IN DREAMS BEGIN RESPONSIBILITIES

Data Science as a Service—using AI to Risk Stratify a Medicare Population and Build a Culture

Erich S. Huang, MD, PhD

erich.huang@duke.edu

Duke | FORGE
HEALTH DATA IN ACTION

Disclosures:

Founder, kēlaHealth (Startup)

Founder, Stratus Medicine (Startup)

Founder, MedBlue Data (Startup)

Duke | FORGE

Duke Forge promotes a culture of learning health by engaging partners to curate, analyze, and disseminate reliable and actionable information that leads to improved health for individuals and populations.

—Rob Califf, Amy Herring & Erich Huang

Duke | FORGE

Erich Huang, MD, PhD
School of Medicine

Amy Herring, PhD
Duke University

School of Medicine

Duke University

Focus on *Actionable* data problems in health

Bring the best methodological approaches to these problems

Focus on *Actionable* data problems in health

Bring the best methodological approaches to these problems

Clinicians & Health System Staff with deep subject matter expertise

Quantitative Faculty & Trainees with methodological expertise health

approaches to these problems

Clinicians & Health System Staff with deep subject matter expertise

Quantitative Faculty & Trainees with methodological expertise

Software Architecture & Engineering

Crucible

DATA SCIENCE ACCELERATED

Clinicians & Health System Staff with deep subject matter expertise

Quantitative Faculty & Trainees with methodological expertise

Software Architecture & Engineering

Motivated by a framework of *Value-Based* Healthcare Delivery and *addressing societal inequities* in health

Building Culture of Data Science Inquiry

Betty Jones

Duke | FORGE
HEALTH DATA IN ACTION

Betty Jones

- 65 year old woman
- previous medical history of a myocardial infarction
- history of cardiac bypass surgery
- congestive heart failure
- history of substance abuse
- rarely visits her doctor

Duke | FORGE
HEALTH DATA IN ACTION

- comes to the emergency room at Duke University Medical Center
- feels like she can't breathe
- she is diagnosed with pulmonary edema
- admitted to Duke Hospital for 10 days and discharged
- *WHAT HAPPENS NEXT?*

$$\text{VALUE} = \frac{\text{HEALTH OUTCOMES}}{\text{COST}}$$

Michael Porter, Harvard School of Business

$ ~ VALUE = HEALTH OU
COS

Michael

Duke | FORGE
HEALTH DATA IN ACTION

$\underbrace{\$ \sim \text{VALUE}}_{}$

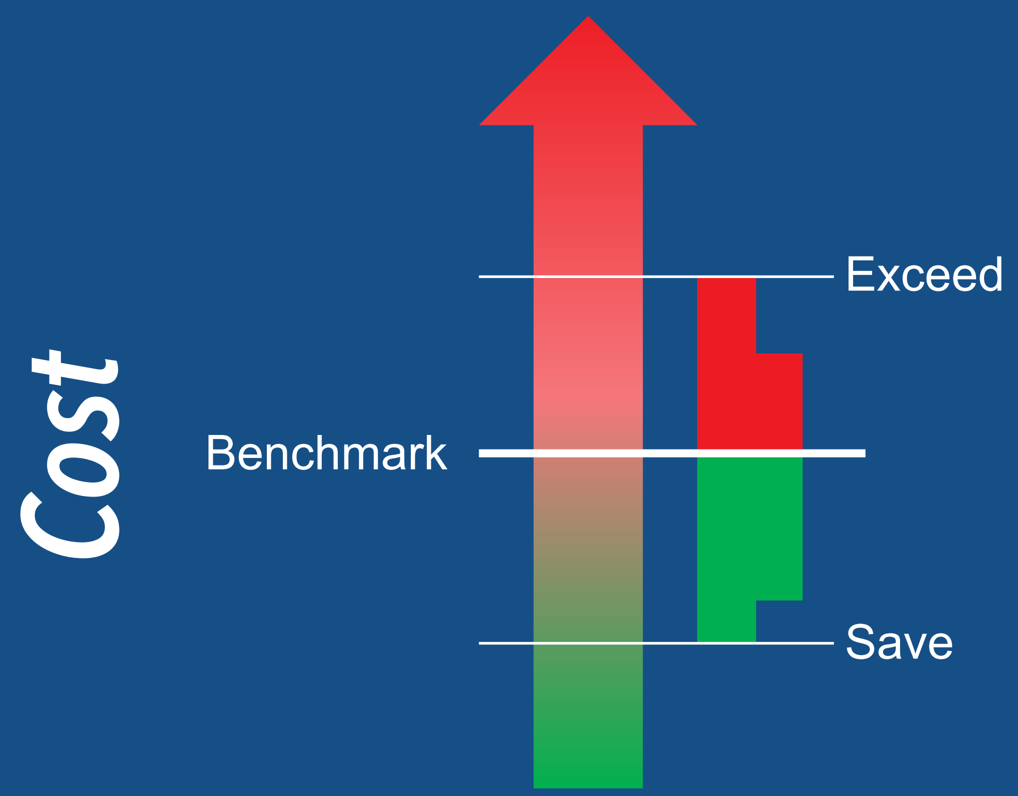
Value-Based Reimbursement Models

Duke | FORGE
HEALTH DATA IN ACTION

Value-Based Reimbursement Models

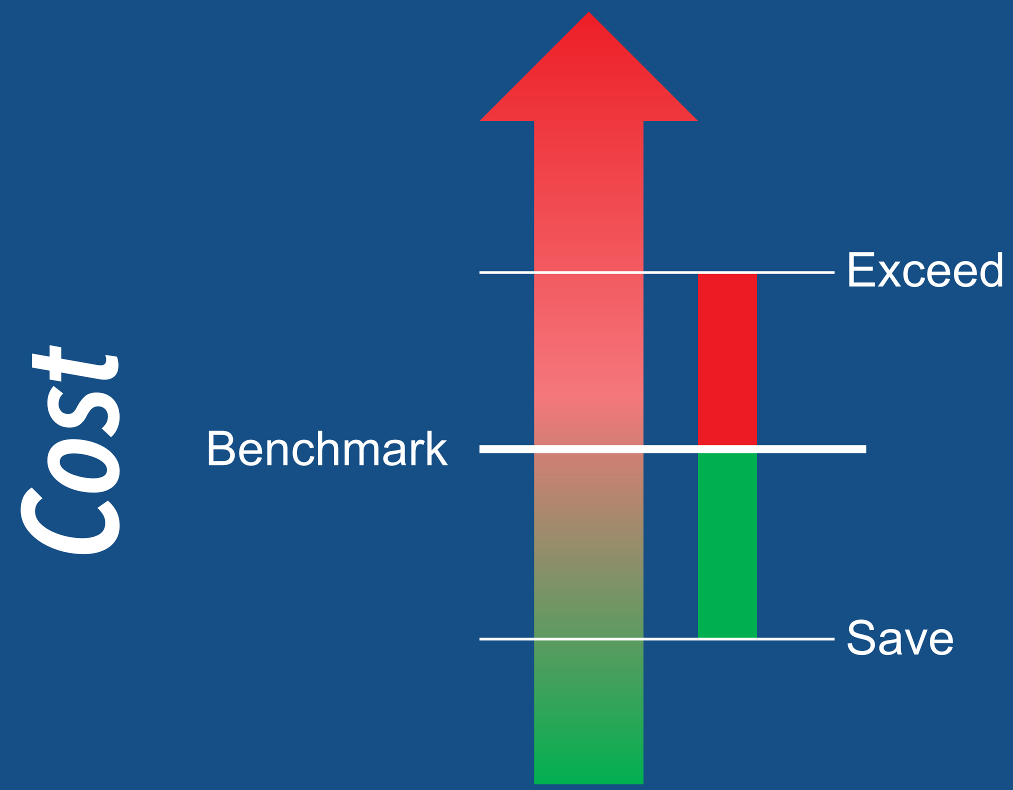
"Obamacare" Medicare Shared Savings Program

"Obamacare" Medicare Shared Savings Program
"Accountable Care Organization (ACO)"



ACO shares 40%-75% of **losses** based on quality performance

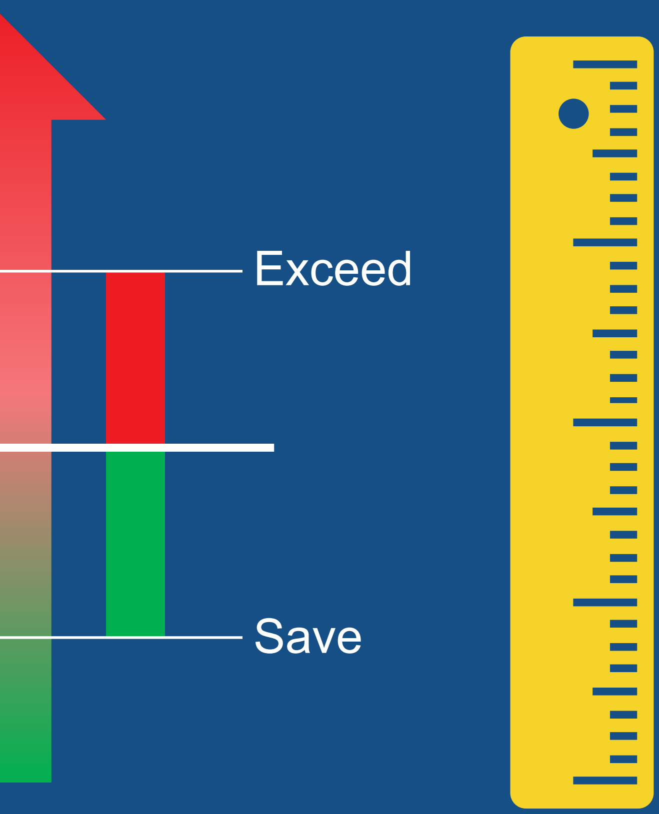
ACO shares up to 75% of **savings**



We are assuming "upside" and "downside" risk for healthcare delivery

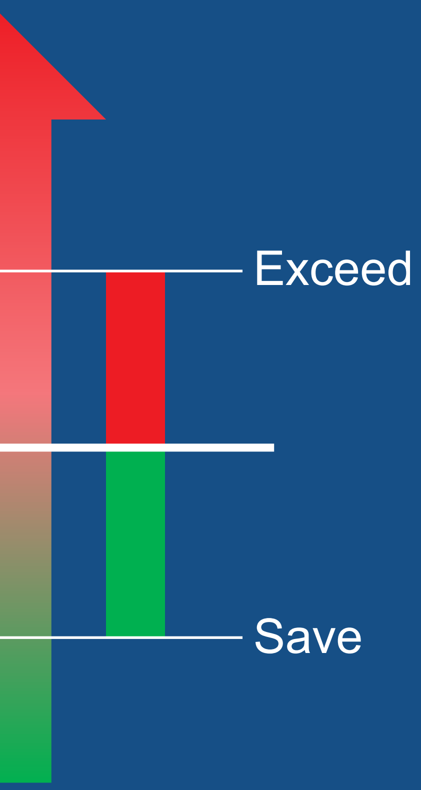
Duke | FORGE
HEALTH DATA IN ACTION

risk



Assess & Quantify Upside/Downside

risk

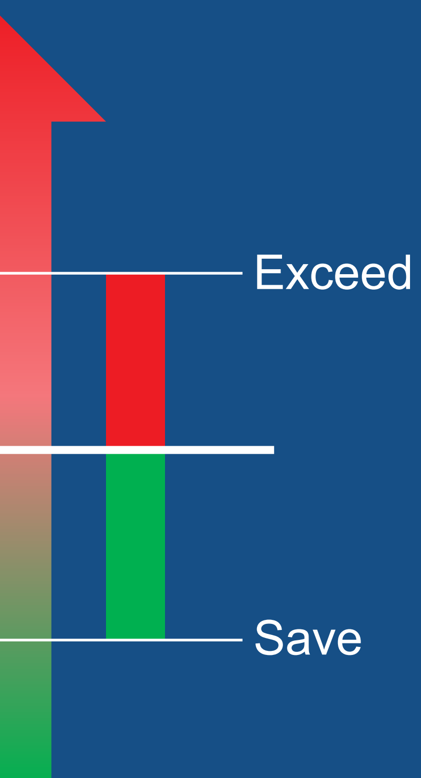


Assess & Quantify Upside/Downside

our ACO saved the US $22M last year and kept $9.5M upside

Hospitalizations

Duke's ACO: "Duke Connected Care"



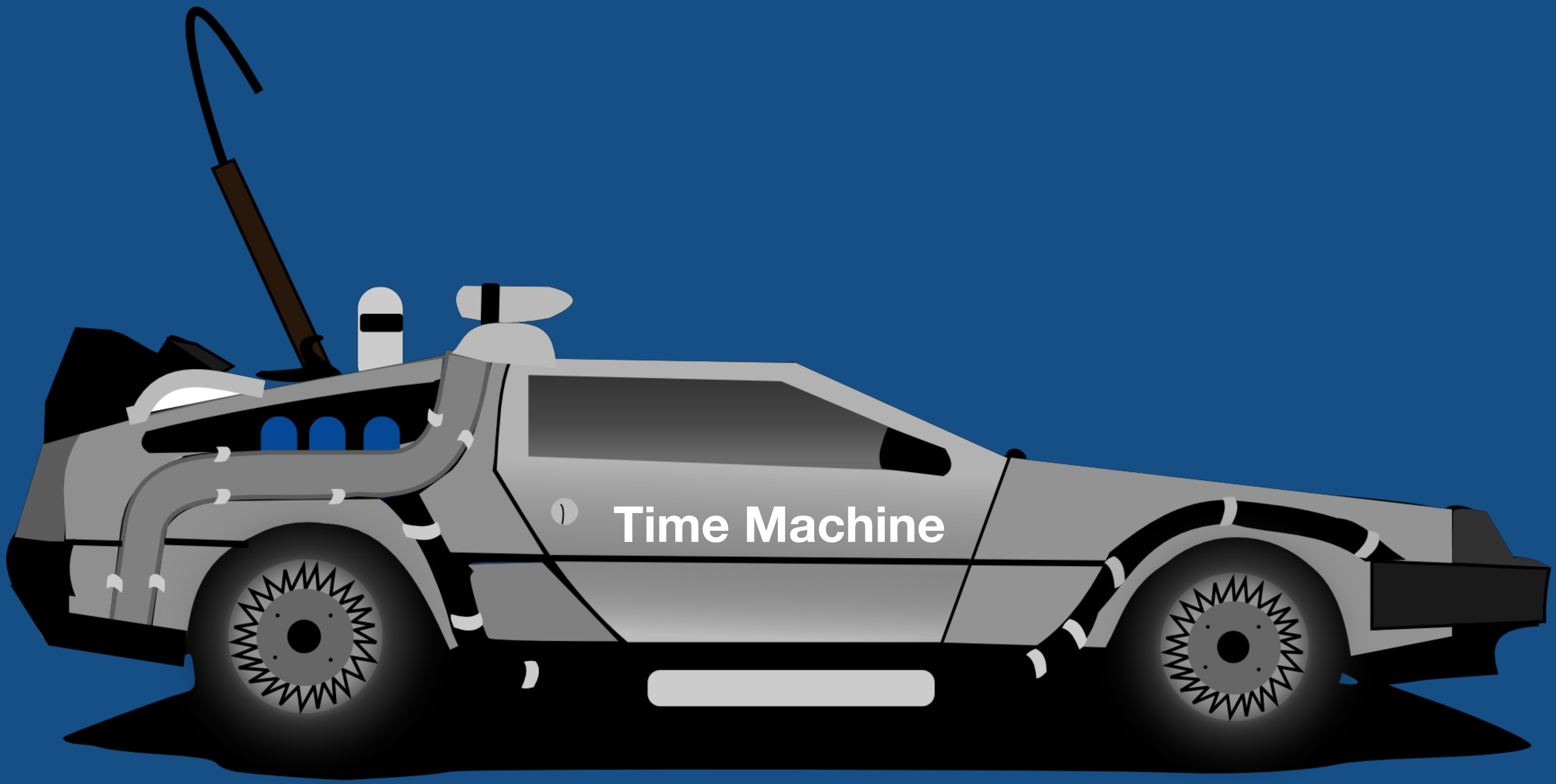
30%

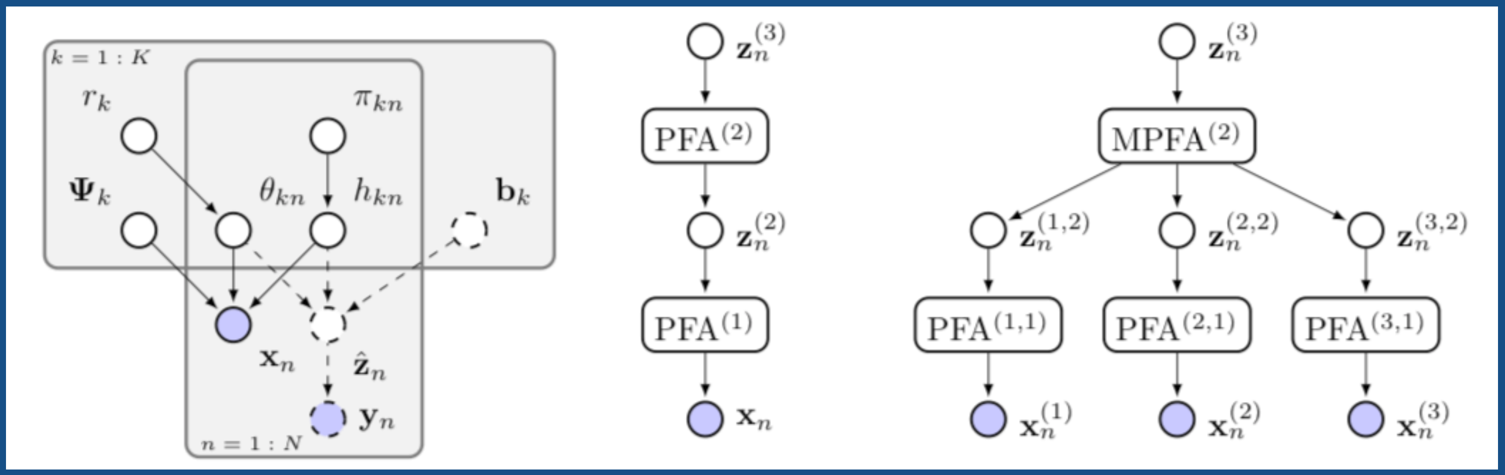
admission rate 7/17-7/18

15%

readmission rate 7/17-7/18

Betty Jones



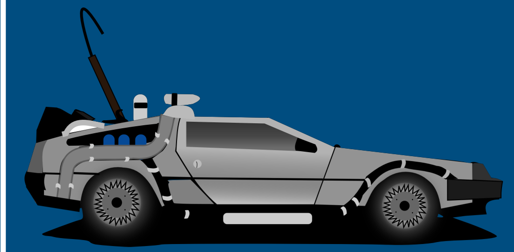


Deep Poisson Factor Model Architecture

Duke | FORGE
HEALTH DATA IN ACTION

54,000 Medicare Patients

BEGINNING OF MONTH

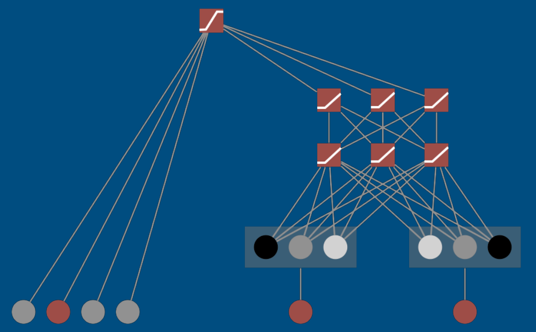


12-MONTH RETROSPECTIVE CMS CLAIMS DATA

12-MONTH RETROSPECTIVE EHR (COMMON DATA MODEL) DATA

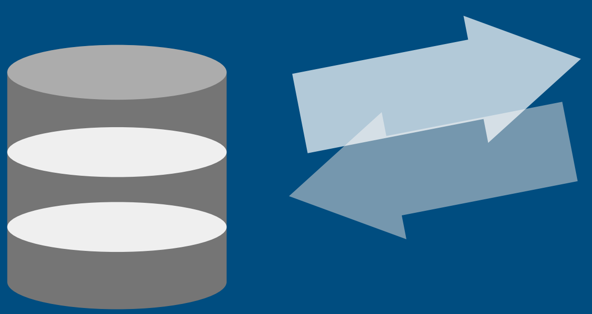
6 MONTH PREDICTION WINDOW

Probability of Unplanned Admissions for:
Any Cause & **31 Diagnostic Categories**
Absolute and Percentile Rank



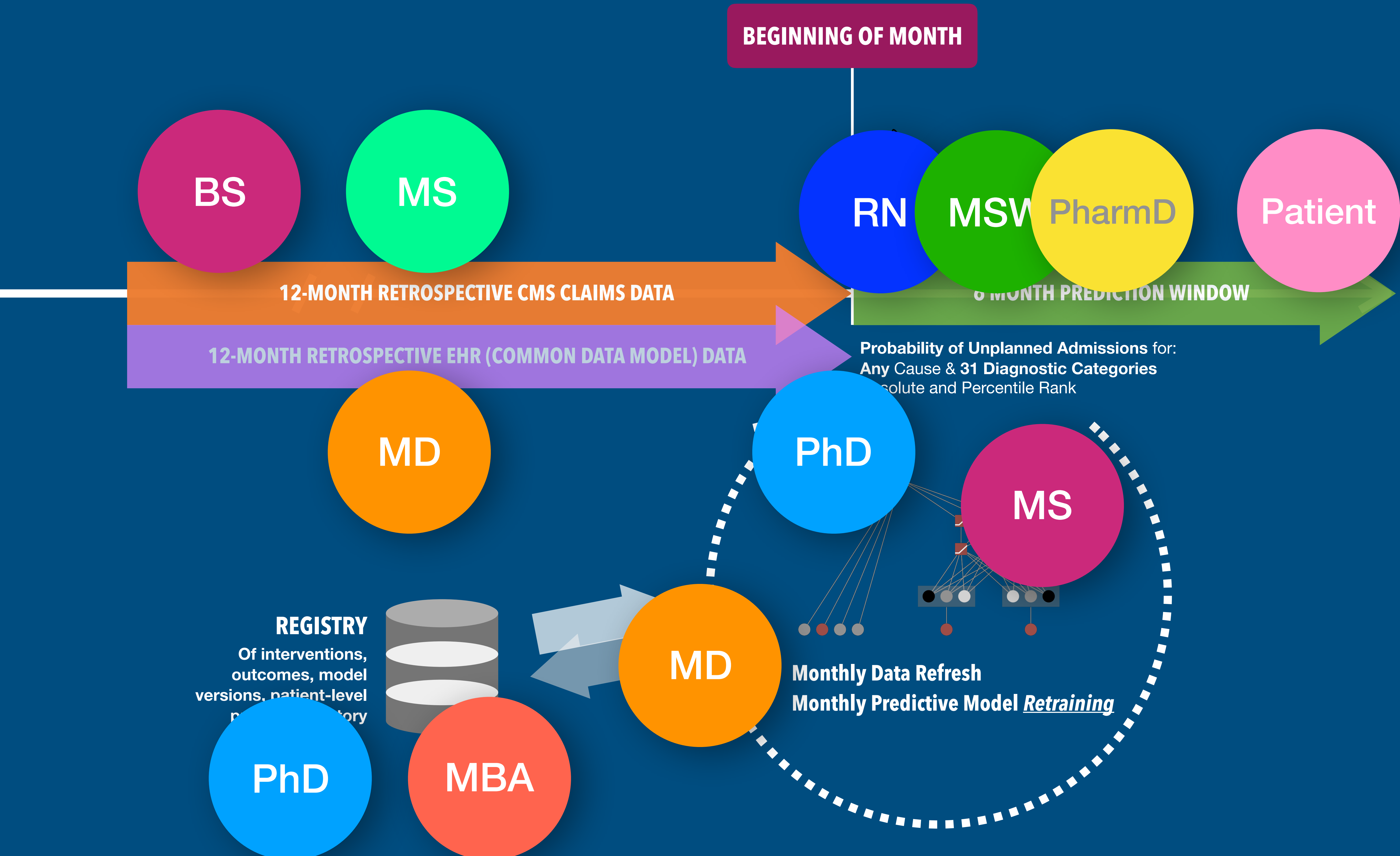
Monthly Data Refresh
Monthly Predictive Model *Retraining*

REGISTRY
Of interventions, outcomes, model versions, patient-level prediction history

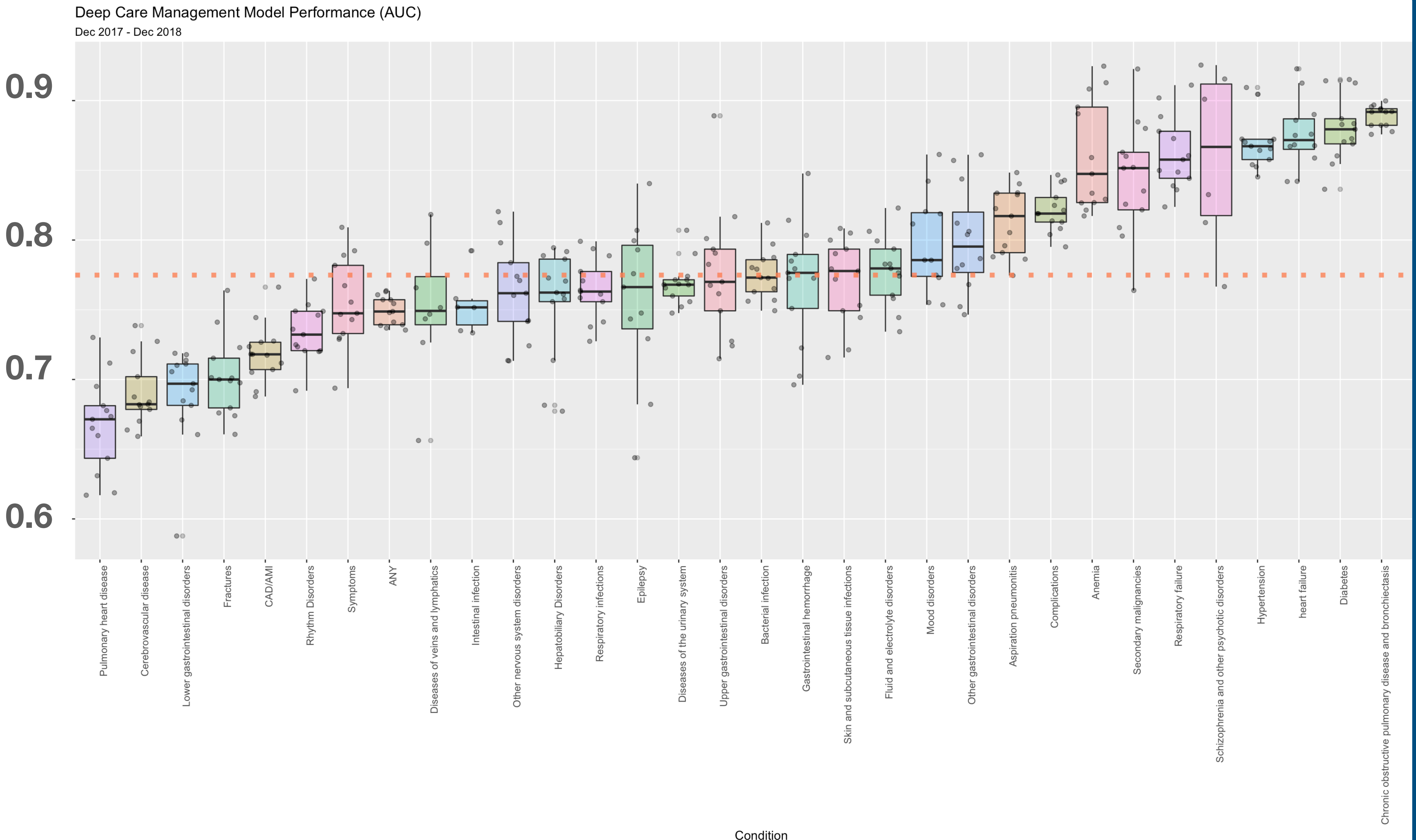


Independent Evaluation

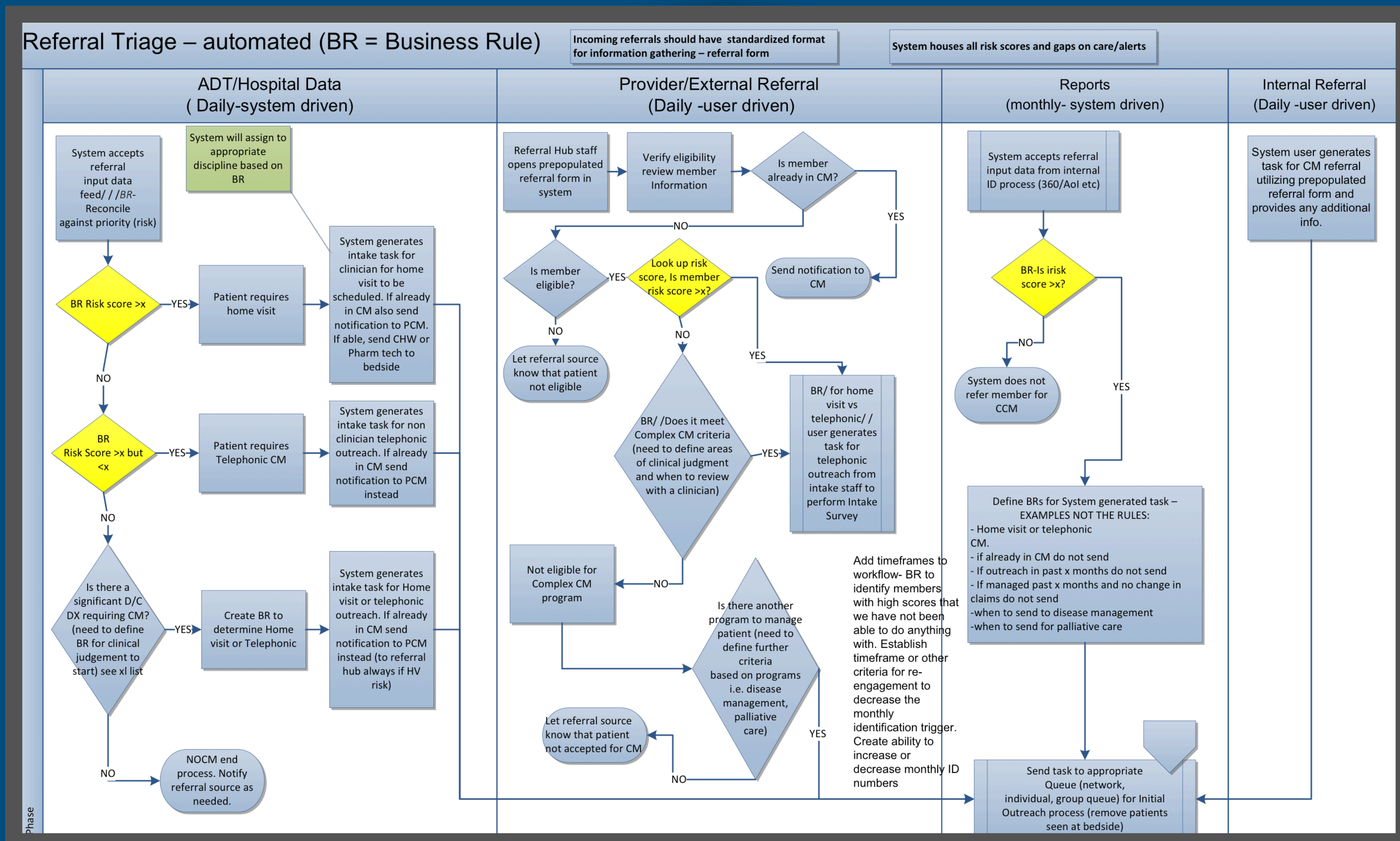
Duke | FORGE
HEALTH DATA IN ACTION


BEGINNING OF MONTH
BS
MS
RN
MSW
PharmD
Patient
12-MONTH RETROSPECTIVE CMS CLAIMS DATA
12-MONTH RETROSPECTIVE EHR (COMMON DATA MODEL) DATA
PREDICTION WINDOW
Probability of Unplanned Admissions for:
Any Cause & 31 Diagnostic Categories
and Percentile Rank
MD
PhD
MS
REGISTRY
Of interventions, outcomes, model versions, patient-level
MD
Monthly Data Refresh
Monthly Predictive Model *Retraining*
PhD
MBA

Boxplot of monthly AUCs for 31 Diagnostic Categories over 12 months



A patient & clinician-centric project & workflow



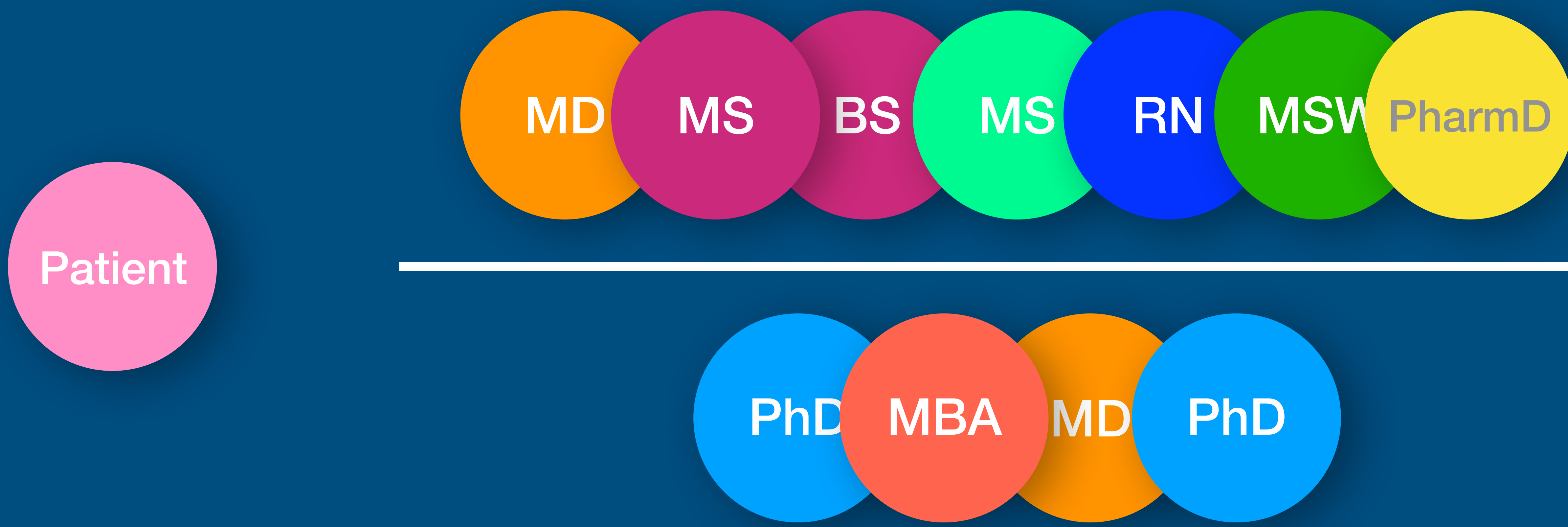
Duke | FORGE HEALTH DATA IN ACTION

SCAN AND PREDICT RISK FOR 32 DIAGNOSES ACROSS 52000 PATIENTS EVERY MONTH

RISK OF HOSPITALIZATION IN THE NEXT 6 MONTHS	DUKE CONNECTED CARE CARE MANAGEMENT TEAM	INTERVENTIONS
Any Cause	above risk threshold	home visit by pharmacy tech, $ assistance for meds. social work assistance with transportation
Heart Attack		
Congestive Heart Failure	above risk threshold	
Pneumonia		
Psychosis		
⋮		
Hip Fracture		

https://pxhere.com/en/photo/861253

A patient & clinician-centric project & workflow



A high “real world user” to “PhD” quotient

Duke | FORGE
HEALTH DATA IN ACTION

jupyter

NVIDIA
CUDA

S

git

R

redhat
LINUX

RUNNER

python

pcornet

openstack

APACHE
Spark

ORACLE
EXADATA

TensorFlow

kubernetes

Sure we’re building apps, but we’re really building ***ecosystems***…

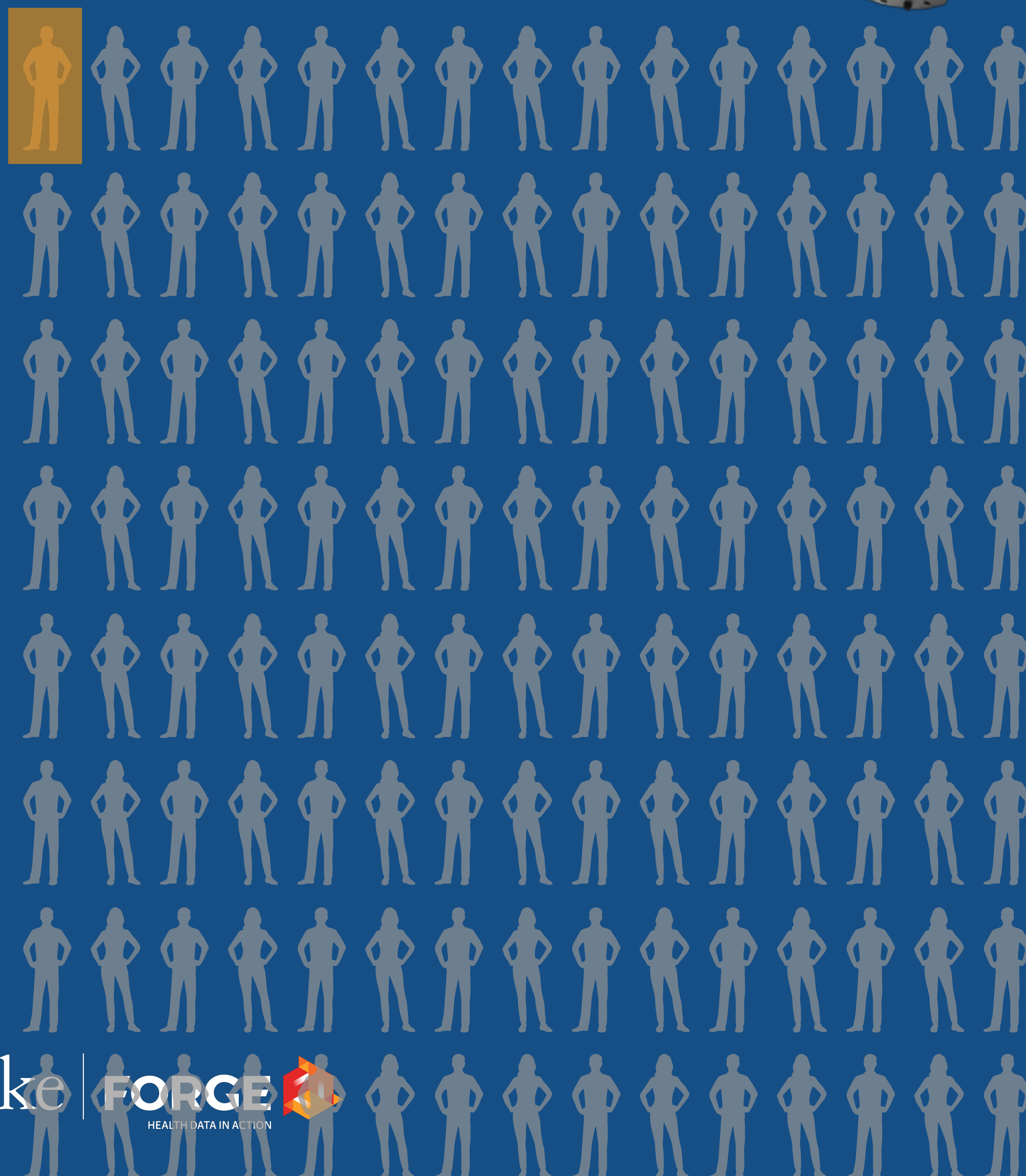
If we're trying to build ***ecosystems***…

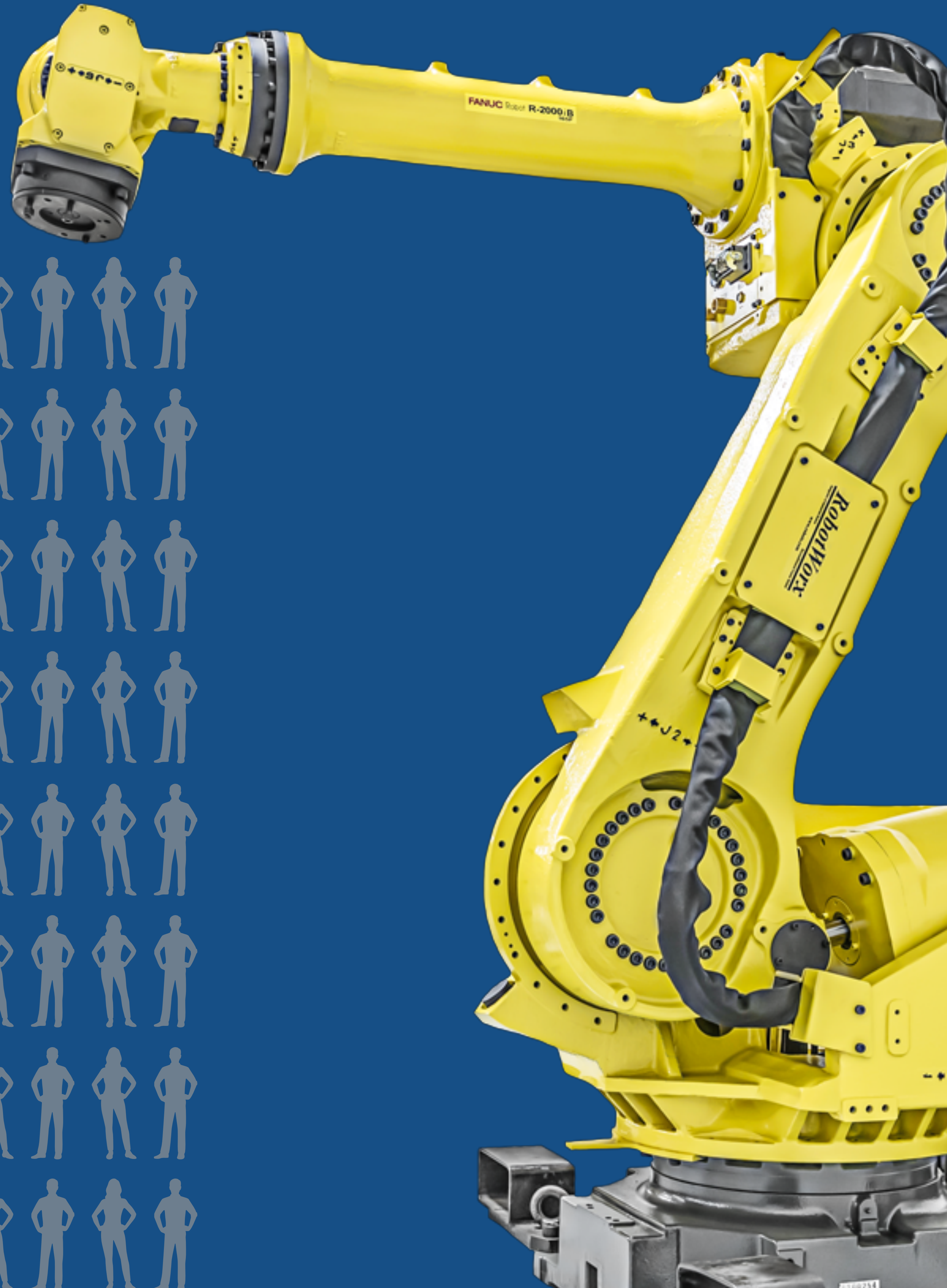
…then an EHR needs to be evaluated by whether it is truly participatory in this ecosystem.

Or, we need to remediate its deficiencies

Duke | FORGE
HEALTH DATA IN ACTION

54,000 Medicare Patients

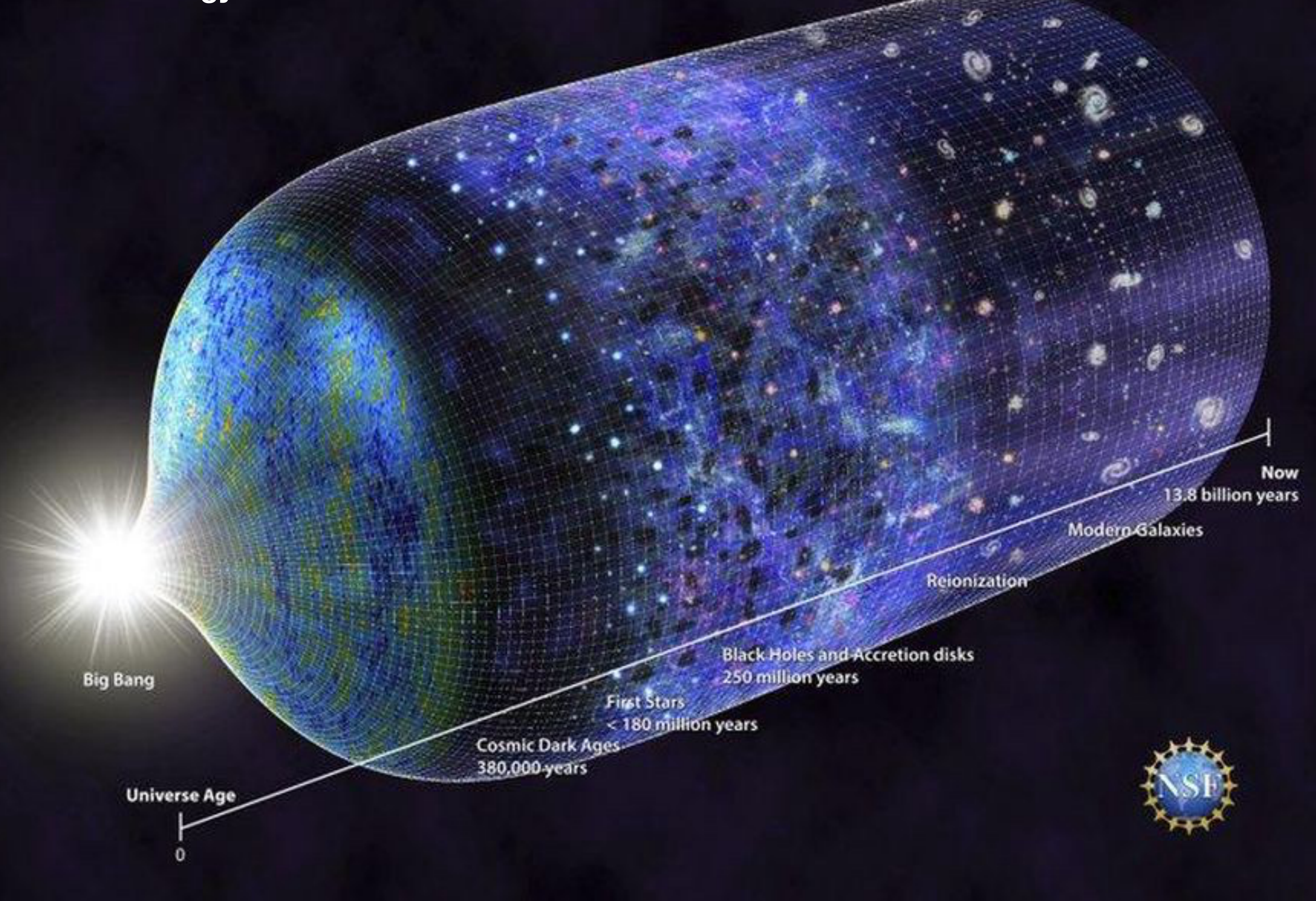


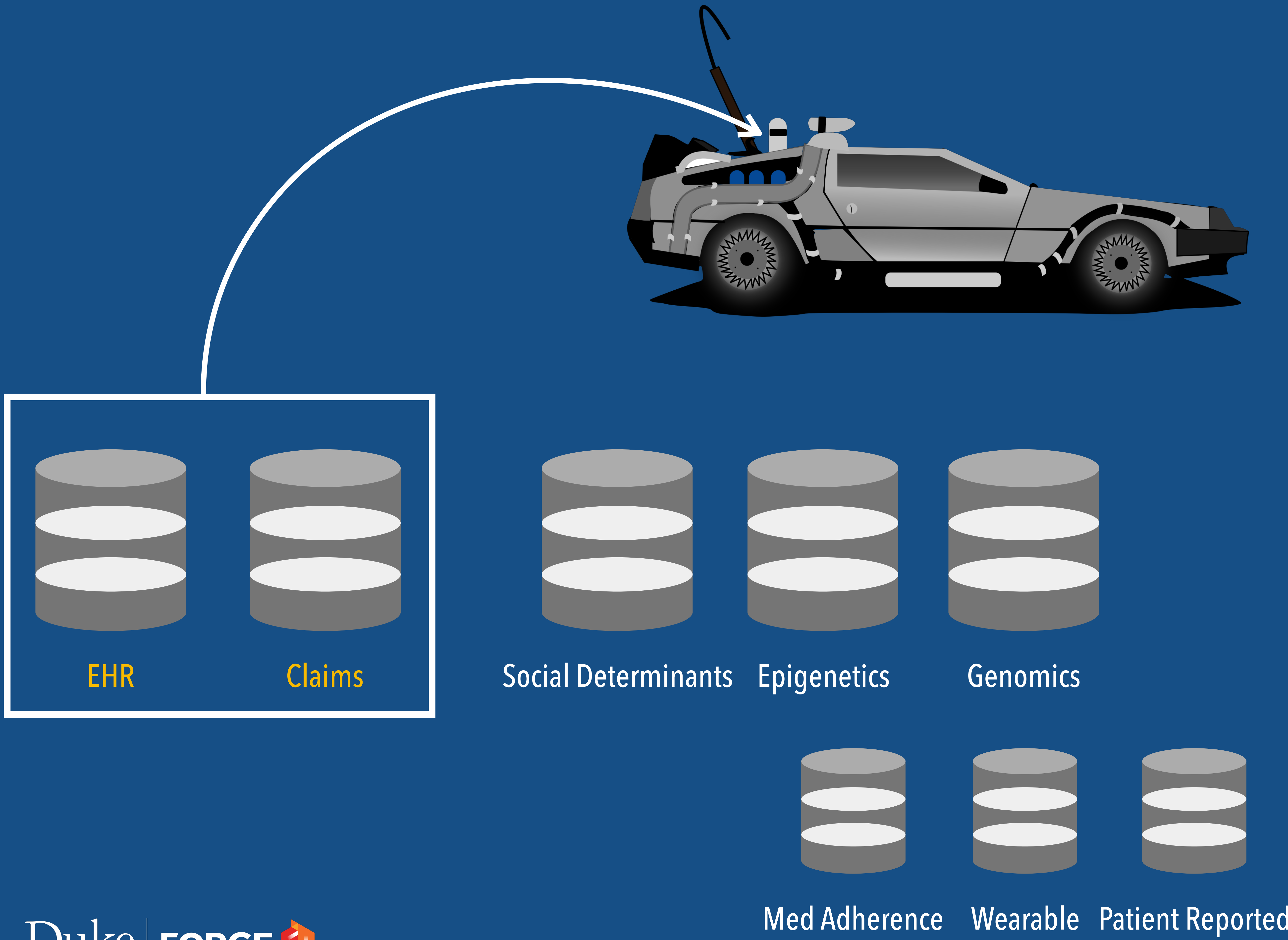


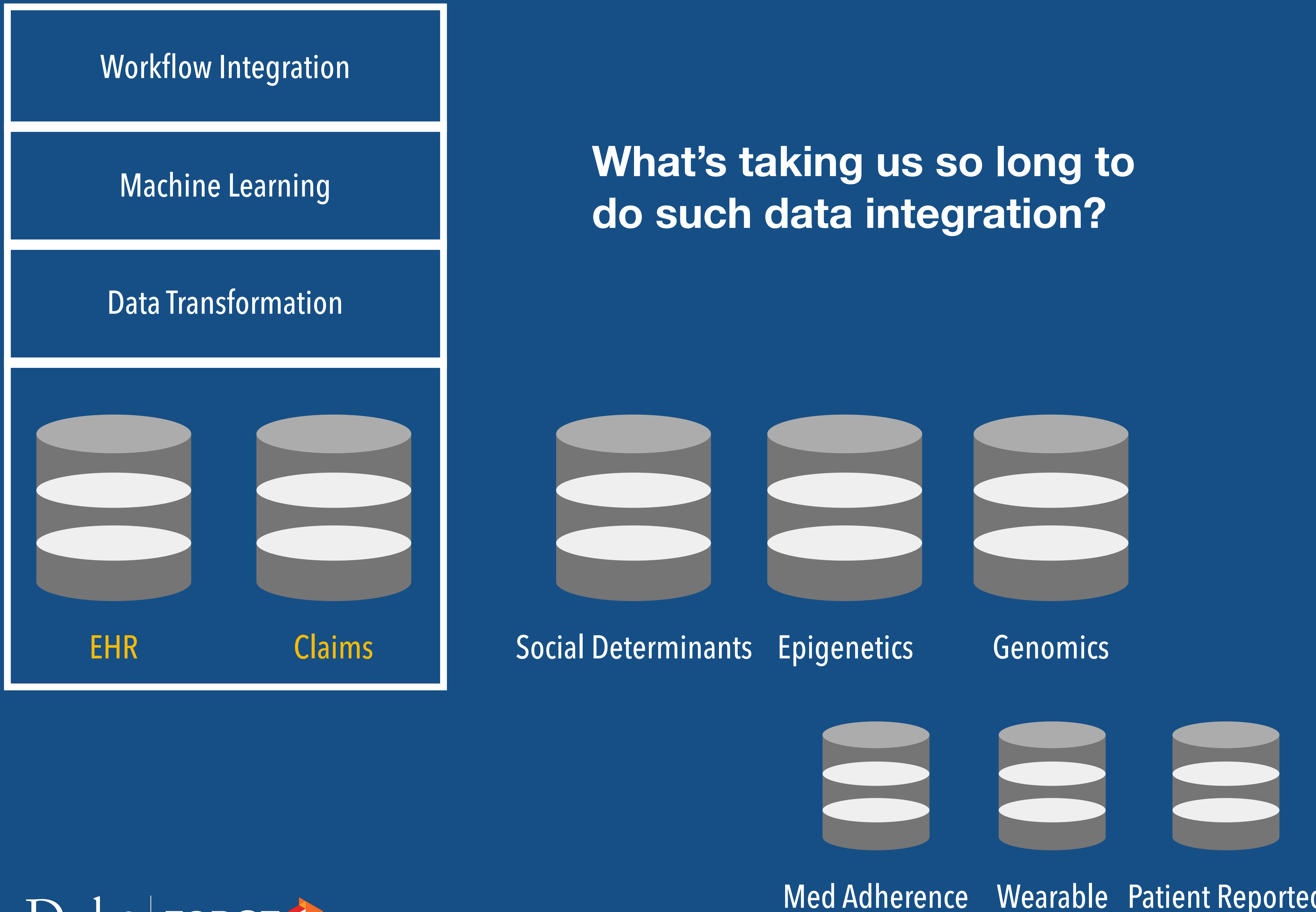
Duke | FORGE
HEALTH DATA IN ACTION

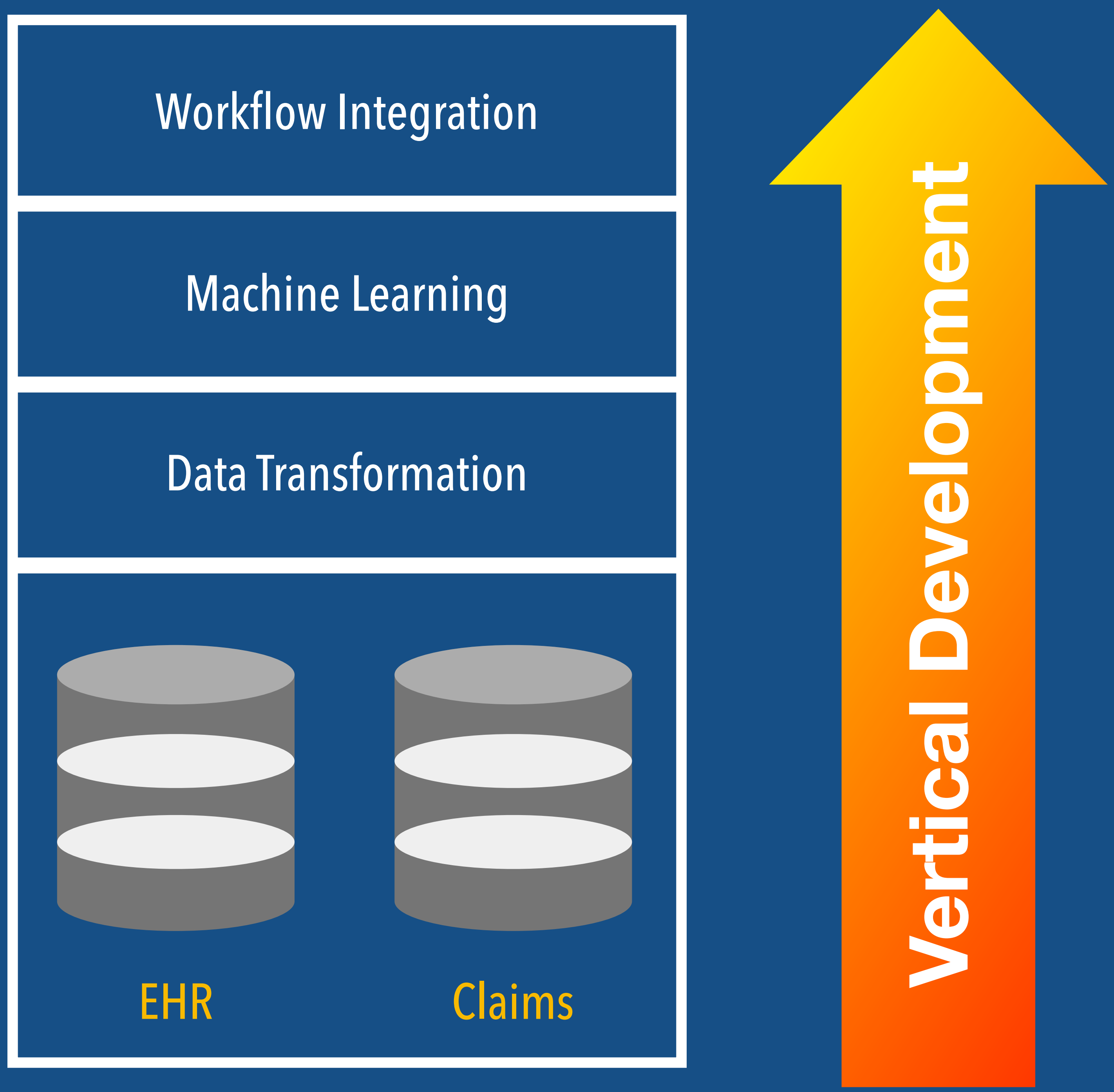
What is the data platform that gets us there?

Dark Energy & Matter




EHR
Claims
Social Determinants
Epigenetics
Genomics
Med Adherence
Wearable
Patient Reported


Workflow Integration
Machine Learning
Data Transformation
EHR
Claims
What’s taking us so long to do such data integration?
Social Determinants
Epigenetics
Genomics
Med Adherence
Wearable
Patient Reported



What's taking us so long to do such data integration?

Duke | FORGE
HEALTH DATA IN ACTION

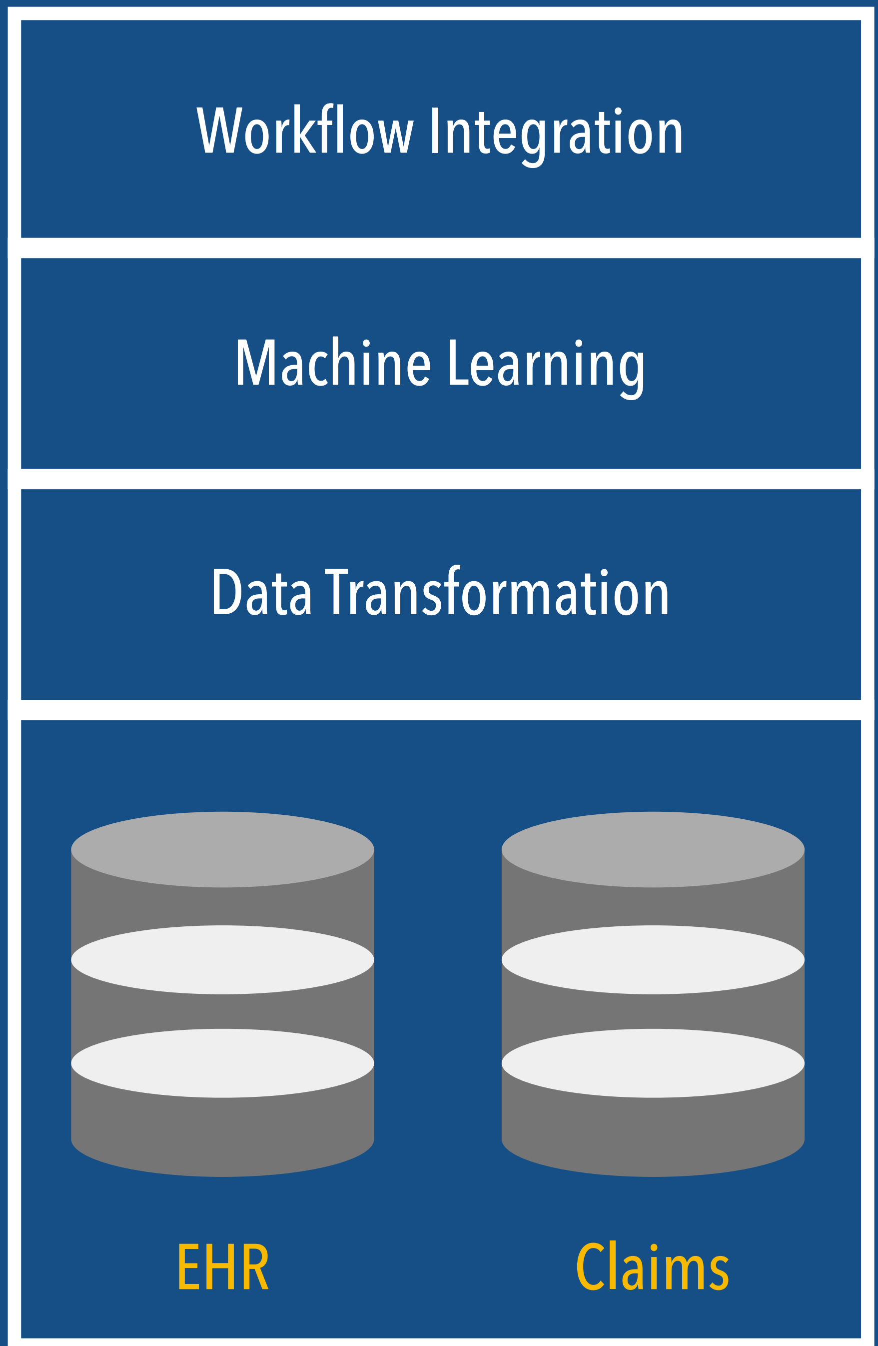
What's taking us so long to do such data integration?

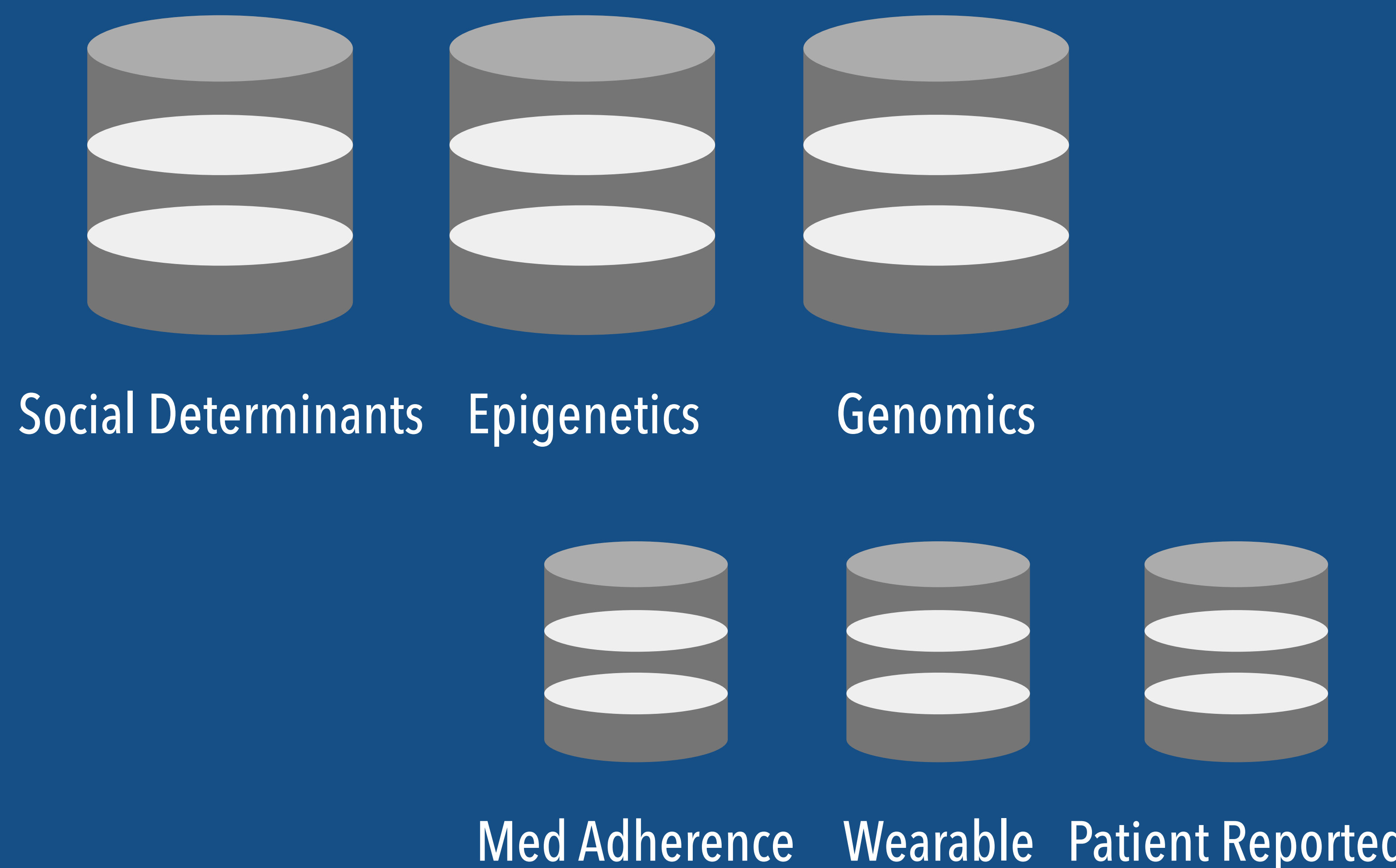
Duke | FORGE
HEALTH DATA IN ACTION

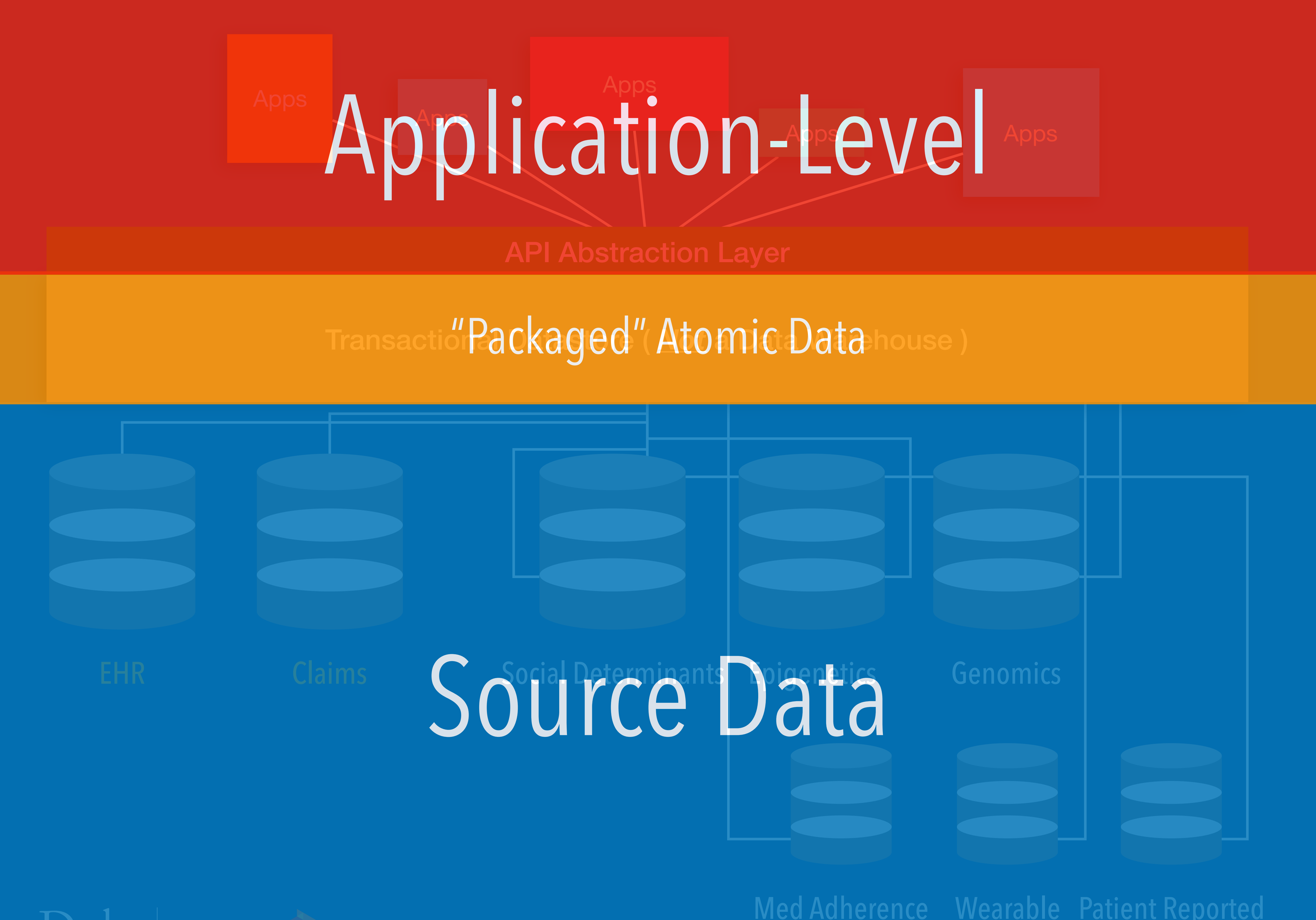
"There's a lot of undifferentiated heavy lifting that stands between your idea and that success… 70% of your time, energy, and dollars go into the undifferentiated heavy lifting and only 30% of your energy, time, and dollars gets to go into the core kernel of your idea."

–Jeff Bezos

What’s taking us so long to do such data integration?

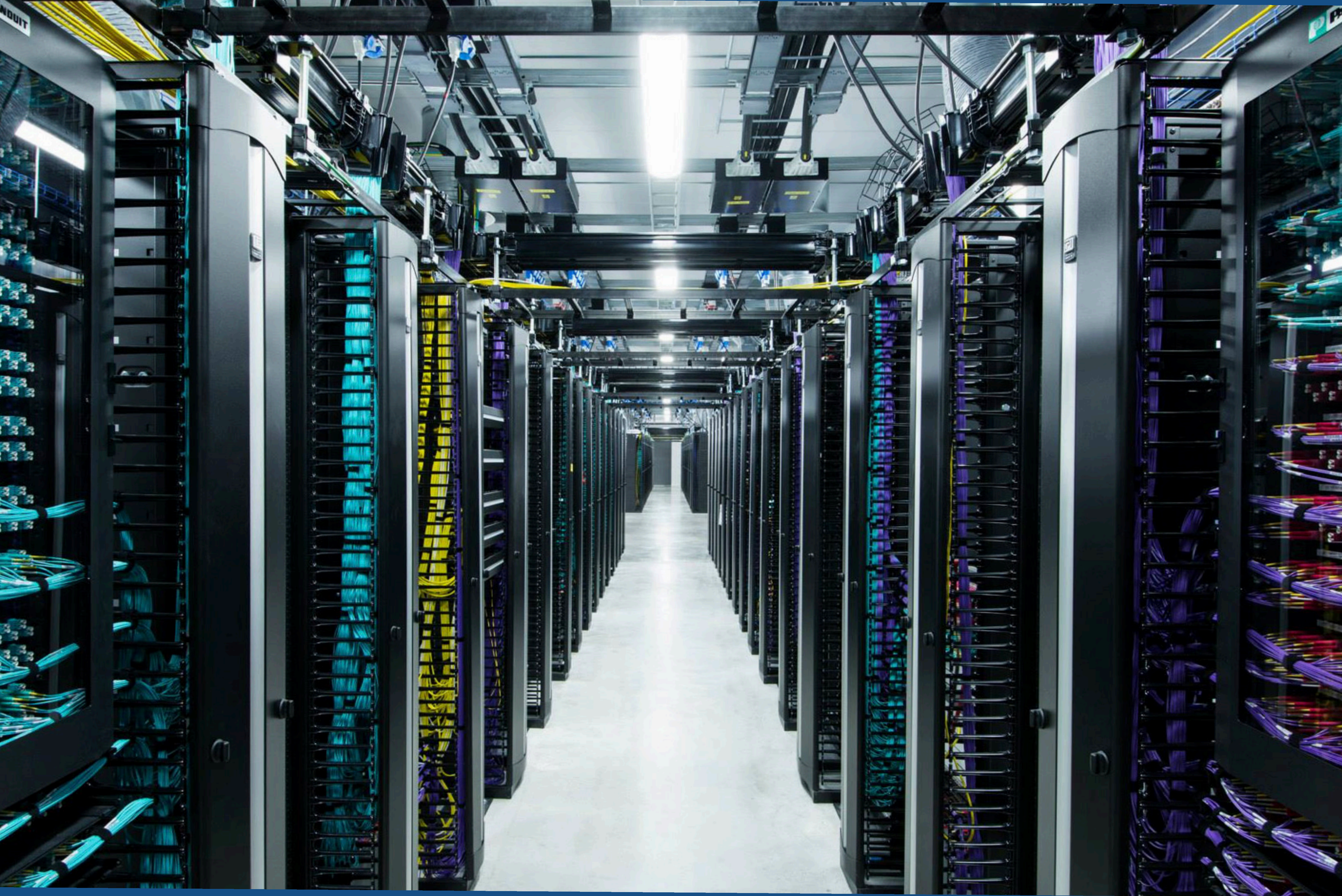






Duke | FORGE
HEALTH DATA IN ACTION

The Cloud



Duke | FORGE
HEALTH DATA IN ACTION

https://forge.duke.edu/blog/what-health-data-science-and-raising-chickens-have-common

What Health Data Science and Raising Chickens Have in Common

Photo credit: Manfred Richter via Pixabay

April 29, 2019 **BY: ERICH S. HUANG, MD, PHD**

In my most recent blog post, I wrote about Susan Jones*, a 65 year-old patient and Medicare recipient, and described how a multidisciplinary team at Duke constructed an AI-powered workflow to help patients like her.

Building that workflow was a laborious, 18-month-long process. We assembled a band of explorers comprising clinicians, flatiron

Cancer is smart. Together, we can be smarter.

Accelerating the fight against cancer requires the entire industry to work together. Our products connect community oncologists, academics, hospitals, life science researchers and regulators on a shared technology platform. Together, we can learn from the experience of every patient.

MARKETS
Big Water Companies Set to Merge

MARKETS
Do You Own Bitcoin? The IRS Is Coming for You

BUSINESS
Cancer Panel Calls for Measures to Lower Drug Costs

OPINION
The GOP's Internet Tax

WORLD N
The Dark
Europe's
System

USINESS

Roche to Acquire Healthcare-Software Company Flatiron for $1.9 Billion

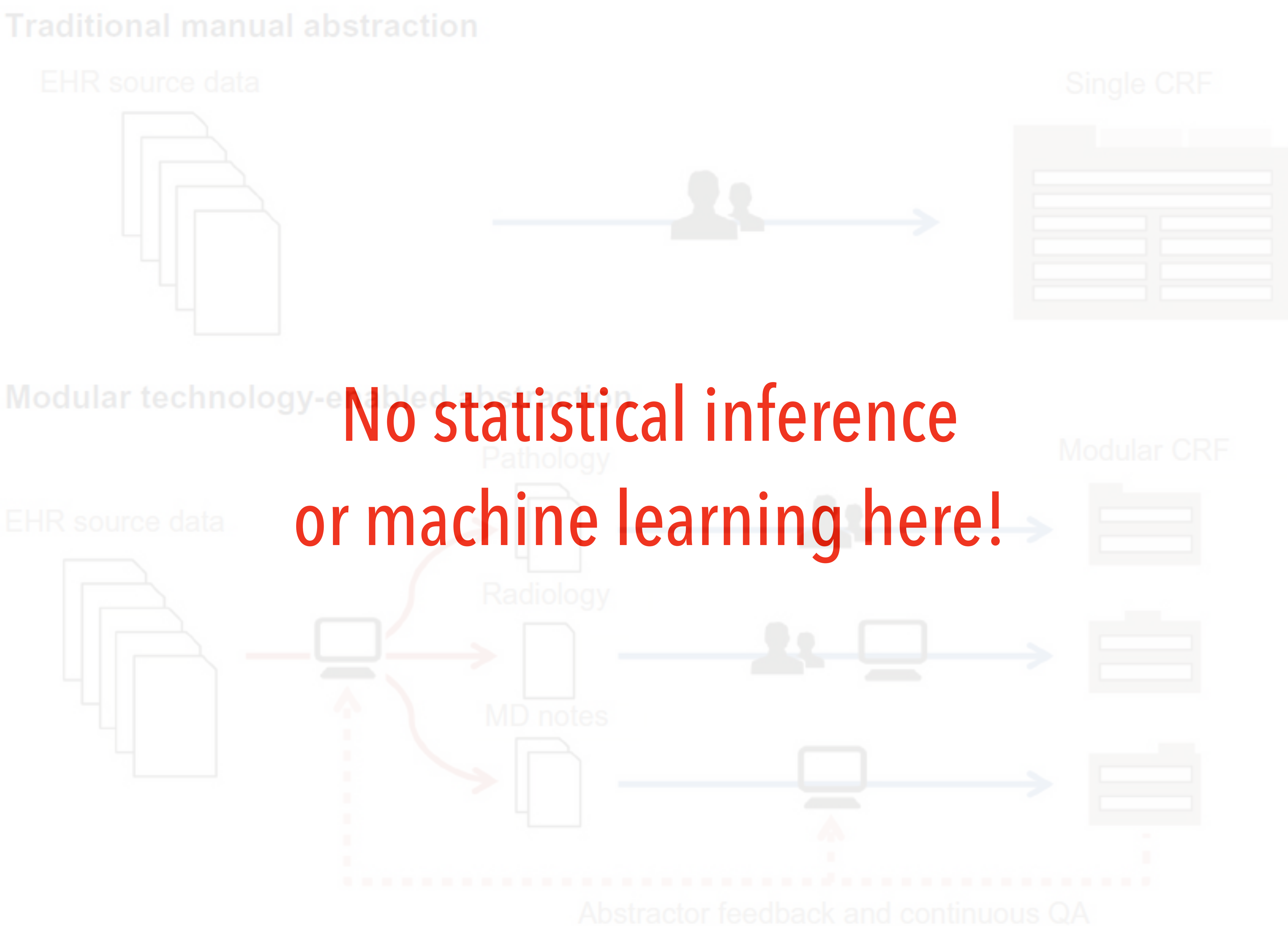
Roche led a round of financing for Flatiron and already owns 12.6% of the company

By *Aisha Al-Muslim*
Feb. 15, 2018 5:13 p.m. ET

Pharmaceuticals firm Roche Holding AG has agreed to buy the shares it doesn't already own of Flatiron Health Inc., an oncology software company, for $1.9 billion, the companies said Thursday.

Most Popular Videos

1. Meet the Man Who Has Spent $70,000 Playing a Mobile



No statistical inference
or machine learning here!

flatiron

npj | Digital Medicine

www.nature.com/npjdigitalmed

ARTICLE OPEN

Scalable and accurate deep learning with electronic health records

Alvin Rajkomar[1,2], Eyal Oren[1], Kai Chen[1], Andrew M. Dai[1], Nissan Hajaj[1], Michaela Hardt[1], Peter J. Liu[1], Xiaobing Liu[1], Jake Marcus[1], Mimi Sun[1], Patrik Sundberg[1], Hector Yee[1], Kun Zhang[1], Yi Zhang[1], Gerardo Flores[1], Gavin E. Duggan[1], Jamie Irvine[1], Quoc Le[1], Kurt Litsch[1], Alexander Mossin[1], Justin Tansuwan[1], De Wang[1], James Wexler[1], Jimbo Wilson[1], Dana Ludwig[2], Samuel L. Volchenboum[3], Katherine Chou[1], Michael Pearson[1], Srinivasan Madabushi[1], Nigam H. Shah[4], Atul J. Butte[2], Michael D. Howell[1], Claire Cui[1], Greg S. Corrado[1] and Jeffrey Dean[1]

Predictive modeling with electronic health record (EHR) data is anticipated to drive personalized medicine and improve healthcare quality. Constructing predictive statistical models typically requires extraction of curated predictor variables from normalized EHR data, a labor-intensive process that discards the vast majority of information in each patient's record. We propose a representation of patients' entire raw EHR records based on the Fast Healthcare Interoperability Resources (FHIR) format. We demonstrate that deep learning methods using this representation are capable of accurately predicting multiple medical events from multiple centers without site-specific data harmonization. We validated our approach using de-identified EHR data from two US academic medical centers with 216,221 adult patients hospitalized for at least 24 h. In the sequential format we propose, this volume of EHR data unrolled into a total of 46,864,534,945 data points, including clinical notes. Deep learning models achieved high accuracy for tasks such as predicting: in-hospital mortality (area under the receiver operator curve [AUROC] across sites 0.93–0.94), 30-day unplanned readmission (AUROC 0.75–0.76), prolonged length of stay (AUROC 0.85–0.86), and all of a patient's final discharge diagnoses (frequency-weighted AUROC 0.90). These models outperformed traditional, clinically-used predictive models in all cases. We believe that this approach can be used to create accurate and scalable predictions for a variety of clinical scenarios. In a case study of a particular prediction, we demonstrate that neural networks can be used to identify relevant information from the patient's chart.

npj Digital Medicine (2018)1:18 ; doi:10.1038/s41746-018-0029-1

INTRODUCTION

The promise of digital medicine stems in part from the hope that, by digitizing health data, we might more easily leverage computer information systems to understand and improve care. In fact, routinely collected patient healthcare data are now approaching the genomic scale in volume and complexity.[1] Unfortunately, most of this information is not yet used in the sorts of predictive statistical models clinicians might use to improve care delivery. It is widely suspected that use of such efforts, if successful, could provide major benefits not only for patient safety and quality but also in reducing healthcare costs.[2–6]

In spite of the richness and potential of available data, scaling the development of predictive models is difficult because, for traditional predictive modeling techniques, each outcome to be predicted requires the creation of a custom dataset with specific variables.[7] It is widely held that 80% of the effort in an analytic model is preprocessing, merging, customizing, and cleaning datasets,[8,9] not analyzing them for insights. This profoundly limits the scalability of predictive models.

Another challenge is that the number of potential predictor variables in the electronic health record (EHR) may easily number in the thousands, particularly if free-text notes from doctors, nurses, and other providers are included. Traditional modeling approaches have dealt with this complexity simply by choosing a very limited number of commonly collected variables to consider.[7] This is problematic because the resulting models may produce imprecise predictions: false-positive predictions can overwhelm physicians, nurses, and other providers with false alarms and concomitant alert fatigue,[10] which the Joint Commission identified as a national patient safety priority in 2014.[11] False-negative predictions can miss significant numbers of clinically important events, leading to poor clinical outcomes.[11,12] Incorporating the entire EHR, including clinicians' free-text notes, offers some hope of overcoming these shortcomings but is unwieldy for most predictive modeling techniques.

Recent developments in deep learning and artificial neural networks may allow us to address many of these challenges and unlock the information in the EHR. Deep learning emerged as the preferred machine learning approach in machine perception problems ranging from computer vision to speech recognition, but has more recently proven useful in natural language processing, sequence prediction, and mixed modality data settings.[13–17] These systems are known for their ability to handle large volumes of relatively messy data, including errors in labels

[1]Google Inc, Mountain View, CA, USA; [2]University of California, San Francisco, San Francisco, CA, USA; [3]University of Chicago Medicine, Chicago, IL, USA and [4]Stanford University, Stanford, CA, USA
Correspondence: Alvin Rajkomar (alvinrajkomar@google.com)
These authors contributed equally: Alvin Rajkomar, Eyal Oren

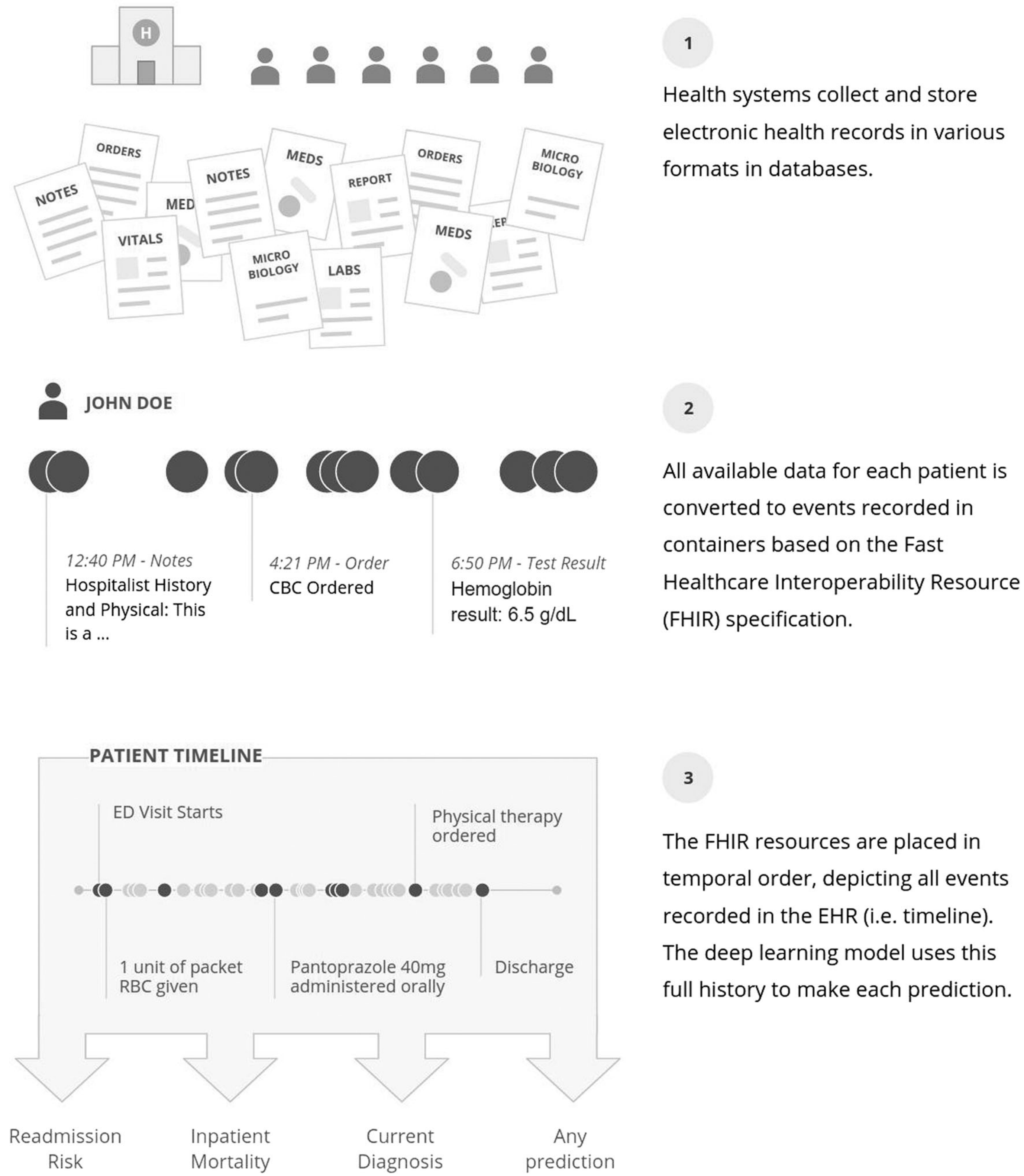


Fig. 4 Data from each health system were mapped to an appropriate FHIR (Fast Healthcare Interoperability Resources) resource and placed in temporal order. This conversion did not harmonize or standardize the data from each health system other than map them to the appropriate resource. The deep learning model could use all data available prior to the point when the prediction was made. Therefore, each prediction, regardless of the task, used the same data

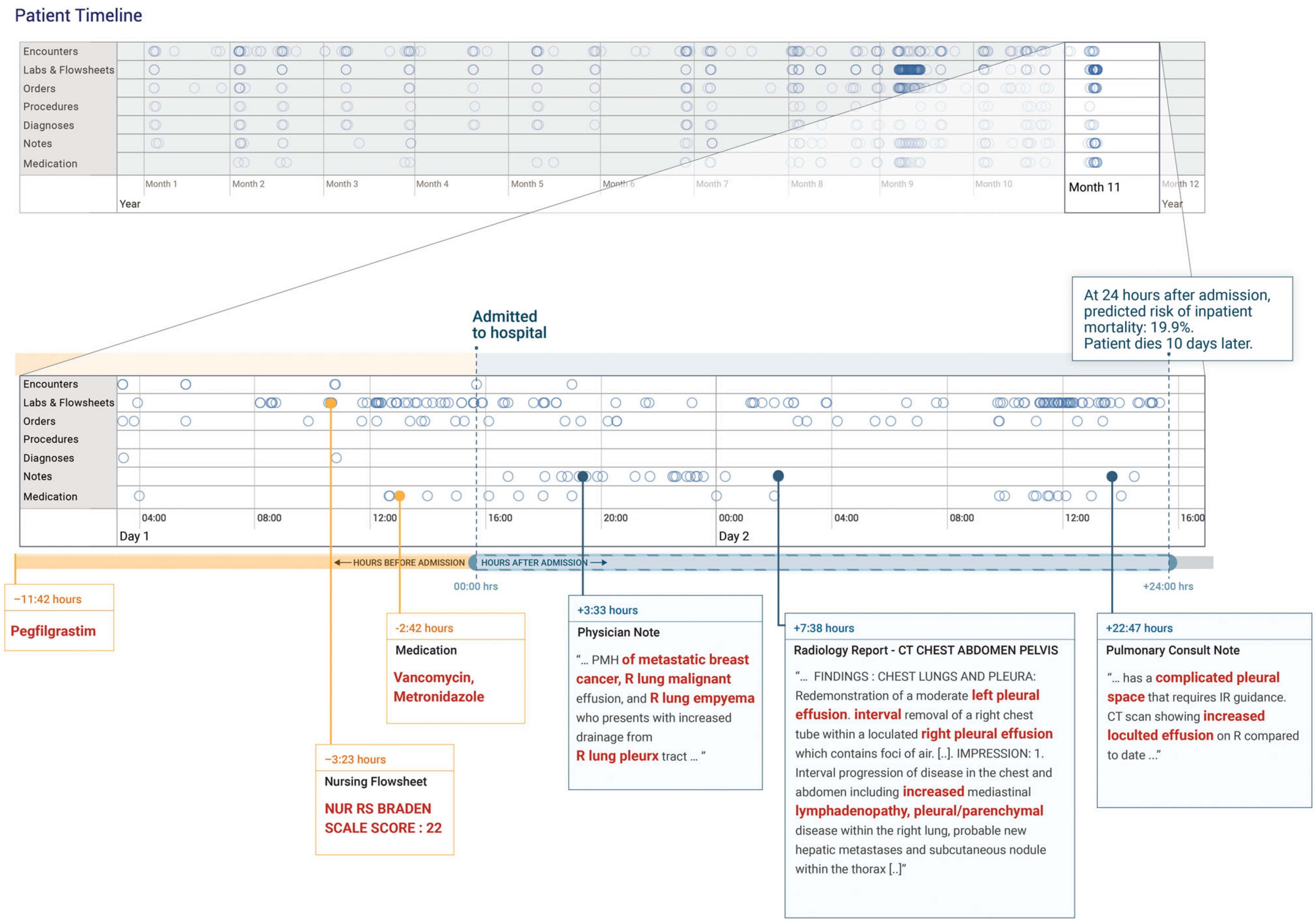
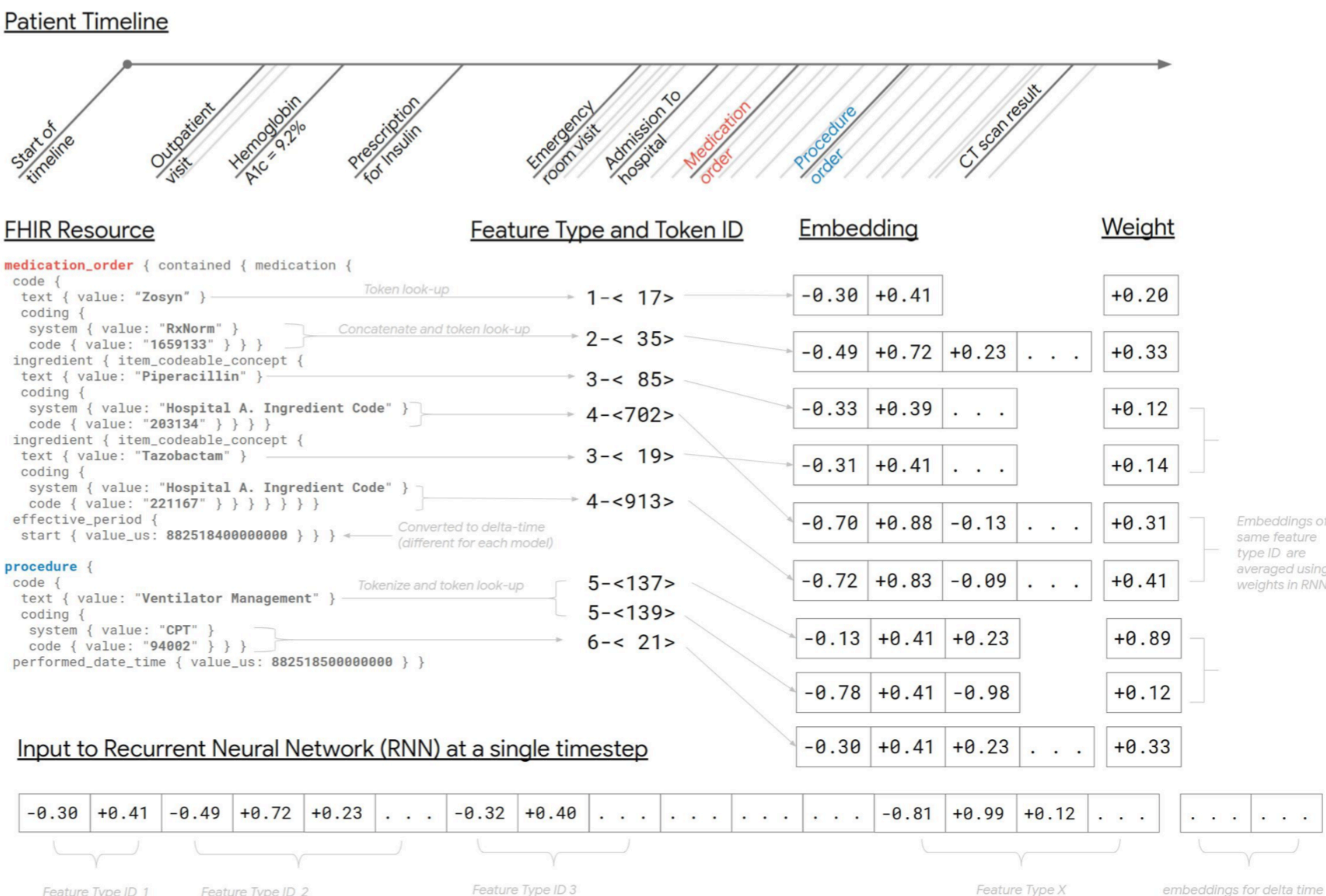
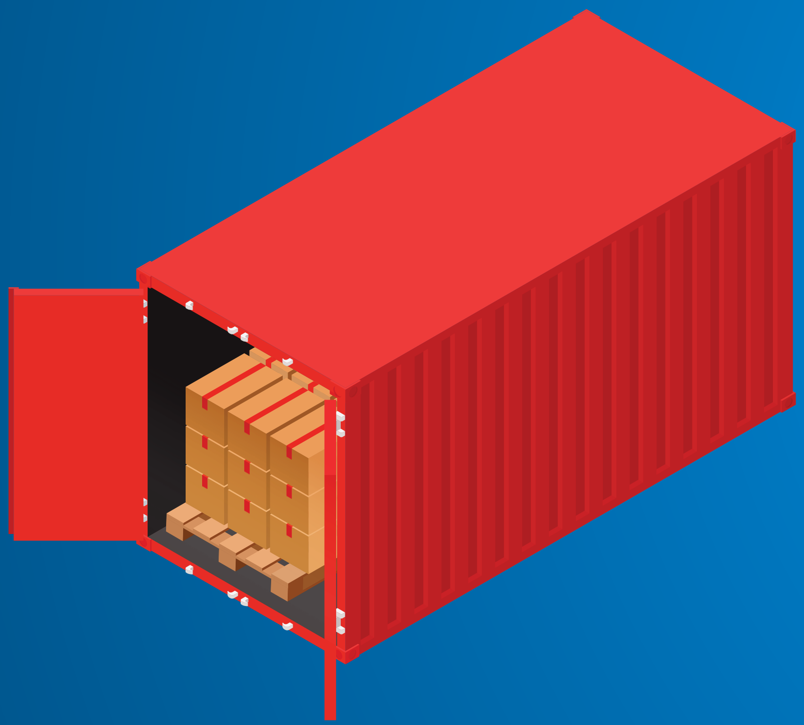


Fig. 3 The patient record shows a woman with metastatic breast cancer with malignant pleural effusions and empyema. The patient timeline at the top of the figure contains circles for every time-step for which at least a single token exists for the patient, and the horizontal lines show the data type. There is a close-up view of the most recent data points immediately preceding a prediction made 24 h after admission. We trained models for each data type and highlighted in red the tokens which the models attended to—the non-highlighted text was not attended to but is shown for context. The models pick up features in the medications, nursing flowsheets, and clinical notes relevant to the prediction

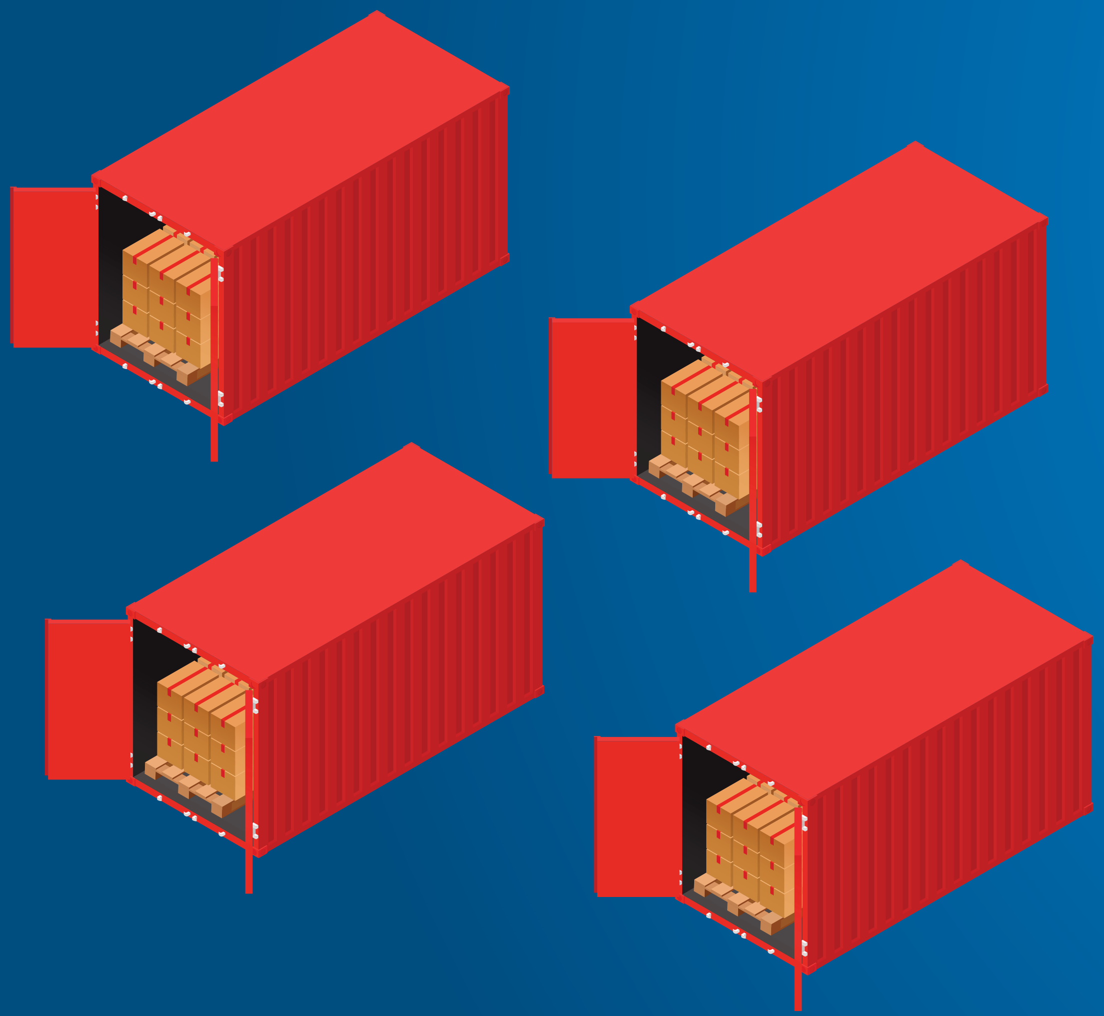

Patient Timeline
Start of timeline
Outpatient visit
Hemoglobin A1c = 9.2%
Prescription for Insulin
Emergency room visit
Admission To hospital
Medication order
Procedure order
CT scan result
FHIR Resource
Feature Type and Token ID
Embedding
Weight
medication_order { contained { medication {
code {
text { value: "Zosyn" }
coding {
system { value: "RxNorm" }
code { value: "1659133" } } }
ingredient { item_codeable_concept {
text { value: "Piperacillin" }
coding {
system { value: "Hospital A. Ingredient Code" }
code { value: "203134" } } } }
ingredient { item_codeable_concept {
text { value: "Tazobactam" }
coding {
system { value: "Hospital A. Ingredient Code" }
code { value: "221167" } } } } } } }
effective_period {
start { value_us: 882518400000000 } } }
procedure {
code {
text { value: "Ventilator Management" }
coding {
system { value: "CPT" }
code { value: "94002" } } }
performed_date_time { value_us: 882518500000000 } }
Token look-up
Concatenate and token look-up
Converted to delta-time (different for each model)
Tokenize and token look-up
1-< 17>
2-< 35>
3-< 85>
4-<702>
3-< 19>
4-<913>
5-<137>
5-<139>
6-< 21>
-0.30 +0.41
-0.49 +0.72 +0.23 . . .
-0.33 +0.39 . . .
-0.31 +0.41 . . .
-0.70 +0.88 -0.13 . . .
-0.72 +0.83 -0.09 . . .
-0.13 +0.41 +0.23
-0.78 +0.41 -0.98
-0.30 +0.41 +0.23 . . .
+0.20
+0.33
+0.12
+0.14
+0.31
+0.41
+0.89
+0.12
+0.33
Embeddings of same feature type ID are averaged using weights in RNN
Input to Recurrent Neural Network (RNN) at a single timestep
-0.30 +0.41 -0.49 +0.72 +0.23 . . . -0.32 +0.40 . . . . -0.81 +0.99 +0.12 . . . .
Feature Type ID 1
Feature Type ID 2
Feature Type ID 3
Feature Type X
embeddings for delta time

Data Liquidity



Consider the shipping container or "intermodal freight container"

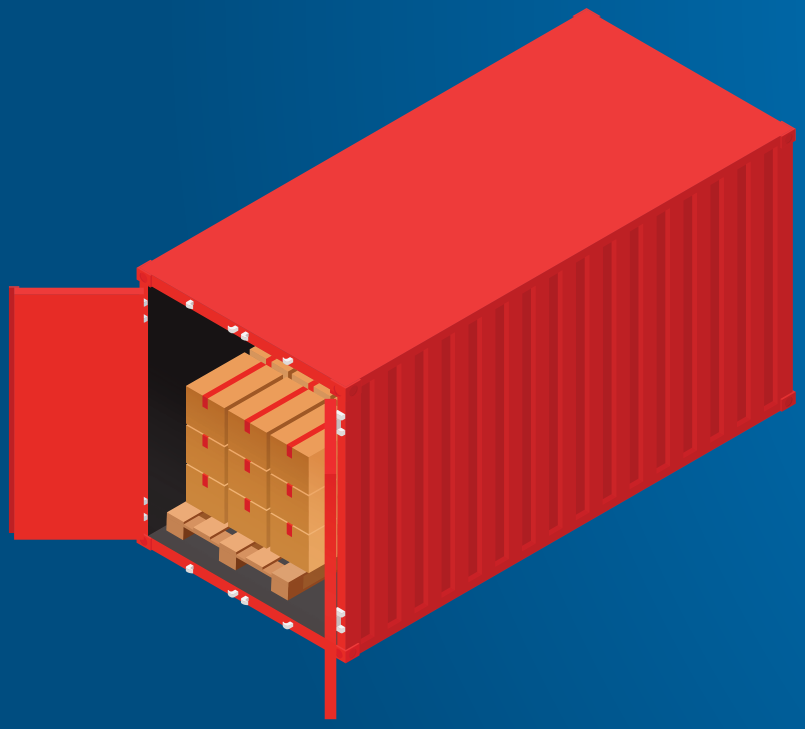
Data Liquidity



Such containers revolutionized shipping because the standardized intermodal container can be easily stacked on a freighter, moved to a train, and transferred to a truck. Every major port in the world has the infrastructure to handle these containers.

And this solves the problem of how to pack bananas next to grand pianos and safely get them to their destinations

Data Liquidity



We need to do the same for health-relevant data…

i.e. have standardized containers that makes any type of data easy to pack, grab, combine and move around.

Your "shipping manifest" may show that you have a breast MRI inside, or a 24-hour urine creatinine, but this is easily manageable because the container is standardized and clearly labeled

Data Liquidity

Examples of health data "containerization" standards include the "Fast Healthcare Interoperability Resource" (FHIR) standard

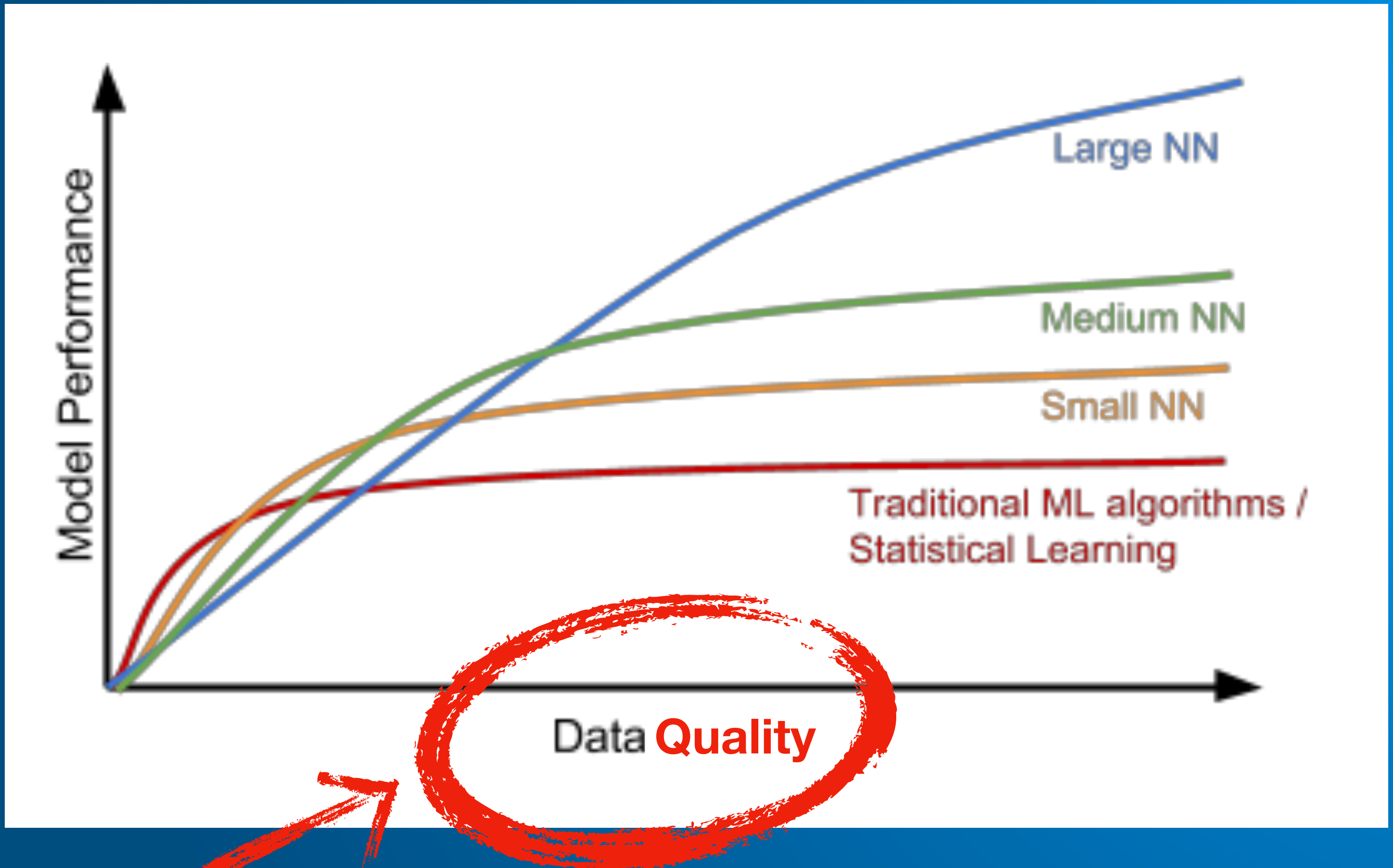
Whether it's this standard, alone, or in conjunction with other "webservice-friendly" standards,

Our aim should be to build the "data liquidity ecosystem" equivalent to freighters, cranes, trains and trucks that facilitate the logistics of health data transport

Data Liquidity

And the data liquidity infrastructure to deliver these data to an artificial intelligence algorithm, app, or a user facilitates rapid and safe shipment of these data containers to their destinations

Data Liquidity



How do we move along the X-axis?

Adapted from Andrew Ng Talk: https://youtu.be/F1ka6a13S9I

Betty Jones

Duke | FORGE
HEALTH DATA IN ACTION

Caution

Algorithms don’t have ethics

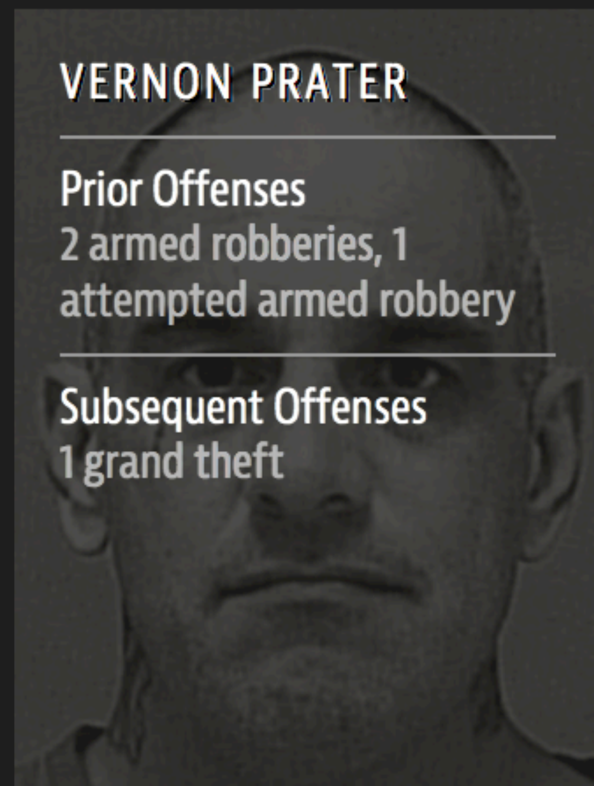
Bernard Parker, left, was rated high risk; Dylan Fugett was rated low risk. (Josh Ritchie for ProPublica)

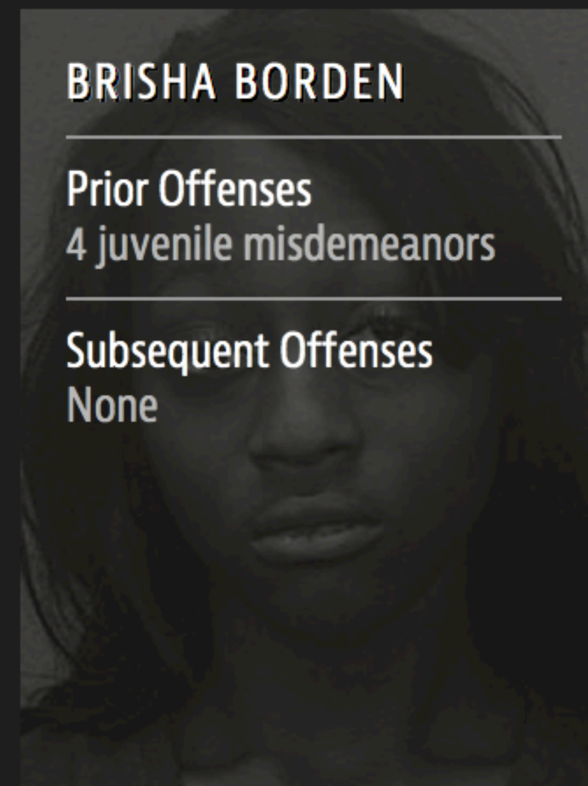
Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

PROPUBLICA

Two Petty Theft Arrests



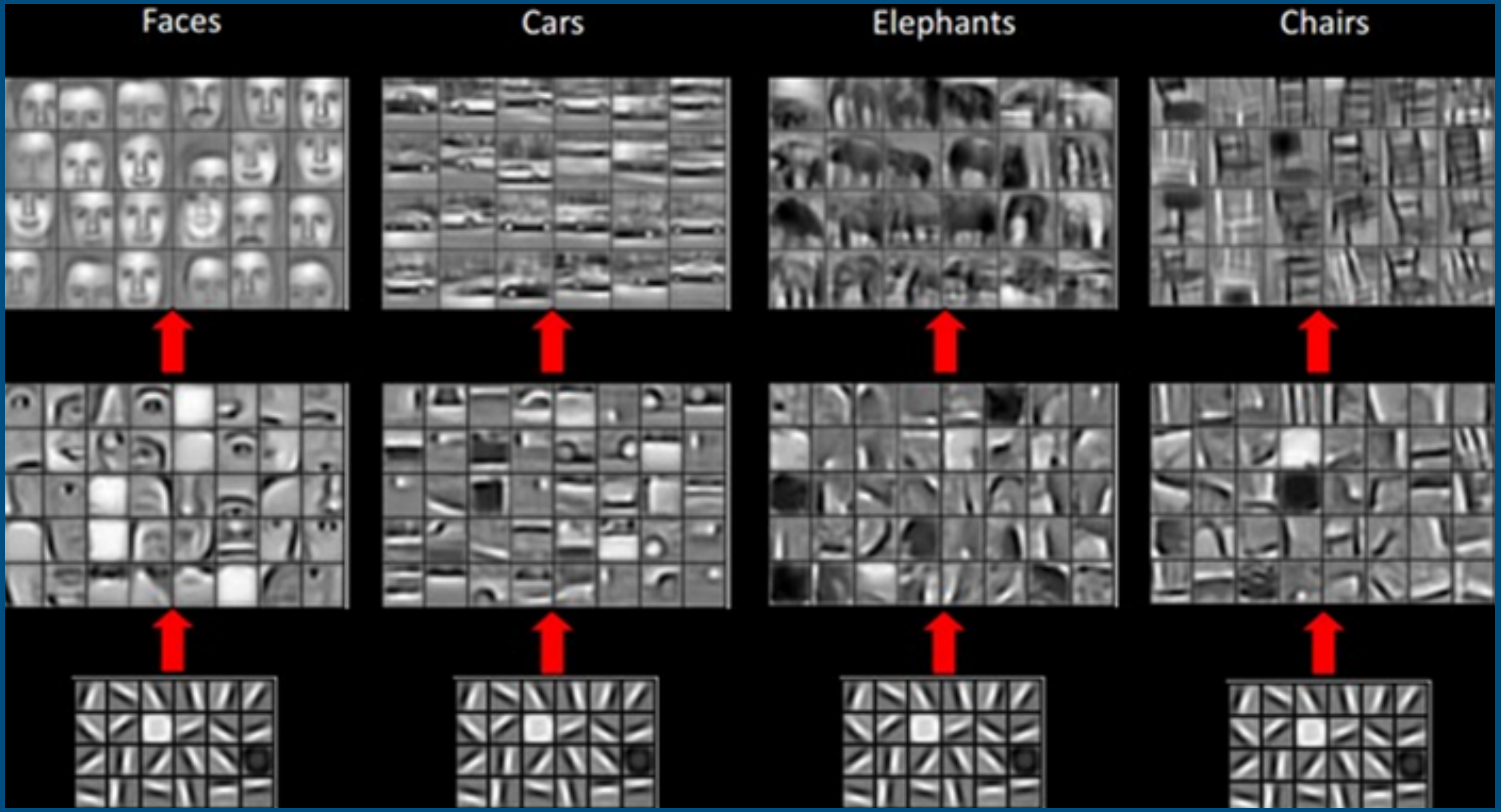


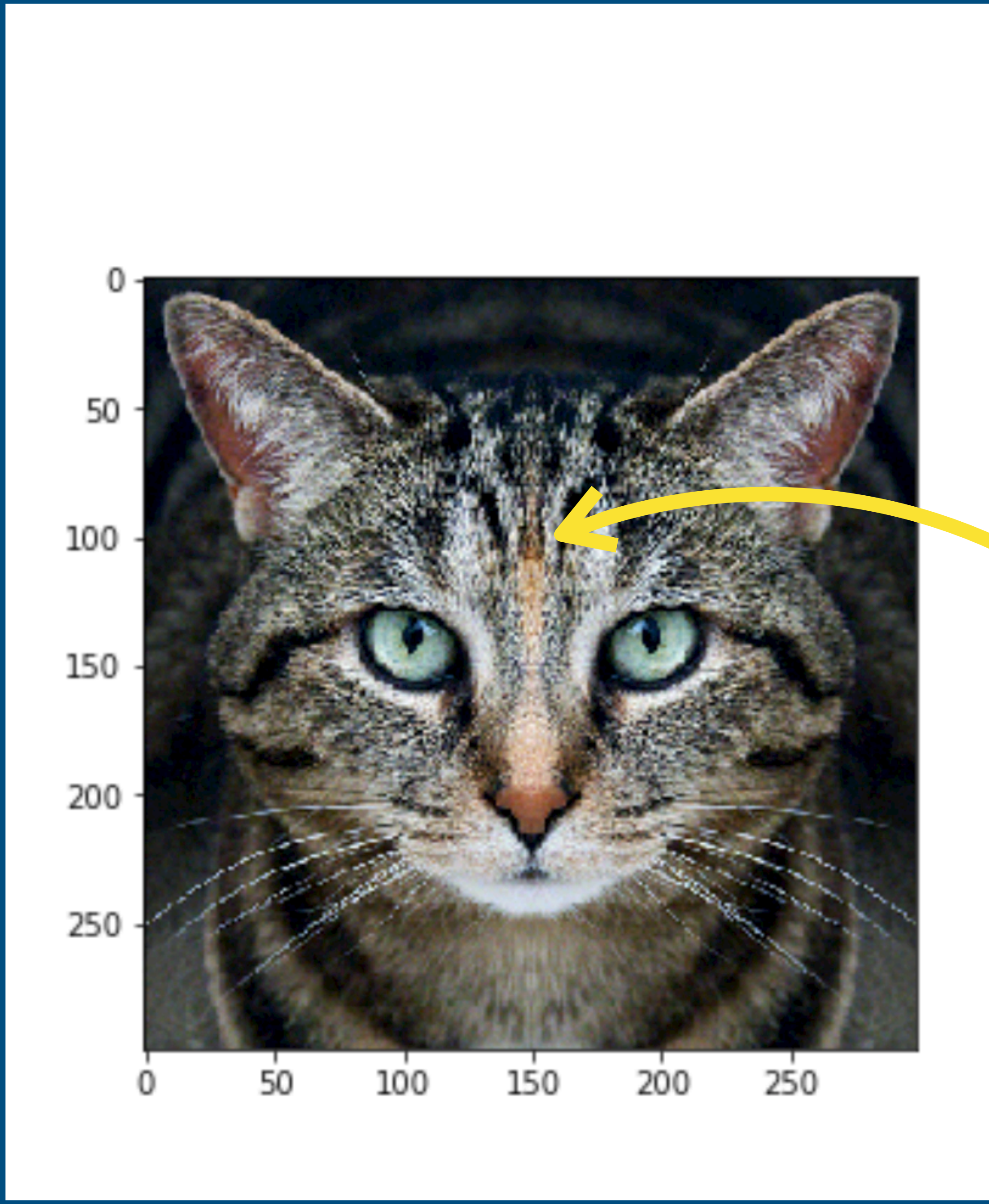
VERNON PRATER	BRISHA BORDEN
RISK: **3**	RISK: **8**

Borden was rated high risk for future crime after she and a friend took a kid's bike and scooter that were sitting outside. She did not reoffend.

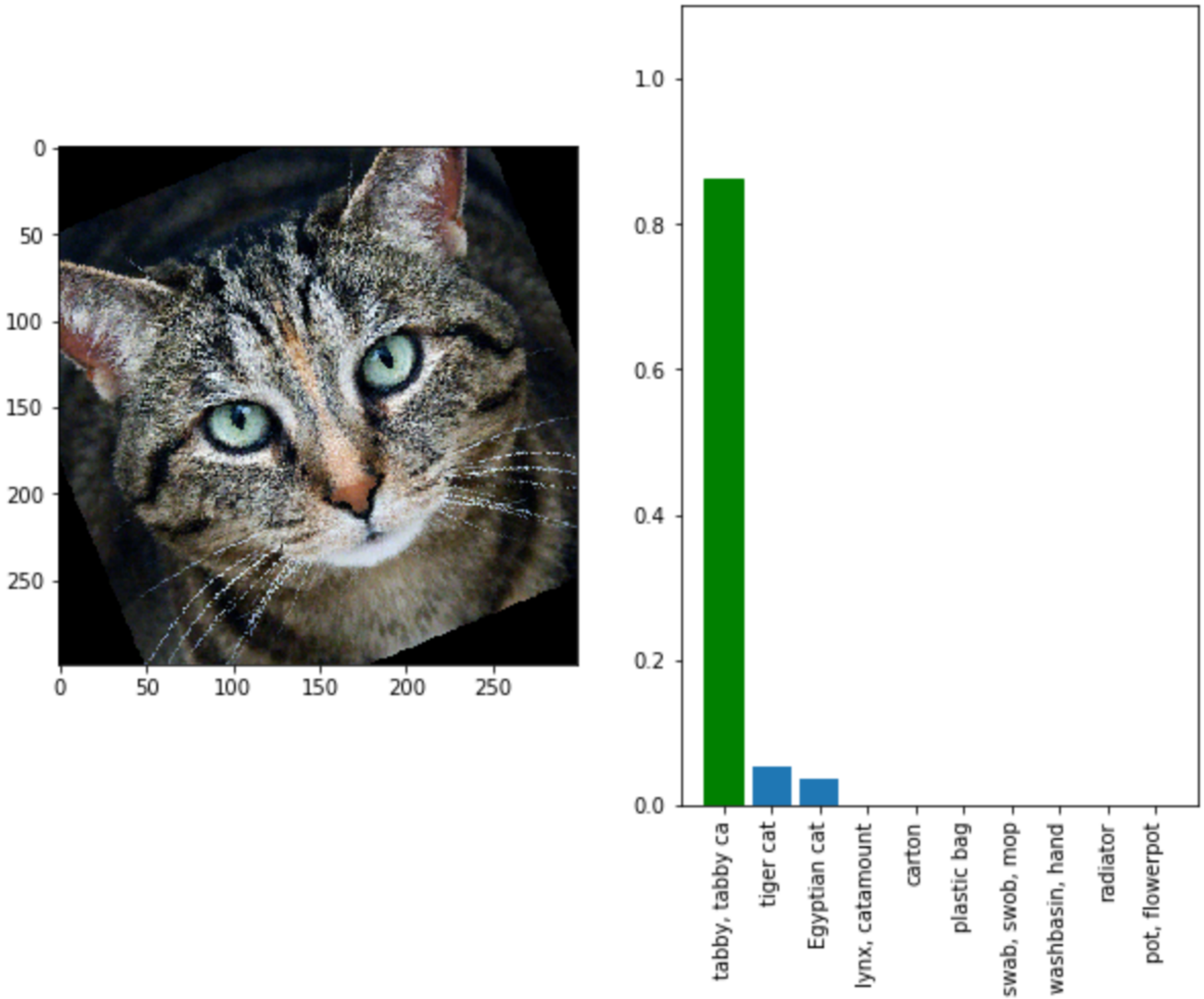
Algorithms don’t think like we do

What a deep convolutional neural net learns





Inject random pixel noise



0
50
100
150
200
250
0
50
100
150
200
250
1.0
0.8
0.6
0.4
0.2
0.0
tabby, tabby ca
tiger cat
Egyptian cat
lynx, catamount
carton
plastic bag
swab, swob, mop
washbasin, hand
radiator
pot, flowerpot

https://www.labsix.org/physical-objects-that-fool-neural-nets/

rnn.py — ~/Desktop

Project

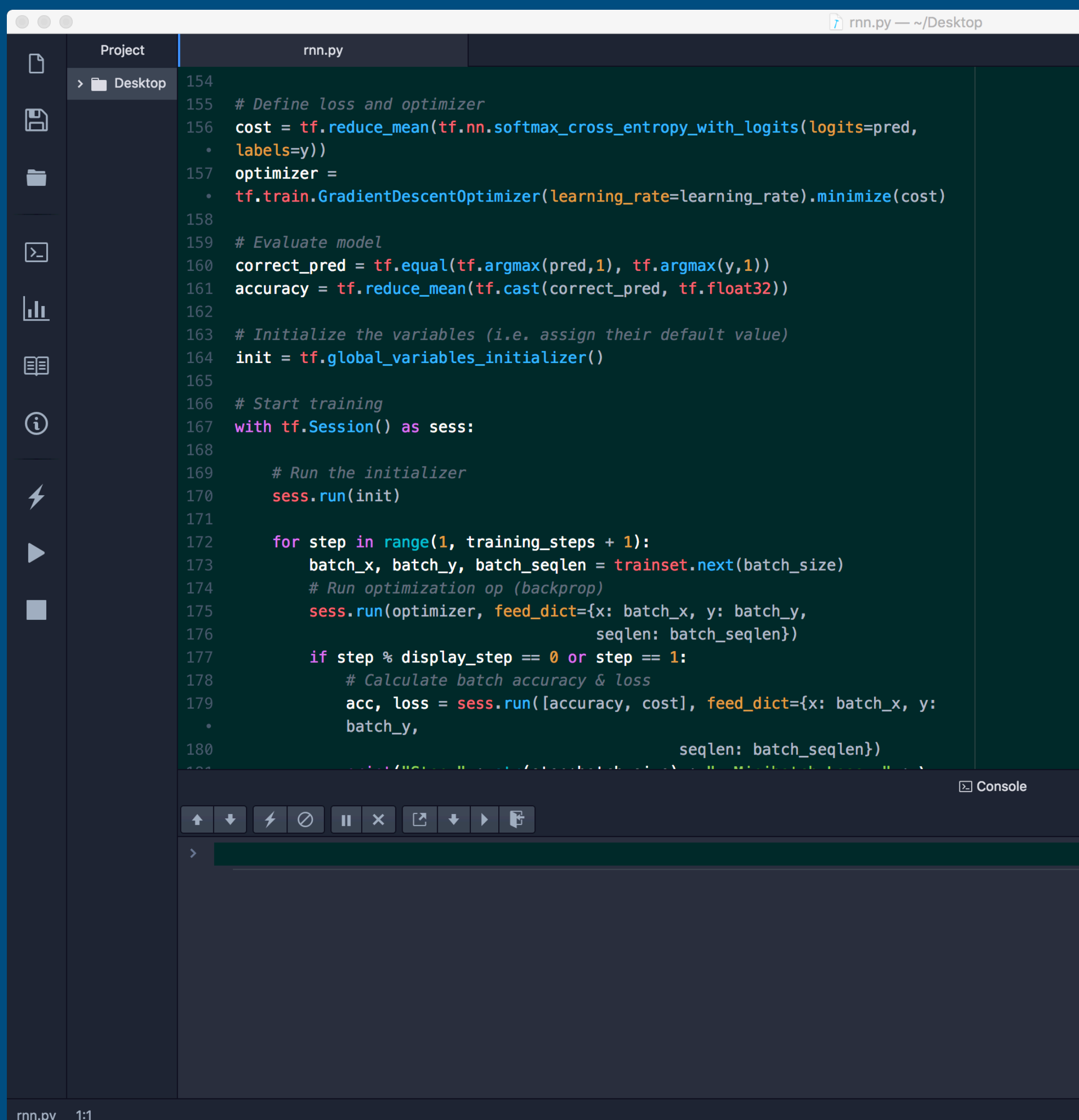
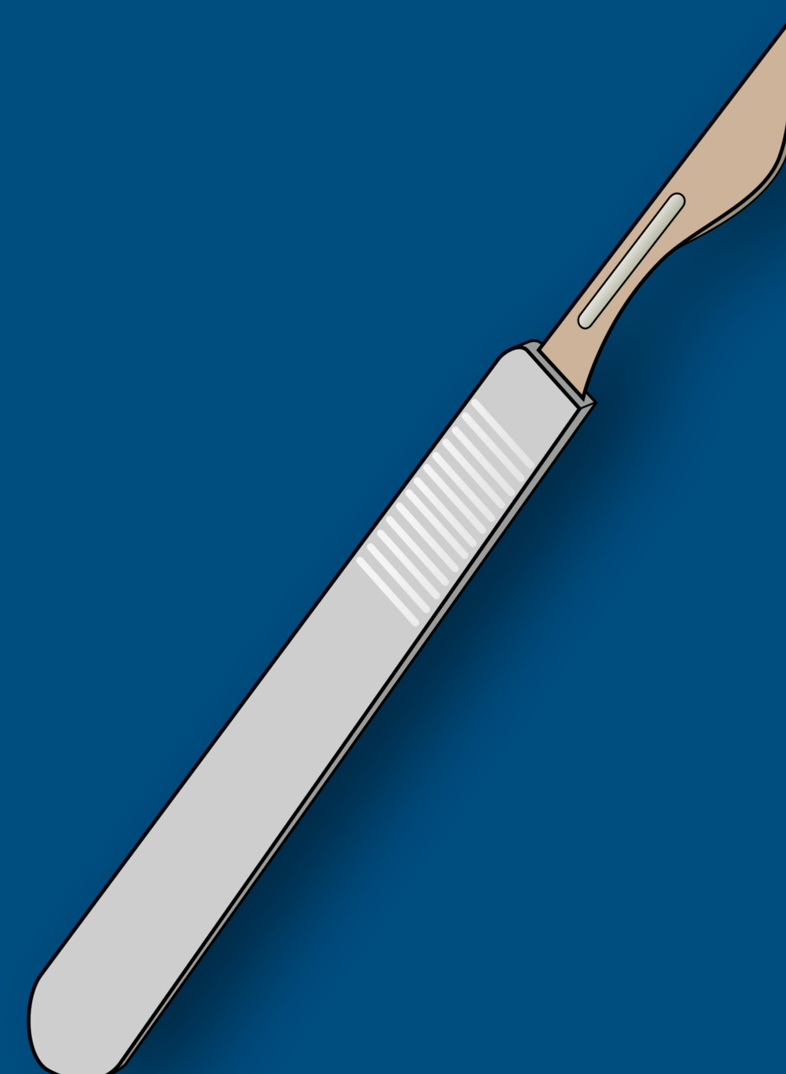
Desktop

rnn.py

```python

# Define loss and optimizer
cost = tf.reduce_mean(tf.nn.softmax_cross_entropy_with_logits(logits=pred, labels=y))
optimizer = tf.train.GradientDescentOptimizer(learning_rate=learning_rate).minimize(cost)

# Evaluate model
correct_pred = tf.equal(tf.argmax(pred,1), tf.argmax(y,1))
accuracy = tf.reduce_mean(tf.cast(correct_pred, tf.float32))

# Initialize the variables (i.e. assign their default value)
init = tf.global_variables_initializer()

# Start training
with tf.Session() as sess:

    # Run the initializer
    sess.run(init)

    for step in range(1, training_steps + 1):
        batch_x, batch_y, batch_seqlen = trainset.next(batch_size)
        # Run optimization op (backprop)
        sess.run(optimizer, feed_dict={x: batch_x, y: batch_y,
                                       seqlen: batch_seqlen})
        if step % display_step == 0 or step == 1:
            # Calculate batch accuracy & loss
            acc, loss = sess.run([accuracy, cost], feed_dict={x: batch_x, y: batch_y,
                                                seqlen: batch_seqlen})
```

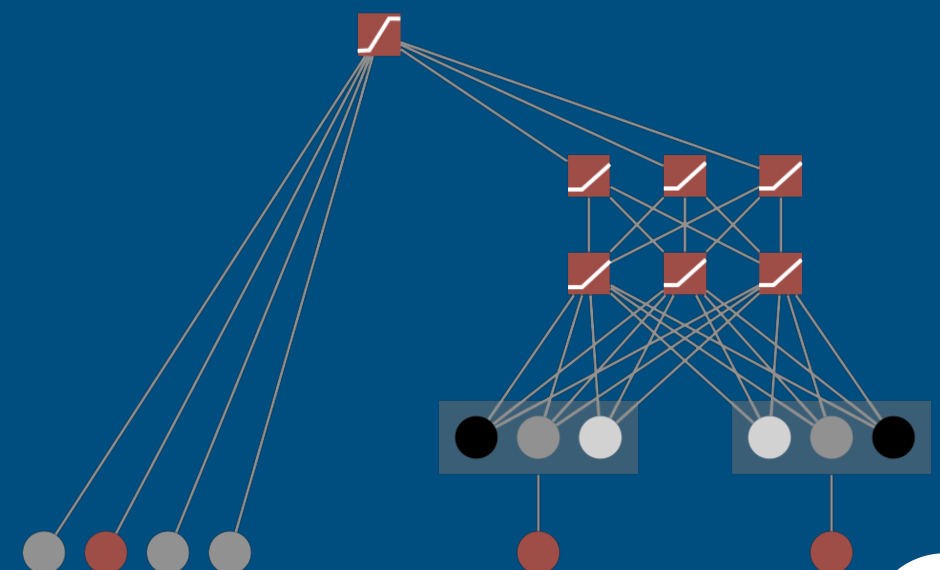
Console

https://github.com/aymericdamien/TensorFlow-Examples/blob/master/examples/3_NeuralNetworks/dynamic_rnn.py


rnn.py — ~/Desktop
Project
rnn.py
Desktop
# Define loss and optimizer
cost = tf.reduce_mean(tf.nn.softmax_cross_entropy_with_logits(logits=pred,
labels=y))
optimizer =
tf.train.GradientDescentOptimizer(learning_rate=learning_rate).minimize(cost)
# Evaluate model
correct_pred = tf.equal(tf.argmax(pred,1), tf.argmax(y,1))
accuracy = tf.reduce_mean(tf.cast(correct_pred, tf.float32))
# Initialize the variables (i.e. assign their default value)
init = tf.global_variables_initializer()
# Start training
with tf.Session() as sess:
# Run the initializer
sess.run(init)
for step in range(1, training_steps + 1):
batch_x, batch_y, batch_seqlen = trainset.next(batch_size)
Run optimization op (backprop)
sess.run(optimizer, feed_dict={x: batch_x, y: batch_y,
seqlen: batch_seqlen})
if step % display_step == 0 or step == 1:
# Calculate batch accuracy & loss
acc, loss = sess.run([accuracy, cost], feed_dict={x: batch_x, y:
batch_y,
seqlen: batch_seqlen})
Console
rnn.py
1:1


Security

Samples



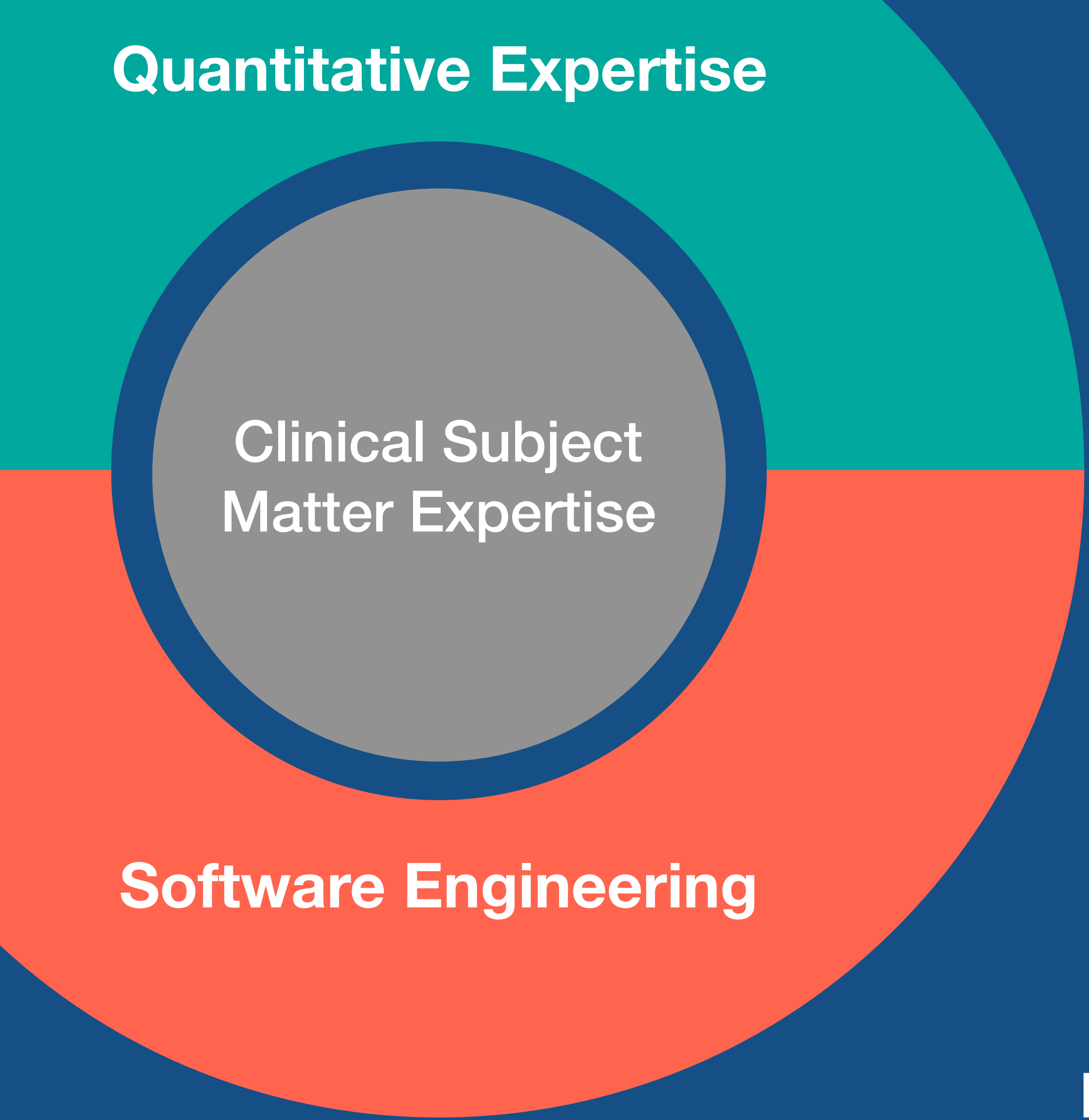
Algorithm

What is Data Science?

Knowledge among computer scientists about how to think of and approach the analysis of data is limited, just as the knowledge of computing environments by statisticians is limited. A merger of the knowledge bases would produce a powerful force for innovation.

–Bill Cleveland on *Data Science* (2001)


Quantitative Expertise
Clinical Subject
Matter Expertise
Software Engineering



Machine Learning

Reinforcement Learning

Deep Neural Networks

Class Discovery Methods

Clinical Informatics

Biostatistics

Experimental Design

R

Python

Clinical Trial Design

Cloud

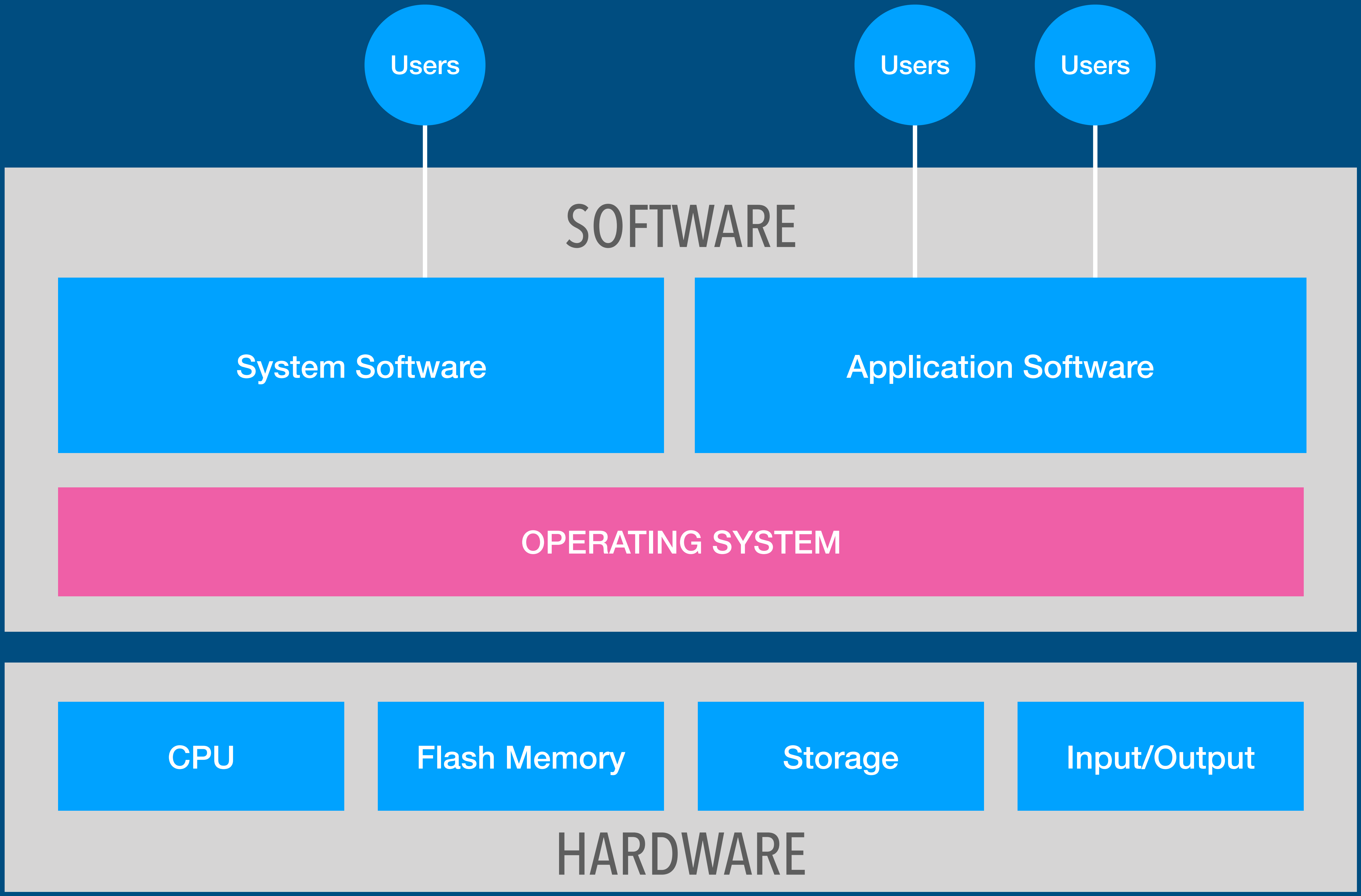
Application Programming Interfaces

Relational Databases

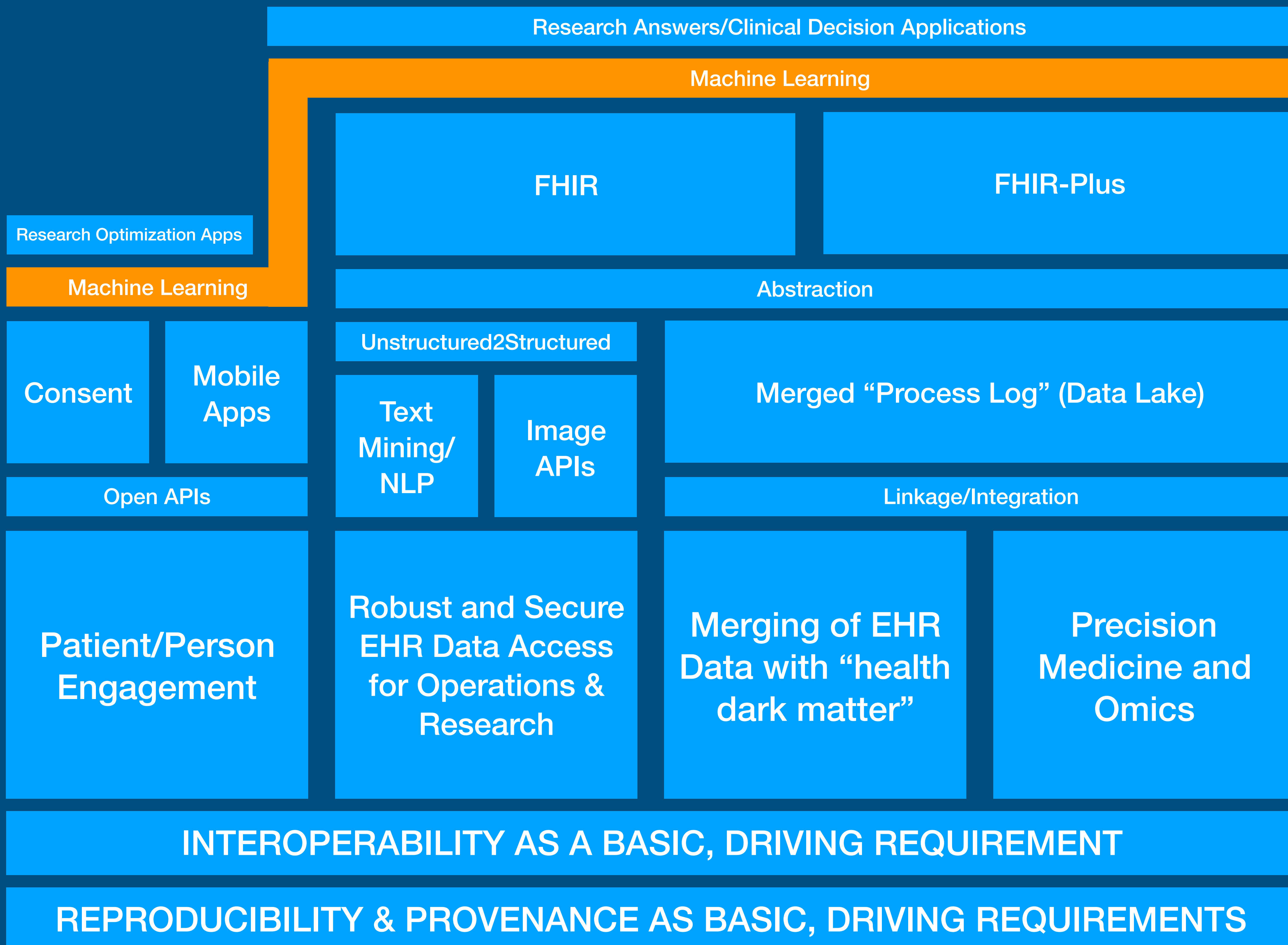
NoSQL Datastores

Containerization

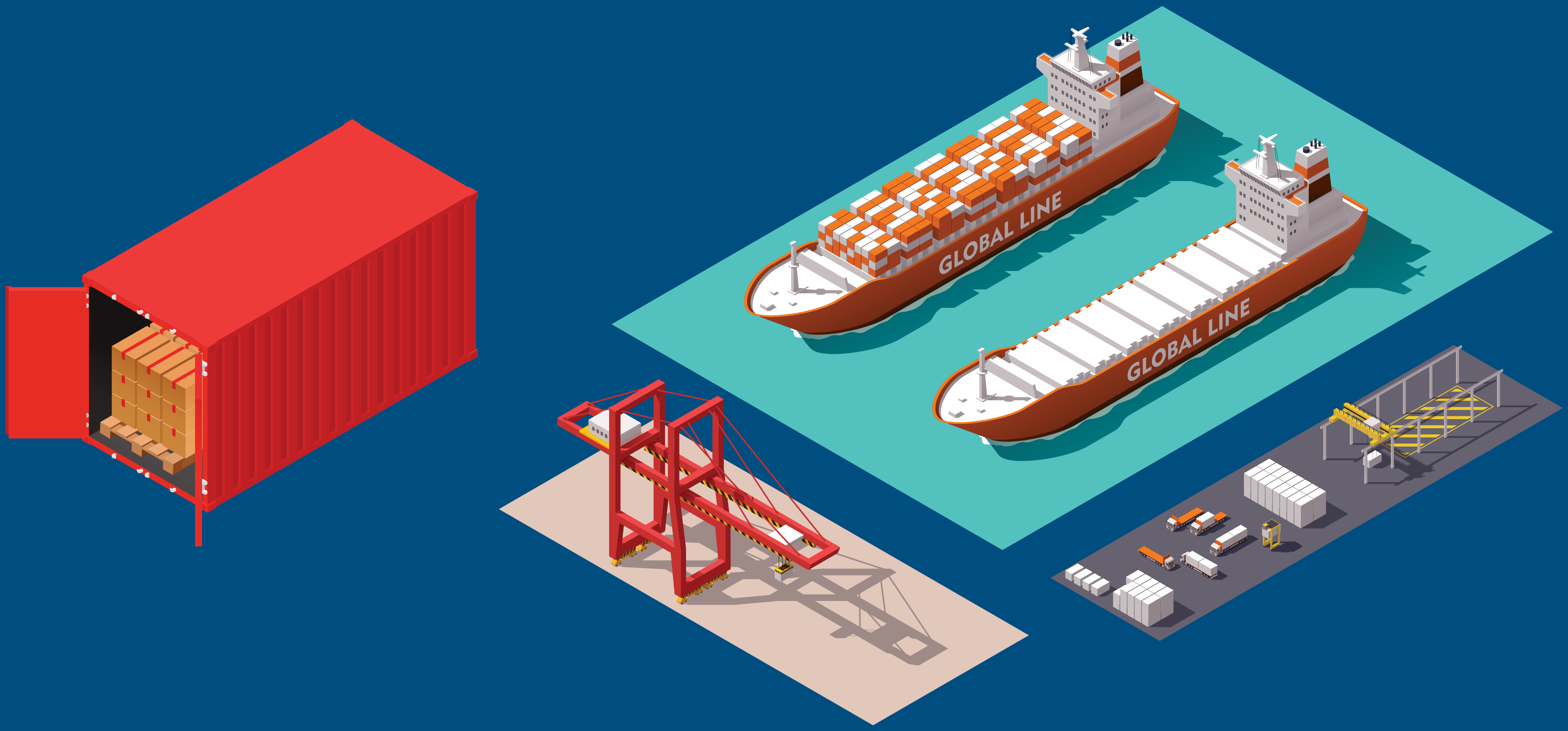
We need an "operating system" for healthcare delivery



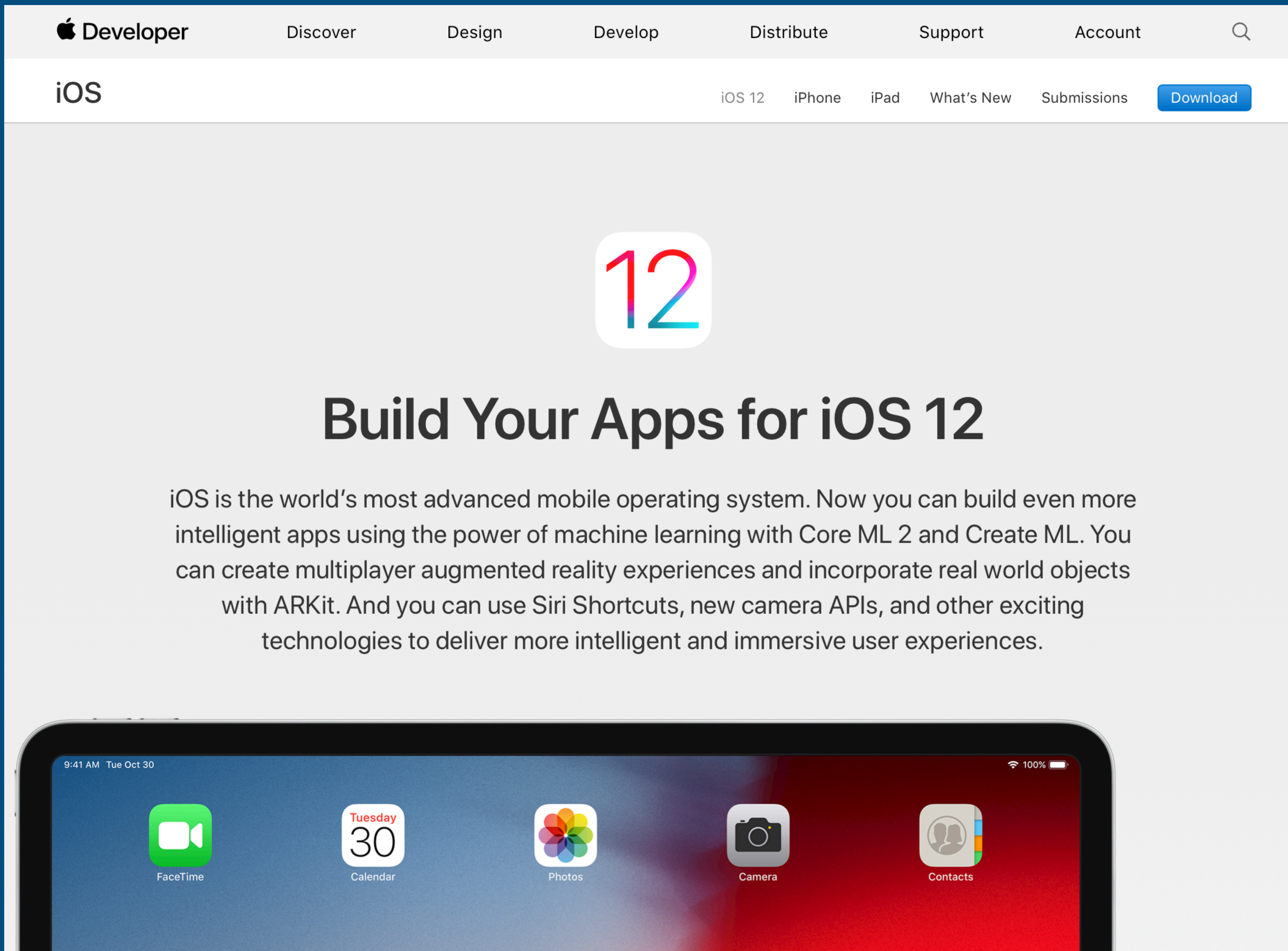
We need an "operating system" for healthcare delivery



We need Data Liquidity



Rapid Prototyping & Development



Rapid Prototyping & Development

Rapid Feedback Loops

Users

patients

clinicians

administrators

Data Science Teams

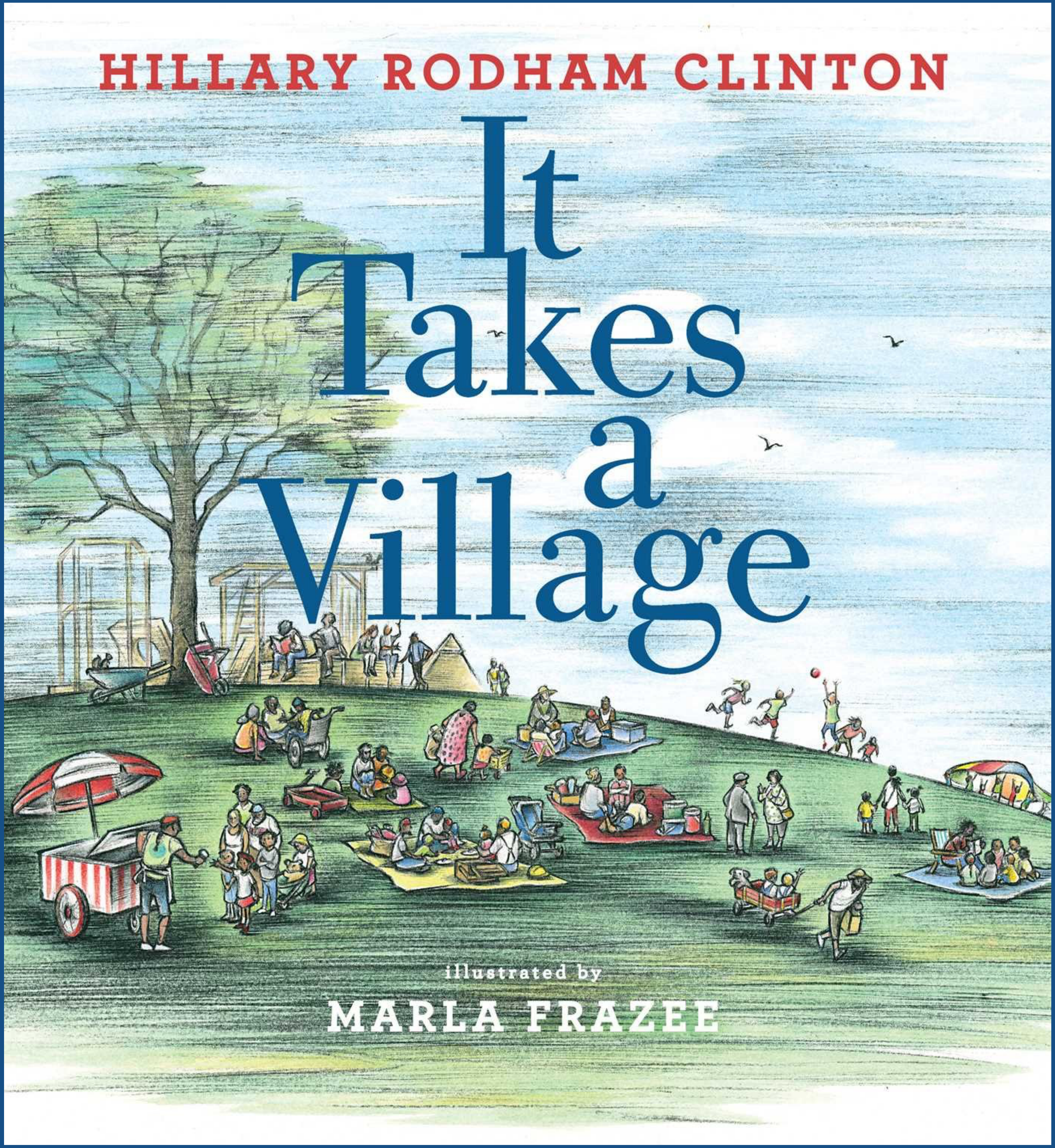
Application-Level

"Packaged" Atomic Data

What should the EHR of the future look like?

Source Data

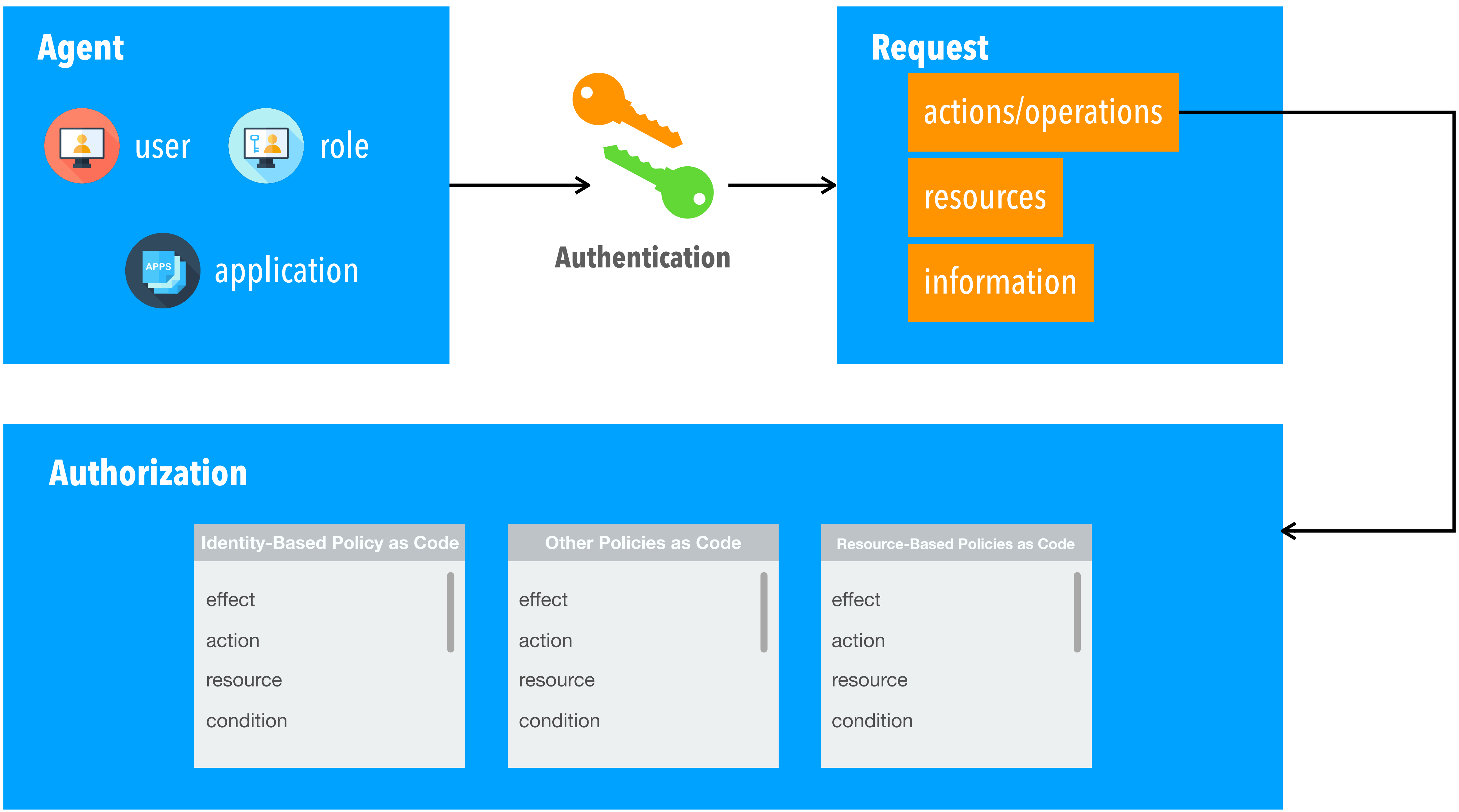
Betty Jones

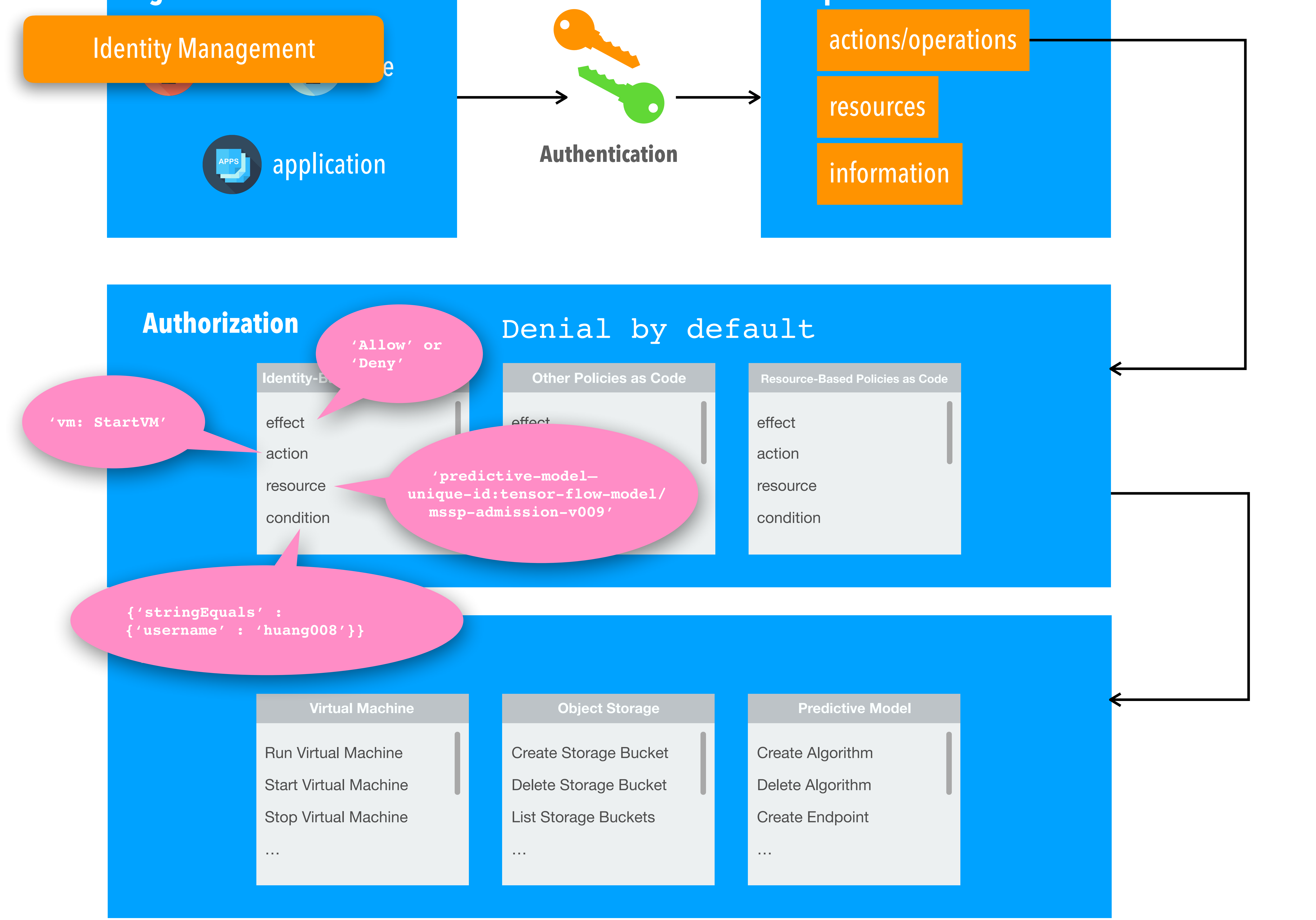


Duke | FORGE
HEALTH DATA IN ACTION

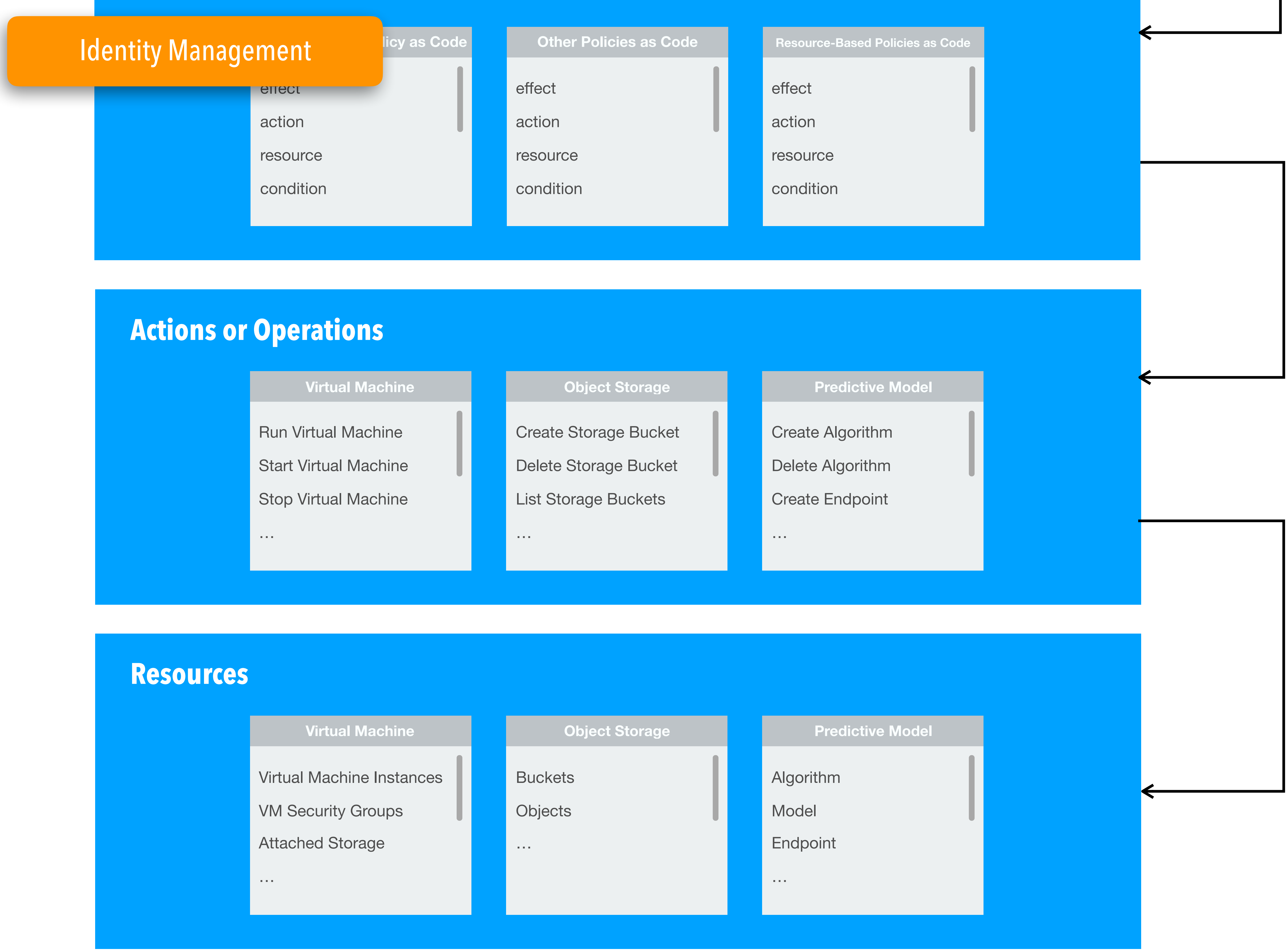
Thank You

Identity Management





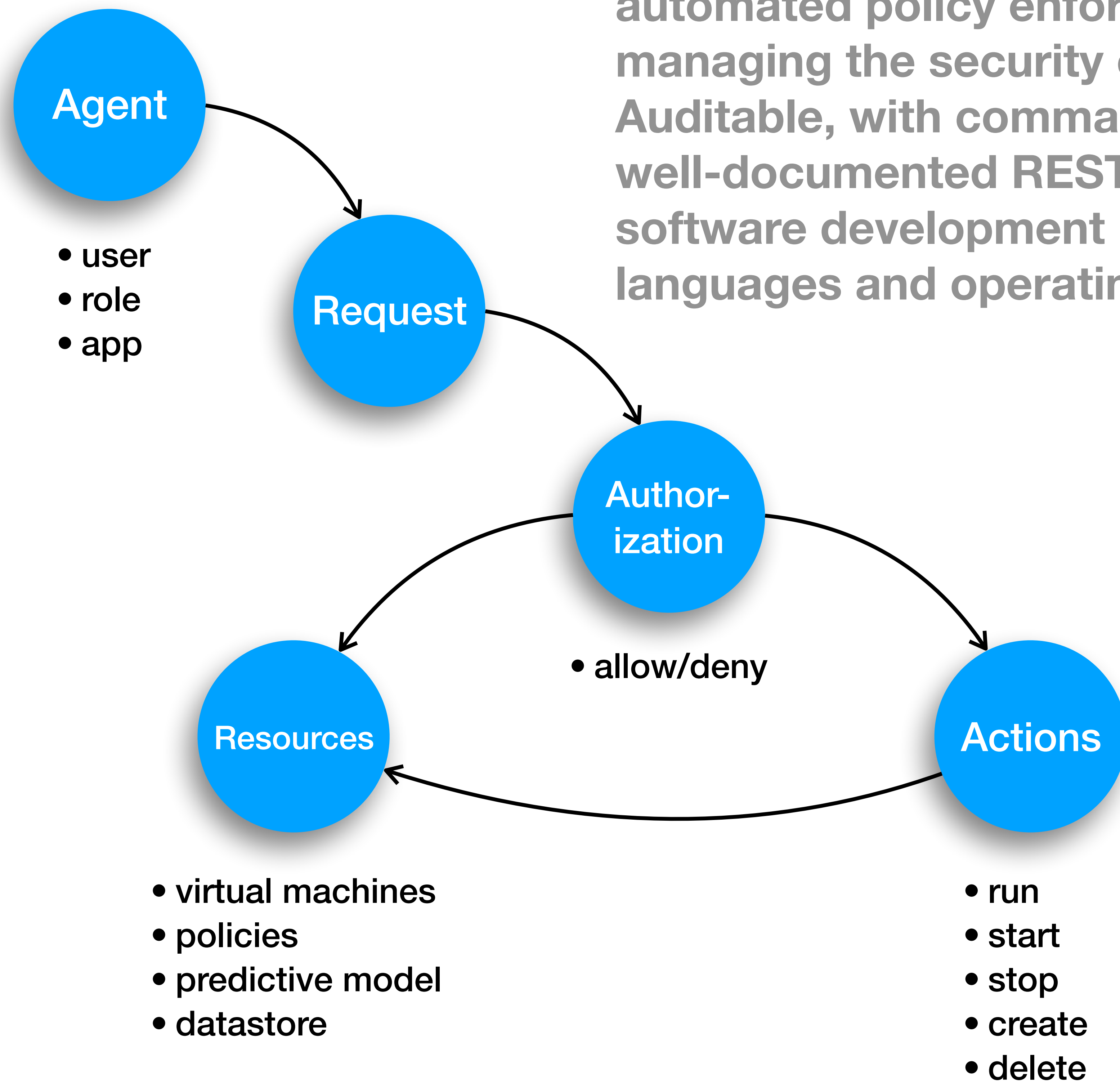
Identity Management
application
Authentication
actions/operations
resources
information
Authorization
Denial by default
'Allow' or 'Deny'
'vm: StartVM'
Identity-B
effect
action
resource
condition
'predictive-model-unique-id:tensor-flow-model/mssp-admission-v009'
{'stringEquals' : {'username' : 'huang008'}}
Other Policies as Code
effect
Resource-Based Policies as Code
effect
action
resource
condition
Virtual Machine
Run Virtual Machine
Start Virtual Machine
Stop Virtual Machine
...
Object Storage
Create Storage Bucket
Delete Storage Bucket
List Storage Buckets
...
Predictive Model
Create Algorithm
Delete Algorithm
Create Endpoint
...

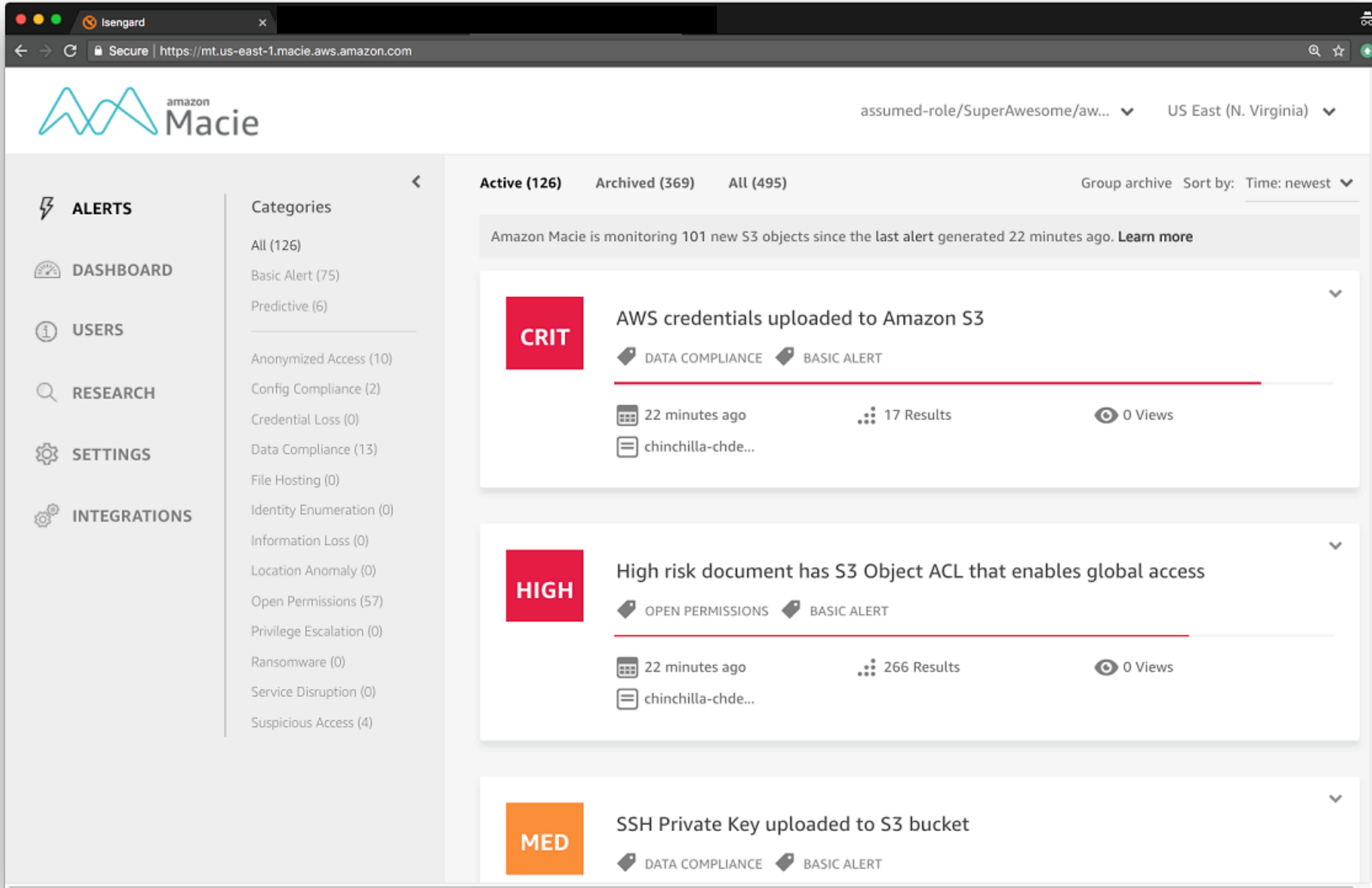


Identity Management
icy as Code
effect
action
resource
condition
Other Policies as Code
effect
action
resource
condition
Resource-Based Policies as Code
effect
action
resource
condition
Actions or Operations
Virtual Machine
Run Virtual Machine
Start Virtual Machine
Stop Virtual Machine
…
Object Storage
Create Storage Bucket
Delete Storage Bucket
List Storage Buckets
…
Predictive Model
Create Algorithm
Delete Algorithm
Create Endpoint
…
Resources
Virtual Machine
Virtual Machine Instances
VM Security Groups
Attached Storage
…
Object Storage
Buckets
Objects
…
Predictive Model
Algorithm
Model
Endpoint
…

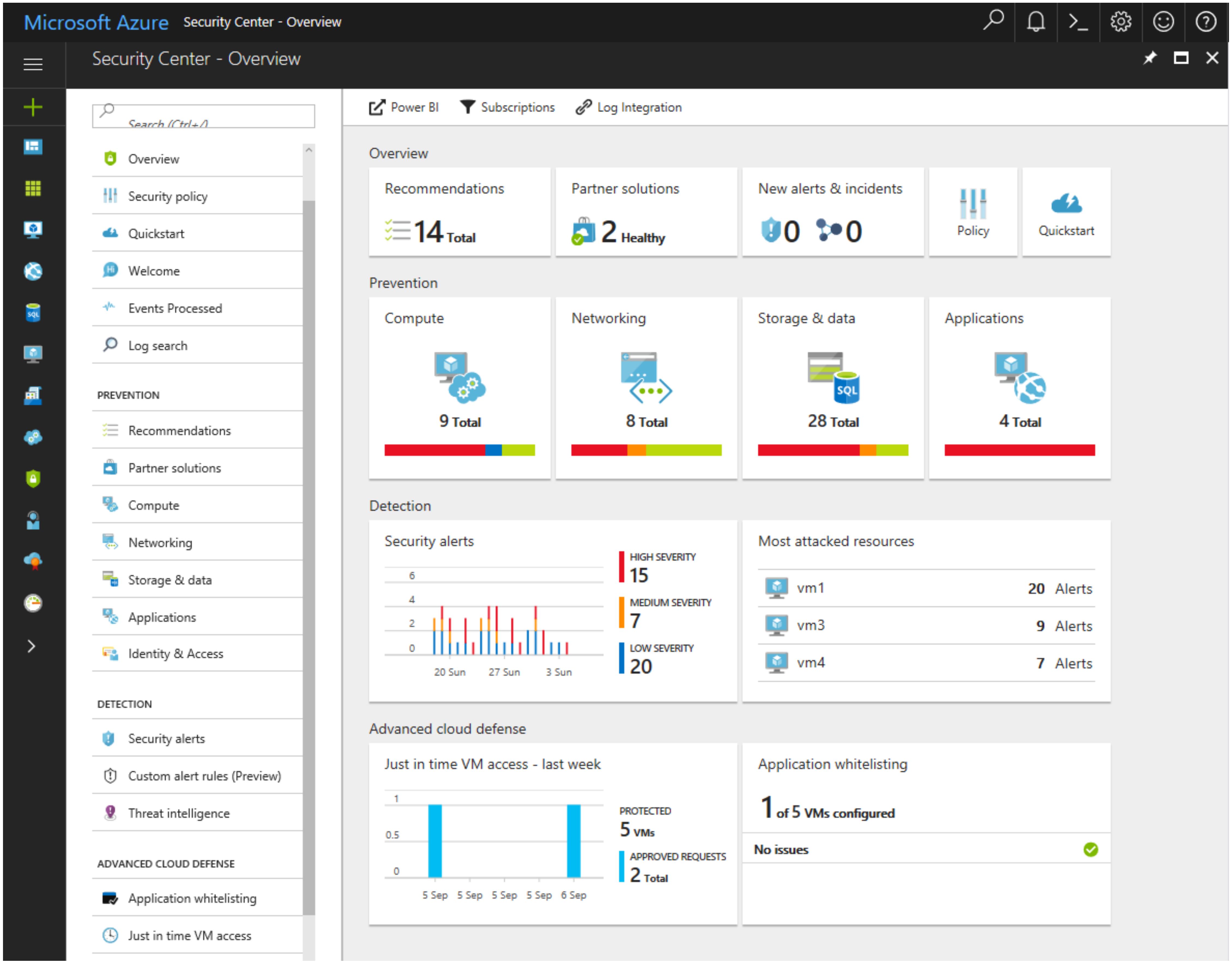
Identity Management

A rich, robust and flexible Identity Management data model that enables automated policy enforcement in managing the security of resources. Auditable, with command line tooling, a well-documented RESTful API, and software development kits for standard languages and operating systems




Isengard
Secure | https://mt.us-east-1.macie.aws.amazon.com
amazon Macie
assumed-role/SuperAwesome/aw...
US East (N. Virginia)
ALERTS
DASHBOARD
USERS
RESEARCH
SETTINGS
INTEGRATIONS
Categories
All (126)
Basic Alert (75)
Predictive (6)
Anonymized Access (10)
Config Compliance (2)
Credential Loss (0)
Data Compliance (13)
File Hosting (0)
Identity Enumeration (0)
Information Loss (0)
Location Anomaly (0)
Open Permissions (57)
Privilege Escalation (0)
Ransomware (0)
Service Disruption (0)
Suspicious Access (4)
Active (126)
Archived (369)
All (495)
Group archive
Sort by:
Time: newest
Amazon Macie is monitoring 101 new S3 objects since the last alert generated 22 minutes ago. Learn more
CRIT
AWS credentials uploaded to Amazon S3
DATA COMPLIANCE
BASIC ALERT
22 minutes ago
17 Results
0 Views
chinchilla-chde...
HIGH
High risk document has S3 Object ACL that enables global access
OPEN PERMISSIONS
BASIC ALERT
22 minutes ago
266 Results
0 Views
chinchilla-chde...
MED
SSH Private Key uploaded to S3 bucket
DATA COMPLIANCE
BASIC ALERT


Minimum Risk: 6
Adjust the slider below to view only documents above a certain risk level.
Total Matching Events
unique risky Events over past 60 days
11
AWS CloudTrail events - minRisk: (6)
daily
weekly
Event Counts
35
30
25
20
15
10
5.0
0.0
May 28
Jun 04
Jun 11
Jun 18
Jun 25
Jul 02
Jul 09
Jul 16
Jul 23
UpdateAssumeRolePolicy (2)
UpdateAccessKey (2)
CreateUser (2)
PutRolePolicy (1)
CreateKey (1)
DeleteTrail (1)
PutUserPolicy (8)
CreateAccessKey (6)
AttachUserPolicy (3)
CreateRole (2)
ConsoleLogin (587)



Microsoft Azure
Security Center - Overview
Security Center - Overview
Search (Ctrl+/)
Overview
Security policy
Quickstart
Welcome
Events Processed
Log search
PREVENTION
Recommendations
Partner solutions
Compute
Networking
Storage & data
Applications
Identity & Access
DETECTION
Security alerts
Custom alert rules (Preview)
Threat intelligence
ADVANCED CLOUD DEFENSE
Application whitelisting
Just in time VM access
Power BI
Subscriptions
Log Integration
Overview
Recommendations
14 Total
Partner solutions
2 Healthy
New alerts & incidents
0
0
Policy
Quickstart
Prevention
Compute
9 Total
Networking
8 Total
Storage & data
SQL
28 Total
Applications
4 Total
Detection
Security alerts
6
4
2
0
20 Sun
27 Sun
3 Sun
HIGH SEVERITY
15
MEDIUM SEVERITY
7
LOW SEVERITY
20
Most attacked resources
vm1 20 Alerts
vm3 9 Alerts
vm4 7 Alerts
Advanced cloud defense
Just in time VM access - last week
1
0.5
0
5 Sep
5 Sep
5 Sep
5 Sep
6 Sep
PROTECTED
5 VMs
APPROVED REQUESTS
2 Total
Application whitelisting
1 of 5 VMs configured
No issues

Google Cloud Audit - Network and Security
LIVE MODE
[From Panel]
Location of Authorization Failures
Map
Satellite
Show In Search
91574
59124
10033
Google
Map data ©2018
Terms of Use
Network and Security Operations Over Time
Last 24 Hours
400
200
0
04:00 PM
08:00 PM
12:00 AM
19 Feb 18
04:00 AM
08:00 AM
v1.compute.firewalls.delete
v1.compute.routes.insert
beta.compute.networks.insert
v1.compute.firewalls.insert
v1.compute.networks.delete
v1.compute.subnetworks.delete
beta.compute.subnetworks.delete
v1.compute.subnetworks.insert
v1.compute.routes.delete
Authorization Failures Over Time
10K
5,000
0
06:00 PM
12:00 AM
19 Feb 18
06:00 AM
Recent Firewall Changes
Last 3 Hours
timestamp
user
method
ranges
2018-02-19T21:18:31.448UTC
player5@bmlabs.com
v1.compute.firewalls.insert
"0.0.0.0/0"
2018-02-19T21:15:11.451UTC
player6@bmlabs.com
v1.compute.firewalls.insert
"0.0.0.0/0"
2018-02-19T21:15:11.450UTC
player6@bmlabs.com
v1.compute.firewalls.insert
"0.0.0.0/0"
2018-02-19T21:15:11.449UTC
player3@bmlabs.com
v1.compute.firewalls.insert
"0.0.0.0/0"
Network and Security Operations
v1.compute.firewalls.insert
v1.compute.firewalls.delete
v1.compute.subnetworks.delete
v1.compute.routes.delete
v1.compute.networks.delete
1/3
Recent Authorization Failures
timestamp
user
action
2018-02-19T21:18:31.643UTC
player5@bmlabs.com
compute.instances.cr
2018-02-19T21:18:31.643UTC
player7@bmlabs.com
compute.networks.us
Firewall Rules with All Allowed
timestamp
user
method
ranges
2018-02-19T21:15:11.448UTC
player4@bmlabs.com
v1.compute.firewalls.insert
"10.128.0.0/20"
2018-02-19T21:01:51.578UTC
player3@bmlabs.com
v1.compute.firewalls.insert
"0.0.0.0/0"
Short Lived Network Resources
creator
terminator
resource_name
player6@bmlabs.com
player2@bmlabs.com
projects/bmlabs-loggen/global/routes
player2@bmlabs.com
player6@bmlabs.com
projects/bmlabs-loggen/global/routes

Outline:

- Betty Jones
- Value-Based Care
- Multidisciplinary Conference
- Hacking versus Products
- Fairness
- Interpretability
- Security
- Joining Data
- Ethics of Data Science (Corroboration, Continuous QI)