(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 326 538**

(51) Int. Cl.:
H04L 29/08 (2006.01)
H04L 12/56 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Número de solicitud europea: **05103034 .4**
(96) Fecha de presentación : **15.04.2005**
(97) Número de publicación de la solicitud: **1603307**
(97) Fecha de publicación de la solicitud: **07.12.2005**

(54) Título: **Sistema y método para la gestión de las características de funcionamiento en un entorno informático de varios niveles.**

(30) Prioridad: **04.06.2004 US 576805 P**

(45) Fecha de publicación de la mención BOPI:
14.10.2009

(45) Fecha de la publicación del folleto de la patente:
14.10.2009

(73) Titular/es: **OpTier Ltd.**
Abba Hillel Silver 14, Beit Oz, 6th Floor
Ramat Gan 52506, IL

(72) Inventor/es: **Alon, Amir;**
Lavi, Yoram Yori y
Tal, Mordechai

(74) Agente: **Tomás Gil, Tesifonte Enrique**

DESCRIPCIÓN

Sistema y método para la gestión de las características de funcionamiento en un entorno informático de varios niveles.

Campo de la técnica descrita

La técnica descrita se refiere a administrar entornos informáticos distribuidos en general, y a un sistema y un método para la gestión de rendimiento de la aplicación en un entorno informático multinivel, en particular.

Antecedentes de la técnica descrita

La tecnología de la información (IT) es responsable de entregar servicios de aplicación usando un medio de producción multinivel cada vez más complejo con una mezcla de aplicación heterogénea. Las operaciones de IT están luchando para encontrar los niveles de servicio requeridos en cuanto a rendimiento y disponibilidad, mientras que son agilizadas para aumentar la eficiencia y utilización de los recursos. La consolidación de los recursos IT, junto con la empresa comercial, exacerba este efecto, extendiendo la capacidad de las operaciones IT para encontrar cualquier cambio en las peticiones para recursos informáticos. Los enfoques tradicionales y herramientas para el rendimiento y gestión de disponibilidad son variaciones del ciclo perpetuo "control-tono-fijo", que implica la identificación de que existe un problema (es decir, control), aumentar el rendimiento global para superar el problema (es decir, sintonía), y realizar el análisis de la causa primera para descubrir la causa precisa de cada ejemplo específico de un problema (es decir, fijación). Los enfoques de este tipo son incapaces de afrontar la complejidad y variabilidad del medio IT en continuo cambio.

Ahora se hace referencia a la figura 1, que es una ilustración esquemática de un entorno informático multinivel, generalmente referenciados con el 50, que es conocido en la técnica. El entorno informático 50 incluye un primer cliente 62 que ejecuta una primera aplicación, un segundo cliente 64 que ejecuta una segunda aplicación, un primer nivel 52, un segundo nivel 54, un tercer nivel 56, un cuarto nivel 58, y un quinto nivel 60. El primer nivel 52 es un servidor web. El segundo nivel 54 es un servidor de aplicación, servidor de aplicación A. El tercer nivel 56 es otro servidor de aplicación, servidor de aplicación B. El cuarto nivel 58 es otro servidor de aplicación, servidor de aplicación C. El quinto nivel 60 es una base de datos. El primer nivel 52 es acoplado al primer cliente 62, con el segundo cliente 64, y con el segundo nivel 54. El segundo nivel 54 es posteriormente acoplado al tercer nivel 56 y al cuarto nivel 58. El tercer nivel 56 es posteriormente acoplado al cuarto nivel 58 y al quinto nivel 60. El cuarto nivel 58 es posteriormente acoplado al quinto nivel 60.

Un "nivel" representa un tipo determinado de procesamiento que es parte de la entrega global de un servicio IT (p. ej., procesamiento de niveles de presentación en un nivel de servidor web o procesamiento de datos informática en un nivel de base de datos). Cada nivel normalmente se ejecuta en una máquina huésped diferente.

La primera aplicación inicia una petición del usuario R1 y manda la petición del usuario R1 al primer nivel 52. La petición del usuario R1 es parte de una transacción global iniciada por el usuario. La petición del usuario R1 puede ser, por ejemplo, una web basada en preguntas para recuperar información de una aplicación determinada. La petición del usuario R1 puede requerir los servicios de diferentes niveles en el entorno informático 50 y puede generar peticiones adicionales para obtener estos servicios. El nivel que recibe una petición sea contesta al nivel que envía la petición, o sea manda una nueva petición a un nivel diferente. Finalmente una respuesta es devuelta en respuesta a la petición del usuario original R1. Un nivel dado sólo puede pedir un servicio de otro nivel en el entorno informático 50 si los dos niveles son directamente acoplados entre sí.

La gestión global del entorno informático distribuido 50 requiere el conocimiento de cómo procesa cada nivel su carga de trabajo. Por ejemplo, si se da una escasez de recursos en un nivel, un administrador del sistema puede escalar este nivel creando clones del nivel, tanto verticalmente (es decir, dentro de la misma máquina huésped) como horizontalmente (es decir, a través de máquinas huéspedes múltiples). Por ejemplo, en el entorno informático 50, el administrador del sistema puede añadir un servidor de aplicación adicional A2 (no mostrado) al servidor de aplicación A del segundo nivel 54, donde el servidor de aplicación A2 es un clon del servidor de aplicación A. Por la misma señal, si existe una sobreabundancia de recursos en un nivel, el administrador de sistema puede transferir recursos libres a otro nivel que tenga una escasez de recursos. El administrador del sistema puede además configurar un nivel determinado para mejorar el rendimiento global o indicar modificaciones para optimizar la ejecución de la aplicación en el nivel. Éste es un ejemplo de control de la aplicación específica de nivel para la gestión de rendimiento. Se observa que una petición puede alcanzar sólo determinados niveles en el entorno informático 50. Además, la misma petición puede alcanzar determinados niveles usando recorridos múltiples. Por ejemplo, en entorno informático 50, una petición puede alcanzar la base de datos del quinto nivel 60 mediante servidor de aplicación B del tercer nivel 56, o mediante el servidor de aplicación C del cuarto nivel 58. Como los recorridos de petición no son constantes a través del medio entero, la resolución de la escasez de recursos en un nivel no garantiza necesariamente el rendimiento de la aplicación global, que puede extenderse a múltiples niveles. Un cuello de botella de procesamiento en cualquier nivel retrasa todas las funciones de aplicación que dependen de ese nivel.

El primer nivel 52 recibe la petición del usuario R1. El primer nivel 52 asigna un enclave de procesamiento X1 para procesar la petición del usuario R1. Mientras que se procesa la petición del usuario R1, la lógica de aplicación

que se está ejecutando en el enclave de procesamiento X1 determina que no puede completar el procesamiento de la petición del usuario R1 sin información u operaciones adicionales que han de ser proporcionadas por el segundo nivel 54. El primer nivel 52 manda entonces una petición posterior R2 al segundo nivel 54, solicitando la información u operaciones adicionales. El segundo nivel 54 asigna un enclave de procesamiento X2 para procesar la petición R2. La lógica de aplicación que se está ejecutando en el enclave de procesamiento X2 determina que la petición R2 requiere información u operaciones adicionales que han de ser proporcionadas por el cuarto nivel 58. El segundo nivel 54 manda entonces una petición posterior R3 al cuarto nivel 58. El cuarto nivel 58 asigna un enclave de procesamiento X4 para procesar la petición R3.

El enclave de procesamiento X4 completa la ejecución. El cuarto nivel 58 devuelve una respuesta R3' al segundo nivel 54, en respuesta a la petición anterior R3 del segundo nivel 54. El enclave de procesamiento X2 recibe la respuesta R3' y recupera el procesamiento. Un vez que el enclave de procesamiento X2 ha completado su ejecución, el segundo nivel 54 devuelve una respuesta R2' al primer nivel 52, en respuesta a la petición anterior R2 del primer nivel 52. El enclave de procesamiento X2 recibe la respuesta R2' y resume el procesamiento. Una vez que el enclave de procesamiento X1 ha completado su ejecución, el primer nivel devuelve una respuesta R1' a la petición del usuario R1, cuyo servicio ahora ha sido completado.

En el entorno informático 50, cada uno de los diferentes niveles se aíslan de los niveles que no están acoplados directamente entre sí. Por ejemplo, la petición R3 del segundo nivel 54 al cuarto nivel 58, directamente acoplado con él, no incluye necesariamente información acerca de una petición anterior R2, que fue recibida en el segundo nivel 54 desde el primer nivel 52, ni la petición R3 incluye información acerca de la petición del usuario R1. Un nivel dado no tiene manera de obtener determinada información relacionada con la petición que es procesada en ese nivel, tal como que el usuario inició la transacción, estas peticiones precedieron la petición que está siendo procesada en el nivel dado, o características de las peticiones que precedieron esa petición. Por ejemplo, el segundo nivel 54 no puede identificar características de petición R2, como si la petición fuera precedida por la petición del usuario R1 enviada a la primer nivel 52, o que la transacción se originara a petición del usuario R1 de la primera aplicación 62. Como resultado, si un nivel de prioridad es asignado a un enclave de procesamiento que procesa una petición dentro de un nivel determinado, ese nivel de prioridad es asignado teniendo en cuenta sólo la información mínima que está disponible en el nivel. Esta información incluye las características de petición (p. ej., los credenciales de registro de nivel usados por la petición) y quizás información sobre el enclave de procesamiento que procesa esa petición (p. ej., la identificación de sesión de la base de datos). Las peticiones son generalmente procesadas en una base de igual prioridad (p. ej., el que llega primero, se sirve primero), aunque los mecanismos operativos para diferenciar los niveles de prioridad están disponibles localmente en un nivel dado. La gestión de rendimiento debe ser hecha en una base de nivel individual, puesto que el otro nivel en entorno informático 50 no puede ser contabilizado cuando se trata de un nivel específico. Normalmente, un administrador del sistema que es responsable de administrar un entorno informático multinivel tal como el entorno informático 50 intenta mejorar el rendimiento ajustando la asignación de recursos a un nivel dado.

La patente estadounidense nº. 5,958,010 de Agarwal *et al*. titulada "Systems and methods for monitoring distributed applications including an interface running in an operating system kernel", se refiere a sistemas y métodos para controlar la operación amplia en una empresa de un sistema informático distribuido para desarrollar datos de gestión de nivel de transacción de negocios para el rendimiento del sistema, tendencias de uso, auditoria de seguridad, planificación de capacidad, y excepciones. Un sistema que tiene una arquitectura informática distribuida incluye estaciones de trabajo múltiples, servidores, y dispositivos de red. Cada estación de trabajo es representativa de una configuración acoplada a una red. Cada estación de trabajo es capaz de solicitar servicio de cualquiera de los servidores. Cada estación de trabajo tiene una pila de comunicación para intercambiar datos con la red. El sistema además incluye una pluralidad de agentes de control, y un módulo de consola con una base de datos conectada a él. Cada agente de control tiene una interfaz de evento externo que proporciona información de eventos sobre distintos componentes de una empresa. Cada uno de los agentes de control está unido con su respectivo de las estaciones de trabajo o servidores.

El agente de control puede residir físicamente en el cliente asociado o servidor del mismo. Los agentes de control controlan y recogen datos que son intercambiados entre un cliente y la red, y entre un servidor y la red. Cada agente de control puede ser un módulo de software, un dispositivo de hardware, o una combinación de los mismos. Cada agente de control pasa información representativa de los datos recogidos al módulo de consola. El módulo de consola almacena esta información dentro de la base de datos para su análisis por un operador. Un programa de aplicación que se ejecuta en el módulo de consola puede ver los datos recogidos para mostrar el rendimiento del sistema de cualquier proceso o componente de la empresa. Un administrador de sistema puede desarrollar estadísticas de uso de nivel de empresa y tiempos de respuesta, desarrollar tablas e informes, y desempeñar otro análisis de datos pertinente para determinar estadísticas definidas de usuario pertinentes a la operación de la empresa.

La patente estadounidense nº. 6,108,700 de Maccabee *et al* titulada "Application end-to-end response time measurement and decomposition", se refiere a un método y sistema para medir y reportar la disponibilidad y el rendimiento de las transacciones comerciales de extremo a extremo. El sistema funciona en una arquitectura de aplicación cliente-servidor. El sistema incluye tres componentes lógicos: generación de eventos, generación de transacciones, y generación de informes, al igual que control del servicio global mediante la administración del sistema.

El componente de generación de eventos existe en todos los ordenadores que son medidos en la arquitectura. Cada ordenador tiene un agente, una pluralidad de sensores y una pluralidad de procesadores. Los sensores interactúan con componentes de la plataforma sobre los cuales tienen lugar las aplicaciones de negocios, se controlan las actividades

de aplicación, y se detectan cambios de estado. Cuando es apropiado, cada uno de los sensores genera un evento que describe el cambio de estado, cuándo y dónde ocurrió el evento, y cualquier dato extra necesario para identificar únicamente el evento. Un evento contiene una marca de tiempo y los datos de correlación usados más tarde por el sistema para asociar el evento con otros eventos en las transacciones. Los sensores envían los eventos generados a sus respectivos agentes. Los agentes almacenan temporalmente los datos y pueden distribuir los datos a otros componentes del sistema que han registrado interés en el evento. Un procesador analiza los eventos, además de deducir los cambios en estado. Los cambios en estado pueden estar directamente relacionados con acciones que ocurren dentro de los componentes de la plataforma de transacciones de negocios o derivados combinando los eventos previamente generados de sensores u otros procesadores para describir los estados conseguidos. Los procesadores envían los eventos generados a sus respectivos agentes.

El componente de generación de transacciones normalmente existe en uno de los ordenadores en la red e incluye un director. El director recibe eventos de los agentes bajo control del mismo. Los eventos son examinados, y correlacionados y verificados en transacciones basadas en reglas de generación de transacciones. El administrador del sistema determina qué transacciones generar.

El componente de generación de informes incluye un administrador. El administrador recolecta las transacciones de los directores. Las transacciones recogidas son manipuladas para obtener información acerca de la disponibilidad y el rendimiento de las transacciones comerciales. Un informe o gráfico continuo puede ser producido sobre una petición específica o periódica de una interfaz gráfica de usuario (GUI). La generación del informe incluye una definición de la selección inicial y procesamiento de transacciones, al igual que la clasificación y métodos de agregación usados para consolidar los datos de evento de transacciones en disponibilidad e información de rendimiento.

La solicitud de patente estadounidense nº. 2002/0129137 A1 de Mills *et al*. titulada “Method and system for embedding correlated performance measurements for distributed application performance decomposition”, se refiere a técnicas para insertar mediciones de rendimiento correlacionadas en transacciones asociadas con una aplicación distribuida. Las técnicas son usadas conforme a la descomposición de rendimiento de la aplicación. Los datos son introducidos en un protocolo de comunicaciones usado para llevar a cabo una transacción entre componentes de aplicación en una red de computación distribuida, mejor que alterando los mismos datos de transacción actuales. Los datos introducidos pueden incluir una marca temporal y datos de medición de la duración. El formato de los datos introducidos combina un prefijo clave bien definido con un sufijo variable que identifica la fuente de señales de sincronización, seguida de unos puntos suspensivos y espacios en blanco, y seguidos del sello del tiempo e información de la duración.

Los estadios de procesamiento posterior de la aplicación distribuida pueden interpretar el protocolo de comunicaciones para recabar duraciones de procesamiento de estadios precedentes, para tomar decisiones que consideren el procesamiento de la transacción. La información de la medición es introducida dentro de la misma aplicación distribuida descrita por la información de medición, de modo que la finalización de la transacción ocurre de forma simultánea o contemporánea con disponibilidad de conocimiento de las características del rendimiento de transacción.

Un protocolo de comunicaciones posible es el Protocolo de transferencia de Hipertexto (HTTP). Una red de computación posible distribuida es la Red Global Mundial (WWW). Los componentes de la aplicación pueden ser una aplicación de cliente que tiene lugar en un cliente y una aplicación de servidor que tiene lugar en un servidor de aplicación. Por ejemplo, la aplicación de cliente es un navegador web, y la aplicación de servidor tiene lugar en un servidor web. Una transacción de aplicación es la aplicación del cliente que solicita contenido del servidor de aplicación y el servidor de aplicación que responde. La información de rendimiento es generada para medir el tiempo de respuesta desde la perspectiva de la aplicación del cliente, al igual que para descomponer el tiempo de respuesta en el tiempo tomado por la aplicación del servidor al servicio la petición y generar una respuesta. En particular, las líneas son agregadas a las cabeceras HTTP para llevar datos de medición de rendimiento, permitiendo al cliente recibir la duración de medición del servidor en la cabecera de respuesta HTTP.

El documento de la técnica anterior US 2004/0006602 expone una arquitectura de sistema que comprende un proxy de web y uno o más servidores web. El sistema habilita aplicaciones para optimizar su tráfico.

Resumen de la técnica descrita

Conforme a la técnica descrita, se proporciona por consiguiente un aparato para controlar un nivel seleccionado en un entorno informático multinivel. El aparato incluye un agente de contexto y una extensión del nivel dinámico.

Así, la invención proporciona un aparato según la reivindicación 1 que es un aparato de control de nivel medio en el entorno informático multinivel.

Así, la invención también proporciona un aparato como se reivindica en la reivindicación independiente 23 que es un aparato de control de nivel inicial en el entorno informático multinivel.

Conforme a la técnica descrita, se proporciona además un sistema para la gestión de rendimiento de la aplicación en un entorno informático multinivel incluyendo una pluralidad de niveles. El sistema incluye, para cada uno de al menos dos niveles controlados de la pluralidad de niveles, una respectiva extensión de nivel dinámico y un agente de contexto

respectivo. La extensión de nivel dinámico es acoplada a al menos un puerto de entrada de petición del nivel controlado. El agente de contexto es acoplado con la extensión de nivel dinámico y con otros agentes de contexto asociados a los niveles que son directamente acoplados con el nivel controlado. La extensión de niveles dinámicos controla el tráfico de peticiones que pasa a través del nivel seleccionado, el tráfico de peticiones controladas incluyendo al menos una petición de introducción recibida en un puerto de introducción de petición de un nivel contiguo. La extensión del nivel dinámico identifica cada petición en el tráfico de peticiones controladas y manda al menos el identificador de peticiones al agente de contexto. El agente de contexto también recibe información acerca del contexto de peticiones de la petición de entrada desde el agente de contexto asociado al nivel contiguo. El agente de contexto asocia la información acerca del contexto de la petición de entrada con la petición de entrada, conforme al identificador de peticiones recibidas. El sistema además incluye un servidor de gestión de red de contexto. El servidor de gestión de red de contexto es acoplado a los agentes de contexto. El servidor de gestión de red de contexto recoge y analiza los datos de rendimiento recibidos de los agentes de contexto.

Conforme a la técnica descrita, se proporciona un método adicional para la gestión de rendimiento de la aplicación en un entorno informático multinivel incluyendo una pluralidad de niveles. El método incluye, para cada uno de al menos dos niveles controlados de la pluralidad de niveles, el procedimiento de recepción de información acerca del contexto de la petición de al menos una petición de entrada, la información incluyendo al menos un identificador de petición y un identificador de transacción. El método incluye adicionalmente el procedimiento de controlar el tráfico de peticiones que pasa a través del nivel controlado, el tráfico de peticiones gestionadas incluyendo al menos la petición de entrada. El método además incluye los procedimientos de identificación de la petición de entrada conforme al identificador de peticiones, y asociación de la petición de entrada con una transacción conforme al identificador de transacciones.

Conforme a la técnica descrita, se proporciona adicionalmente otro método para la gestión del rendimiento de aplicación en un entorno informático multinivel incluyendo una pluralidad de niveles. El método incluye, para cada uno de al menos dos niveles controlados de la pluralidad de niveles, el procedimiento de controlar el tráfico de peticiones que pasa a través del nivel controlado, el tráfico de peticiones controladas incluyendo al menos una petición de entrada y una petición de salida, la petición de salida enviada desde el nivel controlado a un nivel contiguo. El método además incluye los procedimientos de determinar la información acerca del contexto de la petición de la petición de entrada, y de identificar cada petición en el tráfico de peticiones controladas. El método además incluye los procedimientos de asociar la petición de entrada con la petición de salida, y el envío de información acerca del contexto de la petición de la petición de salida a un agente de contexto asociado en el nivel contiguo.

Breve descripción de los dibujos

La técnica descrita se entenderá y apreciará de forma más completa por la siguiente descripción detallada tomada conjuntamente con los dibujos donde:

La figura 1 es una ilustración esquemática de un entorno informático multinivel, que es conocido en la técnica;

La figura 2 es una ilustración esquemática de un sistema de gestión de rendimiento de la aplicación, construido y operativo conforme a una forma de realización de la técnica descrita;

La figura 3 es una ilustración esquemática de transmisión de información entre dos de los agentes de contexto del sistema de la figura 2;

La figura 4 es una ilustración esquemática de un ciclo de vida de la petición de muestra sobre dos de los niveles del sistema de la figura 2;

La figura 5 es una ilustración esquemática de un sistema de gestión de rendimiento de aplicación, construido y operativo conforme a otra forma de realización de la técnica descrita;

La figura 6 es una ilustración esquemática de dos de los niveles del entorno informático multinivel de la figura 5;

La figura 7 es una ilustración esquemática de una extensión de nivel dinámico del sistema de la figura 5;

La figura 8 es una ilustración esquemática de un agente de contexto del sistema de la figura 5;

La figura 9 es un diagrama de bloques que demuestra los estadios implicados para capturar un contexto petición y procesamiento posterior, operativos conforme a otra forma de realización de la técnica descrita;

La figura 10 es un diagrama de bloques que demuestra los estadios implicados para capturar una asignación UOW en un nivel local del sistema de la figura 5 y asociar una petición con la UOW, operativos conforme a otra forma de realización adicional de la técnica descrita; y

La figura 11 es un diagrama de bloques que demuestra los estadios implicados para capturar una petición de salida enviada a un nivel remoto del sistema de la figura 5, y asociar la petición enviada con el contexto de la petición, operativos conforme a otra forma de realización adicional de la técnica descrita.

Descripción detallada de las formas de realización

La técnica descrita supera las desventajas de la técnica anterior suministrando un sistema y método para la gestión de rendimiento de la aplicación en un entorno informático multinivel. El sistema controla los puertos de entrada de petición y los puertos de salida de petición de cada nivel, y detecta la entrada o salida de peticiones hacia o desde un nivel dado, mediante una pluralidad de agentes de contexto. Cada agente de contexto está unido con un nivel en el entorno informático multinivel, y es capaz de comunicarse con otros agentes de contexto. Un agente de contexto recoge información acerca de la ejecución de pedidos en el nivel asociado con el mismo. El agente de contexto identifica el contexto de la petición de una petición del usuario. El agente de contexto clasifica la petición del usuario en una clase de petición. El agente de contexto transfiere las características de una petición que existe en el nivel asociado con el mismo, a un agente de contexto posterior asociado al nivel al que se envía la petición.

El agente de contexto asocia una petición con una petición del usuario y con otras peticiones precedentes en la misma transacción. El agente de contexto asigna una clase de servicio a la petición conforme a la clase de petición y un servicio de política activo almacenado localmente. El agente de contexto puede realizar una intervención para influir en el procesamiento de la petición, tal como ajustar el orden de la petición en la cola en un puerto de entrada de petición al nivel, alterar la prioridad de un enclave de procesamiento que ejecuta la petición, alterar el tipo de procesamiento de un enclave de procesamiento que ejecuta la petición, instruir el nivel para asignar, o para denegar, recursos computacionales (p. ej. unidad de procesamiento central - CPU, memoria, y similares) para procesar la petición, poner la petición en retención y liberar el enclave de procesamiento, o terminar la petición. Un servidor de gestión de red de contexto puede perfilar el comportamiento de diferentes tipos de peticiones a través de diferentes niveles y pueden establecer una política de clase de servicio de nivel cruzado apropiado. De este modo, el sistema proporciona al contexto una gestión de recurso relacionado a un nivel de transacción, a través de los diferentes niveles en el entorno informático multinivel.

La técnica descrita proporciona la capacidad de gestión de la carga de trabajo de transacción preventiva a través de todos los niveles en una cadena de infraestructura IT. El sistema integra con la infraestructura IT los niveles, tales como red, aplicación, base de datos y servidores de software personalizados. El sistema automáticamente perfila cargas de trabajo, ayuda a clasificar la carga de trabajo, y habilita a un usuario para crear políticas de rendimiento de clases de servicios apropiadas. El sistema aplica continuamente estas políticas para transacciones a través del nivel en el entorno informático. El sistema utiliza la infraestructura IT existente y aumenta la infraestructura IT existente para permitir la entrega equilibrada de servicios a niveles de servicios óptimos consistentes con intereses de negocio. Los términos siguientes son usados en todas partes de la descripción de las formas de realización:

El término “nivel” aquí abajo, se refiere a una entidad que entrega un determinado tipo de servicio, donde el servicio es parte de la entrega global de una transacción IT. El servicio puede ser el procesamiento de niveles de presentación en un nivel de servidor web, la funcionalidad de la aplicación en un nivel de servidor de aplicación, el procesamiento de datos en un nivel de base de datos, y similares. Cada nivel normalmente se encuentra en una máquina huésped diferente, aunque puede haber más de un nivel operativo en una única máquina huésped, y un único nivel puede incluir componentes múltiples que residen en más de una máquina huésped. La máquina huésped sobre la que se encuentra al menos un nivel, es denominada aquí abajo como un “nivel huésped”. Algunos ejemplos de un nivel incluyen pero no se limitan sólo a eso: un Java 2 Platform, ejemplo de servidor de aplicación Enterprise Edition (J2EE); un cluster de ejemplos de servidor de aplicación J2EE; un ejemplo de servidor de base de datos incluyendo los componentes de acceso al servidor de base de datos tal como controladores de Conectividad de Base de Datos Java/Conectividad Abierta de Base de Datos (JDBC/ODBC); una base de datos de clusters, y similares.

El término “transacción” representa un único proceso iniciado por un usuario, tal como una fase de un proceso de negocio dentro de una aplicación de negocio. Un ejemplo de una transacción es la asignación de una puja en un servicio de subasta online o la abertura de una cuenta de un cliente nuevo en una institución financiera. Una transacción se compone de una cadena de peticiones entre niveles, partiendo con una petición del usuario. En consecuencia cada petición es únicamente asociada a una petición del usuario (es decir, la petición del usuario de la transacción). Cada transacción es identificada mediante un único identificador, conocido como una “identificación de la transacción”. Se observa que una “serie de transacciones relacionadas” se refiere a diferentes transacciones que son interrelacionadas (p. ej., cada transacción representa estadios diferentes de un único proceso para la empresa). La manipulación de una petición dentro de una transacción puede tener en cuenta no sólo la transacción, sino también la serie de transacciones relacionadas a la que pertenece la petición.

El término “petición” aquí abajo, se refiere a una petición de sistema de un nivel a otro nivel, para proporcionar un servicio determinado que es parte de la transacción. Cada petición se identifica mediante un único identificador, conocido como una “identificación de la petición”. Cada petición resulta en una unidad de trabajo (UOW) en el nivel invocado. Algunos ejemplos de una petición incluyen, pero no se limitan sólo a estos: un navegador web de cliente que emite una petición de Protocolo de Transferencia de Hipertexto (HTTP) a un servidor web; un programa java que emite una llamada de Invocación de Método Remoto (RMI) a un servidor de aplicación; una sesión de servidor de aplicación J2EE que invoca un bean de entidad en un servidor de aplicación remoto (vía RMI), y similares.

El término “petición de usuario” aquí abajo, se refiere a la petición inicial iniciada por un usuario o por una aplicación, que se origina en un nivel no vigilado por la técnica descrita. La petición de usuario es la primera petición en la cadena de peticiones que completa una transacción. La cadena de peticiones puede ser representada como una estructura con la petición del usuario en el nodo de la raíz del árbol.

El término “UOW” aquí abajo se refiere al código de aplicación que se ejecuta en el enclave de procesamiento asignado a la petición aplicable en ese nivel (es decir, una invocación UOW). Una UOW está unida con una fuente y un destino, puede tener parámetros (que son directivas para el comportamiento del código de aplicación), y recursos de niveles de usos dentro de un único nivel.

El término “enclave de procesamiento” aquí abajo, se refiere a cualquier thread, sub-proceso, sesión de base de datos, y similares, que ejecuta una UOW en un nivel dado. Una petición es puesta a la cola en el nivel hasta que un enclave de procesamiento disponible es asignado y el código de aplicación (es decir, una UOW) es asignado al enclave de procesamiento. Las unidades del enclave de procesamiento son ejecuciones genéricas que a su vez ejecutan diferentes códigos de aplicación.

El término “contexto de la petición” aquí abajo se refiere a un grupo de características que son inicialmente capturadas de la petición de usuario, enviadas a peticiones posteriores a lo largo de la cadena de peticiones de la transacción, y pueden ser modificadas en cualquier nivel a lo largo del recorrido. El contexto de la petición habilita la técnica descrita para identificar, seguir la pista y dar prioridad a la cadena de peticiones resultante como parte de la única transacción iniciada por una petición de usuario. El contexto de la petición puede incluir, por ejemplo, las características del usuario que presentó la petición, las características del artículo que es el sujeto de la petición, la ubicación geográfica desde la cual la petición se originó, la hora y fecha en el que se hizo la petición, el grupo de transacciones relacionadas a las que pertenece la petición, y similares. Determinadas partes del contexto de la petición pueden ser modificadas a niveles posteriores. Por ejemplo, la clase de servicio de la petición de usuario que es añadida al contexto de la petición en el primer nivel, puede ser invalidada por un nivel posterior (es decir, según otra forma de realización de la técnica descrita).

El término “clase de petición” aquí abajo, se refiere a una categoría de transacciones que comparten una o más características de contexto de la petición predefinidas. Por ejemplo, una “búsqueda de resumen de catálogo en existencia” puede ser clasificada como una petición de “búsqueda de resumen de catálogo en existencia”, o puede ser parte de una petición más grande “búsqueda de catálogo en existencia” junto con otra transacción, tal como una “búsqueda de antecedentes de catálogo en existencia”. Cada clase de petición es procesada conforme a una política de servicio activo. Una vez una clase de petición es asignada a la petición del usuario, esa clase de petición es automáticamente asignada a cada una de las peticiones posteriores en la transacción iniciada por esa petición del usuario.

El término “clase de servicio” aquí abajo, se refiere a un grupo de clasificaciones para varios parámetros que indican el nivel de importancia para el procesamiento de la petición. Los parámetros pueden incluir: la prioridad que se le ha de asignar a la petición, el porcentaje de CPU que se le ha de asignar a la petición, la memoria que se le ha de asignar a la petición, la prioridad para asignar y acceder a los dispositivos de entrada/salida (I/O) de la petición, y similares. La clase de servicio es asignada a una petición que se ejecuta en un nivel dado por el agente de contexto respectivo, conforme a la política de clase de servicio activo apropiada.

El término “política de clase de servicio” aquí abajo, se refiere a una regla que asigna una clase de servicio a una petición dentro de una clase de petición, con respecto al nivel sobre el que la petición está siendo procesada. Cada agente de contexto contiene un conjunto de políticas de clases de servicios de niveles específicos, cada una de las cuales traza un mapa de una clase de servicio para una clase de petición para un nivel específico asociado a ese agente de contexto. Una “base de datos de políticas de clases de servicios de niveles cruzados” describe el conjunto de rutas de clases de servicios para clases de peticiones para todos los niveles en el entorno informático multinivel. Se observa que un usuario puede definir un conjunto de políticas de clases de servicios. Las políticas de este tipo son denominadas aquí abajo como “políticas de servicio definidas por el usuario”.

El término “política de servicio activo” contiene la clase de petición para la clase de servicio trazada que está habitualmente vigente. Las políticas de clases de servicios múltiples son soportadas y una política de servicio diferente puede ser programada a tiempos diferentes de operación de sistema para reflejar cargas de trabajo variables o varios eventos de sistema, o simplemente como una decisión *ad hoc*.

Ahora se hace referencia a la figura 2, que es una ilustración esquemática de un sistema de gestión de rendimiento de aplicación, generalmente referenciado 100, construido y operativo conforme a una forma de realización de la técnica descrita. El sistema 100 funciona en un entorno informático multinivel, generalmente referenciado 132. El entorno informático 132 incluye un primer cliente 112 que lleva a cabo una primera aplicación, un segundo cliente 114 que lleva a cabo una segunda aplicación, un primer nivel 102, un segundo nivel 104, un tercer nivel 106, un cuarto nivel 108, y un quinto nivel 110. El primer nivel 102 es un servidor web. El segundo nivel 104 es un servidor de aplicación, el servidor de aplicación A. El tercer nivel 106 es otro servidor de aplicación, el servidor de aplicación B. El cuarto nivel 108 es otro servidor de aplicación, el servidor de aplicación C. El quinto nivel 110 es una base de datos.

El primer nivel 102 es acoplado con el primer cliente 112, con el segundo cliente 114, y con el segundo nivel 104. El segundo nivel 104 es posteriormente acoplado con el tercer nivel 106. El tercer nivel 106 es posteriormente acoplado con el cuarto nivel 108 y con el quinto nivel 110. El cuarto nivel 108 es posteriormente acoplado con el quinto nivel 110. La primera aplicación que se ejecuta en el primer cliente 112 inicia una petición de usuario R1. La segunda aplicación que se ejecuta en el segundo cliente 114 inicia una petición de usuario 118.

El sistema 100 incluye una pluralidad de agentes de contexto 122, 124, 126, 128 y 130, y un servidor de gestión de red de contexto (CNMS) 120. En el ejemplo expuesto en la figura 2, hay un único agente de contexto asociado con cada nivel. En particular, los agentes de contexto 122, 124, 126, 128 y 130 están asociados al primer nivel 102, el segundo nivel 104, el tercer nivel 106, el cuarto nivel 108 y el quinto nivel 110, respectivamente. Los agentes de contexto 122, 124, 126, 128 y 130 son acoplados con CNMS 120. Cada agente de contexto es también acoplado con otros agentes de contexto conforme al acoplamiento del nivel en el entorno informático 132. En particular, el agente de contexto 122 es acoplado con el agente de contexto 124, el agente de contexto 124 es posteriormente acoplado con el agente de contexto 126, el agente de contexto 126 es posteriormente acoplado con el agente de contexto 128 y con el agente de contexto 130, y el agente de contexto 128 es posteriormente acoplado con el agente de contexto 130.

El primer cliente 112 requiere un servicio del primer nivel 102 y el primer cliente 112 manda una petición de usuario R1 al primer nivel 102. La petición de usuario R1 espera en una cola en un puerto de entrada de petición del primer nivel 102. El primer nivel 102 asigna un enclave de procesamiento disponible X1 para procesar la petición del usuario R1. Mientras se procesa la petición del usuario R1, la lógica de aplicación que se ejecuta en el enclave de procesamiento X1 determina que el enclave de procesamiento X1 no puede completar el procesamiento de la petición del usuario R1 sin información adicional u operaciones que han de ser proporcionadas por el segundo nivel 104. En consecuencia, el primer nivel 102 manda una petición nueva R2 al segundo nivel 104, solicitando la información adicional u operaciones. El segundo nivel 104 asigna un enclave de procesamiento disponible X2 a la petición de proceso R2. La lógica de aplicación que se ejecuta en el enclave de procesamiento X2 determina que el enclave de procesamiento X2 requiere información adicional u operaciones que han de ser proporcionadas por el tercer nivel 106. En consecuencia, el segundo nivel 104 manda una petición nueva R3 al tercer nivel 106. El tercer nivel 106 asigna un enclave de procesamiento disponible X3 para procesar la petición R3. Se observa que cada una de las peticiones R1, R2, y R3 son parte de una única transacción que se origina de la aplicación ejecutada en el primer cliente 112.

El enclave de procesamiento X3 completa el procesamiento. El tercer nivel 106 devuelve una respuesta R3' al segundo nivel 104, en respuesta a la petición anterior R3 del segundo nivel 104. La lógica de aplicación que se ejecuta en el enclave de procesamiento X2 recibe la respuesta R3' y recupera la ejecución. Una vez que el enclave de procesamiento X2 ha completado el procesamiento, el segundo nivel 104 le devuelve una respuesta R2' al primer nivel 102 en respuesta a la petición anterior R2 del primer nivel 102. La lógica de aplicación que se ejecuta en el enclave de procesamiento X1 recibe la respuesta R2' y continua la ejecución. Una vez que el enclave de procesamiento X1 ha completado el procesamiento, el primer nivel 102 le devuelve una respuesta R1' a la petición del usuario R1, que se ha completado ahora.

Cada agente de contexto controla el nivel asociado al mismo en los puertos de entrada de petición y en los puertos de salida de petición del nivel (representados como círculos pequeños en las figuras 2 y 3). El agente de contexto controla el tráfico de petición que pasa a través del nivel asociado, para detectar que una petición ha entrado o salido del nivel asociado. Si la petición es una petición de usuario (es decir, la petición inicial en una cadena de peticiones), el agente de contexto del primer nivel identifica el contexto de la petición de la petición del usuario, clasifica la petición del usuario en una clase de petición, y asigna una clase de servicio a la petición del usuario basada en los contenidos de la política de servicio activo. Cada agente de contexto tiene un caché de política (no mostrado) que contiene el grupo de políticas de clase de servicio específico de niveles para el nivel asociado al agente de contexto respectivo. CNMS 120 actualiza periódicamente cada agente de contexto con las políticas de clase de servicio activo de nivel específico. Si la petición no es una petición de usuario, el agente de contexto recibe información acerca del contexto de la petición de la petición, junto con información adicional relacionada con la petición (es decir, "información de contexto"), del agente de contexto asociado al nivel donde esa petición se originó. Se observa que la información mínima incluida en la información de contexto que un agente de contexto transmite a otro agente de contexto es al menos: la identidad de la petición, la identidad de la transacción, la clase de petición, y los datos relacionados con el contexto asociados a la petición. Los datos relacionados con el contexto pueden incluir el contexto de la petición mismo, o una indicación (p. ej., un indicador) para el contexto de la petición que reside en otra asignación.

El agente de contexto asocia la información de contexto recibida con la petición que se ejecuta en el nivel. El agente de contexto puede influenciar en el procesamiento de la petición, por el nivel respectivo, conforme a la clase de servicio asignada a la petición. Por ejemplo, el agente de contexto puede ajustar el orden de la petición en la cola en un puerto de entrada de petición al nivel, o puede instruir el nivel para asignar, o de forma alternativa, para denegar, recursos computacionales del nivel para ejecutar la petición. Si el agente de contexto detecta que una petición ha salido del nivel asociado, el agente de contexto transmite información de contexto a otro agente de contexto asociado al nivel al que la petición ha sido enviada. Este otro agente de contexto asocia la información de contexto recibida con la petición pertinente, y con el enclave de procesamiento ejecutando la petición en el nivel asociado a este otro agente de contexto.

Se observa que el agente de contexto controla los puertos de entrada de petición y los puertos de salida de petición del nivel, más que controlar extensamente la actividad que ocurre dentro del nivel mismo (p. ej., el enclave de procesamiento que ejecuta una petición). Como resultado, el sistema 100 no interfiere en la operación real de un nivel dado o el código de aplicación de usuario que se ejecuta en el nivel desde una perspectiva de software, y el sistema 100 añade una carga mínima complementaria a los niveles.

El agente de contexto también es acoplado con el nivel asociado mediante una extensión de nivel dinámico (DTE - no mostrado en la figura 2). La DTE habilita al agente de contexto para recoger datos acerca de la ejecución de UOWs en ese nivel. Los agentes de contexto pueden enviar datos sin procesar al CNMS 120 para archivar objetivos. Los

agentes de contexto pueden además enviar a CNMS 120 datos estadísticos para el análisis agregado. Los agentes de contexto pueden recibir información de CNMS 120 así como perfiles de actividad (definidos aquí abajo con referencia a la figura 6) y nuevas políticas de clases servicios activos para la manipulación de diferentes tipos de clases de peticiones. El agente de contexto es elaborado con detalle en la figura 8 aquí abajo.

En particular, el agente de contexto 122 controla el primer nivel 102 y detecta que la petición de usuario R1 ha entrado en el primer nivel 102. El agente de contexto 122 identifica el contexto de la petición de la petición de usuario R1 y asocia la petición de usuario R1 con el enclave de procesamiento X1 que procesa la petición. El agente de contexto 122 clasifica la petición de usuario R1 en una clase de petición apropiada. El agente de contexto 122 determina la clase de servicio de petición de usuario R1 en el primer nivel 102, recuperando la política de clase servicio activo apropiado en el grupo de políticas de clase de servicio que el agente de contexto 122 ha almacenado, y asigna a la petición del usuario R1 la clase de servicio determinado. El agente de contexto 122 añade la clase de servicio asignado al contexto de la petición. Cuando la nueva petición R2 sale del primer nivel 102 en dirección al segundo nivel 104, el agente de contexto 122 detecta que la petición R2 está relacionada con la petición del usuario R1. El agente de contexto 122 entonces manda al agente de contexto 124 información acerca del contexto de la petición de la petición de usuario R1, con la identidad de petición, la clase de petición, y la identidad de transacción asociada a la petición R2.

Ahora se hace referencia a la figura 3, que es una ilustración esquemática de la transferencia de información entre dos de los agentes de contexto del sistema de la figura 2. El agente de contexto 122 le manda al agente de contexto 124 un mensaje 134. El mensaje 134 incluye la identidad de petición de la petición R2, la identidad de transacción de la petición R2, la clase de petición en la que el agente de contexto 122 clasificó la petición R2, y el contexto de la petición de la petición R2. El agente de contexto 124 recibe el mensaje 134 y determina la clase de servicio de la petición R2 que ha de ser ejecutada en el segundo nivel 104, recuperando la política de clase de servicio activo apropiada en el conjunto de políticas de clase de servicio cuyo agente de contexto 124 ha almacenado. El agente de contexto 124 asigna a la petición R2 la clase de servicio determinado. Por ejemplo, la clase de petición de la petición R2 es el grupo "15". El agente de contexto 124 recupera la política de servicio activo que traza un mapa de una clase de servicio para las peticiones de la clase de petición "15" que son ejecutadas en el segundo nivel 104. La política de clase de servicio apropiada asigna una prioridad de "5" para tal petición, una asignación de CPU de "90", una asignación de memoria de "48", y prioridad de acceso de dispositivo I/O de "2". El agente de contexto 124 puede entonces influir en el procesamiento de la petición R2 conforme a la clase de servicio asignada.

El sistema 100 ejecuta la gestión de rendimiento de aplicación en una base de contexto de la petición. El sistema 100 identifica las peticiones, y las características referentes a cada petición están disponibles en el agente de contexto asociado al nivel. Estas características pueden incluir donde se inició esa petición, qué peticiones precedieron a la petición en la transacción, y qué tipo de petición es. Por ejemplo, el agente de contexto 124 identifica que la petición R2 que opera en el segundo nivel 104 está unida con la petición de usuario R1 que fue procesada por el primer nivel 102 e iniciada en el primer cliente 112. Puesto que el agente de contexto de un nivel dado es consciente del contexto de la petición y de la clase de petición de cada petición que está siendo ejecutada en el nivel respectivo, el agente de contexto puede determinar la clase de servicio específico del nivel apropiado de la petición respectivo basada en la política de clase de servicio. CNMS 120 puede establecer la política de gestión global a través de diferentes niveles, respectivas de diferentes clases de petición, y por consiguiente actualizar los agentes de contexto.

Ahora se hace referencia a la figura 4, que es una ilustración esquemática de un ciclo de vida de petición de muestra, generalmente referenciado 140, sobre dos de los niveles del sistema de la figura 2. El ciclo de vida de muestra 140 representa los estadios que sufre una petición puesto que la petición está siendo servida en un entorno informático multinivel 132 (figura 2). Hay que recordar que una petición causa una invocación de una UOW, que puede además generar peticiones adicionales, bien internamente dentro del mismo nivel (enviando una petición al mismo nivel sobre el que está siendo ejecutada la UOW), o externamente (enviando peticiones a otros niveles). Por lo tanto, una petición del usuario normalmente genera una serie de invocaciones de UOWs, cada una de las cuales pueden ser realizadas en un nivel diferente. Las invocaciones de UOWs pueden ser sincrónicas (es decir, el enclave de procesamiento que ejecuta la invocación UOW espera una respuesta de la UOW invocada antes de reanudar el procesamiento) o asincrónico (es decir, el enclave de procesamiento que ejecuta la invocación UOW continúa procesando la invocación de UOW sin esperar una respuesta de la UOW invocada). En ambos casos, la UOW en el nivel invocado N+1 está dedicada al servicio solicitado por el nivel de invocación N. En el procesamiento sincrónico hay casos donde la invocación UOW en el nivel N espera a la UOW invocada en el extremo de nivel N+1 (es decir, la UOW en el nivel N+1 es desasignada). En otros casos, la UOW invocada en el nivel N+1 puede ser referenciada múltiples veces por el nivel de invocación, hasta que la UOW invocada acaba en el nivel N+1.

En la fase 142, una primera petición es enviada al nivel N (es decir, cualquier nivel representativo) en el entorno informático 132. La primera petición resulta en una invocación UOW en el nivel N para proporcionar un servicio, bien para un nivel precedente o para una aplicación de usuario. La primera petición espera en una cola 158 en el nivel N.

En la fase 144, la primera petición sale de la cola 158 y se asigna una UOW, UOW-A, en el nivel N. Una asignación de UOW implica asignar un enclave de procesamiento disponible de uno de los enclaves de procesamiento 162 en el nivel N y enviar el código de aplicación de petición para seguir en el enclave del procesamiento. La asignación de UOW ocurre una vez que los recursos de nivel de objetivo están disponibles y es posible asignar el código de aplicación a un enclave de procesamiento disponible en el nivel N. En el ciclo de vida de muestra 140, UOW-A comienza la ejecución en el nivel N.

En la fase 146, UOW-A emite una segunda petición al nivel N+1. El nivel N+1 luego invoca a UOW-B para ejecutar esta petición del nivel N. En la fase 148, el nivel N+1 invoca a UOW-B para ejecutar la segunda petición enviada por el nivel N. La segunda petición espera en una cola 160 en el nivel N+1. En la fase 150, la segunda petición sale de la cola 160 y el UOW-B se asigna a la segunda petición. La asignación UOW-B resulta en la atribución de un enclave de procesamiento disponible de uno de los enclaves de procesamiento 164 en el nivel N+1 al código de aplicación de la invocación UOW y enviar el código de aplicación de petición para seguir en ese enclave de procesamiento. La UOW- B entonces comienza la ejecución en el nivel N+1. Se observa que la invocación de UOW-B es sincrónica, y así el enclave de procesamiento que procesa la UOW-A no continúa procesando mientras que espera una respuesta de la UOW-B.

En el caso de que la invocación de UOW-B sea asincrónica, el enclave de procesamiento que procesa UOW-A recibe un acuse de que el nivel N+1 que la segunda petición envió de UOW-A al nivel N+1 fue aceptado. Al recibir el acuse, el enclave de procesamiento que procesa UOW-A recupera la ejecución hasta que el enclave de procesamiento finalmente devuelve una respuesta a la primera petición. Después de que el nivel N+1 acepte la segunda petición asincrónica, la segunda petición espera en la cola 160. La segunda petición es posteriormente leída por uno o más enclaves de procesamiento que manipulan la segunda petición, hasta que uno de esos enclaves de procesamiento también retira la segunda petición de la cola 160. Cada enclave de procesamiento que manipula la segunda petición puede también convertir la segunda petición en una petición del usuario nuevo, el cual puede iniciar a su vez otra cadena de peticiones, comenzando de este modo una transacción nueva.

Por ejemplo, una transacción que implica una petición asincrónica puede ser un usuario que confirma la compra de un libro en una página web de comercio electrónico. La petición de compra devuelve al usuario a una pantalla informando de que la orden está siendo procesada y se le notificará al usuario (p. ej., vía e-mail o mensaje de texto). La misma petición de compra es simultáneamente colocada en una cola de mensajes donde la petición de compra es procesada más tarde por: un enclave de procesamiento que envía una petición de aprobación final a la empresa de tarjetas de crédito; un enclave de procesamiento que envía una petición de orden de compra al almacén; un enclave de procesamiento que envía una petición de contabilidad al sistema de facturación; y similares.

En la fase 152, UOW-B devuelve una respuesta a UOW-A en respuesta a la invocación anterior de UOW-A, y la ejecución de UOW-B es ahora completada. La segunda petición ahora ha terminado. En la fase 154, UOW-A recibe la respuesta y recupera la ejecución, punto en el que UOW-B es liberado por el nivel N+1. UOW-A puede entonces continuar ejecutándose. La duración de tiempo entre cuando una UOW de invocación hace una petición y cuando UOW de invocación recibe una respuesta de la UOW invocada, es conocida como el periodo “latencia”, o el tiempo de respuesta para una petición de UOW dada.

En la fase 156, UOW-A completa la ejecución y la primera petición termina. Se observa que antes de la finalización, UOW-A puede requerir los servicios de otro nivel y puede invocar otra petición para proporcionar ese servicio.

Se observa que después de que UOW-B sea asignado en la fase 150, y comience la ejecución en el nivel N+1, puede ocurrir un error no recuperable (p. ej., una excepción del programa). Conforme a una forma de realización de la técnica descrita, el agente de contexto asociado al nivel N+1 registrará el error y asociará el error a la transacción que comenzó con la petición del usuario que invocó UOW-A en el nivel N, proporcionando información como en cuanto a la naturaleza del error que ocurrió en el nivel N+1.

Ahora se hace referencia a la figura 5, que es una ilustración esquemática de un sistema de gestión de rendimiento de aplicación, generalmente referenciado 200, construido y operativo conforme a otra forma de realización de la técnica descrita. El sistema 200 funciona en un entorno informático multinivel, generalmente referenciado 248. El entorno informático 248 incluye un cliente 214, un primer nivel 202, un segundo nivel 204, un tercer nivel 206, un cuarto nivel 208, y un quinto nivel 210. El primer nivel 202 es un servidor web. El segundo nivel 204 es un servidor de aplicación A. El tercer nivel 206 es un servidor de aplicación B. El segundo nivel 204 y tercer nivel 206 residen ambos en una única máquina huésped 212. El cuarto nivel 208 es otro servidor de aplicación, el servidor de aplicación C. El quinto nivel 210 es una base de datos. El primer nivel 202 es acoplado al cliente 214, y al huésped 212. El huésped 212 es posteriormente acoplado al cuarto nivel 208 y al nivel quinto 210. El cuarto nivel 208 es posteriormente acoplado al quinto nivel 210.

El sistema 200 incluye una pluralidad de extensiones de nivel dinámicos 222, 224, 226, 228 y 230, una pluralidad de agentes de contexto 232, 234, 236 y 238, una pluralidad de registros locales 240, 242, 244 y 246, un servidor de gestión de red de contexto (CNMS) 216, una base de datos de políticas de objetivo de nivel de servicio (SLO) 218, y una estación de trabajo de supervisor 220. Cada nivel contiene una extensión de nivel dinámico (DTE). Hay un agente de contexto asociado a cada nivel. Hay un registro local asociado a cada agente de contexto. Un agente de contexto de un nivel dado es acoplado al DTE (o diferentes DTEs) dentro del nivel, con el registro local asociado al agente de contexto, y con otros agentes de contexto conforme al acoplamiento del nivel en el entorno informático. Cada agente de contexto es también acoplado a CNMS 216. El CNMS 216 es acoplado con la base de datos de políticas de SLO 218. La estación de trabajo de supervisor 220 es acoplada al CNMS 216 y a la base de datos de políticas de SLO 218.

En particular, el primer nivel 202 incluye DTE 222. El agente de contexto 232 está unido con el primer nivel 202. El registro local 240 está unido con el agente de contexto 232. El segundo nivel 204 incluye DTE 224. El tercer nivel 206 incluye DTE 226. Puesto que el segundo nivel 204 y el tercer nivel 206 residen ambos en el huésped 212, hay

sólo un único agente de contexto 234 asociado al segundo nivel 204 y al tercer nivel 206. Se observa que el agente de contexto 234 está directamente acoplado al DTE 224 y DTE 226. El registro local 242 está unido al agente de contexto 234. El cuarto nivel 208 incluye DTE 228. El agente de contexto 236 está unido al cuarto nivel 208. El registro local 244 está unido al agente de contexto 236. Finalmente, el quinto nivel 210 incluye DTE 230. El agente de contexto 238 está unido al quinto nivel 210. El registro local 246 está unido al agente de contexto 238. El agente de contexto 232 está acoplado al agente de contexto 234. El agente de contexto 234 está posteriormente acoplado al agente de contexto 236 y al agente de contexto 238. El agente de contexto 236 está posteriormente acoplado al agente de contexto 238.

Una extensión de nivel dinámico es acoplada al nivel en puntos específicos predeterminados. Estos puntos predeterminados son: los puertos de entrada de petición del nivel, los puertos de salida de petición del nivel, y posiblemente áreas adicionales dentro del nivel (p. ej., un puerto de control del nivel). Un puerto de petición según la técnica descrita es un módulo dentro de un nivel que administra peticiones, bien antes de ser procesadas por el nivel, o después de ser procesadas por el nivel. Un puerto de petición de este tipo puede ser un punto de la interfaz (es decir, entrada, salida o cualquier otro mecanismo de acceso) para una cola de petición en la entrada de un nivel. Puesto que una petición requiere el servicio de un código de aplicación que tiene lugar en el nivel mediante un enclave de procesamiento, el puerto de petición respectivo reside en un nivel de aplicación y no en un nivel de conexión de redes. Se observa que los puertos de petición según la técnica descrita, no están a un nivel de red (p. ej., no en puertos TCP/IP o UDP).

La DTE está localizado en el mismo nivel huésped que el nivel asociado a éste (es decir, la DTE está localizada en al menos una de las máquinas huésped en la que el nivel tiene lugar). Entre las responsabilidades de la DTE está la de capturar un contexto de la petición. La DTE además observa las aberturas de introducción de petición y las aberturas de salida de petición de un nivel, para detectar las peticiones entrantes y salientes. La DTE asigna una identidad de transacción a una petición del usuario, y obtiene la identidad de la petición de cada petición que entra o sale del nivel. La DTE es explicada con detalle en la figura 7 descrita aquí abajo.

El agente de contexto mantiene asociaciones entre una petición dada, el UOW invocada de la petición, y el contexto de la petición de la petición de la usuario en la misma transacción que la petición. El agente de contexto transmite el contexto de la petición asignado a cada petición (es decir, los datos relacionados con el contexto) a otros agentes de contexto que procesan otros niveles. El agente de contexto puede transmitir todo el contexto de la petición, o una parte del contexto de la petición. Además, el agente de contexto puede transmitir el contexto de la petición mismo, o una indicación (p. ej., un indicador) al contexto de la petición que reside en otra asignación. Se observa que el agente de contexto no necesita necesariamente residir en la misma máquina huésped que el nivel, pero éste es el caso en una forma de realización preferida de la técnica descrita. El agente de contexto es explicado con detalle en la figura 8 descrita aquí abajo. CNMS 216 recoge y analiza los datos de rendimiento. La base de datos de políticas de SLO 218 almacena políticas de clase de servicio de nivel cruzado, y se actualiza continuamente.

Ahora se hace referencia a la figura 6, que es una ilustración esquemática de dos de los niveles del entorno informático multinivel de la figura 5. Se observa que tanto el cuarto nivel 208 como el quinto nivel 210 ilustrados en la figura 6 son representativos de dos niveles consecutivos cualquiera (p. ej., el nivel N y el nivel N+1) en el entorno informático 248.

Una petición que entra en el cuarto nivel 208 espera a un puerto de entrada de petición del cuarto nivel 208 en una cola 262. El cuarto nivel 208 invoca una UOW para ejecutar la petición. La petición sale de la cola 262 y el cuarto nivel 208 asigna la UOW a la petición, asignando un enclave de procesamiento disponible a la UOW de los enclaves de procesamiento 252 y despachando el código de aplicación de la petición para llevar a cabo el enclave de procesamiento. La UOW se ejecuta en el cuarto nivel 208. La UOW puede entonces pedir un servicio del quinto nivel 210. La petición nueva sale del cuarto nivel 208 a un puerto de salida de petición y espera en un puerto de entrada de petición del quinto nivel 210 en una cola 264. El quinto nivel 210 invoca una UOW para ejecutar la petición nueva. La petición nueva sale de la cola 264 y el quinto nivel 210 asigna la UOW a la petición nueva, para asignar un enclave de procesamiento disponible a la UOW del enclave de procesamiento 254, y despachar el código de aplicación de la petición nueva para llevar a cabo el enclave de procesamiento. La DTE 228 observa los puertos de entrada de petición y los puertos de salida de petición del cuarto nivel 208 para detectar la petición que entra y sale del cuarto nivel 208, respectivamente.

Las extensiones de nivel dinámico están implicadas en trazar una petición en todo su ciclo de vida, sin cambiar el código de aplicación. La DTE se conecta dinámicamente al entorno del nivel donde la DTE intercepta el contexto de la petición externa al código de aplicación. El rastreo de la petición incluye capturar el contexto de la petición, asociar la petición a una UOW en un nivel, y desasociar la petición de una UOW en un nivel. La DTE además recoge el rendimiento, la disponibilidad, y el error métrico del nivel. La DTE también puede ajustar dinámicamente el procesamiento de petición en el nivel, ajustando el orden de una petición en la cola en un puerto de entrada de petición del nivel, asignando recursos computacionales para procesar la petición (p. ej., CPU, memoria, I/O, y similares) o alterando la prioridad del enclave de procesamiento o los recursos asignados. Estas tareas son explicadas con referencia a las figuras 9, 10 y 11 descritas aquí abajo.

Se observa que hay dos alternativas para el contexto de la petición transmitidas entre los agentes de contexto: dentro de banda y fuera de banda. En la transmisión de contexto dentro de banda, el contexto de la petición se añade a la petición misma (es decir, sobre la carga útil), cuando la petición sale de un nivel determinado en dirección al siguiente nivel. En consecuencia, como una petición y peticiones descendientes de los mismos son procesados entre

niveles diferentes, el contexto de la petición actualizado se añade a las invocaciones de petición. Por el contrario, la transmisión fuera de banda no implica que el contexto de la petición se añada a las invocaciones de petición. Más bien los agentes de contexto envían la información de contexto directamente entre sí. Un agente de contexto envía un contexto de la petición a otro agente de contexto. Una DTE recupera el contexto de la petición del agente de contexto. Se observa que cada agente de contexto de un sistema similar al sistema 200 (figura 5) transmite el contexto de la petición a otro agente de contexto usando la técnica de fuera de banda.

Haciendo referencia de nuevo a la figura 5, el sistema 200 ejecuta un perfilado de la actividad. El perfilado de la actividad implica crear un perfil de actividad. El perfil de actividad incluye la transacción integrada, el nivel, y métrica de rendimiento de nivel del sistema y análisis estadístico, que se obtienen durante el rastreo de la petición. Por ejemplo, la métrica de rendimiento puede incluir: tiempo transcurrido de petición, tiempo de servicio de petición, tiempo de CPU consumido en el nivel, y similares. Los datos del perfil de la actividad son recogidos en el tiempo y usados para controlar (es decir, presentar en el GUI) y para soportar la creación de una política de servicio definida para un usuario. Una consola de perfilado de la actividad (no mostrada) es un componente del GUI que muestra el rendimiento y la disponibilidad de los datos agregados recogidos por el agente de contexto. El rendimiento y la disponibilidad de los datos agregados incluyen vistas de resumen de los datos de perfil de actividad por varias categorías tales como: nivel, clase de petición, transacción y similares. Una actividad que perfila el motor (no mostrado) localizado en CNMS 216 ejecuta el perfil de la actividad.

Cada agente de contexto tiene la tarea de recoger información sobre los detalles de ejecución en cada nivel. La DTE habilita el agente de contexto para recopilar datos acerca de la ejecución de UOWs en ese nivel. Los agentes de contexto entonces envían la información recogida a CNMS 216. La información almacenada para cada UOW incluye: hora de comienzo, identidad de petición de la petición a la que la UOW es asignada, identidad de transacción de la petición a la que la UOW es asignada, clase de petición a la que la UOW es asignada, detalles de usuario, detalles de la dirección de la red de origen, clase de servicio de la petición a la que la UOW es asignada, hora de finalización, consumo de recursos (tal como una CPU), y similares.

El agente de contexto almacena la información acerca de las UOWs ejecutadas habitualmente en una memoria (no mostrada). Una vez que la UOW ha terminado de ejecutarse en el nivel, el agente de contexto transfiere la información a un almacén de datos de historial reciente (no mostrado), que es almacenado en un disco (no mostrado) localmente en el mismo huésped de nivel del agente de contexto. Después de un determinado periodo de tiempo, el agente de contexto mueve las entradas del almacén de datos de historial reciente a un almacén de resumen de datos (no mostrado). Esta información es almacenada como un resumen en un periodo de tiempo dado (p. ej., un promedio de recogida métrica sobre un periodo de media hora). La información en el depósito de resumen de datos es almacenada por cambios. Los cambios se guardan en una base periódica (es decir, se reciclan después de que haya acumulado un número de cambios).

El sistema 200 incluye además políticas de clase de servicio. Cabe recordar que una política de servicio se refiere a una regla que asigna una clase de servicio a una petición dentro de una clase de petición, con respecto al nivel sobre el que la petición está siendo procesado. Un motor de política de servicio a tiempo real, localizado en el agente de contexto, asigna una clase de servicio apropiado a la petición conforme a la información en el contexto de la petición, tal como la clase de petición, las características de rendimiento de ejecuciones precedentes de peticiones con el mismo contexto de la petición, y de acuerdo con una política de servicio activo. Además, la atribución de una clase de servicio puede tener en cuenta consideraciones adicionales, tales como un grupo de métricas de rendimiento para cada clase de petición. Este grupo de métricas de rendimiento caracteriza la clase de petición y crean una línea base para el comportamiento de rendimiento típico. Este proceso utiliza los datos de registro creados a través de todos los niveles y las clases de peticiones asociadas a los mismos. Las clases de petición son almacenadas en una estructura de árbol en la base de datos de políticas de SLO 218. El motor de la política de servicio añade y actualiza información acerca de cada clase de petición, y por consiguiente actualiza la base de datos de políticas de SLO 218.

La clase de servicio es determinada para la petición del usuario en el primer nivel y luego pasada de nivel a nivel con la petición en el contexto de la petición (de agente de contexto a agente de contexto). No hay necesidad de acceder a CNMS 216 para determinar la política de servicio activo y el mapeo para la clase de servicio. Más bien, cada agente de contexto tiene un caché de política en su interior (no mostrado) almacenado hasta la fecha, de modo que el mapeo para la clase de servicio apropiado se desempeña localmente.

El motor de perfilado de actividad emite a los agentes de contexto locales información sobre las peticiones completadas y llevadas a cabo actualmente en intervalos periódicos, o siempre que sea necesario y esté disponible (p. ej., cuando un usuario desea ver el estado actual de las peticiones en la consola de rendimiento). De forma alternativa, el motor de perfilado de la actividad puede instruir a cada agente de contexto a iniciar el envío de registros nuevos al motor de perfilado de actividad a intervalos fijos o cuando un umbral (es decir, número de registros) es alcanzado. El motor de perfilado de actividad recoge los datos de cada petición, realiza cálculos (p. ej., promedio, varianza, y similares) y almacena los resultados a varios intervalos en la base de datos de políticas de SLO 218, como una línea base para el análisis de cada clase de petición. En base a los datos disponibles almacenados en la base de datos de política de SLO 218, un motor de generación de políticas de servicio localizado en CNMS 216 crea un grupo de reglas que sirven de recomendaciones para políticas de clase de servicio nuevas. CNMS 216 determina las políticas de clase de servicio que usan estas recomendaciones. Las políticas de clase de servicio son almacenadas en la base de datos de políticas de SLO 218.

Se observa que la base de datos de políticas de SLO 218 almacena políticas que son generadas automáticamente al igual que las políticas de clase de servicio de usuario definido. Un usuario puede crear un grupo de políticas de clase de servicio (es decir, políticas de clase de servicio de usuario definido) mediante la interfaz de usuario (no mostrado) de sistema 200, o editando un fichero de configuración del agente de contexto. La creación de políticas de clase de servicio de usuario definido implica el análisis de usuario de los perfiles de la actividad, y la obtención de la aprobación de la política de servicio sugerida por CNMS 216.

El agente de contexto recibe políticas de clase de servicio actualizadas de CNMS 216. Se observa que si la política de servicio nueva es generada automáticamente o de usuario definido es transparente para el agente de contexto. El agente de contexto asigna a una petición la clase de servicio designada en la política de servicio activo del nivel especifico apropiado, localizada dentro del caché de política local del agente de contexto.

De forma alternativa, el agente de contexto puede asignar a la petición una clase de servicio diferente de aquella designada por la política de servicio activo en determinadas situaciones (p. ej., si el nivel es habitualmente incapaz de proporcionar todos los recursos requeridos para ejecutar la clase de servicio, si una cantidad significante de peticiones de alta prioridad introducen el nivel y pueden suponer la extinción de recursos para peticiones de prioridad baja, y similares). Además de forma alternativa, el agente de contexto puede alterar la clase de petición de la petición, y posteriormente asignar a la petición la clase de servicio designada por la política de servicio activo del nivel específico apropiada para la clase de petición nueva.

El sistema 200 también realiza la clasificación de la petición. Esto implica la clasificación de peticiones múltiples (cada una designada por su identidad de transacción y el contexto de la petición) en clases de petición según varios criterios. Un proceso de clasificación de petición automático reúne la información genérica asociada a una petición (p. ej., información de cabecera, parámetros tales como serie de preguntas, parámetros de Localizador Uniforme de Recursos (URL) en el caso de un tipo de petición HTTP, y similares). Cuando llega otra petición de la misma clase de petición, esta petición será procesada de una manera similar a la otra petición dentro de la clase de petición. Las clases de petición pueden ser reagrupadas en varias categorías conforme a las características de rendimiento de clase de petición en los perfiles de actividad pertinentes (p. ej., un grupo de clases de petición que tienen un tiempo de respuesta mayor que dos segundos). El agrupamiento de clases de peticiones se usa para reportar objetivos, permitir un nivel más alto de vistas resumidas de datos de rendimiento de clases de petición.

El sistema 200 también soporta un proceso de clasificación de usuario definido donde el usuario crea reglas que clasifican las peticiones basadas en la misma información de petición usada para el proceso de clasificación automática. Un usuario puede crear reglas de clasificación mediante la interfaz de usuario (no mostrado) del sistema 200. El proceso de clasificación de peticiones automático usa un algoritmo "gestión de caché basada en la clase", como se describe por H. Zhu y T. Yang ("*Class-based cache management for dynamic web contents*", Tech. Rep. TRCS00-13, Dept. de Ciencias informáticas, Universidad de California, Santa Bárbara, 2000). La salida del proceso de clasificación es un árbol que representa la clase de petición y las clases parentales (es decir, en un proceso de clasificación de petición automático) o un grupo de reglas de clasificación (es decir, en un proceso de clasificación de usuario definido). Todos los resultados de los procesos de clasificación son almacenados en la base de datos de políticas de SLO 218.

El sistema 200 es también operativo para aplicar la clase de servicio, tal y como se define en la política de servicio activo, en todos y cada uno de los niveles controlados. La aplicación de política es realizada por la DTE y el agente de contexto. La aplicación puede ser implementada bien controlando la cola de peticiones en cada nivel (es decir, el orden en que la petición es en realidad procesada dentro del nivel), o cambiando temporalmente la prioridad de procesamiento del enclave de procesamiento que realiza la petición durante la ejecución.

La implementación depende de la arquitectura del nivel particular. Por ejemplo, la implementación en un nivel de base de datos puede implicar el uso de un esquema de gestión de recursos de base de datos para manipular la asignación de recursos de sesión según la política de servicio apropiada. Otro ejemplo es la implementación en un nivel de servidor de aplicación implementado usando una aplicación de servidor de aplicación J2EE que puede implicar: extensión de la cola de servidor web de aplicación para sostener la prioridad de la petición, extensión de la cola de Enterprise JavaBeans (EJB) para soportar la prioritización, controlar la prioridad de threads en JAVA, y similares.

Ahora se hace referencia a la figura 7, que es una ilustración esquemática de una extensión de nivel dinámico del sistema de la figura 5. Se observa que DTE 228 del cuarto nivel 208, representada en la figura 5, es representativa de todas las extensiones de nivel dinámico en el sistema 200. DTE 228 incluye una serie de ganchos blandos o puntos de intercepción en el cuarto nivel 208. Estos ganchos, referenciados con 266 y 268, sirven para recopilar información relacionada con la petición y datos del rendimiento. Los ganchos también pueden alterar la prioridad de un enclave de procesamiento que ejecuta una UOW. Un gancho posiciona la información recogida en una cola de mensaje dedicada 270. El gancho blando y la técnica de intercepción dependen del medio particular. Por ejemplo, el medio puede ser: un servidor web, un servidor de aplicación J2EE basado en JAVA, una base de datos, un servidor de mensajes, y similares. Se observa que los puntos de intercepción 266 y 268 pueden ser instantáneamente activados o desactivados por un operador.

DTE 228 incluye además un proceso llevado a cabo dentro del cuarto nivel 208. El proceso se encarga de los mensajes de los puntos de intercepción, comunica los mensajes al agente de contexto 236, devuelve los mensajes a los puntos de intercepción, y ejecuta funciones de control administrativas de la DTE (p. ej. peticiones de inicio/parada de

seguimiento, instalación y eliminación de ganchos blandos). DTE 228 incluye un demonio DTE 272, un administrador DTE 274, un módulo de mensajería DTE 276, un manipulador de eventos registrados de DTE 278 y una interfaz de comunicación de DTE 280.

El demonio DTE 272 es un enclave de procesamiento creado artificialmente que opera dentro del nivel. El demonio DTE 272 ejecuta un procesamiento asincrónico asociado a eventos proporcionados donde la petición no necesita ser detenida. Hay dos tipos de escenarios en lo que se refiere a informe de eventos. En el primer escenario, no hay necesidad de detener la petición hasta que se reciba una respuesta. Por ejemplo, cuando se informa de que una petición ha terminado, no hay ninguna necesidad de retrasar la petición hasta después de que el agente de contexto haya sido notificado en realidad. En el segundo escenario, la petición necesita ser mantenida durante un periodo determinado antes de que el procesamiento se reanude. Por ejemplo, cuando se obtiene la clase de servicio de una petición o cuando se realiza la clasificación de una petición, el procesamiento no puede comenzar hasta que la clase de petición y clase de servicio sea determinada por el agente de contexto, de lo contrario el procesamiento puede ser hecho usando una clase de servicio incorrecto o una clase de petición incorrecta.

El demonio DTE 272 procesa los eventos del primer escenario, donde la petición no necesita ser detenida. El procesamiento es hecho asíncronamente, de manera que la petición no se retrasa. De este modo la petición sale rápidamente, casi instantáneamente. El demonio DTE 272 tiene una cola 270 asociada a él. Después de que las entradas de petición y las salidas de la petición relacionadas sean notificadas por puntos de intercepción 266 y 268, el demonio DTE 272 recoge estas notificaciones de la cola 270 y ejecuta cualquier procesamiento adicional que sea necesario.

El administrador de DTE 274 habilita a la DTE 228 para recibir mensajes acerca de cómo las peticiones deberían ser procesadas. Por ejemplo, los mensajes de este tipo pueden incluir: parar el seguimiento de una petición, continuar el seguimiento de la petición pero parar de priorizar, y similares. El módulo de mensajes de DTE 276 se comunica con el agente de contexto 236 usando mensajes. Por ejemplo, los mensajes de este tipo pueden incluir: inicio o final de una UOW, asociar una UOW con una petición dada, y similares. El manipulador de eventos de registro de DTE 278 registra información de trazado en lo que se refiere a operaciones de DTE y registra alertas formuladas por DTE 228. Estos eventos registrados podrían ser enviados a destinos múltiples tales como un fichero local, una consola de mensajes de sistema, un registro del sistema, y similares. El manipulador de eventos de registro de DTE 278 soporta protocolos estándar de industria múltiple tal como Protocolo Simple de Gestión de Red (SNMP), y similares. La interfaz de comunicación de DTE 280 sirve como interfaz entre DTE 228 y el agente de contexto 236. La interfaz de comunicación de DTE 280 transmite mensajes enviados desde un módulo de mensajería de agente 286 de agente de contexto 236 a DTE 228. La interfaz de comunicación de DTE 280 también transmite mensajes enviados del módulo de mensajería de DTE 276 de DTE 228 al agente de contexto 236. Múltiples protocolos de comunicación son soportados, y cada DTE usa el método de comunicación más eficaz disponible dentro de su arquitectura, tal como una comunicación entre procesos, Protocolo de control de transmisión/Protocolo de Internet (TCP/IP), y similares.

Ahora se hace referencia a la figura 8, que es una ilustración esquemática de un agente de contexto del sistema de la figura 5. Se observa que el agente de contexto 236 del cuarto nivel 208, representado en la figura 5, es representativo de todos los agentes de contexto en el sistema 200. El agente de contexto 236 recibe notificaciones de DTE 228 mediante una variedad de mecanismos, tales como TCP/IP, canales de comunicación entre procesos, y similares. Las notificaciones son de eventos que ocurren dentro del nivel, tales como: la captura de un contexto de petición, el inicio de una UOW, el fin de una UOW, el consumo de recursos de una UOW, la invocación/asignación de una UOW en un nivel remoto, la respuesta/liberación de una UOW en un nivel remoto, y similares. El agente de contexto 236 incluye un manipulador de eventos de registros de agente 282, una interfaz de comunicación de agente 284, un módulo de mensajería de agente 286, un gestor de tabla de contexto 288, un gestor de la clasificación 290, un gestor de políticas 292, y un administrador de agente 294.

El manipulador de eventos de registro de agente 282 se usa por el agente de contexto 236 para los fines de tareas domésticas internas y para alertas de registros aumentadas por el agente de contexto 236. El manipulador de eventos de registro de agente 282 registra información que entra en la tabla de contexto (como se describe aquí abajo), pero es también usado para el trazado interno y los fines de mensajería, así como para detectar irregularidades operacionales (es decir, problemas o errores) que pueden ocurrir dentro del agente de contexto 236. Estos eventos registrados pueden ser enviados a múltiples destinos tales como un fichero local, una consola de mensajes del sistema, un registro del sistema, y similares. El manipulador de eventos de registro de agente 282 soporta protocolos estándares de industria múltiple tales como SNMP, y similares.

La interfaz de comunicación de agente 284 sirve como una interfaz entre DTE 228 y el agente de contexto 236. La interfaz de comunicación de agente 284 transmite mensajes enviados desde el módulo de mensajería de DTE 276 de DTE 228 al agente de contexto 236. La interfaz de comunicación de agente 284 también transmite mensajes enviados desde el módulo de mensajería de agente 286 del agente de contexto 236 a DTE 228. Pueden haber diversos canales que conectan la DTE 228 y el agente de contexto 236, para asegurar una comunicación rápida y fiable entre los dos, y al menos diversos canales se quedan abiertos en todo momento, por ejemplo un canal de prioridad alta y un canal administrativo. También pueden haber diferentes conexiones de cada tipo de canal, para diferentes tipos de mensajes. Como resultado, la interfaz del agente de comunicación 284 es operativa para considerar estas diferentes posibilidades.

El módulo de mensajería de agente 286 notifica a otros agentes de contexto asociados a los niveles remotos que una petición fue enviada al nivel remoto. El módulo de mensajería de agente 286 además se comunica con DTE 228

usando mensajes. Por ejemplo, este tipo de mensajes incluyen: iniciar o finalizar una UOW, asociar una UOW con una petición, y similares. El módulo de mensajería de agente 286 se comunica con DTE 228 mediante la interfaz del agente de comunicación 284.

El gestor de la tabla de contexto 288 funciona como el contable del agente de contexto. El gestor de la tabla de contexto 288 mantiene una tabla de referencia cruzada, conocida como una "tabla de contexto", usada para asociar UOWs llevadas a cabo en el nivel para su contexto de petición. El contexto de petición puede ser del nivel corriente (es decir, en el caso de una petición de usuario), transmitido desde un nivel precedente de la cadena de ejecución de petición, o ambos (es decir, el contexto de la petición es modificado o se añade información nueva en su interior). La tabla de contexto almacena información asociada con cada petición (p. ej., identidad de transacción, clase de petición, clase de servicio, origen de petición, y similares). El módulo de mensajería de agente 286 accede a la tabla de contexto y busca un registro de interés después de que el módulo de mensajería de agente 286 haya recibido la información de DTE 228. El gestor de la tabla de contexto 288 identifica una petición basada en la información asociada a la petición y los datos almacenados en la tabla de contexto. Así, el agente de contexto 236 obtiene información acerca de la petición que entra en el nivel, tal como la clase de petición, clase de servicio, y otra información pertinente asociada a la petición.

El gestor de la clasificación 290 y el gestor de políticas 292 procesan cada petición de usuario que entra en el primer nivel. La primera vez que una petición de usuario entra en el entorno informático no hay ninguna información en la tabla de contexto en lo que se refiere a esta petición de usuario. Como consecuencia, la petición de usuario es requerida para someterse a la clasificación. Durante el proceso de clasificación, toda la información conocida sobre la petición de usuario es recogida en un puerto de entrada de petición. Por ejemplo, si la petición de usuario es una petición HTTP, entonces la información de este tipo incluye el encabezamiento HTTP, la serie de preguntas, los parámetros URL, y similares. Para cada tipo de protocolo usado en cada una de los niveles, hay un plug-in genérico que en efecto clasifica la petición del usuario.

El proceso de clasificación esencialmente extrae la perspectiva de negocio desde la petición del usuario, traduciendo la información relacionada con la petición técnica (p. ej., una petición HTTP) en una clasificación de petición formal relacionada con un proceso de negocio (p. ej., recuperar el equilibrio de una cuenta bancaria). Una petición de usuario es colocada en una clase de petición específica. La petición de usuario puede ser identificada como parte de un proceso de negocio o una serie de transacciones relacionadas. Por ejemplo, la recuperación del saldo de una cuenta puede ser parte de un proceso mayor para solicitar una hipoteca. Cuando la información es transferida del agente de contexto 236 a CNMS 216, CNMS 216 puede determinar el perfil de actividad y detectar tendencias de comportamiento de petición para clases de petición.

El gestor de políticas 292 asigna una clase de servicio a las peticiones. El gestor de políticas 292 recibe la salida del gestor de la clasificación 290, y basándose en la clase de petición, el contexto de la petición, y la política de servicio activo, determina la clase de servicio de una petición dada. Por ejemplo, el gestor de políticas 292 puede establecer todas las peticiones de una clase de petición determinada para tener una clase de servicio con una prioridad más alta que todas las peticiones de una clase de petición diferente. Un agente de contexto asigna una clase de servicio a una petición que es procesada en el nivel asociado al mismo, actualizando el contexto de la petición con la clase de servicio apropiado según la política de servicio activo.

La base de datos de políticas de servicio de nivel cruzado incluye el grupo de mapeos de clases de servicios para clases de peticiones para todos los niveles en el entorno informático multinivel. A cada petición perteneciente a una clase de petición determinada se le puede asignar una clase de servicio diferente dependiendo del nivel en el que la petición está siendo procesada. Con referencia a la figura 2, a la petición del usuario R1 se le puede asignar una clase de servicio con una prioridad baja en el primer nivel 102, a la petición R2 se le puede asignar una clase de servicio con una alta prioridad en el segundo nivel 104 y a la petición R3 se le puede asignar una clase de servicio con una prioridad media en el tercer nivel 106. Las políticas de clase de servicio del nivel cruzado son almacenadas en la base de datos de políticas de SLO 218. Las políticas de clase de servicio de nivel cruzado pueden ser generadas automáticamente (es decir, sistema definido) o definidas por un usuario del sistema.

Además, un supervisor de un nivel local (p. ej., un administrador de base de datos) tiene el control administrativo del nivel y puede decidir anular una política de servicio en ese nivel si se cree necesario. El supervisor de nivel puede alterar la clase de servicio asignada a una petición por el agente de contexto de este nivel. El supervisor del nivel tiene una vista global sustancialmente de todo lo que se encuentra en el nivel. Se observa que también puede haber un supervisor global del sistema, que es normalmente una persona que lleva a cabo la aplicación de usuario y le interesa la imagen global.

Se observa que después de la clasificación inicial y la atribución de política de una petición de usuario, la clase de petición es mantenida para posteriores peticiones de la misma transacción, puesto que los agentes de contexto transfieren esta información de nivel a nivel (es decir, almacenada en la tabla de contexto de cada nivel). De esta manera, en cada nivel el agente de contexto identifica a qué clase de petición pertenece la petición, cuál es el servicio específico de la clase de nivel de la petición, y otras Informaciones asociadas a la petición.

El agente administrador 294 es una interfaz para CNMS 216. El agente administrador 294 informa de los datos históricos a CNMS 216. Por ejemplo, cuando una DTE indica que una UOW ha terminado, no es necesario que la

información asociada con la UOW permanezca en la tabla de contexto. La información es entonces enviada desde la tabla de contexto al agente administrador 294, que archiva la información y periódicamente manda la información a CNMS 216. El agente administrador 294 también recibe de CNMS 216 nuevas políticas de clase de servicio activo, nuevas configuraciones del agente de contexto, y similares. El agente administrador 294 puede también ser interrogado en tiempo real para obtener una indicación de estado. La indicación de estado puede incluir qué información está habitualmente en la tabla de contexto, qué UOWs se llevan a cabo ahora, y similares.

Ahora se hace referencia a la figura 9, que es un diagrama de bloques que demuestran los estadios implicados al capturar un contexto de la petición y su procesamiento posterior, operativo conforme a otra forma de realización de la técnica descrita. En el procedimiento 310, un contexto de la petición es capturado. Una petición del usuario nuevo es identificado y el contexto de la petición de éste es interceptado por el gancho blando de DTE. Con referencia a la figura 5, DTE 222 captura un contexto de la petición de una petición de usuario que ha entrado en el primer nivel 202.

En el procedimiento 312, el contexto de la petición capturado es normalizado. La DTE convierte el contexto de la petición en un formato estándar, idéntico para todos los tipos de nivel (p. ej., bases de datos, servidores de aplicación, y similares). La DTE asigna a la petición del usuario una única identificación (es decir, identidad de transacción) que identificará la petición de usuario y sus peticiones posteriores en la transacción iniciada por la petición de usuario. La DTE además obtiene la identidad de petición de la petición de usuario desde el nivel. Con referencia a la figura 5, DTE 222 convierte el contexto de la petición capturado en un formato estándar, asigna una identidad de transacción a la petición de usuario, y obtiene la identidad de petición de la petición de usuario.

En el procedimiento 314, la DTE manda el contexto de la petición, la identidad de transacción, y la identidad de petición al agente de contexto asociado a ese nivel. La DTE notifica al agente de contexto que un contexto de la petición nueva ha sido enviado. Con referencia a la figura 5, la DTE 222 manda el contexto de la petición, la identidad de transacción, y la identidad de petición de la petición de usuario al agente de contexto 232.

En el procedimiento 316, la petición de usuario es clasificada. El agente de contexto se aplica a un esquema de clasificación específica de nivel que determina la clase de petición basada en el contexto de la petición. Con referencia a la figura 5, el agente de contexto 232 clasifica la petición de usuario en una clase de petición determinada.

En el procedimiento 318, la política de servicio activo apropiada es recuperada de las políticas de clase de servicio almacenadas en el caché de política local. El agente de contexto recupera la política de servicio activo usando la clase de petición, y posiblemente otros campos de contexto de la petición. Con referencia a la figura 5, el agente de contexto 232 recupera la política de servicio activo apropiada para la petición de usuario, conforme a la clase de petición de la petición de usuario, y específica para el primer nivel 202.

En el procedimiento 320, la clase de servicio para la petición de usuario es determinada según la política de servicio recuperada, y asignada a la petición. El agente de contexto puede entonces añadir la clase de servicio asignado al contexto de la petición. Con referencia a la figura 5, el agente de contexto 232 asigna a la petición de usuario la clase de servicio de la política de servicio apropiado recuperada, y añade la clase de servicio al contexto de la petición.

En el procedimiento 322, la prioridad de nivel es determinada para la UOW invocada por la petición del usuario. Una prioridad local para el enclave de procesamiento que ejecuta la UOW es extraída de la clase de servicio. Con referencia a la figura 5, el agente de contexto 232 extrae la prioridad local para el enclave de procesamiento asignado a la petición del usuario para ejecutar la UOW, de la clase de servicio asignado.

En el procedimiento 324, el agente de contexto le envía la clase de petición e información de prioridad al DTE. Esta información es necesaria para ciertas tareas, tales como alterar la prioridad de un enclave de procesamiento, que se desarrolla dentro de la DTE. Con referencia a la figura 5, el agente de contexto 232 manda la clase de petición asignada y la prioridad local para el enclave de procesamiento ejecutando la UOW para DTE 222.

En el procedimiento 326, el contexto de la petición y la información relacionada (tales como clase de petición e identidad de transacción) es reenviado por otros agentes de contexto y por componentes internos del agente de contexto, sobre todo la tabla de contexto que almacena la petición, lo cual indexa el contexto de la petición y la información relacionada para una referencia adicional. Con referencia a la figura 5, el agente de contexto 232 adelanta el contexto de la petición, con información adicional acerca de la petición (p. ej., la identidad de petición, identidad de transacción, y clase de petición) a otros agentes de contexto acoplados a éste (p. ej., agente de contexto 234), al igual que para componentes internos del agente de contexto 232, tales como gestor de la tabla de contexto 288, gestor de la clasificación 290, y gestor de políticas 292 (con referencia a la figura 8).

Ahora se hace referencia a la figura 10, que es un diagrama de bloques que demuestra los estadios implicados al capturar una asignación UOW en un nivel local del sistema de la figura 5 y asocia una petición con la UOW, operativa conforme a otra forma de realización de la técnica descrita. Cabe destacar que una UOW es la lógica de aplicación que se ejecuta en el enclave de procesamiento asociado a la petición en ese nivel. En el procedimiento 340, una asignación UOW es capturada por la DTE. La captura puede ocurrir cuando la UOW es inicialmente invocado por el nivel, puesto que una petición de entrada entra en la cola de un puerto de entrada de la petición del nivel, mientras la petición de entrada permanece en la cola, o una vez que los recursos están disponibles en el nivel y la petición sale de la cola. Con referencia a la figura 5, la DTE 228 captura una asignación UOW mediante el cuarto nivel 208.

En el procedimiento 342, la DTE determina la identidad de la petición de la petición de entrada asociada a la asignación de UOW. La identidad de la petición se determina en base a la información enviada desde el nivel precedente, incluyendo el contexto de la petición junto con una clave de asociación que enlaza la UOW a la petición, tal como identificadores de zócalo, y similares. Con referencia a la figura 5, la DTE 228 determina la identidad de petición de la petición de introducción asociada a la asignación de UOW capturada, basada en la información recibida de un nivel acoplado precedente (p. ej., tercer nivel 206).

Pueden haber situaciones en las que el contexto de la petición y la clave de asociación no alcancen la DTE en el momento en que el procedimiento 342 se desarrolla. Como consecuencia el sistema podría ser configurado para proceder esperando esta información o asignando una identidad de petición temporal y asociando la petición de introducción con una asignación de UOW en una fase posterior.

En el procedimiento 344, la DTE envía información acerca de la identificación de UOW así como la identidad de petición determinada de la petición de entrada, al agente de contexto. Una identificación de UOW es un conjunto de características que únicamente identifican el enclave de procesamiento que ejecuta la UOW. La identificación de UOW es usada por el agente de contexto para seguirle la pista a la UOW. Con referencia a la figura 5, la DTE envía información de asociación para la UOW (p. ej., una clave de asociación que enlaza la UOW con la petición, tal como los identificadores de zócalo), al igual que la identidad de petición determinada de la petición entrante, para el agente de contexto 236.

En el procedimiento 346, el agente de contexto recupera la entrada en la tabla de contexto asociada a la identificación de UOW o la identidad de petición. Con referencia a la figura 5, el agente de contexto 232 localiza la entrada en la tabla de contexto mediante un gestor de la tabla de contexto 288 (figura 8) asociado a la asignación de UOW o la identidad de petición determinada.

Si se encuentra la entrada, entonces en el procedimiento 348, esa entrada es actualizada con la identificación de UOW e información relacionada. Con referencia a la figura 5, el agente de contexto 232 actualiza la entrada pertinente en la tabla de contexto mediante el gestor de la tabla de contexto 288 (figura 8) con la identificación de la asignación UOW capturada e información relacionada.

Si la entrada no es encontrada, entonces en el procedimiento 350, se añade una entrada nueva en la tabla de contexto. La entrada nueva incluye la identidad de petición de la petición de entrada y la identificación de la asignación de UOW asociada. La clase de petición por defecto y la clase de servicio están asociadas a la entrada recién añadida. Con referencia a la figura 5, el agente de contexto 232 añade una entrada nueva a la tabla de contexto mediante un gestor de la tabla de contexto 288 (figura 8) que incluye la identidad de petición de la petición de entrada y la identificación de la asignación de UOW, y asocia una clase de petición por defecto y una clase de servicio a la entrada recién añadida.

En el procedimiento 352, el agente de contexto determina la prioridad local para el enclave de procesamiento que ejecuta la UOW, y reúne una estadística de petición si es necesario. Con referencia a la figura 5, el agente de contexto 232 extrae la prioridad local para el enclave de procesamiento que ejecuta la UOW invocada por la petición de la clase de servicio asignada a la petición.

En el procedimiento 354, el agente de contexto impone la clase de servicio asignada a la petición de entrada. Con referencia a la figura 5, el agente de contexto 232 influye en el procesamiento de la petición entrante en el primer nivel 202, por ejemplo, alterando el nivel de prioridad del enclave de procesamiento que ejecuta la UOW invocada por la petición, alterando el tipo de ejecución del enclave de procesamiento, o asignando o denegando recursos computacionales para procesar la petición.

Se observa que el agente de contexto puede modificar posteriormente la clase de servicio, clase de petición, u otros parámetros del contexto de la petición, si el agente de contexto recibe información de asociación nueva acerca de la petición. Esto ocurre en una situación donde el contexto de la petición llega al agente de contexto asociado al nivel desde un agente de contexto remoto (p. ej., con referencia a la figura 5, el agente de contexto 234 recibe información sobre una petición de un agente de contexto 232). Debido a las salidas temporizadas, un agente de contexto puede capturar la asignación de UOW asociada a una petición, y luego recibir el contexto de la petición de la petición en una fase posterior.

Ahora se hace referencia a la figura 11, que es un diagrama de bloques que demuestra los estadios implicados al capturar una petición saliente enviada a un nivel remoto del sistema de la figura 5, y asociando la petición enviada con el contexto de la petición, operativo conforme a otra forma de realización de la técnica descrita. En el procedimiento 370, una petición enviada a un segundo nivel de un primer nivel es capturada por la DTE del primer nivel. Con referencia a la figura 5, una petición es enviada del primer nivel 202 al segundo nivel 204. La DTE 222 captura la petición saliente en el primer nivel 202.

En el procedimiento 372, la DTE determina la identidad de petición de la petición saliente. Con referencia a la figura 5, la DTE 222 determina la identidad de petición de una petición saliente enviada desde el primer nivel 202 al segundo nivel 204.

En el procedimiento 374, la DTE envía información acerca de la identificación de UOW, al igual que la identidad de petición determinada de la petición saliente, al agente de contexto local. Con referencia a la figura 5, la DTE 228 envía información acerca de la identificación de UOW, al igual que la identidad de petición determinada de la petición saliente, al agente de contexto 236.

En el procedimiento 376, el agente de contexto local recupera la entrada en la tabla de contexto asociada a la identidad de petición de la petición saliente o la identificación de UOW. Con referencia a la figura 5, el agente de contexto 232 localiza la entrada en la tabla de contexto mediante el gestor de la tabla de contexto 288 (figura 8) asociado con la identidad de la petición saliente determinada o mediante la identificación de UOW.

Si la entrada es encontrada, entonces en el procedimiento 378, esa entrada es actualizada con la identificación de UOW y la información relacionada. Por ejemplo, la entrada puede ser actualizada con información en el contexto de la petición de la petición saliente que habitualmente no está en la entrada. Con referencia a la figura 5, el agente de contexto 232 actualiza la entrada pertinente en la tabla de contexto mediante el gestor de la tabla de contexto 288 (figura 8) con la identificación de UOW y con información en el contexto de la petición de la petición saliente que no fue previamente almacenada en la entrada pertinente.

Si la entrada no es encontrada, entonces en el procedimiento 380, se añade una entrada nueva a la tabla de contexto. La entrada nueva incluye la identidad de petición de la petición saliente y la identificación de la asignación de UOW asociada. La clase de petición, la clase de servicio, y otras características de la petición almacenada en el contexto de la petición son agregadas a la entrada nueva. Si determinadas características no están presentes en el contexto de la petición, entonces las características por defecto (p. ej., la clase de petición y la clase de servicio) son asociadas a la entrada recién añadida. Con referencia a la figura 5, el agente de contexto 232 añade una entrada nueva a la tabla de contexto mediante el gestor de tabla de contexto 288 (figura 8) que incluye la identidad de petición de la petición saliente y características de la petición almacenada en el contexto de la petición de la petición saliente.

En el procedimiento 382, el agente de contexto local manda un mensaje al agente de contexto remoto (es decir, el agente de contexto al que la petición saliente fue enviada). El mensaje incluye la identidad de petición de la petición saliente, al igual que la identidad de transacción y la clase de petición. Con referencia a la figura 5, el agente de contexto 232 asociado al primer nivel 202 envía un mensaje al agente de contexto 235 asociado al segundo nivel 204. El mensaje incluye la identidad de petición de la petición saliente, al igual que la identidad de transacción y la clase de petición de la misma.

En otra forma de realización de la técnica descrita, puede haber un único agente de contexto asociado a huéspedes de nivel múltiple. Se observa que la manera en la que un agente de contexto se comunica con un DTE, hace posible que el agente de contexto resida en cualquier sitio en la red. También se observa que el agente de contexto sigue la pista de la petición y UOWs uniendo un identificador de nivel al mismo, de ese modo manteniendo la contabilidad para cada nivel. Tal forma de realización puede ser usada, por ejemplo, debido a las preocupaciones de seguridad o preocupaciones de extra superioridad de un usuario que resulta de la adición del agente de contexto al huésped del nivel.

Los expertos en la técnica apreciarán que la técnica descrita no se limita a lo que se ha mostrado particularmente y descrito anteriormente. Más bien el objetivo de la técnica descrita se define sólo por las reivindicaciones a continuación.

Documentos citados en la descripción

Esta lista de documentos citados por el solicitante ha sido recopilada exclusivamente para la información del lector y no forma parte del documento de patente europea. La misma ha sido confeccionada con la mayor diligencia; la OEP sin embargo no asume responsabilidad por eventuales errores u omisiones.

Documentos de patente citados en la descripción

- US 5958010 A [0010]
- US 6108700 A [0012]
- US 20020129137 A1 [0016]
- US 20040006602 A [0019]

REIVINDICACIONES

1. Aparato para controlar un nivel seleccionado en un entorno informático multinivel, comprendiendo dicho aparato:

un agente de contexto asociado a dicho nivel seleccionado de dicho entorno informático, dicho agente de contexto siendo acoplado a otros agentes de contexto, cada uno de dichos agentes de contexto asociados a un nivel respectivo de dicho entorno informático; y

una extensión de nivel dinámico, dicha extensión de nivel dinámico acoplada a dicho agente de contexto y con al menos un puerto de entrada de petición de dicho nivel seleccionado, al menos para controlar el tráfico de peticiones que pasa a través de dicho nivel seleccionado, dicho tráfico de petición controlado incluyendo al menos una petición de entrada recibida en dicho puerto de entrada de petición desde un nivel contiguo de dicho entorno informático, dicha extensión de nivel dinámico suministrando información de petición a dicho agente de contexto, dicha información de petición al menos identificando cada petición incluida en dicho tráfico de petición controlado,

donde dicho agente de contexto recibe información de contexto, de otro agente de contexto asociado a dicho nivel contiguo, dicha información de contexto recibida acerca del contexto de la petición de dicha al menos una petición entrante, dicha información de contexto recibida incluyendo al menos un único identificador de transacción; y

donde para cada una de dicha al menos una petición entrante, dicho agente de contexto asocia dicha información de contexto con dicha información de petición, respectivamente.

2. Aparato según la reivindicación 1, donde dicha extensión de nivel dinámico es posteriormente acoplada con al menos un puerto de salida de petición de dicho nivel seleccionado, y donde dicho tráfico de petición controlado incluye además al menos una petición saliente recibida en dicho al menos un puerto de salida de petición, dicho nivel seleccionado suministrando dicha al menos una petición saliente a otro nivel contiguo de dicho entorno informático,

donde dicho agente de contexto asocia entre dicha al menos una petición saliente y dicha al menos una petición entrante; y envía información de contexto acerca del contexto de la petición de dicha al menos una petición saliente a un agente de contexto adicional asociado a dicho otro nivel contiguo.

3. Aparato según la reivindicación 1, donde dicha extensión de nivel dinámico es posteriormente acoplada a un puerto de control de nivel de dicho nivel seleccionado.

4. Aparato según la reivindicación 3, donde dicho agente de contexto altera además el nivel de prioridad de un enclave de procesamiento que ejecuta dicha al menos una petición entrante.

5. Aparato según la reivindicación 3, donde dicho agente de contexto altera además el tipo de ejecución de un enclave de procesamiento que ejecuta dicha al menos una petición entrante.

6. Aparato según cualquiera de las reivindicaciones 1 ó 3, donde dicho agente de contexto ajusta además el orden de al menos dicha petición de entrada en una cola en dicho puerto de entrada de petición.

7. Aparato según cualquiera de las reivindicaciones 1 ó 3, donde dicho agente de contexto instruye además dicho nivel seleccionado para asignar recursos computacionales para procesar dicha al menos una petición entrante.

8. Aparato según cualquiera de las reivindicaciones 1 ó 3, donde dicho agente de contexto instruye además dicho nivel seleccionado para denegar recursos computacionales para procesar dicha al menos una petición entrante.

9. Aparato según la reivindicación 1, donde dicha información de contexto incluye además al menos un identificador de petición, una clase de petición, y datos relacionados con el contexto.

10. Aparato según la reivindicación 9, donde dichos datos relacionados con el contexto incluyen todo el contexto de la petición de dicha al menos una petición entrante.

11. Aparato según la reivindicación 9, donde dichos datos relacionados con el contexto incluyen una parte de dicho contexto de dicha al menos una petición entrante.

12. Aparato según la reivindicación 9, donde dichos datos relacionados con el contexto incluyen una indicación para dicho contexto de la petición de dicha al menos una petición entrante.

13. Aparato según la reivindicación 2, donde dicho agente de contexto comprende:

una interfaz de un agente de comunicación, transmisión de mensajes de entre dicha extensión de nivel dinámico y dicho agente de contexto;

un módulo de mensajería de agente, que notifica a dicho agente de contexto adicional que dicha al menos una petición saliente ha sido enviada a dicho otro nivel contiguo, dicho módulo de mensajería de agente además se comunica con dicha extensión de nivel dinámico usando mensajes; y

un gestor de tabla de contexto, manteniendo una tabla de referencia que asocia al menos una unidad de trabajo ejecutada en dicho nivel seleccionado con dicho contexto de la petición de al menos dicha petición entrante para la cual dicha unidad de trabajo ha sido asignada, dicha tabla de referencia siendo una tabla de contexto, dicha tabla de contexto además almacena información asociada con dicha al menos una petición entrante,

dicho gestor de la tabla de contexto además identifica dicha al menos una petición entrante basada en información almacenada en dicha tabla de contexto.

14. Aparato según la reivindicación 13, donde dicho agente de contexto además comprende:

un gestor de clasificación, que realiza la clasificación de dicha al menos una petición entrante basada en dicha información de contexto recibida, dicho gestor de clasificación además identifica dicha al menos una petición entrante como parte de una transacción.

15. Aparato según la reivindicación 13, donde dicho agente de contexto además comprende:

un gestor de políticas, asignando una política de servicio para dicha al menos una petición entrante según una política de servicio activo y dicha clasificación de dicha al menos una petición entrante.

16. Aparato según la reivindicación 13, donde dicho agente de contexto además comprende:

un manipulador de eventos de registro de agente, que realiza los registros de eventos, trazado interno y mensajería, y detección de irregularidades operacionales dentro de dicho agente de contexto.

17. Aparato según la reivindicación 13, donde dicho agente de contexto además comprende:

un administrador de agente, archivo de información, envío de datos históricos a un servidor de gestión de red de contexto (CNMS) y recepción de políticas de clase de servicio activo y datos adicionales de dicho CNMS.

18. Aparato según la reivindicación 3, donde dicha extensión de nivel dinámico comprende:

un demonio de extensión de nivel dinámico, que realiza un procesamiento asincrónico asociado a eventos proporcionados donde dicha al menos una petición entrante no es detenida;

un módulo de mensajería de la extensión de nivel dinámico, que se comunica con dicho agente de contexto usando mensajes; y

una interfaz de comunicación de extensión de nivel dinámico, transfiriendo mensajes entre dicha extensión de nivel dinámico y dicho agente de contexto.

19. Aparato según la reivindicación 18, donde dicha extensión de nivel dinámico además comprende:

un administrador de extensión de nivel dinámico, que permite a dicha extensión de nivel dinámico recibir mensajes acerca de cómo se debería procesar dicha al menos una petición entrante.

20. Aparato según la reivindicación 18, donde dicha extensión de nivel dinámico además comprende:

un manipulador de eventos de registro de extensión de nivel dinámico, información de trazado de registro referente a la operación de dicha extensión de nivel dinámico y registro de alertas aumentadas por dicha extensión de nivel dinámico.

21. Aparato según la reivindicación 1, donde dicho agente de contexto reside en la misma máquina huésped que dicho nivel seleccionado.

22. Aparato según la reivindicación 3, donde dicha extensión de nivel dinámico recoge rendimiento, disponibilidad, y error métrico de dicho nivel seleccionado.

23. Aparato para controlar un nivel seleccionado en un entorno informático multinivel, dicho aparato comprendiendo:

un agente de contexto asociado a dicho nivel seleccionado, dicho agente de contexto siendo acoplado con al menos otro agente de contexto, cada uno de los otros agentes de contexto siendo asociados a un nivel respectivo de dicho entorno informático; y

una extensión de nivel dinámico, dicha extensión de nivel dinámico acoplada con dicho agente de contexto y con al menos un puerto de entrada de petición y un puerto de salida de petición de dicho nivel seleccionado, dicha extensión de nivel dinámico al menos controlando el tráfico de petición que pasa a través de dicho nivel seleccionado, dicho tráfico de petición controlado incluyendo al menos una petición entrante recibida en dicho puerto de entrada de petición y al menos una petición saliente respectiva en dicho puerto de salida de petición, dicho nivel seleccionado suministrando dicha al menos una petición saliente respectiva a un nivel contiguo, dicha extensión de nivel dinámico capturando un contexto de la petición de dicha al menos una petición entrante, dicha extensión de nivel dinámico asignando un único identificador de transacción a dicho contexto de la petición capturado, dicha extensión de nivel dinámico suministrando información de petición e información de contexto a dicho agente de contexto, dicha información de petición al menos identificando cada petición incluida en dicho tráfico de petición controlado, dicha información de contexto acerca de dicho contexto de petición capturado de dicha al menos una petición entrante, dicha información de contexto incluyendo al menos dicho único identificador de transacción,

donde dicho agente de contexto asocia dicha al menos una petición entrante con al menos una petición saliente, y

donde dicho agente de contexto proporciona información de contexto acerca del contexto de la petición de dicha al menos una petición saliente respectiva a otro agente de contexto asociado a dicho nivel contiguo.

24. Aparato según la reivindicación 23, donde dicha al menos una petición entrante es una petición del usuario recibida de una aplicación de usuario.

25. Aparato según la reivindicación 23, donde dicha información de contexto incluye además al menos un identificador de petición, una clase de petición, y datos relacionados con el contexto.

26. Aparato según la reivindicación 23, donde dicho agente de contexto clasifica dicha al menos una petición entrante en una clase de petición conforme a dicha información de contexto.

27. Aparato según la reivindicación 26, donde dicho agente de contexto contiene un grupo de políticas de clase de servicio específicas del nivel, cada una de dichas políticas de clase de servicio específicas del nivel mapeando una clase de servicio a una clase de petición para dicho nivel seleccionado,

donde una política de servicio activo contiene la clase de petición para mapeo de clase de servicio que está habitualmente vigente; y

donde dicho agente de contexto asigna una clase de servicio a dicha al menos una petición entrante basada en dicha clase de petición y la política de servicio activo apropiada en dicho grupo de políticas de clase de servicio de nivel específico.

28. Aparato según la reivindicación 26, donde dicho agente de contexto contiene un grupo de políticas de clase de servicio específicas del nivel, cada una de dichas políticas de clase de servicio específicas del nivel trazando una clase de servicio para una clase de petición para dicho nivel seleccionado,

donde una política de servicio activo contiene la clase de petición para la clase de servicio trazado que está habitualmente vigente; y

donde dicho agente de contexto asigna una clase de servicio para dicha al menos una petición entrante diferente de la clase de servicio designada por dicha clase de petición y la política de servicio activo apropiada en dicho grupo de políticas de clase de servicio específicas del nivel.

29. Aparato según cualquiera de las reivindicaciones 27 ó 28, donde dicha clase de servicio incluye la prioridad que debe ser asignada a dicha al menos una petición entrante.

30. Aparato según cualquiera de las reivindicaciones 27 ó 28, donde dicha clase de servicio incluye el porcentaje de unidad de procesamiento central que debe ser asignado a dicha al menos una petición entrante.

31. Aparato según cualquiera de las reivindicaciones 27 ó 28, donde dicha clase de servicio incluye la memoria que debe ser asignada a dicha al menos una petición entrante.

32. Aparato según cualquiera de las reivindicaciones 27 ó 28, donde dicha clase de servicio incluye la prioridad para asignar y acceder a los dispositivos de entrada y de salida para dicha al menos una petición entrante.

33. Aparato según cualquiera de las reivindicaciones 27 ó 28, donde dichas políticas de clase de servicio son actualizadas periódicamente.

34. Aparato según cualquiera de las reivindicaciones 27 ó 28, donde dichas políticas de clase de servicio son definidas por un usuario de dicho entorno informático multinivel.

35. Aparato según cualquiera de las reivindicaciones 27 ó 28, donde dicho agente de contexto añade dicha clase de servicio asignada a dicho contexto de petición.

36. Aparato según la reivindicación 23, donde dicha extensión de nivel dinámico es además acoplada a un puerto de control de nivel de dicho nivel seleccionado.

37. Aparato según la reivindicación 36, donde dicho agente de contexto altera además el nivel de prioridad de un enclave de procesamiento que ejecuta dicha al menos una petición entrante.

38. Aparato según la reivindicación 36, donde dicho agente de contexto altera además el tipo de ejecución de un enclave de procesamiento que ejecuta dicha al menos una petición entrante.

39. Aparato según cualquiera de las reivindicaciones 23 ó 36, donde dicho agente de contexto ajusta además el orden de dicha al menos una petición entrante en una cola en dicho puerto de entrada de petición.

40. Aparato según cualquiera de las reivindicaciones 23 ó 36, donde dicho agente de contexto instruye además dicho nivel seleccionado para asignar recursos computacionales para procesar dicha al menos una petición entrante.

41. Aparato según cualquiera de las reivindicaciones 23 ó 36, donde dicho agente de contexto instruye además dicho nivel seleccionado para denegar recursos computacionales para procesar dicha al menos una petición entrante.

42. Aparato según la reivindicación 23, donde dicha extensión de nivel dinámico asigna un identificador de transacción para cada petición incluida en dicho tráfico de petición controlado.

43. Aparato según la reivindicación 23, donde dicha extensión de nivel dinámico obtiene un identificador de petición de cada petición incluida en dicho tráfico de petición controlado.

44. Aparato según la reivindicación 23, donde dicho agente de contexto comprende:

una interfaz de agente de comunicación, transmisión de mensajes entre dicha extensión de nivel dinámico y dicho agente de contexto;

un módulo de mensajería de agente, que notifica a dicho otro agente de contexto que dicha al menos una petición saliente ha sido enviada a dicho otro nivel contiguo, dicho módulo de mensajería de agente además se comunica con dicha extensión de nivel dinámico usando mensajes;

un gestor de tabla de contexto, manteniendo una tabla de referencia que asocia al menos una unidad de trabajo ejecutada en dicho nivel seleccionado con dicho contexto de petición de dicha al menos una petición entrante a la que dicha unidad de trabajo ha sido asignada, dicha tabla de referencia siendo una tabla de contexto, dicha tabla de contexto almacenando información asociada a dicha al menos una petición entrante, dicho gestor de tabla de contexto además identificando al menos una petición entrante basada en información almacenada en dicha tabla de contexto;

un gestor de clasificación, que realiza la clasificación de dicha al menos una petición entrante basada en dicha información de contexto recibida, dicho gestor de clasificación además identifica a dicha al menos una petición entrante como parte de una transacción; y

un gestor de políticas, que asigna una política de servicio a dicha al menos una petición entrante según una política de servicio activo y dicha clasificación de dicha al menos una petición entrante.

45. Aparato según la reivindicación 44, donde dicho agente de contexto además comprende:

un manipulador de eventos de registro de agente, que realiza registros de eventos, trazado interno y mensajería, y detección de irregularidades operacionales dentro de dicho agente de contexto.

46. Aparato según la reivindicación 44, donde dicho agente de contexto además comprende:

un administrador de agente, archivo de información, envío de datos históricos a un servidor de gestión de red de contexto (CMNS) y recepción de políticas de clase de servicio activo y datos adicionales de dicho CNMS.

47. Aparato según la reivindicación 36, donde dicha extensión de nivel dinámico comprende:

un demonio de extensión de nivel dinámico, que realiza un procesamiento asincrónico asociado a eventos proporcionados donde al menos una petición entrante no es detenida;

un módulo de mensajería de extensión de nivel dinámico, que se comunica con dicho agente de contexto usando mensajes; y

una interfaz de comunicación de extensión de nivel dinámico, que transfiere mensajes entre dicha extensión de nivel dinámico y dicho agente de contexto.

48. Aparato según la reivindicación 47, donde dicha extensión de nivel dinámico además comprende:

un administrador de extensión de nivel dinámico, que le permite a dicha extensión de nivel dinámico recibir mensajes sobre cómo al menos una petición entrante debería ser procesada.

49. Aparato según la reivindicación 47, donde dicha extensión de nivel dinámico además comprende:

un manipulador de registro de eventos de nivel dinámico, que traza el registro de información en lo que se refiere a la operación de dicha extensión de nivel dinámico y registro de alertas aumentado por dicha extensión de nivel dinámico.

50. Aparato según la reivindicación 23, donde dicho agente de contexto reside en la misma máquina huésped que dicho nivel seleccionado.

51. Aparato según la reivindicación 36, donde dicha extensión de nivel dinámico recoge rendimiento, disponibilidad, y error métrico de dicho nivel seleccionado.

52. Entorno informático multinivel incluyendo una pluralidad de niveles, un sistema para la gestión de rendimiento de aplicación, asociada a al menos los niveles seleccionados vigilados de dicha pluralidad de niveles, comprendiendo el sistema:

para cada uno de dichos al menos dos niveles vigilados, una extensión de nivel dinámico respectiva que es acoplada con al menos un puerto de entrada de petición de dicho nivel controlado, dicha extensión de nivel dinámico al menos controlando el tráfico de petición que pasa a través de dicho nivel respectivo controlado, dicho tráfico de petición controlado incluyendo al menos una petición entrante recibida en dicho puerto de entrada de petición; y

para cada uno de dichos al menos dos de los niveles controlados, un agente de contexto respectivo, acoplado con dicha extensión de nivel dinámico respectiva, dicho agente de contexto siendo acoplado con otros dichos agentes de contexto asociados al nivel que son directamente acoplados con dicho nivel respectivo, dicha información de petición de la extensión de nivel dinámico, dicha información de petición al menos identificando cada petición incluida en dicho tráfico de petición controlado, dicho agente de contexto respectivo recibe además información de contexto acerca del contexto de la petición de al menos una petición entrante de otro agente de contexto asociado a dicho nivel contiguo, dicha información de contexto recibida incluyendo al menos un único identificador de transacción, para cada uno al menos una petición entrante, dicho agente de contexto asocia dicha información de contexto respectiva con dicha información de petición, respectivamente.

53. Entorno informático multinivel incluyendo una pluralidad de niveles, un sistema para la gestión de rendimiento de aplicación, asociada con al menos niveles vigilados seleccionados de dicha pluralidad de niveles, el sistema comprendiendo:

para cada uno de al menos dos de dichos niveles vigilados, una extensión de nivel dinámico respectiva que es acoplada con al menos un puerto de entrada de petición y un puerto de salida de petición de dicho nivel controlado, dicha extensión de nivel dinámico al menos controlando el tráfico de petición que pasa a través de dicho nivel respectivo vigilado, dicho tráfico de petición controlado incluyendo al menos una petición de entrada recibida en dicho puerto de entrada de petición y al menos una petición saliente respectiva en dicho puerto de salida de petición, dicho nivel seleccionado suministrando al menos una petición saliente respectiva a un nivel contiguo respectivo vigilado, dicha extensión de nivel dinámico capturando un contexto de petición de al menos una petición entrante, dicha extensión de nivel dinámico asignando un único identificador de transacción para dicho contexto de la petición capturado; y

para cada una de al menos dos de dichos niveles vigilados, un agente de contexto respectivo, acoplado con dicha extensión de nivel dinámico respectiva, dicho agente de contexto siendo acoplado con otros de dichos agentes de contexto asociados a niveles que son directamente acoplados con dicho nivel respectivo, dicho agente de contexto respectivo recibiendo información de petición de dicha extensión de nivel dinámico, dicha información de petición al menos identificando cada petición incluida en dicho tráfico de petición controlado, dicho agente de contexto respectivo recibe además información de contexto acerca del contexto de la petición de dicha al menos una petición entrante de dicha extensión de nivel dinámico respectiva, dicha información de contexto recibida incluye al menos dicho identificador único de transacción, para cada uno al menos una petición entrante, dicho agente de contexto respectivo asocia al menos una petición entrante con dicha al menos una petición saliente, y dicho agente de contexto respectivo proporciona información de contexto acerca del contexto de la petición de al menos una petición saliente respectiva a otro agente de contexto asociado a dicho nivel contiguo, dicha información de contexto proporcionada incluyendo al menos dicho único identificador de transacción.

54. Sistema según cualquiera de las reivindicaciones 52 ó 53, comprendiendo además:

un servidor de gestión de red de contexto (CMNS) acoplado con dichos agentes de contexto, dicho CNMS al menos recopilando y analizando los datos de rendimiento recibidos de dichos agentes de contexto.

55. Sistema según cualquiera de las reivindicaciones 52 ó 53, donde al menos una petición entrante es una petición de usuario recibida de una aplicación de usuario.

56. Sistema según la reivindicación 52, donde cada una de dichas extensiones de nivel dinámico son posteriormente acopladas con al menos un puerto de salida de petición de dicho nivel respectivo, y donde dicho tráfico de petición controlado incluye además al menos una petición saliente recibida en dicho puerto de salida de petición, dicho nivel seleccionado suministra al menos una petición saliente a otro nivel contiguo de dicho entorno informático,

donde dicho agente de contexto asocia al menos una petición que existe con al menos una petición entrante; y manda información de contexto acerca del contexto de la petición de al menos una petición saliente a un agente de contexto adicional asociado a dicho otro nivel contiguo.

57. Sistema según la reivindicación 54, que además incluye una base de datos de políticas acoplada con dicho CNMS, dicha base de datos de políticas almacena al menos políticas de clase de servicio.

58. Sistema según la reivindicación 57, que además incluye una estación de trabajo de supervisor acoplado con dicho CNMS y con dicha base de datos de políticas.

59. Sistema según la reivindicación 54, donde dicho CNMS ejecuta además un perfilado de la actividad, creando un perfil de actividad incluyendo transacción integrada, nivel, y métrica de rendimiento de nivel de sistema y análisis estadístico.

60. Método de gestión de rendimiento de aplicación para controlar, en un entorno informático multinivel incluyendo una pluralidad de niveles, un nivel seleccionado vigilado de dicha pluralidad niveles, el método comprendiendo los procedimientos de:

recibir información de contexto acerca del contexto de la petición de al menos una petición entrante, dicha información de contexto incluyendo al menos un identificador de petición y un único identificador de transacción;

controlar el tráfico de petición que pasa a través de dicho nivel controlado, dicho tráfico de petición controlado incluyendo al menos dicha petición entrante;

identificar dicha al menos una petición entrante conforme a dicho identificador de petición; y

asociar ésta con dicha al menos una petición entrante con una transacción conforme a dicho identificador de transacción.

61. Método de gestión de rendimiento de aplicación para controlar, en un entorno informático multinivel incluyendo una pluralidad de niveles, un nivel seleccionado controlado de dicha pluralidad de niveles, el método comprendiendo los procedimientos de:

control de tráfico de petición que pasa a través de dicho nivel controlado, dicho tráfico de petición controlado incluyendo al menos una petición entrante y al menos una petición saliente respectiva, dicho nivel seleccionado suministrando al menos una petición saliente respectiva a un nivel contiguo;

determinar información de contexto acerca del contexto de la petición de al menos una petición entrante, dicha información de contexto incluyendo al menos un único identificador de transacción;

identificar cada petición incluida en dicho tráfico de petición controlada;

asociar al menos una petición entrante con al menos una petición saliente; y enviar información de contexto acerca del contexto de la petición de dicha petición saliente a otro agente de contexto asociado a dicho nivel contiguo, dicha información de contexto incluyendo al menos dicho único identificador de transacción.

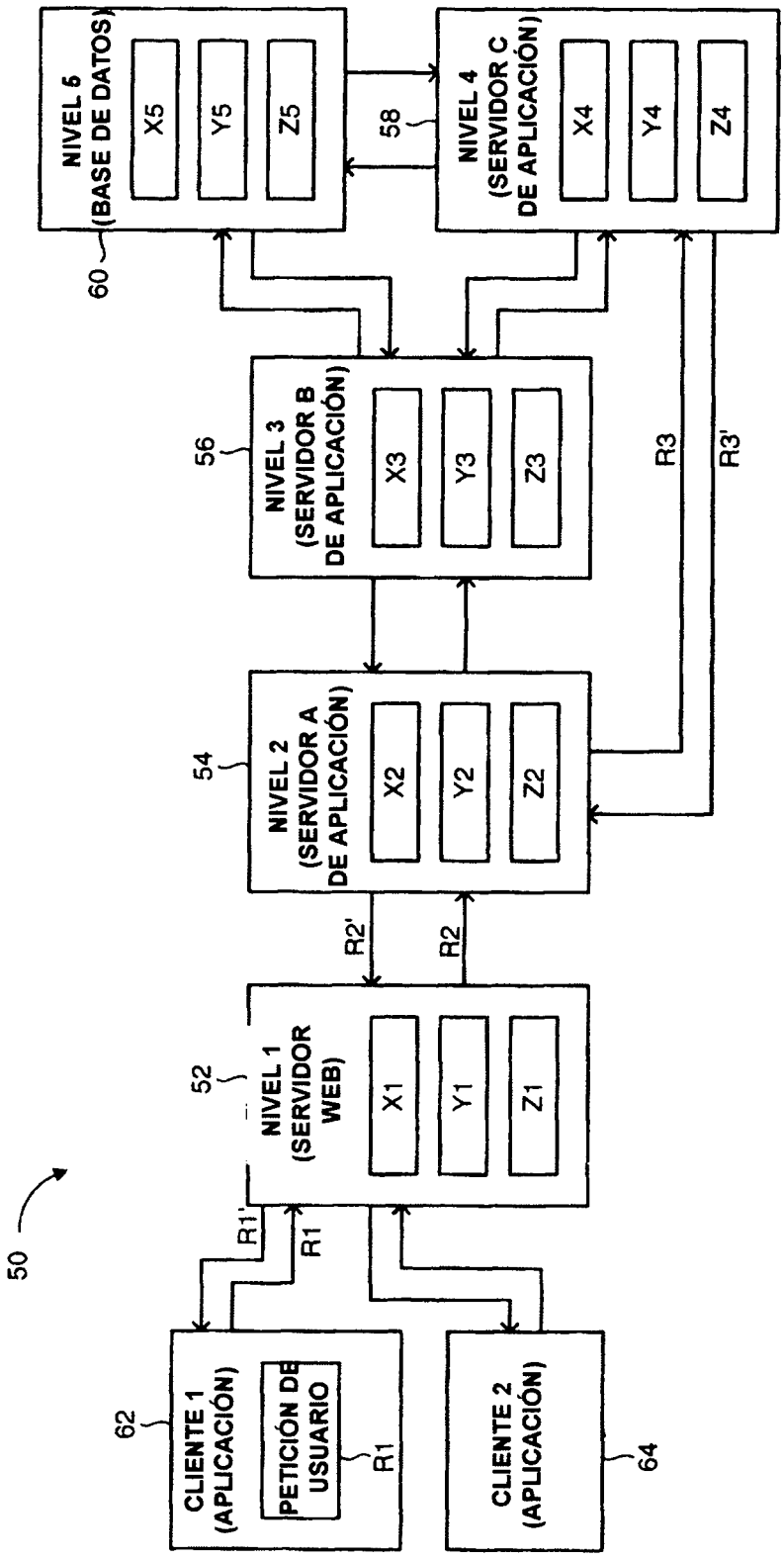

FIG. 1
TÉCNICA ANTERIOR

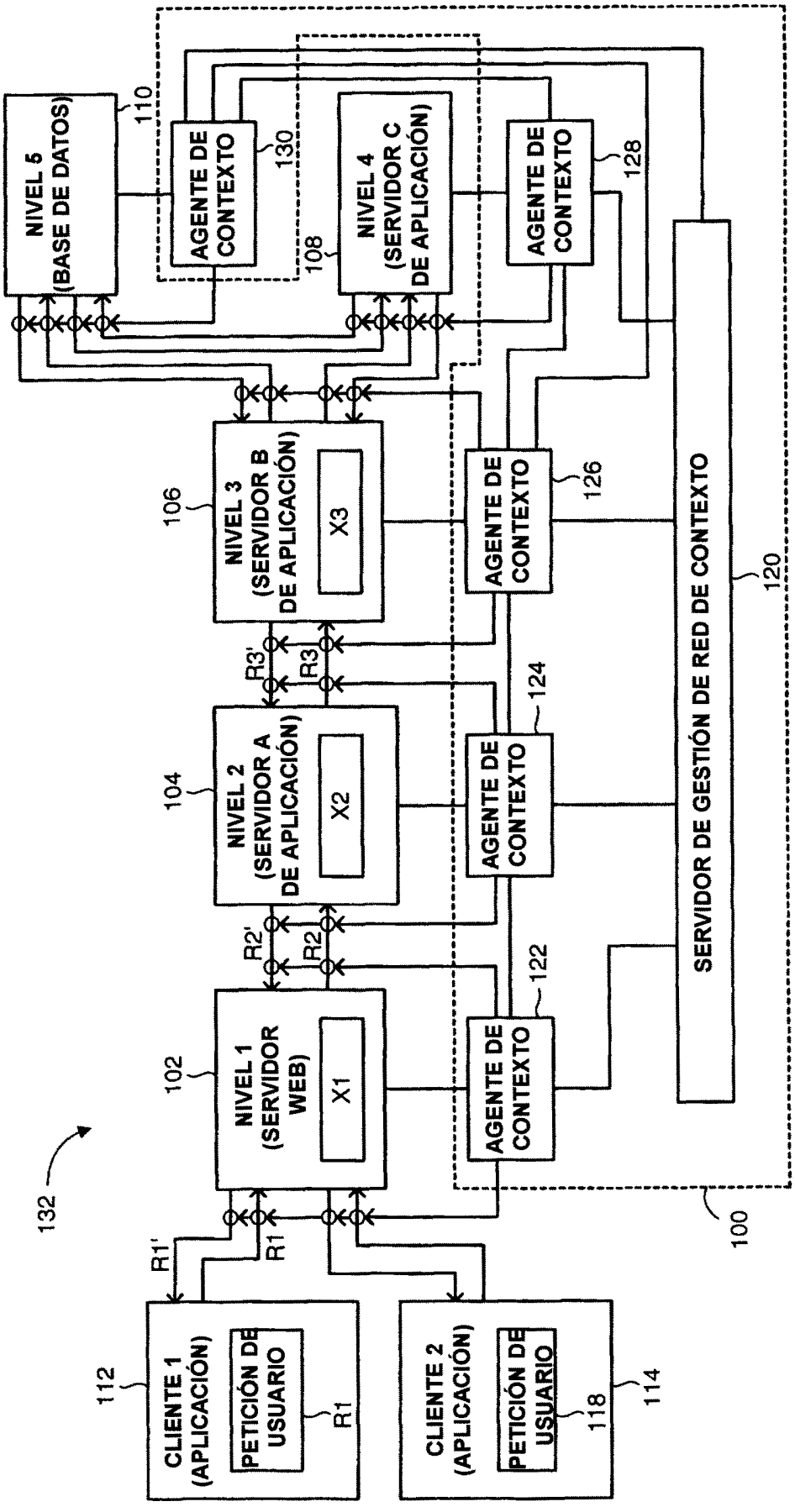
NIVEL 5 (BASE DE DATOS)
110
AGENTE DE CONTEXTO
130
NIVEL 4 (SERVIDOR C DE APLICACIÓN)
108
AGENTE DE CONTEXTO
128
NIVEL 3 (SERVIDOR B DE APLICACIÓN)
X3
106
AGENTE DE CONTEXTO
126
R3'
R3
NIVEL 2 (SERVIDOR A DE APLICACIÓN)
X2
104
AGENTE DE CONTEXTO
124
R2'
R2
NIVEL 1 (SERVIDOR WEB)
X1
102
AGENTE DE CONTEXTO
122
SERVIDOR DE GESTIÓN DE RED DE CONTEXTO
120
132
R1'
R1
100
CLIENTE 1 (APLICACIÓN)
PETICIÓN DE USUARIO
R1
112
CLIENTE 2 (APLICACIÓN)
PETICIÓN DE USUARIO
118
114

FIG. 2

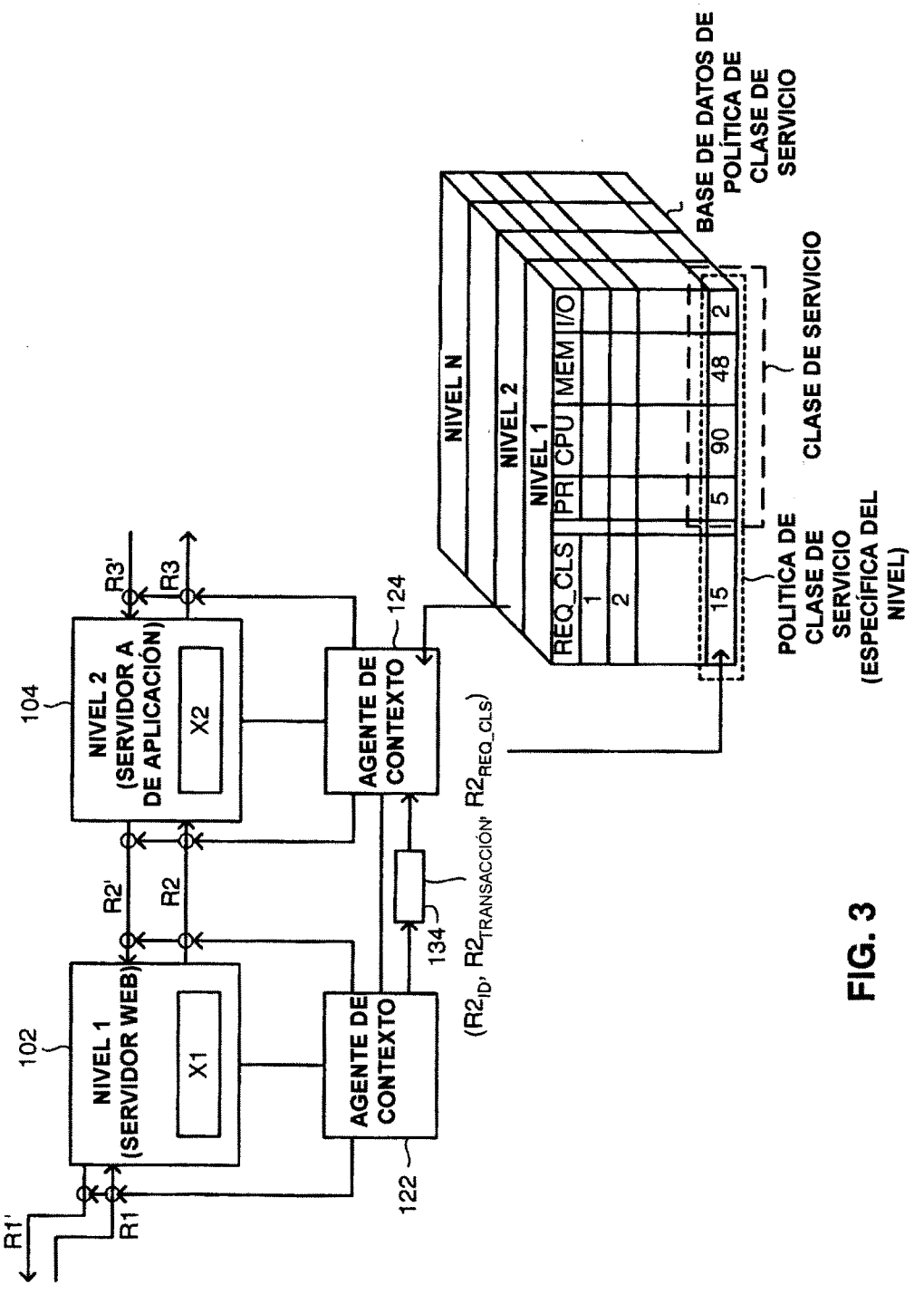
R1'
R1
102
NIVEL 1
(SERVIDOR WEB)
X1
R2'
R2
104
NIVEL 2
(SERVIDOR A
DE APLICACIÓN)
X2
R3'
R3
AGENTE DE
CONTEXTO
122
AGENTE DE
CONTEXTO
124
134
(R2ID, R2TRANSACCIÓN, R2REQ_CLS)
NIVEL N
NIVEL 2
NIVEL 1
REQ_CLS
PR
CPU
MEM
I/O
1
2
15
5
90
48
2
BASE DE DATOS DE
POLÍTICA DE
CLASE DE
SERVICIO
CLASE DE SERVICIO
POLITICA DE
CLASE DE
SERVICIO
(ESPECÍFICA DEL
NIVEL)

FIG. 3

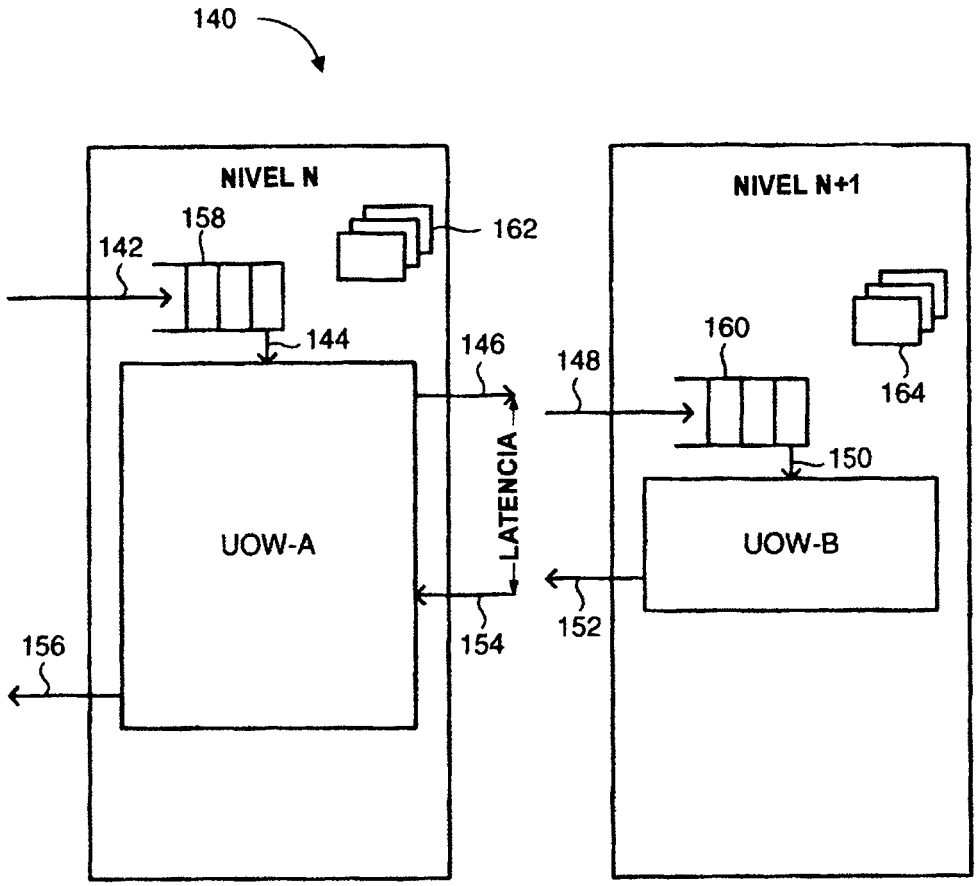
140
NIVEL N
NIVEL N+1
158
162
142
144
146
148
160
164
150
UOW-A
LATENCIA
UOW-B
156
154
152

FIG. 4

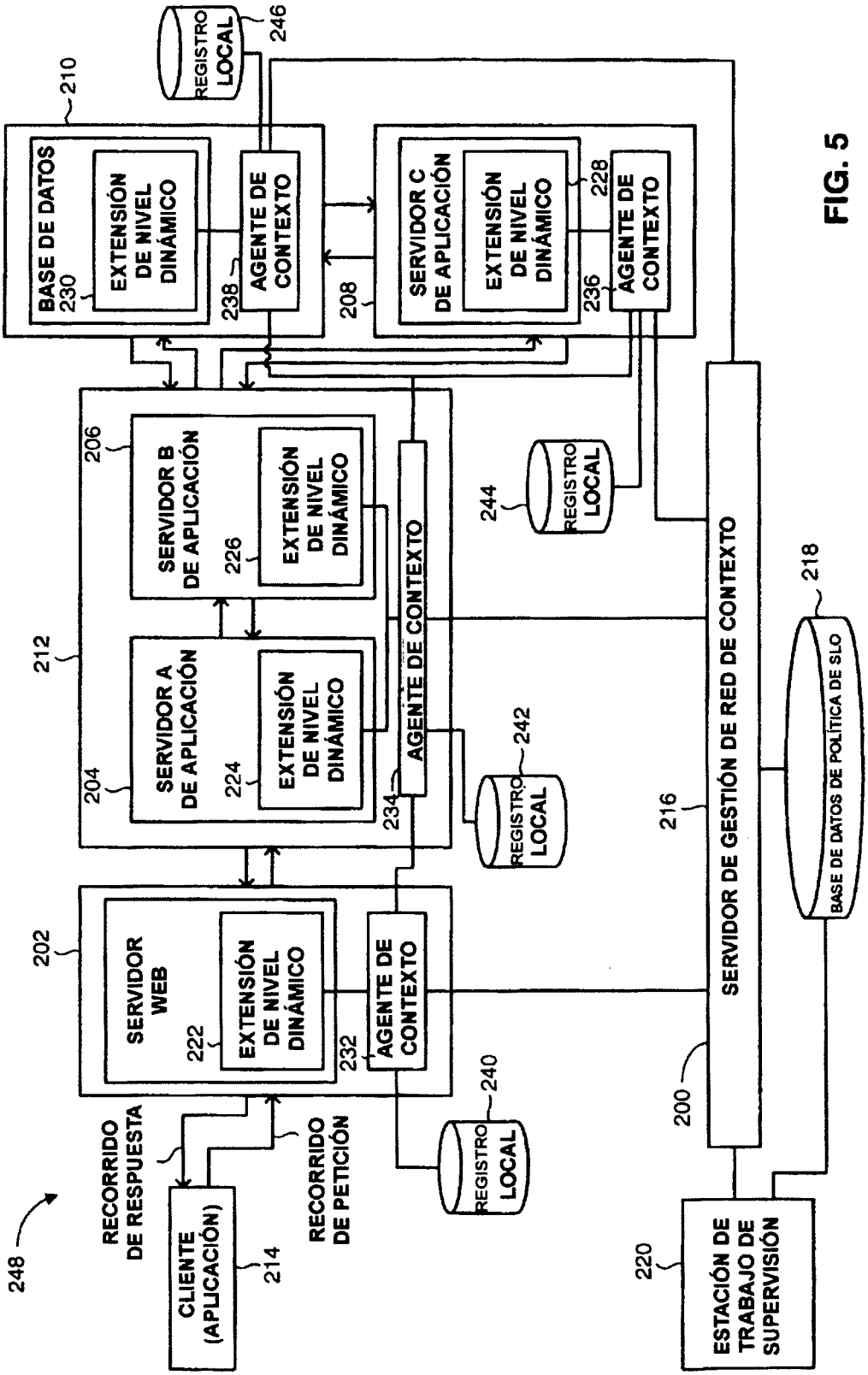
REGISTRO LOCAL
246
210
BASE DE DATOS
230
EXTENSIÓN DE NIVEL DINÁMICO
AGENTE DE CONTEXTO
238
208
SERVIDOR C DE APLICACIÓN
EXTENSIÓN DE NIVEL DINÁMICO
228
AGENTE DE CONTEXTO
236
206
SERVIDOR B DE APLICACIÓN
226
EXTENSIÓN DE NIVEL DINÁMICO
212
SERVIDOR A DE APLICACIÓN
224
EXTENSIÓN DE NIVEL DINÁMICO
204
AGENTE DE CONTEXTO
234
244
REGISTRO LOCAL
242
REGISTRO LOCAL
216
SERVIDOR DE GESTIÓN DE RED DE CONTEXTO
218
BASE DE DATOS DE POLÍTICA DE SLO
202
SERVIDOR WEB
222
EXTENSIÓN DE NIVEL DINÁMICO
232
AGENTE DE CONTEXTO
240
REGISTRO LOCAL
200
248
RECORRIDO DE RESPUESTA
RECORRIDO DE PETICIÓN
CLIENTE (APLICACIÓN)
214
220
ESTACIÓN DE TRABAJO DE SUPERVISIÓN

FIG. 5

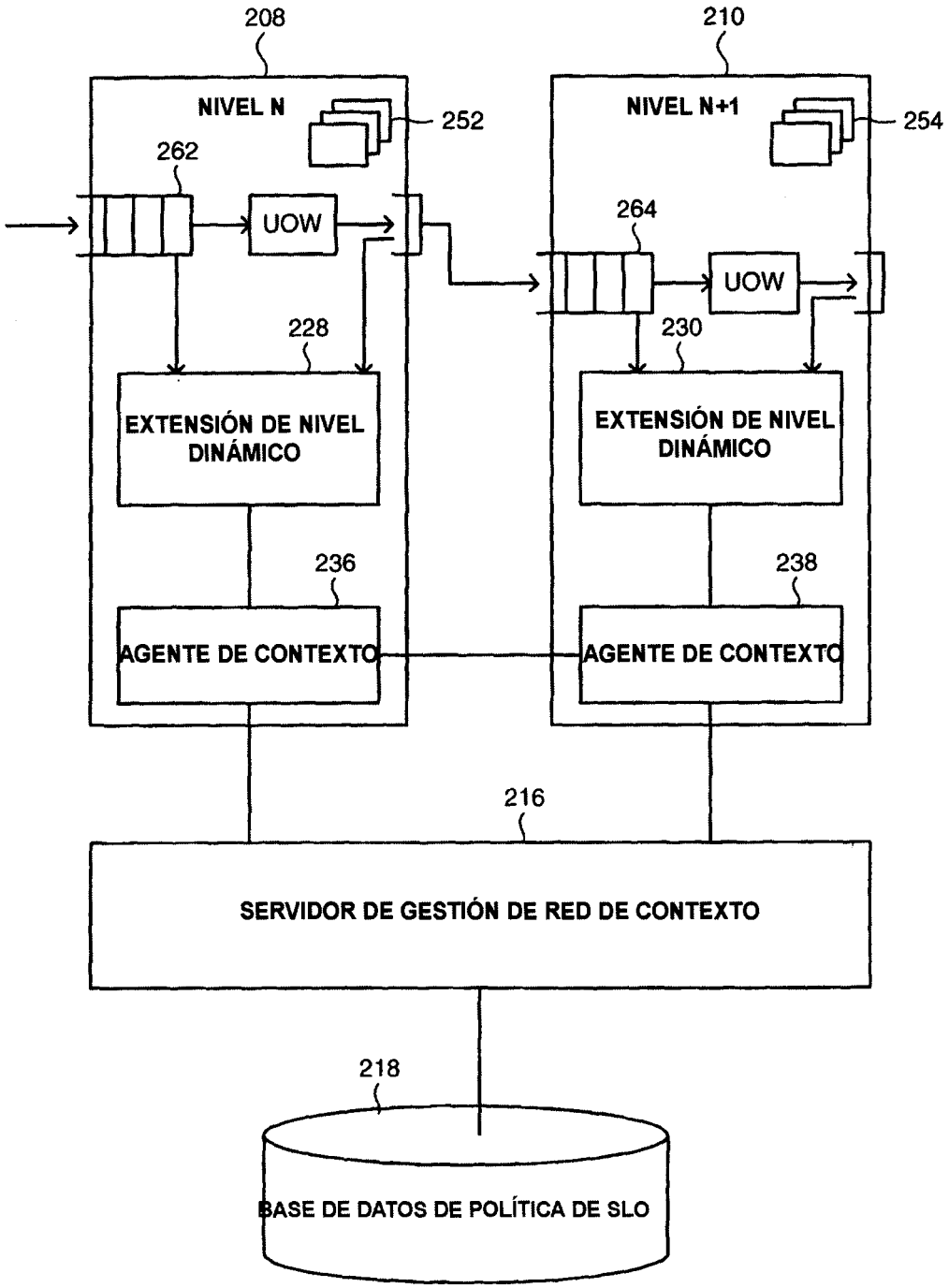
208
210
NIVEL N
252
NIVEL N+1
254
262
UOW
264
UOW
228
230
EXTENSIÓN DE NIVEL DINÁMICO
EXTENSIÓN DE NIVEL DINÁMICO
236
238
AGENTE DE CONTEXTO
AGENTE DE CONTEXTO
216
SERVIDOR DE GESTIÓN DE RED DE CONTEXTO
218
BASE DE DATOS DE POLÍTICA DE SLO

FIG. 6

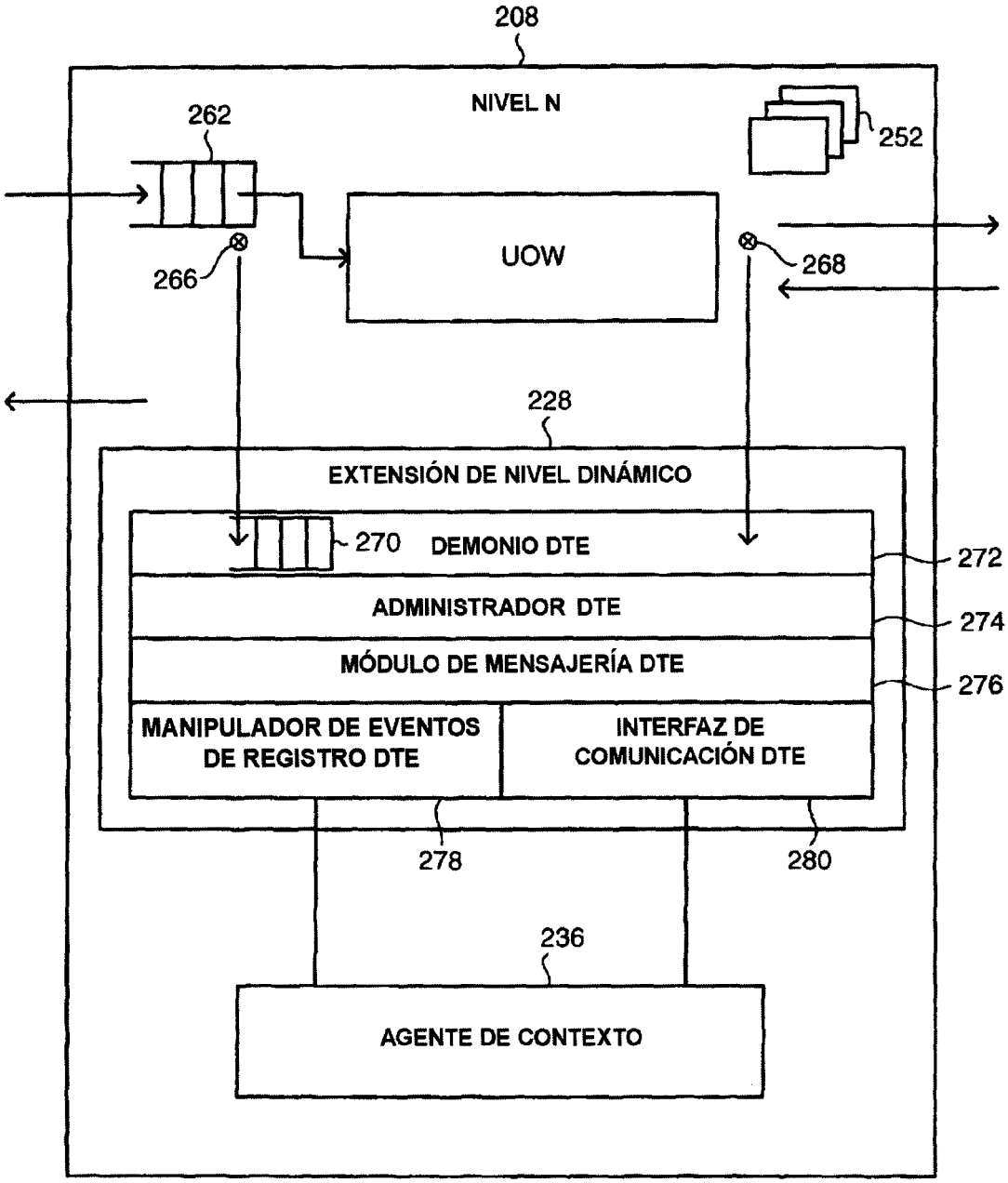
208
NIVEL N
262
252
UOW
266
268
228
EXTENSIÓN DE NIVEL DINÁMICO
270
DEMONIO DTE
272
ADMINISTRADOR DTE
274
MÓDULO DE MENSAJERÍA DTE
276
MANIPULADOR DE EVENTOS DE REGISTRO DTE
INTERFAZ DE COMUNICACIÓN DTE
278
280
236
AGENTE DE CONTEXTO

FIG. 7

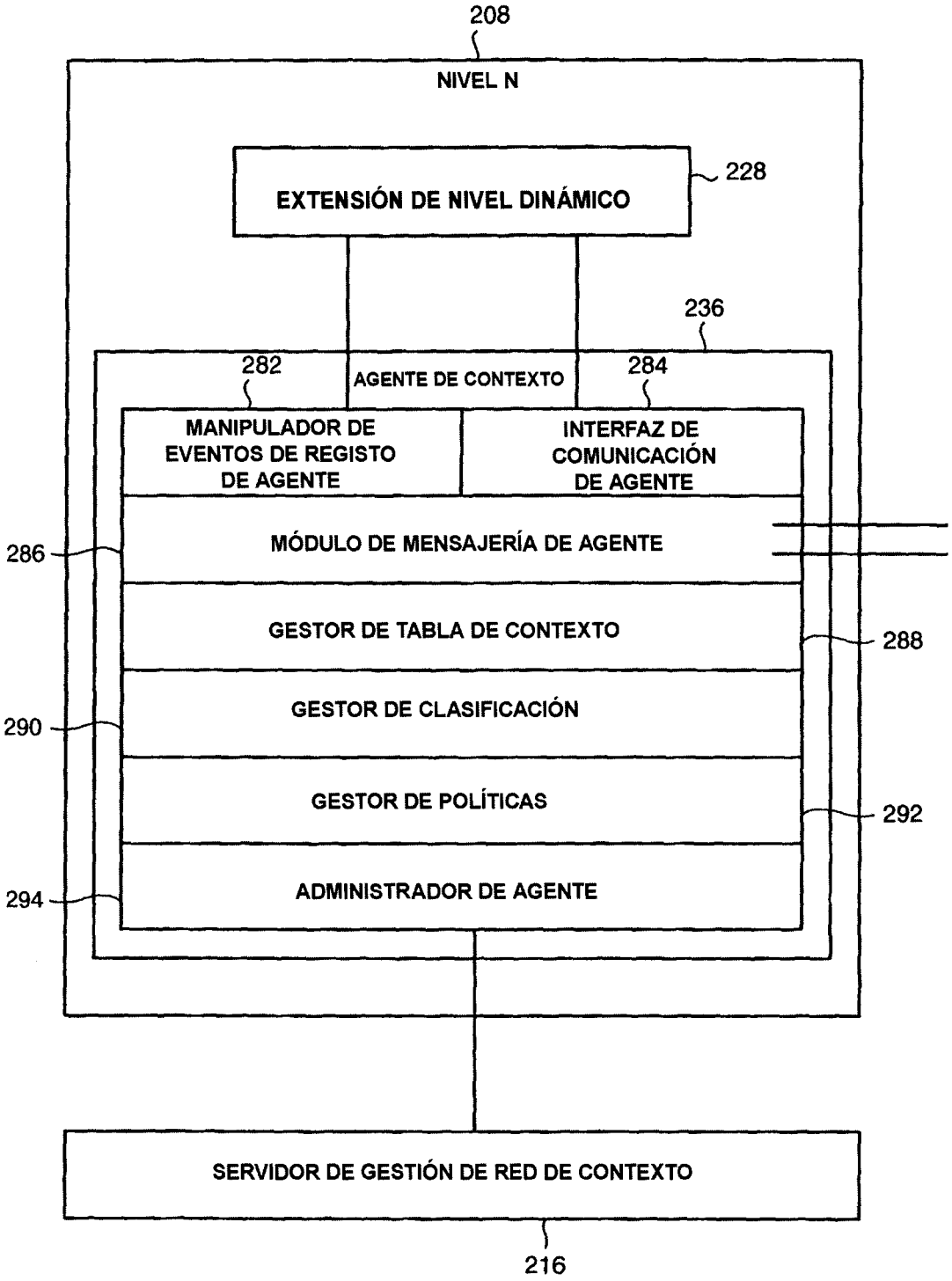
208
NIVEL N
EXTENSIÓN DE NIVEL DINÁMICO
228
236
282
AGENTE DE CONTEXTO
284
MANIPULADOR DE EVENTOS DE REGISTO DE AGENTE
INTERFAZ DE COMUNICACIÓN DE AGENTE
286
MÓDULO DE MENSAJERÍA DE AGENTE
GESTOR DE TABLA DE CONTEXTO
288
290
GESTOR DE CLASIFICACIÓN
GESTOR DE POLÍTICAS
292
294
ADMINISTRADOR DE AGENTE
SERVIDOR DE GESTIÓN DE RED DE CONTEXTO
216

FIG. 8

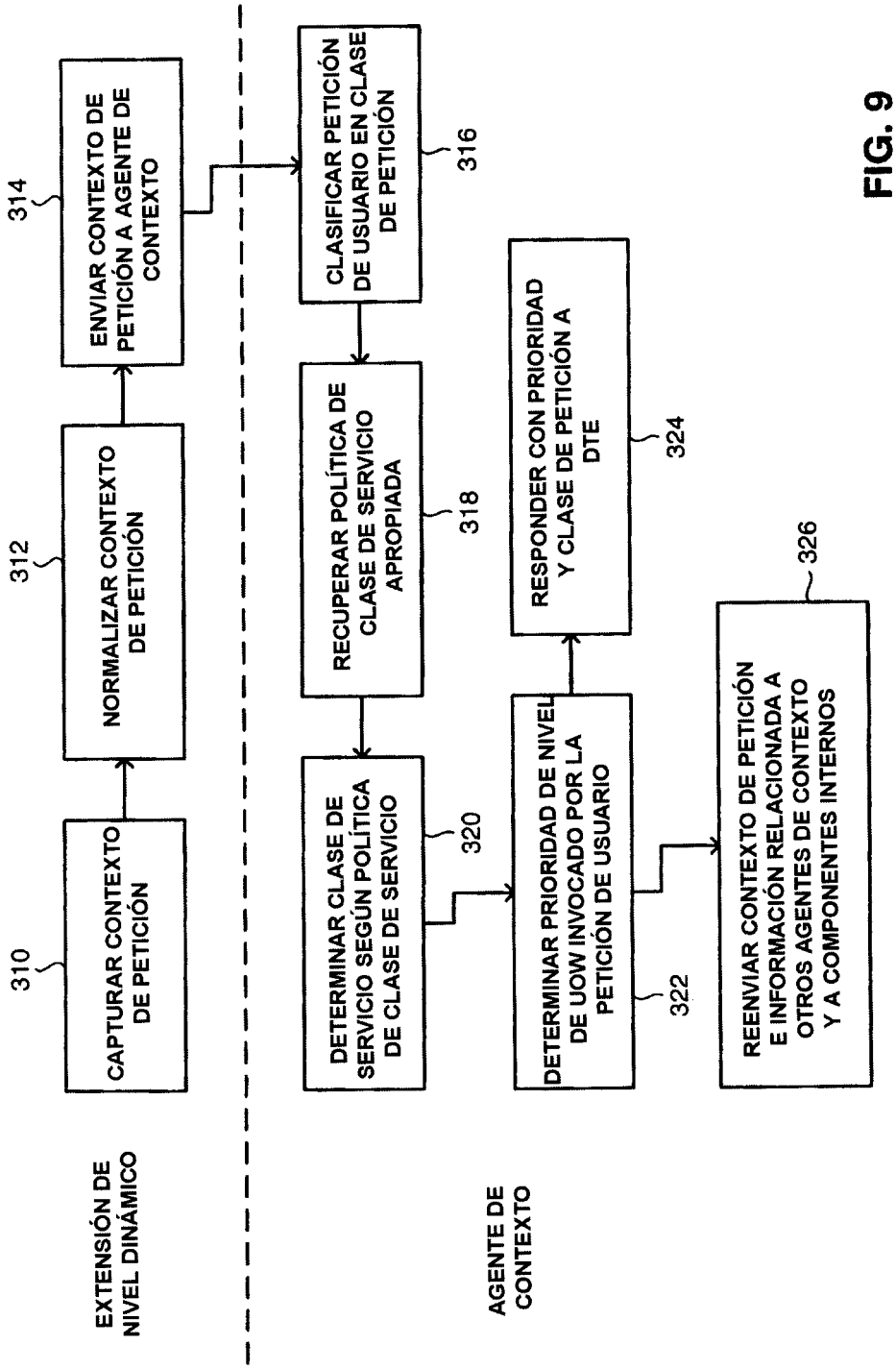
EXTENSIÓN DE NIVEL DINÁMICO
AGENTE DE CONTEXTO
310
CAPTURAR CONTEXTO DE PETICIÓN
312
NORMALIZAR CONTEXTO DE PETICIÓN
314
ENVIAR CONTEXTO DE PETICIÓN A AGENTE DE CONTEXTO
316
CLASIFICAR PETICIÓN DE USUARIO EN CLASE DE PETICIÓN
318
RECUPERAR POLÍTICA DE CLASE DE SERVICIO APROPIADA
320
DETERMINAR CLASE DE SERVICIO SEGÚN POLÍTICA DE CLASE DE SERVICIO
322
DETERMINAR PRIORIDAD DE NIVEL DE UOW INVOCADO POR LA PETICIÓN DE USUARIO
324
RESPONDER CON PRIORIDAD Y CLASE DE PETICIÓN A DTE
326
REENVIAR CONTEXTO DE PETICIÓN E INFORMACIÓN RELACIONADA A OTROS AGENTES DE CONTEXTO Y A COMPONENTES INTERNOS

FIG. 9

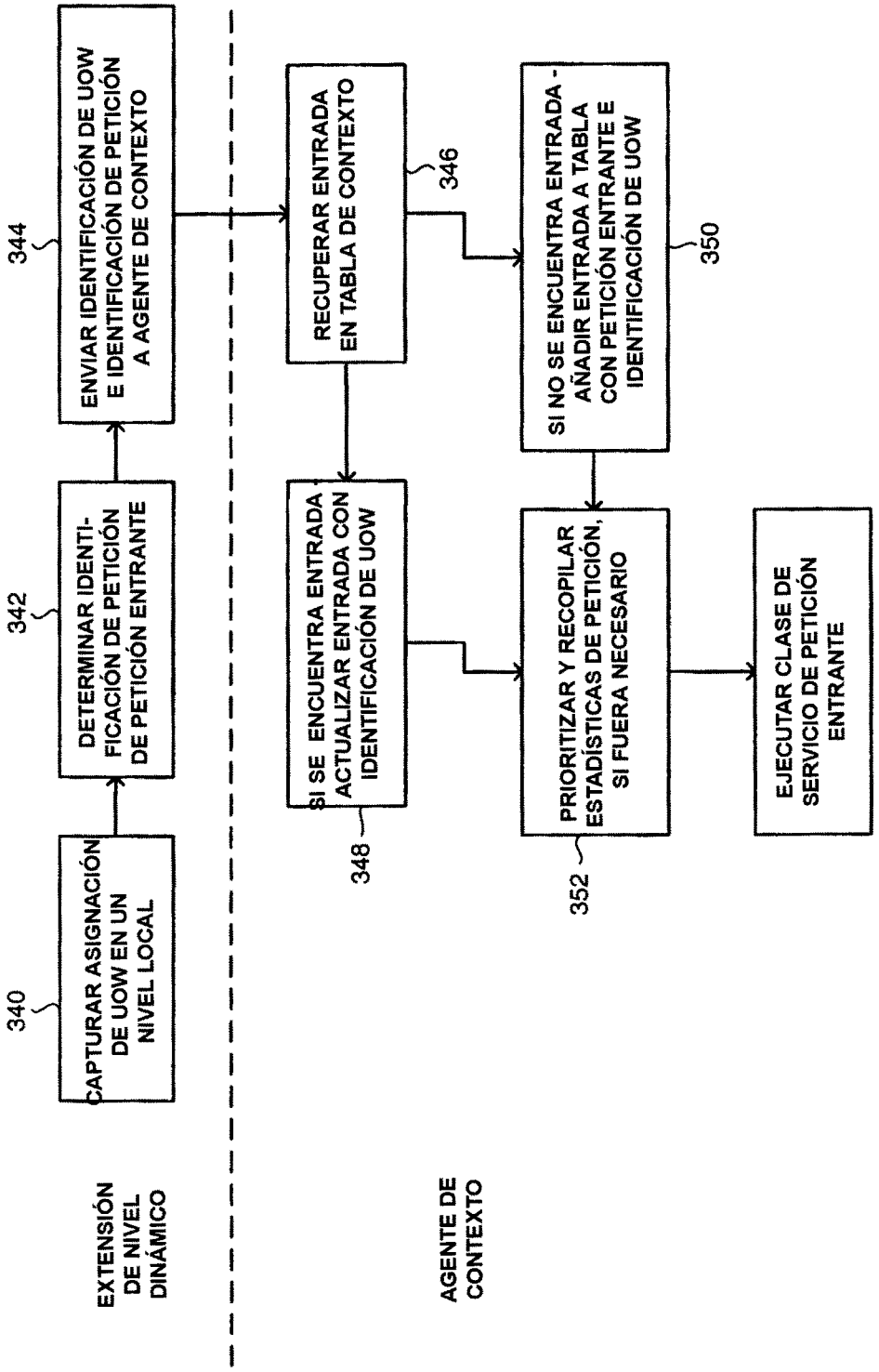
EXTENSIÓN DE NIVEL DINÁMICO
AGENTE DE CONTEXTO
340
CAPTURAR ASIGNACIÓN DE UOW EN UN NIVEL LOCAL
342
DETERMINAR IDENTIFICACIÓN DE PETICIÓN DE PETICIÓN ENTRANTE
344
ENVIAR IDENTIFICACIÓN DE UOW E IDENTIFICACIÓN DE PETICIÓN A AGENTE DE CONTEXTO
346
RECUPERAR ENTRADA EN TABLA DE CONTEXTO
348
SI SE ENCUENTRA ENTRADA ACTUALIZAR ENTRADA CON IDENTIFICACIÓN DE UOW
350
SI NO SE ENCUENTRA ENTRADA - AÑADIR ENTRADA A TABLA CON PETICIÓN ENTRANTE E IDENTIFICACIÓN DE UOW
352
PRIORITIZAR Y RECOPILAR ESTADÍSTICAS DE PETICIÓN, SI FUERA NECESARIO
EJECUTAR CLASE DE SERVICIO DE PETICIÓN ENTRANTE

FIG. 10

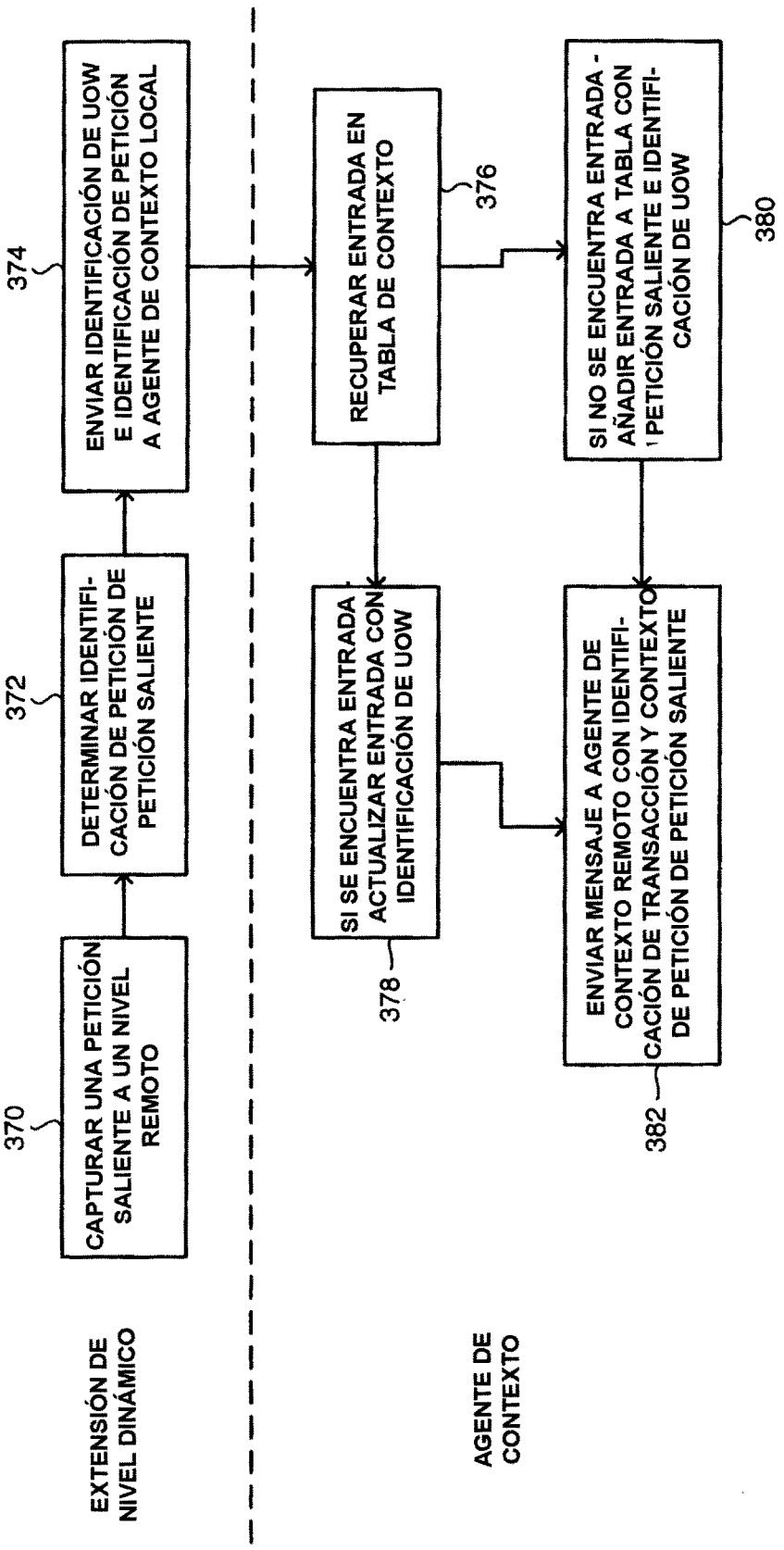
EXTENSIÓN DE NIVEL DINÁMICO
AGENTE DE CONTEXTO
370
CAPTURAR UNA PETICIÓN SALIENTE A UN NIVEL REMOTO
372
DETERMINAR IDENTIFICACIÓN DE PETICIÓN DE PETICIÓN SALIENTE
374
ENVIAR IDENTIFICACIÓN DE UOW E IDENTIFICACIÓN DE PETICIÓN A AGENTE DE CONTEXTO LOCAL
376
RECUPERAR ENTRADA EN TABLA DE CONTEXTO
378
SI SE ENCUENTRA ENTRADA ACTUALIZAR ENTRADA CON IDENTIFICACIÓN DE UOW
380
SI NO SE ENCUENTRA ENTRADA - AÑADIR ENTRADA A TABLA CON PETICIÓN SALIENTE E IDENTIFICACIÓN DE UOW
382
ENVIAR MENSAJE A AGENTE DE CONTEXTO REMOTO CON IDENTIFICACIÓN DE TRANSACCIÓN Y CONTEXTO DE PETICIÓN DE PETICIÓN SALIENTE

FIG. 11