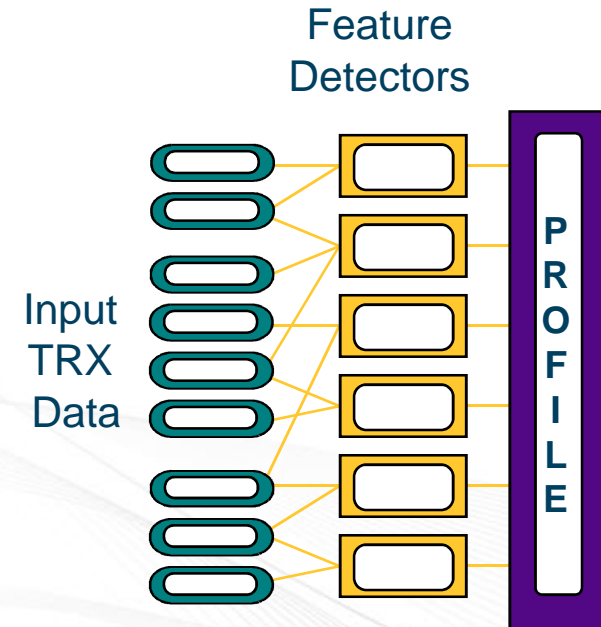
Profile variables include features such as:

- » Patterns of behavior
- » Recent trends and deviation from norm
- » Significant events

Profiles updated by every transaction

- » Relevant and most recent information used to make predictions and decisions.
- » Earlier identification of changes in customer behavior and allows timely scoring decisions

Custom designed for each problem



Profiles give the model the power to compare recent patterns with historical behavior, or undesirable behavior with normal, desirable behavior

Deeper Understanding

Behavior Sorted Lists

Why A Deeper Understanding of the Customer?

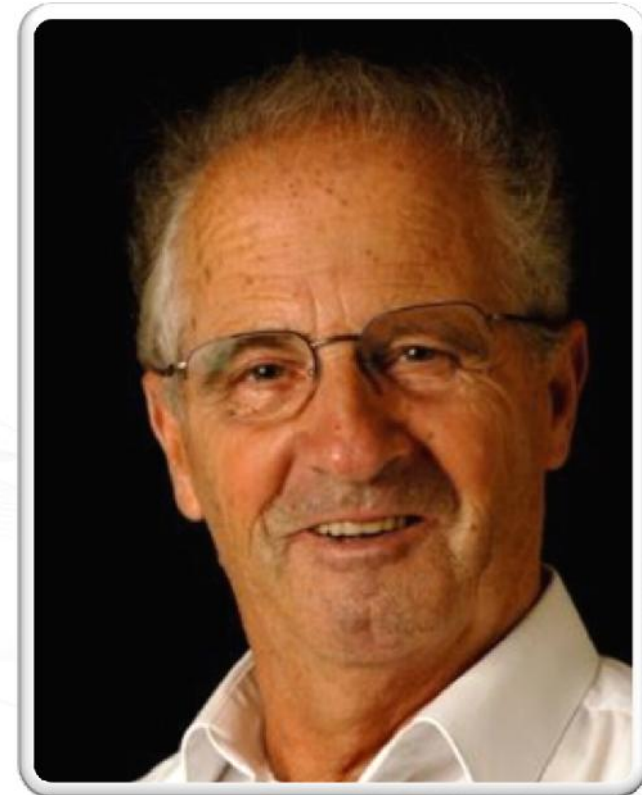


Reduce False Positives

- » Regularly do target behavior
- » Propensity to do target behavior

Eliminate Embarrassment

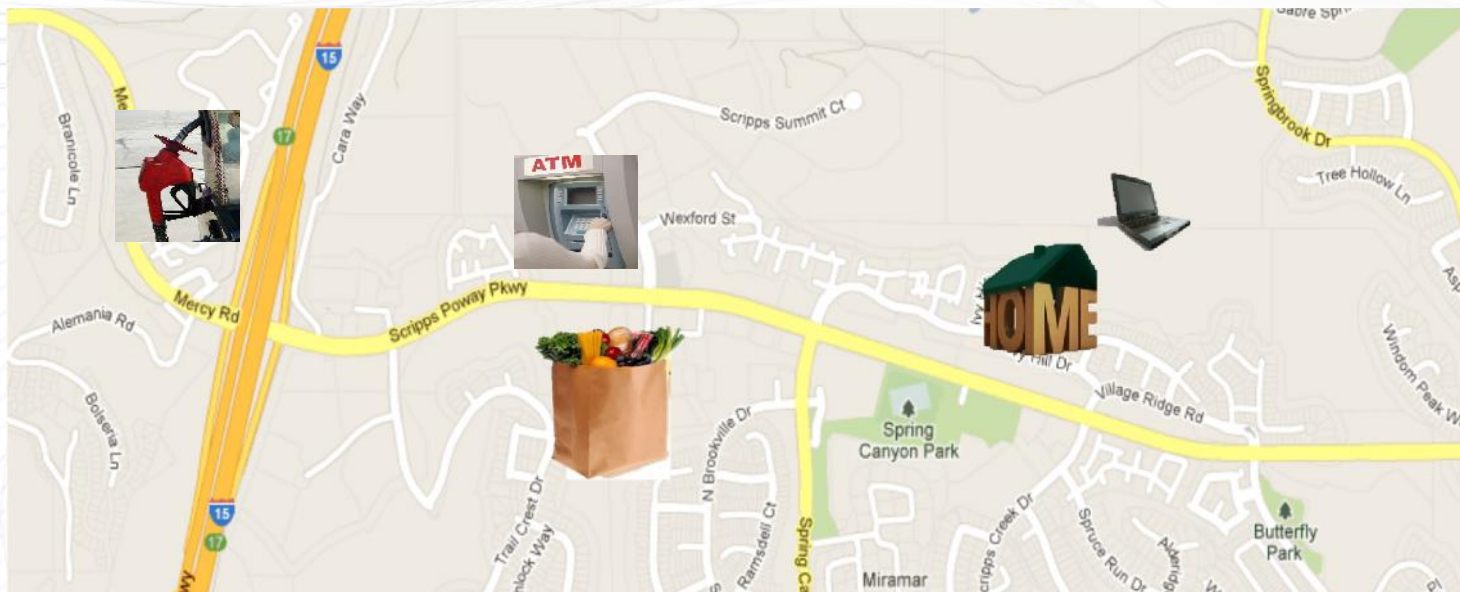
- » “I use that gas station every week”
- » “I pay that person every month”
- » “I travel to that country every year”
- » “Based on my spending pattern you should know my likely transactions”



Individuals are Creatures of Habit

Cardholders use

- » Favorite ATMs close to work or home
- » Favorite gas stations along the way of daily commute
- » Preferred grocery stores just for knowing where to find things
- » Preferred online stores for internet shopping
- » Other preferred merchants/localities (postal codes)



Frequent Behavior Tracking Algorithm



**Frequent ATM
BLIST table**

| Index | Entity Value | Frequency |
|-------|--------------|-----------|
| 1 | ATM_77 | F1=3.2 |
| 2 | ATM_318 | F2=9.2 |
| 3 | ATM_291 | F3=0.3 |
| 4 | ATM_54 | F4=2.7 |

| Index | Rank |
|-------|------|
| 1 | 2 |
| 2 | 1 |
| 3 | 4 |
| 4 | 3 |

**Rank
Table**

A New ATM withdrawal at ATM_54

| Index | Entity Value | Frequency |
|-------|--------------|-----------|
| 1 | ATM_77 | F1*wt |
| 2 | ATM_318 | F2*wt |
| 3 | ATM_291 | F3*wt |
| 4 | ATM_54 | F4*wt+Fo |

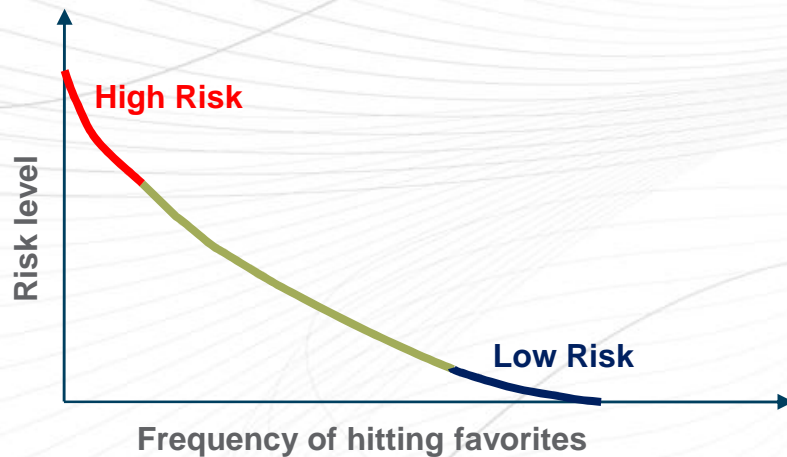
| Index | Rank |
|-------|------|
| 1 | 3 |
| 2 | 1 |
| 3 | 4 |
| 4 | 2 |

ATM_318 and ATM_54 are Favorites after update

BLIST Differentiating Fraud Risks



- » **In-Pattern** means low risk while **Out-Pattern** means high risk
- » BLIST helps to reduce Falcon scores for normal cardholder **In-Pattern** spending even in risky categories



| T&E Merchant | Rank |
|--------------|------|
| V-Liquor | 1 |
| S-Casino | 2 |

Legacy FFM

| | |
|----------|-----|
| V-Liquor | 820 |
| S-Casino | 930 |
| V-Liquor | 870 |

In Pattern

Reduce scores

FFM with BLIST

| | |
|----------|-----|
| V-Liquor | 670 |
| S-Casino | 710 |
| V-Liquor | 630 |

Beyond a Single View

Connecting the Holistic Customer

Why Connect Decisions?

Better Customer Experience

- » Eliminate conflicting decisions
- » Reduce customer impact due to multiple inquiries

Better Predictions

- » Leverage additional information across more channels
- » Customer activity is correlated

Better Business Decisions

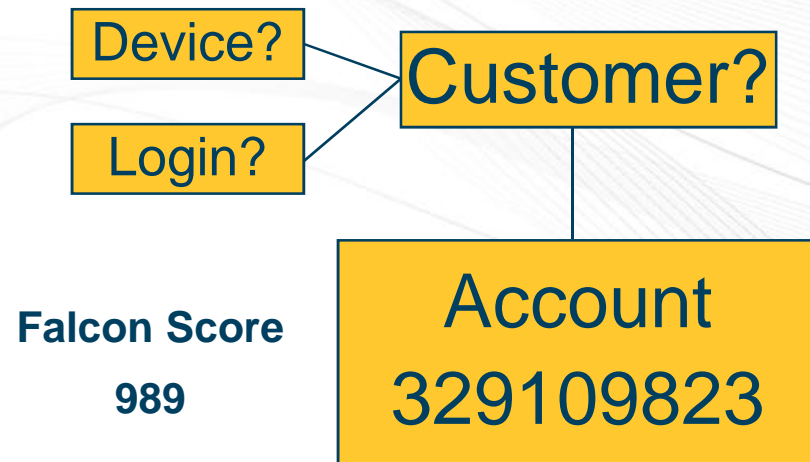
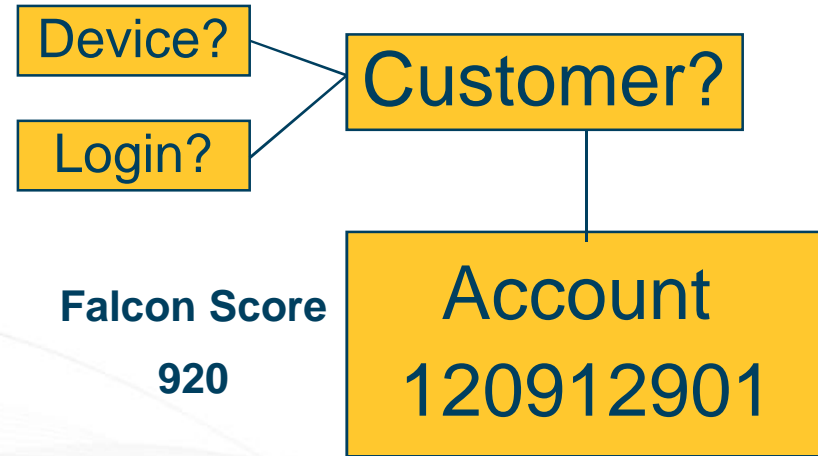
- » Know Your Customer (KYC)



Customer within a Silo – limited view / understanding



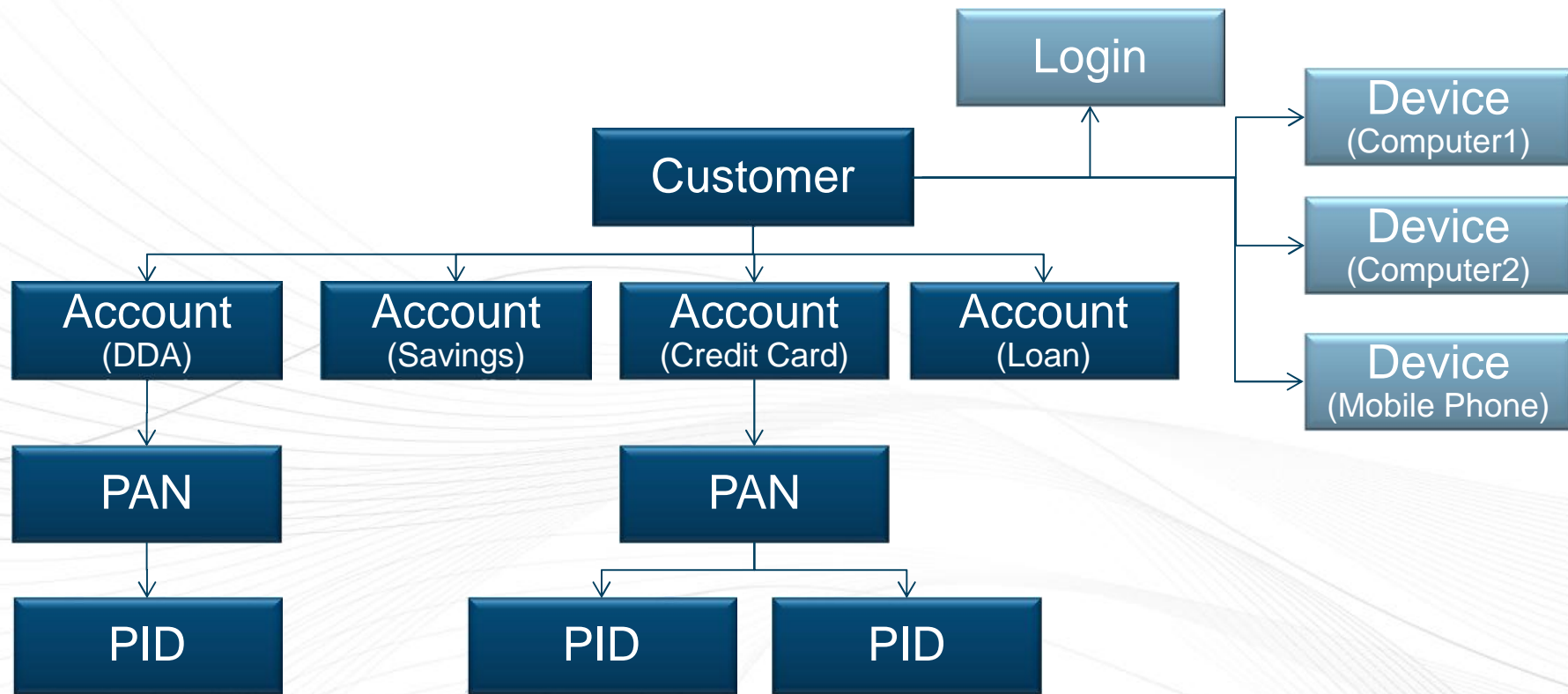
- » Often effective scores exist in silos
- » The complete customer view requires connecting silos/decisions
- » Must link multiple accounts, devices, logins to the customer



Connecting Profiles: A Correct Data Model



| Field Name | Type | Definition |
|---------------------------------|-------------------|---|
| recordType | CHAR | This Record's Type. "ACH" "Check" "Credit Card " "Customer Information" "Account Information" "Application" Etc. |
| recordCreationDate | YYYYMMDD | Date of the record creation |
| recordCreationTime | HHMMSS | Time of the record creation |
| customerIdFromHeader | CHAR | The Primary Customer Number. Financial institution's unique identifier for the customer. |
| customerAcctNumber | CHAR | Customer Account Number. Identifier for the account used in the transaction, |
| Transaction Record Format | Per Specification | Field Level Definitions |



Customer = The unique identifier for a given individual across all channels.

Account = An identifier for an account which may be one of many held by the associated customer.

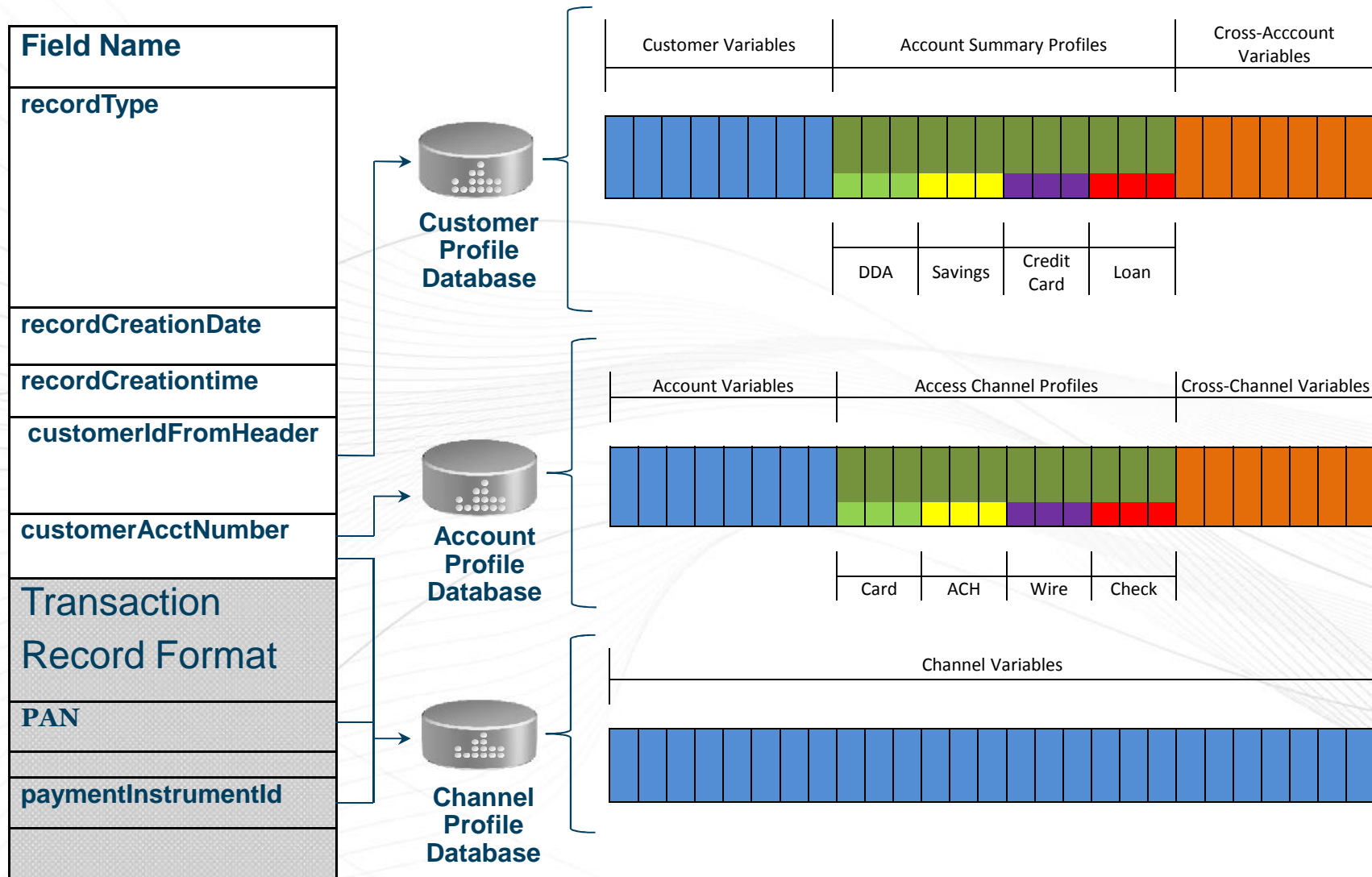
PAN = The primary account number account number encoded or embossed on the payment instrument

Payment Instrument Id (PID) = Identifier to distinguish between payment instruments with the same PAN.

Device = Identifier for the device the customer uses access his accounts at the financial institution.

Login = The login user identifier the customer uses to access his

Connecting Profiles: Multiple Profiles



Customers have Different Mixes of Accounts



| Credit | Debit | DDA | ONLINE | Unsecured | Checking |
|--------|-------|-----|--------|-----------|----------|
| 1 | 0 | 0 | 0 | 0 | 0 |

| Credit | Debit | DDA | ONLINE | Unsecured | Checking |
|--------|-------|-----|--------|-----------|----------|
| 2 | 0 | 0 | 0 | 0 | 0 |



| Credit | Debit | DDA | ONLINE | Unsecured | Checking |
|--------|-------|-----|--------|-----------|----------|
| 1 | 1 | 2 | 1 | 0 | 1 |

How to train models with this variability / complexity?

Will models be relevant?!

Self-Calibrating Analytics

Essential In-Stream Self-Learning

Why Self Calibrating Models?

Data availability issue

- » Data export restrictions
- » Gathering of data impractical
- » Unreliable targets/no targets

Diverse Customers

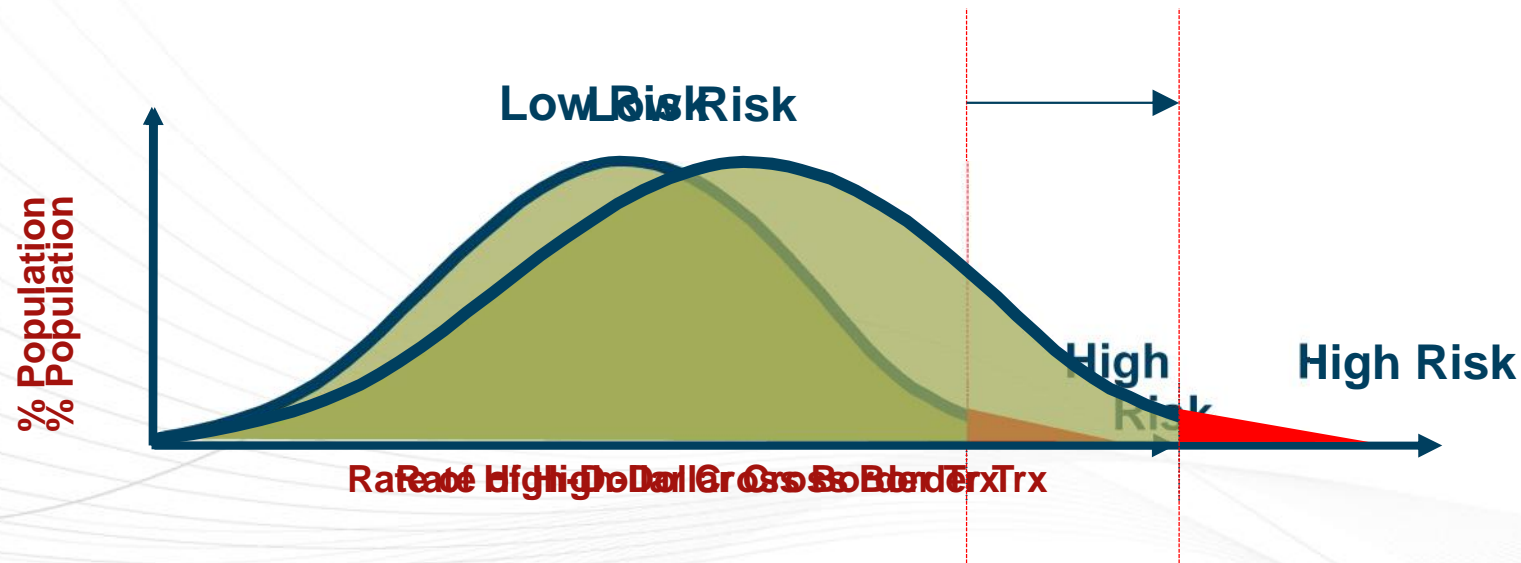
- » Peer grouping addresses diversity
 - » NOT 100+ poorly trained models

Dynamic Environment

- » Real-time adjustment to changes

**Emerging Markets Require
Flexible Streaming Analytics**





Scaling function

$$q(x_i | s) = \frac{x_i - s_p}{(s_L - s_R) / c} \in [0, C]$$

where s_p s_L s_R are quantiles estimated real-time

Self-Calibrating Analytics



$$Score = \sum w_i q(x_i | t, s)$$



$(t_1, \dots, t_m) \in T$: segment vector (m segments)

$(x_1, \dots, x_p) \in X$: profile variables (p variables)

$(s_1, \dots, s_l) \in S$: quantile estimates

w_i : weights

Weights can be assigned

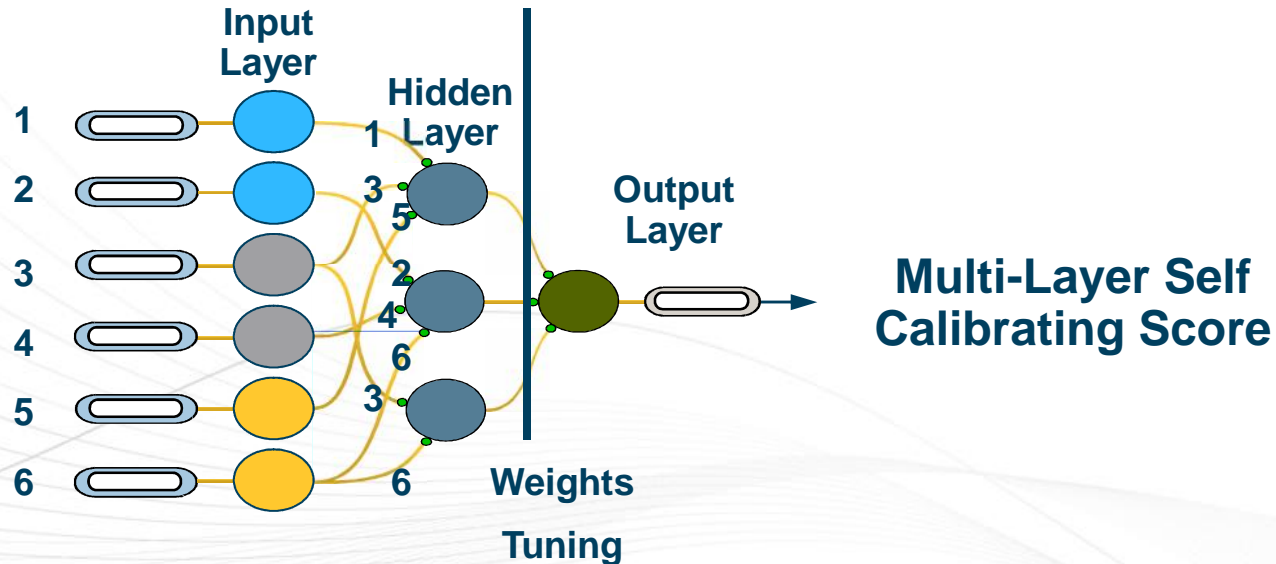
- » Uniformly
- » Using Expert Knowledge
- » Based on limited data

Peer Group
Distributions



Multi-layered Self-Calibrating Models

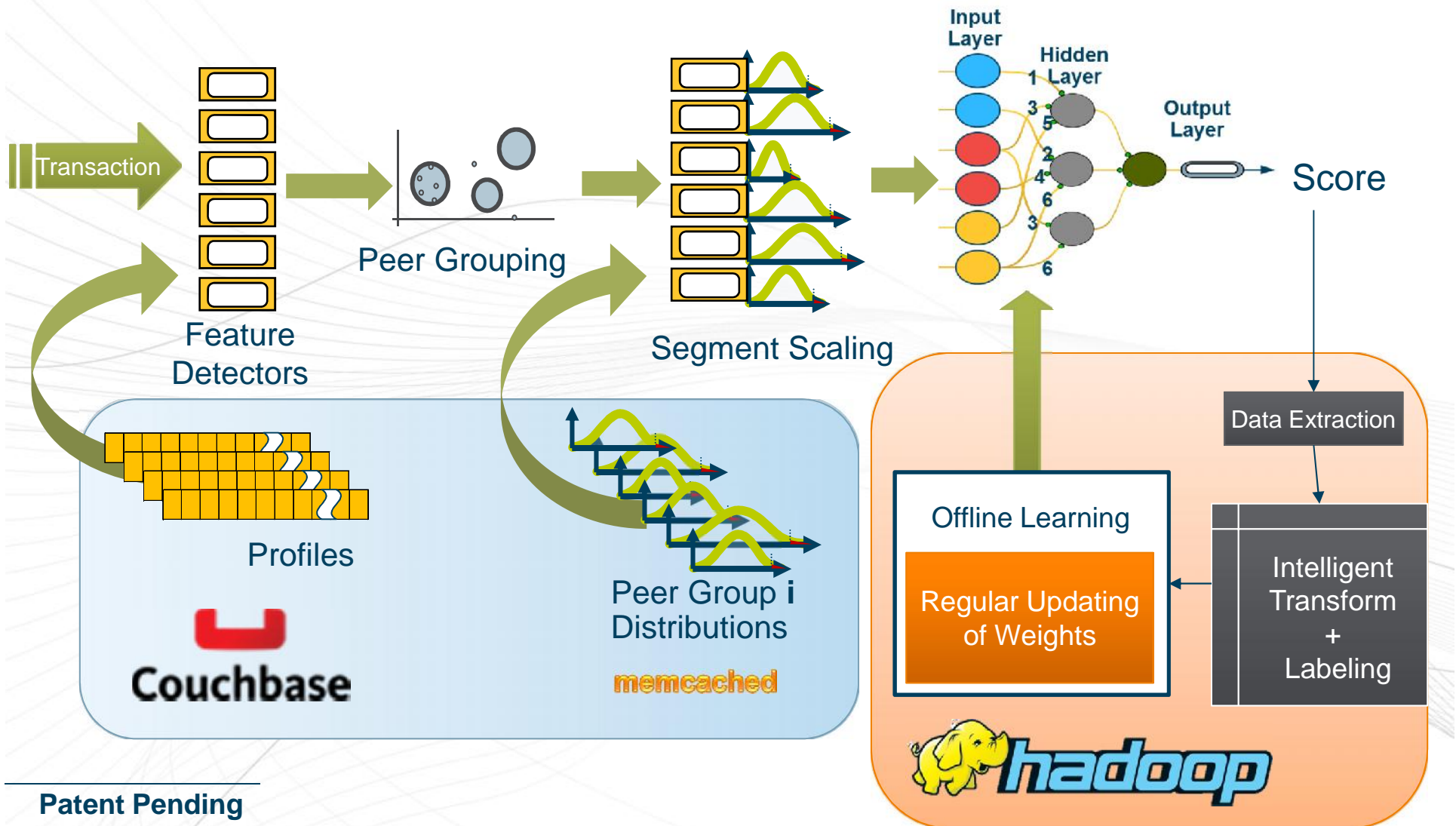
US 13/367,344 Pending



Inspired by Neural Networks

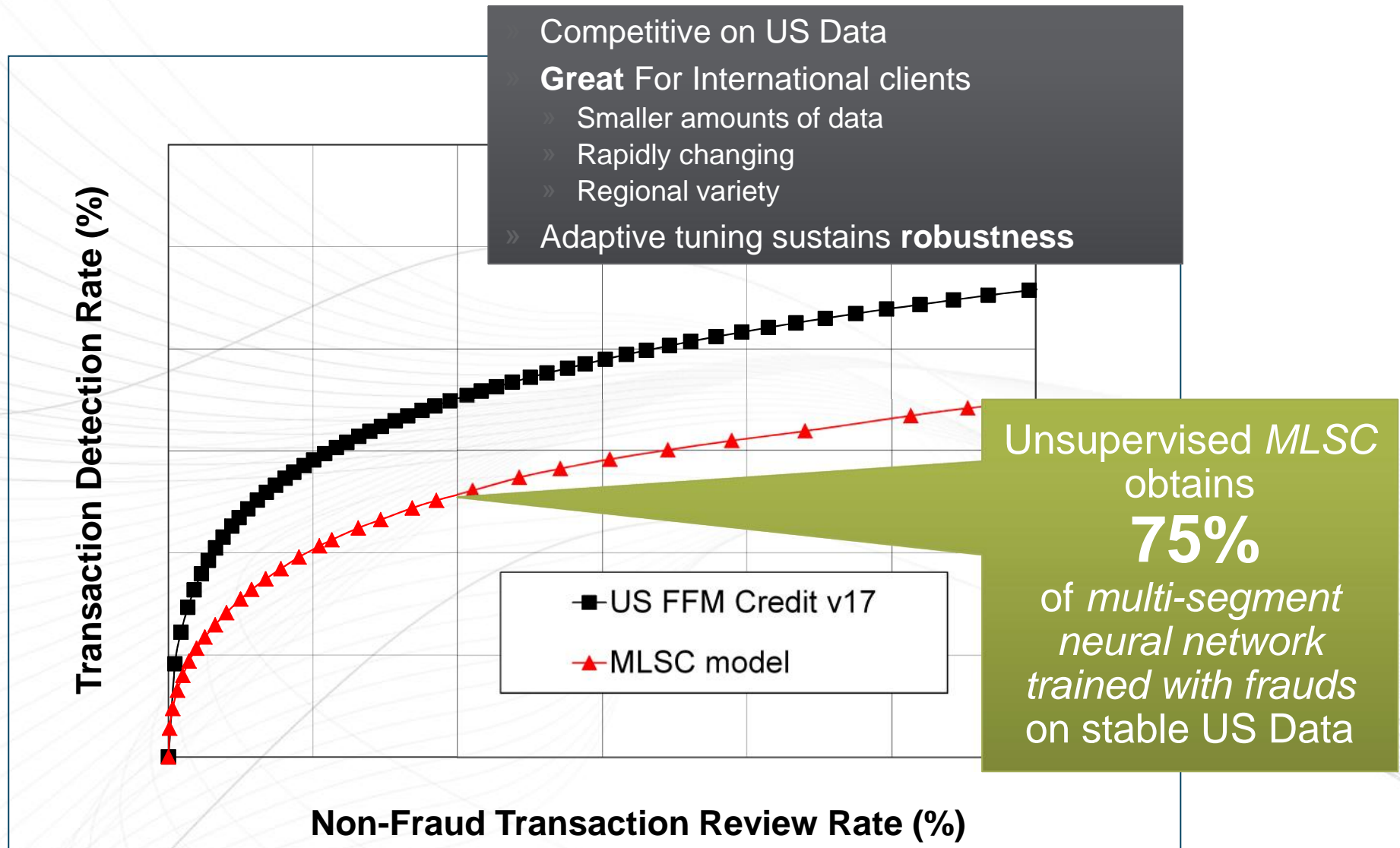
- » Each hidden node is a separate self-calibrating model
- » Variables in the hidden node are selected based on Factor group analysis
- » Ability to do simple tuning of output weighting of hidden nodes to improve performance or study effectiveness of different hidden nodes
- » Can experiment with new hidden nodes/variables in production, dial up promotion of nodes

Adaptive MLSC Model Architecture



Patent Pending

Performance Results US Data



THANK YOU

Scott M Zoldi, PhD
858-369-8891
ScottZoldi@fico.com

Credit Scoring and Credit Control XIII
August 28-30, 2013

This presentation is provided for the recipient only and cannot be reproduced or shared without Fair Isaac Corporation's express consent.

© 2013 Fair Isaac Corporation.

