

**ESCUELA SUPERIOR POLITÉCNICA DEL**

**LITORAL**

**SISTEMA DE BASE DE DATOS AVANZADO**

**II Término, 2021**

**INTEGRANTES**

• Gary Barzola

• Tommy Villagómez

**PROYECTO DEL SEGUNDO PARCIAL**

Apache Kylin

Tabla de contenido

[Objetivos del proyecto 1](#_Toc312039528)

[Descripción del dataset 2](#_Toc1784075784)

[Instalación del apache kylin 2](#_Toc2024059974)

[Importación de datos 4](#_Toc859098450)

[Creación del modelo 6](#_Toc971066531)

[Creación del cubo (debe de ser un cubo diferente al creado en clases) 9](#_Toc165109835)

[Reportes: crear tres reportes con el cubo de información utilizando: roll-up, drill-down y dice. 11](#_Toc1743806886)

[Roll-up (sql) 12](#_Toc866819208)

[￼Roll-up (Visualization) 12](#_Toc1646813520)

[Dice(sql) 13](#_Toc2126248085)

[Dice(visualization) 13](#_Toc1699775999)

# Objetivos del proyecto

Crear un cubo de información con el dataset de AdventureWorks

Generar reportes a partir del cubo de información

# Descripción del dataset

El dataset contiene información de la empresa Adventure Works que se dedica a la venta de bicicletas y partes de estas a clientes y distribuidores. Debido a su gran crecimiento y demanda, la compañía está buscando la implementación de una solución que les permita a los gerentes tomar decisiones para su expansión global. Las dimensiones que se tiene en el dataset actualmente son: dimcustomer , dimcurrency, dimdate, dimsalesterritory, dimproduct ,dimgeography, dimproductcategory, dimproductsubcategory y la tabla de hechos factinternetsales.

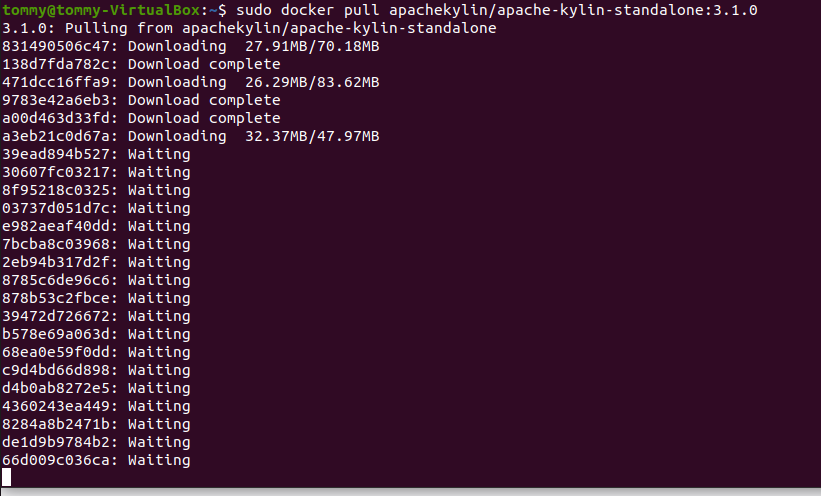
# Instalación del apache kylin

Para la instalación de Apache Kylin se utilizó Docker, el manual de instalación de esta herramienta se lo puede encontrar en el siguiente enlace: <https://docs.docker.com/engine/install/ubuntu/>

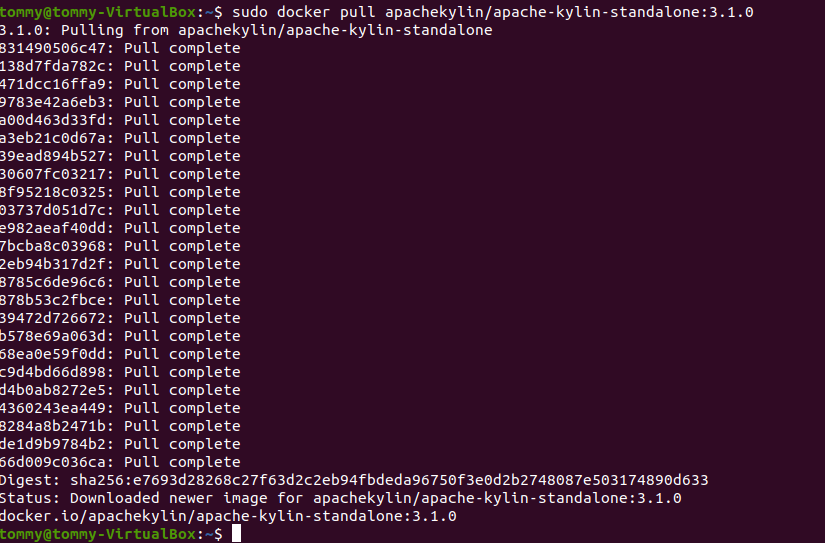
Una vez se instale Docker se procede a instalar kylin en un contenedor Docker, siguiendo los siguientes pasos:

1. Descargar la imagen de kylin versión 3 en Docker con el comando:

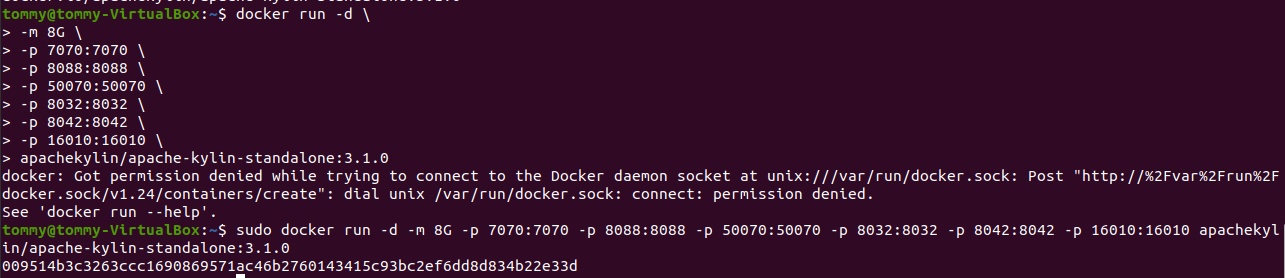
docker pull apachekylin/apache-kylin-standalone:3.1.0



1. Cuando la instalación se complete deberá aparecer algo como la imagen de abajo

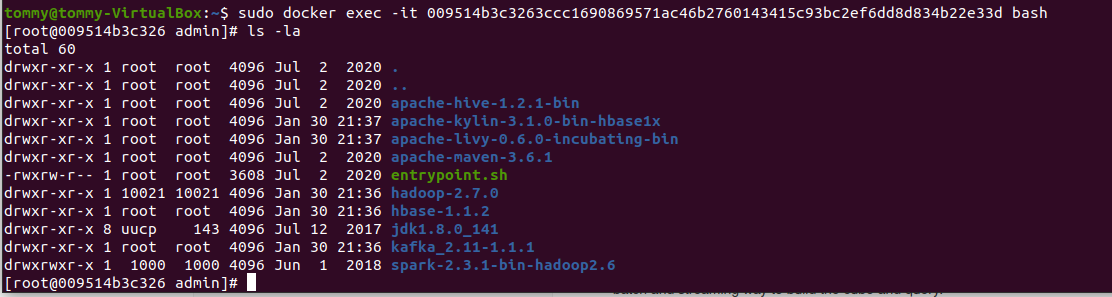


1. Luego de tener la imagen de Kylin en nuestra máquina, ejecutamos un contenedor con la imagen, como lo podemos ver la siguiente imagen

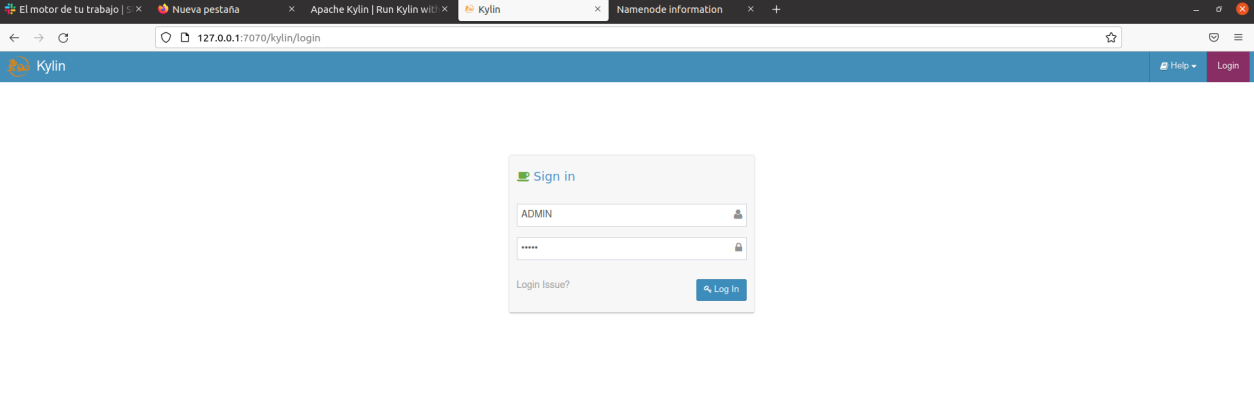


1. Copiamos el código hash que retorna la ejecución del container, este es el id del contenedor que contiene al kylin.
2. Ejecutamos el siguiente comando para acceder a la instancia del contenedor:

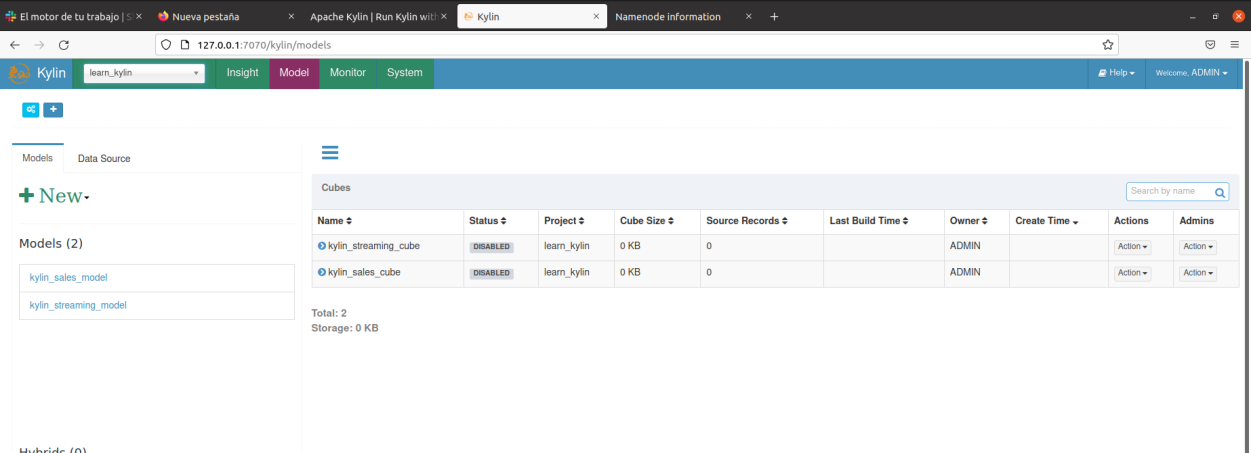
*sudo exec –it <id> bash*



1. Podemos comprobar la ejecución del Kylin si ingresamos desde el navegador al siguiente enlace: <http://127.0.0.1:7070/kylin/login,> y debe aparecer el login



1. El usuario es ADMIN y la contraseña KYLIN
2. Una vez logeado aparecerá la siguiente interfaz



# Importación de datos

Para la importación de datos se utilizará SQOOP, el manual de instalación de este se encuentra en el siguiente recurso <https://www.tutorialspoint.com/sqoop/sqoop_installation.html>

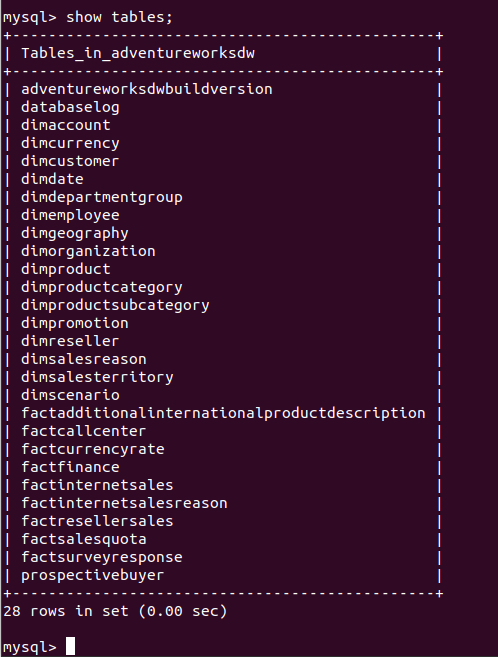
Una vez instalado sqoop vamos a proceder a importar los datos que ya deben estar previamente cargados en un servidor de mysql, este proceso se puede realizar dentro del contenedor de kylin, con el comando:

mysql –u root –p123456 <base\_de\_dato\_existente> < <nombre\_del\_archivoSQL\_a\_cargar>

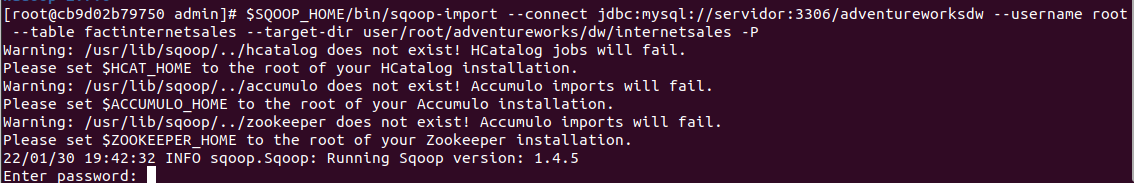
Si se cargó todo correctamente ingresamos a mysql y debería aparecer la información de la base de dato cargada:

Comando para ingresar a mysql:

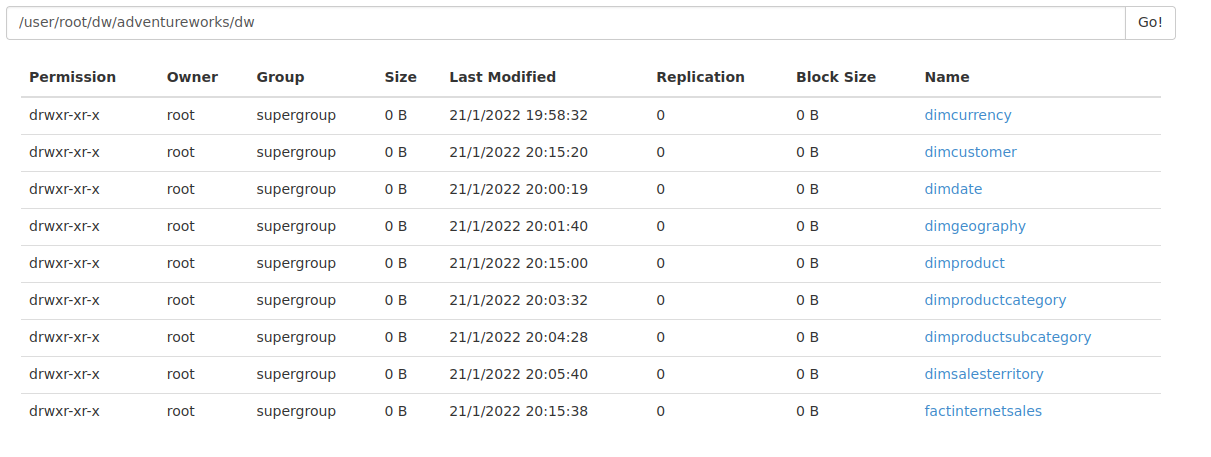
mysql –u root –p123456



1. Ahora, volvemos al contenedor de kylin y vamos a ejecutar el comando que nos permitirá hacer la importación al HDFS



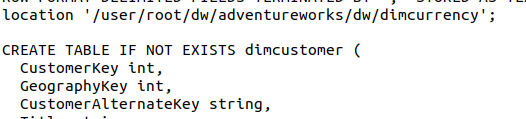
1. Pedirá la contraseña para confirmar, una vez hecho esto se empezará con la importación
2. Este proceso se deberá repetir con cada una de las tablas de interés (hechos y dimensiones)
3. Una vez se haga toda la importación se deberá acceder al hdfs para corroborar que la información se subió correctamente. [http://127.0.0.1:50070](http://127.0.0.1:50070/)
4. La información se encontrará en el target file establecido en el comando



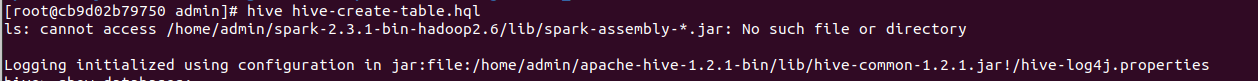
# Creación del modelo

Para la creación del modelo primero tenemos que crear una base de datos en hive

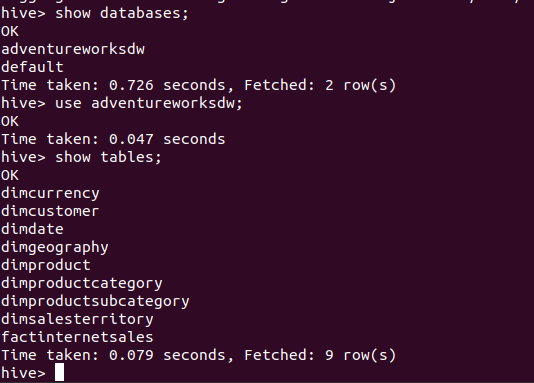
1. Para esto se usará un archivo del repositorio de github llamado [hive-create-table.hql](https://github.com/vanechev/ejemplo-dw-kylin/blob/main/hive/hive-create-table.hql)
2. Antes de ejecutar el script se deberá realizar una configuración adicional al archivo, en este caso la location, y se pondrá la ubicación del directorio HDFS en donde se cargaron los datos previamente.



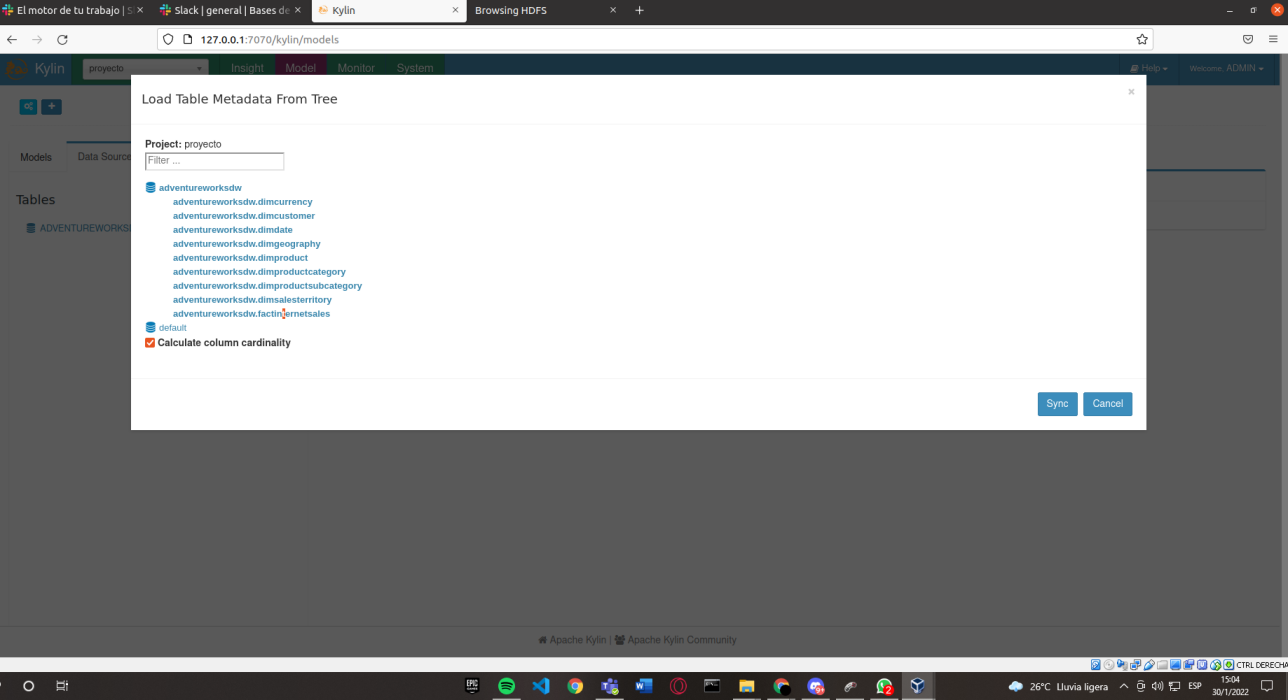
1. Se deberá cambiar la locación a todas las tablas que se crean en el archivo
2. Una vez realizado esto podemos crear la base de datos en hive, con el siguiente comando hive <nombredelarchivoHQL>



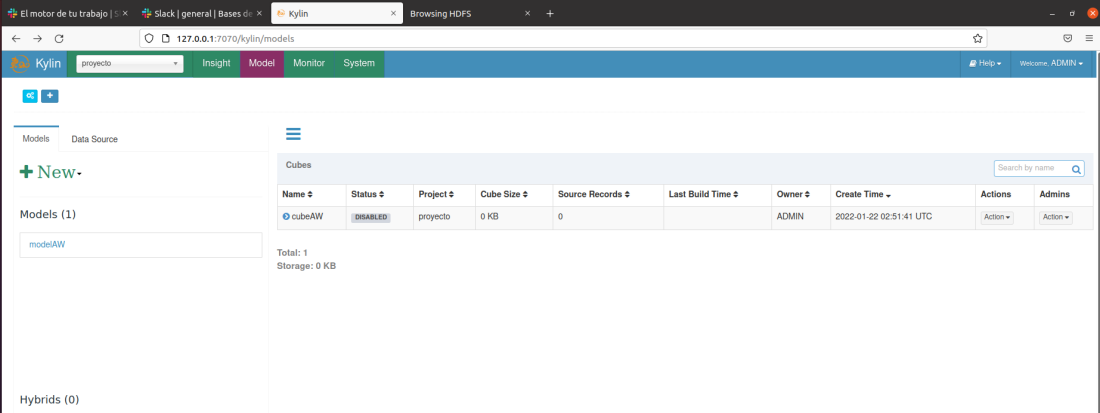
1. Una vez hecho esto procedemos a revisar si se creó la base correctamente en hive con las tablas respectivas.



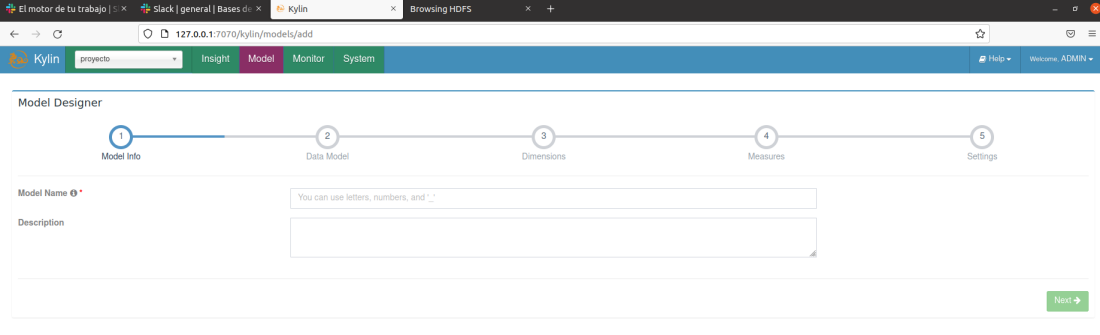
1. Una vez tenemos todo esto, debemos cargar las tablas de hive en Kylin. Esto se hace en la interfaz de Kylin. Vista Model --> Pestaña Data Source --> Opcion Load-table from tree--> Seleccionar todas las tablas --> Sync



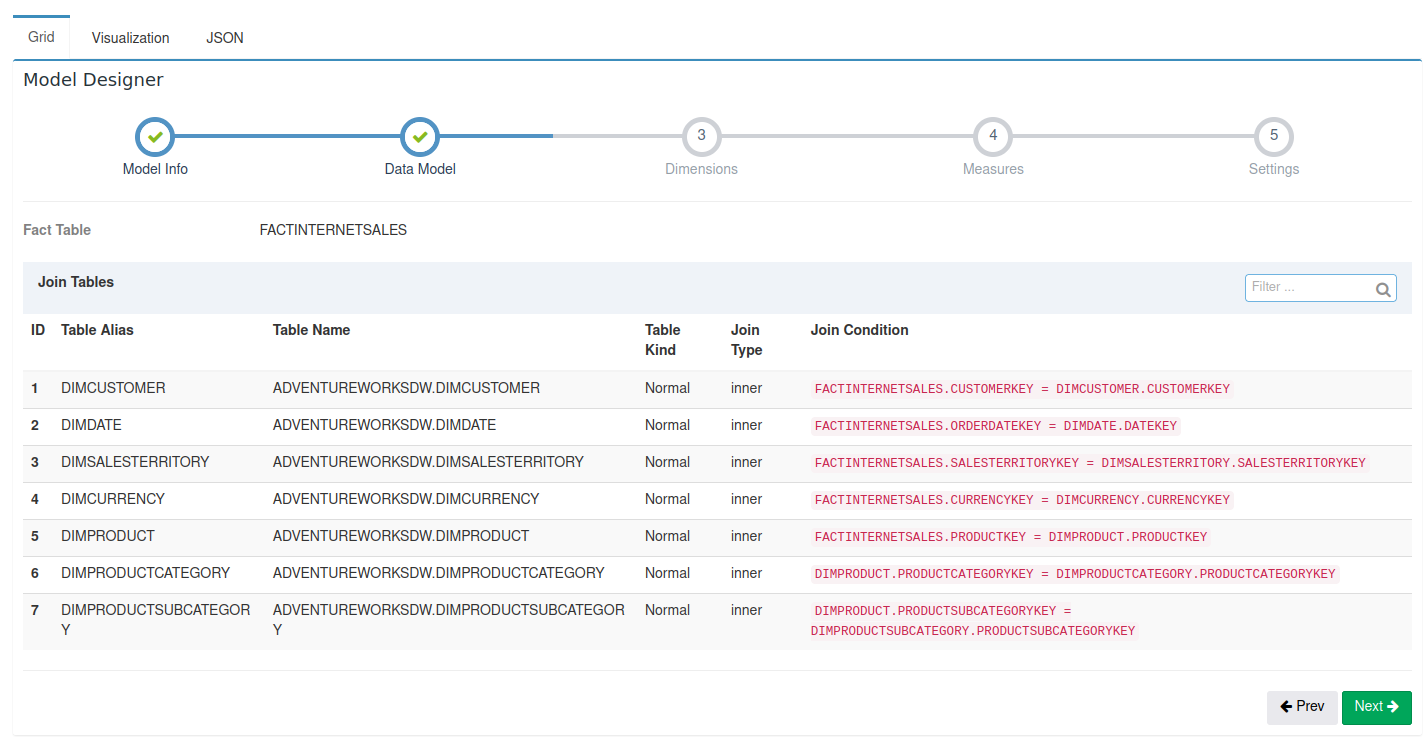
1. Para la creación del modelo, nos dirigimos a model-> new model



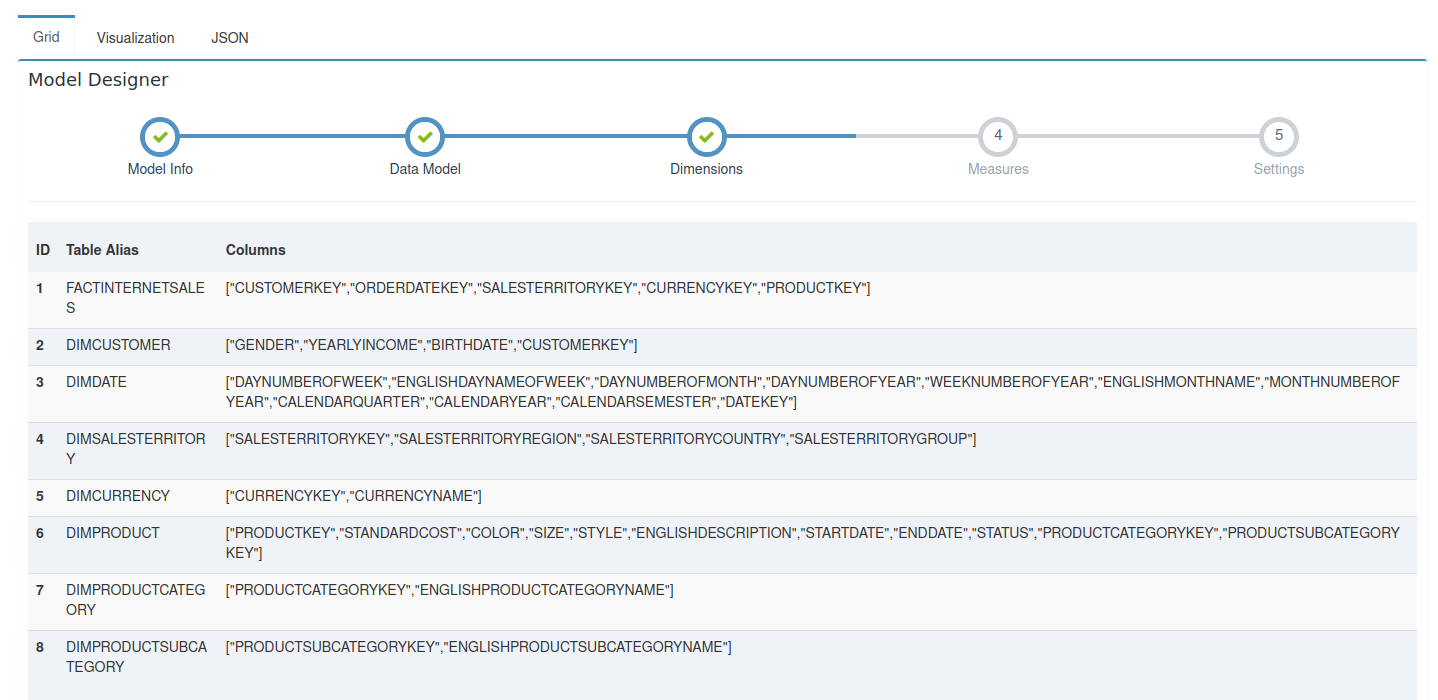
1. Aparecerá la siguiente ventana de configuración para crear el modelo.



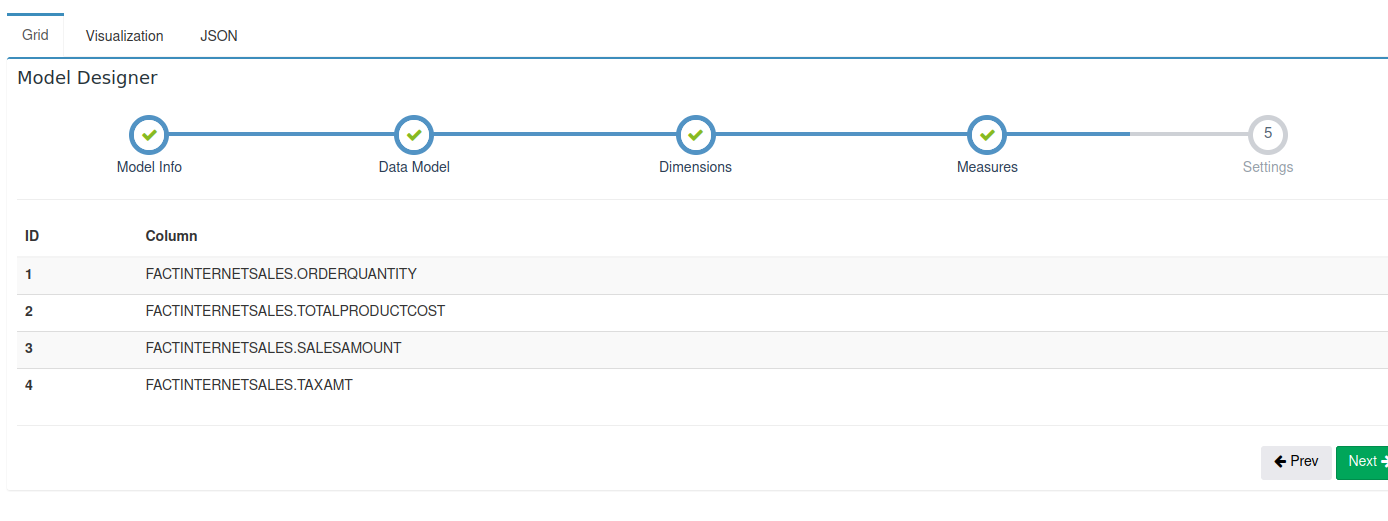
1. La ventana data model deberá quedar así, con las foreign keys relacionadas



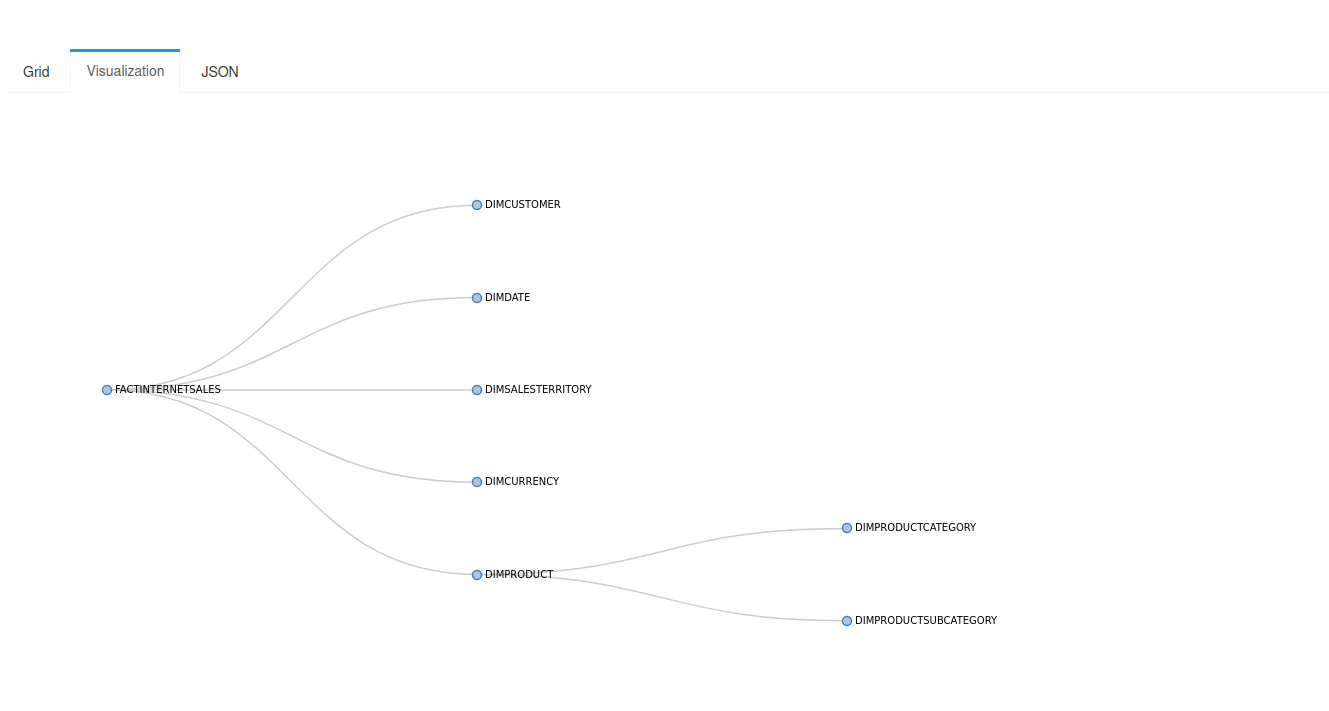
1. Las dimensiones deberán contener los siguientes atributos



1. Las measures

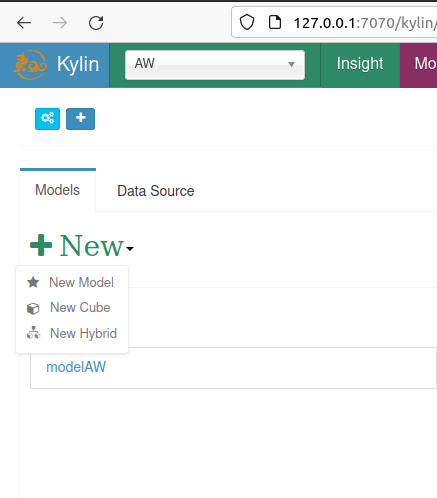


1. En settings no se debe modificar nada
2. Finalmente, en la ventana visualization se debe observar algo así

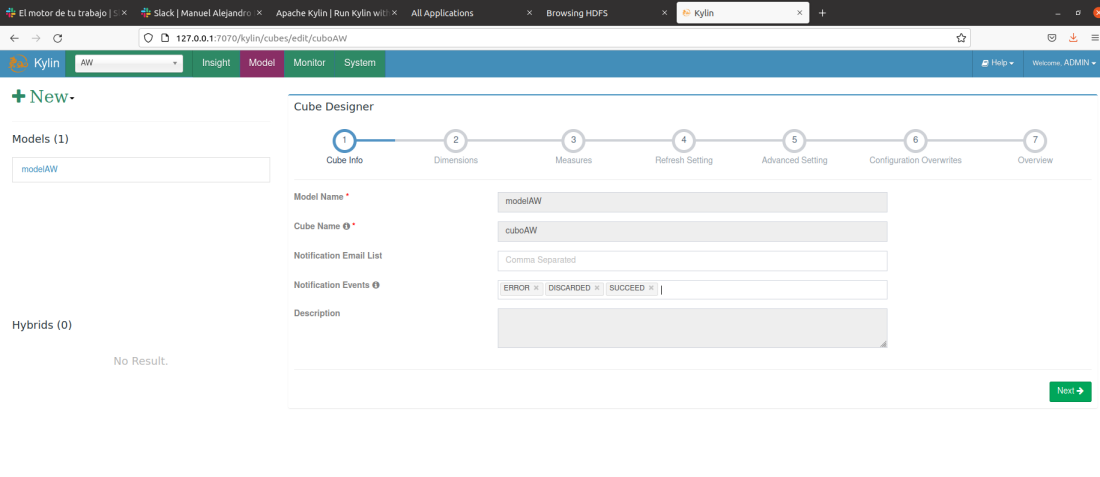


# Creación del cubo (debe de ser un cubo diferente al creado en clases)

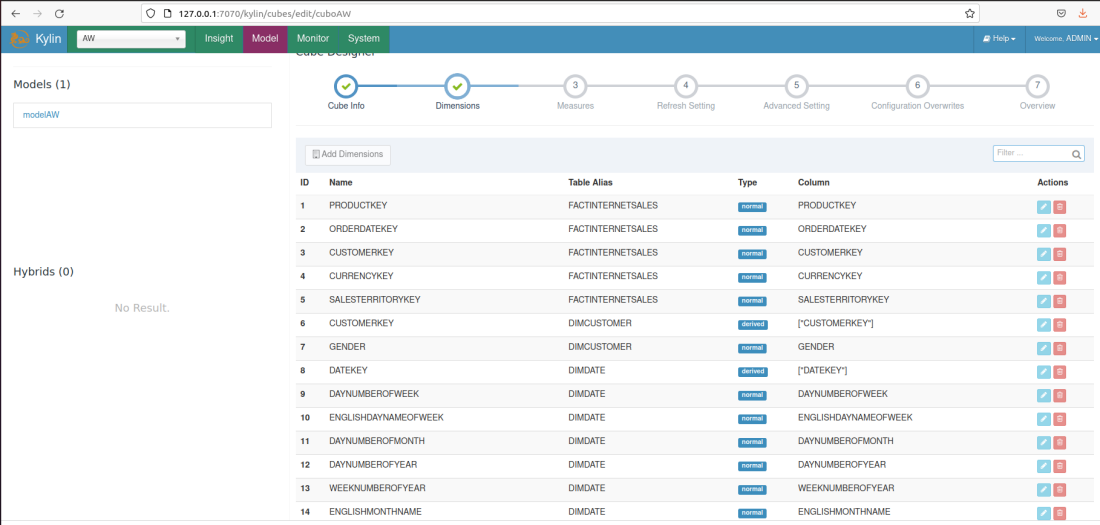
1. Para la creación del cubo, nos dirigimos a model-> new cube



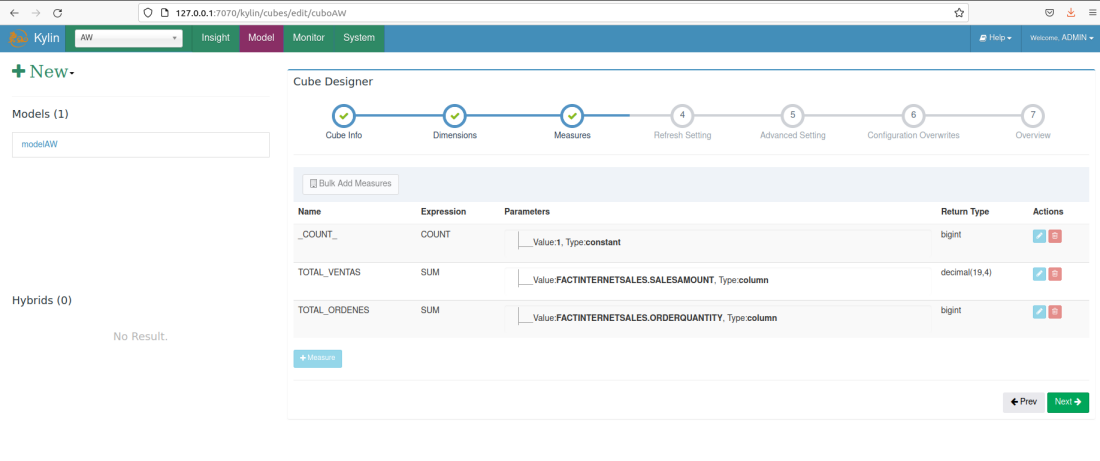
1. Escogemos el modelo creado anteriormente y escribimos el nombre para nuestro cubo



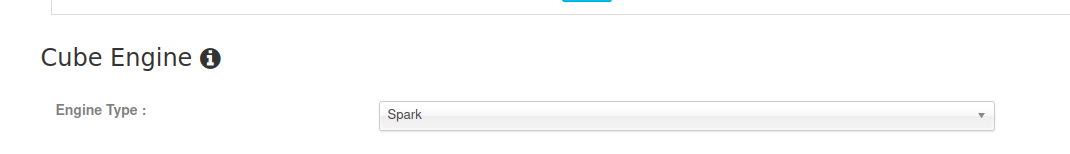
1. En el apartado de dimensiones se seleccionan las dimensiones que queremos cargar en nuestro cubo y las enlazamos con la tabla de hechos: factinternetsales



1. En measures, agregamos las siguientes



1. En refresh setting, se deja la configuración por defecto.
2. Advanced setting, se modifica el Engine Type del cubo a Spark

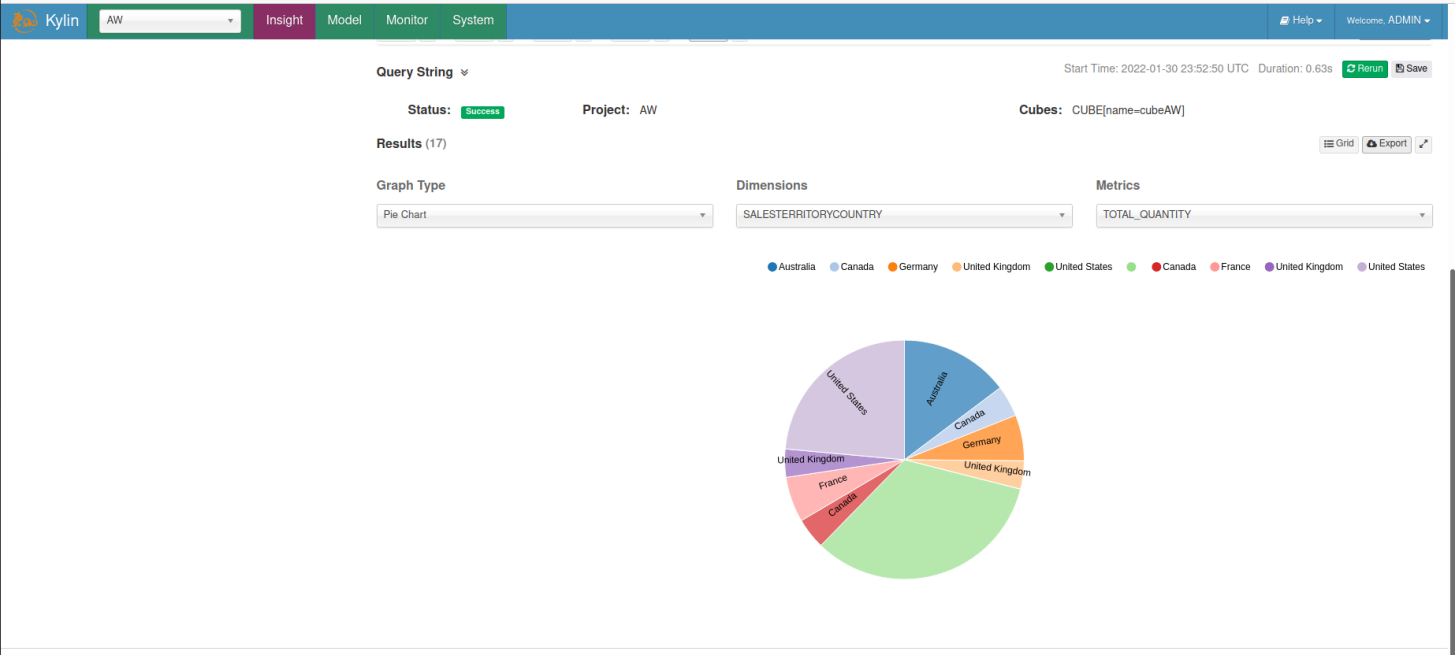


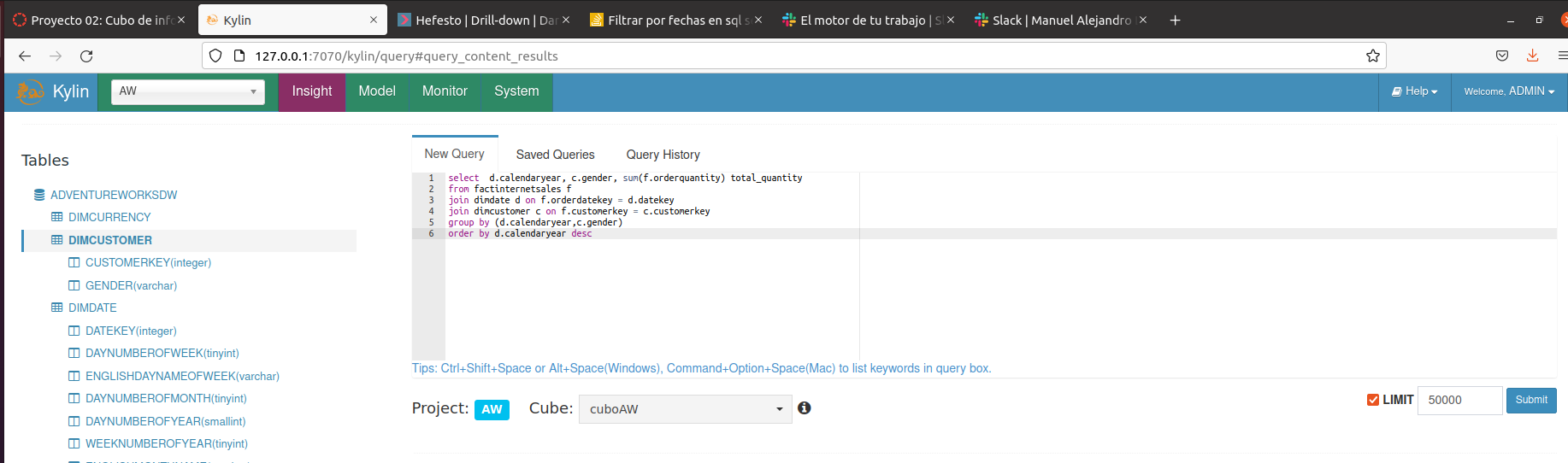
1. Configuration overwrites, se deja la configuración por defecto
2. Overview, resumen del cubo a construir

# Reportes: crear tres reportes con el cubo de información utilizando: roll-up, drill-down y dice.

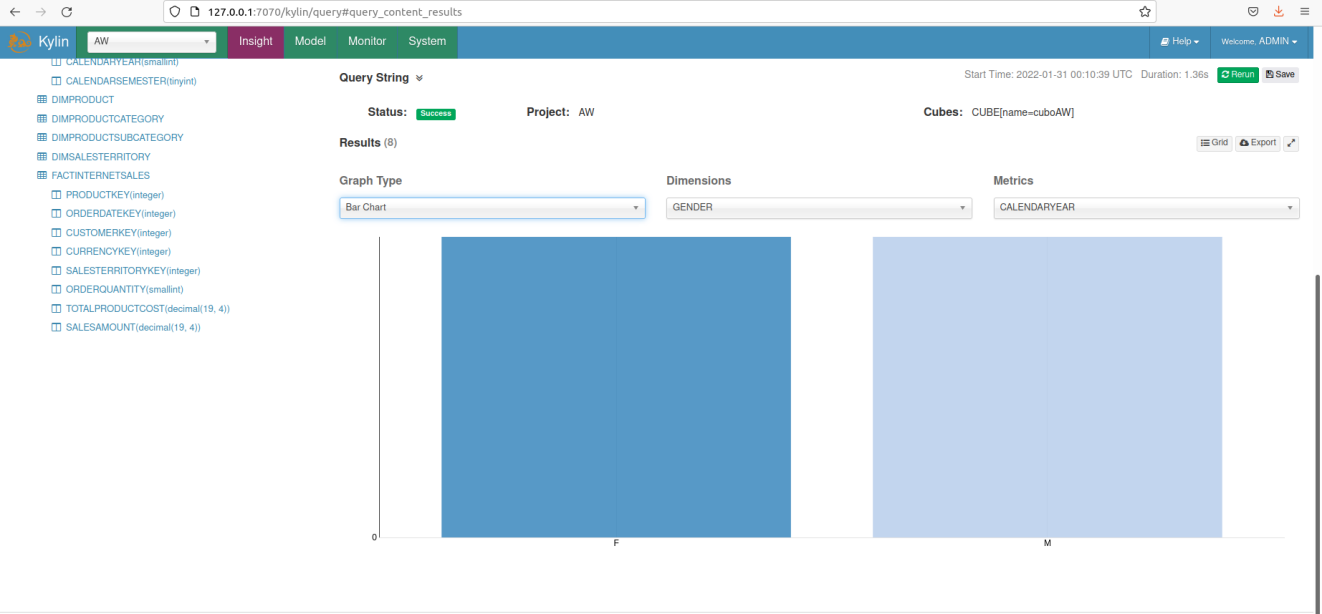
## Roll-up (sql)

## Roll-up (Visualization)

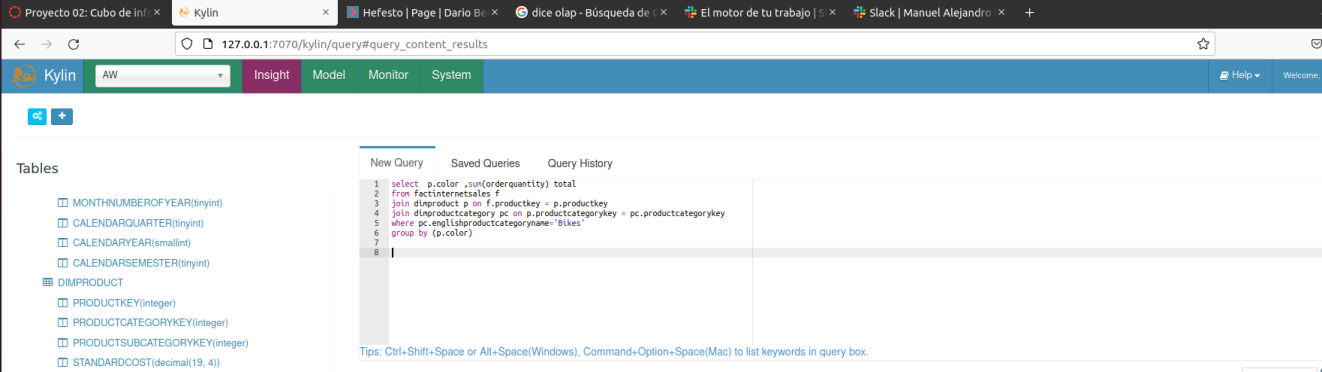
Drill-down (sql)



Drill-down (Visualization)



## Dice(sql)



## Dice(visualization)

