

Bodega de datos con alta capacidad de análisis para el desempeño académico de Universidades

Data warehouse with high capacity for analysis for the academic performance of Universities

Daniel Urrea Pito  Eduardo Cerón Moreno  Martha Mendoza Becerra 
Universidad del Cauca, Colombia

OPEN  ACCESS

Recibido: 22/09/2020

Aceptado: 22/10/2020

Publicado: 23/11/2020

Correspondencia de autores:

durrea@unicauca.edu.co



Copyright 2020
by Investigación e
Innovación en Ingenierías

Resumen

Objetivo: DProponer una bodega de datos con alta capacidad de análisis para el desempeño de estudiantes universitarios; analizando las diferentes situaciones de diseño que pueden ser comunes para instituciones de educación superior, permitiendo obtener modelos más efectivos y facilitando la adaptación de estos. **Metodología:** Se identificaron dimensiones y medidas comunes encontradas en la literatura sobre bodegas de datos relacionados con el desempeño académico de los estudiantes, luego se incluyeron elementos propios de las universidades del país. **Resultados:** Dos modelos dimensionales para el desempeño académico de estudiantes en instituciones de educación superior y las situaciones comunes analizadas en el proceso de diseño de estos. Además, un prototipo de bodega de datos de estos modelos para la Universidad del Cauca y una evaluación basada en la norma ISO/IEC 25022, del Nivel de satisfacción de los usuarios con los modelos implementados. **Conclusiones:** Los resultados de la evaluación mostraron que el 87.72% de los usuarios quedaron satisfechos con los datos brindados por el modelo a través del prototipo, observando la importancia de este tipo de sistemas en las universidades.

Palabras clave: Bodegas de datos, modelado dimensional, analítica de datos, evaluación del desempeño, instituciones educativas.

Abstract

Objective: To propose a data warehouse with high capacity for analysis for the performance of university students; analyzing the different design situations that may be common for higher education institutions, allowing to obtain more effective models and facilitating their adaptation. **Methodology:** Common dimensions and measures found in the literature on data warehouses related to students' academic performance were identified, then elements typical of the country's universities were included. **Results:** Two dimensional models for the academic performance of students in higher education institutions and the common situations analyzed in the design process of these. In addition, a prototype of the data warehouse of these models for the University of Cauca and an evaluation based on the ISO / IEC 25022 standard, of the level of user satisfaction with the models implemented. **Conclusions:** The results of the evaluation showed that 87.72% of the users were satisfied with the data provided by the model through the prototype, observing the importance of this type of system in universities.

Keywords: Data warehouses, dimensional model, data analytics, performance Evaluation, educational institutions.

Como citar (IEEE): D. Urrea-Pito., E. Cerón-Moreno., M. Mendoza-Becerra. "Bodega de datos con alta capacidad de análisis para el desempeño académico de Universidades", vol. 8, n°3, pp. 102-118, 2020. DOI: <https://doi.org/10.17081/invinno.8.3.4707>

Introducción

Las instituciones educativas de nivel superior han presentado un crecimiento significativo en el número de estudiantes, debido a la ampliación de la cobertura del sector educativo. Según el Ministerio de Educación Nacional [1] la cantidad de estudiantes matriculados en el año 2003 fue de 470.532 y en 2015 de 1.167.888, lo cual implica un incremento del 148%. El crecimiento de esta población conlleva un aumento de los datos personales y académicos de los estudiantes, los cuales son almacenados en diferentes fuentes de información, por lo general, bases de datos y archivos EXCEL.

Estas instituciones necesitan mejorar la capacidad de análisis con respecto al rendimiento académico de los estudiantes para caracterizar la población y generar estrategias que le permitan mejorar las condiciones y el desempeño académico de los mismos. Sin embargo, hasta ahora muchas de estas instituciones no han logrado una consolidación de esta cantidad de datos, en un sistema que favorezca la toma de decisiones y la proyección institucional. Debido a esto, el horizonte de planeación se ha limitado a períodos de tiempo cortos, generando así un gran inconveniente para estas instituciones [2].

En este sentido, las bodegas de datos permiten consolidar diferentes fuentes de datos con el objetivo de facilitar las tareas de consulta y análisis, sin embargo, estas han sido mayoritariamente usadas por el sector financiero y de ventas, generando impacto en la relación existente con el cliente y permitiendo la identificación de los productos más significativos [3]. Aun conociendo los beneficios que brinda una bodega de datos en los procesos de análisis y planificación, no es común su aplicación en el ámbito de la educación superior pública del país, esto por lo general, se debe a que las instituciones del sector público se ven limitadas en su capacidad de adquisición, por lo cual no pueden costear o acceder a aquellas soluciones ofertadas por terceros. Por otro lado, la creación de una solución de bodegas de datos por parte del personal interno de la institución resulta compleja ya que en la mayoría de los casos no están capacitados en el diseño e implementación de este tipo de soluciones.

Uno de los aspectos de mayor dificultad al abordar este tipo de soluciones es la creación del modelo dimensional, esto debido a que su enfoque es diferente al de un modelo relacional, contemplando elementos y relaciones particulares. En la literatura se encuentran trabajos relacionados con bodegas de datos para matrícula académica y registro de notas [4,5,6,7,8,9,10,11,12,13,14]. Sin embargo, la mayoría de estos modelos presentan pocas dimensiones y/o atributos, y poca información socio-demográfica del estudiante, lo cual limita la capacidad de análisis sobre los mismos. También algunos modelos presentan inconsistencias de diseño incluyendo atributos de una dimensión en las tablas de hechos. Además, no analizan las situaciones de diseño comunes presentadas, lo cual permite obtener modelos de diseño más efectivos [3]. Por lo anterior, se dificulta la apropiación de estos modelos por las instituciones de educación superior.

En este artículo se proponen dos modelos dimensionales relacionados con el rendimiento académico de los estudiantes de universidades públicas, que incluyen información para aumentar la capacidad de análisis: socio-demográfica y académica del estudiante, de materias y de docentes; y analiza situaciones de diseño comunes que facilitan la adaptación para estas instituciones. Además, se construye un prototipo de bodega de datos implementado los dos modelos y se realiza una evaluación del nivel de satisfacción de los usuarios con los datos brindados por el modelo a través del prototipo.

Metodología

Para el desarrollo de este trabajo se tomó como guía el Patrón de Investigación Iterativa (PII) propuesto por Pratt [15], el cual ha sido creado para proyectos de investigación que involucran una solución computacional. Esta metodología propone la realización de ciclos compuestos por cuatro etapas: observación, identificación del problema, desarrollo de la solución y prueba de la solución; y al final de cada ciclo se entrega un producto específico.

Para esta investigación se realizó un ciclo, orientado al modelado dimensional del desempeño académico de estudiantes de instituciones de educación, incluyendo el registro de las notas semestrales y de notas parciales.

Observación

En esta etapa se realizó una revisión de la literatura involucrando bodegas de datos orientadas al desempeño académico. Para esta revisión se siguieron algunos elementos de las pautas presentadas por [16] Las actividades realizadas en esta etapa fueron:

Planeación de los siguientes elementos:

- Objetivo de la investigación. Proponer una bodega de datos con alta capacidad de análisis para el desempeño de estudiantes universitarios.
- Recursos. Las bases de datos científicas a usar para esta investigación fueron: Scopus, ScienceDirect, IEEE Xplore Digital Library, Springer Link.
- Pregunta de investigación. Las preguntas que se pretendían resolver con los trabajos encontrados en la literatura fueron:

P1: ¿Cuáles procesos respecto al desempeño académico se han modelado de manera dimensional?

P2: ¿Qué capacidad de análisis presentan estos modelos?

P3: ¿Qué situaciones de diseño se analizan en estos modelos?

- Cadena de búsqueda. Se usando términos claves relevantes para la investigación, para permitir encontrar los trabajos relacionados con el desempeño académico de los estudiantes.

("Data warehouse" OR "Data warehousing" OR "Business Intelligence") AND ("Multidimensional model" OR "Dimensional model") AND (Academic OR University OR Education)

- Procesos de selección de artículos. Se definieron los criterios de inclusión y exclusión de los artículos.

Inclusión: Cualquier autor, publicado en los últimos 5 años y que estuvieran relacionados con el desempeño académico.

Exclusión: No presentaban un modelo dimensional relacionado con desempeño académico o estaban enfocados solo en uso de técnicas de minería.

Aplicación de la cadena de búsqueda.

La cadena se aplicó en las bases de datos científicas previamente definidas sobre los campos: título, resumen y palabras clave. Se encontraron 109 artículos, los cuales fueron revisados en mayor detalle para aplicar los criterios de exclusión, obteniendo como resultado la selección de 11 artículos relevantes para esta investigación.

Estudio y análisis de los artículos relevantes.

A continuación, se presenta el detalle de los artículos encontrados y luego el análisis de estos.

En el año 2014 fue propuesta una bodega de datos para organizaciones académicas [4], que incluye las tablas de hechos (TH): Matrícula, con las dimensiones: curso, estudiante, fecha, tipo de matrícula, ciudad y sin medidas; y Examen, con las dimensiones: curso, estudiante, fecha, curso de enseñanza y con las medidas calificación curso y cumlaude. En este artículo las dos TH presentan pocas dimensiones y la mayoría de las dimensiones tiene muy pocos atributos. La bodega de datos fue implementada pero no presenta una evaluación por parte de los usuarios.

En el 2015, se propone una bodega de datos para determinar características de los estudiantes que influyen en el rendimiento académico [5], que presenta una TH Alumnos con las dimensiones: situación laboral (De la madre, del padre y del estudiante), residencia, procedencia, estudios secundarios, dedicación al estudio, importancia al estudio, importancia a las TICs y como medidas: primera nota, segunda nota, nota extraordinaria, situación final. Aunque este artículo presenta bastante información socio-demográfica del estudiante (género, estado civil, localización) y laboral de los padres; presenta inconsistencias de diseño, ya que la TH almacena información como: género, edad, estado civil, entre otros; que debería estar en una dimensión. Además, menciona notas obtenidas por materias, pero en el modelo no se presenta una dimensión de este tipo ni un atributo que se relacione. Este trabajo no fue implementado y solo contempla la etapa de diseño de la bodega de datos y el proceso ETL.

En el mismo año, se propone monitorear la educación primaria y secundaria por medio de indicadores almacenados en una bodega de datos [6]; presentando una TH de Matrícula con dimensiones que almacenan un solo atributo: año, sede, regional, municipio, estado, escuela, segmento y medidas relacionadas con cantidad de matriculados. Este trabajo fue implementado, pero no fue evaluado por usuarios.

También en el 2015 año, se propone evaluar la educación básica por medio de una bodega de datos de indicadores sobre Matrícula y Notas [7]. El artículo presenta dos TH con las dimensiones: año académico, curso, estudiante, escuela, grado, disciplina, país y como medidas cantidad de matriculados y nota del curso. Sin embargo, ninguna dimensión presenta los atributos. Este trabajo fue implementado, pero no fue evaluado por usuarios.

En el año 2016, proponen una bodega de datos para la seguimiento del estudiante [8], por medio de dos TH con solo dos dimensiones, Reporte por facultad: facultad, semestre y la medida número de estudiantes; y Reporte estudiante: registro académico, curso y medidas de número de matriculados. Y otra TH Seguimiento estudiante con las dimensiones: registro académico, curso, semestre, departamento, colegio, universidad, empleados, información personal del estudiante, resultado primer semestre y una dimensión de resultado por cada semestre (del segundo al décimo) y sin medidas. Sin embargo, ninguna dimensión presenta los atributos. Este trabajo fue implementado, pero no fue evaluado por usuarios.

En el mismo año, se implementó una bodega de datos para analizar la deserción estudiantil [9], proponiendo dos TH, Notas con las dimensiones: programa, tiempo, estudiante, asignatura, docente y como medidas: notas de los diferentes cortes, definitiva, aprobadas y asignaturas cursadas. La otra TH enfocada en la Deserción con las dimensiones: programa, tiempo, estudiante, estado civil, estrato, lugar residencia, promedio estudiantil, sexo y medidas: matriculado, desertor potencial y desertor. En este trabajo solo la dimensión estudiante y tiempo presentan varios atributos, las demás dimensiones solo tienen un atributo descripción. Además, la TH Notas presenta algunas medidas de diferente granularidad de la TH (asignaturas aprobadas, asignaturas cursadas). Esta bodega de datos fue implementada, pero no fue evaluada por usuarios.

En el año 2017, se propone una bodega de datos para Matrícula [10], que presenta una TH con las dimensiones: facultad, campo de estudio, género, grupo, período de entrada y como medida cantidad de estudiantes. Este artículo, no muestra la TH sino una interfaz, por lo tanto, solo algunas dimensiones presentan atributos. Este trabajo fue implementado y evaluado.

En el año 2018, se propone una bodega de datos para el registro de Notas [11], presentando una TH con las dimensiones: docente, curso, fecha, estudiante, sección y como medida la nota. Este trabajo fue implementado, pero no fue evaluado por usuarios.

En el mismo año, para generar estadísticas de información académica se propone una bodega de datos [12], con dos TH relacionadas con desempeño académico, una de ellas con cuatro dimensiones, Promedio: cohorte, período, semestre, género y medida promedio semestre. Y otra TH con tres dimensiones, Nota examen: cohorte, curso, clase y medida nota promedio curso. Sin embargo, de las seis dimensiones cinco solo tienen un atributo y la otra (período) tiene dos atributos. Esta bodega de datos fue implementada y evaluada.

También en 2018, se propone una base de datos del registro académico [13], con tres TH, Estudiante con las dimensiones: estado, programa, padre, provincia, sin medidas y presenta atributos (nombre, dirección, ciudad, lugar y fecha de nacimiento, género, religión, escuela). Índice logro semestre: programa, período y medidas código estudiante, horas créditos, índice logro semestre; Índice logro carrera: programa y medidas código estudiante, total créditos, índice logro carrera. Solo la dimensión padre y programa tienen más de un atributo. En este artículo se presentan inconsistencias de diseño en la TH Estudiante ya que contiene atributos que deberían ser parte de una dimensión estudiante. Este trabajo fue implementado, pero no fue evaluado por usuarios.

Finalmente en el mismo año, se propone una bodega de datos como servicio [14], que contempla una TH para el registro de Notas con cinco dimensiones: universidad, curso, estudiante, semestre, tipo registro y las medidas nota del curso, porcentaje de asistentes. Esta bodega de datos fue implementada y evaluada.

En la Tabla 1 se presenta la comparación de estos trabajos, la primera columna presenta la referencia del artículo y el año de publicación; el nombre del proceso que se modela en la segunda; en la tercera se menciona si el modelo presenta alguna inconsistencia a nivel de diseño o si relaciona pocas dimensiones o atributos, lo cual limita la capacidad de análisis del modelo; en la cuarta columna se muestra la información socio-demográfica del modelo; la quinta indica si el modelo analiza situaciones de diseño (Sit. Dis.); la sexta si el modelo fue implementado (Imp.) y la última si el modelo fue evaluado (Eval.).

Tabla 1. Comparación trabajos relacionados

Ref. Año	Proceso	Inconsistencias de Diseño/ Pocas Dimensiones o Atributos	Inf. Soc.	Sit. Dis.	Imp.	Eval.
[4] 2014	Matrícula, Notas	Pocas dimensiones y Muy pocos atributos.	Localización	No	Si	No
[5] 2015	Alumnos	TH con atributos como una dimensión. No existe dimensión materia.	Localización, Situación laboral	No	No	No
[6] 2015	Matrícula	Dimensiones con un solo atributo.	No	No	Si	No
[7] 2015	Matrícula, Notas	Dimensiones sin atributos.	País	No	Si	No
[8] 2016	Matrícula, Notas	Dimensiones sin atributos.	Información personal del estudiante	No	Si	No
[9] 2016	Notas, Deserción	Dimensiones con pocos atributos. Medidas de diferente granularidad de la TH.	Sexo, Estado civil, Estrato, Localización.	No	Si	No
[10] 2017	Matrícula	Algunas dimensiones no presentan atributos.	Género	No	Si	Si
[11] 2018	Notas	Ninguna	Género, Nacionalidad, Ciudad	No	Si	No
[12] 2018	Notas curso, Promedio semestre	Pocas dimensiones y Muy pocos atributos	Género	No	Si	Si
[13] 2018	Matrícula, Promedio semestre, Promedio carrera	Pocas dimensiones y Muy pocos atributos. TH con atributos como una dimensión.	Provincia, Género, Religión	No	Si	No
[14] 2018	Notas	Pocas dimensiones	Género	No	Si	Si

Fuente: Elaboración Propia.

Como se aprecia en la Tabla 1, se han realizado diferentes investigaciones que apuntan a obtener modelos dimensionales del rendimiento académico de los estudiantes en instituciones de educación. Sin embargo, estos modelos en su mayoría presentan pocas dimensiones y/o atributos, y poca información socio-demográfica del estudiante, lo cual limita la capacidad de análisis de los usuarios sobre el modelo. Además, algunos modelos presentan inconsistencias de diseño y solo tres fueron evaluados por usuarios.

Por otra parte, ninguna investigación analiza las diferentes situaciones de diseño comunes que permiten obtener modelos de diseño más efectivos [3] y facilitan la apropiación por parte de las instituciones educativas. Por ejemplo, el modelo debe contemplar el manejo de información que cambie en el tiempo y sobre la cual es pertinente realizar análisis (como: estado civil, edad, estrato, etc.), sin tener que colocar en el modelo una dimensión por cada uno de estos atributos.

Identificación

En esta etapa se identificaron elementos de los modelos dimensionales de los artículos encontrados, que serían tomados como punto de partida. En la Tabla 2 se muestran las dimensiones y en la Tabla 3 las medidas.

Tabla 2. Dimensiones identificadas en el estado del arte

Dimensiones Relacionadas con	Nombre Dimensiones
Localización	Sede, Regional, Ciudad, Provincia, Región, País.
Tiempo	Año académico, Fecha, Semestre, Período, Cohorte
Programa	Facultad, Departamento, Programa, Campo de estudio, Tipo registro.
Asignatura	Asignatura, Curso, Grupo, Estado.
Personas	Estudiante, Padre, Docente, Empleado.
Datos Socio-Demográficos	Lugar de residencia y procedencia, Nacionalidad, Género, Estado civil, Estrato, Edad, Religión. Escuela, Colegio. Situación laboral estudiante y padres.

Fuente: Elaboración Propia.

Tabla 3. Medidas identificadas en el estado del arte

Medidas
Cantidad de estudiantes, Nota definitiva curso, Notas parciales, Promedio semestre, Promedio curso.

Fuente: Elaboración Propia.

Desarrollo de la solución

Para la construcción de la bodega de datos se siguió la metodología de desarrollo de bodegas de datos para MiPymes (MBD) [17] mendeley : { «formattedCitation» : «[17]», «plainTextFormattedCitation» : «[17]», «previouslyFormattedCitation» : «[17]» }, «properties» : { «noteIndex» : 0 }, «schema» : «https://github.com/citation-style-language/schema/raw/master/csl-citation.json» }, que presenta un enfoque iterativo e incremental, participación del usuario y detalle de las fases. Esta metodología fue seleccionada ya que es adecuada para grupos de trabajo pequeños, facilitando el proceso de desarrollo debido al uso de pocos roles y artefactos. El ciclo de trabajo incluye las siguientes fases:

- **Iniciación.** Esta fase se seleccionó el proceso de desempeño académico para universidades públicas del país.
- **Planeación.** Se estableció como alcance de la bodega de datos el registro de notas de los estudiantes, y se realizó la planeación del proyecto por medio del cronograma de actividades.

- Análisis y diseño. Partiendo de los elementos encontrados en el estado del arte se procedió a enriquecer el modelo con información específica del país, para lo cual se tomó la Universidad del Cauca (Universidad pública colombiana). Para recolectar esta información específica, se elaboraron y aplicaron dos cuestionarios, uno dirigido a los decanos y coordinadores de programa, y otro a los ingenieros de la división de tecnologías de la información encargados del registro académico. Después, se identificaron y priorizaron las necesidades analíticas del proceso de registro de notas. Finalmente, teniendo en cuenta la recolección de requerimientos y los elementos encontrados en el estado del arte se proponen los dos modelos dimensionales, el de notas semestrales y el de notas por componente.
- Desarrollo. Se construyó un prototipo de bodega para los dos modelos dimensionales propuestos. Esta fase se compone de las siguientes subfases:
 - Definición de la arquitectura: se diseñó la arquitectura técnica teniendo en cuenta los requerimientos identificados en las entrevistas, el plan de infraestructura (requisitos de hardware) y un plan de seguridad (acceso a los datos), necesarios para desplegar el prototipo.
 - Desarrollo del BackRoom: se realizó el diseño físico de la bodega de datos relacional (estándares de nombrado, estimación del tamaño de la base de datos, plan de indexación, agregación y particionamiento), la implementación de la bodega de datos y el proceso de extracción, transformación y cargue de los datos (ETL)
 - Desarrollo del FrontRoom: se identificaron y priorizaron los reportes candidatos; se definió la estrategia de navegación y el formato de estos reportes.
 - Integración: se integró el BackRoom y el FrontRoom, se realizaron las pruebas a los datos y se corrigieron las inconsistencias.
 - Despliegue: se realizó el despliegue del prototipo y la capacitación a los usuarios en el uso del prototipo.
- Mantenimiento y crecimiento. En esta fase se realiza un seguimiento a la puesta en producción de la bodega de datos, lo cual no era objeto de esta investigación.
- Gestión del proyecto. Se realizó el seguimiento a la ejecución del proyecto y se controlaron los cambios que se realizaban.

Prueba de la solución

En esta investigación era importante medir el grado de Satisfacción de un usuario con el uso de un producto software (en este caso, la implementación de los modelos) en un contexto de uso específico [18]. En el modelo de calidad de uso de la norma ISO/IEC 25022, la Satisfacción es una característica que permite realizar esta medición [18].

Para hacer esto, se elaboraron dos cuestionarios como se recomienda en [19], uno para decanos de facultad y otro para coordinadores de programa; los cuales permitían evaluar algunos de los reportes estándar y consultas Ad-hoc. Para la calificación que harían los usuarios se utilizó la escala de Likert [20], que permite seleccionar un estado de satisfacción, iniciando con el valor de 5 para Totalmente de acuerdo y finalizando con el valor 1 para Totalmente en desacuerdo.

Resultados

Los resultados de esta investigación son los dos modelos dimensionales de desempeño académico y las situaciones de diseño analizadas. Además, el prototipo realizado y la evaluación de este.

Modelos Dimensionales

Los modelos dimensionales propuestos son el de Notas semestrales y el de Notas por componente, que se explican a continuación.

Modelo dimensional de Notas semestrales

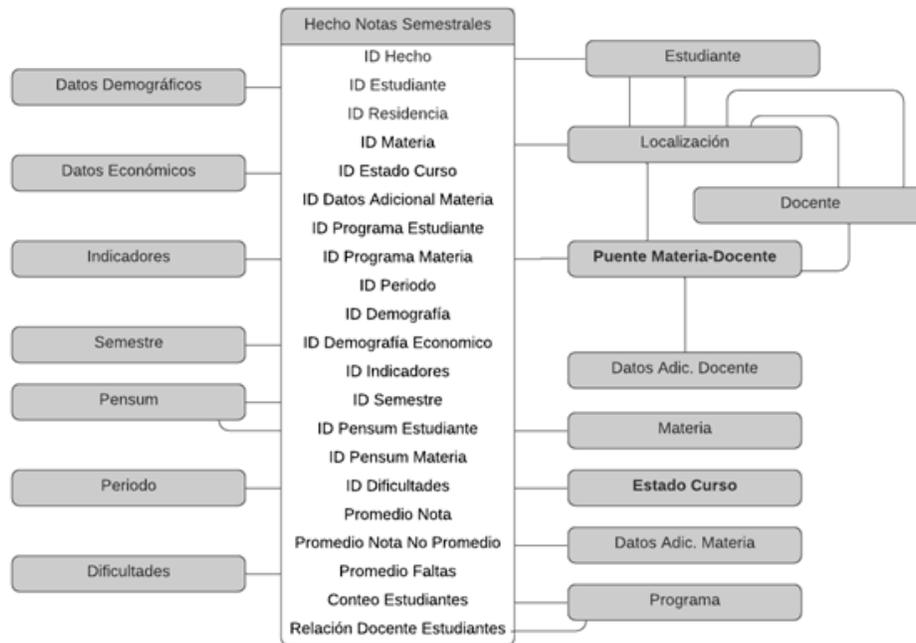
Para aumentar la capacidad de análisis del modelo se incluyó información (entre paréntesis se colocan algunos atributos) con respecto al Estudiante (Ver Figura 1), relacionada con: Estudiante (nombres, número de identificación, tipo de identificación, código, lugar de nacimiento, lugar de procedencia, teléfono, dirección, correo, género, edad actual, factor RH, Grupo étnico, pueblo-Comunidad, período académico de ingreso, tipo de admisión, puesto en la prueba de admisión, tipo de ingreso, indicador de ingreso, indicador procedencia programa regional, andrés bello, pensum actual, énfasis, institución educativa de procedencia, tipo institución de procedencia, indicador sector institución educativa de procedencia), Datos Económicos, Datos Demográficos, Indicadores, Semestre, Localización, Dificultades (problemas de conducta, vandalismo, alcoholismo, drogadicción, padre separados).

Con respecto a las materias: Materia (materia, curso, tipo materia, periodicidad, área Conocimiento, área conocimiento lcfes, indicador nota, indicador promedio, indicador materia electiva), Datos adicionales materia (semestre, créditos, horas semana, máximo faltas), Pensum, Programa, Periodo (periodo completo, año, periodo, tipo período, fecha inicio, fecha fin) y Estado (estado curso, estado semestre, estado de cancelación, estado materia).

Y dimensiones relacionadas con docentes: Docente y Datos adicionales docente. Las dimensiones que no se le colocaron los atributos son explicadas más adelante.

Las medidas modeladas son: promedio notas que afectan (Promedio Nota) y no afectan el promedio (Promedio Nota No Promedio); promedio de faltas, conteo de estudiantes y relación docente estudiantes.

Figura 1. Esquema estrella Notas Semestrales

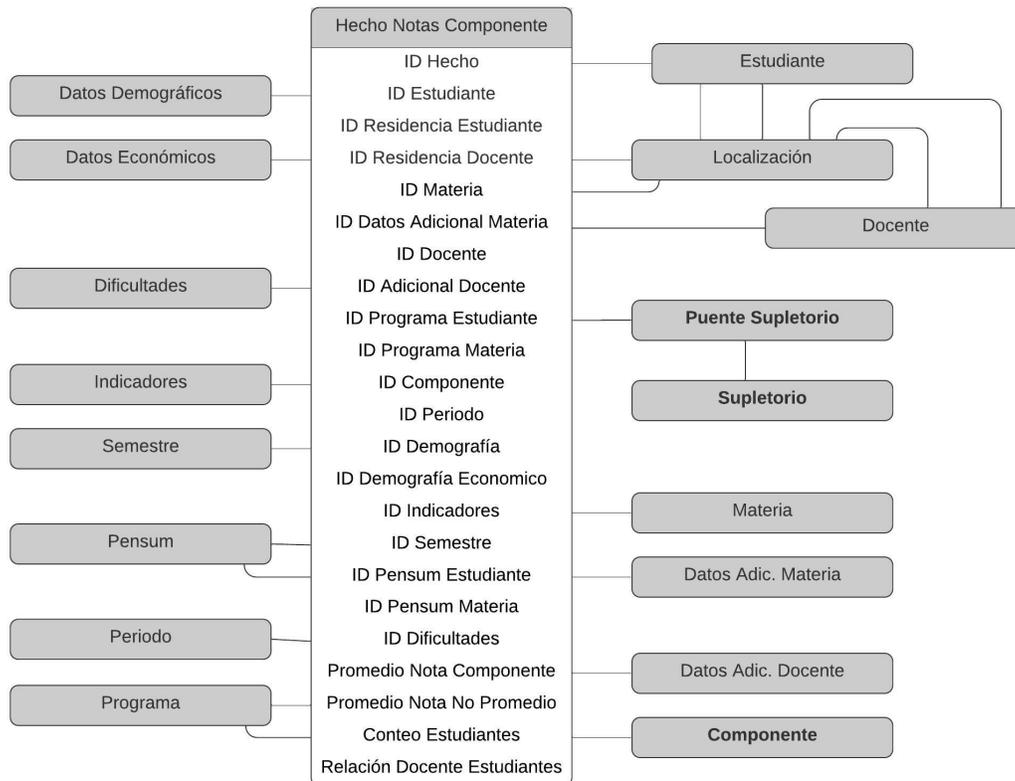


Fuente: Elaboración Propia.

Modelo dimensional de Notas por componente

Este modelo permite el análisis de las notas parciales de los estudiantes (Componente). La Figura 2 presenta dimensiones que fueron explicadas en el modelo anterior (concepto de dimensiones conformadas [3]) y dos dimensiones propias de este modelo que son: Componente (descripción, tipo) y Supletorio (motivo). La dimensión Docente en este modelo se relaciona directamente con la TH, porque, aunque una materia se dicte por varios docentes, un componente específico es responsabilidad de un único docente. En esta TH se modelan las medidas del modelo anterior (Conteo Estudiantes y Relación Docente Estudiantes) y promedio nota componente que afecta (Promedio Nota Componente) y no afecta el promedio (Promedio Nota No Promedio).

Figura 2. Esquema estrella Notas por Componente



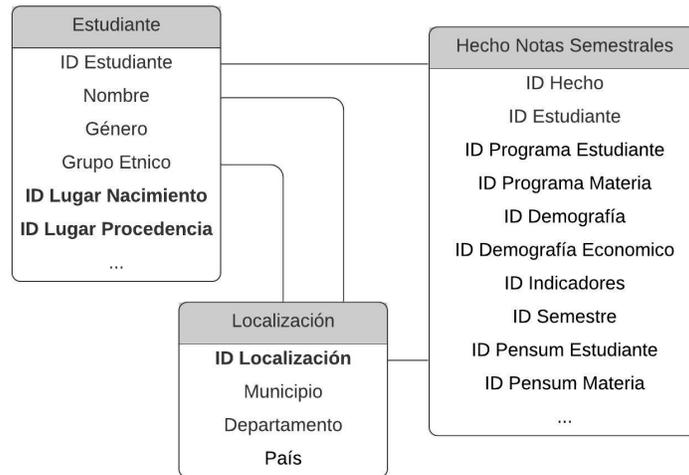
Fuente: Elaboración Propia

Situaciones de diseño

Existen situaciones de diseño que permiten obtener modelos más efectivos [3], lo cual facilita la adaptación de los modelos dimensionales por otras universidades. Las situaciones analizadas fueron [3]:

- El lugar de nacimiento y de procedencia de los estudiantes es un aspecto importante en los análisis, para reducir espacio de almacenamiento se debe utilizar el concepto de Subdimensión, que permite modelar el lugar en una sola tabla física (Localización) la cual se relaciona con la dimensión Estudiante, como se presenta en la Figura 3. La misma situación ocurre con los docentes.

Figura 3. Subdimensión

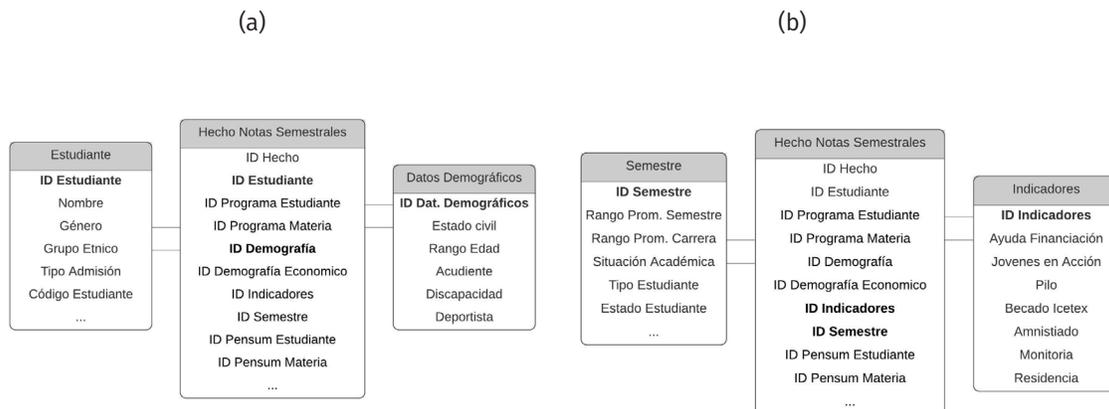


Fuente: Elaboración Propia

- Existen datos del estudiante que cambian en el tiempo y que son importantes para el análisis, por lo tanto, usando el concepto de Minidimensión se saca esta información del estudiante a una tabla separada que se relaciona directamente con la TH. Esto ocurre con la dimensión Datos Demográficos, como se aprecia en la Figura 4(a), y con Datos Económicos (estrato, número de hermanos, posición entre hermanos, vivienda propia, desplazado, madre cabeza, situación laboral, rango de horas trabajadas, dependencia económica).

También información de promedios y situación académica (Semestre), e indicadores de ayudas del estado o de la institución (Indicadores), como se observa en la Figura 4(b); además problemas de conducta (Dificultades). En el caso del docente también es necesario tener información adicional relacionada con el contrato (Datos Ad. Docente), al igual que datos adicionales para las materias (Datos Ad. Materia).

Figura 4. Minidimensión

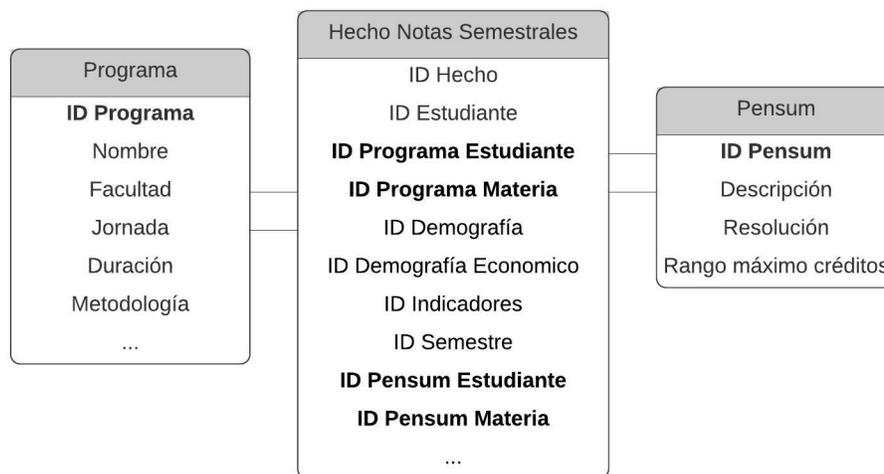


Fuente: Elaboración Propia

El estudiante puede cursar materias que pertenecen a un programa (Programa) diferente al que se encuentra matriculado (Ver Figura 5). Para modelarlo se relaciona dos veces la dimensión con la TH, de esta forma, solo se crea una tabla física (usando el concepto de Juego de roles). Lo mismo ocurre con el plan de estudios (Pensum).

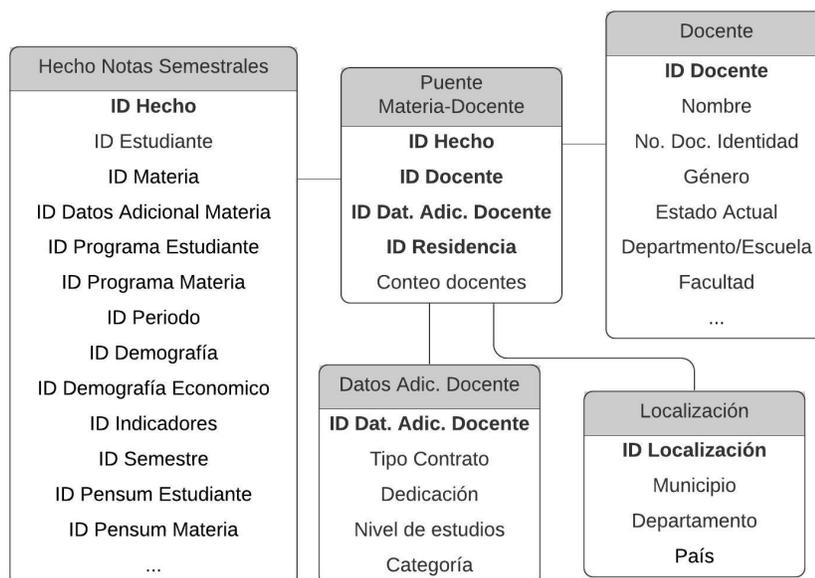
En las universidades los cursos pueden ser dictados por varios docentes (Docente), por esto se debe crear una tabla intermedia (Materia-Docente) entre la TH y la dimensión Docente. Además, otra información relacionada con el docente como el contrato y su ubicación de residencia debe incluirse en esta tabla intermedia, como se presenta en la Figura 6. Lo mismo ocurre con la dimensión Supletorio para permitir que se almacenen varios supletorios para el mismo componente, por eso se modela la tabla intermedia Puente-Supletorio.

Figura 5. Juego de roles



Fuente: Elaboración Propia

Figura 6. Tabla Puente



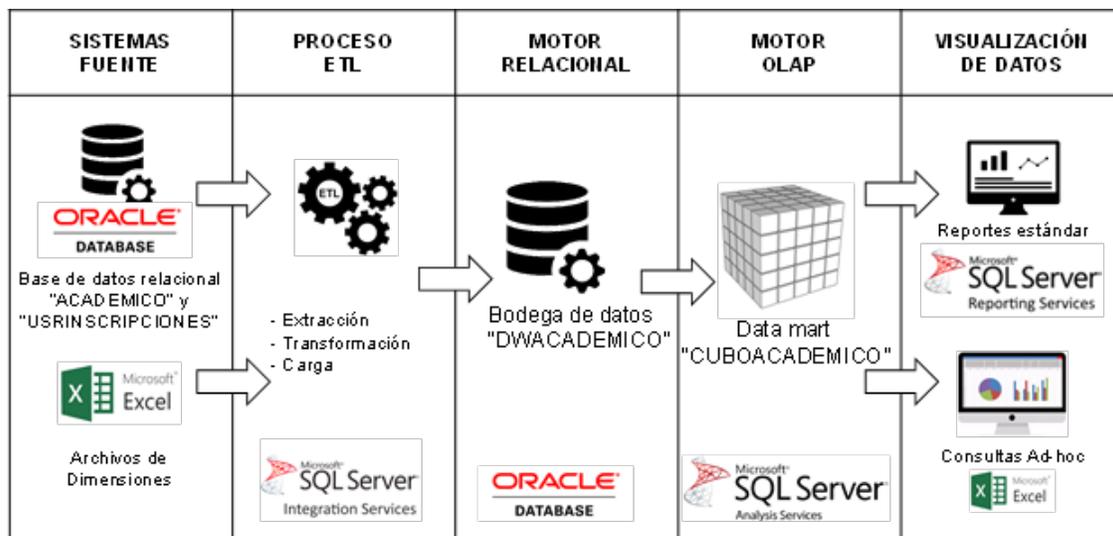
Fuente: Elaboración Propia

Prototipo

En el prototipo de bodega de datos se implementaron los dos modelos dimensionales Notas semestrales y Notas por componente para la Universidad del Cauca, teniendo en cuenta solo la estructura de la base de datos relacional que almacena esta información, ya que por políticas de confidencialidad de los datos el cargue no se pudo realizar con los datos reales.

En la Figura 7 se presenta la arquitectura del prototipo de bodega de datos, las fuentes de datos fueron dos bases de datos relacionales almacenadas en un motor Oracle ("ACADEMICO" y "USRINSCRIPCIONES") y algunos archivos EXCEL usados para cargar algunas de las dimensiones. El proceso ETL se realizó por medio del servicio de integración de datos de SQL Server, la bodega de datos (DWACADEMICO) se creó en el motor relacional de Oracle y el cubo ("CUBOACADEMICO") en el motor OLAP de SQLServer.

Figura 7. Arquitectura del prototipo de bodega de datos



Fuente: Elaboración propia

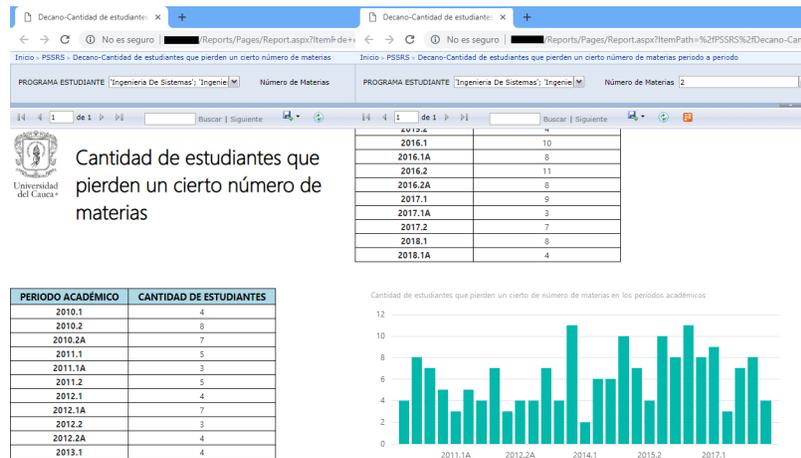
La aplicación web de usuario final fue realizada usando el servidor de reportes de SQL Server, el cual permite presentar los reportes estándar que acceden tanto a la bodega de datos relacional como al cubo OLAP (On-Line Analytic Processing). Estos reportes son organizados en carpetas para cada usuario para manejar los privilegios de acceso a estos.

Los reportes se visualizan de forma matricial y gráfica, en la Figura 8, se presenta un reporte que muestra la cantidad de estudiantes por período académico que han perdido un número específico de materias, el cual es dado como un parámetro, en este caso es dos materias. Además, la exploración de la información del cubo se realiza con consultas Ad-hoc por medio de Excel.

Evaluación del modelo

El resultado de la evaluación del Nivel de satisfacción se presenta en la Tabla 4, en la columna Fórmula las variables X_1 y X_2 corresponden a la cantidad de preguntas calificadas con 5 y 4 respectivamente, corresponde al número total de preguntas (57), y X a la proporción de preguntas calificadas con 5 y 4 respectivamente; la columna Valor deseado corresponde a la interpretación de la variable, donde 1 es el mayor valor de satisfacción, y la columna Valor total obtenido corresponde a la suma de X_1 y X_2 para obtener como resultado el valor de X , para una mejor interpretación del grado de satisfacción, este valor se ajusta a una escala de 0 a 100%.

Figura 8. Reporte Prototipo de bodega de datos



Fuente: Elaboración propia

Tabla 4. Resultado evaluación métrica

Métrica	Fórmula	Valor obtenido por variable	Valor total obtenido	Valor deseado
Nivel de Satisfacción	$X_1 = A_1/B$	$A_1 = 40$ $X_1 = 0.702$	$X = X_1 + X_2$ $X = 0.877$ 87.7%	1
	$X_2 = A_2/B$	$A_2 = 10$ $X = X_1 + X_2$		

Fuente: Elaboración propia

En la Tabla 4 se observa que los usuarios en el 70.2% (de las preguntas manifestaron estar Totalmente de acuerdo y en el 17.5% (estar Parcialmente de acuerdo, obteniendo un valor total de 87.7% (). En la escala de medición del grado de satisfacción, el valor obtenido se encuentra en el rango de 80.75% a 100%, que corresponde a Muy satisfactorio. Esto muestra que en general los reportes y consultas analíticas obtenidas del modelo dimensional fueron percibidos como importantes para los usuarios.

Conclusiones

En este artículo se proponen los modelos dimensionales de Notas semestrales y Notas por componente, que incluyen información socio-demográfica y académica de los estudiantes, de las materias y de contratación de los docentes, para aumentar la capacidad de análisis del modelo; y se explican las situaciones de diseño analizadas, las cuales permiten obtener modelos más efectivos y facilitan la adaptación por parte de las instituciones de educación superior públicas del país.

El grado de satisfacción de los modelos dimensionales por parte de los usuarios fue evaluado utilizando la medida de Nivel de satisfacción, cuyo resultado fue Muy satisfactorio (87.7%), lo cual permite reconocer el poder analítico de los modelos dimensionales creados y la utilidad que puede brindar en una institución educativa de nivel superior.

Se espera desarrollar otros modelos dimensionales relacionados con la investigación realizada por estudiantes y docentes; y con los egresados, buscando que el futuro análisis de esta información ayude a las instituciones educativas públicas a mejorar sus indicadores también en estos aspectos.

Referencias bibliográficas

1. Ministerio de Educación, Compendio Estadístico Educación Superior Colombiana. Bogotá, Colombia: Imprenta Nacional de Colombia, 2016.
2. Ministerio de Educación, "Documento metodológico MIDE Universitario 2017," 2017. [Online]. Available: [http://aprende.colombiaaprende.edu.co/ckfinder/userfiles/files/Documento Metodológico MIDE T 2018.pdf](http://aprende.colombiaaprende.edu.co/ckfinder/userfiles/files/Documento%20Metodol%C3%B3gico%20MIDE%20T%202018.pdf).
3. R. Kimball and M. Ross, *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, Third Edition. Indianapolis, Indiana: John Wiley & Sons, Inc, 2013.
4. F. Di Tria, E. Lefons, and F. Tangorra, "Academic data warehouse design using a hybrid methodology," *Comput. Sci. Inf. Syst.*, vol. 12, no. 1, pp. 135–160, 2014.
5. D. Martinez, M. Karanik, M. Giovannini, and N. Pinto, "Estudio de perfil de rendimiento académico: Un abordaje desde data warehousing," *Rev. Tecnol. y Cienc.*, vol. 13, no. 27, pp. 89–99, 2015.
6. T. Graca, J. Cristian, F. Machado, and B. P. Vieira, "Primary Education Evaluation in Brazil using Big Data and Cluster Analysis," in *Information Technology and Quantitative Management (ITQM 2015)*, 2015, vol. 55, pp. 1031–1039, DOI: 10.1016/j.procs.2015.07.061.
7. A. Ferreira, F. Portela, and M. F. Santos, "Pervasive Business Intelligence Platform to Improve the Quality of Decision Process in Primary and Secondary Education – A Portuguese Case Study," *New Contrib. Inf. Syst. Technol.*, pp. 167–176, 2015, DOI: 10.1007/978-3-319-16528-8.

8. R. Singh and K. Singh, "Design and Research of Data Analysis System for Student Education Improvement (Case Study: Student Progression System in University)," in 2016 International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE), 2016, pp. 508–512, DOI: 10.1109/ICMETE.2016.80.
9. L. Canchila and J. Sánchez, "Análisis de la Deserción Estudiantil de Cecar Utilizando Herramientas de Inteligencia de Negocios con Licencia Libre," (Trabajo de pregrado en Ingeniería de Sistemas), Facultad de Ciencias Básicas, Ingeniería y Arquitectura, Corporación Universitaria del Caribe, Sincelejo, Colombia, 2016.
10. L. W. Santoso, "Data Warehouse with Big Data Technology for Higher Education," in 4th Information Systems International Conference 2017, ISICO, 2017, vol. 124, pp. 93–99, doi: 10.1016/j.procs.2017.12.134.
11. W. Boulila, M. A. Mohammed, and F. Hamzah, "A business intelligence based solution to support academic affairs : case of Taibah University," in Wireless Networks, 2018, vol. 3, DOI: 10.1007/s11276-018-1880-3.
12. R. Rahutomo, R. A. Putri, and B. Pardamean, "Building Datawarehouse for Educational Institutions in 9 Steps," in 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), 2018, pp. 128–133, DOI: 10.1109/INAPR.2018.8627010.
13. R. Somya, D. Manongga, and M. A. Ineke Pakereng, "Service-oriented Business Intelligence (SoBI) for Academic and Financial Data Integration in University," in 2018 International Seminar on Application for Technology of Information and Communication, 2018, pp. 1–5, DOI: 10.1109/ISEMANTIC.2018.8549762.
14. A. K. B, S. Ghosh, and S. K. Ghosh, "eDWaaS : A Scalable Educational Data Warehouse as a Service," in Intelligent Systems Design and Applications, 2018, pp. 998–1007, DOI: 10.1007/978-3-319-76348-4.
15. K. Pratt, "Design Patterns for Research Methods: Iterative Field Research," AAAI Spring Symp. Exp. Des. Real, no. 1994, pp. 1–7, 2009, [Online]. Available: http://www.kpratt.net/wp-content/uploads/2009/01/research_methods.pdf.
16. B. Kitchenham and P. Brereton, "A systematic review of systematic review process research in software engineering," Information and Software Technology, vol. 55, no. 12. Elsevier B.V., pp. 2049–2075, 2013, DOI: 10.1016/j.infsof.2013.07.010.
17. M. E. Mendoza, L. Durán, and N. Rivera, "MBD 1.0 - Metodología de Desarrollo de Bodegas de Datos Para Micro, Pequeñas y Medianas Emp resas," UIS Ing., no. 1, pp. 85–101, 2010.
18. M. Zarour, "A rigorous user needs experience evaluation method based on software quality standards," vol. 18, no. 5, 2020, DOI: 10.12928/TELKOMNIKA.v18i5.16061.
19. H. Nakai, N. Tsuda, K. Honda, H. Washizaki, and Y. Fukazawa, "Initial Framework for Software Quality Evaluation based on ISO / IEC 25022 and ISO / IEC 25023," in 2016 IEEE International Conference on Software Quality, Reliability and Security-Companion, 2016, pp. 410–411, DOI: 10.1109/QRS-C.2016.66.
20. J. García, J. Aguilera, and A. Castillo, "Guía técnica para la construcción de escalas de actitud," Rev. electrónica Pedagog., vol. 8, no. 16, 2011, [Online]. Available: <https://studylib.es/doc/5379377/guía-técnica-para-la-construcción-de-escalas-de-actitud>.