Java

1. What is Encapsulation?
2. What is Abstraction?
3. What is Inheritance?
4. What is Polymorphism?
5. What is Casting?
6. How many objects on the heap if I cast?
7. What is an Interface?
8. Can I instantiate an Abstract class? Constructor?
9. Why would you use an Abstract class over an Interface?
10. equals() vs ==?
11. Can I force garbage collection?
12. Which method does the garbage collector call?
13. What is the finally block?
14. Is a catch block needed?
15. What are Generics for?
16. Comparable vs Comparator
17. Hashtable vs Hashmap
18. How do I start a thread implementing Runnable?
19. Thread methods. (run, start, sleep, wait, notify, notifyAll, join).
20. Difference between the run and start method?
21. What is a deadlock?
22. Which is the fastest File IO available and why?
23. Which are the scopes of a variable?
24. Can you override static methods?
25. What is shadowing?
26. What are Wrapper classes?
27. What is Varargs used for?
28. What is the difference between protected and default?
29. What is the final keyword used for?
30. What is the difference between StringBuilder and Buffer?
31. What is synchronization?
32. How do you go about starting a thread?
33. What is the difference between a List of a Set?
34. Some concrete implementations of Set.
35. LinkedList vs ArrayList
36. How do you insert elements in a Map?
37. Different between Exception and Error?
38. What is Serialization?
39. What is a Singleton?
40. What is the IS-A rule?
41. Where are variable references stored?
42. When is an object ready for garbage collection?

43. What is reflection?
44. What makes the String class special?
45. What is the difference between an Exception and a RuntimeException?
46. Rules of the catch block?
47. What does the Iterable interface do?
48. HashSet vs TreeSet
49. What is a Map?
50. Can I sort a Map?
51. How do I start a Thread extending the Thread class
52. What is Starvation?
53. InputStream vs Reader
54. When is a class fully synchronized?
55. Can an Interface have variables? What are they?
56. What is an Abstract class?
57. What are default methods in an interface?
58. What's the first line in a constructor?
59. What is constructor chaining?
60. What is a short circuit operator?
61. Where are Strings stored in memory?
62. What is hashCode() for?
63. What's the parent of all exceptions?
64. Can I catch an Error? Does it make sense?
65. Is there any case the finally block won't execute?
66. Array vs ArrayList
67. How do I insert elements in a map?
68. What is a Thread?
69. What is a Factory?
70. How do I make an object Serializable?

SQL

1. All SQL Sub Languages.
2. Truncate vs Delete.
3. DATE vs TIMESTAMP.
4. What is Referential Integrity?
5. WHERE vs HAVING.
6. Scalar Functions.
7. INNER JOIN vs INTERSECT.
8. Can I do a sub query in an INSERT statement?
9. What does the 1NF say?
10. Properties of a transaction.
11. What is Atomicity?
12. What is a Cursor? What can I use it for?
13. Which are the main interfaces of JDBC?

14. ALTER vs UPDATE.
15. CLOB vs BLOB.
16. What is a Primary Key?
17. GROUP BY vs ORDER BY.
18. Aggregate Function.
19. FULL JOIN vs UNION.
20. Can I do a sub query in the FROM clause of a SELECT statement? Why?
21. What does the 2NF say?
22. What is Isolation?
23. What is the Read Committed Isolation level?
24. What is a Trigger? When can it execute?
25. What do I need to connect to a DB with JDBC?
26. CREATE vs INSERT.
27. What does the difference between two DATES return?
28. What is a Foreign Key?
29. WHERE vs HAVING.
30. Scalar Function.
31. LEFT JOIN vs MINUS.
32. Can I do a sub query in an UPDATE statement?
33. What does the 3NF say?
34. What is Consistency?
35. What is the Serializable Isolation level?
36. Stored Procedure vs UDF.
37. Statement vs PreparedStatement.
38. DROP vs DELETE.
39. VARCHAR vs VARCHAR2.
40. What is a composite Primary Key?
41. GROUP BY vs ORDER BY.
42. Scalar Function.
43. What is a SELF JOIN?
44. UNION vs UNION ALL.
45. Can I do a sub query in a DELETE statement?
46. What is Normalization?
47. What is Durability?
48. What is a Phantom read?
49. What is a Dirty read?
50. What is a Sequence?
51. Connection vs DriverManager.
52. INNER JOIN vs OUTER JOIN.
53. What is a tetha join?
54. What is a natural join?
55. All types of joins. You can get asked differences between any of them.
56. What is an index?

57. Clustered vs Unclustered index.
58. What is a view?
59. View vs Materialized view.
60. Unary, Binary and Ternary relations.
61. All sections of the SELECT statement.

Business Intelligence

1. What is OLTP?
2. What is OLAP?
3. What is a Data Warehouse?
4. What is the standard structure of tables in a Data Warehouse?
5. What is a KPI?
6. A Data Warehouse should follow which properties?
7. What is ETL?
8. What is a Staging Database?
9. What is a Data Mart?

Hadoop

1. What is Big Data?
2. What kind of data do we have available?
3. Mention a few Big Data problems (Common Hadoopable Problems)
4. Which are the major Big Data Unix distributions?
5. What is Hadoop?
6. Mention a few components that you know of that exist within the Hadoop ecosystem (and their purpose)
7. What is HDFS?
8. What is YARN?
9. Which are the Hadoop 1.x Daemons? (Just mention)
10. Which are the HDFS Daemons? (Explained)
11. Which are the Hadoop 2.x Daemons (YARN MRv2)? (Explained)
12. How does Hadoop achieve fault tolerance?
13. How does Hadoop distribute the processing? What does this mean?
14. What is the standard replication strategy used in Hadoop?
15. What is the default block size of files in HDFS? Can you change this?
16. What is the command to run a jar within a Hadoop cluster?
17. Mention a few HDFS commands
18. What is Data Locality and how does Hadoop apply it?
19. What is speculative execution?
20. How do you set up/implement a MapReduce job?
21. What does the generic types in Mapper and Reducer classes mean?
22. How do you test a MapReduce job?
23. What is a Combiner?
24. How does Hadoop IO work?

25. Which are some of the Hadoop IO classes?
26. Which are the different input formats available in Hadoop?
27. What's the default input format in Hadoop?
28. What is a map-side join?
29. How does sorting work in Hadoop?
30. What is a Hadoop counter?
31. What is a Partitioner?
32. What is a Combiner?
33. What is MR Unit?
34. How do you test a MapReduce job with MR Unit?

Hive

1. What is Hive?
2. What kind of Database is Hive?
3. Does Hive support al CRUD operations?
4. What is the Hive MetaStore and where is it located?
5. Where is the Hive Warehouse located?
6. Which are the different query engines supported in Hive QL?
7. Which are some downsides of using Derby?
8. Which are the Hive data types?
9. What is the difference between a Managed and an External table?
10. How do you create a table with data in Hive?
11. How do you load data into Hive?
12. What if I have the data locally and want to load the data into the table and copy the data to HDFS? (Copy into HDFS)
13. Which are the different INSERT and LOAD into Hive procedures?
14. What happens if the table already has data in Hive?
15. How do you specify within the create command that a table is external?
16. When loading, what happens if Hive finds an discrepancy between the data and the defined schema?
17. How do you export a SELECT statement into a file?
18. What are partitions in Hive? How do they improve queries?
19. How do you specify a partition in the CREATE command and then load data?
20. What is dynamic partitioning in Hive and how do we enable it?
21. What is bucketing in Hive? How do they improve queries?
22. Can you use bucketing without partitioning?
23. How do you specify a bucket in the CREATE command?
24. How do you insert into a bucket or partition?
25. How do you enable Bucketing in Hive?
26. What does strict mode not allow you to do in Hive?

Pig

1. What is Pig?
2. Which are the main components of Pig?
3. Is Pig case sensitive?
4. Is Pig strongly typed?
5. Which are the Pig datatypes?
6. What is the syntax for Pig complex datatypes?
7. Can you actually declare a datatype in Pig?
8. What is the usual execution of commands in Pig?
9. What is the syntax of the load command in Pig?
10. What is PigStorage?
11. Does Pig mandatorily need to run on top of HDFS/MapReduce? If no, which are the different modes?
12. How do you run Pig in a different mode?
13. Mention a few relation operators in Pig
14. How do you foreach through elements in Pig?
15. Can you define your own functions in Pig? If so, which languages are supported?
16. When loading, what happens if Pig finds an discrepancy between the data and the defined schema?
17. What does the history command do in Pig?
18. What do describe, illustrate and explain do in Pig?
19. Pig vs HiveQL vs SQL
20. Can you perform joins in Pig?
21. When performing load and relation operator commands, does data actually get loaded?

Zookeeper

1. What is Zookeeper?
2. Explain the Zookeeper architecture
3. What is the Zookeeper Data Model?
4. Explain the Zookeeper workflow
5. What are some Zookeeper CLI commands?
6. How do you install Zookeeper and manage its lifecycle?
7. What is a Zookeeper Ensemble?

Oozie

1. What is Oozie?
2. Give me some examples of Oozie jobs you could run?
3. Which are some different ways that Oozie jobs can run? (Workflow and Bundle jobs)
4. How do you configure Oozie (the Service and a Workflow)?
5. What is a Control Node in Oozie?

6. What is a Decision Node (Action Node) in Oozie?
7. What is an Oozie Coordinator?
8. What is an Oozie properties file?
9. How do you run an Oozie workflow?
10. Which are the different actions that can happen in an Oozie workflow?

Sqoop

1. What is Sqoop?
2. How do you import data using Sqoop and why do we do it?
3. How do you export data using Sqoop and why do we do it?
4. Can Sqoop perform reduce jobs?
5. What is the default amount of mappers run in Sqoop and how do you change it?
6. What is a Sqoop join?
7. What are some advantages of using a Sqoop job instead of import/export?
8. What happens if the output directory of a Sqoop import already exists?
9. What is an Incremental append in Sqoop? How to use it through commands?
10. What is the Sqoop Metastore? How can you specify the location in a command?

HBase

1. What is NoSQL?
2. Which are some of the major NoSQL technologies?
3. What is the Primary Key structure in a NoSQL database?
4. On the queries Spectrum, what are a few things that need to be taken in consideration?
5. Row vs Columnar Storage (Advantages and Disadvantages)
6. What is HBase?
7. Explain the HBase architecture
8. Explain the table structure in HBase
9. Which are the different modes in which HBase can be started?
10. Is it mandatory to have Zookeeper running for HBase? Why?
11. Drop vs Truncate
12. What kind of permissions can be granted or revoked in an HBase table?
13. Alter vs Delete
14. What does the Enable/Disable command do?
15. How do you interact with HBase as a client?
16. List the available DDL and DML CLI commands available in HBase.

Spark

1. What is Spark?
2. What are some advantages and disadvantages of using Spark over Hadoop?
3. Which are the different kind of Spark Client libraries?
4. Which are the different ways in which you can commuicate with a Spark Master?
5. Which are the different Cluster Managers available for use with Spark?

6. Hadoop vs Spark (Naming)
7. What does RDD stand for? What is it? What are some of its features?
8. What is the default size of an RDD in Spark?
9. How does Spark make RDDs fault tolerant?
10. RDD Transformation vs Action (Mention a few)
11. Which are the most famous programming languages used to develop Spark?
12. Which are a few considerations that need to be taken before running a Spark application?
13. Which are the different ways in which you can run a Spark application?
14. What happens if the output folder of a Spark application already exists?

EMR

1. What kinds of Services does AWS emphasize on? (IaaS, PaaS)
2. Mention different kinds of AWS services
3. What is EC2?
4. Which are the different kinds of EC2?
5. What is Virtualization?
6. Which are the different kinds of Virtualization?
7. Is it possible to reassign a new Key to an EC2 after created? If no, how would you do it?
8. Horizontal Scaling vs Vertical Scaling vs Elasticity
9. What is EMR?
10. Which are the different kind of platforms in EMR?
11. Which are the different launch modes available for EMR? Explain
12. Which are the different kinds of EC2 types available for EMR? Any limitations?
13. What is the command syntax to connect to an EMR cluster?
14. What is the command syntax to transfer a local file into an EMR cluster?
15. Is there any use of having an EMR cluster with no key?

Soft Questions

- Tell me about yourself.
- Tell me about your most recent project.

Problem Solving

1. Calculate the second maximum value of a given array of numbers.
2. Sort an array with bubble sort algorithm.
3. Return all Fibonacci numbers of a given a number.
4. Return the Factorial of a number, recursively.
5. Determine all prime factors of a given number.
6. Rotate the elements of a given array to the right, n given times.
7. Reverse a String without using StringBuilder or Buffer.
8. Check if a String is palindrome (reads the same back and fort).
9. Obtain a substring from an array of characters without using the String class.

10. Find the first occurrence of a given word within another String.
11. Count how many times does each character in a String repeat.
12. Reverse the order of elements in a List without using a helper data structure.
13. Transform a decimal number to binary without using Java API methods.
14. Check if a number is palindrome without transforming it to a String.
15. Calculate the max amount of occurrences existent in a List with repeated values.
16. Flip the order of elements in a Stack using stack methods only (pop/push).
17. Calculate the sum of both diagonals existent in a matrix of any size.
18. Find the element in the middle of a LinkedList in a single traverse.
19. Remove duplicates of a given array of numbers without altering its order.
20. Sort a TreeSet of Integers in descending order.

For SQL, practice and remember syntax for all sub languages statements, aggregate/scalar functions, joins, set operators, and heavily review everything about the SELECT statement (GROUP BY, ORDER BY, HAVING, SUB QUERIES, ETC.) Be ready to write Java MapReduce and any particular syntax of the Hadoop ecosystem.

Don't forget to check the past quizzes to increase the size of your pool of questions, there is a lot more! Go above and beyond, but master what I taught you first.