

Huawei Cloud: Podpora podmínek pro výpočetní techniku a průkopníků umělé inteligence v průmyslových odvětvích

21.9.2025 - Komerční sdělení | ČTK

Druhý den konference HUAWEI CONNECT 2025 přednesl Zhang Ping'an, výkonný ředitel představenstva společnosti Huawei a generální ředitel Huawei Cloud, hlavní projev s názvem „All Intelligence: Empowering AI Pioneers for Industries“. Informoval o inovacích a postupech společnosti Huawei Cloud v oblasti výpočetních služeb AI, základních modelů, ztělesněné AI, agentů AI a mnoha dalších.

Neustálé inovace v oblasti AI Compute Service: možnosti výkonných počítačů v inteligentní éře

Letos oznámila společnost Huawei Cloud službu AI Compute Service založenou na technologii CloudMatrix384. Specifikace supernode Huawei CloudMatrix se zvýší z 384 karet na 8192 karet. Supernody mohou podporovat hyperškálový cluster s 500.000 až 1 milionem karet, a poskytovat tak robustní výpočetní systém AI, který je v inteligentní éře zdrojem nedožitého významu. Společnost Huawei Cloud uvedla také inovativní paměťové úložiště se svou službou Elastic Memory Service (EMS), která jako první v oboru rozšířila video RAM o paměť. Tím se výrazně snižuje latence vícekolových konverzací na modelech základny, což výrazně zlepšuje pohodlí uživatelů.

Společnost Huawei Cloud nasadila datová centra AI plně chlazená kapalinou v čínských provinciích Kuej-čou, Vnitřní Mongolsko a An-chuej. Tato datová centra s AI podporují odvod tepla o výkonu 80 kW na skříň, snižují indikátor energetické efektivity (PUE) na 1,1 a nabízejí provoz a údržbu s podporou AI. Podniky tak nemusejí rekonstruovat tradiční datová centra ani budovat nová. Budou potřebovat pouze pár optických vláken, aby se mohly připojit k datovému centru a získat přístup k efektivním výpočetním službám AI a také k plnohodnotným dedikovaným cloudovým službám AI v cloudu Huawei Cloud.

Zhang Ping'an zdůraznil, že služba AI Token Service společnosti Huawei Cloud eliminuje technickou komplikovanost a přímo poskytuje uživatelům konečné výsledky výpočtů AI. To uživatelům umožňuje využívat výpočetní výkon inference co nejefektivněji. Supernode CloudMatrix384 realizuje plně sdružování výpočetních, paměťových a úložných zdrojů, odděluje výpočetní úlohy, úložné úlohy a expertní systémy AI a převádí sériové úlohy na distribuované paralelní úlohy, čímž výrazně zvyšuje výkonnost systému při inferencích. Ve scénářích zahrnujících inferenční úlohy s různými požadavky na latenci, jako jsou online, nearline a offline inference, poskytuje CloudMatrix384 průměrný inferenční výkon na kartu, který je 3 až 4krát vyšší než u H20.

Zhang Ping'an oznámil na konferenci oficiální spuštění služby AI Token Service powered by CloudMatrix384. Ta poskytuje zákazníkům vynikající výkon, služby a kvalitu.

Přímé řešení problémů, aby si podniky mohly budovat své vlastní modely

Společnost Huawei Cloud zdokonaluje své modely Pangu tím, že se hloubkově zabývá do specifickými odvětvovými scénáři, spolupracuje se svými zákazníky na řešení jejich nejnaléhavějších problémů a proměňuje představy o tom, co je v těchto odvětvích možné. Společnost Huawei využívá openPangu k poskytování osvědčených postupů pro trénování AI a inferenci, což vývojářům

usnadňuje efektivní využívání výpočetního výkonu AI. Zhang Ping'an poznamenal, že společnost Huawei zároveň vyvíjí model Pangu s uzavřeným zdrojovým kódem. Společnost Huawei bude neustále zvyšovat investice do modelů Pangu, studovat odvětvové scénáře, aby lépe porozuměla požadavkům zákazníků, a podporovat je při vývoji jejich vlastních modelů specifických pro dané odvětví, čímž urychlí inteligentní transformaci napříč odvětvími.

Modely Pangu byly použity ve více než 500 scénářích ve více než 30 odvětvích. Hrají významnou roli v oblastech, jako jsou státní správa, finance, výroba, zdravotnictví, těžba uhlí, ocelářství, železnice, autonomní řízení a meteorologie.

Nad rámec terminálů: cesta k nekonečnému vývoji inteligence v cloudu

V letošním roce spustila společnost Huawei Cloud platformu CloudRobo Embodied AI Platform, která využívá komplexní algoritmy a inteligentní logiku v cloudu a umožňuje realizovat lehčí roboty. Díky obrovskému výpočetnímu výkonu a pokročilých modelů umělé inteligence v cloudu umožňuje tato platforma inteligentnější provádění robotických činností. Cloudová inteligence překonává omezení, která roboty dosud brzdila, a umožňuje jejich nasazení ve větším počtu scénářů.

V zájmu vytvoření jednotného, otevřeného a bezpečného komunikačního kanálu mezi roboty a cloudem spustila společnost Huawei Cloud protokol R2C (Robot to Cloud). Zhang Ping'an uvedl, že se oficiálně připojilo prvních 20 partnerů protokolu R2C.

Kunpeng Cloud Services: podpora průmyslových inovací díky synergii softwaru a hardwaru a otevřenému ekosystému

Jednou z klíčových strategií společnosti Huawei Cloud je vyvíjet cloudové služby ARM poháněné společností Kunpeng, které poskytují výkon, bezpečnost a spolehlivost. V uplynulém roce se počet výpočetních jader Kunpeng ve službě Huawei Cloud zvýšil z 9 na 15 milionů, tedy o 67 %. Kromě toho byla platforma Kunpeng neustále vylepšována, aby podporovala kompatibilitu s běžným softwarem, a byla také přizpůsobena více než 25 000 aplikacím. Platforma Kunpeng poskytuje spolehlivou podporu pro cloudové služby Kunpeng, které lze vedle překódování, databází, webových aplikací a cloudových telefonů použít i pro obecnější výpočetní scénáře.

GaussDB: budování efektivních a spolehlivých datových základů založených na supernodech a plném sdílení dat (full pooling)

Databáze GaussDB společnosti Huawei Cloud jsou založeny na univerzálních výpočetních supernodech, které umožňují vrstvené sdílení výpočetních, paměťových a úložných zdrojů, a umožňují vícečetné čtení a zápis na libovolném uzlu současně, čímž eliminují omezení tradiční architektury, kdy čtení/zápis dat podporuje pouze primární uzel. Databáze GaussDB také podporují dynamické plánování zátěže, což výrazně zvyšuje výkon souběžného zpracování transakcí. Cluster GaussDB nasazený na základě výpočetních supernodů dokáže zpracovat 5,4 milionu transakcí za minutu, což představuje 2,9násobný nárůst výkonu oproti clusterům bez supernodů.

Cloud distribuovaný cloud všechny scénáře: všudypřítomný a nejlepší možný výpočet s místním přístupem

Společnost Huawei Cloud vytvořila distribuované cloudové řešení pokrývající všechny scénáře, včetně CloudOcean, CloudSea, CloudLake a CloudPond. To pokrývá centrální regiony, hotspoty a okrajové lokality a přináší ucelené služby Huawei Cloud všude tam, kde její zákazníci vyvíjejí podnikatelskou činnost.

Vytvoření snadno použitelné, efektivní a otevřené platformy pro vývoj a provoz agentů

Společnost Huawei Cloud spustila Versatile, agentskou platformu vhodnou pro firemní sektor. Ta má sloužit jako snadno použitelná, efektivní a otevřená platforma pro vývoj a provoz agentů na bázi AI. Díky této platformě budou zákazníci lépe vybaveni k rychlému vývoji agentů AI, kteří vyhovují daným scénářům aplikací.

S využitím Versatile stačí, když uživatelé připraví a zadají dokumenty s popisem činnosti a vývojové diagramy. Po jednoduchém potvrzení lze agenta vygenerovat ve dvou krocích, což výrazně zvyšuje efektivitu generování.

Vedle hlavních projevů na akci přinesla společnost Huawei Cloud celou řadu dalších iniciativ, jako jsou fóra a kulaté stoly. Společnost Huawei Cloud také spolupracuje se zákazníky a partnery na prezentaci široké škály inovativních technologií a postupů v oblastech, jako jsou cloudová infrastruktura, velké modely, databáze, agenti AI a ztělesněná AI, a ukazuje, jak může technologie usnadnit digitální a inteligentní transformaci průmyslových odvětví.

Foto -

https://mma.prnewswire.com/media/2777251/Zhang_Ping_an_Huawei_s_Executive_Director_Board_CEO_Huawei_Cloud.jpg

KONTAKT: Liang Xu, cyril.xuliang@huawei.com

<https://www.ceskenoviny.cz/tiskove/zpravy/huawei-cloud-podpora-podminek-pro-vypocetni-techniku-a-prukopniku-umele-inteligence-v-prumyslovych-odvetvich/2723972>