System S

O. Verscheure

IBM T.J. Watson Research Center, Hawthorne, NY 11532; 914-784-7369, Fax: 914-784-6040,
   E-mail: chitrav@us.ibm.com

As the amount of data available to enterprises and other organizations dramatically increases, more and more organizations are looking to turn this data into actionable information and knowledge. IBM has developed the System S platform as a Research initiative, to address these requirements by enabling efficient extraction of knowledge and information from potentially enormous volumes and varieties of continuous data streams. System S is designed to scale from systems that acquire, analyze, interpret, and organize continuous streams on a single processing node, to high performance clusters of hundreds of processing nodes. System S provides an execution platform and services for user-developed stream processing applications. It supports the composition of new applications in the form of stream processing data flow graphs that can be created on the fly, mapped to a variety of hardware configurations, and adapted as requests come and go, and relative priorities shift. This allows for adaptive, hypothesis-based analysis of data, simultaneous evaluation of alternate analysis models, and discovery of new information and intelligence from data streams. Stream processing applications can be found in areas as diverse as radio-astronomy, manufacturing, and energy-trading. In this presentation, we will describe the System S programming paradigm, its capabilities, as well as a few applications that researchers at IBM T.J. Watson Research Center are engaged in.

The System S architecture represents a significant change in computing system organization and capability. Users compose stream-processing applications as a stream-processing dataflow graph, as shown in the figure. The System S runtime environment accepts these specifications, determines how it might reorganize itself in order to best meet the requirements of newly submitted and already executing specifications, and automatically effects the changes required. The runtime continually monitors and adapts to the state and utilization of its computing resources, as well as the information needs expressed by the users, and availability of data to meet those needs.
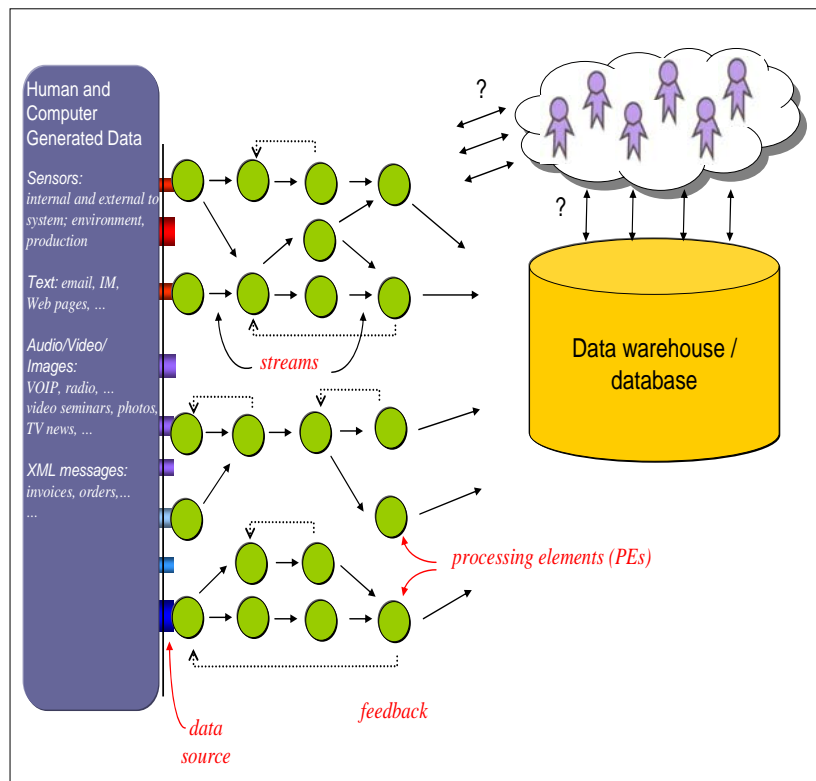

Figure 1. System S architecture.