

生成式 AI 時代 Generative AI Era (AI 2.0)

生成式人工智慧是未來創新的驅動力
Generative AI is the driving force for future innovation.

主權AI: 自建可掌控的人工智慧系統為重要的關鍵
Owning your AI systems is crucial to building sovereign AI.



10³X~10⁵X
Computing power
requirement



Model Training Phase

CNN / RNN (< million parameters)

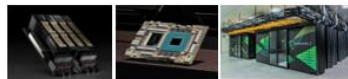


Model Inference Phase

CPU / GPU / NPU



Transformer - based LLM (> billion parameters)



GPU / NPU (Memory is key)



台智雲和華碩集團攜手合作提供AI算力基礎建設和平台營運服務
TWSC and ASUS collaborate to provide an end-to-end total AI data center solution.

提供AIHPC超級電腦平台解決方案，整合軟硬體架構，推動企業、城市乃至國家建立自己的AI主權能力，提升技術實力與競爭力，實現永續的管理目標。

We have created AIHPC supercomputer platforms that drive enterprises, cities, and even nations to establish their own AI capabilities, enhance technological prowess and competitiveness, and achieve sustainable administrative goals.



IaaS (Infrastructure-as-a-Service)

以人為本的設計思考

Human-centered Design Thinking

從主機軟硬體系統架構、整體資訊系統安全、機房設計與用電節能、網路與儲存系統、軟體與資安工具，全方位考量打造最前瞻的基礎設施。

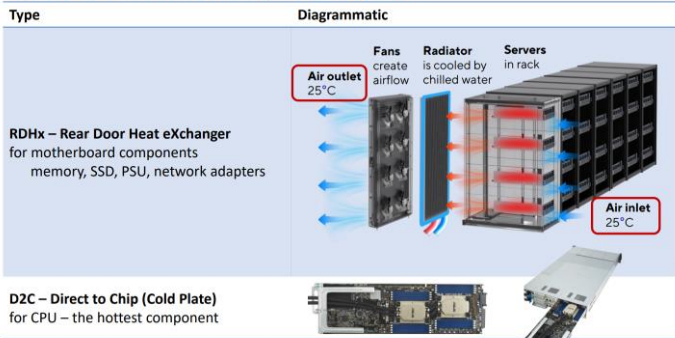
We comprehensively consider from the architecture of the host hardware and software systems, overall information system security, energy-efficient data center design and power usage, networking and storage systems, to software and cybersecurity tools, in order to create the most forward-thinking infrastructure.

Machine Room and Rack layout

- no raised access floor
- overhead cable tray
- warm water cooling



Hybrid Solution of Air-cooling & Liquid-cooling



Advanced Security Protection

- Real-time detection of abnormal traffic and DDoS attacks by monitoring border router Netflow data.
- Mitigate the impact of DDoS attacks on TWCC by rerouting and cleansing traffic
- Further protected by the joint defense with the UTRS(Unwanted Traffic Removal Service)
- ISO/IEC 27000-series comprises information security certified
- User decentralization management, change management review
- Periodic system vulnerability scans
- Notification of information security incidents
- Automatic abnormal event ticketing and tracking
- Malicious IP automatic blocking
- SIEM (Security Information and Event Management) platform to collect and analyze heterogeneous device logs
- SOC monitors network traffic and provides alerts and notifications of information security incidents for threat behaviors
- Micro Segmentation: Avoid unauthorized access by virtual machine segmentation
- Distributed storage architecture: Ensuring data integrity and high availability
- NGFW: Fine grained access and flow control through high performance firewall
- WAF: Protect websites, web apps and web data from network attacks
- Intrusion Detection system: Real-time detection and identification of network threat characteristics
- Server Security Maintenance: Periodic server credential and log check and warning
- APT: Automatic sandbox to analyze captured malwares and botnet activities
- Sending APT attack warning through the SOC integration



PaaS (Platform-as-a-Service)

嚴謹營運的智慧化管理

Rigorous Operational Smart Management

從營運管理、操作使用、功能設計、升級更新、多元軟體選擇、資料授權與維護，全方位考量打造最便利的使用和管理平台。

Taking into account operational management, user operation, feature design, upgrade updates, diverse software selection, data licensing, and maintenance, we aim to create the most convenient and user-friendly management platform.

Main Feature – User Side



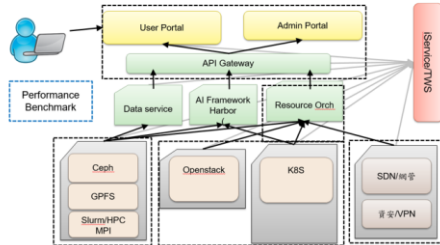
Technology Transfer and Operation

- Three-stage transfer planning:

- (1) Acceptance test
- (2) Test and trial period
- (3) operation period

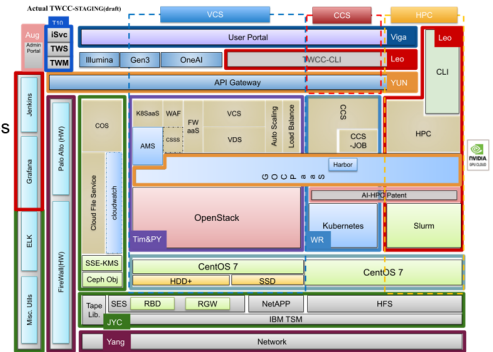
- Operation Requirements :

- (1) Undertake the whole software and system
- (2) Update in staging and production Environments
- (3) Plan the release management process
- (4) Rolling Upgrade
- (5) Monitor resource usage and performance



Software Feature & Stack

- Integrated account system
- E-wallet
- Long-term archiving storage services
- OneAI one-stop machine learning
- UX enhanced / I18n Web Interface
- TWCC Command-line
- Continuous software update



AIHPC-as-a-Service

生產力革命-高效能駕馭AI應用發展

Productivity Revolution : Highly Efficient Control of AI Application Development

專為AI高速運算設計的架構和功能，可以建置專屬的大語言模型和生成式AI應用，掌握AI主權能力和競爭力，再創智慧城市高峰。

The architecture and functions specially designed for high-speed AI computing can build exclusive large language models and generative AI applications, master AI sovereignty capabilities and competitiveness, and create new heights in smart cities.

AI HPC : Essential for Large Scale LLM Training

176B Model both Training and Inference can be run on **Taiwania-2**.

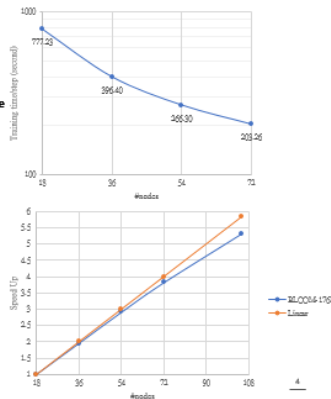
- Training time gradually decreases as the number of nodes increases.
- 18 Nodes (144 GPU), 36 Nodes (288 GPU), 54 Nodes (432 GPU), 72 Nodes (576 GPU), 90 Nodes (720 GPU), 108 Nodes (864 GPU), can achieve close to linear acceleration.



- Performance demanding HPC Workloads - High Performance Computing
- Cloud native
- High-speed network
- High-Performance Storage

We can linearly scale out, with each GPU node reasonably and efficiently allocating computing work. There will be no idling. Every dollar spent by the customer is effectively used on each GPU, without any waste.

- Linear = Theoretical
- Blue BLOOM-176B = Real training accelerate performance on Taiwania-2



Analytics

Smart File System

Effective storage for hot and cold data
High-speed storage designed for AIHPC



GPU computing

AFS + OneAI + CCS + HPC

One-stop solution to AI development
A fully-managed AI container service
Distributed parallel computing



Availability

Proactive & Predictive

Woodpecker Project for Instant Failure Detection
Sub-Second Failover
End-to-end Based User scenarios health monitoring
Max. Availability Architecture



Security

End-to-End Protection

Advanced Intrusion Detection
Full-Stack Security Scanning
Isolation and Virtualization
Secured cloud networking solutions



AIHPC Services on Cloud and on Prem.

Research, Development and Fine-tuning LLM

LLM Production

OneAI (no-code)

Customer can try various transformer-based models via a no-code UI.



HPC (Job)

Customer can train a LLM with hundreds of GPUs, high-speed networking and storage.



CCS (Container)

Customer can select pre-build images to deploy LLM codes with Multi-node configuration.



AFS (Workbench)

Customer can select pre-trained models, fine-tune demo by no-code UI, and deploy the models on cloud and on premise.



AI 2.0 New Ecosystem

指導單位



經濟部中小及新創企業署
SMALL AND MEDIUM ENTERPRISE AND STARTUP ADMINISTRATION
MINISTRY OF ECONOMIC AFFAIRS



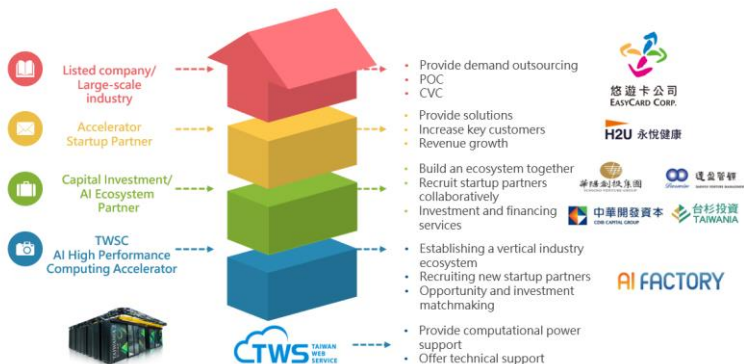
AI超算加速器

AI High Performance Computing Accelerator

台智雲專為AI新創夥伴技術、商品、市場提供全面升級的加速平台，集結華碩集團的品牌、產品創新能量以及全球市場，連結國際及跨領域平台合作，促進新創產業投資及衍生商機。

TWSC is dedicated to providing a comprehensive upgrade acceleration platform for AI new partners' technologies, products, and markets. It gathers ASUS Group's brand, product innovation energy, and global markets, connects international and cross-domain platform cooperation, and promotes investment in new industries and derivative business opportunities.

Industrial Innovation Consortium



Solution for AI 2.0 · AIHPC · New Ecosystem

AIHPC Core Technology

HPC technology team with over 200 engineers and 6 years of experiences.

We provide cutting-edge supercomputer solutions, featuring leading **Hardware and Software** configurations.

Generative AI

We develop **Enterprise-grade generative AI** solutions this season for better flexibility, security, and portability.

Precision Medicine

We design **computational genomics** services based on AIHPC.

Digital Twin

Over AIHPC, we deploy high-fidelity and **Digital Twin** projects for government.



Smart Transportation Smart Governance Smart Buildings

Smart tourism Smart Energy Smart Security Smart Factory

Internet of Things Applications Smart Innovation Unmanned store

AI High Performance Computing Accelerator Program



Global AI Foundry



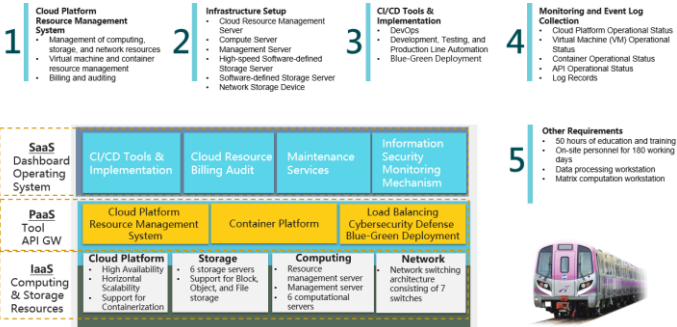
領航台灣 鏈結國際

Navigating Taiwan and Linking to the International

台智雲和華碩集團擁有台灣自主雲端平台技術能量最高、雲原生技術最成熟的自研自建專業團隊，提供台灣多個縣市政府打造智慧城市，並和東南亞國家協會 (ASEAN) 的多個成員國合作，展開AI算力基礎建設和平台的建置，建立未來國際算力鏈版圖。

TWSC and ASUS Group boast a self-researched and self-built professional team with the highest technical expertise and the most mature cloud-native technology in Taiwan's independent cloud platform. We collaborate with numerous county and municipal governments in Taiwan to develop smart cities. Additionally, we partner with multiple organizations of the Association of Southeast Asian Nations (ASEAN) to construct AI computing power infrastructure and platforms, establishing the future landscape of Global AI Foundry.

metro Taoyuan Taoyuan Metro Cloud Platform setup project



結合數位學生 協助城市治理

系統整合架構



Smart City for Tainan Government

Video/Image Data Management & Recognition System

- Tainan City has been promoting the development of a smart city and establishing a digital government to deepen data governance.
- Using AI technology to provide precise feedback data to strengthen decision-making quality.

