

стек технических средств для реализации методов ML

Евгений Борисов

Технические средства ML

общее описание стека технологий

прикладные программные средства

вычислительные библиотеки

программный интерфейс с аппаратурой

аппаратные вычислительные средства

Технические средства ML

аппаратная часть : про модельные данные

набор с изображениями цифр MNIST - 70 000 примеров - 53MB

набор картинок CIFAR-10 - 60 000 примеров - 163MB

набор изображений PASCAL-2011 - 11 530 - 2GB

набор изображений ILSVRC-2015: CLS-LOC - 100 000 - 155GB

необходимо использовать средства соответствующей производительности

- высокопроизводительные параллельные вычисления

Технические средства ML

аппаратная часть :
параллельные вычисления

классификация по двум параметрам

- способ организации памяти
- способ организации управления

Технические средства ML

аппаратная часть :

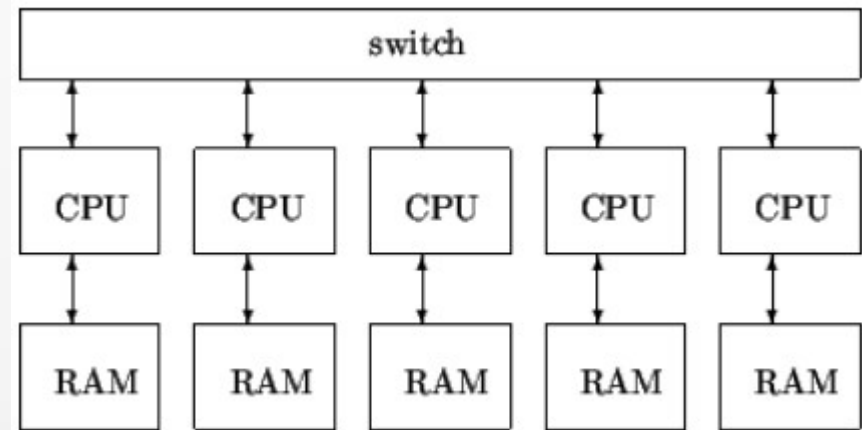
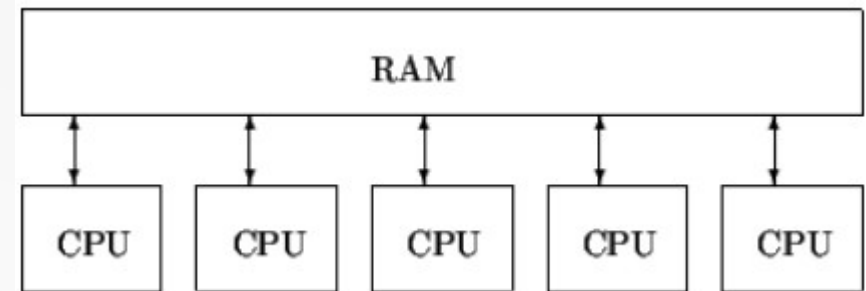
основные типы организации памяти параллельных вычислителей

общая память - все процессора работают в едином адресном пространстве с равноправным доступом к памяти

- + относительно просто программировать,
- ограниченная масштабируемость

распределенная память - каждый процессор имеет собственную локальную память, прямой доступ к этой памяти других процессоров невозможен.

- + хорошая масштабируемость
- сложно программировать
(больше вычислений, меньше обменов)



Технические средства ML

аппаратная часть :

способ организации управления параллельными вычислителями

классификация Флинна

SIMD "single instruction multiple data" (один поток команд много потоков данных)

SISD

MISD

MIMD "multiple instruction multiple data" (много потоков команд много потоков данных)



Технические средства ML

аппаратная часть :

задачи при организации параллельных вычислений

автоматическое распараллеливание последовательных программ

автоматическая балансировка загрузки процессоров

Технические средства ML

аппаратная часть :

основные типы параллельных вычислителей

SMP - симметричные мультипроцессорные системы (общая память)

PVP - параллельные векторные системы (специальные векторно-конвейерные процессоры)

MPP - системы массового параллелизма. несколько однородных вычислительных узлов; связаны специальной высокоскоростной сетью. (распределенная память)

NUMA - системы с неоднородным доступом к памяти. память физически распределена, но логически общедоступна.

Cluster - «упрощённый» вариант MPP

Grid - «кластер из кластеров»

GPGPU - General-Purpose Graphics Processing Units

Технические средства ML

аппаратная часть

HPC технологии можно комбинировать

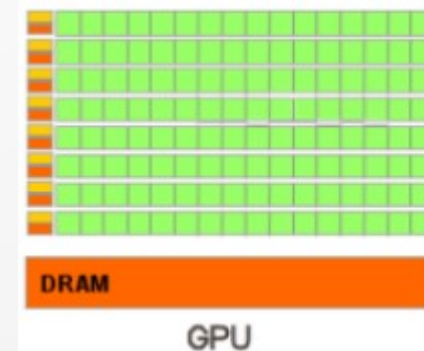
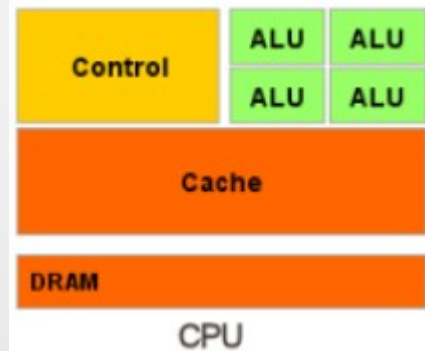
Пример: Грид состоящий из нескольких кластеров,
кластеры состоят из вычислительных узлов SMP
вычислительные узлы имеют GPU

Технические средства ML

аппаратная часть : подробней про GPGPU

параллельная программа GPGPU состоит из двух частей

1. основная - выполняется на CPU (host)
2. kernel - выполняется на GPU (device)



Технические средства ML

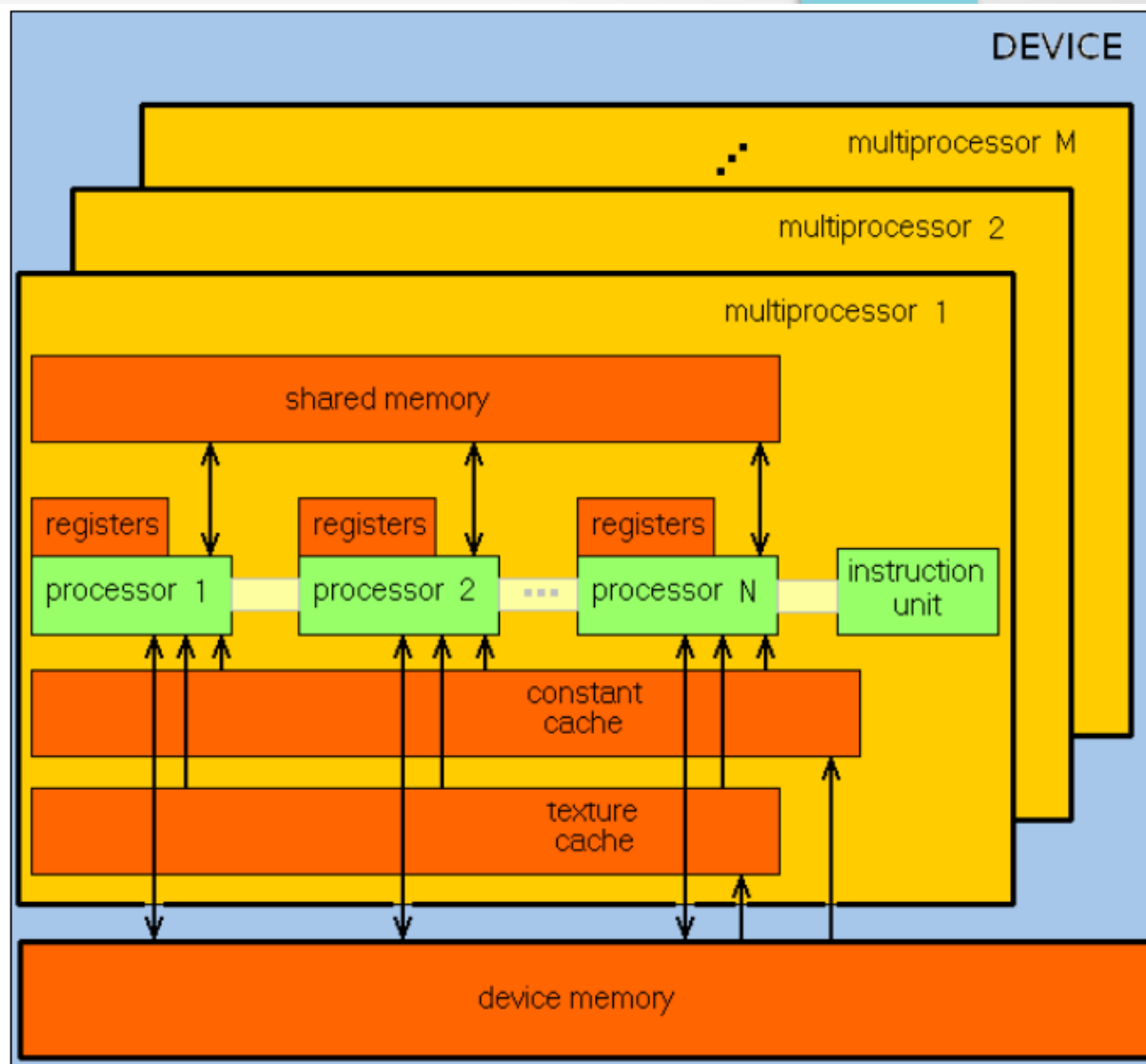
аппаратная часть : подробней про GPGPU

GPU - массив потоковых процессоров (Streaming Processor Array), состоит из Texture Processor Clusters (TPC)

TPC состоит из Streaming Multi-processor (SM)

SM содержит несколько Streaming Processors (SP) или ядер.

ядра мультипроцессора работают по схеме SIMD



Технические средства ML

программный интерфейс с аппаратурой

распределенная память: MPI, Spark

общая память: OpenMP, OpenCL

GPU: OpenCL, CUDA, DirectML

прикладные программные средства

вычислительные библиотеки

программный интерфейс с аппаратурой

аппаратные вычислительные средства

Технические средства ML

программный интерфейс с аппаратурой :

вычислительные библиотеки для систем с распределенной памятью

MPI - стандарт параллельного программирования для распределённых систем

на каждом узле кластера запускаем вычислительный процесс

процессы могут обмениваться данными

есть механизм синхронизации процессов

Apache Spark - средство для обработки данных в распределённых хранилищах

Технические средства ML

программный интерфейс с аппаратурой :

вычислительные библиотеки для систем с общей памятью

OpenMP - стандарт параллельного программирования для SMP

OpenCL - стандарт параллельного программирования для SMP и GPU

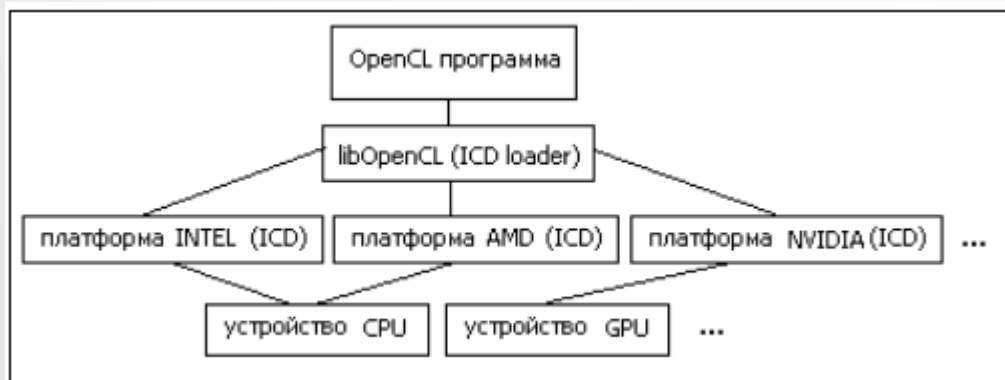
Технические средства ML

программный интерфейс с аппаратурой :

вычислительные библиотеки для GPGPU

CUDA - библиотека параллельного программирования для устройств nVidia
compute capability 7.5, cuBLAS, cuDNN

OpenCL - стандарт параллельного программирования для SMP и GPU



DirectML - API MS Windows для реализации вычислений на GPU и не только

Технические средства ML

прикладные средства

TensorFlow - пакет ML от Google

CNTK - пакет ML от Microsoft

Keras - библиотека-надстройка над TensorFlow, CNTK

PyTorch - пакет ML от Facebook

Технические средства ML

TensorFlow - пакет ML от Google

пакет предназначенный для символьных вычислений

может выполнять аналитическое дифференцирование и оптимизацию описанных пользователем функций

позволяет компилировать и выполнять заданные пользователем функций в машинный код для CPU и/или GPU

содержит TensorBoard - инструмент для визуализации графа вычислений

Технические средства ML

TensorFlow - пакет ML от Google

пакет предназначенный для символьных вычислений

может выполнять аналитическое дифференцирование и оптимизацию описанных пользователем функций

позволяет компилировать и выполнять заданные пользователем функций в машинный код для CPU и/или GPU

содержит TensorBoard - инструмент для визуализации графа вычислений

Keras - библиотека-надстройка над Theano, TensorFlow, CNTK

содержит примитивы для реализации моделей Deep Learning

использует TensorFlow или CNTK в качестве основы (backend)

Технические средства ML

TensorFlow - пакет ML от Google

пакет предназначенный для символьных вычислений

может выполнять аналитическое дифференцирование и оптимизацию описанных пользователем функций

позволяет компилировать и выполнять заданные пользователем функций в машинный код для CPU и/или GPU

содержит TensorBoard - инструмент для визуализации графа вычислений

Keras - библиотека-надстройка над Theano, TensorFlow, CNTK

содержит примитивы для реализации моделей Deep Learning

использует TensorFlow или CNTK в качестве основы (backend)

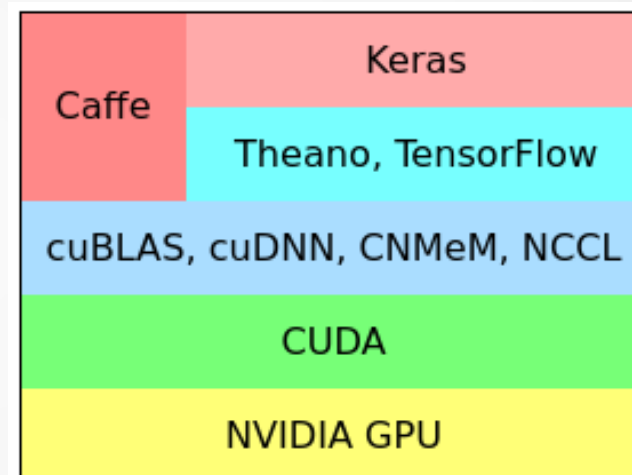
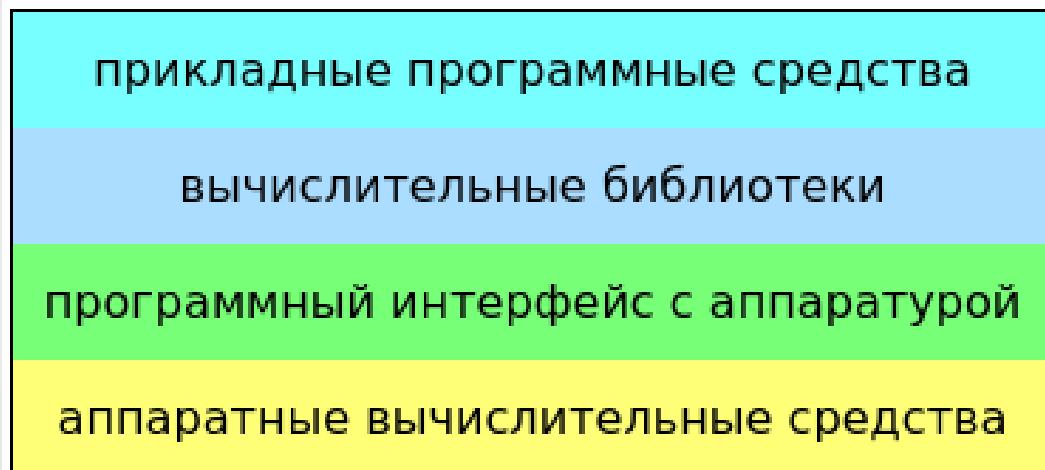
PyTorch - пакет ML от Facebook

пакет аналогичный TensorFlow/Keras предназначенный для символьных вычислений

содержит примитивы для реализации моделей Deep Learning

Технические средства ML

описание стека технологий



Технические средства ML: литература

Борисов Е.С. Методы машинного обучения. 2024

https://github.com/mechanoid5/ml_lectorium_2024_I

Евгений Борисов О технических средствах для реализации методов Deep Learning.

<http://mechanoid.su/ml-deep-learning-tools.html>

Евгений Борисов Технология параллельного программирования CUDA.

<http://mechanoid.su/parallel-cuda.html>

Евгений Борисов Технология параллельного программирования OpenCL.

<http://mechanoid.su/parallel-opencl.html>