LAKSHMI M R, 2025, 13:5
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

Data Retrieval During Failure of Virtual Machine

LAKSHMI M R, SHASHANK SHANKAR M, VINAY KURDEKAR, KARTHIK D

,Department of Computer Science and Engineering
Dayananda Sagar Academy of Technology and Management, Bangalore, India

Abstract Virtualization has become a cornerstone of modern computing, enabling multiple Virtual Machines (VMs) to operate on a single physical server, thereby enhancing resource utilization and scalability. However, one of the significant challenges in virtualized environments is the retrieval of data from VMs that are in a shut-down state. Traditional recovery methods are often inefficient, prone to data corruption, or unable to maintain system integrity. This project aims to explore and implement effective techniques for accessing and recovering data from VMs in various shutdown conditions, considering factors such as storage formats, hypervisor types, and security configurations. By analyzing existing approaches and developing optimized recovery workflows, the project seeks to ensure data integrity, reduce recovery time, and minimize system impact. The outcomes will contribute to improving virtual data resilience in enterprise and cloud environments, offering valuable solutions for system administrators, forensic analysts, and disaster recovery teams.

Keywords – Virtual Machines (VMs) ,Data Recovery, VM Shutdown State ,Virtualization

I. INTRODUCTION

Virtualization technology allows multiple Virtual Machines (VMs) to run on a single physical server, optimizing hardware usage and system flexibility. However, retrieving data from VMs after shutdown poses challenges due to storage complexity, security controls, and potential data corruption. This project focuses on analyzing and developing efficient methods to recover data from shut-down VMs while ensuring integrity and minimal disruption.

II. PROJECT OVERVIEW

The paper introduces HeatSnap, a system designed to enhance the efficiency of continuous snapshots in virtual machines (VMs) within web infrastructure environments. Traditional snapshot methods often treat all memory pages equally, leading to performance bottlenecks and inefficient storage usage. HeatSnap addresses this by distinguishing between frequently accessed "hot" pages and less-used "cold" pages. By applying specialized snapshotting and storage strategies to these different memory regions, HeatSnap aims to optimize snapshot performance and storage efficiency. The system was implemented on QEMU/KVM and demonstrated significant improvements in VM performance loss, snapshot duration, and storage efficiency compared to existing methods, as evidenced by evaluations on common web and cloud-based workloads

Advantages and Disadvantages

HeatSnap offers several advantages. By focusing on hot pages, it reduces the amount of data processed during snapshots, leading to faster snapshot creation and reduced storage requirements. This targeted approach minimizes the performance impact on running VMs, ensuring that applications remain responsive during snapshot operations. However, there are potential disadvantages. The

system's effectiveness relies on accurately identifying hot pages; misclassification could lead to important data being snapshot less frequently. Additionally, implementing such a system adds complexity to the VM management infrastructure, which could pose challenges during deployment and maintenance.

Gap Identified

Traditional VM snapshot systems often do not differentiate between memory pages based on access frequency, leading to inefficiencies in both performance and storage. HeatSnap addresses this gap by introducing a hot page-aware approach to continuous snapshots, optimizing the process by focusing on the most frequently accessed memory regions. This innovation provides a more efficient and performance-friendly solution for maintaining VM state in dynamic web infrastructure environments.

Data Center Virtualization and Secure Data Storage Architecture Project Overview

Data center virtualization enables multiple virtual machines (VMs) to run on shared physical servers, improving resource utilization,

scalability, and flexibility in managing IT infrastructure. However, this consolidation raises significant concerns about data security and storage integrity within virtual environments. This project proposes a secure data storage architecture tailored for virtualized data centers, integrating encryption, access control, and backup mechanisms to ensure that data remains confidential, available, and intact even in the event of VM failures or cyberattacks. The architecture is designed to work alongside popular hypervisors and supports multi-tenant environments common in cloud deployments.

Advantages and Disadvantages

The main advantages of this secure storage architecture include enhanced protection of sensitive data through encryption, prevention of unauthorized access using role-based controls, and improved reliability through regular backups and recovery plans. It also incorporates monitoring and auditing features to detect suspicious activities early. However, the system introduces complexity in management, as configuring security policies across multiple VMs and storage nodes requires expertise. Performance overhead is another challenge, as encryption and real-time monitoring can consume CPU and memory resources, potentially impacting VM performance. Moreover, scalability in large data centers can be difficult without adequate automation.

Gap Identified

Existing virtualization platforms often lack comprehensive data storage security measures that cover the full data lifecycle within virtual environments. While many focus on network security or hypervisor isolation, there is a gap in unified solutions that simultaneously address encryption, access control, backup, and auditing in an integrated architecture. This project aims to fill that gap by proposing a holistic approach to secure data storage in virtualized data centers, ensuring data confidentiality, integrity, and availability without sacrificing the benefits of virtualization.

Comparison of type-2 hypervisor performance on the example of VirtualBox, VMware Workstation player and MS Hyper-V

Project Overview

This project focuses on evaluating and comparing the performance of three widely used Type-2 hypervisors: Oracle VirtualBox, VMware Workstation Player, and Microsoft Hyper-V (Client version). These hypervisors operate on top of a host operating system and are popular choices for desktop virtualization, especially among developers, students, and testers. The study involves running identical virtual machines across each platform and measuring key performance indicators such as CPU usage, RAM consumption, disk I/O speed, and boot-up time. The objective is to assess how efficiently each hypervisor utilizes system resources and how it handles virtual workloads under real-world conditions. The outcome helps users select the most suitable hypervisor based on specific performance needs and host system constraints.

Advantages and Disadvantages

One of the main advantages of this project is that it provides a practical and measurable comparison of virtualization platforms under consistent test environments. It helps users understand how different hypervisors behave under similar workloads and offers clarity on resource optimization. VirtualBox is known for its flexibility and open-source nature, VMware Workstation Player excels in stability and performance, while MS Hyper-V provides seamless integration with Windows environments. However, there are some limitations. Each hypervisor has different levels of hardware support and feature sets, which might affect test fairness. Additionally, performance may vary across different host configurations, making it hard to generalize results. Another disadvantage is that some hypervisors may require additional configuration or licenses to access advanced features, which could affect usability and cost.

Gap Identified

Most existing comparisons of virtualization platforms are either outdated, incomplete, or based on theoretical data rather than empirical testing. While performance benchmarks exist, they often fail to use identical workloads and system setups, leading to inconsistent conclusions. Furthermore, many users select hypervisors based on popularity or convenience, without understanding how it impacts actual system performance. This project addresses this gap by offering a standardized and side-by-side evaluation of VirtualBox, VMware, and Hyper-V using consistent test criteria. It fills the need for updated, hands-on performance data that can guide users in choosing the right hypervisor based on factual metrics rather than assumption

Serverless Snapshot-Resume Performance in the Real-World

Project Overview

The paper titled "Serverless Snapshot-Resume Performance in the Real-World" delves into theperformance implications of using snapshot-resume techniques in serverless computing environments. Serverless computing, characterized by its event-driven execution model, often suffers from cold start latency—delays experienced when initializing functions. To mitigate this, snapshot-resume methods capture the state of a function after initialization, allowing for quicker subsequent invocations by restoring this state. The study evaluates the effectiveness of these techniques in real-world scenarios, analyzing their impact on function startup times and overall system performance

Advantages and Disadvantages

One of the primary advantages highlighted is the significant reduction in cold start latency achieved through snapshot-resume methods. By restoring a pre-initialized function state, systems can bypass repetitive setup processes, leading to faster response times. This enhancement is particularly beneficial for applications requiring rapid scalability and responsiveness. However, the paper also notes potential drawbacks. Maintaining and managing snapshots can introduce storage overhead and complexity, especially when dealing with numerous functions or frequent updates. Additionally, ensuring the consistency and security of restored states poses challenges, particularly in multi-tenant environments where isolation is paramount.

Gap Identified

While snapshot-resume techniques offer promising improvements in reducing cold start times, the paper identifies a gap in understanding their performance across diverse workloads and environments. Most existing studies focus on controlled or synthetic benchmarks, lacking insights into real-world applications with varying resource demands and execution patterns. This research addresses that gap by providing empirical data on the performance of snapshot-resume methods in practical settings, offering valuable guidance for optimizing serverless platforms.

eHotSnap: An Efficient and Hot Distributed Snapshots System for Virtual Machine Cluster

Project Overview

The paper introduces eHotSnap, an advanced system designed to efficiently capture distributed snapshots

of virtual machine clusters (VMCs) in Infrastructure-as-a-Service (IaaS) cloud environments. Traditional snapshot mechanisms often incur significant downtime and performance degradation, particularly in large-scale deployments. eHotSnap addresses these challenges by decoupling the coordination phase from the snapshotting process. It employs a two-phase approach: a lightweight transient snapshot followed by a full memory snapshot. This method ensures that the snapshot process is logically completed within a second, minimizing system disruption. Additionally, eHotSnap incorporates optimizations such as memory deduplication and a priority queue mechanism to enhance efficiency during the memory snapshot phase. Implemented on the QEMU/KVM platform, eHotSnap demonstrates significant improvements in snapshot performance and system reliability.

Advantages and Disadvantages

The primary advantage of eHotSnap is its ability to perform distributed snapshots with minimal system downtime, making it suitable for applications requiring high availability. By separating the coordination from the snapshotting process, it reduces the risk of network interruptions and ensures a consistent global state across the VMC. The integration of memory deduplication further optimizes storage usage, and the priority queue mechanism ensures that critical guest write operations are handled promptly. However, the system's complexity may increase due to the additional components and coordination required. Moreover, the effectiveness of eHotSnap is contingent on the underlying infrastructure's performance, and its benefits may vary depending on the specific workload and environment.

Gap Identified

Prior to eHotSnap, existing distributed snapshot systems often struggled with balancing the need for consistency with the requirement for minimal disruption. Many approaches either compromised on consistency to achieve faster snapshots or incurred significant downtime to ensure consistency. eHotSnap bridges this gap by introducing a coordinated yet efficient snapshot mechanism that maintains global consistency without substantial performance penalties. This advancement is particularly valuable in cloud environments where maintaining service continuity is paramount.

Be United in Actions: Taking Live Snapshots of Heterogeneous Edge–Cloud Collaborative Cluster With Low Overhead

Project Overview

This paper presents a system designed to efficiently take live snapshots of heterogeneous clusters that combine edge devices and cloud resources. In modern distributed computing, such edge-cloud collaborative clusters support latency-sensitive and bandwidth-heavy applications, but their diverse hardware and software configurations make consistent snapshotting a challenge. The proposed solution unifies the snapshot process across different platforms with minimal interruption to ongoing services. By coordinating snapshot operations intelligently, the system ensures consistency while maintaining low overhead, enabling better fault tolerance and disaster recovery in edge-cloud environments.

Advantages and Disadvantages

The main advantage of this system is its ability to handle heterogeneous environments, which are typical in edge-cloud collaboration, without imposing heavy performance penalties. It achieves efficient snapshotting by leveraging lightweight coordination and optimized data handling strategies that minimize network congestion and resource use. This allows applications to continue running smoothly during snapshot operations. However, the complexity of coordinating snapshots across diverse devices and network conditions introduces challenges in implementation and maintenance. Furthermore, some edge devices with limited resources may still experience performance degradation during snapshot capture.

Gap Identified

Existing snapshot techniques often focus separately on either cloud or edge environments but rarely address the combined, heterogeneous nature of edge-cloud clusters. This results in inefficiencies and higher overheads when applied to such integrated systems. The paper identifies this gap and contributes a unified, low-overhead approach that ensures consistent snapshots across varied hardware and software platforms. This is a crucial advancement for maintaining reliability and availability in emerging distributed applications that span from edge to cloud.

Comparison of containerization and virtualization in cloud architectures Project Overview
This project explores the fundamental differences and performance implications of containerization and virtualization

technologies within cloud computing architectures. Virtualization uses hypervisors to create multiple virtual machines (VMs), each with its own operating system, on a single physical server. In contrast, containerization packages applications and their dependencies into lightweight containers that share the host OS kernel but operate in isolated user spaces. Both approaches enable efficient resource utilization and scalability but differ in overhead, portability, and deployment speed. This comparison evaluates factors such as startup time, resource efficiency, security, and management complexity to help cloud architects choose the optimal technology for specific use cases.

Advantages and Disadvantages
Containerization offers rapid deployment, lower overhead, and better resource efficiency since containers share the host OS and require fewer resources than full VMs. This makes containers ideal for microservices architectures, continuous integration/continuous deployment (CI/CD), and cloud-native applications. However, containers may face security risks due to shared kernels and can be less suitable for running multiple diverse operating systems. Virtualization, on the other hand, provides stronger isolation by running separate OS instances, which enhances security and compatibility with legacy applications. But VMs typically require more resources, have longer startup times, and impose greater management complexity compared to containers.

Gap Identified
While both containerization and virtualization are well-established, there remains a lack of comprehensive comparative studies focusing on their performance and security trade-offs in large-scale, multi-tenant cloud environments. Existing research often evaluates them separately or under limited workloads. This project addresses the gap by providing a side-by- side analysis under identical cloud infrastructure settings, offering clearer insights into how each technology performs in real- world scenarios. This understanding is critical for enterprises planning cloud migration or hybrid cloud deployments.

Hyper-V as type-2 hypervisor virtualization: guest file system performance examination Project Overview
This project investigates the performance of the guest file system running on virtual machines hosted by Microsoft Hyper-V configured as
a Type-2 hypervisor. Unlike traditional Type-1 hypervisors that run directly on hardware, Type-2 hypervisors run on a host operating system, which can affect performance characteristics. The study focuses on how well Hyper-V manages file system operations within guest VMs, including read/write speeds, latency, and I/O throughput. Through benchmarking and comparative testing, it aims to understand the impact of virtualization overhead on file system performance and to identify potential bottlenecks when running storage-intensive applications on Hyper-V virtual machines.

Advantages and Disadvantages
One key advantage of using Hyper-V as a Type-2 hypervisor is its tight integration with the Windows ecosystem, providing ease of use and compatibility for Windows-based host and guest systems. It supports advanced virtualization features such as dynamic memory allocation and snapshotting, which enhance VM management. However, running as a Type-2 hypervisor introduces additional overhead compared to bare-metal solutions, potentially affecting file system responsiveness and throughput.

The layered architecture can lead to increased latency and resource contention, especially under heavy disk I/O workloads. Another disadvantage is that tuning performance requires careful configuration and may not be straightforward for all users.

Gap Identified

While Hyper-V's overall virtualization capabilities are well documented, detailed performance analysis specifically targeting guest file system operations under Type-2 configurations is limited. Most research focuses on Type-1 Hyper-V deployments or general VM performance without isolating file system metrics. This project fills that gap by providing targeted insights into how virtualization affects file storage and retrieval tasks on Hyper-V Type-2 VMs. The findings help optimize VM configurations for workloads with demanding file system requirements, contributing to better performance tuning and resource planning.

Comparison of VMware Workstation, VirtualBox and MS Hyper-V hypervisor performance with MS Windows OS based guests

Project Overview

This project compares the performance of three popular Type-2 hypervisors—VMware Workstation, Oracle VirtualBox, and Microsoft Hyper-V—when running Microsoft Windows operating systems as guest virtual machines. Each hypervisor runs on a host OS and manages resources to create isolated virtual environments. The study involves benchmarking key performance indicators such as CPU utilization, memory consumption, disk I/O, and network throughput to evaluate how each hypervisor handles Windows-based workloads. The goal is to provide practical insights for users and organizations to select the best hypervisor based on Windows guest performance and resource efficiency.

Advantages and Disadvantages

VMware Workstation is widely praised for its stability, comprehensive features, and strong performance with Windows guests, making it a preferred choice for professional and enterprise users. VirtualBox offers an open-source and user-friendly alternative with broad hardware support, though it may lag behind VMware in raw performance. MS Hyper-V integrates deeply with Windows hosts, providing benefits like seamless management and strong compatibility, but sometimes suffers from higher resource overhead. Each hypervisor presents trade- offs between ease of use, performance optimization, and feature availability. Limitations include variability in support for advanced guest features and differences in how each handles resource allocation under heavy workloads.

Gap Identified

Although many benchmarks compare hypervisors, few focus explicitly on Windows guests under identical conditions across VMware Workstation, VirtualBox, and Hyper-V. Existing studies often use varied host systems or lack comprehensive metrics, making it difficult to draw clear conclusions. This project fills that gap by providing a controlled, side-by-side evaluation tailored for Windows environments, enabling users to make informed decisions when deploying Windows VMs in different virtualization platforms. This insight is especially valuable for environments relying heavily on Windows software compatibility and performance.

Comparison between common virtualization solutions: VMware Workstation, Hyper-V and Docker

Project Overview

This project compares three widely-used virtualization solutions: VMware Workstation, Microsoft Hyper-V, and Docker containers. VMware Workstation and Hyper-V are hypervisor-based virtualization platforms that create virtual machines (VMs) with

separate operating systems, providing strong isolation. Docker, on the other hand, uses containerization technology that packages applications and their dependencies in lightweight containers sharing the host OS kernel. The study evaluates these solutions on factors such as resource efficiency, startup time, security, ease of management, and suitability for different workloads, aiming to help users and organizations choose the best fit based on their infrastructure needs.

Advantages and Disadvantages

VMware Workstation offers robust VM management and compatibility with multiple OSes, excelling in scenarios where complete OS isolation is required. Hyper-V integrates seamlessly with Windows environments and provides enterprise-grade features suitable for Windows-centric data centers. Docker shines in rapid deployment and resource efficiency, making it ideal for microservices and cloud- native applications. However, VMware and Hyper-V involve higher overhead due to running full OS instances, resulting in longer startup times and more resource consumption. Docker's shared kernel model reduces overhead but introduces potential security concerns and limits support for running diverse operating systems within containers.

Gap Identified

While many comparisons exist between hypervisors or container platforms individually, there is a lack of comprehensive studies comparing these three technologies side-by-side under consistent test conditions. Most research focuses on either traditional VMs or containers separately, without exploring the trade-offs users face when choosing between full virtualization and containerization. This project fills this gap by providing a balanced comparison that highlights the strengths and weaknesses of VMware Workstation, Hyper-V, and Docker, helping decision-makers understand which technology best fits their specific application requirements and operational contexts.

III. CONCLUSIONS

This project successfully explored the challenges and solutions related to data retrieval from shut-down Virtual Machines. Through analysis of existing techniques and implementation of optimized recovery workflows, it demonstrates that data can be efficiently recovered without compromising system integrity or security. The developed methods enhance reliability in virtual environments, offering practical value for system administrators, forensic investigators, and disaster recovery operations. Future improvements can focus on automation, support for diverse hypervisors, and real-time integrity verification during recovery.

REFERENCES

Academic Papers & Journals

1."Virtual machine introspection: Techniques and applications"
Garfinkel, T., & Rosenblum, M. (2003). ACM SIGOPS Operating Systems Review. Focuses on VM introspection for analyzing VMs externally.

2. "Forensic data recovery from Virtual Machines: Challenges and approaches" International Journal of Digital Crime and Forensics (IJDCF)
Explains methods to access data from shut-down or corrupted VM images.

3. "A survey of data recovery techniques in cloud environments" International Journal of Computer Applications (IJCA)
Provides an overview of recovery methods in virtualized environments.

4. "Secure and reliable VM image management in cloud computing" IEEE Transactions on Cloud Computing.
Addresses storage and recovery mechanisms for VM data.

5. "Live Forensics for Hypervisor-based Virtual Machines" Dolan-Gavitt et al., 2011.
Explores data retrieval without starting the VM.

Books & Chapters

6. "Virtualization Essentials" by Matthew Portnoy
Covers VM architecture, shutdown behavior, and data handling. Wiley Publishing.

7. "Cloud Computing: Concepts, Technology & Architecture" by Thomas Erl Includes VM data recovery, state transitions, and security implications.

Technical Documentation & Whitepapers

8. VMware Knowledge Base –Articles on data recovery from suspended and powered-off VMs.

9. Microsoft Hyper-V Documentation – learn.microsoft.com/en- us/windows-server/virtualization/hyper-v/)
Official recovery and management procedures for shut-down VMs.

10. Red Hat KVM Virtualization Guide – access.redhat.com/documentation/en-us/) Recovery techniques and VM image handling for KVM-based systems

Industry Blogs & Case Studies

11. "Data Recovery from Virtual Machines – A Guide" DiskInternals Blog Practical insights on recovering VHD/VMDK files from powered-off VMs.

12. "Digital Forensics in Virtual Environments" – SANS Institute Whitepaper Discusses the implications of VM shutdown states on forensic investigations.