

BDS ASSIGNMENT 1

MANISH KUMAR
MEGHARAJ S HOLIYAPPA

Table of Contents

a) The Schema / Structure	2
○ MongoDB	2
○ PostgreSQL.....	7
○ Cassandra.....	15
b) Queries used for each of the below operation	24
○ MongoDB	24
● WRITE.....	24
● READ.....	24
● DELETE	24
● GROUP BY & ORDER BY	25
○ PostgreSQL.....	25
● WRITE.....	25
● READ.....	26
● DELETE	26
● GROUP BY & ORDER BY	27
○ Cassandra.....	27
● WRITE.....	27
● READ.....	28
● DELETE	28
● GROUP BY & ORDER BY	29
c) Performance parameters considered	30
d) Visualizations based on the performance parameters	31
e) Tabular , quick summary of the prototyping exercise	33
f) Clearly spelt out recommendation.....	34
MongoDB is the best choice for this usecase	34
g) Any other observation / finding that you feel is important for this exercise	35

a) The Schema / Structure

○ MongoDB

• ORDERS

└ _id	ObjectId("5f72442e5d0f5c4fa9375495")	ObjectId
└ akeed_order_id	163240.0	Double
└ customer_id	QS68UD8	String
└ item_count	7	String
└ grand_total	8.7	Double
└ payment_mode	1	Int32
└ promo_code		String
└ vendor_discount_amount	0.0	Double
└ promo_code_discount_percentage		String
└ is_favorite		String
└ is_rated	No	String
└ vendor_rating		String
└ driver_rating	0.0	Double
└ deliverydistance	0.0	Double
└ preparationtime		String
└ delivery_time		String
└ order_accepted_time		String
└ driver_accepted_time		String
└ ready_for_pickup_time		String
└ picked_up_time		String
└ delivered_time		String
└ delivery_date	2019-07-31 05:30:00	String
└ vendor_id	294	Int32
└ created_at	2019-08-01 05:31:10	String
└ LOCATION_NUMBER	0	Int32
└ LOCATION_TYPE	Work	String
└ CID X LOC_NUM X VENDOR	QS68UD8 X 0 X 294	String

• TEST_CUSTOMERS

└ _id	ObjectId("5f7244272eda3981ee79e5c7")	ObjectId
└ akeed_customer_id	LRX7BCH	String
└ gender	Male	String
└ dob		String
└ status	1	Int32
└ verified	1	Int32
└ language	EN	String
└ created_at	2018-04-17 20:01:15	String
└ updated_at	2018-04-17 20:01:15	String

• TEST_FULL

	_id	ObjectId("5f724880da331e216540e38e")	ObjectId
↳	customer_id	ICE2DJP	String
↳	gender	Male	String
↳	status_x	1.0	Double
↳	verified_x	1.0	Double
↳	created_at_x	2018-02-07 16:45:36	String
↳	updated_at_x	2018-02-07 16:45:36	String
↳	location_number	0.0	Double
↳	location_type	String	String
↳	latitude_x	-96.44	Double
↳	longitude_x	-67.2	Double
↳	id	4.0	Double
↳	authentication_id	118597.0	Double
↳	latitude_y	-0.5884	Double
↳	longitude_y	0.7544	Double
↳	vendor_category_en	Restaurants	String
↳	vendor_category_id	2.0	Double
↳	delivery_charge	0.0	Double
↳	serving_distance	6.0	Double
↳	is_open	1.0	Double
↳	OpeningTime	11:00AM-11:30PM	String
↳	OpeningTime2	-	String
↳	preparation_time	15.0	Double
↳	commission	0.0	Double
↳	is_akeed_delivering	Yes	String
↳	discount_percentage	0.0	Double
↳	status_y	1.0	Double
↳	verified_y	1.0	Double
↳	rank	11.0	Double
↳	language	EN	String
↳	vendor_rating	4.4	Double
↳	sunday_from_time1	00:00:00	String
↳	sunday_to_time1	00:30:00	String
↳	sunday_from_time2	08:00:00	String
↳	sunday_to_time2	23:59:00	String
↳	monday_from_time1	00:00:00	String
↳	monday_to_time1	00:30:00	String
↳	monday_from_time2	08:00:00	String
↳	monday_to_time2	23:59:00	String
↳	tuesday_from_time1	00:00:00	String
↳	tuesday_to_time1	00:30:00	String
↳	tuesday_from_time2	08:00:00	String
↳	tuesday_to_time2	23:59:00	String
↳	wednesday_from_time1	00:00:00	String
↳	wednesday_to_time1	00:30:00	String
↳	wednesday_from_time2	08:00:00	String
↳	wednesday_to_time2	23:59:00	String
↳	thursday_from_time1	00:00:00	String
↳	thursday_to_time1	00:30:00	String
↳	thursday_from_time2	08:00:00	String
↳	thursday_to_time2	23:59:00	String
↳	friday_from_time1	00:00:00	String
↳	friday_to_time1	00:30:00	String
↳	friday_from_time2	10:00:00	String
↳	friday_to_time2	23:59:00	String
↳	saturday_from_time1	00:00:00	String
↳	saturday_to_time1	00:30:00	String
↳	saturday_from_time2	10:00:00	String
↳	saturday_to_time2	23:59:00	String
↳	primary_tags	{"primary_tags":"4"}	String
↳	open_close_flags	1.0	Double
↳	vendor_tag	2,4,5,8,91,22,12,24,16,23	String
↳	vendor_tag_name	Arabic,Breakfast,Burgers,Desserts,Free Delivery,Grills,Lebanese,Salads,Sandwiches,Sha...	String
↳	one_click_vendor	Y	String
↳	country_id	1.0	Double
↳	city_id	1.0	Double
↳	created_at_y	2018-01-30 14:42:04	String
↳	updated_at_y	2020-04-07 15:12:43	String
↳	device_type	3	Int32
↳	display_orders	1	Int32
↳	location_number_obj	0	Int32
↳	id_obj	4	Int32
↳	CID X LOC_NUM X VENDOR	TCHWPBT X 0 X 4	String
↳	target	0	Int32

• TEST_LOCATIONS

_id	ObjectId("5f7244134327eb7c1f70fa3c")	ObjectId
customer_id	0JP29SK	String
location_number	0	Int32
location_type	Home	String
latitude	0.278708653846154	Double
longitude	-78.623847386364	Double

• TRAIN_CUSTOMERS

_id	ObjectId("5f724408603d72e1f48f70eb")	ObjectId
akeed_customer_id	TCHWPBT	String
gender	Male	String
dob		String
status	1	Int32
verified	1	Int32
language	EN	String
created_at	2018-02-07 19:16:23	String
updated_at	2018-02-07 19:16:23	String

• TRAIN_FULL

_id	ObjectId("5f7244408430538ec239f644")	ObjectId
customer_id	TCHWPBT	String
gender	Male	String
status_x	1	Int32
verified_x	1	Int32
created_at_x	2018-02-07 19:16:23	String
updated_at_x	2018-02-07 19:16:23	String
location_number	0	Int32
location_type	Work	String
latitude_x	-96.44	Double
longitude_x	-67.2	Double
id	4	Int32
authentication_id	118597.0	Double
latitude_y	-0.5884	Double
longitude_y	0.7544	Double
vendor_category_en	Restaurants	String
vendor_category_id	2.0	Double
delivery_charge	0.0	Double
serving_distance	6.0	Double
is_open	1.0	Double
OpeningTime	11:00AM-11:30PM	String
OpeningTime2	-	String
preparation_time	15	Int32
commission	0.0	Double
is_akeed_delivering	Yes	String
discount_percentage	0.0	Double
status_y	1.0	Double
verified_y	1	Int32
rank	11	Int32
language	EN	String
vendor_rating	4.4	Double
sunday_from_time1	00:00:00	String
sunday_to_time1	00:30:00	String
sunday_from_time2	08:00:00	String
sunday_to_time2	23:59:00	String
monday_from_time1	00:00:00	String
monday_to_time1	00:30:00	String
monday_from_time2	08:00:00	String
monday_to_time2	23:59:00	String
tuesday_from_time1	00:00:00	String
tuesday_to_time1	00:30:00	String
tuesday_from_time2	08:00:00	String
tuesday_to_time2	23:59:00	String
wednesday_from_time1	00:00:00	String
wednesday_to_time1	00:30:00	String
wednesday_from_time2	08:00:00	String

wednesday_to_time2	23:59:00	String
thursday_from_time1	00:00:00	String
thursday_to_time1	00:30:00	String
thursday_from_time2	08:00:00	String
thursday_to_time2	23:59:00	String
friday_from_time1	00:00:00	String
friday_to_time1	00:30:00	String
friday_from_time2	10:00:00	String
friday_to_time2	23:59:00	String
saturday_from_time1	00:00:00	String
saturday_to_time1	00:30:00	String
saturday_from_time2	10:00:00	String
saturday_to_time2	23:59:00	String
primary_tags	{"primary_tags":"4"}	String
open_close_flags	1.0	Double
vendor_tag	2,4,5,8,91,22,12,24,16,23	String
vendor_tag_name	Arabic,Breakfast,Burgers,Desserts,Free Delivery,Grills,Lebanese,Salads,Sandwiches,Sha...	String
one_click_vendor	Y	String
country_id	1.0	Double
city_id	1.0	Double
created_at_y	2018-01-30 14:42:04	String
updated_at_y	2020-04-07 15:12:43	String
device_type	3	Int32
display_orders	1	Int32
location_number_obj	0	Int32
id_obj	4	Int32
CID X LOC_NUM X VENDOR	TCHWPBT X 0 X 4	String
target	0	Int32

• TRAIN_LOCATIONS

_id	ObjectId("5f7243fa532dc8a6ee8fd35f")	ObjectId
customer_id	02SFNJH	String
location_number	0	Int32
location_type	NA	String
latitude	1.68239153846158	Double
longitude	-78.7897371818186	Double

● VENDORS

<code>_id</code>	ObjectId("5f7243ccb95f752d1ef6a11d")	ObjectId
<code>id</code>	4	Int32
<code>authentication_id</code>	118597.0	Double
<code>latitude</code>	-0.588596153846155	Double
<code>longitude</code>	0.754434090908845	Double
<code>vendor_category_en</code>	Restaurants	String
<code>vendor_category_id</code>	2.0	Double
<code>delivery_charge</code>	0.0	Double
<code>serving_distance</code>	6.0	Double
<code>is_open</code>	1.0	Double
<code>OpeningTime</code>	11:00AM-11:30PM	String
<code>OpeningTime2</code>	-	String
<code>preparation_time</code>	15	Int32
<code>commission</code>	0.0	Double
<code>is_akeed_delivering</code>	Yes	String
<code>discount_percentage</code>	0.0	Double
<code>status</code>	1.0	Double
<code>verified</code>	1	Int32
<code>rank</code>	11	Int32
<code>language</code>	EN	String
<code>vendor_rating</code>	4.4	Double
<code>sunday_from_time1</code>	00:00:00	String
<code>sunday_to_time1</code>	00:30:00	String
<code>sunday_from_time2</code>	08:00:00	String
<code>sunday_to_time2</code>	23:59:00	String
<code>monday_from_time1</code>	00:00:00	String
<code>monday_to_time1</code>	00:30:00	String
<code>monday_from_time2</code>	08:00:00	String
<code>monday_to_time2</code>	23:59:00	String
<code>tuesday_from_time1</code>	00:00:00	String
<code>tuesday_to_time1</code>	00:30:00	String
<code>tuesday_from_time2</code>	08:00:00	String
<code>tuesday_to_time2</code>	23:59:00	String
<code>wednesday_from_time1</code>	00:00:00	String
<code>wednesday_to_time1</code>	00:30:00	String
<code>wednesday_from_time2</code>	08:00:00	String
<code>wednesday_to_time2</code>	23:59:00	String
<code>thursday_from_time1</code>	00:00:00	String
<code>thursday_to_time1</code>	00:30:00	String
<code>thursday_from_time2</code>	08:00:00	String
<code>thursday_to_time2</code>	23:59:00	String
<code>friday_from_time1</code>	00:00:00	String
<code>friday_to_time1</code>	00:30:00	String
<code>friday_from_time2</code>	10:00:00	String
<code>friday_to_time2</code>	23:59:00	String
<code>saturday_from_time1</code>	00:00:00	String
<code>saturday_to_time1</code>	00:30:00	String
<code>saturday_from_time2</code>	10:00:00	String
<code>saturday_to_time2</code>	23:59:00	String
<code>primary_tags</code>	{"primary_tags":"4"}	String
<code>open_close_flags</code>	1.0	Double
<code>vendor_tag</code>	2,4,5,8,9,12,12,24,16,23	String
<code>vendor_tag_name</code>	Arabic,Breakfast,Burgers,Desserts,Free Delivery,Grills,Lebanese,Salads,Sandwiches,Sha...	String
<code>one_click_vendor</code>	Y	String
<code>country_id</code>	1.0	Double
<code>city_id</code>	1.0	Double
<code>created_at</code>	2018-01-30 14:42:04	String
<code>updated_at</code>	2020-04-07 15:12:43	String
<code>device_type</code>	3	Int32
<code>display_orders</code>	1	Int32

○ PostgreSQL

• ORDERS

#	column_name	data_type	is_nullable	check	column_default	foreign_key	comment
1	akeed_order_id	float8	YES	NULL	NULL	EMPTY	→ NULL
2	customer_id	text	YES	NULL	NULL	EMPTY	→ NULL
3	item_count	float8	YES	NULL	NULL	EMPTY	→ NULL
4	grand_total	float8	YES	NULL	NULL	EMPTY	→ NULL
5	payment_mode	int4	YES	NULL	NULL	EMPTY	→ NULL
6	promo_code	text	YES	NULL	NULL	EMPTY	→ NULL
7	vendor_discount_amount	float8	YES	NULL	NULL	EMPTY	→ NULL
8	promo_code_discount_percentage	text	YES	NULL	NULL	EMPTY	→ NULL
9	is_favorite	text	YES	NULL	NULL	EMPTY	→ NULL
10	is_rated	text	YES	NULL	NULL	EMPTY	→ NULL
11	vendor_rating	text	YES	NULL	NULL	EMPTY	→ NULL
12	driver_rating	float8	YES	NULL	NULL	EMPTY	→ NULL
13	deliverydistance	float8	YES	NULL	NULL	EMPTY	→ NULL
14	preparationtime	text	YES	NULL	NULL	EMPTY	→ NULL
15	delivery_time	text	YES	NULL	NULL	EMPTY	→ NULL
16	order_accepted_time	text	YES	NULL	NULL	EMPTY	→ NULL
17	driver_accepted_time	text	YES	NULL	NULL	EMPTY	→ NULL
18	ready_for_pickup_time	text	YES	NULL	NULL	EMPTY	→ NULL
19	picked_up_time	text	YES	NULL	NULL	EMPTY	→ NULL
20	delivered_time	text	YES	NULL	NULL	EMPTY	→ NULL
21	delivery_date	text	YES	NULL	NULL	EMPTY	→ NULL
22	vendor_id	int4	YES	NULL	NULL	EMPTY	→ NULL
23	created_at	text	YES	NULL	NULL	EMPTY	→ NULL
24	LOCATION_NUMBER	int4	YES	NULL	NULL	EMPTY	→ NULL
25	LOCATION_TYPE	text	YES	NULL	NULL	EMPTY	→ NULL
26	CID X LOC_NUM X VENDOR	text	YES	NULL	NULL	EMPTY	→ NULL

• TEST_CUSTOMERS

#	column_name	data_type	is_nullable	check	column_default	foreign_key	comment
1	akeed_customer_id	text	YES	NULL	NULL	EMPTY	→ NULL
2	gender	text	YES	NULL	NULL	EMPTY	→ NULL
3	dob	text	YES	NULL	NULL	EMPTY	→ NULL
4	status	int4	YES	NULL	NULL	EMPTY	→ NULL
5	verified	int4	YES	NULL	NULL	EMPTY	→ NULL
6	language	text	YES	NULL	NULL	EMPTY	→ NULL
7	created_at	text	YES	NULL	NULL	EMPTY	→ NULL
8	updated_at	text	YES	NULL	NULL	EMPTY	→ NULL

• TEST_FULL

#	column_name	data_type	is_nullable	check	column_default	foreign_key	comment
1	customer_id	text	YES	NULL	NULL	EMPTY	→ NULL
2	gender	text	YES	NULL	NULL	EMPTY	→ NULL
3	status_x	float8	YES	NULL	NULL	EMPTY	→ NULL
4	verified_x	float8	YES	NULL	NULL	EMPTY	→ NULL
5	created_at_x	text	YES	NULL	NULL	EMPTY	→ NULL
6	updated_at_x	text	YES	NULL	NULL	EMPTY	→ NULL
7	location_number	float8	YES	NULL	NULL	EMPTY	→ NULL
8	location_type	text	YES	NULL	NULL	EMPTY	→ NULL
9	latitude_x	float8	YES	NULL	NULL	EMPTY	→ NULL
10	longitude_x	float8	YES	NULL	NULL	EMPTY	→ NULL
11	id	float8	YES	NULL	NULL	EMPTY	→ NULL
12	authentication_id	float8	YES	NULL	NULL	EMPTY	→ NULL
13	latitude_y	float8	YES	NULL	NULL	EMPTY	→ NULL
14	longitude_y	float8	YES	NULL	NULL	EMPTY	→ NULL
15	vendor_category_en	text	YES	NULL	NULL	EMPTY	→ NULL
16	vendor_category_id	float8	YES	NULL	NULL	EMPTY	→ NULL
17	delivery_charge	float8	YES	NULL	NULL	EMPTY	→ NULL
18	serving_distance	float8	YES	NULL	NULL	EMPTY	→ NULL
19	is_open	float8	YES	NULL	NULL	EMPTY	→ NULL
20	OpeningTime	text	YES	NULL	NULL	EMPTY	→ NULL
21	OpeningTime2	text	YES	NULL	NULL	EMPTY	→ NULL
22	prepration_time	float8	YES	NULL	NULL	EMPTY	→ NULL
23	commission	float8	YES	NULL	NULL	EMPTY	→ NULL
24	is_akeed_delivering	text	YES	NULL	NULL	EMPTY	→ NULL
25	discount_percentage	float8	YES	NULL	NULL	EMPTY	→ NULL
26	status_y	float8	YES	NULL	NULL	EMPTY	→ NULL
27	verified_y	float8	YES	NULL	NULL	EMPTY	→ NULL
28	rank	float8	YES	NULL	NULL	EMPTY	→ NULL
29	language	text	YES	NULL	NULL	EMPTY	→ NULL
30	vendor_rating	float8	YES	NULL	NULL	EMPTY	→ NULL
31	sunday_from_time1	text	YES	NULL	NULL	EMPTY	→ NULL
32	sunday_to_time1	text	YES	NULL	NULL	EMPTY	→ NULL
33	sunday_from_time2	text	YES	NULL	NULL	EMPTY	→ NULL
34	sunday_to_time2	text	YES	NULL	NULL	EMPTY	→ NULL
35	monday_from_time1	text	YES	NULL	NULL	EMPTY	→ NULL

#	column_name	data_type	is_nullable	check	column_default	foreign_key	comment
36	monday_to_time1	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
37	monday_from_time2	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
38	monday_to_time2	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
39	tuesday_from_time1	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
40	tuesday_to_time1	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
41	tuesday_from_time2	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
42	tuesday_to_time2	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
43	wednesday_from_time1	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
44	wednesday_to_time1	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
45	wednesday_from_time2	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
46	wednesday_to_time2	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
47	thursday_from_time1	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
48	thursday_to_time1	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
49	thursday_from_time2	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
50	thursday_to_time2	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
51	friday_from_time1	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
52	friday_to_time1	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
53	friday_from_time2	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
54	friday_to_time2	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
55	saturday_from_time1	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
56	saturday_to_time1	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
57	saturday_from_time2	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
58	saturday_to_time2	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
59	primary_tags	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
60	open_close_flags	float8	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
61	vendor_tag	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
62	vendor_tag_name	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
63	one_click_vendor	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
64	country_id	float8	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
65	city_id	float8	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
66	created_at_y	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
67	updated_at_y	text	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
68	device_type	float8	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
69	display_orders	float8	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
70	location_number_obj	int4	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL
71	id_obj	int4	✓	YES	✗ NULL	NULL	✗ EMPTY → NULL

• TEST_LOCATIONS

#	column_name	data_type	is_nullable	check	column_default	foreign_key	comment
1	customer_id	text	YES	NULL	NULL	EMPTY	→ NULL
2	location_number	int4	YES	NULL	NULL	EMPTY	→ NULL
3	location_type	text	YES	NULL	NULL	EMPTY	→ NULL
4	latitude	float8	YES	NULL	NULL	EMPTY	→ NULL
5	longitude	float8	YES	NULL	NULL	EMPTY	→ NULL

• TRAIN_CUSTOMERS

#	column_name	data_type	is_nullable	check	column_default	foreign_key	comment
1	akeed_customer_id	text	YES	NULL	NULL	EMPTY	→ NULL
2	gender	text	YES	NULL	NULL	EMPTY	→ NULL
3	dob	text	YES	NULL	NULL	EMPTY	→ NULL
4	status	int4	YES	NULL	NULL	EMPTY	→ NULL
5	verified	int4	YES	NULL	NULL	EMPTY	→ NULL
6	language	text	YES	NULL	NULL	EMPTY	→ NULL
7	created_at	text	YES	NULL	NULL	EMPTY	→ NULL
8	updated_at	text	YES	NULL	NULL	EMPTY	→ NULL

• TRAIN_FULL

#	column_name	data_type	is_nullable	check	column_default	foreign_key	comment
1	customer_id	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
2	gender	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
3	status_x	int4	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
4	verified_x	int4	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
5	created_at_x	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
6	updated_at_x	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
7	location_number	int4	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
8	location_type	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
9	latitude_x	float8	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
10	longitude_x	float8	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
11	id	int4	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
12	authentication_id	float8	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
13	latitude_y	float8	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
14	longitude_y	float8	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
15	vendor_category_en	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
16	vendor_category_id	float8	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
17	delivery_charge	float8	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
18	serving_distance	float8	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
19	is_open	float8	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
20	OpeningTime	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
21	OpeningTime2	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
22	preparation_time	int4	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
23	commission	float8	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
24	is_akeed_delivering	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
25	discount_percentage	float8	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
26	status_y	float8	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
27	verified_y	int4	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
28	rank	int4	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
29	language	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
30	vendor_rating	float8	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
31	sunday_from_time1	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
32	sunday_to_time1	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
33	sunday_from_time2	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
34	sunday_to_time2	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
35	monday_from_time1	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL

#	column_name	data_type	is_nullable	check	column_default	foreign_key	comment
36	monday_to_time1	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
37	monday_from_time2	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
38	monday_to_time2	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
39	tuesday_from_time1	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
40	tuesday_to_time1	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
41	tuesday_from_time2	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
42	tuesday_to_time2	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
43	wednesday_from_time1	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
44	wednesday_to_time1	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
45	wednesday_from_time2	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
46	wednesday_to_time2	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
47	thursday_from_time1	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
48	thursday_to_time1	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
49	thursday_from_time2	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
50	thursday_to_time2	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
51	friday_from_time1	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
52	friday_to_time1	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
53	friday_from_time2	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
54	friday_to_time2	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
55	saturday_from_time1	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
56	saturday_to_time1	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
57	saturday_from_time2	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
58	saturday_to_time2	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
59	primary_tags	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
60	open_close_flags	float8	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
61	vendor_tag	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
62	vendor_tag_name	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
63	one_click_vendor	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
64	country_id	float8	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
65	city_id	float8	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
66	created_at_y	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
67	updated_at_y	text	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
68	device_type	int4	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
69	display_orders	int4	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
70	location_number_obj	int4	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL
71	id_obj	int4	◊ YES	◊ NULL	NULL	◊ EMPTY	→ NULL

• TRAIN_LOCATIONS

#	column_name	data_type	is_nullable	check	column_default	foreign_key	comment
1	customer_id	text	YES	NULL	NULL	EMPTY	→ NULL
2	location_number	int4	YES	NULL	NULL	EMPTY	→ NULL
3	location_type	text	YES	NULL	NULL	EMPTY	→ NULL
4	latitude	float8	YES	NULL	NULL	EMPTY	→ NULL
5	longitude	float8	YES	NULL	NULL	EMPTY	→ NULL

• VENDORS

#	column_name	data_type	is_nullable	check	column_default	foreign_key	comment
1	id	int4	YES	NULL	NULL	EMPTY	→ NULL
2	authentication_id	float8	YES	NULL	NULL	EMPTY	→ NULL
3	latitude	float8	YES	NULL	NULL	EMPTY	→ NULL
4	longitude	float8	YES	NULL	NULL	EMPTY	→ NULL
5	vendor_category_en	text	YES	NULL	NULL	EMPTY	→ NULL
6	vendor_category_id	float8	YES	NULL	NULL	EMPTY	→ NULL
7	delivery_charge	float8	YES	NULL	NULL	EMPTY	→ NULL
8	serving_distance	float8	YES	NULL	NULL	EMPTY	→ NULL
9	is_open	float8	YES	NULL	NULL	EMPTY	→ NULL
10	OpeningTime	text	YES	NULL	NULL	EMPTY	→ NULL
11	OpeningTime2	text	YES	NULL	NULL	EMPTY	→ NULL
12	preparation_time	int4	YES	NULL	NULL	EMPTY	→ NULL
13	commission	float8	YES	NULL	NULL	EMPTY	→ NULL
14	is_akeed_delivering	text	YES	NULL	NULL	EMPTY	→ NULL
15	discount_percentage	float8	YES	NULL	NULL	EMPTY	→ NULL
16	status	float8	YES	NULL	NULL	EMPTY	→ NULL
17	verified	int4	YES	NULL	NULL	EMPTY	→ NULL
18	rank	int4	YES	NULL	NULL	EMPTY	→ NULL
19	language	text	YES	NULL	NULL	EMPTY	→ NULL
20	vendor_rating	float8	YES	NULL	NULL	EMPTY	→ NULL
21	sunday_from_time1	text	YES	NULL	NULL	EMPTY	→ NULL
22	sunday_to_time1	text	YES	NULL	NULL	EMPTY	→ NULL
23	sunday_from_time2	text	YES	NULL	NULL	EMPTY	→ NULL
24	sunday_to_time2	text	YES	NULL	NULL	EMPTY	→ NULL
25	monday_from_time1	text	YES	NULL	NULL	EMPTY	→ NULL
26	monday_to_time1	text	YES	NULL	NULL	EMPTY	→ NULL
27	monday_from_time2	text	YES	NULL	NULL	EMPTY	→ NULL
28	monday_to_time2	text	YES	NULL	NULL	EMPTY	→ NULL
29	tuesday_from_time1	text	YES	NULL	NULL	EMPTY	→ NULL
30	tuesday_to_time1	text	YES	NULL	NULL	EMPTY	→ NULL
31	tuesday_from_time2	text	YES	NULL	NULL	EMPTY	→ NULL
32	tuesday_to_time2	text	YES	NULL	NULL	EMPTY	→ NULL
33	wednesday_from_time1	text	YES	NULL	NULL	EMPTY	→ NULL
34	wednesday_to_time1	text	YES	NULL	NULL	EMPTY	→ NULL
35	wednesday_from_time2	text	YES	NULL	NULL	EMPTY	→ NULL

36	wednesday_to_time2	text	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
37	thursday_from_time1	text	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
38	thursday_to_time1	text	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
39	thursday_from_time2	text	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
40	thursday_to_time2	text	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
41	friday_from_time1	text	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
42	friday_to_time1	text	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
43	friday_from_time2	text	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
44	friday_to_time2	text	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
45	saturday_from_time1	text	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
46	saturday_to_time1	text	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
47	saturday_from_time2	text	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
48	saturday_to_time2	text	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
49	primary_tags	text	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
50	open_close_flags	float8	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
51	vendor_tag	text	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
52	vendor_tag_name	text	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
53	one_click_vendor	text	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
54	country_id	float8	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
55	city_id	float8	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
56	created_at	text	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
57	updated_at	text	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
58	device_type	int4	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL
59	display_orders	int4	◊	YES	◊ NULL	NULL	◊ EMPTY	→ NULL

- Cassandra

- ORDERS

Name	Data Type
customer_id	text
akeed_order_id	double
CID X LOC_NUM X VENDOR	text
created_at	text
delivered_time	text
delivery_date	text
delivery_time	text
deliverydistance	double
driver_accepted_time	text
driver_rating	double
grand_total	double
is_favorite	boolean
is_rated	boolean
item_count	double
location_number	int
location_type	text
order_accepted_time	text
payment_mode	int
picked_up_time	text
preparationtime	text
promo_code	text
promo_code_discount_percentage	double
ready_for_pickup_time	text
vendor_discount_amount	double
vendor_id	int
vendor_rating	double

135303 rows imported from 1 files in 6.011 seconds (0 skipped).

• TEST_CUSTOMERS

Name	Data Type
akeed_customer_id	text
created_at	text
dob	text
gender	text
language	text
status	int
updated_at	text
verified	int

• TEST_FULL

Name	Data Type
CID X LOC_NUM X VENDOR	text
authentication_id	double
city_id	double
commission	double
country_id	double
created_at_x	text
created_at_y	text
customer_id	text
delivery_charge	double
device_type	double
discount_percentage	double
display_orders	double
friday_from_time1	text
friday_from_time2	text
friday_to_time1	text
friday_to_time2	text
gender	text
id	double
id_obj	double
is_akeed_delivering	boolean
is_open	boolean
language	text
latitude_x	double
latitude_y	double
location_number	double
location_number_obj	double

location_type	text
longitude_x	double
longitude_y	double
monday_from_time1	text
monday_from_time2	text
monday_to_time1	text
monday_to_time2	text
one_click_vendor	text
open_close_flags	boolean
openingtime	text
openingtime2	text
prepration_time	text
primary_tags	text
rank	double
saturday_from_time1	text
saturday_from_time2	text
saturday_to_time1	text
saturday_to_time2	text
serving_distance	double
status_x	double
status_y	double
sunday_from_time1	text
sunday_from_time2	text
sunday_to_time1	text
sunday_to_time2	text
thursday_from_time1	text

thursday_from_time2	text
thursday_to_time1	text
thursday_to_time2	text
tuesday_from_time1	text
tuesday_from_time2	text
tuesday_to_time1	text
tuesday_to_time2	text
updated_at_x	text
updated_at_y	text
vendor_category_en	text
vendor_category_id	double
vendor_rating	double
vendor_tag	text
vendor_tag_name	text
verified_x	double
verified_y	double
wednesday_from_time1	text
wednesday_from_time2	text
wednesday_to_time1	text
wednesday_to_time2	text

Processed: 1672000 rows; Rate: 5744 rows/s; Avg. rate: 9557 rows/s

1672000 rows imported from 1 files in 2 minutes and 54.947 seconds (0 skipped).

• TEST_LOCATIONS

Name	Data Type
customer_id	text
latitude	double
location_number	int
location_type	text
longitude	double

• TRAIN_CUSTOMERS

Name	Data Type
akeed_customer_id	text
created_at	text
dob	text
gender	text
language	text
status	int
updated_at	text
verified	int

• TRAIN_FULL

Name	Data Type
CID X LOC_NUM X VENDOR	text
authentication_id	double
city_id	double
commission	double
country_id	double
created_at_x	text
created_at_y	text
customer_id	text
delivery_charge	double
device_type	double
discount_percentage	double
display_orders	double
friday_from_time1	text
friday_from_time2	text
friday_to_time1	text
friday_to_time2	text
gender	text
id	double
id_obj	double
is_akeed_delivering	boolean
is_open	boolean
language	text
latitude_x	double
latitude_y	double
location_number	double

location_number_obj	double
location_type	text
longitude_x	double
longitude_y	double
monday_from_time1	text
monday_from_time2	text
monday_to_time1	text
monday_to_time2	text
one_click_vendor	text
open_close_flags	boolean
openingtime	text
openingtime2	text
prepreration_time	text
primary_tags	text
rank	double
saturday_from_time1	text
saturday_from_time2	text
saturday_to_time1	text
saturday_to_time2	text
serving_distance	double
status_x	double
status_y	double
sunday_from_time1	text
sunday_from_time2	text
sunday_to_time1	text

	sunday_to_time2	text
	target	double
	thursday_from_time1	text
	thursday_from_time2	text
	thursday_to_time1	text
	thursday_to_time2	text
	tuesday_from_time1	text
	tuesday_from_time2	text
	tuesday_to_time1	text
	tuesday_to_time2	text
	updated_at_x	text
	updated_at_y	text
	vendor_category_en	text
	vendor_category_id	double
	vendor_rating	double
	vendor_tag	text
	vendor_tag_name	text
	verified_x	double
	verified_y	double
	wednesday_from_time1	text
	wednesday_from_time2	text
	wednesday_to_time1	text
	wednesday_to_time2	text

Processed: 5802400 rows; Rate: 6884 rows/s; Avg. rate: 10068 rows/s

5802400 rows imported from 1 files in 9 minutes and 36.297 seconds (0 skipped).

• TRAIN_LOCATIONS

Name	Data Type
customer_id	text
latitude	double
location_number	int
location_type	text
longitude	double

- VENDORS

Name	Data Type
id	double
authentication_id	double
city_id	double
commission	double
country_id	double
created_at	text
delivery_charge	double
device_type	double
discount_percentage	double
display_orders	double
friday_from_time1	text
friday_from_time2	text
friday_to_time1	text
friday_to_time2	text
is_akeed_delivering	boolean
is_open	boolean
language	text
latitude	double
longitude	double
monday_from_time1	text
monday_from_time2	text
monday_to_time1	text
monday_to_time2	text
one_click_vendor	text
open_close_flags	boolean

openingtime	text
openingtime2	text
prepration_time	text
primary_tags	text
rank	double
saturday_from_time1	text
saturday_from_time2	text
saturday_to_time1	text
saturday_to_time2	text
serving_distance	double
status	double
sunday_from_time1	text
sunday_from_time2	text
sunday_to_time1	text
sunday_to_time2	text
thursday_from_time1	text
thursday_from_time2	text
thursday_to_time1	text
thursday_to_time2	text
tuesday_from_time1	text
tuesday_from_time2	text
tuesday_to_time1	text
tuesday_to_time2	text
updated_at	text
vendor_category_en	text
vendor_category_id	double
vendor_rating	double
vendor_tag	text
vendor_tag_name	text
verified	double
wednesday_from_time1	text
wednesday_from_time2	text
wednesday_to_time1	text
wednesday_to_time2	text

b) Queries used for each of the below operation

○ MongoDB

● WRITE

Record Exist

```
Local MongoDB  localhost:27017  bds_restaurant
db.orders.update({ "akeed_order_id" : 163240 }, { $set : { "item_count" : "10" }}, {upsert : true})  
  
0.002 sec.  
Updated 1 existing record(s) in 1ms
```

Record Not Exist

```
Local MongoDB  localhost:27017  bds_restaurant
db.orders.update({ "akeed_order_id" : 200213 }, { $set : { "item_count" : "10" }}, {upsert : true})  
  
0.093 sec.  
Updated 1 new record(s) in 93ms
```

● READ

```
db.getCollection('train_full').find({customer_id: 'DFHGZF9', gender: 'Male'})  
  
train_full  0.45 sec.  


| Key                                         | Value         | Type   |
|---------------------------------------------|---------------|--------|
| ▶ (1) ObjectId("5f72446a8430538ec244ba84")  | { 74 fields } | Object |
| ▶ (2) ObjectId("5f72446a8430538ec244ba85")  | { 74 fields } | Object |
| ▶ (3) ObjectId("5f72446a8430538ec244ba86")  | { 74 fields } | Object |
| ▶ (4) ObjectId("5f72446a8430538ec244ba87")  | { 74 fields } | Object |
| ▶ (5) ObjectId("5f72446a8430538ec244ba88")  | { 74 fields } | Object |
| ▶ (6) ObjectId("5f72446a8430538ec244ba89")  | { 74 fields } | Object |
| ▶ (7) ObjectId("5f72446a8430538ec244ba8a")  | { 74 fields } | Object |
| ▶ (8) ObjectId("5f72446a8430538ec244ba8b")  | { 74 fields } | Object |
| ▶ (9) ObjectId("5f72446a8430538ec244ba8c")  | { 74 fields } | Object |
| ▶ (10) ObjectId("5f72446a8430538ec244ba8d") | { 74 fields } | Object |
| ▶ (11) ObjectId("5f72446a8430538ec244ba8e") | { 74 fields } | Object |
| ▶ (12) ObjectId("5f72446a8430538ec244ba8f") | { 74 fields } | Object |


```

● DELETE

```
Local MongoDB  localhost:27017  bds_restaurant
db.train_full.remove({id: 4})  
  
7.76 sec.  
Removed 58024 record(s) in 7759ms
```

- GROUP BY & ORDER BY

```
db.orders.aggregate([
  { $group : { _id : "$customer_id", count: { $sum: 1 } } },
  { $sort:{ "count": -1} }
])
```

orders 0.758 sec.

	_id	count
1	XW90EAP	262.0
2	TL7Z2DM	151.0
3	VDEJEMP	146.0
4	HJFTTGW	128.0
5	B28LJKM	119.0

- PostgreSQL

- WRITE

```
1
2 INSERT INTO "bds_restaurant"."orders"
3 (akeed_order_id, customer_id, item_count, created_at)
4 VALUES
5 ('190213', 'QS68UD8','1', now()),
6 ('163240', 'QS68UD8','2', now())
7     ON CONFLICT (akeed_order_id) DO UPDATE
8     SET item_count = EXCLUDED.item_count;
9
```

⚙ line 9, column 1, location 252

akeed_order_id	customer_id	item_count	grand_total	payment_mode	promo_code
163240	QS68UD8	2	8.7	1	EMPTY
190213	QS68UD8	1	NULL	NULL	NULL

588 µs

●READ

```
1 select * from "bds_restaurant"."train_full" where customer_id='DFHGZF9';  
2
```

⚙️ line 1, column 72, location 71

customer_id	gender	status_x	verified_x	created_at_x	updated_at_x	location_number
DFHGZF9	Male		1	1 2018-10-1...	2018-10-1...	0
DFHGZF9	Male		1	1 2018-10-1...	2018-10-1...	0
DFHGZF9	Male		1	1 2018-10-1...	2018-10-1...	0
DFHGZF9	Male		1	1 2018-10-1...	2018-10-1...	0
DFHGZF9	Male		1	1 2018-10-1...	2018-10-1...	0
DFHGZF9	Male		1	1 2018-10-1...	2018-10-1...	0
DFHGZF9	Male		1	1 2018-10-1...	2018-10-1...	0
DFHGZF9	Male		1	1 2018-10-1...	2018-10-1...	0
DFHGZF9	Male		1	1 2018-10-1...	2018-10-1...	0

8.346 s

●DELETE

```
3 delete from "bds_restaurant"."train_full" WHERE id=4;  
4
```

⚙️ 53 characters selected

Query 1 OK: DELETE 58024, 58024 rows affected

8.756 s

● GROUP BY & ORDER BY

```
3 delete from "bds_restaurant"."train_full" WHERE id=4;
4
5
6 select customer_id, count(*) from "bds_restaurant"."orders" group by customer_id order by count(*) desc;
```

line 6, column 105, location 234

customer_id	count
XW90EAP	262
TL7Z2DM	151
VDEJEMP	146
HJFTTGW	128
B28LJKM	119
2JJAO46	115
	129 ms
	27,4

○ Cassandra

● WRITE

- Existing record got updated

```
update orders set item_count=2
where akeed_order_id = 163240 and customer_id='QS68UD8' ;
```

)

Execution time: 0.008 s

- New record got created, which was not there.

```
update orders set item_count=3
where akeed_order_id = 193240 and customer_id='QS68UD8';
```

Execution time: 0.012 s

●READ

```
3 CREATE INDEX ON train_full (customer_id);
4
5 SELECT * from train_full where customer_id='DFHGZF9';
6
```

Result Log



CID X LOC_NUM X VENDOR	authentication_id	city_id	commission	country_id	created_at_x	created_at_y	customer_id
► DFHGZF9 X 1 X 459	119063	1	0	1	2018-10-18 22:25:00	2019-09-10 19:04:26	DFHGZF9
DFHGZF9 X 1 X 20	118616	1	0	1	2018-10-18 22:25:00	2018-05-04 22:28:22	DFHGZF9
DFHGZF9 X 0 X 85	118682	1	0	1	2018-10-18 22:25:00	2018-09-16 20:01:04	DFHGZF9
DFHGZF9 X 0 X 113	118710	1	0	1	2018-10-18 22:25:00	2018-10-24 18:11:35	DFHGZF9
DFHGZF9 X 1 X 4	118597	1	0	1	2018-10-18 22:25:00	2018-01-30 14:42:04	DFHGZF9
DFHGZF9 X 1 X 78	118675	1	0	1	2018-10-18 22:25:00	2018-08-26 21:47:55	DFHGZF9
DFHGZF9 X 0 X 231	118830	1	0	1	2018-10-18 22:25:00	2019-04-24 20:04:18	DFHGZF9
DFHGZF9 X 0 X 679	124147	1	null	1	2018-10-18 22:25:00	2019-11-17 13:27:17	DFHGZF9
DFHGZF9 X 1 X 90	118687	1	0	1	2018-10-18 22:25:00	2018-10-04 19:41:36	DFHGZF9
DFHGZF9 X 0 X 216	118815	1	0	1	2018-10-18 22:25:00	2019-04-09 12:50:49	DFHGZF9
DFHGZF9 X 1 X 221	118820	1	0	1	2018-10-18 22:25:00	2019-04-13 19:09:58	DFHGZF9
DFHGZF9 X 1 X 149	118746	1	0	1	2018-10-18 22:25:00	2019-01-06 20:42:17	DFHGZF9

Rows fetched: 200

Result Log

```
1 Execution time: 0.032 s
2 |
```

●DELETE

```
delete from train_full where "CID X LOC_NUM X VENDOR" in ('MZKKKGJU X 1 X 4','MZKKKGJU X 1 X 4',
'L7C8HC4 X 0 X 4','9BKV79P X 1 X 4','BYEWRGZ X 0 X 4','BU9FASD X 0 X 4','6VIYGDF X 0 X 4','EP524LN X 1 X 4',
'SARTXD7 X 0 X 4','W9ZUJ4U X 4 X 4','USBZ15E X 2 X 4','BXNL20R X 0 X 4','ZCRXM7E X 0 X 4','IO6EVDV X 0 X 4',
'QGOR3NL X 0 X 4','AVNXNSJ X 3 X 4','89SR3X6 X 2 X 4','VLW998E X 3 X 4','9MT9IVA X 1 X 4','SGZD9ZF X 0 X 4',
'JALMVNE X 0 X 4','94UGQC7 X 3 X 4','AXPWOOE X 0 X 4','LRD01SQ X 0 X 4','BHD9DDJ X 0 X 4','17YRR00 X 0 X 4');
```

```
Execution time: 0.012 s
```

* Delete in Cassandra is only possible with including all keys in where condition.

• GROUP BY & ORDER BY

```
12 | select customer_id, count(*) as count from orders group by customer_id;
13 |
```

Result Log

◀ ▶ ⏪ ⏩ ⏴ ⏵

customer_id	count
PDJ89E2	5
Q8H9EHH	5
1JUJQDZ	3
X2UGA04	5
AVSABNG	1
YX08GH3	3
O01ZSLF	2
ZDCOKNP	5
EJT3MWI	9
9JJJSU2J	4
MZPBM89	14
WASR1QY	5
A30O8W0	1

Rows fetched: 27444

```
1 | Execution time: 0.179 s
2 |
```

ORDER BY :

```
cqlsh:cass_asign1_database> PAGING OFF;
Disabled Query paging.
cqlsh:cass_asign1_database> select akeed_order_id, customer_id from orders where customer_id
... in('BTB20BZ', 'RFFMC2S', 'BKECLR4', 'TFO0E97', 'IXWXK1L')
... and akeed_order_id in (185156, 218961, 278519, 365975, 253442)
... order by customer_id ASC;

akeed_order_id | customer_id
-----+-----
 2.7852e+05 |      BKECLR4
 1.8516e+05 |      BTB20BZ
 2.5344e+05 |      IXWXK1L
 2.1896e+05 |      RFFMC2S
 3.6598e+05 |      TFO0E97

(5 rows)
cqlsh:cass_asign1_database>
```

c) Performance parameters considered

MongoDB is the leading document database. It is built on a distributed, scale-out architecture and has become a comprehensive cloud-based platform for managing and delivering data to applications. MongoDB handles **transactional, operational, and analytical workloads** at scale. If your concerns are time to market, developer productivity, supporting DevOps and agile methodologies, and building stuff that scales without operational gymnastics, MongoDB is the way to go.

PostgreSQL is a rock solid, open source, enterprise-grade SQL database that has been expanding its capabilities for 30 years. Everything you would ever want from a relational database is present in PostgreSQL, which relies on a scale-up architecture. If your concerns are compatibility, serving up thousands of queries from hundreds of tables, taking advantage of existing SQL skills, and pushing SQL to the limit, PostgreSQL will do an awesome job.

Several key performance factors influence database performance. Understanding these factors helps identify performance opportunities:

- **Query Performance**

As we compared with all the DB, MongoDB performance is better. Cassandra DB performance is much better compared with PostgreSQL. Cassandra DB does not provide the ability to run the query with Non-key.

- **Performance of Import CSV file into DB**

MongoDB provides the database tool called “mongoimport” which took ~5mins of to load ~4GB of data into DB but PostgreSQL took around ~1hour time. Cassandra import functionality better compared with PostgreSQL.

- **Tweaking of DB Schema**

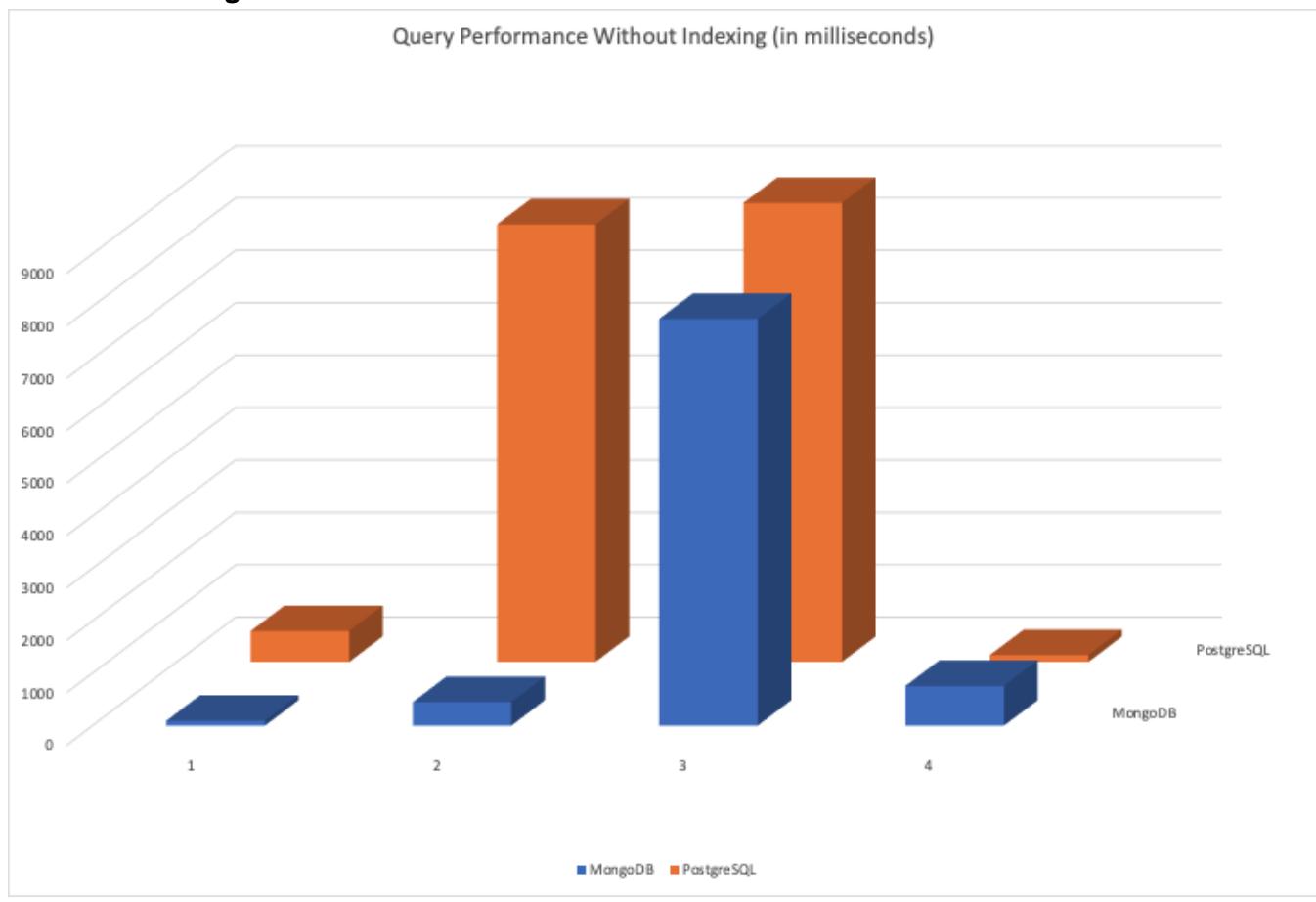
If we are at the beginning of a development project and are seeking to figure out your needs and data model by using an agile development process, MongoDB will shine because developers can reshape the data on their own, when they need to. MongoDB enables you to manage data of any structure, not just tabular structures defined in advance. In a typical relation database this involves lot of effort to make it happen.

- Even Though we have not evaluated below metrics as part of this Assignment but found the parameters which helps to finalize the database on this

- Scalability
- Resilience
- Security
- System Resources

d) Visualizations based on the performance parameters

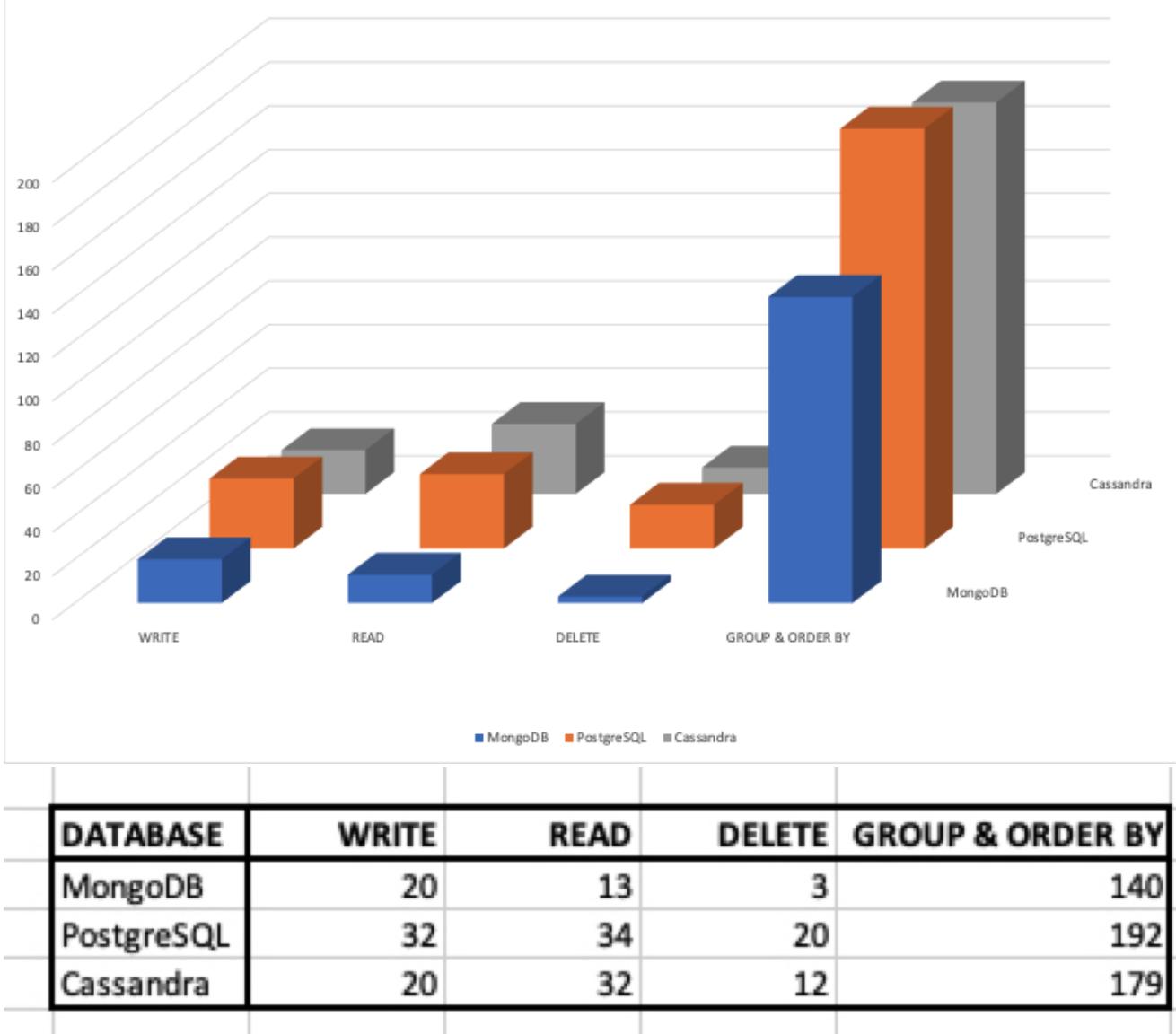
- Without Indexing



DATABASE	WRITE	READ	DELETE	GROUP & ORDER BY
MongoDB	93	450	7759	758
PostgreSQL	588	8346	8756	129

- With Indexing

Query Performance With Indexing (in milliseconds)



NOTE: Cassandra does not provide ability to execute query without index. Hence not included the cassandra comparison in the above chart

e) Tabular , quick summary of the prototyping exercise

MongoDB vs PostgreSQL vs Cassandra Terminology and Concepts

MongoDB	PostgreSQL	Cassandra
Collection	Table	Table
Document	Row	Row
Field	Column	Column
Embedded documents, \$lookup & \$graphLookup, \$unionWith	JOINS, UNIONs	It's not supported, but it can be achieved by using SparkSQL with Cassandra
Aggregation Pipeline	GROUP_BY	GROUP_BY
Sort	ORDER_BY	ORDER_BY

f) Clearly spelt out recommendation

MongoDB is the best choice for this usecase

- If we are supporting an application will have to scale in terms of volume of traffic or size of data (or both) and that needs to be distributed across regions for data locality or data sovereignty, MongoDB's scale-out architecture will meet those needs automatically.
- If you want a multi-cloud database that works the same way in every public cloud, can store customer data in specific geographic regions, and support the latest serverless and mobile development paradigms, MongoDB Atlas is the right choice.
- With the current emphasis on "Big Data", NoSQL databases have surged in popularity. These databases are claimed to perform better than SQL databases.
- MongoDB's document data model maps naturally to objects in application code, making it simple for developers to learn and use. Documents give you the ability to represent hierarchical relationships to store arrays and other more complex structures easily. JSON documents can store data in fields, as arrays, or even as nested sub-documents. In this way, related information can be stored together for fast query access through the rich and expressive MongoDB query language.
- This paper compares key-value store implementations on NoSQL and SQL databases. While NoSQL databases are generally optimized for key-value stores, SQL databases are not. Yet, we find that not all NoSQL databases perform better than the SQL database.
- MongoDB stores data as documents in a binary representation called BSON (Binary JSON). Fields can vary from document to document; there is no need to declare the structure of documents to the system – documents are self-describing. If a new field needs to be added to a document, then the field can be created without affecting all other documents in the collection, without updating a central system catalog, updating an ORM, and without taking the system offline. Optionally, schema validation can be used to enforce data governance controls over each collection.
- MongoDB allows you to store data in almost any structure, and each field – even those deeply nested in subdocuments and arrays – can be indexed and efficiently searched.
- MongoDB adds elements to the document model and the query engine to handle both geospatial and time series tagging of data. This expands the type of queries and analytics that can be performed on a database.
- BSON includes data types not present in JSON data (e.g., datetime, int, long, date, floating point, and decimal128, and byte array) offering type-strict handling for multiple numeric types instead of a universal "number" type. Schema validation enables you to apply governance and data quality controls to your schema.
- Query performance in MongoDB can be accelerated by creating indexes on fields in documents and subdocuments. MongoDB allows any field of a document, including those deeply nested in arrays and subdocuments, to be indexed and efficiently queried.
- We observe that even with in NoSQL databases there is a wide variation in performance based on the type of operation (such as read and write). We also observe little correlation between performance and the data model each database uses.
- Casandra is slow on read operations, but is reasonably good for write and delete operations. MongoDB are the fastest two overall for read, write and delete operations.
- Like any application software, NoSQL implementations gothrough changes and thus performance improvements and degradations are likely to happen through these changes.
- Consequently, one will need to compare databases not only at the application design stage but also at regular intervals to enable switching to the most suitable database implementation.
- Tiered software development processes would help isolate the database backend to facilitate such switching when deemed necessary

g) Any other observation / finding that you feel is important for this exercise

- While importing CSV files into all the databases observed that, the MongoDB performance is high compared to PostgreSQL & Cassandra.
 - MongoDB provides the database tool “mongoimport” to import CSV files into DB. For PostgreSQL used the database GUI tool “TablePlus” to import all the CSV files to DB.
 - When performing “WRITE” operation for ORDERS table, the PostgreSQL was not allowing and throwing the unique ID not exist hence we needed to create the Unique index for the ORDERS table then PostgreSQL allowed us to perform the operation. But in MongoDB the unique ID issue not happened and able to perform the operation successfully.
 - There are many limitations in cassandra DB in crud operations.
 - 1) Select: It only allows key column or index column for where condition in select query. Any non key or non indexed column can't be used in where condition.
 - 2) Delete: Only key columns can be used to delete records in where condition.
 - 3) Group by : Group by option only accept as arguments primary key column names in the primary key order.
 - 4) Order by: ORDER BY clauses can select a single column only. That column has to be the second column in a compound PRIMARY KEY.
ORDER BY is only supported when the partition key is restricted by an EQ or an IN.
 - 5) Table creation: Table can't be created without a Primary key. Need to plan all operations to be done in the table so that respective key combination and cluster ordering can be defined, because it has limitations in all operations.
 - 6) Nested queries are not allowed in Cassandra CQL
 - Cassandra DB import was quick and its query retrieval is fast because it works mostly on key attributes but due to so many limitations, it's less user friendly in data extraction and analysis.
-