DP-accurate DAU/MAU Counter Under Deletions

Dev Sanghvi   Lazeen Manasia

October 9, 2025
580-Probability Algorithms and Data Structures

Abstract

We built a small, end-to-end system that reports Daily Active Users (DAU) and 30-day Monthly Active Users (MAU) while protecting people's privacy and honoring "please delete me" requests. The service accepts simple JSON events in real time. To keep identities safe, it scrambles user IDs using a secret key, and that key is rotated on a schedule so long-lived identifiers are harder to trace.

For each day, we keep tiny summaries that let us estimate how many unique users showed up. We provide both an exact version (good for testing) and compact, fast versions that trade a little accuracy for speed and cost; operators can switch between them by setting {{SKETCH_IMPL}}. If someone asks to be erased, we record it in a lightweight database and automatically rebuild the affected days so future DAU/MAU numbers no longer count that person.

To add differential privacy, we mix in a small amount of carefully tuned noise (Laplace or Gaussian) to each reported count. We also set a clear, documented bound {{W_BOUND}} on how often one person can appear, and we track a simple monthly privacy budget. When that budget would be exceeded, the system refuses the request and explains why.

The interface is straightforward: a web API for sending events and asking for DAU/MAU, plus a command-line tool for chores like loading datasets, resetting budgets, and practicing key rotations. Evaluation scripts generate realistic and stress-test workloads, compare accuracy versus privacy settings, and export easy-to-read plots; a short Jupyter notebook ties results to plain-language takeaways.

Operationally, we include automated checks, tests, and container images to make local runs and deployment easy. The documentation (README, HANDOFF, and AGENTS) outlines day-to-day operations, how to extend the system, and how to move to larger databases and queues when traffic grows. In short, the design balances sound privacy with practical engineering so teams can publish reliable, privacy-aware user metrics.