FIN301

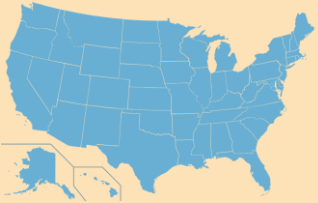# Fraud Detection and Machine Learning on AWS

Pawan Agnihotri, Principal Solutions Architect, Global Financial Services

November 28, 2016

Payments fraud is an ongoing concerns for
Financial Services (FS) organizations.

In 2015…

$8.5b of fraud losses
in the US*

$21.8b of fraud losses
globally*

$31.7b of *projected*
fraud losses globally in 2020*

* From The Nilson Report (https://www.nilsonreport.com/constant_contact_promo.php?id_promo=8)

Solving for Multiple Layers Simultaneously

Layer 1

Addresses endpoint authentication

Layer 2

Analyzes session behavior

Layer 3
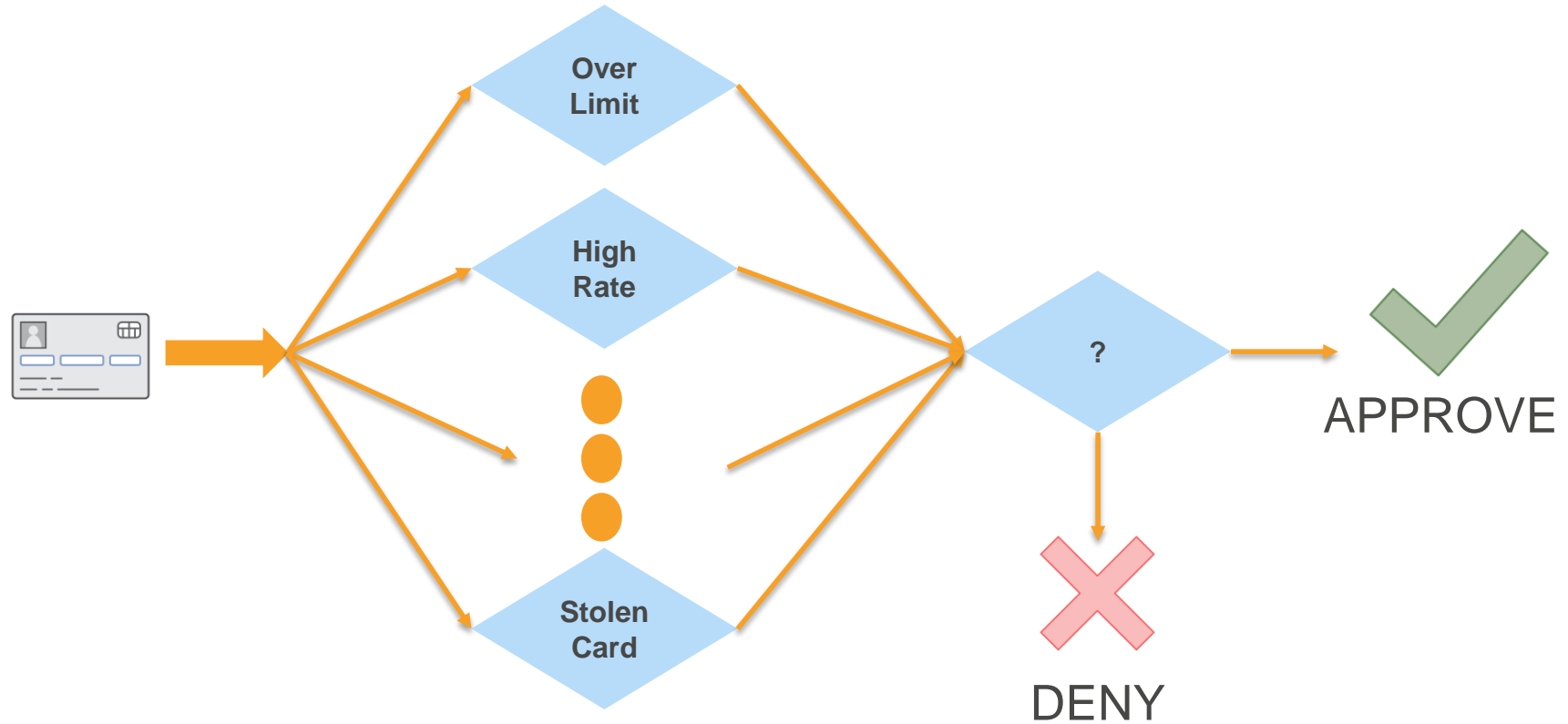
Monitors account behavior for a channel

Layer 4

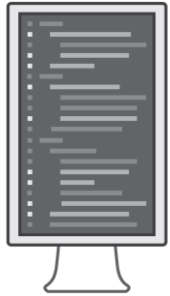Monitoring account behavior across multiple channels

Layer 5

Monitoring multiple account behaviors across multiple channels
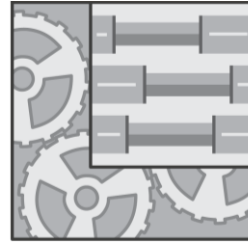
Rule-Based Fraud Detection

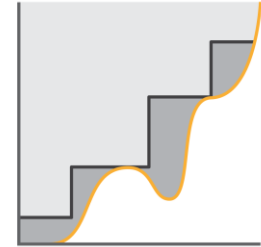# Rules Fall Short for Fraud Detection

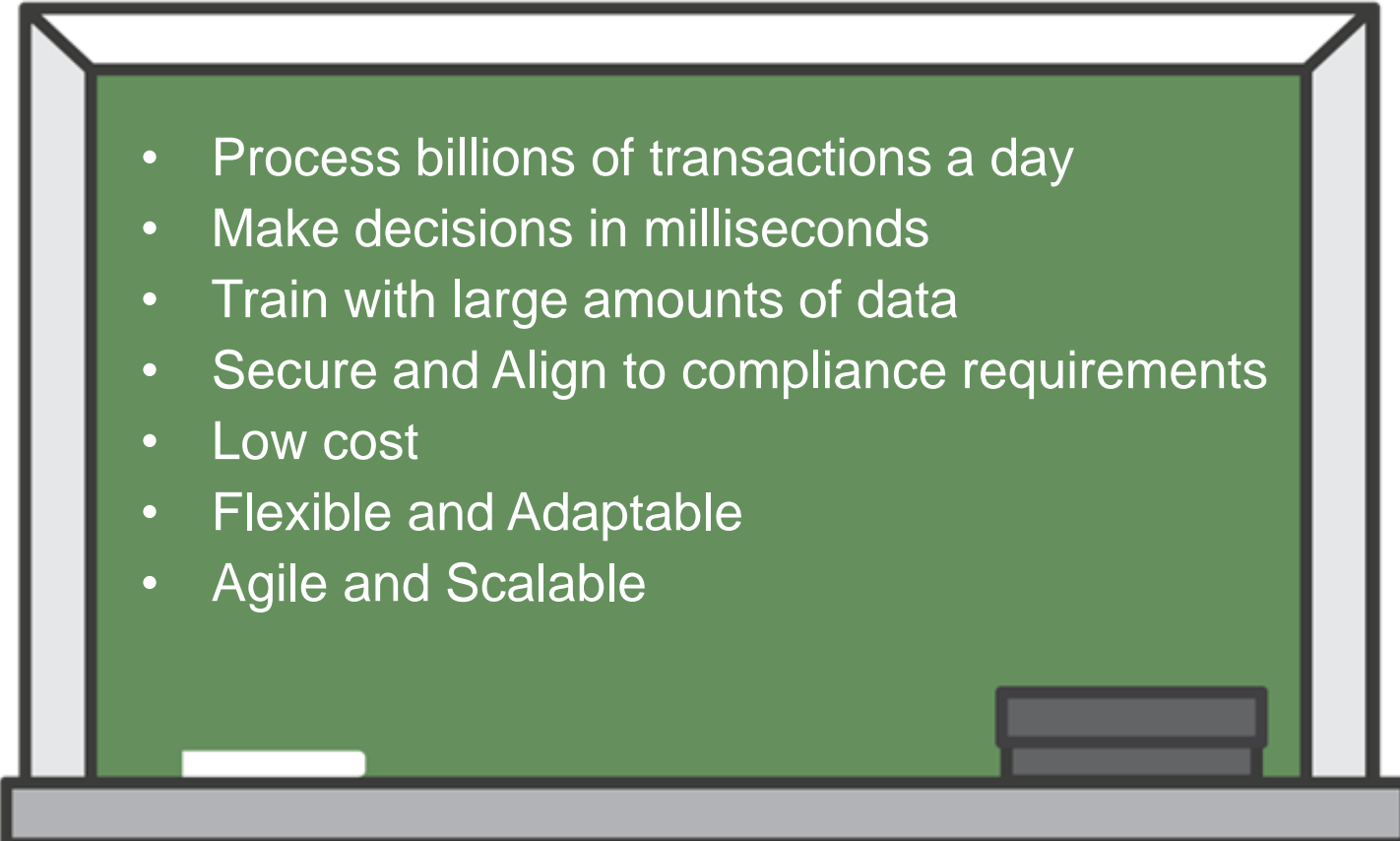Static set of rules

Human errors & bias
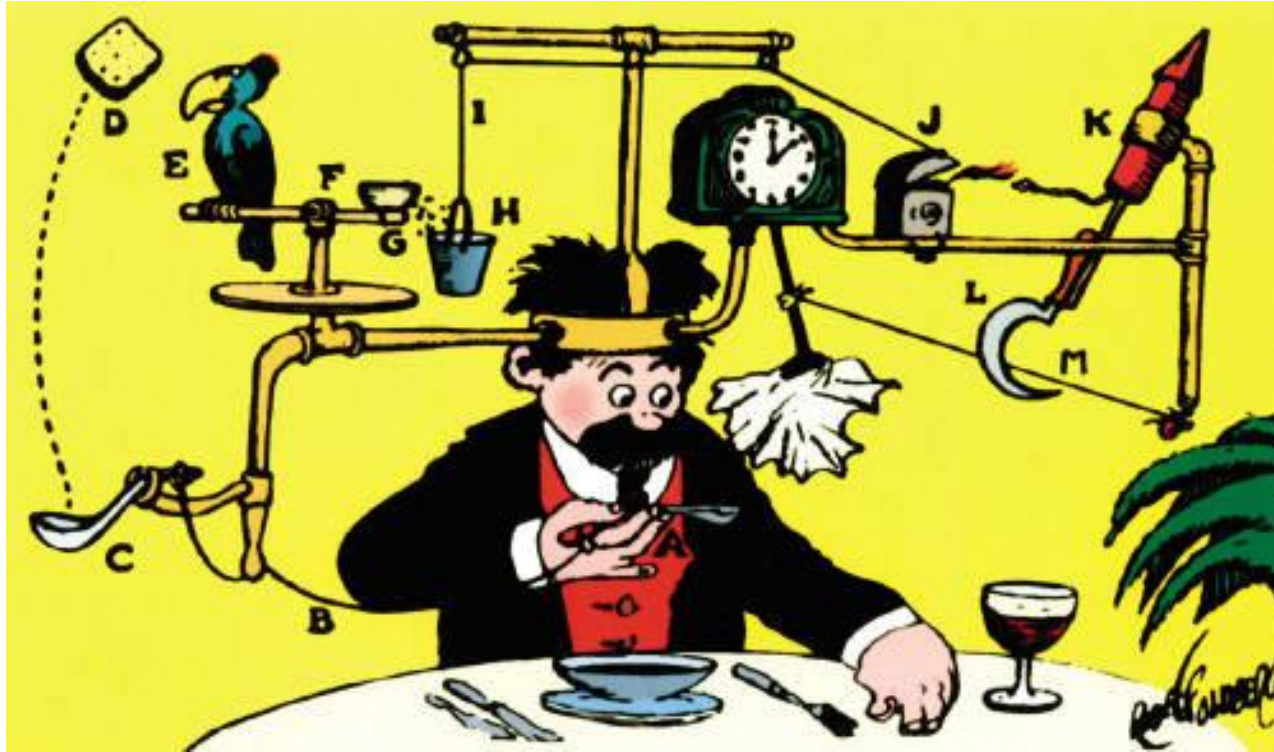

Difficult to manage

Cannot scale

Solution Requirements

- Process billions of transactions a day
- Make decisions in milliseconds
- Train with large amounts of data
- Secure and Align to compliance requirements
- Low cost
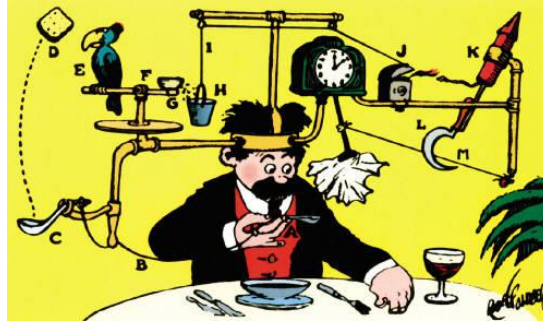- Flexible and Adaptable
- Agile and Scalable

Overview of Machine Learning

Supervised Learning

Supervised Learning

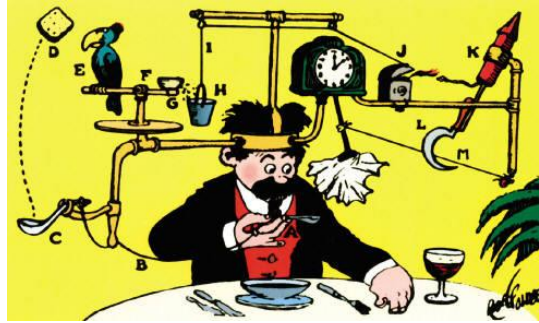Input → → Outcome

Supervised Learning

Input   →

Input   →
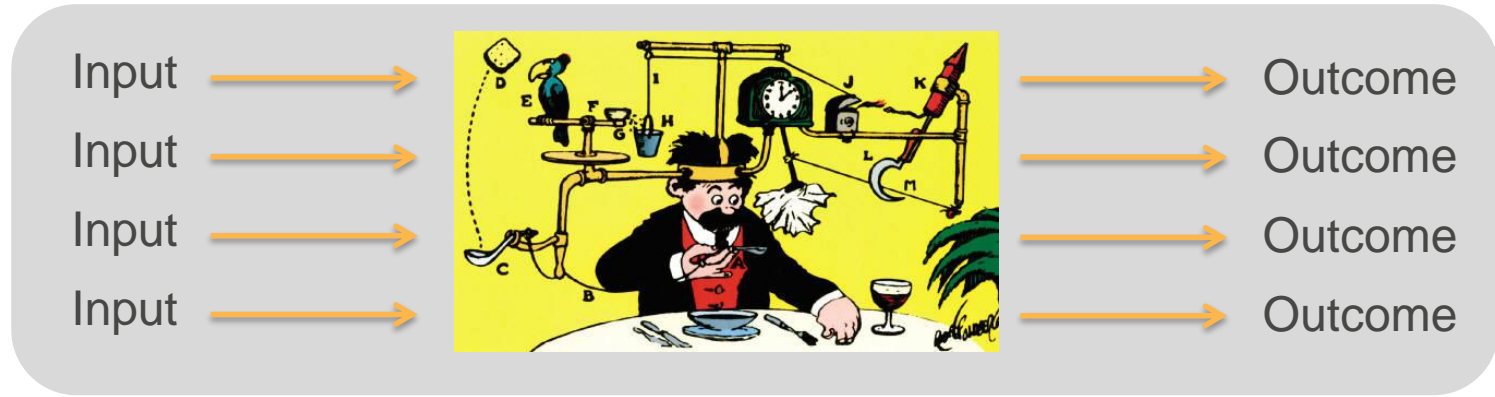
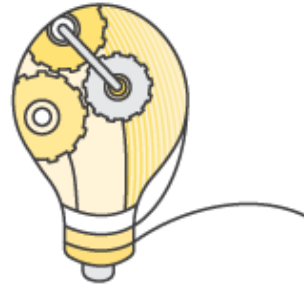Input   →

Input   →

→   Outcome

→   Outcome

→   Outcome

→   Outcome

Supervised Learning

Input →

Input →

Input →

Input →

→ Outcome

→ Outcome

→ Outcome

→ Outcome

Known Historical Data

Supervised
Learning ↓

Supervised Learning

Input → → Outcome
Input → → Outcome
Input → → Outcome
Input → → Outcome

Known Historical Data

Supervised Learning ↓

Unseen Input → → Same Outcome
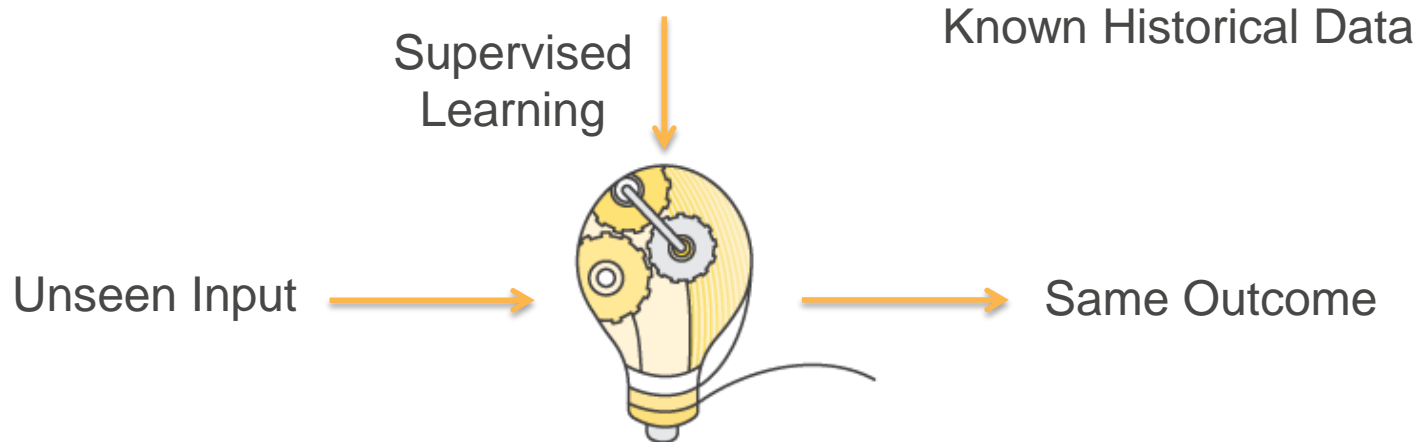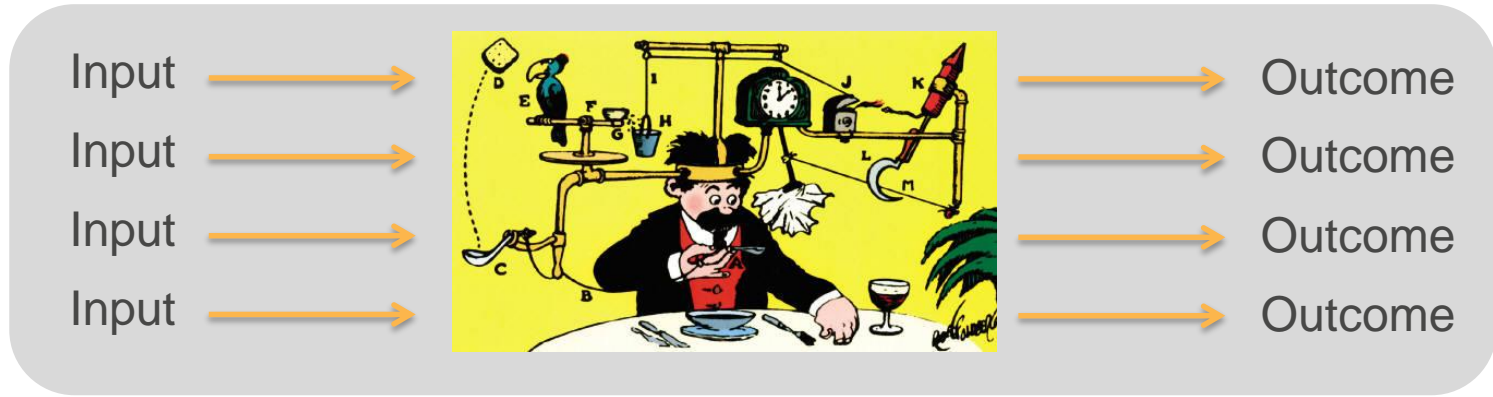
Tools of the Trade

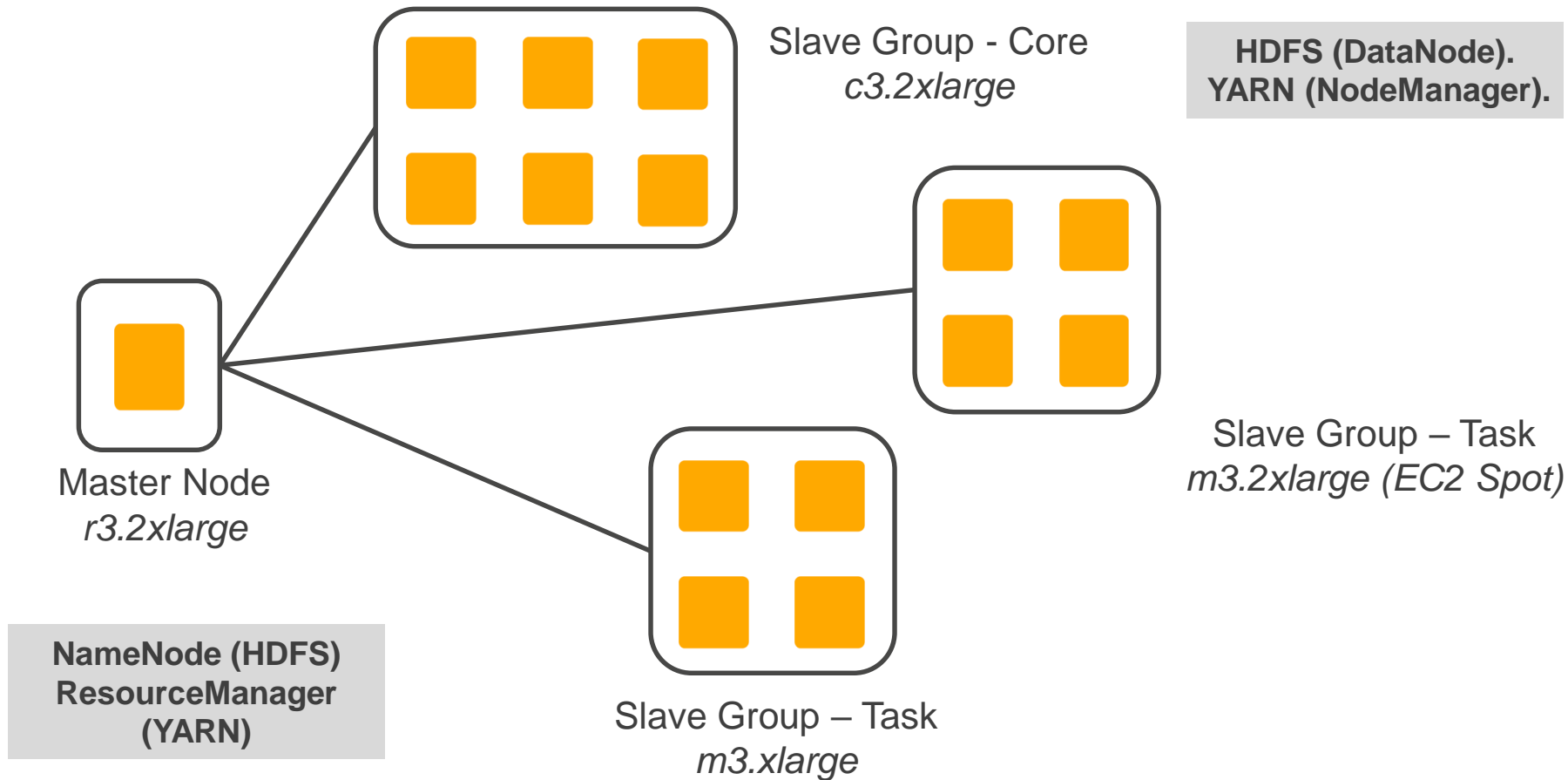# Amazon Simple Storage Service (S3)

- Highly scalable object storage
- Files are stored as objects and organized into high-level folders called buckets
- Store and retrieve data from anywhere on the web
- Native support data encryption at rest
- Data in transit to and from the service is encrypted using SSL.
- Pay for exactly what you use
- Highly durable (99.999999999% design)
- Limitlessly scalable

Amazon Elastic Map Reduce (EMR)

- *Managed* platform
- MapReduce, Apache Spark, Presto
- Launch a cluster in minutes
- Open source distribution & MapR distribution
- Elasticity of the cloud
- Built in security features
- Support for encryption of data at rest and in transit
- Pay by the hour and save with Spot
- Flexibility to customize

An Example EMR Cluster

Slave Group - Core
c3.2xlarge

HDFS (DataNode).
YARN (NodeManager).

Slave Group – Task
m3.2xlarge (EC2 Spot)

Master Node
r3.2xlarge

NameNode (HDFS)
ResourceManager
(YARN)

Slave Group – Task
m3.xlarge

Flexibility to add Hadoop applications to Amazon EMR

Processing
- Apache Apex™
- Flink
- Spark

Databases
- presto
- HIVE
- APACHE HBASE

Analytics
- Zeppelin
- mahout

Amazon Machine Learning
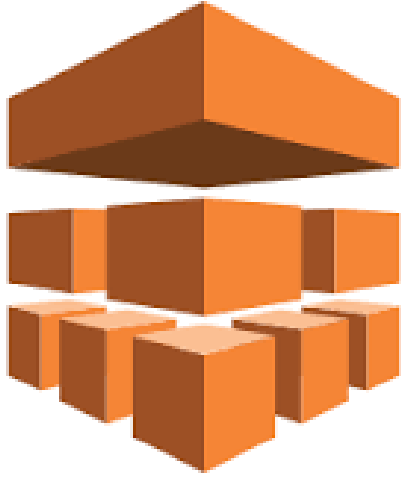
- Easy-to-use service built for developers
- Robust, powerful, and technology-based
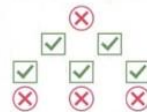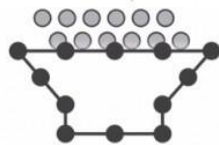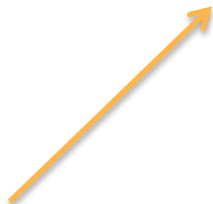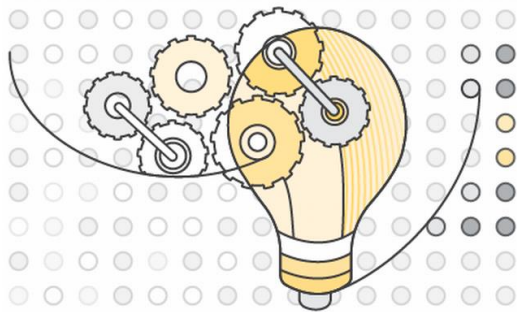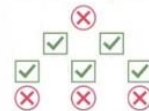- Ability to create models using your data
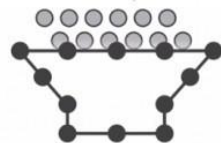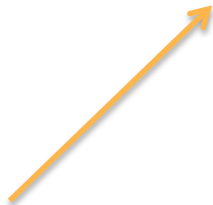- Deployable to production in seconds
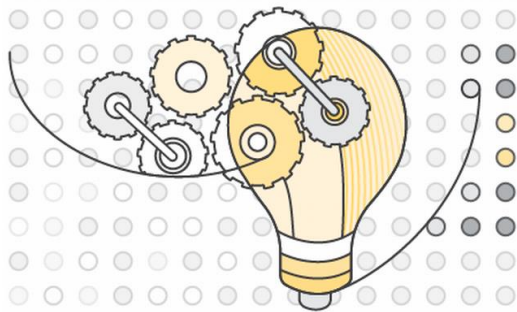
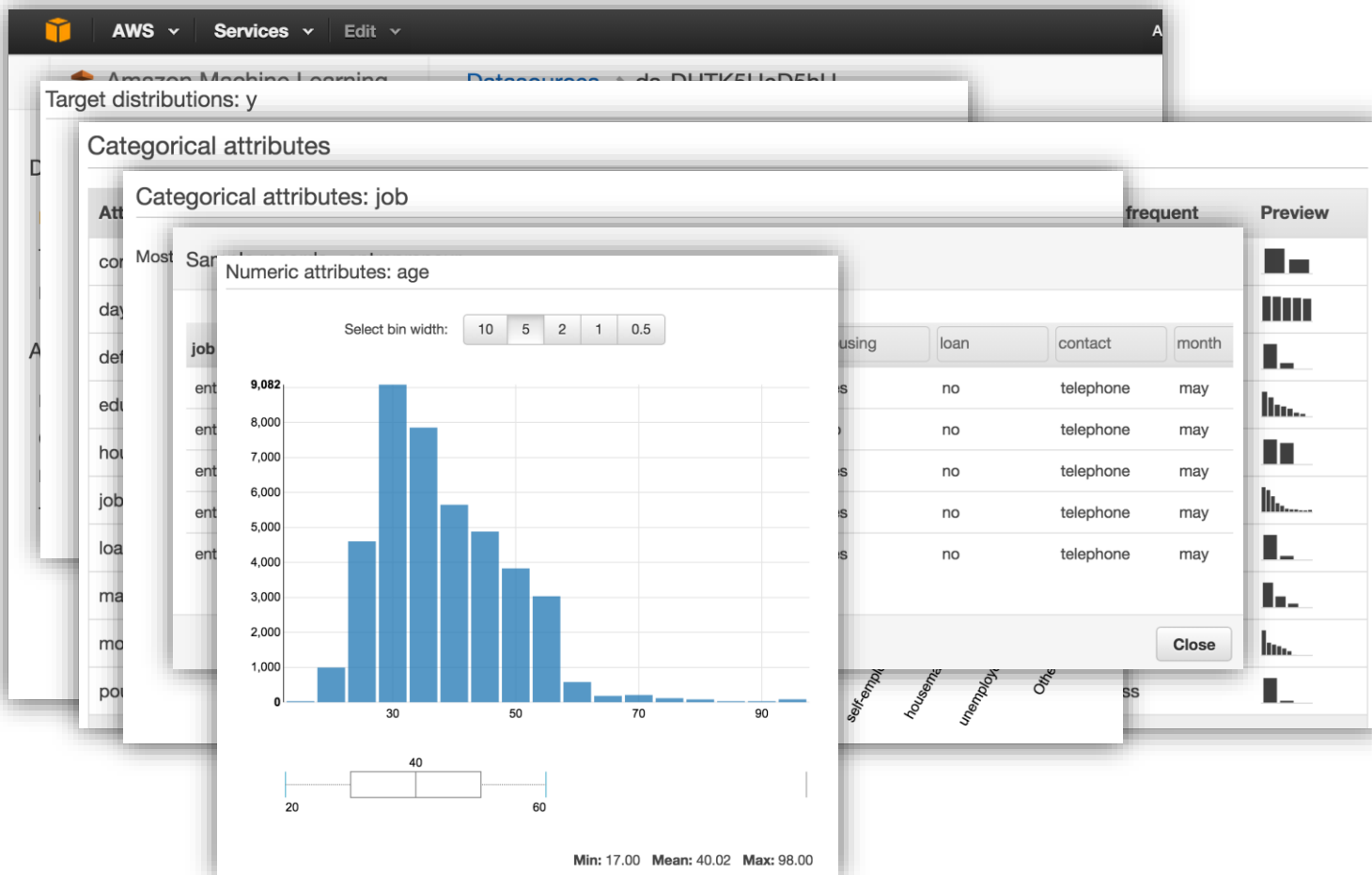# Amazon Machine Learning Service
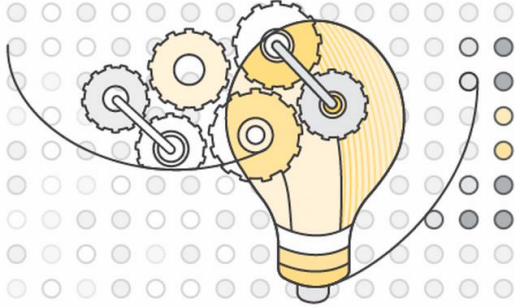
Amazon Machine Learning Service
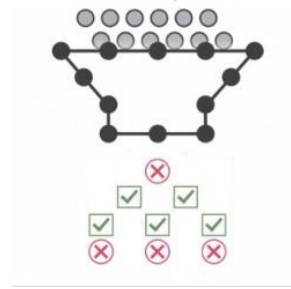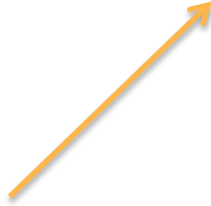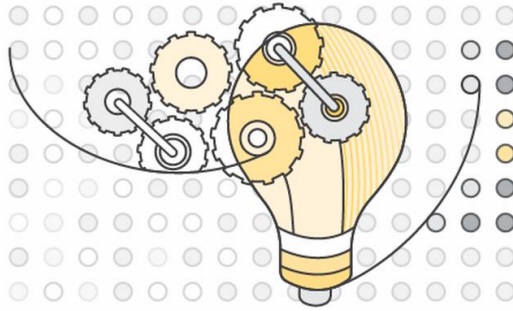
Amazon Machine Learning Service
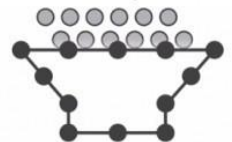
Amazon Machine Learning Service
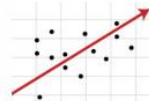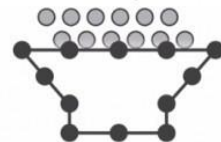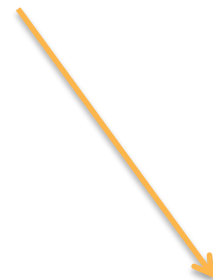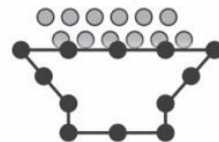
Explore and Understand Your Data

Evaluate and Explore Model Performance
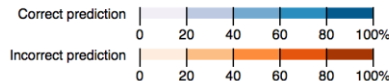
Putting It Together

Credit Card Transaction Dataset
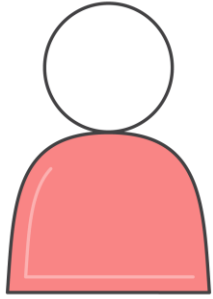
Customer Profile + **Store Profile** + **Transaction Details**

Model Creation and Training – Reference Architecture

PCI Data Security Standard – High Level Overview

|	
Build and Maintain a Secure Network and Systems	1. Install and maintain a firewall configuration to protect cardholder data
	2. Do not use vendor-supplied defaults for system passwords and other security parameters
Protect Cardholder Data	3. Protect stored cardholder data
	4. Encrypt transmission of cardholder data across open, public networks
Maintain a Vulnerability Management Program	5. Protect all systems against malware and regularly update anti-virus software or programs
	6. Develop and maintain secure systems and applications
Implement Strong Access Control Measures	7. Restrict access to cardholder data by business need to know
	8. Identify and authenticate access to system components
	9. Restrict physical access to cardholder data
Regularly Monitor and Test Networks	10. Track and monitor all access to network resources and cardholder data
	11. Regularly test security systems and processes
Maintain an Information Security Policy	12. Maintain a policy that addresses information security for all personnel
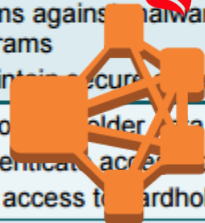
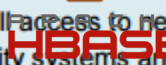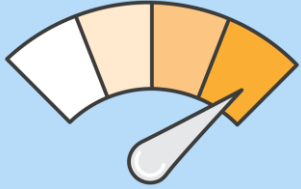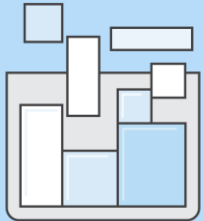# Online Fraud Detection – Reference Architecture
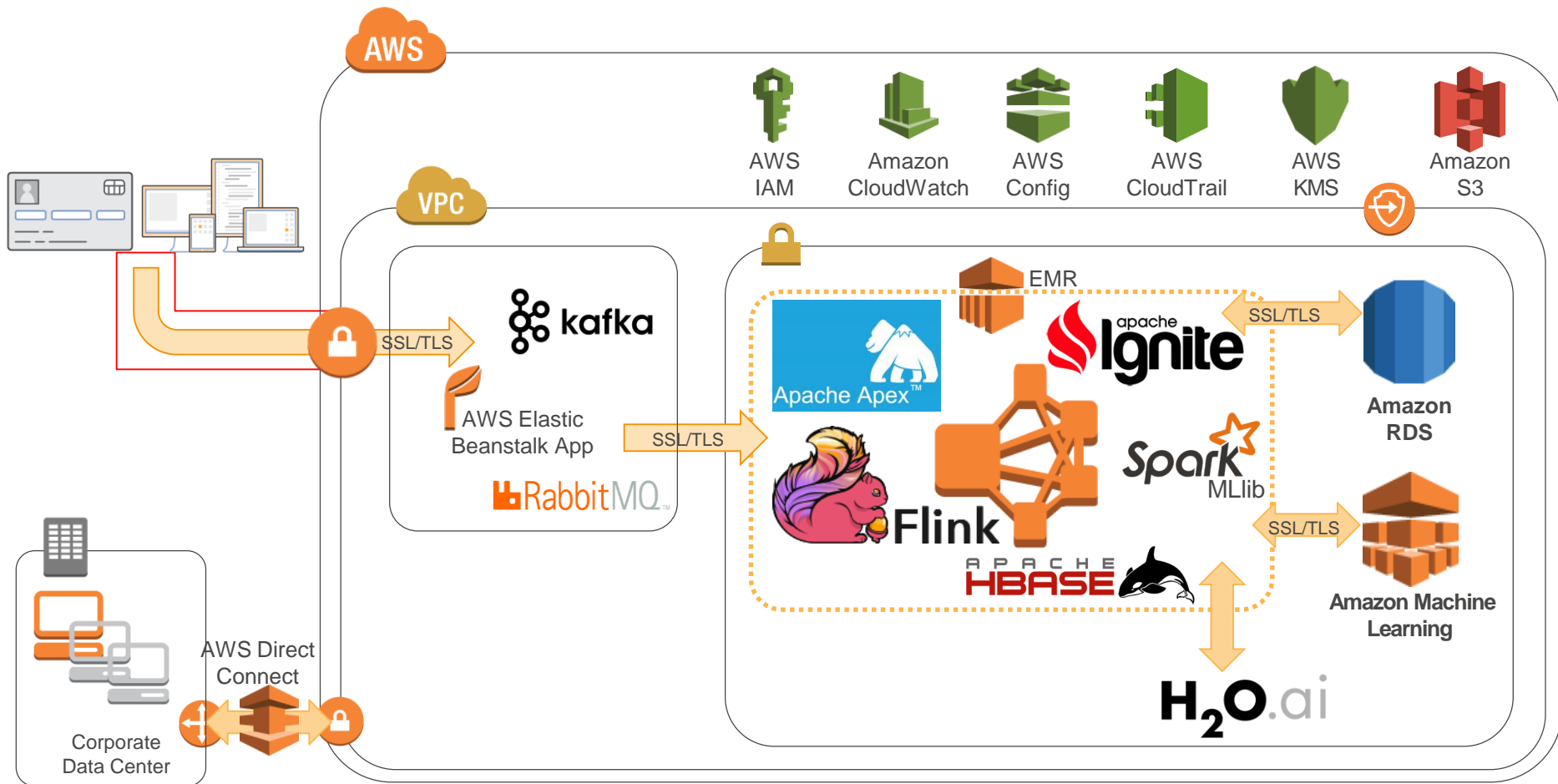
The Outcomes of the AWS Solution

Cost: Solution price down from $100K to $10K
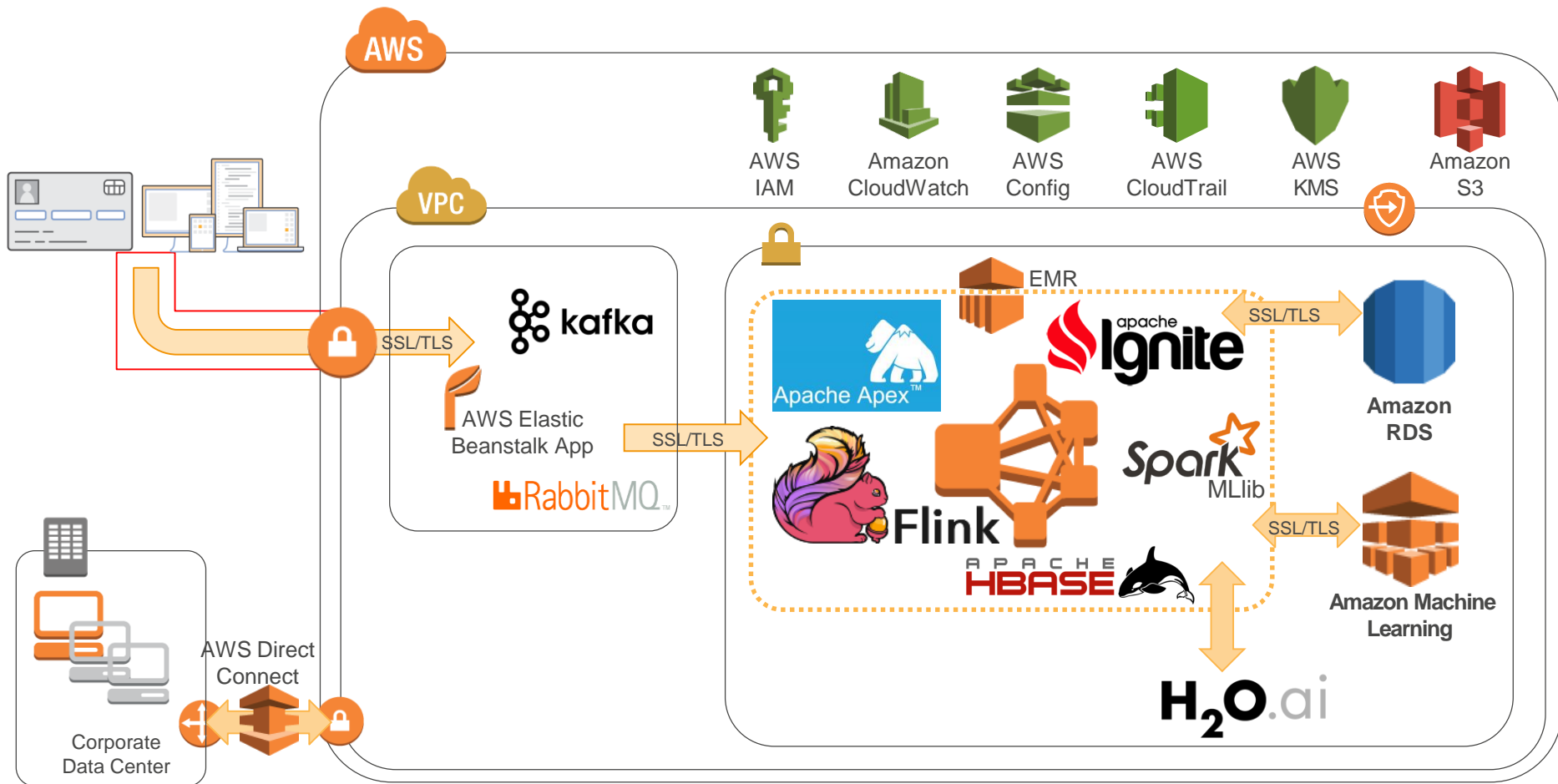

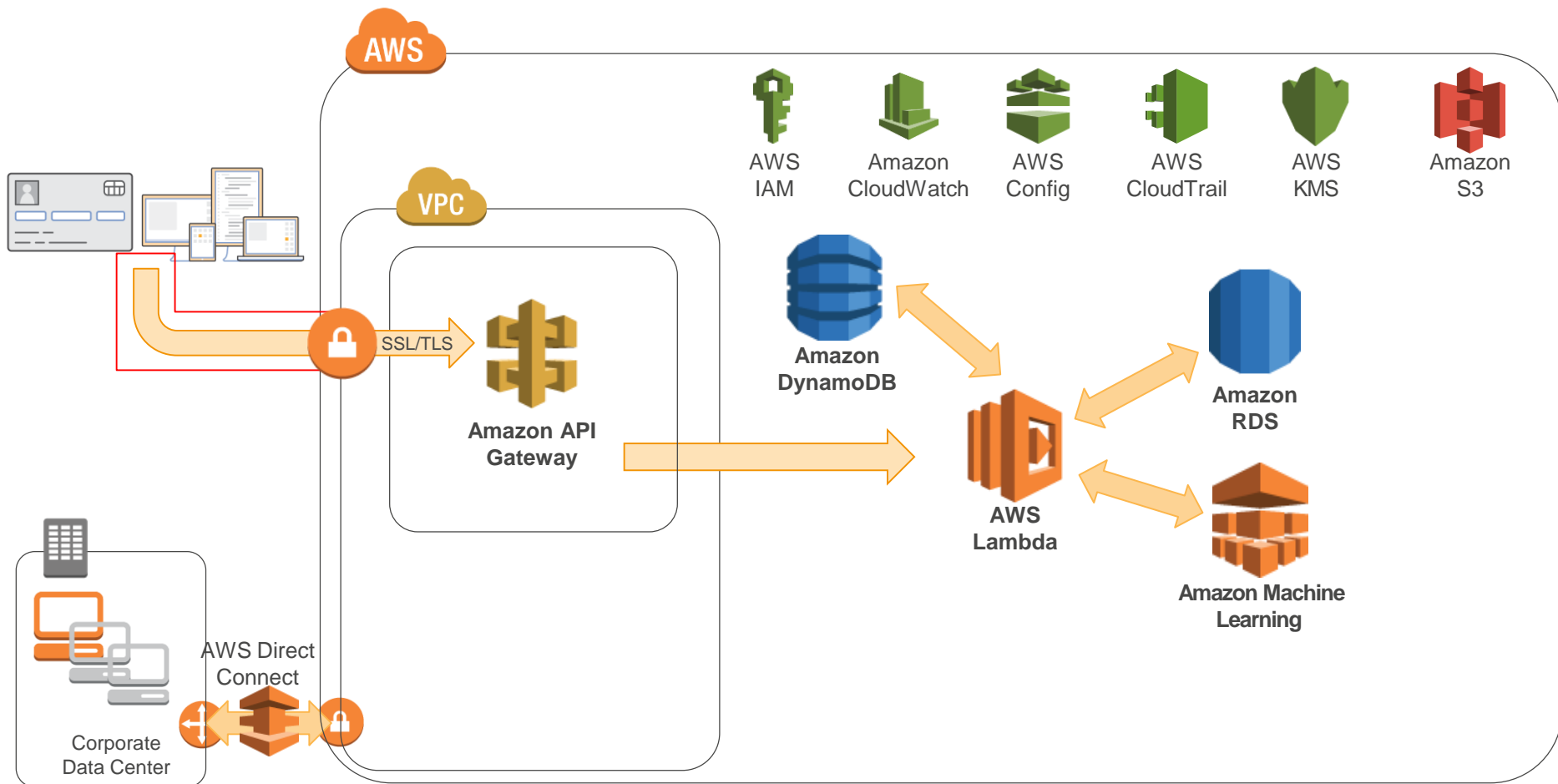**Speed:** Development down from months to days

Resources: Focus shift from management to development

Next Evolution of the Platform

Online Fraud Detection – Reference Architecture

Online Fraud Detection – Future State

Other Sessions on Machine Learning

CMP314 - Bringing Deep Learning to the Cloud with Amazon EC2

MAC206 - Machine Learning State of the Union

MAC303 - Developing Classification and Recommendation Engines with Amazon EMR and Apache Spark

Thank you!

Remember to complete your evaluations!