Skyline Query Processing in P2P Netze

Seminar Ringvorlesung - WS 2010/2011

Julissa Cusi Juarez

Department Informatik
HAW Hamburg

02.12.2010

Informatik @ HAW HAMBURG

Informatik @ HAW HAMBURG

Agenda

1. Skyline Query Processing
2. Masterarbeit
   - 2.1. Motivation
   - 2.2. Szenario
   - 2.3. Vorarbeiten
   - 2.4. Methodischer Ansatz
   - 2.5. Risiken
3. Zusammenfassung und Ausblick

a. Literatur

Informatik @ HAW HAMBURG

1. Skyline Query Processing

Beispiel 1:
Suche nach Hotels in Hamburg

- ✓ In der Nähe der Alster
- ✓ Preisgünstig

1. Skyline Query Processing

Beispiel 2:
Suche nach Flügen

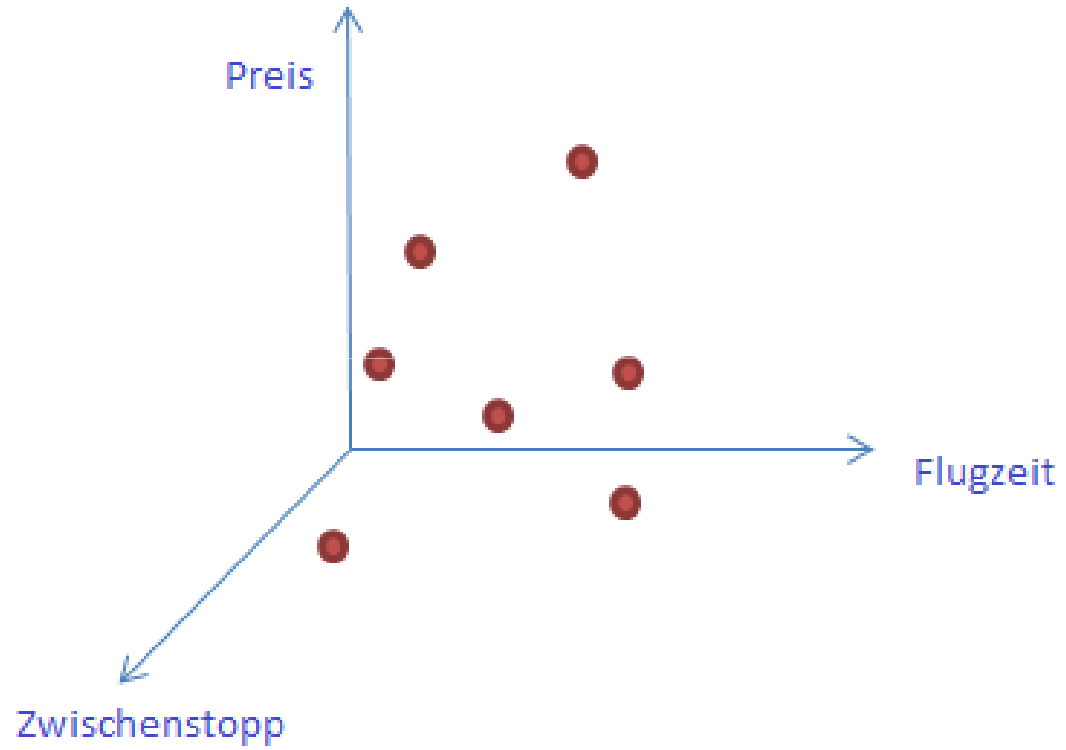
- ✓ Preisgünstig
- ✓ Wenig Flugzeit
- ✓ Wenige Zwischenstopps

1. Skyline Query Processing

Beispiel 2: Suche nach Flügen

Dimensionen :

- ✓ Preis : minimun
- ✓ Flugzeit : minimun
- ✓ Zwischenstopp : minimun



Ergebniss:

- ✓ Datenpunkte -Menge

2. Masterarbeit

Informatik @ HAW HAMBURG

2.1. Motivation

- Verteilte Systeme: Verteilte Datenbanken in P2P Umgebungen.
- Ein Bereich der P2P Datenbanken ist Anfragebearbeitung (query processing)
- Daten in einem P2P System sind nicht permanent verfügbar.
- Man kann spezielle Anfragen wie Skyline Queries anwenden.
- Skyline Queries sind für multikriterielle Entscheidungsunterstützung geeignet.

2. Masterarbeit

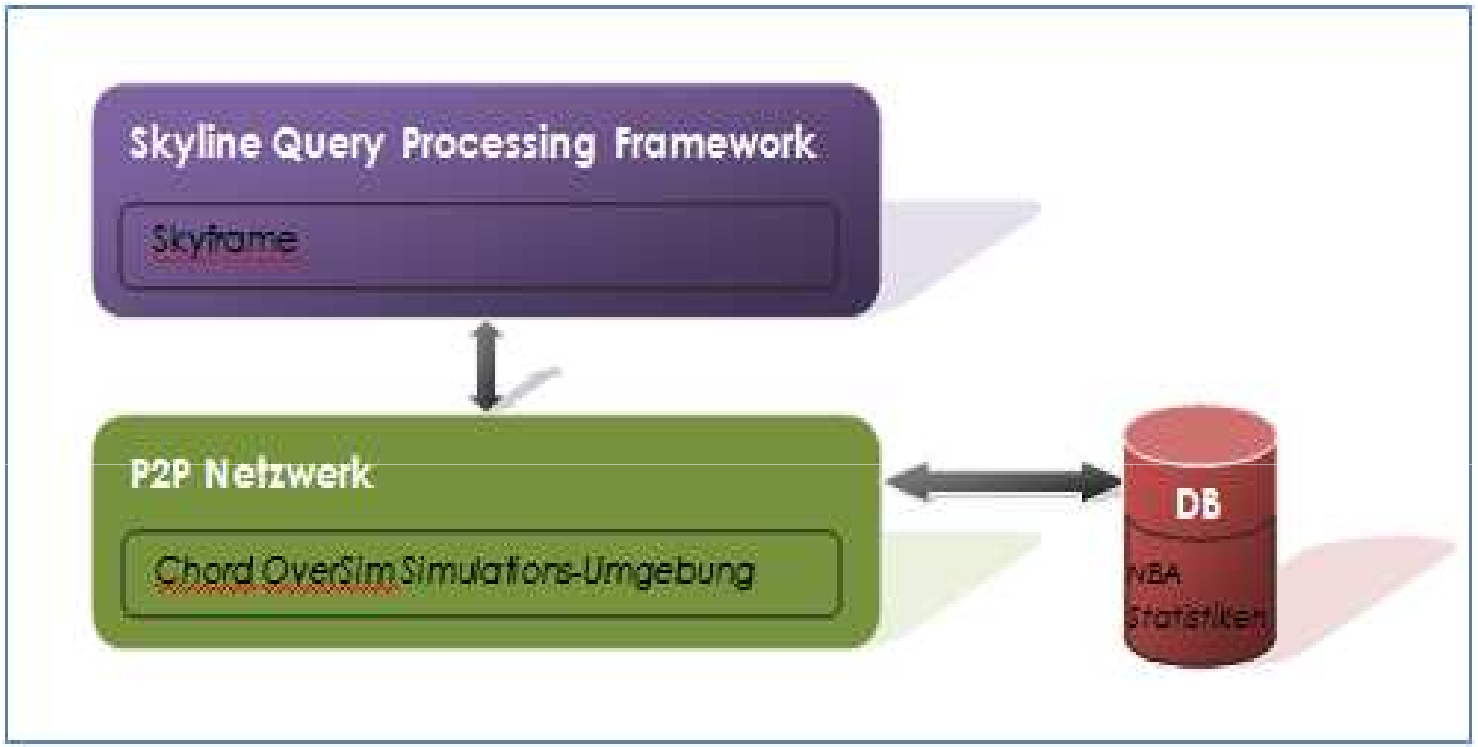
Informatik @ HAW HAMBURG

2.2. Szenario

- Implementierung eines Skyline Query Processing Frameworks über eine P2P Simulationsumgebung, als experimentelles Szenario.
- Verwendung einer vordefinierten Datenbank.
- Ziel:
  Skyline Query Processing Verfahren evaluieren.
- Basis für die Masterarbeit.

2. Masterarbeit

2.2. Szenario



Architektur

2. Masterarbeit

Informatik @ HAW HAMBURG

2.3. Vorarbeiten (Projekt 1 und 2)

Implementierung simulierter P2P Umgebung:
Overlay Framework OverSim

- Erstellt 2007 bei dem Institut für Telematik der Universität Karlsruhe.
- Strukturierte und unstrukturierte Overlay-Protokolle: Chord, Kademlia, Pastry, Bamboo, Koorde, Broose, Gia und Vast.
- Auf OMNeT++ basierendes Overlay-Framework für Linux, Windows und Mac OS X.
- OMNeT++: Simulationsbibliothek und ein Framework, um Netzwerk Simulatoren aufzubauen.

OverSim
The Overlay Simulation Framework

Quelle [3, 4]

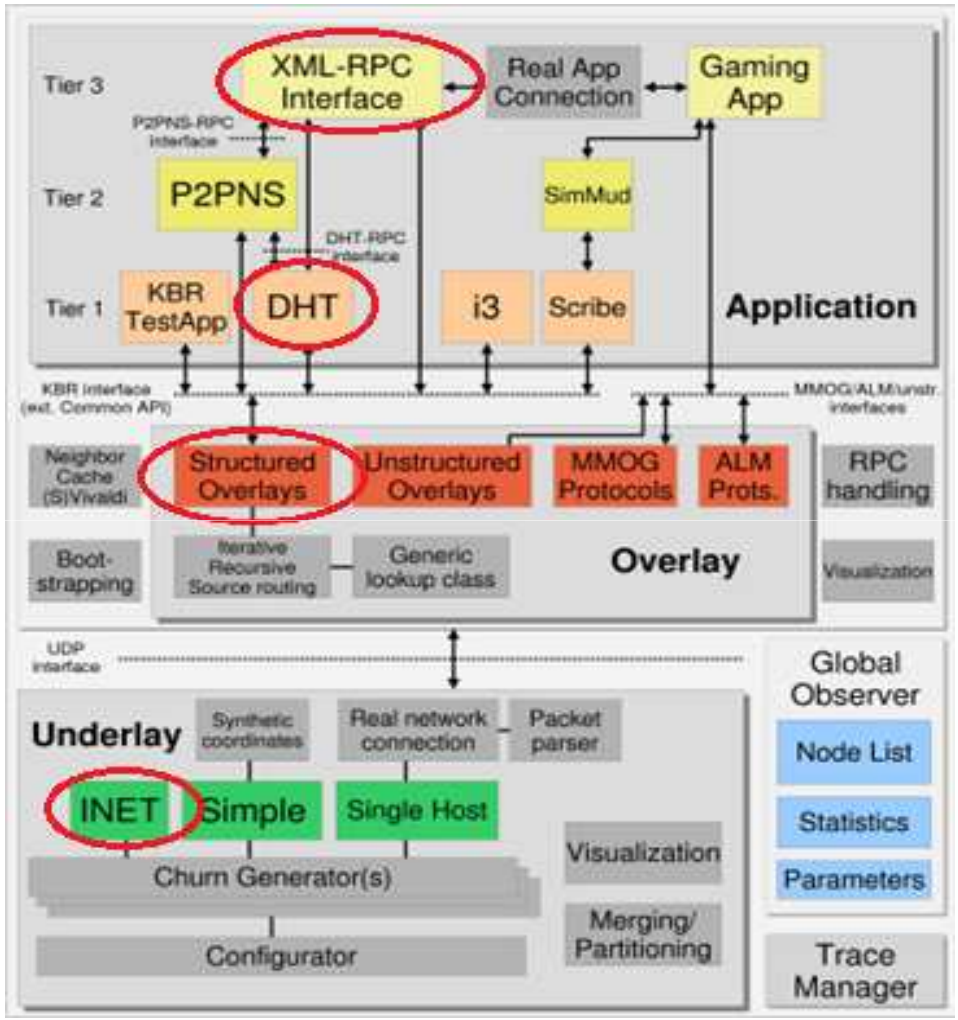
2. Masterarbeit

Informatik @ HAW HAMBURG

2.3. Vorarbeiten (Projekt 1 und 2)

Implementierung simulierter P2P Umgebung :
Overlay Framework OverSim

- Drei Schichten: Underlay, Overlay und Application
- Konfiguration:
  - Underlay: INET
  - Overlay : Chord
  - Application: DHT, XML-RPC



Quelle [3]

2. Masterarbeit

Informatik @ HAW HAMBURG

2.3. Vorarbeiten (Projekt 1 und 2)

Implementierung simulierter P2P Umgebung :
Overlay Framework OverSim

- Die Konfiguration soll die Simulation mehrerer Knoten und die Kommunikation zwischen ihnen und einer externen Application ermöglichen.

- Kommunikation durch Tunneling:
  Simulierte Umgebung baut ein virtuelles Device, das mit externen Applikationen mittels TUN Interface von Linux-Kernel kommuniziert.

2. Masterarbeit

Informatik @ HAW HAMBURG

2.3. Vorarbeiten (Projekt 1 und 2)

Verteilung der Datenbank:
Testdatenbank von NBA

- NBA Player und Spiele Statistiken von 1946 bis 2009 aus www.databasebasketball.com
- CSV Dateien
- Verteilt auf 10 Knoten, ohne Redundanz

2. Masterarbeit

Informatik @ HAW HAMBURG

2.3. Vorarbeiten (Projekt 1 und 2)

Implementierung Skyline Query Processing Framework: Skyframe

- Framework für effiziente Skyline Query Processing in P2P Netze.

- Ziel:

  Den Zeitverlauf von Anfrageverarbeitung zu optimieren, die Netzwerkkommunikationskosten zu reduzieren und die Query-load durch die Peers zu balancieren [1].

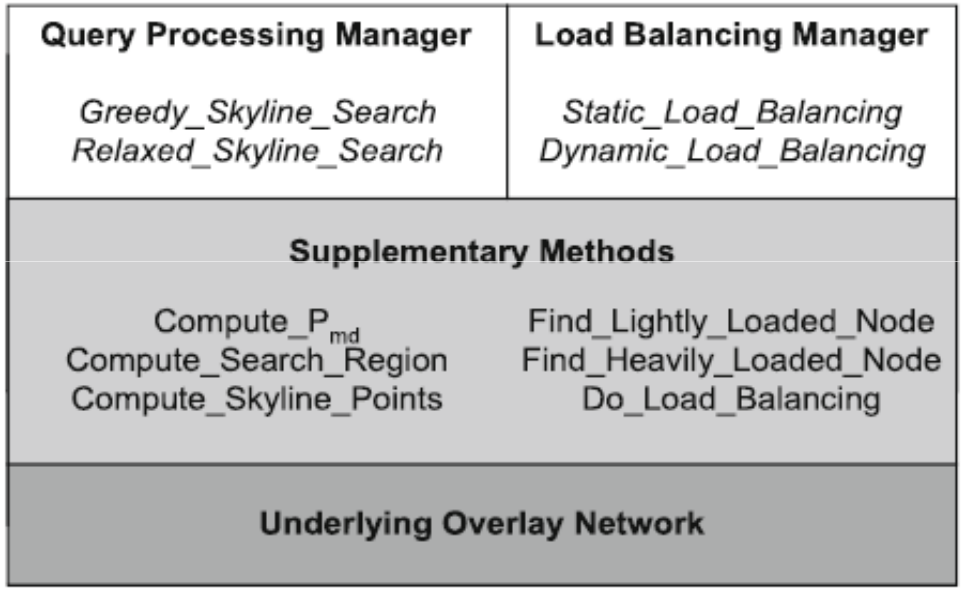
2. Masterarbeit

Informatik @ HAW HAMBURG

2.3. Vorarbeiten (Projekt 1 und 2)

Implementierung Skyline Query Processing Framework: Skyframe

- Keine verfügbare Version. Nur GSS und RSS Algorithmen im Paper verfügbar.
- Selbst implementiert und getestet.



Skyframe [1]

2. Masterarbeit

Informatik @ HAW HAMBURG

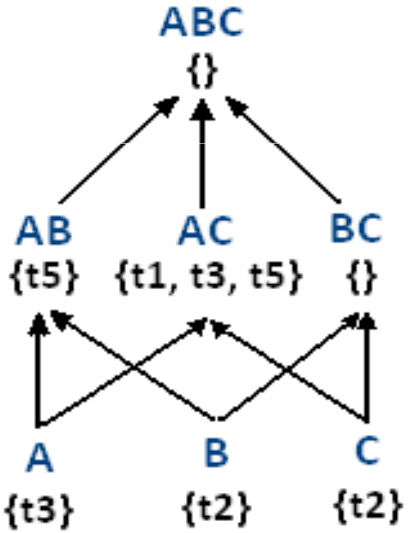
2.4. Methodischer Ansatz

Neue Ansätze

- Compressed Skycubes:
  Die verteilten Daten in jedem Knoten werden strukturiert durch den Aufbau von Skycubes gemäß ihren Dimensionen.

	A	B	C
t1	40	30	4
t2	50	10	5
t3	10	40	2
t4	30	50	1
t5	20	20	3

Dataset



Skycube
Lattice Struktur

Cuboid	Skyline
AB	{t5}
AC	{t1, t3, t5}
A	{t3}
B	{t2}
C	{t2}

Compressed Skycube

Quelle [2]

2. Masterarbeit

Informatik @ HAW HAMBURG

2.4. Methodischer Ansatz

Neue Ansätze

- Skyline Operator:
  SQL-Extension durch Implementierung Skyline Operator „SKYLINE OF“.

  Die verteilten Daten in jedem Knoten können in mehreren Tabellen gespeichert sein. Skyline Operator versucht, die join-Berechnungskosten zu minimieren.

```
SELECT *
FROM Tabelle
SKYLINE OF Dimension1 MAX, Dimension2 MIN
```

Quelle [5]

2. Masterarbeit

Informatik @ HAW HAMBURG

2.4. Methodischer Ansatz

Neue Ansätze

- Skyline Operator:
  Forschungsprojekt von PostgreSQL Community hat Probeversion von Skyline Operator implementiert [6].

2. Masterarbeit

2.4. Methodischer Ansatz

Herausforderung

- Anwendung in einem echten P2P Netz
- Anwendung in anders Overlay-Netzen außer Chord
- Beliebige Knotenanzahl
- Hochdimensionaler Raum
- Verteilung der Datenbank mit Redundanz

2. Masterarbeit

2.4. Methodischer Ansatz

Kriterien zur Evaluierung

- Korrekheit der Ergebnisse
- Antwortzeit
- Kommunikationskosten
- Arbeitslast in jedem Peer

2. Masterarbeit

Informatik @ HAW HAMBURG

2.5. Risiken

- Kein optimales Funktionieren in anderen Overlay-Netzen.
- Simulationsergebnisse zu weit von der Realität entfernt.
- Simulator Bibliotheken nicht ausreichend.
- Verteilte Daten nicht optimal strukturiert.

Informatik @ HAW HAMBURG

3. Zusammenfassung und Ausblick

- Ein experimentelles Szenario wird in den Vorarbeiten erstellt, es wird die Basis der Masterarbeit bilden.
- Nach der Evaluierung des experimentellen Szenarios wird ein Skyline Query Processing Verfahren weiterentwickelt,
  - mit Berücksichtigung neuer Ansätze,
  - um die Skyline Suche zu optimieren und
  - die Last in jedem Knoten zu minimieren.

Informatik @ HAW HAMBURG

A. Literatur

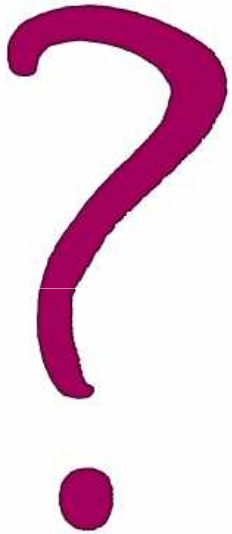
[1] J. Cusi Juarez. *P2P Datenbanken*. Ausarbeitung im Rahmen der Vorlesung Anwendungen I im Studiengang Informatik Master of Science am Studiendepartment Informatik der Fakultät. Technik und Informatik der Hochschule für Angewandte Wissenschaften Hamburg, 2010.

[2] J. Cusi Juarez. *Skyline Query Processinng*. Ausarbeitung im Rahmen der Vorlesung Anwendungen II im Studiengang Informatik Master of Science am Studiendepartment Informatik der Fakultät. Technik und Informatik der Hochschule für Angewandte Wissenschaften Hamburg, 2010.

[3] www.oversim.org

[4] www.omnetpp.org

[5] S. Borzsonyil, D. Kossmann, K. Stocker. *The Skyline Operator*. In 17th International Conference on Data Engineering, 2001.

[6] http://randdataset.projects.postgresql.org/

Vielen Dank für Ihre Aufmerksamkeit !