

# Olympics Statistics

---

## **Group 5**

Ashita Shetty  
Ashutosh Somaiya  
Mansi Tandel  
Shivani Tergaonkar  
Surya Kiran Udaya Kumar

# Contents

- Introduction
  1. Dataset description
  2. Dataset
  3. ER Diagram
- Table Definitions
- Queries execution using oracle and hive.
  1. Query runtime comparison.
  2. Oracle Explain Plans.
  3. Proposed query plan
- Hadoop Configuration
- Comparison Graph
- Challenges
- Conclusion.

# Introduction

## Dataset description:

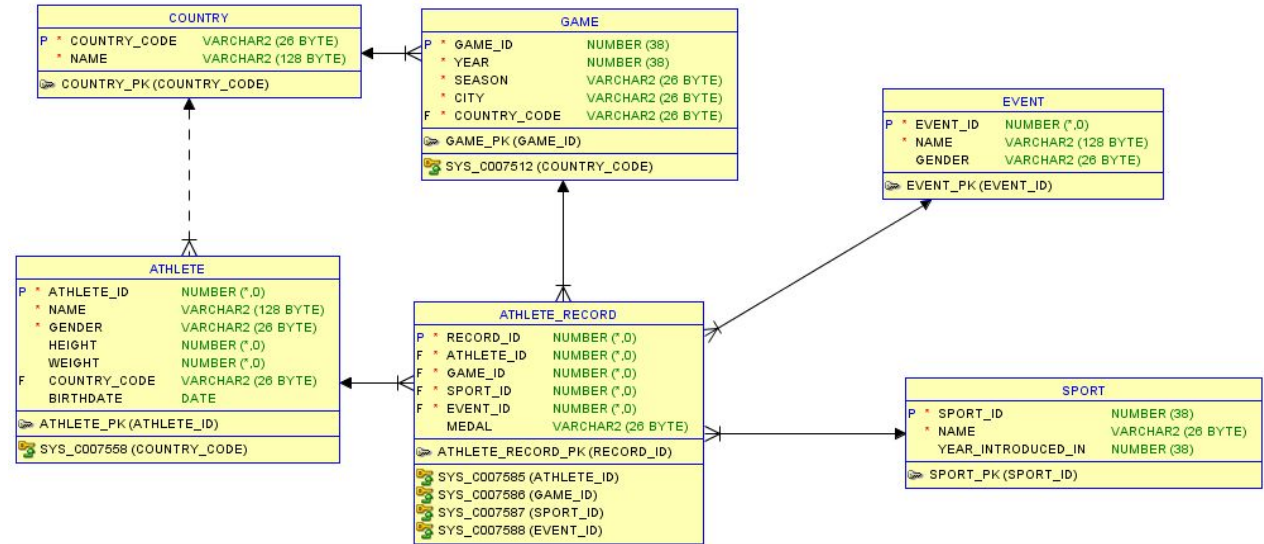
- **Country:** Contains country code and country name.
- **Sport:** Contains data of sports and the date they were introduced.
- **Game:** Contains the year in which the games were held along with the country and the city where they were played.
- **Event:** Contains different games according to their genders.
- **Athlete:** Description of the athlete along with the country\_id they play for.
- **Athlete\_record:** Contains the primary key of every table along with the medals for each player.

# Dataset

The Olympics dataset comprises of following tables:

<b>Table Name</b>	<b>Total no of Rows</b>
Country	219
Game	52
Event	765
Sport	67
Athlete	134729
Athlete_record	271114

# Er diagram



# Table definitions I

## 1. Country

	❖ COLUMN_NAME	❖ DATA_TYPE	❖ NULLABLE	DATA_DEFAULT	❖ COLUMN_ID	❖ COMMENTS
1	COUNTRY_CODE	VARCHAR2(26 BYTE)	No	(null)	1	(null)
2	NAME	VARCHAR2(128 BYTE)	No	(null)	2	(null)

## 2. Game

	❖ COLUMN_NAME	❖ DATA_TYPE	❖ NULLABLE	DATA_DEFAULT	❖ COLUMN_ID	❖ COMMENTS
	GAME_ID	NUMBER(38,0)	No	(null)	1	(null)
	YEAR	NUMBER(38,0)	No	(null)	2	(null)
	SEASON	VARCHAR2(26 BYTE)	No	(null)	3	(null)
	CITY	VARCHAR2(26 BYTE)	No	(null)	4	(null)
	COUNTRY_CODE	VARCHAR2(26 BYTE)	No	(null)	5	(null)

# Table definitions II

## 1. Event

❖	COLUMN_NAME	❖	DATA_TYPE	❖	NULLABLE	DATA_DEFAULT	❖	COLUMN_ID	❖	COMMENTS
1	EVENT_ID		NUMBER(38,0)		No	(null)		1		(null)
2	NAME		VARCHAR2(128 BYTE)		No	(null)		2		(null)
3	GENDER		VARCHAR2(26 BYTE)		Yes	(null)		3		(null)

## 2. Athlete

❖	COLUMN_NAME	❖	DATA_TYPE	❖	NULLABLE	DATA_DEFAULT	❖	COLUMN_ID	❖	COMMENTS
1	ATHLETE_ID		NUMBER(38,0)		No	(null)		1		(null)
2	NAME		VARCHAR2(128 BYTE)		No	(null)		2		(null)
3	GENDER		VARCHAR2(26 BYTE)		No	(null)		3		(null)
4	HEIGHT		NUMBER(38,0)		Yes	(null)		4		(null)
5	WEIGHT		NUMBER(38,0)		Yes	(null)		5		(null)
6	COUNTRY_CODE		VARCHAR2(26 BYTE)		Yes	(null)		6		(null)
7	BIRTHDATE		DATE		Yes	(null)		7		(null)

# Table definitions III

## 1. Athlete\_record

❖	COLUMN_NAME	❖	DATA_TYPE	❖	NULLABLE	DATA_DEFAULT	❖	COLUMN_ID	❖	COMMENTS
1	RECORD_ID		NUMBER(38,0)		No	(null)		1		(null)
2	ATHLETE_ID		NUMBER(38,0)		No	(null)		2		(null)
3	GAME_ID		NUMBER(38,0)		No	(null)		3		(null)
4	SPORT_ID		NUMBER(38,0)		No	(null)		4		(null)
5	EVENT_ID		NUMBER(38,0)		No	(null)		5		(null)
6	MEDAL		VARCHAR2(26 BYTE)		Yes	(null)		6		(null)

## 2. Sport


❖	COLUMN_NAME	❖	DATA_TYPE	❖	NULLABLE	DATA_DEFAULT	❖	COLUMN_ID	❖	COMMENTS
1	SPORT_ID		NUMBER(38,0)		No	(null)		1		(null)
2	NAME		VARCHAR2(26 BYTE)		No	(null)		2		(null)
3	YEAR INTRODUCED IN		NUMBER(38,0)		Yes	(null)		3		(null)



# Queries - Select without predicate(Oracle and Hive)

Fetches all of the data from the athlete table.

*SELECT \* from athlete;*



ATHLETE_ID	NAME	GENDER	HEIGHT	WEIGHT	COUNTRY_CODE	BIRTHDATE
134717	133890 Zenons Popovs	M	(null)	(null)	LAT	(null)
134718	133900 Zewdie Hailemariam	F	170	51	ETH	(null)
134719	133920 Zhang Anzhi	M	(null)	(null)	CHN	(null)
134720	133921 Zhang Banglun	M	(null)	(null)	CHN	(null)
134721	133972 Zhang Jinhai	M	(null)	(null)	CHN	(null)
134722	134281 Zhu Yongqiang	M	(null)	(null)	CHN	(null)
134723	134292 Ziaeddin Shademan	M	(null)	(null)	IRI	(null)
134724	134332 Ziya Azak	M	(null)	(null)	TUR	(null)
134725	134477 Zolner, Emil	M	(null)	(null)	RUS	(null)

Oracle SQL Developer T(Oracle) = 5.918 sec

38608	Friedrich Lthi	M	NULL	NULL	SUI	NULL		
17846	"Carl Emil "Carl-Emil" Johansson"	M	NULL	NULL	SWE	NULL		
95575	Olaf Emil Frydenlund	M	NULL	NULL	NOR	NULL		
1130	Adolf Karl Oskar Davids	M	NULL	NULL	GER	NULL		
54439	Jacques Baudrier	M	NULL	NULL	FRA	NULL		
17913	Carl Johan Eldh	M	NULL	NULL	SWE	NULL		
61647	John Richardson	M	NULL	NULL	RSA	NULL		
63629	Josef (Jaroslav) Suk	M	NULL	NULL	TCH	NULL		
82529	Marie-Charles-Andr Corvington	M	NULL	NULL	HAI	NULL		
63975	"Joseph "Joe" Jackson"	M	183	NULL	USA	NULL		
129987	William Morton	M	NULL	NULL	CAN	NULL		
61649	John Riegel DeWitt	M	NULL	NULL	USA	NULL		
116783	Stanley Howard Shoveller	M	NULL	NULL	GBR	NULL		
130032	William Philo	M	NULL	NULL	GBR	NULL		
17947	Carl Marinus Jensen	M	NULL	85	DEN	NULL		
60041	"Johan August "Janne" Gustafsson"	M	NULL	NULL	SWE	NULL		

Time taken: 1.616 seconds, Fetched: 134731 row(s)

Hive T(Hive) = 1.616 sec

# Queries - Select without predicate(Oracle explain plan and proposed query plan)

Query Result x Explain Plan x

SQL 0.111 seconds

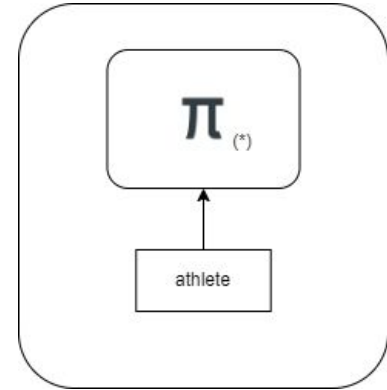
OPERATION	OBJECT_NAME	OPTIONS	CARDINALITY	COST
SELECT STATEMENT				254
TABLE ACCESS	ATHLETE	FULL	134731	254

Other XML

{info}

- info type="has\_user\_tab"
  - yes
- info type="db\_version"
  - 19.0.0.0
- info type="parse\_schema"
  - "SYSTEM"
- info type="plan\_hash\_full"
  - 2460843779
- info type="plan\_hash"
  - 81540165
- info type="plan\_hash\_2"
  - 2460843779
- stats type="compilation"
  - stat name="bg"
    - 0
- q o="2" f="y"
  - {n}
  - SEL \$1
    - {t}
    - ATHLETE
    - SEL \$1

Oracle explain plan



# Queries - Select with predicate(Oracle and Hive)

## List all female athletes

*SELECT \* from athlete where gender = "F";*

 All Rows Fetched: 33778 in 1.072 seconds

ATHLETE_ID	NAME	GENDER	HEIGHT	WEIGHT	COUNTRY_CODE	BIRTHDATE
33770	132230 Yolande Kada-Gango	F	(null)	(null)	CGO	(null)
33771	132232 Yolande Plancke	F	(null)	(null)	FRA	(null)
33772	133483 Yvonne Conte	F	(null)	(null)	FRA	(null)
33773	133501 Yvonne Makouala	F	(null)	(null)	CGO	(null)
33774	133506 Yvonne Meley	F	(null)	(null)	FRA	(null)
33775	133512 Yvonne Prvost (-Boppe)	F	(null)	(null)	FRA	(null)
33776	133522 Yvonne Van Bets	F	(null)	(null)	BEL	(null)
33777	133900 Zewdie Hailemariam	F	170	51	ETH	(null)
33778	134552 Zoya Blyuvas	F	(null)	(null)	URS	(null)

Oracle SQL Developer T(Oracle) = 1.072  
sec

132230	Yolande Kada-Gango	F	NULL	NULL	CGO	NULL
132232	Yolande Plancke	F	NULL	NULL	FRA	NULL
133483	Yvonne Conte	F	NULL	NULL	FRA	NULL
133501	Yvonne Makouala	F	NULL	NULL	CGO	NULL
133506	Yvonne Meley	F	NULL	NULL	FRA	NULL
133512	Yvonne Prvost (-Boppe)	F	NULL	NULL	FRA	NULL
133522	Yvonne Van Bets	F	NULL	NULL	BEL	NULL
133900	Zewdie Hailemariam	F	170	51	ETH	NULL
134552	Zoya Blyuvas	F	NULL	NULL	URS	NULL

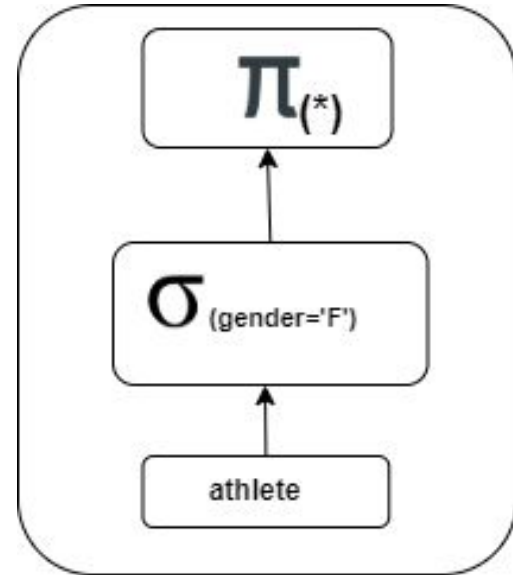
Time taken: 0.811 seconds, Fetched: 33779 row(s)

Hive T(Hive) = 0.811 sec

# Queries - Select with predicate(Oracle explain plan and proposed query plan)

OPERATION	OBJECT_NAME	OPTIONS	CARDINALITY	COST
SELECT STATEMENT				33779 254
TABLE ACCESS	ATHLETE	FULL		33779 254
Filter Predicates				
GENDER='F'				
Other XML				
{info}				
info type="has_user_tab"				
yes				
info type="db_version"				
19.0.0.0				
info type="parse_schema"				
"SYSTEM"				
info type="plan_hash_full"				
2460843779				
info type="plan_hash"				
81540165				
info type="plan_hash_2"				
2460843779				
stats type="compilation"				
stat name="bg"				
0				
q.o="2" f="y"				


Oracle explain plan



# Queries - Select with predicate IN(Oracle and Hive)

List all athletes from country - IND and USA

*SELECT \* from athlete where country\_code IN ('USA','IND');*

 All Rows Fetched: 10481 in 0.308 seconds

ATHLETE_ID	NAME	GENDER	HEIGHT	WEIGHT	COUNTRY_CODE	BIRTHDATE
10455	119374 Taj Mohammed	M	(null)	(null)	IND	(null)
10456	119988 Tarashkeswar Pandey	M	(null)	(null)	IND	(null)
10457	120437 Temba Sherpa	M	(null)	(null)	IND	(null)
10458	120867 Thazhatheri Abdul Rehman	M	(null)	(null)	IND	(null)
10459	121168 Thomas "Tom" Hunter	M	(null)	(null)	USA	(null)
10460	124971 Valli Asari Mookan	M	(null)	55	IND	(null)
10461	127776 W. A. Hersey	M	(null)	(null)	USA	(null)
10462	127785 W. J. Ross	M	(null)	(null)	USA	(null)
10463	127789 W. R. Gibson	M	(null)	(null)	USA	(null)

Oracle SQL Developer T(Oracle) = 0.308 sec

119374	Taj Mohammed	M	NULL	NULL	IND	NULL		
119988	Tarashkeswar Pandey	M	NULL	NULL	IND	NULL		
120437	Temba Sherpa	M	NULL	NULL	IND	NULL		
120867	Thazhatheri Abdul Rehman	M	NULL	NULL	IND	NULL		
121168	Thomas Tom Hunter	M	NULL	NULL	USA	NULL		
124971	Valli Asari Mookan	M	NULL	55	IND	NULL		
127776	W. A. Hersey	M	NULL	NULL	USA	NULL		
127785	W. J. Ross	M	NULL	NULL	USA	NULL		
127789	W. R. Gibson	M	NULL	NULL	USA	NULL		
Time taken: 0.367 seconds, Fetched: 10481 row(s)								

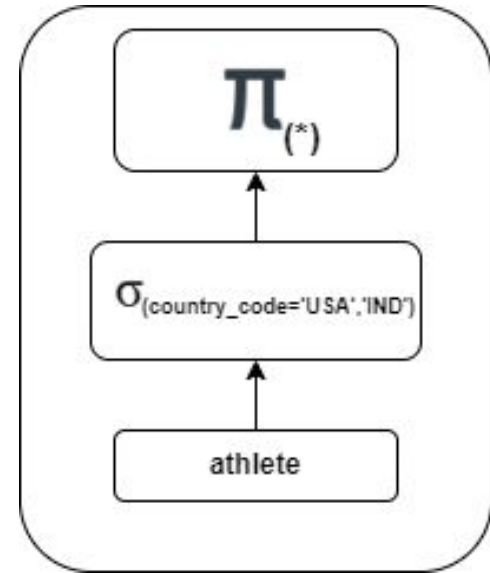
Hive T(Hive) = 0.367 sec

# Queries - Select with predicate IN(Oracle explain plan and proposed query plan)

SQL | 0.026 seconds

OPERATION	OBJECT_NAME	OPTIONS	CARDINALITY	COST
SELECT STATEMENT			10481	274
TABLE ACCESS	ATHLETE	FULL	10481	274
Filter Predicates				
VOR				
COUNTRY_CODE='IND'				
COUNTRY_CODE='USA'				

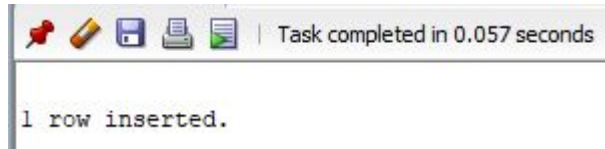
Oracle explain plan



# Queries- Insert (Oracle and Hive)

*INSERT INTO event (event\_id,name,gender) VALUES (766,'Test event 1','M');*

*SELECT \* FROM event WHERE event\_id = 766;*



	EVENT_ID	NAME	GENDER
1	766	Test event 1	M

**Oracle SQL Developer T(Oracle) = 0.057s**

OK  
Time taken: 2.063 seconds

766 Test event 1 M  
Time taken: 0.188 seconds, Fetched: 1 row(s)

**Hive T(Hive) = 0.188s**

# Queries- Insert (Oracle explain plan)

OPERATION	OBJECT_NAME	OPTIONS	CARDINALITY	COST
 INSERT STATEMENT			1	1
 LOAD TABLE CONVENTIONAL	<a href="#">EVENT</a>			

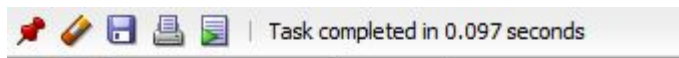
Oracle explain plan



# Queries- Update (Oracle and Hive)

*Update event set name='Test event 2' where event\_id = 766;*

*select \* from event where event\_id = 766;*



1 row updated.

```
hive> Update event set name='Test event 2' where event_id = 766;  
FAILED: SemanticException [Error 10294]: Attempt to do update or delete using tr  
ansaction manager that does not support these operations.
```

Hive doesn't support

A screenshot of the Oracle SQL Developer interface showing a query result. At the top, there is a toolbar with icons for a pin, a print icon, a refresh icon, a delete icon, and a document icon. To the right of the toolbar, a status bar indicates "All Rows Fetched: 1 in 0.002 seconds". Below the toolbar is a table with three columns: EVENT\_ID, NAME, and GENDER. The table contains one row with the values 1, 766 Test event 2, and M.

EVENT_ID	NAME	GENDER
1	766 Test event 2	M

Oracle SQL Developer T(Oracle) = 0.097 sec

Hive

# Queries- Update (Oracle explain plan)

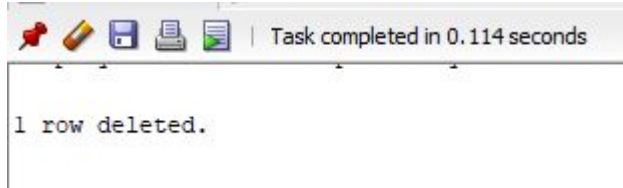
OPERATION	OBJECT_NAME	OPTIONS	CARDINALITY	COST
UPDATE STATEMENT			1	2
UPDATE	<a href="#">EVENT</a>			
INDEX	<a href="#">SYS_C0012863</a>	UNIQUE SCAN	1	1
Access Predicates	<a href="#">EVENT_ID=766</a>			

Oracle explain plan

# Queries- Delete (Oracle and Hive)

*Delete event where event\_id = 766;*

*select \* from event where event\_id = 766;*








Hive doesn't support

Oracle SQL Developer T(Oracle) 0.114 sec

Hive

# Queries- Delete (Oracle explain plan)

 SQL  | 0.046 seconds

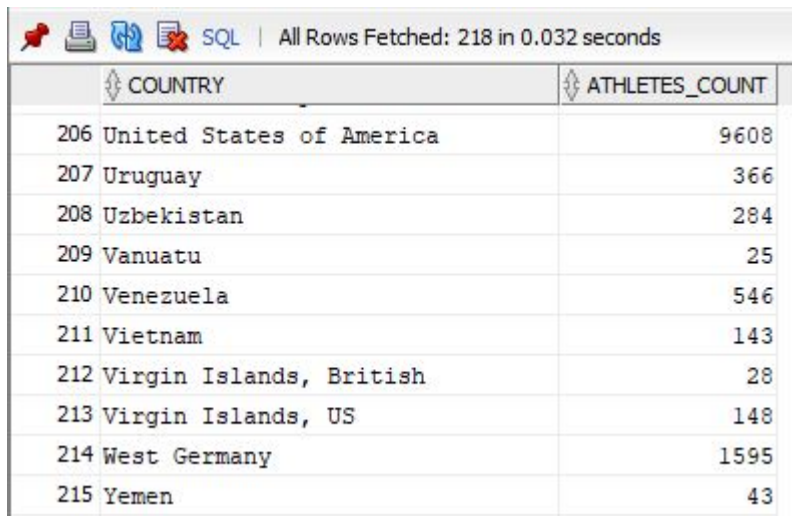
OPERATION	OBJECT_NAME	OPTIONS	CARDINALITY	COST
 DELETE STATEMENT			1	1
 DELETE	<a href="#">EVENT</a>			
 INDEX	<a href="#">SYS_C0012863</a>	UNIQUE SCAN	1	1
 Access Predicates	<a href="#">EVENT_ID=766</a>			

Oracle explain plan

# Queries - Aggregate (Oracle and Hive)

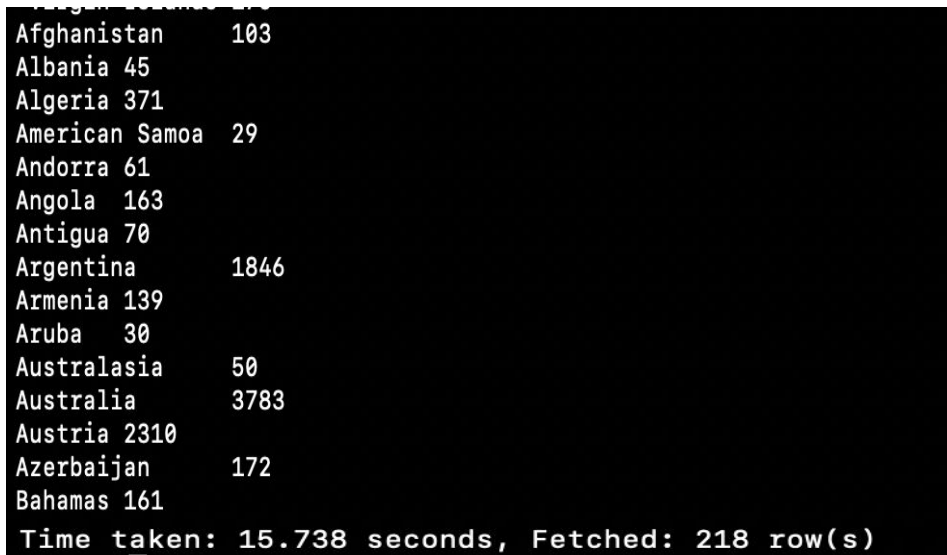
Count of all athletes grouped by country.

*SELECT c.name as country, count(a.athlete\_id) as athletes\_count FROM athlete a inner join country c on a.country\_code=c.country\_code group by c.name order by c.name;*



	COUNTRY	ATHLETES_COUNT
206	United States of America	9608
207	Uruguay	366
208	Uzbekistan	284
209	Vanuatu	25
210	Venezuela	546
211	Vietnam	143
212	Virgin Islands, British	28
213	Virgin Islands, US	148
214	West Germany	1595
215	Yemen	43

Oracle SQL Developer T(oracle) 0.032 sec



Afghanistan	103
Albania	45
Algeria	371
American Samoa	29
Andorra	61
Angola	163
Antigua	70
Argentina	1846
Armenia	139
Aruba	30
Australasia	50
Australia	3783
Austria	2310
Azerbaijan	172
Bahamas	161
Time taken: 15.738 seconds, Fetched: 218 row(s)	

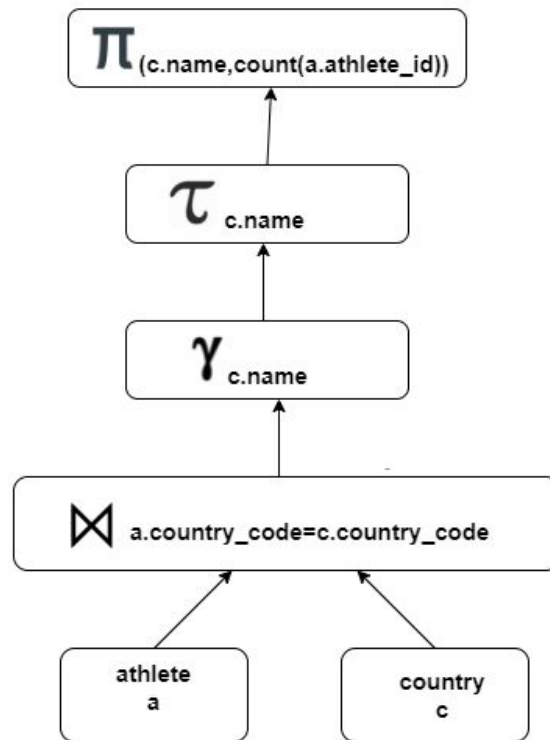
Hive T(Hive) 15.738 sec

# Queries - Aggregate (Oracle explain plan and query optimizer)

SQL | 0.074 seconds

OPERATION	OBJECT_NAME	OPTIONS	CARDINALITY	COST
SELECT STATEMENT			219	281
SORT		GROUP BY	219	281
HASH JOIN			134731	277
Access Predicates				
A.COUNTRY_CODE=C.COUNTRY_CODE				
TABLE ACCESS	COUNTRY	FULL	219	3
TABLE ACCESS	ATHLETE	FULL	134731	273

Oracle explain plan



# Queries - Inner join (Oracle and Hive)

List of countries with a total number of female gold medalists.

```
SELECT c.name, MAX(ar.medal) AS type_of_medal, COUNT(ar.medal) AS total_female_gold_medalists FROM athlete_record ar INNER JOIN athlete a ON a.athlete_id = ar.athlete_id INNER JOIN country c ON c.country_code = a.country_code WHERE ar.medal = 'Gold' AND a.gender = 'F' GROUP BY c.name ORDER BY total_female_gold_medalists DESC;
```

📌 🖨️ 🔄 🗑️ SQL | All Rows Fetched: 77 in 0.303 seconds

NAME	TYPE_OF_MEDAL	TOTAL_FEMALE_GOLD_MEDALISTS
Azerbaijan	Gold	1
66 Morocco	Gold	1
67 Costa Rica	Gold	1
68 Uzbekistan	Gold	1
69 Hongkong	Gold	1
70 Latvia	Gold	1
71 Puerto Rico	Gold	1
72 Individual Olympic Athletes	Gold	1
73 Albania	Gold	1
74 Australasia	Gold	1
75 Nigeria	Gold	1
76 Serbia	Gold	1
77 Syria	Gold	1

Oracle SQL Developer T(Oracle) = 0.303 s

```
United States of America      Gold      852
Soviet union      Gold      379
China      Gold      208
East Germany      Gold      203
Germany      Gold      194
Russia      Gold      185
Australia      Gold      178
Canada      Gold      158
Netherlands      Gold      140
Romania      Gold      121
South Korea      Gold      110
UK      Gold      99
Norway      Gold      77
Japan      Gold      72
Hungary      Gold      67
Italy      Gold      60
France      Gold      55
Time taken: 16.11 seconds, Fetched: 77 row(s)
```

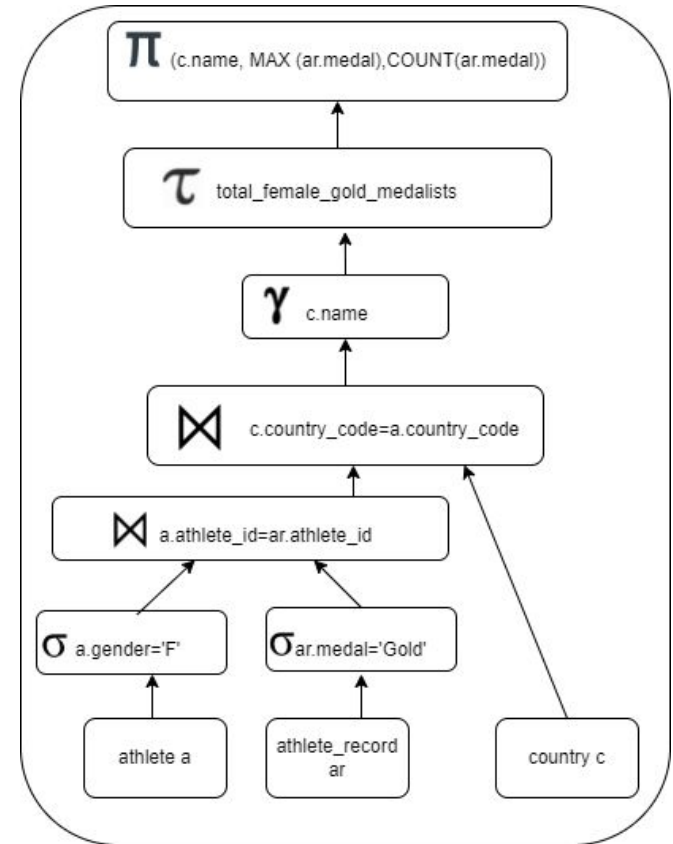
Hive T(Hive) = 16.11 s

# Queries - Inner Join(Oracle explain plan and proposed query plan)

SQL 0.076 seconds

OPERATION	OBJECT_NAME	OPTIONS	CARDINALITY	COST
SELECT STATEMENT			219	588
SORT		ORDER BY	219	588
HASH		GROUP BY	219	588
HASH JOIN			13373	587
Access Predicates				
ITEM_1=A.COUNTRY_CODE				
VIEW	SYS.VW_GBF_10		219	3
TABLE ACCESS	COUNTRY	FULL	219	3
HASH JOIN			13373	584
Access Predicates				
A.ATHLETE_ID=AR.ATHLETE_ID				
NESTED LOOPS			13373	584
NESTED LOOPS				
STATISTICS COLL				
TABLE ACCESS	ATHLETE_RECORD	FULL	13373	310
Filter Predicates				
AR.MEDAL='Gold'				
INDEX	SYS_C0012867	UNIQUE SCAN		
Access Predicates				
A.ATHLETE_ID=AR.ATHLETE_ID				
TABLE ACCESS	ATHLETE	BY INDEX ROWID	1	273
Filter Predicates				
A.GENDER='F'				
TABLE ACCESS	ATHLETE	FULL	67366	273
Filter Predicates				
A.GENDER='F'				

Oracle explain plan

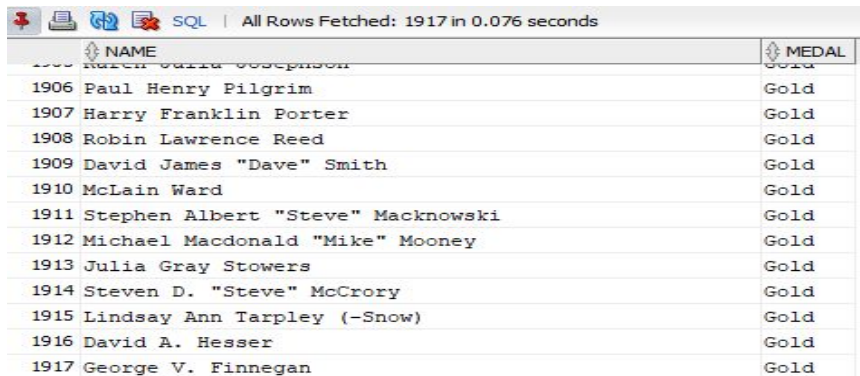




# Queries - Equi join (Oracle and Hive)

List all athletes from USA who won gold

*select distinct a.name, ar.medal from athlete a, athlete\_record ar where a.athlete\_id = ar.athlete\_id and a.country\_code = 'USA' and ar.medal = 'Gold';*



	NAME	MEDAL
1906	Paul Henry Pilgrim	Gold
1907	Harry Franklin Porter	Gold
1908	Robin Lawrence Reed	Gold
1909	David James "Dave" Smith	Gold
1910	McLain Ward	Gold
1911	Stephen Albert "Steve" Macknowski	Gold
1912	Michael Macdonald "Mike" Mooney	Gold
1913	Julia Gray Stowers	Gold
1914	Steven D. "Steve" McCrory	Gold
1915	Lindsay Ann Tarpley (-Snow)	Gold
1916	David A. Hesser	Gold
1917	George V. Finnegan	Gold

Oracle SQL Developer T(oracle) = 0.076 s



Aaron Wells Peirsol	Gold
Abbey Weitzeil	Gold
Abel Richard Kiviat	Gold
Adam McCright Nelson	Gold
Adelaide T. Lambert (-Ballard)	Gold
Adolph Gustav Kiefer	Gold
Adrian Delano Dantley	Gold
Adrian Howard Smith	Gold
Aeriwentha Mae Faggs (-Starr)	Gold
Aileen Muriel Riggin (-Soule)	Gold
Alan Boone Helffrich	Gold
Alan C. Morgan	Gold
Alan Chester Valentine	Gold
Alan Frank Williams	Gold
Alan Morgan Al Washbond	Gold
Alana Olga Blahoski	Gold

Time taken: 16.575 seconds, Fetched: 1917 row(s)

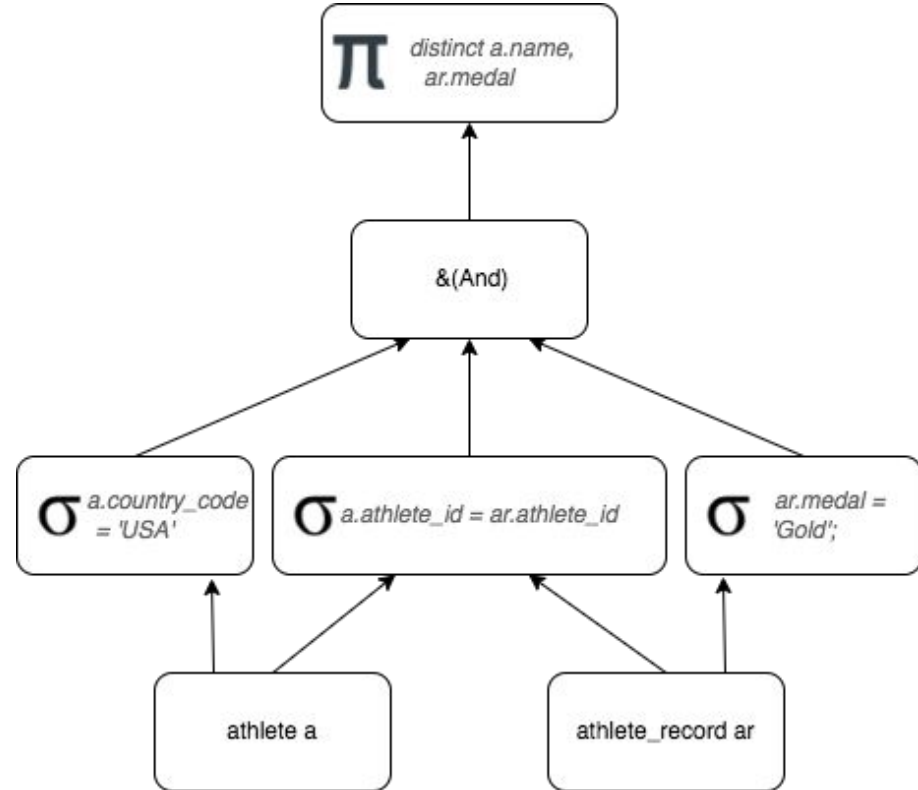
Hive T(Hive) = 16.575 s

# Queries - Equi Join(Oracle explain plan and proposed query plan)

SQL | 0.046 seconds

OPERATION	OBJECT_NAME	OPTIONS	CARDINALITY	COST
SELECT STATEMENT			6789	666
HASH		UNIQUE	6789	666
HASH JOIN		RIGHT SEMI	9608	584
Access Predicates				
A.ATHLETE_ID=AR.ATHLETE_ID				
TABLE ACCESS	ATHLETE_RECORD	FULL	13373	310
Filter Predicates				
AR.MEDAL='Gold'				
TABLE ACCESS	ATHLETE	FULL	9608	274
Filter Predicates				
A.COUNTRY_CODE='USA'				

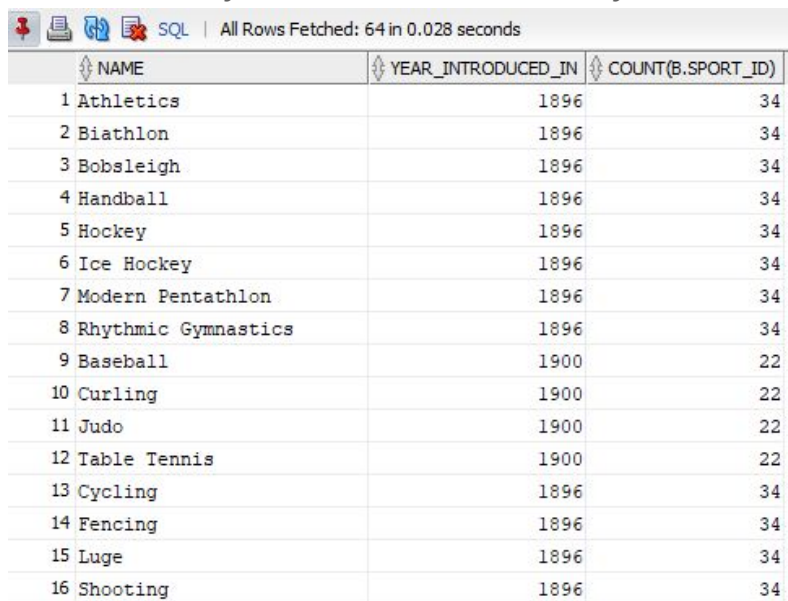
Oracle explain plan



# Queries - Self join (Oracle and Hive)

Count of sports which were played on the same year

```
SELECT A.name , A.year_introduced_in, count(B.sport_id) FROM sport A, sport B WHERE A.sport_id <> B.sport_id AND A.year_introduced_in = B.year_introduced_in group BY A.year_introduced_in, A.name;
```



SQL | All Rows Fetched: 64 in 0.028 seconds

NAME	YEAR INTRODUCED_IN	COUNT(B.SPORT_ID)
1 Athletics	1896	34
2 Biathlon	1896	34
3 Bobsleigh	1896	34
4 Handball	1896	34
5 Hockey	1896	34
6 Ice Hockey	1896	34
7 Modern Pentathlon	1896	34
8 Rhythmic Gymnastics	1896	34
9 Baseball	1900	22
10 Curling	1900	22
11 Judo	1900	22
12 Table Tennis	1900	22
13 Cycling	1896	34
14 Fencing	1896	34
15 Luge	1896	34
16 Shooting	1896	34

Oracle SQL Developer T(oracle) = 0.028 s



Alpine Skiing	1896	34
Alpinism	1904	1
Archery	1900	22
Art Competitions	1896	34
Athletics	1896	34
Badminton	1900	22
Baseball	1900	22
Basketball	1896	34
Beach Volleyball	1900	22
Biathlon	1896	34
Bobsleigh	1896	34
Boxing	1896	34
Canoeing	1896	34
Cricket	1924	1

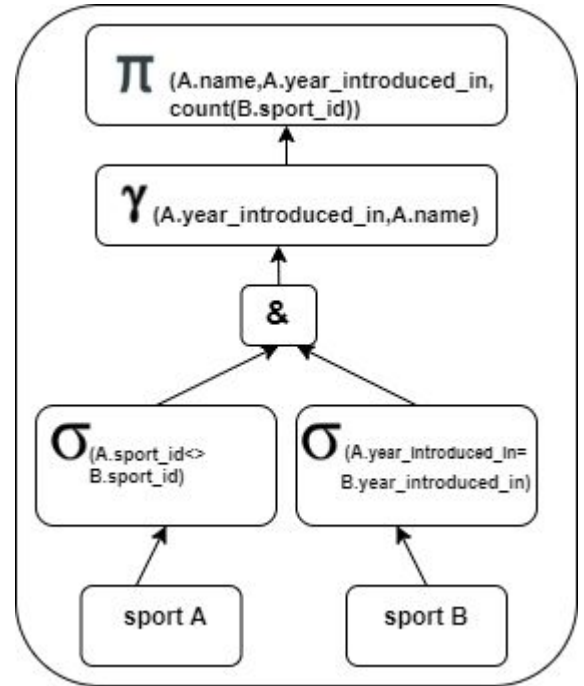
Time taken: 19.572 seconds, Fetched: 64 row(s)

Hive T(Hive) = 19.572 s

# Queries - Self Join(Oracle explain plan and proposed query plan)

OPERATION	OBJECT_NAME	OPTIONS	CARDINALITY	COST
SELECT STATEMENT				5
HASH				5
HASH JOIN		GROUP BY		4
Access Predicates				
A.YEAR INTRODUCED_IN=B.YEAR INTRODUCED_IN				
Filter Predicates				
A.SPORT_ID<>B.SPORT_ID				
TABLE ACCESS	SPORT	FULL		2
TABLE ACCESS	SPORT	FULL		2
Other XML				
{info}				
info type="has_user_tab"				
yes				
info type="db_version"				
19.0.0.0				
info type="parse_schema"				
"SYSTEM"				
info type="plan_hash_full"				
249407563				
info type="plan_hash"				

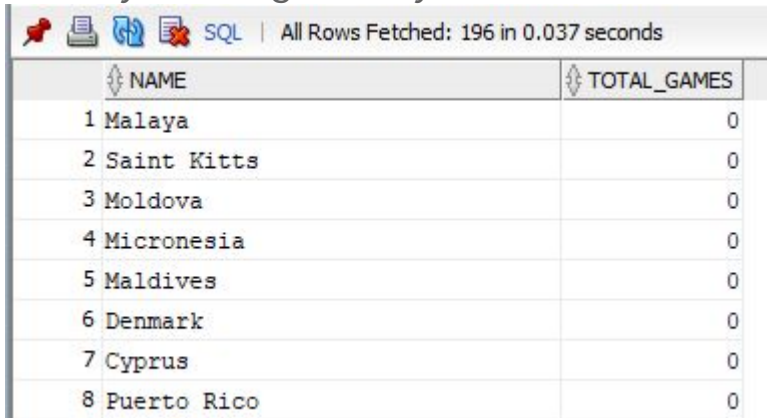
Oracle explain plan



# Queries - Left Join(Oracle and Hive)

List out all the countries where no games have been held

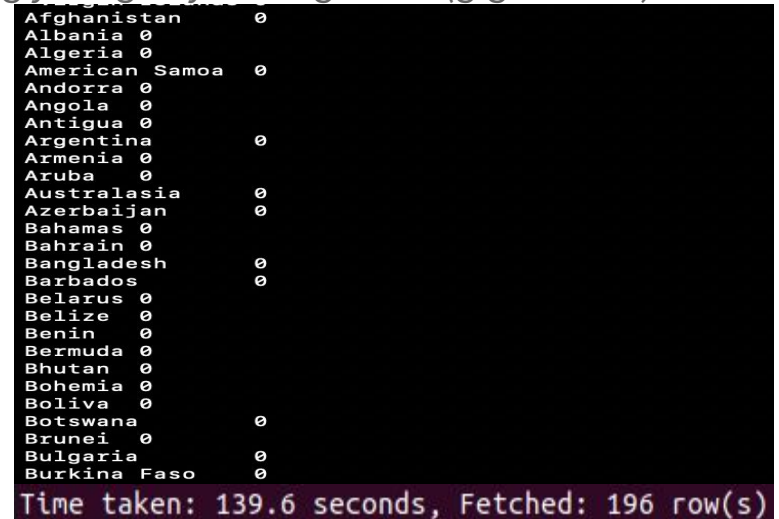
```
SELECT c.name,count(g.game_id) as total_games FROM country c LEFT JOIN game g ON  
c.country_code=g.country_code GROUP BY c.name,g.year,g.city having count(g.game_id) = 0;
```



Oracle SQL Developer interface showing a query result table. The table has two columns: NAME and TOTAL\_GAMES. The results are as follows:

	NAME	TOTAL_GAMES
1	Malaya	0
2	Saint Kitts	0
3	Moldova	0
4	Micronesia	0
5	Maldives	0
6	Denmark	0
7	Cyprus	0
8	Puerto Rico	0

Oracle SQL Developer T(Oracle) = 0.037sec



Hive command line interface showing a list of countries and their total games. The results are as follows:

COUNTRY	TOTAL_GAMES
Afghanistan	0
Albania	0
Algeria	0
American Samoa	0
Andorra	0
Angola	0
Antigua	0
Argentina	0
Armenia	0
Aruba	0
Australasia	0
Azerbaijan	0
Bahamas	0
Bahrain	0
Bangladesh	0
Barbados	0
Belarus	0
Belize	0
Benin	0
Bermuda	0
Bhutan	0
Bohemia	0
Boliva	0
Botswana	0
Brunei	0
Bulgaria	0
Burkina Faso	0

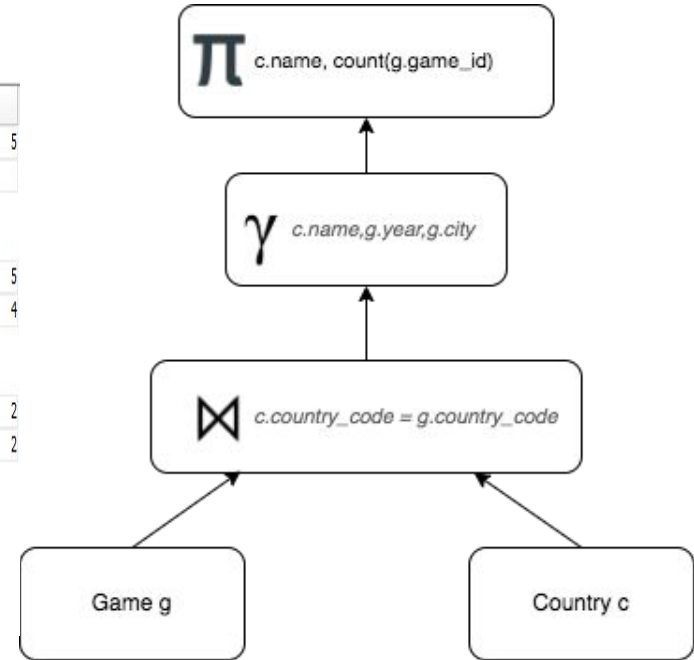
Time taken: 139.6 seconds, Fetched: 196 row(s)

Hive T(Hive) = 139.6 sec

# Queries - Left Join(Oracle explain plan and proposed query plan)

OPERATION	OBJECT_NAME	OPTIONS	CARDINALITY	COST
SELECT STATEMENT				5
FILTER				
Filter Predicates COUNT(G.GAME_ID)=0				
HASH		GROUP BY	3	5
HASH JOIN		OUTER	248	4
Access Predicates C.COUNTRY_CODE=G.COUNTRY_CODE(+)				
TABLE ACCESS	COUNTRY	FULL	219	2
TABLE ACCESS	GAME	FULL	52	2
Other XML				
{info}				
info type="has_user_tab"				
yes				
info type="db_version"				

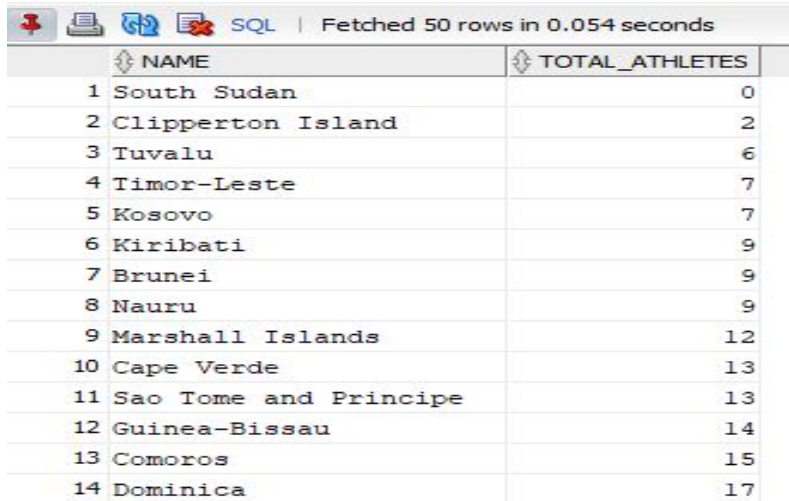
Oracle explain plan



# Queries -Right Join (Oracle and Hive)

List down all the countries and total no of athletes each country has

*select c.name, count(a.athlete\_id) as total\_athletes from athlete a right join country c on c.country\_code = a.country\_code group by c.name order by total\_athletes;*



SQL | Fetched 50 rows in 0.054 seconds

	NAME	TOTAL_ATHLETES
1	South Sudan	0
2	Clipperton Island	2
3	Tuvalu	6
4	Timor-Leste	7
5	Kosovo	7
6	Kiribati	9
7	Brunei	9
8	Nauru	9
9	Marshall Islands	12
10	Cape Verde	13
11	Sao Tome and Principe	13
12	Guinea-Bissau	14
13	Comoros	15
14	Dominica	17

Oracle SQL Developer T(Oracle) = 0.043 sec for 219 total records



South Sudan	0
Clipperton Island	2
Tuvalu	6
Timor-Leste	7
Kosovo	7
Nauru	9
Brunei	9
Kiribati	9
Marshall Islands	12
Cape Verde	13
Sao Tome and Principe	13
Guinea-Bissau	14
Comoros	15
Dominica	17

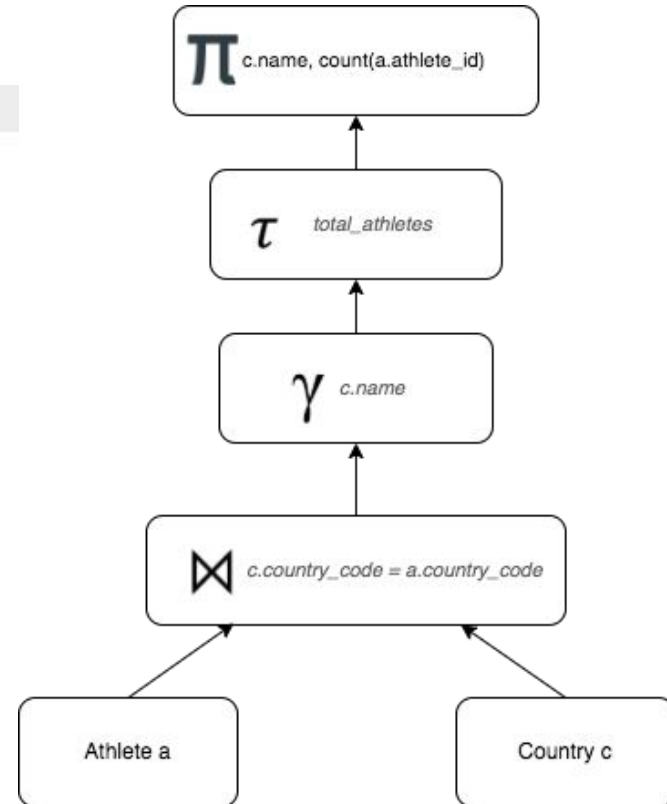
Time taken: 15.104 seconds, Fetched: 218 row(s)

Hive T(Hive) = 15.104

# Queries - Right Join (Oracle explain plan and proposed query plan)

OPERATION	OBJECT_NAME	OPTIONS	CARDINALITY	COST
SELECT STATEMENT				219 267
SORT		ORDER BY		219 267
HASH		GROUP BY		219 267
HASH JOIN		OUTER		219 265
Access Predicates				
C.COUNTRY_CODE=ITEM_1(+)				
TABLE ACCESS	COUNTRY	FULL		219 2
VIEW	VW_GRC_5			218 263
HASH		GROUP BY		218 263
TABLE ACCESS	ATHLETE	FULL	134729	254

Oracle explain plan





# Queries - Theta join (Oracle and Hive)

**Get athletes and athletes who are younger than them. Limit the result to 10000 rows.**

*select \* from(select a.name as older\_athlete, a.birthdate as dob1, b.name as younger\_athlete, b.birthdate as dob2 from athlete a inner join athlete b on a.athlete\_id <> b.athlete\_id and a.birthdate < b.birthdate) where rownum < 10000;*



The screenshot shows the Oracle SQL Developer interface. At the top, a status bar indicates 'All Rows Fetched: 9999 in 0.431 seconds'. Below this is a table with 4 columns: OLDER\_ATHLETE, DOB1, YOUNGER\_ATHLETE, and DOB2. The table contains 10 rows of data, each representing a pair of athletes where the first is older than the second. The names of the athletes are truncated in the YOUNGER\_ATHLETE column.

	OLDER_ATHLETE	DOB1	YOUNGER_ATHLETE	DOB2
9987	Friedrich Lthi	23-08-1851	Mikul Krn	08-12-1947
9988	Friedrich Lthi	23-08-1851	Kenji Soda	23-04-1948
9989	Friedrich Lthi	23-08-1851	Tatsuo Jihira	17-10-1948
9990	Friedrich Lthi	23-08-1851	Ljubodrag Simonovi	03-12-1949
9991	Friedrich Lthi	23-08-1851	Maria Canins	22-01-1949
9992	Friedrich Lthi	23-08-1851	Ole Benediktson	03-09-1949
9993	Friedrich Lthi	23-08-1851	Tsutomu Hanzawa	12-02-1949
9994	Friedrich Lthi	23-08-1851	Alberto Blanco Fernandez	26-05-1950
9995	Friedrich Lthi	23-08-1851	Franciszek Niemiec	11-05-1950
9996	Friedrich Lthi	23-08-1851	Anna Boena Wodarczyk	15-01-1951
9997	Friedrich Lthi	23-08-1851	Kaliq Singh	17-07-1951
9998	Friedrich Lthi	23-08-1851	Oga Koziov (-Gubzov)	03-06-1951
9999	Friedrich Lthi	23-08-1851	Ricardo Daniel Ibarra	10-01-1951

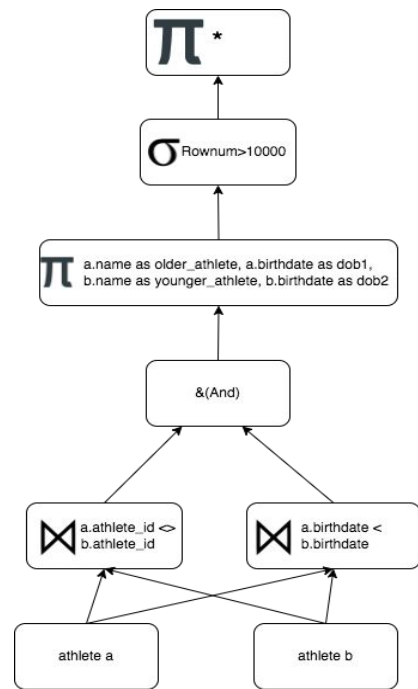
Hive doesn't support

**Oracle SQL Developer T(Oracle) = 0.431  
sec**

# Queries - Theta join (Oracle explain plan and proposed query plan)

OPERATION	OBJECT_NAME	OPTIONS	CARDINALITY	COST
SELECT STATEMENT				1101 46
COUNT		STOPKEY		
Filter Predicates				
ROWNUM<10000				
NESTED LOOPS				1101 46
TABLE ACCESS	ATHLETE	FULL	22017	2
TABLE ACCESS	ATHLETE	FULL	1101	44
Filter Predicates				
AND				
A.BIRTHDATE<B.BIRTHDATE				
A.ATHLETE_ID<>B.ATHLETE_ID				

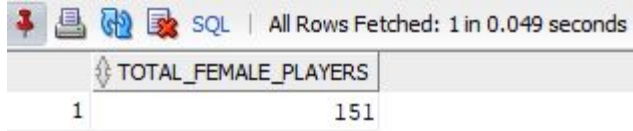
Oracle explain plan



# Queries -Subquery (Oracle and Hive)

**Total number of female athletes who play for India.**

*select count(athlete\_id) as total\_female\_players from Athlete where gender='F' and country\_code in(Select country\_code from country where name='India');*



The screenshot shows the Oracle SQL Developer interface. At the top, there is a toolbar with icons for pinning, printing, refreshing, and saving, followed by the text "SQL | All Rows Fetched: 1 in 0.049 seconds". Below this, a table with the column header "TOTAL\_FEMALE\_PLAYERS" is displayed. The table contains one row with the value "151".

TOTAL_FEMALE_PLAYERS
151

151

Time taken: 13.951 seconds, Fetched: 1 row(s)

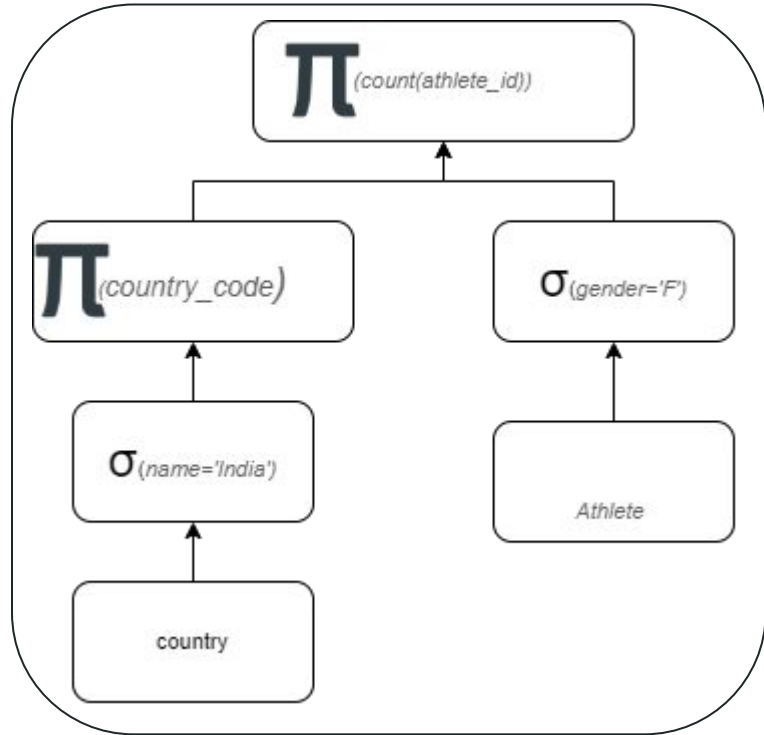
Oracle SQL Developer T(oracle) = 0.049 sec

Hive T(oracle) = 13.951 sec

# Queries - Subquery (Oracle explain plan and proposed query plan)

OPERATION	OBJECT_NAME	OPTIONS	CARDINALITY	COST
SELECT STATEMENT			1	256
SORT		AGGREGATE	1	
HASH JOIN			155	256
Access Predicates				
COUNTRY_CODE=COUNTRY_CODE				
TABLE ACCESS	COUNTRY	FULL	1	2
Filter Predicates				
NAME='India'				
TABLE ACCESS	ATHLETE	FULL	33778	254
Filter Predicates				
GENDER='F'				

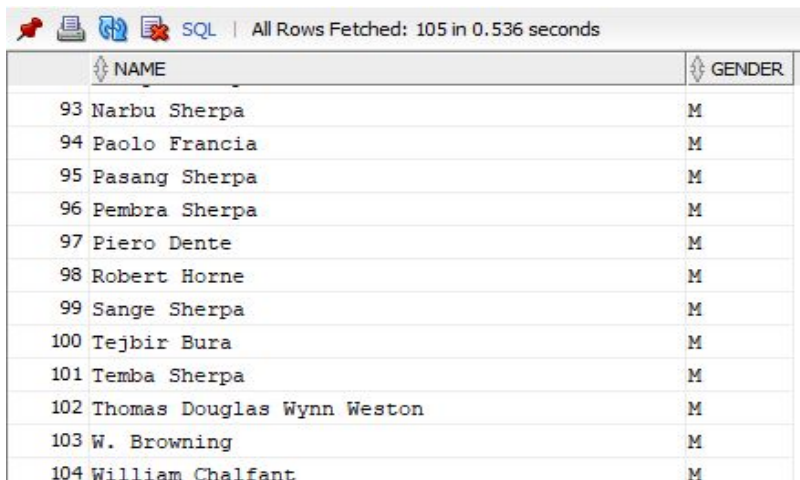
Oracle explain plan



# Queries - Nested Query (Oracle and Hive)

**Name and gender of players who play sports introduced after 1900.**

*select a.name, a.gender from athlete a where a.athlete\_id in (select ar.athlete\_id from Athlete\_record ar where ar.sport\_id in (Select sport\_id from sport where YEAR INTRODUCED\_IN > 1900));*



NAME	GENDER
93 Narbu Sherpa	M
94 Paolo Francia	M
95 Pasang Sherpa	M
96 Pembra Sherpa	M
97 Piero Dente	M
98 Robert Horne	M
99 Sange Sherpa	M
100 Tejbir Bura	M
101 Temba Sherpa	M
102 Thomas Douglas Wynn Weston	M
103 W. Browning	M
104 William Chalfant	M

Oracle SQL Developer T(Oracle) = 0.536 sec



Francisco Villota y Baquiola	M
Philip Humphreys Tomalin	M
Tom George Longstaff	M
Eustace Hamilton Miles	M
Josef Bm	M
Alfred James Bowerman	M
Edwin Anthony Biedermann (-Best)	M
Paul Camille Albert Mandrillon	M
Isaac Thomas Thornycroft	M
Colin Grant Crawford	M
Alfred George Fentiman	M
"Harriet Pauline "Hettie" Dyhrenfurth (Heyman-)"	F
Gnter Oskar Dyhrenfurth	M
"Anton "Toni" Schmid"	M
Vane Hungerford Pennell	M
Thomas Evelyn Ellis (-Scott-Ellis)	M
Karel Buchta	M
Alfons Julien	M
Edward Felix Norton	M

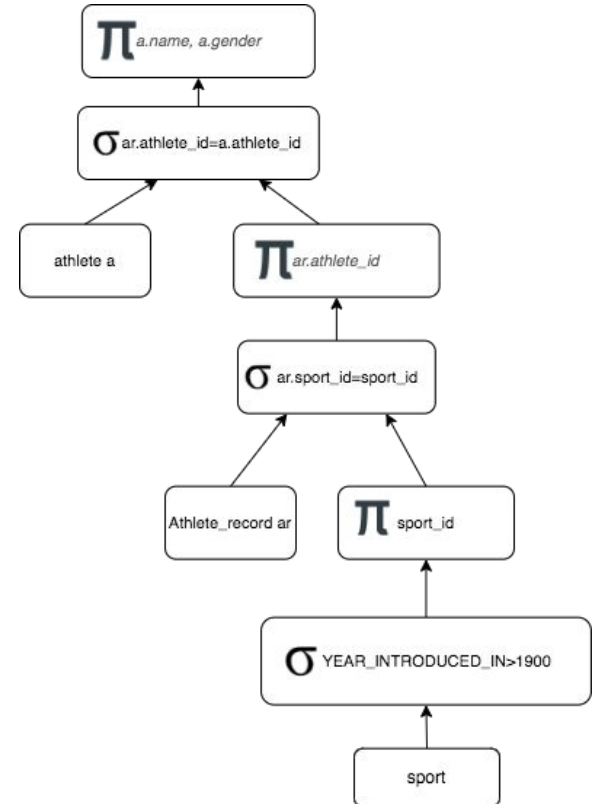
Time taken: 28.217 seconds, Fetched: 105 row(s)

Hive T(Hive) = 28.217 sec

# Queries - Nested Query (Oracle explain plan and proposed query plan)

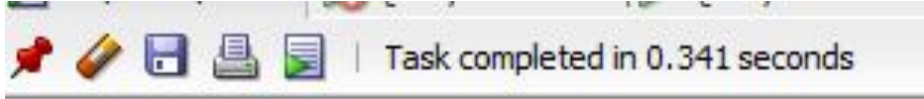
OPERATION	OBJECT_NAME	OPTIONS	CARDINALITY	COST
SELECT STATEMENT				34916 549
HASH JOIN		RIGHT SEMI	34916	549
Access Predicates				
A.ATHLETE_ID=ATHLETE_ID				
VIEW	VW_NSQ_1		34916	294
HASH JOIN			34916	294
Access Predicates				
AR.SPORT_ID=SPORT_ID				
TABLE ACCESS	SPORT	FULL	9	2
Filter Predicates				
YEAR INTRODUCED_IN>1900				
TABLE ACCESS	ATHLETE_RECORD	FULL	271114	290
TABLE ACCESS	ATHLETE	FULL	134729	253

Oracle explain plan



# Queries - Index (Oracle and Hive)

*create index ar\_athlete\_id on athlete\_record ( athlete\_id );*



Index AR\_ATHLETE\_ID created.

Hive doesn't support

Oracle SQL Developer T(Oracle) = 0.341  
sec

# Queries - Correlated subquery using Exists (Oracle and Hive)

List of all the athletes who have won more than 2 gold medals.

*SELECT a.athlete\_id, a.name FROM athlete a WHERE EXISTS (SELECT 1 FROM athlete\_record ar WHERE ar.athlete\_id = a.athlete\_id and ar.medal='Gold' GROUP BY ar.athlete\_id HAVING COUNT (1) > 2);*



	ATHLETE_ID	NAME
559	48035	Hendrika Wilhelmina "Rie" Mastenbroek (-Kuiper...
560	20579	Cherie Piper
561	95067	Norman DeMille Ross
562	2164	Akinori Nakayama
563	105071	Regla Radameris Torres Herrera
564	113050	Seimone Delicia Augustus
565	99400	Paul Vincent Costello
566	10225	Anthony Lee Ervin
567	123161	Torsten Ren Gutsche

Oracle SQL Developer T(Oracle) = 1.217 sec



```
63975 "Joseph "Joe" Jackson"
15001 Bernhard Germeshausen
74753 Lenny Krayzelburg
96251 Olga Valentinovna Korbut (-Bortkevich -Voynich)
118611 Sven Fischer
9169 Anke Nothnagel-von Seck
39172 Gabriella Timea Szab
77848 Lucien Alphonse Paul Gaudin
100058 Pedro Pablo Morales Jr.
72040 Kristin Otto
86520 "Melissa Jeanette "Missy" Franklin"
121907 Tibor Berczelly (Berczeller-)
57402 Jean-Claude Killy
127294 Vladimir Valeryevich Salnikov
39177 "Gabrielle Christina Victoria "Gabby" Douglas"
24912 Danuta Kozk
Time taken: 15.504 seconds, Fetched: 571 row(s)
```

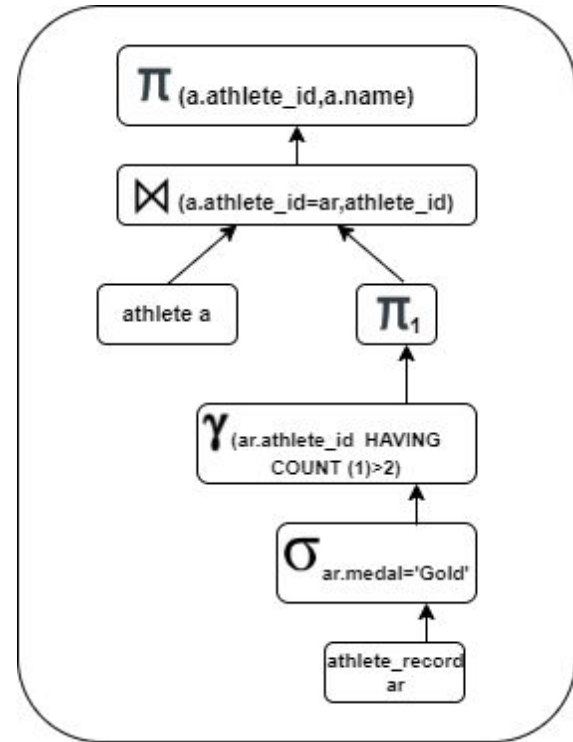
Hive T(Hive) = 15.504 sec



# Queries - Correlated subquery using Exists (Oracle explain plan and proposed query plan)

OPERATION	OBJECT_NAME	OPTIONS	CARDINALITY	COST
SELECT STATEMENT			1	495756
FILTER				
Filter Predicates				
EXISTS (SELECT 0 FROM ATHLETE_RECORD AR WHERE AR.ATHLETE_ID=:B1 AND AR.MEDAL='Gold' GROUP BY AR.ATHLETE_ID)				
TABLE ACCESS	ATHLETE	FULL	134729	253
SORT		GROUP BY NOSORT	1	4
Filter Predicates				
COUNT(*)>2				
TABLE ACCESS	ATHLETE_RECORD	BY INDEX ROWID	1	4
Filter Predicates				
AR.MEDAL='Gold'				
INDEX	AR_ATHLETE_ID	RANGE SCAN	2	3
Access Predicates				
AR.ATHLETE_ID=:B1				

Oracle explain plan



# Queries - View (Oracle and Hive)

## Create a view to access athlete information

*CREATE OR REPLACE VIEW athlete\_bio AS SELECT athlete\_id, name, height, weight FROM Athlete;*

*Select \* from athlete\_bio;*

Task completed in 0.75 seconds

Displays the execution plan only.

View ATHLETE\_BIO created.

	ATHLETE_ID	NAME	HEIGHT	WEIGHT
1	38608	Friedrich Lthi	(null)	(null)
2	17846	Carl Emil "Carl-Emil" Johansson	(null)	(null)
3	95575	Olaf Emil Frydenlund	(null)	(null)
4	1130	Adolf Karl Oskar Davids	(null)	(null)
5	54439	Jacques Baudrier	(null)	(null)
6	17913	Carl Johan Eldh	(null)	(null)
7	61647	John Richardson	(null)	(null)
8	63629	Josef (Jaroslav) Suk	(null)	(null)
9	82529	Marie-Charles-Andr Corvington	(null)	(null)
10	63975	Joseph "Joe" Jackson	183	(null)
11	129987	William Morton	(null)	(null)
12	61649	John Riegel DeWitt	(null)	(null)
13	116783	Stanley Howard Shoveller	(null)	(null)
14	130032	William Philo	(null)	(null)

Oracle SQL Developer T(oracle) = 0.75 sec

38608	Friedrich Lthi	NULL	NULL		
17846	"Carl Emil "Carl-Emil" Johansson"	NULL	NULL	NULL	NULL
95575	Olaf Emil Frydenlund	NULL	NULL		
1130	Adolf Karl Oskar Davids	NULL	NULL		
54439	Jacques Baudrier	NULL	NULL		
17913	Carl Johan Eldh	NULL	NULL		
61647	John Richardson	NULL	NULL		
63629	Josef (Jaroslav) Suk	NULL	NULL		
82529	Marie-Charles-Andr Corvington	NULL	NULL	NULL	NULL
63975	"Joseph "Joe" Jackson"	183	NULL		
129987	William Morton	NULL	NULL		
61649	John Riegel DeWitt	NULL	NULL		
116783	Stanley Howard Shoveller	NULL	NULL	NULL	NULL
130032	William Philo	NULL	NULL		
Time taken: 0.453 seconds, Fetched: 134731 row(s)					

Hive T(Hive) = 0.453 sec

# Queries - View (Oracle explain plan)

```
-----  
Select * from athlete_bio
```

```
Plan hash value: 81540165
```

```
-----  
| Id | Operation          | Name    | E-Rows |  
-----  
|  0 | SELECT STATEMENT   |         |        |  
|  1 | TABLE ACCESS FULL | ATHLETE | 134K   |  
-----
```

OPERATION	OBJECT_NAME	OPTIONS	CARDINALITY	COST
SELECT STATEMENT			134729	254
TABLE ACCESS	ATHLETE	FULL	134729	254

Oracle explain plan

# Queries - Trigger on View (Oracle and Hive)

## Trigger to update table when view is updated

```
CREATE OR REPLACE TRIGGER update_athlete_bio_trigger INSTEAD OF UPDATE ON athlete_bio
BEGIN UPDATE athlete SET height = :NEW.height, weight =:NEW.weight WHERE athlete_id =
:OLD.athlete_id; END;
```

---


Trigger UPDATE\_ATHLETE\_BIO\_TRIGGER compiled

Hive doesn't support

Oracle SQL Developer T(oracle) = 0.75 sec

# Queries - Update trigger on View (Oracle and Hive)


*Select \* from athlete where athlete\_id= 766;*

 All Rows Fetched: 1 in 0.001 seconds


ATHLETE_ID	NAME	GENDER	HEIGHT	WEIGHT	COUNTRY_CODE	BIRTHDATE	TOTAL_MEDALS
1	766 Achraf Kharroubi M		160	52 MAR		10-AUG-91 00:00:00	0

```
hive> Update athlete_bio set weight=58 where athlete_id= 766;  
FAILED: SemanticException [Error 10294]: Attempt to do update or delete using transaction manager that does not support these operations.
```

*Update athlete\_bio set weight=58 where athlete\_id= 766;*

 Task completed in 0.027 seconds

1 row updated.

 All Rows Fetched: 1 in 0.008 seconds

ATHLETE_ID	NAME	GENDER	HEIGHT	WEIGHT	COUNTRY_CODE	BIRTHDATE	TOTAL_MEDALS
1	766 Achraf Kharroubi M		160	58 MAR		10-AUG-91 00:00:00	0

Hive doesn't support

Oracle SQL Developer T(Oracle) = 0.027

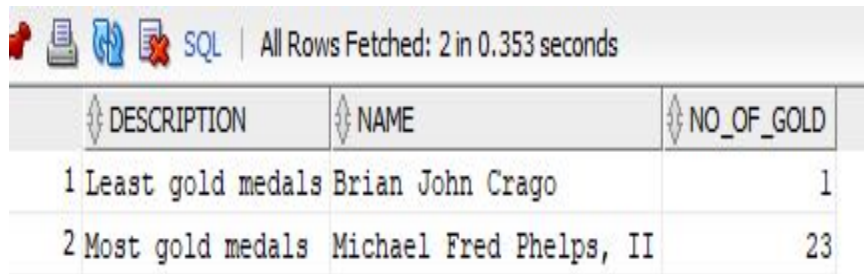
Hive

# Queries - Union (Oracle and Hive)

**List down the athlete with the most and the least number of gold medals won.**

*select \* from (select 'Most gold medals' as description,a.name ,count(ar.record\_id)as no\_of\_gold from athlete\_record ar inner join athlete a on a.athlete\_id = ar.athlete\_id where ar.medal = 'Gold' group by a.name order by no\_of\_gold desc) where rownum=1 **UNION***

*select \* from (select 'Least gold medals' as description,a.name, count(ar.record\_id)as no\_of\_gold from athlete\_record ar inner join athlete a on a.athlete\_id = ar.athlete\_id where ar.medal = 'Gold' group by a.name order by no\_of\_gold asc) where rownum=1;*



	DESCRIPTION	NAME	NO_OF_GOLD
1	Least gold medals	Brian John Crago	1
2	Most gold medals	Michael Fred Phelps, II	23

Oracle SQL Developer T(Oracle) = 0.35 sec



```
Total MapReduce CPU Time Spent: 37 seconds 440 msec
OK
Least gold medals      Brian John Crago      1
Most gold medals      Michael Fred Phelps  23
Time taken: 306.364 seconds, Fetched: 2 row(s)
hive>
```

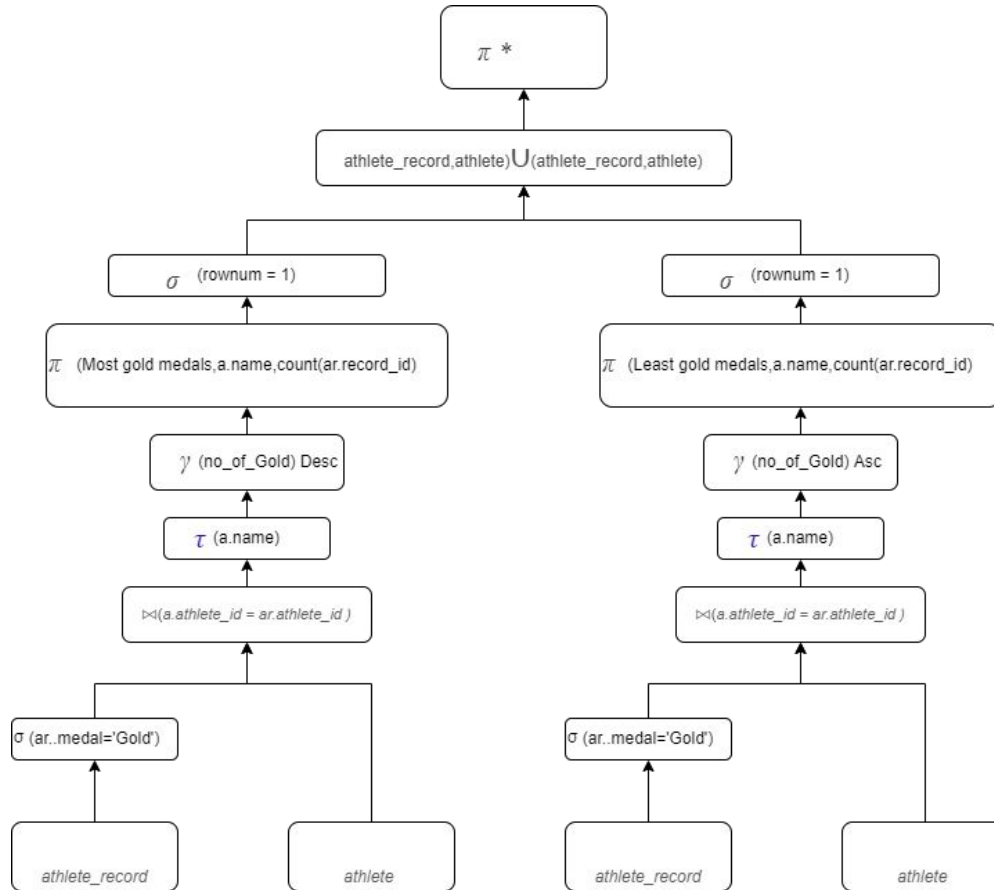
Hive T(Hive) = 306.364 sec

## Queries - Union (Oracle explain plan)

	SELECT STATEMENT			2	1696
	SORT			2	1696
	UNION-ALL				
	COUNT				
	Filter Predicates ROWNUM=1				
	VIEW			13372	548
	SORT			13372	548
	Filter Predicates ROWNUM=1				
	HASH			13372	548
	HASH JOIN			13372	546
	Access Predicates ITEM_1=AR.ATHLETE_ID				
	TABLE ACCESS ATHLETE_RECORD	FULL		13372	292
	Filter Predicates AR.MEDAL='Gold'				
	VIEW VW_GBF_8			134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS ATHLETE	FULL		134729	253
	TABLE ACCESS				

## Oracle explain plan

# Queries - Union (Proposed plan)





# System Configuration

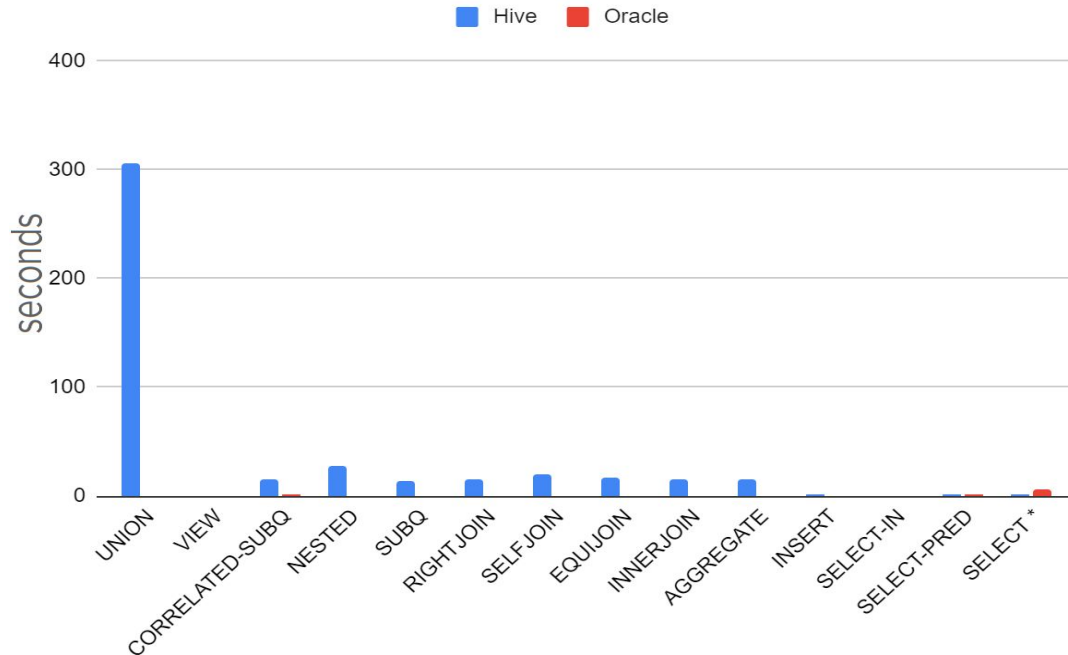
## Oracle Configuration

- Version - Oracle DB 20.4.0.379
- 2 GB RAM
- 1 core CPU

## Hive Configuration

- Version - 3.1.2
- 4 GB RAM
- 3 core CPU

# Comparison Graph



Avg Execution Time:

Oracle - 0.79 seconds\*

Hive - 11.23 seconds\*

\*excludes UNION  
(outlayer)

# Challenges

## **Oracle:**

- Creation of the proposed query plan and comparison with oracle query plan.
- More amount of time taken by oracle to load the data.

## **Hive:**

- Hadoop and Hive installations.
- Loading data into Hive tables.
- Workaround to delete data from hive table.

# Conclusion

1. Queries ran more efficiently on Oracle SQL as compared to Hive.
2. Some queries like “select \* from” ran faster on hive.
3. Join, aggregate, nested query are some of the examples that took much more time in hive.
4. List of operations not supported by Hive:
  - Update
  - Delete
  - Trigger
  - Rownum
  - Update on View
  - Index
  - Theta Join

Thank You!

---