

## General Guidelines

**Report format:** SIGCONF two-column format, 4 pages (excluding references). You must use the latex template provided here

<https://www.overleaf.com/read/hvcsdsxqkty#bbba90>

Any changes to the templates, including but not limited to font size, margins, document class parameters etc. will lead to penalties. You may use latex packages as convenient though.

## Expectations (Benchmarking Project)

- **Understand the system:** Demonstrate a clear grasp of the chosen PostgreSQL extension/feature: architecture, key primitives, and typical use-cases.
- **Working prototype:** Deliver working code/scripts that set up the system, load data, and run the experiments on a laptop (or in a provided Docker image). **The repo must be runnable end-to-end.**
- **Baseline & variations:** Include at least one sensible baseline (e.g., vanilla Postgres, JSONB vs hstore, exact COUNT vs HLL) and vary  $\geq 1$  important parameter(s) (e.g., data size, index type, HLL precision, chunk size, etc.).
- **Reproducibility:** Provide exact install commands, SQL snippets, and scripts. Any non-trivial external dependency (APIs, models) must be documented with instructions or a local fallback.
- **Measurement rigor:** Run each experiment multiple times ( $\geq 5$ ), report central tendency and variability (median/mean  $\pm$  std or CI), and explain measurement methodology (what you measured and how).
- **Honest limitations:** Record failures, anomalies, and practical issues (e.g., long runtime, extension bugs, memory limits) — these are important findings.

## Expectations (E&A report)

- **Title & Authors**
- **Introduction** (motivation and one-paragraph statement of objectives, short summary of goals, method, and key takeaways.).
- **Background & System Overview** (concise description of the chosen PostgreSQL feature/extension and how it works). Include architecture diagram if useful (small figure).
- **Methodology**

- Datasets used (sources, sizes, brief characteristics).
- Experimental setup (hardware: laptop spec(s), PostgreSQL version, extension versions, OS).
- Workloads / queries / operations executed (explicit list or representative ones).
- Metrics collected (latency, throughput, memory, storage size, accuracy/error where applicable).
- Reproducibility notes (scripts, seed values).
- **Results**
  - Present key graphs/tables (use clear captions). Every result must explicitly identify experimental parameters.
  - Include at least one plot showing scaling (data size or workload intensity) vs one key metric.
  - For approximate methods (HLL, vectors, learned indexes), include accuracy/error plots (e.g., error vs memory or latency vs recall).
- **Discussion**
  - Interpret results, highlight trade-offs, practical considerations (ease of use, integration pain points).
  - Explain surprising results or anomalies.
- **Limitations & Future Work** (short).
- **Conclusion** (one short paragraph).
- **References** (not counted toward 4 pages).

## **Deliverables**

- **4-page E&A paper** (PDF) following sections above (upload to Moodle).
- **Repro package** (GitHub repo) containing:
  - README.md with exact install commands and how to run experiments.
  - Scripts for data loading, experiment execution, and plotting.
  - Any generated datasets used or links/instructions to download (small sample if large).
  - A manifest file with environment details: OS, CPU, RAM, PostgreSQL version, extension versions, Python versions, Docker image ID if used.
  - **Your report should include the link to your Github repo.**

### Experimental Checklist (must be addressed in your E&A report)

- State laptop hardware used (CPU model, cores, RAM, disk type).
- State Postgres and extension versions and exact commands used to enable them.
- For each experiment:
  - Identify the workload (SQL statements, parameters).
  - Report number of runs and statistical measure (median/mean + confidence interval) — run each experiment  $\geq 5$  times and report median or mean+std.
- For measurement:
  - Report wall-clock latency, throughput (ops/sec) where relevant, memory and disk usage.
  - For approximate methods: report error metrics (e.g., relative error, recall@k, precision@k, F1 as appropriate).
- Parameter sweep: vary at least **one** important parameter (e.g., HLL precision, vector index lists, Timescale chunk size, index type, query concurrency) and show its effect.
- Figures & Tables
  - Include at least 2 high-information figures (e.g., a throughput/latency plot and an accuracy/size trade-off plot) and 1 table summarising environment and dataset sizes. Keep figures readable in two-column layout.

### Evaluation Rubric

- **Correctness & completeness of setup** (20%) — working scripts and clear environment manifest, working code, reproducible experiments.
- **Clarity of methodology** (15%) — clear workloads, datasets, metrics.
- **Quality of experiments** (40%) — appropriate baselines, parameter sweeps, proper statistics.
- **Analysis & insight** (25%) — interpretation of results and practical takeaways.

**You may use AI-assisted tools for your research, but the paper must be written by you. We will use professional tools to detect AI-generated text and plagiarism. Any violations will result in failing the assignment and will incur additional penalties.**