Uses Estimate probability of satisfying QoS if BufferSize is set to 6k? $P(Th > 40/s \mid do(Buffersize = 6k))$ Stage 1- Specify 3- Determine Performance Query **Next Configuration** Causal Inference Engine 5- Estimate 4- Update Causal **Causal Queries** Performance Model Decoder Muxer **Performance Tasks** Buffer Enable Batch **Bitrate** Size **Padding** Size What is the root-cause of fault? • How do I **fix** misconfiguration? • How do I optimize perf.? How do I understand perf.? Branch Cache No of Initial Misses Cycles Misses Software: DeepStream Perf. Data Middleware: TF, TensorRT Hardware: Nvidia Xavier 2- Learn Causal **Configuration**: Default Performance Model **Performance Fault/Issue** Energy Throughput)

Causal Performance Model

QoS : Th > 40/s

System Stack

Observed : $Th < 30/s \pm 5/s$