

COMMERZBANK



## Performance durch Parallelisierung - mit dem SPD-Server und dem SAS® Grid Manager

Am Beispiel des Performance Managers der Commerzbank AG

Lothar Nottekämper Bonn / 07.10.2010

COMMERZBANK

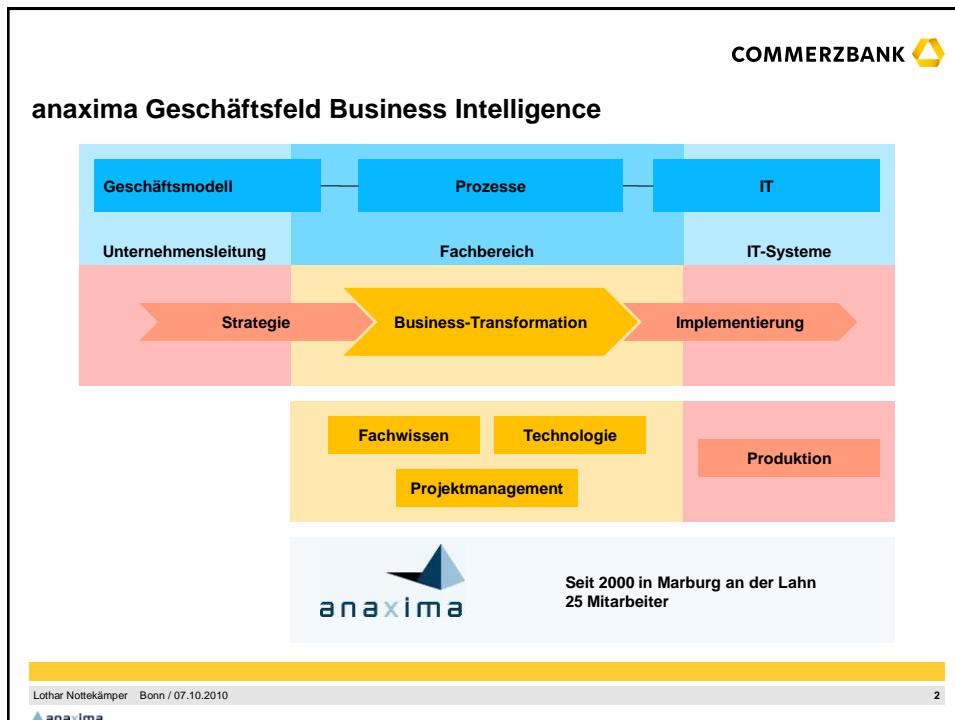
## Privat- und Geschäftskunden

Stategische Positionierung	Eine der führenden Banken im Privatkundengeschäft in Deutschland mit mehr als 11 Millionen Kunden Commerzbank und Dresdner Bank bilden ein umfassendes deutschlandweites Netzwerk mit über 1200 Filialen
Erfolgs-faktoren	<b>Nachhaltige Ergebnisentwicklung und Ausnutzung von Synergieeffekten:</b> <ul style="list-style-type: none"><li>› Optimierung des Filialnetzes</li><li>› Nachhaltige Ertrags situation: Die Commerzbank profitiert von ihrem festen Kundenstamm, einem High-End Produktp o r t f o l i o s i e r e n d e r e n g e n P a r t n e r s c h a f t m i t d e r A l l i a n z g r u p p e</li><li>› Kostenreduzierung durch Vereinfachung der Verwaltungsabläufe</li><li>› Umfassende Betreuung durch 10.000 Berater und Spezialisten</li></ul>
Strategische Ziele	<ul style="list-style-type: none"><li>› Durch die erfolgreiche Integration der Dresdner Bank sollen anspruchsvolle Privatkunden gebunden werden</li><li>› Überdurchschnittliche Partizipation an langfristigen Wachstumstrends (Demografische Entwicklung, Entwicklung der Sparquote)</li></ul>

Lothar Nottekämper Bonn / 07.10.2010

1

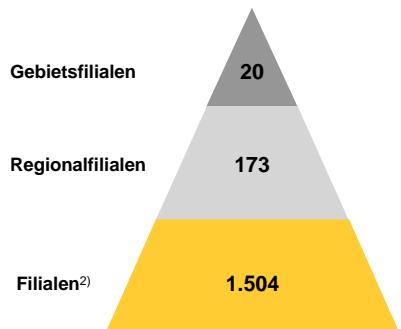
 anaxima



### Abbildung Filialstruktur und Betreuungsmodell

Filialstruktur der Commerzbank AG in Deutschland per 1. Januar 2010

**Privat- und Geschäftskunden**  
1.695 Standorte<sup>1)</sup>



#### Komplexität durch

- › Steuerungsinformationen über alle Hierarchieebenen bereitstellen
- › Abbildung unterschiedlicher Filialtypen, wie "Filiale der Zukunft" und Beratungszentren
- › Integration der Dresdner Bank Filialen
- › Abbildung Kundensegmente
- › Berücksichtigung Betreuungsmodell (Generalisten und Spezialisten)
- › Mitarbeiter-Vorgesetzten Beziehung und Vertretungsregelungen
- › Dynamische Anpassung an Änderungen im Filialnetz – Erfordert Datenreorganisation

<sup>1)</sup> Dresdner Bank Filialen inkl. <sup>2)</sup> inkl. Beratungszentren

Lothar Nottetkämper Bonn / 07.10.2010

4

anaxima

### Integrierte Steuerung des gesamten Produktportfolios

#### Anforderungen

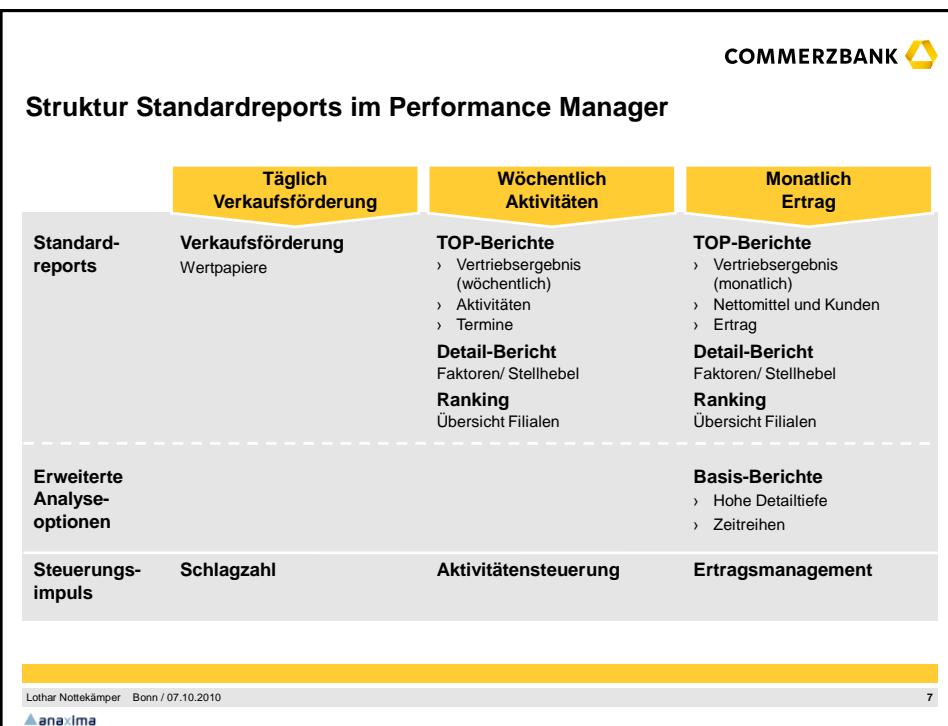
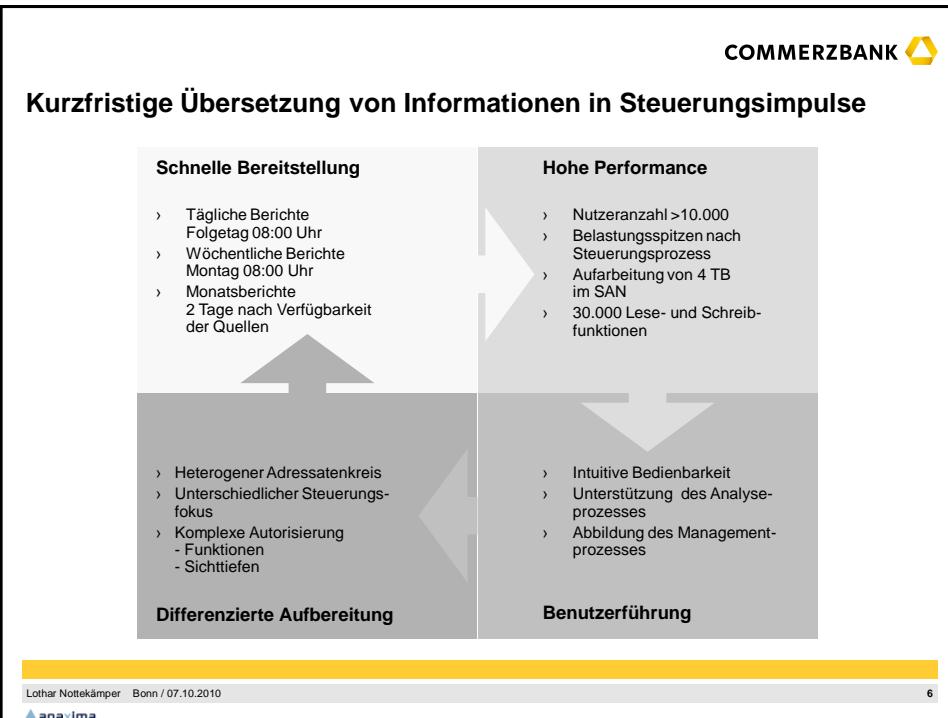
**Ziel**  
Vollständige und einheitliche Abbildung aller steuerungsrelevanten Informationen

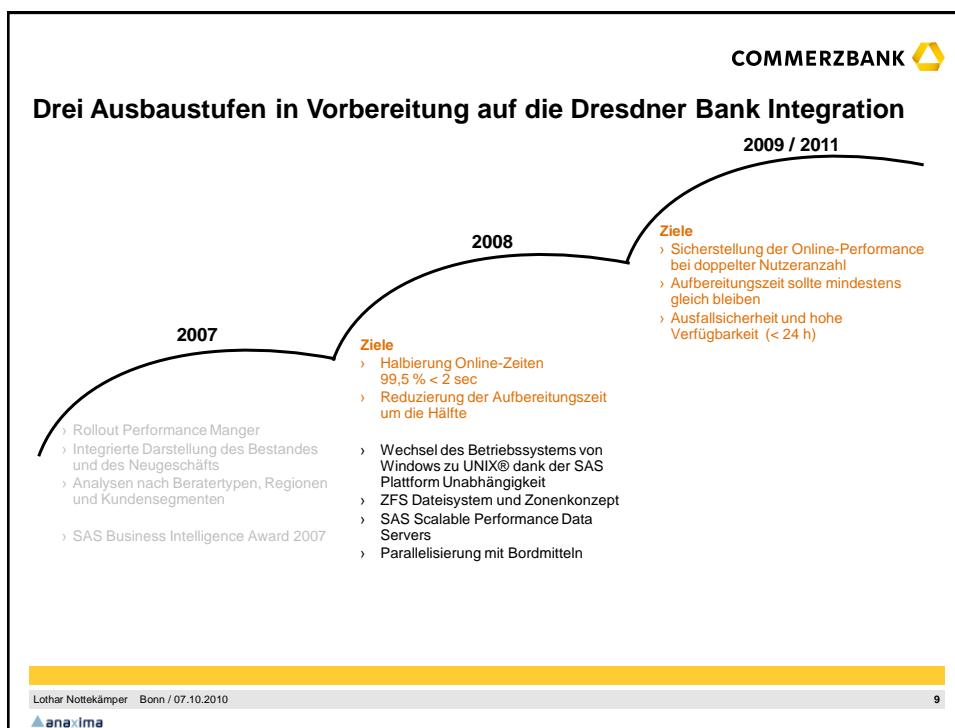
- › Abbildung der strategischen Produktfelder
  - Vermögensanlage
  - Vorsorge
  - Einlagen
  - Kreditgeschäft
  - Zahlungsverkehr
- › Hohe Flexibilität bei Einführung neuer Produkte
- › Abbildung von Neugeschäfts- und Bestanderträgen
- › Bereitstellung von Erträgen, Volumina und Margen
- › Hohe Breite und Tiefe der verfügbaren Informationen
- › Fokussierung auf Key Performance Indicators

Lothar Nottetkämper Bonn / 07.10.2010

5

anaxima





**COMMERZBANK**

### SPD – Scalable Performance Data Server - Grundprinzip

High Performance Data Server-Anwendung von SAS

- › Parallelverarbeitende SQL-Engine (Multi-Threading) mit speziellen Optimierungen für schnellste Abfrage (Star Join Optimizer, ...)
- › SAS Tabellen werden physikalisch aufgeteilt in n Datenkomponenten, die wiederum verteilt abgelegt werden können über mehrere Filesysteme (Indexes separat von Daten und Work-Bereich z.B.)
- › Anschluss über schnelle Netzcontroller stellen hohen Durchdurchsatz sicher



› Clustertabellen-Konzept für schnellstes paralleles Laden

› Beschleunigte Abfragen, da nur relevante Cluster Teil-Tabellen verarbeitet werden

Jan 2003	Feb 2003	Mar 2003	Apr 2003	Jan 2005	ETL1
May 2003	Jun 2003	Jul 2003	Aug 2003	Feb 2005	ETL2
Sep 2003	Oct 2003	Nov 2003	Dec 2003	Mar 2005	ETL3
Jan 2004	Feb 2004	Mar 2004	Apr 2004	Apr 2005	ETL4
May 2004	Jun 2004	Jul 2004	Aug 2004	May 2005	ETL5
Sep 2004	Oct 2004	Nov 2004	Dec 2004	Jun 2005	ETL6

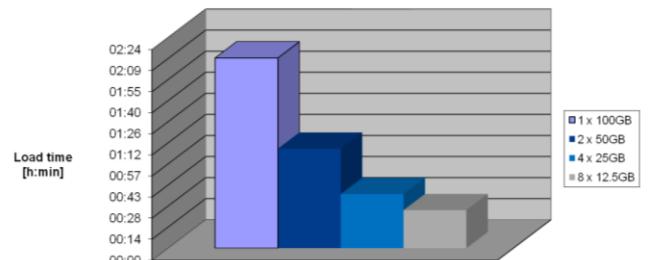
Lothar Nottetkämper Bonn / 07.10.2010

10



**COMMERZBANK**

### SPD-Server – Clustertabellen Ladezeiten

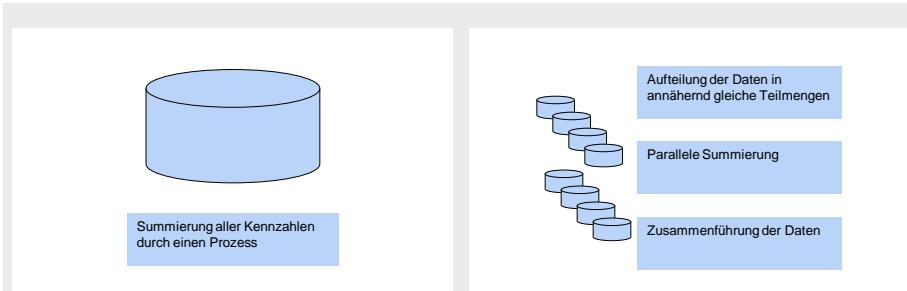


Lothar Nottetkämper Bonn / 07.10.2010

11



## Parallelisierung durch Aufteilung

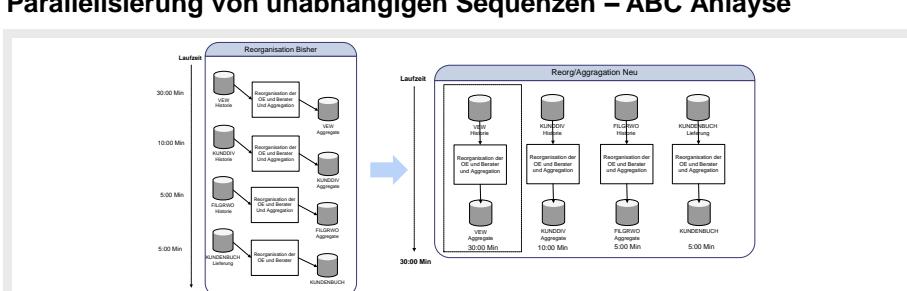


- > Abhängigkeiten zwischen den Teilmengen dürfen nicht existieren
- > Gleichverteilung der resultierenden Mengen muss möglich sein
- > Overhead für die Aufteilung und das Zusammenführen darf den Parallelisierungseffekt nicht übersteigen.

Lothar Nottetkämper Bonn / 07.10.2010 12



## Parallelisierung von unabhängigen Sequenzen – ABC Analyse



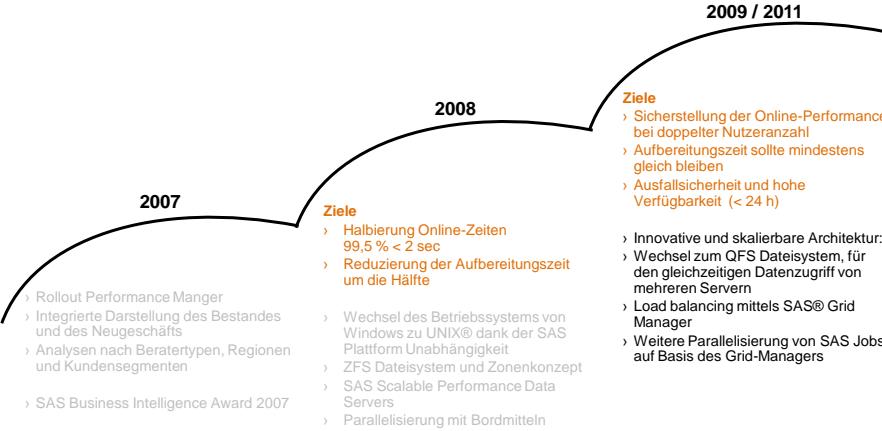
- > Abhängigkeiten innerhalb der Jobketten identifizieren
- > Die Verteilung der Teilschritte verkürzt die Laufzeit auf die Dauer des längsten Einzelschritts
- > Die Schritte wurden nach dem erwarteten Einsparungseffekt und dem Umsetzungsaufwand priorisiert.
- > Maßnahmen mit dem größten Effekt und dem geringsten Umsetzungsaufwand werden als erstes getestet und umgesetzt.
- > Werden die Ziele für die Laufzeit erreicht, wird eine Neubewertung der weiteren Maßnahmen vorgenommen und die Umsetzung geprüft.

Lothar Nottetkämper Bonn / 07.10.2010 13



**COMMERZBANK**

### Drei Ausbaustufen in Vorbereitung auf die Dresdner Bank Integration



2007	2008	2009 / 2011
<ul style="list-style-type: none"> <li>› Rollout Performance Manager</li> <li>› Integrierte Darstellung des Bestandes und des Neugeschäfts</li> <li>› Analysen nach Beratertypen, Regionen und Kundensegmenten</li> <li>› SAS Business Intelligence Award 2007</li> </ul>	<p><b>Ziele</b></p> <ul style="list-style-type: none"> <li>› Halbierung Online-Zeiten 99,5 % &lt; 2 sec</li> <li>› Reduzierung der Aufbereitungszeit um die Hälfte</li> </ul>	<p><b>Ziele</b></p> <ul style="list-style-type: none"> <li>› Sicherstellung der Online-Performance bei doppelter Nutzeranzahl</li> <li>› Aufbereitungszeit sollte mindestens gleich bleiben</li> <li>› Ausfallsicherheit und hohe Verfügbarkeit (&lt; 24 h)</li> </ul>
		<ul style="list-style-type: none"> <li>› Innovative und skalierbare Architektur:</li> <li>› Wechsel zum QFS Dateisystem, für den gleichzeitigen Datenzugriff von mehreren Servern</li> <li>› Load balancing mittels SAS® Grid Manager</li> <li>› Weitere Parallelisierung von SAS Jobs auf Basis des Grid-Managers</li> </ul>

Lothar Nottetkämper Bonn / 07.10.2010

 14

**COMMERZBANK**

### Entscheidungssituation

**Ausgangssituation**

- › Anzahl Anfragen und Datenmengen werden sich verdoppeln
- › Gute Erfahrungen mit SUN Fire X4600 AMD Plattform
- › Hoher Datendurchsatz unter ZFS erlaubt schnelle Laufzeiten
- › Allerdings war keine weitere Skalierung der Plattform möglich
- › Hohe Kosten bei der Umstellung auf notwendige Sparc Architektur

**Entscheidung**

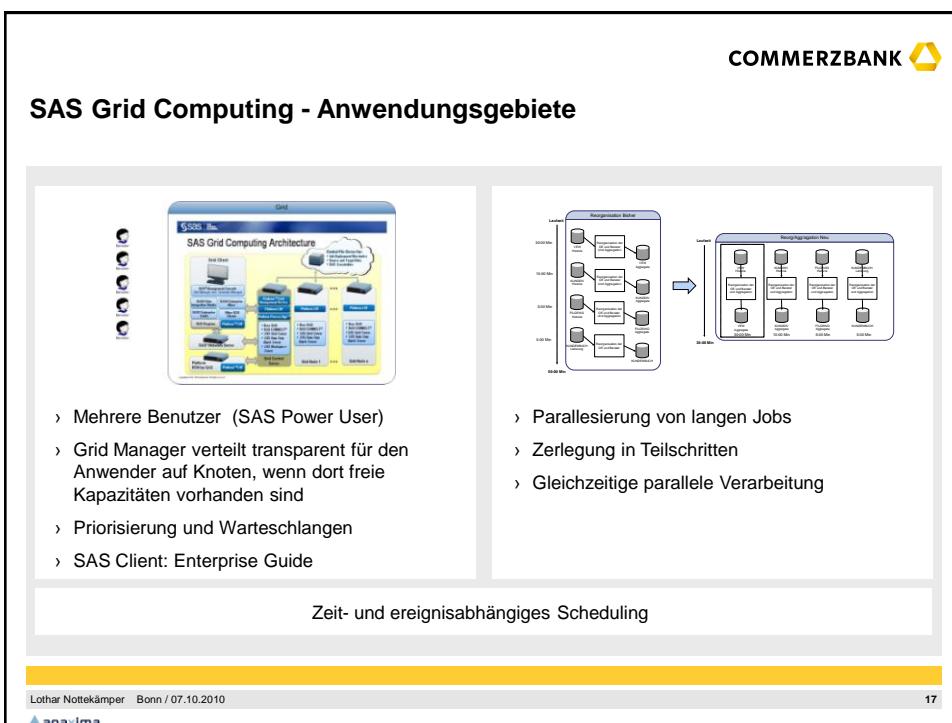
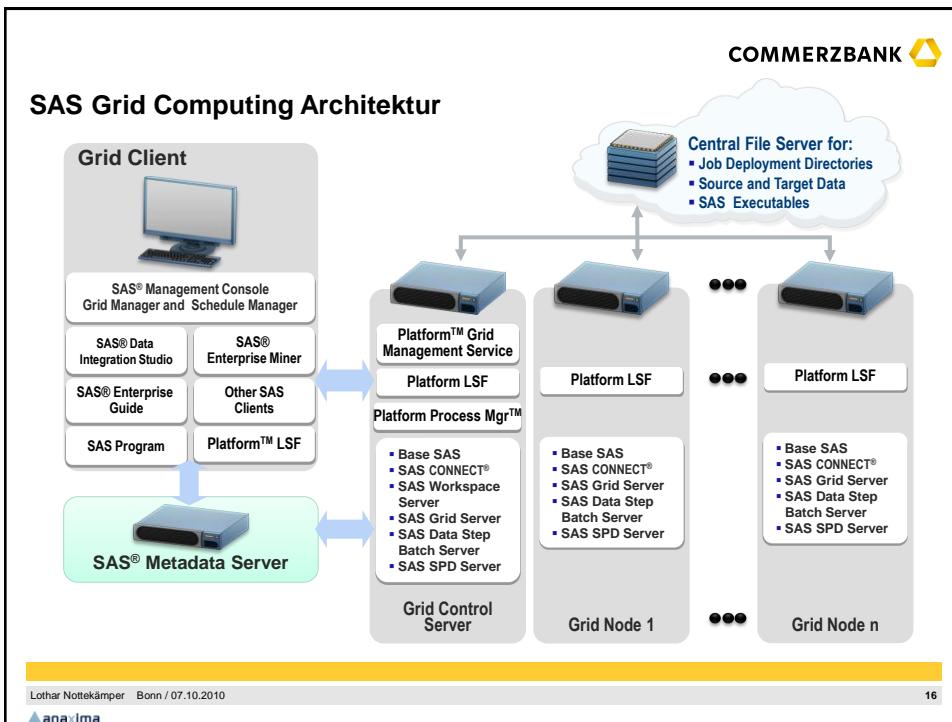
- › Anschaffung von 4 Servern
- › QFS erlaubt direkten Zugriff von allen 4 Servern
- › Erwartung: Parallele Verarbeitung mit hohem Datendurchsatz

**Lösung**

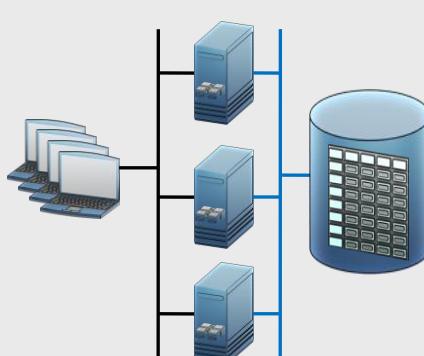
SPD-Server mit Grid-Deployment

Lothar Nottetkämper Bonn / 07.10.2010

 15



## SPD-Server Grid Deployment – Gemeinsamer Zugriff auf Daten



- › Mehrere SMP-Hosts jeweils bestückt mit einem SPD-Server
- › Multipler Zugriff auf Daten im SAN ist erforderlich
  - Filesystem = QFS
- › Client Zugriffe werden über mehrere Knoten skaliert
- › Gemeinsamer lesender Zugriff auf gleiche Tabellen
- › Schreibender Zugriff nur auf verschiedenen Tabellen
- › Hochverfügbarkeit und Transparenz für den Anwender
- › Anspruchsvolles Deployment

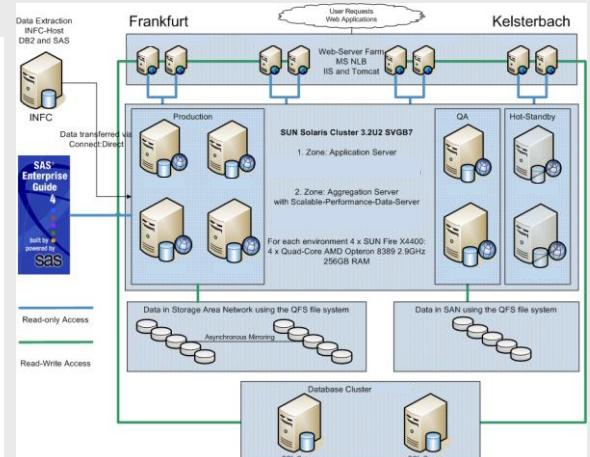
## SPD-Server Grid Deployment – Vorteile der Lösung

- › System mit hohem Datendurchsatz auf allen Servern
- › Kostengünstige Hardware
- › Ausfallsicherheit und Verfügbarkeit durch Zonenkonzept und QFS
- › Keine umfangreichen Programmumstellungen notwendig
- › Alle Metadaten für die Parallelisierung im Metadatenserver
- › Leichte Erweiterbarkeit durch das Hinzufügen von kleinen Einheiten falls notwendig

## Übersicht über die SAS Plattform als Ganzes

### Architekturprinzipien

- Asynchrone Spiegelung an zwei Standorten**
  - > Disaster management
  - > Verfügbarkeit
- Schichtenmodell mit klarer Trennung**  
Einfacher Ausbau und Upgrading
- Zonenkonzept**
  - > Im Falle einer Fehlfunktion Schwenk auf anderen Cluster
  - > Maintenance am "live"-System
- QFS**  
Mehrerer Server können auf der gleichen Datenbasis arbeiten
- SAS® Grid Manager mit 4 SPD Servern**  
Optimierte parallele Verarbeitung



Frankfurt

Kelsterbach

User Requests Web Applications

Web-Server Farm MS NLB IIS and Tomcat

INFC Data Extraction INF-C-Host DB2 and SAS

SUN Solaris Cluster 3.2U2 S VGB7

1. Zone: Application Server

2. Zone: Aggregation Server with Scalable-Performance-Data-Server

For each environment 4 x SUN Fire X4600: 4 x Quad-Core AMD Opteron 8389 2.9GHz 256GB RAM

Data in Storage Area Network using the QFS file system Asynchronous Mirroring

Data in SAN using the QFS file system

Database Cluster

SQL-Server

SQL-Server

Read-only Access

Read-Write Access

SAS Enterprise Guide 4

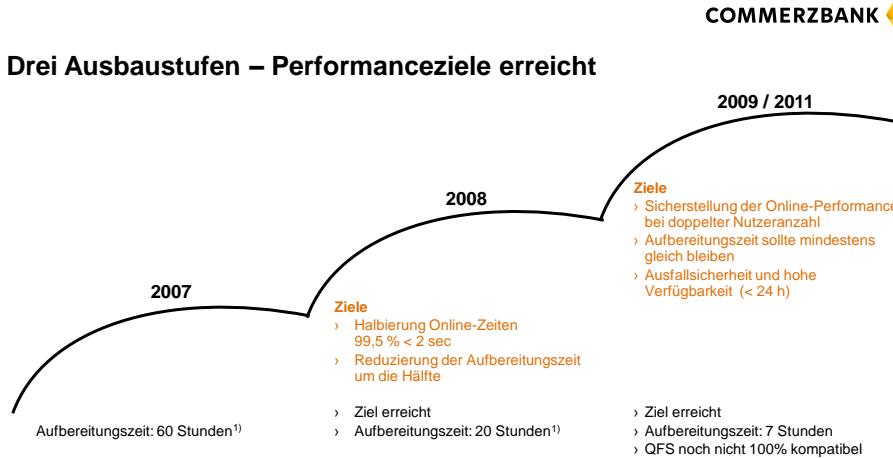
Not live yet covered by SAS

Lothar Nottekämper Bonn / 07.10.2010

ana:ima

20

## Drei Ausbaustufen – Performanceziele erreicht



2007

Aufbereitungszeit: 60 Stunden<sup>1)</sup>

2008

Ziele

- > Halbierung Online-Zeiten  
99,5 % < 2 sec
- > Reduzierung der Aufbereitungszeit um die Hälfte

> Ziel erreicht

> Aufbereitungszeit: 20 Stunden<sup>1)</sup>

2009 / 2011

Ziele

- > Sicherstellung der Online-Performance bei doppelter Nutzeranzahl
- > Aufbereitungszeit sollte mindestens gleich bleiben
- > Ausfallsicherheit und hohe Verfügbarkeit (< 24 h)

> Ziel erreicht

> Aufbereitungszeit: 7 Stunden

> QFS noch nicht 100% kompatibel mit Zonenkonzept

<sup>1)</sup> Aufbereitungszeit ist auf das heutige Volumen hochgerechnet

Lothar Nottekämper Bonn / 07.10.2010

ana:ima

21

## Umsetzung

### Analyse der Ausgangssituation

- › Anzahl Anfragen und Datenmengen
- › Erwartete Zuwachsraten
- › Art der Performance Herausforderungen an Hand der Anwendungsszenarien

### Entscheidung

- › Hardware
- › Softwarekomponenten
- › Test und Durchstich

### Lösungen

- › Code-Optimierung
- › SPD-Server
- › Grid-Manager

### Kosten-Nutzen-Analyse

Business Case

Lothar Nottetkämper Bonn / 07.10.2010

22

 anaxima