

Data Virtualization at Nationwide



Nationwide®

DAMA – October 13, 2011



Nationwide®
On Your Side

Agenda

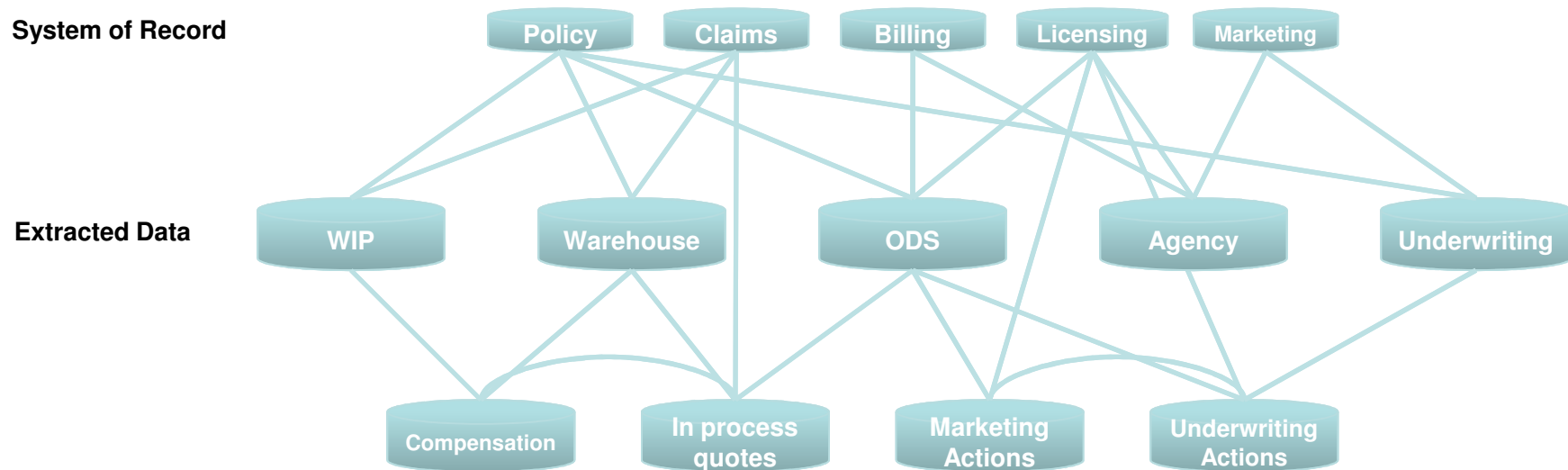
- Background
- What is Virtual Data – Isn't all data real?
- Virtual Data and the Architectural Fit
- Example Use Cases
- Must Do's Before Implementation

Background

- Many things drive the need to duplicate and or extract specific pieces of data from the authoritative data store.
 - Per Forrester, one of the key requirements that will drive increased adoption of virtual data approaches is the tactical demand for better data management and data integration alternatives to ETL or DBMS consolidation.
 - Data virtualization facilitates the abstraction, transformation, and federation of data from a variety of disparate sources.
 - Data is presented via a single access point to a consumer regardless of the physical location or nature of the various data sources. Some benefits include:
 - ❑ Greater flexibility and agility due to faster creation of virtual data stores.
 - ❑ Improved data quality because we reduce the number of physical copies, eliminating redundant data stores.
 - ❑ Improved usage through creation of subject-oriented, business-friendly data objects.
- Once we begin duplicating data, without a strong governance structure, we tend to loose control of the data and its meaning.

Oh what a mess we make

A look at traditional data movement



Data Virtualization – What is it?

Data virtualization is the process of abstracting, transforming, federating and delivering data contained within a variety of information sources so that they may be accessed by a consuming application or users when requested without regard to their physical storage or heterogeneous structure.

This concept and software is commonly used within data integration, business intelligence, service-oriented architecture data services, cloud computing, enterprise search and master data management.

Key Characteristic

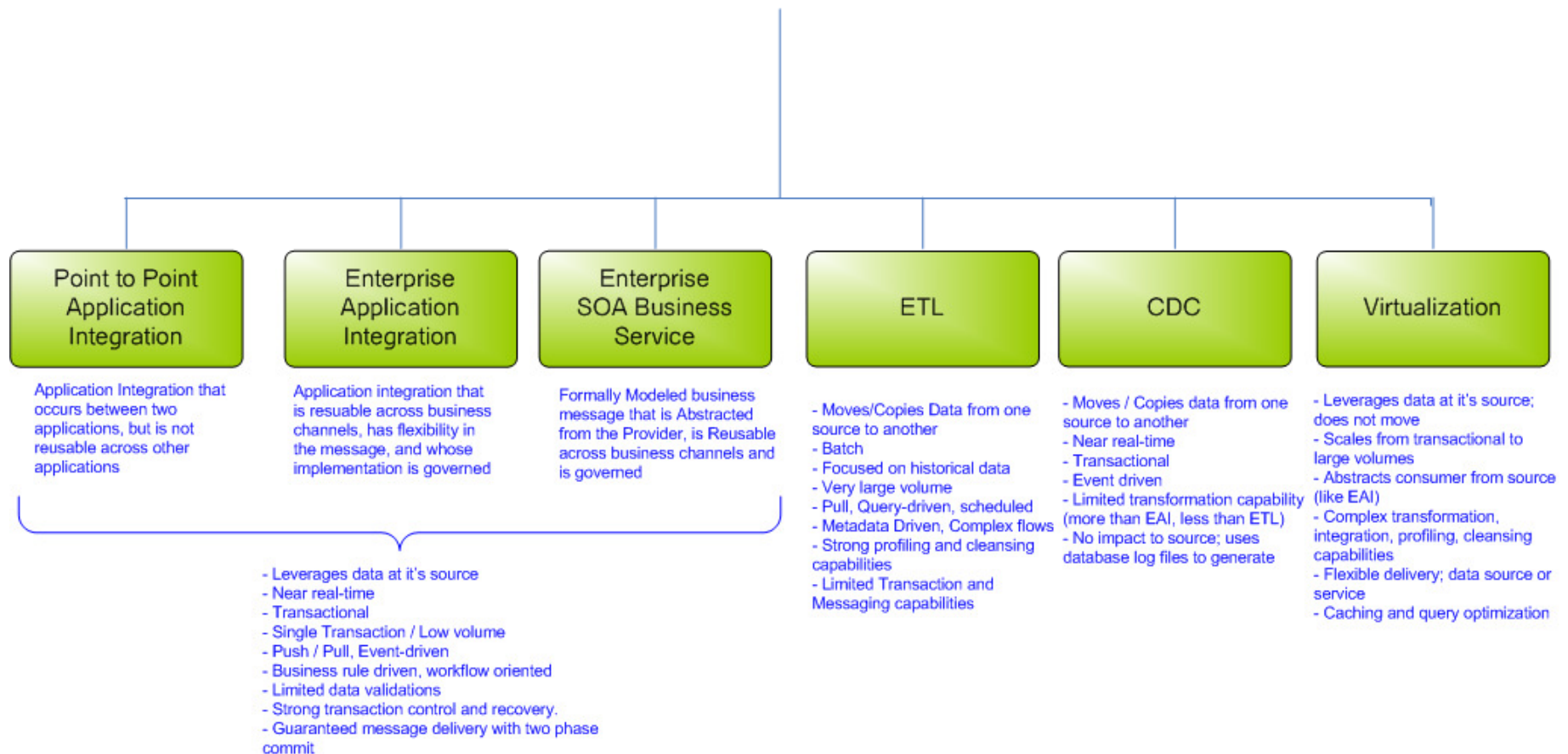
It keeps the data where it belongs (persisting data in its native data store as opposed to relocating it) . It creates a façade and exposes it to consumers, but doesn't move it. It can be used in read-only or read/write scenarios.

So, how is this different than a SOA implementation?

Integration Landscape

High Level Integration Implementation Patterns

I am a consumer application with a need to integrate with another application



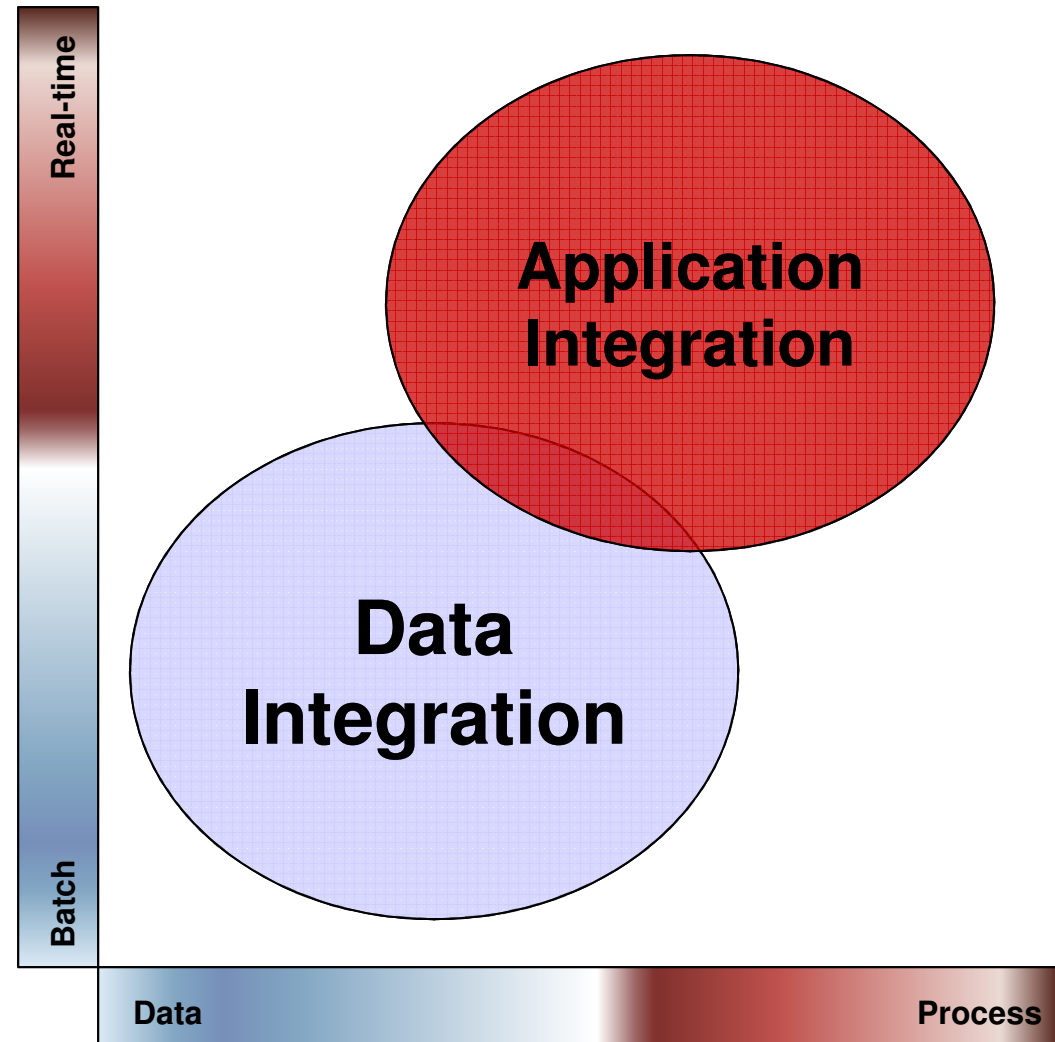
Data Integration and Application Integration

Data Virtualization tools are Data Integration tools that provide functionality similar to Enterprise Application Integration.

That overlap is the gray area where confusion arises around when to use which integration tool.

Other tools in this gray area are Change Data Capture and Informatica Web Service Gateway.

Because of this gray area, it's difficult to provide clear cut prescriptive guidance.



Example Use Case – Single Underwriting Desktop

- Quick delivery timelines: POC was conducted in 4 days
- Multiple data source types: IMS, Teradata, DB2, etc.
- Some large volume data sources (ie, Teradata)
- Users hypersensitive to response time challenges
- 11 Data Sources
- 2-3 Attributes per source required
- Read-Only requirement

The screenshot displays the Nationwide Underwriting Desktop interface. At the top left is the Nationwide logo with the tagline "On Your Side". The top right corner shows "User ID:" followed by links for "Feedback", "Help", and "Switch Agency", and a "READ ONLY" status indicator. Below this is a "Producer:" dropdown menu set to "340000062(Agency)".

The main dashboard consists of six panels:

- Top Left Panel (Yellow):** A table with columns "Name", "Date", and "Description". It contains the text "Click 'Search/Reassign' to view all follow ups." and a "Search/Reassign>>" link.
- Top Middle Panel (Yellow):** A table with columns "Name" and "Description". It lists several names with their corresponding "Cross Sell" types: GRASSO, JOSEPH (Auto); MCCULLOUGH, STEVEN (Auto); BARR, SAM (Auto); TAYLOR, VICKIE (Simple Life); MILLER, WAYNE (Simple Life); KITCHEN, LYNETTE (Simple Life); ZELLERS, JAMES (Simple Life); WEST, CINDY (Simple Life). It includes a "Create Offers List>>" link.
- Top Right Panel (Yellow):** A table with columns "Name" and "Description". It contains the text "No high priority alerts exist. Click 'Create List' to view all alerts." and "Create Alerts List>>" link.
- Bottom Left Panel (Green):** A search form titled "Choose a search method" with radio buttons for "Customer", "Policy", and "Prospect". It has input fields for "Last Name:" and "First Name:", a "Submit" button, and a note: "Asterisk (*) indicates required field."
- Bottom Middle Panel (Blue):** A panel titled "Marketing Storefront" with sub-sections: "Moments Of Truth:" (links to "On Your Side Policy Reviews" and "Premium Change List"); "Reports:" (link to "Agent Gateway Marketing Reports"); and "Training:" (link to "Agent Gateway Marketing Training").
- Bottom Right Panel (Yellow):** A table with columns "Prospect Name" and "Description". It contains the text "No prioritized offers exist."

Single Underwriting Desktop R1 Compare

Release Integration Needs

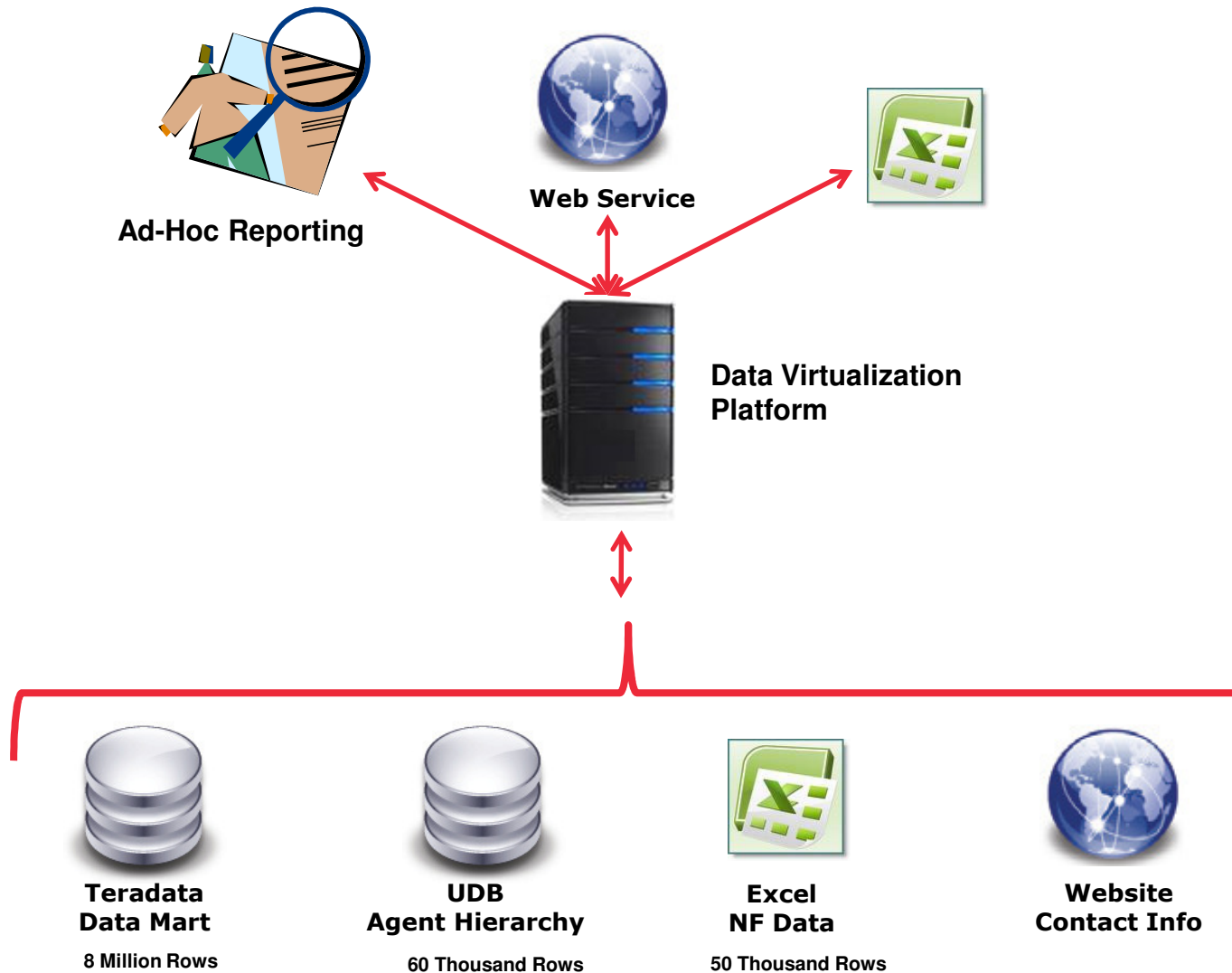
- Agency Data: Need agency contact and contract information
- Agency Performance Metrics: Need agency profitability and growth metrics as well as sales territory metrics
- Underwriting Organization: Need underwriter to agency assignments and an underwriter's manager
- Sales Organization: Need sales manager to agency assignments and sales manager phone number

Release Integration Approach

- Agency Data: Data is retrieved from the agent source application once a day and stored locally in the database. Agency data is flushed and refreshed daily
- Agency Performance Metrics: Report extracts from data warehouse are stored in the database via an ETL process in order provide point in time views of the agencies performance
- Underwriting Organization: Underwriter assignment to an agency is extracted from the source application for Allied and a spreadsheet for Nationwide. The underwriters manager is retrieved from call-up. Underwriting organization data is refreshed daily
- Sales Organization: Sales manager assignment is extracted from the source application, and the sales manager's phone number is retrieved from call-up. Sales organization data is refreshed daily

At least four of the five integration points listed were definitive candidates for data virtualization. The 5th source would be dependent on negotiation of non-functional requirements around availability.

Example Use Case – Business Intelligence



Performance Was Met or Exceeded

Category	Configuration
Server	Windows 2008 R2 Standard
Processor	interl® Xeon® CPU L5530 ® 2.40GHz 2.40GHz (2 processors)
Installed Memory	3.00 GBS
System Type	64-bit Operating System

Rows processed	Execution time
join: 142,162,589,; projection: 141,287,438	1 hour and 21 minutes and 4 seconds.
projection: 237,411,854	3 hours and 51 minutes
join: 8,383,234; projection: 5,102,815	10 minutes and 53 seconds.
4,297,869	3 minutes
128,130,455	1 hour and 25 minutes
Netezza: 260,634 ; Teradata: 4358057 ; join: 115319 ; projection: 115319	2 minutes and 23 seconds
Netezza: 2309123; Teradata: 7885758 ; join: 82524 ; Projection: 82524	2 minutes and 58 seconds

Must Do's Prior to Implementation – Governance

- Very difficult to provide definitive prescriptive guidance
- Characteristics from earlier should drive general direction, but each case will have to be evaluated for “right fit”
- Must understand roles and responsibilities around source access – this can cause performance issues on the source if not careful! This needs to be balanced with Guiding Principles around “Data is Accessible”.
- Writes – when should this be allowed and when not. Very powerful, but need to understand possible implication to source business rules