

Przetwarzanie Danych Masowych

Laboratorium – wprowadzenie

Krzysztof Galus, Piotr Giedziun, Mateusz Gniewkowski

`krzysztof.galus@pwr.edu.pl`

`piotr.giedziun@pwr.edu.pl`

`mateusz.gniewkowski@pwr.edu.pl`

Wydział Informatyki i Telekomunikacji

Politechnika Wrocławska

semestr 2025/2026

Bezpieczeństwo i Higiena Pracy

- ▶ Na zajęciach obowiązują zasady BHP pracowni oraz regulamin kursu.
- ▶ W kursie może uczestniczyć osoba, która zaakceptowała zasady BHP oraz reguły zaliczenia.
- ▶ Sprzęt i środowiska uruchamiamy tak, aby nie przeszkadzać innym i nie niszczyć wspólnej infrastruktury.

Słowem wstępu...

O czym jest ten kurs? Czym będziemy się tutaj zajmować?

- ▶ przetwarzanie danych w skali większej niż “na jednym laptopie”,
- ▶ narzędzia: Linux, Docker, kolejki, Spark, Flink, Kubernetes,
- ▶ druga część semestru: **mini-projekt** łączący kolejne etapy pracy z danymi.

Najpierw poznajemy klocki, potem składamy z nich działający system.

Organizacja zajęć

- ▶ Plan typowego laboratorium:
 - krótkie wprowadzenie do następnego zadania,
 - prezentowanie i oddawanie bieżącej listy.
- ▶ Pracujemy przez **ePortal Politechniki Wrocławskiej**.
- ▶ Rozwiązania zadań trzeba zamieszczać na ePortalu w miejscu wskazanym przez prowadzących.
- ▶ Plagiaty nie są dopuszczalne.
- ▶ Oddanie listy wymaga osobistego zaprezentowania rozwiązania.
- ▶ Prowadzący może sprawdzić, czy student rozumie oddany kod, decyzje projektowe i uzyskane wyniki.

Ocenianie

Poszczególne listy: wyniki $\in [2,0; 5,0]$. Każda lista musi być zaliczona, czyli mieć wynik co najmniej 3,0.

$$\text{ocena końcowa} = 0,5 \cdot \bar{L}_{\text{zwykłe}} + 0,5 \cdot \bar{P}_{\text{mini-projekt}}$$

- ▶ $\bar{L}_{\text{zwykłe}}$ – średnia z ocen za zwykłe listy laboratoryjne,
- ▶ $\bar{P}_{\text{mini-projekt}}$ – średnia z ocen za etapy mini-projektu.

Zasady zaliczenia

1. Warunkiem koniecznym zaliczenia kursu jest zaliczenie wszystkich przewidzianych zajęć laboratoryjnych.
2. Maksymalnie dwie nieobecności mogą być nieusprawiedliwione; każda kolejna wymaga usprawiedliwienia.
3. Zaległą listę trzeba oddać na najbliższym laboratorium.
4. W trakcie jednego laboratorium można oddać maksymalnie dwie listy.
5. Etapy mini-projektu mogą być oceniane osobno.

Tematy laboratoriów: listy zwykłe

- Lab 1. Wprowadzenie do zajęć
- Lab 2. Podstawy wykorzystania infrastruktury chmurowej – Unix, Shell
- Lab 3. Podstawy wykorzystania kontenerów – Docker, containerd
- Lab 4. Ewaluacja metod zrównoleglania zadań – Python, wątki i procesy
- Lab 5. Wykorzystanie przetwarzania opartego o broker wiadomości
- Lab 6. MapReduce w przetwarzaniu danych masowych – Spark
- Lab 7. MapReduce w przetwarzaniu danych masowych – Flink

Tematy laboratoriów: mini-projekt

Lab 8. Rozproszone pozyskiwanie danych

Lab 9. Czyszczenie danych i ekstrakcja cech

Lab 10. Wybór i ewaluacja modeli uczenia maszynowego

Lab 11. Kubernetes w przetwarzaniu danych masowych – K3S

Lab 12. Uruchomienie metody w środowisku Kubernetes

Lab 13. Udostępnienie metody – API gRPC

Lab 14. Wykorzystanie metody – API REST/GraphQL + SPA

Lab 15. Prezentacja całościowego rozwiązania

Wymagania sprzętowe i oprogramowanie

- ▶ Najwygodniej pracować na własnym laptopie.
- ▶ Alternatywnie można używać maszyn wirtualnych w chmurze.
- ▶ Spróbujemy przygotować klaster do dalszych laboratoriów.
- ▶ **Docker musi być zainstalowany i wstępnie przetestowany.**
- ▶ Polecamy Linuxa; Windows na własną odpowiedzialność.

Pozostałe informacje

- ▶ Instrukcje dotyczące poszczególnych zajęć będą pojawiać się w trakcie semestru.
- ▶ Realizacja zadań wymaga osobistego zaprezentowania wykonanej pracy oraz krótkiego omówienia; prowadzący weryfikuje, czy student rozumie oddane rozwiązanie.
- ▶ Wszystkie rezultaty prac należy zamieszczać na ePortalu.
- ▶ Bieżące wyniki będą udostępniane w postaci zanonimizowanej.
- ▶ Wszelkie sprawy ważne będą komunikowane przez prowadzących pocztą uczelnianą.

Ostatnia ważna uwaga

Osoby, które ze względu na stan zdrowia, niepełnosprawność lub inne obiektywne przesłanki mogą mieć szczególne potrzeby związane ze sposobem realizacji zajęć, zaliczeniem bądź przygotowaniem materiałów, proszone są o kontakt z prowadzącymi.

Będziemy starać się, aby na zajęciach każdy miał równe prawo do zdobycia wiedzy i rozliczenia się z niej.

W uwzględnieniu szczególnych potrzeb nie chodzi o obniżenie wymagań, a o wyrównanie szans.

To wszystko na dziś.

Pytania?

Dziękuję za uwagę!