



# AI HARDWARE INFRASTRUCTURE REPORT UAE

**NOV 2020** 



www.ai.gov.ae

#### **Table of Contents**

- 04 Forward
- 05 Executive Summary
- 06 Brief and Methodology
- 07 Introduction
- 09 CPU vs GPU for AI workloads
- 10 Existing AI Hardware infrastructure in the UAE
- 11 Distribution of computation resources in the UAE

3

- 12 International Position
- 13 Key Participants
- 14 Closure
- 15 Other Publications

#### **FORWARD**



**H. E. Omar Sultan Al Olama** Minister of State for Artificial Intelligence, Digital Economy and Remote Work Applications

The United Arab Emirates (UAE) government is in a phase of rapid growth in terms of the adoption and development of AI services. Government entities in the UAE are utilizing and implementing AI systems with one goal in mind; which is to continuously better serve individuals. Moreover, local universities, research institutes and the private sector are producing more work in AI than ever before, supported by a fast-growing pool of AI talent and experts in the UAE.

In addition to the availability of talent, another major key enabler of the development and deployment of AI systems is the computational hardware infrastructure also known as supercomputers. The trend to lead in this domain is heading towards more complex algorithms and larger datasets which require specialized hardware to process. Breakthroughs in AI in the UAE require that top talent is equipped powerful supercomputers with that foster growth and innovation,

feeding into the UAE's AI strategy vision of being a leading nation in AI by 2031.

As part of the UAE's commitment to remove barriers that challenge the advancement of AI in the country, the government of the UAE strongly believes in enabling its growing ecosystem with the tools necessary to excel and lead in the field. In line with that commitment, it pleases me to launch the first edition (2020) of the AI Hardware Infrastructure Report UAE series. In this edition, we highlighted the current state of the computational hardware infrastructure that exists in the UAE by surveying over 40 leading entities in the field of AI in the UAE and understanding the computational hardware they possess and the challenges these entities tackle using it.

We hope that this report sheds some light on the maturity level of the hardware available for AI in the UAE.

#### **EXECUTIVE SUMMARY**

The branches of AI including machine learning, computer vision and NLP have seen unprecedented growth in the past 10 years. The rapid development of these technologies was a result of the growth in computation power, memory, and their affordability. In addition, there is a unanimous agreement from governments, industries and academia across the world on the fact that developing AI holds significant benefits to society and the economy. The market size for AI is rapidly expanding and is expected to grow even faster. In anticipation for that growth, the UAE government took the initiative by appointing the first ever minister of Artificial Intelligence in October of 2017. This was followed by releasing the UAE Al strategy which outlines a vision for the UAE to be one of the leading nations in AI by the year 2031.

With such an ambitious vision comes a number of initiatives designed to accelerate up skilling talent, developing expertise, setting up a governance framework for AI and finally building up an infrastructure capable of supporting the growing demand for computational power for both developing and deploying AI solutions in the UAE. Being a leader in AI requires pushing the envelope with regards to research. However, the algorithms and data are growing in complexity which requires more and more computational power than ever. For example, an autonomous vehicle can generate on average over a petabyte of data per month if operated for only 4 hours a day. Storing, processing and learning from this data is not possible on consumer grade hardware. In fact, working on such applications require specialized hardware generally referred to as High-Performance Computing platforms (HPCs) or also known as Super Computers, which are specialized machines that are an aggregate of many highperformance processing, memory units and storage units.

Given that computational power is a key enabler of AI, the UAE has developed this report as an initial step to fully understand and compile a comprehensive understanding of the existing hardware infrastructure used for AI applications in the UAE. Thus, this report is set to define a baseline for the current state of the hardware infrastructure for AI in the UAE, then compare it to other leading nations in the field. The report will also highlight the main characteristics of a typical AI high performance computing cluster.

# **BRIEF AND METHODOLOGY**

This report is the initial step to understand the current state of the AI hardware infrastructure in the UAE. The data presented in this report is a result of a survey that was carried out with over 40 government, private and academic entities as well as meetings with key players to discuss the details of their infrastructure. The survey covered the following:

- Hardware specifications of the platforms used for Al applications
  - CPU count and specifications
  - GPU count and specifications
  - Total system memory

6

- > The **benchmark** used to measure the performance
- > Whether it is physically located in the UAE
- > How the entity utilizes these supercomputers
  - The problems these systems solve
  - Whether it is available for the general community to use or not

Due to the difficulty of standardising the performance metrics of different hardware specifications, benchmarks and the data reported by these entities, the report normalised and approximated certain performance metrics based on publicly available information and reports (We plan to standardise the main performance metrics for the next iteration of the report to be released in April 2021).



Figure 1: Summit Supercomputer The second most powerful system in the world -USA-



Figure 2: Tianhe-2 The sixth most powerful system in the world -China-

**CPU**: Central Processing Unit **GPU**: Graphics processing unit

**Benchmark**: A software or set of tasks that is run by computers to evaluate their performance.

### INTRODUCTION

There are a number of key enablers to build a successful and efficient ecosystem to support innovation and research in AI. Some of the key enablers come in the form of the infrastructure to support entities developing, deploying and researching AI. The main pillars of the AI infrastructure can be categorized into three main items

- > **Data infrastructure**: Generally, refers to the availability of data and the high-performance storage platforms that synergizes with AI applications.
- > **Network**: Specialized high-performance networking systems that connect the compute servers amongst themselves and to storage units.
- Hardware infrastructure: The computing platforms and computer chips that accelerate the process of training and deploying AI applications as well as supporting a large amount of memory.

The trend in AI development is leaning towards more complex and larger models that require large computational capabilities. In fact, it is becoming more difficult to train and deploy AI models on home desktops. Thus, this report focuses on the compute infrastructure as a key enabler for pushing the boundaries of AI.

The process of training AI algorithms is a very complex task that requires a significant amount of domain expertise. However, in its simplest form,



Figure 3: Al Infrastructure Reference Architecture

HPC: High-Performance Computer

training a machine learning model is simply the process of feeding a large amount of data to a set of algorithms. A set of simple computations are repeated numerous times where in every cycle of this process, the algorithm slowly converges towards more accurate predictions (learns to better perform the task). Thus, in its nature training algorithms is a very repetitive task and is highly parallelizable. This makes it synergetic with some types of computer hardware.

8

The hardware components designed for accelerating the learning process of AI algorithms are very different from typical data center hardware designed for service centers and cloud platforms. Compute units typically emphasizes a large number of processing units (by coupling CPUs and GPUs) to achieve very high processing capabilities measured in number of operations per second (FLOPS). In loose terms, the higher the FLOPS count, the more capable the system is for performing the operations needed to train and develop AI algorithms. When designing or upgrading hardware infrastructure, it is important to make a choice between expanding the CPU count or GPU count of the system. These two components are similar in that they perform the processing tasks. However, the similarities end there as these two components are designed with different purposes in mind.

**FLOPS**: Floating point operations per second. It is one of the indicators of the system's performance.

#### **CPU vs GPU FOR AI WORKLOADS**

Results

Results

Res

#### CPU

- Optimised for serial tasks

#### GPU

- Optimised for many parallel tasks



CPUs (Central Processing Unit) are the main processing units of any computing machine, these components are designed to be a general-purpose processing unit that are able to execute a wide variety of workloads. Building a platform that is mainly composed of CPUs adds a large degree of versatility to the platform making it suitable for data center and generic cloud workloads. This is mainly due to its design choices making it well suited for serial tasks by having limited number of more capable **cores**. Furthermore, CPUs can also perform AI algorithm training tasks albeit at a much lower performance than preferred. CPUs can also be provisioned/ assigned in finer grained resources to perform smaller tasks. Overall, expanding a platform CPU count is a safe choice to steadily increase its processing performance (given that there are no other bottlenecks in the system).

On the other hand, GPUs (Graphical Processing Unit) play a different role in a system, these components are not as versatile as CPUs. However, they make up for it with a much higher performance in very specialized tasks such as 3D rendering and training deep learning models<sup>2</sup>. GPUs are composed of a very large number of computational units (cores) specialized in performing simple mathematical tasks rapidly in parallel which is very suited for training machine learning models. In fact, a single GPU can have thousands of cores compared to tens of cores on a CPU, which makes GPUs vastly superior to CPUs in AI applications. However, the cost of a GPU is significantly higher than a CPU and it remains a highly specialized component that is not suitable for all work loads.

**core**: The building block or processing unit of a modern day CPU

<sup>2</sup> Please refer to the Al Guide



# EXISTING AI HARDWARE INFRASTRUCTURE IN THE UAE

The UAE is an growing leader when it comes to AI infrastructure. In fact, the UAE (at the time of writing) has the 36<sup>th</sup> most powerful highperformance computer in the world<sup>5</sup>. The UAE's AI infrastructure ecosystem is dominated by the private sector owning 89% of the total AI related processing power in the UAE. The academic community comes in second and is home to some large-scale state of the art systems, for example NYUAD hosts "Dalma", a supercomputer composed of more than 870 CPUs and 30 GPUs. These are not the only AI systems in the UAE according to our survey. On the contrary, our survey identified more than 10 high-end systems owned by the government, academia and the private sector. Below is a compiled summary of their combined resources.

<sup>5</sup> Top500 list for June 2020 https://www.top500.org/lists/ top500/list/2020/06/



# DISTRIBUTION OF COMPUTATION RESOURCES IN THE UAE





CPU CORES III	GPU MEMORY GIGABYTES
16,904	
117,968	292403111111111111111111111111111111111111
2,318	

## **INTERNATIONAL POSITION**

The international benchmark and ranking for supercomputers focus mainly on the total processing power of the system. The most notable list containing information on the most powerful supercomputers in the world is "Top500"<sup>3</sup> list. **The UAE currently ranks 36** globaly with the "Artemis" high performance computing cluster owned by G42.

<sup>3</sup> The Top 500 list project was started by HANS WERNER MEUER and ERICH STROHMAIER in the University of Mannheim and uses the Linpack Benchmark for rating the HPCs around the world.

#### **COUNTRY RANKING PER NUMBER OF SYSTEMS ON THE TOP500 LIST**



(Countries having more supercomputers on the top500 list will rank higher)

# RANKING OF COUNTRIES ON THE TOP500 LIST ORDERED BY COUNTRY'S FASTEST SUPERCOMPUTERS<sup>4</sup>

1	2	3	4	5 ()			8	9 (*) 8	<b>10</b>	11	12 C	13 *	14	15	16 +
JAPAN	NSA	CHINA	GERMANY	ITALY	SAUDI ARAB	SWITZERLA	FRANCE	SOUTH KOR	AUSTRALIA	TAIWAN	UAE	UK	RUSSIA	SPAIN	FINLAND

<sup>4</sup> Repeating countries were filtered out.

### **KEY PARTICIPANTS**













#### جامعـة نيويورك ابوظـي NYU ABU DHABI

**G42**: Group 42 is an Abu Dhabi based artificial intelligence (AI) and cloud computing company that owns the 36<sup>th</sup> "Artemis" and the 304<sup>th</sup> "POD3" most powerful supercomputers in the world and is currently being used to fight COVID 19 as well as other scientific projects.

**RTA**: The Road and Transportation Authority of Dubai hosts a CPU and GPU hybrid system which is utilised for conducting Transport and Corporate related analytics and projections. This system is currently the largest local government platform in terms of the number of CPU and GPUs currently in use.

**ADDA**: Abu Dhabi Digital Authority is currently in the process of building one of the largest local government platforms to support the development of their projects and services.

**Morohub**: A Dubai based private entity that offers data center service. Its "Moro Cloud" is composed of both CPU and high-end GPU units.

**ADNOC**: The Abu Dhabi National Oil Company is a global leader in energy production. ADNOC currently houses a powerful supercomputer that is performing scientific oil reservoir simulations and seismic processing.

**UAEU**: United Arab Emirates University is based in Al Ain with over 13,000 students. The university owns 3 high performance platforms that are used by various departments for research. These platforms are also used for scientific Al research such as training deep learning models, medical image processing, big data processing as well as faculty working in physics and math.

**NYUAD**: A university based in Abu Dhabi that currently hosts the most powerful supercomputer in the academic domain in the UAE. The platform is being used for various research tasks including AI, machine learning, NLP, Data Analytics, ML, data visualization, big data & data manipulation, climate and ocean modeling, Sea Level Rise research, Computational Astrophysics, Computational Chemistry, Bioinformatics & Genomics (sequencing), Applied Math, Psychology (fMRI).

البرنامج الوطني للذكاء الاصطناعي NATIONAL PROGRAM FOR ARTIFICIAL INTELLIGENCE **National Program for Artificial Intelligence**: The Artificial Intelligence Office at the Prime Ministers Office is a federal government entity mandated with delivering the UAE's strategy for Artificial Intelligence for 2031. The Artificial intelligence Office announced the AI Lab which houses a GPU based supercomputer that is made available to all government entities, researchers and AI start-ups in the UAE and is the UAE federal government's most powerful super computer.

# CLOSURE

The UAE has a vision to be a leader in AI by the year 2031. Consistent with that vision, we find that the UAE is rapidly growing in terms of the hardware infrastructure for AI with plans to expand many of the surveyed platforms to match the rapid adoption and development of these technologies in the UAE.

In the end we would like to thank the entities that participated in the survey and shared the details of their systems with us.



#### **OTHER PUBLICATIONS**

To read other publications released by the National Program for Artificial Intelligence please scan the relevant barcode.

#### **AI GUIDE**







**BLOCKCHAIN** 

GUIDE

UAE NATIONAL STRATEGY FOR ARTIFICIAL INTELLIGENCE 2031



The National Program for Artificial Intelligence has put extensive effort in creating an accurate report. However, this topic is continuously developing and we strive to maintain accuracy as of date of publishing.

Al Hardware Infrastructure Report UAE Document Number: 001-2020

Version: 01

Published on Nov 2020

Copyright © Minister of State for Artificial Intelligence, Digital Economy & Remote Work Application Office

#### **Contact us**

Minister of State for Artificial Intelligence, Digital Economy & Remote Work Applications Office

email: info@ai.gov.ae



@UAEAI



@uaeai\_2031

www.ai.gov.ae