University of Zurich[UZH]

Department of Informatics

University of Zürich
Department of Informatics
Binzmühlestr. 14
CH-8050 Zürich
Phone. +41 44 635 43 11
Fax +41 44 635 68 09
www.ifi.uzh.ch/dbtg

UZH, Dept. of Informatics, Binzmühlestr. 14, CH-8050 Zürich

Jerinas Gresch
Triemlistrasse 117
8047 Zürich
Switzerland

Prof. Dr. Michael Böhlen
Professor
Phone +41 44 635 43 33
Fax +41 44 635 68 09
boehlen@ifi.uzh.ch

Zürich, January 19, 2016

Bachelor Thesis in Informatik
Datenbanktechnologie

Topic: Temporal Primitives on Partitioned Relations

Overview:

Given an outer and an inner relation, many operators (such as joins, antijoins, and aggregations) can be computed. When a timestamp $T = [T_s, T_e)$ (that is part of the schema of the input relations) is involved in the operators, we refer to them as *Temporal Operators*. We are interested in computing temporal operators efficiently. Up to now, we have developed a partitioning technique called DIP (Disjoint Interval Partitioning), that partitions an input relation into a set of partitions, each composed by not overlapping tuples, i.e., by tuples whose interval does not intersect any other interval in the same partition. The goal of this project is to design and implement efficient algorithms applying the temporal operators to the partitions. The obtained results should then be used for computing the total result, i.e., the temporal operator between the initial relations.

Detailed description: A DIP-partition is a set of tuples whose interval does not overlap with any other tuple in the same partition. In the following picture, two input relations ($\mathbf{r}$ and $\mathbf{s}$) are split into two DIP-partitions each. In the context of this project, given an algorithm partitioning the input relations into the minimum number of DIP-partitions, the student is required to compute *efficiently* a temporal operator (such as a temporal join, antijoin, or aggregation) on the input relations, using the DIP-partitions.

$\mathbf{r}_1$	R	T
r_1	b	$[2,3)$
r_2	a	$[6,8)$
r_4	a	$[10,12)$

$\mathbf{r}_2$	R	T
r_3	b	$[7,11)$

$\mathbf{s}_1$	S	T
s_1	a	$[0,10)$

$\mathbf{s}_2$	S	T
s_2	b	$[1,2)$
s_3	c	$[3,4)$
s_4	a	$[5,6)$
s_5	d	$[8,11)$

Given two DIP-partitions, they offer the nice property that it is possible to find all the pairs with an overlapping interval, doing just one scan of the partitions. Using this property, many temporal operators can be efficiently computed. The following figure displays the result of a temporal join, anitjoin, and aggregation applied to our example relations.

$\mathbf{r} \bowtie_T \mathbf{s}$

R	S	T
b	a	$[2,3)$
a	a	$[6,8)$
a	a	$[6,8)$
a	a	$[7,9)$
a	a	$[7,9)$

$\mathbf{r} \triangleright_T \mathbf{s}$

R	T
a	$[11,12)$

${}_T\vartheta_{MAX(S)}\mathbf{s}$

S	T
a	$[0,1)$
b	$[1,2)$
a	$[2,3)$
c	$[3,4)$
a	$[4,5)$
a	$[5,6)$
a	$[6,8)$
d	$[8,11)$

Tasks:

- **Jan 16**: Read and understand in full detail the report on the Disjoint Interval Partitioning.

Given a memory-based algorithm computing the DIP-partitions of an input relation, implement:

- **Jan 16**: a disk-based algorithm for computing the same thing, and that places each DIP-partition in a different file.

Given the equivalence rules in the report and the partitioning algorithm, build a memory-based and a disk-based implementation for computing:

- **Feb 16**: A Temporal Join between two temporal relations

- **Feb - Mar 16**: A Temporal AntiJoin between two temporal relations

- **Mar 16**: A Temporal Aggregation on a temporal relation

- **Apr 16**: In your thesis, for each operator, define it and present your solution precisely; design a representative running example to illustrate the operator computation. Describe your algorithms using pseudo code.

Present progress and plans once every week to your supervisor.

All algorithms must be integrated in the same framework computing the partitioning.

All algorithms must output the runtime for computing the partitioning, the runtime for computing the operator, and the number of *unproductive comparisons* done (as defined in the given report).

Supervisor: Francesco Cafagna

Starting date: 13 January 2016

Ending date: 13 July 2016

Department of Informatics, University of Zurich

Prof. Dr. Michael Böhlen