

ДОСЛІДЖЕННЯ МЕТОДІВ МАСШТАБУВАННЯ ОЗНАК ДЛЯ ПЕРЕДАЧІ ДАНИХ ЧЕРЕЗ БРОКЕРИ ПОВІДОМЛЕНЬ ТАКИХ ЯК RABBITMQ ТА КАФКА

Рамазанов Р. Ш.

Науковий керівник – к.т.н., доц. Ревенчук І. А.

Харківський національний університет радіоелектроніки, каф. ПІ

м. Харків, Україна

e-mail: rasul.ramazanov@nure.ua

This study delves into the exploration of traditional feature scaling methods within the context of RabbitMQ and Kafka message brokers. It meticulously examines vertical, horizontal, and sharding scaling techniques, highlighting their respective advantages and limitations in enhancing the efficiency and reliability of data transmission in microservices architectures. Through an in-depth analysis, it evaluates the applicability of these methods in optimizing the performance of RabbitMQ and Kafka, crucial components for facilitating communication among various system elements. Furthermore, the study discusses optimization possibilities, emphasizing the importance of a nuanced approach tailored to specific system requirements.

Зі зростанням темпу розвитку технологій та обсягів даних у сучасних системах зростає попит на використання ефективного масштабування. При передачі даних через брокери повідомлень, такі як RabbitMQ та Kafka, важливо розглянути методи масштабування ознак. Ці методи є базовою складовою для забезпечення продуктивності та надійності систем, що використовують брокери повідомлень для обміну даними між різними компонентами. В даному дослідженні будуть розглянуті основні методи масштабування ознак для RabbitMQ та Kafka, будуть визначені їх переваги та недоліки, а також розглянуті можливості оптимізації процесу масштабування в контексті цих брокерів повідомлень.

Традиційні методи масштабування ознак для RabbitMQ та Kafka є таких видів як нормалізація (Min-Max Scaling), стандартизація та масштабування до одиничної дисперсії (Unit Variance Scaling) [1].

Для RabbitMQ нормалізація за допомогою Min-Max Scaling може бути реалізована в додатках-виробниках та споживачах даних перед відправленням або прийомом повідомлень, що дозволяє просто і ефективно застосовувати цей метод без необхідності внесення значних змін у сам брокер повідомлень. Зміна діапазону значень ознак на відрізок від 0 до 1 корисна для підготовки даних перед їх відправленням через RabbitMQ, що допомагає зберегти важливу інформацію у вхідних даних та зменшує вплив різниці в шкалі значень ознак. Однак нормалізація Min-Max Scaling може бути чутливою до викидів у вихідних даних, оскільки

змінює діапазон значень на основі мінімального та максимального значень в оригінальному наборі даних, що може призвести до зміни розподілу ознак у разі наявності великої кількості викидів [2].

У системах, які використовують Kafka, нормалізація може бути реалізована на рівні споживачів та виробників даних перед відправленням або прийомом повідомлень, що дозволяє гнучко керувати процесом нормалізації та адаптувати його до конкретних потреб системи. Оскільки Kafka базується на потоках даних, необхідно забезпечити стійкість до змін у форматі повідомлень при застосуванні нормалізації, що означає, що метод нормалізації повинен бути стійким до змін у структурі даних та забезпечувати коректну обробку даних у реальному часі. Нормалізація Min-Max Scaling у Kafka може допомогти зберегти характеристики даних у вхідних потоках даних, забезпечуючи збереження важливої інформації та зменшуючи вплив різниці в шкалі значень ознак на продуктивність системи [3].

Стандартизація в контексті RabbitMQ та Kafka передбачає процес приведення значень ознак до стандартного розподілу з середнім значенням 0 та стандартним відхиленням 1. У випадку RabbitMQ, цей процес може бути реалізований в додатках-виробниках та споживачах даних перед відправленням або прийомом повідомлень. Це дозволяє зберегти важливі характеристики даних та зменшити вплив викидів, що робить стандартизацію корисною для оптимізації продуктивності та надійності систем обробки повідомлень. У випадку Kafka, стандартизація може бути реалізована на рівні споживачів та виробників даних перед відправленням або прийомом повідомлень, забезпечуючи гнучкість та адаптованість до конкретних потреб системи. Збереження характеристик даних та зменшення впливу викидів є ключовими перевагами стандартизації як для RabbitMQ, так і для Kafka, забезпечуючи оптимальну обробку та передачу даних у великих та складних системах обробки повідомлень.

Масштабування до одиничної дисперсії також передбачає приведення значень ознак до однакового масштабу, але відмінність полягає в тому, що в цьому випадку середнє значення може бути будь-яким, але стандартне відхилення має бути рівним 1. Для RabbitMQ, цей процес може бути реалізований в додатках-виробниках та споживачах даних перед відправленням або прийомом повідомлень, що забезпечує простоту впровадження методу без значних змін у брокері повідомлень. Масштабування до одиничної дисперсії дозволяє уникнути впливу великих числових значень на модель та забезпечити стабільність та ефективність її роботи. У випадку Kafka, цей процес може бути реалізований на рівні споживачів та виробників даних перед відправленням або прийомом повідомлень, забезпечуючи гнучкість та адаптованість до конкретних потреб системи. Збереження характеристик даних та зменшення впливу викидів є ключовими перевагами масштабування до одиничної дисперсії

як для RabbitMQ, так і для Kafka, забезпечуючи оптимальну обробку та передачу даних у великих та складних системах обробки повідомлень [4].

У ході дослідження традиційних методів масштабування ознак для RabbitMQ та Kafka було виявлено, що кожен з цих методів має свої переваги та недоліки. Нормалізація, стандартизація та масштабування до одиничної дисперсії дозволяють оптимізувати продуктивність та ресурси систем, проте їх ефективність залежить від конкретного випадку використання та вимог до системи.

Підсумовуючи, важливо підкреслити, що оптимізація процесу масштабування в контексті цих брокерів повідомлень вимагає комплексного підходу та урахування специфіки системи. Правильний вибір методів масштабування, їхнє налаштування та постійний моніторинг дозволять підтримувати високу продуктивність та надійність обміну даними у сучасних мікросервісних системах.

Список використаних джерел

1. Рамазанов Р.Ш., Ревенчук І.А. Підходи до масштабування ознак для передачі даних через меседж брокери на базі мікросервісної архітектури // Матер VII Міжнар наук.-практ конф. "Scientific Research: Theoretical Foundations and Practical Applications". – Відень 24-26.01.2024. – Р.158-162.
2. Alvaro Videla. RabbitMQ in Action. – Manning Publications Co, 2012. – 16 p.
3. Dylan Scott. Kafka Streams in Action. – Manning Publications Co, 2022. – 18 p.
4. Sergiy Zagorodnyuk, Bohdan Sus, Ilona Revenchuk, Oleksandr Bauzha Information Security of Users Rights Assignment via the Software Solutions Based on LDAP // Problem of Infocommunications. Science and Technolpgy (PIC S&T'2020), Kharkiv, Ukraine- 6-9 October 2020.