

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 714 676**

51 Int. Cl.:

G06F 17/30 (2006.01)

G06Q 30/00 (2012.01)

G06F 12/08 (2006.01)

G06Q 10/00 (2012.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **09.08.2013 PCT/EP2013/002390**

87 Fecha y número de publicación internacional: **20.02.2014 WO14026753**

96 Fecha de presentación y número de la solicitud europea: **09.08.2013 E 13750848 (7)**

97 Fecha y número de publicación de la concesión europea: **12.12.2018 EP 2885725**

54 Título: **Actualización de resultados de consulta de base de datos almacenados en memoria caché**

30 Prioridad:

14.08.2012 EP 12368020
14.08.2012 US 201213585286

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
29.05.2019

73 Titular/es:

AMADEUS S.A.S. (100.0%)
485 Route du Pin Montard, Sophia Antipolis
06410 Biot, FR

72 Inventor/es:

CIABRINI, DAMIEN;
LEGRAND, GUILLAUME;
JANIN, BENOIT;
ISNARDY, LUC;
MAILLOT, NICOLAS;
ROBELIN, CHARLES-ANTOINE y
DANIELLO, RUDY

74 Agente/Representante:

SUGRAÑES MOLINÉ, Pedro

ES 2 714 676 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Actualización de resultados de consulta de base de datos almacenados en memoria caché

5 **Campo de la invención**

La presente invención se dirige al campo de la tecnología de bases de datos. Más en concreto, la misma se refiere al precálculo y el almacenamiento en memoria caché de los resultados de consulta de base de datos y a las estrategias de mantener actualizados estos resultados.

10

Antecedentes

Un problema común en la tecnología de bases de datos es asegurar unos tiempos de respuesta cortos a las consultas de base de datos que requieren el procesamiento de unos volúmenes de datos grandes. Por ejemplo, un procesamiento como este, que supone un consumo de potencia informática, se ha de llevar a cabo en respuesta a las así denominadas "consultas abiertas" que solo contienen una poca información de entrada (por ejemplo, solo se especifican uno o dos parámetros de entre una docena de parámetros posibles y/o los intervalos de valor especificados de los parámetros son amplios) y, en consecuencia, conducen en general a un gran número de resultados. Las posibilidades para acelerar el procesamiento de datos mediante el aumento del rendimiento de soporte físico son limitadas. Por lo tanto, se pone el foco en la mejora de los mecanismos que subyacen al procesamiento de unos volúmenes de datos grandes.

15

20

Un enfoque general para acortar los tiempos de consulta es precalcular las consultas esperadas y mantener los resultados de consulta correspondientes en un sistema de caché. Entonces, las consultas no se procesan en la práctica en las bases de datos grandes, sino que se dirigen al sistema de memoria caché.

25

No obstante, otro problema que surge junto con tales enfoques de almacenamiento en memoria caché, es mantener los resultados de búsqueda precalculados actualizados con el fin de asegurar que las consultas que son respondidas por los resultados almacenados en memoria caché reflejan correctamente el estado de las bases de datos grandes correspondientes. En el caso en el que cambian los datos subyacentes, los resultados de consulta almacenados en memoria caché quedan desactualizados y el sistema de memoria caché devolvería unos resultados incorrectos. Por lo tanto, se necesitan unas estrategias acerca de cómo se puede mantener actualizado el sistema de memoria caché.

30

En la técnica anterior se conocen diversas estrategias de actualización relativamente sencillas como, por ejemplo, recalcular la totalidad del dominio de datos con frecuencia, establecer y mantener unas programaciones de recálculo de forma manual y recalcular los datos cuando los mismos se estén volviendo demasiado antiguos.

35

Se han desarrollado estrategias de actualización algo más sofisticadas, tal como se describe, por ejemplo, por los documentos WO 01/33472 y WO 02/25557.

40

El documento WO 01/33472 se refiere a un sistema de disponibilidad que se usa en un sistema de planificación de viajes. El sistema incluye una memoria caché que tiene unas entradas de una información de disponibilidad con respecto a los asientos de línea aérea. Un gestor de memoria caché gestiona una información de entrada en la memoria caché con el fin de mantener la información en la memoria caché correcta, actual, completa o, por lo demás, tan útil como sea posible. En respuesta a una consulta que está dirigida a la memoria caché, el gestor de memoria caché determina si una respuesta almacenada es obsoleta y, si este es el caso, envía una consulta de disponibilidad a una fuente de la información de disponibilidad. Las entradas de memoria caché que se van a modificar se obtienen por medio de unas notificaciones asíncronas a partir de sistemas externos y se determinan por medio de un modelo determinista, predictivo o estadístico.

45

50

De forma similar, el documento WO 02/25557 se refiere a un sistema de recuperación de información en el que una información que se recibe a partir de unas fuentes de información se almacena en memoria caché para su uso futuro, tal como para unas solicitudes de cliente futuras. Se pueden generar unas consultas proactivas para rellenar una memoria caché y / o para actualizar la información actualmente almacenada en memoria caché. En un sistema de información de línea aérea, las consultas proactivas se ordenan en función de unas estadísticas o indicaciones predictivas tales como una cercanía de la hora de salida, la antigüedad de los datos almacenados en memoria caché, los asientos restantes en una aeronave, las vacaciones o eventos especiales o el tipo de equipo. Además, se reciben actualizaciones por medio de unas notificaciones externas procedentes de las líneas aéreas, tales como mensajes de AVS.

55

60

Además, el documento WO 99/22315 describe un mecanismo para refrescar de forma automática los documentos en una memoria caché mediante el uso de un modelo probabilístico basado en estadísticas. Para cada documento, la memoria caché determina una probabilidad $P_{\psi}(t)$ de que un objeto almacenado en memoria caché i esté obsoleto en un tiempo particular t (es decir, el servidor ha cambiado ese objeto) y una probabilidad $P_{\psi}(h)$ de que ese objeto i sea solicitado por un usuario en el tiempo de solicitud h . La memoria caché refresca esos objetos con el producto

65

más alto $P_i = \Psi(t) \times P_r(h)$, es decir la probabilidad de que se devuelva un objeto desactualizado al usuario con la siguiente solicitud. Para mantener estos valores de probabilidad, la memoria caché mantiene y realiza un seguimiento de unas estadísticas históricas para los objetos almacenados en memoria caché tales como un intervalo medio estimado entre las actualizaciones de servidor EUI. El EUI de un objeto se actualiza, por ejemplo, cuando el propio objeto es actualizado por el servidor o el objeto no se actualiza después de que haya transcurrido su tiempo de refresco medio estimado.

El documento US 6.725.333 se refiere a la gestión de entidades, tales como un objeto, un archivo de imagen, o una página web, que están almacenadas en una memoria caché, y / o entidades que se pueden almacenar en una memoria caché. Se analiza el código de programa para determinar si hay al menos una instrucción que afecte a la conveniencia de realizar al menos una transacción de memoria caché y, si se desea, se lleva a cabo la al menos una transacción de memoria caché. Además, una herramienta de análisis de programas analiza por medios estadísticos el código de programa para ubicar unos puntos en los que tienen lugar cambios de estado de objetos, en los que se crean objetos y en los que se suprimen objetos, y entonces genera unas dependencias regularizadas en tales puntos para, y empleando, las dependencias para anular las consultas almacenadas en memoria caché dependientes. Más en concreto, el programa se analiza para identificar una o más instrucciones que puedan modificar un valor de una o más entidades almacenables en memoria caché durante el tiempo de ejecución del programa. Para cada una de las instrucciones detectadas, se determina una probabilidad que representa la verosimilitud de que se ejecuten las instrucciones detectadas, es decir, la verosimilitud de que una o más entidades almacenables en memoria caché cambien debido a la ejecución de las instrucciones. Para determinar si se llevará a cabo una transacción de memoria caché, por ejemplo, insertando un objeto en una memoria caché o suprimiendo o actualizando un objeto almacenado en memoria caché, se realiza una determinación en lo que respecta a si la probabilidad de cambio de una o más entidades almacenables en memoria caché satisface un umbral previamente definido. Si esto es así, el sistema se puede encontrar a favor de no almacenar en memoria caché una o más entidades no almacenadas en memorias caché y / o encontrarse a favor de anular o actualizar una o más entidades almacenadas en memorias caché. Si esto no es así, el sistema se puede encontrar a favor de almacenar en memoria caché una o más entidades no almacenadas en memorias caché y / o no encontrarse a favor de anular o actualizar una o más entidades almacenadas en memorias caché.

Sumario de la invención

De acuerdo con la presente invención, se proporciona un método de actualización de unos resultados de consulta de base de datos precalculados en un sistema de base de datos distribuido tal como se define en la reivindicación independiente 1.

De acuerdo con otro aspecto, se proporciona una plataforma de memoria caché de base de datos tal como se define en la reivindicación independiente 13.

De acuerdo con otro aspecto, se proporciona un medio de almacenamiento legible por ordenador no transitorio tal como se define en la reivindicación independiente 15.

En las reivindicaciones dependientes se exponen algunos aspectos adicionales.

Breve descripción de las figuras

La presente invención se describirá con referencia a las figuras adjuntas. En general, unos números de referencia similares indican unos elementos idénticos o funcionalmente similares.

La figura 1 ilustra una visión de conjunto general del sistema de base de datos distribuido;

la figura 2 muestra una vista más detallada del sistema de base de datos distribuido de acuerdo con una forma de realización;

la figura 3 ilustra los componentes de la plataforma de memoria caché de acuerdo con una forma de realización;

la figura 4a visualiza un efecto de volatilidad a lo largo del tiempo de acuerdo con un modelo probabilístico;

la figura 4b visualiza una rectificación del modelo probabilístico en respuesta a un evento en tiempo real;

la figura 5 ilustra un diagrama de flujo de acuerdo con una forma de realización del método;

la figura 6 muestra una disponibilidad de recursos a modo de ejemplo para el recálculo de acuerdo con una forma de realización;

la figura 7 muestra una representación esquemática de un ordenador de plataforma de memoria caché de acuerdo con una forma de realización.

Descripción general

Antes de pasar a la descripción detallada en función de las figuras, unos pocos aspectos más generales se expondrán en primer lugar con respecto a la figura 1.

Con el fin de ser capaz de manejar consultas de base de datos o solicitudes de cálculos por lotes que requieren cálculos en función de unos volúmenes de los datos subyacentes, en general los resultados de consulta de base de datos que se corresponden con las consultas esperadas se precálculan y se almacenan en memoria caché (posteriormente, la expresión “consulta” se usa como una expresión general que incluye cualquier tipo de solicitud de recuperación de información tal como consultas transaccionales, solicitudes de cálculos por lotes y otras formas). Los resultados almacenados en memorias caché se almacenan y se devuelven a la entidad que realiza la consulta en respuesta a las consultas que tienen lugar en la práctica. La figura 1 ilustra un sistema de base de datos 1 de ese tipo a un nivel abstracto. Los datos básicos se mantienen en una plataforma de cálculo 3 que está conectada con una plataforma de memoria caché 2. Esta última emite unas órdenes recalculadas a la plataforma de cálculo 3 que, a su vez, transmite los resultados correspondientes de vuelta a la plataforma de memoria caché 2 en donde se almacenan los resultados de consulta precálculados.

El presente enfoque de almacenamiento en memoria caché de unos resultados de consulta precálculados conduce al problema general de que los datos del dominio de datos subyacente pueden cambiar a lo largo del tiempo y, por lo tanto, los resultados de consulta precálculados almacenados en memoria caché quedan obsoletos. Los resultados de consulta almacenados en memoria caché que siguen estando actualizados, es decir, que coinciden con los equivalentes de cálculo en tiempo real correspondientes (unos resultados que se calcularían en la práctica a petición sin haber almacenado en memoria caché los resultados precálculados que se encuentran disponibles), se denominan resultados almacenados en memoria caché “precisos” en lo sucesivo en el presente documento. Por lo tanto, cuando la memoria caché representa correctamente el estado actual del dominio de datos que subyace a los resultados de consulta almacenados en memoria caché, la memoria caché es - en general - precisa.

En general, con el fin de devolver unos resultados correctos en función de la memoria caché, se desea mantener un alto grado de correlación entre los resultados de consulta de base de datos almacenados en memoria caché que se proporcionan a la entidad que realiza la consulta en respuesta a las consultas de base de datos y sus equivalentes de cálculo en tiempo real. Al mismo tiempo, no obstante, es deseable reducir al mínimo el consumo de recursos de cálculo que es causado por los recálculos, es decir, evitar todo recálculo innecesario tal como el recálculo de unos resultados de consulta almacenados en memoria caché aún precisos. Los recursos de cálculo son limitados y, en general, no hay suficientes recursos de cálculo para recalcular todos los resultados de consulta almacenados en memoria caché en todo momento. Por lo tanto, es necesario hallar una compensación recíproca entre la precisión de la memoria caché y la utilización de la potencia de cálculo disponible.

Los enfoques sencillos de mantener actualizada una memoria caché de resultados de consulta precálculados presentan varios inconvenientes.

El recálculo de la totalidad del dominio de datos con frecuencia, dependiendo de la cantidad de datos y los recursos de cálculo disponibles, por ejemplo, una vez al día, podría asegurar un equilibrio razonable entre la precisión de la memoria caché y las respuestas en tiempo real. No obstante, el presente enfoque es tanto a duras penas ajustable a escala como ineficiente en términos del consumo de recursos de soporte físico. En particular, también se recalculan aquellos resultados de consulta que siguen siendo válidos debido a que los datos subyacentes correspondientes no han cambiado.

La confección de programaciones de recálculo para determinar qué resultados de consulta se van a recalcular en qué instante de forma manual, por parte de un administrador humano, se puede mostrar eficiente para fines específicos, pero la misma es rígida y nada flexible. Es necesario volver a confeccionar la programación de nuevo cuando las suposiciones y condiciones que subyacen al cambio de programación. Asimismo, esto tampoco puede realizar, de forma dinámica, un seguimiento de un deterioro repentino de la calidad de la memoria caché que podría surgir en el evento de cambios masivos en el dominio de datos subyacente. También es difícil diseñar una programación de ese tipo de forma manual, por ejemplo, debido a la ausencia de unos criterios de calidad objetivos, y mantener la misma en términos de costes de mano de obra.

Aún otro enfoque más es recalcular los datos cuando los mismos se están volviendo demasiado antiguos. Dependiendo de la naturaleza de los datos subyacentes y los resultados de consulta que se van a precalcular, puede ser difícil, no obstante, evaluar unos umbrales buenos de “antigüedad”.

Con el fin de volver el recálculo más eficiente, se debería definir una métrica para evaluar cómo de “innecesario” es un recálculo. Por ejemplo, no merece la pena relanzar la totalidad de un precálculo masivo cada día si sucede que menos de la mitad de los resultados de consulta calculados están obsoletos. Por otro lado, si se sabe que unas clases particulares de resultados de consulta cambian con frecuencia, el recálculo de las mismas varias veces al día podría ser beneficioso para la precisión. En consecuencia, es necesaria una forma eficaz de evaluar o estimar la precisión de los resultados de consulta, teniendo en cuenta tanto la ganancia asociada sobre la precisión como el

coste de recálculo. De acuerdo con la estrategia de actualización de memoria caché que se presenta en el presente documento, los recálculos de los resultados de consulta de base de datos se deciden basándose en las probabilidades de que las consultas de base de datos almacenadas en memoria caché estén obsoletas, es decir, difieren potencialmente de los resultados que se obtienen por medio de otro recálculo. Solo aquellos resultados de consulta almacenados en memoria caché que tienen al menos una cierta probabilidad previamente determinada de inexactitud se recalculan, mientras que otros resultados de consulta almacenados en memoria caché que, probablemente, sigue reflejando con precisión los datos subyacentes, es decir, los mismos tienen una probabilidad menor de estar obsoletos, no se recalculan.

La estrategia de actualización de la memoria caché que se presenta en el presente documento se basa, como un primer aspecto, en unos medios para estimar la precisión de la totalidad de la memoria caché de los resultados de consulta de base de datos precalculados en función de un modelo predictivo. Como un segundo aspecto, también se comprueba si esas estimaciones se encuentran, en general, en consonancia con la realidad, verificando que la estrategia de recálculo basado en modelos sigue siendo válida cuando tienen lugar eventos reales en tiempo real (y de la vida real) que pueden servir como indicaciones de que, por ejemplo, se ha cambiado una parte significativa de los datos que subyacen a las consultas almacenadas en memoria caché y - debido a estos cambios - también las consultas almacenadas en memoria caché correspondientes están obsoletas.

El modelo predictivo, en el que se basan en general las estimaciones de la precisión de la memoria caché, modela las discrepancias entre los resultados de consulta almacenados en memoria caché y los resultados de consulta de base de datos reales supuestos, es decir, el mismo aproxima la precisión o inexactitud de cualquier resultado de consulta almacenado en memoria caché. El modelo modela, por ejemplo, la volatilidad probable de los resultados almacenados en memorias caché a lo largo del tiempo. Las suposiciones acerca de la volatilidad de los resultados almacenados en memorias caché se concluyen y se extrapolan a partir de las (pasadas) experiencias en el mundo real sobre la materia objeto del dominio de datos respectivo. Por lo tanto, el modelo predictivo es, en general, una representación de la realidad en términos de la volatilidad y / o la precisión de los resultados de consulta de base de datos precalculados.

Se pueden emplear diversos tipos de modelos. Por ejemplo, el modelo predictivo puede ser un modelo conceptual que se ha generado por medio de un proceso de modelado estructurado que se conoce como modelado científico. El modelado requiere identificar y seleccionar aquellos aspectos y eventos del entorno en el mundo real (en el presente caso, por ejemplo: los aspectos y eventos que tienen influencia sobre la volatilidad y la disminución de precisión a lo largo del tiempo de los resultados de consulta precalculados y almacenados en memoria caché) y dejando a un lado otros aspectos que pueden no ser particularmente relevantes para el fin del modelo. Además, los aspectos identificados y seleccionados se pueden abstraer y formular / poner en práctica en una forma procesable tal como un programa informático y / o unas ecuaciones matemáticas. Un modelo conceptual de ese tipo también puede servir como un medio para una simulación que simula el comportamiento (es decir, la volatilidad / precisión) de los resultados de consulta de base de datos precalculados a lo largo del tiempo. Como alternativa, el modelo predictivo empleado puede ser un modelo histórico - estadístico que se basa en supervisar y evaluar, por ejemplo, la volatilidad / precisión de los resultados de consulta precalculados en el periodo dado de tiempo en el pasado (por ejemplo, tres meses) y extrapola las tendencias determinadas al futuro. Un ejemplo particular de un modelo histórico - estadístico de ese tipo se da adicionalmente en lo sucesivo.

Por ejemplo, los datos subyacentes se pueden ubicar en el dominio de los viajes por aire y pueden contener una información acerca de vuelos tal como el aeropuerto de salida y de destino, la línea aérea, las fechas de salida y de vuelta, las tarifas, las clases de reserva y similares. Estos datos relacionados con viajes por aire se mantienen en la plataforma de cálculo y son consultados por los clientes con el fin de obtener un conocimiento de la disponibilidad y los precios de los vuelos aéreos. El cálculo de los precios en función de los datos de vuelo básicos consume una gran cantidad de recursos y de tiempo. Por lo tanto, los precios reales se precálculan y se almacenan en memoria caché en la presente plataforma de memoria caché. En el presente ejemplo, el modelo probabilístico modela la volatilidad de los precios de vuelo a lo largo del tiempo.

El conocimiento requerido para construir un modelo como este se puede tomar a partir de las experiencias en el mundo real sobre el comportamiento y la evolución de los precios de vuelo antes de la fecha de salida. Por ejemplo, se podría saber que los precios de vuelo permanecen relativamente estables a lo largo del periodo de tiempo antes de un mes antes de las fechas de salida respectivas, pero se vuelven más volátiles durante el mes antes de la fecha de salida. Por lo tanto, el modelo probabilístico indica que los precios precalculados y almacenados en memoria caché que pertenecen a los vuelos venideros en el mes siguiente se deberían recalculan más a menudo que tales precios precalculados que están asociados con los vuelos en un futuro más distante.

Además de modelar la precisión de la memoria caché usando el modelo probabilístico, se evita una caída acusada de la precisión de la memoria caché al presentar reactividad frente a los eventos en tiempo real. Las decisiones de recálculo se refinan a la recepción de unos eventos en tiempo real previamente determinados que podrían tener un impacto sobre la exactitud de los resultados de consulta almacenados en memoria caché. Los eventos en tiempo real son asíncronos, es decir, el instante en el tiempo de su aparición no está predeterminado - estos pueden tener lugar en cualquier momento. Con el fin de ser capaz de recibir y procesar los eventos en tiempo real entrantes, la

plataforma de datos de memoria caché 2 se equipa con unas interfaces externas a unas fuentes de comunicación que notifican en consecuencia la información relevante a la plataforma de datos de memoria caché.

5 Tales eventos en tiempo real se pueden referir a situaciones particulares que no se consideran en el modelo predictivo. Por ejemplo, una porción de los precios almacenados en memoria caché se puede ver afectada por una promoción, mientras que otros precios de pueden hacer más volátiles en un instante particular del año (tal como temporadas de vacaciones, Navidad, etc.). También podrían cambiar las suposiciones que subyacen a las causalidades “normales” del modelo unas situaciones “excepcionales” como una exposición industrial, un evento deportivo o similares, eventos aleatorios tales como golpes o desastres naturales. Se pueden considerar estas influencias particulares cuando se determinan las probabilidades de que los resultados de consulta almacenados en memoria caché estén obsoletos en respuesta a los eventos en tiempo real respectivos que representan tales situaciones excepcionales. Como alternativa, el impacto de los eventos programados tales como una exposición industrial, temporadas de vacaciones, eventos deportivos y similares se puede introducir en el modelo probabilístico en un buen momento antes de la fecha del evento.

15 Es importante observar que la estrategia de actualización que se presenta en el presente documento es capaz de tener en cuenta los eventos “incierto”, es decir, aquellos eventos que no anulan uno o más resultados de consulta almacenados en memoria caché precalculados con certeza, sino que solo indican que se podría aumentar la probabilidad de que estén obsoletos los resultados de consulta de base de datos almacenados en memoria caché. En otras palabras, estos eventos son indeterministas con respecto a la precisión de los resultados de consulta almacenados en memoria caché y solo tienen una influencia probabilística sobre las discrepancias entre los resultados de consulta almacenados en memoria caché que se mantienen en la plataforma de memoria caché 2 y los resultados de consulta de base de datos reales supuestos que resultan de un recálculo hipotético. Esto es diferente de las propuestas que se describen en los documentos WO 01/33472 y WO 02/25557 en donde los mensajes de AVS, por ejemplo, indican que se ha cancelado un vuelo particular. En consecuencia, a la recepción de un mensaje de AVS de ese tipo, se conoce con certeza que los asientos de avión respectivos han dejado de encontrarse disponibles.

30 Por ejemplo, haciendo referencia al escenario de almacenamiento de datos relacionados con viajes tal como se ha mencionado en lo que antecede, un evento en tiempo real que tiene solo un impacto potencial sobre la precisión de los resultados de consulta almacenados en memoria caché podría ser una actualización de tarifas. Una tarifa es un conjunto de datos que incluye parámetros como la ciudad de salida y de destino, la clase de reserva, el tipo de vuelo (de ida o de ida y vuelta), una cantidad de dinero y reglas que definen las restricciones que se han de satisfacer para que la tarifa sea de aplicación en la práctica. Por lo tanto, las tarifas representan los datos básicos para el cálculo de precios de un vuelo particular. Si las tarifas para un par de ciudades de origen - destino específico son actualizadas por la línea aérea, se puede aumentar la verosimilitud de que un precio de vuelo precalculado y almacenado en memoria caché con respecto a este par de ciudades haya dejado de ser correcto. No obstante, desde la perspectiva de la plataforma de memoria caché de datos 2, esto no es seguro debido a que la tarifa que fue aplicada en la práctica por la plataforma de precálculo 3 cuando se precalcula el precio almacenado en memoria caché no es conocida por la plataforma de memoria caché de datos 2. Por ejemplo, la tarifa que se aplica para el precálculo previo podría no haberse cambiado en la práctica y los cambios de tarifas que son indicados por el evento de cambio de tarifas no cambian el hecho de que la tarifa previamente pertinente sigue siendo de aplicación y, por lo tanto, el precio que se ha calculado en lo que antecede sigue siendo válido. O, la tarifa previamente aplicada se cambia en la práctica, pero - debido al cambio - otra tarifa es de aplicación ahora para el cálculo del precio de vuelo en cuestión que, al final, tiene el efecto de que el precio almacenado en memoria caché sigue siendo válido en la práctica.

50 Por lo tanto, al observar tales eventos en tiempo real, la plataforma de memoria caché de datos 2 solo puede conjeturar con una verosimilitud indeterminista que determinados resultados de consulta almacenados en memoria caché están ahora obsoletos y sería ventajoso recalcular estos con el fin de mantener precisa la memoria caché. No obstante, esto no es un hecho cierto y bien puede ser que los resultados de consulta almacenados en memoria caché respectivos - a pesar de que ha aumentado su probabilidad de estar obsoletos - sigan siendo precisos en la práctica.

55 La determinación de las probabilidades de que estén obsoletos los resultados de consulta de base de datos almacenados en memoria caché se lleva a cabo en dos etapas lógicas: En general, en el primer nivel lógico, las probabilidades se identifican mediante el uso del modelo predictivo probabilístico. Posteriormente, en el segundo nivel lógico, estas probabilidades determinadas se pueden rectificar en respuesta a los eventos en tiempo real entrantes.

60 Basándose en las probabilidades que se determinan de esta forma, la plataforma de memoria caché de datos 2 genera y emite de forma automática unas órdenes recalculadas a la plataforma de recálculo 3 por medio de una interfaz de red apropiada entre las dos entidades (consúltese la figura 1). En general, se generan unas órdenes recalculadas con respecto a aquellos resultados de consulta almacenados en memoria caché que tienen una probabilidad más alta de estar obsoletos que la de otros resultados de consulta almacenados en memoria caché que tienen una probabilidad menor de estar obsoletos. Esta regla práctica general se puede poner en práctica mediante

el uso de valores umbral de las probabilidades: Los resultados de consulta almacenados en memoria caché con una probabilidad determinada de estar obsoletos por encima de un valor umbral de ese tipo, necesitan ser recalculados. Por consiguiente, se emiten unas órdenes recalculadas respectivas. Los resultados de consulta almacenados en memoria caché con una probabilidad determinada de estar obsoletos a o por debajo de un valor umbral de ese tipo, se consideran como que es verosímil que sigan siendo precisos y, en consecuencia, no es necesario que se recalculen los mismos. Por consiguiente, no se emite orden recalculada alguna con respecto a estos resultados de consulta almacenados en memoria caché.

Las capacidades de cálculo disponibles en instantes particulares son tenidas en cuenta por la plataforma de memoria caché de datos 2 antes de emitir las órdenes recalculadas. Con el fin de ser capaz de considerar los recursos disponibles, es necesario que la plataforma de memoria caché de datos 2 tenga un cierto conocimiento acerca del grado y / o la programación de la utilización de capacidad de la plataforma de cálculo 3 y los recursos de cálculo libres, de forma respectiva. La información relevante se rellena por medio del enlace de comunicación entre las dos plataformas.

En respuesta a la recepción de una orden recalculada, la plataforma de recálculo 3 recalcula los resultados de consulta respectivos y los devuelve a la plataforma de memoria caché de datos 2 en donde los mismos se almacenan y la supervisión y la determinación de las probabilidades reaparece.

Es preferible considerar la relación entre el modelo probabilístico y los eventos en tiempo real que tienen lugar antes de decidir si se debería rectificar o anular, o no, una decisión de recálculo en respuesta a un evento en tiempo real particular. Básicamente, los eventos en tiempo real se deberían analizar para determinar si, y en qué medida, se encuentran ya presentes los mismos en el modelo probabilístico.

Esta determinación de si, en el modelo probabilístico, se encuentra presente, o no, y en qué medida, un evento depende, en general, del tipo del modelo que se usa. Si, por ejemplo, el modelo es un modelo conceptual que es el resultado de un proceso de generación de modelos científicos (tal como se ha bosquejado brevemente en lo que antecede), el experto que ha generado este modelo está al tanto de qué aspectos, eventos y suposiciones se han considerado como relevantes para y se han incorporado en el modelo y qué otros aspectos, eventos y suposiciones se han considerado como de menor relevancia y, por lo tanto, no se han incluido en el modelo. En este caso, en general se conoce, por lo tanto, qué eventos en tiempo real asíncronos están representados en el modelo y cuáles no lo están.

Si, por otro lado, el modelo es uno de tipo histórico - estadístico que, principalmente, se basa en evoluciones y tendencias pasadas de la volatilidad, la precisión y / o la popularidad de los resultados de consulta de base de datos recalculados, puede que no se sepa, desde el inicio, qué eventos han conducido a este comportamiento particular de la volatilidad / precisión / popularidad de los resultados de consulta recalculados en el pasado. No obstante, en este caso también es posible realizar una aproximación de si se incluye, o no, uno asíncrono particular en el modelo probabilístico. Un enfoque es, por ejemplo, determinar la precisión real de los eventos de consulta de base de datos recalculados (por ejemplo, mediante el recálculo de una muestra representativa de los mismos y la comparación de los resultados de recálculo con los resultados de consulta almacenados en memoria caché) y comparar la misma con la precisión predicha que es indicada por el modelo probabilístico. Si la precisión (de muestra) real es sustancialmente menor que la precisión predicha, por ejemplo, por debajo de un umbral dado, se puede suponer que el modelo probabilístico no está al tanto del evento en tiempo real asíncrono particular, es decir, el evento no está representado en el modelo (de lo contrario, el modelo también habría indicado una precisión menor). Si, por otro lado, la precisión (de muestra) real se encuentra en la región de (es decir, solo apenas por debajo de, a o incluso superior a) la precisión que se predice por medio del modelo probabilístico, se puede suponer que el evento en tiempo real asíncrono se ha incluido en el modelo probabilístico (o este solo es un evento despreciable que solo tiene una influencia limitada sobre la volatilidad o la precisión de los resultados de consulta de base de datos recalculados). Un ejemplo más detallado del presente enfoque se da adicionalmente en lo sucesivo.

Para tales eventos que están suficientemente representados en el modelo, no es necesaria rectificación alguna de las probabilidades debido a que su aparición ya ha sido tenida en cuenta cuando se determinan las probabilidades de los resultados de consulta de base de datos almacenados en memoria caché respectivos en función del modelo probabilístico. Si, por otro lado, se halla que un evento en tiempo real no está representado en el modelo probabilístico, este se tiene en cuenta inmediatamente y las probabilidades se rectifican (es decir, se aumentan) y, potencialmente, debido a las probabilidades rectificadas, se emiten unas órdenes recalculadas con respecto a los resultados de consulta de base de datos almacenados en memoria caché respectivos.

De forma opcional, los eventos en tiempo real que tienen lugar que se encuentran presentes en el modelo probabilístico en cierta medida se acumulan con el fin de evaluar las tendencias. Si una acumulación de los eventos en tiempo real que aparecen en la práctica que se modela, en general, mediante el modelo probabilístico indica una ráfaga, la medida de la cual está más allá de lo que es considerado por las probabilidades del modelo se rectifica y, si es aplicable, las órdenes recalculadas se anulan en consecuencia.

De forma opcional, los eventos también se acumulan y se analizan en grupos con el fin de eliminar por filtrado los eventos que podrían desactualizar demasiado pocos resultados de consulta almacenados en memoria caché y / o que se podrían considerar irrelevantes. También por esta razón, los eventos se registran, se recopilan a lo largo del tiempo y se manejan de una forma agregada. De esta forma, se evita la generación de demasiadas órdenes recalculadas en respuesta a los eventos de bajo impacto y, por lo tanto, se evita un aumento desproporcionado en los costes de los recursos de cálculo.

En resumen, tener en cuenta unos eventos en tiempo real que influyen potencialmente sobre la precisión de los resultados de consulta de base de datos almacenados en memoria caché al menos por encima de una medida previamente determinada proporciona una reactividad alta frente al deterioro de la memoria caché.

La presente estrategia de actualización de memoria caché se puede usar, por ejemplo, junto con la Plataforma de cálculo masivo (MCP, *Massive Computation Platform*) de Amadeus que es el objeto del documento EP 2521074 A1. Al emplear la presente plataforma de memoria caché de datos de forma que coexista con la MCP, se encuentra disponible un subsistema potenciado para desencadenar un recálculo de MCP. Los resultados de consulta de base de datos tales como las recomendaciones de viajes que son generadas por la MCP se duplican y se almacenan en la plataforma de memoria caché de datos para su análisis adicional. Las decisiones de recálculo se realizan en función de un modelo probabilístico que se puede componer, él mismo, en función de unos datos de estadísticas que se toman de otros servicios de Amadeus. Además, se tienen en cuenta eventos en tiempo real tales como cambios de tarifas de vuelo, cambios de disponibilidad de asientos de avión, anulaciones de clase de reserva, solicitudes de billetes de vuelo de clientes, eventos de realimentación de calidad de usuario, cancelaciones de vuelos y / o similares.

Una la aplicación a modo de ejemplo para el presente enfoque de actualización de memoria caché es la precompra. Antes de reservar en la práctica un viaje, los usuarios finales de la industria del turismo normalmente desean informarse a sí mismos acerca de los vuelos disponibles, incluyendo los precios de vuelo actuales, sin compromiso alguno de reservar en la práctica el vuelo. Muy a menudo, tales solicitudes no vinculantes de información de prerreserva adoptan la forma de unas consultas de base de datos abiertas y amplias que requerían una cantidad enorme de recursos de cálculo cuando se calculan solo en el momento de la consulta. Además, los clientes esperan que la información solicitada se entregue de forma virtualmente instantánea en respuesta a sus consultas. Por lo tanto, los resultados de consulta de precompra tales como las recomendaciones de viajes por aire con precio habitualmente se precalcular y se almacenan en memoria caché. En consecuencia, la precompra en la industria del turismo forma una aplicación conveniente para la estrategia de actualización de memoria caché que se propone en el presente documento.

Descripción detallada

Pasando a continuación a la descripción más detallada, la figura 2 muestra una visión de conjunto del sistema de base de datos distribuido 1 de acuerdo con una forma de realización a modo de ejemplo. Las formas de realización que se describen posteriormente se refieren a las bases de datos en la industria del turismo. En concreto, se presenta una forma de realización en la que la plataforma de cálculo 3 mantiene unos datos acerca de las ofertas de viajes por aire y la plataforma de datos de memoria caché 2 almacena precios en relación con estas ofertas de viajes por aire que la plataforma de cálculo 3 calcula en función de unas reglas de cálculo, en particular las tarifas de vuelo y sus reglas de cálculo asociadas. No obstante, se debería hacer notar que estas formas de realización son ejemplos solo para el fin de ilustrar la presente estrategia de actualización de memoria caché con mayor detalle. La estrategia de actualización de memoria caché que se presenta en el presente documento se puede aplicar a cualquier tipo de datos y resultados de consulta de base de datos independientes de la estructura y / o la semántica de los datos y los resultados almacenados en memorias caché.

Tal como se ha descrito en lo que antecede, las entidades principales del sistema de base de datos distribuido 1 son la plataforma de memoria caché de datos 2 (a la que se hace referencia brevemente en lo sucesivo en el presente documento como DCP) y la plataforma de cálculo 3. En el ejemplo de la figura 2, la plataforma de cálculo 3 es una plataforma de cálculo masivo (MCP, *Massive Computation Platform*) tal como se divulga en la solicitud de patente europea 11305518. La DCP 2 y la MCP 3 se acoplan por medio de al menos un enlace de comunicación que se utiliza / utilizan para transmitir unas órdenes recalculadas desde la DCP 2 a la MCP 3 y, en respuesta, las recomendaciones de viajes con precio precalculadas (a las que también se hace referencia brevemente en lo sucesivo en el presente documento como "precios") de vuelta desde la MCP 3 a la DCP 2.

La DCP 2 se equipa con unas interfaces de comunicación adicionales para incorporar datos que esta usa para la determinación de las probabilidades de precisión de los precios almacenados en memoria caché. Estas interfaces incluyen, por ejemplo, unos enlaces de comunicación para incorporar unos datos estadísticos que forman la base del modelo probabilístico y para recibir eventos en tiempo real asíncronos tales como cambios de tarifas y anuncios de disponibilidad de vuelos que son rellenados por líneas aéreas o campañas de promoción orientadas a clientes.

Además, el sistema de base de datos distribuido 1 puede comprender unas plataformas de aplicaciones 4 que organizan y mantienen datos que pueden ser consultados por usuarios finales o clientes externos tales como

agencias de viajes. Las plataformas de aplicaciones 4 se rellenan y se actualizan por medio de la MCP 3 por medio de unos enlaces de comunicación respectivos entre la MCP 3 y las plataformas de aplicaciones 4. Este relleno y actualización se desencadena por medio de las órdenes recalculadas que son emitidas por la DCP 2.

5 Tal como se ha descrito en general en lo que antecede y como se describe más en concreto en lo sucesivo, en respuesta a las órdenes recalculadas que se reciben a partir de la DCP 2, la MCP 3 recalcula los precios de las recomendaciones de viajes y los devuelve a la DCP 2. De forma simultánea, no obstante, la MCP 3 también reenvía las recomendaciones de viajes con precio recalculadas a las plataformas de aplicaciones 4 y que también almacenan (tal como se indica mediante “integración de recomendaciones de viajes” en la figura 2). En consecuencia, también las plataformas de aplicaciones 4 almacenan en memoria caché las recomendaciones de viajes con precio precalculadas que son consultadas por los usuarios, en función de la estrategia de actualización de memoria caché que es puesta en práctica por la DCP 2. Por lo tanto, la presente estrategia de actualización de memoria caché se aprovecha para aplicaciones que proporcionan beneficios a los usuarios, por ejemplo, en forma de respuestas instantáneas a las consultas abiertas. En una disposición de ese tipo, la plataforma de memoria caché de datos 2 actúa como una plataforma de control que está dispuesta para controlar y desencadenar las actualizaciones de las memorias caché de las plataformas de aplicaciones 4. Por lo tanto, ningún usuario o cliente accede a o consulta en la práctica los resultados de consulta de base de datos almacenados en memoria caché que se almacenan en la plataforma de memoria caché de datos 2, sino que estos meramente forman la base de datos de control sobre la que se lleva a cabo la estrategia de actualización de memoria caché. No obstante, en otras disposiciones, la plataforma de memoria caché de datos 2 también puede ser consultada directamente por los usuarios o clientes, o - en otras palabras - la presente estrategia de actualización de memoria caché se puede poner en práctica directamente en una plataforma o varias plataformas de aplicaciones 4 en contraposición a una entidad de control independiente.

25 Las plataformas de aplicaciones 4 comprenden, por ejemplo, una plataforma de aplicaciones de precompra, una plataforma de aplicaciones de análisis de tarifas, y otras plataformas. La plataforma de aplicaciones de precompra es consultada por usuarios finales que desean información acerca de los precios y la disponibilidad de vuelos. Por ejemplo, un usuario final podría dirigir una consulta a la aplicación de precompra con el fin de obtener una visión de conjunto de precios de ofertas de viajes durante la temporada de vacaciones que salen de Niza por debajo de 500 Euros. Debido a las recomendaciones de viajes con precio precalculadas que se almacenan en memoria caché en la aplicación de precompra que se actualizan en consonancia con la presente estrategia de actualización de memoria caché, no es necesario que se calculen precios de los vuelos respectivos en el momento de la aparición de la consulta. Más bien, una lista de ofertas de viajes que cumplen estas restricciones un tanto indeterminadas se puede devolver muy rápido en función de las recomendaciones de viajes con precio que se almacenan en memoria caché en la plataforma de aplicaciones de precompra. El usuario puede seleccionar entonces un viaje de entre la lista devuelta que mejor le convenga y puede emitir entonces una solicitud adicional para reservar en la práctica este viaje. La segunda solicitud se procesa entonces por medio de un motor de reserva (que no se muestra) el cual calcula el precio actual y real y sugiere una oferta vinculante al usuario.

40 Observando a continuación más de cerca la estructura de la plataforma de memoria caché de datos 2 que se ilustra por medio de la figura 3, la plataforma de memoria caché de datos 2 se compone de tres módulos:

- 45 - El gestor de entradas 5 es responsable de recibir entradas tales como unos resultados de consulta de base de datos precalculados a partir de la MCP 3, eventos en tiempo real asíncronos y otra información tal como datos estadísticos para alimentar y actualizar el modelo probabilístico.
- El analizador 6 utiliza el modelo probabilístico en función del cual el mismo se dispone para determinar los resultados de consulta almacenados en memoria caché candidatos que se van a actualizar.
- 50 - Por último, el consolidador 7 rectifica las probabilidades que se determinan por medio del analizador 6 y - cuando sea necesario - también el modelo probabilístico en función de los eventos en tiempo real observados (esto último no se muestra en la figura 3).

55 Además, la DCP 2 incluye una base de datos interna 8 que mantiene los datos de recomendaciones de viajes con precio almacenados en memoria caché. Esta representación solo conserva los atributos de la información de precios que son relevantes para evaluar las probabilidades de obsolescencia y tomar una decisión acerca de la decisión recalculada, tal como, por ejemplo: par de ciudades, la fecha de salida, la duración de la estancia y la fecha del último cálculo, siendo la totalidad de las cuales unas salidas de cálculos que son devueltos por la MCP 3. Otros datos que son utilizados por la MCP 3 con el fin de llevar a cabo sus cálculos tales como tarifas no se reflejan en la DCP 2 debido a que los mismos no son necesarios para llevar a cabo la estrategia de actualización de memoria caché. Por otro lado, no obstante, la DCP 2 enriquece sus datos con atributos de metadatos (que no son parte de los conjuntos de datos que son devueltos por la MCP 3), tal como una precisión supuesta inicial (las probabilidades de que un precio que acaba de ser recalculado por la MCP 3 difiera del cálculo para la reserva), una volatilidad (una indicación de la probabilidad de que un precio difiera con respecto al cálculo para la reserva desde su último cálculo) y una popularidad (con qué frecuencia se busca y se reserva el vuelo). Los datos que son necesarios para establecer estos atributos se mantienen en unas bases de datos externas tales como una base de datos de

volatilidad 10, una base de datos de precisiones iniciales 11 y unos servidores de estadísticas 9. Los atributos de metadatos representan el modelo probabilístico (y, por lo tanto, dependen de las especificidades del modelo probabilístico) en función del cual el analizador 6 determina las probabilidades de si los precios almacenados en memoria caché podrían estar obsoletos (tal como se explicará con mayor detalle en lo sucesivo).

El gestor de entradas 5 se dispone para convertir e integrar la totalidad de las fuentes heterogéneas de información en la base de datos de representaciones locales 8 de los precios que son devueltos por la MCP 3. Esta registra los eventos y las acciones que potencialmente inciden sobre unos precios modelados. Estos eventos incluyen una realimentación de promociones orientadas a clientes y de discrepancias de clientes. Además, los eventos de disponibilidad de vuelos como cancelaciones de vuelos no solo anulan las recomendaciones de viajes almacenadas en memoria caché que se basan directamente en el vuelo cancelado, sino que también pueden influir sobre los datos almacenados en memoria caché en paralelo tales como vuelos del mismo par de ciudades que se programan en el mismo instante del vuelo cancelado. Estos eventos en tiempo real se reenvían entonces al consolidador 7 que los procesa adicionalmente con el fin de rectificar las probabilidades de obsolescencia y las decisiones de recálculo.

Debido a la cantidad de información que está involucrada en el almacenamiento en memoria caché de las recomendaciones de viajes con precio, el empleo de técnicas de codificación es beneficioso. Por medio de esto, los datos con precio que están almacenados en memoria caché en la DCP 2 se correlacionan de forma global con el dominio de datos subyacente que se mantiene en la MCP 3 al tiempo que se reducen los costes para los recursos de almacenamiento de forma significativa. Una codificación probabilística se pone en práctica, por ejemplo, mediante el uso de filtros de Bloom. El efecto de tal codificación es doble. En primer lugar, los filtros de Bloom son conservadores. Los mismos permiten realizar, positivamente, un seguimiento de, al menos y en cualquier caso, aquellos precios que probablemente se ven afectados, por ejemplo, por un evento en tiempo real que indica unos cambios de tarifas, pero estos no son erróneos a la inversa, aquellos precios que se consideran no afectados no se ven afectados en la práctica. Por lo tanto, no hay riesgo alguno de no reconocer precios que se ven potencialmente influenciados por un evento de cambio de tarifas de ese tipo. En segundo lugar, la cantidad de indicaciones de falsos positivos depende, en un sentido estricto, del tamaño asignado del filtro de Bloom, por lo tanto se puede limitar su aparición según sea necesario.

El segundo módulo, el analizador 6, lleva a cabo el primer nivel general de determinación de las probabilidades de si las recomendaciones de viajes con precio almacenadas en memoria caché están obsoletas basándose en el modelo de deterioro probabilístico de la precisión de los precios precalculados que se mantienen en la DCP 2. Este examina y evalúa los metadatos que son añadidos a los precios por el gestor de entradas 5, tal como se ha explicado en lo que antecede. Por lo tanto, el modelo probabilístico que se representa por medio de estos metadatos incluye una métrica con respecto a la volatilidad de precios que se incluye por medio de la base de datos de volatilidad 10, la precisión inicial de los precios que se incorporan a partir de la base de datos de precisiones iniciales 11 y la popularidad de las recomendaciones de vuelos tal como son devueltas por los informes de popularidad a partir de los servidores de estadísticas 9. Esta emite una información de probabilidad y de prioridad con respecto a los precios almacenados en memoria caché al consolidador 7, es decir, unas indicaciones acerca de qué precios necesitan ser recalculados con prioridad basándose solo en la información probabilística estática (es decir, sin considerar evento alguno).

Más en concreto, un modelo probabilístico a modo de ejemplo se basa en los siguientes parámetros para un resultado de consulta de base de datos precalculado x_i :

- La antigüedad t_i del resultado de consulta precalculado: el tiempo desde el último cálculo de este resultado de consulta precalculado por la MCP 3.
- La volatilidad λ_i del resultado de consulta precalculado que se puede obtener a partir de la aparición y los resultados de los cálculos pasados.
- La volatilidad λ_i se puede emplear para proporcionar una estimación de la probabilidad de que un resultado de consulta precalculado permanezca sin cambios después de un tiempo dado: **P (sin cambios después de t) = $e^{-\lambda_i t}$** . También se hace referencia a esto como la probabilidad de que la precisión esperada **prec_i^t = $e^{-\lambda_i t}$** o, más en general, como la probabilidad de que un resultado de consulta de base de datos precalculado, esté obsoleto. Dos funciones a modo de ejemplo de esta disminución de precisión probable a lo largo del tiempo se ilustran en la figura 4a.
- La precisión inicial: puede que los resultados de consulta de base de datos precalculados que son generados por la MCP 3 no sean necesariamente precisos, incluso en tiempo de cálculo si, por ejemplo, la propia MCP 3 basa sus cálculos en unos datos almacenados en memoria caché (y, por lo tanto, obsoletos). Esto conduce a discrepancias adicionales entre los resultados de consulta de base de datos precalculados que son calculados por la MCP 3 y los resultados de cálculo que hipotéticamente se han generado por medio de unos datos subyacentes precisos. Esta discrepancia se puede medir si se encuentra disponible una realimentación respectiva. Se puede inferir, por ejemplo, a partir de los cálculos previos, que el resultado de consulta de base de datos precalculado x_i tiene una probabilidad α_i de ser preciso en el momento del cálculo por la MCP 3. Esto quiere decir que la probabilidad de que un resultado de consulta de base de datos precalculado sea preciso después de un tiempo dado t es $\alpha_i e^{-\lambda_i t}$.
- La precisión de los resultados de consulta de base de datos precalculados globales que se mantienen en la DCP

2 de acuerdo con este modelo se puede considerar entonces como la precisión media:

$$\text{Precisión Global} = \text{Promedio} (a_i e^{-\lambda_i t})$$

- 5 - La popularidad p_i del resultado de consulta de base de datos precalculado: esta es la frecuencia de acceso promedio a este resultado de consulta de base de datos precalculado por los usuarios finales. Tal como también se presenta adicionalmente en lo sucesivo brevemente, puede ser deseable lograr una precisión mejor para los resultados de consulta de base de datos precalculados que son solicitados a menudo por los usuarios. Por lo tanto, la precisión de la totalidad de la memoria caché tal como es vista por los usuarios también se puede definir como que cada valor de precisión se pondera con la popularidad. Por lo tanto, se incluye la proporción de los accesos precisos a los resultados de consulta de base de datos precalculados en contraposición a la proporción esperada de los resultados de consulta de base de datos precalculados precisos:

$$\text{Precisión Usuario} = \sum \frac{p_i}{p_{tot}} a_i e^{-\lambda_i t}$$

15 Estos parámetros del modelo de precisión probabilístico se establecen, por ejemplo, de acuerdo con unos datos históricos estadísticos que se obtienen a partir del servidor de estadísticas 9 basándose en varios días, semanas o meses de historia. Cada uno de los resultados de consulta de base de datos precalculados se modela con estos parámetros con el fin de predecir el estado de los resultados de consulta de base de datos precalculados y, por lo tanto, de predecir la calidad de la totalidad de la memoria caché.

Hay varias formas de utilizar la información del modelo probabilístico con el fin de priorizar y de decidir qué precios se van a recalcular a continuación. El analizador 6 se puede configurar para aplicar esa estrategia o una mezcla de estrategias dependiendo de la situación (por ejemplo, de acuerdo con un acuerdo con el cliente debido a los datos de viajes subyacentes en la MCP 3, dependiendo de la cantidad de datos, dependiendo de los recursos de cálculo disponibles, dependiendo de la objeción en cuanto a la forma en la que se debería optimizar la memoria caché). Se pueden aplicar las siguientes estrategias:

- Precisión de los precios: Esto tiene por objeto el aumento al máximo de la precisión global del dominio de las fechas. En primer lugar se calculan los precios supuestamente inexactos.
- Precisión de los precios, ponderada con la popularidad: Entre los precios que es verosímil que sean inexactos, los precios más populares se recalcularán con una prioridad más alta que los precios menos populares.
- Precisión de los precios, ponderada con la popularidad y por su antigüedad: Como la estrategia previa, pero también teniendo en cuenta el tiempo del último recálculo. Esta estrategia evita la falta de recálculos que es causada por unos precios muy volátiles, en particular en el contexto en donde los recursos de recálculo son limitados en comparación con la cantidad de precios que, en general, se deberían recalcular.
- Modular los pares de ciudades populares basándose en su ubicación geográfica y en la hora de recálculo: Esta estrategia tiene en cuenta adicionalmente unas estadísticas de qué vuelos de par de ciudades se consultan más a menudo en instantes particulares de un día. Como un efecto, se evitan recálculos frecuentes en aquellos instantes en los que rara vez se accede a los vuelos de un par de ciudades específico (debido a que unos datos almacenados en memoria caché inexactos no tienen un efecto perjudicial siempre que las consultas respectivas no tengan lugar virtualmente en la práctica).

Como un efecto secundario, el analizador 6 actualiza la base de datos de modelos de volatilidad 10 basándose en los valores de los precios recientemente recalculados que se reciben a partir de la MCP 3 y que se incorporan en la DCP 2. Debido a que el analizador no puede realizar un seguimiento de la volatilidad real de los precios almacenados en memoria caché basándose en la repetición de recálculos, este puede realimentar esta información estadística de vuelta a la base de datos de modelos de volatilidad 10. Para actualizar el modelo de volatilidad, el analizador 6 cuenta el número de diferencias entre los resultados de precios recién calculados y los valores de precios previamente recibidos. A partir de estas diferencias, este actualiza los parámetros de volatilidad para las partes respectivas de los precios analizados.

De forma similar, el analizador 6 puede actualizar la base de datos de precisiones iniciales 11 de la misma forma. Este también puede pedir otros informes de popularidad, por ejemplo, si se han integrado por primera vez precios a partir de nuevos pares de ciudades en la DCP 2.

En el caso en el que no hay datos históricos y estadísticos, de forma respectiva, para la volatilidad, la precisión o la popularidad de los precios, el analizador 6 lleva a cabo su procesamiento con unos parámetros por defecto para que sea tan conservador como sea posible.

Pasando a continuación al tercer módulo, el consolidador 7, lleva a cabo el segundo nivel de la determinación de probabilidad al tener en cuenta los eventos en tiempo real entrantes. Además, esta es la entidad que genera las órdenes recalculadas y las emite a la MCP 3. Esta toma las salidas del analizador 6 como algo en lo que basar su decisión. Estas proporcionan una primera estimación de las prioridades de recálculo para todos los precios del dominio de datos. Entonces, la misma superpone la totalidad de la información que se recopila a partir de las

diversas fuentes de eventos en tiempo real para rectificar las prioridades de recálculo. Esto da como resultado unas prioridades de recálculo potenciadas.

De forma opcional, este puede considerar cualquier acuerdo de nivel de servicio de cliente tal como, por ejemplo, “garantizar que todos los precios se recalcularán al menos una vez a la semana”, y rectificar las prioridades en consecuencia. El mismo selecciona aquellas entradas en la representación de datos de precios interna 8 con las prioridades más altas en primer lugar y las marca para su recálculo. Debido a que, preferiblemente, el consolidador tiene conocimiento de los recursos de cálculo que se encuentran disponibles en la MCP 3, este es capaz de marcar tantos precios almacenados en memoria caché como puedan ser recalculados por la MCP 3 en un intervalo de tiempo particular. Entonces, la misma envía la orden recalculada resultante a la MCP 3.

La información a partir de los eventos en tiempo real es un medio para mejorar la precisión de los datos almacenados en memoria caché frente a un modelado estadístico estricto. Estos se pueden usar para realizar un seguimiento de lo que está teniendo lugar realmente en lugar de solo lo que se espera. Esto es un medio para controlar las predicciones del modelo estadístico y rectificar las mismas / la misma cuando se muestra que las predicciones son erróneas o inapropiadas. Se pueden contemplar varias clases de eventos en tiempo real con respecto a la presente forma de realización:

Los eventos de los actores tienen lugar, en general, de forma selectiva (es decir, de vez en cuando), pero pueden tener una influencia drástica sobre las decisiones recalculadas. Los clientes externos pueden proporcionar una realimentación acerca de las discrepancias entre la memoria caché y la compra que este está experimentando en su propia plataforma. Esta realimentación se puede usar para rectificar la precisión que se predice por medio del modelo estadístico y, por lo tanto, forzar recálculos antes cuando sea necesario. Cuando un proveedor de datos que se almacenan en la MCP 3 (por ejemplo, un proveedor de viajes que ofrece viajes) lleva a cabo una campaña de promoción que promueve vuelos en determinados pares de ciudades, se puede suponer que estos precios son más volátiles y quedan obsoletos más a menudo. Por lo tanto, podría ser necesario que se aumentara la frecuencia de recálculo de estos precios durante la campaña de promoción. Como otro ejemplo, de vez en cuando puede ser necesaria una operación de mantenimiento para la MCP 3, o se pueden experimentar problemas operativos dentro del sistema 1. En tales casos, se puede ordenar a la DCP 2 que genere menos órdenes recalculadas, o ninguna, hasta que se ha realizado el mantenimiento y se ha subsanado el problema, de forma respectiva.

Los eventos de disponibilidad indican un estado en tiempo real acerca de la precisión de los vuelos almacenados en memoria caché. Dependiendo de la instrucción del evento, se puede saber con certeza que un precio específico del dominio de datos subyacente en la MCP 3 ha cambiado y, por lo tanto, el precio que es almacenado en memoria caché por la DCP 2 se ha vuelto no válido. No obstante, también se pueden ver afectados otros precios, en donde el efecto es incierto y, por lo tanto, se puede aumentar la probabilidad de que estos precios estén obsoletos. Por ejemplo, un evento de “clase cerrada” indica que una clase de reserva particular ha quedado completa para un vuelo particular. Los asientos en este vuelo y clase ya no pueden reservarse y, por lo tanto, los precios respectivos que son almacenados en memoria caché por la DCP 2 se han vuelto no válidos con certeza. No obstante, esto se podría considerar como una indicación de que otras clases del mismo vuelo y / o los asientos en la misma clase en otros vuelos que salen poco antes o después de este vuelo se podrían volver más volátiles. Por lo tanto, sus probabilidades de quedar obsoletos podrían aumentar y el recálculo de estos precios podría ser beneficioso. Como otro ejemplo, se experimenta que las aerolíneas de bajo coste establecen el precio de sus asientos dependiendo de la ocupación del vuelo. Tras las notificaciones de los cambios de ocupación, los precios almacenados en memoria caché respectivos se pueden recalcular con rapidez y, por lo tanto, se mejora / recupera la precisión de la memoria caché.

Las implicaciones de los eventos de cambios de tarifas son difíciles de estimar. Expuesto de una forma sencilla, las tarifas contienen una información y una lógica tal como unas reglas que se usan para calcular un precio de un vuelo particular. Por lo tanto, cuando se calcula el precio real de un vuelo particular, se accede a un conjunto de tarifas, se decide qué tarifa es relevante y se va a aplicar en la práctica y, por último, se calcula el precio. Por lo tanto, hay una relación “vuelo → tarifa o tarifas” (que, no obstante, también puede cambiar a lo largo del tiempo, debido a que pueden cambiar las restricciones sobre qué tarifa es de aplicación a un vuelo particular). No obstante, en general no se realiza un seguimiento de las relaciones en el otro sentido “tarifa → vuelos”, es decir, no queda claro qué tarifa se aplica, de hecho, a qué vuelos. Además, un cambio en una tarifa incide potencialmente sobre una enorme cantidad de precios que se calculan a partir del dominio de datos subyacente.

Con el fin de determinar el impacto de los eventos de tarifas, la comunicación entre la MCP 3 y la DCP 2 se puede utilizar para dotar a la DCP 2 de una correlación con las tarifas que son aplicadas por la MCP 3 para calcular precios. Cuando se calculan precios en respuesta a las órdenes recalculadas, la MCP 3 registra todas las tarifas a las que se accede para calcular los precios solicitados. Esta información se almacena entonces de forma global en una correlación tarifas ↔ vuelos y se mantiene durante cada cálculo por la MCP 3. En el momento en el que se recibe un evento de cambio de tarifas, el gestor de entradas 5 explora esta correlación global con el fin de determinar los vuelos que se ven afectados por el evento de cambio de tarifas y los marca como “actualizados”. Obsérvese que un cambio de tarifas no implica necesariamente un cambio de precios para una recomendación de viajes, tal como se ha explicado brevemente en lo que antecede.

El consolidador 7 evalúa en primer lugar la influencia potencial del evento en tiempo real sobre los precios almacenados en memoria caché en lugar de iniciar un recálculo de las recomendaciones de viajes almacenadas en memoria caché sin haber considerado la relación del evento con el modelo probabilístico básico. Tales eventos se analizan en primer lugar con respecto a su representación en el modelo probabilístico.

Tal como se ha bosquejado en lo que antecede, las especificidades de este análisis dependen del tipo de modelo probabilístico que se usa. En el caso de un modelo conceptual, se conoce, a partir del proceso de generación de modelos, qué eventos se han considerado en el modelo probabilístico. Por lo tanto, los eventos entrantes se comprueban frente a las características del modelo probabilístico que se conocen a partir del proceso de generación de modelos con el fin de determinar si, en el modelo, se encuentra presente, o no, un evento particular.

Por otro lado, se puede usar un modelo histórico - estadístico y, en consecuencia, puede que no se sepa a priori qué eventos han conducido a una base de datos histórica - estadística del modelo. En este caso, el siguiente enfoque se puede adoptar con el fin de determinar si un evento particular está representado, o no, en el modelo probabilístico (por razones de completión: el siguiente enfoque no se limita a tipo particular alguno de modelo probabilístico y también se puede emplear, por ejemplo, para los modelos conceptuales).

Dependiendo del tamaño del dominio de datos que es formado por todos los resultados de consulta de base de datos precalculados almacenados en memoria caché, los resultados de consulta de base de datos almacenados en memoria caché se subdividen en unas porciones más pequeñas, a las que se hace referencia como "cuotas" en lo sucesivo en el presente documento. En el ejemplo de los resultados de consulta de base de datos precalculados que son unas recomendaciones de viajes con precio, las cuotas están formadas, por ejemplo, por unos pares de origen - destino, es decir, una cuota incluye todas las recomendaciones de viajes con precio precalculadas que tienen la misma ubicación de origen y de destino. Como alternativa, si el dominio de datos solo es de un tamaño limitado, no se lleva a cabo subdivisión alguna. En otras palabras, en este caso, solo hay una cuota que cubra todos los resultados de consulta de base de datos precalculados que se mantienen en la DCP 2.

Después de que se haya detectado un evento en tiempo real asíncrono, se identifican una cuota o más cuotas que potencialmente se ven afectadas por este evento, es decir, que contienen unos resultados de consulta de base de datos precalculados que pueden tener una probabilidad aumentada de estar obsoletos debido al evento detectado. Una precisión real de la al menos una cuota se determina, por ejemplo, mediante la estimación de su precisión mediante el recálculo de una muestra particular o aleatoria de los resultados de consulta de base de datos precalculados en la cuota. Por ejemplo, si la cuota contiene 10.000 resultados de consulta de base de datos precalculados, se recalcula una muestra de 20 resultados de consulta de base de datos precalculados. Este recálculo se lleva a cabo de forma similar a otros recálculos. La DCP 2 genera y emite unas órdenes recalculadas a la MCP 3. La MCP 3 lleva a cabo los recálculos y devuelve sus resultados, es decir, los resultados de consulta de base de datos precalculados de muestra actualizados a la DCP 2. La DCP 2 compara los resultados de los recálculos de muestra con los resultados de consulta de base de datos correspondientes que previamente se han almacenado en memoria caché y, por ejemplo, determina el porcentaje de los resultados de consulta de base de datos precalculados de muestra que han cambiado (es decir, que se han quedado obsoletos o que han sido inexactos) y / o el porcentaje de los resultados de consulta de base de datos precalculados de muestra que permanecen sin cambios (es decir, que no se han quedado obsoletos y seguían siendo precisos). Estos porcentajes de los resultados de consulta de base de datos precalculados de muestra se podrían transferir a la cuota global como una estimación de la precisión de la cuota global, es decir, en general se podría concluir, a partir de estos porcentajes de muestra, que la cuota global presenta una inexactitud / precisión similar a la de la muestra.

En una etapa siguiente, la precisión real determinada de la cuota se compara con la precisión de la cuota que se predice por medio del modelo probabilístico. Dependiendo del resultado de esta comparación, se concluye si el evento en tiempo real asíncrono particular está representado, o no, en el modelo probabilístico y, posiblemente, en qué medida lo está. Si la precisión real determinada se encuentra por debajo de la precisión que se predice por medio del modelo probabilístico en una medida superior a una dada, es decir, la precisión de muestra es sustancialmente menor que la precisión que se predice por medio del modelo, no se considera que el evento esté incluido en el modelo probabilístico. Si, por otro lado, la precisión real determinada se encuentra por debajo de la precisión que se predice por medio del modelo probabilístico menos que la medida dada, o ambas precisiones son idénticas o la precisión real determinada es incluso superior a la precisión predicha, se considera que el evento está representado en el modelo probabilístico. La diferencia entre ambas precisiones también es un indicador de la medida en la que el evento se ha incorporado en el modelo.

Se hace notar que el propio muestreo se puede reconocer como una forma de detectar los eventos en tiempo real asíncronos que han tenido lugar en primer lugar. Además de o como alternativa a las indicaciones explícitas de eventos en tiempo real asíncronos tales como una indicación procedente de una línea aérea de que se ha cancelado un vuelo particular o una indicación de una campaña de promoción, el muestreo se puede utilizar para detectar de forma implícita los eventos en tiempo real asíncronos que no son indicadas de forma explícita por una tercera parte. Para este fin, se puede emplear un proceso de muestra continuo u ordinario a lo largo de todas las cuotas con el fin de estimar la precisión real de todas las cuotas de los resultados de consulta de base de datos precalculados. Esto permite una detección implícita de los eventos que no se reflejan en el modelo probabilístico si, de la comparación

de la precisión real determinada y la precisión que se predice por medio del modelo probabilístico, resulta que la primera es sustancialmente más baja que esta última.

Se hace notar adicionalmente que la identificación de una cuota o más cuotas que potencialmente se ven afectadas por un evento detectado depende de las características y el conocimiento acerca de los eventos. Si, por ejemplo, un evento se ha detectado de forma implícita por medio de un proceso de muestreo cíclico, las cuotas potencialmente afectadas se han determinado al mismo tiempo (en concreto, aquellas cuotas cuya precisión real determinada es sustancialmente menor que la precisión predicha). Por otro lado, una señalización explícita de un evento también puede contener una indicación de unos resultados de consulta de base de datos precalculados potencialmente afectados. Si, por ejemplo, una línea aérea ha señalado la cancelación de un vuelo particular, se podría considerar que los precios de los vuelos entre el mismo origen y destino durante los instantes antes y después de los vuelos cancelados se ven potencialmente afectados. Además, a partir de la experiencia pasada se puede saber qué eventos (señalizados de forma explícita) tienen un impacto potencial sobre la precisión de qué resultados de consulta de base de datos precalculados. Por ejemplo, es posible volver a realizar un seguimiento de qué eventos en tiempo real y de la vida real han sido responsables de unas precisiones disminuidas de las cuotas que se detectan por medio de un proceso de muestreo continuo y volver a obtener un cierto conocimiento acerca del impacto habitual de determinados eventos en tiempo real asíncronos de esta forma.

El análisis anterior de estimación de la precisión real de las cuotas y de comparación de esta con la precisión que se predice por medio del modelo probabilístico se describe a continuación de una forma más detallada y más formal.

Se hace referencia a la precisión real de un resultado de consulta de base de datos precalculado x_i como **Prec_i**. El resultado de consulta de base de datos precalculado x_i es, en un instante particular en el tiempo, o bien preciso (**Prec_i = 1**), es decir, igual al resultado de su recálculo, o inexacto (**Prec_i = 0**), es decir, diferente de los resultados de su recálculo. La recopilación de todos los resultados de consulta de base de datos precalculados se puede dividir, desde un punto de vista lógico, en las cuotas D_j . Las siguientes consideraciones se refieren a una cuota **D** a modo de ejemplo específica. La precisión promedio de los resultados de consulta de base de datos precalculados que están contenidos en la cuota **D** tal como se predice por medio del modelo probabilístico se denota como **Prec_{modelo}(D)**. Por otro lado, se hace referencia a la proporción de los resultados de consulta de base de datos precalculados precisos en la práctica en **D** como **Prec(D)**.

Tal como se ha descrito en lo que antecede, con el fin de determinar o estimar **Prec(D)**, se lleva a cabo un recálculo de muestra de **N** resultados de consulta de base de datos precalculados en **D**. Se hace referencia al número de resultados precisos entre los **N** como **AC_N**, es decir, después de haber recalculado los **N** resultados de consulta de base de datos precalculados de muestra, **AC_N** resultados de consulta de base de datos recalculados son idénticos a sus contrapartidas previamente almacenadas en memoria caché mientras que (**N - AC_N**) resultados de consulta de base de datos recalculados difieren de sus contrapartidas previamente almacenadas en memoria caché y, por lo tanto, se ha hallado que son inexactos.

Tal como se ha indicado en lo que antecede, la probabilidad de precisión para un resultado de consulta de base de datos precalculado x_i dentro de la cuota **D** es **Prec(D)**.

Esto se puede generalizar para los **N** resultados de consulta de base de datos precalculados de muestra dentro de **D** mediante el empleo de la ley binomial $AC_N \sim B(N, \text{Prec}(D))$. Por lo tanto, la probabilidad de que **K** resultados de consulta de base de datos precalculados de entre las **N** muestras sean precisos es

$$P(AC_N = K) = \binom{N}{K} \times \text{Prec}(D)^K \times (1 - \text{Prec}(D))^{N-K}$$

La esperanza estadística de esta ley es $E(AC_N) = N \times \text{Prec}(D)$.

Por lo tanto, la relación $\frac{AC_N}{N}$ es una estimación de la precisión real de la cuota **D**, **Prec(D)**, que es transmitida por el resultado del recálculo de muestra de los **N** resultados de consulta de base de datos precalculados de muestra. Por otro lado, **Prec_{modelo}(D)** es la estimación de precisión de **Prec(D)** que se da por medio del modelo probabilístico.

Con el fin de determinar si, en el modelo probabilístico, se encuentra presente, o no, un evento en tiempo real dado, es necesario que se decida si $\frac{AC_N}{N}$ o si **Prec_{modelo}(D)** es la aproximación más fidedigna de **Prec(D)** en el caso

verosímil en el que ambos valores difieren entre sí, en particular si $\frac{AC_N}{N} \ll \text{Prec}_{\text{modelo}}(D)$. Para preparar Esta decisión, se pueden especificar dos hipótesis **H₀** y **H₁**: **H₀**: La predicción del modelo probabilístico indica de forma

apropiada la precisión de la cuota D , es decir, $Prec(D) \cong Prec_{\text{modelo}}(D) \cdot \frac{AC_N}{N}$ es menor que $Prec_{\text{modelo}}(D)$ debido a las desviaciones estadísticas que son causadas por el proceso de muestreo (es decir, de forma accidental, los N resultados de consulta de base de datos precalculados seleccionados son más inexactos que el promedio de todos los resultados de consulta de base de datos precalculados en D y, por lo tanto, las N muestras no se reflejan de forma representativa $Prec(D)$).

H_1 : Tuvo lugar algo que no era esperado por el modelo probabilístico y, por lo tanto, $Prec_{\text{modelo}}(D)$ es una estimación inapropiada de $Prec(D)$. Más bien, la estimación de precisión real que se logra por medio del recálculo de muestra es la mejor estimación que se encuentra disponible, es decir, $Prec(D) \cong \frac{AC_N}{N}$.

Tomar la decisión D_0 de que H_0 es cierta o la decisión D_1 de que H_1 es cierta se puede hacer de la siguiente forma: Si la precisión real es menor que la precisión predicha en una medida superior a una dada, hay una verosimilitud aumentada de que la precisión que se predice por medio del modelo probabilístico sea inapropiada y de que sea necesario reducirla. Este es, por ejemplo, el caso si $Prec_{\text{modelo}}(D)$ indica una precisión $Prec(D)$ de un 85 %, mientras que la precisión real que es recibida por el recálculo de muestra solo indica únicamente un 20 % de resultados de consulta de base de datos precalculados precisos dentro de D . En este caso, D_0 es la decisión correcta. Si, no obstante, la precisión real que es recibida por el muestreo y la precisión que se predice por medio del modelo son similares, es decir, la desviación entre ambas indicaciones es relativamente pequeña, la precisión que se predice por medio del modelo probabilístico es probablemente apropiada. Por dar un ejemplo, este puede ser el caso si $Prec_{\text{modelo}}(D)$ indica una precisión $Prec(D)$ de un 85 % y $\frac{AC_N}{N}$ asciende a un 80 % de precisión. En este caso, D_1 es la decisión correcta.

Con el fin de definir la frontera entre D_0 y D_1 , se introduce un umbral K_{Th} : Si $AC_N > K_{Th}$, entonces la precisión que se predice por medio del modelo probabilístico $Prec_{\text{modelo}}(D)$ se considera como una aproximación apropiada de $Prec(D)$, se elige D_0 . Si $AC_N \leq K_{Th}$, entonces no se considera que la precisión que se predice por medio del modelo probabilístico $Prec_{\text{modelo}}(D)$ sea una aproximación apropiada de $Prec(D)$, se elige D_1 .

Con el fin de ajustar K_{Th} a un valor apropiado, es ventajoso tener información acerca de la calidad del modelo probabilístico. Habitualmente, tal información se encuentra disponible a partir de un proceso de verificación de modelos que se ha llevado a cabo antes de implantar el modelo y / o durante su uso continuado. Por ejemplo, se puede haber observado durante un año de uso del modelo probabilístico que los eventos en tiempo real asíncronos que no son reflejados por el modelo probabilístico son bastante raros y tienen lugar, por ejemplo, en solo tres días de cada 365 días. En este caso, se puede suponer que $Prec_{\text{modelo}}(D)$ es apropiado (= H_0) aproximadamente un 99 % del tiempo e inapropiado solo aproximadamente un 1 % del tiempo (= H_1). En otro ejemplo, los eventos en tiempo real asíncronos que no se incluyen en el modelo probabilístico pueden tener lugar relativamente a menudo tal como en 180 días de un año. Por lo tanto, se puede suponer que $Prec_{\text{modelo}}(D)$ es apropiado solo en uno de cada dos casos, es decir, hasta un 50 % del tiempo (= H_0) e inapropiado de lo contrario (= H_1).

Una forma a modo de ejemplo de establecer K_{Th} al tener en cuenta la calidad / fiabilidad pasada del modelo probabilístico es considerar la probabilidad de una decisión errónea D_1 , es decir, no se considera que $Prec_{\text{modelo}}(D)$ sea una estimación apropiada de $Prec(D)$ a pesar de que la misma lo es en la práctica. La probabilidad de que la predicción del modelo probabilístico sea correcta se puede expresar como:

$$P(AC_N = K | H_0) = \binom{K}{N} \times Prec_{\text{modelo}}(D)^K \times (1 - Prec_{\text{modelo}}(D))^{N-K}$$

La probabilidad α de que la hipótesis H_0 sea cierta y de que se tome una decisión errónea D_1 es la probabilidad de que AC_N sea menor que K_{Th} :

$$\alpha = P(D_1 | H_0) = P(AC_N \leq K_{Th} | H_0) = \sum_{K \leq K_{Th}} \binom{K}{N} \times Prec_{\text{modelo}}(D)^K \times (1 - Prec_{\text{modelo}}(D))^{N-K}$$

Esta fórmula permite elegir α y definir K_{Th} de acuerdo con la fiabilidad pasada del modelo probabilístico. Si, por ejemplo, el modelo probabilístico ha sido correcto un 99 % de es tiempo, α se establece a 0,01. El valor umbral de K_{Th} se puede calcular entonces usando la fórmula previa. Si el número de K resultados de recálculo de muestra precisos es menor que K_{Th} , es decir, se considera que la estimación de precisión real $\frac{AC_N}{N}$ que se determina por medio de los recálculos de muestra es una estimación más fiable de $Prec(D)$ que $Prec_{\text{modelo}}(D)$:

Número N de recálculos de muestra con el fin de estimar la precisión real de la cuota D	N = 10	N = 20	N = 50	N = 100
Número K de los N resultados de consulta de base de datos precalculados de muestra que se predice que son precisos por medio del modelo probabilístico con Prec_{modelo} = 0,85	K = 8	K = 17	K = 42	K = 85
Umbral K_{Th} (con $\alpha = 0,01$)	6	14	36	76
Rango para AC_N (recálculos de muestra precisos) en el que no se considera que el evento está representado en el modelo probabilístico (con $\alpha = 0,01$ y una K_{Th} correspondiente)	AC_N ≤ 5	AC_N ≤ 13	AC_N ≤ 35	AC_N ≤ 75

Por lo tanto, para resumir, si **AC_N** es igual a o se encuentra por encima de **K_{Th}**, se considera que el evento en tiempo real asíncrono está representado en el modelo probabilístico y no es necesaria rectificación alguna de las probabilidades de los resultados de consulta de base de datos precalculados respectivos. Si, no obstante, si **AC_N** se encuentra por debajo de **K_{Th}**, se supone que el evento no está representado en el modelo probabilístico.

Para los eventos que no se pueden predecir y, por lo tanto, no se incluyen en el modelo probabilístico en absoluto como, por ejemplo, campañas de promociones o eventos de mantenimiento, los eventos se procesan tan pronto como sea posible. En general, esto quiere decir que las probabilidades de aquellos resultados de consulta de base de datos precalculados que potencialmente se ven afectados por un evento en tiempo real particular se rectifican, en particular se disminuyen. Continuando con el ejemplo anterior de determinación de si, en el modelo probabilístico, se encuentra presente, o no, un evento, opcionalmente esta rectificación se logra mediante la aplicación de la

estimación para **Prec (D)** que se obtiene por medio de los recálculos de muestra, es decir, $\frac{AC_N}{N}$, a las probabilidades de estar obsoletos de los resultados de consulta de base de datos precalculados en la cuota **D**. Por

ejemplo, si ha resultado que $\frac{AC_N}{N}$ es un 25 % menor que la **Prec_{modelo} (D)**, las predicciones de precisión actuales del modelo probabilístico para los resultados de consulta de base de datos precalculados en la cuota **D** se disminuyen, por ejemplo, también en un 25 %. Esta rectificación a modo de ejemplo se visualiza por medio de la figura 4b (los ejes x e y en la figura 4 tienen el mismo significado que en la figura 4a, es decir, el eje x ilustra el tiempo desde el último recálculo (= la antigüedad **t**) en horas, el eje y muestra la probabilidad de precisión / obsolescencia).

Después de que se haya realizado esta rectificación, la decisión de qué resultados de consulta de base de datos precalculados se van a recalcularse se toma tal como es habitual, es decir, de la misma forma que si no hubiera tenido lugar rectificación alguna de las probabilidades. Haciendo referencia de nuevo al modelo histórico - estadístico a modo de ejemplo que se ha bosquejado en lo que antecede, la probabilidad expuesta por el modelo probabilístico para la precisión / obsolescencia es $e^{-\lambda_i t}$. No obstante, debido a que la misma se puede rectificar en respuesta a los eventos en tiempo real asíncronos reconocidos, por ejemplo, mediante la aplicación de la estrategia que se ha descrito con detalle en lo que antecede, en la práctica se puede recalcularse paso a paso: en el instante τ_n la precisión / obsolescencia probable $Prec_i^{\tau_n}$ se recalcula a partir de su valor previo en el instante τ_{n-1} :

$$Prec_i^{\tau_n} = Prec_i^{\tau_{n-1}} \times e^{-\lambda_i(\tau_n - \tau_{n-1})}$$

El siguiente cálculo de precisión por el analizador 6 comienza de nuevo a partir del valor previo (que se ha rectificado potencialmente en respuesta a los eventos en tiempo real).

La rectificación de las probabilidades en respuesta a los eventos en tiempo real asíncronos entrantes, por ejemplo, la precisión que se predice por medio del modelo probabilístico, se restablece después de que los resultados de consulta de base de datos precalculados respectivos hayan sido recalculados por la MCP 3 la siguiente vez.

Los eventos que están representados en el modelo probabilístico, al menos hasta un cierto punto, como tarifas o cambios de disponibilidad, se acumulan opcionalmente y su apariencia se compara habitualmente con las predicciones del modelo probabilístico. Cuando los picos de evento no coinciden con el modelo a nivel local, es decir, una ráfaga de eventos se encuentra significativamente fuera de la estadística que subyace al modelo probabilístico, los precios afectados se marcan como potencialmente obsoletos con el fin de recalcularse los mismos tan pronto como sea posible. Por medio de esto, se elimina por filtrado el "ruido" que es causado por los eventos que se encuentran ya presentes en el modelo probabilístico y que, por lo tanto, ya han sido considerados por las determinaciones que son realizadas por el analizador 6 de antemano.

De forma opcional, el manejo de los eventos en tiempo real asíncronos entrantes se realiza por medio de un componente dedicado que es funcionalmente independiente de o un subcomponente del consolidador 7, un módulo de gestión de eventos. Los eventos son recibidos por el gestor de entradas 5 e inicialmente se almacenan sin

agregación en el depósito de datos interno. El gestor de eventos recibe cualquier evento en tiempo real asíncrono entrante, procesa el evento, por ejemplo, mediante la aplicación de la estrategia que se ha descrito con detalle en lo que antecede, y posiblemente lleva a cabo la rectificación de las probabilidades de que los resultados de consulta de base de datos precalculados afectados estén obsoletos / sean inexactos. De forma opcional, la estrategia que se ha descrito en lo que antecede solo se aplica si un número dado de eventos ha sido reconocido por el gestor de eventos aplica la estrategia que se ha descrito en lo que antecede y la precisión probable de cada resultado de consulta de base de datos precalculado se rectifica en consecuencia solo entonces, si es necesario.

De forma opcional, para fines de optimización, el consolidador 7 trabaja sobre una vista de cuadrícula de la representación de datos interna 8, es decir, el mismo considera grupos de precios adyacentes de forma conjunta durante su algoritmo en lugar de trabajar sobre el conjunto de precios de una forma aislada. En el presente enfoque, los conjuntos de datos de precios adyacentes se ven como un único conjunto de datos con unos valores de las propiedades agregadas. Trabajar sobre conjuntos de datos agregados limita la generación de unas órdenes recalculadas dispersas, y, por lo tanto, aumenta las oportunidades de mutualización y de optimización en la MCP 3. Esto reduce los costes de cálculo.

La figura 5 resume la descripción detallada anterior y da una visión de conjunto del método de actualización de memoria caché que se presenta en el presente documento.

El proceso de mantener actualizada la memoria caché de los resultados de consulta de base de datos precalculados comienza con la determinación 14 de las probabilidades de inexactitud de los resultados de consulta almacenados en memoria caché. Esta determinación se compone de dos actividades que están ubicadas en dos niveles lógicos: En primer lugar y en general, se emplea un modelo predictivo basándose en datos estadísticos y probabilísticos con el fin de estimar la verosimilitud de que un resultado de consulta almacenado en memoria caché particular no se corresponda con el resultado de consulta que (hipotéticamente) se ha recalculado. En segundo lugar y más en concreto, se tienen en cuenta unos eventos en tiempo real que potencialmente inciden sobre y aumentan, de forma respectiva, la probabilidad de que los resultados de consulta almacenados en memoria caché estén obsoletos. Estos eventos en tiempo real se caracterizan por el hecho de que los mismos, en general, no indican una inexactitud de unos resultados de consulta almacenados en memoria caché particulares con certeza, sino que son indeterministas a este respecto. Por lo tanto, cuando tienen lugar los mismos, solo se pueden extraer unas conclusiones probabilísticas acerca de la verosimilitud de precisión e inexactitud, de forma respectiva.

En función de las probabilidades determinadas de que los resultados de consulta de base de datos almacenados en memoria caché estén obsoletos, las órdenes recalculadas de base de datos son emitidas de forma automática 15 por la DCP 2. Estas órdenes son recibidas por la MCP 3 que entonces recalcula los resultados respectivos y devuelve estos 16 a la DCP 2. A su vez, la DCP 2 recibe los resultados y los almacena 17 en una representación local 8. Esto concluye un ciclo de actualización y el siguiente ciclo vuelve a realizar una iteración con la determinación de probabilidad 14.

A continuación, se describe con respecto a la figura 6 un ejemplo particular con respecto a la temporización del procedimiento de la presente estrategia de actualización de memoria caché. En el presente ejemplo, la DCP 2 está configurada para generar órdenes recalculadas cada 20 minutos, es decir, una ronda o ciclo de determinación de las probabilidades de si los datos almacenados en memoria caché están obsoletos y de generación y emisión de las órdenes recalculadas lleva 20 minutos. Los recursos en la MCP 3 se conocen a priori durante todo un día y la DCP 2 está al tanto de los recursos de cálculo que se encuentran disponibles en la MCP 3 y, por lo tanto, es capaz de sincronizar la cantidad de recálculos con los recursos disponibles de la MCP 3.

Al principio de un ciclo de recálculo, la DCP 2 analiza la precisión actual de los resultados de consulta de base de datos almacenados en memoria caché, es decir, las recomendaciones de viajes con precio que se almacenan en la base de datos interna 8. La ronda produce un conjunto de órdenes recalculadas que se procesarán por medio de la MCP 3 al final de la ronda de 20 minutos. Mientras tanto, en el lado de la MCP 3, las órdenes que proceden del último ciclo se están calculando y unas nuevas recomendaciones de precios se generan y se transmiten de vuelta a la DCP, en donde las mismas se almacenan y se encuentran disponibles para el análisis y la actualización de la información periódica en el siguiente ciclo.

La figura 6 muestra que la MCP tiene una cantidad significativa de recursos disponibles en el intervalo de tiempo entre las 04:00 y las 05:00 a. m., por lo tanto se puede llevar a cabo una gran cantidad de recálculos en esta hora. Después de lo anterior, no obstante, no se encuentra disponible recurso alguno hasta las 9:00 a. m., por lo tanto no es posible recálculo alguno en ese instante. Más tarde durante el día, de 09:00 a. m. a 7:00 p. m., algunos recursos se encuentran disponibles en la MCP 3.

Durante el ciclo que comienza a las 04:20 a. m., el analizador 6 analiza la precisión de la memoria caché, mientras que el consolidador 7 genera unas órdenes recalculadas en consecuencia. Esas órdenes serán puestas en práctica por la MCP 3 a las 04:40 a. m. El analizador 6 se centra en la recomendación de precios de MCP que el mismo recibió al principio de la ronda. Este cuenta las diferencias entre los precios recibidos y los precios previos cuyos valores se han almacenado en un depósito interno. Basándose en las diferencias, el mismo rectifica la fuente,

periódica en cuanto a la “volatilidad”, de información. El gestor de entradas 5 guarda los precios de MCP recibidos para su inspección adicional.

5 En el ciclo de 4:40 a 5:00 a. m., la MCP 3 procesa las órdenes recalculadas que se reciben a partir de la DCP 2 durante el intervalo de 04:20 a 4:40 a. m. La DCP 2 está al tanto de que la misma no puede generar orden recalculada alguna para el sector de tiempo por venir (05:00 a. m.) y sus sucesores hasta las 09:00 a. m. No obstante, esta sigue analizando el modelo de datos de forma continua para actualizar la precisión actual de todas las recomendaciones de viajes con precio almacenadas en memoria caché. Esta realizará lo mismo para cada ciclo futuro hasta las 08:40 a. m.

10 A las 08:40 a. m., el analizador 6 determina que la precisión de la memoria caché disminuyó durante las cuatro horas previas sin recálculo alguno. Este genera unas órdenes recalculadas en consecuencia a lo largo de los siguientes ciclos, pero solo hasta una medida menor debido a la cantidad limitada de recursos disponibles en la MCP 3 desde las 09:00 a. m. hasta las 7:00 p. m. Entonces, a las 09:00 a. m., la MCP 3 comienza a procesar las nuevas órdenes recalculadas que la misma recibió en el intervalo anterior (es decir, de 08:40 a 09:00 a. m.) y se detendrá después del final de la ronda que dura de 6:40 a 7:00 p. m.

15 Después de eso, no se encuentra disponible recurso adicional alguno en la MCP 3 de principio a fin de la noche. Por lo tanto, la DCP 2 no generará orden recalculada adicional alguna, sino que continuará analizando la precisión de la memoria caché en función del modelo probabilístico y posiblemente los eventos en tiempo real entrantes.

20 Por último, la figura 7 es una representación esquemática de un sistema informático que proporciona la funcionalidad de la plataforma de memoria caché 2 de la figura 2. Dentro de la plataforma de memoria caché 2 se puede ejecutar un conjunto de instrucciones, para dar lugar a que el sistema informático lleve a cabo cualquiera de las metodologías que se analizan en el presente documento. La plataforma de memoria caché 2 incluye un procesador 101, una memoria principal 102 y un dispositivo de interfaz de red 103, que se comunican entre sí por medio de un bus 104. De forma opcional, esta puede incluir adicionalmente una memoria estática 105 y una unidad de disco 106. Una pantalla de vídeo 107, un dispositivo de entrada alfanumérico 108 y un dispositivo de control de cursor 109 pueden formar una interfaz de usuario de navegador de lista de distribución. El dispositivo de interfaz de red 103 conecta la plataforma de memoria caché de datos 2 con la plataforma de cálculo 3, las fuentes de los datos estadísticos que son necesarios para rellenar el modelo predictivo tales como los servidores de estadísticas 9, la base de datos de volatilidad 10 y la base de datos de precisión inicial 11, las fuentes de los eventos en tiempo real, Internet y / o cualquier otra red. Un conjunto de instrucciones (es decir, un soporte físico) 110 que materializa una cualquiera, o la totalidad, de las metodologías que se han descrito en lo que antecede, residen completa, o al menos parcialmente, dentro de, o en, un medio legible por máquina, por ejemplo la memoria principal 102 y / o el procesador 101. Un medio legible por máquina en el que reside el soporte físico 110 también puede ser un soporte de datos no volátil 111 (por ejemplo, un disco duro magnético no extraíble o un disco extraíble óptico o magnético) que sea parte de la unidad de disco 106. El soporte físico 110 se puede transmitir o recibirse adicionalmente como una señal propagada 112 por medio de Internet a través del dispositivo de interfaz de red 103.

35 40 La presente estrategia de actualización de memoria caché proporciona un medio para generar de forma automática unas decisiones de recálculo. Esta determina qué resultados de consulta almacenados en memoria caché se han de recalcular y controla el recálculo también según el tiempo, al tener en cuenta los recursos de cálculo disponibles en la plataforma de cálculo. Por lo tanto, en general, la precisión de los resultados de consulta almacenados en memoria caché se estima en el modelo probabilístico que modela la actualización y desactualización, respectivamente, a lo largo del tiempo. Estos análisis de desactualización posibilitan el procesamiento de varios miles de millones de conjuntos de datos por hora, que subyacen al recálculo de los resultados de consulta de base de datos almacenados en memoria caché.

REIVINDICACIONES

1. Un método de actualización de unos resultados de consulta de base de datos precalculados en un sistema de base de datos distribuido (1), en donde el sistema de base de datos distribuido (1) comprende una plataforma de memoria caché de datos (2) que mantiene los resultados de consulta de base de datos precalculados y una plataforma de cálculo (3) para calcular los resultados de consulta de base de datos precalculados basándose en datos que se mantienen en la plataforma de cálculo (3), en donde la plataforma de memoria caché de datos (2) mantiene un modelo probabilístico que modela las discrepancias entre los resultados de consulta de base de datos precalculados que se mantienen en la plataforma de memoria caché de datos (2) y los resultados de consulta de base de datos reales supuestos, comprendiendo el método:
- determinar, por la plataforma de memoria caché de datos (2), las probabilidades de estar obsoletos de los resultados de consulta de base de datos precalculados basándose en el modelo probabilístico;
 - detectar, por la plataforma de memoria caché de datos (2), un evento en tiempo real asíncrono que tiene una influencia probabilística sobre las discrepancias entre los resultados de consulta de base de datos precalculados que se mantienen en la plataforma de memoria caché de datos (2) y los resultados de consulta de base de datos reales supuestos;
 - analizar, en la plataforma de memoria caché de datos (2), si el evento en tiempo real asíncrono está representado en el modelo probabilístico;
 - aumentar, por la plataforma de memoria caché de datos (2), las probabilidades determinadas de estar obsoletos de los resultados de consulta de base de datos precalculados si se determina que el evento en tiempo real asíncrono no está representado en el modelo probabilístico;
 - emitir de forma automática (15), por la plataforma de memoria caché de datos (2), unas órdenes recalculadas a la plataforma de cálculo (3) para actualizar unos resultados de consulta de base de datos precalculados en función de las probabilidades determinadas de estar obsoletos de los resultados de consulta de base de datos precalculados, en donde se ordena que se recalculen los resultados de consulta de base de datos precalculados que tienen una probabilidad de estar obsoletos por encima de un umbral dado; y
 - recibir (17), en la plataforma de memoria caché de datos (2), los resultados de consulta de base de datos precalculados actualizados como resultados de las órdenes recalculadas.
2. Método de acuerdo con la reivindicación 1, en donde detectar el evento en tiempo real asíncrono y / o analizar si el evento en tiempo real asíncrono está representado en el modelo probabilístico comprende:
- determinar una precisión de una cuota de los resultados de consulta de base de datos precalculados que se mantienen en la plataforma de memoria caché de datos, viéndose potencialmente afectada la cuota de los resultados de consulta de base de datos precalculados por el evento en tiempo real asíncrono;
 - comparar la precisión determinada con una predicción de la precisión de la cuota de los resultados de consulta de base de datos precalculados por el modelo probabilístico; y
 - determinar que el evento en tiempo real asíncrono no está representado en el modelo probabilístico si la precisión determinada se encuentra por debajo de la precisión predicha hasta una medida dada.
3. Método de acuerdo con la reivindicación 2, en donde determinar la precisión de la cuota de los resultados de consulta de base de datos precalculados comprende emitir unas órdenes recalculadas de muestra a la plataforma de cálculo y comparar los resultados de las órdenes recalculadas de muestra con los resultados de consulta de base de datos precalculados respectivos previamente mantenidos en la plataforma de memoria caché de datos.
4. Método de acuerdo con la reivindicación 2 o 3, en donde la medida dada se define por medio de un valor umbral que se define de acuerdo con la fiabilidad pasada del modelo probabilístico.
5. Método de acuerdo con cualquiera de las reivindicaciones 2 a 4, en donde aumentar las probabilidades determinadas de los resultados de consulta de base de datos precalculados comprende aumentar las probabilidades de los resultados de consulta de base de datos precalculados en la cuota a la precisión determinada de la cuota.
6. Método de acuerdo con cualquiera de las reivindicaciones 1 a 5, que comprende adicionalmente:
- para los eventos en tiempo real que se determinan como que no están representados en el modelo probabilístico, emitir unas órdenes recalculadas con respecto a los resultados de consulta de base de datos precalculados potencialmente afectados tan pronto como sea posible.
7. Método de acuerdo con cualquiera de las reivindicaciones 1 a 6, que comprende adicionalmente:
- para los eventos en tiempo real que se determinan como que están representados en el modelo probabilístico, acumular tales eventos en tiempo real a lo largo de un determinado periodo de tiempo, comparar los eventos en tiempo real que han tenido lugar y que se han acumulado en la práctica con su representación en el modelo probabilístico y, si los eventos en tiempo real que han tenido lugar y que se han acumulado en la práctica se desvían con respecto a su representación en el modelo probabilístico hasta una medida previamente

determinada, emitir unas órdenes recalculadas con respecto a los resultados de consulta de base de datos precalculados potencialmente afectados tan pronto como sea posible.

- 5 8. Método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde la plataforma de memoria caché de datos (2), cuando se determinan las probabilidades de estar obsoletos de los resultados de consulta de base de datos precalculados y se emite el recálculo, considera unas cuadrículas de los resultados de consulta de base de datos precalculados que se corresponden con grupos de conjuntos adyacentes de datos que se mantienen en la plataforma de cálculo (3).
- 10 9. Método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde la plataforma de memoria caché de datos (2) emite las órdenes recalculadas basándose en la cantidad de recursos de cálculo disponibles en la plataforma de cálculo (3).
- 15 10. Método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde el sistema de base de datos distribuido (1) es un sistema de reservas de viajes, en donde la plataforma de cálculo (3) mantiene una información acerca de la disponibilidad de viajes y las tarifas y la plataforma de memoria caché de datos (2) mantiene unas recomendaciones de viajes con precio que se calculan a partir de la información de disponibilidad de viajes y las tarifas.
- 20 11. Método de acuerdo con la reivindicación 10, en donde los eventos en tiempo real comprenden cambios de tarifas de vuelo, cambios de disponibilidad de asientos de avión, solicitudes de billetes de vuelo de clientes y/o cancelaciones de vuelos.
- 25 12. Método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde el sistema de base de datos distribuido (1) comprende al menos una plataforma de aplicaciones (4) que está conectada con la plataforma de cálculo (3), manteniendo y organizando la al menos una plataforma de aplicaciones (4) los resultados de consulta de base de datos precalculados, siendo rellenados y / o actualizados los resultados de consulta de base de datos que están almacenados en la al menos una plataforma de aplicaciones (4) por la plataforma de cálculo (3) como un resultado de las órdenes recalculadas que son emitidas por la plataforma de memoria caché de datos (2).
- 30 13. Plataforma de memoria caché de datos (2) que mantiene unos resultados de consulta de base de datos precalculados que son calculados por una plataforma de cálculo (3) basándose en datos que se mantienen en la plataforma de cálculo (3), estando configurada la plataforma de memoria caché de datos (2) para:
- 35 - mantener un modelo probabilístico que modela las discrepancias entre los resultados de consulta de base de datos precalculados que se mantienen en la plataforma de memoria caché de datos (2) y los resultados de consulta de base de datos reales supuestos,
 - determinar las probabilidades de estar obsoletos de los resultados de consulta de base de datos precalculados basándose en el modelo probabilístico,
- 40 - detectar un evento en tiempo real asíncrono que tiene una influencia probabilística sobre las discrepancias entre los resultados de consulta de base de datos precalculados que se mantienen en la plataforma de memoria caché de datos (2) y los resultados de consulta de base de datos reales supuestos;
 - analizar si el evento en tiempo real asíncrono está representado en el modelo probabilístico;
- 45 - aumentar las probabilidades determinadas de estar obsoletos de los resultados de consulta de base de datos precalculados si se determina que el evento en tiempo real asíncrono no está representado en el modelo probabilístico;
 - emitir de forma automática unas órdenes recalculadas a la plataforma de cálculo (3) para actualizar unos resultados de consulta de base de datos precalculados en función de las probabilidades determinadas de estar obsoletos de los resultados de consulta de base de datos precalculados, en donde se ordena que se recalculen
- 50 los resultados de consulta de base de datos precalculados que tienen una probabilidad de estar obsoletos por encima de un umbral dado; y
 - recibir los resultados de consulta de base de datos precalculados actualizados como resultados de las órdenes recalculadas.
- 55 14. Plataforma de memoria caché de datos (2) de acuerdo con la reivindicación 13, que está configurada adicionalmente para
- 60 - para los eventos en tiempo real que no están representados en el modelo probabilístico, emitir unas órdenes recalculadas con respecto a los resultados de consulta de base de datos precalculados particulares respectivos tan pronto como sea posible;
 - para los eventos en tiempo real que están representados en el modelo probabilístico, acumular tales eventos en tiempo real a lo largo de un determinado periodo de tiempo, comparar los eventos en tiempo real que han tenido lugar y que se han acumulado en la práctica con su representación en el modelo probabilístico y, si los eventos en tiempo real que han tenido lugar y que se han acumulado en la práctica se desvían con respecto a su
- 65 representación en el modelo probabilístico hasta una medida previamente determinada, emitir unas órdenes recalculadas con respecto a los resultados de consulta de base de datos precalculados potencialmente afectados

tan pronto como sea posible.

15. Un programa informático que, cuando se ejecuta en un sistema informático, da lugar a que el sistema informático (2):

- 5
- mantenga un modelo probabilístico que modela las discrepancias entre los resultados de consulta de base de datos precalculados que se mantienen en el sistema informático (2) y los resultados de consulta de base de datos reales supuestos;
- 10
- determine las probabilidades de estar obsoletos de los resultados de consulta de base de datos precalculados basándose en el modelo probabilístico;
 - detecte un evento en tiempo real asíncrono que tiene una influencia probabilística sobre las discrepancias entre los resultados de consulta de base de datos precalculados que se mantienen en el sistema informático (2) y los resultados de consulta de base de datos reales supuestos;
- 15
- analice si el evento en tiempo real asíncrono está representado en el modelo probabilístico;
 - aumente las probabilidades determinadas de estar obsoletos de los resultados de consulta de base de datos precalculados si se determina que el evento en tiempo real asíncrono no está representado en el modelo probabilístico;
- 20
- emita de forma automática unas órdenes recalculadas para actualizar unos resultados de consulta de base de datos precalculados en función de las probabilidades determinadas de estar obsoletos de los resultados de consulta de base de datos precalculados, en donde se ordena que se recalculen los resultados de consulta de base de datos precalculados que tienen una probabilidad de estar obsoletos por encima de un umbral dado; y
 - reciba los resultados de consulta de base de datos precalculados actualizados como resultados de las órdenes recalculadas.

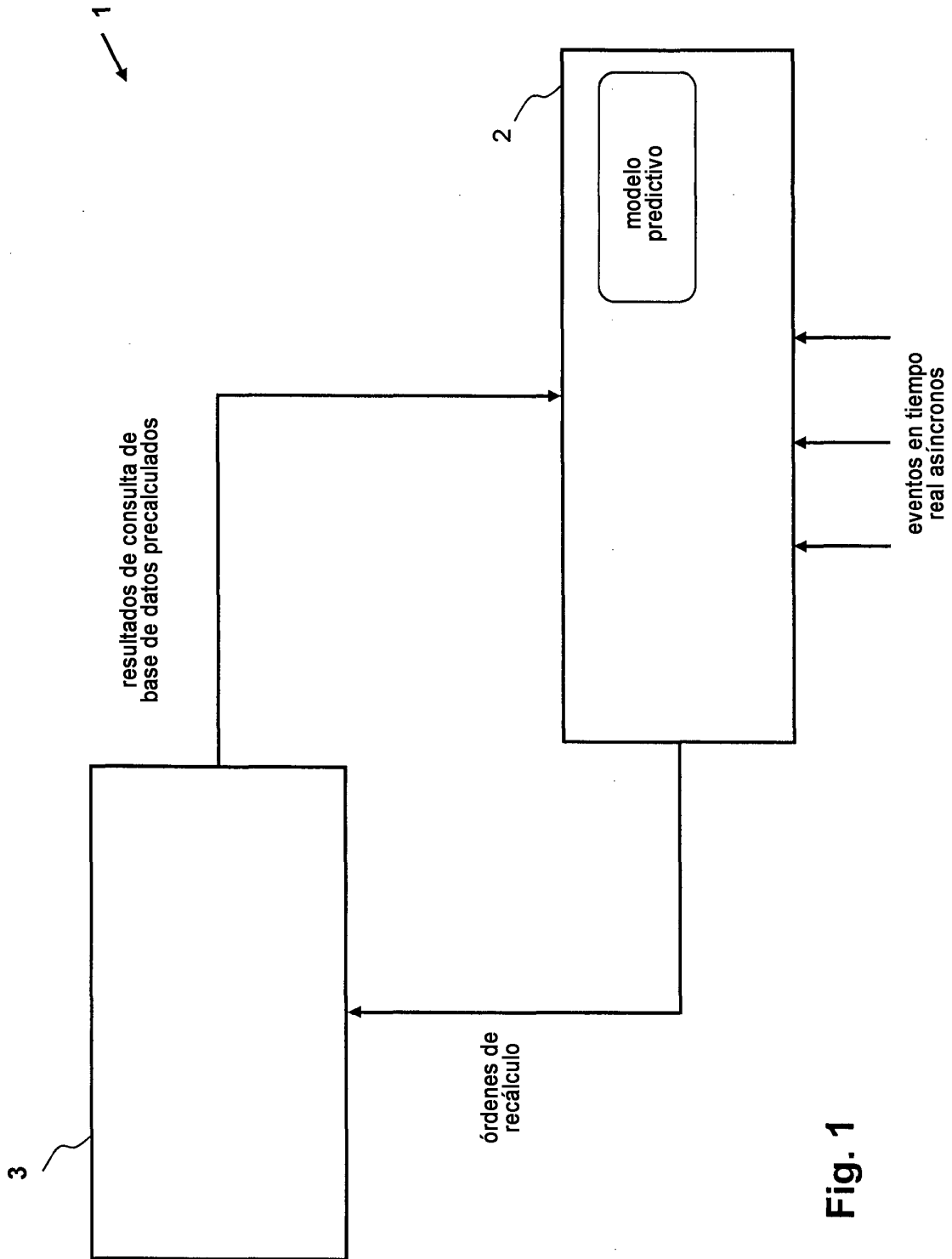


Fig. 1

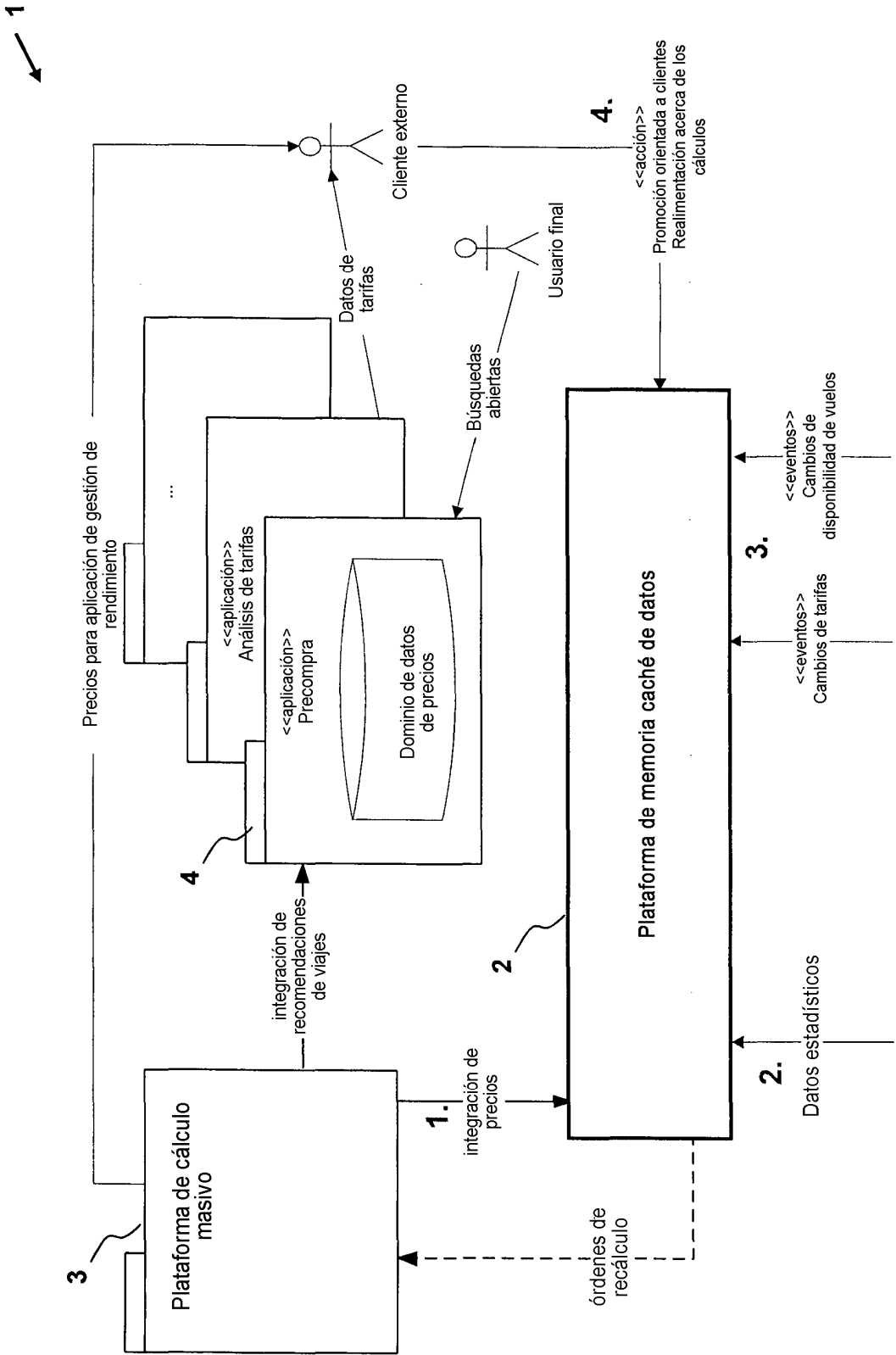


Fig. 2

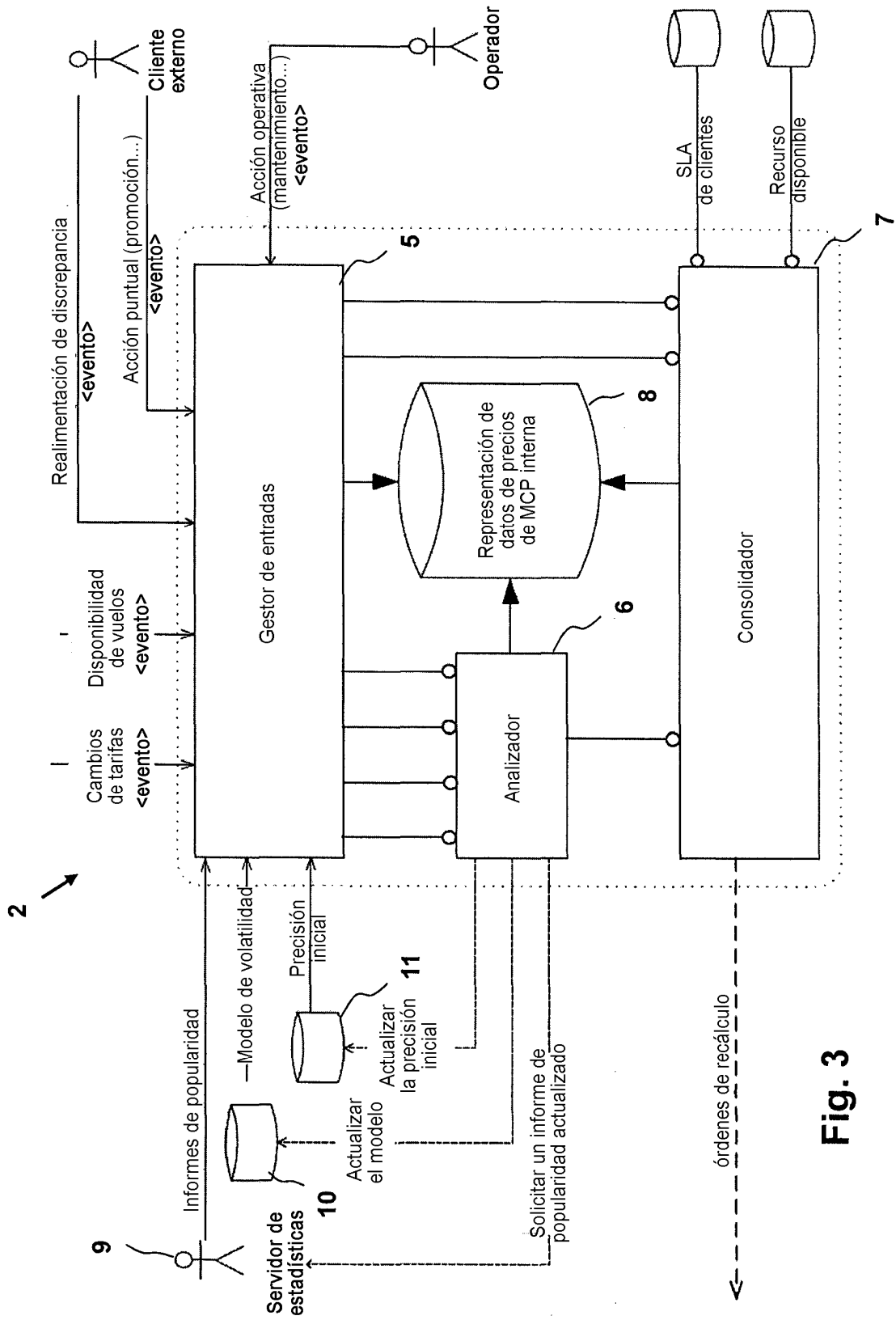


Fig. 3

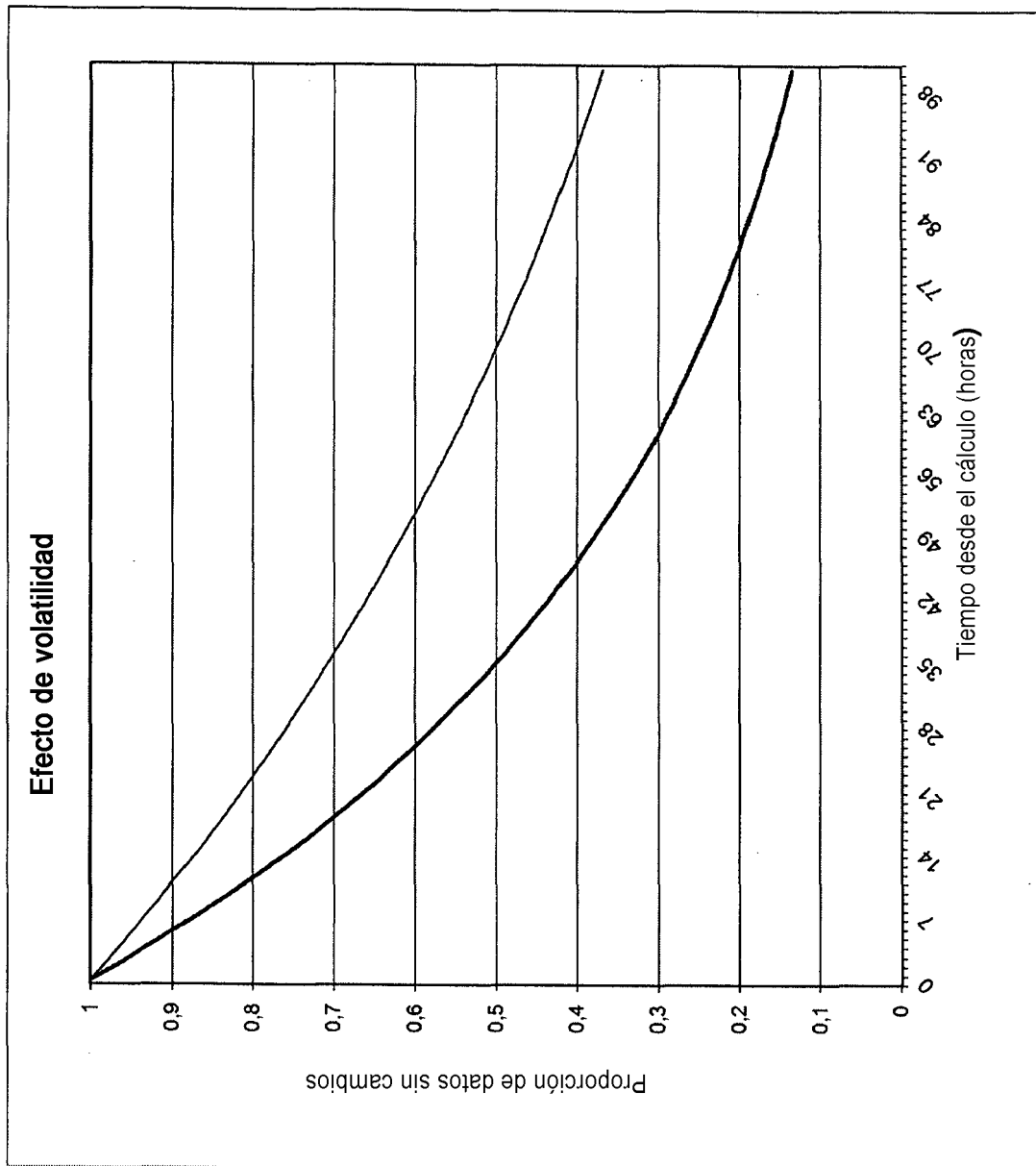


Fig. 4a

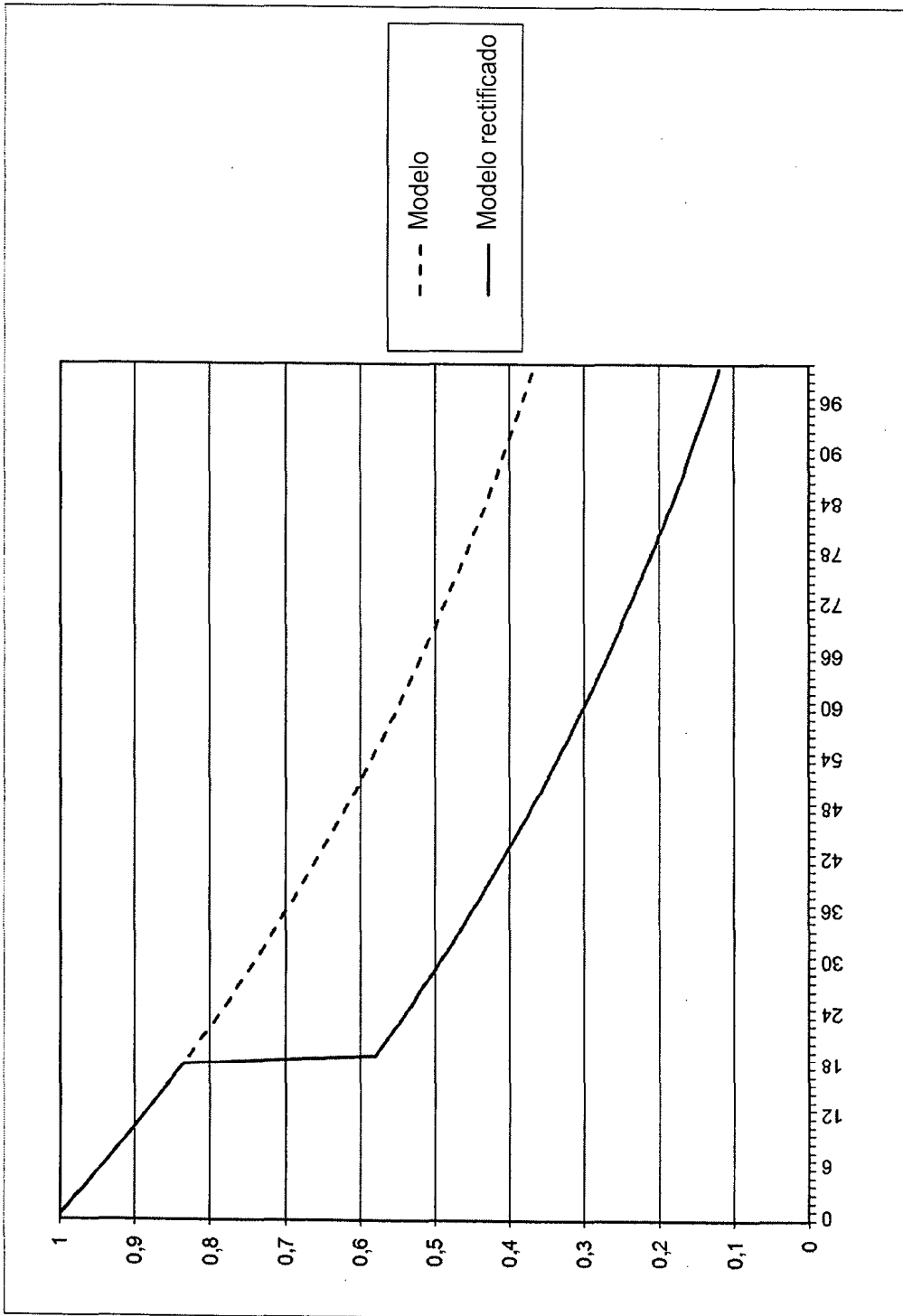


Fig. 4b

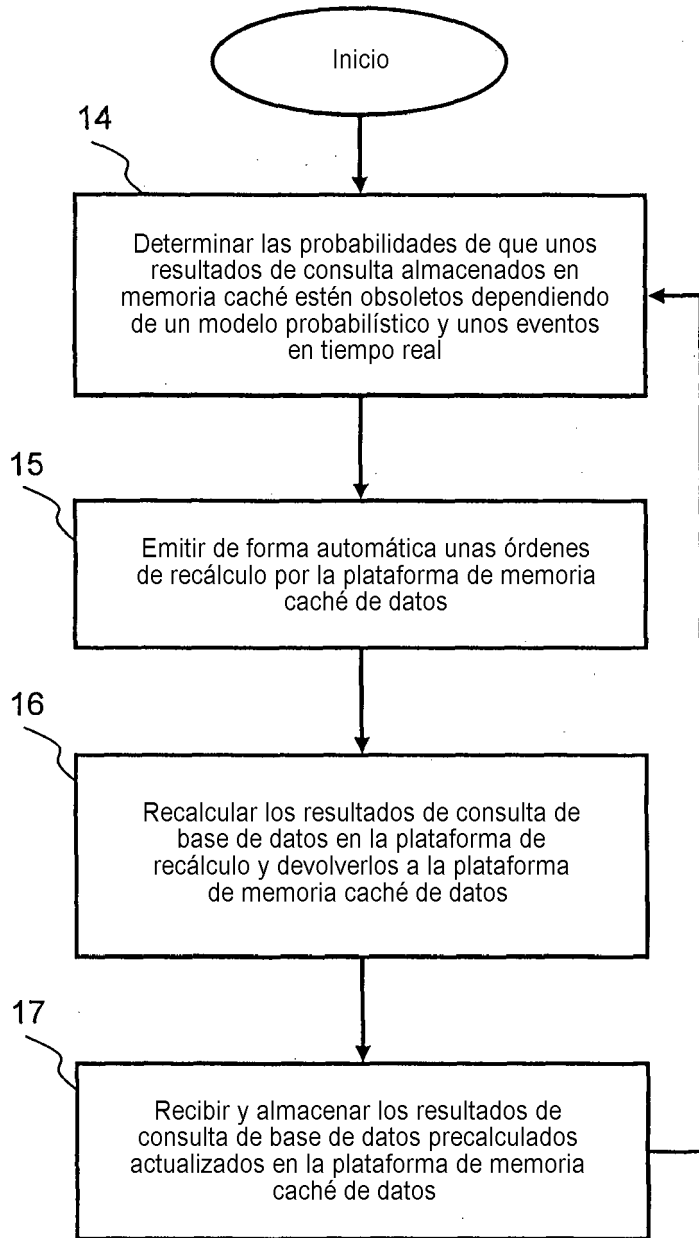


Fig. 5

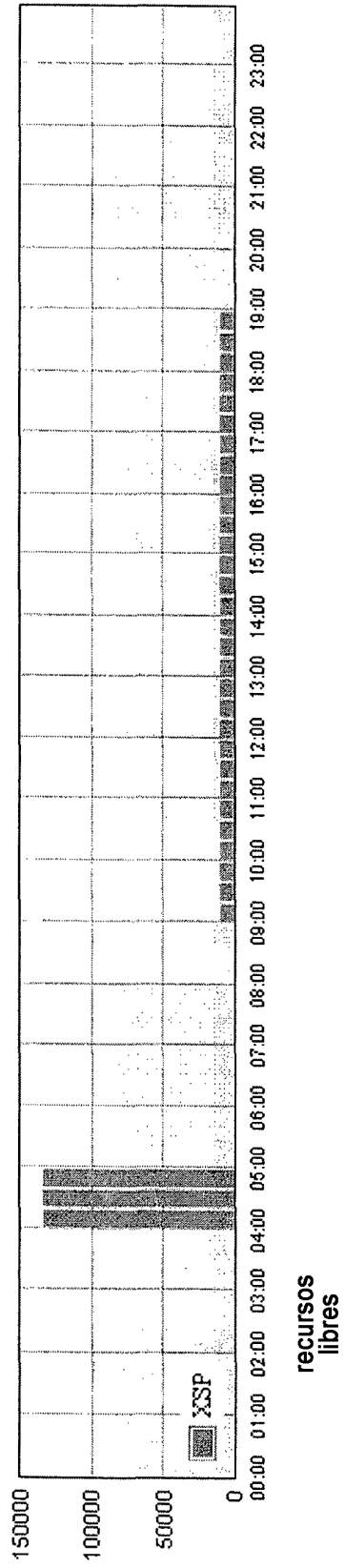


Fig. 6

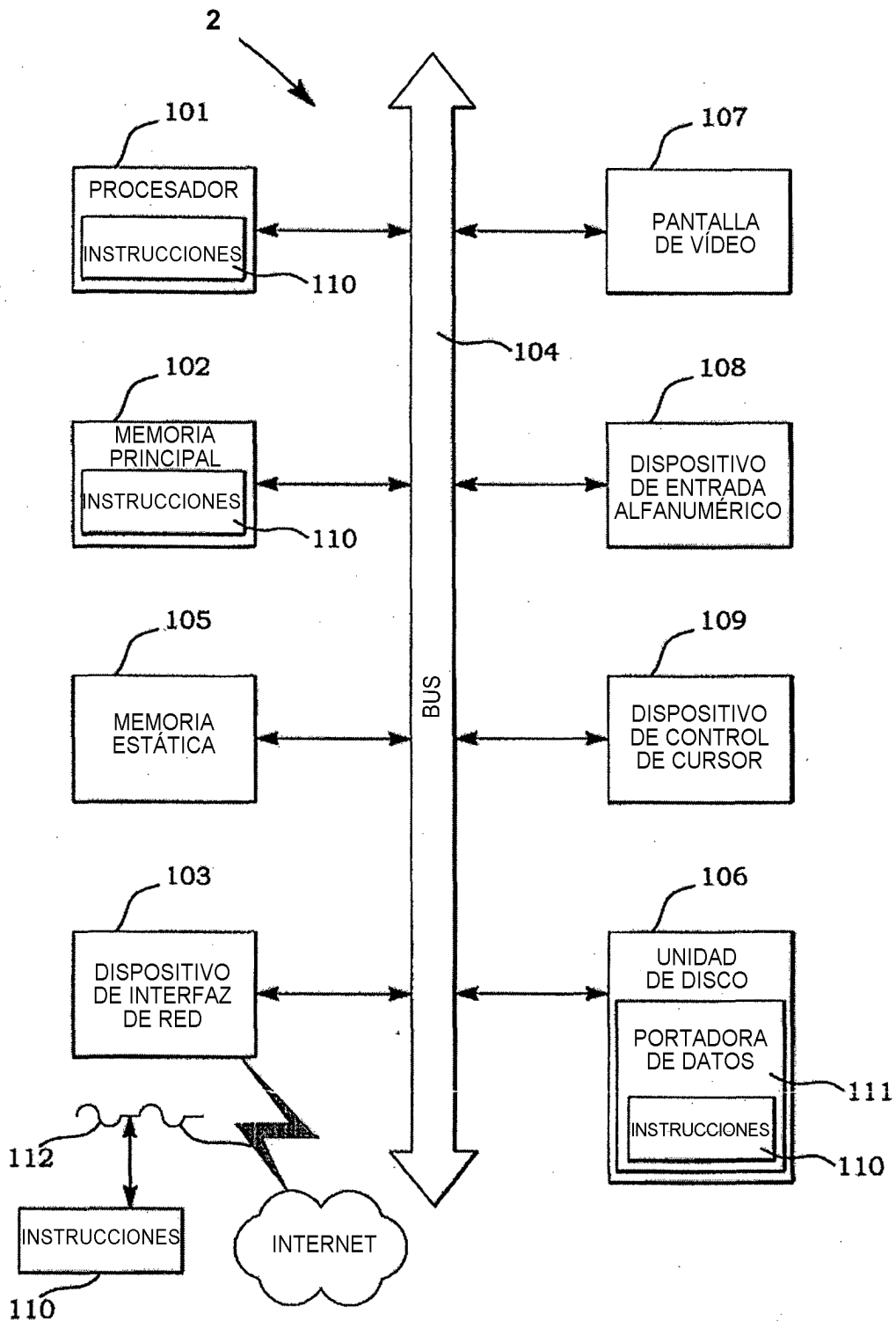


Fig. 7