

PARTH VIPUL SHAH

San Francisco, California | (661) 514-7214 | parthvshah@gmail.com | parthvshah.me

Performance engineer with 2+ years of experience optimizing Linux kernel at Oracle, specializing in workload profiling, benchmarking infrastructure and statistical analysis. Proven ability to build tools for production, analyze complex workloads and drive improvements at a kernel level for large enterprise systems. Looking for platform and infrastructure engineering roles.

EXPERIENCE

Oracle

Linux Performance Engineer; Feb 2024–May 2026

- Identified a newidle balancing backoff algorithm regression in the v6.17 kernel release using a performance agent built with codex that builds kernels, runs micro-benchmarks and collects hardware counters to flag regressions or improvements at a per-commit level.
- Quantified the performance cost of virtualization on bare metal hosts by measuring host kernel efficiency across hand-crafted qemu-kvm VMs running production workloads. Designed cgroup-based CPU and memory pressure monitoring tools to support ongoing analysis.
- Improved overall throughput by 17% and 40% for database and web server workloads respectively by identifying logging bottlenecks and recommending scheduler preemption tunings through root cause analysis using perf and Dtrace.
- Analyzed performance data from complex real-world workloads and OCI service proxies such as object store using numpy, scipy, pandas and matplotlib. Verified a zombie cgroup fix in the kernel using an advanced memory soak test that measured and plotted the rate of memory reclaim per second.
- Closed the feedback loop between hardware vendors and performance teams by delivering memory latency and bandwidth characterization data on new hardware platforms using benchmarks such as lat_mem_rd and himeno.
- Reduced measurement-to-report time from weeks to days by building and scaling automation infrastructure for kernel performance measurement, enabling routine per-release comparison of Oracle Linux against upstream.
- Created tools and methodologies to characterize, analyze, and profile workloads used to characterize the performance of Oracle Linux. Identified the total amount of kernel code that was exercised by our suite of workloads and plugged gaps using new workloads.
- Worked on a Principal Component Analysis (PCA) based approach to identify signatures of certain key workloads to reduce the total amount of workload runtime without sacrificing coverage.

Commvault

Associate Software Engineer; Jan 2021–Dec 2021

- Spearheaded development of the PostgreSQL data agent — multi-stream file system/dump-based backups and restores.
- Enabled protection of on-prem and cloud (AWS, Azure, GCP) PostgreSQL databases for 25+ environments by working on 90+ enhancements and defects.

Samsung R&D Institute

Research Intern; May 2020–Jul 2020

- Created a deep learning model with TensorFlow Lite to detect device overheating as part of the On Device AI team.
- Processed raw handset images to determine ambient temperature — average RMSE of 3.79 (CNN) and 2.32 (LSTM) for single and multi-image prediction.

SKILLS

AI tools: Codex, Claude Code, Cursor, Wispr Flow

Technologies: Linux, Perf, Dtrace, NUMA, ARM's Statistical Profiling Extension, cgroups, qemu-kvm, Oracle Database, Apache, Phoronix Test Suite, SPEC CPU 2017, SPECjbb2015, PostgreSQL, Docker, Airflow, React, MongoDB, .NET, Git

Clouds: OCI, AWS, Azure, GCP

Languages: Bash, C/C++, Python, JavaScript, C#, HTML/CSS, SQL

EDUCATION

University of Southern California (USC)

Master of Science in Computer Science; Jan 2022–Dec 2023

PES University

Bachelor of Tech. in Computer Science and Engineering; Aug 2017–Jul 2021

RESEARCH

Information Sciences Institute *Nov 2022–Dec 2023*

- The unequal opportunities of large language models: Examining demographic biases in job recommendations by ChatGPT and LLaMA. Studying the biases in Large Language Models (LLMs). 98 citations. Published at ACM's EAAMO'23. [Link](#).

- Creating a Bayesian model to represent Ranked Choice Voting responses. Can predict the outcome of elections in the United States and segment populations based on survey responses.

PES University *Jun 2020–Sep 2020*

- Prediction of the Peak, Effect of Intervention and Total Infected by the Coronavirus Disease in India: Forecasted using the SEIR compartmental model. 4 citations. Published with the Cambridge University Press. [Link](#).

SSCU, Indian Institute of Science (IISc) *Jun 2019–May 2020*

- Conceived 3 parallel algorithms for computing correlation functions using the MPI standard in C — achieved super linear speedups. Released package on a 120-node HPC cluster.

PROJECTS

- Kernel tuning for fast GPU-based inference: Measured the performance impact of kernel tunables such as transparent huge pages, preemption modes and CPU frequency governors on LLM inference.
- Video Indexer: Indexes a video into scenes, shots, and sub-shots based on visual and audio features. Achieved 99% accuracy.
- Nutritional KG: A recommendation system that uses a knowledge graph with 10K+ entities to help minutely alter your diet.
- Converting Black-box Neural Networks into Interpretable Decision Trees, Explainable AI: Processed using layer-wise relevance propagation and perturbations. Model and data agnostic methods.
- Naturalization of Text by the Insertion of Pauses and Filler Words: Used bigram frequency and an RNN. 55% convincing.
- Database as a Service: Created using Docker, RabbitMQ, ZooKeeper. High availability and scalability. Tested on AWS. 100% uptime and supported 500+ concurrent reads/writes.