Maximize Your Return on Big Data

May 15-18, 2012
What’s New with Informatica Data Services & PowerCenter Data Virtualization Edition

Kevin Brady, Integration Team Lead
Bonneville Power

Wei Zheng, Product Management
Informatica

Ash Parikh, Product Marketing
Informatica
Agenda

• The Problem
• Data Virtualization Can Help
• Informatica’s Data Virtualization Solution
• Customer Example
• What’s New in v9.5 & Roadmap?
• Next Steps & Q&A
Safe Harbor

The information being provided today is for informational purposes only. The development, release and timing of any Informatica product or functionality described today remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision. Statements made today are based on currently available information, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products or functionality in the future.
What is the Problem?
BI will be the top priority for the CIO, in 2012!

“Demands by users of business intelligence (BI) applications to "just get it done" are turning typical BI relationships, such as business/IT alignment and the roles that traditional and next-generation BI technologies play, upside down. As business users demand more control over BI applications, IT is losing its once-exclusive control over BI platforms, tools, and applications.” – Boris Evelson, Forrester Research, Blog - “Top 10 BI Predictions for 2012”

Had this discussion yet?
- Need for a new BI infrastructure
- Replacing spreadsheets
- Faster data access & reporting

Our world will be turned upside down...
How Long Does it Take to Deliver New Critical Data or Reports to the Business?

On average, how long does it take to add a new source of data to your data warehouse?

- 1 week
- 2 weeks
- 3 weeks
- 1 month
- 2 months
- 3 months
- 4-6 months
- 6 months or more

On average, how long does it take to create a complex report or dashboard with about 20 dimensions, 12 measures, and 6 user access roles?

- 1 week
- 2 weeks
- 3 weeks
- 1 month
- 2 months
- 3 months
- 4-6 months
- 6 months or more

On average, how long does it take to change a hierarchy (e.g., a new way of classifying products or organizing sales regions)?

- 1 week
- 2 weeks
- 3 weeks
- 1 month
- 2 months
- 3 months
- 4-6 months
- 6 months or more

2011 TDWI BI BENCHMARK REPORT
Organizational and Performance Metrics for Business Intelligence Teams
It Takes too Long to “Get the Data Out”

“Business”

IT

Data Warehouse

Unstructured Data

Spread Marts

Applications

Mainframe

Trust

It Takes too Long to “Get the Data Out”
Fast, Direct Access in “Days NOT Months”

Empower the Analyst to validate results & define business rules

Turn many sources into ONE with Data Virtualization
Using a BI Layer Just Won’t Scale

- No Common Data Access Layer
- No Easy Way to Handle Change
- No Data Quality & No Data Consistency

Unstructured Data  Spread Marts  EDW  Applications

No Reuse

Informatica World
Simple Approaches Don’t Solve it Either

“Simple” Data Virtualization
(Built-On Data Federation)

- Limited or Data Source Profiling Only
- SQL/XQuery Only Transformations & No Data Quality
- Cannot Easily Move to Persistent Store or Reuse
- Delivers
- Merger
- Access

- Addresses specific use cases
- No data movement / no copies / only federation
- Code heavy / not model-based / no reuse
- Not tools for business self-service
- SQL/XQuery-only transformations
- No data profiling / no data quality

Time GAINED upfront is nullified by Time SPENT on more processing

It’s like ONE step forward & TWO steps backward
What is Data Virtualization & How Can it Help?
What is Needed to Solve these Problems?

Think Virtual Machines for DATA!

**Logical View of All Underlying Data**

**Data Consumers**

**Data Abstraction**

**Enterprise Data Sources**

**COMMON ACCESS LAYER ACROSS MANY DATA SOURCES**

**FAST, DIRECT ACCESS TO DATA THE BUSINESS TRUSTS**

**DATA ABSTRACTION & REUSE OF SKILLS/LOGIC**

**SUPPORT ALL USE CASES**

- BI / DW
- MDM
- SOA
Are We Talking About TWO Separate Tools?

Which of the following DI techniques are you using in your DI solutions today? Click one or more of the following answers, in priority order from most used to least used.

**Priority Order**

- Extract, transform, and load (ETL)
- Data federation or virtualization
- Extract, load, and transform (ELT)
- Messaging or application integration
- Replication or data synchronization
- Event processing

**Percentage of Respondents**

Which of the following best describes your organization’s portfolio of DI tools today? For your organization’s next generation DI implementation, how would you prefer that the DI portfolio be?

**TODAY**

- Using multiple DI tools from multiple vendors: 44%
- Using just one DI tool: 22%
- Mostly hand-coded without much use of vendor DI tools: 18%
- Using a DI tool that's part of an integrated suite of data management tools from one vendor: 9%
- Using multiple DI tools from one vendor: 3%
- Other: 4%

**WOULD PREFER**

- Using multiple DI tools from multiple vendors: 25%
- Using just one DI tool: 24%
- Mostly hand-coded without much use of vendor DI tools: 6%
- Using a DI tool that's part of an integrated suite of data management tools from one vendor: 42%
- Using multiple DI tools from one vendor: 6%
- Other: 2%
What Are the Top 3 Key Capabilities for a Project that Needs Data Virtualization?

If “PERFORMANCE” is a given...

Which technologies do you think are critical in your data virtualization project? (Select all that apply)

- Data Quality
- ETL
- ESB/EAI
- BPM
- Data modeling
- Caching

Dataset - 600

Source – Informatica Data Virtualization Expert’s Forum, 2011
What is Informatica’s Data Virtualization Solution?
What Does the Ideal Solution Look Like?

“Simple Data Virtualization” Addresses ONLY TWO Boxes. THINK ABOUT IT!
1. Model

- Represent underlying data as business entities (CUSTOMER)
- Provide a common logical view or abstraction of all data
- Import logical model from 200+ modeling tools (ERWIN)
- Use visual and metadata based mapping language
- Instantly reuse logical data object for all applications

Common Data Access Layer – Logical Data Object

CUSTOMER  ORDER  PRODUCT  INVOICE

Unstructured Data  Spread Marts  Data marts  EDW  Applications
2. Access and Merge

Turn many data sources into ONE with Data Virtualization

Analytical Data  Transactional Data  Master Data  Archived Data  Interactional Data

Social  Warehouses  NoSQL

Cloud Computing  Application  Database  Unstructured  Partner Data
3. Profile in RT

Rich set of integrated profiling capability to find data anomalies and to discover keys and hidden relationships:

- Column & Rule Profiling
- Midstream or Comparative Profiling
- Join & Overlap Analysis
- Primary Key / Foreign Key Profiling
- Dependency Profiling
4. Transform in RT

- Metadata-driven, codeless, graphical environment
- Rich, pre-built library of advanced transformation
- Integrated Data Quality transformations
- Define policies to mask sensitive data in real time
5. Reuse Instantly

- Instantly reuse LDOs for any mode/protocol (SQL, WS)
- Single click deployment to batch
- Execution & optimization separate from design-time
- No re-development & re-building of LDOs
6. Move or Federate

**Data Federation**

- Specific use cases
- No data movement / no copies
- Real-time federation
- SQL/XQuery-only transformations
- No data quality / business validation

Single-click deployment to PowerCenter (batch)

**Data Integration**

- Majority of use cases
- Physical data movement
- Bulk/batch, near real-time, real-time
- Advanced transformations
- Built-in data quality
7. Scale & Perform

- Leverage the proven, high-performance Informatica engine
- Optimized SQL Query engine & graphical Query Plan
- High-performance Web services server
- Rich set of optimizations & caching mechanisms
  - Rule Based, Cost Based, Push Down, Early Projection, Early Selection, Semi-Join, Virtual Table & Result Set Caching
- Fine grained access control, WS-Security & pass-through security
- Database, Schema, Table, Column, Row-Level (v9.5) security
Ready to Win?

“Simple Data Virtualization” Addresses ONLY TWO Boxes. TELL US & WIN!
Self-Service – The Keystone!

- Involve Business Early & Often in Design & Validation
- Rapidly Prototype & Iterate to Meet Requirements
- Apply Advanced Data Transformations & Data Quality in Real Time
- Analyze & Profile Data & Logic Anytime
- Access
- Merge
- Deliver
- 1-Click Move to DW or Provision to Portal or BI
- Time
- Cost
- Risk

CUSTOMER
What Does Informatica’s Data Virtualization Solution Look Like?

✅ New PowerCenter Edition for AGILITY & PRODUCTIVITY

✅ Combines:
  - Data integration (PowerCenter SE)
  - Data Virtualization (IDS Full Use)
  - Data Profiling (IDE Full Use)
  - Business-IT Collaboration (Analyst)

✅ Packaged for simplicity and attractively priced

✅ Reuses existing skills and resources

Informatica proprietary/confidential. Informational purposes only. No representation, warranty or commitment regarding future functionality. Not to be relied upon in making purchasing decision.
Let's Hear from Kevin!
Modernization Can Take Down Your Business

A Week Lost When Apps. Go Down =

Many Weeks Lost in IT Productivity
So...

ARCHITECT FOR CHANGE

Build-In Flexibility Only WHERE it is Needed!
Bonneville Power Reduces Cost & Accelerates Data Delivery with Informatica

**The Challenge**
- High data storage costs due to copying of data
- Set up many simple Web services that don’t address DI issues
- Application integration increased number of integration points
- Complex SQL – lots of rules and joins for data quality
- No way for analysts and developers to do data validation
- Need real-time data processing vs. 3 hour batch window

**The Solution**
- Common data access layer for both BI and SOA projects
- Data abstraction layer for insulating consuming applications
- Reusable data services layer for SQL and Web services
- Rapidly create a virtual data layer to reduce the 1-day reporting delay
- Data profiling for use by analysts and developers for data validation
- Physically move the virtual view into a DW without extra overhead

**The Benefits**
- Flexible data architecture that supports change
- Reduction in code with model and metadata drivel tools
- 3-10x reduction in cost of servers, duplication and staleness of data
- Leveraging existing Informatica resources - up to speed in 1 hour
- Faster data processing – from 4 hours to 15 minutes
- Adhere to changing regulations and quick identification of data issues
Next Steps & Q&A
Have the Conversation!

1. Identify a Critical Project in Your Company
2. Involve the Business Early & Often
3. Bake-In Quality & Support Advanced Logic
4. Demonstrate Business Value Early
5. Self-Service + Data Virtualization = ROI

New data & reports take too long...

“YOU” can now do it in DAYS!
Plan Your Roadmap

1. DW/Business Intelligence (BI)
   Prototype DW & accelerate new data & reports from months to days

2. MDM
   Deliver a complete view of master & transactional data in real-time

3. SOA
   Deliver the missing data services layer to SOA & applications

Applications
- BPM
- Biz. Services
- Data Abstraction

Data Sources
- Registry
- ESB
Get the Leading Data Virtualization Solution from the Leader in Data Integration

Gartner Magic Quadrant for Data Integration Tools, 2011

“The ability to switch seamlessly & transparently between delivery modes (bulk / batch vs. granular real-time vs. federation) with minimal rework will be key for IT organizations seeking to develop a successful data integration strategy.”

Ted Friedman, VP Distinguished Analyst, Gartner

Forrester Wave: Data Virtualization, Q1 ‘12

“With v9, Informatica advanced its capabilities with on-the-fly data quality & profiling, a model-driven approach to provisioning data services, performance enhancements, cloud integration, common metadata, & role-specific tools.”

The Forrester Wave: Data Virtualization, Q1 2012
Upgrade to the New PowerCenter Data Virtualization Edition

**BASIC EDITION**
FOR DW

**AGILITY & PRODUCTIVITY**

- PowerCenter Data Virtualization Edition
- Partitioning
- Data Profiling
- Data Services
- Developer Tool
- Analyst Tool
- 2 Adapters (PWX for Relational)

**REAL-TIME & OPERATIONAL DI**

- PowerCenter Real Time Edition
- 2 Adapters (PWX for Relational)

**DATA GOVERNANCE**

- PowerCenter Advanced Edition
- 2 Adapters (PWX for Relational)

**ETL**

INFORMATICA WORLD
Learn More…

Informatica.com > Products > PowerCenter > Data Virtualization Edition

Informatica.com > Products > Data Virtualization

Sign-Up

Expert Roundtables
Customer Webinars
Whitepapers, Articles, Blogs…

Data Virtualization Corner

Data Virtualization & Data Services Corner

Architect-to-Architect & Business Value Series

2012 Roundtable Series

Why Agile BI, Why Now and How Data Virtualization Can Help

- David Maloney, Informatica CTO
- Ignite 2012 Keynote Address
- Data virtualization enables agile BI and decision support
- JOIN & DISCUSS 2000+ Strong “Data Virtualization & Data Services Architecture” Group

Data Virtualization and Data Services Architecture

JOIN & DISCUSS 2000+ Strong “Data Virtualization & Data Services Architecture” Group

Google Data Virtualization and Data Services Corner

Data Virtualization Corner

Informatica World
Thank You