FACULTY OF ENGINEERING AND IT

42891 CLOUD COMPUTING INFRASTRUCTURE

Assignment 2: Group Major Project

Group 4

Designing a Cloud Computing Infrastructure

STUDENT NAME	STUDENT NUMBER
PIOUS ALEX	14228389
Mohammadamin Fotouhi	14359537
Jaynil Pandya	14348282
Akriti Srivastava	13820194

Table of Contents

Group Contribution Table

Task No.	Task Name	Contributor
1	Business Requirements	Mohammadamin, Akriti
2	Cloud Architecture and Design	Mohammadamin, Akriti
3	Considerations and Challenges	Jaynil
4	Evolution of Technology	Pious

Business Requirements

The business requirements plan has been devised for the next 30 days of public video streaming.

Functional requirements:

- Provide subscription requirements
- Implemented search services
- Provide streaming platform requirements
- Secure authentication
- Facilitating the video streaming service through CDN and Cache servers
- Controlling the users' access levels
- Managing the video files uploaded
- Editing and processing videos using implemented applications on the server
- Encryption of users' data

Non-functional requirements:

- **Daily active users:** 10
- **Maximum number of videos uploaded by each user:** 2
- **Maximum video file size:** 50 MB
- **Maximum File uploads each day:** 20
- **Speed:** 10 Mbps
- **Channel Bandwidth:** 100 Mbps
- **Total number of users:** 10
- **Hardware resources**
    - **NAS storage capacity needed:** 30 GB
    - **CPU:** Total 8 units of Intel Core i9-10900kf Processors (Preferred)
    - **RAM:** Total of 128 GB
    - **Require IP addresses:** At least 4 Public IP addresses – At least 7 Private IP addresses

- **High availability:** Despite any failure, our services should continue running and video streaming should not stop working.
- **Reliability:** There should be no or minimum downtime or failure in streaming the videos and no interruptions during the upload phase of videos.
- **Scalability:** Our platform might need to be expanded using the same design and implementation. We must be able to increase our user's capacity, bandwidth, speed, storage and other hardware resources anytime.

Cloud Architecture and Design

Hybrid Cloud Architecture:
We have taken the Hybrid cloud architecture design principle for our video service cloud design. The primary reason to utilise the Hybrid cloud is the ease of data availability anytime, anywhere. As video streaming should be accessible from anywhere around the world. For video

streaming structure, cost optimization is highly essential, and with Hybrid cloud will get more security and control over data in a cost-effective way. With increasing automation made possible by this cloud structure, we can modify our cloud settings to react automatically to variations in demand, improving performance and efficiency. Hybrid cloud architectures enhance business continuity while lowering the possibility of downtime and related expenses.

Cost Comparison: Cloud Vs Premises for video processing over cloud
On-Premises:
Considering all of the things, an on-premises solution necessitates funding for:

- Insurance and building space rental/payment
- Facilities for cooling, power, and UPS
- Installing numerous server racks
- Modifications include increasing RAM and memory to each individual server, swapping out weak CPUs for stronger ones, updating software, etc.
- Improvement
- after year five, start again

Over Cloud:
Particularly helpful for new and expanding businesses, the cloud provider manages all usual data centre capital expenditure costs, saving you from having to make the investment or assume the associated obligation. Whereas video service applications on the cloud can involve the cost of labour, CDN and traffic.

Computation, Security and Backup nodes: There are three clusters named Computation cluster, Security and Monitoring cluster and Backup Cluster.

1. The Computation cluster includes the ESXi 1 and ESXi 4 (Backup) Virtual Machines that are handling the main processing of the video streaming.
2. The Backup Cluster includes the cloned version of the computing machines.
3. The Security Cluster contains two machines to control and monitor the traffic.

Storage nodes: We have considered three data stores to cover our needs.

- ESXi 1 Videos datastore
- ESXi 2 Backup and Load Balancing Videos datastore
- ESXi 3 User Database and logs and security
- ESXi 4 Backup Storage

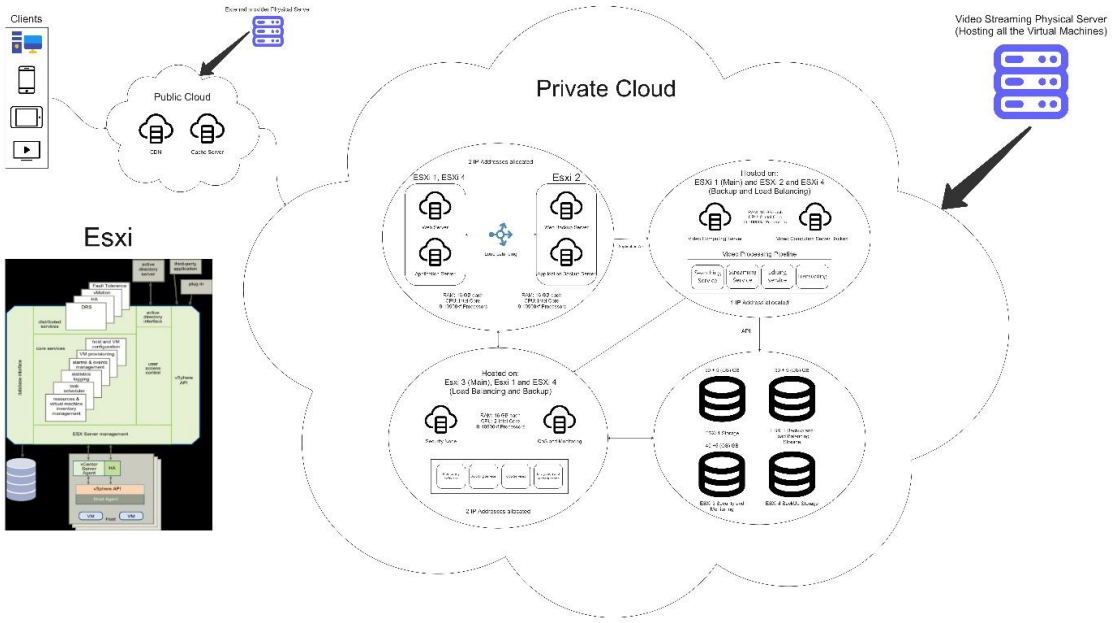

Figure 1. Design Figure

[Original Image]

Video API:
An application programming interface (API) for creating, storing, and delivering video is known as a video API. You can often obtain analytics on the films and live streams you show visitors using several video APIs. A video API has the advantage of being entirely configurable, allowing you to create the API for whatever use case you need.

An individual step in the production and delivery of videos may be the subject of a video API. A video API, for instance, could only support live streaming. Or perhaps it merely converts movies for use elsewhere.

Storage Plan:
As mentioned in the business requirement, video streaming service on the cloud requires at least 30 GB of capacity in case of even 10 active users though this can increase over the period. Therefore, to manage a future request, we must consider a storage plan that can cover storage to 50GB or more. In this case, Amazon S3 will be the best option to store the video files, for cache storage over SSD (solid state drive) will be a cost-effective plan, and to store user-related data and file metadata MYSQL can be an efficient option for the storage.

Video Processing Steps:

- **Video Transcoding:**

The act of taking an encoded video file, decoding it, changing it in some way, and then re-encoding it is known as transcoding. The bitrate and resolution of the video, as well as the kind of compression, are all variables that may be altered during transcoding because they are determined by the data's encoding.

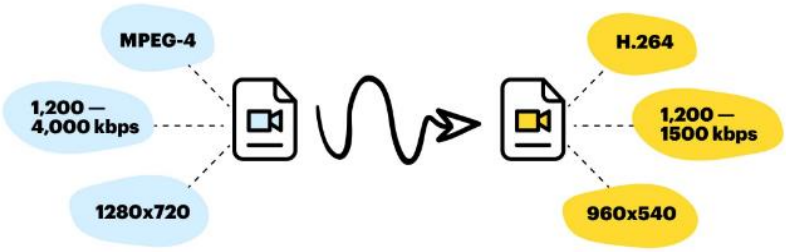

Figure 2. Video transcoding steps (Restream, .n.d)

- **Video Editing:**

When you edit and encode videos using a cloud service, you upload the original video files first. Then, you edit and encode the videos on your browser. You can edit from a laptop or low-powered computer because the actual changes are carried out on computers provided by the cloud provider.

- **Video Searching:**

The architecture of the Video Search is shown in the diagram bottom down. Any ONVIF camera may be given intelligence thanks to this framework. The Bridge transmits crucial photos and videos to the MOBOTIX CLOUD data centre when motion is detected. In order to process these crucial photos in real-time, MOBOTIX CLOUD has equipped its data centres with AI capabilities. Real-time data extraction from the most important photos and video tagging is performed by several AI models operating on cloud servers. When a search term is an input, the metadata is searched, and the corresponding key photos are shown.

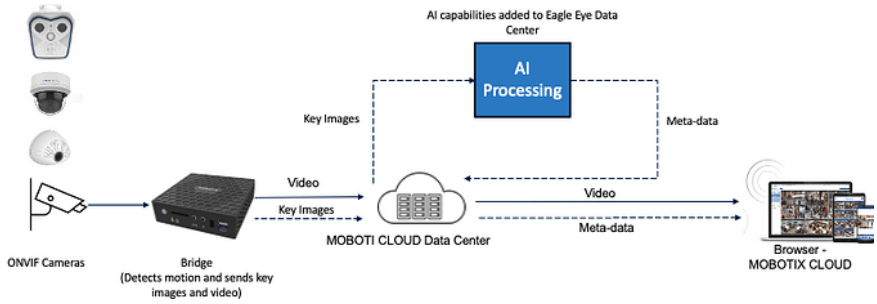

Figure 3. Architecture of video search over cloud (MOBOTIX, .n.d)

High availability: Overall, there are five VMs that should have a backup in terms of possible failures. These three Virtual machines are the Web Server, Application Server and Computation Processing server, QoS, and Security Server. We considered three separate ESXi hosts to increase the redundancy and also balance and spread the process between these three hosts. ESXi 1 will be the master and ESXi 4 will be the slave in the computing cluster. Although we cannot implement this in the vSphere yet.

DRS (Distributed Resource Scheduler): In our project, DRS is essential for video streaming as if we can allocate and distribute the service accordingly to the Hosts which is essential for us in order to benefit from all the available resources on the server. In other words, DRS provides balanced resources and mages with the workload of the machines. (VMware, n.d.)

Resource Control: On vSphere, there will be 3 resource pools to handle our computations on ESXi 1, 2 resource pools on ESXi 2 and 2 resource pools on ESXi 3. The priority of resources is as the table below illustrates. All the machines on ESXi 4 will have the same configuration as ESXi 1 and its Machines.

VM_Computing_Server_Esxi_1	High
VM_Application_Server_Esxi_1	High
VM_Web_Server_Esxi_1	Medium
VM_Security_Server_BackUp_Esxi_1	Low
VM_QoS_Monitoring_Server_BackUp_Esxi_1	Low
VM_Computing_Server_BackUp_Esxi_2	High
VM_Application_Server_BackUp_Esxi_2	Medium
VM_Web_Server_BackUp_Esxi_2	Medium
VM_Security_Server_Esxi_3	High
VM_QoS_Monitoring_Server_Esxi_3	Low

Table 1. Virtual Machines

Updating: In order to upgrade vSphere and ESXi in the cloud service, we can easily turn off the backup or main servers and upgrade them. If we did not have backup servers, we had to use vMotion to migrate our VMs from one host to another to progress in the update procedure.

VMware vSphere or OpenStack: The main difference between these two virtualizations is in costs. In terms of hardware and support, Implementing VMware usually costs more than OpenStack. Both VMware and OpenStack licencing and supports and deployments have become pretty much similar in the past years (Ubuntu, n.d). We can also consider using VMware Integrated OpenStack (VIO) which gives the developers access to the OpenStack APIs. So, the VMware infrastructure can also benefit from the OpenStack capabilities and modularity traits (ubackup.com, 2022). Therefore, we prefer to benefit from the vSphere and the specific virtualization it offers to stream and process the videos in the cloud.

Considerations and Challenges

Distributing the traffic: CDN

For the distribution of traffic CDN is used for faster loading speed for users of SmartV. Number of providers of CDN are available in the market, for example Akamai, Amazon CloudFront, CacheFly. Google Cloud CDN, etc. (*Content delivery networks*)

Web server connection method

There are plenty of connection types in Web server like Web Sockets, Server Sent Events, Web Real-time Connection (RTC), etc. the following table summarise some of them with their differences (*Connection Types In Web Apps*)

Connection Type	Transfer Mode	Common Usage
XHR/Fetch	Client to Server	Most dynamic web apps, pull to refresh, Gmail, Instagram etc
Web Sockets	Bi-Directional	Text messaging applications, broadcasting, real-time applications
Server Sent Events	Server to Client	Streaming data, tracking, timers
Web RTC	Client to Client	Screen sharing, video calling apps
Push Notifications	Server to Client	Notifications, alerts, background updates
Web Transport	Bi-Directional	Low latency, unordered messaging, media streaming, online games

Table 2. APIs

Geographical boundaries:
Determining the physical location of an object, such as a cloud computing server, is generally known as geolocation. Geolocation enables the identification of a cloud server's approximate location by adding that information to the server's root of trust. The hardware root of trust is seeded by the organization with the host's unique identifier and platform metadata stored in tamperproof hardware. This information is accessed using secure protocols to assert the integrity of the platform and confirm the location of the host (*Erin K. Banks et al.*). So, to simplify this, geolocation is connected with multiple CDNs.

Traffic Control: QoS
In a network with many computers, the volume of data moving through the firewall can be very large. You can use Traffic Management and Quality of Service (QoS) actions to prevent data loss. Traffic Management uses priority queues for outbound packets on an interface. There is one queue per interface for traffic management. QoS Uses eight queues per interface for different types of QoS packets. Each queue can hold 1000 packets. Lower-priority packets are only sent if there are no higher-priority packets in the queue. When the queue is full, subsequent packets are discarded. (*Traffic Management and QoS*)

Resource monitoring and consideration:
Resource monitoring and consideration are the foundation to achieve resource automation and high-performance management in a cloud computing environment. Resource monitoring consists of Data centre optimisation to minimise response time which gives a better experience to users, to monitor real-time temperature and to avoid overheating, monitor the power usage, etc.

Data Security Challenges:
The most common data security challenges in cloud computing:

- Lack of data visibility and control
  Solution: - Knowi is a tool used by many companies that allows them not only to see their data in one place and allows better visibility and control for data managers across multiple databases.
- Cloud misconfiguration and how it can leave data open and unprotected

Solution: - enabled encryption, monitoring cloud database access.

- Unauthorized access to cloud data and hijacking of accounts
  Solution: - encryption of data, strong password policy with two-factor or multifactor authentication, kipping and monitoring user log and
- Cyberattacks and data breaches
  Solution: - firewalls, VPNs, updates of software, etc.
- Denial of service (DoS) attacks flood
  Solution: - more bandwidth, redundancy in infrastructure,
- Data loss in cloud computing
  Solution: - viruses or malware protection, Auto-backup at secure place, software check and regular updates to avoid software corruption. (*Security Challenges*)

Cloud reliability: Cloud Challenges

The basic five challenges in cloud computing:

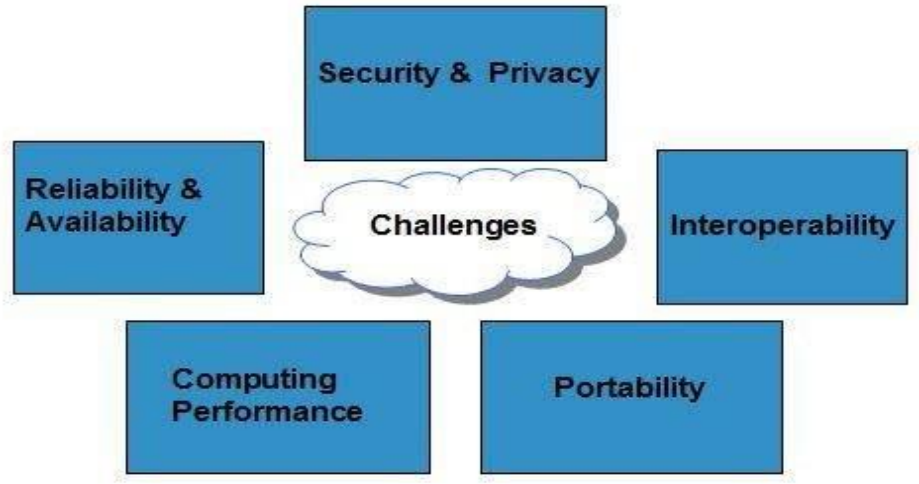

Figure 4. Challenges (*Cloud Computing Challenges, .n.d*)

- **Security and Privacy**: Information security and privacy are cloud computing's main problems. By using encryption, security hardware, and security software, security and privacy concerns can be resolved.
- **Portability**: Application migration from one cloud provider to another should be simple, which is another challenge for cloud computing. Neither should there be vendor lock-in. Because each cloud provider utilises a separate standard language for their systems, it is currently not practicable.
- **Interoperability**: It implies that an application running on one platform ought to be able to use services from other platforms. Web services enable it; however, creating such web services is really difficult.
- **Computing Performance**: Applications that use many data in the cloud need a lot of network bandwidth, which is expensive. Low bandwidth falls short of the cloud application's expected computing performance.

- **Reliability and Availability**: It is vital for cloud structures to be reliable and strong due to the fact maximum of the companies at the moment are turning to structure on offerings supplied with the aid of using third-party. (*Cloud Computing Challenges*)

High-Availability: Auto scaling

This term refers to a cloud computing feature that lets you automatically manage the different types of cloud scalability and automatically enable consistent performance regardless of the current demand on resources which gives high availability using auto-scaling.

Cost Management:

The cloud can have several business units with associated costs. These costs include Object storage costs, Data transfer costs, Networking costs, and Computing costs- associated with CPUs, RAM, and compute capacity. So, to maintain, we need cloud cost management, and this also has some challenges cost management:

- **Lack of Visibility into Cloud Spend**:
  Some businesses may find hidden costs because they don't have the right tools to determine their cloud spending. Cloud users may not interpret expense reports correctly, which can lead to spending more than necessary. To solve these problems, grant access to cloud expense reports so everyone can understand how their activities are impacting their spending. The right cloud management tools can help users uncover and eliminate hidden costs associated with the cloud. If users want to manage their cloud costs, choose a management tool that gives them a complete view of all cost centres in their cloud.
- **Inaccurate Budget Forecasts**:
  Cloud spending is difficult to predict, especially when cloud user has a large number of cloud resources. As user change his roadmap or perform certain tasks over time, his cloud costs will rise quickly. Inaccurate budget forecasts can negatively impact user's cloud cost management process. If he falls short of his budget, the active applications may be at risk. If he overestimates his cloud costs, he is paying for what he is not using. Instead of relying on guesswork, use his test environment to determine what he will use in the future. Some cloud cost management tools analyse resource health to produce accurate forecasts. They provide a general idea of what to expect based on historical data and trends.
- **Poor Cloud Architecting**:
  AWS has published the AWS Well-Architected Framework to help cloud engineers and developers build secure, reliable, and scalable workloads. This process creates a reliable, efficient and secure app. Without an AWS Well-Architected process, cloud application is vulnerable to crashes and attacks by hackers. However, most AWS users do not build their resources properly, resulting in unauthorized data access and data loss.
- **Complex Billing**:
  Billing is complex and cannot be left to the finance team alone. Companies building workloads in the cloud need to make sure every department understands the bill. For example, development teams often focus on efficiency and forget to manage cloud costs. All chargebacks must be transparent to avoid billing issues. (*Cost Management Challenges*)

Evolution of Technology

Role of Backup and High Availability

In today's business needs, backup and high availability are equally and crucially important. Every organisation has a tonne of data that needs to be backed up every day due to the development of technology and the expanding customer base. In addition, High Availability is turning into a requirement for all enterprises to limit their losses due to operational downtime as attacks by hackers using new and different methods continue to interrupt a business' day-to-day operations (*What's the Difference between High Availability and Backup Again? | Webroot*, n.d.).

Businesses should implement backups for all forms of data to meet expectations for document retention, discoverability, and recoverability. For the worst cases of data loss, backup protection should enable IT to do both simple file and folder restore and full-system recovery. Additionally, businesses must make a second backup copy and store it offshore in a safe environment, like the cloud (*How to Blend Backup and High Availability*, n.d.).

By building clusters, high availability can be accomplished in the cloud. A collection of servers that function as one server in a high-availability cluster offers constant uptime. These servers will all have access to the same shared data storage, so if one server goes down, the workload is distributed across the other servers (*High Availability in the Cloud | Ekco*, n.d.).

Data backup and high availability have undergone numerous improvements throughout time to accommodate shifting business requirements. Hard drives were the original high availability and backup technology, but it has since developed to include flash storage, storage area networks (SAN), and, more recently, cloud storage. The initial data backup method employed hard drives, but the read-and-write procedure was extremely slow and time-consuming, required much space, and was expensive to maintain. Flash storage, SAN, and the emergence of Big Data replaced this, enabling both the storage and intelligent analysis of data (*A Brief History of High Availability*, n.d.). Thus, the use of cloud storage for backup and high availability has been the recent trend.

Alternatives for Backup and High Availability

Continuous Data Protection: A technique known as continuous backup or continuous data protection (CDP) backs up data on a computer system each time a change is performed. By continuously recording data changes, CDP allows for system restoration to any earlier time point (*Continuous Data Protection*, n.d.).

Cloud Snapshots: Snapshots focus on speed. They make it possible to take a live copy of a dataset nearly instantly. Then, this copy can be made accessible for recovery as well as to other systems, frequently as a clone copy for things like testing and development. To ensure that we obtain quick and accurate data copies, snapshots are a common feature in most enterprise storage systems and are frequently seamlessly integrated with tools like Windows VSS(*Cloud Snapshots and Backups: How to Protect Data in the Cloud*, n.d.).

Image-based Backups: A computer or virtual machine operating system and all associated data, such as the system state and application configurations, are copied during an image-based backup procedure. The backup is stored as an image, which is a single file. The benefit of image-based backups is that all of the data can be gathered in one pass, giving each file-based backup an updated bare metal restore (BMR) capability(*What Is Image-Based Backup? - Definition from WhatIs.Com*, n.d.).

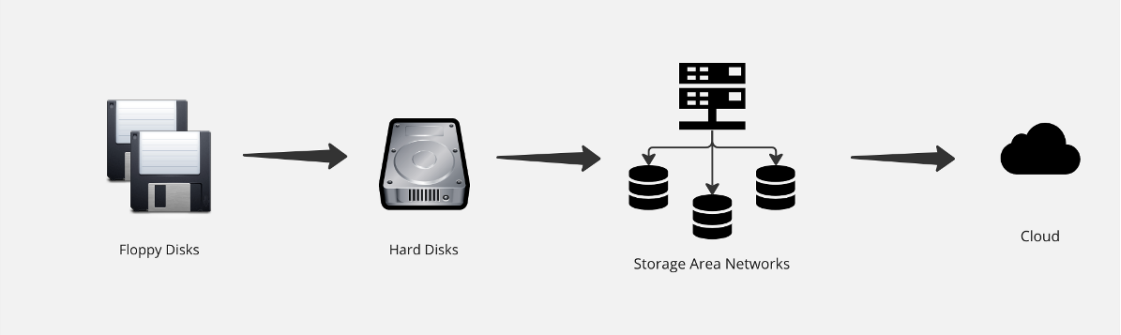

Figure 5. Evolution of Backup and High Availability

Recommendation: Continuous Data Protection would be the most appropriate approach to backup and high availability. The "backup window" issue, where businesses risk losing data created in the period between two scheduled backups, is now resolved. Additionally, it offers a strong defence against dangers like malware and ransomware, as well as sabotage and unintentional data destruction. Compliance with strict data protection standards may call for this. Furthermore, the data backup would take less time and be more economical because it is done in smaller blocks. Additionally, since the backup is performed at frequent and brief intervals, the restoration procedure would be quick and would be the best to minimise the time and money wasted as a result of the interruption. Additionally, CDP is the only substitute that may operate on any system and is not constrained to a virtual machine.

Comparisons and Recommendations of Different Virtualisation Platforms

Virtual Box: It is a free type 2 virtualisation software for Mac and Windows. It builds virtual machines for nearly every Windows, Linux, OS/2, Solaris, and Unix-based system. High levels of customizability but few pre-set integrations are required for commercial services, making it easier for amateur users. Virtual Box also supports snapshots to store the current VM state.

However, less capable graphics than VMware or Parallels. macOS or OS X guest systems are not easily supported (*Oracle VM VirtualBox for Mac - Review 2022 - PCMag Australia*, n.d.). This is recommended for amateurs and not for commercial use.

Hyper V: The Microsoft virtualisation program Hyper-V is a significant competitor in the market. One of the main rivals to VirtualBox is this. Hyper-V is a bare metal hypervisor, while VirtualBox is a hosted hypervisor. One of the most significant drawbacks of Hyper-V is that Only Windows is able to run Hyper-V.

Hyper-V only allows hardware virtualization in contrast to VirtualBox, although all of its other features are quite similar (migration, remote control, etc.) However, the program can be a bit

difficult to set up and use, and support isn't always the best (*Best Virtualization Software Comparison - DNSstuff*, n.d.). Hyper-V can only be recommended if you are using a window only or primarily windows-based system.

VM workstation Pro: Another key participant in the Windows virtualization software market is VMware, which also offers well-regarded products for a variety of host operating systems. However, it offers VMware fusion as an alternative for Mac users. VMware Workstation is type 2 hypervisor and is installed on the operating system and used as an application.

The GUI for VMware Workstation is excellent, simple to use, and straightforward. A bare-metal hypervisor for Windows computers called VMware vSphere is also produced by VMware. This is used to create virtual data centres rather than virtual desktops. VMware solutions can be on the expensive side and can be recommended for commercial and enterprise use (*Best Virtualization Software Comparison - DNSstuff*, n.d.).

Recommendation: The most highly recommended virtualization solution for business use is VMware Workstation Pro since it is an established and well-recognized enterprise-grade virtualization technology. There are commercial licences accessible. It also has got reputable third-party development and distribution. It is well worth the money and features a well-polished UI.

References

- *A Brief History of High Availability. (n.d.). Retrieved November 9, 2022, from https://www.cockroachlabs.com/blog/brief-history-high-availability/*
- *Benefits of switching to a hybrid cloud infrastructure. (2020, March 10). Lucidchart. https://www.lucidchart.com/blog/hybrid-cloud-benefits*
- *Best Virtualization Software Comparison - DNSstuff. (n.d.). Retrieved November 9, 2022, from https://www.dnsstuff.com/virtualization-software*
- *Cloud snapshots and backups: How to protect data in the cloud. (n.d.). Retrieved November 9, 2022, from https://www.computerweekly.com/feature/Cloud-snapshots-and-backups-How-to-protect-data-in-the-cloud*
- *Cloud Computing Challenges, https://www.tutorialspoint.com/cloud_computing/cloud_computing_challenges.htm*
- *Cost Management Challenges, https://www.nops.io/what-are-the-common-challenges-to-cloud-cost-management*
- *Connection Types In Web Apps, by: Nicholas Mendez, Issue 1 Nov 2021 https://dev.to/snickdx/connection-types-in-web-apps-15ie*
- *Content delivery networks (CDNs), by: IBM Cloud Education, 23 December 2020, https://www.akamai.com/our-thinking/cdn/what-is-a-cdn#:~:text=A%20CDN%20is%20a%20network,and%20stored%20elsewhere%20as%20needed.*
- *Continuous Data Protection. (n.d.). Retrieved November 9, 2022, from https://cloudian.com/guides/data-protection/continuous-data-protection/#*
- *Distributed Resource Scheduler | VMware vSphere. (2022, June). VMware. https://www.vmware.com/au/products/vsphere/drs-dpm.html*
- *Erin K. Banks et al., "Trusted Geolocation in the Cloud: Proof of Concept Implementation" (draft), NIST Interagency Report 7904, U.S. Dept. of Commerce, December 2012. http://nist.gov\publications\drafts\ir7904\draft_nistir_7904.pdf*
- *High availability in the cloud | Ekco. (n.d.). Retrieved November 9, 2022, from https://www.ek.co/publications/high-availability-in-the-cloud/*
- *How to blend backup and high availability. (n.d.). Retrieved November 9, 2022, from https://www.carbonite.com/blog/article/2017/10/how-to-blend-backup-and-high-availability*
- *Jan Ozer. (2021, November 11). Video editing in cloud. https://www.streamingmedia.com/Articles/Editorial/Featured-Articles/Video-Editing-in-the-Cloud-150047.aspx*
- *OpenStack vs VMware: Bringing costs down. (n.d.). Ubuntu. https://ubuntu.com/blog/openstack-vs-vmware*
- *Oracle VM VirtualBox for Mac - Review 2022 - PCMag Australia. (n.d.). Retrieved November 9, 2022, from https://au.pcmag.com/system-utilities/92188/oracle-vm-virtualbox-for-mac*
- *Pellen, A. (2020, November 25). Cost comparison: On-premises vs cloud computing. Harmonic | Faster Video Streaming | Virtualized Broadband Solutions. https://www.harmonicinc.com/insights/blog/on-prem-vs-cloud*
- *Security Challenges, 10 DATA SECURITY CHALLENGES & SOLUTIONS IN CLOUD COMPUTING, 1st may,2022 https://www.knowi.com/blog/10-data-security-challenges-solutions-in-cloud-computing/*

- *Traffic Management and QoS, WatchGuard Technologies, https://www.watchguard.com/help/docs/help-center/en-US/Content/en-US/Fireware/qos_trafficmanagement/traffic_mgmt_qos_about_c.html#:~:text=You%20can%20use%20Traffic%20Management,Guarantee%20or%20limit%20bandwidth*
- *VIDEO search - Cloud based AI search - FAQ (ENG). (2022, August 30). MOBOTIX Community. https://community.mobotix.com/t/video-search-cloud-based-ai-search-faq-eng/5579*
- *VMware vSphere ESXi vs OpenStack | What's the difference? (n.d.). Www.ubackup.com. Retrieved November 9, 2022, from https://www.ubackup.com/enterprise-backup/esxi-vs-openstack.html*
- *What is image-based backup? - Definition from WhatIs.com. (n.d.). Retrieved November 9, 2022, from https://www.techtarget.com/searchdatabackup/definition/image-based-backup*
- *What's the difference between high availability and backup again? | Webroot. (n.d.). Retrieved November 9, 2022, from https://www.webroot.com/blog/2021/07/15/whats-the-difference-between-high-availability-and-backup-again/*
- *What is Transcoding and why you need it. (2022, October 4). Ultimate Live Streaming Hub – Restream Blog. https://restream.io/blog/what-is-transcoding/*
- *L, N. (2018, September 8). Netflix system design. Medium. https://medium.com/@narengowda/netflix-system-design-dbec30fede8d*
- *What is a video API? (2021, February 8). Video API for hosting, encoding and delivery. Video as a Service | api.video. https://api.video/blog/video-trends/what-is-a-video-api*