

Diseño de almacén de datos para el análisis eficiente de la información de incidentes informáticos y mantenimientos.

Ing. Corso Cynthia, Ing. Luque Claudio, Ing. Ciceri Leonardo, Sr Donnet Matías

Grupo de Investigación, Desarrollo y Transferencia de Sistemas de Información
Facultad Regional Córdoba
Universidad Tecnológica Nacional
Maestro M. López esq. Cruz Roja.
cynthia@bbs.frc.utn.edu.ar/cluque@prominente.com.ar/leocic@bbs.frc.utn.edu.ar/55017@sistemas.frc.utn.edu.ar

Resumen

El objetivo de este proyecto de investigación es la caracterización de incidentes informáticos y mantenimientos en el contexto de un laboratorio informático. Para ello se propone la utilización en conjunto de herramientas de la Inteligencia de Negocios y la Minería de Datos, con el propósito de analizar de manera exhaustiva los incidentes informáticos y mantenimientos. En este trabajo se describe un modelo de Data Mart para la determinación de perfiles de los incidentes informáticos y mantenimientos con el fin de instrumentar estrategias que permitan disminuir los niveles de incidencias, como así también optimizar la gestión de mantenimientos.

Palabras clave: Data Mart, Inteligencia de Negocios, Incidente Informático, Mantenimiento

Contexto

Este trabajo presenta el resultado de los avances del proyecto de investigación Generación de Modelo Descriptivo para la prevención de incidentes de equipos

informáticos en el contexto de laboratorio de sistemas” , cuyo código es UTN 1683 homologado por la Secretaría de Ciencia, Tecnología y Posgrado de la Universidad Tecnológica Nacional.

Actualmente este proyecto forma parte del grupo de I&D GIDSI (Grupo de Investigación, Desarrollo y Transferencia de Sistemas de Información) radicado en la UTN-Facultad Regional Córdoba.

Introducción

Uno de los pilares fundamentales para determinar la confiabilidad de los recursos informáticos es el uso de indicadores de carácter estadístico. El cálculo de muchos de ellos requiere información histórica de los incidentes informáticos presentados y que en la actualidad la mayoría de las organizaciones registra y almacena en grandes bases de datos. Gracias al crecimiento tecnológico, la capacidad de almacenamiento de los dispositivos informáticos se ha incrementado de manera significativa, lo que hace que el volumen de información que manejan las organizaciones complejiza el proceso de análisis e interpretación de los datos.

Surge entonces, la necesidad de implementar un mecanismo capaz de

determinar las características propias de los incidentes informáticos, analizando la existencia de posibles relaciones de comportamiento que permita la definición clara de ciertos perfiles de incidentes informáticos. El estudio de incidentes es interesante complementarlo con el análisis de los mantenimientos. [1] La posibilidad de poder clasificar los mantenimientos correctivos realizados a los equipos informáticos posibilita la creación de planes especiales de carácter preventivo aumentando la disponibilidad del servicio prestados por estos.

Por este motivo resulta interesante disponer de una serie de mecanismos que permitan el acceso de una serie de técnicas de pre procesamiento, análisis de los datos en línea y generación de modelos de los datos contenidos en grandes almacenes de datos y que sea un complemento para la gestión eficiente de incidentes informáticos y mantenimiento. La Inteligencia de Negocios en conjunto con la Minería de Datos resultan ser un complemento ideal permitiendo y facilitando el análisis de los datos y la detección de tendencias y asociaciones entre variables que son interesantes para un estudio exhaustivo e integral de los incidentes informáticos y mantenimiento. Para dar soporte a un gran volumen de información se necesita del diseño e implementación de un almacén de datos.

[2] La definición clásica de un almacén de datos fue acuñada por Immon como una colección de datos históricos, orientados por temas, no volátiles, integrados, diseñados para apoyar el proceso de toma de decisiones de una organización.

Los almacenes de datos se clasifican en dos tipos: Corporativos (Data Warehouse) y Departamentales (Data Mart). Este trabajo hace foco en los almacenes de datos de tipo Departamental comúnmente conocidos como Data Mart. [3] Un Data

Mart “se refiere a una vista del Data Warehouse orientada solo a un aspecto de la organización.” Contiene menos volumen de datos que un Data Warehouse y es el objeto del procesamiento analítico por parte del usuario final.

La implementación de un Data Mart no es una condición excluyente para la posterior aplicación de técnicas de Minería de Datos. [4] Sin embargo, el mismo brinda una estructura óptima de datos que facilita no solo el análisis de la información de manera detallada sino que también un contexto apropiado para la generación de modelos de conocimiento.

La Minería de Datos es un proceso de descubrimiento de nuevas relaciones y tendencias significativas en grandes volúmenes de datos.

En este trabajo se expone el diseño conceptual del Data Mart para el tratamiento integral de incidentes informáticos y mantenimientos, guiados bajo los lineamientos de una determinada metodología.

El proceso de desarrollo de un almacén de datos es diferente al desarrollo de sistemas operacionales clásicos, es por ello que existen diversas metodologías entre las más conocidas son la propuesta por Ralph Kimball y Bill Inmon. Para el desarrollo del Data Mart la metodología seleccionada fue la de Kimball que describe un flujo de tareas de alto nivel que son consideradas necesarias para el diseño, desarrollo e implementación de un almacén de datos.

El diseño del Data Mart está condicionado por la identificación de los requisitos de negocio. Estos mismos han sido relevados mediante el establecimiento de entrevistas y reuniones dirigidas al personal directivo y de las áreas operativas involucradas. El propósito de las mismas es la detección de información estratégica, operativa e indicadores que el Data Mart debe

responder. A continuación se enumeran algunos de los principales indicadores detectados:

- Número de incidentes informáticos a través del tiempo.
- Número de incidentes informáticos por aula y año.
- Reincidencia de incidentes por equipo informático a través del tiempo.
- Reincidencia de incidentes informáticos por equipo informático y componente a través del tiempo.
- Número de incidentes informáticos reportados que no fueron incidentes por área de reporte y tipo de diagnóstico.
- Número de mantenimientos por equipo informático, por componente de hardware a través del tiempo.
- Tiempo transcurrido entre la realización del mantenimiento y el reporte de incidencia informático.

Uno de los ejes principales en la propuesta de Ralph Kimball es el Modelo de Datos Dimensional.

Una de las principales motivaciones que da origen a la creación un nuevo modelo de datos radica en las carencias sobre el bajo rendimiento de las consultas en el modelo relacional, que fundamentalmente está orientado a transacciones que manejan pocos registros simultáneamente. [3] Otra de las limitaciones es la propia estructura de la base de datos, las consultas realizadas en los DSS (Sistema de Soporte de Decisión) son muy complejas y su definición no está fijada de antemano.

En el modelo de datos multidimensional, los datos se organizan en torno a los conceptos de la empresa y la estructura de datos manejada en este modelo son matrices multidimensionales o cubos. Un cubo consiste en un conjunto de celdas, de tal manera que cada una está identificada por la combinación de los miembros de las diferentes dimensiones y

contiene el valor de la medida analizada para dicha combinación de dimensiones.

Las variables o medidas son aquellas características del negocio que pueden ser cuantificadas y son seleccionadas para el análisis. Por ejemplo: ventas, compras, costos entre otros se corresponden con los datos numéricos.

Las dimensiones se definen como los atributos categóricos que caracterizan a una variable. El tiempo se considera una dimensión más del modelo, puesto que los análisis de las cifras de un negocio se especifican para un periodo de tiempo concreto.

Tomando como base la tabla de indicadores relacionados con los Incidentes Informáticos y Mantenimientos, se elabora el Modelo Conceptual del Data Mart que finalmente se traducirá en tablas de Dimensiones y Hechos. Para lograr este modelo se cumplimentó los siguientes pasos:

Establecimiento de nivel de granularidad:

En esta fase del proceso se trató de establecer cuál ha de ser el grado de detalle de las celdas que constituirán el cubo OLAP (*On-Line Analytical Processing*). En una primera aproximación se pretende establecer una granularidad que permita el nivel de detalle más preciso que los datos puedan ofrecer y de hecho esto es un principio a tener en cuenta para un buen diseño. Un exceso de detalle puede provocar que el sistema tenga que manejar una cantidad de datos que bien sea excesiva desde un principio, o bien crezca de una manera que provoque que las consultas se vuelvan demasiado lentas y poco operativas.

En este caso de estudio tanto los incidentes informáticos como los mantenimientos están registrados por día y hora. En un primer análisis de las consultas requeridas pueden ser llevadas a cabo con la granularidad original de los

datos tal como son facilitados por el laboratorio informático.

Selección de Dimensiones: una tabla de dimensión permite almacenar detalles acerca de los hechos del negocio que se desea analizar y estudiar. Por lo general las columnas de una tabla de dimensión se utilizan para crear informes o para mostrar resultados de consultas.

Las tablas de dimensiones contienen generalmente una llave o clave simple y atributos que la describen.

Identificación de Medidas y Tabla de Hechos: este último paso consistió en identificar las medidas que surgen del proceso de negocio a estudiar. Una medida es un atributo (campo) de una tabla que se desea analizar, agrupando sus datos, usando los criterios de corte conocidos como dimensiones. Las medidas generalmente se vinculan con el nivel de granularidad.

La tabla de hechos se caracteriza por almacenar medidas que son interesantes para nuestro negocio asociadas a los incidentes informáticos y mantenimientos y que serán medio de análisis para las diferentes perspectivas o dimensiones.

Líneas de Investigación, Desarrollo e Innovación

- Herramientas de Inteligencia de Negocios.
- Metodologías para el diseño e implementación de Almacén de Datos.
- Técnicas para la categorización de datos.
- Herramientas OLAP para la implementación y explotación de cubos OLAP.
- Metodología para el tratamiento integral de incidentes y mantenimientos.
- Minería de Datos

Resultados y Objetivos

El objetivo central de esta investigación es aumentar la disponibilidad de los servicios prestados por los equipos informáticos de un laboratorio mediante la optimización los procesos de mantenimiento. Para cumplimentar este objetivo se procede a la caracterización y clasificación de incidentes informáticos y mantenimientos aplicando recursos de Inteligencia de Negocios y la Minería de Datos.

Los resultados alcanzados hasta el momento relacionados con el diseño y desarrollo del Data Mart son:

- Establecimiento de los requisitos del negocio dirigido al Responsable del Laboratorio de Sistemas y del usuario final.
- Identificación de indicadores que para este caso de estudio se definió una codificación para indicar si pertenece al subproceso de Gestión de Incidente Informático o al de Gestión de Mantenimiento.
- Identificación de Fuentes de origen de datos.
- Diseño e implementación de herramienta web para la gestión integral de incidentes informáticos y mantenimientos que facilita la etapa de recolección de datos y su posterior tratamiento y procesamiento.
- Realización de Modelo Conceptual del Data Mart en el que se enfatiza cada una de las dimensiones de análisis y de los indicadores del modelo (hechos). La propuesta de dicho modelo se resume en el siguiente gráfico.

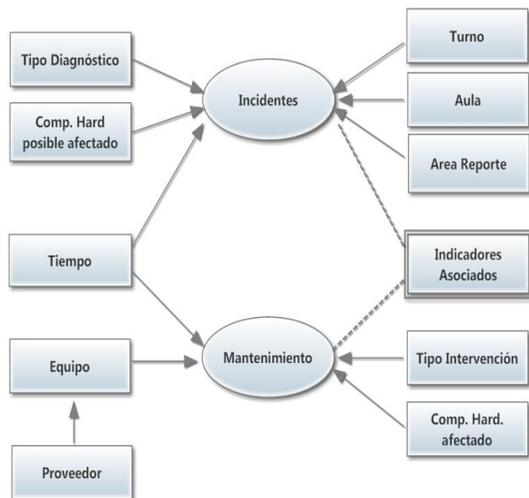


Fig 1. Diseño Conceptual de DM

La diagramación de actividades propuestas a corto plazo son las que se detallan a continuación:

- Análisis y selección de herramienta de Inteligencia de Negocios para la gestión e implementación del Data Mart.
- Diseño del Modelo Físico del Data Mart, en la que se prevé la especificación de las estructuras necesarias para soportar las necesidades de información.
- Definición de los procesos de extracción, transformación y carga que facilitarán el poblamiento del Data Mart.
- Fase Explotación en este proceso se definirá como se utilizará el Data Mart, especificando reportes y gráficos multidimensionales necesarios para dar respuesta a los indicadores detectados.

Formación de Recursos Humanos

Este proyecto está conformado por docentes-investigadores pertenecientes a la carrera de grado de Ingeniería en Sistemas de Información y un graduado. Todos los integrantes docentes del PID han participado del proceso de categorizaciones en investigación dentro del Programa de Incentivos del MECyT;

así como en la categorización interna que posee la U.T.N.

Una de los integrantes está desarrollando su tesis para la obtención del título Magister en Ingeniería en Sistema de Información, cuya temática está relacionada con la temática del proyecto. Además colaboran de manera activa en este proyecto tres becarios, dos alumnos y un graduado.

Referencias

- [1] Gómez Torrejano Manuel de Jesús, Aplicación de Minería de Datos para describir la ocurrencia de fallas en los equipos y sistemas, 2009.
- [2] Inmon W, Building the Data Warehouse, 2005.
- [3] Di Vasta Concettina, DataWarehouse (Almacenes de Datos). Métodos de Diseño de DataWarehouse, Universidad Central, Venezuela, 2014.
- [4] Yalán Castillo Julio, Palomino Panoira Luis, Implementación de un Data Mart como una solución de Inteligencia de Negocios para el área de logística de T-Impulso, Revista de Ingeniería de Sistemas e Informática, ISSN:1816-3823, 2012.
- [5] Bosquet Isabel Dapena, San Roque Antonio Muñoz, Sistemas de Información orientados a la toma de decisiones: el enfoque multidimensional, Anales de mecánica y electricidad, p18-23, 2005
- [6] Quintana Ramírez María José, Orallo José Hernández, Extracción Automática de conocimiento en Base de Datos e Ingeniería de Software, España, 2005.
- [7] Basilio Serra Araujo, Aprendizaje Automático: conceptos básicos y avanzados. Aspectos prácticos usando software Weka, Pearson Educación, Madrid, 2006.
- [8] Rodríguez Sanz Miguel, Análisis y Diseño de un Data Mart para el

seguimiento académico de alumnos en un entorno universitario, 2010.

[9] Kimball Ralph, The Datawarehouse Staging Toolkit, IDG Books, 2004.