

Planificación de la Capacidad

Néstor Espíndola

njespindola@cantv.net

0416-2751684

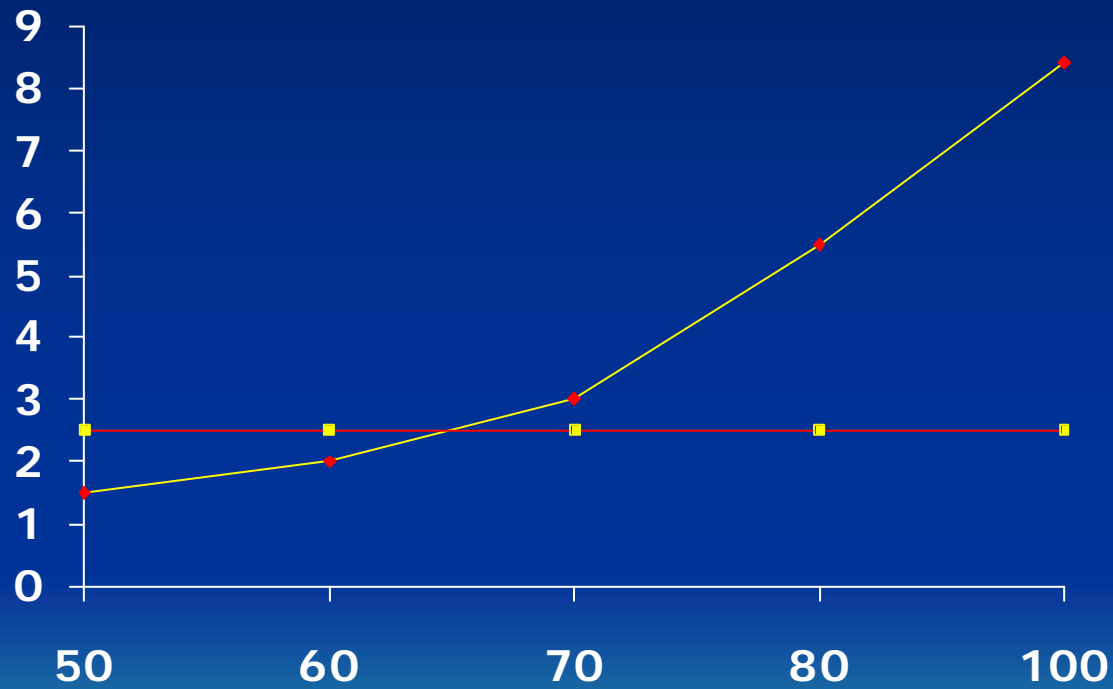
Consultor

Planificación de la Capacidad

Planificación de la capacidad es el proceso de *predecir* cuándo los niveles de servicio serán violados como función de la evolución de la *carga de trabajo*, así como la determinación de la forma más efectiva de demorar dicha saturación

Ejemplo

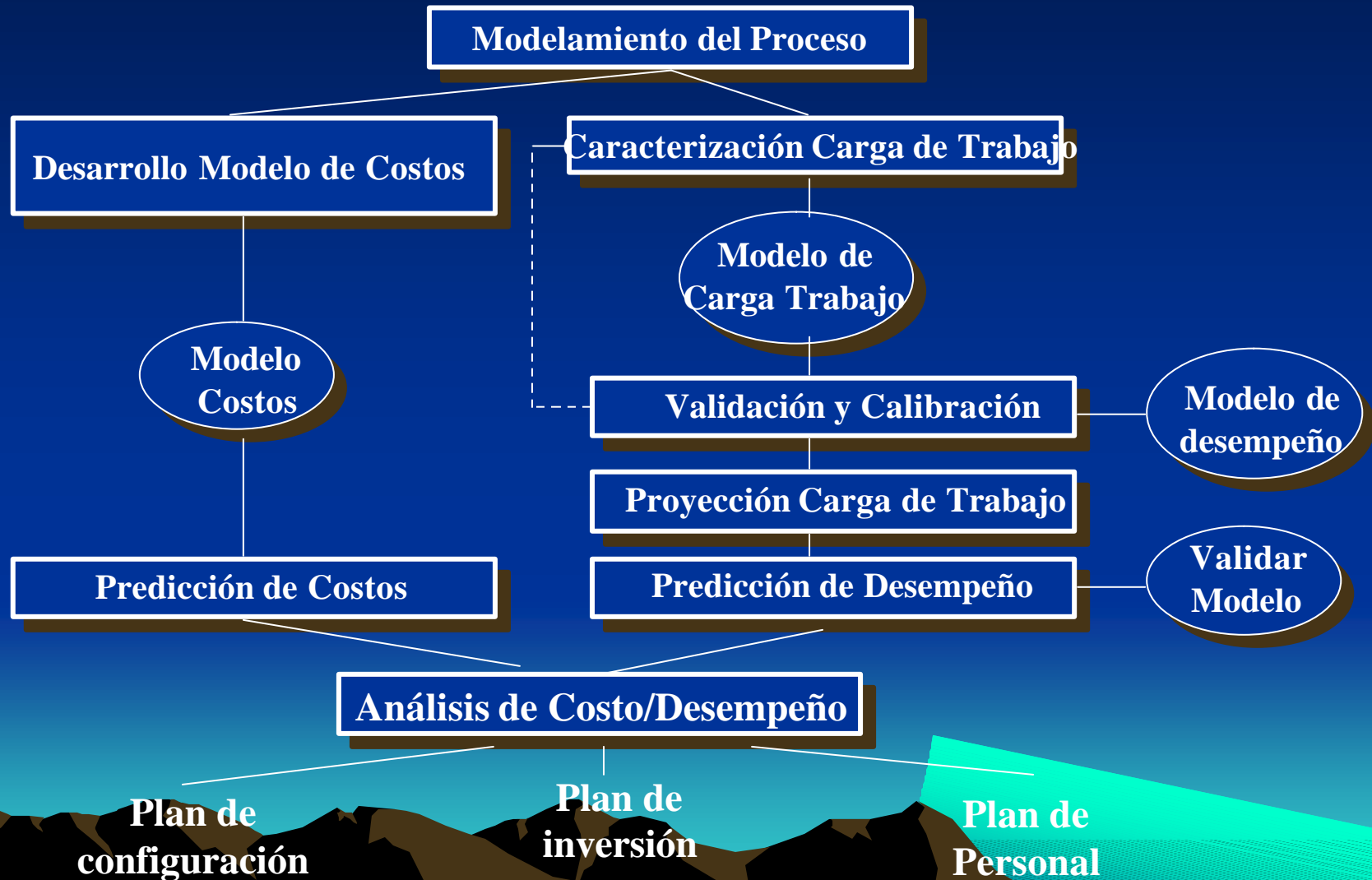
Tiempo de respuesta
(seg)



Nivel de servicio

Número de clientes

Metodología



Modelamiento de Procesos

- Herramientas gráficas que permiten describir (diagramar y documentar): funciones, operaciones, toma de decisiones, etc., que se realizan en el proceso.
- Existe un gran número de dichas herramientas, algunos son:
 - Modelo de Ciclos de Negocios
 - Modelo de Regulación
 - Modelo de Análisis Estructurado (SADT, IDEF0)

Modelamiento del Proceso

- **Recursos Productivos:** corresponden a los elementos que son utilizados para el desarrollo del proceso.
- **Configuración del Proceso:** corresponde a los procedimientos, reglas y distribución física de las operaciones.
- **Carga de Trabajo:** corresponde a los parámetros que describen la carga de trabajo, es decir, el tipo, frecuencia y comportamiento.
- **Niveles de Servicio:** representa la definición de aquellos indicadores de calidad del proceso.

Metodología



Caracterización de la carga de trabajo

- Es el proceso de particionar la carga de trabajo global en subconjuntos llamados componentes. Ejemplos de componentes son:
 - Transacciones de clientes, DB u otros,
 - Atención a clientes,
 - Trabajos con características similares.
- Los componentes de la carga de trabajo están compuestos de ***componentes básicos***.

Componentes

Componentes	Componentes Básicos
<i>e-mail</i>	<ul style="list-style-type: none">- <i>enviar mensaje</i>- <i>recibir mensaje</i>
<i>Acceso a servidor DB</i>	<ul style="list-style-type: none">- <i>query</i>- <i>update</i>
<i>Solicitud de saldo</i>	<ul style="list-style-type: none">- <i>identificación</i>- <i>acceso a DB</i>- <i>respuesta solicitud</i>

Parámetros Componentes Básicos

- **Parámetros de Intensidad**
 - número de mensajes enviados/hora
 - número de llegas de clientes/min
- **Parámetros de Servicio**
 - Largo promedio del mensaje
 - Tiempo promedio de atención.

Caracterización

- Identificación de los componentes de la carga de trabajo.
- Identificación de componentes básicos.
- Selección de parámetros.
- Recolección de datos: medidas, benchmarks y otros.
- Partición de la carga de trabajo: promedios y clustering.

Alternativas de obtener datos

- usa *“benchmarks”*
y *experiencia*

- usa *medidas,*
“benchmarks” y
experiencia

- usa *medidas*



simple

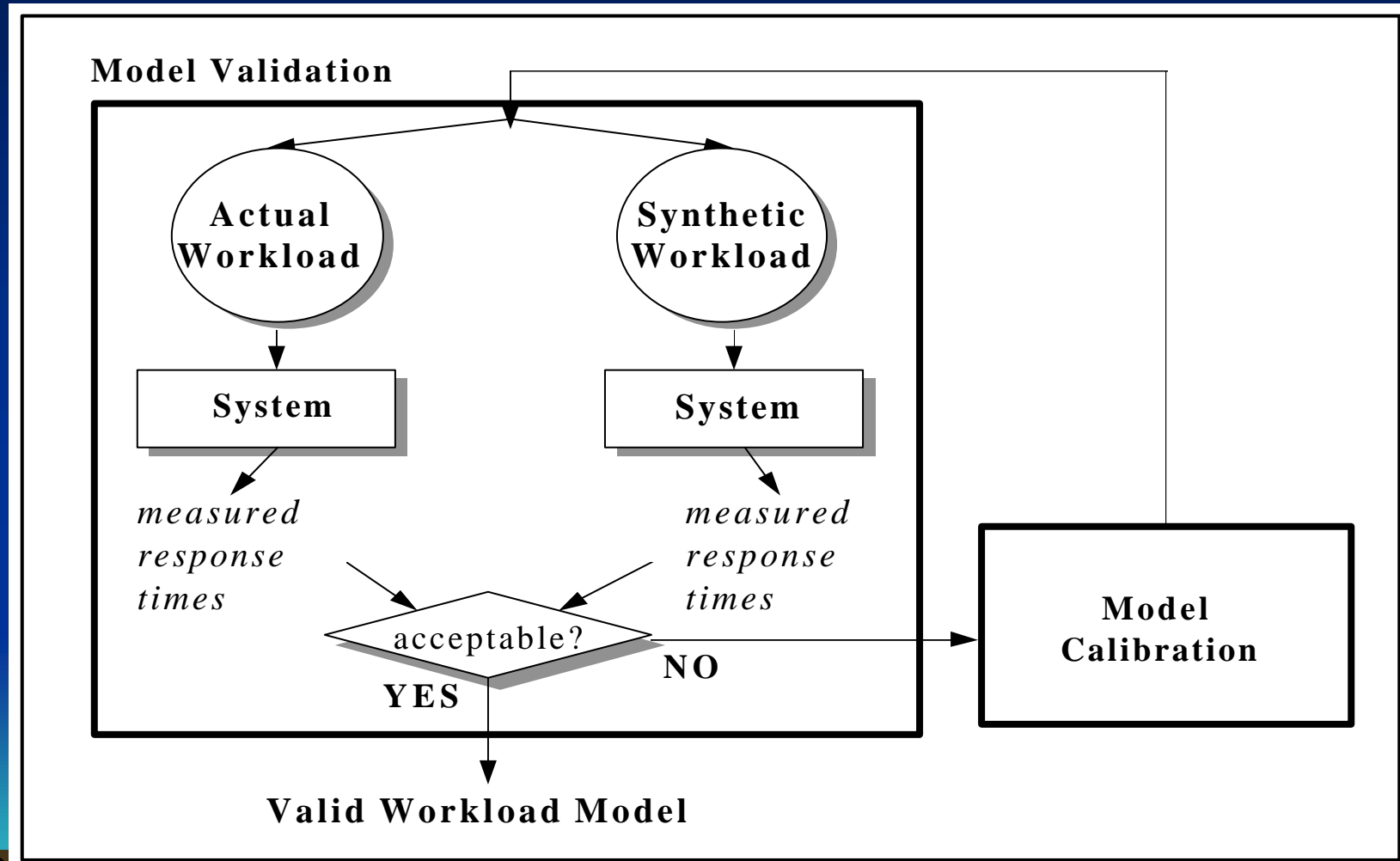
medio

detallado

Metodología



Validación



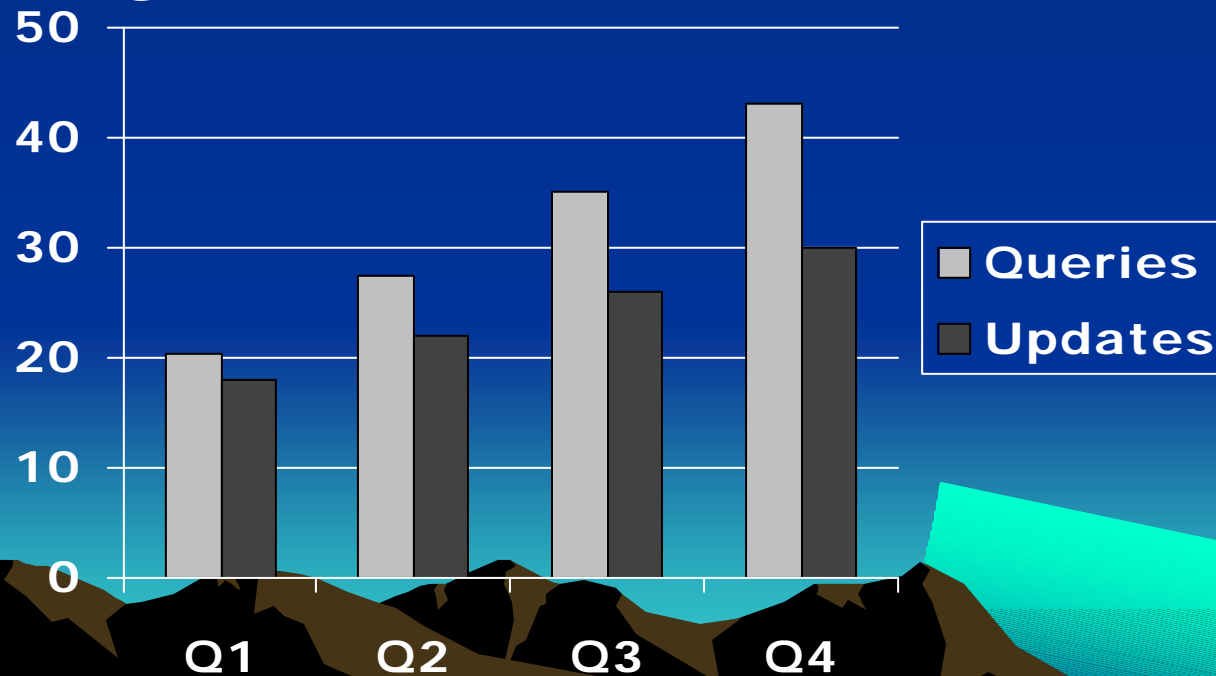
Metodología



Proyección de la Carga de Trabajo

Proceso de predecir la intensidad de la carga de trabajo

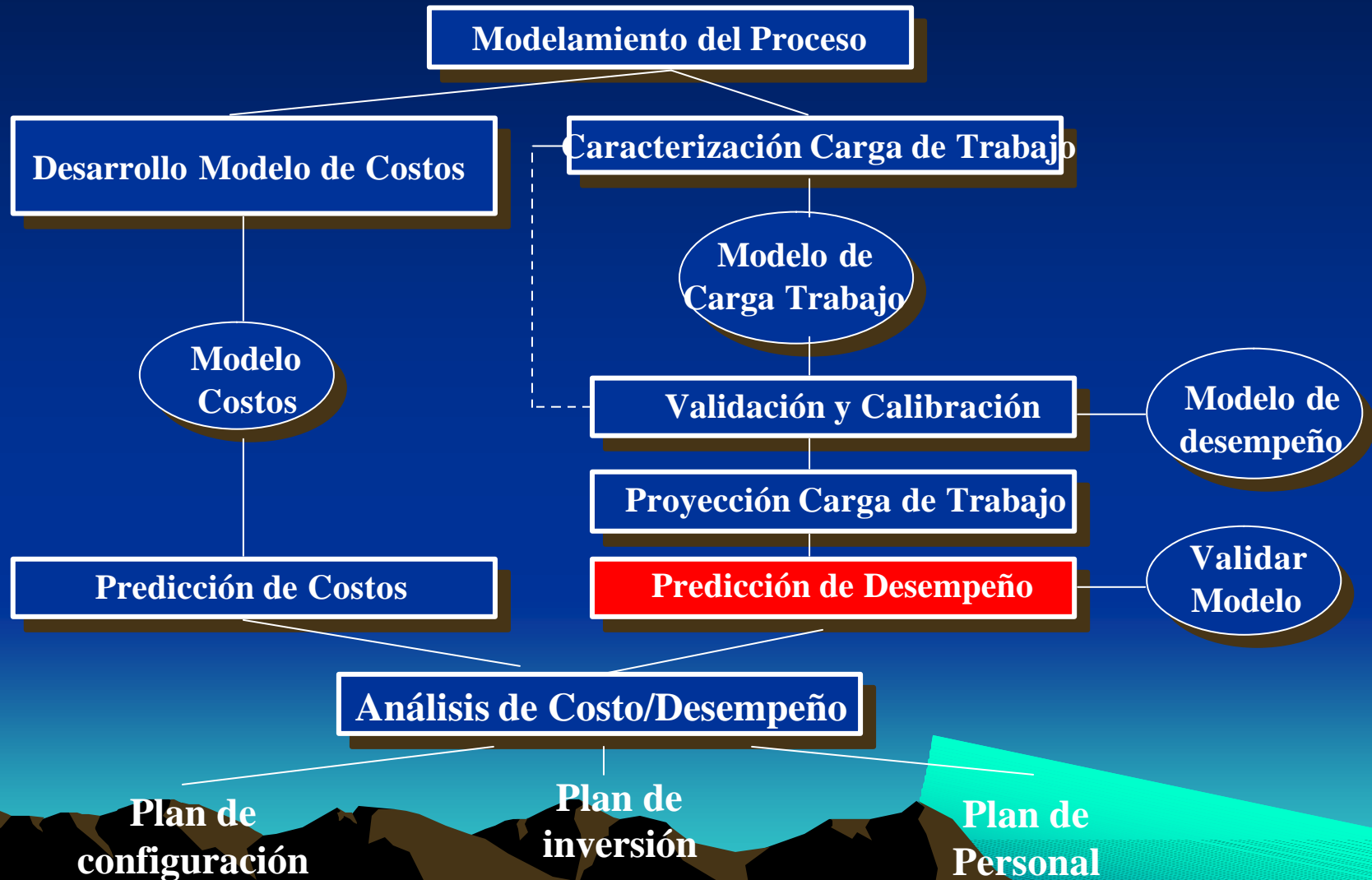
Trans. por seg



Metodología

- Recopilación de antecedentes (datos, planes estrategicos, competencia, etc.).
- Uso de metodologías de pronóstico (regresión lineal, promedios móviles, ajuste exponencial, etc.) ajustados por planes estrategicos.

Metodología



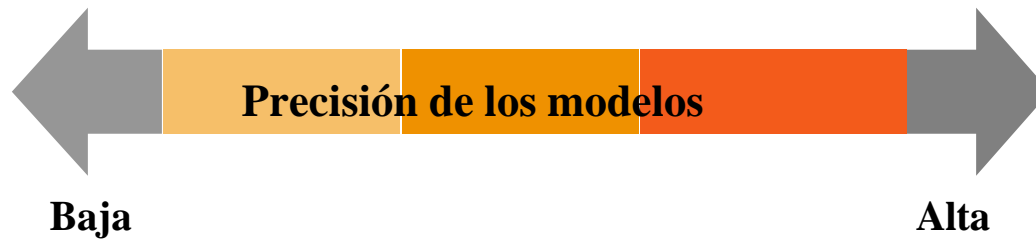
Modelos de Predicción

- Modelos de simulación: permite una alta representación de los fenómenos, pero son difíciles de desarrollar.
- Modelos analíticos basados en redes de colas (Queuing Networks (QNs))
 - eficientes
 - permiten un rápido análisis de un gran número de escenarios
 - buenos para fines de identificación de tradeoffs

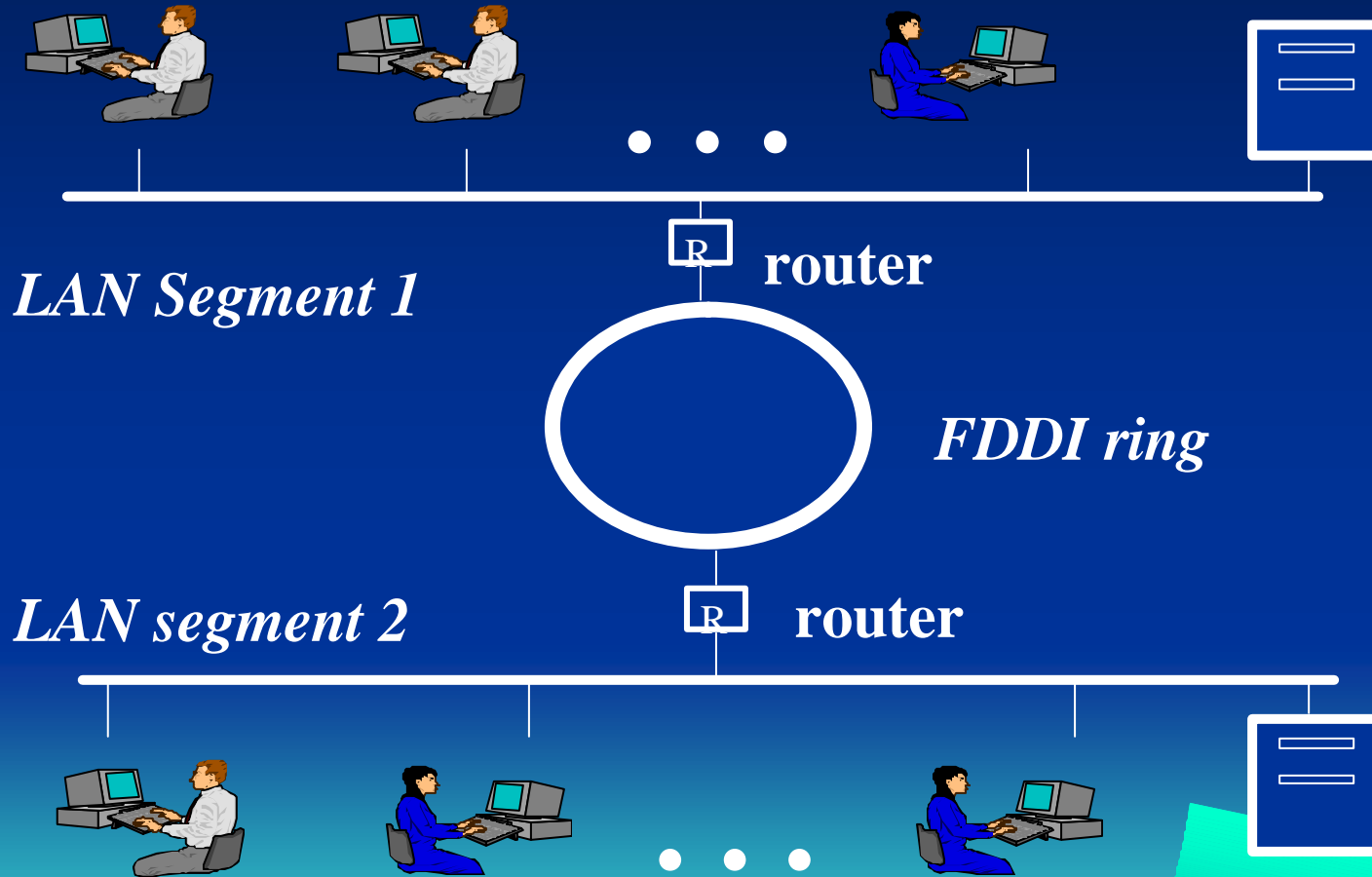
Precisión de los Modelos

- *modelos generales*
- *poco esfuerzo en manejo de datos*

- *modelo de detalles*
- *intensivo en uso de datos*



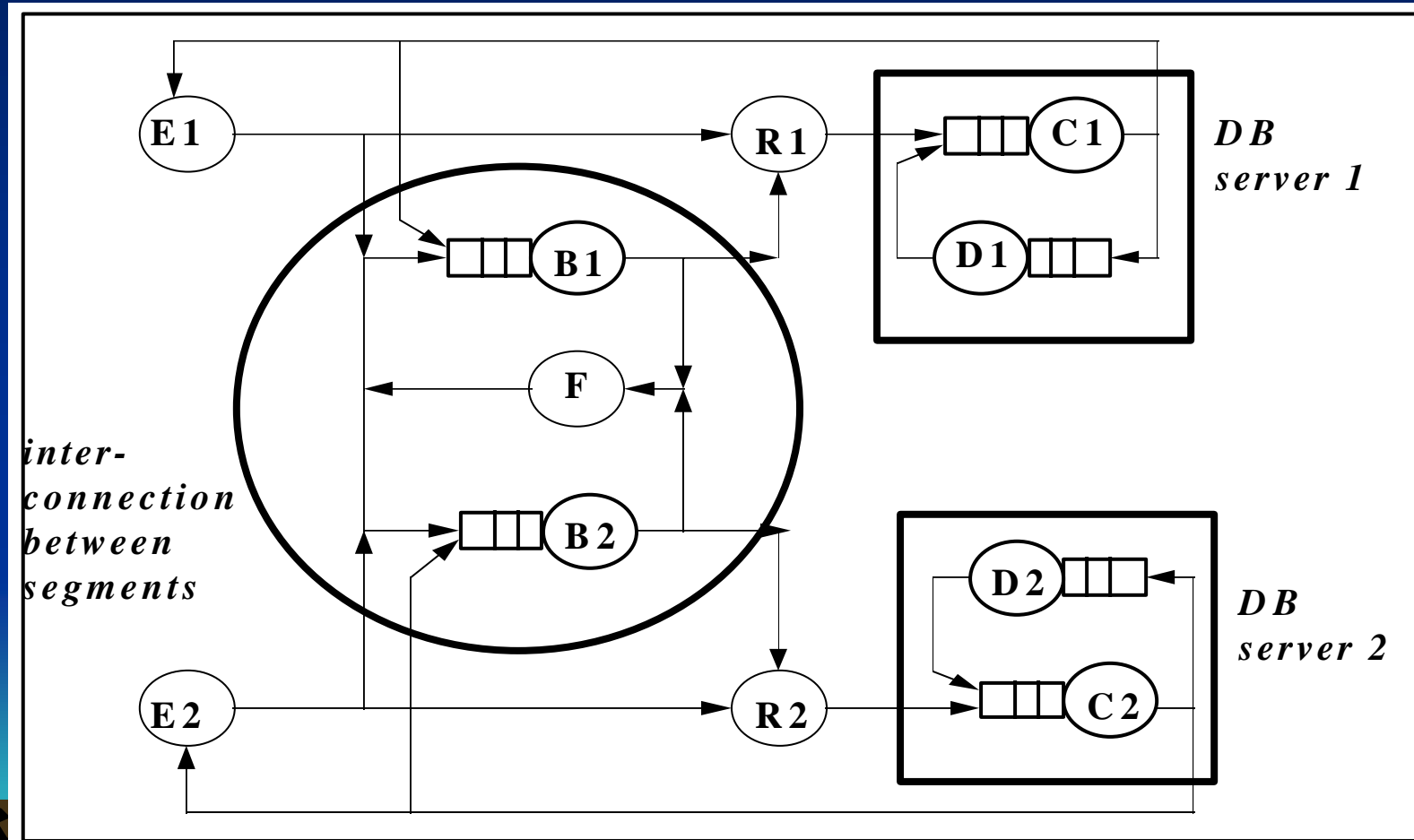
Ejemplo de Predicción



Factores que afectan el desempeño

- Estaciones de trabajo (clientes)
- Servidores
- Medios de comunicación
- Protocolos
- Mecanismos de interconexión
(routers and gateways)

Ejemplo



Tiempos de Respuestas



Metodología



Modelo de Costos

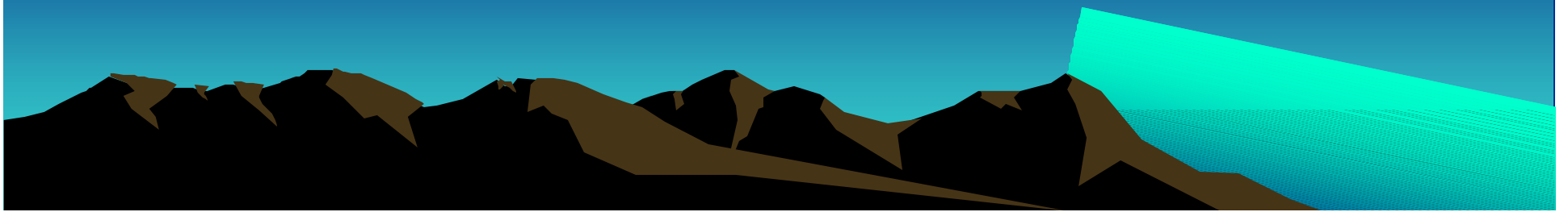
- Pocas empresas cuantifican los costos de provisión de servicio. Ejemplo, en EE. UU. menos del 5% cuantifican costos de PC y LAN.
- Algunos costos *escondidos* en los procesos
 - administración y control gerencial
 - entrenamiento continuo
 - mantención y upgrades de hardware y software
 - costos de personal indirecto

Algunos Costos

- El costo de upgrades de software y hardware es 10% del precio de compra anual.
- Un administrador de LAN varía entre US\$500 y US\$700 por cliente de WS/mes.
- Costos de entrenamiento varían entre US\$1,500 y US\$3,000 por técnico al año.
- 40% de los costos de personal corresponden a administración, 40% en desarrollo y 20% en otros aspectos.

Ejemplo

Caso de Telemarketing



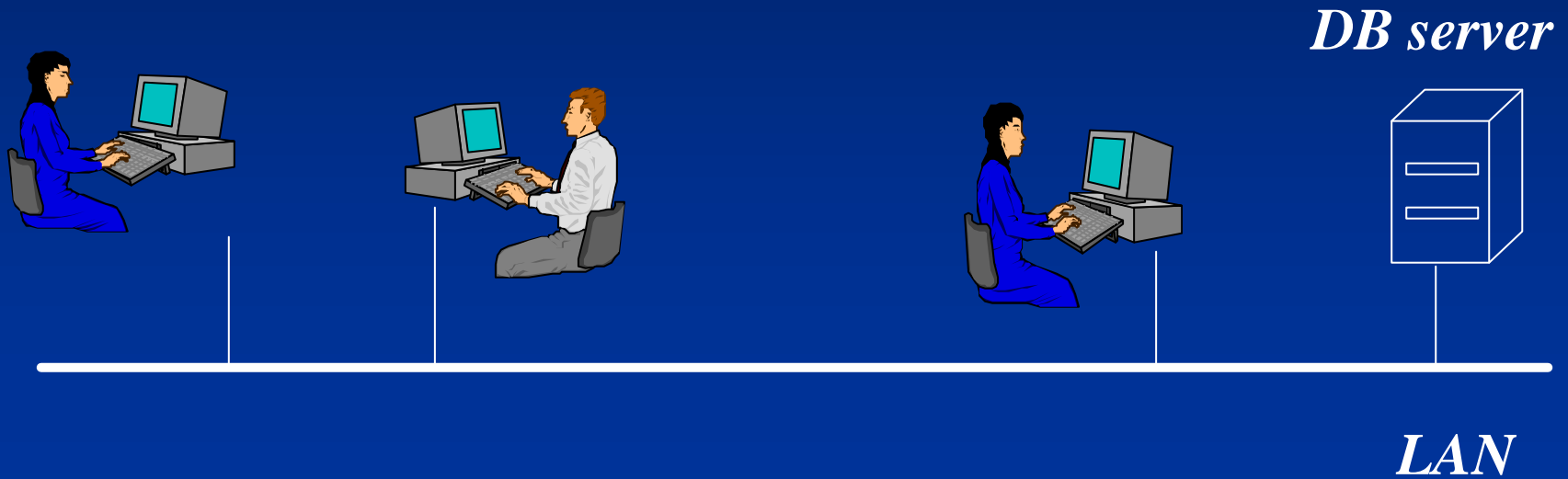
Ejemplo

- Consumidores ordenan los productos mediante el uso de un catálogo.
- Las ordenes son realizadas mediante el teléfono usando tarjeta de crédito.
- 30,000 ordenes son recibidas al día.
- Las llamadas son colocadas en hold hasta que un asistente está disponible.

Respuesta Buscada

Cuántos vendedores son necesarios para garantizar que una llamada que entra no espere más de 5 segundos en promedio?

Ejemplo

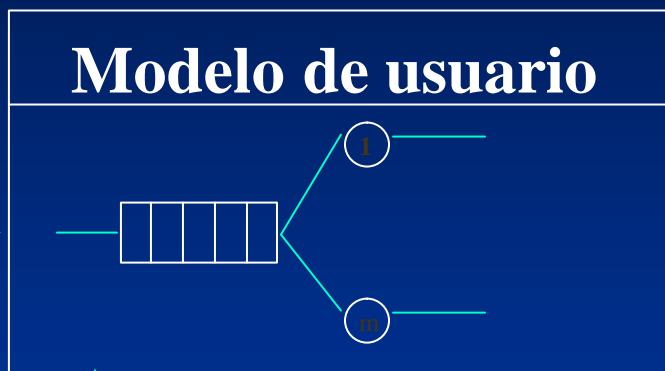


Ejemplo ...

- m (a ser determinado) estaciones de trabajo y un servidor SQL.
- LAN Ethernet (10 Mbps)
- SQL server: una CPU y un disco.

Modelo Jerárquico

tasa de
llegada de
llamadas



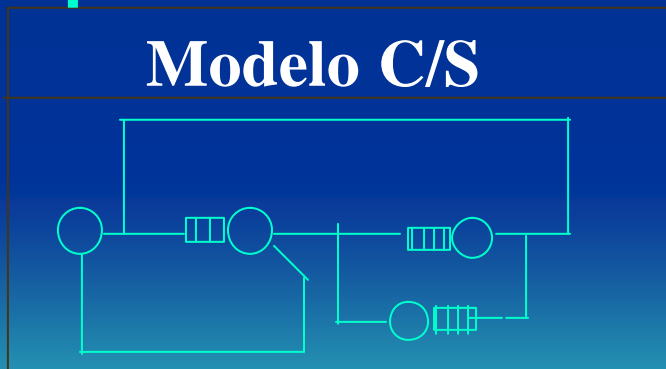
Tiempo de
espera
promedio
por
llamada



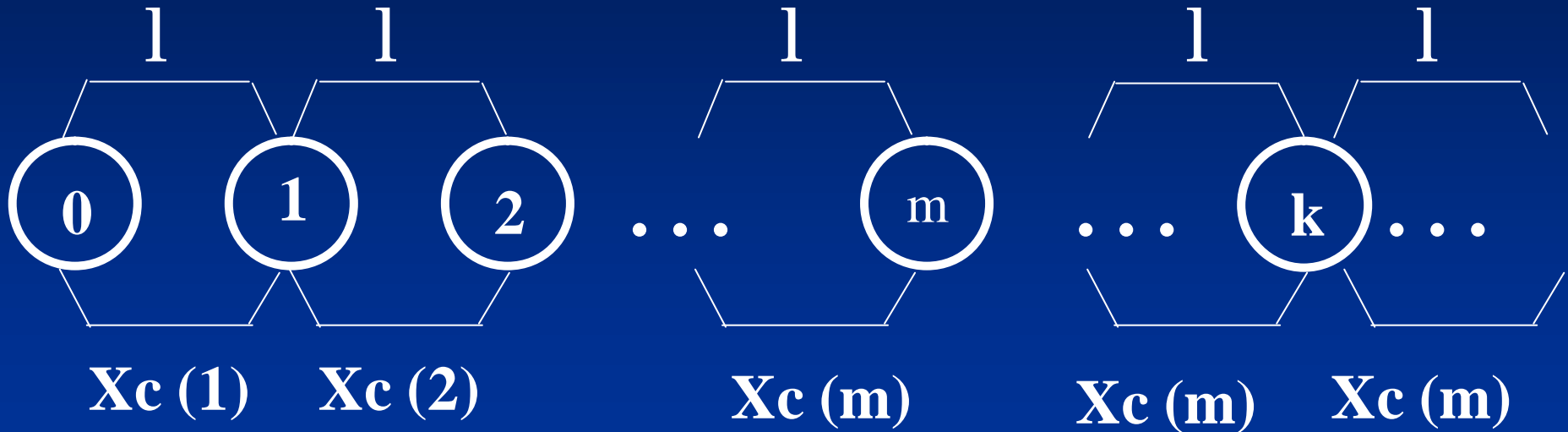
$X_c(k) \quad k=0, \dots, m$



Parámetros de
aplicación,
servidor y
LAN



Modelo de Usuarios



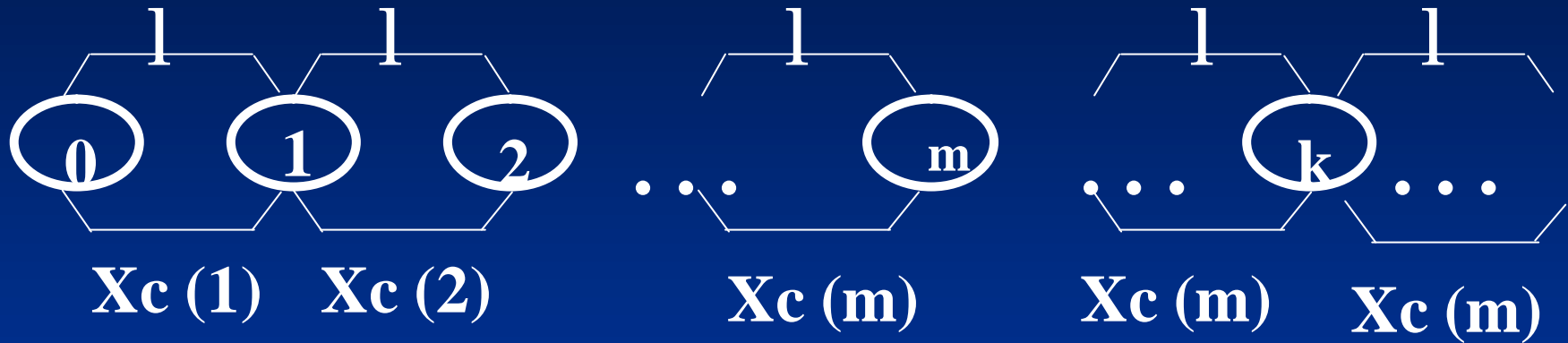
k = número de llamadas en el sistema

Modelo de Usuarios ...

- Cálculo de tasa de llamadas promedio I :
 - 30,000 llamadas/día
 - 12 horas de operación día
 - tráfico balanceado durante el día:

$$I = \frac{30.000 \cdot \text{llamadas} / \text{día}}{12 \times 3.600} = 0.69 \cdot \text{llamadas} / \text{sec}$$

Modelo de Usuario ...



- **Solución:**

$$p(k) = P(0) \cdot \prod_{i=0}^{k-1} \frac{1}{X_c(i+1)}$$

$$X_c(j) = X_c(m) \forall j \geq m$$

$$W = N_w / I = (1/I) \sum_{k=1}^{\infty} k P_k^w$$

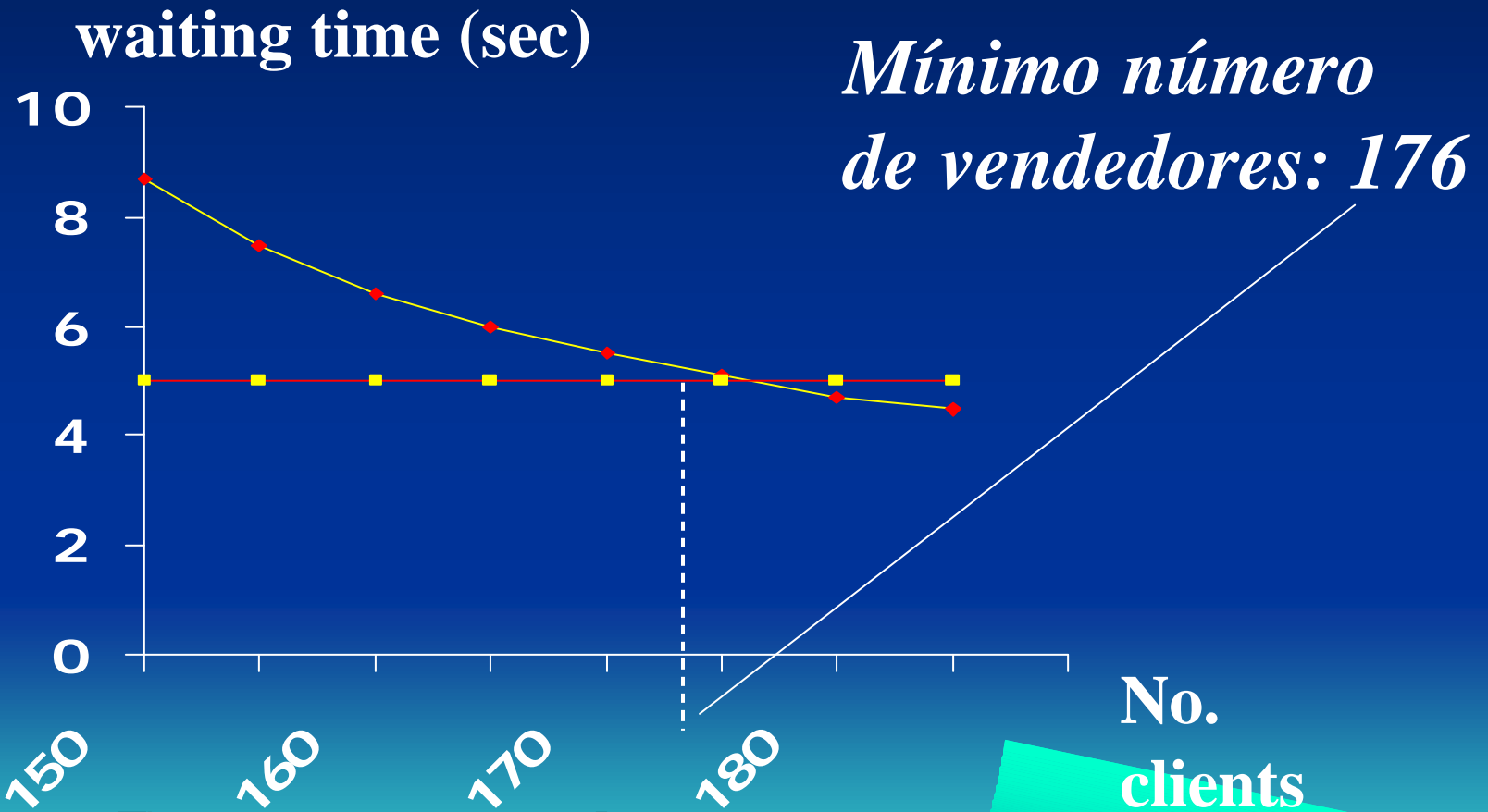
$$P_k^w = \sum_{j=0}^m P_j \quad k=0$$

$$= P_{k+m} \quad k > 0$$

Modelo Cliente/Servidor

- Si la utilización de la red LAN es muy baja, se puede considerar como un mecanismo de demora (e.g., alto ancho de banda).
- Si la utilización es mayor que el 20%, se puede modelar como un mecanismo dependiente de la carga.
- Los elementos internos (routers) se deberían modelar como demoras (segundos/paquete).

Tiempo promedio de espera por llamada



Conclusiones

- Muchos modelos de procesos comparten recursos computacionales, los que pueden ser críticos al momento de evaluar la capacidad, comunicación y capacidad de red.
- El dimensionamiento de los procesos requieren del uso de modelos predictivos de desempeño.
- Los modelos analíticos son una buena alternativa para entregar soluciones rápidas de los tradesoffs de los parámetros del sistema.

Consultoria

- Estudios de Capacidades.
- Analisis de Performance.
- Proyectos de Infraestructura
- Gerencia de Proyectos
- Estudios de Riesgos/ Seguridad
Informacion Unix /NT
- Gerencia de Datacenter Client/server
- Contacto: njespindola@cantv.net

Cargos

- Subgerente de Tecnologia CITIBANK.
- Gerente de Tecnologia y Planificacion BANCO DE VENEZUELA.
- Gerente de QA/TQM BANCO UNION
- Director Asociado de Procesamiento BANCO DEL CARIBE.

Proyecto

- Cambio de Plataforma Tecnologica de Mainframe IBM (SAFE) a Cliente/Servidor y migracion de data y oficinas bancarias.
- Servidor: Sun/Solaris/Sybase/ Cobis
- Cliente: Wintel/NT/SQL/VB6.0
- Red: TCP/IP, Cisco 5500/WAN-Frame Relay/Router/LAN Ethernet
- **Cliente: Banco del Caribe**