

Розробка системи оркестрації розподілених обчислень на кластері

науковий керівник Черкасов Д.І., ст. викладач, к.н.

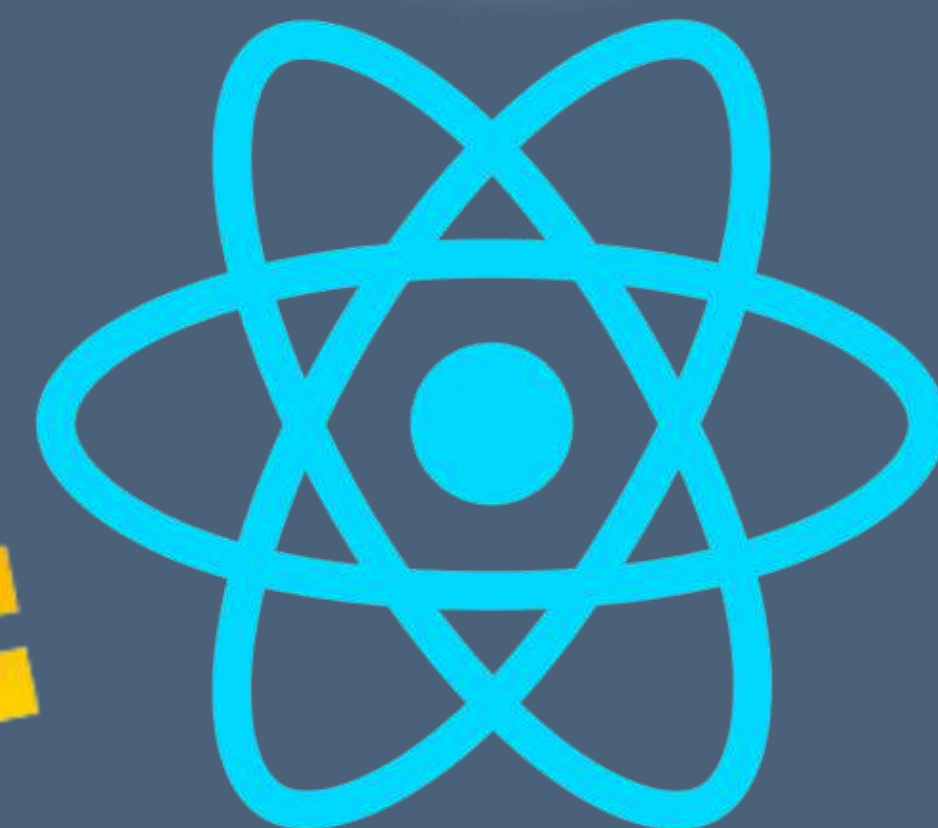
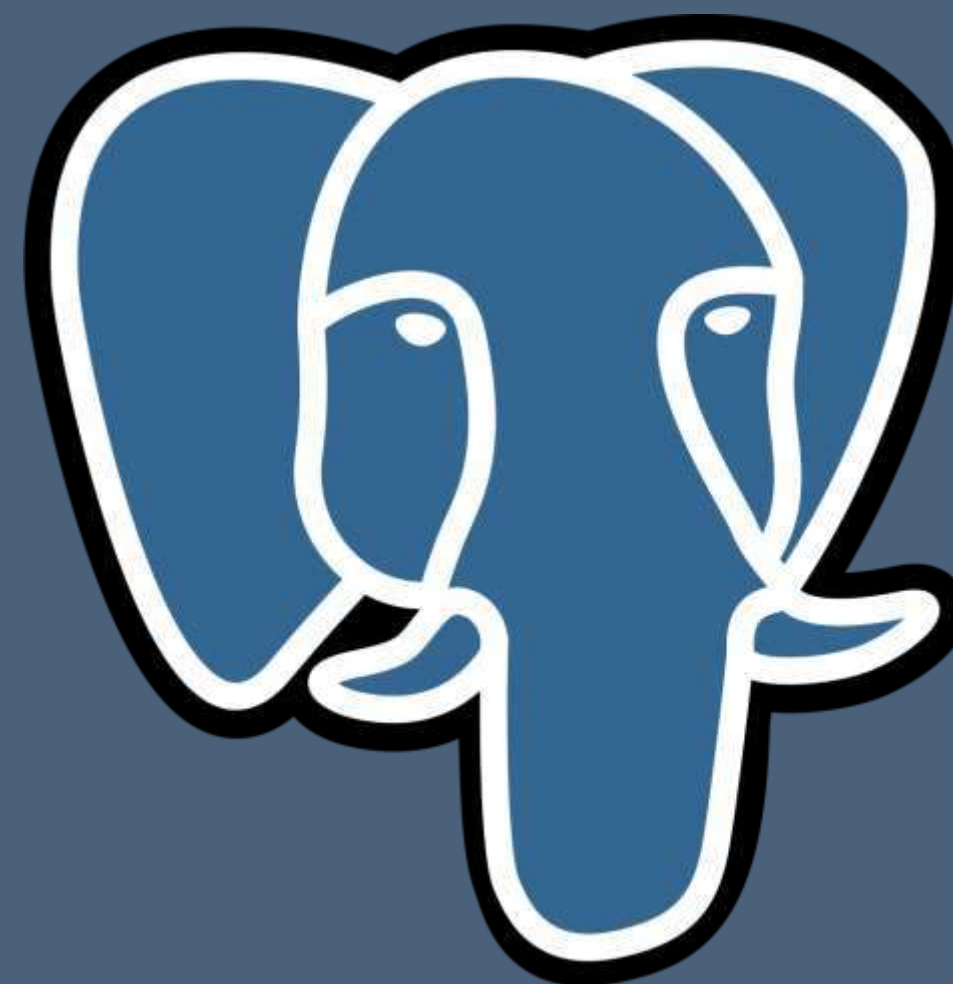
Актуальність дослідження

- Об'єкт дослідження – розподілені обчислювальні процеси у кластерному середовищі
- Предмет дослідження – методи та засоби автоматизованого планування і запуску задач у системі оркестрації

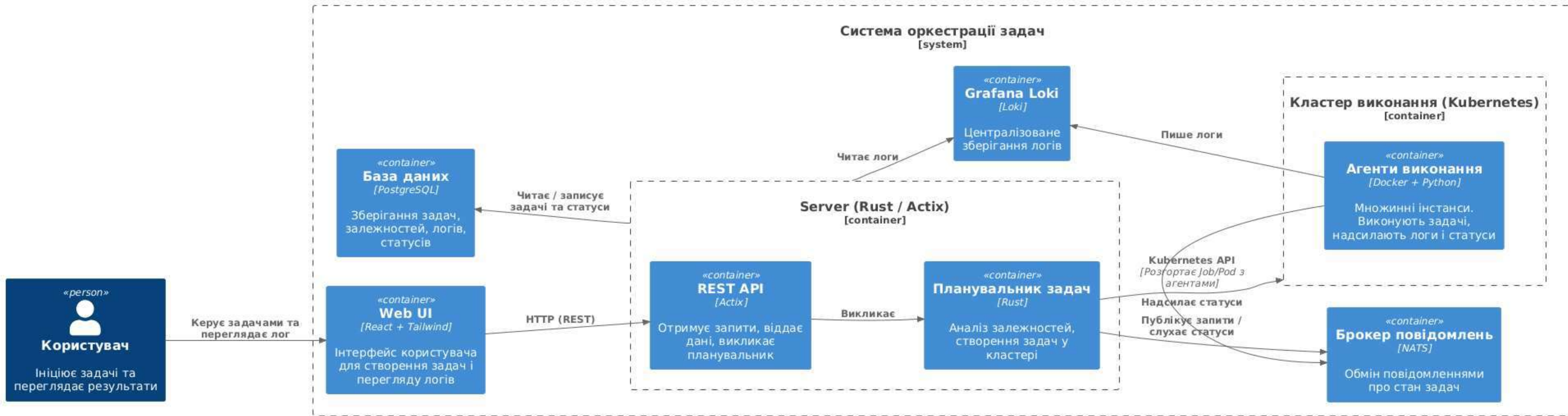
Мотивація для створення власного рішення

- Відсутність рішень, що поєднують гнучкість конфігурації задач із доступністю для користувача
- Необхідність попередньої підготовки контейнерних образів або налаштування CI/CD процесів користувачем
- Надмірна складність циклу розробки та розгортання прикладного ПЗ у кластерному середовищі

Використані технології



Архітектура системи



Створення програм

- Інтегрований редактор коду
- Система контролю версій через створення ревізій
- Кожна програма – Python скрипт, що має доступ до бібліотеки сервісу
- Компоненти взаємодії з оркестратором ініціалізуються агентом виконання без втручання користувача

Programs

 New

Example Program

Cool Program

New Program

Sort

Calculate Stats

Present stats

Calculate Stats

 Details Code

```
1 import flows
2 import json
3 import os
4
5 flows.logger.info("Script 2 (Calculate Statistics) started")
6
7 try:
8     # Load the sorted list (assuming JSON list of numbers)
9     intermediate_data_bytes = flows.load_input()
10    intermediate_data_str = intermediate_data_bytes.decode('utf-8')
11    flows.logger.info(f"Script 2 loaded raw intermediate input: {intermediate_data_str}")
12
13    # Parse the JSON
14    sorted_list = json.loads(intermediate_data_str)
15    if not isinstance(sorted_list, list) or not all(isinstance(n, (int, float)) for n in sorted_list):
16        flows.logger.error("Intermediate input is not a list of numbers.")
17        raise Exception("Invalid intermediate input format")
18
19    # Calculate sum and average
20    total_sum = sum(sorted_list)
21    average = total_sum / len(sorted_list) if sorted_list else 0
22    flows.logger.info(f"Script 2 calculated sum: {total_sum}, average: {average}")
23
```

Flows

+ New

Search flows...

Sample Flow

JSON Pipeline

Available Programs

Example Program

Cool Program

New Program

Sort

Calculate Stats

Present stats

JSON Pipeline

 Save Changes

⋮ 1. Sort



⋮ 2. Calculate Stats

⋮ 3. Present stats

ЖИТТЄВИЙ ЦИКЛ ЗАДАЧІ

1. Підготовка середовища виконання
2. Запуск клієнтської програми
3. Моніторинг виконання, збереження результатів або повідомлення про помилку
4. Оновлення статусу задачі
5. Вивільнення ресурсів

Агент виконання

- Універсальний Docker образ
- Встановлює користувацькі залежності через менеджер пакетів
- Імпортує та виконує клієнтський модуль коду
- Сповіщає систему на всіх етапах життєвого циклу програми

Flow Run Details



Details for run ID: 547ffeb7-978d-47e9-9207-477ed446b646

Flow:

JSON Pipeline

Timing:

Started: May 27th, 2025 1:57 PM

Completed: May 27th, 2025 1:57 PM

Status:

Completed

Program Execution Status:



Program Logs:

```
13:57:41.571 [STDERR]: INFO:flows:Status sent successfully.
13:57:41.571 [STDERR]: INFO:flows:Executing third-party script...
13:57:41.572 [STDERR]: INFO:flows:Script 2 (Calculate Statistics) started
13:57:42.051 [STDERR]: INFO:flows:Script 2 loaded raw intermediate input: [3, 5, 7, 10, 12, 18, 20, 25, 33, 45]
13:57:42.051 [STDERR]: INFO:flows:Script 2 calculated sum: 178, average: 17.8
13:57:42.393 [STDERR]: INFO:flows:Saved result to S3 bucket flows-diploma with key 547ffeb7-978d-47e9-9207-477ed446b646/b9661919-9aab-4c14-972c-5862ab9cf7a5/number_statistics.json
13:57:42.393 [STDERR]: INFO:flows:Script 2 (Calculate Statistics) finished
13:57:42.394 [STDERR]: INFO:flows:Script executed successfully.
13:57:42.394 [STDERR]: INFO:flows:Creating status object with status='completed' and error='None'
13:57:42.394 [STDERR]: INFO:flows:Sending status update: flow_exec_id='547ffeb7-978d-47e9-9207-477ed446b646'
program_exec_id='b9661919-9aab-4c14-972c-5862ab9cf7a5' status='completed' result_ref='547ffeb7-978d-47e9-
```

Get Results

Висновки

- Створено прототип системи автоматизованого керування обчислювальними задачами
- Розроблене рішення не потребує значних інфраструктурних ресурсів для розгортання та експлуатації
- Система ефективно забезпечує прототипування пайплайнів та виконання ресурсоємних обчислень