

Na veletrhu MWC 2024 v Barceloně se představila infrastruktura pro AI s desítkou systematických inovací

27.2.2024 - | PROTEXT

Společnost Huawei Cloud na summitu představila deset inovací pro AI, díky nimž je její infrastruktura optimálním prostředím pro její využívání.

Prezidentka globálního marketingu a prodejních služeb Huawei Cloud Jacqueline Shi ve svém projevu uvedla: *"Huawei Cloud je jedním z nejrychleji rostoucích poskytovatelů cloudových služeb na světě. Ve společnosti Huawei Cloud se snažíme posouvat hranice a přinášet zákazníkům po celém světě špičkové technologie. V posledních letech jsme spustili řadu lokálních cloudových regionů, například v Irsku, Turecku, Indonésii nebo Saúdské Arábii, a poskytli tak zákazníkům snadný přístup k nejvýkonnějšímu cloudu. Díky více než 120 bezpečnostním certifikacím z celého světa se můžete spolehnout, že vaše podnikání a data jsou skutečně v bezpečí. Nejde však jen o technologie. Naším posláním je pomáhat našim partnerům růst spolu s námi, k čemuž nyní slouží také naše programy GoCloud a GrowCloud. Nezapomeňme ani na umělou inteligenci – ta dnes mění celý náš svět, a my stojíme v čele dění. Budujeme pevný cloudový základ pro každého, pro každé odvětví, abychom umožnili akcelarovat inteligenci."*

Dnešní základní modely přetvářejí výrobu, interakce, paradigmaty služeb i obchodní modely tradičních aplikací. Díky nim se umělá inteligence stává novým motorem růstu cloud computingu. A ačkoli je jejich potenciál ohromný, implementace AI v souladu s obchodními cíli vyžaduje systematické inovace. Jak uvedl technický ředitel Huawei Cloud Bruno Zhang: *"Huawei Cloud pomáhá prostřednictvím dvou strategií. 'AI pro Cloud' využívá AI a základní modely k vylepšení vašich uživatelských zkušeností. Výsledkem je revoluce ve vývoji softwaru, produkci digitálního obsahu a dalších oblastech. 'Cloud pro AI' pak umožňuje hladké a efektivní zavádění AI. Architektonické inovace, úložiště s podporou AI a konvergence dat a AI vám umožní trénovat a používat umělou inteligenci lépe, než kdykoli předtím."*

KooVerse: Huawei Cloud disponuje 85 AZ ve 30 regionech napříč více než 170 zeměmi a oblastmi. Tato globální cloudová infrastruktura zahrnující výpočetní výkon, úložiště, síť a zabezpečení snižuje latenci na 50 ms.

Distribuovaná architektura QingTian: Základní modely přinášejí desetinásobný nárůst poptávky po výpočetních zdrojích každých 18 měsíců, což dalece překonává Moorův zákon. Pro vyřešení této výzvy se tato architektura vyvinula nad rámec konvenční primární/sekundární architektury. Architektura QingTian překonává omezení výpočetní techniky, úložišť a sítí pro prvotřídní výpočetní páteř AI s heterogenní, peer-to-peer, full-mesh strukturou.

AI výpočty: Hyperškálovatelná a stabilní služba AI Cloud Service podporuje trénování modelů s biliony parametrů, přičemž tréninkové úlohy mohou nepřetržitě běžet v clusteru na tisícovkách karet po dobu 30 dní, a to po 90 % času. Doba výpadku služby nepřekračuje 10 minut. Poskytuje více než 100 sad funkcí modelů Pangu a 100 adaptovaných velkých modelů s otevřeným zdrojovým kódem ihned připravených k použití.

AI-Native úložiště: Školení modelů vyžaduje kvanta dat a Huawei Cloud tento požadavek zvládá díky trojímu přístupu: Služba paměti EMS ukládá petabajty parametrů s ultrarozsáhlou šířkou pásma

220 TB a ultranízkou latencí až na úroveň mikrosekundy; služba SFS Turbo cache pro vysokou propustnost a souběžnost desítek milionů IOPS umožňuje přípravu (warm-up) jedné miliardy datových záznamů za pouhých pět hodin namísto 100; a služba Object Storage Service (OBS) knowledge lake snižuje náklady na ukládání tréninkových a inferenčních dat o 30 %.

E2E zabezpečení: Kompletní životní cyklus zahrnuje prostředí pro chod modelů, tréninková data, samotné modely, generovaný obsah a aplikace. To vše společně zajišťuje robustní, bezpečné a kompatibilní modely a aplikace.

GaussDB: Tato databáze nové generace se vyznačuje vysokou dostupností, zabezpečením, výkonem, flexibilitou a inteligencí a také jednoduchým a inteligentním nasazením a migrací. Konkrétně její distribuovaná architektura na firemní úrovni zajišťuje vysokou dostupnost díky nulovému RPO v rámci dvou clusterů, úplné izolaci softwarových a hardwarových závad a nulovému výpadku služeb. V oblasti zabezpečení je certifikována na nejvyšší možnou úroveň CC EAL4+. Co se týče automatizace, GaussDB zdokonaluje jako první AI-native databáze na světě migraci, nasazení a přesun databází.

Konvergence dat a AI: Vzhledem k boomu základních modelů se z přístupu "data + AI" dnes stává "data pro AI a AI pro data" (Data4AI, AI4Data). Huawei Cloud LakeFormation sjednocuje datová jezera z více jezer nebo skladů, což umožňuje sdílet jednu kopii dat mezi více analytickými enginy a AI enginy bez nutnosti migrace dat. Tři kooperativní procesy (pipelines) - DataArts, ModelArts a CodeArts - pak koordinují a plánují datové a AI pracovní postupy (workflows). Současně řídí online trénování modelů a interferenci s daty v reálném čase. EngineAI4data přináší inteligentnější správu dat - od integrace a vývoje po řízení kvality a aktiv.

Mediální infrastruktura: Pro současnou éru AIGC a 3D internetu vybudovala společnost Huawei Cloud mediální infrastrukturu charakterizovanou efektivitou, uživatelským zážitkem a vyspělostí. Efektivitu ztělesňuje systém pro produkci obsahu Huawei Cloud MetaStudio, který obsahuje mj. Workspace a AIGC (AI generované) virtuální postavy pro rychlejší a kvalitnější tvorbu. Co se týče uživatelské zkušenosti, o hladší a dokonalejší interakce se stará Huawei Cloud Live, Low Latency Live a SparkRTC. Vyspělé prvky představuje AIGC (Artificial Intelligence Generated Content, AI generovaný obsah) a 3D služby s interakcí v reálném čase. To vše společně posouvá obchodní možnosti a uživatelský zážitek na vyšší úroveň.

Landing Zone: Jednotná správa účtů, identit, oprávnění, sítí, compliance a nákladů umožňuje společností v prostředí Huawei Cloud lepší využívání a správu zdrojů. Multitenance a spolupráce personálu, financí, zdrojů, povolení a compliance s bezpečnostními předpisy je nyní zcela bezproblémová.

Flexibilní nasazení: Všechny výše uvedené funkce a služby modelů Pangu mohou fungovat ve veřejném, specializovaném nebo hybridním cloudu. Zákazníci mohou například vytvářet a provozovat dedikované AI platformy a základní modely ve svých stávajících datových centrech pomocí hybridního cloudového řešení Huawei Cloud Stack.

Veletrh Mobile World Congress (MWC) 2024 se koná v Barceloně od 26. do 29. února. Společnost Huawei Cloud je zde připravena prezentovat spolu se svými zákazníky a partnery širokou škálu zajímavých témat. Kromě toho se během akce představí inovativní produkty a reálné projekty využívající modely Pangu, GaussDB, konvergenci data-AI, virtuální postavy a vývoj softwaru.

Foto -

https://mma.prnewswire.com/media/2346836/Jacqueline_Shi_President_Huawei_Cloud_Global_Marketing_Sales_Service.jpg

Foto - https://mma.prnewswire.com/media/2346837/Bruno_Zhang_CTO_of_Huawei_Cloud.jpg

Video - https://mma.prnewswire.com/media/2346838/Ten_Huawei_Cloud_Innovations_EN.mp4

KONTAKT: cyril.xuliang@huawei.com

<https://www.ceskenoviny.cz/tiskove/zpravy/huawei-cloud-na-veletrhu-mwc-2024-v-barcelone-se-preds-tavila-infrastruktura-pro-ai-s-desitkou-systematickych-inovaci/2485161>