Push to Pull: How Lean Concepts Improve a Data Migration

Agile 2007 Experience Report

Rand Bradley
Projectway, LLC
Data Migration

Data Migration is a subproject within a much larger project

• Central and critical to the success of the whole project

• High degree of complexity due to:
  • Inexperience, not performed often
  • Legacy systems knowledge
  • Multiple data sources
  • Unclean data

If not managed and executed properly, Data Migration will be a bottleneck for the entire project!
Project overview

- Large-scale ERP implementation. (Siebel integrating with SAP)
- Complex: migrating 14 data sources to Siebel
- Data migration team goals:
  - Minimize go-live system down time
  - Migrate all assets per business requirements

Phase-1 (9 months)
Americas, Europe

Planned data loads, integration points

Phase-2 (3 months)
Asia Pacific

Testing improvements implemented

Change to pull system
### Architectural Overview

1. **STEP 1**
   - Load source copy database

2. **STEP 2**
   - Load STAGE database

3. **STEP 3**
   - Load vendor dataloading tables

4. **STEP 4**
   - Run vendor dataloading process

5. **STEP 5**
   - Post dataload clean-up
## Phase comparisons…

<table>
<thead>
<tr>
<th>Phase-1</th>
<th>Phase-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 large batch – process all data per job step</td>
<td>Process data in small batches, overlapping steps</td>
</tr>
<tr>
<td>Process data as fast as possible at each step (sub-optimization)</td>
<td>Process data only as fast as the slowest step consumed</td>
</tr>
<tr>
<td>Load processing time 1+ weeks (some loads 10-15 days)</td>
<td>Turnover full load under 24 hours</td>
</tr>
<tr>
<td>9 practice loads in 9 months</td>
<td>25+ practice loads in 2 months</td>
</tr>
<tr>
<td>Load process extremely labor intensive (~200 manual steps)</td>
<td>Automated, very little labor required to run data load</td>
</tr>
<tr>
<td>Unrepeatable process – no two loads ever the same</td>
<td>Automated and fully repeatable</td>
</tr>
</tbody>
</table>
Phase-1 Issues

- Poor performance
- Sub-optimization
- Resource conflicts
- Unrepeatable processes
- Overworked team
- System crashes
- Costly restarts
- Poor quality
Phase-1 Assessment

- Poor quality due to employed testing processes
- Labor-intensive migration processes
- Informatica “development bottleneck”

... and most important...

- Processing data in one large batch though the system
Black Box Testing

- Extremely wasteful – high costs, low quality
- Slow feedback loop
- Ability to find root cause of issues? Extremely low!
- Difficult to change direction
- Changes are risky, especially late in the project
Phase-1 Success: Inline Testing

Unit tests at each step

- Immediate feedback
- Easier to discover root causes to problems
- Reduced cycle time, easier to make changes
- Reduces the risk of change, even late in the project

“Quality control should check the process, not the product.”
– Goldratt, “The Race”, commenting on Dr. Deming’s approach to quality control
Phase-1 Effects of large batch size

Push: Force data through system, one step at a time

STEP 2 (stage)
- Long load times, labor intensive
- Failures costly, difficult to recover
- Creating your own bottlenecks
- Overloaded databases
- Extra work to improve performance
- Difficult to test

STEP 3 (eim)

STEP 4 (base)

STEP 5 (post)

Quality & Throughput
Development costs

Copyright © 2007 Projectway, LLC
Phase-2  Small batch size effect

Pull: Flow data through the system at the pace of the slowest step.

- Fast loads, automated
- Easy to recover from failure
- Less hardware resources required
- Easier to tune bottlenecks
- Much easier to develop and test

Queuing Theory: Get more done faster if utilize less than 100% of resources
## Phase-2 Pull system example

<table>
<thead>
<tr>
<th>BATCH_NUM</th>
<th>SRC</th>
<th>FIRST_ID</th>
<th>LAST_ID</th>
<th>STAGE</th>
<th>EIM</th>
<th>BASE</th>
<th>POST</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SAPJP</td>
<td>374952</td>
<td>384982</td>
<td>COMPLETE</td>
<td>COMPLETE</td>
<td>COMPLETE</td>
<td>READY</td>
</tr>
<tr>
<td>2</td>
<td>SAPJP</td>
<td>364938</td>
<td>374951</td>
<td>COMPLETE</td>
<td>COMPLETE</td>
<td>IN_PROGRESS</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>SAPJP</td>
<td>354916</td>
<td>364937</td>
<td>COMPLETE</td>
<td>COMPLETE</td>
<td>IN_PROGRESS</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>SAPJP</td>
<td>344888</td>
<td>354915</td>
<td>COMPLETE</td>
<td>COMPLETE</td>
<td>IN_PROGRESS</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>SAPJP</td>
<td>324811</td>
<td>334847</td>
<td>COMPLETE</td>
<td>COMPLETE</td>
<td>IN_PROGRESS</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>SAPJP</td>
<td>314752</td>
<td>324810</td>
<td>COMPLETE</td>
<td>IN_PROGRESS</td>
<td>IN_PROGRESS</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>SAPJP</td>
<td>294617</td>
<td>304656</td>
<td>COMPLETE</td>
<td>IN_PROGRESS</td>
<td>IN_PROGRESS</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>SAPJP</td>
<td>284570</td>
<td>294616</td>
<td>COMPLETE</td>
<td>IN_PROGRESS</td>
<td>IN_PROGRESS</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>SAPJP</td>
<td>274509</td>
<td>284569</td>
<td>COMPLETE</td>
<td>IN_PROGRESS</td>
<td>IN_PROGRESS</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>SAPJP</td>
<td>264435</td>
<td>274507</td>
<td>COMPLETE</td>
<td>IN_PROGRESS</td>
<td>IN_PROGRESS</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>SAPJP</td>
<td>254171</td>
<td>264434</td>
<td>COMPLETE</td>
<td>IN_PROGRESS</td>
<td>IN_PROGRESS</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>SAPJP</td>
<td>241422</td>
<td>254170</td>
<td>IN_PROGRESS</td>
<td>IN_PROGRESS</td>
<td>IN_PROGRESS</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>SAPJP</td>
<td>226359</td>
<td>241417</td>
<td>IN_PROGRESS</td>
<td>IN_PROGRESS</td>
<td>IN_PROGRESS</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>SAPJP</td>
<td>216284</td>
<td>226358</td>
<td>READY</td>
<td>IN_PROGRESS</td>
<td>IN_PROGRESS</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>SAPJP</td>
<td>206243</td>
<td>216283</td>
<td>READY</td>
<td>IN_PROGRESS</td>
<td>IN_PROGRESS</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>SAPJP</td>
<td>196207</td>
<td>206242</td>
<td>READY</td>
<td>IN_PROGRESS</td>
<td>IN_PROGRESS</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>SAPJP</td>
<td>186025</td>
<td>196206</td>
<td>READY</td>
<td>IN_PROGRESS</td>
<td>IN_PROGRESS</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>SAPJP</td>
<td>175983</td>
<td>186024</td>
<td>READY</td>
<td>IN_PROGRESS</td>
<td>IN_PROGRESS</td>
<td></td>
</tr>
</tbody>
</table>
## Summary of principles

<table>
<thead>
<tr>
<th>Principle</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-inventory</td>
<td>Small batch size, overlap processing</td>
</tr>
<tr>
<td>Eliminate waste</td>
<td>Minimize defects, delays, handoffs, unnecessary complexity</td>
</tr>
<tr>
<td>Build quality in</td>
<td>Shorten feedback, unit testing, test each step, process testing, repeatable</td>
</tr>
<tr>
<td>Optimize the whole</td>
<td>Pull system, average cycle time, limit to capacity, fast as slowest step</td>
</tr>
</tbody>
</table>
Recommended resources

Theory of Constraints
- *The Goal*, Goldratt
  - Everyone should read this book…
- *The Race*, Goldratt, Fox
  - Companion book to *The Goal*
- *Viable Vision*, Kendall
  - Good summary of the TOC components and how they relate

Lean
- *Implementing Lean Software Development*, Poppendieck
- *Lean Practitioners Course*, Poppendieck
Questions?

Thank you!

Rand Bradley
rbradley@projectway.com
http://www.projectway.com
415-430-5535

data migration • project rescue • project management & training •
software development environment consulting • software testing