# ESCUELA SUPERIOR POLITECNICA DEL LITORAL



**FACULTAD DE INGENIERÍA EN ELECTRICIDAD Y COMPUTACIÓN BASES DE DATOS AVANZADAS**

PROYECTO:

MANUAL DE USUARIO ADVENTURE WORKS

PROFESORA:

VANESSA I. ECHEVERRIA BARZOLA

INTEGRANTES:

LUIS CARLOS SANCHEZ PLAZA

MANUEL ALEJANDRO LOOR MACIAS

TÉRMINO:

II PAO 2021 – 2022

## Objetivos

* Aprender como instalar Apache Kylin con Docker, para poder hacer uso de sus herramientas en el proyecto.
* Crear un Data Warehouse con Apache Kylin y construir un modelo de datos haciendo del dataset Adventure Works.
* Comprender el funcionamiento de los Cubos de Apache Kylin y su relación con la eficiencia en la consulta de datos.

## Descripción del dataset

Adventure Works es una dataset, proporcionado por la profesora, el cual almacena los datos correspondientes a cada una de las 12 tablas presentes en la base de datos. El modelo de la base de datos es el siguiente:

Diagrama, Esquemático

Descripción generada automáticamente

El modelo entidad-relación simula los datos que debería de manejar una empresa que dedica a la venta de productos por internet; como se puede observar en la imagen, cada una de las tablas cuenta con claves y atributos que pueden ser usados en las diferentes consultas que se realicen a través de Apache Kylin.

## Instalación Apache Kylin

Para la instalación de Apache Kylin se hice uso de Docker; por ende, previamente se debió de haber instalado Docker en nuestro ordenador. Una vez instalado Docker, los pasos a seguir son sumamente sencillos:

1. Abrimos el terminal y ejecutamos el comando: **docker pull apachekylin/apache-kylin-standalone:3.1.0**, con lo cual se creará un nuevo contenedor en docker el cual ya contendrá todo cada uno de los servicios que Apache Kylin necesita para funcionar, algunos de estos son: JDK, Hadoop, Hive, Spark, MySQL, etc.
2. Una vez el proceso de creación del contenedor haya finalizado, procederemos a ejecutarlo haciendo uso del comando**: docker run -d -m 8G -p 7070:7070 -p 8088:8088 -p 50070:50070 -p 8032:8032 -p 8042:8042 -p 16010:16010 apachekylin/apache-kylin-standalone:3.1.0**, con lo cual ya dejaremos en funcionamiento el programa Apache Kylin.
3. Una vez hecho esto, accederemos al programa a través del navegador, para esto nos dirigiremos al siguiente enlace: <http://127.0.0.1:7070/kylin/login>.

## Importación de datos

Antes de empezar la importación de datos al repositorio hdfs de Hadoop, es necesario ejecutar el script ‘adventureworksdw2008.sql’ del enlace [https://github.com/vanechev/ejemplo-dw-kylin/tree/main/dataset]. El script sql se encontrará comprimido como .bz2, por lo que deberá utilizar el comando ‘sudo bzip2 adventureworksdw2008.sql.bz2’. Por consiguiente, podrá ejecutar el script con el comando ‘mysql -u root -p123456 mysql<adventureworksdw2008.sql’.

Posteriormente, se recomienda utilizar la herramienta de Sqoop, la cual puede ser instalada siguiendo los pasos 3,4,5,6 y 7 del apartado de ‘Descargando Sqoop’ del enlace [https://www.tutorialspoint.com/sqoop/sqoop\_installation.htm].

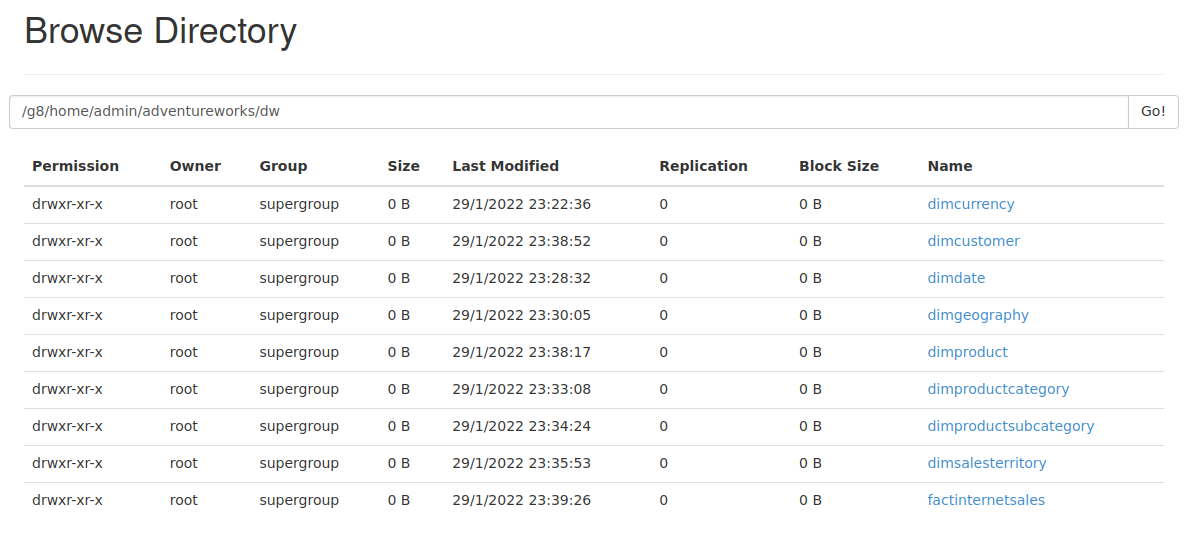
Con la herramienta de Sqoop ya instalada, se podrá importar los datos ejecutando las siguientes líneas de comandos, que representan la importación de los datos de cada tabla:

* SQOOP\_HOME/bin/sqoop-import --connect jdbc:mysql://localhost:3306/adventureworksdw --username root --driver com.mysql.jdbc.Driver --table factinternetsales --as-parquetfile --target-dir /g8/home/admin/adventureworks/dw/internetsales -P
* SQOOP\_HOME/bin/sqoop-import --connect jdbc:mysql://localhost:3306/adventureworksdw --username root --driver com.mysql.jdbc.Driver --table dimcurrency --as-parquetfile --target-dir /g8/home/admin/adventureworks/dw/dimcurrency -P
* SQOOP\_HOME/bin/sqoop-import --connect jdbc:mysql://localhost:3306/adventureworksdw --username root --driver com.mysql.jdbc.Driver --table dimcustomer --as-parquetfile --target-dir /g8/home/admin/adventureworks/dw/customer -P
* SQOOP\_HOME/bin/sqoop-import --connect jdbc:mysql://localhost:3306/adventureworksdw --username root --driver com.mysql.jdbc.Driver --table dimdate --as-parquetfile --target-dir /g8/home/admin/adventureworks/dw/dimdate -P
* SQOOP\_HOME/bin/sqoop-import --connect jdbc:mysql://localhost:3306/adventureworksdw --username root --driver com.mysql.jdbc.Driver --table dimgeography --as-parquetfile --target-dir /g8/home/admin/adventureworks/dw/dimgeography -P
* SQOOP\_HOME/bin/sqoop-import --connect jdbc:mysql://localhost:3306/adventureworksdw --username root --driver com.mysql.jdbc.Driver --table dimproduct --as-parquetfile --target-dir /g8/home/admin/adventureworks/dw/product -P
* SQOOP\_HOME/bin/sqoop-import --connect jdbc:mysql://localhost:3306/adventureworksdw --username root --driver com.mysql.jdbc.Driver --table dimproductsubcategory --as-parquetfile --target-dir /g8/home/admin/adventureworks/dw/dimproductcategory -P
* SQOOP\_HOME/bin/sqoop-import --connect jdbc:mysql://localhost:3306/adventureworksdw --username root --driver com.mysql.jdbc.Driver --table factinternetsales --as-parquetfile --target-dir /g8/home/admin/adventureworks/dw/dimproductsubcategory -P
* SQOOP\_HOME/bin/sqoop-import --connect jdbc:mysql://localhost:3306/adventureworksdw --username root --driver com.mysql.jdbc.Driver --table dimsalesterritory --as-parquetfile --target-dir /g8/home/admin/adventureworks/dw/dimsalesterritory -P

Cabe mencionar que para el tutorial se utilizó la versión 1.4.7 de Sqoop, dado que el indicador –as-parquetfile, el cual está disponible para esta versión, permite los importar los archivos de datos con el formato **parquet**.

Finalmente, deberá descargar el script ‘hive-create-table.hql’ que se encuentra en el enlace [https://github.com/vanechev/ejemplo-dw-kylin/tree/main/hive] y ejecutarlo dentro de su contenedor Docker, con el comando **hive -f hive-create-table.hql;** donde Hive es un sistema gestor de base de datos parecido a MySql, y además es el sistema de base de datos utilizado para la creación de modelos y cubos en Apache Kylin.

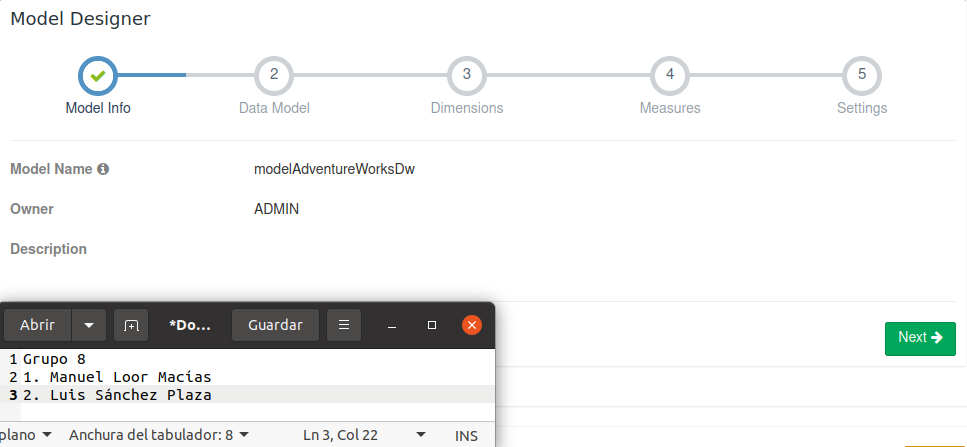
Como resultado, en el repositorio hdfs se tendrá lo siguiente:

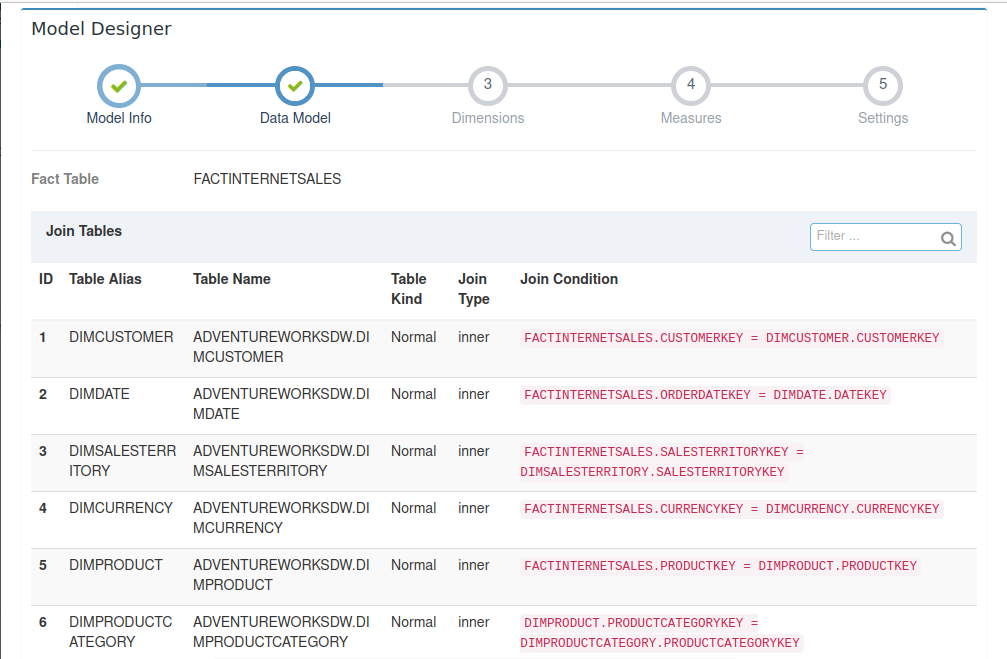


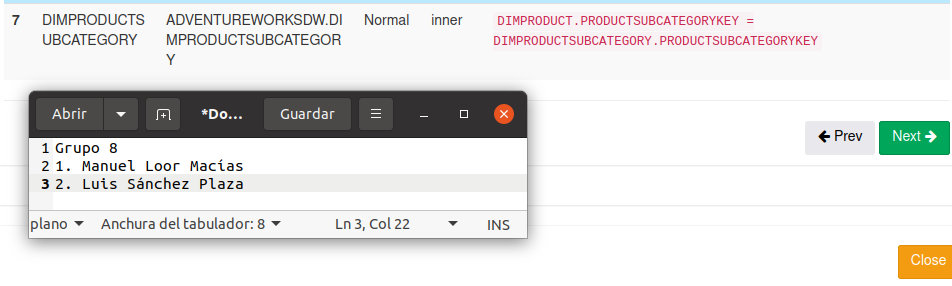
## Creación del modelo

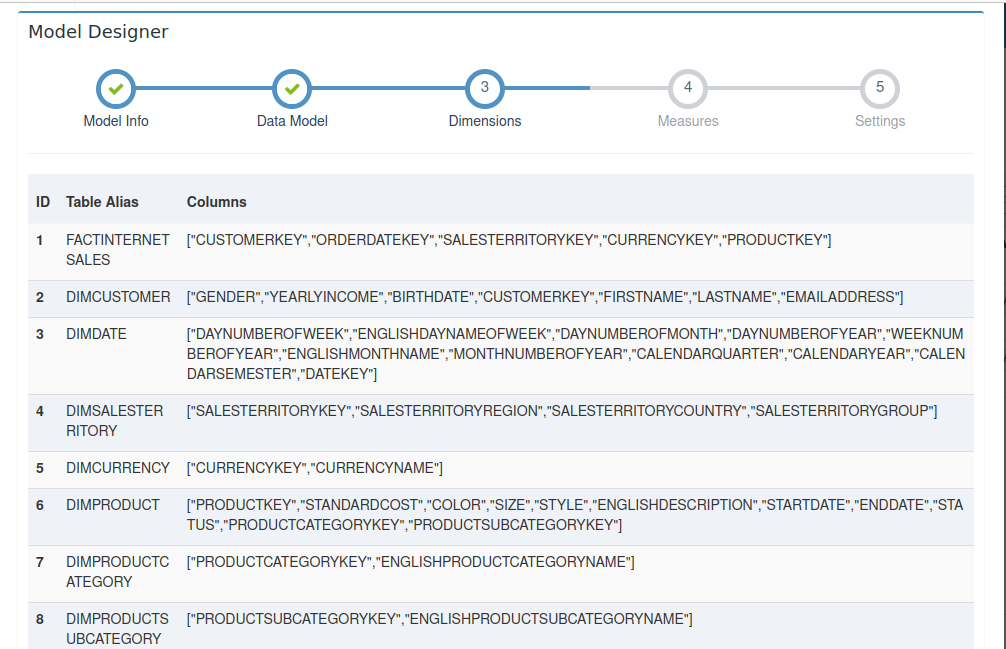
Para crear un modelo en Apache Kylin, es necesario haber seleccionado un proyecto y haber elegido una fuente de datos (Data Source), dado que se realizaron los pasos anteriormente mencionados, deberá aparecer en su interfaz la base de datos de ‘ADVENTUREWORKSDW’, seleccionar todas sus tablas y dar clic en Sincronizar (Sync).

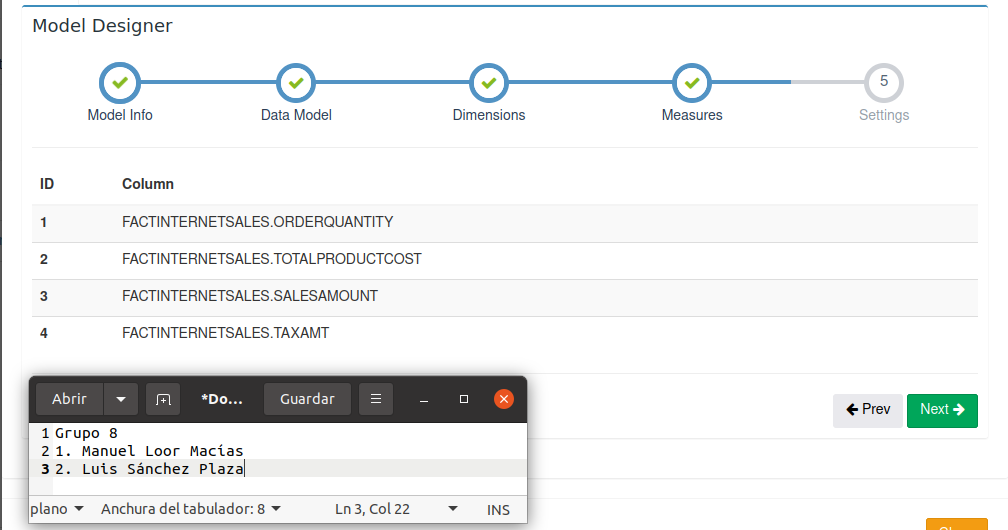
Posteriormente, cree el modelo con el wizard de Apache Kylin, con las siguientes configuraciones:

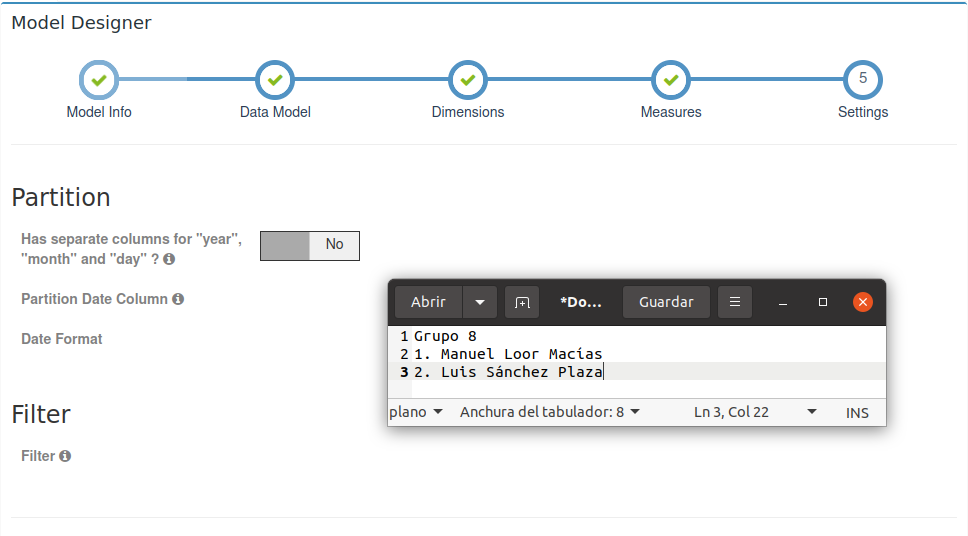








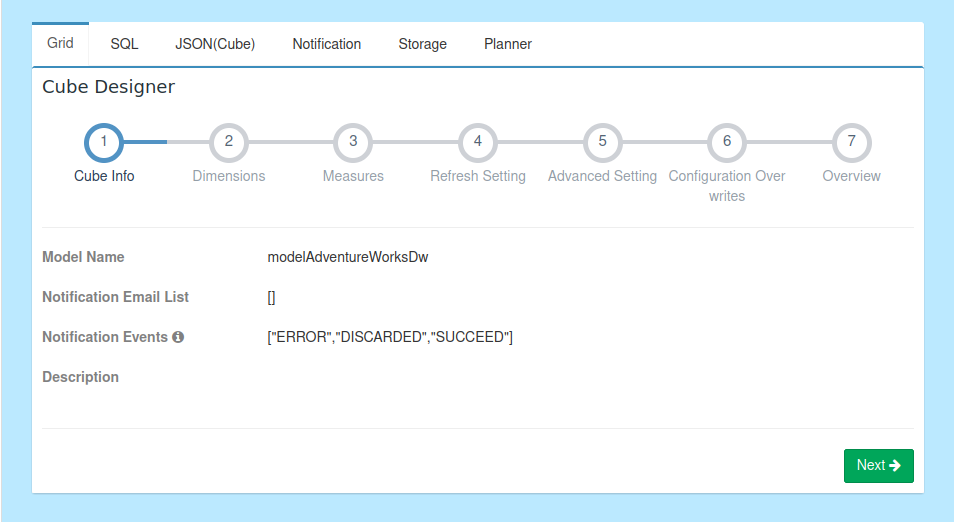
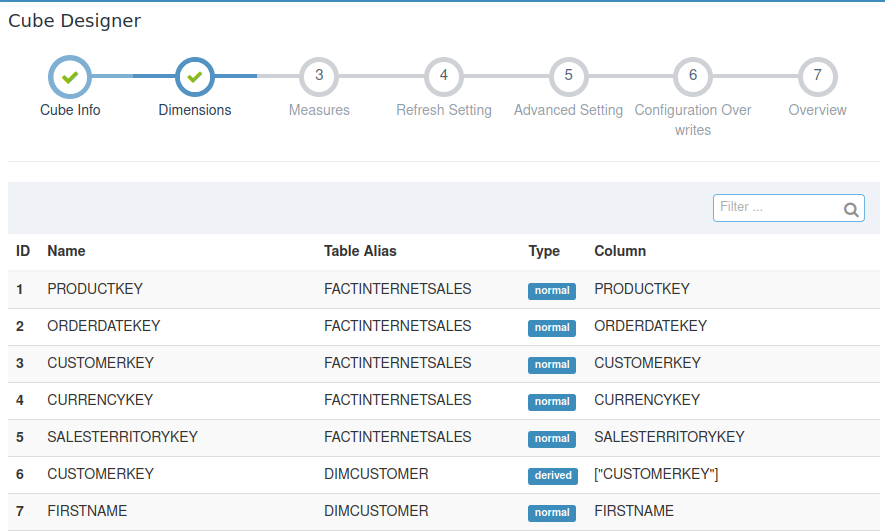


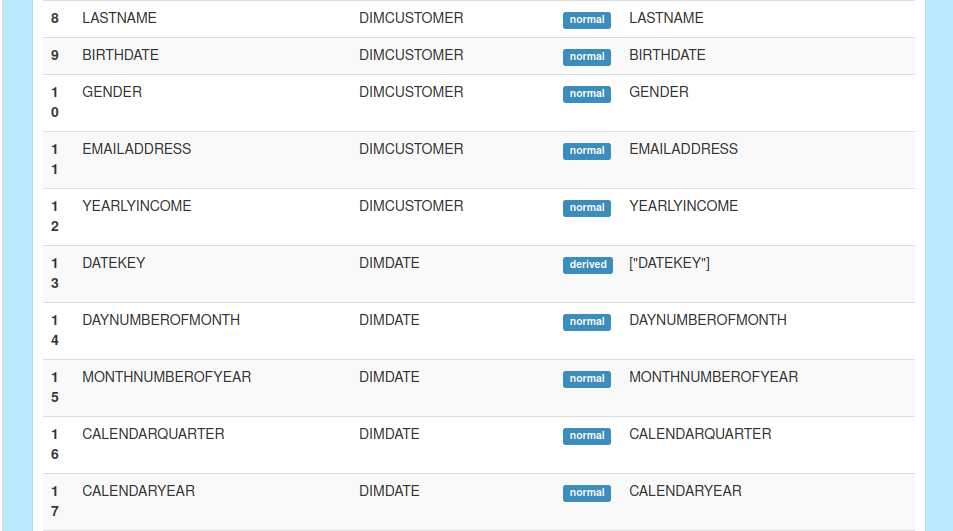


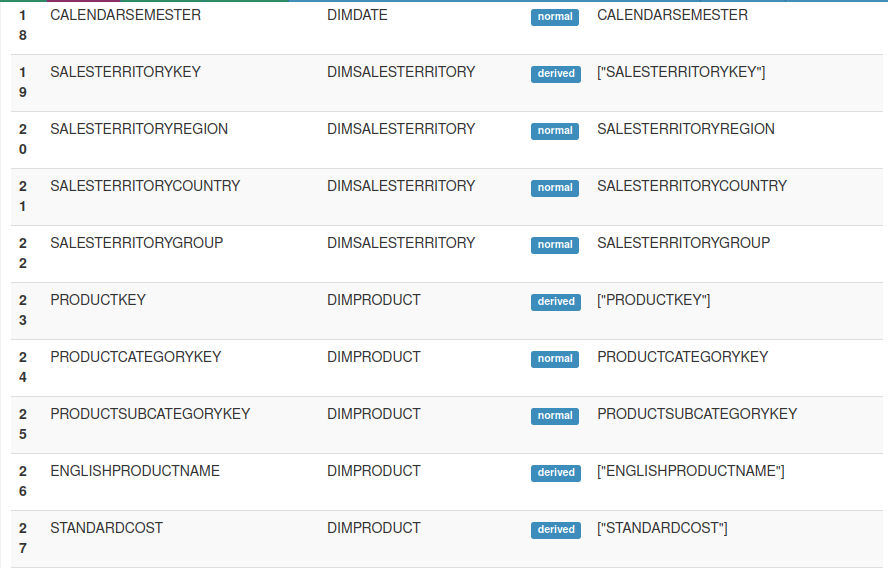
## Creación del cubo

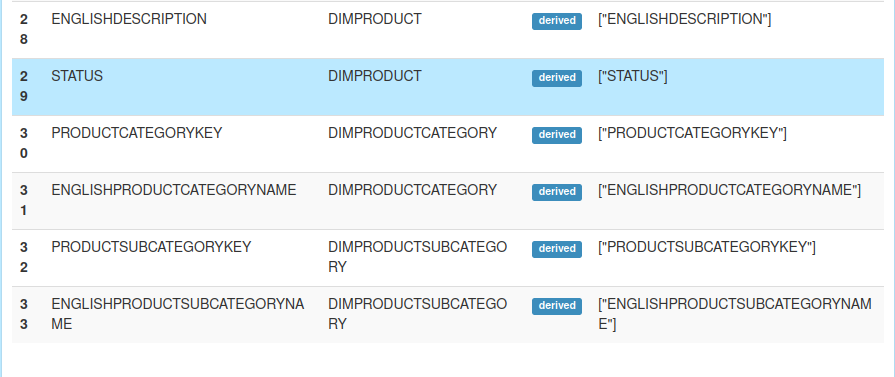
Para crear un cubo en Apache Kylin, es necesario haber seleccionado un modelo existente.

Posteriormente, cree el cubo OLAP con el wizard de Apache Kylin, con las siguientes configuraciones:

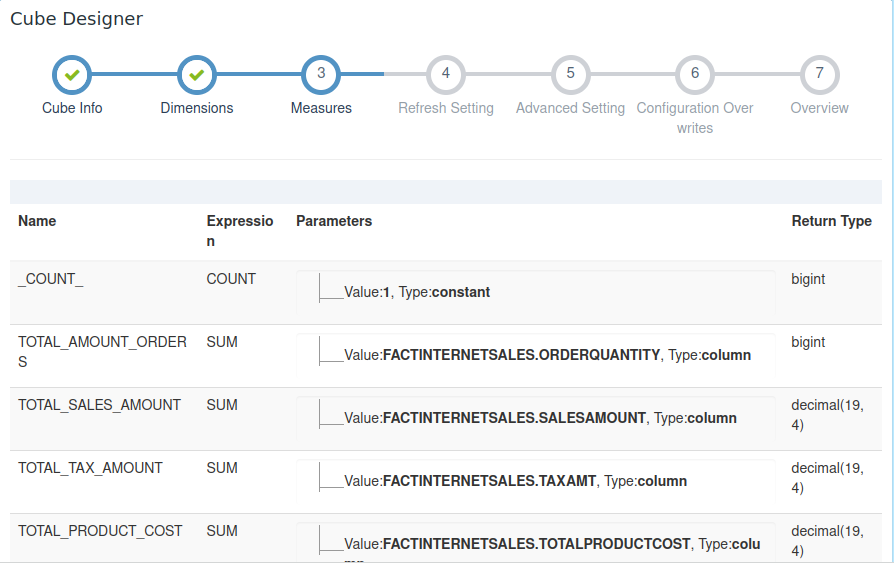
1. Selección de un modelo existente.
2. Agregación de las dimensiones (33 dimensiones seleccionadas). Se omitieron las dimensiones de la tabla Currency y Geography.

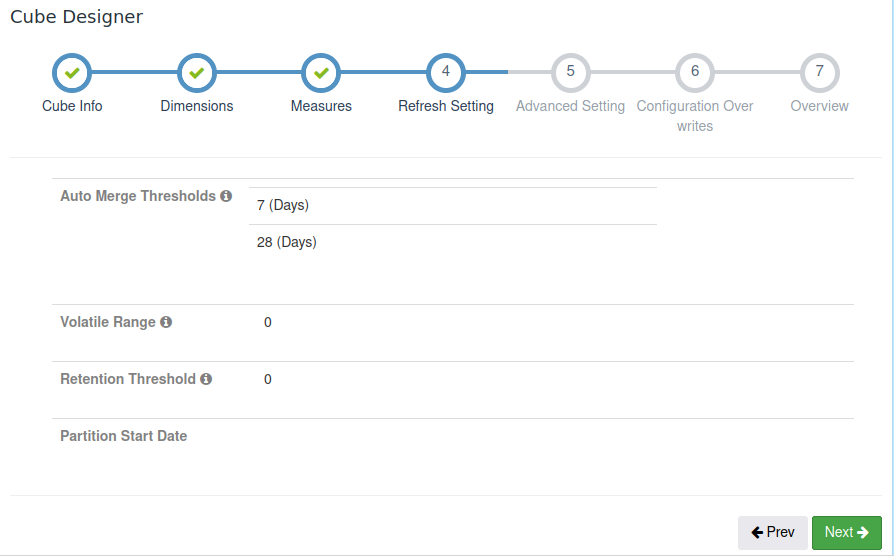


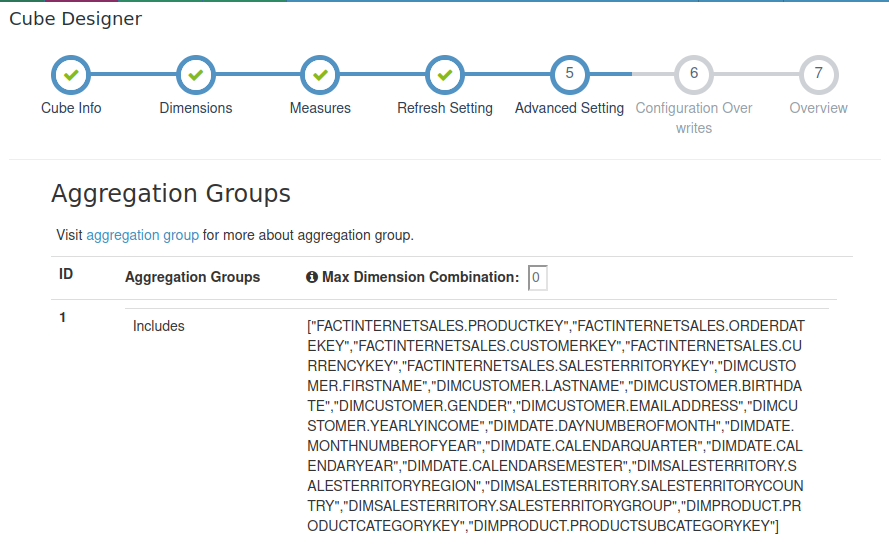


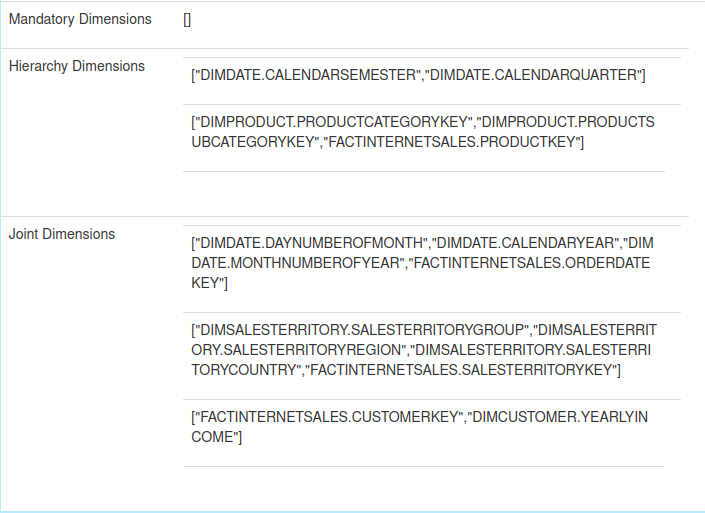


1. Agregación de las mediciones.



1. Actualizar configuraciones (por defecto)
2. Configuraciones avanzadas

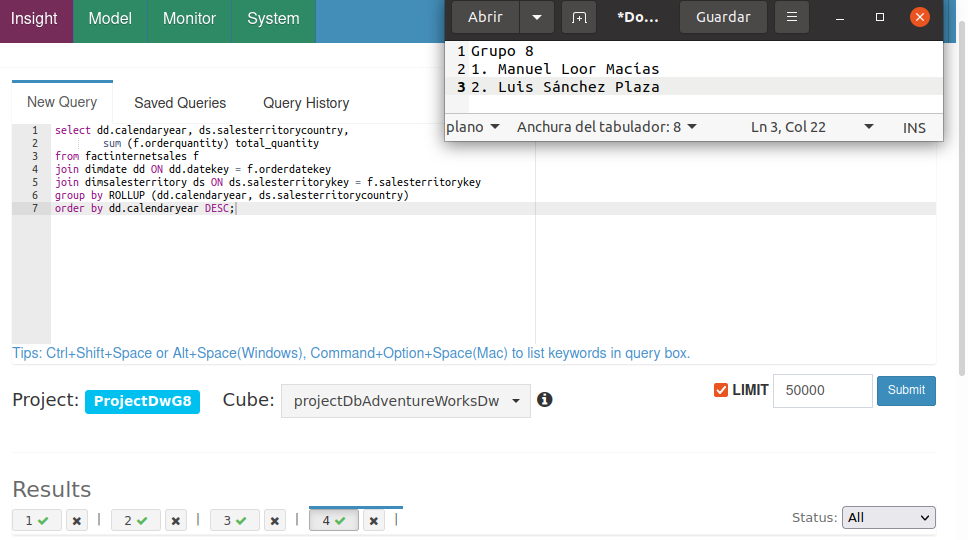


* Para las dimensiones conjunto (joint dimensions) se seleccionaron las dimensiones de las fechas, territorio, y la dimensión de ingreso anual de Customer, para poder clasificar la información de mejor manera en los gráficos.
* Para el motor de cubo (cube engine), se seleccionó el tipo Spark.

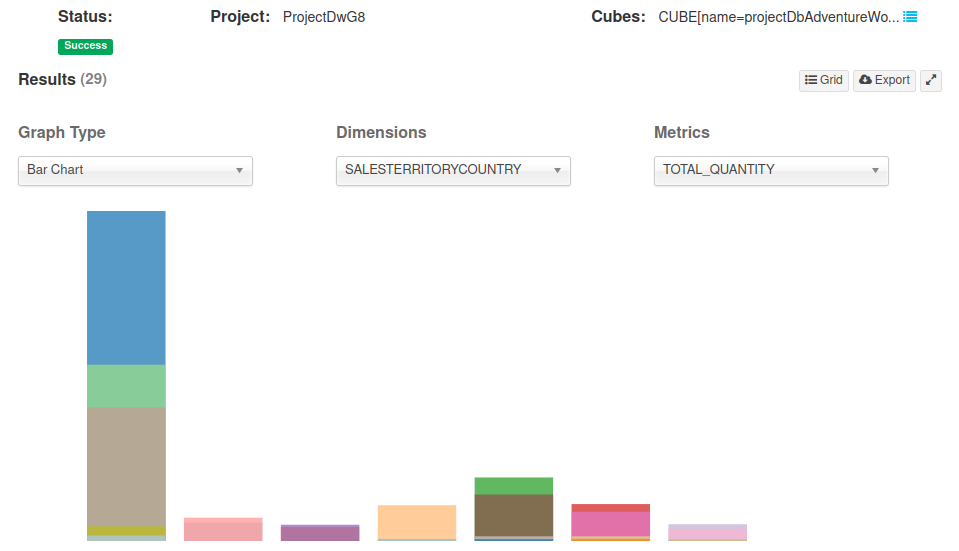


## Reportes

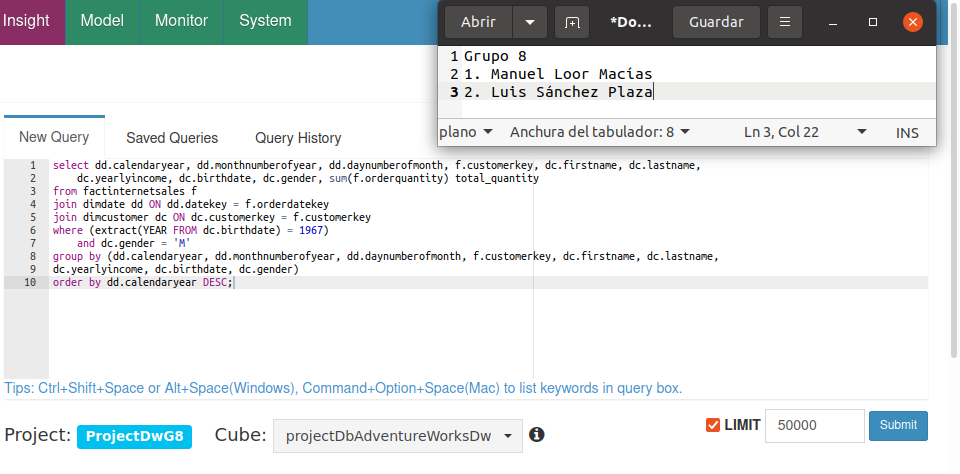
1. **Roll-up**
2. Sentencia SQL

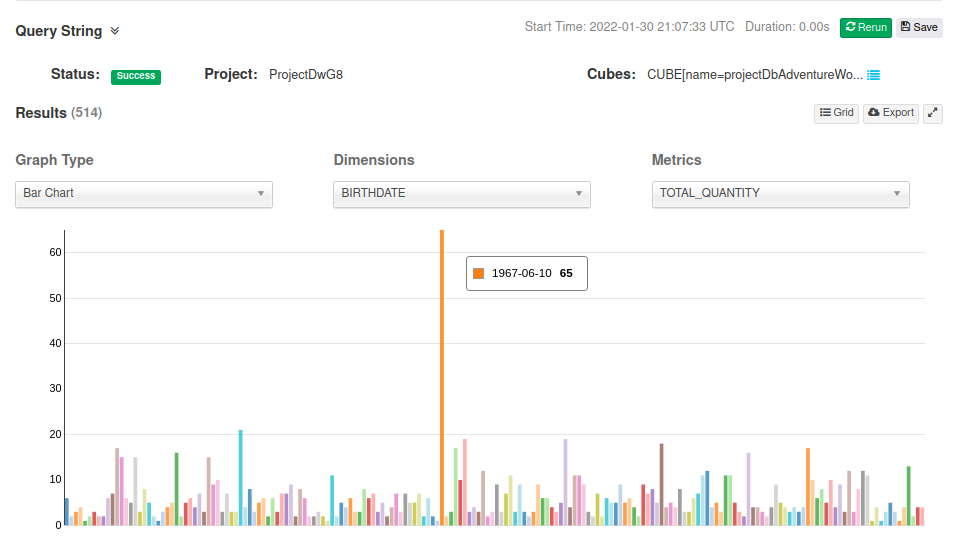


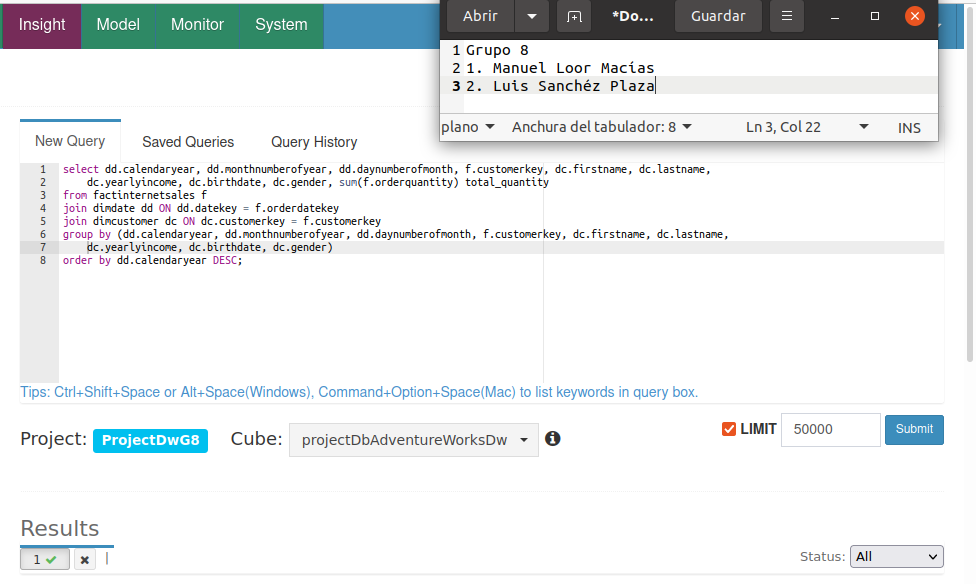
1. Visualización de Datos (Bar Chart)



1. **Dice**
2. Sentencia SQL



1. Visualización de datos (Bar Chart)
2. **Drill-down**
3. Sentencia SQL



1. Visualización de datos (Bar Chart)

