

Infraestructura computacional

Infraestructura de TI: planeamiento de la capacidad

Planeamiento de la Capacidad en TI

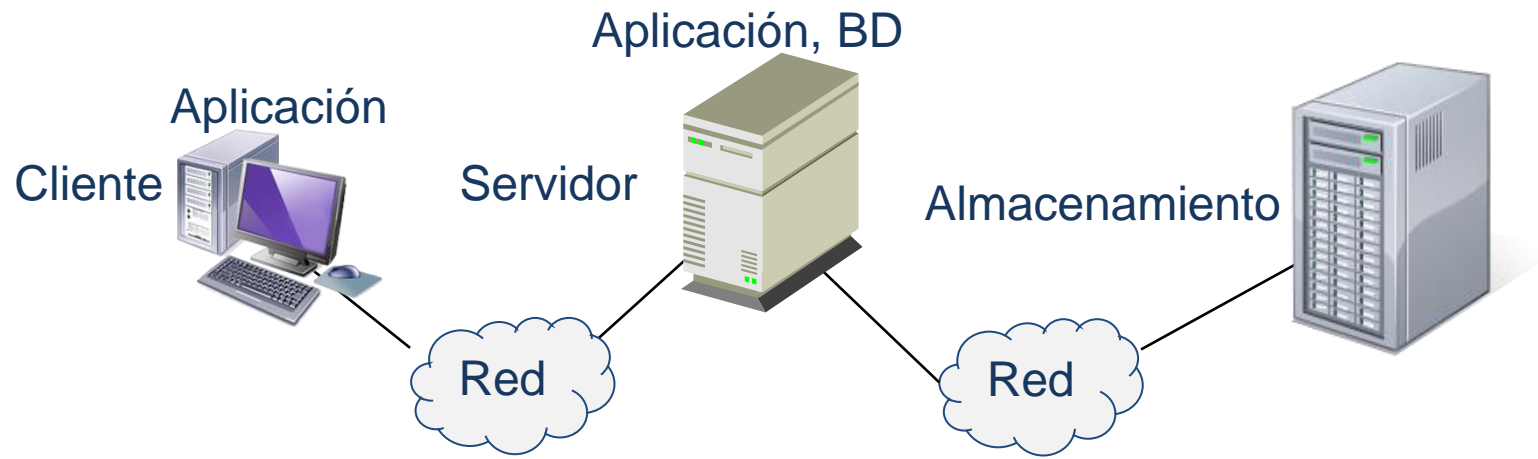
- Determinar la **infraestructura de TI** (procesamiento, almacenamiento, comunicaciones, etc,) necesaria para prestar **eficientemente** un servicio **a lo largo del tiempo** a un **cierto nivel de efectividad** considerado satisfactorio
 - ❶ Definir el nivel de servicio requerido: precisar los servicios, sus alcances y características
 - ❷ Analizar la capacidad actual
 - ❸ Hacer prospectiva

Planeamiento de la Capacidad en TI

- ① Definir los requerimientos: caracterizar el servicio prestado y determinar las expectativas de los usuarios
 - Definir la carga de trabajo (*workload*): clasificación del trabajo realizado por el sistema según un cierto criterio :
 - quién lo hace
 - qué hace
 - cómo lo hace, etc.
 - Definir la unidad de carga
 - En términos del negocio
 - No son los recursos utilizados
 - Establecer los niveles de servicio
 - Acuerdo entre proveedor y cliente para definir “aceptable”
 - En términos del cliente (puede ser en términos de la carga de trabajo)
 - Típicamente tiempos de respuesta, o productividad

Planeamiento de la Capacidad en TI

- ① Determinar requerimientos:
Componentes de una solución informática



Planeamiento de la Capacidad en TI

- ① Determinar requerimientos:
Tipos de requerimientos:
 - En cuanto al procesamiento
 - Capacidad
 - Escalabilidad
 - Desempeño
 - En cuanto a la seguridad:
 - Confidencialidad
 - Integridad
 - Disponibilidad

Planeamiento de la Capacidad en TI

- 1 Determinar requerimientos:

	Aplicación	Servidor	BD	Almacenamiento	Red
Capacidad					
Escalabilidad					
Desempeño					
Confidencialidad					
Integridad					
Disponibilidad					

Planeamiento de la Capacidad en TI

- ① Determinar requerimientos:

Tipos de requerimientos:

- Capacidad: carga que debe soportar el sistema. Volumen de información procesada, almacenada o enviada.
- Escalabilidad: capacidad de asignar recursos adicionales (procesamiento, almacenamiento, comunicación) perturbando al mínimo el funcionamiento de la infraestructura
- Desempeño: tiempo de respuesta. Velocidad con la que se desempeña la infraestructura
- Disponibilidad: accesibilidad ininterrumpida a la información (o con mínimos trastornos)

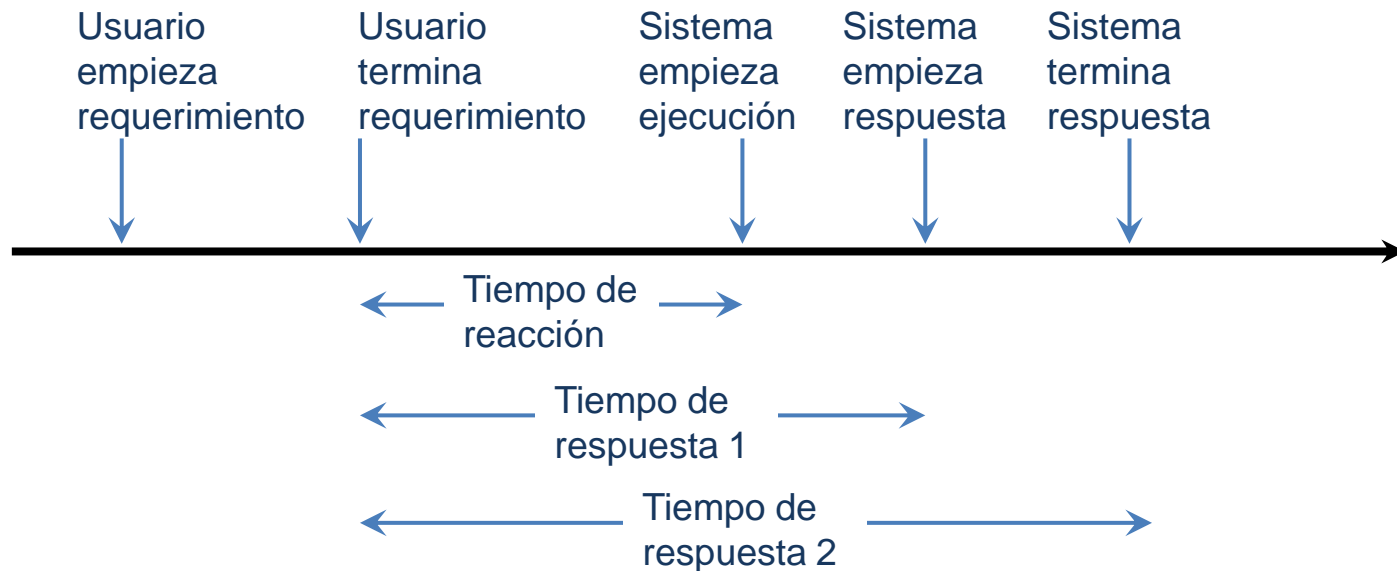
Planeamiento de la Capacidad en TI

- ① Determinar requerimientos:
Métricas :
 - Reactividad - Productividad
 - Particular – Global
 - Métricas usuales:
 - Tiempo de respuesta: tiempo desde que se emite el requerimiento hasta que se recibe la respuesta
 - Tiempo de retorno (turnaround): tiempo desde que se emite una tarea (*batch*) hasta el retorno de la salida completa
 - Tiempo de reacción: tiempo desde que se emite el requerimiento hasta que empieza su ejecución
 - Rendimiento: rata a la cual se procesan los requerimientos (requerimientos / tiempo)

Planeamiento de la Capacidad en TI

- 1 Determinar requerimientos:

Tiempo de respuesta:



Planeamiento de la Capacidad en TI

- ① Determinar requerimientos:
Medida de la capacidad de procesamiento:
 - Productividad = Número de tareas despachadas por unidad de tiempo
 - Carga = Número tareas activas al mismo tiempo
 - Tarea: transacción, job, etc.
 - Unidad de tiempo: depende de lo demorado de la tarea (segundo, minuto, etc.)

Planeamiento de la Capacidad en TI

- ① Determinar requerimientos:
Medida de la capacidad de almacenamiento:
 - Datos operativos
 - Archivo
 - Lapso cubierto (*desde - hasta*)
 - Ciclo de vida de la información
 - Niveles de archivo
 - Backup
 - Frecuencia
 - Número de versiones

Planeamiento de la Capacidad en TI

- ① Determinar requerimientos:
Escalabilidad:
 - Crecimiento sostenido
 - Variabilidad en la demanda (dinámica)
 - Picos eventuales
 - Predecibles
 - No predecibles

Planeamiento de la Capacidad en TI

- ① Determinar requerimientos:
Disponibilidad:
 - Medida de la frecuencia o del tiempo que un servicio o componente está disponible
 - Se suele medir como porcentaje de tiempo en que un servicio o componente del sistema está disponible:
 - $\text{disponibilidad} = T_d / (T_d + T_{nd})$
 T_d = Tiempo disponible
 T_{nd} = Tiempo no disponible

Planeamiento de la Capacidad en TI

- ① Determinar requerimientos:
Disponibilidad:
 - Confiabilidad (Reliability): probabilidad de que un sistema se mantenga funcionando continuamente. Se suele expresar por:
 - $MTBF = \text{Mean Time Between Failures}$
 - Facilidad de mantenimiento (Serviceability): medida de la facilidad para reparar un servicio. Se puede expresar por:
 - $MTTR = \text{Mean Time To Repair}$
 - Disponibilidad (Availability):
 - $\text{Disponibilidad} = MTBF / (MTBF + MTTR)$

Planeamiento de la Capacidad en TI

- 1 Determinar requerimientos:
Porcentaje de disponibilidad (“número de nueves”):
 - Disponibilidad anual:

Disponibilidad	Tiempo de indisponibilidad
90.0 %	36 días 12 horas
95.0 %	18 días 6 horas
99.0 %	87 horas 36 minutos
99.50 %	43 horas 48 minutos
99.90 %	8 horas, 45 minutos, 36 segundos
99.99 %	52 minutos 33 segundos
99.999 %	5 minutos 15 segundos
99.999 %	32 segundos

Planeamiento de la Capacidad en TI

- ② Analizar la capacidad actual:
 - Mediciones:
 - Carga de trabajo (*qué*)
 - Niveles de servicio (*cómo*)
 - Consumo de recursos (*con qué*):
 - Determinar nivel de agregación del consumo de recursos: global, local
 - Discriminar los componentes del tiempo de respuesta: procesamiento (¿dónde?), tránsito, entrada/salida
 - Determinar frecuencia de muestreo y periodo de observación
 - Tendencia (histórico)
 - Estacionalidad
 - Picos

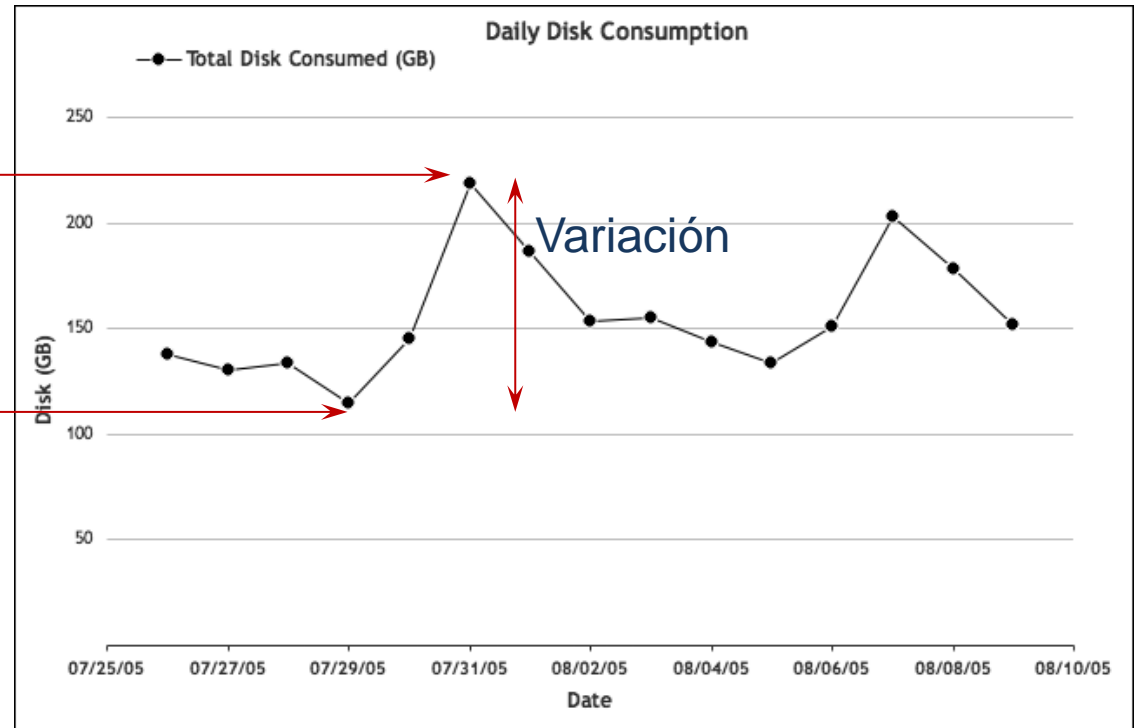
Planeamiento de la Capacidad en TI

- ② Analizar la capacidad actual

Mediciones:

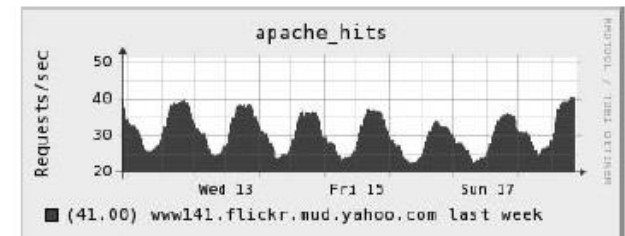
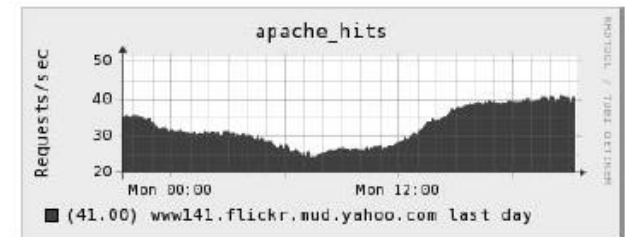
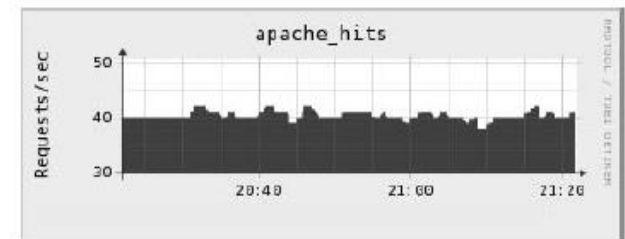
Máximo (domingos)

Mínimo (viernes)



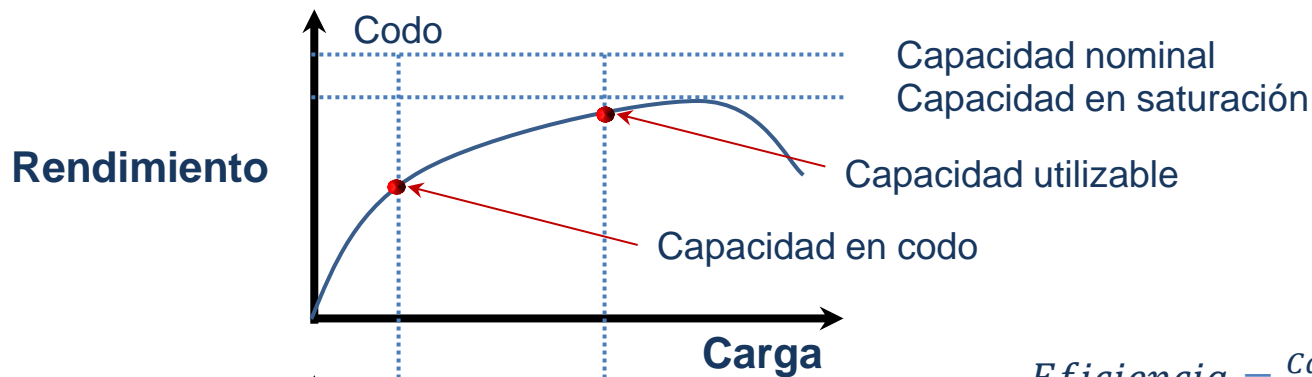
Planeamiento de la Capacidad en TI

- ② Analizar la capacidad actual
Mediciones:
Determinar frecuencia de muestreo
 - Poco frecuente:
Se pueden pasar cosas por alto
 - Muy frecuente:
Se puede atafagar de datos inútiles
Se puede perder perspectiva
Se genera recarga en el sistema



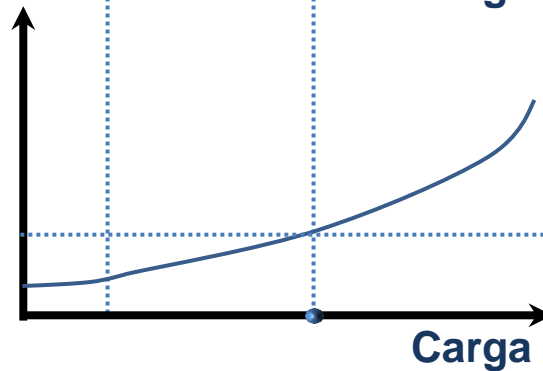
Planeamiento de la Capacidad en TI

- ② Analizar la capacidad actual
Tiempo de respuesta - Rendimiento:



$$Eficiencia = \frac{Capacidad\ utilizable}{Capacidad\ nominal}$$

Tiempo de respuesta
Tiempo límite



Planeamiento de la Capacidad en TI

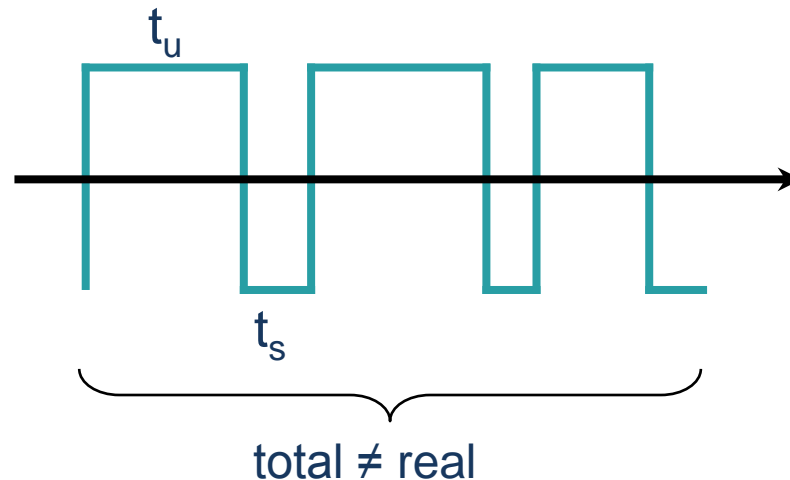
- ② Analizar la capacidad actual:

Tiempo de procesamiento:

- Usuario (t_u)
- Sistema (t_s)
- Total (o de CPU) = $t_u + t_s$
- Real: transcurrido (“wall clock”)

Usuario

Supervisor



Planeamiento de la Capacidad en TI

- ② Analizar la capacidad actual

Tiempo de procesamiento:

- Alarmas:

$$t_u + t_s \ll t_r$$

$$t_s > 10\%$$

- Unix:

```
$ time comando
```

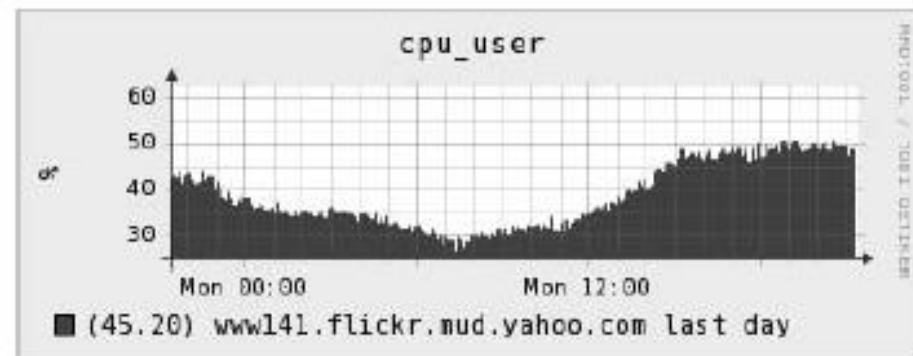
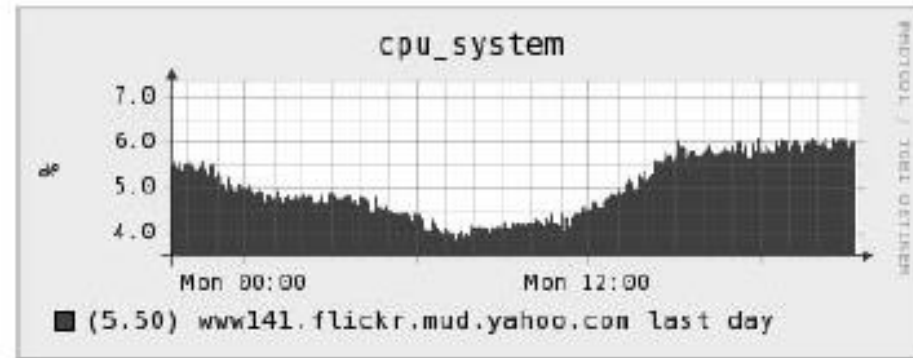
```
s.mmmu  s.mmmS  m:ss.cc  n.n%
```

Planeamiento de la Capacidad en TI

- ② Analizar la capacidad actual

Uso del procesador:

- Eficiencia = $(t_u + t_s) / t_r$
- Nivel de uso:
 - < 80%: hay disponibilidad
 - = 80%: saturación
 - > 80%: recargado

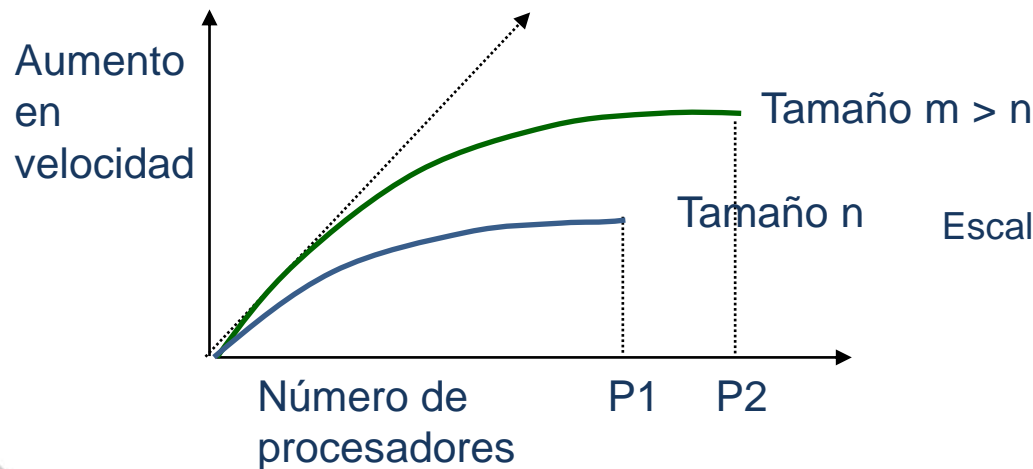


Planeamiento de la Capacidad en TI

- ② Analizar la capacidad actual

Uso del procesador: multiprocesadores

- Aumento de velocidad (“speedup”) = $T_{\text{secuencial}} / T_{\text{paralelo}}$
 - Debe ser mayor que 1
 - Deseable que sea proporcional al número de procesadores



Escalabilidad:

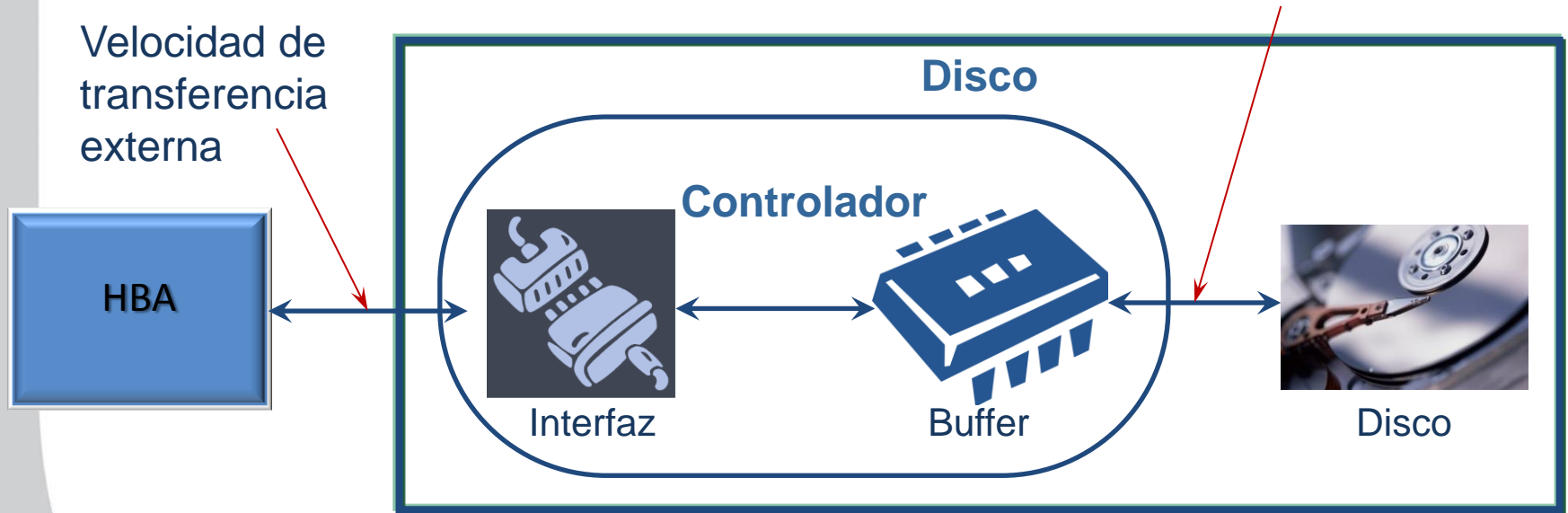
- P_i (punto saturación) grande
- $m \gg n$, entonces $P2 \gg P1$

Planeamiento de la Capacidad en TI

- ② Analizar la capacidad actual
Uso del procesador: multiprocesadores
 - $\text{Costo} = t_r * \text{Número procesadores}$
 - Total de tiempo de cómputo invertido
 - $\text{Eficiencia} = \Sigma t_{it} / \text{Costo}$
 - No puede ser mayor que 1

Planeamiento de la Capacidad en TI

- ② Analizar la capacidad actual :
Tiempo de E/S en disco:



Planeamiento de la Capacidad en TI

- ② Analizar la capacidad actual

Desempeño de discos:

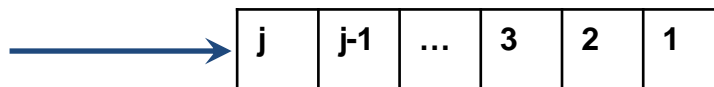
– Ley de Little:

$$N = aR$$

- N: Número total de requerimientos en el sistema
- a: tasa de llegada (número de requerimientos que llegan por unidad de tiempo)
- R: tiempo de retorno

Llegada de
requerimientos
de E/S

Cola de E/S



Controlador
de E/S

Salida de
requerimientos
de E/S

Planeamiento de la Capacidad en TI

- ② Analizar la capacidad actual

Desempeño de discos:

- Ley de uso:

$$U = aR_s$$

- U : uso del controlador

- R_s : tiempo de servicio (tiempo promedio que se gasta en el controlador)

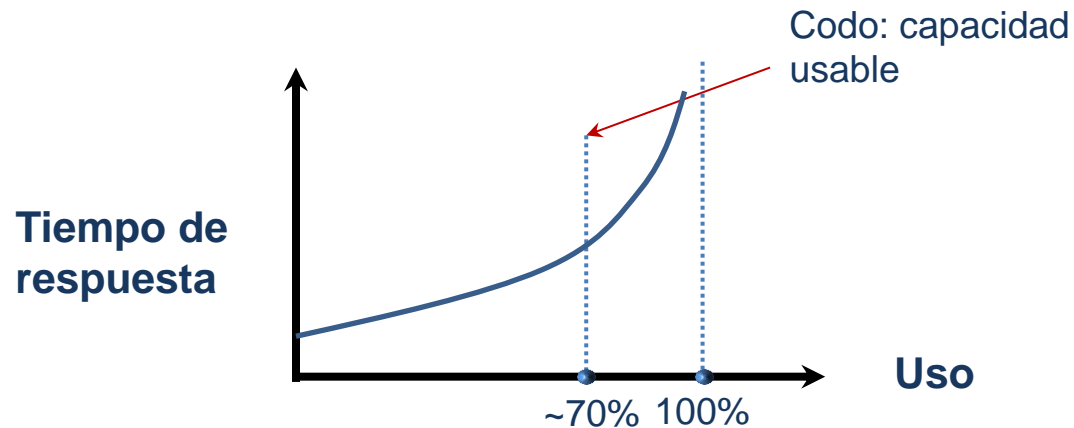
- Tiempo promedio entre llegadas: $R_a := 1 / a$

- $U = R_s / R_a$

- $R = R_s / (1-U)$

Planeamiento de la Capacidad en TI

- ② Analizar la capacidad actual
Desempeño de discos:



Planeamiento de la Capacidad en TI

- ② Analizar la capacidad actual
Capacidad de almacenamiento:
 - Datos operativos:
 - Tipos
 - Número de cada tipo
 - Tamaño de cada tipo
 - Archivo
 - Determinar niveles del ciclo de vida
 - Qué se conserva de lo operativo
 - En general: qué se conserva entre niveles
 - Cuántos periodos
 - Backup
 - Frecuencia
 - Número de versiones

Planeamiento de la Capacidad en TI

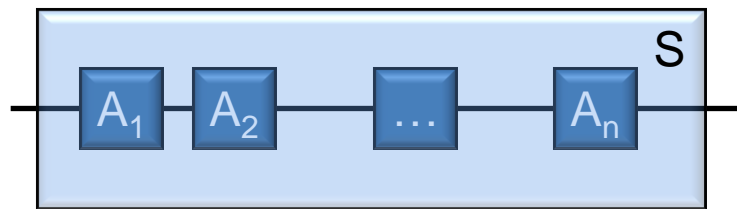
- ② Analizar la capacidad actual
- Disponibilidad - Principios:
- Robustez: disminuir el potencial de fallas
“Tan simple como sea posible, pero tan complejo como sea necesario”
 - Redundancia: recursos adicionales de respaldo
“Pluralitas non est ponenda sine necessitate”
 - Prevenir puntos únicos de fallo
 - Es una decisión del negocio
 - No existe la seguridad total
 - Alta disponibilidad: protección contra o recuperación ante fallas menores en corto tiempo
 - Disponibilidad continua: prestación ininterrumpida del servicio
 - Tolerancia a fallas: las fallas no se hacen visibles al usuario

Planeamiento de la Capacidad en TI

- ② Analizar la capacidad actual
Disponibilidad - análisis:

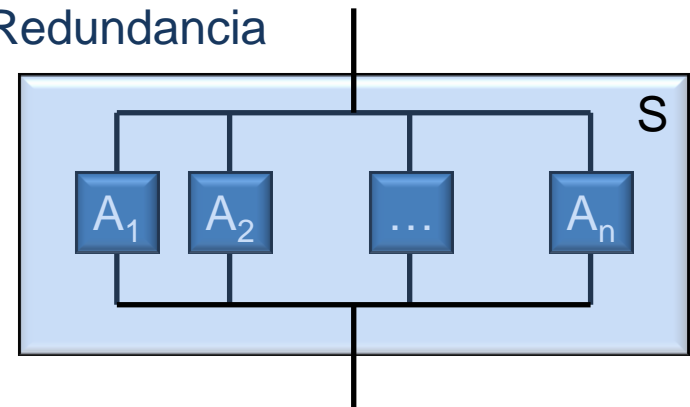
$P(A_i)$ = disponibilidad del elemento A_i

Dependencia secuencial



$$P(S) = P(A_1 \text{ and } A_2 \dots \text{ and } A_n) = \prod_i P(A_i)$$

Redundancia



$$P(A \text{ or } B) = P(A) + P(B) - P(A)P(B)$$

$$P(S) = P(A_1 \text{ or } A_2 \dots \text{ or } A_n) = \sum_i P(A_i) - \sum P(\text{todas las intersecciones})$$

Planeamiento de la Capacidad en TI

- ② Analizar la capacidad actual

Disponibilidad - análisis:

– La disponibilidad es una cadena

- 99.99 % = 53 minutos
- 99.99⁷ % = 99.93 % = 6:08 h
- 99.99⁶ % * 99 % = 98.94 % = 92:48 h

– Dan rendimientos decrecientes

$$\begin{array}{ccccc}
 99\% & & 99\% & & 99.99\% \\
 \boxed{\text{Equipo 1}} & + & \boxed{\text{Equipo 2}} & = & \\
 \$\$ & & \$\$ & & 2*\$\$
 \end{array}$$

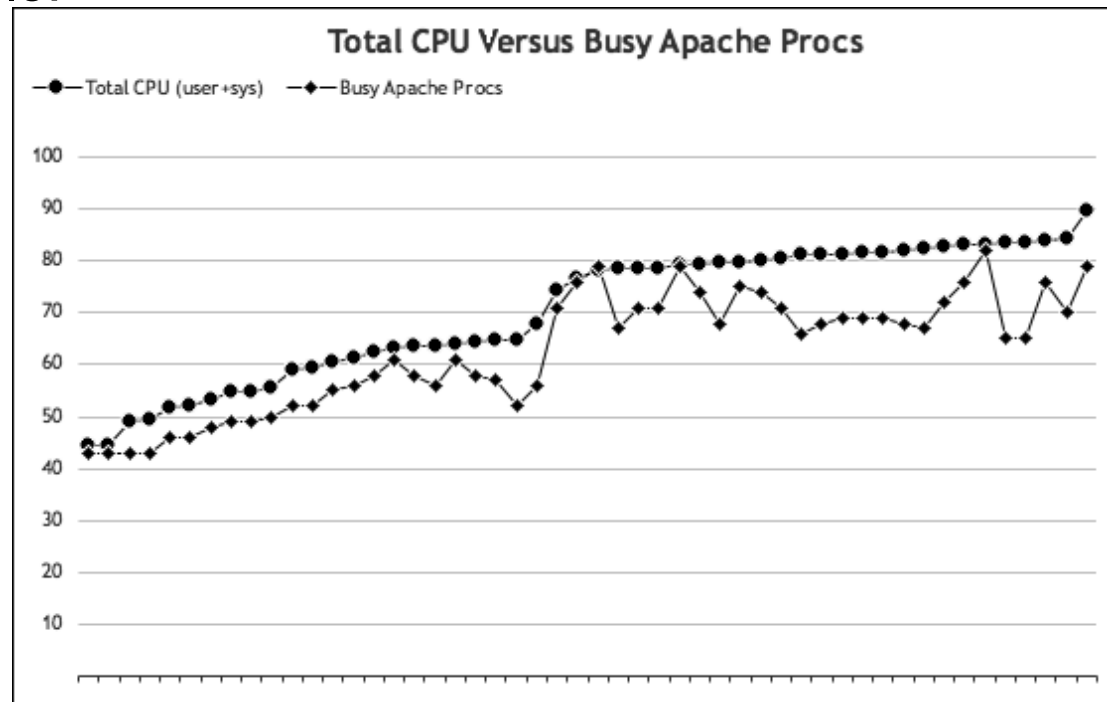
$$\begin{array}{ccccccc}
 p\% & * & p\% & * & p\% & = & p^n \% \\
 \boxed{\text{Equipo 1}} & & \boxed{\text{Equipo 2}} & & \dots & & \boxed{\text{Equipo n}}
 \end{array}$$

Planeamiento de la Capacidad en TI

- ③ Hacer prospectiva:
 - Elaborar pronóstico de necesidades de procesamiento:
 - Crecimiento del volumen de negocio
 - Planeado
 - Tendencia
 - Nuevas aplicaciones
 - Expansión del negocio
 - Limitaciones presupuestales
 - Elaborar modelo:
 - niveles del servicio = $f(\text{carga, recursos})$
 - Definir acciones

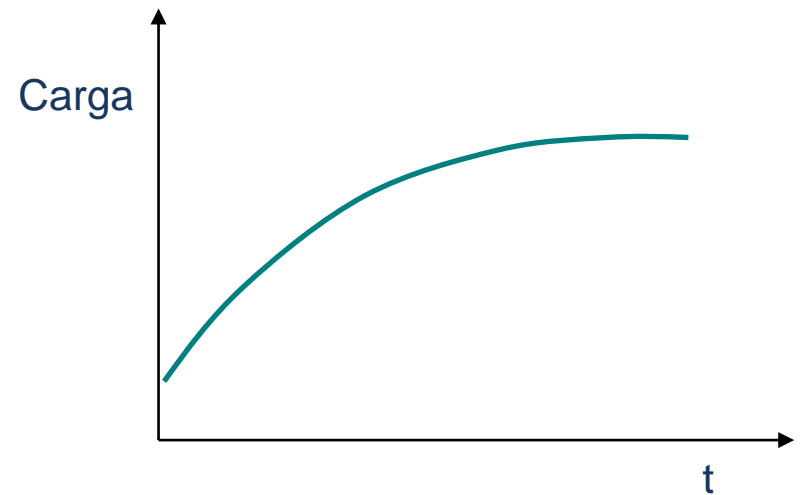
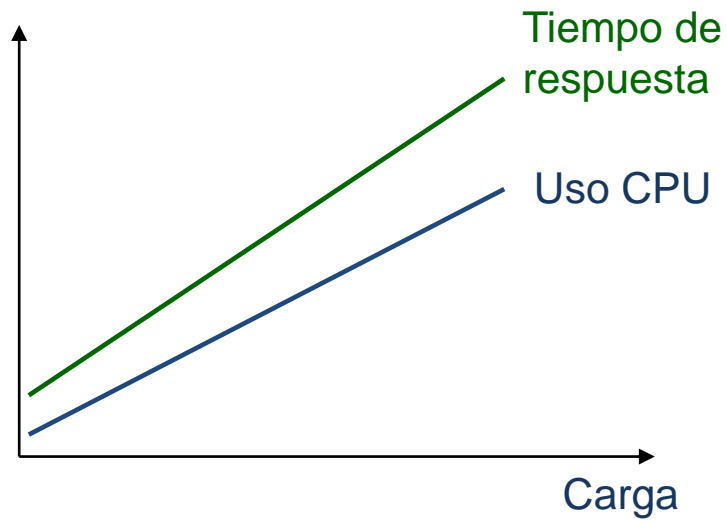
Planeamiento de la Capacidad en TI

- ③ Hacer prospectiva:
Elaborar modelo:



Planeamiento de la Capacidad en TI

- ③ Hacer prospectiva:
Elaborar modelo:

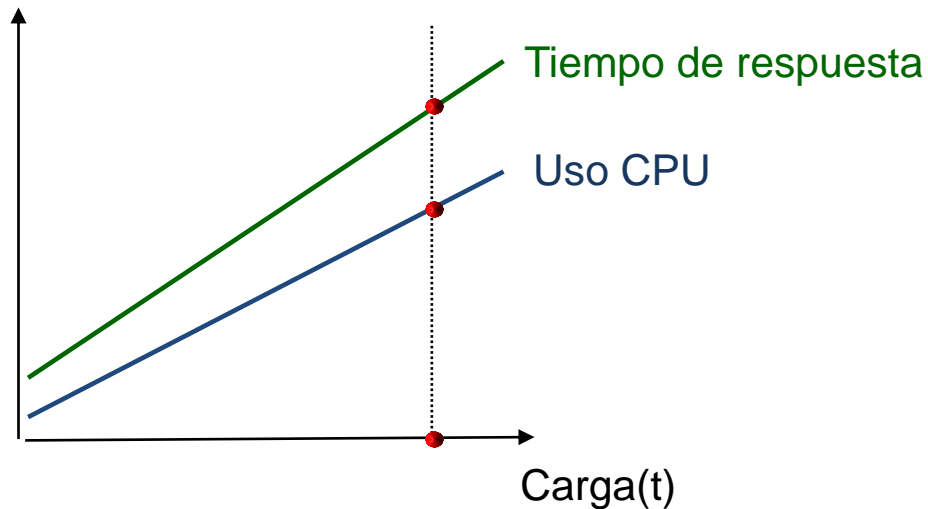


Planeamiento de la Capacidad en TI

- ③ Hacer prospectiva:
Definir acciones:
 - Determinar configuración:
 $\text{niveles del servicio} = f(\text{carga}, \text{recursos})$
 - Encontrar límites de crecimiento:
 $\text{niveles del servicio} = f(\text{carga}, \text{recursos})$
 - Redefinir niveles de servicio:
 $\text{niveles del servicio} = f(\text{carga}, \text{recursos})$

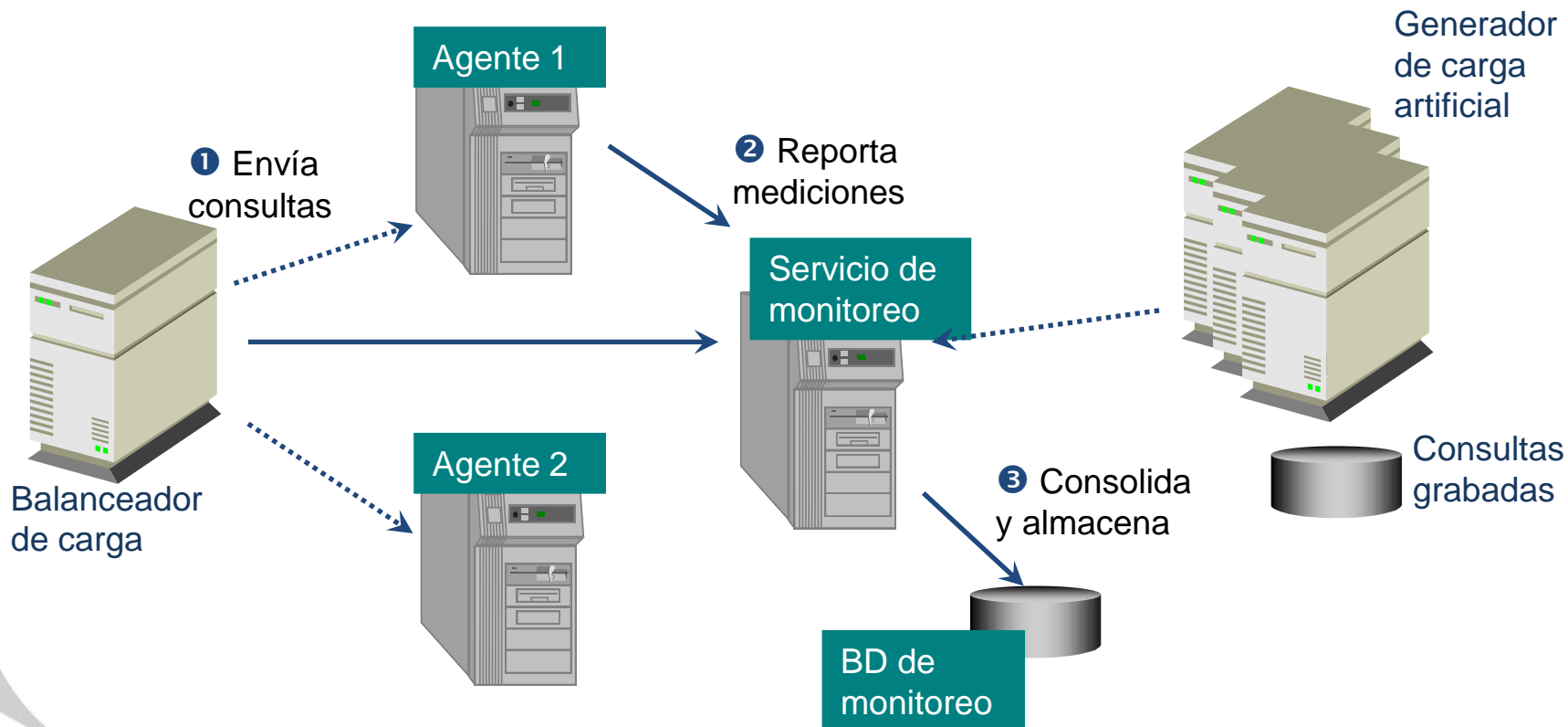
Planeamiento de la Capacidad en TI

- ③ Hacer prospectiva:
Elaborar modelo:
 - Aumento de velocidad



Planeamiento de la Capacidad en TI

- Sistemas de recolección de métricas



Referencias

Capacity planning. The art of capacity planning; John Allspaw.
Capítulo 3 (hasta pág. 39)

Capacity planning. Métricas. Art of Computer Systems Performance
Analysis; Raj Jain. Secciones 3.2-3.3.