

# Prioritizing Computation and Analyst Resources in Large-scale Data Analytics

Host: Abhishek Bhattacharjee



Tuesday, March 16, 2021  
4:00 p.m.  
Speaker: Kexin Rong

Zoom Presentation

**Abstract:** Data volumes are growing exponentially, fueled by an increased number of automated processes such as sensors and devices. Meanwhile, the computational power available for processing this data – as well as analysts’ ability to interpret it – remain limited. As a result, database systems must evolve to address these new bottlenecks in analytics. In my work, I ask: how can we adapt classic ideas from database query processing to modern compute- and analyst-limited data analytics?

In this talk, I will discuss the potential for this kind of systems development through the lens of several practical systems I have developed. By drawing insights from database query optimization, such as pushing workload- and domain-specific filtering, aggregation, and sampling into core analytics workflows, we can dramatically improve the efficiency of analytics at scale. I will illustrate these ideas by focusing on two systems — one designed to optimize visualizations for streaming infrastructure and application telemetry and one designed for high-volume seismic waveform analysis — both of which have been field-tested at scale. I will also discuss lessons from production deployments at companies including Datadog, Microsoft, Google and Facebook.

**Bio:** Kexin Rong is a Ph.D. student in Computer Science at Stanford University, co-advised by Professor Peter Bailis and Professor Philip Levis. She designs and builds systems to enable data analytics at scale, supporting applications including scientific analysis, infrastructure monitoring, and analytical queries on big data clusters. Prior to Stanford, she received her bachelor’s degree in Computer Science from California Institute of Technology.