DataFlux DataVision Technology Demonstration

Joe Rademacher
DataFlux Senior Solutions Engineer
Clay Harter
OpenSpirit Chief Technology Officer

Agenda

- Presenters
  - Joe Rademacher, DataFlux Senior Solutions Engineer
  - Clay Harter, OpenSpirit Chief Technology Officer
- Why does Data Quality matter?
- How do I Identify Data Quality Issues?
- How do I improve on the reliability and accuracy of my data?
- How do I keep my data clean after I've corrected it?
- How do I connect to my Geotechnical applications and databases?
- Brief Demo
- Q&A

DataFlux
A sas COMPANY

What we're hearing: Broad domain of activities

- Today's large, integrated oil and gas companies face long term data management obstacles arising from the broad spectrum of activities involved in:
  - Subsurface (understanding and managing reservoirs)
  - Surface (construction and management of facilities)
  - The Business Domain (back office systems supporting many other functions.
- Short term solution requirements must be weighed against more flexible, longer term solutions which will fully support the environment as assets mature over time.

DataFlux
A sas COMPANY

More of what we're hearing…

- … looking for Data Quality solutions to empower business partners and enable more efficient decision making.
- … must be able to be leveraged by key stakeholders to improve process and DQ in the Drilling, Operations and Production focus areas.
- Ultimately, it should be an Enterprise Solution covering all business lines.
- ROI –beyond time saved and productivity gains via the DQ solution, real value measured by increased speed of identifying and resolving actual business issues

DataFlux
A sas COMPANY

Some of the drivers…

- Compliance: regulatory reporting of pipeline assets. Fines being levied in the multi-millions.
- System Conversions: Numerous refineries to be converted over next 5 years
- Product Information Management (PIM): match/merge vendor files, invoices (shorten AR cycle), etc.
- Employee Master Data Management: employee data access and consolidation of HR systems
- Vendor Management cost soaring

DataFlux
A sas COMPANY

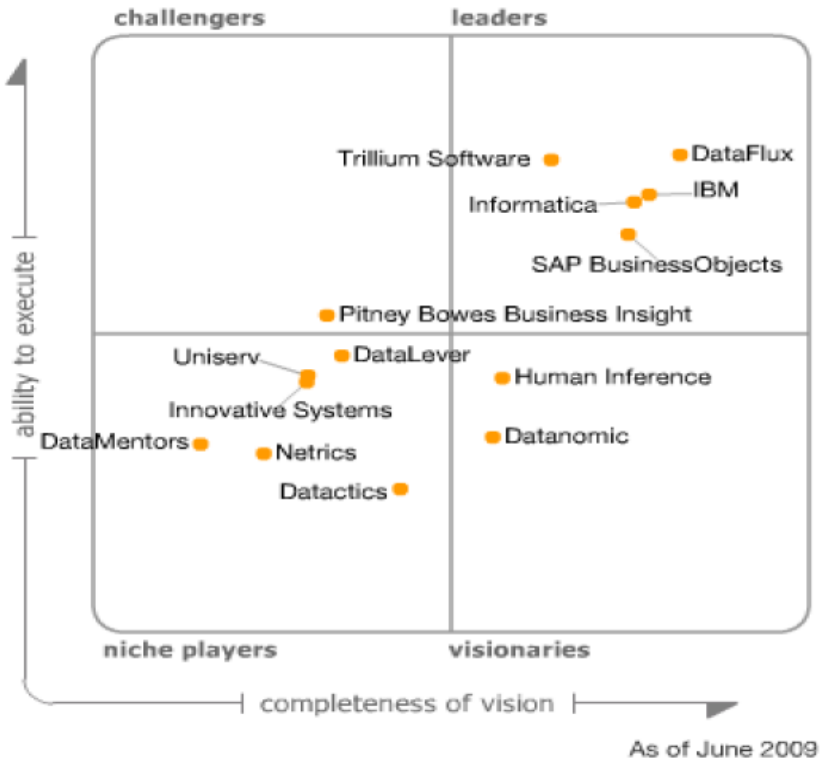
DataFlux Differentiators

- Recognized Market Leader
  - Customer References / Analyst Endorsement
- Solution Ease of Use
  - Designed for business analysts – Consultants quickly productive
  - Single integrated solution – all modules built by DataFlux (no acquisitions)
- Applicable Across Master Data Types
  - Customers, products, financials, HR, etc.
- Flexible / Extensible / Proven integration approach
  - SOA Architecture
- Client Growth Path – Flexible Starting Points
  - Start with Core Platform, grow to Accelerators and/or CDI
- SAS Subsidiary
  - R&D Depth, Worldwide Reach

DataFlux
A sas COMPANY

Gartner: Data Quality Magic Quadrant 2009

Figure 1. Magic Quadrant for Data Quality Tools



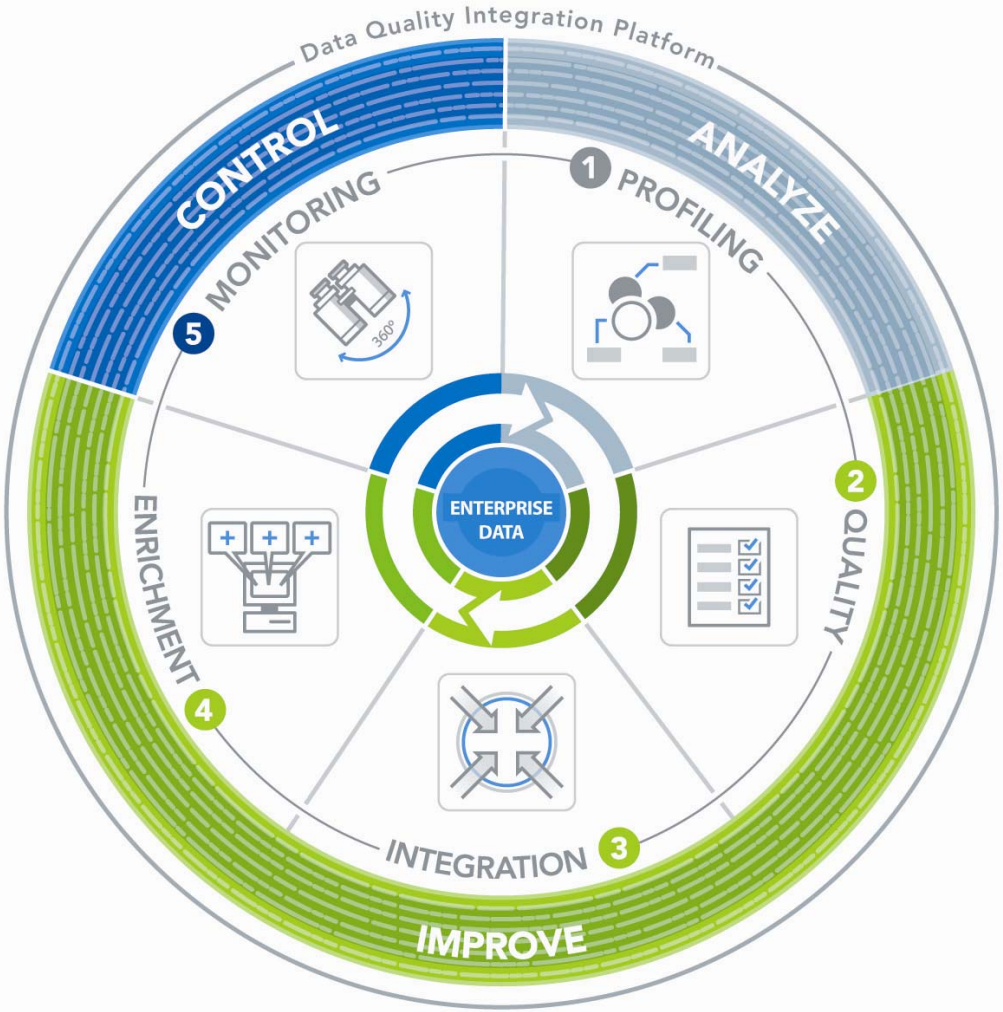
Source: Gartner (June 2009)

The Magic Quadrant is copyrighted 2009 by Gartner, Inc. and is reused with permission. The Magic Quadrant is a graphical representation of a marketplace at and for a specific time period. It depicts Gartner's analysis of how certain vendors measure against criteria for that marketplace, as defined by Gartner. Gartner does not endorse any vendor, product or service depicted in the Magic Quadrant, and does not advise technology users to select only those vendors placed in the "Leaders" quadrant. The Magic Quadrant is intended solely as a research tool, and is not meant to be a specific guide to action. Gartner disclaims all warranties, express or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

Gartner, Inc. "Magic Quadrant for Data Quality Tools" by Ted Friedman and Andreas Bitterer. June 2009.

DataFlux
A sas COMPANY

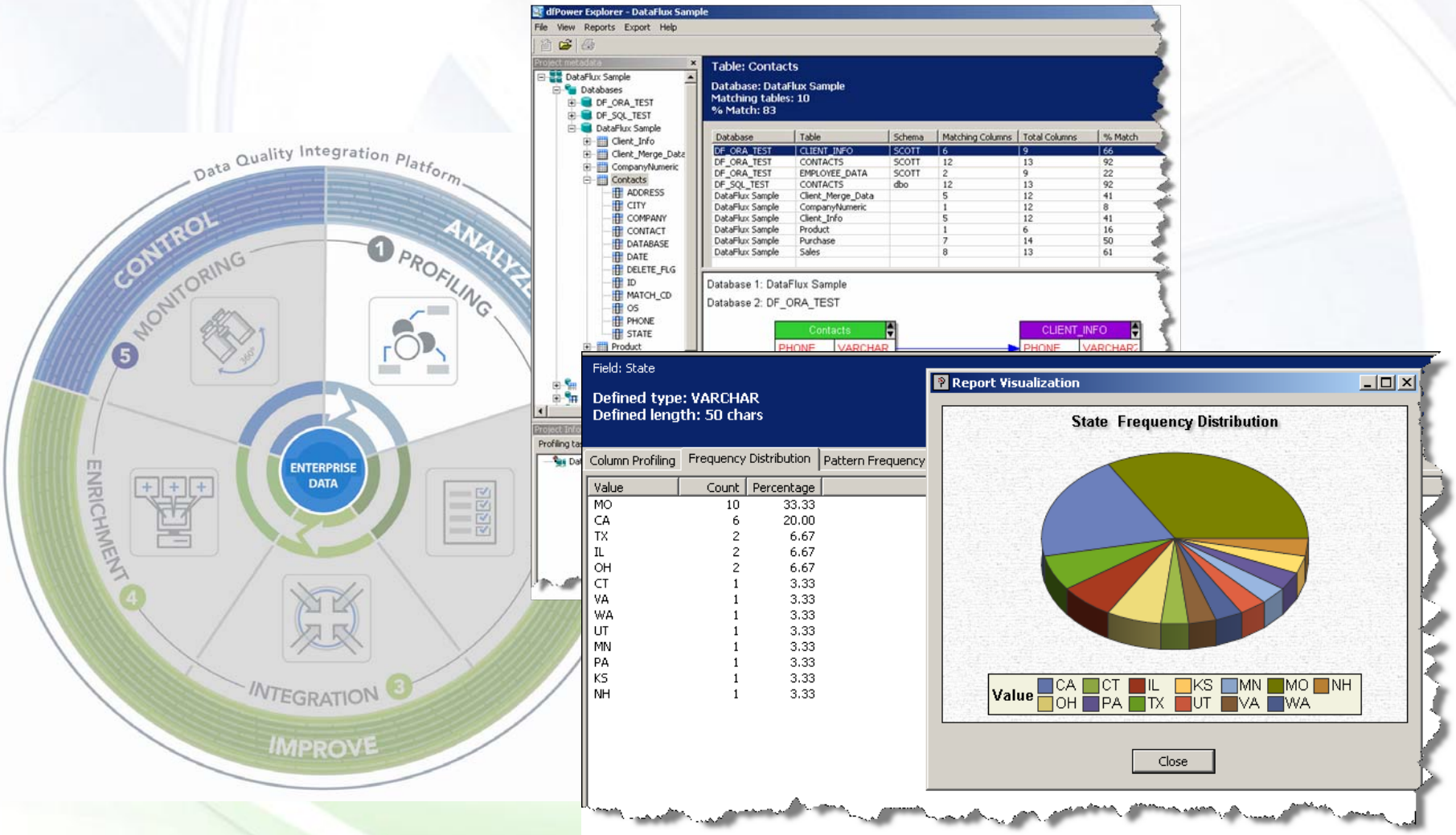
Data Quality Integration Platform



DataFlux
A sas COMPANY

DataFlux Methodology: Analyze

Metadata Profiling and Data Profiling

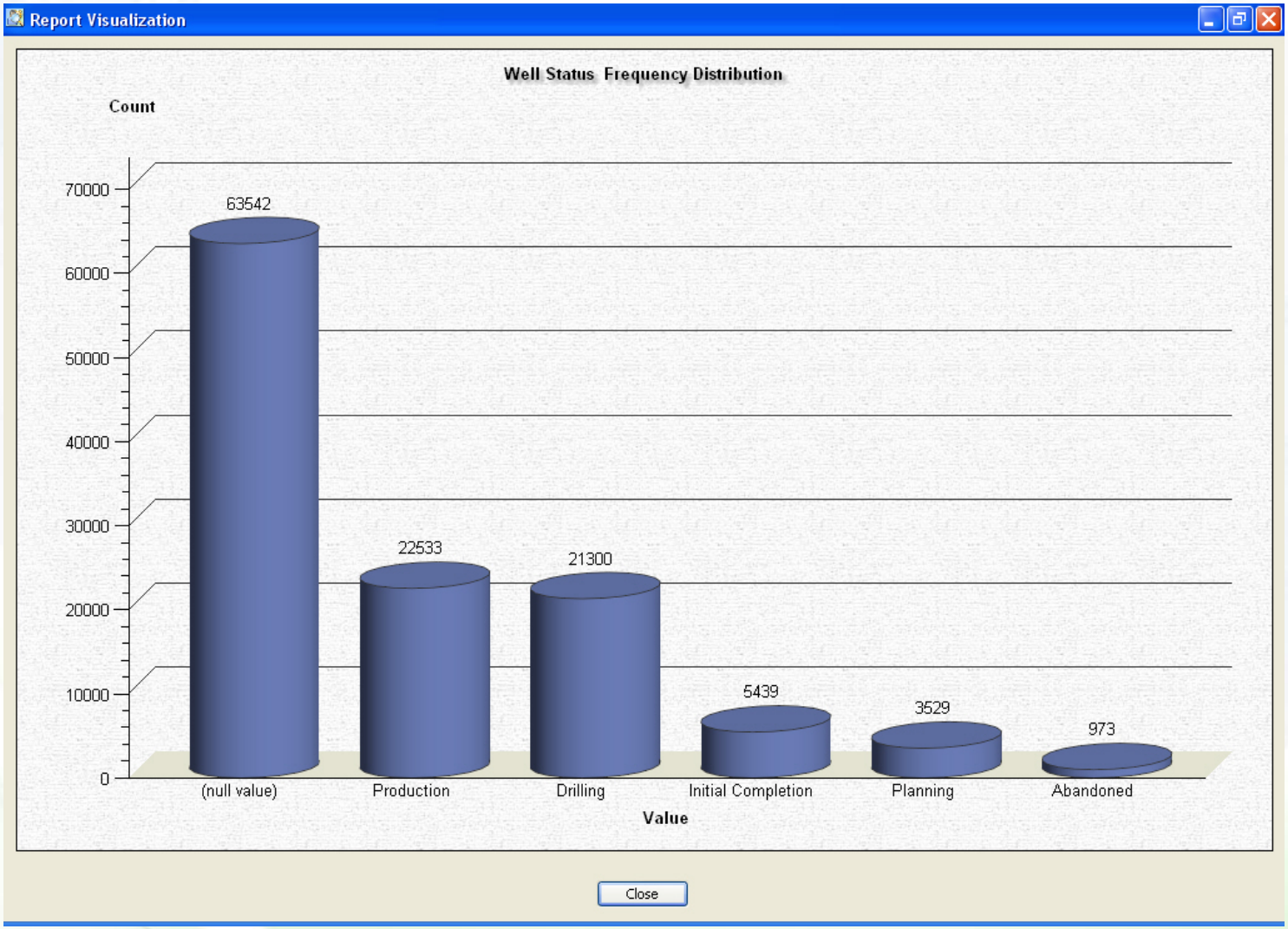


DataFlux A sas COMPANY

Discovery of Data Issues

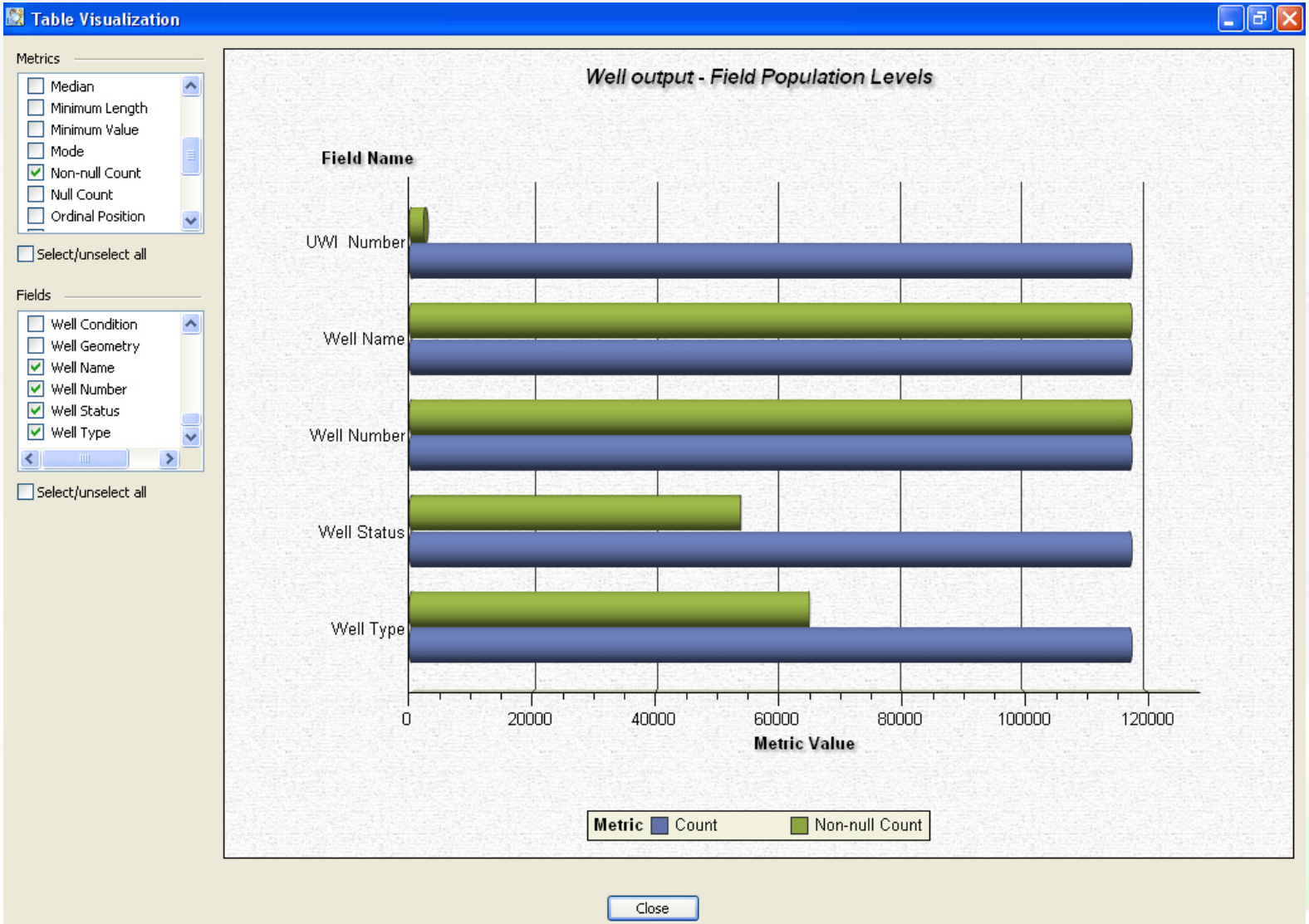
- Auto Profiling of tables to enable discovery of DQ issues
- The first step in Data Analysis should be included in standard Upstream processing.
- Quite often is the foundation for decisions around Data Models, Data Quality Standardizations, Cleansing Routines, and ongoing Data Quality Monitoring
- Needs to be presented in an easy to use interface that allows interaction with the profiling results, and quickly sheds light on anomalies in the data.
- Cross-Source data validations

Visualizations



DataFlux
A sas COMPANY

Field population levels



DataFlux
A SAS COMPANY

DataFlux Methodology: Improve

The improvement phase includes processes for correcting, consolidating and enriching data.

- Three components:
  1. Quality
  2. Integration
  3. Enrichment

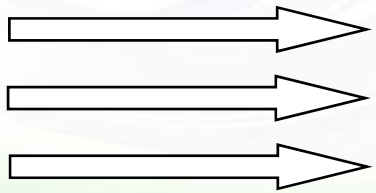
DataFlux
A sas COMPANY

Creation of a Customer Rollup

Derived Name	ERP Customer or I.H.S Operator Name
CONOCOPHILLIPS	MERIDIAN RESOURCES USA INCORPORATED
CONOCOPHILLIPS	MERIDIAN RES. & EXPLOR. LLC, THE
CONOCOPHILLIPS	CONOCOPHILLIPS-PARENT
CONOCOPHILLIPS	CONOCOPHILLIPS COMPANY
CONOCOPHILLIPS	CONOCO INCORPORATED
CONOCOPHILLIPS	BURLINGTON RESOURCES OIL & GAS CO - MIDL
CONOCOPHILLIPS	BURLINGTON RESOURCES INC
CONOCOPHILLIPS	BURLINGTON RESOURCES

And can also assign values based upon previously defined rules

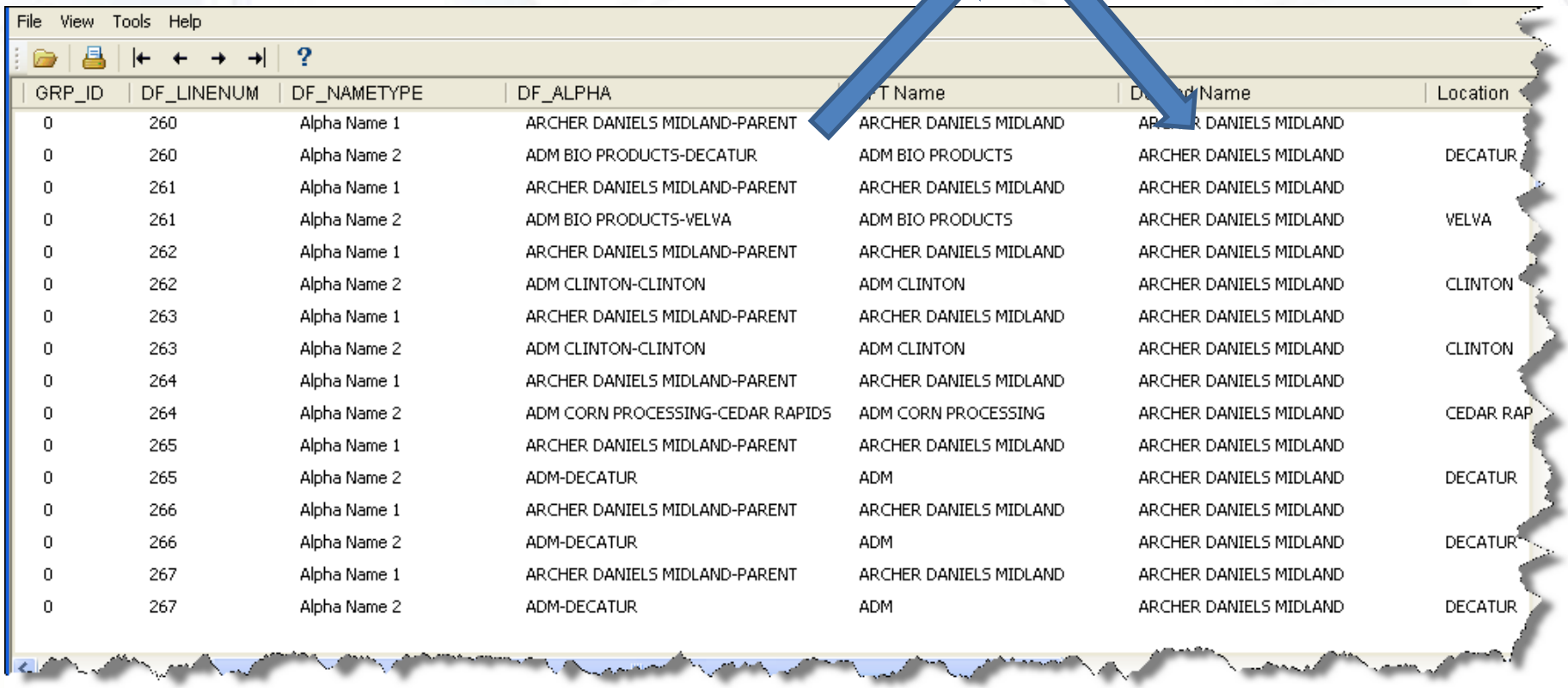
County	State
Baldwin	Alabama
Graham	Kansas
Grant	Kansas



Area
Louisiana - Land (south)
Oklahoma / Kansas
N. Texas - Panhandle

DataFlux
A sas COMPANY

Deriving standardized names from within the data



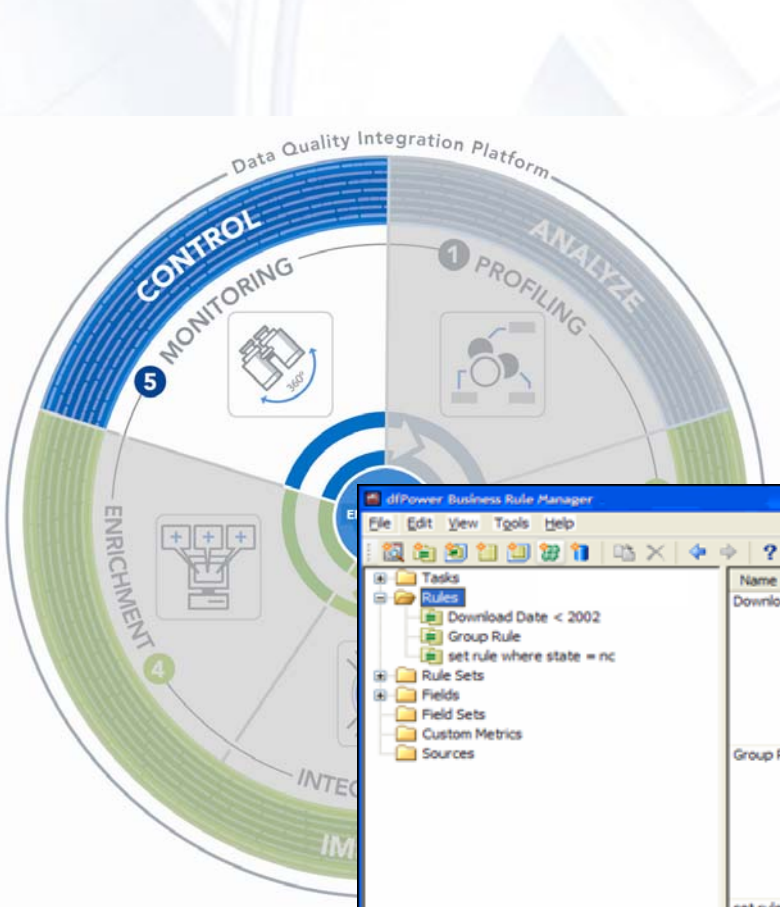
File View Tools Help

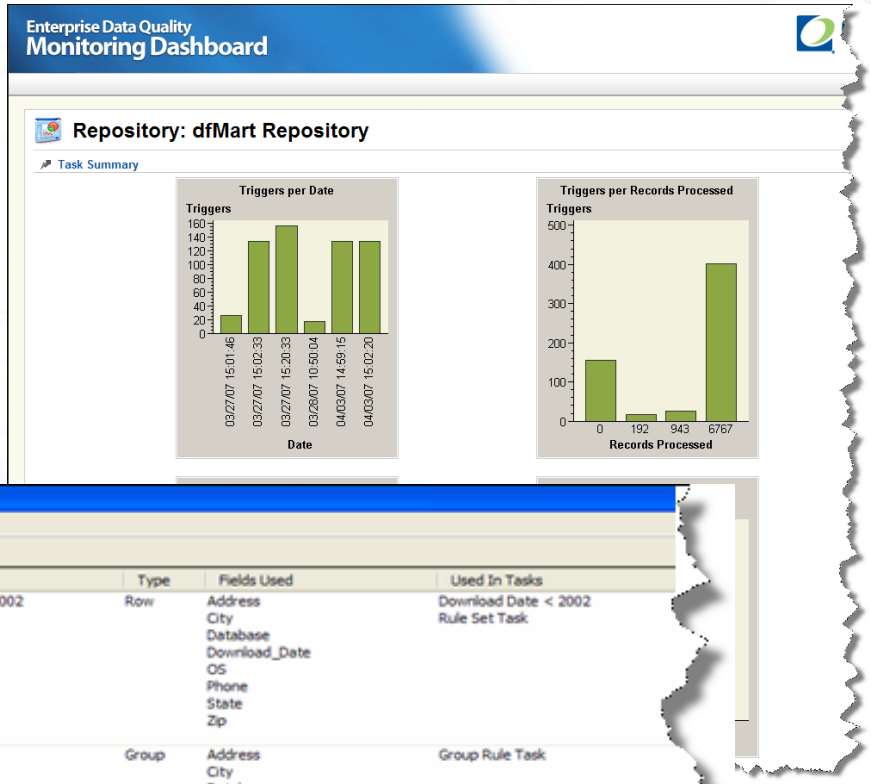
GRP_ID	DF_LINENUM	DF_NAMETYPE	DF_ALPHA	FT Name	D Name	Location
0	260	Alpha Name 1	ARCHER DANIELS MIDLAND-PARENT	ARCHER DANIELS MIDLAND	AR R DANIELS MIDLAND	
0	260	Alpha Name 2	ADM BIO PRODUCTS-DECATUR	ADM BIO PRODUCTS	ARCHER DANIELS MIDLAND	DECATUR
0	261	Alpha Name 1	ARCHER DANIELS MIDLAND-PARENT	ARCHER DANIELS MIDLAND	ARCHER DANIELS MIDLAND	
0	261	Alpha Name 2	ADM BIO PRODUCTS-VELVA	ADM BIO PRODUCTS	ARCHER DANIELS MIDLAND	VELVA
0	262	Alpha Name 1	ARCHER DANIELS MIDLAND-PARENT	ARCHER DANIELS MIDLAND	ARCHER DANIELS MIDLAND	
0	262	Alpha Name 2	ADM CLINTON-CLINTON	ADM CLINTON	ARCHER DANIELS MIDLAND	CLINTON
0	263	Alpha Name 1	ARCHER DANIELS MIDLAND-PARENT	ARCHER DANIELS MIDLAND	ARCHER DANIELS MIDLAND	
0	263	Alpha Name 2	ADM CLINTON-CLINTON	ADM CLINTON	ARCHER DANIELS MIDLAND	CLINTON
0	264	Alpha Name 1	ARCHER DANIELS MIDLAND-PARENT	ARCHER DANIELS MIDLAND	ARCHER DANIELS MIDLAND	
0	264	Alpha Name 2	ADM CORN PROCESSING-CEDAR RAPIDS	ADM CORN PROCESSING	ARCHER DANIELS MIDLAND	CEDAR RAP
0	265	Alpha Name 1	ARCHER DANIELS MIDLAND-PARENT	ARCHER DANIELS MIDLAND	ARCHER DANIELS MIDLAND	
0	265	Alpha Name 2	ADM-DECATUR	ADM	ARCHER DANIELS MIDLAND	DECATUR
0	266	Alpha Name 1	ARCHER DANIELS MIDLAND-PARENT	ARCHER DANIELS MIDLAND	ARCHER DANIELS MIDLAND	
0	266	Alpha Name 2	ADM-DECATUR	ADM	ARCHER DANIELS MIDLAND	DECATUR
0	267	Alpha Name 1	ARCHER DANIELS MIDLAND-PARENT	ARCHER DANIELS MIDLAND	ARCHER DANIELS MIDLAND	
0	267	Alpha Name 2	ADM-DECATUR	ADM	ARCHER DANIELS MIDLAND	DECATUR

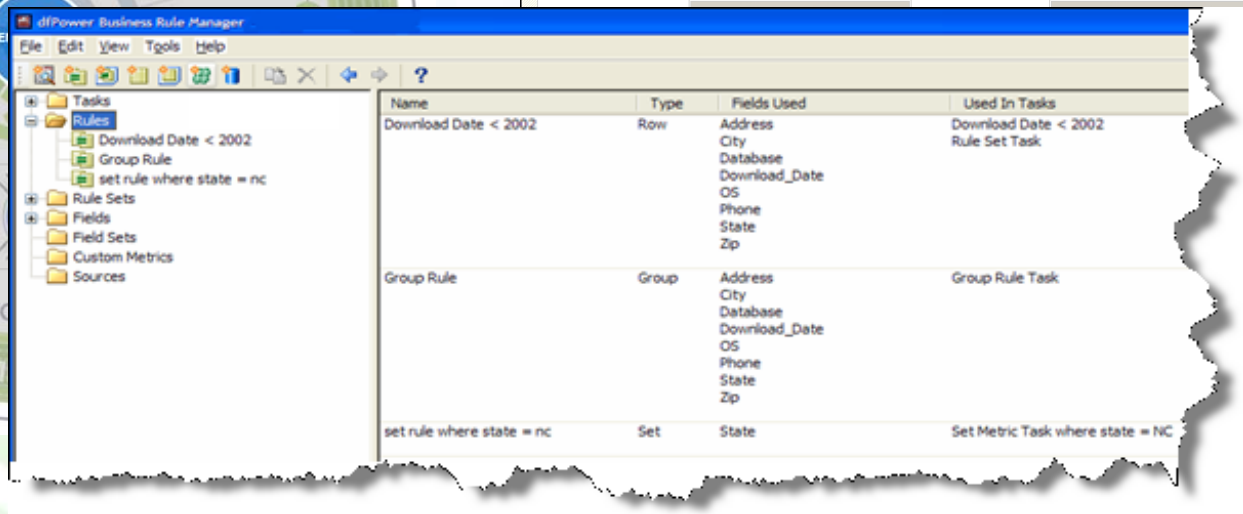
DataFlux
A sas COMPANY

DataFlux Methodology: Control

Business Rules Monitoring



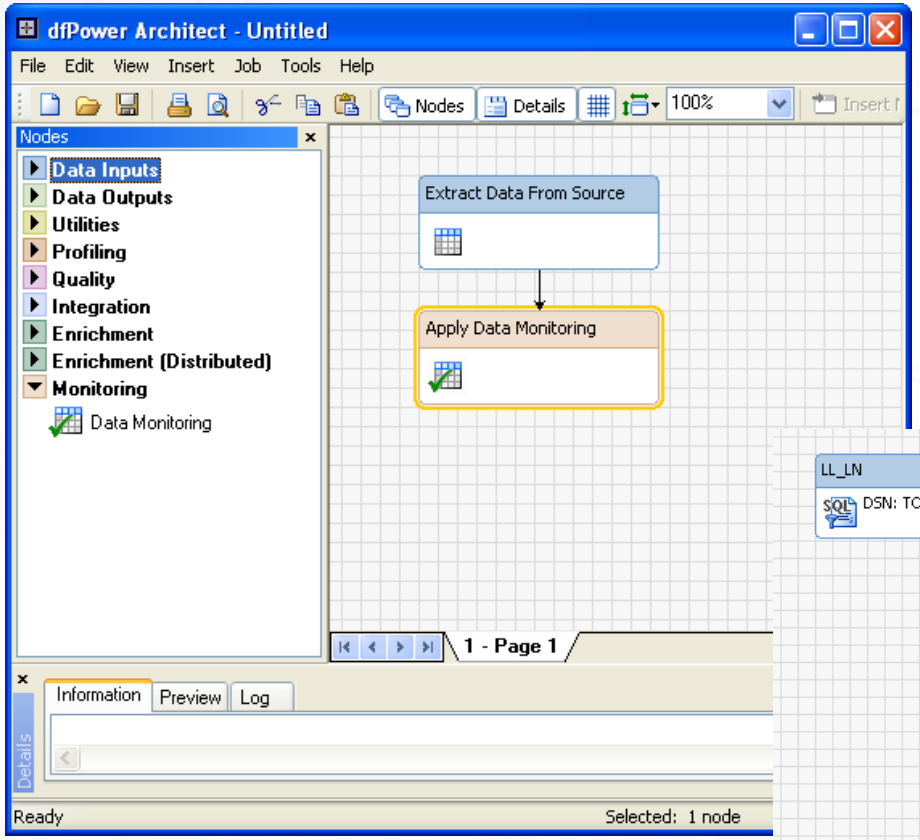


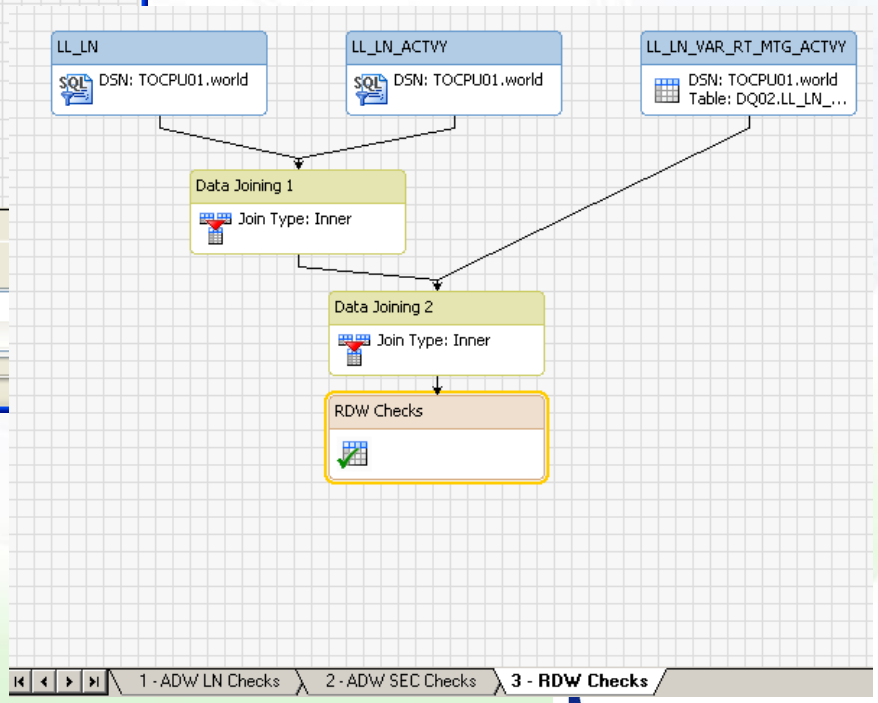


DataFlux
A sas COMPANY

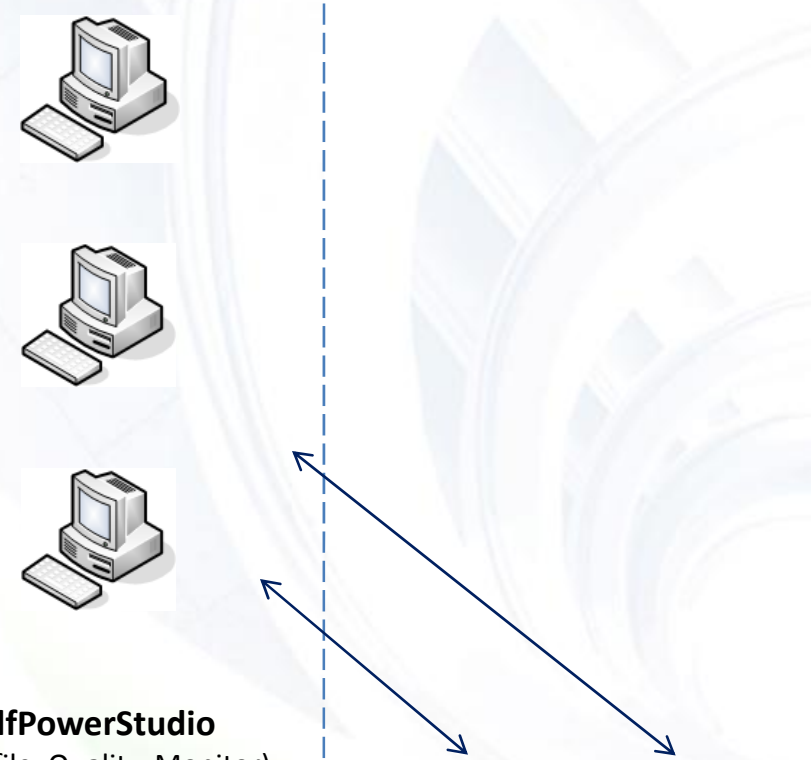
Flexible Deployment models

Including Batch & SOA for application integration





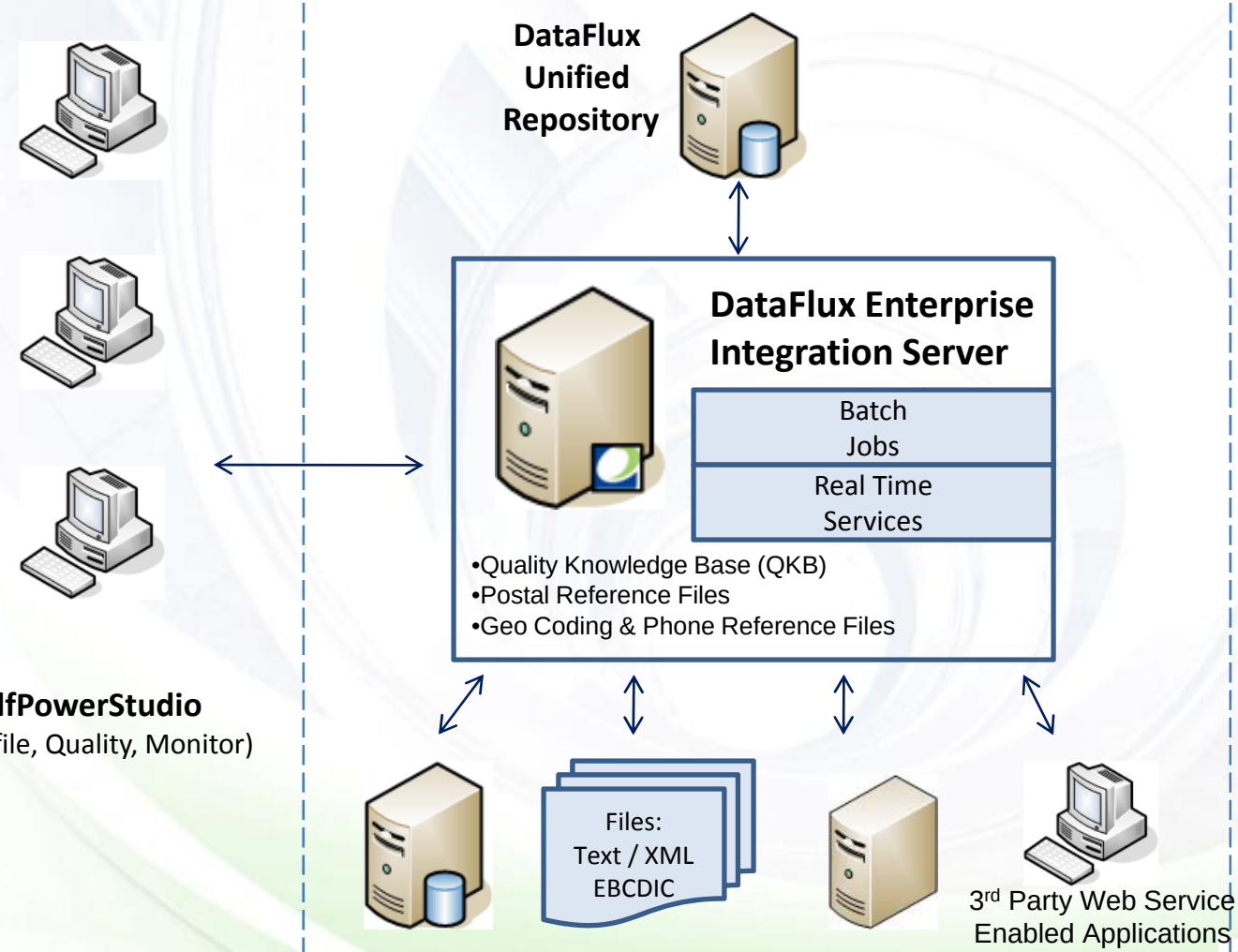
Design



DataFlux
A sas COMPANY

Design

Run



DataFlux
A sas COMPANY

Design

Run

Display

dfPowerStudio
(Profile, Quality, Monitor)

DataFlux Unified Repository

Java Application Server

DataFlux Enterprise Integration Server

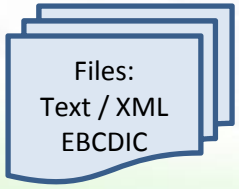
Batch Jobs	Accelerator Jobs
Real Time Services	Accelerator Services

- Quality Knowledge Base (QKB)
- Postal Reference Files
- Geo Coding & Phone Reference Files

Web Based Interfaces and Reports:

- *Master Data Management*
- *Customer Data Analysis*
- *Watch List Compliance*
- *Materials Classification*
- *Commodity Coding*
- *Customer Data Control*

RDBMS



Enterprise Applications:
SAP, Oracle, CRM

3rd Party Web Service Enabled Applications

DataFlux
A SAS COMPANY

Challenges for Integration with Geotechincal Applications and Data

- Diverse Applications
  - Varying vendors
  - Varying operating systems, computer languages
  - User interface inconsistencies
- Diverse Data
  - Different storage mechanisms
    - Propriety file formats, various RDBMS
    - Proprietary data access API tied to specific language and operating system
  - Varying data models
  - Varying units and coordinate system designations
  - Data types with special performance issues (e.g. seismic trace data)

DataFlux
A sas COMPANY

OpenSpirit Enables Multi-Vendor Workflows

Multi vendor, cross discipline, user defined workflows

Used by 40+ vendors, 250+ oil companies in 60+ countries

Partners

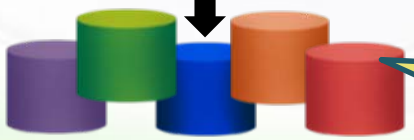
Subsurface Interpretation | GIS | Data Management | Well Planning | Other

Proprietary

OpenSpirit Multi-Vendor Tools

OpenSpirit Integration Framework

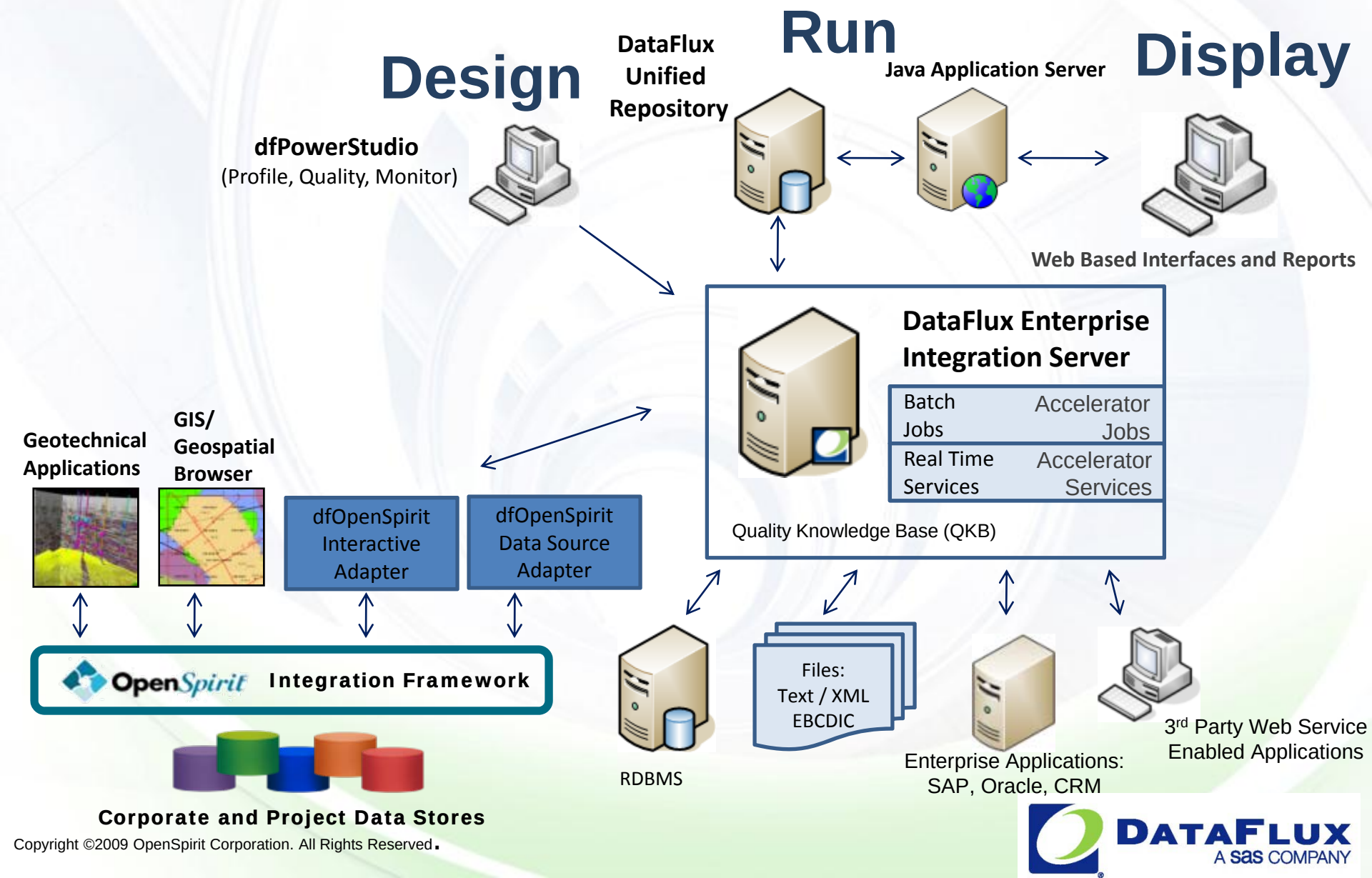
SOA Services
- Data
- Unit
- Coordinate
- Event
- Metamodel



Corporate and Project Data Stores

OpenWorks PPDM
GeoFrame Managed SEGY
Kingdom Recall
Petra EPOS(Geolog)
Finder Seabed

DataFlux A sas COMPANY



DataFlux
A SAS COMPANY

dfOpenSpirit Interactive Adapter

- What it does it do?
  - Allow user to select DataFlux job to run in response to different data types selected
  - Listen for OpenSpirit data selection events (sourced from GIS map, Data Selector, other OpenSpirit enabled applications) and then trigger DataFlux job
- Typical Usage
  - User selects wells from map and runs quality check (e.g. kb exists, deviation survey consistent with td, etc…) before sending to G&G application
  - User selects well picks in Data Selector and runs quality check (e.g. names consistent with standard list and no crazy isochors implied) before sending to G&G application

Status: Successful prototype implemented

DataFlux
A sas COMPANY

dfOpenSpirit DataSource Adapter

- What does it do?
  - Allow user to select one or more OpenSpirit data sources (project or master datastore) from inside dfPowerStudio and optionally assert desired coordinate and unit system
  - Also allows access to array data (e.g. deviation surveys, checkshots, well logs, etc..)
  - Then use OpenSpirit accessed data in any DataFlux job
- Typical Usage
  - Batch runs of profiling or quality jobs
  - Project comparisons (e,g. match wells or seismic lines between projects based on fuzzy name matches and spatial location)

Status: Preliminary development scoping done. Soliciting additional customer input

DataFlux
A sas COMPANY

Technology Demonstration

DataFlux
A sas COMPANY

DataFlux
A sas COMPANY