

Panel: Data Streams go Mainstream

Panel Organizer:

Mike Lefler

Northrop Grumman Corporation

Invited Panel Participants:

Johannes Gehrke
Cornell University
(chairman)

Mike Spicer
IBM

Jonathan Goldstein
Microsoft

Dieter Gawlick
Oracle

Abstract

Data Stream Processing (DSP) has moved rapidly from being the focus of an invited talk at PODS 2002 to start-up companies in 2005 and 2006. Today the giant RDBMS vendors such as IBM, Microsoft, and Oracle have added, or are seeking to add, DSP capabilities to their respective product lines. This panel will discuss the likely trajectory of Data Stream Processing as it moves forward as a component of the market-leading RDBMS product lines.

Categories and Subject Descriptors

H.3.3: Information Search and Retrieval – *information filtering, retrieval model*.

General Terms

Algorithms

Questions to explore among the panel:

1. First of all let's consider the name of this technology. Research projects such as STREAM (Stanford) and Aurora (MIT/Brandeis/Brown), among others, described this area of research using the term "continuous data streams," but today it is more typical to see the technology advertised under the rubric of "Complex Event Processing" instead of "Data Stream Processing." How does Complex Event Processing as it is coming to market today differ from Data Stream Processing as originally researched circa 2000-2002? Is this just a name change? Or does it reflect some deeper distinction

between the technology as originally explored by the database research community and the technology as brought to market? If the latter, can you summarize the changes?

2. To what degree is there any consensus on a query language for Data Stream Processing/ Complex Event Processing? And if there is one, is the consensus forming around StreamSQL or is it going in a different direction? Specifically, is there sufficient consensus to justify establishing a new project part for the ISO SQL standard?
3. Most of the early major research projects such as the two (STREAM and Aurora) mentioned earlier are completed. Does that mean that most of the important research is finished? What do you see as the areas that are still ripe for original research?
4. Does data stream processing really belong as part of a DBMS, or is the "outboard accelerator" approach taken by some competing products a reasonable one?
5. Obviously your respective corporations feel that there is a good market for Complex Event Processing or they would not be performing R&D in this area. How would you characterize the market for CEP? Niche but high margin? Large?
6. What does the future hold for the start-ups in this space, given your companies' domination of the database market?