The activities in this project have focused on four major areas: feedback punctuation, Frames (data-driven windows), prioritization and the Data Streams course.

In the area of feedback punctuation, both theoretical and implementation work has been completed. The theoretical work focuses on properties of stream queries and the usage of feedback punctuation to improve query processing while maintaining stream query correctness and completeness. Feedback punctuation has been implemented and tested in the NiagaraST stream system. Performance results are encourage and are part of Rafael Fernandez’s thesis and will be presented at his thesis defense. As part of that work, we have developed a notion of “contract” for stream operators. A contract indicates under which (forward) punctuation an operator can effectively operate, and what punctuation the operator can emit in that case. An operator can publish multiple contracts, and we have developed techniques to select a compatible set of contracts for all operators in a query, if such exist. We have also extended contracts to handle feedback punctuation. This work was presented at the Scalable Stream Processing Systems workshop 2010.

Graduate student James Whiteneck has led the Frames work. James took the Data Streams course (described below) and has continued his work on his project for that course as a research assistant on this project. He has worked on development of models and execution techniques for Frames. Frames are an alternative for windows in a stream processing system. Whereas windows are typically defined by a fixed duration or a fixed number of tuples, the span of a Frame is based on the data content of the stream. Examples are max speed – min speed > 5 mph on a traffic-sensor stream, and packet-loss rate > 0.3% for 5 minutes on a stream of router reports. In addition, an initial implementation and initial performance tests have been completed based on a scenario proposed by Dr. Ted Johnson of AT&T labs. A paper on Frames with spatial extensions was written and submitted to the ACM SIGSPATIAL International Workshop on GeoStream (IWGS) 2010.

There have been two aspects of prioritization in stream queries that we have investigated. One is on early production of partial results (possible in conjunction with later production of complete results). Amit Bhat, a Masters student, undertook this work, under the co-direction of Rafael Fernandez. We have introduced the idea of a “prod” that travels through a stream query, and triggers production of results relative to a time period. Mr. Bhat’s MS thesis showed that this technique can be effective, in some situations, for dealing with late data and bursts of data. Shreemoyee Sarkar, also a Masters student is developing a capability within NiagaraST to prioritize the results of aggregate operators by controlling the frequency of window production. Currently we have the ability to follow a static priority policy, but are currently working on a new flavor of feedback punctuation that would allow dynamic setting of priorities.

Dr. Tufte and Dr. Maier taught the Data Streams course in Winter Quarter 2010. The course was updated to include recent research and to incorporate Microsoft StreamInsight as an example system for the students to use in assignments and the course project. Rafael Fernandez assisted as a TA in this course, particularly leveraging his experience with and knowledge of Microsoft StreamInsigh; this experience and knowledge was gained at his summer 2009 internship at Microsoft. Many course projects were done in the data streams course based on research problems associated with this NSF project. Three of the students from the course have continued to work on this project. We intend to offer the course again in Winter Quarter 2012.

*Year 2:* In year two, work focused on advancement of frames and prioritization.

*Frames:* We are developing a theoretical and implementation framework for frames. James Whiteneck is handling the implementation and testing. Frames provide a more sophisticated mechanism than windows for segmenting streams. It seems intuitive that there are situations in which a static, fixed window specification (i.e. range 5 minutes, slide 1 minute) would provide a poorer approximation to the user’s desired result than a more specialized mechanism such as frames. However, demonstrating the power of frames with performance results has been more difficult than expected; we have explored many avenues for demonstrating the effectiveness of frames including combining frames, various data sets. Recently, we have found data and objectives that benefit significantly from frame constructs including traffic data and most recently ocean observing data. It appears that in cases where the data is highly variable and where there are significant periods of time where the data values are stables, frames can provide a better representation of the stream using fewer result elements than windows.

Prioritization: Prioritization work has continued. Prioritization implementation has leveraged the existing feedback framework.