

MODELO PARA LA EVALUACION DE RENDIMIENTO DE SISTEMAS MANEJADORES DE BASE DE DATOS DISTRIBUIDAS¹

JUDITH BARRIOS ALBORNOZ

ESCUELA DE INGENIERIA DE SISTEMAS

UNIVERSIDAD DE LOS ANDES

MERIDA, VENEZUELA

RESUMEN

Se presenta una vision intermedia para la evaluacion del rendimiento de los Sistemas Manejadores de Base de Datos Distribuidas (SMBDD). Para realizarla se define un modelo conceptual de rendimiento que permite el establecimiento de las relaciones entre los componentes internos del sistema distribuido tales como traductores y procesadores de comandos de usuario, procesadores locales de datos y, los componentes externos a este: aquellos que apoyan o soportan las actividades del sistema de base de datos, pero que no forman parte integral del mismo, estos son: el sistema de comunicaciones y transmisión de datos y el equipo computacional. La medida de rendimiento evaluada es el tiempo total de respuesta.

El estudio no es exhaustivo, sin embargo, permite observar las relaciones cambiantes entre los componentes del SMBDD cuando es sometido a la influencia de las aplicaciones que manipulan la base de datos en un ambiente distribuido.

Se designa como factor critico de rendimiento a la visión distribuida de la base de datos y a las características particulares de cada una de las aplicaciones que accesan la base de datos distribuida. En este caso, su importancia no radica en el tiempo tomado para la definición de estrategia distribuida, sino en el tiempo total ahorrado por una buena decisión en la forma procesamiento de las solicitudes de los usuarios demandadas por las aplicaciones particulares.

1-INTRODUCCION

Las bases teóricas sobre las cuales se apoya un estudio de rendimiento, permiten establecer el contexto estudiado, las variables implicadas en el proceso, las restricciones y limitaciones tomadas en cuenta en la realización

¹ Trabajo realizado como Tesis de Maestria en Sistemas de Informacion en el Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM), Mexico 1990.

de la evaluación. La explicación del contexto, es definida en términos de modelos que reflejan la realidad, los cuales, pueden ser matemáticos, teóricos, de simulación, o, de cualquier otro tipo que permita mostrar lo que ocurre en el sistema.

En el área de sistemas distribuidos, de acuerdo a la investigación bibliográfica realizada, no se ha desarrollado hasta el momento, ningún modelo general de rendimiento que permita explicar el comportamiento global de los sistemas manejadores de base de datos. Esto se debe, especialmente, a que los estudios de rendimiento sobre SMBDD's, han tenido dos orientaciones principales, ubicadas en los extremos de especialización y generalización de la evaluación.

La primera orientación, en la cual se ha concentrado la mayoría de las investigaciones, se avoca a la medición cuantitativa de tiempos de respuesta y de ejecución en modelos muy especializados. Dichos modelos explican comportamientos particulares y toman en cuenta, solamente, algunos de los componentes internos o de apoyo del sistema distribuido. Este enfoque es muy técnico y puntualiza los detalles de comportamiento de elementos críticos pertenecientes a la estructura interna del SMBDD. Entre algunas de las referencias que tienen este enfoque especializado estan [Bhar 89], [Bodo 88], [Casa 89], [Garc 88]. En este tipo de estudios, no se toma en cuenta la posibilidad de un enfoque más amplio; es decir, visualizar el contexto en el cual está inmerso el sistema distribuido. Al perder el concepto de globalidad, por la misma especialización, se dejan de incluir en el estudio elementos que contribuyen fuertemente en el rendimiento total del sistema. Sin embargo, no debe olvidarse que a medida que en un estudio se amplía el marco de referencia, se va perdiendo la oportunidad de profundizar en los detalles componentes de los elementos del sistema.

El segundo enfoque, opuesto al primero, es aquel en el que no se consideran los componentes especializados del sistema; sino que toma al sistema como una caja negra, la cual se somete a diferentes valores en sus factores de entrada y, se mide la respuesta o la velocidad de reacción del sistema ante los comandos introducidos. Esta salida es vista como tiempos de procesamiento y de respuesta, así como características de interfaz que el sistema evaluado proporciona al usuario. Algunos estudios incluyen las facilidades externas y las habilidades que ofrece el sistema para una

interacción hombre/máquina más amigable.(miden las cualidades de la interfaz proporcionada al usuario para facilitar su trabajo de desarrollo de aplicaciones) [Robi 88], [Youn 88], [Nich 88]. Las últimas referencias señalan evaluaciones para sistemas centralizados. Dado que el área de sistemas distribuidos es joven, las evaluaciones realizadas siguen la primera orientación (especializada) y los SMBDD's que existen, se encuentran aun en la fase experimental.

Tomando en consideración las referencias anteriores, en este proyecto se propone la definición de un modelo general de rendimiento para SMBDD. Dicho modelo señala los parámetros determinantes del comportamiento global de los SMBDD y las relaciones existentes entre ellos, colocando así esta investigación, en un punto intermedio entre los dos enfoques mencionados previamente.

Dado que el parámetro central de la evaluación, es el **tiempo total de respuesta**, la medición comienza con la introducción de una solicitud de información por parte del usuario del sistema y finaliza cuando se despliega dicha información al usuario. Por medio del modelo establecido, se analizan los cambios potenciales sobre la respuesta del sistema, al variar los valores de elementos externos al SMBDD (la frecuencia de consulta, el volumen de datos implicados, el tipo de requerimientos del usuario, etc.). El cambio en el tiempo de respuesta, es causado por el grado en que varían las relaciones internas del SMBDD cuando se enfrentan a valores diferentes de la influencia del medio ambiente.

Para llevar a cabo este estudio, se toman como base fundamental, algunas de las investigaciones especializadas en el área de SMBDD. Dichas investigaciones proporcionan información sobre las características de comportamiento aislado de los diversos componentes del sistema distribuido específicamente en el nivel especializado. A través de esta información, se definen tanto las características de los componentes del SMBDD, como las del conjunto de elementos externos sobre los cuales se apoya dicho sistema manejador para poder funcionar.

2- CONCEPTOS BASICOS sobre SISTEMAS MANEJADORES DE BASE DE DATOS DISTRIBUIDAS

La arquitectura de un SMBDD permite entender la configuración e interfaz entre los componentes del sistema, dando la posibilidad de convertir datos y programas de un formato a otro, haciendo más flexible la comunicación entre los distintos nodos del sistema distribuido. La arquitectura de referencia seleccionada como base del estudio [Lars 85a], maneja dos tipos de componentes: un conjunto de procesadores y un conjunto de esquemas. Los procesadores son partes de programas -software- ejecutables que tienen la responsabilidad de realizar una o más tareas. Pueden tener subprocesadores asociados para realizar algunas de las tareas más especializadas. Los esquemas son descripciones de algún aspecto lógico o físico de la base de datos. Son utilizados para que los procesadores realicen sus tareas eficientemente. Por lo tanto, un sistema de base de datos es el resultado de la cooperación entre diferentes procesos que siguen ciertos esquemas y manipulan un conjunto de datos locales. Extendiendo este concepto a sistemas distribuidos, los procesadores(usuario-datos) y los esquemas (externo, conceptual y de distribución) ya no interactúan a nivel local, sino que se comunican por medio de una red de transmisión de datos con otros procesadores y esquemas localizados en otros nodos; ampliando así, el marco de procesamiento para las diferentes tareas implicadas en una solicitud de usuario. La complejidad de cooperación está oculta bajo un sistema de niveles y las interfaces entre ellos se establecen a través de un conjunto de reglas y protocolos, los cuales instruyen al elemento de nivel inferior sobre cómo estructurar los mensajes hacia los niveles superiores.

Ahora bien, teniendo en cuenta que el tipo de relaciones de cooperación e intercambio que existen entre los componentes del SMBDD pueden variar, es necesario, clasificar a los SMBDD en dos grandes grupos: los sistemas homogéneos y los sistemas heterogéneos:

1) Sistemas homogéneos presentan cierto grado de similitud entre sus componentes. De esta manera se hace más fácil el establecimiento del grado de interacción existente entre los elementos del modelo cuando se ejecutan las actividades de procesamiento distribuido. Las diferencias entre las características de implantación de cada uno de ellos, así como las diferencias

de equipo y de programas de apoyo que los soportan, no son demasiado relevantes para la clasificación.

Dentro de esta categoría, las investigaciones han llegado a producir prototipos funcionales bastante precisos a pesar de que la evaluación de su desempeño y de su habilidad para cumplir los objetivos establecidos aún se encuentra en experimentación. Actualmente, existen a nivel comercial algunos productos catalogados como sistemas manejadores de base de datos distribuidas y homogéneas. Aunque ninguno cumple cabalmente con todos los requerimientos y funciones establecidas, puede afirmarse que representan un gran apoyo para la construcción de bases de datos distribuidas en ambientes organizacionales. Quedan excluidos de este grupo, algunos productos comerciales centralizados a los cuales se les extiende con módulos especiales para que puedan funcionar sobre redes, especialmente de área local. La exclusión se debe a que no proporcionan las principales funciones de transparencia de la distribución y localización de datos, características principales de los SMBDD. Referencias sobre estos sistemas se encuentran en los siguientes autores [Ceri 84], [Escu 85]. Los sistemas homogéneos se subdividen en centralizados y altamente acoplados²:

2) Los sistemas heterogéneos están compuestos por SMBD locales que manejan modelos implantables diferentes. El SMBDD integra las diferentes visiones individuales en una sola por medio del esquema de distribución. No obstante, este objetivo de integración es muy difícil de lograr y le agrega gran complejidad de construcción al sistema manejador, especialmente para el logro de una interfaz común entre usuarios. El nivel de heterogeneidad de estos sistemas, varía según se determinan grados de alejamiento³ entre los nodos integrantes del sistema [Ceri 84], [Date 85]. A nivel de investigación, los sistemas heterogéneos aún presentan problemas para cumplir con todos

² No debe confundirse con el concepto de sistemas altamente acoplados, utilizado en sistemas distribuidos. En dicha área, el concepto se utiliza para caracterizar los sistemas de procesamiento en paralelo (con varios procesadores) que comparten la misma unidad de memoria principal. En este caso se utiliza para definir la integración resultante entre SMBD similares.

³ Relacionados con diversidad de SMBD, modelos implantables de datos, equipos, sistemas de apoyo, protocolos, etc. A medida que hay mas diversidad hay mas alejamiento entre los componentes del sistema distribuido.

los objetivos planteados. Sin embargo, ya existen modelos experimentales que se acercan mucho a lo esperado en el manejo óptimo de dicha heterogeneidad. Como se mencionó, el problema central está en la complejidad de implantación de sus componentes individuales y en la dificultad de encontrar modelos lo suficientemente generales como para soportar cualquier tipo de SMBD local sin degradar el desempeño global del sistema distribuido. Entre los modelos típicos heterogéneos se encuentran, el Multibase, el Sirius-Delta, DDTS, Mermaid y algunos productos comerciales que apoyan la construcción de BDD heterogéneas, tales como Linda y Cics/Iscl. Las referencias bibliográficas [Ceri 84], [Date 85], [Virg 85], [Glig 85], [Temp 85], [Wolk 89] describen detalladamente ambos tipos de sistemas.

3- MODELO GENERAL DE RENDIMIENTO PARA SMBDD

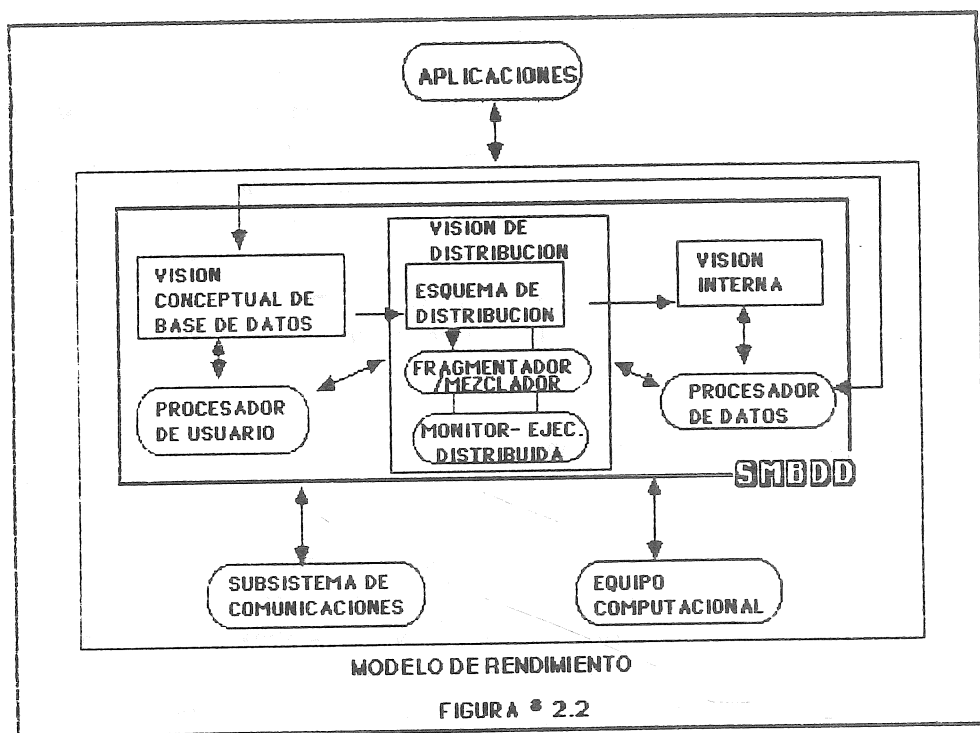
Con el fin de estructurar mejor las ideas seguidas en el estudio, se establece un modelo general de rendimiento que permite mostrar, en forma clara y precisa, los parámetros, relaciones y grado de contribución de cada uno de ellos en el comportamiento (tiempo de respuesta) total del sistema.

Para la construcción de dicho modelo se toma como base la arquitectura mencionada anteriormente y el modelo de rendimiento para bases de datos centralizadas establecido por Ignacio Casas y Kenneth Sevcik [Casa 86]. En la figura anexa, se muestran gráficamente cada uno de sus componentes y las vías de comunicación e interacción entre ellos.

La característica principal del modelo radica en la agrupación de los componentes del SMBDD atendiendo a las funciones lógicas de operación de los componentes que conforman la arquitectura. También se toman en cuenta los factores del medio ambiente (externos) que influyen en el tiempo total de respuesta del sistema. A continuación se describen cada una de las secciones que constituyen el modelo:

3.1-Aplicaciones

Define los parámetros de utilización generados por el ambiente en el cual presta apoyo el sistema distribuido. Se modela en términos de factores o características de los programas que accedan la base de datos distribuida.



Para efectos de rendimiento del sistema, en este grupo se colocan elementos tanto cuantitativos como cualitativos; entre otros están el grado de utilización, tipos de datos manejados, volúmen de datos implicados en cada solicitud y frecuencia de uso de la aplicación. Estos factores son determinados, directamente por el usuario y su ambiente. La influencia de este grupo dentro del comportamiento del sistema, se refleja a través de la traducción de dichas características al modelo conceptual global de datos y a las operaciones realizadas sobre el conjunto de datos; pero muy especialmente, se refleja en el diseño de la base de datos distribuida, ya que determina la forma óptima de distribución, el grado de fragmentación y de replicación de datos; así como también, interviene en la selección de la estrategia distribuida más adecuada en un momento dado.

3.2- Visión Conceptual -esquemas externo y conceptual de la base de datos-

En este grupo del modelo de rendimiento, se ubican los elementos de la base de datos que están relacionados con la representación conceptual y externa del modelo de datos. Dicha representación dependerá del modelo lógico⁴ seleccionado y de características particulares de los manejadores de base de datos locales. Los usuarios de la base de datos pueden tener diferentes representaciones lógicas de los datos, estas son llamadas visiones externas. Por lo tanto, el modelo conceptual debe contemplar la integración de estas visiones individuales y debe proporcionar medios de conversión y manipulación de datos cuando las visiones concentradas difieren mucho unas de otras.

3.3- Visión de Distribución

Relacionada con el esquema de distribución y el conjunto de factores requeridos para la determinación de la estrategia de ejecución distribuida. Por lo tanto, se incluyen en ella el esquema de localización y distribución de datos, el cual es originado por el diseño de la BDD [Ceri 84], [Ceri 87]. Dicho esquema contiene las direcciones de los nodos que almacenan los conjuntos de datos requeridos para satisfacer las diferentes solicitudes de acceso del usuario. Se incluye la descripción de tipo y grado de fragmentación y replicación de datos. Este esquema es utilizado por otro elemento incluido dentro de esta sección, el fragmentador, el cual se encarga de hacer efectiva la ejecución distribuida. El monitor de ejecución distribuida es el encargado de monitorear y controlar el procesamiento distribuido. Debido a su relación con los elementos anteriores es incluido como tercer componente de la visión distribuida. Estos elementos son considerados de gran repercusión sobre el comportamiento esperado en el SMBDD.

3.4- Visión Interna

El esquema interno corresponde al diseño físico de los datos almacenados en cada uno de los nodos locales. Este está estrechamente relacionado con la implantación de la base de datos y con los mecanismos de acceso a dichos datos. En un sistema distribuido son particulares del SMBD local, por lo tanto

⁴ Revisar alguna de las siguientes referencias, en las cuales se describen dichos modelos [Kort 87], [Ever 86].

las estrategias para selección y acceso al conjunto de datos también se incluyen dentro de la visión interna. Dado que corresponde a la representación de los datos en los dispositivos de almacenamiento, su organización efectiva depende de los formatos, agrupamientos, enlaces y códigos de representación que permite un equipo dado.

3.5- Procesador de datos

Se refiere a las características particulares del SMBD local para la realización de las manipulaciones necesarias sobre los comandos físicos y canónicos. Es decir, lo relacionado a programas de conversión de datos. También la validación de consistencia e integridad de su base de datos local y de informar sobre el monitoreo de la ejecución llevada a cabo en su nodo.

Existe una relación directa con el equipo computacional y las facilidades que brinda para el manejo y acceso de datos y archivos a nivel físico. El SMBD debe optimizar el uso del equipo aprovechando sus ventajas para obtener mejor rendimiento de procesamiento.

3.6- Procesador de usuario

Contempla todo lo relacionado con los mecanismos de traducción y formateo de comandos en el nivel de usuario. Su punto clave son los algoritmos de manipulación de consultas. La forma de generar subconsultas a partir de una consulta global tiene fuerte influencia en el rendimiento de todo el sistema. El procesador de usuario (cliente) tiene una estrecha relación con los esquemas conceptual y externo de la base de datos. Y su salida va directamente a la visión distribuida a través de la cual será encaminada apropiadamente.

3.7- Equipo Computacional

Se refiere al equipo físico de apoyo al sistema. Son los procesadores sobre los cuales corren los programas del sistema distribuido. Por lo tanto, de su desempeño depende el grado de respuesta del SMBDD. Se incluyen en este grupo, procesadores, controladores, dispositivos de entrada/salida, de almacenamiento, junto con sus respectivas características particulares de manipulación. También forman parte los programas de apoyo del sistema, tales como sistemas operativos los cuales son básicos para el funcionamiento

normal del sistema computacional. Establecen limitantes fuertes en las implantaciones de SMBD, las cuales pueden extenderse a los SMBDD.

3.8- Subsistema de comunicaciones

Dentro de esta clasificación se colocan los elementos necesarios, tanto de equipo (hardware) como de programas (software), para el transporte de los datos o resultados de procesamiento entre las diferentes localidades. Y además, todos los elementos encargados de las comunicaciones de cualquier otro tipo de mensajes entre los componentes del sistema distribuido.

Las características de implantación y manejo del equipo y protocolos de comunicaciones son excluidas de este grupo. Para los efectos del SMBDD, solo se requieren a nivel de recepción y entrega de datos, por lo tanto sólo se incluyen los detalles referentes a la interfaz de comunicación. Se supone la existencia de un front-end de comunicaciones en cada localidad.

4- DEFINICION DE LAS ECUACIONES BASICAS DE RENDIMIENTO

Los parámetros seleccionados para la evaluación son aquellos que influyen en la medida del tiempo de respuesta. Por lo tanto, el conjunto de parámetros están relacionados con el tiempo de procesamiento ocupado por el SMBDD para la realización de cada una de las actividades relacionadas con la manipulación de la BDD. La suma de todas esas cantidades individuales conforman el tiempo global de respuesta medido para el sistema distribuido. Se utiliza al modelo de rendimiento, definido en la sección anterior, como base para la determinación y agrupamiento de los parámetros de evaluación.

Para derivar las ecuaciones del tiempo de respuesta se toma como base la secuencia de procesamiento que sigue una solicitud de acceso a la base de datos distribuida y el conjunto de parámetros derivados para cada una de los componentes del modelo de rendimiento. De ésta manera, el tiempo total de respuesta se fragmenta en:

- | | |
|----------------------------------|----------|
| 1-Tiempo de visión conceptual | (TVC) |
| 2-Tiempo de visión distribuida | (TVD) |
| 3-Tiempo de procesamiento local | (TPL) |
| 4-Tiempo de equipo computacional | (TEQ) |
| 5-Tiempo de comunicaciones | (TCOM) |

(1)- El (**TVC**) **tiempo de visión conceptual** esta conformado por los factores [Gill89], [Hevn 85], [Lars 85b]:

- tiempo de lectura al esquema conceptual depende del tamaño del esquema **T**, del método de acceso al esquema **A**:

$$t_{lectura} (T, A) \quad (1)$$

- tiempo de revisión de integridad depende del momento en que se realice la revisión, previo, en paralelo o posterior a la ejecución de la solicitud

$$t_{integridad} \quad (2)$$

- tiempo de traducción depende del lenguaje manipulador de datos **LMD** y del algoritmo utilizado para la traducción **AL**

$$-t_{traducción} (AL, LMD) \quad (3)$$

- tiempo de localización de los datos - contención por bloqueo de la base de datos realizada por los mecanismos de control de concurrencia

$$t_{localización} \quad (4)$$

- tiempo de acceso a esquema externo

$$t_{externo} (\text{tamaño del esquema, método de acceso}) \quad (5)$$

- tiempo de formateo de resultados depende del lenguaje manipulador de datos

$$t_{formateo} \quad (6)$$

La ecuación total del tiempo para la visión conceptual es:

$$TVC = t_{lectura} + t_{integridad} + t_{traducción} + t_{localización} + t_{externo} + t_{formateo} \quad (7)$$

(2)- El (**TVD**) **tiempo de visión distribuida** está dado por los siguientes factores [Jark 85], [Daya 85], [Chu 82], [Yu 85]:

- tiempo de acceso distribuido, depende del tamaño del esquema de distribución **TD**, de la cantidad de datos implicados **C**, métodos de acceso **A**

$$t_{acceso\ distribuido} (TD, C, A) \quad (8)$$

- tiempo de estrategia distribuida, depende de los parámetros definidos **C**, de la distancia entre nodos **D**, grado de replicación **R**, número de nodos que contienen datos requeridos **N**, secuencia de operaciones **S** y materialización de las mismas **M**:

$$t_{\text{estrategia distribuida}} (C, N, D, R, S, M)^5 \quad (9)$$

- tiempo de mezcla de resultados, depende del número de operaciones distribuidas OD y, orden de ejecución O. Se determinan dos tiempos importantes:

$$* \text{tiempo de mezcla: } t_{\text{mezcla}} (OD, O) \quad (10)$$

*tiempo de mezcla-real : implica el tiempo que debe esperar para completar la actividad de mezcla de resultados, desde que comienzan a llegar los resultados parciales de los diferentes nodos hasta que obtiene el resultado total en el formato deseado

$$t_{\text{mezcla-real}} = t_{\text{local-último}} - t_{\text{local-primer}} + t_{\text{mezcla}} \quad (11)$$

Así el tiempo de visión distribuida queda definido como:

$$TVD = t_{\text{acceso-distribuido}} + t_{\text{estrategia}} + t_{\text{mezcla-real}} \quad (12)$$

(3)- El (TPL) tiempo de procesamiento local, esta definido por los siguientes factores [Lars 85c], [Jark 85], [Onue 88], [Marc 85]:

-tiempo de acceso a esquema externo

$$t_{\text{acceso-externo}} = t_{\text{externo}} + t_{\text{integridad}} \quad (13)$$

-tiempo conversión de comandos esta es una definición general, luego a nivel de cada SMBDD, se determinan las peculiaridades de cada uno de ellos, depende de los siguientes tiempos

*tiempo de acceso a esquemas internos

$$t_{\text{interno}} (TI, A) \quad (14)$$

*tiempo de manipulación sobre los datos, el cual depende del lenguaje manipulador de datos LMD y de la cantidad de datos implicados C :

$$t_{\text{manipulación}} (LMD, C) \quad (15)$$

$$t_{\text{conversión de comandos}} = t_{\text{interno}} + t_{\text{manipulación}}$$

(16)

-tiempo de acceso a la base de datos física, el cual depende del tamaño real de los archivos internos TBD, de la cantidad de

⁵ En los estudios realizados por [Mukk 88] , [Ceri 84] se definen fórmulas muy específicas para manipular las variables mencionadas. Son válidas únicamente para casos especiales, donde se ha restringido deliberadamente el contexto; para obtener estimaciones mucho más precisas.

datos implicados **C** y de las estructuras de apoyo para la implantación y acceso más rápido a la bd física **EA**.

$$t_{\text{acceso físico}} (\text{TBD, C, EA}) \quad (17)$$

- tiempo de formateo a comandos canónicos, depende directamente de las expresiones canónicas **EC** utilizadas y del lenguaje manipulador de datos **LMD** que se tenga

$$t_{\text{formateo}} (\text{LMD, EC}) \quad (18)$$

$$\text{TPL} = t_{\text{acceso-externo}} + t_{\text{conversión}} + t_{\text{acceso físico}} + t_{\text{formateo}} \quad (19)$$

(4)- El (**TEQ**) **tiempo de equipo computacional** esta dado por los siguientes elementos [Kein 86], [Vija 85], [Papa 86]:

- tiempo de ejecución de transacciones, dado por el ciclo de instrucción del procesador central de la máquina que apoya la ejecución distribuida

$$t_{\text{cpu}} \quad (20)$$

- tiempo de servicio o acceso al disco para acceder la base de datos, depende de los tiempos de rotación, latencia y posicionamiento de la cabeza lectora, así como del número de bloques que lee por cada acceso y, finalmente por el número de veces que debe repetir la actividad para completar una transacción de lectura o escritura sobre la base de datos (**n**). Para los efectos de este estudio se considera como un factor integrado de todos estos elementos:

$$t_{\text{servicio total}} = n * t_{\text{servicio}} \quad (21)$$

- tiempo de retraso por bloqueo del medio, causado por mecanismos de control de concurrencia y de atomicidad de las transacciones

$$t_{\text{bloqueo-disco}} \quad (22)$$

$$\text{TEQ} = t_{\text{cpu}} + t_{\text{servicio total}} + t_{\text{bloqueo-disco}} \quad (23)$$

este tiempo afecta a cada uno de los elementos del modelo que realiza algún tipo de procesamiento.

(5)- El (**TCOM**) **tiempo de comunicaciones** esta definido por elementos tales como [Hopp 89], [Hutc 88]:

- tiempo de transmisión, los factores más relevantes que lo afectan son: el volumen de datos transmitidos **V**, de la distancia **D** que debe recorrerse y la capacidad del canal de comunicaciones **CC**:

$$t_{\text{transmisión}} (V, CC, D) \quad (24)$$

- tiempo de acceso al canal de comunicaciones, depende de los protocolos P que se establecen y de la topología de la red TR que apoya la BDD

$$t_{\text{acceso canal}} (P, TR) \quad (25)$$

- tiempo de retraso introducido por congestión en la red debido a sobrecargas en el medio de comunicación

$$t_{\text{retraso}} \quad (26)$$

- tiempo de servicio de comunicaciones, utilización de front-end de comunicaciones el cual se encarga de todos los factores de interfaz que son ajenos al SMBDD, pero que son necesarios para lograr el procesamiento distribuido. Depende directamente de la velocidad del CPU y de los protocolos P utilizados para lograr la interfaz:

$$t_{\text{ser.comunicación}} (\text{CPU}, P) \quad (27)$$

$$TCOM = t_{\text{transmisión}} + t_{\text{acceso canal}} + t_{\text{retraso}} + t_{\text{ser.comunicación}} \quad (28)$$

Tomando en cuenta las ecuaciones individuales definidas, se establece la ecuación final para el tiempo de respuesta del SMBDD. Para definirlo se ordena en base al ciclo de procesamiento lógico seguido por las solicitudes. El **tiempo total TT** es:

$$TT = [(\text{Tiempo procesador-cliente}) + TCOM + (\text{Tiempo proceso local})] * TEQ \quad (29)$$

El tiempo en procesador cliente, está constituido por los tiempos necesarios para ejecutar las actividades en el nodo procesador de la solicitud de acceso a la base de datos distribuida.

El tiempo de proceso local también está conformado por la suma de los tiempos de las actividades requeridas, para la producción de resultados parciales a la solicitud de usuario.

5- EVALUACION DE RENDIMIENTO PARA CADA UNO DE LOS DIFERENTES TIPOS DE SMBDD [Barr 90]

El modelo anteriormente expuesto, fue ajustado a las características de procesamiento de cada uno de los modelos de SMBDD. Se hicieron las diferenciaciones sobre cada una de las ecuaciones involucradas en dicho procesamiento. Estos modelos particulares fueron sometidos a la influencia

de las aplicaciones, relacionadas con Sistemas de Información (administrativos y de apoyo a la toma de decisiones), Sistemas Integrados (manufactura, control de procesos y de producción) y Sistemas Transaccionales.

De la evaluación se destacan los siguientes puntos:

1- El tiempo de estrategia distribuida, ha sido denotado como el más relevante para la obtención de un tiempo mínimo de respuesta, independientemente del tipo de SMBDD. Sin embargo, su contribución al tiempo total, no está definida como el tiempo mínimo que toma efectuar una decisión estratégica de operación distribuida; sino en el tiempo que dicha decisión ahorra en términos de mayor efectividad en el uso de los recursos que componen el SMBDD. Así, se entiende que si un TVD, es ligeramente alto al compararlo con otros tiempos del modelo, éste se justifica completamente; dado que al producir un efecto reductor en otros tiempos más impactantes, tales como procesamiento local y comunicaciones, el tiempo global de respuesta será menor.

2- El caso contrario, implica que si se dedica menos tiempo a la definición de estrategia, puede ahorrarse tiempo de TVD, pero la falta de optimización requerirá mayor utilización de otros recursos del SMBDD y si estos recursos son limitados, el rendimiento total del sistema irá degradando su valor hasta caer en un estado de completa ineficiencia desde todo punto de vista.

3- Debe quedar muy claro que lo esencial en un SMBDD es la cooperación entre componentes. Debe lograrse el equilibrio entre ellos, para garantizar la supervivencia⁶ del sistema completo, tal como ocurre con cualquier otro tipo de sistema. El rango aceptable de valores para este tiempo total es establecido por las características operativas demandadas por las aplicaciones.

4- Es muy cierto que los factores externos tales como equipo y sistema de comunicaciones son importantes para el logro de un buen rendimiento de SMBDD; pero se han destacado factores que limitan esta

⁶ Dentro de este contexto supervivencia significa utilidad máxima. Proporcionar los beneficios de utilización y operatividad adecuados y establecidos por el ambiente en el cual se encuentra y al cual presta apoyo.

• aceleración del tiempo de respuesta, tales como la estrategia distribuida y optimización de consultas. Los cuales causan una subutilización de recursos, deficiencia en la asignación de tareas que al manejarse en forma aislada sin importar los demás componentes del sistema originan un incremento notable en el tiempo de respuesta del sistema.

Los resultados obtenidos pueden ser resumidos en las siguientes tablas (la nomenclatura utilizada es la misma definida en el modelo de rendimiento):

SMBDD's	SISTEMAS DE INFORMACION	SISTEMAS INTEGRADOS	APLICACIONES TRANSACCIONALES
Hom. Alt. Acoplado	Adecuado	Adecuado	Adecuado
Hom. Centralizado	No - Adecuado	No - Adecuado	Adecuado *
Heterogéneo	Adecuado *	Adecuado	Adecuado *

TABLA 1: SMBDD vs. APLICACIONES

(Adecuado : presenta las características necesarias para que la aplicación funcione apropiadamente.

Adecuado * : puede ser utilizado por la aplicación pero con algunas limitaciones.

No-adequado: no presenta las características básicas requeridas para soportar la aplicación.)

	SISTEMAS DE INFORMACION	SISTEMAS INTEGRADOS	APLICACIONES TRANSACCION.
TVC	RELEVANTE	IMPORTANTE	NORMAL
TVD	RELEVANTE	IMPORTANTE	NORMAL
TPL	NORMAL	NORMAL	NORMAL
TEQ	NORMAL	RELEVANTE	NORMAL
TCOM	NORMAL	RELEVANTE	IMPORTANTE

TABLA 2 : SMBDD Homogéneo Altamente Acoplado

TVC	R	R	I
TVD	I	I	N
TPL	I	I	N
TEQ	N	I	I
TCOM	N	R	I

TABLA 3: SMBDD Heterogeneo

	Central	Servicio	Central	Servicio	Central	Servicio
TVC	I	-	M	I	M	M
TVD	I	-	M	I	M	I
TPL	M	-	M	-	M	-
TEQ	R	I	R	R	M	M
TCOM	M	I	R	R	I	I

TABLA 4: SMBDD Homogéneo Centralizado

- (**Relevante** : afecta fuertemente el tiempo total de respuesta del sistema.
- Importante:** no debe descuidarse es básico en la determinación del tiempo total.
- Normal** : su influencia sobre el tiempo de respuesta no es superior a la de los demás elementos.
- (**No es importante:** influencia mínima sobre el tiempo de respuesta).

6- CONCLUSIONES Y RECOMENDACIONES

1) A través del análisis efectuado se determinó que los factores principales de rendimiento de un SMBDD están dentro de la sección de visión distribuida perteneciente al modelo de evaluación definido.

2) La optimización del acceso remoto se fundamenta en el diseño conceptual de la BDD, por lo tanto la relación de rendimiento es directamente proporcional a la implantación de un buen diseño de la misma.

3) Del punto anterior se concluye que no puede haber un alto rendimiento en un sistema de BDD, cuyo diseño de datos sea deficiente. Las estrategias de optimización de consultas y de ejecución distribuida solo pueden mejorar hasta cierto grado la respuesta del sistema. Si la distribución de los datos no es la óptima para enfrentar las características de la aplicación que la utiliza, el rendimiento será minoritario o al menos no podrá superar ciertos límites establecidos por otros parámetros del ambiente en el cual se encuentra el sistema de base de datos.

4) La influencia de los medios de comunicación no es tan fuerte en el rendimiento de los SMBDD. Aún contando con un canal de comunicación de gran capacidad, la estrategia puede determinar un tiempo de respuesta bajo si esta hace un uso excesivo de los medios de comunicación.

Mientras no se saturan los canales de comunicación la influencia en el tiempo de respuesta es solamente determinada por el número de veces que estos se utilicen. La limitación en el tiempo de respuesta ocasionada por los medios de comunicación, solo se alcanza cuando los canales no son apropiados o no son capaces de manipular el volumen y tráfico de datos requerido por las aplicaciones.

5) Las diferencias, con respecto a los parámetros de evaluación de rendimiento entre los sistemas homogéneos y los heterogéneos, se concentra en hacer mas énfasis en la optimización de consultas. La diferencia real se establece a nivel de implantación o construcción de los mecanismos que garanticen un funcionamiento transparente del sistema ante los usuarios. La complejidad está en la forma de lograrlo, no en la forma de evaluarlo.

6) El estudio presentado a través del modelo de rendimiento permitió identificar un conjunto de elementos individuales que contribuyen a la obtención de un tiempo global de respuesta. Las relaciones entre dichos componentes no ha sido exhaustivamente estudiada y analizada aquí, por lo que constituyen un punto de inicio para estudios posteriores.

7) A través de modelos de simulación se validaron algunos de los resultados teóricos señalados en el estudio. Actualmente se está trabajando en la validación y profundización de los conceptos expuestos en el modelo.

7- BIBLIOGRAFIA

[Barr 90] Modelo para la evaluación de rendimiento de Sistemas Manejadores de Base de Datos Distribuidas. Tesis de Maestría. ITESM. Monterrey, Mexico 1990.

[Bhar 89] Bhargava Bharat; Mafla Enrique; Riedl Jonh; Sauder Bradley. Implementation and measurements of efficient Communication Facilities for Distributed Database Systems. IEEE Fifth International Conference on Data Engineering 1989.

[Bhid 88] Bhide Anupam.; Stonebraker Michael. A performance comparison of two architectures for fast transaction processing. IEEE Fourth International Conference on Data Engineering 1988.

[Bodo 88] Bodorik P.; Riordon J. Distributed query processing optimization objectives. IEEE Fourth International Conference on Data Engineering 1988.

[Casa 86] Casas Ignacio; Sevcik Kenneth. Performance modeling of Database Systems: a framework and a case of study. Computer Systems Research Institute, University of Toronto, Canada. Tech. Report. 1986.

[Casa 89] Casas Ignacio; Sevcik Kenneth. A buffer Management Model for Use in Predicting Overall Database System Performance. IEEE Fifth International Conference on Data Engineering 1989.

- [Ceri 84] Ceri Stefano, Pelagatti Giuseppe. Distributed Databases. Principles & systems Mc Graw Hill 1984.
- [Ceri 87] Ceri Stefano; Percini Barbara; Widerhold GIo. Distributed Database Systems 1987.
- [Chu 82] Chu W. W.; Hurley P. Optimal query processing for distributed database systems. IEEE Transactions on computers. September 1982.
- [Date 85] Date, C. J. An introduction to Database systems. Volumen II. Addison Wesley 1985.
- [Daya 85] Dayal Umeshwar. Query processing in a multidatabase system. Query processing in Database Systems. Editado por Kim Won, Reiner David, Batory Don. 1985.
- [Ever 86] Everest Gordon. Database management objectives, systems functions & administration McGraw Hill. Computer sciences series. 1986.
- [Escu 85] Esculier Christian. The Sirius -Delta architecture: a framework for co-perating database systems. Tutorial : Distributed database management. IEEE Computer Society 1985.
- [Garc 88] Garcia Molina Hector; Kabbott Robert. Reliable distributed database management. Proceedings of IEEE. May 1987.
- [Gill 89] Gillenson, Mark. Introducción a las base s de datos. McGraw Hill 1989.
- [Hevn 85] Hevner Alan. Methods for data retrieval in distributed systems. Tutorial. Distributed database management. IEEE computer Society 1985.
- [Hevn 87] Hevner Alan. R., Yao S. Bing. : Quering distributed database on local area network. Proceedings of the IEEE. May 1987.
- [Hopp 89] Hopper; Temple; Williamson. Diseño de redes locales. Adisson Wesley. Iberoamericana S.A. 1989.
- [Hutc 88] Hutchinson David. Local area network architectures. International computer science series. Addison Wesley 1988.
- [Jark 85] Jarke Matthias; Jurgen Koch; Schmidt Joachim. Introduction to query processing. Query processing in Database Systems. Editado por Kim Won, Reiner David, Batory Don. 1985.
- [Kein 86] Kleinrock Leonard. Performance models for distributed systems. Teletraffic Analysis and computer performance evaluation. Elseviers Science Publisher B.V. 1986.

- [Kort 87] Korth, Henry; Siverschatz, Abraham. Fundamentos de base de datos. McGraw Hill. 1987.
- [Lai 85a] Larson James. Introduction. Tutorial. Distributed database management. IEEE computer Society 1985.
- [Lars 85b] Larson James. Bringing the gap between network and relational database. Special feature. Tutorial. Distributed database management. IEEE computer Society 1985.
- [Lars 85] Larson James; Rahimi Saeed. Transforming database command. Tutorial. Distributed database management. IEEE computer Society 1985.
- [Marc 85] March Salvatore; Carlis John. Physical Database Design: Techniques for improved Database Performance. Query processing in Database Systems. Editado por Kim Won, Reiner David, Batory Don. 1985..
- [Mukk 88] Mulkamala Ravi. Measuring the effect of data distribution and replication models on performance evaluation of distributed database systems. IEEE Fourth International Conference on Data Engineering 1988.
- [Nich 88] Nicholls Bill. The "B" world. BYTE June 88.
- [Onue 88] Onuegbu Emmanuel; Rahimi Saeed; Hevner Alan. Local query translation and optimization. IEEE Fourth International Conference on Data Engineering 1988.
- [Papa 86] Papadimitrou Christos. The theory of Database Concurrency Control. Computer Science Press. 1986.
- [Robi 88] Robie Jonathan. Fast data access. Byte January 1988.
- [Vija 85] Vijay Ahuja. Design and analysis of Computer Communication Networks. International Student Edition. McGraw Hill 1985.
- [Virg 85] Virgil Gligor; Luckenbaugh Gary. Interconnecting heterogeneous database management systems. Tutorial Distributed database management. IEEE computer Society 1985.
- [Youn 88] Young Jeffrey; Warren Carl; Mac user Lab staff. Database comparison. MacUser December 1988.
- [Yu 85] Yu Clement. Distributed query processing. Query processing in Database Systems. Editado por Kim Won, Reiner David, Batory Don. 1985.