COMPUTATIONAL RESEARCH
in BOSTON and BEYOND SEMINAR

AN OPEN DATACENTER DATASET FOR AI ENABLED OPTIMIZATION

SIDDHARTH SAMSI *and* VIJAY GADEPALLY
MIT-Lincoln Laboratory

ABSTRACT:

The first step in training an AI is to get the right data. In order to apply AI to the problem of data center optimization, such as identifying faults with servers, energy or cooling systems, before they become critical, the MIT Lincoln Laboratory Supercomputing Center is developing a state-of-the-art dataset. This dataset contains rich information such as: physical information about building management; system information such as scheduler and filesystem logs; and node-level information such as utilization, memory, GPU activity (both job level statistics as well as time-series monitoring collected via NVIDIA's DCGM tool), energy utilization, etc. In this talk, we will describe the dataset, detail how developers can get access to this data, and discuss a number of open problems associated with datacenter analytics.

FRIDAY, APRIL 2, 2021
12:00 PM – 1:00 PM

ZOOM MEETING info:

https://mit.zoom.us/j/96155042770

Meeting ID: 961 5504 2770

https://math.mit.edu/sites/crib/

MIT Massachusetts Institute of Technology