Exercises for the book

Semantic Web Databases

by Sven Groppe

Chapter 8 – Parallel Databases

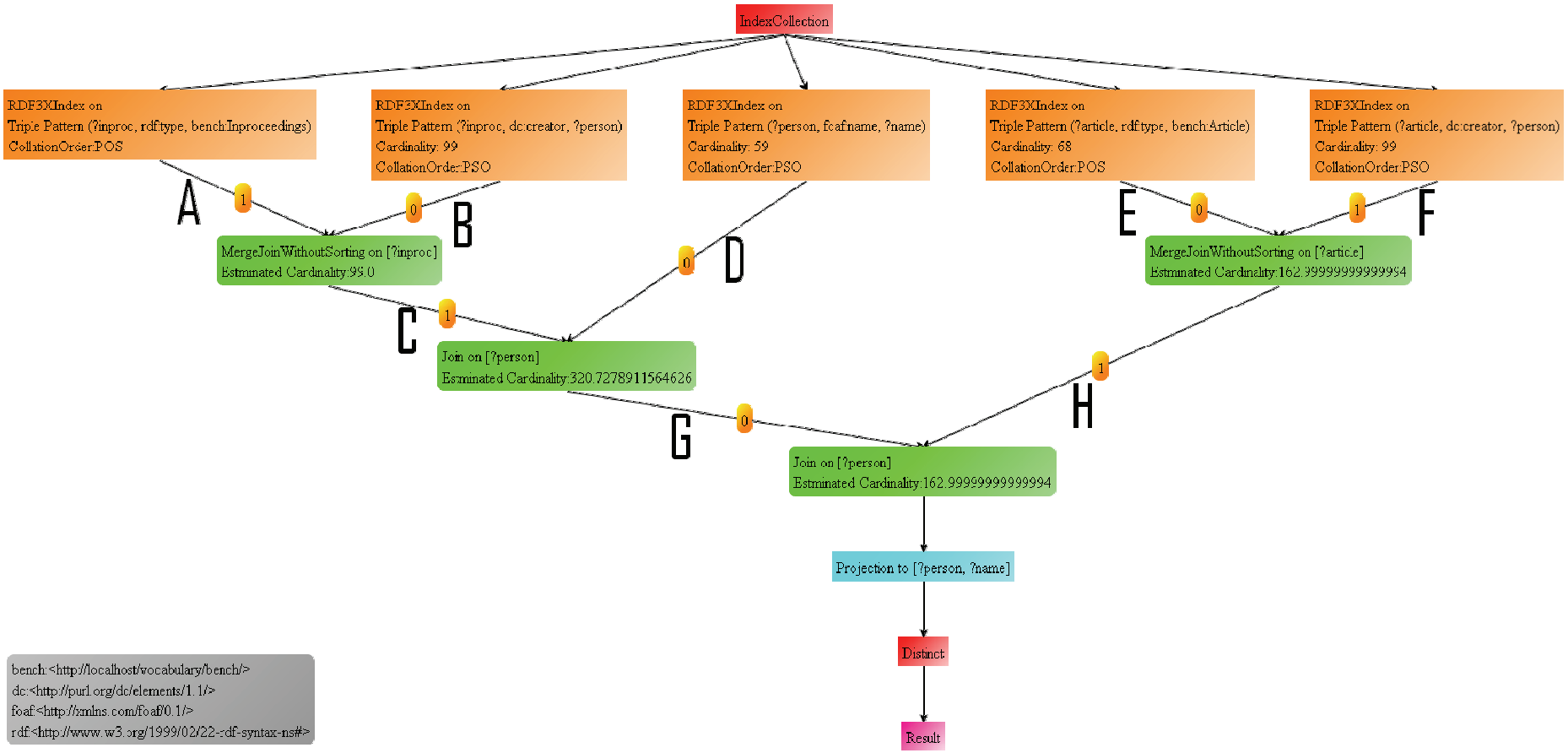
1. Let H be the finite set of hash functions with signature $U \rightarrow \{ 0, 1, \ldots, m\text{-}1 \}$. H is called universal, if for every pair of different keys $x, y \in U$ the number of hash functions $h \in H$ with $h(x) = h(y)$ is exactly $|H|/m$. In other words, the chance of a collision between x and y with $x \neq y$ is exactly $1/m$ for an arbitrarily chosen hash function $h \in H$, i.e., the chance is as big as the chance of a collision when using hash functions, which map all elements from U arbitrarily to $\{ 0, 1, \ldots, m\text{-}1 \}$. Note that for the last kind of hash functions a table of size $|U|$ would be necessary for the administration of the mapping from U to $\{ 0, 1, \ldots, m\text{-}1 \}$.

   The following class H_{prim} of hash functions is universal:

   Let m be a prime. We divide a key x in $r + 1$ subcomponents, such that $x = < x_0, x_1, \ldots, x_r >$. The only condition is, that the maximal value of a subcomponent must be smaller than m. Let $a = < a_0, a_1, \ldots, a_r >$ be a sequence of $r + 1$ elements, which are arbitrarily chosen from $\{ 0, 1, \ldots, m\text{-}1 \}$. Then we define the corresponding hash function $h_a \in H_{prim}$ by $h_a(x) = (\Sigma_{i=0}^{r} a_i * x_i) \bmod m$. Note that we can prove the universality of H_{prim} by using number theory.

   a) Calculate the hash values of 3, 20 and 82 using the hash function $h_{<5, 8, 9>}(x) = (\Sigma_{i=0}^{r} a_i * x_i) \bmod 13$. Let the range of a subcomponent x_i be $\{0, \ldots, 7\}$. The subcomponent can be calculated using $x_i = (x \operatorname{div} 8^i) \bmod 8$.
   b) Prove: If we require $a_i > 0$ for each $a_i \in \{ a_0, a_1, \ldots, a_r \}$, then H_{prim} is not universal any more.

2. Let us assume we want to use 3 threads for parallel operand result computation and the following operatorgraph. Which operand results (A to H) are optimal to be computed in parallel?


IndexCollection
RDF3XIndex on
Triple Pattern (?inproc, rdf:type, bench:Inproceedings)
CollationOrder:POS
A
1
RDF3XIndex on
Triple Pattern (?inproc, dc:creator, ?person)
Cardinality: 99
CollationOrder:PSO
0
B
RDF3XIndex on
Triple Pattern (?person, foaf:name, ?name)
Cardinality: 59
CollationOrder:PSO
0
D
RDF3XIndex on
Triple Pattern (?article, rdf:type, bench:Article)
Cardinality: 68
CollationOrder:POS
E
0
RDF3XIndex on
Triple Pattern (?article, dc:creator, ?person)
Cardinality: 99
CollationOrder:PSO
1
F
MergeJoinWithoutSorting on [?inproc]
Estmimated Cardinality:95.0
C
1
MergeJoinWithoutSorting on [?article]
Estmimated Cardinality:162.99999999999994
Join on [?person]
Estmimated Cardinality:320.7278911564626
G
0
1
H
Join on [?person]
Estmimated Cardinality:162.99999999999994
Projection to [?person, ?name]
Distinct
Result
bench:<http://localhost/vocabulary/bench/>
dc:<http://purl.org/dc/elements/1.1/>
foaf:<http://xmlns.com/foaf/0.1/>
rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>