

*Скороход О.М., студентка 1 курсу спеціальності 122 «Комп'ютерні науки»  
Горяшин А.С., асистент кафедри інформаційних технологій*

## **ВИКОРИСТАННЯ МОЖЛИВОСТЕЙ МОВИ PYTHON ДЛЯ РОБОТИ З ВЕЛИКИМИ ОБ'ЄМАМИ ДАНИХ**

*Донецький національний університет імені Василя Стуса*

У цьому науковому документі розглядаються можливості мови програмування Python для обробки великих обсягів даних. Постійно зростаючий обсяг даних, доступних у сучасному світі, ставить перед галузями науки та промисловості нові виклики. Python, як популярна мова програмування, надає велику кількість бібліотек і інструментів, які допомагають ефективно аналізувати й обробляти великі масиви даних.

У цій статті розглядаються такі функції Python, як бібліотеки для роботи з даними, зокрема Pandas, NumPy і SciPy. Будуть досліджені методи використання цих бібліотек для інтерактивного аналізу даних, візуалізації та маніпулювання великими наборами даних. Також буде проаналізовано можливість використання бібліотеки Dask для паралельних обчислень і розподіленої обробки даних.

Дослідження також включатиме огляд інструментів і технік для роботи з базами даних, включаючи SQLite, PostgreSQL і MongoDB, з використанням мови Python. Будуть розглянуті можливості створення та керування базами даних, виконання запитів та аналізу результатів.

Результати цієї роботи покажуть, що Python є потужним інструментом для обробки великих обсягів даних. Він надає широкий спектр бібліотек і інструментів для полегшення аналізу даних, обробки та візуалізації. Python також підтримує паралельні обчислення та розподілену обробку даних, що дозволяє ефективно обробляти великі набори даних.

### **Бібліотеки для обробки даних:**

Python має багато потужних бібліотек, які дозволяють легко й ефективно маніпулювати великими обсягами даних. Однією з найпопулярніших бібліотек є Pandas. Він забезпечує зручні структури даних, такі як DataFrame, які забезпечують швидкий доступ і обробку великих обсягів даних. Бібліотека NumPy надає потужні функції для математичних обчислень на масивах даних. Це важливий момент при роботі з великими обсягами даних. Бібліотека SciPy надає багато функцій, корисних для наукових обчислень. В. Статистика, оптимізація, обробка сигналів.

### **Інтерактивний аналіз та візуалізація даних:**

Python також надає потужні інструменти для інтерактивного аналізу та візуалізації великих обсягів даних. Такі бібліотеки, як Matplotlib і Seaborn,

дозволяють створювати графіки та візуалізації, які полегшують розпізнавання та розуміння ваших даних. Ноутбуки Jupyter популярні серед вчених для виконання інтерактивного аналізу даних. Поєднайте код, обчислення та пояснення в одному документі та використовуйте Python для створення звітів та опитувань.

### **Паралельні обчислення та розподілена обробка даних:**

Завдяки бібліотеці Dask Python пропонує можливість паралельного обчислення та розподіленої обробки великих обсягів даних. За допомогою Dask ви можете розподіляти обчислення на своєму кластері та використовувати багатоядерні процесори для ефективної обробки даних. Це особливо важливо під час роботи з великими наборами даних, оскільки обчислення можуть зайняти багато часу. Проведення інтерактивного аналізу даних є дуже важливим для науковців. Поєднайте код, обчислення та пояснення в одному документі та використовуйте Python для створення звітів та опитувань.

### **Паралельні обчислення та розподілена обробка даних:**

Завдяки бібліотеці Dask Python пропонує можливість паралельних обчислень і розподіленої обробки великих обсягів даних. За допомогою Dask ви можете розподіляти обчислення на своєму кластері та використовувати багатоядерні процесори для ефективної обробки даних. Це особливо важливо під час роботи з великими обсягами даних, оскільки обчислення можуть зайняти багато часу.

### **Оснвні операції з базою даних:**

Python також надає зручні інструменти баз даних для ефективного читання, запису та аналізу великих обсягів даних. Бібліотека SQLite дозволяє створювати локальні бази даних і керувати ними. З іншого боку, такі бібліотеки, як PostgreSQL і MongoDB, пропонують потужні можливості для роботи з реляційними та нереляційними базами даних.

Таким чином, використання функцій мови Python для обробки великих обсягів даних є ефективним і потужним підходом. Python надає зручні бібліотеки, інструменти та методи для аналізу, обробки та візуалізації даних. Паралельні обчислення та розподілена обробка даних дають змогу ефективно обробляти великі обсяги даних і скорочувати час обчислення. Крім того, робота з базами даних на мові Python спрощує доступ до даних і аналіз даних.

Таким чином, використання можливостей мови Python є ключовим фактором успішної роботи з великими обсягами даних, що дозволяє ефективно вирішувати завдання в науці, промисловості та інших сферах, де великі обсяги даних є нормою.

### **Список літератури:**

1. McKinney, W. (2018). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media.
2. VanderPlas, J. (2016). *Python Data Science Handbook: Essential Tools for Working with Data*. O'Reilly Media.
3. Електронний ресурс: <https://towardsdatascience.com/>
4. Електронний ресурс: <https://www.python.org/>
5. Електронний ресурс: <https://dask.org/>