Project Proposal
Project Name: Cloud Odyssey
Date: 28/03/2025
Version: 1.0

21116034	Vaishnavi Virat Dave
21116048	Kaashvi Jain
23114043	Kajal
23114052	Kumud
21323003	Abhijna Raghavendra
23114008	Anushka Jangid

---

1. Introduction

Cloud Odyssey is a high-performance, distributed cloud computing framework that leverages idle computational resources by orchestrating them into a centralized Beowulf cluster. A Beowulf cluster is a computer cluster of what are normally identical computers networked into a LAN with libraries and programs installed which allow processing to be shared among them. The result is a high-performance parallel computing cluster from inexpensive personal computer hardware. This system enables seamless execution of parallel workloads through OpenMPI (Open Message Passing Interface) and SLURM (Simple Linux Utility for Resource Management) while dynamically scaling worker nodes to optimize resource utilization. By integrating robust monitoring and remote access capabilities, Cloud Odyssey fosters efficient collaboration across networks, ensuring an adaptable and cost-effective alternative to traditional cloud infrastructure. Cloud infrastructure essentially requires huge server rooms that are obviously costly to build and maintain and also suffer from single point failures. This distributed cloud system solves this issue and also does a better resource utilization as we are ultimately using the idle compute of the worker nodes.

2. Objectives

- Establish a scalable, **decentralized cloud computing framework** for parallel task execution.

- Facilitate secure SSH-based access to a Master Node for workload submission.
- Implement OpenMPI and SLURM for efficient job scheduling and execution across worker nodes.
- Enable dynamic scaling of worker nodes based on real-time resource availability.
- Integrate Prometheus and Grafana for comprehensive system monitoring and analytics.
- Ensure secure remote accessibility via LocalTunnel or ngrok to circumvent NAT/firewall restrictions.

3. Problem Statement

The demand for high-performance computing (HPC) resources has significantly increased across research institutions and enterprises, yet existing commercial cloud solutions impose prohibitive costs. Many organizations or even individual systems possess underutilized computational resources that, if efficiently harnessed, could serve as a cost-effective alternative. Cloud Odyssey addresses this gap by providing a distributed framework that seamlessly aggregates and optimizes idle machines for parallel computing, mitigating financial and infrastructural constraints.

4. Methodology

- **Phase 1:** Develop the foundational system, including Master Node setup, SSH-based access, and preliminary workload scheduling.
- **Phase 2:** Implement dynamic scaling mechanisms, enabling worker nodes to join and leave the cluster seamlessly.
- **Phase 3:** Integrate system monitoring and logging solutions to enhance operational transparency and performance evaluation.
- **Final Phase:** Conduct extensive testing, refine security measures, optimize execution efficiency, and finalize deployment documentation.

5. Expected Outcomes

- A fully functional, decentralized cloud computing framework optimized for HPC applications.

- Efficient utilization of underutilized computational resources to reduce infrastructure costs.
- Scalable and extensible system architecture capable of supporting diverse workload requirements.
- Secure and reliable remote-access capabilities for seamless user interaction.
- Open-source availability with a roadmap for future enhancements, including Kubernetes integration.

7. Technical Requirements

- **Software:** OpenMPI, SLURM, OpenSSH, NFS (Network File System), FastAPI (for high performance HTTP based service APIs), Fiber (Golang) for low memory footprint, PostgreSQL as it supports both relational and non-relational queries, Redis Streams for dynamic data analysis, Prometheus for event monitoring, Grafana for data visualization, Docker for OS-level virtualization and Docker Compose for multi-container applications.
- **Hardware:** Multi-core processors, a minimum of 8GB RAM per node, and high-speed network connectivity.

8. Possible Challenges and Mitigation Strategies

Challenge	Mitigation Strategy
Security vulnerabilities in SSH access	Enforce key-based authentication, firewall rules, and encryption protocols
Network latency affecting parallel execution	Optimize scheduling algorithms and implement efficient data transfer mechanisms
Unpredictable worker node availability	Develop fault-tolerance strategies and intelligent load balancing
Scalability limitations	Integrate container orchestration (e.g., Kubernetes) in future iterations

9. Future Enhancements

- Implementing advanced fault-tolerant mechanisms to ensure continuous system availability.

- Introducing auto-scaling policies that dynamically adjust resources based on real-time computational demand.
- Enhancing orchestration capabilities through Kubernetes for improved resource management and deployment flexibility.

Conclusion: Cloud Odyssey represents a paradigm shift in distributed computing, leveraging idle computational resources to construct an efficient, scalable, and cost-effective cloud infrastructure. By integrating advanced workload scheduling, dynamic resource allocation, and comprehensive monitoring, this system provides a viable alternative to conventional cloud services while fostering collaborative high-performance computing. With a future-oriented roadmap, Cloud Odyssey is well-positioned to evolve into a robust, autonomous cloud computing solution for diverse research and enterprise applications.