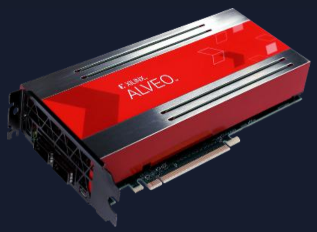
Wide & Deep
CTR Model for Recommender System

雪湖科技 snowlake-tech.com XILINX

SOLUTION BRIEF

XILINX ALVEO™



- 13Bucketized key+ 24Indicator + 6Embedding (60K Look Up Table)
- 14.7k QPS
- 1.8ms Latency

INTRODUCTION

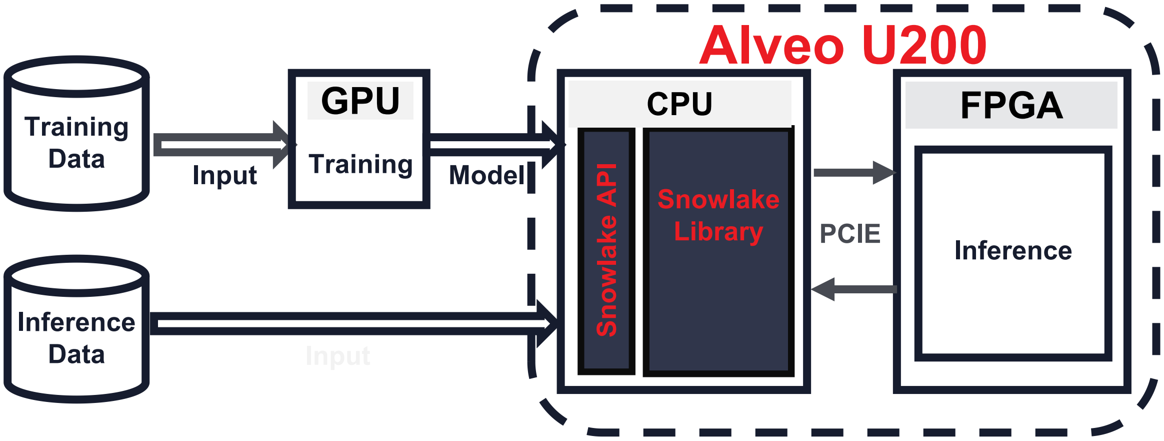
The product is designed to improve the performance of Wide & Deep model of Recommender System by using FPGA. It can achieve the lower latency and higher throughput than CPU to meet the requirement of rapidly increasing data to be dealt with.

KEY BENEFITS

- 2.4x higher QPS
- 5.7x lower latency
- 82% lower TCO

SOLUTION OVERVIEW

Explanation and/or image to show what is important about the solution. If there is an image, put it here below. Imagery on page1 will help entice readers to learn more and turn the page.



Wide & Deep
CTR Model for Recommender System

雪湖科技 snowlake-tech.com

SOLUTION DETAILS

Google published its implementation of wide & deep learning on tensorflow in 2016 and widely used by internet companies to significantly increases CTR (click-through rate) of advertisement.

The product is designed to improve the performance of Wide & Deep model of Recommender System by using FPGA. It can achieve the lower latency and higher throughput than CPU to meet the requirement of rapidly increasing data to be dealt with.

According to operators of model in tensorflow, we utilize distributed pipeline calculation to implement this system on Alveo U200. With these optimizations, the inference performance can be dramatically improved.

In addition, a tool kit is developed for model deployment. It can be easily applied to renew model and deploy the system. At the same time, our system supports various models based on Google open source to guarantee system compatibility.

RESULTS

Compared with CPU, the implementation of system can achieve the higher throughput and lower latency. Comprehensively, it can achieve more than 5 times TCO (total cost of ownership).

TAKE THE NEXT STEP

Learn more about Xilinx Alveo accelerator cards
Learn more about Snowlake-Tech http://www.snowlake-tech.com/
Reach out to Snowlake-Tech sales – (yinqingyu@snowlake-tech.com)