Getting Real About Data Virtualization
Informatica Data Services

Ash Parikh
May 3, 2011

Defining Data Virtualization

- Makes multiple heterogeneous data sources appear as one
- Federates data in real-time & also supports physical materialization to DW
- Abstracts data sources from consumers and insulates from change
- Hides & handles complexity (quality, rich transformations, bus-IT collaboration)
- Let’s the business own the data & define the rules while IT retains control

Simple Data Federation Does Not Cut It
Here’s Why…

The Traditional DI Process

1. Source
2. Target
3. Mapping
4. Workflow
5. Execute/Test
6. Deploy

Typical Value Stream Map - Too Much WAIT & WASTE

The biz gets involved late & does not get what is needed

The Problem(s)
It Takes too Long to Deliver Data the Business Needs!

Change Request ... Approve & Prioritize ... Analyze & Design ... Build ... Test ... Deploy

Data Is Everywhere & Growing!
The Impact

Reports Take too Long

Reporting Scenario: On-going requests for data that is NOT in the DW

Business  What if?  IT
Change Request  Deploy to Production

Weeks/Days

Change Request ... Approve & Prioritize ... Analyze & Design ... Build ... Test ... Deploy

• 66% of BI requirements change on between a daily and monthly basis
• 71% of the respondents said they have to ask data analysts to create custom reports for them
• 36% of custom report requests require a custom cube or data mart to answer the request
• 77% of respondents cited that it takes between days and months to get their BI requests fulfilled

**HealthNow Case Study**

No Reuse of Data Services for BI, MDM & SOA

Agreement on MEMBER & Attributes Time-Consuming & Painful

Different Price Info in Each BU & 1700 Dev Hours to Add 1 Product

16 Heterogeneous Enterprise Stores With Large Volumes of Data

- **30,000 Data Marts (MS Access)**
- **Data Warehouse (DB2)**
- **Facets (Benefits, Products) (Sybase ASE)**
- **Product Config Mgmt (MS SQL Server)**

---

**Lean Integration & Data Virtualization**
Data Virtualization Built on Lean Integration Principles

The Traditional DI Process
1. Source
2. Target
3. Mapping
4. Workflow
5. Execute/Test
6. Deploy

The New Agile DI Process
1. Logical Data Object
2. Source
3. Integrate
   - Preview
   - Profile at any stage
   - Apply rich transformations
   - Apply DQ & masking rules on-the-fly
   - Federate data without data movement
4. Comment/Tag
   - Debug
5. Deploy as Reusable Data Services
   - Web services or SQL

Original Value Stream Map - Too Much WAIT & WASTE
Optimized Value Stream Map - Cut the WAIT & WASTE

Self Service – Analyst Empowerment & Business-IT Collaboration
- Easily map sources to physical & virtual targets
- Quickly find data via integrated business glossary
- Specify transformations with reusable expressions
- Include pre-built rules and mappings (e.g. ETL, DQ)
- Collaborate, test and validate specification results
- Automatically generate ETL mappings and SQL views

Improved analyst & developer productivity
5x Faster Direct Data Access, Increased Reuse, Improved Governance & Agility

**REUSE**

1 week
(vs. 3 months)

"Virtual Table"

**PRODUCT ORDER MEMBER CLAIM**

**VIRTUAL TABLE**

30,000 Data Marts (MS Access)
Data Warehouse (DB2)
Facets [Benefits, Products] (Pybase ASE)
Product Config Mgmt (MS SQL Server)

**INFORMATICA**

Case for Making Data Virtualization as part of Your Information Management Best Practices

**INFORMATICA**
Next Steps

Let’s Talk!
Architect to Architect Webinar Series
Data Virtualization Architecture & Best Practices for Agile Data Integration
May 19, 2011

“SOA Data Integration Architecture Group”

Forrester IaaS
(Data Services) Wave

INFORMATICA
The Data Integration Company™