Information Lifecycle Management at Beth Israel Deaconess

John D. Halamka MD
CIO
The Statistics

• We generate 1 terabyte of clinical text data (structured and unstructured) per year
• We generate 19 terabytes of image data per year
• We have 250,000 active patients
• 20 terabytes/250,000 = 80 megabytes per patient per year
The Cost of Storage

• Standard Performance
  – Unreplicated .34 per Gigabyte per year
  – Replicated .68 per Gigabyte per year

• High Performance
  – Unreplicated .55 per Gigabyte per year
  – Replicated to Standard Performance is .89 per Gigabyte per year
Cost for Storage in Regulatory timeframes

• Text storage growing at 4 megabytes per year, retained for 15 years
  – Use arithmetic progression formula $n*(n+1)/2$
  – Assume replicated storage of .89 cents per gigabyte
  – 4 megs*15*16/2*.89/1000 = 42 cents per patient for first 15 years
  – 4 megs*15*.89/1000 = 5 cents per year thereafter

• Image storage growing at 76 megabytes per year, retained for 7 years
  – 76 megs*7*8/2*.89/1000 = $1.89 per patient for first 7 years
  – 76 megs*7*.89/1000 = 47 cents per year thereafter
Other considerations

• The definition of the official medical record is in flux
• BIDMC’s costs are low because of a stable set of technologies and vendors
• Multimedia may impact storage volumes
• Disaster recovery, high availability, and application life cycle requirements can impact storage volumes
• Emerging factors such as e-discovery will impact retention times
SAN, NAS, CAS

Application Data Usage

- **High Update Rates**
  - e.g., TRANSACTIONAL
- **Moderate Update Rates**
  - e.g., COLLABORATIVE
- **Low Update Rates**
  - e.g., ARCHIVAL

Common Everything
- Components • Connectivity • Management • Support
Migration from DVD to Atmos
Storage Tiering

- Tiering allows for cost appropriate storage choices through the complete data lifecycle.
- We have many tiers, across several platforms all designed to deliver a specific $/GB appropriate target for each different type of storage requirement.
Enterprise Storage Categories

- **SAN** (Storage Area Network): Performance
- **NAS** (Network Attached Storage): Access
- **CAS** (Content Addressable Storage): Immutable
- **COS** (Cloud Optimized Storage): Policy

<table>
<thead>
<tr>
<th>Transaction</th>
<th>Unstructured</th>
<th>Compliance</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symmetrix</td>
<td>Celerra</td>
<td>Centera</td>
<td>Atmos (software)</td>
</tr>
<tr>
<td>CLARiiON</td>
<td></td>
<td></td>
<td>Atmos (hardware)</td>
</tr>
</tbody>
</table>
SAN Block Storage

Data Center Host
• CCC
• SQL
• Oracle
• Exchange

Transport
• Cisco FC Switch
• iSCSI to FC bridge

Fibre Channel Array
• CLARiiON
• DMX
Network Access Storage

Client Host
- Web App.
- Desktop
- Server

Transport
- CIFS
- NFS
- SOAP
- REST

Storage Platform
- Celerra
- DataDomain
- Atmos
Content Addressable Storage

Applications
• PACs
• Medical Records

Transport
• CenteraAPI
• CIFS

Storage Array
• Centera
Cloud Optimized Storage

Application/Host
- PACs
- Medical Records
- Unstructured Data
- Other Image Formats

Transport
- CIFS
- NFS
- SOAP
- REST

Storage Platform
- Atmos (SW)
- Any disk (HW)
Business Rules

- Seamless data migration from tier to tier
- Policy based movement of data based on date, file type, size
- Tiering for infrequently used data to be removed from the backup process saving resources
Questions?

• jhalamka@caregroup.harvard.edu
• http://geekdoctor.blogspot.com