

Internal load balancer

10 minutes

In addition to balancing requests from users to front-end servers, you can use Azure Load Balancer to distribute traffic from front-end servers evenly among back-end servers.

In your healthcare organization, front-end servers call business logic that's hosted on a middle tier. You want to ensure that the middle tier is as scalable and resilient as the front end. You want to use a load balancer to distribute requests from the front-end servers evenly among the middle-tier servers. This way, you'll scale out the middle-tier servers to achieve the highest capacity possible. You'll also ensure that the middle tier is resilient to failure. When a server fails, the load balancer automatically reroutes traffic.

Here, you'll learn how to use load balancers to distribute internal traffic.

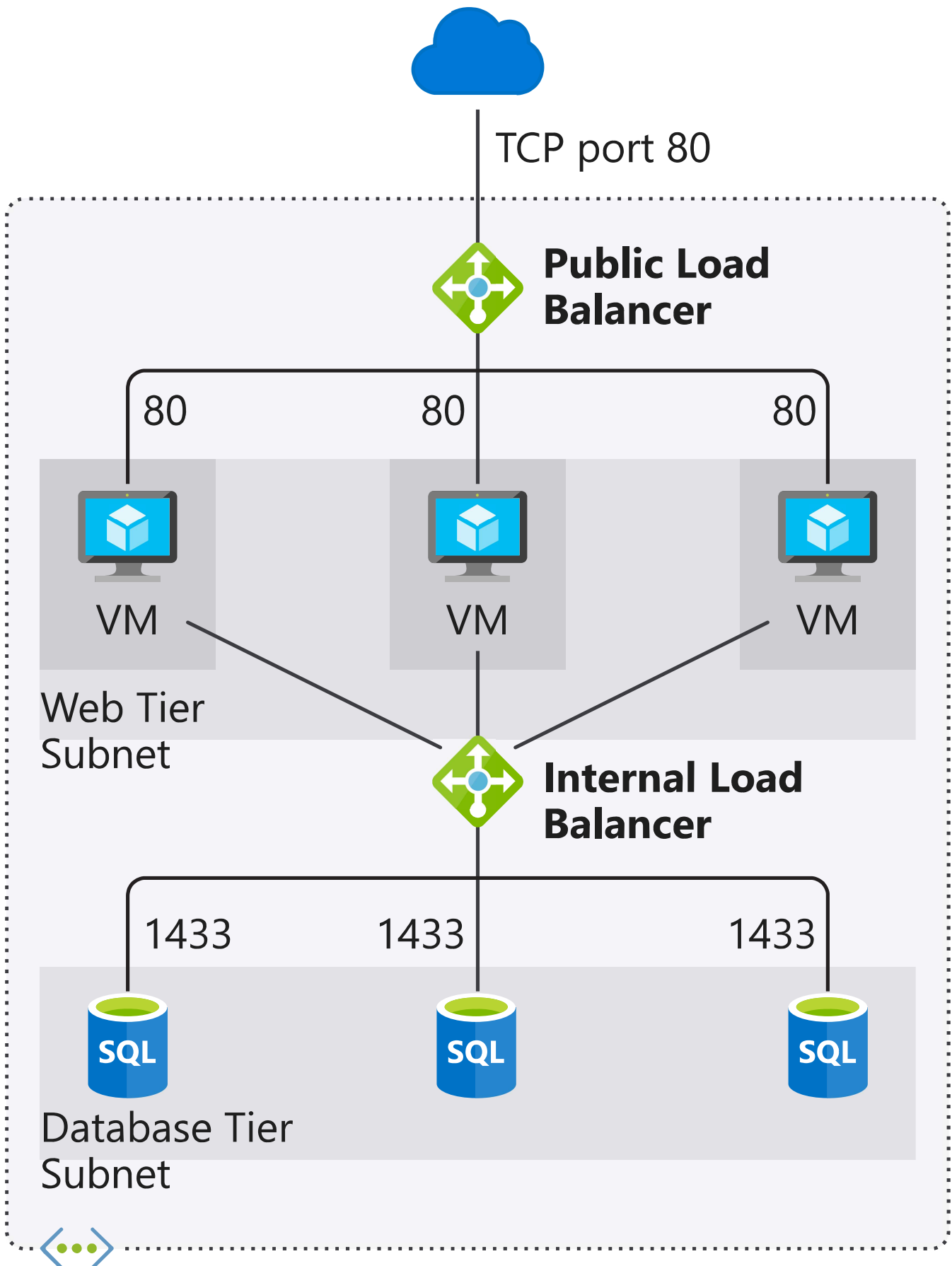
Configure an internal load balancer

In the healthcare portal scenario, a web tier handles requests from users. The web tier connects to databases to retrieve data for users. The database tier is also deployed on two virtual machines. To allow the front-end web portal to continue to serve client requests if a database server fails, you can set up an internal load balancer to distribute traffic to the database servers.

You can configure an internal load balancer in almost the same way as an external load balancer, but with these differences:

- When you create the load balancer, for the **Type** value, select **Internal**. When you select this setting, the front-end IP address of the load balancer isn't exposed to the internet.
- Assign a private IP address instead of a public IP address for the front end of the load balancer.
- Place the load balancer in the protected virtual network that contains the virtual machines you want to handle the requests.

The internal load balancer should be visible only to the web tier. All the virtual machines that host the databases are in one subnet. You can use an internal load balancer to distribute traffic to those virtual machines.



Choose the distribution mode

In the healthcare portal, the application tier is stateless, so you don't need to use source IP affinity. You can use the default distribution mode of a five-tuple hash. This mode offers the

greatest scalability and resilience. The load balancer routes traffic to any healthy server.

Check your knowledge

1. Which configuration is required to configure an internal load balancer?

- ☐ Virtual machines should be in the same virtual network.
- ☐ Virtual machines must be publicly accessible.
- ☐ Virtual machines must be in an availability set.

2. Which of the following statement about external load balancers is correct?

- ☐ They have a private, front-facing IP address.
- ☐ They don't have a listener IP address.
- ☐ They have a public IP address.

Check your answers