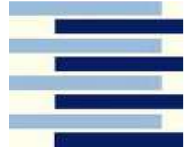


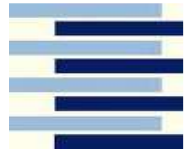
# WP Datenbankdesign



## Kapitel 1: Einführung

- **State of the Art**
- **Aktuelle Trends**
- **Datenbankentwurf**

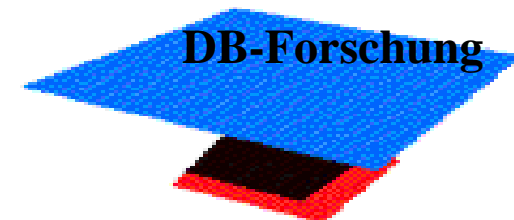
## Sie haben gelernt ...



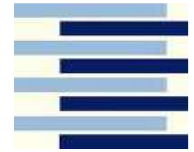
- 1. bis 3. Normalform  
kann man vergessen: NF<sup>2</sup> oder 4. NF
- Transaktionsprinzip ACID  
kann man vergessen: Flashback
- Views  
Kann man (fast) vergessen: Materialized Views
- Relationale Datenbank  
kann man bei Geo-Datenbanken vergessen
- **Trotzdem sollen Sie Ihr Wissen aus DB nicht über Bord werfen**



**Was bewegt die Datenbankszene?**



# Stand der Technik



**Sun Oracle Database Machine**

Datenbankserver Quad Core Xenon mit bis zu 72 GB HSP

1 Rack mit max. 8 Servern

1 – 8 Exadata Storage Server mit bis zu 100 TB Kapazität (**S**erial **A**ttached **S**CSI)

Bis zu 5 TB Flash Cache

Bis zu 1 Mio. I/O Operationen / Sek.

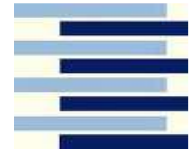
I/O-Bandbreite bis zu 50 GB/Sek.

Spezielle Indizes und optimierte Scans.

Kosten: > 1 Mio \$ pro Rack ohne DBMS

**big data**

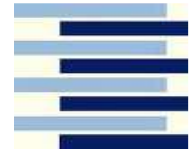
# Die größten Datenbanken der Welt



10. **Library of congress** mehr als 130 Mio. Bücher, Fotos, ..., ca. 1.000 Km Regale,
8. **Amazon** ca 50. TB, regelmäßig 60 Mio. aktive User
7. **Youtube** > 60 Std. Videoupload/Min., 45 TB, 100 Mio. Video-Aufrufe/Tag, 2 TB/Monat neu
3. **at&t** 1,9 Billionen Gesprächs-Datensätze/Tag, 323 TB
2. **NERSC** National Energy Research Scientific Computer Center, 2,8 PB über Atomenergie-Forschung, Simulationen des Universums, ...
1. **MPI für Klimaforschung und Meteorologie**, 220 TB mit Internet-Zugriff + 110 TB Klimasimulation + 6 PT auf Band

<http://www.comparebusinessproducts.com/fyi/10-largest-databases-in-the-world>, Stand März 2010

# Es geht inzwischen noch größer (2013)



## Bluffdale – Die größte Datenhalde der Welt Prism

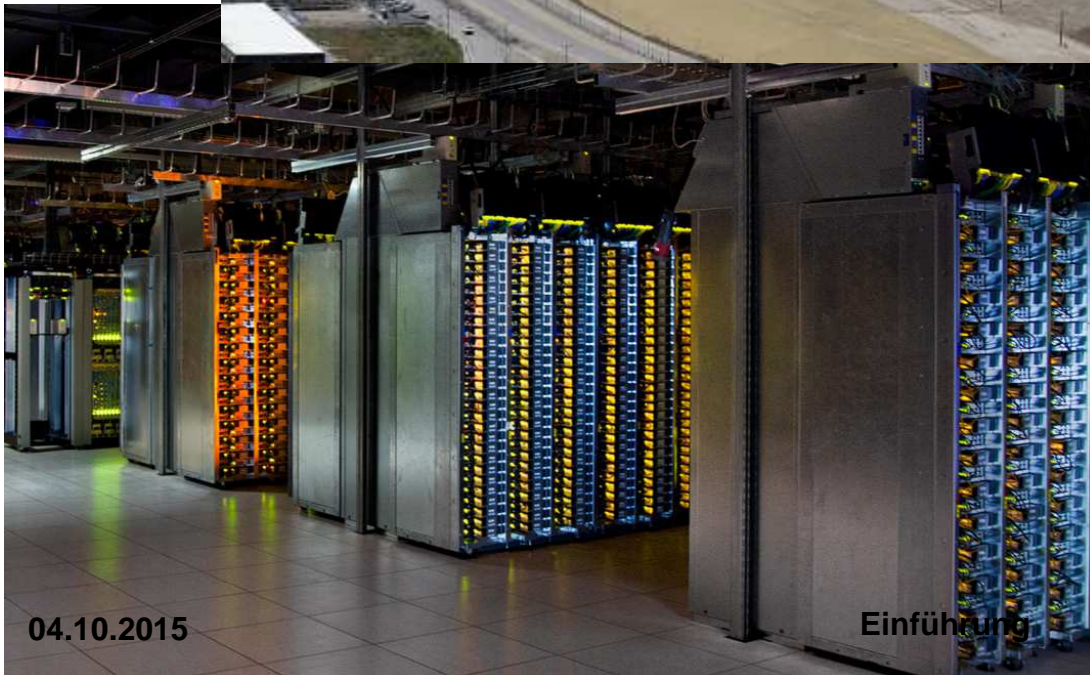
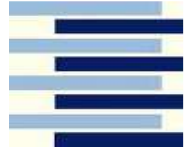


Im September 2013 will die National Security Agency (NSA) einen gigantischen Serverkomplex in Betrieb nehmen, der derzeit in der unwirtlichen Weite Utahs nahe dem Städtchen Bluffdale entsteht. Der „Utah Data Center“ genannte Komplex soll weltweit möglichst viele Daten zur Auswertung sammeln.

Auf insgesamt 300.000 Quadratmetern rund 30.000 Quadratmeter Stellfläche für Server bieten und will Daten in der Größenordnung von einem Yottabyte vorhalten können. Das entspricht der schwer vorstellbaren Menge von einer Billion Terabytes. Die NSA braucht für ihr Vorhaben solche Kapazitäten, denn laut einer Analyse von Cisco Systems soll sich der globale Datenverkehr zwischen 2010 und 2015 auf rund 966 Exabytes pro Jahr vervierfachen. In dem rund 2 Milliarden US-Dollar teuren Komplex – unklar ist, ob hier die Computerhardware eingerechnet ist – soll möglichst der gesamte Datenfluss des Planeten gesammelt und ausgewertet werden. Dazu zählen die Inhalte von E-Mails, Telefongesprächen, Sprach- und Videochats sowie aller Google-Suchen. Die Speicherrate ist ca. 20 TB/Min. Der selbstversorgende Komplex wird für drei Tage Energie vorhalten und 1,7 Millionen Liter Wasser täglich pumpen können. Die Stromrechnung der Anlage wird in der Größenordnung von 40 Millionen US-Dollar pro Jahr liegen.

<http://www.computerbase.de/2013-06/bluffdale-die-groesste-datenhalde-der-welt/>

# So sieht das heute aus

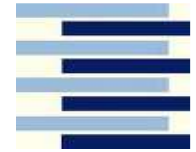


04.10.2015

Einführung



# Beispiel Geo-Datenbanksysteme



Speicherung von Geo-Daten in einem RDBMS, Abfrage, Darstellung und Verknüpfung mit strukturierten Daten.

## Datentyp SDO\_GEOMETRY

Tabelle Laender

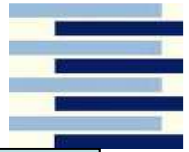
ID	NAME	GRENZE
NUMBER	VARCHAR2	SDO_GEOMETRY



Welche Bundesländer grenzen an NRW?

```
select l1.name
from laender l1, laender l2
where l2.name='NRW' and
sdo_relate( l1.grenze, l2.grenze, 'mask=touch')='TRUE'
```

# Beispiel Hochverfügbarkeit



## 1. Welche Ausfallzeiten können Sie tolerieren?

- Bedenken Sie: Es handelt sich womöglich um den Zeitpunkt, an dem die Hauptlast erzeugt wird (z.B. Handel: Samstagmittag in der Weihnachtszeit)

## 2. Welchen Datenverlust können Sie tolerieren?

- Gar keinen – ist kaum realisierbar!

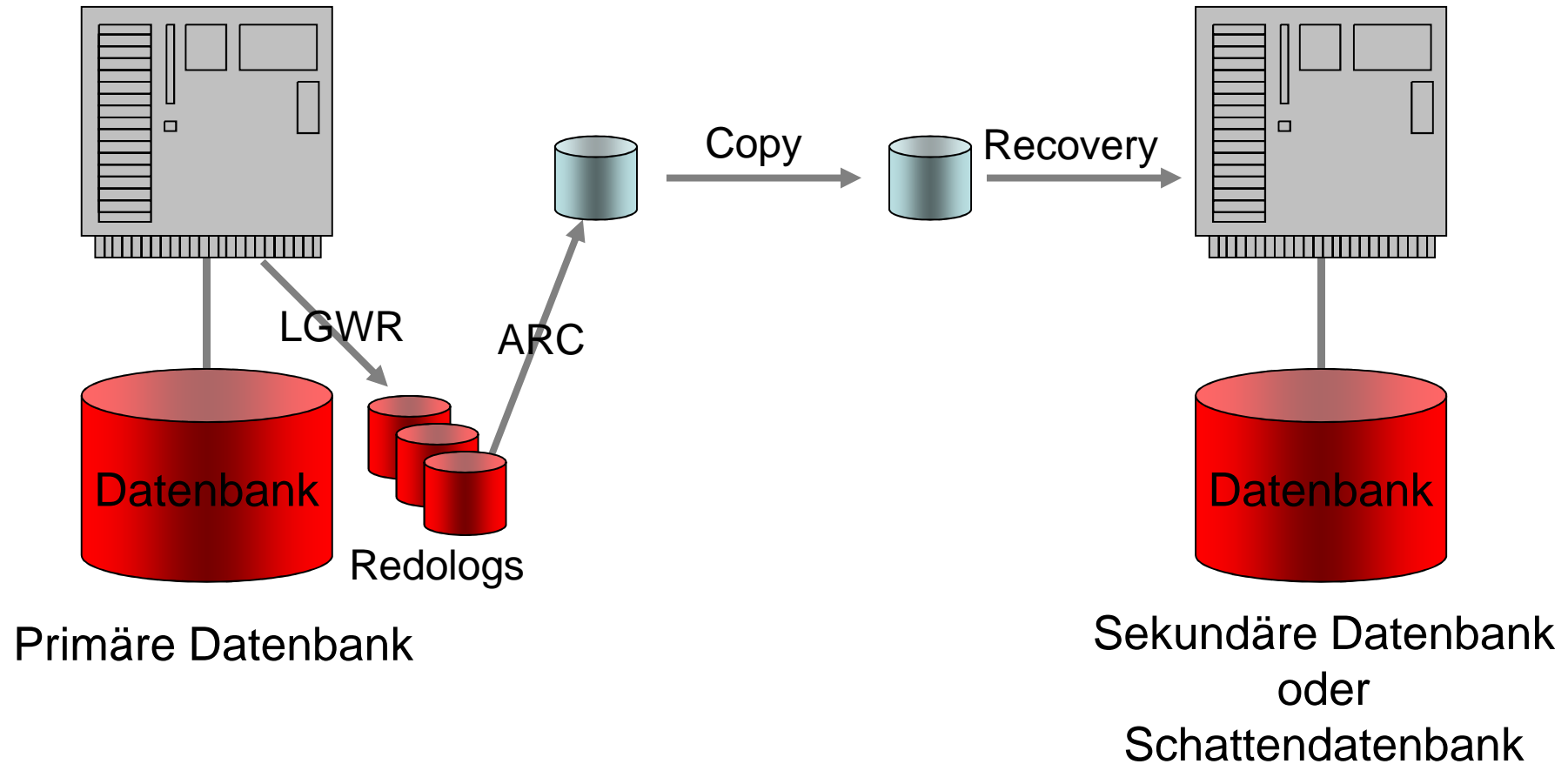
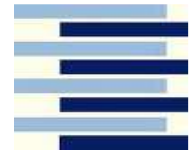


### Einschränkung der Verfügbarkeit

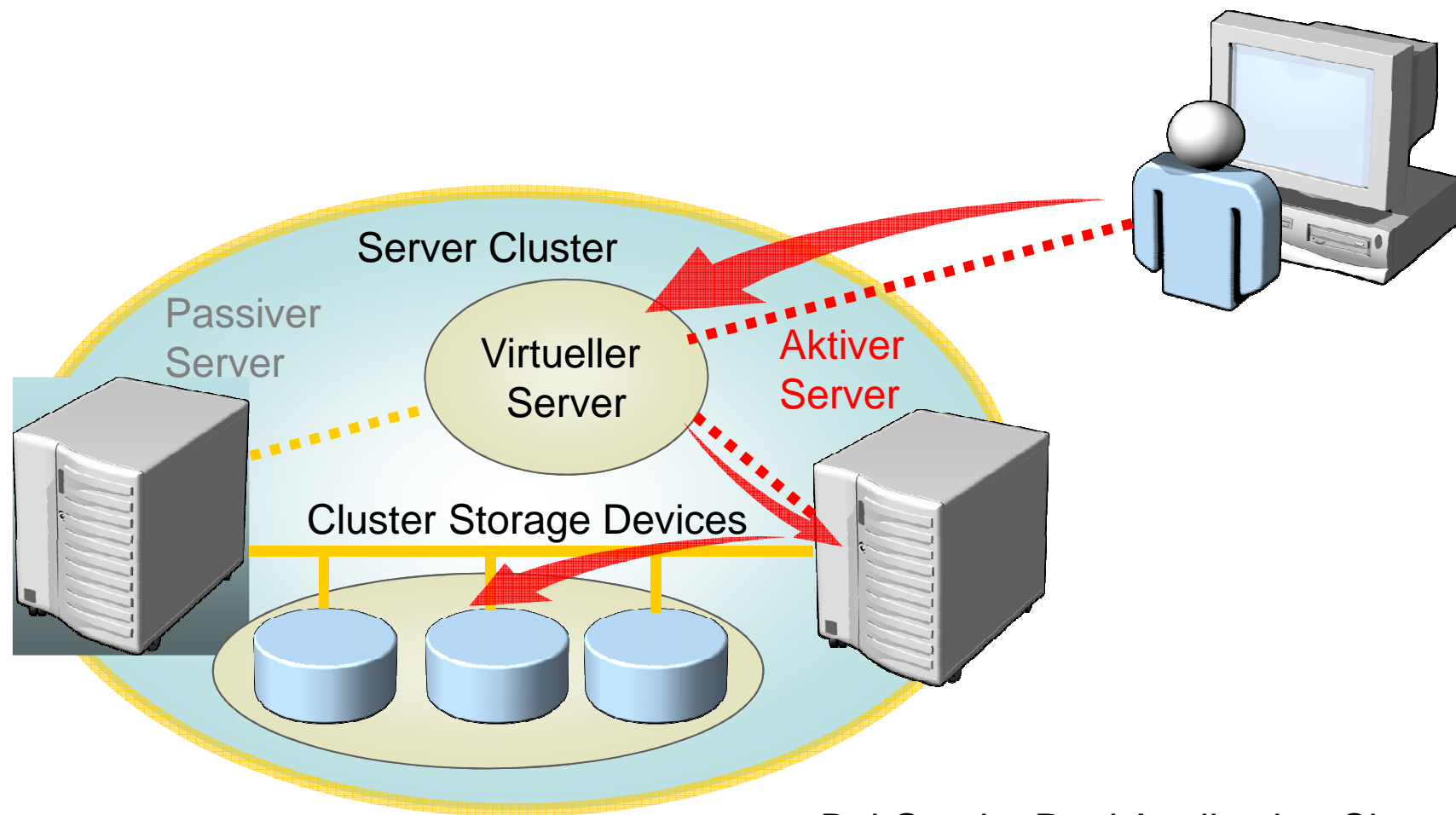
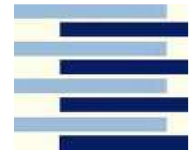
- Hardwarefehler
- Softwarefehler
- Anwenderfehler
- Maintenance



# Von der Stand-by Datenbank ...

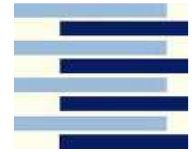


## ... zum Failover-Clustering



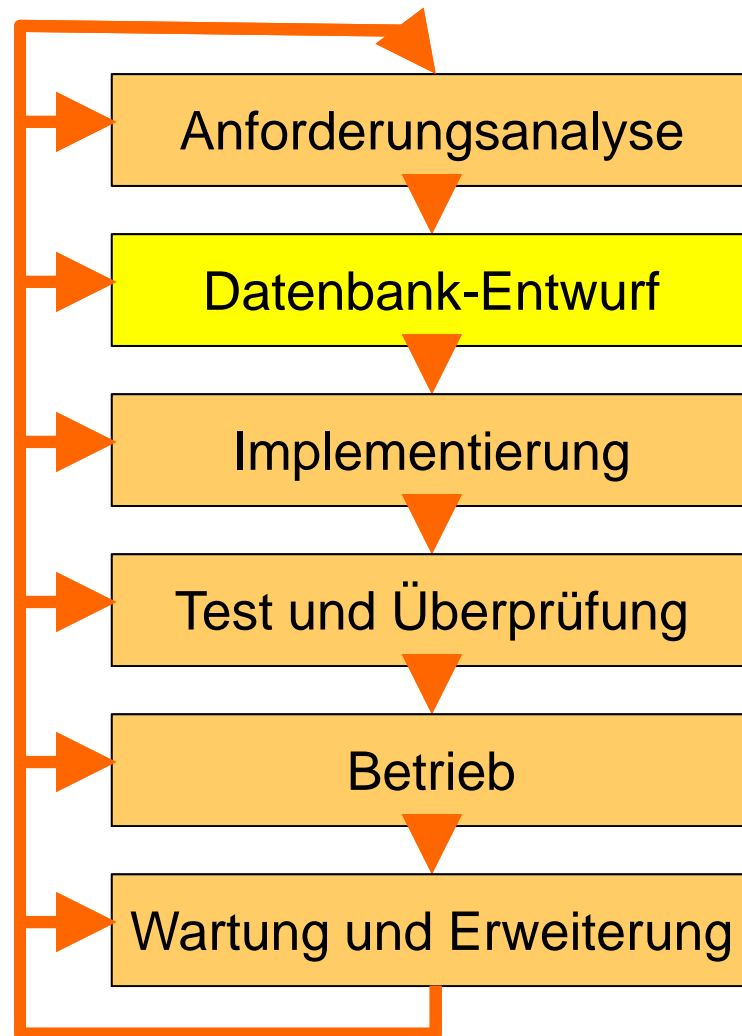
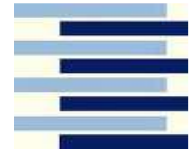
Bei Oracle: Real Application Cluster

# NoSQL Datenbanken



- Nicht alle Daten relational
- Glaubenskrieg bei DB-Profis: Wieviel Struktur muss eine Datenbank haben (Schema)?
- Integration von DB-Systemen mit File-Systemen
- Traditionelle File-Systeme nicht mehr ausreichend (in großen Unternehmen Millionen von Dateien) → Weiterentwicklung zu Datenbanksystemen
- Alternative Datenbankmodelle:
  - Dokumentenorientierte DB (z.B. CouchDB)
  - Graphen-DB
  - Key Value Stores
  - Column Stores

# Von der Anforderungsanalyse zur Maintenance



Definiere das Problem  
Definiere die Anwender

Definiere ein Modell zur Lösung des Problems

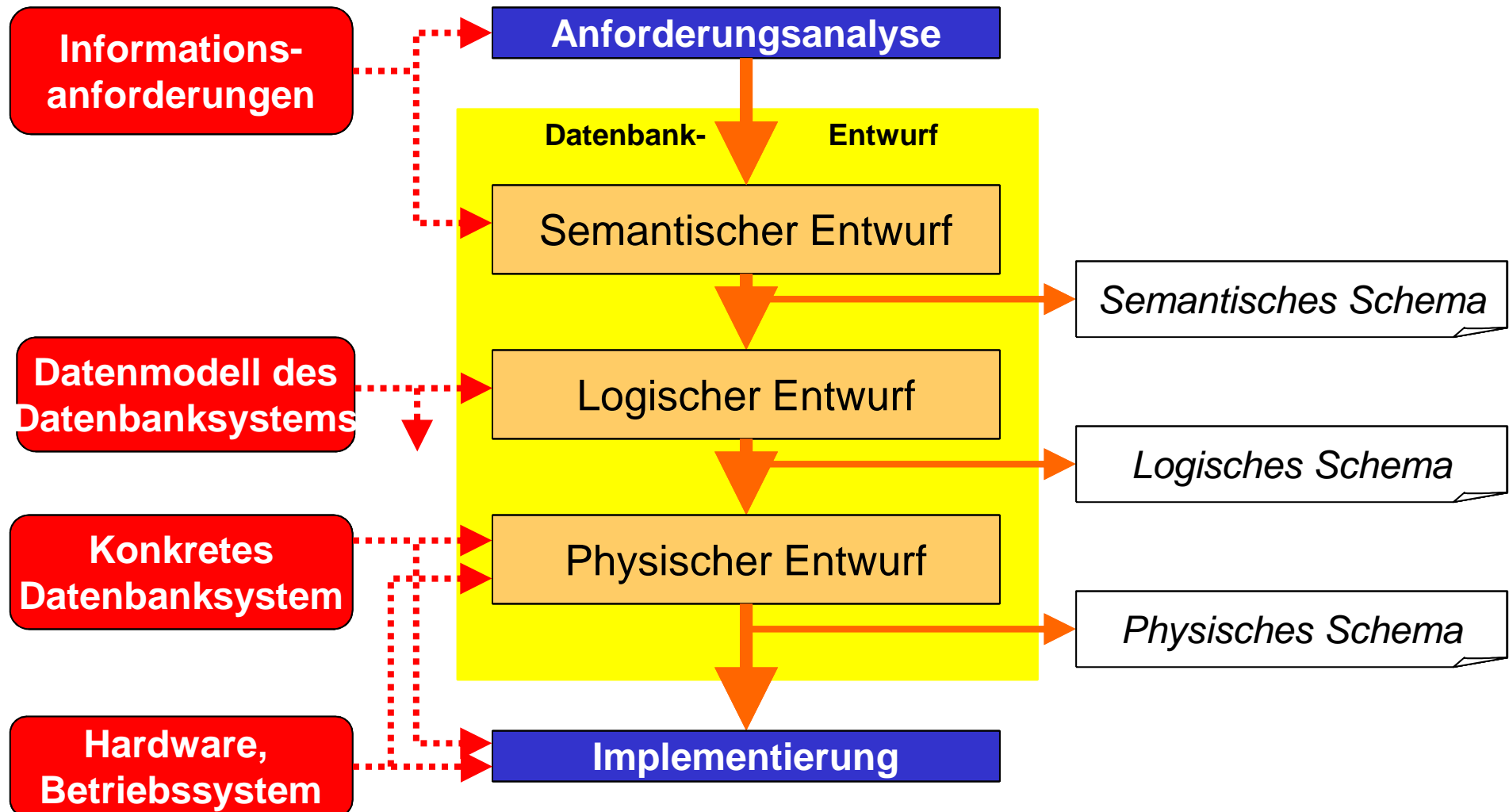
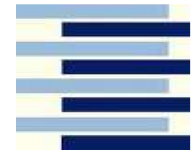
Installiere das DBMS  
Erzeuge die Datenbank  
Fülle/Lade die Daten

Teste die Datenbank  
Verbessere die Datenbank (Feintuning)  
Überprüfe die Datenbank und die Anwendungen

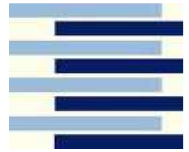
Produziere den benötigten Informationsfluss  
Halte alles am Laufen

Führe Änderungen ein  
Führe Erweiterungen durch

# Der Datenbankentwurf



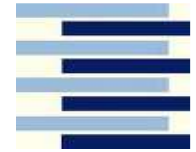
# Datenbanken und Anwendungen



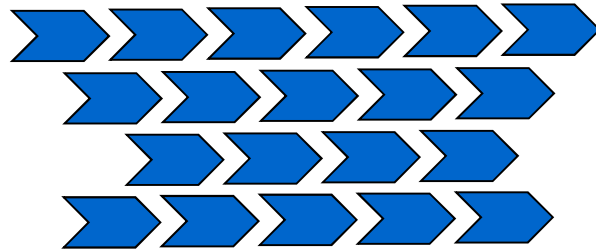
- **Datengetriebener Ansatz** (in dieser Vorlesung):
  - Entwickle erst Datenbank, dann Anwendungen
  - Daten bzw. Informationen bilden Kapital vieler Firmen.
  - Daten überleben i.d.R. die Anwendungen
  - Typisches Szenario: Zentrale Firmen-Datenbanken
- Andere Szenarien erfordern u.U. einen anderen Fokus:
  - Ausgangspunkt: Anwendung
  - Datenbank der Anwendung unterordnen NoSQL?
- In der Praxis tritt häufig eine Mischform auf:
  - DB-Designer muss über den „Tellerrand“ hinaussehen
  - existierende Datenbanken müssen erweitert oder migriert werden.
- Referenzmodelle



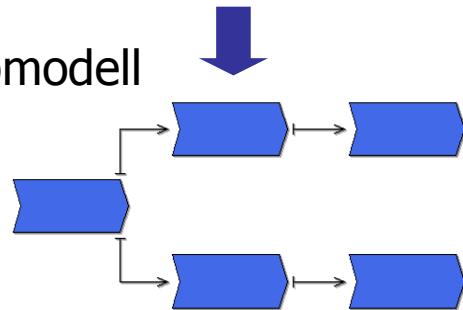
# Verwendung von Referenzmodellen



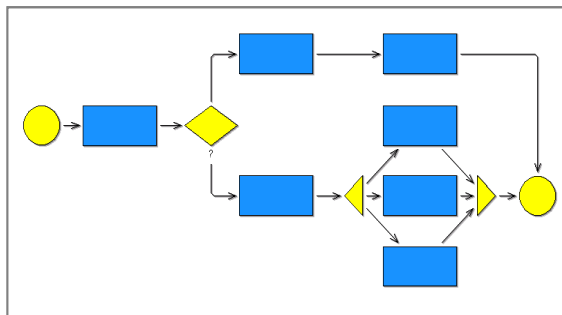
Menge "bewährter" Prozesse



Grobmodell



Detailmodell



## Vorgehen

- Wähle das am besten passende Referenzmodell
- Anpassung an das Unternehmen
- Detailliere es für das Unternehmen

## Vorteile

- Geringerer Modellierungsaufwand
- Vorgegebene „bewährte“ Prozesse und Strukturen
- Geringere Gefahr, wichtige Bereiche unberücksichtigt zu lassen
- Bewertung der eigenen Leistung im Vergleich zu anderen Firmen

