

# Real-Time Data Warehousing

René Wicht

10.12.2009



FAKULTÄT FÜR  
INFORMATIK



# Gliederung

- Motivation
- Begriffe
- Ansätze
- Zusammenfassung
- Quellen

# Motivation

- Data Warehouses in Unternehmen wichtig
  - Informationen schneller benötigt
  - Aktualität von Daten
  - Vermeidung von Verzögerungen
- 
- Zeit ist Geld

# Begriffe

- Real-Time
  - Echtzeit
  - Synchron zur realen Welt
  - Daten sind vorhanden,  
wenn sie gebraucht werden
  - Nicht zwingend sofort

# Begriffe

- Data Warehousing
  - Füllen des Data Warehouses (ETL)
    - Extracting
    - Transforming
    - Cleansing
    - Loading
  - Analyse
    - OLAP Systeme
    - Data Mining
    - ...

# Begriffe

- Real-Time Data Warehousing
  - Data Warehouse enthält aktuelle Daten
  - OLAP auf aktuellen Daten
  - Analyse-, Entscheidungsprozesse arbeiten auf aktuellen Daten
  
- Data Warehouse ist Zeitnah

# Ansätze

- Bisher
  - „Downtime“ bei der Datenintegration
    - Periodische Aktualisierung (nächtlich)
    - Daten nicht aktuell
- Real-Time Ansatz
  - Kontinuierliche Datenintegration
    - Keine „Downtime“
    - Aktuellere Daten

# Ansätze

- Kontinuierliche Datenintegration
  - „Near Real-Time“ ETL
  - „Trickle & Flip“
    - Zwischengestellte Tabelle
  - Externer Real-Time Daten Cache
    - Weitere Datenbank für Real-Time Daten
- Immer Auswirkungen auf Analyse

# Ansätze

- Auswirkungen auf Datenanalyse (OLAP)
  - Daten ändern sich während Bearbeitung
- Ansätze
  - „Near Real-Time“ Ansatz
  - Risikominimierung
  - Externer Daten Cache
    - Trennt Real-Time Daten von historischen Daten

# Ansätze

- **Externer Daten Cache**
  - Kein Konflikt bei der Analyse
  - Anfragen auf Real-Time Daten und historische Daten problematisch
- **Just-in-time information merging**
  - Verknüpfung von Daten Cache und DWH
  - Anfrageanalyse
  - „Snapshot“ aus Daten Cache

# Ansätze

- Umwandlung eingehender Daten
  - Häufige Änderung der Daten
  - Anpassungen nötig
- Wrapper
  - Wandeln Daten um
  - Müssen flexibel sein
  - Temporal constraints

# Zusammenfassung

- Real-Time Data Warehouse
  - Leifert aktuelle Daten (just-in-time)
- Datenbeschaffung kritisch
  - Kontinuierliche Datenintegration
    - Keine „downtime“
    - Auswirkungen auf Analyse
- Verhältnis Nutzen/Aufwand muss stimmen

# Quellen

- Mukesh Mohania; Ullas Nambiar; Michael Schrefl; Millist Vincent: „Active and Real-Time Data Warehousing“, 2008
- Francisco Araque: „Real-Time Data Warehousing with temporal requirements“
- Josef Schiefer, Jun-Jang Jeng, Robert M. Bruckner: „Real-Time Workflow Audit Data Integration into Data Warehouse Systems“
- Justin Langseth: „Real-Time Data Warehousing: Challenges and Solutions“, DSSResources.COM, 02.08.2004