**ROMÂNIA**

**MINISTERUL APĂRĂRII NAŢIONALE**

**ACADEMIA TEHNICA MILITARĂ „FERDINAND I”**

**Specializarea: Calculatoare și sisteme informatice pentru apărare și securitate națională**



**Student: Sd. Sg. Maj. Pesu Ciprian-George**

**Cuprins**

[**1. Introducere** 3](#_Toc92124777)

[**1.1 Descrierea proiectului** 3](#_Toc92124778)

[**2. Cerințe software** 4](#_Toc92124779)

[**2.1 Actori** 4](#_Toc92124780)

[**2.2 Cerințe funcționale** 4](#_Toc92124781)

[**2.2 Cerințe nefuncționale** 4](#_Toc92124782)

[**3. Arhitectura sistemului** 5](#_Toc92124783)

[**3.1 Diagrama sistem crawling si clasificare știri** 5](#_Toc92124784)

[**3.2 Arhitectura platforma web** 6](#_Toc92124785)

[**3.2.1 Client** 6](#_Toc92124786)

[**3.2.2 Server** 6](#_Toc92124787)

[**4. Diagrama cazurilor de utilizare** 7](#_Toc92124788)

[**5. Diagrama de secvență** 8](#_Toc92124789)

[**6. Diagrama de activități** 9](#_Toc92124790)

[**7. Testare platforma web** 10](#_Toc92124791)

[**8. Bibliografie** 11](#_Toc92124792)

# **1. Introducere**

## **1.1 Descrierea proiectului**

Scopul proiectului este realizarea unei aplicații WEB și a unor servicii de fundal pentru procesarea de știri în limba engleză publicate în internet (fluxuri RSS preluate de pe site-uri consacrate de știri).

Principalele obiective sunt: construirea unui corpus de date uniform, extracția de entități, teme, sentiment și topice, clasificarea, sumarizarea conținutului, indexarea într-un motor de căutare și realizarea diferitelor analize statistice.

Această platformă va fi compusă din:

* Un serviciu de crawling ce va parsa fluxurile RSS ale platformelor de știri .
* O componenta de ingestie a acestor știri folosind cozi de mesaje (ex. Apache Kafka)
* Un serviciu de crawling ce va extrage textul si conținutul știrilor extrase din fluxurile RSS.
* Sistem de învățare automată care va fi antrenat folosind un set de date existent și se va folosi pentru clasificarea articolelor in funcție de sentimentele exprimate (Sentiment Analysis).
* O componentă de stocare si centralizare a datelor bazata pe Elasticsearch
* O interfață de tip web care va permite vizualizarea știrilor cat si a datelor statistice în timp real.

# **2. Cerințe software**

## **2.1 Actori**

Actorii acestui sistem sunt utilizatorii ce interacționează direct cu interfaţa grafică cu scopul de a vizualiza știri si date statistice provenite de la mai multe publicații active in domeniul online.

## **2.2 Cerințe funcționale**

Sistemul trebuie să fie capabil să satisfacă următoarele cerințe:

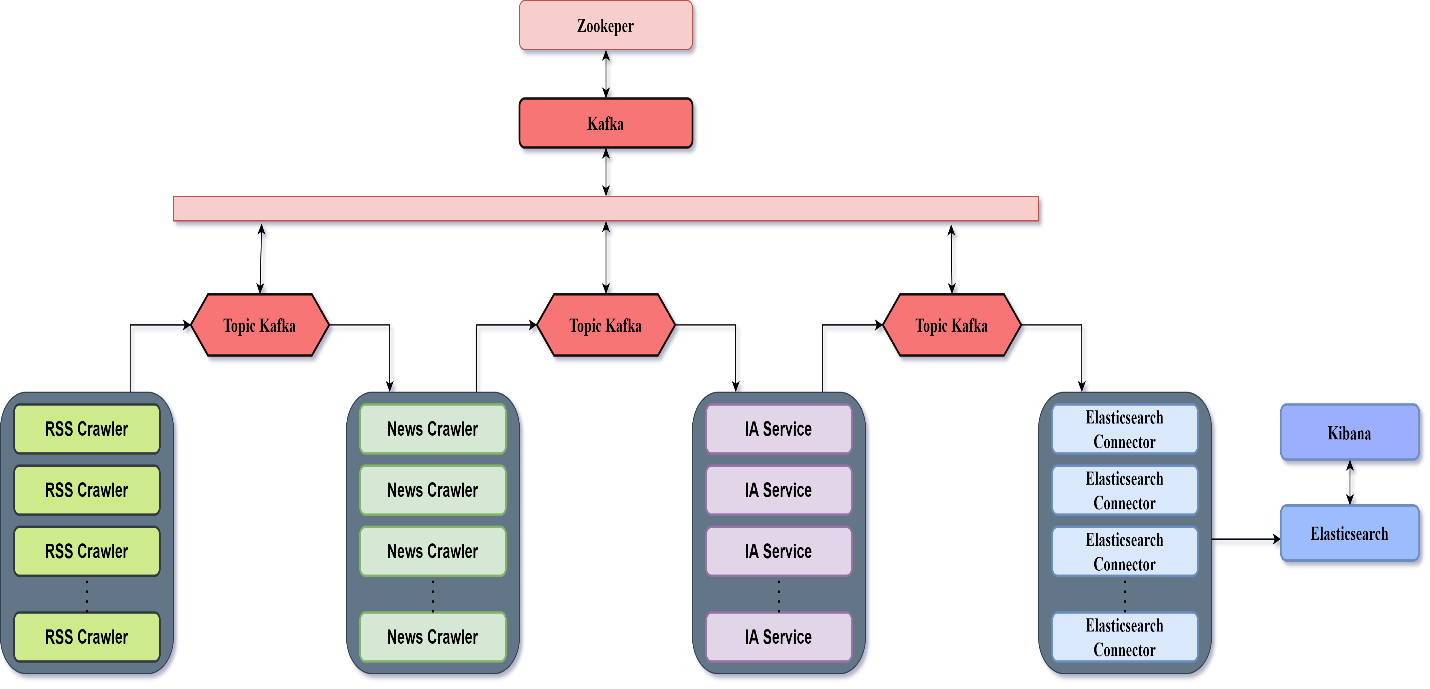
* Posibilitatea cererii unui nou cont de utilizator cat si accesarea unui cont existent din interfața grafica.
* Centralizarea știrilor in timp real si stocarea acestora intr-o baza de date.
* Efectuarea analizei de sentiment asupra știrilor prin intermediul unui model de I.A. dezvoltat.
* Posibilitatea filtrării știrilor in funcție de diferite caracteristici.
* Vizualizarea statisticilor din interfața grafica
* Posibilitatea personalizării unui flux de știri in funcție de preferințele utilizatorului
* O interfața accesibila doar administratorului sistemului pentru gestionarea serviciilor si a fluxurilor de știri urmărite.

## **2.2 Cerințe nefuncționale**

* **Scalabilitate** – structura sistemului trebuie sa fie ușor modificabila pentru a permite gestionarea unui volum mai mare de date si utilizatori.
* **Disponibilitate -**  serviciile sistemului trebuie sa aibă un procentaj al timpului de disponibilitate ridicat.
* **Performanță** – pentru a oferii o experiența plăcută sistemul trebuie să fie rapid și receptiv la comenzile date de către utilizatori.
* **Utilizare ușoară** – aplicația este ușor de folosit de către utilizatori, iar toate funcționalitățile trebuiesc prezentate si accesate într-un mod intuitiv.

# **3. Arhitectura sistemului**

## **3.1 Diagrama sistem crawling si clasificare știri**



Sistemul de crawling si clasificare a știrilor este bazat pe patru servicii containerizate conectate intre ele prin framework-ul pentru gestionara fluxurilor de date, Kafka.

* RSS Crawler – are rolul de parsa fluxurile RSS ale platformelor de știri si de a extrage link-urile cate cele mai recente știri publicate cat si alte informații relevate disponibile.
* News Crawler – reprezintă serviciul ce extrage conținutul articolelor ale căror link-uri sunt extrase de către RSS Crawler.
* IA Service – are rolul de a clasifica textul extras prin intermediul modelor de A.I dezvoltate.
* Elastcsearch Connector – introduce in componenta Elasticsearch documente in format JSON conținând datele extrase si procesate de serviciile anterioare.

Elasticsearch este un motor de căutare și analiză distribuit care permite stocarea, căutarea și analiza a volume uriașe de date rapid, oferind răspunsuri la interogări în milisecunde. Acesta a fost ales pentru a reprezenta componenta de sotare a datelor in arhitectura prezentata datorita scalabilității pe care o oferă sistemului.

Kibana permite căutarea și vizualizarea datelor indexate în cluster Elasticsearch și analizarea acestora prin crearea de grafice, diagrame circulare, tabele, histograme și hărți prin intermediul unei interfețe web. Aceasta componeta a fost introdusa in sistem pentru a permite o mai buna vizualizare a datelor stocate si a log-urilor provenite de la restul componentelor.

## **3.2 Arhitectura platforma web**

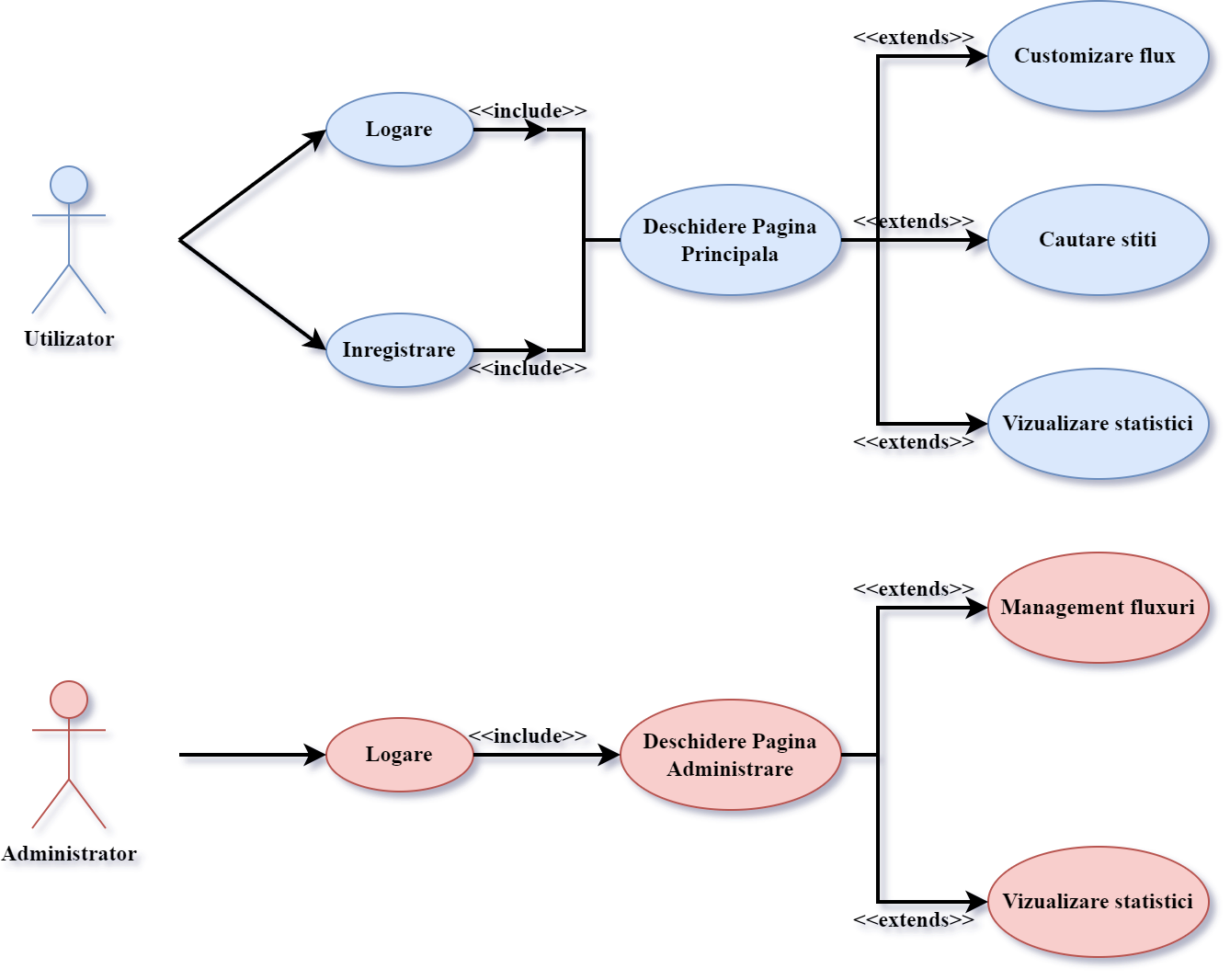
### **3.2.1 Client**

Clientul v-a fii reprezentat de o interfața grafica implementata cu ajutorul framework-ului React Js care v-a permite atât utilizatorilor cat si administratorilor sa acceseze funcționalitățile dorite într-un mod intuitiv.

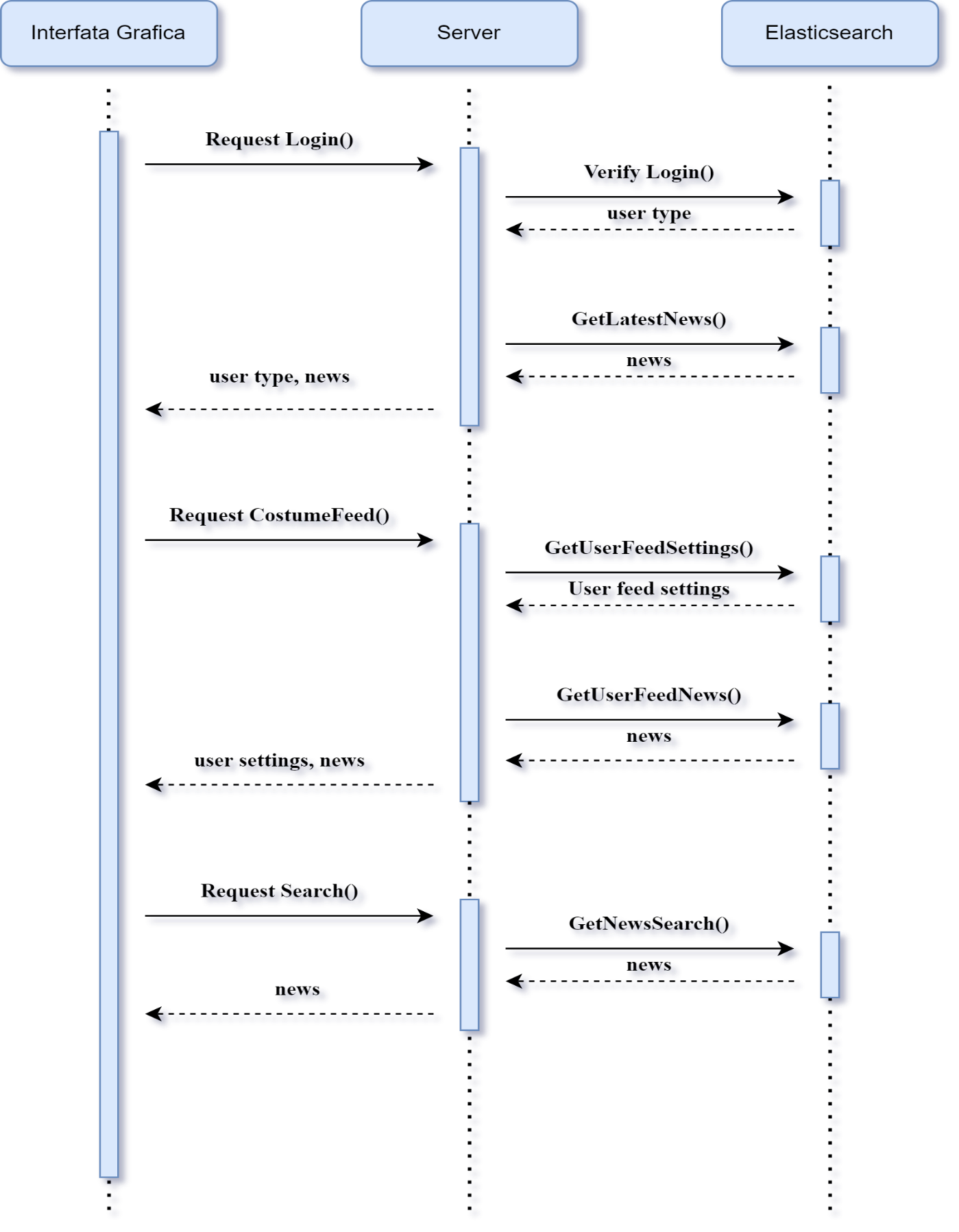
### **3.2.2 Server**

Server-ul v-a fii implementata prin intermediul tehnologiei Node Js si are ca scop principal sa gestionarea request-urile trimise de către interfața web. Acesta preia datele necesare din componenta de stocare reprezentata de Elasticsearch si le trimite către client intr-un mod securizat pentru a asigura integritatea si confidențialitatea datelor.

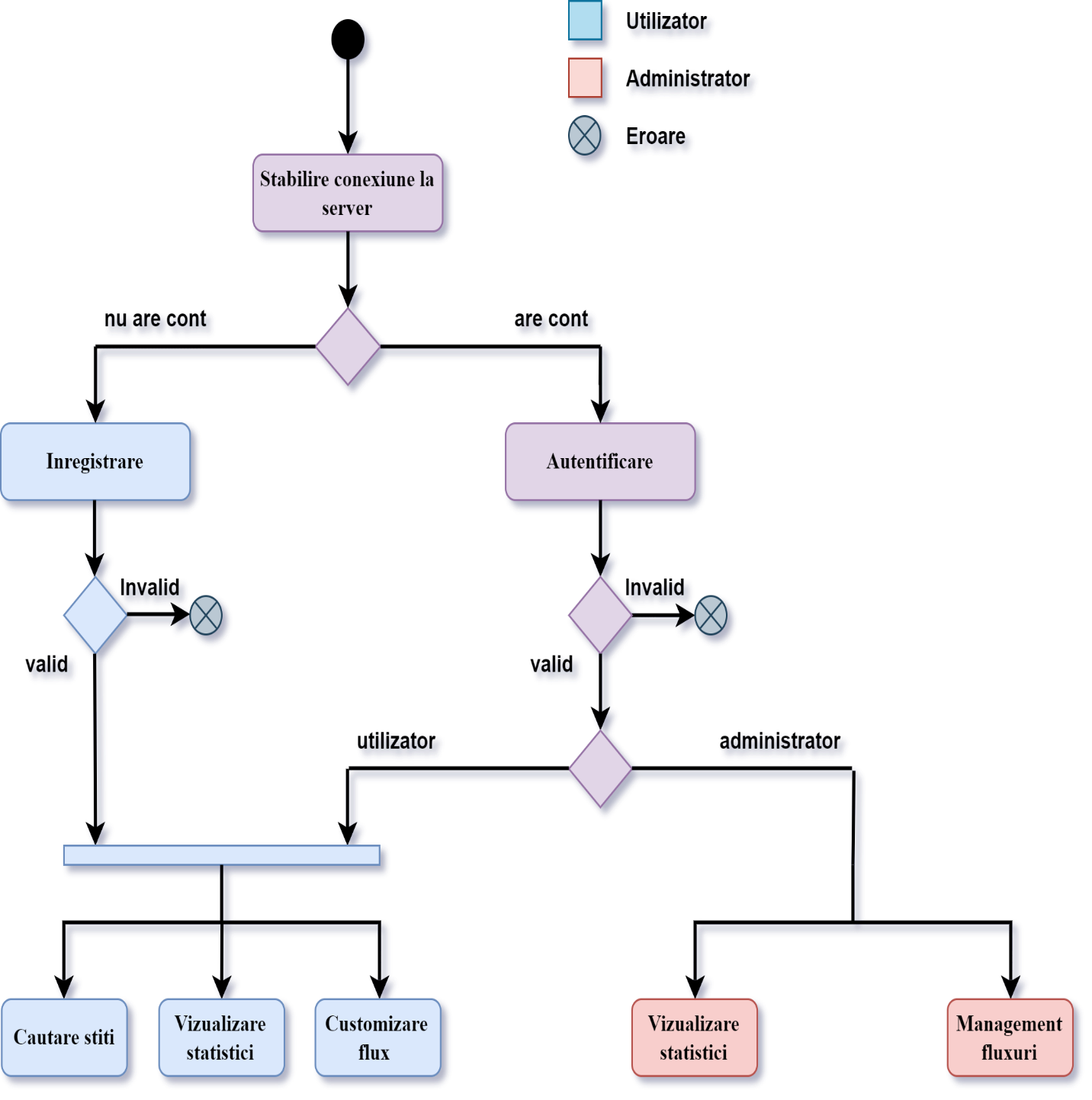
# **4. Diagrama cazurilor de utilizare**

****

# **5. Diagrama de secvență**



# **6. Diagrama de activități**



# **7. Testare platforma web**

Funcționalitățile testate pe parcursul testării, se va ține cont de un tabel cu funcționalități, prezentat mai jos și ce are următoarele coloane:

* ID
* Tipul utilizatorului
* Funcționalitatea testata
* Metodologie de testare
* Erori posibile

Testele vor fii rulate in mod repetat pentru a detecta orice comportament imprevizibil al programului. Pentru notare rezultatului testării se vor folosi trei criterii de evaluare :

* Acceptat – aplicația se comportă in modul dorit
* Respins – aplicația nu se comportă in modul dorit sau prezinta erori
* Amânat – aplicația prezinta un comportament imprevizibil pe parcursul rulării consecutive a testelor

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | Tipul utilizatorului | Funcționalitatea testata | Metodologie de testare | Erori posibile |
| T1 | Utilizator/Administrator | Logare | Se completează câmpurile necesare cu datele corecte | Credențele Incorecte |
| T2 | Utilizator | Inregistrare | Se completează câmpurile necesare | Parola prea slaba sau credențele indisponibile |
| T3 | Utilizator | Căutare știri | Se selectează filtrele dorite | Nu exista știri corespunzătoare |
| T4 | Utilizator | Vizualizare statistici | Se selectează un fluxul RSS | Nu exista date necesare pentru a genera statistici |
| T5 | Utilizator | Configurare flux | Se selecteaza filtrele personale | Nu exista știri corespunzătoare |
| T6 | Administrator | Vizualizare statistici | Se selectează filtrele dorite | Nu exista date necesare pentru a genera statistici |
| T7 | Administrator | Management fluxuri | Se adaugă sau se elimina un nou flux | Fluxul adăugat este invalid |

# **8. Bibliografie**

* [Elastic Cloud on Kubernetes | Deploy and Orchestrate Elasticsearch on Kubernetes](https://www.elastic.co/elastic-cloud-kubernetes)
* [Deploy a Kibana instance | Elastic Cloud on Kubernetes [1.9] | Elastic](https://www.elastic.co/guide/en/cloud-on-k8s/current/k8s-deploy-kibana.html)
* [MicroK8s - Zero-ops Kubernetes for developers, edge and IoT](https://microk8s.io/)
* [Apache ZooKeeper](https://zookeeper.apache.org/)
* [Apache Kafka](https://kafka.apache.org/)