

SAS DI/DM Roadmap 2011

Thomas Rocks
Competence Center Data Integration
SAS Deutschland

Agenda

- Data Integration Studio 4.3
- Data Management Platform 2.1
- Ausblick/Roadmap DI
- Modernisierung: Mehrwerte für SAS Base Kunden

**SAS Data Integration Studio
wird nicht durch
Dataflux Data Management Studio
ersetzt!**



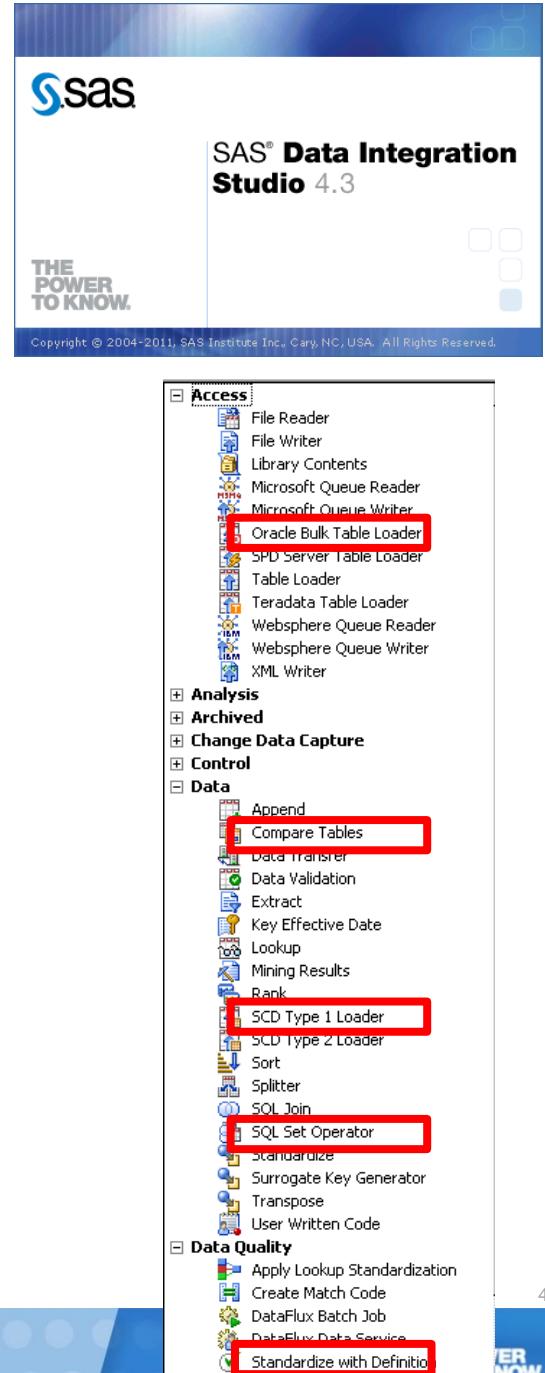
SAS Data Integration Studio 4.3

■ Bedienerfreundlichkeit & Handhabung

- Versionierung
- Jobstatus und Performance Reporting
- Erweiterung des SAS Code Importers
- Erweiterte Metadatensuche
- Source Code Deployment im Batch
- Erweiterter Grid Support

■ Transformationen

- ELT & DBMS Pushdown Support
 - SQL Union
 - Expression Builder
 - » Import von User Defined Functions (UDF)
 - » Erweiterte DBMS Unterstützung
- Oracle Bulk Table Loader
- Tabellen Vergleich um Änderungen zu erkennen
- Type 1 SCD
- Neue Datenquellen:
 - Greenplum
 - Aster Data
 - Sybase IQ



Versionierung

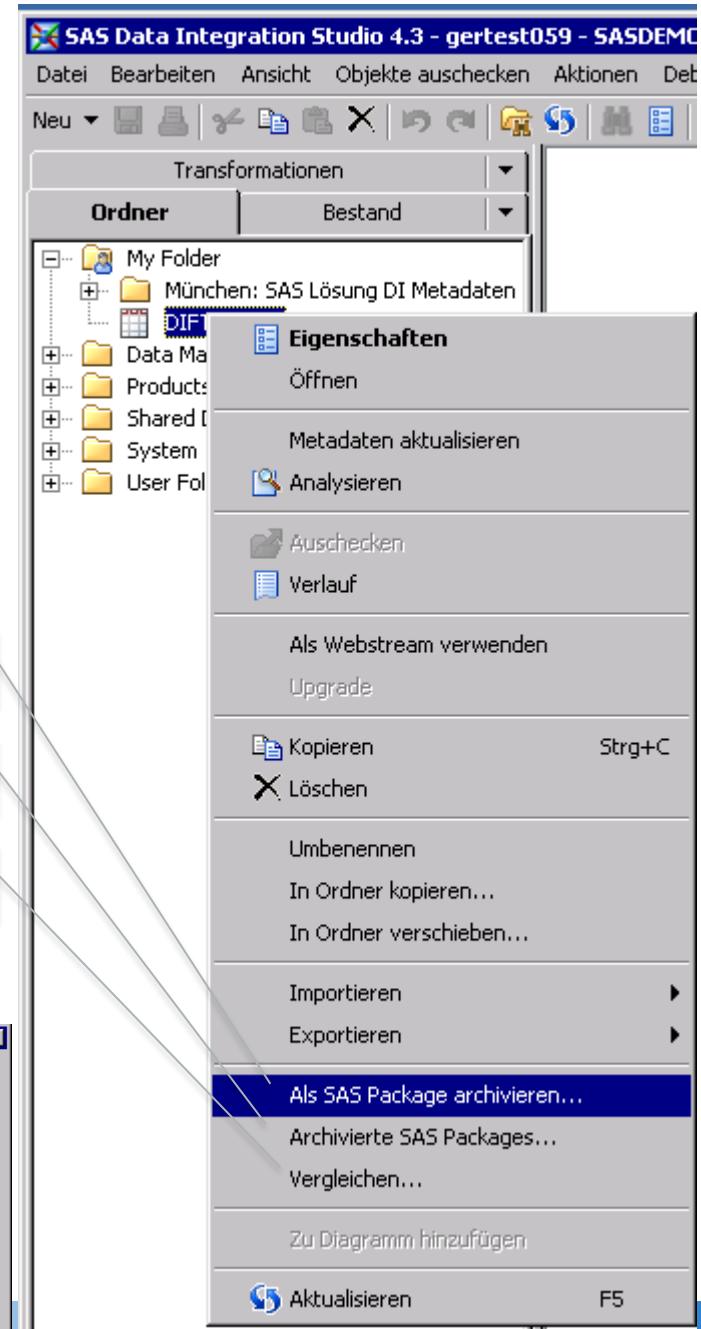
- 3rd Party Source Code Management Systeme:
 - CVS
 - SubVersion
 - Erweiterbar für weitere Produkte durch offen API

Archivieren

Navigieren

Vergleichen

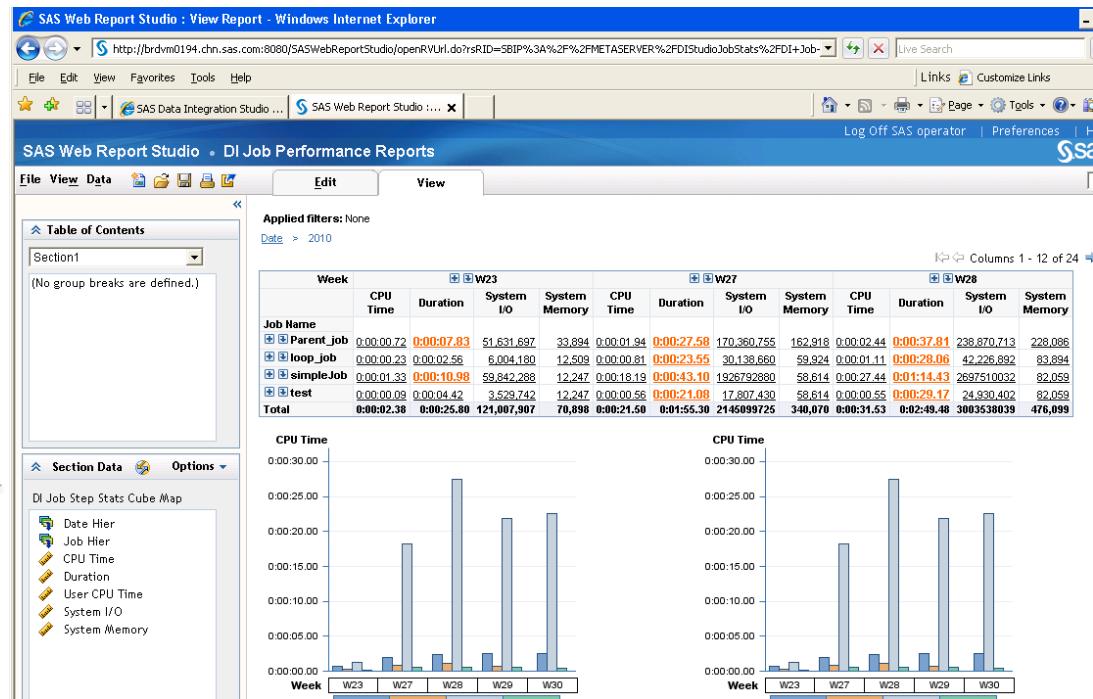
| Archivierte SAS Packages | | | | | |
|--------------------------|---------|----------------|---------------------|------------------|------------------------------------------|
| Anzeigen: Alle Packages | | | | | |
| SAS Packages: | | | | | |
| Package-Name | Version | Beschreibung | Datum und Zeit | Archiviert von | Package-Inhalt: |
| Initiale Release Mü... | 1.1 | Erste Freigabe | 13.04.2011 09:30:12 | sasdemo@GERTE... | Initiale Release München User Folders |
| test | 1.1 | | 11.04.2011 16:02:50 | sasdemo@RDCES... | |



Performance Monitoring

- Web Report Studio
- Stored Processes

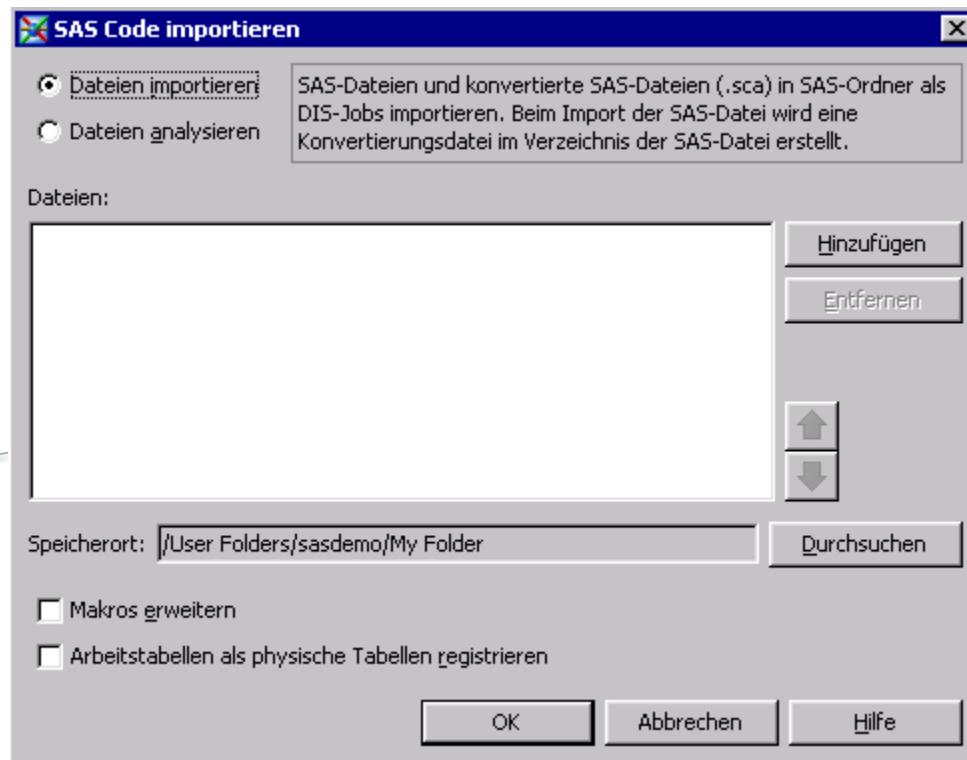
Reports für
Batch Processing



SAS Code Importer

- Macro Support
- Analyse Modus
- Kollisionserkennung (Library-Konflikte)
- Erweiterte Protokollierung
- Registrierung von Worktables als Tabellen

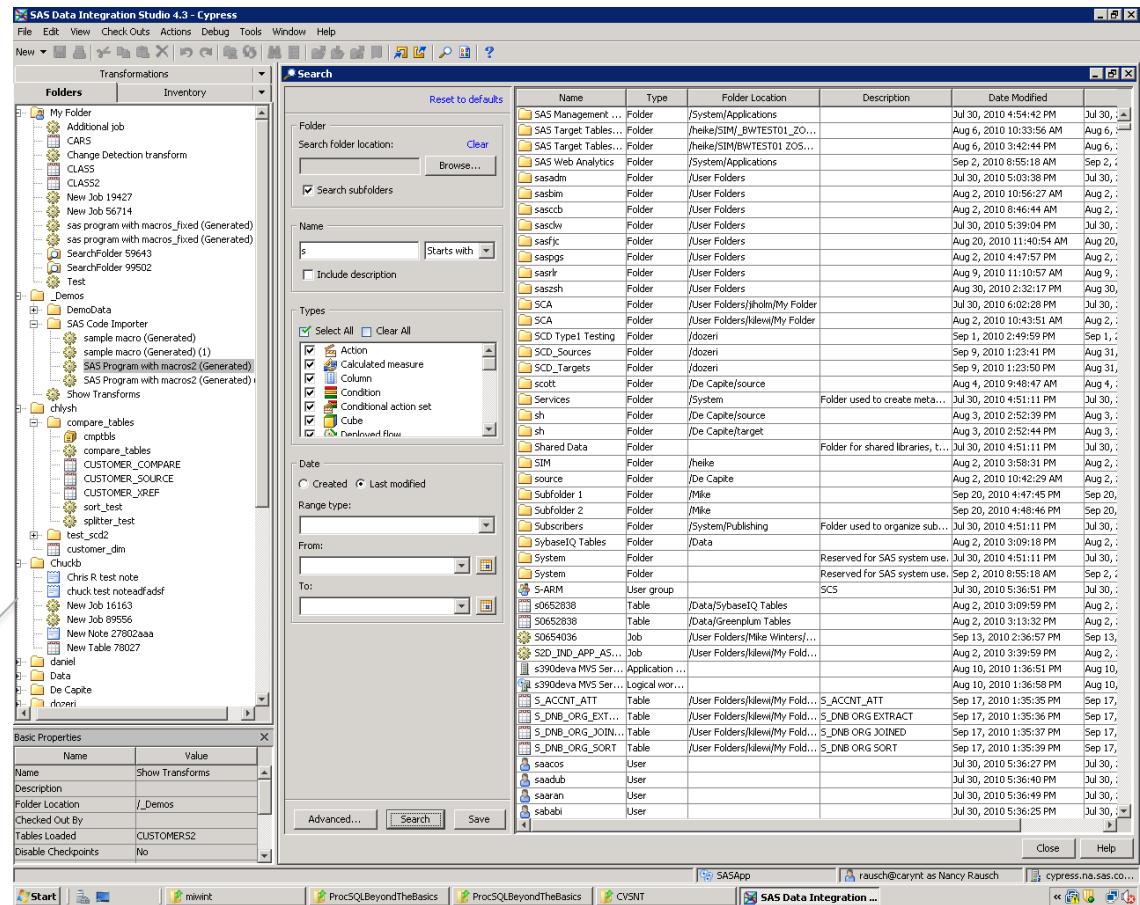
Modernisierung
Von bestehenden
BASE SAS
Programmen.



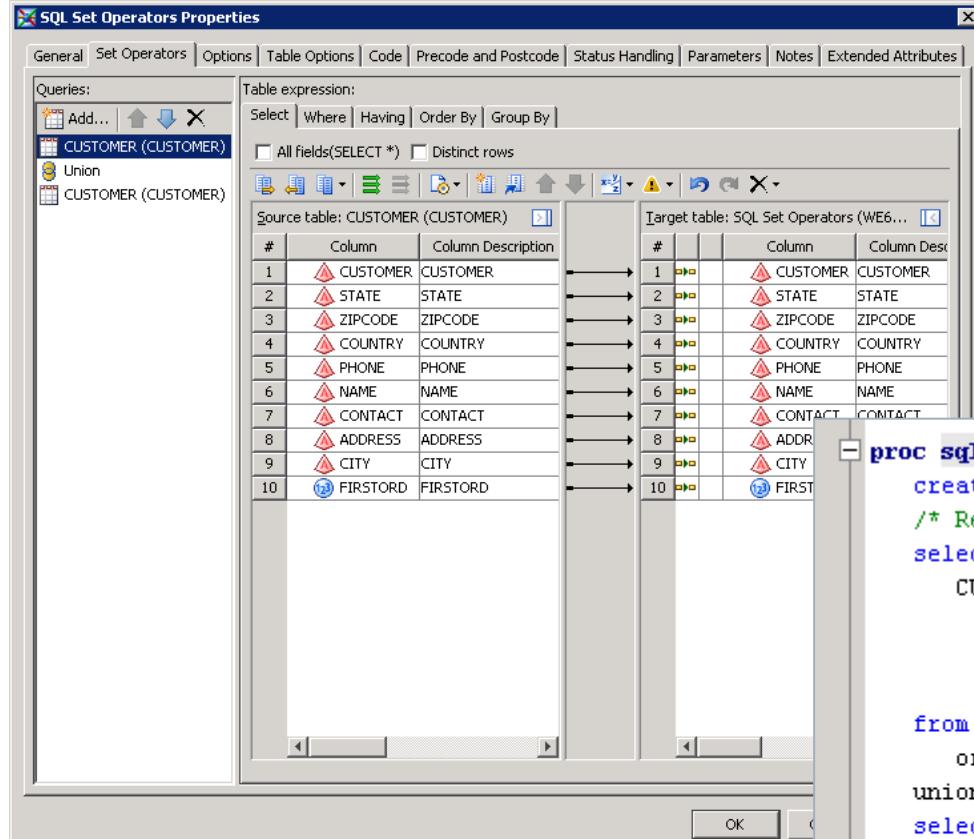
Integrierte Metadatensuche

- Suchen von Metadatenobjekten
- Komplexe Filtermöglichkeiten
- Wiederverwendbarkeit durch Speicherungsmöglichkeit

Metadaten-Analyse
Nicht nur für
Experten!



Erweiterte ELT Unterstützung

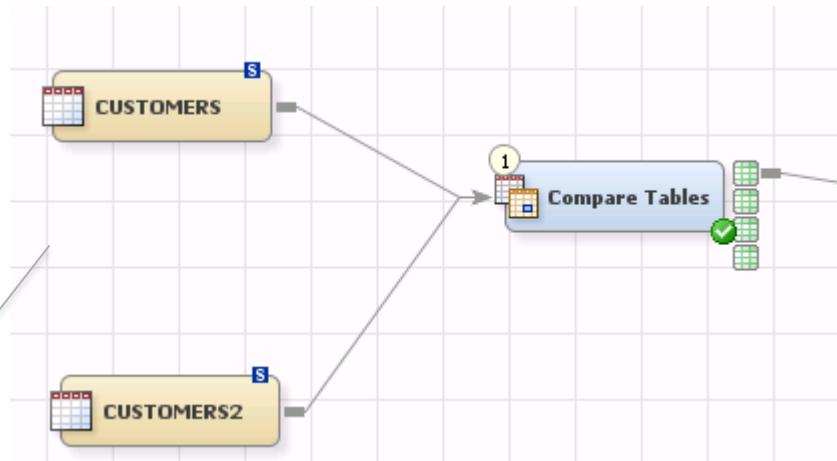


- SQL Union Transformation:
 - UNION
 - UNION ALL
 - INTERSECT
- Optimiert für DBMS

```
proc sql;
  create table orabk.CUSTOMER1 as
  /* Returns all unique rows from the two query results. */
  select
    CUSTOMER.CUSTOMER length = 8
    format = $8.
    informat = $8.
    label = 'CUSTOMER'
  from
    orabk.CUSTOMER
  union all
  select
    "" as CUSTOMER length = 8
    format = $8.
    informat = $8.
    label = 'CUSTOMER'
  from
    orabk.CUSTOMER
  ;
quit;
```

Compare Tables Transformation

- Vergleichen von 2 Tabellen auf Unterschiede
- Performance-Optimierung durch Nutzen von Hashing

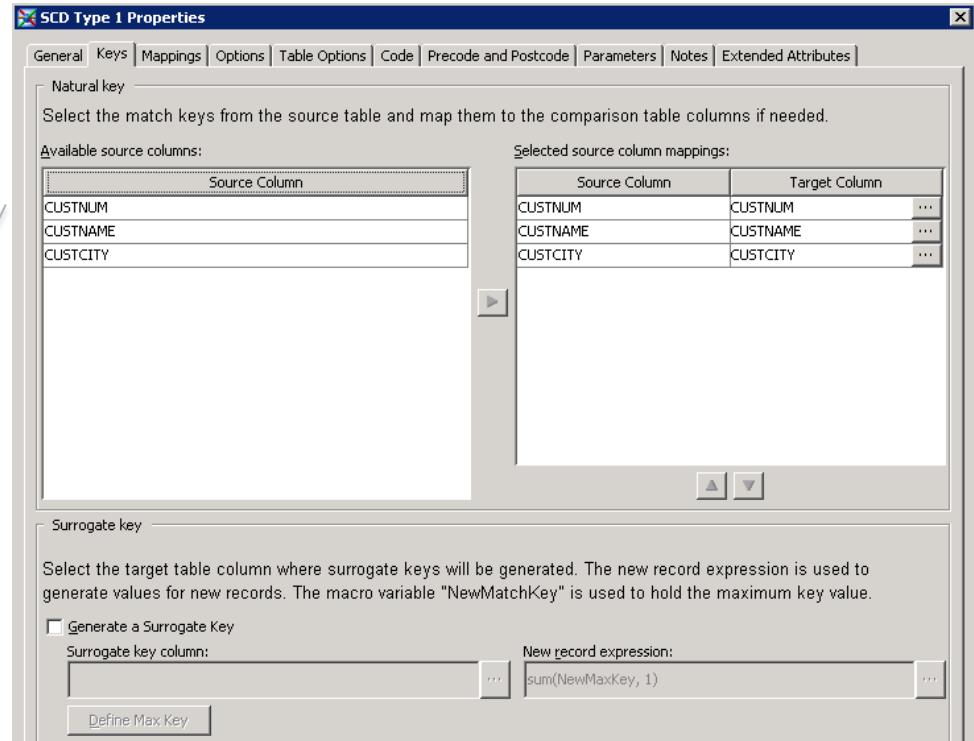


Effiziente Ermittlung
von Änderungen in
Tabellen

SCD Type 1 Transformation

- Aktualisierung ohne Historisierung

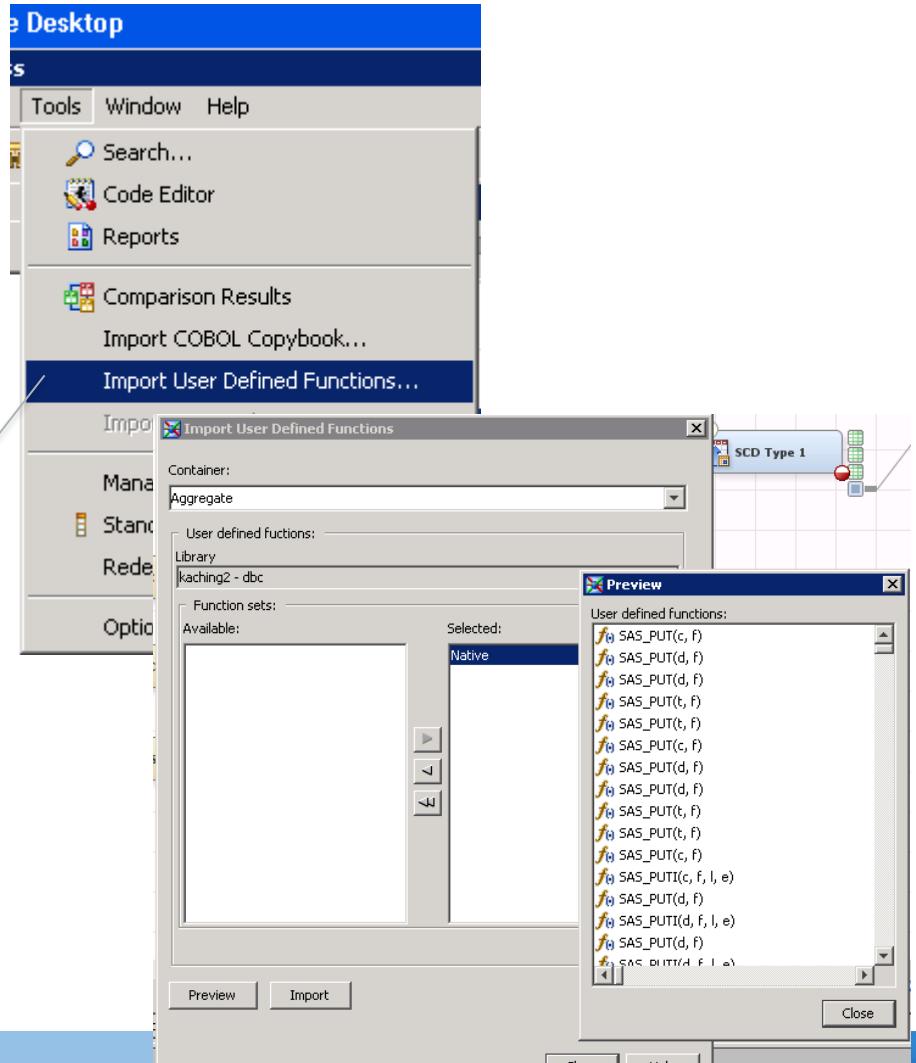
Effiziente Übernahme von Änderungen



Importieren von DBMS User Defined Functions

- Importieren und Nutzen im Expression Builder
- Unterstützt Standard DBMS UDFs und Enterprise Miner publishing Mining Analytics UDFs
- Parametererkennung

Erweiterte
Unterstützung von
DBMS Pushdown



Was ist Data Management?

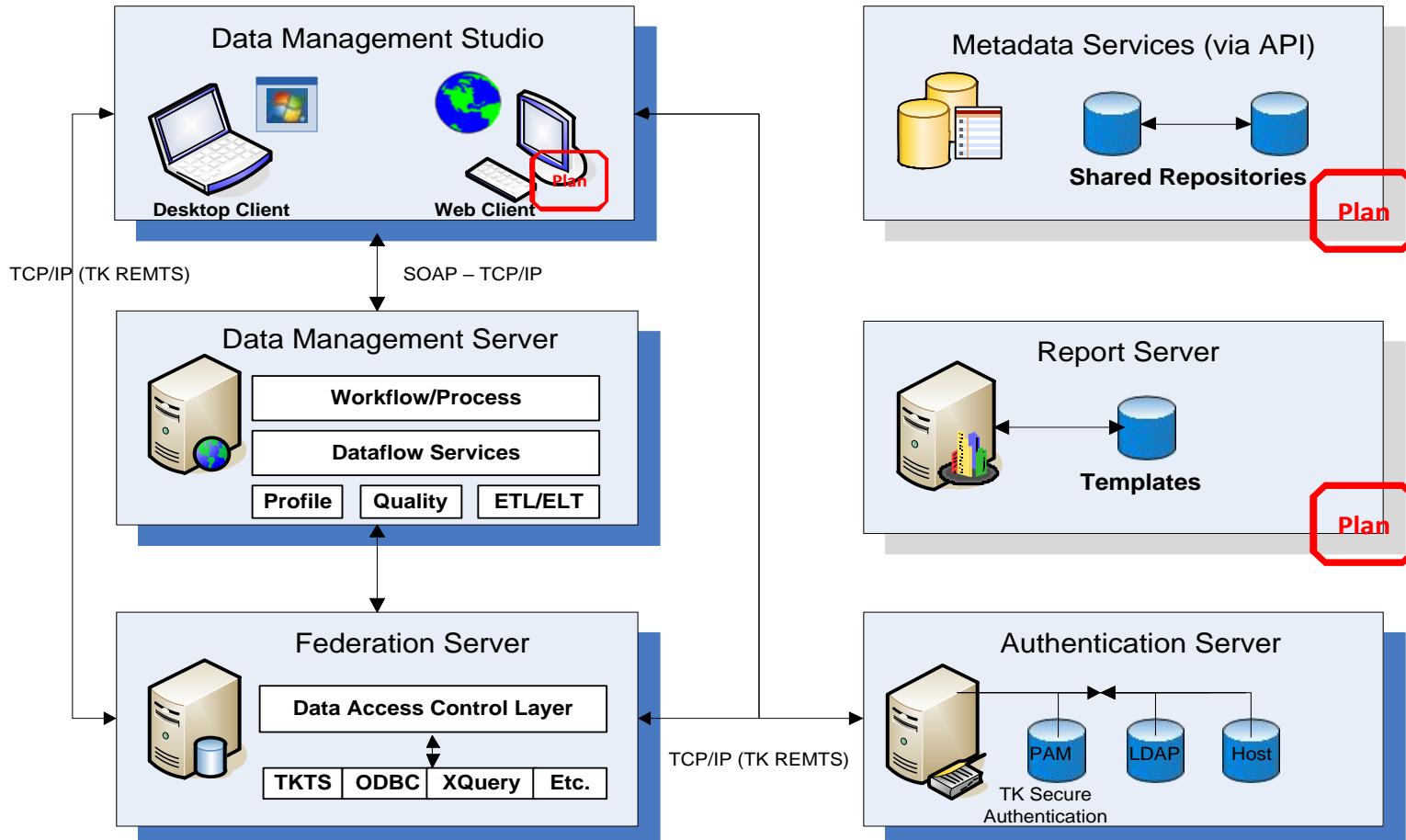
Data Management ist die von Analysten geforderte Zusammenführung von

- Data Quality
- Data Integration
- Master Data Management
- Metadata Management
- Enterprise Data Connectivity & Federation



Die Data Management Platform - Architektur

Data Management Platform - Architektur



Data Management Studio – Die Steuerzentrale

Informationen in einer Oberfläche

Navigation

Zuletzt bearbeitet

Zugriff auf Dataflux Blog

Einstellungen

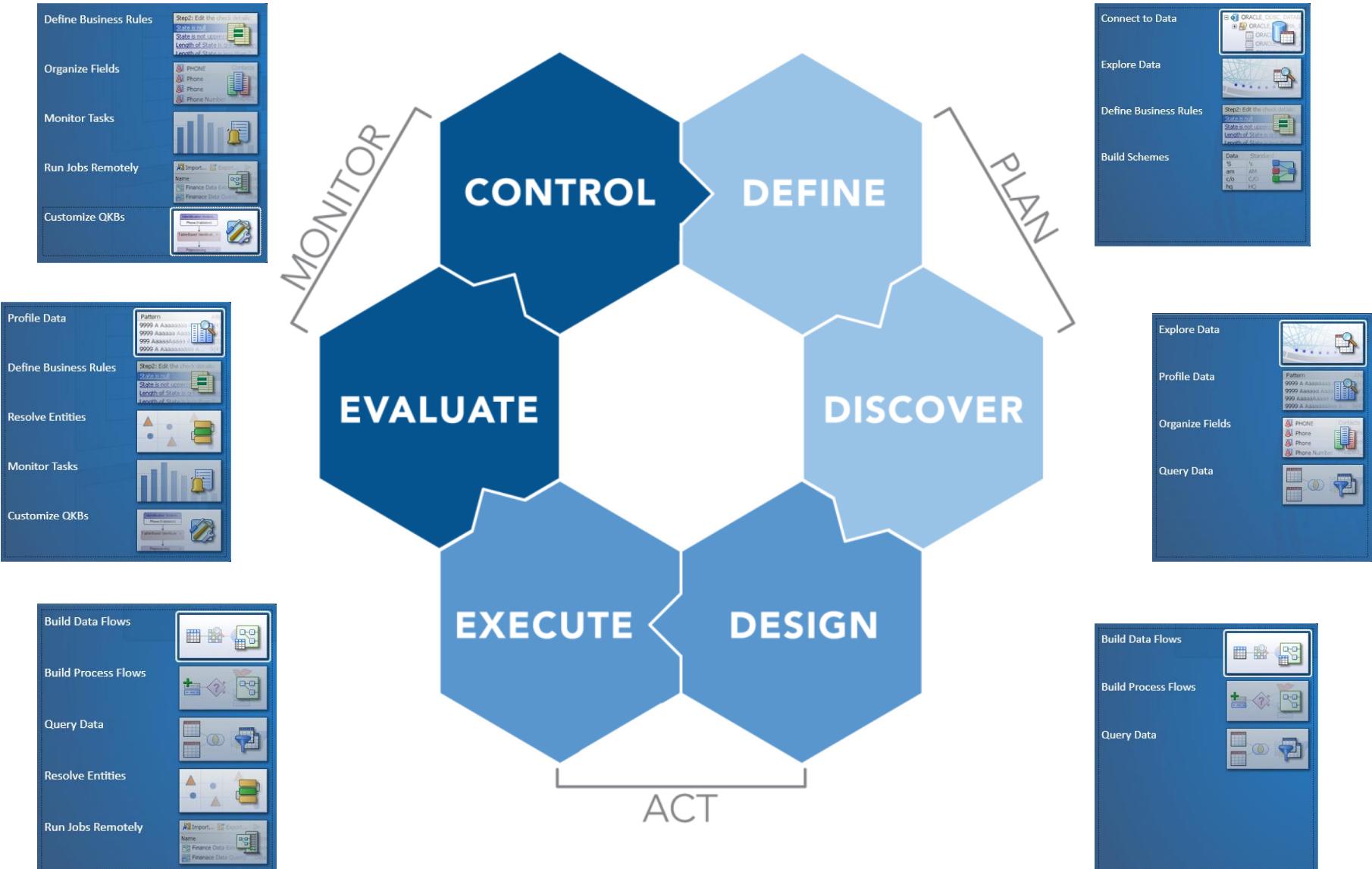
Web-Links

Dokumentation

Data Management Methodology

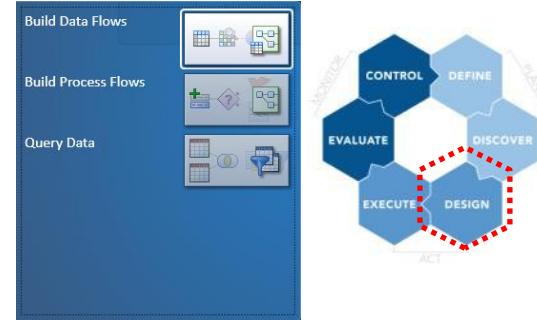
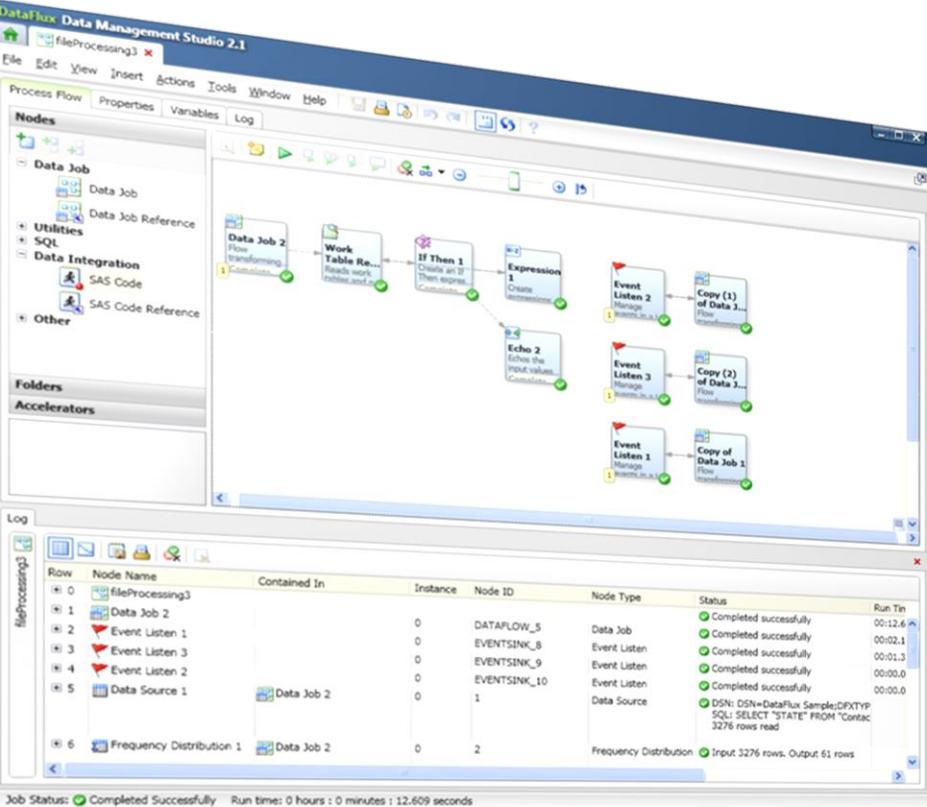
The screenshot shows the DataFlux Data Management Studio 2.1 interface. The main window displays various sections: 'Information' (Overview, Monitor), 'Default Settings' (Host ID, Active QKB, License Primary Method, etc.), and a central pane showing recent files and a blog feed from the DataFlux Community of Experts. Overlaid on the interface are several large, semi-transparent gray boxes containing text labels. One box labeled 'Zuletzt bearbeitet' is positioned over the 'Information' section. Another box labeled 'Zugriff auf Dataflux Blog' is positioned over the central pane. A third box labeled 'Einstellungen' is positioned over the 'Default Settings' section. A fourth box labeled 'Navigation' is positioned over the left sidebar. A fifth box labeled 'Web-Links' is positioned over the right sidebar. A sixth box labeled 'Dokumentation' is positioned over the bottom right sidebar. A seventh box labeled 'Data Management Methodology' is positioned over the top right area, containing a circular diagram with hexagonal nodes labeled MONITOR, CONTROL, DEFINE, PLAN, EVALUATE, DISCOVER, EXECUTE, and DESIGN.

Data Management Lifecycle Methodology



Data Management Studio

Highlights



- **Business Rules**
 - Erstellen und Verwalten von Regeln
⇒ **Business Rule Manager**
- **Data Management**
 - Erstellen und Verwalten von Schemas
 - Nativer Zugriff auf DBMS Typen
 - “ELT” basierte Daten Manipulation (z.B. In-Database Processing)
⇒ **Data Job Editor**
- **Process Modelling**
 - Parallel Processing Orchestration
 - Bedingte Ausführung
 - Event Triggering
 - Erstellen und Ausführen von Workflows
⇒ **Process Job Editor**

DataFlux Data Management Studio 2.1

File View Tools Window Help | New | ?

Data Connections

Total count: 1

DB2TRO

ODBC Verbindungen

Federation Server Verbindungen

Lokale DSN für Federation Server Verbindungen

Information

Data Connections

Collections

Folders

Data Management Servers

Administration

Locked by gertro

gertro-1:21030 | EUROPE\gertro as gertro

The screenshot shows the DataFlux Data Management Studio 2.1 interface. The left sidebar contains a tree view of 'Data Connections' with various entries like 'DataFlux Sample', 'DB2ODBC', 'DB2TRO', 'dBASE Files', etc. A callout box labeled 'ODBC Verbindungen' points to the 'DB2TRO' entry. Another callout box labeled 'Federation Server Verbindungen' points to the 'FEDTRO' entry. A third callout box labeled 'Lokale DSN für Federation Server Verbindungen' points to the 'SASTRO' entry under 'FEDTRO'. The bottom navigation bar includes links for 'Information', 'Data Connections' (which is highlighted in green), 'Collections', 'Folders', 'Data Management Servers', and 'Administration'. The status bar at the bottom shows 'Locked by gertro' and connection details 'gertro-1:21030 | EUROPE\gertro as gertro'.

DataFlux Data Management Studio 2.1

File View Tools Window Help | New | ?

Collections

New Collection

Total count: 106

Selbst erstelltes Dictionary von Tabellen und Spalten

| Name | Table | Database | Catalog | Schema | Type | Length | Description |
|--------------------|----------------|----------|-----------|---------|-----------|--------|-------------|
| NAME | BLCLASS | DB2TRO | geretro-1 | SASDEMO | Character | 8 | |
| SEX | BLCLASS | DB2TRO | geretro-1 | SASDEMO | Character | 1 | |
| HEIGHT | BLCLASS | DB2TRO | geretro-1 | SASDEMO | Numeric | 53 | |
| WEIGHT | BLCLASS | DB2TRO | geretro-1 | SASDEMO | Numeric | 53 | |
| NAME | DB2CLASS | DB2TRO | geretro-1 | SASDEMO | Character | 8 | |
| SEX | DB2CLASS | DB2TRO | geretro-1 | SASDEMO | Character | 1 | |
| AGE | DB2CLASS | DB2TRO | geretro-1 | SASDEMO | Numeric | 53 | |
| HEIGHT | DB2CLASS | DB2TRO | geretro-1 | SASDEMO | Numeric | 53 | |
| WEIGHT | DB2CLASS | DB2TRO | geretro-1 | SASDEMO | Numeric | 53 | |
| NAME | DB2CLASS2 | DB2TRO | geretro-1 | SASDEMO | Character | 8 | |
| SEX | DB2CLASS2 | DB2TRO | geretro-1 | SASDEMO | Character | 1 | |
| AGE | DB2CLASS2 | DB2TRO | geretro-1 | SASDEMO | Numeric | 53 | |
| HEIGHT | DB2CLASS2 | DB2TRO | geretro-1 | SASDEMO | Numeric | 53 | |
| WEIGHT | DB2CLASS2 | DB2TRO | geretro-1 | SASDEMO | Numeric | 53 | |
| CUSTOMER_ID | DB2_ORDER_FACT | DB2TRO | geretro-1 | SASDEMO | Numeric | 12 | |
| EMPLOYEE_ID | DB2_ORDER_FACT | DB2TRO | geretro-1 | SASDEMO | Numeric | 12 | |
| STREET_ID | DB2_ORDER_FACT | DB2TRO | geretro-1 | SASDEMO | Numeric | 12 | |
| ORDER_DATE | DB2_ORDER_FACT | DB2TRO | geretro-1 | SASDEMO | Temporal | 10 | |
| DELIVERY_DATE | DB2_ORDER_FACT | DB2TRO | geretro-1 | SASDEMO | Temporal | 10 | |
| ORDER_ID | DB2_ORDER_FACT | DB2TRO | geretro-1 | SASDEMO | Numeric | 12 | |
| ORDER_TYPE | DB2_ORDER_FACT | DB2TRO | geretro-1 | SASDEMO | Numeric | 53 | |
| PRODUCT_ID | DB2_ORDER_FACT | DB2TRO | geretro-1 | SASDEMO | Numeric | 12 | |
| QUANTITY | DB2_ORDER_FACT | DB2TRO | geretro-1 | SASDEMO | Numeric | 53 | |
| TOTAL_RETAIL_PRICE | DB2_ORDER_FACT | DB2TRO | geretro-1 | SASDEMO | Numeric | 53 | |
| COSTPRICE_PER_UNIT | DB2_ORDER_FACT | DB2TRO | geretro-1 | SASDEMO | Numeric | 53 | |
| DISCOUNT | DB2_ORDER_FACT | DB2TRO | geretro-1 | SASDEMO | Numeric | 53 | |
| CUSTOMER_ID | DB2_ORDER_STAR | DB2TRO | geretro-1 | SASDEMO | Numeric | 12 | |
| EMPLOYEE_ID | DB2_ORDER_STAR | DB2TRO | geretro-1 | SASDEMO | Numeric | 12 | |
| STREET_ID | DB2_ORDER_STAR | DB2TRO | geretro-1 | SASDEMO | Numeric | 12 | |
| ORDER_DATE | DB2_ORDER_STAR | DB2TRO | geretro-1 | SASDEMO | Temporal | 10 | |
| DELIVERY_DATE | DB2_ORDER_STAR | DB2TRO | geretro-1 | SASDEMO | Temporal | 10 | |
| ORDER_ID | DB2_ORDER_STAR | DB2TRO | geretro-1 | SASDEMO | Numeric | 12 | |

Information

Data Connections

Collections

Folders

Data Management Servers

Administration

geretro-1:21030 | EUROPE\geretro as geretro

DataFlux Data Management Studio 2.1

File View Tools Window Help | New | ?

Folders

- DataFlux Sample
 - Jobs
 - DB2_ORDER_FACT Query
 - Exploration
 - First_Job
 - First_Job
 - New Data Exploration
 - New Process Job
 - New Query
 - New Rule
 - New SAS File
 - Second_Job
 - Third_Job
 - sample
 - Data Jobs
 - dfPower Architect
 - Process Jobs
 - Profile Jobs
 - Shared Data

Objekteigenschaften & Ausführungsstatus

Start time:
End time:
Run time:
[View Log](#)

Lineage

> First_Job

CUSTOMER_DIM
GEOGRAPHY_DIM
ORGANIZATION...
PRODUCT_DIM
TIME_DIM
DB2_ORDER_FACT
DB2_ORDER_STAR

Ordnersicht der erstellten Objekte

Impact Analyse

Information
Data Connections
Collections
Folders
Data Management Servers
Administration

Locked by gertro

gertro-1:21030 | EUROPE\gertro as gertro

File View Tools Window Help | New Save Print Refresh ?

Administration

- [-] Authentication Servers
 - [+] gertro-1
- [-] Federation Servers
 - [+] gertro-1
- [-] Macro Files
 - [+] macros
- [-] Quality Knowledge Bases
 - [+] 2010A
- [-] Reference Sources
- [-] Repository Definitions
 - [+] DataFlux Sample

Administration
Authentication Server

Administration
Federation Server

Dataflux Macros

Administration
QKB

Administration von Fremddaten
(USPS & Geocode)

Administration
Dataflux Repositories

Information

Data Connections

Collections

Folders

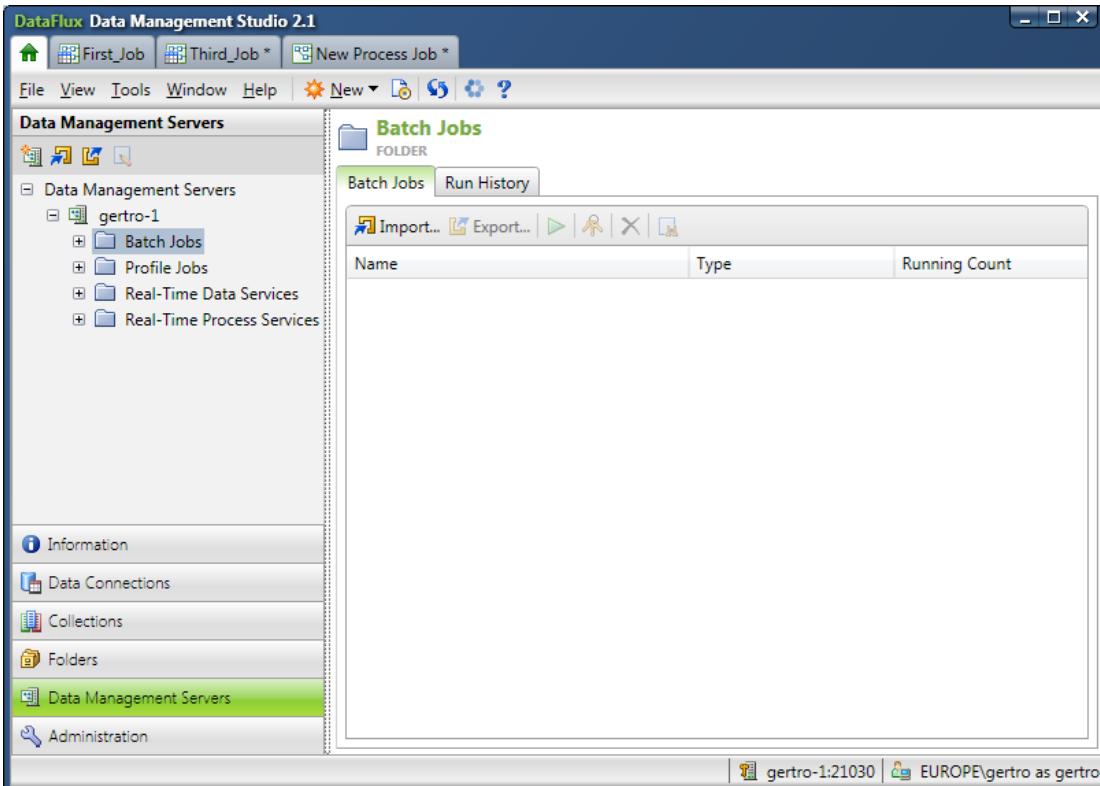
Data Management Servers

Administration

Locked by gertro

Data Management Server

Der „Dienstleister“



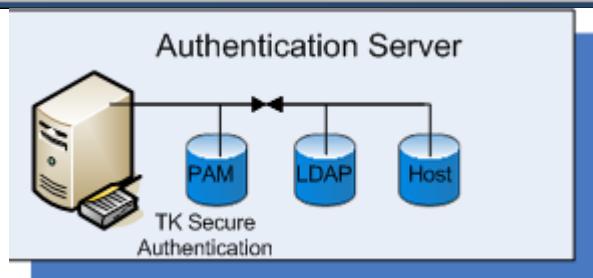
- Registrierung von
 - Jobs
 - Services (SOA)
- Vergleichbar Dataflux Integration Server
- Plattformen:
 - Windows 32 bit
 - Windows 64 bit
 - Solaris x86 64-bit
 - Solaris Sparc 64-bit
 - Linux x86 32-bit
 - Linux x86 64-bit
 - HP-UX Itanium 64-bit
 - AIX 64-bit
 - *HP-UX Pa-Risc geplant*

Authentication Server

Der „Türsteher“

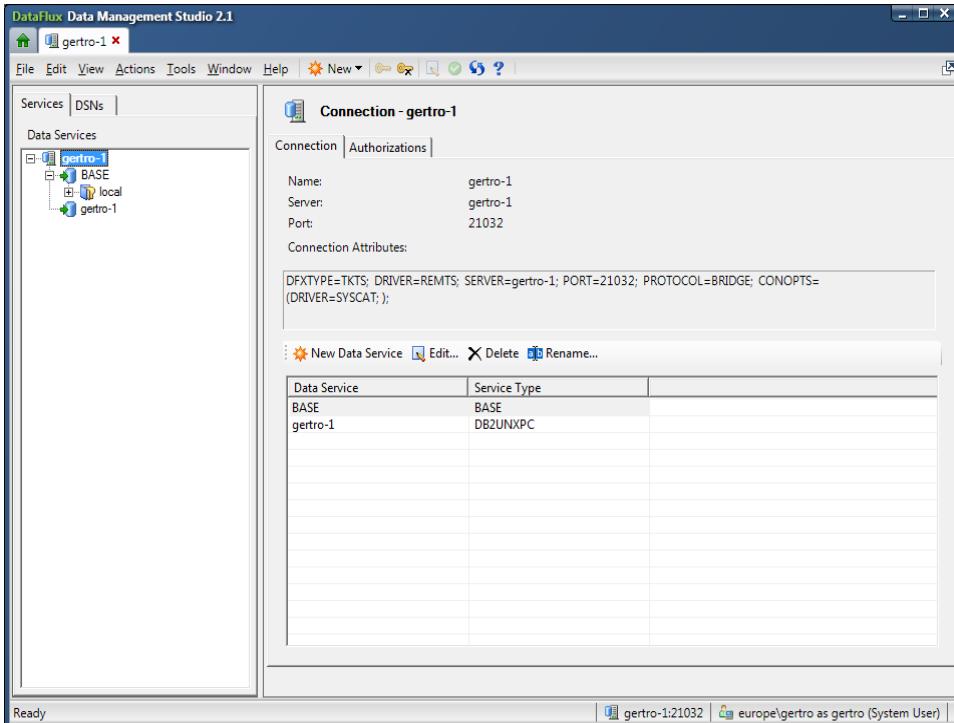
The screenshot shows the DataFlux Data Management Studio 2.1 interface. The title bar reads "DataFlux Data Management Studio 2.1". The menu bar includes "File", "View", "Tools", "Window", "Help", and "New". The main window is titled "Domains" and shows "All Domains". It lists two domains: "europe" and "gertro-1". The "europe" domain has a description "Down-level logon name (domain\userid)" and "Case Sensitive" set to "No". The "gertro-1" domain has a description "User name only (userid)" and "Case Sensitive" set to "No". A toolbar at the top has icons for "First Job", "Third Job", "New Process Job", and "gertro-1". The left sidebar has tabs for "Domains" (selected), "Users", "Groups", and "Shared Logins". The bottom status bar shows "gertro-1:21030" and "europe\gertro as gertro (Administrator)".

- Verwaltung von:
 - Domänen
 - Benutzern
 - Gruppen
 - Gemeinsam genutzten Logins
- Supportet Authentication Provider:
 - Host
 - LDAP
 - Active Directory
 - PAM



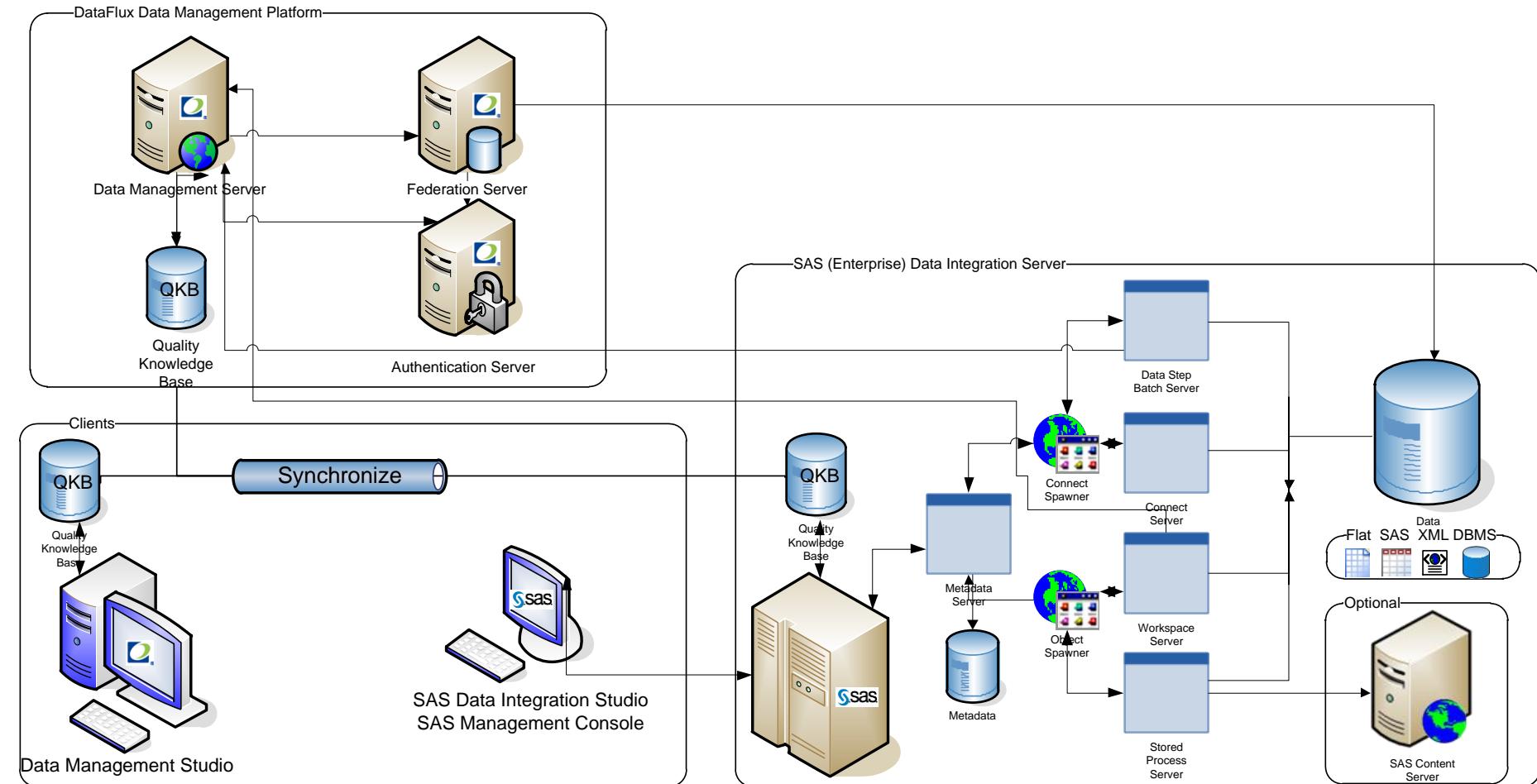
Federation Server

Die „Datenzentrale“



- Technologie (SAS Table Server):
 - Skalierbar
 - Threaded
 - Multi-user
 - Einheitlich
 - Federation Server TSSQL
 - DBMS native SQL
 - Sicher (über Authentication Server)
- Datenquellen:
 - SAS
 - DBMS:
 - DataFlux Driver for Base SAS
 - DataFlux Driver for Oracle
 - DataFlux Driver for Teradata
 - DataFlux Drive for DB2 CS
 - DataFlux Driver for DB2 z/OS
 - DataFlux Driver for MySQL
 - DataFlux Driver for ODBC
 - DataFlux Driver for SAP

SAS DI/DMP Architektur



Zusammenfassung Data Management Platform

Die Data Management Platform Version 2.1 ist der Nachfolger von dfPower Studio:

- Geführter Einstieg über Methodology
- Integrierte Oberfläche mit einheitlichem Look & Feel
- Einsteigerfreundlich

Die Data Management Platform 2.1 (**GA seit 11/2010**) oder aktueller wird dfPower Studio 8.x unter SAS 9.3 ersetzen.

Folgende Erweiterungen sind für SAS DI/DM geplant:

- Short-Term:

- Erweiterter ELT (Pushdown) Support
- JMS Support
- Hadoop Support
- Weitere DBMS-spezifische Loader

- Long-Term:

- Erweiterte Möglichkeiten zur gegenseitigen Nutzung von DI Studio und der DMP Platform:
 - Business Rules
 - DI Prozess Modellierung
 - Scheduling
- Event Stream Processing
- Solution Integration
- Performance Optimierung
 - Pipelining
 - Grid Support