



Nerds for Nerds

SQL Interview Prep Doc

By [Deepanshu Kalra](#) (Data Engineer, Google, Ex-Microsoft)

Some topics to revise	2
Data pipeline should address	3
Important Internet Sources	3
Doubts	11
SQL Questions from Top 200 Data Engineer Interview Questions and Answers	12

Expected time to read: 1 day - 1.5 day
(excluding practice on HackerRank/other resources shared below)

Some topics to revise

S.no	Topic	Links
0	My Linkedin Posts	<p>Top SQL Topics https://www.linkedin.com/posts/deepanshuk_sql-interview-d-ata-activity-6843425149660758016-HZxA</p> <p>How to answer a question (helped me crack MS and G and FB) https://www.linkedin.com/posts/deepanshuk_strategy-to-answering-the-interview-questions-activity-6842676001684656128-D4UZ</p> <p>Resume building https://www.linkedin.com/posts/deepanshuk_jobsearch-resume-interview-activity-6841760821543079937-eWc7</p>
1	SCD	https://datawarehouse4u.info/SCD-Slowly-Changing-Dimensions.html
2	Joins	<p>https://i.stack.imgur.com/4zjxm.png</p> <p>Also read self joins</p> <p>Practice self join: https://www.w3resource.com/sql/joins/perform-a-self-join.php</p>
3	Physical Joins	https://www.linkedin.com/pulse/loop-hash-merge-join-types-eitan-blumin/
4	Star/Snowflake	https://www.guru99.com/star-snowflake-data-warehousing.html
5	Indexes	<p>https://docs.microsoft.com/en-us/sql/relational-databases/indexes/heaps-tables-without-clustered-indexes?view=sql-server-2017</p> <p>https://www.red-gate.com/simple-talk/sql/learn-sql-server/sql-server-index-basics/</p> <p>https://www.red-gate.com/simple-talk/sql/database-administration/brads-sure-guide-to-indexes/</p>

6	Data warehouse concepts	https://www.1keydata.com/datawarehousing/dimensional.html
7	Practicing sql	HackerRank (SQL) Leetcode (SQL) (Worth paying for premium for sql as many questions are in premium. I took premium for a month) https://pgexercises.com/
8	Complex sql queries	http://www.complexsql.com/complex-sql-queries-examples-with-answers/
9	Dimension Types	https://www.edureka.co/blog/types-of-dimension-table/
10	Cheat sheet	https://intellipaat.com/mediaFiles/2019/02/SQL-Basic-Cheat-Sheet-1.png

Data pipeline should address

- Partial loads (A scenarios where Partial processing of the files or records or any failures of ETL Jobs occurred; to clean up a few records and re-run the job)
- Restart-ability (You have to re-run from a previous successful run because a downstream dependent job failed or reprocess process some data from history. for e.g. We need to run since last Monday or a random date)
- Re-processing the same files (A source issue where they sent multiple files; We need to pick the right records)
- Catch-up loads (In case you missed executing jobs for specific runs and playing catch up; Batch Processing)

Important Internet Sources

- Pivot:
<https://docs.microsoft.com/en-us/sql/t-sql/queries/from-using-pivot-and-unpivot?view=sql-server-ver15>
- <https://www.teamblind.com/post/Facebook-DE-decision-wzQRWoCS> (Do topics from here as well)

- Analytical function
 - <https://www.red-gate.com/simple-talk/sql/oracle/introduction-to-analytic-functions-part-1-2/>
 - <https://www.red-gate.com/simple-talk/sql/oracle/introduction-to-analytic-functions-part-2/>
- Windows function
 - <https://www.red-gate.com/simple-talk/sql/learn-sql-server/window-functions-in-sql-server/>
 - <https://www.red-gate.com/simple-talk/sql/learn-sql-server/window-functions-in-sql-server-part-2-the-frame/>
- Indexes
- Columnstore indexes
- Datawarehouse
 - Star, snowflake
 - Types of dimension
 - Types of facts
 - Modeling of databases
 - OLAP vs OLTP -
<https://academy.vertabelo.com/blog/oltp-vs-olap-whats-difference/>
 - <https://www.imaginarycloud.com/blog/oltp-vs-olap/>
 - <https://www.vertabelo.com/blog/a-unified-view-on-database-normal-forms-from-the-boyce-codd-normal-form-to-the-second-normal-form-2nf-3nf-bcnf/>
- Basics of Redshift
 - <https://s3-eu-west-1.amazonaws.com/cdn.jefclaes.be/amazon-redshift-fundamentals/aws-redshift-fundamentals.html>
 - <https://www.youtube.com/watch?v=TFLoCLXuU0>
 - <https://aws.amazon.com/blogs/big-data/top-10-performance-tuning-techniques-for-amazon-redshift/>
- Working of sql queries
- Already asked in interviews
 -

Glassdoor:

https://www.glassdoor.co.in/Interview/Facebook-Data-Engineer-Interview-Questions-EI_IE40772.0,8_KO9,22_IP3.htm?filter.jobTitleFTS=Data+Engineer

[Must do]

products sales

+-----+-----+ +-----+-----+

| product_id | int |----->| product_id | int |

| product_class_id | int | +---->| store_id | int |

| brand_name | varchar | | +->| customer_id | int |

```

| product_name | varchar | | | promotion_id | int |
| price | int | | | store_sales | decimal |
+-----+-----+ | | store_cost | decimal |
| | units_sold | decimal |
| | transaction_date | date |
| | +-----+-----+
| |
stores | | customers
+-----+-----+ | | +-----+-----+
| store_id | int | -+ +-- | customer_id | int |
| type | varchar | | first_name | varchar |
| name | varchar | | last_name | varchar |
| state | varchar | | state | varchar |
| first_opened_date | datetime | | birthdate | date |
| last_remodel_date | datetime | | education | varchar |
| area_sqft | int | | gender | varchar |
+-----+-----+ | date_account_opened | date |
+-----+-----+

```

Question 1:

What brands have an average price above \$3 and contain at least 2 different products?

Question 2:

To improve sales, the marketing department runs various types of promotions. The marketing manager would like to analyze the effectiveness of these promotion campaigns.

In particular, what percent of our sales transactions had a valid promotion applied?

Question 3:

We want to run a new promotion for our most successful category of products (we call these categories "product classes").

Can you find out what are the top 3 selling product classes by total sales?

Question 4:

We are considering running a promo across brands. We want to target customers who have bought products from two specific brands.

Can you find out which customers have bought products from both the "Fort West" and the "Golden" brands?

- o
- o One table has date and salesamount. Output a table which has both the above columns with cumulative month s sales amount as an additional column
- o Relational data modelling and dimensional data modelling diff
- o how to distribute storage while creating the table

- o if I have a data model which has a lot of dimension how can I simplify it
<https://stackoverflow.com/questions/27690617/star-schema-structure-to-many-dimensions>
- o SCD types. if I have a table which has a lot of attributes column but only few changes frequently how can I capture these changes
- o Diff between oltp and master data
<https://metamug.com/article/difference-between-master-and-transaction-table.html>
- o how can we implement normalization
- o Table Questions
 - Find cumulative sum of values from a table of dept, item and value
 - From same table, find item with maximum value in each dept?
- o Create table of fixtures from below table of countries

Country

Ind

Aus

SA

Result:

c1 | c2

ind | aus

aus | sa

sa | ind

- o INPUT:

Asin day is_instock

A1 1 0

A1 2 0

A1 3 1

A1 4 1

A1 5 0

Output:

asin	start_day	end_day	is_instock
a1	1	2	0
a1	3	4	1
a1	5	5	0

- o There is a list of countries say IND, PAK, CHN, AFG, SRI, BNG. Create a combination of countries with the help of this list using one query
How about IND-PAK & PAK-IND duplicate, this is where people get stuck? Could not arrive at the solution or approach
- o Which range has most visitors
 - TBL1: <start_dt> <end_dt>
 - TBL2: <date> <num_of_visitors>
- o How to delete Duplicate Records from a table considering there is no primary key. For example, consider the table below

id
1
1
1
2
2

- o You have two tables:

A
id
1
1
1
1
1

B
id
1
1

- Select count(*) from A INNER JOIN B On A.id = B.Id [ans] 2 correct is 10
- Select count(*) from A LEFT OUTER JOIN B On A.id = B.Id [ans] 5 correct is 10

- Select count(*) from A RIGHT OUTER JOIN B On A.id = B.Id [ans]2 correct is 10

o You have table i.e. customer with details

cust_id	mem_start_date	mem_end_date
114	2015-01-01	2015-02-15
116	2014-12-01	2015-03-15
120	2015-02-15	2015-04-01
221	2015-01-15	2015-10-01
120	2015-05-15	2015-07-01

- Give me SQL QUERY that can produce list of active customers till date?
- Give me SQL Query that can Produce list of active customers for month of January 2015?

o You have a table i.e shipments_details

Shipments Table:

shipment_id	shipment_date	delvry_date
114	2015-01-01	2015-01-02
116	2015-02-01	2015-02-01
120	2015-02-15	2015-02-16
221	2015-03-15	2015-03-18
120	2015-05-15	2015-06-01

- Give me SQL QUERY that can give produce output to draw graph between DeliveredShipment v/s ShippedShipment for last 7 Days?

o Write a SQL query that can give following output in two columns.

- Count of negative numbers | | Count of the positive numbers

id
1
-1
1
-1
1
1
-1
1
-1
1

o Sum of salaries per department for current and previous month

Dept1	PreviousMonthTotal	CurrentMonthTotal
1	100	2000
2	..	

o

- Complex queries example
 - o Second highest salaried person in each dept – Done
 - o Backfilling problem
 - o Rank – Done
 - o Dense rank – Done
 - o Row number – Done
 - o Running sum – Done
 - o Delete rows in table so that out of duplicate rows only singled value rows are left
 - o DML DDL DQL
 - o Diff between truncate delete and drop
 - o Fragmentation
 - o Types of constraints
 - o Acid property
 - o Diff between temp table and cte, table variables
 - o Which is more efficient? CTE or temp tables?
 - o Recursive CTE – To find the hierarchy levels
- Partitioning of table
 - o <https://www.cathrinewilhelmsen.net/2015/04/12/table-partitioning-in-sql-server/>
- Normalization
 - o Normalization of OLTP
 - o Normalization of star and snowflake schema
 - o http://www.sqa.org.uk/e-learning/MDBS01CD/page_01.htm
- <https://mindmajix.com/data-modeling-interview-questions>
- <https://mindmajix.com/sql-server-interview-questions>
- <https://www.softwaretestinghelp.com/data-modeling-interview-questions-answers/>
- Output Clause
 - o <http://www.sqlservercentral.com/articles/T-SQL/156204/>
- <https://www.codeproject.com/Articles/34372/Top-10-steps-to-optimize-data-access-in-SQL-Server>
- <https://biginterview.com/blog/2014/09/sql-interview-questions.html>
- <https://www.upwork.com/i/interview-questions/sql/>
- General architectural questions around Data Pipelines
 - o <https://medium.com/@mrashish/design-strategies-for-building-big-data-pipeline-s-4c11affd47f3>
- <https://www.agent.media/grow/sql-interview-questions/>

- <https://www.toptal.com/sql/interview-questions>
- <http://www.java67.com/2013/04/10-frequently-asked-sql-query-interview-questions-answers-database.html>
- <https://begriffs.com/posts/2018-01-01-sql-keys-in-depth.html>
- https://www.youtube.com/watch?v=9gOw3joU4a8&list=PL9ooVrP1hQOEDSc5QEbl8WYVV_EbWKJwX
- <https://docs.aws.amazon.com/redshift/latest/dg/c-the-query-plan.html>
- <https://aws.amazon.com/blogs/big-data/top-10-performance-tuning-techniques-for-amazon-redshift/>
- <https://aws.amazon.com/blogs/big-data/amazon-redshift-engineerings-advanced-table-design-playbook-preamble-prerequisites-and-prioritization/>
- <https://365datascience.com/data-architect-interview-questions/>

Doubts

- <https://msbiskills.com/2015/03/22/t-sql-query-gold-rate-puzzle/>
- <https://msbiskills.com/2015/03/25/t-sql-query-the-work-order-puzzle/>
- <https://msbiskills.com/2015/03/24/t-sql-query-consecutive-wins-for-india-puzzle/>
- <https://msbiskills.com/2015/03/23/t-sql-query-the-candidate-joining-puzzle/>
- <https://msbiskills.com/2015/03/25/t-sql-query-normalize-divide-amount-between-months/>
- <https://msbiskills.com/2015/03/23/473/>
- <https://msbiskills.com/2015/03/25/t-sql-query-fruit-count-puzzle/>

SQL Questions from Top 200 Data Engineer Interview Questions and Answers

1. Write a SQL Query to find Max salary and Department name from each department.
2. Write a SQL query to find records in Table A that are not in Table B without using NOT IN operator.
3. Write SQL Query to find employees that have same name and email.
4. Write a SQL Query to find Max salary from each department.
5. Write SQL query to get the nth highest salary among all Employees.
6. How can you find 10 employees with Odd number as Employee ID?
7. Write a SQL Query to get the names of employees whose date of birth is between 01/01/1990 to 31/12/2000.
8. Write a SQL Query to get the Quarter from date.
9. Write Query to find employees with duplicate email.
10. Is it safe to use ROWID to locate a record in Oracle SQL queries?
11. What is a Pseudocolumn?
12. What are the reasons for de-normalizing the data?
13. What is the feature in SQL for writing If/Else statements?

14. What is the difference between DELETE and TRUNCATE in SQL?
15. What is the difference between DDL and DML commands in SQL?
16. Why do we use Escape characters in SQL queries?
17. What is the difference between Primary key and Unique key in SQL?
18. What is the difference between INNER join and OUTER join in SQL?
19. What is the difference between Left OUTER Join and Right OUTER Join?
20. What is the datatype of ROWID?
21. What is the difference between where clause and having clause?
22. How will you calculate the number of days between two dates in MySQL?
23. What are the different types of Triggers in MySQL?
24. What are the differences between Heap table and temporary table in MySQL?
25. What is a Heap table in MySQL?
26. What is the difference between BLOB and TEXT data type in MySQL?
27. What will happen when AUTO_INCREMENT on an INTEGER column reaches MAX_VALUE in MySQL?
28. What are the advantages of MySQL as compared with Oracle DB?
29. What are the disadvantages of MySQL?
30. What is the difference between CHAR and VARCHAR datatype in MySQL?
31. What is the use of 'i_am_a_dummy flag' in MySQL?
32. How can we get current date and time in MySQL?
33. What is the difference between timestamp in Unix and MySQL?
34. How will you limit a MySQL query to display only top 10 rows?
35. What is automatic initialization and updating for TIMESTAMP in a MySQL table?
36. How can we get the list of all the indexes on a table?
37. What is SAVEPOINT in MySQL?
38. What is the difference between ROLLBACK TO SAVEPOINT and RELEASE SAVEPOINT?
39. How will you search for a String in MySQL column?
40. How can we find the version of the MySQL server and the name of the current database by SELECT query?
41. What is the use of IFNULL() operator in MySQL?
42. How will you check if a table exists in MySQL?
43. How will you see the structure of a table in MySQL?
44. What are the objects that can be created by CREATE statement in MySQL?
45. How will you see the current user logged into MySQL connection?
46. How can you copy the structure of a table into another table without copying the data?
47. What is the difference between Batch and Interactive modes of MySQL?
48. How can we get a random number between 1 and 100 in MySQL?

Image up: input, image down: output. Write SQL

100 %					
Results Messages					
	name	address	email	floor	resources
1	A	Bangalore	A@gmail.com	1	CPU
2	A	Bangalore	A1@gmail.com	1	CPU
3	A	Bangalore	A2@gmail.com	2	DESKTOP
4	B	Bangalore	B@gmail.com	2	DESKTOP
5	B	Bangalore	B1@gmail.com	2	DESKTOP
6	B	Bangalore	B2@gmail.com	1	MONITOR

	name	total_visits	most_visited_floor	resources_used
1	A	3	1	CPU,DESKTOP
2	B	3	2	DESKTOP,MONITOR