Task 2 Advanced - Створення Real-Time Stream Processing Pipeline with Hazelcast Jet
(15 балів)
Автор: Потужний Богдан

https://docs.hazelcast.com/hazelcast/5.3/pipelines/overview

Зустрічаються задачі, коли необхідно обробляти та реагувати на потік вхідних даних чи подій в режимі близькому до реального часу.
Для цього існують різні платформи класу Real-Time Stream Processing Pipeline, які дозволяють визначати джерело вхідних даних, кроки по обробки кожної з подій та місце куди оброблені події перенаправляються чи зберігаються.

- One or more sources: Where you take your data comes from
- Processing stages: What you do to your data
- At least one sink: Where you send the results of the last processing stage

Hazelcast реалізує такий функціонал за допомогою Hazelcast Jet (https://docs.hazelcast.com/hazelcast/5.3/pipelines/overview), який повністю інтегрується з усіма іншими структурами даних які є на даній платфомі

Завдання

Розберіть і запустіть приклад (на основі якого далі будете виконувати завдання) https://docs.hazelcast.com/hazelcast/5.3/pipelines/stream-processing-embedded

У якості завдання пропонується на базі Hazelcast Jet побудувати невкликий Data pipeline для Real-Time обробки нових записів які з'являються у логах Веб-сервера.

Для цього необхідно запустити будь-який Веб-сервер з довільним Веб-застосуванням, який би робив стандарте логування запитів.

Далі, використовуючи `Sources.fileWatcher("/home/data")` https://docs.hazelcast.com/hazelcast/5.3/integrate/legacy-file-connector#file-watcher необхідно підписатись на зміни які будуть відбуватись у файлі (додавання нового запису).

Отримавши послідовність записів з логу, необхідно написати наступну логіку на основі https://docs.hazelcast.com/hazelcast/5.3/pipelines/transforms:
.

1. Для кожного ресурсу (сторінки) зберігати у Hazelcast IMap - `*RequestsCountMap*` кількість запитів до цієї сторінки, де відповідь була 200 (ключ

- URL-адреса сторінки, значення - каунтер, скільки раз заертались). Див https://docs.hazelcast.com/hazelcast/5.3/integrate/map-connector#map-as-a-sink
2. Використовуючи операцію *window* (https://docs.hazelcast.com/hazelcast/5.3/pipelines/transforms#window) виводити на на екран, скільки успішних запитів було оброблено за останні 30 сек.

У протоколі має бути вміст *RequestsCountMap,* вивід логу, а також скріншот DAG де представлений Data pipeline (можна також отримати його з Management center)

Для виконання даної роботи було створено програму, що генерує подібні логи, більш детально з функцією можна ознайомитися в коді GenerateLogs.kt

```
181.137.32.182 - - [23/Dec/2023:20:23:17 +0100] "GET /index.html HTTP/1.1"
500 448
111.89.34.183 - - [23/Dec/2023:20:23:18 +0100] "GET /contact.html HTTP/1.1"
500 523
2.219.223.42 - - [23/Dec/2023:20:23:19 +0100] "GET /contact.html HTTP/1.1"
403 537
42.239.217.143 - - [23/Dec/2023:20:23:20 +0100] "GET /index.html HTTP/1.1"
500 224
43.74.61.198 - - [23/Dec/2023:20:23:21 +0100] "GET /about.html HTTP/1.1" 200
992
161.15.33.50 - - [23/Dec/2023:20:23:22 +0100] "GET /index.html HTTP/1.1" 301
489
249.181.216.252 - - [23/Dec/2023:20:23:23 +0100] "GET /about.html HTTP/1.1"
301 491
103.190.144.160 - - [23/Dec/2023:20:23:24 +0100] "GET /products.html
HTTP/1.1" 500 492
```

В результаті створення пайплайнів було отримано наступні результати

RequestCountMap:

```
/about.html - 9
/products.html - 3
/index.html - 7
/api/data - 5
/contact.html - 2
```

Log:

Тут знаходиться частина логу що дозволяє оцінити правильність роботи, повний лог доступний у файлі hazelcast-log.txt

```
2023-12-23 20:25:57 INFO  PipelineLogger:52 - Updating URL - Old Value: 3,
New Value: 4
2023-12-23 20:25:58 INFO  PipelineLogger:41 - Procesing URL: /index.html,
Status: 200
2023-12-23 20:25:58 INFO  PipelineLogger:52 - Updating URL - Old Value: 4,
New Value: 5
2023-12-23 20:25:59 INFO  PipelineLogger:41 - Procesing URL: /contact.html,
Status: 200
2023-12-23 20:25:59 INFO  PipelineLogger:52 - Updating URL - Old Value: 1,
```

```
New Value: 2
2023-12-23 20:26:00 INFO  WriteLoggerP:65 - [192.168.56.1]:5701 [dev] [5.3.6]
[0afa-47a2-8680-0001/loggerSink#0] WindowResult{start=20:25:30.000,
end=20:26:00.000, value='9', isEarly=false}
2023-12-23 20:26:06 INFO  PipelineLogger:41 - Procesing URL: /products.html,
Status: 200
2023-12-23 20:26:06 INFO  PipelineLogger:52 - Updating URL - Old Value: 1,
New Value: 2
2023-12-23 20:26:10 INFO  WriteLoggerP:65 - [192.168.56.1]:5701 [dev] [5.3.6]
[0afa-47a2-8680-0001/loggerSink#0] WindowResult{start=20:25:40.000,
end=20:26:10.000, value='8', isEarly=false}
2023-12-23 20:26:12 INFO  PipelineLogger:41 - Procesing URL: /about.html,
Status: 200
2023-12-23 20:26:12 INFO  PipelineLogger:52 - Updating URL - Old Value: 7,
New Value: 8
2023-12-23 20:26:16 INFO  PipelineLogger:41 - Procesing URL: /index.html,
Status: 200
2023-12-23 20:26:16 INFO  PipelineLogger:52 - Updating URL - Old Value: 5,
New Value: 6
```

DAG:

Перший DAG описує pipeline з функцією window, другий описує збереження до IMap