IAE ACADEMY INSTITUTE FOR ENGINEERS

CORE JOBS · INTERVIEW QUESTIONS

IAE Academy free study material · Updated 30 October 2026

Database Management and Warehousing Interview Questions

INSIDE

- 24 questions with short answers
- 6 numericals with the working
- Grouped by topic

Companies do not publish their interview questions; these are ours.

Online, with links to the notes: **iaeacademy.in/study/databases/interview-questions**

FREE COUNSELLING

Preparing for GATE or a core job? Talk to IAE Academy.

Or call **+91 84204 36524**

Every day, 10:30 am to 7:30 pm

Chat on WhatsApp

The entity-relationship (ER) model

1. What is a weak entity, and how is it identified?

A weak entity has no key of its own, so its attributes alone cannot identify it. It is identified by the key of an owner entity together with its own partial key (discriminator). It always has total participation in the identifying relationship, and in a relation it carries the owner's key as a foreign key.

2. What is the difference between a one-to-many and a many-to-many relationship, and how does each become relations?

In one-to-many, each entity on the one side relates to many on the other side, but each of those relates to only one. In many-to-many, both sides can relate to many. One-to-many puts the one side's key as a foreign key in the many side's relation. Many-to-many needs a separate relation whose key combines both entities' keys.

3. Why does total participation affect how a one-to-many relationship is converted to relations?

When the many side takes part totally, every one of its entities has exactly one partner. The relationship can then be folded into that entity's relation as a foreign key declared NOT NULL, so no extra relation is needed. With partial participation the foreign key would hold nulls for entities with no partner, so a separate relation is sometimes preferred.

The relational model

4. What is the difference between the Cartesian product and the natural join of two relations?

The Cartesian product pairs every tuple of R with every tuple of S, giving as many tuples as the two sizes multiplied. The natural join keeps only the pairs that agree on all common attributes and shows each common attribute once. It equals a selection on those equalities over the product, followed by a projection that drops the duplicate columns.

5. Why can a projection in relational algebra return fewer tuples than the relation it starts from?

A relation is a set, so duplicates cannot exist. Once columns are dropped, different tuples can become identical, and the projection keeps one copy of each. If the projected attributes contain a key, no duplicates arise and the count stays the same. SQL SELECT without DISTINCT works on bags and keeps the duplicates.

6. Relation R(A, B) has 12 tuples and S(B, C) has 8 tuples. B is the primary key of S, and B in R is a foreign key to S that is never null. How many tuples does the natural join of R and S have?

The natural join has 12 tuples. B is the key of S, so each tuple of R matches at most one tuple of S, and the non-null foreign key means it matches at least one. So each tuple of R gives exactly one joined tuple.

> **WORKING**
>
> 1. The join attribute is B, so a tuple of R joins with the tuples of S that have the same B value.
> 2. B is the key of S, so each B value occurs at most once in S: at most 1 match per R tuple.
> 3. B in R is a non-null foreign key, so every R tuple has a match: at least 1 match per R tuple.
> 4. Each R tuple gives exactly one joined tuple, so the count is $|R| = 12$.

SQL

7. What is the difference between WHERE and HAVING in SQL?

WHERE filters individual rows before grouping, so it cannot use aggregate functions. HAVING filters whole groups after GROUP BY and can use aggregates such as COUNT or SUM. The logical order is FROM, WHERE, GROUP BY, HAVING, SELECT, ORDER BY, which is why the two clauses see different data.

8. Why does WHERE x = NULL return no rows, even for rows where x is NULL?

Any comparison with NULL evaluates to UNKNOWN, not TRUE. WHERE keeps only rows for which the condition is TRUE, so x = NULL never selects a row. Test for nulls with x IS NULL instead. For the same reason, NOT IN returns no rows when the list contains a NULL.

9. What is the difference between an INNER JOIN and a LEFT OUTER JOIN?

An inner join returns only the row pairs that satisfy the join condition. A left outer join also returns every row of the left table that has no match, filling the right table's columns with NULL. So the left outer join always has at least as many rows as the left table.

Normal forms

10. What is a functional dependency, and what is the attribute closure used for?

An FD $X \to Y$ holds when any two tuples that agree on X also agree on Y. The closure X^+ is the set of all attributes that X determines through the given FDs. If X^+ holds every attribute of the relation, X is a superkey, and a minimal such X is a candidate key.

11. What is the difference between 3NF and BCNF?

In BCNF, for every nontrivial FD $X \to A$, X must be a superkey. 3NF also allows the case where A is a prime attribute. So BCNF is stricter, and every BCNF relation is in 3NF. A lossless, dependency-preserving 3NF decomposition always exists, while a BCNF decomposition may lose some dependencies.

12. Why must a decomposition of a relation be lossless-join?

A decomposition is lossless if joining the parts returns exactly the original relation. Without this, the join can produce spurious tuples that were never stored, so the original information cannot be recovered. For a split into R_1 and R_2, it is lossless when $R_1 \cap R_2$ determines all of R_1 or all of R_2.

13. Relation R(A, B, C, D, E) has the functional dependencies A → B, BC → D, D → E and E → C. How many candidate keys does R have?

R has 3 candidate keys: AC, AD and AE. A is on no right-hand side, so every key contains A, and $\{A\}^+ = \{A, B\}$. Adding any one of C, D or E gives all five attributes, while adding B does not.

WORKING

1. A appears on no right-hand side, so every key contains A. $\{A\}^+ = \{A, B\}$, which is not enough.
2. $\{A, C\}^+$: B from A, then D from BC, then E from D. All five attributes, so AC is a key.
3. $\{A, D\}^+$: B from A, E from D, then C from E. All five, so AD is a key.
4. $\{A, E\}^+$: B from A, C from E, then D from BC. All five, so AE is a key.
5. $\{A, B\}^+ = \{A, B\}$ fails, and each key above is minimal, so the keys are AC, AD and AE: 3.

File organisation and indexing

14. What is the difference between a dense and a sparse index?

A dense index has an entry for every search-key value in the file. A sparse index has entries for only some values, typically one per data block. A sparse index is smaller but needs the data file sorted on the search key. A dense index also works on an unordered field, as in a secondary index.

15. Why are B+ trees preferred to binary search trees for indexes on disk?

Each B+ tree node fills one disk block and holds many keys, so the fan-out is large and the tree stays very shallow, often 3 or 4 levels. Each level costs one block access. A balanced binary search tree needs about $\log_2 n$ accesses. B+ tree leaves are also linked, which makes range scans cheap.

16. A file has 10,000 fixed-length records of 100 bytes each, stored unspanned in blocks of 1024 bytes and sorted on its key. How many block accesses does a binary search on the key need in the worst case?

A binary search needs at most 10 block accesses. Each 1024-byte block holds 10 records of 100 bytes, so the file fills 1000 blocks, and a binary search over 1000 blocks needs $\lceil \log_2 1000 \rceil = 10$ accesses in the worst case.

WORKING

1. Blocking factor: $bf = \lfloor 1024/100 \rfloor = 10$ records per block.
2. Number of blocks: $b = \lceil 10000/10 \rceil = 1000$.
3. Worst-case accesses: $\lceil \log_2 1000 \rceil$.
4. Since $2^{10} = 1024 \geq 1000$, this is 10.

17. A B+ tree uses blocks of 512 bytes, search keys of 8 bytes and pointers of 8 bytes. An internal node holds p child pointers and p - 1 keys, and there is no other overhead in the block. What is the largest p?

An internal node can have at most 32 child pointers. The node must fit in one block, so $8p + 8(p-1) \leq 512$, which gives $p \leq 32.5$. With 32 pointers and 31 keys the node uses 504 bytes, while 33 pointers would need 520.

WORKING

1. The node must fit in one block: $8p + 8(p-1) \leq 512$.
2. Simplify: $16p - 8 \leq 512$, so $16p \leq 520$.
3. $p \leq 32.5$, and p must be a whole number, so $p = 32$.

Data types and data transformation

18. What are the four measurement scales for attributes, and how do they differ?

Nominal values are labels with no order, such as blood group. Ordinal values have an order but no fixed gap, such as low, medium and high. Interval values have equal gaps but no true zero, such as temperature in Celsius. Ratio values have equal gaps and a true zero, so ratios make sense, as with weight.

19. An attribute has a minimum of 20 and a maximum of 70 in a data set. Using min-max normalisation, what value does 56 take when the data are rescaled to the range 0 to 10?

The value 56 becomes 7.2 on the 0 to 10 scale. It lies 36 above the minimum in a range of 50, which is 0.72 of the range, and 0.72 of the new range of 10 is 7.2.

WORKING

1. Min-max formula: $v' = \dfrac{v - \min}{\max - \min}(\text{new}_{\max} - \text{new}_{\min}) + \text{new}_{\min}$.
2. Substitute: $v' = \dfrac{56 - 20}{70 - 20} \times (10 - 0) + 0$.
3. $v' = \dfrac{36}{50} \times 10 = 7.2$.

Data warehouse modelling

20. What is the difference between a star schema and a snowflake schema?

A star schema has one central fact table joined directly to one denormalised table for each dimension. A snowflake schema normalises the dimension tables further along their hierarchy, such as city, state and country. A star needs fewer joins and queries faster. A snowflake removes redundancy and saves space but needs more joins.

21. A data cube has three dimensions, with concept hierarchies of 3, 2 and 4 levels respectively, not counting the top level "all". How many cuboids does the full lattice have?

The lattice has 60 cuboids. Each dimension can appear at any of its own levels or at "all", giving 4, 3 and 5 choices, and the number of cuboids is their product: $4 \times 3 \times 5 = 60$.

WORKING

1. A dimension with L levels can appear in a cuboid at any of its L levels or at "all": $L + 1$ choices.
2. The choices for the three dimensions are 4, 3 and 5.
3. Number of cuboids: $(3 + 1)(2 + 1)(4 + 1) = 4 \times 3 \times 5 = 60$.

Transactions and concurrency control

22. What do the ACID properties of a transaction mean?

Atomicity means all of a transaction's actions happen or none do. Consistency means it takes the database from one valid state to another. Isolation means concurrent transactions behave as if they ran one after another. Durability means that once committed, the changes survive a crash.

23. Why does two-phase locking guarantee conflict serializability?

In two-phase locking, a transaction takes all its locks before releasing any, so each transaction has a lock point. Ordering the transactions by their lock points gives an equivalent serial schedule, so the precedence graph has no cycle. Basic 2PL can still deadlock, and only strict 2PL also avoids cascading rollbacks.

24. What is the difference between a dirty read and a non-repeatable read?

A dirty read sees data written by a transaction that has not yet committed and may still roll back. A non-repeatable read happens when a transaction reads the same row twice and gets different values because another transaction committed an update in between. READ COMMITTED stops dirty reads but allows non-repeatable ones.

FROM IAE ACADEMY

Preparing for GATE or a core job?

IAE Academy teaches GATE in seven papers in Ballygunge, Kolkata, and online. Counselling is free: ask us about batches, fees or your preparation.

FREE COUNSELLING

Chat on WhatsApp

Call +91 84204 36524

Open every day, 10:30 am to 7:30 pm.

Scan to chat on WhatsApp

GATE COURSES

Classroom batches in Ballygunge and online batches. Fees and details are on the course pages.

Scan: courses

See the courses

FREE ON IAEACADEMY.IN

Notes and solved previous-year questions for Database Management and Warehousing, topic by topic.

Scan: the notes

Open the notes

Found this useful? Send it to a friend preparing for core-job interviews.

Send to a friend on WhatsApp

IAE Academy free study material. Companies do not publish their interview questions; these are ours. Prepared and published by IAE Academy, Kolkata. Free to read and free to share, unaltered and with attribution.

IAE ACADEMY INSTITUTE FOR ENGINEERS

GATE coaching in Kolkata since 2012 · EC, EE, IN, CS and IT, DA, BT and CE · in the classroom and online

227A Rash Behari Ave, Ballygunge, Kolkata 700019

+91 84204 36524 · iaeacademy.in · Updated 30 October 2026

Get directions