PosAm

Dôveryhodné úložisko
Trusted Storage

Peter Smák
produktový manažér PosAm

Pretvárame informačné technológie na úžitok pre zákazníkov

Agenda

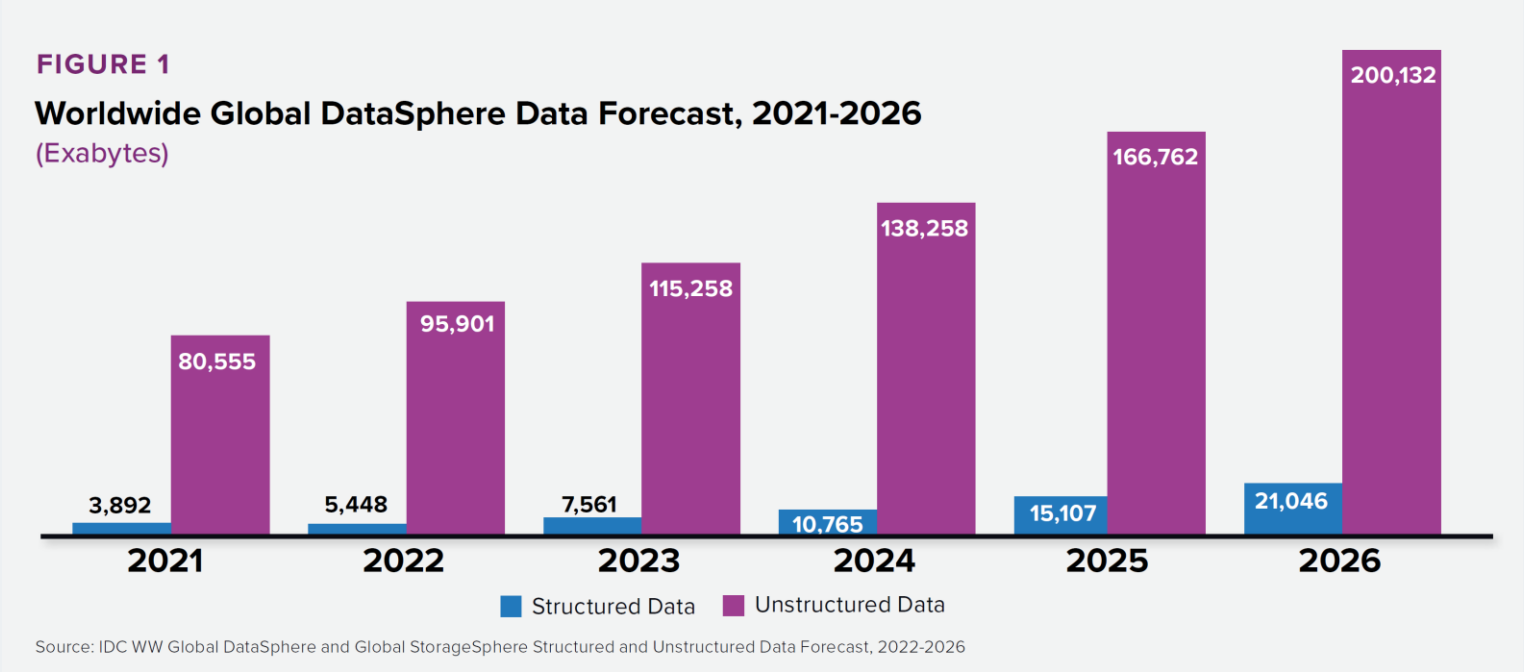
- Trendy v oblasti neštruktúrovaných dát
- Regulácie a uchovávanie neštruktúrovaných dát
- Dôveryhodné úložisko a jeho vlastnosti
- Scenáre možného použitia

PosAm

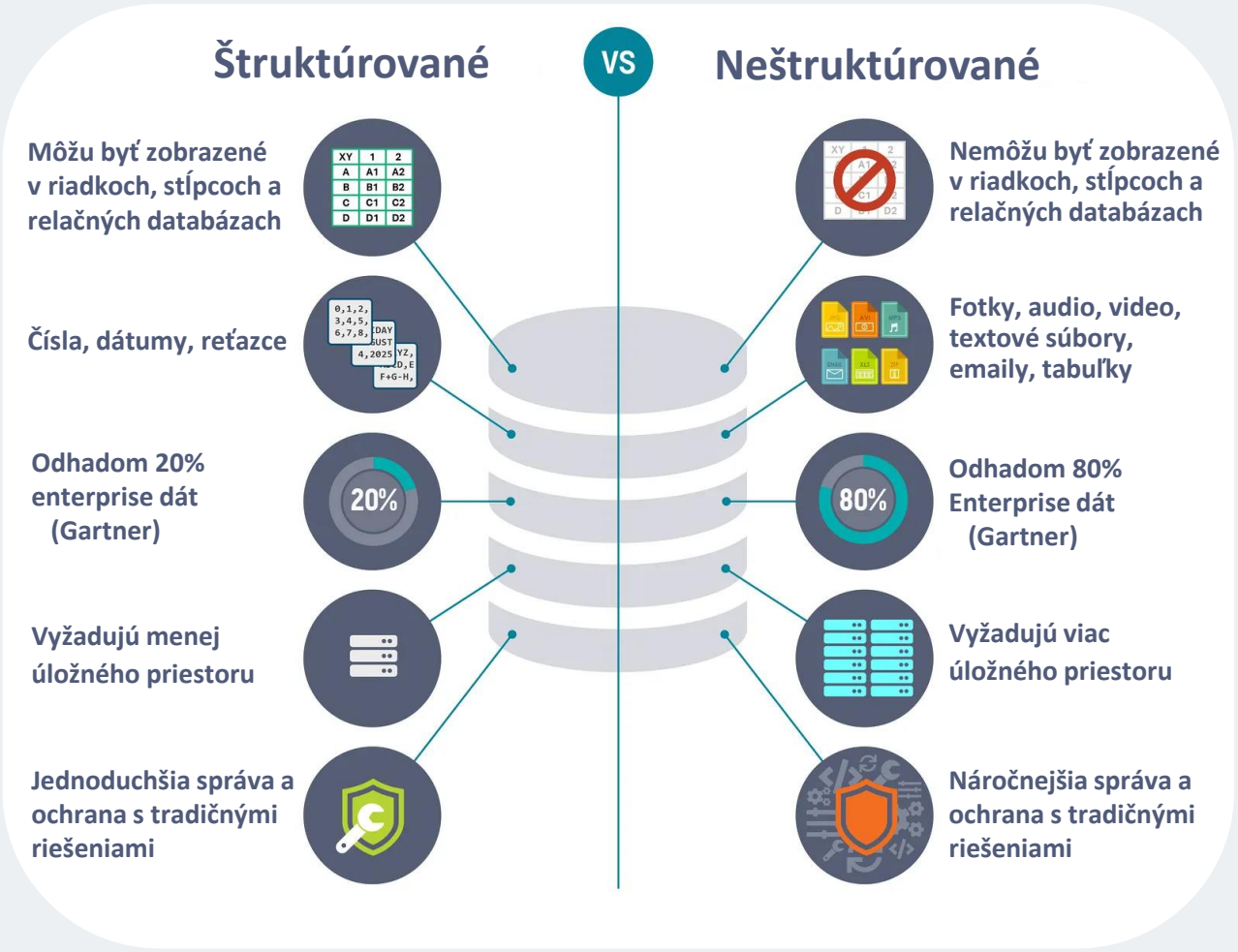
Edge-to-Core-to-Cloud Infrastructure
Hitachi Vantara
PARTNER

Čo spôsobuje rast neštruktúrovaných dát

- Rozmach internetu a konektivity
- Rast počtu užívateľov, zariadení a veľkosti dát
- Internet of Things (IoT), mobilné aplikácie
- Sociálne siete a marketing, okamžité posielanie správ
- Nové technológie (senzory, snímkovanie...)
- Digitálna transformácia
- Potreba rozhodovania založeného na údajoch (BI)
- Inovácie v oblasti AI



Čo sú to neštruktúrované dáta



Vytvorené človekom

- Textové súbory a emaily
- Súbory kancelárskych aplikácii (pdf, doc, xls...)
- Mobilná komunikácia: SMS, IM, chat, audio a video správy
- Príspevky či tweety na sociálnych sieťach
- Médiá: digitálne fotografie, zvukové nahrávky, videosúbory a pod
- Obsah webových stránok

Strojovo generované

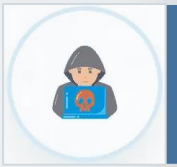
- IoT dáta - senzorické dáta, dáta aplikácii a logy
- Dohľadové systémy - napr. kamerové
- Snímkové údaje – meteo, geo, satelitné, medicínske
- Vedecké dáta (prieskum Zeme, seizmické, atómové...)

Trendy a výzvy v oblasti neštruktúrovaných dát (podľa analytikov)

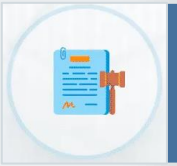
Trend č. 1: Hromadenie neštruktúrovaných údajov núti mnohé organizácie zvážiť nové prístupy k riadeniu rastu údajov, životného cyklu údajov a súvisiacich nákladov

Trend č. 2: Umelá inteligencia a ML budú analyzovať trendy a vlastnosti údajov s cieľom zabezpečiť proaktívnu (a prípadne automatizovanú) správu neštruktúrovaných údajov

Trend č. 3: Dátové silo efekty a stále prevažujúce ukladanie neštruktúrovaných údajov v systémoch NAS



Trend č. 4: Zvýšená zraniteľnosť voči kybernetickým útokom



Trend č. 5: Právne obmedzenia si vyžadujú odlišné zaobchádzanie s rôznymi typmi údajov

Príklady údajov, u ktorých spôsob digitálneho uchovávania podlieha reguláciám

Verejný sektor

- Údaje vznikajúce pri poskytovaní elektronických služieb štátu
- Údaje vznikajúce v konaniach pred štátnymi, regionálnymi alebo miestnymi orgánmi
- Citlivé údaje obsahujúce kritické, osobné, súkromné, dôverné, utajované alebo inak privilegované skutočnosti
- Záznamy trvalých kultúrnych a historických aspektov – národné archívy a kultúrne dedičstvo
- Digitálne registre štátu
- Geografické dáta – mapy, letecké snímky...

Finančný sektor

- Elektronická kontraktácia (na diaľku), plnenie informačných povinností
- Záznamy z obchodných vzťahov a operácii
- Záznamy z finančných operácii, pri obchodovaní s cennými papiermi, z kolektívneho investovania

Zdravotnícky priemysel

- Medicínske informácie – zdravotné údaje, genetické, biometrické, diagnostické, údaje zo sekvenovania genómov, patologické, laboratórne údaje...

Ostatné sektory

- Preukazovanie integrity a pravosti údajov pred súdmi, regulačnými orgánmi alebo pri predkladaní podkladov príslušným orgánom, právna dokumentácia
- Na preukázanie vlastníckeho práva, duševného vlastníctva, zmluvných alebo iných právne vymožiteľných práv a povinností
- Záznamy o platobných operáciách
- Účtovné a daňové informácie, podnikové výkazníctvo, záznamy o zamestnancoch, dávkach, sociálnom a dôchodkovom zabezpečení
- Záznamy, u ktorých spôsob uchovávania údajov podlieha lokálnej legislatíve a európskym smernicam ako GDPR

Dôveryhodné úložisko

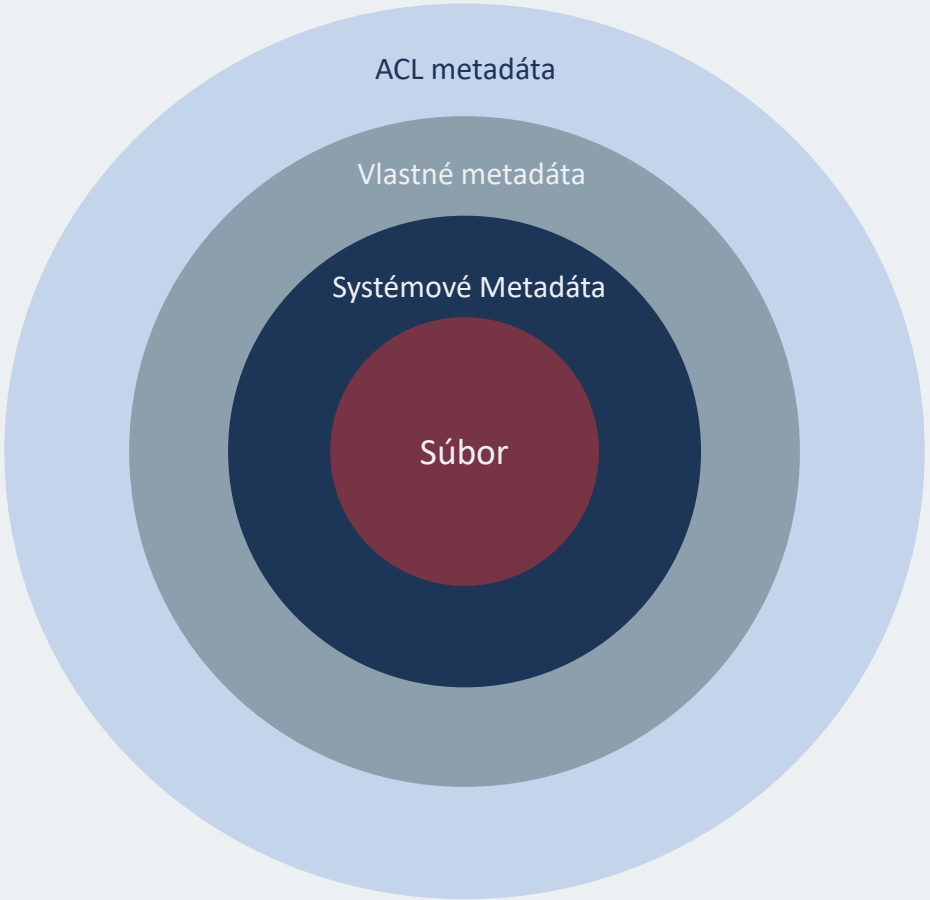
Dôveryhodné úložisko je riešenie **aktívnej archivácie**, ktoré umožňuje organizáciám ukladať, chrániť, uchovávať a vyhľadávať **veľké objemy neštruktúrovaného obsahu** v rámci **jednotnej online platformy úložiska**. Je vhodné najmä na uchovávanie citlivého obsahu, ktorý podlieha **zabezpečeniu súladov s predpismi a bezpečnostnými požiadavkami**.

Voľba technológie pre Dôveryhodné úložisko

	Block	File	Object
Garancia nemennosti údajov (hardvérovo)	☹	😐	😀
Manažment politík na úrovni údajov	☹	😐	😀
Inteligentné vyhľadávanie prostriedkami úložiska	☹	😐	😀
Náklady pri veľkých objemoch dát	☹	😐	😀

Čo je to Objekt

Súbor + Metadáta + Nastavenia politík k objektu + ACL Metadáta = Objekt



Systémové Metadáta

- 28 systémových vlastností
- Nastavenia politík k objektu
- POSIX Metadáta

Vlastné Metadáta

- Pridanie relevantných informácií k objektu (anotácie) a vytvorenie asociácii so súvisiacimi objektmi
- Metadáta objektov (XML) sú indexované do škálovateľnej databázy, v ktorej možno vyhľadávať

ACL Metadáta

- oprávnenia a prístupy k údajom na úrovni jednotlivých objektov pre používateľov alebo skupiny používateľov

Vlastné metadáta

Source Object

Annotation5
Annotation10
Annotation4
Annotation9
Annotation3
Annotation8
Annotation2
Annotation7
Annotation1
Annotation6

```
<xml>
 <data0>123</data0>
 <data1>abc</data1>
 <data2>xyz</data2>
ml>
```

system metadata

name
date created
owner
size
date modified
path

Medical (DICOM)

```
<record>
<doctor>
<name>John Smith</name>
</doctor>
<patient>
<name>John Smith</name>
<age>48</age>
</patient>
</record>
```

Image (EXIF)

```
<taken>11/17/12</taken>
<aperture>5</aperture>
<ISO>400</ISO>
```

Billing

```
<cost>$1,500</cost>
<insurance>yes</insurance>
```

Súbor + Metadáta = Objekt

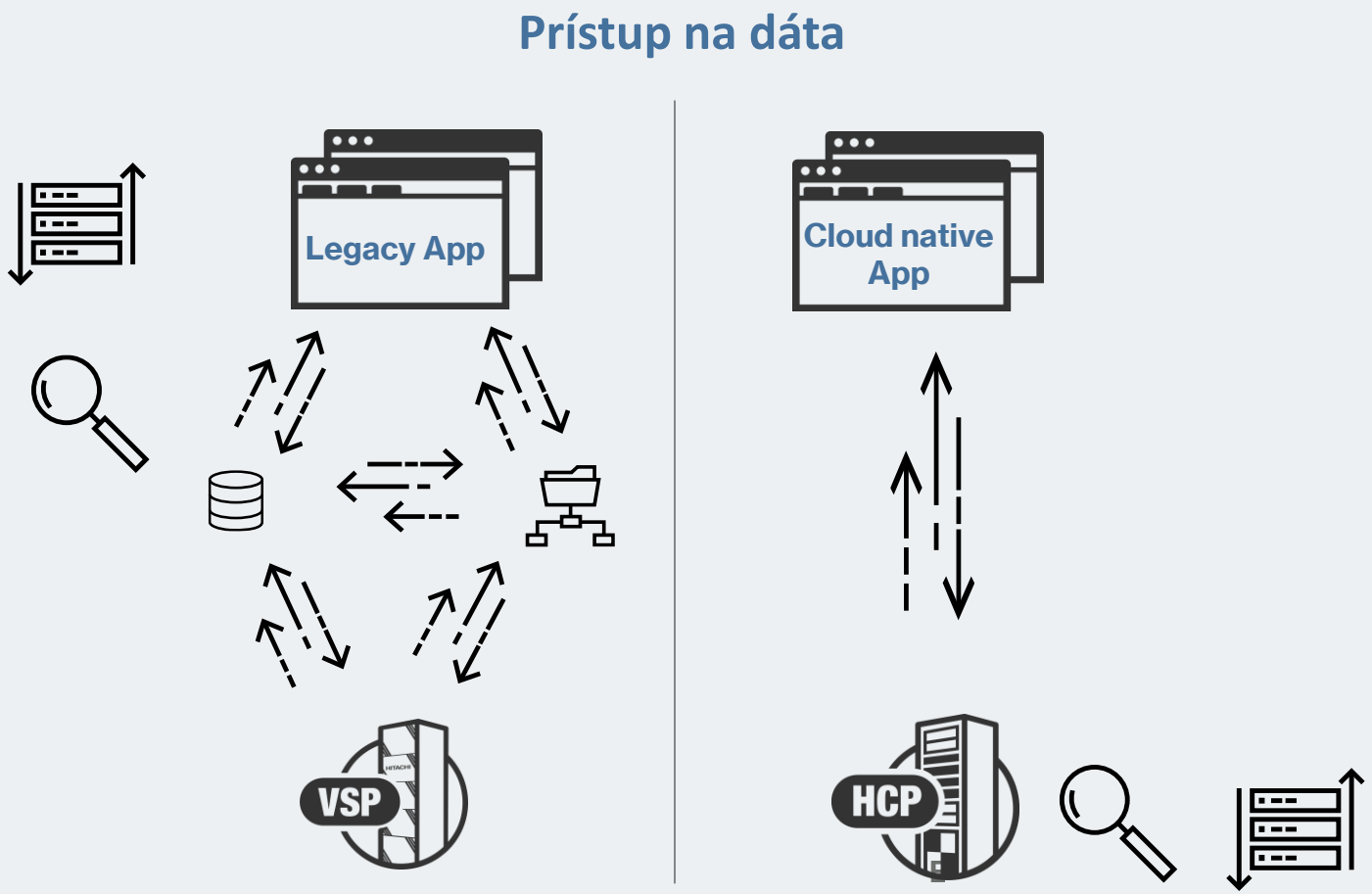
Podpora aplikačných protokolov na HCP

Prístup na dáta

- **RESTful API** (resp. REST API for HCP)
- **Amazon S3 API**
- NFS/CIFS/SMTP/WebDAV*
- OpenStack SWIFT API
- Kubernetes (CSI)

REST API for HCP – až 10 anotácii / objekt, rozšírený náhľad do fyzickej infraštruktúry HCP, do politík uchovávania, multitenancie, vyhľadávania a metadát.

Pre **S3 API** ponúka HCP vlastné hlavičky, aby S3 aplikácie mohli využívať compliance funkcie HCP, ktoré nie sú súčasťou špecifikácie S3 API.



Vlastnosti Dôveryhodného úložiska – Splnenie regulačných požiadaviek

Autenticita údajov

- Zapisované dáta sú presnou digitálnou kópiou zapísaného súboru (digitálny odtlačok na báze hash algoritmu)
- Digitálny odtlačok je trvalo spojený s objektom – HCP ho počas životného cyklu pravidelne porovnáva s pôvodnou hodnotou

Nemennosť údajov

- Funkcionalita WORM (Write-Once-Read-Many)
- Zmenám alebo zmazaniu zabraňuje aj ochrana verzii objektov

Nezmazateľnosť údajov

- Režim **Compliance** – objekty nie je možné odstrániť žiadnym mechanizmom a nie je možné zmazať parametre uchovávania alebo skrátiť dobu uchovávania.
- Režim **Enterprise** - objekty môže odstrániť len oprávnený používateľ, pričom môže tiež zmazať parametre uchovávania alebo skrátiť dobu uchovávania.

Trvanlivosť integrity údajov

- Mechanizmy samo-opravy dát počas životného cyklu – zabránenie degradácii
- Digitálne odtlačky sa porovnávajú s replikami údajov
- Ak je objekt prečítaný aplikáciou a aplikácia ho odošle znovu na uloženie, jeho hash HCP znovu prepočíta a porovná

Politiky uchovávania na úrovni objektov

- Doba uchovávania
- Privilegované vymazanie (GDPR)
- Legal Hold (právne podržanie)
- Verzionovanie
- Indexovanie
- Skartovanie (v súlade so špecifikáciou DoD 5220.22-M)

Vyhľadávanie údajov

- Ako súčasť dátového manažmentu
- Ako súčasť právneho alebo auditného procesu
- Podpora pre 370 súborových formátov a 77 jazykov
- Vstavaný Metadata Query Engine (konzola alebo API)

Auditovateľnosť

- Úplné a komplexné monitorovanie a audit všetkých udalostí počas životného cyklu informácii
- Sledovanie objektov (object tracking) a zaznamenávanie udalostí (event logging) vrátane všetkých akcii zmazania objektov
- HCP vie generovať auditné prehľady

Vlastnosti Dôveryhodného úložiska – Bezpečnosť Dát

DPL/MDPL – Data & Metadata Protection Level

- Redundancia pre kritické alebo vysoko hodnotné údaje
- Všetky kópie údajov pre objekty sú uložené vo všetkých uzloch v „ochranných setoch“
- Každý HCP klaster umožňuje ukladať až 4 repliky pôvodného dátového objektu (2 default) a až 2 repliky metadát
- V prípade poškodenia je ich možno ich obnoviť pomocou druhej kópie

Šifrovanie

- Viac-úrovňové šifrovanie (DEK / KEK)
- Integrácia s externými KMS riešeniami (KMIP API 1.4 alebo novším)
- Šifrovací kľúč sa generuje v čase inštalácie systému a je distribuovaný medzi uzly HCP (nutné kvórum)
- HCP šifruje automaticky in-line – bez potreby meniť prístupy alebo procesy
- Všetky metódy šifrovania sú podľa AES-256 (súlad s FIPS 140)
- Kľúče nie sú nikdy uložené v HCP a nie sú prístupné správcovi HCP

Šifrovanie komunikácie

- IP filtrovanie a SSL pre HTTP prístupy (REST, S3)
- Manažment prístupov k IP adresám
- Každá prístupová brána HCP má svoje bezpečnostné mechanizmy
- Vstavaný firewall pre nepoužité porty
- Vzdialená správa cez SSH
- Administrácia cez samostatné porty alebo VLAN

Overovanie používateľov (RBAC) a Access Control List (ACL)

- Lokálne používateľské účty
- Integrácia s MS Active Directory a RADIUS
- ACL metadáta - riadenie prístupu k údajom na úrovni jednotlivých objektov

Vlastnosti Dôveryhodného úložiska – Multitenantná architektúra

Logické oddelenie správy, údajov a politík

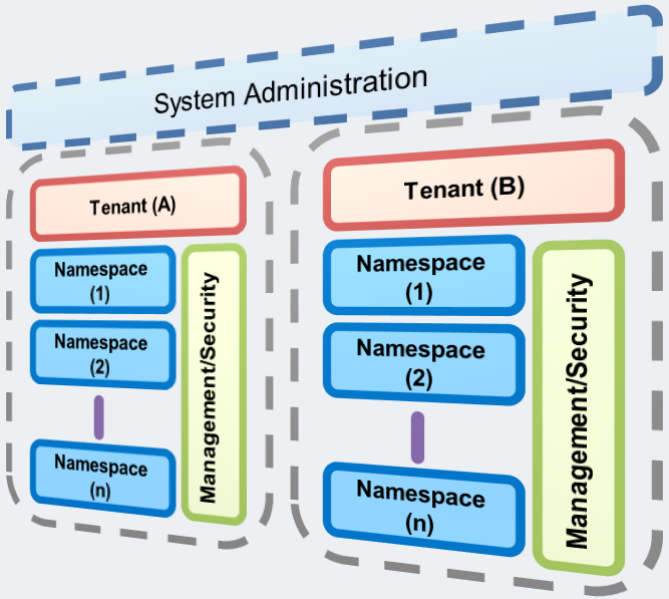
Tenant

- Oddelenie na úrovni správy
- Typicky organizačná jednotka alebo workload – virtuálne HCP
- Až 1000 tenantov / HCP

Namespace

- Oddelenie priestorov pre údaje – až 10 000 namespace / HCP
- Oddelenie údajov uložených - aplikáciami, obchodnými jednotkami, používateľmi alebo zákazníkmi.
- Objekty uložené v jednom namespace nie sú viditeľné v žiadnom inom namespace

Chargeback – HCP vie generovať prehľady za spotrebovanú kapacitu - užitočné napr. pri využívaní prostriedkov úložiska rôznymi organizačnými jednotkami



Example of Namespace adress
https://t1n1.t1.hcp.company.sk
https://t1n2.t1.hcp.company.sk
https://t2n3.t2.hcp.company.sk

Example of object representation
https://t5n1.t5.hcp.company.sk/rest/images/image1.jpg
https://t5n1.t5.hcp.company.sk/hs3/images/image1.jpg

Vlastnosti Dôveryhodného úložiska – Škálovateľnosť

Kľúčové atribúty

- Výkonnostná architektúra HCP je schopná adresovať požiadavky malých aj veľkých objektov
- Multilokalitné riešenie - až 80 uzlov, 160 rackov, >1EB na lokalitu, >6EB (Geo-EC), >100B objektov na systém
- Nezávislé škálovanie všetkých vrstiev a bezvýpadkové rozširovanie

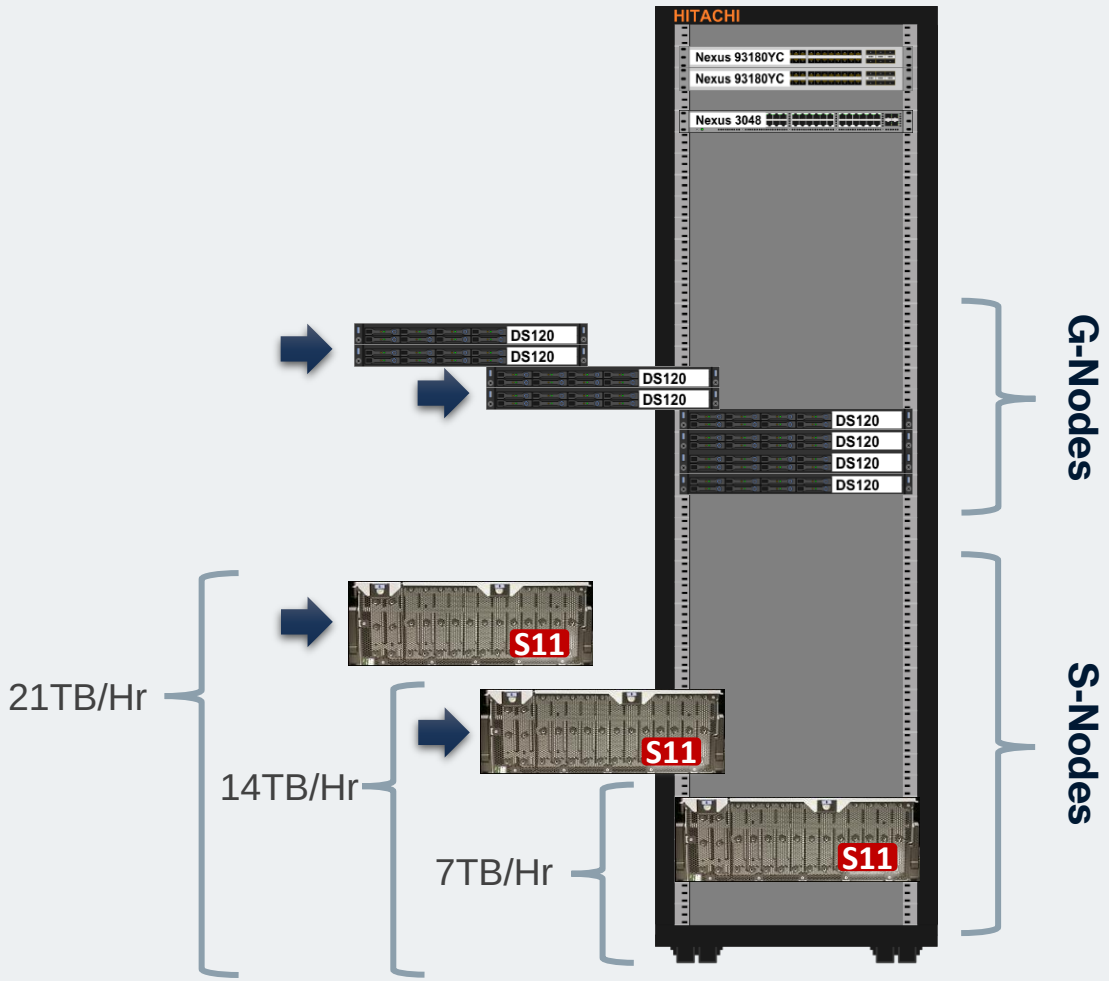
Prístupové G-uzly

- Klaster (min. 4 uzly)
- Spracovanie I/O od klientov, služby pre protokoly, ukladanie metadát, indexov, beží tu vstavaný search engine a OS
- Automatické rozkladanie záťaže na všetky prístupové uzly
- Môže slúžiť aj na ukladanie objektov (interná kapacita s RAID6)

Úložné S-uzly

- Úložná kapacita pre objekty a metadáta
- S11: max 3.2PB per node
- S31: max 15.1PB per node

Ďalšie vrstvy na:
Vyrovnávanie záťaže – Load Balancers
Dedikované Search uzly na vyhľadávanie – Hitachi Content Intelligence



Vlastnosti Dôveryhodného úložiska – Odolnosť

1. Na úrovni údajov

- Redundantné kópie údajov a metadát (DPL / MDPL)
- Overovanie obsahu pomocou hashov a automatická oprava objektov
- Ochrana dát pred zmazaním – WORM a verzionovanie

2. Na úrovni komponentov

- Redundantná HW architektúra, RAID6 na G-uzloch, EC (20+6) na S-uzloch
- HCP dáta zostávajú na S-uzloch dostupné aj pri výpadku 6 diskov naraz
- HCP dáta sú dostupné pri výpadku celého G-uzla

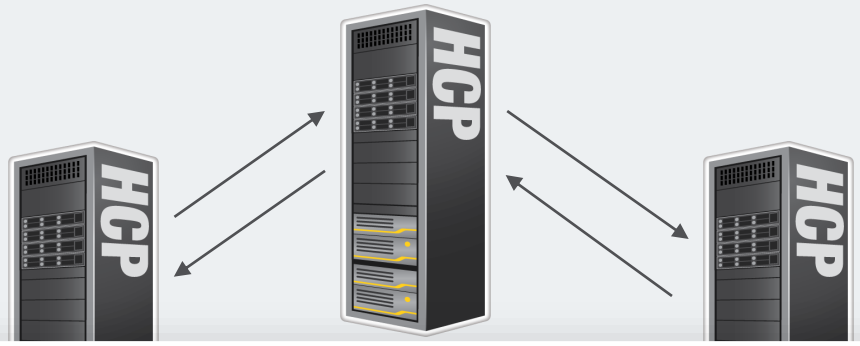
3. Na úrovni celej lokality (viac-lokalitné riešenie

- Automatická oprava údajov pomocou replík zo záložných lok
- Topológia globálneho prístupu - obojsmerná async repl active, pri ktorej sú systémy HCP synchronizované spôsobo používateľom a aplikáciám prístup k údajom z najbližšej lok

Geodistribuovaný Erasure Coding

- Ochranný mechanizmus kódovania dát
- Minimálne 3 lokality, maximálne 6
- Jeho benefitom je zvyšenie efektívnej kapacity až o 40%

Bezvýpadková (online) migrácia dát na novú generáciu úložiska bez vplyvu na existujúce aplikácie a so zachovaním pôvodných metadát objektov



Výsledkom je mimoriadna ochrana údajov:

Prístupnosť HCP je 99.99999999% (10 deviatok) = nedostupnosť 3ms ročne

Trvanlivosť dát na HCP je 99.9999999999999% (15 deviatok) = milión x lepšia ako AWS S3

HCP nie je nutné zálohovať

Dostupnosť = prístupnosť (môžete sa dostať k svojej aplikácii a k svojim údajom)
+ trvanlivosť (či sú údaje neporušené a konzistentné)

Capacity Savings	0%	25%	33%	37%	40%

PosAm

Vlastnosti Dôveryhodného úložiska – Sumarizácia

Hitachi Content Platform spĺňa všetky atribúty dôveryhodného úložiska:

- ✓ Garancia autenticity, nemennosti, nezmazateľnosti a trvalej integrity uchovávaných informácii
- ✓ Schopnosť nastavovať politiky ako doby uchovávania, právneho podržania, verzionovania a skartovania na úrovni objektov
- ✓ Rýchle vyhľadávanie údajov v rámci právnych alebo auditných procesov vďaka metadátam
- ✓ Ochrana citlivých dokumentov pomocou pokročilých techník šifrovania, vstavané schémy šifrovania dát s integráciou na externé KMS
- ✓ Každá činnosť je v systéme je plne kontrolovateľná – auditovateľná
- ✓ Multitenantná architektúra - pobočky, užívatelia, aplikácie a dáta majú vlastné „virtuálne“ HCP
- ✓ Takmer neobmedzená škálovateľnosť, nezávislé a bezvýpadkové rozširovanie a údržba všetkých HW vrstiev
- ✓ Bezvýpadková migrácia na novú generáciu úložiska bez vplyvu na aplikácie / používateľov a so zachovaním pôvodných metadát
- ✓ Mimoriadna ochrana údajov a odolnosť, ktorá eliminuje potrebu zálohovania

Hitachi Content Platform – ďalšie možnosti HCP ekosystému

HCP Anywhere Enterprise

- Smart Cache riešenie Edge úložísk pre SMB/NFS pobočky s protokolovým mostíkom do centrálnych S3 HCP
- Globálne dostupný súborový systém s End-to-End bezpečnosťou
- Privátne zdieľanie a synchronizácia citlivých dokumentov medzi pobočkami

HCP Content Monitor

- Rozširujúce riešenie na monitorovanie HCP úložísk
- Analýza trendov v podobe grafických vizualizácii s cieľom zlepšiť plánovanie výkonu a kapacity
- Prispôsobenie monitorovania metrík, aby boli relevantné pre potreby organizácie

HCP Content Intelligence

- Automatizuje extrakciu, klasifikáciu, obohacovanie a kategorizáciu štruktúrovaných alebo neštruktúrovaných údajov.
- Workflow Designer (wizard driven, drag & drop)
- Možnosť integrácie s treťostrannými aplikáciami

HCP Anywhere Enterprise

Práca na diaľku

Mobilná aplikácia

Regionálna kancelária

HCP AW Edge Filer

Verejný cloud

Dátové centrá

HCP Anywhere Enterprise

Globálny súborový systém

HCP

Dôveryhodné úložisko

Práca z domu

Mobilná aplikácia

Pobočka

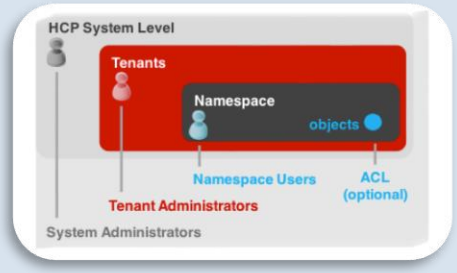
HCP AW Edge Filer

Verejný cloud

PosAm

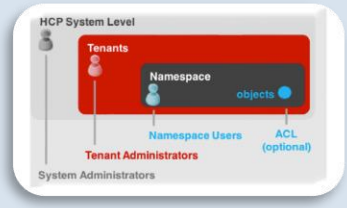
HCP – scenáre použitia a ekonomika škálovania

Dôveryhodné úložisko

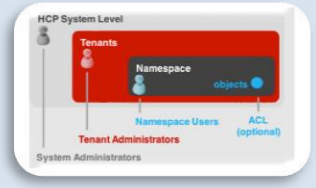


ECM/DMS

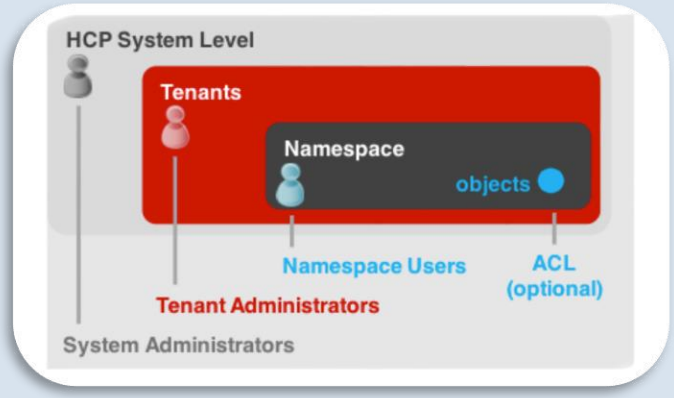
Hadoop / Splunk



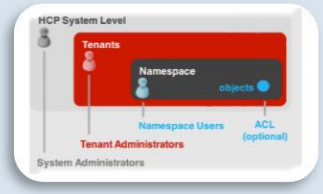
IoT dáta, snímky



Ochrana záloh

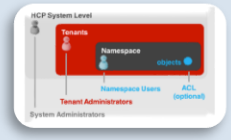


Distribúcia obsahu

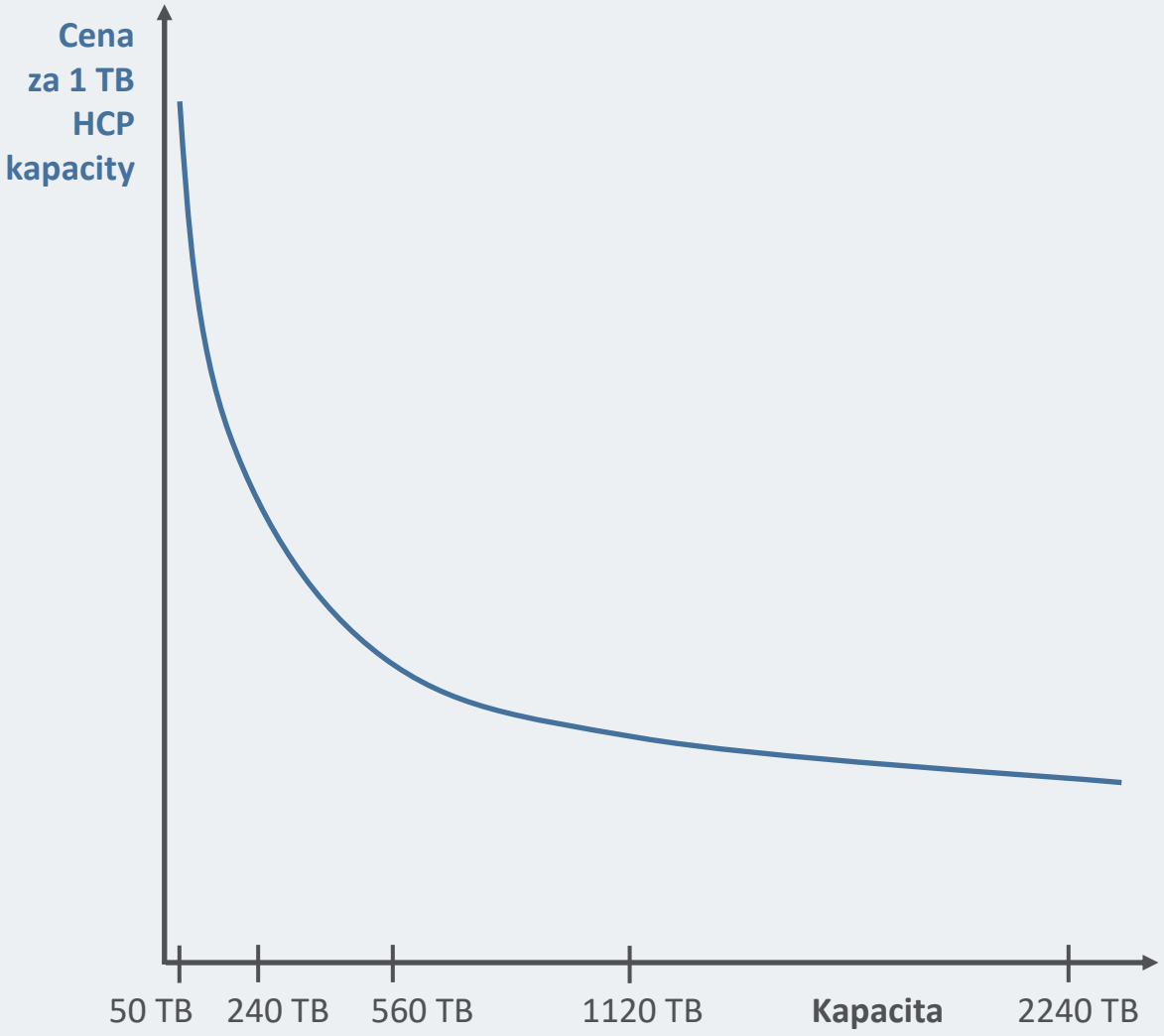


Cloudové aplikácie

Cloud úložisko



...a ďalšie



Ďalšie verejné zdroje:

K splneniu právnych požiadaviek:

- Hitachi Content Platform and Compliance Obligations in the European Union, BEP SYSTEMS, 2019
- Hitachi Content Platform (HCP): SEC 17a-4(f), FINRA 4511(c), and MiFID II Compliance Assessment, Cohasset Associates, 2019

Technická architektúra:

- Hitachi Content Platform Architecture Fundamentals, Hitachi Vantara, 2020
- HCP Security Hardening Whitepaper, Hitachi Vantara, 2020
- The Path to 10 Nines Availability with Hitachi Content Platform (HCP), Hitachi Vantara, 2017
- Secure File Services for Government and Defense, HCP Anywhere Enterprise, Hitachi Vantara, 2023

Posúdenie od tretích strán:

- 2022-23 DCIG TOP 5 On-Premises SDS Object Storage Solutions, Hitachi Content Platform, DCIG, 2023
- GigaOm Radar for Unstructured Data Management (UDM), GigaOm, 2023
- Object storage for digital-age challenges, MIT Technology Review Insights, 2020

PosAm

Ďakujem za pozornosť

Peter Smák
produktový manažér PosAm
peter.smak@posam.sk