

# **ESCUELA POLITÉCNICA NACIONAL**

**FACULTAD DE INGENIERÍA DE SISTEMAS**

**APLICACION DE DATA MINING PARA ESTABLECER  
PATRONES DE COMPORTAMIENTO DE DATOS  
METEOROLOGICOS PARA LA CIUDAD DE QUITO**

**PROYECTO PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN SISTEMAS  
INFORMÁTICOS Y DE COMPUTACIÓN**

**BLANCA JEANNETH AMORES HURTADO**

**blaqui\_amores@hotmail.com**

**DIEGO FERNANDO CRUZ CASA**

**diegocruz@samerica.com**

**DIRECTOR: Ing. Francisco Villavicencio**

**fvillavicencio@epn.edu.ec**

**Quito, Marzo del 2008**

## **DECLARACIÓN**

Nosotros, Blanca Jeanneth Amores Hurtado y Diego Fernando Cruz Casa, declaramos bajo juramento que el trabajo aquí descrito es de nuestra autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que hemos consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedemos nuestros derechos de propiedad intelectual correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

---

Blanca Jeanneth Amores Hurtado

---

Diego Fernando Cruz Casa

## **CERTIFICACIÓN**

Certifico que el presente trabajo fue desarrollado por Blanca Jeanneth Amores Hurtado y Diego Fernando Cruz Casa, bajo mi supervisión.

---

**Ing. Francisco Villavicencio**  
**DIRECTOR DE PROYECTO**

## INDICE

ÍNDICE DE FIGURAS .....	6
ÍNDICE DE TABLAS .....	7
INTRODUCCIÓN .....	8
CAPITULO 1. ANALISIS DEL CASO DE ESTUDIO .....	10
1.1 OBSERVATORIO ASTRONÓMICO DE QUITO COMO FUENTE DE INFORMACIÓN METEOROLÓGICA .....	10
1.1.1 INICIOS .....	10
1.1.2 PARTICULARIDADES CLIMATOLÓGICAS DE LA CIUDAD DE QUITO.....	11
<b>2.1.3.1 Cambios Estacionales De La Humedad</b> .....	11
<b>2.1.3.2 Cambios de Temperatura</b> .....	12
1.1.3 OBSERVACIONES METEOROLÓGICAS .....	13
1.2 ELEMENTOS METEOROLÓGICOS .....	18
1.2.1 TEMPERATURA.....	18
1.2.2 PRESIÓN ATMOSFÉRICA .....	18
1.2.3 HUMEDAD.....	19
<b>2.1.3.1 Saturación</b> .....	19
<b>2.1.3.2 Punto de Rocío</b> .....	20
1.2.4 PRECIPITACIONES. ....	20
CAPITULO 2. IMPLEMENTACIÓN DEL DATA WAREHOUSE .....	22
2.1. ANÁLISIS DE REQUERIMIENTOS .....	22
2.1.1. ANTECEDENTES .....	22
2.1.2. SELECCIÓN DE ELEMENTOS METEOROLÓGICOS .....	22
2.1.3. METODOLOGÍA DE DESARROLLO DEL DATA WAREHOUSE..	23
<b>2.1.3.1 Metodología en Espiral</b> .....	25
2.2. CONSTRUCCIÓN DE UN DATA WAREHOUSE .....	34
2.2.1. DISEÑO PRELIMINAR .....	34
2.2.2. MODELADO DE LOS DATOS .....	38
2.2.2.1. Modelo para la fuente de datos.....	38
2.2.2.2. Modelo para el Data Warehouse .....	44
2.2.3. DISEÑO FISICO PARA EL DATA WAREHOUSE.....	47
2.2.4. DEFINICIÓN DE LAS FUENTES DE DATOS .....	49
2.2.5. EXTRACCIÓN, TRANSFORMACIÓN Y CARGA DE DATOS.....	51
2.2.6. ANÁLISIS DE RESULTADOS.....	53
2.2.7. NUEVA ESPECIFICACIÓN DE DISEÑO .....	57
2.2.8. INGRESO DE NUEVOS DATOS .....	60
CAPITULO 3. MINERIA DE DATOS .....	67
3.1. EXPLORACIÓN Y SELECCIÓN DE DATOS .....	67
3.1.1. VISTA MINABLE .....	68
3.1.2. RECONOCIMIENTO DEL DOMINIO Y DE LOS USUARIOS.-.....	68
3.2. MINERÍA DE DATOS .....	69
3.2.1. TAREAS DE MINERÍA DE DATOS .....	70

3.2.2.	TÉCNICAS DE MINERÍA DE DATOS .....	71
3.2.3.	CONSTRUCCIÓN DEL MODELO Y VALIDACIÓN. ....	72
3.3.	RECONOCIMIENTO DE PATRONES DE COMPORTAMIENTO.....	73
3.4.	REPORTE CON LOS RESULTADOS OBTENIDOS .....	79
	Temperatura.....	79
	Humedad.....	79
	Presión .....	80
	Cantidad de Lluvia .....	80
CAPITULO 4.	CONCLUSIONES Y RECOMENDACIONES .....	82
4.1.	CONCLUSIONES .....	82
4.2.	RECOMENDACIONES .....	83
BIBLIOGRAFIA	.....	84
ANEXOS.	.....	85
ANEXO 1:	REPORTES METEOROLÓGICO.....	86
ANEXO 2:	DETALLE DE LOS PROBLEMAS .....	93
ANEXO 3:	CREACIONES DE ESQUEMAS .....	95
ANEXO 4:	CARGA DE DATOS BASE TRANSACCIONAL .....	98
ANEXO 5:	REPORTES DATA WAREHOUSE.....	105
ANEXO 6:	GUIA DE USO DE CLEMENTINE 11.1 .....	109
ANEXO 7:	VISTAS MINABLES.....	117
ANEXO 8:	PLANTILLA PARA EL INGRESO DE DATOS .....	119

## ÍNDICE DE FIGURAS

<i>Figura 1.1 Barómetro.....</i>	<i>14</i>
<i>Figura 1.2 Evaporímetro.....</i>	<i>14</i>
<i>Figura 1.3 Pluviómetro.....</i>	<i>15</i>
<i>Figura 1.4 Psicrómetro.....</i>	<i>16</i>
<i>Figura 1.5 Termógrafo e Higrógrafo.....</i>	<i>16</i>
<i>Figura 1.6 Garita meteorológica.....</i>	<i>17</i>
<i>Figura 2.1 Metodología para el desarrollo del Data Warehouse.....</i>	<i>24</i>
<i>Figura 2.2 Entidad.....</i>	<i>27</i>
<i>Figura 2.3 Relación.....</i>	<i>27</i>
<i>Figura 2.4 Atributo y Conector.....</i>	<i>27</i>
<i>Figura 2.5 Esquema Estrella.....</i>	<i>29</i>
<i>Figura 2.6 Esquema Copo de Nieve.....</i>	<i>30</i>
<i>Figura 2.7 Bandas Meteorológicas.....</i>	<i>35</i>
<i>Figura 2.8 Hojas de Cálculo.....</i>	<i>35</i>
<i>Figura 2.9 Reportes Estadísticos.....</i>	<i>36</i>
<i>Figura 2.10 Hoja de Cálculo.....</i>	<i>38</i>
<i>Figura 2.11 Modelo Entidad Relación.....</i>	<i>40</i>
<i>Figura 2.12 Modelo OLTP.....</i>	<i>40</i>
<i>Figura 2.13 Modelo Dimensional con esquema copo de nieve.....</i>	<i>44</i>
<i>Figura 2.14 Nivel de detalle Tiempo.....</i>	<i>46</i>
<i>Figura 2.15 Extracción, Transformación y Carga de Datos [6].....</i>	<i>51</i>
<i>Figura 3.1 Proceso para tener datos y Aplicar una herramienta de Minería de Datos.....</i>	<i>67</i>
<i>Figura 3.2 Conexión al la Vista Minable Vista.....</i>	<i>74</i>
<i>Figura 3.3 Nodo Seleccionar.....</i>	<i>75</i>
<i>Figura 3.4 Nodo Gráfico 3D.....</i>	<i>75</i>
<i>Figura 3.5 Temperatura y Humedad.....</i>	<i>76</i>
<i>Figura 3.6 Temperatura Máxima.....</i>	<i>77</i>
<i>Figura 3.7 Temperatura Promedio.....</i>	<i>78</i>
<i>Figura 3.8 Temperatura Mínima.....</i>	<i>78</i>
<i>Figura 3.9 Temperatura Mínima, Promedio y Máxima por año.....</i>	<i>79</i>
<i>Figura 3.10 Humedad Mínima, Promedio y Máxima por año.....</i>	<i>79</i>
<i>Figura 3.11 Presión Mínima, Promedio y Máxima por año.....</i>	<i>80</i>
<i>Figura 3.12 Cantidad de Lluvia caída sobre el Quito por año.....</i>	<i>80</i>
<i>Figura 3.13 Cantidad de Lluvia caída sobre el Quito por separado por trimestres.....</i>	<i>81</i>

## ÍNDICE DE TABLAS

<i>Tabla 2.1 Elementos Meteorológicos</i> .....	37
<i>Tabla 2.2 Descripción de los atributos de la Entidades</i> .....	39
<i>Tabla 2.3 Descripción de los atributos de la Relación</i> .....	39
<i>Tabla 2.4 Usuarios</i> .....	40
<i>Tabla 2. 5 Elementos Meteorológicos</i> .....	61
<i>Tabla 2.6 Ingreso de datos</i> .....	63

## INTRODUCCIÓN

El presente proyecto, describe el proceso para aplicar Data Mining sobre datos meteorológicos que posee el Observatorio Astronómico de Quito (OAQ) desde el año de 1995 hasta el año 2005, con el objetivo de obtener información inferida a partir de datos almacenados en un Data Warehouse.

Para minar los datos se usará una herramienta de minería de datos tratando de extraer información oculta e identificar los patrones de comportamiento en base a la información del Data Warehouse, de esta forma se podrá obtener información para las personas interesadas en el tema, especialmente el personal encargado del Departamento Meteorológico del OAQ, ya que esta información es usada para la elaboración de boletines meteorológicos.

El Primer Capítulo realiza un análisis del caso de estudio, en este caso como el OAQ ha recolectado los datos y los instrumentos meteorológicos que se han empleado para estos fines, además se hace una breve explicación de las particularidades climatológicas de la ciudad de Quito y de los elementos meteorológico.

Los elementos meteorológicos de mayor influencia para nuestro proyecto de titulación son: la temperatura del aire, la humedad relativa, la cantidad de lluvia y la presión atmosférica, los mismos que son descritos en este capítulo.

El Segundo Capítulo desarrolla la implementación del Data Warehouse, para lo cual se debe realizar el análisis de requerimientos, la selección y limpieza de los datos meteorológicos los cuales se usaran para cargarlos en la base de datos, además se hace una breve descripción de la metodología a usarse, en este caso la Metodología en Espiral. Al mismo tiempo se realiza el modelado de datos y la descripción de cada elemento involucrado en este, así como los modelos disponibles para diseñar el Data Warehouse.

La construcción del Data Warehouse involucra la recopilación de información que en este caso proviene de archivos planos, bandas meteorológicas, reportes estadísticos. Este conjunto de datos se los denomina datos históricos los cuales han sido cargados en una base de datos transaccional, con el fin de transferirlos a una base de datos temporal la cual permite realizar la extracción, transformación y carga de datos al Data Warehouse, una vez realizado este proceso se procede a realizar el análisis de resultados, a través de reporte utilizando una herramienta de reporte.

En el Tercer Capítulo se describe la exploración y selección de datos con el objetivo de escoger un conjunto de datos (vista minable) que permitan realizar un análisis de los mismos, con la finalidad de obtener patrones de comportamiento, además se realiza una descripción de los modelos de minería de datos, así como las tareas y técnicas que pudieran ser usados por el conjunto de datos.

La construcción del modelo depende de que técnica y algoritmo de minería de datos se aplique sobre la vista minable que se menciona en este capítulo y la forma como se interpreta los resultados. La herramienta que se utiliza es CLEMENTINE 11.1, como se podrá notar esta herramienta tiene opciones que facilita el desarrollo del proceso de modelado e interpretación de patrones.

## **CAPITULO 1. ANALISIS DEL CASO DE ESTUDIO**

### **1.1 OBSERVATORIO ASTRONÓMICO DE QUITO COMO FUENTE DE INFORMACIÓN METEOROLÓGICA**

#### **1.1.1 INICIOS**

El Observatorio Astronómico de Quito (OAQ) fue fundado en la Presidencia del Dr. Gabriel García Moreno en el año de 1873. El 1ero de enero de 1891 se instaló una estación meteorológica, la cual ha trabajado de forma casi interrumpida hasta hoy en día, poseyendo de esta forma el banco de datos más completo del país.

Durante muchos años el OAQ se centró en recopilar información únicamente en el centro histórico de la capital del Ecuador pero al paso de los años y mediante acuerdos con otras entidades, se logró que se implementen estaciones meteorológicas en las capitales de provincia. El número de estaciones meteorológicas se incrementó en las distintas regiones del país con el objetivo de que sus habitantes conozcan las características reales de su clima y las puedan relacionar con la agricultura, ganadería y sus actividades productivas.

Cabe mencionar que desde 1891 el OAQ ha manejado la información meteorológica completa, siguiendo disposiciones internacionales para efectuar los trabajos meteorológicos, de esta forma se puede dar a los interesados, la información que requieran para poder realizar sus actividades productivas, industriales, deportivas, de navegación aérea, marítima y terrestre, etc. [1] [23]

---

[1] Álvaro, Marco; Boletín Meteorológico OAQ

[23] Observatorio Astronómico de Quito; Historia del Observatorio

### 1.1.2 PARTICULARIDADES CLIMATOLÓGICAS DE LA CIUDAD DE QUITO

La capital del Ecuador se encuentra situada al sur del Ecuador geográfico y a 2800 m de altitud sobre el nivel del mar, como consecuencia de estos dos factores, el centro de la ciudad presenta una temperatura media de 13 a 13.5 grados centígrados durante todo el año, pero debido a las condiciones topográficas locales y a otros factores, el valle de Quito presenta algunas características climatológicas especiales, que no se observan en otros valles interandinos del Ecuador, ni aun a igual altitud. [15] *Ver anexo 1*

Considerando que el conocimiento de dichas particularidades climáticas de la ciudad es de interés especial para todos los residentes de Quito, exponemos a continuación un resumen de las mismas.

#### 2.1.3.1 Cambios Estacionales De La Humedad

Se trata de una de las características climáticas más notorias del clima de esta ciudad. La alta humedad relativa del aire (normal en todo el trópico) está presente gran parte del año, pero entre junio y septiembre, y en especial durante los meses de julio y agosto; ocurren drásticos descensos en la humedad tanto relativa como absoluta, lo cual no se presenta en otras ciudades, ni aun en períodos de poca lluvia.

La mayoría de las ciudades del Ecuador presentan una humedad relativa media que siempre oscila entre 77 y 80%, tanto en períodos lluviosos como de sequía. Sin embargo, en Quito, los promedios alcanzan 83 a 85% en abril, pero en julio y agosto son de solo 66 a 69%, con mínimas diurnas que alcanzan cifras tan bajas como 25 o 30%. [3] [7]

---

[3] Blandin Landivar Carlos; Análisis y Estudios Climatológicos en el Ecuador

[7] Inédito; Condiciones Climáticas de lugares turísticos del Ecuador

[15] López Ericson; 132 años de historia del OAQ

Este descenso es aun más significativo si consideramos que la temperatura media no aumenta, por lo cual esto implica que no solamente baja la humedad relativa sino también la humedad absoluta, es decir, la cantidad de vapor de agua realmente contenida en el aire.

Esta característica climática debe ser tomada en cuenta para muchos propósitos incluso de salud, pues el aire seco de mediados de año resulta muy beneficioso para la mayor parte de las condiciones médicas, en especial si se descarta la gran cantidad de polvo presente en esa época, mayormente en el árido norte de la ciudad [1]

### 2.1.3.2 Cambios de Temperatura

Al considerar que Quito tiene una temperatura media de 13 grados durante todo el año, podría pensarse que se trata de un clima monótono y uniforme. Sin embargo no sucede así, pues ocurren grandes cambios de temperatura no solamente entre el día y la noche, sino también entre un día y otro a la misma hora.

Por ejemplo, al mediodía y siempre refiriéndonos a temperaturas a la sombra y medidas según normas de la Organización Meteorológica Mundial (OMM), en un día nublado puede haber 13 grados centígrados, pero al día siguiente o pocos días después, puede haber 23 grados centígrados a la misma hora, si el día es soleado. De igual forma de madrugada y en este caso citemos un ejemplo concreto, el día 22 de diciembre de 1998 a las 05h00, se registraron 5 grados centígrados en el aeropuerto, mientras que el día 27 a la misma hora se registraban 12.

***Ver anexo 1***

Esto significa que la ciudad, pese a su latitud ecuatorial, presenta cambios derivados de condiciones locales de nubosidad y otros factores; pero en este caso concreto, los mencionados cambios térmicos de un día a otro no ocurren solamente en esta ciudad, sino también son observados en Ambato, Riobamba, Cuenca y casi todo el país. [3]

---

[1] Álvaro, Marco; Boletín Meteorológico OAQ

[3] Blandin Landivar Carlos; Análisis y Estudios Climatológicos en el Ecuador

En resumen, la ciudad de Quito presenta características climatológicas especiales que deben ser consideradas para efectos de salud, diseños arquitectónicos y otros propósitos.

### 1.1.3 OBSERVACIONES METEOROLÓGICAS

Una de las actividades de la meteorología es la toma de datos cada hora diariamente. En las respectivas horas de observación se dispone de formatos especiales para llenar los casilleros que corresponden a los valores de los múltiples parámetros atmosféricos. De esta forma la Estación Meteorológica del OAQ efectúa sus observaciones diarias.

Se necesitan todos los elementos atmosféricos con sus continuos cambios, ya que mediante la visualización permanente de lo que ocurre con estos elementos, se organiza el registro aproximado de cada uno de los parámetros tales como: temperatura del aire, precipitaciones, humedad y la presión atmosférica.

La segunda fase de una observación meteorológica es la parte instrumental, para esto el OAQ cuenta con diversos instrumentos de medición tales como: barómetro, evaporímetro, psicrómetro, hiterógrafo, pluviómetro, entre otros para sus observaciones diarias. [4] [24]

#### **Barómetro.**

En la figura 1.1 se puede apreciar el barómetro, con el cual se obtiene los valores de la presión atmosférica, valores que son tomados en las observaciones de 07h00, 13h00 y 19h00 en mm. de Hg.

---

[4] Donn William; Meteorología

[24] Organización Meteorológica Internacional; Intercambio de datos y productos hidrológicos



**Figura 1.1 Barómetro**

### **Evaporímetro**

En la figura 1.2 se puede observar el evaporímetro cuya graduación es de 0-30 mm., con mediciones de la evaporación del agua cada 6 horas principalmente, las información obtenida es usada con fines industriales. [14]



**Figura 1.2 Evaporímetro**

### **Pluviógrafo.**

La figura 1.3 muestra el pluviómetro que registra en forma continua la cantidad total y la duración de lluvia caída en milímetros (mm.), de los registros puede definirse no sólo la altura de la precipitación caída sino también, cuanto ha caído, permitiendo analizar la distribución de la lluvia en el tiempo. El pluviógrafo que se utiliza normalmente en las estaciones es de sistema Hellman de Sifón. [25]

---

[14] LLauge Felix; La Meteorología

[25] Pluviógrafo

La altura de la boca del pluviógrafo será de 1.50 m., sobre el suelo, y su superficie quedará perfectamente horizontal, es muy importante la nivelación del aparato, para que su funcionamiento sea correcto. Las bandas que se ajustan al tambor son diarias ya que el clima de Quito es muy variable. **Ver anexo 1**



**Figura 1.3 Pluviómetro**

### **Psicrómetro**

En la figura 1.4 consta el psicrómetro el cual tiene de un termómetro llamado de bulbo seco y de un termómetro de bulbo húmedo, los dos termómetros son utilizados en la medición psicrométrica de la humedad, datos que se los consigue mediante la utilización apropiada de los cálculos matemáticos de los que se obtendrán los valores de humedad relativa, tensión de vapor del agua y temperatura del punto de rocío.



**Figura 1.4 Psicrómetro**

### **Hiterógrafo**

En la figura 1.5 se encuentra el hiterógrafo que es una combinación del termógrafo e higrógrafo. El termógrafo registra de forma continua la temperatura del aire en la hoja que se encuentra dividida en días y horas, horizontalmente, y en grados de temperatura en escala vertical, mientras el higrógrafo registra la humedad relativa, el papel de registro está calibrado verticalmente en humedades relativas de 0 a 100%. [1] [4]



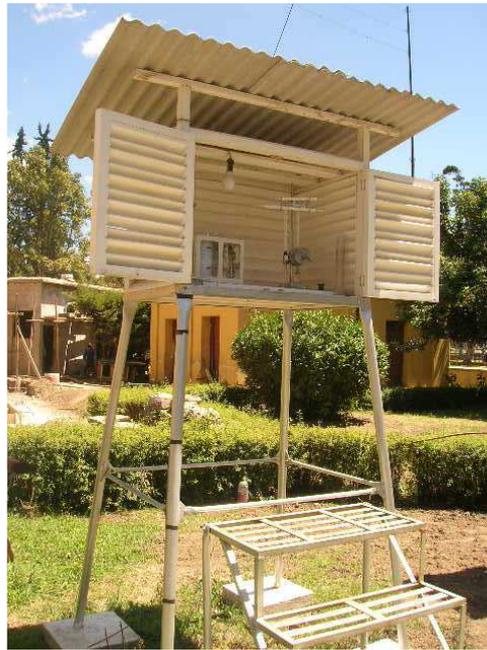
**Figura 1.5 Termógrafo e Higrógrafo**

---

[1] Álvaro, Marco; Boletín Meteorológico OAQ

[4] Donn William; Meteorología

Tanto el psicrómetro como el hiterógrafo están instalados en la garita meteorológica como se puede apreciar en la figura 1.6, la misma que está provista en todo su alrededor de persianas dobles superpuestas que permiten el ingreso solamente del aire suficiente para que exista la variabilidad de datos en aparatos y termómetros, para el registro de acuerdo con las condiciones atmosféricas imperantes durante el momento de efectuar una observación meteorológica. [1]



**Figura 1.6 Garita meteorológica**

Además con el debido chequeo, mantenimiento y calibración, se puede obtener con máxima precisión los valores señalados en los aparatos registradores, como también de los termómetros que son de lectura directa.

Todo lo señalado anteriormente es la norma que debe guiar al observador meteorólogo, para el buen desempeño al momento de realizar una observación meteorológica, disponer de información ágil y oportuna y lograr una estadística confiable.

---

[1] Álvaro, Marco; Boletín Meteorológico OAQ

## **1.2 ELEMENTOS METEOROLÓGICOS**

La Temperatura, presión, humedad y precipitaciones son los elementos de mayor influencia para la vida humana y dentro de ella su accionar múltiple en todas las actividades diarias.

### **1.2.1 TEMPERATURA**

#### **La temperatura y la sensación térmica**

La temperatura atmosférica es el indicador de la cantidad de energía calorífica acumulada en el aire. Aunque existen otras escalas para otros usos, la temperatura del aire se suele medir en grados centígrados y, para ello, se usa un instrumento llamado "termómetro".

La temperatura depende de diversos factores, por ejemplo, la inclinación de los rayos solares, el tipo de sustratos (la roca absorbe energía, el hielo la refleja), la dirección y fuerza del viento, la latitud, la altura sobre el nivel del mar, la proximidad de masas de agua, etc.

Sin embargo, hay que distinguir entre temperatura y sensación térmica. Aunque el termómetro marque la misma temperatura, la sensación que percibimos depende de factores como la humedad del aire y la fuerza del viento. Por ejemplo, se puede estar a 15 grados centígrados en un lugar soleado y sin viento. Sin embargo, a esta misma temperatura a la sombra o con un viento de 80 Km. /h, sentimos una sensación de frío intenso. [15]

### **1.2.2 PRESIÓN ATMOSFÉRICA**

La presión atmosférica es el peso de la masa de aire por cada unidad de superficie. Por este motivo, la presión suele ser mayor a nivel del mar que en las cumbres de las montañas, aunque no depende únicamente de la altitud.

Las grandes diferencias de presión se pueden percibir con cierta facilidad. Con una presión alta nos sentimos más cansados, por ejemplo, en un día de verano, con una presión demasiado baja (por ejemplo, por encima de los 3.000 metros) nos sentimos más ligeros, pero también respiramos con mayor dificultad.

La presión "normal" a nivel del mar es de unos 1.013 milibares y disminuye progresivamente a medida que se asciende. Para medir la presión utilizamos el "barómetro". [12]

### **1.2.3 HUMEDAD**

La humedad indica la cantidad de vapor de agua presente en el aire. Depende, en parte, de la temperatura, ya que el aire caliente contiene más humedad que el frío.

La humedad relativa se expresa en forma de tanto por ciento (%) de agua en el aire. La humedad absoluta se refiere a la cantidad de vapor de agua presente en una unidad de volumen de aire y se expresa en gramos por centímetro cúbico (gr/cm<sup>3</sup>). Para medir la humedad se utiliza un instrumento llamado "higrómetro".

#### **2.1.3.1 Saturación**

La saturación es el punto a partir del cual una cantidad de vapor de agua no puede seguir creciendo y mantenerse en estado gaseoso, sino que se convierte en líquido y se precipita.

Cuando una masa de aire contiene la máxima cantidad de vapor de agua admisible a una determinada temperatura, es decir, que la humedad relativa llega al ciento por ciento, el aire está saturado.

Si estando la atmósfera saturada se le añade más vapor de agua, o se disminuye su temperatura, el sobrante se condensa.

### **2.1.3.2 Punto de Rocío**

Si una masa se enfría lo suficiente, alcanza una temperatura llamada punto de rocío, por debajo de la cual no puede mantener toda su humedad en estado de vapor y éste se condensa, convirtiéndose en líquido, en forma de gotitas de agua. Si la temperatura es suficientemente baja se originan cristales de hielo.

Sin embargo, casi siempre se necesita algo sobre lo que el vapor pueda condensarse, es decir, superficies o cuerpos apropiados donde depositarse, y en la atmósfera, ese algo son partículas diminutas, impurezas procedentes de la Tierra. La mayoría de esas partículas son tan pequeñas que no pueden verse a simple vista y se conocen como núcleos de condensación.

### **1.2.4 PRECIPITACIONES.**

Si la temperatura, la humedad y la presión son los elementos que determinan el clima, el viento y las precipitaciones son sus más evidentes consecuencias. El viento es la circulación del aire de un lugar a otro, con más o menos fuerza. Su principal efecto es el de mezclar distintas capas o bolsas de aire. Cuando se concentra la humedad en una zona y esta asciende hasta una capa de aire más fría, se producen las precipitaciones.

Cuando la humedad del aire supera el punto de saturación, se condensa alrededor de pequeñas partículas sólidas que flotan en la atmósfera y se forman las nubes. Algunas de ellas se desarrollan en vertical, corrientes internas hacen que el aire ascienda hacia zonas más frías, mientras las gotas aumentan de tamaño ya que, al descender la temperatura, el agua en estado gaseoso tiende a convertirse en líquida. Si las gotas de agua o hielo superan en peso a las fuerzas que las sostienen, caen por la fuerza de la gravedad y forman lo que llamamos una "precipitación".

Dependiendo de la temperatura y el grado de condensación, el agua se puede precipitar en forma de lluvia líquida, pero también puede hacerlo en forma de cristales de hielo (nieve) o de masas densas de hielo de diverso tamaño (granizo).

Cuando las diferencias de temperatura entre dos masas de aire son muy grandes, la condensación se produce con enorme rapidez y abundancia, hay precipitaciones intensas, acompañadas de movimientos bruscos del aire y de intercambio eléctrico entre las masas (rayos y relámpagos). Es lo que llamamos "tormentas" y, en algunos casos, pueden llegar muy violentas [4]

---

[4] Donn William; Meteorología

## **CAPITULO 2. IMPLEMENTACIÓN DEL DATA WAREHOUSE**

### **2.1.ANÁLISIS DE REQUERIMIENTOS**

#### **2.1.1. ANTECEDENTES**

La estación meteorológica del Observatorio Astronómico de Quito (OAQ), al haber funcionado de forma casi ininterrumpida en más de cien años posee un gran banco de datos de los distintos elementos climatológicos, pero si los datos registrados en las observaciones meteorológicas son analizados de forma directa perderían su valor y sería muy difícil encontrar patrones de comportamiento y peor aún pronosticar parámetros meteorológicos.

#### **2.1.2. SELECCIÓN DE ELEMENTOS METEOROLÓGICOS**

Los parámetros meteorológicos que el OAQ registra actualmente son:

- Brillo de sol
- Cantidad de Lluvia
- Heliofanía
- Humedad
- Nubosidad
- Presión Atmosférica
- Temperatura del aire
- Temperatura del césped
- Velocidad y dirección del viento

Pero al ser la temperatura, la presión atmosférica y la humedad los elementos que determinan el clima y las precipitaciones las más notorias consecuencias que influyen en las actividades diarias de la sociedad, se ha escogido estos elementos y se han recopilado los registros diarios de los mismos desde el año 1995 hasta el 2005 para la implementación del data warehouse. [1]

### **2.1.3. METODOLOGÍA DE DESARROLLO DEL DATA WAREHOUSE**

Los sistemas de bases de datos son una herramienta en la administración y gerencia de la información, las cuales han evolucionado a tal punto que se han convertido en la estructura ideal para almacenar cualquier tipo abstracto de dato, que se pueda automatizar y digitalizar, de tal manera que la implementación del sistema de información sea una tarea sencilla.

En la actualidad tanto los investigadores, científicos y administradores necesitan tecnologías que integren y analicen toda la información por ellos requerida, pero no como una simple colección, sino como una estructura completa y variante en el tiempo. Diseñando un nivel superior en las bases de datos, conocido como DataWarehouse.

El desarrollo del Data Warehouse es diferente al desarrollo clásico de aplicaciones, ya que estas usualmente son desarrolladas siguientes pasos: Obtención de requerimientos, análisis y diseño, programación, pruebas e implementación. Sin embargo, en el mundo de los sistemas para soporte de decisiones (DSS), los requerimientos son frecuentemente la última cosa que se conoce en el proceso de desarrollo. La metodología propuesta es un desarrollo iterativo, centrado en los datos, conocido con el nombre de metodología de desarrollo en espiral. Como lo muestra la figura 2.1

---

[1] Álvaro, Marco; Boletín Meteorológico OAQ

## Metodología de Desarrollo de un Data Warehouse en Espiral

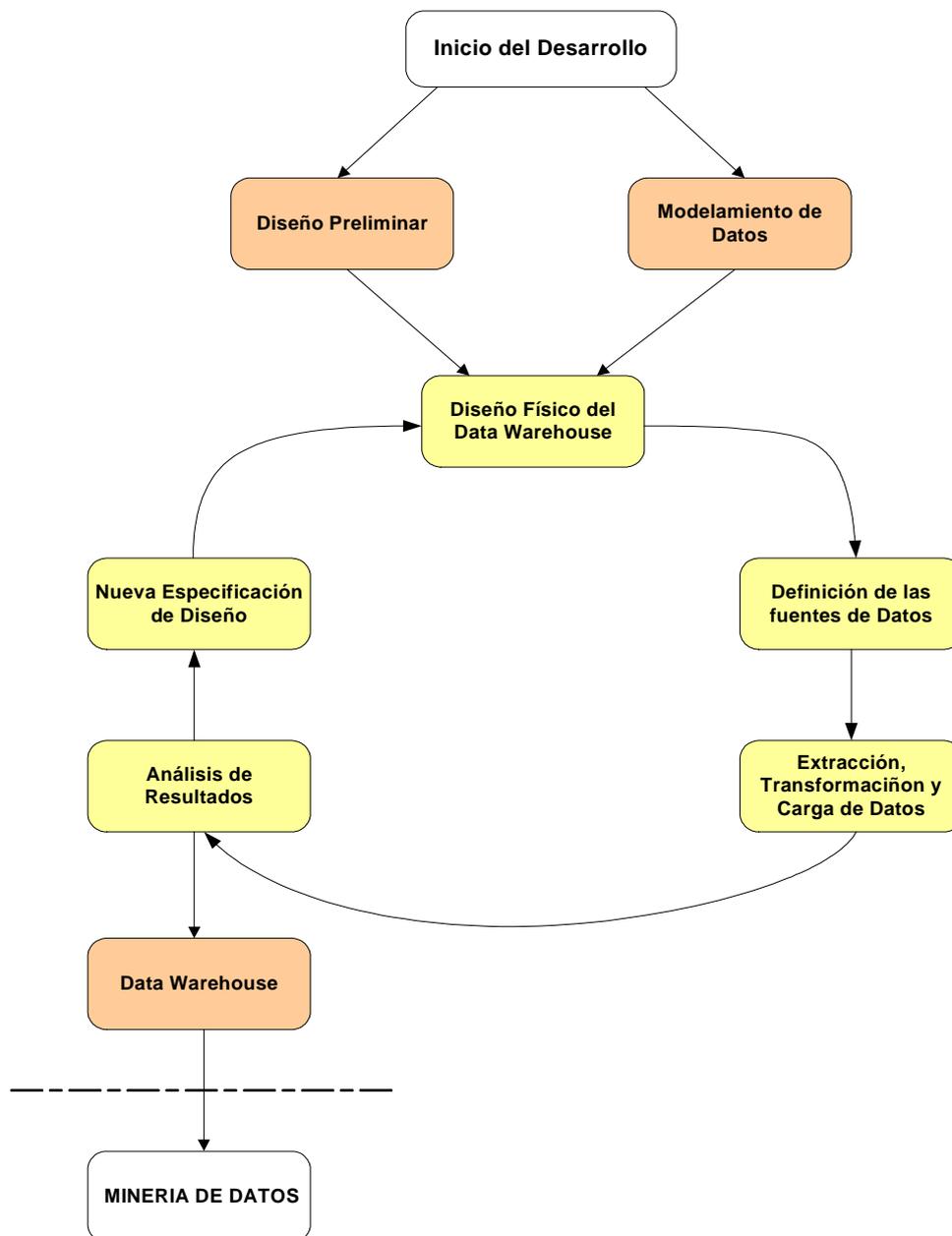


Figura 2.1 Metodología para el desarrollo del Data Warehouse

[14]

### 2.1.3.1 Metodología en Espiral

En este método el énfasis está en la velocidad y la culminación, con reconocimiento de que los requerimientos no se pueden identificar con claridad o especificar al inicio. El enfoque se basa en la observación de que es más fácil redirigir un sistema desplegado, con base a nuevos requerimientos, que construir una solución completa basada en requerimientos inadecuados o no disponibles.

El método espiral se presta principalmente al desarrollo de aplicaciones de bases de datos, al desarrollo de un Data Warehouse.

#### **Diseño Preliminar:**

Es la etapa inicial, en la que se realiza un análisis de los requerimientos del caso de estudio. En esta fase se deberá tomar en cuenta los siguientes conceptos:

- **Granularidad:** Se refiere al nivel de detalle de la información requerida. La granularidad tiene una relación directa con las actividades de resumen y adición que deben realizarse sobre los datos fuente. A menor granularidad, mayor cantidad de detalle. En general, los datos operacionales se consideran como el nivel más bajo de granularidad. Para incrementar su granularidad, los datos operacionales deben resumirse y acumularse todavía más. Por lo regular a mayor granularidad, mayor será la cantidad de procesamiento requerido para convertir y resumir los datos operacionales. Al mismo tiempo los datos con alta granularidad requieren de menos volumen de almacenamiento y se pueden también consultar con rapidez y conveniencia. El costo de resumen y adición es continuo para cada actualización del Data Warehouse cuando la fuente de datos operacionales reúne nueva información.

- **Variables:** Las variables representan algún aspecto cuantificable o medible de los datos almacenados o recopilados durante un periodo de tiempo.

### **Modelado de Datos:**

Los modelos de datos proveen una estructura para identificar, nombrar, describir y asociar los componentes de una base de datos. En general se necesitan modelos para la fuente de datos como también para los datos seleccionados para el almacenamiento en el Data Warehouse. [21]

En esta fase se realiza la abstracción del caso de estudio basándose en los siguientes conceptos:

**Sujeto de Datos.-** Un sujeto de datos es una categoría grande de datos relevantes para el negocio, que consiste de un conjunto de entidades y sus relaciones entre sí y permite mapear los sujetos de datos.

### **Modelo entidad-relación**

El modelo entidad-relación es el modelo conceptual más utilizado para el diseño conceptual de bases de datos. Fue introducido por Peter Chen en 1976. El modelo entidad-relación está formado por un conjunto de conceptos que permiten describir la realidad mediante un conjunto de representaciones gráficas y lingüísticas

**Entidades.-** Cualquier tipo de objeto o concepto sobre el que se recoge información: cosa, persona, concepto abstracto o suceso. Por ejemplo: autos, casas, empleados, clientes, empresas, oficios, diseños de productos, conciertos, excursiones, etc. Las entidades se representan gráficamente como se muestra en la figura 2.2, mediante rectángulos y su nombre aparece en el interior. Un nombre de entidad sólo puede aparecer una vez en el esquema conceptual.

---

[21] Implementando un Data Warehouse



**Figura 2.2 Entidad**

**Relación.-** Es una correspondencia o asociación entre dos o más entidades. Cada relación tiene un nombre que describe su función. Las relaciones se representan gráficamente mediante rombos y su nombre aparece en el interior como se muestra en la figura 2.3.



**Figura 2.3 Relación**

**Atributo.-** Es una característica de interés o un hecho sobre una entidad o sobre una relación. Los atributos representan las propiedades básicas de las entidades y de las relaciones. Toda la información extensiva es portada por los atributos. Gráficamente, se representan mediante óvalos que se unen a las entidades o relaciones a las que pertenecen, mediante conectores, ver figura 2.4. [17] [18]



**Figura 2.4 Atributo y Conector**

### **Normalización:**

Normalización es un proceso que clasifica relaciones, objetos, formas de relación y demás elementos en grupos, en base a las características que cada uno posee.

---

[17] Diseño Conceptual de la Base de datos

[18] Modelo Entidad-Relación

La normalización convierte una relación en varias sub-relaciones, cada una de las cuales obedece a reglas. Estas reglas se describen en términos de dependencia. Una vez que hayamos examinado las distintas formas de dependencia, encontraremos procedimientos a aplicar a las relaciones de tal modo que las mismas puedan descomponerse de acuerdo a la dependencia que prevalece.

### **Primera Forma Normal (1FN)**

Sea  $\alpha$  un conjunto de atributo perteneciente ( $\in$ ) a la relación R, en donde R está en la Primera Forma Normal si todos los atributos  $\alpha[n]$  son atómicos, es decir no pueden seguir dividiéndose. Por ejemplo:

La Relación:

- cursos: nombre, código, vacantes, horario, bibliografía

Queda después de aplicar la Forma Normal 1 de la siguiente manera:

- cursos1: nombre, código, vacantes
- horario1: código, día, módulo
- bibliografia1: código, nombre, autor

Una columna no puede tener múltiples valores. Los datos están atómicos (Si a cada valor de X le pertenece un valor de Y, entonces a cada valor de Y le pertenece un valor de X).

### **Segunda Forma Normal (2FN)**

Dependencia completa. Esta en 2FN si esta en 1FN y si sus atributos no principales dependen de forma completa de la clave principal. Toda columna que no sea clave debe depender por completo de la clave primaria. Los atributos dependen de la clave. Varía la clave y varían los atributos. Dependencia completa. Sus atributos no principales dependen de forma completa de la clave principal.

### **Tercera Forma Normal (3FN)**

Está en segunda forma normal y todo atributo no primo es implicado por la clave primaria en una secuencia no transitiva. Se eliminan las dependencias transitivas. Todos los atributos no llave dependen de manera no transitiva de

la llave primaria (no deben depender de ninguna otra columna que no sea llave)

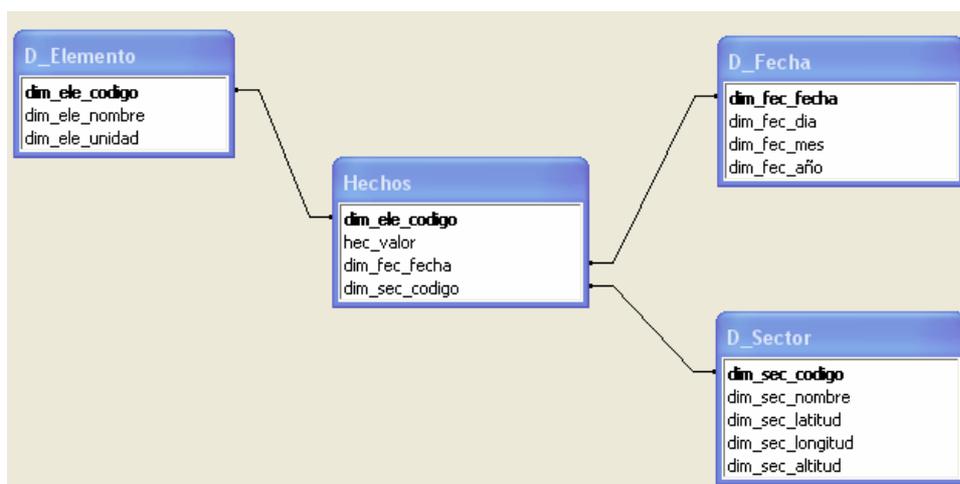
**Modelo dimensional.-** El modelado dimensional es muy diferente al modelado en tercera forma normal (3FN). El modelado en tercera forma normal es una técnica que busca remover datos redundantes. La principal diferencia entre los dos modelos es el grado de normalización.

Un modelo dimensional contiene la misma información que un modelo normalizado pero en paquetes de datos en un formato cuyos objetivos de diseño son la comprensibilidad del usuario, el rendimiento de las consultas y la elasticidad a los cambios.

Los datos se modelan con una estructura multidimensional, donde cada dimensión corresponde a un atributo o conjunto de atributos en el esquema entorno a unos “Hechos”, que almacenan el valor de alguna medida agregada. [22]

### Esquema Estrella

El esquema estrella es un prototipo de modelado que tiene un solo objeto en medio conectado con varios objetos de manera radial. Ver figura 2.5



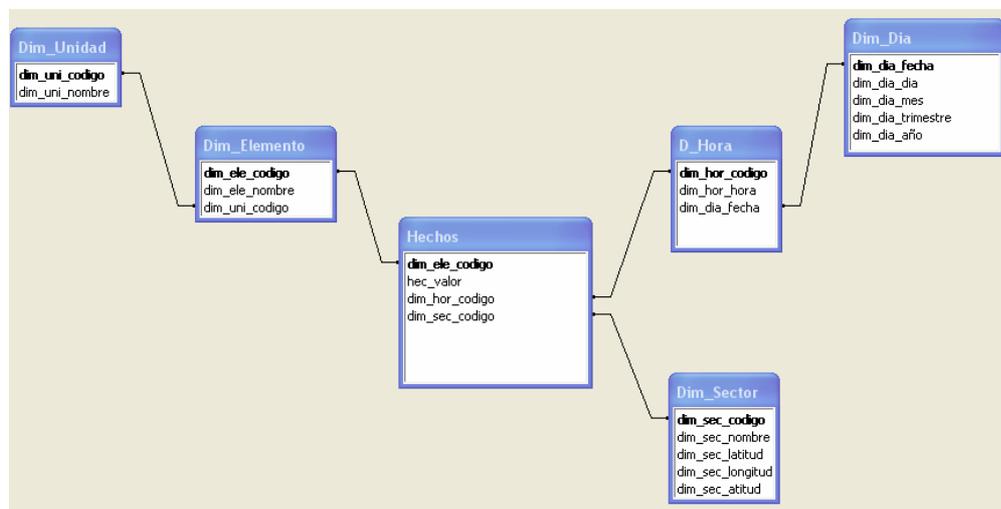
**Figura 2.5 Esquema Estrella**

El esquema estrella refleja la visión del usuario final de una consulta. El objeto en el centro de la estrella se denomina tabla de hechos, los objetos conectados a ella se llaman tablas de dimensión.

Las tablas de dimensión contienen atributos que describen los valores de la dimensión, las consultas en la tabla de hechos simplemente conectan listas de apuntadores a las tablas de dimensión. Cuando las consultas en la tabla de hechos se unen son consultas en la tabla de dimensión, se recupera información diversa. Las tablas de dimensión permiten hacer adiciones en la búsqueda por medio de uniones.

### Esquema Copo de Nieve.

El esquema de copo de nieve es una extensión del esquema estrella, en donde cada uno de los puntos de la estrella se divide en más puntos, ver la figura 2.6



**Figura 2.6 Esquema Copo de Nieve**

En esta forma de esquema, las tablas de dimensión del esquema estrella contienen más normas. Las ventajas que proporciona el esquema copo de nieve son mejorar el desempeño de consultas debido a un mínimo almacenamiento de disco para los datos y mejorar el desempeño mediante la unión de tablas más pequeñas, en lugar de las grandes tablas. El esquema

copo de nieve incrementa la flexibilidad de las aplicaciones, por lo tanto disminuye la granularidad de las dimensiones.

El esquema copo de nieve aumenta el número de tablas con que el usuario debe tratar e incrementa la complejidad de algunas de las consultas que debe localizar. [5][21]

### **Diseño Físico del Data Warehouse:**

En la fase de diseño físico, los modelos lógicos desarrollados en el modelamiento de datos se convierten en modelos físicos. Los procesos identificados en el modelamiento de datos son conectar las fuentes de datos con el Data Warehouse y los datos con las herramientas de la estación de trabajo del usuario final.

Las actividades del diseño comprenden lo siguiente:

Desarrollo del modelo físico de datos para la base de datos de almacenamiento del Data Warehouse. En forma adicional, pudiera también requerirse el desarrollo de modelos físicos para el almacenamiento local necesario para las herramientas del usuario final.

Correspondencia de los modelos físicos de las fuentes de datos con los modelos físicos del Data Warehouse. Esta correspondencia ayuda a los procesos de extracción, refinamiento y reingeniería a efectuar sus funciones dentro del Data Warehouse.

### **Definición de las Fuentes de Datos:**

Un Data Warehouse es un repositorio de información coleccionada desde varias fuentes, almacenadas bajo un esquema unificado. Fuentes de datos como:

---

[5] Gill, Harjinder; Data Warehousing

[21] Implementando un Data Warehouse

- Reportes.
- Hojas de Cálculo
- Archivos de Texto
- OLTPs

**OLTP: On line Transaction Processing:** Son suficientes para realizar las transacciones diarias de una organización. Sin embargo resultan insuficientes para funciones más complejas, como el análisis, planificación y la predicción, es decir para tomar decisiones estratégicas a largo plazo.

### **Extracción Transformación y Carga de Datos.**

La calidad de los datos es de suma importancia en la construcción de un Data Warehouse, por ello, después de la recopilación el siguiente paso es seleccionar y preparar el subconjunto de datos que va a cargar en el Data Warehouse. Este paso es necesario ya que algunos datos en la etapa anterior son irrelevantes o innecesarios para la tarea que se desea realizar.

Pero además de la irrelevancia, existen otros problemas que afectan a la calidad de los datos, uno de ellos es la presencia de valores que no se ajustan al comportamiento general de los datos.

Otra es la presencia de datos faltantes o perdidos ya que puede ser un problema grave que puede conducir a resultados pocos precisos, no obstante, es necesario analizar primero el significado de los valores faltantes antes de tomar alguna decisión sobre cómo tratarlos ya que estos pueden deberse a causas muy diversas como un mal funcionamiento del dispositivo que registro el valor, al los cambios efectuados en los procedimiento usados durante la recolección de los datos, o al hecho de que los datos se recopilen desde diversas fuentes.

Otra tarea de preparación de los datos es la construcción de atributos, la cual consiste en construir automáticamente nuevos atributos, aplicando alguna operación o función a los atributos originales con el objeto de que estos

nuevos atributos hagan más fácil las consultas. La razón principal para esta tarea es cuando los atributos originales no tiene mucho poder predictivo por si solos o los patrones dependen de variaciones lineales de las variables originales.

El tipo de los datos pueden también modificarse para facilitar el uso de técnicas que requieren tipos de datos específicos. Así, algunos atributos de pueden sumarizar lo que reduce el espacio y permite usar técnicas numéricas

Si deseamos el proceso inverso (discretizar los atributos) es decir, transformar valores numéricos en atributos discretos o nominales. Los atributos discretizados pueden tratarse como atributos categóricos con un número más pequeño de valores. La idea básica es partir los valores de un atributo continuo en una pequeña lista de intervalos, tal que cada intervalo es visto como un valor discreto del atributo.

#### **Análisis de Resultados:**

En esta etapa comprobamos la calidad de los datos cargados en el Data Warehouse, los cuales debe ser precisos, comprensibles y relevantes.

#### **Nueva Especificación del Diseño:**

Si al analizar los resultados se encuentran datos que no fueron tomados en cuenta, o surgen nuevos requerimientos. Se deberá especificar un nuevo diseño y de esta manera empezar nuevamente el proceso de construcción, es decir un proceso en espiral, el cual va depurando el Data Warehouse.

#### **Data Warehouse:**

Es un conjunto de datos históricos, internos o externos y descriptivos de un contexto o área de estudio, que están integrados y organizados de tal forma que permiten aplicar eficientemente herramientas para resumir, describir y analizar los datos con el fin de ayudar en la toma de decisiones estratégicas.

La ventaja fundamental de un Data Warehouse es su diseño específico y su separación de la base de datos transaccional. [14][8]

## 2.2.CONSTRUCCIÓN DE UN DATA WAREHOUSE

### 2.2.1. DISEÑO PRELIMINAR

#### Recopilación de la Información

La recopilación de los requerimientos existentes se realizó basándose en los reportes, hojas de cálculo, hojas de registro de muestras, bandas y los boletines meteorológicos generados en un periodo de diez años por el área de Meteorología del OAQ.

Cabe recalcar que cuando se recopiló la información el OAQ no contaba con ningún sistema ni base de datos que almacenara las muestras recolectadas cada hora por el personal del departamento de Meteorología del OAQ.

Por lo que más adelante se describirá el diseño y la construcción de un modelo de datos transaccional el cual se recomienda ser usado para almacenar los futuros datos.

Ejemplos:

**Bandas Meteorológicas:** Formato que representa a escala las distintas fluctuaciones de los diferentes elementos meteorológicos que registra un instrumento de medición. En la figura 2.7, se muestra la banda meteorológica correspondiente a la temperatura, la escala de la banda meteorológica está dividida en grados y horas.

---

[14] Artículo Bodega de Datos

[8] Jarke, Matthias; Fundamentals of Data Warehouses

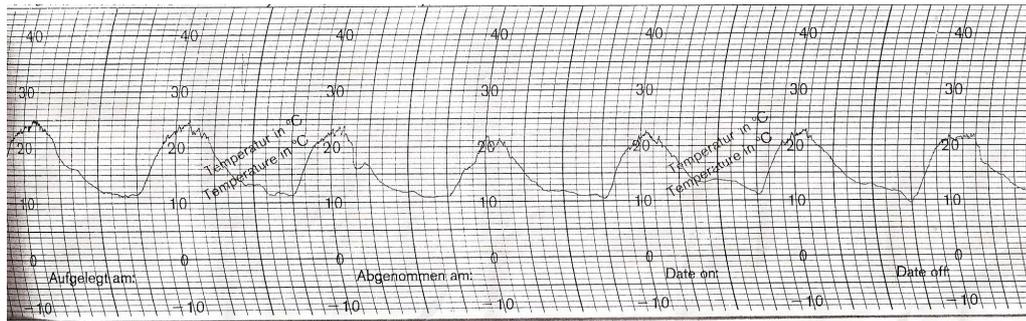


Figura 2.7 Bandas Meteorológicas

Los datos recogidos de las bandas meteorológicas con transcritos a una hoja de cálculo diariamente, según lo informó el OAQ.

Hojas de Cálculo: El OAQ registra las mediciones diarias en hojas electrónicas, como se ve en la figura 2.8, en las que almacenan datos para luego procesarlas a través de histogramas con la finalidad de visualizar el comportamiento del clima y obtener reportes.

ESCUELA POLITÉCNICA NACIONAL																															
OBSERVATORIO ASTRONÓMICO DE QUITO																															
DEPARTAMENTO DE METEOROLÓGIA																															
HUMEDAD ATMOSFERICA																															
LONGITUD 78°29'54" W																															
LATITUD 0°12'57" S																															
ALTURA 2818																															
EMERGO 2800																															
2010																															
MADRUGADA																															
MAÑANA																															
TARDE																															
NOCHE																															
FECHA	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	SUMA	MEAN	MAX.	MIN.	DESC.		
01	100	100	100	100	100	100	93	93	100	80	77	70	56	51	48	55	62	85	94	100	100	100	100	100	100	2875	84	100	48	52	
02	100	100	99	99	99	99	99	100	100	71	65	64	58	56	54	54	70	84	98	96	96	97	97	99	99	2850	85	100	54	46	
03	98	98	98	98	98	98	98	100	69	67	63	57	46	46	44	63	65	90	87	98	98	98	98	98	99	1992	83	100	44	56	
04	98	98	98	97	97	97	98	98	98	70	71	67	46	42	65	62	68	85	88	93	96	97	97	97	97	2846	85	100	46	53	
05	6	80	83	92	94	95	94	98	100	95	61	63	56	56	52	52	48	46	50	77	90	90	90	90	90	1844	76	100	46	54	
06	7	92	92	93	93	93	96	98	95	74	65	67	61	64	73	71	70	89	94	96	98	98	98	98	98	2802	86	100	41	37	
07	8	98	98	98	98	98	98	100	47	53	42	41	51	51	44	51	55	62	80	85	94	94	94	94	94	1844	76	100	41	59	
08	9	98	100	99	99	99	99	95	76	55	51	50	48	49	51	65	65	76	92	95	95	98	99	99	99	1940	80	100	48	52	
09	10	100	100	99	99	99	99	98	100	100	100	77	78	78	79	77	74	84	84	88	98	99	99	99	99	2182	83	100	74	26	
10	11	95	95	95	95	95	96	98	97	99	100	84	69	68	66	68	63	70	74	87	96	96	98	98	98	2893	87	100	68	48	
11	12	98	99	99	98	96	96	97	98	79	54	48	51	51	52	54	53	58	69	89	93	96	96	97	97	1916	80	100	48	51	
12	13	97	97	97	96	96	96	96	98	98	95	75	63	49	46	51	64	71	80	84	95	96	96	96	96	2826	84	100	46	52	
13	14	97	97	97	98	98	98	100	91	57	48	46	35	38	42	42	47	75	82	79	87	100	100	100	100	1854	80	100	35	45	
14	15	100	100	100	99	99	97	97	97	93	90	66	61	60	59	55	56	70	72	91	98	97	100	100	100	2864	86	100	55	45	
15	16	99	99	98	98	98	98	99	99	100	100	100	88	86	74	72	76	95	96	94	97	98	98	98	98	2268	94	100	72	28	
16	17	98	98	98	98	98	98	99	99	100	65	57	64	67	52	65	74	86	87	87	83	99	96	97	97	2858	85	100	52	48	
17	18	97	97	96	96	95	95	97	98	99	62	69	48	47	50	43	47	48	54	63	65	68	77	87	87	1754	73	100	42	56	
18	19	90	93	95	97	96	96	98	100	54	42	46	40	40	37	36	39	56	65	82	87	94	90	91	91	1755	87	100	36	64	
19	20	91	92	97	97	95	96	97	99	78	41	35	40	40	35	41	45	50	61	72	72	71	81	85	70	1838	78	100	35	64	
20	21	85	94	100	100	100	99	98	98	78	62	45	41	41	40	42	41	41	39	54	59	61	63	62	64	1579	80	100	39	61	
21	22	81	96	100	100	100	94	94	79	78	49	44	39	40	37	41	43	47	50	54	57	55	56	57	55	56	1839	80	100	37	63
22	23	60	62	62	66	62	71	97	63	60	45	42	43	44	47	49	50	50	59	69	88	81	96	98	98	98	1532	63	100	42	54
23	24	98	98	98	97	96	95	98	69	72	59	49	95	44	46	43	42	49	50	67	88	97	100	100	100	1856	80	100	42	59	
24	25	100	100	100	100	100	99	98	100	100	62	56	47	46	47	52	58	87	93	97	98	98	98	98	98	2835	84	100	46	54	
25	26	98	98	98	98	98	98	100	87	73	50	55	54	54	44	54	53	88	92	96	94	94	97	97	97	2828	84	100	44	54	
26	27	97	97	97	98	98	98	98	100	100	77	70	57	59	62	88	82	75	94	98	98	98	98	98	98	98	2135	80	100	57	43
27	28	98	98	98	98	98	98	100	100	100	100	100	89	83	85	87	98	97	96	96	96	97	97	97	97	2346	96	100	83	17	
28	29	97	97	98	98	98	98	98	98	98	97	97	97	97	92	92	94	95	95	96	96	96	97	97	97	2355	96	100	92	4	
29	30	97	97	97	98	98	98	98	100	100	72	62	72	72	78	81	76	85	94	98	98	98	98	98	98	98	2191	91	100	62	38
30	31	98	98	98	98	98	98	98	98	100	100	100	96	96	98	89	87	87	84	96	99	94	95	95	95	2271	88	100	84	16	
31	32	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	2271	88	100	84	16
32	33	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	2271	88	100	84	16
33	34	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	2271	88	100	84	16
34	35	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	2271	88	100	84	16
35	36	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	2271	88	100	84	16
36	37	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	2271	88	100	84	16
37	38	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	2271	88	100	84	16
38	39	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	2271	88	100	84	16
39	40	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	2271	88	100	84	16
40	41	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	2271	88	100	84	16
41	42	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	2271	88	100	84	16
42	43	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	2271	88	100	84	16
43	44	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	2271	88	100	84	16
44	45	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	2271	88	100	84	16
45	46	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	2271	88	100	84	16
46	47	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	2271	88	100	84	16
47	48	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	2271	88	100	84	16
48	49	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	2271	88	100	84	16
49	50	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	2271	88	100	84	16
50	51	98	98	9																											

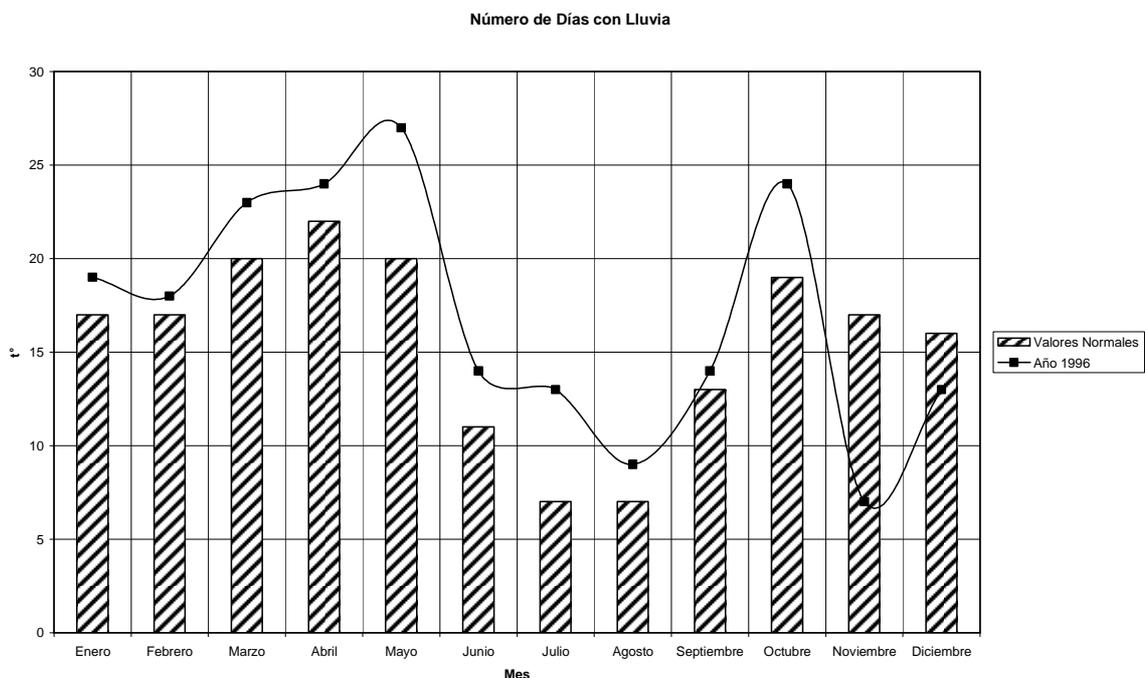
- Departamento Responsable
- Elemento
- Año

Latitud, longitud y altitud de donde fueron tomadas las muestras. Dicho datos corresponden a la ubicación de la estación meteorológica.

Los libros de registro de cada elemento están divididos en horas que son las columnas y días las filas, las muestras se registran de acuerdo a la hora y día correspondiente.

Según este formato se tiene la posibilidad de obtener promedios, máximos y mínimos según la conveniencia del departamento de meteorología del OAQ. La consolidación de estos datos son mecuales y anuales,

**Reportes Estadísticos:** Representaciones gráficas de los datos obtenidos y almacenados en hojas de cálculo, que permiten visualizar los resultados conseguidos de las muestras recogidas en un determinado periodo, como se muestra en la figura 2.9.



**Figura 2.9 Reportes Estadísticos**

Los reportes estadísticos en este caso muestran la cantidad de lluvia (mm) en cada mes para tener una mejor visualización se realizó una comparación con los valores normales que se debieron obtenerse para ese año.

**Variabes.-** Las variables son representadas por valores detallados y numéricos para cada instancia del elemento medido. En la tabla 2-1 se puede apreciar las variables escogidas con sus respectivas unidades y el número de muestras que son tomadas diariamente. [1]

<b>Elementos Meteorológicos</b>	<b>Unidades</b>	<b>Número de muestras diarias</b>
Temperatura del Aire	Grados Centígrados	24
Humedad Relativa	Tanto por Ciento	24
Presión Atmosférica Media	Milímetros de Mercurio	1
Cantidad de Lluvia	Milímetros cúbicos	24

**Tabla 2.1 Elementos Meteorológicos**

**Ejemplo:** En la figura 2.10 se muestra un segmento de las muestras tomadas para la temperatura, en la cual se registran los valores en grados centígrados y se comprueba las 24 muestras registradas.

ESCUELA POLITECNICA NACIONAL							
OBSERVATORIO ASTRONOMICO DE QUITO							
DEPARTAMENTO DE METEOROLOGIA							
TEMPERATURA DEL AIRE							
en grados centígrados a la sombra y a 2 metros sobre el suelo							
ENERO 1996							
MADRUGADA							
FECHA	1	2	3	4	5	6	7
1	11.2	11.0	10.9	10.6	10.0	9.4	9.0
2	11.5	11.4	10.8	10.8	10.8	10.6	10.1
3	10.6	10.7	10.6	10.6	10.3	10.8	10.6
4	10.3	10.3	10.1	10.2	10.1	10.1	10.4
5	11.1	11.0	10.5	10.2	10.2	10.1	10.4
6	10.3	10.5	10.6	10.6	10.3	10.2	10.4
7	9.4	9.4	9.2	8.6	8.3	8.3	8.2
8	11.0	10.7	10.3	10.0	9.8	9.2	9.1
9	11.6	11.3	11.1	10.9	10.7	10.7	10.6
10	11.2	11.3	10.9	10.6	10.4	10.4	10.7
11	10.9	10.9	10.0	10.0	9.4	9.4	9.8
12	11.4	11.3	11.1	11.0	10.4	10.1	10.7
13	11.8	11.8	11.9	11.2	11.2	11.1	11.2
14	11.0	11.0	10.8	10.8	10.4	10.3	10.6
15	10.9	10.9	10.9	10.8	10.1	10.2	10.6
16	12.1	12.1	11.5	11.3	11.3	11.2	11.2
17	11.8	11.5	11.1	11.2	11.1	11.1	11.3
18	10.9	10.5	10.3	10.1	10.0	10.0	9.9
19	10.9	10.6	9.9	9.8	8.9	8.7	8.8
20	10.8	10.7	10.4	10.3	10.2	10.1	10.4

Figura 2.10 Hoja de Cálculo

## 2.2.2. MODELADO DE LOS DATOS

### 2.2.2.1. Modelo para la fuente de datos

**Sujeto de Datos.-** Un sujeto de datos es una categoría grande de datos relevantes para las observaciones meteorológicas, que consiste de un conjunto de entidades y sus relaciones entre sí y permite mapear los sujetos de datos.

En este caso el Sujeto de datos son las Observaciones Meteorológicas.

**Entidades.-** Cualquier objeto concreto o abstracto del que se desea almacenar información. En el caso de estudio las entidades son:

- ELEMENTO\_METEOROLÓGICO
- MUESTRA\_DIARIA

Para la construcción de modelo relacional se uso la siguiente nomenclatura

**Relación.-** Correspondencia o asociación entre ELEMENTO\_METEOROLÓGICO Y MUESTRA \_DIARIA.

- OBSERVACIÓN

**Atributo.-** Propiedad o característica de interés que describe a las entidades en estudio y a su relación

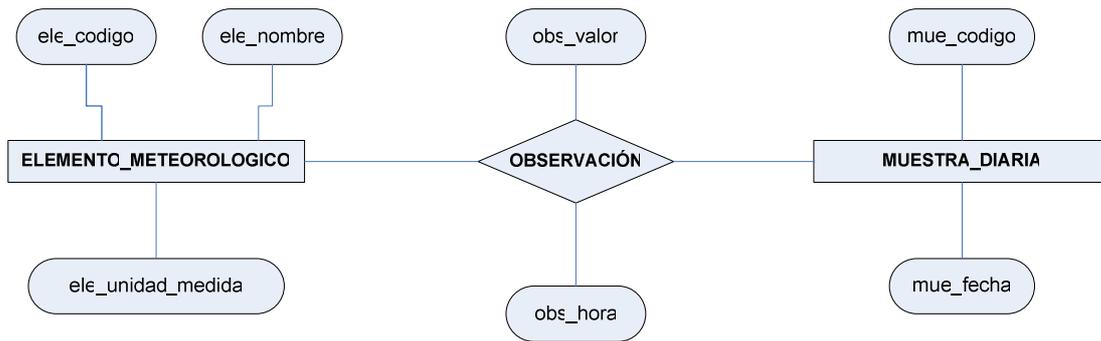
Entidad	Atributo	Tipo de Dato	Detalle
ELEMENTO METEOROLÓGICO	ele_codigo	number	Código del elemento meteorológico
	ele_nombre	text(50)	Nombre del elementos meteorológico
	ele_unidad_medida	text(10)	Unidad de medida del elemento meteorológico
MUESTRA DIARIA	mue_codigo	number	Código de la muestra diaria
	mue_fecha	date	Día en que la muestra es recolectada

**Tabla 2.2 Descripción de los atributos de la Entidades**

Relación	Atributo	Tipo de Dato	Detalle
OBSERVACIÓN	obs_hora	number	Hora en que se tomó la muestra
	obs_valor	number	Valor de la muestra

**Tabla 2.3 Descripción de los atributos de la Relación**

**Diagrama Entidad Relación.-** La representación del modelo Entidad-Relación se muestra en la figura 2.9



**Figura 2.11 Modelo Entidad Relación**

En base a lo detallado anteriormente se procede a construir la base de datos Observaciones\_Meteorológicas, Se anexa el Scrip de la Base de Datos Transaccional.



**Figura 2.12 Modelo OLTP**

[2] [18]

### Creación de los Usuarios

Para una adecuada creación de los diferentes elementos que fueron requeridos, fue necesaria la creación de tres usuarios, ver tabla 2.4, los cuales permitieron definir los esquemas, los cuales se especificarán más adelante.

USUARIOS		
Nombre	Perfil	Esquema
Transaccional	Administrador	Transaccional
Temporal	Administrador	Temporal
Data_Warehouse	Administrador	Data_Warehouse

**Tabla 2.4 Usuarios**

[2] Anahory, Sam; Data Warehousing in the Real World

[18] Modelo Entidad Relación

## Base de Datos Transaccional

Para la implementación de esta base de datos, fue requerida la creación y configuración de Tables Spaces y en usuario que va a trabajar sobre la bases de datos. Para nuestro caso de estudio se ha definido el table space OAQ, el cual tiene un tamaño de 100 MB. y el usuario transaccional el cual tiene los permisos necesarios para poder trabajar en la base de datos Transaccional , a continuación se describe el código fuente para la creación de lo mencionado anteriormente.

### Tables Spaces

```
CREATE TABLESPACE "OAQ"
LOGGING
DATAFILE 'E:\ORACLE\ORADATA\EXPI\OAQ.ora' SIZE 100M EXTENT
MANAGEMENT LOCAL SEGMENT SPACE MANAGEMENT AUTO
```

### Usuario Transaccional

```
CREATE USER "TRANSACCIONAL" PROFILE "DEFAULT"
IDENTIFIED BY "vaio" DEFAULT TABLESPACE "USERS"
ACCOUNT UNLOCK;
GRANT "AQ_ADMINISTRATOR_ROLE" TO "TRANSACCIONAL" WITH ADMIN OPTION;
GRANT "AUTHENTICATEDUSER" TO "TRANSACCIONAL" WITH ADMIN OPTION;
GRANT "CONNECT" TO "TRANSACCIONAL" WITH ADMIN OPTION;
GRANT "DBA" TO "TRANSACCIONAL" WITH ADMIN OPTION;
GRANT "OLAP_DBA" TO "TRANSACCIONAL" WITH ADMIN OPTION;
```

Una vez creado el table space y el usuario, procedemos a crear las tablas, las cuales están basadas en el modelo OLPT, el cual se aprecia en la figura 2.12, en donde se almacenarán los datos meteorológicos recolectados, los cuales ya están en archivos planos, listos para se importados a la base de datos.

## CREACIÓN DE TABLAS

**ELEMENTO\_METEOROLOGICO:** Esta tabla permite almacenar la información de los elementos meteorológicos, como son:

Un identificador, el nombre del elemento y su unidad de medida.

```
CREATE TABLE "TRANSACCIONAL"."ELEMENTO_METEOROLOGICO"
("ELE_CODIGO" VARCHAR2(10) NOT NULL, "ELE_NOMBRE" VARCHAR2(40)
NOT NULL, "ELE_UNIDAD_MEDIDA" VARCHAR2(40) NOT NULL,
CONSTRAINT "PK_ELE_MET" PRIMARY KEY("ELE_CODIGO"))
TABLESPACE "OAQ"
```

**MUESTRA\_DIARIA:** Esta tabla permite almacenar la información de la fecha en que los datos meteorológicos son tomados.

```
CREATE TABLE "TRANSACCIONAL"."MUESTRA_DIARIA" ("MUE_CODIGO"
VARCHAR2(10) NOT NULL, "MUE_FECHA" DATE NOT NULL,
CONSTRAINT "PK_MUE_DIA" PRIMARY KEY("MUE_CODIGO"))
TABLESPACE "OAQ"
```

**OBSERVACION:** Esta tabla almacena las mediciones de los elementos meteorológicos tomadas cada hora diariamente.

```
CREATE TABLE "TRANSACCIONAL"."OBSERVACION" ("ELE_CODIGO"
VARCHAR2(10) NOT NULL, "MUE_CODIGO" VARCHAR2(10) NOT NULL,
"OBS_HORA" NUMBER(10) NOT NULL, "OBS_VALOR" FLOAT(10) NOT NULL,
CONSTRAINT "FK_ELE_MET" FOREIGN KEY("ELE_CODIGO")
REFERENCES
"TRANSACCIONAL"."ELEMENTO_METEOROLOGICO"("ELE_CODIGO"),
CONSTRAINT "FK_MUE_DIA" FOREIGN KEY("MUE_CODIGO")
REFERENCES "TRANSACCIONAL"."MUESTRA_DIARIA"("MUE_CODIGO"))
TABLESPACE "OAQ"
```

## CARGA DE DATOS AL MODELO TRANSACCIONAL

Importamos a la base de datos los archivos planos depurados, para ello utilizamos el comando SQL Loader (sqlldr) y trabajaremos con los datos de elementos meteorológicos para explicar como cargaremos los datos, a la base de datos Transaccional y para las el ingreso de datos a las demás tablas se seguirá el mismo proceso. [26]

**Sqlldr** tiene la siguiente sintaxis:

***sqlldr miusuario/mipasword@mibase ejemplo.ctl rows=1 errors=30000***

*miusuario: Usuario de la Instancia*

*mipassword: Password asignado*

*mibase: Nombre de la base de Datos que se utilizando*

*ejemplo.ctl: nombre que se le dio al archivo de control que contiene los datos a ser importados*

*rows=1 : Significa que cada 1 registros ira haciendo COMMIT*

*errors=30000 : Significa el número de errores que puede soportar*

Como se aprecia, sqlldr requiere de un archivo de control (control file), el cual especifica cómo deben cargarse los datos en la base de datos; y un fichero de datos (data file), que especifica qué datos deben ser cargados.

El archivo de control tiene el siguiente formato:

```
LOAD DATA
INFILE *
INTO TABLE <TABLA>
APPEND
FIELDS TERMINATED BY '<DELIMITADOR>' OPTIONALLY ENCLOSED ""
TRAILING NULLCOLS
<ATRIBUTOS> BEGINDATA
Datos separados con el delimitador ";"
```

Y la extensión con la que se guarda el archivo de control .ctl

Al ejecutar sqlldr, Oracle usará ".bad" si hubo errores al cargar los datos. Si no se especifica archivo de log y un archivo de tuplas malas, sqlldr usará el nombre del archivo de control y generará un archivo. log el cual indicará las novedades en la carga de datos y el archivo .bad, el cual contendrá las tuplas que no pudieron cargarse. **Ver anexo 4.**

### 2.2.2.2. Modelo para el Data Warehouse

#### Modelo dimensional.-

Las observaciones meteorológicas están almacenadas en una base de datos relacional, estas se abstraen y modelan de tal manera que sea posible tener un modelo dimensional.

Estas tablas modeladas dimensionalmente se conocen como asociaciones de estrella. Para poder analizar los datos almacenado se emplea scripts para poder realizar las consultas y representarlos en cubos.

Una tabla de hecho es la tabla primaria en un modelo dimensional donde las medidas del rendimiento cuantitativo del registro meteorológico almacenadas, es decir los hechos que suceden en el OAQ. Una medida es tomada de la intersección de todas las dimensiones. Al registrarse las muestras de algunos elementos cada hora es necesario dividir en diferentes dimensiones, por lo que se implemento el esquema copo de nieve o híbrido. [4]

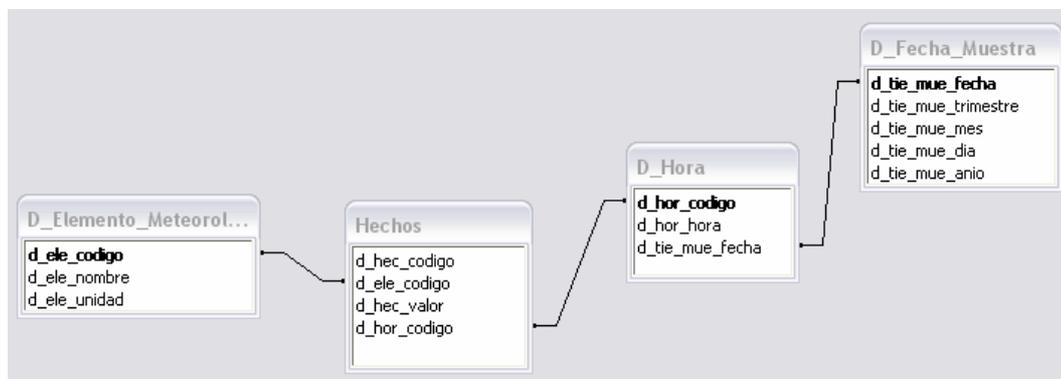


Figura 2.13 Modelo Dimensional con esquema copo de nieve

**D\_Elemento\_Meteorológico:** Esta tabla representa la Dimensión Elemento Meteorológico, la cual permite registrar la información referente a los elementos meteorológicos más importantes. Esta tabla tiene los siguientes campos:

**d\_ele\_codigo:** Almacena el código de cada uno de los elementos meteorológicos del OAQ.

**d\_ele\_nombre:** Almacena el nombre del elemento meteorológico.

**d\_ele\_unidad:** Almacena la unidad de medida del elemento meteorológico.

**D\_Fecha\_Muestra:** Esta Tabla representa a la Dimensión Fecha Muestra, la misma que almacena los datos correspondientes a los días en que el OAQ ha registrado las muestras tomadas. Esta tabla contiene los siguientes campos:

**d\_tie\_mue\_fecha:** Almacena las fechas en que el OAQ ha registrado las muestras.

**d\_tie\_mue\_trimestre:** Almacena el trimestre (1, 2, 3, 4) correspondiente al día resgistrado.

**d\_tie\_mue\_mes:** Almacena el mes del año (enero, febrero,.. diciembre) correspondiente al día resgistrado.

**d\_tie\_mue\_dia:** Almacena el día resgistrado.

**d\_tie\_mue\_anio:** Almacena el año (1995 al 2005)) correspondiente al día registrado.

**D\_Hora:** Esta Tabla representa a la Dimensión Hora, la misma que almacena los datos correspondientes a las horas en que las muestras son tomadas. Esta tabla contiene los siguientes campos:

**d\_hor\_codigo:** registra el código de las horas de cada día en que se ha registrado las muestras meteorológicas.

**d\_hor\_hora:** registra las horas de cada día en que se ha registrado las muestras meteorológicas.

**d\_tie\_mue\_fecha:** permite relacionar esta dimensión con la dimensión D\_Fecha\_Muestras.

**Hechos:** Esta tabla almacena todas las muestras de los diferentes elementos meteorológicos, el cual cuenta con los siguientes campos:

**d\_hec\_codigo:** Clave primaria única la cual permite controlar el ingreso de datos en la tabla de hechos.

**d\_ele\_codigo:** Campo que permite relacionar la dimensión Elemento Meteorológico con la tabla de Hechos.

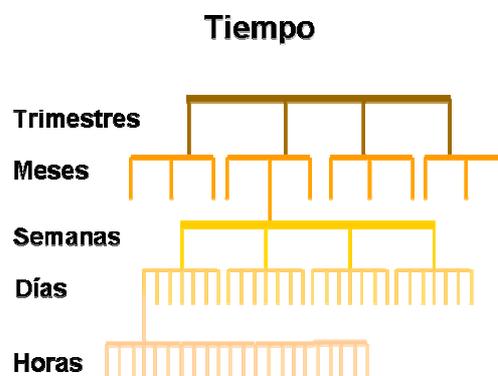
**d\_hec\_valor:** Almacena los valores de cada muestra tomada cada hora y cada día.

**d\_hora\_codigo** Campo que permite relacionar la dimensión Hora con la tabla de Hechos.

**Granularidad.-** Define el nivel de detalle o de sumariación de los datos que se encuentran en el Data Warehouse. Se debe hacer una selección del nivel de granularidad del Data Warehouse, teniendo en cuenta que:

- A menor nivel de granularidad, mayor funcionalidad, a mayor nivel de granularidad menor funcionalidad.
- A menor nivel de granularidad mayor el volumen de datos, a mayor nivel de granularidad menor el volumen de los datos.
- A menor nivel de granularidad mayores tiempos de respuesta, a mayor nivel granularidad, menores tiempos de respuestas.
- A menor nivel granularidad, mayor el costo del Data Warehouse, a mayor nivel de granularidad menor el costo de la bodega de datos.

Para el caso de estudio la granularidad (ver figura 2.14) de tiempo llegará hasta la hora, debido a que es el tiempo mínimo en el que se registra una muestra. [18]



**Figura 2.14 Nivel de detalle Tiempo**

### 2.2.3. DISEÑO FÍSICO PARA EL DATA WAREHOUSE

La creación Física del Data Warehouse llevó un largo proceso desde la recolección de los datos, hasta la carga y comprobación de la calidad de los mismos.

Para la creación del Data Warehouse, se trabajó sobre motor de base de datos: Oracle 9i y se realizaron las siguientes actividades:

#### Creación de los TableSpaces

Una base de datos Oracle consta de una o más unidades de almacenamiento lógicas denominadas tablespaces, que almacenan de forma colectiva todos los datos de la base de datos.

Se han creado tres tablespaces para las tres bases datos que fueron requeridas para la consolidación de la información

**Tablespace OAQ:** Creada para el almacenamiento lógico de la Base de Datos Transaccional. El tamaño de 100M fue escogido debido a que se considero como el adecuado para manipular los datos en la base de datos.

```
CREATE TABLESPACE "OAQ"
LOGGING
DATAFILE 'E:\ORACLE\ORADATA\DM\OAQ.ora' SIZE 100M EXTENT
MANAGEMENT LOCAL SEGMENT SPACE MANAGEMENT AUTO
```

**Tablespace TOAQ:** Creada para el almacenamiento lógico de la Base de Datos Temporal, la cual se usó para la Extracción, transformación y carga de datos.

```
CREATE TABLESPACE "TOAQ"
LOGGING
DATAFILE 'E:\ORACLE\ORADATA\DM\TOAQ.ora' SIZE 100M EXTENT
MANAGEMENT LOCAL SEGMENT SPACE MANAGEMENT AUTO
```

**Tablespace DWOAQ:** Creada para el almacenamiento lógico del Data Warehouse. Debido a la cantidad de datos a ingresarse se considero 100M, de esta forma es posible realizar las operaciones necesarias en este tablespace.

```
CREATE TABLESPACE "DWOAQ"
LOGGING
```

```
DATAFILE 'E:\ORACLE\ORADATA\MDM\DWOAQ.ora' SIZE 100M EXTENT
MANAGEMENT LOCAL SEGMENT SPACE MANAGEMENT AUTO
```

**Usuario DATAWAREHOUSE:** Este usuario se creo para administrar la instancia de la base de datos del Datawarehouse tiene contraseña y permisos que le permiten administrar en su totalidad la instancia de la base de datos.

```
CREATE USER "DATAWAREHOUSE" PROFILE "DEFAULT"
  IDENTIFIED BY "vaio" DEFAULT TABLESPACE "USERS"
  ACCOUNT UNLOCK;
GRANT "AQ_ADMINISTRATOR_ROLE" TO " DATAWAREHOUSE ";
GRANT "AQ_USER_ROLE" TO " DATAWAREHOUSE "
GRANT "CONNECT" TO " DATAWAREHOUSE "
GRANT "DBA" TO " DATAWAREHOUSE ";
GRANT "OLAP_DBA" TO " DATAWAREHOUSE ";
```

## ESQUEMA DATA\_WAREHOUSE

### D\_ELEMENTO\_METEOROLOGICO

```
CREATE TABLE "DATA_WAREHOUSE"."D_ELEMENTO_METEOROLOGICO"
("D_ELE_CODIGO" VARCHAR2(10) NOT NULL, "D_ELE_NOMBRE"
  VARCHAR2(40) NOT NULL, "D_ELE_UNIDAD" VARCHAR2(40) NOT NULL,
  CONSTRAINT "PK_D_ELE_MET" PRIMARY KEY("D_ELE_CODIGO"))
  TABLESPACE "DWOAQ"
```

### D\_TIEMPO\_MUESTRA

```
CREATE TABLE "DATA_WAREHOUSE"."D_TIEMPO_MUESTRA"
("D_TIE_MUE_FECHA"
  VARCHAR2(10) NOT NULL, "D_TIE_MUE_TRIMESTRE" NUMBER(10) NOT NULL,
  "D_MES" VARCHAR2(15) NOT NULL, "D_TIE_MUE__DIA" CHAR(10) NOT NULL,
  "D_TIE_MUE__ANIO" NUMBER(4) NOT NULL,
  CONSTRAINT "PK_D_TIE_MUE" PRIMARY KEY("D_TIE_MUE__FECHA"))
  TABLESPACE "DWOAQ";
```

### D\_HORA

```
CREATE TABLE "DATA_WAREHOUSE"."D_HORA" ("D_HOR_CODIGO" NUMBER(10)
  NOT NULL, "D_HOR_HORA" VARCHAR2(10) NOT NULL,
  "D_TIE_MUE_FECHA" VARCHAR2(10) NOT NULL,
  CONSTRAINT "PK_D_HOR" PRIMARY KEY("D_HOR_CODIGO"),
  CONSTRAINT "FK_D_TIE_MUE" FOREIGN KEY("D_TIE_MUE_FECHA")
  REFERENCES
  "DATA_WAREHOUSE"."D_TIEMPO_MUESTRA"("D_TIE_MUE_FECHA"))
  TABLESPACE "DWOAQ"
```

## HECHOS

```
CREATE TABLE "DATA_WAREHOUSE"."HECHOS" ("D_ELE_CODIGO"
  VARCHAR2(10) NOT NULL, "D_HOR_CODIGO" NUMBER(10) NOT NULL,
  "D_HEC_VALOR" FLOAT(10) NOT NULL,
  CONSTRAINT "FK_D_HOR" FOREIGN KEY("D_HOR_CODIGO")
  REFERENCES "DATA_WAREHOUSE"."D_HORA"("D_HOR_CODIGO"),
  CONSTRAINT "FK_D_ELE_MET" FOREIGN KEY("D_ELE_CODIGO")
  REFERENCES
  "DATA_WAREHOUSE"."D_ELEMENTO_METEOROLOGICO"("D_ELE_CODIGO"))
  TABLESPACE "DWOAQ"
```

### 2.2.4. DEFINICIÓN DE LAS FUENTES DE DATOS

El OAQ, posee una variada fuente de datos como:

- Boletines meteorológicos
- Bandas Meteorológicas
- Reportes Estadísticos
- Hojas de Cálculo
- Base de datos Transaccional

Por lo que la información meteorológica se ha obtenido de los reportes digitales, impresos y escritos de los archivos del OAQ, dichos reportes han sido revisados, rectificados y completados para contar con una fuente confiable de datos para la implementación del data warehouse.

Para poder consolidar la información de una mejor manera, se tuvieron que cargar de todas las fuentes los datos a la base de datos transaccional.

A continuación describimos como se realizó el proceso de carga de datos a la base de datos transaccional.

1. El OAQ, entregó una serie de hojas de cálculo en el cual se encontraban almacenadas todas las muestras registradas de los diferentes elementos meteorológicos.

2. Se procedió a revisar cada registro de cada año, encontrando las siguientes novedades.

### **Valores faltantes**

En algunos años no se registraron valores debido a que los instrumentos no se encontraban operando o los responsables de la toma de datos lo hicieron de forma ineficiente.

### **Datos inconsistentes y duplicación**

- Al examinar los datos se descubrió que los valores no estaban dentro de un rango normal correspondiente al mes.
- Se notó que se duplicó la información mensual en un determinado año.
- El ingreso de los datos también presentó problemas ya que hubo errores al momento de transcribir de las fajas de registro de los instrumentos y boletines a las hojas de cálculo.

### **Soluciones**

Para poder verificar los datos inconsistentes y faltantes se revisó las fajas de registro y las observaciones en los boletines para verificar las causas de los problemas mencionados anteriormente y realizar los respectivos correctivos.

Los datos faltantes debido a la inactividad de los instrumentos de medición no pueden ser reemplazados debido a que son únicos y no pueden ser calculados.

Para aquellos casos en que se encontraron valores nulos aislados, se procedió a realizar una aproximación de los datos, en base a los datos contiguos y valores referenciales de acuerdo a la hora de otros. **Ver anexo 2**

Para la consolidación de todos los datos se procedió a crear archivos de texto con el delimitador “ ; ”. Y se procedió a cargar los datos en la base de datos transaccional, para visualizar de una mejor manera todos los datos. Ver anexo 3 para visualizar el script de creación de la base de datos transaccional

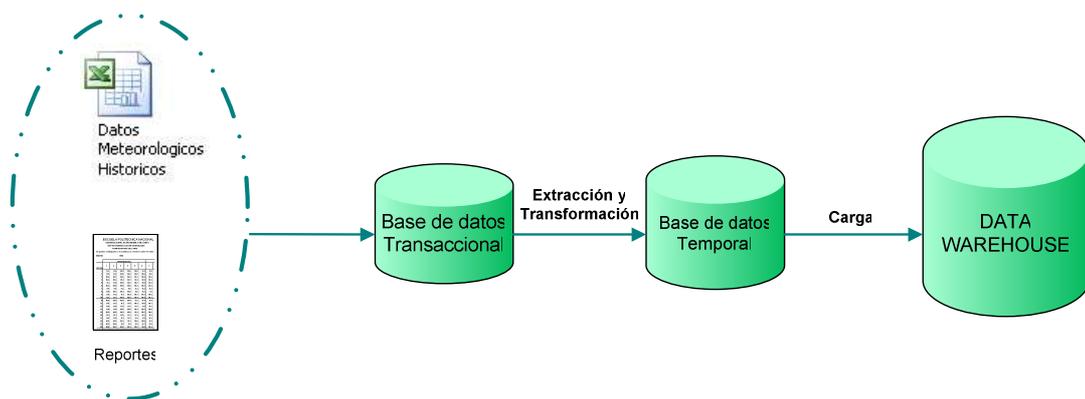
**Usuario TEMPORAL:** Este Usuario fue creado para realizar la depuración y la carga de datos, posee una contraseña y permisos de administrador que le permiten administrar en su totalidad la instancia de la base de datos creada.

```
CREATE USER "TEMPORAL" PROFILE "DEFAULT"
  IDENTIFIED BY "vaio" DEFAULT TABLESPACE "USERS"
  ACCOUNT UNLOCK;
GRANT "AQ_ADMINISTRATOR_ROLE" TO "TEMPORAL";
GRANT "AQ_USER_ROLE" TO "TEMPORAL"
GRANT "CONNECT" TO "TEMPORAL"
GRANT "DBA" TO "TEMPORAL";
GRANT "OLAP_DBA" TO "TEMPORAL";
```

### 2.2.5. EXTRACCIÓN, TRANSFORMACIÓN Y CARGA DE DATOS

Fue necesaria la creación de la base de datos temporal ya que no se contaba con un atributo necesario en la tabla Observación, gracias a la base de datos temporal se pudo suplir el problema de no tener una clave primaria para Observación. *Ver anexo 3*

La figura 2.15 muestra el proceso de extracción, transformación y carga de datos al Data Warehouse que se siguió, para continuar con el proceso final que es aplicar minería de datos sobre esta base.



**Figura 2. 15 Extracción, Transformación y Carga de Datos [6]**

Para poder facilitar el ingreso de 293387 muestras meteorológicas que no tenían un código primario, fue necesario crear una secuencia y enlazarlo mediante un trigger a la tabla T\_OBSERVACION.

## Secuencia Incremento

```
CREATE SEQUENCE incremento
  MINVALUE 1
  MAXVALUE 9999999999999999999
  START WITH 1
  INCREMENT BY 1
  NOCACHE
;
```

## Trigger Incrementar

```
CREATE OR REPLACE TRIGGER Incrementar
  BEFORE INSERT ON T_OBSERVACION
  FOR EACH ROW
  BEGIN
    IF :new.T_OBS_CODIGO IS NULL THEN
      SELECT incremento.nextval INTO :new.T_OBS_CODIGO FROM DUAL;
    END IF;
  END;
/
COMMIT;
```

## Inserciones a la base de datos Temporal

Inserción de datos en la Tabla temporal T\_ELEMENTO\_METEOROLOGICO

```
INSERT INTO T_ELEMENTO_METEOROLOGICO
SELECT * FROM TRANSACCIONAL.ELEMENTO_METEOROLOGICO;
```

Inserción de datos en la Tabla temporal T\_MUESTRA\_DIARIA

```
INSERT INTO T_MUESTRA_DIARIA
SELECT * FROM TRANSACCIONAL.MUESTRA_DIARIA;
SELECT * FROM T_MUESTRA_DIARIA;
```

Inserción de datos en la Tabla temporal T\_OBSERVACION

```
INSERT INTO T_OBSERVACION
(T_ELE_MET_CODIGO, T_MUE_CODIGO, T_OBS_HORA, T_OBS_VALOR)
SELECT * FROM TRANSACCIONAL.OBSERVACION;
```

Cuando se inserto en la tabla T\_OBSERVACION el Trigger nos permitió tener el código único de cada muestra meteorológica automáticamente

En este momento tenemos los datos listos para ser insertados en el Data Warehouse.

## Carga de Datos en el Data Warehouse

Para la carga de datos del Data Warehouse fue necesario ejecutar el siguiente Script

### **D\_TIEMPO\_MUESTRA**

```
INSERT INTO D_TIEMPO_MUESTRA
SELECT DISTINCT TO_CHAR(T_MUE_FECHA,'YYYYMMDD') D_TIE_MUE_FECHA,
TO_CHAR(T_MUE_FECHA,'Q') D_TIE_MUE_TRIMESTRE,
TO_CHAR(T_MUE_FECHA,'MONTH')D_TIE_MUE_MES,
TO_CHAR(T_MUE_FECHA,'DAY')D_TIE_MUE_DIA,
TO_CHAR(T_MUE_FECHA,'YYYY')D_TIE_MUE_ANIO
FROM TEMPORAL.T_MUESTRA_DIARIA;
```

### **D\_ELEMENTO\_METEOROLOGICO**

```
INSERT INTO D_ELEMENTO_METEOROLOGICO
SELECT * FROM TEMPORAL.T_ELEMENTO_METEOROLOGICO;
```

### **D\_HORA**

```
INSERT INTO D_HORA
SELECT
TO_NUMBER(OS.T_OBS_CODIGO) D_HOR_CODIGO,
TO_CHAR(OS.T_OBS_HORA) D_HOR_HORA,
TO_CHAR(MD.T_MUE_FECHA,'YYYYMMDD') D_TIE_MUE_FECHA
FROM TEMPORAL.T_MUESTRA_DIARIA MD,TEMPORAL.T_OBSERVACION OS
WHERE MD.T_MUE_CODIGO = OS.T_MUE_CODIGO
```

### **HECHOS**

```
INSERT INTO HECHOS
SELECT
TO_CHAR(EM.T_ELE_MET_CODIGO) D_ELE_CODIGO,
TO_NUMBER(OS.T_OBS_CODIGO) D_HOR_CODIGO,
OS.T_OBS_VALOR D_HEC_VALOR
FROM TEMPORAL.T_ELEMENTO_METEOROLOGICO
EM,TEMPORAL.T_OBSERVACION OS
WHERE EM.T_ELE_MET_CODIGO = OS.T_ELE_MET_CODIGO
```

## 2.2.6. ANÁLISIS DE RESULTADOS

Para verificar la consistencia de los datos almacenados en el Data Warehouse se han propuesto varias consultas a las que llamaremos gerenciales, las cuales tienen como objetivo extraer información que sean de utilidad al OAQ y de esta forma también comprobar que los datos almacenados sean válidos.

Para visualizar las consultas realizadas se obtuvo reportes utilizando la Herramienta de reporte Report Builder de Oracle 10g, la cual se conecta directamente a la base de Datos proporcionando una mejor visualización de los resultados obtenidos ya que permite mostrar y diseñar los reportes.

## CONSULTAS GERENCIALES

### Cantidad de lluvia total y trimestralmente caída en el observatorio de la ciudad de Quito

Esta consulta muestra la cantidad de lluvia caída trimestralmente desde 1995 hasta 2005

```
SELECT DECODE(M.D_TIE_MUE_TRIMESTRE,'1', 'PRIMER TRIMESTRE', '2',
'SEGUNDO TRIMESTRE',
      '3', 'TERCER TRIMESTRE', '4', 'CUARTO TRIMESTRE', 'TODO EL AÑO')
TRIMESTRE,
SUM(H.D_HEC_VALOR) VALOR
FROM HECHOS H, D_TIEMPO_MUESTRA M, D_HORA HO
WHERE M.D_TIE_MUE_CODIGO = HO.D_TIE_MUE_CODIGO AND
HO.D_HOR_CODIGO = H.D_HOR_CODIGO AND
H.D_ELE_CODIGO = 'L'
GROUP BY CUBE(M.D_TIE_MUE_TRIMESTRE);
```

### Cantidad de lluvia

La consulta muestra la cantidad de lluvia caída cada año, además se puede conocer la cantidad total de lluvia desde 1995 hasta 2005.

```
SELECT M.D_TIE_MUE_ANIO,
SUM(H.D_HEC_VALOR) VALOR
FROM HECHOS H, D_TIEMPO_MUESTRA M, D_HORA HO
WHERE M.D_TIE_MUE_CODIGO = HO.D_TIE_MUE_CODIGO AND
HO.D_HOR_CODIGO = H.D_HOR_CODIGO AND
H.D_ELE_CODIGO = 'L'
GROUP BY CUBE(M.D_TIE_MUE_ANIO)
ORDER BY VALOR DESC;
```

### El año más lluvioso

Esta consulta permite conocer el año más lluvioso y a partir de este el resto de años aparecen en forma descendente, mostrando la cantidad de lluvia caída en cada uno de ellos.

```
SELECT M.D_TIE_MUE_ANIO,
MAX(H.D_HEC_VALOR) VALOR
FROM HECHOS H, D_TIEMPO_MUESTRA M, D_HORA HO
WHERE M.D_TIE_MUE_CODIGO = HO.D_TIE_MUE_CODIGO AND
HO.D_HOR_CODIGO = H.D_HOR_CODIGO AND
H.D_ELE_CODIGO = 'L'
GROUP BY CUBE(M.D_TIE_MUE_ANIO)
```

ORDER BY VALOR DESC;

### Los meses más lluviosos

Esta consulta muestra la cantidad de lluvia caída mensualmente en cada año de esta manera es posible conocer en que mes de que año se tuvieron mayores o menores precipitaciones y en año.

```
SELECT M.D_TIE_MUE_ANIO,M.D_TIE_MUE_MES,
SUM(H.D_HEC_VALOR) VALOR
FROM HECHOS H, D_TIEMPO_MUESTRA M, D_HORA HO
WHERE M.D_TIE_MUE_CODIGO = HO.D_TIE_MUE_CODIGO AND
HO.D_HOR_CODIGO = H.D_HOR_CODIGO AND
H.D_ELE_CODIGO = 'L'
GROUP BY CUBE(M.D_TIE_MUE_ANIO, M.D_TIE_MUE_MES )
ORDER BY VALOR DESC;
```

### Presión máxima, temperatura y humedad promedio

Esta consulta despliega la cantidad de lluvia total anual, presión máxima, temperatura y humedad promedio anual

```
SELECT ANIO, NOMBRE, VALOR
FROM
(SELECT M.D_TIE_MUE_ANIO ANIO ,EM.D_ELE_NOMBRE NOMBRE,
SUM(H.D_HEC_VALOR) VALOR
FROM HECHOS H, D_TIEMPO_MUESTRA M, D_HORA HO,
D_ELEMENTO_METEOROLOGICO EM
WHERE M.D_TIE_MUE_CODIGO = HO.D_TIE_MUE_CODIGO AND
HO.D_HOR_CODIGO = H.D_HOR_CODIGO AND
H.D_ELE_CODIGO = 'L' AND
EM.D_ELE_CODIGO = H.D_ELE_CODIGO
GROUP BY (M.D_TIE_MUE_ANIO,EM.D_ELE_NOMBRE)
ORDER BY VALOR DESC)
UNION
SELECT ANIO,NOMBRE, VALOR
FROM(SELECT M.D_TIE_MUE_ANIO ANIO ,EM.D_ELE_NOMBRE NOMBRE,
MAX(H.D_HEC_VALOR) VALOR
FROM HECHOS H, D_TIEMPO_MUESTRA M, D_HORA HO,
D_ELEMENTO_METEOROLOGICO EM
WHERE M.D_TIE_MUE_CODIGO = HO.D_TIE_MUE_CODIGO AND
HO.D_HOR_CODIGO = H.D_HOR_CODIGO AND
H.D_ELE_CODIGO = 'P' AND
EM.D_ELE_CODIGO = H.D_ELE_CODIGO
GROUP BY (M.D_TIE_MUE_ANIO,EM.D_ELE_NOMBRE)
ORDER BY VALOR DESC)
UNION
SELECT ANIO,NOMBRE, VALOR
FROM (SELECT M.D_TIE_MUE_ANIO ANIO ,EM.D_ELE_NOMBRE NOMBRE,
AVG(H.D_HEC_VALOR) VALOR
FROM HECHOS H, D_TIEMPO_MUESTRA M, D_HORA HO,
D_ELEMENTO_METEOROLOGICO EM
WHERE M.D_TIE_MUE_CODIGO = HO.D_TIE_MUE_CODIGO AND
```

```

HO.D_HOR_CODIGO = H.D_HOR_CODIGO AND
H.D_ELE_CODIGO = 'T' AND
EM.D_ELE_CODIGO = H.D_ELE_CODIGO
GROUP BY (M.D_TIE_MUE_ANIO,EM.D_ELE_NOMBRE )
ORDER BY VALOR DESC)
UNION
SELECT ANIO,NOMBRE, VALOR
FROM (SELECT M.D_TIE_MUE_ANIO ANIO ,EM.D_ELE_NOMBRE NOMBRE,
AVG(H.D_HEC_VALOR) VALOR
FROM HECHOS H, D_TIEMPO_MUESTRA M, D_HORA HO,
D_ELEMENTO_METEOROLOGICO EM
WHERE M.D_TIE_MUE_CODIGO = HO.D_TIE_MUE_CODIGO AND
HO.D_HOR_CODIGO = H.D_HOR_CODIGO AND
H.D_ELE_CODIGO = 'H' AND
EM.D_ELE_CODIGO = H.D_ELE_CODIGO
GROUP BY (M.D_TIE_MUE_ANIO,EM.D_ELE_NOMBRE)
ORDER BY VALOR DESC)
ORDER BY VALOR DESC;

```

### **AÑO MAS CALUROSO**

Esta consulta muestra el año más caluroso, se obtiene el valor máximo de temperatura de cada año para una mejor visualización se ha ordenado en forma descendente.

```

SELECT M.D_TIE_MUE_ANIO,
MAX(H.D_HEC_VALOR) VALOR
FROM HECHOS H, D_TIEMPO_MUESTRA M, D_HORA HO
WHERE M.D_TIE_MUE_CODIGO = HO.D_TIE_MUE_CODIGO AND
HO.D_HOR_CODIGO = H.D_HOR_CODIGO AND
H.D_ELE_CODIGO = 'T'
GROUP BY CUBE(M.D_TIE_MUE_ANIO)
ORDER BY VALOR DESC;

```

### **VISTAS**

Esta vista muestra el valor mínimo de humedad registrada por día.

```

CREATE OR REPLACE VIEW VISTA_HUM_MIN_MENSUAL
AS
(
SELECT
M.D_TIE_MUE_MES MES,
M.D_TIE_MUE_ANIO ANIO,
M.D_TIE_MUE_CODIGO,
MIN(H.D_HEC_VALOR) VALOR
FROM HECHOS H, D_TIEMPO_MUESTRA M, D_HORA HO
WHERE M.D_TIE_MUE_CODIGO = HO.D_TIE_MUE_CODIGO AND
HO.D_HOR_CODIGO = H.D_HOR_CODIGO AND
H.D_ELE_CODIGO = 'H'
GROUP BY (M.D_TIE_MUE_MES, M.D_TIE_MUE_ANIO,M.D_TIE_MUE_CODIGO)
)
ORDER BY M.D_TIE_MUE_CODIGO DESC

```

Se procedido a realizar diferentes consultas gerenciales como se mostró anteriormente para verificar la consistencia de lo datos, al desplegar dichas consultas en SQLPlus, se encontró un problema con el tipo de dato especificado en el año en la tabla **D\_TIEMPO\_MUESTRA**, por lo que se tuvo que realizar una modificación del tipo de dato.

### 2.2.7. NUEVA ESPECIFICACIÓN DE DISEÑO

Para que la modificación del tipo de dato en la tabla antes mencionada, se tuvo que realizar las siguientes acciones:

- Borrar el contenido de la tabla D\_TIEMPO\_MUESTRA
- Realizar una modificación a través de un Alter Table

```
ALTER TABLE "DATA_WAREHOUSE"."D_TIEMPO_MUESTRA"
      MODIFY("D_TIE_MUE_TRIMESTRE" VARCHAR2(10), "D_TIE_MUE_ANIO"
      VARCHAR2(4))
```

- Se volvió a cargar los datos

```
D_TIEMPO_MUESTRA
INSERT INTO D_TIEMPO_MUESTRA
SELECT DISTINCT TO_CHAR(T_MUE_FECHA,'YYYYMMDD') D_TIE_MUE_CODIGO,
TO_CHAR(T_MUE_FECHA,'Q') D_TIE_MUE_TRIMESTRE,
TO_CHAR(T_MUE_FECHA,'MONTH')D_TIE_MUE_MES,
TO_CHAR(T_MUE_FECHA,'DAY')D_TIE_MUE_DIA,
TO_CHAR(T_MUE_FECHA,'YYYY')D_TIE_MUE_ANIO
FROM TEMPORAL.T_MUESTRA_DIARIA;
```

Luego de haber realizado los cambios y cargar nuevamente los datos se pudieron realizar las siguientes consultas gerenciales.

- La escala de años más calurosos.
- La escala de los más en que se ha precipitado mayor cantidad de lluvia.
- Cantidad de lluvia caída trimestralmente cada año

- La información consolidada de la cantidad de lluvia caída, la presión atmosférica máxima, la humedad relativa media y la temperatura media por cada año.

Para poder contentarlas fue necesaria la creación de las siguientes vistas, para extraer la información que cumpla con las consultas gerenciales.

**Año más caluroso:** Está vista permite obtener en una escala descendente las temperaturas más altas registradas en la ciudad de Quito cada año.

```
CREATE OR REPLACE VIEW ANIO_MAS_CALUROSO
AS
(
SELECT DECODE(M.D_TIE_MUE_ANIO,','TEMPERATURA MÁS
ALTA',M.D_TIE_MUE_ANIO) AÑO,
MAX(H.D_HEC_VALOR) VALOR
FROM HECHOS H, D_TIEMPO_MUESTRA M, D_HORA HO
WHERE M.D_TIE_MUE_CODIGO = HO.D_TIE_MUE_CODIGO AND
HO.D_HOR_CODIGO = H.D_HOR_CODIGO AND
H.D_ELE_CODIGO = 'T'
GROUP BY CUBE(M.D_TIE_MUE_ANIO)
)
```

**Año más lluvioso:** Está vista permite adquirir en una escala descendente la mayor cantidad de lluvia registradas al año en la ciudad de Quito.

```
CREATE OR REPLACE VIEW ANIO_MAS_LLUVIOSO
AS
(
SELECT M.D_TIE_MUE_ANIO,
SUM(H.D_HEC_VALOR) VALOR
FROM HECHOS H, D_TIEMPO_MUESTRA M, D_HORA HO
WHERE M.D_TIE_MUE_CODIGO = HO.D_TIE_MUE_CODIGO AND
HO.D_HOR_CODIGO = H.D_HOR_CODIGO AND
H.D_ELE_CODIGO = 'L'
GROUP BY CUBE(M.D_TIE_MUE_ANIO)
)
```

**Cantidad de Lluvia:** Está vista permite conseguir por trimestre la cantidad de lluvia registradas por año en la ciudad de Quito.

```
CREATE OR REPLACE VIEW CANTIDAD_DE_LLUVIA
AS
(
SELECT
M.D_TIE_MUE_ANIO,
DECODE(M.D_TIE_MUE_TRIMESTRE,'1', 'PRIMER TRIMESTRE', '2', 'SEGUNDO
TRIMESTRE',
```

```

        '3', 'TERCER TRIMESTRE', '4', 'CUARTO TRIMESTRE', 'TODO EL AÑO')
    TRIMESTRE,
    SUM(H.D_HEC_VALOR) VALOR
    FROM HECHOS H, D_TIEMPO_MUESTRA M, D_HORA HO
    WHERE M.D_TIE_MUE_CODIGO = HO.D_TIE_MUE_CODIGO AND
    HO.D_HOR_CODIGO = H.D_HOR_CODIGO AND
    H.D_ELE_CODIGO = 'L'
    GROUP BY CUBE(M.D_TIE_MUE_ANIO, M.D_TIE_MUE_TRIMESTRE)
)

```

**Elementos Meteorológicos:** Esta vista permite consolidar la información de todos los elementos meteorológicos, dependiendo del cada uno de ellos:

- Lluvia: Cantidad total de lluvia registrada por año
- Presión Atmosférica: El valor máximo de la presión atmosférica registrada al año.
- Humedad relativa: el promedio de la humedad relativa por año.
- Temperatura: El promedio de la temperatura del aire por año.

```

CREATE OR REPLACE VIEW ELEMENTOS_METEOROLOGICOS
AS
(
    SELECT ANIO, NOMBRE, VALOR
    FROM
    (SELECT M.D_TIE_MUE_ANIO ANIO ,EM.D_ELE_NOMBRE NOMBRE,
    SUM(H.D_HEC_VALOR) VALOR
    FROM HECHOS H, D_TIEMPO_MUESTRA M, D_HORA HO,
    D_ELEMENTO_METEOROLOGICO EM
    WHERE M.D_TIE_MUE_CODIGO = HO.D_TIE_MUE_CODIGO AND
    HO.D_HOR_CODIGO = H.D_HOR_CODIGO AND
    H.D_ELE_CODIGO = 'L' AND
    EM.D_ELE_CODIGO = H.D_ELE_CODIGO
    GROUP BY (M.D_TIE_MUE_ANIO,EM.D_ELE_NOMBRE)
    ORDER BY VALOR DESC)
    UNION
    SELECT ANIO,NOMBRE, VALOR
    FROM(SELECT M.D_TIE_MUE_ANIO ANIO ,EM.D_ELE_NOMBRE NOMBRE,
    MAX(H.D_HEC_VALOR) VALOR
    FROM HECHOS H, D_TIEMPO_MUESTRA M, D_HORA HO,
    D_ELEMENTO_METEOROLOGICO EM
    WHERE M.D_TIE_MUE_CODIGO = HO.D_TIE_MUE_CODIGO AND
    HO.D_HOR_CODIGO = H.D_HOR_CODIGO AND
    H.D_ELE_CODIGO = 'P' AND
    EM.D_ELE_CODIGO = H.D_ELE_CODIGO
    GROUP BY (M.D_TIE_MUE_ANIO,EM.D_ELE_NOMBRE)
    ORDER BY VALOR DESC)
    UNION
    SELECT ANIO,NOMBRE, VALOR
    FROM (SELECT M.D_TIE_MUE_ANIO ANIO ,EM.D_ELE_NOMBRE NOMBRE,
    AVG(H.D_HEC_VALOR) VALOR
    FROM HECHOS H, D_TIEMPO_MUESTRA M, D_HORA HO,
    D_ELEMENTO_METEOROLOGICO EM

```

```

WHERE M.D_TIE_MUE_CODIGO = HO.D_TIE_MUE_CODIGO AND
HO.D_HOR_CODIGO = H.D_HOR_CODIGO AND
H.D_ELE_CODIGO = 'T' AND
EM.D_ELE_CODIGO = H.D_ELE_CODIGO
GROUP BY (M.D_TIE_MUE_ANIO,EM.D_ELE_NOMBRE )
ORDER BY VALOR DESC)
UNION
SELECT ANIO,NOMBRE, VALOR
FROM (SELECT M.D_TIE_MUE_ANIO ANIO ,EM.D_ELE_NOMBRE NOMBRE,
AVG(H.D_HEC_VALOR) VALOR
FROM HECHOS H, D_TIEMPO_MUESTRA M, D_HORA HO,
D_ELEMENTO_METEOROLOGICO EM
WHERE M.D_TIE_MUE_CODIGO = HO.D_TIE_MUE_CODIGO AND
HO.D_HOR_CODIGO = H.D_HOR_CODIGO AND
H.D_ELE_CODIGO = 'H' AND
EM.D_ELE_CODIGO = H.D_ELE_CODIGO
GROUP BY (M.D_TIE_MUE_ANIO,EM.D_ELE_NOMBRE)
ORDER BY VALOR DESC)
)

```

Para una ver los resultados de las consultas realizadas sobre las vistas creadas, **ver anexo 5**

### 2.2.8. INGRESO DE NUEVOS DATOS

Si se desean registrar más datos al Data Warehouse se deberá considerar las siguientes particularidades, restricciones y estándares para el ingreso de datos:

- **Código del elemento meteorológico**
- **Código de la hora en que se ha tomado la muestra**
- **Valor de la muestra**
- **El código de la observación**

**Código del elemento meteorológico (D\_ELE\_CODIGO):** Para los elementos meteorológicos se usó como código único la primera letra del elemento, seguido del nombre y la unidad de medida, tal como se muestra en la tabla 2.5, en el caso de que se desee ingresar más elementos meteorológicos hay que tomar en cuenta que el identificador del elemento no puede repetirse, si ese fuese el caso se deberá agregar una segunda letra para diferenciarlo de los elementos que ya están creados.

Identificador Elemento Meteorológico	Nombre Elemento Meteorológico	Unidad de Medida
T	Temperatura del Aire	Grados Centígrados
H	Humedad Relativa	Tanto por Ciento
P	Presión Atmosférica Media	Milímetros de Mercurio
L	Cantidad de Lluvia	Milímetros cúbicos

**Tabla 2. 5 Elementos Meteorológicos**

**Código de la hora en que se ha tomado la muestra (D\_HORA\_CODIGO):**

Para facilidad del usuario el código de la hora tendrá el siguiente estándar:

- **<hora>-<fecha>:**
  - **Hora:** número entero comprendido entre 1 y 24
  - **Fecha:** La fecha tiene el siguiente formato: día, mes, año (ddmmaaaa). Para mayor comprensión ver el siguiente ejemplo: si una muestra fue registrada el 1 de enero del 2006, el código a ingresarse será: 01012006
  - Si la muestra fue registrada a la una de la mañana, entonces el código deberá ser:1-01012006

**Valor de la muestra:** Este es el valor más importante a ingresar ya que de este dependen los resultados que se puedan extraer del Data Warehouse a través de las diferentes herramientas expuestas en este proyecto de titulación.

Cuando se ingresen **datos de la temperatura del aire** estos deberán oscilar entre los 3°C y 30°C ya que en la ciudad de Quito la temperatura no sobrepasa estos límites.

Cuando se ingresen los **datos de la humedad relativa** estos no pueden tener valores negativos, ni sobrepasar el 100% y si por alguna razón no se tiene valores se deberá poner 0.

Al ingresar los **datos de la cantidad de lluvia** estos no pueden ser negativos.

Al ingresar los **datos de la presión atmosférica** estos deberán ser registrados cada hora para que de esta forma se pueda obtener resultados de los modelos a aplicarse.

**El código de la observación:** debe ser único para evitar errores o inconsistencia en la base de datos ante la posibilidad de duplicar los registros. El formato establecido es el que se muestra a continuación:

- **< D\_ELE\_CODIGO >-< D\_HORA\_CODIGO >:**
  - Siguiendo el ejemplo anterior ingresamos el valor de la humedad (H) el código de la observación deberá ser: H-1-01012006.

Debido a que el presente proyecto de titulación tiene como alcance el análisis de los datos meteorológicos en un periodo de 10 años, únicamente se sugiere una forma de como se debe importar nuevos datos en caso de ser requeridos.

1. Para migrar los datos al Data Warehouse se debe llenar la plantilla que se muestra en el **anexo 8**. Un ejemplo de la plantilla lista con los datos a ingresarse al Data Warehouse se puede ver en la tabla 2.6

D_HEC_CODIGO	D_ELE_CODIGO	D_HORA_CODIGO	D_HEC_OBS_VALOR
H-1-01012006	H	1-01012006	100
H-2-01012007	H	2-01012007	100
H-3-01012008	H	3-01012008	100
H-4-01012009	H	4-01012009	100
H-5-01012010	H	5-01012010	99
H-6-01012011	H	6-01012011	98
H-7-01012012	H	7-01012012	97
H-8-01012013	H	8-01012013	99
H-9-01012014	H	9-01012014	81
H-10-01012015	H	10-01012015	66
H-11-01012016	H	11-01012016	52
H-12-01012017	H	12-01012017	49
H-13-01012018	H	13-01012018	43
H-14-01012019	H	14-01012019	39
H-15-01012020	H	15-01012020	36
H-16-01012021	H	16-01012021	44

H-17-01012022	H	17-01012022	45
H-18-01012023	H	18-01012023	49
H-19-01012024	H	19-01012024	71
H-20-01012025	H	20-01012025	83
H-21-01012026	H	21-01012026	98
H-22-01012027	H	22-01012027	100
H-23-01012028	H	23-01012028	100
H-24-01012029	H	24-01012029	100
H-1-02012030	H	1-02012030	100
H-2-02012031	H	2-02012031	100
H-3-02012032	H	3-02012032	99
H-4-02012033	H	4-02012033	99
H-5-02012034	H	5-02012034	98
H-6-02012035	H	6-02012035	98
H-7-02012036	H	7-02012036	97
H-8-02012037	H	8-02012037	99
H-9-02012038	H	9-02012038	70
H-10-02012039	H	10-02012039	51
H-11-02012040	H	11-02012040	46
H-12-02012041	H	12-02012041	41
H-13-02012042	H	13-02012042	40
H-14-02012043	H	14-02012043	39
H-15-02012044	H	15-02012044	41
H-16-02012045	H	16-02012045	42
H-17-02012046	H	17-02012046	53
H-18-02012047	H	18-02012047	72
H-19-02012048	H	19-02012048	79
H-20-02012049	H	20-02012049	97
H-21-02012050	H	21-02012050	100
H-22-02012051	H	22-02012051	100
H-23-02012052	H	23-02012052	100
H-24-02012053	H	24-02012053	99

**Tabla 2.6 Ingreso de datos**

2. Para mayor facilidad al momento de ingresar los datos se recomienda usar una hoja de cálculo.
3. Debido a que en este proyecto de titulación se uso la herramienta SQLLOUDER para la carga de datos (ver capítulo 2, modelado de datos), se sugiere usarla para el ingreso de datos al Data Warehouse.
4. El archivo de control para el ingreso de datos tiene el siguiente formato:

```
LOAD DATA
INFILE *
INTO TABLE <TABLA>
APPEND
```

*FIELDS TERMINATED BY '<DELIMITADOR>' OPTIONALLY ENCLOSED ''''*  
*TRAILING NULLCOLS*  
*<ATRIBUTOS> BEGINDATA*  
*Datos separados con el delimitador “;”*

Y la extensión con la que se guarda el archivo de control tiene la extensión .ctl

```
LOAD DATA
INFILE *
INTO TABLE OBSERVACION
APPEND
FIELDS TERMINATED BY ',' OPTIONALLY ENCLOSED ''''
TRAILING NULLCOLS
(D_HEC_ELE_CODIGO,D_ELE_CODIGO,D_HEC_OBS_VALOR)
H-1-01012006;H;1-01012006;100
H-2-01012007;H;2-01012007;100
H-3-01012008;H;3-01012008;100
H-4-01012009;H;4-01012009;100
H-5-01012010;H;5-01012010;99
H-6-01012011;H;6-01012011;98
H-7-01012012;H;7-01012012;97
H-8-01012013;H;8-01012013;99
H-9-01012014;H;9-01012014;81
H-10-01012015;H;10-01012015;66
H-11-01012016;H;11-01012016;52
H-12-01012017;H;12-01012017;49
H-13-01012018;H;13-01012018;43
H-14-01012019;H;14-01012019;39
H-15-01012020;H;15-01012020;36
H-16-01012021;H;16-01012021;44
H-17-01012022;H;17-01012022;45
H-18-01012023;H;18-01012023;49
H-19-01012024;H;19-01012024;71
H-20-01012025;H;20-01012025;83
H-21-01012026;H;21-01012026;98
H-22-01012027;H;22-01012027;100
H-23-01012028;H;23-01012028;100
H-24-01012029;H;24-01012029;100
H-1-02012030;H;1-02012030;100
```

H-2-02012031;H;2-02012031;100  
 H-3-02012032;H;3-02012032;99  
 H-4-02012033;H;4-02012033;99  
 H-5-02012034;H;5-02012034;98  
 H-6-02012035;H;6-02012035;98  
 H-7-02012036;H;7-02012036;97  
 H-8-02012037;H;8-02012037;99  
 H-9-02012038;H;9-02012038;70  
 H-10-02012039;H;10-02012039;51  
 H-11-02012040;H;11-02012040;46  
 H-12-02012041;H;12-02012041;41  
 H-13-02012042;H;13-02012042;40  
 H-14-02012043;H;14-02012043;39  
 H-15-02012044;H;15-02012044;41  
 H-16-02012045;H;16-02012045;42  
 H-17-02012046;H;17-02012046;53  
 H-18-02012047;H;18-02012047;72  
 H-19-02012048;H;19-02012048;79  
 H-20-02012049;H;20-02012049;97  
 H-21-02012050;H;21-02012050;100  
 H-22-02012051;H;22-02012051;100  
 H-23-02012052;H;23-02012052;100  
 H-24-02012053;H;24-02012053;99

5. Una vez creado el archivo de control se deberá correr el siguiente comando en una ventana DOS para el ingreso a la base de datos:

***sqlldr miusuario/mipasword@mibase ejemplo.ctl rows=1 errors=30000***

*miusuario: Usuario de la Instancia*

*mipassword: Password asignado*

*mibase: Nombre de la base de Datos que se utilizando*

*ejemplo.ctl: nombre que se le dio al archivo de control que contiene los datos a ser importados*

*rows=1 : Significa que cada 1 registros ira haciendo COMMIT*

*errors=30000 : Significa el número de errores que puede soportar*

Como se aprecia, sqldr requiere de un archivo de control (control file), el cual especifica cómo deben cargarse los datos en la base de datos; y un fichero de datos (data file), que especifica qué datos deben ser cargados.

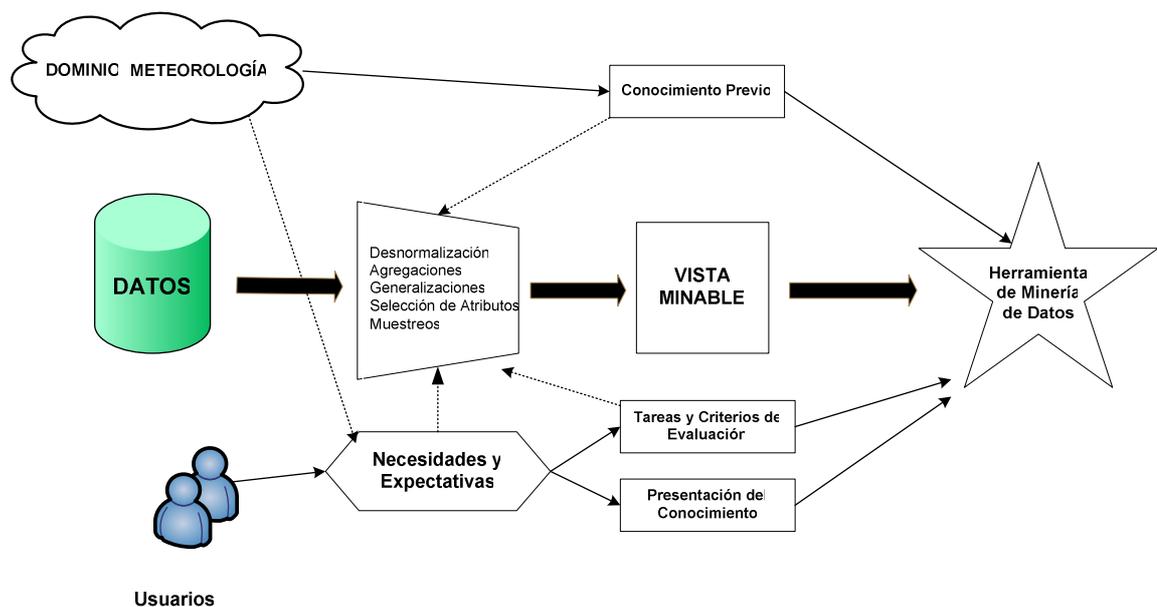
6. Una vez ingresados los datos al Data Warehouse se pueden aplicar las herramientas de reportes y Clementine para el análisis de los datos, a través de las rutas que se detallan en el capítulo siguiente.

## CAPITULO 3. MINERIA DE DATOS

### 3.1.EXPLORACIÓN Y SELECCIÓN DE DATOS

Al tener una fuente de datos, tan completa como es un Data Warehouse, el siguiente paso es explorar y seleccionar los datos para realizar un análisis exploratorio con el objetivo de poder establecer patrones de comportamiento.

No solo es necesario conocer los datos y el dominio del que provienen, sino que se debe explorar los datos, el contexto y los usuarios. Por otra parte si no se selecciona y se estudia un subconjunto de datos, no se puede saber que validez pueden tener y si finalmente van hacer útiles, solo al determinar el método se puede saber si hay ciertos atributos que harán falta cambiar o eliminar.



**Figura 3. 1 Proceso para tener datos y Aplicar una herramienta de Minería de Datos**

La figura 3.1, muestra el proceso que se debería seguir para poder tener un conjunto de datos que puedan ser procesados por una herramienta de minería de datos. No solo es necesario tener el conjunto de datos con los atributos relevantes, sino que debe ir acompañada de la tarea a realizar sobre ella y como evaluarla, así como la forma de presentar el resultado final.

El primer paso para la preparación de datos es conocer el problema a resolver para lo cual se deberá incluir como actividad preliminar la comprensión del dominio o negocio: el propósito es asegurar el entendimiento del negocio y objetivos del proyecto, o al menos hacia que objetivo se quiere llegar, sin esto, nos resulta imposible conocer los datos que debemos extraer.

### **3.1.1. VISTA MINABLE**

La vista minable es un conjunto de datos que recoge toda la información necesaria para realizar la tarea de minería de datos. La preparación de los mismos no es un proceso automático, por lo cual es necesario aplicar nuestro conocimiento para generar el conjunto de datos necesario para poder aplicar un modelo de explotación. Por lo antes mencionado se puede definir que el principal objetivo de la preparación de datos, es tomar información manipularla, transformarla y presentarla para que pueda ser procesada por un modelo de minería de datos.

### **3.1.2. RECONOCIMIENTO DEL DOMINIO Y DE LOS USUARIOS.-**

Es importante determinar las decisiones que se toman frecuentemente y a partir de que modelo se toman, si estos tienen una base sólida o simples reglas de negocio. Es importante también determinar quien usará el conocimiento obtenido y que tipo de presentación puede ser más aconsejable. De los datos obtenidos es necesario seguir transformando y seleccionándolos con el objetivo de conseguir una vista minable, lista para ser tratada por las herramientas de minería de datos. Para ello se puede utilizar distintas técnicas para obtener o refinar dicha vista: visualización, descripción, generalización, agregación y selección.

Se debe distinguir el término análisis exploratorio de datos definido como una serie de técnicas para investigar los datos, ver tendencias, patrones, errores y características.

**Visualización:** las técnicas de visualización de datos se utilizan fundamentalmente por dos objetivos:

- Aprovechar la gran capacidad humana de ver patrones, anomalías y tendencias a partir de imágenes y facilitar la comprensión de datos
- Ayudar al usuario a comprender más rápidamente patrones descubiertos automáticamente por un sistema KDD (descubrimiento de conocimiento en base de datos).

### 3.2.MINERÍA DE DATOS

Es el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos. Es decir, la tarea fundamental de la minería de datos es encontrar modelos explicables a partir de los datos. Para que este proceso sea efectivo debería ser automático o semiautomático y el uso de los patrones descubiertos debería ayudar a la investigación de planes de contingencia o proyectos que reporten beneficios al OAQ.

El conocimiento puede ser en forma de relaciones, patrones o reglas inferidas de los datos y desconocidos o bien en forma de una descripción más concisa. Estas relaciones constituyen el modelo de los datos analizados. Existen muchas formas de representar los modelos y cada una de ellas determina el tipo de técnica que puede usarse para explicarlos.

Los modelos pueden ser de dos tipos: Predictivos y Descriptivos

**Los modelos Predictivos:** pretenden estimar valores futuros o desconocidos de variables de interés que se denominan variables dependientes, usando otras variables de las bases de datos a las que nos referimos como variables predicativas.

**Los modelos Descriptivos:** identifican patrones que explican o resumen los datos, es decir sirven para explorar las propiedades de los datos examinados, no para predecir nuevos datos.

Para el tipo de información que se debe obtener en este proyecto de titulación se usará el modelo descriptivo.

### 3.2.1. TAREAS DE MINERÍA DE DATOS

Dentro de la minería de datos se puede distinguir tipos de tareas, cada una de las cuales puede considerarse como un tipo de problema a ser resuelto por un algoritmo de minería de datos. Esto significa que cada tarea tiene sus propios requisitos y que el tipo de información obtenida con una tarea puede diferir mucho de la obtenida con otra.

Para el caso de estudio se aplicara alguno de los modelos descriptivos, ya que el objetivo de este proyecto es identificar los patrones meteorológicos que se puedan obtener con los datos consolidados.

Entre las tareas descriptivas encontramos el agrupamiento (clustering), reglas de asociación, reglas de asociación secuencial y correlacional.

El agrupamiento consiste en obtener grupos “naturales” a partir de los datos. Los datos son agrupados basándose en el principio de maximizar la similitud entre los elementos de un grupo y minimizando la similitud entre los distintos grupos.

Las correlaciones son una tarea descriptiva que se usa para examinar el grado de similitud de los valores de dos variables numéricas. Una fórmula estándar para medir la correlación lineal es el coeficiente de correlación  $r$ , el cual es un valor comprendido entre 1 y -1. Esto quiere decir que cuando  $r$  es positivo, las variables tienen un comportamiento similar (ambas crecen o decrecen al mismo tiempo) y cuando  $r$  es negativo si una variable crece la otra decrece.

Las reglas de asociación identifican relaciones no explícitas entre atributos categóricos no implican una relación causa efecto, es decir, puede no existir una causa para que los datos estén asociados. Un caso especial de las reglas de asociación, son las reglas secuenciales, se usa para determinar patrones secuenciales en los datos.

### **3.2.2. TÉCNICAS DE MINERÍA DE DATOS**

Existen varias técnicas utilizadas para la fase de minería de datos, tales como: técnicas de inferencia estadística, árboles de decisión, redes neuronales, inducción de reglas, algoritmos genéticos.

#### **Redes Neuronales**

Una red neural artificial es un sistema computacional de elementos de proceso (neuronas artificiales) altamente interconectados, los cuales contienen algoritmos que le permiten responder a ciertos estímulos. Como se explicó, consiste de elementos de proceso interconectados con una cierta función de ponderación.

El proceso de aprendizaje consiste básicamente en indicarle las salidas que debe producir para determinadas entradas y la red auto adapta las ponderaciones de las interconexiones para producir el efecto deseado. Una vez logrado el aprendizaje, la red responderá asociando las señales de entrada, a través de procesos múltiples, para dar respuesta a sus salidas de acuerdo con la enseñanza obtenida.

Con lo que corresponde a la minería de datos las redes neuronales permiten modelizar problemas complejos en el que puede haber interacciones no lineales entre variables. Las redes neuronales pueden usarse en problemas de clasificación, de regresión y de agrupamiento, las redes neuronales trabajan directamente

#### **Árboles de Decisión**

Son una serie de condiciones organizadas en forma jerárquica a forma de árbol. Son muy útiles para encontrar estructuras en espacios de alta

dimensionalidad y en problemas que mezclen datos categóricos y numéricos, esta técnica se usa en tareas de clasificación, agrupamiento y regresión, los árboles de decisión usados para predecir variables categóricas reciben el nombre de arbole de clasificación ya que distribuyen las instancias en clases. Cuando los árboles de decisión se usan para predecir variables continuas se llaman árboles de regresión.

Los árboles de decisión siguen una aproximación “divide y vencerás” y pueden considerarse una forma de aprendizaje de reglas, ya que cada rama puede interpretarse como una regla, donde los nodos internos en el camino desde la raíz hasta las hojas definen los términos de la conjunción que constituye el antecedente de la regla, en la clase asignada en la hoja es el consecuente.

### **Algoritmos Genéticos**

Son métodos de búsqueda colectiva en el espacio de soluciones dada una población de potenciales soluciones a un problema, la computación evolutiva expande esta población con nuevas y mejores soluciones.

Los algoritmos genéticos se pueden usar para el agrupamiento, la clasificación y las reglas de asociación, así como para la selección de atributos.

Para ello, se usa una función de adaptación, que selecciona los mejores modelos que sobrevivirán o serán cruzados. Los distintos algoritmos genéticos difieren en la forma en que se representan los modelos, como se combinan los individuos en los modelos, si existen mutaciones y como se estas, y como se usa la función de adaptación.

### **3.2.3. CONSTRUCCIÓN DEL MODELO Y VALIDACIÓN.**

La construcción del modelo requiere que los datos preparados puedan ser usados iterativamente, en otras palabra poder aplicar algoritmos y técnicas sobre las diferentes vistas minables que sean necesarias, y de esta manera descubrir patrones de comportamiento. Para poder conseguir los mejores patrones se tendrá inclusive que retrocede a fases realizadas anteriormente y

hacer cambios en los datos usados, esto con el objetivo de obtener una información consistente para poder validarla.

Para el presente proyecto de titulación, nos servimos de la herramienta de minería de datos, Clementine 11.1, el cual se caracteriza por:

**Acceso a Datos:** Fuente de Datos ODBCs, tablas Excel y archivos planos.

**Preprocesado de Datos:** Muestreo, particiones, reordenación de campos, nuevas estrategias para la fusión de tablas.

**Técnicas de Aprendizaje:** Árboles de Decisión, C5.0 y el C&RT, Redes Neuronales (Redes de Kohonen), Agrupamiento (K Medias), Reglas de Asociación (GRI, A priori, etc.), Regresión Lineal y Logística, Combinación de modelos.

**Visualización de Resultados:** Clementine ofrece un soporte gráfico que permite al usuario tener una visión global de todo el proceso, que comprende desde el análisis del problema, hasta la imagen final del modelo aprendido. Se dispone de:

**Gráficos estadísticos:** Histogramas, diagramas de dispersión, etc.

**Gráficos 3D y Animados.**

Navegadores de Árboles de Decisión, Reglas de Asociación, Redes Neuronales de Kohonen, etc.

Exportación: Generación automática de informes (HTML y texto).

Para la construcción del modelo se han creado vistas y analizado elementos meteorológicos relacionados entre sí.

### **3.3.RECONOCIMIENTO DE PATRONES DE COMPORTAMIENTO**

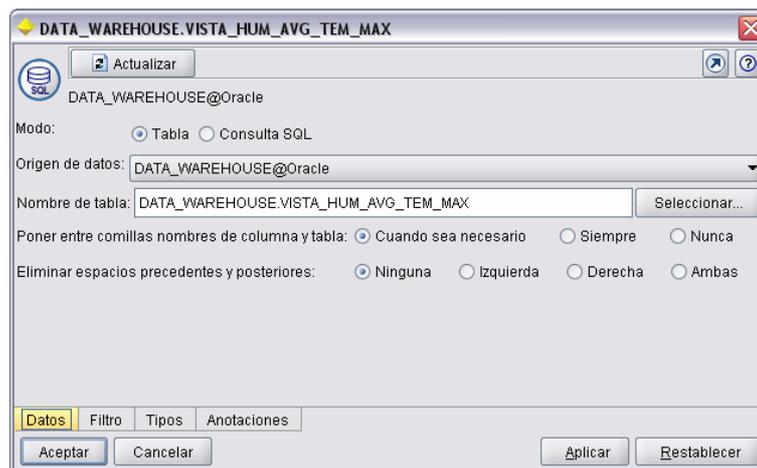
***Temperatura del Aire y Humedad Relativa:***

Conocimiento Previo: A Mayor temperatura, mayor humedad Relativa

Para poder obtener este patrón fue requerido analizar tanto la temperatura máxima, media y mínima, contra el promedio de la humedad relativa. Por lo que se creó una vista que contenga esa información. **Ver Anexo 7**

Al tener las diversas Vistas procedemos a diseñar y esquematizar en la herramienta Clementine. **Ver anexo 6**

Se empieza a diseñar la ruta, se debe conectar a la base de datos para obtener el conjunto de datos a ser analizados, como se puede apreciar en el figura 3.2 Conexión a Vista Minable.



**Figura 3.2 Conexión a la Vista Minable Vista**

Para este ejemplo por tener una gran cantidad de datos como se aprecia en la figura 3.3 Nodo Seleccionar, procederemos a filtrarlos.

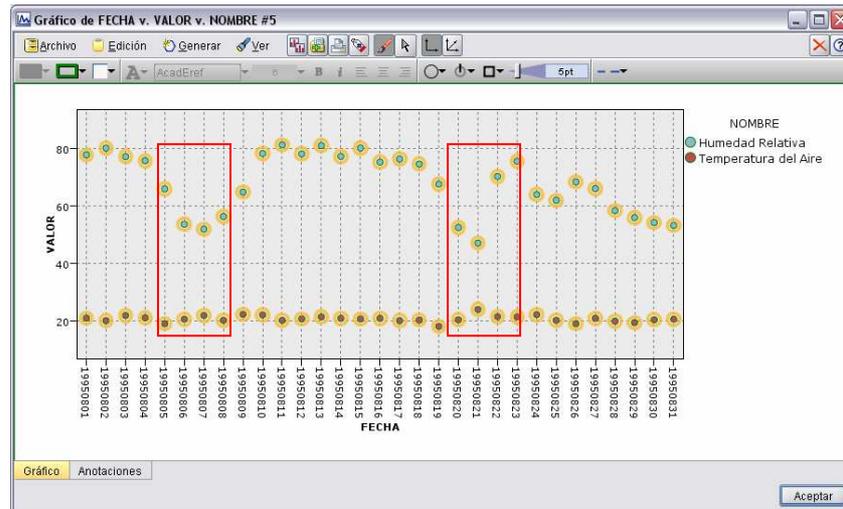


**Figura 3.3 Nodo Seleccionar.**

Aplicar una herramienta para graficar en tres dimensiones, como lo indica la figura 3.4 y verificamos el comportamiento de los elementos relacionados, temperatura y humedad relativa en base al gráfico obtenido, ver figura 3.5 Temperatura y Humedad



**Figura 3.4 Nodo Gráfico 3D**



**Figura 3.5 Temperatura y Humedad**

Ver la figura 3.5, se puede apreciar que la humedad relativa disminuye cuando la temperatura aumenta, comprobando de esta forma un patrón de comportamiento.

### ***Temperatura:***

Conocimiento Previo: No se tiene conocimientos previos, por lo que se espera obtener un patrón de comportamiento de la temperatura desde el 1995 hasta el 2005

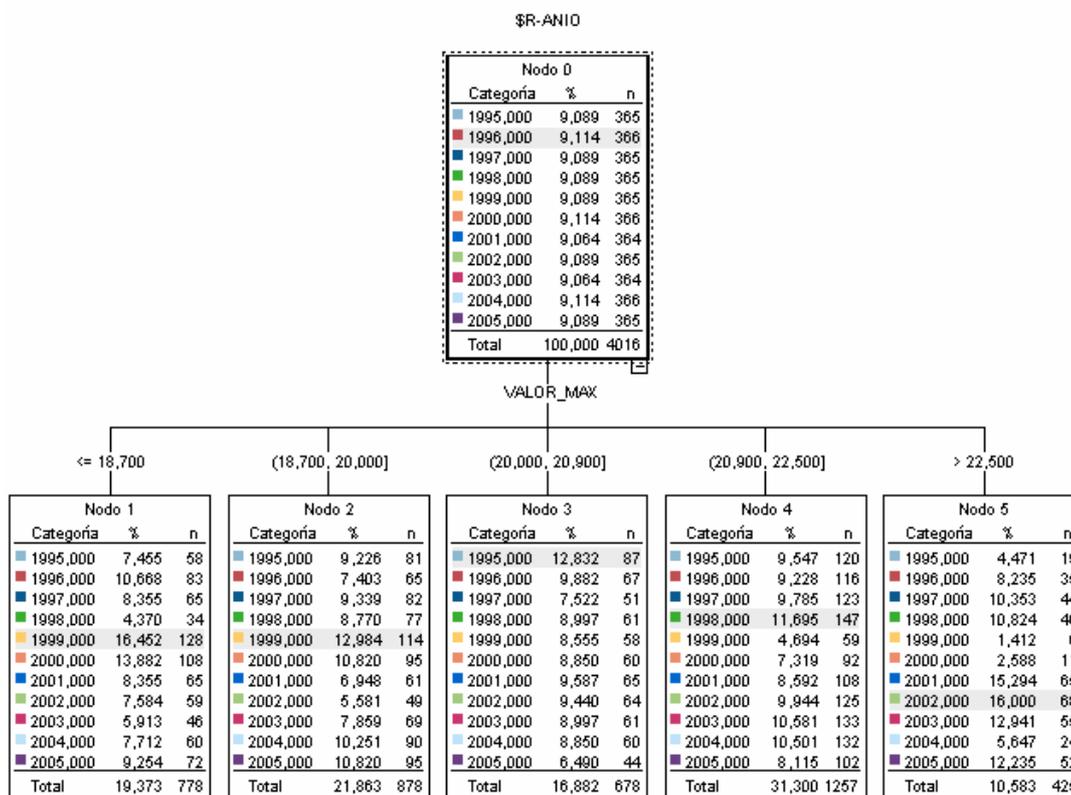
Para poder obtener este patrón fue requerido analizar la temperatura máxima, media y mínima, aplicando el modelo CHAID sobre la vista Minable Temperatura.

CHAID, o detección automática de interacciones mediante chi-cuadrado (del inglés Chi-squared Automatic Interaction Detection), es un método de clasificación para generar árboles de decisión mediante estadísticos de chi-cuadrado para identificar divisiones óptimas.

CHAID examina en primer lugar las tablas de tabulación cruzada entre las distintas variables y los resultados para, a continuación, comprobar la significación mediante una comprobación de independencia de chi-cuadrado. Si varias de estas relaciones son estadísticamente importantes, CHAID seleccionará la

variable de mayor relevancia (el valor P más pequeño). Si la variable cuenta con más de dos categorías, se compararán estas categorías y se contraerán las que no presenten diferencias en los resultados. Para ello, se unirá el par de categorías que presenten menor diferencia, y así sucesivamente. Este proceso de fusión de categorías se detiene cuando todas las categorías restantes difieren entre sí en el nivel de comprobación especificado.

Como se describe en los gráficos de las figuras 3.6, 3.7 y 3.8 sobre la aplicación del modelo CHAID sobre la temperatura, muestra la disminución de las temperaturas desde 1995 hasta el 2005



**Figura 3.6 Temperatura Máxima**

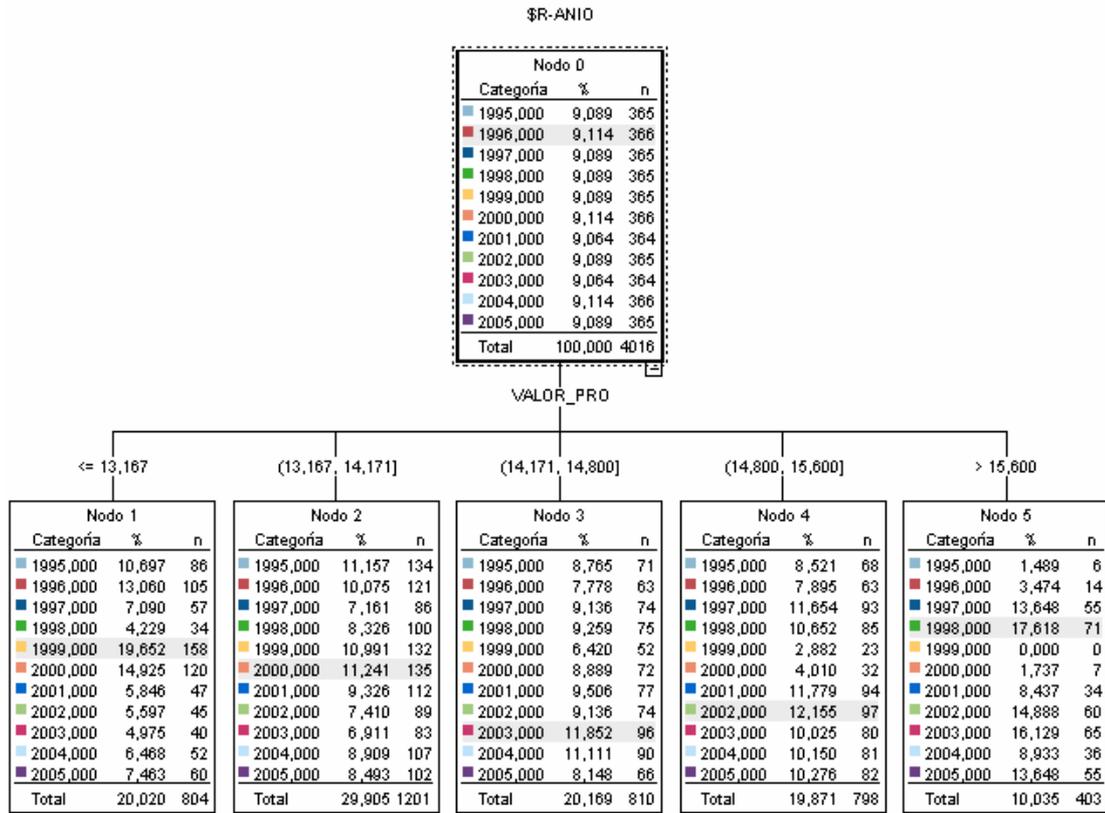


Figura 3.7 Temperatura Promedio

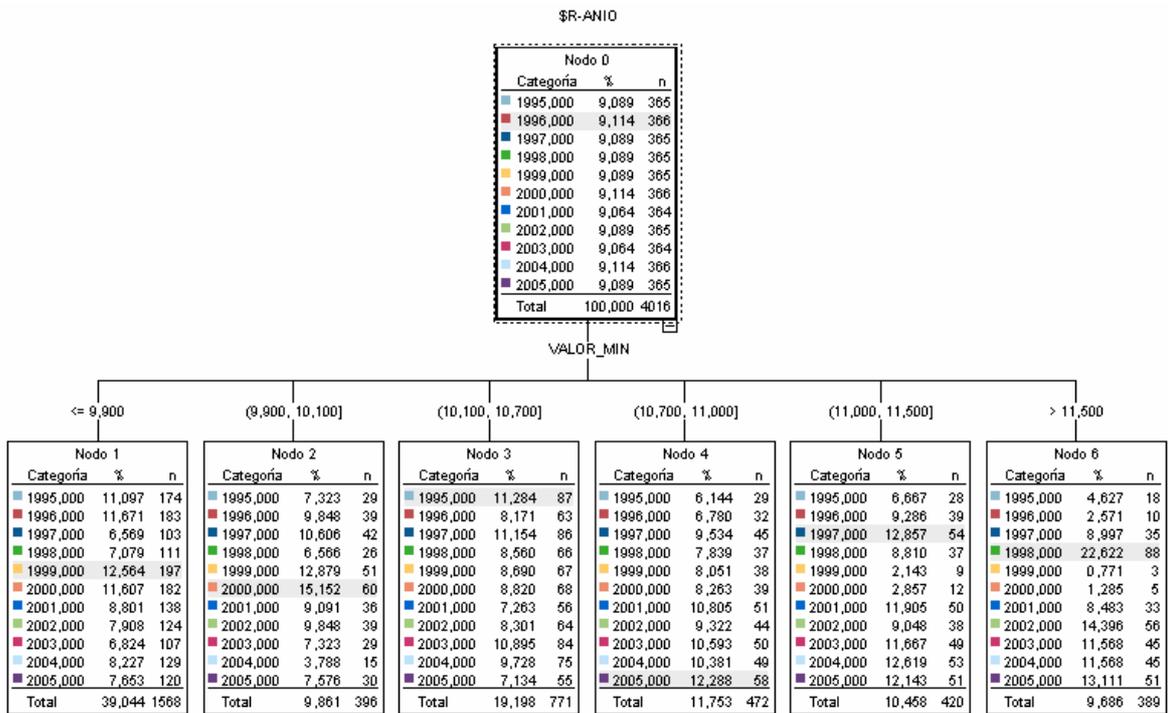


Figura 3.8 Temperatura Mínima

### 3.4.REPORTE CON LOS RESULTADOS OBTENIDOS

#### Temperatura

La figura 3.9 muestra la variación de la temperatura mínima, promedio y máxima, desde el año 1995 hasta el año 2005.

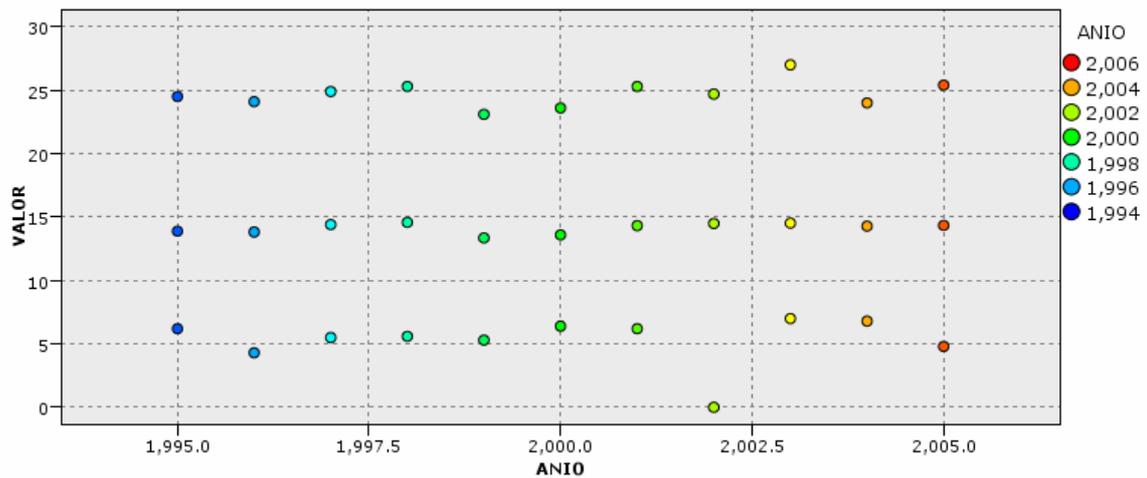


Figura 3.9 Temperatura Mínima, Promedio y Máxima por año

#### Humedad

La figura 3.10 muestra la variación de la humedad mínima, promedio y máxima, desde el año 1995 hasta el año 2005.

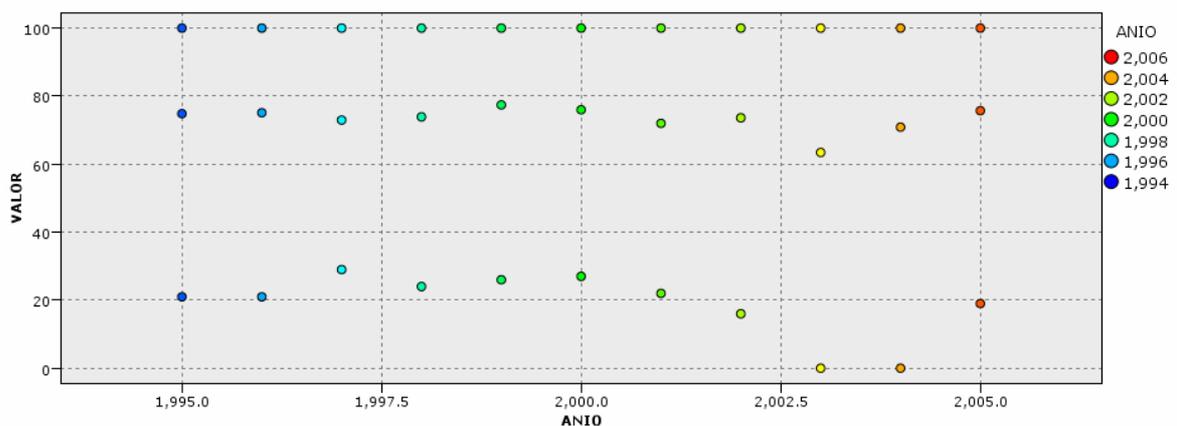
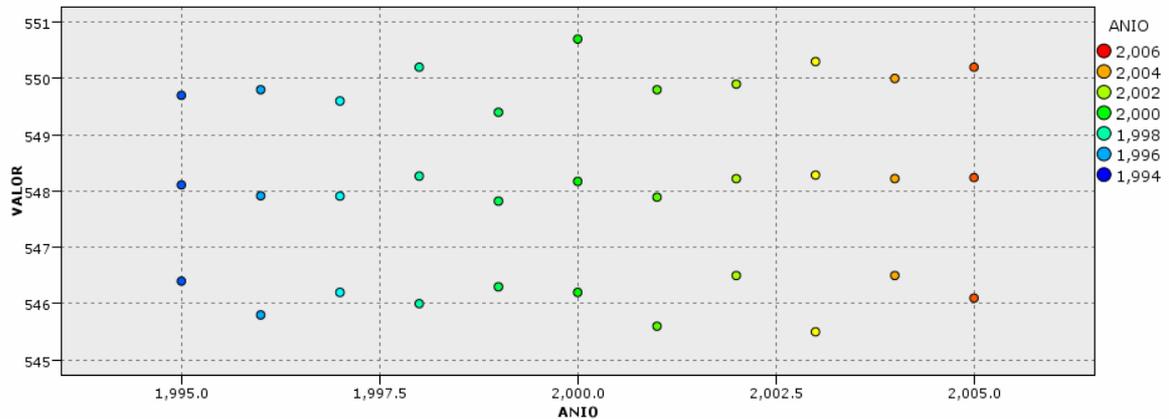


Figura 3.10 Humedad Mínima, Promedio y Máxima por año

## Presión

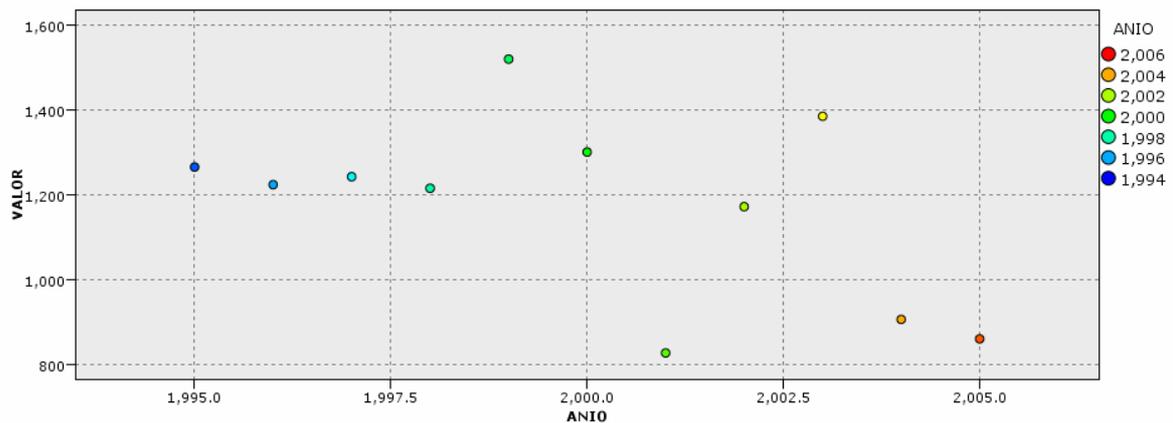
La figura 3.11 muestra la variación de la presión mínima, promedio y máxima, desde el año 1995 hasta el año 2005.



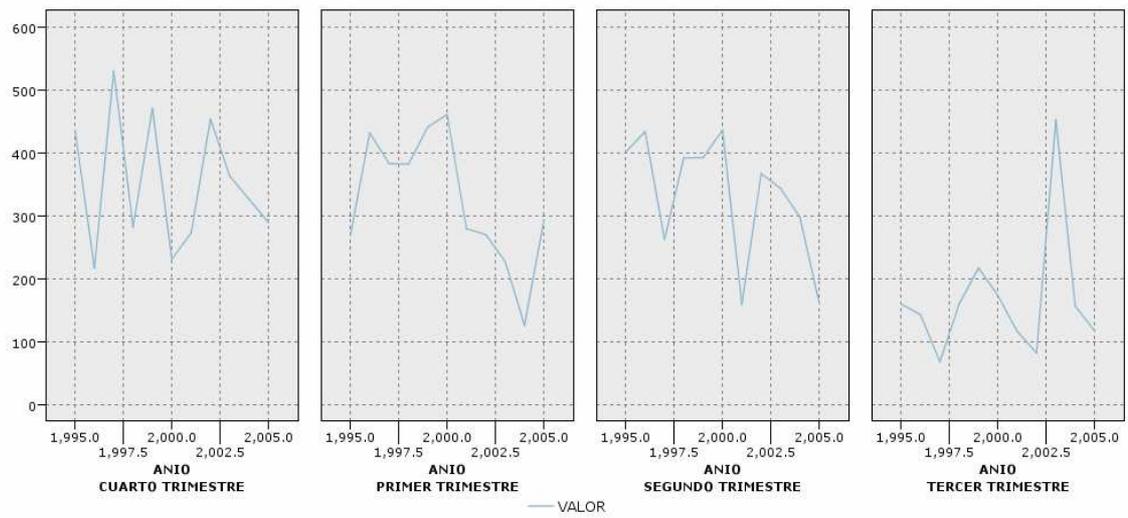
**Figura 3.11 Presión Mínima, Promedio y Máxima por año**

## Cantidad de Lluvia

La figura 3.12 muestra la cantidad de lluvia caída sobre la ciudad de Quito, desde el año 1995 hasta el 2005, para poder comprender mejor el comportamiento de la lluvia, se ha separado por trimestres como se muestra en la figura 3.13



**Figura 3.12 Cantidad de Lluvia caída sobre el Quito por año**



**Figura 3.13 Cantidad de Lluvia caída sobre el Quito por separado por trimestres**

## **CAPITULO 4. CONCLUSIONES Y RECOMENDACIONES**

### **4.1. CONCLUSIONES**

Existen muchas fuentes de datos meteorológicos pero no se ha realizado una publicación formal actual del comportamiento climatológico de la ciudad de Quito.

La fase más importante es la de limpieza y carga de datos, ya que de la calidad de los mismos dependerá que el análisis que se realicen sobre las vistas creadas, den como resultado patrones de comportamiento, que se ajusten a la información requerida.

Al momento de la creación de la base de datos es requerido que se realicen pruebas y cálculos para poder asignar tables spaces con un tamaño razonable, ya que como es de esperarse un DataWarehouse seguirá creciendo conforme los datos sean ingresados.

Al haber sido ingresado los datos a Excel manualmente se tuvo que revisar las fajas hora por hora ya que se evidenció problemas al momento de la carga de datos

Este caso en particular de DataMining se debió tomar en cuenta un análisis de los datos para la migración conociendo de antemano el número total de datos, para poder verificar en la BDD transaccional

Si los datos se encuentran en diferentes tables spaces es necesario ejecutar commit después de cada transacción para que se puedan visualizar los cambios realizados en diferentes esquemas

Para el análisis de las vistas minables no es preciso aplicar modelos o algoritmos para poder detectar patrones, ya que como se comprobó al analizar la temperatura y la humedad solo fue necesaria la graficación de la vista.

Hallar el modelo que se acople para encontrar los patrones de comportamiento, podrán necesitar redefinir las consultas y hasta las fases previas al aplicar los modelos.

Los patrones de comportamiento generados se han basado en conocimiento previo, o es el resultado de aplicar un modelo como el CHAID, pero no ha sido posible con la cantidad de datos almacenados actualmente obtener más patrones de comportamiento, ya que según el personal del OAQ, se necesita datos almacenados de cincuenta años como mínimo y que los datos recolectados sean registrados diariamente, ya que para determinar más patrones de comportamiento se necesita periodos de tiempos largos además de registros de datos completos.

## **4.2.RECOMENDACIONES**

De acuerdo al motor de base de dato se debe tomar en cuenta el formato que se usa ya que puede presentar problemas al momento de la carga de datos

Se recomienda seguir siempre una metodología de trabajo en todo el proceso, ya que se pueden encontrar con situaciones o nuevos requerimiento a medida que se desarrolla los análisis para determinar en este caso patrones o algún tipo de proyecto.

Se debe unificar los datos de las tres estaciones meteorológicas de la ciudad de Quito (OAQ, INAMHI, DAC) con el fin de tener datos consolidados.

Para poder utilizar de una mejor manera las herramientas de minería de datos es recomendable el registro continuo de las muestras, ya que actualmente por no tener toda la información no es posible aplicar más modelos.

## BIBLIOGRAFIA

- [1] ALVARO, Marco; Boletín Meteorológico del Observatorio Astronómico de Quito, 2001.
- [2] ANAHORY, Sam, MURRAY Dennis; Data Warehousing in the Real World; ADDISON-WESLEY; Harlow 1997
- [3] BLANDIN LANDIVAR, Carlos; Análisis y Estudios Climatológicos en el Ecuador; INAMHI; Quito 1989.
- [4] DONN, Willian L. ;Meteorología ; REVERTE; Barcelona 1978.
- [5] GILL, Harjinder , Data Warehousing; PRENTICE HALL; México
- [6] HERNANDEZ, Jose, RAMIREZ, María José; Instrucción a la Minería de Datos; PRENTICE HALL; Madrid 2004
- [7] INEDITO; Condiciones Climáticas de Lugares Turísticos del Ecuador; INAMHI; Quito 1999.
- [8] JARKE, Matthias; Fundamentals of Data Warehouses; SPRINGER; Berlin 2000.
- [9] LLAUGE, Félix; La Meteorología; MARCOMBO; Barcelona 1971.
- [10] LOPEZ, Ericson; 132 Años de Historia del Observatorio Astronómico de Quito; NINA Comunicaciones; Edición 1, 2005.
- [11] SPSS; Manual de Usuario de Clementine 11.1; Integral Solutions Limited; 2007
- [12] Weber, Robert L; Física General Moderna; REVERTÉ; Barcelona; Edición 1; 1953
- [13] WISLEY C. O; Hydology; Jonh Willey & Sons Inc; Edición 2, 1959

## ENLACES WEB

- [14] Artículo BodegasDeDatos.html; <http://quimbaya.utp.edu.co:9673/kernel/proyectos-de-grado/bodegas-de-datos/articulo-bodegasdedatos.html>; 2006
- [15] AstroMía, Elementos del Clima; <http://www.astromia.com/tierraluna/elemclima.htm>; 2006.
- [16] Comandos SQL; <http://sql.1keydata.com>; 2000
- [17] Diseño conceptual de bases de Datos; <http://www.cs.us.es/cursos/bd/tema-BD-2.pdf>; 2005
- [18] El Modelo Entidad-Relación; <http://www3.uji.es/~mmarques/f47/apun/node83.html>; 2001.
- [19] Elementos del Clima; <http://www.astromia.com/tierraluna/elemclima.htm>; 2006.
- [20] <http://www.itba.edu.ar/capis/rtis/rtis-6-2/metodologia-de-transformacion-de-datos.pdf>
- [21] Implementando un DataWarehouse; <http://www.inf.udec.cl/revista/edicion5/cwolff.htm>; 2002
- [22] Normalizacion de Bases de Datos; <http://www.monografias.com/trabajos5/norbad/norbad2.shtml>; 1997.
- [23] Observatorio Astronómico de Quito; Historia del Observatorio; <http://www.epn.edu.ec/OAQ/index.htm>; 2005.
- [24] Organización Meteorológica Intenacional, Intercambio De Datos Y Productos Hidrológicos, <http://www.wmo.ch/web-sp/wmofact.html>; 2006.
- [25] Pluviógrafo; <http://www.ineter.gob.ni/Direcciones/meteorologia/Red%20Meteorologica/pluviografo.htm>; 2002.
- [26] Sql Louder; <http://eisc.univalle.edu.co/~mauferna/IntrodOracleBulkLoader.pdf>; 2005

# **ANEXOS**

# ANEXO 1: REPORTES METEOROLÓGICO

ENERO	DEPARTAMENTO DE METEOROLOGIA		DEPARTAMENTO DE MATEMATICAS		TEMPERATURA DEL AIRE		VALORES HORARIOS VERDADEROS		LONGITUD 78°29'56"W		LATITUD 0°12'57"S		ALTURA 2818 m																
	1998				en grados centígrados a la sombra y a 2 metros sobre el suelo																								
			MADRUGADA		MAÑANA		TARDE		NOCHE		SUMA		MEDIA		MAX.		MIN.		OSC.										
FECHA	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	SUMA	MEDIA	MAX.	MIN.	OSC.
1	13.1	12.9	12.8	12.0	11.9	12.0	12.8	14.9	18.0	19.7	21.1	22.0	22.6	22.0	19.7	18.8	17.5	16.4	15.7	14.9	14.0	13.4	12.3	12.1	382.6	15.9	22.6	11.9	10.1
2	12.1	12.0	12.1	12.0	11.5	11.4	12.2	15.0	17.0	20.1	21.6	22.1	23.2	23.4	23.9	22.6	20.1	17.9	15.9	15.1	14.5	14.1	13.9	13.5	397.2	16.6	23.9	11.4	12.1
3	13.1	12.5	12.2	12.1	12.1	12.2	12.9	15.2	17.9	20.3	22.5	22.3	23.4	23.4	23.7	21.1	20.5	18.3	16.6	15.1	14.7	14.0	14.1	14.0	404.2	16.8	23.7	12.1	11.1
4	13.5	13.3	13.2	13.7	13.6	13.3	13.4	16.2	18.6	20.4	21.1	21.5	22.0	21.1	20.8	21.2	19.6	17.8	17.0	16.1	15.5	14.7	14.2	14.2	407.1	17.0	22.0	13.2	8.1
5	14.0	14.0	13.7	13.7	13.4	13.3	13.6	14.8	16.8	18.9	17.6	20.0	19.6	20.7	20.5	20.1	19.5	18.1	16.2	15.9	15.4	14.5	13.9	13.2	391.4	16.3	20.7	13.2	7.1
6	12.6	12.5	12.8	12.5	12.5	16.6	12.6	13.1	15.9	18.4	18.2	18.7	21.2	21.8	19.5	18.2	16.9	16.6	14.8	14.1	14.1	14.8	13.5	13.2	375.1	15.6	21.8	12.5	9.1
7	13.0	12.9	12.9	12.8	12.7	12.5	12.4	12.5	14.0	16.0	17.5	18.8	18.2	14.0	14.4	13.8	13.5	13.2	13.0	12.9	12.6	12.5	12.4	12.1	330.6	13.8	18.8	12.1	6.1
8	12.1	12.2	11.9	11.7	11.4	11.2	11.2	12.4	14.4	17.5	19.2	20.7	19.2	16.0	16.7	16.3	14.3	13.3	13.0	12.9	12.0	11.7	11.5	10.9	333.7	13.9	20.7	10.9	9.1
9	10.2	10.1	10.0	9.9	9.4	9.1	9.2	10.8	14.3	18.1	21.0	21.9	22.6	23.7	21.7	21.0	17.7	16.0	15.4	14.9	14.9	14.7	13.6	13.6	363.8	15.2	23.7	9.1	14.1
10	13.1	13.4	13.4	12.5	11.0	10.5	11.0	12.0	14.2	17.9	20.6	21.9	20.2	20.0	17.5	14.1	13.9	13.5	12.4	12.1	11.7	11.7	11.4	11.4	341.4	14.2	21.9	10.5	11.1
11	11.0	10.3	9.9	9.8	9.7	9.9	9.4	12.0	14.8	17.2	19.3	21.0	21.6	21.6	21.6	22.6	20.2	18.7	16.4	11.8	11.0	14.9	14.1	11.3	360.1	15.0	22.6	9.4	13.1
12	13.6	13.5	13.6	13.9	14.0	13.7	12.2	13.7	16.2	17.6	19.4	18.7	17.8	17.2	18.0	18.5	17.4	16.9	16.2	15.3	14.9	14.4	13.2	13.7	373.6	15.6	19.4	12.2	7.1
13	12.2	13.0	12.9	12.8	12.6	12.1	12.4	13.1	15.1	18.0	19.5	20.1	20.4	20.8	21.0	19.5	18.8	17.7	16.6	15.1	14.3	14.0	13.6	12.2	377.8	15.7	21.0	12.1	8.1
14	13.1	13.2	13.1	11.8	11.1	11.0	11.2	14.0	15.9	18.6	20.0	21.1	21.6	22.1	20.7	19.1	17.1	15.9	16.0	14.1	13.8	13.7	14.0	13.7	375.9	15.7	22.1	11.0	11.1
15	12.5	12.1	11.9	12.0	12.1	11.8	12.2	14.1	15.9	18.1	19.5	21.4	22.4	21.0	19.0	16.1	16.0	15.6	15.0	14.0	13.9	14.2	13.9	13.4	368.1	15.3	22.4	11.8	10.1
16	12.3	11.7	10.6	10.6	10.6	10.7	10.8	14.0	16.5	17.6	20.6	21.1	22.4	22.0	20.5	17.0	16.6	15.4	15.0	14.2	13.2	13.2	12.8	12.8	362.2	15.1	22.4	10.6	11.1
17	12.8	12.8	12.8	12.8	12.8	12.7	12.8	13.1	14.9	17.3	18.6	19.0	19.6	20.0	21.4	19.2	19.1	17.8	16.5	15.0	14.6	14.2	13.9	13.1	376.8	15.7	21.4	12.7	8.1
18	13.1	13.0	11.9	11.4	11.1	11.1	11.2	13.1	15.1	17.1	19.1	21.0	21.6	21.0	20.9	21.0	16.6	16.6	16.4	14.9	14.2	13.9	13.2	13.0	371.5	15.5	21.6	11.1	10.1
19	12.5	12.6	11.9	11.5	11.4	11.5	11.6	13.0	15.0	17.1	20.2	20.9	22.2	20.3	19.8	16.0	14.9	14.7	14.2	13.7	13.2	12.8	12.2	11.5	354.7	14.8	22.2	11.4	10.1
20	10.5	10.9	10.6	11.0	10.2	10.2	10.2	12.0	15.1	17.9	21.0	21.4	21.4	21.9	20.1	18.4	17.0	16.9	15.6	14.6	14.1	13.7	13.3	13.1	361.1	15.0	21.9	10.2	11.1
21	13.1	13.1	13.0	13.0	12.7	12.2	12.2	13.4	14.9	16.1	18.3	18.4	18.8	18.8	21.0	19.5	19.0	18.6	17.6	15.2	14.8	14.5	14.1	13.9	377.4	15.7	21.0	12.2	8.1
22	13.5	13.0	12.9	13.0	13.0	12.9	13.0	14.0	15.6	18.1	19.8	20.8	21.2	21.2	19.9	18.9	18.0	17.2	16.6	15.4	15.3	15.0	15.0	14.9	388.2	16.2	21.2	12.9	8.1
23	13.7	12.8	12.5	12.0	12.0	11.3	14.4	13.0	15.5	17.9	20.6	22.0	21.6	20.8	20.9	20.9	19.5	18.7	17.4	16.8	15.2	14.2	13.2	13.0	389.9	16.2	22.0	11.3	10.1
24	13.0	12.1	12.0	11.2	11.0	11.8	12.0	14.0	16.0	18.9	20.9	22.1	22.6	23.1	23.1	22.0	20.3	19.0	17.4	16.0	15.1	15.0	14.0	13.4	396.0	16.5	23.1	11.0	12.1
25	12.9	12.9	13.0	13.0	12.5	12.5	12.0	13.0	14.8	17.0	20.0	21.0	22.8	24.0	22.0	22.8	19.8	18.5	16.2	15.2	14.8	14.0	13.8	13.5	382.0	16.3	24.0	12.0	12.1
26	13.0	12.2	11.8	11.0	10.8	10.2	10.0	12.2	15.5	18.0	20.0	21.0	21.6	20.4	21.1	21.2	19.2	17.5	17.4	15.8	15.0	14.5	14.6	13.9	377.9	15.7	21.6	10.0	11.1
27	13.9	13.1	12.2	12.0	11.2	11.2	12.1	12.7	15.2	17.0	20.0	21.4	20.2	20.0	20.2	19.4	20.4	18.2	16.6	15.7	15.0	14.2	14.0	13.7	379.6	15.8	21.4	11.2	10.1
28	13.2	13.2	13.2	13.2	13.2	12.7	12.4	13.0	14.9	16.6	17.2	18.0	19.2	18.7	19.7	21.0	20.7	19.2	17.2	16.4	15.2	14.7	14.2	14.0	381.0	15.9	21.0	12.4	8.1
29	13.6	13.9	12.0	11.6	11.0	11.0	12.4	13.0	14.9	16.6	17.2	18.0	19.2	18.7	19.7	21.0	20.7	19.2	17.3	16.4	15.2	14.7	14.2	14.0	375.5	15.6	21.0	11.0	10.1
30	13.4	12.8	12.4	12.9	12.4	11.7	10.8	12.2	14.1	16.1	19.2	20.8	22.6	21.8	21.6	18.1	18.7	18.7	16.6	15.6	14.2	13.7	13.3	13.4	377.1	15.7	22.6	10.8	11.1
31	13.2	13.1	12.9	12.1	11.1	10.4	9.8	12.0	15.0	17.7	20.0	21.5	21.4	22.2	21.3	21.4	20.4	19.4	17.2	16.2	15.2	14.3	13.3	13.1	384.2	16.0	22.2	9.8	12.1

**CANTIDAD DE LLUVIA**

Cantidad de Lluvia promedio año 2000

<b>Mes</b>	<b>Cantidad total de Lluvia (mm)</b>	<b>Días con Lluvia</b>
Enero	140,3	25
Febrero	157,7	24
Marzo	163,6	26
Abril	211,5	29
Mayo	151,4	28
Junio	73,1	18
Julio	63	9
Agosto	15,3	6
Septiembre	94,8	18
Octubre	97,2	12
Noviembre	58,2	11
Diciembre	75,4	12
Total anual	1301,5	218
Promedio anual	108,5	

Cantidad de Lluvia promedio año 2001

<b>Mes</b>	<b>Cantidad total de Lluvia (mm)</b>	<b>Días con Lluvia</b>
Enero	60	13
Febrero	100	15
Marzo	119	22
Abril	78	14
Mayo	68	14
Junio	13	7
Julio	28	12
Agosto	0	0
Septiembre	63	17
Octubre	29	8
Noviembre	113	14
Diciembre	156	25
Total anual	827	161
Promedio anual	68,9	

## Cantidad de Lluvia promedio año 2002

<b>Mes</b>	<b>Cantidad total de Lluvia (mm)</b>	<b>Días con Lluvia</b>
Enero	65,5	13
Febrero	50,8	11
Marzo	153,5	22
Abril	238	26
Mayo	89,8	12
Junio	39,1	11
Julio	11,8	10
Agosto	18,8	4
Septiembre	51,2	12
Octubre	179,3	18
Noviembre	108,9	25
Diciembre	165	19
Total anual	1171,7	
Promedio anual	97,64	

## Cantidad de Lluvia promedio año 2003

<b>Mes</b>	<b>Cantidad total de Lluvia (mm)</b>	<b>Días con Lluvia</b>
Enero	19,6	7
Febrero	81,5	13
Marzo	124,5	14
Abril	280,4	22
Mayo	19,9	10
Junio	43	15
Julio	22,5	6
Agosto	11,3	4
Septiembre	75,8	15
Octubre	139,1	21
Noviembre	130,5	24
Diciembre	92,9	14
Total anual	1041	165
Promedio anual	86,75	

## Cantidad de Lluvia promedio año 2004

<b>Mes</b>	<b>Cantidad total de Lluvia (mm)</b>	<b>Días con Lluvia</b>
Enero	19	8
Febrero	29,3	9
Marzo	77,7	14
Abril	172,5	20
Mayo	95,4	23
Junio	29,7	7
Julio	11,5	9
Agosto	24,4	4
Septiembre	120,3	19
Octubre	83,4	24
Noviembre	143,2	21

Diciembre	99,2	17
Total anual	905,6	175
Promedio anual	75,5	

Cantidad de Lluvia promedio año 2005

Mes	Cantidad total de Lluvia (mm)	Días con Lluvia
Enero	26,8	10
Febrero	138,4	12
Marzo	127,5	22
Abril	93,6	20
Mayo	39,4	11
Junio	28	13
Julio	23,3	5
Agosto	41,2	9
Septiembre	52,5	26
Octubre	83,5	15
Noviembre	75,6	18
Diciembre	129,9	22
Total anual	859,7	183
Promedio anual	71,6	

Cantidad de Lluvia promedio año 2006

Mes	Cantidad total de Lluvia (mm)	Días con Lluvia
Enero	30,7	10
Febrero	101,5	18
Marzo	250,4	21
Abril	254,2	23
Mayo	123,7	16
Junio	64,8	12
Julio	3,6	2
Agosto	3	3
Septiembre	19,6	13
Octubre	129,9	16
Noviembre	172,3	23
Diciembre	226,8	23
Total anual	1380,5	180
Promedio anual	115	

## TEMPERATURAS MEDIAS MENSUALS

Temperaturas promedio año 2000

Mes	Temperatura (°K) Promedio por mes	Máxima (°K)	Minima (°K)
Enero	286	296,2	278,8
Febrero	285,9	295,4	279,8
Marzo	286,2	294,8	279,9
Abril	286,2	295,4	280,8
Mayo	286,3	294,3	281,7
Junio	287,1	295,7	281
Julio	283,3	284,9	279,9
Agosto	287,3	297,5	280
Septiembre	286,3	296,6	280,2
Octubre	284,1	285,4	281,1
Noviembre	283,4	284,4	280,1
Diciembre	286,8	295,8	278,6
Promedio anual	285,7	293,0	280,2

Temperaturas promedio año 2001

Mes	Temperatura (°K) Promedio por mes	Máxima (°K)	Minima (°K)
Enero	286,2	297,2	279,2
Febrero	287	296	280,2
Marzo	286,7	295,7	275,1
Abril	287,6	295,7	275,4
Mayo	287,4	297,2	280,9
Junio	287,7	296,7	280,6
Julio	287,4	298	280,4
Agosto	288,1	298,5	280
Septiembre	286,8	297,2	279,8
Octubre	288,1	298,9	280,5
Noviembre	287,2	297,9	279,3
Diciembre	287,6	297,4	280,3
Promedio anual	287,3	297,2	279,3

Temperaturas promedio año 2002

Mes	Temperatura (°K) Promedio por mes	Máxima (°K)	Minima (°K)
Enero	287,5	297,4	280,1
Febrero	287,5	298,4	280,8
Marzo	287,4	296	281,8
Abril	286,8	296	280,5
Mayo	288,0	297,2	280,1
Junio	287,7	297,2	280,3
Julio	288,2	297,8	280,3
Agosto	288,1	297,6	279,9
Septiembre	287,9	298,4	280,2
Octubre	287,4	297,7	281,2

Noviembre	286,9	296,7	280,8
Diciembre	287,3	297	281
Promedio anual	287,6	297,3	280,6

## Temperaturas promedio año 2003

Mes	Temperatura (°K) Promedio por mes	Máxima (°K)	Minima (°K)
Enero	288	300	281,8
Febrero	288,3	297,7	280,6
Marzo	287,3	295,5	281
Abril	287,2	296,3	281,4
Mayo	288,1	296,4	282,4
Junio	287	295,7	281
Julio	287,6	299	280,6
Agosto	288,1	297,6	280,6
Septiembre	287,4	296,4	280
Octubre	286,9	296,3	281,4
Noviembre	286,8	296,3	280,4
Diciembre	287	295,8	280,8
Promedio anual	287,5	296,9	281

## Temperaturas promedio año 2004

Mes	Temperatura (°K) Promedio por mes	Máxima (°K)	Minima (°K)
Enero	287,14	296,4	279,8
Febrero	287,17	296,4	279,9
Marzo	287,98	295,7	281,7
Abril	286,97	296,4	281,6
Mayo	287,29	297	280,9
Junio	287,93	295,9	281,2
Julio	287,25	294,9	280,6
Agosto	288,15	297	279,8
Septiembre	287,11	296,2	280,2
Octubre	286,69	295,9	280,6
Noviembre	286,85	295,4	274,5
Diciembre	286,88	294,9	279,8
Promedio anual	287,3	296,0	280,1

## Temperaturas promedio año 2005

Mes	Temperatura (°K) Promedio por mes	Máxima (°K)	Minima (°K)
Enero	287,2	296,2	280,4
Febrero	287,8	297,0	282,2
Marzo	287,1	294,8	282,4
Abril	287,8	295,8	282,2
Mayo	287,6	295,8	280,0
Junio	287,7	295,7	282,1
Julio	288,0	297,2	280,6
Agosto	288,1	298,4	280,2

Septiembre	288,0	297,6	281,8
Octubre	286,5	296,0	280,0
Noviembre	286,5	296,0	277,8
Diciembre	285,9	296,0	279,4
Promedio anual	287,3	296,4	280,8

Temperaturas promedio año 2006

Mes	Temperatura (°K) Promedio por mes	Máxima (°K)	Minima (°K)
Enero	287,3	296,4	282,2
Febrero	287,3	295,6	282,0
Marzo	286,8	295,6	281,6
Abril	286,8	295,2	281,4
Mayo	287,7	295,6	282,0
Junio	287,3	295,3	281,0
Julio	287,8	296,4	280,4
Agosto	288,0	297,0	281,2
Septiembre	287,8	297,4	281,0
Octubre	287,4	297,4	279,6
Noviembre	286,6	295,8	281,8
Diciembre	287,0	296,7	281,7
Promedio anual	287,3	296,2	281,3

## ANEXO 2: DETALLE DE LOS PROBLEMAS

Año	Elemento	Fecha y Hora	Problema	Solución
1995	Humedad	20H 30/06 hasta las 6H del 30/06	datos no fueron ingresados	Se revisaron las fajas del levantamiento de los datos
1996	Humedad	18/05 a las 8	dato ingresado erróneamente	Se revisaron las fajas del levantamiento de los datos
1997	Humedad	el 31/12 a las 5	el dato ingresado es muy bajo (7) del nivel normal	Corrección mediante aproximación
		el 7/9 a las 12	no se ingresó el valor	Se revisaron los reporte (44%)
	Temperatura	el 23/11 a las 4	el valor era demasiado grande 110	Se revisaron las fajas del levantamiento de los datos dato 9.5 C
1998	Presión Atmosférica	27/05/2006	valor de 5410.12	Se Revisó el boletín meteorológico
		28/05/2006	valor de 541.59	Se Revisó el boletín meteorológico
1999	Temperatura	4/7 a las 21	valor muy bajo (1)	Se revisaron las fajas del levantamiento de los datos 11.2
		28/7 a las 3	valor muy bajo (1.2)	Se revisaron las fajas del levantamiento de los datos 11.4
		1/10 a las 2	valor muy bajo (1.8)	Se revisaron las fajas del levantamiento de los datos 10.6
2000	humedad	9/5/00 a las 23	valor muy bajo (9)	Se revisaron las fajas del levantamiento de los datos 99%
		5/10/00 de 11 a 12	valor muy bajo (0.6)	Se revisaron las fajas del levantamiento de los datos 99%
		5/10/00 de 11 a 13	valores muy bajos (0.8)	Se revisaron las fajas del levantamiento de los datos 38%
	Temperatura	5/10/00 de 13	valores muy alto (31.8)	
2001	humedad	13/2/01 a las 6	valor muy alto (725)	Se revisaron las fajas del levantamiento de los datos 68%
2002	Temperatura	26/02/2002 a las 4	valor muy alto (100.8)	10,08
		02/06/2002 a las 23	valor muy alto (112)	11,2
		17/06/2002 a las 11	no hay valores	IGUAL
		17/06/2002 a las 12	no hay valores	IGUAL
		01/08/2002 a las 16	no hay valores	IGUAL
		05/09/2002 a las 16	valor muy alto (214)	21,4

		11/09/2002 a las 14	valor de 0.2	20.7 Corrección mediante aproximación
		04/10/2002 a las 7	valor de 4.2 pero valores cercanos + de 10	
		21/10/2002 a las 3	valor muy alto (113)	11,3
	<b>Humedad</b>	julio y Agosto	Datos iguales	
		6/6 a la 7	hay una coma	
<b>2003</b>	<b>Temperatura</b>	31/8 a las 24	no hay valor	Se revisaron las fajas del levantamiento de los datos 11.6
	<b>Humedad</b>	desde la 8 del 15/9 hasta las 6 del 10/11 el instrumento no estuvo en funcionamiento	no hay valores	se trabajará con valores nulos
		08/3 a las 16	valor muy alto (378)	Se revisaron las fajas del levantamiento de los datos 30%
		15/1 a las 10	valor muy alto (150)	Se revisaron las fajas del levantamiento de los datos 99%
	<b>Lluvia</b>	mes de noviembre	hay datos en 31 de noviembre pero este día no existe	datos fueron eliminados tras de revisar y comprobar que fue un error de observación de fajas de levantamiento de datos
<b>2004</b>	<b>Humedad</b>	desde las 14 del 21/4 hasta las 24 del 27/4 el instrumento de medición estuvo en reparación	no hay valores	se trabajará con valores nulos
		15/02/2004 a las 6	valor muy bajo (7)	Corrección mediante aproximación
		21/02/2004 a las 18	valor muy alto (662)	Corrección mediante aproximación
		06/03/2004 a las 3	valor muy bajo (9)	Corrección mediante aproximación
		13/03/2004 a las 5	valor muy bajo (9)	Corrección mediante aproximación
		14/03/2004 a las 18	valor muy alto (363)	Corrección mediante aproximación
		22/03/2004 a las 12	valor muy alto (363)	Corrección mediante aproximación
		29/04/2004 a las 18	valor muy bajo (7)	Corrección mediante aproximación

## ANEXO 3: CREACIONES DE ESQUEMAS

### ESQUEMA TRANSACCIONAL

Creación de tablas para la base de datos transaccional

**ELEMENTO\_METEOROLOGICO:** Tabla que almacena los datos de los elementos meteorológicos con sus respectivas unidades de medida.

```
CREATE TABLE "TRANSACCIONAL"."ELEMENTO_METEOROLOGICO"
  ("ELE_CODIGO" VARCHAR2(10) NOT NULL, "ELE_NOMBRE"
  VARCHAR2(40) NOT NULL, "ELE_UNIDAD_MEDIDA" VARCHAR2(40)
  NOT
  NULL,
  CONSTRAINT "PK_ELE_MET" PRIMARY KEY("ELE_CODIGO"))
  TABLESPACE "OAQ"
```

**MUESTRA\_DIARIA:** Tabla que almacena la fecha en que los datos fueron tomados.

```
CREATE TABLE "TRANSACCIONAL"."MUESTRA_DIARIA"
  ("MUE_CODIGO"
  VARCHAR2(10) NOT NULL, "MUE_FECHA" DATE NOT NULL,
  CONSTRAINT "PK_MUE_DIA" PRIMARY KEY("MUE_CODIGO"))
  TABLESPACE "OAQ"
```

**OBSERVACION:** Tabla que almacena los registros de medición con su respectiva hora.

```
CREATE TABLE "TRANSACCIONAL"."OBSERVACION" ("ELE_CODIGO"
  VARCHAR2(10) NOT NULL, "MUE_CODIGO" VARCHAR2(10) NOT
  NULL,
```

```
"OBS_HORA" VARCHAR2(10), "OBS_VALOR" FLOAT(10),
CONSTRAINT "FK_ELE_MET" FOREIGN KEY("ELE_CODIGO")
REFERENCES
```

```
"TRANSACCIONAL"."ELEMENTO_METEOROLOGICO"("ELE_CODIGO"),
CONSTRAINT "FK_MUE_DIA" FOREIGN KEY("MUE_CODIGO")
REFERENCES
```

```
"TRANSACCIONAL"."MUESTRA_DIARIA"("MUE_CODIGO"))
TABLESPACE "OAQ"
```

### ESQUEMA TEMPORAL

Fue necesario la creación esquema temporal ya que no se contaba con el con un atributo necesario en la tabla Observación, gracias a la base de datos temporal se pudo suplir el problema de no tener una clave primaria para Observación.

```
T_ELEMENTO_METEOROLOGICO
CREATE TABLE "TEMPORAL"."T_ELEMENTO_METEOROLOGICO"
("T_ELE_MET_CODIGO" VARCHAR2(10) NOT NULL,
"T_ELE_MET_NOMBRE"
VARCHAR2(40) NOT NULL, "T_ELE_MET_UNIDAD_MEDIDA"
VARCHAR2(40)
NOT NULL,
CONSTRAINT "PK_T_ELE_MET" PRIMARY
KEY("T_ELE_MET_CODIGO"))
TABLESPACE "TOAQ"
```

T\_MUESTRA\_DIARIA

```
CREATE TABLE "TEMPORAL"."T_MUESTRA_DIARIA"
("T_MUE_CODIGO"
  VARCHAR2(10) NOT NULL, "T_MUE_FECHA" DATE NOT NULL,
  CONSTRAINT "PK_T_MUE" PRIMARY KEY("T_MUE_CODIGO"))
  TABLESPACE "TOAQ"
```

T\_OBSERVACION

```
CREATE TABLE "TEMPORAL"."T_OBSERVACION" ("T_OBS_CODIGO"
  NUMBER(10) NOT NULL, "T_ELE_MET_CODIGO" VARCHAR2(10)
  NOT NULL,
  "T_MUE_CODIGO" VARCHAR2(10) NOT NULL, "T_OBS_HORA"
  NUMBER(10)
  NOT NULL, "T_OBS_VALOR" FLOAT(10) NOT NULL,
  CONSTRAINT "PK_T_OBS" PRIMARY KEY("T_OBS_CODIGO"),
  CONSTRAINT "FK_T_ELE_MET" FOREIGN
  KEY("T_ELE_MET_CODIGO")
  REFERENCES
```

```
"TEMPORAL"."T_ELEMENTO_METEOROLOGICO"("T_ELE_MET_CODIG
  O"),
```

```
  CONSTRAINT "FK_T_MUE" FOREIGN KEY("T_MUE_CODIGO")
  REFERENCES
```

```
"TEMPORAL"."T_MUESTRA_DIARIA"("T_MUE_CODIGO"))
  TABLESPACE "TOAQ"
```

## ANEXO 4: CARGA DE DATOS BASE TRANSACCIONAL

Carga de datos a la tabla ELEMENTO\_METEOROLOGICO

Como se vio en la sintaxis del sqlldr, es requerido el archivo plano, el cual contiene la información de los elementos meteorológicos a ser importados, con la estructura detallada anteriormente el archivo .ctl, quedo de la siguiente manera:

elemento\_meteorologico.ctl

```
LOAD DATA
INFILE *
INTO TABLE ELEMENTO_METEOROLOGICO
APPEND
FIELDS TERMINATED BY ';' OPTIONALLY ENCLOSED ""
TRAILING NULLCOLS
(ELE_CODIGO, ELE_NOMBRE, ELE_UNIDAD_MEDIDA)
BEGINDATA
H;Humedad Relativa;Tanto por Ciento
L;Cantidad de Lluvia;Milímetros cúbicos
P;Presión Atmosférica Media;Milímetros de Mercurio
T;Temperatura del Aire;Grados Centígrados
```

Una vez guardado, se apertura una interfase de línea de comandos y se ejecuta el comando sqlldr para cargar los datos a la tabla ELEMENTO\_METEOROLOGICO,

```
D:\>sqlldr transaccional/vaio@exp elemento_meteorologico.ctl rows=4 errors=4
```

```
SQL*Loader: Release 9.2.0.1.0 - Production on Dom Mar 16 11:34:32 2008
```

```
Copyright (c) 1982, 2002, Oracle Corporation. All rights reserved.
```

```
Commit point reached - logical record count 3
```

```
Commit point reached - logical record count 4
```

En nuestro caso el archivo generado solo fue el .log el cual tiene la siguiente información:

## elemento\_meteorologico.log

SQL\*Loader: Release 9.2.0.1.0 - Production on Dom Mar 16 13:26:48 2008

Copyright (c) 1982, 2002, Oracle Corporation. All rights reserved.

Control File: elemento\_meteorologico.ctl

Data File: elemento\_meteorologico.ctl

Bad File: elemento\_meteorologico.bad

Discard File: none specified

(Allow all discards)

Number to load: ALL

Number to skip: 0

Errors allowed: 4

Bind array: 4 rows, maximum of 256000 bytes

Continuation: none specified

Path used: Conventional

Table ELEMENTO\_METEOROLOGICO, loaded from every logical record.

Insert option in effect for this table: APPEND

TRAILING NULLCOLS option in effect

Column Name	Position	Len	Term	Encl	Datatype
ELE_CODIGO	FIRST	*	;	O(")	CHARACTER
ELE_NOMBRE	NEXT	*	;	O(")	CHARACTER
ELE_UNIDAD_MEDIDA	NEXT	*	;	O(")	CHARACTER

Table ELEMENTO\_METEOROLOGICO:

4 Rows successfully loaded.

0 Rows not loaded due to data errors.

0 Rows not loaded because all WHEN clauses were failed.

0 Rows not loaded because all fields were null.

Space allocated for bind array: 3096 bytes(4 rows)

Read buffer bytes: 1048576

*Total logical records skipped: 0*  
*Total logical records read: 4*  
*Total logical records rejected: 0*  
*Total logical records discarded: 0*  
*Run began on Dom Mar 16 13:26:48 2008*  
*Run ended on Dom Mar 16 13:26:49 2008*

*Elapsed time was: 00:00:00.19*

Comprobación: para poder realizar la comprobación realizamos la consulta a la base de datos elemento\_meteorológico a través del SQL PLUS.

*select \* from elemento\_meteorologico;*

<i>ELE_CODIGO</i>	<i>ELE_NOMBRE</i>	<i>ELE_UNIDAD_MEDIDA</i>
<i>H</i>	<i>Humedad Relativa</i>	<i>Tanto por Ciento</i>
<i>L</i>	<i>Cantidad de Lluvia</i>	<i>Milímetros cúbicos</i>
<i>P</i>	<i>Presión Atmosférica Media</i>	<i>Milímetros de Mercurio</i>
<i>T</i>	<i>Temperatura del Aire</i>	<i>Grados Centígrados</i>

Para las demás tablas se procederá de la misma forma.

Carga de datos a la tabla MUESTRA\_DIARIA

Muestra\_diaria.ctl

```

LOAD DATA
INFILE *
INTO TABLE MUESTRA_DIARIA
APPEND
FIELDS TERMINATED BY ';' OPTIONALLY ENCLOSED ""
TRAILING NULLCOLS
(MUE_CODIGO, MUE_FECHA date(10) 'MM/DD/YYYY')
BEGINDATA
1;1/1/1995
2;1/2/1995
3;1/3/1995
4;1/4/1995
.....

```

.....  
 .....  
 4013;12/26/2005  
 4014;12/27/2005  
 4015;12/28/2005  
 4016;12/29/2005  
 4017;12/30/2005  
 4018;12/31/2005

### Ejecución de sqlldr:

*D:\>sqlldr transaccional/vaio@exp muestra\_diaria.ctl rows=1000 errors=4018*

*SQL\*Loader: Release 9.2.0.1.0 - Production on Dom Mar 16 11:34:32 2008*

*Copyright (c) 1982, 2002, Oracle Corporation. All rights reserved.*

*Commit point reached - logical record count 3*

*Commit point reached - logical record count 4*

*...*

*..*

*Commit point reached - logical record count 4000*

*Commit point reached - logical record count 4004*

*Commit point reached - logical record count 4008*

*Commit point reached - logical record count 4012*

*Commit point reached - logical record count 4016*

*Commit point reached - logical record count 4017*

*Commit point reached - logical record count 4018*

### Archivo Muestra\_diaria.log generado:

*SQL\*Loader: Release 9.2.0.1.0 - Production on Dom Mar 16 13:42:40 2008*

*Copyright (c) 1982, 2002, Oracle Corporation. All rights reserved.*

*Control File: muestra\_diaria.ctl*

*Data File: muestra\_diaria.ctl*

*Bad File: muestra\_diaria.bad*

*Discard File: none specified*

*(Allow all discards)*

*Number to load: ALL*

*Number to skip: 0*

*Errors allowed: 4018*

*Bind array: 4 rows, maximum of 256000 bytes*

*Continuation: none specified*

*Path used: Conventional*

*Table MUESTRA\_DIARIA, loaded from every logical record.*

*Insert option in effect for this table: APPEND*

*TRAILING NULLCOLS option in effect*

<i>Column Name</i>	<i>Position</i>	<i>Len</i>	<i>Term</i>	<i>Encl</i>	<i>Datatype</i>
<i>MUE_CODIGO</i>	<i>FIRST</i>	<i>*</i>	<i>;</i>	<i>O(")</i>	<i>CHARACTER</i>
<i>MUE_FECHA</i>	<i>NEXT</i>	<i>10</i>	<i>;</i>	<i>O(")</i>	<i>DATE MM/DD/YYYY</i>

*Table MUESTRA\_DIARIA:*

*4018 Rows successfully loaded.*

*0 Rows not loaded due to data errors.*

*0 Rows not loaded because all WHEN clauses were failed.*

*0 Rows not loaded because all fields were null.*

*Space allocated for bind array: 1080 bytes(4 rows)*

*Read buffer bytes: 1048576*

*Total logical records skipped: 0*

*Total logical records read: 4018*

*Total logical records rejected: 0*

*Total logical records discarded: 0*

*Run began on Dom Mar 16 13:42:40 2008*

*Run ended on Dom Mar 16 13:42:42 2008*

*Elapsed time was: 00:00:02.78*

*CPU time was: 00:00:00.30*

Comprobación: para poder realizar la comprobación realizamos la consulta a la base de datos muestra\_diaria a través del SQL PLUS.

```
SQL> select * from muestra_diaria
      2* order by mue_fecha;
MUE_CODIGO      MUE_FECHA
-----
1                01-ENE-95
2                02-ENE-95
3                03-ENE-95
4                04-ENE-95
5                05-ENE-95
.
.
.
4014            27-DIC-05
4015            28-DIC-05
4016            29-DIC-05
4017            30-DIC-05
4018            31-DIC-05
```

4018 rows selected.

### Carga de Datos OBSERVACION

Para la carga a la tabla OBSERVACIÓN se ha separado los distintos elementos y de esta forma controlar su ingreso.

```
D:\>sqlldr transaccional/vaio@exp ingresohumedad.ctl rows=1000 errors=96432
D:\>sqlldr transaccional/vaio@exp ingresolluvia.ctl rows=1000 errors=96432
D:\>sqlldr transaccional/vaio@exp ingresotemperatura.ctl rows=1000 errors=96432
D:\>sqlldr transaccional/vaio@exp ingresopresion.ctl rows=1000 errors=96432
```

### Comprobación

```
SQL> select * from observacion
      2 ;
ELE_CODIGO      MUE_CODIGO      OBS_HORA      OBS_VALOR
-----
H                20              12             47
```

H	20	13	48
H	20	14	51
H	20	15	50
.	....	...	.....
.	....	...	.....
P	3992	15	548.2
P	3992	16	548.2
P	3992	17	548.2
P	3992	18	548.2
P	3992	19	548.2
P	3992	20	548.2
P	3992	21	548.2
P	3992	22	548.2
P	3992	23	548.2
P	3992	24	548.2
P	3993	1	547.2
P	3993	2	547.2
P	3993	3	547.2

385728 rows selected.

**ANEXO 5: REPORTES DATA WAREHOUSE.****Cantidad lluvia caída por mes**

<b>AÑO</b>		1995
<b>MES</b>	<b>DIAS LLUVIOSOS</b>	<b>CANTIDAD (mm)</b>
Enero	9	18.5
Febrero	9	88
Marzo	13	162.3
Abril	18	191.2
Mayo	20	143.3
Junio	15	66.7
Julio	15	54.1
Agosto	16	88.6
Septiembre	6	16.9
Octubre	20	122.1
Noviembre	24	239.3
Diciembre	17	74.2
<b>AÑO</b>		1996
Enero	19	155.4
Febrero	18	103
Marzo	23	173.1
Abril	24	150.8
Mayo	27	179.2
Junio	14	103.5
Julio	13	29.5
Agosto	9	52.2
Septiembre	14	60.9
Octubre	24	121.1
Noviembre	7	28.4
Diciembre	13	66.6
<b>AÑO</b>		1997
Enero	20	162.1
Febrero	13	35
Marzo	21	185.6
Abril	20	91.4
Mayo	17	78.2
Junio	11	92.7
Julio	2	0.8
Septiembre	17	67.1
Octubre	20	123.2
Noviembre	29	313.1
Diciembre	14	93.2

<b>AÑO</b> 1998		
<b>MES</b>	<b>DIAS LLUVIOSOS</b>	<b>CANTIDAD (mm)</b>
Enero	13	64.9
Febrero	18	142.3
Marzo	21	174.5
Abril	26	168.5
Mayo	20	186.8
Junio	14	36.4
Julio	15	34.2
Agosto	9	40.7
Septiembre	19	85.1
Octubre	20	114.4
Noviembre	21	130.3
Diciembre	20	37.2
<b>AÑO</b> 1999		
Enero	22	87.9
Febrero	24	192.3
Marzo	21	160.7
Abril	22	191.2
Mayo	18	77.3
Junio	22	123.8
Julio	12	19.4
Agosto	15	63.1
Septiembre	22	134.2
Octubre	22	157.9
Noviembre	15	85.1
Diciembre	26	227.2
<b>AÑO</b> 2000		
Enero	25	140.3
Febrero	24	157.7
Marzo	26	162.7
Abril	29	211.5
Mayo	28	151.4
Junio	18	73.1
Julio	9	63
Agosto	6	15.3
Septiembre	18	94.8
Octubre	12	97.2
Noviembre	11	58.2
Diciembre	12	75.4
<b>AÑO</b> 2001		
Enero	13	60.2
Febrero	15	99.9
Marzo	22	119
Abril	14	78.2

Mayo	14	68.1
Junio	7	12.6
Julio	12	27.6
Septiembre	17	88.4
Octubre	8	29.2
Noviembre	14	113.2
Diciembre	24	130.2
<b>AÑO</b>	<b>2002</b>	
<b>MES</b>	<b>DIAS LLUVIOSOS</b>	<b>CANTIDAD (mm)</b>
Enero	13	65.5
Febrero	11	50.8
Marzo	22	153.5
Abril	24	238
Mayo	12	89.8
Junio	11	39.1
Julio	10	11.8
Agosto	4	18.8
Septiembre	12	51.2
Octubre	18	179.3
Noviembre	25	108.9
Diciembre	19	165
<b>AÑO</b>	<b>2003</b>	
Enero	7	21
Febrero	13	81.5
Marzo	14	124.5
Abril	22	280.4
Mayo	10	19.9
Junio	14	43
Julio	6	22.5
Agosto	5	354
Septiembre	15	75.8
Octubre	22	139.1
Noviembre	23	130.5
Diciembre	14	92.9
<b>AÑO</b>	<b>2004</b>	
Enero	8	19
Febrero	9	29.3
Marzo	14	77.7
Abril	20	172.5
Mayo	23	95.4
Junio	7	29.7
Julio	9	11.5
Agosto	4	24.4
Septiembre	19	120.3
Octubre	24	83.4
Noviembre	21	143.2
Diciembre	17	99.2

<b>AÑO</b>	<b>2005</b>	
Enero	10	26.8
Febrero	12	138.4
Marzo	22	127.5
Abril	20	93.6
Mayo	11	39.4
Junio	13	28
Julio	5	23.3
Agosto	9	41.2
Septiembre	26	52.5
Octubre	15	83.5
Noviembre	18	75.6
Diciembre	22	129.9

## **ANEXO 6: GUIA DE USO DE CLEMENTINE 11.1**

### Antecedentes

Clementine proporciona una interfaz visual de los datos, las estadísticas y los algoritmos complejos. Cada proceso se representa con un icono, o nodo, que se conecta para formar una ruta que representa el flujo de datos a través de una serie de procesos.

Como aplicación de minería de datos, Clementine ofrece un método para encontrar relaciones útiles entre grandes conjuntos de datos. Al contrario de los métodos estadísticos, no es necesario saber lo que se está buscando al comenzar. Puede explorar los datos, mediante el ajuste de diferentes modelos y la investigación de diferentes relaciones, hasta que encuentre la información que resulte útil.

Una amplia variedad de organizaciones utilizan Clementine para ayudar a analizar grandes repositorios de datos. A continuación se ofrece una muestra de los tipos de problemas que la minería de datos puede ayudar a resolver.

**Sector público:** Las instituciones gubernamentales de todo el mundo utilizan la minería de datos para explorar enormes almacenes de datos, mejorar las relaciones con los ciudadanos, detectar fraudes, como el lavado de dinero o la evasión de impuestos, detectar patrones de crimen y terrorismo y mejorar el dominio en expansión del e-government.

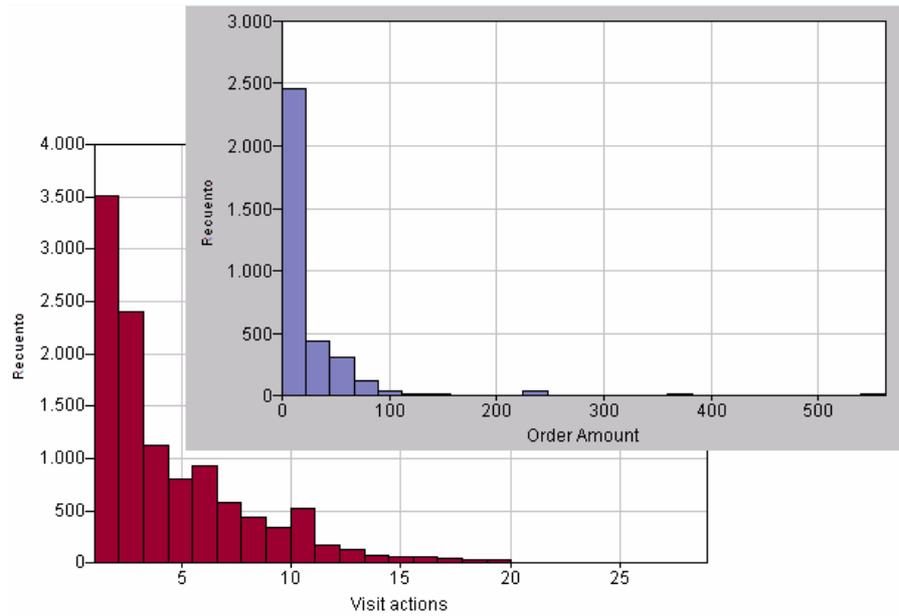
**CRM.client relationship management** La administración de relaciones con los clientes se puede mejorar gracias a la clasificación inteligente de tipos de clientes y los pronósticos exactos de producción. Clementine ha ayudado con éxito a las empresas de distintos sectores a atraer y retener los clientes más valiosos. [11]

---

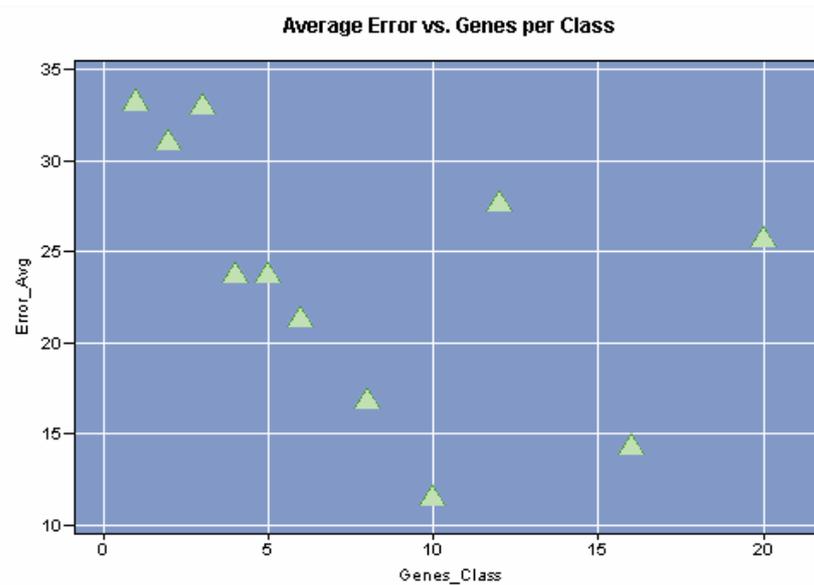
[11] SPSS, Manual de Usuario de Clementine 11.1



Minería Web. Mediante eficaces algoritmos de pronóstico y secuencias, Clementine contiene las herramientas necesarias para descubrir exactamente lo que hacen los invitados de un sitio Web y ofrecerles exactamente los productos o la información que deseen. Desde la preparación de datos hasta el modelado, el proceso completo de minería de datos se puede administrar dentro de Clementine.



Investigación farmacéutica y bioinformática. La minería de datos resulta útil en la investigación farmacéutica al analizar los enormes almacenes de datos que se obtienen de la creciente automatización de los laboratorios. Los modelos de clasificación y conglomerados de Clementine permiten generar pistas a partir de bibliotecas compuestas y la detección de secuencias que permiten descubrir patrones.

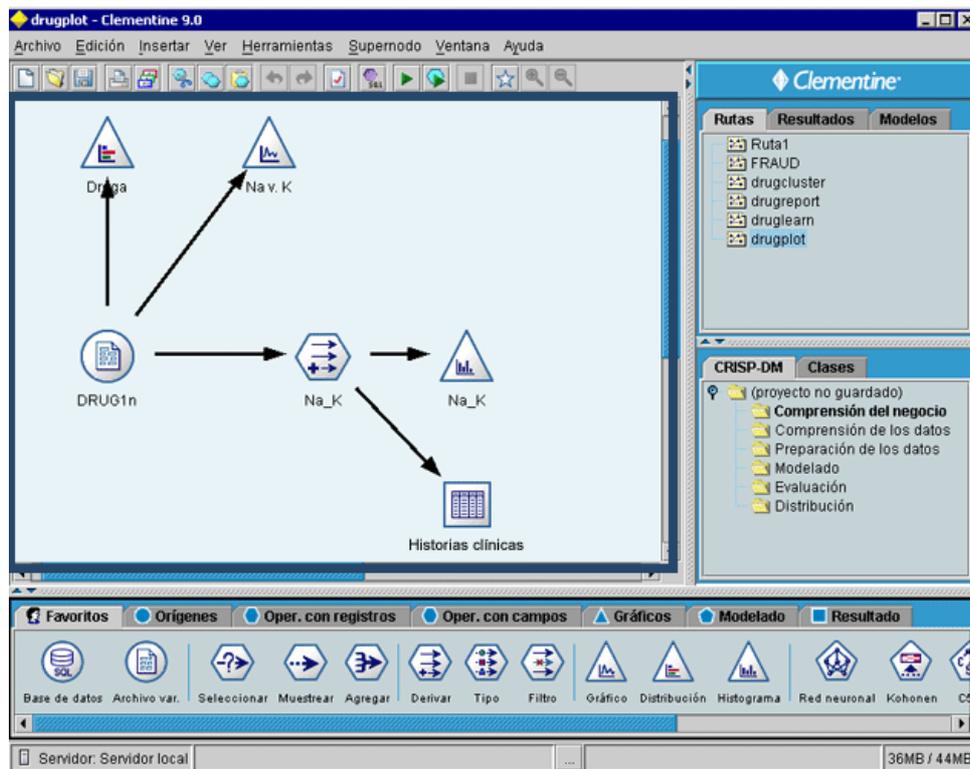


Interfase

Clementine 11.1 cuenta con una gran ayuda visual para poder trabajar y estructurar el modelo e interactuar con los datos de la vista minable. A continuación se describe cada uno de ellas:

### Lienzo de rutas

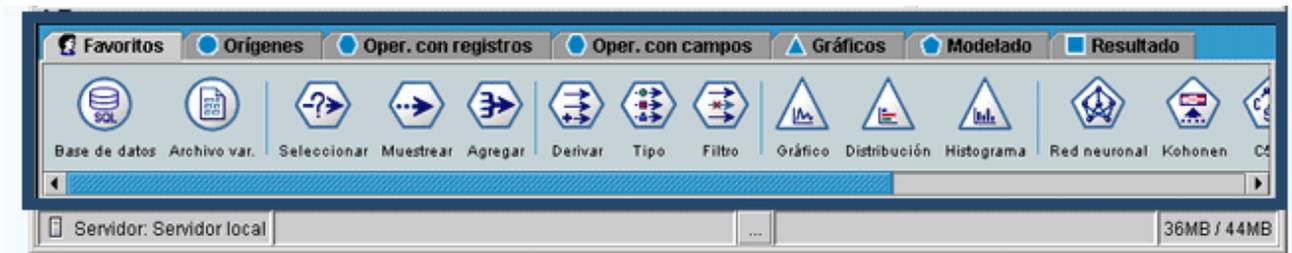
Cuando inicia Clementine por primera vez, el espacio de trabajo se abrirá en la vista por defecto. El área del centro se denomina lienzo de rutas. Ésta es el área principal que utilizará para trabajar con Clementine.



### Paletas

La mayoría de las herramientas de modelado y de datos de Clementine se encuentran en las paletas, el área situada bajo el lienzo de rutas. Cada ficha contiene grupos de nodos que representan gráficamente las *tareas de minería de datos*, como el acceso y el filtrado de datos, la creación de gráficos y la generación de modelos.

Para añadir nodos al lienzo, pulse dos veces en los iconos de las paletas de nodos o arrástrelos y suéltelos en el lienzo. A continuación, conéctelos para crear una ruta, que represente el flujo de datos.



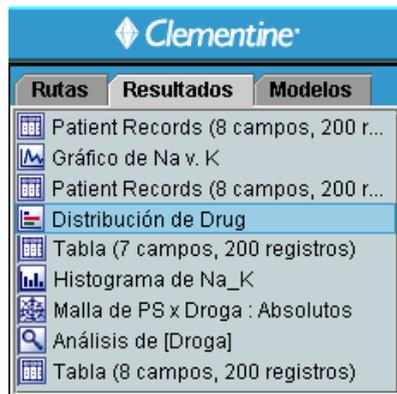
## Administradores

En la parte superior derecha de la ventana se encuentran los administradores de objetos y resultados. Estas fichas se utilizan para ver y administrar una serie de objetos de Clementine.



La ficha Rutas contiene todas las rutas abiertas en la sesión actual. Puede guardar las rutas y cerrarlas, así como añadirlas a un proyecto.

La ficha Resultados contiene una serie de archivos generados mediante operaciones de rutas con Clementine. Puede mostrar, cambiar el nombre y cerrar las tablas, gráficos e informes que se enumeran aquí.



La ficha Modelos es una herramienta eficaz que contiene todos los modelos generados (modelos que se han generado con Clementine) para una sesión. Los modelos se pueden examinar más detenidamente, añadir a la ruta, exportar o anotar.

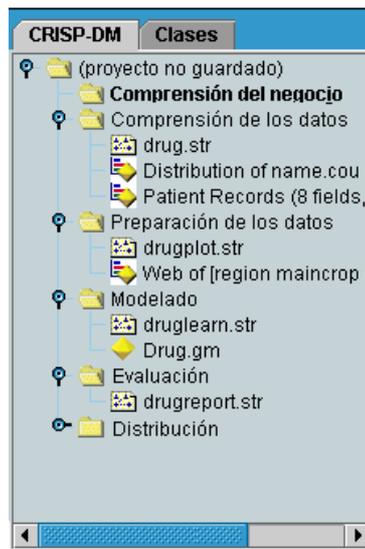


## Proyectos

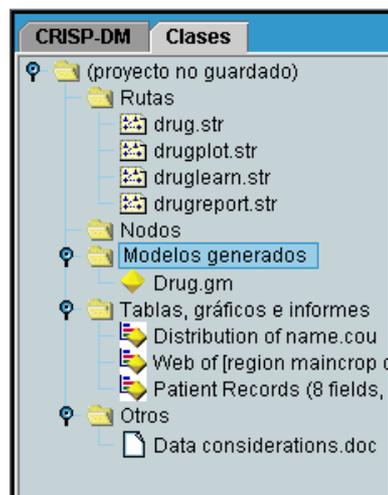
En la parte inferior derecha de la ventana se encuentra la herramienta proyectos, que se utiliza para crear y administrar los proyectos de minería de datos.

Existen dos formas de ver los proyectos que se crean con Clementine: la vista Clases y la vista CRISP-DM.

La ficha CRISP-DM permite organizar los proyectos según el proceso CRISP-DM (Cross-Industry Standard Process for Data Mining), una metodología independiente y probada en el sector. Los analizadores de datos con o sin experiencia pueden utilizar la herramienta CRISP-DM para mejorar la organización y la comunicación de los esfuerzos.



La ficha Clases permite organizar el trabajo en Clementine de forma categórica, por los tipos de los objetos que se hayan creado. Esta vista resulta útil al realizar un inventario de datos, rutas, modelos, etc.



## Plantillas

Clementine ofrece plantillas para muchas de estas aplicaciones de minería de datos. Las plantillas de aplicaciones de Clementine, también denominadas CAT, están disponibles para los siguientes tipos de actividades:

- Minería Web
- Detección de fraude
- CRM analítico

- CRM analítico de telecomunicaciones
- Análisis de micromatriz
- Detección y prevención de delitos

## ANEXO 7: VISTAS MINABLES

### *Humedad Media y Temperatura Media*

```
CREATE VIEW VISTA_HUM_AVG_TEM_MAX
```

```
AS
```

```
(
```

```
SELECT MES, ANIO, NOMBRE, FECHA, VALOR
```

```
FROM
```

```
(SELECT M.D_TIE_MUE_MES MES,
```

```
M.D_TIE_MUE_ANIO ANIO ,
```

```
M.D_TIE_MUE_CODIGO FECHA,
```

```
EM.D_ELE_NOMBRE NOMBRE,
```

```
AVG(H.D_HEC_VALOR) VALOR
```

```
FROM HECHOS H, D_TIEMPO_MUESTRA M, D_HORA HO,
```

```
D_ELEMENTO_METEOROLOGICO EM
```

```
WHERE M.D_TIE_MUE_CODIGO = HO.D_TIE_MUE_CODIGO AND
```

```
HO.D_HOR_CODIGO = H.D_HOR_CODIGO AND
```

```
H.D_ELE_CODIGO = 'H' AND
```

```
EM.D_ELE_CODIGO = H.D_ELE_CODIGO
```

```
GROUP BY (M.D_TIE_MUE_MES , M.D_TIE_MUE_ANIO, M.D_TIE_MUE_CODIGO
```

```
,EM.D_ELE_NOMBRE)
```

```
ORDER BY FECHA DESC)
```

```
UNION
```

```
SELECT MES, ANIO,NOMBRE,FECHA, VALOR
```

```
FROM(
```

```
SELECT M.D_TIE_MUE_MES MES,
```

```
M.D_TIE_MUE_ANIO ANIO ,
```

```
M.D_TIE_MUE_CODIGO FECHA,
```

```
EM.D_ELE_NOMBRE NOMBRE,
```

```
AVG(H.D_HEC_VALOR) VALOR
```

```
FROM HECHOS H, D_TIEMPO_MUESTRA M, D_HORA HO,
```

```
D_ELEMENTO_METEOROLOGICO EM
```

```
WHERE M.D_TIE_MUE_CODIGO = HO.D_TIE_MUE_CODIGO AND
```

```
HO.D_HOR_CODIGO = H.D_HOR_CODIGO AND
```

```
H.D_ELE_CODIGO = 'T' AND
```

```
EM.D_ELE_CODIGO = H.D_ELE_CODIGO
```

```
)
```

## *Temperatura*

```
CREATE VIEW VISTA_TEMPERATURA
AS
(
SELECT
M.D_TIE_MUE_CODIGO,
M.D_TIE_MUE_MES MES,
M.D_TIE_MUE_ANIO ANIO,
MAX(H.D_HEC_VALOR) VALOR_MAX,
AVG (H.D_HEC_VALOR) VALOR_PRO,
MIN(H.D_HEC_VALOR) VALOR_MIN
FROM HECHOS H, D_TIEMPO_MUESTRA M, D_HORA HO
WHERE M.D_TIE_MUE_CODIGO = HO.D_TIE_MUE_CODIGO AND
HO.D_HOR_CODIGO = H.D_HOR_CODIGO AND
H.D_ELE_CODIGO = 'T'
GROUP BY (M.D_TIE_MUE_MES, M.D_TIE_MUE_ANIO,M.D_TIE_MUE_CODIGO)
)
ORDER BY M.D_TIE_MUE_CODIGO DESC
```

