Załącznik nr 2

OPIS PRZEDMIOTU ZAMÓWIENIA

Rozbudowa macierzy o wielkość 300 TB

w ramach projektu:

„Hereditas. Digitalizacja i udostępnianie zbiorów Muzeum Narodowego w Warszawie”

Projekt współfinansowany jest ze środków Programu Operacyjnego Polska Cyfrowa w ramach działania 2.3 „Cyfrowa dostępność i użyteczność informacji sektora publicznego”, poddziałanie 2.3.2 „Cyfrowe udostępnienie zasobów kultury”

1. Słownik pojęć

Administrator - pracownik Muzeum zarządzający instancją systemu wewnątrz Muzeum;

Magazyn Worm – magazyn przechowywania danych na zasadzie write-once-read-many;

Muzeum/Zamawiający/MNW - Muzeum Narodowe w Warszawie wraz z oddziałami;

2. Przedmiot zamówienia

Rozbudowa poprzez dostawę, instalację i konfigurację niezbędnych elementów obiektowego systemu składowania danych o efektywnej, dostępnej dla danych i metadanych pojemności nie mniejszej niż: 700TB, tj. rozbudowa o 300TB przestrzeni użytecznej (base10) dla archiwizacji środowiska Informatycznego.

Rozbudowa obecnego Systemu Składowania danych Obiektowych o dodatkowe 300TB (Base10) przestrzeni fizycznej wraz z wymaganymi licencjami do zarządzania tą przestrzenią (wsparcie producenta na rok od daty dostawy).

OPIS Równoważny: **System Przechowywania danych Obiektowych**

Docelowa (wraz z rozbudową) konfiguracja Bezpiecznego magazynu WORM (Write-Once-Read-Many) (wsparcie producenta na rok od daty dostawy).

1. Ogólne wymagania techniczne dla Bezpiecznego magazynu WORM.

1.1. Przedmiotem zamówienia jest rozbudowa poprzez dostawę, instalację i konfigurację niezbędnych elementów obiektowego systemu składowania danych o efektywnej, dostępnej dla danych i metadanych pojemności nie mniejszej niż: 700TB, tj. rozbudowa o 300TB przestrzeni użytecznej (base10) dla archiwizacji środowiska Informatycznego – środowisko dalej nazywane produkcyjnym (wymagany sprzętowy appliance, nie dopuszcza się rozwiązań zbudowanych w oparciu o maszyny wirtualne VM)

1.2. Wymagana pojemność 700TB dla środowiska produkcyjnego nie uwzględnia wykorzystania mechanizmów redukcji danych (przed procesem deduplikacji i kompresji)

1.3. Wymagana pojemność musi być dostarczona i zainstalowana w 1 ośrodku, z możliwością zbudowania lustrzanej kopii w drugim ośrodku z taką samą ilością 700TB użytecznej (base10) przestrzeni w każdym ośrodku.

1.4. System musi zapewnić mechanizm asynchronicznej replikacji obiektów pomiędzy ośrodkami za pomocą istniejących łącz Ethernet.

1.5. Całe rozwiązanie musi być dostarczone wraz z kompatybilnym rozwiązaniem i infrastrukturą u Zamawiającego szafami Rack 19”, zainstalowane maksymalnie w jednej szafie per ośrodek.

1.6. Dostarczane rozwiązanie musi być produktem rozpoznawalnym na rynku, co oznacza, że powinno być wymieniane w raportach niezależnych organizacji, takich jak Gartner, IDC, Gigaom lub ESG (Enterprise Strategy Group).

1.7. Dostarczane rozwiązanie obiektowego magazynu dokumentów musi być obecne na rynku od co najmniej 3 lat.

1.8. Oferowane rozwiązanie musi być produktem gotowym, posiadającym na moment składania oferty wszystkie wymagane przez Zamawiającego funkcjonalności. Do oferty należy załączyć listę wszystkich komponentów urządzenia. Lista ma zawierać co najmniej nazwy urządzeń, modeli oraz inne

informacje pozwalające w sposób jednoznaczny zidentyfikować poszczególne komponenty sprzętowe i programowe.

1.9. Oferowane urządzenia i wszystkie jego elementy składowe muszą być fabrycznie nowe i wyprodukowane nie wcześniej niż pół roku przed terminem dostawy do Zamawiającego.

1.10. Dostarczony sprzęt musi być zakupiony w oficjalnym kanale sprzedaży producenta. Oznacza to, że dostarczony sprzęt będzie sprzętem nowym, nieużywanym wcześniej w innych projektach i posiadającym stosowny pakiet usług gwarancyjnych kierowanych do użytkowników z obszaru Rzeczpospolitej Polskiej, na co Wykonawca przedstawi stosowne oświadczenia.

1.11. Oferowane urządzenia i wszystkie jego elementy muszą pochodzić od autoryzowanego dostawcy producenta.

1.12. Urządzenia muszą być oznakowane przez producenta w taki sposób, aby możliwa była identyfikacja zarówno produktu jak i producenta.

1.13. Wraz z rozwiązaniem musi być dostarczony komplet dokumentacji w formie papierowej lub elektronicznej. Dokumentacja papierowa powinna być czytelna. Zamawiający dopuszcza dostawę dokumentacji producenta rozwiązania w językach polskim lub angielskim.

1.14. Wraz z rozwiązaniem musi być dostarczony komplet nośników umożliwiający odtworzenie oprogramowania systemowego urządzeń, z których zbudowane jest dostarczone rozwiązanie wraz z procedurami disaster recovery których zakres zostanie ustalony w porozumieniu z Zamawiającym.

1.15. Rozwiązanie musi mieć możliwość podłączenia go do centrum serwisowego producenta, w celu zdalnego monitorowania poprawności funkcjonowania komponentów rozwiązania.

1.16. Wszystkie oferowane urządzenia muszą być publicznie dostępne. Zamawiający nie dopuszcza stosowania urządzeń dedykowanych, stworzonych na potrzeby niniejszego zamówienia.

1.17. Oferowane oprogramowanie i urządzenia w dniu składania ofert nie mogą być przeznaczone przez producenta do wycofania z produkcji, do wycofania ze sprzedaży lub wsparcia technicznego.

1.18. Zamawiający wymaga, by dostarczone oprogramowanie systemowe (firmware) było oprogramowaniem w wersji aktualnej na dzień poprzedzający dzień składania ofert.

2. Wymagania dotyczące skalowalności, budowy i architektury obiektowego systemu składowania dokumentów.

2.1. Wszystkie elementy dostarczonego rozwiązania muszą być redundantne, a jego architektura musi zapewniać odporność na wystąpienie pojedynczego punktu awarii w obrębie poszczególnych grup elementów, to jest co najmniej: interfejsów dostępowych kontrolerów, serwerów, zasilaczy, wentylatorów, dysków. Odporność na awarię oznacza, że dostęp do urządzenia oraz do składowanych na nim danych musi być realizowany bez przerywania pracy korzystającej z niego aplikacji/systemu, zapewniając możliwość odczytów wszystkich składowanych danych oraz wykonywania zapisów na urządzenie nawet
w przypadku awarii lub wymiany pojedynczego elementu urządzenia z ww. grup urządzeń.

2.2. Rozwiązanie powinno być dostarczone w postaci klastra wysokiej dostępności (HA) zbudowanego z odpowiedniej liczby węzłów wraz z zainstalowanym na nich oprogramowaniem układowym i systemowym, które zapewnią realizację wymienionych w punkcie 3 funkcjonalności.

2.3. Architektura rozwiązania musi zapewniać umieszczenie interfejsów dostępowych i dyskowych wewnątrz wszystkich węzłów klastra, realizujących funkcję obiektowego systemu składowania dokumentów.

2.4. Rozwiązanie dla środowiska produkcyjnego musi posiadać możliwość pracy zarówno w architekturze, w której węzły klastra HA korzystają z przestrzeni dyskowej zainstalowanej w blokowej macierzy dyskowej (scale-up, SAIN), jak i w architekturze, w której przestrzeń dyskowa jest dostarczana na dyskach zainstalowanych wewnątrz węzłów klastra (scale-out, RAIN).

2.5. Architektura dostarczonego rozwiązania dla środowiska produkcyjnego powinna uwzględniać:

2.5.1. dedykowane węzły dla realizacji funkcji dostępu do obiektowego magazynu składowania danych oraz przechowywania metadanych i danych najczęściej używanych.

2.5.2. możliwość rozbudowy o dedykowane węzły dla realizacji funkcji przechowywania danych w obiektowym magazynie składowania danych.

2.6. Wszystkie elementy opisanej powyżej architektury muszą być ze sobą zintegrowane w taki sposób, aby zapewnić automatyczny przepływ danych pomiędzy różnymi warstwami architektury.

2.7. Dostarczone rozwiązanie dla środowiska produkcyjnego powinno umożliwiać rozbudowę do co najmniej 500PB przestrzeni bez konieczności zatrzymywania pracy rozwiązania i bez przerywania dostępu do danych.

2.8. Dostarczone rozwiązanie dla środowiska produkcyjnego powinno umożliwiać rozbudowę do co najmniej 80 węzłów dostępowych oraz 80 węzłów przechowywania danych.

2.9. Dostarczone rozwiązanie dla środowiska produkcyjnego powinno być zbudowane z co najmniej 4 węzłów dostępowych oraz 1 części ekonomicznej składowania.

2.10. Przestrzeń dyskowa w dostarczonym rozwiązaniu dla środowiska produkcyjnego musi w każdym ośrodku zostać dostarczona w węzłach dostępowych. Musi istnieć możliwość rozbudowy tej przestrzeni zarówno

poprzez dodawanie węzłów dostępowych, jak i poprzez dodawanie węzłów przechowywania.

2.11. Komunikacja pomiędzy węzłami klastra (wewnętrzna) musi być realizowana za pomocą interfejsów 10GbE. Wraz z rozwiązaniem należy dostarczyć redundantne przełączniki LAN dedykowane do obsługi wewnętrznej sieci klastra.

2.12. Komunikacja na zewnątrz (czyli dostęp do rozwiązania) musi być realizowana za pomocą interfejsów 10GbE(Base-T/RJ45).

2.13. Wykonawca zobowiązany jest do dostarczenia pełnego okablowania niezbędnego do uruchomienia systemu, a długość okablowania musi zostać zatwierdzona przez Zamawiającego przed rozpoczęciem prac montażowych, w fazie przygotowania dokumentacji wykonawczej.

2.14. Architektura rozwiązania dla środowiska produkcyjnego musi zapewniać możliwość elastycznej rozbudowy poprzez co najmniej:

- dodawanie tylko przestrzeni dyskowej przy niezmienionej ilości węzłów dostępowych do obiektowego magazynu składowania danych.
- dodawanie niezależnie węzłów dostępowych oraz węzłów przechowywania danych.
- dodawania dysków do istniejących węzłów w klastrze.

3. Szczegółowe wymagania funkcjonalne dla obiektowego systemu składowania danych.

3.1. Dane w obiektowym magazynie danych muszą być składowane na napędach dyskowych. Nie dopuszcza się rozwiązań zbudowanych w oparciu o napędy taśmowe.

3.2. Dostarczone rozwiązanie musi posiadać wbudowane mechanizmy przechowywania zarówno danych, jak i metadanych (informacji opisujących dane). Nie dopuszcza się wykorzystania rozwiązań plikowych (NAS) jako warstwę przechowywania w systemie składowania danych.

3.3. Rozwiązanie powinno posiadać możliwość integracji z aplikacjami za pomocą co najmniej następujących protokołów i interfejsów: HTTP, S3, SWIFT, REST API, WebDAV, CIFS, NFS. Jeżeli wykorzystanie któregokolwiek z wymienionych protokołów i interfejsów wymaga zastosowania dodatkowej licencji lub oprogramowania, to należy je dostarczyć wraz z rozwiązaniem.

3.4. Rozwiązanie powinno posiadać wbudowane mechanizmy protekcji danych, które gwarantują odczyt wszystkich składowanych danych w przypadku awarii pojedynczego, losowego komponentu architektury (dysku, karty sieciowej, przełącznika LAN, serwera i kontrolera urządzenia).

3.5. W przypadku dysków Zamawiający wymaga, aby dostarczone rozwiązanie wykorzystywało następujące mechanizmy protekcji danych: RAID-6 lub Erasure Coding (EC) dla dysków SAS i SAS-NL

3.6. Dostarczone rozwiązanie powinno być wyposażone w dyski nie większe niż 14TB.

3.7. Dostarczone rozwiązanie musi zapewniać i gwarantować niezmienność składowanych w nim obiektów, między innymi poprzez wykorzystanie

wbudowanej technologii WORM (Write Once Read Many). W przypadku rozwiązania dla środowiska produkcyjnego Zamawiający wymaga, aby funkcjonalność WORM była realizowana wewnątrz dostarczonego gotowego rozwiązania sprzętowego (appliance) w jego oprogramowaniu systemowym. Dla środowiska produkcyjnego Zamawiający nie dopuszcza, aby funkcjonalność WORM realizowana była poprzez rozwiązania programowe i rozwiązania uruchamiane w warstwie wirtualizacyjnej (VMware, Hyper-V, KVM i inne).

3.8. Rozwiązanie musi posiadać możliwość definiowania różnych poziomów retencji przechowywania danych, gwarantujących brak możliwości skasowania danych przed upływem zdefiniowanego czasu.

3.9. Funkcjonalności WORM oraz retencja muszą działać dla wszystkich wspieranych przez rozwiązanie protokołów dostępowych.

3.10. Retencja powinna być ustawiana zarówno dla danych jak i dla własnych (custom) metadanych. W przypadku własnych metadanych musi istnieć możliwość zdefiniowania przez administratora co najmniej następujących operacji:

3.10.1. Pełna retencja, czyli brak możliwości jakichkolwiek zmian w metadanych.

3.10.2. Możliwość dopisania nowych rekordów do metadanych, ale bez możliwości zmiany i kasowania już istniejących.

3.10.3. Możliwość dopisania nowych rekordów do metadanych z możliwością zmiany i kasowania już istniejących.

3.11. Rozwiązanie musi posiadać możliwość ustrukturyzowania metadanych. To oznacza, że musi istnieć możliwość podziału metadanych na co najmniej 10 odrębnych grup (adnotacji) w taki sposób, aby z każdej z tych grup (adnotacji) mogły korzystać niezależne aplikacje bez konieczności duplikowania obiektów w magazynie składowania danych.

3.12. Rozwiązanie musi posiadać możliwość tworzenia logicznych partycji oraz przestrzeni nazw definiowanych wewnątrz tych partycji. Zamawiający wymaga, aby dostarczone rozwiązanie dla środowiska produkcyjnego posiadało możliwość zdefiniowania co najmniej 1000 logicznych partycji oraz co najmniej 10000 przestrzeni nazw. W przypadku środowiska testowego wymagana jest możliwość zdefiniowania co najmniej 5 logicznych partycji oraz co najmniej 25 przestrzeni nazw. Musi istnieć możliwość mapowania i wykorzystania różnych przestrzeni nazw dla różnych aplikacji, w taki sposób, aby dla każdej z tych aplikacji możliwe było definiowanie różnych i niezależnych parametrów i kryteriów składowania danych, w tym co najmniej: retencji, nieodwracalnego niszczenia danych, wersjonowania, indeksowania i replikacji.

3.13. Rozwiązanie musi pozwalać na zdefiniowanie partycji, w których istnieje możliwość usuwania danych przed upływem retencji oraz partycji, w których usuwanie danych przed upływem retencji jest niemożliwe. Rozwiązanie powinno pozwalać na definiowanie i uruchamianie jednocześnie obydwu typów partycji.

3.14. W przypadku partycji, w której istnieje możliwość usuwania danych przed upływem retencji wymagane jest, aby taką operację mógł wykonywać jedynie

administrator z odpowiednimi uprawnieniami oraz aby operacja ta była audytowalna, co oznacza, że czynności związane z usuwaniem muszą być rejestrowane w wewnętrznych dziennikach dostarczonego rozwiązania.

3.15. Każda ze zdefiniowanych partycji musi mieć możliwość zarządzana przez różnych administratorów.

3.16. Rozwiązanie musi posiadać wbudowane mechanizmy pozwalające na rozliczanie kosztów wykorzystania jego zasobów.

3.17. Rozwiązanie powinno posiadać wbudowany mechanizm zatrzymania retencji danych, co oznacza, że w przypadku, gdy taki mechanizm zostanie włączony dla danego obiektu, retencja danych musi być utrzymywana dla tego obiektu do momentu jego wyłączenia, niezależnie od zadanego parametru czasu w definicji polityki retencji.

3.18. Rozwiązanie musi posiadać wbudowany mechanizm nieodwracalnego niszczenia danych, dla których okres retencji został przekroczony lub nie został zdefiniowany.

3.19. Rozwiązanie musi posiadać wbudowane mechanizmy zapewniające możliwość potwierdzania autentyczności składowanych danych. Mechanizmy te muszą opierać się o wyliczenie przez urządzenie sumy kontrolnej dla każdego składowanego obiektu. Administrator rozwiązania musi mieć możliwość wyboru algorytmu, który będzie wykorzystany do wyliczenia sumy kontrolnej. Wymagane jest wsparcie dla co najmniej następujących algorytmów kryptograficznych: MD5, SHA-1, SHA-256 i SHA-512.

3.20. Rozwiązanie musi posiadać swoje własne wbudowane mechanizmy weryfikacji sum kontrolnych składowanych obiektów.

3.21. Rozwiązanie powinno posiadać wbudowane mechanizmy redukcji danych, w tym co najmniej deduplikację i kompresję danych.

3.22. W przypadku, gdy oferowane rozwiązanie nie posiada dowolnej z funkcjonalności opisanych w poprzednim punkcie, wówczas należy dostarczyć rozwiązanie o pojemności dwukrotnie większej niż pierwotnie opisana przestrzeń wymagana.

3.23. Rozwiązanie powinno posiadać wbudowany mechanizm wersjonowania obiektów wraz z funkcjonalnością kasowania poprzednich wersji po określonym przez administratora czasie.

3.24. Rozwiązanie dla środowiska produkcyjnego musi posiadać możliwość szyfrowania danych. Szyfrowanie powinno być realizowane: na dyskach obiektowego magazynu składowania danych, na połączeniu do replikacji pomiędzy ośrodkami i w przypadku tieringu danych do zewnętrznej warstwy w szczególności do chmury publicznej.

3.25. Rozwiązanie musi posiadać natywnie wbudowane mechanizmy umożliwiające replikację składowanych danych pomiędzy różnymi lokalizacjami z wykorzystaniem sieci LAN/WAN i protokołu HTTP. Zastosowanie niniejszego mechanizmu musi również spełniać wymagania replikacji metadanych, uprawnień, polityki retencji oraz niezmienności danych tzn. awaria urządzenia w lokalizacji podstawowej nie może eliminować gwarancji niezmienności danych na platformie zdalnej.

3.26. Replikacja powinna być możliwa zarówno w trybie Active/Passive, czyli w trybie, w którym do odczytu i zapisu udostępniona jest replikowana przestrzeń nazw tylko w jednym ośrodku, jak i w trybie Active/Active, w którym do odczytu i zapisu udostępnione są replikowane przestrzenie nazw w każdym ośrodku.
3.27. Replikacja powinna być możliwa pomiędzy co najmniej 5 ośrodkami. W każdym z tych ośrodków replikowana przestrzeń nazw musi być jednocześnie dostępna do zapisu i odczytu w przypadku replikacji w trybie Active/Active.
3.28. Rozwiązanie powinno wspierać różne topologie replikacji danych w tym co najmniej: 1-do-wielu, 1-do-1, wiele-do-1.
3.29. Rozwiązanie powinno posiadać wbudowany mechanizm tieringu danych, realizowanego automatycznie i w sposób przeźroczysty dla aplikacji użytkowników.
3.30. Tiering powinien być realizowany pomiędzy węzłami dostępowymi, a:
    3.30.1. węzłami przechowywania danych w obiektowym magazynie danych.
    3.30.2. zewnętrznymi urządzeniami NAS za pomocą protokołu NFS.
    3.30.3. chmurą publiczną, co najmniej takich producentów jak Microsoft, Amazon i Google.
3.31. Rozwiązanie powinno posiadać możliwość zarządzania co najmniej poprzez graficzny interfejs użytkownika oraz poprzez API.
3.32. Rozwiązanie powinno posiadać interfejs API dla protokołów dostępowych, co najmniej S3, http i SWIFT oraz do zarządzania.
3.33. Rozwiązanie powinno posiadać wbudowany silnik wyszukiwania metadanych oraz dostępne dla tego silnika API.
3.34. Wyszukiwanie metadanych powinno być realizowane m.in. w oparciu o tzw. content classy oraz wyrażenia regularne.
3.35. W celu weryfikacji funkcjonalności oferowanych przez proponowany system, Zamawiający zastrzega sobie możliwość wezwania do przeprowadzenia wybranych testów funkcjonalnych potwierdzających zadeklarowane funkcjonalności, potwierdzenie funkcjonalności musi zostać zrealizowane w ciągu 5 dni od daty wezwania. W razie odmowy przeprowadzenia testów lub w przypadku niepotwierdzenia funkcjonalności w ramach przeprowadzonych testów, Zamawiającemu odrzuci proponowaną ofertę jako niespełniającą wymagań.

4. Posiadane rozwiązanie sprzętowo-programowe przez Zamawiającego, które jest przedmiotem rozbudowy w ramach niniejszego Zamówienia:

   Hitachi Content Platform o następujących parametrach:

   - Zabudowa – 1 szafa RACK
   - Sprzętowy HCP
   - Węzły dostępowe 4xG11
   - Ilość i typ węzłów dyskowych 1xS11
   - Obecna licencja PLC – 400TB
   - Pojemność dyskowa per ośrodek - 342TB (DPL=1)

- Punkt styku ze środowiskiem produkcyjnym – 8x10GbE (Base-T/RJ45)
- Interfejs do środowiska zarządzania – 4x1GbE
- Obsługiwana ilość obiektów – 3.6mld, przy maksymalnej liczbie 100mld
- Szyfrowanie danych - Data in flight, data at rest
- Wsparcie standard SEC17-a4(f)
- Algorytmy dla wyliczania sum kontrolnych – MD5, SHA-1, SHA-256, SHA-512
- Obsługa modelu silnej spójności (strong consistency)
- Kompresja i deduplikacja
- Kompresja i szyfrowanie danych na łączach replikacyjnych
- Ilość partycji (tenant) – 1.000
- Ilość przestrzeni nazw (bucket) – 10.000