

COMPUTER ARCHITECTURE OF WAREHOUSE SCALE COMPTS

the total storage capacity

DOMAIN SPECIFIC ARCHITECTURE

O "tensiflow a system for large scale machine arch."

Useni'r nov 2016.

@ norman joppy, hishant patil, davaid patterson.
motivation for evalution of ist tensor processing unit.

Gielly, diffyon.

a golden age; empowering the ML revolution.

any data center perf. analysis of 9 TPG.

6 norman p joupy.

@ a domain specific arch, for ML apps.

4 DSAS discussed in book

- TPU - cotapult - - Crust by intel.

- pixel visual core Intel. 7.4--7.7.

GUIDELINE

over which data is moved.

data movement is managed by sw.

- invest in resources

domain.

- reduce the data size and type to simplest

needed for the domain.

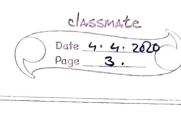
ruse domain specific prog. long. to port the

read host memory -> unified buffer.

reads weights from weights memory into weight fifo for matrix unit.

- matrix multiply + stroke convolve multiply. dot. convolution.

5 main instructions. TPU



+ activate non-linear functions for NN, relu, sigmoid, tanh. write host memory TPU, 28 nm, 700 MHZ, 113 rd space = unified buffies. 114 Space = matrix multiply unit. rogisters