|  |
| --- |
| Universidad del Quindío |
| Definición de la Arquitectura técnica. |
|  |
|  |

|  |
| --- |
| David Felipe Hernández  21 octubre 2016 |

**Tabla de Contenido.**

[HISTORIAL DE REVISIONES 2](#_Toc464843003)

[INTRODUCCIÓN 3](#_Toc464843004)

[**Definiciones, acrónimos y abreviaciones.** 3](#_Toc464843006)

[Arquitectura Técnica. 4](#_Toc464843007)

[**Fuentes de datos.** 4](#_Toc464843008)

[**Herramientas para el desarrollo de la implementación.** 10](#_Toc464843009)

[**Reglas ETL** 11](#_Toc464843010)

[Reglas de Extracción. 11](#_Toc464843011)

[Reglas de Transformación. 12](#_Toc464843012)

[Reglas de Carga de datos 14](#_Toc464843013)

[**Políticas de Uso** 14](#_Toc464843014)

[Calendarios de Carga 14](#_Toc464843015)

[Responsabilidades de los funcionarios. 15](#_Toc464843016)

# HISTORIAL DE REVISIONES

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fecha | | Versión | Descripción | Autor |
| 21/Octubre/2016 | 1.0 | | Desarrollo de la arquitectura | David Hernández |
|  |  | |  |  |
|  |  | |  |  |
|  |  | |  |  |

# INTRODUCCIÓN

## 

## En este documento se determinan las fuentes de datos, las reglas de extracción, transformación y carga (ETL) y las políticas de uso del Datamart.

## **Definiciones, acrónimos y abreviaciones.**

**Indicador:** Representan lo que se desea analizar concretamente, por ejemplo: saldos, promedios, cantidades, sumatorias, entre otros.

**Perspectiva:** Se refieren a los objetos mediante los cuales se requiere examinar los indicadores, por ejemplo: Estudiante, Facultad, Programa, Proveedores entre otros.

# Arquitectura Técnica.

## **Fuentes de datos.**

En esta sección se especifican las fuentes de datos que se tendrán en cuenta para la implementación del Datamart, su ubicación, tipo de información, motor de base de datos y estructura.

**Fuente de datos 01.**

**Nombre:** Base de datos del sistema Emisor y facturador de entradas.

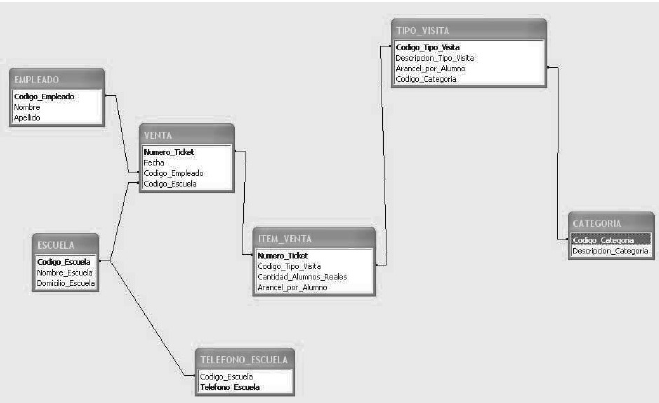
**Ubicación**: Localhost

**Tipo de Información**:

* Información de ventas de entradas al parque a escuelas

**Motor de base de datos**: MySQL

**Estructura**:



**Fuente de datos 02.**

**Nombre:** Base de datos del Sistemas de facturación 1.

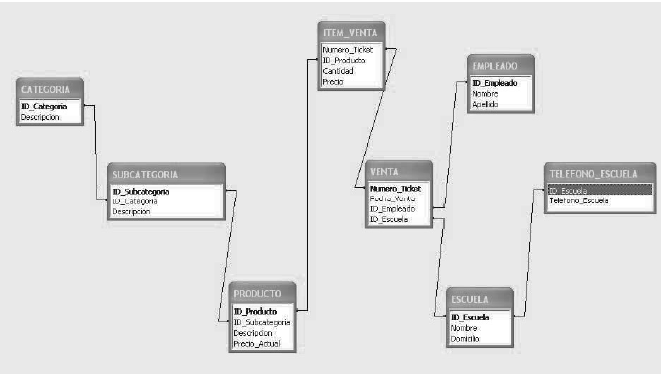
**Ubicación**: Localhost

**Tipo de Información**:

* Información de ventas de productos a escuelas

**Motor de base de datos**: MySQL

**Estructura**:



**Fuente de datos 03.**

**Nombre:** Base de datos del Sistemas de facturación 2.

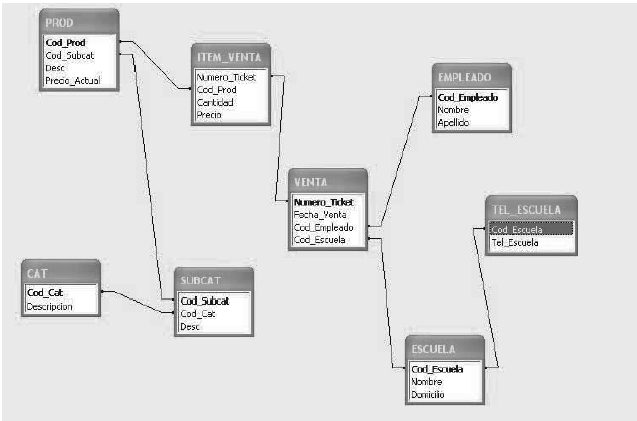
**Ubicación**: Localhost

**Tipo de Información**:

* Información de ventas de productos a escuelas

**Motor de base de datos**: MySQL

**Estructura**:



**Fuente de datos 04.**

**Nombre:** Base de datos del Sistema de recursos humanos.

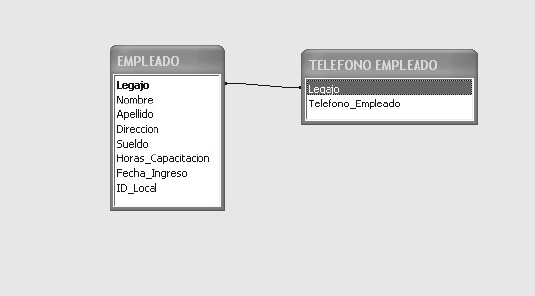
**Ubicación**: Localhost

**Tipo de Información**:

* Información de los empleados del parque

**Motor de base de datos**: MySQL

**Estructura**:



**Fuente de datos 05.**

**Nombre:** Parques.

**Ubicación**: Local

**Tipo de Información**:

* Información de los parques

**Tipo de archivos**: Excel

**Estructura**:



**Fuente de datos 06.**

**Nombre:** Empleados.

**Ubicación**: Local

**Tipo de Información**:

* Información de los empleados

**Tipo de archivos**: Excel

**Estructura**:



**Fuente de datos 07**

**Nombre:** Regiones

**Ubicación**: Local

**Tipo de Información**:

* Información de las regiones.

**Tipo de archivos**: Excel

**Estructura**:



**Fuente de datos 08**

**Nombre:** Locales

**Ubicación**: Local

**Tipo de Información**:

* Información de los locales de los parques.

**Tipo de archivos**: Excel

**Estructura**:



**Fuente de datos 09**

**Nombre:** Escuelas

**Ubicación**: Local

**Tipo de Información**:

* Información adicional de las escuelas.

**Tipo de archivos**: Excel

**Estructura**:



**Fuente de datos 10**

**Nombre:** Clientes

**Ubicación**: Local

**Tipo de Información**:

* Información de los clientes.

**Tipo de archivos**: Excel

**Estructura**:



**Fuente de datos 11**

**Nombre:** VentasProductosClientes

**Ubicación**: Local

**Tipo de Información**:

* Información de las ventas de productos que se hacen a clientes.

**Tipo de archivos**: Excel

**Estructura**:



**Fuente de datos 12**

**Nombre:** VentasEntradasClientes

**Ubicación**: Local

**Tipo de Información**:

* Información de las ventas a entradas que se hacen a clientes.

**Tipo de archivos**: Excel

**Estructura**:



**Fuente de datos 13**

**Nombre:** VentasProductosEscuelas

**Ubicación**: Local

**Tipo de Información**:

* Información de las ventas de productos que se hacen a escuelas.

**Tipo de archivos**: Excel

**Estructura**:



**Fuente de datos 14**

**Nombre:** VentasEntradasEscuelas

**Ubicación**: Local

**Tipo de Información**:

* Información de las ventas de entradas que se hacen a escuelas.

**Tipo de archivos**: Excel

**Estructura**:



**Fuente de datos 15**

**Nombre:** Stock

**Ubicación**: Local

**Tipo de Información**:

* Stock disponible diario de cada local.

**Tipo de archivos**: Excel

**Estructura**:



## **Herramientas para el desarrollo de la implementación.**

En la siguiente tabla se listan las herramientas seleccionadas para la implementación del Datamart.

|  |  |  |
| --- | --- | --- |
| Id. | Herramienta | Uso |
| 1 | SQL server management Studio 2012 | Es un entorno integrado para acceder, configurar, gestionar, administrar, y el desarrollo de todos los componentes de SQL Server. |
| 2 | SQL Server datatools | Permite diseñar e implementar cualquier tipo de base de SQL Server con la misma facilidad como lo haría desarrollar una aplicación en Visual Studio. |

## **Reglas ETL**

En esta sección se indican las características puntuales que se deben tener en cuenta al realizar la extracción de datos, la transformación y su carga a la bodega de datos.

### Reglas de Extracción.

**Fecha y horario**

**Código**: 01

**Descripción**: Se establece un horario para realizar la extracción de los datos, estos datos se cargaran al área de espera únicamente los lunes a partir de las 12 del mediodía y con un plazo hasta las 6 de la tarde, deben ser cargados a un área de espera. Una vez estén en este sitio, se puede proceder a la transformación en cualquier momento de la semana.

**Fuente**: Requisito de parte del cliente.

**Datos nulos**

**Código**: 02

**Descripción**: Cuando se reciba un dato nulo, se enviara a un archivo de texto plano o a una hoja de cálculo donde deberá ser revisado por una persona y luego se vuelve a poner en el área de espera donde volverá a ser procesado [regla 03].

**Fuente**: por requisito de parte del cliente.

**Datos con otros errores**

**Código**: 03

**Descripción**: Cuando se reciba un dato con cualquier tipo de error o un dato no reconocido se enviara a un archivo de texto plano o a una hoja de cálculo donde será revisado por una persona la cual lo puede corregir si es posible o lo rechaza completamente y deberá reportar el error y dejarlo escrito.

**Fuente**: por requisito de parte del cliente.

**Valores duplicados**

**Código**: 04

**Descripción**: Cuando se reciba un valor duplicado será tomado como error y se procesara como lo dice la regla 03. Hay que tener en cuenta que algunos registros con la clave primaria duplicada pueden tener información distinta, por tanto se puede almacenar en la bodega pero con fecha de vigencia diferente.

**Fuente**: por requisito de parte del cliente.

**Datos incompletos**

**Código**: 05

**Descripción**: cuando un dato este incompleto, por ejemplo la fecha, se enviara a un archivo de texto plano o a una hoja de cálculo donde deberá ser revisado por una persona y luego se vuelve a poner en el área de espera donde volverá a ser procesado [regla 03].

**Fuente**: por requisito de parte del cliente.

**Datos sin sentido**

**Código**: 06

**Descripción**: los datos sin sentido, se enviara a un archivo de texto plano o a una hoja de cálculo donde deberá ser revisado por una persona y luego se vuelve a poner en el área de espera donde volverá a ser procesado [regla 03].

**Fuente**: por requisito de parte del cliente.

**Volumen**

**Código**: 07

**Descripción**: El tamaño de la carga o la cantidad de registros no está restringida, en este caso se deben pasar todos los registros de la semana los días lunes como se indicó en la regla 01, independientemente de la cantidad de datos, todos deben ser pasados al área de espera.

**Fuente**: por requisito de parte del cliente.

### Reglas de Transformación.

**Orden de las transformaciones**

**Código**: 08

**Descripción**: Primero se realizaran las transformaciones a las dimensiones y luego a la tabla de hechos, en caso de que se encuentren copos de nieve, se realizara la transformación primero a estas tablas, el orden de las transformaciones será de la siguiente manera: primero venta de productos a escuelas, ventas de entradas a escuelas, luego venta de productos a clientes, venta de entradas a clientes y finalmente el stock de productos. Se cargaran primero los datos más recientes hasta llegar a los más antiguos.

**Fuente**: por requisito de parte del cliente.

**Formatos a tomar**

**Código**: 09

**Descripción**: Los formatos a tomar son solo siguientes:

* **Género**: el género se representa con el nombre completo ejemplo : Masculino o Femenino
* **Nombre completo**: primero el nombre , luego el apellido ejemplo : David Hernández
* **Mes**: nombre completo ejemplo: octubre
* **Año**: 4 dígitos ejemplo: 2016
* **Fecha:** separado por el símbolo “/” ejemplo 02/noviembre/2016
* **Día** **de** **la** **semana**: nombre completo ejemplo: lunes
* **Costos y valores**: se tomaran los dígitos enteros sin punto y para decimales con “.” ejemplo: 23000
* **Horas capacitación:** se tomara un entero que represente el número de horas, este campo no debe llevar ninguna letra, solo un máximo de tres dígitos enteros.
* **Zona\_geografica:** es el nombre completo de la zona geográfica, en el proceso de extracción se verifica que corresponda a un nombre de una zona geográfica ej. Sur America,norte America,Centro América ,Europa,Asia,Africa,Oceania,

**Fuente**: por requisito de parte del cliente.

**Rango de datos**

**Código**: 10

**Descripción**: Los rangos son los siguientes

* **Rango de edad**: fecha nacimiento entre el año 1900 hasta el año actual, en caso de encontrarse un valor fuera del rango o ser null se asignara el año de nacimiento promedio actual y se omitirá en este caso la regla 3.
* **Rango cantidad de horas de capacitación:** el rango de horas de capacitación es de 0 a 1000, si llega un valor nulo se cambiara a cero, de lo contrario debe seguir el protocolo como error [regla 3].
* **Rango de ventas de un día:** el rango de ventas de un día está entre 0 y 10000000,valores por fuera del rango deberán seguir el protocolo como error [regla 3].
* **distanciaBoleteria:** el rango de distancia a la boletería es de 1 a 80, valores por fuera del rango serán tomados como error y deberán ser analizados por una persona [regla 3].
* **Tamaño:** el rango de tamaño de un local es de 1 a 200, valores por fuera del rango serán tomados como error y deberán ser analizados por una persona [regla 3].
* **fechaIngreso**: el año de la fecha de ingreso debe estar en el rangos de 1990 hasta el año actual, valores por fuera del rango serán tomados como error y deberán ser analizados por una persona [regla 3].
* **Sueldo:** el sueldo de un empleado debe estar en el rango 200000 a 10000000, valores por fuera del rango serán tomados como error y deberán ser analizados por una persona [regla 3].
* **costoActual de un producto:** El valor del costo de un producto debe estar entre 100 y 200000, valores por fuera del rango serán tomados como error y deberán ser analizados por una persona [regla 3].

**Fuente**: por requisito de parte del cliente.

### Reglas de Carga de datos

**Horarios de carga de datos**

**Código**: 11

**Descripción**: los datos se cargaran a la bodega de datos una vez cumplan todos los pasos sin generar error, el proceso de enviar los datos a la bodega se puede realizar cualquier día de la semana.

**Fuente**: por requisito de parte del cliente.

### Otras reglas

**Otras reglas importantes**

**Código**: 12

**Descripción**: se listan otras reglas importantes a tener en cuenta en el proceso del ETL

* El **valor de entrada** por alumno será sacado de la tabla item\_venta la columna Arancel\_por\_alumno
* **Fk\_escuela:** de la dimensión cliente, este valor corresponde al id de la escuela donde estudia el cliente, en caso de que el id no se encuentre o sea nulo, se le asignara la escuela “sin escuela” por defecto.

**Fuente**: por requisito de parte del cliente.

## **Políticas de Uso**

En esta sección se describen las políticas de uso que se tienen para el momento en que el Datamart este implementado.

### Calendarios de Carga

|  |  |  |  |
| --- | --- | --- | --- |
| Código | Fecha | Descripción de los datos a Cargar | Regla Asociada |
| 01 | 07-noviembre-2016 | Se cargaran los datos de las bases de datos nombradas anteriormente, este proceso puede tardar más de dos días ya que es la primera carga, en tal caso se destinaran los días 07 y 08 de noviembre para esta tarea. | Regla 01 |
| 02 | 10-noviembre-2016 | En esta fecha se realizara una carga normal de datos, como ya se ha hecho una carga anteriormente pueden surgir errores como datos repetidos, en caso de que ocurran errores considerables se podrá tomar el día 11 de noviembre para completar la carga. | Regla 01  Regla 04 |

### Responsabilidades de los funcionarios.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Id | Responsabilidad | Nombre del funcionario | Rol del funcionario | Dependencia del funcionario |
| 01 | Aseguramiento de la calidad de los datos correspondientes a la venta de productos en los locales del parque | Juan Carlos | Administrador de los locales de ventas de productos | Ventas |
| 02 | Manejo del control de stock de los locales | Juan Carlos | Administrador de los locales de ventas de productos | Ventas |
| 02 | Aseguramiento de la calidad de datos correspondientes a la venta de entradas al parque | Daniela Marín | Jefe encargado de ventas de boletas de entrada al parque | Ventas |