HEALTHCARE DOES HADOOP: AN ACADEMIC MEDICAL CENTER’S FIVE-YEAR JOURNEY

Charles Boicey, MS, RN-BC, CPHIMS
Chief Innovation Officer
Clearsense

UC Irvine Health
Stony Brook Medicine
The doctor of the future will give no medicine, but instead will interest his patients in the care of human frame, in diet, and in the cause and prevention of disease.

Thomas Edison (1847 – 1931)
PHR Centric Health

Modern HDP

HIE

EMR
NowTrending 2012
Current Environment

Electronic Medical Record
• Not designed to process high volume/velocity data
• Not intended to handle complex operations
  • Such as:
    • Anomaly detection
    • Machine learning
    • Building complex algorithms
    • Pattern set recognition

Enterprise Data Warehouse
• Suffer from a latency factor of up to 24 hours
• The EDW serves all of the following retrospectively as opposed to in real time
  • Clinicians
  • Operations
  • Quality and research
Big Data = Interoperability

- Big Data Ecosystem that Supports:
  - Hadoop (HDFS)
  - Hbase
  - Hive
  - Pig
  - MapReduce
  - Mahout
  - MongoDB (NoSQL)

- Neo 4j (Graph Database)
- Relational Data Base
- R
- Spark
- Storm
- Weka
Big Data = Complete Data

• The Electronic Medical Record is primarily transactional taking feeds from source systems via an interface engine

• The Enterprise Data Warehouse is a collection of data from the EMR and various source systems in the enterprise

• In both cases decisions are made concerning data acquisition

• A Big Data system is capable of ingesting and storing healthcare data in total and in real time
Modern Healthcare Data Platform

A healthcare information ecosystem built on “Big Data” technologies should:

- Be capable of serving the needs of clinicians, operations, quality and research
- And should do so in real time and in one environment

**Should be:**
- Able to ingest all healthcare generated data both internal and external in native format

**Should be:**
- A platform for advanced analytics such as early detection of sepsis & hospital acquired conditions
- Be enabled to predict potential readmissions
- Leverage complex algorithms and be a machine learning platform
Architecture Guiding Principles

- Architecture to minimize encumbrance on IT staff
- Ability to store all healthcare date in native form and complete
- Use of supported open source code
- Ensure architectural compatibility with commercial applications
Infrastructure

• Low Cost of Entry & Scalable
  • Open Source
  • Commodity Hardware
    • UCI Hadoop Ecosystem
      • 10 nodes
      • 5 terabytes
    • Yahoo Hadoop Ecosystem
      • 60K nodes
      • 160 petabytes
• Cloud Ready
Data Sources

- Legacy Systems
  - Print to Text or Delimited String
- All HL7 Feeds (EMR source systems)
- All EMR Initiated Data (Stored Procedures)
- Device Data (in one minute intervals)
- Physiological Monitors (HL7)
- Ventilators (HL7)

- Smart Pumps
- Social Media (POC)
- Healthcare Organization Sentiment Analysis
- Patient Engagement
- Home Monitoring (POC)
- Real Time Location System (RFID)
- Hospital Sensors
Newer Data Sources

• External Streaming Device Data
• Wearables
• Home Devices
• Social Media
• Geographic Information System (GIS) Data
• Omic Data

• Open Data
  • www.data.gov
• Adverse Drug Event
  • www.researchae.com
• Internet of Things (IoT)
  • Telematics
  • 5G
Use Cases

- Legacy System Retirement
- **Patient Condition Changes**
  - RRT
- **Early Sepsis Detection**
- **Environmental Response**
- Real Time Nursing Unit Utilization
  - Staffing and Resource Allocation
- Social Media Sentiment Analysis
- Research

- Cohort Discovery
- Data Science
- Clinician Aware Applications
- Patient Monitoring External to Traditional Healthcare Setting
- **Event Driven Care & Real Time Quality Monitoring**
- **Personal Health Record**
Future Use Cases

- Ventilator Management
  - Vent dashboard in EMR
- Hospital Acquired Infections (HAI)
- VTE Surveillance
- Sensium Vitals Digital Patch
- Patient-Generated Data
  - Home Devices (Scale, Vital Signs, Glucose)
  - Exercise & Diet (Fit Bit, Jawbone, Nike)
- Combining Phenotype Data with Genotype Data
- Patient Threat Analysis
- Edge and Vertices Analysis
  - Patient caregivers and outcomes
Imaging Analytics

- NIH Funded U24 Grant
- Joel Saltz, PhD

This project is to develop, deploy, and disseminate a suite of open source tools and integrated informatics platform that will facilitate multi-scale, correlative analyses of high resolution whole slide tissue image data, spatially mapped genetics and molecular data for cancer research.
Patient Persona

- Surveys
- Questionnaires
- Clinic Notes
- External Sources
  - IoT
  - Social Media
  - Credit
  - Telemetrics
Hadoop for Healthcare

ANALYSIS
- Cohort discovery
- Predicting readmission
- Detection of sepsis pathways
- Analyzing test variances
- Rapid bedside response

Tracking patient wait times
Home health monitoring
Chronic disease management
Patient scorecards

DATA REPOSITORIES
- EDW
- Surgical Data Mart
- Diagnosis Data Mart
- Quality Data Mart
- Clinical Info Data Mart
- Neo4j

Governance & Integration
- Batch
- Script
- SQL
- NoSQL
- Stream
- Search
- Others

YARN: Data Operating System

HDFS
(Hadoop Distributed File System)

TRADITIONAL SOURCES
- LEGACY EMR
- FINANCIAL
- RADIOLOGY
- PHARMACY POS
- PACS
- RTLS
- CLINICAL TRIALS
- TRANSCRIPTIONS

EMERGING & NON-TRADITIONAL SOURCES
- SOCIAL MEDIA
- MEDICATION
- HOME DEVICES
- DEVICE INTEGRATION
- LABORATORY
- BIO REPOSITORY
- GENOMICS
- QUANTIFIED SELF
FOSS Driven Protean

- Is a centrally-hosted, instrumented “Smart and Connected” platform servicing real time business event streams using high-speed MPP Compute and Storage Grids
- Primarily based on the concepts and principles of Event Driven Architecture (EDA), Complex Event Processing (CEP) and Multi-Agent-Systems (MAS)
- Support for high speed data ingestion - Structured and Unstructured (Textual)
- Core Advanced Analytics enabled through Model Building, Data Mining and Machine Learning techniques (Supervised and Unsupervised)
- Context modeling creation across Time-Space-Value dimensions
- Enables creation of a Central Enterprise Data Refinery to enable “Source of Truth” for transactional information within the Healthcare Enterprise
A reference architecture blueprint for realizing the Big Data platform leveraging Free and Open Source Software (FOSS)....The platform has been deployed successfully across 4 large client implementations across various business domains....
FHIR – The “Public API” for Healthcare?

FHIR = Fast Health Interoperability Resource
- Emerging HL7 Standard (DSTU 2 soon)
- More powerful & less complex than HL7 V3

ReSTful API
- ReST = Representational State Transfer – basis for Internet Scale
- Resource-oriented rather than Remote Procedure Call (nouns > verbs)
- Easy for developers to understand and use

FHIR Resources
- Well-defined, simple snippets of data that capture core clinical entities
- Build on top of existing HL7 data types
- Resources are the “objects” in a network of URI reference links

Huff, S., McCallie, D HIMSS 2015
SMART Platform – Open Specification for Apps

• “Substitutable Medical Apps”
  • Kohane/Mandl – NEJM (2009)

• A SMART App is a Web App
  • HTML5 + JavaScript
  • Remote or embedded in EHR
  • URL passes context & FHIR link

• EHR Data Access via FHIR

• OAuth2 / OIDC for security

Huff, S., McCallie, D HIMSS 2015
Some SMART Hotbeds

Huff, S., McCallie, D HIMSS 2015
Boston Childrens: SMART Growth Chart

Huff, S., McCallie, D HIMSS 2015
DSRIP

- 8 billion dollar grant (Medicaid waiver) from CMS to NY State
  - 25% reduction over five years in avoidable hospitalizations and ER visits in the Medicaid and uninsured population
  - Collaborative effort to implement innovative projects focused on
    - System transformation
    - Clinical improvement
    - Population health improvement
5 Year Goals

• Create integrated Suffolk County care delivery system for 387K lives anchored by safety net providers
• Engage partners across the care delivery spectrum to create a countywide network of care
• After five years, transition this network to an ACO which will contract with insurance providers on an at risk basis
The Future of Genomic Medicine

Gavin Stone, edico genome 5G Summit May, 14, 2015
New Team Members

- Data Scientist
- Developers
- Cognitive and Behavioral Psychology
- User Experience
- Human & Computer Interaction
  - Devices
  - Wearables
- Patients & Family
Trends: Big Data

- **Definition:** Evolving
- **Creation & Management:** Distributed and augmented
- **Information Governance:** Shared
- **Meaningful Analysis:** Beyond PnL, Reporting, Connections, Correlations, Pattern Recognition, Machine Learning, Natural Language Processing
- **Business Requirements:** Blank Page; We don’t know what we want we will figure it out once we look at the data, the data will lead the way, AKA, Data Science
Trends: Healthcare

- Content Analytics – Suggestive Analytics* – Prescriptive Analytics
- Imaging Analytics
- Moving Analytics out of the EMR Environment
- Graph Data Mart
  - Edge and Vertices Analysis
- Omic & Phenotype Combines
- Sentiment Analysis

Dale Sanders
Takeaways

• Underpinning platforms may change but concept is here to stay, abstract where possible.

• Machine learning will lead to the evolution of Data Science and eventual use of AI in Healthcare.

• Get used to source now, ask questions later: Healthcare evolves with data and it is not a point in time construct any longer.

• Get used to working with constant change, disruptive trends and something new that will make your “frameworks” obsolete.
Contact Me @

Charles Boicey

cboicey@uci.edu
charles.boicey@stonybrookmedicine.edu
cboicey@clearsense.com

1+904-373-0831

@N2InformaticsRN