C-MR: A Continuous MapReduce Implementation for Stream Processing

Nathan Backman, Karthik Pattabiraman, Uğur Çetintemel

C-MR Architecture

The Combine Phase

Deployed on multi-core architectures
• Intermediate disk buffers eliminated
• Core/Processors treated as generic computing nodes
• Nodes asynchronously receive/process data for any operation

The addition of a Combine phase allows for:
• Incremental processing
• Fewer redundant computations
• Shared streams w/ different window sizes

Window Materialization
• Inserted punctuations when the stream is sorted
• Replicate punctuations to parallel operators
• Gathered punctuations at downstream buffer prior to aggregation
• Materialize window on receipt of relevant punctuations

Buffer Timeline

Arrived:
Materialize: