



rchn
community health foundation

Path2Analytics Project Introduction

David Hartzband, D.Sc.

Director, Technology Research
RCHN Community Health Foundation

May 2014

Path2Analytics: What is It?

- **Path2Analytics is a project developed by the RCHN Community Health Foundation with the goal of working with community health centers to integrate the capacity to productively use contemporary analytics capabilities into the work of the health center**
- **The project will work with both rural & urban health centers to meet this goal**
 - The first two centers have been selected & the project has started at both
 - The project will consist of both a process & a set of software to assist in meeting this goal

Path2Analytics: Assumptions 1

- **Much of current healthcare is increasingly an information management effort.**
- **Analytics (i.e. the systematic analysis of data focused on answering a specific question or set of questions) must be an integral part of a healthcare organization's administrative & clinical plans.**
- **Analytics is not an IT function, a software package or a technical methodology.**
 - **It is a way of thinking about an organization's goals that makes use of IT, analytic (software) packages & technical methodologies.**

Path2Analytics: Assumptions 2

- **Analytics is not necessarily about “Big Data”.**
 - But it must be about “right data”
 - It can be very effective making use of “little data”, that is the data that an organization already has.
 - Not all organizations have multiple petabytes of data, but even multiple gigabytes of data can produce productive results.
- **An analysis is only as good as:**
 - The quality of the data used, &
 - The quality of the questions asked.
- **Questions asked are only as good as their alignment with:**
 - The quality & content of the available data, &
 - The strategy & goals of the organization .

Path2Analytics: Process

- **The project consists of five units:**
 - Landscape - what are other healthcare organizations doing with respect to analytics? how does this relate to CHCs?
 - Analysis Planning - what resources are available? what questions to ask? how to address them analytically? what tools to use, setting goals & metrics
 - Data Organization - what data is available, extraction & aggregation, ETL & normalization
 - Analysis & Interpretation - operation of analytic process
 - Outcome & Outreach - how to organize & convey results, who to tell, communication process etc.

Path2Analytics: Landscape - 1

- **Data Warehousing**

- Almost all projects require some data extraction or warehousing

- **Point-of-Care Recommendations**

- **Mayo Clinic** - Large-scale warehousing: semantic normalization, data dictionary, 5M clinical records covering 15 years, 15-25PB of data (2500x the size of the content of the Library of Congress)
 - AWARE “bedside consulting”: delivering diagnosis & best-practice treatment for individual patients based on analysis of anonymized data set
- **BIDMC** - Clinical Query: system used directly by providers to analyze 2.2M clinical records to determine best-practice treatment in real-time
- **Kaiser** - Clinical, pharma & lab data on 9.1M patients over 10 years
 - Natural Language query system allows providers to get recommendations for best-practice treatment
- **Partners** - combined clinical, ops, financial for real-time PoC data & best practice recommendations (Queriable Inference Patient Dossier)

Path2Analytics: Landscape - 2

- **Outcome & Population Characterizations**

- Intermountain (Deloitte as partner) - 90M patient records, Outcome Miner = factors affecting outcome, Population Miner = relationship between treatment & outcomes
- McKinsey (BeyondCore) - “next 5% analysis”, 30M claims characterize “microsegments for next 5% of patients wrt cost, assign care managers

- **Prediction**

- Express Scripts - 1.5B prescriptions/year, modeling to predict which patients most likely to modify drug use behavior, proactive intervention

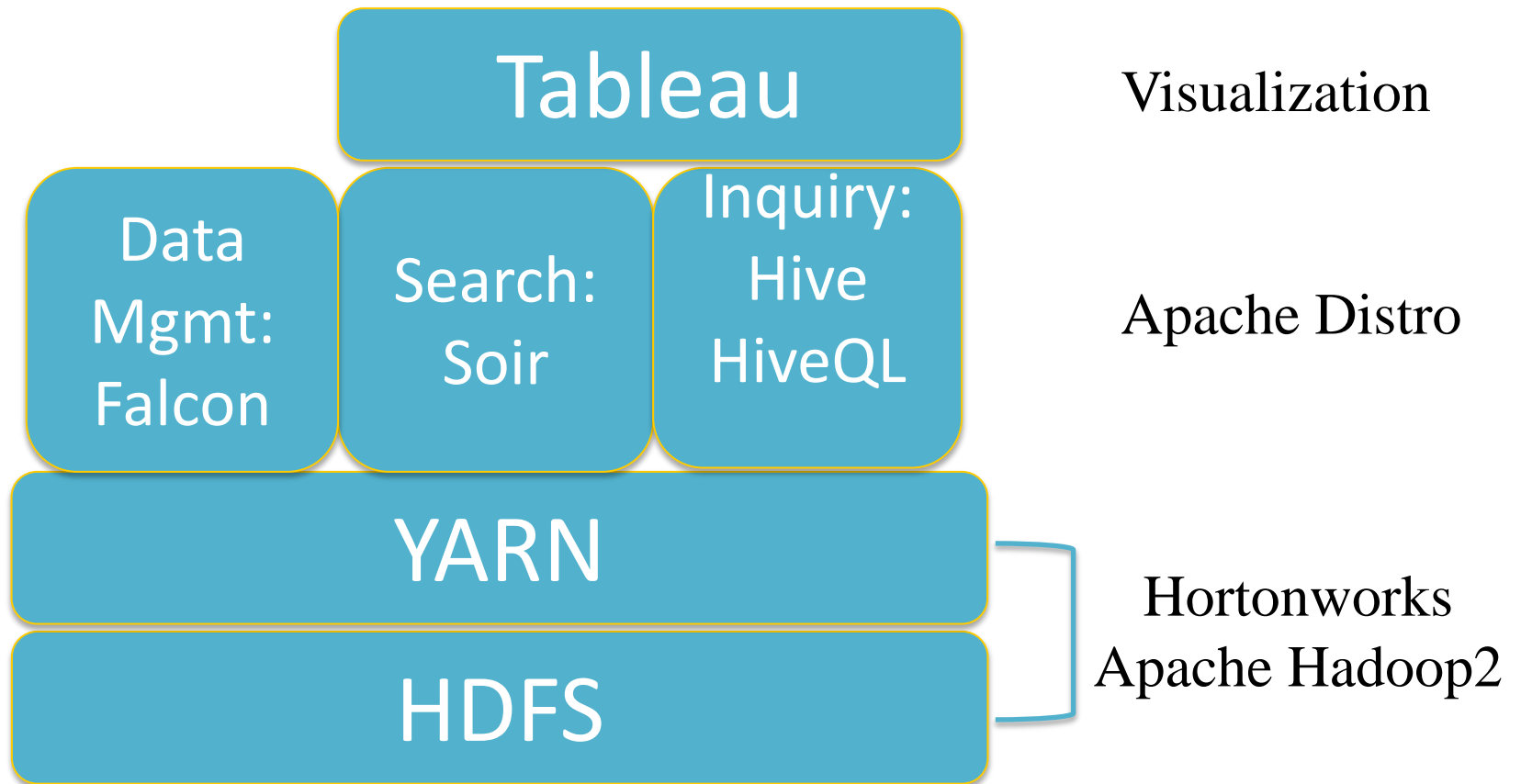
- **Research**

- Mt Sinai Medical Center (Ayasdi) - topological data analysis, analysis of entire e. coli (1M DNA variants) to determine bacteria’s response to antibiotics

Proposed Technical Approach

- **Preliminary characterization of data using existing reporting/BI system**
 - Can be used to determine if normalization/ETL is needed
- **Use of existing data extracts/warehouse if available**
 - Analysis of data quality needs to be done
- **Layering of existing data into Hadoop**
 - Data from multiple sources & of different types can be loaded
 - Normalization can be done entirely in Hadoop
- **Use of Yarn/MapReduce to generate results, &/or**
- **Use of HiveQL to do query**
- **Use of Tableau to visualize raw data**

Proposed Analytic Stack



Path2Analytics: Areas of Inquiry & Query Identification

- **Several Areas that other healthcare organizations have explored:**
 - Cost/outcome - requires integration of financial & clinical data
 - Trends in service provision: per location, per provider, per cost, per diagnosis, per outcome
 - Trends in demographics aligned with trends in encounters: per location, per diagnosis
 - Geolocational trends: per diagnosis, per demographic data
- **CHC has to identify areas of inquiry based on health center goals**
- **Once a goal & rationale for an inquiry is decided on, queries to address that inquiry can be planned**

- **Some ideas from initial discussions with pilot sites:**
 - Understand patient's outcomes, environment & components of their health
 - Will need to define these terms before relationships can be explored
 - Investigate the cost of patient care management
 - Cost of care is different than cost/outcome
 - Cost/outcome also requires careful definitions
 - Also requires availability of financial/cost data

Path2Analytics:CHC Issues/Inquiries 2

- Understand all quality indicators as required to manage patients
- Understand the coding that lies behind the reports
- Understand & utilize the EHR as a tool
 - Other work has shown that individual providers use the EHR differently & that this greatly influences the statistical characterization of everything from quality indicators to cost
 - This can be as straightforward as preferring to use notes to convey clinical information (instead of the EHR forms) or as subtle as coding preferences that may differ among providers



Current Status (3 months)

● Rural CHC

- 25K patients in area with population density of 11/sq. mi., service area = 24K sq. mi., 4 doctors & ~25 Nurse Practitioners & PAs
- EHR vendor will not release clinical data for analysis or make access to underlying DB available (as per license agreement)
- Vendor recommends using captive reporting tool supplied with EHR for any analysis or using contractor recommended by vendor
- Negotiations underway to make data available

Current Status (continued)

● Urban CHC

- 25K patients in an area with population density of 848/sq. mi., service area = 2K sq. mi., 40 doctors & ~ 110 Nurse Practitioners, PAs, Counselors, etc.
- Center owns it data & has already aggregated a normalized extract of clinical, financial & ops data
- Currently in query design phase
 - Most interested in cost/service & cost/outcome analysis along with trends in service utilization per location, & per patient – especially non-clinical services

Initial Lessons Learned

- **Semantic & Syntactic normalization of data essential for analysis**
 - Even if normalization is done at time of analysis (as in Hadoop), there must be standards for the data to be normalized against
- **Substantial time is needed for selection of analysis areas & inquiry development**
- **Large opportunity for operational optimization if data is available**



Thank You

Please feel free to contact me
for more information

David Hartzband

RCHN Community Health Foundation

55 Broadway, Suite 1502

New York, New York 10006

Phone: (212) 246-1122 x 722

(617) 501-4611 (m)

Email: dhartzband@rchnfoundation.org

dhartz@mit.edu