

Effect of KV Cache Performance on 4 HPU's  
for LLaMA-3-70B Model using Batch Size = 1

