

Keynote

Silent Data Corruptions at Scale

Harish Dixit (Meta)



Harish Dixit is a Principal Engineer (Release to Production) at Meta. Harish and team work on reliability, analytics and performance evaluation for all of deployed fleet of servers. Harish leads the efforts to deal with silent data corruptions within Meta infrastructure across CPUs, GPUs and ASICs, and has been working across different layers of the stack to mitigate the effects of silent data corruption on production applications. Harish has over 20 patent filings across system architecture and communication domains. As part of this talk, Meta will be giving an overview of silent data corruptions, their prevalence in large scale infrastructure with a case study, and efforts to mitigate them at scale. These efforts are also published as “Silent Data Corruptions at Scale” and “Detecting silent data corruptions in the wild” for the curious reader..

Abstract: Silent data corruptions (SDC) in hardware impact computational integrity for large-scale applications. Sources of corruptions include datapath dependencies, temperature variance, and age among other silicon factors. These errors do not leave any record or trace in system logs. As a result, silent errors stay undetected within workloads, and can propagate across the stack to the applications. Silent errors can result in data loss and can require months of debug engineering time. In our large-scale infrastructure, we have run a vast library of silent error test scenarios across hundreds of thousands of machines in our fleet. This has resulted in hundreds of CPUs detected for these errors, showing that SDCs are a systemic issue across device generations. Based on this experience, we determine that reducing silent data corruption requires not only hardware resiliency and production detection mechanisms, but also robust fault-tolerant software architectures.

