

Pohrana podataka za AI i ve

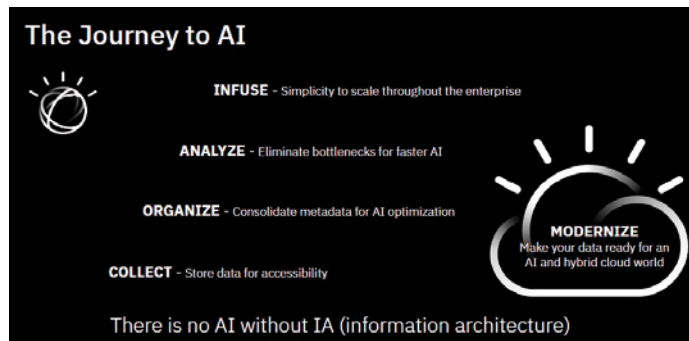
Umjetna inteligencija (Artificial Intelligence) predstavlja cijelo jedno putovanje, čija ishodišna točka počinje podacima. Podaci su gorivo za AI, a AI ne može postojati bez IA (informacijske arhitekture). Razne organizacije, tvrtke i općenito industrije, često se susreću s raznim izazovima pri izvlačenju analiza i uvida iz vlastitih podataka

Vedran Podubski, konzultant, Megatrend poslovna rješenja

Silos s podacima (Data silo) otežavaju cjelovit uvid u sve podatke, ograničavajući vrijednost AI-a. Neke infrastrukture nisu izgrađene, točnije, prilagođene za potrebe korištenja umjetne inteligencije pa samim time nisu dovoljno fleksibilne da podrže u potpunosti nove zahtjeve poslovanja, bez dodavanja određene složenosti u cijeli proces.

Trenutačni izazovi s kojima se suočavaju podatkovno vođene organizacije potreba su za unosenjem i analizom (ingest and analyze) eksponencijalno sve veće količine podataka u sve kraćim vremenskim intervalima, kao i potreba za kreiranjem sve složenijih modela i simulacija, gdje su često potrebni djelotvoran uvid i odgovori u stvarnom vremenu. Kako bi se udovoljilo zahtjevima tih trenutačnih i novih izazova s podacima, potrebno je učinkovito povećati podatkovne centre i općenito infrastrukturu koja se nalazi u pozadini navedenih alata.

Prema raznim vanjskim studijama, jedan od najčešćih problema koji se predviđi prilikom kreiranja



AI i rješenja za pohranu velikih podataka

sustava za umjetnu inteligenciju dio je infrastrukture za pohranu podataka (data storage) na kojem se stvara tzv. "bottleneck", te je baš zbog toga IBM kreirao svoje specijalizirane sustave za pohranu podataka kod sustava za umjetnu inteligenciju i "big data".

Kao jedan od lidera na Gartnerovom magičnom kvadrantu za distribuirane datotečne sustave i pohranu objekata za 2020. godinu (2020 Gartner Magic Quadrant for Distributed File Systems and Object Storage), sa svojim sustavima pohrane IBM pruža najveću skalabilnost, kapacitet

i propusnost te jednu od najbogatijih funkcionalnosti za velike (enterprise) sustave s testiranim količinama podataka u razini egzabajta (exabyte).

U portfoliju rješenja za pohranu podataka za umjetnu inteligenciju i "big data" sustave možemo izdvojiti sljedeće: IBM Spectrum Scale; IBM Elastic Storage System (ESS) 5000 i 3000; te IBM Cloud Object Storage (COS).

IBM SPECTRUM SCALE

Dio IBM Spectrum Storage™ obitelji rješenja, IBM Spectrum Scale paralelni je datotečni

sustav poslovnog razreda koji pruža vrhunsku elastičnost, skalabilnost i kontrolu. Zasnovan na IBM-ovom općenitom paralelnom datotečnom sustavu (GPFS™), IBM Spectrum Scale pruža skalabilni kapacitet i performanse za rukovanje zahtjevnom analitikom podataka, spremištima sadržaja i ostalim radnim opterećenjima.

Omogućuje objedinjavanje slučajeva virtualizacije, analitike, datoteka i objekata u jedinstveno rješenje za pohranu podataka. Pruža jedinstvenu unificiranu platformu, nudeći jednu točku upravljanja.

IBM ELASTIC STORAGE SYSTEM (ESS)

IBM Elastic Storage System (IBM ESS) implementacija je softverski definiranog sustava pohrane podataka (software-defined storage), olakšava postavljanje brze, visoko skalabilne pohrane za AI i "big data". Uz brzu NVMe tehnologiju i vodeće mogućnosti upravljanja datotekama, u kombinaciji sa softverom IBM Spectrum Scale, IBM ESS 3000 i 5000 mogu narasti do preko jotabajta (yotabyte – YT) s mogućnošću integracije u ostale sustave za pohranu podataka.

U srcu oba sustava IBM ESS 3000 i 5000 leži softver IBM Spectrum Scale, koji osigurava jedinstveni sustav datoteka, pojednostavljuje upravljanje pohranom i pruža visoke performanse.

IBM SPECTRUM SCALE

Neke od prednosti IBM Spectrum Scalea koje se mogu navesti su:

Nova razina performansi i učinkovitosti pohrane

- Dramatična poboljšanja u I/O izvedbi
- Znatno povećana brzina protoka podataka među čvorovima
- Poboljšane performanse za radna opterećenja s promjenjivom veličinom bloka
- Poboljšane performanse rada metapodataka

Jednostavnija i moćnija administracija sustava

- Brže i jednostavnije iskustvo s podešavanjem za optimalne performanse
- Poboljšane GUI značajke

Sigurnosna poboljšanja za upravljanje osjetljivim podacima

- Nova sposobnost evidentiranja revizije datoteka za praćenje pristupa korisnika datotečnom sustavu i ostalih događaja
- Analizirani podaci pohranjeni u zaštićenom skupu datoteka
- REST sučelje treće generacije API-ja s modernim, visoko automatiziranim upravljanjem i daljinskom kontrolom klastera
- Proširena podrška za "Transparent Cloud Tiering" s boljom granularnošću podataka, s više računava u oblaku i više kontejnera
- Bolje sigurnosno kopiranje / vraćanje putem nove podrške za *snapshot* te bolje performanse load balancinga
- Brža implementacija u okruženjima IBM Elastic Storage Server

elike sustave Big data

IBM ESS 3000

Secure access. Anywhere.
Enterprise data services. Everywhere
Hybrid cloud. Anyone

Use Cases	AI / ML	IBM Cloud Paks	Cloudera/ Hadoop	HPC	Video & Images	Big Data Analytics	NVIDIA	Data Collaboration
Integration	IBM, SAS, Splunk, Radar, Spark	Red Hat OpenShift	NVIDIA, Cloudera, SAS, Aspera					
Organize	IBM Spectrum Discover							
Hybrid Cloud File System	IBM Spectrum Scale AWS Spectrum Scale IBM Cloud Spectrum Scale OpenShift Red Hat OpenShift Container Platform ESS 5000 ESS 3000							
Edge Data Center Private Cloud Public Cloud								

IBM ESS sustav

IBM CLOUD OBJECT STORAGE (COS)

IBM Cloud Object Storage vodeće je softverski definirano visokoskalabilno i isplativo rješenje za pohranu podataka u podatkovnom centru, privatnom ili javnom oblaku. Koristi se za modernizaciju infrastrukture za AI, analitiku, IoT, itd. Smanjuje troškove pohrane i do 12%, koristeći nove diskovne pogone od 18 TB i povećava protok do 55 GB / sa za klaster od 12 čvorova. Lako se spaja na bilo koji Windows ili Linux sustav datoteka, koristeći SMB ili NFS pristup.

IBM Cloud Object Storage može se kupiti samo kao softver, kao uređaj (appliance) ili kao privatni ili javni oblak u IBM Cloudu. Ne zahtijeva replikaciju, već koristi algoritme za širenje informacija na geodisperzirani sustav, raspoređen u tri ili više podatkovna centra.

Neke od prednosti IBM Cloud Object Storagea koje se mogu navesti su: vodeća platforma za pohranu objekata u industriji, na prvom mjestu prema nezavisnim ocjenama, jednostavno pokretanje s 3 čvora i jednostavno skaliranje, izvorni S3 API s preko 100 potvrđenih početnih aplikacija, patentirana lokalna i geodisperzirana tehnolo-

IBM ESS 3000 nudi:

- Jednostavnost: kontejnerizirana instalacija i nadogradnja softvera, moćan GUI za upravljanje
- Visoke performanse: NVMe *flash* pohrana podataka s protokom od 40 GB / sekundi
- AI optimizacija: Pohrana visokih performansi za maksimiziranje performansi preko GPU-a
- Operativna učinkovitost: Gusta pohrana podataka (dense storage) unutar 2U racka smanjuje iskorištenje prostora podatkovnog centra.

Nudi vrlo guste mogućnosti pohrane, smještajući do 24 NVMe pogona od 15,4 TB (ukupno 369 TB) u 2U kućištu za vrhunske ukupne troškove vlasništva (TCO), zahvaljujući smanjenim potrebama za hlađenjem i snagom.

IBM ESS 5000 samostalni je poslužitelj za pohranu, kapaciteta u petabajtima (PB) sa softverski definiranom pohranom (IBM

Spectrum Scale) na bazi procesora IBM POWER9 za pohranu podataka u masivne "Data Lakeove". Omogućuje integraciju s prethodnim IBM ESS modelima i nudi opcije s nižim troškovima, poput pohrane u oblaku.

Neke od prednosti IBM ESS 5000 poslužitelja su:

- Instalacija u nekoliko minuta s kontejneriziranim softverom
- Stvaranje jedinstvenog "namespacea" u jotabajt (YB) kapacitetima
- Performanse do 55 GB / s u jednom sustavu
- Gustoća od 370 TB / u jednom kućištu veličine 36U
- Integriran IBM Spectrum Discovery softver
- Dvije osnovne konfiguracije od 552 TB do 13,5PB



investment), povrat investicije u roku od 8 mjeseci, mogućnost zaključavanja WORM (Write Once, Read Many) segmenata, ugrađena enkripcija za vrijeme rada i mirovanja, dostupnost do "8 devetki" (99,99999999%) i pouzdanost od 15, višestruki uspješno odrađeni projekti i implementacije na svjetskoj razini. Ako želite smanjiti troškove učinkovitijom, bržom obradom podataka i analitikom s jedinstvenim setom proizvoda i kvalitetnijim utroškom resursa, podići produktivnost, modernizirati vlastito poslovanje korištenjem metodologije "AI ljestvice", ili doznati više podataka, javite se našim stručnjacima na adresu poslovanje.rjesenja@megatrend.com

2020 Gartner Magic Quadrant for Distributed File Systems and Object Storage