Maximize Your Return on Big Data

May 15-18, 2012
Automating the Integration Factory

Nagaraj Sastry
HCL
HCL

$6.2
BILLION

32
COUNTRIES

90000
EMPLOYEES

Employees First

Trust, Transparency and Flexibility

Value Centricity
Agenda

• Engagement Overview
  • Challenge & Objectives

• Lean Integration Principles
  • Characteristics of an Integration Factory
  • Building a Successful Factory Model

• Solution Architecture

• Value Proposition
Engagement Overview

- Finance & Risk System Transformation (FiRST) is a major strategic change initiative undertaken to deliver “best of breed” application functionality in the Finance, Risk, Treasury and Tax functions
  - FiRST is intended to deliver operational and financial benefits to the Finance, Treasury, Risk and associated support functions of the bank.
  - Key objectives of FiRST is to meet statutory obligations, support decision making, support organizational changes and managing assets and liability positions

- Data Transformation Program (DTP) solution comprises of a “smart pipe” that feeds raw data extracts from many source systems and provides standardized feeds of data into the FiRST architecture
Challenge – AS IS has evolved in an ad hoc fashion

Inconsistent results
DTP Objectives

- I need to standardize the data to the format required by downstream FiRST applications.
- The data should be traceable back to the originating “True Source” and forward through the FiRST architecture to business data requirements.
- Enriched – with reference data to meet the needs of downstream applications.
- Reconciled - back to source and forward to target.
- Adaptive – allowing flexibility to handle changing data requirements.
- The data should be validated and cleansed - to a standard of data quality which meets the needs of FiRST.
DTP – Solution Architecture
Lean Integration Principles
# Lean Integration in DTP

## Built in Quality – Test driven development and continuous testing
- Early test case preparation and then coding to fulfill tests. Fixing problems as they are discovered
- Ensuring only the required functionality is developed following standards

## Sustaining knowledge - Factory assets creation
- Creating re-usable components, patterns and artifacts
- Maintaining comprehensive, unambiguous and up to date documents with version control.

## Delivering fast - Automation
- Automating the load to stage and publication mappings using MAV, MAE and Custom SDK’s
- Using identified and already built components and patterns rather than re-inventing the wheel.

## Planning for change and empowering the team – Factory delivery model
- Keeping the core Re-use team and Assembly team. Robust yet loosely coupled code.
- Adopting variable staffing model for rapid ramp up and ramp down of extended Assembly teams

## Eliminating waste – Eliminate non value add efforts
- Considering and estimating only the value added effort
- Project maturity, reusability and automation factors to eliminate waste and reduce overall effort
Accelerate

Test Driven Approach
Test Driven Development Methodology

1. Prepare Test Cases based on Mapping Specs and Sample Source Data
2. Create Test Schedule
3. Continuously execute tests till it pass

Code Refactoring
- Review Mapping Spec., sample source data

Code Integrating
- Refer Test cases and Develop scheduled tests

Code Check-in
- Code Check-in
reuse

Factory Asset Creation
Reusable Components

- Process Management
- Data Exception Management
- Reference Data Management
- Monitoring
- Archival
Reusable Templates

- Load to Stage
- Data Transformation
- Publication
- Detail Design Template
- Standards and Best Practices
- Deployment Guide
Reusable Framework

- Process Management
- Data Exception Management
- Reference Data Management
- Monitoring
- Archival
- Load to Stage
- Data Transformation
- Publication
- Detail Design Template
- Standards and Best Practices
- Deployment Guide
Reconciliation – A Special Consideration

Transformation Level
Stage Level
Performance
Scalable
Detailed Logs
User Interface
Workflow Management (Approval Process)
Standardization Schema
Rejected Records
Staging
Reference Data
Execution Data
Reconciliation Database
Reconciliation Engine
File System
ETL Layer (Informatica)
Setting up the Right Team
DTP Factory Setup Model

**Assembly line 1 (Core Team)**

**Mapped roles**
- ETL Architects
- ETL Designers/Developers

**Activities**
- Data Mapping
- Data Modeling
- Detailed Design
- Reusable Components creation
- Priority Build and Testing

**ETL Design (Reuse Team)**

**Activities**
- Data Mapping
- Detailed Design
- Design Review
- Priority Build

---

**Assembly line 2**

**ETL Development**

**Mapped skills**
- ETL Leads – Assembly Line
- ETL Developers

**Activities**
- ETL Development
- Unit Testing

**Assembly line 3**

**Mapped skills**
- ETL Testers

**Activities**
- Fix defects
- Link Testing

---

**Reusable Component Library (Factory Assets)**

**System Test Support**

- Continuous Test
- Unit Test
- Coding

---

- Assembly line 1
- Assembly line 2
- Assembly line 3
- Onsite
- Offshore
DTP Delivery Model – Ways of Working

**Onsite**

**Reuse Group**
- Assembly members periodically rotate in & out of Reuse group

**Assembly Team**
- Consumes reusable components, works on initial instances and transfers knowledge to offshore assembly teams

**Offshore**

**Assembly Group – 1 (HCL)**
- Developers

**Assembly Group – 2 (HCL)**
- Developers

**Assembly Group – n(HCL)**
- Developers

**Reuse Group** are point of contact for other work streams, ensure Environment and Test data availability for build, Identify, design & build reusable components, create functional specification

Feed specific assembly teams, consume knowledge from onsite assembly teams, works on instance build replication, provides feedback for continuous improvement

- One Tech Lead for every 5-7 developers
- One Team lead/manager for every 20 members
Automate

Enabling Consistency
Automate ETL Code Generation

Lean Automation

- Focus to improve the efficiency by automating the manual activities
- Aim to reduce the development time
- Automate deployment process
- Automate the process of continuous testing to improve quality of delivery

Methodology to automate code generation using Informatica

<table>
<thead>
<tr>
<th>Using Framework</th>
<th>MAV</th>
<th>MAE</th>
<th>SDK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Framework Mapping</td>
<td>Mapping Templates which are Published Using Mapping Architect for Visio</td>
<td>Mapping Analyst for Excel(MAE) to Generate Source and Target Structures</td>
<td>Java application to generate source And Target definitions into Informatica Repository And generate mappings using MAV</td>
</tr>
</tbody>
</table>
Automation - IDP – SDK Solution

- Informatica developer platform has been identified to automate the source and target creation
- Creation of customized flat file sources and Oracle target definitions creation is automated using Java SDK
- Parameter file generation for MAV is automated using Java taking flat file as input
- Using the above two files, mappings will be generated from command line of developers PC using the existing MAV templates

- Using MAV, MAE and SDK, Load to Stage and Publication ETL process is completely automated
Value Proposition
Key Learning

- Lean Integration Principles
- Re-Usable Components
- Core Team concept
- Tool capability to support Lean principles
- ROI
Estimation Methodology

- Projection of Capacity – Person Days
- Projection of FTE Count – Existing Vs New
- Resource Ramp Up Plan

Note: DTP estimation is based on Informatica ICC estimation methodology, above methodology was used by HCL based on its experiences, to validate ICC estimation methodology.
Value Proposition

Estimated values for a new release development of about 700 ETL mappings:

- 10-20% upfront reduction in build effort due the knowledge base and reusable components built in prior release
- 30-50% further reduction in build effort can be achieved post automating Load to Stage and Publication mappings using MAV, MAE and SDK
Thank You

Hi I’m HCL

We are at Booth G1

$6.2
BILLION

32
COUNTRIES

90000
EMPLOYEES

Nagaraj Sastry
Business Intelligence Principal
Enterprise Transformation Services
NSastry@hcl.com

HCL ETS-BIS Practice
Enterprise Transformation Services
ets@hcl.com