PARADIGM SHIFT IN QA WHEN TESTING DATA WAREHOUSES

Deepthi Krishnamaneni Sr. Analyst
Janaki Narasimhappa Sr. Analyst
Howard C Shang Director

DHTS
Agenda

- Profiles
- Why QA?
- Testing process
- Example
- Summary
- Q&A
• Located in central North Carolina
• 3 hospitals (in Durham and Raleigh)
• Dozens of Duke clinics and over 200 affiliated clinics
• 60,000 annual hospital visits
• 1.4 million annual clinic visits
• 2300 employed & affiliated physicians
• Not for profit
Information Management (IM)
Duke Health Technology Solutions (DHTS)

- Duke’s Data Warehouse Group (DWG)
  - 24 full-time employees
  - 2 Resource Managers, 1 Project Manager, 4 Architects, 3 Data Analysts, 7 ETL Developers, 2 BI Developers, 3 QA Analysts

- Decision Support Repository (DSR)
  - 12 major data sources
  - 2 Terabyte of data
Why QA?

- Exponentially increasing cost associated with finding software defects later
- Additional business costs of using incorrect data to make critical business decisions
Testing Process

- Data Completeness
- Data Transformation
- Data Quality
- Performance and Scalability
- Integration Testing
- User-Acceptance Testing
- Regression Testing
Data completeness

Ensures that all expected data is loaded

- Record counts and rejected records
- Comparing unique values of key fields
- Verify possible data errors
Data transformation

Ensures that all data is transformed correctly according to business rules and/or design specifications

- Create a Test plan
- Create test data
- Validate
  - surrogate keys
  - the data types
  - Referential Integrity
  - parent-to-child relationships (Orphans)
Data quality

Ensures that the ETL application correctly rejects, substitutes default values, Corrects or ignores and reports invalid data.

- There may be Rejections?
  - We correct the data
  - Notify the users
  - Skip the record
Performance and scalability
Ensures that data loads and queries perform within expected time frames and that the technical architecture is scalable.

- Volume Test
- Load Test
- Performance Test
Integration testing

Ensures that the multiple ETL processes function well together loading the same target tables

- End to End Testing
- Build Test data
- Interaction of ETL processes
User-acceptance testing

Ensures the solution meets users' current expectations and anticipates their future expectations.

- Use data from production
- Test database views comparing view contents to what is expected
- Includes more team members then QA
Regression Testing

Ensures existing functionality remains intact

- Modifications
- Enhancements
- Upstream systems changes
Example: Patients location lookup

Data Warehouse Testing:

Source file → Patient_id → DE#_Loc → Magic happens here

Target Table → Patient_key → Location_key
Example: Patients location lookup

Source
De #_Loc

Lkp1
De #_Loc
Location_code

Lkp2
Location_code
Location_name
Location_key

Source
Patient_id

Lkp3
Patient_id
MRN

Lkp4
MRN
Patient_key
Example: Patients location lookup

<table>
<thead>
<tr>
<th>Patient_id</th>
<th>De #_Loc</th>
<th>MRN</th>
<th>Location_Code</th>
<th>Patient_key</th>
<th>Location_key</th>
</tr>
</thead>
<tbody>
<tr>
<td>13645456</td>
<td>94</td>
<td>M001234</td>
<td>N36</td>
<td>123456718</td>
<td>EYE CENTER</td>
</tr>
<tr>
<td>23645457</td>
<td>45</td>
<td>M002365</td>
<td>N54</td>
<td>234567890</td>
<td>N58 GYN</td>
</tr>
<tr>
<td>33632456</td>
<td>94</td>
<td>M003363</td>
<td>N36</td>
<td>345678912</td>
<td>EYE CENTER</td>
</tr>
</tbody>
</table>
Example: Patients location lookup

<table>
<thead>
<tr>
<th>Source Lkp result</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient_id</strong></td>
<td><strong>De #</strong></td>
</tr>
<tr>
<td>13645456</td>
<td>94</td>
</tr>
<tr>
<td>23645457</td>
<td>45</td>
</tr>
<tr>
<td>33632456</td>
<td>94</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Target table</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient_key</strong></td>
<td><strong>Location_key</strong></td>
</tr>
<tr>
<td>123456718</td>
<td>EYE CENTER</td>
</tr>
<tr>
<td>234567890</td>
<td>N58 GYN</td>
</tr>
<tr>
<td>345678912</td>
<td>EYE CENTER</td>
</tr>
</tbody>
</table>
Example: Patients location lookup

### Source Lkp result

<table>
<thead>
<tr>
<th>Patient_id</th>
<th>De #</th>
<th>Patient_key</th>
<th>Location_key</th>
</tr>
</thead>
<tbody>
<tr>
<td>13645456</td>
<td>94</td>
<td>123456718</td>
<td>EYE CENTER</td>
</tr>
<tr>
<td>23645457</td>
<td>45</td>
<td>234567890</td>
<td>N58 GYN</td>
</tr>
<tr>
<td>33632456</td>
<td>94</td>
<td>345678912</td>
<td>EYE CENTER</td>
</tr>
</tbody>
</table>

### Target table

<table>
<thead>
<tr>
<th>Patient_key</th>
<th>Location_key</th>
</tr>
</thead>
<tbody>
<tr>
<td>123456718</td>
<td>EYE CENTER</td>
</tr>
<tr>
<td>234567890</td>
<td>N58 GYN</td>
</tr>
<tr>
<td>345678912</td>
<td>EYE CENTER</td>
</tr>
</tbody>
</table>

VERIFY THAT THERE IS NO DIFFERENCE
Application Testing

- Test Case: Verify that Patient has correct location in the report as per the spec.

- Patient Key: 1234

- Expected Location Name: Eye Center

- Actual Results? If Eye Center > Pass
Summary

- Importance of Testing
- Data Warehouse Testing process
- How different is Data Warehouse from Traditional Testing process
Q&A???

Contact Info:-
Deepthi Krishnamaneni
Phone no: 919-668-6587
Deepthi Krishnamaneni/DHTS.mc/Duke

Janaki Narasimhappa
Phone no: 919-613-6192
Janaki Narasimhappa/DHTS.mc/Duke