Aurora Borealis

Borealis Stream Processing

“Distributed streaming processing engine with more flexible stream processing model” to meet the requirements of streaming applications.

Stream-Processing Engine

• Real-time processing of large volumes of streaming data.
  – Store then process → Process data before store
  – Relatively small set of operators (e.g. filter, aggregates, correlations) on windows of data that move with time

• Aurora Stream-Processing Engine (SPE)
  – Time-series ops. for streaming applications
  – Real-time latency (i.e. process before storing)
  – Gracefully deal with bursty message load
Aurora Stream Processing

Arc: tuple queue that represents stream

CP: buffer stream message history required by operators

Box: pre-defined operator

QoS: important metrics for resource management

Extended in Borealis

A: tuples are appended to streams only.
B: revision – tuples can be inserted, deleted, or replaced

A: on the outputs
B: any point

B: time travel – run queries in the past or future

B: control line - change box semantics on the fly

B: this giant network of operators are distributed over multiple sites

Cetintemel et al., The Aurora and Borealis Stream Processing Engines. Data Stream Management: Processing High-Speed Data Streams, Springer-Verlag (2006)
Borealis System Architecture

- Each site runs a Borealis server ...

Borealis Stream Processing

“Distributed streaming processing engine with more flexible stream processing model” to meet the requirements of streaming applications.

• Built on top of Aurora + Medusa
• Advanced features for flexible stream proc.
  – Dynamic revision of query/results
  – Dynamic operator network optimization