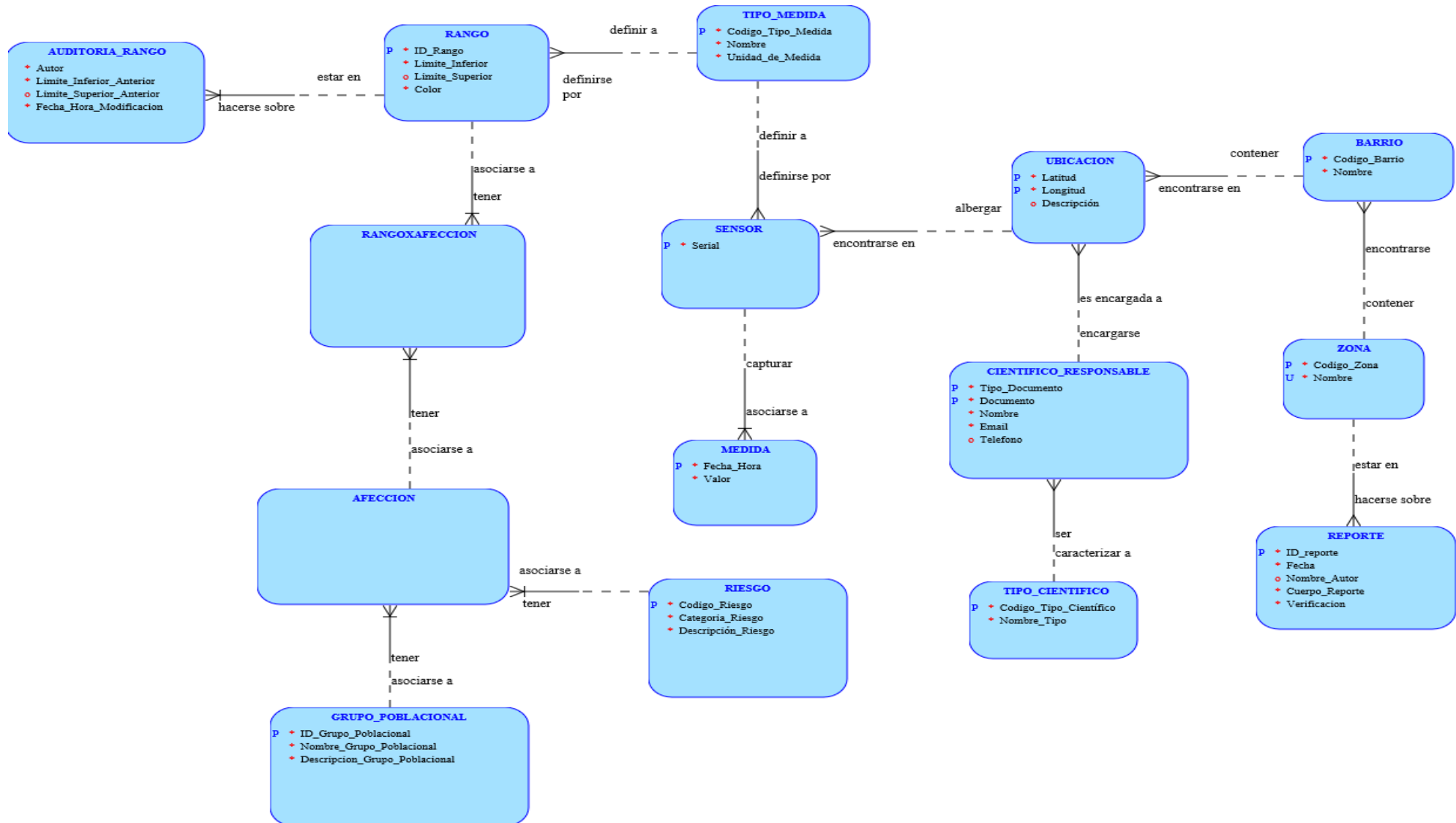


DISEÑO FÍSICO

Historia de Condiciones Climatológicas y Ambientales en el Valle de Aburrá

Felipe Rodríguez Ángel - Juan Luis Rojas Rincón

MODELO ENTIDAD RELACIÓN



DESCRIPCIÓN DE TRANSACCIONES CRÍTICAS

Transacción 1: Registrar Medición

Esta transacción permite registrar las mediciones capturadas por los sensores. Es la transacción más crítica para cumplir los objetivos del sistema porque es necesario registrar constantemente múltiples mediciones provenientes de múltiples sensores en múltiples ubicaciones, que luego son utilizadas por otros procesos del sistema como la identificación de afecciones en rangos.

Es, además, una transacción muy frecuente debido al elevado número de sensores registrando mediciones en lapsos de tiempo regulares.

En esta transacción solo se debe insertar un nuevo registro a la tabla Medida, ya que a partir del atributo Sensor podemos conocer el Tipo de Medida que se capturó y la ubicación en la que se realizó.

Tablas Utilizadas: MEDIDA.

Transacción 2: Registrar Reporte

Esta transacción permite registrar los reportes elaborados por los usuarios acerca de una consulta. Un Reporte posee información relacionada con una Zona. Es la comunicación entre el Usuario de la Comunidad y el Sistema, debido a que las otras transacciones están orientadas a la manipulación de mediciones o de sensores.

Los reportes, debido a los requisitos de tamaño en su cuerpo, son de una carga de almacenamiento alta en el sistema, además, el volumen de reportes realizados es directamente proporcional a la cantidad de medidas existentes en la base de datos, debido a que los reportes se realizan sobre estas.

Esta transacción solo debe insertar un nuevo registro en la tabla Reporte, ya que a partir del atributo Zona podemos conocer la demás información pertinente para una consulta.

Tablas Utilizadas: REPORTE.

Transacción 3: Consultar Mediciones en una Zona, Intervalo Temporal y Tipo de Medida.

Las mediciones en una Zona están dadas por todas las mediciones capturadas en todos los barrios pertenecientes a la Zona. Esta información es requerida constantemente en el sistema, pues es necesaria para las estadísticas de mediciones en sus divisiones geopolíticas y para la definición de los riesgos de afección en grupos poblacionales.

Es una consulta común en el sistema.

Para conocer toda la información correspondiente a las medidas de una Zona en un intervalo de tiempo es necesario hacer un recorrido a partir de la tabla Medida, por medio de la clave foránea que apunta al Sensor que la capturó accedo a los datos completos sobre el Tipo de Medida y su Ubicación, la Ubicación se encontrará en un Barrio que pertenece a una Zona, proyectando solo aquellas mediciones que cumplan con los criterios de la consulta, es decir, la Zona, Intervalo de Tiempo y Tipo de Medida estipulados.

Tablas Utilizadas: MEDIDA, SENSOR, TIPO DE MEDIDA, UBICACIÓN, BARRIO, ZONA

Transacción 4: Consultar Afección en una Zona e Intervalo Temporal

En esta transacción se obtienen las afecciones que se dan en una Zona (recordemos que el Rango está relacionado por un Tipo de Medida) y en un Intervalo temporal.

Esta consulta es crítica para soportar uno de los objetivos más importantes del sistema, la Alerta de Riesgos en Grupos Poblacionales.

Realizar una consulta sobre posibles Afecciones existentes en determinada Zona requiere hacer un recorrido desde la tabla intersección Rango por Afección, donde a partir de la clave foránea compuesta de la Afección se accede a la información completa correspondiente al Riesgo y Grupo Poblacional que afecta, por su parte, con la clave foránea que apunta al Rango y sus atributos se puede llegar al Tipo de Medida. Se proyectan entonces aquellas medidas que tengan el Tipo y se encuentren en el Rango estipulado para una Afección.

Tablas Utilizadas: GRUPO POBLACIONAL , AFECCIÓN , RIESGO , RANGO X AFECCION, RANGO , MEDIDA, SENSOR , TIPO DE MEDIDA , UBICACIÓN , BARRIO , ZONA

Transacción 5: Registrar una nueva Ubicación

Esta transacción registra nuevas ubicaciones donde pueden situarse sensores utilizados para llevar a cabo los objetivos del sistema. Se considera una transacción crítica porque constantemente se vincularán nuevas ubicaciones al sistema.

Esta transacción afecta a la tabla Ubicación y a la tabla Científico Responsable, ya que es posible que el Científico Responsable de una nueva Ubicación aún no se encuentre en la Base de Datos, esta transacción no afecta a la tabla Sensor debido a que, y en base al modelo Entidad Relación, una Ubicación puede o no tener Sensores asociados en determinado tiempo

Tablas Utilizadas: UBICACIÓN , CIENTIFICO RESPONSABLE

TRANSACCIONES CRÍTICAS

Tabla / Transacción	Registrar	Registrar	Consultar Mediciones en	Consultar Afección	Registrar una
	Medición	Reporte	una Zona, Intervalo Temporal y Tipo de Medida.	en una Zona e Intervalo Temporal	nueva Ubicación
Afección					
id_grupo_poblacional	-	-	-	K	-
riesgo_codigo_riesgo	-	-	-	K	-
Auditoria_Rango					
autor	-	-	-	-	-
limite_inferior_anterior	-	-	-	-	-
limite_superior_anterior	-	-	-	-	-
rango_id_rango	-	-	-	-	-
fecha_hora_modificacion	-	-	-	-	-
Barrio					
codigo_barrio	-	-	K	K	-
nombre	-	-	-	-	-
zona_codigo_zona	-	-	S	S	-
Cientifico_Responsable					
tipo_documento	-	-	-	-	I
documento	-	-	-	-	I
nombre	-	-	-	-	I
email	-	-	-	-	I
telefono	-	-	-	-	I
codigo_tipo_cientifico	-	-	-	-	I
Grupo_Poblacional					
id_grupo_poblacional	-	-	-	K	-
nombre_grupo_poblacional	-	-	-	S	-

descripcion_grupo_poblacional	-	-	-	S	-
Medida					
sensor_serial	I	-	K-S	K-S	-
fecha_hora	I	-	K-S	K-S	-
valor	I	-	S	-	-
Rango					
id_rango	-	-	-	K	-
limite_inferior	-	-	-	S	-
limite_superior	-	-	-	S	-
color	-	-	-	S	-
tipo_medida_codigo_tipo_medida	-	-	-	S	-
RangoxAfeccion					
rango_id_rango	-	-	-	K-S	-
afeccion_id_grupo_poblacional	-	-	-	K-S	-
afeccion_codigo_riesgo	-	-	-	S	-
Reporte					
id_reporte	-	I	-	-	-
fecha	-	I	-	-	-
nombre_autor	-	I	-	-	-
cuerpo_reporte	-	I	-	-	-
zona_codigo_zona	-	I	-	-	-
Verificacion	-	I	-	-	-
Riesgo					
codigo_riesgo	-	-	-	K	-
categoria_riesgo	-	-	-	S	-
descripción_riesgo	-	-	-	S	-
Sensor					

serial	-	-	K	K	-
ubicacion_latitud	-	-	S	S	-
ubicacion_longitud	-	-	S	S	-
tipo_medida_codigo_tipo_medida	-	-	S	S	-
Tipo_Cientifico					
codigo_tipo_cientifico	-	-	-	-	-
nombre_tipo	-	-	-	-	-
Tipo_Medida					
codigo_tipo_medida	-	-	K	K	-
nombre	-	-	S	S	-
unidad_de_medida	-	-	S	S	-
Ubicación					
latitud	-	-	K	K	I
longitud	-	-	K	K	I
descripción	-	-	-	-	I
barrio_codigo_barrio	-	-	S	S	I
cientifico_tipo_documento	-	-	-	-	I
cientifico_documento	-	-	-	-	I
Zona					
codigo_zona	-	-	K	K	-
nombre	-	-	S	S	-

Frecuencia de Tx por DIA	87696	87696	30000	30000	580
--------------------------	-------	-------	-------	-------	-----

MAGNITUDES Y PROYECCIONES

En este segmento se presentan las magnitudes obtenidas de algunas de las Tablas más representativas en el Sistema. Para realizar estos cálculos se tomaron como referencia las Fórmulas vistas en clase y las proyecciones iniciales realizadas en la Fase del Análisis del Sistema. Todos estos cálculos están disponibles en [esta hoja de cálculo de Google Sheet](#). Pero también están incluidos a continuación:

CIENTIFICO_RESPONSABLE	VALOR EJEMPLO	LONGITUD			
Tipo_Documento (var)	CC	2			
Documento (var)	1037669179	10			
Nombre (var)	Jose Jesus del Mar Ramirez	26			
Email (var)	josefitocorreo1817912@yahoo.com	31			
Telefono (var)	573217544505	12			
Código_Tipo_Cientifico (var)	101	3	Registros por Bloque	Número de Registros En 20 Años	TAMAÑO TABLA
	total bytes registro	100	80	104,357	10,435,700 bytes
				(Nuevos Científicos por Semana * # Semanas en 20 años)	9.952 Megabytes
					0.010 Gigabytes
MEDIDA	VALOR EJEMPLO	LONGITUD			
Fecha_Hora	1970-01-01 0:00:01	18			
Valor	242.0387	8			
SENSOR_serial (var)	CMK16GX4M2B3000C15	18	Registros por Bloque	Número de Registros En 20 Años	TAMAÑO TABLA
	total bytes registro	52	151	640,618,403	33,312,156,956 bytes
				(Mediciones por Semana * # Semanas en 20 años)	31,768.949 Megabytes
					31.024 Gigabytes

SENSOR	VALOR EJEMPLO	LONGITUD			
Serial (var)	CMK16GX4M2B3000C 15	18			
Ubicacion_Latitud	-75.56359	9			
Ubicacion_Longitud	6.25184	7			
TIPO_MEDIDA_Codigo_Tipo_Medida (var)	10055	5	Registros por Bloque	(Nuevos Sensores por Semana * # Semanas en 20 años)	
	total bytes registro	50			
				Número de Registros En 20 Años	TAMAÑO TABLA
				417,428	20,871,400 bytes
					19.905 Megabytes
					0.019 Gigabytes
REPORTE	VALOR EJEMPLO	LONGITUD			
ID_reporte	1897856541	10			
Fecha	1970-01-01 0:00:01	18			
Nombre_Autor (var)	Jose Jesus del Mar Ramirez	26			
Cuerpo_Reporte (var)		512			
ZONA_codigo_zona (var)	15027	5			
Verificacion (var)	En Verificacion	15	Registros por Bloque	(Mediciones por Semana * # Semanas en 20 años)	
	total bytes registro	603			
				Número de Registros En 20 Años	TAMAÑO TABLA
				640,618,403	386,292,897,009 bytes
					359.7633 Gigabytes
					0.3513 Terabytes

UBICACIÓN	VALOR EJEMPLO	LONGITUD			
Latitud	-75.56359	9			
Longitud	6.2518465	9			
Descripción (var)	Ubicación de mediciones estándar	32			
BARRIO_Codigo_Barrio (var)	1234567890	10		Número de Registros En 20 Años	TAMAÑO TABLA
CIENTIFICO_Tipo_Documento (var)	CC	2		104,357	8,870,345 bytes
CIENTIFICO_Documento (var)	1037669179	10	Registros por Bloque	(Ubicaciones nuevas por Semana * # Semanas en 20 años)	8.459 Megabytes
	total bytes registro	85	94		0.008 Gigabytes

UBICACIÓN	VALOR EJEMPLO	LONGITUD			
Latitud	-75.56359	9			
Longitud	6.2518465	9			
Descripción (var)	Ubicación de mediciones estándar	32			
BARRIO_Codigo_Barrio (var)	1234567890	10		Número de Registros En 20 Años	TAMAÑO TABLA
CIENTIFICO_Tipo_Documento (var)	CC	2		104,357	8,870,345 bytes

CIENTIFICO_Documento (var)	1037669179	10	Registros por Bloque	(Ubicaciones nuevas por Semana * # Semanas en 20 años)	8.459	Megabytes
	total bytes registro	85	94		0.008	Gigabytes

ZONA	VALOR EJEMPLO	LONGITUD		Número de Registros En 20 Años	TAMAÑO TABLA	
Codigo_Zona	AB123	5		4,174	116,872	bytes
Nombre (var)	SUROCCIDENTAL	13	Registros por Bloque	(Estimación máxima zonas nuevas a la semana * # Semanas en 20 años)	114.133	Kilobytes
	total bytes registro	28	272		0.111	Megabytes

BARRIO	VALOR EJEMPLO	LONGITUD		Número de Registros En 20 Años	TAMAÑO TABLA	
codigo_barrio (var)	AB00050043	10		26,089	1,200,094	bytes
nombre (var)	PICACHO	7				
zona_codigo_zona (var)	CMK16GX4M2B3 000C15	18	Registros por Bloque	(Estimación barrios nuevos a la semana * # Semanas en 20 años)	1.144	Megabytes
	total bytes registro	46	170		0.001	Gigabytes

FÓRMULAS Y VARIABLES UTILIZADAS

FÓRMULA PARA OBTENER LOS REGISTROS POR PÁGINA		
DB_BLOCK SIZE	8192	
LONGITUD TIPICA REGISTRO (multiplos 4)	" = IF(RESIDUO(SUMA(Cx:Cy), 4) = 0 , SUMA(Cx:Cy) , SUMA(Cx:Cy) + (4 - RESIDUO(SUMA(Cx:Cy) , 4))) + #campos_longitud_variable + byte_longitud_registro"	Donde Cx:Cy son los atributos de la Tabla.
R	Registros por Página	
ECUACIÓN	$DB_BLOCK_SIZE = 4 + 1 + (2 * R) + (R * L)$	
REGISTROS EN EL PERIODO DE TIEMPO Y TAMAÑO DE LA TABLA		
MEDICIONES POR SEMANA	613872	Obtenido del análisis de sistema
REPORTES POR SEMANA	613872	Magnitud del Sistema. Análisis del Sistema
SEMANAS EN PERIODO DE 20 AÑOS	520.29	
TAMAÑO DE LA TABLA	$Tamaño_Tabla = Longitud_Registro * (Numero_Semanas_20_años * Registros_por_Semana)$	

LINK a la Hoja de Cálculo: <https://docs.google.com/spreadsheets/d/1i7OLOEpw9S9uj2h2qOOeYkdye6JDLWZWNFahvWKd-dk/edit?usp=sharing>

JUSTIFICACIÓN DE ESQUEMA DE ALMACENAMIENTO USADO

A continuación se explicará la elección de las Estructuras Físicas de las Tablas más relevantes y posteriormente se calcularán los tamaños (o alturas) de los Índices utilizados. Las tablas de intersección utilizarán un Árbol B+ con las páginas en sus hojas.

NOMBRE TABLA	ESTRUCTURA ELEGIDA	JUSTIFICACIÓN
SENSOR	CLUSTER	Estas Tablas serán guardadas en un CLUSTER debido a que se realizan muchas veces JOINS entre ellas. Al guardarlas en un mismo CLUSTER esta operación se agiliza. Facilita también la búsqueda por Rangos de Sensor y medida.
MEDIDA	CLUSTER	Además, debido a la naturaleza de los datos no se realizarán actualizaciones de la clave cluster.
CIENTÍFICO_RESPONSABLE	ARCHIVO ALEATORIO	Las Transacciones en esta tabla son mayormente consultas de igualdad para identificar a un Científico Responsable a través del Tipo de Documento y Documento del Científico, este tipo de búsqueda por igualdad es priorizada en esta tabla sobre búsquedas por rango. Por lo tanto, se elige la estructura de archivo aleatorio.
REPORTE	HEAP	Se utiliza la estructura HEAP debido al alto número de Inserciones en la Tabla Reporte, aunque esta tabla maneja una gran cantidad de registros las operaciones de búsqueda sobre estos son de mínima frecuencia.
UBICACIÓN	CLUSTER	Se crea un CLUSTER en el que están las tablas UBICACIÓN y BARRIO.
BARRIO	CLUSTER	Tanto para obtener estadísticas sobre zonas geopolíticas como para visualizar localizaciones geográficas es necesario conocer las ubicaciones presentes en un barrio. Esto se traduce a un JOIN entre ambas tablas, por lo que el CLUSTER facilita esta operación.
ZONA	HEAP	Se utiliza un HEAP para la tabla ZONA debido a que no es una tabla muy masiva y los registros totales de la tabla pueden almacenarse en pocos bloques (como fue calculado en las magnitudes), sin embargo la tabla se considera crítica debido a su estrecha interrelación con las transacciones relacionadas con la información geopolítica y de reportes

CLUSTER Y ÁREA CLUSTER

Se decidió utilizar un CLUSTER que contenga las tablas MEDIDA y SENSOR.

Para calcular el ÁREA DEL CLUSTER utilizamos la siguiente fórmula:

$$\begin{aligned} \text{Área Cluster} = & \\ & \text{número_de_Tablas} + \# \text{registros_tabla_n} + \text{desplazamiento_tabla_n+1} + \# \text{registros_tabla_n+1} + \text{desplazamiento_tabla_n} + \\ & \text{longitud_clave_cluster} + \text{clave_cluster} + \text{espacio_registros_tabla_n} + \text{espacio_registros_tabla_n+1} \end{aligned}$$

Donde n , n+1... n+m; es el número de tablas que pertenecen al cluster y m es el número de tablas y la unidad de medida es BYTES.

Así, con los siguientes datos calculamos el tamaño del CLUSTER:

CLUSTER MEDIDA-SENSOR	TAMAÑO(BYTES)
Fecha_Hora	18
SENSOR_serial (var)	18
Serial (var)	18
Cantidad de Tablas	2
Número Registros Tabla Medida	4
Número Registros Tabla Sensor	4
Desplazamiento Tabla Medida	2,562,473,612
Desplazamiento Tabla Sensor	1,669,712
Longitud de la Clave	1
Clave Cluster	18

Espacio Registros Tabla Medida	33,312,156,956	
Espacio Registros Tabla Sensor	20,871,400	UNIDADES
Sumatoria	35,897,171,763	bytes
	34234.21074	megabytes
	33.43184643	gigabytes

CLUSTER UBICACIÓN-BARRIO	TAMAÑO(BYTES)	
Latitud	9	
Longitud	9	
codigo_barrio (var)	10	
código_zona	5	
Cantidad de Tablas	3	
Número Registros Tabla Ubicación	6	
Número Registros Tabla Barrio	3	
Desplazamiento Tabla Ubicación	417428	
Desplazamiento Tabla Barrio	104356	
Longitud de la Clave	1	
<u>Clave Cluster</u>	10	
Espacio Registros Tabla Ubicación	8,870,345	
Espacio Registros Tabla Barrio	1,200,094	UNIDADES
Sumatoria	10,592,279	bytes
	10344.02246	kilobytes
	10.10158443	megabytes

Árbol B+

1. Para los Árboles B+ se asume que un bloque es igual a una página del árbol y que

$$2d - 1 = \text{tamaño_pagina} / (\text{id_tamaño} + 4)$$

En donde d es el orden del árbol.

2. Posteriormente, se calcula el número de páginas con la siguiente Fórmula:

$$\text{Páginas} = \text{Registros_en_20_años} / d$$

3. Por último, se calcula la altura del árbol con la Fórmula:

$$\text{Altura} = \log_{\text{Orden_Árbol}}(\text{Páginas})$$

Utilizando estas fórmulas para calcular en Tabla REPORTE el B+ del índice:

- Para obtener el orden del árbol:
 - Para obtener el número de páginas:
 - Para obtener la altura del Árbol:
- $$2d - 1 = 8192 / (10 + 4) \implies \mathbf{293 \text{ (Orden del árbol)}}$$
$$\text{Páginas} = 640618403 / 293 \implies \mathbf{2186411 \text{ (Número de Páginas)}}$$
$$\text{Altura} = \log_{293}(640618403) \implies \mathbf{3.569 \text{ (Altura del Árbol)}}$$

ZONAS DE ALMACENAMIENTO

Para calcular el tamaño de las zonas de almacenamiento hacemos uso de la siguiente fórmula:

$$Zona\ de\ almacenamiento = \frac{número\ de\ registros}{registros\ por\ bloque * número\ zonas\ almacenamiento}$$

Así, el tamaño de las zonas de almacenamiento para la tabla Científico Responsable es:

$$1.3031 = \frac{104357}{80 * 1001}$$

Las cubetas no pueden ser menores a un bloque físico, cuando es menor se aproxima al tamaño de bloque definido (8192).

MATRIZ DE TABLAS Y ESQUEMA DE ALMACENAMIENTO USADO

Tabla.	Columnas más usadas (1)	Esquema de almacenamiento usado. (2)	Clave cluster o Clave Hash (3)	Indice primario (4)
SENSOR	Serial, Ubicacion_Latitud, Ubicación_Longitud, Codigo_Tipo_Medida	CLUSTER	Serial	
MEDIDA	Sensor_Serial, Fecha_Hora, Valor	CLUSTER	Sensor_Serial	
REPORTE	Id_reporte, fecha, nombre_autor, cuerpo_reporte, ZONA_codigo_zona, Verificacion	HEAP		ID_Reporte
CIENTÍFICO RESPONSABLE	Tipo_documento, documento, nombre, email, telefono, codigo_tipo_cientifico	ARCHIVO ALEATORIO	CLAVE HASH: 1001 Función Propuesta: $h_1(\text{tipo_documento}, \text{documento}) = (\text{TO_NUMBER}(\text{tipo_documento}) + \text{TO_NUMBER}(\text{documento})) \bmod 1001$	
UBICACIÓN	Latitud,Longitud, Barrio_Codigo_Barrio	CLUSTER	Barrio_Codigo_Barrio	
BARRIO	Codigo_barrio, ZONA_codigo_zona	CLUSTER	Codigo_Barrio	
ZONA	Codigo_zona nombre	HEAP		Codigo_zona

MATRIZ DE ÍNDICES DEFINIDOS POR TABLA

Nombre del Índice	Tabla asociada	Tamaño páginas del índice	Modalidad de índice B+, Bitmap, etc.	Altura de árbol (si aplica)	Columnas que conforman el índice
I_reporte	REPORTE	2186411	B+	3.569	ID_reporte
I_reporte_verificación	REPORTE	2965825	B+	2.772	Verificación
I_zona	ZONA	58	B+	1.66	codigo_zona
I_reporte_fecha_zona	REPORTE	4187048	B+	3.031	Fecha, Zona_codigo_zona

INSTRUCCIONES SQL

Debido a la longitud de las instrucciones SQL, se adjunta con este trabajo el **.sql** correspondiente.

Algunos comentarios relevantes sobre este son:

- Se reemplazó el índice que define Oracle automáticamente en las Tablas Reporte y Zona por uno personalizado.
- Se utilizó el índice que define Oracle automáticamente en las Tablas de Referencia (Tipo Medida y Tipo Científico)
- Se usan Tablas Indexadas (Árboles B+ con Páginas en sus hojas) para las tablas de intersección (Afección RangoXAfeccion)
- Se omitieron los comentarios de las tablas y los atributos en este SQL para facilitar la lectura de Docente (la longitud del archivo disminuye considerablemente)
- Se definieron múltiples TableSpace según la lógica del negocio.