maincubes
SECURE DATACENTERS

WHITE PAPER

NETWORK LATENCY

Network latency is more than just latency rates

Data center interconnects: network latency is more than just latency rates

This white paper on latency was produced in response to questions posed in conjunction with latency-related requests for proposals (RFPs) received by maincubes from prospects looking for a European colocation provider to house their IT infrastructures. Through these data center RFPs we discovered that many companies are not aware of the complexity of 'network latency', or of the consequences of unbalanced and incorrect network latency assumptions on their data center outsourcing decisions.

Table of contents

page

Introduction . . . 3

The cost of fiber optics . . . 4

Client case: remote storage . . . 5

CHECKLIST 1 – Measuring true network latency performance between data centers.
Which parameters have to be taken into account? . . . 6

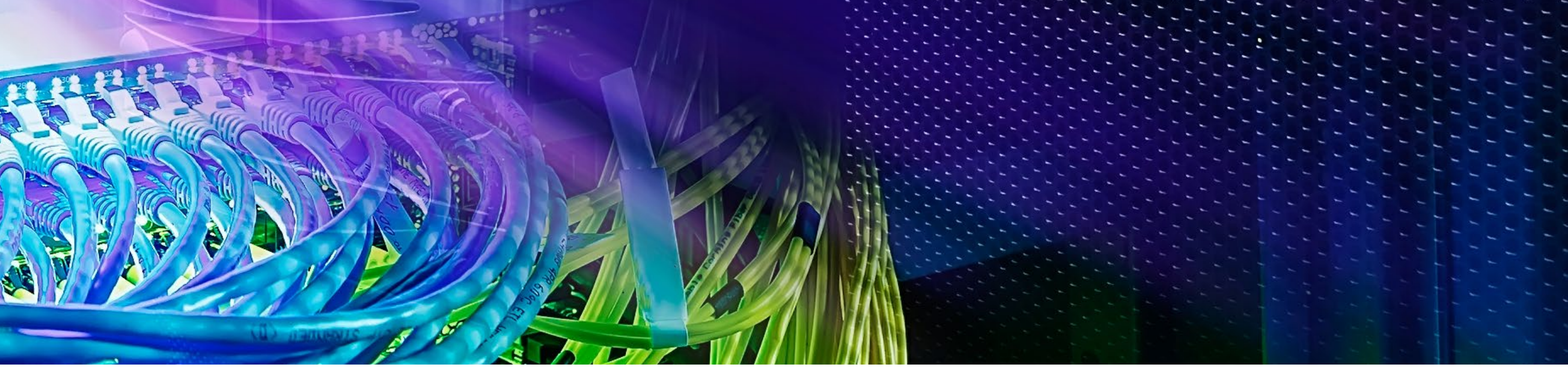
Typical latency figures in metro areas . . . 7

Latency vs. redundancy . . . 8

Cloud connect, storage, gaming, finance . . . 9

Achieving reliable network redundancy . . . 9

CHECKLIST 2 – Ensuring reliable network redundancy between data centers.
Which parameters have to be taken into account? . . . 10



Introduction

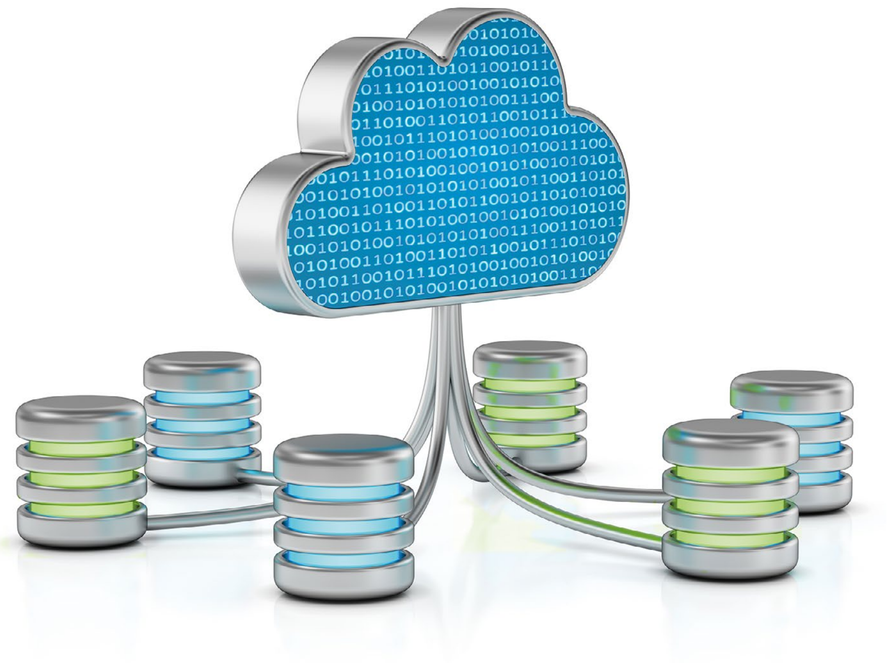
We are a German-engineered data center provider with a keen eye for data center details such as the facilities' security levels, cooling efficiency and infrastructural redundancy, but also data center connectivity – including the quantity and quality of network options available onsite.

Under the influence of market developments such as cloud, SaaS applications, video streaming, the Internet of Things (IoT) and big data, the network latency aspect of data center interconnects (DCIs) is becoming increasingly important.

Organizations globally are looking for 'edge' data centers to fulfill their low-latency network requirements and asking data center providers such as maincubes for 'network latency rates' to ensure the flawless availability of applications in a third party data center environment.

However, network latency is a complex issue that cannot be defined in terms of latency rates alone. In this white paper, maincubes' connectivity experts share their insights into network latency.

Aimed at ICT and business decision makers at enterprises, SMBs, cloud service providers (CSPs), managed service providers (MSPs) and systems integrators, this white paper communicates the business-oriented 'network latency' knowledge which is necessary for a well-informed data center selection, and for 'latency-optimized' as well as 'redundancy-optimized' data center requests for proposals (RFPs).



The cost of fiber optics

The proximity of a third party data center provider can be a consideration when the objective is latency enhancement in combination with network cost optimization because it means short, rather than long (and expensive) fiber optic distances. These short fiber optic distances automatically translate into low-latency customer experiences. It is possible to further enhance latency by taking Ethernet and IP out of the data transmission path, thereby eliminating the need for the data in flight to be processed on OSI Layer 2 and 3. This latency-enhancing strategy simply involves sourcing nearby data centers within a certain geographical radius. What also has to be taken into account is that on redundant routes, the active and failover path never have the same length and therefore never result in the same latency.

Even when third party data center providers are being selected based on proximity to the company's offices, merely asking them to provide latency figures is not necessarily sufficient to obtain a balanced view of latency performance. The issue of latency is more complicated than that. In fact, a series of questions is necessary to obtain a fully balanced picture of the network latency performance being offered by data center and network providers.

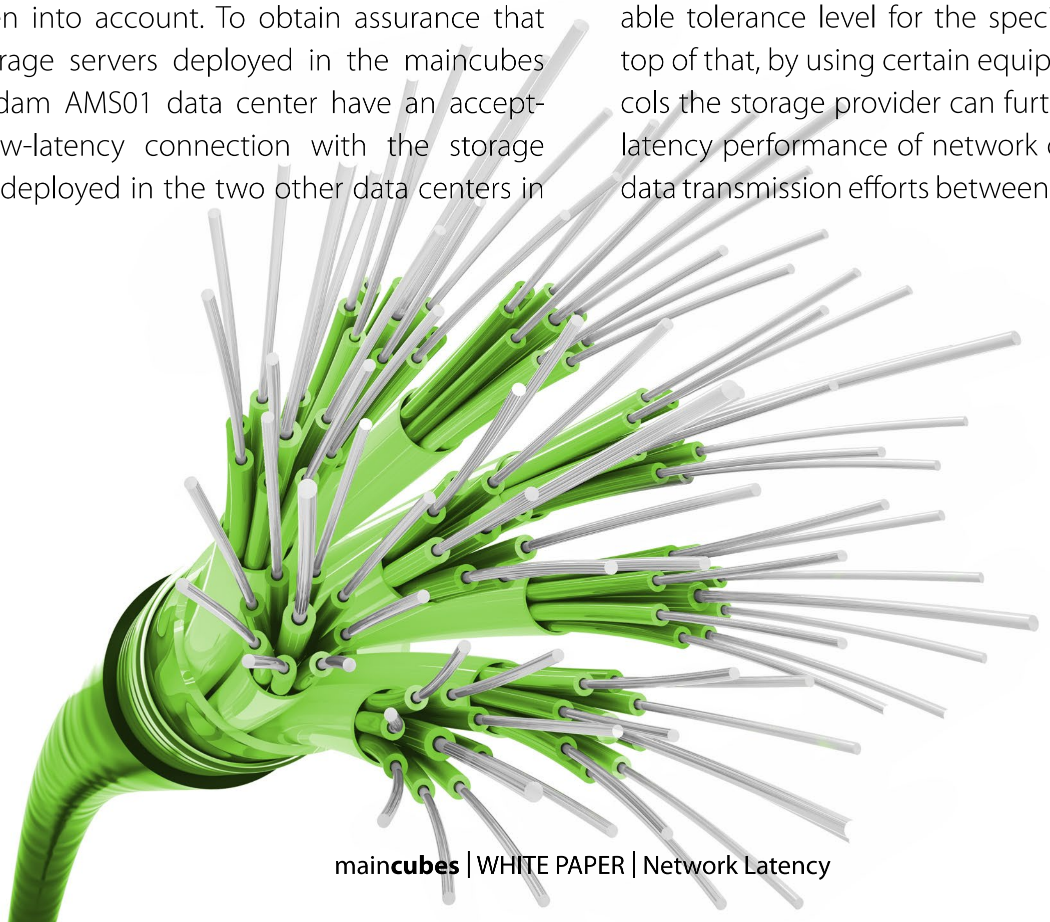
Client case: remote storage

A remote storage provider recently visited the maincubes AMS01 data center in Amsterdam Schiphol-Rijk to ask about the network latency performance of our connection with two other third party data centers (two well-known brands with multiple data centers worldwide) located in the Amsterdam area. One was located 830 meters away and the other nearly 19 kilometers away. The storage provider in question only asked for network latency based on dark fiber and wavelength – decimals with four digits after the dot.

This is an example of a data center customer aiming for an active-active data center architecture who is asking for low latency performance guarantees (since the performance of remote storage is mainly driven by network latency). In the case of this remote storage provider, statistics showed that distance plays an important role in latency performance. Therefore, the connection to the data center located at a distance of 830 meters performed much better in terms of network latency than the other facility.

In addition to the requested information on network latency, other relevant factors must also be taken into account. To obtain assurance that the storage servers deployed in the maincubes Amsterdam AMS01 data center have an acceptable low-latency connection with the storage servers deployed in the two other data centers in the Amsterdam area, it is necessary to take a number of additional connectivity parameters into account. One such parameter is the protocol stack being used. Is the storage customer using fiber channel, iSCSI, or perhaps even a proprietary solution? Specific protocols that operate on different layers or have been designed for a specific purpose can have a significant influence on the latency rate. Furthermore, the question of the transfer rate arises, e.g. 10G / 100G, which influences the serialization delay and thus has a further influence on the total latency.

Therefore, before any conclusion can be drawn on network latency performance, it is essential to know what the client use case is, and what impact factors are involved. Even with the same fiber distance, the same equipment and the same protocols, there will be deviations in the latency times if different wavelengths are used. Although the impact of this is very low in metro applications, it may be significant when ultra-low-latency requirements on long-distance networks come into play (for example international algorithmic trading on different exchanges). Furthermore, latency is never a given fact, so it is necessary to know the acceptable tolerance level for the specific use case. On top of that, by using certain equipment and protocols the storage provider can further optimize the latency performance of network connections and data transmission efforts between the data centers.



CHECKLIST 1 – Measuring true network latency performance between data centers. Which parameters have to be taken into account?

1 What are acceptable limits for standard tolerance levels?

Latency is never a given fact, so the acceptable tolerance level for each specific use case must be defined. Is 5 percent the maximum tolerance level, or would 10 percent also be acceptable?

2 Which interface speed and what technology will be used?

A 1 Gigabit network interface has a different serialization delay than a 10G or a 100G interface. In addition, the use of coherent DWDM or flexible grid technology affects latency.

3 What protocol stack is being used?

Specific protocols can have a significant influence on the network latency. Is fiber channel, iSCSI, or maybe even a proprietary solution being used?

4 What is the real fiber distance between the data centers?

To achieve the lowest latency DCI results the obvious solution is to keep the distance between data centers as low as possible. A distance of 1 kilometer is always better than a distance of 10 kilometers if you are truly looking for the lowest latency rates available. However, the acceptable distance between data centers depends on the actual use case. In addition, geographical distance is not as important as fiber distance. The real distance is measured in terms of the routes taken by the carriers' fiber optic cables between the data centers and not the distance 'as the crow flies'.

This checklist can help to refine your data center request for proposals and add substantiated specifications for the required data center interconnect latency performance in your specific use case.

5 Are the interconnections between the data centers encrypted?

Encryption does add a lot of latency when it takes place on the Ethernet layer, and especially on the IP layer. With L1 optical encryption, very little latency is added. Therefore, if the intention is end-to-end encryption of the interconnection, this factor may have to be taken into account when determining network latency performance.

6 Does the network interconnection have a redundant design?

If intelligent network equipment is being used and one of the two network routes fails, it may automatically redirect to another path. However, these paths will never have the same lengths. It is therefore necessary to establish whether latency will remain within the acceptable tolerance range if this happens.

7 Do I have to optimize wavelength frequency; does every nanosecond count?

Wavelength frequency can matter in certain ultra-low latency applications as mentioned above, so if the latency sensitivity of the application is extremely high, latency can be optimized by using lower wavelengths (as long as the technology in use supports it).

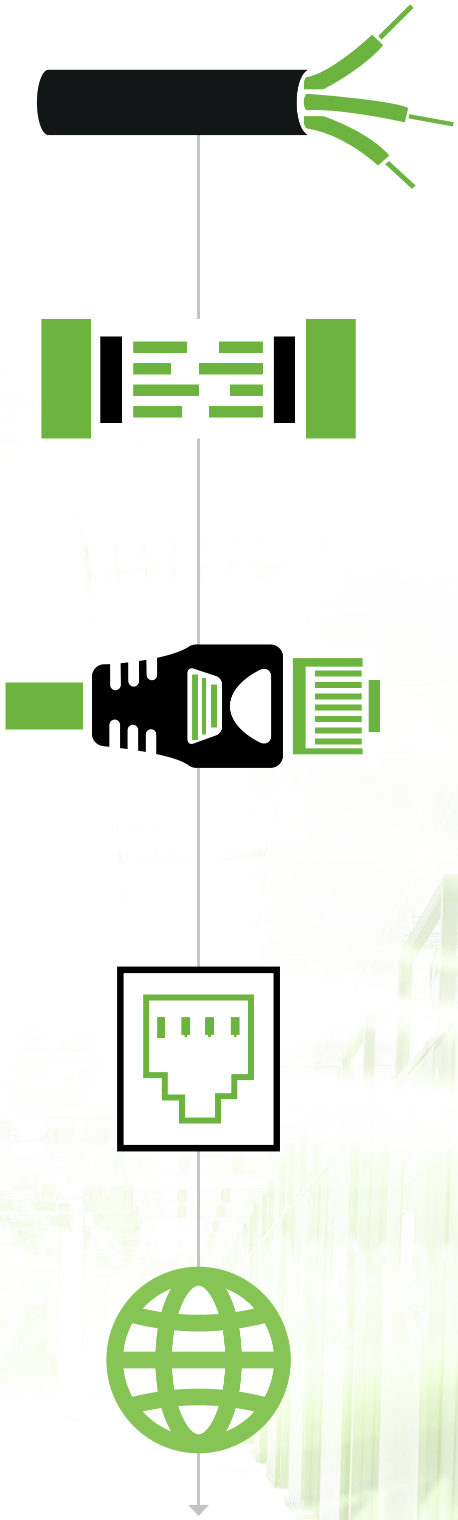
8 Does your network equipment fit the latency requirements?

Your optical and network equipment is the final and, on small distances in certain cases, the most important determinant of latency. Latency can be optimized by

- Keeping the connection 100% optical for as much of the total distance possible
- Using low-latency switching equipment (low port-to-port latency)
- Avoiding excessive routing

Typical latency figures in metro areas

To provide you with some benchmark figures on latency (one-way, not RTT which would be two way), here is an overview of what we consider to be typical latency figures for metro areas in Northern Europe. Be aware that these figures always consist of two components, fiber latency and equipment latency (including serialization delay, which decreases with interface bandwidth). Therefore, these figures cannot be reduced linearly. Also please note that the latency of the equipment inside the data center (e.g. from server to router to network uplink) has to be added on top. Finally, always refer to the provider's SLA and remember that the latency that you can expect will always be lower than the guaranteed latency figures stated in the SLA, in some cases even significantly.



Dark fiber
< 0.5ms per 100km
100% throughput
Guaranteed 99.9xx% SLA

Dense Wavelength Division Multiplexing (DWDM)
< 0.6ms per 100km (no signal regeneration)
100% throughput
Guaranteed 99.9xx% SLA

Ethernet
1-2ms per 100km, depending on equipment and number of devices in the transmission path
100% throughput
Guaranteed 95-99% SLA – depending on equipment and network provider configuration, SLA and user policies

Internet exchange port
1-3ms per 100km, depending on equipment and number of devices in the transmission path
Throughput limited by network provider SLA and user policies

Internet
1-3ms per 100km depending on equipment and number of devices in the transmission path
Fair use policy with practically no throughput or latency guarantees

Latency vs. redundancy

Network reliability and performance do not only depend on the network latency rates between data centers. As mentioned in the latency checklist, network redundancy also plays a role. If a redundant network setup is the objective, another important requirement should be taken into consideration, namely the time it takes for traffic to be shifted to a secondary route after signaling that the primary network route is not working properly. In other words, what would be an acceptable packet loss for your use case?

Real time protocols such as those used for streaming, Voice over IP and gaming are particularly susceptible to packet loss. Other protocols including HTTP and FTP are less susceptible to packet loss. They usually only slow down as the packet losses increase transmission times due to the requirement to retransmit a significant number of packets.

When selecting data centers for an active-active data center setup, long distances between facilities may negatively impact latency performance and cause cost escalation. On the other hand, distances between data centers and the associated latency and cost attributes have to be weighed up against redundancy requirements.

With an active-active connection between two data centers, it can be useful to select fiber optic routes with the same distance. This is because lines of the same length tend to provide similar latency for both lines, which leads to smaller latency deviations between line A and line B. This in turn provides greater stability for the redundantly configured network connection. In addition, network engineers prefer identical, redundant configurations – since this causes fewer network problems, at least in theory.

Cloud connect, storage, gaming, finance

Applications that particularly benefit from low latency configuration include remote storage, cloud storage, virtual machines with storage attached, cloud connects, audio/video rendering, gaming, content delivery networks, ecommerce, as well as financial applications including payment services. Low latency is actually important to all applications that may benefit from the highest I/O (input/output) possible. This includes also practically all applications that use remote storage.

The network latency checklist in this white paper can be used to obtain a more detailed idea of latency requirements and prepare an optimized data center request for proposals.

Achieving reliable network redundancy

DCIs can be seen as lifelines in complex application networks but they also have their pitfalls if the objective is data center interconnect redundancy. Using two different connectivity providers for DCIs is not a guarantee of 100% uptime and business continuity can still be at risk, particularly since some network providers use the same physical infrastructure or dark fiber routes. In a disaster scenario, both the primary and the secondary (backup) connection could be lost.

To ensure true redundancy on your DCI, it is best to start with the optical transmission layer. For example, you can order dark fiber routes with geographically separated paths between data centers. With optical wavelengths, you can also ensure appropriate security through redundancy of the transmission equipment and appropriate switching between the geographically separated paths.

The following network redundancy checklist may come in handy to ensure a reliable DCI and minimize the risk of losing connectivity.

CHECKLIST 2 – Ensuring reliable network redundancy between data centers. Which parameters have to be taken into account?

1 Redundancy requirements

Write down all network and business redundancy requirements – not forgetting to include a clear network diagram.

2 Network overlaps

Request proposals that include digital mapping ('kmz') files to check for possible network overlaps.

3 Route requirements

Validate that the routes being offered match your company's requirements.

4 Penalties < outage

Don't fully rely on the 5-nines availability in an SLA because the penalties vendors pay for failing to meet SLA thresholds are usually minimal compared to the impact an outage will have on your business.

5 Network risk analysis

Many companies do not have direct control over their network and rely on third parties to maintain connectivity, so include data center cross connects, OTDR measurement validation and operating procedures in your network risk analysis.

6 Network providers shortlist

Shortlist network providers and ask for formal offers that are explicit on your redundancy requirements.

7 Requirements in contract

Include your redundancy requirements in the contract with your chosen network providers and have the providers add their network designs as an annex. Include termination clauses in contracts with your network providers, particularly covering their failure to meet your redundancy requirements.

8 Network equipment verification

Verify configuration of optical and active network equipment to accommodate two or more physical fiber connections.

9 Regular audits

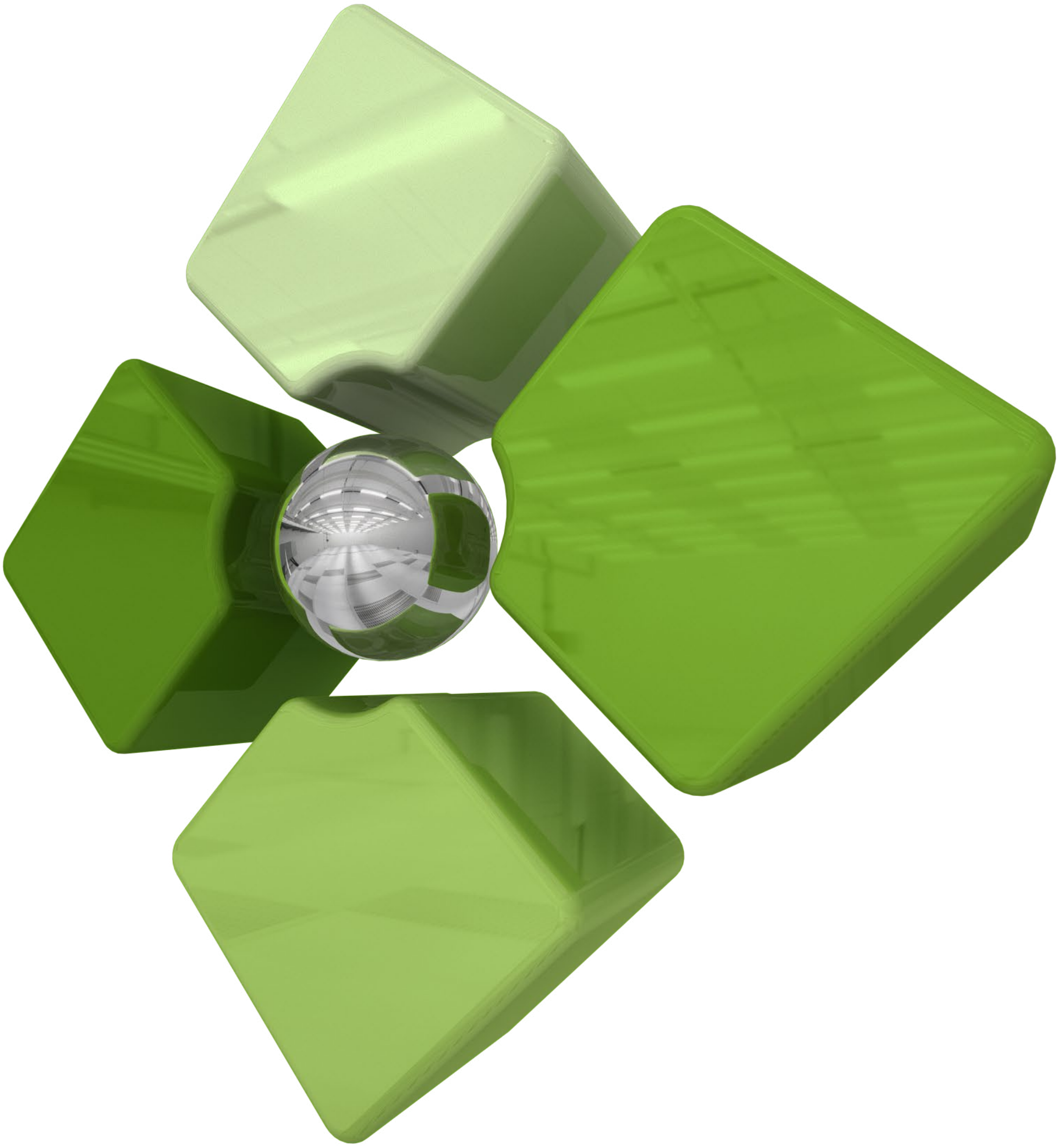
Set up regular audits to verify compliance with agreed network requirements and vendor stability. This is an often overlooked but crucial step if your network has to switch to a redundant path.

This data center knowledge white paper is offered to you by main**cubes**

About maincubes – maincubes is part of German investor and real estate developer Art-Invest which is part of the German construction conglomerate Zech Group. maincubes has data centers in Frankfurt and Amsterdam, and a network of high-availability data centers of various sizes and types in Europe, enabling it to provide colocation services and secure ecosystems for the digital future of customers across various industries. Via the secureexchange® digital platform, customers and partners of maincubes can use IT services worldwide such as IoT, (cyber) security and connectivity as well as cloud services to expand their business opportunities. maincubes offers secure, efficient and user-friendly services – and a secure home for your data.

To learn more about maincubes, visit http://www.maincubes.com

maincubes
SECURE DATACENTERS



STAY UP TO DATE!

Sign up for the maincubes newsletter.

maincubes.com

Are you interested in IT migration, colocation 4.0 and other topics?

You can find other maincubes white papers here!

maincubes one GmbH · Frankfurt am Main · +49 69 6976810-0 · info-de@maincubes.com