Data Science Infrastructures

Assignment I – Computer Architecture

a. Research on the characteristics of CPUs, GPUs, and TPUs. What purpose have they been designed for? Which data science problems benefit most?

Central Processing Units (CPUs):

Characteristics:

- General-purpose computing: CPUs can be widely used for servers, PCs, laptops, and for higher ranges of tasks including arithmetic, logic, and control functions.
- Fewer cores: Although modern computers have multiple cores, CPUs generally have fewer cores compared to GPUs and TPUs.
- Strong single-threaded performance: CPUs excel at executing sequential instructions efficiently.

Designed Purpose:
CPUs are designed for general-purpose computing tasks, such as running operating systems, applications, and background processes.

Data Science Problems Benefitting Most:

- Basic data preprocessing: Tasks like data parsing and simple data manipulation can be performed by CPUs.
- Sequential algorithms: Problems involving sequential processing such as statistical analysis or linear regression, can benefit from CPUs' strong single-threaded performance.

Graphics Processing Units (GPUs):

Characteristics:

- Optimized for parallel processing: GPUs excel at performing many calculations simultaneously, which makes them highly efficient for parallelizable tasks.
- Thousands of smaller cores: GPUs are equipped with numerous smaller processing cores for massive parallelism.
- Better Efficiency: GPUs have higher performance compared to CPUs. For instance, NVIDIA GTX 10170 has theoretical performance between 100 GFLOPs & 6.5 TFLOPS depending on precision.

Designed Purpose:
It was initially designed for rendering graphics. However, nowadays is specifically built for Data Science. It is used as a parallel processor capable of handling complex computational tasks.

Data Science Problems Benefitting Most:

- Deep Learning: GPUs are exceptionally well-suited for training deep learning models due to their ability to perform parallel matrix operations efficiently.

- Images and video processing: GPUs are efficient and are highly used for image recognition, object detection, and video analysis as they benefit greatly from the parallel processing capabilities of GPUs.

Tensor Processing Units (TPUs):

Characteristics

- Optimized for tensor operations: TPUs are specifically designed to accelerate tensor calculations used for machine learning algorithms.
- Specialized for deep learning: TPUs are used to handle unique computational requirements of deep learning models.

Designed Purpose:

TPUs are highly specialized hardware that are developed to accelerate machine learning workloads, particularly deep learning models, for tensor operations.

Data Science Problems Benefiting Most:

- Large-scale deep learning: TPUs excel in training large-scale deep learning models, offering significant speedup compared to CPUs and GPUs.
- Tensor-intensive computations: Tasks that involve heavy use of tensor operations, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), benefit greatly from TPUs' optimized architecture.

In summary, CPUs are versatile and used for a wide range of tasks, GPUs excel in parallel processing and are ideal for deep learning and image/video processing, while TPUs are specialized hardware designed for accelerating deep learning models and tensor-intensive computations.

b. Choose a current model each (one CPU, one GPU, and one TPU) and compare

- **Performance**
- **Power consumption**
- **Cost**

CPU: Intel Core i9-12900K

Performance:

- Multi-core: 16 Cores
- Excellent single-threaded performance

Power Consumption:

- Thermal Design Power (TDP): 125W
  It represents the maximum amount of heat the CPU is expected to produce under normal operating conditions.

Costs:

- Approximately $500 - $600 (depending on the retailer and the region)

GPU: NVIDIA GeForce RTX 3090

Performance:

- CUDA (Compute Unified Device Architecture) Cores: 10496
- Memory: 24 GB GDDR6X

Power Consumption:

- TDP: 350W

Cost:

- Around $1500 - $2000 (depending upon the manufacturer and variant).

TPU: Google Cloud TPU v4

Performance:

- Matrix Multiply Units: 2048 (specific architecture details may vary)

Power Consumption:

- Not directly comparable due to its cloud-based nature.

Cost:

• Approximately $32 per TPU per hour (Google Cloud pricing model).

Comparison Summary:

- Performance: The GPU (NVIDIA GeForce RTX 3090) typically outperforms both the CPU (Intel Core i9-12900K) and TPU (Google Cloud TPU v4) in tasks that require parallel processing, such as deep learning model training.
- Power Consumption: The CPU (Intel Core i9-12900K) has the lowest power consumption (TDP) among three, followed by the TPU (Google Cloud TPU v4), while the GPU (NVIDIA GeForce RTX 3090) consumes the most power.
- Cost: The CPU (Intel Core i9-12900K) generally has the lowest upfront cost, followed by GPU (NVIDIA GeForce RTX 3090), while the TPU (Google Cloud TPU v4) tends to have a higher ongoing costs due to its cloud-based pricing model.

References:

1. Chat gpt -3
2. Intel Core i9-12900K (https://www.intel.com/content/www/us/en/products/sku/134599/intel-core-i912900k-processor-30m-cache-up-to-5-20-ghz/specifications.html)
3. NVIDIA GeForce RTX 3090 (https://www.nvidia.com/de-de/geforce/graphics-cards/30-series/rtx-3090-3090ti/)
4. Google Cloud TPU v4 (https://cloud.google.com/tpu/pricing#regional-pricing)

Assignment III – Cost of Cloud Computing

For a data science project, you need the following IaaS resources:

- **Virtual Machine with at least 4vCPUs & 8 GB RAM (no special instance needed)**
- **Object Storage up to 100 TB of data (standard)**
- **Hosted in Europe**

a. How much will this setting cost you per year? Choose one of the providers presented in the lecture.

Let us calculate the cost per year for the specified IaaS sources hosted in Europe. We will use the AWS (Amazon Web Services) as the service provider for this calculation. Let's select the t3.xlarge instance type, which offers 4 vCPUs and 16 GB RAM.

- Price/hour = $0.3328 (as per AWS pricing).
  Therefore, price/year = 0.3328 * 24 * 30 * 12 = $2875.392
- Let us use Amazon S3 (Simple Storage Service) for object storage.
  For the first 50 TB/month, the price is $0.0245 per GB
  For the next 50 TB/month, the price is $0.0235 per GB
  Therefore, the storage price per year = ((0.0245 * 1000 * 50) + (0.0235 * 1000 * 50)) * 12 = $ 28,800

- Total cost per per year = $2875.392 + $28800 = $31,675.392

b. You put your project in operation and monitor a mean value of 50TB/month data transfer (outgoing). How much will that cost you per month?

Again, we will use AWS to calculate the charges for data transfer of 50TB/month.

- First 10 TB/month = $0.09 per GB
- First 40 TB/month = $0.085 per GB
- Total cost/month = (0.09 * 1000 * 10) + (0.085 * 1000 * 40) = $4,300