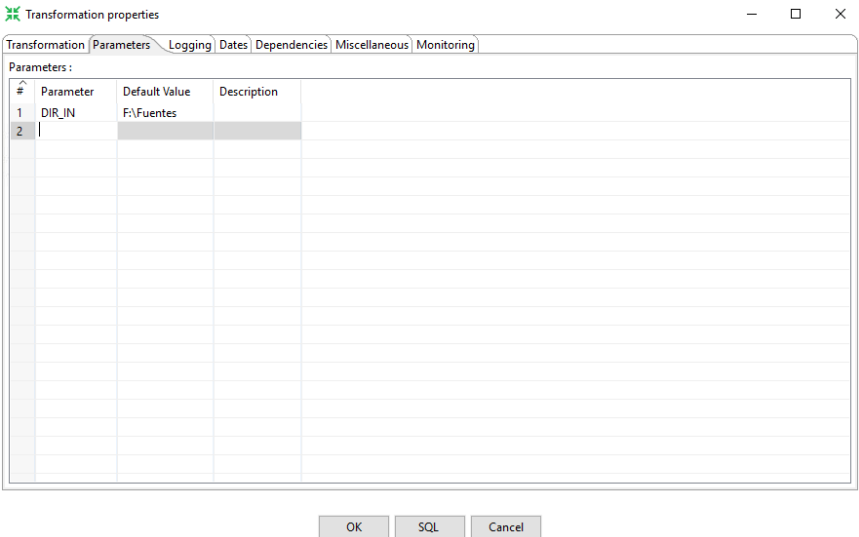


2.2. Ejercicio 1. Carga de un fichero en la base de datos (10%)

2.2.1. Configuración de Spoon



2.2.2. Análisis de las fuentes de datos

2.2.3. Creación de una tabla en la base de datos en la capa de Staging

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC

1 SELECT *FROM dbo.stg_athletes

	name character varying (100)	sex character varying (1)	age integer	height integer	weight integer	team character varying (50)	noc character varying (5)	year integer	season character varying (6)	city character varying (50)	sport character varyin
1	A Dijiang	M	24	180	80	China	CHN	1992	Summer	Barcelona	Basketball
2	A Lamusi	M	23	170	60	China	CHN	2012	Summer	London	Judo
3	Gunnar Nielsen Aaby	M	24	[null]	[null]	Denmark	DEN	1920	Summer	Antwerpen	Football
4	Edgar Lindenu Aabye	M	34	[null]	[null]	Denmark/Sweden	DEN	1900	Summer	Paris	Tug-Of-War
5	Christine Jacobs Aaftink	F	21	185	82	Netherlands	NED	1988	Winter	Calgary	Speed Skating
6	Christine Jacobs Aaftink	F	21	185	82	Netherlands	NED	1988	Winter	Calgary	Speed Skating
7	Christine Jacobs Aaftink	F	25	185	82	Netherlands	NED	1992	Winter	Albertville	Speed Skating
8	Christine Jacobs Aaftink	F	25	185	82	Netherlands	NED	1992	Winter	Albertville	Speed Skating
9	Christine Jacobs Aaftink	F	27	185	82	Netherlands	NED	1994	Winter	Lillehammer	Speed Skating
10	Christine Jacobs Aaftink	F	27	185	82	Netherlands	NED	1994	Winter	Lillehammer	Speed Skating
11	Per Knut Aaland	M	31	188	75	United States	USA	1992	Winter	Albertville	Cross Country
12	Per Knut Aaland	M	31	188	75	United States	USA	1992	Winter	Albertville	Cross Country
13	Per Knut Aaland	M	31	188	75	United States	USA	1992	Winter	Albertville	Cross Country
14	Per Knut Aaland	M	31	188	75	United States	USA	1992	Winter	Albertville	Cross Country
15	Per Knut Aaland	M	33	188	75	United States	USA	1994	Winter	Lillehammer	Cross Country
16	Per Knut Aaland	M	33	188	75	United States	USA	1994	Winter	Lillehammer	Cross Country
17	Per Knut Aaland	M	33	188	75	United States	USA	1994	Winter	Lillehammer	Cross Country
18	Per Knut Aaland	M	33	188	75	United States	USA	1994	Winter	Lillehammer	Cross Country
19	John Aalberg	M	31	183	72	United States	USA	1992	Winter	Albertville	Cross Country
20	John Aalberg	M	31	183	72	United States	USA	1992	Winter	Albertville	Cross Country
21	John Aalberg	M	31	183	72	United States	USA	1992	Winter	Albertville	Cross Country
22	John Aalberg	M	31	183	72	United States	USA	1992	Winter	Albertville	Cross Country
23	John Aalberg	M	33	183	72	United States	USA	1994	Winter	Lillehammer	Cross Country
24	John Aalberg	M	33	183	72	United States	USA	1994	Winter	Lillehammer	Cross Country
25	John Aalberg	M	33	183	72	United States	USA	1994	Winter	Lillehammer	Cross Country

Total rows: 1000 of 271116 Query complete 00:00:00.881 Ln 1, Col 30

2.2.4. Carga del fichero *athletes.csv*

Spoon - Athlete_Transformation

File Edit View Action Tools Help

View Design

Search

Transformations

Athlete_Transformation

Run configurations

Database connections

Steps

Hops

Partition schemas

Slave server

Kettle cluster schemas

VFS Connections

Data Services

Hadoop clusters

Welcome!

JOB_DM_PARTICIPAT...

Transformación Di...

Athlete_Transform...

100%

CSV file input

Value mapper

Select values

Split fields

Table output

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL-UOC* X

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL-UOC

No limit

Query Query History

1 SELECT

2 name, age, height, weight, year, sport,event

3 FROM

4 dbo.stg_athletes

Data Output Messages Notifications

name

age

height

weight

year

sport

event

1	A Dijiang	24	180	80	1992	Basketball	Basketball Men's Basketball
2	A Lamusi	23	170	60	2012	Judo	Judo Men's Extra-Lightweight
3	Gunnar Nielsen Aaby	24	[null]	[null]	1920	Football	Football Men's Football
4	Edgar Lindenau Aabye	34	[null]	[null]	1900	Tug-Of-War	Tug-Of-War Men's Tug-Of-War
5	Christine Jacobsa Aaftink	21	185	82	1988	Speed Skating	Speed Skating Women's 500 metres
6	Christine Jacobsa Aaftink	21	185	82	1988	Speed Skating	Speed Skating Women's 1,000 metres
7	Christine Jacobsa Aaftink	25	185	82	1992	Speed Skating	Speed Skating Women's 500 metres
8	Christine Jacobsa Aaftink	25	185	82	1992	Speed Skating	Speed Skating Women's 1,000 metres
9	Christine Jacobsa Aaftink	27	185	82	1994	Speed Skating	Speed Skating Women's 500 metres
10	Christine Jacobsa Aaftink	27	185	82	1994	Speed Skating	Speed Skating Women's 1,000 metres
11	Per Knut Aaland	31	188	75	1992	Cross Country Skiing	Cross Country Skiing Men's 10 kilometres
12	Per Knut Aaland	31	188	75	1992	Cross Country Skiing	Cross Country Skiing Men's 50 kilometres
13	Per Knut Aaland	31	188	75	1992	Cross Country Skiing	Cross Country Skiing Men's 10/15 kilometres Pursuit
14	Per Knut Aaland	31	188	75	1992	Cross Country Skiing	Cross Country Skiing Men's 4 x 10 kilometres Relay
15	Per Knut Aaland	33	188	75	1994	Cross Country Skiing	Cross Country Skiing Men's 10 kilometres
16	Per Knut Aaland	33	188	75	1994	Cross Country Skiing	Cross Country Skiing Men's 30 kilometres
17	Per Knut Aaland	33	188	75	1994	Cross Country Skiing	Cross Country Skiing Men's 10/15 kilometres Pursuit
18	Per Knut Aaland	33	188	75	1994	Cross Country Skiing	Cross Country Skiing Men's 4 x 10 kilometres Relay
19	John Aalberg	31	183	72	1992	Cross Country Skiing	Cross Country Skiing Men's 10 kilometres
20	John Aalberg	31	183	72	1992	Cross Country Skiing	Cross Country Skiing Men's 50 kilometres
21	John Aalberg	31	183	72	1992	Cross Country Skiing	Cross Country Skiing Men's 10/15 kilometres Pursuit
22	John Aalberg	31	183	72	1992	Cross Country Skiing	Cross Country Skiing Men's 4 x 10 kilometres Relay
23	John Aalberg	33	183	72	1994	Cross Country Skiing	Cross Country Skiing Men's 10 kilometres

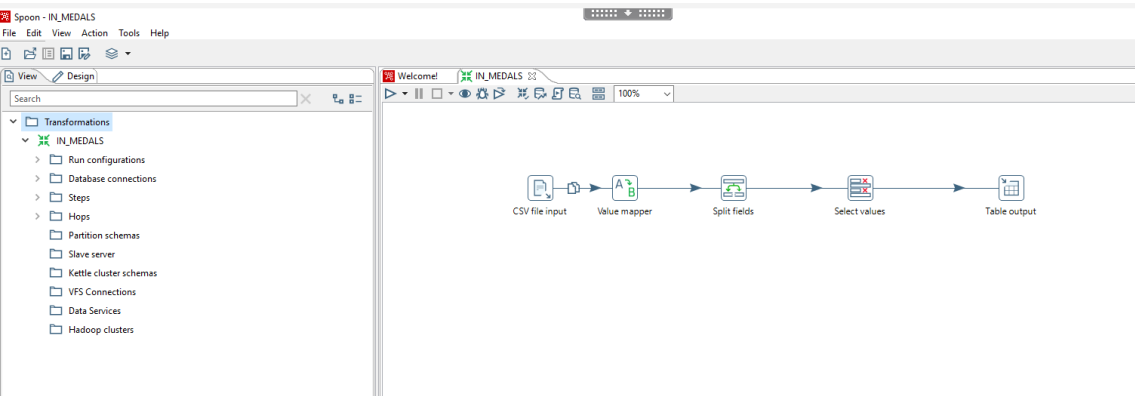
Total rows: 1000 of 271116 Query complete 00:00:00.515

Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

2.2.5. Ejercicios a realizar



Value mapper

Step name:

Fieldname to use:

Target field name (empty=overwrite):

Default upon non-matching:

Field values:

#	Source value	Target value
1	NA	

Split fields

Step name:

Field to split:

Delimiter:

Enclosure:

Fields

#	New field	ID	Remove ID?	Type	Length	Precision	Format	Group	Decimal	Currency	Nullif	Default	Trim type
1	year		N	String									none
2	month		N	String									none
3	day		N	String									none

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

Table output

Step name:

Connection:

Target schema:

Target table:

Commit size:

Truncate table: ☒

Ignore insert errors: ☐

Specify database fields: ☒

Main options Database fields

Fields to insert:

#	Table field	Stream field
1	name	Name
2	event	Event
3	medal	Medal
4	day	day
5	month	month
6	year	year

Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

The screenshot shows the PostgreSQL interface with a query window and a results table. The query is:

```
1 SELECT
2   name, event, medal, day, month, year
3 FROM
4   dbo.stg_medals
```

The results table has the following columns: name, event, medal, day, month, year. The data is as follows:

name	event	medal	day	month	year
Samuel Aguiar	Rowing Mens Lightweight Coxless Fours				
Jos Luiz Aguiar e Ramalho	Handball Mens Handball				
Barry Aguibou	Boxing Mens Flyweight				
Claudio S. guila	Rowing Mens Coxless Pairs				
Judith guila Hernandez	Basketball Womens Basketball				
Judith guila Hernandez	Basketball Womens Basketball				
Alberto Martin Aguiar Suarez	Athletics Mens 4 x 400 metres Relay				
Alessandra Aguiar Morn	Athletics Womens Marathon				
Alessandra Aguiar Morn	Athletics Womens Marathon				
Alessandra Aguiar Morn	Athletics Womens Marathon				
Andreas Aguiar	Gymnastics Mens Individual All-Around				
Andreas Aguiar	Gymnastics Mens Team All-Around				
Andreas Aguiar	Gymnastics Mens Floor Exercise				
Andreas Aguiar	Gymnastics Mens Horse Vault				
Andreas Aguiar	Gymnastics Mens Parallel Bars				
Andreas Aguiar	Gymnastics Mens Horizontal Bar				
Andreas Aguiar	Gymnastics Mens Rings				
Andreas Aguiar	Gymnastics Mens Pommel Horse				
Andrey Alberto Aguiar Komissarov	Swimming Mens 200 metres Breaststroke				
Andrey Alberto Aguiar Komissarov	Swimming Mens 100 metres Butterfly				
Andrey Alberto Aguiar Komissarov	Swimming Mens 200 metres Butterfly				
Andrey Alberto Aguiar Komissarov	Swimming Mens 400 metres Individual Medley				

2.3. Ejercicio 2. Análisis de la calidad de los datos (5%)

Tabla stg_athletes

- Determinación Nº registros

The screenshot shows the PostgreSQL interface with a query window and a results table. The query is:

```
1 SELECT
2   count(1) as reg_athletes
3 FROM
4   dbo.stg_athletes
```

The results table has the following columns: reg_athletes. The data is as follows:

reg_athletes
271116

- Calcular numero de registros que no tienen algún valor en los campos name, age year o sport

Query Query History

```

22 );
23 sport_field as (
24 SELECT 'stg_athletes' as table_name, count(1) as sport_null
25 FROM
26 dbo.stg_athletes
27 WHERE
28 sport is null
29 )
30 SELECT
31 name_null, age_null, year_null, sport_null
32 FROM
33 name_field, age_field, year_field, sport_field
34 WHERE
35 name_field.table_name = age_field.table_name and
36 name_field.table_name = year_field.table_name and
37 name_field.table_name = sport_field.table_name

```

Data Output Messages Notifications

	name_null bigint	age_null bigint	year_null bigint	sport_null bigint
1	0	9474	0	0

- Análisis columnas name, age, year y sport

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC* x

Query Query History

```

1 SELECT
2 min(length(name)) as min_length, max(length(name)) as max_length
3 FROM
4 dbo.stg_athletes

```

Data Output Messages Notifications

	min_length integer	max_length integer
1	2	100

Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC* x

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC

Query Query History

```
1 SELECT distinct name
2 FROM
3 dbo.stg_athletes
```

Data Output Messages Notifications

	name character varying (100)	
1	Frantiek Konvika	
2	Roel Hubertus Braas	
3	Adolfo Belmonte Heredia	
4	Frederick Robert "Robin" Lindsay	
5	Josef Hantych	
6	Edna Roxanne "Roxy" Atkins (Smith-, -Campbell, -Andersen)	
7	Andrei Cadar	
8	Mauricio de la Lama	
9	Ricardo Burgos Spangen	
10	David Evans	
11	Diego Mularoni	
12	Idora Hegel	
13	Chung Pao-Ming	
14	Corinne Miller (-von Rohr)	
15	Susan Cushman	
16	Ern Kolczonay	
17	Francis James "Frank" Donovan	
18	William Cesar "Willie" Farrell	
19	Maurice Apeang	
20	Chen Hsiao-Nan	
21	Rugil Mileiyt	
22	Lesley Attwell	
23	Eino Johannes Evert Kentt	
24	Roberto Brunamonti	

Total rows: 1000 of 134731 Query complete 00:00:00.367

The screenshot shows a PostgreSQL client window titled "SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC". The query editor contains the following SQL query:

```
1 SELECT
2 min(length(sport)) as min_length, max(length(sport)) as max_length
3 FROM
4 dbo.stg_athletes
```

The "Data Output" tab is active, displaying the results of the query in a table:

	min_length integer	max_length integer
1	4	25

Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC* X

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC

Query Query History

```
1 SELECT
2 distinct sport
3 FROM dbo.stg_athletes
```

Data Output Messages Notifications

	sport character varying (50)
1	Aeronautics
2	Alpine Skiing
3	Alpinism
4	Archery
5	Art Competitions
6	Athletics
7	Badminton
8	Baseball
9	Basketball
10	Basque Pelota
11	Beach Volleyball
12	Biathlon
13	Bobsleigh
14	Boxing
15	Canoeing
16	Cricket
17	Croquet
18	Cross Country Skiing
19	Curling
20	Cycling
21	Diving





Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas





Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC* x

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC v



No limit v



Query Query History








1 SELECT



2 min(age) as min_value, max(age) as max_value

3 FROM

4 dbo.stg_athletes

Data Output Messages Notifications



	min_value integer 	max_value integer 
1	10	97

Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC* x

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC

Query Query History

```
2 distinct age
3 FROM
4 dbo.stg_athletes
5 ORDER BY
6 age
```

Data Output Messages Notifications

	age integer
1	10
2	11
3	12
4	13
5	14
6	15
7	16
8	17
9	18
10	19
11	20
12	21
13	22
14	23
15	24
16	25
17	26
18	27
19	28
20	29
21	30
22	31
23	32

Total rows: 75 of 75 Query complete 00:00:00.179

Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC* X

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC

Query Query History

```
1 SELECT
2 min(year) as min_value, max(year) as max_value
3 FROM
```

Data Output Messages Notifications

	min_value integer	max_value integer
1	1896	2016

Total rows: 1 of 1 Query complete 00:00:00.185

Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC* X

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC

Query Query History

```
1 SELECT
2 distinct year
3 FROM
4 dbo.stg_athletes
5 ORDER BY
6 year
```

Data Output Messages Notifications

	year integer
1	1896
2	1900
3	1904
4	1906
5	1908
6	1912
7	1920
8	1924
9	1928
10	1932
11	1936
12	1948
13	1952
14	1956
15	1960
16	1964
17	1968
18	1972
19	1976
20	1980
21	1984

Total rows: 35 of 35 Query complete 00:00:00.208

Ejercicio a realizar:

De forma análoga al análisis de la table *stg_athletes* se solicita el análisis de la calidad de los datos cargados en la tabla *stg_medals*.

1. Revisión de valores faltantes

Query		Query History	
1	SELECT		
2	COUNT(*) AS total,		
3	SUM(CASE WHEN name IS NULL THEN 1 ELSE 0 END) AS missing_name,		
4	SUM(CASE WHEN event IS NULL THEN 1 ELSE 0 END) AS missing_event,		
5	SUM(CASE WHEN medal IS NULL THEN 1 ELSE 0 END) AS missing_medal,		
6	SUM(CASE WHEN year IS NULL THEN 1 ELSE 0 END) AS missing_year,		
7	SUM(CASE WHEN month IS NULL THEN 1 ELSE 0 END) AS missing_month,		
8	SUM(CASE WHEN day IS NULL THEN 1 ELSE 0 END) AS missing_day		
9	FROM dbo.stg_medals;		
10			
Data Output		Messages	
		Notifications	
	total bigint	missing_name bigint	missing_event bigint
1	271116	0	0
	missing_medal bigint	missing_year bigint	missing_month bigint
	231333	231333	231333
	missing_day bigint		
	231333		

El análisis de valores nulos muestra que:

- **Total, de registros:** 271.116
- **Columnas sin valores nulos:**
 - name y event.
- **Columnas con valores nulos:**
 - medal: 231,333 registros (85%).
 - year, month, day: Cada una con 231,333 valores nulos (85%).

Conclusión:

- La alta cantidad de valores faltantes en medal y las columnas de fecha (year, month, day) podría deberse a eventos sin medallas asignadas o falta de información histórica.

2. Análisis de las columnas alfanuméricas (name y event)

Query Query History

```
1 SELECT
2     COUNT(DISTINCT name) AS unique_name_count,
3     COUNT(DISTINCT event) AS unique_event_count
4 FROM dbo.stg_medals;
5
6
7
8
9
10
```

Data Output Messages Notifications

	unique_name_count bigint	unique_event_count bigint
1	134731	765

Query Query History

```
1 SELECT *
2 FROM dbo.stg_medals
3 WHERE LENGTH(name) < 3 OR LENGTH(name) > 100
4     OR LENGTH(event) < 3 OR LENGTH(event) > 100;
5
6
7
8
9
10
```

Data Output Messages Notifications

	name character varying (100)	event character varying (100)	medal character varying (20)	day integer	month integer	year integer
1	Cr	Fencing Men's epee, Masters, Individual	[null]	[null]	[null]	[null]
2	Lu	Swimming Men's 1,000 metres Freestyle	[null]	[null]	[null]	[null]
3	Lu	Swimming Men's 4,000 metres Freestyle	[null]	[null]	[null]	[null]
4	Lu	Swimming Men's 200 metres Backstroke	[null]	[null]	[null]	[null]

Longitud de los campos:

- **name:**
 - Longitud mínima: 2 caracteres.
 - Longitud máxima: 100 caracteres.
 - Promedio: 19.34 caracteres.
- **event:**
 - Longitud mínima: 15 caracteres.
 - Longitud máxima: 85 caracteres.
 - Promedio: 32.06 caracteres.

Valores únicos:

- Nombres únicos (name): 134,731.
- Eventos únicos (event): 765.

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

Se encontraron registros duplicados en la combinación de las columnas clave (name, event, year, month, day):

- Ejemplos:
 - A. Wynne Rogers: 2 duplicados.
 - Aage Albert Leidersdorff: 2 duplicados en diferentes eventos.

Conclusión:

- Los duplicados podrían deberse a errores en la carga de datos. Si no son intencionados, se deben eliminar o consolidar según el caso. Aunque parece ser que es debido a múltiples participaciones de un atleta en un evento.

4. Valores inválidos

Medallas (medal):

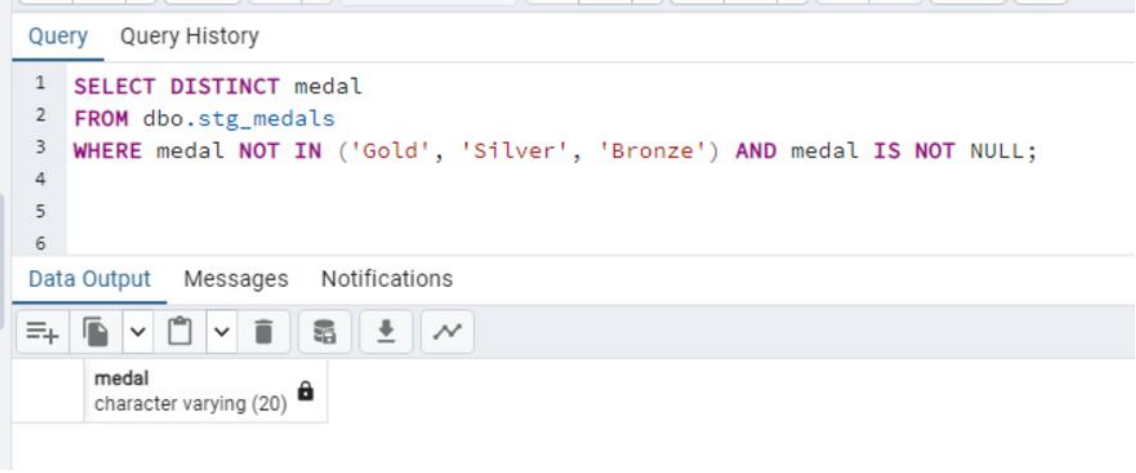
- Se verificó que los valores fueran únicamente Gold, Silver, Bronze o NULL.
- No se encontraron valores inválidos en esta columna.

Fechas (year, month, day):

- No se detectaron registros con años fuera del rango válido (1896 a la fecha actual), ni meses o días fuera de rango.

Conclusión:

- Las columnas medal y de fecha no presentan valores inválidos, pero las fechas faltantes pueden afectar el análisis.



```
1 SELECT DISTINCT medal
2 FROM dbo.stg_medals
3 WHERE medal NOT IN ('Gold', 'Silver', 'Bronze') AND medal IS NOT NULL;
4
5
6
```

The screenshot shows a SQL query editor with a query window and a data output window. The query is: `SELECT DISTINCT medal FROM dbo.stg_medals WHERE medal NOT IN ('Gold', 'Silver', 'Bronze') AND medal IS NOT NULL;`. The data output window shows the result: `medal` character varying (20).

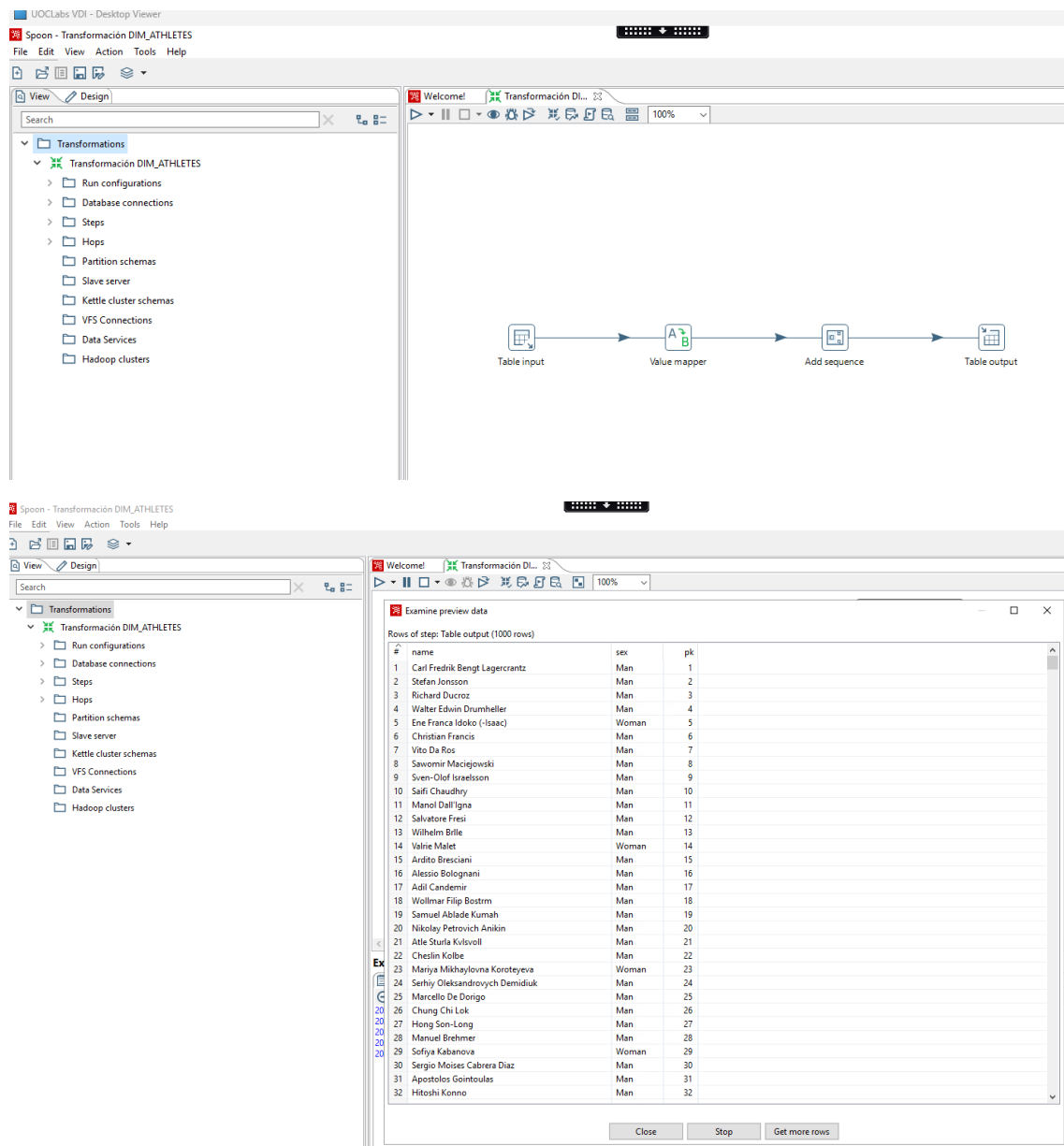
Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

2.4. Ejercicio 3. Carga de una tabla de hechos y sus dimensiones en la base de datos (20%)

2.4.1. Transformación DIM_ATHLETES



The screenshot displays the Spoon IDE interface for the transformation workflow. The left pane shows the project structure under 'Transformations' > 'Transformación DIM_ATHLETES'. The main workspace shows a workflow with four steps: 'Table input', 'Value mapper', 'Add sequence', and 'Table output'. The 'Examine preview data' window is open, showing the output of the 'Table output' step. The data is presented in a table with columns: #, name, sex, and pk. The table contains 32 rows of athlete data.

#	name	sex	pk
1	Carl Fredrik Bengt Lagercrantz	Man	1
2	Stefan Jonsson	Man	2
3	Richard Ducroz	Man	3
4	Walter Edwin Drumheller	Man	4
5	Ene Franca Idoko (-Isaac)	Woman	5
6	Christian Francis	Man	6
7	Vito Da Ros	Man	7
8	Sawomir Maciejowski	Man	8
9	Sven-Olof Irselsson	Man	9
10	Saifi Chaudhry	Man	10
11	Manol Dall'igna	Man	11
12	Salvatore Friesi	Man	12
13	Wilhelm Brille	Man	13
14	Valerie Malet	Woman	14
15	Ardito Bresciani	Man	15
16	Alessio Bolognani	Man	16
17	Adil Candemir	Man	17
18	Wollmar Filip Bostrom	Man	18
19	Samuel Ablade Kumah	Man	19
20	Nikolay Petrovich Anikin	Man	20
21	Atte Stauri Kulevov	Man	21
22	Cezlin Kolbe	Man	22
23	Mariya Mikhaylovna Koroteyeva	Woman	23
24	Serhiy Oleksandrovych Demidiuk	Man	24
25	Marcello De Dorigo	Man	25
26	Chung Chi Lok	Man	26
27	Hong Son-Long	Man	27
28	Manuel Brehmer	Man	28
29	Sofiya Kabanova	Woman	29
30	Sergio Moises Cabrera Diaz	Man	30
31	Apostolos Gointoulas	Man	31
32	Hitoshi Konno	Man	32

Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC* x

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC

Query Query History

```
1 SELECT *FROM dbo.dim_athletes
```

Data Output Messages Notifications

	pk_athlete [PK] integer	name character varying (100)	sex character varying (5)
1	1	Carl Fredrik Bengt Lagercrantz	Man
2	2	Stefan Jonsson	Man
3	3	Richard Ducroz	Man
4	4	Walter Edwin Drumheller	Man
5	5	Ene Franca Idoko (-Isaac)	Woman
6	6	Christian Francis	Man
7	7	Vito Da Ros	Man
8	8	Sawomir Maciejowski	Man
9	9	Sven-Olof Israelsson	Man
10	10	Saifi Chaudhry	Man
11	11	Manol Dall'Igna	Man
12	12	Salvatore Fresi	Man
13	13	Wilhelm Brille	Man
14	14	Valrie Malet	Woman
15	15	Ardito Bresciani	Man
16	16	Alessio Bolognani	Man
17	17	Adil Candemir	Man
18	18	Wollmar Filip Boström	Man
19	19	Samuel Ablade Kumah	Man
20	20	Nikolay Petrovich Anikin	Man
21	21	Atle Sturla Kvistvoll	Man
22	22	Cheslin Kolbe	Man
23	23	Mariya Mikhaylovna Koroteyeva	Woman
24	24	Serhiy Oleksandrovych Demidiuk	Man
25	25	Mikhail G. G. G. G. G.	Man

Total rows: 1000 of 2000 Query complete 00:00:00.216

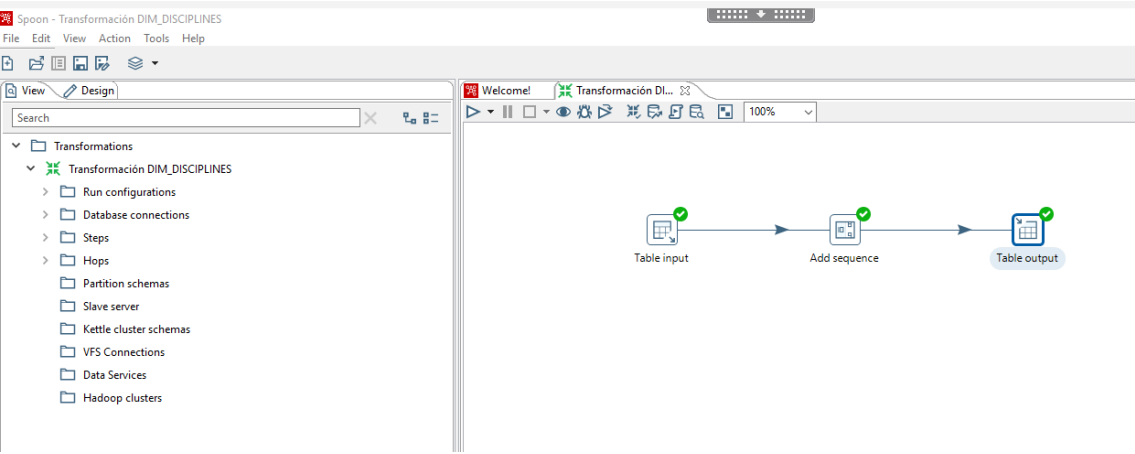
Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

2.4.2. Transformación DIM_DISCIPLINES

**Añadido columna sport debido a la necesidad del ejercicio posterior 2.4.6*



Examine preview data

Rows of step: Table output (765 rows)

#	event	sport	pk
1	Equestrianism Men's Three-Day Event, Individual	Equestrianism	1
2	Cycling Men's Team Pursuit, 1,980 yards	Cycling	2
3	Tennis Mixed Doubles, Covered Courts	Tennis	3
4	Cross Country Skiing Women's 15 km Skiathlon	Cross Country Skiing	4
5	Judo Women's Extra-Lightweight	Judo	5
6	Canoeing Men's Kayak Singles, 200 metres	Canoeing	6
7	Shooting Men's Small-Bore Rifle, Moving Target, 25 yards	Shooting	7
8	Art Competitions Mixed Music, Compositions For Solo Or Chorus	Art Competitions	8
9	Canoeing Men's Kayak Singles, 500 metres	Canoeing	9
10	Wrestling Men's Light-Flyweight, Freestyle	Wrestling	10
11	Rowing Men's Coxed Pairs (1 mile)	Rowing	11
12	Cycling Men's 5,000 metres	Cycling	12
13	Volleyball Men's Volleyball	Volleyball	13
14	Art Competitions Mixed Sculpturing, Reliefs	Art Competitions	14
15	Weightlifting Women's Lightweight	Weightlifting	15
16	Athletics Women's 3,000 metres Steeplechase	Athletics	16
17	Roque Men's Singles	Roque	17
18	Swimming Men's 400 metres Individual Medley	Swimming	18
19	Cross Country Skiing Women's 15 kilometres	Cross Country Skiing	19
20	Rowing Men's Coxed Eights	Rowing	20
21	Swimming Women's 100 metres Freestyle	Swimming	21
22	Badminton Women's Singles	Badminton	22
23	Sailing Mixed 10-20 Ton	Sailing	23
24	Figure Skating Mixed Ice Dancing	Figure Skating	24
25	Canoeing Men's Kayak Singles, 10,000 metres	Canoeing	25
26	Cross Country Skiing Men's 15 kilometres	Cross Country Skiing	26
27	Handball Men's Handball	Handball	27
28	Alpine Skiing Women's Combined	Alpine Skiing	28
29	Speed Skating Women's 5,000 metres	Speed Skating	29
30	Archery Women's Double National Round	Archery	30
31	Beach Volleyball Women's Beach Volleyball	Beach Volleyball	31
32	Short Track Speed Skating Women's 1,500 metres	Short Track Speed Skating	32

Close

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC*

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC

No limit

Query Query History

1 SELECT *FROM dbo.dim_disciplines

Data Output Messages Notifications

	pk_discipline [PK] integer	event character varying (100)	sport character varying (50)
1	1	Equestrianism Men's Three-Day Event, Individual	Equestrianism
2	2	Cycling Men's Team Pursuit, 1,980 yards	Cycling
3	3	Tennis Mixed Doubles, Covered Courts	Tennis
4	4	Cross Country Skiing Women's 15 km Skiathlon	Cross Country Skiing
5	5	Judo Women's Extra-Lightweight	Judo
6	6	Canoeing Men's Kayak Singles, 200 metres	Canoeing
7	7	Shooting Men's Small-Bore Rifle, Moving Target, 25 yards	Shooting
8	8	Art Competitions Mixed Music, Compositions For Solo Or Chorus	Art Competitions
9	9	Canoeing Men's Kayak Singles, 500 metres	Canoeing
10	10	Wrestling Men's Light-Flyweight, Freestyle	Wrestling
11	11	Rowing Men's Coxed Pairs (1 mile)	Rowing
12	12	Cycling Men's 5,000 metres	Cycling
13	13	Volleyball Men's Volleyball	Volleyball
14	14	Art Competitions Mixed Sculpturing, Reliefs	Art Competitions
15	15	Weightlifting Women's Lightweight	Weightlifting
16	16	Athletics Women's 3,000 metres Steeplechase	Athletics
17	17	Roque Men's Singles	Roque
18	18	Swimming Men's 400 metres Individual Medley	Swimming
19	19	Cross Country Skiing Women's 15 kilometres	Cross Country Skiing
20	20	Rowing Men's Coxed Eights	Rowing
21	21	Swimming Women's 100 metres Freestyle	Swimming
22	22	Badminton Women's Singles	Badminton
23	23	Sailing Mixed 10-20 Ton	Sailing
24	24	Figure Skating Mixed Ice Dancing	Figure Skating
25	25	Canoeing Men's Kayak Singles, 1,000 metres	Canoeing

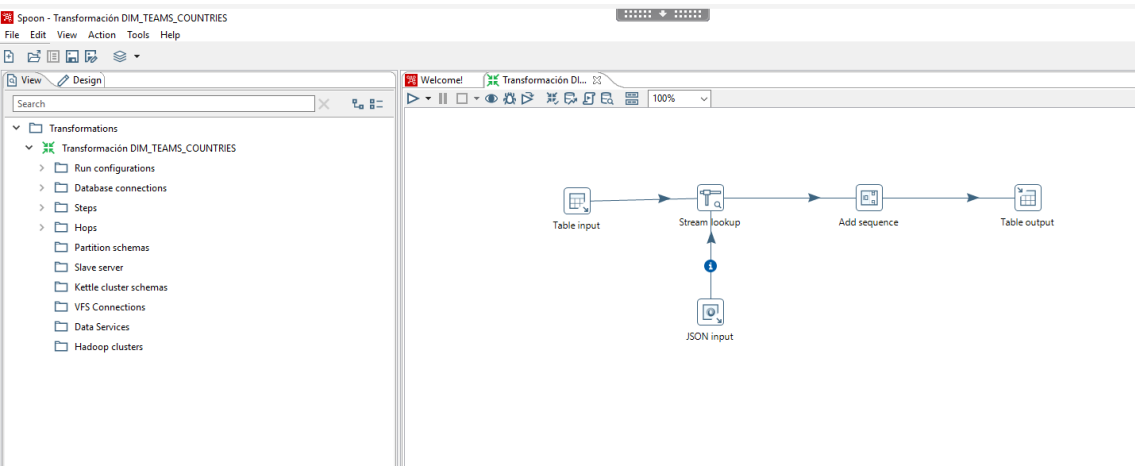
Total rows: 765 of 765 Query complete 00:00:00.105

Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

2.4.3. Transformación DIM_TEAMS_COUNTRIES



UOCLabs VDI - Desktop Viewer

Spoon - Transformación DIM_TEAMS_COUNTRIES

File Edit View Action Tools Help

View Design

Search

Transformations

- Transformación DIM_TEAMS_COUNTRIES
 - Run configurations
 - Database connections
 - Steps
 - Hops
 - Partition schemas
 - Slave server
 - Kettle cluster schemas
 - VFS Connections
 - Data Services
 - Hadoop clusters

Examine preview data

Rows of step: Table output (1000 rows)

#	team	noc	country	pk
1	New York Athletic Club-4	USA	USA	1
2	Boer Team	RSA	South Africa	2
3	Sass-2	SWE	Sweden	3
4	Cha-Cha III	YUG	Serbia	4
5	Shrew II	USA	USA	5
6	Ciocca II	ITA	Italy	6
7	Laurea-1	GBR	UK	7
8	Arcturus	ARG	Argentina	8
9	Lyn-2	NOR	Norway	9
10	EA II	FRA	France	10
11	Indonesia-2	INA	Indonesia	11
12	Chuckles	GBR	UK	12
13	Camelia	POR	Portugal	13
14	Bosnia and Herzegovina	BIH	Bosnia and Herzegovina	14
15	Roklubb Kbhavn-2	DEN	Denmark	15
16	Pimm	BRA	Brazil	16
17	Guatemala	GUA	Guatemala	17
18	Cushaven	TUR	Turkey	18
19	Complex II	USA	USA	19
20	K.S.S.S. 1912-2	SWE	Sweden	20
21	Whitini Star	FRA	France	21
22	Marie-Tim	GRE	Greece	22
23	Saint Kitts and Nevis	SKN	Saint Kitts	23
24	Rambo	JPN	Japan	24
25	Vervine-19	FRA	France	25
26	Kosovo	KOS	Kosovo	26
27	Armenia	ARM	Armenia	27
28	Venilia	ITA	Italy	28
29	Umberta V	ITA	Italy	29
30	Latvia-2	LAT	Latvia	30
31	Falcon IV	AUS	Australia	31
32	Switzerland-1	SUI	Switzerland	32

Close Stop Get more rows

Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC X

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC

Query Query History

```
1 SELECT *FROM dbo.dim_teams_countries
```

Data Output Messages Notifications

	pk_team_country [PK] integer	team character varying (50)	noc character varying (3)	country character varying (50)
1	1	New York Athletic Club-4	USA	USA
2	2	Boer Team	RSA	South Africa
3	3	Sass-2	SWE	Sweden
4	4	Cha-Cha III	YUG	Serbia
5	5	Shrew II	USA	USA
6	6	Ciocca II	ITA	Italy
7	7	Laurea-1	GBR	UK
8	8	Arcturus	ARG	Argentina
9	9	Lyn-2	NOR	Norway
10	10	EA II	FRA	France
11	11	Indonesia-2	INA	Indonesia
12	12	Chuckles	GBR	UK
13	13	Camelia	POR	Portugal
14	14	Bosnia and Herzegovina	BIH	Bosnia and Herzegovina
15	15	Roklubb Kbnhavn-2	DEN	Denmark
16	16	Pimm	BRA	Brazil
17	17	Guatemala	GUA	Guatemala
18	18	Cuxhaven	TUR	Turkey
19	19	Complex II	USA	USA
20	20	K.S.S.S. 1912-2	SWE	Sweden
21	21	Whitini Star	FRA	France
22	22	Marie-Tim	GRE	Greece
23	23	Saint Kitts and Nevis	SKN	Saint Kitts
24	24	Rambo	JPN	Japan

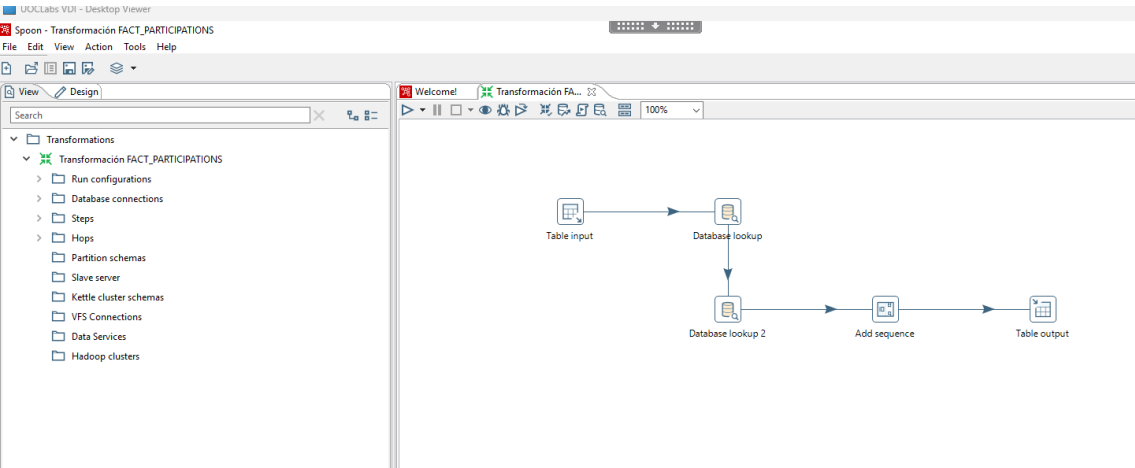
Total rows: 1000 of 1231 Query complete 00:00:00.140

Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

2.4.4. Transformación FACT_PARTICIPATIONS



#	name	age	height	weight	event	noc	year	season	city	pk_athlete	medal	fx_discipline	fx_team_country	pk
1	Alaaeldin Ahmad El Sayid Abouelkassem	21	188	87	Fencing Men's Foil, Individual	EGY	2012	Summer	London	1806	Silver	238	350	1
2	Emelita "Shella" Agteribbe (Bito)	19	<null>	<null>	Gymnastics Women's Team All-Around	NED	1928	Summer	Amsterdam	1175	Gold	724	70	2
3	Sandra Aguilar Navarro	23	167	50	Rhythmic Gymnastics Women's Group	ESP	2016	Summer	Rio de Janeiro	588	Silver	246	163	3
4	Franz Aigner	32	<null>	107	Weightlifting Men's Heavyweight	AUT	1924	Summer	Paris	659	Silver	717	235	4
5	Yutaka Aihara	21	156	55	Gymnastics Men's Team All-Around	JPN	1992	Summer	Barcelona	1770	Bronze	434	24	5
6	Benjamin Alexandro "Ben" Agoston	24	178	75	Figure Skating - Mixed Ice Dancing	USA	2006	Winter	Torino	1799	Silver	24	1	6
7	Thomas Valtter Ilin	24	182	90	Ice Hockey Men's Ice Hockey	SWE	1984	Winter	Sarajevo	365	Bronze	601	3	7
8	Renzo Alver	23	178	83	Bobsleigh Men's Two	ITA	1956	Winter	Cortina d'Ampezzo	988	Silver	737	6	8
9	Renzo Alver	23	178	83	Bobsleigh Men's Four	ITA	1956	Winter	Cortina d'Ampezzo	988	Silver	58	6	9
10	Reidar Andersen	24	180	<null>	Ski Jumping Men's Normal Hill, Individual	NOR	1936	Winter	Garmisch-Partenkirchen	671	Bronze	263	9	10
11	Haimrich "Yaeli" Angert	40	<null>	<null>	Bobsleigh Men's Four	SUI	1956	Winter	Cortina d'Ampezzo	1938	Gold	58	32	11
12	Nikolay Petrovich Ankin	24	<null>	<null>	Cross Country Skiing Men's 4 x 10 kilometres Relay	URS	1956	Winter	Cortina d'Ampezzo	20	Gold	101	92	12
13	Nikolay Petrovich Ankin	28	<null>	<null>	Cross Country Skiing Men's 30 kilometres	URS	1960	Winter	Squaw Valley	20	Bronze	685	92	13
14	Nikolay Petrovich Ankin	28	<null>	<null>	Cross Country Skiing Men's 4 x 10 kilometres Relay	URS	1960	Winter	Squaw Valley	20	Bronze	101	92	14
15	Katrin Apel	24	172	62	Biathlon Women's 7.5 kilometres Sprint	GER	1998	Winter	Nagano	283	Bronze	323	50	15
16	Katrin Apel	24	172	62	Biathlon Women's 4 x 7.5 kilometres Relay	GER	1998	Winter	Nagano	283	Gold	493	50	16
17	Katrin Apel	28	172	62	Biathlon Women's 4 x 7.5 kilometres Relay	GER	2002	Winter	Salt Lake City	283	Gold	493	50	17
18	Katrin Apel	32	172	62	Biathlon Women's 4 x 6 kilometres Relay	GER	2006	Winter	Torino	283	Silver	741	50	18
19	Luk Bauer	28	181	73	Cross Country Skiing Men's 15 kilometres	CZE	2006	Winter	Torino	1586	Silver	26	111	19
20	Luk Bauer	32	181	73	Cross Country Skiing Men's 15 kilometres	CZE	2010	Winter	Vancouver	1586	Bronze	26	111	20
21	Luk Bauer	32	181	73	Cross Country Skiing Men's 4 x 10 kilometres Relay	CZE	2010	Winter	Vancouver	1586	Bronze	101	111	21
22	Jaroslav Benk	21	181	82	Ice Hockey Men's Ice Hockey	TCH	1984	Winter	Sarajevo	922	Silver	601	135	22
23	Pr Jonas Bergqvist	25	183	83	Ice Hockey Men's Ice Hockey	SWE	1988	Winter	Calgary	267	Bronze	601	3	23
24	Pr Jonas Bergqvist	31	183	83	Ice Hockey Men's Ice Hockey	SWE	1994	Winter	Lillehammer	267	Gold	601	3	24
25	nderson Lus de Abreu Oliveira	20	176	77	Football Men's Football	BRA	2008	Summer	Beijing	1487	Bronze	316	16	25
26	Andreas Arlene Anderson (- Bolder)	22	164	67	Athletics Women's 4 x 400 metres Relay	USA	2000	Summer	Sydney	1189	Gold	152	1	26
27	Craig Anthony Anderson	23	187	86	Baseball Men's Baseball	AUS	2004	Summer	Athina	1582	Silver	712	31	27
28	Sune Isidor Andersson	27	<null>	<null>	Football Men's Football	SWE	1948	Summer	London	770	Gold	316	3	28
29	Thure Eloff Andersson	29	<null>	<null>	Wrestling Men's Welterweight, Freestyle	SWE	1936	Summer	Berlin	1466	Silver	330	3	29
30	Akadi Andriyashyn	25	174	70	Football Men's Football	URS	1972	Summer	Munich	1578	Bronze	316	92	30
31	Hannelore Anke (- Hofmann)	18	160	52	Swimming Women's 100 metres Breaststroke	GDR	1976	Summer	Montreal	95	Gold	619	295	31
32	Hannelore Anke (- Hofmann)	18	160	52	Swimming Women's 4 x 100 metres Medley Relay	GDR	1976	Summer	Montreal	95	Gold	619	295	32

Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC* x

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC

No limit

Query Query History

```
1 SELECT
2 count(1) as total_participations
3 FROM
4 dbo.fact_participations p, dbo.dim_athletes a
5 WHERE
6 p.fk_athlete = a.pk_athlete
7
8
```

Data Output Messages Notifications

	total_participations	
1	bigint	271414

2.4.5. Job JOB_PARTICIPATIONS

JOB_PR3 Transformación DIM_ATHLETES

Start Truncate FACTs DIM_ATHLETES DIM_DISCIPLINES DIM_TEAMS_COUNTRIES FACT_PARTICIPATIONS Success

Execution Results

Logging History Job metrics Metrics

2024/11/24 18:46:39 - Database lookup 2.0 - linenr 100000
2024/11/24 18:46:39 - Add sequence.0 - linenr 100000
2024/11/24 18:46:40 - Table output.0 - linenr 100000
2024/11/24 18:46:40 - Table input.0 - [TableInput.Log.LineNumber!]
2024/11/24 18:46:51 - Database lookup 2.0 - linenr 150000
2024/11/24 18:46:51 - Add sequence.0 - linenr 150000
2024/11/24 18:46:51 - Table output.0 - linenr 150000
2024/11/24 18:47:01 - Table input.0 - [TableInput.Log.LineNumber!]
2024/11/24 18:47:03 - Database lookup 2.0 - linenr 200000
2024/11/24 18:47:03 - Add sequence.0 - linenr 200000
2024/11/24 18:47:03 - Table output.0 - linenr 200000
2024/11/24 18:47:14 - Table input.0 - [TableInput.Log.LineNumber!]
2024/11/24 18:47:16 - Database lookup 2.0 - linenr 250000
2024/11/24 18:47:16 - Add sequence.0 - linenr 250000
2024/11/24 18:47:16 - Table output.0 - linenr 250000
2024/11/24 18:47:19 - Table input.0 - [TableInput.Log.FinishedReadingQuery!]
2024/11/24 18:47:19 - Table input.0 - Finished processing (I=271414, O=0, R=0, W=271414, U=0, E=0)
2024/11/24 18:47:22 - Database lookup 2.0 - Finished processing (I=271414, O=0, R=271414, W=271414, U=0, E=0)
2024/11/24 18:47:22 - Add sequence.0 - Finished processing (I=0, O=0, R=271414, W=271414, U=0, E=0)
2024/11/24 18:47:22 - Table output.0 - Finished processing (I=0, O=271414, R=271414, W=271414, U=0, E=0)
2024/11/24 18:47:22 - JOB_PR3 - Starting entry (Success)
2024/11/24 18:47:22 - JOB_PR3 - Finished job entry (Success) (result=[true])
2024/11/24 18:47:22 - JOB_PR3 - Finished job entry (FACT_PARTICIPATIONS) (result=[true])
2024/11/24 18:47:22 - JOB_PR3 - Finished job entry (DIM_TEAMS_COUNTRIES) (result=[true])
2024/11/24 18:47:22 - JOB_PR3 - Finished job entry (DIM_DISCIPLINES) (result=[true])
2024/11/24 18:47:22 - JOB_PR3 - Finished job entry (DIM_ATHLETES) (result=[true])
2024/11/24 18:47:22 - JOB_PR3 - Finished job entry (Truncate FACTs) (result=[true])
2024/11/24 18:47:22 - JOB_PR3 - Job execution finished
2024/11/24 18:47:22 - Spoon - Job has ended.

2.4.6. Ejercicios a realizar

- Ampliar los datos de la tabla ***DIM_DISCIPLINES***

The screenshot displays the SQL Server Data Tools (SSDT) interface. At the top, a status bar shows 'Welcome!' and 'Transformación DI...'. Below it, a toolbar contains various icons for running and debugging tasks. A small window in the top-left corner displays execution statistics:

Errors	
Active	Finished
Time	0.2s
Speed (r/s)	4,274
Input/output	-

The main workspace shows a data flow diagram with three steps: 'Table input', 'Add sequence', and 'Table output'. The 'Table input' step is selected, and its configuration window is open. The configuration window has the following fields:

- Step name: `Table input`
- Connection: `postgres`
- SQL: `SELECT DISTINCT event, sport
FROM dbo.stg_athletes`

Below the configuration window, the 'Execution Results' pane is visible, showing a 'Logging' tab with a list of timestamps: 2024/12/07 19:02:55, 2024/12/07 19:02:55, 2024/12/07 19:02:55, 2024/12/07 19:02:55, and 2024/12/07 19:02:55.

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC* x

SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC

No limit

Query Query History

```
1 SELECT sport, COUNT(event) AS total_events
2 FROM dbo.dim_disciplines
3 GROUP BY sport
4 ORDER BY total_events DESC;
5
```

Data Output Messages Notifications

	sport character varying (50)	totalEvents bigint
1	Shooting	83
2	Athletics	83
3	Swimming	55
4	Cycling	44
5	Sailing	38
6	Wrestling	30
7	Archery	29
8	Art Competitions	29
9	Canoeing	27
10	Gymnastics	27
11	Rowing	25
12	Cross Country Skiing	23
13	Weightlifting	21
14	Fencing	18
15	Equestrianism	18
16	Boxing	15
17	Judo	15
18	Speed Skating	13
19	Biathlon	13
20	Snowboarding	12
21	Diving	10

Total rows: 66 of 66 Query complete 00:00:00.103 Ln 5, Col 1

- Incorporación de **MEDALLAS**

Modifico código y transformación:

SELECT

a.name,

a.age,

a.height,

a.weight,

a.event,

a.noc,

a.year,

a.season,

a.city,

b.pk_athlete,

m.medal

FROM dbo.stg_athletes a

JOIN dbo.dim_athletes b

ON a.name = b.name

LEFT JOIN dbo.stg_medals m

ON a.name = m.name

AND a.year = m.year

AND a.event = m.event;

Table output

Step name: Table output

Connection: postgres

Target schema: dbo

Target table: fact_participations

Commit size: 1000

Truncate table: ☒

Ignore insert errors: ☐

Specify database fields: ☒

Main options | Database fields

Fields to insert:

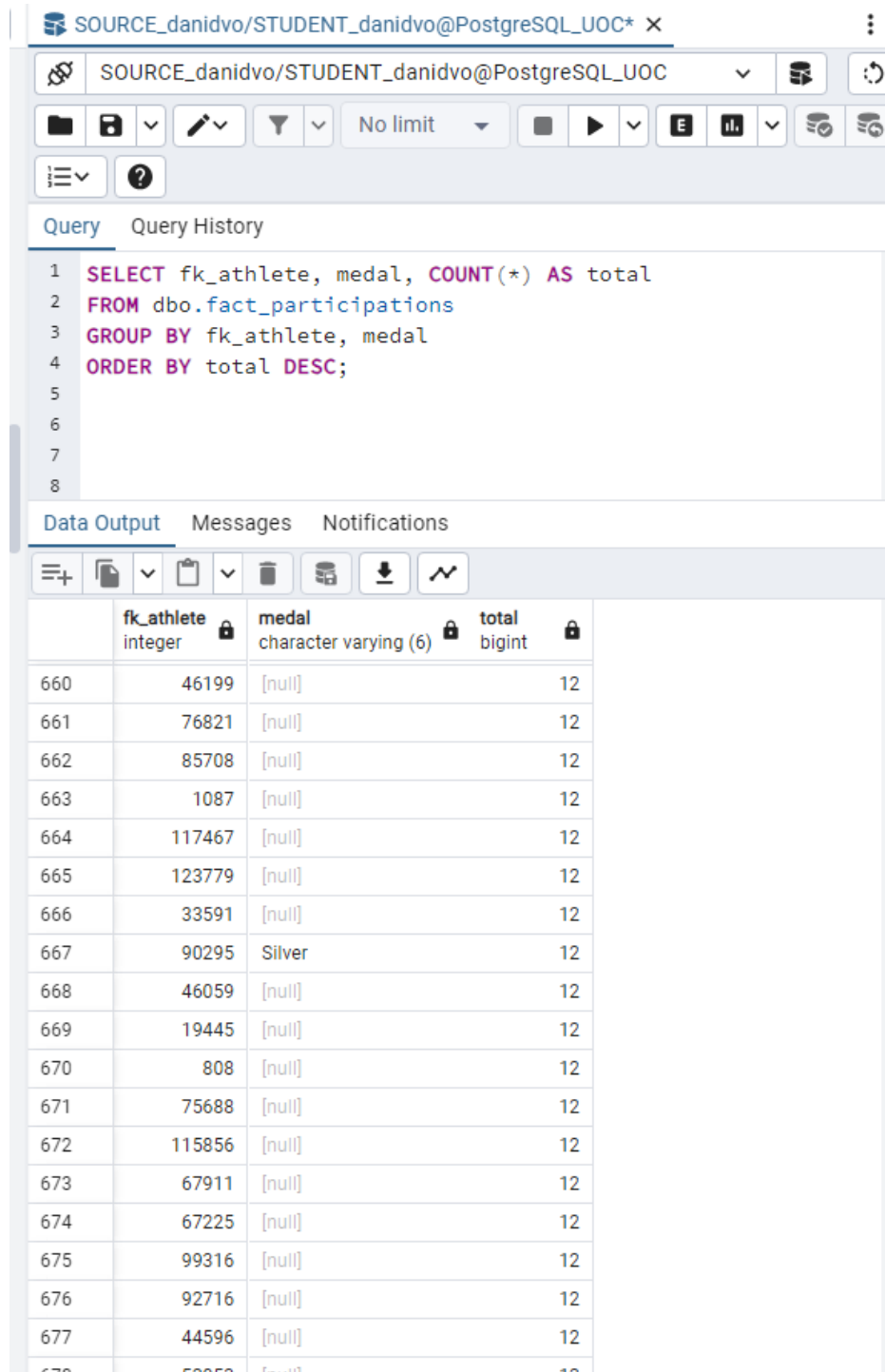
#	Table field	Stream field
1	age	age
2	height	height
3	weight	weight
4	year	year
5	season	season
6	city	city
7	fk_discipline	fk_discipline
8	fk_team_co...	fk_team_cou...
9	pk	pk
10	fk_athlete	pk_athlete
11	medal	medal

Get fields

Enter field mapping

Help OK Cancel SQL

VERIFICO:



The screenshot shows a PostgreSQL query editor interface. The title bar indicates the connection is 'SOURCE_danidvo/STUDENT_danidvo@PostgreSQL_UOC'. The query editor contains the following SQL query:

```
1 SELECT fk_athlete, medal, COUNT(*) AS total
2 FROM dbo.fact_participations
3 GROUP BY fk_athlete, medal
4 ORDER BY total DESC;
```

The 'Data Output' tab is active, displaying the results of the query. The table has three columns: 'fk_athlete' (integer), 'medal' (character varying (6)), and 'total' (bigint). The results are ordered by 'total' in descending order.

	fk_athlete integer	medal character varying (6)	total bigint
660	46199	[null]	12
661	76821	[null]	12
662	85708	[null]	12
663	1087	[null]	12
664	117467	[null]	12
665	123779	[null]	12
666	33591	[null]	12
667	90295	Silver	12
668	46059	[null]	12
669	19445	[null]	12
670	808	[null]	12
671	75688	[null]	12
672	115856	[null]	12
673	67911	[null]	12
674	67225	[null]	12
675	99316	[null]	12
676	92716	[null]	12
677	44596	[null]	12
678	52052	[null]	12

The screenshot shows a PostgreSQL query editor interface. The query is as follows:

```
1 SELECT medal, COUNT(*) AS total
2 FROM dbo.fact_participations
3 WHERE medal IS NOT NULL
4 GROUP BY medal
5 ORDER BY total DESC;
```

The results are displayed in a table with two columns: medal and total.

	medal	total
1	Gold	13440
2	Bronze	13373
3	Silver	13226

- Trasladar los datos del Data Warehouse al Data Mart.

Tabla creada:

The screenshot shows a SQL query in a query editor. The query is as follows:

```
SELECT *
FROM INFORMATION_SCHEMA.TABLES
WHERE TABLE_NAME IN ('DIM_athletes', 'DIM_disciplines', 'DIM_teams_countries', 'FACT_participations');
```

The results are displayed in a table with five columns: TABLE_CATALOG, TABLE_SCHEMA, TABLE_NAME, and TABLE_TYPE.

	TABLE_CATALOG	TABLE_SCHEMA	TABLE_NAME	TABLE_TYPE
1	SOURCE_danidvo	dbo	DIM_athletes	BASE TABLE
2	SOURCE_danidvo	dbo	DIM_disciplines	BASE TABLE
3	SOURCE_danidvo	dbo	DIM_teams_countries	BASE TABLE
4	SOURCE_danidvo	dbo	FACT_participations	BASE TABLE

Tablas copiadas->

DIM_ATHLETES

SQLQuery1.sql - U...DENT_danidvo (59))* X

```
SELECT *FROM dbo.DIM_athletes
```

91 %

Results Messages

	pk_athlete	name	sex
1	1	Eric Greenwood Avila	Man
2	2	Cai Zelin	Man
3	3	Kurt Homfischer	Man
4	4	Lineo Mochesane	Woman
5	5	Karl Sture Lennart Hansson	Man
6	6	Tak Chuen "Terence" Lam	Man
7	7	James McKenzie	Man
8	8	John Joseph "Johnny" Kelley	Man
9	9	Alberto A. Llorens	Man
10	10	Sugiono Katijo	Man
11	11	Genevieve LaCaze	Woman
12	12	Emerich Alexandru Jenei	Man
13	13	John Mugabi	Man
14	14	Joseph "Joe" Laporte	Man
15	15	Chisato Nagaoka	Woman
16	16	Jrg Helmdach	Man
17	17	Marco Cimatti	Man
18	18	Carl Fredrik Bengt Lagercrantz	Man
19	19	Stefan Jonsson	Man
20	20	Richard Ducroz	Man
21	21	Walter Edwin Drumheller	Man
22	22	Ene Franca Idoko (-Isaac)	Woman
23	23	Christian Francis	Man
24	24	Vito Da Ros	Man
25	25	Sawomir Maciejowski	Man
26	26	Sven-Olof Israelsson	Man
27	27	Saifi Chaudhry	Man
28	28	Manol Dall'Igna	Man
29	29	Salvatore Fresi	Man
30	30	PI Victor hling Norberg	Man
31	31	Mika Sakari Nieminen	Man
32	32	Wilhelm Brlle	Man
33	33	Sean Nicholas Olsson	Man

IOCSQL03 (14.0 RTM) | STUDENT_danidvo (59) | SOURCE_danidvo | 00:00:01 | 134786 rows

SQLQuery1.sql - U...DENT_danidvo (59))* ✕

```
SELECT COUNT(*) AS total_records
FROM dbo.DIM_athletes;
```

91 %

Results Messages

	total_records
1	134786

DIM_DISCIPLINES

SQLQuery1.sql - U...DENT_danidvo (59))* ✕

```
SELECT *FROM dbo.DIM_disciplines
```

91 %

Results Messages

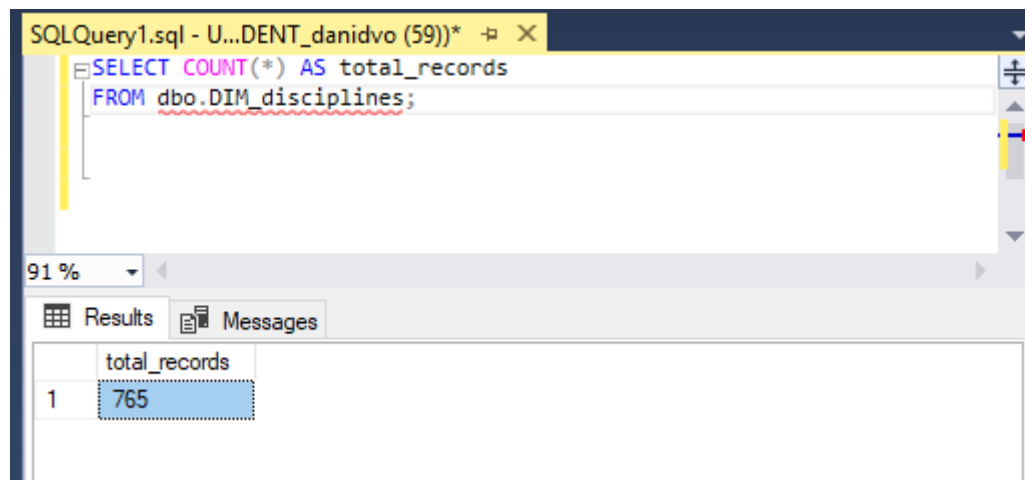
	pk_discipline	event	sport
1	1	Equestrianism Men's Three-Day Event, Individual	Equestrianism
2	2	Cycling Men's Team Pursuit, 1,980 yards	Cycling
3	3	Tennis Mixed Doubles, Covered Courts	Tennis
4	4	Cross Country Skiing Women's 15 km Skiathlon	Cross Country Skiing
5	5	Judo Women's Extra-Lightweight	Judo
6	6	Canoeing Men's Kayak Singles, 200 metres	Canoeing
7	7	Shooting Men's Small-Bore Rifle, Moving Target, ...	Shooting
8	8	Art Competitions Mixed Music, Compositions For S...	Art Competitions
9	9	Canoeing Men's Kayak Singles, 500 metres	Canoeing
10	10	Wrestling Men's Light-Flyweight, Freestyle	Wrestling
11	11	Rowing Men's Coxed Pairs (1 mile)	Rowing
12	12	Cycling Men's 5,000 metres	Cycling
13	13	Volleyball Men's Volleyball	Volleyball
14	14	Art Competitions Mixed Sculpturing, Reliefs	Art Competitions
15	15	Weightlifting Women's Lightweight	Weightlifting
16	16	Athletics Women's 3,000 metres Steeplechase	Athletics
17	17	Roque Men's Singles	Roque
18	18	Swimming Men's 400 metres Individual Medley	Swimming
19	19	Cross Country Skiing Women's 15 kilometres	Cross Country Skiing
20	20	Rowing Men's Coxed Eights	Rowing
21	21	Swimming Women's 100 metres Freestyle	Swimming
22	22	Badminton Women's Singles	Badminton
23	23	Sailing Mixed 10-20 Ton	Sailing
24	24	Figure Skating Mixed Ice Dancing	Figure Skating
25	25	Canoeing Men's Kayak Singles, 10,000 metres	Canoeing
26	26	Cross Country Skiing Men's 15 kilometres	Cross Country Skiing
27	27	Handball Men's Handball	Handball
28	28	Alpine Skiing Women's Combined	Alpine Skiing
29	29	Speed Skating Women's 5,000 metres	Speed Skating
30	30	Archery Women's Double National Round	Archery
31	31	Beach Volleyball Women's Beach Volleyball	Beach Volleyball
32	32	Short Track Speed Skating Women's 1,500 metres	Short Track Spee...
33	33	Speed Skating Women's 3,000 metres	Speed Skating
34	34	Weightlifting Men's Middleweight	Weightlifting
35	35	Wrestling Women's Light-Heavyweight, Freestyle	Wrestling
36	36	Weightlifting Men's Heavyweight	Weightlifting

R1UOCSQL03 (14.0 RTM) | STUDENT_danidvo (59) | SOURCE_danidvo | 00:00:00 | 765 rows

Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos



The screenshot shows a SQL Server Enterprise Manager window with a query executed. The query is: `SELECT COUNT(*) AS total_records FROM dbo.DIM_disciplines;`. The results pane shows a single row with the value 765 for the column total_records.

	total_records
1	765

DM DIM TEAMS COUNTRIES

SQLQuery1.sql - U...DENT_danidvo (59))* - X

SELECT *FROM dbo.DIM_teams_countries

91 %

Results Messages

	pk_team_country	team	noc	country
1	1	New York Athletic Club-4	USA	USA
2	2	Boer Team	RSA	South Africa
3	3	Sass-2	SWE	Sweden
4	4	Cha-Cha III	YUG	Serbia
5	5	Shrew II	USA	USA
6	6	Ciocca II	ITA	Italy
7	7	Laurea-1	GBR	UK
8	8	Arcturus	ARG	Argentina
9	9	Lyn-2	NOR	Norway
10	10	EA II	FRA	France
11	11	Indonesia-2	INA	Indonesia
12	12	Chuckles	GBR	UK
13	13	Camelia	POR	Portugal
14	14	Bosnia and Herzegovina	BIH	Bosnia and Herzegovina
15	15	Roklubb Kbnhavn-2	DEN	Denmark
16	16	Pimm	BRA	Brazil
17	17	Guatemala	GUA	Guatemala
18	18	Cuxhaven	TUR	Turkey
19	19	Complex II	USA	USA
20	20	K.S.S.S. 1912-2	SWE	Sweden
21	21	Whitini Star	FRA	France
22	22	Marie-Tim	GRE	Greece
23	23	Saint Kitts and Nevis	SKN	Saint Kitts
24	24	Rambo	JPN	Japan
25	25	Verveine-19	FRA	France
26	26	Kosovo	KOS	Kosovo
27	27	Amenia	ARM	Amenia
28	28	Venilia	ITA	Italy
29	29	Umberta V	ITA	Italy
30	30	Latvia-2	LAT	Latvia
31	31	Falcon IV	AUS	Australia
32	32	Switzerland-1	SUI	Switzerland
33	33	Bermudes	BER	Bermuda
34	34	Samoa	SAM	Samoa
35	35	Partenope	INA	Indonesia
36	36	Monaco	MON	Monaco
37	37	A...	GBR	UK

SQLUCSQL03 (14.0 RTM) | STUDENT_danidvo (59) | SOURCE_danidvo | 00:00:00 | 1231 rows

SQLQuery1.sql - U...DENT_danidvo (59))* - X

SELECT COUNT(*) AS total_records
FROM dbo.DIM_teams_countries;

91 %

Results Messages

	total_records
1	1231

DIM FACT PARTICIPATIONS

Object Explorer

Connect

UCS1R1UOCSQL03 (SQL Server 14.0.2060.1 - STUDENT)

Databases

System Databases

Database Snapshots

SOURCE_danidvo

Database Diagrams

Tables

System Tables

FileTables

External Tables

Graph Tables

dbo.DIM_athletes

dbo.DIM_disciplines

dbo.DIM_teams_countries

dbo.event

dbo.FACT_participations

dbo.sport

Views

External Resources

Synonyms

Programmability

Service Broker

Storage

Security

Security

Server Objects

Replication

PolyBase

Always On High Availability

Management

Integration Services Catalogs

XEvent Profiler

SQLQuery1.sql - U...DENT_danidvo (59))

SELECT * FROM dbo.FACT_participations;

91 %

Results Messages

pk	year	season	city	fk_athlete	fk_discipline	fk_team_country	age	height	weight	medal
1	1936	Summer	Berlin	114079	86	549	22	NULL	67	NULL
2	1980	Summer	Moskva	98503	139	885	28	178	72	NULL
3	2016	Summer	Rio de Janeiro	124429	589	410	27	165	58	NULL
4	1980	Winter	Lake Placid	95794	275	9	20	183	80	Bronze
5	2004	Summer	Athina	101384	352	7	24	165	55	NULL
6	2004	Summer	Athina	101384	709	7	24	165	55	NULL
7	1964	Summer	Tokyo	107045	513	591	25	181	89	NULL
8	1960	Summer	Roma	109049	445	24	25	174	68	NULL
9	1960	Summer	Roma	109049	665	24	25	174	68	NULL
10	1968	Winter	Grenoble	107312	101	18	NULL	172	72	NULL
11	1956	Summer	Melbourne	118109	552	92	28	NULL	NULL	Silver
12	1960	Summer	Roma	81108	71	18	NULL	179	71	NULL
13	1984	Summer	Los Angeles	105141	599	18	19	170	62	NULL
14	2000	Summer	Sydney	107056	330	379	26	180	69	NULL
15	1936	Summer	Berlin	115833	100	235	45	NULL	NULL	NULL
16	1996	Summer	Atlanta	86254	80	30	16	176	69	NULL
17	2000	Summer	Sydney	86254	80	30	20	176	69	NULL
18	2008	Summer	Beijing	116746	156	30	20	186	77	NULL
19	1996	Summer	Atlanta	89900	330	31	29	182	76	NULL
20	1924	Winter	Paris	86418	34	30	27	NULL	75	NULL
21	2014	Winter	Sochi	90544	601	30	41	190	95	NULL
22	1992	Summer	Barcelona	132198	139	30	25	178	74	Bronze
23	1996	Summer	Atlanta	132198	139	30	29	178	74	NULL
24	2000	Summer	Sydney	132198	139	30	34	178	74	NULL
25	2000	Summer	Sydney	132198	750	30	34	178	74	NULL
26	1912	Summer	Stockholm	116673	324	85	23	NULL	NULL	NULL
27	1948	Summer	London	90528	529	18	NULL	NULL	NULL	NULL
28	1984	Summer	Los Angeles	114399	507	18	20	196	80	NULL
29	1924	Summer	Paris	114089	316	713	28	NULL	NULL	NULL
30	1956	Summer	Stockholm	94439	730	52	31	185	75	NULL
31	1908	Summer	London	97954	434	52	24	NULL	NULL	Bronze
32	1952	Summer	Helsinki	129837	516	52	19	188	90	NULL
33	2016	Summer	Rio de Janeiro	129056	316	76	28	168	68	NULL
34	2010	Winter	Vancouver	92291	359	1	29	170	77	Bronze
35	1976	Summer	Montreal	97881	84	1	19	180	64	Gold

Query executed successfully... UCS1R1UOCSQL03 (14.0 RTM) STUDENT_danidvo (59) SOURCE_danidvo 00:00:03 271501 rows

SQLQuery1.sql - U...DENT_danidvo (59))

SELECT COUNT(*) AS total_records
FROM dbo.FACT_participations;

91 %

Results Messages

total_records
271501

Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

JOB:

**EXECUTE SQL SCRIPT en TRUNCATE, dado que no permite truncar*

The image displays two screenshots from SQL Server Enterprise Manager. The top screenshot shows the 'Execute SQL script' dialog box. The 'Step name' is 'Execute SQL script' and the 'Connection' is 'SQL'. The SQL script to be executed is:

```
DELETE FROM dbo.FACT_participations;
DELETE FROM dbo.DIM_teams_countries;
DELETE FROM dbo.DIM_disciplines;
DELETE FROM dbo.DIM_athletes;
```

The bottom screenshot shows the 'Execution Results' window. The job 'JOB_DM_PARTICIPATIONS' is shown as successful. The execution log indicates that the job was started at 2024/11/24 21:19:22 and finished at 2024/11/24 21:19:51. The log also shows the execution of the SQL script and the truncation of the tables.

Execution Results

Logging | History | Job metrics | Metrics

2024/11/24 21:19:22 - COPY_DM_DIM_ATHLETES - Dispatching started for transformation [COPY_DM_DIM_ATHLETES]
2024/11/24 21:19:22 - Table output.0 - Connected to database [SQL] (commit=1000)
2024/11/24 21:19:22 - Table input.0 - [TableInput.Log.FinishedReadingQuery]
2024/11/24 21:19:22 - Table input.0 - Finished processing (I=1231, O=0, R=0, W=1231, U=0, E=0)
2024/11/24 21:19:22 - Table output.0 - Finished processing (I=0, O=1231, R=1231, W=1231, U=0, E=0)
2024/11/24 21:19:22 - JOB_DM_PARTICIPATIONS - Starting entry [COPY_DM_FACT_PARTICIPATIONS]
2024/11/24 21:19:22 - COPY_DM_FACT_PARTICIPATIONS - Using run configuration [Pentaho local]
2024/11/24 21:19:22 - COPY_DM_FACT_PARTICIPATIONS - Running transformation using the Kettle execution engine
2024/11/24 21:19:22 - COPY_DM_FACT_PARTICIPATIONS - Dispatching started for transformation [COPY_DM_FACT_PARTICIPATIONS]
2024/11/24 21:19:22 - Table output.0 - Connected to database [SQL] (commit=1000)
2024/11/24 21:19:26 - Table input.0 - [TableInput.Log.LineNumber!]
2024/11/24 21:19:27 - Table output.0 - linenr 50000
2024/11/24 21:19:32 - Table input.0 - [TableInput.Log.LineNumber!]
2024/11/24 21:19:33 - Table output.0 - linenr 100000
2024/11/24 21:19:37 - Table input.0 - [TableInput.Log.LineNumber!]
2024/11/24 21:19:38 - Table output.0 - linenr 150000
2024/11/24 21:19:42 - Table input.0 - [TableInput.Log.LineNumber!]
2024/11/24 21:19:43 - Table output.0 - linenr 200000
2024/11/24 21:19:48 - Table input.0 - [TableInput.Log.LineNumber!]
2024/11/24 21:19:49 - Table output.0 - linenr 250000
2024/11/24 21:19:50 - Table input.0 - [TableInput.Log.FinishedReadingQuery!]
2024/11/24 21:19:50 - Table input.0 - Finished processing (I=271501, O=0, R=0, W=271501, U=0, E=0)
2024/11/24 21:19:51 - Table output.0 - Finished processing (I=0, O=271501, R=271501, W=271501, U=0, E=0)
2024/11/24 21:19:51 - JOB_DM_PARTICIPATIONS - Starting entry [Success]
2024/11/24 21:19:51 - JOB_DM_PARTICIPATIONS - Finished job entry [Success] (result=[true])
2024/11/24 21:19:51 - JOB_DM_PARTICIPATIONS - Finished job entry [COPY_DM_FACT_PARTICIPATIONS] (result=[true])
2024/11/24 21:19:51 - JOB_DM_PARTICIPATIONS - Finished job entry [COPY_DM_DIM_TEAMS_COUNTRIES] (result=[true])
2024/11/24 21:19:51 - JOB_DM_PARTICIPATIONS - Finished job entry [COPY_DM_DIM_DISCIPLINES] (result=[true])
2024/11/24 21:19:51 - JOB_DM_PARTICIPATIONS - Finished job entry [COPY_DM_DIM_ATHLETES] (result=[true])
2024/11/24 21:19:51 - JOB_DM_PARTICIPATIONS - Finished job entry [TRUNCATE_SQL_TABLES] (result=[true])
2024/11/24 21:19:51 - JOB_DM_PARTICIPATIONS - Job execution finished
2024/11/24 21:19:51 - Spoon - Job has ended.

Después del job

SQLQuery1.sql - U...DENT_danidvo (59))*

```
SELECT COUNT(*) AS total_records FROM dbo.FACT_participations;
SELECT COUNT(*) AS total_records FROM dbo.DIM_teams_countries;
SELECT COUNT(*) AS total_records FROM dbo.DIM_disciplines;
SELECT COUNT(*) AS total_records FROM dbo.DIM_athletes;
```

91 %

Results Messages

	total_records
1	271501
1	1231
1	765
1	134786

Query executed successfully. UCS1R1UOCSQL03 (14.0 RTM) STUDENT_danidvo (59) SOURCE_danidvo 00:00:00 4 rows

Ln 5 Col 1 Ch 1 INS

dmin 4 SQLQuery1.sql - UCS... Spoon - JOB_DM_PAR... 21:21

2.5. Ejercicio 4. Uso de consultas *Common Table Expressions (CTE)* para analizar los datos (15%)

Ejercicios a realizar:

Se solicitan las sentencias SQL que den respuesta a las siguientes cuestiones:

- 1) Conocer el número de registros almacenados en la tabla fact_participations agrupando por los campos year, country y sport, y ordenados por los mismos campos de forma ascendente.

The screenshot shows a SQL query window with the following CTE query:

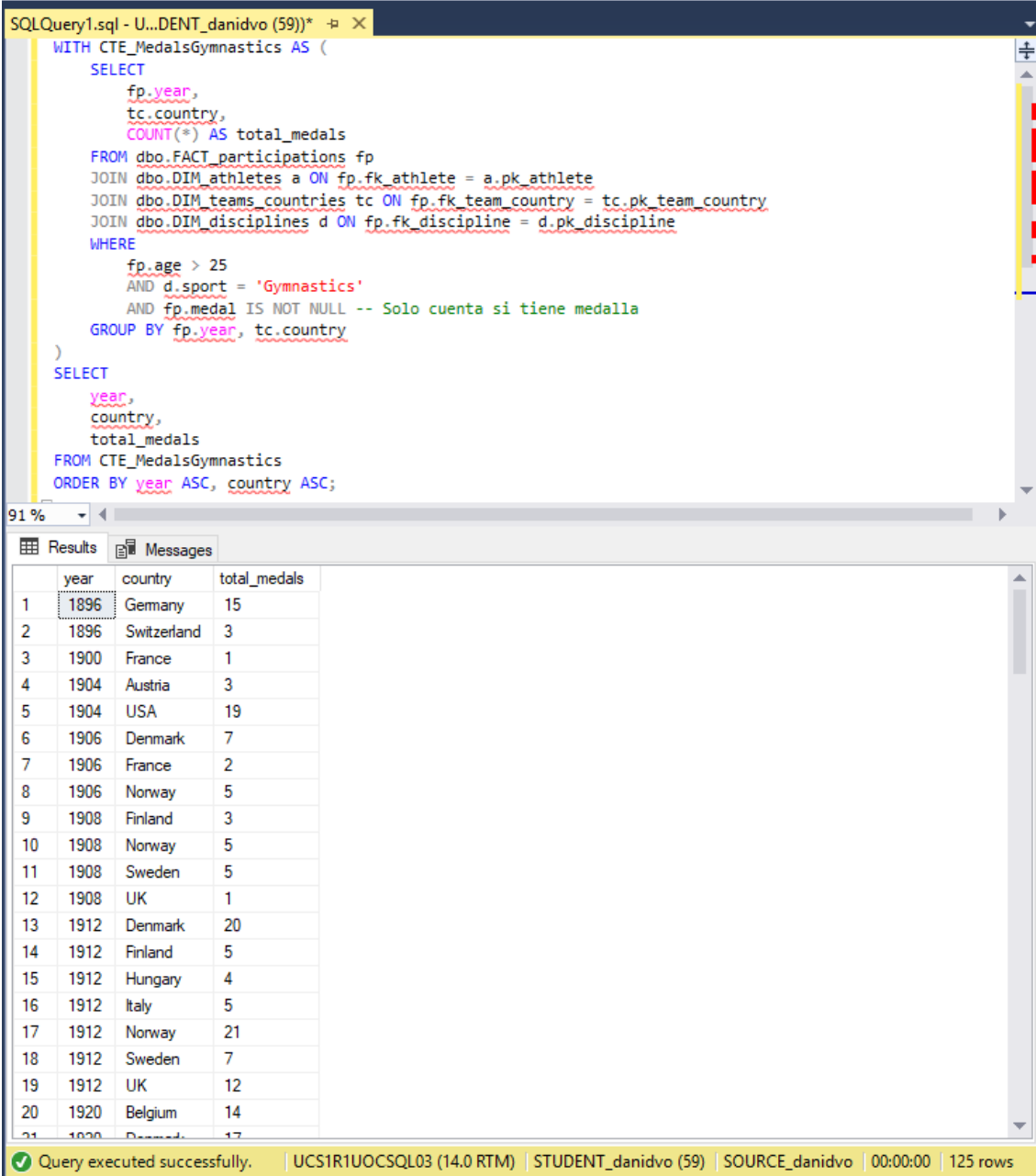
```
WITH CTE_FactParticipations AS (
    SELECT
        fp.year,
        tc.country,
        d.sport,
        COUNT(*) AS num_records
    FROM dbo.FACT_participations fp
    JOIN dbo.DIM_teams_countries tc ON fp.fk_team_country = tc.pk_team_country
    JOIN dbo.DIM_disciplines d ON fp.fk_discipline = d.pk_discipline
    GROUP BY fp.year, tc.country, d.sport
)
SELECT
    year,
    country,
    sport,
    num_records
FROM CTE_FactParticipations
ORDER BY year ASC, country ASC, sport ASC;
```

The query results are displayed in a table with the following columns: year, country, sport, and num_records. The results are ordered by year, country, and sport.

	year	country	sport	num_records
1	1896	Australia	Athletics	3
2	1896	Australia	Tennis	2
3	1896	Austria	Cycling	4
4	1896	Austria	Fencing	1
5	1896	Austria	Swimming	3
6	1896	Denmark	Athletics	4
7	1896	Denmark	Fencing	1
8	1896	Denmark	Gymnastics	1
9	1896	Denmark	Shooting	7
10	1896	Denmark	Weightlifting	2
11	1896	France	Athletics	12
12	1896	France	Cycling	7
13	1896	France	Fencing	4
14	1896	France	Gymnastics	1
15	1896	France	Shooting	1
16	1896	France	Tennis	1
17	1896	Germany	Athletics	14
18	1896	Germany	Cycling	10
19	1896	Germany	Gymnastics	66
20	1896	Germany	Tennis	2
21	1896	Germany	Weightlifting	1
22	1896	Germany	Wrestling	1
23	1896	Greece	Athletics	26

Query executed successfully. | UCS1R1UOCSQL03 (14.0 RTM) | STUDENT_danidvo (59) | SOURCE_danidvo | 00:00:00 | 26560 rows

2) Conocer el total de medallas obtenidas por los atletas mayores de 25 años, en el deporte Gymnastics, por año y país.



The screenshot displays a SQL query window titled 'SQLQuery1.sql - U...DENT_danidvo (59))' with a query that uses a Common Table Expression (CTE) to filter for gymnasts aged 25 and older, then groups the results by year and country to show the total number of medals. Below the query, the 'Results' tab shows a table with 21 rows of data, including columns for year, country, and total_medals. The status bar at the bottom indicates the query was executed successfully, returning 125 rows.

```
WITH CTE_MedalsGymnastics AS (
    SELECT
        fp.year,
        tc.country,
        COUNT(*) AS total_medals
    FROM dbo.FACT_participations fp
    JOIN dbo.DIM_athletes a ON fp.fk_athlete = a.pk_athlete
    JOIN dbo.DIM_teams_countries tc ON fp.fk_team_country = tc.pk_team_country
    JOIN dbo.DIM_disciplines d ON fp.fk_discipline = d.pk_discipline
    WHERE
        fp.age > 25
        AND d.sport = 'Gymnastics'
        AND fp.medal IS NOT NULL -- Solo cuenta si tiene medalla
    GROUP BY fp.year, tc.country
)
SELECT
    year,
    country,
    total_medals
FROM CTE_MedalsGymnastics
ORDER BY year ASC, country ASC;
```

	year	country	total_medals
1	1896	Germany	15
2	1896	Switzerland	3
3	1900	France	1
4	1904	Austria	3
5	1904	USA	19
6	1906	Denmark	7
7	1906	France	2
8	1906	Norway	5
9	1908	Finland	3
10	1908	Norway	5
11	1908	Sweden	5
12	1908	UK	1
13	1912	Denmark	20
14	1912	Finland	5
15	1912	Hungary	4
16	1912	Italy	5
17	1912	Norway	21
18	1912	Sweden	7
19	1912	UK	12
20	1920	Belgium	14
21	1920	Denmark	17

Query executed successfully. | UCS1R1UOCSQL03 (14.0 RTM) | STUDENT_danidvo (59) | SOURCE_danidvo | 00:00:00 | 125 rows

3) Analizar desde un punto de vista estadístico, para cada país, el número de atletas que han participado a lo largo de la historia de los Juegos Olímpicos, el total de medallas que ha conseguido el país, y el porcentaje de atletas que representan sobre el total de participantes del país. Ordenar los resultados por el porcentaje de forma descendente.

SQLQuery1.sql - U...DENT_danidvo (59) * X

```

WITH CTE_ParticipationStats AS (
    SELECT
        tc.country,
        COUNT(DISTINCT fp.fk_athlete) AS num_athletes, -- Número de atletas únicos
        COUNT(CASE WHEN fp.medal IS NOT NULL THEN 1 END) AS total_medals -- Total de medallas
    FROM dbo.FACT_participations fp
    JOIN dbo.DIM_teams_countries tc ON fp.fk_team_country = tc.pk_team_country
    GROUP BY tc.country
),
CTE_TotalAthletes AS (
    SELECT
        SUM(num_athletes) AS total_athletes -- Total de atletas en todos los países
    FROM CTE_ParticipationStats
)
SELECT
    ps.country,
    ps.num_athletes,
    ps.total_medals,
    CAST(ps.num_athletes AS FLOAT) / (SELECT total_athletes FROM CTE_TotalAthletes) * 100 AS participation_percentage -- Cálculo
FROM CTE_ParticipationStats ps
ORDER BY participation_percentage DESC; -- Orden descendente por porcentaje
  
```

	country	num_athletes	total_medals	participation_percentage
1	USA	9652	5638	7.11358745320008
2	Germany	7541	3766	5.55776657527785
3	UK	6273	2076	4.62324223932078
4	France	6161	1870	4.54069750302173
5	Russia	5597	3947	4.12502579523009
6	Italy	4921	1640	3.62680935113941
7	Canada	4810	1352	3.54500162141446
8	Japan	4037	914	2.97529553963621
9	Australia	3868	1349	2.85074142861354
10	Sweden	3782	1536	2.78735886324106
11	Poland	2964	567	2.18448748562837
12	Netherlands	2937	1048	2.16458830812771
13	Switzerland	2868	719	2.1137348545149
14	China	2861	1071	2.10857580849621
15	Czech Republic	2741	644	2.02013501960437
16	Hungary	2681	1135	1.97591462515846
17	Spain	2637	489	1.94348633589812
18	Finland	2346	900	1.72901742283541
19	Austria	2334	452	1.72017334394623
20	South Korea	2296	641	1.69216709413048

Query executed successfully. UCS1R1UOCSQL03 (14.0 RTM) | STUDENT_danidvo (59) | SOURCE_danidvo | 00:00:00 | 207 rows

Ln 14 Col 2 Ch 2 INS

Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

2.6. Ejercicio 5. Creación y explotación de un cubo OLAP (20%)

2.6.1. Creación del proyecto

The screenshot displays the Microsoft Visual Studio interface for a project named 'PR3_ProyectoMultidimensional'. The main workspace shows a data model diagram with a central fact table 'FACT_participations' and four dimension tables: 'DIM_disciplines', 'DIM_athletes', 'DIM_teams_countries', and 'DIM_teams_countries'. The 'weight' column in the fact table is highlighted. The right pane shows the 'Explorador de soluciones' (Solution Explorer) with the project structure, including data sources, views, cubes, and dimensions. The bottom pane shows the 'Propiedades' (Properties) window for the 'weight' data column.

Propiedades

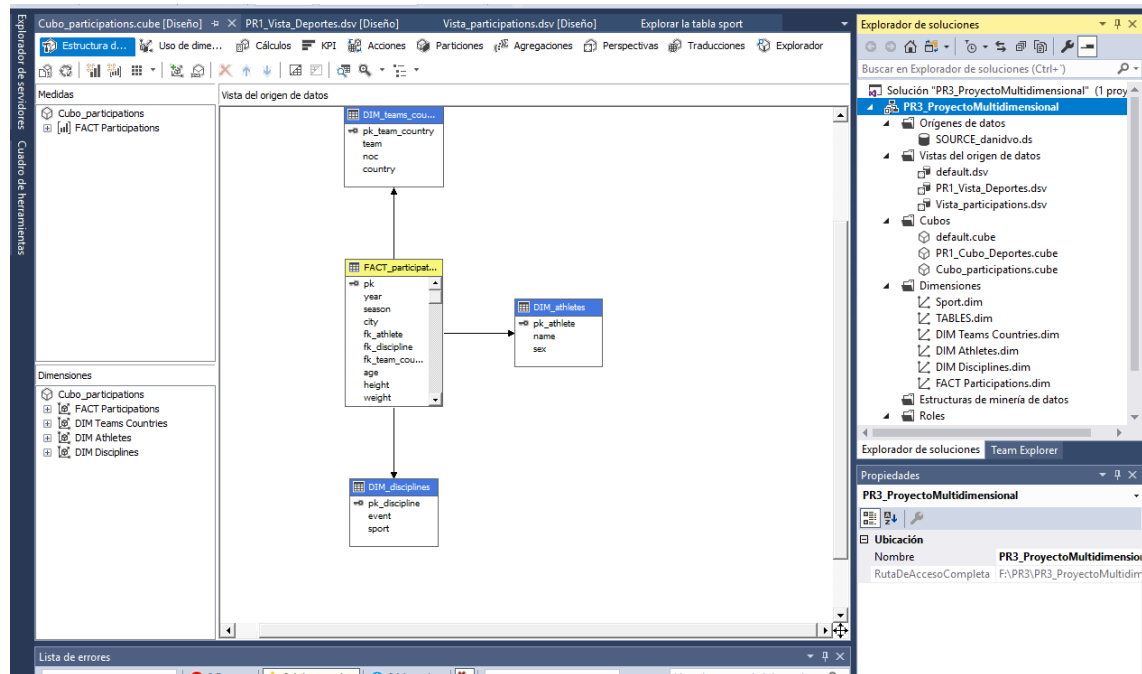
Propiedad	Valor
weight	DataColumn
AllowNull	True
DataType	System.Int32
DateTimeMode	UnspecifiedLocal
Description	
FriendlyName	weight
Length	-1
Name	weight

Daniel Vargus Olivencia

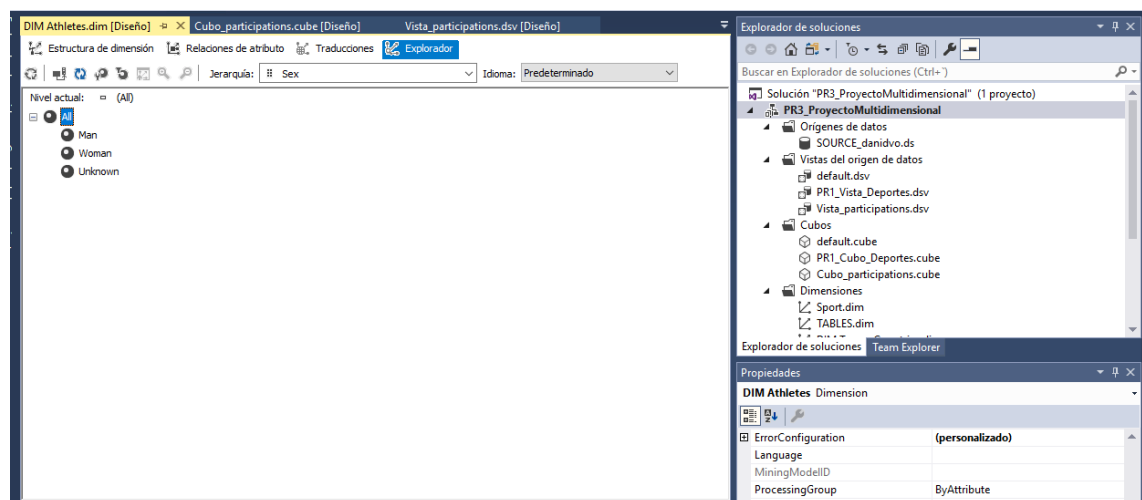
PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

Creación del Cubo



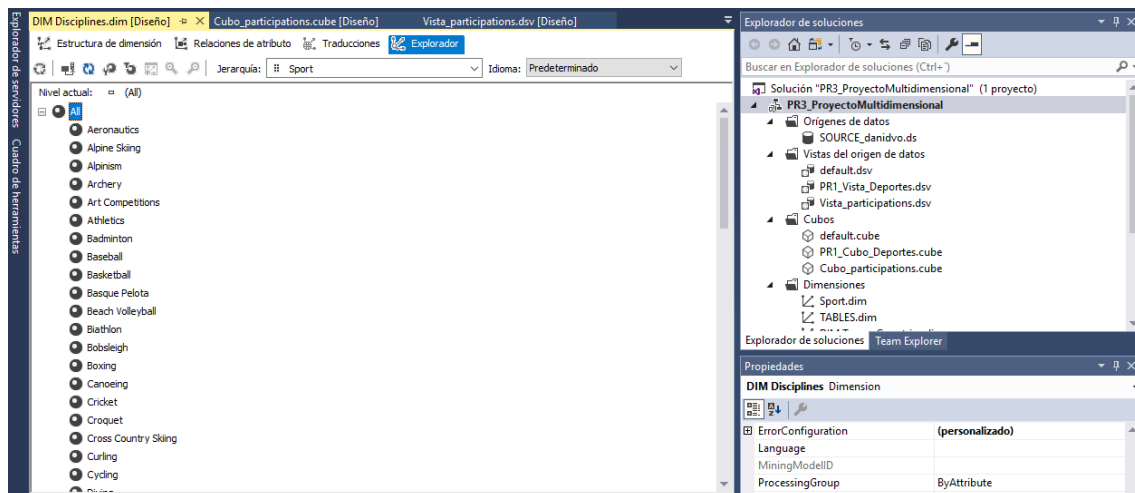
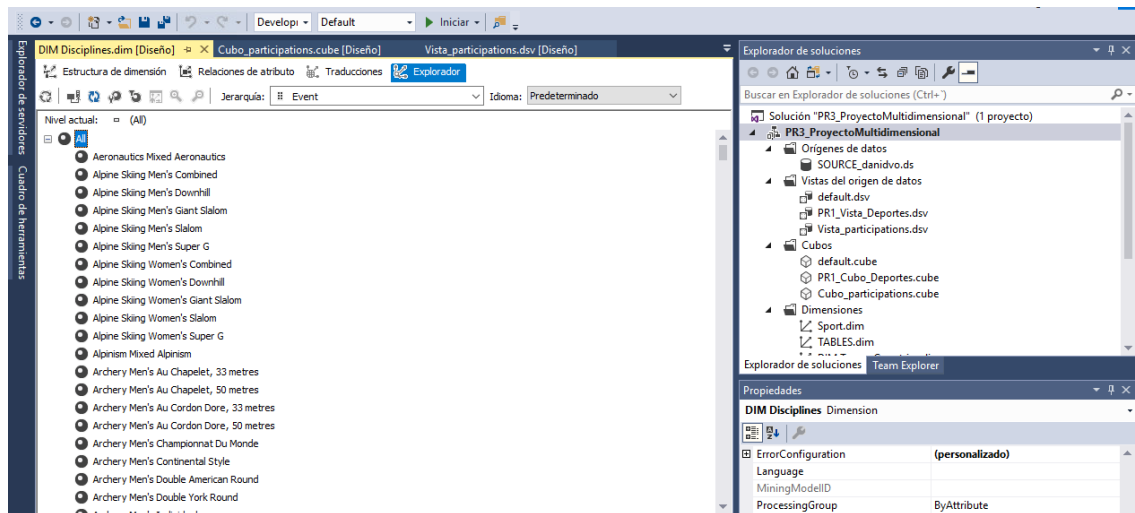
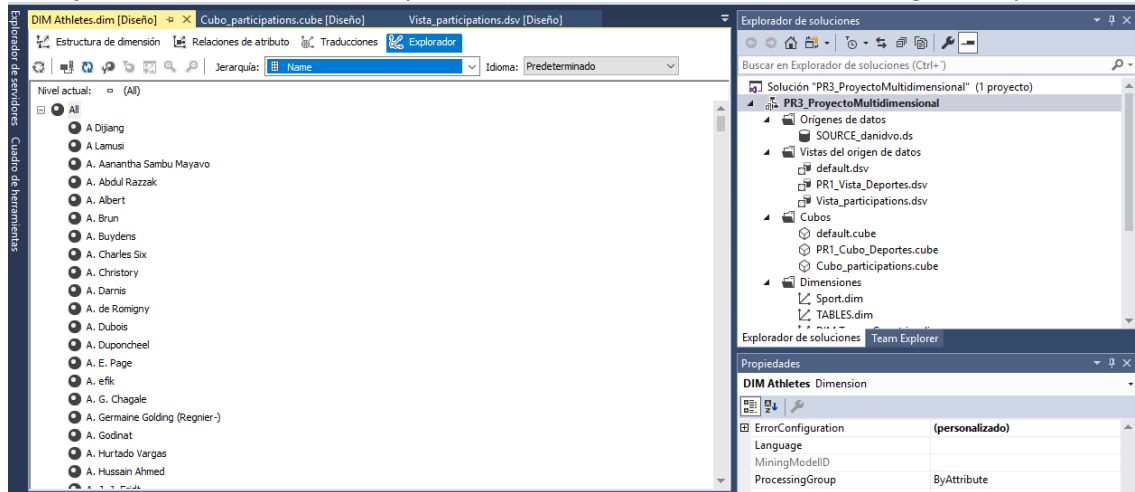
2.6.6. Adaptación de dimensiones



Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

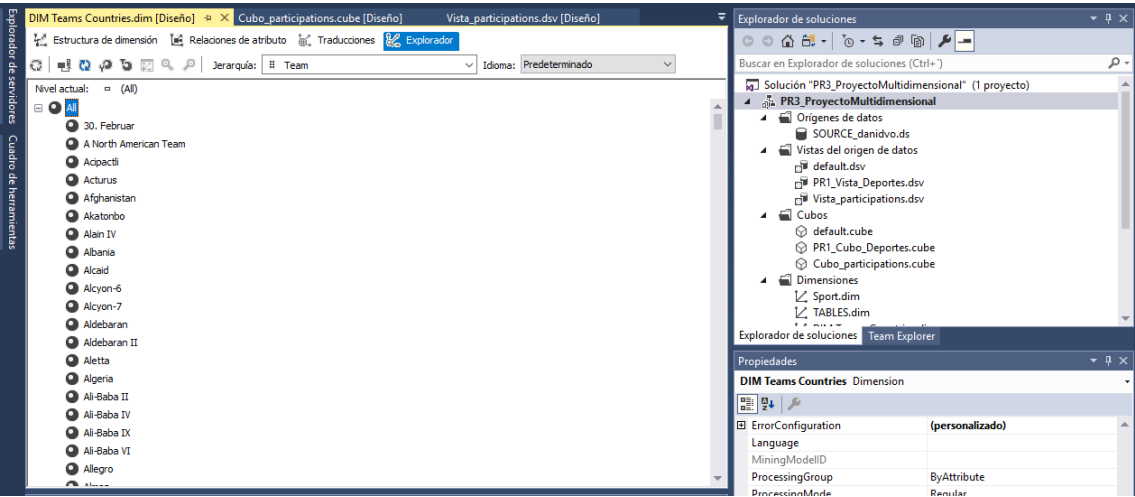
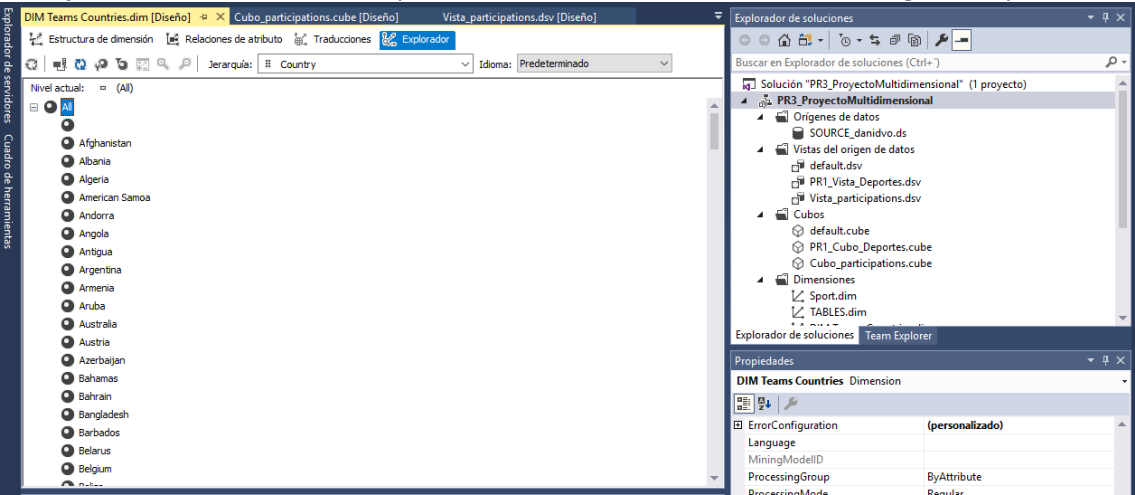
Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos



Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos



Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

Añadir DIM_YEAR

Editar consulta con nombre

Nombre: DIM_YEAR

Descripción:

Origen de datos: SOURCE_danidvo (principal)

Definición de la consulta:

FACT_participations

- ☐ * (Todas las columnas)
- ☐ pk
- ☒ year
- ☐ season
- ☐ city

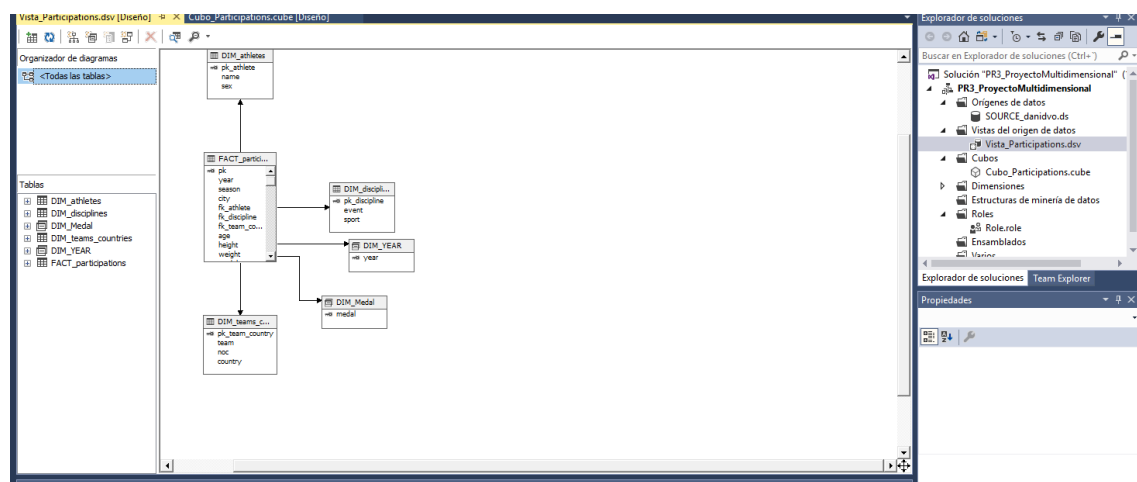
Columna	Alias	Tabla	Resul...	Tipo de orden	Criterio de or...	Filtro	O...
[year]		FACT_parti...	<input checked="" type="checkbox"/>				

SELECT DISTINCT [year]
FROM FACT_participations

year
1998
1904
2004
2010
1924

1 de 35 Celda de solo lectura.

Aceptar Cancelar Ayuda

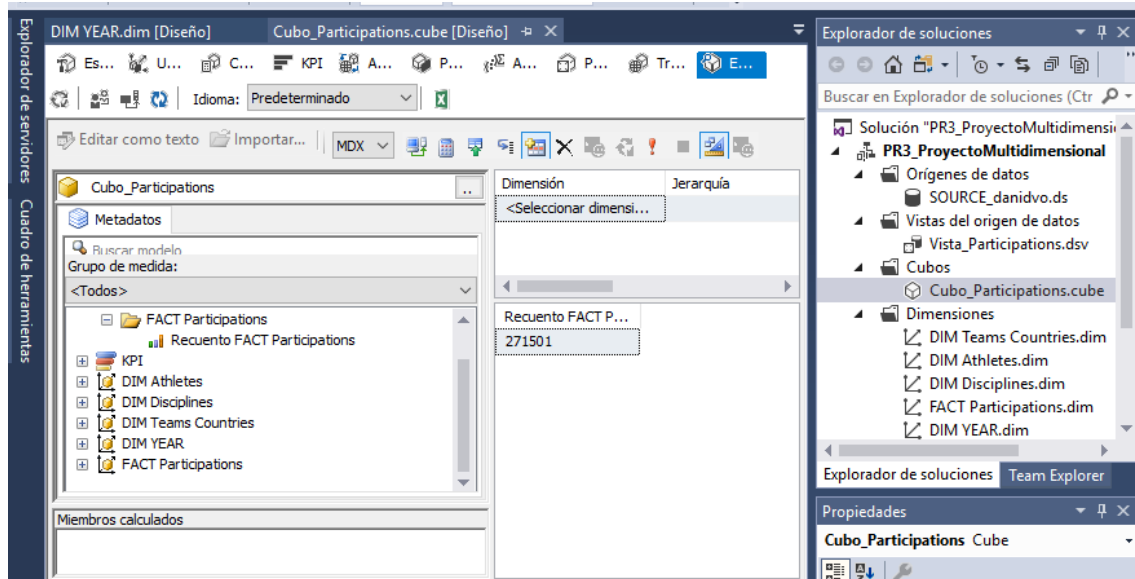


Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

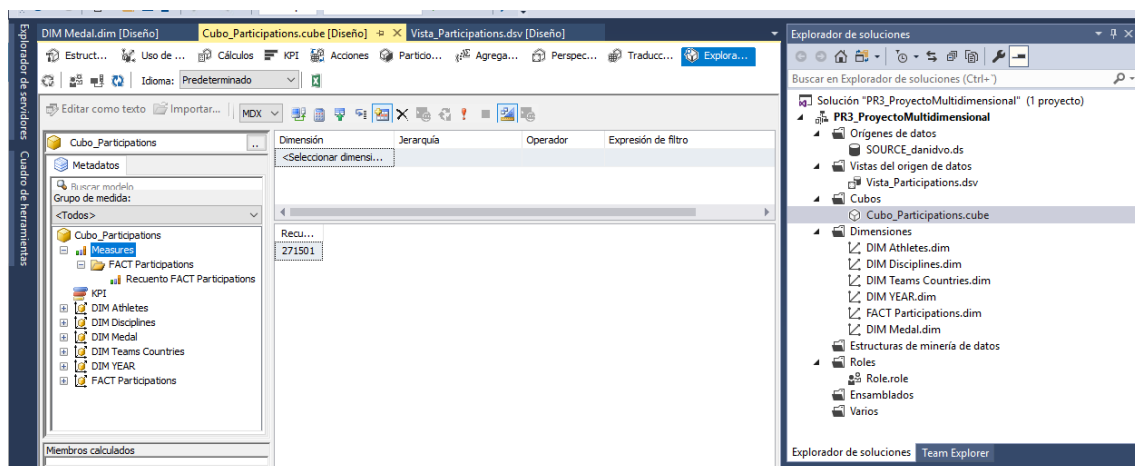
Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

PROCESADO DEL CUBO



2.6.7. Ejercicios a realizar

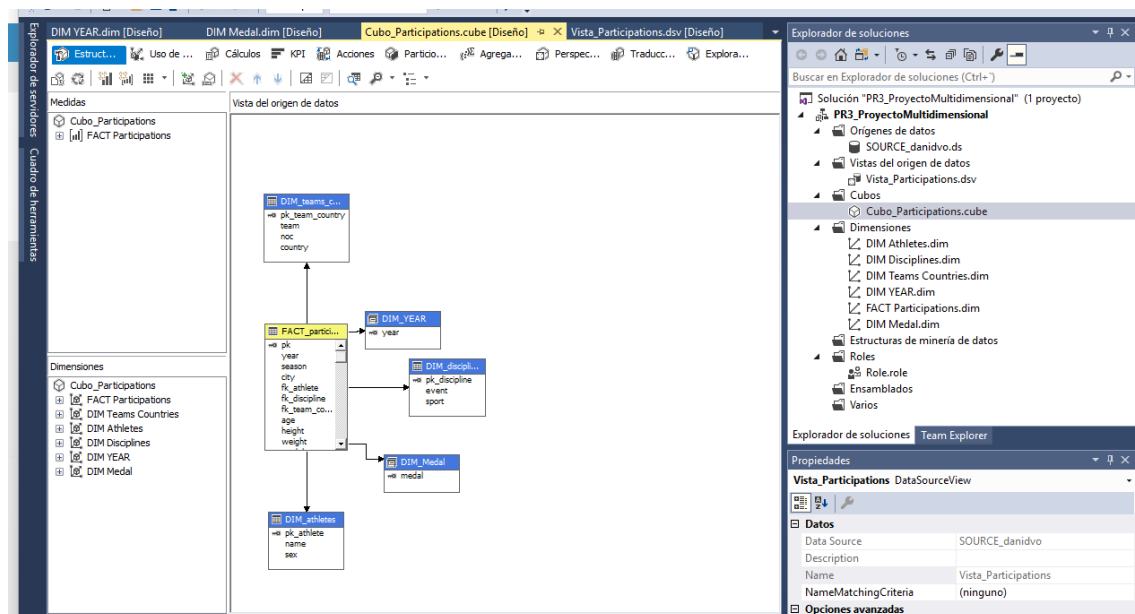
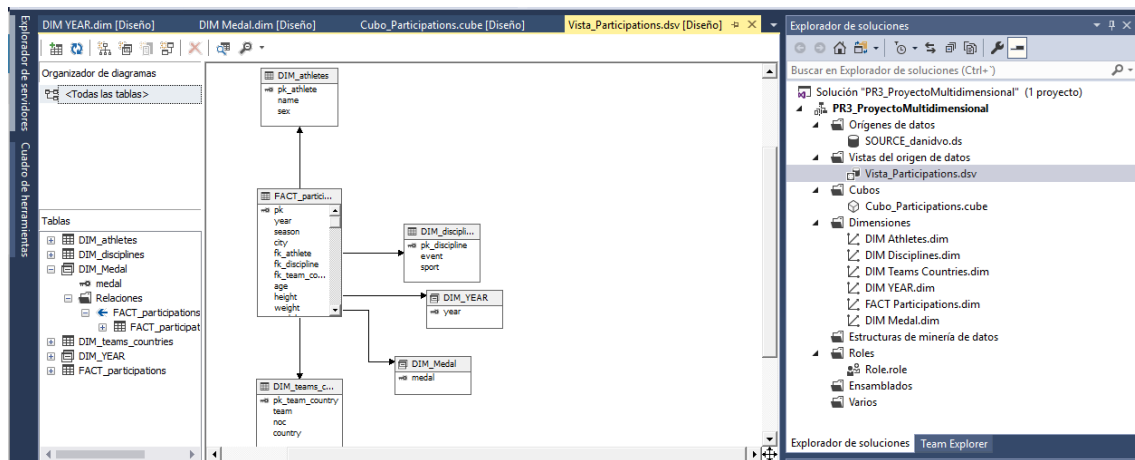
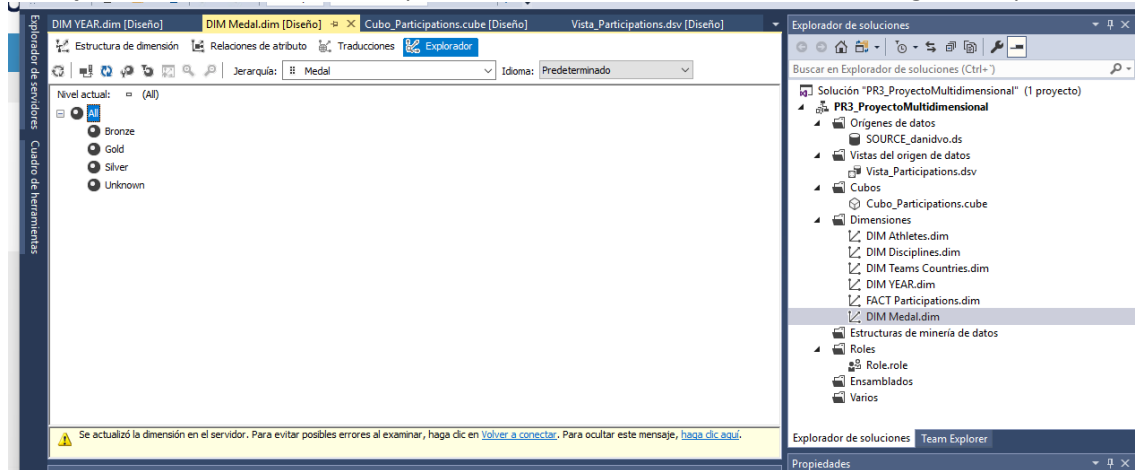
1) Se deberá extender las dimensiones del cubo agregando una dimensión más DIM_Medal para realizar el análisis de las medallas obtenidas. Esta dimensión se agregará siguiendo los mismos pasos que los indicados en la creación de la dimensión DIM_Year del enunciado.



Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos



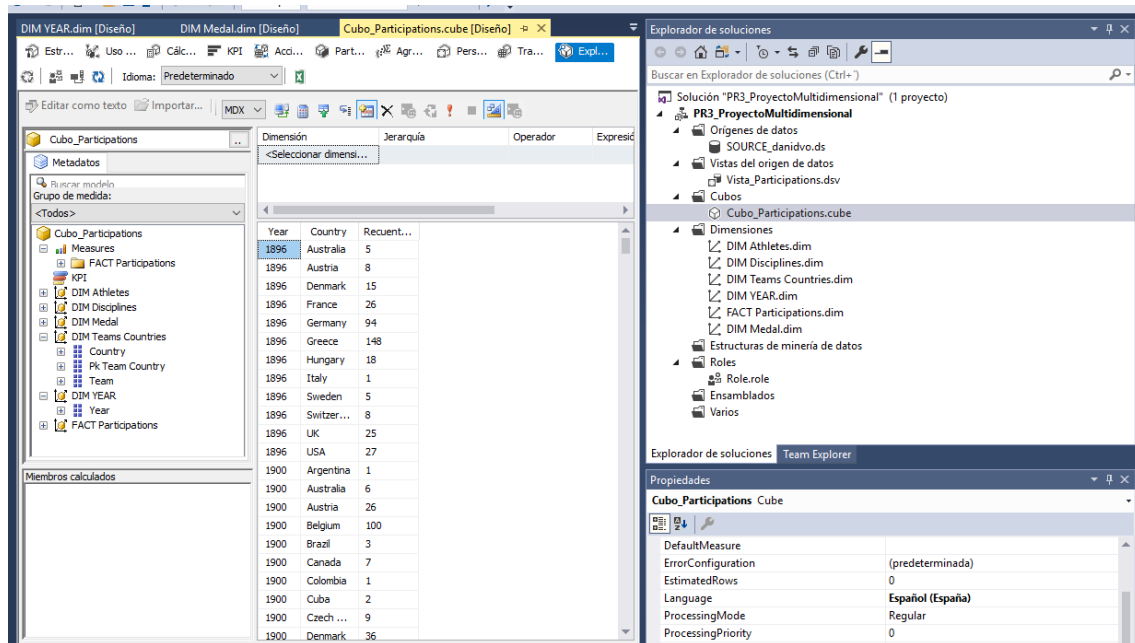
Daniel Vargas Olivencia

PR3 – Sistemas de bases de datos analíticas

Caso práctico: Almacén de datos para el análisis de los resultados de los Juegos Olímpicos

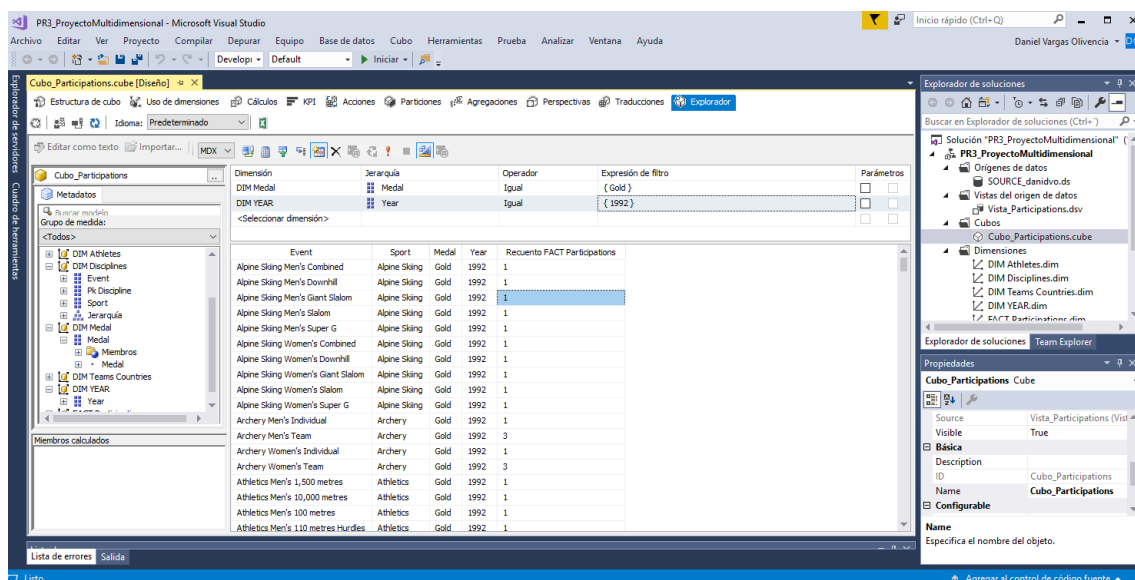
2) Se realizarán las siguientes consultas sobre el cubo OLAP que den respuestas a las siguientes cuestiones:

a) Conocer, por año y país, el total de participantes de las diferentes ediciones de los juegos olímpicos.



Year	Country	Recuent...
1896	Australia	5
1896	Austria	8
1896	Denmark	15
1896	France	26
1896	Germany	94
1896	Greece	148
1896	Hungary	18
1896	Italy	1
1896	Sweden	5
1896	Switzer...	8
1896	UK	25
1896	USA	27
1900	Argentina	1
1900	Australia	6
1900	Austria	26
1900	Belgium	100
1900	Brazil	3
1900	Canada	7
1900	Colombia	1
1900	Cuba	2
1900	Czech ...	9
1900	Denmark	36

b) Conocer, por deporte y especialidad, el total de participantes que han obtenido una medalla de Oro en la edición de 1992.



Event	Sport	Medal	Year	Recuento FACT Participations
Alpine Skiing Men's Combined	Alpine Skiing	Gold	1992	1
Alpine Skiing Men's Downhill	Alpine Skiing	Gold	1992	1
Alpine Skiing Men's Giant Slalom	Alpine Skiing	Gold	1992	1
Alpine Skiing Men's Slalom	Alpine Skiing	Gold	1992	1
Alpine Skiing Men's Super G	Alpine Skiing	Gold	1992	1
Alpine Skiing Women's Combined	Alpine Skiing	Gold	1992	1
Alpine Skiing Women's Downhill	Alpine Skiing	Gold	1992	1
Alpine Skiing Women's Giant Slalom	Alpine Skiing	Gold	1992	1
Alpine Skiing Women's Slalom	Alpine Skiing	Gold	1992	1
Alpine Skiing Women's Super G	Alpine Skiing	Gold	1992	1
Archery Men's Individual	Archery	Gold	1992	1
Archery Men's Team	Archery	Gold	1992	3
Archery Women's Individual	Archery	Gold	1992	1
Archery Women's Team	Archery	Gold	1992	3
Athletics Men's 1,500 metres	Athletics	Gold	1992	1
Athletics Men's 10,000 metres	Athletics	Gold	1992	1
Athletics Men's 100 metres	Athletics	Gold	1992	1
Athletics Men's 110 metres Hurdles	Athletics	Gold	1992	1