



Trabajo Práctico N°2

Middleware y Coordinación de Procesos
Reddit Memes Analyzer

Alumna	Ailén Magalí García
Padrón	100560
Repositorio	https://github.com/ailugarcia98/reddit_memes_analyzer
Fecha de entrega	26/05/2022
Fecha de re-entrega	09/06/2022



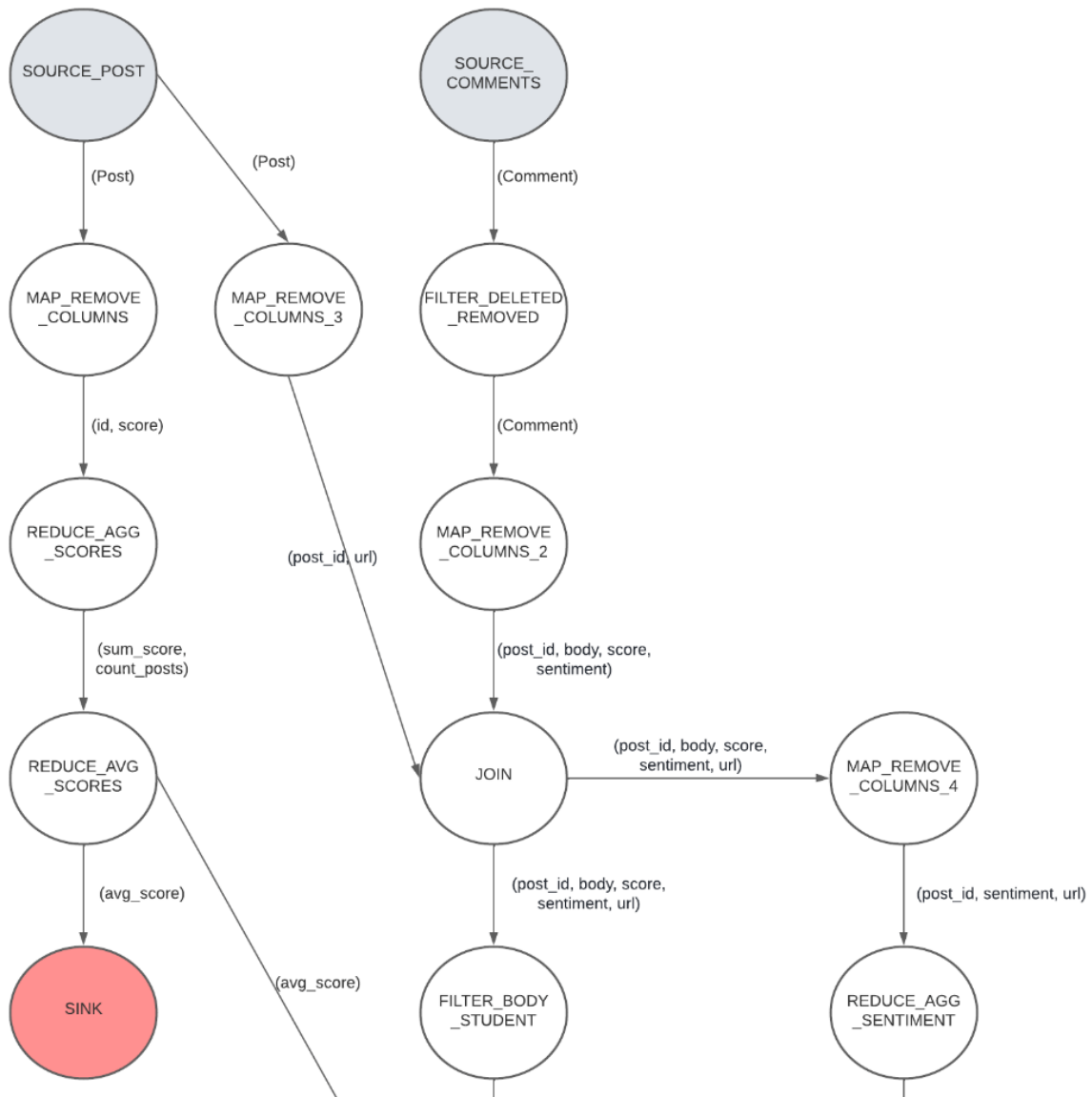
ÍNDICE

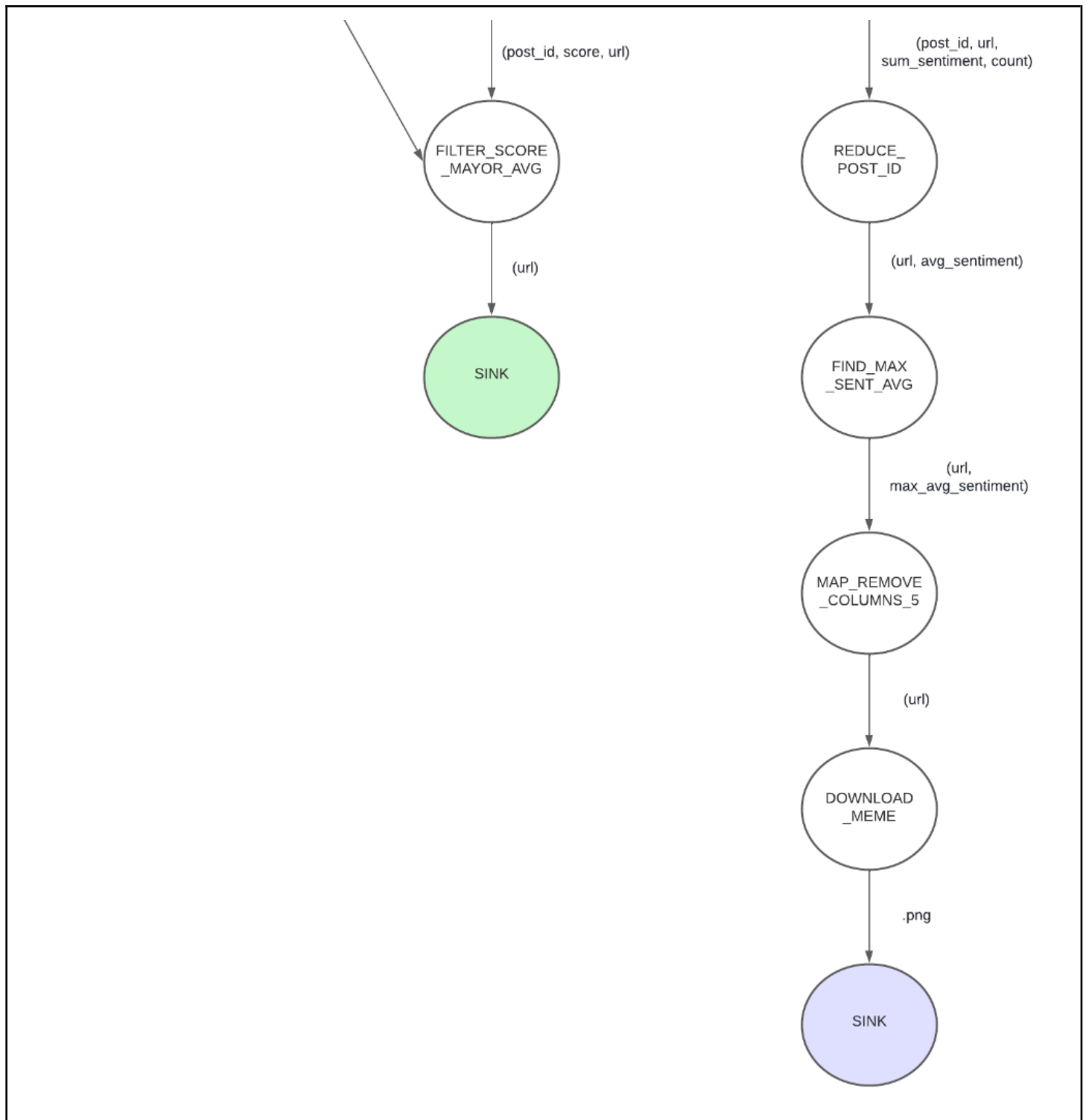
ÍNDICE	1
VISTA LÓGICA	2
DAG	2
VISTA FÍSICA	4
Diagrama de robustez	4
Diagrama de despliegue	6
VISTA DE PROCESOS	7
Diagramas de actividades	7
Cálculo del promedio de score de posts	8
Camino desde consumo de posts y comments hasta join	9
Camino desde join hasta que se envía al producir las urls de estudiantes con score mayor al promedio	10
VISTA DE DESARROLLO	11
Diagrama de paquetes	11



VISTA LÓGICA

DAG





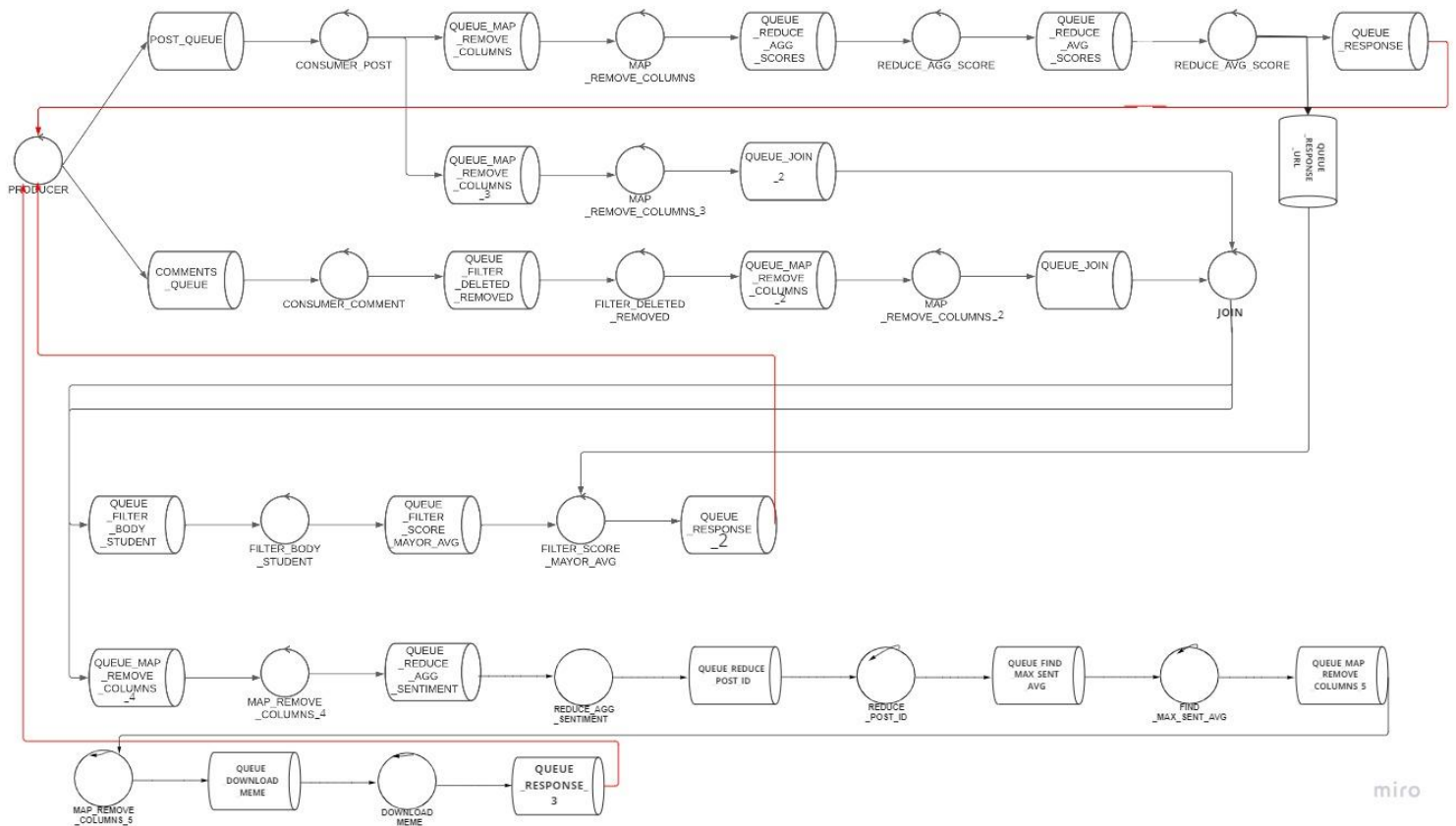
Se puede observar que hay 3 flujos para resolver las 3 queries:

El SINK **rojo** refiere al promedio de score de todos los posts, el **verde** a URLs de memes que gustan a estudiantes (con comments sobre university, college, student, teacher, professor y con score mayor al promedio) y el **lila** a la descarga del meme con mejor sentiment promedio.



VISTA FÍSICA

Diagrama de robustez



Este diagrama representa todos los flujos de la aplicación, en donde procesos se mandan mensajes a través de queues de RabbitMQ. A continuación, se describe el comportamiento de cada proceso:

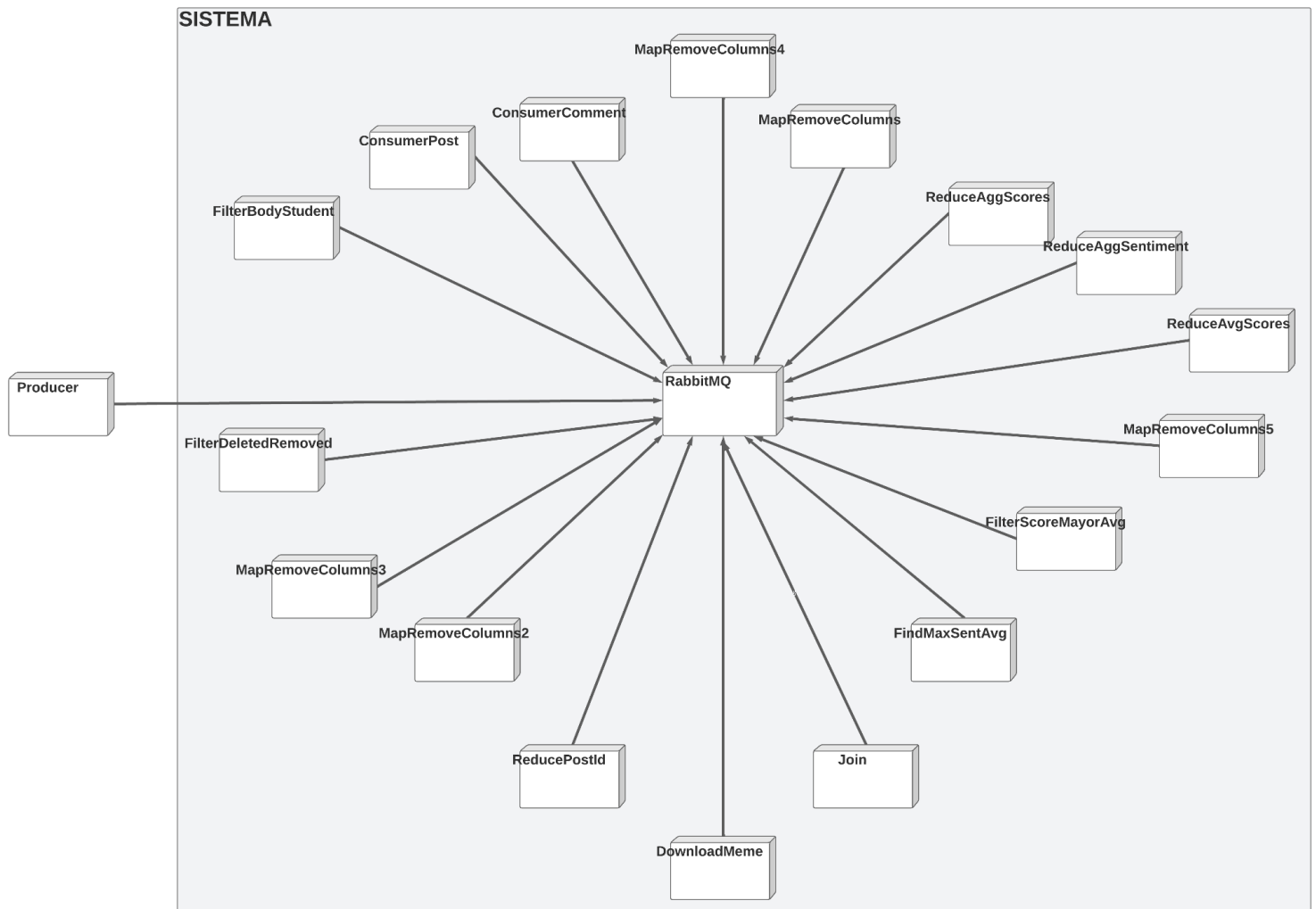
- **Producer:** Es el encargado de leer los archivos de comentarios y posts, mandar las líneas leídas a los consumers y recibir las respuestas de las 3 queries.
- **Consumer (post y comments):** Es el que recibe lo leído por el productor y se encarga de encolarlo a donde corresponda.
- **Map Remove Columns:** Cada proceso de este nombre, realiza una selección de columnas a enviar.
- **Reduce Agg Score:** Proceso que interviene en el cálculo del promedio de score de todos los posts. Se encarga de calcular la suma de los scores y la cantidad total para mandársela al proceso encargado de calcular el promedio.
- **Reduce Avg Score:** Es el que calcula el promedio con la información que le brinda el proceso anterior.
- **Filter Deleted Removed:** Se encarga de filtrar aquellos comentarios cuyo body [DELETED] o [REMOVED].
- **Join:** Es el encargado de buscar en los posts la url correspondiente a cada comment a través de la key `post_id`.
- **Filter Body Student:** Es el encargado de seleccionar aquellos registros cuyo body contenga palabras o frases relacionadas con estudiantes.
- **Filter Score Mayor Avg:** Recibiendo el promedio de score de los posts, filtra aquellas urls cuyo score sea mayor al promedio, y se las manda al productor.
- **Reduce Agg Sentiment:** Es el encargado de calcular la cantidad y suma de los sentiments, para que Reduce Post ID calcule el promedio de sentiment.
- **Reduce Post ID:** Calcula el promedio de sentiment por url (meme).



- Find Max Sent Avg: Dentro de todos los sentiments promedio, encuentra la url que posee el mayor valor del mismo.
- Download Meme: Es el encargado de descargar el meme recibiendo la url del proceso anterior, y mandárselo al producer.



Diagrama de despliegue



Este diagrama busca mostrar los nodos que constituyen el sistema, y sus dependencias. Podemos observar que todos dependen de RabbitMQ, debido a que es el nodo que permite la comunicación del resto.



VISTA DE PROCESOS

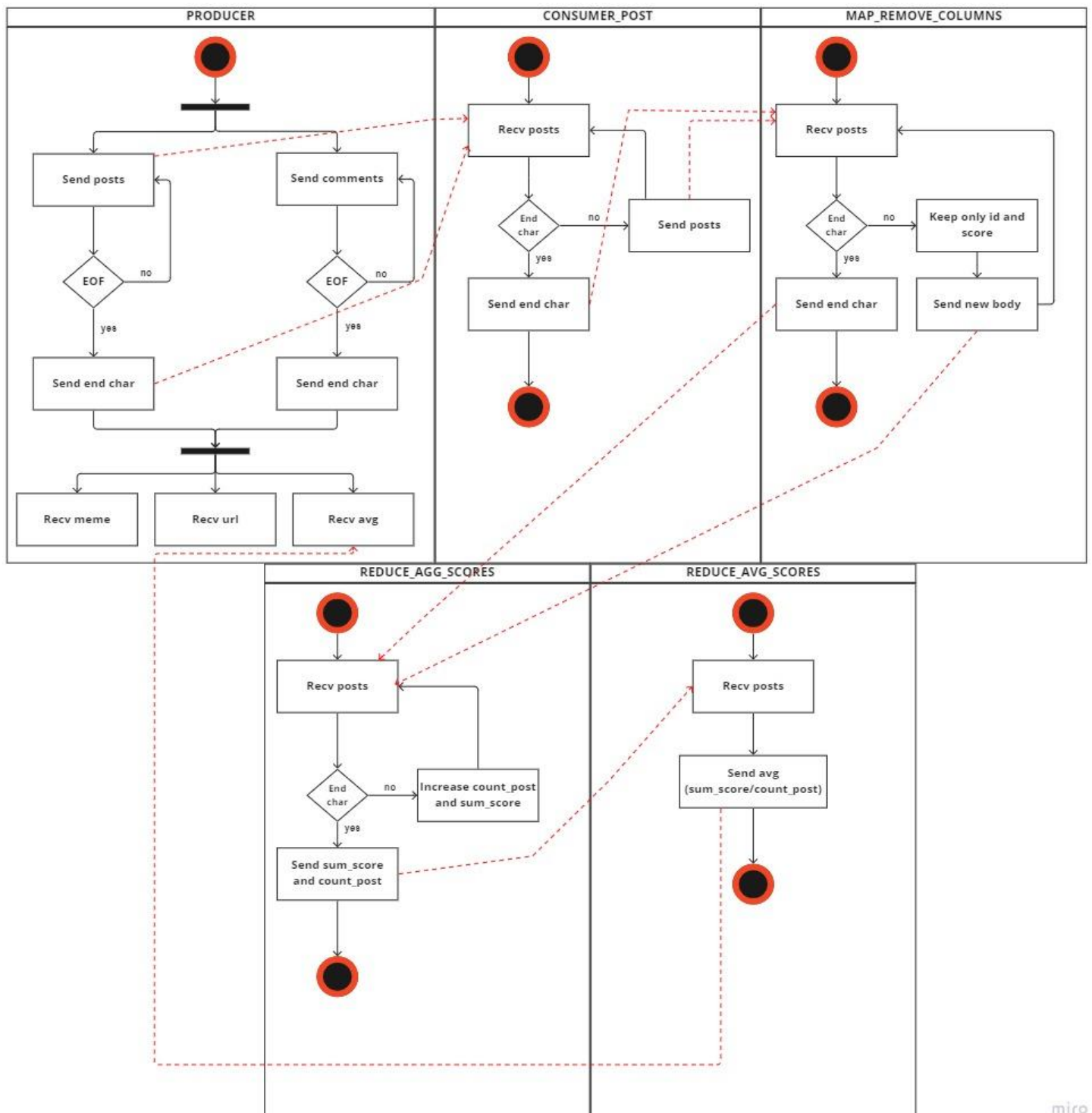
Diagramas de actividades

A continuación, se pueden observar distintos diagramas de actividades. Se ha decidido dividirlos en varios, por un tema de espacio y claridad.

Los mismos corresponden a los flujos de cálculo del promedio de score de todos los posts, y de obtención de las urls de memes que gustan a estudiantes.



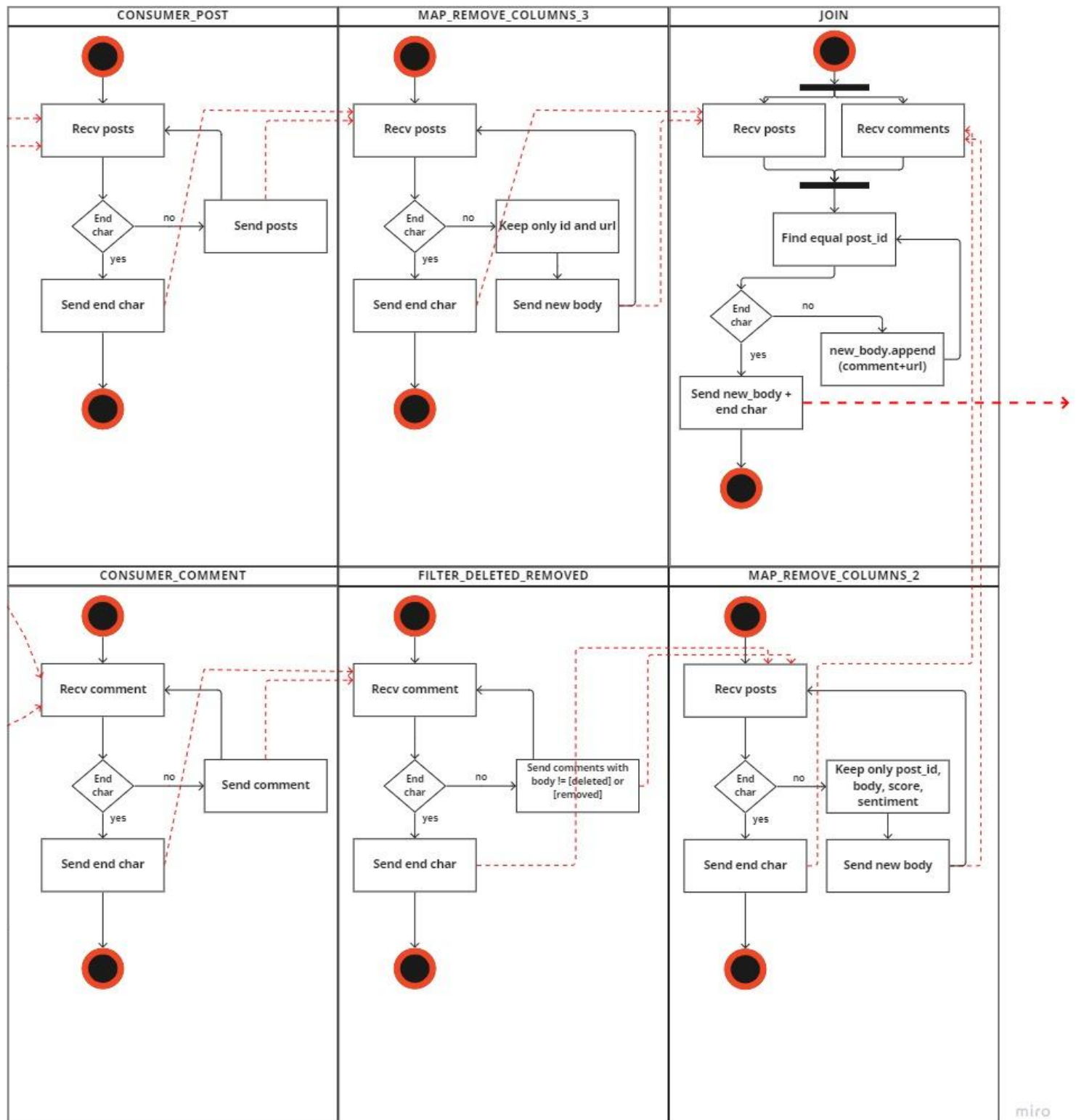
Cálculo del promedio de score de posts



miro



Camino desde consumo de posts y comments hasta join

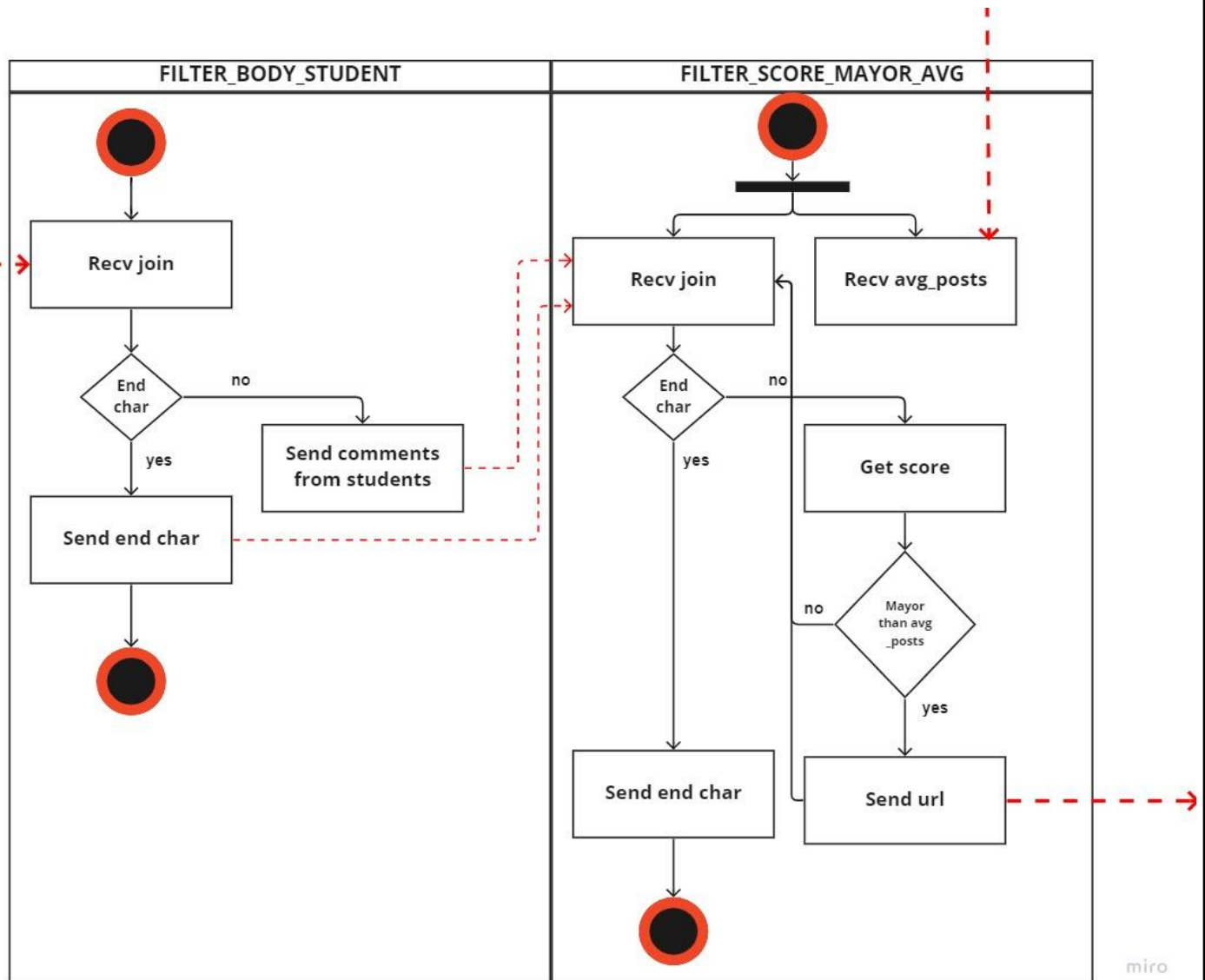


miro

- Las flechas ingresando a Recv posts y Recv comments de los Consumers vienen de Producer.
- La fecha que sale del Join es la entrada al Recv join del diagrama siguiente.



Camino desde join hasta que se envía al producir las urls de estudiantes con score mayor al promedio



Filter Score Mayor Avg recibe el promedio de score de todos los post que genera el proceso Reduce Avg Scores. A su vez, Send url lo hace hacia el Producer.



VISTA DE DESARROLLO

Diagrama de paquetes

