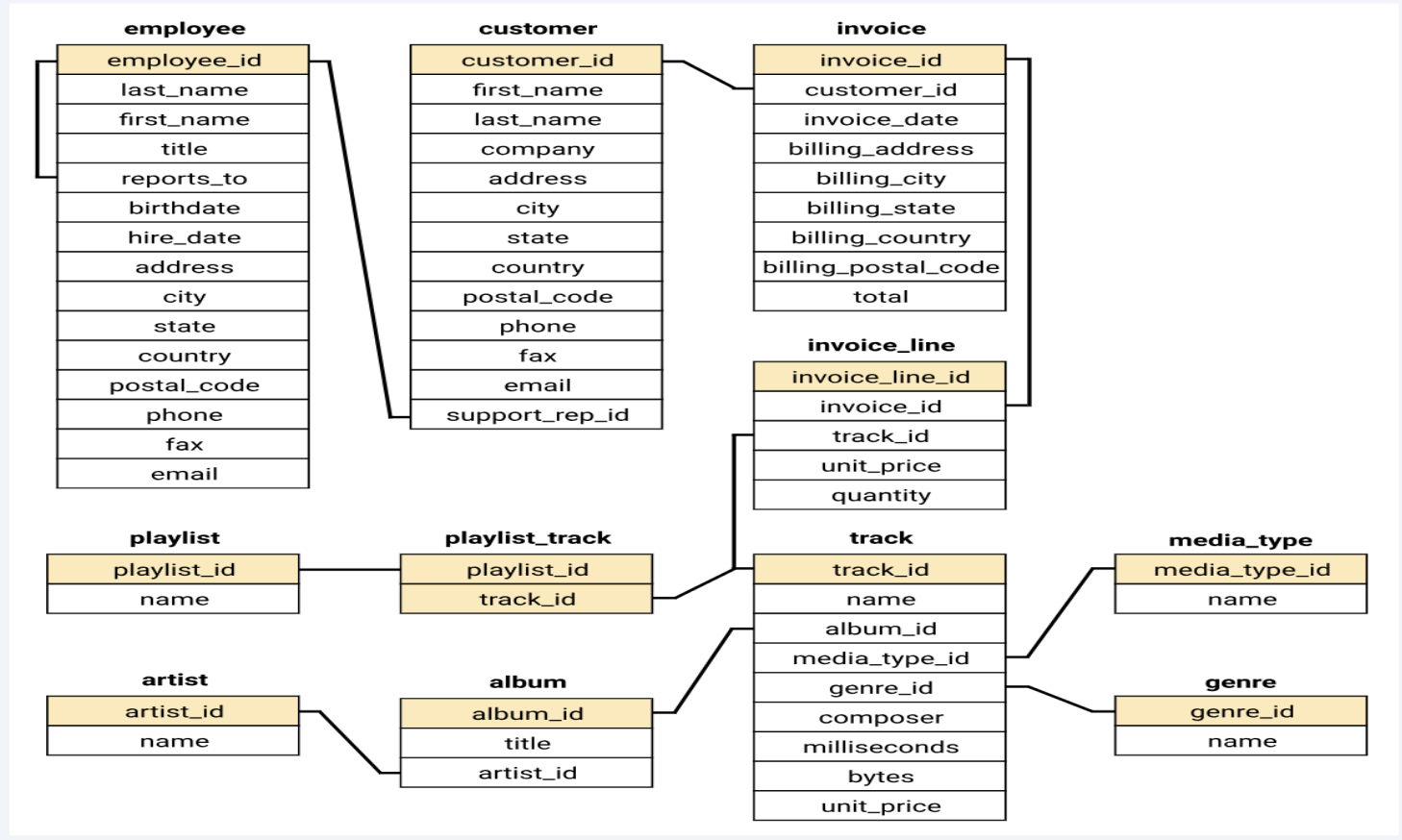
MUSIC STORE SQL PROJECT REPORT

By Gaurav Chandra

The tables are connected in this following manner



--Q1 A: Who is the senior most employee based on age ?

```
select first_name,last_name,title,(extract(year from current_date)-extract(year from birthdate)) as Age
 from employee order by Age desc limit 1;
```

	first_name character	last_name character	title character varying (50)	age numeric
1	Margaret	Park	Sales Support Agent	76

--Q1 B: Who is the senior most employee based on jobtitle ?

```
select first_name,last_name,title,levels
 from employee order by levels desc limit 1;
```

	first_name character	last_name character	title character varying (50)	levels character varying (10)
1	Mohan	Madan	Senior General Manager	L7

--Q2 : Which country has the most invoices ?

```
select billing_country,count(invoice_id) from invoice
group by billing_country
order by 2 desc limit 1;
```

	billing_country character varying (30)	count bigint
1	USA	131

--Q3 : What are top 3 values of total invoice ?

```
select invoice_id,total from invoice
order by total desc limit 3;
```

	invoice_id [PK] integer	total double precision
1	183	23.759999999999998
2	92	19.8
3	31	19.8

--Q4. Which city has the best customers? We would like to throw a promotional Music Festival in the city we made the most money. Write a query that returns one city that has the highest sum of invoice totals. Return both the city name & sum of all invoice totals

```
select billing_city,sum(total) as Total_invoices from invoice
group by billing_city order by Total_invoices desc limit 1;
```

	billing_city character varying (30)	total_invoices double precision
1	Prague	273.24000000000007

--Q5. Who is the best customer? The customer who has spent the most money will be declared the best customer. Write a query that returns the person who has spent the most money

```
select c.customer_id,c.first_name,c.last_name,sum(i.total) as total
from customer c join invoice i
on c.customer_id = i.customer_id
group by c.customer_id
order by 4 desc limit 1;
```

	customer_id [PK] integer	first_name character	last_name character	total double precision
1	5	R ...	Madhav	144.54000000000002

--Q6. Write query to return the email, first name, last name, & Genre of all Rock Music

--listeners. Return your list ordered alphabetically by email starting with A

```
select distinct c.first_name,c.last_name,c.email from customer c
join invoice i on c.customer_id = i.customer_id
join invoice_line il on i.invoice_id = il.invoice_id
join track t on t.track_id = il.track_id
join genre g on g.genre_id = t.genre_id
where g.name like 'Rock'
order by email;
```

	first_name character	last_name character	email character varying (50)
1	Aaron	Mitchell	aaronmitchell@yahoo.ca
2	Alexandre	Rocha	alero@uol.com.br
3	Astrid	Gruber	astrid.gruber@apple.at
4	Bjørn	Hansen	bjorn.hansen@yahoo.no
5	Camille	Bernard	camille.bernard@yahoo.fr
6	Daan	Peeters	daan_peeters@apple.be
7	Diego	Gutiérrez	diego.gutierrez@yahoo.ar
8	Dan	Miller	dmiller@comcast.com
9	Dominique	Lefebvre	dominiquelefebvre@gmail.c...
10	Edward	Francis	edfrancis@yachoo.ca
11	Eduardo	Martins	eduardo@woodstock.com.br
12	Ellie	Sullivan	ellie.sullivan@shaw.ca
13	Emma	Jones	emma_jones@hotmail.com
14	Enrique	Muñoz	enrique_munoz@yahoo.es
15	Fernanda	Ramos	fernadaramos4@uol.com.br

Total rows: 59 of 59 Query complete 00:00:00.127

--#A better and optimised method for the same problem :

```
select email,first_name,last_name from customer
join invoice on customer.customer_id = invoice.customer_id
join invoice_line on invoice.invoice_id = invoice_line.invoice_id
where track_id in(
        select track_id from track
        join genre on track.genre_id = genre.genre_id
        where genre.name like 'Rock'
)
group by customer.customer_id
order by email;
```

	email character varying (50)	first_name character	last_name character
1	aaronmitchell@yahoo.ca	Aaron	Mitchell
2	alero@uol.com.br	Alexandre	Rocha
3	astrid.gruber@apple.at	Astrid	Gruber
4	bjorn.hansen@yahoo.no	Bjørn	Hansen
5	camille.bernard@yahoo.fr	Camille	Bernard
6	daan_peeters@apple.be	Daan	Peeters
7	diego.gutierrez@yahoo.ar	Diego	Gutiérrez
8	dmiller@comcast.com	Dan	Miller

--Q7. Let's invite the artists who have written the most rock music in our dataset. Write a query that returns the Artist name and total track count of the top 10 rock bands

```
select a.name,count(track_id) as number_of_songs from artist a
join album ab on ab.artist_id=a.artist_id
join track t on t.album_id = ab.album_id
join genre g on g.genre_id = t.genre_id
where g.name like 'Rock'
group by a.name
order by 2 desc limit 10;
```

	name character varying (120)	number_of_songs bigint
1	Led Zeppelin	114
2	U2	112
3	Deep Purple	92
4	Iron Maiden	81
5	Pearl Jam	54
6	Van Halen	52
7	Queen	45
8	The Rolling Stones	41

Total rows: 10 of 10 Query complete 00:00:00.198

--Q8. Return all the track names that have a song length longer than the average song length. Return the Name and Milliseconds for each track. Order by the song length with the longest songs listed first

```
select name ,milliseconds from track where milliseconds > (select avg(milliseconds) from track)
order by milliseconds desc
```

	name character varying (150)	milliseconds integer
1	Occupation / Precipice	5286953
2	Through a Looking Glass	5088838
3	Greetings from Earth, Pt. 1	2960293
4	The Man With Nine Lives	2956998
5	Battlestar Galactica, Pt. 2	2956081
6	Battlestar Galactica, Pt. 1	2952702
7	Murder On the Rising Star	2935894
8	Battlestar Galactica, Pt. 3	2927802

Total rows: 494 of 494 Query complete 00:00:00.225

--Q9. Find how much amount spent by each customer on artists? Write a query to return customer name, artist name and total spent

--I made a CTE so that we can get the customers who have spent on the top n best selling artist

```
with best_selling_artist as(
	select a.artist_id,a.name,sum(il.quantity * il.unit_price) as total_spent from artist a
	join album ab on ab.artist_id = a.artist_id
	join track t on t.album_id = ab.album_id
	join invoice_line il on il.track_id = t.track_id
	group by 1
	order by 3 desc
	limit 1
)
select c.customer_id,c.first_name,c.last_name,bsa.name,sum(il.quantity * il.unit_price) as amt_spent from customer c
join invoice i on c.customer_id = i.customer_id
join invoice_line il on il.invoice_id = i.invoice_id
join track t on t.track_id = il.track_id
join album ab on ab.album_id = t.album_id
join artist a on a.artist_id = ab.artist_id
join best_selling_artist bsa on bsa.artist_id = a.artist_id
group by 1,2,3,4
order by 5 desc
```

	customer_id integer	first_name character	last_name character	name character varying (120)	amt_spent double precision
1	46	Hugh	O'Reilly	Queen	27.719999999999985
2	38	Niklas	Schröder	Queen	18.81
3	3	François	Tremblay	Queen	17.82
4	34	João	Fernandes	Queen	16.830000000000002
5	53	Phil	Hughes	Queen	11.88
6	41	Marc	Dubois	Queen	11.88
7	47	Lucas	Mancini	Queen	10.89
8	33	Ellie	Sullivan	Queen	10.89

Total rows: 43 of 43 Query complete 00:00:00.106

--Q10. We want to find out the most popular music Genre for each country. We determine the

--most popular genre as the genre with the highest amount of purchases. Write a query

--that returns each country along with the top Genre. For countries where the maximum

--number of purchases is shared return all Genres

```
with popular_genre as (
        select c.country ,g.name,count(il.quantity) as purchase ,
        dense_rank() over (partition by c.country order by count(il.quantity) desc) as rank
        from customer c
        join invoice i on i.customer_id = c.customer_id
        join invoice_line il on i.invoice_id = il.invoice_id
        join track t on t.track_id = il.track_id
        join genre g on g.genre_id = t.genre_id
        group by 1,2
        order by 3 desc,1 asc
)
select * from popular_genre where rank = 1;
```

	country character varying (50)	name character varying (120)	purchase bigint	rank bigint
1	USA	Rock	561	1
2	Canada	Rock	333	1
3	France	Rock	211	1
4	Brazil	Rock	205	1
5	Germany	Rock	194	1
6	United Kingdom	Rock	166	1
7	Czech Republic	Rock	143	1
8	Portugal	Rock	108	1

Total rows: 24 of 24 Query complete 00:00:00.152

--Here is another method to solve the above question but now using recursion

```
with recursive
        sales_per_country as(
        select c.country ,g.name,count(il.quantity) as purchase_per_genre
        from customer c
        join invoice i on i.customer_id = c.customer_id
        join invoice_line il on i.invoice_id = il.invoice_id
        join track t on t.track_id = il.track_id
        join genre g on g.genre_id = t.genre_id
        group by 1,2
        order by 1 asc,3 desc
        ),
        max_genre_per_country as(
        select max(purchase_per_genre) as max_genre_number,country
                from sales_per_country
                group by 2
                order by 2
        )
select sales_per_country.* from sales_per_country
join max_genre_per_country on sales_per_country.country = max_genre_per_country.country
where sales_per_country.purchase_per_genre = max_genre_per_country.max_genre_number
order by sales_per_country.purchase_per_genre desc;
```

	country character varying (50)	name character varying (120)	purchase_per_genre bigint
1	USA	Rock	561
2	Canada	Rock	333
3	France	Rock	211
4	Brazil	Rock	205
5	Germany	Rock	194
6	United Kingdom	Rock	166
7	Czech Republic	Rock	143
8	Portugal	Rock	108

Total rows: 24 of 24 Query complete 00:00:00.124

--Q11: Write a query that determines the customer that has spent the most on music for each

--country. Write a query that returns the country along with the top customer and how

--much they spent. For countries where the top amount spent is shared, provide all

--customers who spent this amount

```
WITH Customter_with_country AS (
        SELECT c.customer_id,c.first_name,c.last_name,c.country,SUM(total) AS total_spending,
    DENSE_RANK() OVER(PARTITION BY c.country ORDER BY SUM(total) DESC) AS Rank
        FROM invoice i
        JOIN customer c ON c.customer_id = i.customer_id
        GROUP BY 1,2,3
        ORDER BY 4 DESC,3 asc)
SELECT * FROM Customter_with_country WHERE Rank = 1
```

	customer_id [PK] integer	first_name character	last_name character	country character varying (50)	total_spending double precision	rank bigint
1	17	Jack	Smith	USA	98.01	1
2	53	Phil	Hughes	United Kingdom	98.01	1
3	51	Joakim	Johansson	Sweden	75.24	1
4	50	Enrique	Muñoz	Spain	98.01	1
5	34	João	Fernandes	Portugal	102.96000000000001	1
6	49	Stanisław	Wójcik	Poland	76.22999999999999	1
7	4	Bjørn	Hansen	Norway	72.27000000000001	1
8	48	Johannes	Van der Berg	Netherlands	65.34	1

Total rows: 24 of 24 Query complete 00:00:00.144

--Here is another method to solve the above question but now using recursion

```
WITH RECURSIVE
    customer_with_country AS (
        SELECT customer.customer_id,first_name,last_name,country,SUM(total) AS total_spending
        FROM invoice
        JOIN customer ON customer.customer_id = invoice.customer_id
        GROUP BY 1,2,3,4
        ORDER BY 2,3 DESC),

    country_max_spending AS(
        SELECT country,MAX(total_spending) AS max_spending
        FROM customer_with_country
        GROUP BY country)
```

```
SELECT cc.country, cc.total_spending, cc.first_name, cc.last_name, cc.customer_id
FROM customer_with_country cc
JOIN country_max_spending ms
ON cc.country = ms.country
WHERE cc.total_spending = ms.max_spending
ORDER BY 1;
```

	country character varying (50)	total_spending double precision	first_name character	last_name character	customer_id [PK] integer
1	Argentina	39.6	Diego	Gutiérrez	56
2	Australia	81.18	Mark	Taylor	55
3	Austria	69.3	Astrid	Gruber	7
4	Belgium	60.38999999999999	Daan	Peeters	8
5	Brazil	108.89999999999998	Luís	Gonçalves	1
6	Canada	99.99	François	Tremblay	3
7	Chile	97.02000000000001	Luis	Rojas	57
8	Czech Republic	144.54000000000002	R	Madhav	5

Total rows: 24 of 24 Query complete 00:00:00.267