



# **Business Intelligence**

## **Data Warehouse**

Jan Weinschenker

# Inhaltsverzeichnis

- Einleitung
- Motivation
- Architektur eines Data Warehouse
- Data Warehouse im Ferienclub
- Zusammenfassung
  
- Fragen

# Einleitung

## Definition: Data Warehouse

- „A data warehouse is a subject oriented, integrated, non-volatile and time variant collection of data in support of management's decisions.“

*aus [Inmon 1996]*

# Einleitung

Definition: Data Warehouse

- Ein Data Warehouse ist eine physikalische Datenbank, die eine integrierte Sicht auf beliebige Daten zu Analysezwecken ermöglicht.

*aus [Bauer/Günzel 2004]*

# Motivation

Wozu braucht  
man eigentlich ein  
Data Warehouse?

# Wozu eigentlich DW?

- Analyse von Kennzahlen:
  - Ging lange Zeit ohne DW
  - Ging lange Zeit sogar ohne Computer
- Unternehmen waren trotzdem erfolgreich!

# Wozu eigentlich DW?



# Wozu eigentlich DW?

„Wenn einen die große Masse (an Daten) erschlägt!“

- Daten sind oft geschäftsobjekt-orientiert
- Zusammentragen der Informationen wird mühselig
- Excel, SPSS und Anwender sind irgendwann überfordert

# Deswegen!

- DW machen große Datenmengen überschaubar
  - Sie schaffen eine einheitliche Sicht auf unterschiedliche Datenquellen
- DW machen Analysen performanter
- Weitere neue Impulse durch die KI
  - Finden neuer Zusammenhänge  
Datamining

# Meine Motivation

- DW kann wichtige Beiträge liefern für
  - Planung, Steuerung, Kontrolle
- Hilft überall, wo Projekte kontrolliert zum Erfolg gebracht werden sollen
- Gewinnt in Unternehmen an Bedeutung

# Architektur eines Data Warehouse

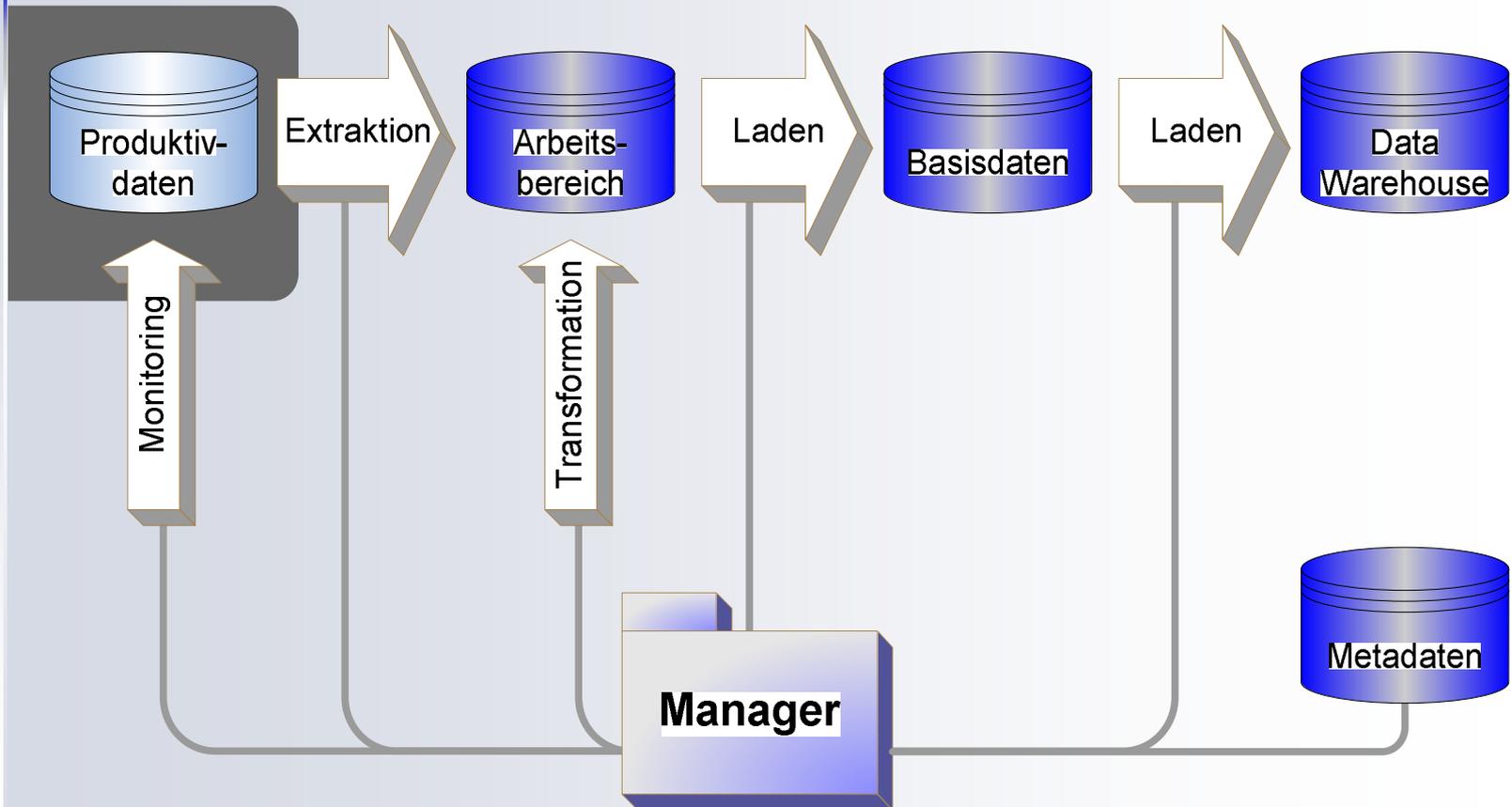
Wie baut man ein  
Data Warehouse?

# Ein Data Warehouse

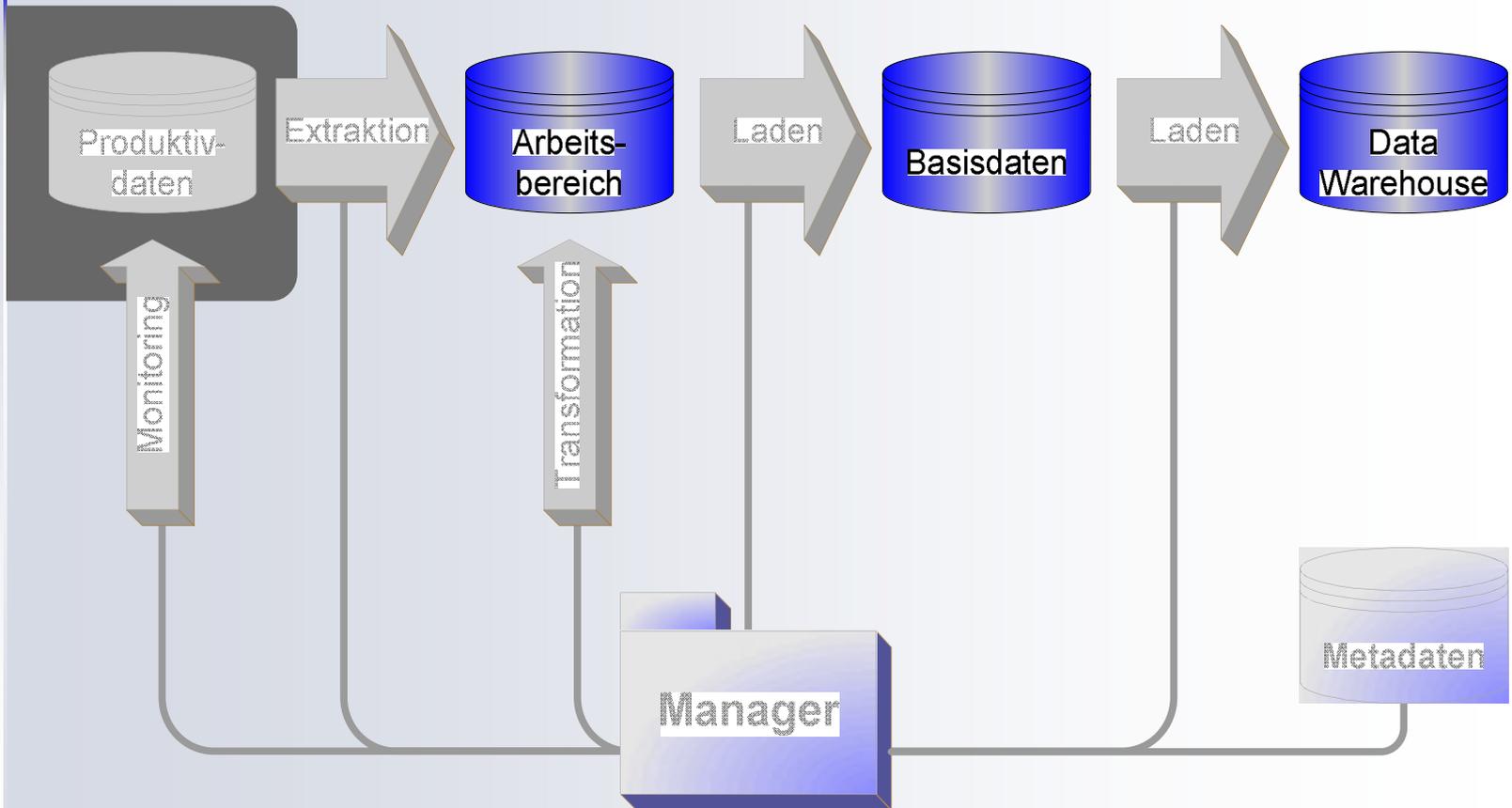
- Orientierung an Referenzarchitektur
- Komponenten:
  - Arbeitsbereich, Basisdatenbank, Data Warehouse
  - Extraktion, Transformation, Laden, Metadaten
  - Manager

Aus [Bauer/Günzel 2004]

# Referenzarchitektur



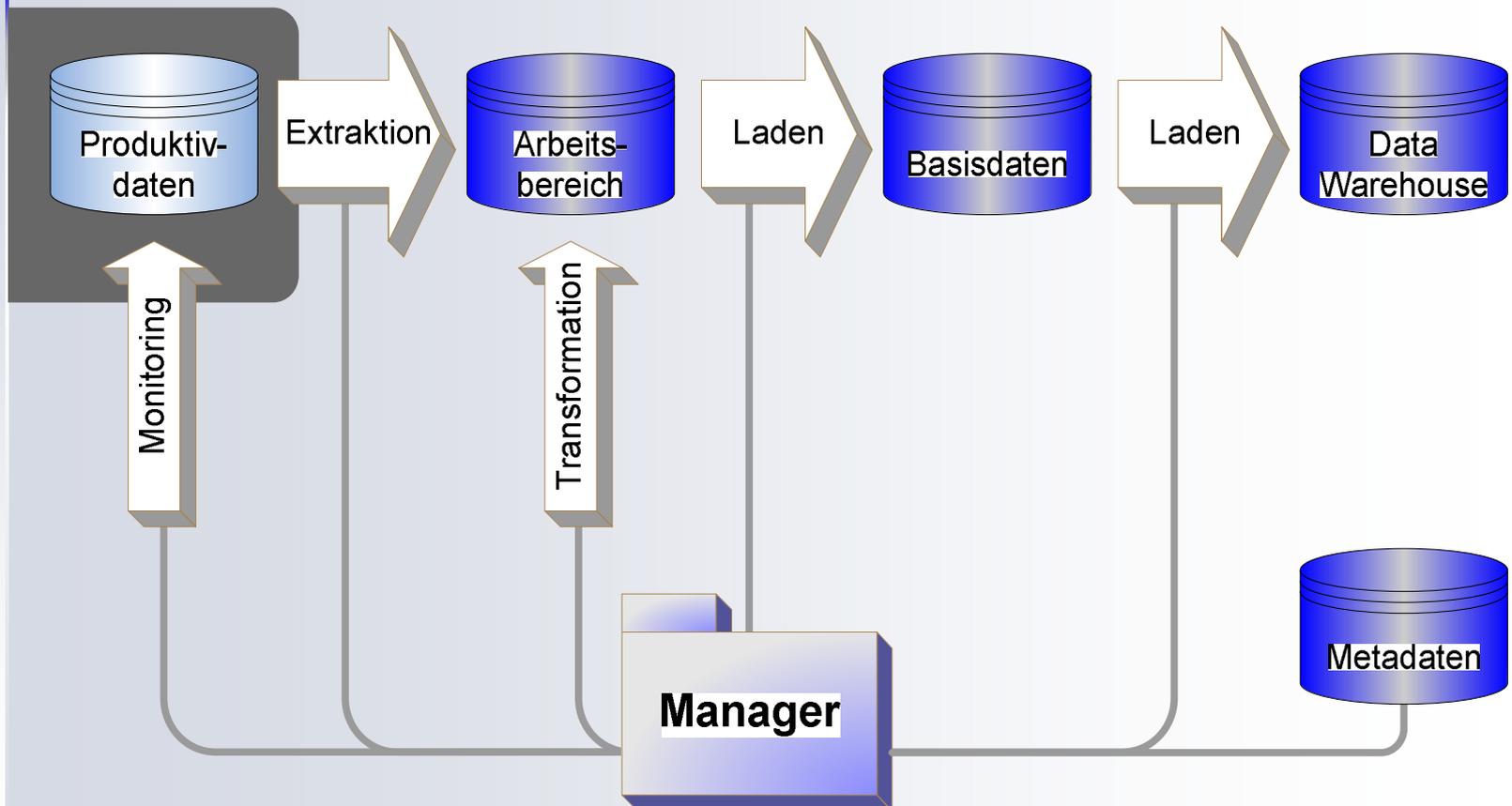
# Datenhaltung



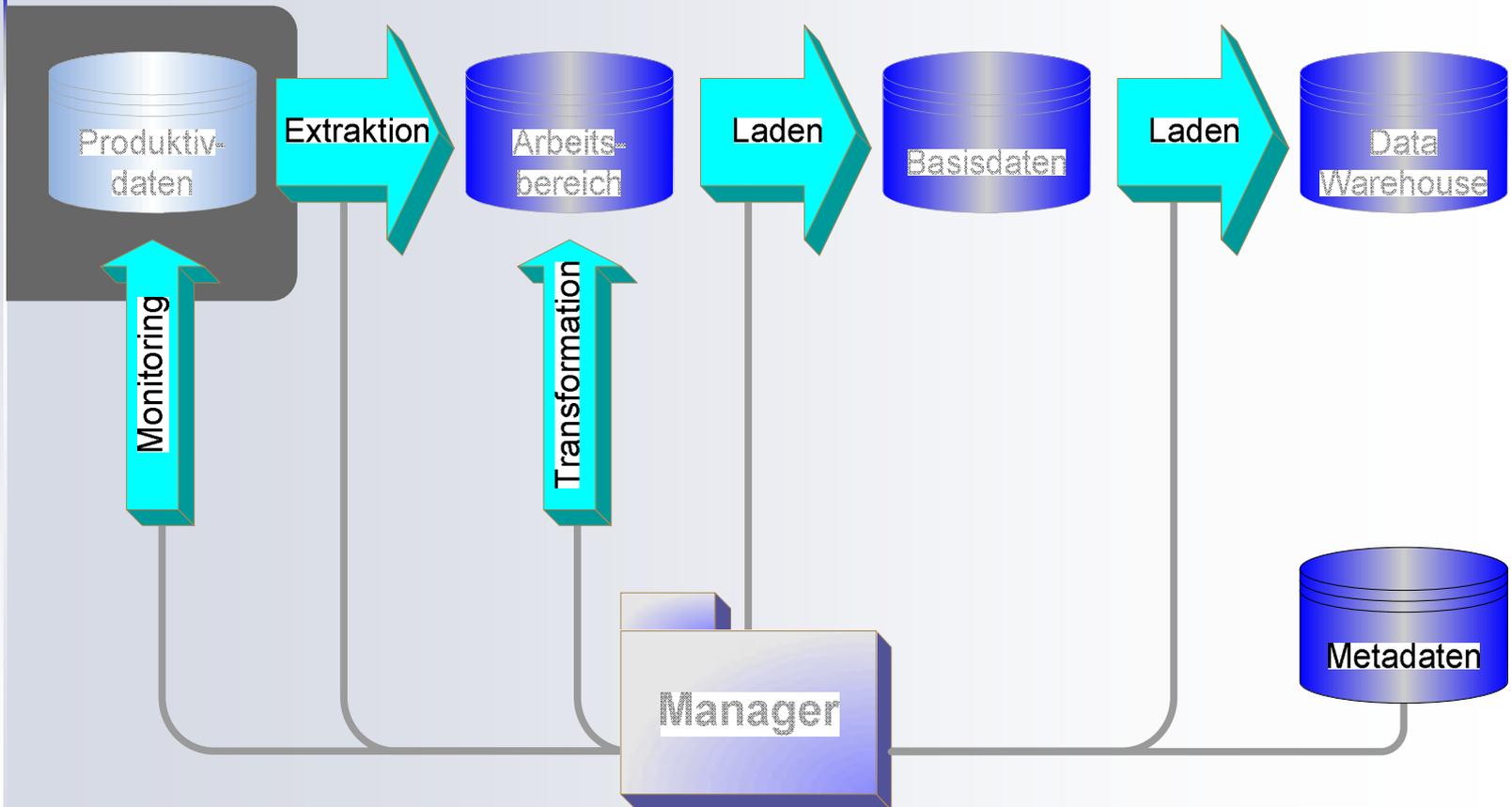
# Datenhaltung

- In der Regel mit einem RDBMS realisiert
- Arbeitsgrundlagen für ETL-Prozesse,
- Datenhaltung getrennt vom Produktivsystem

# Prozesse und Metadaten



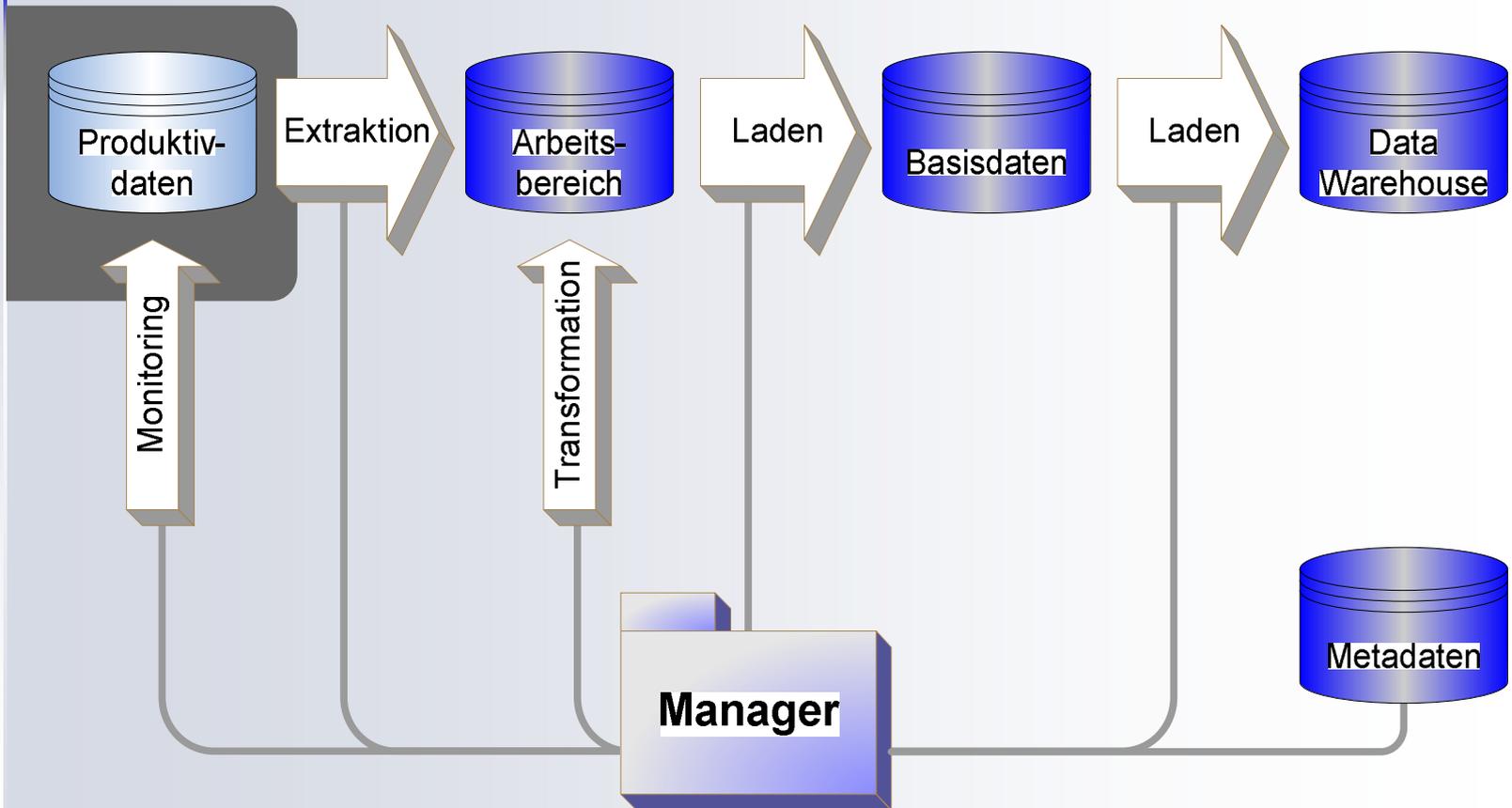
# Prozesse und Metadaten



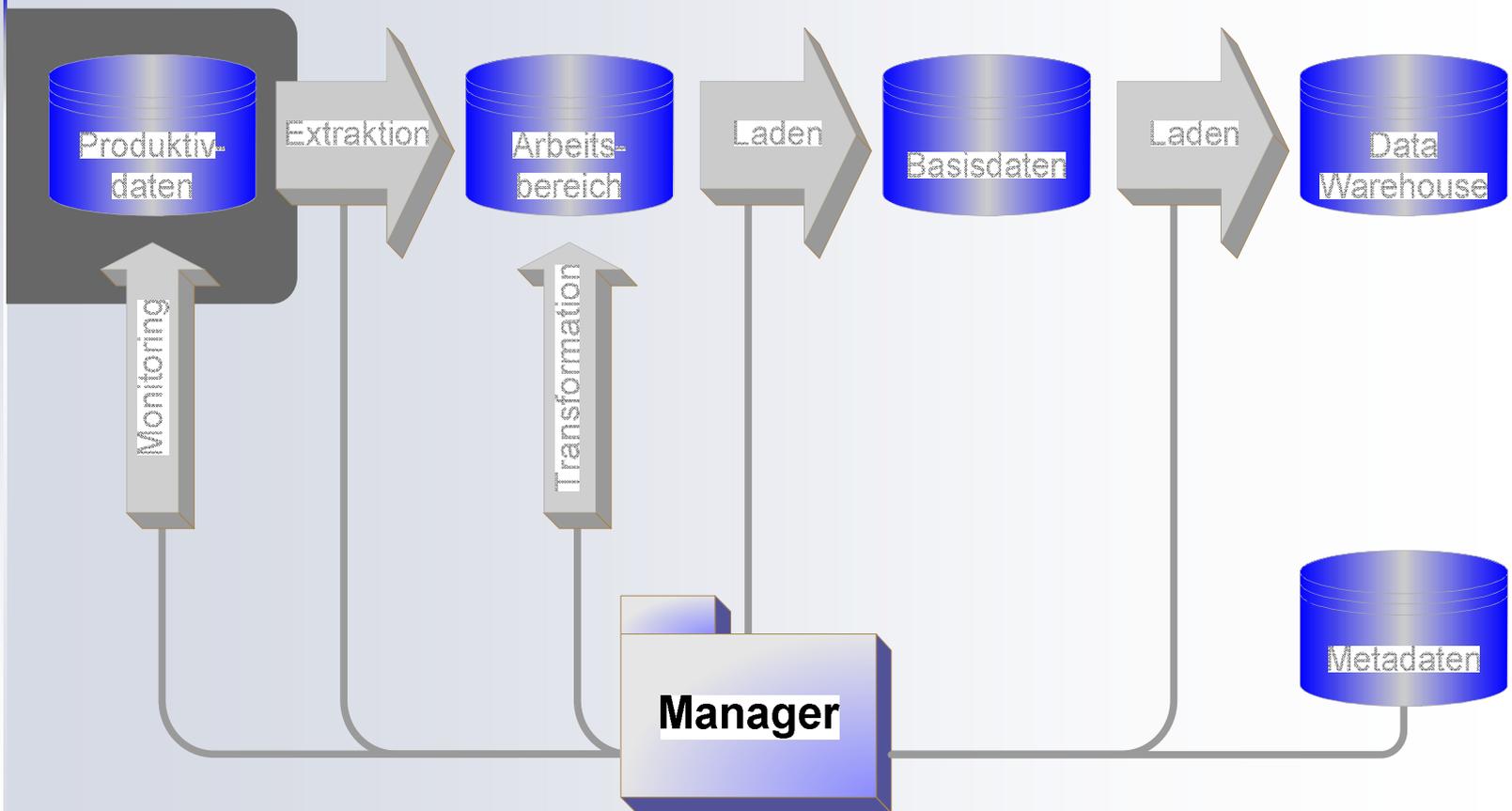
# Prozesse und Metadaten

- Überführung der Daten aus dem „Produktivsystem“ in das Data Warehouse
- Umwandlung der Daten
  - Von Anwendungs- nach Analyseorientiert
  - In einheitliche Datentypen
- Laden von großen Datenmengen
  - Performancelastig

# Manager



# Manager



# Manager

- Steuerungs- und Kontrollkomponente
  - Datenbeschaffung
  - Interner Datenfluss
- Fehlerbehandlung
- Logging der eigenen Aktivitäten

# Architektur

- Ausrichtung an gewünschten Analysen
- Vereinbarungen mit Teilgruppen
  - Anwendungsfälle
  - Anwendungsdaten
- Schnittstellen
  - Persistenz
  - Analyse

# DW im Ferienclub

Was kann ich beitragen?

# Wozu brauchen wir DW?

## Unser Ferienclub

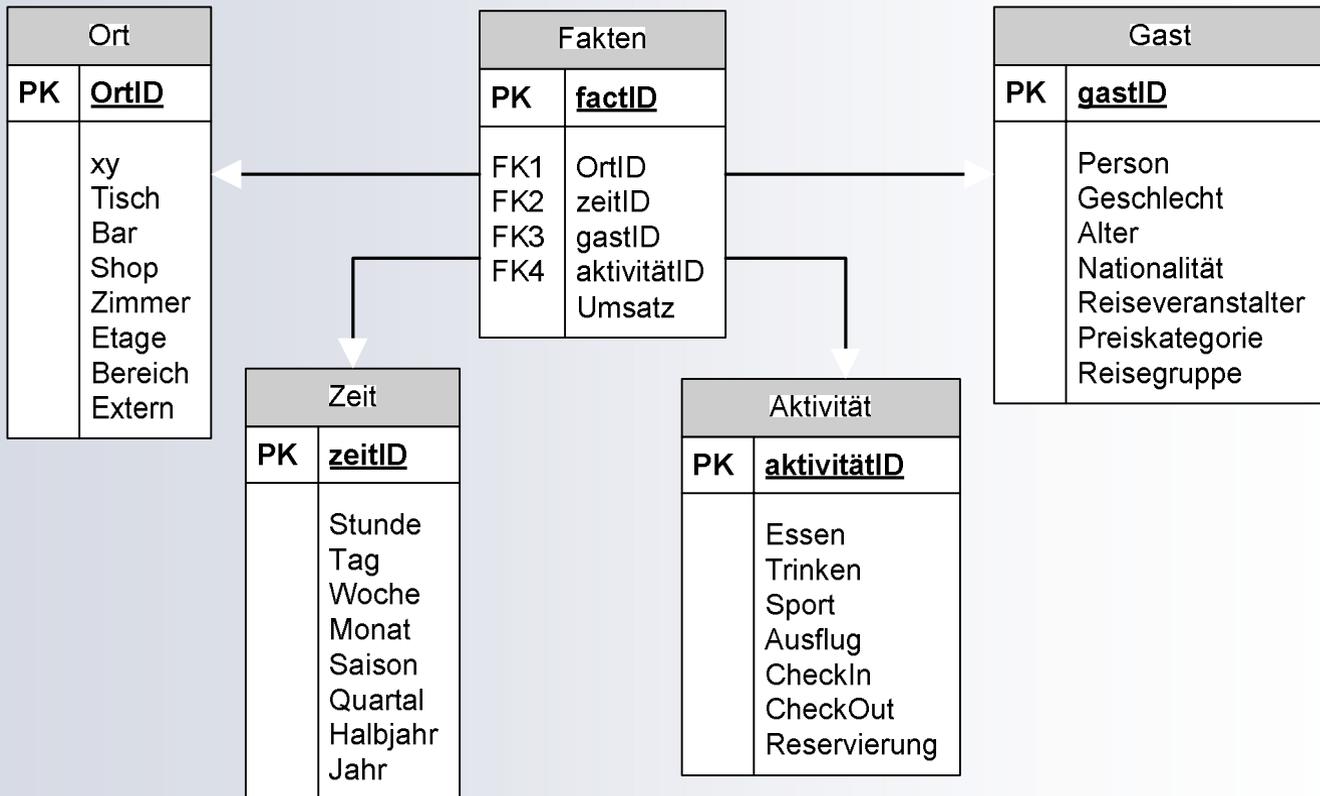
- Heterogene Anwendungen und Daten
- Einzelne Datentöpfe im Persistenzdienst
- Viele interessante Datenquellen
  - RFID, Semantic Web, ...
- also ...?

# Wozu brauchen wir DW?

... für eine integrierte Sicht  
auf die vorhandenen Daten

- Daten und Informationen vorhanden
  - Müssen nur in Wissen umgewandelt werden
- Analyse mit Data Mining und OLAP
  - Dazu mehr im nächsten Vortrag

# Vorschlag Starschema



# Werkzeuge

ETL: Octopus

- Basiert auf Java und XML
- Zugriff auf ODBC und JDBC-Datasources, CSV und Excel
- <http://octopus.objectweb.org>

# Werkzeuge

## Data Warehouse: DeepGreen

- Open Source Data Warehouse
  - Basiert auf BisGres DMBS (PostgreSQL-Fork)
  - Benötigt Fedora/Redhat-Linux
- <http://www.greenplum.com>
- <http://www.bisgres.org>

# Fazit

- Einheitliche Sicht auf unsere Daten
- Grundlage für Analysen
- Benötigt werden:
  - Anwendungsdaten
  - Vereinbarungen über Geschäftsprozesse

# Fragen?





**Vielen Dank!**