Nachiket Sirsikar

+1(857) 269-6402 | nachiketsirsikar01@gmail.com | linkedin.com/in/nachiketsirsikar | github.com/nachiketS | Boston, MA

TECHNICAL SKILLS

Data & Pipelines: PySpark, HBase, Postgres, SQLAlchemy, Airflow, ETL/ELT Pipelines, Data Warehouse

Languages: Python, SQL, Java, JavaScript, C/C++, R

Cloud & Infra: AWS (EMR, S3, EC2, Athena, EKS), Terraform, Kubernetes, Helm, Docker, GitLab CI, KEDA, Prometheus, Grafana

Frameworks: Flask, Django, FastAPI, Pandas, NumPy, scikit-learn, Keras, OpenCV, matplotlib

EXPERIENCE

BitSight Technologies Inc. **Boston, MA**

Senior Software Engineer I (Data Infrastructure) *Feb 2026 – Present*

- Architected and built a webhook callback architecture from the ground up, engineered to support 200K+ events/day, implementing idempotency keys and exponential backoff to guarantee reliable delivery at that scale
- Led a high-priority feature replacing external API calls with direct HBase queries, eliminating third-party dependencies and rate limits and cutting a full inventory run 67% (3 days to 1 day).
- Set up Grafana dashboards on HBase JMX metrics, giving the team cluster observability that enabled optimizations saving ~$2K/month

Software Engineer II (Data Infrastructure) *Feb 2024 – Jan 2026*

- Developed a region-aware, multithreaded scanning layer for HBase's query engine, reducing p90 scan latency by 75% (from 3 hours to 45 minutes)
- Built an ETL pipeline ingesting and normalizing data from Slack, Jira, Confluence, and GitLab APIs into a unified store, powering automated first-pass triage classification and cutting manual triage overhead by 40%
- Optimized Kubernetes resource utilization by switching to an event-driven architecture (KEDA) with custom queues for specific tasks, processing 1.5x the daily tasks without additional cloud cost
- Redesigned a long-running batch job as an asynchronous, highly parallel pipeline, cutting end-to-end runtime 69% (90 to 28 days)
- Offloaded filtering to HBase’s server side instead of Python memory, cutting in-memory data ~62% and query time ~33% (up to 8x on large queries) and eliminating worker OOM failures.

Software Engineer *Jun 2023 – Feb 2024*

- Productionized an XGBoost model via PMML for the customer portal, building optimized feature computation for inference and integrating it end to end with a 100% success rate
- Diagnosed and fixed a quadratic query pattern in a Django service, reducing query complexity from O(N^2) to O(N) and eliminating a major performance bottleneck.
- Profiled a Python service and cut task latency, reducing task latency by as much as 92% (3 min to 15 sec) by using better data structures and memoization

Software Engineering Co-Op *Jun 2022 – Dec 2022*

- Cut per-query lookups 5x by prototyping a denormalized schema and PySpark ingestion pipeline over ~5 TB of data
- Designed a failure-recovery mechanism for the ELT pipeline, reducing potential service downtime from 24 hours to 10 minutes with no additional cloud cost

ZS Associates **Pune, India**

Associate Data Analyst *Dec 2020 – Aug 2021*

- Queried and processed large-scale electronic health records and physician survey datasets using AWS Athena, supporting downstream analytics and reporting pipelines
- Collaborated with domain experts on feature engineering, reducing model false negative rate from 6% to 4%

EDUCATION

Northeastern University **Boston, MA**

Master of Science in Computer Science (Artificial Intelligence) *Sep 2021 – May 2023*

Maharashtra Institute of Technology **Pune, India**

Bachelor of Engineering in Information Technology *Aug 2016 – May 2020*