Tripcom Meeting on Scalability

Elena Simperl, DERI Innsbruck

Agenda

- 05.07.2007 10:00 - 18:00
  - 10:15 - 12:30 Building Scalable Systems
    - 10:15 - 11:00 Scalable Tuplespaces (Antony)
    - 11:00 - 11:45 Scalable Semantic P2P (Ronny)
    - 11:45 - 12:30 SwarmLinda (Robert)
    - 12:30 - 13:15 Scalability in HPC (Hans)
  - 13:15 - 14:15 Lunch break
  - 14:15 - 15:00 Scalability guide for TripCom (Daniel M.)
  - 15:00 - 16:00 Scalability in TripCom (Manfred)
  - 16:00 - 16:15 Coffee break
  - 16:15 - 18:00 TripCom Scalability Task (Elena)
  - 19:00 Dinner
- 06.07.2007 10:00 – 18:00
  - Agreement on definitions, objectives and action plan

TripCom Scalability Task

Elena Simperl, DERI Innsbruck

with contributions from TUW and USTUTT

Overview

- Step 1: Agreement on overall objectives of the scalability task
- Step 2: Agreement on general assumptions, concepts and definitions
- Step 3: Identification of scalability factors and trade-offs
- Step 4: Realization
- Step 4: Evaluation of the results

Step 1: Overall objectives

- Triplespace is meant to provide an abstraction layer to coordinate access and management to various Semantic Web data sources
- The general objective is to built a triplespace platform that should scale at Internet level
- Use cases (WPs 8a and 8b) are show cases

- M24

- M36

Sixth Framework Programme Priority 2
Information Society Technologies (IST)
Specific Targeted Research Project

Scalability guidelines

- Exclude
  - Transactions
  - Locking
  - Deletion
  - Global state
  - Consistency
  - Completeness
- Define levels of scalability which provide some of the aforementioned features at the cost of scalability

Levels of scalability

- Level 0 (scalable)
  - Read and write tuples
  - Issues
    - Read: where to find a tuple?
      - Build distributed index
      - Query routing
      - Read does not guarantee completeness
    - Write: where to store a tuple?
      - Distribution strategy
      - Users do not have control over where the information is stored or cached (Google-like)
    - Signature of the operations
      - URIs vs. content
  - Possible implementation based on Gnutelly or RDF Peers

Levels of scalability (ii)

- Level 1 and higher
  - Write graphs
  - Multiple spaces
  - Deletion (semantics TBD)
  - Security
  - Alerts (problematic to implement)

Sixth Framework Programme Priority 2
Information Society Technologies (IST)
Specific Targeted Research Project

Step 2: Assumptions, concepts and definitions

- What are our assumptions?
  - Bandwidth, openess, decentralization, dynamism, orders of magnitude for data volume, nodes, machines, number of concurrent queries, query complexity
- How we define the size of the problem?
  - Number of tuples, nodes, spaces, users, query expressivity
- What do we measure and how do we measure scalability?
- What is a scalable triple space?
- What additional hardware and software is necessary to demonstrate scalability?

Assumptions

- Dynamic environment (join and leave)
- Decentralized
- Bandwidth will be ignored (network delays)
- Arbitrarily high number of nodes
- Arbitrarily large data sets and numbers of (concurent) queries
- Query complexity levels to be defined
- For testing purposes
  - distributed environment (4 machines)
  - 100 000 000 triples/machine

How to measure and demonstrate scalability

- No analytical model
- Measurements of the time of basic operations
- Distributed environment (4 machines)

Step 3: Factors and trade-offs

- Collect assumptions from use cases
- Agree on procedure model how to identify scalability rules / guidelines to be followed in TripCom
- Identify general factors and trade-offs at system level and component level

Steps 4/5: Realization and evaluation

- Accomplish tasks taking into account results of Step 3
- Prepare test environment
- Evaluate and analyze

TO DO's: Agree on resources

- Totally 24 PM
  - Sirma 2
  - USTUTT 4
  - TUW 5
  - NUIG 2
  - Telefonica 1
  - CEFRIEL 2
  - FU Berlin 4
  - DERI 4
- 2nd year: 14 PM
- 3rd year: 10 PM

TO DO‘s: Schedule

- Step 3: Factors and trade-offs
  - Use case assumptions
    - August, 15
  - General factors and trade-offs
    - September, 15
  - Component-relevant
    - September, 29
  - Final discussion at the general assembly meeting in October in Milan
- Step 4:
  - Refinement of the architecture and implementation of the components
    - October 2007 – March 2008
  - Preparation of the test environment
    - October 2007 – March 2008
- Step 5:
  - Evaluation and analysis
    - Starting with April, 2008

Sixth Framework Programme Priority 2
Information Society Technologies (IST)
Specific Targeted Research Project

TO DO‘s: Scalability deliverable

- V1: M24
  - M18: Documentation of the results of Steps 1, 2 and 3
  - M21: Documentation of the results of Step 4
- V2: M36
  - Final document including evaluation results