



DÍGAME: A Vision of an Active Multidatabase with Push-based Schema and Data Propagation

Christopher Popfinger · Cristian Pérez de Laborda · Stefan Conrad

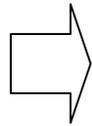
12. Februar 2004



Institute of Computer Science
Databases and Information Systems

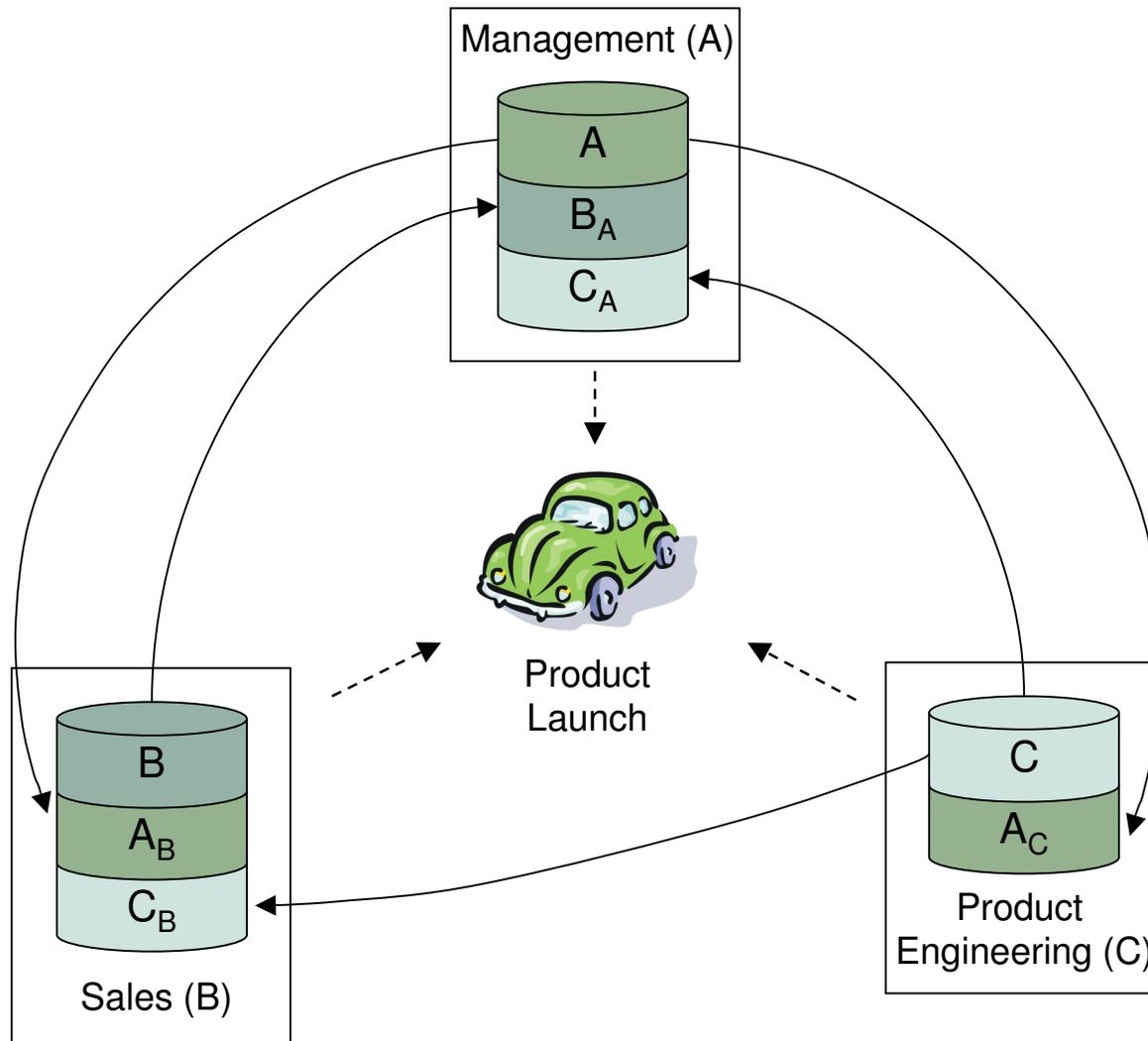
Motivation

- Viele Unternehmen und Organisationen besitzen große Anzahl von Datenbanken
- Typischerweise autonome und heterogene Datenquellen
- Zugriff auf gemeinsamen Datenbestand bei einer (zeitweisen) Kooperation (Virtuelles Unternehmen)
- Informationsplattform angepasst an Datenpolitik
- Integration bestehender Datenbestände
- Erhalt der lokalen Autonomie



DÍGAME: A **D**ynamic **I**nformation **G**rid in an
Active **M**ultidatabase **E**nvironment

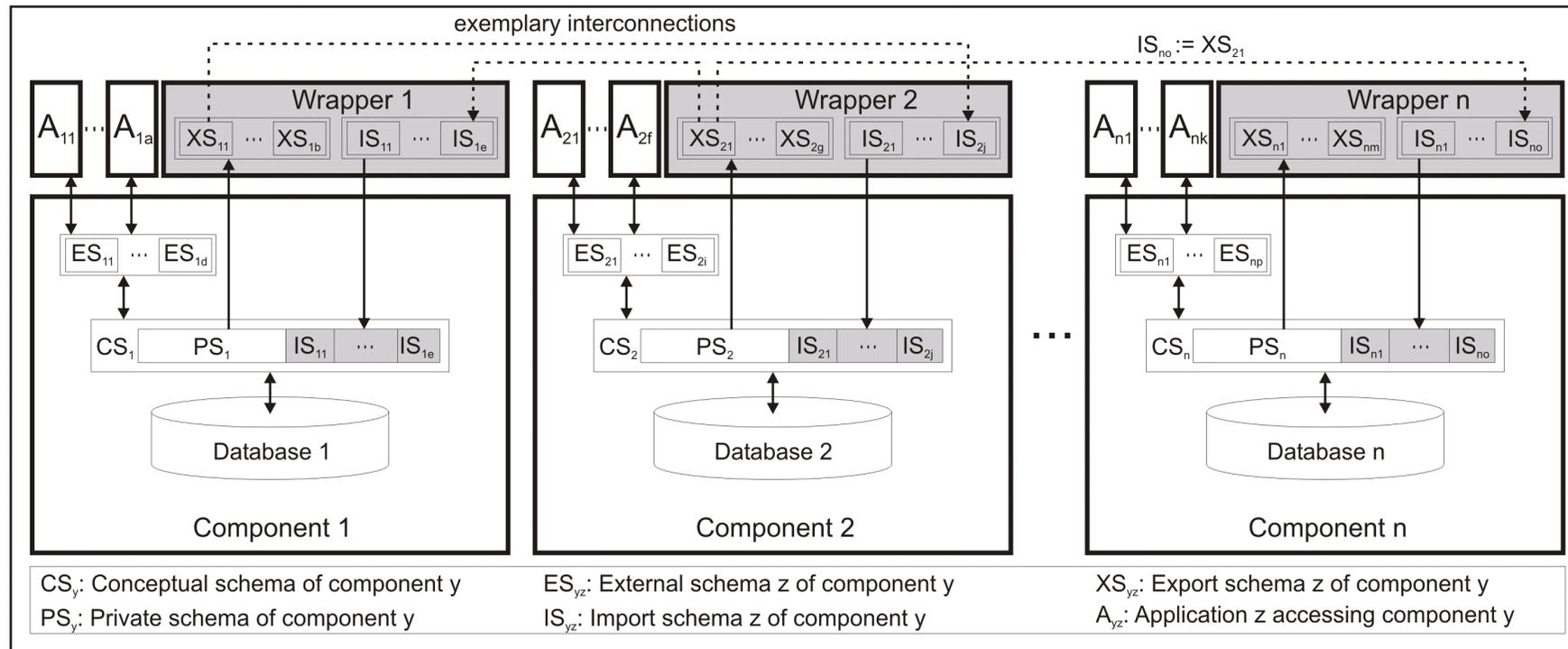
Anwendungsszenario



Funktionsweise

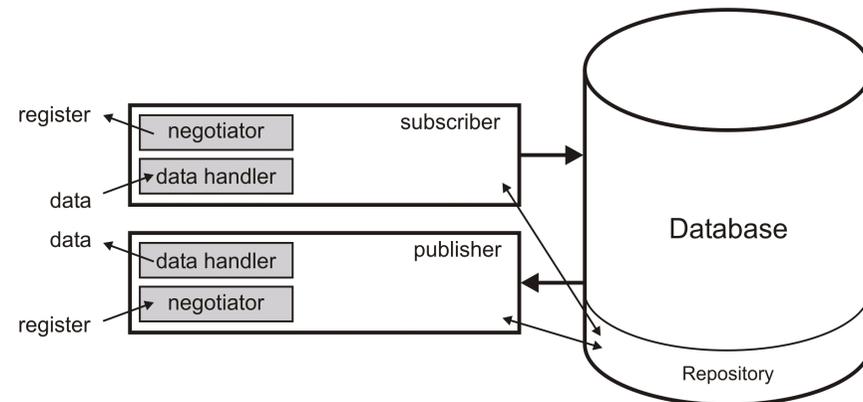
- Administrator macht eine Sicht seiner Daten zugänglich
- Interessierte Knoten *abonnieren* diese Daten direkt von der Quelle
- Initiale Übertragung der gesamten Daten
- Änderungen von Daten und Schema werden von der Quelle an die Abonnenten geschickt
- Abonnent hält eine lokale Kopie aller abonnierten Daten
- Angepasste Replikationsstrategien: sofort, aggregiert, periodisch,...

DÍGAME Architektur



Wrapper-Komponente

- Implementiert Schnittstelle zur Datenbank
- Separate Kommunikationskomponenten zum Senden und Empfangen von Daten
- Metadatenverwaltung in lokaler Datenbank integriert
- Unterschiedliche Module zur Vermittlung (Negotiator) und zur Datenübertragung (Data Handler)
- Zugriffskontrolle



Eigenschaften von DIGAME

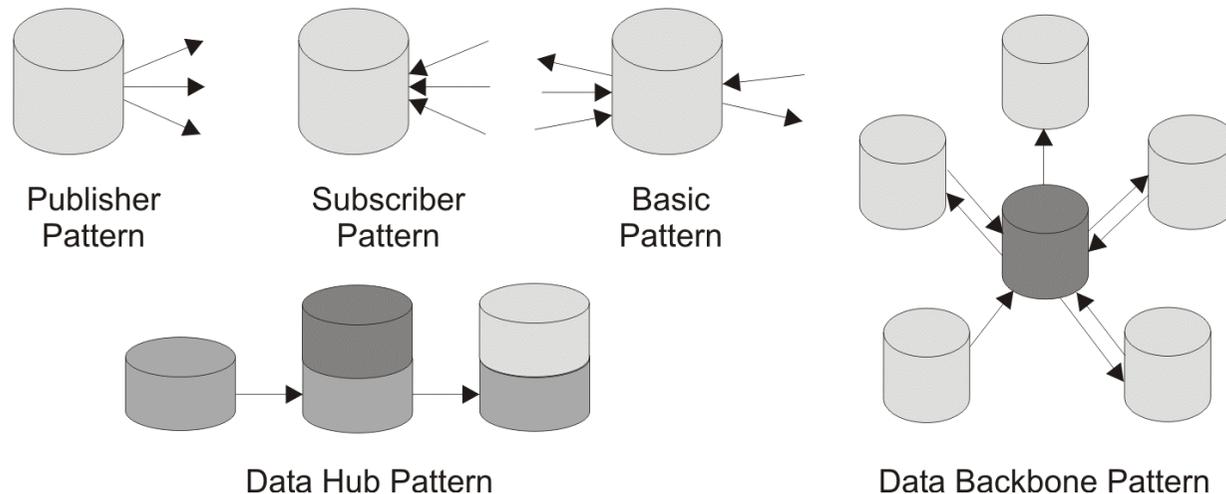
- Lose gekoppelte, aktive Multidatenbank
- Dynamischer Verbund autonomer und heterogener Datenquellen (P2P-Konzepte)
- Erhalt der Autonomie und der Datenhoheit auf jedem teilnehmenden Peer
- Keine zentrale Kontrollinstanz
- Administrator des Peers entscheidet über Art und Umfang der angebotenen Daten

Eigenschaften von DIGAME (II)

- Daten können nur durch den Dateneigner verändert werden \Rightarrow Datenqualität
- Aktive (push-based) Propagation von Änderungen an Daten und Schema durch die Datenquelle
- Pull-based Fallback
- Standardisiertes Austauschformat für Schema und Daten, z.B. Resource Description Framework (RDF)
- Replikation der importierten Daten:
 - Lokaler Zugriff auf alle benötigten Daten
 - Überbrückung von Netzwerkausfällen
 - Reduktion der Netzlast bei überwiegend lesenden Zugriffen

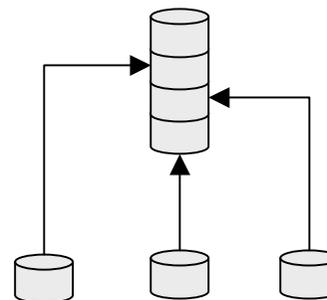
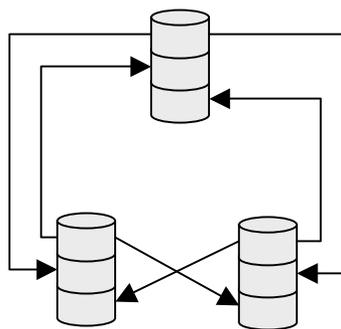
DÍGAME Link Patterns

- Zusammensetzung des Information Grid entweder geplant oder dynamisch entstanden
- Identifizierung von Verknüpfungsmuster (Link Patterns) zwischen den Peers



Weitere Anwendungsgebiete

- **e-health:** Zugriff auf gesamte Krankengeschichte eines Patienten durch Hausarzt, Krankenhaus, Labor
- **e-government:** Integrierter Zugriff auf verteilte Daten von Verwaltungseinheiten (Städte, Gemeinden, Polizei, Finanzbehörde)
- **e-science:** Kontrollierter integrierter Zugriff auf Resultate kooperierender Forschungseinheiten



Nächste Schritte

- Verfeinerung der Architektur
- Kommunikationsprotokolle, Austauschformat
- Erkennung und Signalisierung lokaler Ereignisse
- Datenschutz und Zugriffskontrolle
- Mechanismen für Replikation und Recovery
- Vermeidung von Anomalien
- Aufsetzen einer Testumgebung
- Analyse der Kommunikationskosten

Vielen Dank für die Aufmerksamkeit!

