Big Data in Health Care
Strata – New York
September, 28th 2016

Sabrina Dahlgren, Director, Health care Delivery and Innovation, Kaiser Permanente - @sabridahlgren
Taposh Dutta Roy, Health Lead, Innovation and Data Science, Decision Support, Kaiser Permanente - @taposhdr
Rajiv Synghal, Chief Big Data Strategy, Kaiser Permanente
Approaching 11 million members, Kaiser Permanente is the largest integrated delivery system in the US. The organization employs 200,000 employees, including 19,000 physicians, operates 38 hospitals, 608 medical offices and other outpatient facilities, and generates $61 billion in operating revenue (2016).

Mission: Kaiser Permanente exists to provide high-quality, affordable health care services and to improve the health of our members and the communities we serve.
1960s | Dr. Sidney Garfield & Dr. Morris Collen

“We should begin to take advantage of electronic digital computers.”

Supporting Health Care with Technology

“Continuing total health care requires a continuing life record for each individual... The content of that life record, now made possible by computer information technology, will chart the course to be taken by each individual for optimal health.”

Sidney Garfield, MD
Hospital Computer Systems, 1974
OUR TECHNOLOGY JOURNEY

- Foundational (Facilities)
- EHR (KP Health Connect)
- Circle of Support (continuous availability, ancillary systems)
- Online (KP.org)
- Mobile App (Video Visits)

Transforming Health Care
- Precision Medicine
- Machine learning
- Patient 360 View (Cloud, IOT)
HEALTH CARE INDUSTRY

The future is almost here ...
HEALTHCARE ECOSYSTEM

Providers

Pharmaceutical Life Sciences

Payers

Patients/Consumer

Employers

Tech Players

Retail Pharmacies

Source: Bain and Co.

Medicity
A Healthagen Business

Allscripts

KAISER PERMANENTE

thrive
Percentage of consumers with at least one medical, health or fitness app on their mobile devices doubled from 2013 to 2015*
TECHNOLOGY ADOPTION

Consumers
- 60% willing to have a video visit with a physician through a mobile device
- 21% have used a mobile device to order a refill of a prescription
- 88% willing to share personal data with their doctor to find new treatments
- 67% “very satisfied” with experience at a retail clinic

Clinicians
- 81% say mobile access to medical information helps coordinate patient care
- 38% use email to stay connected with their chronic disease patients
- 58% would rather provide a portion of care virtually
- 74% say non-traditional venues (e.g., retail clinics) improve access to care

*HRI Consumer Survey, PwC, 2015 and HRI Clinician Workforce Survey, PwC, 2014 and 2015*
Health 2050 Vision

Traditional Model
- Population-wide demographics

Emerging Model
- Cohort-relevant measures

Future Model
- Individual N=1

Figure 1, Health 2050: The Realization of Personalized Medicine through Crowdsourcing, the Quantified Self, and the Participatory Biocitizen, Journal of Personalized Medicine
Analytics and Enterprise Information Strategy…

**Approach**

- **decision maker centered**
- Leaders & Managers
- Care and Service Providers
- Patients, Members, Groups

- **information as a strategic asset**

- Frame right questions
- Make better decisions fast
- Link decisions to action

**Strategy**

- **good data**
  - Data sources and platforms

- **right insights**
  - Business intelligence and analytics

- **right time and form**
  - Delivery and decision aids

**Outcome**

- **right information**
- **at the right time**
- **in the right hands**

- Easy to
  - Frame right questions
  - Know and do right things
  - Make better decisions
Making the RIGHT Architecture Choices...

### RDBMS Ecosystem -
*Fragmented Data and Processing on RDBMS Platforms.*

<table>
<thead>
<tr>
<th>Feature</th>
<th>RDBMS Ecosystem</th>
<th>Distributed Ecosystem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Governance</td>
<td>✓ Easy within a domain. Difficult across domains.</td>
<td>✓ Easy across systems. Easy to align fields and definitions.</td>
</tr>
<tr>
<td>Organizational Preparedness</td>
<td>✓ Mature People, Process.</td>
<td>➤<strong>Maturing People, Processes</strong></td>
</tr>
<tr>
<td>Operational Readiness</td>
<td>✓ Mature</td>
<td>➤<strong>Maturing. 24 x 7 Support. Active-Active DR. Class of Service 0/1.</strong></td>
</tr>
<tr>
<td>Row Vs. Columnar</td>
<td>□ Maturing</td>
<td>✓ Mature</td>
</tr>
<tr>
<td>Data Partitioning</td>
<td>□ Add on component. After thought.</td>
<td>✓ Partitioning and hashing are first class citizens.</td>
</tr>
<tr>
<td>OLAP</td>
<td>□ Design time and pre-built aggregations.</td>
<td>✓ On demand, real time aggregations. Agility.</td>
</tr>
<tr>
<td>Alignment</td>
<td>□ Version mismatch across functional components</td>
<td>✓ Functional components are an integral part of the ecosystem</td>
</tr>
<tr>
<td>Search &amp; Data Mining</td>
<td>□ Missing. Add on component.</td>
<td>✓ Native</td>
</tr>
<tr>
<td>h/w Scalability</td>
<td>□ Each tier has to independently scale - CPU/Memory.</td>
<td>✓ Built in scalability.</td>
</tr>
<tr>
<td>System Scalability</td>
<td>□ Vertical</td>
<td>✓ Horizontal</td>
</tr>
</tbody>
</table>
Landing Zone (Home to Secure and Organized Data)
- A Self Service Data Platform hosting both the raw and prepared data sets for quick business consumption.

- **Security and Data Organization are both First Class Citizens:** Five Layers of Security. Data Organized by domains and use cases.

- **Unified, Interconnected Data:** One place to store all data, in any format, large volumes.
  - Import external data to comingle and corroborate with internal data
  - De-Identify internal data to share with and leverage external partners.

- **Self-Service Exploratory BI:** Allows users to explore, discover, and mine data, with full security, using interactive analytical tools

- **Advanced Analytics:** Simple tabular data can mix with more complex and multi-structured data in ways that were never before possible
Systematic Data Liberation...

<table>
<thead>
<tr>
<th>Collect</th>
<th>Curate</th>
<th>Enrich</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Collect</strong></td>
<td><strong>Curate</strong></td>
<td><strong>Enrich</strong></td>
</tr>
<tr>
<td>Ingest Pre-checks</td>
<td>Verification</td>
<td>Enrich Pre-checks</td>
</tr>
<tr>
<td><strong>Pass</strong></td>
<td><strong>Pass</strong></td>
<td><strong>Pass</strong></td>
</tr>
<tr>
<td>2016-08-25 5:35</td>
<td>2016-08-25 5:35</td>
<td>2016-08-25 5:35</td>
</tr>
<tr>
<td>Ingest Process</td>
<td>Data Profiling</td>
<td>Enrich Process</td>
</tr>
<tr>
<td><strong>Failed</strong></td>
<td><strong>Pass</strong></td>
<td><strong>NA</strong></td>
</tr>
<tr>
<td>2016-08-25 5:35</td>
<td>2016-08-25 5:35</td>
<td></td>
</tr>
</tbody>
</table>

- **40 tables** (Q1: 35 tables)
- **40 tables**
- **12 tables**

Solving data repurposing – single copy, multiple use, enrich & context

Growth of run time
Growth of storage
Total tables processed

- archive
- continuous archive
- ingest
- true ELT
- semantic equivalence
Consistent User Experience: One Platform

Liberate  Collect  Curate  Enrich  Consume

episodes  pharmacy  membership  lab  kp.org  employee  hr  help desk  system logs

episode groupings  high utilizers  actionable findings  semantic  search prescriptions  hr analytics  risk intelligence

Data Platform – Landing Zone

Exploratory Intelligence

Visualize

Discover  Refine  Analyze  Mine

JDBC - Impala Lookup Query / Arcadia OLAP Query

Smart Data Zone (Semantic Layer)

Raw Data Zone  User Defined Zone  Refined Data Zone  Master Data Zone  Meta Data Zone  Reference Data Zone  Usage Data Zone

HDFS  All Data Encrypted @ Rest

Analyze  Machine  Redefine  Discover  Smart Data Zone (Semantic Layer)  Visualize  Collect  Curate  Enrich  Consume
360 Member View

- **Personal Behaviors**
  (Life Style Choices, Preferences, Activities, QoL)

- **Demographic Factors**
  (Age, Address, Employer, Industry)

- **Social Factors**
  (Friends, Family, Affiliations, Communication, Activities)

- **Medical Care**
  (encounter, labs, Rx, medical devices, etc.)

- **Family History and Genetics**
  Personal “-omics”
  (Genomics, Proteomics, Transcriptomes, Metabolomics)

- **Environmental Factors**

  - **Environment**
    (Temperature, Humidity, Pollen Count,..)

  - **Geographic**
    (Closest Hospital, Pharmacy, Care Clinic,..)
Analytics - Current State

Self Serve Reporting

Machine Learning
Machine Learning - Current Process

Getting relevant data

Applying variety of algorithms – training & testing datasets

Refining data (clean/impute)

Feature engineering

Dashboard for model performance

Continuous improvement

Data pipeline

Data pipeline
Our Analytics Strategy

**Mentoring**
Developed our people through training.

**Enabling**
Provided an infrastructure to explore and develop.

**Challenging**
Developed an internal crowd sourced machine learning challenge.
Mentoring

Training Programs
Training our people on big data and statistical tools.

In Person Events
Developed learning forum and venues to support machine learning.

Ongoing Learning
Lunch and learn seminars
What was the first problem solved through crowd sourcing and when?

1714 – British Board of Longitude Prize
Problem : To determine the longitude of a ship at sea.

Data Science Challenge

Competition enabled us to test an analytics ecosystem

Competition prompted us to develop a process to manage open source tools.

The data science team provided access to participants to de-identified data following HIPAA compliance.
Results from Challenge

Fast facts
• Over 100 KP data scientists participated in the competition
• 1000 models were submitted in 6 weeks
• Model performance improved by more than 5%
• … in less than <10% of time

Learnings
• Discussion forums were lively and collaboration increased
• New algorithm strategies were discovered
• Papers are planned to be published to share the learning on the algorithm strategies
STEP WISE APPROACH FOR CHANGE

Data Science
Training
Competition

Seminars

Leaderboard
Improvement

Culture Change

Model Selection and Model tuning

The challenge has been closed. Thanks all the organizers. I learn a lot in this challenge.

I know the winner is going to share their code at the end. However, the following questions are the ones I want the most from the winners.

1. Among all prediction methods, which one I shall start with?
2. How do I tune the parameters to get better result? There is no way to run all combination of the parameters. I would like to know if there is a general guil on this.
3. When one model can't be turned anymore, how to choose another method to seek better performance?

Thank you.
July 03 1 Like Reply —
Analytical Opportunities

1. Biometric Monitoring
2. Alerts/Dashboards
3. Resource Recommendations
4. Triage Recommendations
5. Expert Advice
6. Food Recommendations
7. Environment Monitoring
8. Event Prediction
9. Drive Recommendation
10. Brand Sentiment
Acknowledgement

We wish to acknowledge the contribution of many to this work:

• The Permanente Medical Group Physicians and the Permanente Federation

• Health Plan and Hospital Operations, Quality and Finance teams

• Kaiser Permanente Information Technology

It takes a “virtual” village … !