

SUCCESS STORY

AI Hardware Accelerator Benchmarking

Analysis of Power-Performance-Accuracy (PPA Metrics)



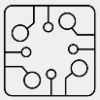


INTRODUCTION

A North America based AI Hardware (HW) accelerator chip company approached Ignitarium to perform power-performance-accuracy (PPA) benchmarking on new age AI accelerator devices to measure and compare PPA metrics.

Service in Focus

AI HW ACCELERATOR BENCHMARKING



Industry

Semiconductors → AI applications



Challenge

- Understand the quirks & nuances of the SDK for each of the HWs, perform model conversions and write inference scripts



Scope

- Find latency and throughput of state-of-the-art (SoTA) DL models using hardware accelerators
- Compare DL model accuracy with SoTA models from original framework
- Measure power consumption under DL load and verify the FPS/watt numbers published

THE CHALLENGE

While performing the task, Ignitarium was required to address the following challenges:

- The team had to ramp up quickly on the technical expertise to use the newly launched software SDK and APIs
- Bring-up various AI accelerator carrier boards in a timely manner
- Ramping up quickly on various aspects of SDK to ensure model conversions and inference scripts are done accurately

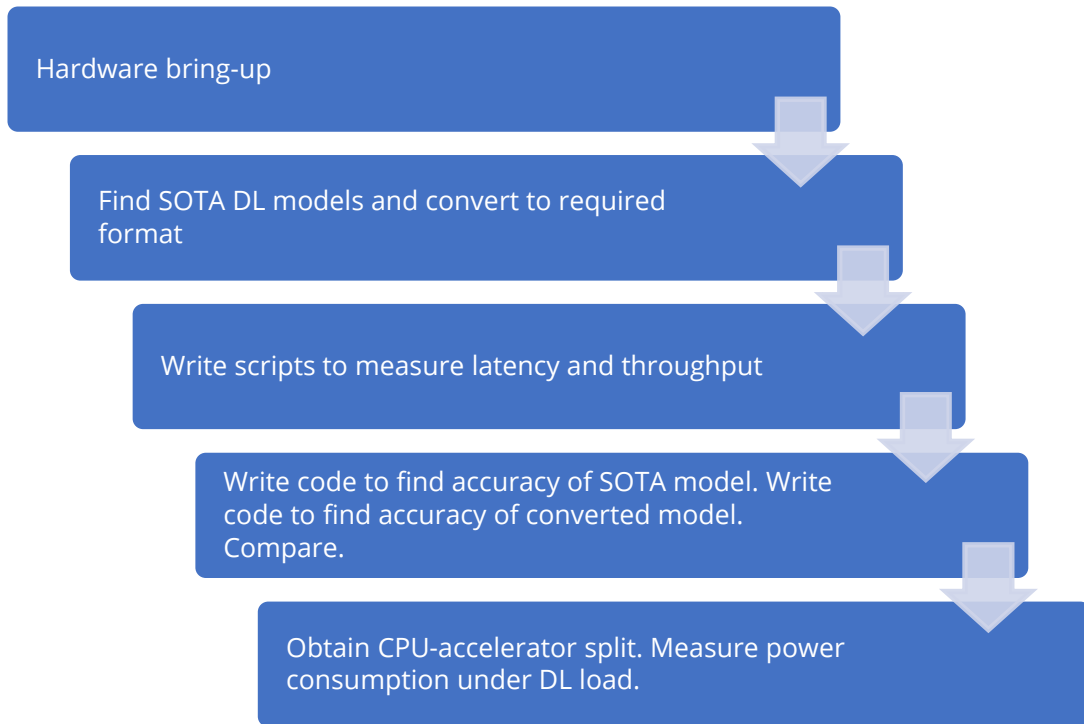




IGNITARIUM'S APPROACH

Ignitarium's approach consisted of following steps:

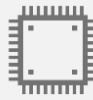
- Once the reference application code ran on the device, inference scripts were written for the standard SoTA models
- To maintain the transparency, SOTA models were converted to proprietary format and latency throughput was calculated
- Accuracy, host-accelerator computation split and power requirements under DL load were calculated



BUSINESS IMPACT



Detailed analysis report helped the customer leverage the PPA data to position their AI HW appropriately with potential customers



Customer's Product Marketing Team got relevant inputs for planning the specifications of their next generation AI HW chip



By leveraging our vast experience in benchmarking of AI HW accelerators, customer was able to get the analysis done well within the time and cost budgeted

Looking to transform your product engineering?

Drop us a line to get in touch with our experts.