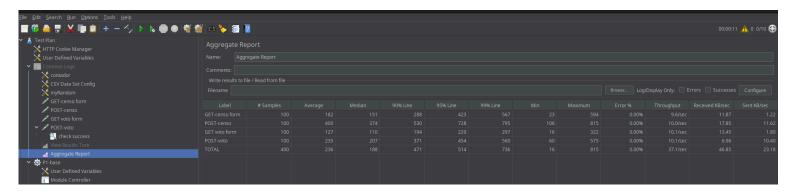
el servidor Django o que la gestión de sesiones no es relevante para el funcionamiento de la aplicación. Memoria de Prácticas Práctica 2

Sistemas Informáticos II

Autores: Ignacio González Porras y Daniel Wentworth Fernández

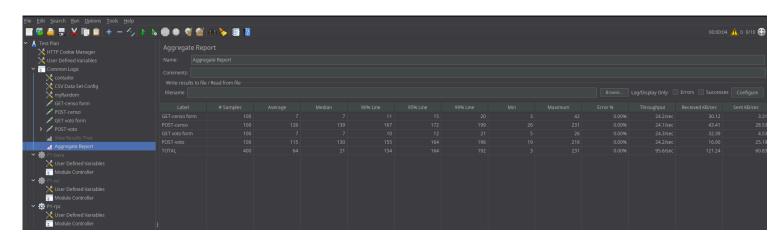
Ejercicio 1.



El fichero P2_P1-base.jmx se ha adjuntado con el resto de "tests plans" y código.

Ejercicio 2.

Test plan de P1-rpc:



Test plan de P1-ws:



Cuestión 1.

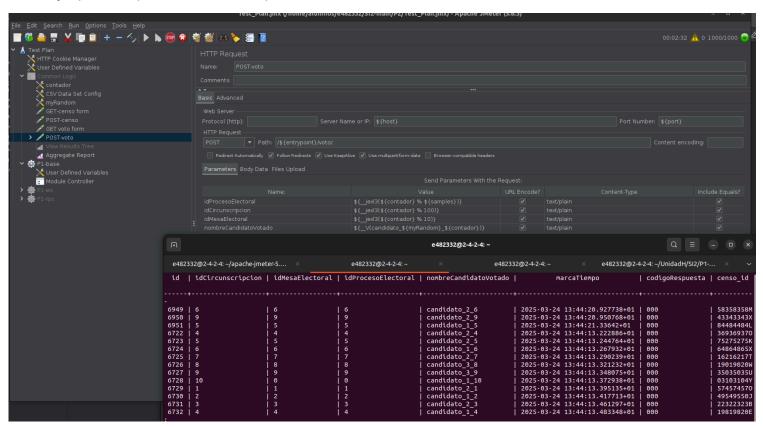
Las diferencias principales entre los dos "Thread Groups" son los entrypoints que se utilizan para P1-rpc y P1-ws (que son diferentes al de P1-base) y el puerto usado, que se cambia al 38000, a diferencia del puerto 28000 que se utiliza en P1-base.

Ejercicio 3.

Según lo que podemos observar en Jmeter al ejecutar el plan de pruebas, Jmeter no genera todas las peticiones a al vez visualmente, pero vemos que si ha sido capaz de crearlas, simplemente el procesamiento de todas es costoso, y no es inmediato, luego van apareciendo las solicitudes progresivamente en la GUI de Jmeter.

Ejercicio 4.

A la hora de tratarlo dentro del bucle, sucede lo mismo. Las solicitudes se pueden crear sin problema, incluso antes de que sean atendidas la anterior iteración, respaldando nuestra respuesta en el anterior apartado de la independencia de las capacidades del sistema para realizar ambas tareas. Esto lo vemos ya que en la salida en Jmeter, vemos completamente intercaladas los ids de las solicitudes lo que nos lleva a pensar que hay disparidad entre cuando se crean y cuando se procesan, ya que el id puede haber sido creado después que otro y aparecer procesado antes que un id menor.



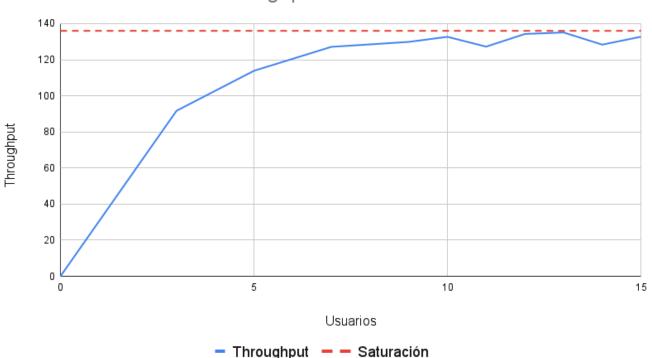
Ejercicio 5.

Para llevar a cabo estas medidas, aplicamos la variable \${samples} como el número de iteraciones para cada uno de los datapoints relativo al número de usuarios. Tomamos medidas similares a las vistas en el gráfico dado en el enunciado, tomando medidas distanciadas de dos en dos cuando el crecimiento era lineal, y luego de uno en uno para que la curva de saturación que se pretende mostrar sea lo más visible posible. También indicamos con una línea punteada dicho umbral de saturación, a modo visual y de manera aproximada que se tomó basándose exclusivamente en los datos experimentales.

En cuanto al rendimiento de cada uno, según los fuimos ejecutando en el orden propuesto, fuimos viendo un aumento en el throughput de salida. Respectivamente nuestras implementaciones de P1-base, P1-WS y P1-RPC, llegaron a una saturación máxima de en torno a 135, 160 y 190, siendo esta última, con RPC, la que dio mayor throughput. Esto se debe a que usa una comunicación binaria eficiente y conexiones persistentes, reduciendo la sobrecarga de serialización y latencia de red.

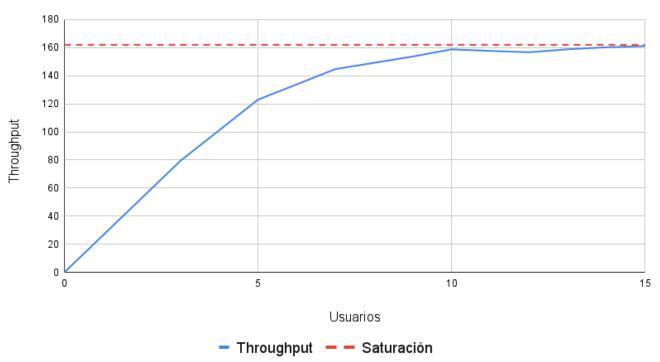
Throughput P1-base:





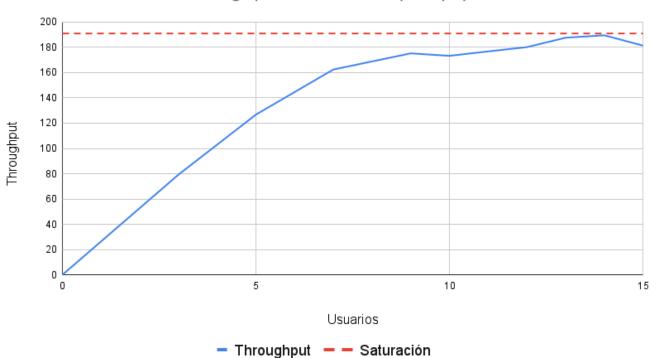
Throughput P1-ws:





Throughput P1-rpc:

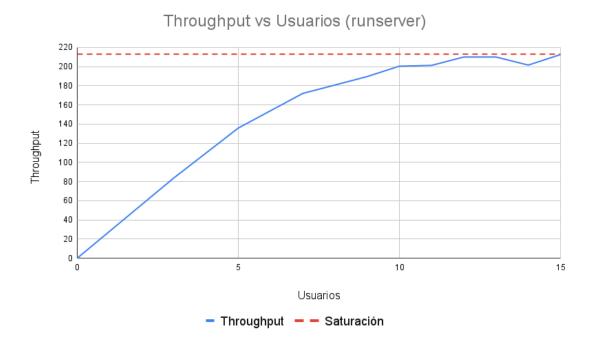
Throughput vs Usuarios (P1-rpc)



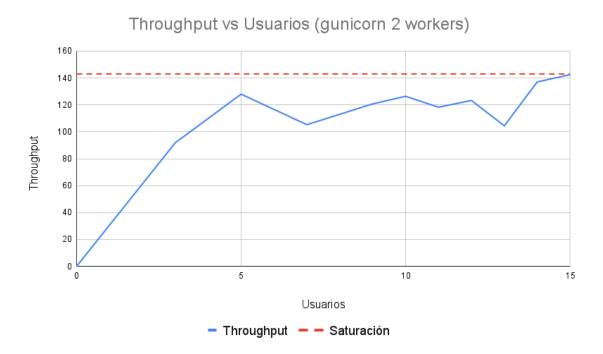
Ejercicio 6.

A la hora de comparar gunicorn, con un worker, usando P1-base como prueba común para ambos, con la ejecución del servidor de django usando runserver, obtuvimos que el P1-base como ya comentamos alcanza su saturación hacía casi un throughput de 140. Con runserver, se obtiene una medición significativamente mayor, resultando en un throughput de más de 210. El throughput de runserver es mayor porque es un servidor síncrono y ligero, sin la sobrecarga de gestión de workers. Gunicorn, aunque más robusto para producción, introduce overhead al manejar múltiples procesos concurrentes.

Throughput de runserver:



Throughput de gunicorn con un worker:



Ejercicio 7.

Cuando realizamos la medida con 2 workers de gunicorn, tras haber rebooteado la máquina virtual asignándole otro core, su rendimiento aumenta sin lugar a duda. Sin embargo, nosotros esperábamos ver un aumento del 100% al ser el doble de workers ejecutándose, pero la salida real fue un throughput que se saturó hacía 145, tan solo marginalmente más que con un worker. Uno podría pensar que es la misma ejecución pero con variación experimental en los datos, ya que la diferencia no es muy significativa. Esto concluimos que se debe a que gunicorn usa un modelo pre-fork, y cada worker maneja solicitudes de forma independiente. Si el servidor no está altamente concurrido, el segundo worker no aporta mucha mejora, ya que el cuello de botella suele estar en la CPU, I/O o la base de datos.

Cuestión 2.

(d) en ninguno, ya que la gestión de sesiones en este proyecto no ha sido relevante para el funcionamiento de la aplicación