

**UNIVERSIDAD SAN CARLOS DE GUATEMALA**

**CENTRO UNIVERSITARIO DE OCCIDENTE**

**DIVISIÓN CIENCIAS DE LA INGENIERÍA**

**INGENIERÍA CIENCIAS Y SISTEMAS**



**LABORATORIO SEMINARIO DE SISTEMAS 2**

**NOVENO SEMESTRE**

**ING.: MAX ALEJANDRO CERNA**

**ESTUDIANTES:**

201830007 - Mario Moisés Ramírez Tobar

201730246 - José Carlos Soberanis Ramírez

**TEMA: “Proyecto final”**

**FECHA: 20 de mayo de 2,022**

# Herramientas utilizadas

Durante el análisis y diseño así como la realización del sistema, se utilizaron las siguientes herramientas con sus respectivas versiones:

- ETL: Pentaho - 9.2
- Sistema de base de datos: SQL Server - 15.0
- SQL management Studio - 18
- Visual Studio 2019
- Power Bi - 2.105.664

## Definición de diccionario de datos

### Descripción del sistema

Sistema de análisis de datos sobre los mundiales y sus estadísticas: Sistema de información que permite realizar consultas estadísticas sobre los datos recopilados de los últimos 3 mundiales con la finalidad de poder comprender cuáles son los valores más relevantes o que han sido determinantes para conocer quién es el ganador de cada mundial.

- Nombre de la base de datos: Mundial

Es de resaltar que para toda la lectura de los datos se utilizaron archivos csv, manipulados en excel, por lo que todas las fuentes de datos originales de datos se pueden hallar [aquí](#).

Nombre destino	Hoja - Campo origen	Campo destino	Descripción	Tipo de dato
DimTiempoGol	Tiempos - id	id_tiempo	Identificador del valor registrado en la dimensión	int
	Tiempos - minuto	minuto	Valor de 1 - 120 que indica el minuto que puede llegar a durar un partido en periodo normal	int
	Tiempos - extra	agregado	Valor de 1 - 8 que indica los minutos agregados/reposición	int
	Tiempos - es_tiempo_extra	tiempo_extra	Valor que determina si es tiempo extra o no	Boolean
	Tiempos - es_penal	es_penal	Valor que determina si es tiempo de penales	Boolean
	Tiempos - numero_de_penal	numero_penal	Valor que indica el número de penal si está en periodo de penales	int
DimTiempoPartido	Dias - id	id_tiempo_partido	Identificador para cada registro único de la tabla DimTiempoPartido	int
	Dias - año	año	Valor entero que identifica el año en el que se	int

			realizará un partido	
	Dias - dia	dia	Valor entero que representa el día (del mes) en el que se realizará un partido	int
	Dias - mes	mes	Valor entero que representa el mes (del año) en el que se realiza un partido	int
	Dias - dia_semana	dia_semana	Cadena que representa el día de la semana (Lunes - Domingo) en el que se realizó un partido	string
	Dias - fecha_equivalente	fecha_equivalente	Fecha equivalente a la unión de los atributos anteriores	Date
DimSeleccion	Secuencia - Añadida en transformación	id_selección	Identificador para cada registro único de la tabla DimSeleccion	int
	Países - país	nombre	Nombre de la selección	string (nchar)
	Selecciones en los mundiales - mundiales	mundiales	Cantidad de mundiales que ha disputado	int
	Selecciones en los mundiales - campeon	campeon	Cantidad de mundiales en los que ha quedado campeón	int

	Selecciones en los mundiales - pts	pts	Cantidad de puntos totales que ha tenido en los diferentes mundiales	int
	Selecciones en los mundiales - partidos_jugados	partidos_jugados	Cantidad de partidos que ha jugado en los diferentes mundiales	int
	Selecciones en los mundiales - ganados	ganados	Cantidad de partidos ganados en los diferentes mundiales	int
	Selecciones en los mundiales - empatados	empatados	Cantidad de partidos que ha empatado en los diferentes mundiales	int
	Selecciones en los mundiales - perdidos	perdidos	Cantidad de partidos que ha perdido en los diferentes mundiales	int
	Selecciones en los mundiales - goles a favor	goles_a_favor	Cantidad de goles totales que ha tenido a favor en los diferentes mundiales	int
	Selecciones en los mundiales - goles en contra	goles_contra	Cantidad de goles en contra que ha tenido en los diferentes mundiales	int
	Selecciones en los mundiales - diff	dif	Diferencia entre goles a favor y goles en contra	int
DimMundial	Añadido en transformación	id_mundial	Identificador para cada registro de los diferentes	int

			mundiales	
	Añadido en transformación	cede	Pais en el que se llevó a cabo el mundial	string (nchar)
	Añadido en transformación	año	Año en el que se llevó a cabo el mundial	int
	Añadido en transformación	participantes	Cantidad de participantes que hubo en el mundial	int
	Añadido en transformación	participantes_fase_final	Cantidad de participantes que hubo desde octavos en adelante	int
	Calculado de Penales	participantes_penales	Cantidad de participantes que se fueron a tanda de penales	int
	Calculado de Penales	penales_marcados	Cantidad de penales marcados en el mundial	int
	Calculado de Penales	penales_lanzados	Cantidad de penales lanzados en el mundial	int
	Añadido en transformación	id_seleccion_campeon	Identificador de la selección campeona del mundial	int
DimEquipo	Añadido en transformación	id_equipo	Identificador para cada registro de un equipo	int
	Paises - pais	equipo	Nombre de la selección	string (nchar)

	Jugadores - dónde posición es DT, filtrado en Transformación	DT	Director técnico del equipo	string (nchar)
	Participantes - año	año	Año en que se formó el equipo	int
	DimSelección - id_selección	seleccion	Referencia a qué selección representaba dicho equipo	int
	Participantes - partidos_jugados	partidos_jugados	Cantidad de partidos que jugó determinado el equipo	int
	Participantes - goles_favor	goles_favor	Cantidad de goles a favor que tuvo determinado equipo	int
	Participantes - goles_contra	goles_contra	Cantidad de goles en contra que tuvo determinado equipo	int
	Participantes - amarillas	amarillas	Cantidad de tarjetas amarillas que tuvo determinado equipo	int
	Participantes - rojas	rojas_directas	Cantidad de tarjetas rojas que tuvo determinado equipo	int
	Países - código	abreviatura	Código de país que identifica al equipo	string (nchar)

	Participantes - confederación	confederación	Confederación a la que pertenece determinado equipo	string (nchar)
	Calculado a través del año del equipo y un partido que haya jugado ese año en fase de grupos	grupo	Grupo en el que se encontró determinado equipo durante la fase de grupos	char
	Calculado en Transformación	es_local	Valor que indica si dicho equipo era local en el año respectivo	Boolean
	DimMundial - id_mundial	id_mundial	Referencia a mundial en el que se armó dicho equipo	int
DimPartido	Partidos - no	id_partido	Identificador para cada registro de partido que se realizó en un mundial	int
	DimTiempoPartido - id_tiempo	id_tiempo_partido	Referencia a la dimensión tiempo en que se llevó a cabo dicho partido	int
	DimEquipo - id_equipo	id_equipo1	Referencia a dimensión equipo para el equipo 1	int
	DimEquipo - id_equipo	id_equipo2	Referencia a dimensión equipo para el equipo 2	int
	Calculado en transformación en base a	goles_equipo2	Cantidad de goles que se marcaron en el encuentro	int



	archivo Goles por partido y penal		a favor del equipo 2	
	Calculado en transformación en base a archivo Goles por partido	goles_equipo1	Cantidad de goles que se marcaron en el encuentro a favor del equipo 1	int
	Partidos - fase	fase	Fase en la que se llevó a cabo el partido	string (nchar)
	Partidos - grupo	grupo	Grupo en el que se llevó a cabo dicho partido (si aplica)	char
	Calculado en transformación en base a archivo penales	penales_ anotados_equipo 1	Penales que anotó el equipo 1 (si aplica)	int
	Calculado en transformación en base a archivo penales	penales_ anotados_equipo 2	Penales que anotó el equipo 2 (si aplica)	int
	Calculado en transformacion en base a archivo penales	penales_fallados_equipo1	Penales que falló el equipo 1 (si aplica)	int
	Calculado en transformación en base a archivo penales	penales_fallados_equipo2	Penales que falló el equipo 2 (si aplica)	int
DimJugador	Añadido en transformación	id_jugador	Identificador para cada registro de un jugador	int

	Todos jugadores - jugador	nombre	Nombre de cada jugador	int
	Todos jugadores - posición	posición	Posición en la que juega dicho jugador	string (nchar)
	Todos jugadores - edad	edad	Edad que tiene el jugador	int
	Todos jugadores - pj	pj	No de la camiseta que portaba el jugador para el equipo	int
	DimEquipo - id_equipo	equipo	Referencia al equipo para el que jugaba dicho jugador (selección y año)	int
	Todos jugadores - club	club	Club para el que jugaba dicho jugador en el mundial respectivo	string (nchar)
Fact_Gol	DimPartido - id_partido	id_partido	Referencia al partido en el que se anotó dicho gol	int
	DimEquipo - id_equipo	id_equipo	Referencia al equipo que anotó dicho gol	int
	DimJugador - id_jugador	id_jugador	Referencia al jugador que anotó dicho gol	int
	DimTiempoGol - id_tiempo	id_tiempo	Referencia a la DimTiempoGol en el que se anotó el gol	int

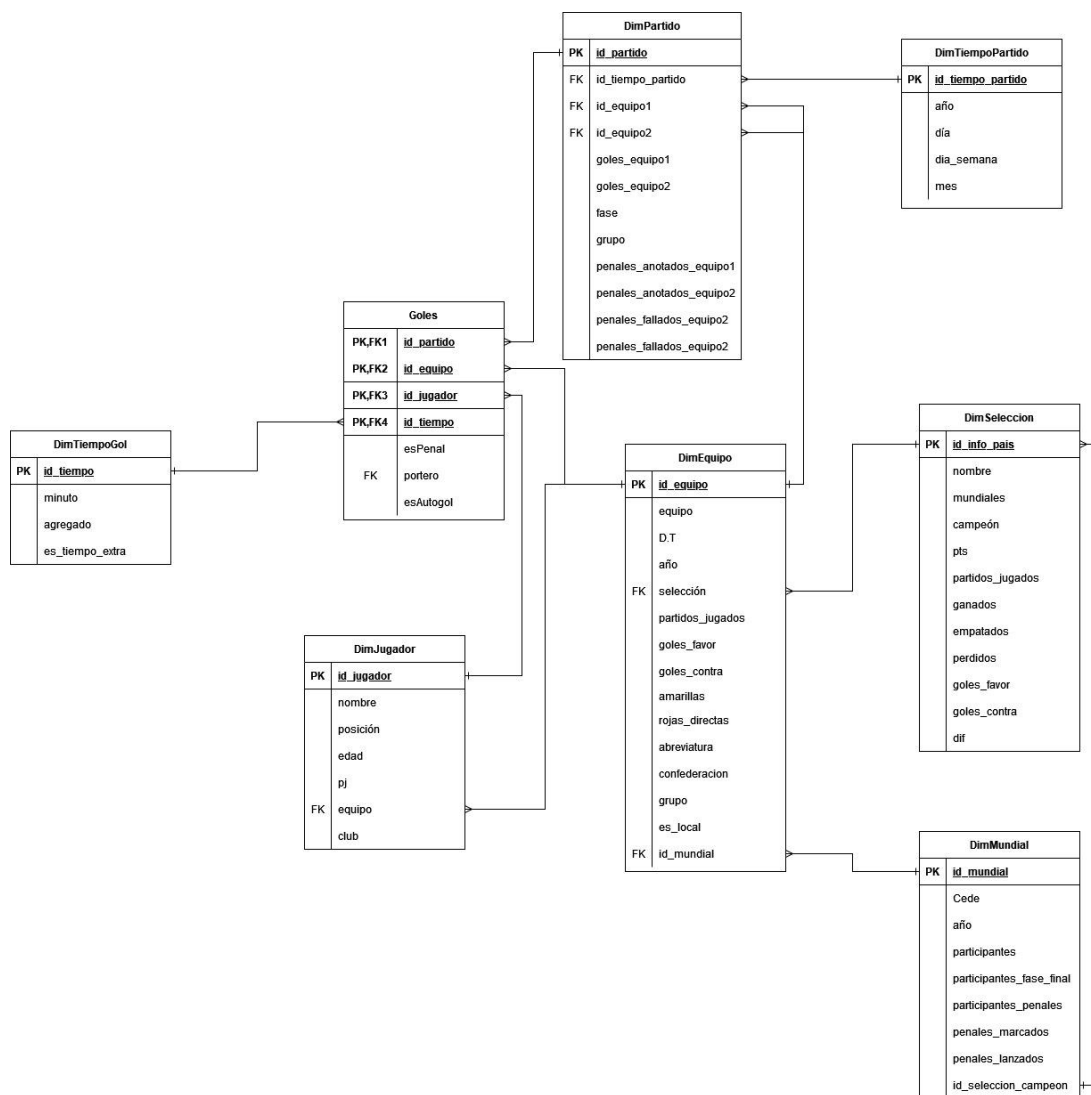
	Goles por partido/Penal - es_penal	esPenal	Valor que identifica si dicho gol fue penal o no	boolean
	Penal - portero	portero	Nombre del portero que atajó (Si aplica, solo para penales)	string (nchar)
	Goles por partido/Penal - es_autogol	es_autogol	Valor que identifica si dicho gol fue autogol o no	Boolean

# Definición de arquitectura de DW y procesos de ETL identificados

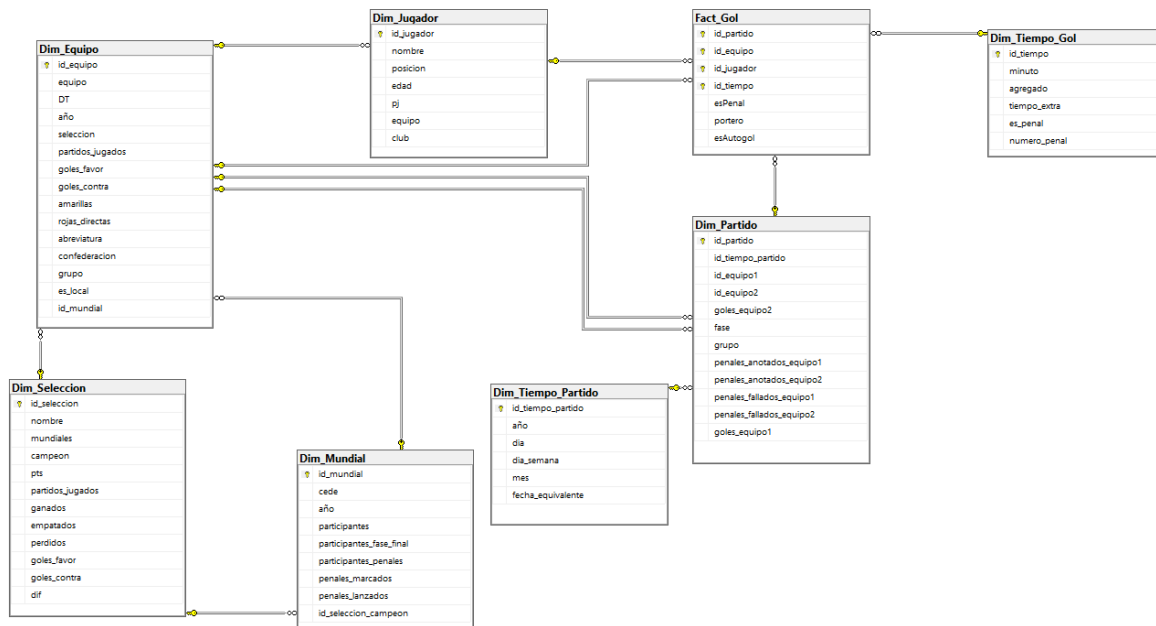
## Modelo de datos

### Modelo inicial

Se planteó en primera instancia las relaciones que existirán en el modelo copo de nieve de la base de datos utilizada para el DataWarehouse.



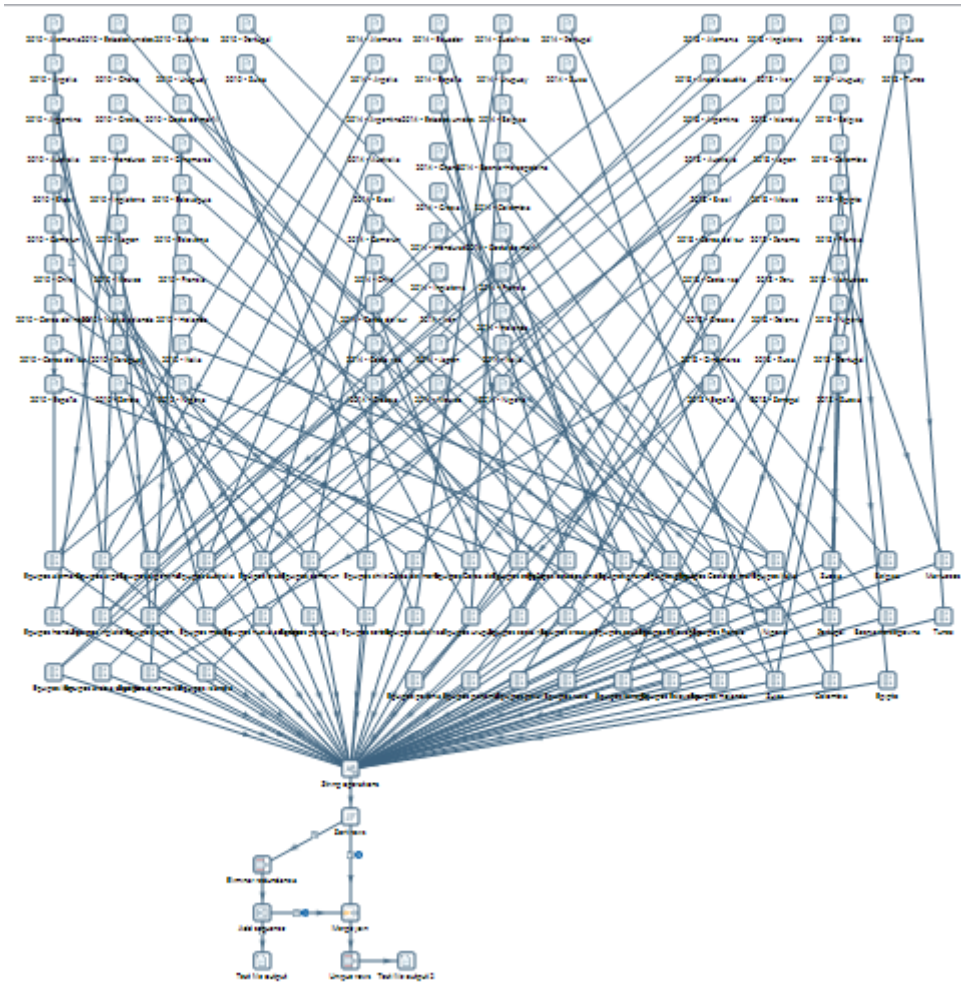
## Modelo de datos - Copo de nieve



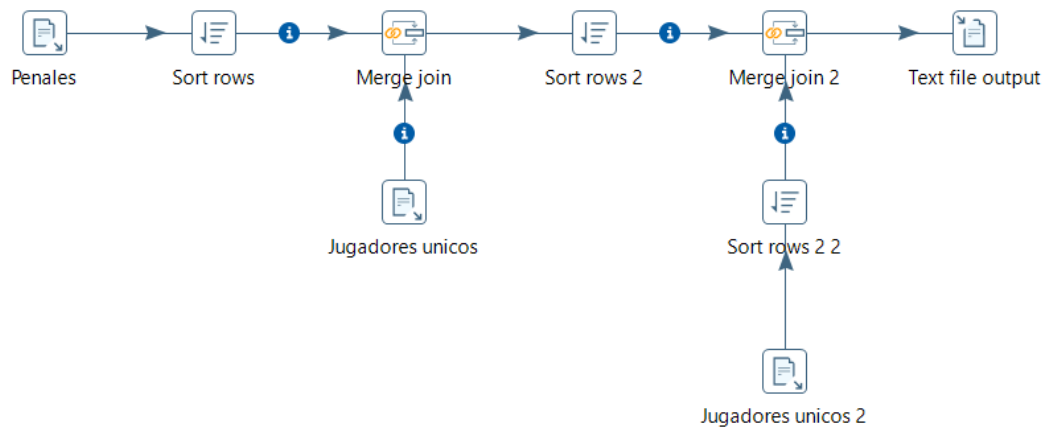
## Pentaho data integration - Procesos ETL

- Normalización de datos
  - Agregar una tabla de 'Países', en la que se encuentra su abreviatura y su nombre, siendo ahora, en cada tabla extra referenciado a su abreviatura.
  - Agregar una tabla de 'Goles', en la que se encuentra quien hizo el gol, el tiempo, el tiempo 'agregado' si hubo y su tipo; haciendo referencia al No. de partido en el que se ejecutó.
  - Los registros de penales, teniendo un registro por cada 'penal' y no por cada tanda de penales.
- Lectura de cada plantilla de 2010, 2014 y 2018, ordenando por equipo (país) y posteriormente ordenando cada uno de ellos por nombre.
  - Este con el fin de eliminar jugadores repetidos, verificar un único registro del nombre del jugador y obtener una 'tabla' de jugadores.
- Lectura de tabla de 'goles', ordenando por equipo (país) y posteriormente ordenando por nombre, de esta manera verificando que los nombres repetidos sean iguales (verificar la integridad de la data) haciendo referencia ahora a la 'tabla' de jugadores.
- Normalización

- Normalización de jugadores
  - Eliminar redundancia uniendo en un mismo lugar a todos los jugadores de todos los equipos para los que se tenía acceso.

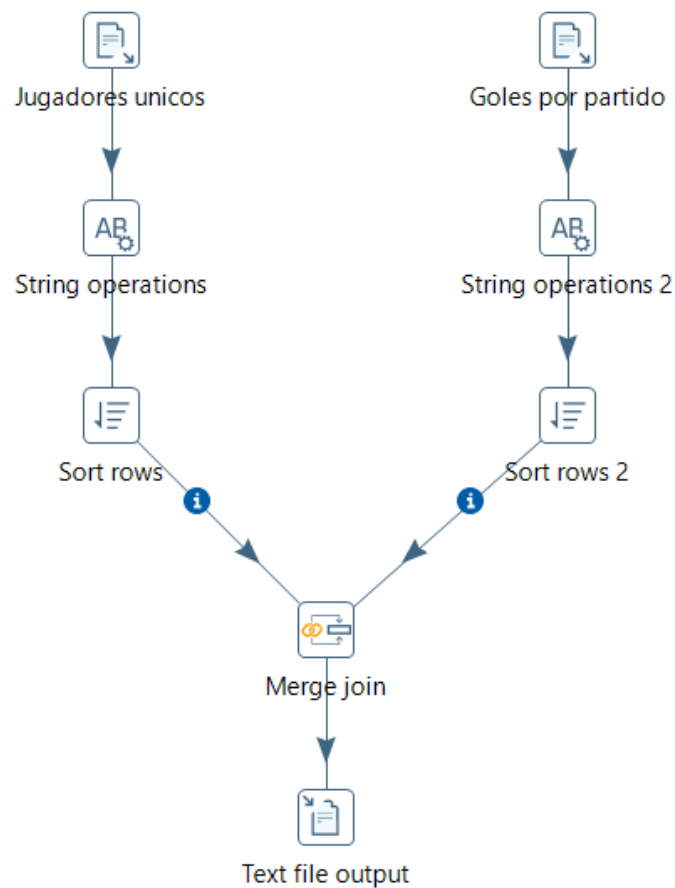


- Normalización penales
  - Añadir la referencia respectiva a qué jugador había realizado qué tiro, qué portero había atajado qué tiro y estandarizar nombres en base a la tabla ‘jugadores’



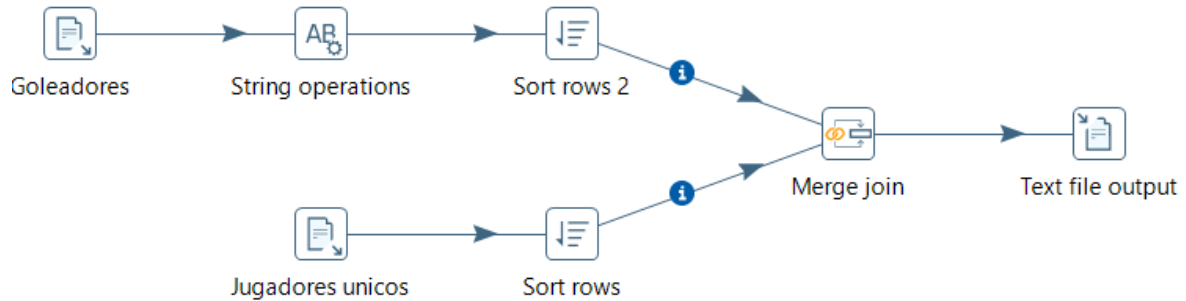
- Normalización goles

- Añadir la referencia respectiva a qué jugador había realizado el gol estandarizando nombres en base a la tabla 'jugadores'



- Normalización goleadores

- Añadir la referencia respectiva a qué jugador poseía qué estadísticas en cuánto a la cantidad de goles y partidos jugados.



- Migración de datos

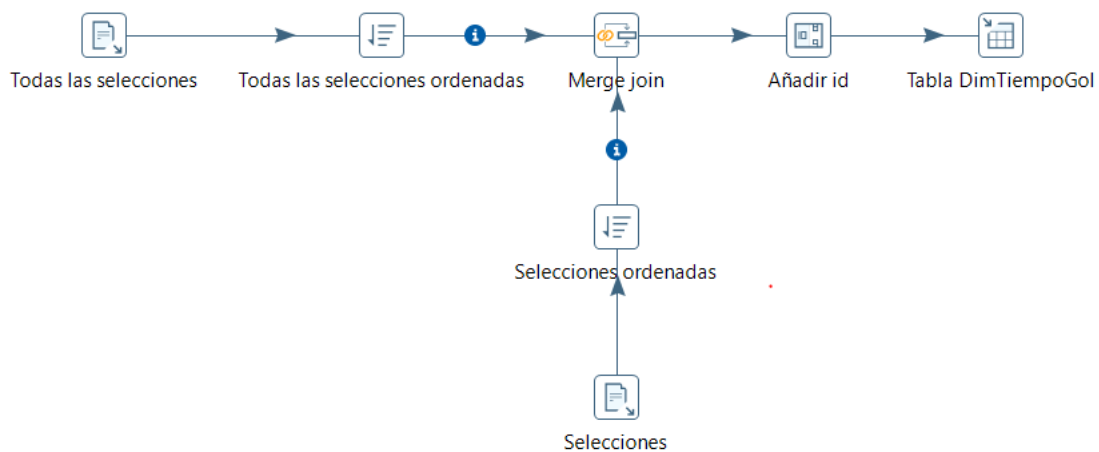
- DimTiempoPartido

- Se realizó un csv con los posibles valores que debían registrarse en la base de datos, exportando al modelo concreto de la dimensión.

- DimTiempoGol

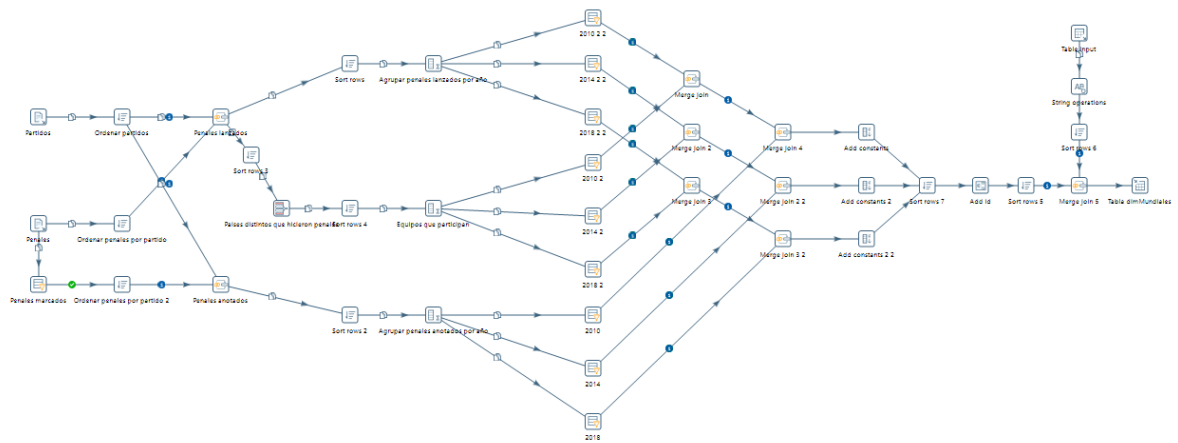
- Se realizó un csv con los posibles valores que debían registrarse en la base de datos, exportando al modelo concreto de la dimension.

- DimSeleccion

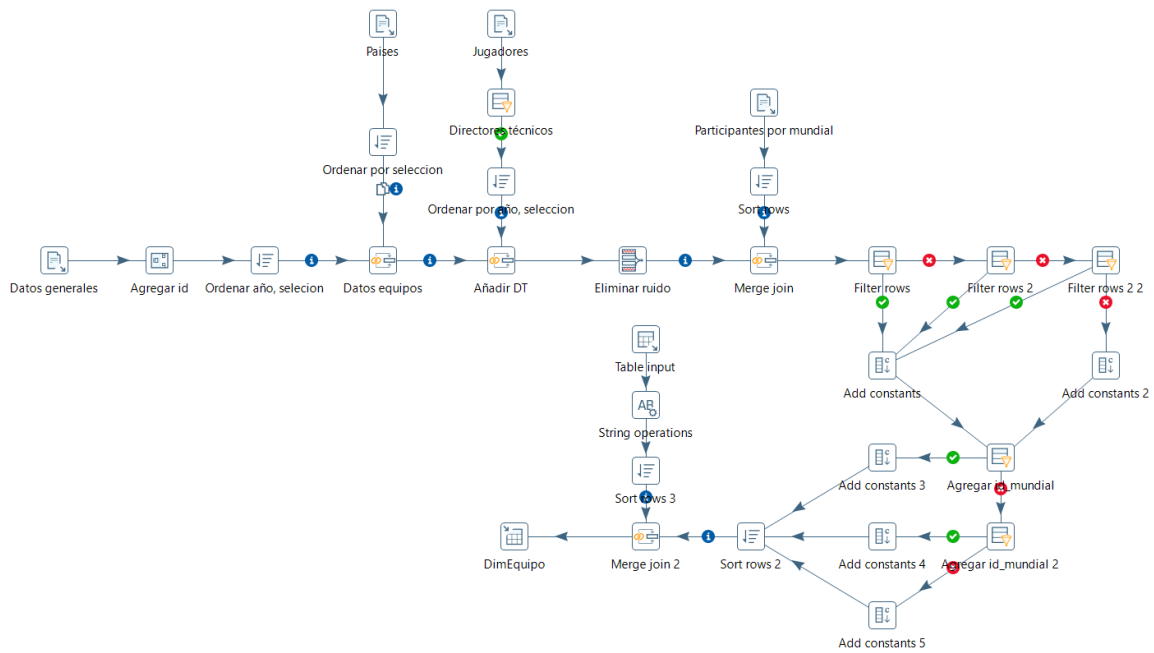


- DimMundial

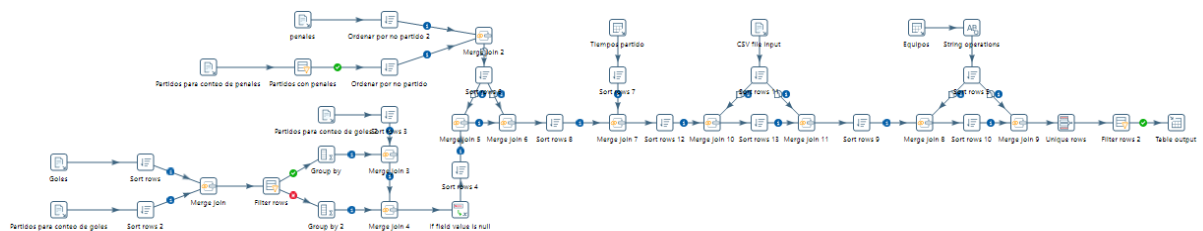




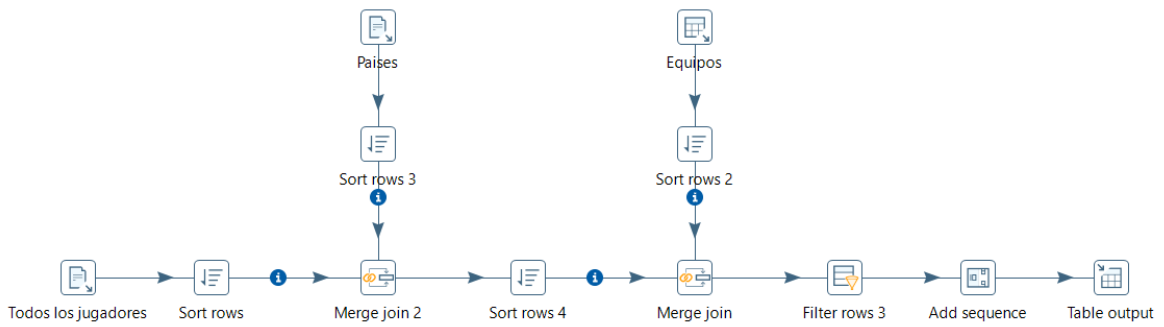
## ○ DimEquipo



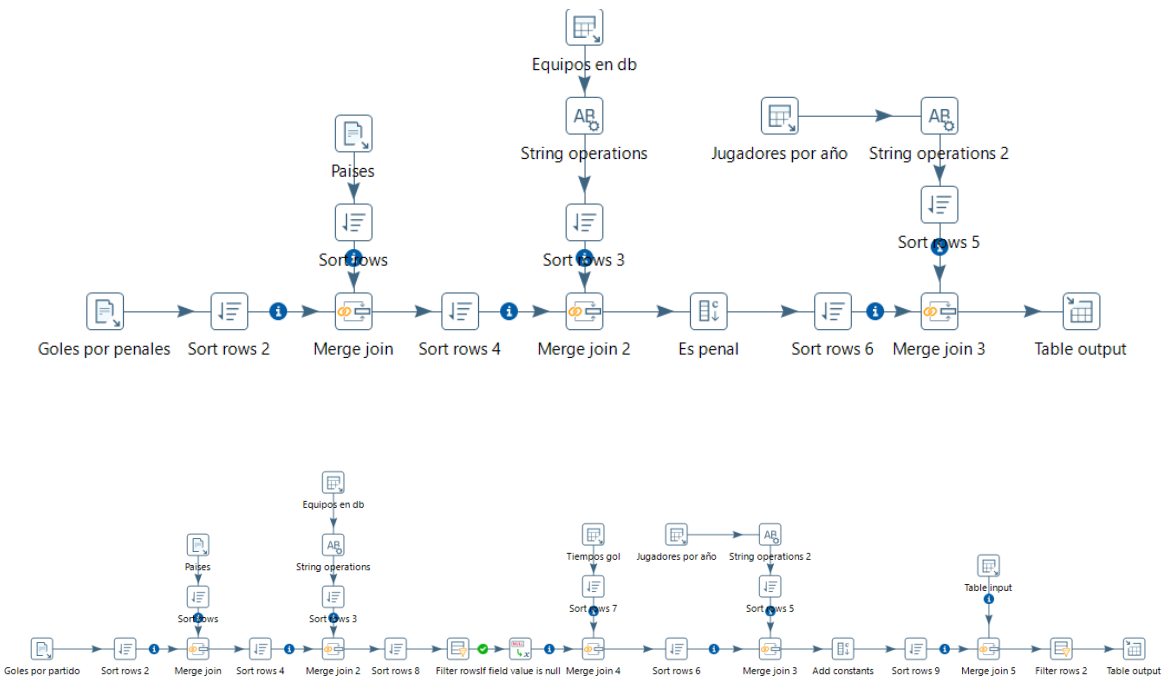
## ○ DimPartido



## ○ DimJugador



## ○ FactGoles



## Cronograma de actividades

Actividad	Periodo de tiempo				
	19/04 - 24/04	25/04 - 01/05	02/05 - 08/05	09/05 - 15/05	15/05 - 20/05
Análisis y diseño (Redacción inicial)					
Estudio y limpieza de los datos					
Transformaciones de datos mediante el uso de ETL					
Diseño copo de nieve/estrella para el datawarehouse					
Dashboards, queries, power bi					

## Definición de Cubos, perspectivas y KPI 's.

### Cubos

- Cubo Mundial: Cubo utilizado versatilmente para aprovechar al máximo las relaciones que existen en sus tablas de dimensiones, así como el establecer distintas tablas de hechos, según sea necesario.
  - FactGol (Tabla de hechos)
  - DimEquipo (Dimensión) que en determinada situación se utilizó como FactEquipo (Tabla de hechos) para determinar cuántas tarjetas amarillas y rojas había por confederación y equipo.
  - DimJugador (Dimensión) que en determinada situación se utilizó como FactJugador (Tabla de hechos) para determinar cuántos jugadores eran ....

### Perspectivas

- Confederación

- Dimensiones
  - DimJugador
  - DimPartido
  - DimSeleccion
  - DimMundial
- KPI's
  - Top amarillas
  - Top rojas
  - Total rojas amarillas
  - Confederación menor goles
- País
  - Dimensiones
    - DimPartido
    - DimEquipo
    - DimTiempoGol
    - DimJugador
  - KPI's
    - Confederación menor goles
    - Top amarillas pais
    - Top rojas pais
    - Top rojas amarillas pais
    - Selección menor goles
- Club
  - Dimensiones
    - DimJugador
    - DimEquipo
    - DimGoles
  - KPI's
- Torneo
  - Dimensiones
    - DimJugador
  - KPI's

## KPI's

- Top amarillas
  - Indicador: 20%
  - Lectura: posee más del 20% de tarjetas amarillas
- Top rojas
  - Indicador: 20%
  - Lectura: posee más del 20% de tarjetas rojas
- Total rojas amarillas
  - Indicador: 10%
  - Lectura: posee más del 10% de tarjetas rojas y amarillas
- Confederación menor goles
  - Indicador: 1
  - Lectura: posee más de 1 gol de promedio son verdes, el resto rojos
- Top\_Amarillas\_Pais
  - Indicador: 3%
  - Lectura: El país posee más del 3% de tarjetas amarillas
- Top\_Rojas\_Pais
  - Indicador: 3%
  - Lectura: El país posee más del 3% de tarjetas rojas
- Top\_Rojas\_Amarillas\_Pais
  - Indicador: 3%
  - Lectura: El país posee más del 3% de tarjetas
- Seleccion\_Menor\_Goles
  - Indicador: 1
  - Lectura: El país que posee más de 1 gol de promedio verde, resto rojos

## Extras

### Campos calculados

- Amarillas rojas: La suma de tarjetas amarillas y rojas
- Gol país partido: El promedio de goles por partido