Data Federation und Virtualisierung mit dem SAS Federation Server

Leistungsumfang & Architektur

Günter Greulich Solution Architekt, SAS DACH

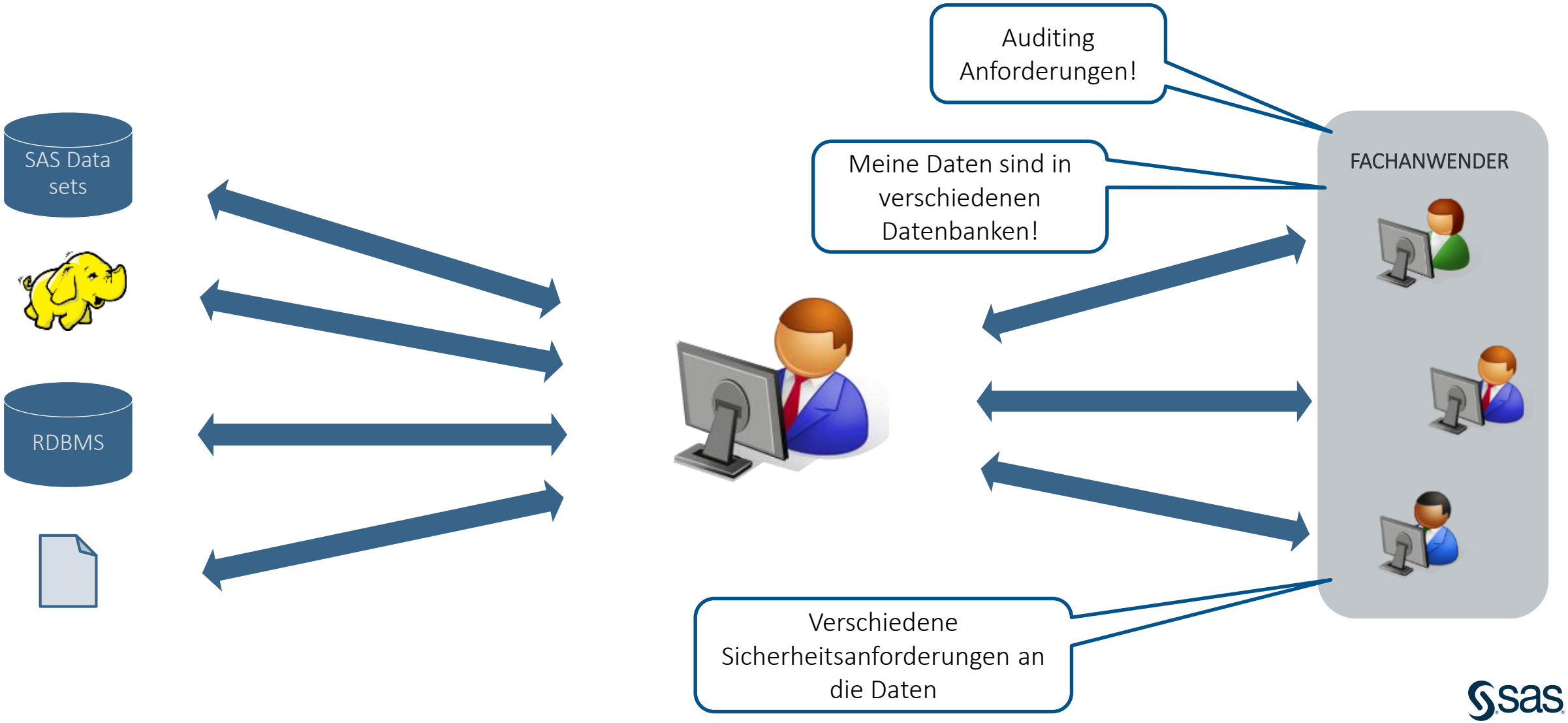
28.06.2017

Copyright © SAS Institute Inc. All rights reserved.

sas
THE POWER TO KNOW.

Data Federation

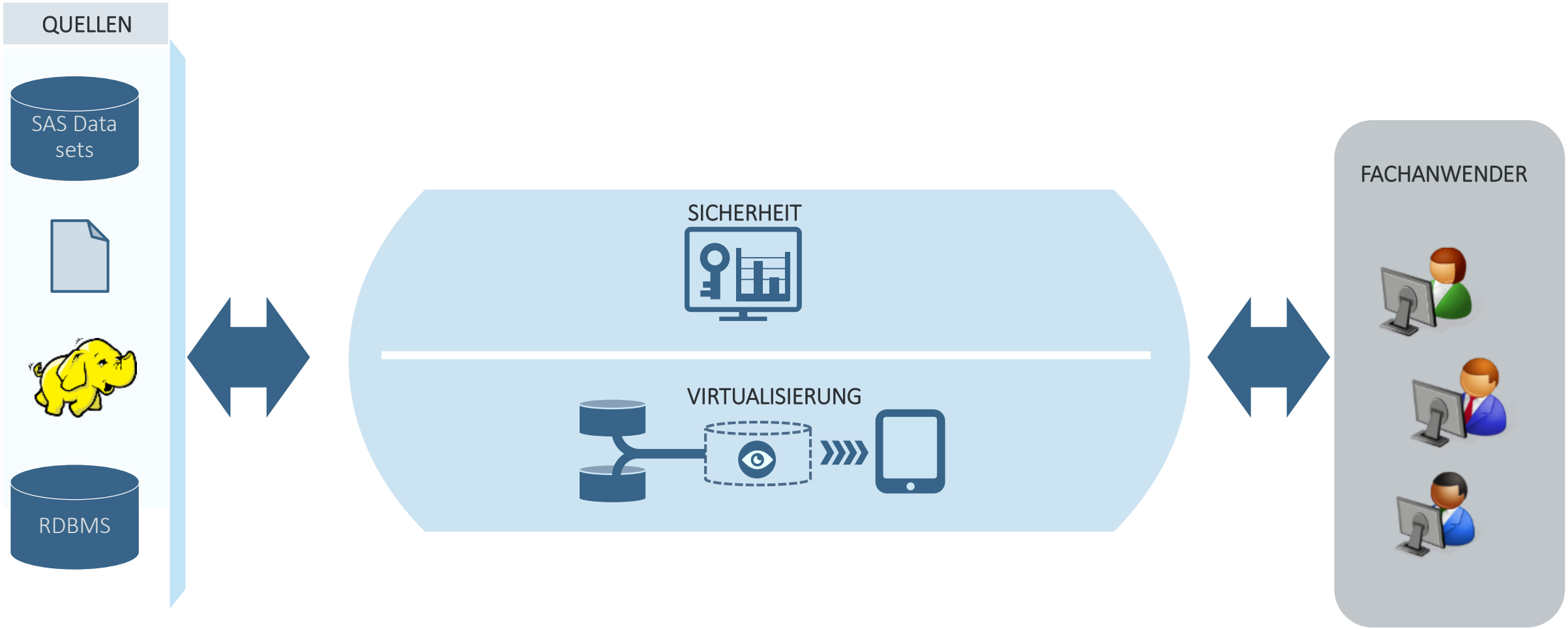
Warum? Weshalb?



Copyright © SAS Institute Inc. All rights reserved.

Data Federation

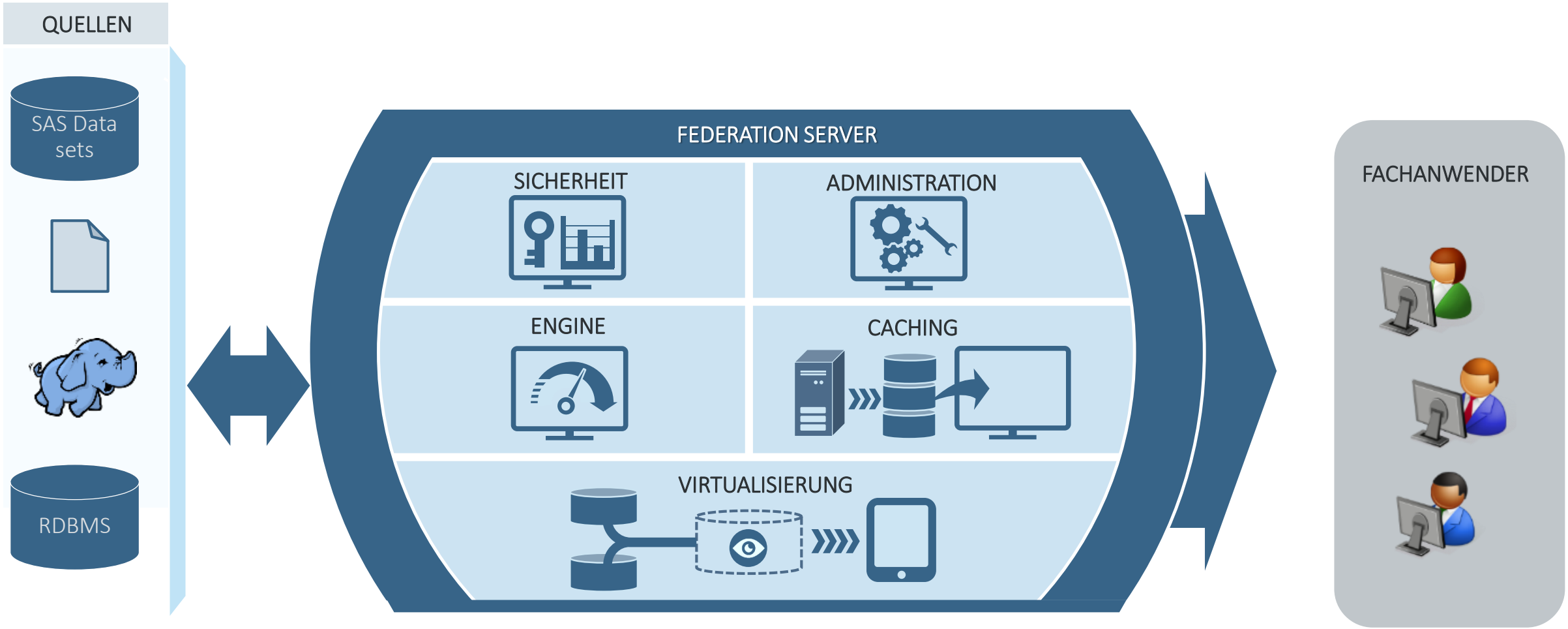
Das allgemeine Konzept



Copyright © SAS Institute Inc. All rights reserved.

SAS Federation Server

Data Federation, Governance, und mehr...



Copyright © SAS Institute Inc. All rights reserved.

Data Federation

Anforderung an Geschäftsdaten

Anwender

Fachanwender
Power Users
Contractors
Auditors
3rd Parties

Prozesse

Besteller
Call Center
Kunden Service
Kunden Onboarding
Auditing
DSGVO

Technologie

Data Warehouse
Master Data Management
Analytic Applikationen
Data Visualisierung
Mobile Apps
Query und Reporting, BI Tools
Spreadsheets

zusammegefasste Daten	Personal Daten	Kunden	neue Aktivitäten	Dashboard Analyse Daten
Öffentliche Daten	Private Daten	Produkte	Lieferanten	Komplette Detail Daten

Copyright © SAS Institute Inc. All rights reserved.

Data Federation

Anforderung an Geschäftsdaten

Anwender

Fachanwender
Power Users
Contractors
Auditors
3rd Parties

Prozesse

DSGVO
Besteller
Call Center
Kunden Service
Kunden Onboarding
Auditing

Technologie

Data Warehouse
Master Data Management
Analytic Applikationen
Data Visualisierung
Mobile Apps
Query und Reporting, BI Tools
Spreadsheets

Zusammegefasste Daten
Öffentliche Daten

Personal Daten
Private Daten

Kunden
Produkte

Neue Aktivitäten
Lieferanten

Dashboard Analyse Daten
Komplette Detail Daten

Verwaltet den Zugriff zu den Quelldaten

SAS Federation Server

SAS Federation Server Drivers

Data Warehouse

ERP

CRM

Big Data

Interne Daten

Externe Daten

Weitere Daten

Copyright © SAS Institute Inc. All rights reserved.

Personenbezogene Daten: Definition

Kapitel I Artikel 4: Begriffsbestimmungen

Merkmale aus denen man eine Person direkt oder indirekt identifizieren kann

- **physischen,**
- **physiologischen,**
- **genetischen,**
- **psychischen,**
- **wirtschaftlichen,**
- **kulturellen**
- **sozialen Identität oder**
- **Online Kennungen**

Artikel 4

Begriffsbestimmungen

Im Sinne dieser Verordnung bezeichnet der Ausdruck:

1. „personenbezogene Daten" alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person (im Folgenden „betroffene Person") beziehen; als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind, identifiziert werden kann;

N~512

N~9

N~4

N=1

Obs	ID	Name	Firma	KSFE17_Teilnahme	Vortrag17	KSFE_Historie
1	20148xJ02	Ulrich Reincke	SAS GmbH	1	1	16,15,14, ,12,11,10,9,8,7,6,5,4,3,2, ,0,99

Copyright © SAS Institute Inc. All rights reserved.

Die DSGVO-Auswirkungsanalyse mit SAS Lösungen

Schritt 4: Absichern und Loggen

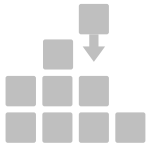
Data Quality	Lineage	Business Data Network	Federation Server	Visual Analytics

Erkennen

- Identifizieren personenbezog. Feldinhalte (DQ- Exploration und Scan)
- Inventurliste PD-kritischer Felder

Identifizieren

- Verbinden mit den Datenfluss-Metadaten zur Dokumentation der Prozesse
- Visualisierung der Lineage



Zuordnen

- Rollenbasiertes Glossar mit Verantwortlichen
- Verknüpfung fachlicher Verfahren mit technischem Data Dictionary

Absichern

- Zentrale Zugriffs-beschränkung mit dynamischen User-Rechten
- Federation
- Pseudonymisierung
- Anonymisierung
- Zugriff Logging

Überwachen

- DSB-Dashboard
- Auditing-Logs
- Monitoring
- Eskalieren und Nachhalten von Verstößen per Workflow

Copyright © SAS Institute Inc. All rights reserved.

SAS Federation Server

Leistungsumfang

Leistungsumfang - SAS Federation Server

1 DATEN ZUGRIFF

- Relationale Datenbanken
- SAS Datasets
- Hadoop (Hive)
- SAP
- ODBC Quellen
- weitere SAS Federation Servers

Gesicherter Datenzugriff & Bereitstellung über Data Services

2 VIRTUALISIERUNG

- Erstellung von Views
- Zeilen / Spalten Filter
- Daten Maskierung
- Granulare Kontrolle des authorisierten CRUD Zugriff
- Native SQL dialect
- DS2 Support
- Erstellung von Views mit heterogenen Daten Quellen

Views Erstellung mit den entsprechenden Daten und Anforderungen für den Endanwender

3 DATEN MANAGEMENT

- Extraction
- Parse
- Standardize
- Change Case
- Gender Analysis
- Pattern Analysis
- Matching
- Identification Analysis
- Analytic Model Scoring

Daten transformieren und die Datenqulität verbessern / standardisieren

4 OPTIMIERUNG

- Pushdown SQL
- Memory Data Store
- Data Caching
- Administration von User, Gruppen und Rollen
- Monitoren der User & Anfragen
- Logging

Managed Datenzugriff und Monitoring der Daten

5 SCHNITTSTELLEN

- SAS Libname
- JDBC
- ODBC
- REST API
- Proc FedSQL
- SASHDAT Driver

Daten den Usern und den Anwendungen zur Verfügung stellen

Copyright © SAS Institute Inc. All rights reserved.

SAS Federation Server

Benutzer und Anwender APPLIKATIONEN

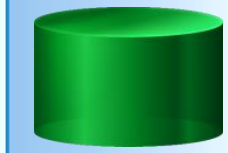
FEDERATION CLIENT

sas

JDBC REST ODBC

SAS FEDERATION SERVER

ADVANCED FEATURES



Business Views

Monitor

SMART CACHE

Multi-cache Support

MANAGED ACCESS

Data Services

Federated DSNs

SECURITY

Authentication

Authorization

Row & Column

DATA ACCESS

Native & ODBC Drivers

QUERY ENGINE

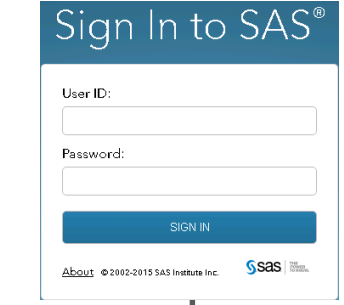
Federation Engine

Optimizer

HIGH THROUGHPUT READ/WRITE DATA ACCESS

DB2 Greenplum Oracle SAS Data Sets SAP Teradata Hadoop Netezza PostgreSQL Other (via ODBC)

SOURCE SYSTEMS



Federation Server Manager Mid Tier

SAS Metadata Server

LDAP

Active Directory

sas

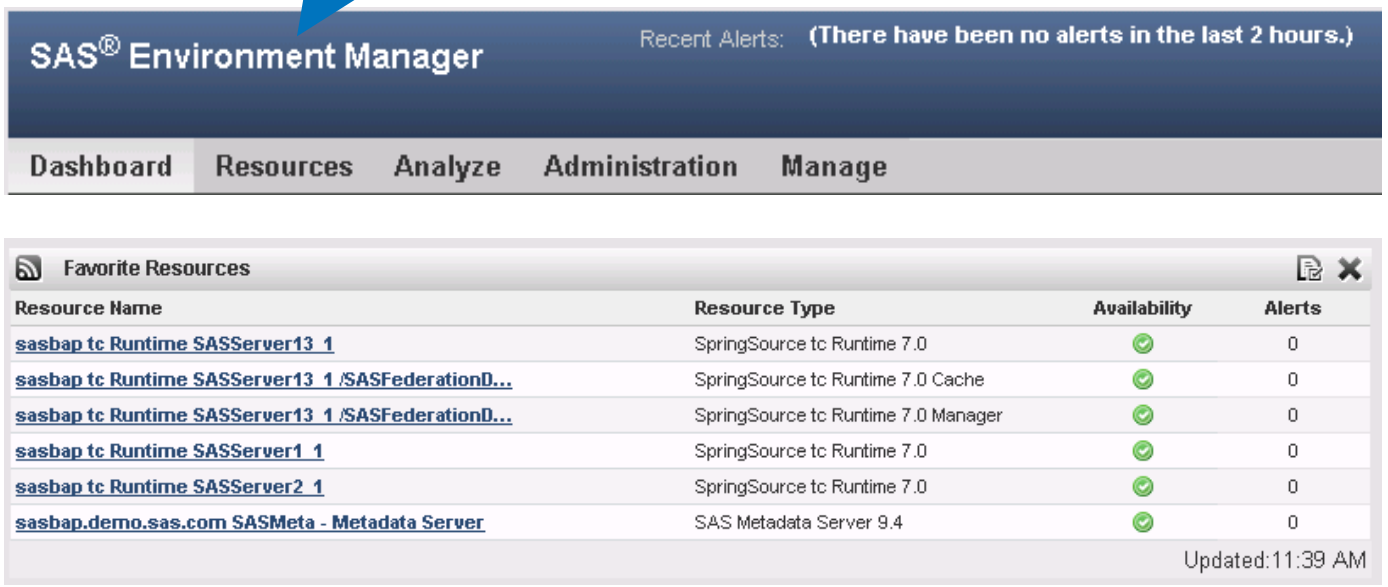
Copyright © SAS Institute Inc. All rights reserved.

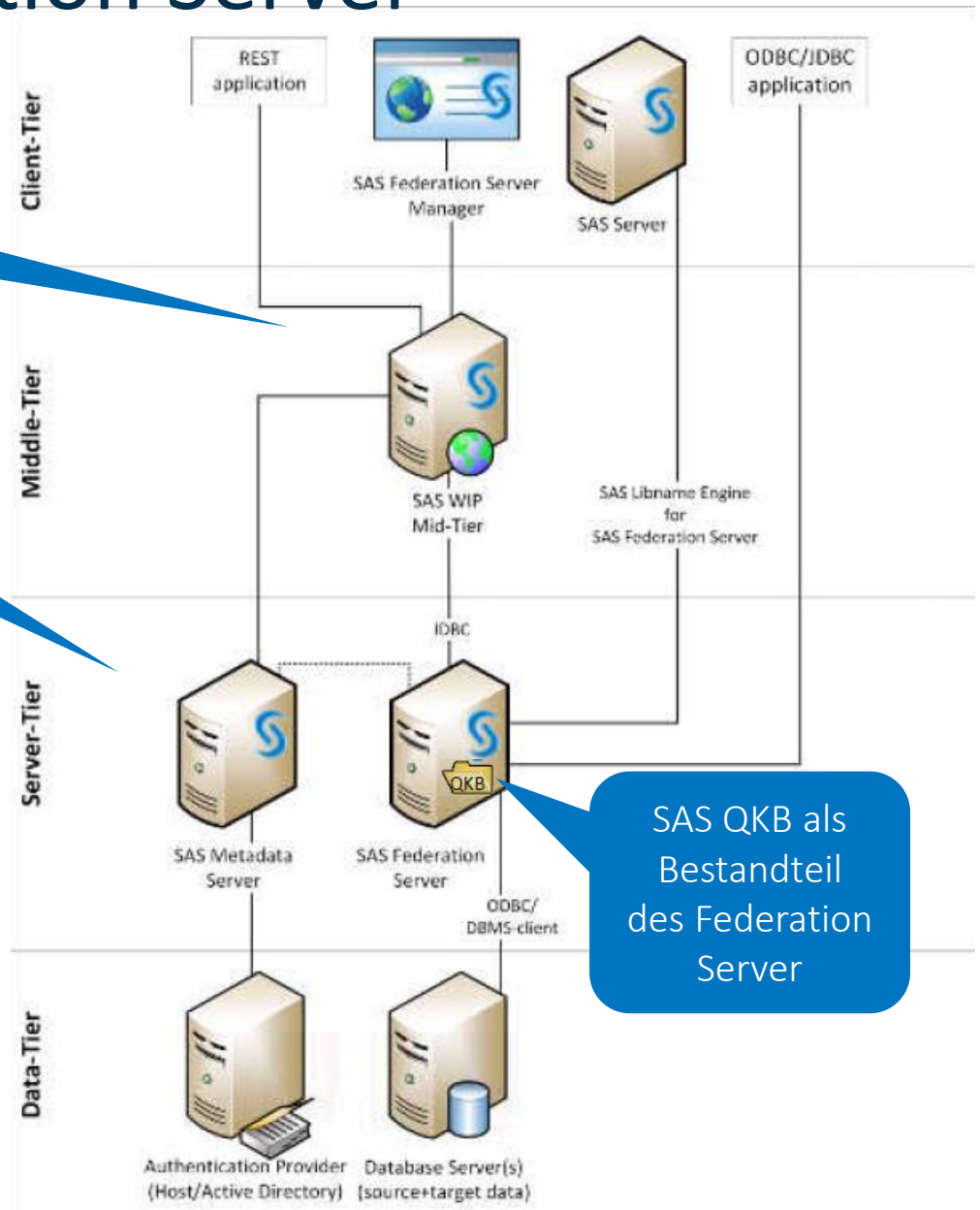
Architektur SAS Federation Server

Federation Server Manager ist ein Teil des SAS WIP Mid-Tier

SAS Metadaten Server Integration

Federation Server Architektur kann über SAS EV überwacht werden



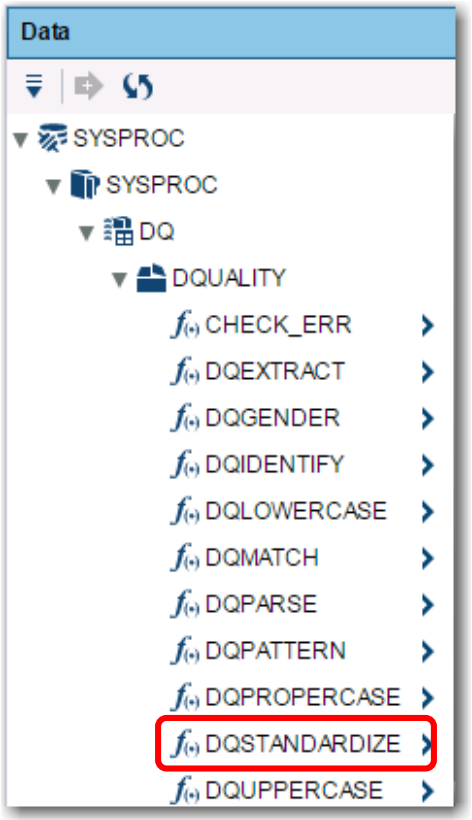


SAS QKB als Bestandteil des Federation Server

Copyright © SAS Institute Inc. All rights reserved.

SAS Federation Server

Data Quality Funktionen




```
CREATE VIEW "TPCAT"."TPSCHEMA"."BANK_STANDARDIZE_STATE" AS
      SELECT
         ID ,
         ACCOUNT_NUM ,
         TRANSACTION_TYPE ,
         TRANSACTION_AMOUNT ,
         TELLER_TYPE ,
         SYSPROC.DQ.DQUALITY.DQSTANDARDIZE (
            STATE ,
            'State/Province (Full Name)',
            'ENUSA' )  AS "STANDARD STATE",
         PUT(TRAN_DATE , DATE9.)  AS "DATE"
      FROM
         "TPCAT"."TPSCHEMA"."BANK_TRAN"
```

#	ID	ACCOUNT_NUM	TRANSACTION_TYPE	TRANSACTION_AMOUNT	TELLER_TYPE	STANDARD STATE	DATE
1	1111	1234567890	DEPOSIT	500.0	ATM	Pennsylvania	10DEC2015
2	2222	2345678901	DEPOSIT	100.0	MOBILE	Virginia	21JAN2016
3	3333	3456789012	INQUIRY	1433.45	ATM	California	02FEB2016
4	4444	1234567890	WITHDRAWAL	-250.0	BANK	Maryland	19NOV2015
5	5555	9991112222	PAYMENT	1200.5	ONLINE	Massachusetts	01JAN2016

Copyright © SAS Institute Inc. All rights reserved.

SAS Federation Server

Datenmaskierung: Verschiedene Methoden

```
SELECT ID, ACCOUNT_NUM, TRANSACTION_TYPE,
       TRANSACTION_AMOUNT, TELLER_TYPE, STATE, TRAN_DATE,
       CAST( SYSCAT.DM.MASK(
             'RANDATE',
             TRAN_DATE,
             'VARY',
             30,
             'UNITS',
             'D',
             'KEY',
             '212e8ba6b7f84796a87a985d54277f2f') AS DATE)
            AS MASK_DATE
  FROM "TPCAT"."TPSCHEMA"."BANK_TRAN"
```

#	ID	ACCOUNT_N...	TRANSACTIO...	TRANSACTIO...	TELLER_TYPE	STATE	TRAN_DATE	MASK_DATE
1	1111	1234567890	DEPOSIT	500.0	ATM	PA	12/10/15	11/29/15
2	2222	2345678901	DEPOSIT	100.0	MOBILE	VA	01/21/16	02/06/16
3	3333	3456789012	INQUIRY	1433.45	ATM	CA	02/02/16	02/18/16
4	4444	1234567890	WITHDRAWAL	-250.0	BANK	MD	11/19/15	11/28/15
5	5555	9991112222	PAYMENT	1200.5	ONLINE	MA	01/01/16	01/28/16

SAS Federation Server

Cache Funktionalität

COUNTRIES_ORACLE_4.2_VIEW CACHED FEDSQL VIEW

Summary | Cache | Query | Data | Authorizations | Row Authorizations

Cached Start: 01/14/16 12:24:21 PM End: 01/14/16 12:24:22 PM Run time: 00:00:01

Cache Definition ▶ Refresh Cache

Cache table's location:	Federation Server sasbap.In_Memory.In_Memory.Oracle_Data
FORCE:	NO
SET:	RESTART='REFRESH'

Nach Fed. Server Restart werden die Views im Hintergrund wieder automatisiert aktualisiert

MEMORY SYSTEM TABLE

Summary | Data | Authorizations | Row Authorizations

View Data | Go to: 1

#	DB_NAME	SCHEMA_NAME	TABLE_NAME	ROW_
1	In_Memory	(null)	(null)	(null)
2	In_Memory	Oracle_Data	CACHEcountries_ora769058673913	25
3	In_Memory	Federated_Data	CACHECUSTOMER_DIME769064320239	1000

CUSTOMER_DIMENSION_VIEW_MASKED CACHED FEDSQL VIEW

Summary | Cache | Query | Data | Authorizations | Row Authorizations

Cached Start: 01/22/16 06:45:20 AM End: 01/22/16 06:45:28 AM Run time: 00:00:08

Cache Definition ▶ Refresh Cache

Cache table's location:	Federation Server sasbap.In_Memory.In_Memory.Federated_Data
Description:	PII Masked Customer Dimension data
FORCE:	NO
SET:	RESTART='REFRESH'

Abfragen können im MDS (Memory Data Store) gespeichert werden = FAST PERFORMANCE

In_Memory DATA SERVICE

Summary | Data Source Names | Authorizations

Properties... | Update Memory Limit...

Name:	In_Memory
Type:	MDS
Catalog:	In_Memory
Attributes:	CASE_SENSITIVITY=(OBJECT=T;COLUMN=T)
Memory size :	3200
Memory peak :	3352
Memory limit :	0 (No memory limit)

Copyright © SAS Institute Inc. All rights reserved.

SAS Federation Server

Logging

Standard Reports sind enthalten und erfüllen in der Regel die regulatorischen Anforderungen

SAS ® Federation Server Manager - SQL Log

File Help Sign Out

Federation Server sasbap | Federation Server sasbap

Filter: Time and Date: All Data Maximum Rows Returned: 5000

Individual Requests | Open | Filter

User ID	Requests	Start Time	Request Lifetim...	Rows Fetched	Data Fetched (kb)	Cache Ac...	Total Work Time (s)
sas	SELECT data_service, catalog_name, schema_name, grantor_id, gra...	01/14/16 10:57:08 AM	0.04	15	2.6992	false	0.024
sas	SELECT data_service, catalog_name, schema_name, grantor_id, gra...	01/14/16 10:57:08 AM	0.048	15	2.6992	false	0.03
sas	SELECT grantee, source_grantee, inherited FROM PRIVILEGES ...	01/14/16 10:57:08 AM	0.042	30	1.0117	false	0.033
sas	SELECT DISTINCT grantee, grantee_id, grantee_type FROM PRIV...	01/14/16 10:57:08 AM	0.046	2	0.0957	false	0.039
sas	SELECT A.DSN_NAME, A.DATA_SERVICE_NAME, A.DESC, A.FORMA...	01/14/16 10:57:08 AM	0.106	3	0.7871	false	0.095
sas	SELECT PREDICATE FROM OBJECT_PRIVILEGES WHERE PRIVILE...	01/14/16 10:56:31 AM	0.019	0	0	false	0.013
sas	SELECT data_service, catalog_name, schema_name, grantor_id, gra...	01/14/16 10:56:30 AM	0.038	15	2.9629	false	0.023
sas	SELECT data_service, catalog_name, schema_name, grantor_id, gra...	01/14/16 10:56:30 AM	0.041	0	0	false	0.033
sas	SELECT data_service, catalog_name, schema_name, grantor_id, gra...	01/14/16 10:56:30 AM	0.053	15	2.9629	false	0.035
sas	SELECT data_service, catalog_name, schema_name, grantor_id, gra...	01/14/16 10:56:30 AM	0.038	0	0	false	0.028

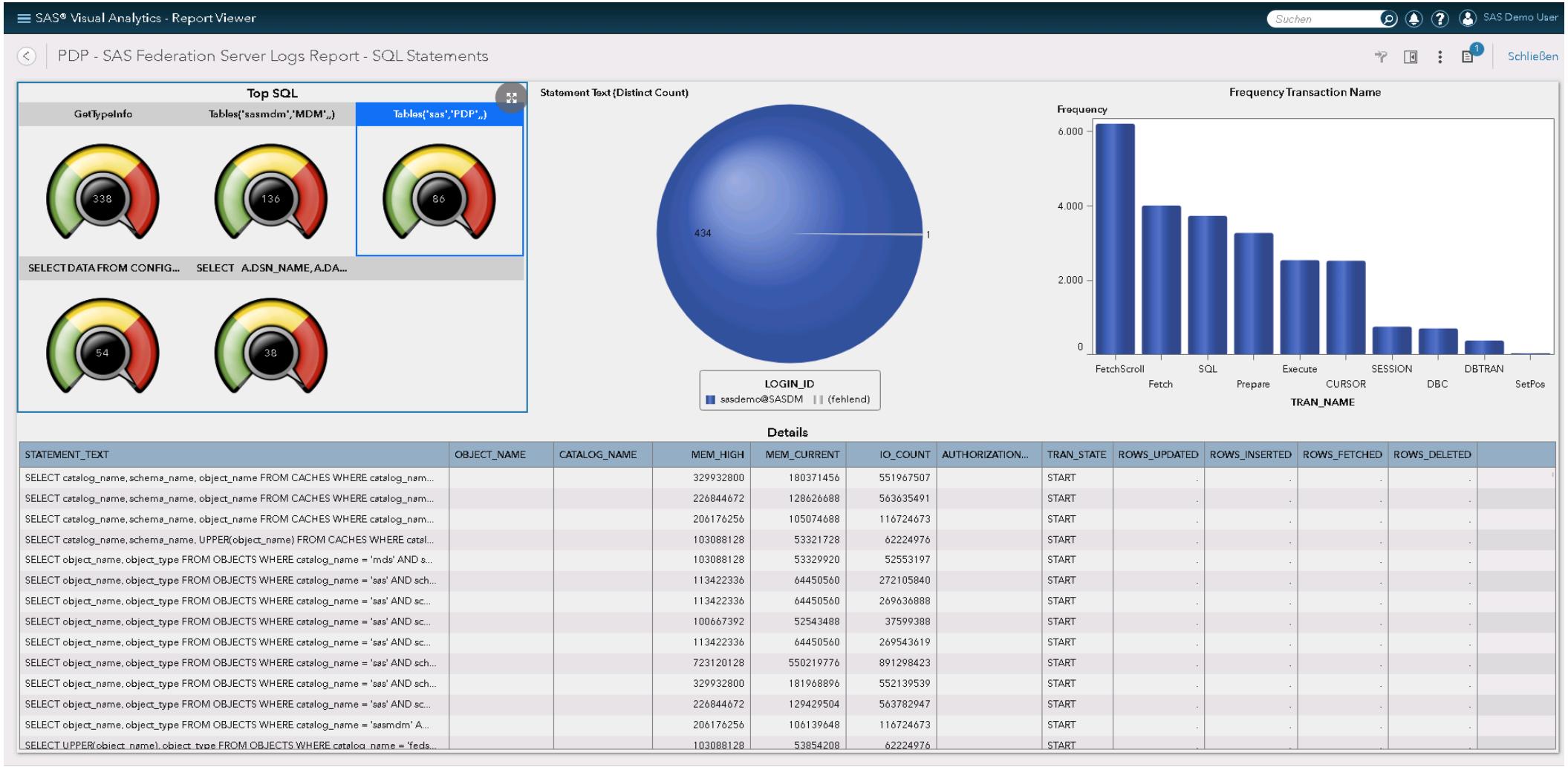
Details | Request | Plan | Cache Access

User ID:	sas	Request lifetime (s):	0.04	Number of executions:	1	Session ID:	ACj/AEQwNjExMzA0LTQ2MkMtNDMxOS1COUVCLUE2NUY2NzAyRjEzOQ==
Login ID:	sas@SASBAP	Cursor lifetime (s):	0.011	Rows fetched:	15		
IP address:	127.0.0.1	Total work time (s):	0.024	Rows inserted:	0		
Host name:	sasbap	-Total prepare time (s):	0.002	Rows updated:	0	Connection ID:	ACj/ADRDMUJFQ0JFLTMzMEQtNEU4NC1BMkMwLUJCODRDMDQ0QTc3QQ==
Host port:	59211	-Total execute time (s):	0.022	Rows deleted:	0		
Statement ID:	1612459565	-Total cursor time (s):	0	Data fetched (kb):	2.6992	Connection driver:	FEDSQL
Type:	DQL	-Total fetch time (s):	0			Dialect:	FEDSQL.Text
Status:	GOOD	-Total fetch scroll time (s):	0	Data inserted (kb):	0	Connection string:	CATALOG=*;DRIVER=FEDSQL;CONOPTS=

Copyright © SAS Institute Inc. All rights reserved.

SAS Federation Server

Erweitertes Logging via Visual Analytics



STATEMENT_TEXT	OBJECT_NAME	CATALOG_NAME	MEM_HIGH	MEM_CURRENT	IO_COUNT	AUTHORIZATION...	TRAN_STATE	ROWS_UPDATED	ROWS_INSERTED	ROWS_FETCHED	ROWS_DELETED	
SELECT catalog_name, schema_name, object_name FROM CACHES WHERE catalog_nam...			329932800	180371456	551967507		START	.	.	.	.	
SELECT catalog_name, schema_name, object_name FROM CACHES WHERE catalog_nam...			226844672	128626688	563635491		START	.	.	.	.	
SELECT catalog_name, schema_name, object_name FROM CACHES WHERE catalog_nam...			206176256	105074688	116724673		START	.	.	.	.	
SELECT catalog_name, schema_name, UPPER(object_name) FROM CACHES WHERE catal...			103088128	53321728	62224976		START	.	.	.	.	
SELECT object_name, object_type FROM OBJECTS WHERE catalog_name = 'mds' AND s...			103088128	53329920	52553197		START	.	.	.	.	
SELECT object_name, object_type FROM OBJECTS WHERE catalog_name = 'sas' AND sch...			113422336	64450560	272105840		START	.	.	.	.	
SELECT object_name, object_type FROM OBJECTS WHERE catalog_name = 'sas' AND sc...			113422336	64450560	269636888		START	.	.	.	.	
SELECT object_name, object_type FROM OBJECTS WHERE catalog_name = 'sas' AND sc...			100667392	52543488	37599388		START	.	.	.	.	
SELECT object_name, object_type FROM OBJECTS WHERE catalog_name = 'sas' AND sc...			113422336	64450560	269543619		START	.	.	.	.	
SELECT object_name, object_type FROM OBJECTS WHERE catalog_name = 'sas' AND sch...			723120128	550219776	891298423		START	.	.	.	.	
SELECT object_name, object_type FROM OBJECTS WHERE catalog_name = 'sas' AND sch...			329932800	181968896	552139539		START	.	.	.	.	
SELECT object_name, object_type FROM OBJECTS WHERE catalog_name = 'sas' AND sc...			226844672	129429504	563782947		START	.	.	.	.	
SELECT object_name, object_type FROM OBJECTS WHERE catalog_name = 'sasmdm' A...			206176256	106139648	116724673		START	.	.	.	.	
SELECT UPPER(object_name), object_type FROM OBJECTS WHERE catalog_name = 'feds...			103088128	53854208	62224976		START	.	.	.	.	

Copyright © SAS Institute Inc. All rights reserved.

SAS Federation Server

Scripting Batchverarbeitung in SAS Base für Deployment / automatisierte Updates

Batch Base SQL - create Data Service, Catalog & Schema

```
proc sql;
   connect to FEDSVR(server="sasbap.demo.sas.com" user=Student "{sas002}00403C1F0021646C0E1E965D3A1DC093" PORT=24141 dsn='ADMIN' schema=SYSCAT);
   execute (CREATE DATA SERVICE ORCL TYPE ORACLE CATALOG ORACLE_ORCL DOMAIN OracleAuth
        {OPTIONS CONOPTS (DRIVER ORACLE, PATH ORCL), CONOPTS (DRIVER ODBC, ODBC_DSN Oracle)}) by FEDSVR;
   execute (CREATE CATALOG "Test" UNDER BASE {OPTIONS conopts (COMPRESS YES)}) by FEDSVR;
   execute (CREATE SCHEMA "Test"."Dift" {OPTIONS (primarypath 'D:\Workshop\dift\data')}) by FEDSVR;
   disconnect from FEDSVR;
quit;
```

Batch Base SQL – create Security für User

```
proc sql;
   connect to FEDSVR(server="sasbap.demo.sas.com" user=Student password=Metadata0 PORT=24141 dsn=ADMIN schema=SYSCAT);

   execute (GRANT SELECT ON SCHEMA "ORCLINTG"."SCOTT" TO "sasdemo" as Administrator) by FEDSVR;
disconnect from FEDSVR;
quit;

proc sql;
   connect to FEDSVR(server="sasbap.demo.sas.com" user=Student password=Metadata0 PORT=24141 dsn=ADMIN schema=SYSCAT);

   execute (GRANT SELECT ON HR.EMPLOYEES to SASUSER where (syscat.rls.current_user() in ("Name","MgrName"))
            or syscat.rls.member_of("Payroll",'DEEP') as Administrator) by FEDSVR;
disconnect from FEDSVR;
quit;
```

SAS Federation Server

Scripting via Federation Server SQL Editor für Deployment / automatisierte Updates

Data Service, Catalog & Schema löschen und neu anlegen

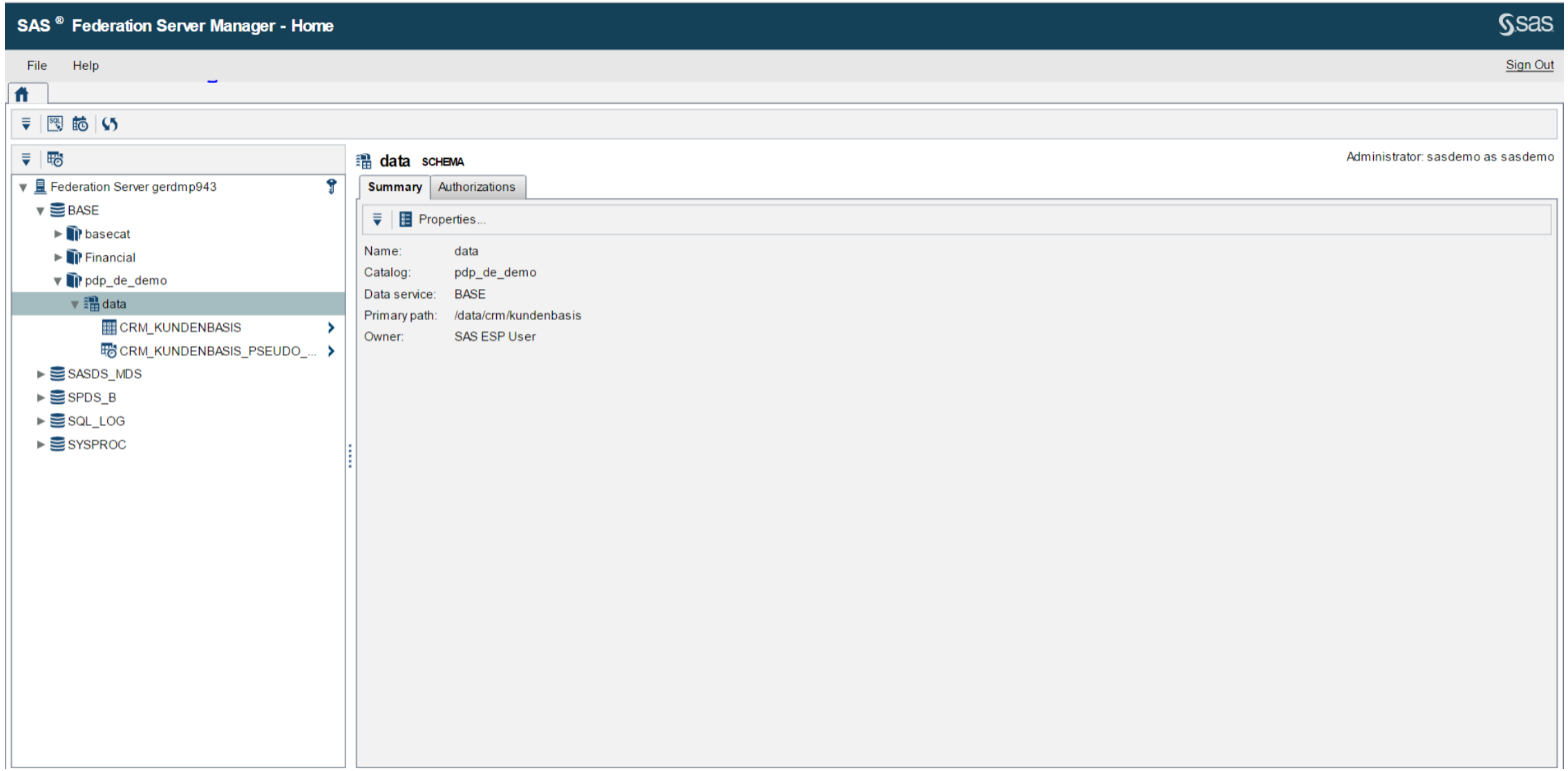
```
drop DSN "ORCL";
drop data service "ORCL" force;
drop schema "Test"."Dift";
drop catalog "Test";

CREATE DATA SERVICE ORCL TYPE ORACLE CATALOG ORCL DOMAIN OracleAuth
{OPTIONS CONOPTS (DRIVER ORACLE, PATH ORCL), CONOPTS (DRIVER ODBC,
ODBC_DSN Oracle)};
CREATE CATALOG "Test" UNDER BASE {OPTIONS conopts (COMPRESS YES)};
CREATE SCHEMA "TEST"."Dift" {OPTIONS (primarypath
'D:\Workshop\dift\data')};
```

Copyright © SAS Institute Inc. All rights reserved.

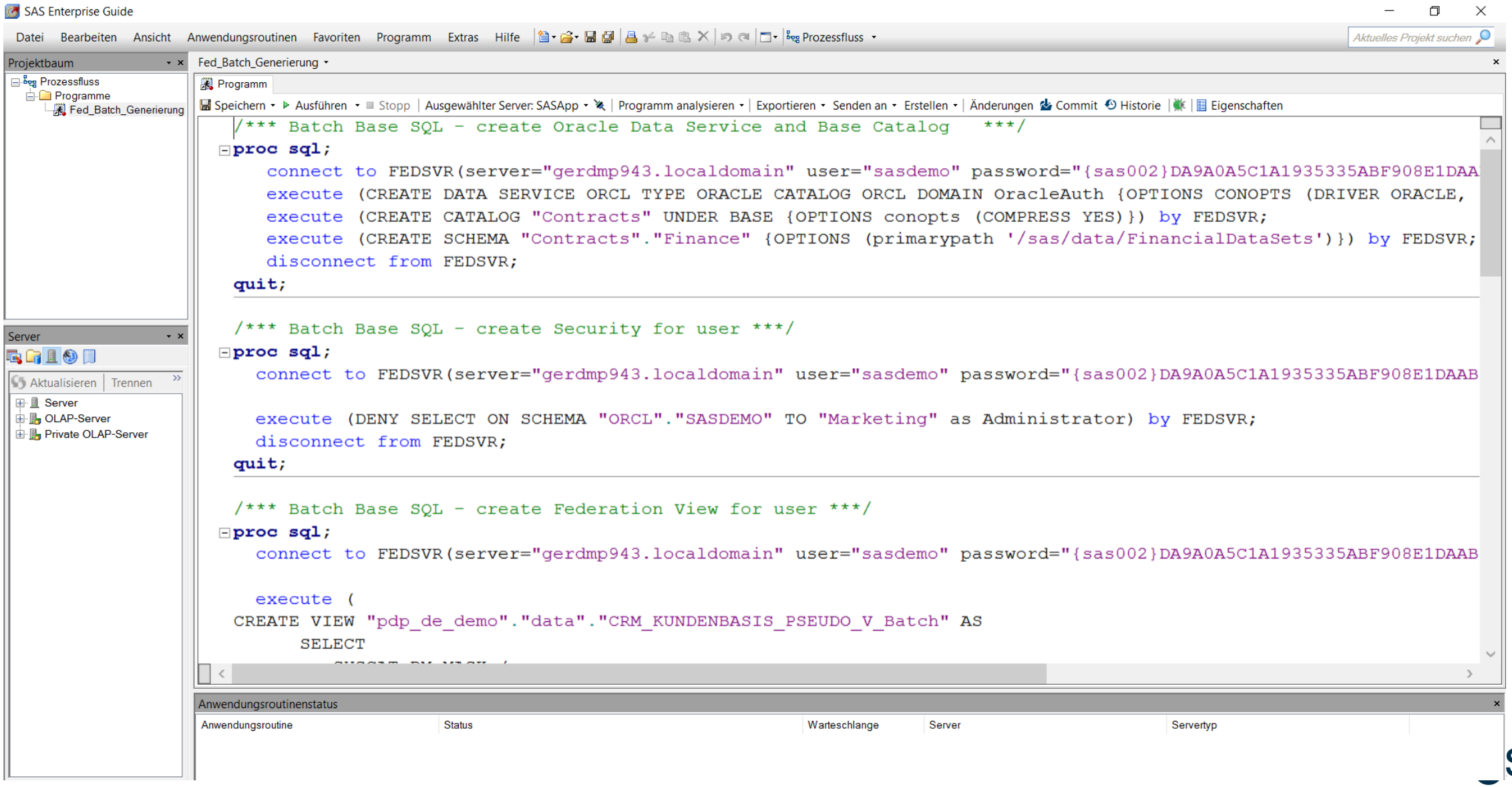
Copyright © SAS Institute Inc. All rights reserved.

Zugriff für Oracle, SAS und View anlegen per Batch



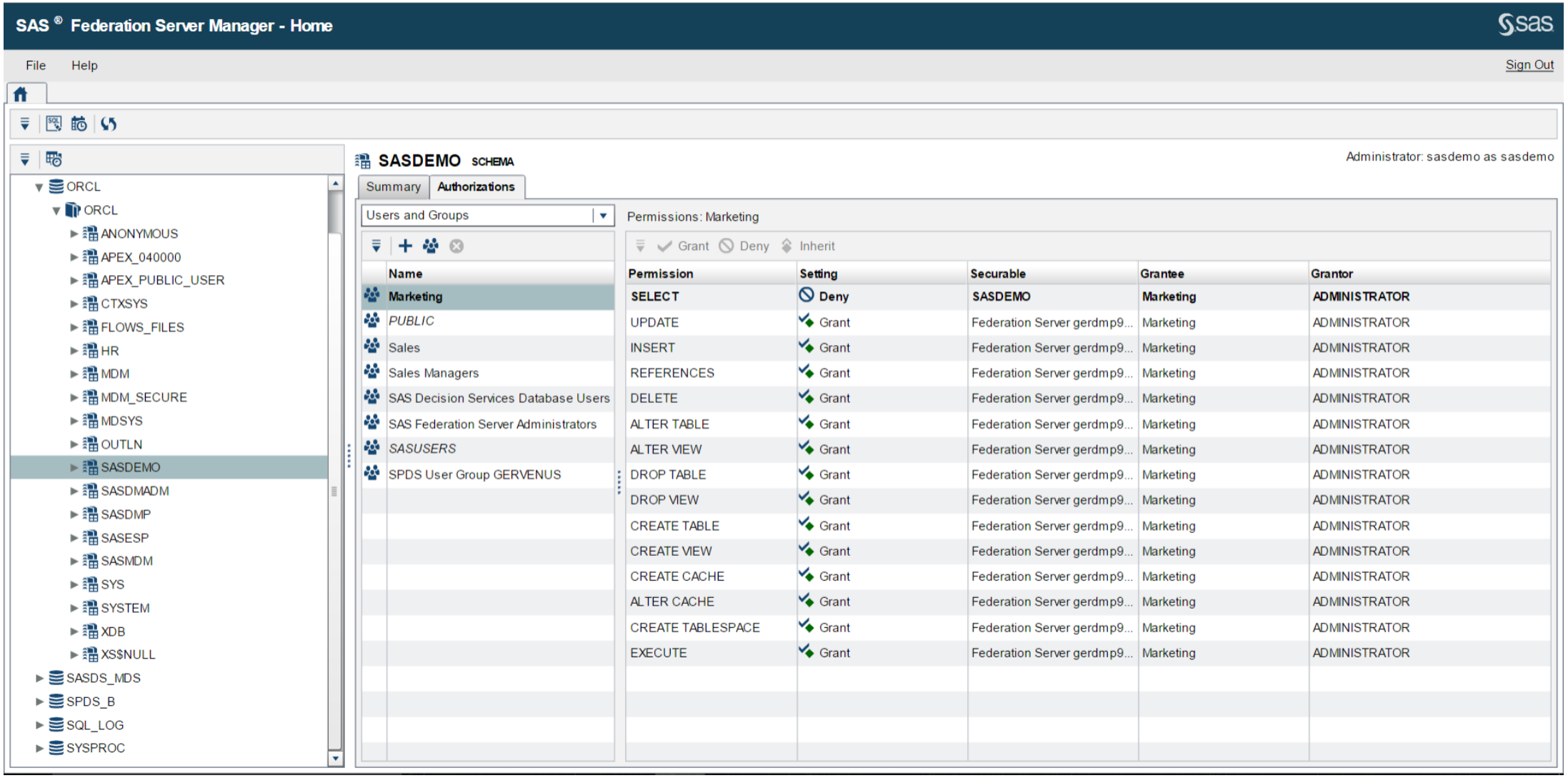
Copyright © SAS Institute Inc. All rights reserved.

Erstelltes Programm im EG ausführen



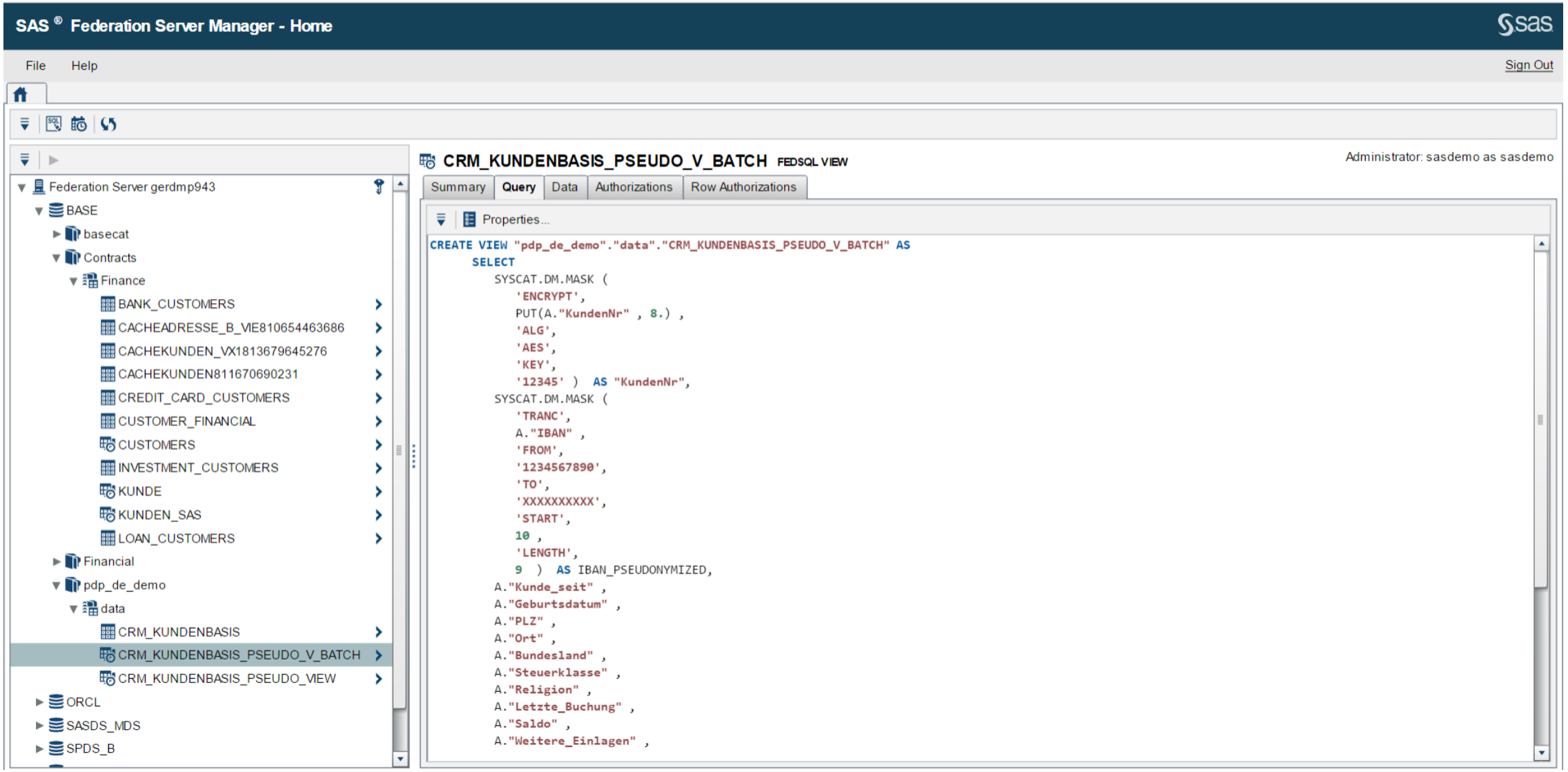
Copyright © SAS Institute Inc. All rights reserved.

Angelegter Oracle Data Service mit Security



Copyright © SAS Institute Inc. All rights reserved.

Angelegtes Base Schema und Federation View



Copyright © SAS Institute Inc. All rights reserved.

Zusammenfassung

SAS Federation Server

- Integriert in die SAS Platform
  - Deployment / automatisierte Updates per SAS Batch
  - Metadata Authentication
  - SAS WebServer

- Unterstützt die DSGVO durch folgende Funktionen:
  - Datenmaskierung (Anonymisierung und Pseudonymisierung)
  - Datenqualität Funktionen (Quality Knowledge Base)

- Zentrale Administration / Zugriff auf die Daten
  - durch Virtualisierung über Datenbanken hinweg
  - Logging der Zugriffe (Nachvollziehbarkeit)

https://www.sas.com/en_us/software/federation-server.html

Copyright © SAS Institute Inc. All rights reserved.