SIM NETWORKS A DIVISION OF Netversor

10 Steps

Towards a resilient infrastructure

Let’s Get Started

Table of contents

On-premise server or fault-tolerant infrastructure?

You will learn the following:

- Define your business strategy
- Equipment redundancy check
- Infrastructure monitoring system setup
- Infrastructure accessability check
- Ensure your website is protected
- Choose a certified data-centre
- Time to set up your backup schedule
- Organizing communication channels
- Organize your working environment correctly
- Set up your data replication scenario

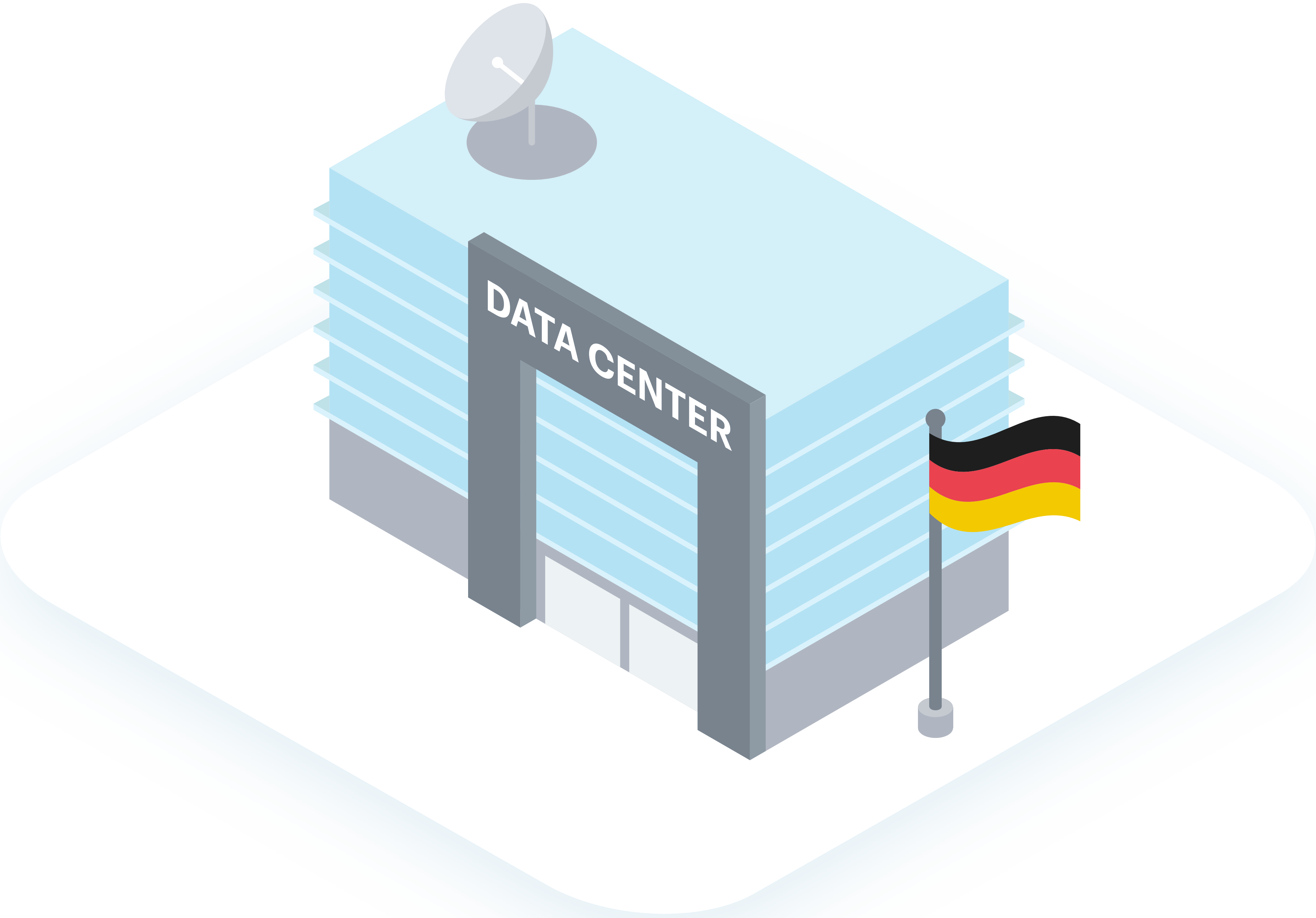
Foreword

Do you still believe that your on-premises server is the best solution for corporate infrastructure?

Try to answer these questions:

- How much does your business lose in 1 hour of downtime due to equipment failure, non-routine maintenance or Internet disconnection?
- How much time do your employees spend on hardware repairs or maintenance?
- Are systematic hardware upgrades cost-effective for your business?
- How easy, fast, and inexpensive is it to scale on-premises servers as your business grows?
- Is your local server protected from fires, floods, power surges, theft, or confiscation?
- Are you and your employees willing put up with the constant noise and heat produced by the server?
- Why is infrastructure rental a better option that on-premises infrastructure?

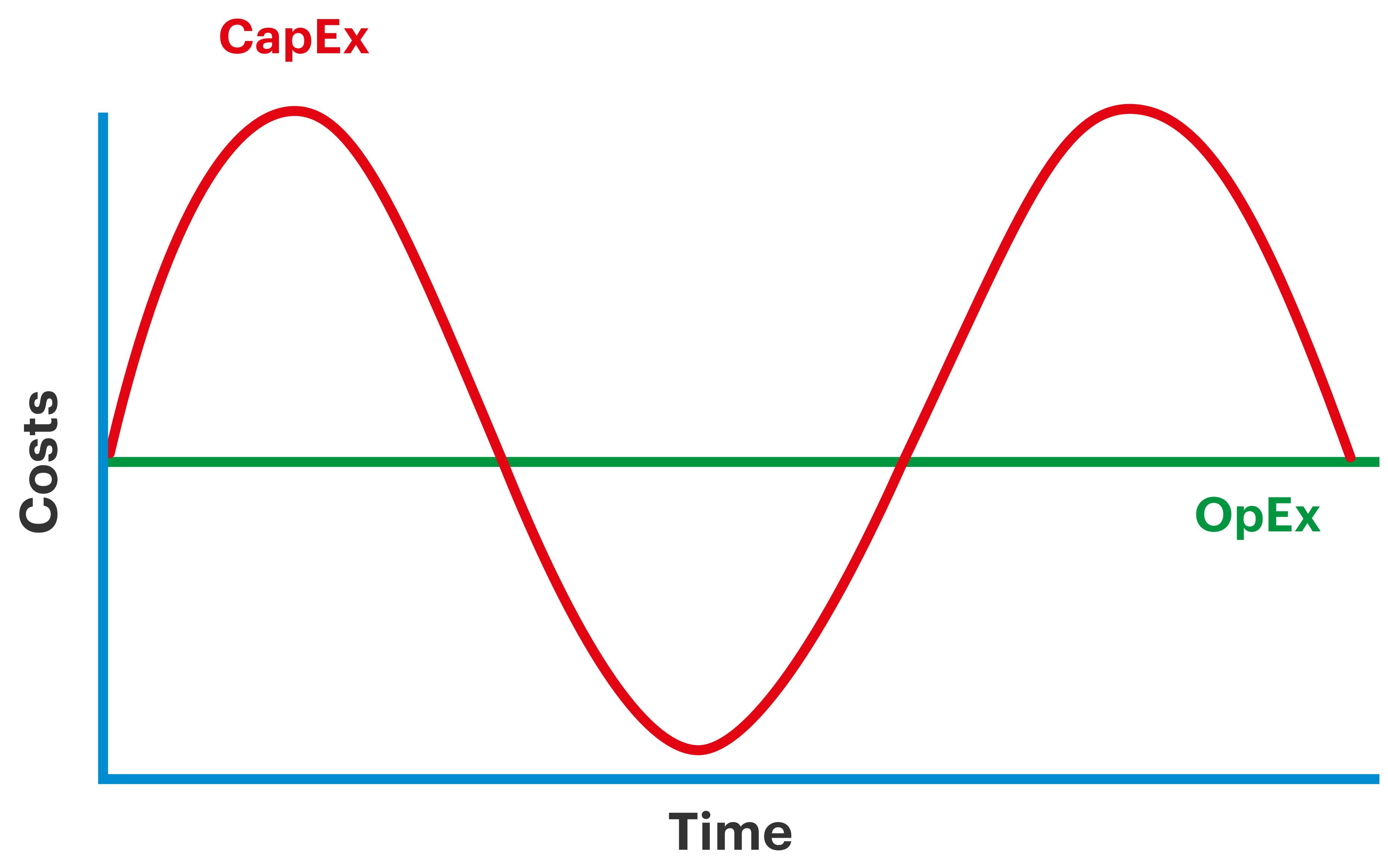
Foreword



Why is infrastructure rental a better option that on-premises infrastructure?

- Renting equipment from a provider is cheaper than buying, maintaining and updating on-premises infrastructure.
- The infrastructure provider ensures high-level information security that complies with international standards.
- Professional equipment capable of working under high loads for years ensures continuous operation and constant availability of services.
- The provider is responsible for equipment replacement and updates, not you.
- A specialized provider has much broader technological capabilities –enterprise-class equipment, connection to powerful communication providers, high-end data centers, top-notch support experts.
- You do not need to spend time and resources on maintenance and support – this is the responsibility of your provider.
- Top-notch infrastructure providers can easily organize a high-availability data center or a redundant server cluster.

Foreword



We will briefly tell you how to create fault-tolerant IT infrastructure.

We assume that you use resources rented from an IT provider, meaning that you know how much more profitable it is:

- to delegate infrastructure management and support to a specialized provider, concentrating on your core business instead;
- to choose operational expenses instead of capital expenses on equipment and the costs of its depreciation, placement, maintenance and support.

SIM NETWORKS A DIVISION OF Netversor

1. Define your business strategy

1. Define your business strategy

Which is better – to rent dedicated server or a cloud solution from a provider?

Both options are viable depending on the specifics of your business. However, keep in mind that you are making a strategic choice that will determine the evolutionary logic of your business IT infrastructure over the next five years.

Regardless of whether you deploy your infrastructure in the cloud or prefer a dedicated server, high availability is necessary. This crucial characteristic directly affects the performance of your business and its profitability.

There are three possible options for deploying corporate infrastructure – in a public cloud, in a private cloud or in a dedicated server. Let's review the specifics and benefits of each option.

1. Define your business strategy

Public cloud

Specifics

- Fast access to all services
- Resource scalability and flexibility
- The provider ensures the workability of the infrastructure
- Guaranteed technological relevance
- The most inexpensive cloud solution
- Advanced data and infrastructure protection

How is high availability achieved?

A well-built public cloud is designed with critical equipment duplication in mind. The microservice architecture of cluster solutions within a cloud allows you to maintain the availability of the entire infrastructure, even if several components fail.

Distributing hardware between remote data centers significantly improves availability. However, not many providers can build their clouds in more than one data center.

1. Define your business strategy

Private cloud

Specifics

The single-tenancy of the private cloud is added to the standard list of benefits of cloud solutions. A private cloud is a custom-tailored solution based on the specifics of your business, on your preferences hardware vendors and other individual needs. You can also install any software, improve data security, organize additional duplication of nodes for redundancy, etc. This solution is more expensive than public cloud rental. However, every aspect of the private cloud's operation depends on your business needs and goals.

How is high availability achieved?

Equipment duplication, if the project allows it.

Deployment in different availability zones, if geographical distribution is feasible for your provider.

1. Define your business strategy

Bare-metal, or dedicated servers

Specifics

Renting a bare-metal server from a provider allows you to choose either a readv-made server configured to handle typical tasks (e.g., SIM-Networks offers 4 standard configurations), or a custom dedicated server, or an entire cluster of servers with components of your choice (CPU, RAM, storage, network infrastructure, etc.).

Single-tenancy: you manage resources as you wish, since you get full root rights.

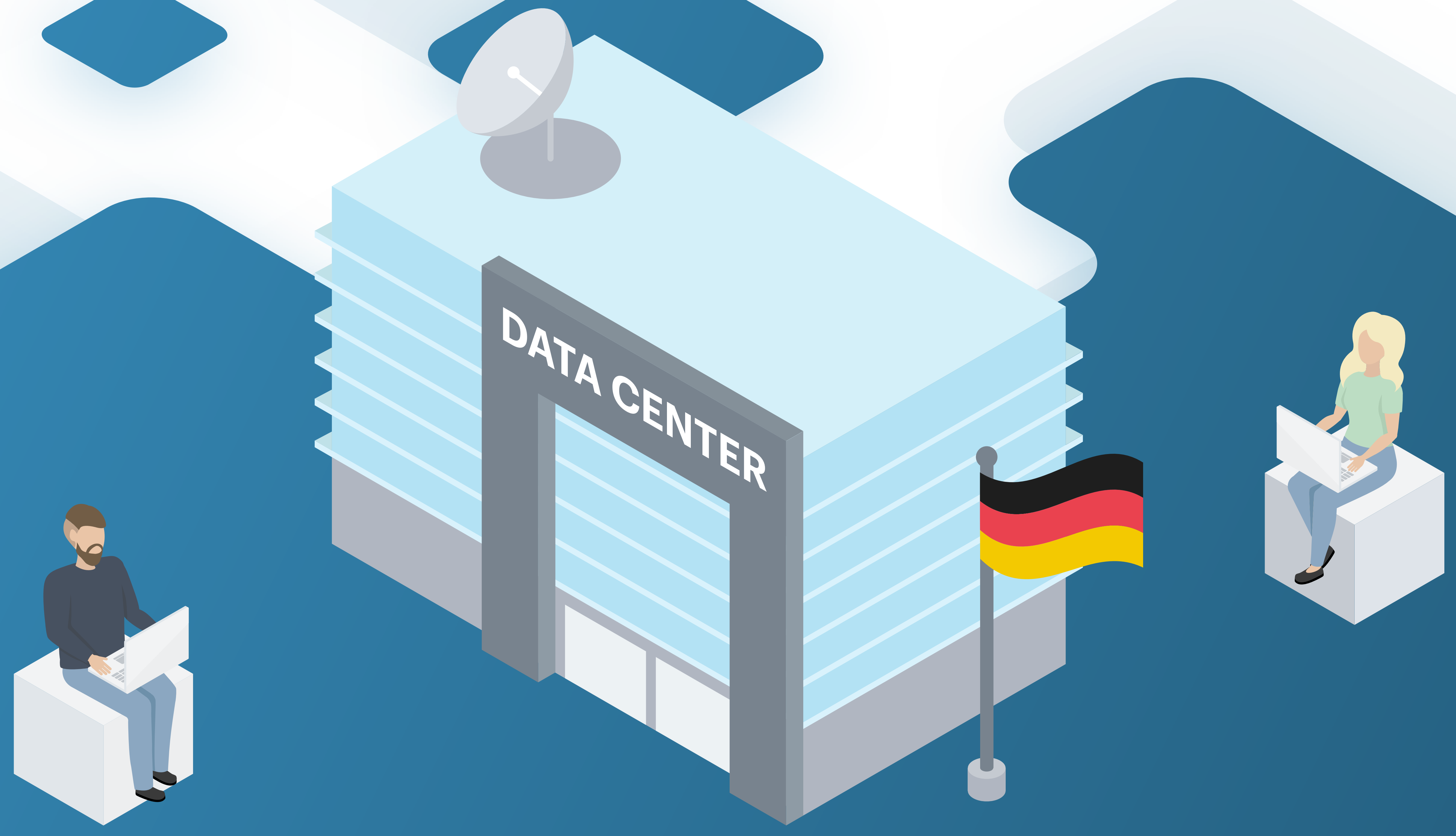
Price: Dedicated server equipment costs a little less than a comparable pool of re- sources in the cloud. But if you want to build a high-available infrastructure with redundancy, the cost of the hardware doubles.

The scalability of a dedicated server is limited by its hardware configuration. When the server's resources are exhausted, you will need to rent another server for flexible scaling.

How is high availability achieved?

Duplication of resources, if the project's configuration covers it.

2. Choose a certified data center



2. Choose a certified data center

The ANSI/TIA-942 standard sets the criteria of fault-tolerance and availability in data centers:

Tier I

Scheduled or emergency maintenance stops the data center's operation. This is due to the fact that Tier I datacenters do not offer redundancy: they do not have a backup power supply system or raised flooring. The engineering infrastructure and communications are not duplicated

Tier II

Short-term power outages do not affect the availability of Tier II class datacenters, but the datacenter must be stopped for repairs. Reservation of infrastructure elements in such a datacenter is organized at the N+1 level: there is a redundant power supply and raised flooring.

Tier III

Redundancy is implemented according to the 2N scheme. This means that Tier III class data centers are highly available and fault-tolerant. All critical systems are duplicated: power supply units, power distribution and cooling distribution channels, communication channels, raised flooring, climate control systems, etc. Maintenance can be carried out can be carried out without stopping the equipment in the data centers.

Tier IV

The most reliable and highly available class of data centers. Resources are reserved according to the 2(N+1) scheme. This means that system components are duplicated twice. Tier IV data centers never stop due to maintenance or various disasters. There are only a few Tier IV data centers in the world. Renting an area in a Tier IV data center is very expensive since huge investments in infrastructure are required to ensure such a high level of availability and fault tolerance.

2. Choose a certified data center

Choose an infrastructure provider with equipment located in Tier III data centers and above.

These data centers won’t be hard to find. However, there are also so-called Tier III+ data centers that offer an «advanced version» of Tier III.

Tier III+ data centers provide higher security and availability. Data security is one of the most critical components of reliability: the higher the level of protection of customer data, the more you can put your trust in a data center. The criteria for data security systems are thoroughly described in the international ISO/IEC standard. ISO/IEC 27001: 2013 is the most recent version, where all the key conditions to ensure the reliability of the datacenter are provided. The data center must be audited for compliance with this standard every three years.

We recommend that you keep your equipment in a **Tier III or a Tier III+ data center with an ISO 27001 certificate** or its equivalent. This way, you will get the proper availability for your projects.

3. Check equipment redundancy

3. Check equipment redundancy

Equipment redundancy is necessary for the stable operation of your business and data security.

When launching a project, you need to be sure that it will keep working even if one of the nodes fails.

In the cloud, equipment redundancy should be implemented by default. You can get the information about duplication in the cloud on the provider's website or request it from the provider. A trustworthy customer-focused provider with a highly available cloud IaaS will have no issue sharing that information.

Consider the redundancy of critical elements of the cloud to avoid problems if a component fails. When choosing a cloud, ask what duplication scheme the provider uses. **The optimal formula is N+1.**

If your infrastructure is based on a dedicated server, make sure that the configuration includes RAID (Redundant Array of Independent Disks), several hard drives combined into a logical element (array).

The array speeds your work up, increases the safety of your data, or both. This depends on the configuration of the array.

3. Check equipment redundancy

Characteristics and functionality of various types of RAID:

RAID 0 (Stripe)

Maximal performance. Data is evenly distributed across the disks of the array, and disks are combined into one, which can be partitioned into several. Distributed read/write operations significantly increase the speed of work since several drives simultaneously read/write their portion of data. The user has access to the entire volume of drives, but this reduces the reliability of data storage. If one of the drives fails, the array is usually destroyed, and it is almost impossible to recover data. Scope: applications requiring high speeds of exchange with the disk (for example, working with video). Recommended for use with highly reliable drives.

RAID 1 (Mirror)

Several disks (usually 2) work simultaneously on writing, so they completely duplicate each other. Performance improves only while reading. This is the most reliable way to protect information from hardware failures. Due to its high costs, RAID 1 is often used store critical data only. The high price point is explained by the fact that only half of your storage space is available for use.

RAID 10 (1+0)

The combination of the first two types: RAID 0 made from RAID 1 arrays. It has all the speed advantages of RAID 0 and the high reliability of RAID 1. The disadvantage is the high cost of the disk array, since the effective capacity of the array is equal to half the capacity of the drives. The minimal required number (always even) of disks to create an array is 4.

3. Check equipment redundancy

Characteristics and functionality of various types of RAID:

RAID 0+1

RAID 1 from RAID 0 arrays. Almost never used due to a lack of advantages compared to RAID 10 and lesser fault tolerance.

RAID 1E

RAID 10-like type of array for distributing data across disks, allowing the use of an odd number of drives (at least 3).

RAID 2, 3, 4

Various options for distributed data storage with disks allocated for parity codes and different block sizes. They are rarely used due to low performance and the need to allocate a lot of disk capacity for ECC and/or parity code storage.

3. Check equipment redundancy

Characteristics and functionality of various types of RAID:

RAID 5

An array that uses distributed data storage similar to RAID 0 (and combined into one large logical drive) + distributed storage of parity codes to recover data. Relative to previous configurations, the size of the Stripe block is increased. Both simultaneous reading and writing are possible. The advantage of this type of array is that the RAID 5 capacity available to the user only decreases by one drive, However, the reliability of data storage is lower than that of RAID 1. It is a compromise between RAID 0 and RAID 1, providing high speed with good storage reliability. If one drive fails, the data can be restored without losses in automatic mode. The minimal amount of drives within this type of array is 3. The «software» RAID 5 implementations built into the south bridges of the motherboards do not have a high write speed, so their use is limited.

RAID 5EE

An array similar to RAID 5, but in addition to the distributed storage of parity codes, the distribution of spare areas is also used – a hard drive can be added to a RAID 5 array as a spare (such arrays are called 5+ or 5+spare). In a RAID 5 array, the backup drive is idle until one of the main hard drives fails, while in a RAID 5EE array this drive is shared with other HDDs all the time, which positively affects the performance of the array. For example, a RAID 5EE array from 5 HDDs can perform 25% more I/O operations per second than a RAID 5 array of four primary and one standby HDD. The minimal amount of drives for this type of array is 4.

3. Check equipment redundancy

Characteristics and functionality of various types of RAID:

RAID 50

The combination of two (or more, which is extremely rare) RAID 5 arrays in a stripe, i.e., a combination of RAID 5 and RAID 0. This partially corrects the main disadvantage of RAID 5, the low data writing speed due to the parallel use of several arrays. The total capacity of the array decreases by the capacity of two disks. However, unlike RAID 6, such an array can only withstand one drive failure without data loss, and the minimum required amount of drives to create a RAID 50 array is 6. Along with RAID 10, this is the most recommended RAID level for use in applications requiring high performance combined with reasonable reliability.

RAID 6

A solution similar to RAID 5 with higher redundancy. Data is not lost when any two drives fail; the total capacity of the array decreases by the capacity of two drives. The minimal number of disks to create an array is 4. The speed of operation is more or less the same as RAID 5. Recommended for cases where the highest possible reliability is essential.

RAID 60

A combination of two RAID 6 arrays in a stripe. The write speed is approximately doubled relative to the write speed of RAID 6. The minimal amount of drives for creating such an array is 8. Information is not lost if two disks from each RAID 6 array fail.

3. Check equipment redundancy

Characteristics and functionality of various types of RAID:

Matrix RAID

The technology implemented by Intel in its south bridges, starting with ICH6R, allows you to organize two disks on multiple RAID 0 and RAID 1 arrays, while simultaneously creating partitions with both increased speed and increased reliability of data storage.

JBOD

JBOD (Just a Bunch Of Disks) is a sequential combination of several physical disks into one logical one that does not affect performance (reliability decreases like in RAID 0), while drives can have different sizes. Rarely used nowadays.

SIM NETWORKS A DIVISION OF Netversor

4. Set up your backup schedule

4. Setup your backup schedule

Backups are a necessary component of the information security policy of any business.

When you build a highly available and fault-tolerant infrastructure, make sure that your critical data is backed up. By systematically and regularly backing up your data, you protect yourself from data loss in the event of system failures, cyberattacks, human error, etc. If your system loses data due to accidents, you can quickly restore your data from a backup. Your backup archives need to be valid and up-to-date.

The 3-2-1 formula for data backup is widely known – 3 backups are stored on 2 different physical carriers, at least 1 of which is located remotely from the core IT infrastructure. This means that your backup storage must be geographically distributed for the best reliability.

If you need backups in the cloud, you can use a XaaS solution, **Backup-as-a-Service**. It allows you to store backups in a remote datacenter. Many providers offer it as a ready-for-use solution that can be instantly activated, for example, SIM-Cloud BaaS. You can configure this service through the SIM-Cloud Dashboard.

On dedicated servers, backups are implemented using specialized software, e.g., popular Veeam products, and additional data storage. You select and customize backups on bare-metal servers yourself. Standard configurations often include additional storage space for backups.

5. Set up infrastructure monitoring

5. Setup infrastructure monitoring

Monitoring IT systems is essential for effective management of the company's infrastructure.

Its purpose is to constantly collect information about the state of the IT infrastructure as a whole and each IT service, track the changes that occur in them in real-time, and analyze them.

The monitoring system forms a catalog of IT services, determines the indicators of availability and quality of each service, as well as its dependence on other components in the infrastructure. Based on these data, the system calculates the quality indicators of the services. This way, specialists receive operational information about its status on all levels, including servers, storage, network equipment, operating systems, business applications, etc.

It allows them to solve problems quickly and effectively, as well as re-allocate resources when necessary, which ensures high-availability and fault-tolerance of the infrastructure.

The IT services monitoring system helps the company:

- increase the availability of services and applications;
- educe downtime of IT infrastructure components;
- reduce the cost of support;
- carry out proactive problem analysis;
- increase the efficiency of resource usage;
- increase staff efficiency and quality of service.

SIM NETWORKS A DIVISION OF Netversor

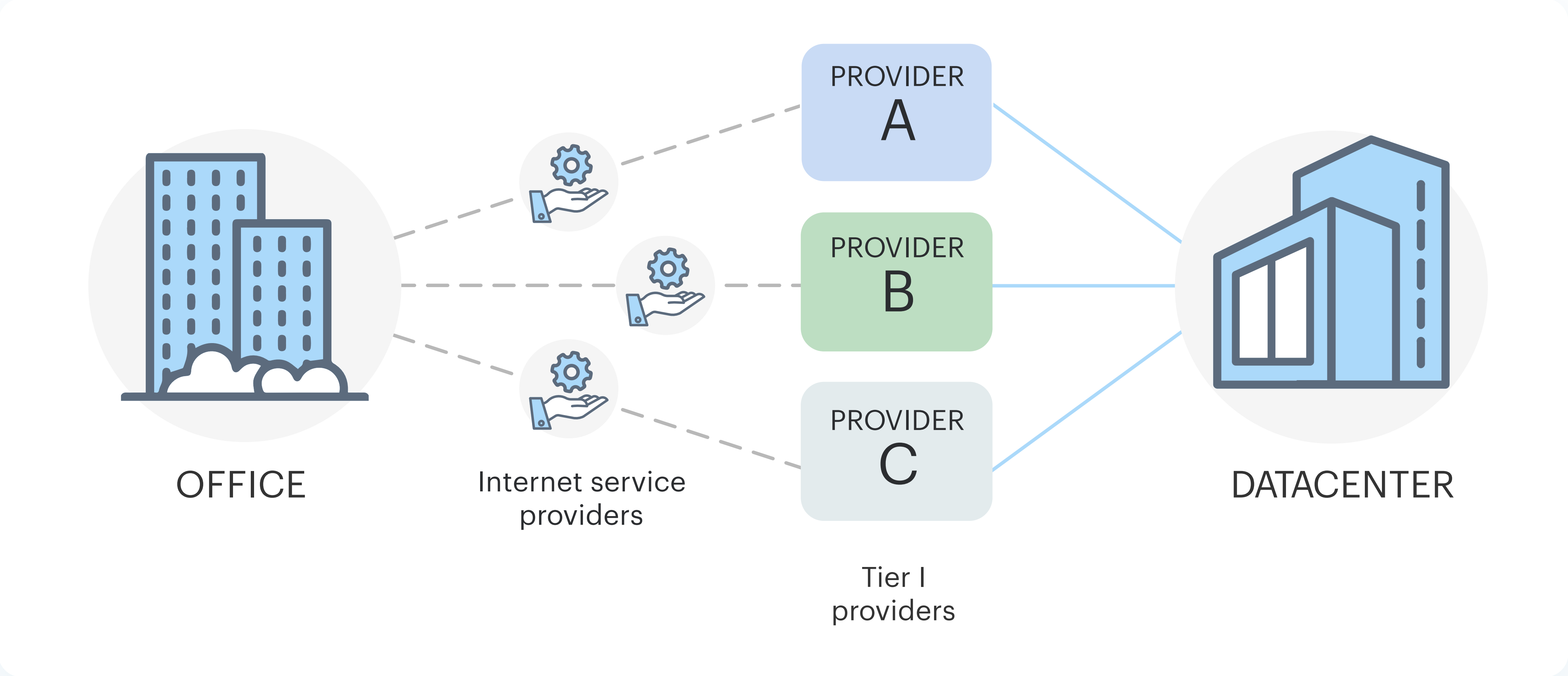
6. Check communication channels

6. Check communication channels

Your infrastructure's availability should not depend on a single communication channel.

If you want to organize a high-available and fault-tolerant IT infrastructure, you need to use additional communication channels.

A good Tier III/III+ class data center is connected to several backbone providers, which means that uninterrupted communication is guaranteed. However, you should connect to two or three providers. This ensures constant access to the Internet for your office environment and for your infrastructure.



7. Check access to Tier-I providers

7. Check access to Tier-I providers

High-availability of your services provides permanent access to the global network.

One more crucial factor in ensuring the availability of your IT infrastructure is connecting to different local ISPs, but also using different lines to connect to the Tier-I backbone.

For example, you need to reach out to the traffic exchange point in Frankfurt, but suddenly the direct communication line to this point is down. So, you need an alternative. Is it available to you?

If your infrastructure solutions provider is responsible for its business, it thinks through workarounds in advance, connecting to several backbones to deliver its customers a high level of their IT resources availability.

Nowadays, there are about two dozen backbone providers in the world (they are also called **Tier-1 operators, or Tier-2 providers**).

There are also regional Tier-2 providers, but it is better to connect to international providers, like CenturyLink, Telia Carrier, etc. Reliable IT system providers have their own backbones with connection to backbone operators.

SIM NETWORKS A DIVISION OF Netversor

8. Organize your workflow

8. Organize your workflow

Your employees' workplaces inside the office and outside it are also elements of this complex, same as the location of your infrastructure.

Ensuring fault tolerance and high availability consists of a whole range of tools, methods, and activities.

The risks of unavailability are not limited by power outages, drops in communication channels or equipment breakdown. There is also the risk that your employees won't be able to connect to corporate systems and business applications due to various reasons.

How to reduce the possible consequences of these risks?

- **Provide a stable electric power supply** in the workplace. Oddly enough, a vast number of companies spend time and money on developing BCP/DRP strategies but neglect to take care of this.
- **Use cheap laptops to get access to the terminal server via RDP** instead of a local network of office computers. In this case, the office only contains a minimal kit of equipment, including some thin clients, a router, and peripherals. This will help you organize your company's work and reduce its sensitivity to road gridlocks or the weather. Secondly, you will make any data within the company's working environment inaccessible to an unauthorized user. Thirdly, you'll save money on the purchase and maintenance of office computers.
- **All applications** containing business-critical data (mail, CRM, accounting, etc.) **should be stored remotel**y and must be backed up.

9. Protect your company’s website

9. Protect your company’s website

No matter the purpose of your site, be it a simple promo page or a full-fledged e-commerce project, its performance is an essential component of your company’s IT infrastructure availability.

First of all, the most reasonable decision is to **deploy your website on a separate virtual machine**. For this purpose, you can either choose a cloud instance or a VDS deployed on your dedicated server. Deploying a site separately from the main infrastructure ensures your corporate IT system won’t be affected by cyber-attacks on your site.

If your revenue depends on the site's performance, **take care of protection against viruses**, spam, and DDoS attacks. The simplest solution for DDoS resistance is cloudflare.com. You will also have to work on your firewall and web server settings, adapting the site to the intended flow of valid requests.

To prevent spam attacks, you should **use patches for security vulnerabilities**. Through vulnerabilities, spam bots can overload your site’s data base. There are many methods – use form fields visible to bots only, captcha/recaptcha, limit the number of form fillings per unit of time, etc.

Sites based on microservice architecture have gained a decent reputation recently. **Microservices reduce the risk of non-working modules causing downtime for the entire site.** If you want to add another layer of protection, place each microservice on a separate virtual machine.

10. Ensure data replication

10. Ensure data replication

Replication is an essential component of the IT infrastructure's availability.

Data replication is a complex mechanism, usually implemented within a cluster architecture. The essence of replication is the following: according to a schedule, data deployed on one cluster node is copied to the backup node in real-time. If the first node fails, the second one picks up its workloads without data loss or downtime. When the functionality of the first node is restored, it receives updated data from the second node and continues to work as usual. Replication keeps the infrastructure constantly running.

In the provider's cloud, replication of the provider’s data is set up by default. It ensures high availability and fault tolerance as a whole for the cloud services. However, data and infrastructure replication of the customer’s project is the responsibility of the client. They can use the capabilities of the cloud to set up replication.

On dedicated servers, replication can be implemented in a similar way: when one of the servers in the cluster fails, full copies of the virtual machines hosted on this server are automatically deployed to the backup infrastructure capacities. For this, you can use solutions developed, e.g., by VMware.

However, it is worth remembering that creating a redundant failover cluster based on dedicated servers can get expensive: the budged for the project will at least double.

Conclusion

Cloud solutions for corporate IT infrastructure will cost much less. Besides, fault tolerance in the cloud can become even more effective if the provider can configure data replication between different availability zones (various data centers).

Now you know that organizing a highly available, fault-tolerant infrastructure for a business is expensive, challenging, and critically important.

We know how to do it quickly, professionally and cost-effectively.

SIM NETWORKS A DIVISION OF Netversor

Thank You

for getting interested

Contact us right away!

Headquarters (Karlsruhe, Germany):

+49 721 781 79601

Customer Care

ask@sim-networks.com