# G2-8 Design & Testing II — Performance Tests

# Contenido

G2-8 Design & Testing II – Performance Tests	1
Definición de Buen Rendimiento:	2
HU-01: Agregar Paciente	2
HU-02: Borrar Paciente	5
HU-03: Modificar Paciente	8
HU-04: Listar Pacientes	10
HU-05: Mostrar Datos de Paciente	13
HU-06: Listar Pacientes personales	16
HU-07: Crear Cita	19
HU-08: Borrar Cita	22
HU-09: List Citas	24
HU-10: Listar Citas por fecha	26
HU-11: Crear Historia Clínica	29
HU-12: Editar Historia Clínica	32
HU-13: Mostrar Historia Clínica	34
HU-14: Buscar Pacientes	36
HU-15: Añadir Informe a Cita	38
HU-16: Editar Informe	41
HU-17: Borrar Informe	43
HU-18: Añadir Tratamiento a Informe	45
HU-19: Editar Tratamiento	48
HU-20: Borrar Tratamiento a Informe	50
HU-21: Informe Privado	53
HU-22: Añadir/Fliminar Informe a una Historia C	linica 55

## Definición de Buen Rendimiento:

Definimos como condiciones aceptables de rendimiento estos dos criterios:

- El tiempo de respuesta medio es menor a 1 segundo (1000 ms)
- El 95% de las peticiones deben acabar en OK
- Existe un tercer assertion que comprueba que la petición que más tarde en realizarse en el sistema no supere los 5 segundos (5000ms), pero debido a posibles eventualidades no se considerará un criterio de peso para determinar si hay buen rendimiento.

# Documentación performance realizadas:

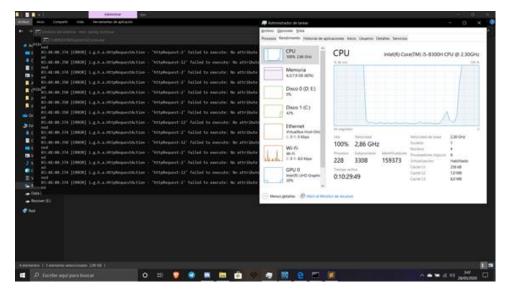
# **HU-01: Agregar Paciente**

Para la siguiente HU se proponen los siguientes dos escenarios:

- Creación exitosa de un paciente en el sistema:
  - El usuario accede a Home
  - El usuario se identifica en el sistema
  - El usuario crea un paciente nuevo con éxito cumplimentando todos los campos requeridos
- Creación fallida de un paciente en el sistema:
  - El usuario accede a Home
  - El usuario se identifica en el sistema
  - El usuario intenta crear un paciente sin datos de contacto (Email, Domicilio o Teléfono) pero es devuelto al formulario.

#### Prueba de estrés:

Se realizó una prueba de estrés que introdujo 9300 usuarios en cada escenario (Un total de 18600 usuarios), con un límite de tiempo de 10 segundos para introducirlos todos.



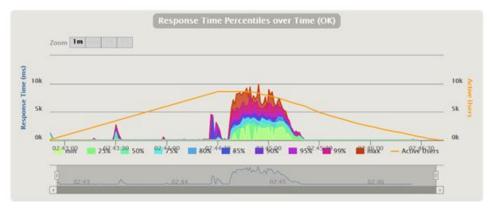
Como observamos, la CPU al alcanzar su máximo uso provoca que el sistema comience a generar tickets csrf incorrectos durante las operaciones post (Como se observa en la ejecución de Gatling) o directamente no sea capaz de generarlos. Seguramente el alto uso de memoria RAM también influyese en la generación de dichos tickets.

#### Prueba de Buen Rendimiento:

Se realizó una prueba que buscaba obtener cual era el número de usuarios máximo bajo los que el sistema era capaz de funcionar en condiciones aceptables.

Los resultados indicaron que el máximo número de usuarios que estos escenarios podían soportar con buen rendimiento era de 4300 usuarios en cada escenario (8600), con un límite de tiempo de 100 segundos para introducirlos todos.







02:43:30

02:44:00

02:44:30

02:45:00

02:43:00

Dado los datos generados por el report, podemos observar que sólo un 2% de las peticiones acabaron en KO (Cuando durante el pico de usuarios, los últimos intentaron acceder al Home) pero todos los usuarios pudieron realizar los escenarios. La media de tiempo de aceptación de una request fue de 516ms, siendo los valores más altos durante la ejecución de la creación de paciente. Aumentar el número de usuarios hubiera resultado en un tiempo medio de aceptación de request mayor a un segundo.

02:46:00

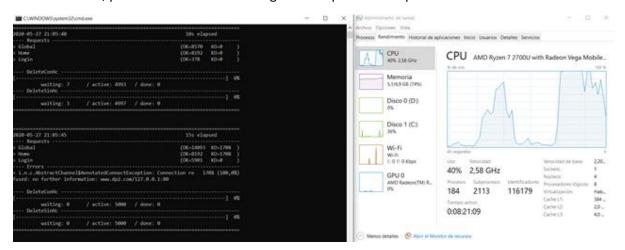
02:46:30

## **HU-02: Borrar Paciente**

Esta Historia de Usuario trata sobre la posibilidad de un médico de borrar del sistema un paciente asociado a él. Para sus pruebas, se han escogido dos escenarios, ambos negativos. El primero comenzará borrando correctamente un paciente, pero el resto de peticiones ejecutarán el mismo proceso por lo que intentarán borrar un paciente que no existe. El segundo escenario tratará de eliminar un paciente que tiene una historia clínica asociada, algo que no está permitido en el sistema.

#### Prueba de estrés

A continuación, procedemos a analizar la siguiente captura de las pruebas realizadas:



En esta captura podemos observar el cuello de botella producido en la CPU del sistema cuando se alcanzan los 10.000 usuarios simultáneos, produciendo una excepción en Gatling y haciendo que el sistema falle. Por ello, podemos concluir que para la CPU en la que se está realizando la prueba, el límite de usuarios simultáneos que puede soportar es de 10.000.

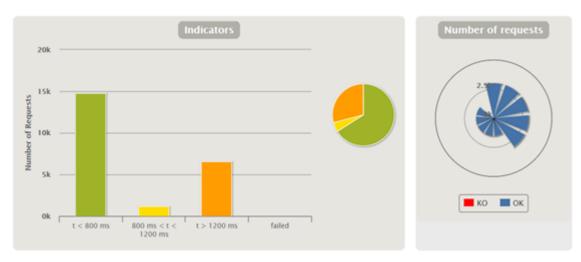
#### Prueba de rendimiento

A continuación, procedemos a analizar la siguiente captura de las pruebas realizadas:

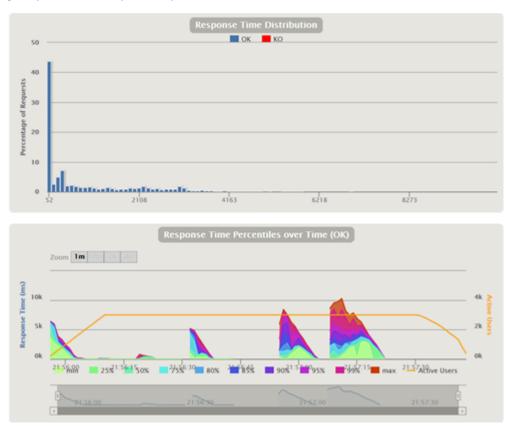


En esta captura podemos observar el comportamiento del sistema cuando existen 3000 usuarios simultáneos ejecutando la historia de usuario actual. Podemos observar que la primera de las tres

restricciones que estamos controlando no se cumple, siendo el tiempo máximo de espera el doble del límite establecido. Aunque el tiempo medio entra dentro del límite impuesto, nos encontramos con 940ms de media, muy cercano al valor establecido. Podemos observar que el comportamiento del percentil 99 es en gran parte de los casos muy poco deseable, añadiendo además que en el percentil 95 los valores son en la gran mayoría el límite de tiempo máximo establecido. Podemos concluir que, para este caso, 3000 usuarios simultáneos es el límite que nos proporcionaría un rendimiento favorable en la aplicación. A continuación, se muestran las gráficas que detallan el comportamiento descrito.



Podemos observar que, aunque no existan KOs, el rendimiento del sistema es desfavorable pues una gran parte del tiempo de espera ha sido excesivo.



Aquí podemos observar que aun no habiendo KOs el tiempo de respuesta es excesivo en las últimas peticiones.



En estas últimas imágenes existe poco que analizar más allá de todo lo que se ha comentado previamente.

## HU-03: Modificar Paciente

#### Prueba de estrés

Vamos a comprobar en las siguientes capturas como con un número de 50.000 usuarios concurrentes (mínimum number of concurrent users that cannot be supported by our scenarios) recibiremos un mensaje de error en el cual se nos notifica que no existe ningún atributo "stoken" definido. Esto indica que el bottleneck para este caso es nuestra CPU para la historia de usuario HU-03.



#### Prueba de rendimiento

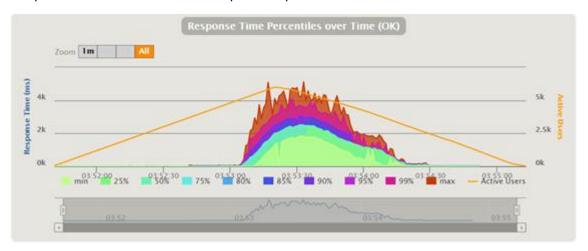
Vamos a distinguir dos escenarios. En primer lugar, listaremos los pacientes personales para modificar un paciente con éxito y para el otro escenario listaremos también la lista de pacients personales para modificar un paciente con errores en el formulario (en este caso se deja un campo vacío, lo cual acabaría en excepción).

Para que este caso de uso se comporte de una manera aceptable, el número de usuarios concurrentes **debe de ser 6.000** (maximun number of concurrent users that our scenarios supports having a proper performance). Consideraremos que el sistema se comporta de una manera aceptable si el número de eventos realizados correctamente sea mayor del 95% y el tiempo medio de respuesta sea menor a 1 segundo. El mayor tiempo de respuesta será de 5 segundos (60 segundos de time-out).



En el siguiente grafico podemos observar cómo al llegar al máximo de usuarios concurrentes el tiempo de respuesta aumenta considerablemente.

Este aumento del tiempo de respuesta es debido a que, con 6.000 usuarios, el sistema proporciona un peor rendimiento, por lo que para un número mayor de usuarios activos en el sistema empezaría a experimentar un aumento del tiempo de respuesta.



En este gráfico podemos observar el número de respuestas por segundo. En el podemos observar que, aunque llegamos al pico de usuarios activos, no nos devuelve respuestas erróneas, sino que estas siguen siendo satisfactorias.

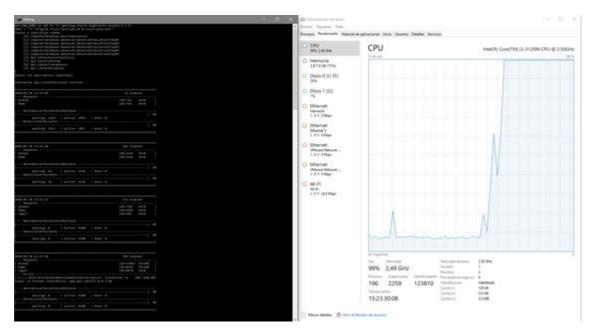


# **HU-04: Listar Pacientes**

# Prueba de estrés

El caso de mínimo nivel de estrés se consigue con 8400 usuarios concurrentes durante 10 segundos.

El cuello de botella se produce en la CPU, como se muestra en la captura, al igual que en todos los casos.

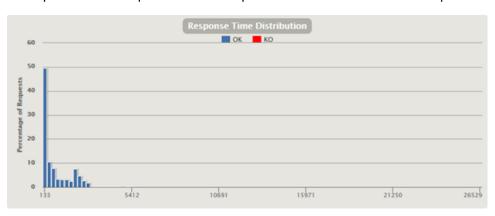


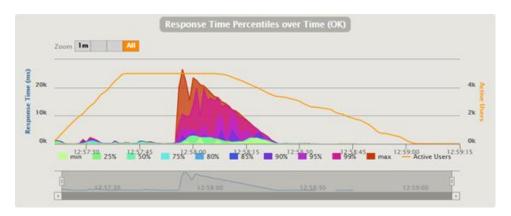
#### Prueba de rendimiento

Por otro lado, el máximo número de usuarios concurrentes que permite el sistema para funcionar correctamente es de 5000.

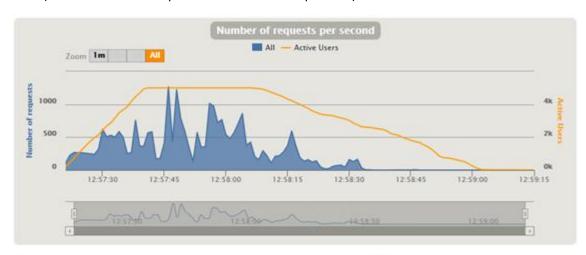


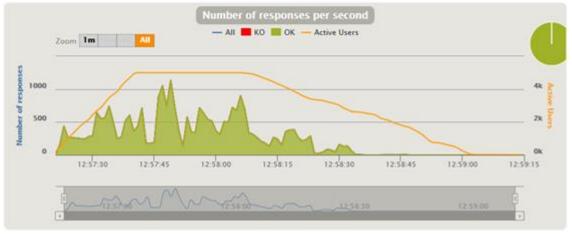
Listar pacientes es la que más recursos parece consumir debido a al tiempo medio de respuesta.





Se alcanzan máximos de tiempo de respuesta superiores a 20000 ms, en el punto medio donde se encuentran todos los usuarios concurrentes, a partir de ahí se empiezan a desconectar la mitad de ellos (Uno de los casos de prueba termina antes que otro).





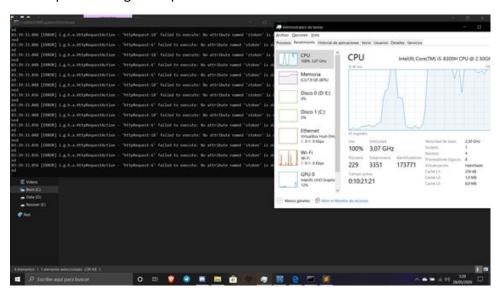
## HU-05: Mostrar Datos de Paciente

Para la siguiente HU se proponen los siguientes escenarios:

- Acceso a los datos de un paciente:
  - El usuario accede al Home del sistema
  - El usuario se identifica en el sistema
  - El usuario accede al listado de pacientes
  - El usuario accede a los datos del paciente 1
- Acceso a los datos de un paciente no existente en BBDD:
  - El usuario accede al Home del sistema
  - El usuario se identifica en el sistema
  - El usuario accede al listado de pacientes
  - El usuario accede a los datos del paciente 1
  - El usuario intenta hacer un acceso ilegal a los datos del paciente 9999 que no existe en el sistema.

## Prueba de estrés:

Se realizó una prueba de estrés que introdujo 12000 usuarios en cada escenario (24000), con un límite de tiempo de 10 segundos para introducirlos todos.



Cuando la CPU alcanzó un 100% de uso continuado, comenzó a generar incorrectamente los tickets csrf o directamente a no generarlos como se puede observar en la traza de ejecución de Gatling. Seguramente el alto uso de memoria RAM también influyese en la generación de dichos tickets.

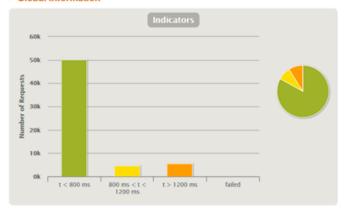
#### Prueba de buen rendimiento:

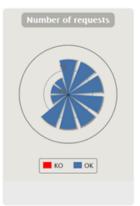
Se realizó una prueba que buscaba obtener cual era el número de usuarios máximo bajo los que el sistema era capaz de funcionar en condiciones aceptables. Definimos como condiciones aceptables estos dos criterios:

- El tiempo de respuesta medio es menor a 1 segundo (1000 ms)
- El 95% de las peticiones deben acabar en OK
- Existe un tercer assertion que comprueba que la petición que más tarde en realizarse en el sistema no supere los 5 segundos (5000ms), pero debido a posibles eventualidades no se considerará un criterio de peso para determinar si hay buen rendimiento.

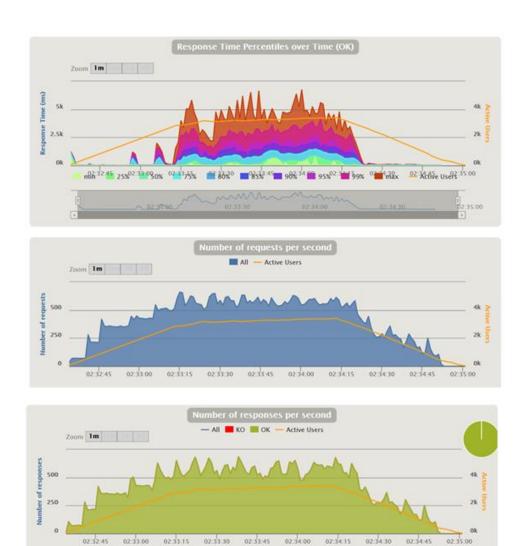
Se descubrió que el número de usuarios concurrentes era de un máximo aproximado de 3550 en cada escenario (7100), con un límite de tiempo de 100 segundos para introducirlos todos.

#### > Global Information









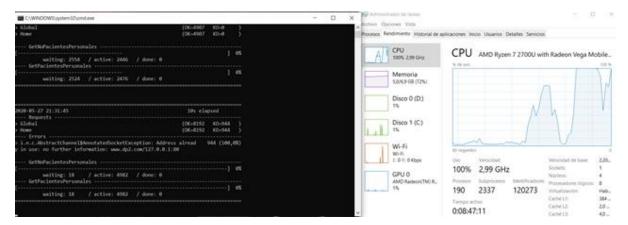
No se rechazó ninguna request durante esta ejecución y el tiempo de asistencia medio de requests fue de 408ms, puede que parezca que existe bastante espacio para introducir más usuarios, pero otras pruebas con un número mayor comenzaban a rechazar más de un 5% de las request realizadas. Las operaciones más costosas correspondieron al envío del login y al acceso a los datos de un paciente. Aumentar el número de usuarios hubiera resultado en un tiempo medio de aceptación de request mayor a un segundo.

## **HU-06: Listar Pacientes personales**

Esta Historia de Usuario trata sobre listar todos los pacientes que posee el médico activo en el sistema. Para sus pruebas, se han escogido dos escenarios, uno positivo y uno negativo. El positivo devolverá todos los pacientes asociado al médico en cuestión, en cambio, el negativo es realizado por un médico sin pacientes asociados, por lo que será redirigido a una lista de todos los pacientes del sistema.

#### Prueba de estrés

A continuación, procedemos a analizar la siguiente captura de las pruebas realizadas:



En esta captura podemos observar el cuello de botella producido en la CPU del sistema cuando se alcanzan los 10.000 usuarios simultáneos, produciendo una excepción en Gatling y haciendo que el sistema falle. Por ello, podemos concluir que para la CPU en la que se está realizando la prueba, el límite de usuarios simultáneos que puede soportar es de 10.000.

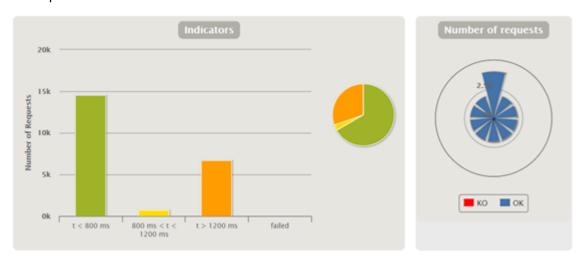
## Prueba de rendimiento

A continuación, procedemos a analizar la siguiente captura de las pruebas realizadas:



En esta captura podemos observar el comportamiento del sistema cuando existen 4.000 usuarios simultáneamente realizando la historia de usuario en cuestión. En este caso, observamos en el tiempo

medio es prácticamente 1 segundo, lo cual es una medida aceptable, además de que no se observan KOs en el sistema. En cambio, el tiempo máximo supera con creces los 5 segundos, llegando a sobrepasar los 13. Podemos observar que en el percentil 95 el sistema comienza a dar en las últimas acciones un comportamiento muy poco adecuado, y en el percentil 99, completamente inadecuado. Podemos deducir que, si se aumentan un poco más los usuarios simultáneos en el sistema, éste comenzará a generar un comportamiento inaceptable incluso sobrepasando el segundo de media. Por lo tanto, concluimos que en este caso 4.000 usuarios simultáneos es el límite con el que el sistema proporcionará un comportamiento adecuado. A continuación, se muestran las gráficas que detallan el comportamiento descrito.



En esta gráfica observamos que el tiempo de respuesta del sistema no es demasiado favorable en gran parte de los casos, sobrepasando los 1200ms.



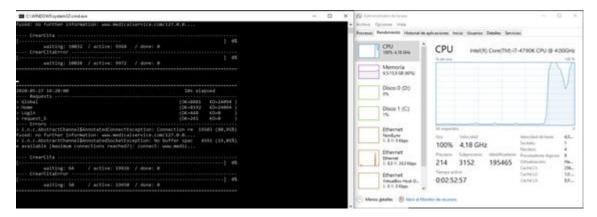
En esta imagen observamos que a pesar de no haber KOs, existe un significativo pico donde el tiempo de respuesta es excesivamente alto.



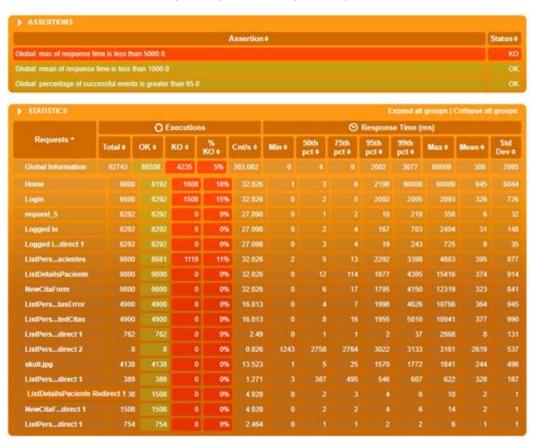
Por último, en esta imagen no existe nada remarcable que comentar más allá de lo analizado anteriormente, pues no se muestran KOs en el test ejecutado.

## HU-07: Crear Cita

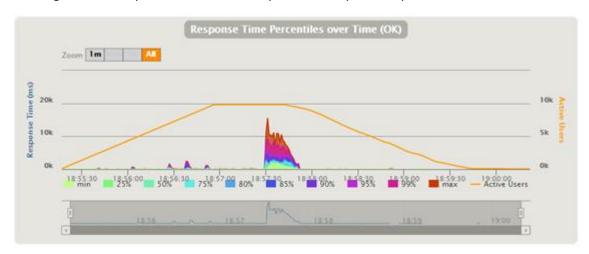
En la siguiente captura podemos observar que con un número de 40.000 usuarios concurrentes (minimum number of concurrent users that cannot be supported by our scenarios) se nos muestra un mensaje de error en el cual nos dice que el espacio del buffer es insuficiente. Esto indica que el bottleneck para este caso es nuestra CPU para la historia de usuario HU-07.



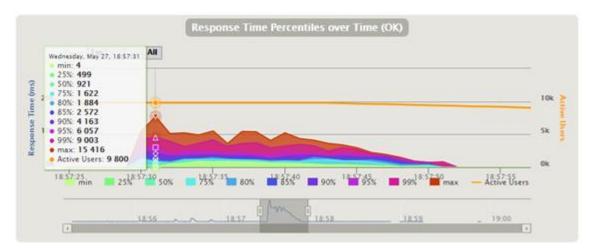
Para que este caso de uso se comporte de una manera aceptable, el número de usuarios concurrentes debe de ser 8.192 (maximun number of concurrent users that our scenarios supports having a proper performance). Para considerar que el sistema se comporta de una manera aceptable, comprobamos que el número de eventos realizados correctamente sea del 95% y el tiempo medio de respuesta sea menor de 1 segundo. El tiempo mayor de respuesta es mayor de 5 segundos (60 segundos de timeout). El escenario ha sido realizado con 9.800 usuarios, pero podemos observar cómo solo 8.192 del total reciben respuesta del servidor. El resto de los usuarios recibe time-out tras 60 segundos al intentar acceder al sistema, por lo que deben esperar a que el resto de usuarios termine.



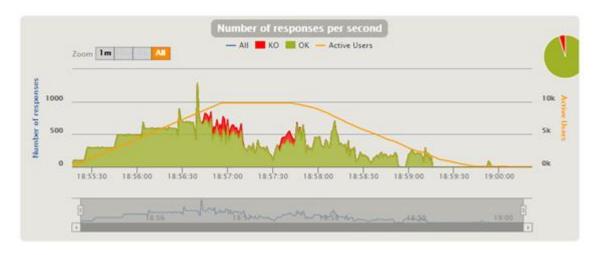
En la siguiente tabla podemos observar un pico del tiempo de respuesta.



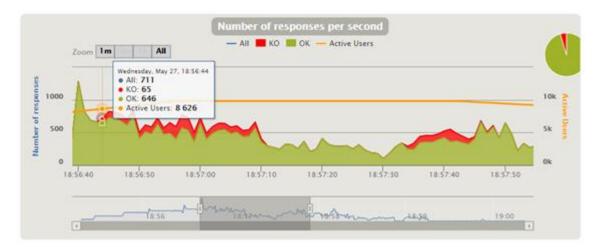
Este pico se debe a que, con 9.800 usuarios, el tiempo de respuesta comienza a incrementar, por lo que para un número igual o superior a 9.800 usuarios activos el sistema empezaremos a experimentar un aumento del tiempo de respuesta en el sistema.



Esta tabla nos muestra el número de respuestas por segundo. Observamos como para casi el límite de usuarios conectados concurrentemente, el sistema no es capaz de devolver una respuesta satisfactoria.

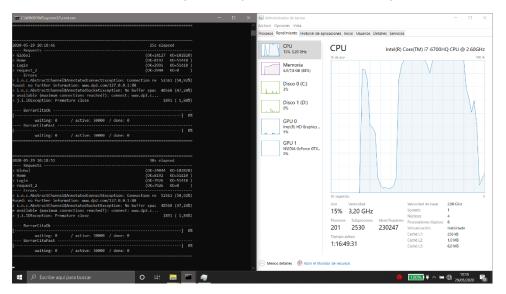


Observamos que el límite de usuarios concurrentes para los que el sistema devuelve respuestas satisfactorias es de unos 8.626 usuarios activos.

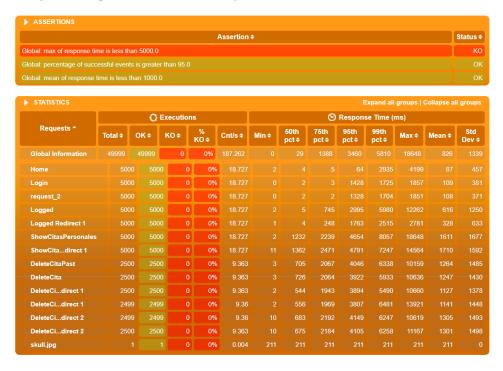


## HU-08: Borrar Cita

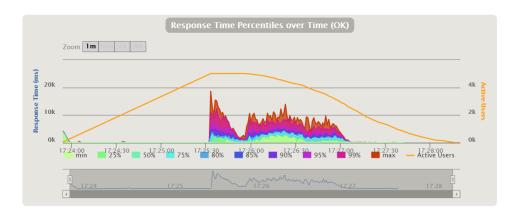
En una primera captura de la ejecución del test de estrés, podemos observar que son 30000 usuarios concurrentes los que ponen en compromiso serio la estabilidad del sistema, recibiendo errores de espacio en búfer, entre otros. En este punto se produce un cuello de botella por el procesador, que está al 100% de rendimiento y no es capaz de atender todas las peticiones adecuadamente.



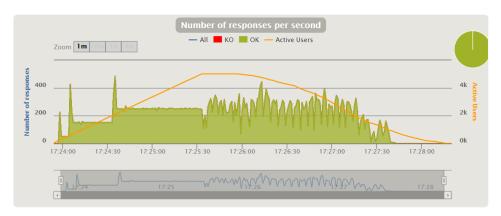
Encontramos un desempeño límite en los 5000 usuarios concurrentes, punto en el cual, aunque el sistema sobrepasa los 5 segundos de tiempo de respuesta máxima (18648ms), aún se mantiene por debajo de 1 segundo de media de respuesta en todos los eventos.



En esta tabla observamos el tiempo de respuesta por percentil. Se puede ver como al llegar al número máximo de usuarios concurrentes el tiempo de respuesta aumenta drásticamente para cada petición debido a la creciente saturación del sistema.

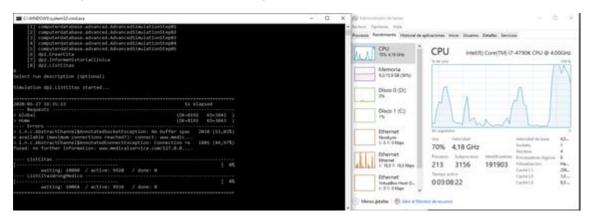


En la siguiente captura de pantalla observamos el número de respuestas por segundo. Se responden dos picos de peticiones cuando la carga de usuarios está aumentando, luego se produce una estabilización. Cuando la carga de usuarios llega a su máximo, se producen picos sucesivos relativos a la respuesta de los diferentes eventos de borrado que se están intentando llevar a cabo en el sistema.



## HU-09: List Citas

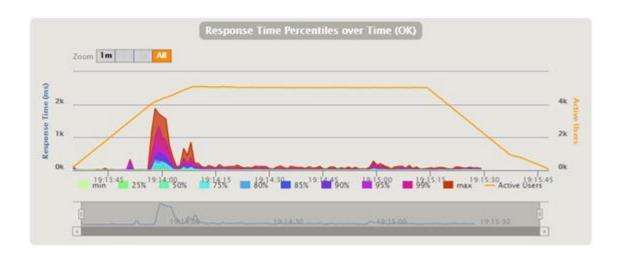
En la siguiente captura podemos observar que con un número de 40.000 usuarios concurrentes (minimum number of concurrent users that cannot be supported by our scenarios) se nos muestra un mensaje de error en el cual nos dice que el espacio del buffer es insuficiente. Esto indica que el bottleneck para este caso es nuestra CPU para la historia de usuario HU-09.



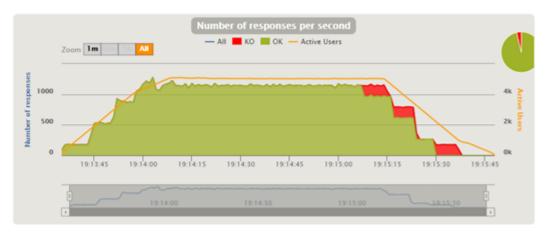
Para que este caso de uso se comporte de una manera aceptable, el número de usuarios concurrentes debe de ser 17.600 (maximun number of concurrent users that our scenarios supports having a proper performance). Para considerar que el sistema se comporta de una manera aceptable, comprobamos que el número de eventos realizados correctamente sea del 95%, el tiempo medio de respuesta sea menor de 1 segundo y el tiempo máximo de respuesta es menor de 5 segundos.



En la siguiente gráfica de respuestas en el tiempo podemos observar algunos picos, pero no muy notorios. El más significativo se encuentra al principio y se debe al inicio masivo de usuarios de esta prueba.



En la gráfica número de respuestas por segundo se observa cómo se mantiene en el tiempo el número de respuestas satisfactorias. En ella nos encontramos con algunas repuestas no satisfactorias al final de la gráfica.



Podemos observar que para 5.036 usuarios concurrentes el sistema no responde adecuadamente para algunas de las peticiones.

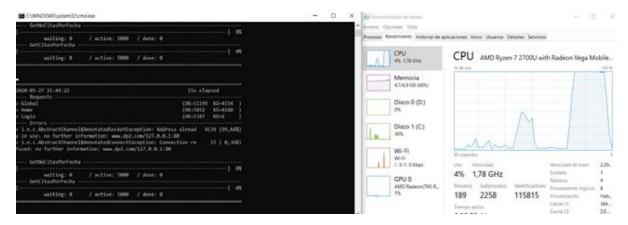


## HU-10: Listar Citas por fecha

Esta Historia de Usuario trata sobre listar todas las citas de un determinado médico que posean la fecha que se indique. Para sus pruebas, se han escogido dos escenarios, uno positivo y uno negativo. El positivo devolverá las citas que posean la fecha introducida, en cambio, el negativo introducirá una fecha que no tenga ninguna cita, por lo que devolverá una lista vacía.

#### Prueba de estrés

A continuación, procedemos a analizar la siguiente captura de las pruebas realizadas:



En esta captura podemos observar el cuello de botella producido en la CPU del sistema cuando se alcanzan los 10.000 usuarios simultáneos, produciendo una excepción en Gatling y haciendo que el sistema falle. Por ello, podemos concluir que para la CPU en la que se está realizando la prueba, el límite de usuarios simultáneos que puede soportar es de 10.000.

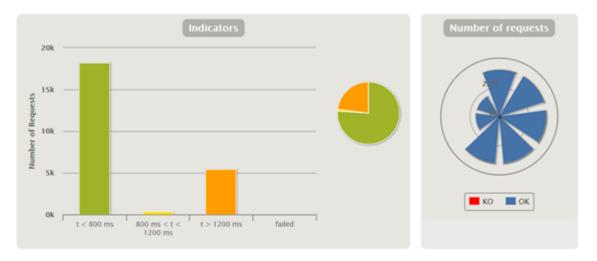
#### Prueba de rendimiento

A continuación, procedemos a analizar la siguiente captura de las pruebas realizadas:

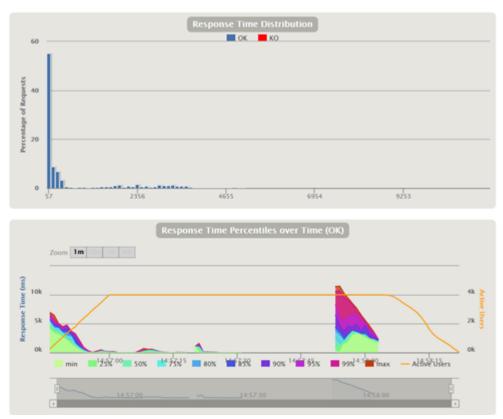


En esta captura podemos observar el comportamiento del sistema cuando existen 4.000 usuarios simultáneamente realizando la historia de usuario en cuestión. Aunque el tiempo medio no resulta desfavorable, el tiempo máximo en los percentiles 95 y 99 se sobrepasa. Podemos observar que el comportamiento es muy similar al de la HU-06. Esto se debe a que ambas historias de usuarios son

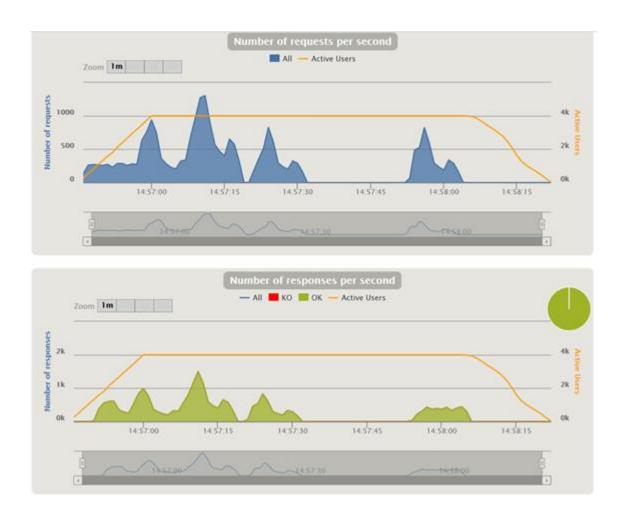
similares y no son altamente complejas, pues podemos observar que no sobrecarga el sistema de manera significativa. Así pues, podemos concluir que en este caso 4.000 usuarios simultáneos es el límite con el que el sistema proporcionará un comportamiento adecuado, puesto que aumentarlo un poco más haría que el tiempo máximo de respuesta se disparase. A continuación, se muestran las gráficas que detallan el comportamiento descrito.



En esta gráfica observamos que, en parte de los casos, a pesar de no existir KOs, se muestra un tiempo de respuesta desfavorable que supera los 1200ms.



En esta imagen podemos observar que, además de no haber KOs, existe un pico donde el tiempo de respuesta es excesivo, sobrepasando los 10.000ms.



En esta última imagen no existe nada interesante de remarcar más allá de lo descrito previamente, puesto que no existen KOs en este test.

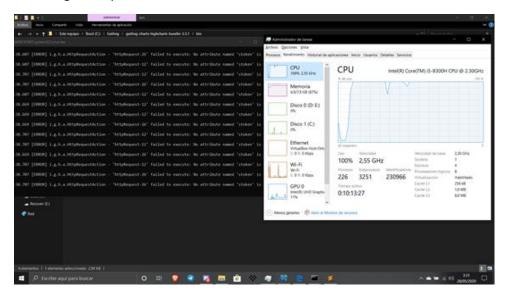
## HU-11: Crear Historia Clínica

Se definen los siguientes escenarios para esta Historia de Usuario:

- Un usuario crea una historia clínica:
  - Usuario accede a Home
  - Usuario se identifica en el sistema
  - Usuario accede a sus pacientes personales
  - Usuario accede a los datos de un paciente sin Historia Clínica
  - Usuario crea una Historia Clínica en el sistema
  - Dado que un Paciente sólo puede tener una Historia Clínica, sólo un usuario es capaz de crearla. El resto realizarían un acceso ilegal (No existe el botón en la vista) al formulario de creación y serían devueltos a la página de error.
  - El paciente sólo tiene una Historia Clínica al final de la ejecución.
- Un usuario Médico intenta crear una Historia Clínica para un paciente que no tiene asignado:
  - Usuario accede a Home
  - Usuario se identifica en el sistema
  - Usuario accede al listado de pacientes completo
  - Usuario accede a un paciente que no tiene asignado
  - Usuario intenta crear una Historia Clínica mediante un acceso ilegal (No existe el botón en la vista) y es redirigido al Home.

#### Prueba de estrés

Se realizó una prueba de estrés con 25000 usuarios en cada escenario (50000), con un límite de tiempo de 10 segundos para introducirlos todos.



Como vemos, cuando la CPU al alcanzó un uso continuado del 100%, los tickets csrf comenzaron a fallar en su generación impidiendo así el uso de una operación post como podemos observar en la traza de Gatling. Seguramente el alto uso de memoria RAM también influyese en la generación de dichos tickets.

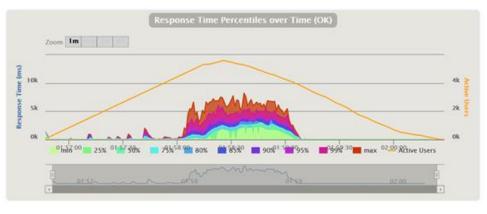
#### Prueba de Buen Rendimiento:

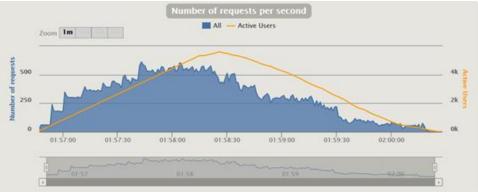
Se realizó una prueba que buscaba obtener cual era el número de usuarios máximo bajo los que el sistema era capaz de funcionar en condiciones aceptables. Definimos como condiciones aceptables estos dos criterios:

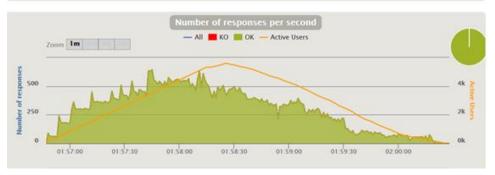
- El tiempo de respuesta medio es menor a 1 segundo (1000 ms)
- El 95% de las peticiones deben acabar en OK
- Existe un tercer assertion que comprueba que la petición que más tarde en realizarse en el sistema no supere los 5 segundos (5000ms), pero debido a posibles eventualidades no se considerará un criterio de peso para determinar si hay buen rendimiento.

Se descubrió que el número máximo de usuarios concurrentes era de aproximadamente 3000 en cada escenario (6000 usuarios), con un límite de tiempo de 100 segundos para introducirlos todos.







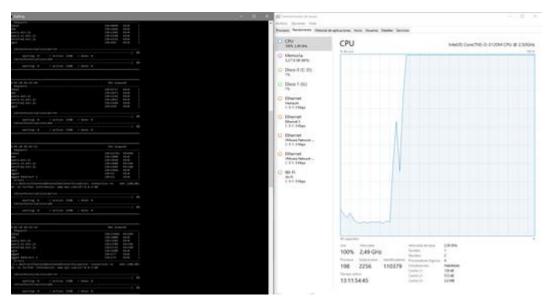


Todas las request fueron aceptadas, el tiempo medio de aceptación de las request fue de 645ms. Si el número de usuarios hubiese sido superior, aunque quizá si se hubieran atendido casi todas las request, el tiempo medio de aceptación hubiera superado el segundo. Observamos que 2999 usuarios del primer escenario acabaron en una página de error debido a la restricción de Historia Clínica. Las operaciones más costosas fueron el acceso a los datos de un Paciente perteneciente a un médico distinto, el acceso ilegal a la creación de una Historia Clínica para este y el acceso fallido a la creación de una Historia Clínica para un médico asignado al usuario (Fallido dado que fueron dirigidos a una página de error).

## HU-12: Editar Historia Clínica

## Prueba de estrés

El caso de mínimo nivel de estrés se consigue con 3000 usuarios concurrentes.

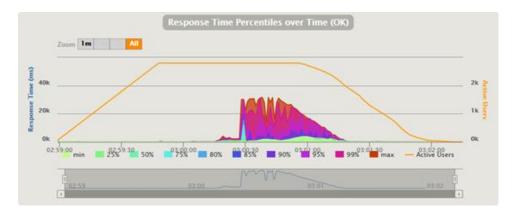


## Prueba de rendimiento

Por otro lado, el máximo número de usuarios concurrentes que el sistema permite en esta historia de usuario es de <u>2800</u> usuarios.



Se experimentaron algunos fallos en los request que más CPU consumen.



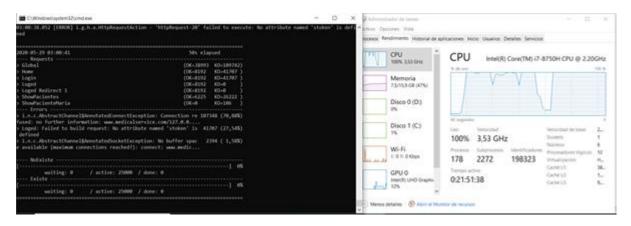
Los tiempos de respuesta máximo llegan a ser de hasta 30000 ms, pero se dan de forma reducida.



## HU-13: Mostrar Historia Clínica

#### Prueba de estrés

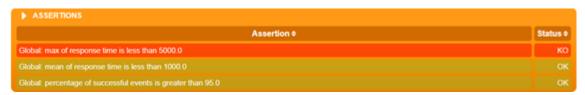
Vamos a comprobar en las siguientes capturas como con un número de 50.000 usuarios concurrentes (mínimum number of concurrent users that cannot be supported by our scenarios) recibiremos un mensaje de error en el cual se nos notifica que el espacio del buffer es insuficiente y que no existe ningún atributo "stoken" definido. Esto indica que el bottleneck para este caso es nuestra CPU para la historia de usuario HU-13.



#### Prueba de rendimiento

Vamos a distinguir dos escenarios. En primer lugar, mostraremos la lista de pacientes para mostrar así un paciente en concreto, entonces le pediremos al sistema que nos muestre una historia clínica existente, mientras que en el otro escenario mostraremos también la lista de pacientes para mostrar así un paciente en concreto, pero en este caso le pediremos al sistema que nos muestre una historia clínica inexistente, lo cual acabaría en excepción.

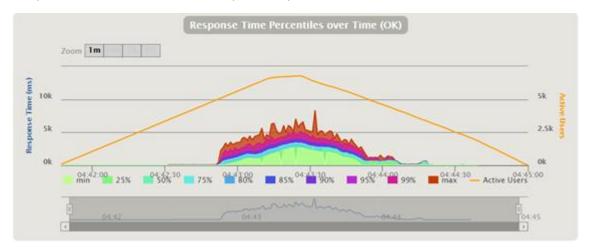
Para que este caso de uso se comporte de una manera aceptable, el número de usuarios concurrentes **debe de ser 7.800** (maximun number of concurrent users that our scenarios supports having a proper performance). Consideraremos que el sistema se comporta de una manera aceptable si el número de eventos realizados correctamente sea mayor del 95% y el tiempo medio de respuesta sea menor a 1 segundo. El mayor tiempo de respuesta será de 5 segundos (60 segundos de time-out).



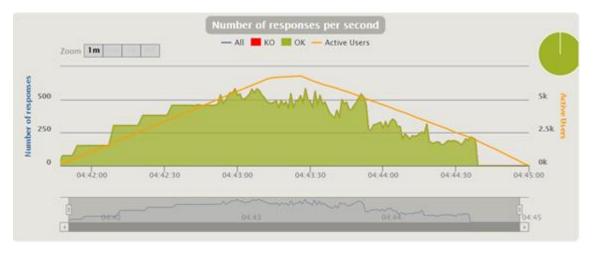


En el siguiente grafico podemos observar cómo, al llegar al máximo de usuarios concurrentes, el tiempo de respuesta aumenta considerablemente.

Este aumento del tiempo de respuesta es debido a que, con **7.800 usuarios**, el sistema proporciona un peor rendimiento, por lo que para un número mayor de usuarios activos en el sistema empezaría a experimentar un aumento del tiempo de respuesta.

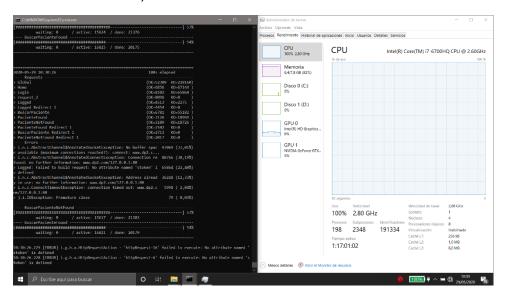


En este gráfico podemos observar el número de respuestas por segundo. En el podemos observar que, aunque llegamos al pico de usuarios activos, no nos devuelve respuestas erróneas, sino que estas siguen siendo satisfactorias.



## **HU-14: Buscar Pacientes**

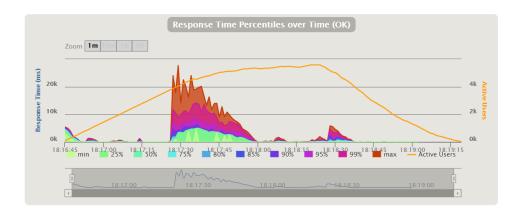
En el caso de esta historia de usuario se pudo aumentar ligeramente el número de usuario concurrentes límite del sistema. En total son 25000 usuarios los que producen un cuello de botella sobre las capacidades del procesador, produciendo errores de búffer, así como la memoria se acerca también a su máximo, saturándose en un 85%.



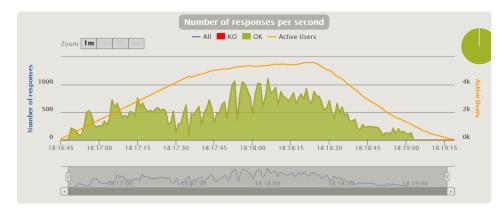
En esta ejecución, al ser más ligera, el sistema responde de manera aceptable con un mayor número de usuarios concurrentes. En este caso son 9000 los que se pueden manejar cumpliendo al menos la restricción de no tardar más de un segundo en dar respuesta de media. El tiempo máximo de respuesta es de 27654ms, mientras que el tiempo medio es de 828ms.



Si observamos en tiempo de respuesta en percentil, de nuevo observamos que el 90% de las peticiones aumentan su tiempo de respuesta en el momento en que los usuarios empiezan a aumentar, quedando tiempos de respuesta entre 10000ms y 20000ms, superando estos últimos en alguna ocasión, en el momento de la ejecución en el que se están dando las respuestas a las peticiones de buscar el paciente.



En el gráfico inferior, de forma complementaria al anterior, podemos observar como el número de respuestas por segundo producido aumenta de forma paralela al tiempo de respuesta durante la rampa de inyección de usuarios, aunque esta vez de una manera más suavizada y con unas 1000 respuestas por segundo en sus mayores picos.

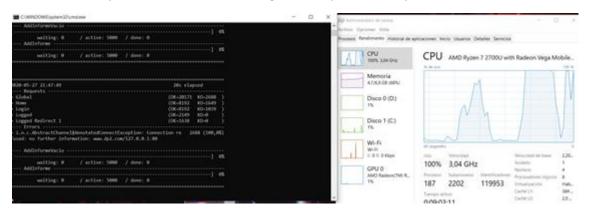


## HU-15: Añadir Informe a Cita

Esta historia de usuario trata sobre crear un informe asociado a una cita existente. Para la realización de sus pruebas se han escogido dos escenarios: uno positivo y uno negativo. El positivo creará un informe correctamente y lo asociará a una cita, en cambio, el escenario negativo intentará crear un informe con sus propiedades en blanco, por lo que notificará de un error pues sus parámetros "motivo" y "diagnóstico" no pueden estar vacíos. Cabe destacar que, para llevar a cabo esta prueba, previamente se ha creado una cita con fecha del día de hoy. Esto se debe a que solo pueden añadirse informes a citas cuya fecha sea el día actual, así que para poder realizar esta prueba correctamente, debe existir una cita cuya fecha sea el mismo día de realización de la prueba.

#### Prueba de estrés

A continuación, procedemos a analizar la siguiente captura de las pruebas realizadas:



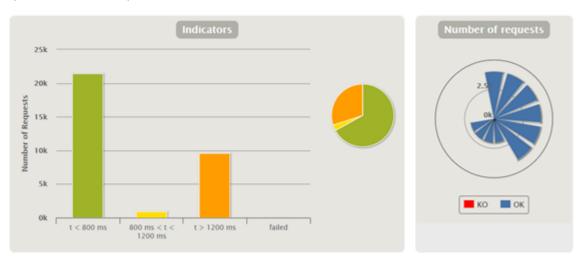
En esta captura podemos observar el cuello de botella producido en la CPU del sistema cuando se alcanzan los 10.000 usuarios simultáneos, produciendo una excepción en Gatling y haciendo que el sistema falle. Por ello, podemos concluir que para la CPU en la que se está realizando la prueba, el límite de usuarios simultáneos que puede soportar es de 10.000.

# Prueba de rendimiento

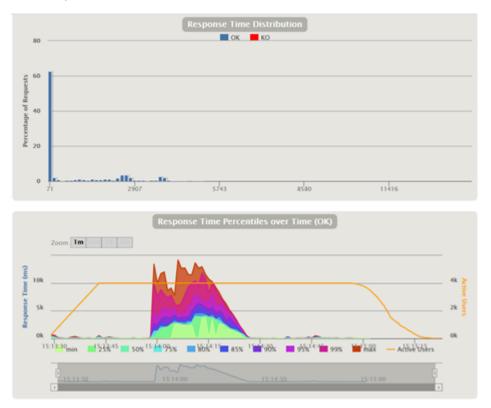
A continuación, procedemos a analizar la siguiente captura de las pruebas realizadas:



Tal como se observa en la captura, con 4.000 usuarios simultáneos el sistema comienza a no tener un comportamiento óptimo, comenzando a producir KOs (aunque ignorables dada su poca frecuencia, puede ser un símbolo de que está comenzando a llegar a su límite) además de tener un tiempo máximo de respuesta inaceptable en muchos de los casos. Aunque el tiempo medio entra dentro del límite establecido, incrementar el número de usuarios simultáneos haría que se elevara y por ello sobrepasara dicho límite. En definitiva, podemos concluir que el límite de usuarios concurrentes que soportará con un comportamiento medianamente aceptable (pues en el percentil 99 ya comienza a superar el tiempo máximo establecido) es de 4.000 usuarios. A continuación, se muestran las gráficas que detallan el comportamiento descrito.



En esta primera gráfica observamos un tiempo de respuesta desfavorable en más de un cuarto de los casos, superando los 1200ms.



En esta segunda imagen observamos que, a pesar de existir un número prácticamente despreciable de KOs, existen múltiples picos indicando un tiempo de respuesta muy desfavorable.



En esta última imagen no existe nada remarcable que analizar más allá de lo comentado previamente, puesto que el número de KOs es tan pequeño que resulta inapreciable.

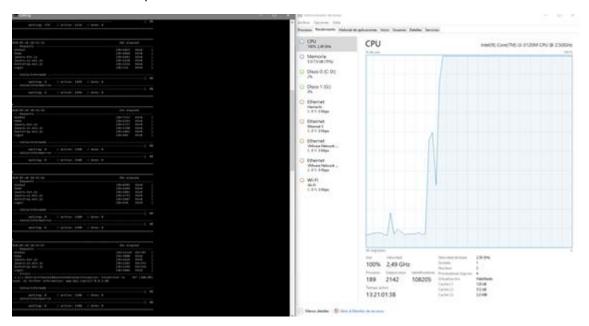
## HU-16: Editar Informe

## Prueba de estrés

Para hacer posible la edición de un informe ha debido cambiarse la fecha de la cita a la que se asocia al día actual para no tener problemas con dicha validación.

El servidor empieza a fallar y negar el servicio cuando se alcanzan los 3000 usuarios concurrentes.

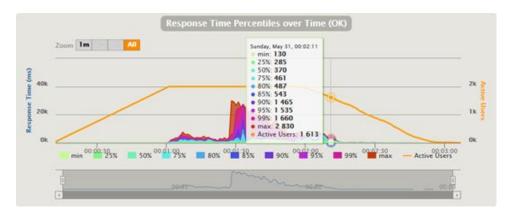
El cuello de botella se produce en la CPU.



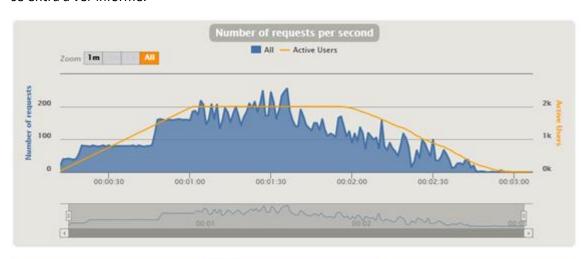
## Prueba de rendimiento

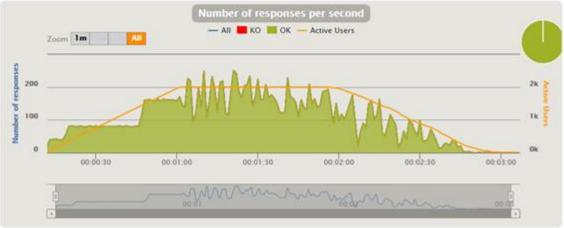
El número máximo de usuarios concurrentes para obtener una buena experiencia es de 2000 usuarios.





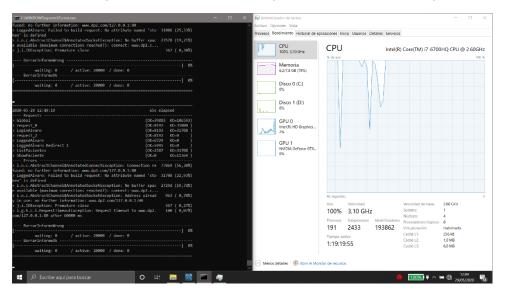
Los máximos tiempos de respuesta se experimentan cerca del punto medio de la ejecución, cuando se entra a ver informe.





## HU-17: Borrar Informe

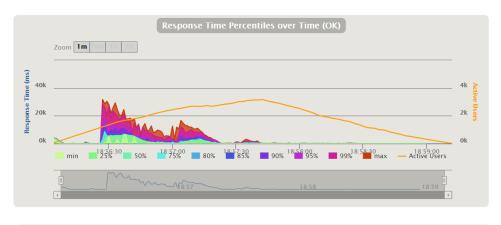
Para esta historia de usuario hemos observado el límite del sistema en 20000 usuarios concurrentes, punto en el cual el rendimiento de la máquina no da más de sí. De nuevo se produce cuello de botella en el procesador mayormente, que no es capaz de atender el alto número de peticiones de la prueba. También se colapsa altamente la memoria, que alcanza el 80% de su capacidad total.

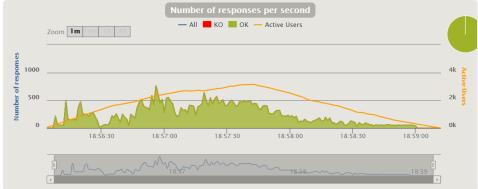




En el caso del comportamiento general en esta historia de usuario observamos que el tiempo máximo de respuesta de eventos es de 31740ms, quedando en KO un 0% de los eventos. Para esta historia de usuario, estas condiciones son prácticamente el límite que el sistema puede dar manteniendo unos valores razonables.

Por otro lado, la media del tiempo de respuesta de todos los eventos se establece en 953ms.





Respecto al tiempo de respuesta por percentiles, se puede observar un pico el inicio, correspondientes a las operaciones de obtención de las citas personales del médico durante la prueba, situándose la mayoría de estos eventos en el rango 20000-40000ms.

Más adelante vemos como esta tendencia se relaja un poco, aunque vuelve a subir hasta los casi 20000ms cuando se pide que se muestre un informe en concreto.

En el caso del número de respuestas por segundo que da el sistema, se puede identificar una rápida subida desde los primeros eventos, manteniéndose en torno al máximo de nuevo hasta el pico para bajar en el descenso del pico a menor velocidad que antes y establecerse en torno a 200 respuestas por segundo hasta casi el final de la operación.

## HU-18: Añadir Tratamiento a Informe

## Prueba de estrés

Vamos a comprobar en las siguientes capturas como con un número de 50.000 usuarios concurrentes (mínimum number of concurrent users that cannot be supported by our scenarios) recibiremos un mensaje de error en el cual se nos notifica que no existe ningún atributo "stoken" definido. Esto indica que el bottleneck para este caso es nuestra CPU para la historia de usuario HU-18.



#### Prueba de rendimiento

Vamos a distinguir dos escenarios. En primer lugar, listaremos las citas personales para mostrar el informe de una, entonces añadiremos un tratamiento con éxito al informe. Para el otro escenario, listaremos las citas personales para mostrar también el informe de una, entonces añadiremos un tratamiento con un error en las fechas del tratamiento, de forma que nos aparezca un error en el formulario.

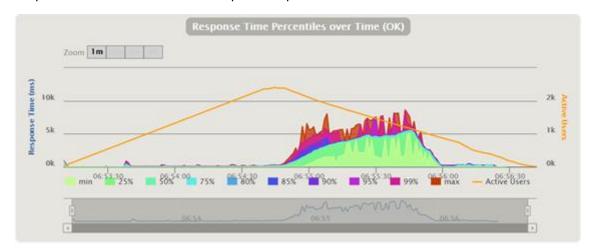
Para que este caso de uso se comporte de una manera aceptable, el número de usuarios concurrentes **debe de ser 2.600** (maximun number of concurrent users that our scenarios supports having a proper performance). Consideraremos que el sistema se comporta de una manera aceptable si el número de eventos realizados correctamente sea mayor del 95% y el tiempo medio de respuesta sea menor a 1 segundo. El mayor tiempo de respuesta será de 5 segundos (60 segundos de time-out).



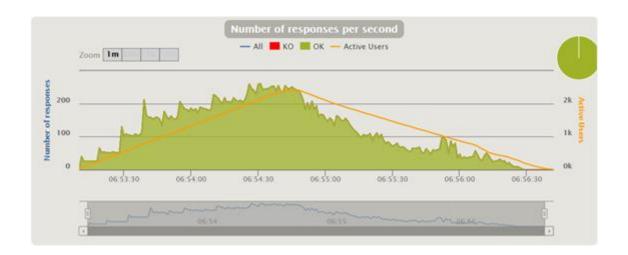


En el siguiente grafico podemos observar cómo, al llegar al máximo de usuarios concurrentes, el tiempo de respuesta aumenta considerablemente.

Este aumento del tiempo de respuesta es debido a que, con 2.600 usuarios, el sistema proporciona un peor rendimiento, por lo que para un número mayor de usuarios activos en el sistema empezaría a experimentar un aumento del tiempo de respuesta.



En este gráfico podemos observar el número de respuestas por segundo. En él podemos observar que, aunque llegamos al pico de usuarios activos, no nos devuelve respuestas erróneas, sino que estas siguen siendo satisfactorias.



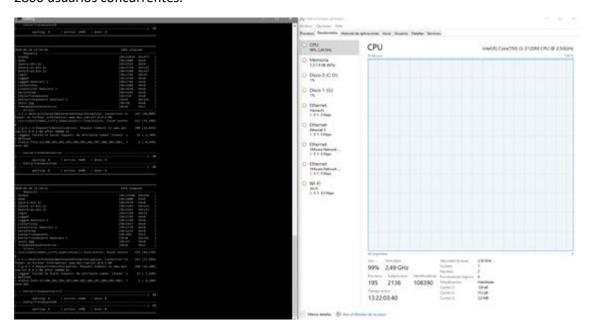
En esta historia de usuario hemos encontrado la peculiaridad de que, a partir de un cierto número de usuarios, el tiempo medio de respuesta a cada petición aumentaba considerablemente. Para que la prueba funcionara con un rendimiento normal, tras una ejecución se debías dropear las tablas de la base de datos, o en cambio los resultados no serían concluyentes.

Por lo tanto, se observó que el proceso que tomaba más tiempo de media al sistema era *ShowInforme*, pues al no existir **paginación** en dicha vista, a partir de cierto número de tratamientos creados en un mismo informe el tiempo de carga de la vista afecta a las siguientes creaciones de tratamientos.

# **HU-19: Editar Tratamiento**

Para hacer posible la edición de un tratamiento ha debido cambiarse la fecha de la cita al día actual para no tener problemas con dicha validación.

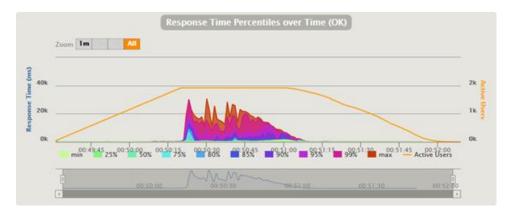
En este caso el nivel mínimo de usuarios concurrentes con el que el servidor experimenta estrés es de 2800 usuarios concurrentes.



El servidor es capaz de funcionar de manera adecuada con 1930 usuarios concurrentes.



Se experimenta mucha inestabilidad cuando se añaden pocos usuarios más, pasando de media de tiempo de respuesta de 700 ms a medias de incluso 3000 ms solo añadiendo 100 usuarios más.



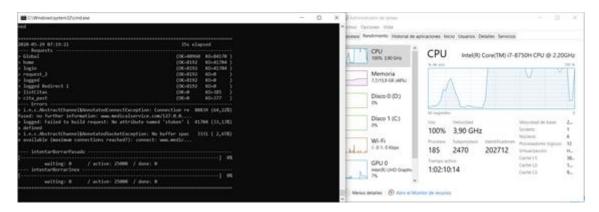
En listar citas y ver el informe es donde se experimentan mayores tiempos de respuesta en toda la ejecución y no en el post de editar tratamiento como podría esperarse.



## HU-20: Borrar Tratamiento a Informe

#### Prueba de estrés

Vamos a comprobar en las siguientes capturas como con un número de 50.000 usuarios concurrentes (mínimum number of concurrent users that cannot be supported by our scenarios) recibiremos un mensaje de error en el cual se nos notifica que el espacio del buffer es insuficiente y que no existe ningún atributo "stoken" definido. Esto indica que el bottleneck para este caso es nuestra CPU para la historia de usuario HU-20.



#### Prueba de rendimiento

Vamos a distinguir dos escenarios. En primer lugar, listaremos las citas personales para mostrar el informe de una, entonces el primer usuario borrará el tratamiento con éxito al informe, haciendo así que el resto de usuarios no puedan borrar ese mismo tratamiento al haber sido ya borrado, redirigiéndoles a una página de error. Para el otro escenario, listaremos las citas personales para mostrar también el informe de una cita con fecha pasada, entonces el usuario intentará borrar el tratamiento mediante un acceso ilegal (ya que no aparecería botón en la vista), por lo que sería redirigido de vuelta a la vista del informe.

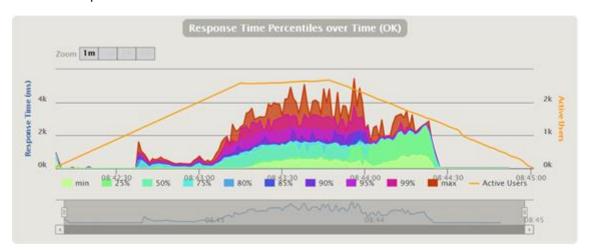
Para que este caso de uso se comporte de una manera aceptable, el número de usuarios concurrentes **debe de ser 3.800** (maximun number of concurrent users that our scenarios supports having a proper performance). Consideraremos que el sistema se comporta de una manera aceptable si el número de eventos realizados correctamente sea mayor del 95% y el tiempo medio de respuesta sea menor a 1 segundo. El mayor tiempo de respuesta será de 5 segundos (60 segundos de time-out).



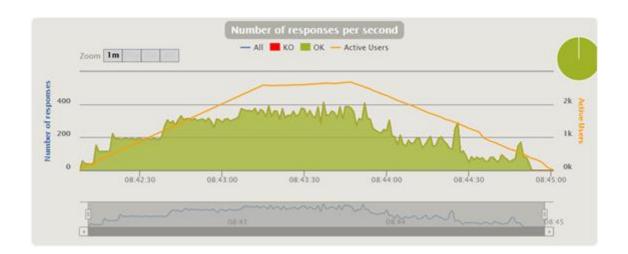


En el siguiente grafico podemos observar cómo, al llegar al máximo de usuarios concurrentes, el tiempo de respuesta aumenta considerablemente.

Este aumento del tiempo de respuesta es debido a que, con 3.800 usuarios, el sistema proporciona un peor rendimiento, por lo que para un número mayor de usuarios activos en el sistema empezaría a experimentar un aumento del tiempo de respuesta. Se ha observado que, en esta historia de uso, al superar dicho número de usuarios el tiempo de respuesta de las peticiones aumentaría exponencialmente, haciendo que con pocos usuarios más los tiempos de respuesta se vuelvan insostenibles por el sistema.



En este gráfico podemos observar el número de respuestas por segundo. En él podemos observar que, aunque llegamos al pico de usuarios activos, no nos devuelve respuestas erróneas, sino que estas siguen siendo satisfactorias.

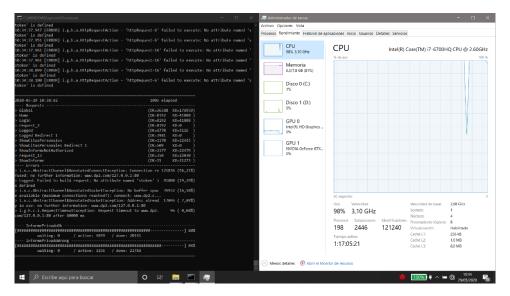


En conclusión, hemos observado que 3.800 usuarios sería el límite aceptable para considerar que el sistema sigue comportándose de manera aceptable. A partir de dicho número, el sistema respondería con éxito a las peticiones, pero empezaría a demorarse en exceso para responderlas.

## **HU-21: Informe Privado**

En el caso de esta historia de usuario encontramos el número máximo de usuarios concurrentes en 25000 usuarios. Es en este caso en el que la CPU alcanza el 100% de ocupación constante, produciendo un cuello de botella en la ejecución de las peticiones que no ven su respuesta a tiempo.

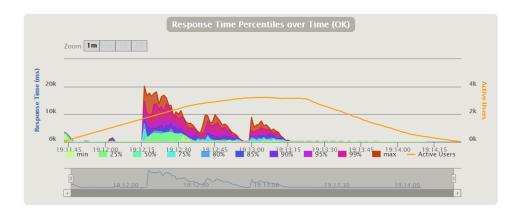
Así, también hay fallos de búfer y la memoria RAM supera el 80% de capacidad consumida por este proceso.



Es con los 5000 usuarios concurrentes con los que el sistema responde de manera adecuada, cumpliendo las restricciones impuestas. El tiempo máximo de respuesta es de 20187ms, mientras que la media se sitúa en 700ms.

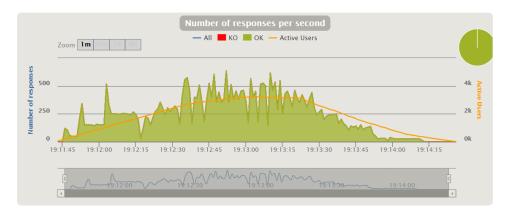


Observamos de nuevo en la gráfica de tiempo de respuesta por percentil como durante la rampa de inyección de usuarios se produce la mayor saturación en el sistema, relativa a las llamadas a listas de Citas, que provoca que el 90% de las peticiones se sitúen en torno a los 20000ms de tiempo de respuesta. Tras esto, se producen picos posteriores más pequeños del orden de 10000ms para el 90% de las peticiones.



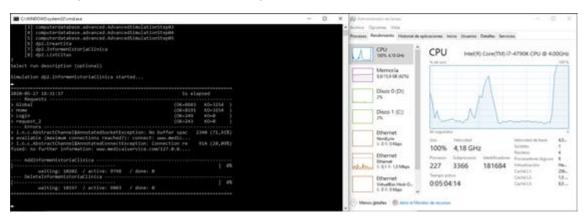
De nuevo el número de respuestas por segundo aumenta durante la rampa de inyección de usuarios a la par que lo hace el tiempo de respuesta por percentil. A mayor número de respuestas por segundo, mayor es el tiempo de respuesta en el que se encuentra el percentil 90%.

En este caso sin embargo se puede apreciar que los mayores picos los encontramos en el momento en que la actividad sobre la aplicación es la más alta, sobrepasando así las 500 respuestas por segundo en esos instantes.



# HU-22: Añadir/Eliminar Informe a una Historia Clínica

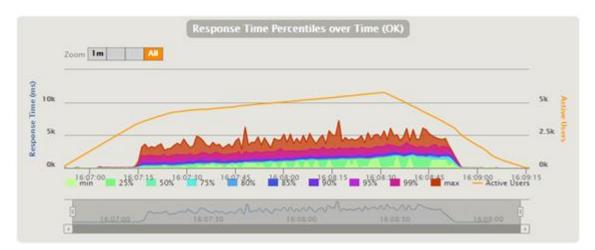
En la siguiente captura podemos observar que con un número de 40.000 usuarios concurrentes (minimum number of concurrent users that cannot be supported by our scenarios) se nos muestra un mensaje de error en el cual nos dice que el espacio del buffer es insuficiente. Esto indica que el bottleneck para este caso es nuestra CPU para la historia de usuario HU-22.



Para que este caso de uso se comporte de una manera aceptable, el número de usuarios concurrentes debe de ser 14.400 (maximun number of concurrent users that our scenarios supports having a proper performance). Para considerar que el sistema se comporta de una manera aceptable, comprobamos que el número de eventos realizados correctamente sea del 95%, el tiempo medio de respuesta sea menor de 1 segundo y el tiempo máximo de respuesta es menor de 5 segundos. En este caso el tiempo máximo de repuesta es de unos 7 segundos, mayor a la comprobación realizada, pero podríamos darlo por aceptable.



En la siguiente gráfica podemos ver como el tiempo de respuesta de la aplicación va incrementado conforme el número de usuarios concurrentes aumenta.



La gráfica de número de respuestas por segundo nos muestra como todas las peticiones son respondidas adecuadamente para el número de usuarios de esta prueba. El principal inconveniente de esta prueba de performance es el tiempo medio de respuesta, dado que con un mayor número de usuarios no podríamos asegurar que se cumpliera.

