Enabling ETL Test Automation in Solution Delivery Teams

Subu Iyer
Oct 21, 2014
INTRODUCTION
What is ETL?

Extract Transform and Load

ETL

Source DB

Source DB

Source Files

Target DB

Target DB

Target Files
Examples of ETL Projects

- Import member data
- Import provider data
- Import LOV type data
- Copy claims, billing data
Challenges to ETL Testing

• Complex data requirements
• Project based approach (not Product based)
• Lack of testing tools
• Lengthy and error prone manual tasks
ETL Testing Misconceptions

• ETL tests are different
  • Tests need to reproduce the ETL job
  • Tests need to validate data (rather than the functionality)
• ETL projects don’t need regression tests
  • Since these are 1-time custom solutions
  • Once built, don’t need to be enhanced or optimized
• ETL test automation is prohibitively expensive
Then and Now…

2 Years Ago

• No belief in Test Automation for ETL
• Zero automated tests

Today

• 5 out of 8 teams have automated tests
WHY ENABLE ETL TEST AUTOMATION?
Typical ETL Development Process

- New Business Requirement
- Design and Data Modelling
- ETL Implementation
- Testing
- Release
- Maintenance
Typical Problems…

**Rigid: Design-Implement-Test**

- New Business Requirement
- Design and Data Modelling
- ETL Implementation
- Testing
- Release
- Maintenance

**Looks perfect until implementation starts**

- Long design phase
- Short testing phase
- Defects are found late and are expensive to fix

**Requirement changes derail the project**
Agile Approach

New Business Requirement

Design Increment

Release

Testing

ETL Implementation

Increment
Agile Approach

- New Business Requirement
- Design Increment
- Release
- ETL Implementation
- Testing

Continuous integration needs testing

Design changes need testing

Implementation changes need testing
Benefits of ETL Test Automation

- Running automated tests is easy and cheap
- Find defects as soon as they are introduced
- Tests monitor and track our progress
- Safety net for development
- Makes incremental development possible
- Quicker defect fixes
- Shorter project implementation time
Enabling Teams

• Having a dedicated team build and maintain automated tests for all teams does not scale

• Educating team members helps them identify and fix problems themselves

• Setting up automated processes establishes a platform for other teams
HOW ARE WE ENABLING ETL TEST AUTOMATION?
Collaborate with Teams

Solution Delivery Team
- Data Model
- ETL Process Details
- SQL Queries

Quest Team
- Test Framework
- Testing Process

© 2014 Cambia Health Solutions, Inc.
Best Practices Exchange

• Present test automation solution
• Demonstrate success
• Pitch ideas
• Present problems
• Discuss test plans
Build a Better Tool

Basic Requirements

- Getting started is easy
- No disruption to current processes
- Platform independent
- Tests - public and readable
Build a Better Tool

Tool Development Approach

- Start simple, start small
- Collaborate with 1 team to build 1 test
- Grow the tool functionality organically
Build a Better Tool

Picking a Test Runner

- Use existing, mature tools/libraries
- Tests are readable
- Tests can be shared easily
Build a Better Tool

Advantages of using FitNesse

- It’s a wiki!
- Read and run tests from a browser
- Great for collaboration
Start with Simple Tests

Set Up
● Establish connections
● Get database to known state

Test Methods
● Run ETL jobs
● Verify data

Tear Down
● Clean up database
● Close connections
Educate

• Demonstrate success
• Training sessions
• Involve multiple teams and departments
RESULTS
## Then and Now...

### 2 Years Ago
- No belief in Test Automation for ETL
- Zero automated tests
- Visual inspection
- Defects found late in the development cycle

### Today
- 5 out of 8 teams have automated tests
- Tests run daily or kicked off by clicking a button
- Defects are found as soon as the job is run
- 2 teams are writing tests before the product functionality is built!
What’s next…

• Propagate and gather feedback from all
  • UAT, System Analysts, Product Owners, Business
• Improve capabilities of the testing framework
• Add functionality to trigger ETL jobs
• May be a GUI to create tests

• Continue improving the framework to improve tests and test coverage