## StreamPy:

## Python Event-Stream Processing for Real-Time Business Intelligence

### K. Mani Chandy, Julian Bunn, Rahul Bachal, Ker Lee Yap

### California Institute of Technology

StreamPy is an open-source platform for developing real-time business intelligence (RTBI) applications by processing event streams that are endless sequences of data generated by phones, sensors, social media, and monitoring systems. The platform leverages Python’s open-source libraries. StreamPy is being developed by a group of students and staff at Caltech.

We prefer to use the phrase “right time business intelligence” because what matters is getting information to the right places at the right time. . By “right time” we mean applications that typically respond in a few seconds or minutes.

Applications of event-stream processing [1, 2, 3] and real-time business intelligence (RTBI) have been described extensively in the literature [4, 5, 6] and include:.





Enterprises, especially medium and small ones, don’t use RTBI for three main reasons:

1. Many examples of RTBI, such as those in the previous slide, deal with operations in different lines of business rather than with corporate business intelligence. Some operating units in medium and small enterprises can benefit from RTBI but don’t have the budget or the time to acquire and integrate proprietary software into event-stream processing apps that generate RTBI.
2. Many enterprises carry out analytics in databases and the cloud. Event-stream processing is most efficient, however, when processing is distributed across phones and smart sensors, local servers, and the cloud. Business units don’t have tools that give them a holistic view of their entire distributed processing capability for event-stream processing.
3. IT staff have deep expertise in their businesses operations and they need tools that enable them to use their expertise to build event-stream processing apps.



Today, more than ever, the resources for building RTBI apps are free or inexpensive. Streaming data sources such as sensors and phones are inexpensive. Hardware platforms, including microcomputers such as the Raspberry Pi and Intel Galileo, connected to sensors, are inexpensive and getting more powerful. Many powerful Python libraries for analytics and science are freely available. StreamPy is a platform that integrates inexpensive streaming data sources, open-source software that operates on data at rest, and inexpensive hardware, to obtain RTBI apps that operate on data in motion.



StreamPy is a platform that addresses challenges faced by enterprises in building RTBI apps.

1. Operating units, especially in medium and small businesses, don’t have large data-analytics staff and budgets. StreamPy enables them to leverage extensive Python libraries to obtain RTBI by analyzing data streams.
2. Businesses often obtain intelligence by processing in the cloud and in corporate databases. RTBI processing is most efficient when it occurs at multiple stages from initial data acquisition to decision support. Processing in hundreds or thousands of small computers, such as the Raspberry Pi and Intel Galileo, connected to sensors, reduces the time and cost of computation. StreamPy helps provide a holistic view of stream processing across the entire distributed system.
3. Businesses can benefit from a rich and growing collection of free Python libraries that cover analytics, business, mathematics, science, engineering and displays. StreamPy enables IT staff to encapsulate Python programs to obtain modules that operate on data streams. They can then build complex RTBI apps by connecting modules together.