Guía de desarrollo, Anexo 03.07

Procedimiento de Pruebas de Rendimiento

Autor: Oficina de Pruebas

GERENCIA INFORMÁTICA
JOSEFA VALCÁRCEL, 44
28027-MADRID

Indice General

1 INTRODUCCIÓN 3

1.1 OBJETO 3
1.2 ALCANCE 3

2 PROCEDIMIENTO DE PRUEBAS DE RENDIMIENTO 4

2.1 ANÁLISIS DE PRUEBAS 4
2.1.1 Criterios de certificación de pruebas de rendimiento 4
2.1.2 Requisitos de rendimiento 5
2.2 ESPECIFICACIÓN DE PRUEBAS 7
2.3 EJECUCIÓN DE PRUEBAS 10
2.3.1 Configuración de pruebas 10
2.3.2 Ejecución de pruebas 11
2.3.2.1 Tipos de pruebas de rendimiento 11
2.3.2.1.1 Pruebas de carga 11
2.3.2.1.2 Pruebas de sobrecarga 11
2.3.2.1.3 Pruebas de escalabilidad 13
2.3.2.2 Tipos de pruebas adicionales 14
2.3.2.2.1 Pruebas de Estabilidad 14
2.3.2.2.2 Pruebas de Sobrecarga y Recuperación 14
2.3.2.2.3 Pruebas de Tolerancia a Fallos 15
2.3.2.2.4 Pruebas Ad-Hoc 15
2.3.3 Análisis de resultados 15
2.3.4 Registrar defectos 17

1 Introducción

1.1 Objeto

El presente documento establece un procedimiento que detalla los pasos que son necesarios para realizar pruebas de rendimiento de una aplicación en la DGT.

1.2 Alcance

La normativa que aquí se expone es **de obligado cumplimiento** para todas aquellas personas que vayan a realizar un proceso de pruebas de rendimiento sobre una aplicación en la DGT: La Oficina de Pruebas, Empresas externas de Desarrollo, etc…

2 Procedimiento de Pruebas de Rendimiento

2.1 Análisis de pruebas

Esta actividad consiste en realizar un análisis de la aplicación para conocer cuáles son los posibles escenarios de rendimiento que se desean simular en la fase de pruebas.

2.1.1 Criterios de certificación de pruebas de rendimiento

Se proporcionan criterios que permitan a los responsables del proyecto DGT establecer de una forma objetiva y normalizada la necesidad que tiene una aplicación de realizar pruebas de rendimiento.
Los criterios son los siguientes:

- Valoración subjetiva: considera adecuada la ejecución de pruebas de rendimiento
- Volumen de usuarios al que da/dará servicio
- Nivel de relación con otros sistemas de información o elementos tecnológicos
- Contempla procesos de una alta demanda de información
- Contempla procesos que habitual o puntualmente pueden requerir una importante demanda de información
- Proporciona datos a otros sistemas críticos con una importante demanda de información, tanto de volumen como de frecuencia
- Incorpora procesos complejos o costosos o que gestionan gran variedad de información
- Proporciona información que requiere ser servida de una manera muy ágil
- Proporciona información que en periodos concretos puede requerir disponibilidad ágil
- El sistema almacena información incremental que pudiera deteriorar su funcionamiento

2.1.2 Requisitos de rendimiento

Los **requisitos de rendimiento** son especificaciones relacionadas con la carga que se espera tenga que soportar el sistema en términos medibles. Por ejemplo, el número esperado de usuarios simultáneamente conectados, número de transacciones por segundo que deberá soportar el sistema, el % de transacciones que deben realizarse en menos de 1 segundo, etc.

Un requisito de rendimiento deberá definirse como un proceso de negocio, entendiendo como proceso de negocio, un conjunto de acciones o pasos de usuario realizados en la aplicación para lograr un objetivo de negocio.

Cada requisito de rendimiento debe relacionarse con un único proceso de negocio, es decir, la relación debe ser de 1:1.

Los requisitos de rendimiento deben estar enlazados a los requisitos funcionales con los cuales se encuentran relacionados.

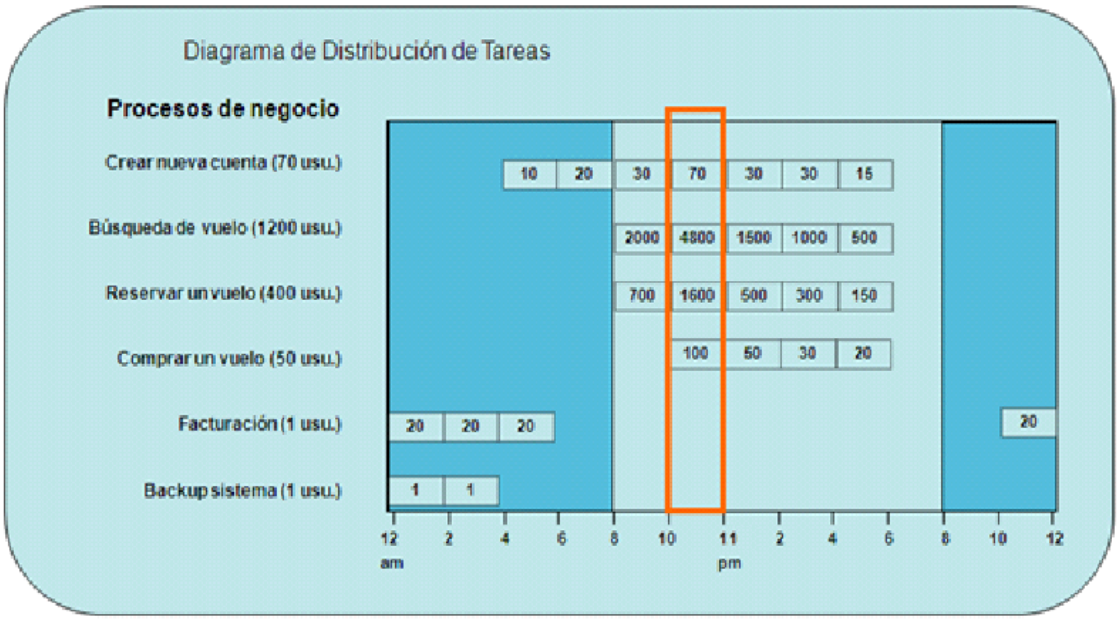
Para proyectos evolutivos y correctivos se deben identificar los requisitos de rendimiento que no han sufrido modificaciones con respecto a la versión anterior y que siguen aplicando en la actual, ya que estos requisitos determinarán los casos de prueba de regresión en la fase de especificación de pruebas.

Es importante identificar correctamente los requisitos de rendimiento/procesos de negocio que posteriormente formarán parte de las pruebas de rendimiento, de ello dependerá que de las pruebas que posteriormente se van a ejecutar, se obtengan datos significativos y que aporten valor o no. Para ello se proponen algunos criterios que se pueden utilizar para elegir unos u otros requisitos de una aplicación:

- Por volumen de usuarios. Seleccionando los procesos de negocio que son utilizados por un mayor número de usuarios.
- Por volumen de transacciones. Seleccionando los procesos de negocio que tienen un número más alto de peticiones alto.
- Por criticidad de los procesos. Seleccionando aquello procesos de negocio que son críticos en las operaciones diarias de la aplicación.

A continuación y a modo de ejemplo, se muestra una tabla con estadísticas de uso diario de una aplicación cualquiera. En la tabla se puede ver cada uno de los distintos procesos de negocio que proporciona la aplicación, con datos de uso estadísticos referentes al número de usuarios que han

usado un determinado proceso de negocio y al número de veces que ese proceso de negocio ha sido utilizado, agrupado por las distintas franjas horarias del un día de trabajo.



Claramente, el momento del día de mayor carga de trabajo es de 10 pm a 11 pm, por lo que es de esta franja de trabajo de la que hay que obtener las estadísticas de uso.
Es fácil de identificar en este caso, que los procesos más utilizados son Búsqueda de vuelo (utilizado por 1200 usuarios, 4800 veces en 1 hora) y Reservar un vuelo (utilizado por 400 usuario, 1600 veces en 1 hora). Por este motivo, estos dos procesos de negocio deberían ser seleccionados como requisitos de rendimiento, para su inclusión en unas pruebas de rendimiento.

Idealmente, estos datos se obtendrán de estadísticas de uso reales de la aplicación, en caso existir y estar disponibles. En caso de aplicaciones nuevas o de las que no se disponga de estas estadísticas, se hará un cálculo estimado en base al uso que se espera de la aplicación en producción, lo más realista posible.

Los requisitos deben tener definidas las siguientes características:

- Nombre.
- Descripción.

- Versión y ciclo de la aplicación.
- Tiempo medio de respuesta.
- Tiempo máximo de respuesta.
- Número de usuarios (máximos)
- Ritmo transaccional por hora.
- Tecnología. Web o webservices.
- Sistemas externos con los que interactúa.
- Trazabilidad con requisitos funcionales.

De todos estos valores y como ya se puede intuir por el ejemplo que se ha visto anteriormente en este apartado, hay 2 de ellos son fundamentales, ya que sirven como base para configurar el comportamiento del caso de prueba asociado a cada requisito durante la ejecución de una prueba de rendimiento. Uno de estos valores es el Ritmo transaccional por hora, que indica cuantas veces se va a ejecutar cada requisito/caso de prueba durante una hora (4800 veces en una hora, 3500, el valor que sea). El otro valor es el Número de usuarios, que indica el número de usuarios que se van a simular de forma concurrente por cada requisito/caso de prueba durante la ejecución de una prueba.

Por ejemplo, un requisito que tenga los siguientes valores: 4800 transacciones/hora y 1200 usuarios máximos, permitirán configurar una prueba de carga que cuando llegue a su punto álgido, 1200 usuarios concurrentes estarán trabajando a un ritmo de 4800 transacciones a la hora.

2.2 Especificación de pruebas

El **plan de pruebas de rendimiento** se define como un conjunto de casos de prueba que permiten verificar si el sistema cumple las necesidades de rendimiento que se le exigen en términos varios como tiempo de respuesta, número de usuarios, estabilidad, escalabilidad, etc.

Cada requisito de rendimiento debe relacionarse con un único proceso de negocio, por lo que debe tener asociado un único caso de prueba que complete un único camino funcional del requisito, es decir, la relación debe ser de 1:1.

A continuación se indican una serie de consideraciones a tener en cuenta en el diseño de los casos de prueba:

- **Transacciones de medida y verificaciones por cada paso.** Cada uno de los pasos de los casos de prueba deben tener definida una medida de tiempo de transacción y un chequeo para verificar que el paso se ha ejecutado correctamente. Como norma se debe aplicar que cada acción de usuario que devuelva una respuesta por parte de la aplicación debe medirse y verificarse.
- **Llamadas a sistemas externos.** Si el proceso de negocio realiza llamadas a sistemas externos se debe indicar en el caso de prueba el paso donde se realiza la llamada, así como detallar el sistema externo al cual se llama.
- **Chequeos dinámicos.** La verificación se debe realizar a chequeos dinámicos que aparezcan como resultado de la acción de usuario que se ha realizado previamente y nunca a chequeos estáticos como títulos de página, frames, tablas, etc, que siempre aparecen independiente de si el paso se ha realizado correcta o incorrectamente.
- **Parametrización.** Los campos que formen parte del proceso de negocio deben estar parametrizados y se debe proporcionar un juego de datos lo suficientemente amplio como para simular una situación real de uso de la aplicación. Es muy importante indicar si los datos utilizados por el caso de prueba son reutilizables, ya que esto hace que las pruebas de rendimiento sean fácilmente repetibles o no.

Los casos de prueba deberán ser escritos con el detalle suficiente para que cualquier persona sin ningún conocimiento funcional sobre la aplicación y sin utilizar la intuición, sea capaz de ejecutar las pruebas.

Todos los requisitos de rendimiento deben quedar cubiertos por casos de prueba y todos los casos de prueba estar enlazados a requisitos de rendimiento, de manera que la cobertura de pruebas sea del 100%.

Un caso de prueba debe tener la siguiente información:

- Descripción. Descripción sobre el aspecto de la funcionalidad que se está probando.
- Objetivo. Propósito del caso de prueba.
- Condiciones de ejecución. Condiciones que se deben cumplir antes de iniciar el caso de prueba.
- Tipo de usuario. Usuario que accede a la aplicación que puede ser persistente en sesión (no se cierra la sesión entre cada ejecución del proceso de negocio) o no persistente en sesión (cada vez que se ejecuta un proceso de negocio se realiza un login previo y una desconexión a la finalización del mismo).
- Datos de entrada. Lista de variables y valores usados para la ejecución de la prueba. Estos valores se escriben por línea de manera que cada línea se corresponde con una ejecución del caso de prueba.
- Diseño de pasos. Es la secuencia exacta de acciones necesarias para completar el caso de prueba. Cada paso debe tener la siguiente información:
    - Descripción. Son las acciones de usuario que se deben realizar sobre el aplicativo.
    - Resultado esperado. Son las verificaciones que se deben realizar tras la ejecución del paso para comprobar que el proceso de negocio se ha realizado correctamente.
    - Captura imagen. Se anexarán las capturas de pantalla necesarias para poder ejecutar correctamente el caso de prueba.
    - Nombre de la transacción de medida de tiempo.
    - Chequeo de texto a verificar.
    - Si aplica, detalle del sistema externo al cual se accede en el paso.

Para proyectos evolutivos y correctivos el plan de pruebas estará formado por:

- Casos de prueba nuevos. Se corresponden con los requisitos nuevos o modificados con respecto a la versión anterior.
- Casos de prueba de regresión. Se corresponden con los requisitos no modificados con respecto a la versión anterior.

2.3 Ejecución de pruebas

En esta fase se realiza la ejecución de los diferentes tipos de pruebas de rendimiento necesarias para alcanzar el objetivo de rendimiento de la aplicación.

2.3.1 Configuración de pruebas

En esta fase, la Oficina de Pruebas realiza la configuración de las de pruebas de rendimiento de manera que sean lo más cercanas a la realidad.

El diseño y creación de las diferentes baterías de pruebas de rendimiento comprende la configuración de las siguientes entidades:

- Procesos de negocio.
- Número de usuarios de cada proceso de negocio.
- Duración de la prueba.
- Rampa de subida de los usuarios.
- Rampa de bajada de los usuarios.
- Ritmo transaccional de cada proceso de negocio.
- Configuración del log.
- Configuración del proxy.
- Configuración de la memoria cache.

- Ancho de banda.
- Medida de las transacciones.
- Maquinas generadoras de carga.
- Configuración de métricas de rendimiento.

La configuración de la prueba de rendimiento se realiza con la información disponible en la fase de requisitos y la información aportada por el jefe de proyecto DGT de la aplicación, a través de la Plantilla Solicitud Lanzamiento Pruebas Rendimiento, en el momento de la solicitud de las pruebas.

2.3.2 Ejecución de pruebas

En esta fase la Oficina de Pruebas realiza la ejecución de la prueba/s de rendimiento configuradas.

La ejecución de una prueba de rendimiento debe estar asociada a una revisión de la aplicación desplegada en el entorno de preproducción, y debe tener asociado un juego de datos de entrada con un volumen y variedad adecuado para que la ejecución de pruebas sea lo más próximo a la realidad.

Durante la ejecución de las pruebas de rendimiento el Departamento de Sistemas realizará la monitorización de la aplicación en la parte servidora.

2.3.2.1 Tipos de pruebas de rendimiento

2.3.2.1.1 Pruebas de carga

El objetivo de esta prueba es verificar que la aplicación es capaz de soportar la carga que se espera sin mostrar ninguna degradación.

La prueba de carga se diseñará con una duración de entre 2 o 3 horas soportando el número de usuarios concurrentes y ritmo transaccional definidos en los requisitos de rendimiento.

2.3.2.1.2 Pruebas de sobrecarga

El objetivo de la prueba de sobrecarga es comprobar que la aplicación es capaz de responder ante un aumento de carga superior al definido en los requisitos de rendimiento comprobando que su comportamiento no se ve afectado.

La prueba de sobrecarga se diseñará con una duración de entre 2 o 3 horas con igual configuración que la de carga pero con un aumento del 20% en el ritmo transaccional. El número de usuarios podrá adaptarse al alza para alcanzar dicho ritmo transaccional.

2.3.2.1.3 Pruebas de escalabilidad

Estas pruebas se realizan insertando de una forma progresiva bloques de usuarios al sistema hasta encontrar los cuellos de botella que impiden la escalabilidad de la aplicación, y nos proporcionen su límite en número de usuarios o transacciones. El escalado disminuye el rendimiento suele ser el resultado de un cuello de botella en uno o más recursos. Con este tipo de pruebas se permite el establecimiento de los umbrales de referencia (en volumen de usuarios), a partir de los cuales la aplicación tendría una degradación del servicio y a partir de la que redundaría en una pérdida del servicio.

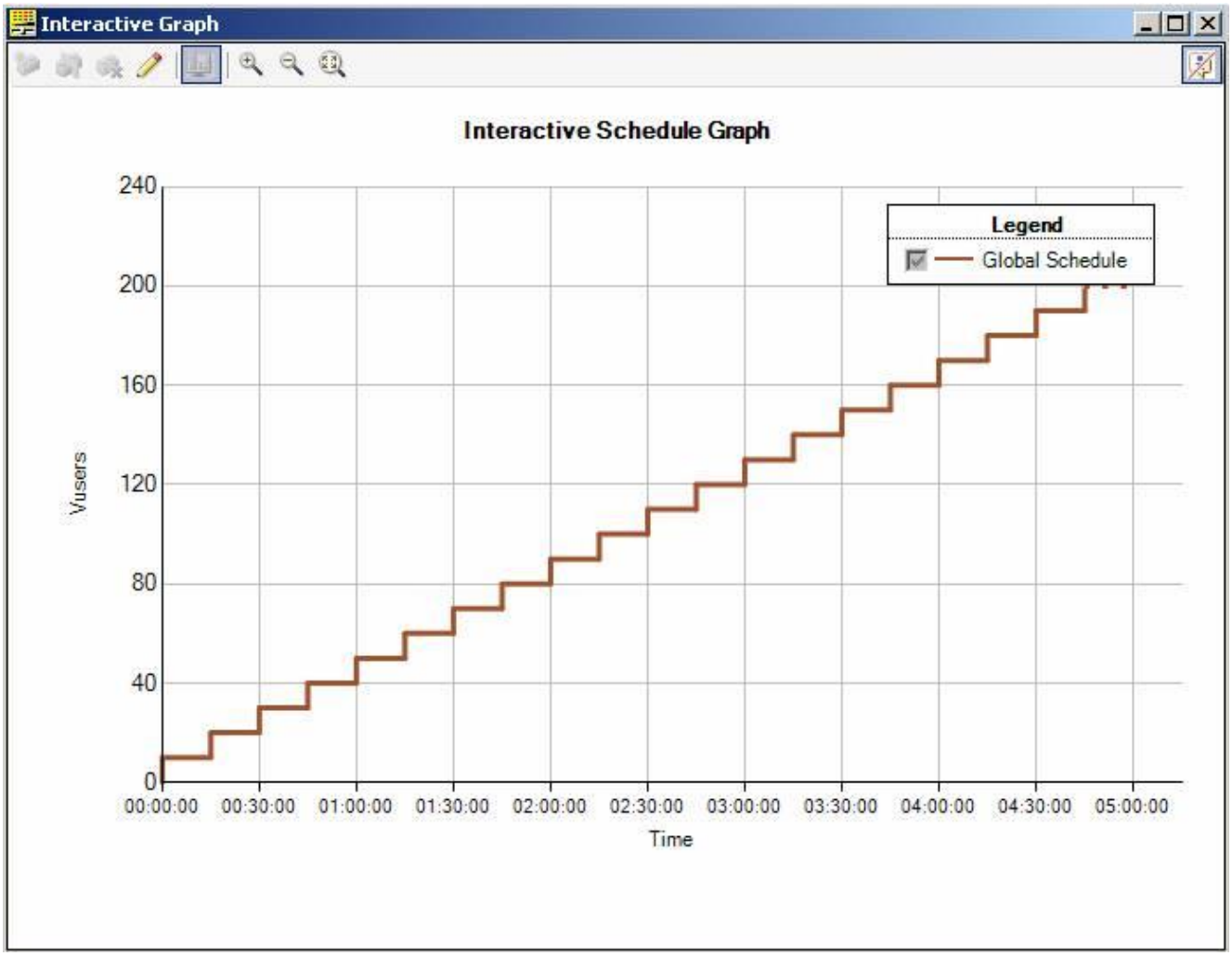


Ilustración 1. Pruebas de escalabilidad

La rampa de usuarios se realiza mediante la inserción de bloques de usuarios (escalones) de manera que se debe asegurar que en cada bloque cada usuario ejecute al menos de tres a cuatro veces el proceso de negocio completo antes de la entrada de un nuevo bloque. Con cada escalón lo que se

pretende es incorporar nueva carga al sistema, cuando ha conseguido estabilizarse entonces se introduce de nuevo otro bloque de usuarios, y así hasta alcanzar la saturación del sistema.

2.3.2.2 Tipos de pruebas adicionales

Por necesidades específicas de cada aplicación se pueden realizar otro tipo de pruebas de rendimiento. Los tipos y objetivos de estas pruebas adicionales se detallan a continuación:

2.3.2.2.1 Pruebas de Estabilidad

Estas pruebas tienen como finalidad comprobar el comportamiento de la aplicación a lo largo de un período prolongado de tiempo, intentando simular un contexto de trabajo real.
Las características de dichas pruebas serán:

- Duración: mínimo 6 horas.
- Volumen de usuarios: igual que en carga.
- Tiempos de espera: simulando la actividad real.

Aunque estos parámetros se ajustarán a las características específicas de cada aplicación.

2.3.2.2.2 Pruebas de Sobrecarga y Recuperación

Estas pruebas se ejecutan con el objetivo de comprobar cómo se recupera una aplicación después de una sobrecarga inesperada de trabajo. Durante una prueba con un nivel de carga de trabajo esperado, lo habitual es aumentar la carga hasta saturar el sistema para inmediatamente después volver a unos niveles de carga de trabajo normal, comprobando la capacidad de recuperación del sistema.

Estas pruebas se realizan insertando la carga en el sistema en forma de “picos” que se irán lanzando en distintos momentos de la prueba y que permitirán comprender el comportamiento de la aplicación ante cambios bruscos de carga. La prueba de picos, como el nombre sugiere, trata de observar el comportamiento del sistema variando el número de usuarios, tanto cuando bajan, como cuando tiene cambios drásticos en su carga.

2.3.2.2.3 Pruebas de Tolerancia a Fallos

Estas pruebas consisten en comprobar que el sistema continúa funcionando aunque se produzcan fallos, como la caída de uno de los nodos del cluster.

2.3.2.2.4 Pruebas Ad-Hoc

Siendo las pruebas anteriormente los tipos más frecuentes, pueden diseñarse y configurarse pruebas a medida que ayuden a solventar problemáticas particulares que requieran un diseño de pruebas más específico y personalizado.

2.3.3 Análisis de resultados

Una vez ejecutada la prueba se realizará un análisis de resultados por parte de la Oficina de Pruebas y el Departamento de Sistemas.
Para la realización de dicho análisis se tendrá en cuenta entre otras las siguientes variables:

- Granularidad. La granularidad nos permite modificar el número de puntos de datos a visualizar en el gráfico. Para el análisis de resultados se utilizará una granularidad de entre 0 y 120 segundos.
- Duración. Para el análisis de resultados se podrá modificar la escala de tiempo en los gráficos para analizar el comportamiento en un determinado tramo.

Los criterios utilizados por la Oficina de Pruebas para dar como correctas o incorrectas las pruebas de rendimiento son los siguientes:

- Los tiempos de respuesta medios obtenidos no deben superar los tiempos medios definidos en los requisitos de rendimiento.
- Los tiempos de respuesta obtenidos en el percentil 98 no deben superar los tiempos máximos definidos en los requisitos de rendimiento.
- El número de transacciones fallidas no puede superar el 2% tanto para cada una de las transacciones individuales como para el total de transacciones.
- El ritmo transaccional obtenido no puede estar por debajo del ritmo transaccional indicado en requisitos permitiendo un margen del 5%.

- Comportamiento estable de la aplicación, es decir, no debiendo existir picos en los tiempos de respuesta que se produzcan reiteradamente cada cierto tiempo. Ver imagen adjunta.

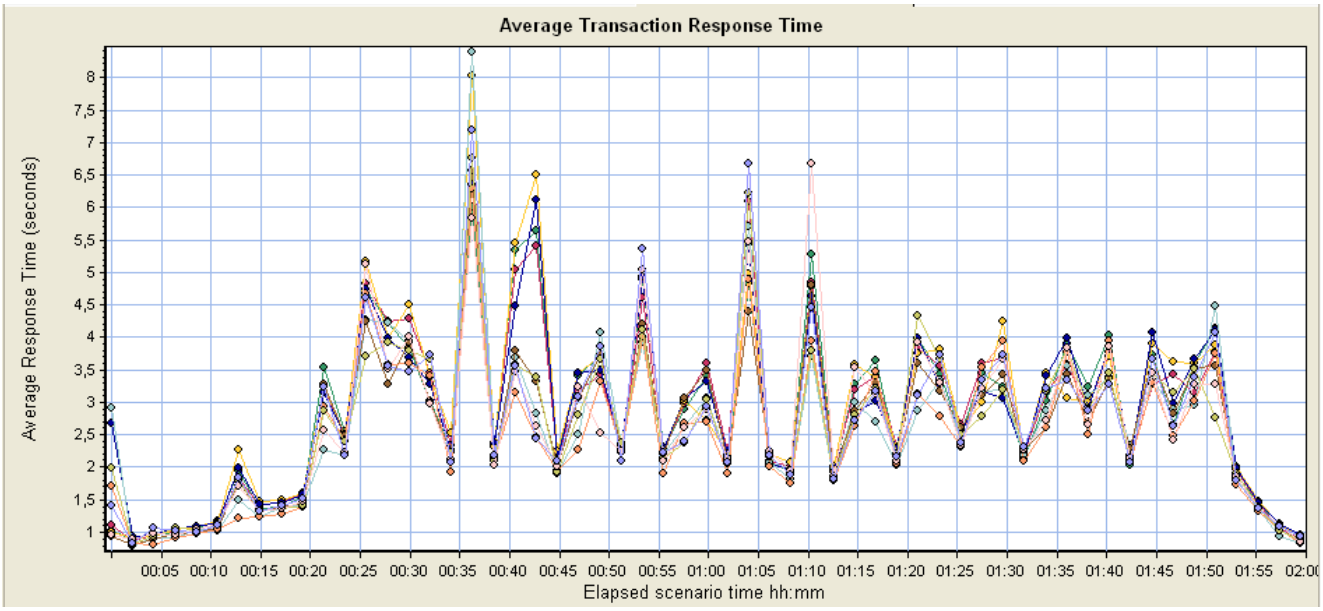


Ilustración 2. Picos en pruebas de carga

- No degradación de los tiempos de respuesta, es decir, no debe producirse un aumento progresivo o disminución en los tiempos de respuesta. Ver imagen adjunta.

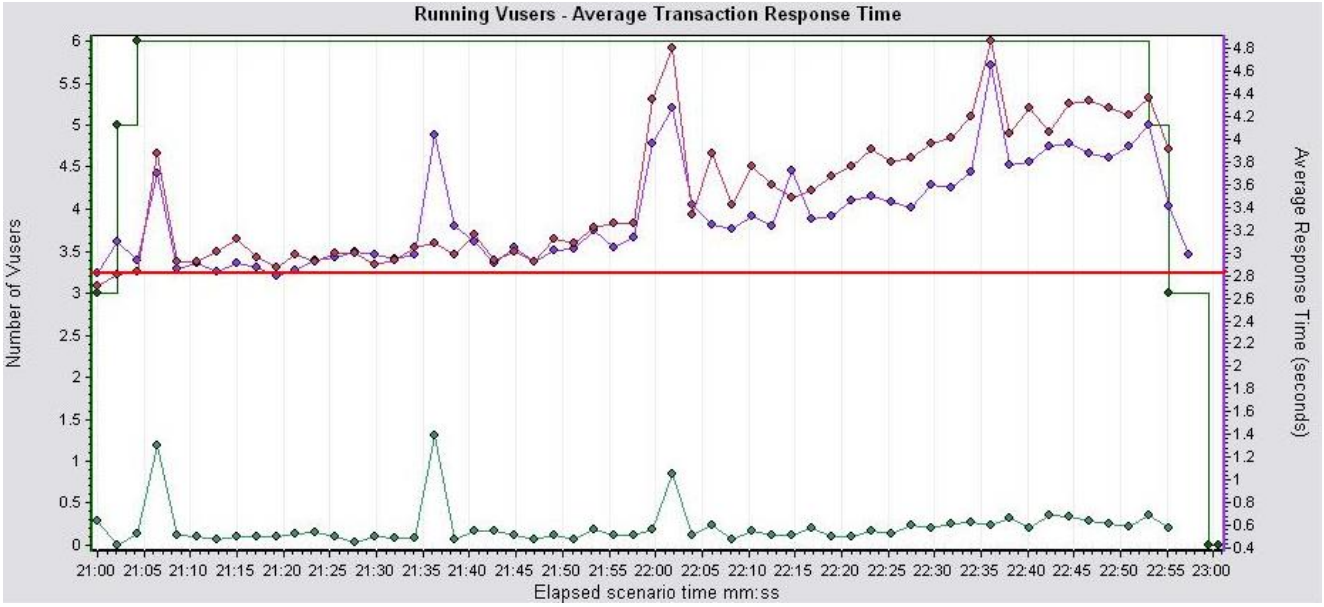


Ilustración 3. Degradación en tiempos de respuesta

- La gráfica del total de transacciones por segundo debe mantenerse constante en el tiempo.
- La gráfica del número de usuarios y throughput debe ir en paralelo.
- La gráfica del número de usuarios y Hits debe ir en paralelo.
- La gráfica Connections per Second y Hits no debe ir en paralelo, el número de conexiones TCP/IP debe ser muy inferior al de hits.

Además de los indicadores anteriormente comentados, existen otros dependientes de cada aplicación, tipo de prueba, tecnología, etc. y que son necesarios analizar a la hora de interpretar los resultados.
Para las pruebas de regresión además se contrastarán los resultados actuales con los obtenidos en la versión anterior, con el objetivo de informar de las posibles variaciones en los indicadores de ambas pruebas.
El Departamento de Sistemas emite un informe de monitorización donde indica el resultado de la prueba de carga desde la parte servidora, y en caso de ser incorrecto detalla los errores encontrados.

2.3.4 Registrar defectos

Una vez analizados los resultados de la ejecución de pruebas se registrarán los errores detectados siguiendo el procedimiento indicado en el documento “Anexo 03.00 Proceso de Pruebas”.