

Konsistenzquantifizierung in Grid-Datenbanksystemen

10. GI-Fachtagung Datenbanksysteme für Business, Technologie und Web

BTW 2003

Leipzig, 27. Februar 2003

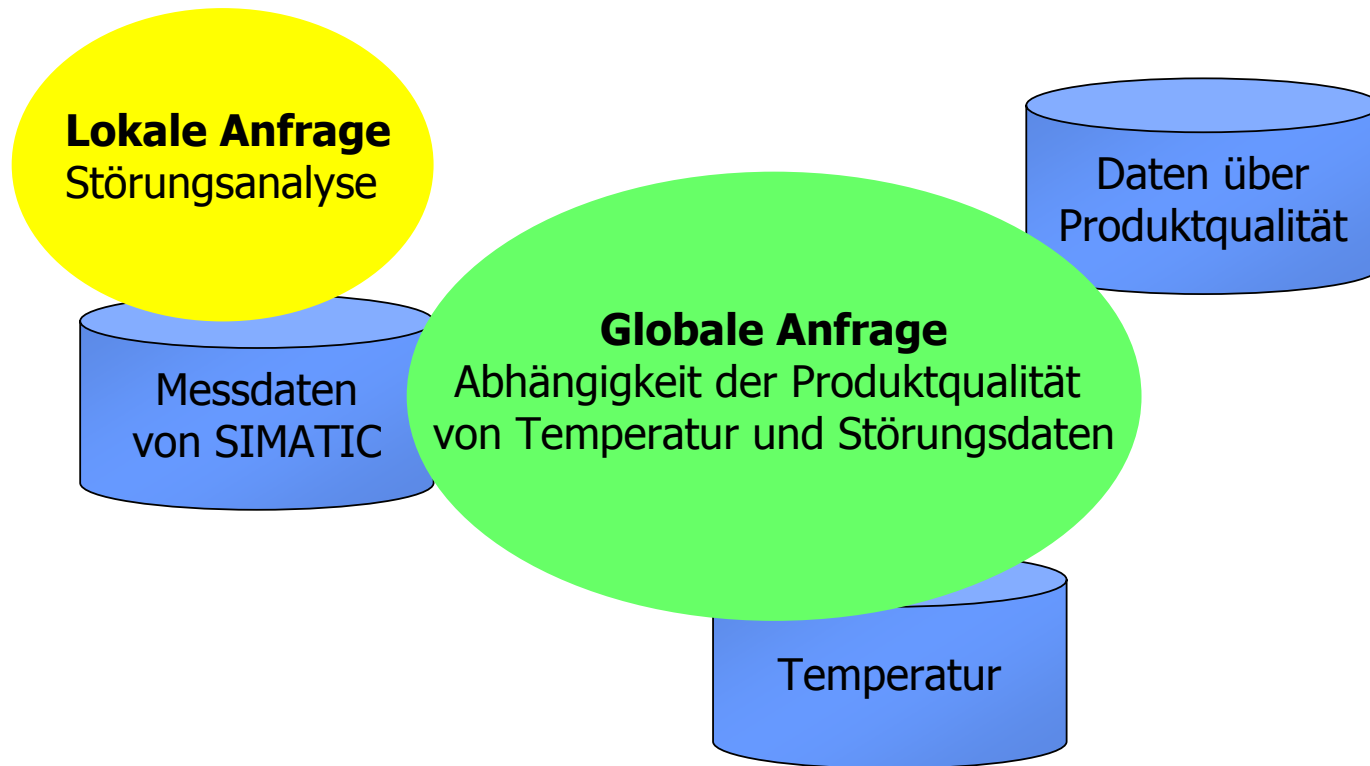
Lutz Schlesinger, Universität Erlangen, Lehrstuhl für Datenbanksysteme

Wolfgang Lehner, Technische Universität Dresden, Arbeitsgruppe Datenbanken

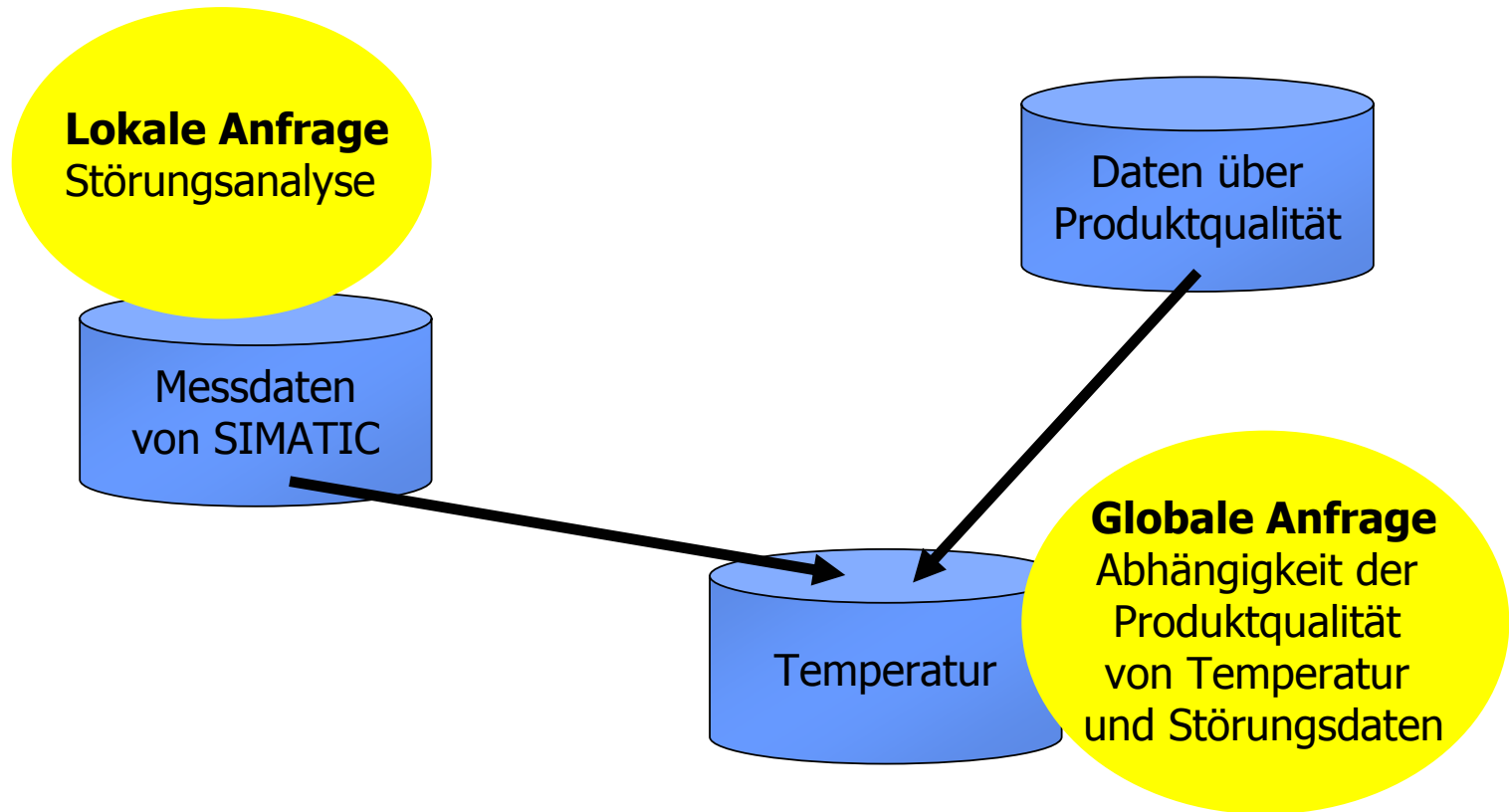
Beispiel: SIMATIC-Steuerung



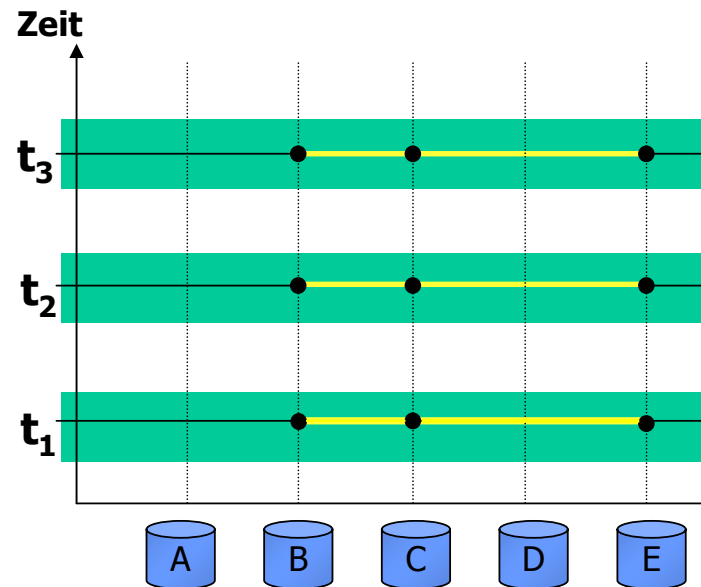
Beispiel: Auswertung von Sensordaten



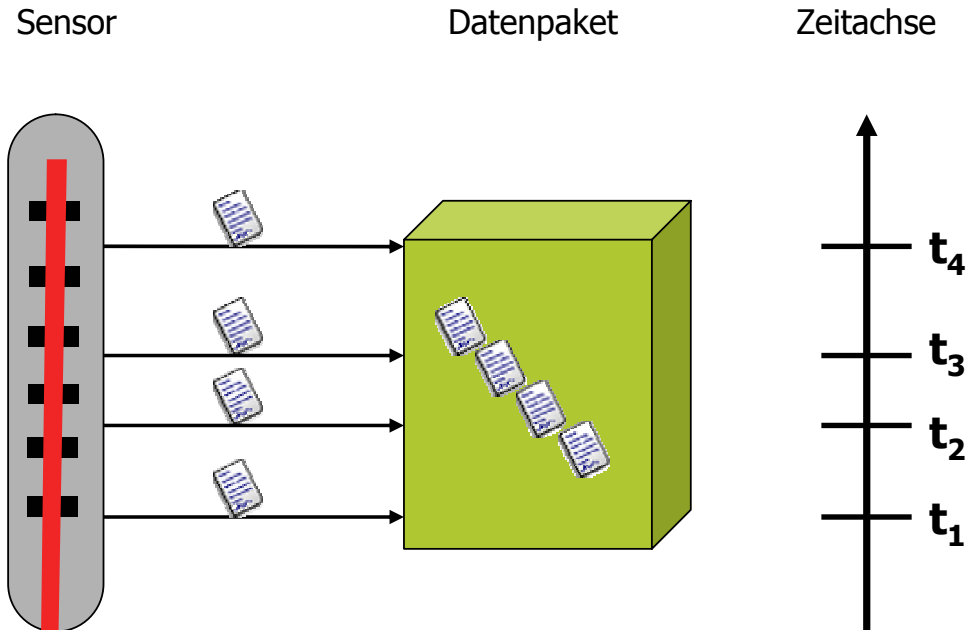
Beispiel: Auswertung von Sensordaten



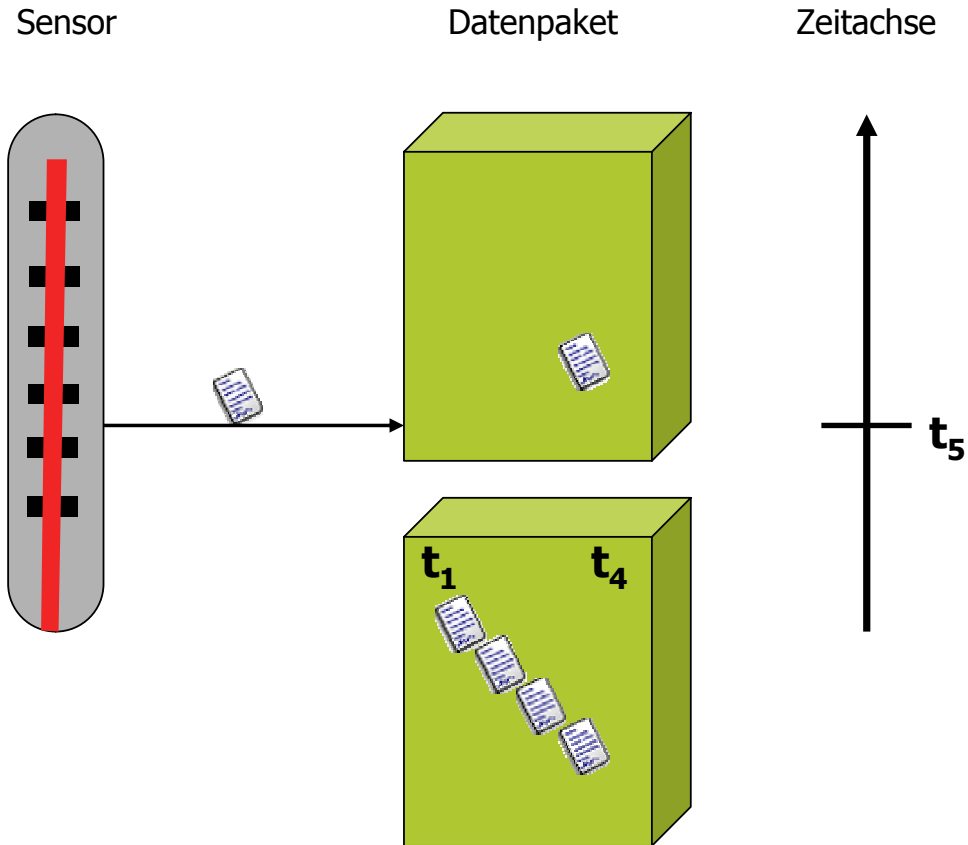
Datenquellen-Zeit-Diagramm



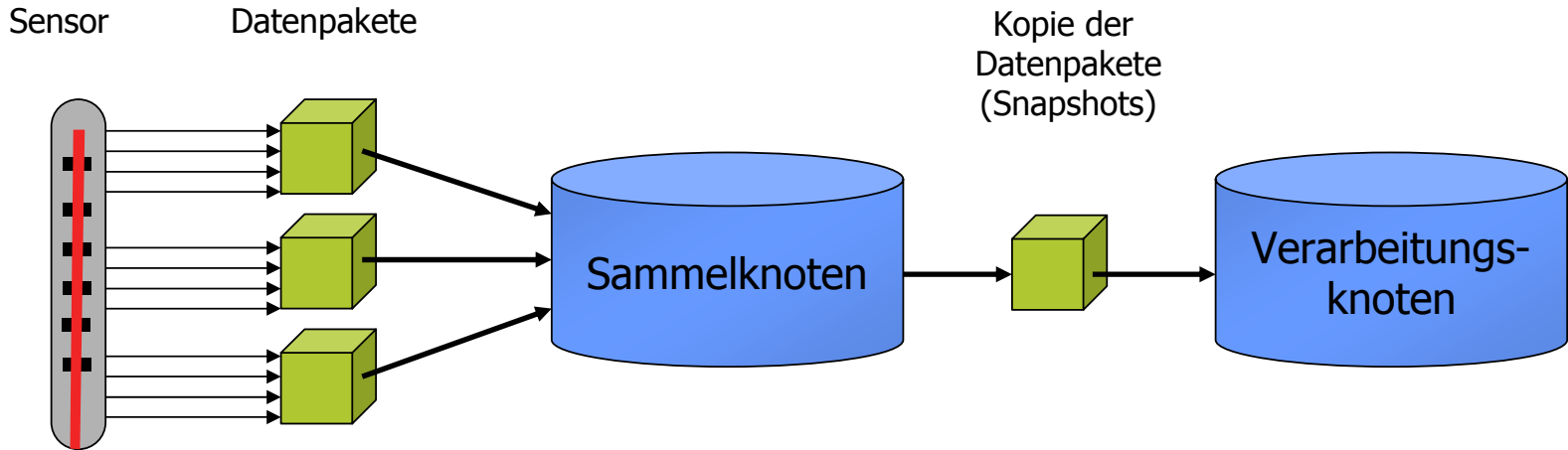
Datenpaketbildung am Sensor



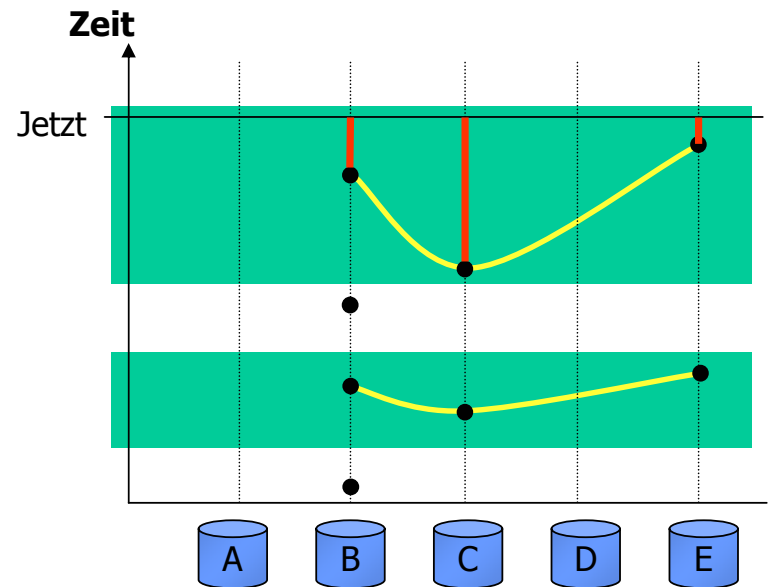
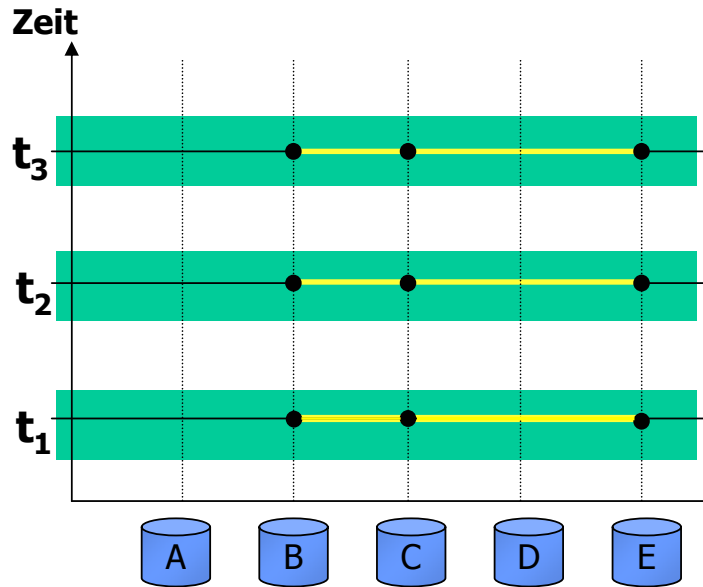
Datenpaketbildung am Sensor



Weiterleitung der Datenpakete



Datenquellen-Zeit-Diagramm



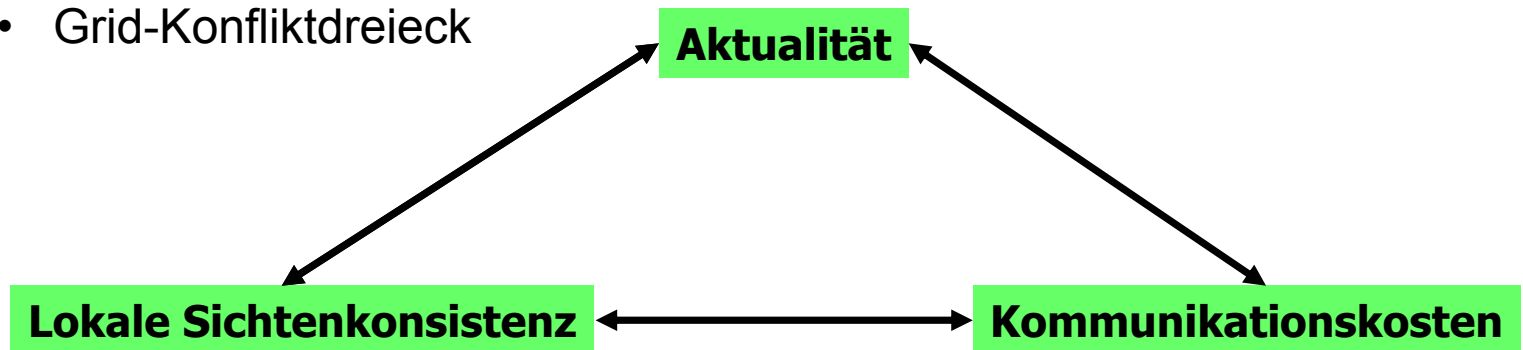
Annahmen und Grid-Konfliktdreieck

- Annahmen

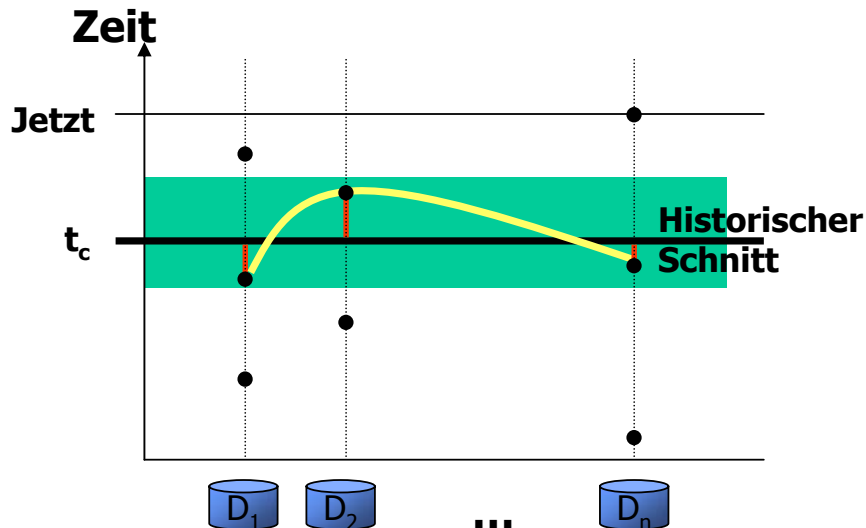
- Autonome, eng gekoppelte (Datenbank-) Systeme
- Einheitliche Begriffswelt
- Existenz eines globalen Schemas
- Keine einheitlichen Gültigkeitszeiten
- Operationaler Fokus
- Übergreifende Auswertungen

Grid-DBS

- Grid-Konfliktdreieck



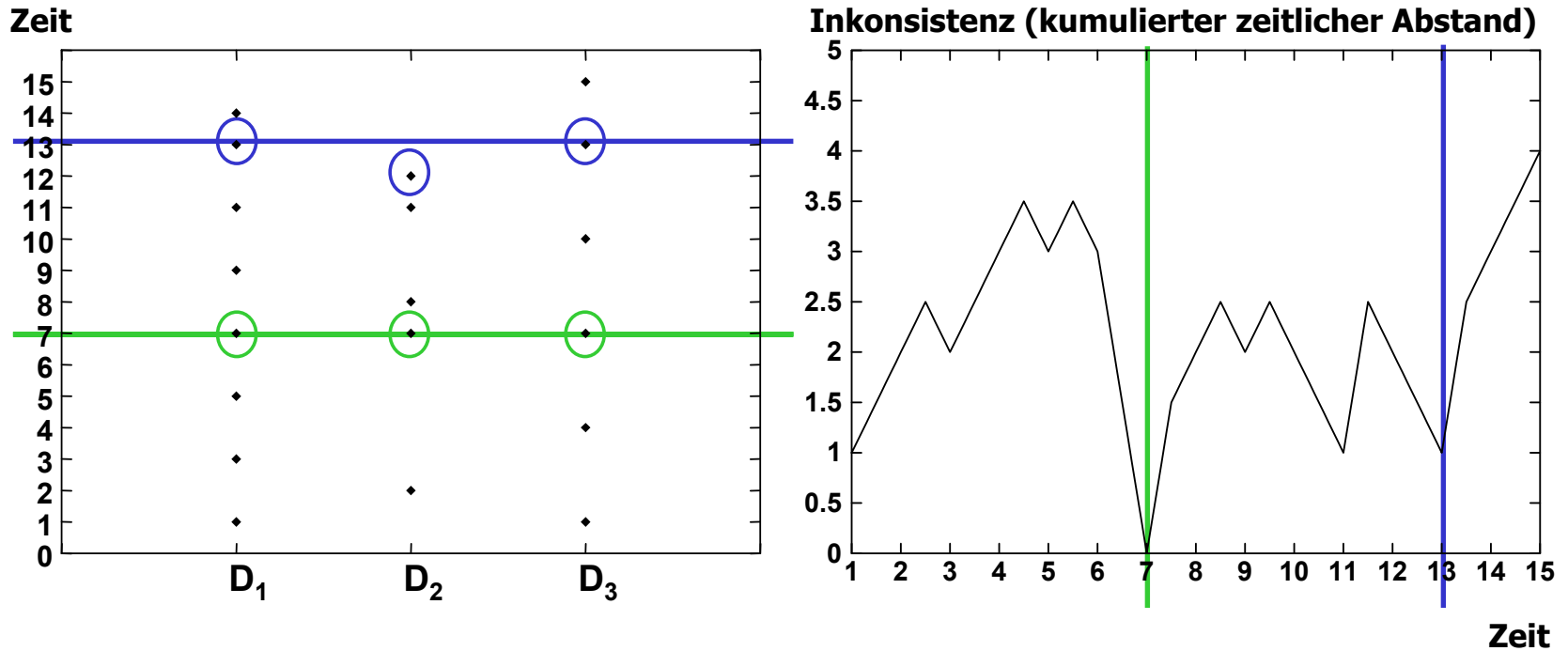
Bestimmung eines konsistenten Schnittes



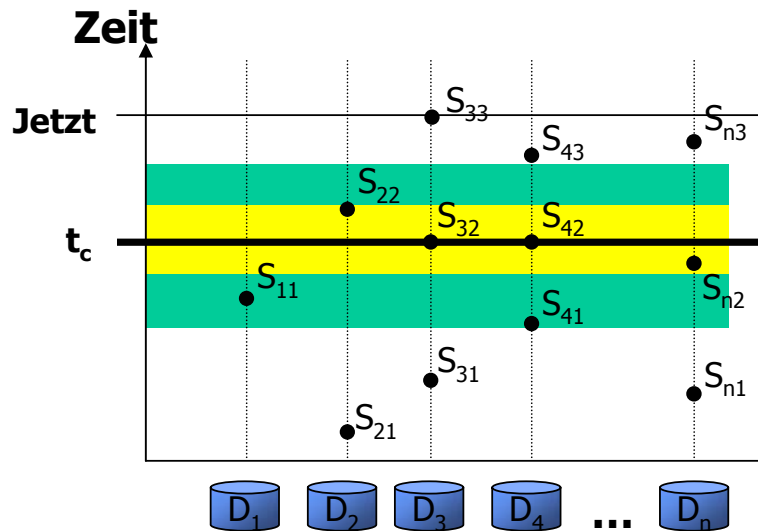
- Idee
 - Spezifikation des Abstandes zwischen t_c und einem Snapshot von D_j in Form von *Kohärenzprädikaten*
 - Bestimmung des *besten* historischen Schnitts: horizontale Linie zum Zeitpunkt t_c
- Bedingungen
 - minimale Verletzung der Kohärenzprädikate
 - möglichst aktuell
- Probleme (nicht orthogonal)
 - Quantifizierung der Inkonsistenz
 - Verbund von Snapshots
 - Auswahl der Snapshots
 - Bestimmung des historischen Schnitts

Quantifizierung der Inkonsistenz

Inkonsistenz definiert über zeitlichen Abstand der ausgewählten Snapshots und t_c



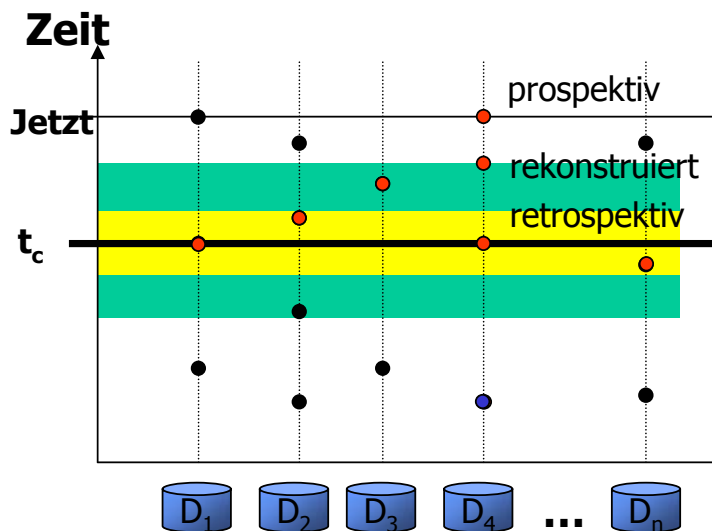
Verbund von Snapshots mittels innerer und äußerer Konsistenzbänder



- Inneres Konsistenzband
 - Betrachtung aller Snapshots innerhalb des inneren Konsistenzband als konsistent
 - Verwendung der inneren Verbundsemantik

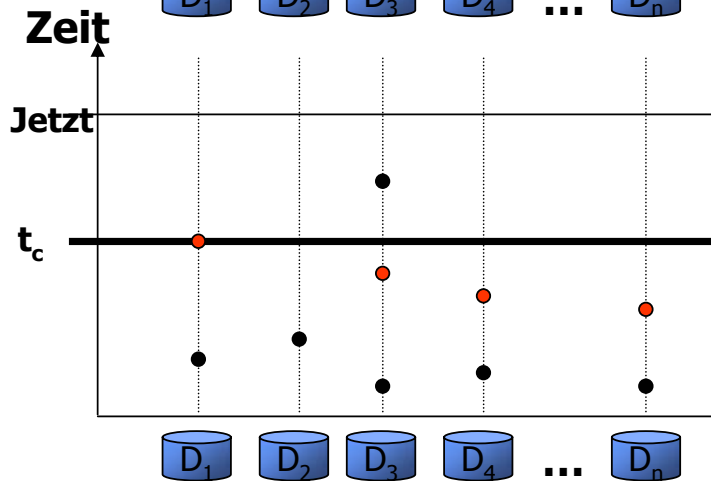
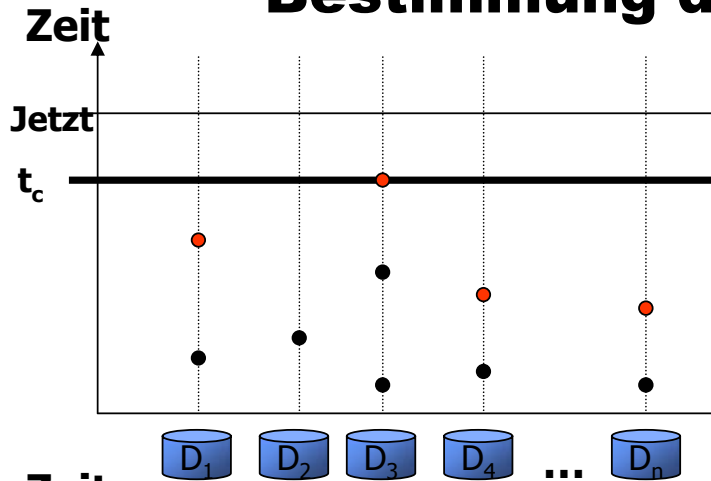
- Äußeres Konsistenzband
 - Maximaler Abstand eines Snapshots zum historischen Schnitt
 - Benutzung äußerer Verbundsemantik

Algorithmische Betrachtung: Auswahl von Snapshots



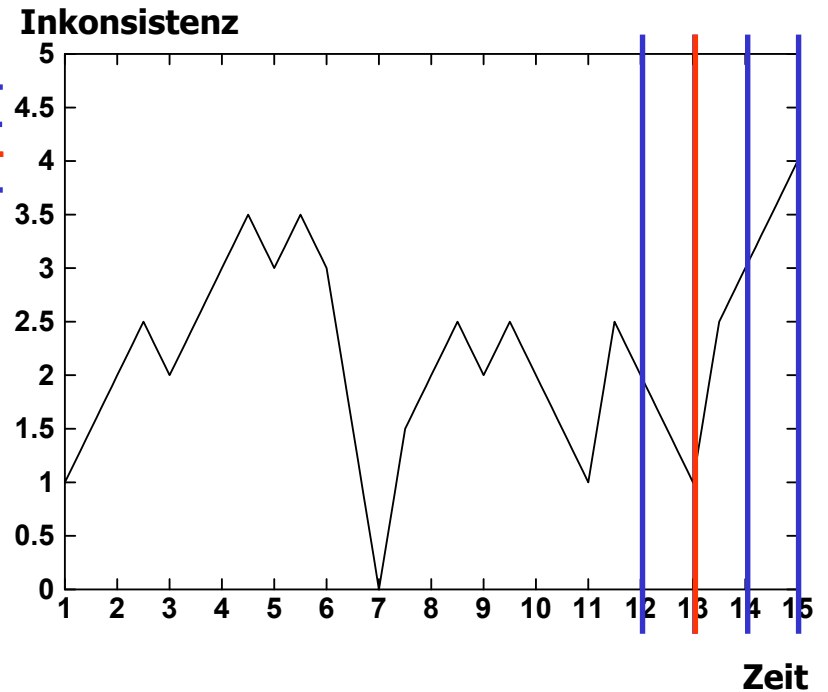
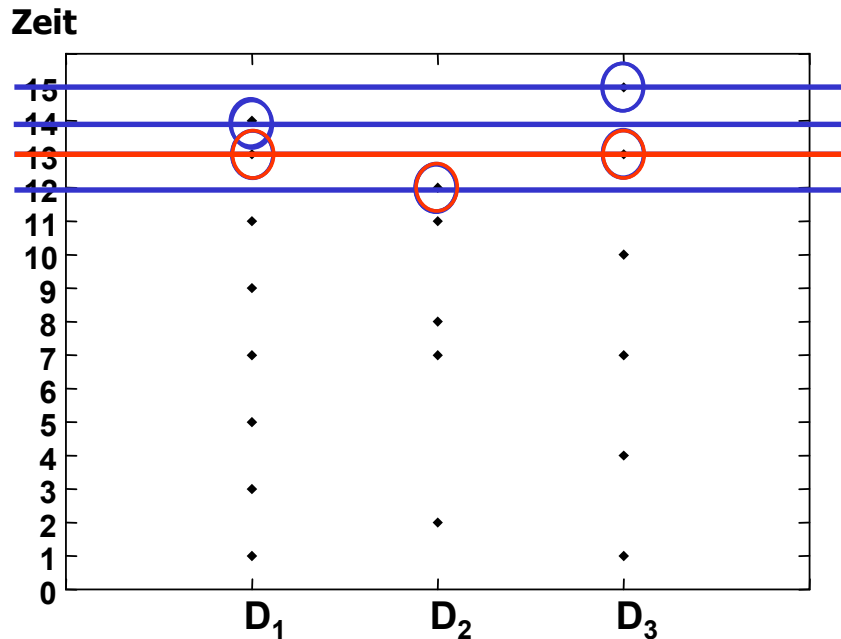
- Eingabe
 - Historischer Schnitt t_c
 - Inneres und äußeres Konsistenzband
- Ausgabe
 - Menge von Snapshots
- Idee
 - Auswahl der am nächsten zum historischen Schnitt liegenden Snapshots
 - Kein Snapshot innerhalb des inneren oder äußeren Konsistenzbandes
→ Aktualisierung
 - Strategien für die Aktualisierung: retro-/prospektiv, rekonstruiert

Algorithmische Betrachtung: Bestimmung des historischen Schnittes



- Eingabe
 - Aktualisierungsstrategie für Snapshotauswahl
- Ausgabe
 - Historischer Schnitt t_c
- Idee
 - Start zum aktuellen Zeitpunkt
 - Berechne Lösungen für verschiedene historische Schnitte t_c
 - Auswahl der besten Lösung

Algorithmische Betrachtung: Bestimmung des historischen Schnitts (2)

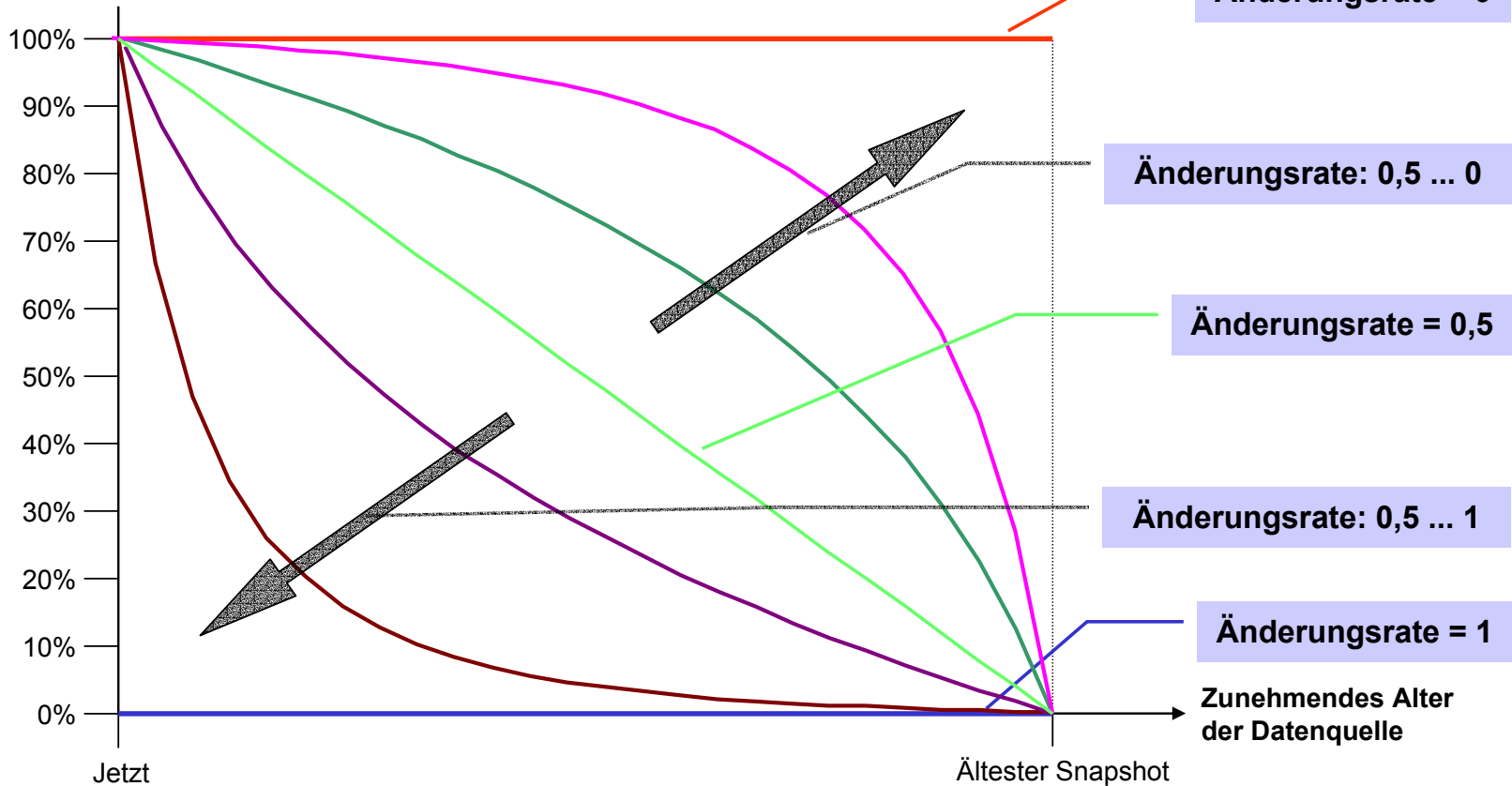


Erweiterung der Konsistenzquantifizierung

- Bisher
 - Berücksichtigung des zeitlichen Abstandes vom historischen Schnitt
- Jetzt
 - Absolutes Alter der Snapshots
 - Änderungsrate der Daten: Durchschnittlicher Prozentsatz der geänderten Daten zwischen zwei Snapshots pro Zeiteinheit
- Wiederverwendungsgrad $\rho(t)$
 - Zusammenführung von Snapshot-Alter und Änderungsrate
 - Idee
 - Hoher Wiederverwendungsgrad alter Snapshots bei niedriger Änderungsrate
 - Niedriger Wiederverwendungsgrad bei hoher Änderungsrate unabhängig vom Alter der Snapshots

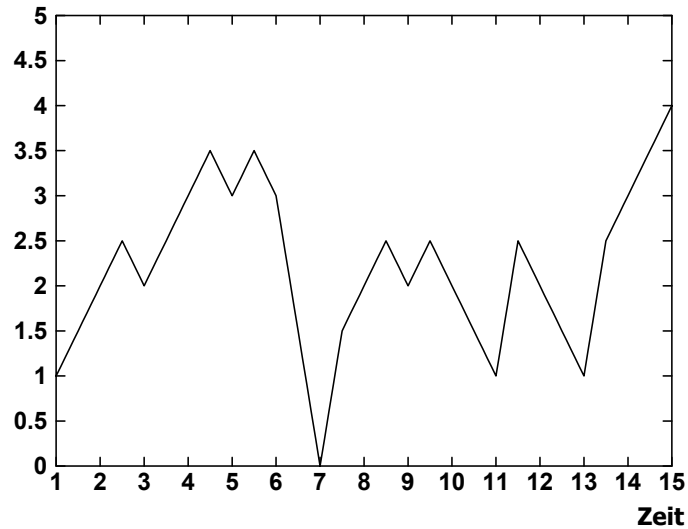
Wiederverwendungsgrad: Graphische Veranschaulichung

Wiederverwendungsgrad

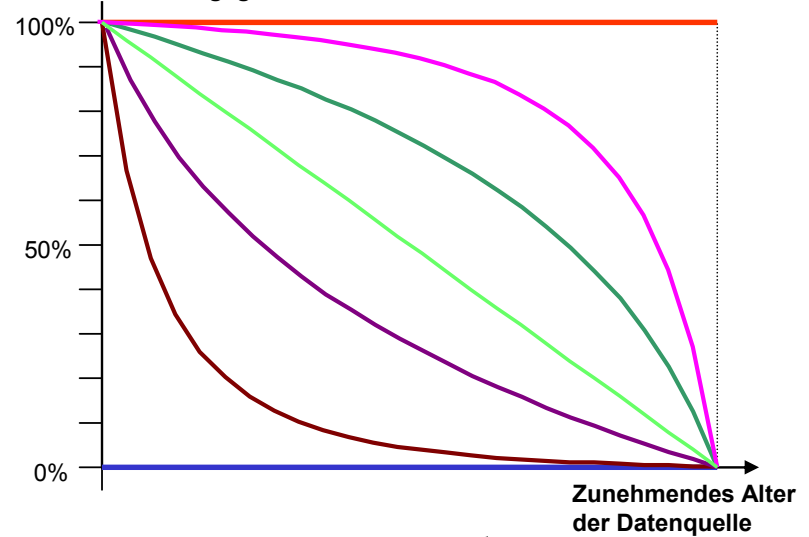


Zusammenführung von Abstand und Wiederverwendungsgrad

Inkonsistenz



Wiederverwendungsgrad



Wie kombinieren?

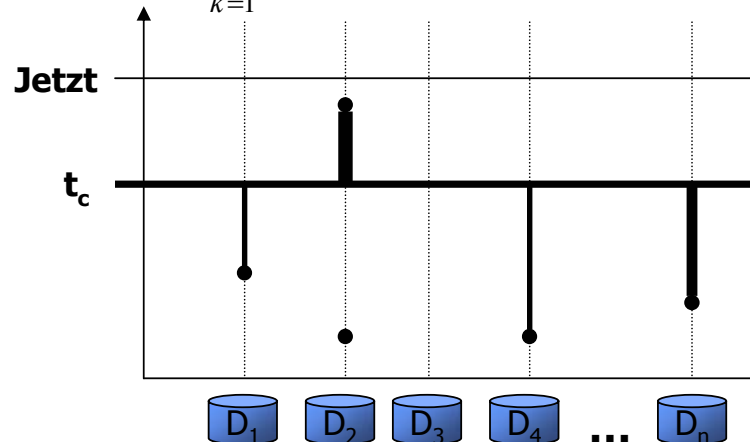
Inkonsistenzmaß für zeitlichen Abstand und Wiederverwendungsgrad

- Inkonsistenzmaß für einen Snapshot einer Datenquelle

$$I_i(t_S) = |t_c - t_{S_i}| \cdot (1 - \rho_i(t_S))$$

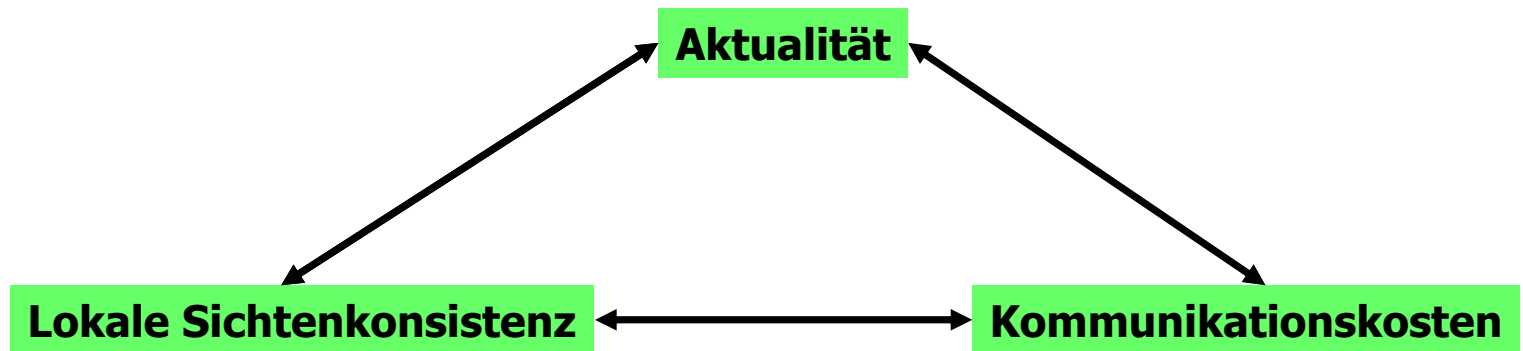
- Inkonsistenz über alle Datenquellen

$$I(t_{S_1}, \dots, t_{S_n}) = \sum_{k=1}^n (\alpha_k \cdot |t_c - t_{S_k}| \cdot (1 - \rho_i(t_{S_k})))$$

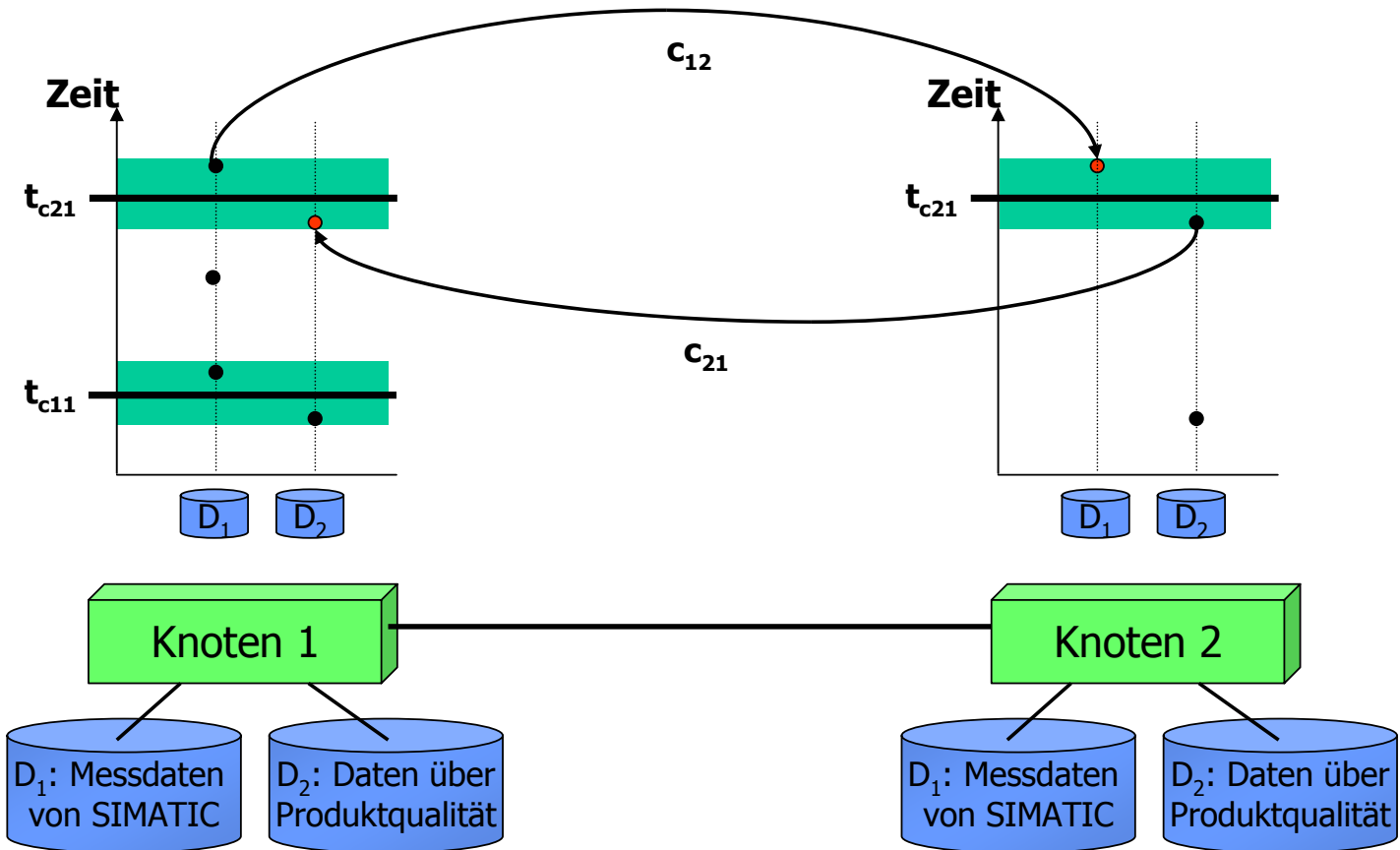


Ausblick: Anfrageverteilung im Grid

Grid-Konfliktdreieck



Ausblick: Anfrageverteilung im Grid



Zusammenfassung

- Ausgangssituation
 - Grid-Datenbanksystem
 - Messdaten mit unterschiedlichen Gültigkeitszeiten
- Idee
 - Rahmenwerk zur Kombination und zur Konsistenzquantifizierung der Daten
- Überblick
 - Konsistenzquantifizierung unter Berücksichtigung von
 - zeitlichem Abstand zum historischen Schnitt
 - absolutem Alter der Snapshots und Änderungsrate → Wiederverwendungsgrad
 - Erläuterung von Verbundsemantiken
 - Algorithmische Betrachtung der Snapshotauswahl
- Ausblick
 - Anfrageverteilung im Grid