Motivation

- Large volumes of available RDF data
- By April of 2014 there were 1091 datasets available online (~terabytes of RDF data)
- Size almost doubling every year

Problem: Isolated computing engines cannot handle the huge load of RDF data

Approach: Massively parallel processing systems to store and query RDF data

Challenge: How to efficiently evaluate an RDF query in a massively parallel context

Query processing overview

Build massively parallel flat (Height Optimal) logical plans using n-ary equality joins

CliqueSquare Optimization algorithm:
- Uses CliqueDecompositions for identifying possible joins
- Exploits CliqueReductions for applying the joins
- Terminates when the variable graph has one node
- The Height Optimality (HO) of the plan depends on the choice of the decomposition function

CliqueSquare: Flat Plans for Massively Parallel RDF Queries

Benjamin Djahandideh, François Goasdoué, Zoi Kaoudi, Ioana Manolescu,
Jorge Quiané-Ruiz, Stamatis Zampetakis

Optimization time (ms) per algorithm & query shape

Inclusions among the plan spaces of CliqueSquare variants

Comparison of CliqueSquare flat plans with binary linear plans and binary bushy plans

Comparison of CliqueSquare with existing systems: SHAPE & H2RDF+

CliqueSquare is available for download:
http://sourceforge.net/projects/cliquesquare/

Experimental Evaluation

SPARQL query

Parser

Logical Optimizer

Physical Optimizer

Job Translator

Hadoop

MSC+ ◊ Maximum Cliques ◊ Minimal Set Cover ◊ HO-partial

MXC+ ◊ Maximum Cliques ◊ Minimum Exact Cover ◊ HO-lossy

XC+ ◊ Partial Cliques ◊ Exact Cover ◊ HO-lossy

MXK ◊ Partial Cliques ◊ Minimum Exact Cover ◊ HO-lossy

SC+ ◊ Partial Cliques ◊ Minimal Set Cover ◊ HO-partial

MSC ◊ Partial Cliques ◊ Minimum Set Cover ◊ HO-partial

SC ◊ Partial Cliques ◊ Set Cover ◊ HO-complete

Basic graph pattern query

Variable Graph

WHERE {
  SELECT ?x ?y
  WHERE {
    :p1 <C1> .
    :p7 ?f .
    :p6 <C4> .
    :p2 <C2> .
    :p11 <C6> .
  }  
  (?x | ?y) \in (\{"a", "b", "c", "d"\})

Contact: stamatis.zampetakis@inria.fr

https://team.inria.fr/oak/projects/cliquesquare/