機器學習於材料資訊的應用
Machine Learning on Material Informatics

陳南佑(NAN-YOW CHEN)

楊安正(AN-CHENG YANG)

NCHC Introduction

Position

MOST 科技部
Ministry of Science and Technology

The Ministry of Science and Technology (MOST)

is the government ministry of the Republic of China (Taiwan) for the promotion and funding of **academic research**, development of science and technology and science parks.

NARLabs
國家實驗研究院
National Applied Research Laboratories

The National Applied Research Laboratories (NARLabs)

is the institution resulted from the combination of national laboratories into an independent nonprofit institute.

One of 8 national-level research laboratories under NARL

NCHC
NARLabs 財團法人國家實驗研究院
國家高速網路與計算中心
National Center for High-performance Computing

Founded In 1991

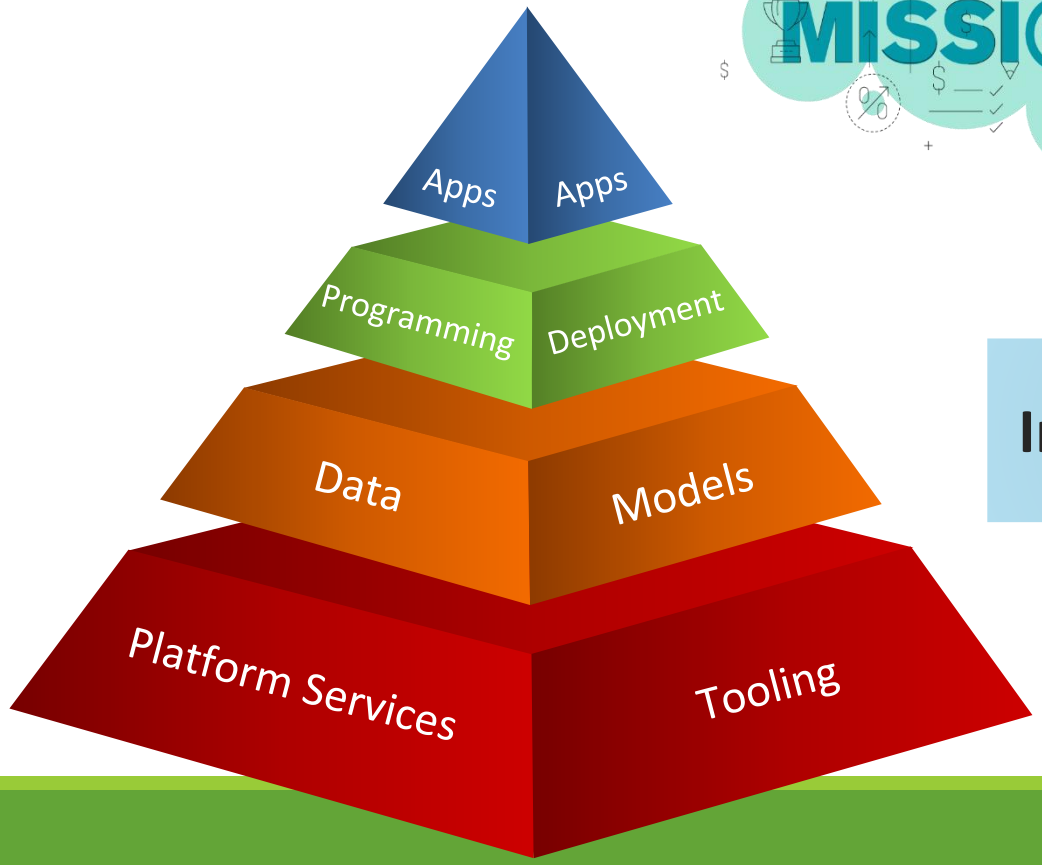
National Center for High-performance Computing (NCHC)

Taiwan's only national-level supercomputing center.
The NCHC possesses a large computing and networking platform facilities for use by domestic academia and the general public.

Vision and Mission

Become a World-Class Supercomputing and Big Data Center

Enable Scientific Discoveries & Technical Innovation through prospective computing technology and platform



Innovations → Business Value

World Class AI and Big Data Platform

Vision with Execution in mind: start with a solid platform

NCHC Milestones

A Member of NARLabs
National Center for High-performance Computing

1991
Taiwan's first national level supercomputer center

1993
Hsinchu Headquarters

2003
NPO under NARLabs

2004
TWAREN Services 10G

2005
Tainan Office

2008
Taichung Office

2011
177 TF ALPS Supercomputer

2016
100G Network Backbone

2017
1.33 PF Peta scale HPC
台灣杉 TAIWANIA

2018
9PF AI-HPC
台灣杉二號 TAIWANIA 2

2019
AI Platform
TWCC TAIWAN COMPUTING CLOUD

Hsin Chu Headquarters

Taichung Office

Tainan Office

Certifications

- ✓ *ISO 9001 (Plus Award)*
- ✓ *ISO 27001*
- ✓ *ISO27018*
- ✓ *ISO20000*
- ✓ *CSA STAR Level 2 Gold Award*
- ✓ *BS 10012*

NCHC Development Structure

Applications

- Engineering and Science
- Biomedical Science
- Environment and Disaster Prevention
- Social Science and Economics
- Digital Cultural Creation
- Emerging Applications

Cloud Services

Infrastructure as a Service
- Real machine/Virtual machine
- Storage
- Networking

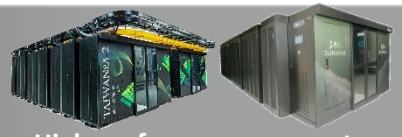
Platform as a Service
- HPC
- Big data analysis
- Render farm
- AI framework

Software as a Service
- NARLabs digital service
- Colife education platform
- Scientific & engineering simulation

Core Technologies

High performance computing | Cloud Middleware | AI & Big data | Network and information security | Visualization

Core Facilities



High performance computers
Big data analysis clusters

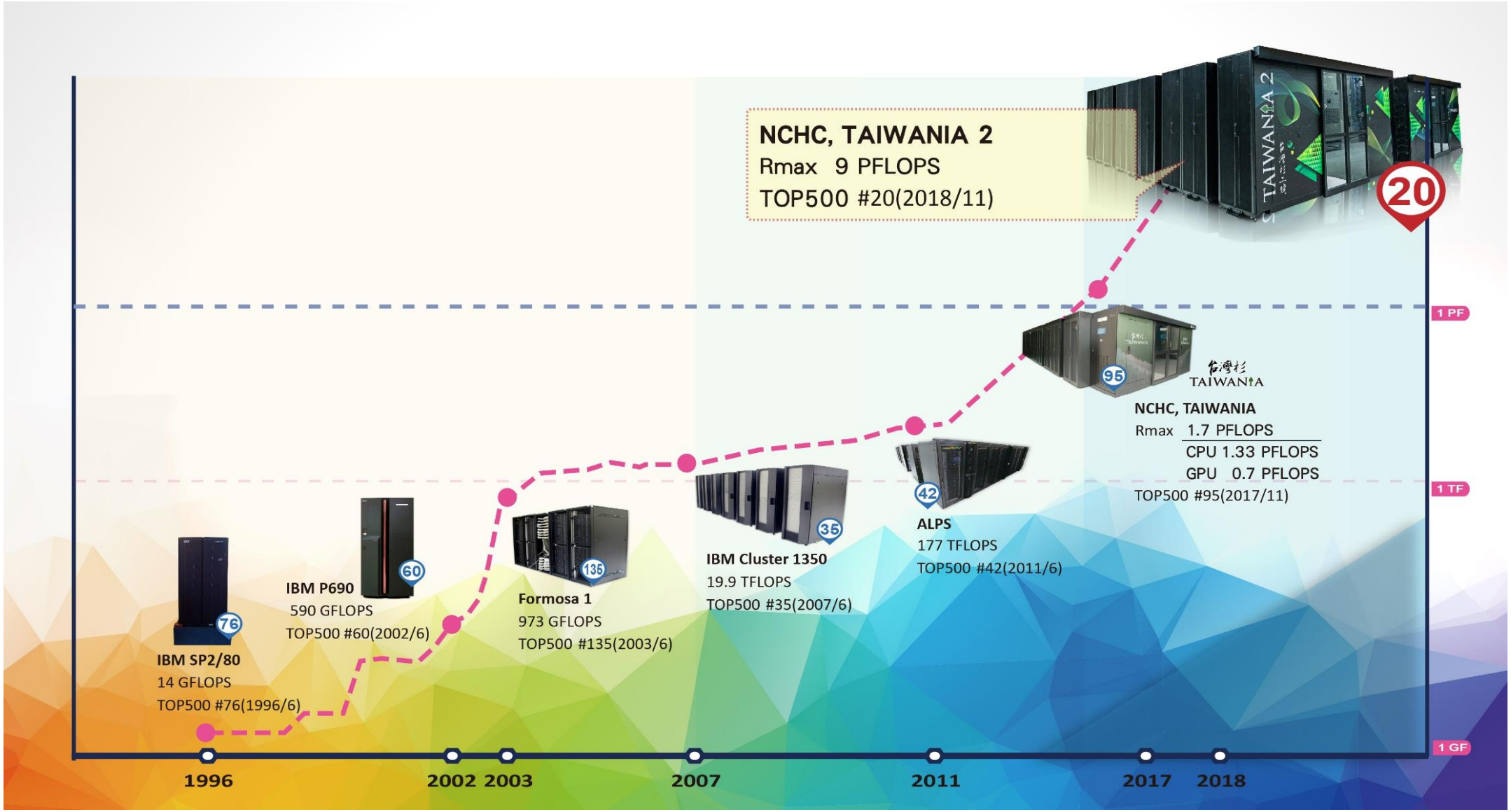
High performance storage equipment

High quality academic research network

Infrastructure

- NCHC HPC History
- Taiwania 1 and Taiwania 2
- Research and Education Network and Storage
- Big Data Analysis Platform and Render Farm
- Cyber Security

NCHC HPC History



Taiwania 1

台灣杉
TAIWANIA

- A general purpose supercomputer with a large number of compute nodes and an independent file system, high security mechanisms
- Supports multiple fields of application, including physics, chemistry, mathematics, atmospheric, engineering and life sciences.

425 million computing hours/year

900+ projects/year

4300+ accounts/year

Launch in May 2018

Supplier : Fujitsu

https://iservice.nchc.org.tw

Computing cores:27,760
CPU ： 1.33 PF
GPU : 0.4 PF
Storage : 3.4 PB
Networks:100Gbps

Efficiency : 4 GFLOPS/W

PUE< 1.3

Taiwania 2

台灣杉二號
TAIWANIA 2

The NCHC is accelerating
AI innovation in Taiwan

TAIWANIA 2
9
petaflops
Rmax

Directed by
MOST 科技部
Ministry of Science and Technology

Specification setting by
NCHC NARLabs 財團法人國家實驗研究院
國家高速網路與計算中心
National Center for High-performance Computing

Manufactured by Taiwanese companies
ASUS 廣達電腦 Quanta Computer Taiwan Mobile 台灣大哥大

20TH 計算能量
TOP 500

10TH 能源效率
Green 500

2018. Nov. ranking

https://www.top500.org/

Rank	System		Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
19	CINECA Italy	**Marconi Intel Xeon Phi** - CINECA Cluster, Lenovo SD530/S720AP, Intel Xeon Phi 7250 68C 1.4GHz/Platinum 8160, Intel Omni-Path Lenovo	348,000	10,384.9	18,816.0	
20	NVIDIA Corporation United States	**DGX SuperPOD** - NVIDIA DGX-2H, Xeon Platinum 8174 24C 3.1GHz, NVIDIA Tesla V100, Mellanox InfiniBand EDR Nvidia	127,488	9,444.0	11,209.1	
21	National Center for High Performance Computing Taiwan	**Taiwania 2** - QCT QuantaGrid D52G-4U/LC, Xeon Gold 6154 18C 3GHz, Mellanox InfiniBand EDR, NVIDIA Tesla V100 SXM2 Quanta Computer / Taiwan Fixed Network / ASUS Cloud	170,352	9,000.0	15,208.2	798
22	DOE/SC/Argonne National Laboratory United States	**Mira** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM	786,432	8,586.6	10,066.3	3,945
23	GSIC Center, Tokyo Institute of Technology Japan	**TSUBAME3.0** - SGI ICE XA, IP139-SXM2, Xeon E5-2680v4 14C 2.4GHz, Intel Omni-Path, NVIDIA Tesla P100 SXM2 HPE	135,828	8,125.0	12,127.1	792

https://www.isc-hpc.com/

http://supercomputing.org/

台灣杉二號
TAIWANIA 2

Hardware - whole system

- 252 nodes / 9072 CPU cores /2016 GPUs
- 193.5 TB memory
- 10 PB storage
- EDR InfiniBand 100 Gbps
- 1.2 PUE (Warm Water Cooling)

Software Environment

- Slurm / Kubernetes
- Nvidia NGC Docker
- Ceph
- Spectrum Scale (GPFS)
- CentOS

Hardware - single node

- Intel Xeon Gold CPU x 2
- Nvidia Tesla V100 w/32GB x 8
- 768 GB memory
- 240 GB SSD + 4TB NVMe

AI Framework

- Tensorflow
- Caffé / Caffé 2
- PyTorch / Torch
- …… and more

Caffe2
PYTORCH
torch

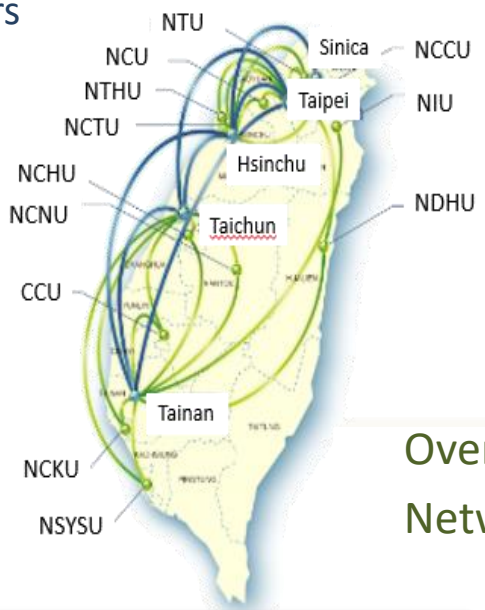
NCHC NARLabs 財團法人國家實驗研究院 國家高速網路與計算中心 National Center for High-performance Computing

台灣杉二號 TAIWANIA 2

Research and Education Network

TaiWan Advanced Research and Education Network

- **TWAREN** (Optical, dedicated bandwidth)
  - ✓ 100G Bandwidth
  - ✓ 12 GigaPOPs
  - ✓ 94 universities & research institutes
  - ✓ 500 K. users



Over 99.99%
Network Availability

TWAREN Domestic Backbone

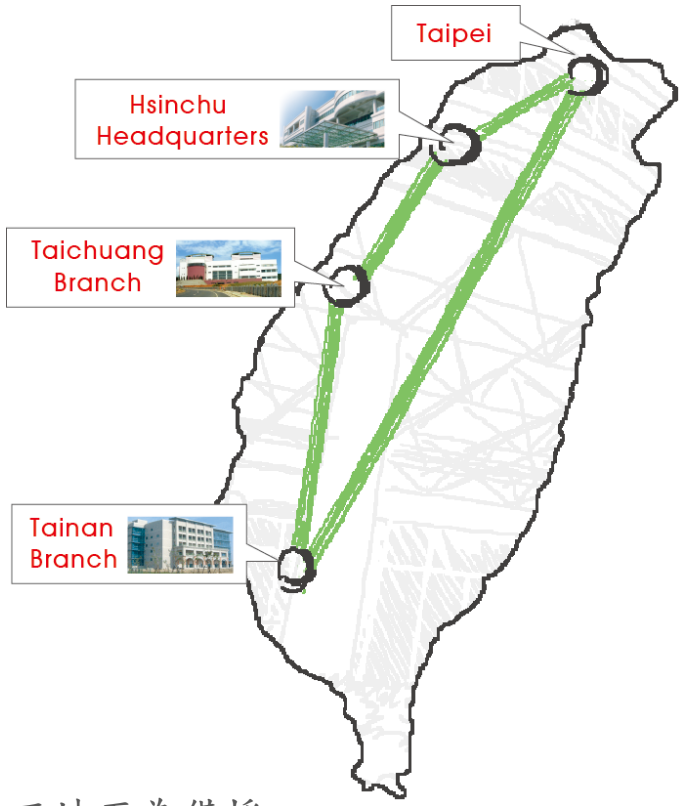
✓Sharing underlying optical network with **TANet** (4000 schools, 4.5 M. users)

✓Network Availability: 99.99% ↑

✓20G **International Bandwidth**

TWAREN International Connection

儲存設施服務



- 建置全國專業領域共用共享大資料儲存資料平台，達到資料共用，研究成果分享之效益
- 每年服務國內腦科學、基因演算、防災、地科、氣候變遷、人文資料庫等30餘重要計畫
- 備份農航所航照圖、太空中心衛星圖、中研院基因體等資料，為國家重要資料提供保障。

EMC Connectrix Director

StorageTek L5500

- 三地互為備援
- ISO9000、ISO27001雙認證、雙保障
- 以自有技術建置負載平衡、異地備份、高擴充性之雲端儲存服務系統
- 提供資料備份、資料保存、資料備援、雲端儲存、及客製化資料儲存服務

- 儲存總容量 50 PB

Big Data Analysis Platform - Braavos

- Launched in January 2016
- 300Nodes,2PB
- Running Apache Hadoop

Service Models

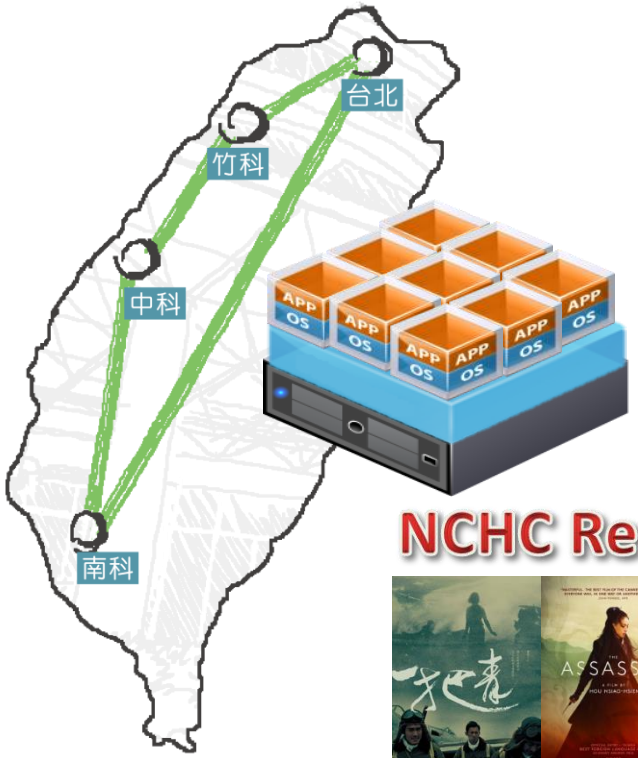
- Time sharing for research and academia
- Dedicated platform for governmental data
  - providing de-identification tools for governmental departments

7 Functions in Service

- MapReduce
- Spark
- Hive
- Hbase
- R-hadoop
- Mahout
- Flume

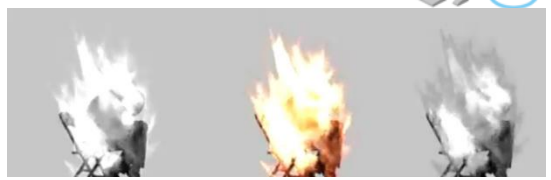
Render Farm

- Launched in 2011 to promote Taiwanese animation film industry
- Adopt market-leading rendering engines & management software
- Establish remote desktop using GPU to speed up film production
- Develop and improve special effect algorithm



NCHC Render Farm Service

Support iconic film productions

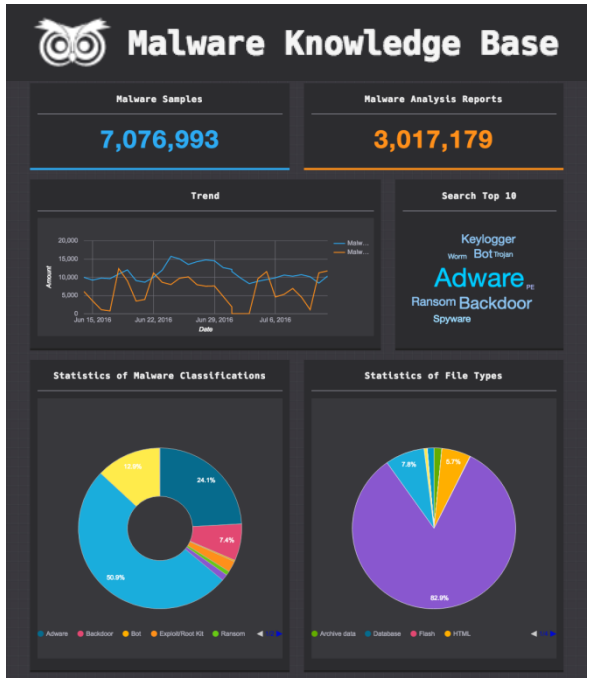


“Below Horizon”special effect of fire

Cyber Security

Malware behavior analysis

- Build the **Malware behavior knowledge database (13+ M malware samples)** , the only one in Taiwan
- Deploy a large-scale Honeynet platform **(6,000+ Honetpot & 65+ GB/Day)**
- Operate **TWCSIRT** (Taiwan Computer Security Incident Response Team) for Intelligence sharing and analysis in the cyberspace
- Organize the **IRCON** (Incident Response Conference) in Taiwan to connect with the other CERT/CC, CSIRT in the global



Taiwan Malware Analysis Net
臺灣惡意程式分析網
TWMAN

TWCSIRT 臺灣電腦安全事件應變中心
Taiwan Computer Security Incident Response Team

From HPC to AI

- NCHC AI Cloud
- National R&D Infrastructure for AI
- AI Platform-TWCC
- TWCC Service
- Impact of TWCC

Grand Plan in Taiwan

5 Strategies

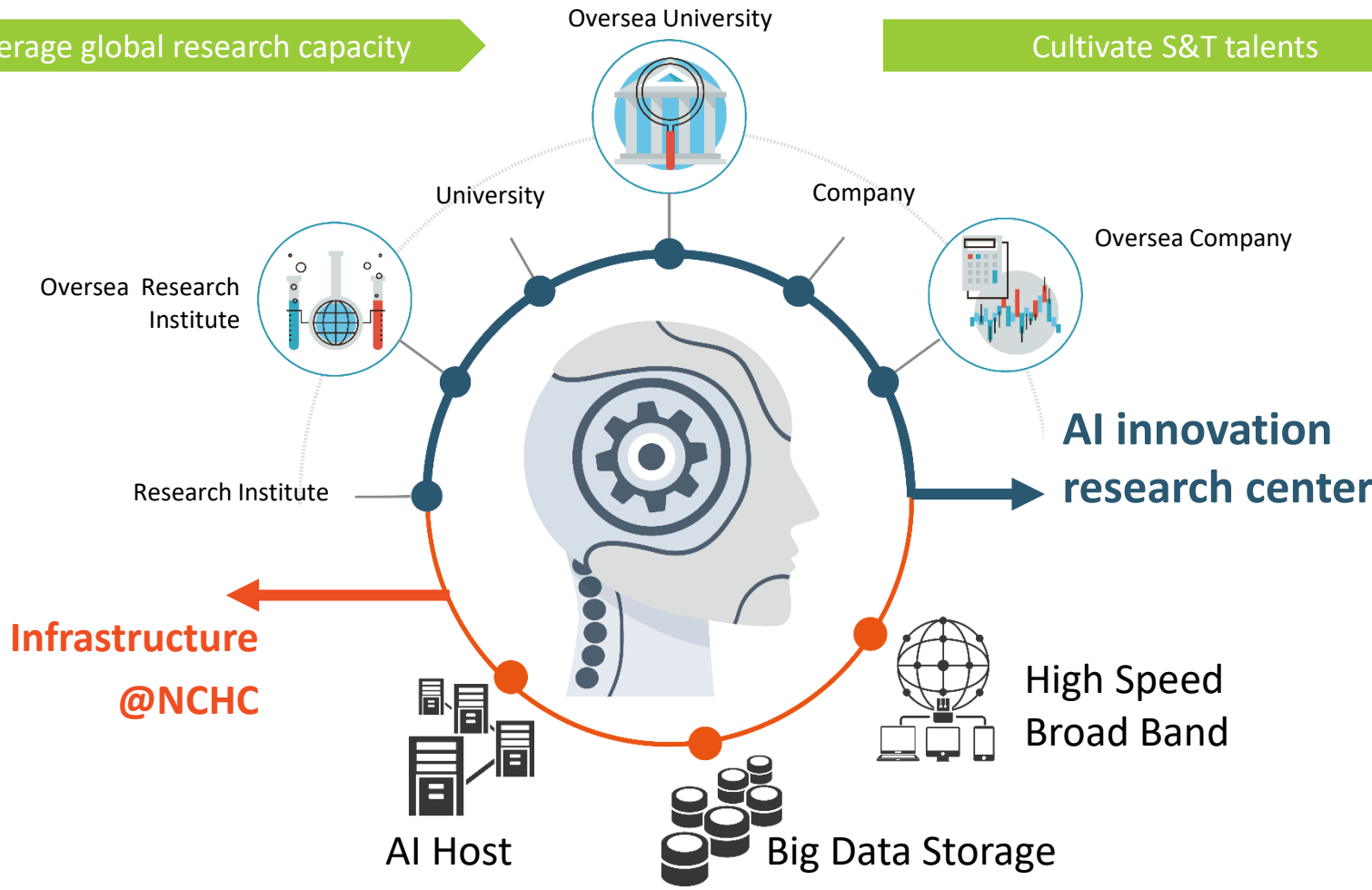
NCHC Supporting AI Ecosystem of Taiwan

Establish cross-disciplinary R&D team

Leverage global research capacity

Provide an integrated platform

Cultivate S&T talents



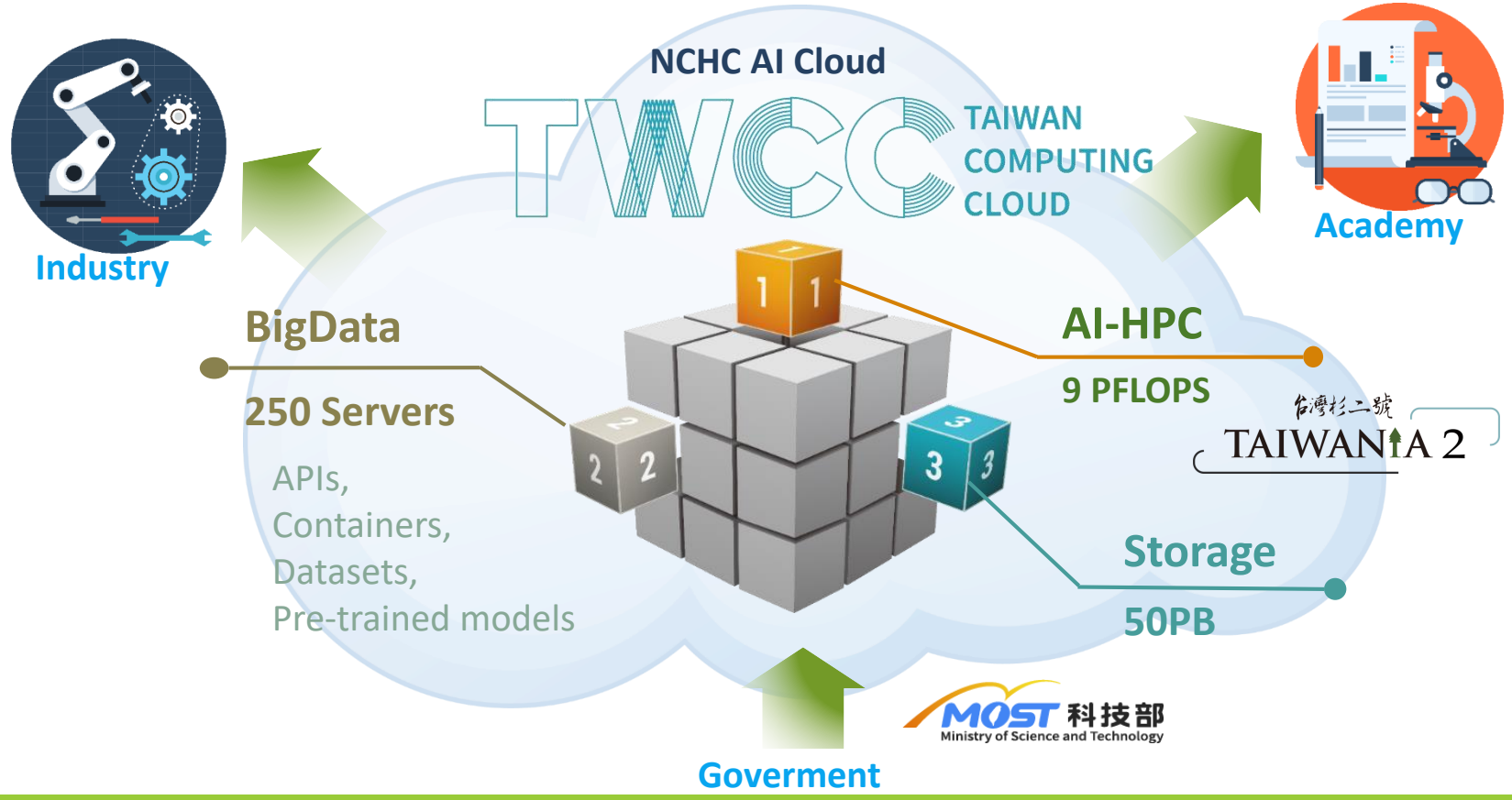
NCHC AI Cloud (4-year National Project)

Target

- Support AI Open Innovation Ecosystem in Taiwan
- Accelerate Taiwan's AI research and industry

Approach

- Leverage Taiwan's ICT superiority
- Take advantage of open AI frameworks, open learned models and open data



NCHC AI Platform-TWCC

On demand
Computing and storage environment

AI/Big Data/HPC
Integrated cloud service platform

AI and cloud resources single service portal

TWCC
TAIWAN COMPUTING CLOUD

台灣杉二號
TAIWANIA 2

Customized fast-form
AI solution

Top performance
Beyond the self-built cloud platform 1.72 times

High convenience and high performance
Containerized computing environment

Multi-level security
Network security protection

TWCC Services

Rapid & Lightweight Deployment

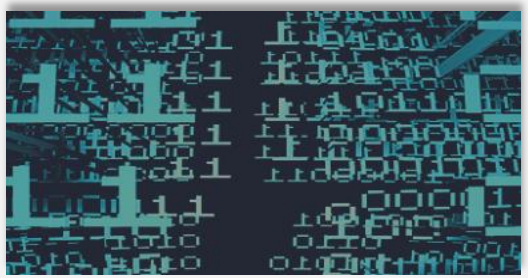
Container Compute Service

Container Compute Service provides rapid deployment of GPU working environment and improves work efficiency by 29%.

Effective Command & Control

High-performance Computing

Deploy a multi-node, multi-GPU, distributed high-performance parallel computing environment that produces at least 30% higher efficiency.



Smart & Speedy Computations in One

Virtual Compute Service

Create a safe, secure, and flexible Virtual Compute Service instance in short order.

Secure Data Repository

Data Storage Service

~50 PB Hyper File System storage so far and still being expanded and upgraded .

台灣杉二號
TAIWANIA 2

Application


Smart City


Disaster Mitigation


Industry 4.0


Autonomous Car


Biomedical Science


Robotics

Service


Image Recognition

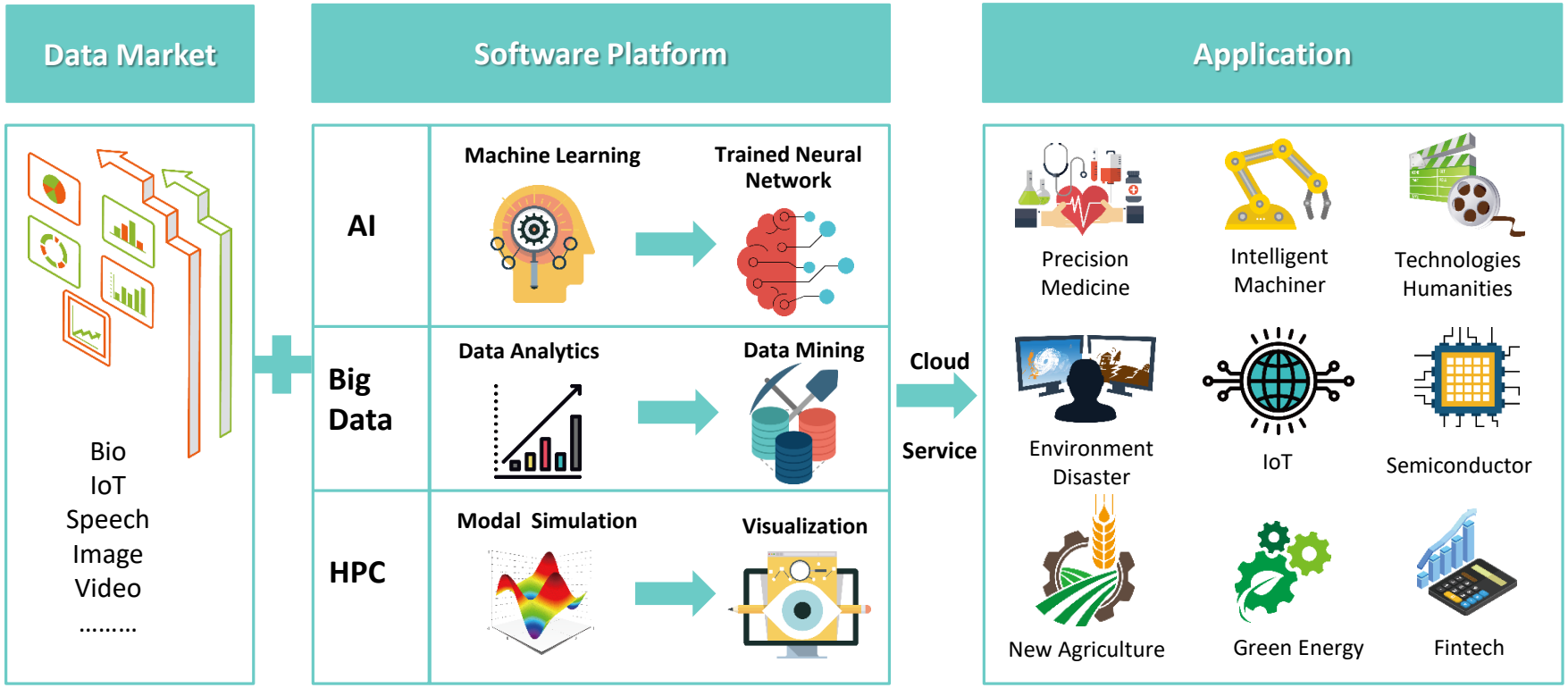

Text Mining


Natural Language Processing


Vehicle Speed Estimation


Bioinformatics Analysis

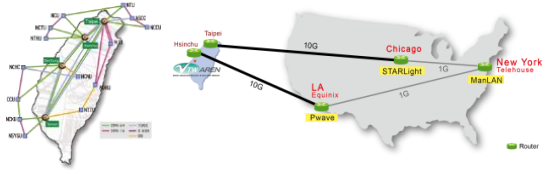
TWCC (Taiwan Cloud Computing)- National Infrastructure for AI



Cloud-based resource management, software and infrastructure

CPU+GPU(>10+ PF)

Storage (150 PB)



TWAREN Backbone (100 Gbps)

Impact of TWCC

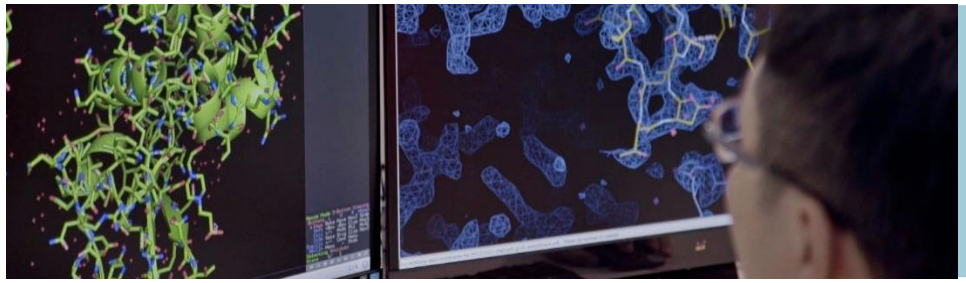
AI Image training

1.76 million Sheets Per second

Pathogen identification

shortened from

7 Days to 5 Hours



Disease detection

shortened from

One week to 12 Hours

Tumor diagnosis

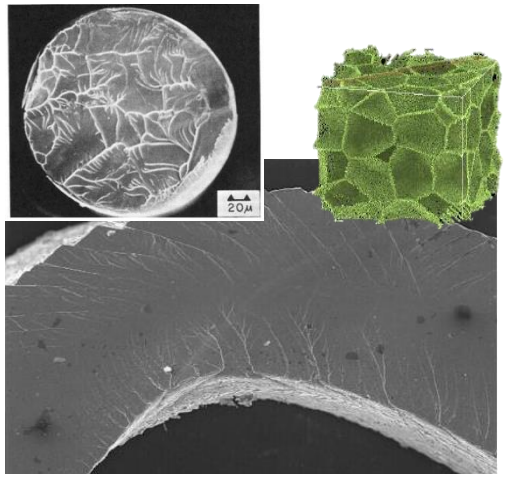
shortened from

48 Hours to 1 Hour

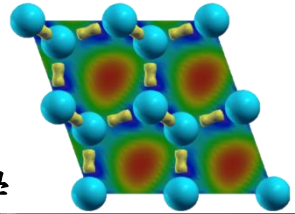
Taiwania2: 16 GPUs*126

材料應用領域

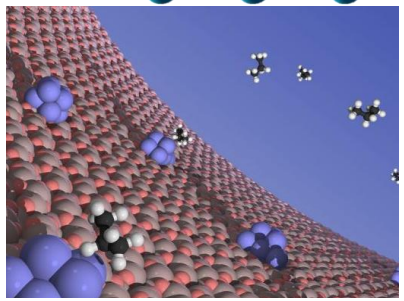
材料



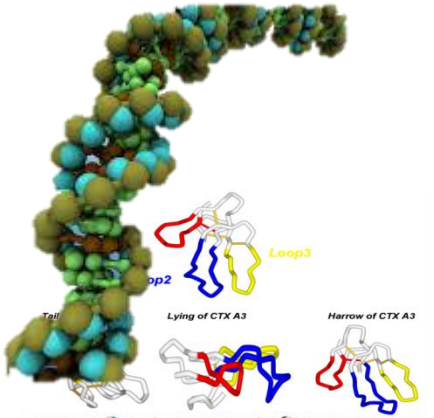
物理

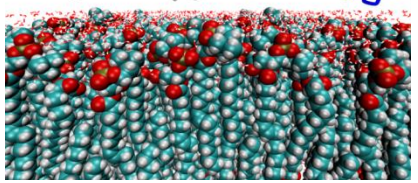


化學

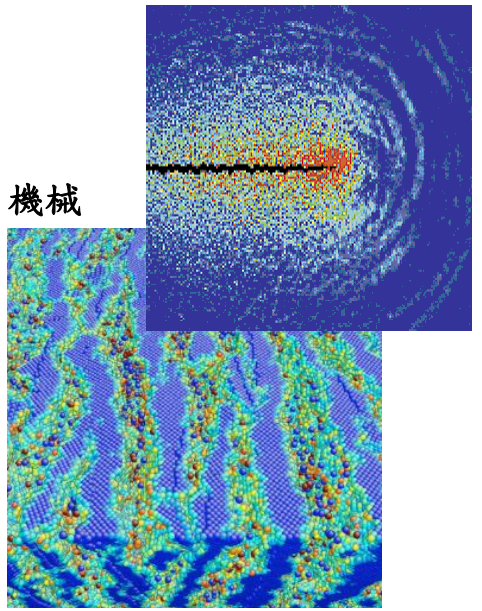


生物



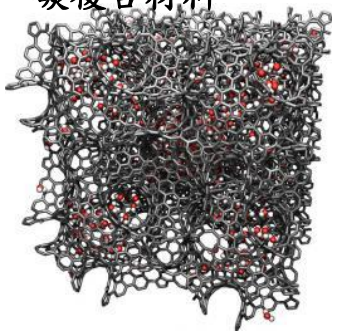


機械



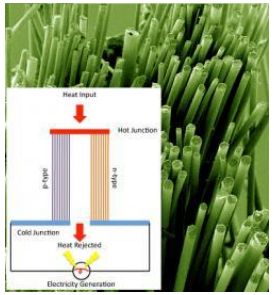
潛力應用: 半導體材料、能源材料、光電磁材料、合金材料...

碳複合材料



高強度合金

熱電材料



奈米結構