



ROMÂNIA

Acord de servicii de asistență tehnică rambursabile privind Consolidarea Sistemului Statistic Național (P167217)

REZULTAT Nr. 4.2 a

Recomandări și îndrumări oferite Beneficiarului în utilizarea sistemului IT integrat cu componentele sale, hardware (inclusiv tablete) și software (licențe), inclusiv stocarea datelor, protecția și securitatea datelor pentru desfășurarea activităților dezvoltate în cadrul RAS (GIS, RPL2021, RGA2020, perioade intercenzitare, SICCA)

Septembrie 2022

Revizuit decembrie 2022



Declinarea responsabilității

Acest raport este un produs elaborat de personalul Băncii Mondiale. Constatările, interpretările și concluziile exprimate în acest document nu reflectă neapărat punctele de vedere ale Directorilor Executivi ai Băncii Mondiale sau ale guvernelor pe care aceștia le reprezintă. Banca Mondială nu garantează acuratețea datelor incluse în prezentul document și nu își asumă responsabilitatea pentru orice erori, omisiuni sau inconsecvențe în ceea ce privește informațiile sau răspunderea cu privire la utilizarea sau neutilizarea informațiilor, metodelor, proceselor sau concluziilor prezentate. Granițele, culorile, denumirile și alte informații prezentate pe orice hartă din acest document nu implică nicio apreciere din partea Băncii Mondiale cu privire la statutul juridic al oricărui teritoriu sau la aprobarea sau acceptarea acestor granițe.

Prezentul raport nu reprezintă neapărat poziția Uniunii Europene sau a Guvernului României.

Declarație privind drepturile de autor

Materialul din această publicație este protejat de drepturi de autor. Copierea și/sau transmiterea unor porțiuni din această lucrare fără permisiune poate constitui o încălcare a legilor aplicabile.

Pentru permisiunea de a fotocopia sau imprima orice parte din această lucrare, vă rugăm să trimiteți o solicitare cu informațiile complete către: (i) Institutul Național de Statistică din România (Bd. Libertății nr. 16, Sector 5, București, România) sau (ii) Grupul Băncii Mondiale România (str. Vasile Lascăr nr. 31, etaj 6, București, România).

Acest raport a fost transmis în luna septembrie 2022 și revizuit în luna decembrie 2022 în temeiul Acordului de servicii de asistență tehnică rambursabile privind Consolidarea Sistemului Statistic Național (P167217), semnat între Institutul Național de Statistică din România și Banca Internațională pentru Reconstrucție și Dezvoltare la data de 17 septembrie 2019. Acesta face parte din Rezultatul 4.2 din acordul menționat mai sus.

Cuprins

Listă de acronime.....	4
Lista figurilor	5
Lista tabelelor.....	5
Introducere	6
1. Sistemul IT integrat al INS	9
2. Infrastructura hardware.....	11
3. Prezentare generală asupra aplicațiilor sistemului IT integrat	13
3.1. RGA 2020	13
3.1.1. Colectarea datelor RGA	13
3.1.2. Soluții de monitorizare și raportare în timpul colectării datelor RGA.....	14
3.1.3. Procesul de validare a datelor colectate de RGA.....	15
3.1.4. Diseminarea rezultatelor RGA.....	15
3.1.5. Georeferențierea exploatațiilor agricole în unități statistice tip grilă.....	15
3.2. RPL 2021	16
3.2.1. Colectarea datelor RPL	16
3.2.2. Soluții de monitorizare și raportare în timpul colectării datelor RPL.....	18
3.2.3. Prelucrarea și validarea datelor RPL.....	18
3.2.4. Editarea și producția de date pentru diseminare rezultatelor RPL.....	18
3.2.5. Sectorizarea pentru RPL	19
3.3. Perioade intercenzitare.....	19
3.4. SICCA.....	20
4. Considerații de securitate și testare ale mediului IT	21
4.1. Securitate.....	21
4.2. Testare.....	21
5. Anexe	22

Listă de acronime

APIA	Agenția de Plăți și Intervenție pentru Agricultură
CAPI	Intervievare personală asistată de calculator
CASS	Sistem de sondaj asistat de calculator
CAWI	Intervievare online asistată de calculator
CE	Comisia Europeană
CNP	Cod numeric personal
DTS	Direcții teritoriale de statistică
ESS	Sistemul Statistic European
ESS-MH	Aplicația Metadata Handler a Sistemului Statistic European
EU INSPIRE	Infrastructura pentru informații spațiale în Comunitatea Europeană
GA	Aplicație tip grilă
GAMSO	Modelul general de activitate pentru organizațiile statistice
GSBPM	Modelul generic statistic al proceselor funcționale
GEOLOC	Aplicație software de locație geografică
GIS	Sistem de informații geografice
GPS	Sistemul de poziționare globală
IIS	Servicii de informare pe internet
INS	Institutul Național de Statistică
IT	Tehnologia informației
RAS	Servicii de asistență tehnică rambursabile
RGA	Recensământul General al Agriculturii
RPL2021	Recensământul populației și locuințelor 2021
SDC	Controlul Divulgării Datelor Statistice
SICCA	Soluție IT integrată pentru Arhitectura și Calculul Costurilor Produselor și Activităților
STS	Serviciul de Telecomunicații Speciale
UAT	Unități administrativ-teritoriale
VPN	Rețea virtuală privată

Lista figurilor

Figura 1 – Schema generală a sistemului IT	12
Figura 2 – Diagrama logică a sistemului IT.....	12
Figura 3 – Mediul de producție RGA (GAC Production environment)	13
Figura 4 – Mediul de producție RPL (PHC Production environment)	17

Lista tabelelor

Tabelul 1 - Resurse IT alocate RGA.....	14
Tabelul 2 – Resurse IT alocate RPL	17

Introducere

Acest raport acoperă **recomandări și îndrumări oferite Beneficiarului în utilizarea sistemului IT integrat cu componentele sale, hardware (inclusiv tablete) și software (licențe), inclusiv stocarea datelor, protecția și securitatea datelor pentru desfășurarea activităților dezvoltate în cadrul RAS (GIS), RPL2021, RGA2020, perioade intercenzitare, SICCA**). Acest livrabil face parte din Rezultatul 4.2 din Acordul de servicii de asistență tehnică rambursabile (RAS) privind *Consolidarea Sistemului Statistic Național din România* (P167217). Proiectul este implementat de Institutul Național de Statistică cu sprijinul Băncii Mondiale.

Raportul prezintă o imagine de ansamblu asupra sistemului IT integrat implementat cu componentele sale dezvoltate în cadrul RAS: RGA, RPL, SICCA și sistemul pentru perioadele intercenzitare. Proiectarea sistemului IT integrat de statistică și implementarea efectivă reflectă capacitățile și structura de producție în vederea satisfacerii nevoilor de date statistice ale țării pentru instituțiile publice, sectorul privat și alte organizații, cetățenii în general, din punctul de vedere al calității și fiabilității stabilite de standardele din domeniu.

Sistemul IT integrat hardware și software, dezvoltat în cadrul acestui proiect, a permis implementarea celor două recensăminte cheie (RGA și RPL) și a pregătit configurația pentru activitățile intercenzitare și pentru managementul activităților și costurilor (SICCA) și va asigura o utilizare continuă și durabilă după finalizarea proiectului, după cum urmează:

- Soluția IT (Survey Solutions și R), constând în componente de pregătire, producție și diseminare a datelor, pentru colectarea și prelucrarea datelor RGA2020, se va utiliza pentru cercetări statistice care trebuie efectuate la fiecare 3 ani în perioada intercenzitară în domeniul statisticii agricole (sondaje structurale în agricultură).
- Soluția IT (Survey Solutions și R/Python), constând în componente de pregătire, producție și diseminare a datelor, pentru implementarea RPL2021, se va utiliza în procesul obișnuit de producere a statisticilor continue în ceea ce privește populația rezidentă (prin preluarea modulelor legate de extragerea datelor din surse administrative și integrarea acestora în baza de date pentru statistica populației, precum și prin continuarea colectării de date pentru clădiri și locuințe, incluzând coordonatele geo-spațiale (GIS)).
- Soluția IT (Survey Solutions și R) pentru colectarea datelor prin metode multimodale (CAPI și CAWI), validarea și prelucrarea datelor, editarea, introducerea și producția de tabele, protecția și securitatea datelor, pentru statisticile din perioada intercenzitară, se va utiliza pentru orice activități de colectare a datelor și incluse în procesul statistic, în vederea producerii de statistici cu o frecvență anuală pentru estimări similare celor din cadrul recensământului. Chestionarele electronice inteligente dezvoltate în cadrul proiectului, precum și cunoștințele dobândite prin sesiunile de instruire și atelierele de lucru vor oferi o bază solidă de cunoștințe și abilități cu privire la proiectarea viitoare a oricăror astfel de sondaje.
- Platforma GIS, utilizată inițial în procesul de colectare și prelucrare a datelor pentru RPL2021, va fi ulterior integrată în procesul obișnuit de producție statistică pentru populația rezidentă, clădiri și locuințe, statistici care vor încorpora coordonatele geo-spațiale (GIS).

- Metodologia de evaluare și asigurare a calității și acoperirii datelor colectate, precum și pentru protecția/securitatea datelor colectate, dezvoltată în cadrul proiectului pentru RGA2020 și RPL2021, se va aplica (împreună cu adaptările corespunzătoare) pentru cercetarea statistică curentă condusă de INS.
- Toate soluțiile sunt construite în jurul sistemelor de sondaj open-source asistat de calculator, Survey Solutions, ale Băncii Mondiale (CASS), precum și prin intermediul aplicațiilor și scripturilor R (Shiny) (și Python, după caz). Codurile sursă pentru toate aceste instrumente au fost partajate cu INS, astfel încât acesta să poată modifica aceste scripturi și să le ajusteze la nevoile viitoare.
- Aplicația SICCA (Soluția IT integrată pentru Arhitectura și Calculul Costurilor Produselor și Activităților) dezvoltată în conformitate cu standardele statistice ale Modelului general de activitate pentru organizațiile statistice (GAMSO) și ale Modelului generic statistic al proceselor funcționale (GSBPM) va fi instrumentul utilizat zilnic pentru fiecare angajat din INS și DTS pentru monitorizarea și evaluarea activităților desfășurate în vederea determinării costurilor cu resursele umane, timp și resurse financiare pentru producția de produse statistice, dar și pentru fiecare activitate sau pentru grupele de activități desfășurate;

Luând în considerare toate funcționalitățile de mai sus și utilizarea lor viitoare, raportul rezumă asistența tehnică acordată pentru utilizarea sistemului IT integrat proiectat și implementat prin proiect, respectiv:

- Infrastructura IT
 - *Rezultatul 4.1b - Recomandări de elaborare a documentației tehnice în vederea organizării achiziției unui sistem IT integrat, inclusiv hardware (inclusiv tablete) și software (licențe), inclusiv stocarea datelor, protecția și securitatea datelor pentru desfășurarea activităților dezvoltate în cadrul RAS (GIS, RPL2021, RGA2020, perioade intercenzitare, SICCA).*
- Implementarea RPL
 - *Rezultatul 4.1c - Plan documentat pentru sistemul integrat pentru implementarea RPL2021 (detaliază modul în care se va realiza implementarea infrastructurii IT pentru RPL2021);*
 - *Rezultatul 7c - Recomandări și bune practici pentru implementarea unui sistem de management al datelor pentru datele geo-spațiale și sectorizarea efectivă a RPL2021;*
 - *Rezultatul 13 - Recomandări și îndrumări privind planul documentat pentru sistemul integrat pentru implementarea RPL2021 (inclusiv managementul utilizatorilor, criptarea ID, recuperarea în caz de dezastru, menținerea disponibilității sistemului în caz de defecțiune).*
- Implementarea RGA
 - *Rezultatul 14 – Recomandări și bune practici pentru implementarea unui sistem IT funcțional pentru RGA2020 (Rapoarte privind (i) logistica legată de colectarea datelor conform combinației optime identificate; și (ii) pilotarea aplicației IT a RGA cu raport și asistență, precum și consiliere privind cele mai bune practici și îndrumări pentru procesele și acțiunile de întreținere a sistemului IT în timpul producției de lansare), inclusiv raportul privind 2 (două) ateliere de 5 (cinci) zile pentru statisticieni, privind metodologiile și instrumentele pentru transformarea chestionarelor în chestionare electronice statistice inteligente;*

- *Rezultatul 7b - Recomandări, bune practici și îndrumări în dezvoltarea unei metodologii de personalizare a unei soluții IT pentru georeferențierea exploatațiilor agricole în unități statistice tip grilă în cadrul RGA2020.*
- Implementarea SICCA
 - *Rezultatul 11 - Recomandări, îndrumări și bune practici pentru dezvoltarea și personalizarea unei Soluții IT integrate pentru Arhitectura și Calculul Costurilor Produselor și Activităților – SICCA.*
- Implementarea sondajelor intercenzitare
 - *Rezultatul 4.2b - Recomandări și îndrumări privind planul documentat (producția de lansare) pentru sistemul integrat pentru perioade intercenzitare.*

și include câteva recomandări (aspecte) care vor fi puse în practică de către specialiștii INS, partenerii acestora (de exemplu, STS) și furnizorii de infrastructură IT, în menținerea și exploatarea sistemului într-o manieră/condiție optimă, respectiv securitatea mediului IT găzduit de INS și a mediului IT găzduit de STS pentru colectarea datelor, ori de câte ori este cazul, precum și procesele de testare a performanței funcționale și a securității care urmează să fie efectuate. Protecția și securitatea datelor, aplicate pentru datele colectate la RPL, sunt detaliate în *Rezultatul 10c: Setul de instrumente proiect ale recensământului statistic care utilizează metode multimodale pentru a promova protecția și securitatea datelor și completată cu Rezultatul 10b: Asistență tehnică și recomandări de bune practici pentru SDC, confidențialitatea datelor și modalități de securizare a microdatelor și a datelor agregate.*

Utilizarea sistemului IT integrat în condiții optime necesită abilități pe domenii specifice ale personalului INS care pot fi dobândite prin implementarea consecventă a planului multianual de formare propus în *Rezultatul 12 – Elaborarea unui plan multianual de nevoi de formare (2021-2024) pentru INS, inclusiv DTS pe baza nevoilor de formare și, de la caz la caz, cu sprijinul unor servicii specializate în parteneriat (de exemplu, STS) sau externalizate pe măsură ce apar situațiile viitoare.*

1. Sistemul IT integrat al INS

Sistemul IT integrat implementat de Institutul Național de Statistică (INS) are în vedere susținerea procesului de producție statistică pentru următorii 10 ani și cuprinde următoarele componente funcționale, inclusiv elemente de aplicație și infrastructură subiacente, respectiv:

1. **Stocarea, prelucrarea și diseminarea datelor pentru Recensământul General al Agriculturii 2020 (RGA2020), inclusiv:**
 - a. Soluția de georeferențiere a exploatațiilor agricole în unități statistice de tip grilă, pe baza amplasamentului, astfel încât să se poată realiza asocierea între suprafața agricolă utilizată și codul de grilă corespunzător, conform prevederilor regulamentului privind statisticile integrate la nivelul fermei; soluția va aborda diferite exploatații care beneficiază de subvenții și sunt incluse în lista APIA (pentru care există coordonate GIS din sursa APIA) și exploatațiile de mici dimensiuni care nu sunt incluse în lista APIA (pentru care trebuie colectate coordonatele GIS).
 - b. Prelucrarea datelor din surse administrative și conversia acestora în baza de date RGA2020, precum și completarea prealabilă a acestor informații în chestionare individuale. Chestionarele individuale au fost completate în prealabil conform informațiilor identificate în sursele administrative (APIA, liste de sectorizare pe baza Registrului de Agricultură etc.).
 - c. Aplicațiile software pentru asigurarea confidențialității datelor și pentru aplicarea tehnicilor de control al divulgării statistice (SDC).
 - d. Întregul sistem include, de asemenea, colectarea datelor CAPI pe baza chestionarelor dezvoltate cu Survey Solution (serverele de colectare a datelor au fost găzduite în STS).
2. **Stocarea, prelucrarea și diseminarea datelor pentru Recensământul Populației și Locuințelor 2021 (RPL2021), inclusiv:**
 - a. Prelucrarea datelor din surse administrative și conversia acestora în baza de date RPL2021 și completarea prealabilă a chestionarelor individuale (chestionarele individuale vor fi completate în prealabil conform informațiilor identificate în sursele administrative).
 - b. Soluția de georeferențiere a locuințelor/clădirilor destinate a fi ocupate de gospodării, unități colective și clădiri (inclusiv pentru cele cu alte destinații) în care cel puțin o persoană își are reședința obișnuită la data de referință a RPL2021.
 - c. Sectorizarea
 - d. Aplicațiile software pentru asigurarea confidențialității datelor și pentru aplicarea tehnicilor de control al divulgării statistice (SDC).
 - e. Colectarea datelor CAPI pe baza chestionarelor dezvoltate cu Survey Solution (găzduite de STS).
 - f. Colectarea datelor CAWI pe baza chestionarelor dezvoltate cu Survey Solution (găzduite de STS).
3. **Soluțiile de prelucrare a datelor în perioadele intercenzitare pentru a produce anual date/indicatori similari/similare celor din cadrul recensământului (pentru clădiri, locuințe și persoane).**
4. **Sondajele eșantion (Survey Solutions) - CAPI și CAWI găzduite în infrastructura IT a INS.**
5. **Soluția IT integrată pentru Arhitectura și Calculul Costurilor Produselor și Activităților (SICCA).**

- 6. Sistemul de informații geografice (GEOLoc) – livrat de un furnizor specializat și care nu face parte din acest proiect.**

2. Infrastructura hardware

Descrierea tuturor componentelor infrastructurii hardware este disponibilă în raportul *Rezultatului 4.1b - Recomandări pentru elaborarea documentației tehnice în vederea organizării achiziției unui sistem IT integrat, inclusiv hardware (inclusiv tablete) și software (licențe), inclusiv stocarea, protecția și securitatea datelor pentru desfășurarea activităților dezvoltate în cadrul RAS (GIS, RPL2021, RGA2020, perioade intercenzitare, SICCA).*

Sistemul IT integrează o structură de hardware capabilă să răspundă nevoilor actuale și viitoare ale procesului de producție statistică, având în vedere următoarele cerințe:

- puterea de procesare - capacitatea proceselor de calcul prin eveniment sau sondaj
- numărul de utilizatori simultani
- timpul de răspuns
- capacitatea de stocare a datelor
- capacitatea de transfer de date
- sistemul de operare și licențele

Sistemul se bazează pe trei clustere cu noduri identice:

- Virtualizare – procesare de date (procesor și unitate grafică) – 2 noduri
- Virtualizare - Windows (Survey Solutions web – IIS, GEOLOC - Server și Portal GIS) - 2 noduri
- Virtualizare – Linux – 4 noduri

Necesarul de stocare este satisfăcut de un dispozitiv extern de stocare. Serverele se conectează la stocare prin adaptoare redundante cu baza la gazdă (HBA) de 16 Gbps și la rețea prin carduri redundante de interfață la rețea (NIC) de 10 Gbps. Sistemul integrat se conectează la LAN-ul INS prin intermediul GBIC-urilor cu fibră optică pe Dispozitivul de securitate integrat și la mediul găzduit de STS prin VPN Site-to-Site terminat, de asemenea, pe Dispozitivul de securitate integrat, confirmat de STS.

Toate mediile sunt conectate la aceeași rețea (inclusiv VLAN-uri, acolo unde este necesar), precum și la echipamentul de stocare.

Utilizatorii INS (locali sau de la departamentele teritoriale de statistică conectate prin INS WAN) ar avea acces la medii bazate pe LAN, deși, din considerente de securitate, ar trebui să se conecteze prin firewall/dispozitivul de securitate integrat.

Schema generală și diagrama logică a sistemului IT integrat sunt prezentate în figurile de mai jos (a se vedea descrierea detaliată din Rezultatul 4.1b menționat mai sus).

Figura 1 – Schema generală a sistemului IT

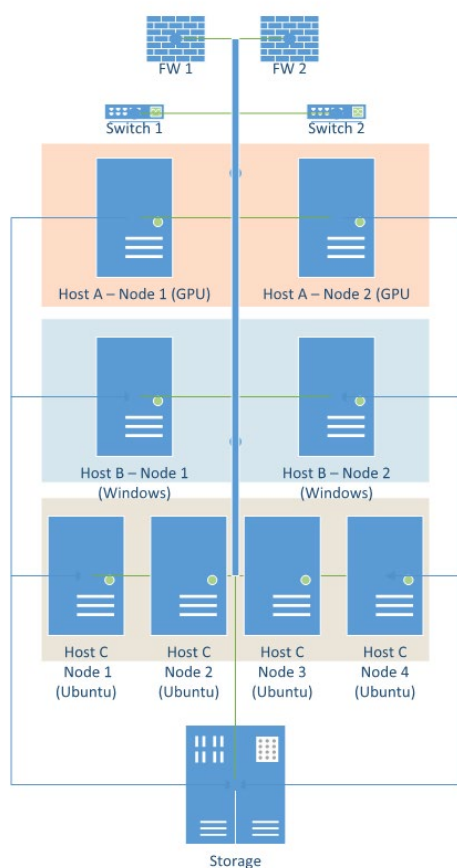
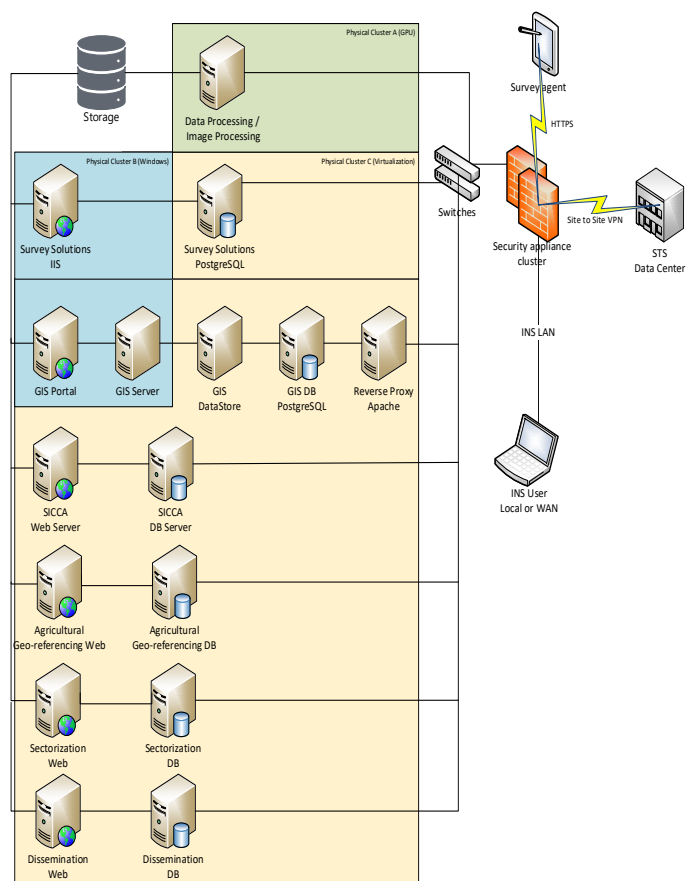


Figura 2 – Diagrama logică a sistemului IT



Descrierea detaliată a sistemului IT este prezentată în *Rezultatul 4.1b - Recomandări pentru elaborarea documentației tehnice în vederea organizării achiziției unui sistem IT integrat, inclusiv hardware (inclusiv tablete) și software (licențe), inclusiv stocarea, protecția și securitatea datelor pentru desfășurarea activităților dezvoltate în cadrul RAS (GIS, RPL2021, RGA2020, perioade intercenzitare, SICCA).*

3. Prezentare generală asupra aplicațiilor sistemului IT integrat

Sistemul IT integrat este compus din diverse componente hardware și software (licențe software, aplicații, baze de date) care, împreună, asigură infrastructura informațională necesară atât pentru activitățile de colectare și prelucrare a datelor, cât și pentru alte activități complementare pe care le desfășoară INS.

Aplicațiile și soluțiile pregătite sunt prezentate în continuare pe scurt. Descrierea detaliată a acestora este prezentată în rezultatele corespunzătoare menționate pentru fiecare componentă.

3.1.RGA 2020

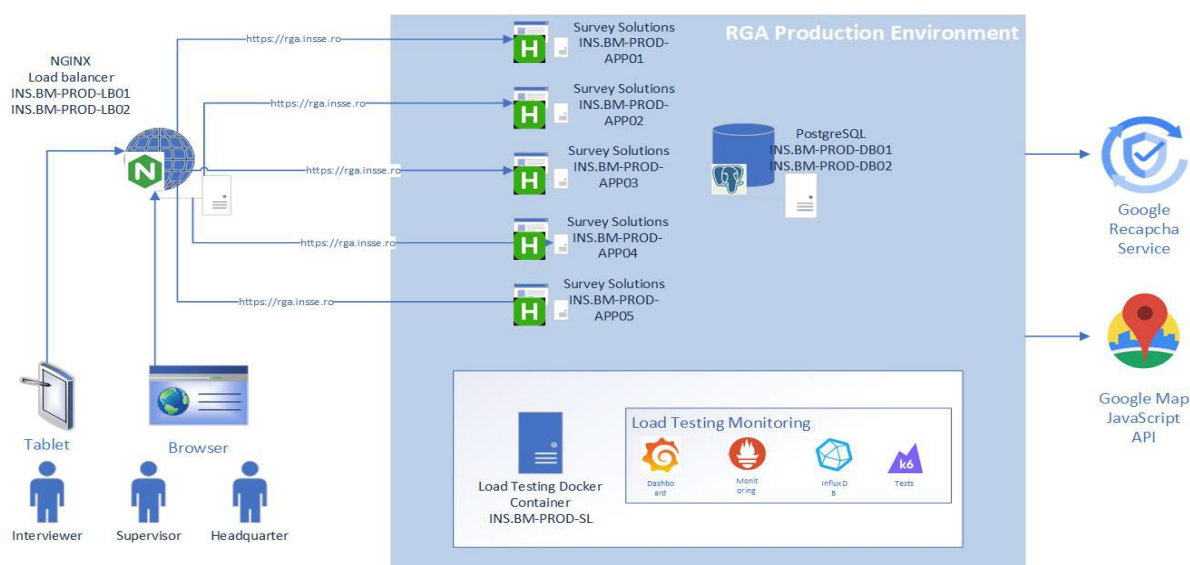
Proiectarea și implementarea RGA 2020 face obiectul *Rezultatului 14 – Recomandări și bune practici pentru implementarea unui sistem IT funcțional pentru RGA2020 (Raport privind (i) logistica legată de colectarea datelor conform combinației optime identificate; și (ii) pilotarea aplicației IT a RGA cu raport și asistență, precum și consiliere privind cele mai bune practici și îndrumări pentru procesele și acțiunile de întreținere a sistemului IT în timpul producției de lansare), inclusiv raportul privind 2 (două) ateliere de 5 (cinci) zile pentru statisticieni privind metodologiile și instrumentele pentru transformarea chestionarelor în chestionare electronice statistice inteligente” și al Rezultatului 7b: Recomandări, bune practici și îndrumări în dezvoltarea unei metodologii de personalizare a unei soluții IT pentru georeferința exploatațiilor agricole în unități statistice tip grilă RGA2020.*

În ceea ce privește acest rezultat, sunt incluse mai jos prezentări scurte ale arhitecturii sistemului de colectare a datelor, soluțiilor de monitorizare și raportare utilizate în timpul colectării datelor, procesului de validare a soluțiilor de date colectate și diseminare a rezultatelor RGA, inclusiv distribuirea tip grilă a fermelor.

3.1.1. Colectarea datelor RGA

Sistemul actual de colectare a datelor RGA a fost construit pe o arhitectură cu trei niveluri, constând într-o pereche de 2 echilibratori de sarcină care deservesc cererile web, 5 noduri de server de aplicații și o pereche de 2 noduri de baze de date. Valorile semnificative sunt captate de la toate nodurile și trimise către un nod de monitorizare.

Figura 3 – Mediul de producție RGA (GAC Production environment)



Sursa: Autorii

Nodurile sunt mașini virtuale găzduite la amplasamentul STS. Următoarele resurse au fost dedicate fiecărui nod:

Tabelul 1 - Resurse IT alocate RGA

Denumire	Spațiu necesar	SO oaspete	Dimensiune memorie	Procesoare	Adresă IP
INS.BM-PROD-APP01	82.09 GB	Microsoft Windows Server 2019 (64-bit)	32 GB	8	
INS.BM-PROD-APP02	82.09 GB	Microsoft Windows Server 2019 (64-bit)	32 GB	8	
INS.BM-PROD-APP03	82.09 GB	Microsoft Windows Server 2019 (64-bit)	32 GB	8	
INS.BM-PROD-APP04	82.09 GB	Microsoft Windows Server 2019 (64-bit)	32 GB	8	
INS.BM-PROD-APP05	82.09 GB	Microsoft Windows Server 2019 (64-bit)	32 GB	8	
INS.BM-PROD-DB01	1050 GB	Ubuntu Linux 20.04 (64-bit)	128 GB	32	
INS.BM-PROD-DB02	1050 GB	Ubuntu Linux 20.04 (64-bit)	128 GB	32	
INS.BM-PROD-LB01	58.08 GB	Ubuntu Linux 20.04 (64-bit)	8 GB	4	
INS.BM-PROD-LB02	58.08 GB	Ubuntu Linux 20.04 (64-bit)	8 GB	4	
INS.VM-PROD-SL	98.09 GB	Ubuntu Linux 20.04 (64-bit)	48 GB	16	

Disponibilitatea ridicată, gestionarea jurnalelor și recuperarea în caz de dezastru au fost implementate și sunt descrise în Rezultatul 14 (i).

Anexa 1 – Sistemul IT de colectare a datelor RGA – raportul as-built oferă toate detaliile despre sistemul implementat.

3.1.2. Soluții de monitorizare și raportare în timpul colectării datelor RGA

Procesul de monitorizare și raportare pentru colectarea datelor a reprezentat o sarcină zilnică realizată pe toată perioada RGA (10 mai – 31 iulie 2021) ca un efort comun al echipelor INS și BM.

Soluțiile oferite au constat în rapoarte pentru:

- a) Monitorizare zilnică
 - Monitorizarea colectării datelor
 - Monitorizarea colectării datelor de către UAT, supraveghetori și recenzori
 - Monitorizarea UAA, bovine și porcine de către UAT (la nivel de județ)
- b) Activitatea personalului de recensământ
 - raport de activitate pentru personalul de recensământ
 - raport de activitate și de plată pentru recenzor
 - raport de activitate și de plată pentru recenzorul-șef
 - raport de activitate și de plată pentru coordonator

Anexa 2 – Exemple de rapoarte de monitorizare zilnică și Anexa 3 – Exemple de rapoarte de activitate pentru personalul de recensământ din Rezultatul 14 (i) oferă detalii despre rapoartele de mai sus.

3.1.3. Procesul de validare a datelor colectate de RGA

Scopul procesului de validare a fost acela de a sprijini transmiterea rezultatelor către Eurostat, în formatul predefinit de fișier Eurofarm (același pentru toate statele membre UE) și diseminarea rezultatelor către alți utilizatori (naționali/internaționali): publicații (tabele predefinite) și tabele produse în urma solicitărilor ad-hoc.

Instrumentele de validare a datelor colectate au constat în:

- script R privind acceptarea și respingerea automată a chestionarului colectat în timpul RGA
- tabele de control (cele 31 de tabele elaborate de INS și echipa BM) utilizate pentru prelucrarea-validarea datelor și pentru a susține compararea cu alte surse de date (de exemplu: RGA2010, ASA 2016, APIA);

Aplicația pentru Tabelele de control RGA2020 este funcțională în mediul IT al INS, „Server Procesare Date” - link http://xx.x.xxx.xx/GAC_Dashboard.

Anexa 4 – Script pentru acceptarea/respingerea automată a chestionarelor și Anexa 6 – Exemplu de tabele de control al datelor (TC nr. 0) din Rezultatul 14(i) oferă detalii despre procesul de validare.

3.1.4. Diseminarea rezultatelor RGA

Pentru diseminare, INS și BM au analizat posibilitatea unei soluții ca interfață aplicație - bază de date care să permită personalului INS să analizeze datele RGA, să le filtreze pe niveluri teritoriale și să le grupeze pe mai multe straturi. Drept urmare, echipa BM a propus și susținut dezvoltarea unei aplicații web Shiny¹ bazată pe limbajul de programare R, permițând integrarea, wrangling și reproductibilitatea rapidă a datelor (a se vedea <https://shiny.rstudio.com/>).

Aplicația urmează să integreze rapid datele brute (la nivel de înregistrare/exploatație agricolă) furnizate de RGA, să rezume variabilele cheie la nivel de UAT, să grupeze variabilele și să permită livrarea rapidă a exporturilor în cazul solicitărilor de date de la INS sau utilizatorii externi de date.

Aplicația pentru diseminarea datelor RGA este implementată pe „Server Procesare Date” INS - link <http://xx.x.xxx.xx/rga-diseminare/>, iar codul sursă a fost partajat cu INS în *Anexa 7 – Script pentru diseminarea datelor RGA* din cadrul Rezultatului 14(i).

3.1.5. Georeferențierea exploatațiilor agricole în unități statistice tip grilă

Georeferențierea exploatației agricole în unități statistice tip grilă este procesul de asociere dintre locația (GPS) exploatației agricole și codul grilei corespunzătoare, conform reglementărilor specifice² relevante pentru statisticile privind fermele integrate, inclusiv RGA2020 și reglementările UE³.

Aplicația GA (aplicația de tip grilă) permite vizualizarea unor seturi mari de date de recensământ, precum și a analizelor spațiale corespunzătoare, fie ca puncte, fie ca seturi de date tip grilă. Este destinată să sprijine munca statisticienilor cu cunoștințe de bază GIS. Aplicația este disponibilă pe „Server Procesare Date” al INS (link: xx.x.xxx.xx/rga-grid/), iar codul sursă, scris în R, a fost partajat cu INS.

¹ Shiny este un pachet R care facilitează construirea de aplicații web interactive direct din R.

² D2.8.1.2 Specificații de date privind sistemele geografice tip grilă – Îndrumări tehnice; INSPIRE, 2014; Directiva 2007/2/CE a Parlamentului European și a Consiliului din 14 martie 2007 de instituire a unei infrastructuri pentru informații spațiale în Comunitatea Europeană (INSPIRE)

³ Regulamentul UE 2018/1098

Anexele *Rezultatului 7b - Recomandări, bune practici și îndrumări în dezvoltarea unei metodologii de personalizare a unei soluții IT pentru georeferențierea exploatațiilor agricole în unități statistice tip grilă RGA2020* furnizează manualul de utilizare complet al aplicației, precum și scriptul R pentru potrivirea datelor și corectarea coordonatelor GPS.

3.2. RPL 2021

Proiectarea sistemului de colectare a datelor RPL 2021 a făcut obiectul *Rezultatului 4.1c – Plan documentat pentru sistemul integrat pentru implementarea RPL2021 (detaliază modul în care se va realiza implementarea infrastructurii IT pentru RPL2021)*. Pe baza acestuia, a fost realizată implementarea pentru sistemul actual de colectare a datelor RPL, iar sistemul as-built este descris în *Rezultatul 13 – Recomandări și îndrumări privind planul documentat pentru sistemul integrat pentru implementarea RPL2021 (inclusiv managementul utilizatorilor, criptarea ID, recuperarea în caz de dezastru, menținerea disponibilității sistemului în caz de defecțiune)* și include, de asemenea, soluțiile de diseminare (hipercuburi, distribuirea datelor RPL sub formă de grilă și diseminarea tabelară a datelor RPL). Soluțiile IT dedicate RPL sunt completate de soluția de sectorizare descrisă în raportul *Rezultatului 7c – Recomandări și bune practici pentru implementarea unui sistem de management al datelor pentru datele geo-spațiale și sectorizarea efektivă a RPL2021*.

Sistemul IT integrat al RPL cuprinde următoarele soluții IT:

- Colectarea datelor:
 - o Survey Solutions pentru procesul de colectare și soluția de monitorizare a datelor CAWI și CAPI
 - o Soluție de autoînregistrare pentru etapa CAWI a RPL
- Prelucrarea și validarea datelor
- Editarea și producția datelor pentru diseminare: hipercuburi, distribuire tip grilă și diseminare tabelară a datelor RPL
- (Sectorizare pentru colectarea datelor)

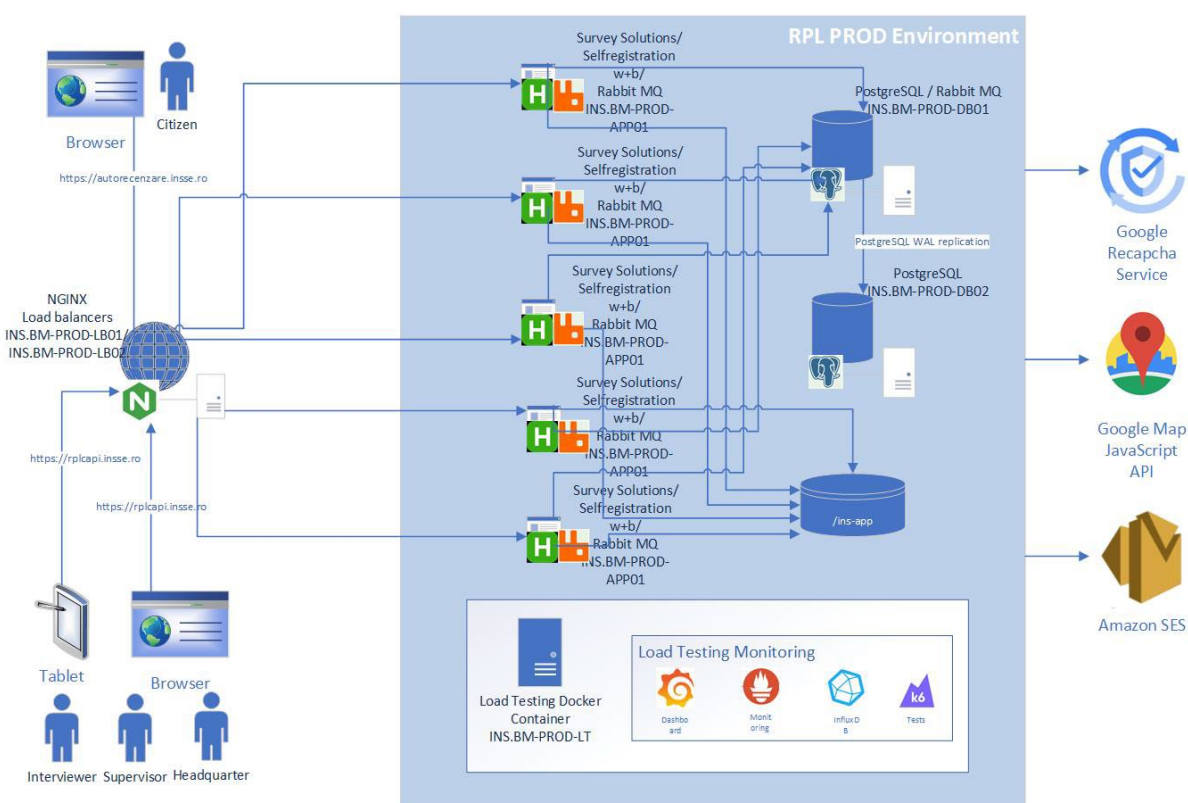
Acest rezultat oferă doar o scurtă descriere a soluției iar orice detalii pot fi regăsite în raportul corespunzător indicat.

3.2.1. Colectarea datelor RPL

Sistemul de colectare a datelor include aplicația Survey Solutions pentru procesul de colectare a datelor CAWI și CAPI, precum și soluția de autoînregistrare pentru etapa CAWI a RPL (a se vedea Figura cu mediul de producție RPL).

Sistemul actual de colectare a datelor RPL a fost construit pe o arhitectură cu trei niveluri, constând într-o pereche de 2 echilibratori de sarcină care deservesc cererile web, 5 noduri de server de aplicații și o pereche de 2 noduri de baze de date. Valorile semnificative sunt captate de la toate nodurile și trimise către un nod de monitorizare.

Figura 4 – Mediul de producție RPL (PHC Production environment)



Sursa: Autorii

Nodurile sunt mașini virtuale găzduite la amplasamentul STS. Următoarele resurse au fost dedicate fiecărui nod:

Tabelul 2 – Resurse IT alocate RPL

Denumire	Spațiu necesar	SO oaspete	Dimensiune memorie	Procesoare	Adresă IP
INS.BM-PROD-APP01	1000 GB	Microsoft Windows Server 2019 (64-bit)	96 GB	24	
INS.BM-PROD-APP02	82.09 GB	Microsoft Windows Server 2019 (64-bit)	96 GB	24	
INS.BM-PROD-APP03	82.09 GB	Microsoft Windows Server 2019 (64-bit)	96 GB	24	
INS.BM-PROD-APP04	82.09 GB	Microsoft Windows Server 2019 (64-bit)	96 GB	24	
INS.BM-PROD-APP05	82.09 GB	Microsoft Windows Server 2019 (64-bit)	96 GB	24	
INS.BM-PROD-DB01	1050 GB	Ubuntu Linux 20.04 (64-bit)	256 GB	48	
INS.BM-PROD-DB02	1050 GB	Ubuntu Linux 20.04 (64-bit)	256 GB	48	
INS.BM-PROD-LB01	58.08 GB	Ubuntu Linux 20.04 (64-bit)	32 GB	16	
INS.BM-PROD-LB02	58.08 GB	Ubuntu Linux 20.04 (64-bit)	32 GB	16	
INS.VM-PROD-SL	98.09 GB	Ubuntu Linux 20.04 (64-bit)	48 GB	16	

Descrierea detaliată a sistemului IT de colectare a datelor RPL este parte a Rezultatului 13, ca raport de implementare (as-built).

Monitorizarea, procesarea și validarea colectării datelor RPL se realizează folosind scripturi limbajul software statistic R implementate în mediul IT al INS, *„Server Procesare Date”* (actual IP xx.x.xxx.xx).

3.2.2. Soluții de monitorizare și raportare în timpul colectării datelor RPL

Procesul de monitorizare și raportare pentru colectarea datelor a reprezentat o sarcină zilnică realizată pe toată perioada RPL (15 martie – 31 iulie 2022) ca un efort comun al echipelor INS și BM. Soluțiile oferite au constat în rapoarte pentru:

- Monitorizarea colectării datelor în faza CAWI (15 martie – 27 mai) și în faza CAPI (31 mai – 31 iulie);
- Monitorizarea colectării datelor pe UAT, recenzori și recenzori șefi;

Analiza paradelor poate fi realizată cu ajutorul unei aplicații dedicate care este disponibilă pentru personalul INS (nu sunt necesare acreditări) pe *„Server Procesare Date”* calea `phc__censustools`.

3.2.3. Prelucrarea și validarea datelor RPL

Prelucrarea și validarea datelor RPL se realizează folosind limbajul statistic de programare R pentru scripturi, implementat în mediul IT al INS.

Scopul procesului de validare a fost de a pregăti datele finale RPL și de a sprijini diseminarea rezultatelor, inclusiv transmiterea acestora către Eurostat.

Instrumentele de validare a datelor colectate au constat în script-ul R privind acceptarea automată a chestionarului colectat în timpul RPL. Validarea post-recenzare a fost pregătită de echipa INS, împreună cu tratarea SDC a datelor finale folosind metoda(ele) selectată(e) dintre cele descrise în Rezultatul 10b: *Asistență tehnică și recomandări de bune practici pentru SDC, confidențialitatea datelor și modalități de securizare a microdatelor și a datelor agregate*.

Detaliile scripturilor R fac obiectul Rezultatului 13 și sunt prezentate în raportul acestuia.

3.2.4. Editarea și producția de date pentru diseminare a rezultatelor RPL

Diseminarea rezultatelor RPL cuprinde trei soluții specifice.

Hipercuburile de date RPL, care trebuie publicate obligatoriu pe portalul Eurostat, respectă liniile directoare furnizate în acest sens de această organizație și vor fi preluate din depozitul de date RPL organizat în mediul IT al INS (echipa INS pregătește mediul IT și instalează aplicațiile necesare).

Distribuirea pe grilă a datelor RPL permite diseminarea a 20 de indicatori (13+7) pe grile (INSPIRE) de 1 kmp pe întreg teritoriul României. Georeferențierea datelor PHC2021 în unități statistice pe grilă este procesul de asociere între locația (GPS) a variabilelor colectate și codul grilei corespunzător, conform Regulamentului UE nr. 1799/2018. Astfel se produce și se transmite către Eurostat un tabel bidimensional care încrucișează setul de celule ale grilei cu selecția de 13 (+7) categorii de indicatori pe tema recensământului specificate în Anexa I la Regulament. Datele RPL diseminate pe grilă sunt pregătite cu ajutorul software statistic R/Shiny.

Diseminarea rezultatelor RPL prin intermediul tabelelor dinamice, este o aplicație scrisă în R folosind cadrul Shiny utilizat pentru diseminarea datelor RPL agregate/grupate la nivel de UAT, județ, regiune și țară.

Detaliile scripturilor R fac obiectul Rezultatului 13 și sunt prezentate în cadrul acestuia și rulează pe *Server Procesare Date*.

3.2.5. Sectorizarea pentru RPL

Sectorizarea în contextul unui recensământ al populației și al locuințelor se referă la procesul de creare a unor segmente ale ariei geografice de recensământ, cu un număr similar de unități de recensământ în interiorul fiecăruia dintre aceste segmente. Pentru RPL2021, aceasta înseamnă că întreaga zonă de recensământ (= teritoriul geografic al României) trebuie segmentată în zone geografice cu o populație de unități locative (de exemplu, 100; 200; 600; $\pm$ 20%). Soluția de sectorizare a fost dezvoltată ca script în R și cu sprijinul unor funcții C++ personalizate și a fost implementată în mediul IT al INS (și anume, serverul de analiză). Scriptul poate fi accesat/rulat ca proiect RStudio de sine stătător prin serverul RStudio disponibil pe *Server Procesare Date* la calea *arhiva2/ins_sectorization*.

Codul de script și metoda detaliată de utilizare pentru a produce segmente sunt prezentate în raportul Rezultatului 7c - *Recomandări și bune practici pentru implementarea unui sistem de management al datelor pentru datele geo-spațiale și sectorizarea efektivă a RPL2021*.

3.3. Perioade intercenzitare

În perioadele intercenzitare, INS va actualiza anual informațiile necesare obținerii datelor recensământului asupra populației și locuințelor, în principal prin colectarea datelor de la respondenți prin Survey Solutions, folosind moduri CAWI și/sau CAPI și din surse administrative, prin actualizarea registrelor. Astfel se vizează un sistem care să permită obținerea de informații actualizate privind mărimea populației și caracteristicile acesteia la momentul extragerii sau la un moment de referință specificat, respectiv de a avea o listă structurată de unități, fiecare dintre acestea conținând un număr de atribute, un mecanism de actualizare regulată și o componentă de producere a rezultatelor.

Folosirea registrelor pentru a produce date pentru populație și locuințe se bazează pe un algoritm de actualizare regulată în perioada intercenzitară. Aceasta va folosi variabile și date de dezagregare din 22 de surse administrative (estimate ca potențiale pentru utilizare) și cu o anumită cronologie (data și frecvența – lunar și sau semestrial, semestrial, anual) în funcție de în funcție de evenimentul statistic și de atributele de actualizare.

Algoritmul este dezvoltat pe R Statistical Software, R Shiny și este rulat pe *Server Procesare Date*.

Instalarea sistemului de colectare a datelor Survey Solutions pentru perioadele intercenzitare pe mediul IT al INS a fost realizată de experții INS.

Registrul populației face obiectul Rezultatului 4.2b – *Recomandări și îndrumări privind planul documentat (lansare în producție) pentru sistemul integrat pentru perioadele intercenzitare* și este prezentat în raportul la acesta.

3.4. SICCA

Proiectarea, instalarea, configurarea și optimizarea SICCA face obiectul „Recomandărilor, orientărilor și bunelor practici pentru dezvoltarea și personalizarea unei Soluții IT Integrate pentru Arhitectura și Calculul Costurilor Produselor și Activităților INS – SICCA”.

Scopul SICCA este de a sprijini managementul costurilor INS și se bazează pe trei obiective: (i) calcularea costurilor directe și indirecte pentru toate produsele statistice la nivel de activități generice, componente/subcomponente și proiecte, standardizate conform GSBPM/GAMSO; (ii) optimizarea costurilor de producție statistică și determinarea costurilor; și (iii) îmbunătățirea bugetului INS pe baza dovezilor/înregistrărilor.

SICCA este proiectată și dezvoltată pentru a face față mai multor cerințe tehnice. Aplicația este accesibilă simultan tuturor celor 2.000 de angajați INS (500 de angajați la sediul central și 1.500 de angajați în direcțiile teritoriale de statistică, amplasate în toată țara). Este o aplicație web care poate fi accesată doar cu un browser de internet, fără a necesita software suplimentar precum plug-in-uri sau altele similare, cu o interfață accesibilă prin cele mai utilizate browsere de internet: Microsoft Edge, Internet Explorer, Firefox, Chrome. Este instalată în LAN a INS și utilizatorii se conectează prin WAN a INS.

Toate diferitele componente software ale aplicației au fost proiectate și implementate folosind tehnologii cu sursă deschisă. Stack-ul de tehnologie este format din:

- Sistem de operare – este acceptat orice sistem de operare de renume, inclusiv Linux, OSX și Windows;
- Limbaje de programare – PHP (backend), Javascript/HTML/CSS (frontend);
- Server de baze de date – MySQL/MariaDB/PostgreSQL;
- Server web – Apache/NGINX;

Detalii despre funcționalitățile aplicației sunt disponibile în anexele furnizate împreună cu Rezultatul 11 - *Recomandări, îndrumări și bune practici pentru dezvoltarea și personalizarea unei Soluții IT integrate pentru Arhitectura și Calculul Costurilor Produselor și Activităților INS – SICCA*, respectiv: *Anexa 1 - Cerințe preliminare (inițiale) comerciale și tehnice; Anexa 2 - Manualul administratorului web; Anexa 3 – Manualul de utilizare; Anexa 4 – Manualul administratorului (backend); Anexa 5 – Baza de date și nomenclaturi; Anexa 6 – Codul sursă al aplicației.*

4. Considerații de securitate și testare ale mediului IT

4.1. Securitate

În ceea ce privește securitatea fizică a mediului IT, INS are propriile proceduri și decide în consecință pentru aplicarea acestora. Acestea sunt completate cu acțiunile logice de securitate (de exemplu, soluție VPN, configurații de sisteme etc., deja existente) pentru a asigura securitatea solidității mediului IT.

Securitatea fizică a infrastructurii IT găzduite în STS și dedicată colectării datelor (și utilizată în proiectul actual) este acoperită de procedurile și protocoalele organizaționale STS. Accesul fizic la acest mediu este permis numai personalului autorizat STS.

Accesul logic este permis numai persoanelor care au un nivel de autorizare eligibil și este posibil doar prin VPN cu autentificare cu doi factori.

Aplicațiile web expuse sunt supuse unei evaluări de securitate efectuate de o echipă dedicată STS. În procesul de colectare a datelor pentru recensămintele efectuate, raportul de evaluare a fost prezentat INS de către echipa STS. Recomandările privind funcționalitatea aplicațiilor și configurarea aplicațiilor sunt aplicate în mod corespunzător.

4.2. Testare

Testarea funcțională este efectuată de echipa INS și scopul acesteia este de a asigura că toate funcționalitățile proiectate sunt acoperite de verificări eficiente.

În procesul de colectare a datelor pentru recensământ, testarea performanței a fost susținută de activitățile experților echipei BM pentru a se asigura că sunt respectate capacitatea planificată și timpii de răspuns conveniți. *Anexa 1* descrie planificarea detaliată a testelor de performanță, execuția și rezultatele.

Testarea de securitate pentru componentele expuse la internet ale sistemului face obiectul activităților experților STS și sunt acoperite de protocolul STS – INS și rapoartele de securitate existente.

5. Anexe

Anexa 1 – Planificarea, executarea și rezultatele de testare ale RPL



Competența face diferența!

Proiect selectat în cadrul Programului Operațional Capacitate Administrativă, cofinanțat de Uniunea Europeană
din Fondul Social European

Anexa 1 – Planificarea, executarea și rezultatele de testare ale RPL

Ipoteze

Pentru a estima numărul maxim de utilizatori concurenți pe care sistemul îi poate gestiona, am vizat 16000000 de utilizatori (mai mult de 70% din populația estimată) în 60 de zile cu o fereastră de 20 de ore pentru înregistrare sau finalizarea interviului. Pe baza constatărilor pilot, vom dezvolta un model de estimare pentru a afla numărul maxim de utilizatori concurenți pe care sistemul ar trebui să-i poată servi.

Constatări pilotării

Pentru a calcula numărul maxim de utilizatori concurenți, folosim informațiile colectate din pilot. Iată cele mai importante constatări ale pilotului:

Tabel 1 – Durata medie a interviului

MEMBRU	3/10/2021	3/11/2021	3/12/2021	3/13/2021	3/14/2021	3/15/2021	3/16/2021	AVERAGE
ECHIPA								
Toate	1 oră, 51	2 ore, 30	3 ore, 46	1 oră, 19	1 oră, 35			
echipele	minute	minute	minute	minute	minute	51 minute	21 minute	
supervisor	1 oră, 51	2 ore, 30	3 ore, 46	1 oră, 19	1 oră, 35	51 minute	21 minute	1 oră, 35
	minute	minute	minute	minute	minute			minute

Tabel 2 – Numărul interviurilor completate

MEMBRU	3/10/2021	3/11/2021	3/12/2021	3/13/2021	3/14/2021	3/15/2021	3/16/2021	AVERAGE	TOTAL
ECHIPA									
Toate									
echipele	710	562	479	485	676	931	822	666.43	4,665
supervisor	710	562	479	485	676	931	822	666.43	4,665

S-ar putea să observați că timpul mediu de finalizare a interviurilor este de 95 de minute și numărul total de interviuri finalizate pentru cele 7 zile ale pilotului este de 4665. Putem presupune că durata medie a înregistrării va fi de aproximativ 5 minute, așa că putem folosi 100 de minute ca timpul total pentru a finaliza atât auto-înregistrarea, cât și interviul.

Pentru a găsi numărul maxim de utilizatori într-o zi, am normalizat distribuția utilizatorilor de-a lungul celor 7 zile, apoi de-a lungul a 20 de minute. În modelul real avem mai multe vârfuri de obicei la mijlocul săptămânii și de mai multe ori pe parcursul unei zile. Asumând un singur vârf, ne asigurăm că niciun alt vârf nu va fi mai mare decât acesta.

Tabel 3 – Model pentru estimarea cel mai înalt vârf de distribuție pe zile - pilot

Zi	Visitator
1	383
2	736
3	1020
4	1020
5	736
6	383
7	143

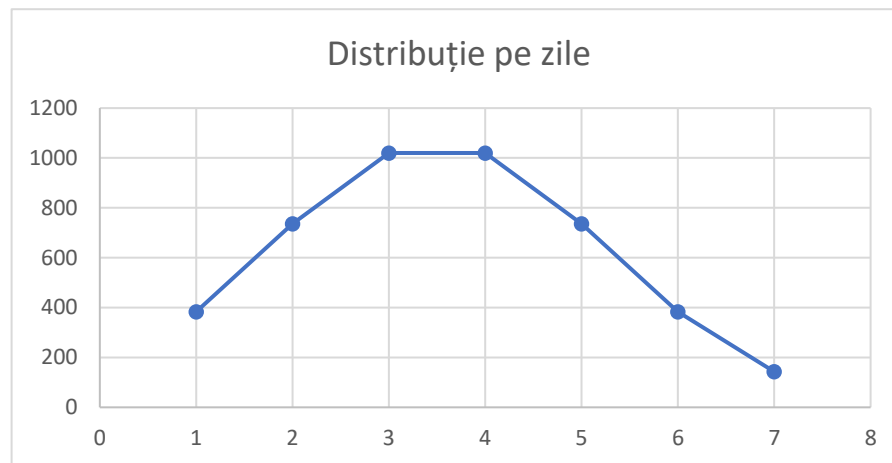
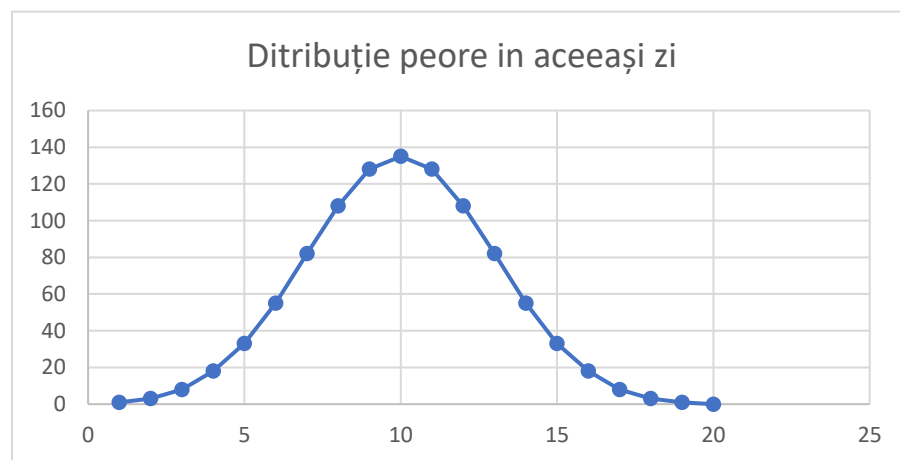


Table 4 – Model to estimate highest peak of distribution by ore - pilot

Oră	Vizitatori
1	1
2	3
3	8
4	18
5	33
6	55
7	82
8	108
9	128
10	135
11	128
12	108
13	82
14	55
15	33
16	18
17	8
18	3
19	1
20	0

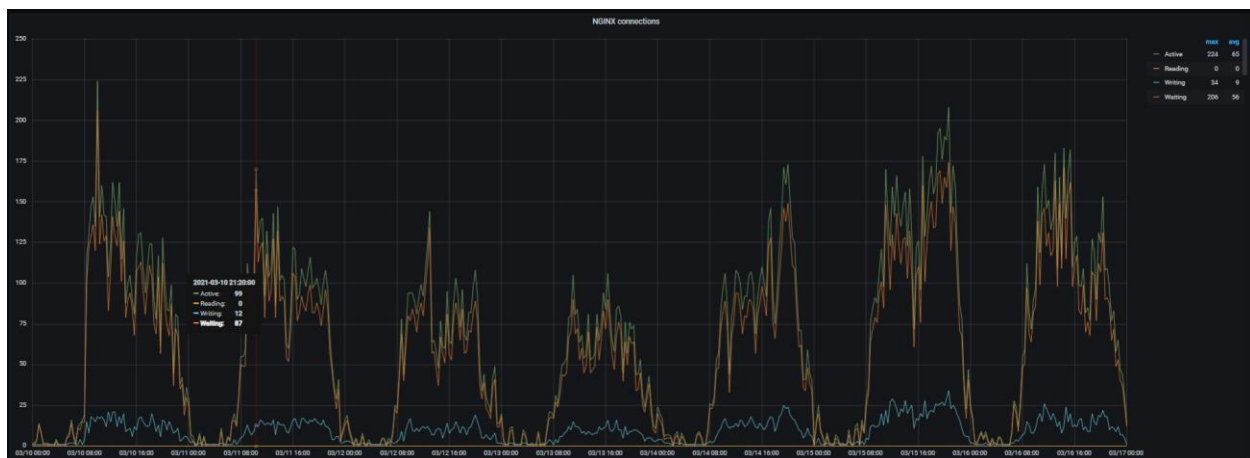


Împărțind numărul maxim de utilizatori pe zi la numărul de interviuri care pot fi realizate într-o oră înmulțit cu intervalul de timp activ în minere, obținem numărul de utilizatori concurenți de vârf:

Table 5 – Număr maxim (vârf) de utilizatori concurenți

Număr maxim (vârf) de utilizatori concurenți / înregistrare	225
Număr Total *	4665
Număr de zile	7
Max număr/zi	1020
Timp mediu de compeltare în minute*	100
Intervalul de activitate în ore*	20
Număr maxim pe oră	135

Acest calcul a furnizat un număr maxim de 225 de utilizatori concurenți în timpul pilotului. Comparând acest număr estimat cu datele reale, validăm modelul nostru de estimare pentru a fi utilizat pentru producție. Numărul real de utilizatori concurenți maximi ar trebui determinat de numărul maxim de conexiuni înregistrat de proxy-urile NGINX, care în cazul nostru este 224:



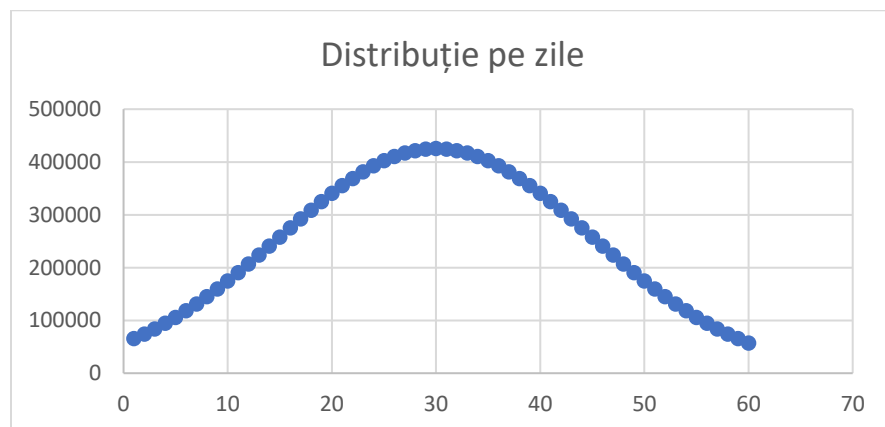
Rețineți că conexiunile concurente ale echilibratoarelor de sarcină (load balancers) includ și utilizatori de la sediul central, observatori și supraveghetori, astfel încât, în realitate, numărul real de utilizatori concurenți care sunt implicați în auto-înregistrare și finalizarea interviului este de fapt mai mic.

Calculule pentru mediul de producție

Folosind modelele de estimare din pilot putem calcula numărul maxim de utilizatori concurenți pentru etapa de producție:

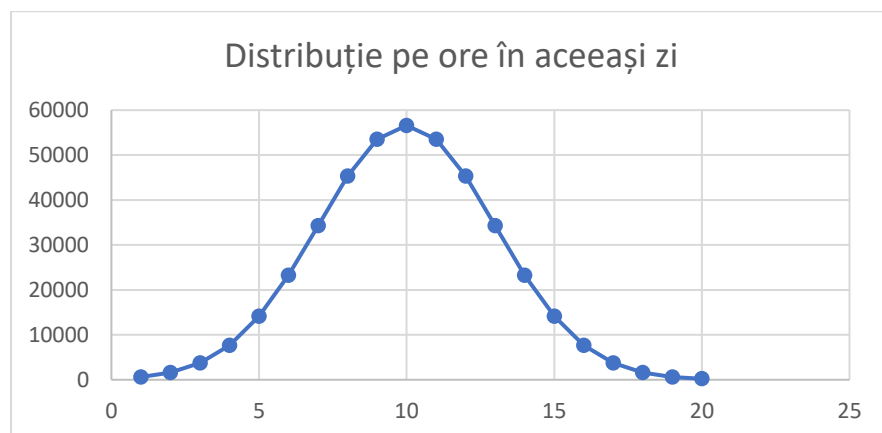
Tabel 6 –Model pentru a estima cel mai înalt vârf de distribuție pe zile - producție

Zi	Visitatori	21	355439	42	309004
1	65658	22	369124	43	292305
2	74524	23	381635	44	275283
3	84213	24	392821	45	258102
4	94739	25	402542	46	240920
5	106108	26	410673	47	223885
6	118315	27	417112	48	207131
7	131341	28	421772	49	190782
8	145154	29	424593	50	174944
9	159709	30	425538	51	159709
10	174944	31	424593	52	145154
11	190782	32	421772	53	131341
12	207131	33	417112	54	118315
13	223885	34	410673	55	106108
14	240920	35	402542	56	94739
15	258102	36	392821	57	84213
16	275283	37	381635	58	74524
17	292305	38	369124	59	65658
18	309004	39	355439	60	57590
19	325208	40	340744		
20	340744	41	325208		



Tabel 7 - Model pentru a estima cel mai înalt vârf de distribuție pe ore - producție

Oră	Vizitatori	8	45312	16	7658
1	628	9	53530	17	3719
2	1616	10	56588	18	1616
3	3719	11	53530	19	628
4	7658	12	45312	20	218
5	14110	13	34322		
6	23264	14	23264		
7	34322	15	14110		



Prin urmare, putem determina:

Tabel 8 – Număr maxim de utilizatori concurenți pentru mediul de producție

Număr maxim (vârf) de utilizatori concurenți / înregistrare	94313
Număr Total	16000000
Număr de zile	60
Număr maxim/zi	425538
Timp mediu de completare în minute	100
Intervalul de activitate în ore*	20
Număr maxim pe oră	56588

Test de stres – scenariu testare

Scenariul de testare va presupune că o persoană se va înregistra pe <https://autorecenzare.insse.ro>, ca o singură persoană cu adresă aleatorie (preluată din lista de disponibile județ/localitate/satul/stradă/stradăNr baza de date) cu 1 gospodărie formată din 1 persoană. Apoi persoana va finaliza interviul răspunzând la 60 de întrebări cu răspunsuri fixe.

Pentru testul de stres, vom lua în considerare o perioadă de 5 minute $\pm$ 30 de secunde pentru autoînregistrare și 60 de minute $\pm$ 6 minute pentru finalizarea interviului. Între aceste perioade va fi o pauză de 5 minute $\pm$ 30 de secunde (pentru a simula așteptarea e-mailului link-ului interviului).

Patru mașini vor fi folosite pentru a genera un total de 100000 de utilizatori care se vor înregistra singuri, apoi vor completa interviurile. Două faze vor fi utilizate pentru comportamentul de generare a utilizatorilor pentru fiecare dintre cele 4 mașini generatoare de sarcină:

- Ramp-Up de la 0 la 25000 pe parcursul unei ore
- Stai la 25000 de utilizatori pentru încă 1 oră

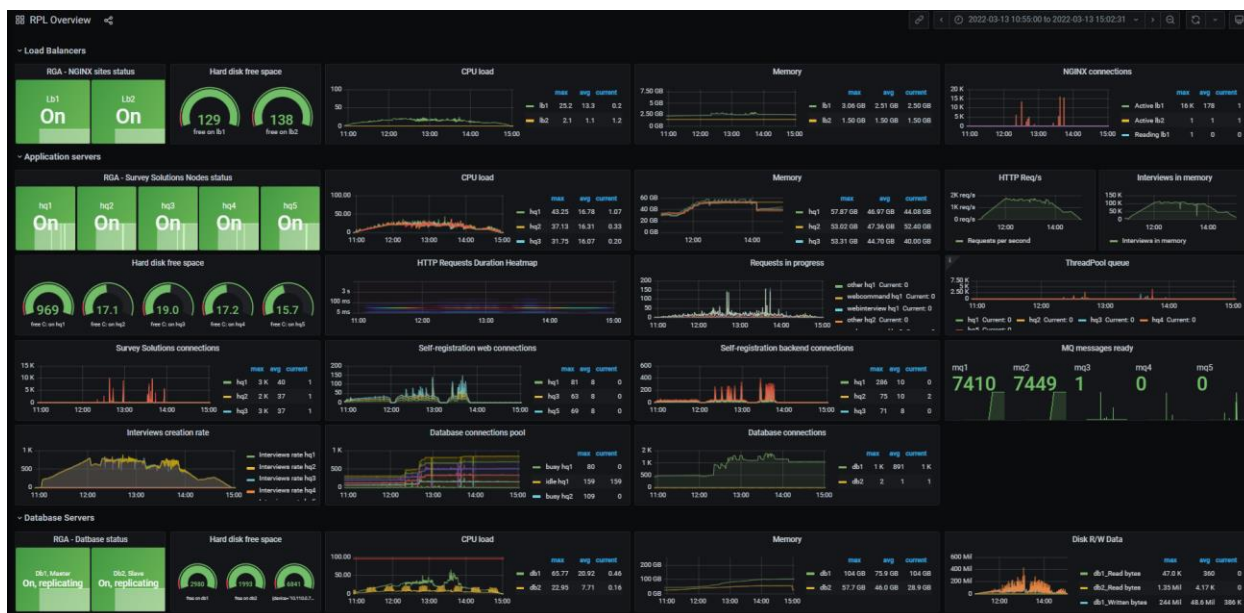
Apoi, pentru a închide corect interviurile, va exista o perioadă de ramp-down cu durată variabilă, calculată luând în considerare timpul maxim necesar pentru finalizarea restului interviurilor. Acest lucru nu este relevant pentru testarea performanței, dar este necesar pentru executarea fără probleme, ceea ce înseamnă că toate interviurile deschise ar trebui finalizate.

Scripturile de testare vor raporta și valori de performanță legate de comportamentul aplicației:

- Solicitare care durează mai mult de 5 secunde - pragul de 99% din răspunsuri să fie sub 5 secunde
- Interviuri cu întrebări fără răspuns – pragul 99% din răspunsuri să fie procesate fără erori

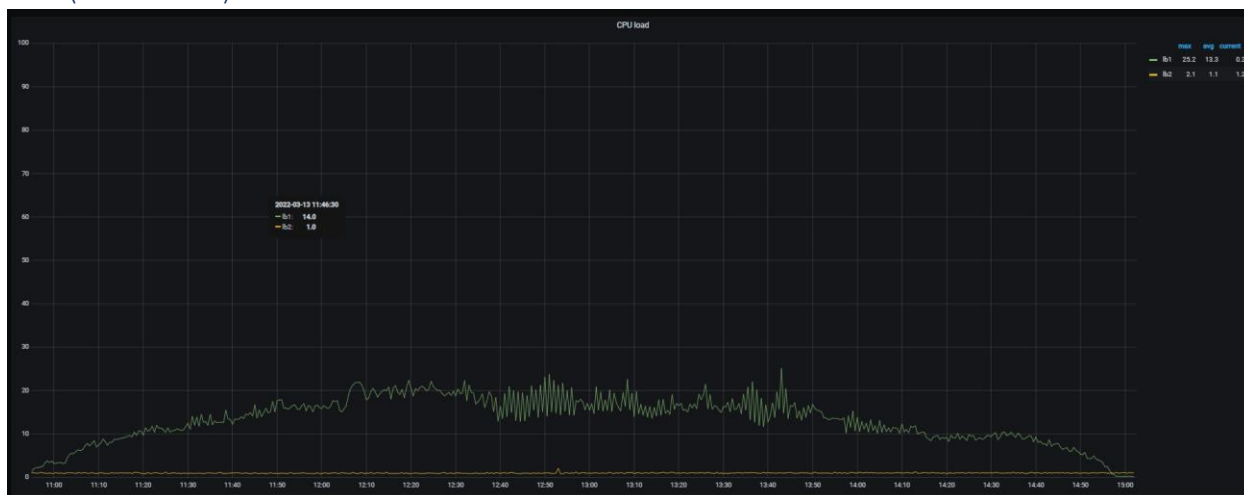
Resultate testare

Mai jos puteți găsi rezultatele testului de stres de performanță și cele mai importante - metrice legate de componentele sistemului. Acesta este panoul general de monitorizare:



Load balancers (echilibratoare de sarcină)

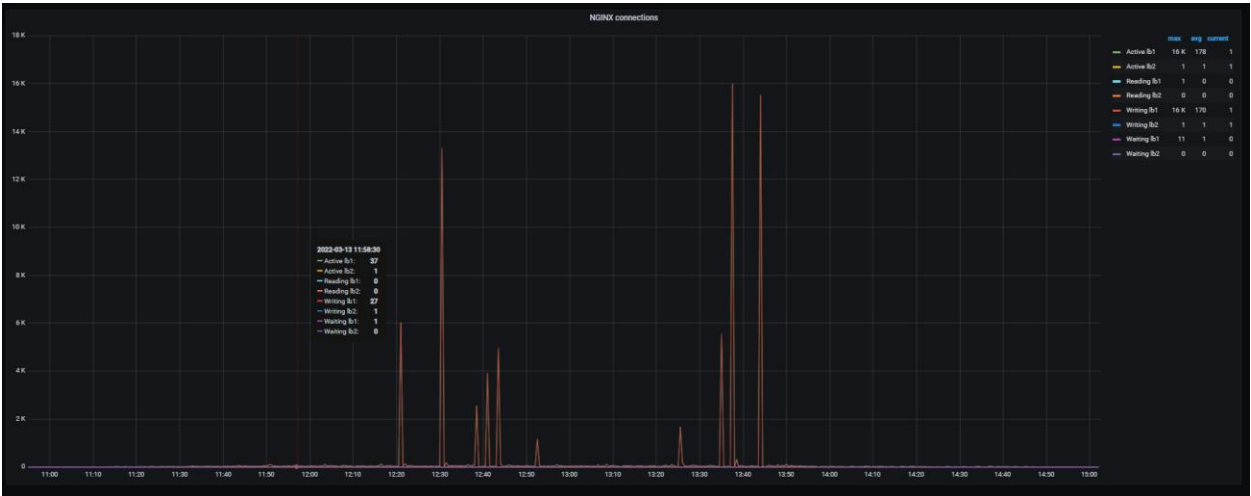
CPU (Procesoare)



Memorie

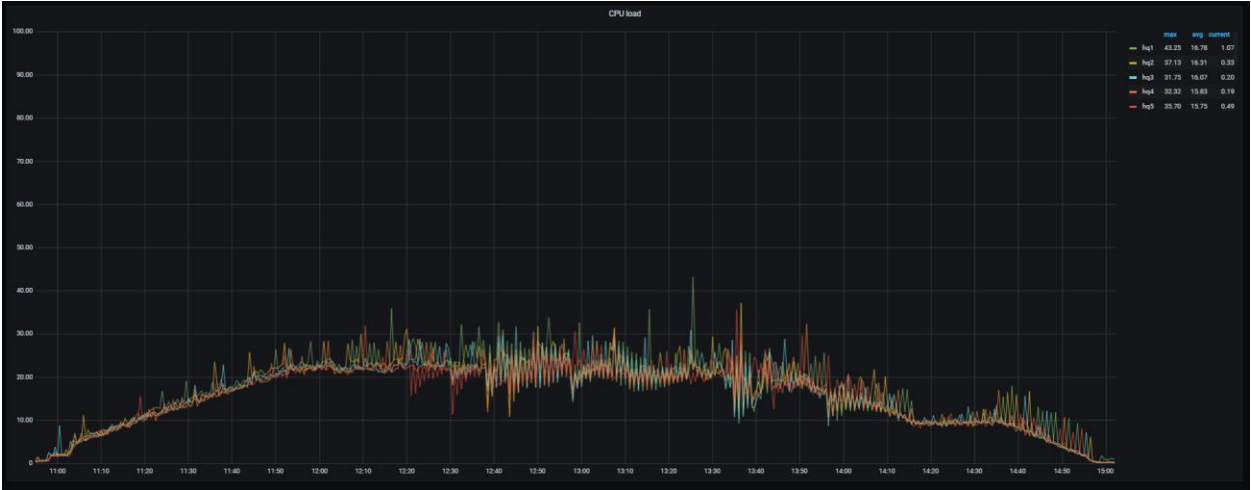


Conexiuni concurente



Servere de aplicație

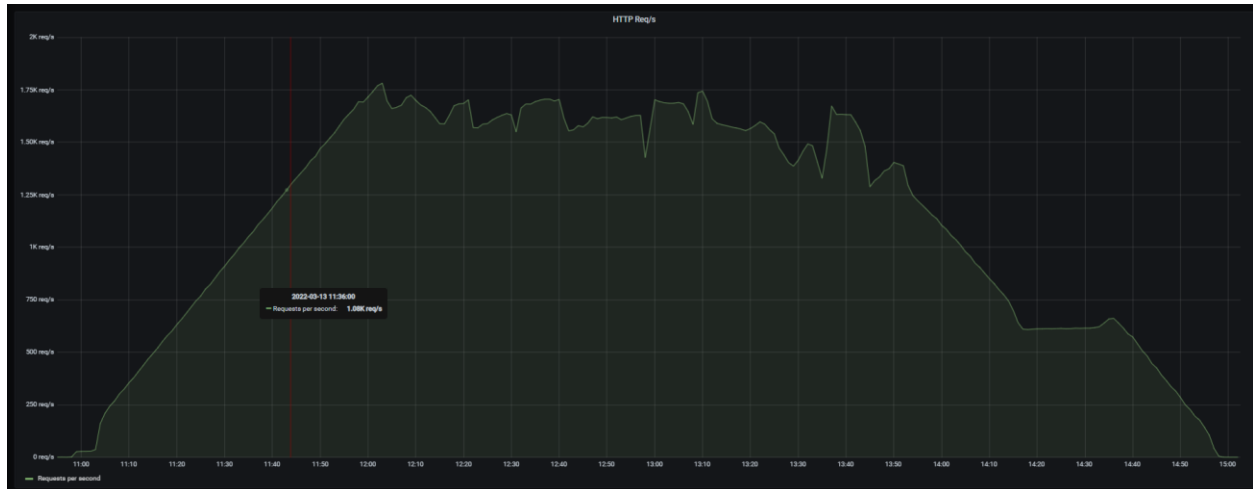
CPU



Memorie

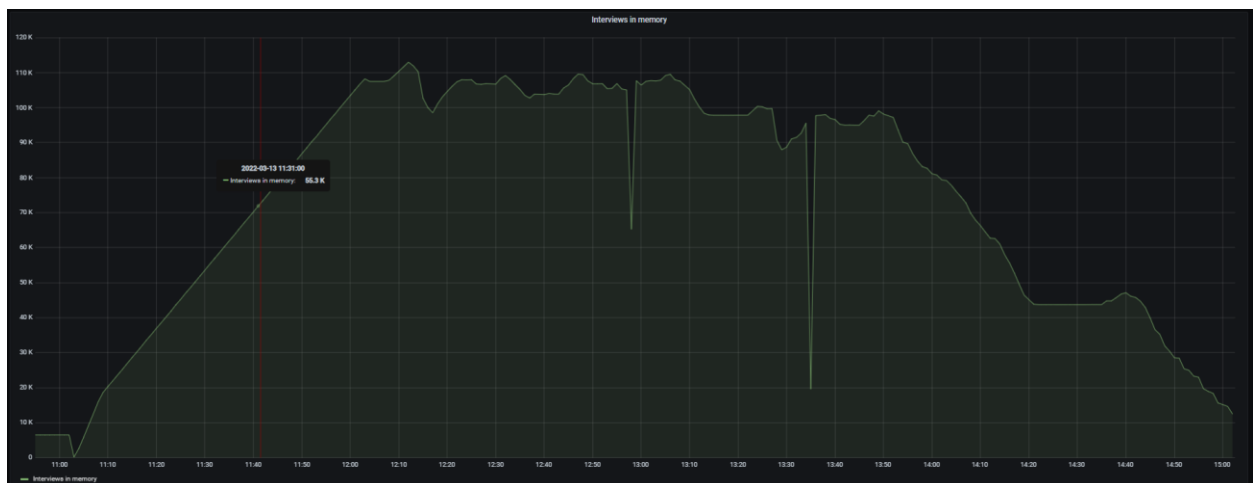


HTTP req/s

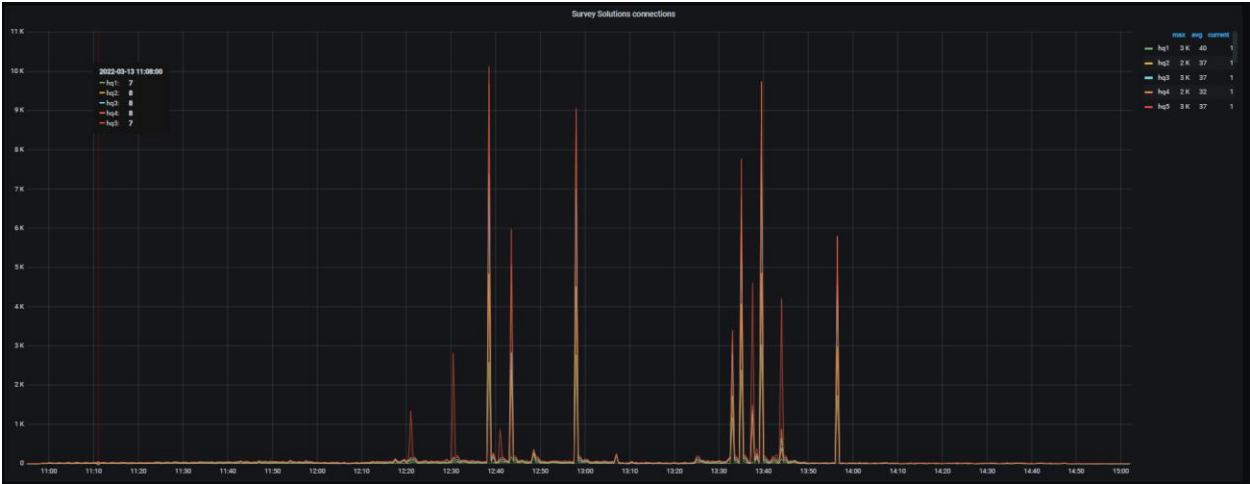


Survey Solutions

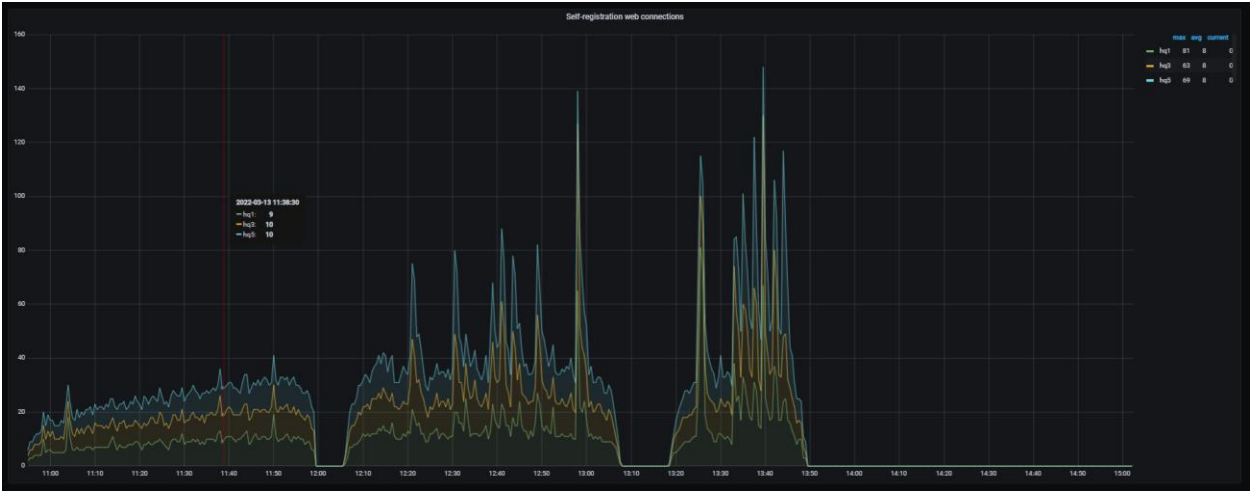
Interviuri în memorie



Conexiuni concurente

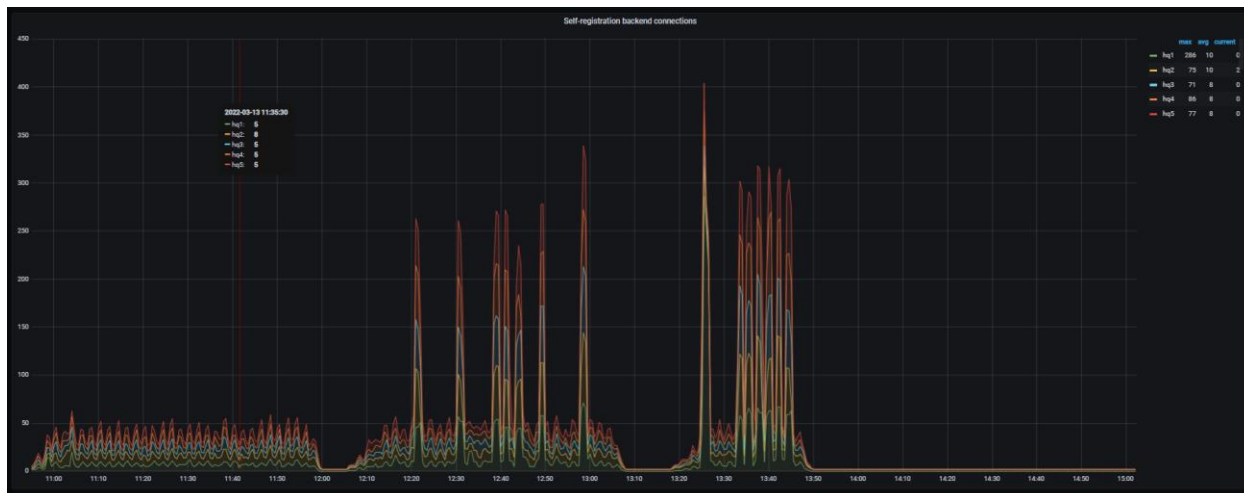


Autoinregistrare web
Conexiuni concurente



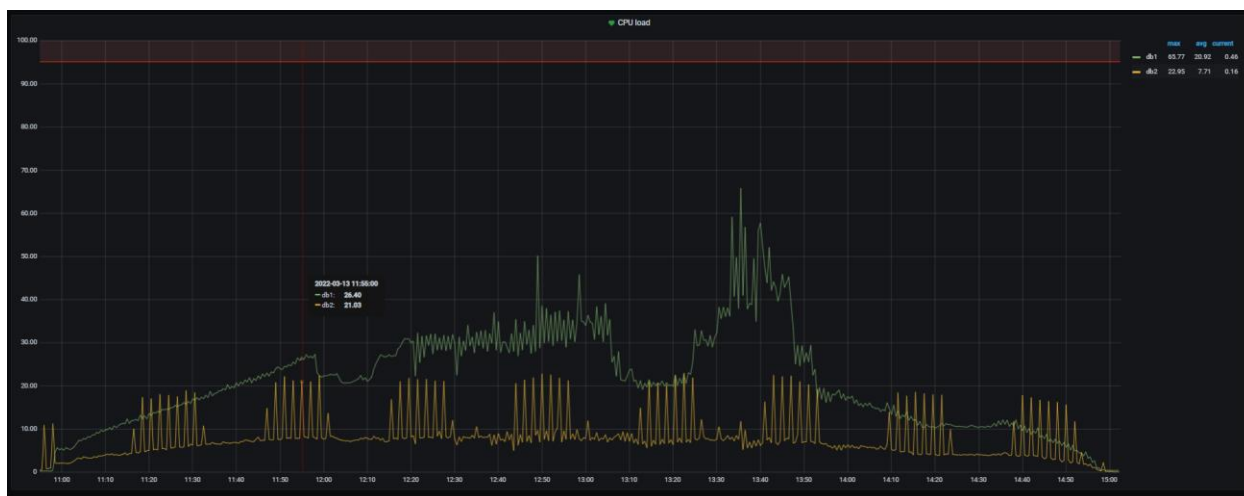
Autoinregistrare backend

Conexiuni concurente

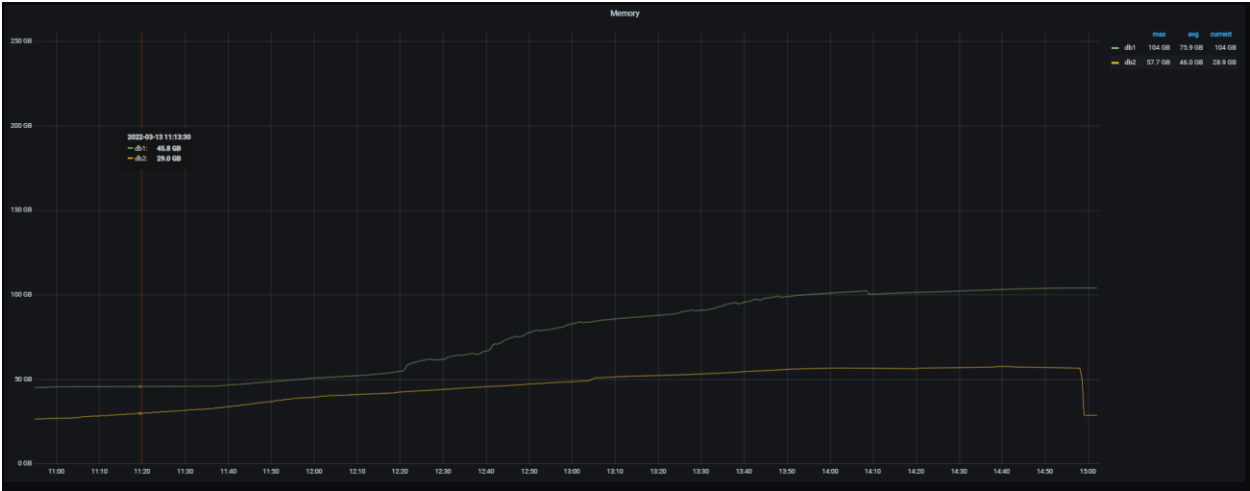


Servere baze de date

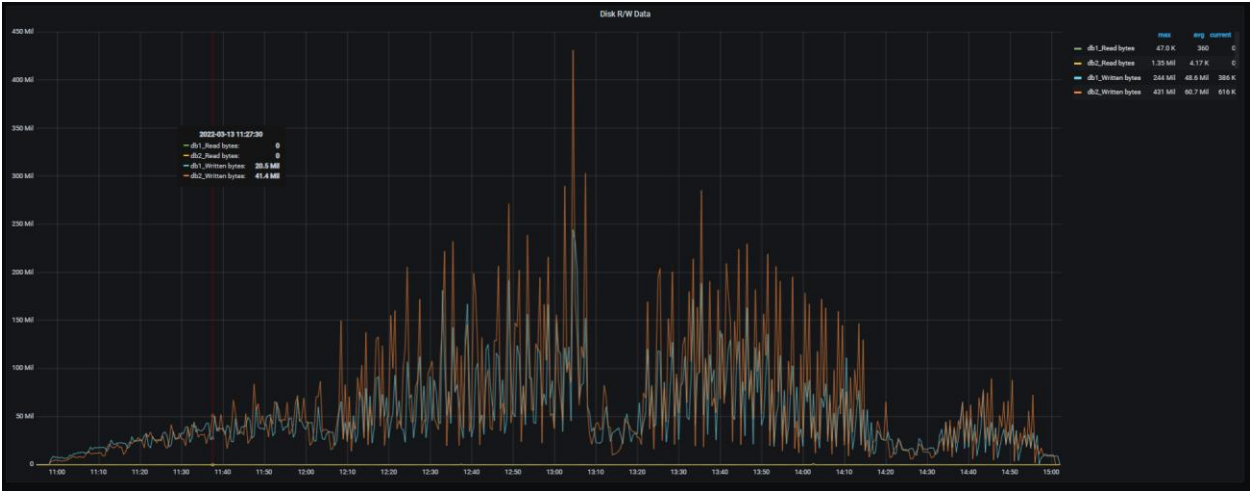
CPU



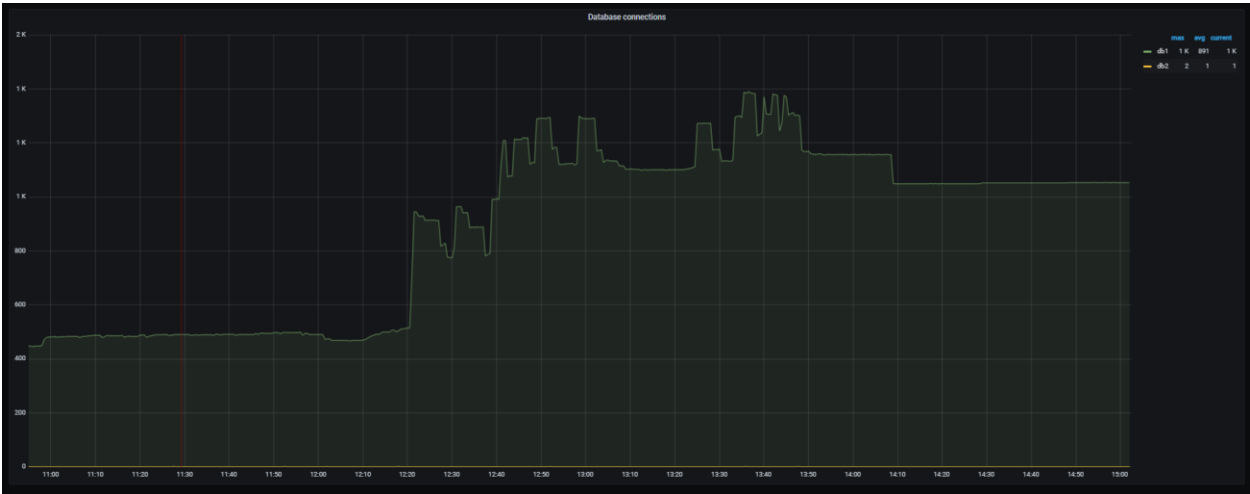
Memorie



Disk R/W



Conexiuni Baza de date



Rezultatele încărcării - scripturi

Scriptul de încărcare pentru auto-înregistrare și finalizarea interviului

Primul load generator:

```
census_simulation ✓ [=====] 00000/25000 VUs 3h17m30s

✓ Valid registration(s)
X Interview(s) started
  □ 99% - ✓ 56802 / X 1
X Question(s) answered
  □ 99% - ✓ 3635219 / X 109
X Interview(s) completed
  □ 99% - ✓ 56799 / X 3

checks.....: 99.99% ✓ 3805623 X 113
data_received.....: 24 GB 1.7 MB/s
data_sent.....: 5.6 GB 380 KB/s
http_req_blocked.....: avg=3.3ms min=1.84ms med=3.01ms max=70.82ms p(95)=5.28ms p(99)=7.24ms
http_req_connecting.....: avg=434.25µs min=148.25µs med=344.26µs max=62.02ms p(95)=1.02ms p(99)=1.49ms
http_req_duration.....: avg=147.01ms min=716.58µs med=19.89ms max=19.14s p(95)=67.51ms p(99)=6.05s
  { expected_response:true }.....: avg=147ms min=1.34ms med=19.89ms max=19.14s p(95)=67.5ms p(99)=6.05s
http_req_failed.....: 0.00% ✓ 119 X 4430443
http_req_receiving.....: avg=108.49µs min=18.21µs med=100.23µs max=42.93ms p(95)=185.04µs p(99)=251.42µs
http_req_sending.....: avg=66.32µs min=21.62µs med=59.53µs max=55.45ms p(95)=107.7µs p(99)=155.31µs
http_req_tls_handshaking.....: avg=2.77ms min=1.59ms med=2.47ms max=70.2ms p(95)=4.65ms p(99)=6.54ms
http_req_waiting.....: avg=146.83ms min=605.8µs med=19.72ms max=19.14s p(95)=67.31ms p(99)=6.05s
http_reqs.....: 4430562 302.15529/s
interviews_completed.....: 56802 3.873781/s
interviews_created.....: 56802 3.873781/s
iteration_duration.....: avg=1h13m21s min=8m58s med=1h13m21s max=1h15m58s p(95)=1h14m23s p(99)=1h14m49s
iterations.....: 56803 3.873849/s
registrations_completed.....: 56803 3.873849/s
registrations_started.....: 56803 3.873849/s
✓ SelfReg_answers_below_response_limit...: 99.82% ✓ 283524 X 491
✓ SelfReg_http_valid_responses.....: 100.00% ✓ 284015 X 0
X SuSo_answers_below_response_limit.....: 98.61% ✓ 3584637 X 50491
✓ SuSo_http_valid_responses.....: 99.99% ✓ 3635219 X 109
vus.....: 1 min=1 max=25000
vus_max.....: 25000 min=25000 max=25000
```

Al doilea load generator:

```
census_simulation ✓ [=====] 00000/25000 VUs 3h17m30s

✓ Valid registration(s)
X Interview(s) started
  □ 99% - ✓ 56783 / X 4
X Question(s) answered
  □ 99% - ✓ 3633995 / X 117
X Interview(s) completed
  □ 99% - ✓ 56782 / X 1

checks.....: 99.99% ✓ 3804347 X 122
data_received.....: 24 GB 1.7 MB/s
data_sent.....: 5.6 GB 380 KB/s
http_req_blocked.....: avg=10.31ms min=7.39ms med=9.94ms max=1.14s p(95)=13.1ms p(99)=15.3ms
http_req_connecting.....: avg=3.47ms min=2.91ms med=3.19ms max=1.12s p(95)=4.67ms p(99)=5.63ms
http_req_duration.....: avg=150.75ms min=3.55ms med=22.72ms max=19.6s p(95)=73.2ms p(99)=6.1s
  { expected_response:true }.....: avg=150.74ms min=3.98ms med=22.72ms max=19.6s p(95)=73.18ms p(99)=6.1s
http_req_failed.....: 0.00% ✓ 124 X 4428974
http_req_receiving.....: avg=98.41µs min=14.36µs med=86.6µs max=23.98ms p(95)=159.65µs p(99)=235.49µs
http_req_sending.....: avg=71.13µs min=17.67µs med=67.02µs max=20.96ms p(95)=106.48µs p(99)=160.13µs
http_req_tls_handshaking.....: avg=6.75ms min=4.3ms med=6.45ms max=263.5ms p(95)=9.25ms p(99)=11.37ms
http_req_waiting.....: avg=150.58ms min=3.44ms med=22.55ms max=19.6s p(95)=73ms p(99)=6.1s
http_reqs.....: 4429098 302.071475/s
interviews_completed.....: 56783 3.87269/s
interviews_created.....: 56783 3.87269/s
iteration_duration.....: avg=1h13m22s min=9m20s med=1h13m22s max=1h16m6s p(95)=1h14m24s p(99)=1h14m51s
iterations.....: 56787 3.872963/s
registrations_completed.....: 56787 3.872963/s
registrations_started.....: 56787 3.872963/s
✓ SelfReg_answers_below_response_limit...: 99.83% ✓ 283471 X 464
✓ SelfReg_http_valid_responses.....: 100.00% ✓ 283935 X 0
X SuSo_answers_below_response_limit.....: 98.60% ✓ 3583506 X 50606
✓ SuSo_http_valid_responses.....: 99.99% ✓ 3633995 X 117
vus.....: 1 min=1 max=25000
vus_max.....: 25000 min=25000 max=25000
```

Al treilea load generator:

```
census_simulation ✓ [=====] 00000/25000 VUS 3h17m30s

✓ Valid registration(s)
X Interview(s) started
  □ 99% - ✓ 56775 / X 2
X Question(s) answered
  □ 99% - ✓ 3633457 / X 143
X Interview(s) completed
  □ 99% - ✓ 56773 / X 2

checks.....: 99.99% ✓ 3803782 X 147
data_received.....: 24 GB 1.7 MB/s
data_sent.....: 5.6 GB 380 kB/s
http_req_blocked.....: avg=10.45ms min=7.36ms med=10.07ms max=1.14s p(95)=13.16ms p(99)=15.34ms
http_req_connecting.....: avg=3.48ms min=2.9ms med=3.2ms max=1.12s p(95)=4.68ms p(99)=5.54ms
http_req_duration.....: avg=151.27ms min=3.63ms med=22.72ms max=19.61s p(95)=73.53ms p(99)=6.11s
  { expected_response:true }.....: avg=151.26ms min=3.98ms med=22.72ms max=19.61s p(95)=73.52ms p(99)=6.11s
http_req_failed.....: 0.00% ✓ 151 X 4428311
http_req_receiving.....: avg=108.51µs min=14.68µs med=96.92µs max=314.82ms p(95)=170.7µs p(99)=254.65µs
http_req_sending.....: avg=77.4µs min=18.37µs med=74.01µs max=5.57ms p(95)=110.89µs p(99)=167µs
http_req_tls_handshaking.....: avg=6.87ms min=4.34ms med=6.54ms max=282.68ms p(95)=9.29ms p(99)=11.43ms
http_req_waiting.....: avg=151.08ms min=3.51ms med=22.54ms max=19.61s p(95)=73.31ms p(99)=6.11s
http_reqs.....: 4428462 302.323947/s
interviews_completed.....: 56775 3.875938/s
interviews_created.....: 56775 3.875938/s
iteration_duration.....: avg=1h13m22s min=8m46s med=1h13m22s max=1h16m12s p(95)=1h14m25s p(99)=1h14m51s
iterations.....: 56777 3.876074/s
registrations_completed.....: 56777 3.876074/s
registrations_started.....: 56777 3.876074/s
✓ SelfReg answers below response limit....: 99.84% ✓ 283447 X 438
✓ SelfReg http valid responses.....: 100.00% ✓ 283885 X 0
X SuSo answers below response limit.....: 98.59% ✓ 3582578 X 51022
✓ SuSo http valid responses.....: 99.99% ✓ 3633457 X 143
vus.....: 0 min=0 max=25000
vus_max.....: 25000 min=25000 max=25000
```

Al patrulea load generator:

```
census_simulation ✓ [=====] 00000/25000 VUS 3h17m30s

✓ Valid registration(s)
✓ Interview(s) started
X Question(s) answered
  □ 99% - ✓ 3634038 / X 138
X Interview(s) completed
  □ 99% - ✓ 56783 / X 1

checks.....: 99.99% ✓ 3804389 X 139
data_received.....: 24 GB 1.7 MB/s
data_sent.....: 5.6 GB 380 kB/s
http_req_blocked.....: avg=10.31ms min=7.37ms med=9.94ms max=1.21s p(95)=13.09ms p(99)=15.28ms
http_req_connecting.....: avg=3.47ms min=2.91ms med=3.18ms max=1.03s p(95)=4.66ms p(99)=5.62ms
http_req_duration.....: avg=150.99ms min=3.66ms med=22.72ms max=19.49s p(95)=73.33ms p(99)=6.08s
  { expected_response:true }.....: avg=150.98ms min=4.05ms med=22.72ms max=19.49s p(95)=73.31ms p(99)=6.08s
http_req_failed.....: 0.00% ✓ 141 X 4429011
http_req_receiving.....: avg=98.53µs min=14.36µs med=86.71µs max=21.78ms p(95)=160.1µs p(99)=238.12µs
http_req_sending.....: avg=71.19µs min=16.87µs med=67.05µs max=10.27ms p(95)=106.57µs p(99)=161.26µs
http_req_tls_handshaking.....: avg=6.75ms min=4.33ms med=6.45ms max=248.1ms p(95)=9.24ms p(99)=11.36ms
http_req_waiting.....: avg=150.82ms min=3.52ms med=22.55ms max=19.49s p(95)=73.14ms p(99)=6.08s
http_reqs.....: 4429152 302.295271/s
interviews_completed.....: 56784 3.87558/s
interviews_created.....: 56784 3.87558/s
iteration_duration.....: avg=1h13m22s min=1h10m58s med=1h13m22s max=1h16m1s p(95)=1h14m24s p(99)=1h14m50s
iterations.....: 56784 3.87558/s
registrations_completed.....: 56784 3.87558/s
registrations_started.....: 56784 3.87558/s
✓ SelfReg answers below response limit....: 99.84% ✓ 283489 X 431
✓ SelfReg http valid responses.....: 100.00% ✓ 283920 X 0
X SuSo answers below response limit.....: 98.60% ✓ 3583307 X 50869
✓ SuSo http valid responses.....: 99.99% ✓ 3634038 X 138
vus.....: 1 min=1 max=25000
vus_max.....: 25000 min=25000 max=25000
```

Concluziile testării

Chiar și loadscript-urile raportează un anumit prag atins, rezultatele testelor reflectă faptul că sistemul este pregătit să facă față sarcinii necesare.

Deoarece sistemul poate fi actualizat, sugestia este de a crește resursele din memoria serverului de aplicații (de la 64G la 96G și cantitatea de nuclee de la 16 la 32. Acest lucru va îmbunătăți timpul de răspuns și fezabilitatea generală a sistemului).