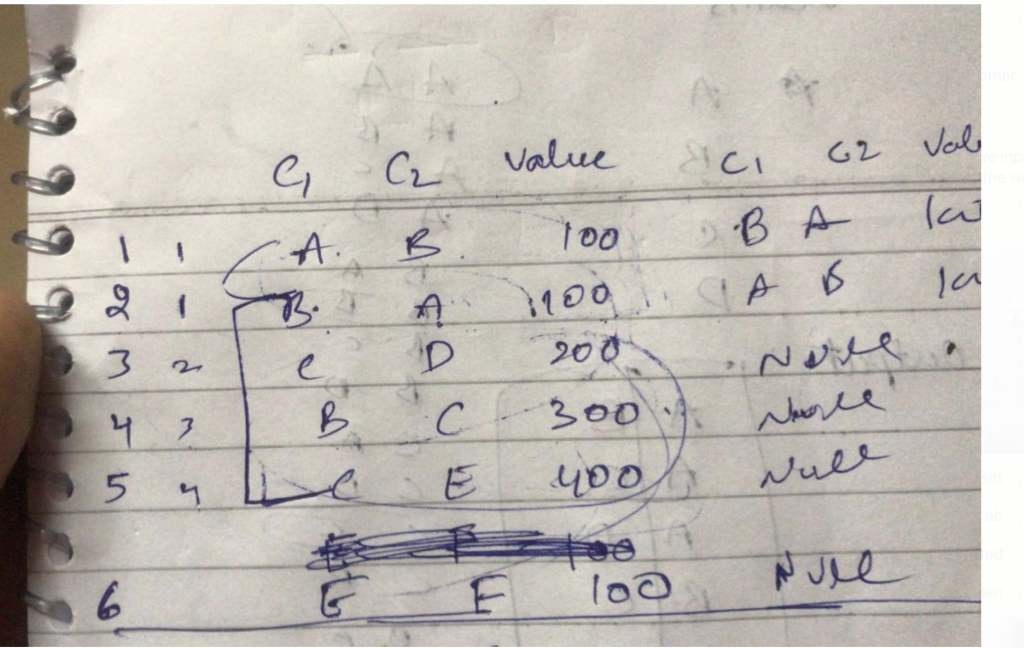
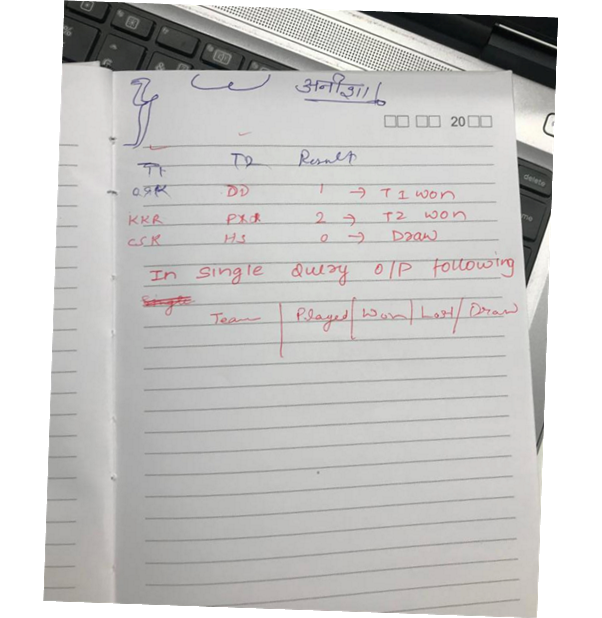
|  |  |  |
| --- | --- | --- |
| S.no | Topic | Links |
| 1. | Indexes | * <https://docs.microsoft.com/en-us/sql/relational-databases/indexes/heaps-tables-without-clustered-indexes?view=sql-server-2017> * <https://www.red-gate.com/simple-talk/sql/learn-sql-server/sql-server-index-basics/> * <https://www.red-gate.com/simple-talk/sql/database-administration/brads-sure-guide-to-indexes/> |
| 2. | Data warehouse concepts | * <https://www.1keydata.com/datawarehousing/dimensional.html> |
| 3. | 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/> |
| 4. | Complex sql queries | * <http://www.complexsql.com/complex-sql-queries-examples-with-answers/> |

**A data pipeline should address these issues:**

* 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)

Topics to study:

* <https://www.teamblind.com/post/Facebook-DE-decision-wzQRWoCS> (Do topics from here as well)
* Analytical function (**DONE**)
  + <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=TFLoCLXulU0>
  + <https://aws.amazon.com/blogs/big-data/top-10-performance-tuning-techniques-for-amazon-redshift/>
* Working of sql queries
* Already asked in interviews
  + 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
  + This is a data set We need to only keep one occurrence of a b So if table has a b and also b a the only one row should be deleted
  + 
  + Diststyle in redshift
  + Relational data modelling and dimensional data modelling diff
  + how to distribute storage while creating the table
  + 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>
  + 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
  + Diff between oltp and master data   
    <https://metamug.com/article/difference-between-master-and-transaction-table.html>
  + how can we implement normalization
  + 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?
  + Create table of fixtures from below table of countries

**Country**

Ind

Aus

SA

**Result:**

c1 | c2

ind | aus

aus | sa

sa | ind

* + 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

* + 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
  + Which range has most visitors
    - TBL1: <start\_dt> <end\_dt>
    - TBL2: <date> <num\_of\_visitors>
  + 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**

* + 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
  + 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?
  + 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?
  + 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

* + Sum of salaries per department for current and previous month

Dept1 PreviousMonthTotal CurrentMonthTotal

1 100 2000

2 ..

* Complex queries example
  + Second highest salaried person in each dept – Done
  + Backfilling problem
  + Rank – Done
  + Dense rank – Done
  + Row number – Done
  + Running sum – Done
  + Delete rows in table so that out of duplicate rows only singled value rows are left
  + DML DDL DQL
  + Diff between truncate delete and drop
  + Fragmentation
  + Types of constraints
  + Acid property
  + Diff between temp table and cte, table variables
  + Which is more efficient? CTE or temp tables?
  + Recursive CTE – To find the hierarchy levels
* Partitioning of table
  + <https://www.cathrinewilhelmsen.net/2015/04/12/table-partitioning-in-sql-server/>
* Normalization
  + Normalization of OLTP
  + Normalization of star and snowflake schema
  + <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
  + <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
  + <https://medium.com/@mrashish/design-strategies-for-building-big-data-pipelines-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=PL9ooVrP1hQOEDSc5QEbI8WYVV_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?