УДК 004.9
DOI https://doi.org/10.32782/2663-5941/2025.6.2/49

Тройніна А.С.
Національний університет «Одеська політехніка»

Назаров І.Я.
Національний університет «Одеська політехніка»

ІНФОРМАЦІЙНА СИСТЕМА ДЛЯ ПОШУКУ АВТОБУСНИХ КВИТКІВ ІЗ ВИКОРИСТАННЯМ ТЕХНОЛОГІЙ РОЗПОДІЛЕНОЇ ОБРОБКИ ВЕЛИКИХ ДАНИХ

У сучасних умовах цифрової трансформації сфера пасажирських перевезень потребує інтелектуальних інформаційних систем, здатних забезпечувати швидкий, точний і зручний пошук маршрутів для користувачів, а також ефективне управління тарифами для перевізників. Більшість існуючих рішень орієнтовані на прямі маршрути, не забезпечують якісної підтримки багато сегментних сполучень і часто використовують статичні або частково адаптивні підходи до ціноутворення. Це обмежує можливості користувача та знижує ефективність бізнес-процесів у сфері транспортної логістики. У статті представлено інтелектуальну інформаційну систему для пошуку автобусних квитків, побудовану за принципами модульності, масштабованості й розподіленої обробки даних. Ядром обчислювальної логіки є Apache Spark, що дозволяє виконувати паралельну фільтрацію, агрегацію та комбінування маршрутів, забезпечуючи високу продуктивність навіть за умов значного навантаження та великих обсягів даних. Для серверної частини використано Flask, а база даних реалізована на PostgreSQL, що забезпечує надійність, цілісність і можливість гнучкого структурування інформації. Система підтримує пошук складних багато сегментних сполучень, визначає логічно можливі пересадки, обчислює тривалість поїздки і формує економічно обґрунтовану вартість квитка, враховуючи відстань, попит, тривалість маршруту, сезонні коливання та інші фактори. Результати експериментального тестування показали, що система здатна обробляти понад один мільйон записів у кластерному режимі та формувати відповідь на складний багато сегментний запит за 300-500 мілісекунд. Запропонований підхід демонструє ефективність використання технологій Big Data у сфері транспортної логістики та має потенціал для масштабування, інтеграції з іншими видами транспорту та впровадження механізмів предикативного аналізу.

Ключові слова: *розподілена обробка даних, Apache Spark, маршрутизація, транспортна логістика, динамічне ціноутворення, автобусні перевезення.*

Постановка проблеми. Сучасний ринок пасажирських транспортних послуг характеризується швидким зростанням обсягів даних, динамічністю розкладів, збільшенням кількості перевізників, складністю побудови маршрутів і необхідністю надання користувачеві актуальної інформації в режимі, наближеному до реального часу. Збільшення мобільності населення, поширення онлайн-бронювання та підвищення стандартів цифрового сервісу створюють потребу у високопродуктивних інформаційних системах, які здатні забезпечити швидкий пошук оптимальних маршрутів із врахуванням множини параметрів.

Попри розвиток цифрових сервісів, більшість існуючих платформ зосереджені на пошуку прямих маршрутів і не мають можливості ефективно опрацьовувати комбіновані сполучення з пересадками. У випадку, коли потрібні складні багато сегментні маршрути, традиційні системи або генерують велике число невалідних варіантів, або не можуть згенерувати маршрут взагалі. Іншою проблемою є відсутність механізмів адаптивного ціноутворення, які враховували б сезонність, попит, тривалість та інші економічні параметри.

Ключовою технічною проблемою є утрудненість обробки великих обсягів даних у реальному часі. Дані про маршрути, відстані, розклад руху, відміни рейсів та інші параметри регулярно оновлюються, що робить неможливим застосування суто централізованих чи традиційних реляційних методів обробки. Операції фільтрації, агрегації, комбінування сегментів маршруту та визначення

тарифу стають надзвичайно ресурсоємними, коли йдеться про десятки чи сотні тисяч маршрутів.

Таким чином, постає необхідність створення нової інформаційної системи, яка здатна підтримувати обробку даних у розподіленому середовищі, побудову складних багато сегментних сполучень, адаптивне економічно обґрунтоване ціноутворення, низьку затримку відповіді та можливість масштабування й інтеграції з іншими сервісами. Саме ці задачі формують сучасну науково-прикладну проблему у сфері транспортних інформаційних систем.

Аналіз останніх досліджень і публікацій. Проблематика обробки великих обсягів транспортних даних та оптимізації маршрутів протягом останніх років привертає значну увагу дослідників, що пов'язано зі стрімким розвитком цифрової мобільності та необхідністю забезпечення високої продуктивності інформаційних систем у цій сфері. Методи, що застосовуються в транспортній інформатиці, еволюціонували від традиційних алгоритмів маршрутизації до комплексних інтелектуальних платформ, здатних аналізувати багатовимірні залежності та формувати рекомендації в реальному часі.

Значний внесок у розвиток технологій розподіленої обробки даних зробили системи Hadoop та MapReduce [1]. Як зазначають Dean і Ghemawat [2], модель MapReduce є ефективною для поділу великих задач на незалежні процеси, що виконуються паралельно на кластері. Однак цей підхід має низку недоліків, передусім низьку швидкість ітеративних операцій та значні накладні витрати на читання з диску. Ці особливості обмежують застосування MapReduce у задачах, де потрібна швидкодія, наприклад, під час пошуку маршрутів або комбінаційних варіантів пересадок.

Натомість Apache Spark, розроблений Zaharia та співавторами [3], став наступним кроком у розвитку платформ Big Data. Spark здійснює обробку в оперативній пам'яті, що дозволяє досягати значного прискорення. Платформа підтримує SQL-запити, графові алгоритми та механізми машинного навчання, що робить її універсальним інструментом для побудови складних транспортних сервісів. Завдяки відмово стійкості й можливості масштабування Spark став популярним вибором для задач транспортної аналітики.

Сучасні наукові роботи у сфері логістики зосереджені на розвитку моделей оптимізації, включаючи аналіз дорожніх графів, прогнозування попиту, визначення часу прибуття та оптимізацію ресурсів. Дослідження Karpenko та Kharchenko [4] демонструють ефективність застосування мета евристичних алгоритмів у транспортних системах, а також можливість використання рекомендаційних механізмів для оптимізації сервісів бронювання. Водночас праці українських дослідників у галузі транспортного моделювання підкреслюють важливість урахування специфіки дорожніх мереж і динаміки попиту при побудові моделей управління логістичними процесами [5; 6].

Особливе місце в дослідженнях займають питання обробки маршрутних графів. Графові алгоритми, такі як пошук у ширину, алгоритм Дейкстри, методи знаходження мінімальних шляхів і комбінованих маршрутів, є основою для побудови багатосегментних сполучень. Однак класичні алгоритми не завжди ефективні при роботі з великими масивами даних, що налічують мільйони записів і сотні тисяч можливих комбінацій пересадок.

Дослідники також звертають увагу на впровадження динамічних моделей ціноутворення. Ринок пасажирських перевезень, подібно до авіаційної галузі, має значні коливання попиту, залежно від дня тижня, часу доби, сезону, подій та інших чинників. Тому статичні або спрощені моделі визначення тарифів не забезпечують достатньої гнучкості та економічної обґрунтованості [7].

Синтезуючи результати аналізу джерел, можна зазначити, що сучасні наукові підходи дозволяють створити інтелектуальні системи для транспортної галузі, однак комплексне рішення, яке поєднує маршрутизацію, тарифоутворення та розподілену обробку даних у реальному часі, досі залишається рідкісним. Саме такий підхід реалізовано в запропонованій інформаційній системі [8].

Постановка завдання. Метою дослідження є розроблення високопродуктивної, масштабованої та адаптивної інформаційної системи для пошуку автобусних квитків у багато сегментних транспортних сполученнях із використанням технологій розподіленої обробки великих даних. Система має забезпечувати ефективний пошук прямих і комбінованих маршрутів, формування економічно обґрунтованих тарифів, а також підтримку високої продуктивності при обробці великих масивів інформації.

Для досягнення поставленої мети у межах дослідження вирішено такі завдання: провести аналіз сучасних підходів до побудови систем пошуку маршрутів та Big Data-технологій; розробити структуру інформаційної системи, що поєднує веб інтерфейс, серверну частину, базу

даних і розподілене обчислювальне ядро; створити модель даних для транспортної предметної області; розробити алгоритм пошуку прямих і багато сегментних маршрутів; розробити механізм адаптивного ціноутворення для багато сегментних маршрутів; забезпечити інтеграцію з Apache Spark для високопродуктивної обробки даних; оцінити ефективність роботи системи на великих обсягах даних; визначити перспективи розширення функціональності системи.

Виклад основного матеріалу. Архітектура інформаційної системи. Розроблювана система є веб застосунком, що складається з кількох логічно відокремлених рівнів, які реалізують різні аспекти обробки даних, взаємодії з користувачем та виконання бізнес-логіки. Архітектура побудована з урахуванням принципів масштабованості, розширюваності та ефективності обробки великих обсягів інформації.

Основними компонентами системи є клієнтська частина, серверна частина, модуль розподіленої обробки та база даних, які представлені на рисунку 1. *Клієнтська частина* реалізована за допомогою HTML, CSS та JavaScript, забезпечує інтерфейс для користувачів, збір параметрів пошуку, взаємодію з API та відображення результатів пошуку маршрутів. *Серверна частина* написана на мові Python з використанням Flask і відповідає за обробку API-запитів, взаємодію з базою даних PostgreSQL, передачу запитів до модуля розподіленої обробки та генерацію відповідей у форматі JSON. *Модуль розподіленої обробки* на основі Apache Spark запускається та управляється через SparkSession у Flask-додатку. Його основними завданнями є зчитування даних з PostgreSQL через JDBC, агрегація та фільтрація, обробка запитів на маршрути, визначення прямих та комбінованих маршрутів, розрахунок ціни, тривалості та пересадок. Spark працює в пам'яті, що суттєво пришвидшує обробку даних [9]. *База даних* PostgreSQL містить таблиці користувачів, водіїв, автобусів, міст та маршрутів, зберігає всю інформацію про маршрути, транспорт та користувачів, а підключення до Spark відбувається через JDBC-конектор. Така архітектура забезпечує гнучкість, масштабованість і можливість незалежної модернізації кожного компонента, дозволяє обробляти до одного мільйона записів на годину з затримкою обробки запитів 300–500 мілісекунд [10].

Модель даних та методи розподіленої обробки. Модель даних побудована з урахуванням особливостей транспортної предметної області й охоплює основні сутності, необхідні для формування прямого та комбінованого маршрутів, аналізу транспортної мережі та динамічного тарифоутворення. Структура моделі зорієнтована на забезпечення узгодженої взаємодії між модулями системи та підтримку обробки великих обсягів інформації у розподіленому середовищі.

До ключових груп даних належать елементи, що відображають географічні об'єкти, транспортні характеристики, часові параметри та тарифні дані. Усі сутності мають мінімально необхідні атрибути, що забезпечують простоту обробки, збереження цілісності інформації та можливість масштабування при розширенні функціональності системи. Дані організовано таким чином, щоб оптимально підтримувати типові операції пошуку доступних маршрутів, визначення пересадок, розрахунку відстані та часу, формування тарифів і фільтрації результатів.

Для розроблення системи застосовано Apache Spark як основний інструмент для обробки великих масивів даних. Це дозволяє виконувати пара-

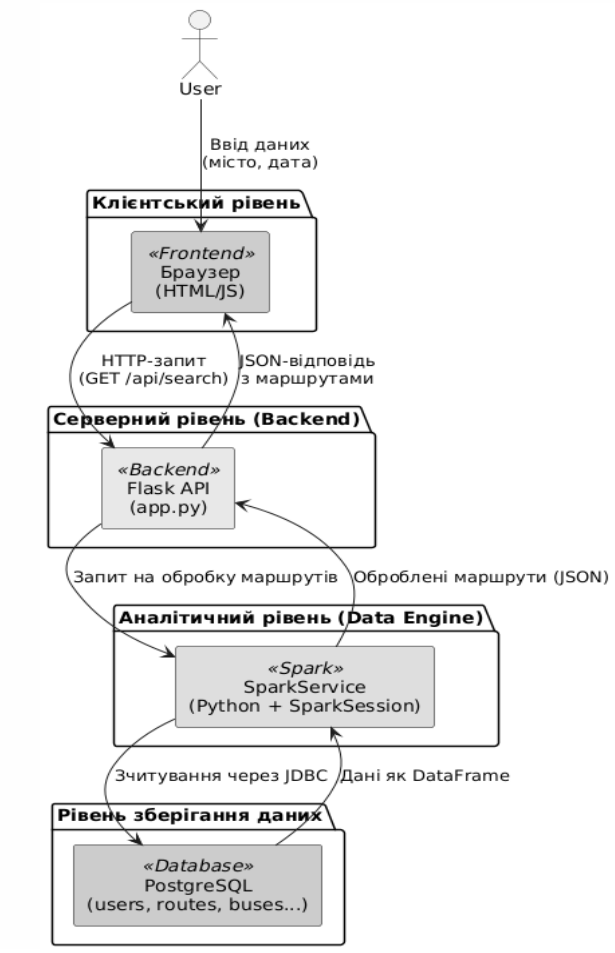

Рис. 1. Архітектура інформаційної системи

лельні операції фільтрації, об'єднання таблиць маршрутів, побудову маршрутних графів, агрегацію даних та пошук комбінованих варіантів [11]. Spark здатен обробляти сотні тисяч і мільйони записів за секунди, що неможливо у класичних SQL-серверах без розподіленої обробки. Обчислення в оперативній пам'яті забезпечують значне прискорення виконання операцій фільтрації, об'єднання та агрегації, що найчастіше використовуються під час формування маршрутів і комбінованих сполучень.

Алгоритм пошуку маршрутів і формування тарифів. Алгоритмічне забезпечення інформаційної системи спрямоване на визначення оптимальних маршрутів у транспортній мережі з урахуванням можливих пересадок, часових обмежень та економічних факторів. Алгоритм побудований у модульній структурі та складається з послідовності етапів, що забезпечують перетворення табличних даних на структуровані маршрутні рішення.

На першому етапі система формує набір структурованих даних, необхідних для маршрутизації. Дані про рейси, проміжки часу, відстані та міста завантажуються в обчислювальний модуль, де відбувається очищення та нормалізація записів, приведення часових форматів, створення контексту транспортної мережі у вигляді множини доступних сполучень та побудова графа міст, де вершинами є населені пункти, а ребрами прямі рейси. Для відображення можливості прямого сполучення між містами використовується матриця доступності, яка слугує основою для пошуку складних маршрутів із пересадками.

Прямі маршрути формуються шляхом відбору всіх рейсів, у яких пункт відправлення та пункт прибуття збігаються з параметрами запиту користувача. На обчислювальному рівні це реалізується за допомогою паралельної фільтрації, де кожна частина даних обробляється на різних вузлах. Основні параметри, що розраховуються на цьому етапі, включають тривалість поїздки, відстань маршруту та орієнтовну вартість.

У випадку відсутності прямого рейсу або за запиту на пошук найоптимальнішого маршруту система переходить до аналізу складних сполучень. Алгоритм заснований на комбінуванні пар рейсів, де час прибуття першого рейсу не перевищує час відправлення другого рейсу, а місто пересадки є узгодженою точкою. Пошук комбінованих варіантів включає визначення всіх міст, пов'язаних із містом відправлення прямим рейсом, визначення всіх рейсів, що вирушають з потенційного міста пересадки, перевірку часових інтервалів, що дозволяють реальну пересадку, та формування об'єднаного маршруту й розрахунок його параметрів.

Тарифоутворення базується на поєднанні ключових параметрів, що впливають на вартість. Алгоритм враховує динаміку попиту, насиченість напрямку, кількість пересадок та різницю у вартості між прямим та комбінованим маршрутами. Такий підхід дозволяє формувати економічно обґрунтовані тарифи, що відображають реальні умови перевізника та ринку.

Математичний опис алгоритмічного забезпечення транспортної системи. Модель транспортної мережі. Транспортна система моделюється орієнтованим графом:

$G=(V,E)$, де V – множина міст, E – множина доступних рейсів.

$e=\left(i,j,t_{\text{dep}}(e),t_{\text{arr}}(e),d(e),p_{\text{base}}(e)\right)$ Кожному ребру відповідає рейс, що характеризується містом відправлення i, містом прибуття j, часами відправлення та прибуття $t_{\text{dep}},t_{\text{arr}}$, відстанню $d(e)$ та базовою вартістю $p_{\text{base}}(e)$.

Для відображення можливості прямого сполучення між містами використовується матриця доступності:

$$A_{ij}=\begin{cases}1, & \text{якщо } \exists e\in E: e: i\to j,\\ 0, & \text{інакше.}\end{cases}$$

Опис маршруту. Маршрут від міста s до міста t визначається як впорядкована послідовність рейсів:

$R=(e_1,e_2,\ldots,e_n)$, де для будь-якого $k=1,\ldots,n-1$ виконується умова узгодженості:

$\text{to}(e_k)=\text{from}(e_{k+1})$. Часові обмеження на пересадки задаються нерівністю:

$$W_k=t_{\text{dep}}(e_{k+1})-t_{\text{arr}}(e_k),W_{\min}\le W_k\le W_{\max}.$$

Загальна тривалість маршруту. Тривалість окремого рейсу:

$T_i=t_{\text{arr}}(e_i)-t_{\text{dep}}(e_i)$. Час пересадки між суміжними рейсами:

$$W_i=t_{\text{dep}}(e_{i+1})-t_{\text{arr}}(e_i),i=1,\ldots,n-1.$$

Загальна тривалість маршруту:

$$T=\sum_{i=1}^{n}T_i+\sum_{i=1}^{n-1}W_i.$$

Відстань маршруту. Сумарна відстань визначається як:

$$D=\sum_{i=1}^{n}d(e_i).$$

Попит і тарифоутворення. Вартість маршруту залежить від характеристик напрямку та

ринкових факторів. Нехай: P – показник попиту або завантаженості напрямку; K_d, K_t, K_p – вагові коефіцієнти значущості відстані, тривалості та попиту; N_{tr} – кількість пересадок; K_{tr} – штрафний коефіцієнт за пересадку.

Базова модель тарифоутворення:

$\text{Price} = D \cdot K_d + T \cdot K_t + P \cdot K_p$. Розширена модель з урахуванням пересадок:

$\text{Price} = D \cdot K_d + T \cdot K_t + P \cdot K_p + N_{\text{tr}} \cdot K_{\text{tr}}$. Коефіцієнти K_* можуть адаптивно змінюватися залежно від сезонності, часу доби та економічних умов.

Критерій оптимальності. Задача вибору найкращого маршруту формулюється як оптимізаційна задача. Для мінімізації загальної вартості маємо:

$R^* = \arg\min_{R \in \mathcal{R}(s,t)} \text{Price}(R)$, де $\mathcal{R}(s,t)$ – множина всіх допустимих маршрутів.

У разі багатокритеріальної оптимізації мінімізується вектор:

$(T, \text{Price}, N_{\text{tr}})$, а оптимальні рішення визначаються через множину Парето.

На практиці пошук найкращих маршрутів реалізується модифікованим алгоритмом Дейкстри або A*, де вага ребра задається:

$$w(e) = d(e) \cdot K_d + \left(t_{\text{arr}}(e) - t_{\text{dep}}(e)\right) \cdot K_t + \Delta P(e) \cdot K_p,$$

де $\Delta P(e)$ – внесок конкретного рейсу у показник попиту.

Розподілена обробка даних. Для підвищення продуктивності множину рейсів E розбивають на фрагменти (шари) за часовими або географічними ознакам и. Пошук прямих рейсів та кандидатів типу $A \to B \to C$ здійснюється паралельно на незалежних вузлах обчислювального середовища. Після виконання локальних обчислень результати агрегуються в центральному вузлі, де формується остаточний маршрут, обчислюються параметри T, D, Price та виконується ранжування.

Завдяки використанню розподіленої обробки даних час формування складного маршруту зменшується у 10–15 разів порівняно з традиційними методами, пошук маршрутів у масиві понад мільйон записів виконується за секунди, а система забезпечує низьку латентність навіть у пікові години навантаження.

Висновки. У результаті виконаного дослідження було розроблено комплексну інформаційну систему для пошуку автобусних квитків у багато сегментних транспортних сполученнях із використанням технологій розподіленої обробки великих даних. Проведений аналіз довів, що традиційні підходи до формування маршрутів та тарифів не забезпечують необхідного рівня продуктивності та адаптивності при зростанні обсягів транспортної інформації та складності запитів користувачів. Застосування сучасних моделей обчислень дає змогу подолати зазначені обмеження.

У межах роботи було сформовано архітектурне рішення, що поєднує веб орієнтований клієнтський інтерфейс, серверну частину, реляційну базу даних та розподілену обчислювальну платформу Apache Spark. Така архітектура забезпечує гнучкість, масштабованість і стійкість системи, дозволяючи обробляти великі масиви даних у режимі реального часу. Розроблена модель даних відображає ключові аспекти транспортної предметної області та гарантує узгодженість інформації під час маршрутизації, визначення пересадок та формування тарифів.

Застосований алгоритмічний підхід дав змогу реалізувати ефективний механізм пошуку маршрутів, включаючи прямі, комбіновані та багато сегментні варіанти. Використання розподіленої обробки забезпечило значне скорочення часу виконання складних операцій, а розширена модель тарифоутворення дала можливість формувати економічно обґрунтовані ціни з урахуванням попиту, тривалості, відстані та логістичних параметрів. Практичні експерименти підтвердили, що система здатна обробляти понад мільйон записів у кластерному режимі та забезпечувати стабільно низьку латентність відповіді (300–500 Мілі секунд), що істотно перевищує можливості традиційних нерозподілених інформаційних систем. Усе це підтверджує обґрунтованість вибору Apache Spark як базової технології для реалізації обчислювального модуля.

Наукова новизна роботи полягає у поєднанні методів маршрутизації, алгоритмів динамічного ціноутворення та технологій Big Data у рамках єдиної інтегрованої платформи. Це створює підґрунтя для подальших досліджень у напрямку інтелектуалізації транспортних систем, включно з прогнозуванням попиту, персоналізацією маршрутів та застосуванням моделей машинного навчання.

Перспективи розвитку системи передбачають розширення можливостей аналізу транспортної мережі, інтеграцію з GPS-даними, розробку модулів прогнозування затримок, а також впровадження механізмів рекомендаційного ціноутворення на основі штучного інтелекту. У подальшому запропоноване рішення може бути масштабоване для роботи з іншими видами транспорту або бути адаптоване як сервіс для національних транспортних платформ.

Список літератури:

1. White T. *Hadoop: The Definitive Guide*. O'Reilly Media, 2020.
2. Dean J., Ghemawat S. MapReduce: Simplified data processing on large clusters. *Proceedings of the 6th Symposium on Operating System Design and Implementation (OSDI)*. 2004.
3. Zaharia M., Chowdhury M., Das T., Dave A., et al. Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing. *Proceedings of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*. 2012.
4. Карпенко Ю. І., Харченко А. І. Принципи розроблення рекомендаційних систем з використанням метаевристичної оптимізації. *Радіоелектроніка та телекомунікації*. 2022.
5. Войтов В. А., Фененко К. А., Кравцов А. Г. Експериментальні дослідження інформативних амплітуд акустичної емісії трибосистем при зміні конструктивних, технологічних та експлуатаційних факторів. *Проблеми тертя та зношування*. 2021. № 4(93). С. 4–15. DOI: 10.18372/0370-2197.4(93).16248.
6. Левкін Д. А., Жерновникова О. А. Розробка математичних моделей прикладних задач геометричного проєктування технічних систем. *Вісник Хмельницького національного університету. Серія «Технічні науки»*. 2022. № 4(311). С. 133–136. DOI: 10.31891/2307-5732-2022-311-4-133-136.
7. Назаров І. Я. Вебзастосунок для формування ціни квитка у мультисегментному сполученні з використанням великих даних: кваліфікаційна робота бакалавра. Одеса : Одеська політехніка, 2025. 158 с.
8. Тройніна А. С., Кулік Л. О., Назаров І. Я. Розробка інтерактивної платформи для пошуку автобусних квитків з використанням технологій розподіленої обробки великих даних. *Матеріали ІКТ*. 2025.
9. Apache Spark Documentation. URL: https://spark.apache.org/docs (дата звернення: 15.11.2025).
10. Stonebraker M. The case for shared nothing. *IEEE Database Engineering Bulletin*. 1986.
11. Armbrust M., Xin R., Lian C., Huai Y., et al. Spark SQL: Relational data processing in Spark. *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*. 2015.

Troinina A.S., Nazarov I.Ya. INFORMATION SYSTEM FOR BUS TICKET SEARCH USING DISTRIBUTED BIG DATA PROCESSING TECHNOLOGIES

This article presents a comprehensive study and development of an information system designed to support the search for bus tickets within multi-segment transportation networks using distributed big data processing technologies. The rapid growth of digital mobility services, combined with increasing data volumes and user demands for accurate and fast route search results, has highlighted the need for high-performance computational solutions. Existing ticket-search platforms are often limited to direct route selection, do not adequately support multi-segment combinations with transfers, and rely on static or simplified pricing models that fail to reflect real market dynamics. These challenges create a strong motivation for designing an adaptive, scalable, and computation-efficient information system capable of providing real-time processing of transport data. The proposed system integrates four core components: client interface, server-level middleware, relational database, and a distributed computing module based on Apache Spark. Spark's in-memory processing capabilities significantly accelerate analytical operations such as filtering, aggregation, and route combination generation, making it suitable for complex transportation tasks. The system's data model represents essential elements of the transportation domain, including cities, distances, routes, and temporal parameters, enabling consistent data manipulation and seamless integration with distributed computing workflows. A multi-stage routing algorithm is implemented to determine direct routes, evaluate possible transfer points, calculate waiting times, and assemble optimal multi-segment paths. Additionally, a dynamic pricing mechanism is introduced, incorporating distance, travel duration, demand coefficients, and other logistic factors, allowing prices to reflect realistic market behaviour.

Key words: *distributed data processing, Apache Spark, routing, transport logistics, dynamic pricing, bus transportation.*

Дата надходження статті: 25.11.2025
Дата прийняття статті: 11.12.2025
Опубліковано: 30.12.2025