

Прикладне програмування в науках про Землю

Практична робота №3

доц. Онищук В.І.
доц. Демидов В.К.
аспір. Охрімчук Р.Ю.

Хід роботи

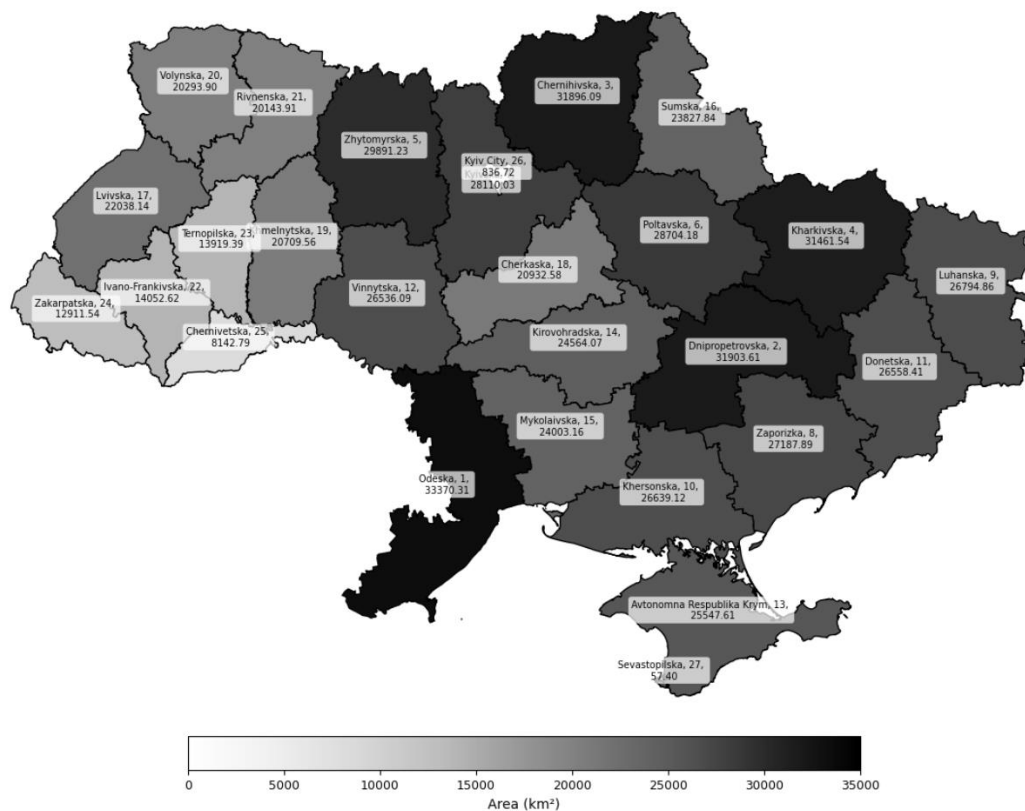
1. знайомство з основами маніпуляцій із векторними даними та їх візуалізацією на прикладі використання бібліотеки GeoPandas;
2. знайомство із паралельним обчисленням багатовимірних растрових геоданих та поняттям «Lazy-loading» на прикладі використання бібліотеки Dask;
3. знайомство із комбінованим використання растрових та векторних даних для комплексного геопросторового аналізу;
4. зробити звіт про виконану роботу в електронному вигляді.

1. Знайомство з основами маніпуляцій із векторними даними та їх візуалізацією на прикладі використання бібліотеки GeoPandas

1.1 Потрібно інсталювати бібліотеку georandas у вже існуюче середовище ods_env, виконавши наступну команду в командному рядку Anaconda Prompt: **pip install geopandas**

1.2 Для виконання практичної роботи потрібно самостійно детальніше ознайомитися з даною бібліотекою на її офіційному веб-ресурсі: **<https://geopandas.org/en/stable/>**

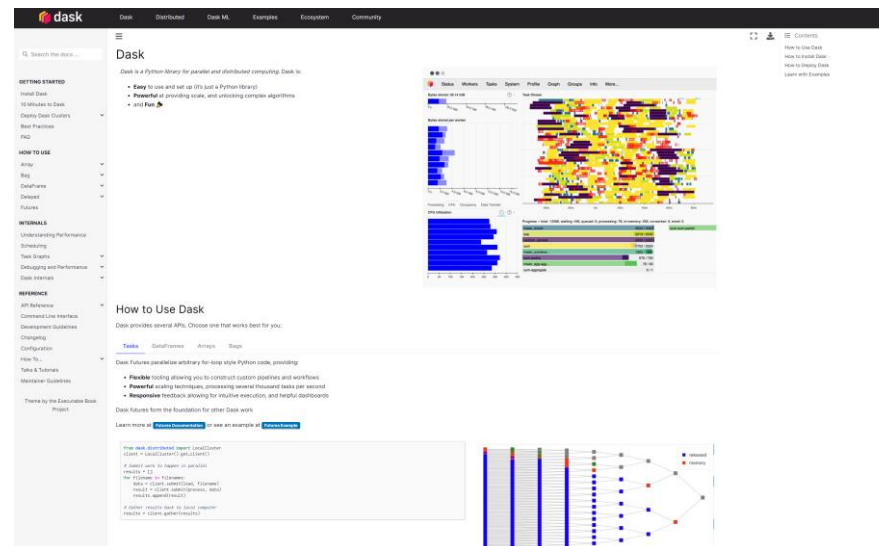
1.3 Необхідно виконати всі завдання у файлі **lab_3_1.ipynb**, який знаходиться у папці - **./lab_3**. Завдання буде вважатися виконаним, якщо буде отримано дане геоображення, але із стар='viridis'. Для прикладу, було використано colormap='Grays'



2. Знайомство із паралельним обчисленням багатовимірних растрових геоданих та поняттям «Lazy-loading» на прикладі використання бібліотеки Dask

Бібліотека Dask підтримує паралельне обчислення та має інструменти для його масштабування. Вона також дозволяє працювати з об'ємом даних, які перевищують розмір оперативної пам'яті, наявний у системі, розбиваючи його на менші частини - **chunks** і виконуючи обчислення паралельно.

'Lazy-loading' - це концепція, за якої дані завантажуються в пам'ять лише тоді, коли це дійсно необхідно для виконання обчислень. Даний підхід дозволяє визначити серію обчислень як граф задач, де фактичні обчислення відкладаються до моменту виклику функції, яка ініціює виконання (наприклад, **compute()**). Даний підхід дозволяє значно оптимізувати використання ресурсів і зменшити час обробки, особливо при роботі з великими наборами даних.



(Детальніше із бібліотекою Dask, можна ознайомитися за посилання: <https://docs.dask.org/en/stable>)



2. Знайомство із паралельним обчисленням багатовимірних растрових геоданих та поняттям «Lazy-loading» на прикладі використання бібліотеки Dask

2.1 Паралельна обробка даних забезпечує одночасне виконання декількох процесів, що допомагає отримувати результати обчислень швидше та проводити їх, маючи обмежений обсяг ресурсів. Даний підхід є реалізованим для роботи із багатовимірними просторово-часовими даними у Open Data Cube (ODC) із використанням Dask. Для виконання практичної роботи потрібно ознайомитися з поняттям **chunks** за посиланням:

<https://docs.dask.org/en/latest/array-chunks.html>);

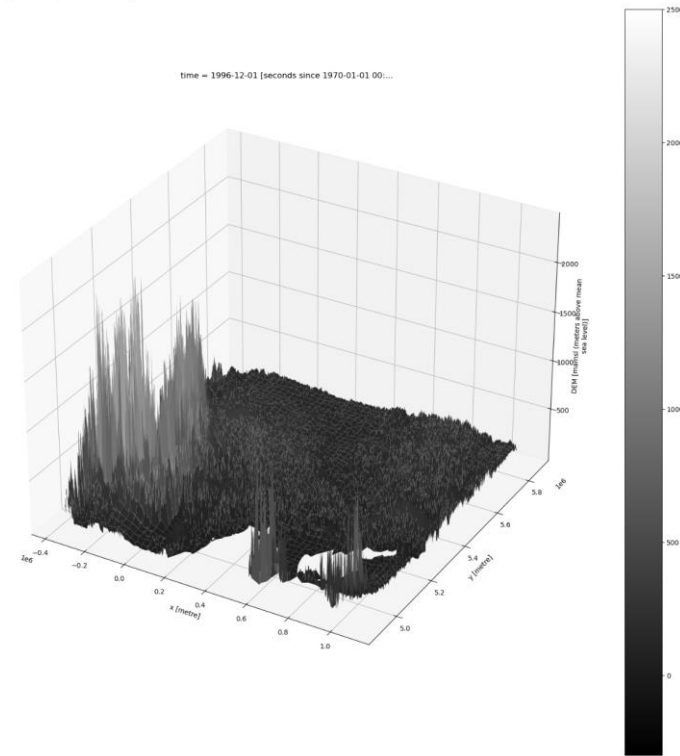
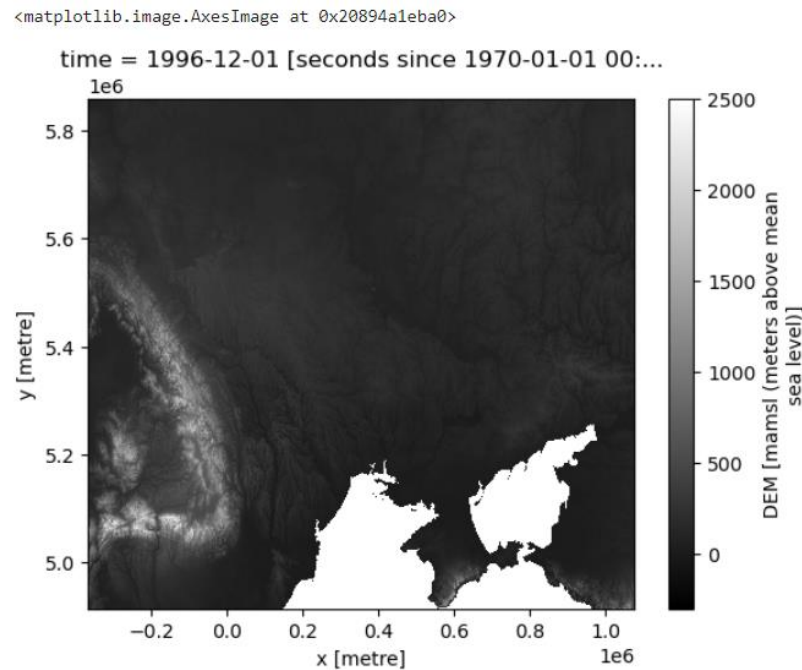


(Інформаційна панель Dask. Джерело: <https://blog.dask.org/2021/11/02/choosing-dask-chunk-sizes>)

2. Знайомство із паралельним обчисленням багатовимірних растрових геоданих та поняттям «Lazy-loading» на прикладі використання бібліотеки Dask

2.2 Потрібно ознайомитися з використанням **chunks** у середовищі ODC для завантаження даних. Для цього необхідно детально ознайомитися з методом **load** класу **Datacube**, який належить до бібліотеки **datacube**, за посиланням: <https://datacube-core.readthedocs.io/en/latest/api/indexed-data/generate/datacube.Datacube.load.html>

2.3 Необхідно виконати всі завдання у файлі **lab_3_2.ipynb**, який знаходиться у папці - **./lab_3**. Завдання буде вважатися виконаним, якщо будуть отримані представлені нижче 2 геозображення, із `star='terrain'`. Для прикладу, було використано `colormap='Grays'`



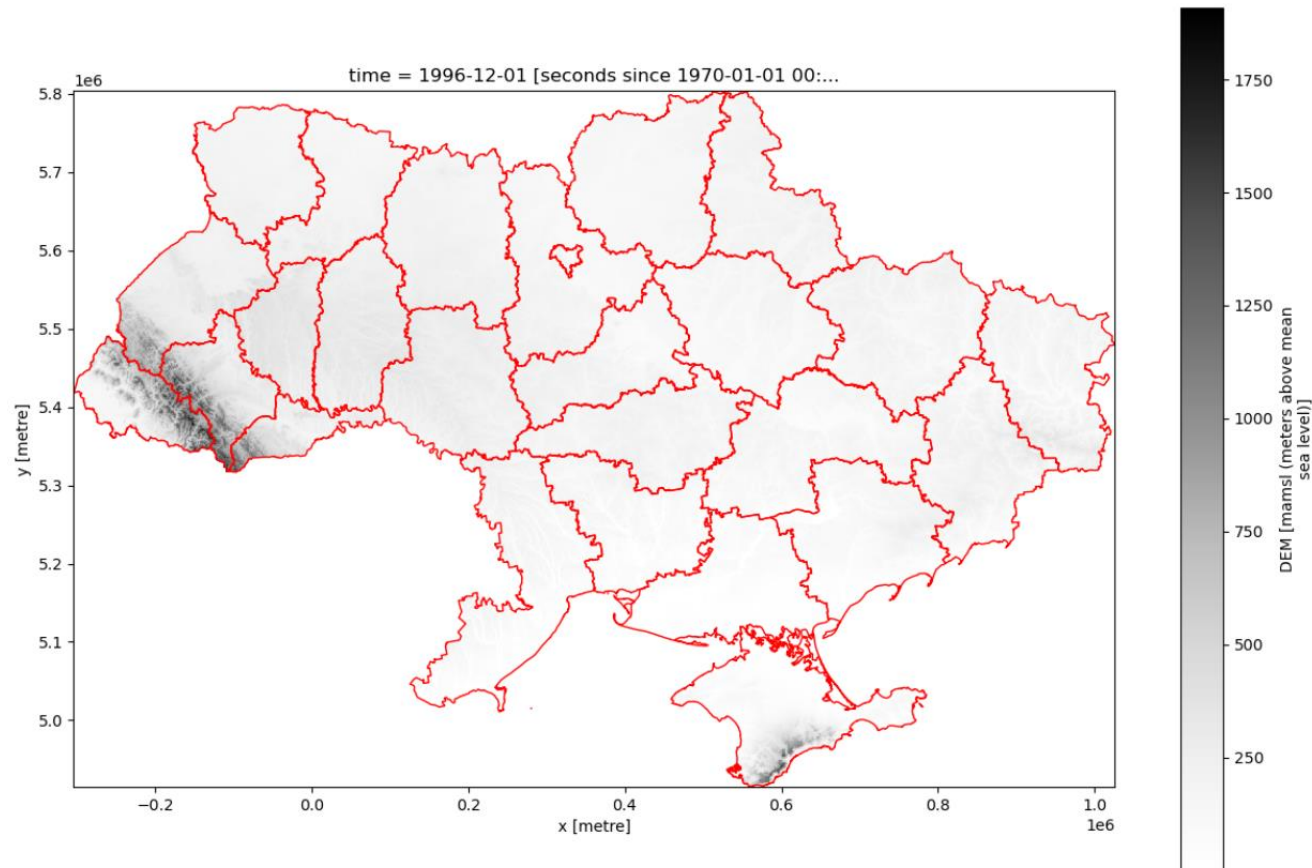
Також рекомендую:

1. Ознайомитися із тим, як отримати доступ до інформаційної панелі Dask (127.0.0.1:8787/status): <https://docs.dask.org/en/latest/dashboard.html>;
2. Встановити Dask JupyterLab Extension: (<https://github.com/dask/dask-labextension#dask-jupyterlab-extension>).

Це допоможе краще зрозуміти, як Dask опрацьовує дані у вирішенні цільової задачі і допоможе вам оптимізувати його використання.

3. Знайомство із комбінованим використання растрових та векторних даних для комплексного геопросторового аналізу

3.1 Даний пункт дозволить закріпити попередньо наданий матеріал і навчить поєднувати растрові та векторні дані для проведення геопросторового аналізу. Для його виконання необхідно виконати всі завдання у файлі **lab_3_3.ipynb**, який прикріплено в **Google Classroom**. Завдання буде вважатися виконаним, якщо буде отримано дане геоображення, але із `map='terrain'`. Для прикладу, було використано `colormap='Grays'`



4. Зробити звіт про виконану роботу в електронному вигляді

4.1 Для успішного виконання практичної роботи необхідно підготувати звіт про виконану роботу. У звіті потрібно детально описати хід виконання кожного завдання та надіслати його у Google Classroom. Також необхідно прикріпити три .ipynb файли з виконаними завданнями.