

Tuesday, February 13, 2024

4:00 PM - 4:20 PM

Don't Let the IO Blender Destroy Your AI Model Training

Chris Zurich

Principal Systems Architect

WEKA Federal

According to one of the largest surveys of AI practitioners, data infrastructure is the most significant hurdle to AI success with 32% of respondents saying it is their top challenge. During AI and GenAI pipeline operations – including ingestion of data, pre-processing, embedding, retuning, and then extensive validation/backtesting – IO patterns are widely varied and may lead to an IO blender issue, a situation where different workloads consisting of transactional and streaming IO contend with a sub-optimal performance level. The IO blender impacts time-to-results in generative AI by reducing the utilization of the compute GPU and CPU. We will show actual IO blender effects on model performance and how you can avoid them and discuss the metadata management challenges at AI scale that can also affect model performance. Finally, we will share some practical tips for how to get maximum utilization for your existing AI infrastructure.