Apache Flink

Jesús Rodríguez Heras Roberto Muras González Juan Pedro Rodríguez Gracia Gabriel Fernando Sánchez Reina

30 de mayo de 2018

Index

- Definiciones
 - Apache Flink
 - RabbitMQ
- 2 Instalación
 - Java
 - Apache Flink
 - RabbitMQ
- Palabras más publicadas en twitter
- 4 Problemas encontrados
- Mejoras futuras
- 6 Referencias

¿Qué es Apache Flink?

Apache Flink

Se trata de un motor de procesamiento de streams o flujos de datos de código abierto que proporciona capacidades de distribución de datos, comunicaciones y, muy importante, tolerancia a fallos a las computaciones.

¿Qué es Rabbit MQ?

RabbitMQ

Es un software de negociación de mensajes de código abierto. Se encuentra dentro de la categoría de middleware de mensajería. Implementa el estándar "Advanced Message Queuing Protocol" (AMQP).

Java

Instalación de java

Para la instalación de java, solo hay que descargarse el jdk de la página oficial e instalarlo en /usr/java/.

Flink

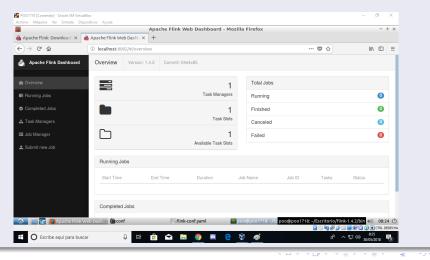
Instalación de Apache Flink

Para su instalación, nos descargamos el software desde su página web, lo descomprimimos donde queramos y ejecutamos el fichero start-cluster.sh.

Flink

Comprobamos que funciona

Para ello entramos en localchost:8082.



RabbitMQ

Instalación de RabbitMQ

Para su instalación, introducimos el siguiente comando en la terminal:

sudo apte-get install rabbitmq-server

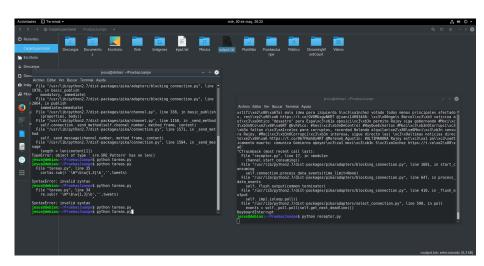
Luego, instalamos la consola de administración con este comando:

sudo rabbitmq-plugins enable rabbitmq_management

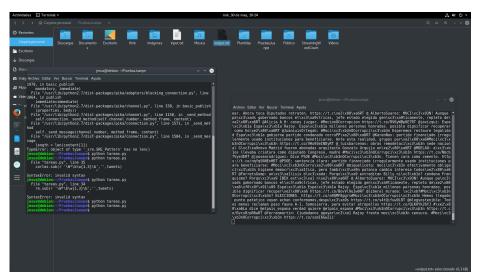
Funcionamiento

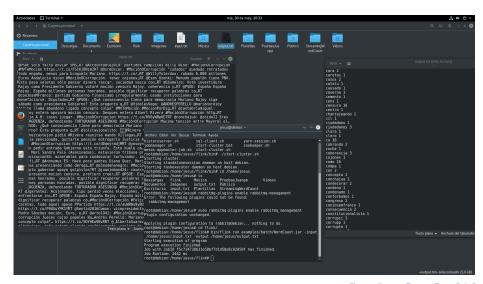
Nuestro trabajo consta de dos archivos Python:

- tareas.py.
- receptor.py.
- Ejecución del servicio de Apache Flink: bin/flink run examples/batch/WordCount.jar -input /home/jesus/input.txt -output /home/jesus/output.txt



10 / 15





Problemas encontrados

Problemas

- No teníamos donde ejecutar Apache Flink.
- Apache Flink lanzaba multitud de excepciones.

Mejoras futuras

Mejoras

- Mejorar la implementación de las palabras cortas (determinantes y preposiciones).
- Utilizar "jpype" para llamar a las clases Java desde Python, lo que nos ahorraría bastante trabajo.

Referencias

- https://flink.apache.org/
- https://www.rabbitmq.com/
- https://en.wikipedia.org/wiki/Apache_Flink
- https://en.wikipedia.org/wiki/RabbitMQ
- https://www.adictosaltrabajo.com/tutoriales/ introduccion-a-apache-flink/
- https://data-flair.training/blogs/install-configureapache-flink-ubuntu/