

R17-1:CliqueSquare: Flat Plans for Massively Parallel RDF Queries

François Goasdoué (University of Rennes 1, INRIA*),

Zoi Kaoudi (IMIS, Athena Research Center),

Ioana Manolescu (INRIA*, France & University of Paris Sud),

Jorge-Arnulfo Quiané-Ruiz (Qatar Computing Research Institute),

Stamatis Zampetakis (INRIA*, France & University of Paris Sud)

発表: 筑波大(北川研) M2 熊本 和正

INRIA - Benjamin Djahandideh

図は論文とポスターから引用しました.

背景

- Problem
 - 一つの計算機で扱えるRDFデータの量には限度がある
- Approach
 - RDFデータを大規模に並列分散して格納・問合せする
- Challenge
 - どうすれば問合せ効率が良くなるか？

提案手法の流れ

- MapReduce処理に適したプラン木をつくる
 1. 問合せを変数グラフに変換
 2. 変数グラフをCliqueという単位にグループ化
 3. Cliqueを更にグループ化して一つのノードになるまで繰り返す
- 木ができたらMapReduceのJobとして実行

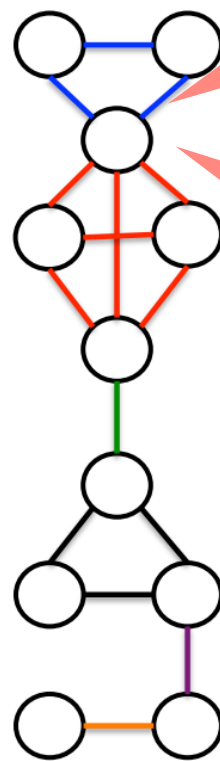


問合せを変数グラフに変換

- これにより n-ary join のグループ(Clique)ができる

```
SELECT ?x ?y
WHERE {
1: ?w :p1 <C1> .
2: ?w :p2 <C2> .
3: ?w :p3 ?x .
4: ?x :p4 <C3> .
5: ?x :p5 <C4> .
6: ?x :p6 ?z .
7: ?z :p7 ?f .
8: ?f :p8 <C5> .
9: ?f :p9 ?y .
10: ?y :p10 ?h .
11: ?h :p11 <C6> .}
```

問合せ
(SPARQLクエリ)



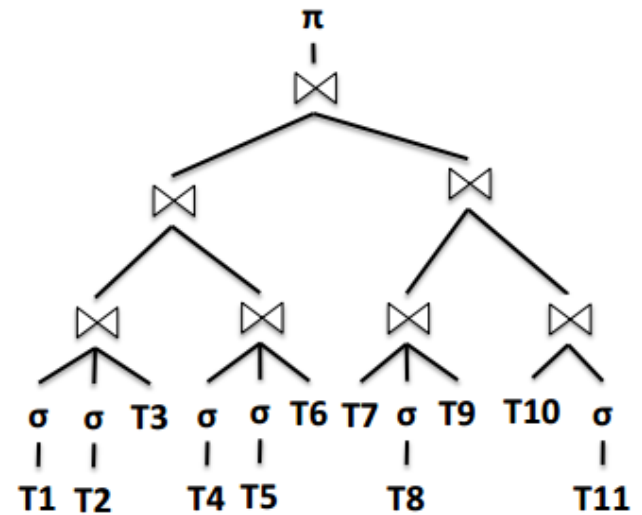
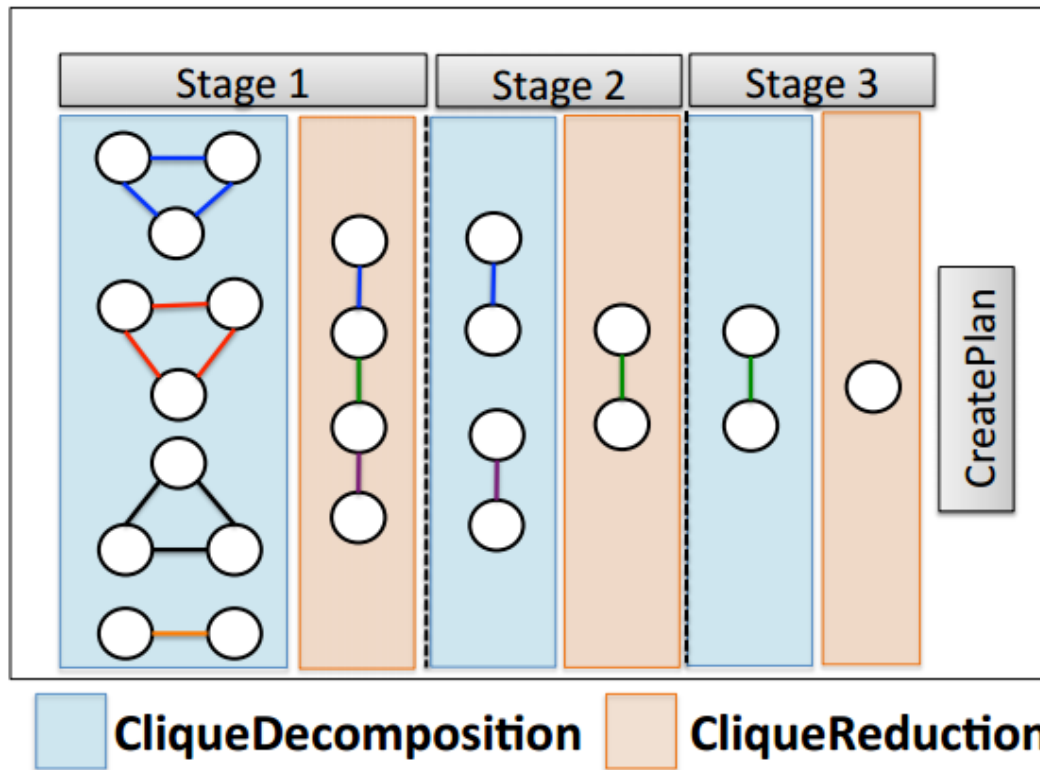
各ノードが
各トリプル

各エッジが
各変数

変数グラフ

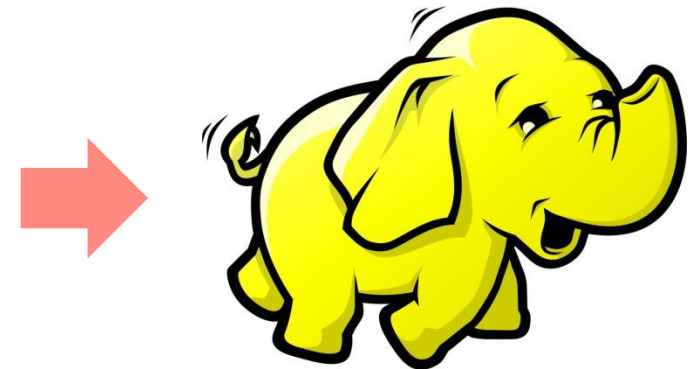
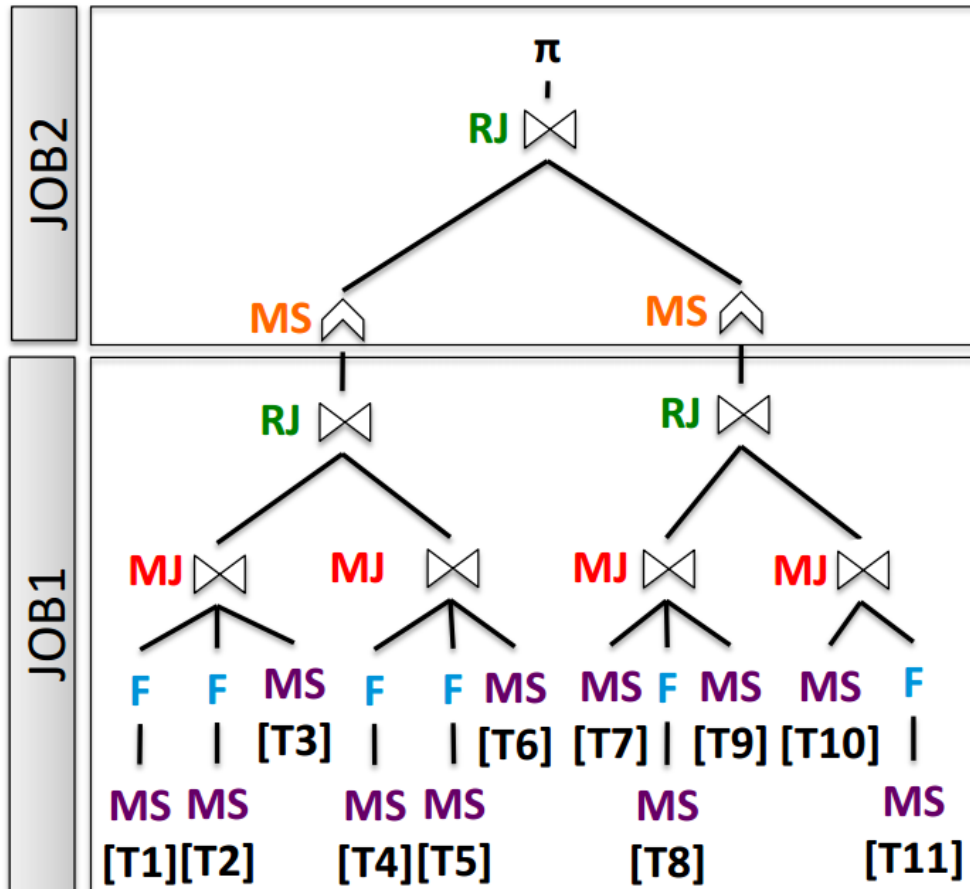
Cliqueに分解→還元

- 変数グラフからCliqueを構成→ノードとして還元
- 再帰的に繰り返してプランを作成



MapReduce実行

- 作成したプランにphysical operatorsを割当てて
- MapReduceのジョブとしてHadoop上で実行

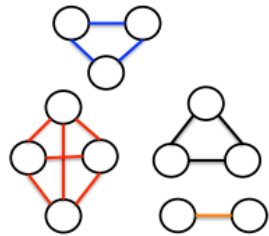


Clique Decompositionの方法の評価

● MSCが一番よかった

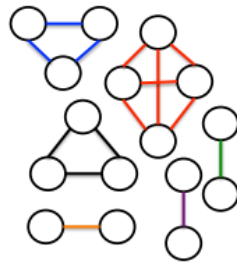
MSC+

- ◇ Maximal Cliques
- ◇ Minimum Set Cover
- ◇ HO-partial



SC+

- ◇ Maximal Cliques
- ◇ Set Cover
- ◇ HO-partial



MXC+

- ◇ Maximal Cliques
- ◇ Minimum Exact Cover
- ◇ HO-lossy

NO
DECOMPOSITION

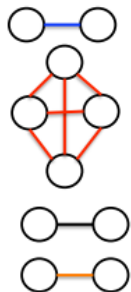
XC+

- ◇ Maximal Cliques
- ◇ Exact Cover
- ◇ HO-lossy

NO
DECOMPOSITION

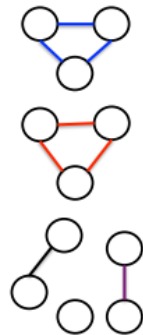
MXC

- ◇ Partial Cliques
- ◇ Minimum Exact Cover
- ◇ HO-lossy



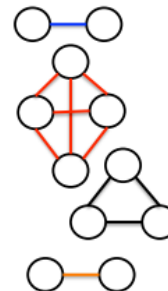
XC

- ◇ Partial Cliques
- ◇ Exact Cover
- ◇ HO-lossy



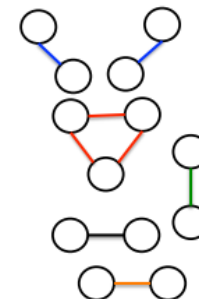
MSC

- ◇ Partial Cliques
- ◇ Minimum Set Cover
- ◇ HO-partial



SC

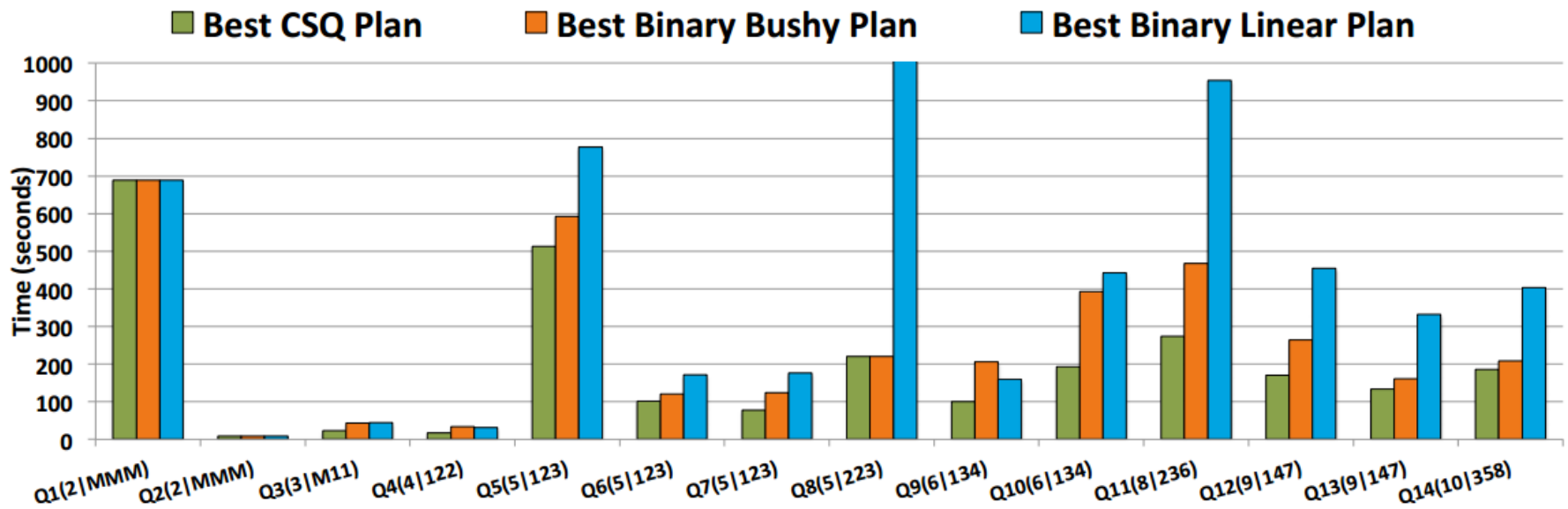
- ◇ Partial Cliques
- ◇ Set Cover
- ◇ HO-complete



BEST

Bushy planとLinear planとの比較

- 提案手法が良い
- LUBMベンチマーク(1G Triples/216GB)を使用



他のMapReduceを使った手法との比較

- 提案手法が良い(一部)
- LUBMベンチマーク(1G Triples/216GB)を使用

