Akshat Maurya

Systems & Tooling Engineer

New Delhi, India • akshat_maurya9@proton.me • +91 9451504938
linkedin.com/in/akshat-maurya • github.com/akshat09867 • makshat.in

SUMMARY

CS undergraduate at JNU (graduating May 2027) and two-time Google Summer of Code contributor to the R Project. Builds performance-focused data tooling (streaming ingestion, query translation, automated joins) and has published four CRAN packages with 8,000+ total downloads. Seeking a software, data, or developer-tools engineering internship.

EXPERIENCE

Open Source Contributor, R Project *(Google Summer of Code)* May 2026 – Aug 2026

- **awkreader** (R package, on CRAN, 432+ downloads): Reads, filters, and aggregates large delimited files by translating R queries into AWK programs that stream data from disk.
- Built the translation engine with regex-based function detection, numeric type guards, and metadata handling for headerless data.
- Benchmarked **22.8**× faster than `readr` and up to **5.7**× faster than `data.table` on multi-file filtered reads and record counting (median of 20 runs).
- Engineered disk-level streaming aggregations (`aggregated.fread`) using the P-Square algorithm for **O(1)-memory median estimation**.
- Architected an ephemeral `.awk` script execution layer to eliminate shell-escaping vulnerabilities; resolved Windows' **8,191-character buffer limit** with dynamic sub-batch chunking; managed the CRAN release.

Open Source Contributor, R Project *(Google Summer of Code)* May 2025 – Aug 2025

- **DBmaps** (R package, on CRAN, 5,000+ downloads): Automates joining of relational tables with differing granularities, so analysts can combine complex databases without manual key matching or duplicated rows.
- Enforced localized pre-aggregation before merging to guarantee zero row inflation and eliminate Cartesian expansion.
- Built a declarative metadata architecture with a dual-discovery engine that infers hidden join paths.
- Validated on a 200,000-row TPC-H star schema: **5.88**× **median speedup** over `dplyr` (669.8 ms vs 3,940.1 ms).

PROJECTS

rPandas *(CRAN, 1,000+ downloads)* R, Python, reticulate

- Built an R-to-Python translation engine using non-standard evaluation, letting users write pandas operations (`rp_filter`, `rp_select`, `rp_mutate`) in native R syntax.
- Translates commands to pandas and executes them via `reticulate`; supports chaining, grouping, and summarisation, and generates copy-pasteable Python code.

grepreaper *(CRAN, 2,000+ downloads)* R, grep, Shell

- R package that translates user input into `grep` commands, executes them, and returns the output.

EnvGuard Go, Cobra, Bubble Tea

- Built a Linux CLI that checks whether essential developer tools (e.g., Git, Docker, Go) are installed and configured, and installs missing ones through package managers such as `dnf`.
- Separated command orchestration from tool-check logic so new checks plug in easily; built an interactive Bubble Tea terminal UI with live status updates, structured error reporting, and guided remediation.

Sentiment Analyzer React, Node.js, Python

- Full-stack sentiment analysis app: React dashboard and Node.js REST API over a Python NLP pipeline (TF-IDF, tuned RBF-kernel SVM) reaching **91% validation accuracy** on 3 classes (5-fold cross-validation).

EDUCATION

B.Tech, Computer Science & Engineering Expected May 2027
Jawaharlal Nehru University, New Delhi
Relevant Coursework: OOPs, Operating Systems, DBMS, Design & Analysis of Algorithms

SKILLS

Languages: C/C++, Go, R, Python, JavaScript, Bash/POSIX Shell
Backend & Web: Node.js, Express, React, REST APIs, MongoDB, MySQL
Tools & Libraries: Linux, Docker, Git, GitHub Actions, testthat, roxygen2, Scikit-learn