Shriyansh Singh

+1-930-333-5141 | shriyansh.singh24@gmail.com | LinkedIn

SUMMARY

Machine Learning Engineer adept at collaboration, problem-solving, and results. Guides cross-functional teams to deliver solutions, ensuring strategic alignment and fostering growth in dynamic research settings.

PROFESSIONAL EXPERIENCE

Machine Learning Intern *April 2024 - Dec2024*

Hyphenova AI *Los Angeles, CA*

- Maintained a **Linux-based HPC cluster** supporting large-scale ML workloads, integrating **CUDA** and **MPI** for distributed model training with a 35% improvement in throughput.
- Optimized **GPU resource allocation** via containerization (Docker, Kubernetes), reducing idle GPU time by 20% and ensuring consistent performance across multiple deep learning projects.
- Designed and implemented **parallel file system** strategies (e.g., Lustre) to address I/O bottlenecks, decreasing average data access latency by 25%.
- Developed **monitoring dashboards** (Python, Grafana) for real-time HPC metrics (GPU load, memory usage), enabling proactive issue detection and faster debugging.
- Collaborated with data scientists and DevOps teams to **profile and debug HPC workloads**, resolving cross-node synchronization issues and improving job runtimes by 30%.
- Implemented best practices for HPC **job scheduling** (SLURM), ensuring high utilization of compute nodes and seamless integration of multi-tenant workloads.

Data Analyst Intern *May 2022 - Oct 2022*

Enterprise Business Technologies Pvt Ltd *Mumbai, India*

- Managed **Azure-based HPC environments** for advanced analytics, configuring container images for large-scale data simulation and iterative ML model training.
- Conducted **performance tests** on distributed time-series forecasting jobs, identifying memory hotspots and optimizing parallel processing tasks, leading to a 25% decrease in runtime.
- Provided **user training** and technical support on HPC best practices (MPI, resource allocation), increasing HPC adoption rates by 30% across data science teams.
- Developed in-house **usage monitoring scripts** (Python) to track cluster resource usage, ensuring fair scheduling policies and enhanced visibility into HPC consumption.
- Resolved **parallel I/O performance** issues by implementing caching strategies and consolidating read/write operations, enhancing cluster-wide efficiency for high-volume data processing.

EDUCATION

Indiana University Bloomington Aug 2023 – May 2025

Master of Science in Data Science *Indiana, United States*

- Relevant Coursework: High-Performance Computing, Cloud Computing, Distributed Systems, Applied Machine Learning, Big Data Applications.

University of Mumbai Aug 2018 – May 2022

Bachelor of Engineering in Electronics *Maharashtra, India*

TECHNICAL SKILLS

Languages and ML Frameworks: Python, C/C++, Bash, CUDA, PyTorch, TensorFlow
HPC and Parallel Computing: MPI, SLURM, Kubernetes, Docker, InfiniBand, Lustre/GPFS
Systems & Cloud: Linux (RHEL/Ubuntu), AWS, Azure, Terraform, Kubernetes, Containerization
Performance Optimization: Code profiling (Nsight, gprof), memory management, GPU kernel tuning
Data Analytics & Visualization: Python (Pandas, NumPy), Grafana, Power BI, SQL, NoSQL

PROJECTS

High-Throughput HPC Pipeline for Genomic Analysis | *MPI, CUDA, Docker* *Jan 2024 – Apr 2024*

- Orchestrated **MPI** workflows on a 40-node HPC cluster to parallelize genomic data analysis, cutting end-to-end processing time by 50%.
- Leveraged **CUDA-accelerated** algorithms for sequence alignment, reducing kernel execution time by 35% on large-scale datasets.
- Deployed HPC applications within **Docker containers** for portable and reproducible environments, streamlining collaborative research efforts.
- Monitored cluster usage via custom **Grafana dashboards**, enhancing visibility into compute, memory, and network utilization.