Abstract:
It’s 2023, we have no manned missions to Mars or flying cars. We have stadium-sized supercomputing facilities which can help us craft a witty memo when it’s not making up fictional court precedents. But analyzing large data sets seems to remain largely trapped behind serial processes after moving data over constrained, expensive networking links. Storage device vendors have talked about ‘computational storage’ for literally decades, we are thus exploring a new approach which we’d appreciate your perspective on. We will discuss the challenges which seemed to have trapped us in a “local maximum” for processing paradigms and explain how we have applied a divide and conquer approach to efficiently store data to enable processing - when called upon in the future. We will explain the initial applications for classic record-oriented data, how we have extended this to support multi-dimensional data sets, and our proposal to extend this to enable AI inference applications for scientific and video applications (proposed with UChicago). This work was supported with government support by the NSF and NOAA.