

# Ejercicio evaluable de Sistemas Distribuidos

Grado en Ingeniería Informática

Curso: 2016/2017

## Sobre este documento

Este documento detalla el enunciado del ejercicio evaluable de la asignatura de Sistemas Distribuidos del Grado en Ingeniería Informática. El objetivo de este ejercicio es evaluar los temas de teoría de la asignatura, a partir del tema 5. La evaluación de este ejercicio supondrá el 60% de la nota de teoría (siendo el otro 40% el resultado de los dos cuestionarios calificables realizados en clase).

## Asignación de Grupos

El ejercicio está planteado para realizarse en grupos de 3 personas. De todas formas, se podrá realizar si se desea en parejas, o incluso de forma individual. Se valorará la carga de trabajo por persona, de tal manera que un trabajo con menos personas, a igual funcionalidad, será valorado más positivamente. Si se desea, se podrá consultar al profesorado la realización en un grupo de cuatro alumnos, y se indicará el grado de complejidad que dicho ejercicio deberá tener para realizarse en un grupo de cuatro.

Los miembros del trabajo deberá de indicarse en la Wiki correspondiente del campus virtual así como en el código y en el documento de presentación.

## Tecnologías disponibles

Las tecnologías disponibles serán las vistas en clase. **NO** es necesario aplicarlas todas, o bien porque las tecnologías son (a efectos de este ejercicio) equivalentes (como almacenar en Dropbox o Google Drive), o porque no son necesarias (no usar Celery porque ZeroMQ ofrece ya tratamiento distribuido, o de entrada usar Twitter y no scrapy o al revés, usar Scrapy y no Twitter).

## Temática

La temática es totalmente libre. Cada grupo podrá realizarla de forma que considere, siempre que tenga en cuenta los objetivos mínimos pedidos en la sección siguiente.

## Requisitos Mínimos

Dada la temática de la asignatura, los alumnos deberán realizar un ejercicio que muestre que se han entendido los conceptos vistos.

- El sistema debe contener de varios procesos, que se comunicarán o bien mediante la librería **ZeroMQ** o mediante un sistema de ejecución asíncrono (usando **Celery** para ello).
- Deberá de leer información en algún momento usando **Twitter** o bien haciendo *web scraping*.
- Almacenar datos intermedios en la nube (como **Dropbox** o **Google Drive**), que serán utilizados como entrada de otro proceso.
- Ofrecer una salida final de datos (estadísticos, gráficas, ...) que será o bien visualizados usando un interfaz web, o almacenados en otro fichero **Dropbox/Google drive**.

### Entrada de datos del Sistema

- La fuente de entrada de datos del sistema será:
  - Cuenta(s) de Twitter, usando la librería que vimos.
  - Web scraping, usando la librería **scrapy** para ello, de una información públicamente disponible.
- La entrada de datos se ejecutará o bien de forma periódica (con **Celery** o **crontab**) o bien usando *Streaming* desde **Twitter**.

Adicionalmente se podría usar un interfaz web (usando **Bottle**) ya sea para iniciar el proceso con ciertos parámetros, o bien para mostrar la información generada (por ejemplo, guardando alguna gráfica en un fichero, que el interfaz web visualice).

### Almacenamiento en la nube

Para compartir datos entre los distintos procesos, éstos deberán estar en la nube. Se deberá de utilizar o bien Dropbox o bien Google Drive, tal y como se ha visto en teoría.

Al menos un proceso deberá de almacenar en la nube información que otro proceso diferente procese.

### Escalabilidad

El sistema deberá de ser escable, en el sentido de que se puedan aumentar los recursos para escalar el sistema. Esto se puede realizar según la tecnología usada:.

**Con ZeroMQ** mediante un mecanismo de comunicación que permita aumentar el número de procesos esclavos.

**Con Celery** aumentando el número de procesos *workers*.

### Peticiones asíncronas

Se deberá de realizar de forma asíncrona (usando Celery o con ZeroMQ), al menos alguna las siguientes tareas:

- Subir los datos a la nube (sea Dropbox o Google Drive).
- Procesar la respuesta a una entrada de Twitter (ya sea usando un polling o mediante Streaming).

- La realización de *web scraping* con Scrapy ya es asíncrona, así que sólo tendría sentido realizar de forma asíncrona su procesamiento, pero no es necesario.

La idea con las peticiones asíncronas es que se puedan ir atendiendo peticiones sin tener que esperar a que las anteriores terminen totalmente.

Se valorará que se realicen todas las tareas anteriores de forma asíncrona, pero para aprobar es necesario realizar la tarea asíncrona al menos en un punto, para demostrar que se conoce dicha técnica.

## Rúbrica de Evaluación

Se recuerda que la nota final de Sistemas Distribuidos = 45% Nota Teoría + 45% Nota Prácticas + 10% Cuestionarios.

La evaluación de teoría será: Teoría = 40%· Nota Cuestionario + 60%· Ejercicio.

- La no entrega en la fecha prevista o no cumplir los objetivos mínimos supondrá un 0 en el ejercicio y, por tanto, la necesidad de ir al examen final para aprobar.
- Para aprobar deberá la aplicación de cumplir todos los objetivos mínimos indicados en la sección anterior.
- Una vez cumplidos los objetivos de la sección anterior, se valorará según el siguiente modelo.

Concepto	Aprobado (5-6)	Notable (7-8)	Sobresaliente (9-10)	Porcentaje
Entrada de datos	Control errores	No control errores o no asíncrono	Gestiona bien errores y asíncrono	30%
Almacenamiento	No controla errores	Algún control o no asíncrono	Control de errores y asíncrono	10%
Peticiones asíncronas	Alguna, para aprobar	Varios, con algún error	Varias y sin errores	10%
Procesamiento salida	Apenas procesamiento	Procesamiento algo	Procesamiento interesante	10%
Dificultad	Demasiado fácil	Complejidad algo baja	Suficientemente complejo	15%
Escalabilidad	Escala muy básica o con errores	Escala con limitaciones (no en lo importante)	Escala bien	15%
Presentación	Falta funcionalidad dudas en presentación	Bien la funcionalidad mejorable la presentación	Bien ambos conceptos	10%

## Proceso de evaluación y Presentación

- Se realizará una presentación de hasta 10 minutos del trabajo con los miembros del grupo.
- Se deberá de poder mostrar la funcionalidad del sistema en dicha presentación.
- Habrá un turno de preguntas, de otros compañeros o del profesor, que deberá de responderse algún miembro del grupo.
- Los compañeros evaluarán los trabajos tras la presentación, pero el profesorado podrá modificar dicha evaluación según su propio criterio, predominando la calificación del profesorado.
- En principio la calificación será la misma a todos los miembros del grupo.