****

**Base de Datos 2**

**- Final -**

**Casos de COVID-19 durante el año 2020**

**Profesor**: Ing. Mariano, GARCÍA MATTÍO

**Integrantes:** Casermeiro Maria Silvia - 2013430@ucc.edu.ar

Chiletti Emanuel - 1921260@ucc.edu.ar

Vietto Santiago - 1802890@ucc.edu.ar

**Índice**

[**Introducción**](#_9nntd3nwtqmw) **3**

[**1) Análisis de requerimientos**](#_fowgxn77w4k6) **4**

[1.1) Preguntas de negocio](#_1ft64ib35d1) 4

[1.2) Indicadores y perspectivas](#_v5tya9he59ta) 4

[1.3) Modelo conceptual](#_7syfbqeewcdg) 6

[**2) Análisis de los OLTP**](#_3g81kuls9rtg) **6**

[2.1) Conformar indicadores](#_bdqwyv71zp6k) 6

[2.2) Establecer correspondencias](#_sb3qgbane8bm) 7

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

[**3) Modelo lógico de DW**](#_na2pslsmmb8k) **11**

[3.1) Tipo de Modelo Lógico del DW](#_4k8iemvjkjgp) 11

[3.2) Tablas de dimensiones](#_eyg8cn8nz5b6) 11

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

[3.4) Uniones](#_llvzwd7y8l1q) 12

[**Cubo multidimensional**](#_3g9lnvg58jkj) **12**

[**4) Bocetos de futuros reportes y tableros**](#_trqx3kwjdyhr) **18**

[4.1) Reportes](#_19bwbpvn6r0) 18

[4.2) Tableros](#_b09f16punvvd) 19

[**Conclusión**](#_74wwm3w7jex2) **21**

# 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 promedio de casos positivos por continente: *¿Cuál es el promedio de casos por continente?*
* Queremos informar la cantidad de hisopados positivos en un determinado periodo de tiempo por cada continente: *¿Qué cantidad de hisopados positivos hubo por tiempo por continente?*
* Queremos informar la cantidad de fallecidos en un determinado periodo de tiempo: *¿Qué cantidad de fallecidos hubo por mes tiempo por continente?*
* 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?*

## 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 continente?*    Indicador perspectiva |
| --- |
| *¿Qué cantidad de fallecidos hubo por mes tiempo por continente?*    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
  + fallecidos
  + pico máximo de fallecidos
  + Promedio de casos positivos
  + Promedio de fallecidos

Hemos identificado a “casos positivos” 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.

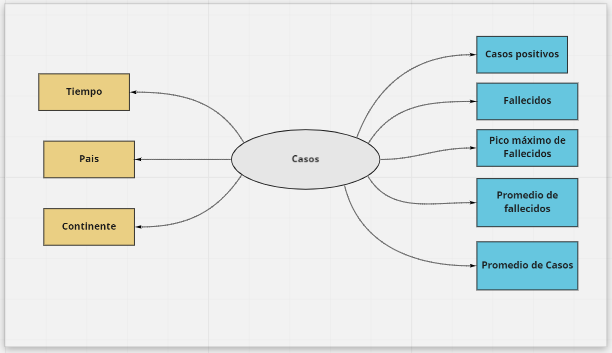
Tiempo: permite obtener respuestas concretas sobre el mes en particular

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.



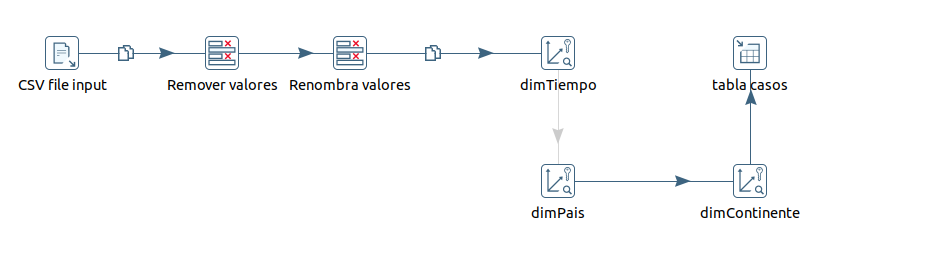
# 2) Análisis de los OLTP

## 2.1) Conformar indicadores

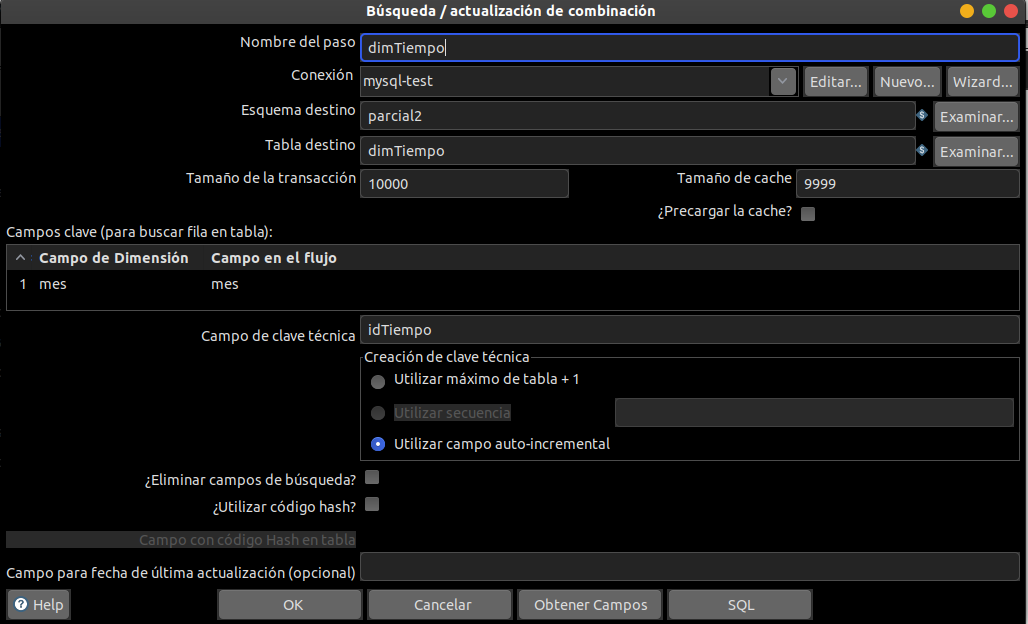
* “Cantidad de casos positivos”
  + Hecho: casos positivos
  + Función de sumarización: SUM
* “Cantidad de fallecidos”
  + Hecho: personas fallecidas
  + Función de sumarización: SUM
* “Pico máximo de fallecidos”
  + Hecho: casos positivos
  + Función de sumarización: MAX
* “Promedio de casos positivos”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

## 2.2) Establecer correspondencias

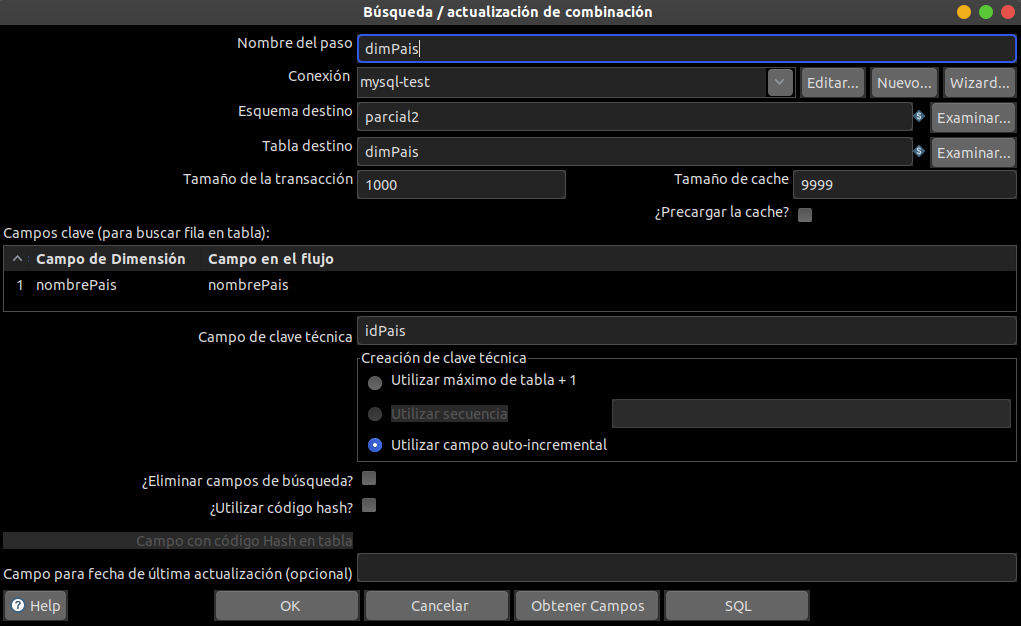
Proceso ETL:



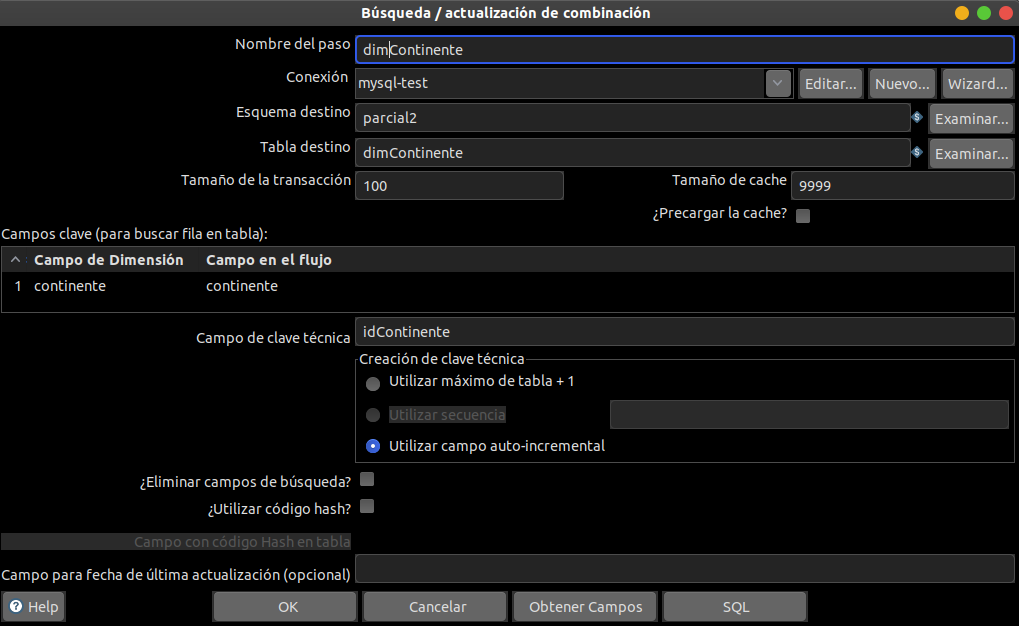
Carga dimensión Tiempo:



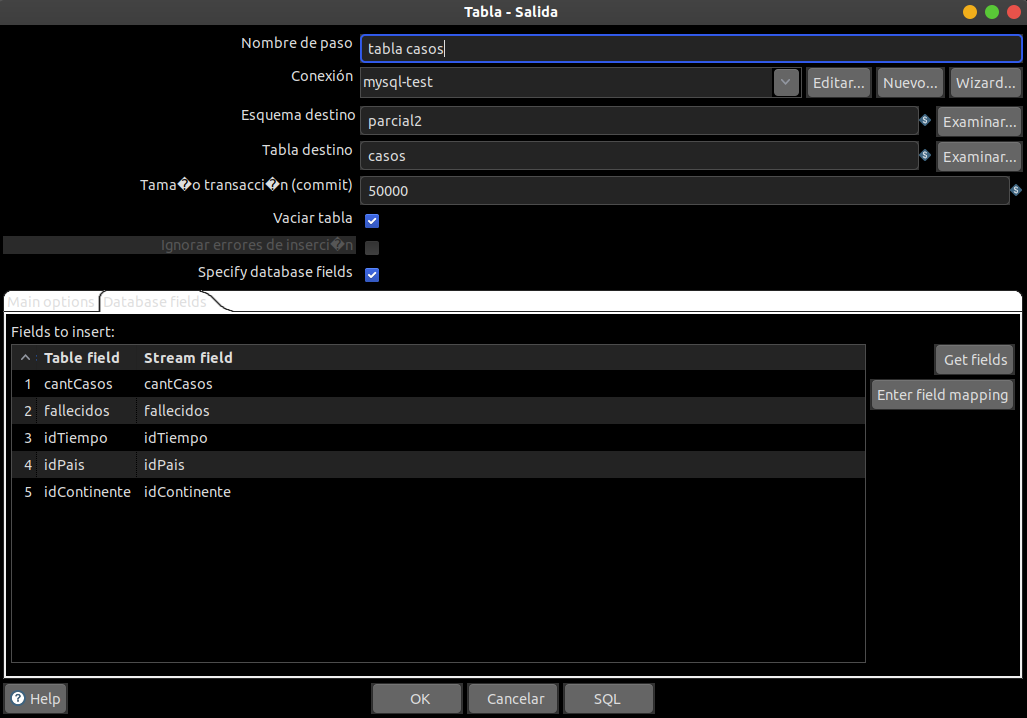
Carga dimensión País:



Carga dimensión Continente:



Carga de tabla de hechos casos:



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 |
| --- | --- |

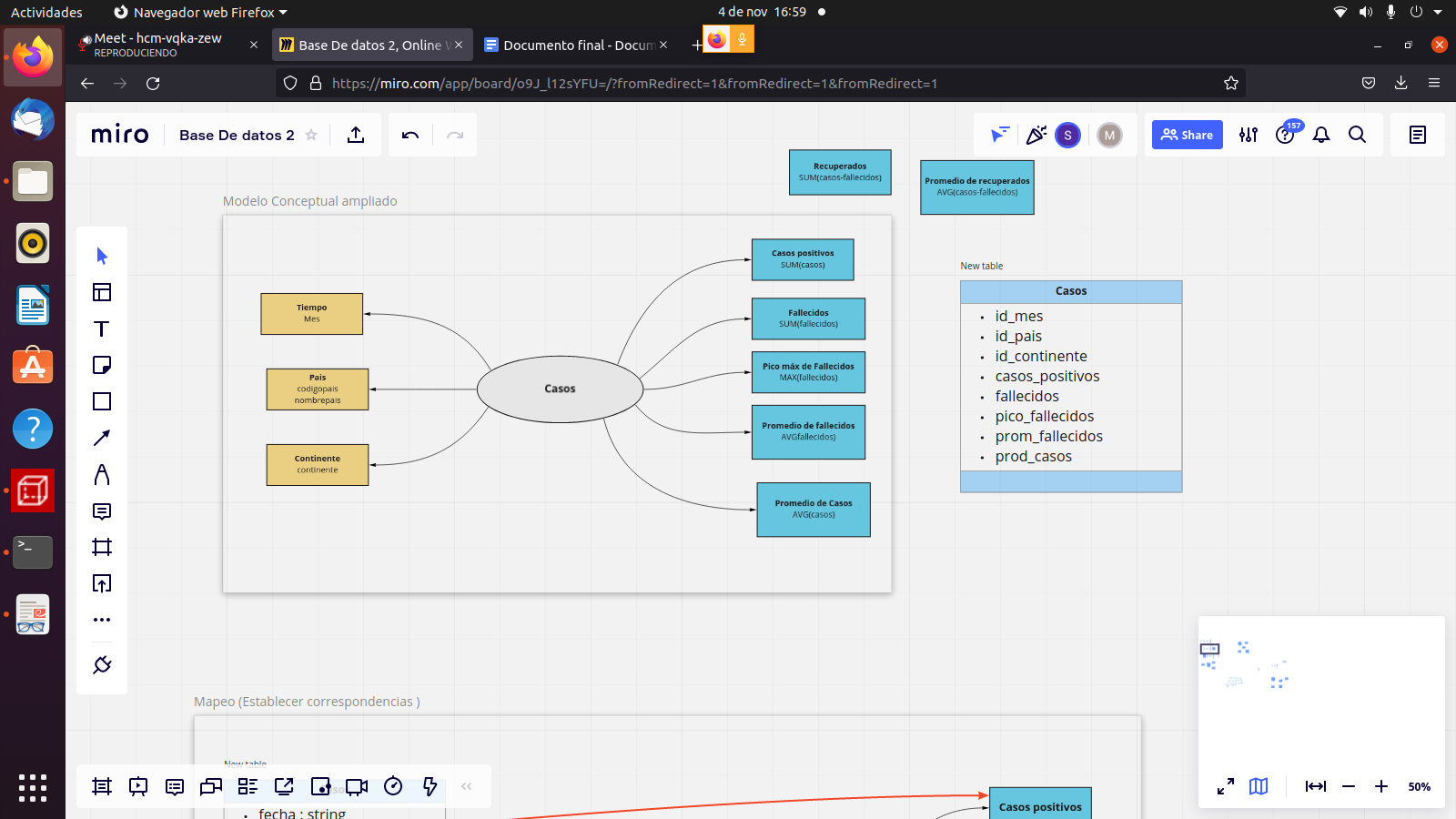
**Continente**,

| 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**

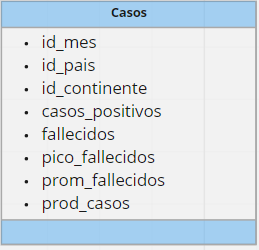
Como ya se mencionó en el punto 2.3, esta perspectiva pasará a llamarse dimContinente, con la clave principal cuyo nombre es Id\_Continente. Vamos a mantener el campo nombre\_Continente

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.

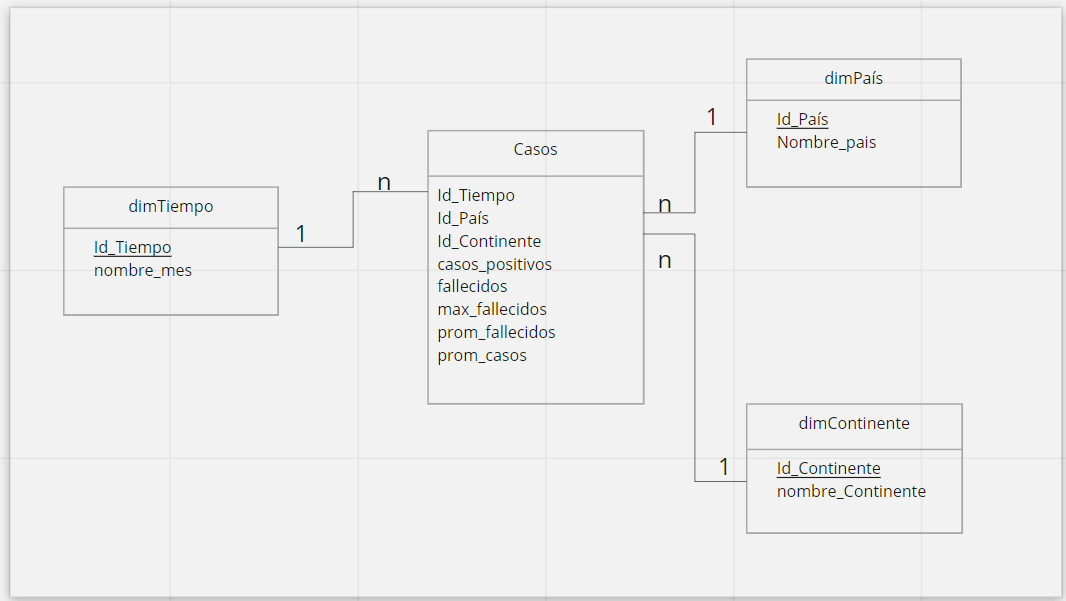
## 3.3) Tablas de hechos

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

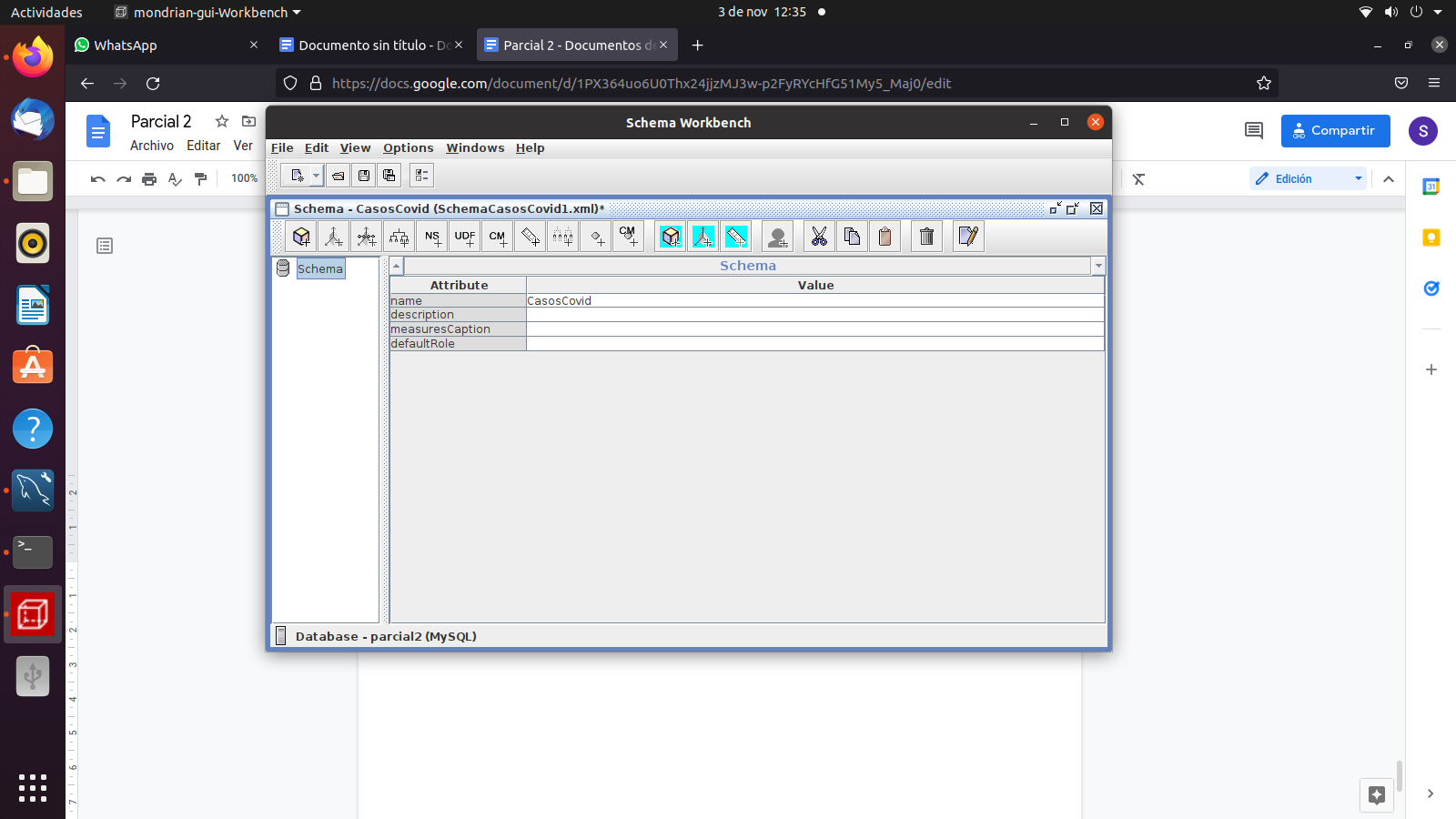
Como la palabra lo dice, mostramos acá las uniones entre el hecho definido y las dimensiones creadas.

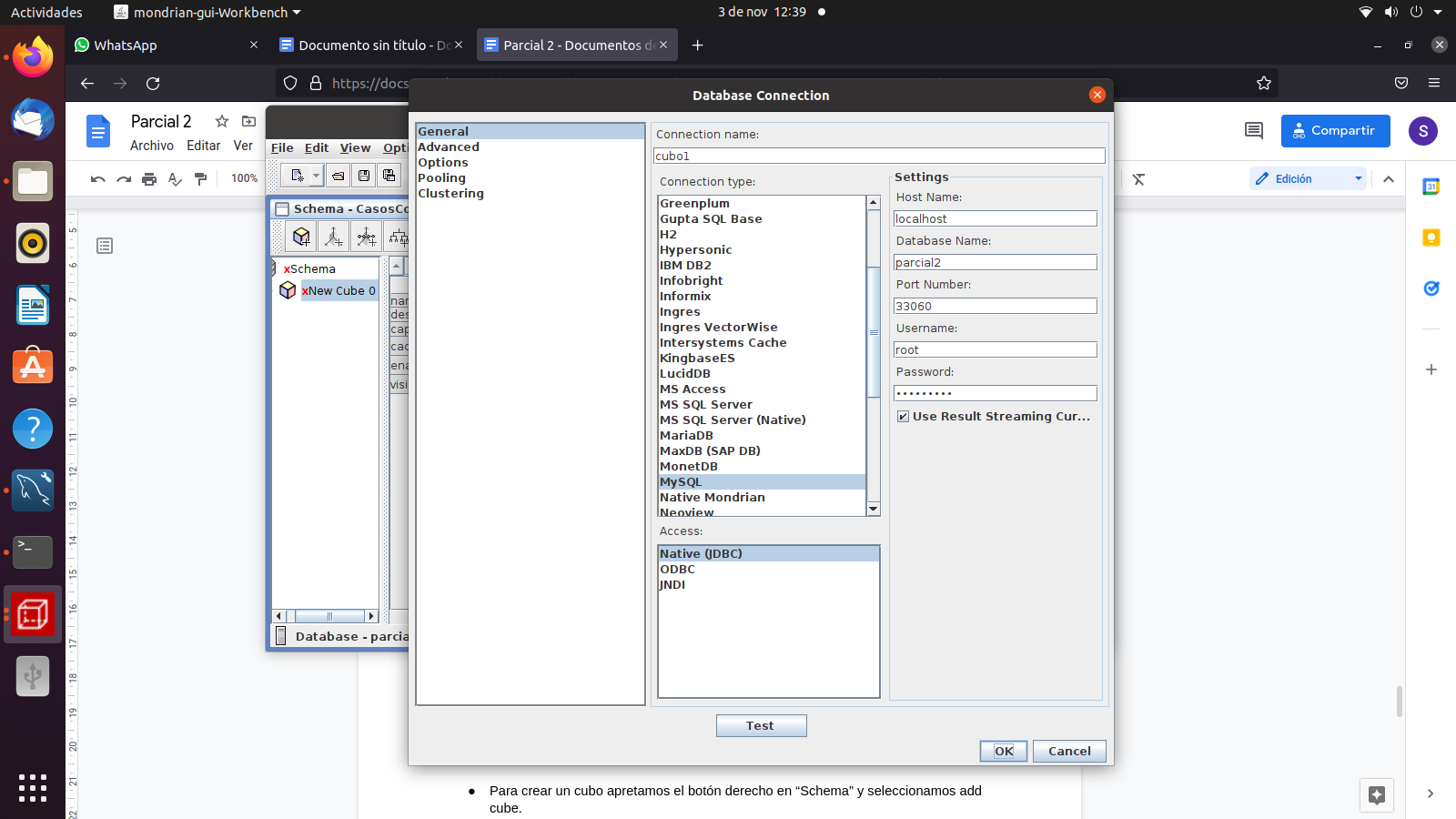
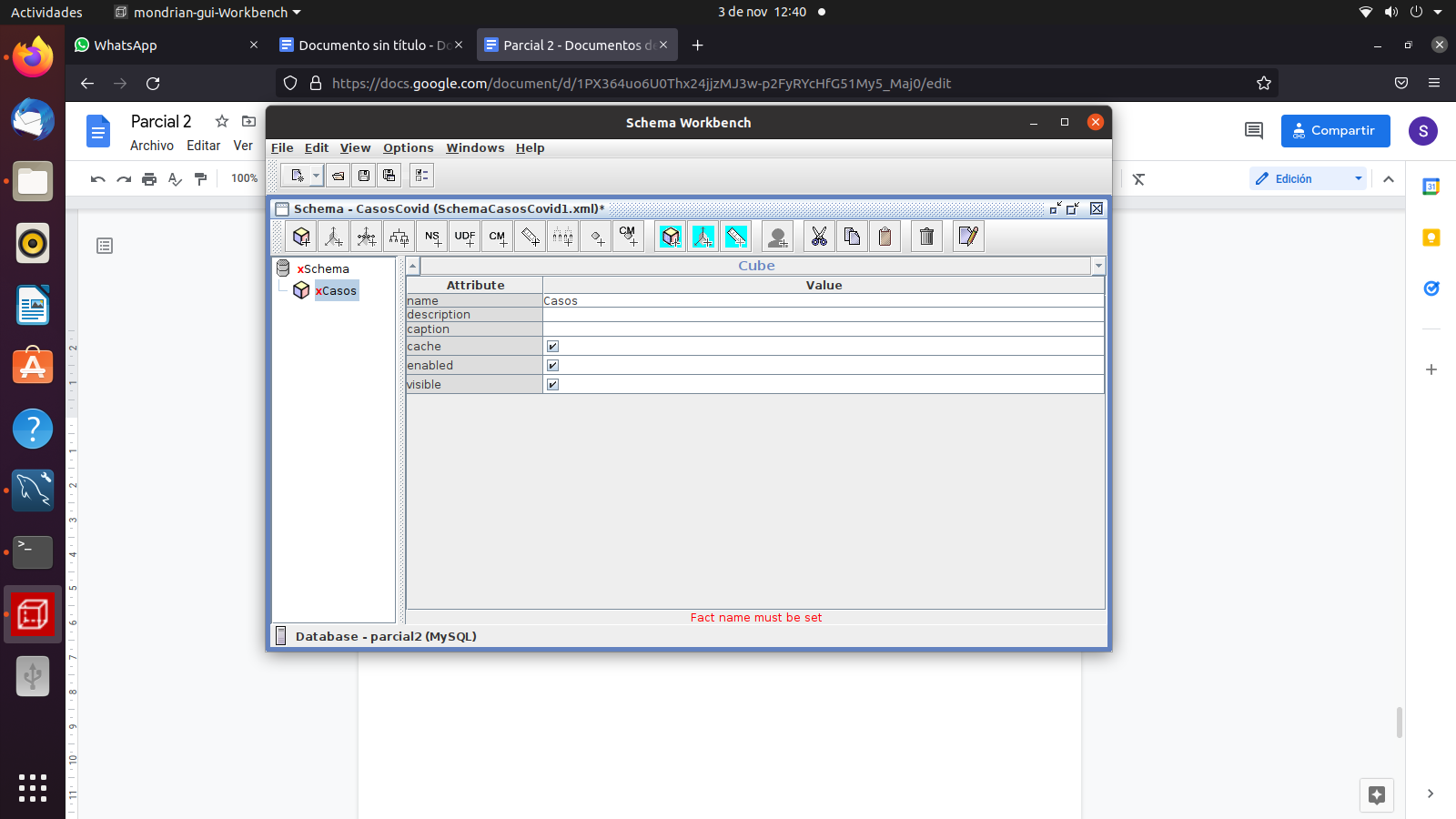
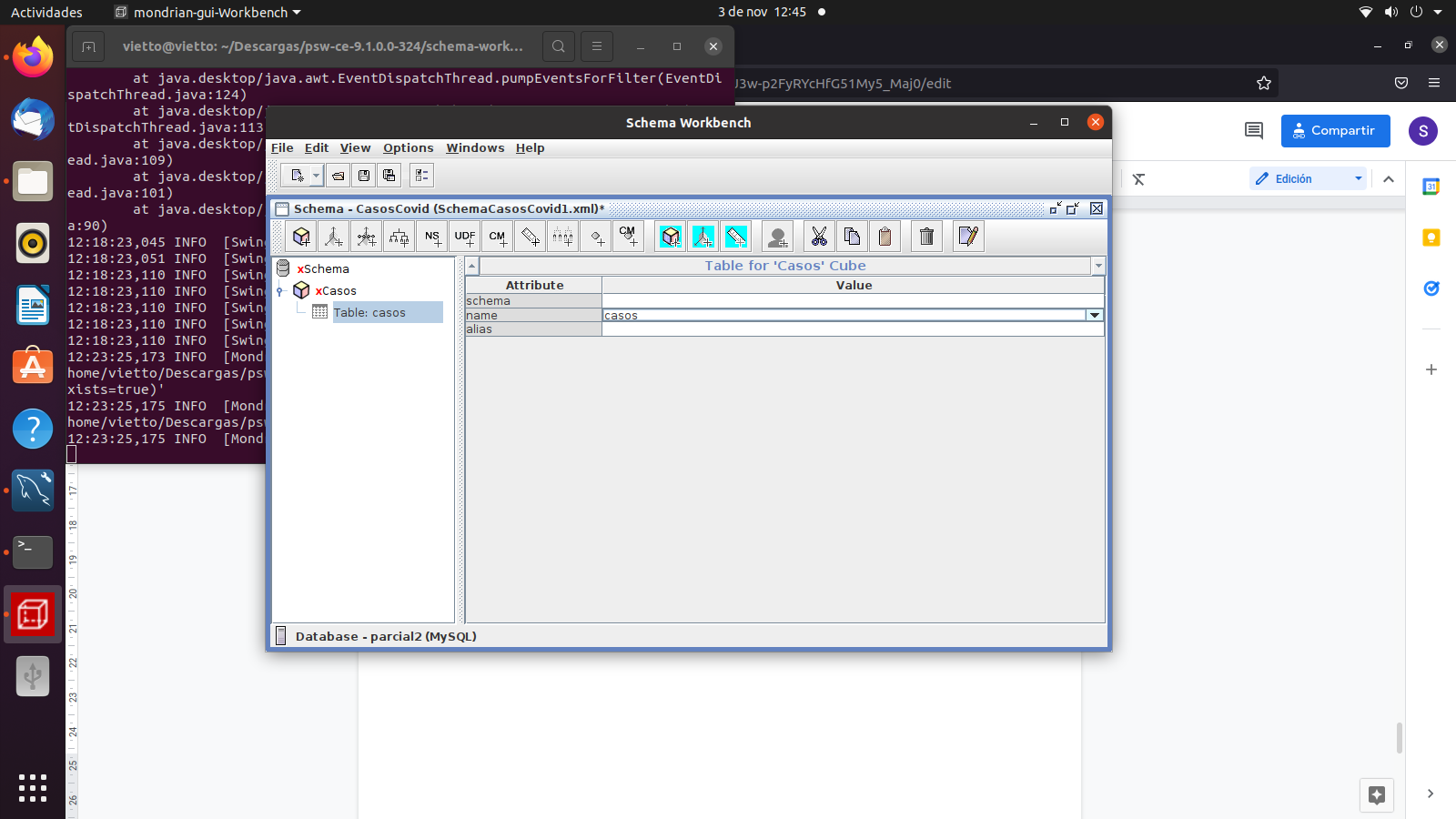
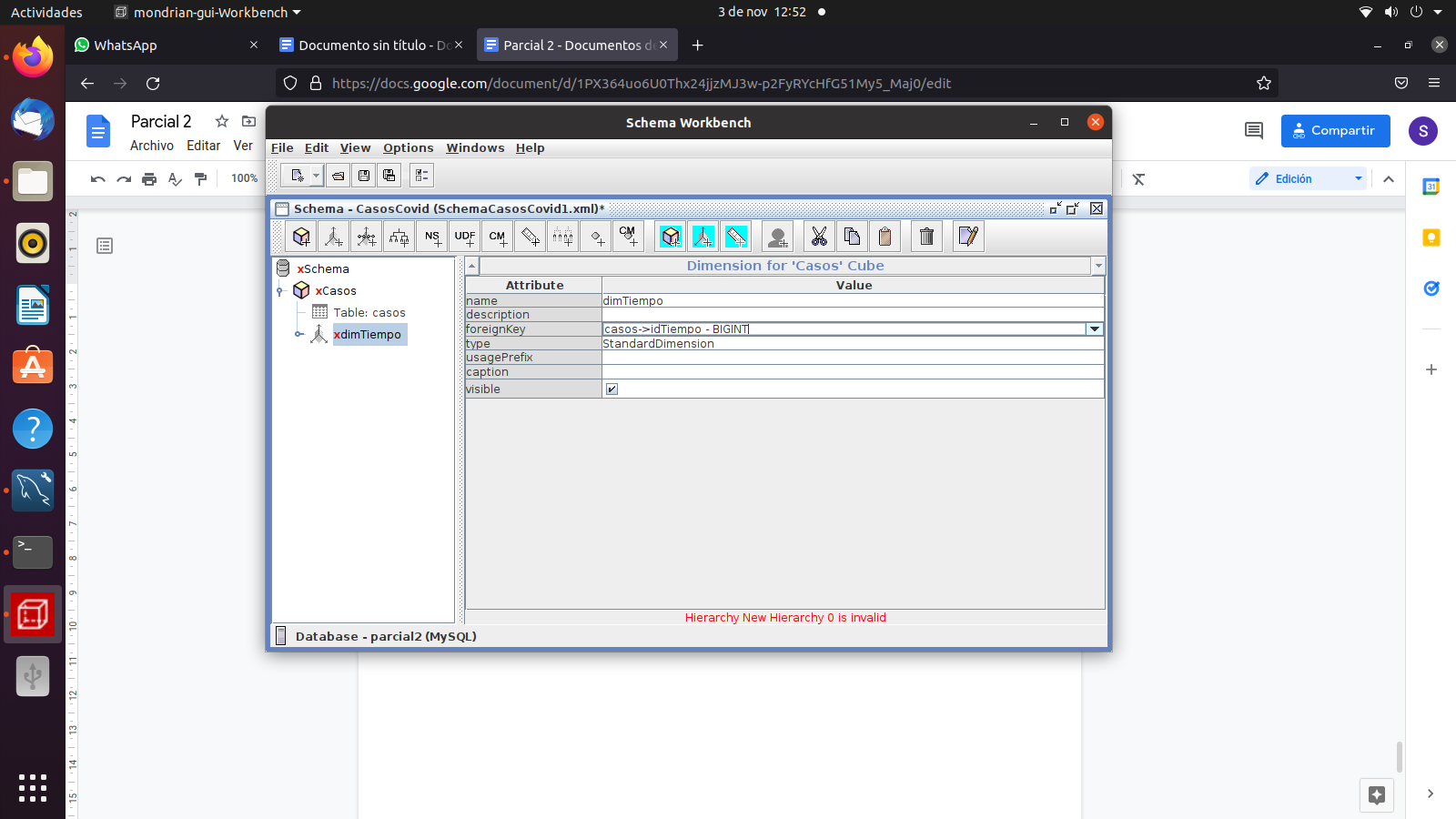
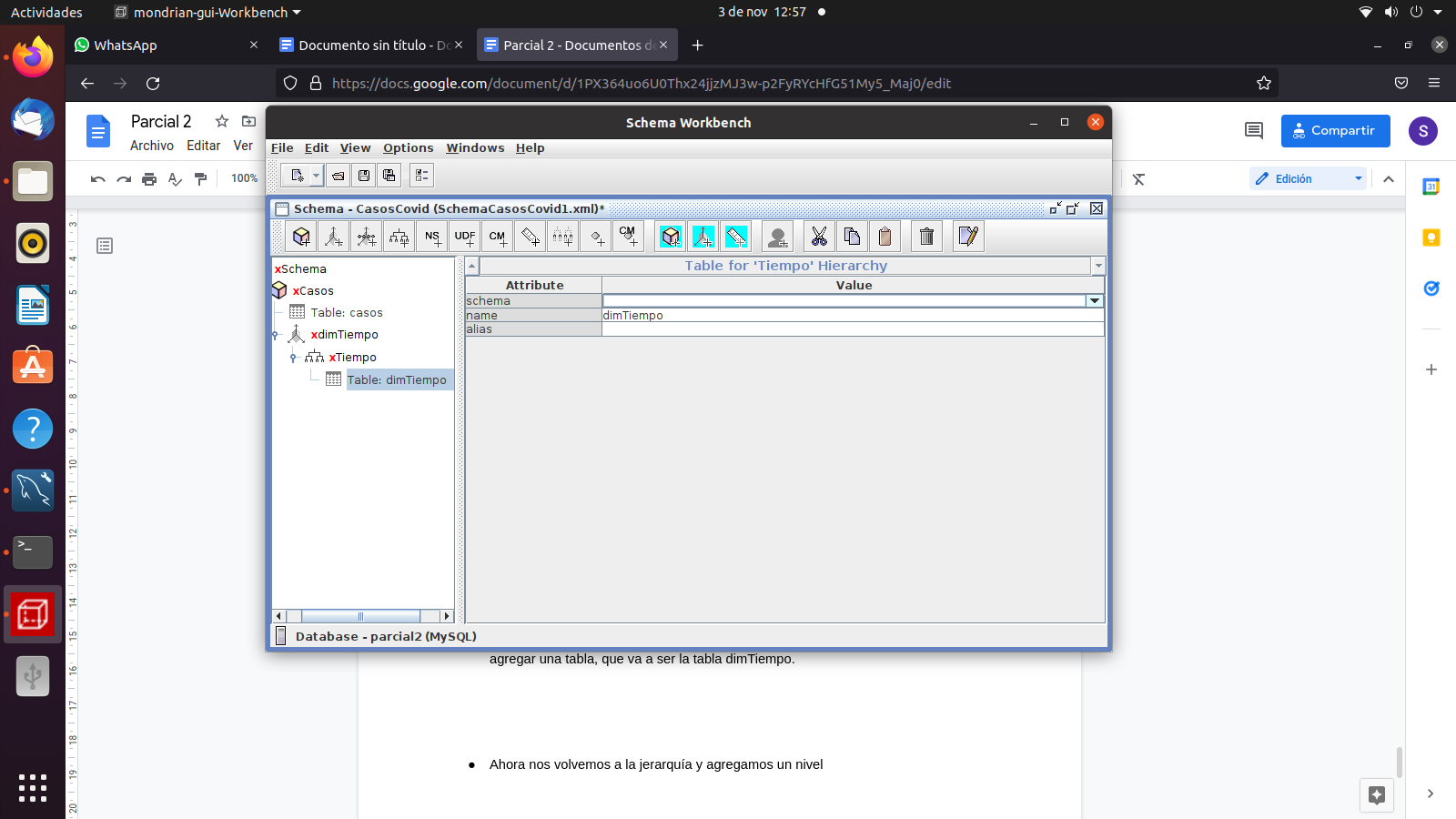
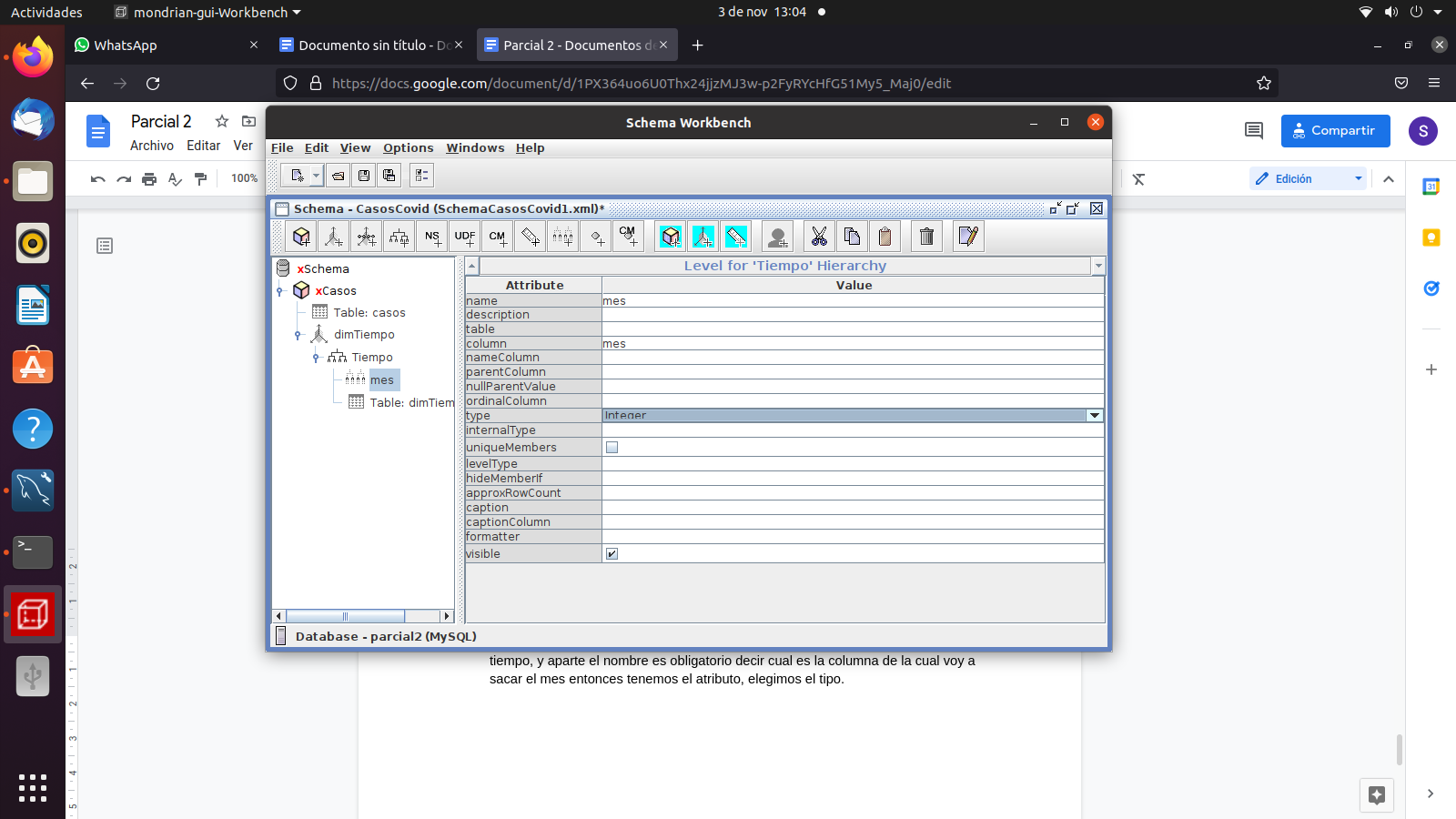
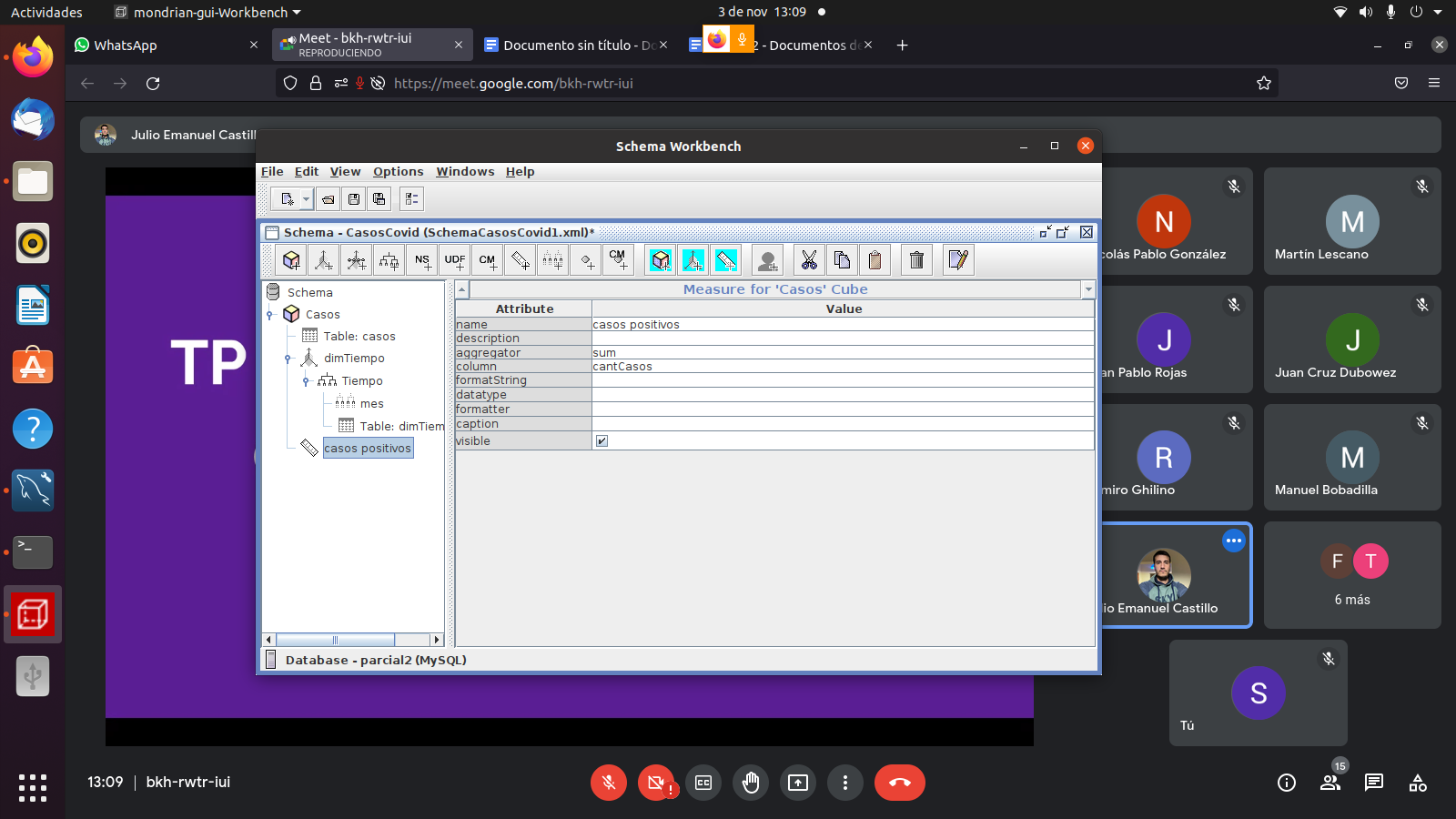


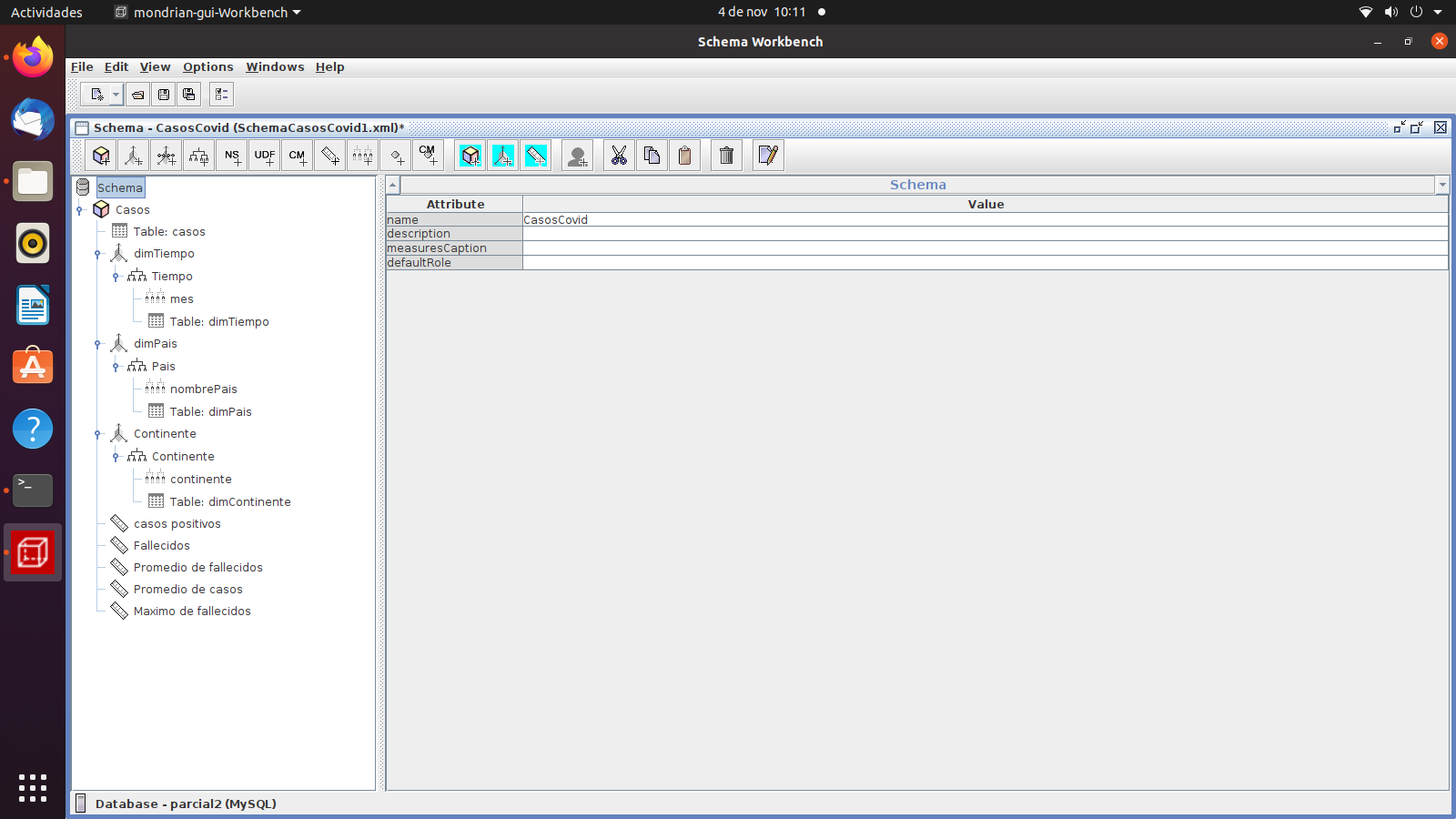
# Cubo multidimensional

Para desarrollar el cubo dimensional utilizamos la herramienta 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 con “new schema”.
* Se nos abre un editor gráfico que por detrás va a ir creando un archivo xml (lenguaje marcado extensible, el cual consiste en un archivo de texto sin formato que utiliza una serie de etiquetas personalizadas con la finalidad de describir tanto la estructura como otras características del documento). Constantemente vamos a poder ir viendo la parte gráfica y el XML. Cada vez que le hacemos click al schema creado donde se va a ir armando un árbol, y si le hacemos click a un nodo del árbol que va a estar representando un componente de ese esquema, 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 que es la vista normal.



* 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, vamos a “options” -> “connections”, seleccionamos de tipo MySQL y completamos los datos.
* 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 o haya un error.
* 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, en donde si estamos bien conectados a la base de datos, 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 cuál es la clave foránea (de la tabla de hechos) que conecta justamente con la tabla de hechos.
* Como 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 un nivel sino agregamos una tabla ya que de esta última van a salir la columna del nivel jerárquico.
* Entonces, en la jerarquía le cambiamos el nombre y hacemos click derecho para “add tabla”, que va a ser la tabla dimTiempo.
* Ahora nos volvemos a la jerarquía y agregamos un nivel, que va a ser en sí 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.
* Un cubo tiene que tener por lo menos una medida, que son los indicadores, porque básicamente un cubo está compuesto por los indicadores y dimensiones. Procedemos a agregar un indicador, apretamos click derecho en el cubo y apretamos “add measure”. Le damos un nombre, seleccionamos la función de agregación, es decir, que vamos a hacer con los casos, hechos o cantidades (las vamos a sumar, calcular el valor medio, el máximo, etc), y seleccionamos las columnas.
* Y de esta forma creamos el cubo con una dimensión, entonces, procedemos a agregarle las demás dimensiones, para luego guardarlo y exportarlo al servidor pentaho. En el schema workbench, vamos a “file” -> “publish”, colocamos la dirección del servidor con la que estamos trabajando. Lo publicamos y si esta todo bien nos conecta.



# 

# Componentes BI

## Consultas Multidimensionales

### *Consultas Pivot4J Analytics 1*

En base a la pregunta que nos hemos hecho,

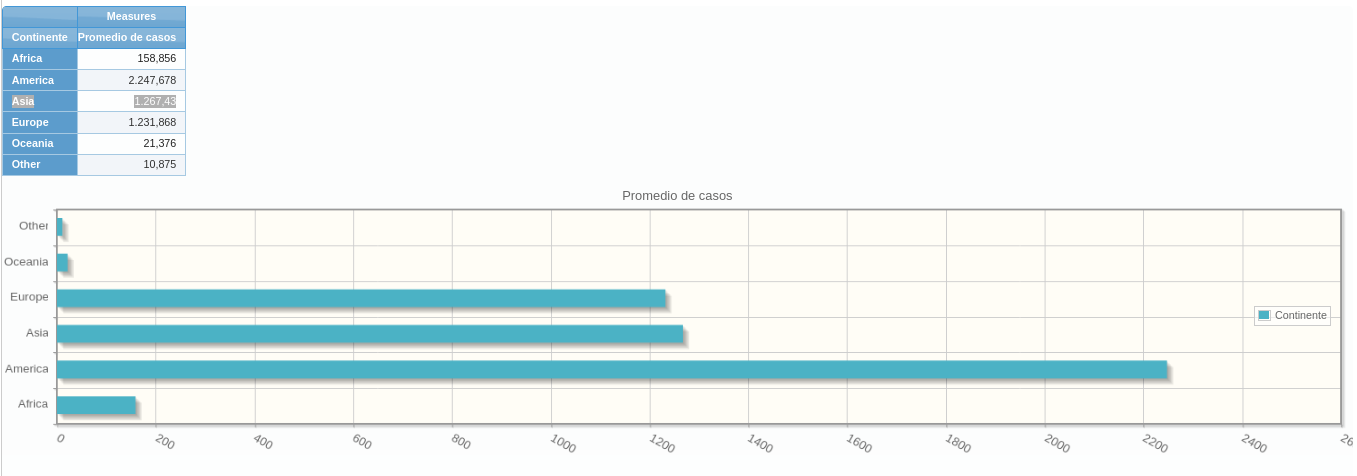
*¿Cuál fue el promedio de casos positivos en cada uno de los continentes por año?*

la respuesta que obtenemos es la cantidad en promedio de aquellos que se han sometido al hisopado dando resultado positivo de COVID. Esto permite conocer la situación sanitaria en cada continente para definir, de manera prioritaria, los siguientes pasos para paliar dicha situación de cada lugar.

La consulta es:

SELECT {[Measures].[Promedio de casos]} ON COLUMNS, {[dimContinente.Continente].[Africa], [dimContinente.Continente].[America], [dimContinente.Continente].[Asia], [dimContinente.Continente].[Europe], [dimContinente.Continente].[Oceania], [dimContinente.Continente].[Other]} ON ROWS FROM [Casos]

Resultado obtenido:



### *Consultas Pivot4J Analytics 2*

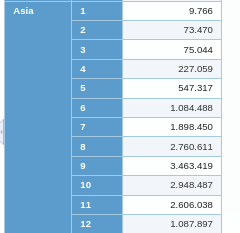
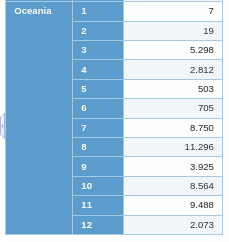
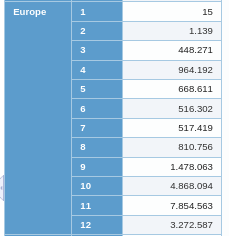
La siguiente pregunta que nos hemos hecho fue

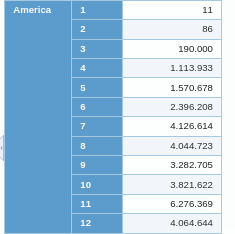
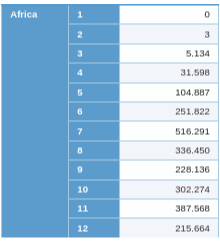
*¿Cuál es la cantidad de positivos por mes por cada continente?*

En base a esta pregunta, la respuesta que obtuvimos es la cantidad de los enfermos contagiados, obteniendo la medida por mes dando como resultado la rapidez en cómo el virus se ha esparcido por cada continente. Esta información, al igual que la anterior, permite tomar la decisión de qué manera ayudar en cada continente.

La consulta es:

SELECT NON EMPTY {[Measures].[casos positivos]} ON COLUMNS, NON EMPTY CrossJoin({[dimContinente.Continente].[Africa], [dimContinente.Continente].[America], [dimContinente.Continente].[Asia], [dimContinente.Continente].[Europe], [dimContinente.Continente].[Oceania]}, {[dimTiempo.Tiempo].[1], [dimTiempo.Tiempo].[2], [dimTiempo.Tiempo].[3], [dimTiempo.Tiempo].[4], [dimTiempo.Tiempo].[5], [dimTiempo.Tiempo].[6], [dimTiempo.Tiempo].[7], [dimTiempo.Tiempo].[8], [dimTiempo.Tiempo].[9], [dimTiempo.Tiempo].[10], [dimTiempo.Tiempo].[11], [dimTiempo.Tiempo].[12]}) ON ROWS FROM [Casos]

Resultado obtenido:



### *Consultas Pivot4J Analytics 3*

La siguiente pregunta fue:

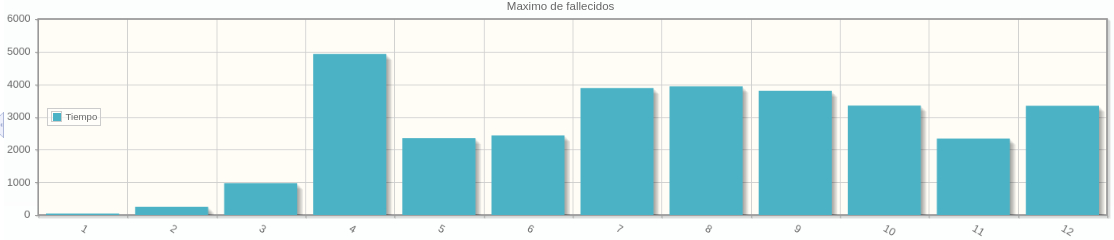
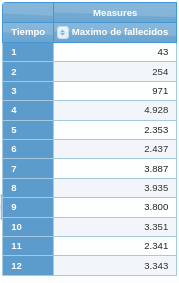
*¿Cuál es el máximo de fallecidos por mes? (pico máximo)*

Aplicando la consulta abajo visualizada, se obtuvo la cantidad máxima de enfermos contagiados que han perdido la vida por mes. Esta información resulta útil para determinar como avanzó el virus por la época del año permitiendo determinar si empeora o no la situación de los contagiados.

La consulta es:

SELECT {[Measures].[Maximo de fallecidos]} ON COLUMNS, {[dimTiempo.Tiempo].[1], [dimTiempo.Tiempo].[2], [dimTiempo.Tiempo].[3], [dimTiempo.Tiempo].[4], [dimTiempo.Tiempo].[5], [dimTiempo.Tiempo].[6], [dimTiempo.Tiempo].[7], [dimTiempo.Tiempo].[8], [dimTiempo.Tiempo].[9], [dimTiempo.Tiempo].[10], [dimTiempo.Tiempo].[11], [dimTiempo.Tiempo].[12]} ON ROWS FROM [Casos]

Resultado obtenido:



### 

### 

### 

### 

### 

### 

### *Consultas Pivot4J Analytics 4*

Al igual que la pregunta anterior,

*¿Cuál fue el promedio de fallecidos por país?*

Para este caso, se obtuvo la cantidad promedio de enfermos contagiados que han perdido la vida en cada país del mundo. Esta información resulta útil para determinar como avanzó el virus en cada país permitiendo analizar y determinar en donde poner mas foco para paliar la situación de los contagiados en cuanto a asistencia y capacidad hospitalaria.

La consulta es:

SELECT NON EMPTY {[Measures].[Promedio de fallecidos]} ON COLUMNS, NON EMPTY {[dimPais.Pais].[Afghanistan], [dimPais.Pais].[Algeria], [dimPais.Pais].[Armenia], [dimPais.Pais].[Australia], [dimPais.Pais].[Austria], [dimPais.Pais].[Azerbaijan], [dimPais.Pais].[Bahrain], [dimPais.Pais].[Belarus], [dimPais.Pais].[Belgium], [dimPais.Pais].[Brazil], [dimPais.Pais].[Cambodia], [dimPais.Pais].[Canada], [dimPais.Pais].[Cases\_on\_an\_international\_conveyance\_Japan], [dimPais.Pais].[Croatia], [dimPais.Pais].[Czechia], [dimPais.Pais].[Denmark], [dimPais.Pais].[Dominican\_Republic], [dimPais.Pais].[Ecuador], [dimPais.Pais].[Egypt], [dimPais.Pais].[Estonia], [dimPais.Pais].[Finland], [dimPais.Pais].[France], [dimPais.Pais].[Georgia], [dimPais.Pais].[Germany], [dimPais.Pais].[Greece], [dimPais.Pais].[Iceland], [dimPais.Pais].[India], [dimPais.Pais].[Indonesia], [dimPais.Pais].[Iran], [dimPais.Pais].[Iraq], [dimPais.Pais].[Ireland], [dimPais.Pais].[Israel], [dimPais.Pais].[Italy], [dimPais.Pais].[Japan], [dimPais.Pais].[Kuwait], [dimPais.Pais].[Lebanon], [dimPais.Pais].[Lithuania], [dimPais.Pais].[Luxembourg], [dimPais.Pais].[Malaysia], [dimPais.Pais].[Mexico], [dimPais.Pais].[Monaco], [dimPais.Pais].[Nepal], [dimPais.Pais].[Netherlands], [dimPais.Pais].[New\_Zealand], [dimPais.Pais].[Nigeria], [dimPais.Pais].[North\_Macedonia], [dimPais.Pais].[Norway], [dimPais.Pais].[Oman], [dimPais.Pais].[Pakistan], [dimPais.Pais].[Philippines], [dimPais.Pais].[Qatar], [dimPais.Pais].[Romania], [dimPais.Pais].[Russia], [dimPais.Pais].[San\_Marino], [dimPais.Pais].[Singapore], [dimPais.Pais].[South\_Korea], [dimPais.Pais].[Spain], [dimPais.Pais].[Sri\_Lanka], [dimPais.Pais].[Sweden], [dimPais.Pais].[Switzerland], [dimPais.Pais].[Taiwan], [dimPais.Pais].[Thailand], [dimPais.Pais].[United\_Arab\_Emirates], [dimPais.Pais].[United\_Kingdom], [dimPais.Pais].[United\_States\_of\_America], [dimPais.Pais].[Vietnam], [dimPais.Pais].[China], [dimPais.Pais].[Anguilla], [dimPais.Pais].[Antigua\_and\_Barbuda], [dimPais.Pais].[Aruba], [dimPais.Pais].[Bahamas], [dimPais.Pais].[Barbados], [dimPais.Pais].[Belize], [dimPais.Pais].[Benin], [dimPais.Pais].[Bermuda], [dimPais.Pais].[Bhutan], [dimPais.Pais].[British\_Virgin\_Islands], [dimPais.Pais].[Central\_African\_Republic], [dimPais.Pais].[Chad], [dimPais.Pais].[Congo], [dimPais.Pais].[Djibouti], [dimPais.Pais].[Dominica], [dimPais.Pais].[El\_Salvador], [dimPais.Pais].[Equatorial\_Guinea], [dimPais.Pais].[Eritrea], [dimPais.Pais].[Eswatini], [dimPais.Pais].[Ethiopia], [dimPais.Pais].[Faroe\_Islands], [dimPais.Pais].[Fiji], [dimPais.Pais].[French\_Polynesia], [dimPais.Pais].[Gibraltar], [dimPais.Pais].[Greenland], [dimPais.Pais].[Grenada], [dimPais.Pais].[Guernsey], [dimPais.Pais].[Guinea], [dimPais.Pais].[Guinea\_Bissau], [dimPais.Pais].[Haiti], [dimPais.Pais].[Holy\_See], [dimPais.Pais].[Isle\_of\_Man], [dimPais.Pais].[Kyrgyzstan], [dimPais.Pais].[Laos], [dimPais.Pais].[Latvia], [dimPais.Pais].[Liberia], [dimPais.Pais].[Libya], [dimPais.Pais].[Liechtenstein], [dimPais.Pais].[Madagascar], [dimPais.Pais].[Maldives], [dimPais.Pais].[Mali], [dimPais.Pais].[Malta], [dimPais.Pais].[Mauritania], [dimPais.Pais].[Mongolia], [dimPais.Pais].[Montserrat], [dimPais.Pais].[Mozambique], [dimPais.Pais].[Myanmar], [dimPais.Pais].[Namibia], [dimPais.Pais].[New\_Caledonia], [dimPais.Pais].[Northern\_Mariana\_Islands], [dimPais.Pais].[Papua\_New\_Guinea], [dimPais.Pais].[Rwanda], [dimPais.Pais].[Saint\_Kitts\_and\_Nevis], [dimPais.Pais].[Saint\_Lucia], [dimPais.Pais].[Saint\_Vincent\_and\_the\_Grenadines], [dimPais.Pais].[Senegal], [dimPais.Pais].[Seychelles], [dimPais.Pais].[Sint\_Maarten], [dimPais.Pais].[Slovakia], [dimPais.Pais].[Somalia], [dimPais.Pais].[Suriname], [dimPais.Pais].[Timor\_Leste], [dimPais.Pais].[Turks\_and\_Caicos\_islands], [dimPais.Pais].[Uganda], [dimPais.Pais].[United\_States\_Virgin\_Islands], [dimPais.Pais].[Zambia], [dimPais.Pais].[Palestine], [dimPais.Pais].[Brunei\_Darussalam], [dimPais.Pais].[Jamaica], [dimPais.Pais].[Cote\_dIvoire], [dimPais.Pais].[Gabon], [dimPais.Pais].[Kenya], [dimPais.Pais].[Kosovo], [dimPais.Pais].[Togo], [dimPais.Pais].[Guatemala], [dimPais.Pais].[Guyana], [dimPais.Pais].[Kazakhstan], [dimPais.Pais].[Uruguay], [dimPais.Pais].[United\_Republic\_of\_Tanzania], [dimPais.Pais].[Gambia], [dimPais.Pais].[Montenegro], [dimPais.Pais].[Guam], [dimPais.Pais].[Nicaragua], [dimPais.Pais].[Cayman\_Islands], [dimPais.Pais].[Costa\_Rica], [dimPais.Pais].[Cape\_Verde], [dimPais.Pais].[Moldova], [dimPais.Pais].[Zimbabwe], [dimPais.Pais].[Cameroon], [dimPais.Pais].[Sudan], [dimPais.Pais].[Curaçao], [dimPais.Pais].[South\_Africa], [dimPais.Pais].[Uzbekistan], [dimPais.Pais].[Paraguay], [dimPais.Pais].[Trinidad\_and\_Tobago], [dimPais.Pais].[Jersey], [dimPais.Pais].[Venezuela], [dimPais.Pais].[Angola], [dimPais.Pais].[Syria], [dimPais.Pais].[Cuba], [dimPais.Pais].[Mauritius], [dimPais.Pais].[Jordan], [dimPais.Pais].[Niger], [dimPais.Pais].[Ghana], [dimPais.Pais].[Bangladesh], [dimPais.Pais].[Chile], [dimPais.Pais].[Saudi\_Arabia], [dimPais.Pais].[Bolivia], [dimPais.Pais].[Tunisia], [dimPais.Pais].[Cyprus], [dimPais.Pais].[Bulgaria], [dimPais.Pais].[Honduras], [dimPais.Pais].[Bosnia\_and\_Herzegovina], [dimPais.Pais].[Democratic\_Republic\_of\_the\_Congo], [dimPais.Pais].[Slovenia], [dimPais.Pais].[Andorra], [dimPais.Pais].[Albania], [dimPais.Pais].[Ukraine], [dimPais.Pais].[Hungary], [dimPais.Pais].[Burkina\_Faso], [dimPais.Pais].[Colombia], [dimPais.Pais].[Serbia], [dimPais.Pais].[Argentina], [dimPais.Pais].[Peru], [dimPais.Pais].[Panama], [dimPais.Pais].[Poland], [dimPais.Pais].[Morocco], [dimPais.Pais].[Puerto\_Rico], [dimPais.Pais].[Portugal], [dimPais.Pais].[Turkey], [dimPais.Pais].[Bonaire, Saint Eustatius and Saba], [dimPais.Pais].[Botswana], [dimPais.Pais].[Burundi], [dimPais.Pais].[Falkland\_Islands\_(Malvinas)], [dimPais.Pais].[Lesotho], [dimPais.Pais].[Western\_Sahara], [dimPais.Pais].[Sao\_Tome\_and\_Principe], [dimPais.Pais].[Comoros], [dimPais.Pais].[Tajikistan], [dimPais.Pais].[Malawi], [dimPais.Pais].[Sierra\_Leone], [dimPais.Pais].[South\_Sudan], [dimPais.Pais].[Yemen], [dimPais.Pais].[Marshall\_Islands], [dimPais.Pais].[Solomon\_Islands], [dimPais.Pais].[Wallis\_and\_Futuna], [dimPais.Pais].[Vanuatu]} ON ROWS FROM [Casos]

Resultado obtenido:



## Reportes

### Reporte 1

En este reporte se realizó una comparación sobre la cantidad de casos positivos y las personas fallecidas, todo estos datos están detallados por mes y por país.

Con este reporte se podría conocer los meses o los países en los cuales se diagnosticaron la mayor cantidad de casos o personas fallecidas.

Preguntas a la que responde:

* Cantidad de casos por tiempo y país
* Cantidad de personas fallecidas por tiempo y país

<https://drive.google.com/drive/folders/1OmvpAmGd0dL2rDEh8NYVJjTEV2dzcKMs?usp=sharing>

### Reporte 2

En este reporte se puede observar con detalle el promedio de casos positivos por mes y en cada continente.

Con este reporte se podría conocer en qué meses y en qué continentes el promedio de casos fue más elevado o más leve.

Preguntas a la que responde:

* Cantidad de casos promedio por tiempo y por continente

<https://drive.google.com/drive/folders/1H4Y_t4QhcjXkO1aHYvlZHYF2yqE4F57K?usp=sharing>

### Reporte 3

En este reporte se puede observar una lista del máximo de casos positivos detallada por país y por mes.

Con este reporte se podría conocer en qué meses y/o países se detectaron la cantidad máxima de casos positivos.

Preguntas a la que responde:

* Cantidad máxima de casos positivos por tiempo y por país.

<https://drive.google.com/drive/folders/1hnLzlrW4hn0OJHbVaM2jCMinapT09do9?usp=sharing>

### Reporte 4

En este reporte se puede observar una lista con el promedio de personas fallecidas detallado por mes y por continente.

Con este reporte se podría observar en qué meses y/o continentes el promedio de personas fallecidas tuvo más repercusión.

Preguntas a la que responde:

* Cantidad promedio de personas fallecidas por tiempo y por continente

<https://drive.google.com/drive/folders/1RmIl4j_lcImjhRLCbRVkAwH2GE4cx46P?usp=sharing>

### Reporte 5

En este reporte se puede observar con detalle una lista detallada por país de: casos positivos, personas fallecidas, personas recuperadas, máximo de casos positivos y máximo de cantidad de personas fallecidas. Todo esta información teniendo en cuenta que el año al cual pertenecen los datos es 2020.

Con este reporte podríamos hacernos todo tipo de preguntas con respecto a todos los ítems de detalle teniendo en cuenta los países y que es todo anual (2020).

Preguntas a las que responde:

* Cantidad de casos por país
* Cantidad de personas fallecidas por país
* Cantidad de personas recuperadas por país
* Cantidad máxima de casos positivos por país
* Cantidad máxima de personas fallecidas por país

<https://drive.google.com/drive/folders/1Ii-P0Of70nqp94XhxMPOVRfyqkqVKaYI?usp=sharing>

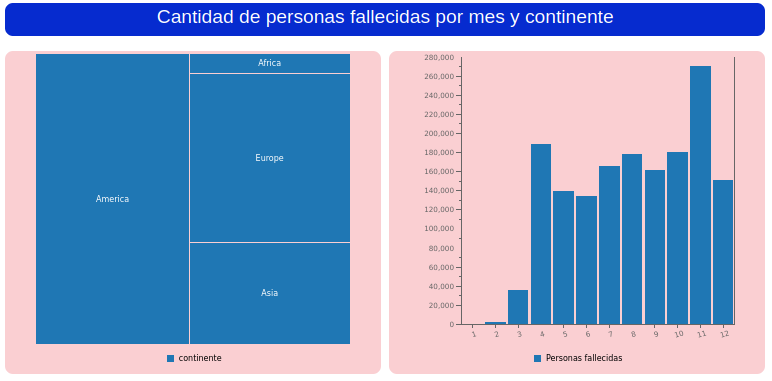
## Tableros

### Tablero 1

Preguntas de negocio a las que da respuesta:

* Cantidad de personas fallecidas por continente.
* Cantidad de personas fallecidas por mes.

El siguiente tablero creado con CDE (Community Dashboard Editor) tiene como objetivo detallar la cantidad de personas fallecidas que hubo por cada mes en cada continente. Como podemos ver, el gráfico de la izquierda, mediante un treemap chart, representa los datos de la dimensión continente, teniendo como base la cantidad de personas fallecidas que hubo en cada uno, y en el gráfico de la derecha, a través de un bar chart, se muestran los datos de la dimensión mes también respecto a las personas fallecidas. Siendo parametrizable según continente, nos permite conocer de forma interactiva y gráficamente la cantidad de personas fallecidas que hubo en cada uno a lo largo del tiempo. Entonces al seleccionar un continente, por ejemplo América, veremos la variación de los fallecidos en el gráfico de barra desde el mes 1 (Enero) hasta el mes 12 (Diciembre) como se muestra a continuación:

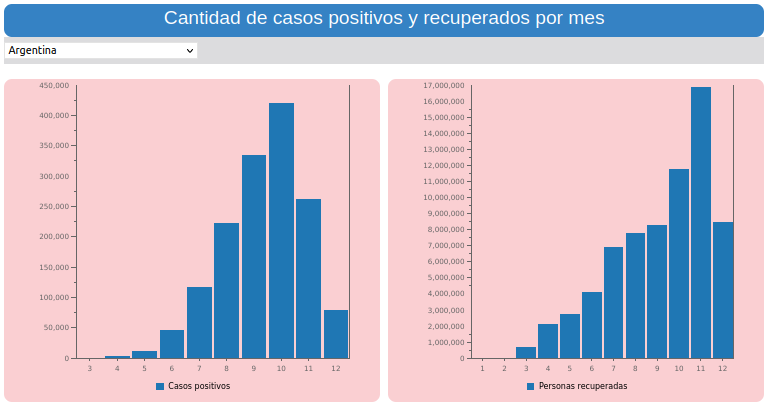


### Tablero 2

Preguntas de negocio a las que da respuesta:

* Cantidad de casos positivos por país por mes.
* Cantidad de personas recuperadas por país por mes.

Nuevamente, el tablero fue desarrollado con CDE (Community Dashboard Editor) y tiene como objetivo detallar la cantidad de casos positivos y recuperados por mes en países de América. A continuación, el gráfico de la izquierda, mediante un bar chart, representa los datos de la dimensión mes, teniendo como base la cantidad de personas fallecidas que hubo en cada uno, y en el gráfico de la derecha, también por medio de un bar chart, se muestran los datos de la dimensión mes pero con respecto a las personas recuperadas, para este caso en particular realizamos un cálculo en su SQL Query restando los casos positivos con las personas fallecidas para obtener los recuperados y poder mostrar los que hubo en cada mes y en un determinado país. Siendo parametrizable según los datos de la dimensión país, utilizando una lista desplegable configurada con una SQL Query donde seleccionamos únicamente los países de América, en donde dependiendo del país que se seleccione, nos permite conocer de forma interactiva y gráficamente la cantidad de casos positivos que hubo en el mismo a lo largo del tiempo como así también las personas recuperadas en ese mismo tiempo. Entonces al seleccionar la lista desplegable podemos elegir por ejemplo Argentina y ver los datos como de muestran a continuación:



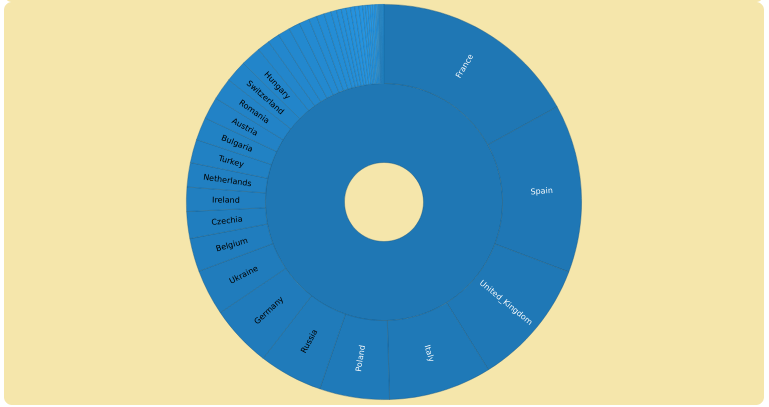
### Tablero 3

Preguntas de negocio a las que da respuesta:

* Pico máximo de casos positivos por países de Europa.
* Pico máximo de personas fallecidas por países de Europa.

Por último, este tablero también fue creado con CDE (Community Dashboard Editor) y tiene como objetivo detallar cuál fue el pico máximo de casos positivos y personas fallecidas que hubo en todos los países de Europa, de forma anual. A continuación, en el primer gráfico, con un pie chart se representan los datos de la dimensión país, en donde se muestran únicamente los países Europeos por lo que para esto se modificó la SQL Query, teniendo como base el pico máximo de casos que hubo en cada uno, y en el gráfico de abajo, ahora por medio de un sunburst chart, se muestran los datos de la dimensión país pero con respecto al pico máximo de personas fallecidas en cada país, para este caso aplicamos la misma lógica del caso anterior en su SQL Query para filtrar por países de Europa. Para este tablero, no se realizaron interacciones por lo que únicamente se detallan gráficamente los resultados de cada análisis. En el pie chart los países se diferencian con colores y según sea su pico máximo de casos, se van a diferenciar en tamaño, además muestra el pico máximo de cada uno y el porcentaje que representan. Y en el caso de el sunburst chart, los países se diferencian con el tamaño del pico máximo de personas fallecidas, en donde si pasamos el mouse sobre algún país, nos va a mostrar el número. A continuación vemos los gráficos:





# Conclusión

Nos gustaría mencionar dos aspectos importantes que se desprenden de este trabajo. El primero, es que nos permitió pensar “out of the box” y pensar de otra manera diferente a la forma estructura en donde los datos se guardan de manera relacional. Aquí hemos tenido la posibilidad de poder extraer y analizar información recibida de diferentes fuentes y con un gran volumen de datos. Estos, completamente desnormalizados y representando la realidad misma en que se presenta la información. Gracias a las bondades del DW podemos acceder a la información y obtener resultados luego de analizarla.

Lo segundo, y como corolario de lo antes mencionado, aplicando la metodología solicitada por el profesor (Metodología Hefesto) y usando las herramientas necesarias hemos podido, no solo observar el poder del manejo de la información y cómo obtener sus resultados sino también conocer de manera resumida, cómo afectó, en nuestro caso de estudio, el covid-19 a la población a nivel mundial en el año 2020, en donde todo comenzó y algunos hemos podido seguir adelante y otros se han quedado atrás sucumbiendo a la enfermedad.

Si el mundo entero pudiera acceder a la información de manera rápida como nosotros, podrían tomar conciencia de qué tan rápido la población se contagia y se recupera o fallece en cada país en cada momento del año.

# Diccionario de datos

| Nombre Tabla | Nombre columna | Descripción | Tipo de datos | Rango de datos | Fórmula | Observaciones |
| --- | --- | --- | --- | --- | --- | --- |
| Casos | id\_tiempo | Clave primaria,  Id identificador  de la tabla  casos | BIGINT | 1…352 | - | Se relaciona con una fecha determinada (dia, mes y año) |
|  | id\_pais | Clave foránea,  Id identificador  de la tabla  dimPais | BIGINT | 1…352 | - | - |
|  | id\_continente | Clave foránea,  Id identificador  de la tabla  dimContinente | BIGINT | 1…352 | - | - |
|  | casos\_positivos | Indicador de  los casos de hisopados positivos | INT | - | SUM(casos\_positivos) | - |
|  | fallecidos | Indicador de  los de los fallecidos por casos positivos | INT | - | SUM(fallecidos) | - |
|  | max\_fallecidos | Indicador de  la cantidad máxima de los fallecidos | INT | - | MAX(fallecidos) | - |
|  | prom\_fallecidos | Indicador del promedio de fallecidos | FLOAT | - | AVG(fallecidos) | - |
|  | prom\_casos | Indicador del promedio de  casos de hisopados positivos | FLOAT | - | AVG(fallecidos) | - |
| dimTiempo | id\_tiempo | Clave primaria,  Id identificador  de la tabla  dimTiempo | BIGINT | 1…352 | - | - |
|  | nombre\_mes | Nombre del mes | TINYTEXT | - | - | - |
| dimPais | Id\_Pais | Clave primaria,  Id identificador  de la tabla  dimPais | BIGINT | 1…352 | - | - |
|  | nombre\_pais | Nombre del País | TINYTEXT | - | - | - |
| dimContinente | Id\_Continente | Clave primaria,  Id identificador  de la tabla  dimContinente | BIGINT | 1…352 | - | - |
|  | nombre\_Continente | Nombre del Continente | TINYTEXT | - | - | - |

# Bibliografía

Dataset utilizado:

* <https://datos.gob.ar/dataset?page=2>

Material teórico utilizado:

* Hefesto Data Warehousing: Guía completa de aplicación teorico-práctica; metodologia Data Warehouse. Por: Mariano García Mattío y Dario R. Bernabeu
* Pentaho 5.0 Reporting by Example: Beginner’s Guide. Por: Mariano García Mattío y Dario R. Bernabe