

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**  
**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ**  
**“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ**  
**ІМЕНІ ІГОРЯ СІКОРСЬКОГО”**

Факультет прикладної математики  
Кафедра програмного забезпечення комп’ютерних систем

**КУРСОВИЙ ПРОЕКТ**  
**ТЕХНІЧНЕ ЗАВДАННЯ**  
з дисципліни “Бази даних”

спеціальність 121 – Програмна інженерія

на тему: **Моніторингова система спостереження за рухом наземного**  
**громадського транспорту**

**Студент**

групи КП-71

Кравчук Аркадій Андрійович

\_\_\_\_\_  
(підпис)

**Викладач**

к.т.н, доцент кафедри

**СПіСКС**

Петрашенко А.В.

\_\_\_\_\_  
(підпис)

Київ – 2020

**ЗМІСТ**

	стор.
1. НАЙМЕНУВАННЯ ТА ГАЛУЗЬ ЗАСТОСУВАННЯ РОЗРОБКИ .....	3
2. ДАТА ПОЧАТКУ ТА ЗАКІНЧЕННЯ ПРОЕКТУ .....	3
3. МЕТА РОЗРОБКИ .....	3
4. ВИМОГИ ДО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ .....	3
5. ОБҐРУНТУВАННЯ ВИБОРУ СУБД .....	5
6. ВИМОГИ ДО ІНТЕРФЕЙСУ КОРИСТУВАЧА.....	6
7. ВИБІР ЗАСОБІВ РОЗРОБКИ.....	6
8. ЕТАПИ РОЗРОБКИ .....	7

## **1. НАЙМЕНУВАННЯ ТА ГАЛУЗЬ ЗАСТОСУВАННЯ РОЗРОБКИ**

Найменування: моніторингова система спостереження за рухом наземного громадського транспорту.

Галузь застосування: диспетчеризація роботи громадського транспорту в режимі реального часу шляхом контролю наявності на маршруті машин та виконання встановленого розкладу руху; планування маршрутів громадського транспорту з оцінкою приблизної тривалості подорожей у встановлений час та день тижня, враховуючи статистичні показники; онлайн система сповіщення про затримки руху громадського транспорту.

## **2. ДАТА ПОЧАТКУ ТА ЗАКІНЧЕННЯ ПРОЕКТУ**

Дата початку проекту: 26.02.2020

Дата закінчення проекту: 21.05.2020

## **3. МЕТА РОЗРОБКИ**

Метою розробки є створення програмного забезпечення, яке буде взаємодіяти з постріляційною базою даних, проводити заміри досліджуваного об'єкта і подальший аналіз, оцінку, порівняння отриманих результатів для виявлення певних закономірностей, тенденцій, змінних і їх динаміки. А також здобуття навичок використання СУБД та інструментальними засобами підтримки розробки додатків для подібних баз даних, оформлення відповідного текстового, програмного та ілюстративного матеріалу у формі проектної документації.

## **4. ВИМОГИ ДО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ**

- Вимоги до генерації даних:

Для генерації даних слід використовувати відкриті набори даних департаменту транспортної інфраструктури КМДА, а також сайт КП Київпастрас. Для цього слід розробити відповідний компонент ПЗ, який буде щонайменше кожні 60 секунд завантажувати актуальні GPS-дані бортового обладнання про місцезнаходження наземного

пасажирського транспорту у форматі GTFS Realtime (докладніше про формат: <https://developers.google.com/transit/gtfs-realtime/reference>). Ці дані доступні у форматі json за наступним посиланням: <http://193.23.225.214:732/api/gps/X/Y>, де X та Y – дати у форматі ISO YYYY-MM-DDThh:mm:ss, які вказують в межах якого часового періоду збирати дані. Приклад: <http://193.23.225.214:732/api/gps/2020-03-10T20:02:00/2020-03-10T20:03:00>.

Також даний компонент повинен одноразово при старті отримувати загальну інформацію про маршрути, їх розклади руху та координати зупинок у форматі GTFS Static (докладніше про формат: <https://developers.google.com/transit/gtfs>). Дані доступні як zip архів за посиланням: <http://193.23.225.214:732/api/static>.

- Вимоги до фільтрації та валідації даних:

Щоразу при отриманні нових даних про місцезнаходження громадського транспорту слід фільтрувати наступним чином: в масиві повинні залишитися записи лише з унікальними ідентифікаторами машин (їх номерних знаків). При виявленні декількох таких записів слід залишити такий, що має найновішу відмітку часу. Також слід відфільтровувати ті транспортні засоби, які знаходяться поза встановленим маршрутом. Окремо взятий запис слід вважати валідним, якщо в ньому обов'язково присутні такі дані: широта та довгота, час зняття замірів, номер маршруту, номерний знак, вид транспортного засобу (трамвай, тролейбус, автобус).

Для перевірки на валідність отриманого розкладу руху треба порівнювати з розміщеним розкладом на сайті КП Київпастрас для конкретного маршруту.

- Вимоги до засобів реплікації та масштабування:

Розроблений програмний комплекс має використовувати СУБД Cassandra, яка повинна горизонтально масштабуватись шляхом додання нових вузлів, обов'язково використовуючи при цьому для

кожного запису partition key; реплікувати дані шляхом налаштування для усіх keyspace значення replication\_factor.

- Вимоги до аналізу даних:

Отримані дані повинні бути проаналізовані створеним програмним забезпеченням в наступний спосіб: оцінювати якість виконання розкладу руху певним транспортним засобом, маршрутом, видом громадського транспорту чи взагалі компанії за заданий проміжок часу. Оцінювання слід проводити у відносній шкалі 0...100 % відповідно до кількості вчасних відправлень від зупинок. Необхідно знаходити зв'язок (кореляцію) отриманої якості з кількістю заторів (та за можливістю з погодними умовами) за обраний період часу, регресію залежності цієї якості в залежності від періоду доби та дня тижня. Також слід оцінювати відсоток серйозних затримок руху громадського транспорту, скласти рейтинг маршрутів з найчастішими затримками у русі. Для реалізації алгоритмів використовувати математичні бібліотеки або методи Big Data тощо.

- Вимоги до резервування та відновлення даних:

Резервування даних здійснювати шляхом використання реплікації. У разі втрачання зв'язку з певним вузлом, при підключенні вже нового вузла дані мають бути відновлені в ньому. За потреби здійснювати backups чи snapshots та відновлення з них.

## 5. ОБҐРУНТУВАННЯ ВИБОРУ СУБД

Для поставленої задачі було обрано СУБД Cassandra — вільна, відкрита, розподілена, wide column store, NoSQL. Створена для роботи з високомасштабованими і надійними сховищами величезних масивів даних. Забезпечує високу доступність даних з асинхронною нецентралізованою реплікацією даних, що дозволяє для всіх користувачів виконувати операції з низькою затримкою. Було обрано саме таку БД з декількох причин, а саме: в режимі реального часу формується дуже багато інформації (адже для цієї галузі

характерна велика кількість транспортних засобів, які обслуговують місто з населенням понад мільйона мешканців), схема даних якої з часом може змінюватись (встановлюються різні GPS-трекери, які можуть відправляти певні додаткові данні тощо), що зумовлює неможливість використання звичайних реляційних баз даних. Тому для цього було обрано NoSQL БД сімейства колоночного типу, щоби пришвидшити зчитування однотипних даних, наприклад потрібно знайти переміщення транспорту за весь маршрут – слід зчитати по чергово усі координати. Також Cassandra найпопулярніша серед аналогів даного підкласу СУБД, а ще їй притаманні дві ознаки з теореми CAP - Availability та Partition tolerance, що є необхідним для моніторингових систем реального часу.

## **6. ВИМОГИ ДО ІНТЕРФЕЙСУ КОРИСТУВАЧА**

Інтерфейс користувача виконати в консольному варіанті. Повинні бути доступні наступні розділи (пункти в меню) в інтерфейсі: налаштування засобів та підсистем (генерації даних, масштабування, резервування та відновлення даних), запуск/завершення їх роботи, генерація звітної інформації (графіків, діаграм) у вигляді збережених файлів-зображень. Звітна інформація стосується візуалізації роботи засобів аналізу даних (визначення трендів, регресії, кореляції).

## **7. ВИБІР ЗАСОБІВ РОЗРОБКИ**

У якості засобів розробки було обрано мову Python 3.6. Для взаємодії з СУБД – бібліотеку DataStax Driver for Apache Cassandra. Для аналізу даних повинні використовуватися відповідні бібліотеки напрямку Data Science чи Big Data, як-от: numpy, pandas, matplotlib, scipy.

## 8. ЕТАПИ РОЗРОБКИ

№	Назва етапів розроблення	Термін виконання
1	Затвердження теми курсової роботи. Опрацювання відповідної літератури. Розроблення та узгодження технічного завдання.	12.03.2020
2	Аналіз постановки задачі	22.03.2020
3	Розробка засобів генерації даних.	05.04.2020
4	Додавання засобів фільтрації та валідації даних.	12.04.2020
5	Реалізація зберігання, реплікації та масштабування інформації розробленої моніторингової системи.	19.04.2020
6	Додавання засобів аналізу даних.	03.05.2020
7	Додавання засобів резервування та відновлення даних (з урахуванням необхідності підключення додаткового комп'ютера як елемента горизонтального масштабування).	10.05.2020
8	Тестування програми	15.05.2020
9	Аналіз результатів. Підготовка матеріалів курсового проекту та оформлення пояснювальної записки	17.05.2020
10	Захист курсової роботи	21.05.2020