Paper Title: Adaptive Flushing Strategy for Multi-State Memory-Intensive Long-Running Queries in a Streaming Environment.

Authors: Justin Levandoski (Group 3)  Mohamed Khalefa

Reviewer: Esten Rye (Group 1)

Date of Review: Sunday, November 12, 2006

Summary:

Focus:
This group clearly identifies the problem they are addressing in their paper. Their paper is on a novel memory flushing algorithm they developed. The goal of the algorithm is to be highly adaptive to the extreme variability of streaming inputs in order to make a globally optimal decision on which partitions to flush to disk. They assert that their algorithm is novel in the fact that it is the first algorithm that uses global output prediction to provide high and reliable throughput while considering multiple join operations. They show their novelty by first providing a brief but detailed summary of the previous works in the field and illustrate the weaknesses of the approaches proposed in those works when compared to their approach. They plan on supporting their claim by providing experimental evidence. The experiment they propose will measure the runtime throughput of their approach and the only other approach that considers a query plan with multiple join operations for performing a join of 4 relations. They factors that they will vary are the join ration and the input rates. The reason for varying these factors is to test the flushing strategy adaptation.

Technical Evaluation:
After reading the paper, I felt that the authors’ literature survey was thorough enough to provide them with a strong background in their problem. The authors presented enough information to the reader to understand and follow their arguments while avoiding from diluting their paper with details that were outside the scope of their approach. I did not take the time to read any of the references listed; however, I do feel the authors presented the previous works in the area well enough to justify their novelty claim. At this point in the paper I can neither agree with or deny the claims of the author as their experimental data has not been completed.

Readability and Organization:
The paper is well organized and easy to follow. It is just about the right length, any length that is needed is assumed to be allocated to the Experimental Data section. The paper is for the most part self-contained. The authors do refer the readers to external papers in Section 3 for the sake of brevity. In general the authors do a good job at deciding when to refer the reader outside their paper, but
I do have some suggested areas in their paper where they should elaborate more before doing so. These are listed in the Suggested Areas for Improvement. The grammar in the paper is for the most part correct, with the few exceptions listed in the Suggested Areas for Improvement below.

**Strengths**

- The paper flows well, making it easier to understand the approach and arguments for it.
- The Abstract adequately and concisely describes the content of the paper and the general reasoning of the approach presented.
- The authors give a concise and thorough evaluation of the previous approaches and give a good presentation of the advantages of their approach.

**Suggested Areas for Improvement**

- When referring to named figures, tables and equations (i.e. Figure 1, Table 3.1.2, and Equation 5) always use a capital letter on the words Figure, Table and Equation. Because they are named, they must be referred to as proper nouns.
- The figures at first glance seem to be low resolution scans from other papers. My suggestion is to scan them at higher resolutions if that is the case or redraw the figures at higher resolutions if it is not.
- Section 3 has some simple grammar mistakes like spelling errors and verb tense. A simple reading of the section on your own should ferret out these problems.
- In Section 3.1.4, the Tuple Input Prediction section mentions $\alpha$ as the decay constant when in Section 3.1.2 $\lambda$ is defined as the decay constant in the paragraph below Equation 2 and used throughout the rest of the paper.
- In Section 3.2, break up Equation 5 into smaller parts and explain those. I found your explanations in the previous sections to be most helpful in understanding the formulas and was slightly disappointed by the explanation of this equation.
- Under the Global Output section of Section 3.2 you refer to $GO_j$. What is it and how is it defined?
- Under the Output Statistics Tracking section of Section 3.2 you define Equation 6. I was very confused by this equation because I couldn’t figure out how the left side of the equation could ever equal the right side of the equation. I would suggest you review the explanation of this equation and consider explaining the derivation of this equation as well.
- Globally, replace “To the author’s knowledge” with “To our knowledge” when speaking about your knowledge. The former phrase signifies someone else’s knowledge while the latter is more assertive, signifies that it is your knowledge that is being quantified and follows the flow of your paper better.