**Parcial 2**

**Base de Datos 2**

**Integrantes:** Casermeiro Maria Silvia, Chiletti Emanuel, Vietto Santiago

**Fecha límite de presentación: Jueves 4 de Noviembre de 2021**

**Objetivo:**

Implementar una solución de Inteligencia de Negocios siguiendo la metodología Hefesto.

**Especificaciones de evaluación:**

El Modelo Multidimensional y relacional deberán estar correctamente documentados con la profundidad suficiente como para que otra persona lo lea y comprenda sin ambigüedades.

La documentación del Modelo Multidimensional y relacional (de implementación) consistirá en lo propuesto por la metodología Hefesto (Modelo Multidimensional, DER, etc) más un diccionario de datos el cual debe tener la información suficiente para cumplir el punto anterior.

**Detalle de las funciones a implementar:**

1. Presentación de la situación problemática que requiere de una solución de Inteligencia de Negocios, plantear problema, objetivos, tecnologías y solución. Se valorará la selección de un set de datos público referido a temas relevantes y actuales.
2. Modelo Multidimensional correctamente documentado siguiendo la metodología Hefesto. Poner especial énfasis en la argumentación de los indicadores implementados.
3. Data Warehouse (DW): esquema de base de datos relacional de implementación del Modelo Multidimensional (esquema estrella, copo de nieve o constelación)
4. Proceso de integración de datos: crear los procesos de transformación de datos necesarios para poblar el DW a partir de la o las fuentes de datos relevadas.
5. Crear algunos bocetos de futuros reportes y tableros.
6. Componentes BI: crear y justificar los siguientes componentes BI:
   1. Al menos 4 consultas (análisis) multidimensionales dinámicas con y sin gráficos que respondan a cada una de las preguntas planteadas en el proceso que propone Hefesto.
   2. 5 Reportes en los cuales deben estar presentes (al menos): agrupamientos, gráficos, subreportes, parámetros y funciones.
   3. 3 Tableros que contengan (al menos): filtros, gráficos, tablas e interacciones.

Plasmar en el documento capturas de los componentes gráficos BI y mapearlos con las preguntas que responden.

1. Se valorará la aplicación de una o más tecnologías que hayan sido desarrolladas en las presentaciones realizadas por los pares, o bien tecnologías propuestas de forma espontánea o por necesidad puntual.

**Aclaraciones:**

* Pueden formarse grupos de hasta 3 estudiantes para presentar el trabajo.
* Se requiere la concreción de los puntos 1 a 5 para el segundo parcial. Cada grupo dispondrá de 15 minutos para exponer su trabajo. Se recomienda dirigir la exposición por medio de una presentación. Se debe hacer especial hincapié en mostrar y demostrar que los indicadores y perspectivas seleccionadas responden a las preguntas planteadas y permitirán realizar análisis sustantivos de la información contenida en el data set. Mostrar los bocetos que propone el punto 5, a fin de afianzar los argumentos que defienden el trabajo.
* Se debe anexar al trabajo, todo material o solución que no haya sido parte formal de la metodología y que se considere como valor agregado.
* El trabajo se defenderá y argumentará en formato online, cada grupo dispondrá de 20 minutos para exponer de forma equitativa por integrante. El grupo deberá guiarse con una agenda o presentación para mostrar el trabajo. La conclusión del trabajo tiene una gran relevancia en el trabajo, debe plantearse una conclusión personal y grupal muy sintética y puntual.
* El trabajo debe presentarse en formato electrónico que posea un backup de la base de datos y los archivos que representan los componentes de solución, el mismo debe contener las siguientes secciones:
* Carátula: Nombre de la Universidad, nombre de la materia, nombre del proyecto, año, nombre del docente (Ing. Mariano, GARCÍA MATTÍO), y nombre del o los estudiantes junto a sus correos electrónicos.
* Implementación de la metodología Hefesto.
* Diagrama de entidad relación (DER): deben verse con claridad las claves primarias y los nombre de todas las Tablas y columnas.
* Diccionario de Datos: debe poseer al menos las siguientes columnas:

1. Nombre Tabla
2. Nombre Columna
3. Descripción
4. Tipo datos
5. Rango de datos (en caso de que el dominio del tipo de datos sea demasiado amplio)
6. Fórmula (de poseerla)
7. Observaciones: notas o observaciones dignas de tener en cuenta o que sirvan para clarificar relaciones o conceptos.

* Conclusión
* Bibliografía (libros, links, revistas, etc)

**Anexo**

**Evaluación Final*:***

La evaluación final constará de la definición e implementación del 100% de los puntos descritos en el presente trabajo, con las correcciones propuestas en el examen parcial, además de la electrónica.

HEFESTO

Paso 1) Análisis de requerimientos - punto 1

1. Identificar preguntas
2. Identificar indicadores y perspectivas
3. Modelo conceptual

Paso 2) Análisis de los OLTP - punto 2

a) Conformar indicadores

b) Establecer correspondencias

c) Nivel de granularidad

d) Modelo Conceptual ampliado

PASO 3) MODELO LÓGICO DEL DW - punto 3

a) Tipo de Modelo Lógico del DW

b) Tablas de dimensiones

c) Tablas de hechos

d) Uniones

PASO 4) INTEGRACIÓN DE DATOS - punto 4

a) Carga Inicial

b) Actualización

Paso 1)

Proceso: Registro de casos

Indicador: Cantidad de personas infectadas por el Covid19

Preguntas de negocio:

Cantidad de personas infectadas en un periodo de tiempo.

Modelo conceptual

**Índice**

[**Introducción**](#_9nntd3nwtqmw) **6**

[**1) Análisis de requerimientos**](#_fowgxn77w4k6) **7**

[1.1) Preguntas de negocio](#_1ft64ib35d1) 7

[1.2) Indicadores y perspectivas](#_v5tya9he59ta) 7

[1.3) Modelo conceptual](#_7syfbqeewcdg) 9

[**2) Análisis de los OLTP**](#_hdmzfcczk54y) **9**

[2.1) Conformar indicadores](#_bdqwyv71zp6k) 9

[2.2) Establecer correspondencias](#_sb3qgbane8bm) 10

[2.3) Nivel de granularidad](#_j2s4nb7c5nol) 10

[2.4) Modelo conceptual ampliado](#_71cwjkhxxo6p) 10

[**3) Modelo lógico de DW**](#_na2pslsmmb8k) **10**

[3.1) Tipo de Modelo Lógico del DW](#_4k8iemvjkjgp) 10

[3.2) Tablas de dimensiones](#_eyg8cn8nz5b6) 10

[3.3) Tablas de hechos](#_l2ie06fmy246) 11

[3.4) Uniones](#_llvzwd7y8l1q) 11

[**Conclusión**](#_74wwm3w7jex2) **11**

# Introducción

El año 2020 fue un año que quedará marcado en la historia y en la vida de miles de millones de personas en todo el mundo. La pandemia de COVID-19, o coronavirus, es una [pandemia](https://es.wikipedia.org/wiki/Pandemia) mundial ocasionada por el virus [SARS-CoV-2](https://es.wikipedia.org/wiki/SARS-CoV-2), que continúa actualmente en varios países.Su primer caso fue identificado en diciembre de 2019 en la ciudad de [Wuhan](https://es.wikipedia.org/wiki/Wuhan), capital de la provincia de [Hubei](https://es.wikipedia.org/wiki/Hubei), en la [República Popular China](https://es.wikipedia.org/wiki/Rep%C3%BAblica_Popular_China), al reportarse casos de un grupo de personas enfermas con un tipo de [neumonía](https://es.wikipedia.org/wiki/Neumon%C3%ADa) desconocida. ​desde ese entonces la [Organización Mundial de la Salud](https://es.wikipedia.org/wiki/Organizaci%C3%B3n_Mundial_de_la_Salud) (OMS) la reconoció como una pandemia el 11 de marzo de 2020, cuando informó que había 4291 muertos y 118 000 casos en 114 países.

El virus [se transmite](https://es.wikipedia.org/wiki/Transmisi%C3%B3n_de_COVID-19) de persona a persona a través de pequeñas gotas de [saliva](https://es.wikipedia.org/wiki/Saliva) que se emiten al hablar, estornudar, toser o [respirar](https://es.wikipedia.org/wiki/Exhalaci%C3%B3n). Se difunde principalmente cuando las personas están en contacto cercano, pero también se puede difundir al tocar una superficie contaminada y luego llevar las manos contaminadas a la cara o las mucosas. Su [período de incubación](https://es.wikipedia.org/wiki/Per%C3%ADodo_de_incubaci%C3%B3n) suele ser de cinco días, pero puede variar de dos a catorce días. Los síntomas más comunes son la [fiebre](https://es.wikipedia.org/wiki/Fiebre), la [tos seca](https://es.wikipedia.org/wiki/Tos) y dificultades para respirar. Las complicaciones pueden incluir neumonía, síndrome respiratorio agudo o [sepsis](https://es.wikipedia.org/wiki/Sepsis).

Las medidas de prevención recomendadas incluyen [lavarse las manos](https://es.wikipedia.org/wiki/Higiene_de_manos), cubrirse la boca al toser, el [distanciamiento físico](https://es.wikipedia.org/wiki/Distanciamiento_f%C3%ADsico) entre las personas y el uso de [mascarillas](https://es.wikipedia.org/wiki/Mascarilla_FFP2), además del autoaislamiento y el seguimiento para las personas sospechosas de estar infectadas. La edad es un factor de riesgo, como por ejemplo mayores a 65 años, así como también la presencia de enfermedades crónicas como [diabetes](https://es.wikipedia.org/wiki/Diabetes), [cardiopatías](https://es.wikipedia.org/wiki/Cardiopat%C3%ADas), [enfermedades respiratorias](https://es.wikipedia.org/wiki/Enfermedades_respiratorias), [hipertensión arterial](https://es.wikipedia.org/wiki/Hipertensi%C3%B3n_arterial) o [inmunodeficiencias](https://es.wikipedia.org/wiki/Inmunodeficiencia) ya que tienen un riesgo mucho mayor de contraer la enfermedad y de llegar a tener complicaciones graves, por lo que se aconseja el aislamiento.

​Para evitar la expansión del virus, los gobiernos han impuesto restricciones de viajes, [cuarentenas](https://es.wikipedia.org/wiki/Cuarentena), confinamientos, cancelación de eventos y el cierre de establecimientos. Por otro lado, la pandemia ha tenido [un efecto socioeconómico disruptivo](https://es.wikipedia.org/wiki/Impacto_socioecon%C3%B3mico_de_la_pandemia_de_COVID-19).Se han cerrado colegios y universidades en más de 124 países, lo que ha afectado a más de 2200 millones de estudiantes. Más de un tercio de la población mundial se confinó, con fuertes restricciones a la libertad de circulación, lo cual ha conducido a una reducción drástica de la actividad económica​ y a un aumento paralelo del [desempleo](https://es.wikipedia.org/wiki/Desempleo). A partir de febrero de 2021, [varias vacunas](https://es.wikipedia.org/wiki/Vacuna_contra_la_COVID-19) estaban siendo desarrolladas y utilizadas por distintos países, lo que hasta el día de hoy ha logrado reducir los contagios y muertes, pero aún la lucha no ha terminado.

En plena crisis pandémica, la tecnología fue un aliado clave para lograr mejores resultados, tanto para reducir su propagación como para la mitigación y reducción de sus impactos, y es por esta razón que nos pareció interesante y una buena idea tomarlo como caso de estudio, ya que es algo que nos ha afectado a todos por igual.

# 1) Análisis de requerimientos

## 1.1) Preguntas de negocio

Nos abocamos al análisis de una situación de actualidad como es el covid-19. Hemos accedido a los datos subidos a la página oficial del gobierno de la Unión Europea. Los casos publicados como confirmados por criterio clínico epidemiológico con residencia en cada continente y cada país corresponden a casos confirmados por laboratorio.

Data set seleccionado: [Casos COVID-19](https://data.europa.eu/data/visualisation/?file=https%3A%2F%2Fopendata.ecdc.europa.eu%2Fcovid19%2Fcasedistribution%2Fcsv#)

El objetivo de nuestro análisis es brindar información acerca de los diferentes escenarios que se han tenido lugar en el último año hasta hoy respecto a la situación epidemiológica relacionada al covid-19. Todo esto, mostrará cómo se fue desarrollando el contagio del virus y cómo afectó a la población a nivel mundial.

Con el fin de obtener e identificar las necesidades de información clave de alto nivel para llevar a cabo el compendio de la información sobre el COVID-19 en el último año, y dado que no estamos trabajando directamente con las personas que diseñaron los datos sino que con un data set formado, simplemente vamos a analizar el mismo y determinar los requerimientos para luego generar las preguntas correspondientes.

* Queremos informar la cantidad de hisopados positivos en un determinado periodo de tiempo: *¿Qué cantidad de hisopados positivos hubo por tiempo por país?*
* Queremos informar la cantidad de fallecidos en un determinado periodo de tiempo: *¿Qué cantidad de fallecidos hubo por mes tiempo por país?*
* Queremos informar la cantidad máxima de fallecidos en un determinado tiempo: *¿Cuál fue el pico máximo de fallecidos por tiempo?*
* Queremos informar la cantidad promedio de fallecidos que ha habido en un país en un determinado tiempo: *¿Cuál es el promedio de fallecidos por país por tiempo?*
* Queremos informar la cantidad promedio de casos positivos por país: *¿Cuál es el promedio de casos por país?*
* Queremos informar la cantidad promedio de casos positivos por continente: *¿Cuál es el promedio de casos por continente?*

## 1.2) Indicadores y perspectivas

Luego de identificar preguntas en base al proceso de hisopado realizado y la identificación de los casos positivos, que luego fueron registrados, obtuvimos los siguientes indicadores y perspectivas:

| *¿Qué cantidad de hisopados positivos hubo por tiempo por país?*    Indicador perspectiva |
| --- |
| *¿Qué cantidad de fallecidos hubo por mes tiempo por país?*    Indicador perspectiva |
| *¿Cuál fue el pico máximo de fallecidos por tiempo?*    Indicador perspectiva |
| *¿Cuál es el promedio de fallecidos por país por tiempo?*    Indicador perspectiva |
| *¿Cuál es el promedio de casos por país?*    Indicador perspectiva |
| *¿Cuál es el promedio de casos por continente?*    Indicador perspectiva |

* Indicadores:
  + casos positivos por mes
  + fallecidos
  + pico máximo de fallecidos
  + Promedio de casos positivos
  + Promedio de fallecidos

Hemos identificado a “casos positivos por mes” como indicador dado que nos dará la cantidad de contagiados de covid-19 por mes a nivel mundial, es decir, sin discernimiento del país de origen del enfermo.

El siguiente, “fallecidos”, se refiere a la cantidad de personas que no han podido sobrellevar la enfermedad y han perdido la vida.

El indicador relacionado con el pico máximo “pico de máximo de fallecidos”, demuestra los valores máximos alcanzados.

Los dos últimos indicadores, “Promedio de contagiados” y “Promedio de fallecidos”, indican los afectados, en uno y otro caso, en promedio.

* Perspectivas:
  + tiempo
  + país
  + continente

La selección de las perspectivas se realizó sobre tiempo, país y continente.

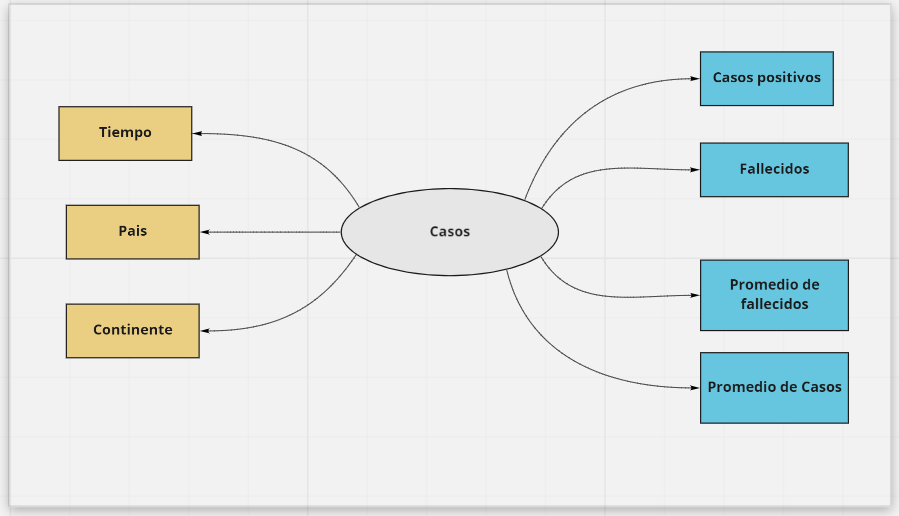
Tiempo: permite obtener respuestas concretas sobre el mes en particular

País: permite obtener información sobre la cantidad de casos por país.

Continente: al igual que el país, permite conocer los resultados de los efectos del virus por continente.

## 1.3) Modelo conceptual

Teniendo en cuenta los indicadores y perspectivas que fueron identificados, procedemos a realizar un modelo conceptual que nos permita describir la base de datos con respecto a los objetos, relaciones y los atributos, para así poder determinar los alcances del proyecto.

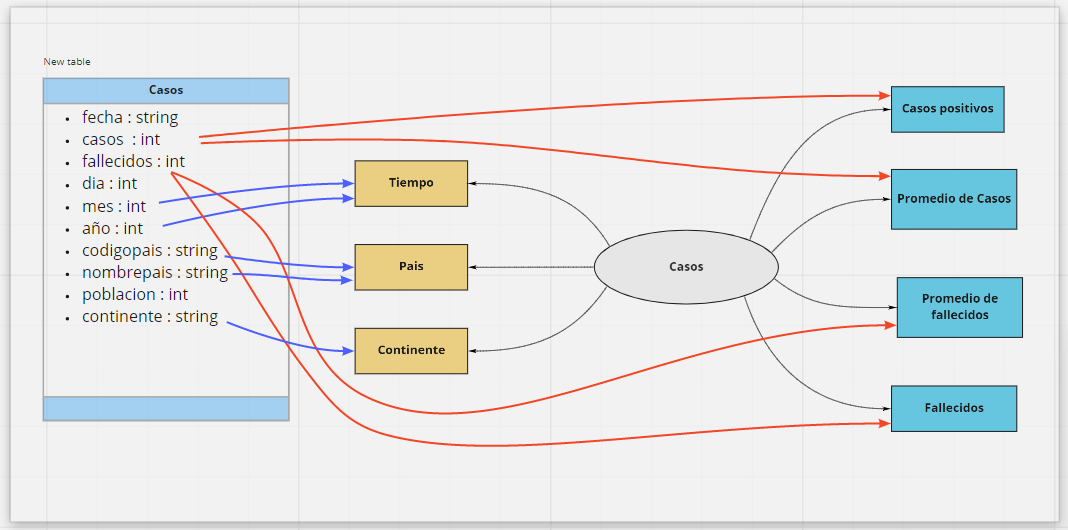


# 2) Análisis de los OLTP

## 2.1) Conformar indicadores

* “Cantidad de casos positivos”
  + Hecho: casos positivos
  + Función de sumarización: SUM
* “Cantidad de fallecidos”
  + Hecho: personas fallecidas
  + Función de sumarización: SUM
* “Promedio de casos positivos”
  + Hecho: casos positivos
  + Función de sumarización: AVG
* “Promedio de fallecidos”
  + Hecho: personas fallecidas
  + Función de sumarización: AVG

## 2.2) Establecer correspondencias



## 2.3) Nivel de granularidad

La especificación de la granularidad nos dará como resultado la determinación de las dimensiones que usaremos para conformar nuestro cubo multidimensional.

Las perspectiva seleccionadas están descritas como siguen:

**País**,

| Id\_Pais = PK -> dimPais | nombre |
| --- | --- |

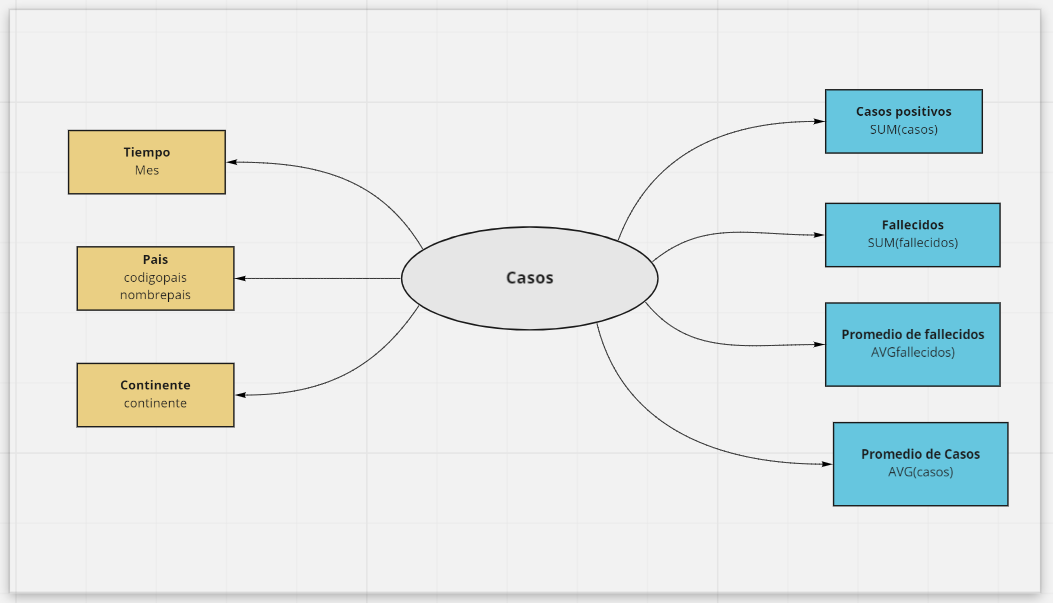
**Continente**,

| Id\_Continente = PK -> dimContinente |
| --- |

**Tiempo**,

| Id\_Mes = PK -> dimTiempo |
| --- |

## 2.4) Modelo conceptual ampliado



# 3) Modelo lógico de DW

## 3.1) Tipo de Modelo Lógico del DW

El esquema sugerido a utilizar es el de tipo Estrella. Cabe aclarar que es el que mejor se adapta al comportamiento de nuestro caso de estudio.

## 3.2) Tablas de dimensiones

Perspectiva: **País**

Como ya se mencionó en el punto 2.3, esta perspectiva pasará a llamarse dimPais, con la clave principal cuyo nombre es Id\_Pais luego de renombrada desde codigopais. Vamos a mantener el campo nombre\_Pais. El último campo será población

Perspectiva: **Continente**

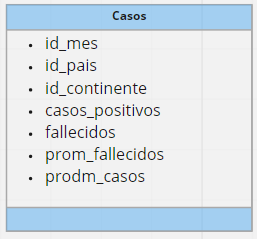
Como ya se mencionó en el punto 2.3, esta perspectiva pasará a llamarse dimContinente, con la clave principal cuyo nombre es Id\_Continente. Vamos a mantener el campo nombre\_Continente

Perspectiva: **Tiempo**

Como ya se mencionó en el punto 2.3, esta perspectiva pasará a llamarse dimTiempo, con la clave principal cuyo nombre es Id\_Tiempo. Vamos a mantener el campo mes.

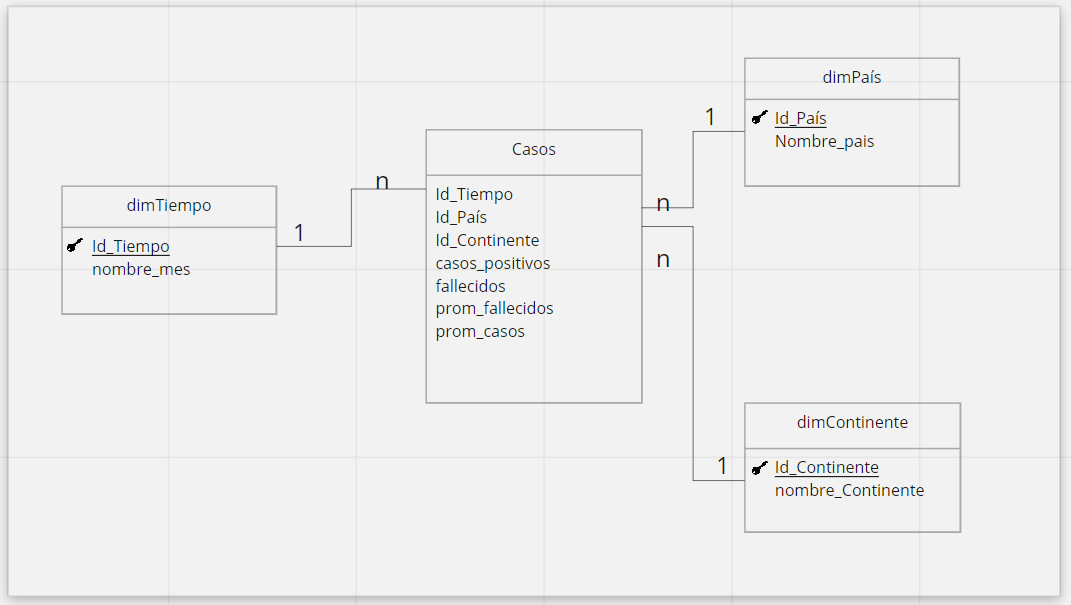
## 3.3) Tablas de hechos

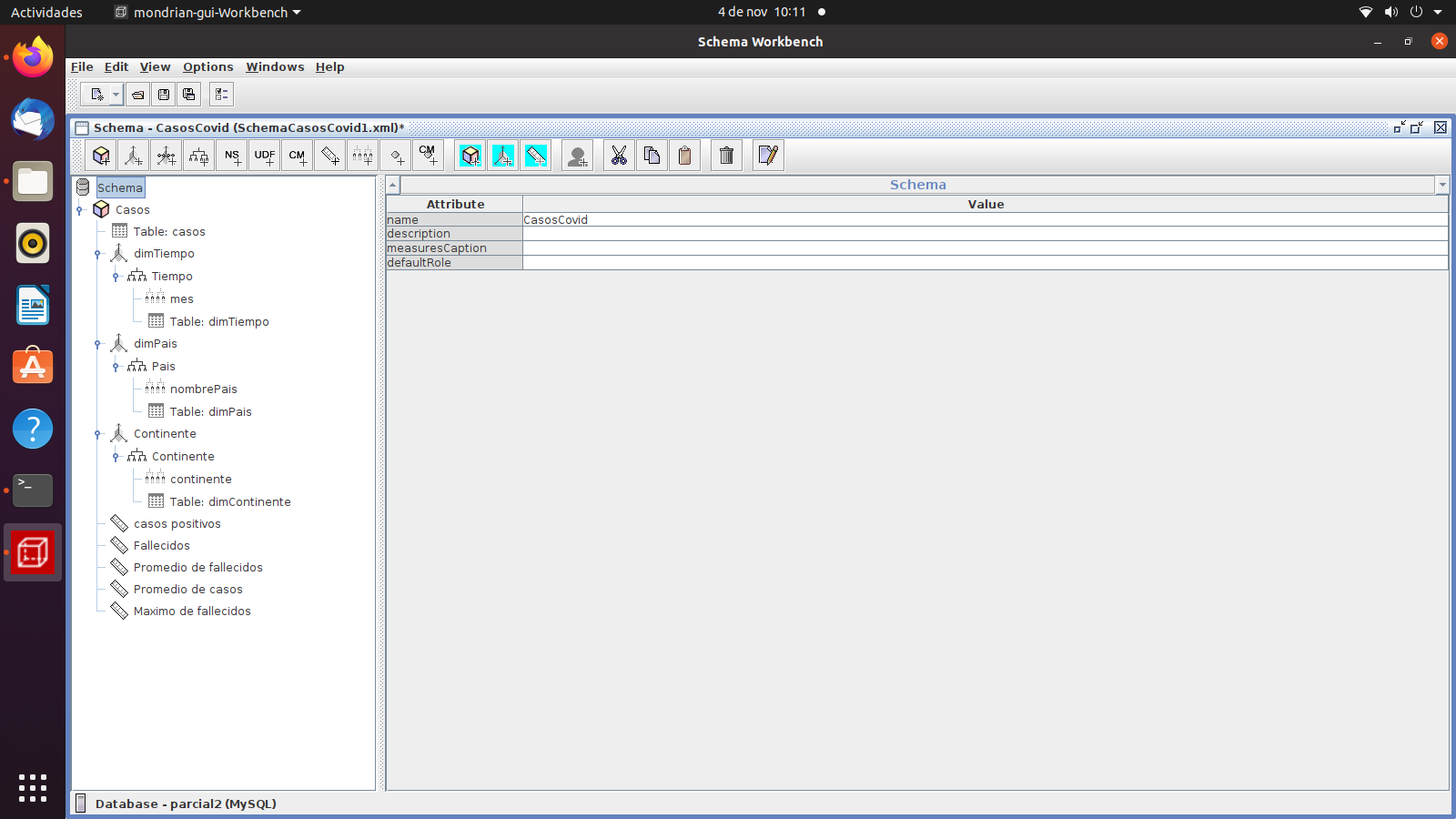
Hemos nombrado nuestra tabla de hechos como Casos. Esta tabla tendrá los atributos definidos en las dimensiones antes mencionadas (ver punto 3.2). Respecto a los hechos, tendremos seis indicadores que son los mencionados en el Modelo Conceptual (ver punto 1.3)



## 3.4) Uniones

Como la palabra lo dice, mostramos acá las uniones entre el hecho definido y las dimensiones creadas.



A continuación, incluímos el cubo multidimensional que hemos creado utilizando Schema Workbench. Aquí, hemos tomado los datos de nuestra base de datos y generamos el cubo al que llamamos Casos. Las tres dimensiones, con una tabla correspondiente a cada una. Esto indica el mapeo entre la base de datos y las dimensiones que hemos definido.

# 4) Bocetos de futuros reportes y tableros

## 4.1) Reportes

*Casos positivos*

Este reporte representará el número de personas que han contraído el virus en un determinado periodo de tiempo permitiendo comparar la situación sanitaria de otros países de acuerdo al número de enfermos de covid-19..

Este reporte da información sobre nuestro indicador “Casos positivos por mes”.

*Fallecidos*

Este reporte representará el número de personas que han contraído el virus y no han sobrevivido, midiendo en un determinado periodo de tiempo permitiendo comparar la cantidad de muertos con respecto a otros países..

Este reporte da información sobre nuestro indicador “Fallecidos por mes por país”

*Picos máximos de afectados*

Este reporte representará el número máximo de personas que han contraído el virus en un determinado periodo de tiempo y que han sobrevivido a la enfermedad. Esto, permitiendo comparar la situación sanitaria de otros países de acuerdo al número de enfermos de covid-19. Estamos tomando en cuenta el año 2020, año sobre el que se realizó el análisis.

Este reporte da información sobre nuestro indicador “Pico máximo de Casos recuperados por país”

*Picos máximos de fallecidos*

Este reporte representará el número máximo de personas que han contraído el virus en un determinado periodo de tiempo y que han sucumbido a la enfermedad. Esto, permitiendo comparar la situación sanitaria de otros países de acuerdo al número de enfermos de covid-19. En este caso, la información será mostrada por mes, dentro del año 2020, año sobre el que se realizó el análisis.

Este reporte da información sobre nuestro indicador “Pico máximo de fallecidos por mes”

*Promedio de afectados*

Este reporte representará el número promedio de personas que han contraído el virus en un determinado periodo de tiempo y han fallecido.

El mismo podrá mostrar, en promedio, las personas que han sobrevidio en un país. Lo mismo se podrá obtener por continente.

Este reporte da información sobre nuestros indicadoress “Promedio de fallecidos por mes”, “Promedio de Casos positivos por país” y “Promedio de casos positivos por continente”.

## 4.2) Tableros

*Panorama general del año 2020*

Con este tablero, se muestra el avance de la enfermedad a nivel mundial durante el año 2020. Aquí, podremos elegir el mes, el país e incluso el continente.

Este tablero da información sobre los indicadores “Casos positivos”, “Fallecidos” y “Recuperados”.

# Conclusión

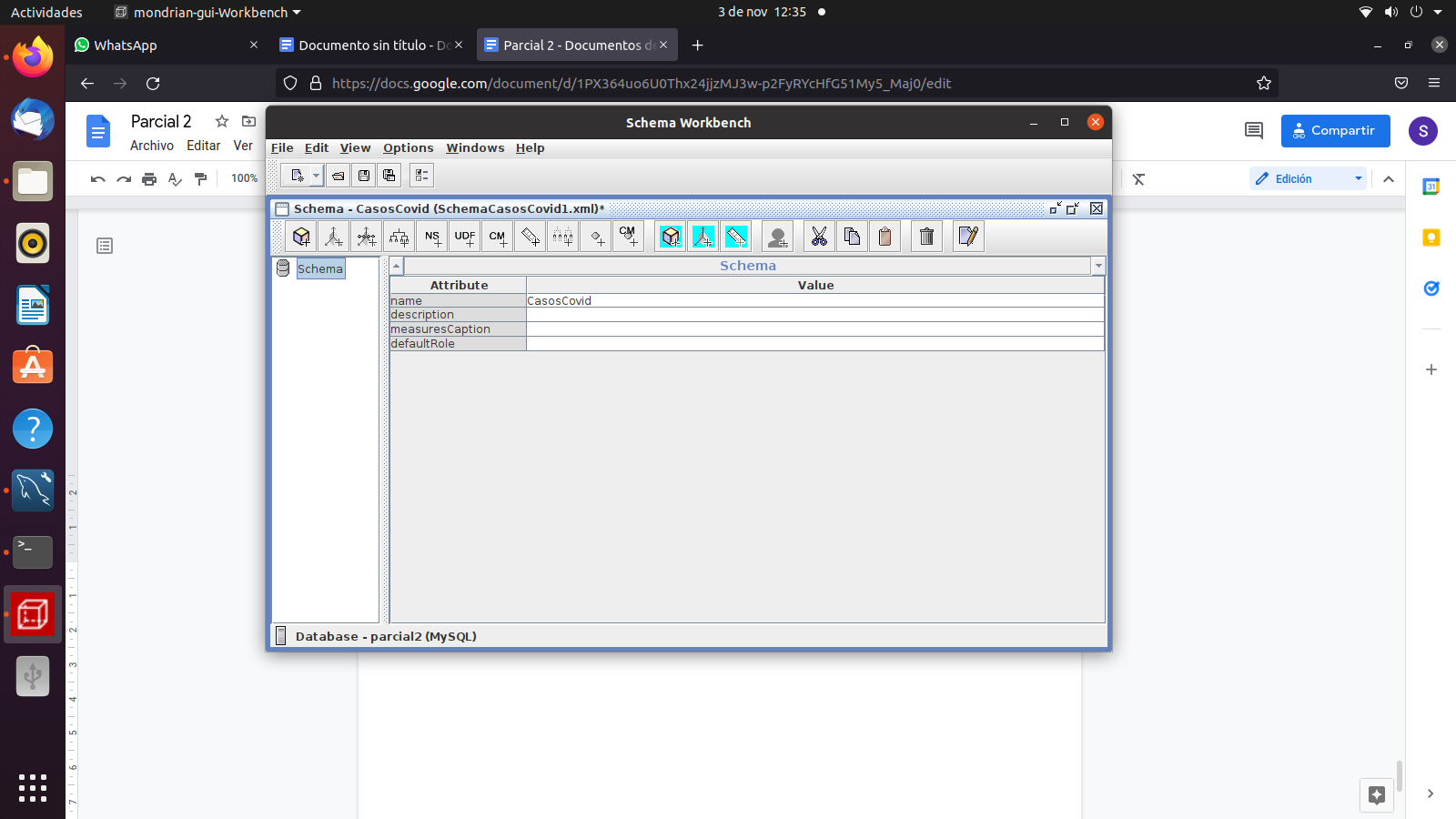
Conexión pentaho:

El que tenemos que conectar ahora a la base de datos, porque finalmente es el que va terminar explotando ese cubo, es el servidor entonces vamos a management data source, vamos a crear una nueva fuente de datos de tipo database tables, nos va a mostrar todas las que existen actualmente pero vamos a crear una nueva. El nombre que nosotros ponemos en connection name, se llama JNDI que es una tecnología de JAva para dar nombres a componentes (por ejemplo a una conexión a una base de datos) y despues para poder cambiar la cuestión básica de los datos de la conexión sin cambiar el nombre. En hostname tenemos que poner el ip gateway de mysql en docker (docker inspect mysql). Ponemos el nombre de la base de datos (parcial2), ponemos el puerto (306), el usuario (root) y la contraseña.

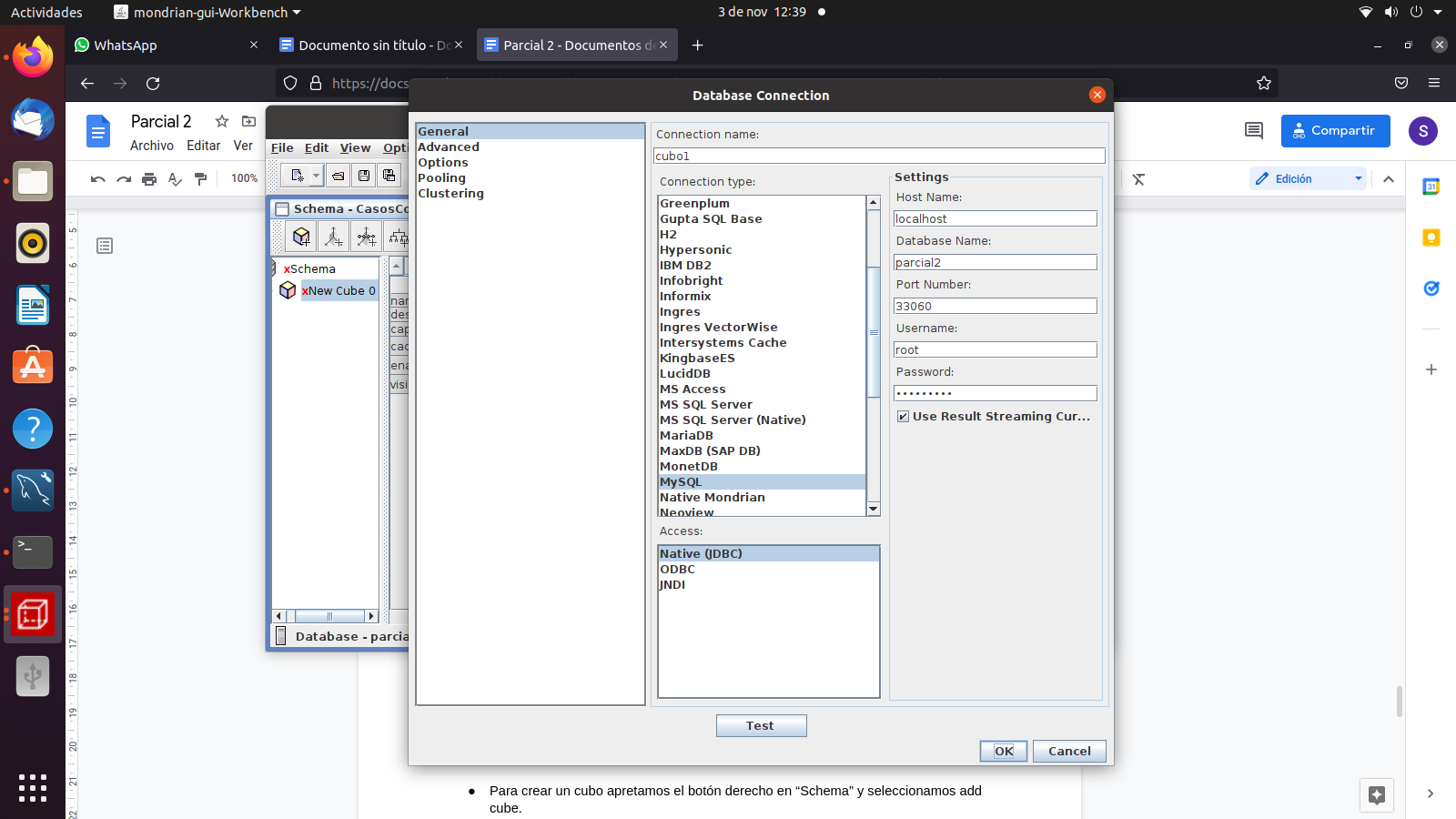
Schema-workbench:

Esto permite crear esquemas que a su vez pueden crear en sí uno o varios cubos, un cubo puede contener dimensiones e indicadores, y una dimensión puede contener una o más jerarquías, y una jerarquía contiene niveles. esto es un resumen del datawarehouse.

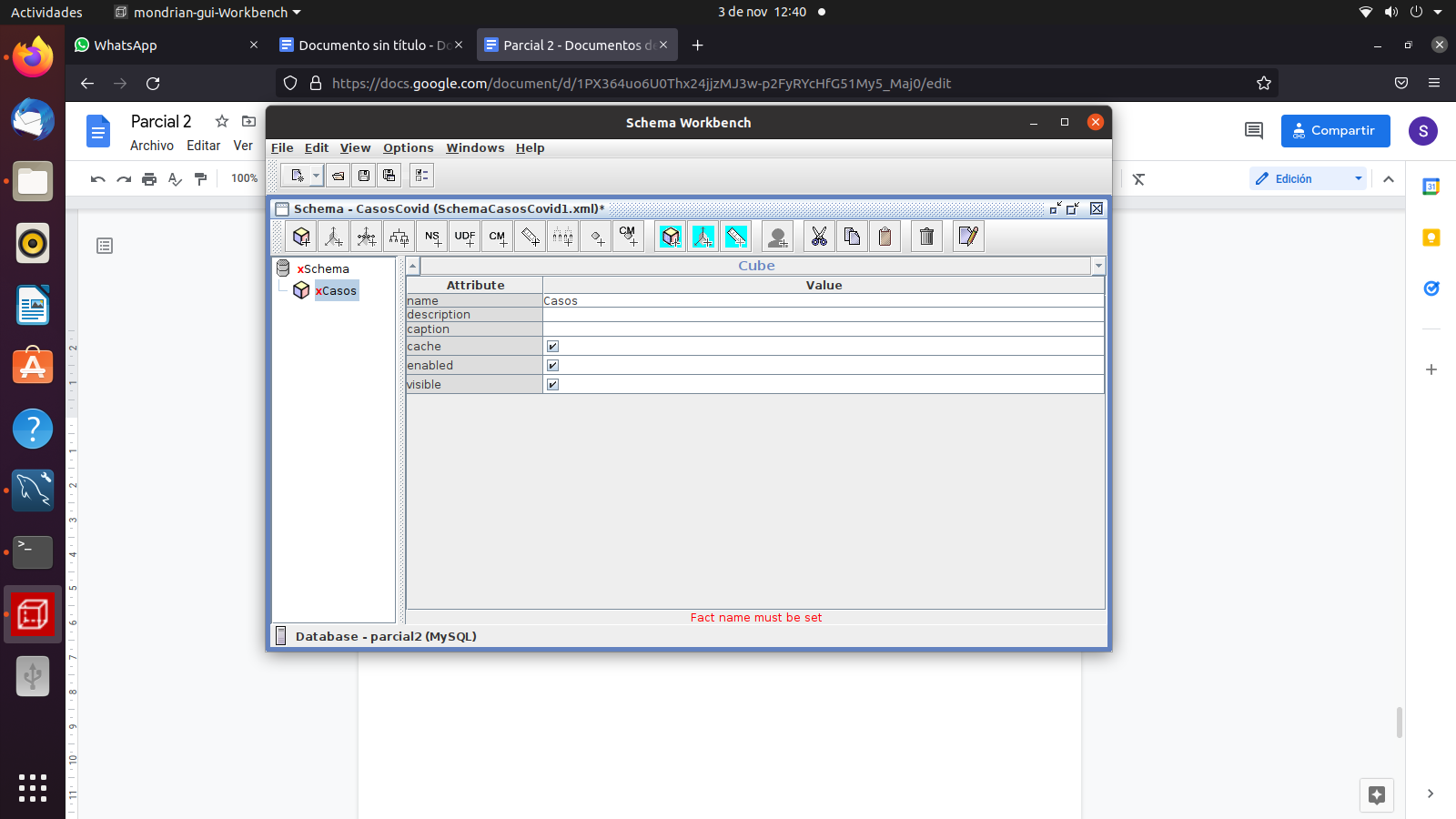
* Procedemos a crear un nuevo esquema
* Se nos abre un editor gráfico que por detrás va a ir creando un archivo xml. Constantemente podemos ver la parte gráfica y el xml. Para saber lo del xml, seleccionamos el “Schema” creado, que simula ser un árbol, en la parte view, seleccionamos view as xml y podemos ver lo que se va haciendo, y para volver vamos a view y ponemos la vista gráfica.

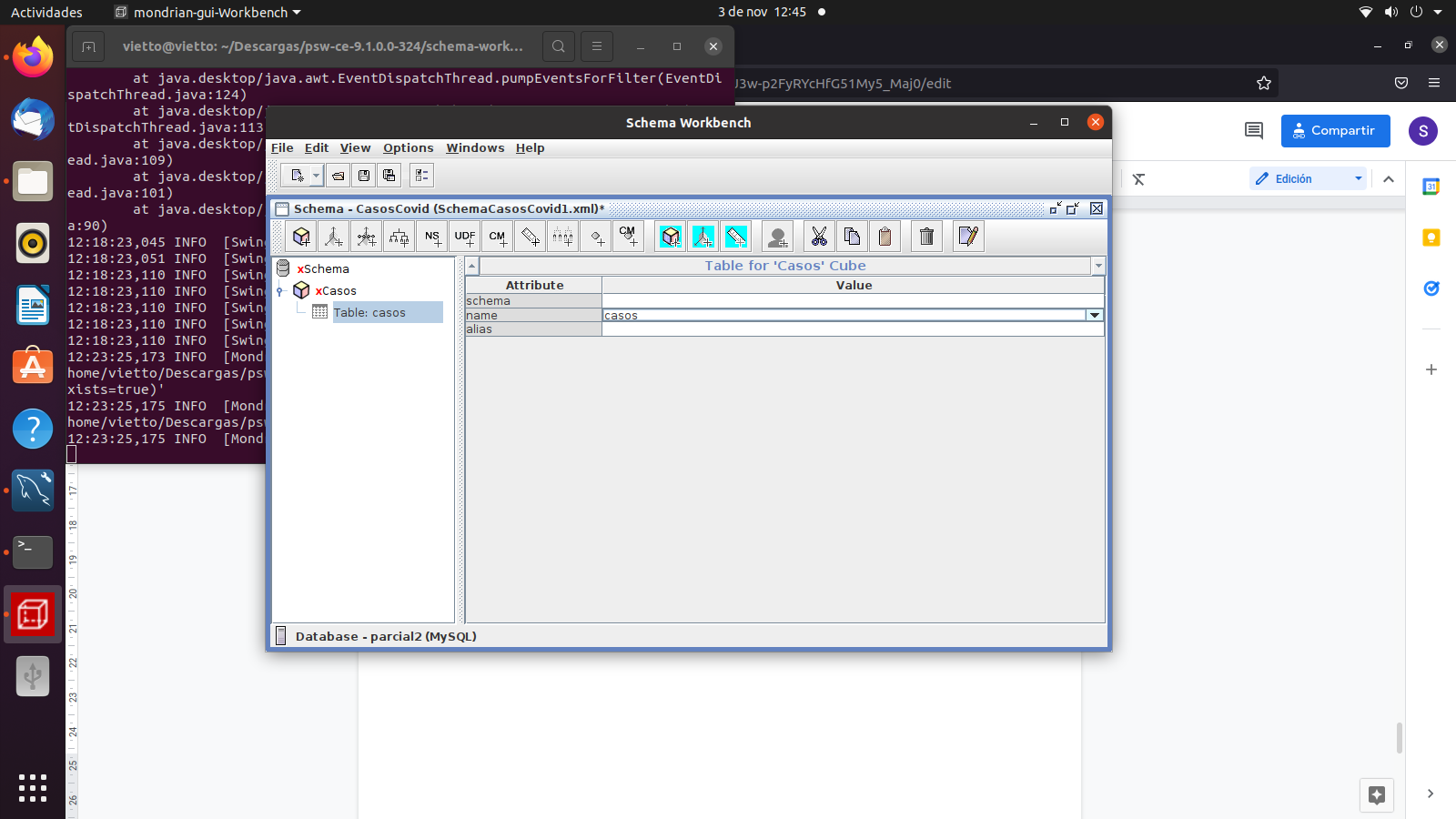
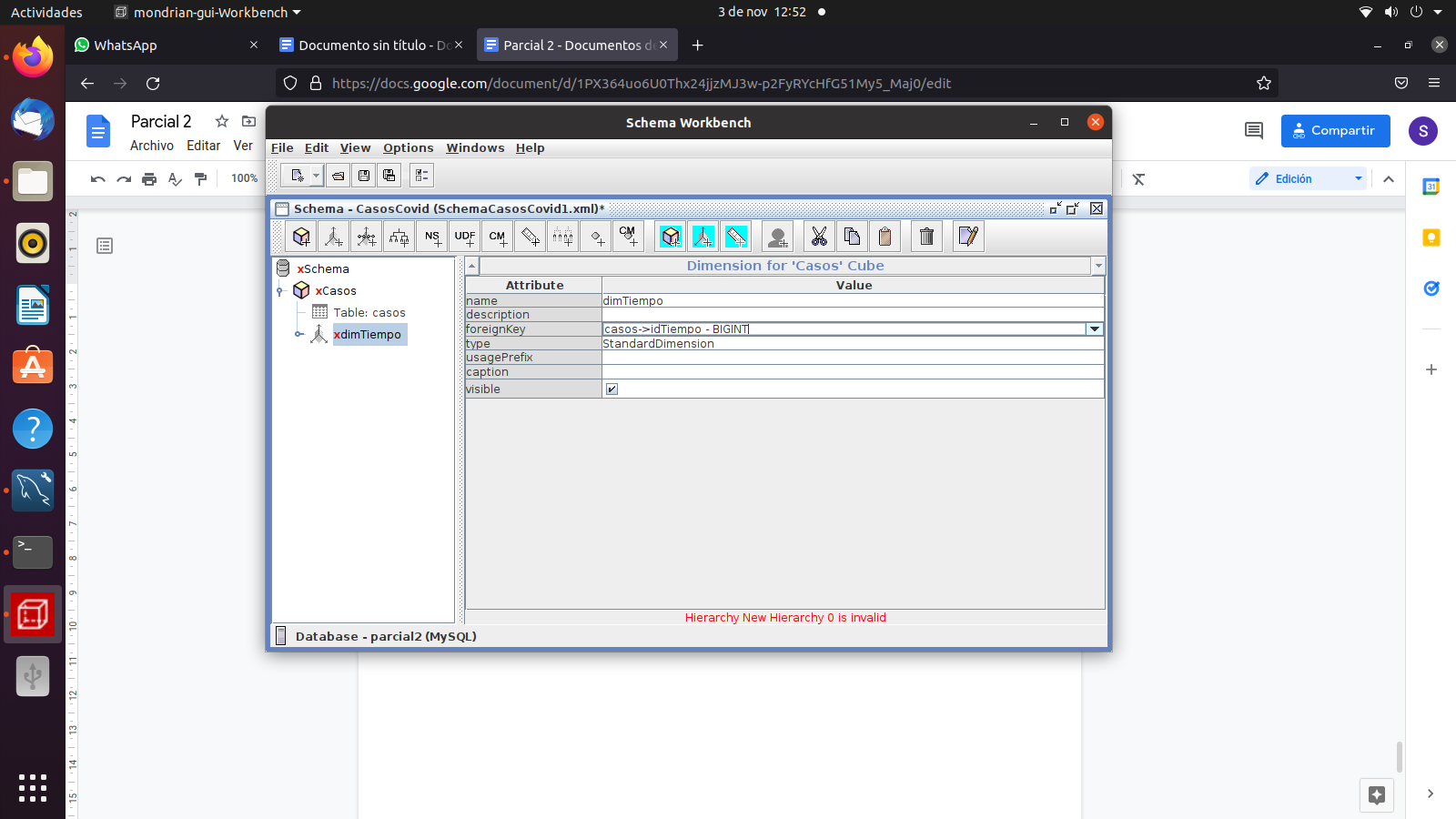


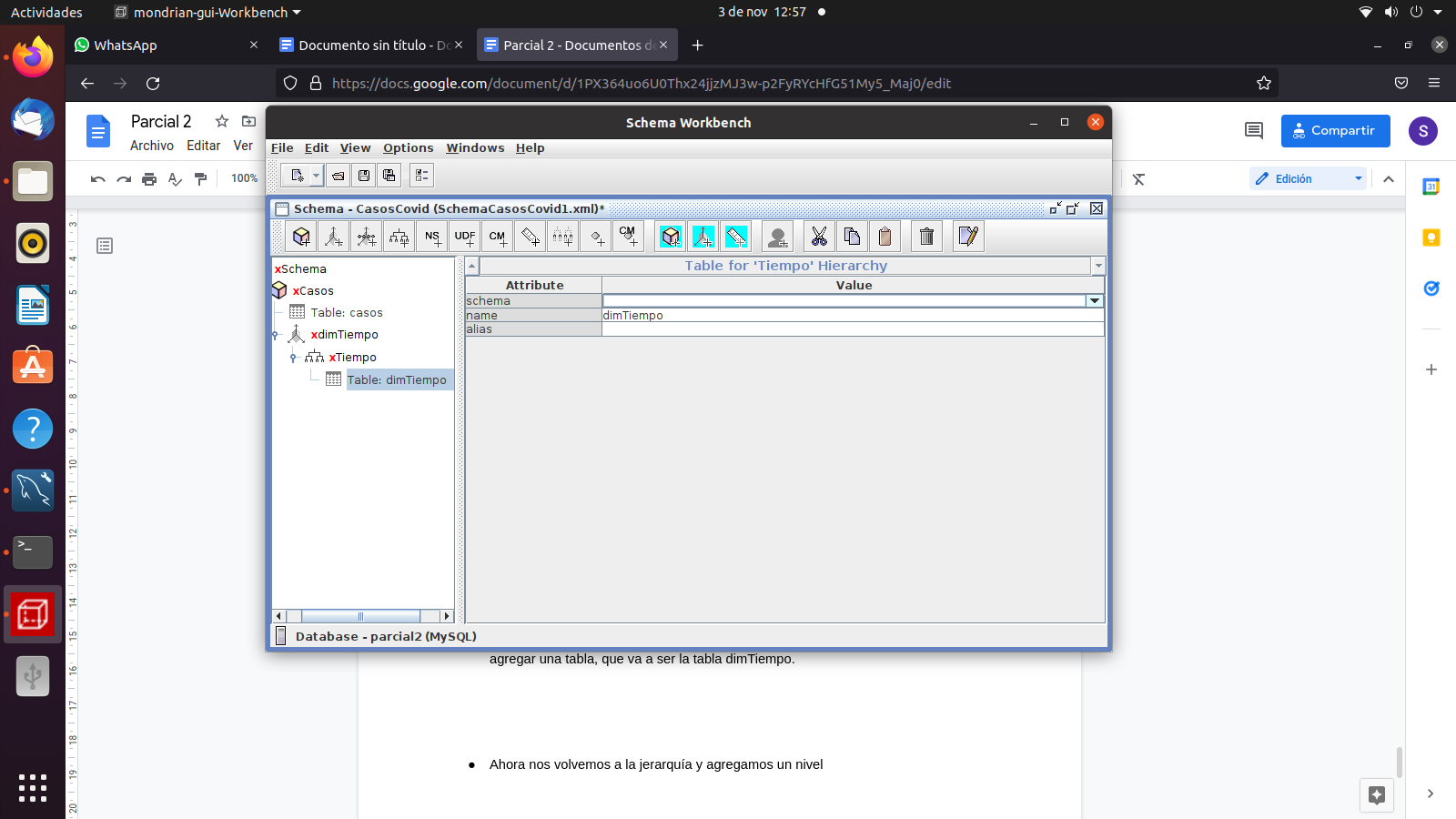
* Cuando apretamos “Schema” nos va a salir el formulario de configuración. Primero tenemos que colocarle un nombre al esquema, y procedemos a guardarlo.
* Tenemos que conectarnos a la base de datos parcial2, de tipo MySQL.

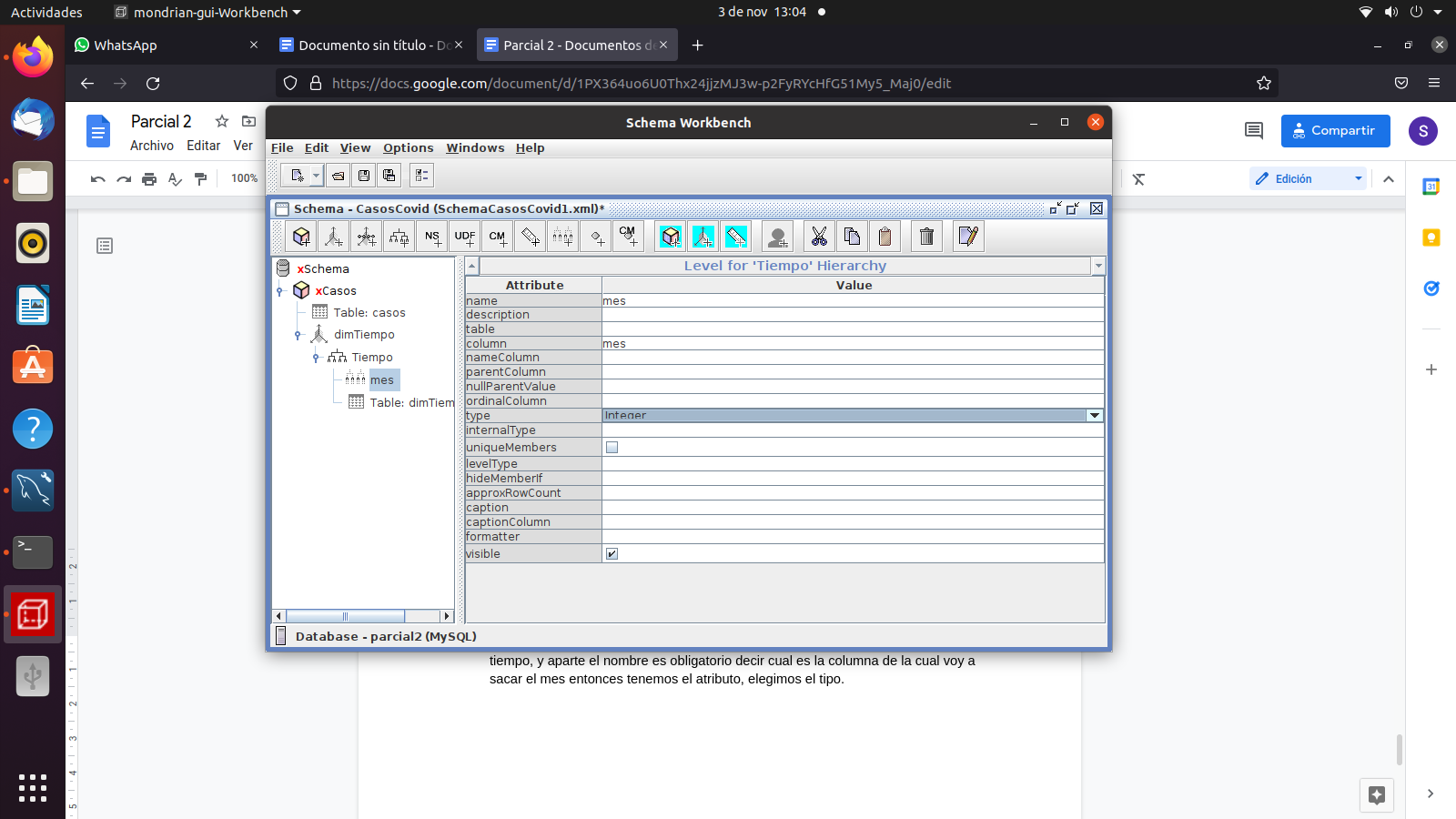


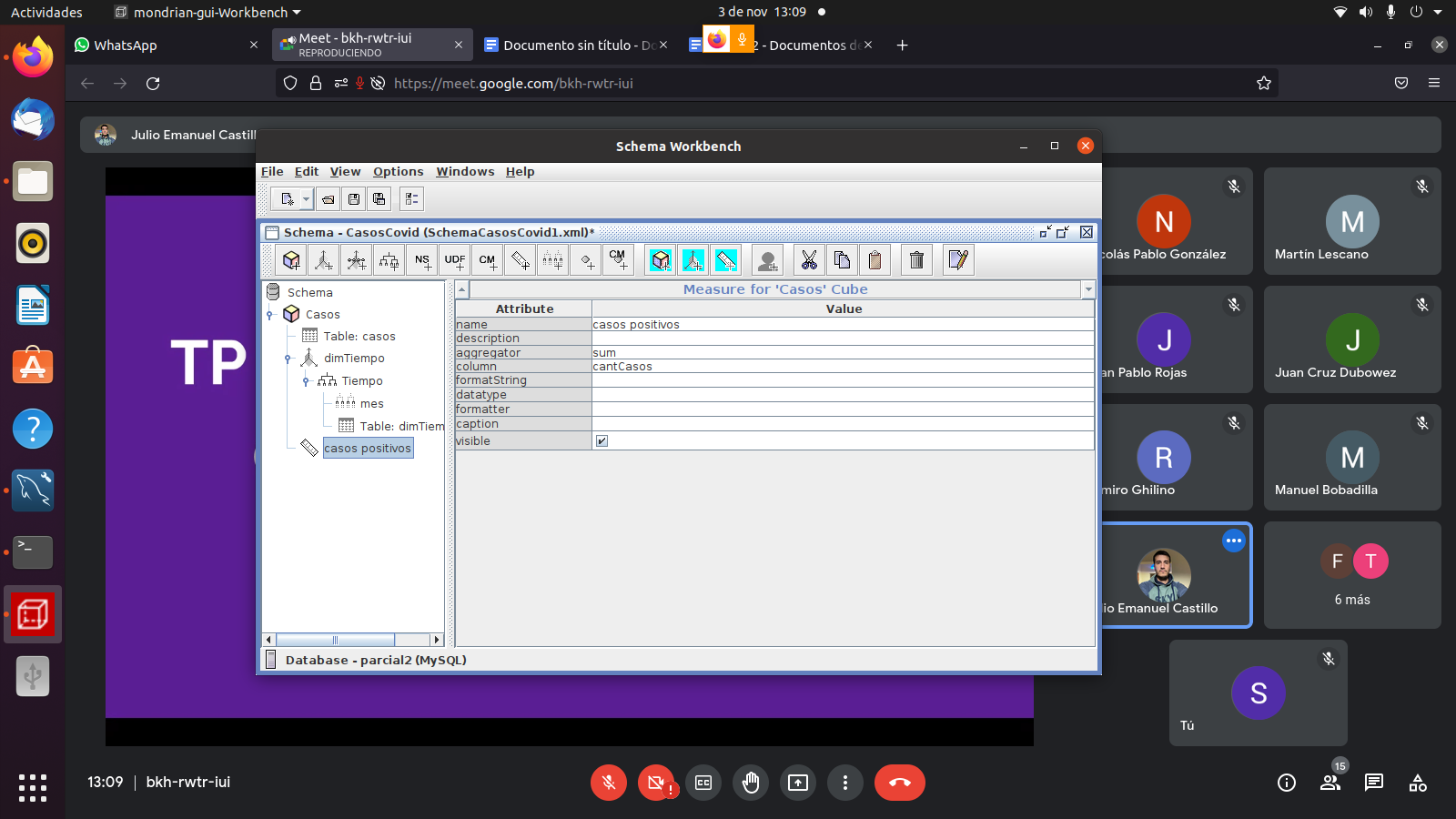
* Para crear un cubo apretamos el botón derecho en “Schema” y seleccionamos add cube.
* Este sistema nos indica con una x roja cada vez que falte rellenar algún dato.
* Tenemos que elegir la tabla de hechos por lo que venimos



* Ahora tenemos que elegir la tabla de hechos, entonces a nivel del cubo hacemos click derecho y seleccionamos add table. Simplemente en el nombre de la tabla tenemos una lista desplegable que nos va a mostrar todas las tablas disponibles de esa conexión. seleccionamos casos porque es la tabla de hechos, pero todavía nos falta agregar al menos una dimensión.
* Ahora procedemos a agregar las dimensiones, haciendo click derecho en el cubo y seleccionamos add dimension. le colocamos el nombre y es obligatorio decir cual es la clave foránea que conecta con la tabla de hechos
* Necesita tener al menos una jerarquía, el programa nos crea una por defecto pero a esta le falta al menos un nivel. cada jerarquía está asociada a una tabla así que antes de agregarle un nivel le vamos a agregar una tabla porque no vamos a poder agregar una jerarquía si no tenemos una tabla.
* Entonces, en la jerarquía le cambiamos el nombre y hacemos click derecho para agregar una tabla, que va a ser la tabla dimTiempo.



* Ahora nos volvemos a la jerarquía y agregamos un nivel, que va a ser en si el tiempo, y aparte el nombre es obligatorio decir cual es la columna de la cual voy a sacar el mes entonces tenemos el atributo, elegimos el tipo.
* Ahora procedemos a agregar los indicadores, con la función de agregación, es decir, que vamos a hacer con los casos, y seleccionamos las columnas.



* Y de esta forma creamos el cubo con una dimensión, lo guardamos y luego lo exportamos al servidor pentaho.