eHEALTH INITIATIVE

AI and Machine Learning: Turn Your Data into Actionable Insights

February 20, 2020

eHEALTH INITIATIVE

Agenda

Welcome and Introductions

- **Claudia Ellison**, Director of Programs, eHealth Initiative

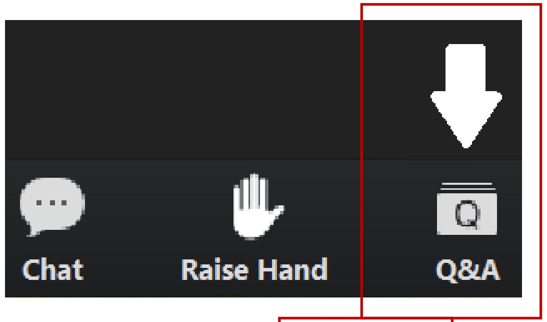
Presenters

- **Cole Erdmann,** Director, Clinical Intelligence, Cerner Corporation
- **Navneet Srivastava**, Senior Solutions Architect, Amazon Web Services
- **William Feaster, MD, MBA,** Vice President, Chief Health Information Officer at CHOC Children's Hospital
- **Janos Hajagos, Ph.D.,** Chief of Data Analytics, Stony Brook Medicine

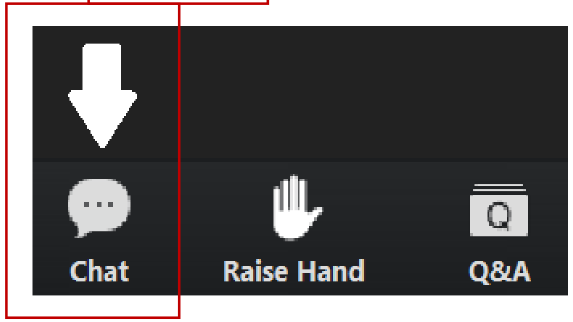
eHEALTH INITIATIVE

Housekeeping

- **All participants are muted**
- Use the **Q&A** box to ask a question related to the presentation



- Use the chat box is for *technical difficulties* and other questions / comments



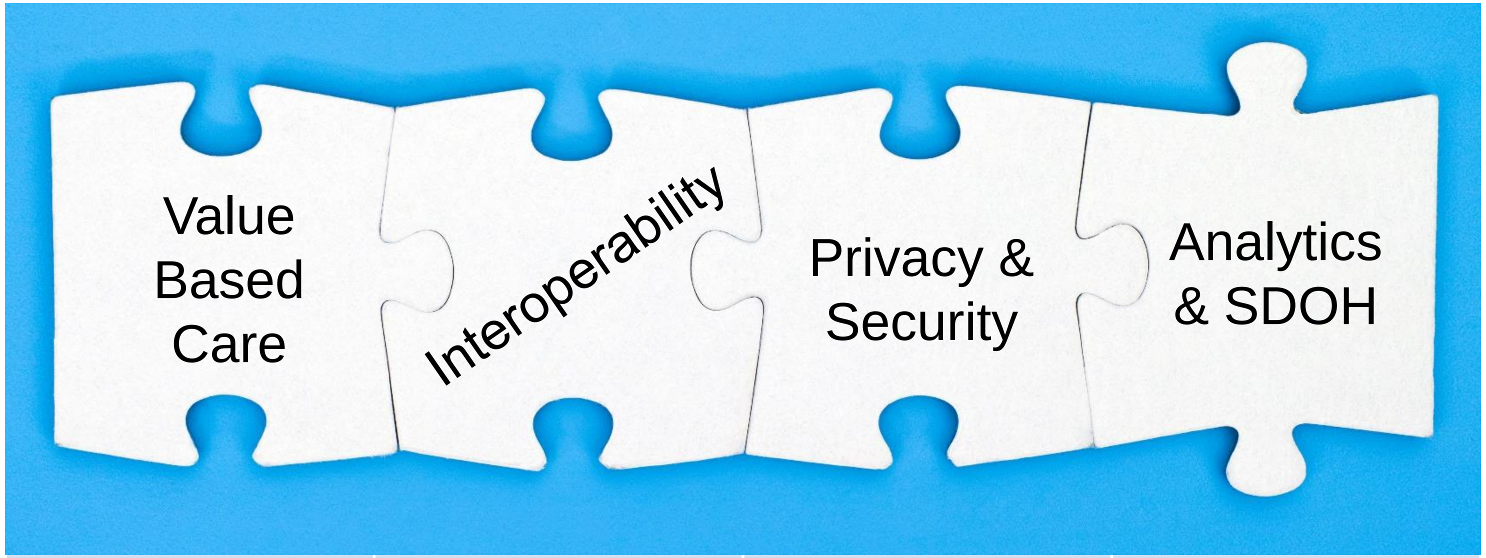
Presentation slides are in the eHI resource Center
https://www.ehidc.org/resources

eHEALTH INITIATIVE

eHI's Mission

To serve as the industry leader in **convening executives** and multi-stakeholder groups to **identify best practices** that **transform healthcare** through the use of t**echnology and innovation**

Current Areas of Focus



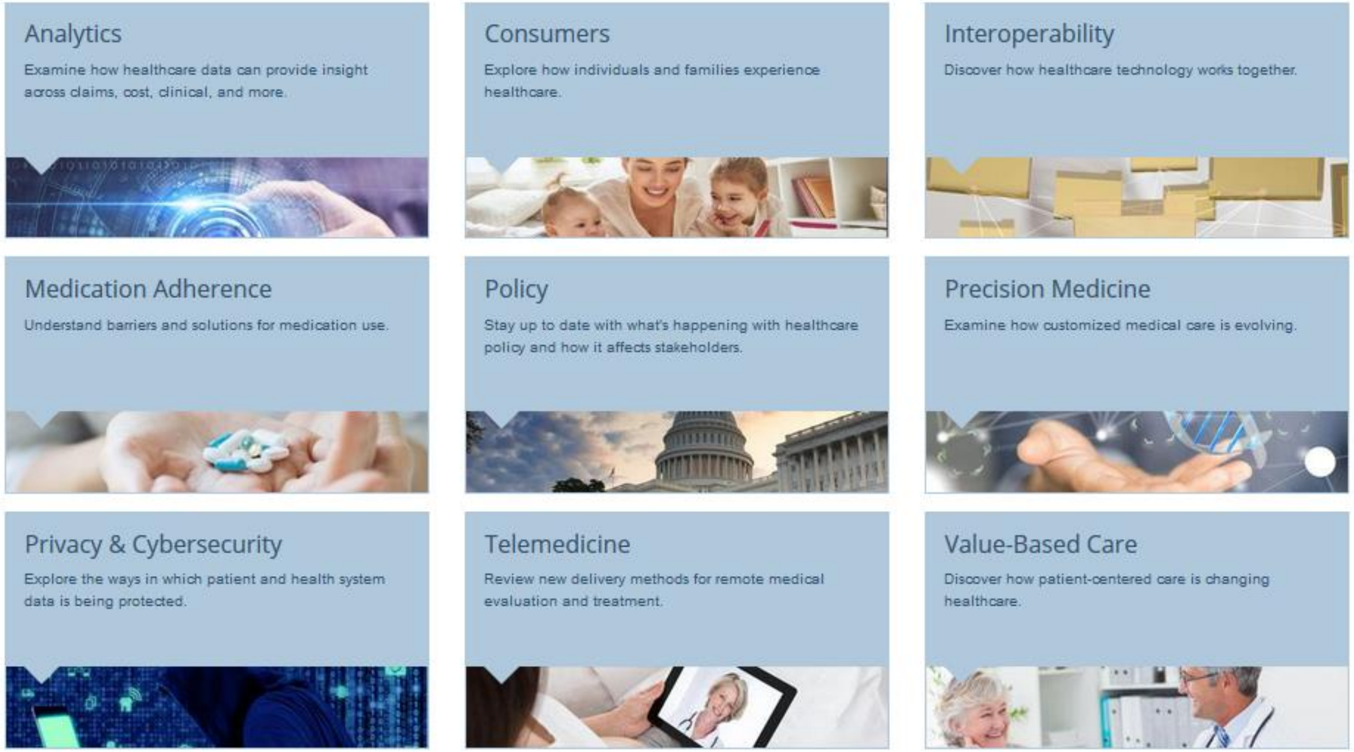
Cost Transparency	Understanding FHIR / APIs / Da Vinci	Protection vs. Access	Non-traditional data sources
Prior Authorization	Policy	HIPAA Part 2	SDOH & PGHD
	Workflow	Cybersecurity	EHR Data for Clinical trials
	Info Blocking	Medical Device Security	

eHEALTH INITIATIVE

eHealth Resource Center

Thousands of Resources

- Best Practices
- Reports
- Surveys
- Policy Briefings
- Comment Letters



eHEALTH INITIATIVE

Who does eHI work with?

In 2019 eHI worked with

- **1,400+ Payers & Providers**
- **3,000+ Stakeholders**

How do we reach stakeholders?

- Webinars
- Roundtables
- Task Forces / Work Groups
- Conferences
- Meetings
- HIMSS

eHI Leadership Council

ACOG The American College of Obstetricians and Gynecologists
AHIMA American Health Information Management Association®
Allscripts

aws
AMERICAN ACADEMY OF FAMILY PHYSICIANS
American Cancer Society®
AMERICAN COLLEGE of CARDIOLOGY
ACP American College of Physicians Leading Internal Medicine, Improving Lives
ACR AMERICAN COLLEGE OF RADIOLOGY
American Heart Association®
American Hospital Association

Availity
BlueCross BlueShield Association
Booz | Allen | Hamilton
CAQH
Cerner
CHANGE HEALTHCARE
DirectTrust

CORHIO COLORADO REGIONAL HEALTH INFORMATION ORGANIZATION
CRISP
EHNAC
Lilly
EPSTEIN BECKER GREEN
fitbit
Google Cloud
GUNDERSEN HEALTH SYSTEM®

HealthCatalyst
HealthCore®
Healthix
HL7
Hogan Lovells
inovalon
InterSystems Health | Business | Government

Johnson & Johnson
LexisNexis® RISK SOLUTIONS
LifeWIRE COMMUNICATING CARE
manatt
Marshfield Clinic
MAYO CLINIC
MedAllies

National Alliance of Healthcare Purchaser Coalitions Driving Innovation, Health and Value
nextgen healthcare
Northwell Health℠
OhioHealth BELIEVE IN WE™
ORION HEALTH
Point-of-Care PARTNERS HEALTH IT MANAGEMENT CONSULTANTS

Providence St.Joseph Health
salesforce
Sonora Quest Laboratories A Subsidiary of Laboratory Sciences of Arizona
STRATEGIC INTERESTS
AT THE FOREFRONT UChicago Medicine
UnitedHealthcare®

eHI eHEALTH INITIATIVE
updox
VALIDIC
verato
welldoc®
Wellmark®
zipnosis®

eHI Events and Meetings (https://www.ehidc.org/events)

- **eHI VIP Networking Reception at HIMSS**

  Tuesday, March 10, 2020
  5:30 – 7:00 pm
  OCCC, EF Overlook

eHEALTH INITIATIVE

This webinar was made possible through the generosity and support of

Today's Speakers

Cole Erdmann
Director, Clinical Intelligence, Cerner Corporation

Navneet Srivastava
Senior Solutions Architect, Amazon Web Services

William Feaster, MD, MBA
Vice President, Chief Health Information Officer at CHOC Children's Hospital

Janos Hajagos, Ph.D.
Chief of Data Analytics, Stony Brook Medicine

Solving data science challenges

Complex data

- Manage multiple data sources
- Aggregate and normalize disparate data models

Disparate infrastructure

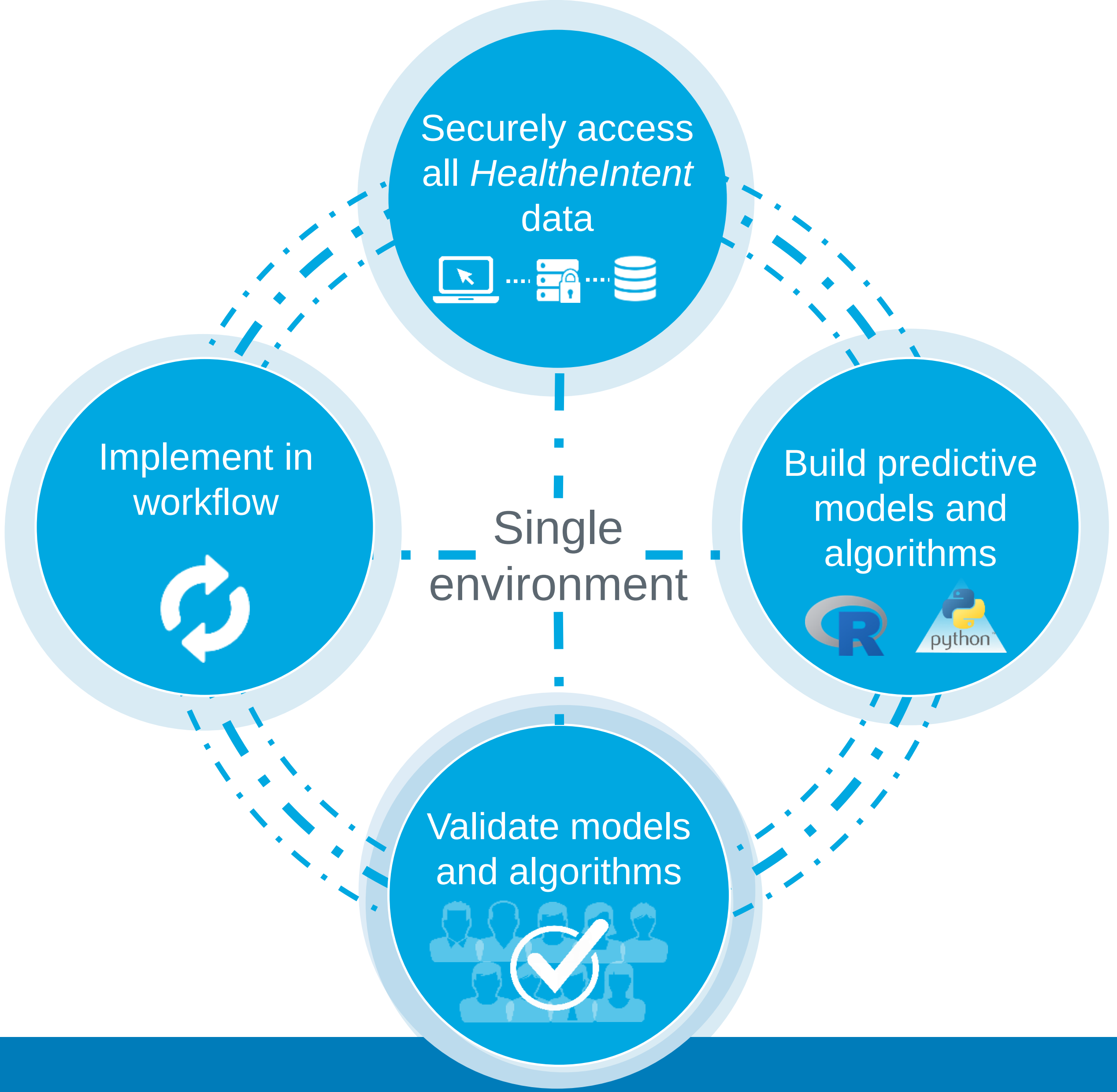
- Meet demand by automatically provisioning and de-provisioning workloads in an elastic, adaptable environment
- Scale from large, complex to smaller, more targeted analysis

Limited adoption

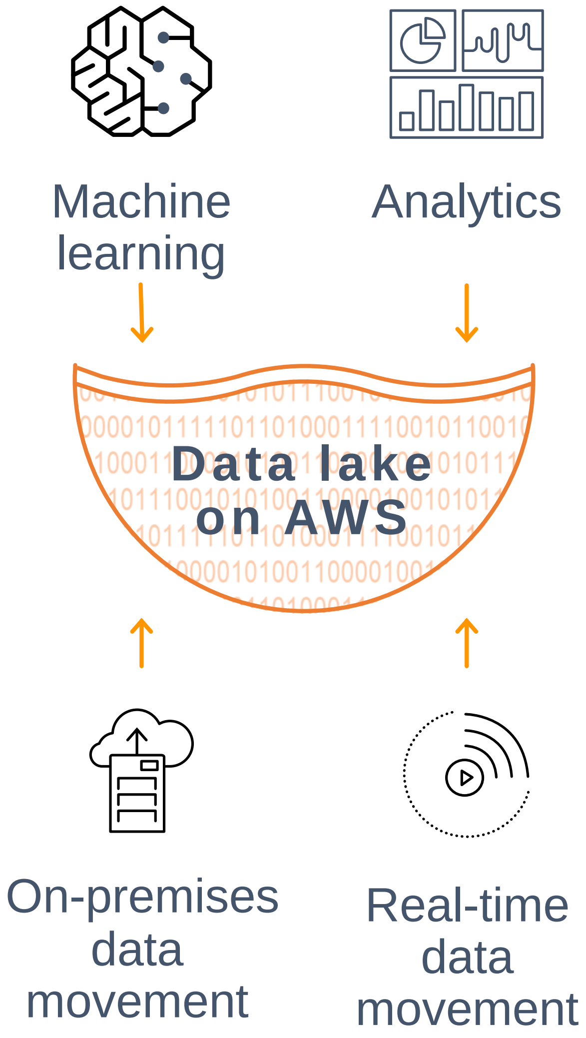
- Utilize industry-leading, sophisticated tools
- Access data and models via an easy-to-use, secure environment

Data science ecosystem

HealtheDataLab



Data lakes and analytics from AWS



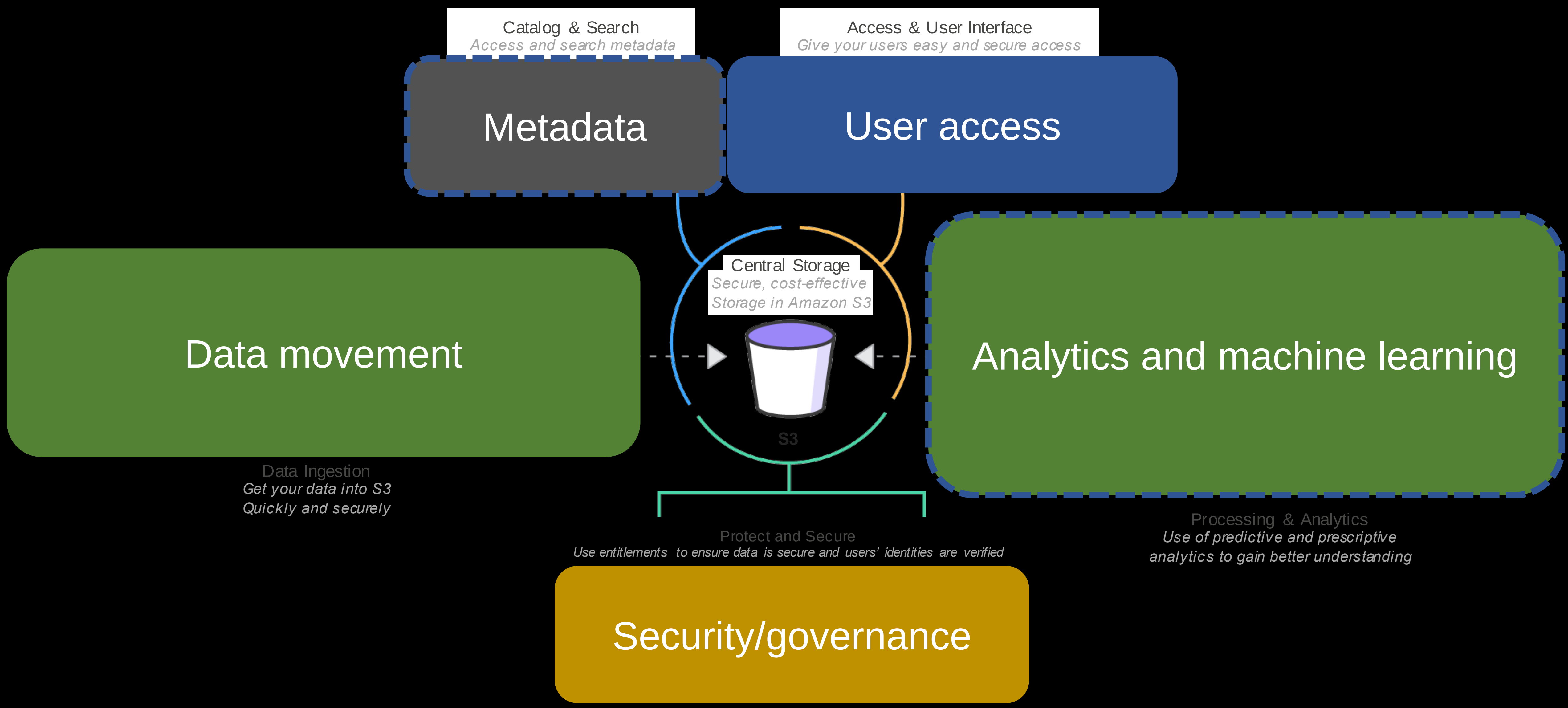
Open and comprehensive

Secure

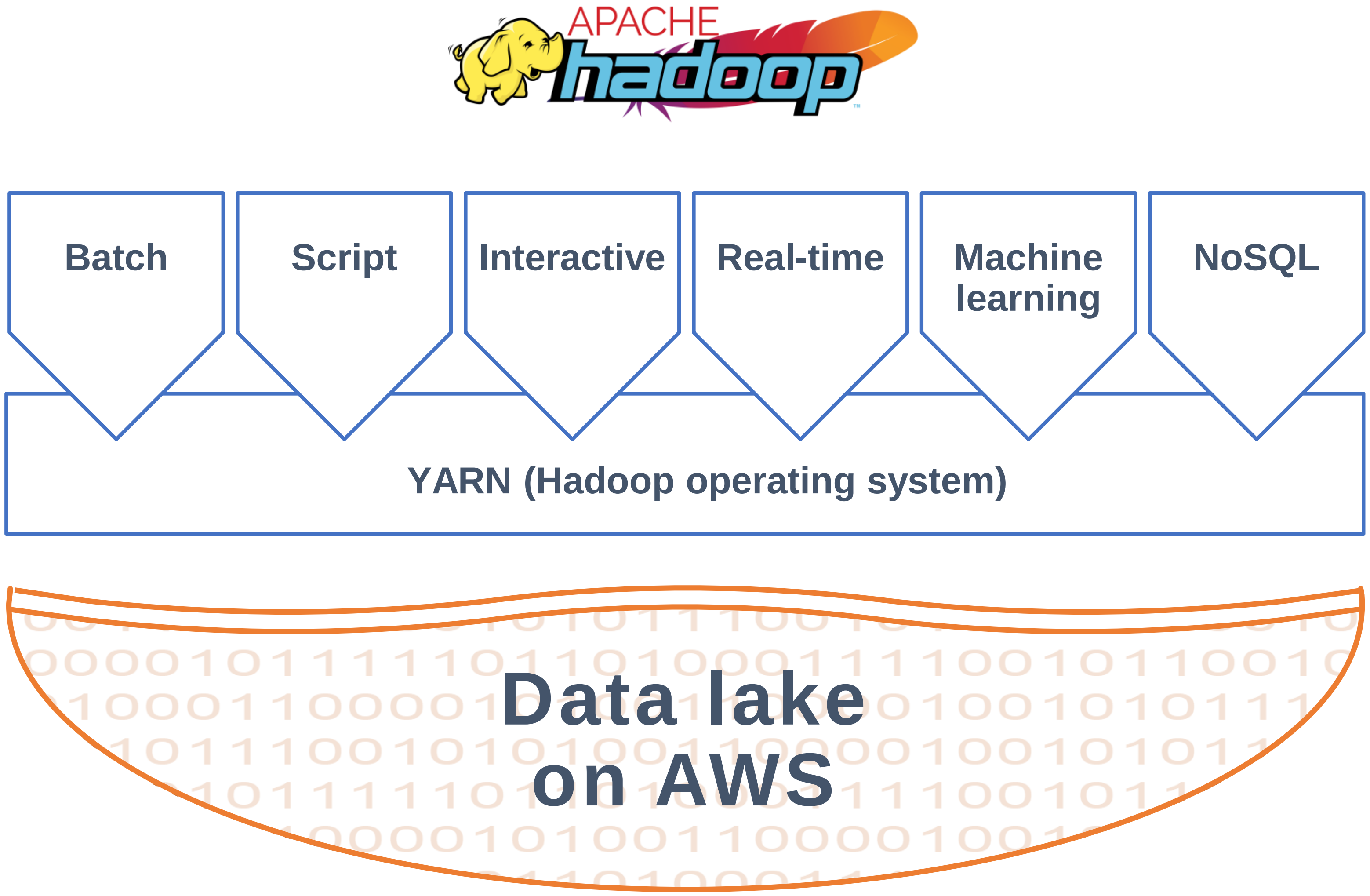
Scalable and durable

Cost-effective

Data lake components



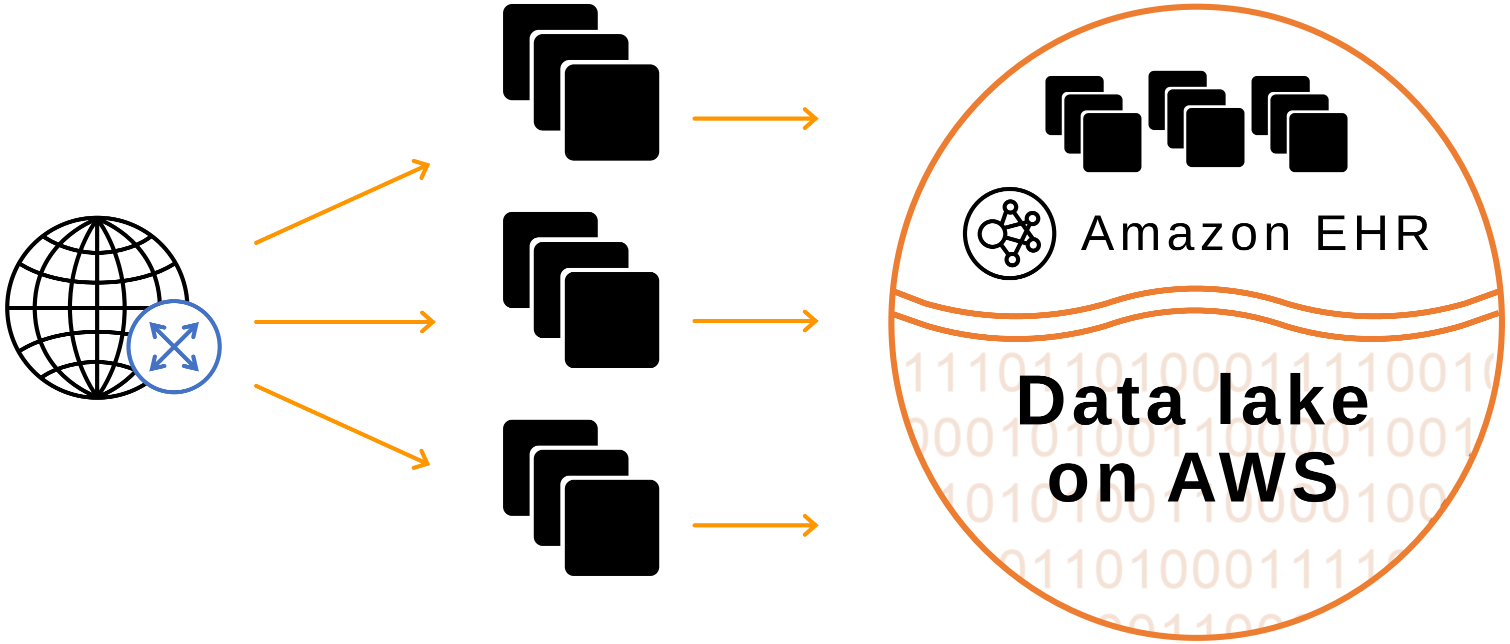
Apache Hadoop on the data lake



- Distributed processing
- Diverse analytics
  - Batch/script (Hive/Pig)
  - Interactive (Spark, Presto)
  - Real-time (Spark)
  - Machine learning (Spark)
  - NoSQL (HBase)
- For many use-cases
  - Log and clickstream analysis
  - Machine learning
  - Real-time analytics
  - Large-scale analytics
  - Genomics
  - ETL

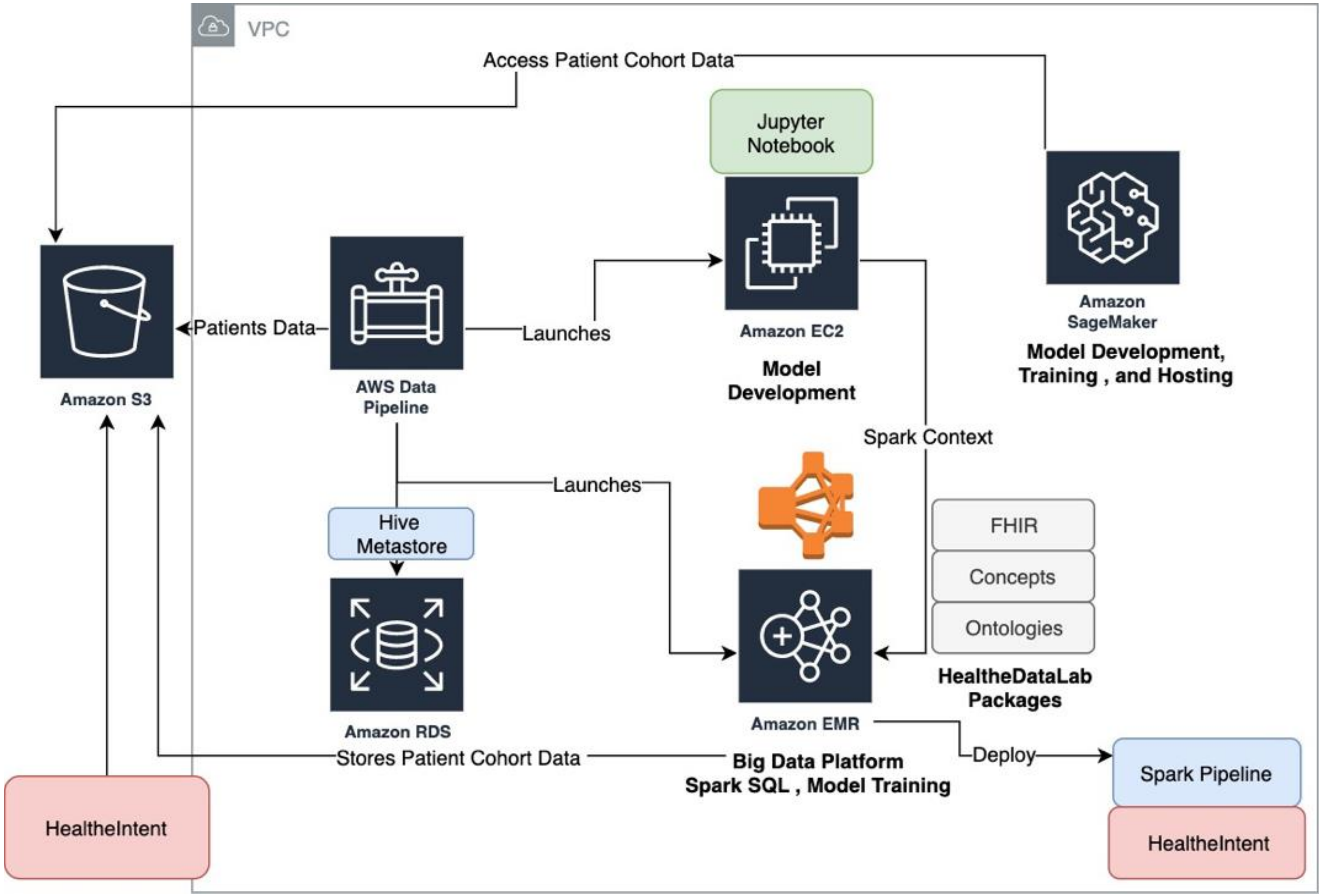
Enterprise-grade Hadoop and Spark

Scale to any size



- Scale compute (EMR) & storage (S3) independently
- Store and process any amount of data—PB to EBs
- Provision one, hundreds or thousands of nodes
- Auto-scaling

HealtheDataLab architecture



Security is job zero

- People & process
- System
- Network
- **Physical**

Familiar security model

Validated and driven by customers' security experts

Benefits all customers

Shared responsibility model

Customers have their choice of security configuration **IN** the cloud

Customer data

Platform, applications, identity & access management

Operating system, network & firewall configuration

Client-side data Encryption & data Integrity authentication	Server-side encryption (file system and/or data)	Networking traffic Protection (encryption integrity, identity)

AWS is responsible for the security **OF** the cloud

Software

Compute	Storage	Database	Networking

Hardware/AWS global infrastructure

Regions	Edge locations
Availability zones	

Implementing a cloud computing solution for advanced analytics and data science in health care

William W. Feaster, MD, MBA

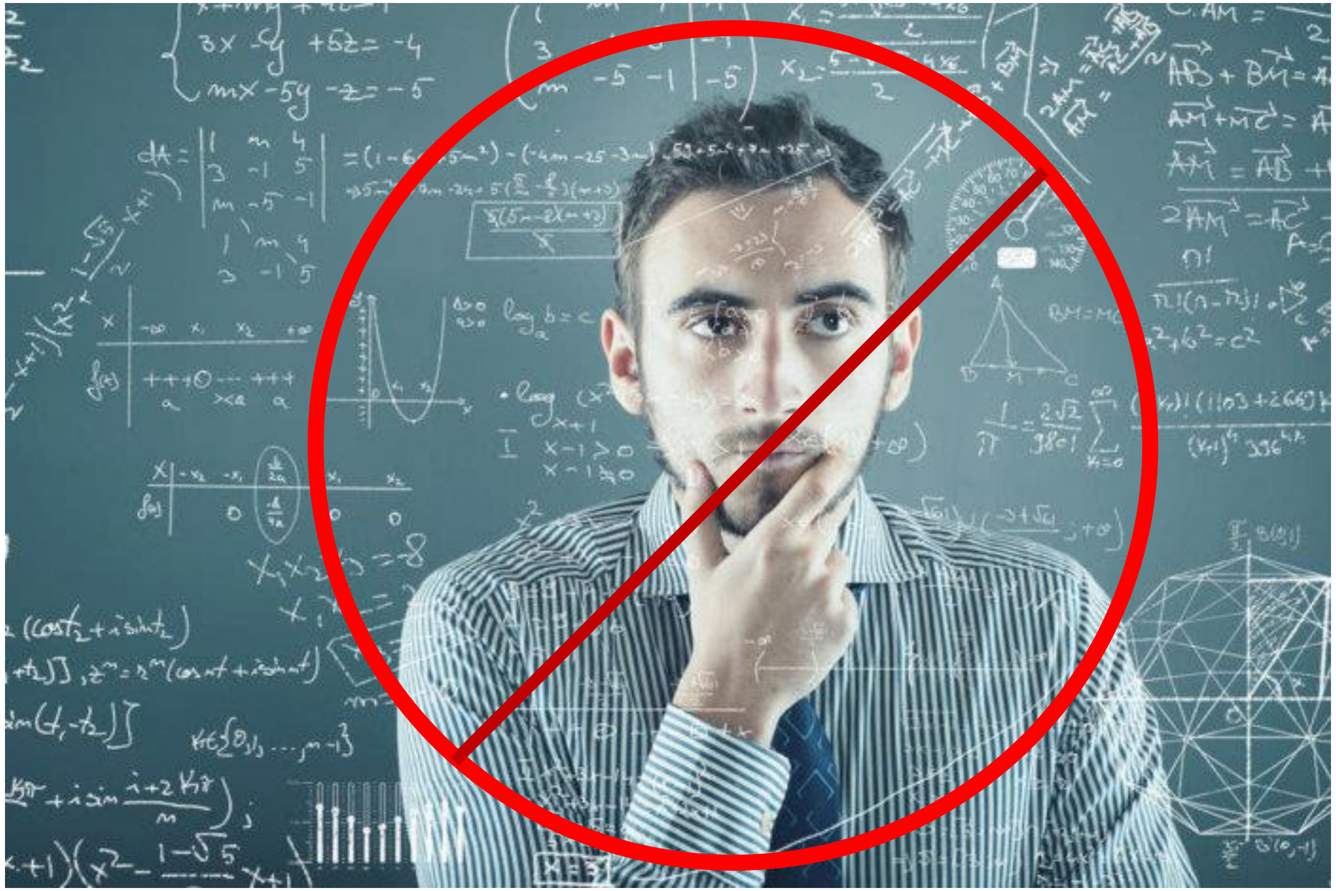
V.P., Chief Health Information Officer

Children’s Hospital Orange County

Objectives

- Explore the scope and challenges of utilizing health care data
- Explain how we have applied advanced analytics and data science to predict readmissions
- Describe our prior on-premises and current cloud-based computational environments that aids the work of the data scientist
- Illustrate how CHOC Children's has applied advanced analytics and data science back into the EHR and process of care

Disclaimer




Health care is awash in a sea of data

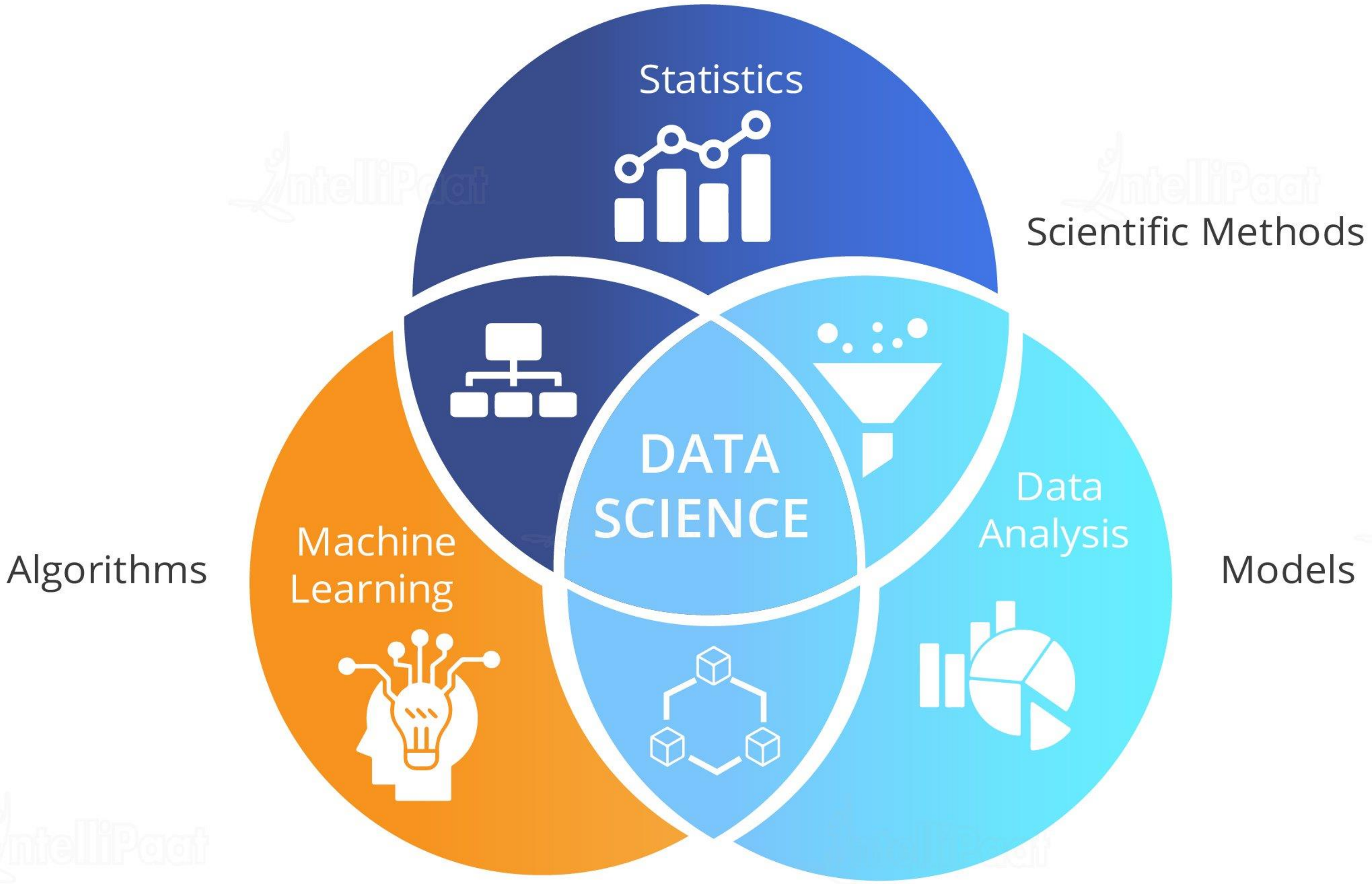
Health care data

- From the EHR
  - The good news is all patient data in the EHR are readily available at the point of care
  - The bad news is all patient data in the EHR are readily available at the point of care
  - 40% in notes
- Images – radiology, CT, MRI, ultrasound
- Lab data
- “Omics” of various types – genomic markers for disease risk, pharmacogenomics to guide prescribing, tumor genomics to guide therapies
- Streaming data from acute care monitors, wearables, continuous glucose monitors, etc.
- Finance data, billing data, claims data, satisfaction data

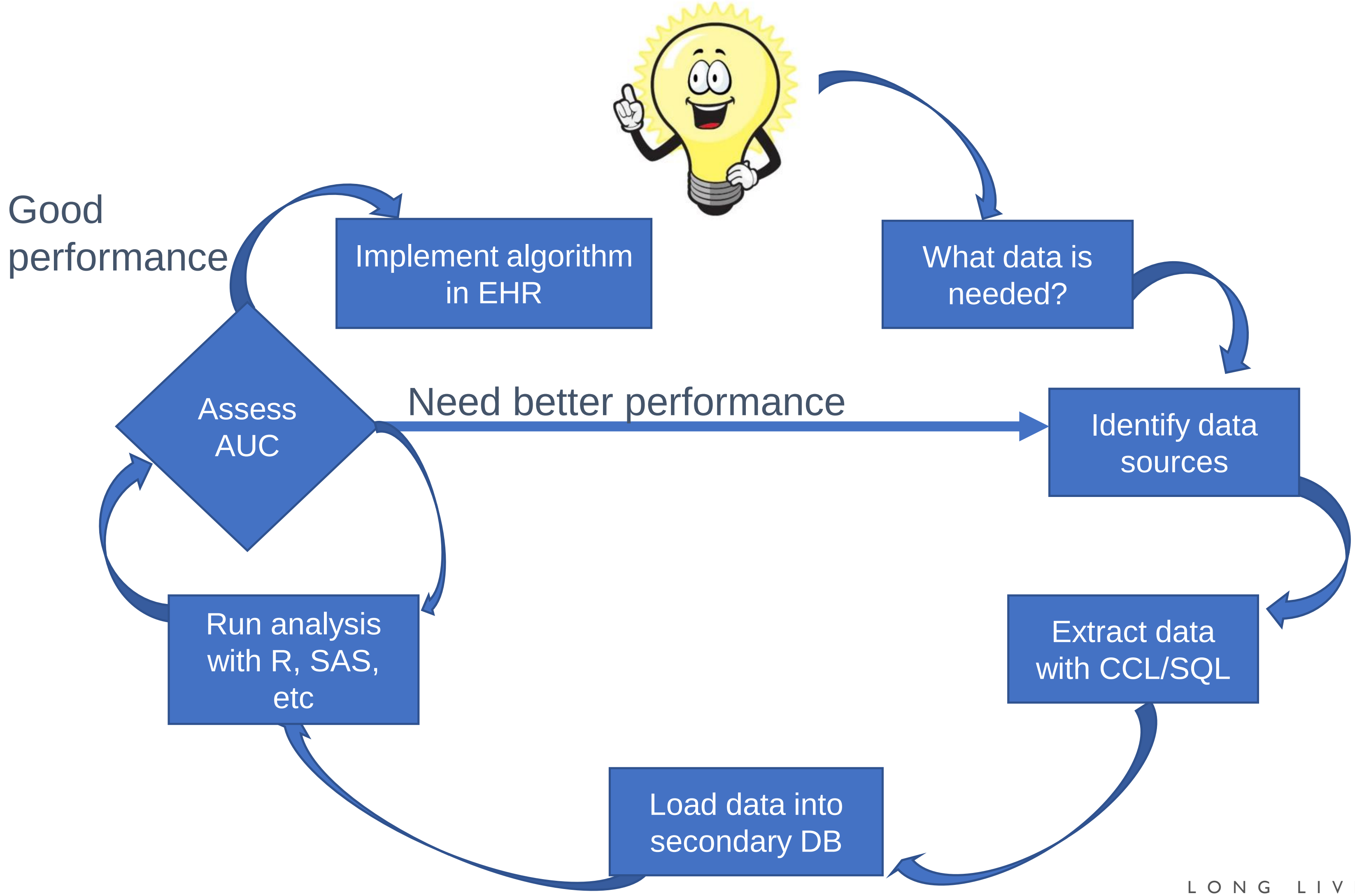
Predictive analytics aim to take advantage of EHR data through computational data science

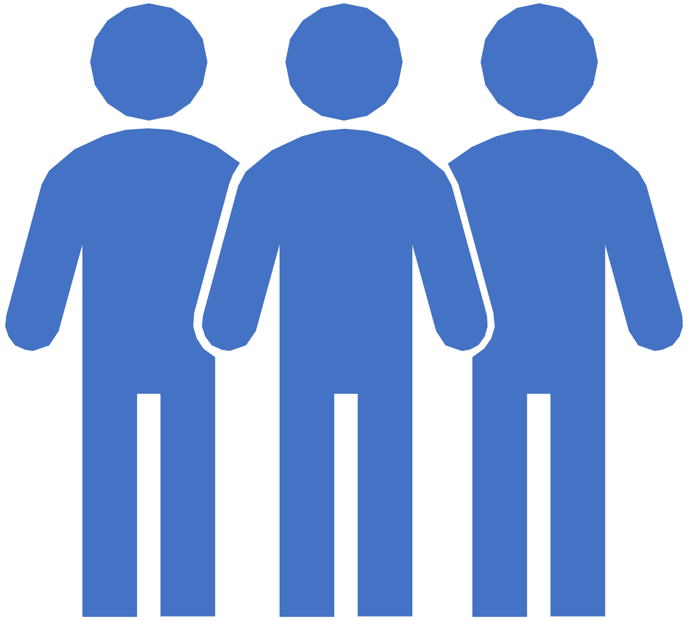
- Takes into account all available data on all qualifying patients
- Explores associations and statistical interactions between all variables
- Uses a data-driven and statistically sound approach to determine the most important set of variables that impact care for the event of interest
- Predicts the future occurrence of an event or classify patients at high risk of poor outcomes
- Guides caregivers to intervene on patients identified to be at high risk and improves quality of care outcomes

Data science



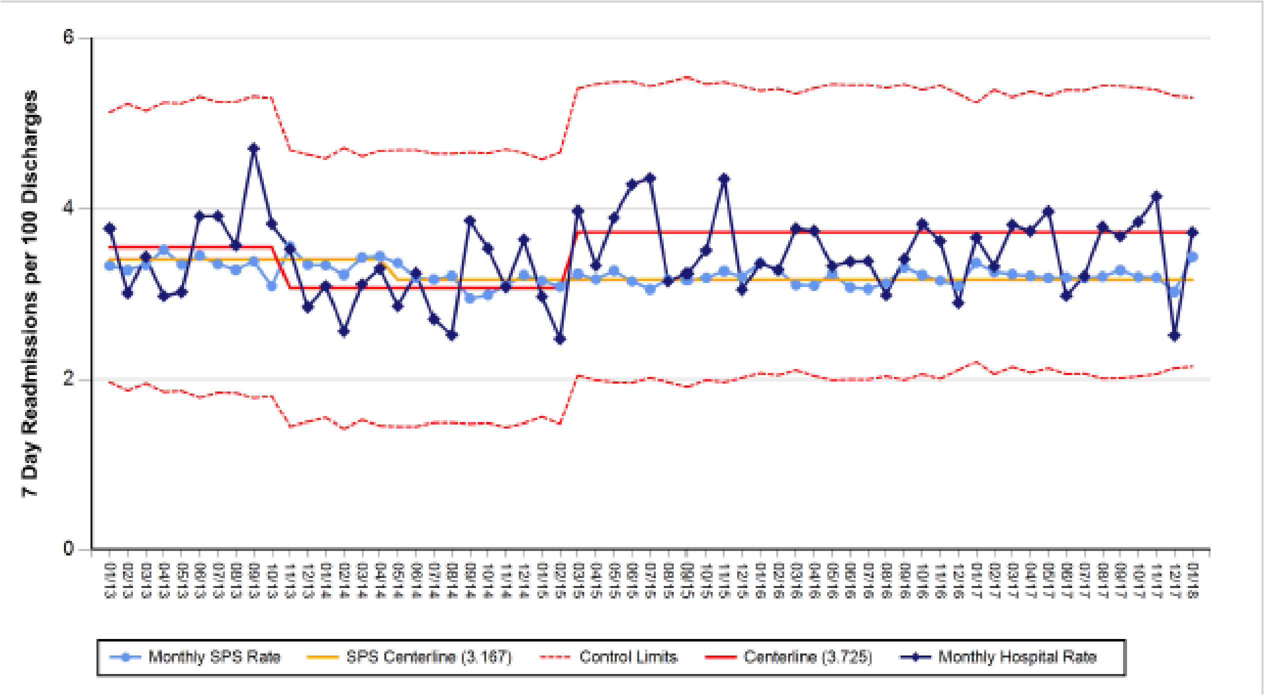
Applying predictive analytics to an identified clinical question (pre HealtheDataLab)





Applying data science to predict readmissions

Reducing readmissions



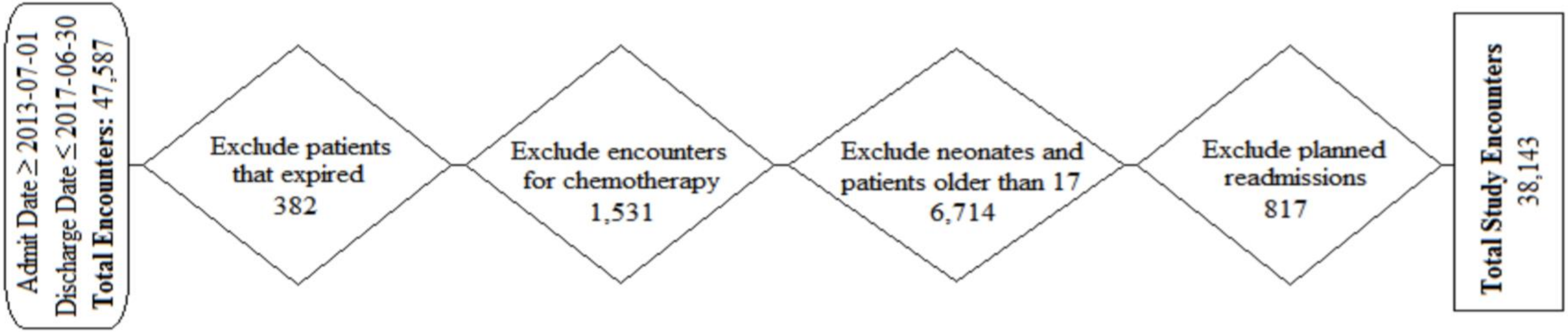
Annotations

09/17	Care Team - ED started
09/16	Care Team - Red started
05/16	Case Mgmt Refer To orders
11/15	Care Team - Blue started
08/15	Orders-F/U appts/tests by PAC
06/15	DC checklist/New Admit Booklet
04/15	DC phone call Cerner form
02/15	Med to Bed housewide
10/14	Med to Bed Pilot
09/14	Seizure Action Plan
03/14	Low risk oncology fever guideline
11/13	High risk groups analyzed
06/13	Readmission events emailed to MDs

- While CHOC has had great success in applying guidelines to care, and have reduced length of stay and readmissions for certain diseases, our overall 7-day readmissions (solutions for Patient Safety metric) hasn't decreased for the overall organization, hovering at around 4% or near the national average for children's hospitals
- We've tried many ways to improve this number without success
- We're now turning to data science…

The goal of our original model

Build the most predictive model for 30-day readmissions under the most stringent set of inclusion/exclusion criteria

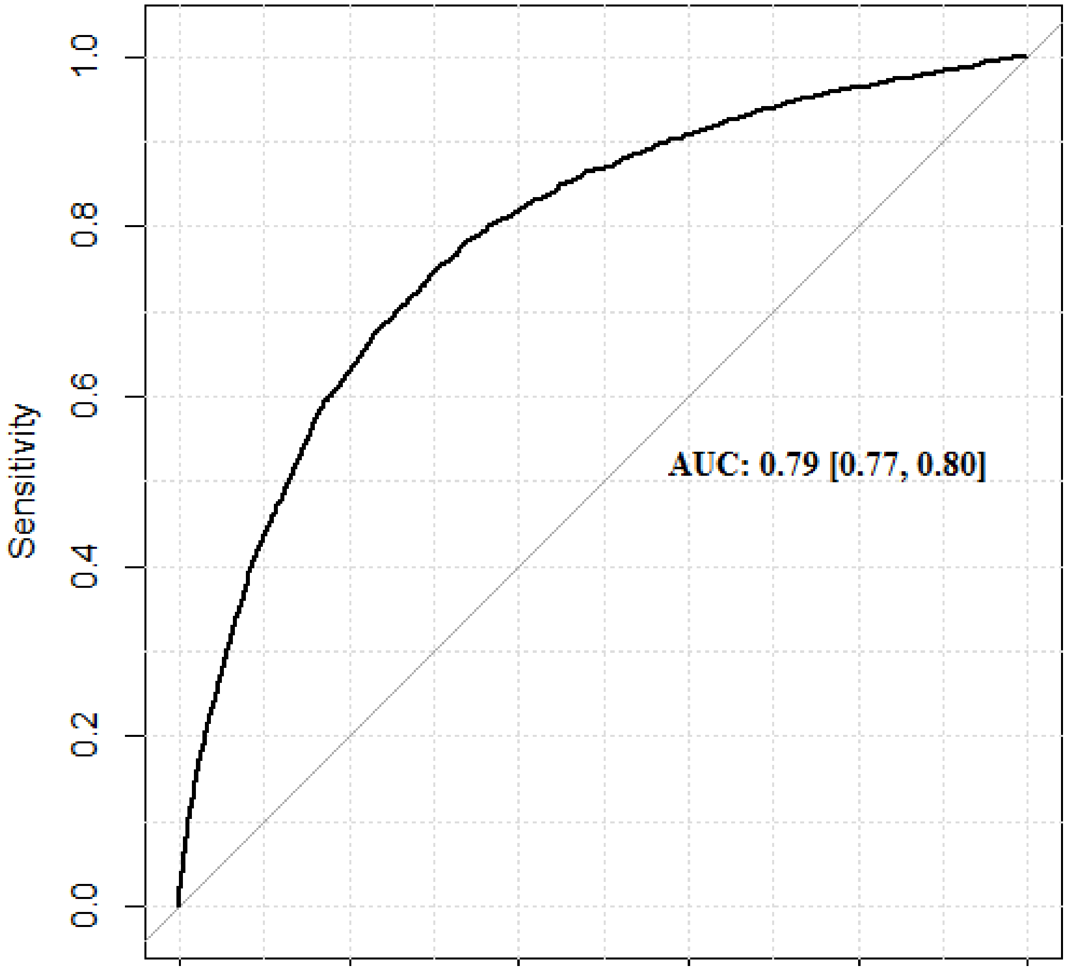


30-day readmissions

- Study takes into account variables novel to those studies, including:
  - Rothman Index
  - Prior readmissions and whether current admission is a readmission
  - Medication classes
- 30 day predictive algorithm actually predicts 7 day readmissions better than a model specific to 7 days

Performance of the 30-day readmission model

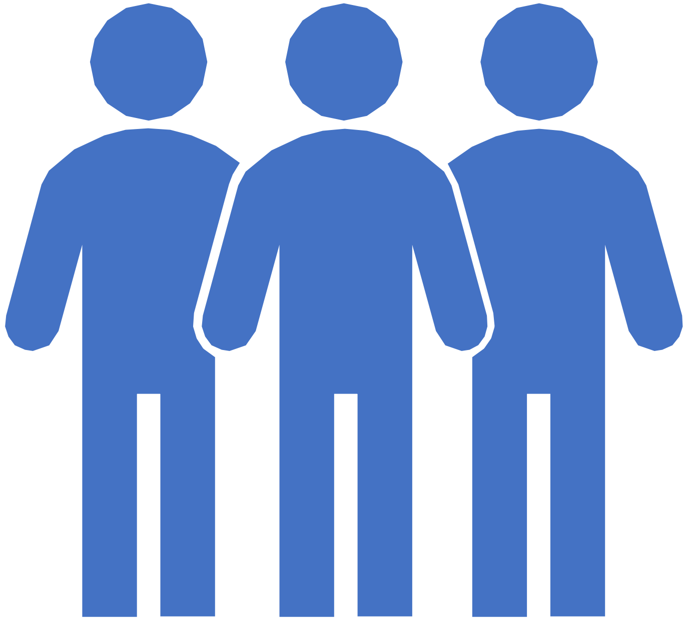
- AUC: 0.79 [0.77, 0.80]
  - Highest for pediatric readmission models with encounters for chemotherapy excluded
- Operating threshold:
  - We chose the following thresholds on analyzing the performance of the model and predicted probabilities:
  - **High risk** – Predicted probability greater than or equal to 0.22
  - **Medium risk** – Predicted probability between 0.11 and 0.22
  - **Low risk** – Predicted probability less than 0.11
- *Among "high risk" patients, predicted 3 will be flagged daily of which 1 is guaranteed to be readmitted in the absence of an intervention*



Initial clinical application of readmission model

- We initiated the use of our published model via spreadsheets utilized by our inpatient case management staff
  - Those at high readmission risk worked daily from admission to discharge
    - They determine interventions required
- Performance of the model is actually better than we predicted
- For patients identified as:
    - High risk, 52% were readmitted in 30 days
    - Medium risk, 26% were readmitted in 30 days
    - Low risk, 5% were readmitted in 30 days
- Follow up measurements of 7-day readmissions show the rate trending down (statistically significant drop)

Exploring data with HealtheDataLab



In our ideal data science environment

- We would not be limited by access to data
- Data elements needed for analysis would be easy to identify
- We would not have to spend an extended time preprocessing this data
- We would not be limited by computing resources
- We would have access to all available big data analysis tools

Enter HealtheDataLab

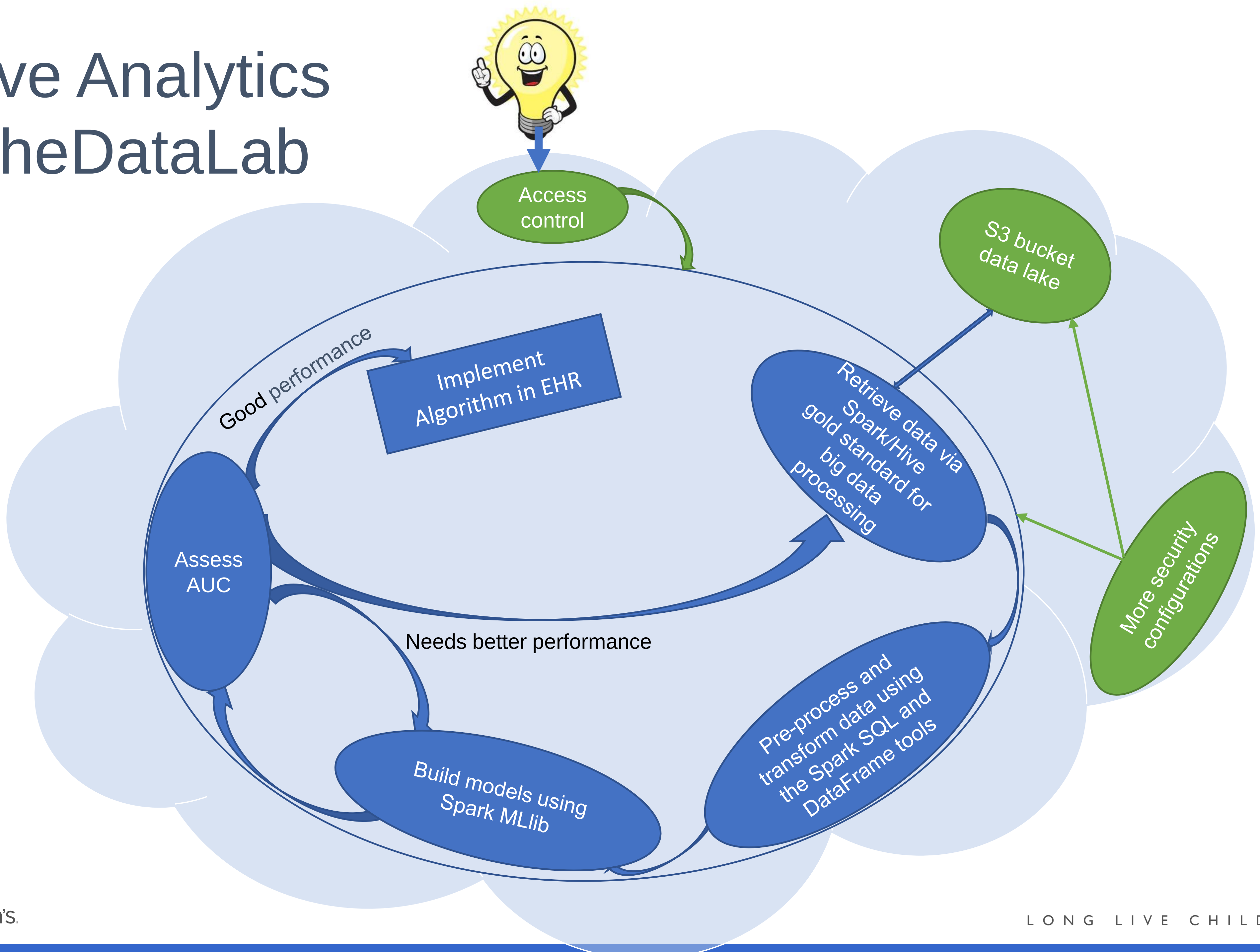
- Cerner approached us to be its development partner for a new, client-facing cloud computing and big data analytics and insights platform, powered by AWS

- Single data science environment
  - Complex and disparate data are easily stored and accessible
  - Common big data analytic tool set is readily available
  - Elastic MapReduce for high performance distributed computing

HealtheDataLab data

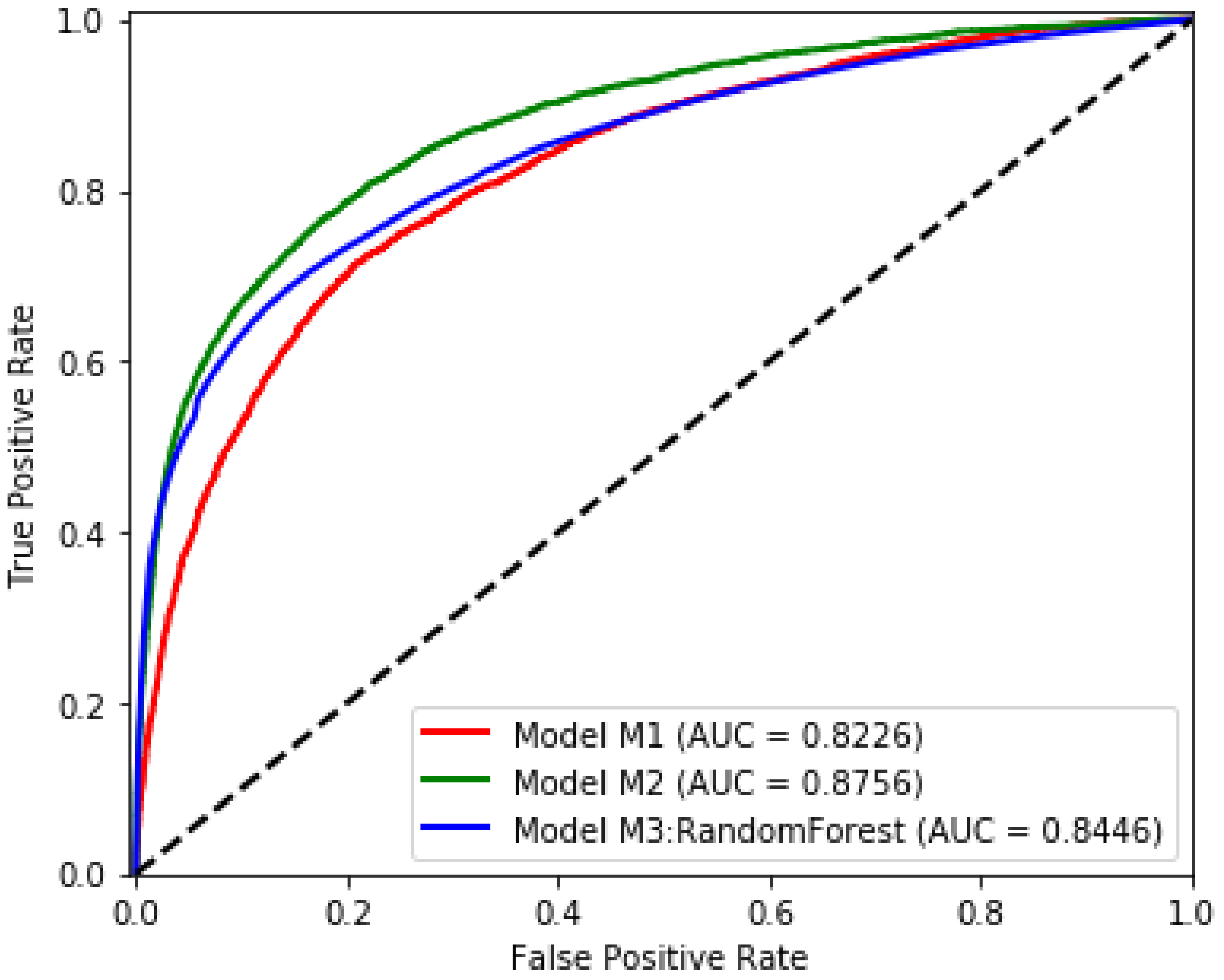
- Our data from our EHR and other data sources in *HealtheIntent*® is in *HealtheDataLab*™ in both identified and deidentified form
  - Data mapped to S3 bucket (encrypted data lake) utilizing FHIR data definitions where available
- *Cerner Real-World Data*® database – deidentified database of 68.7million patients, 503.8 million encounters across 600 health care facilities
- We've loaded other data into the S3 bucket
  - Rothman Index data to validate prior readmission model
  - Mimic III ICU database – deidentified data from MIT, 46,520 ICU patients and 61,532 ICU stays

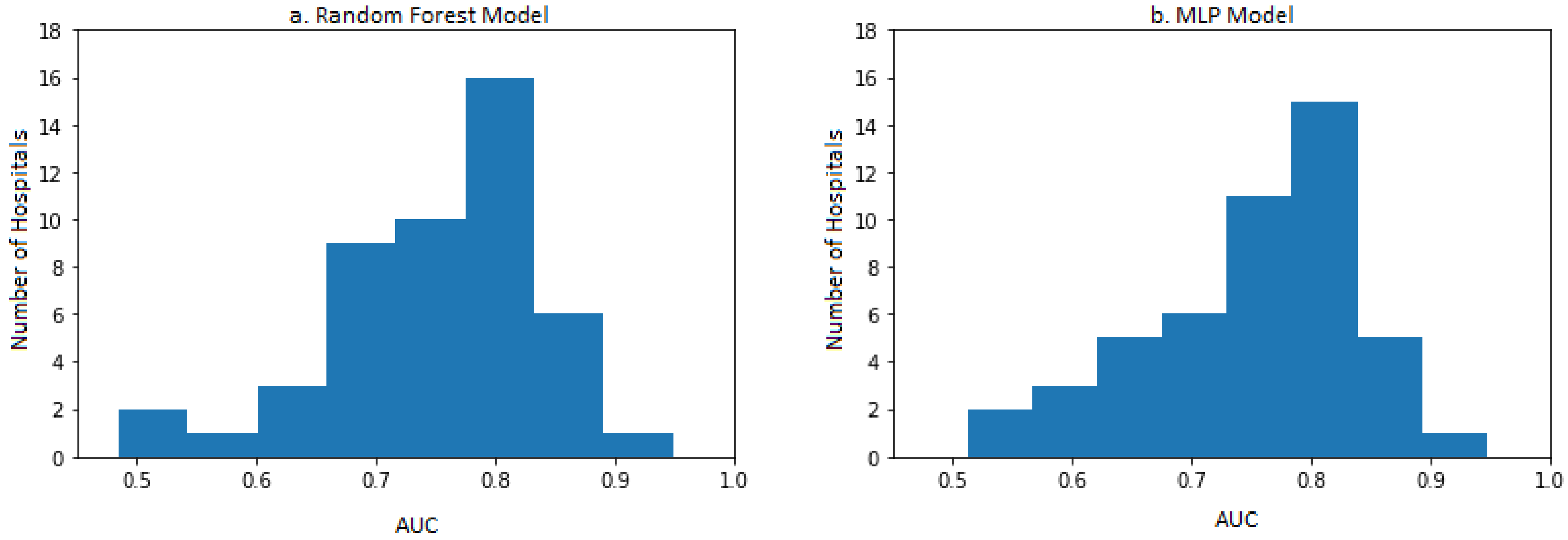
Predictive Analytics in HealtheDataLab



Our readmission work in HealtheDataLab

- Validate the environment by replicating prior published study (M0)
- Provide methodological improvement over our previous study (M1)
  - Exposure to more patients and more data elements
- Re-run the analysis on all readmissions, not just unplanned to prepare for work on the HF database (M2)
- Run a multi-center model for predicting general all-cause 30-day readmissions among pediatric-age patients (patients less than 18 years) using the HF database (M3)
  - 1.4 million children hospitalized at 48 hospitals over a 10 year period with higher pediatric volumes in the HF database


1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate
Model M1 (AUC = 0.8226)
Model M2 (AUC = 0.8756)
Model M3:RandomForest (AUC = 0.8446)

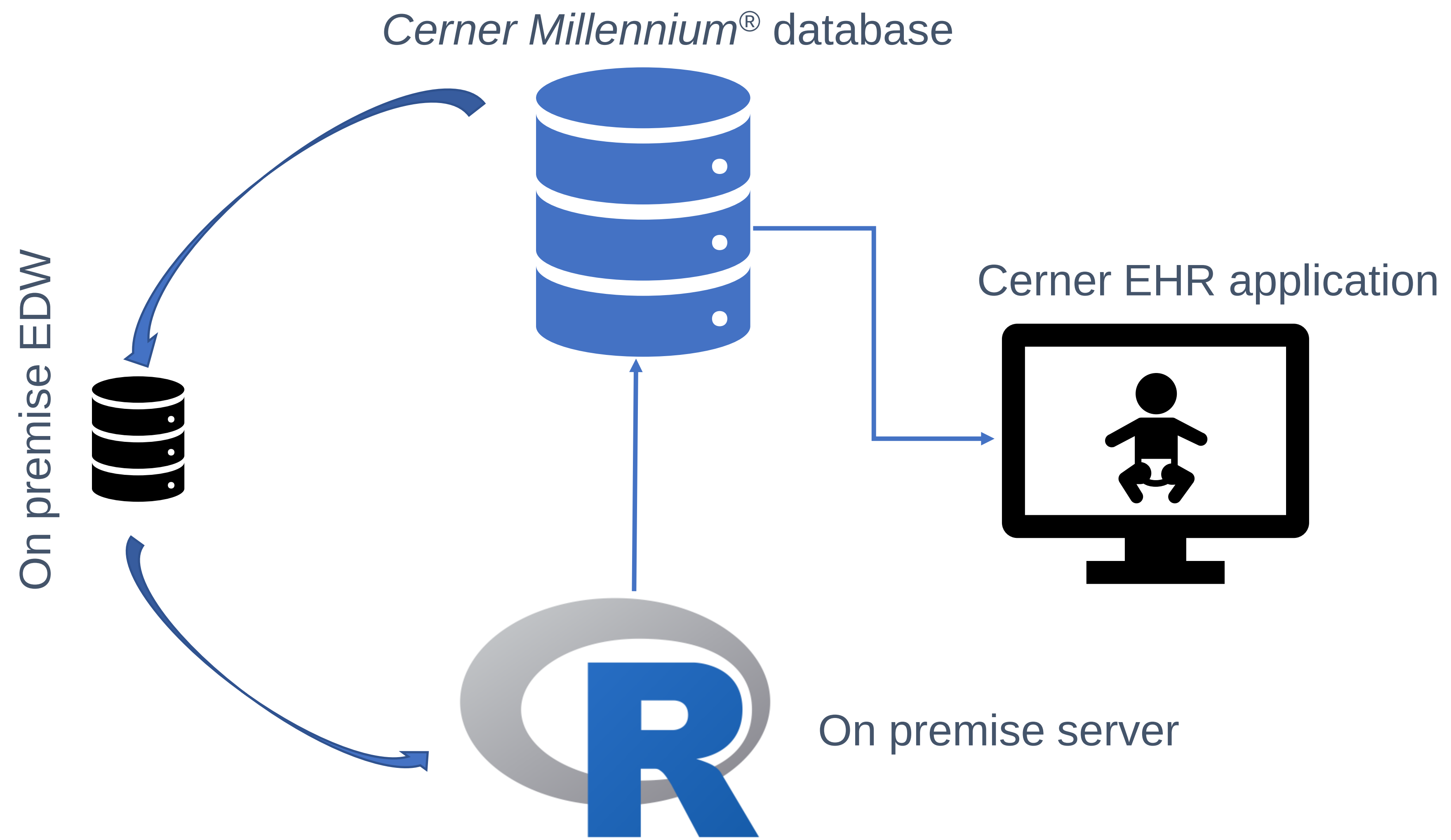


a. Random Forest Model
Number of Hospitals
AUC
0.5
0.6
0.7
0.8
0.9
1.0
0
2
4
6
8
10
12
14
16
18
b. MLP Model
Number of Hospitals
AUC

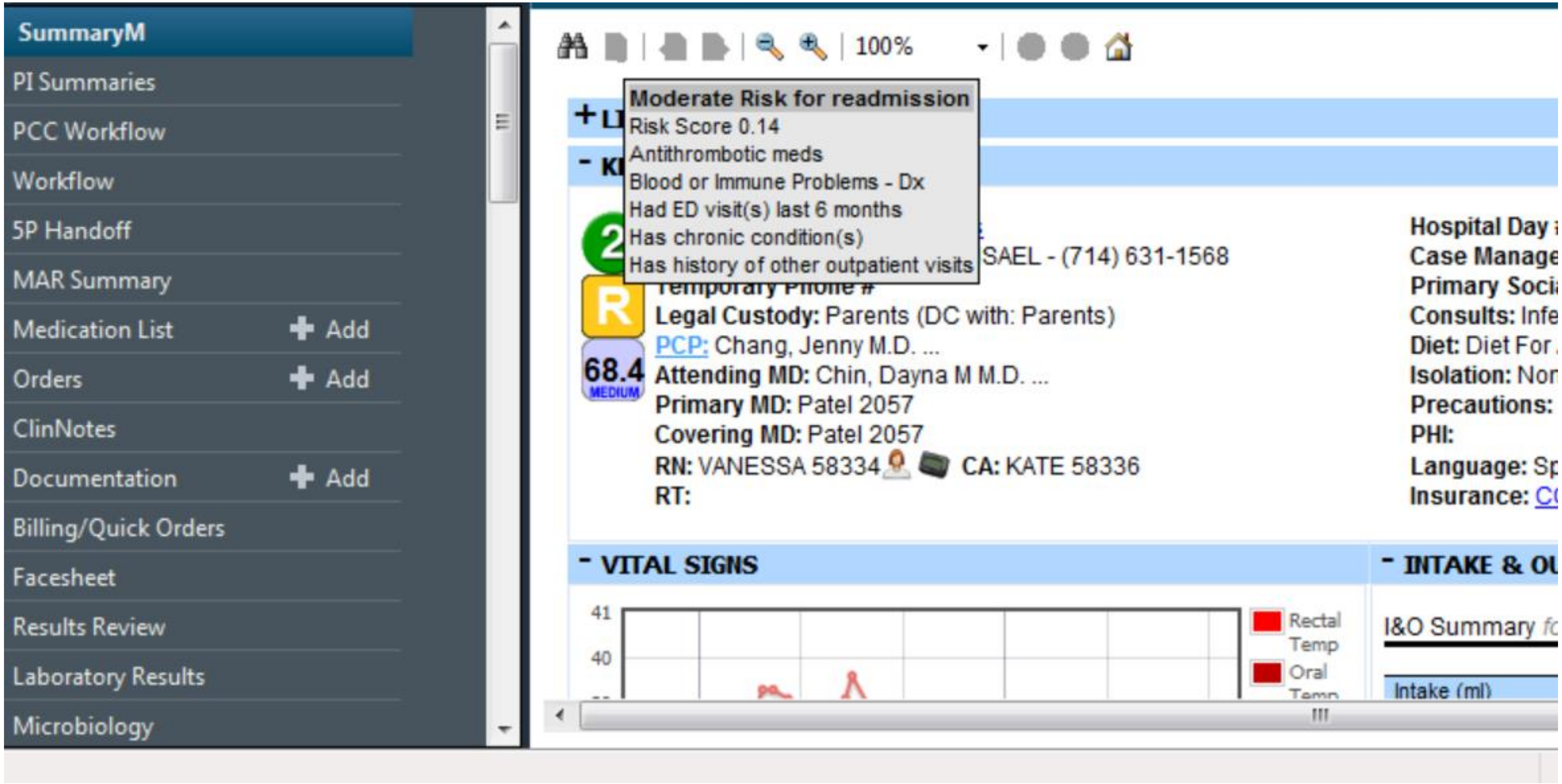
Current implementation of our readmission predictive model

- We've refined our readmission model as described above
- The real challenge is bringing the results of these models back to caretakers in the EHR
  - Patient lists can replace the spreadsheets for care managers
  - Patient specific results can be used in patient care
  - Alerts related to those results can fire as appropriate

Integration into the EHR



Patient example



RESEARCH ARTICLE

A Statistical-Learning Model for Unplanned 7-Day Readmission in Pediatrics

Louis Ehwerhemuepha, PhD, Karen Pugh, MSN, RNC, Alex Grant, BS, Sharief Taraman, MD, Anthony Chang, MD, MBA, MPH, MS, Cyril Rakovski, PhD, William Feaster, MD, MBA.

HOSPITAL PEDIATRICS Volume 10, Issue 1, January 2020

Since onboarding HealtheDataLab

- We’re currently developing numerous predictive models for asthma outcomes, early diagnosis of sepsis after ED triage, ICU transfers, rising risk prediction, among other projects
- 6 more papers under current publisher review
- 8 more will be submitted over the next several months
- The real challenge is bringing the results of these models back to caretakers in the EHR

HealtheDataLab: a next-generation analytic environment

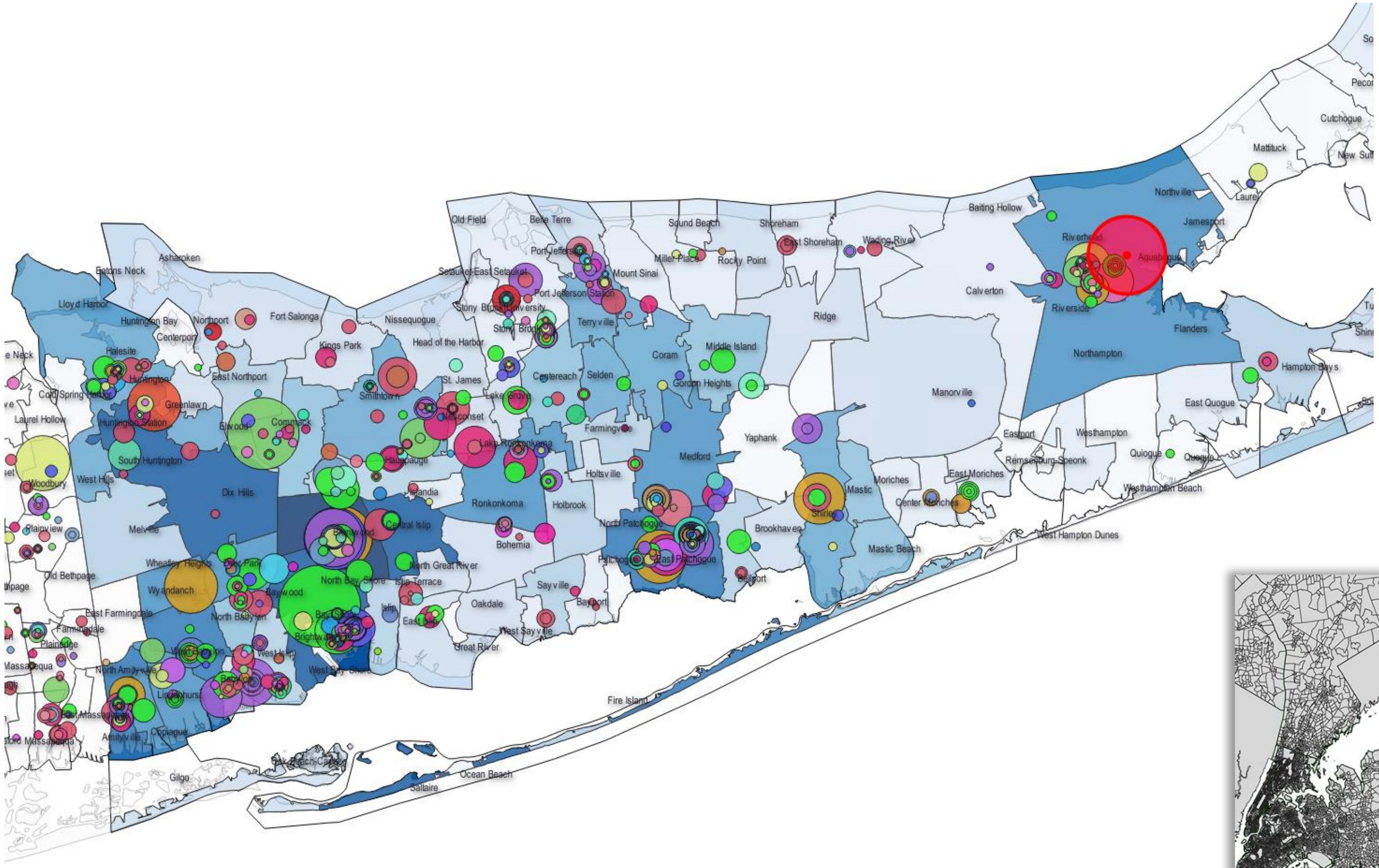
Janos G. Hajagos, Ph.D.
Chief of Data Analytics
Stony Brook Medicine
Research Assistant Professor
Department of Biomedical Informatics
Stony Brook University

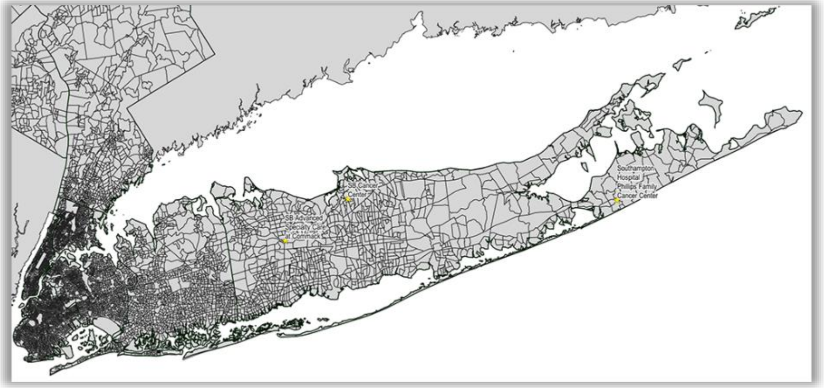
Stony Brook, New York

Stony Brook Medicine

- Academic medical center (600 beds)
  - Two consecutive years “America’s 100 Best Hospitals” by Health Grades
- Two community hospitals on the eastern end of Long Island
- DSRIP award (2015) managing the Medicaid population of Suffolk County, New York
- Research campus of the State University of New York
- Leading department of Biomedical Informatics combining Engineering and Medicine

Population health:
provider map and Medicaid diabetic patients

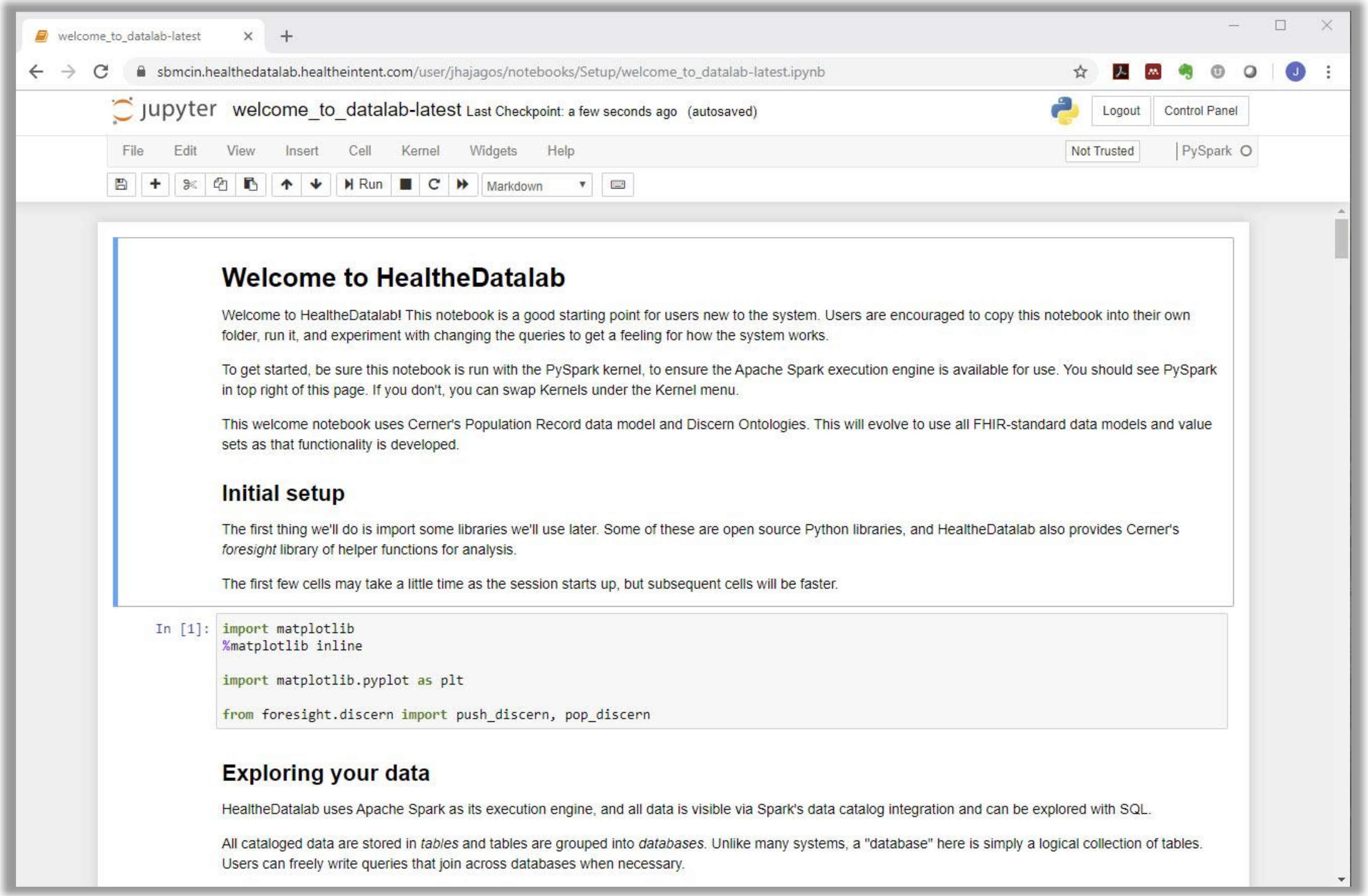




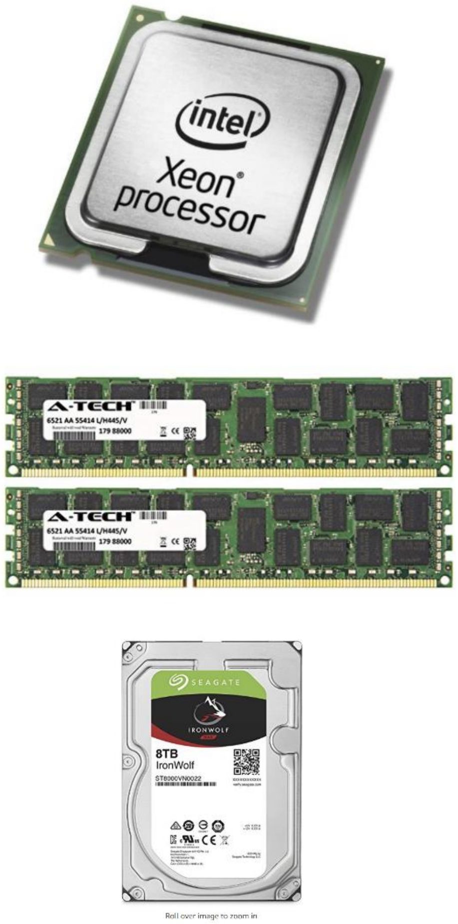
No single tool or environment

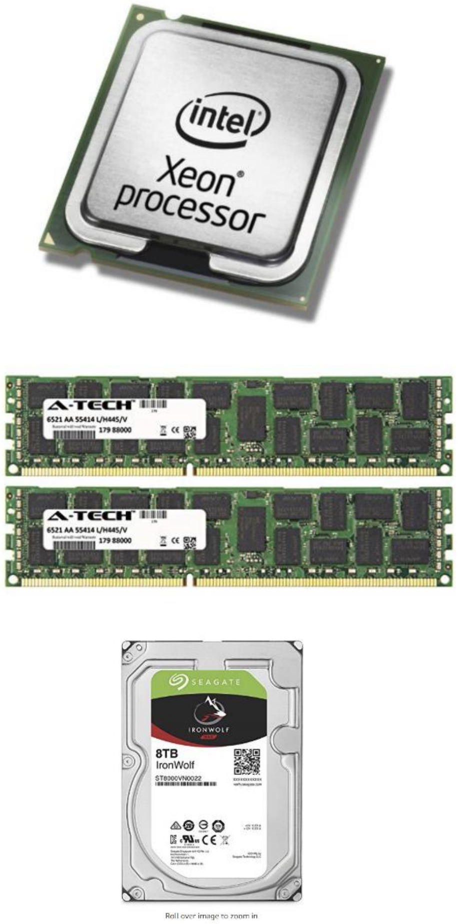
- Data for population health from multiple sources
  - EHRs
  - Administrative systems and claims
  - Public sources
    - ACS: American Community Survey (census data)
    - NPPES: National Plan & Provider Enumeration (NPI database)
- Need tools which can combine data from multiple sources
- Efficiently query large volumes of data
- Build machine learning models with cutting-edge algorithms

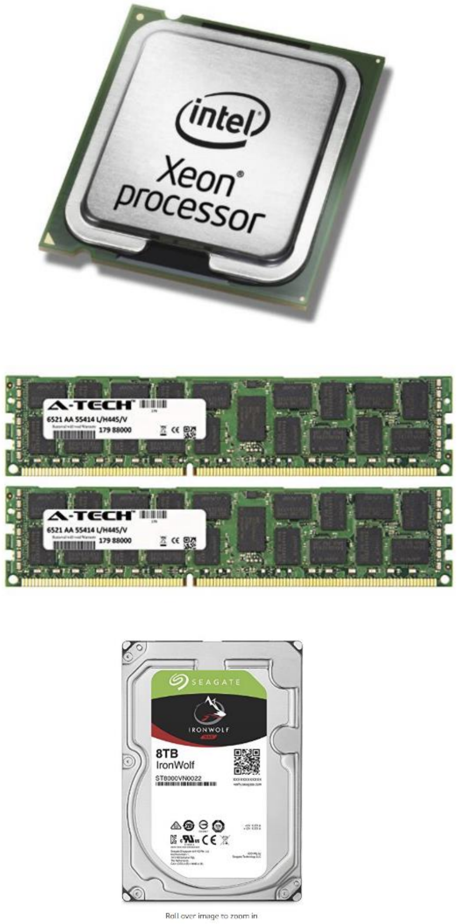
Jupyter notebooks are widely used



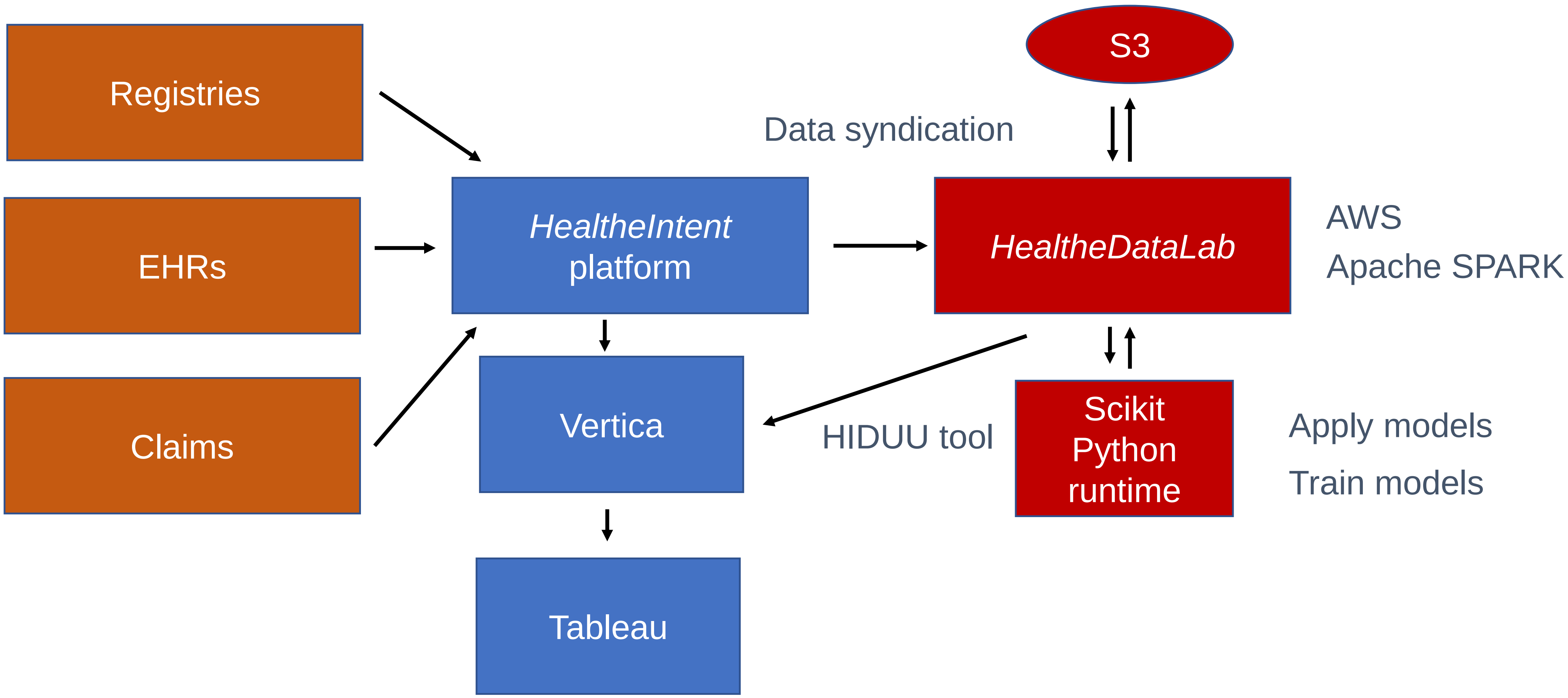
Apache SPARK







Data flow


Registries
EHRs
Claims
HealtheIntent platform
Vertica
Tableau
Data syndication
S3
HealtheDataLab
AWS
Apache SPARK
HIDUU tool
Scikit Python runtime
Apply models
Train models

Hemoglobin A1C registry measure prediction

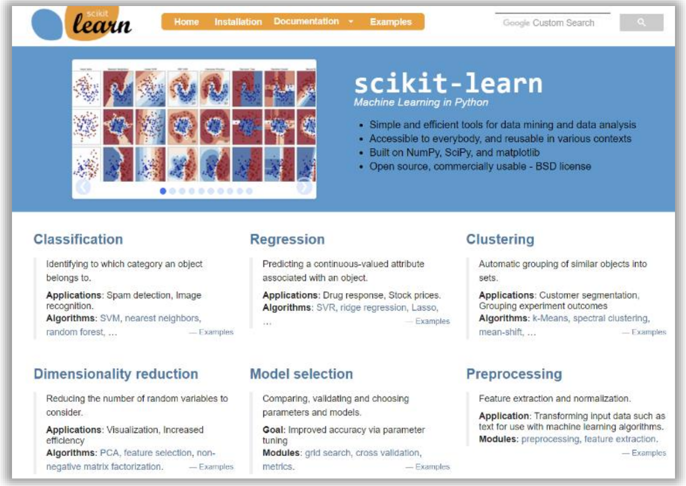
Generate features with Apache Spark

```
newRow = pd.DataFrame({'empi_id':prev_empi_id,
                       'serviceDate_first': first_srvdate,
                       'serviceDate_last': last_srvdate,
                       'serviceDate_diffdays': (last_srvdate - first_srvdate).days,
                       'A1C_first': first_A1C,
                       'A1C_last': last_A1C,
                       'non_diabetic_first': 1 if first_A1C < 6.5 else 0,
                       'diabetic_controlled_first': 1 if first_A1C > 6.5 and first_A1C < 9.0 else 0,
                       'diabetic_uncontrolled_first': 1 if first_A1C > 9.0 else 0,
                       'non_diabetic_last': 1 if last_A1C < 6.5 else 0,
                       'diabetic_controlled_last': 1 if last_A1C > 6.5 and last_A1C < 9.0 else 0,
                       'diabetic_uncontrolled_last': 1 if last_A1C > 9.0 else 0,
                       'gender': gender,
                       'ethinicity': ethinicity,
                       'dob': dob,
                       'age': age,
                       'condition_id_disp': condition_id_disp}, index=[0])
```

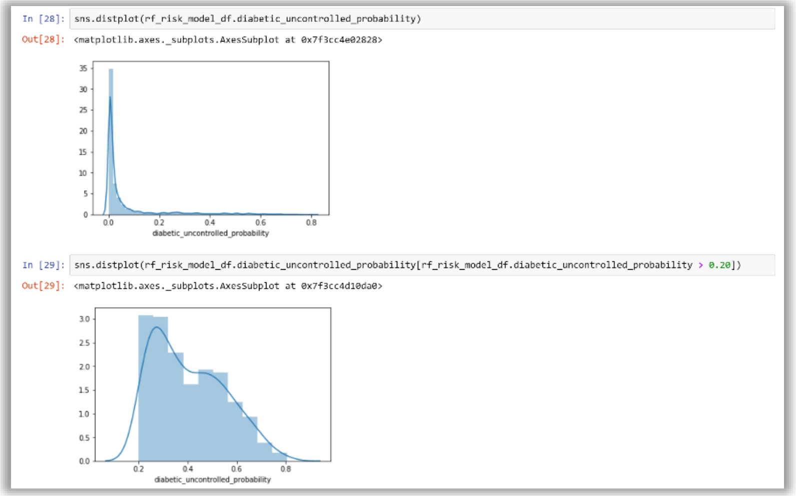
Test model performance

Sl.No.	Model	Parameters	Training AUC-ROC	Testing AUC-ROC
1	DTC Decision Tree Classifier	"StandardScaler(with_std=True), PCA(n_components=4), DecisionTreeClassifier(random_state=42,min_samples_split=9)"	0.9066	0.634
2	RF RandomForest Classifier	StandardScaler(with_std=True), PCA(n_components=4), RandomForestClassifier(random_state=42, max_depth=9, n_estimators=84, min_samples_split=9, criterion='entropy')	0.7294	0.6176
3	AB AdaBoost Classifier	StandardScaler(with_std=True), PCA(n_components=4), AdaBoostClassifier(RandomForestClassifier(max_depth=7, n_estimators=94, min_samples_split=6, criterion='entropy'),algorithm="SAMME.R",n_estimators=200, random_state=42,learning_rate=0.001)	0.7019	0.6057

Fit a model to the data



Use model to make risk predict predictions



Current projects in progress

- Development of inpatient diabetic scoring criteria for at-risk patients
  - Moved prototyped algorithm in Python to *HealtheDataLab*
- Aortic aneurysm screening
  - Determine at risk population and associated documents to extract
  - Working with Cerner engineering on making radiology notes accessible in *HealtheDataLab*
- Development of predictive models for coded diagnosis
  - Data normalization and processing medications, labs, and vitals
  - Future model development and testing with *Cerner Real-World Data*™

Q&A

Cole Erdmann
Director, Clinical Intelligence, Cerner Corporation

Navneet Srivastava
Senior Solutions Architect, Amazon Web Services

Dr. William Feaster, MD, MBA
Vice President, Chief Health Information Officer at CHOC Children's Hospital

Janos Hajagos, Ph.D.
Chief of Data Analytics, Stony Brook Medicine

eHEALTH INITIATIVE

Thank you for participating in today’s webinar!

We would like to thank Cerner for supporting eH’s educational initiatives!