**„Archivní úložiště pro IT4Innovations“**

**Základní požadavky zadavatele na předmět veřejné zakázky**

**Obsah**

[**Základní požadavky zadavatele na předmět veřejné zakázky** 1](#_Toc158734568)

[1. Úvod 1](#_Toc158734569)

[2. Parametry zakázky 1](#_Toc158734570)

[3. Účel a plánované užití úložiště 2](#_Toc158734571)

[4. Požadavky na archivní úložiště 2](#_Toc158734572)

[4.1. Obecné 2](#_Toc158734573)

[4.2. Vlastnosti a funkcionality 2](#_Toc158734574)

[4.3. Licencování 4](#_Toc158734575)

[4.4. Komponenty 4](#_Toc158734576)

[5. Rozsah zakázky 5](#_Toc158734577)

[6. Požadavky na záruku a servisní služby 5](#_Toc158734578)

[7. Obsah Návrhu 6](#_Toc158734579)

[8. Infrastruktura zadavatele 6](#_Toc158734580)

[8.1. Datové centrum 6](#_Toc158734581)

[8.2. Storage síť 6](#_Toc158734582)

## Úvod

Obsahem tohoto dokumentu jsou informace pro připravovanou veřejnou zakázku na pořízení archivního úložiště a předběžné požadavky na vlastnosti, parametry a funkcionality tohoto úložiště.

**Všechny požadavky a parametry uvedené v tomto dokumentu jsou předběžné.**

Pozornost prosím věnujte rovněž požadavkům a pokynům k Návrhu úložiště obsaženým v textu tohoto dokumentu. Požadavky jsou vyznačeny touto modrou barvou písma.

Pozn. V textu jsou uváděny rychlosti a kapacity za použití dekadických jednotek (násobky deseti), pokud není explicitně uvedena binární jednotka (např. KiB, MiB).

## Parametry zakázky

Předpokládaný termín realizace nákupu úložiště je **rok 2024**.

Předpokládaný finanční objem zakázky je **cca 5 mil. Kč bez DPH**.

Předpokládáme, že hodnotícím kritériem bude ekonomická výhodnost zakázky, kde jako hlavní kritéria budou zvolena kapacita úložiště a význačné vlastnosti řešení při stanovení pevné ceny za předmět veřejné zakázky.

## Účel a plánované užití úložiště

Archivní úložiště je pořizováno se záměrem **rozšířit stávající portfolio úložišť** zadavatele, které tvoří zejména centrální úložiště projektových dat PROJECT, uživatelská souborová úložiště clusterů HOME a SCRATCH, disková pole virtualizační infrastruktury, disková pole řešení pro zálohování a další.

Archivní úložiště bude sloužit jako centrální úložiště dat superpočítačového centra IT4Innovations, **k dlouhodobé archivaci dat** laboratoří a výzkumných projektů.

## Požadavky na archivní úložiště

### Obecné

Archivní úložiště je určeno pro archivaci dat – dlouhodobé uložení dat. Motivací pro realizaci archivního úložiště není plnění legislativních požadavků, ale vlastní potřeba superpočítačového centra na uložení dat. Archivní úložiště nemusí splňovat žádné specifické regulatorní požadavky na archivaci.

Pro uložená data se předpokládá minimální potřeba přístupu, data budou dlouhodobě odložena v úložišti a potenciální přístup může nastat za několik let (tři, pět, ... deset let). Zdrojem dat archivního úložiště bude zejména souborové úložiště projektových dat zadavatele PROJECT poskytující souborový přístup protokolem NFSv4 a další úložiště provozované superpočítačovým centrem.

Je kladen značný důraz na ekonomickou stránku celého řešení, náklady a udržitelnost v dlouhodobém časovém horizontu.

Očekáváme, že úložná kapacita bude realizována páskovou technologií LTO aktuální generace (tj. LTO9). Důvody pro páskovou technologii jsou předpokládaná nižší pořizovací cena na jednotku kapacity a nižší náklady na provoz v porovnání s diskovou technologií a delší životnost.

Součástí dodávky je komplexní řešení archivního úložiště včetně softwarového řešení.

### Vlastnosti a funkcionality

#### Souborová data

Archivní úložiště bude pracovat se souborovými daty a bude poskytovat souborový přístup k archivovaným datům.

#### LTFS

Archivní úložiště bude ukládat data na magnetické pásky v páskové knihovně. Pro ukládání dat na pásky bude používat formát Linear tape file system (LTFS) dle mezinárodního standardu ISO/IEC 20919:2021 „Information technology – Linear tape file system (LTFS) Format specification“.

Prosím uveďte, zda navrhované řešení splňuje požadavek na konkrétní verzi standardu ISO/IEC.

Použitím standardního formátu pro ukládání dat chceme zajistit dlouhodobou dostupnost/čitelnost dat (a to rovněž nezávisle na softwarovém řešení samotného archivního úložiště) a otevřenost řešení pro případný budoucí rozvoj.

#### Databáze metadat

Řešení umožní pracovat s metadaty (názvy adresářů, názvy souborů, vlastník, skupina, modification time) dat uložených na páskách ve formátu LTFS, zejména procházet adresářovou strukturu, vypisovat a prohledávat metadata uložených dat, bez nutnosti vložení pásky (pásek) do mechaniky. Metadata dat uložených na páskách tedy budou uložena nejenom na páskách samotných ve formátu LTFS, ale budou rovněž dostupná ve vhodném formátu (databázi) na disku.

Prosím uveďte, jakým způsobem budou uložena a dostupná metadata uložených dat a jak je možno s nimi pracovat.

#### Obsluha knihovny

Řešení bude obsluhovat páskovou knihovnu automaticky, bez intervence obsluhy. Pro běžné operace jako zápis a čtení archivního úložiště nebude potřeba zjišťovat/vyhledávat vhodné médium (pásku), nebude potřeba instruovat knihovnu k vložení potřebné/vhodné pásky do mechaniky, ale vyhledání vhodné pásky a manipulaci s knihovnou zařídí software archivního úložiště.

Prosím uveďte, jak probíhá přístup k datům a obsluha knihovny.

#### Replikace dat

Řešení bude poskytovat replikaci uchovávaných dat. Minimální požadavek je možnost dvou kopií dat, výhodou by byla možnost pro důležitá data nastavit tři kopie. Kopie dat budou umístěny na různých páskách. Vytváření kopií dat bud řídit software archivního úložiště automaticky.

Prosím uveďte, jaké vlastnosti replikace dat navrhované řešení poskytuje.

#### Kontrola integrity

Řešení bude poskytovat kontrolu integrity uchovávaných dat.

Prosím uveďte, jaké možnosti kontroly integrity dat navrhované řešení poskytuje.

#### Rozšířená metadata

Vítanou vlastností řešení by byla možnost doplnit/rozšířit metadata ukládaných dat (spravovaných v databázi na disku) o další atributy, jako jsou:

* štítky (tagy, labely) pro popsání dat
* požadovaná doba uložení/expirace dat
* a případně další údaje

Prosím uveďte, jaká metadata je možno pro uložená data používat a jakým způsobem.

#### Varianta HSM

Jako možné řešení vidíme HSM (Hierarchical storage management) systém, který by obsahoval dva tiery – relativně malý diskový tier a kapacitní páskový tier. Diskový tier by poskytoval buffer pro zpracovávaná data a okamžitou dostupnost metadat (umístěných na disku), páskový tier by zajišťoval umístění dat na páskách a obsluhu knihovny. HSM řešení by poskytovalo jednoduchý, snadno použitelný a transparentní způsob práce s archivním úložištěm. Přivítáme návrhy postavené na HSM systému, avšak uvádíme, že použití HSM systému není podmínkou řešení, a přivítáme i jiné návrhy.

Prosím uveďte, zda je navržené řešení koncipováno jako HSM systém.

#### Uživatelé

Archivní úložiště je určeno pro uživatele národního superpočítačového centra IT4Innovations. V ideálním případě by archivní úložiště měl možnost používat libovolný uživatel centra IT4Innovations.

Uživatelé centra IT4Innovations mají na uzlech výpočetních clusterů (OS Linux) vytvořené uživatelské účty (POSIX účty s uid, gid), alternativně lze použít LDAP databázi, kde uživatelé (a skupiny) jsou reprezentovány záznamy třídy posixAccount (resp. posixGroup).

V kompromisním případě si dovedeme představit, že možnost používání archivního úložiště by byla omezena na menší počet vybraných uživatelů.

Prosím uveďte, jak navržené řešení integruje uživatele centra IT4Innovations a v jakém rozsahu.

#### Klientské počítače

Co se týče počítačů, ze kterých by byl realizován klientský přístup uživatelů k archivnímu úložišti, tak máme představu, že by se jednalo o přístupové (login) uzly superpočítače Velký cluster (aktuálně Karolina) v počtu minimálně dva (ideálně všechny čtyři), pokud by to řešení umožňovalo. Přístupové servery používají OS na bázi RHEL (Rocky Linux 8). Na těchto uzlech je připojeno úložiště PROJECT a další úložiště superpočítače a jsou určeny pro interaktivní práci uživatelů. Klientské počítače by se časem měnily – po zprovoznění nového Velkého clusteru.

Prosím uveďte, jaké klientské počítače navrhujete a jaké požadavky jsou na klientské počítače kladeny.

#### Řízení přístupu

Řešení bude zajišťovat řízení přístupu k datům.

Řešení by mělo poskytovat a uplatňovat standardní unixová práva (rwx, ugo) a ideálně i rozšířená ACL práva (POSIX ACL nebo NFS v4 ACL). Alternativně by řešení poskytovalo a uplatňovalo řízení na per projekt bázi.

Centrum IT4Innovations přiděluje výpočetní a další zdroje pro konkrétní užití na základě schválených projektových žádostí formou tzv. projektů. Archivní úložiště bude sloužit pro uložení dat některých projektů a přidělování a přístup do archivního úložiště bude na per projekt bázi. Členství v projektu se promítá do členství v POSIX skupině.

Řešení by mělo umožňovat nastavit limit na kapacitu a počet souborů v archivním úložišti pro projekt a efektivně sledovat využití téhož.

Prosím uveďte, jaké možnosti řízení přístupu k datům navržené řešení poskytuje.

Prosím uveďte, jakým způsobem vámi navrhované řešení naplňuje uvedené požadavky. Uveďte požadavky, které vaše řešení nesplňuje, či plní jen částečně. Uveďte vlastnosti a funkcionality, které má navržené řešení nad rámec uvedených požadavků.

### Licencování

Pro řešení vyžadujeme trvalou licence všech potřebných komponent – zadavatel bude oprávněn archivační úložiště používat po neomezenou dobu bez dalších nákladů za licence.

Podpora softwarového řešení včetně jeho aktualizací po dobu záruky bude součástí dodávky.

Požadujeme, aby podmínky a finanční náklady na případné navýšení kapacity archivního úložiště a na prodloužení podpory softwarových produktů byly předvídatelné a akceptovatelné. Preferujeme řešení, jehož licencování se neodvíjí od kapacity.

Prosím specifikujte, jaké komponenty jsou předmětem licencování, možnosti licencování, popište licenční model(y) a náklady na licence. Uveďte náklady na software a licence pro případné navýšení kapacity archivního úložiště.

### Komponenty

Předpokládáme, že archivní úložiště bude sestávat z páskové knihovny, či páskových knihoven, obslužných serverů a potřebných sítí. Ze síťového pohledu se předpokládá připojení do storage sítě zadavatele, kde je připojeno rovněž úložiště projektových dat zadavatele PROJECT. Součástí dodávky jsou potřebná média – datové a čistící pásky.

#### Pásková knihovna

* Požadovaná souhrnná kapacita datových pásek **10PB** bez komprese.
* Datové a čistící pásky označené čárovými kódy
* Barcode scanner
* Minimálně čtyři LTO-9 mechaniky.
* Redundantní napájecí zdroje

Prosím uveďte navrhovanou konfiguraci páskové knihovny.

Prosím uveďte, zda pro předpokládaný účel vyhovují Half Height mechaniky, nebo jsou potřeba (či vhodnější) Full Height mechaniky.

#### Servery, sítě

Dle návrhu dodavatele.

* Redundantní napájecí zdroje

Prosím uveďte navrhovanou konfiguraci serverů a sítě.

#### Integrace úložiště do storage sítě

Služby úložiště budou systémům zadavatele poskytovány prostřednictvím storage sítě zadavatele využívající technologii 100Gb/s Ethernet. Na straně storage sítě zadavatele jsou k dispozici dva přístupové Ethernet switche podporující multi-chassis etherchannel vybavené konektory typu QSFP28. Pro připojení úložiště budou určeny dva QSFP28 porty, jeden port v každém switchi zadavatele. Přístupové switche zadavatele nejsou osazeny transceivery, potřebné transceivery jsou předmětem dodávky, stejně tak jako potřebná kabeláž.

Prosím uveďte návrh připojení úložiště do storage sítě zadavatele a jeho redundanci.

Pro připojení managementu úložiště zadavatel poskytne několik 1Gb/s portů ve své síťové infrastruktuře.

## Rozsah zakázky

Součástí zakázky bude:

* dodávka hardware řešení úložiště (pásková knihovna, servery, sítě, a další)
* datové a čistící pásky
* dodávka software řešení úložiště včetně potřebných licencí
* dodávka infrastruktury pro instalaci a provoz úložiště v datovém centru zadavatele – racky a příslušenství potřebné pro umístění zařízení úložiště, řešení napájení a chlazení zařízení úložiště, napojení na infrastrukturu datového centra zadavatele
* připojení úložiště do storage sítě zadavatele
* komplexní implementace celého systému v datovém centru zadavatele (doprava, integrace do datového centra, instalace, konfigurace, ladění, testování, likvidace obalů)
* příprava a provedení akceptačních testů pro prokázání funkčnosti, vlastností a parametrů díla
* zaškolení obsluhy (školení)
* poskytování servisu, podpory a provádění pravidelných softwarových upgradů po dobu záruční doby
* konzultace v průběhu provozu řešení

## Požadavky na záruku a servisní služby

Předpokládáme obdobné požadavky a formulace záručních podmínek jako u dříve realizovaných zakázek. Požadujeme, aby podmínky a finanční náklady na prodloužení podpory softwarových produktů byly předvídatelné a akceptovatelné.

Předpokládáme následující parametry záruky a servisních služeb:

* záruční doba 5 let
* servis v místě instalace
* jediné kontaktní místo pro hlášení poruch
* možnost hlášení poruch a vad 24 hodin denně

Kategorie vad:

* Vada kategorie A – vada, která zcela nebo podstatným způsobem znemožňuje užívání Úložiště,
* Vada kategorie B – vada, která nebrání užívání Úložiště, ale podstatným způsobem jej omezuje, nebo která vytváří riziko znemožnění užívání Úložiště,
* Vada kategorie C – jiná vada, která není vadou kategorie A ani B

Zahájení řešení odstranění vady:

* Vada kategorie A – maximálně do 2 hodin od nahlášení vady
* Vada kategorie B – maximálně do 4 hodin od nahlášení vady
* Vada kategorie C – maximálně do dalšího pracovního dne od nahlášení vady

Odstranění vady – archivní úložiště:

* Vada kategorie A – do 3 pracovních dnů od nahlášení vady
* Vada kategorie B – do 5 pracovních dnů od nahlášení vady
* Vada kategorie C – do 7 pracovních dnů od nahlášení vady

Za odstranění vady kategorie A nebo B lze považovat poskytnutí akceptovatelného náhradního řešení. Náhradní řešení vady kategorie A se považuje za nahlášenou vadu kategorie B a náhradní řešení vady kategorie B se považuje za nahlášenou vadu kategorie C; přípustné je jen to náhradní řešení, které skutečně umožňuje změnu kategorizace vady.

## Obsah Návrhu

Prosím, navrhněte technické řešení archivního úložiště a popište jeho vlastnosti a funkcionality.

V návrhu prosím uveďte:

* Orientační schéma řešení
* Navrhované softwarové řešení, jeho funkcionality a vlastnosti
* Navrhovaná zařízení a jejich počet a parametry
* Předpokládané kapacitní a výkonové parametry

V návrhu prosím uveďte odhadované finanční náklady ve struktuře:

* Hardware
* Software
* Implementace
* Podpora
* Racky a instalace v datovém centru zadavatele

## Infrastruktura zadavatele

### Datové centrum

IT4Innovations provozuje veškeré systémy ve svém datovém centru umístěném v budově IT4Innovations.

Datové centrum zajišťuje redundantní zdroj zálohovaného napájení (dvě nezávislé napájecí větve, redundance 1+1) rozvodné soustavy 400/230 V; 3+N+PE; 50 Hz; TN-S.

Chlazení pro datový sál poskytuje pět samostatných chladících kapalinových okruhů – dva okruhy teplé vody teploty (cca 30 °C) a tři okruhy studené vody (cca 11,5 °C), je použita směs 35% propylenglykol a 65% voda.

Datový sál o výměře 511 m2 poskytuje světlou výšku mezi systémovou podlahou a stropním podhledem 2.9 m.

### Storage síť

Storage síť zadavatele je postavena na technologii 100Gb/s Ethernet a je tvořena dvěma hraničními prvky – L3 switchi Cisco Nexus 9336C-FX2 s porty QSFP28. Hraniční prvky jsou vzájemně propojeny čtyřmi 100Gb/s propoji. Storage síť je připojena do centrální 100Gb/s WAN/LAN sítě zadavatele čtyřmi 100Gb/s propoji.

Hraniční prvky storage sítě nejsou osazeny transceivery, je potřeba je dodat.