Hints for Assignment 1 — Log Analytics & Backup Automation

Use this guide to navigate through parsing logs, creating reports, rotating archives, and scheduling automation.

1. Parsing Logs

1. Understand the structure of log lines. Use `awk` or `cut` to extract IPs, status codes, and paths.
2. For counting unique IPs: `awk '{print $1}' access.log | sort | uniq | wc -l`.
3. To get top 5 IPs: `awk '{print $1}' access.log | sort | uniq -c | sort -nr | head -5`.
4. Extract request paths between quotes using `awk -F"" '{print $2}'`.

2. Generating Reports

1. Redirect all your calculated outputs into a CSV file using `>>` redirection.
2. For timestamps, use `$(date '+%Y-%m-%d_%H-%M-%S')`.
3. Add column headers in CSV to make it structured.

3. Archiving and Checksums

1. Use `tar -czf` to compress logs and `sha256sum` to create integrity files.
2. Remove older archives beyond 7 using: `ls -t archive/*.tar.gz | tail -n +8 | xargs rm -f`.

4. Scheduling

1. Schedule the daily run using `crontab -e` → `55 23 * * * bash ~/loglab/run_daily.sh`.
2. Ensure you use absolute paths because cron runs in limited environments.

5. Error Handling

1. Start scripts with `set -euo pipefail` to handle unexpected errors.
2. Always validate input files and directories before using them.