



Nr. 692.357 / 24.11.2022

RĂSPUNS

la solicitările de clarificări primite din partea unor operatori economici cu privire la anunțul de consultare a pieței ce are ca obiect *produse și servicii pentru implementarea unei platforme de tip BigData și dezvoltarea Sistemului Informatic de date integrate - în cadrul proiectului finanțat din PNRR “Administrație performantă prin informație consolidată - APIC ”*
2022_CP_008

Autoritatea contractantă răspunde după cum urmează:

Întrebarea nr.1

„Cu privire la cerința de migrare a datelor din DW actual în sistemul DW nou/ platforma de tip BigData, vă rugăm să clarificați dacă:

a. noul DW va fi folosit exclusiv pentru scopul proiectului de BigData și Data Warehouse (DW) existent va fi păstrat în continuare după migrare.

sau

b. noul DW se va folosi și în scopuri pre-existente proiectului în discuție.”

Răspuns la întrebarea nr.1

Subsistemul DW este descris în Caietul de Sarcini la cap. 3.4.1.3.8 (pag. 39-41) unde se precizează că o acțiune importantă în cadrul proiectului este migrarea datelor din DW actual în sistemul DW nou/ platforma de tip BigData.

De asemenea, la cap. 3.5.3 Servicii de implementare a sistemului (A#4), pag. 80, din Caietul de Sarcini se menționează că în cadrul proiectului trebuie asigurată integrarea cu sistemele actuale și migrarea tuturor componentelor din DW actual. Aceasta înseamnă că datele din Data Warehouse, job-urile de încărcare (aproximativ 250 joburi de complexitate medie), repository-ul de middleware și catalogul de rapoarte vor fi migrate SI SIDI și se va reconfigura accesul la acestea pentru aplicațiile/API, etc. pentru care sunt surse de date cu care sunt interconectate, astfel încât toate fluxurile existente să fie funcționale.

Întrebarea nr.2

„Vă rugăm să ne furnizați volumele estimate pentru bazele de date a surselor de date dorite a fi integrate în noul DW.”

Răspuns la întrebarea nr.2

În Caietul de Sarcini la cap. 3.1 Descrierea situației actuale la nivelul Autorității contractate, pag.18, sunt prezentate *Sursele de date pentru platforma Data Warehouse, din cadrul Sistemului Informatic al Ministerului Finanțelor.*

Baza de date din cadrul platformei Data Warehouse este de tip Oracle Database Enterprise Edition versiunea 11.2.0.3 și este în configurație RAC cu 2 noduri. Middleware-ul din cadrul platformei este de tip Oracle Business Intelligence Enterprise Edition versiunile 10 și 11. Sursele de date sunt:

- 252 de baze de date Oracle Database Enterprise 8.0.5 distribuite la nivel teritorial, a căror date sunt consolidate zilnic la nivel central cu instrumentul ETL IBM DataStage 8;
- 10 baze de date Oracle Database Enterprise Edition versiunile 10g, 11g și 19c, situate la nivel central, din care datele sunt consolidate prin intermediul conexiunilor de tip dblink și/sau cu instrumentul de replicare Quest SharePlex.

Fondul de date existent care va fi consolidat din mediile de producție este de aproximativ 102TB, care include creșterea anuală estimată pe o perioadă de 5 ani.

Menționăm că în cadrul proiectului sunt solicitate și funcționalități noi, de exemplu 3.4.1.3.3.5 Analiza rețelei sociale (pag. 34 din caietul de sarcini), 3.4.1.3.5 Constituirea (realizarea) Registrul de Risc al Contribuabilului (RRC) (pag. 37 din caietul de sarcini), etc. pentru care nu dispunem de o estimare a volumului datelor care vor trebui integrate în noul DW.

Întrebarea nr.3

„Care este numărul de utilizatori totali ai platformei Big Data?”

Răspuns la întrebarea nr.3

Având în vedere similitudinea cu alte platforme de tip Big Data, implementate la nivelul altor instituții, coroborat cu estimarea privind personalul ANAF implicat în efectuarea analizei de risc, numărul de utilizatori prezumat este cuprins între 2500-3000 utilizatori, cu luarea în calcul a posibilității de extindere cu până la 5000 de utilizatori.

Întrebarea nr.4

„Vă rugăm să confirmați că Beneficiarul va pune la dispoziția implementatorului metodologia pentru analiza și modelarea factorilor de risc.”

Răspuns la întrebarea nr.4

Confirmăm faptul că beneficiarul va pune la dispoziția implementatorului metodologia pentru analiza de risc și modelarea factorilor de risc, după începerea proiectului, în cadrul activității de analiză.

Întrebarea nr.5

„Pentru a estima planul de proiect și serviciile de implementare, vă rugăm să detaliați:

- lista surselor de date care vor constitui date de intrare pentru noul sistem, împreună cu date de dimensionare: tip aplicații, volume de date.

- lista tuturor aplicațiilor care sunt acum în implementare sau re tehnologizare ale căror surse de date vor intra în scopul proiectului și calendarul estimat de implementare a fiecărei din aceste aplicații pentru determinarea suprapunerii cu perioada derulării proiectului BigData.”

Răspuns la întrebarea nr.5

În Caietul de Sarcini la pag. 9 sunt prezentate câteva din sistemele informatice care sunt sursă de date pentru noul sistem. În Caietul de Sarcini la cap. 2.4 Alte inițiative/proiecte/programe asociate cu această achiziție (pag. 12-13) sunt descrise două proiecte ale căror etape de implementare trebuie armonizate cu precădere.

Întrebarea nr.6

„Vă rugăm să detaliați sursele de date ce vor furniza date pentru noul sistem din punct de vedere al tehnologiilor utilizate, tip de integrare: batch/timp real/ near real time. De asemenea, având în vedere diversitatea tehnologică, pentru sursele de date externe vă rugăm să confirmați dacă veți defini un format standardizat ce va trebui implementat, atât din perspectiva furnizorului de date, cât și al noului sistem (fișiere txt, servicii REST, acces direct la baza de date etc.).”

Răspuns la întrebarea nr.6

Tehnologiile utilizate pentru sursele de date sunt detaliate la pag. 18 și 43 din Caietul de sarcini. Tipul de integrare solicitat este „Near real time”. Conform cap. 3.4.1.1.3 „Extensibilitate”, „Utilizarea serviciilor web (web Services) va asigura o interoperabilitate ridicată a soluției, permițând interconectarea cu alte aplicații bazate și pe platforme eterogene”.

Întrebarea nr.7

„Vă rugăm să clarificați dacă în cele două site-uri, cel principal și cel de DR, componentele software vor fi dimensionate identic (cu licențierea aferentă) sau este posibil ca site-ul de DR să fie dimensionat la o capacitate de procesare redusă (spre exemplu 50% din capacitatea site-ului principal). Care sunt cerințele pentru RTO și RPO?”

Răspuns la întrebarea nr.7

În Caietul de Sarcini la cap. 3.6.2 Arhitectura tehnică a sistemului (pag. 90) la pct. 5 și 6 se regăsește răspuns la întrebarea legată de cerințele pentru RTO Recovery Time Objective și RPO Recovery Point Objective:

5. „Soluția propusă, trebuie să asigure obiectivul de timp de recuperare (RTO, în engl. Recovery Time Objective) de 15 (cincisprezece) minute pentru recuperarea în caz de dezastru.
6. Soluția propusă, trebuie să asigure obiectivul (de timp) de punct de recuperare (RPO, în engl. Recovery Point Objective) de 60 (șaizeci) minute pentru recuperarea în caz de dezastru.”

Întrebarea nr.8

„Vă rugăm să specificați numărul de utilizatori care vor folosi componentele software, în special pentru subcomponentele software descrise în capitolele: 3.4.1.3.9 Subcomponenta de Business Intelligence, 3.4.1.3.3 Subsistemul de Analytics, previziune etc, 3.4.1.3.5 Constituirea (realizarea) Registrul de Risc al Contribuabilului (RRC).”

Răspuns la întrebarea nr.8

Majoritatea utilizatorilor menționați la răspunsul de la întrebarea nr. 3 vor putea folosi toate subcomponentele, iar în ceea ce privește Constituirea (realizarea) Registrul de Risc al Contribuabilului (RRC) utilizatorii vor fi în număr de 40 (DGMR), cu luarea în calcul a posibilității de extindere cu până la 100.

Întrebarea nr.9

„Având în vedere Cap 3.4.1.1.4 Prevederi de securitate, “Accesul utilizatorilor - sistemul va fi integrat cu sistemul de management al identității și portalul ANAF”, vă rugăm să ne furnizați mai multe informații despre soluția existentă de management a identității și portal ANAF și să clarificați dacă ofertantul va trebui să includă propria soluție de identitate și acces la nivelul sistemului APIC.”

Răspuns la întrebarea nr.9

Aplicația trebuie să folosească soluția de identitate și acces utilizată la nivelul portalului ANAF. Nu este necesar ca operatorii economici interesați să propună propria soluție de identitate și acces.

Soluția de management al identității utilizată în cadrul sistemului informatic al MF-ANAF este realizată prin intermediul produsului Tivoli Identity Manager - Version: 5.0.0.4. Acesta realizează provizionarea utilizatorilor interni și a rolurilor către SAM-ul, LDAP-ul aferent portalului ANAF (Security Acces Manager, Tivoli Directory Server). Drepturile de acces se dau sub formă de roluri. Un utilizator poate avea unul sau mai multe roluri în funcție de aplicația/aplicațiile la care are acces.

Din punct de vedere al accesului și al integrării cu portalul ANAF (portal ce are la bază tehnologiile IBM, IBM WebSphere Portal Extend, versiunea 8.02), utilizatorul se va autentifica pe portalul ANAF și odată autentificat, în baza rolurilor aferente fiecărui utilizator, va accesa aplicația APIC, conform funcționalităților implementate în cadrul acestei aplicații. Aplicația trebuie să suporte SSO (single sign-on) integrat cu IBM Websphere Portal prin IBM Security Access Manager, versiunea 7.0, IBM WebSeal, versiunea 7.0. Astfel, utilizatorul autentificat în portal prin IBM Security Access Manager, IBM WebSeal, nu trebuie să se mai autentifice iar atunci când accesează aplicația.

Întrebarea nr.10

„Vă rugăm să clarificați conceptul de „middleware” definit în documentația consultării de piață, pe de o parte, în capitolul 3.4.1.2. Funcționalitati ale sistemului / pag 30, apare definit conceptul de Bază de date intermediară (Middleware) - ODS (Online Data Store), iar pe de altă parte în specificațiile tehnice ale echipamentelor hardware (3.4.8 SERVERE MIDDLEWARE) se solicită servere middleware, generând astfel o confuzie de termeni, în final, nefiind astfel clar pe ce echipamente va fi deployat layer-ul aplicativ,

spre exemplu unde vor rula componentele de analytics, clasificare pe registre de risc etc.”

Răspuns la întrebarea nr.10

Baza de date intermediară nu va fi considerată ca făcând parte din categoria „middleware” în contextul Caietului de sarcini.

Întrebarea nr.11

„În contextul punctului ridicat mai sus, vă rugăm să clarificați dacă Baza de date intermediară - ODS va fi deployată pe serverele de baza de date (vezi 3.4.7 SERVERE PENTRU BAZE DE DATE), reprezentând o componentă cheie în partea de preluare, unificare și curățare a datelor, și va trebui să îndeplinească aceleași cerințe tehnice ca și subsistemul de DW, definit în capitolul 3.4.1.3.8 Subsistemul DW.”

Răspuns la întrebarea nr.11

Confirmăm faptul că baza de date intermediară va fi deployată pe serverele de la cap. 3.4.7 SERVERE PENTRU BAZE DE DATE.

Întrebarea nr.12

„Având în vedere cerința exprimată la pag. 18, „Fondul de date care va fi consolidat din mediile de producție este de aproximativ 102TB, care include creșterea anuală estimată pe o perioadă de 5 ani.”, vă solicităm să furnizați adițional informații volumetrice estimative detaliate (dimensiune estimată ODS/DW, creștere anuală/cantitate de date ce va fi procesată, constrângeri existente în procesarea datelor, complexitatea raportărilor etc), astfel încât ofertantul să poată defini o arhitectura corectă din punct de vedere al capacității de stocare și procesare a datelor, corelat cu nevoia reală din cadrul proiectului.”

Răspuns la întrebarea nr.12

Estimarea de aproximativ 102TB se referă la datele extrase din sursele de date existente, pentru a crea o bază de date unitară, pentru a integra informații din surse multiple (de exemplu, conturi contribuabili, procesare declarații, colectare etc.) pentru efectuarea unor operații suplimentare asupra datelor, păstrarea unui istoric, etc.

Din cauza faptului că informația provine din surse multiple, integrarea presupune adesea și curățarea, soluționarea redundanțelor și verificarea comparativă cu regulile definite referitor la încărcarea datelor.

Pentru estimarea volumului de date care se vor modela, vizualiza în diversele componente ale sistemului trebuie ținut cont de faptul că noul sistem solicitat prin acest proiect trebuie să ofere performanță liniară la nivelul platformei Big Data, chiar și în condiții de creștere polinomială sau exponențială a volumului de date. În acest sens, trebuie să utilizeze tehnologii moderne de accesare a volumului mare de date, întâlnite în zona BigData.

Întrebarea nr.13

„Având în vedere faptul ca pentru diferiți producători, mecanismele de licențiere pot fi diferite, recomandăm modificarea cerinței „Infrastructura software a platformei Big Data va fi licențiată, în ansamblul său, pentru utilizarea completă a infrastructurii hardware de procesare alocate, dar nu mai puțin de 709 unități de tip nuclee de procesare (CPU-core).”, astfel încât să existe libertatea de a folosi mecanisme de licențiere diferite, proprii fiecărui furnizor, dar cu obligația de a utiliza infrastructura hardware ce va fi achiziționată în proiectul BigData aflat în consultare de piață. Tot în acest context, autoritatea este rugată să clarifice modalitatea în care a calculat numărul de 709 CPU-core exprimat în cerința de mai sus, în condițiile în care numărul total de core-uri calculat plecând de la cerințele hardware minimale este de: $28 \times 4 \times 8 \text{ cores} + 6 \times 112 \text{ cores} = 1568 \text{ cores}$.”

Răspuns la întrebarea nr.13

Cerința „nu mai puțin de 709 unități de tip nuclee de procesare (CPU-core)” face parte din setul de cerințe pentru una dintre componentele Caietului de sarcini, respectiv „SOFTWARE PENTRU PLATFORMA DE TIP BIG DATA SI SISTEMUL INFORMATIC INTEGRAT”

Autoritatea contractantă reiterează cele precizate la punctul 5 din anunțul de consultare a pieței, încurajând toți potențialii ofertanți să transmită propuneri/sugestii scrise până la termenul limită, 29.11.2022, astfel încât să poată participa ulterior la întâlnirea comună cu toții operatorii economici care participă la consultarea de piață.

Cu stimă,

p. Director general,
Director general adjunct,
Simion ILIE