Abstract - Analysis of data, as it is being generated, enables timely response. Sensors, social media, phones and other devices generate streams of data. “Right time” analytics extracts the information from data streams that enables effective, timely response. This talk describes StreamPy, a tool for analyzing data in motion to obtain right-time analytics.

Many open-source and proprietary analytics packages exist. Most of these packages operate on “data at rest” as opposed to “data in motion” – they operate on arrays, lists, or tables. Though the contents of these data structures change with time, these packages analyze a single snapshot or a small set of snapshots. By contrast, analysis of data streams – or data in motion – treats every value of the stream as a different snapshot. StreamPy enables the use of data-at-rest packages for data-in-motion.

Some applications that monitor sensors require that data in streams be ordered by the time of generation. Generally, some degree of error in time ordering can be tolerated. The talk discusses methods of time stamping and dealing with timing error. StreamPy processes streams in parallel on multicore clusters, but maintains ordering in streams, by using concepts from networks of asynchronous processes.

This talk includes demos using StreamPy to obtain right-time analytics for different types of applications. This work was done with a team of undergraduates – Rahul Bachal, Ker Lee Yap, and Kalyn Chang – and Dr. Julian Bunn of the Center for Data Driven Discovery.

Brief Bio – K. Mani Chandy is the Simon Ramo Professor, Emeritus at the California Institute of Technology. He got his Bachelors in Electrical Engineering at the Indian Institute of Technology, Madras, in 1965; MS in Electrical Engineering at the Polytechnic Institute of New York in 1966; and a PhD in Operations Research at the Massachusetts Institute of Technology in 1969. He taught at the University of Texas at Austin, Computer Science Department, from 1969 to 1987, and at the California Institute of Technology from 1987 to 2014. He has written over 100 papers and 3 books on performance modeling and concurrent systems. He has received several awards including the IEEE Koji Kobayashi Award and the ACM Dijkstra Award. He is an IEEE Fellow and a member of the National Academy of Engineering.