

69 Cassandra interview questions and answers to hire the best talent

Questions

1. What is Apache Cassandra, and what are its key features?
2. Explain how data is distributed across nodes in a Cassandra cluster.
3. What is a keyspace in Cassandra?
4. How does Cassandra ensure high availability and fault tolerance?
5. Can you explain the concept of a column family in Cassandra?
6. What is a partition key, and why is it important?
7. How does Cassandra handle read and write requests?
8. Describe the consistency levels available in Cassandra.
9. What is a Cassandra node, and what role does it play in the cluster?
10. How does Cassandra handle data replication?
11. What are some common use cases for Cassandra?
12. Explain the purpose of the Cassandra Query Language (CQL).
13. What is a tombstone in Cassandra, and how does it affect performance?
14. How do you monitor and maintain a Cassandra cluster?
15. What are some best practices for designing a data model in Cassandra?
16. How does Cassandra achieve scalability?
17. What is eventual consistency in Cassandra?
18. What are secondary indexes in Cassandra?
19. Can you explain the role of a coordinator node in Cassandra?
20. What is the purpose of the Cassandra commit log?
21. How does Cassandra handle schema changes?
22. What are hints in Cassandra, and how do they work?
23. What are the benefits of using Cassandra for big data applications?
24. How does Cassandra's gossip protocol work, and why is it important for cluster communication?
25. Explain the concept of virtual nodes (vnodes) in Cassandra. What are their advantages?
26. What is the difference between a static column and a regular column in Cassandra?
27. How does Cassandra handle read repairs, and when are they triggered?
28. Describe the purpose and functioning of Cassandra's bloom filters.
29. What is the significance of the partition key in Cassandra's data model, and how does it affect query performance?
30. Explain the concept of lightweight transactions (LWTs) in Cassandra. When would you use them?
31. How does Cassandra handle compaction, and what are the different compaction strategies available?
32. What is the purpose of the system keyspace in Cassandra?
33. Describe the process of adding or removing nodes from a Cassandra cluster.
34. How does Cassandra handle token range assignments, and what is the significance of virtual nodes in this process?
35. Explain the concept of materialized views in Cassandra. What are their limitations?
36. What strategies would you use to troubleshoot performance issues in a Cassandra cluster?
37. How would you handle a situation where a node in the Cassandra cluster is consistently underperforming?
38. What would be your approach to migrating data from an existing database to Cassandra?
39. How would you design a disaster recovery plan for a Cassandra cluster?
40. What are the key considerations when scaling a Cassandra cluster?
41. How do you ensure data consistency in a distributed Cassandra environment?
42. How would you address the issue of tombstone accumulation in Cassandra?
43. How would you approach designing a data model for a time-series application in Cassandra?
44. Can you explain the importance of clustering columns in Cassandra and how they impact query performance?
45. What factors would you consider when determining the partition size in your data model?
46. How do you handle data denormalization in Cassandra, and why is it necessary?
47. What is the impact of the chosen primary key on data distribution and access patterns?
48. Can you explain the concept of data modeling for write-heavy applications in Cassandra?
49. In what scenarios would you use a composite partition key versus a simple partition key?
50. How do you manage data growth in your Cassandra data model to prevent performance degradation?
51. What techniques would you use to optimize your data model for read-heavy workloads in Cassandra?
52. How can you leverage materialized views to improve data access patterns in your data model?
53. What steps would you take to identify and resolve performance bottlenecks in a Cassandra cluster?
54. How do you optimize read and write performance in Cassandra?
55. Can you explain the impact of compaction strategies on Cassandra's performance?
56. What tuning parameters do you consider essential for optimizing Cassandra's performance?
57. How do you handle high write workloads in Cassandra to prevent performance degradation?
58. What tools and metrics do you use to monitor the performance of a Cassandra cluster?
59. How does the choice of partition key affect performance, and what strategies do you use to choose it?
60. What are the best practices for configuring garbage collection in Cassandra?
61. Describe a situation where you had to design a data model for a new application. What factors did you consider in your approach?
62. How would you troubleshoot a scenario where your Cassandra cluster is experiencing high latency? What steps would you take?
63. Imagine you need to migrate a large dataset from a relational database to Cassandra. What strategies would you employ to ensure a smooth transition?
64. If you encountered a node that frequently went down in a Cassandra cluster, how would you investigate and resolve the issue?
65. Picture a situation where the application demands a change in consistency level. How would you assess the impact of this change on data access and performance?
66. What steps would you take if you noticed an unexpected spike in tombstone accumulation in your Cassandra database?
67. How would you handle a request for a new feature that requires significant changes to the existing data model? What considerations would you take into account?