



Universitat
de les Illes Balears

Evaluación de Comportamiento de Sistemas Informáticos

Práctica 5 - Evaluación y modelado del rendimiento de los sistemas informáticos

Alberto Pérez Ancín

Professores:

Dr. Carlos Juíz García

Dra. Belén Bermejo González



APLICACIÓN PRÁCTICA

1.1 Objetivo

El objetivo principal es la evaluación práctica, real y completa de un sistema informático. Para ello se aplicarán todos los conceptos, metodologías y técnicas vistos a lo largo de todo el curso. Se tratarán desde los aspectos relacionados con la monitorización y el benchmarking, pasando por el modelado y llegando finalmente a la predicción de la carga.

Un ejemplo real del camino expuesto se ve en el artículo “The Performance Evaluation Journey of a Flight Seats Availability Service: A Real-world Business Case Study of Transactional Workload Running in Virtual Machines”, el cual se recomienda encarecidamente que se tenga a disposición y se comprenda en su totalidad.

1.2 Quinta parte

Para la realización de esta parte volveremos a utilizar la herramienta QNAP y el modelo.qnp, además de programar en el lenguaje que os sea más cómodo y práctico.

Para esta quinta parte realizaremos un modelo que resuelva el problema 5.1 del libro de referencia del curso. Para ello deberás seguir los siguientes pasos:

1. Editad el modelo.qnp con el cuaderno de notas (NO con el wordpad) o cualquier otro **editor que no formatee el texto, ni incluya caracteres invisibles. Salvad el nuevo fichero en otro .qnp** (en formato cuaderno de notas y **en el mismo directorio que el programa ejecutable**).
2. Cambiad el modelo y los parámetros del mismo para que al ejecutar vuestro nuevo modelo en QNAP la tabla de resultados demuestre que es lo mismo que al realizar el problema (ver solución en la tabla 5.2 en el libro).

- a) **Programad el cálculo de las demandas de los 2 dispositivos, y la demanda total (D) y su impresión. ¿Cuál es el cuello de botella (Db)? ¿Cuál es el punto de saturación (N*)? Imprimid todas esas variables.**

En este programa, que es una modificación del modelo.qnp y añade los datos del ejercicio 5.1 al cual se hace referencia en el enunciado de la práctica. Además, sirve para calcular la demanda de la CPU, la demanda del disco y la demanda total, se utilizan las funciones MSERVICE y MRESPONSE, que calculan el tiempo medio de servicio y el tiempo de respuesta medio, respectivamente. Luego, se determina el cuello de botella y el punto de saturación en función de cuál estación tenga la mayor demanda.

```
/DECLARE/ QUEUE CPU,DISC,TERMINAL;
          REAL PROB1(2)=(7.,8.);
          REAL TTR1,TOTAL,WORK;
          INTEGER I,N1;
          REAL D_CPU, D_DISC, D_TOTAL, N_STAR, T_CPU, T_DISC;
          STRING DB;
/STATION/ NAME=CPU;
&        SCHED=PS;
          SERVICE=EXP(0.03);
          TRANSIT=DISC,PROB1(2),TERMINAL,1;
/STATION/ NAME=DISC;
          SERVICE=EXP(0.1);
          TRANSIT=CPU,PROB1(1);
/STATION/ NAME=TERMINAL;
          TYPE=INFINITE;
          INIT=N1;
          SERVICE=EXP(8.);
          TRANSIT=CPU;
/CONTROL/ CLASS=ALL QUEUE;
/EXEC/   FOR N1:=10 STEP 10 UNTIL 10 DO
          BEGIN
            PRINT("");
            PRINT("NUMERO DE USUARIOS=",N1);
            SOLVE;
            PRINT("");
            D_CPU := MSERVICE(CPU) * PROB1(2);
            D_DISC := MSERVICE(DISC) * PROB1(1);
            D_TOTAL := D_CPU + D_DISC;
            IF D_CPU > D_DISC
            THEN BEGIN
              N_STAR := (D_TOTAL + MRESPONSE(TERMINAL)) / (D_CPU);
              DB := "CPU";
            END
            ELSE BEGIN
              N_STAR := (D_TOTAL + MRESPONSE(TERMINAL)) / (D_DISC);
              DB := "DISC";
            END;
          END;
```

Figura 1.1: Parte de código del modeloApartadoA.qnp

Los resultados obtenidos por la ejecución del modelo ApartadoA.qnp son los siguientes:

```
=====
                        APARTADO A
=====
NUMERO DE USUARIOS=      10
DEMANDA CPU=  0.2400      s
DEMANDA DISCO=  0.7000      s
DEMANDA TOTAL=  0.9400      s
CUELLO DE BOTELLA=DISC
PUNTO DE SATURACION=  12.77      trabajos
```

Figura 1.2: Resultado de ejecutar el modelo ApartadoA.qnp

Estos resultados indican que en el sistema, el disco es el cuello de botella, lo que significa que limita la capacidad del sistema para manejar más usuarios. Además, el punto de saturación indica que el sistema puede manejar hasta 12.77 trabajos antes de que se produzca una congestión.

- b) **Programad el cálculo del tiempo de respuesta del sistema (R) y el tiempo TOTAL (R+Z), así como el número de usuarios trabajando y reflexionando (imprimid los valores).**

En este programa, que es una modificación del modelo.qnp y añade también el modelo ApartadoA.qnp, hay además funcionalidades para calcular el tiempo de respuesta y la cantidad de usuarios.

```
PRINT("=====");
PRINT("                        APARTADO B                        ");
PRINT("=====");
R:=(PROB1(2)*MRESPONSE(CPU))+(PROB1(1)*MRESPONSE(DISC));
PRINT("EL TIEMPO DE RESPUESTA DEL SISTEMA (R)=",R:10:4,"s");
TOTAL:=(R+MRESPONSE(TERMIAL));
PRINT("EL TIEMPO TOTAL (R+Z)=",TOTAL:10:4,"s");
PRINT("");
REFLEX:=(MTHRUPUT(TERMIAL))*(MRESPONSE(TERMIAL));
PRINT("NÚMERO DE USUARIOS REFLEXIONANDO=",REFLEX:10:4);
WORK:=(N1-REFLEX);
PRINT("NÚMERO DE USUARIOS TRABAJANDO=",WORK:10:4);
```

Figura 1.3: Parte de código del modelo ApartadoB.qnp

Inicialmente, se calculan los tiempos de respuesta y luego los usuarios, que pertenecen al "APARTADO B". Entonces, se realizan cálculos mediante funciones y operaciones para obtener los valores de rendimiento del sistema, como el tiempo de respuesta y el número de usuarios trabajando y reflexionando.

Durante este código, se utilizaron variables para almacenar los resultados de los cálculos y se emplearon funciones como son MTHRUPUT y MRESPONSE, para obtener los valores necesarios para este problema. Estas funciones representan la productividad media y el tiempo medio de respuesta del sistema.

```
=====
                        APARTADO B
=====
EL TIEMPO DE RESPUESTA DEL SISTEMA (R)=    2.1541s
EL TIEMPO TOTAL (R+Z)=    10.1541s

NÚMERO DE USUARIOS REFLEXIONANDO=    7.6509
NÚMERO DE USUARIOS TRABAJANDO=    2.3491
```

Figura 1.4: Resultado de ejecutar el modeloApartadoB.qnp

- c) **Volved a vuestro modelo original y cread otro disco gemelo al original (7 visitas) y equilibrad las cargas, ¿qué variaciones se observan en los cálculos?**

Primero creo otro disco gemelo al original y equilibrio las cargas, entonces agrego un nuevo dispositivo de almacenamiento al modelo. A continuación, ajusto las razones de visita para distribuir las visitas entre los dos discos.

```
=====
                        APARTADO A
=====
NUMERO DE USUARIOS=    10
DEMANDA CPU=    0.2400    s
DEMANDA DISCO1=    0.3500    s
DEMANDA DISCO2=    0.3500    s
DEMANDA TOTAL=    0.9400    s
CUELLO DE BOTELLA=DISC2
PUNTO DE SATURACION=    25.54    trabajos

=====
                        APARTADO B
=====
EL TIEMPO DE RESPUESTA DEL SISTEMA (R)=    1.2333s
EL TIEMPO TOTAL (R+Z)=    9.2333s

NÚMERO DE USUARIOS REFLEXIONANDO=    7.6509
NÚMERO DE USUARIOS TRABAJANDO=    2.3491
```

Figura 1.5: Resultado de ejecutar el modeloApartadoC.qnp

Al agregar un nuevo disco y equilibrar las cargas, se observan importantes variaciones en los cálculos. La demanda total del sistema ahora se distribuye entre tres dispositivos en lugar de dos, permitiendo una distribución más eficiente de los recursos. Al reducir la demanda individual de cada disco a la mitad, gracias a la distribución equitativa de las visitas entre ambos discos, se alivia el cuello de botella en el sistema. De esta manera, cada disco puede tardar más en sobrecargarse.

Esta reorganización en la distribución de las cargas resulta en una mejora significativa en el tiempo de respuesta del sistema. Con tiempos de espera más cortos, pasando a valer el tiempo de respuesta 1,2333s. Además, el punto de saturación, que indica el número máximo de trabajos que el sistema puede manejar antes de que su rendimiento se vea afectado, también mejora significativamente. Pasa de 12,77 trabajos en el modelo original a 25,54 trabajos en el modelo con el nuevo disco y cargas equilibradas.

- d) **Volved a vuestro modelo original e iterad el modelo hasta 30 usuarios con saltos de 1 y construid una tabla .xls o similar y dos gráficas con líneas, en la que se vea la variación del tiempo de respuesta (R) y la productividad del sistema (X) con el número de usuarios incremental.**

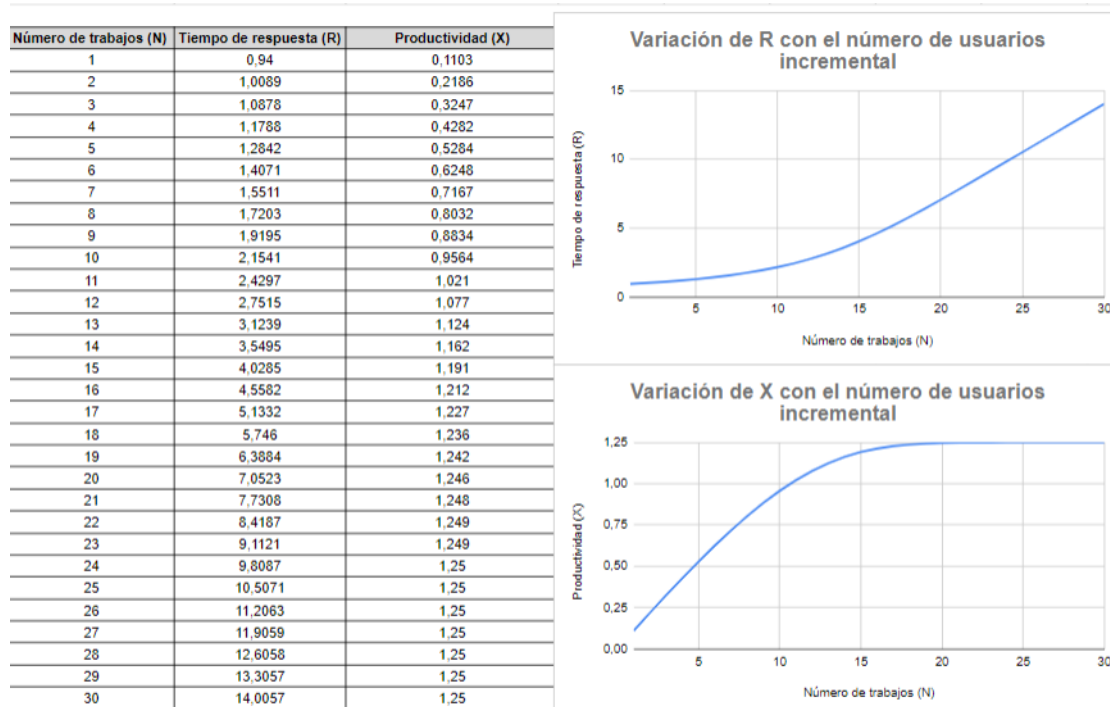


Figura 1.6: Resultado de ejecutar el modeloApartadoD.qnp

1. APLICACIÓN PRÁCTICA

Esta tabla representa una relación entre el número de trabajos (N), el tiempo de respuesta (R) y la productividad (X). A medida que aumenta el número de trabajos, el tiempo de respuesta también aumenta y la productividad disminuye.

Podemos observar que a medida que el número de trabajos aumenta, el tiempo de respuesta y la productividad cambian de manera no lineal. Al principio, el tiempo de respuesta aumenta de manera más lenta, mientras que la productividad disminuye rápidamente. Sin embargo, a medida que el número de trabajos aumenta aún más, el tiempo de respuesta aumenta rápidamente mientras que la productividad disminuye a una tasa más lenta.

Otra observación importante es que después del trabajo número 10, la tasa de cambio de la productividad disminuye significativamente y se mantiene estable en 1.25. Esto sugiere que hay un límite en la productividad que se puede lograr con la cantidad de recursos disponibles.

- e) **Representad las 4 asíntotas del modelo original en sendas gráficas y el N^* (ver libro páginas 140 -141 o transparencias del tema 5).**

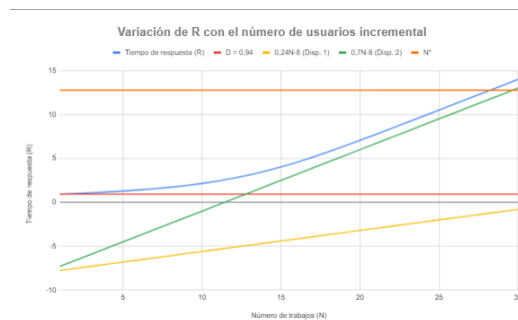


Figura 1.7: Representación de las 4 asíntotas del modelo original en sendas gráficas y el N^*

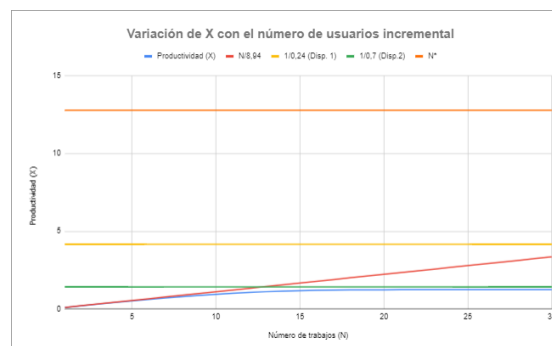


Figura 1.8: Representación de las 4 asíntotas del modelo original en sendas gráficas y el N^*

f) Realizad lo mismo que en el apartado d y e con el modelo de dos discos gemelos (apartado c).

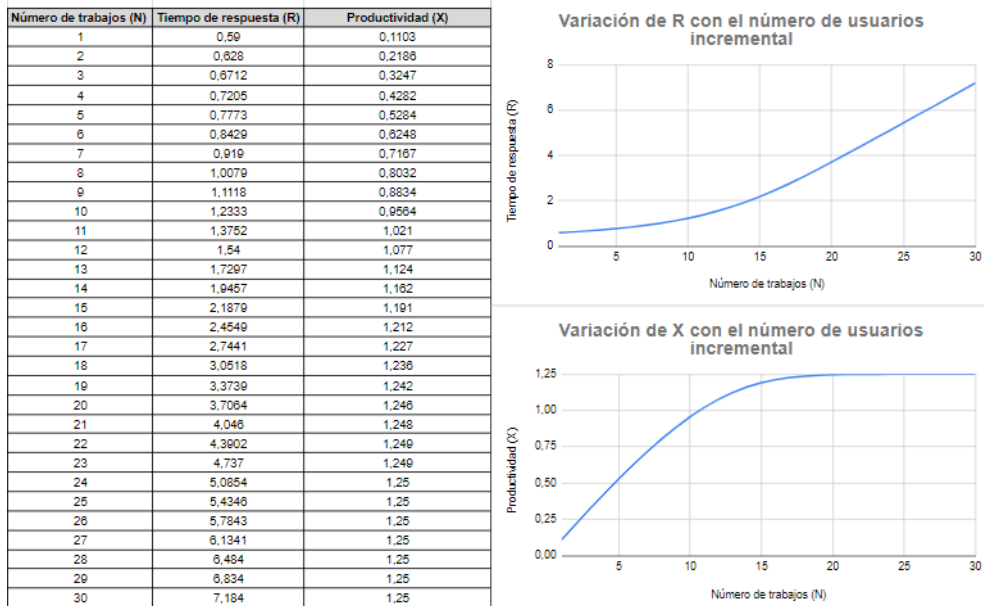


Figura 1.9: Resultado de ejecutar el modeloApartadoC.qnp

Esta tabla representa una relación entre el número de trabajos (N), el tiempo de respuesta (R) y la productividad (X). A medida que aumenta el número de trabajos, el tiempo de respuesta aumenta y la productividad disminuye.

En comparación con la tabla anterior, podemos observar que en esta tabla el tiempo de respuesta y la productividad cambian de manera más lineal a medida que aumenta el número de trabajos. Además, la tasa de cambio en el tiempo de respuesta y la productividad es mucho más lenta en comparación con la tabla anterior.

Otra observación importante es que la productividad nunca llega a ser igual a cero, incluso cuando el tiempo de respuesta es mucho más alto. Esto sugiere que hay un punto óptimo en el que la productividad se estabiliza y el tiempo de respuesta sigue siendo aceptable.

1. APLICACIÓN PRÁCTICA

$$R_{opt} = \max\{D, D_b x N - Z\} = \max\{0, 94, 0, 35xN - 8\}$$

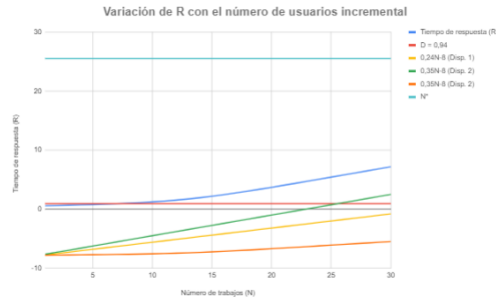


Figura 1.10: Representación de las 4 asíntotas del modelo original en sendas gráficas y el N^*

$$X_{opt} = \min\left\{\frac{N}{D+Z}, \frac{1}{D_b}\right\} = \min\left\{\frac{N}{8,94}, 2,86\right\}$$

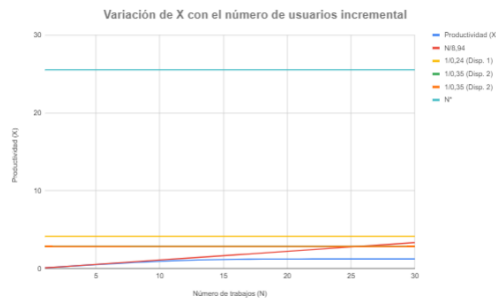


Figura 1.11: Representación de las 4 asíntotas del modelo original en sendas gráficas y el N^*

3. **Programad un pequeño algoritmo que resuelva con MVA** idénticamente al problema y el modelo en QNAP del apartado 2 (ver libro página 136 o transparencias del tema 5). Comprobad que los resultados de vuestro programa salen iguales al ejercicio 5.1 del libro y vuestro modelo en QNAP hasta el apartado b.

Entregables: fichero fuente impreso en .pdf del programa (apartado 3), fichero de resultados final .qnp (apartado 2d), fichero.xls o similar con sus 4 gráficas de líneas (apartados 2d a 2f) y un documento conciso que explique lo observado (apartados 2a a 2f).