

Intro to SQL and Databases Bootcamp

Student Handout



Valeri Analytics 2018 ©

Instructor: Michael Valeri

Handout 1: valerianalytics database schema

ALBUMS TABLE SCHEMA		
field/column	data type	sample data
row	integer	1
artist	text	Red Hot Chili Peppers
album	text	Californication
release date	date	6/8/1999
genre	text	Rock
plays	integer	120
rating	numeric(10,2)	4
org price	money	\$11.99
market value	numeric(10,2)	3.2
burned	boolean	FALSE
playable	boolean	TRUE

NORTH_KOREAN_MISSILE_TESTS TABLE SCHEMA		
field/column	data type	sample data
row id	integer	5
date launched	date	9/1/1984
date entered	date	12/23/2016
launch time utc	date	(null)
missile name	varchar	Scud-B
missile type	varchar	SRBM
launch authority	varchar	(null)
facility name	varchar	Tonghae Satellite
facility location	varchar	Hwadae County
facility other name	varchar	Musudan-ri
facility latitude	float	40.8499966
facility longitude	float	129.666664
landing location	varchar	(null)
apogee	varchar	200 km
distance traveled	varchar	(null)
confirmation status	varchar	Confirmed
test outcome	varchar	Failure
additional info	varchar	(null)

PRODUCTS TABLE SCHEMA		
field/column	data type	sample data
item no	integer	904616
category name	text	TEQUILA
item description	text	Jose Cuervo
vendor	integer	305
vendor name	text	Mhw Ltd
bottle size	integer	750
pack	integer	12
inner pack	integer	1
age	text	(null)
proof	text	40
list date	timestamp	2/11/2009
bottle price	money	\$9.77
shelf price	numeric(10,2)	14.66
case cost	numeric(10,2)	117.22

COUNTIES TABLE SCHEMA		
field/column	data type	sample data
county	text	Adair
population	integer	7682

STORES TABLE SCHEMA		
field/column	data type	sample data
store	integer	2106
name	text	Hillstreet News and Tobacco
store status	text	A
store address	text	217 CollegeCedar Falls, IA
address info	text	(null)

SALES TABLE SCHEMA		
field/column	data type	sample data
date	timestamp	6/27/2014
convenience store	text	Y
store	integer	4771
county number	county_number	57
county	text	Linn
category	text	1081600
category name	text	WHISKEY LIQUEUR
vendor no	text	421
vendor	text	Sazerac Co. Inc.
item	integer	64858
description	text	Fireball Cinnamon Whiskey
pack	integer	1
liter size	integer	3000
state btl cost	money	29.72
btl price	money	44.58
bottle qty	integer	1
total	numeric(10,2)	44.58

Draw the ERD (Entity Relationship Diagram) between all the tables in 'valerianalytics' db

Handout 2: Connecting to 'valerianalytics' database

Step 1 - Download most recent version of PG-Admin 4 and open.

Note: (Mac OS Users - If getting error "PgAdmin4 is damaged and cannot be opened, click troubleshoot link"

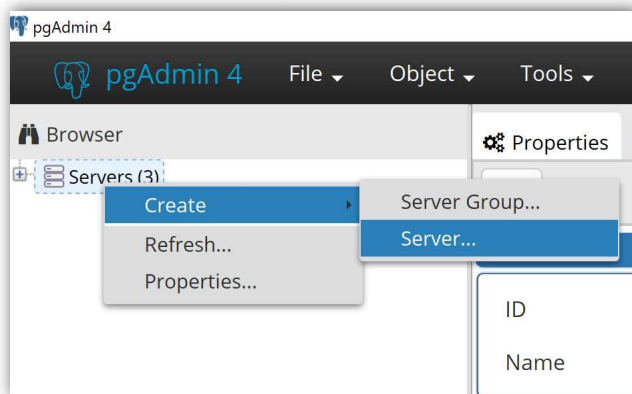
Window <https://www.pgadmin.org/download/pgadmin-4-windows/>

Mac* <https://www.pgadmin.org/download/pgadmin-4-macos/>

MAC Troubleshoot <http://www.tech-recipes.com/rx/45404/mac-downloaded-app-is-damaged-and-cant-be-opened-error-solved/>

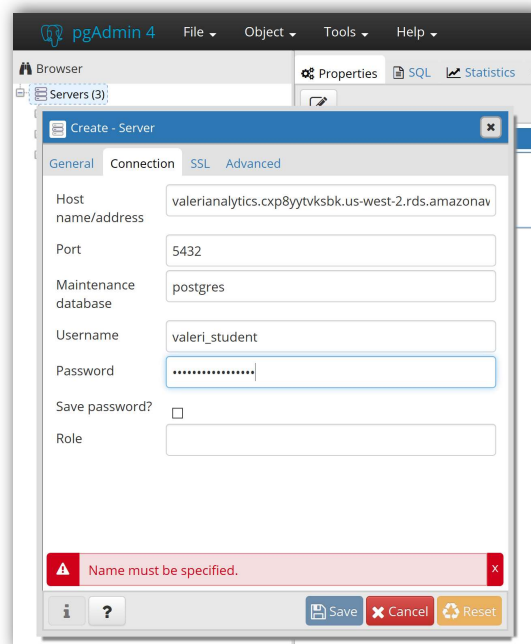
Step 2 - Get connected to 'valerianalytics' db

A. Right Click "Servers" and then select "Create --> Server..."



B. Navigate to the the "Connection" tab, then enter in the following credentials

Hostname/ address	valerianalytics.cxp8yytvksbk.us-west-2.rds.amazonaws.com
Maintenan ce_DB	postgres
Username	valeri_student
Password	analyticsvaleri



C. Navigate to 'General' and name the db

Handout 2 (backup): Connecting to 'analyticsga' database

Step 1 - Download most recent version of PG-Admin 4 and open.

Note: (Mac OS Users - If getting error "PgAdmin4 is damaged and cannot be opened, click troubleshoot link"

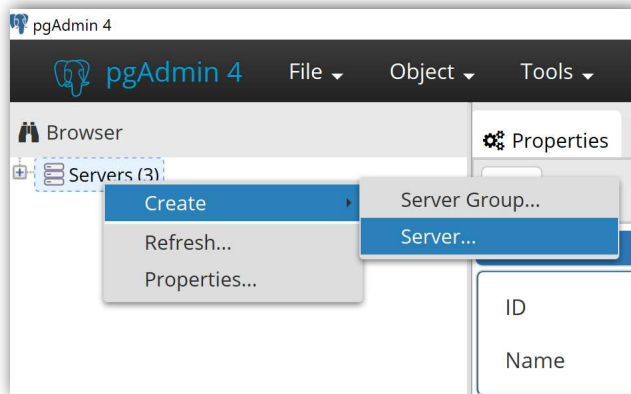
Window <https://www.pgadmin.org/download/pgadmin-4-windows/>

Mac* <https://www.pgadmin.org/download/pgadmin-4-macos/>

MAC Troubleshoot <http://www.tech-recipes.com/rx/45404/mac-downloaded-app-is-damaged-and-cant-be-opened-error-solved/>

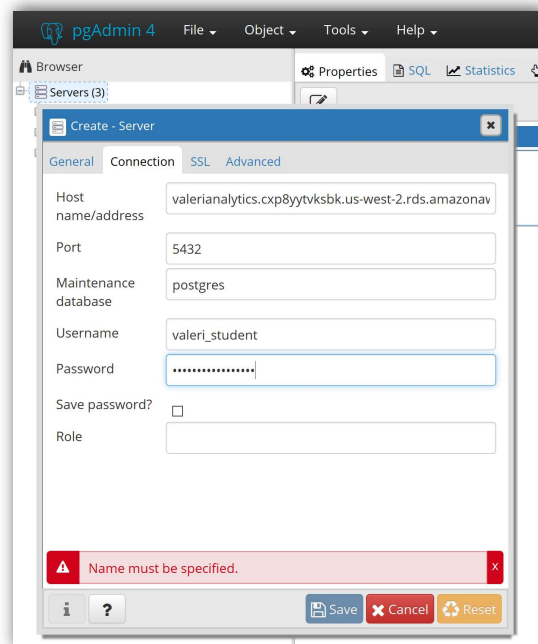
Step 2 - Get connected to 'analyticsga' db

A. Right Click "Servers" and then select "Create --> Server..."



B. Navigate to the the "Connection" tab, then enter in the following credentials

Hostname/ address	Analyticsga.cuwj8wuu6wbh.us- west-2.rds.amazonaws.com
Maintenan ce_DB	postgres
Username	analytics_student
Password	analyticsga



C. Navigate to 'General' and name the db

Handout 3: Albums Table

INT	text	text	date	text	int	numeric	money	numeric	boolean	boolean
Row	Artist	Album	Release_Date	Genre	Plays	Rating	Org_Price	Market_Value	Burned	Playable
1	Red Hot Chili Peppers	Californication	6/8/1999	Rock	120	4	\$11.99	3.2	0	1
2	Red Hot Chili Peppers	By the Way	7/9/2002	Rock	100	3.5	\$11.99	4	1	0
3	Kanye West	College Dropout	2/10/2004	Rap	200	5	\$10.99	5	0	1
4	Kanye West	Late Registration	8/30/2005	Rap	300	4	\$9.99	7	0	1
5	Kanye West	Graduation	9/11/2007	Rap	250	4	\$0.00	1.75	1	1
6	Papa Roach	Infest	4/25/2000	Rock	75	3.5	\$11.99	0.5	0	0
7	Kid Cudi	Man on the Moon II	9/15/2009	Rap	40	4	\$10.99	6	0	0
8	Ratatat	Ratatat	4/20/2004	Electropop	60	5	\$9.99	6	0	(null)
9	Ratatat	Classics	8/22/2006	Electropop	400	4	\$0.00	12.99	(null)	0
10	Dragonforce	Sonic Firestorm	5/11/2004	Rock	500	5	\$2.99	0.01	1	(null)
11	(null)	Summer Mix 08	(null)	(null)	1000	5	\$0.00	0	1	1
12	(null)	Party Mix 07	(null)	(null)	4000	5	\$0.00	0	1	1
13	Common	Be	5/24/2005	Rap	2000	4.5	\$0.00	15	1	1
14	T.I.	Paper Trail	9/26/2008	Rap	300	4	\$0.00	6.99	1	0
15	Children of Bodom	Children of Bodom	(null)	Metal	150	3	\$0.00	0	(null)	0

	Plays	Rating	Org_Price	Market_Value
Sum	9,495	63.5	80.9	68.4
Avg	633.0	4.2	5.4	4.6
Min	40	3.0	-	-
Max	4,000	5.0	12.0	15.0

Handout 4: Basic Syntax of SQL

- 1 **SELECT**
- 2 **FROM**
- 3 **WHERE**
- 4 **GROUP BY**
- 5 **HAVING**
- 6 **ORDER BY**
- 7 **LIMIT**

Definition

Key Takeaways: Basic Syntax of SQL

Handout 5: Filtering with the WHERE Clause

(It's going to be ok, just follow this matrix 1-->2-->3)

Step 1: Choose Condition			STEP 3: Choose data type of column					
USE CASE	OPERATOR	DESCRIPTION	STEP 2: Choose use case	BOOLEAN EXAMPLE	CHARACTER EXAMPLE	NUMERIC EXAMPLE	DATE EXAMPLE	DATE WITH TO_CHAR EXAMPLE
1	=	Equal (single criteria)	You want to filter on 1 column on 1 condition	PLAYABLE = TRUE	ALBUM = 'Californication'	ORG_PRICE = 9.99	RELEASE_DATE = '2004-02-10'	TO_CHAR(RELEASE_DATE, 'YYYY') = '2009'
	>	Greater than (single criteria)				ORG_PRICE > 9.99	RELEASE_DATE > '2004-02-10'	TO_CHAR(RELEASE_DATE, 'YYYY') > '2009'
	<	Less than (single criteria)				ORG_PRICE < 9.99	RELEASE_DATE < '2004-02-10'	TO_CHAR(RELEASE_DATE, 'YYYY') < '2009'
	>=	Greater than or equal (single criteria)				ORG_PRICE >= 9.99	RELEASE_DATE >= '2004-02-10'	TO_CHAR(RELEASE_DATE, 'YYYY') >= '2009'
	<=	Less than or equal (single criteria)				ORG_PRICE <= 9.99	RELEASE_DATE <= '2004-02-10'	TO_CHAR(RELEASE_DATE, 'YYYY') <= '2009'
	<> or !=	Not equal		PLAYABLE != FALSE	ALBUM != 'Californication'	ORG_PRICE != 9.99	RELEASE_DATE != '2004-02-10'	TO_CHAR(RELEASE_DATE, 'YYYY') != '2009'
	LIKE	Look for a specified pattern in a column			ARTIST LIKE '%K'			
	NOT LIKE	Look for a specified pattern in a column (Not Like)			ARTIST NOT LIKE '%K'			
2	BETWEEN	Between two numeric values or dates (multiple criteria)	You want to filter between two ranges (numeric or date)			ORG_PRICE BETWEEN 9.99 AND 12.99	RELEASE_DATE BETWEEN '2001-01-01' AND '2009-01-01'	TO_CHAR(RELEASE_DATE, 'YYYY') BETWEEN '2009' and '2011'
3	AND	Logical operator AND	You want to filter on multiple columns	PLAYABLE = TRUE AND GENRE = 'Rap'	ARTIST = 'Kanye West' AND PLAYABLE = TRUE	9.99	RELEASE_DATE >= '2001-01-01' AND RELEASE_DATE <= '2009-01-01'	(same as left)
	OR	Logical operator OR		PLAYABLE = TRUE OR GENRE = 'Metal'	ARTIST = 'Kanye West' OR RATING = 5	9.99	RELEASE_DATE >= '2001-01-01' OR GENRE = 'Electronic'	(same as left)
4	IS	Logical operator for (null) records	You are looking for null values in a column	PLAYABLE IS NULL	ARTIST IS NULL	ORG_PRICE IS NULL	RELEASE_DATE IS NULL	(same as left)
	IS NOT	Logical operator for not (null) records		PLAYABLE IS NOT NULL	ARTIST IS NOT NULL	ORG_PRICE IS NOT NULL	RELEASE_DATE IS NOT NULL	(same as left)
5	IN	Equal (multiple criteria)	You want to filter on 1 column with multiple conditions	BURNED IN (TRUE, FALSE)	ALBUM IN ('Californication', 'By the Way')	ORG_PRICE IN (9.99, 12.99)	RELEASE_DATE IN ('2012-01-01', '2017-01-02')	TO_CHAR(RELEASE_DATE, 'YYYY') IN ('2009', '2011')
	NOT IN	Not equal (multiple criteria)		BURNED NOT IN (TRUE, FALSE)	ALBUM NOT IN ('Californication', 'By the Way')	ORG_PRICE NOT IN (9.99, 12.99)	RELEASE_DATE NOT IN ('2012-01-01', '2017-01-02')	TO_CHAR(RELEASE_DATE, 'YYYY') NOT IN ('2009', '2011')

Key Takeaways: Filtering with the WHERE Clause

Handout 6: Aggregations and GROUP BY and Fix the Code

Key Takeaways: Aggregations and Group BY

Key Takeaways: Fix the Code

Handout 7a: Joining Tables in SQL

1. Why do we join tables in SQL?

2. What are the 3 requirements to join tables in SQL?

3. What are the 3 rules of the JOIN syntax?

4. How do we visualize a JOIN?

5. How do you join two tables that don't have the same column in common?

5. What do you do if you are getting duplicate records?

Handout 7b: Which JOIN do I use? (Class Exercise)

STEP 1

Table A:
ORDERS

TABLE B:
CUSTOMERS

Make a joined table of order_id, cust_id, and state. Put CUST_IDS in correspondig circle

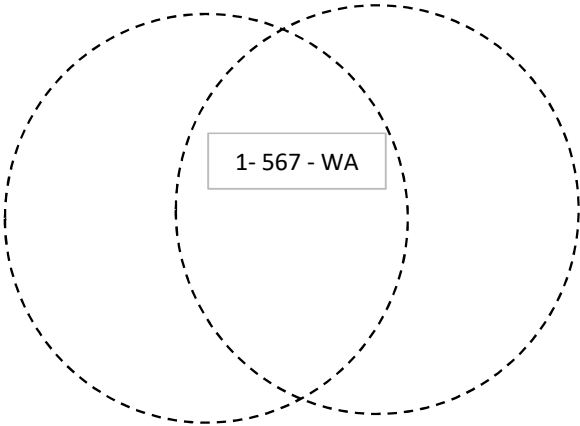


Table A: ORDERS		Table B: CUSTOMERS	
ORDER_ID	CUST_ID	CUST_ID	State
12345	567	567	WA
12346	457	367	SD
12366	123	678	DE
14567	(null)	457	MI
19453	(null)	234	(null)
15678	234		

- 1
- 2
- 3
- 4
- 5
- 6
- 7

ORDER_ID	CUST_ID	STATE

STEP 2

Make a table of the count of orders and customers in each state. Don't count NULL customer_id or order_id

ORDER COUNT	CUSTOMER COUNT	STATE

STEP 3

Tell us the total count of orders and customers overall.

ORDER COUNT	CUSTOMER COUNT

Handout 7c: Supplemental Join Notes

JOIN Syntax for two tables

<pre>SELECT A.ORDERID ,A.CUSTOMERID ,B.CUSTOMERNAME FROM ORDERS A JOIN CUSTOMERS B ON B.CUSTOMERID = A.CUSTOMERID GROUP BY A.ORDERID ,A.CUSTOMERID ,B.CUSTOMERNAME</pre>	<p>1. ASSIGN AN ALIAS to each table and apply it to each distinct column.</p> <p>2. JOIN ON the columns the TWO tables have in common (make sure same data type)</p> <p>3. You need indicate what STYLE of JOIN (more than 1 style)</p>
--	--

Sample Code for Checking for Duplicate Records in Table B

```
-- Sample code to check for duplicate records in Table B.

SELECT
STORE -- This is the column you'll be joining onto in Table B. Eg. customer_id.
,COUNT(*) -- This counts the occurrence of duplicates in column above.

FROM STORES -- Table B goes here

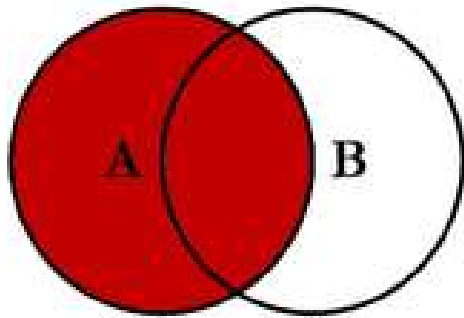
GROUP BY
STORE --Column you'll be joining onto in Table B.

HAVING COUNT(*) > 1 --This returns the primary key records that only have duplicates.
--If there is nothing in result output, then no duplicates for primary key. Clean join.
--If there is records in result output, then duplicate values for primary key. Need to de-deup before JOINing.
```

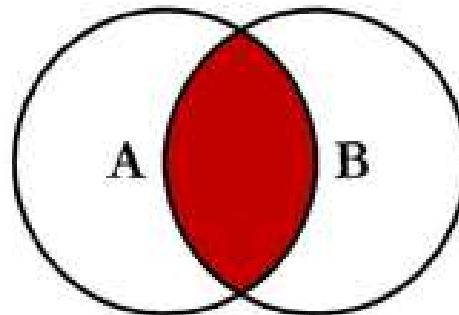
Sample Code for Daisy Chaining (For joining on intermediary table)

```
1 SELECT
2   A.ORDER_ID
3   ,A.CUST_ID
4   ,B.STATE
5 FROM
6   ORDERS A
7 LEFT JOIN ACCOUNT_MAPPING C ON A.CUST_ID = C.CUST_ID
8 LEFT JOIN ACCOUNT_STATES B ON B.ACCOUNT_ID = C.ACCOUNT_ID
```

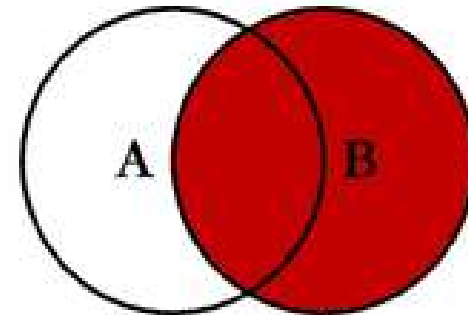
SQL JOINS



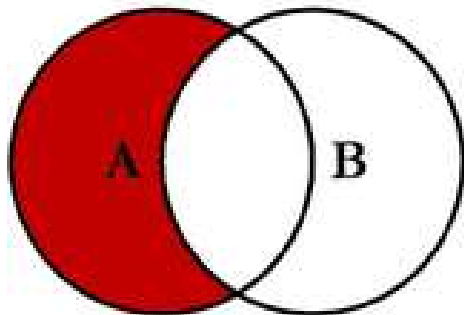
```
SELECT <select_list>  
FROM TableA A  
LEFT JOIN TableB B  
ON A.Key = B.Key
```



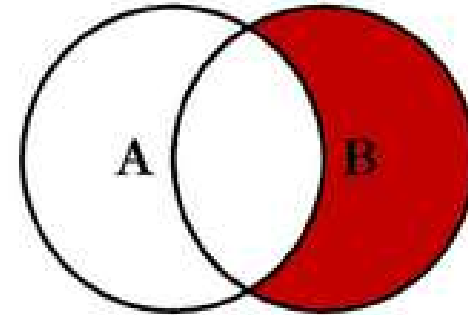
```
SELECT <select_list>  
FROM TableA A  
INNER JOIN TableB B  
ON A.Key = B.Key
```



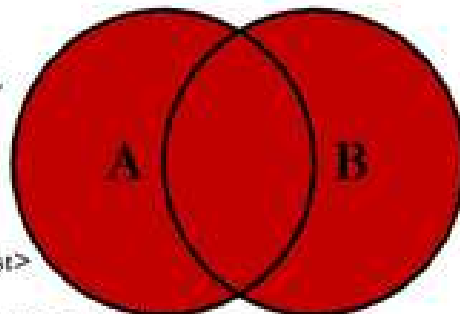
```
SELECT <select_list>  
FROM TableA A  
RIGHT JOIN TableB B  
ON A.Key = B.Key
```



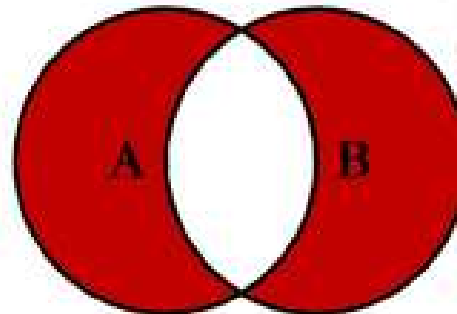
```
SELECT <select_list>  
FROM TableA A  
LEFT JOIN TableB B  
ON A.Key = B.Key  
WHERE B.Key IS NULL
```



```
SELECT <select_list>  
FROM TableA A  
RIGHT JOIN TableB B  
ON A.Key = B.Key  
WHERE A.Key IS NULL
```



```
SELECT <select_list>  
FROM TableA A  
FULL OUTER JOIN TableB B  
ON A.Key = B.Key
```



```
SELECT <select_list>  
FROM TableA A  
FULL OUTER JOIN TableB B  
ON A.Key = B.Key  
WHERE A.Key IS NULL  
OR B.Key IS NULL
```

CONGRATULATIONS!

YOU CAN NOW TALK TO A RELATIONAL DATABASE USING SQL

THANK
YOU!

```
SELECT  
SALUTATION  
FROM ENGLISH  
WHERE SALUTATION_CATEGORY = 'THANK'
```



THANKS FOR SPEAKING WITH US TODAY



PLEASE COMPLETE OUR SURVEY AND HELP IMPROVE OUR COURSE

<https://www.surveymonkey.com/r/NHRDMQK>

Valeri Analytics 2018 ©