

МАТЕМАТИЧНЕ ТА ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ ДЛЯ ОБРОБКИ ПОТОКІВ ДАНИХ У ХМАРНИХ ОБЧИСЛЮВАЛЬНИХ СЕРЕДОВИЩАХ

Козловський М.М.¹⁾, Фатюк В.В.²⁾, Сафандула Н.Ю.³⁾, Романюк М.В.⁴⁾, Мацьків В.І.⁵⁾

Західноукраїнський національний університет

^{1)магістрант;} ^{2)аспірант;} ^{3)аспірант;} ^{4)аспірант;} ^{5)аспірант;}

І. Постановка проблеми

У роботі розглядаються поняття моделі хмарних обчислень та обробки потоків даних у таких обчислювальних середовищах у контексті систем індустріального інтернету речей та цифрових двійників. Обговорюються ключові підходи до організації архітектури програмних систем та методи обробки потоків даних з урахуванням та без урахування стану, інструментів потокової обробки даних, а також особливості застосування платформ управління потоками робіт для вирішення таких завдань. Особливу увагу приділено огляду методів декомпозиції потоків робіт і підходів до проектування програмних систем, орієнтованих на обробку потоків даних у хмарних середовищах.

ІІ. Мета роботи

Метою дослідження є розробка математичного та програмного забезпечення для обробки потоків даних у хмарних обчислювальних середовищах.

ІІІ. Реалізація програмного забезпечення

У роботі був розроблений набір потоків робіт для системи обробки потоків даних у хмарних обчислювальних середовищах. Як типове завдання обробки потоків даних у реальному часі з використанням географічно-розподілених обчислювальних систем використовується реальний набір даних, наданий DEBS GrandChallenge [1,2]. В рамках даного набору представлений масив даних, зібраних з датчиків і клапанів, встановлених на промисловому обладнанні. Затримка між двома послідовними точками даних становить близько 10мс. Кожен елемент вихідних даних являє собою спілкування у форматі Apache Avro, що складається з 66 полів. У полях повідомлення міститься інформація про стан відповідних датчиків у момент зчитування інформації.

Процес обробки даних, необхідних для вирішення цього завдання складається із двох етапів. На першому етапі проводиться ідентифікація зміни стану у вхідних полях між послідовними точками вихідних даних та генерація повідомлень про зміну станів разом з часовими мітками даної події. На другому етапі вирішується завдання кореляції між зміною стану датчиків та зміною стану клапанів, а також розраховується тимчасова відстань між настанням зміни стану.

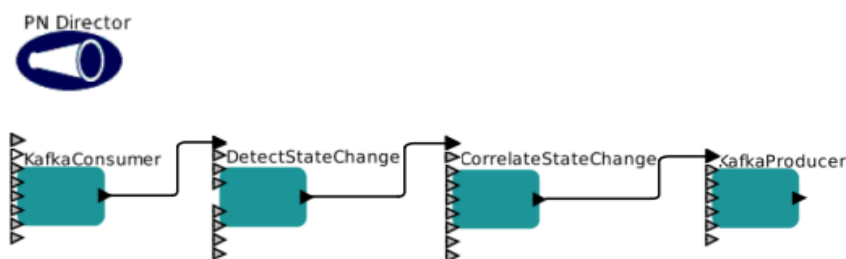


Рисунок 1—Ідентифікація зміни стану та кореляція зміни стану в одному потоці робіт

Перший потік робіт розроблений для включення всіх необхідних етапів обробки, де вихідний потік даних надходить до актора 1, а він забезпечує десеріалізацію вхідних повідомлень і передає їх у вигляді записів токена в актор 2. Цей актор забезпечує

виявлення змін у вхідному потоці даних. Кожна виявлена зміна надсилається як запис до актора 3. Цей актор виконує кореляційний аналіз парсерій даних при отриманні кожного запису. Результати роботи актора 3 є набір потоків даних, що містять інформацію про кореляцію між станами датчиків та клапанів.

Висновок

Для тестування підходу мікро-потоків робіт, на основі розроблених акторів, були спроектовані та реалізовані потоки робіт, що забезпечують обробку даних у потоковому режимі.

Список використаних джерел

1. Erl, Thomas, Ricardo Puttini, and Zaigham Mahmood. Cloud Computing: Concepts, Technology&Architecture, 2013
2. Bahga, Arshdeep, and Vijay Madisetti. Cloud Computing: A Hands-OnApproach, 2014