China's AI compute:
Shifting from single-chip performance to system-level integration

sinolytics

	PRODUCT	AI CHIPS	CARDS PER RACK	SELF-CLAIMED SCALABILITY
Huawei	Atlas 950 SuperPoD	Ascend 950 NPU	**64** cards per rack	**8,192**
Sugon	Sugon 8000	Domestic AI accelerators, e.g. Hygon DCUs	**640** cards per rack	scalable to hundreds of thousands of cards
Alibaba	Pangu AL128	Zhenwu M890	**128** cards per rack	**Not disclosed**
Biren Technology	NPO Optical Interconnect Supernode	BR2xx series GPUs	**No fixed rack density**	**1,024**-card scale up configuration
Moore Threads	MTT C256 Supernode	MTT-series GPUs	**128** cards per rack	scalable to **hundreds of thousands** of cards
TSINGMICRO	REX81 Supernode	TX81	**No fixed rack density**	scalable to clusters with **tens of thousands** of cards
MetaX	Xijing S600 GPGPU	Xiyun C600	**64** cards per rack	scalable to **tens of thousands** of cards
EVAS Intelligence	Epoch RISC-V Supernode	Epoch RISC-V AI	**64–128** chips per system	scalable to **hundreds of thousands** of cards

IMPORTANT
Bars are grouped by metric type; they are not one common scale.
Claimed scalability is shown separately from concrete configurations.

Sources: Sinolytics research, WAIC

Weekly
Facts &
Figures.

- **Supernodes:** China's AI compute race is shifting from chips to integrated systems.

- **Scale:** Huawei showed a 1,024-card system; Sugon showcased a 100,000-card cluster.

- **Reality check:** Only a few vendors have achieved scaled commercial deployment; many remain at prototype stage.

What
This
Means.

"Supernodes may ease China's domestic AI compute constraints, but they are a workaround, not a cure. The real solution still lies in breakthroughs in leading-edge manufacturing, advanced packaging, and software ecosystems."

Mengying Tao
Project Leader

Start the conversation.

mengying.tao@sinolytics.de

sinolytics.de

sinolytics