

Affiliated Distributors Screening Questions

1. *During a database migration how would you handle data loss?*

The first step is always prevention. Database migration should be a gradual process that is performed in segments, planned out by the administrator. Databases must consistently be tested to check for health issues and proof data security. If the administrator fails to perform this process correctly, there may be data loss.

In the event of data loss, an administrator should always have a recovery plan and backups of the database that are up-to-date. The reason behind the loss *must* be studied, whether that be problems with server connections, hardware issues, or software issues. If backups of the data are available (and usable), data needs to be restored using one as soon as possible. If necessary, the administrator must contact shareholders, vendors, and other sources of information to communicate losses and discuss further options.

In special cases, certain shareholders and vendors may have some access to backups or logs which can be used to reconstruct missing data.

2. *We need to build a new database for our employee records. How would you define the system storage requirements?*

Depends on the amount of data being stored (volume), the format of the data being stored, the growth projection of the company, indexes, and the database schema (data blueprint). A single employee may take up to only a few megabytes of memory, but this amount quickly accumulates as more are added. The database should be several times the size of the maximum projected size of the data being logged to account for growth and change. Because databases are mostly unstructured data, they can get *quite* large (in the tens to hundreds, if not thousands of terabytes large).

Assuming the company has 5,000 employees and each employee has basic information (picture, employee ID, first and last name, department, salary, etc.), the system storage requirement may be rather "small" for a database. You will want to add more employees or employee information in the future, so you want to increase storage requirements by 20-30%. This will also leave space for indexes. Then there's backups, which will be exact copies with the *full storage* size of the database. For each backup, the storage requirements increase by 105-110%.

3. *Tell me about your process for troubleshooting database problems*

Immediately pay mind to any error messages that pop up, as they usually detail exact problems. Run through the server logs for more detailed information to check for connection issues, timeouts, low ping, variable errors, etc. Prepare to use a test environment to experiment with fixes.

First check for connection issues. Ensure that the connection to the server isn't being stopped by a firewall or VPN, then check to see if the server is actively running. If you are not signed in or do not have the right permissions, the server may deny you access. Ensure that permissions are configured correctly and that no user has unauthorized access. If all is correct and the connection is established between a client and the server, the server may be stopped or has crashed. Restart the application and / or the client computer.

If that fails, use database monitoring tools to check CPU usage and storage details. If the CPU usage is high, use EXPLAIN ANALYZE to look through the server's plan. Inefficient queries must be optimized (unnecessary JOINS, concatenation, etc. will slow the server down). Make sure that all indexes are efficient and precise. Inefficient querying tends to be the problem.

In rare instances server settings can be the culprit. If needed, tinker with the timeout settings to decrease the chance of hanging, and tinker with memory parameters to allot additional performance to the application. If this does not work then there may be a data integrity issue, data loss, or data corruption, which can be

detected with a keen eye or by using a CHECK TABLE function. Use a backup database to restore the server to a previous point and use records to restore or update missing data.

4. What measurements would you take to protect our databases from external threats?

Firewalls and VPNs are a must. In the past I have come across servers connected to an organization that were not protected by VPN, which shows the importance of it; *an Azure SQL server will be accessible by all who can connect to it if a firewall or VPN is not required*. Luckily these protections are relatively simple to set up, as Windows has built-in functionality for this in its settings. Unauthorized applications should never be able to connect to the database, so an API can be used to prevent such from happening.

A subnet will isolate the server from the cleartnet— the server should stay on the subnet under most, if not possible to make it all, circumstances. Most school and university servers are easily accessible because they are not placed on a subnet and can be accessed in Microsoft Azure with just a student / organization email, which shows why this is necessary.

Ensure that all who can use the database have the proper permissions and that those who should not be allowed access do not. Users must have unique passwords. If Microsoft Azure is used then the Authenticator app will enforce Multi-Factor Authentication, which can and should be used.

5. What types of databases do you work with?

I have worked mostly with relational SQL databases as a shadow under other analysts and administrators to extract, analyze, and present data to shareholders and managers for inventory and procurement. I have plenty of experience with graph, time-series, and object-oriented databases as well, as I have had to work with Microsoft PowerBI, Sharepoint, and other programs. I have been working with Amazon RDS, Oracle, and Microsoft Azure SQL Server for 3 years, and have been working with Microsoft Excel for over 5 years.

6. What is SQL Agent?

SQL Server Agent is an automation tool that will automate tasks for you and optimize your database's performance. Usage of the agent makes the job of the administrators easier.

7. What is DBCC?

Database Console Commands, part of Microsoft SQL Server. These allow for diagnostics and extended operation of the server and are mainly used for debugging or troubleshooting.

8. Explain what a system database and a user database are

System databases in SQL Server are the core databases that come with any server, all which are visible upon login: *master*, *model*, *msdb*, and *tempdb*. Each has its purpose, with *master* being the account and system configurations manager, *model* being the “parent” database which establishes the format of any new databases created by the user, *msdb* keeps record of backups, restores, jobs and alerts, and *tempdb*. Each may have functions within them.

User databases are databases the user may create within the system databases to manage their objects, charts, queries, indexes, etc.

9. What are the operating modes in which Database Mirroring runs? What are the differences between them?

High performance, high safety, and high availability. As database mirroring creates a copy of the database on a different server, each has key differences.

High performance prioritizes system and server performance but increases the risk of data loss if the original server fails. High safety prioritizes data integrity and minimizes defects / data loss, but decreases performance (reading and writing data may be slower). High availability ensures that a copy of the database will always be accessible to read from, but since three servers are used performance will be impacted.

10. Explain the purpose of a model database

Model databases are the parent / template for all user databases created within the server. Whenever CREATE DATABASE is used, the contents of *model* will be copied over. This reduces the time and effort required to create running databases and decreases the risk of errors.

11. What is the difference between a data processor and a data controller?

A data controller directs the data's usage much like a traffic regulator or a flagman. They control how data can be used, and sometimes may be able to control when it can be used. Inventory managers, supervisors and senior analysts may act as data controllers because they regulate what can be used and is presented to those who need the data.

A data processor strictly processes data where the data controller gives them the okay to do so. They do not have control over the data and are restricted to what the controller allows. Most analysts will fall into this boat.

12. Explain what are the Right of Access requests

A Right of Access Request is a request made by an individual to get a copy of their personal information from an organization which withholds it within a short amount of time, usually under a month. For example, if an employee wants a copy of their company identity information (employee id, department id, etc.) they need to make a Right of Access request to get a copy of the information they require. Snapchat has a feature which allows one to receive a copy of data using an online Right of Access request form.

13. Do all businesses need a Data Protection Officer (DPO)?

Only when a business is a public authority, monitors people on a large-scale (i.e. internet service providers), or withholds sensitive information that may fall under protection laws such as OSHA or HIPAA are they required to have a DPO. Otherwise it is not required.

Most small businesses do not need DPOs.

14. Can anyone access the personal data within your company? Or are there different

Incomplete question, but I will answer what is visible. Personal data should *never* be accessible by everyone within the company. Anyone who needs the data must either have a username and password to access the data themselves, or make a Right of Access request to receive a copy. Giving them access poses a massive security and integrity risk.

15. How can you collect data (by email, activity tracking, etc.)?

Data can be collected through almost any medium available, including communication. If the question is how I specifically am able to collect data, I can do so through E-mail, analytics, Microsoft Excel, Microsoft PowerBI, management systems, order numbers, any program which provides a number and / or statistic. I have created forms in the past which users could feed values into, and those forms would spit out information I need. I am also great at communication and have no problem speaking with an associate , a client, or a vendor. I can use activity tracking to collect data as well, though it can be intrusive.

16. How did you learn about new applications / resources?

To give a brief answer, I spend a lot of time researching and implementing solutions into my own work because programming is a hobby to me. Learning one thing is the gateway to learning about another, as there are similarities between each of them.

GitHub is extremely useful for looking at and building upon other people's knowledge, and some GitHubs allow you to play with and experiment with their components. Shadowing others has allowed me to watch someone find a solution which I can hopefully use myself. There are several websites available that you could use for the learning of different applications, with tech forums being some of the best for complex questions. Even Microsoft has several articles for learning the basics of database management on their site.

17. What is the most challenging project you've worked on? Why was it challenging and what was your role?

As a Process Lead I was one of the three people placed in charge of implementation, migrating and translating the vast datasets of an Amazon facility's SAP HANA into Amazon RDS (switching data from a physical server to a cloud-based server). It was challenging because it was a month-long project plagued with hardware limitations and data corruption issues, and it felt neverending because I also had to update inventory and employee records by hand at the end of each day. Unfortunately this is a problem that isn't unheard of as many Fortune 500 companies are old fashioned; I experienced the same thing with Mobis Parts America. The servers of these companies are almost always outdated, their records are handled by a fax machine, and the machines holding the data are usually up to decades old and struggle to handle the large movements of data required during a migration.

18. Can you tell me of a time you made a mistake and how did you resolve it? Was there any way you could prevent it?

I was always taught to use a test environment when experimenting with databases, which is helpful because you will never destroy a database that way. However, I have made a few mistakes when querying (it's quite an easy mistake to make). Sometimes a table has a name which isn't easy to remember or is easy to get wrong and I will make a syntax error, which will usually pop up in an error message or the log. I resolve it by going back and fixing the query.

I have been in charge of fixing a mistake where someone used DROP DATABASE on the wrong user database and deleted it, however. I restored the database with a backup and got with administrators to use data from the facility's local server to replace what was missing. This could've been prevented if the person who made the mistake double checked what they were deleting before performing the function.

19. If you had to work in a team with people with difficult personalities, and constant conflict, how would you manage it?

A team of difficult personalities is always a culture issue. It is important that we try to connect with the difficult associates and seek to understand them. Some people have problems going on within their personal lives and may feel stranded or ostracized. Team-building exercises, managing the workload balance, and setting up workshops often rectifies this issue. Discipline should be a last resort, as everyone who has the job was considered qualified at some point and should not be disregarded.

20. What is your experience working with database servers?

I am a data and inventory associate, so I have worked with and processed information in cloud databases for years. My knowledge of SQL is advanced since I have been expected to make presentations using the information I have been given and have generated professional reports for other departments. I have been lucky enough to shadow database administrators to learn the innards of Amazon RDS and Microsoft SQL Server, and have played with the idea enough to the point where I know how to create and configure a cloud database from scratch. I'm also on my school's Microsoft SQL Server and contribute to it occasionally.

21. What is the highest number of database servers you have worked with?

At one time? 2. Overall I've worked with maybe 11 or 12 medium-to-large active servers.

22. Have you worked with on-premises databases, cloud databases or both?

Both, but I have not worked in a data center. I have more experience with cloud databases than on-premises, with the majority of the experience being with Microsoft SQL Server.

23. Why did you choose database administration?

If I were to put all of my workplace skills together, including hands-on experience with SQL Server, I could serve to be a strong administrator since my skills align with what is required. I have high proficiency in multiple flavors of SQL and am proficient in Python. I have been part of the auditing, migration and implementation teams for Fortune 500 companies from the beginning to the end of their projects. I also have been in the logistics and analytics field for 3 years, meaning I would be great at defining data integrity issues (one of the key parts of being a database administrator) even with large datasets of over 500,000 values. Because of leadership and coordinator experience I have the ability to communicate very effectively. This all serves to make for a well-rounded worker.

24. Describe your workflow without direct supervision

The workflow should be to ensure database health and prevent data loss.

1. Upon login, review alerts, error messages, and logs. Refer to documentation if any issues or discrepancies arise during the initial check; if the issues are not documented, document them.
2. Verify that backups have successfully been completed and are able to be used. If the backups failed, rerun them. If recovery of a backup is possible, do so.
3. Use SQL Server Agent to help check for performance issues. Optimize queries and indexes.
4. Report incidents and performance to the team and management.