

REIVINDICACIONES

1. Un método implementado por computadora para el procesamiento a gran escala de registros de datos, que comprende:

5 recibir registros de datos sin procesar que incluyen un primer tipo de registros de datos que tienen valores estáticos y un segundo tipo de registros de datos relacionados con registros transaccionales con marcas de tiempo de eventos;

filtrar y transformar los registros de datos en formatos estandarizados;

10 hacer persistir los registros de datos transformados en dos almacenamientos de datos diferentes, que incluyen un primer almacenamiento de datos para datos estáticos relacionados con entidades para su recuperación con base en valores estáticos que incluyen un identificador único de una entidad, y un segundo almacenamiento de datos para datos transaccionales para su recuperación con base en marcas de tiempo de eventos;

15 agrupar diferentes categorías de registros transaccionales de acuerdo con los identificadores únicos de entidades y hacer persistir las categorías a diferentes tablas que incluye mantener una tabla de tamaño estable para el acceso eficiente de una categoría de registros transaccionales a la que se accede con mayor frecuencia; y

recuperar características relacionadas con una entidad de múltiples tablas para su procesamiento para proporcionar una salida relacionada con la entidad.

20 2. El método de acuerdo con la reivindicación 1, en donde la recuperación de características relacionadas con una entidad incluye generación de características con dimensiones de comportamiento de la entidad en un punto de referencia dado en el tiempo, e incluye:

recuperar características de diferentes tablas;

unir las características; y

25 transformar las características en el punto de referencia en el tiempo.

3. El método de acuerdo con la reivindicación 1, en donde el segundo almacenamiento de datos se divide por períodos de tiempo y/o por tipo de evento.

4. El método de acuerdo con la reivindicación 1, en donde la categoría de registros transaccionales a la que se accede con mayor frecuencia son transacciones abiertas que están en curso.

30 5. El método de acuerdo con la reivindicación 1: procesar primero los registros de datos estáticos para obtener información de identificación de las entidades.

6. El método de acuerdo con la reivindicación 1, en donde los registros de datos sin procesar se refieren a telecomunicaciones y las entidades son usuarios, y en donde el primer tipo de registros de datos son registros de datos de usuario con valores estáticos de atributos de usuario y el segundo tipo de registros de datos son registros de detalle de llamada de eventos.

35 7. El método de acuerdo con la reivindicación 6, que incluye generar un identificador único para cada usuario como una combinación de un Número de Directorio de Abonado Internacional de Estación Móvil (MSISDN, por sus siglas en inglés) y una fecha de activación del Módulo de Identidad de Abonado

(SIM, por sus siglas en inglés).

8. El método de acuerdo con la reivindicación 6, en donde la filtración y transformación de los registros de datos a formatos estandarizados incluyen aplicar reglas de filtrado para extraer registros relevantes de registros transaccionales que incluyen uno o más del grupo de: registros relacionados con eventos que incluyen: llamadas, mensajes, eventos monetarios, eventos de uso de datos; eventos de ciclo de vida; eventos de uso anticipado; eventos de préstamo; eventos de billetera móvil.

9. El método de acuerdo con la reivindicación 6, que incluye la agrupación de diferentes categorías de registros transaccionales de acuerdo con los identificadores únicos de una entidad incluye categorías de anticipos de créditos abiertos y anticipos de créditos cerrados, y una categoría de registros transaccionales a los que se accede con mayor frecuencia son los anticipos de créditos abiertos disponibles para su análisis posterior.

10. El método de acuerdo con la reivindicación 1, en donde el producto relacionado con una entidad es un perfil de usuario para análisis de riesgo de crédito posterior.

11. Un sistema para el procesamiento a gran escala de registros de datos, el sistema que incluye una memoria para almacenar código de programa leíble por computadora y un procesador para ejecutar el código de programa leíble por computadora, el sistema que comprende:

un componente de recepción de datos para recibir datos sin procesar que incluyen un primer tipo de registros de datos que tienen valores estáticos y un segundo tipo de registros de datos relacionados con registros transaccionales con marcas de tiempo de eventos.

un componente de transformación para transformar registros de datos en formatos estandarizados; un componente de filtrado para filtrar los registros de datos relevantes;

un componente de persistencia de almacenamiento de datos para hacer persistir los registros de datos transformados en dos almacenamientos de datos diferentes, que incluye un primer almacenamiento de datos para datos estáticos relacionados con entidades para su recuperación con base en valores estáticos que incluyen un identificador único de una entidad, y un segundo almacenamiento de datos para datos transaccionales para su recuperación con base en marcas de tiempo de eventos;

un componente de agrupación para agrupar diferentes categorías de registros transaccionales de acuerdo con los identificadores únicos de las entidades y un componente de persistencia de tabla para hacer persistir las categorías a diferentes tablas, que incluye el mantenimiento de una tabla de tamaño estable para el acceso eficiente de una categoría de registros transaccionales a la que se accede con mayor frecuencia; y

un componente de perfil de entidad para recuperar características relacionadas con una entidad de múltiples tablas para su procesamiento para proporcionar un resultado relacionado con la entidad.

12. El sistema de acuerdo con la reivindicación 11, en donde el componente de perfil de entidad para recuperar características relacionadas con una entidad incluye generación de características con dimensiones de comportamiento de la entidad en un punto de referencia dado en el tiempo, que incluye:

un componente de recuperación de características para recuperar características de diferentes tablas;

un componente de unión de características para unir las características; y
 un componente de transformación de características para transformar las características en el punto de referencia en el tiempo.

13. El sistema de acuerdo con la reivindicación 11, en donde el segundo almacenamiento de datos se divide por períodos de tiempo y/o por tipo de evento.

14. El sistema de acuerdo con la reivindicación 11, en donde la categoría de registros transaccionales a la que se accede con mayor frecuencia son transacciones abiertas que están en curso.

15. El sistema de acuerdo con la reivindicación 11, en donde los registros de datos sin procesar se refieren a telecomunicaciones y las entidades son usuarios, y en donde el primer tipo de registros de datos son registros de datos de usuario con valores estáticos de atributos de usuario y el segundo tipo de registros de datos son registros de detalle de llamada de eventos.

16. El sistema de acuerdo con la reivindicación 15, que incluye un componente de identificador único para generar un identificador único para cada usuario como una combinación de un Número de Directorio de Abonado Internacional de Estación Móvil (MSISDN) y una fecha de activación del Módulo de Identidad de Abonado (SIM).

17. El sistema de acuerdo con la reivindicación 15, en donde el componente de filtrado aplica reglas de filtrado para extraer registros relevantes de registros transaccionales que incluyen uno o más del grupo de: registros relacionados con eventos que incluyen: llamadas, mensajes, eventos monetarios, eventos de uso de datos; eventos de ciclo de vida; eventos de uso anticipado; eventos de préstamo; eventos de billetera móvil.

18. El sistema de acuerdo con la reivindicación 15, en donde el componente de agrupación incluye categorías de agrupación de anticipos de créditos abiertos y anticipos de créditos cerrados, y una categoría de registros transaccionales a las que se accede con mayor frecuencia son anticipos de créditos abiertos disponibles para su análisis posterior.

19. Un producto de programa de computadora para el procesamiento a gran escala de registros de datos, el producto de programa de computadora que comprende un medio leíble por computadora que tiene un código de programa leíble por computadora almacenado para realizar los pasos de:

recibir registros de datos sin procesar que incluyen un primer tipo de registros de datos que tienen valores estáticos y un segundo tipo de registros de datos relacionados con registros transaccionales con marcas de tiempo de eventos;

filtrar y transformar los registros de datos en formatos estandarizados;

hacer persistir los registros de datos transformados en dos almacenamientos de datos diferentes, que incluyen un primer almacenamiento de datos para datos estáticos relacionados con entidades para su recuperación con base en valores estáticos que incluyen un identificador único de una entidad y un segundo almacenamiento de datos para datos transaccionales para su recuperación con base en marcas de tiempo de eventos;

agrupar diferentes categorías de registros transaccionales de acuerdo con los identificadores únicos de entidades y hacer persistir las categorías a diferentes tablas que incluye mantener una tabla de

tamaño estable para el acceso eficiente de una categoría de registros transaccionales a la que se accede con mayor frecuencia; y

recuperar características relacionadas con una entidad de múltiples tablas para su procesamiento para proporcionar una salida relacionada con la entidad.

20. Un método implementado por computadora para el procesamiento a gran escala de registros de datos de telecomunicación, que comprende:

recibir registros de datos sin procesar que incluyen un primer tipo de registros de datos que tienen valores estáticos relacionados con una entidad de usuario y un segundo tipo de registros de datos relacionados con registros transaccionales de eventos de telecomunicación con marcas de tiempo de eventos;

filtrar y transformar los registros de datos en formatos estandarizados;

hacer persistir los registros de datos transformados en dos almacenamientos de datos diferentes, que incluyen un primer almacenamiento de datos para datos estáticos relacionados con entidades para su recuperación con base en valores estáticos que incluyen un identificador único de una entidad, y un segundo almacenamiento de datos para datos transaccionales para su recuperación con base en marcas de tiempo de eventos;

agrupar diferentes categorías de registros transaccionales de acuerdo con los identificadores únicos de las entidades y hacer persistir las categorías a diferentes tablas, que incluye el mantenimiento de una tabla de tamaño estable para el acceso eficiente de una categoría de registros transaccionales a la que se accede con mayor frecuencia en relación con anticipos de créditos abiertos a una entidad; y

recuperar características relacionadas con una entidad de múltiples tablas para su procesamiento para proporcionar una salida en la forma de una calificación crediticia relacionada con la entidad.