Baikal – Implementing and Deploying Clinical Models with a Real-Time Data Lake

Wade L. Schulz, MD, PhD
Assistant Professor, Yale School of Medicine
Director of Informatics, Department of Laboratory Medicine

Yale SCHOOL OF MEDICINE

Center for Computational Health

Competing Interests

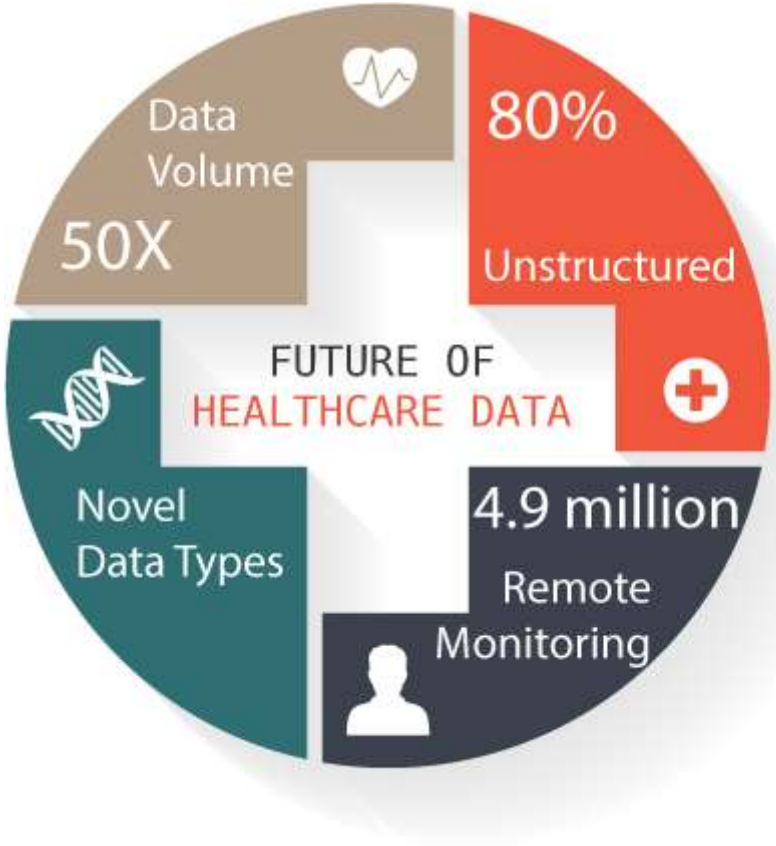
- Consultant, Hugo Health
- Consultant, National Center for Cardiovascular Disease, Beijing, China

Outline

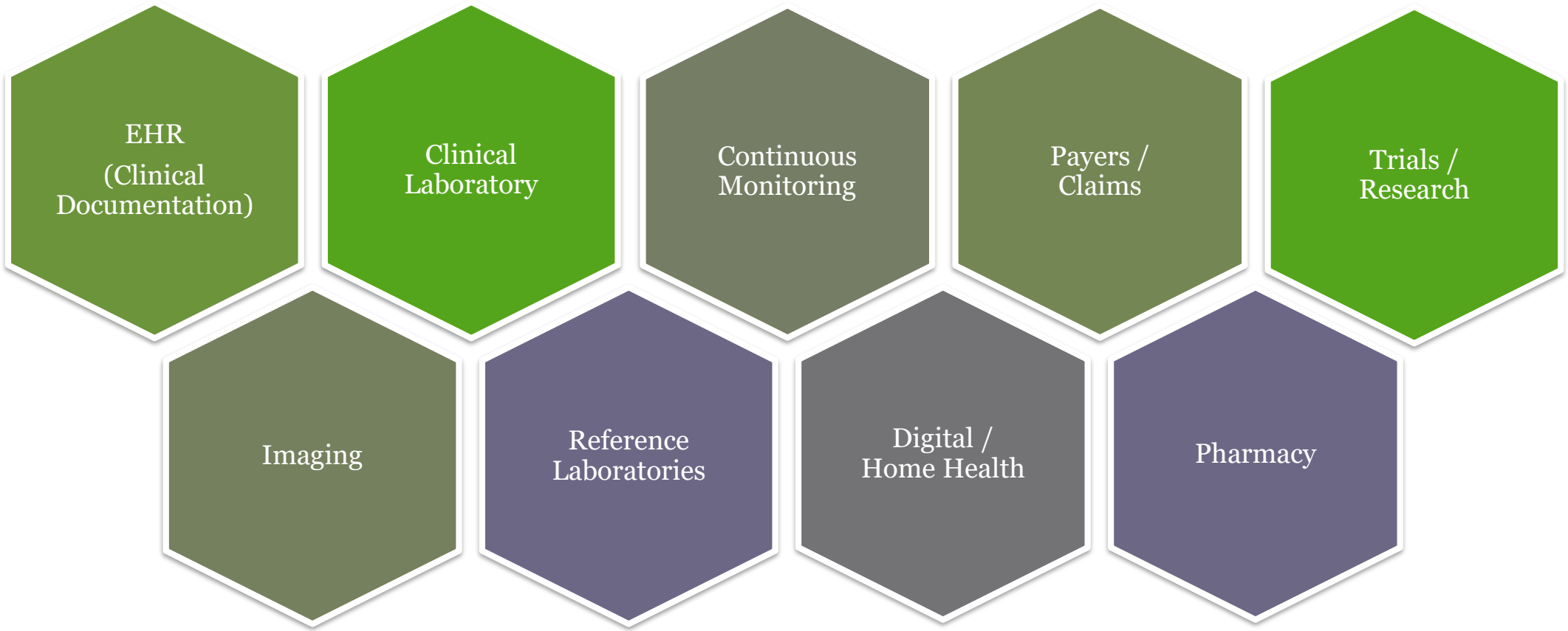
- Data in Healthcare: Sources, Volume, and Quality
- Developing Clinical Predictive Models
- Implementing an Integrated Clinical Data Lake
- Creating and Infrastructure for Real-Time Clinical Analytics
- The Future of Cognitive Support in Medicine

Improving Healthcare with Data

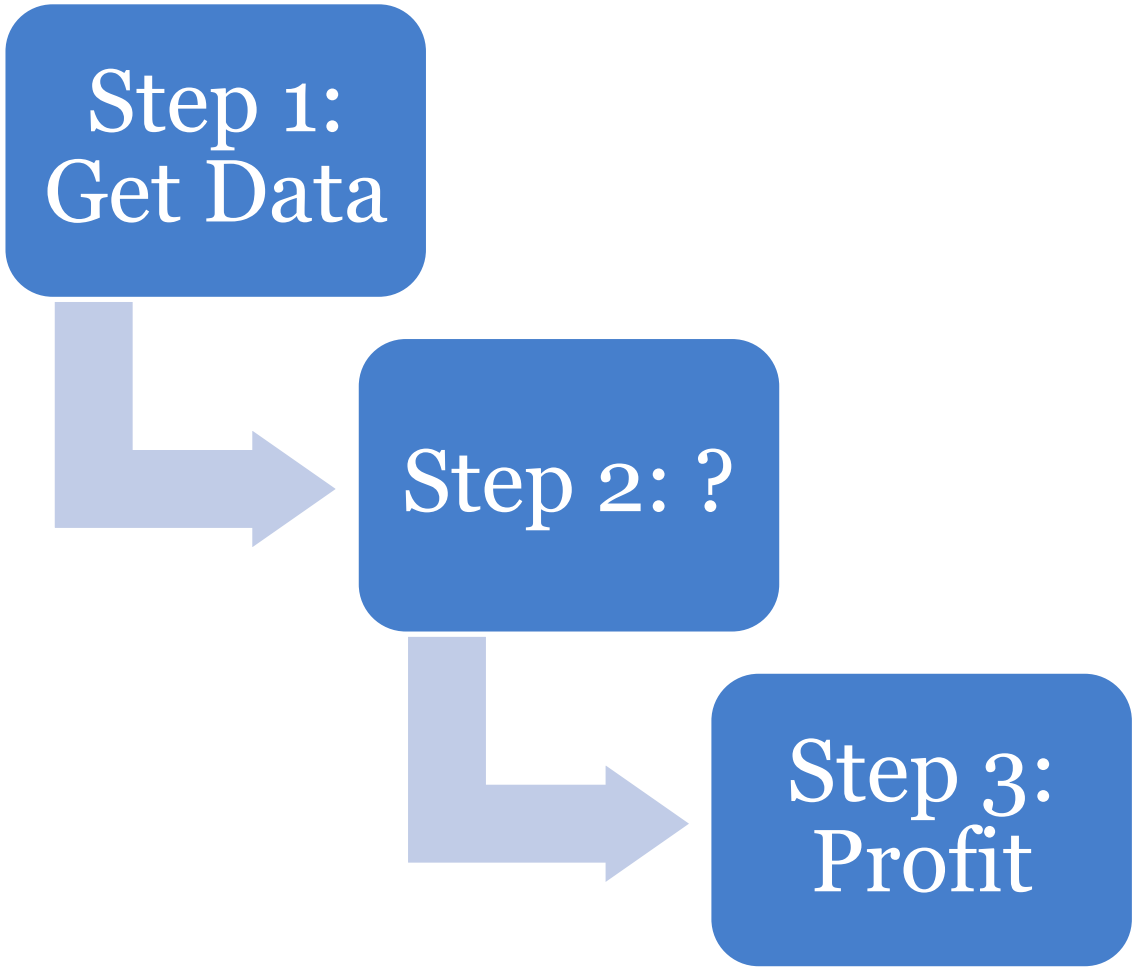
Healthcare Data



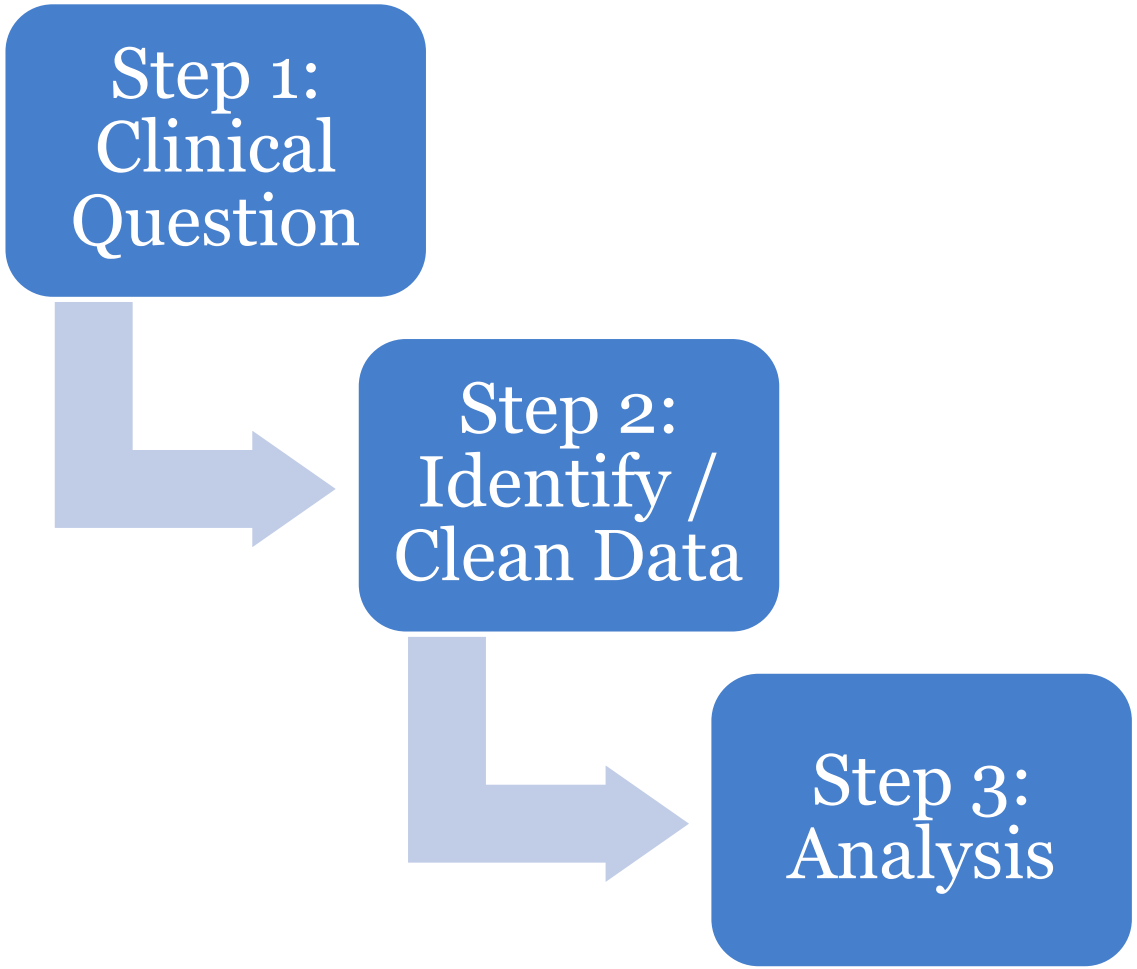
Healthcare Data Sources



Data-Driven Analytic Approach



Problem-Centric Approach



The State of Clinical Predictive Modeling

- Significant attention and interest, with the hope of increased efficiency of healthcare delivery with improved outcomes.

However…

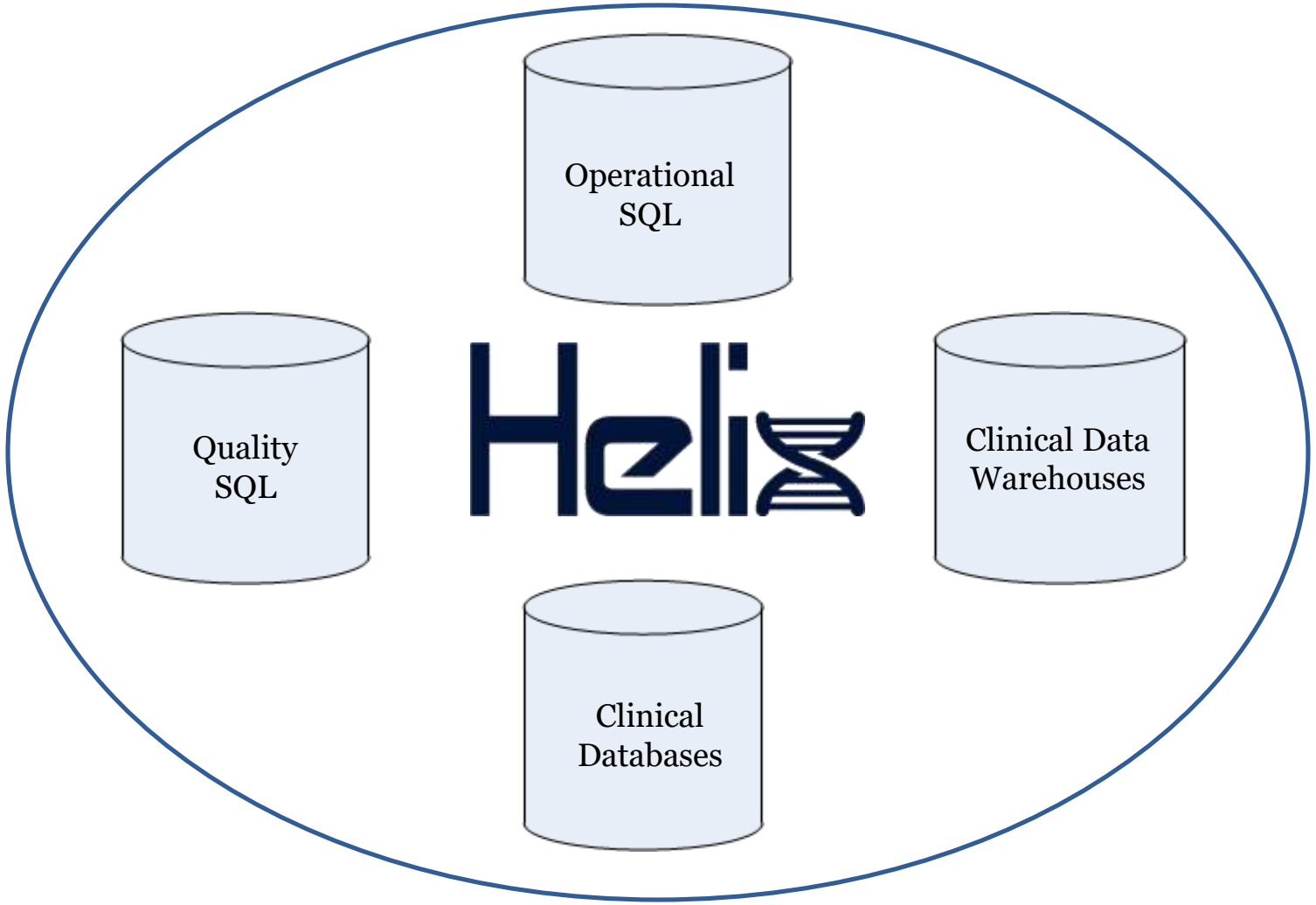
- "Forecasting-based interventions might have uncertain effectiveness, focus on cost savings rather than long-term health, or specifically exclude disadvantaged populations."
- "There is little evidence regarding how or whether [health] forecasting improves healthcare value. This is due to both the modest level of research and what is termed the "impactibility" problem. That is, even if prediction algorithms accurately identify at-risk patients, intervening to achieve desired outcomes is often inhibited by limitations of current disease management approaches or the general state of medical science."

J. Frank Wharam and Jonathan P. Weiner, "The Promise and Peril of Healthcare Forecasting." Am J Manag Care. 2012;18(3):e82-e85)

Moving the Needle for High-Quality Models

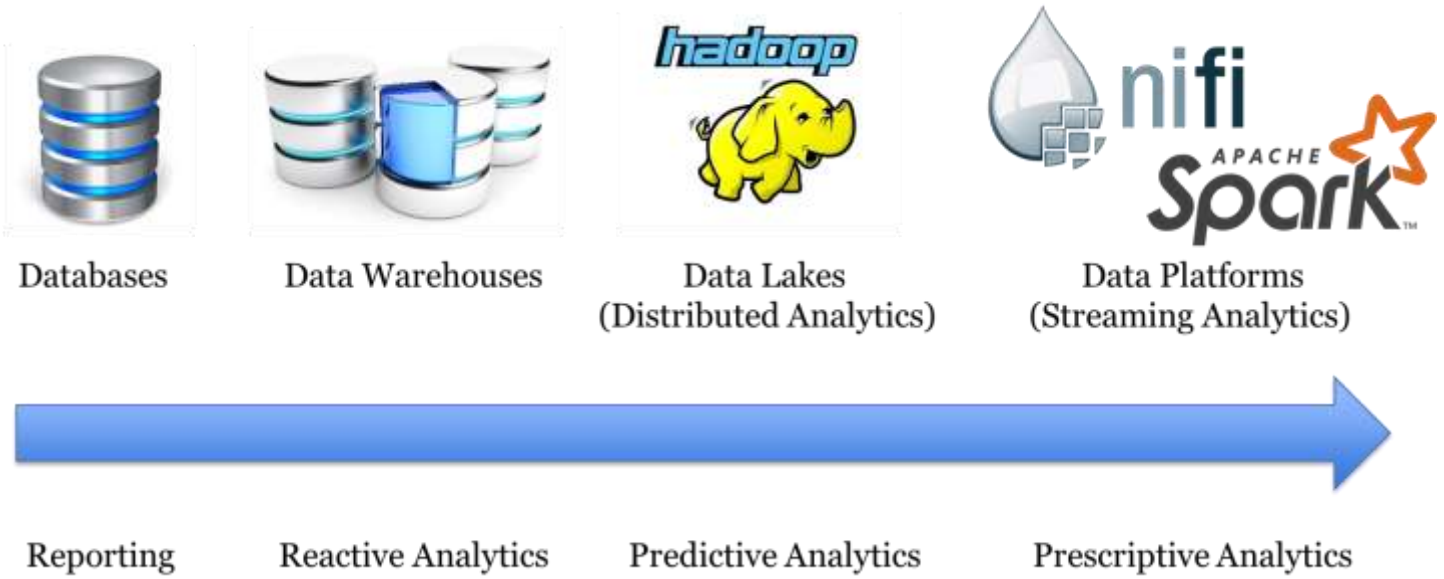
- Clinically impactful questions
- High-quality data
  - Validation of data source / extraction
  - Documentation and validation of data pre-processing
- Experimental approach to model development
  - Quality control of methods and code
- Clinical validation of model on naïve data set
- Ongoing performance validation of deployed models in each clinical setting

Delivering High-Quality Data for Analytics

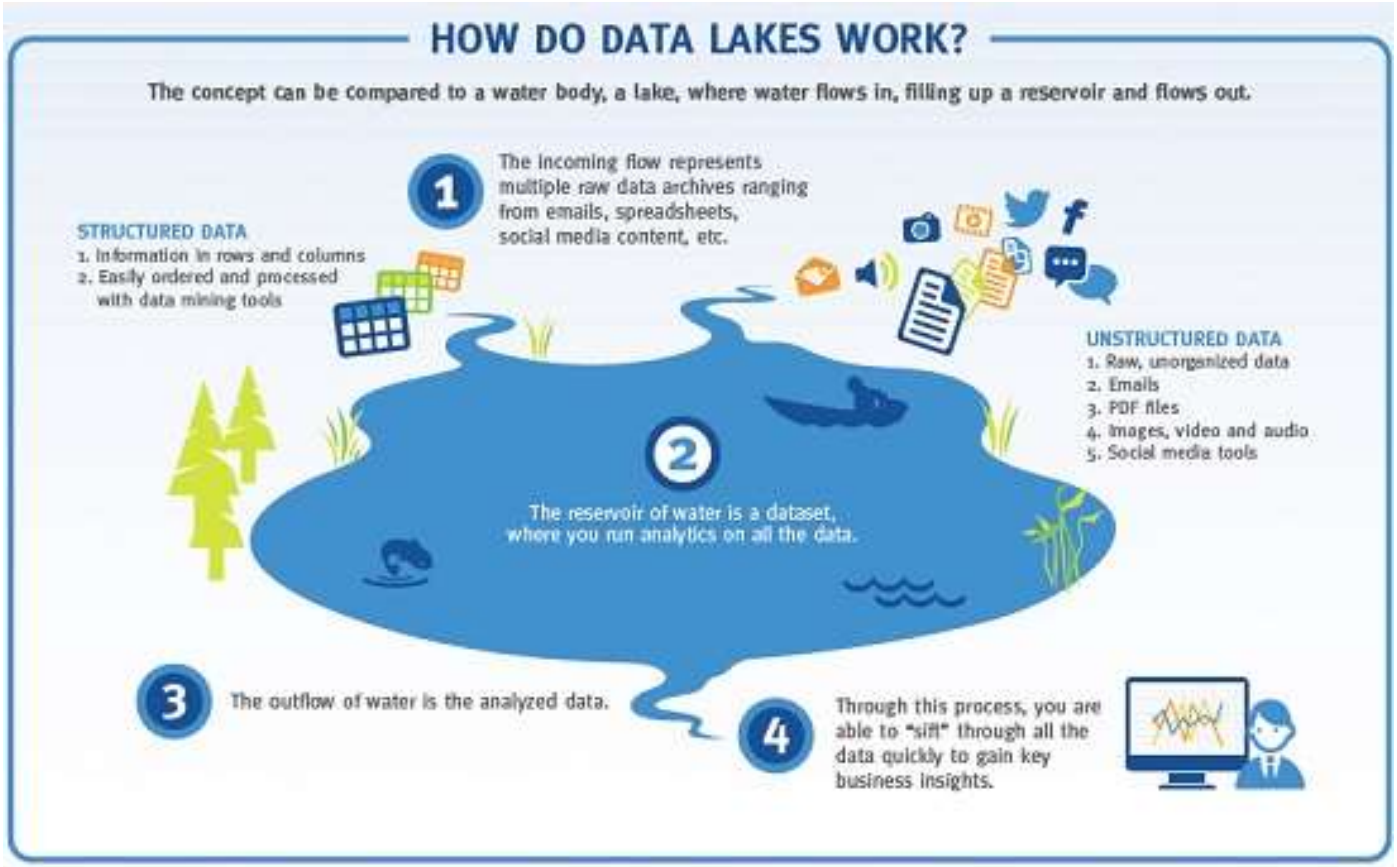


The “Death” of ***?

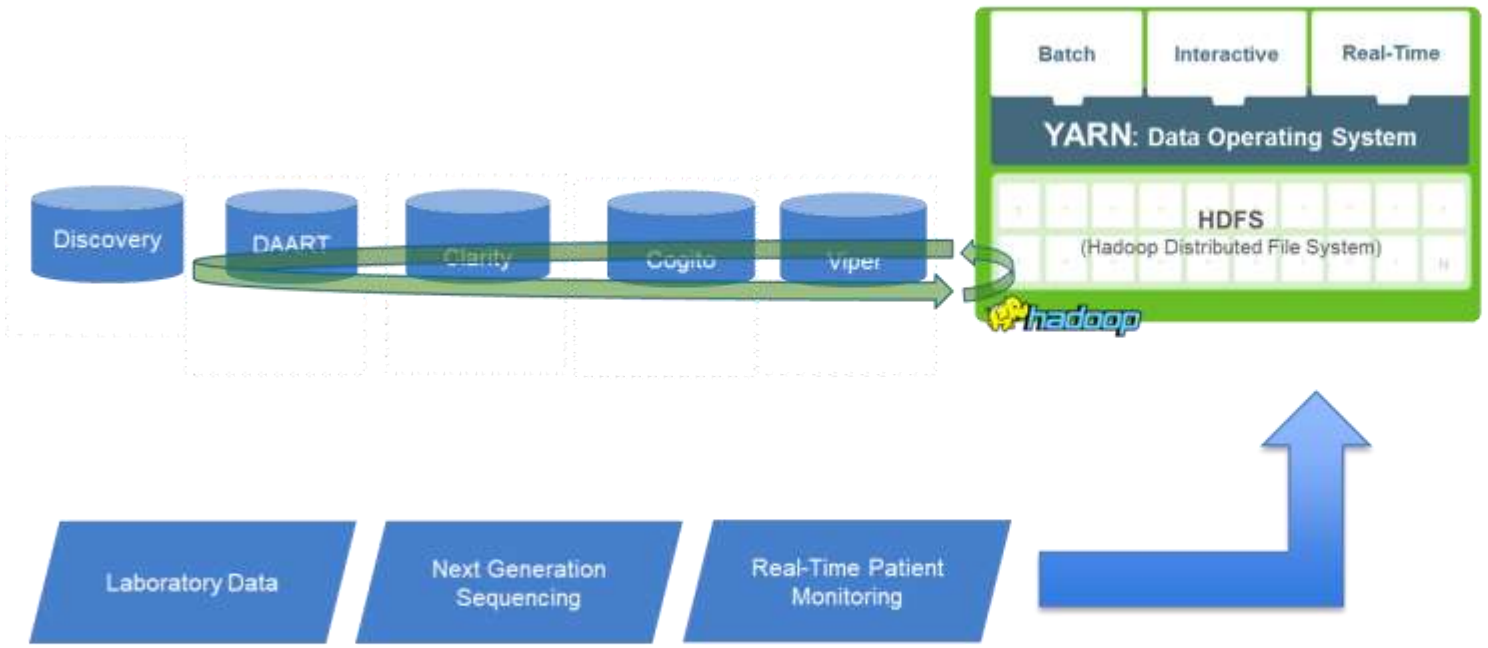
- Software for a data analysis / data science platform:
  - Distributed storage
  - Distributed analytics
  - Data governance / security
  - Development and visualization tools



Migrating to a Data Lake

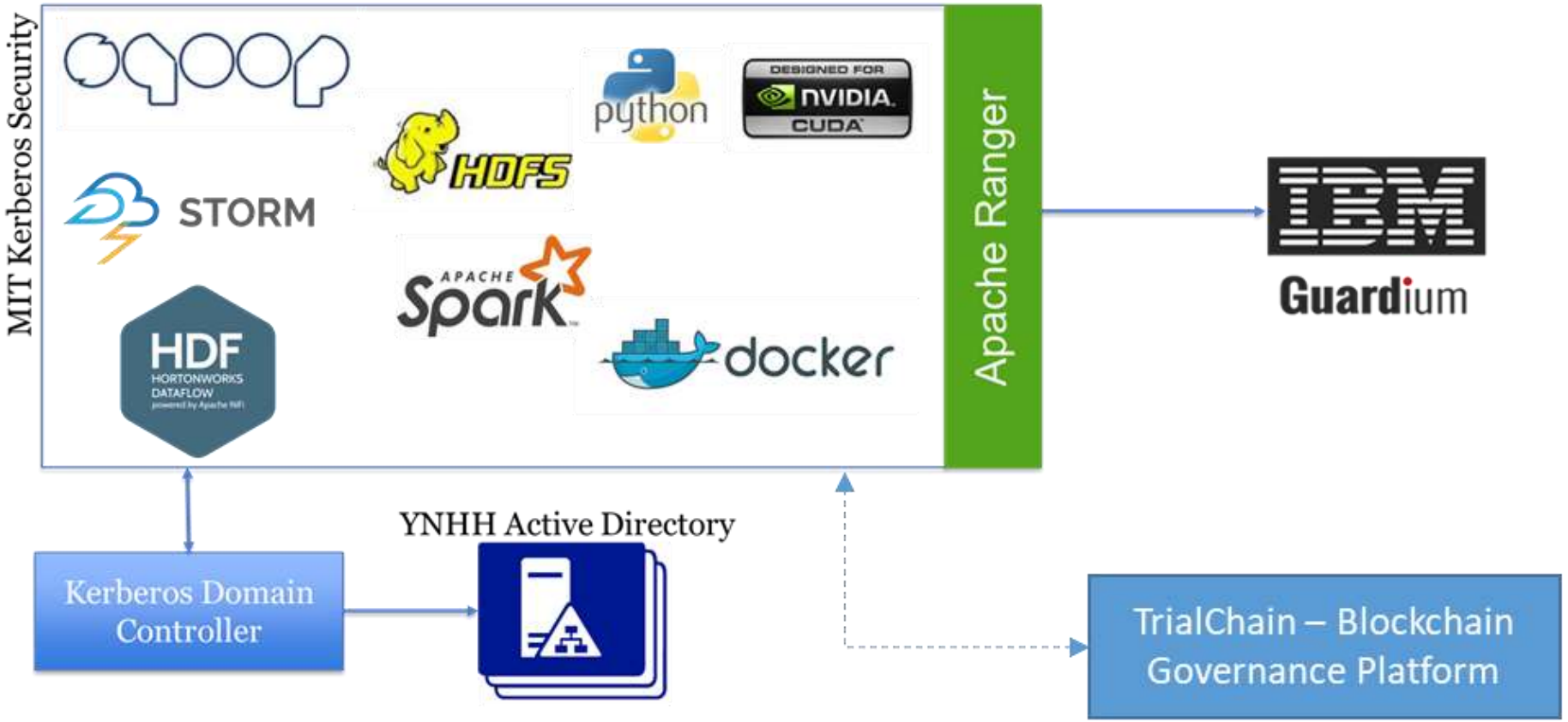


Baikal: The YNHH Data Science Platform



The Data Science Toolkit

Governance and Data Validation



Blockchain-Based Data Validation

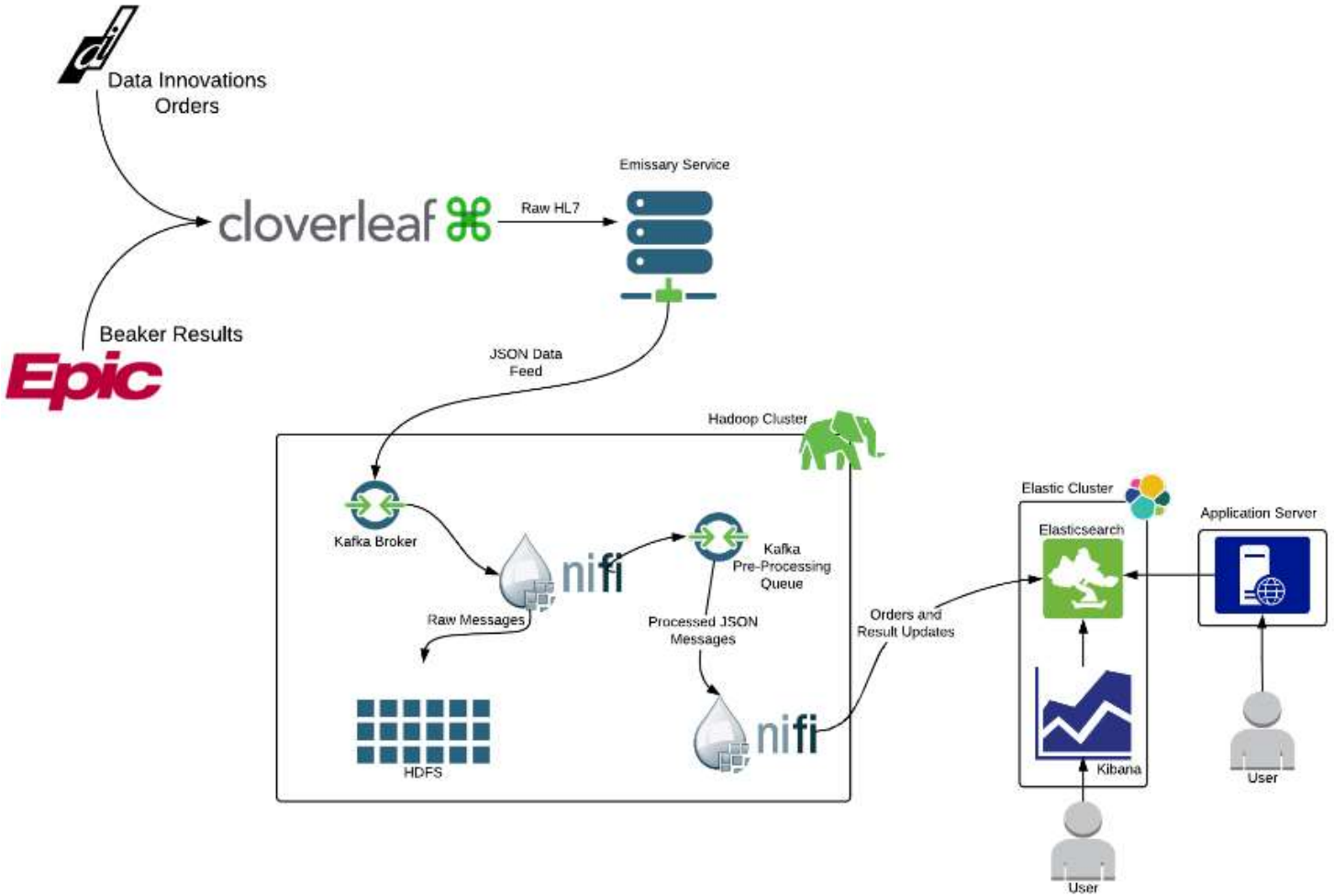
TrialChain: A Blockchain-Based Platform to Validate Data Integrity in Large, Biomedical Research Studies

Hao Dai[1†], H Patrick Young PhD[2,3†], Thomas JS Durant MD[2,4], Guannan Gong MS[5], Mingming Kang PhD[1], Harlan M Krumholz MD SM[2,3,5], Wade L Schulz MD PhD[2,4Δ], Lixin Jiang MD PhD[1Δ*]

[1]National Center for Cardiovascular Disease, Fuwai Hospital, Beijing, China; [2]Yale New Haven Hospital, Center for Outcomes Research and Evaluation, New Haven, CT; [3]Yale University School of Medicine, Department of Internal Medicine, Cardiology, New Haven, CT; [4]Yale University School of Medicine, Department of Laboratory Medicine, New Haven, CT; [5]Yale University School of Public Health, Department of Health Policy and Management, New Haven, CT

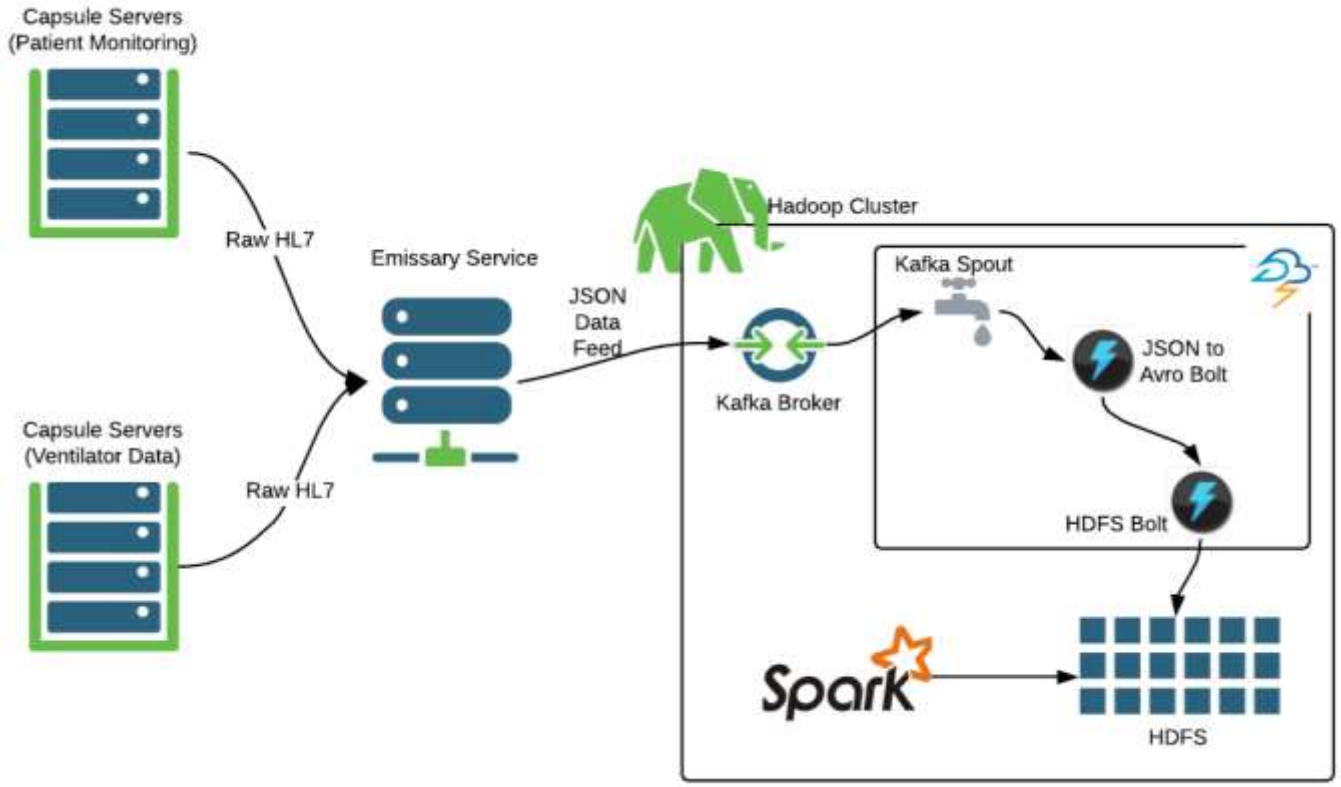
https://arxiv.org/abs/1807.03662

Nucleus: Real-Time Laboratory Business Intelligence



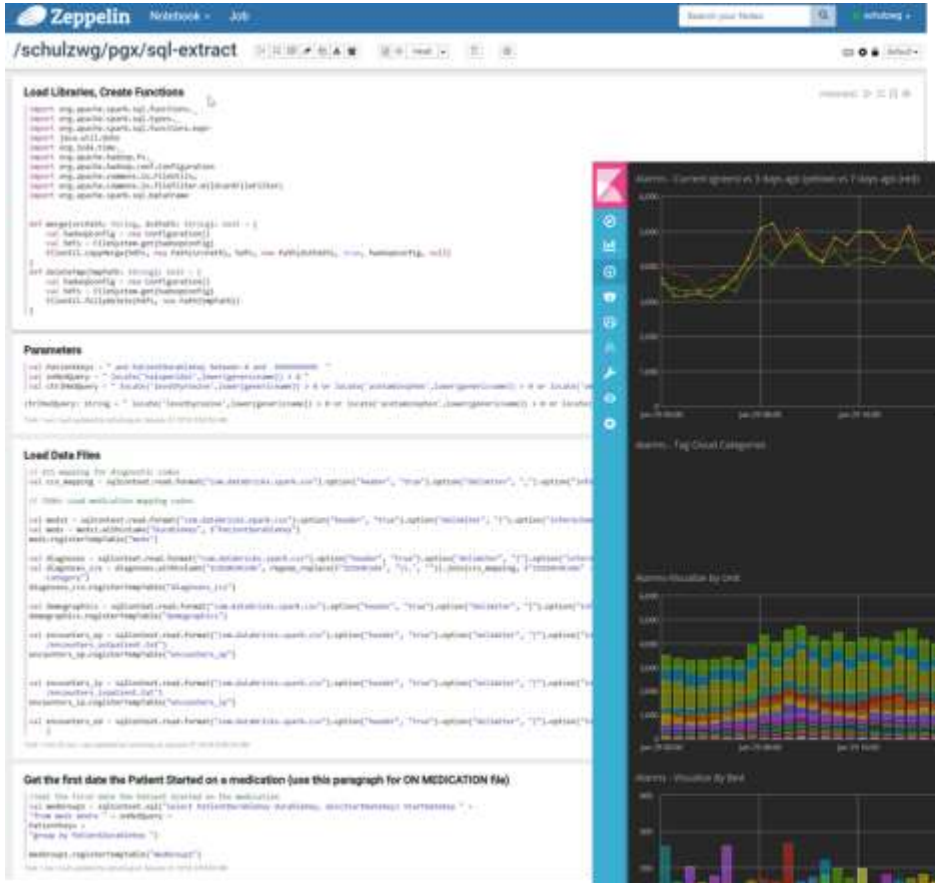
- Archive of electronic laboratory data dating to 1972
- Real-time stream of laboratory orders and results for use in predictive modeling, laboratory business intelligence, and quality control

Electron: Real-Time Physiologic Monitoring Platform

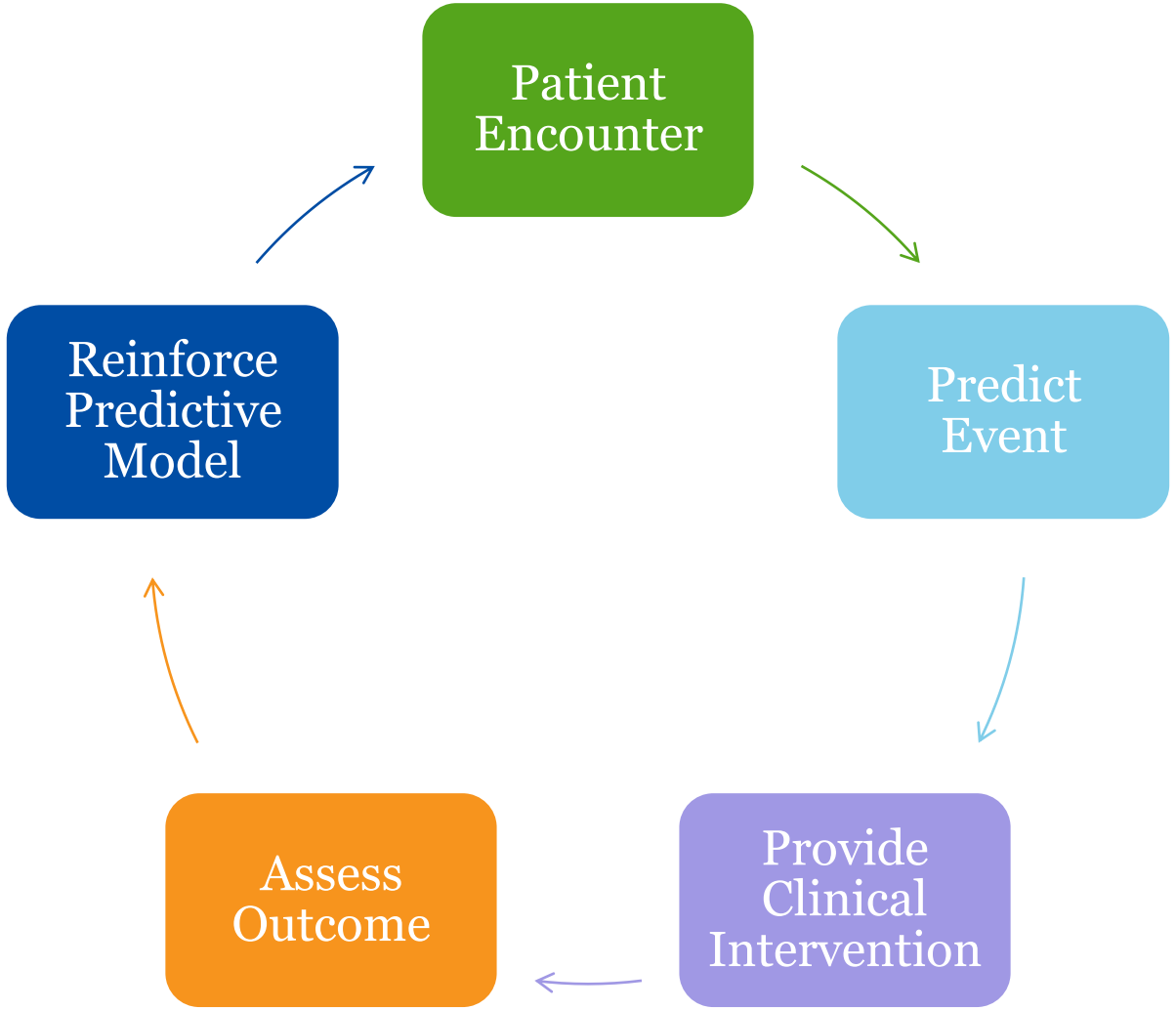


- Capture real-time data from patient monitoring devices throughout health system
- Processes >15 billion data points per month
- Actively used for several ongoing research and clinical redesign projects

Visualizing Big Data



Building a Platform for Real-Time Clinical Analytics

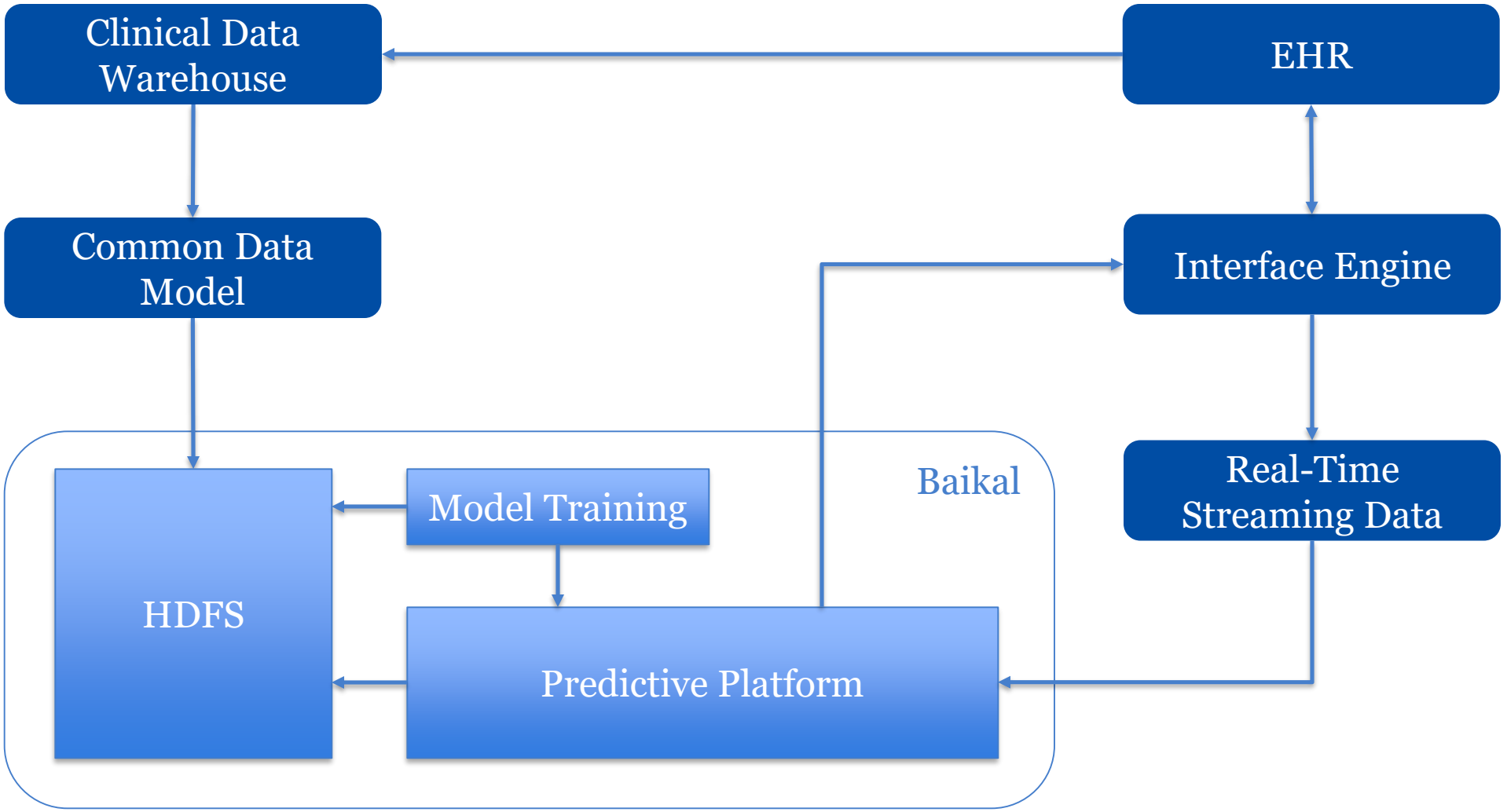


Clinical Use Case: Pharmacogenomic Screening

- Pharmacogenomics has gained renewed attention with push for precision medicine
  - And ability to cheaply test many genetic targets
- Conceptual benefit for decreased cost and adverse events with improved efficacy
- Unclear practical / clinical impact with varied clinical trial results (especially related to cost)

- Key clinical question: Who to screen?

Developing a Clinical Modeling Pipeline

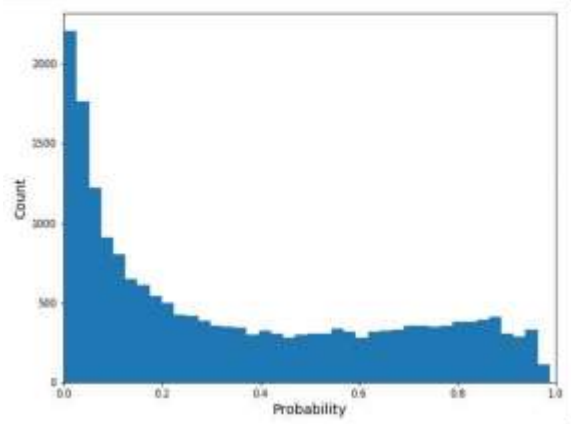


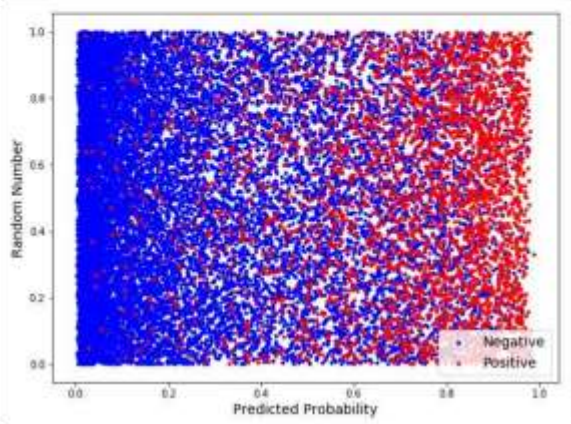
Pharmacogenomics – Developing a Profile

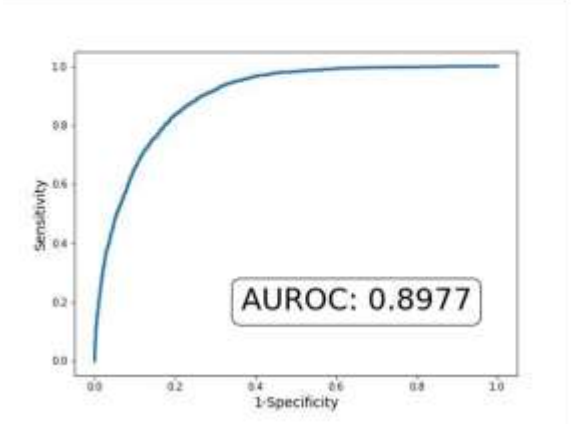
- Retrospective electronic health record data for patients ages 18-89 who were prescribed warfarin or haloperidol between 1-1-2013 and 10-30-2017.
- Demographics, prescribed medications, encounter types, and diagnoses, consolidated into Clinical Classifications Software (CCS) codes, present at least 30 days prior to the medication being prescribed were extracted.
- Encounter types and other medications were summarized by occurrences in periods of 30, 120, 365, and >365 days prior to target medication start.
- A control population was selected from patients on neither warfarin or haloperidol over the same period.

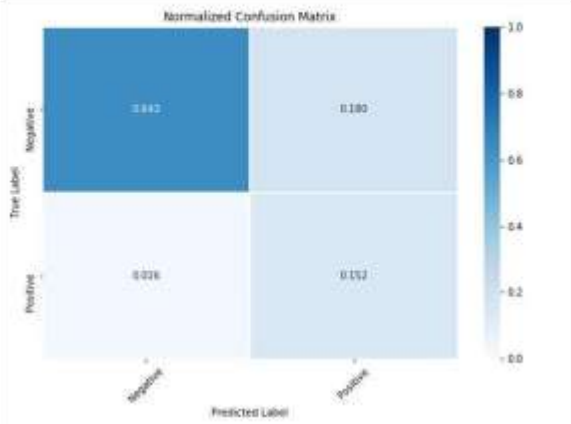
Predicting Warfarin Use

Warfarin XGBoost





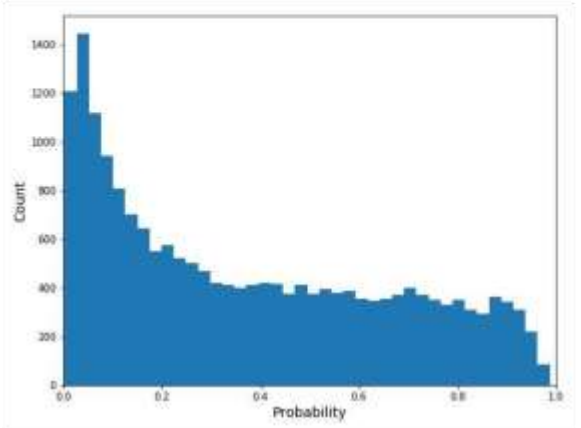


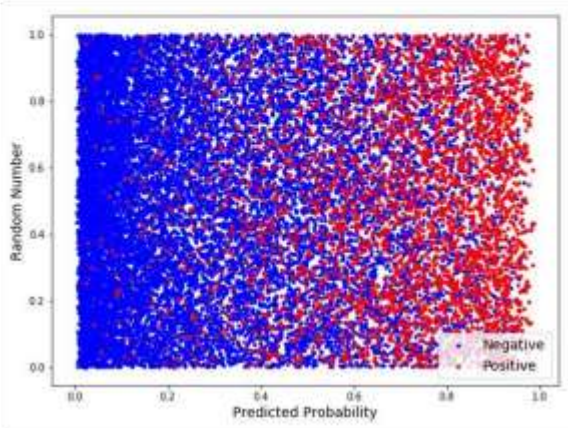


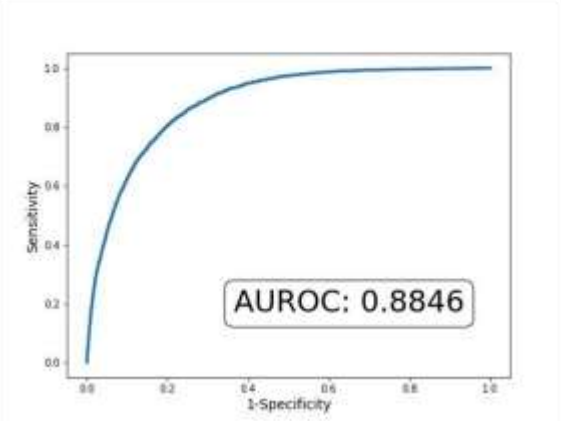
Warfarin
- Train: 14,043
  - Train pos: 6,996
  - Train neg: 7,047
- Test: 19,545
  - Test pos: 3,484
  - Test neg: 16,061

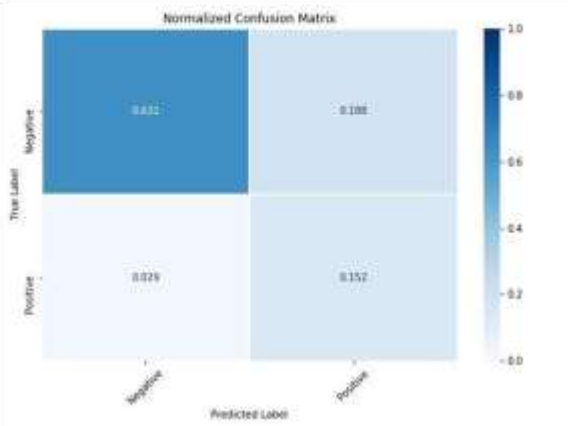
Predicting Warfarin Use

Haloperidol XGBoost









Haloperidol
- Train: 14,305
  - Train pos: 7,153
  - Train neg: 7,152
- Test: 19,479
  - Test pos: 3,523
  - Test neg: 15,956

Next Steps

- Predictive model with good analytic performance
- Beginning prospective assessment of model quality with clinical data feeds
- Finalization of clinical modeling platform for production deployment

Conclusions

- A robust data platform and data modeling / architecture are needed for the development of quality clinical models
  - Identifying reusable tools and workflows is key for training and ongoing development
- Combining multiple toolsets for specific tasks can create a ‘best of both’ infrastructure while keeping within reasonable costs
- Clinical model development is complex, with potential high rewards, but need for ongoing evaluation
  - Treating predictive model development like a clinical trial or development of a laboratory test can provide guidelines for data and model validation and quality control

Questions?

Baikal – Implementing and Deploying Clinical Models with a Real-Time Data Lake

Wade L. Schulz, MD, PhD
Assistant Professor, Yale School of Medicine
Director of Informatics, Department of Laboratory Medicine

Yale SCHOOL OF MEDICINE

Center for Computational Health