**Universidad Marianao Gálvez de Guatemala, Extensión San Pedro Sacatepéquez, San Marcos**

14

**Proyecto Examen Privado Área de Análisis y Diseño de Sistemas**

**Ingeniería en Sistemas**

**Daniel Eduardo Monzón González 0903-07-10075**

**Introducción**

El término inteligencias empresariales se refiere al uso de datos en una empresa para facilitar la toma de decisiones. Abarca la comprensión del funcionamiento actual de la empresa, bien como la anticipación de acontecimientos futuros, con el objetivo de ofrecer conocimientos para respaldar las decisiones empresariales.

Las herramientas de inteligencia se basan en la utilización de un sistema de información de inteligencia que se forma con distintos datos extraídos de los datos de producción, con información relacionada con la empresa o sus ámbitos y con datos económicos.

Mediante las herramientas y técnicas ELT (extraer, cargar y transformar), o actualmente ETL (extraer, transformar y cargar) se extraen los datos de distintas fuentes, se depuran y preparan (homogeneización de los datos) para luego cargarlos en un almacén de datos.

La vida o el periodo de éxito de un software de inteligencia de negocios dependerá únicamente del éxito de su uso en beneficio de la empresa; si esta empresa es capaz de incrementar su nivel financiero, administrativo y sus decisiones mejoran la actuación de la empresa, el software de inteligencia de negocios seguirá presente mucho tiempo, en caso contrario será sustituido por otro que aporte mejores y más precisos resultados.

Finalmente, las herramientas de inteligencia analítica posibilitan el modelado de las representaciones basadas en consultas para crear un cuadro de mando integral que sirve de base para la presentación de informes.

**Aplicación Base**

1. **Definición del problema**

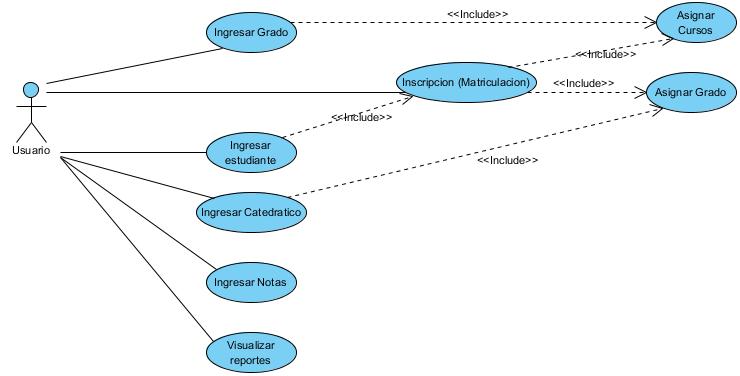
No se cuenta con ninguna aplicación mediante la cual los usuarios del sistema del ministerio de educación de Guatemala puedan realizar tanto las transacciones como la visualización de los reportes que sirven como base para la toma de decisiones en el nivel primario.

Dichas transacciones incluyen principalmente el ingreso o inscripción de estudiantes así como la asignación de cursos y grados a dichos estudiantes.

En dicha aplicación se necesita tener una sección para el ingreso de datos, esto solo es para poder ver como los datos ingresados afectan el resultado de los reportes, también se tendrá una sección de reportes gerenciales, donde se van a poder ver los siguientes reportes:

* El total de estudiantes
  + Por grado
  + Por departamento
  + Por sexo
  + Por catedrático
* El total de estudiantes Normales (Sin Beca)
  + Por grado
  + Por departamento
  + Por sexo
  + Por catedrático
* El total de estudiantes Becados
  + Por grado
  + Por departamento
  + Por sexo
  + Por catedrático
* Total de estudiantes matriculados
  + Por grado
  + Por departamento
  + Por sexo
  + Por Año

1. **Casos de Uso**

****

* 1. **Descripción de casos de Uso**

|  |  |  |  |
| --- | --- | --- | --- |
| **Nombre Caso de uso:** | Ingresar Grado | | |
| **Actor(es):** | Usuario | | |
| **Descripción:** | Permite al usuario ingresar nuevos grados al sistema. | | |
| **Evento desencadenador:** | El usuario ingresa los datos del grado y presiona el botón registrar. | | |
| **Pasos Realizados** |  | | **Información para los pasos** |
| 1. El usuario ingresa a la página e introduce los datos del nuevo grado y presiona el botón aceptar. | | | Formulario web para ingresar datos de los estudiantes. |
| 1. Se validan los datos introducidos. | | | Nombre\_Grado, idCatedratico, idNivel\_Educativo |
| **Pre-condiciones:** | | Ingresar a la página web ingresar nuevo grado. | |
| **Post-condiciones:** | | El usuario ingreso con éxito el nuevo grado. | |
| **Requerimientos cumplidos:** | | Permitir a los usuarios poder ingresar nuevos grados mediante el uso de un sitio web. | |
| **Prioridad:** | | Media | |
| **Riesgo:** | | Medio | |

|  |  |  |  |
| --- | --- | --- | --- |
| **Nombre Caso de uso:** | Ingresar Estudiante | | |
| **Actor(es):** | Usuario | | |
| **Descripción:** | Permite al usuario ingresar nuevos estudiantes al sistema. | | |
| **Evento desencadenador:** | El usuario ingresa a la página web para el ingreso de estudiantes e ingresa los datos del estudiante y presiona el botón registrar. | | |
| **Pasos Realizados** |  | | **Información para los pasos** |
| 1. El usuario ingresa a la página e introduce los datos del estudiante. | | | Formulario web para ingresar estudiantes. |
| 1. Se validan los datos introducidos. | | | Datos del estudiante. |
| **Pre-condiciones:** | | Ingresar a la página web para ingreso de estudiantes. | |
| **Post-condiciones:** | | El usuario ingreso con éxito el nuevo estudiante. | |
| **Requerimientos cumplidos:** | | Permitir a los usuarios ingresar nuevos estudiantes mediante el uso de un sitio web. | |
| **Prioridad:** | | Media | |
| **Riesgo:** | | Medio | |

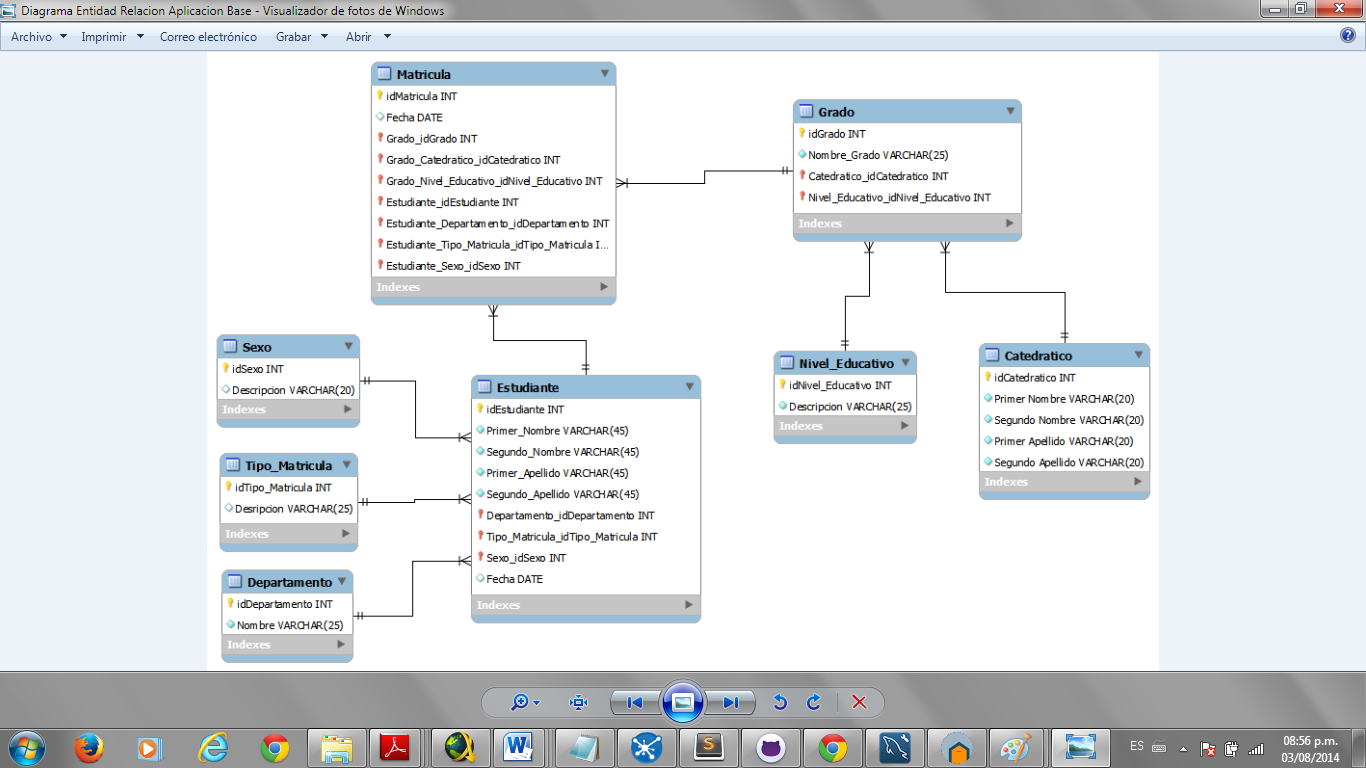
|  |  |  |  |
| --- | --- | --- | --- |
| **Nombre Caso de uso:** | Ingresar Notas | | |
| **Actor(es):** | Usuario | | |
| **Descripción:** | Permite al usuario ingresar las notas de los estudiantes matriculados en los diferentes grados. | | |
| **Evento desencadenador:** | El usuario ingresa las notas de los distintos cursos y presiona el botón registrar notas. | | |
| **Pasos Realizados** |  | | **Información para los pasos** |
| 1. El usuario ingresa a la página e introduce el número de matriculación del estudiante. | | | Idmatriculacion. |
| 1. Se verifica la existencia de la matriculación. | | | Idmatriculacion. |
| 1. Se muestran los cursos asignados al estudiante y se pide que se ingresen las notas. | | | Información estudiante, cursos asignados. |
| 1. El usuario ingresa las notas al formulación web de inserción de notas y presiona el botón registrar. | | | Formulario web para la inserción de notas. |
| **Pre-condiciones:** | | Ingresar a la página web para ingreso de notas. | |
| **Post-condiciones:** | | El usuario ingreso con éxito las notas del estudiante. | |
| **Requerimientos cumplidos:** | | Permitir a los usuarios ingresar las notas de los estudiantes matriculados. | |
| **Prioridad:** | | Media | |
| **Riesgo:** | | Medio | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Nombre Caso de uso:** | | Ingresar Catedrático | | | | | |
| **Actor(es):** | | Usuario | | | | | |
| **Descripción:** | | Permite al usuario ingresar nuevos catedráticos al sistema. | | | | | |
| **Evento desencadenador:** | | El usuario ingresa los datos del catedrático y presiona el botón registrar. | | | | | |
| **Pasos Realizados** | |  | | | | **Información para los pasos** | |
| 1. El usuario ingresa a la página e introduce los datos del nuevo catedrático y presiona el botón aceptar. | | | | | | Formulario web para ingresar datos de los catedráticos. | |
| 1. Se validan los datos introducidos. | | | | | | Datos del nuevo catedrático. | |
| **Pre-condiciones:** | | | | Ingresar a la página web ingresar nuevo catedrático. | | | |
| **Post-condiciones:** | | | | El usuario ingreso con éxito el nuevo catedrático. | | | |
| **Requerimientos cumplidos:** | | | | Permitir a los usuarios poder ingresar nuevos catedráticos mediante el uso de un sitio web. | | | |
| **Prioridad:** | | | | Media | | | |
| **Riesgo:** | | | | Medio | | | |
| **Nombre Caso de uso:** | Inscripción (Matriculación) | | | | | |
| **Actor(es):** | Usuario | | | | | |
| **Descripción:** | Permite al usuario matricular al estudiante en un grado. | | | | | |
| **Evento desencadenador:** | El usuario ingresa los datos de la nueva matriculación y presiona el botón enviar. | | | | | |
| **Pasos Realizados** |  | | | | **Información para los pasos** | |
| 1. El usuario ingresa a la página e introduce los datos del estudiante. | | | | | Formulario web para ingresar datos de los estudiantes. | |
| 1. Se validan los datos introducidos. | | | | | Datos del estudiante | |
| 1. Se introducen los datos de la matriculación. | | | | | Formulario web para la matriculación de estudiantes. | |
| 1. Se validan los datos de matriculación. | | | | | Datos de estudiante y matriculación. | |
| **Pre-condiciones:** | | | Ingresar a la página web ingresar nueva matriculación. | | | |
| **Post-condiciones:** | | | El usuario ingreso con éxito la nueva matricula. | | | |
| **Requerimientos cumplidos:** | | | Permitir a los usuarios poder ingresar nuevas matrículas mediante el uso de un sitio web. | | | |
| **Prioridad:** | | | Alta | | | |
| **Riesgo:** | | | Alto | | | |

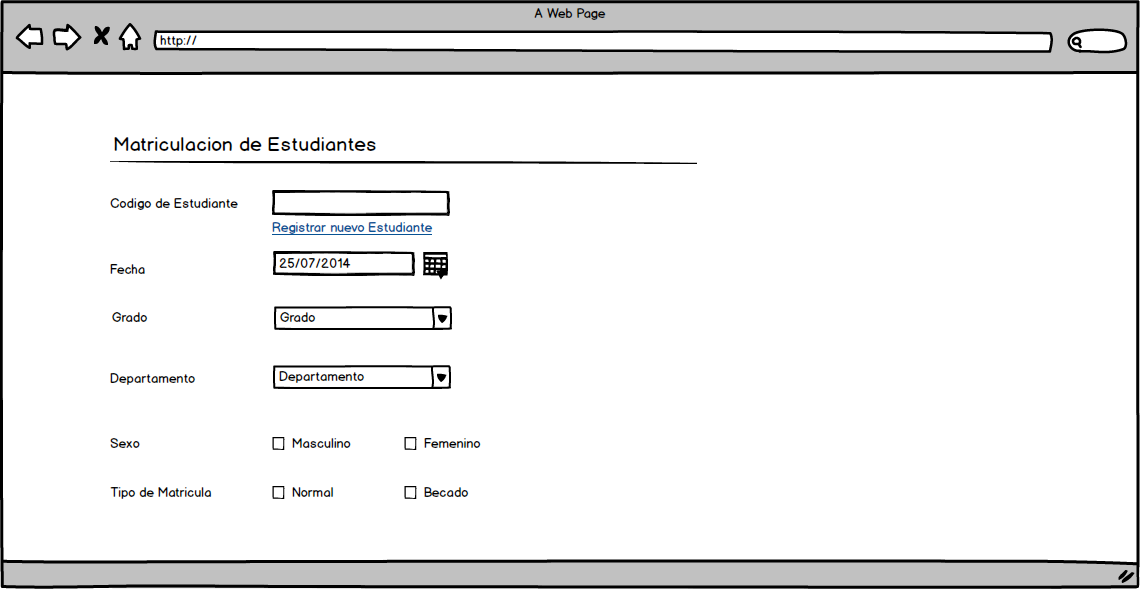
|  |  |  |  |
| --- | --- | --- | --- |
| **Nombre Caso de uso:** | Asignar Grado | | |
| **Actor(es):** | Usuario | | |
| **Descripción:** | Permite al usuario asignar un grado a un estudiante o catedrático. | | |
| **Evento desencadenador:** | El usuario ingresa los datos del grado y del estudiante o catedrático y presiona el botón registrar. | | |
| **Pasos Realizados** |  | | **Información para los pasos** |
| 1. Se ingresa el id del grado | | | idGrado |
| 1. Se validan los datos introducidos. | | |  |
| 1. Se muestra el formulario para asignar el grado al estudiante o catedrático | | | Formulario para la asignación de grado |
| 1. Se ingresan los datos del estudiante o catedrático. | | | Formulario para la asignación de grado. |
| 1. Se validan los datos. | | | Datos del grado. Datos del estudiante o datos del catedrático. |
| **Pre-condiciones:** | | Ingresar a la página web asignar grado. | |
| **Post-condiciones:** | | El usuario realiza con éxito la asignación. | |
| **Requerimientos cumplidos:** | | Permitir a los usuarios poder asignar grados mediante el uso de un sitio web. | |
| **Prioridad:** | | Media | |
| **Riesgo:** | | Medio | |

|  |  |  |
| --- | --- | --- |
| **Nombre Caso de uso:** | Asignar Curso | |
| **Actor(es):** | Usuario | |
| **Descripción:** | Permite al usuario asignar un curso a un estudiante o grado. | |
| **Evento desencadenador:** | El usuario ingresa los datos del curso y del estudiante o grado y presiona el botón registrar. | |
| **Pasos Realizados** |  | **Información para los pasos** |
| 1. Se ingresa el id del curso | | idCurso |
| 1. Se validan los datos introducidos. | |  |
| 1. Se muestra el formulario para asignar el curso al estudiante o grado | | Formulario para la asignación de curso |
| 1. Se ingresan los datos del estudiante o grado. | | Formulario para la asignación de curso. |
| 1. Se validan los datos. | | Datos del grado. Datos del estudiante o datos del grado. |
| **Pre-condiciones:** | Ingresar a la página web asignar curso. | |
| **Post-condiciones:** | El usuario realiza con éxito la asignación. | |
| **Requerimientos cumplidos:** | Permitir a los usuarios poder asignar cursos mediante el uso de un sitio web. | |
| **Prioridad:** | Media | |
| **Riesgo:** | Medio | |

* 1. **Diseño de la base de datos**



* 1. **Diseño de interfaz**

****

* 1. **Script Base de datos**

-- MySQL Script generated by MySQL Workbench

-- 08/03/14 21:20:54

-- Model: New Model Version: 1.0

SET @OLD\_UNIQUE\_CHECKS=@@UNIQUE\_CHECKS, UNIQUE\_CHECKS=0;

SET @OLD\_FOREIGN\_KEY\_CHECKS=@@FOREIGN\_KEY\_CHECKS, FOREIGN\_KEY\_CHECKS=0;

SET @OLD\_SQL\_MODE=@@SQL\_MODE, SQL\_MODE='TRADITIONAL,ALLOW\_INVALID\_DATES';

-- -----------------------------------------------------

-- Schema Aplicacion\_Base

-- -----------------------------------------------------

--

CREATE SCHEMA IF NOT EXISTS `Aplicacion\_Base` DEFAULT CHARACTER SET utf8 COLLATE utf8\_general\_ci ;

USE `Aplicacion\_Base` ;

-- -----------------------------------------------------

-- Table `Aplicacion\_Base`.`Catedratico`

-- -----------------------------------------------------

CREATE TABLE IF NOT EXISTS `Aplicacion\_Base`.`Catedratico` (

`idCatedratico` INT NOT NULL,

`Primer Nombre` VARCHAR(20) NOT NULL,

`Segundo Nombre` VARCHAR(20) NOT NULL,

`Primer Apellido` VARCHAR(20) NOT NULL,

`Segundo Apellido` VARCHAR(20) NOT NULL,

PRIMARY KEY (`idCatedratico`))

ENGINE = InnoDB;

-- -----------------------------------------------------

-- Table `Aplicacion\_Base`.`Nivel\_Educativo`

-- -----------------------------------------------------

CREATE TABLE IF NOT EXISTS `Aplicacion\_Base`.`Nivel\_Educativo` (

`idNivel\_Educativo` INT NOT NULL,

`Descripcion` VARCHAR(25) NOT NULL,

PRIMARY KEY (`idNivel\_Educativo`))

ENGINE = InnoDB;

-- -----------------------------------------------------

-- Table `Aplicacion\_Base`.`Grado`

-- -----------------------------------------------------

CREATE TABLE IF NOT EXISTS `Aplicacion\_Base`.`Grado` (

`idGrado` INT NOT NULL,

`Nombre\_Grado` VARCHAR(25) NOT NULL,

`Catedratico\_idCatedratico` INT NOT NULL,

`Nivel\_Educativo\_idNivel\_Educativo` INT NOT NULL,

PRIMARY KEY (`idGrado`, `Catedratico\_idCatedratico`, `Nivel\_Educativo\_idNivel\_Educativo`),

INDEX `fk\_Grado\_Catedratico1\_idx` (`Catedratico\_idCatedratico` ASC),

INDEX `fk\_Grado\_Nivel\_Educativo1\_idx` (`Nivel\_Educativo\_idNivel\_Educativo` ASC),

CONSTRAINT `fk\_Grado\_Catedratico1`

FOREIGN KEY (`Catedratico\_idCatedratico`)

REFERENCES `Aplicacion\_Base`.`Catedratico` (`idCatedratico`)

ON DELETE NO ACTION

ON UPDATE NO ACTION,

CONSTRAINT `fk\_Grado\_Nivel\_Educativo1`

FOREIGN KEY (`Nivel\_Educativo\_idNivel\_Educativo`)

REFERENCES `Aplicacion\_Base`.`Nivel\_Educativo` (`idNivel\_Educativo`)

ON DELETE NO ACTION

ON UPDATE NO ACTION)

ENGINE = InnoDB;

-- -----------------------------------------------------

-- Table `Aplicacion\_Base`.`Departamento`

-- -----------------------------------------------------

CREATE TABLE IF NOT EXISTS `Aplicacion\_Base`.`Departamento` (

`idDepartamento` INT NOT NULL,

`Nombre` VARCHAR(25) NOT NULL,

PRIMARY KEY (`idDepartamento`))

ENGINE = InnoDB;

-- -----------------------------------------------------

-- Table `Aplicacion\_Base`.`Tipo\_Matricula`

-- -----------------------------------------------------

CREATE TABLE IF NOT EXISTS `Aplicacion\_Base`.`Tipo\_Matricula` (

`idTipo\_Matricula` INT NOT NULL,

`Desripcion` VARCHAR(25) NULL,

PRIMARY KEY (`idTipo\_Matricula`))

ENGINE = InnoDB;

-- -----------------------------------------------------

-- Table `Aplicacion\_Base`.`Sexo`

-- -----------------------------------------------------

CREATE TABLE IF NOT EXISTS `Aplicacion\_Base`.`Sexo` (

`idSexo` INT NOT NULL,

`Descripcion` VARCHAR(20) NULL,

PRIMARY KEY (`idSexo`))

ENGINE = InnoDB;

-- -----------------------------------------------------

-- Table `Aplicacion\_Base`.`Estudiante`

-- -----------------------------------------------------

CREATE TABLE IF NOT EXISTS `Aplicacion\_Base`.`Estudiante` (

`idEstudiante` INT NOT NULL,

`Primer\_Nombre` VARCHAR(45) NOT NULL,

`Segundo\_Nombre` VARCHAR(45) NOT NULL,

`Primer\_Apellido` VARCHAR(45) NOT NULL,

`Segundo\_Apellido` VARCHAR(45) NOT NULL,

`Departamento\_idDepartamento` INT NOT NULL,

`Tipo\_Matricula\_idTipo\_Matricula` INT NOT NULL,

`Sexo\_idSexo` INT NOT NULL,

`Fecha` DATE NULL,

PRIMARY KEY (`idEstudiante`, `Departamento\_idDepartamento`, `Tipo\_Matricula\_idTipo\_Matricula`, `Sexo\_idSexo`),

INDEX `fk\_Matricula\_Departamento1\_idx` (`Departamento\_idDepartamento` ASC),

INDEX `fk\_Matricula\_Tipo\_Matricula1\_idx` (`Tipo\_Matricula\_idTipo\_Matricula` ASC),

INDEX `fk\_Matricula\_Sexo1\_idx` (`Sexo\_idSexo` ASC),

CONSTRAINT `fk\_Matricula\_Departamento`

FOREIGN KEY (`Departamento\_idDepartamento`)

REFERENCES `Aplicacion\_Base`.`Departamento` (`idDepartamento`)

ON DELETE NO ACTION

ON UPDATE NO ACTION,

CONSTRAINT `fk\_Matricula\_Tipo\_Matricula`

FOREIGN KEY (`Tipo\_Matricula\_idTipo\_Matricula`)

REFERENCES `Aplicacion\_Base`.`Tipo\_Matricula` (`idTipo\_Matricula`)

ON DELETE NO ACTION

ON UPDATE NO ACTION,

CONSTRAINT `fk\_Matricula\_Sexo1`

FOREIGN KEY (`Sexo\_idSexo`)

REFERENCES `Aplicacion\_Base`.`Sexo` (`idSexo`)

ON DELETE NO ACTION

ON UPDATE NO ACTION)

ENGINE = InnoDB;

-- -----------------------------------------------------

-- Table `Aplicacion\_Base`.`Matricula`

-- -----------------------------------------------------

CREATE TABLE IF NOT EXISTS `Aplicacion\_Base`.`Matricula` (

`idMatricula` INT NOT NULL,

`Fecha` DATE NULL,

`Grado\_idGrado` INT NOT NULL,

`Grado\_Catedratico\_idCatedratico` INT NOT NULL,

`Grado\_Nivel\_Educativo\_idNivel\_Educativo` INT NOT NULL,

`Estudiante\_idEstudiante` INT NOT NULL,

`Estudiante\_Departamento\_idDepartamento` INT NOT NULL,

`Estudiante\_Tipo\_Matricula\_idTipo\_Matricula` INT NOT NULL,

`Estudiante\_Sexo\_idSexo` INT NOT NULL,

PRIMARY KEY (`idMatricula`, `Grado\_idGrado`, `Grado\_Catedratico\_idCatedratico`, `Grado\_Nivel\_Educativo\_idNivel\_Educativo`, `Estudiante\_idEstudiante`, `Estudiante\_Departamento\_idDepartamento`, `Estudiante\_Tipo\_Matricula\_idTipo\_Matricula`, `Estudiante\_Sexo\_idSexo`),

INDEX `fk\_Inscripcion\_Grado1\_idx` (`Grado\_idGrado` ASC, `Grado\_Catedratico\_idCatedratico` ASC, `Grado\_Nivel\_Educativo\_idNivel\_Educativo` ASC),

INDEX `fk\_Matricula\_Estudiante1\_idx` (`Estudiante\_idEstudiante` ASC, `Estudiante\_Departamento\_idDepartamento` ASC, `Estudiante\_Tipo\_Matricula\_idTipo\_Matricula` ASC, `Estudiante\_Sexo\_idSexo` ASC),

CONSTRAINT `fk\_Inscripcion\_Grado`

FOREIGN KEY (`Grado\_idGrado` , `Grado\_Catedratico\_idCatedratico` , `Grado\_Nivel\_Educativo\_idNivel\_Educativo`)

REFERENCES `Aplicacion\_Base`.`Grado` (`idGrado` , `Catedratico\_idCatedratico` , `Nivel\_Educativo\_idNivel\_Educativo`)

ON DELETE NO ACTION

ON UPDATE NO ACTION,

CONSTRAINT `fk\_Matricula\_Estudiante`

FOREIGN KEY (`Estudiante\_idEstudiante` , `Estudiante\_Departamento\_idDepartamento` , `Estudiante\_Tipo\_Matricula\_idTipo\_Matricula` , `Estudiante\_Sexo\_idSexo`)

REFERENCES `Aplicacion\_Base`.`Estudiante` (`idEstudiante` , `Departamento\_idDepartamento` , `Tipo\_Matricula\_idTipo\_Matricula` , `Sexo\_idSexo`)

ON DELETE NO ACTION

ON UPDATE NO ACTION)

ENGINE = InnoDB;

SET SQL\_MODE=@OLD\_SQL\_MODE;

SET FOREIGN\_KEY\_CHECKS=@OLD\_FOREIGN\_KEY\_CHECKS;

SET UNIQUE\_CHECKS=@OLD\_UNIQUE\_CHECKS;

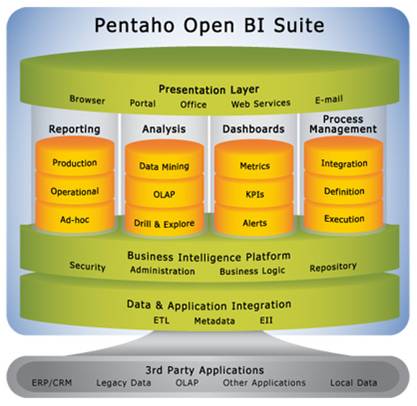
**Inteligencia de Negocios**

1. **Estudio de Pre-Factibilidad de herramientas a utilizar.**

En el presente proyecto se consideran las herramientas open source que permitan realizar todo el proceso de Data warehouse.

Por lo que se seleccionó Pentaho basado en las características que se presentan a continuación.

La solución Business Intelligence OpenSource Pentaho pretende ser una alternativa a las soluciones propietarias tradicionales más completas: Business Objects, Cognos, Microstrategy, Microsoft, IBM, etc., por lo que incluye todos aquellos componentes que se pueden encontrar en las soluciones Business Intelligence (BI) propietarias más avanzadas:



* Reporting.

La solución proporcionada por la plataforma Business Intelligence OpenSource Pentaho e integrada en su suite para el desarrollo de informes se llama Pentaho Reporting



Existen tres productos con diferentes enfoques y dirigidos a diferentes tipos de usuarios:

* Pentaho Report Designer

Es un editor basado en Eclipse con prestaciones profesionales con capacidad de personalización de informes a las necesidades de los negocios destinado a desarrolladores.

Esta herramienta está estructurada de forma que los desarrolladores puedan acceder a sus prestaciones de forma rápida.

Incluye un editor de consultas para facilitar la confección de los datos que serán utilizados en un informe.

* Pentaho Report Design Wizard

Es una herramienta de diseño de informes, que facilita el trabajo y permite a los usuarios obtener resultados de forma inmediata. Está destinada a usuarios con menos conocimientos técnicos.

* Web ad-hoc reporting

Es el similar a la herramienta Pentaho Report Design Wizard, pero vía web.

Esta herramienta extiende la capacidad de los usuarios ﬁnales para la creación de informes a partir de plantillas preconﬁguradas y siguiendo un asistente de creación.

* Dashboards.

Esta solución provee inmediata perspicacia en un rendimiento individual, departamental o empresarial. Pentaho Dashboards facilita a los usuarios de los negocios información crítica que necesitan para entender y mejorar el rendimiento organizacional.

El Pentaho Dashboards es una potente herramienta que cuenta con las siguientes características:

* Identiﬁcacion de métricas clave (KPIs, Key Performance Indicators), mediante la generación de Monitoreo/Métricas.
* Realización de investigaciones de detalles subyacentes, con reportes de soportes.
* Ejecución de seguimientos de excepciones, permitiendo pre-establecer alertas basadas en reglas del negocio.
* DataMining.

La plataforma Business Intelligence OpenSource Pentaho ofrece diferentes soluciones para el desarrollo de un proyecto deBusiness Intelligence.

El Weka (Waikato Enviroment for Knowledge Analysis) es un conjunto de librerías JAVA para la extracción de conocimientos desde bases de datos

Es un software que ha sido desarrollado bajo licencia GPL lo cual ha impulsado que sea una de las suites más utilizadas en el área en los últimos años.

**Características Generales del Weka**Esta herramienta Open Source incluye las siguientes características:

* + - Diversas fuentes de datos (ASCII, JDBC).
    - Interfaz visual basada en procesos / ﬂujos de datos (rutas).
    - Distintas herramientas de minería de datos:
      * Reglas de asociación (a priori, Tertius, etc.).
      * Agrupación / segmentación / conglomerado (cobweb, EM y k-medias).
      * Clasiﬁcación (redes neuronales, reglas y árboles de decisión, aprendizaje bayesiano).
      * Regresión (regresión lineal, SVM, etc.).
      * Manipulación de datos (pick & mix, muestreo, combinación, separación, etc.).
      * Combinación de modelos (bagging, boosting, etc.).
      * Entorno de experimentos, con la posibilidad de realizar pruebas estadísticas (T-test).
* ETL (Pentaho Data Integration).

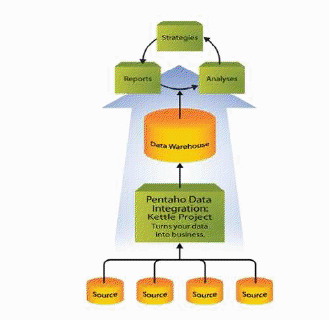
Los datos que alimentan a un sistema data warehouse (DW) proviene de diferentes fuentes, estas fuentes son los distintos sistemas operacionales que la empresa posee, generalmente ni son homogéneos entre sí ni concuerdan exactamente con lo que se necesita, por lo que será necesario realizar todas las adaptaciones pertinentes.

También muchas organizaciones tienen información disponible en aplicaciones y base de datos separadas.

Pentaho Data Integration abre, limpia e integra esta valiosa información y la pone en manos del usuario. Provee una consistencia, una sola versión de todos los recursos de información, que es uno de los más grandes desafíos para las organizaciones TI hoy en día.

Pentaho Data Integration permite una poderosa ETL (Extract, Transform, Load) Extracción, Transformación y Carga.

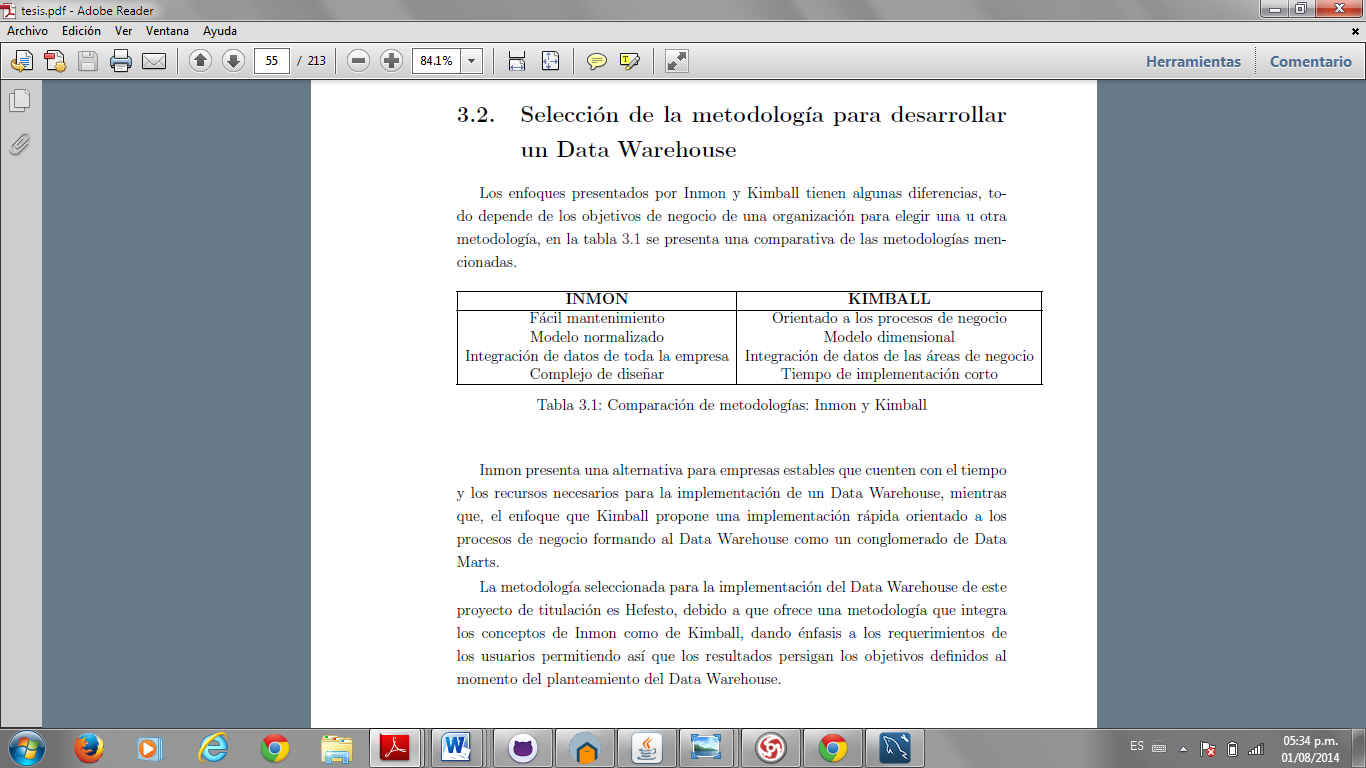
El uso de la solución Kettle permite evitar grandes cargas de trabajo manual frecuentemente difícil de mantener y de desplegar.



1. **Introducción**

Antes de empezar con el análisis para la implementación del Data Warehouse es necesario seleccionar la metodología, a partir de la cual se debe realizar la recolección de requerimientos para comprender las necesidades de información que se tengan y diseñar la solución más adecuada, considerando que los requerimientos estén acorde a la información disponible en las diferentes fuentes de datos que se dispongan.

1. **Selección de la metodología para desarrollar un Data warehouse.**

Inmon presenta una alternativa para empresas estables que cuenten con el tiempo y los recursos necesarios para la implementación de un Data Warehouse, mientras que, el enfoque que Kimball propone una implementación rápida orientada a los procesos de negocio formando al Data Warehouse como un conglomerado de Data Marts. La metodología seleccionada para la implementación del Data Warehouse de este proyecto de titulación es Hefesto, debido a que ofrece una metodología que integra los conceptos de Inmon como de Kimball, dando énfasis a los requerimientos de los usuarios permitiendo asi que los resultados persigan los objetivos definidos al momento del planteamiento del Data Warehouse.

1. **Aplicación de la metodología Hefesto**
   1. **Análisis de Requerimientos**
      1. **Identificar Preguntas**

Se indago con los usuarios del sistema de educación de Guatemala sobre la necesidad de información que tienen para la toma de decisiones. Pero la misma abarca casi todas las actividades del Ministerio de educación por lo cual se le pidió que escogieran un proceso el cual consideraran más importante para la toma de decisiones de dicho Ministerio, pero dicho proceso debería de estar soportado por algún OLTP (Procesamiento de Transacciones en Línea).

Por lo cual eligieron el proceso para el control de los estudiantes, catedráticos y grados en el nivel primario.

Por lo que se procedió a identificar lo que más les interesaba conocer acerca del proceso antes mencionado y establecer cuáles eran las variables y perspectiva que se tendrían en cuanta para poder tomar las decisiones correctas basadas en ellas.

Se procedió a aplicar las técnicas de recopilación de información tales como entrevistas y Encuestas, con dichas técnicas se pretendía que los usuarios nos dieran indicios sobre cual según ellos eran los indicadores que representan de mejor modo el proceso de control del área de primaria y que exactamente era la información que ellos pretendían obtener de ese proceso para realizar una mejor toma de decisiones.

Las respuestas a dichas preguntas se pueden resumir de la siguiente manera:

1. Se desea establecer el total de estudiantes existentes en la base de datos de nivel primario por departamento en un tiempo determinado.
2. Se desea establecer el total de estudiantes existentes en la base de datos de nivel primario por sexo en un tiempo determinado.
3. Se desea conocer el total de estudiantes de nivel primario Becados por departamento en un tiempo determinado.
4. Se desea conocer el total de estudiantes de nivel primario Becados por sexo en un tiempo determinado.
5. Se desea conocer el total de estudiantes de nivel primario Becados por catedrático en un tiempo determinado.
6. Se desea conocer el total de estudiantes Normales (Sin Beca) de nivel primario por grado en un tiempo determinado.
7. Se desea conocer el total de estudiantes Normales (Sin Beca) de nivel primario por departamento en un tiempo determinado.
8. Se desea conocer el total de estudiantes Normales (Sin Beca) de nivel primario por sexo en un tiempo determinado.
9. Se desea conocer el total de estudiantes Normales (Sin Beca) de nivel primario por catedrático en un tiempo determinado.
10. Se desea conocer el total de estudiantes matriculados nivel primario por grado en un tiempo determinado.
11. Se desea conocer el total de estudiantes matriculados nivel primario por departamento en un tiempo determinado.
12. Se desea conocer el total de estudiantes matriculados nivel primario por sexo en un tiempo determinado.
13. Se desea conocer el total de estudiantes matriculados nivel primario por catedrático en un tiempo determinado.

Debido a que la dimensión Tiempo es un elemento fundamental en el DataWarehouse, se hizo hincapié en él. Además, se puso mucho énfasis en dejar en claro a los usuarios, a través de ejemplos prácticos, que es este componente es el que permitirá tener varias versiones de los datos a fin de realizar un correcto análisis posterior.

Como se puede apreciar, las necesidades de información expuestas están acorde a los objetivos y estrategias de la empresa, ya que es precisamente esta información requerida la que proveerá un ámbito para la toma de decisiones, que en este caso permitirá analizar el comportamiento de los estudiantes y catedráticos a los cuales se pretende satisfacer ampliamente para así lograr tomar decisiones que favorezcan al mejoramiento de la educación en Guatemala.

* + 1. **Identificar Indicadores y Perspectivas**

A continuación se analizaran las preguntas obtenidas en el paso anterior y se detallaran con sus respectivos indicadores y perspectivas.

Indicadores

Perspectivas

1. Total de estudiantes existentes en la base de datos de nivel primario por departamento en tiempo determinado.
2. Total de estudiantes existentes en la base de datos de nivel primario por sexo en tiempo determinado.
3. Total de estudiantes matriculados de nivel primario Becados por departamento en tiempo determinado.
4. Total de estudiantes matriculados de nivel primario Becados por sexo tiempo en determinado.
5. Total de estudiantes matriculados de nivel primario Becados por catedrático en tiempo determinado.
6. Total de estudiantes matriculados de nivel primario Normales (Sin Beca) por grado en tiempo determinado.
7. Total de estudiantes matriculados de nivel primario Normales (Sin Beca) por departamento en tiempo determinado.
8. Total de estudiantes matriculados de nivel primario Normales (Sin Beca) por sexo en tiempo determinado.
9. Total de estudiantes matriculados de nivel primario Normales (Sin Beca) por catedrático en tiempo determinado.
10. Total de estudiantes matriculados de nivel primario por grado en tiempo determinado.
11. Total de estudiantes matriculados de nivel primario por departamento en tiempo determinado.
12. Total de estudiantes matriculados de nivel primario por sexo en tiempo determinado.
13. Total de estudiantes matriculados de nivel primario por catedrático en tiempo determinado.

En resumen los Indicadores son los Siguientes:

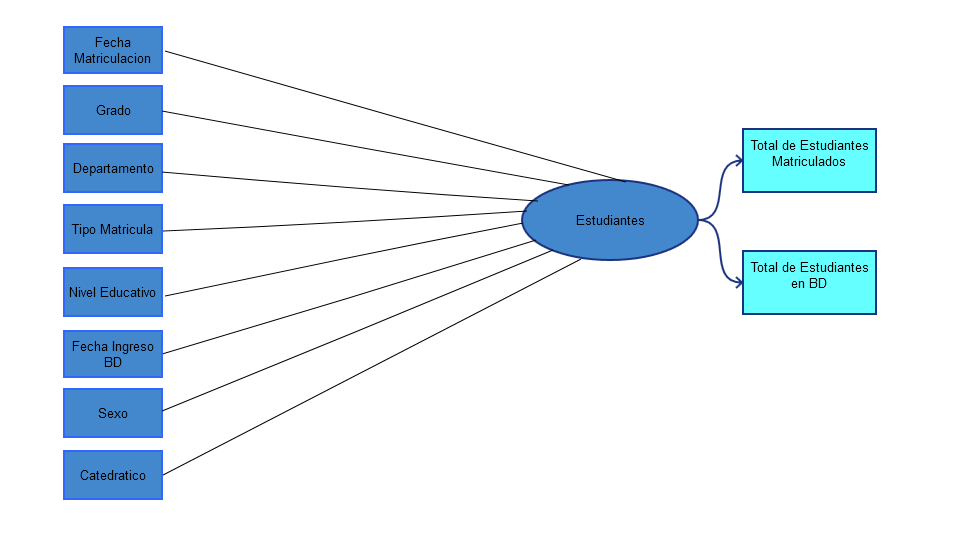
* Total de estudiantes Matriculados
* Total de estudiantes en la base de datos

Y las perspectivas las siguientes

* Fecha Ingreso a la Base de Datos
* Fecha Matriculación
* Grado
* Sexo
* Catedrático
* Nivel educativo
* Departamento
* Tipo de matricula
  + 1. **Modelo Conceptual**

A continuación construiremos el modelo conceptual a partir de los indicadores y perspectivas obtenidas en la etapa anterior.

A través de este modelo, se podrá observar con claridad cuáles son los alcances del proyecto, para luego poder trabajar sobre ellos, además al poseer un alto nivel de definición de los datos, permite que pueda ser presentado ante los usuarios y explicado con facilidad.

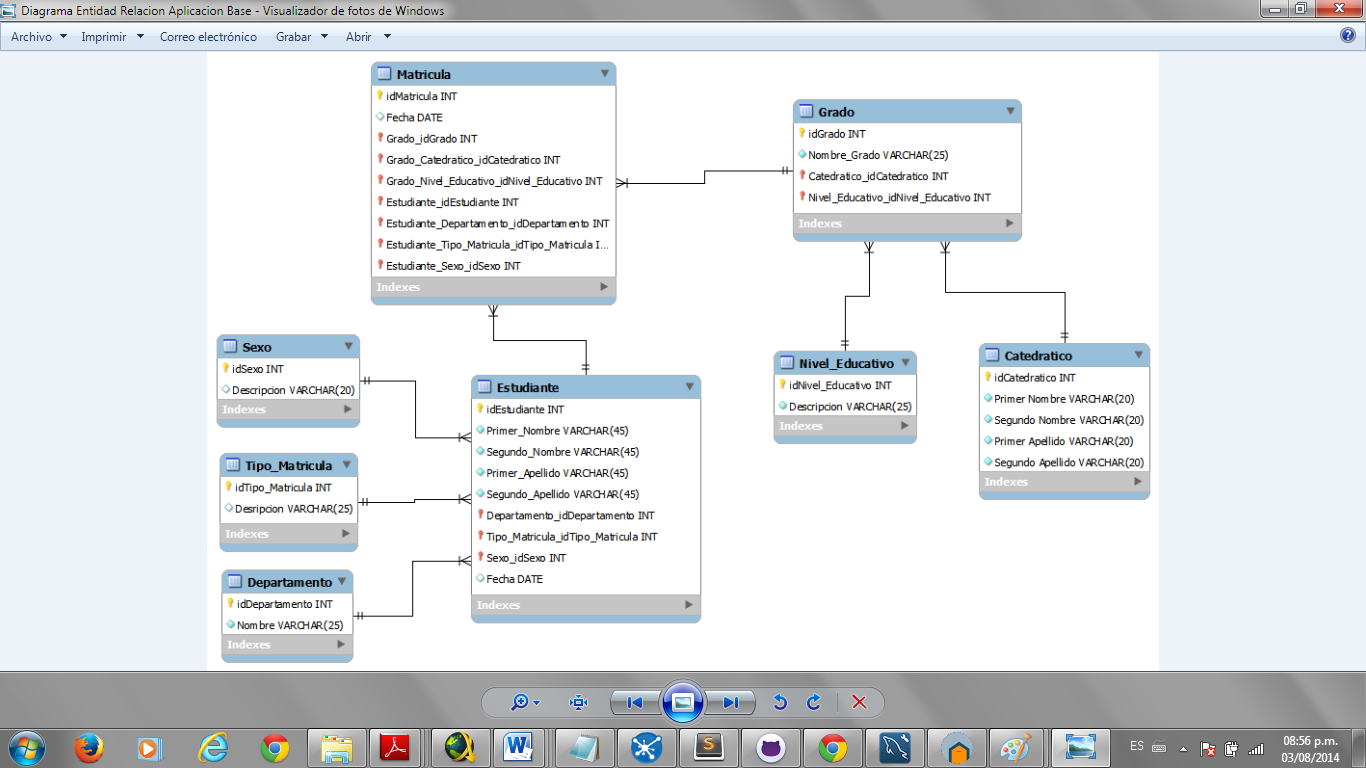


Como se puede observar, “Estudiantes” es la relación mediante la cual se unen las diferentes perspectivas, para obtener como resultado los indicadores requeridos por los usuarios.

* 1. **Análisis de los OLTP**

Primero vamos a definir la fuente de información

* Aplicación Web.
  + Descripción: se dispone de una aplicación web en la cual se lleva a cabo todas las operaciones correspondientes a la inserción de nuevos estudiantes y su respectiva matriculación asi como el ingreso de catedraticos y sus repsectivos grados todo esto realizado sobre el nivel primario a nivel nacional.
  + Dicha aplicación utiliza como repositorio el motor de base de datos MySql donde se almacena toda la informacion recolectada.
  + El diagrama entidad relación de la aplicación web se presenta a continuación:



* + 1. **Conformar Indicadores**

A continuación se explican las operaciones para calcular los indicadores de la tabla de hecho.

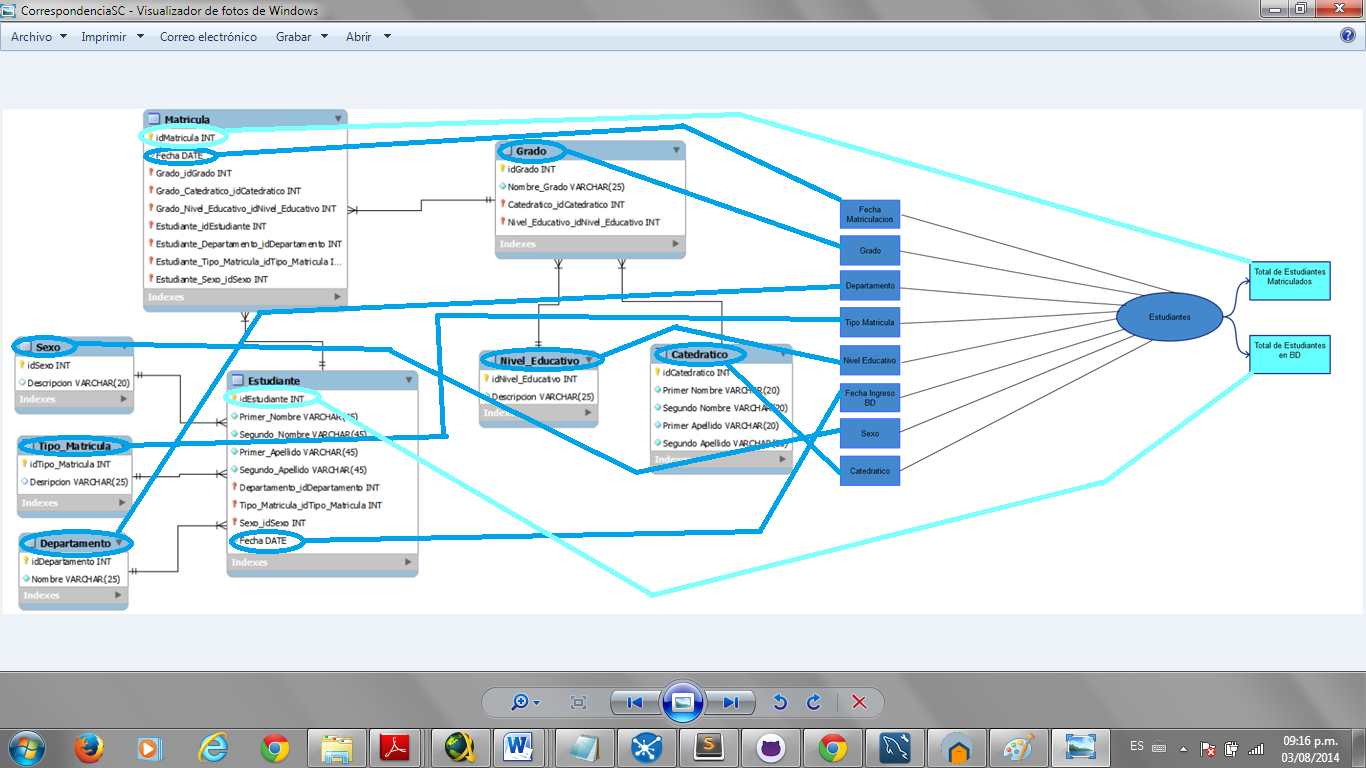
* + Total de Estudiantes en Base de datos
    - Hechos: Total de estudiantes en Base de Datos
    - Función de Sumarizacion: SUM.
    - Aclaración: El indicador “Total de Estudiantes en base de datos” representa la suma de los estudiantes de los estudiantes que se encuentran en la base de datos del sistema.
  + Total de Estudiantes Matriculados
    - Hechos: Total de estudiantes Matriculados
    - Función de Sumarizacion: SUM.
    - Aclaración: El indicador “Total de Estudiantes Matriculados” representa la suma de los estudiantes de los estudiantes matriculados.
    1. **Establecer Correspondencia**

Se examinaron los OLTP involucrados en el proceso para poder identificar y establecer las correspondencias de los mismos con el modelo conceptual.

Las relaciones identificadas fueron las siguientes:

* + La tabla “Grado” se relaciona con la perspectiva “Grado”.
  + La tabla “Sexo” se relaciona con la perspectiva “Sexo”
  + La tabla “Departamento” se relaciona con la perspectiva “Departamento”
  + La tabla “Catedratico” se relaciona con la perspectiva catedratico.
  + La tabla “Nivel Educativo” se relaciona con la perspectiva “Nivel Educativo”
  + La tabla “Tipo Matricula” se relaciona con la perspectiva “Tipo Matricula”.
  + El campo “Fecha ” de la tabla Matricula se relaciona con la perspectiva “Fecha Matriculación”
  + El campo “Fecha ” de la tabla Estudiante se relaciona con la perspectiva “Fecha Ingreso Base de datos”

La correspondencia de indicadores se presenta en la imagen siguiente:



* + 1. **Nivel de Granulidad**

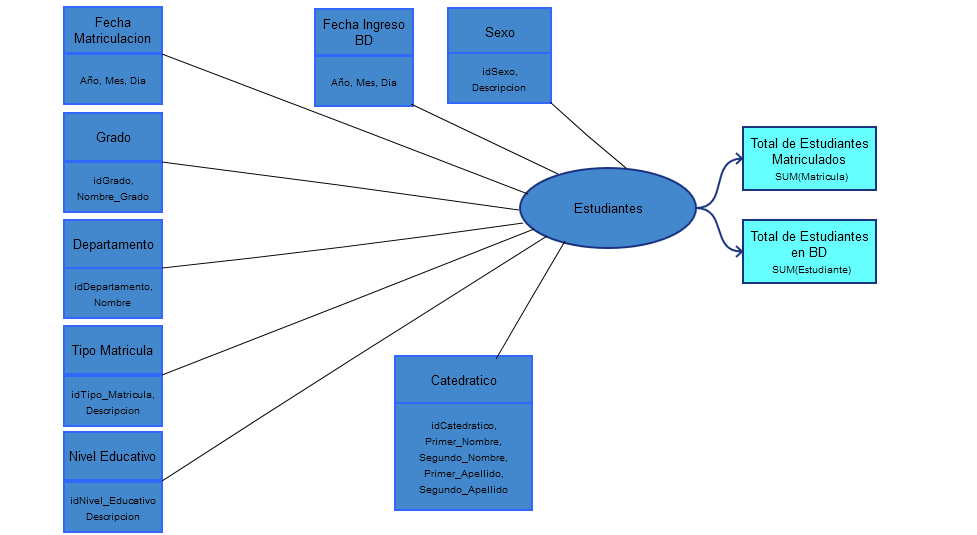
De acuerdo a las correspondencias que se establecieron anteriormente, se analizaron los campos correspondientes a cada tabla a la que se hacer referencia, dicho análisis se llevó a cabo examinando la base de datos para entender los significados de cada campo.

* + En la perspectiva “Grado” los datos disponibles son los siguientes:
    - idGrado: es la clave primaria de la tabla “Grado” y representa a un cliente en particular.
    - Nombre\_Grado: el nombre correspondiente al grado
    - Catedratico\_idCatedratico: representa mediante una clave foránea el catedratico que tiene a su cargo el grado.
    - Nivel\_Educativo\_idNivel\_Educativo: representa mediante una clave foránea el nivel educativo al cual pertenece el grado.
  + En la perspectiva “Sexo” los datos disponibles son los siguientes:
    - idSexo: es la clave primaria de la tabla “Sexo” y representa un género en particular.
    - Descripción: en él se describe el género al cual se quiere representar.
  + En la perspectiva “Catedrático” los datos disponibles son los siguientes:
    - IdCatedratico: es la clave primaria de la tabla “Catedrático” y representa a un catedrático en especial.
    - Primer\_Nombre: representa el primer nombre del catedrático.
    - Segundo\_Nombre: representa el segundo nombre del catedrático.
    - Primer\_Apellido: representa el primer apellido del catedrático.
    - Segundo\_Apellido: representa el segundo apellido del catedrático.
  + En la perspectiva “Nivel educativo” los datos disponibles son los siguientes:
    - idNivel\_Educativo: es la clave primaria de la tabla “Nivel Educativo” y representa de forma única a un nivel educativo.
    - Descripción: en él se describe el nivel educativo correspondiente.
  + En la perspectiva “Departamento” los datos disponibles son los siguientes:
    - IdDepartamento: es la clave primaria de la tabla “Departamento” y representa de forma única a un departamento.
    - Nombre: en él se indica el nombre del departamento.
  + En la perspectiva “Tipo\_Matricula” los datos disponibles son los siguientes:
    - idTipo\_Matricula: es la clave primaria de la tabla Tipo\_Matricuala y representa de forma única a un tipo de matrícula.
    - Descripción: en él se describe el tipo de matrícula.
  + En la perspectiva “Fecha Matriculacion” los datos disponibles son los siguientes:
    - Año
    - Nombre mes
    - Numero mes
    - Nombre día
    - Numero día
  + En la perspectiva “Fecha Ingreso a Base de datos” los datos disponibles son los siguientes:
    - Año
    - Nombre mes
    - Numero mes
    - Nombre día
    - Numero día

Una vez que se recolectó toda la información correspondiente a cada perspectiva se procedió a consultar con los usuarios cuales eran los datos que consideraban de interés para analizar los indicadores ya expuestos, los resultados obtenidos fueron los siguientes:

* + Perspectiva Grado
    - idGrado: es la clave primaria de la tabla “Grado” y representa a un cliente en particular.
    - Nombre\_Grado: el nombre correspondiente al grado
  + Perspectiva Sexo:
    - idSexo: es la clave primaria de la tabla “Sexo” y representa un género en particular.
    - Descripción: en él se describe el género al cual se quiere representar.
  + Perpespectiva Catedratico
    - IdCatedratico: es la clave primaria de la tabla “Catedrático” y representa a un catedrático en especial.
    - Primer\_Nombre: representa el primer nombre del catedrático.
    - Segundo\_Nombre: representa el segundo nombre del catedrático.
    - Primer\_Apellido: representa el primer apellido del catedrático.
    - Segundo\_Apellido: representa el segundo apellido del catedrático.
  + Perspectiva Nivel Educativo:
    - idNivel\_Educativo: es la clave primaria de la tabla “Nivel Educativo” y representa de forma única a un nivel educativo.
    - Descripción: en él se describe el nivel educativo correspondiente.
  + Perpectiva Departamento:
    - IdDepartamento: es la clave primaria de la tabla “Departamento” y representa de forma única a un departamento.
    - Nombre: en él se indica el nombre del departamento.
  + Perspectiva Tipo Matricula:
    - idTipo\_Matricula: es la clave primaria de la tabla Tipo\_Matricuala y representa de forma única a un tipo de matrícula.
    - Descripción: en él se describe el tipo de matrícula.
  + Perspectiva Fecha Matriculacion:
    - Numero año
    - Numero mes
    - Numero día
  + Perspectiva Fecha Ingreso a la Base de datos:
    - Numero año
    - Numero mes
    - Numero día
    1. **Modelo Conceptual ampliado**

En este punto se ampliara el modelo conceptual anterior colocando bajo cada perspectiva los campos elegidos para cada una.



* 1. **Modelo Lógico del Data warehouse**

En este paso se presentan los resultados obtenidos de los anteriores pasos, se ampliara el modelo conceptual colocando bajo cada perspectiva los campos seleccionados.

* + 1. **Tipo de modelo lógico del Data warehouse**

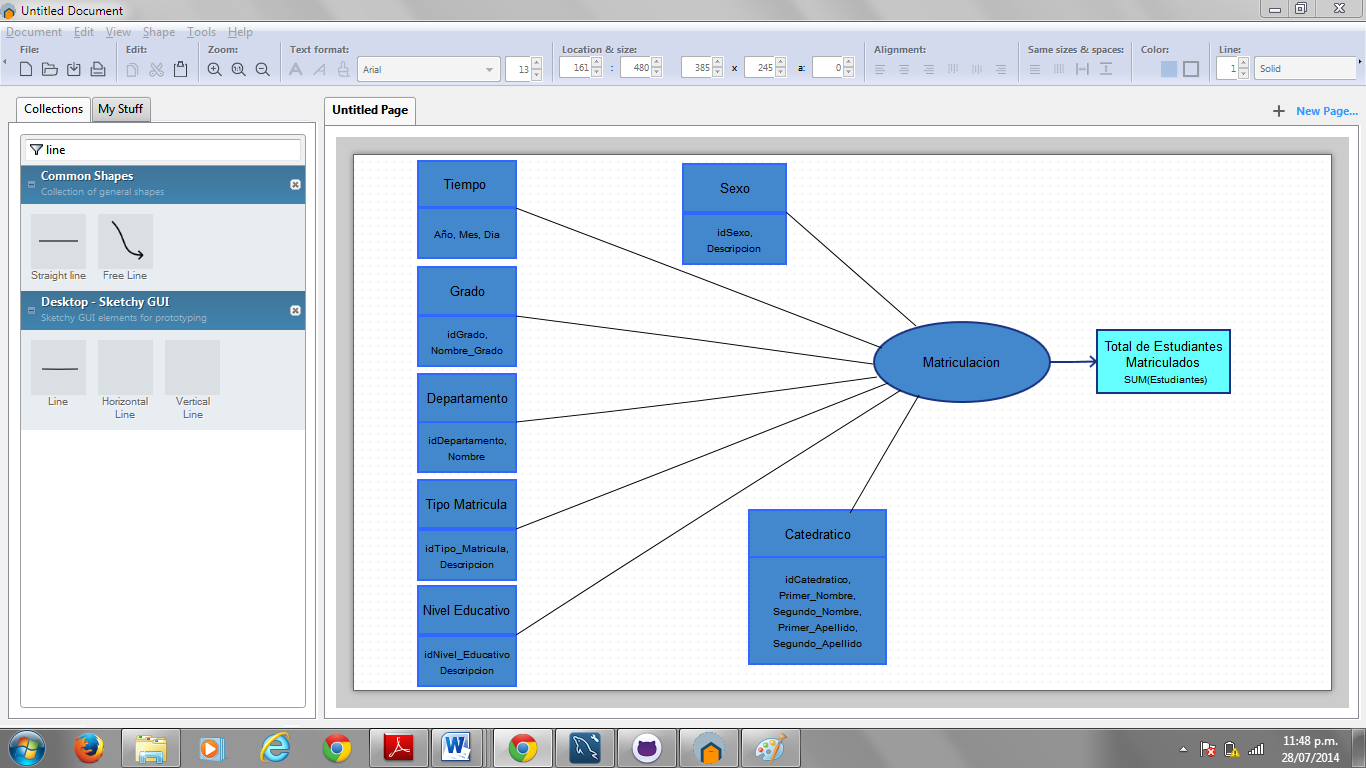
Primero que nada elegimos esquema de para el desarrollo del datawarehouse, en nuestro caso utilizaremos el esquema de estrella debido a su simplicidad tanto en la comprensión como en la aplicación.

* + 1. **Tabla de dimensiones**

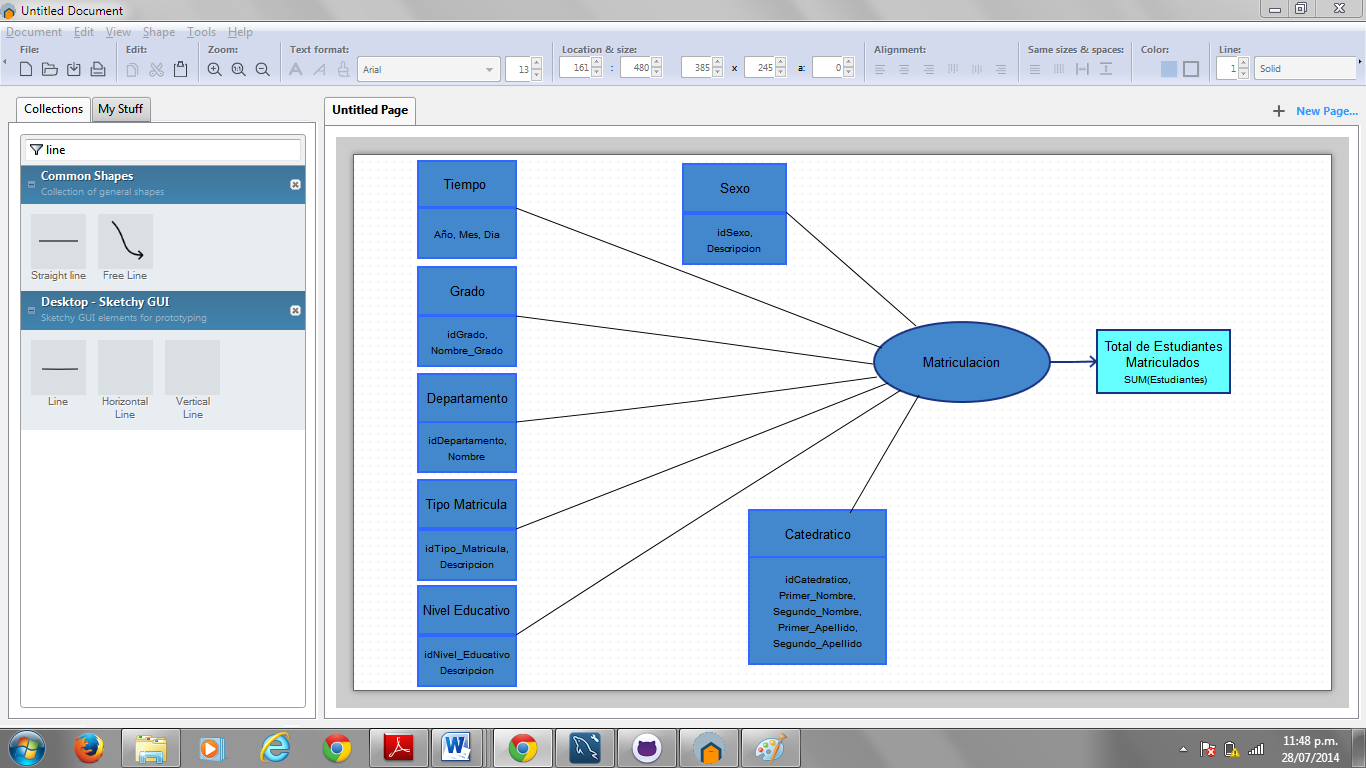
En este punto se diseñan las tablas de dimensiones que forman parte del proceso de matriculación dentro del Data warehouse, para lo cual se debe tomar en cuenta cual será el nombre de la tabla, añadir el campo que será la clave primaria y los nombres a los demás campos de la tabla.

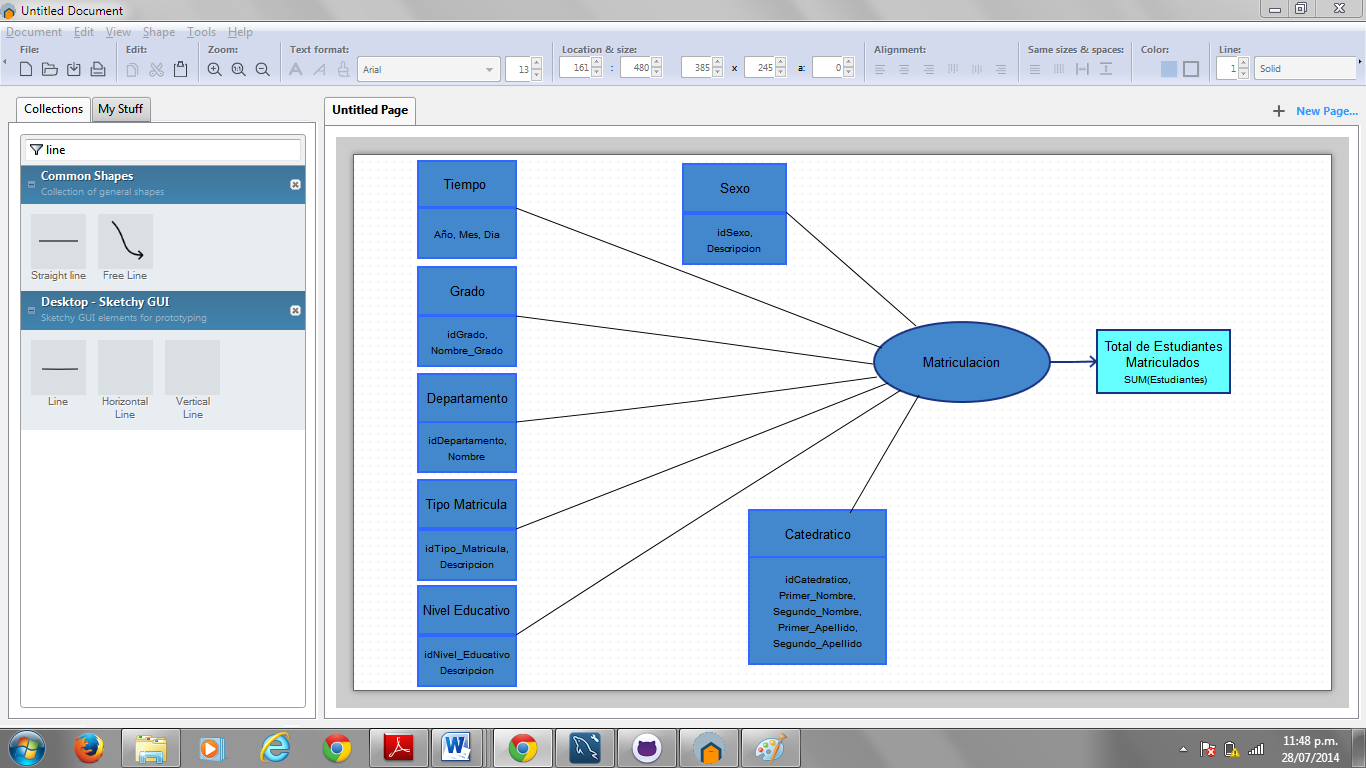
* + Perspectiva Grado: se creara tabla de dimensión con el nombre “Grado” la cual hace referencia a los diferentes grados en los que se pueden matricular los alumnos y contiene los siguientes atributos:
    - La clave primaria “idGrado”

|  |
| --- |
| Grado |
| idGrado |
| Nombre\_Grado |

* + - El campo “Nombre\_Grado” el cual tendrá el nombre del grado.
  + Perspectiva Sexo: se creara la tabla de dimensión con el nombre “Sexo” la cual hace referencia al género del estudiante y contiene los siguientes atributos:
    - La clave primaria “idSexo”
    - El campo “Descripción” el cual contiene la descripción del género al que pertenece el alumno.

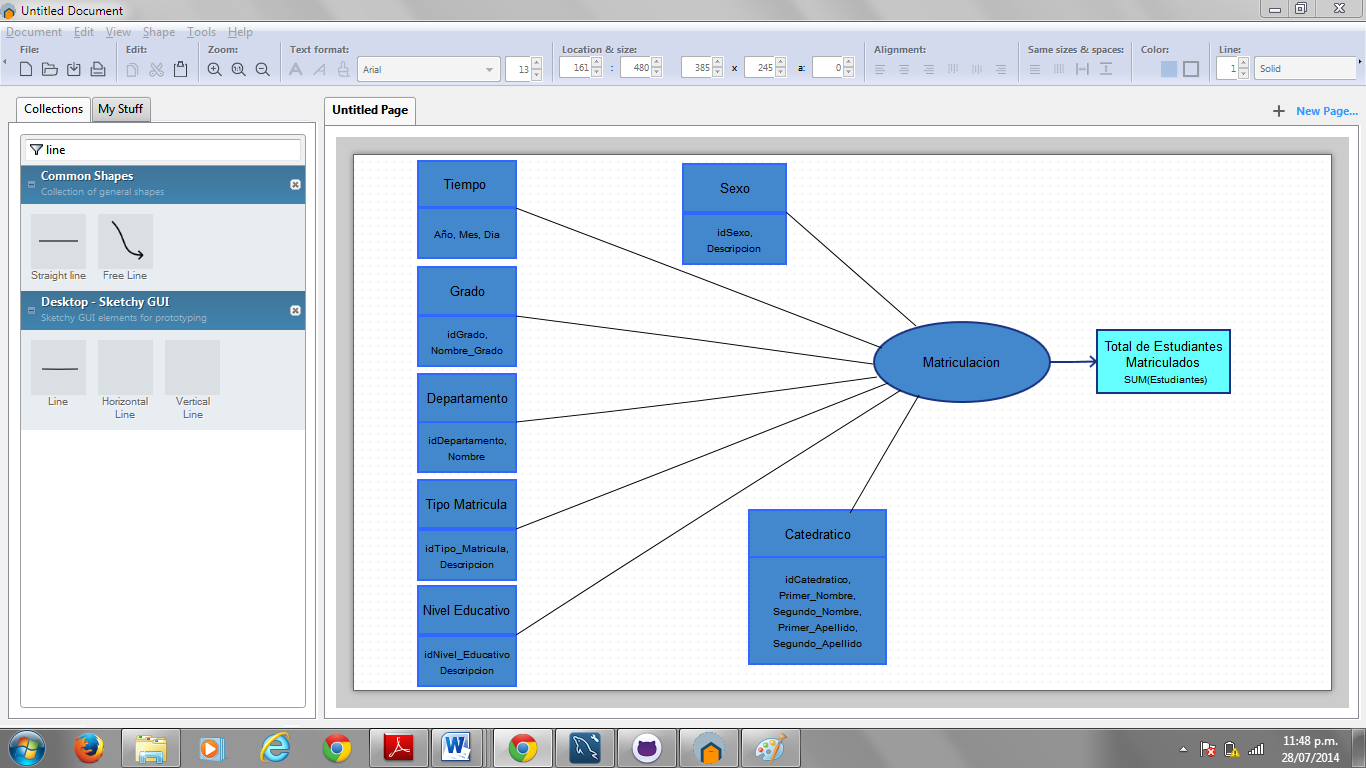
|  |
| --- |
| Sexo |
| idSexo |
| Descripción |



* + Perspectiva Catedrático: se creara la tabla de dimensión con el nombre “Catedrático” la cual hace referencia al genero del estudiante y contiene los siguientes atributos:
    - La clave primaria “idCatedratico”
    - El campo “Primer\_Nombre” representa el primer nombre del catedrático
    - El campo “Segundo\_Nombre” representa el segundo nombre del catedrático
    - El campo “Primer\_Apellido” representa el primer apellido del catedrático.
    - El campo “Segundo\_Apellido representa el Segundo apellido del catedrático.

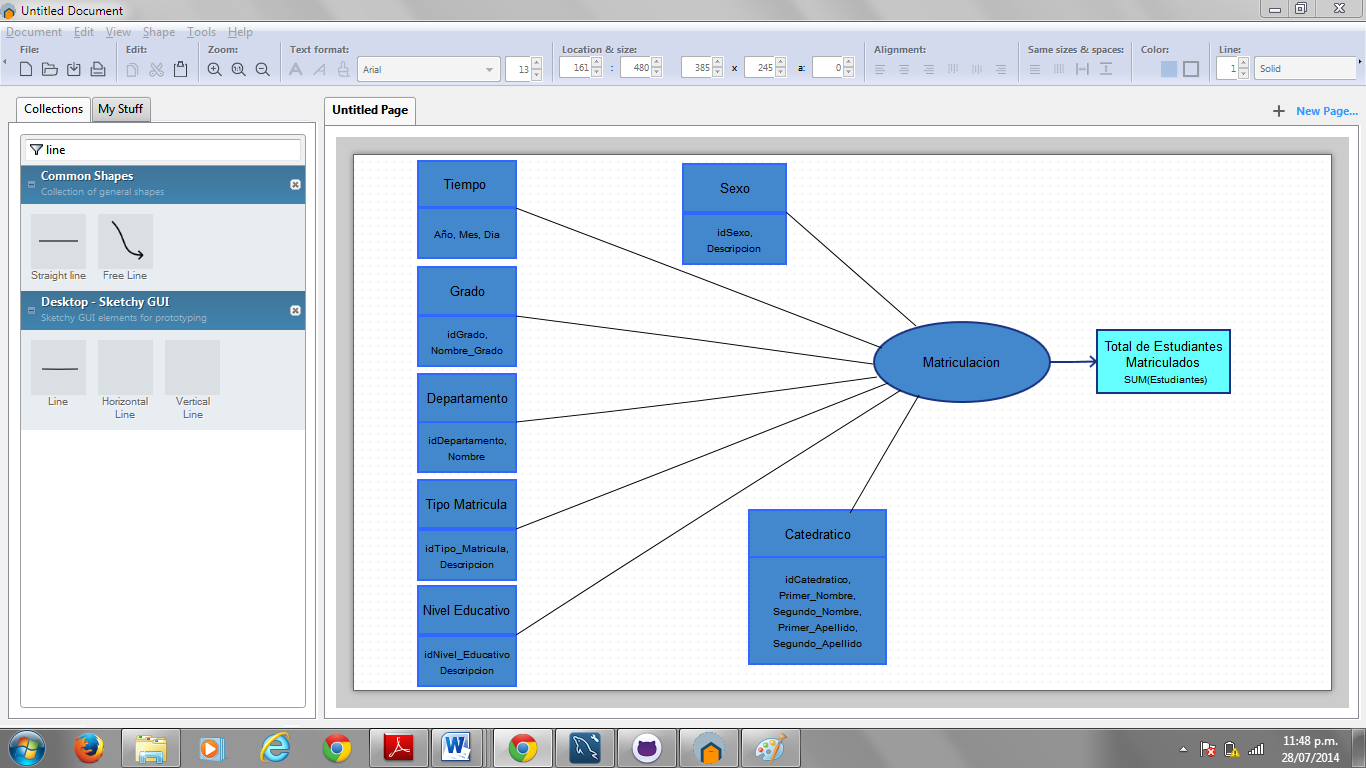
|  |
| --- |
| Catedratico |
| idCatedratico |
| Primer\_Nombre |
| Segundo\_Nombre |
| Primer\_Apellido |
| Segundo\_Apellido |

* + Perspectiva Nivel Educativo: se creara la tabla de dimensión con el nombre “Nivel Educativo” la cual hace referencia al nivel educativo que pertenecen los distintos grados y contiene los siguientes atributos:
    - La clave primaria “idNivel\_Educativo”.
    - El campo “Nombre” que representa el nombre de cada los distintos niveles educativos.



|  |
| --- |
| Nivel\_Educativo |
| IdNivel\_Educativo |
| Descripcion |

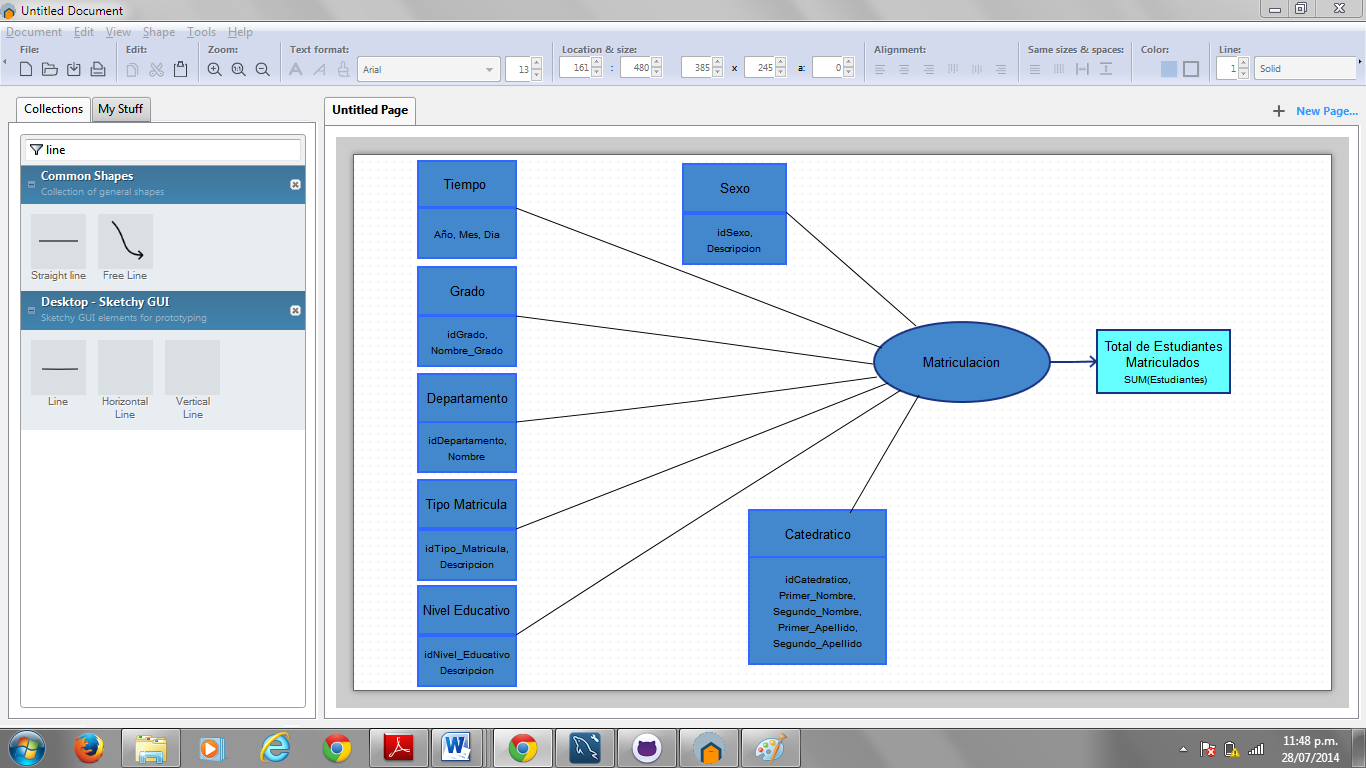
* + Perspectiva Departamento: se creara la tabla de dimensión con el nombre “Departamento” la cual hace referencia al departamento al cual pertenece el estudiante que se matriculara u contiene los siguientes atributos:
    - La clave primaria “idDepartamento”
    - El campo “Nombre” que representa el nombre que tiene cada departamento.



|  |
| --- |
| Departamento |
| idDepartamento |
| Nombre |

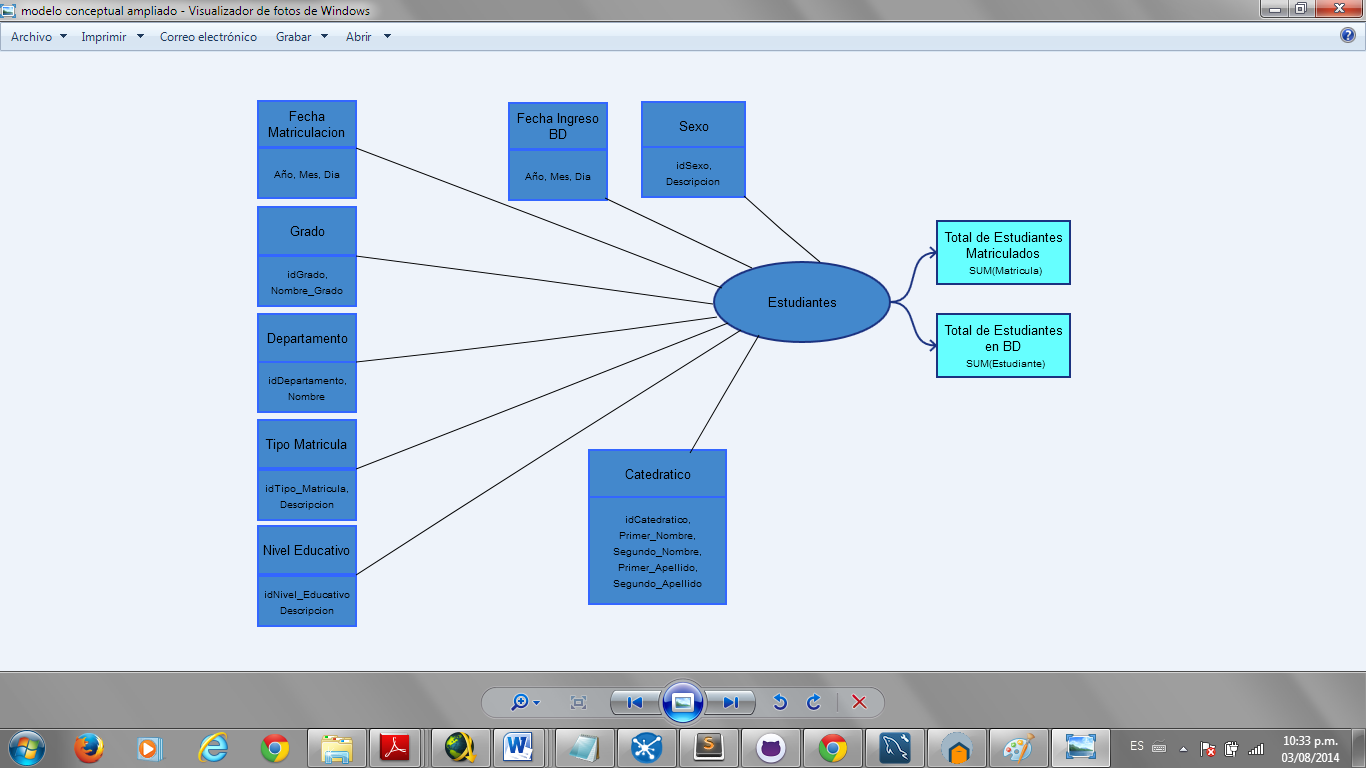
* + Perspectiva Tipo\_Matricula: se creara la tabla de dimensión con el nombre “Tipo\_Matricula” la cual hace referencia al tipo de matrícula con el que se pueden matricular los estudiantes y contiene los siguientes atributos:
    - La clave primaria “idTipo\_Matricula”
    - El campo “Descripcion” que representa la descripción de los tipos de matricualas que existen.

|  |
| --- |
| Tipo\_Matricula |
| idTipo\_Matricula |
| Descripcion |



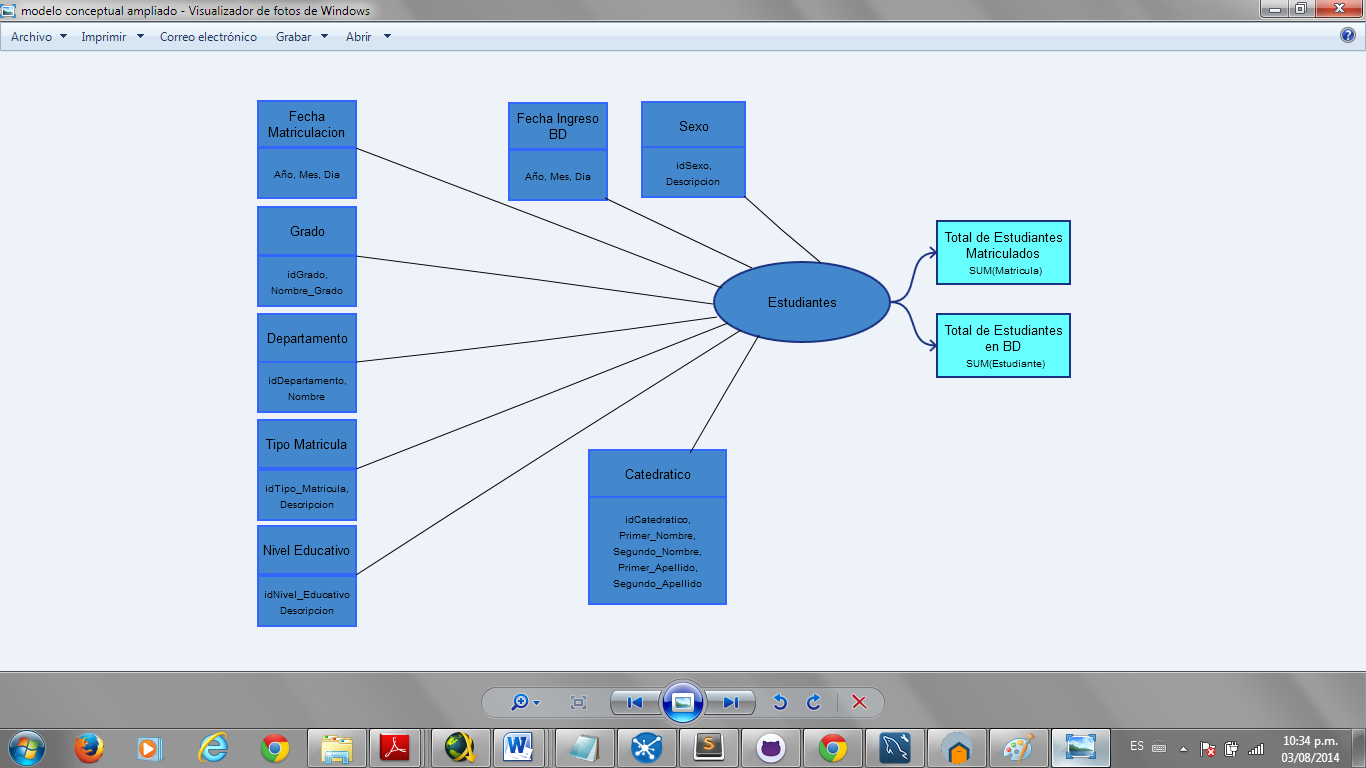
* + Perspectiva Fecha Matriculación: se creara la tabla de dimensión con el nombre “Fecha Matriculación” con los siguientes atributos:
    - Año
    - Mes
    - Día

|  |
| --- |
| Fecha\_Matriculacion |
| idFecha |
| Año |
| Mes |
| Dia |



* + Perspectiva Fecha Ingreso a Base de datos: se creara la tabla de dimensión con el nombre “Fecha Ingreso a base de datos” con los siguientes atributos:
    - Año
    - Mes
    - Día

|  |
| --- |
| Fecha\_Ingreso\_a\_Base\_Datos |
| idFecha |
| Año |
| Mes |
| Dia |



* + 1. **Tabla Hechos**

En este punto se define la tabla de hechos, que es la que contendrá los hechos a través de los cuales se construirá el indicador de estudio. Para la cual, se debe de tomar en cuenta cual será el nombre de la tabla, así como también la clave primaria y los respectivos nombres a los campos.

* + La tabla de hecho tendrá el nombre de Estudiantes:
    - La clave principal será la combinación de las claves primarias de las tablas de dimensiones que establecimos en el paso anterior.
    - Se creara un hecho que corresponde al indicador (Total de estudiantes Matriculados) y será renombrado por Cantidad Matriculados.
    - Se creara un hecho que corresponde al indicador (Total de estudiantes en Base de datos) y será renombrado por Cantidad en Base de datos

|  |
| --- |
| Matriculacion |
| IdGrado |
| idSexo |
| idCatedratico |
| IdNivel\_Educativo |
| idDepartamento |
| IdTipo\_Matricula |
| idFecha |
| Cantidad\_Matriculados |
| Cantidad\_en\_BasedeDatos |

* + 1. **Uniones**

Se realizaron las uniones correspondientes entre las tablas de dimensiones y la tabla de hechos.

