

CoC und DSGVO im DWH

- Oder wie stelle ich mehrere Tausend Tabellen unter
Kontrolle des Federation Server

Thomas Frenzel und Miriam Zand

19.06.2018

Agenda

Ausgangslage DWH

Anforderung CoC und DSGVO

Konzept und Realisierung CoC und DSGVO im DWH

Einführung

Ausblick

Agenda

Ausgangslage DWH

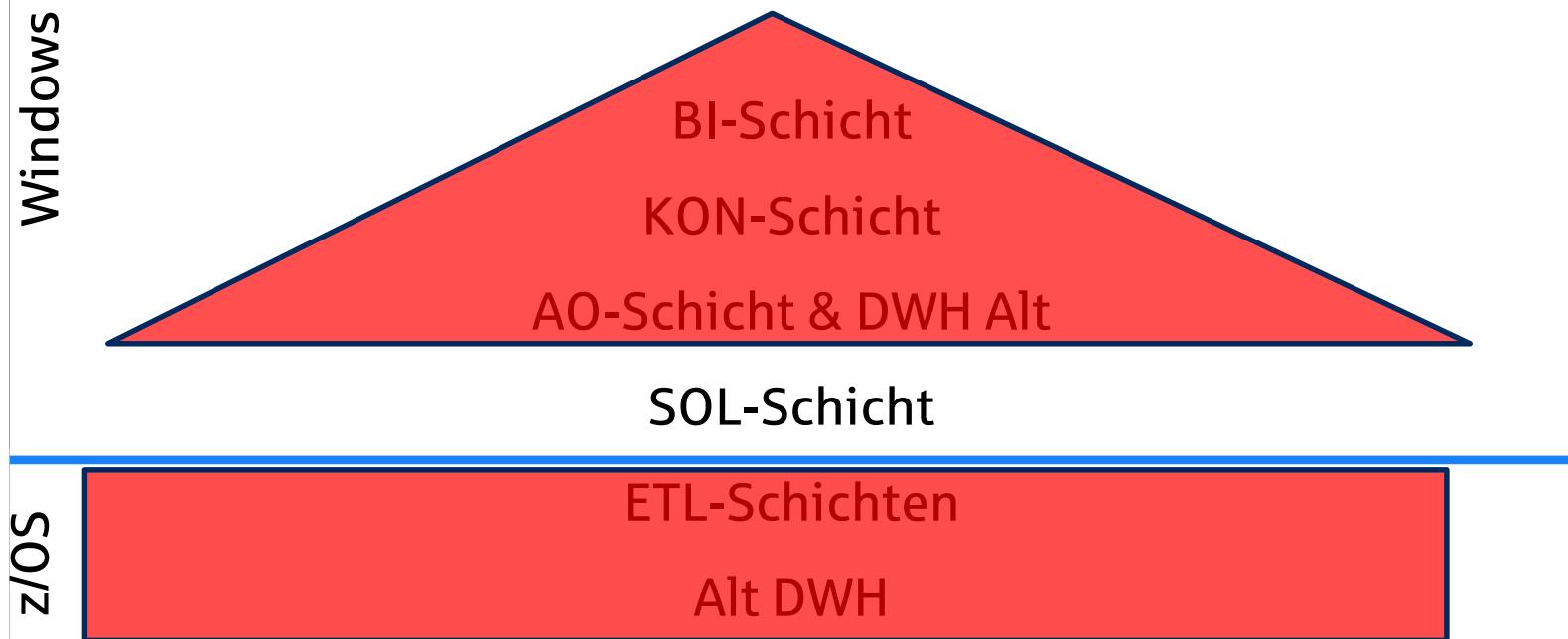
Anforderung CoC und DSGVO

Konzept und Realisierung CoC und DSGVO im DWH

Einführung

Ausblick

Die DWH / BI Landschaft teilt sich auf in mehrere Datenschichten



Ausgangslage DWH

Provinzial NordWest

Schichten	Anzahl Server-kontexte	Daten-verzeich-nisse	Metadaten-bibliothek	Tabellen	SAS Views	Index	Metadaten-registrierte Tabellen &View	Größe in TB
BI-Schicht	15	30	30	769	4	11	511	1,60
KON-Schicht	1	22	22	6.901	5	217	6.028	2,08
AO-Schicht	1	51	51	13.219	7.031	169	19.720	3,15
DWH-Alt	1	6.043	0	38.496	0	4.287	0	3,17
SOL-Schicht	1	7	7	275	3	3	0	0,10
Summe	19	6.153	110	59.660	7.043	4.687	26.259	10

Agenda

Ausgangslage DWH

Anforderung CoC und DSGVO

Konzept und Realisierung CoC und DSGVO im DWH

Einführung

Ausblick

CoC

Anonymisierung von zu sperrenden Spalten in Datensätzen

Verschlüsselung von zu löschen Spalten in Datensätzen

DSGVO

Verschlüsselung von zu löschen Spalten in Datensätzen

Aus dem CoC Projekt der PNW kriegt DWH täglich eine CoC Instanz, bestehend aus drei Tabellen die Verträge, Partner und Schadensnummer den Eigenschaften löschen, sperren,... zuordnen.

Agenda

Ausgangslage DWH

Anforderung CoC und DSGVO

Konzept und Realisierung CoC und DSGVO im DWH

Einführung

Ausblick

Konzept - Rahmenbedingungen

Einsatz SAS Federation Server für dynamische Realisierung der Anonymisierung und Verschlüsselung in Form von Views.

Alle Daten müssen unter Metadatenkontrolle stehen.

→ Auch das sog. DWH-Alt sollte unter Metadatenkontrolle mit einer neuen Namenskonvention für eindeutige Tabellennamen gestellt werden.

Die User haben dabei keine Berechtigungen auf die Basisdaten der Views. Insbesondere gibt es auch keinen Zugriff mehr auf SAS Libraries auf z/OS.

Libref und Metadatenobjekte (Bibliotheksname, Ordner, Tabellename und -struktur, Spalteneigenschaften wie Type, Format, Länge,) für den Anwender ändern sich nicht.

Hauptlast der Umstellung wird durch IT-Betrieb und Anwendungsentwicklung getragen!

Konzept

Übernahme AO-Schicht, DWH-Alt und KON-Schicht in den Federation Server. Weitere Schichten fungieren als Nutzer der Federation Server Bibliotheken.

Dezentral werden je Tabelle zwei Views als Join mit der CoC Instanz generiert. Bei allen in der CoC Instanz vorhandenen 1 – n Keyvariablen werden ausgewählte Variablen nach Vorgabe anonymisiert oder pseudonymisiert. Die Steuerung, welche Variablen anonymisiert oder pseudonymisiert werden, erfolgt über eine Steuertabelle.

AKTIV: Aktive Daten = keine Pseudonymisierung oder Anonymisierung

PASSIV: CoC relevante Daten = Pseudonymisiert oder Anonymisiert

Bestehende fachliche SAS Views werden durch fachliche Federation Server Views ersetzt.

Konzept

Je „alter“ Bibliothek existieren jetzt 3 Verzeichnisse, 3 Federation Server Schemata und 2 Metadatenbibliotheken:

1. Verzeichnis: Federation Server Schemata und Metadatenbibliothek, Metadatenregistrierte Basistabellen, zugriffsberechtigt nur ausgewählte AE, Produktion und Federation Server
2. Verzeichnis: Federation Server Schemata und Metadatenbibliothek Metadatenregistrierte CoC Views, zugriffsberechtigt alle auf die Daten berechtigten Anwender
3. Verzeichnis: Federation Server Schemata, Federation Server Bibliothek für Cache, zugriffsberechtigt alle auf die Daten berechtigten Anwender

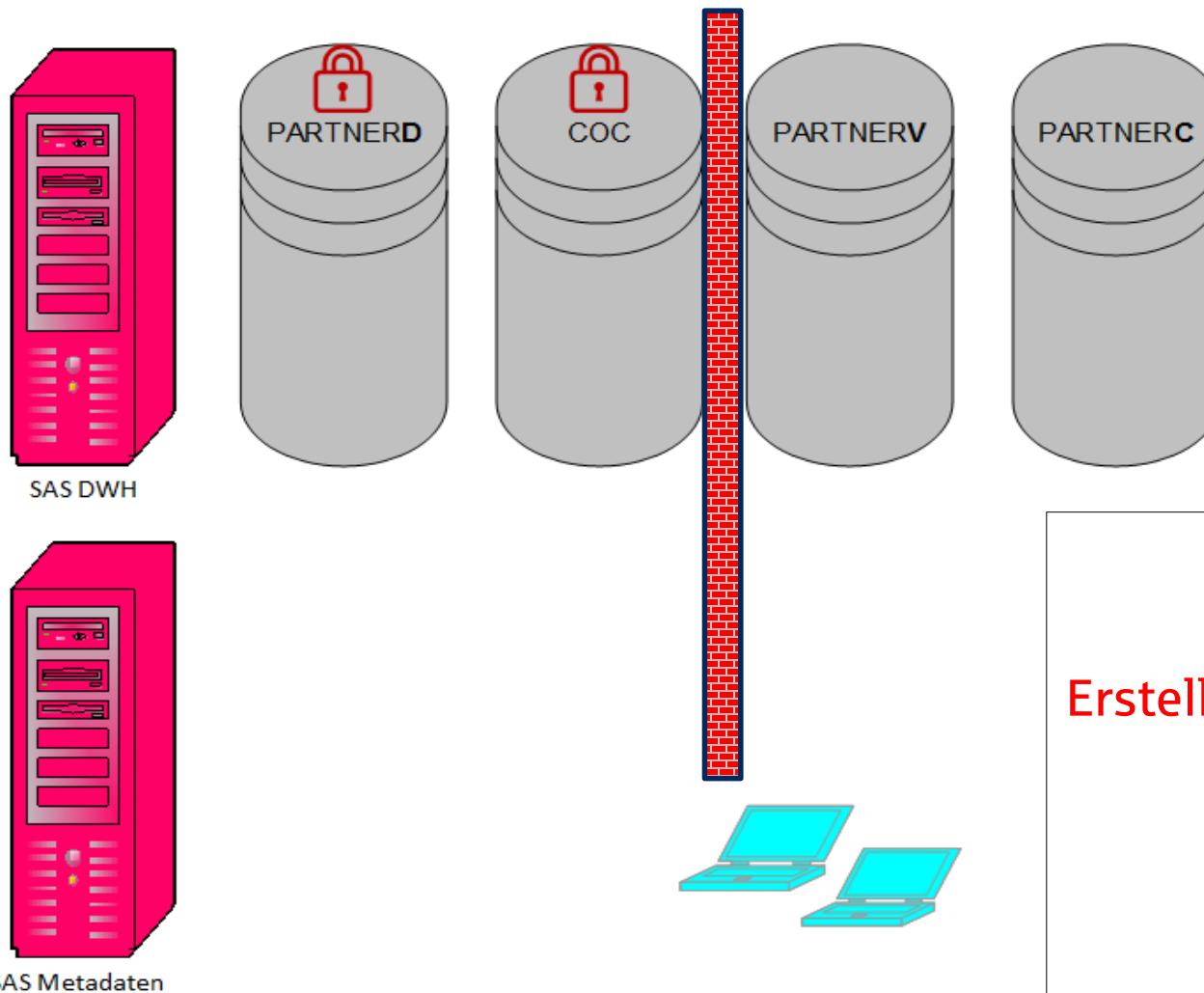
Integration in das bestehende Batchverfahren

Macro für den IT-Betrieb mit den Funktionen

- einheitliche Erstellung der Federation Server Strukturen je Serverkontext
 - Catalog
 - Schema
 - DSN-&serverkontext.-View
 - Authorisierungen auf die
 - Fedserver Instanz
 - Admin DSN
 - Base DSN
 - Catalog
 - Schema
 - Federation Server Optionen setzen (z.B. Tuning Parameter)

Realisierung Federation Server

Provinzial NordWest



Erstellen der Libs

Konzept

Macro für die Anwendungsentwicklung mit den Funktionen

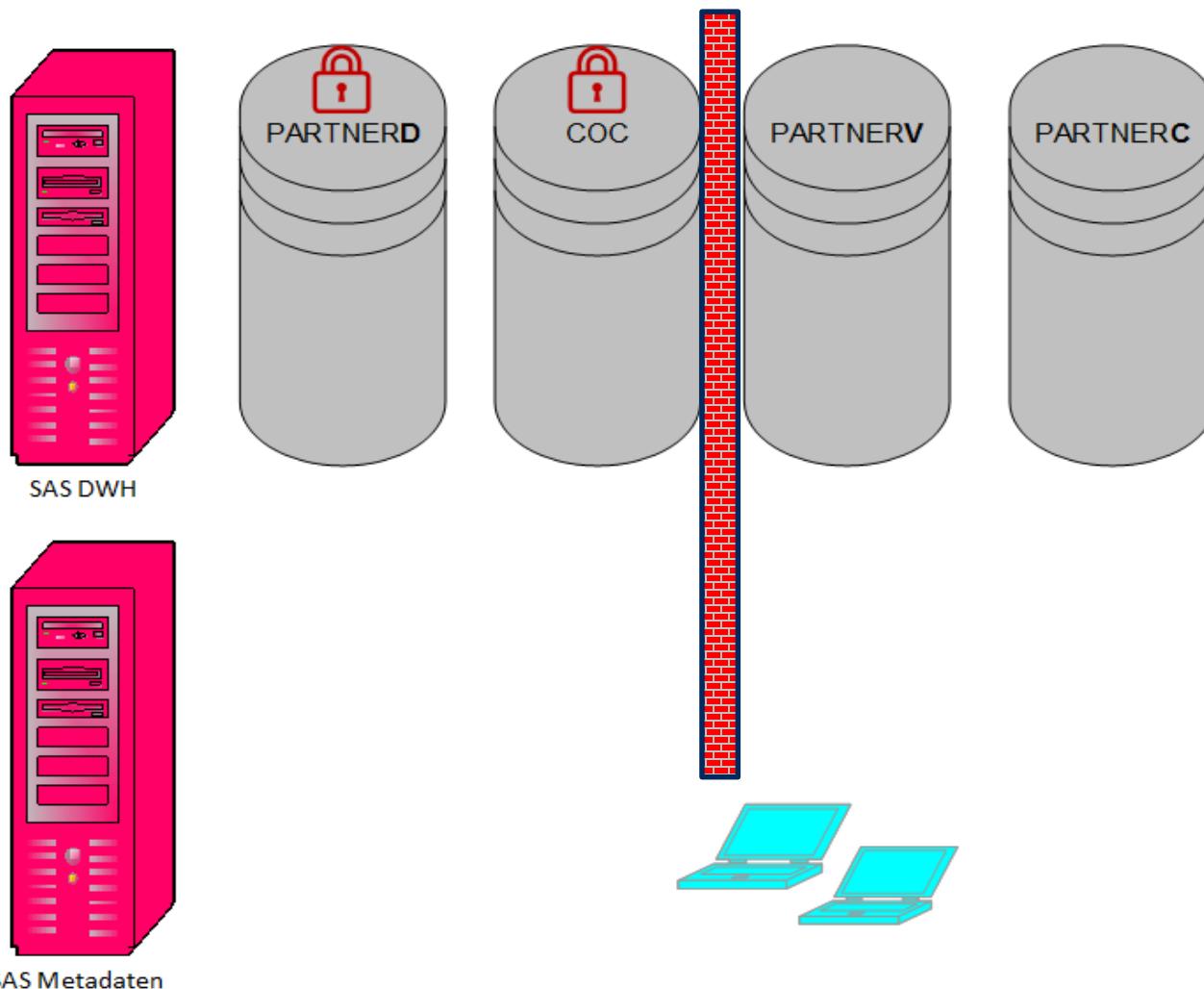
- einheitliche Viewerstellung AKTIV und PASSIV mit SAS
 - 1 – n Joinvariablen für Entscheidung AKTIV und PASSIV
 - Where Abfrage für die Erstellung der fachlichen Views
 - Attribute (Formate) werden aus den Metadaten gelesen
 - Registrieren und Deregistrieren in den Metadaten
 - Delete View
 - Maskieren, Anonymisieren und Pseudonymisieren mit unterschiedlichen Verfahren und für unterschiedliche Datentypen
- Create und Delete Cache
- Prüfung der Ausführungsberechtigung
- Umgebungsunabhängigkeit
- Eigenschaft für Datenzugriff auf Definer setzen

Herausforderungen

- Eineindeutige Anonymisierungen (Joinbarkeit) und Darstellbarkeit der Werte im SAS (HASH,INTEGER Wertebereich 18).
- Beibehaltung der Spaltenattribute (mindestens der Datentyp)
- Übernahme von SAS-Formaten und User defined Formaten
- Darstellung der SAS Federation Server Views als Views im Enterprise Guide (geht technisch, wurde aber zugunsten der preassigned Libs aufgegeben)
- Performance (Enable Resource Cache, Buffsize in Abhängigkeit von Anzahl großer/kleiner Tabellen)
- Korrekte Anzahl Observations in den Views
- Security

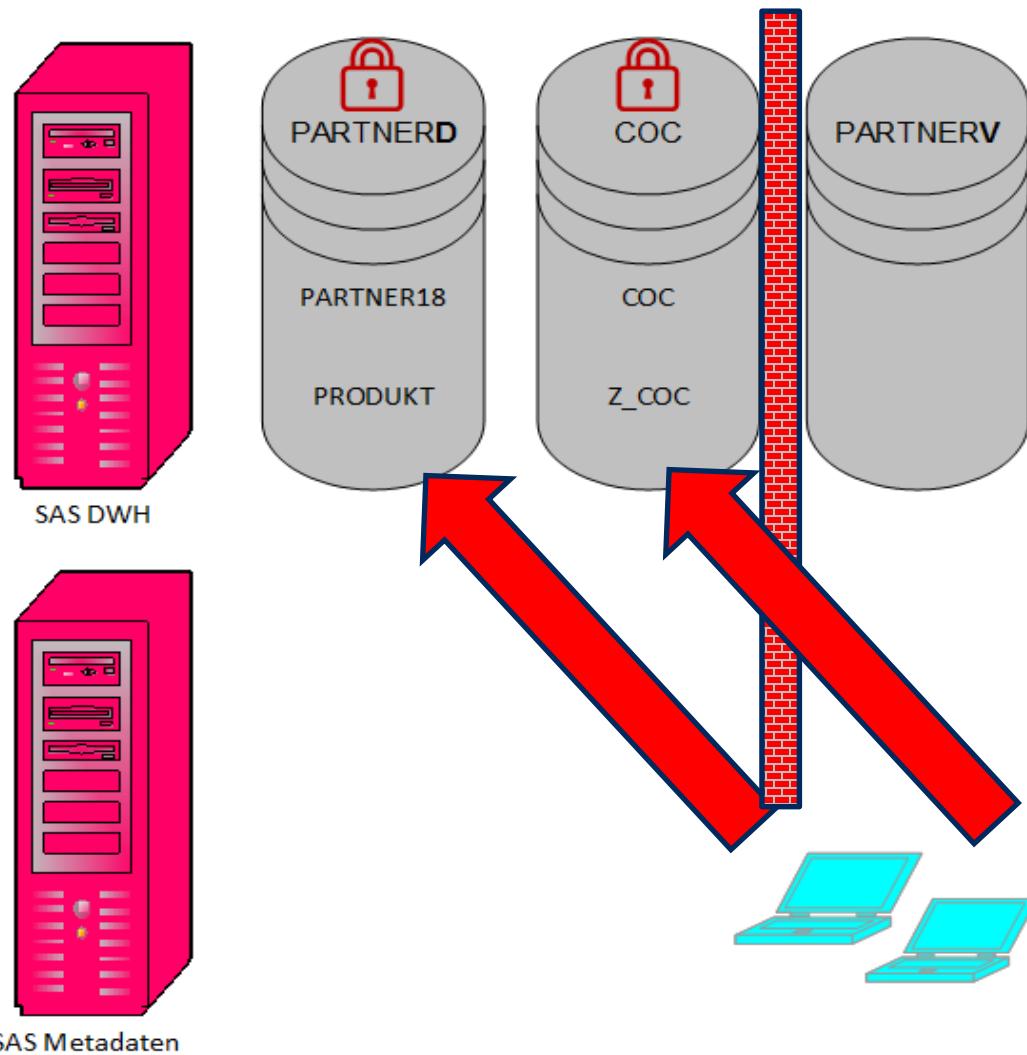
Realisierung Viewerstellung

Provinzial NordWest



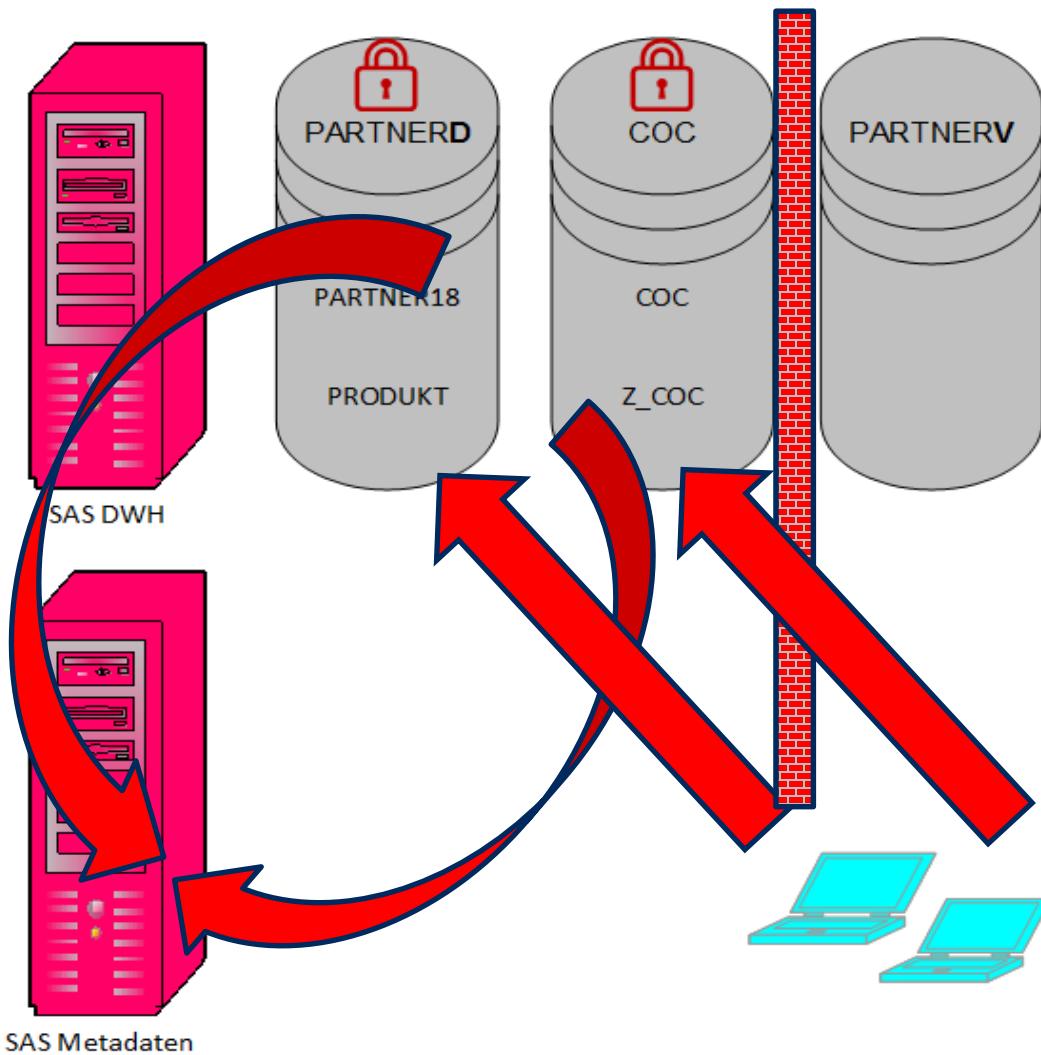
Realisierung Viewerstellung

Provinzial NordWest



Beladen der Schemata
PartnerD und COC mit
Tabellen!

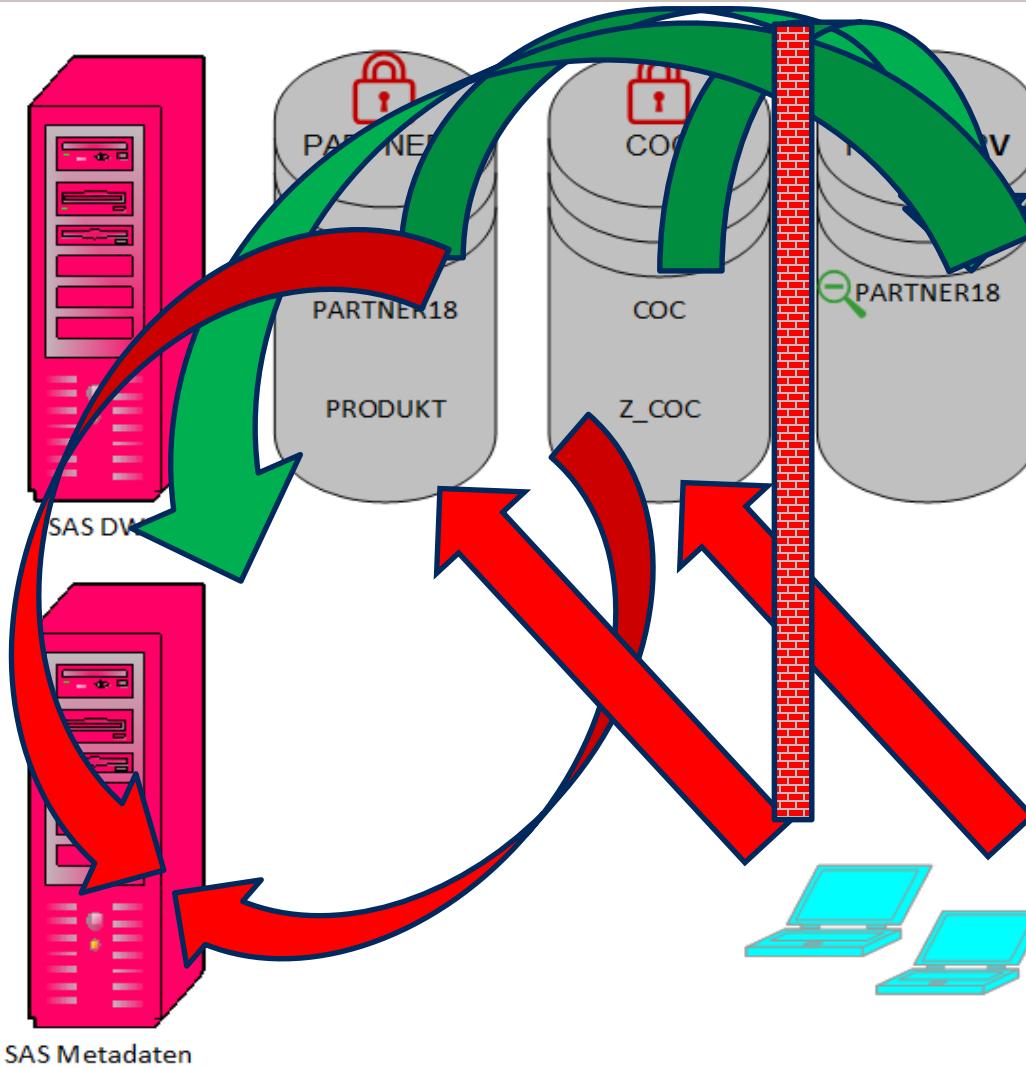
Realisierung Viewerstellung



Registrierung der
Tabellen in den
Metadaten

Realisierung Viewerstellung

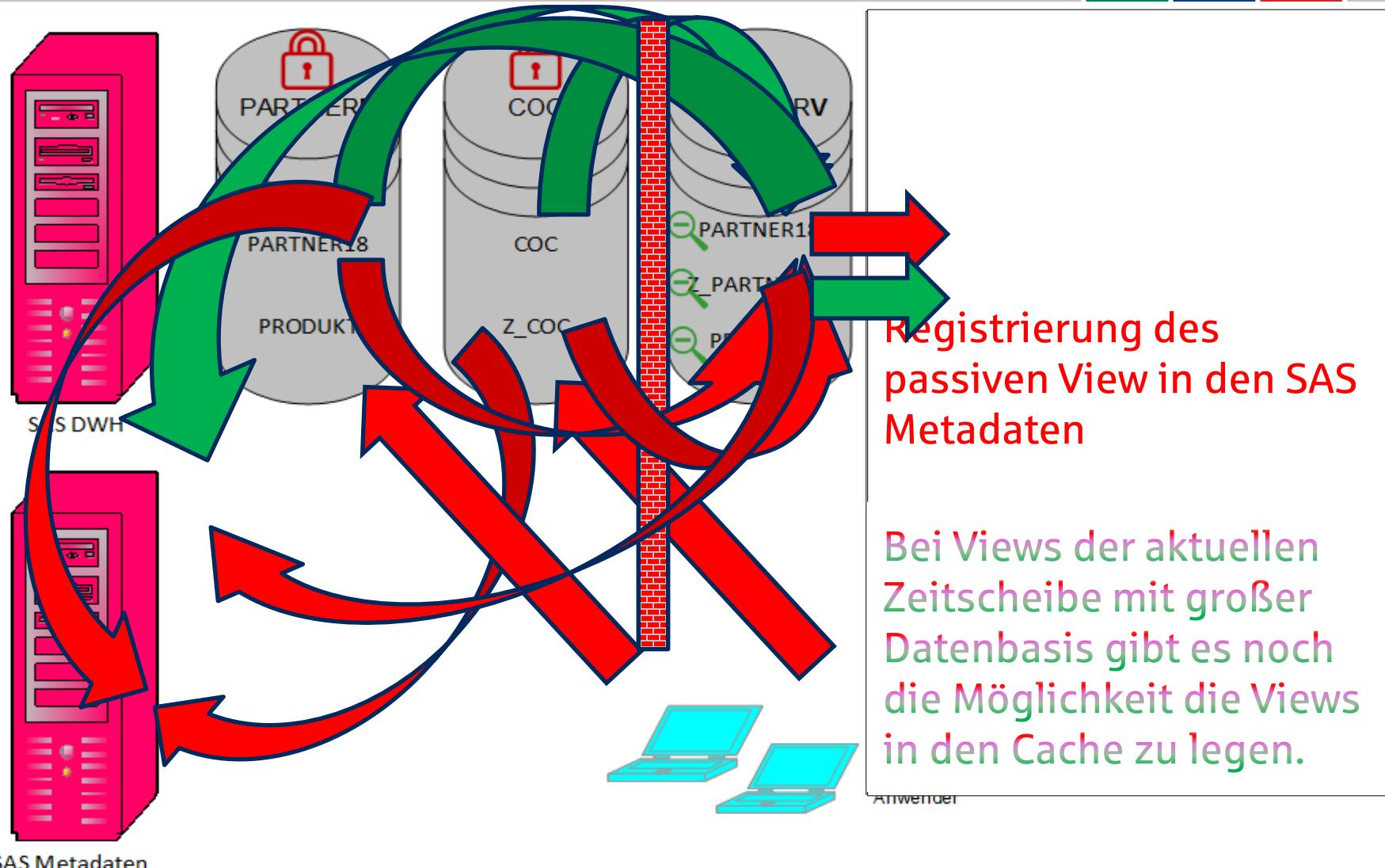
Provinzial NordWest



Erstellen eines Views mit
Wherebedingung
in PartnerV als Inner Join
zwischen COC & der
Tabelle Partner18

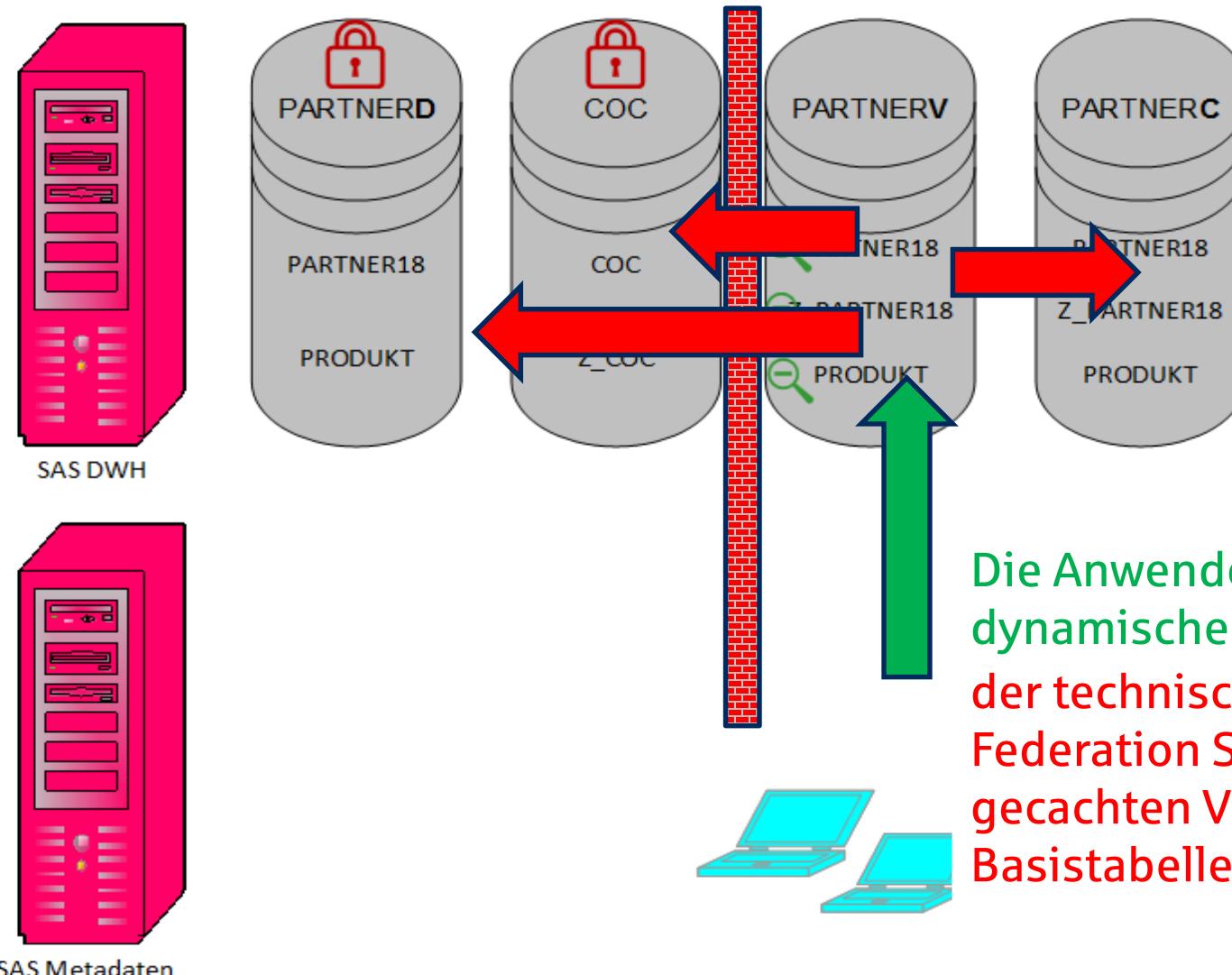
Realisierung Viewerstellung

Provinzial NordWest



Realisierung - Datenzugriff

Provinzial NordWest



Die Anwender greifen auf die dynamischen Fed-Views zu, der technische User des Federation Servers auf die gecachten Views oder die Basistabellen.

Großer Umstellungsaufwand in der Anwendungsentwicklung

Einzelne Data Connections im Data Management nicht mehr nötig

Umstellen STP's auf Federation Server Bibliotheken

Umstellen Zugriffe von Fremdanbietern auf SAS Tabellen auf ODBC to Federation Server

Anbinden der BI- und KON-Schicht an die Federation Server Bibliotheken (Views) als Anwender der Daten

Agenda

Ausgangslage DWH

Anforderung CoC und DSGVO

Konzept und Realisierung CoC und DSGVO im DWH

Einführung

Ausblick

2016	POC – Entscheidung Warten auf SAS 9.4 M3
2017	Aufbau des Prototypen in X für die AO-Schicht
	Hotfixing
	Verfahrensentwicklung zur Viewerstellung
	Verfahrensentwicklung Federation Server Objekterstellung
	Ausrollen in die DWHAlt und KON-Schicht
Jan. 18	Beginn Ausrollen nach Y
Feb. 18	Hotfixing und AO-Schicht mit Datenverschieben
Mrz. 18	DWHAlt

Einführung – Produktivgang in 2018

Provinzial NordWest

1. Mrz.	Plan für das Ausrollen nach P
30. Apr.	Aufbau neues Filesystem und Metadaten für DWHAlt
7. Mai.	Migration der DWHAlt Daten in neues Filesystem und Metadaten
9. Mai.	Hotfixing
10. Mai.	Redeploy Federation Server Manager
11. Mai.	Umbau AO-Schicht mit Datenverschieben und Aufbau Federation Server Objekte
	Erstellung der Federation Server Views nach Datumscheiben
12. Mai.	Aufbau Federations Server Schemata für DWHAlt
13. Mai.	Erstellung der Federation Server Views nach Datumscheiben
14. Mai.	Umstellung weiterer Datenzugriffe auf Federation Server Views
ab 14. Mai.	Performanceprobleme
ab 14. Mai.	Neuerstellte Views werden nicht auf DEFINER gesetzt

Einführung-In Zahlen

Provinzial NordWest

Schichten	Datenverzeichnisse	Metadaten-registriert	Tabellen	Fed Views	1.1 FedViews	fachliche FedViews	Metadaten-registrierte Tabellen & FedViews
BI-Schicht	90	60	769	0	0	0	769
KON-Schicht	66	44	6.901	0	0	0	6.901
AO-Schicht	153	102	13.219	40.500	26.438	14.062	53.719
DWH-Alt	15	10	38.496	76.992	76.992	0	115.488
SOL-Schicht	21	14	275	0	0	0	275
Summe	345	230	59.660	117.492	103.430	14.062	177.152
Summe vor COC	6.153	110	59.660	0	0	0	26.259

Agenda

Ausgangslage DWH

Anforderung CoC und DSGVO

Konzept und Realisierung CoC und DSGVO im DWH

Einführung

Ausblick

=> Bearbeitung der aktuellen Probleme

Ausrollen des Federations Servers in die KON-Schicht

Row Level Security

Column Level Security

Eventuell die BI Daten, SOL-Schicht und angebundene Datenbanken unter Federation Server Kontrolle stellen.

Fachliche Views ohne CoC-Instanz bzw. komplexere fachliche Views

Übernahme weiterer Spalteneigenschaften wie Label, Länge

Vermehrter Einsatz von Caching

Herzlichen Dank für Ihre Aufmerksamkeit!

