Switch - Total Memory Access Time (TMAT)

Figure: Relative TMAT versus Batch Size for ProMoE, ExpertFlow, and FATE, showing Cache Penalty and Dedup Penalty.